tøi : **Institute of Transport Economics**
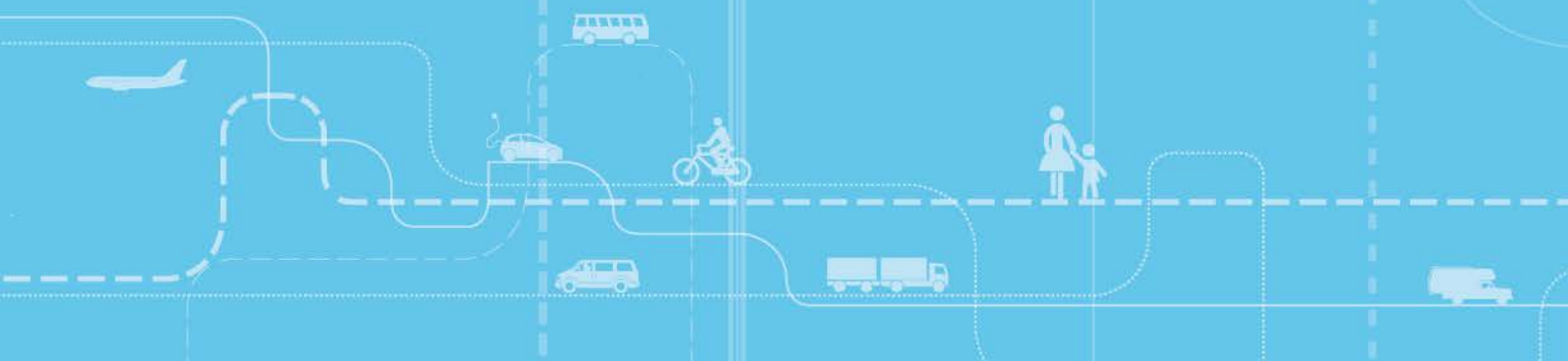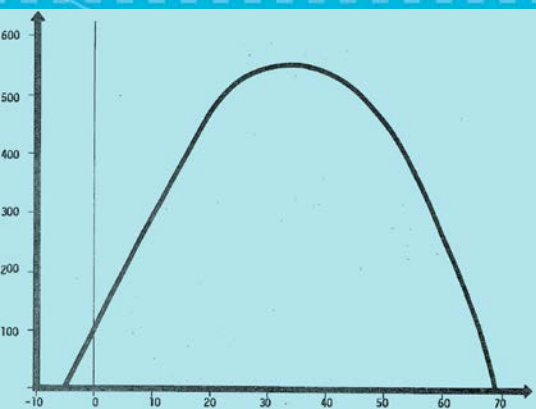**Norwegian Centre for Transport Research**

The Value of Life

The Rise and Fall of a Scientific Research Programme

The Value of Life

The Rise and Fall of a Scientific Research Programme

Rune Elvik

Sammendrag:

Rapporten gir en historisk og vitenskapsfilosofisk rekonstruksjon av forskning om økonomisk verdsetting av trafikksikkerhet ved å bygge på Imre Lakatos' teori om vitenskapelige forskningsprogrammer. Denne teorien viser seg å egne seg godt til en rekonstruksjon av verdsettingsforskningen. Den kan bidra til å forklare at forskning videreføres selv om den gir svært sprikende resultater og resultater som strider mot teorigrunnlaget for forskningen. Teorigrunnlaget er etter hvert reformulert, slik at noen resultater som før var i strid med det ikke lenger er det.

Summary:

The report reconstructs the history of research on the monetary valuation of road safety by applying the methodology of scientific research programmes as developed by Imre Lakatos. This theory of science is found to be highly applicable to the history of valuation research. It can explain why research continues in a field, even if findings are very widely dispersed and to a large extent contradict the theoretical foundations of empirical research. The theoretical foundations have been partly reformulated, so that findings that original contradicted theory no longer are seen as doing so.

Preface

This report presents a historic reconstruction of research on the monetary valuation of road safety, based on the methodology of scientific research programmes, as developed by Imre Lakatos.

Modern research designed to obtain a monetary valuation of the benefits to society of improving road safety started around 1970 when some prominent economists, notably Thomas Schelling and Ezra Mishan, called for basing this valuation on the willingness-to-pay approach. A large number of studies of willingness-to-pay have since been made. These studies have produced very diverse findings, many of which have for a long time been regarded as difficult to reconcile with the theoretical foundations of valuation studies.

The study presented in this report has been funded by the Research Council of Norway, as part of the TRANSIKK-research programme on transport safety. Rune Elvik has been project manager and is the author of this report. Valuable comments on drafts of the report have been given by (alphabetically): Peter Christensen, Beate Elvebakk, Ezra Hauer, Alena Høye, Gunnar Lindberg, Marika Kolbenstvedt and Knut Veisten. The author would like to thank for the comments, which greatly improved the quality of the report. The responsibility for any remaining errors rests with the author.

Research director Michael W. J. Sørensen has been responsible for formal quality control of the report. Secretary Trude Kvalsvik has edited the report and prepared it for electronic and paper publication.

Oslo, December 2016
Institute of Transport Economics

Gunnar Lindberg   *Michael W. J. Sørensen*
Managing Director   *Research Director*

Content

List of Abbreviations

CV = Contingent valuation; a method for eliciting monetary valuations by means of direct questions ("how much are you willing to pay for …?").

GDP = Gross Domestic Product; i.e. the monetary value of all goods and services produced in a country during a specific period, usually one year.

QALY = Quality Adjusted Life Year, a numerical scale for rating health-related quality of life. Death has the value of 0, perfect health the value of 1.

SC = Stated choice; a method for eliciting monetary valuations by asking people to make choices between alternatives that are characterised by certain attributes, one of which is monetary.

SE = Standard error; an indicator of the statistical uncertainty of an estimate.

VSL = Monetary value of a statistical life, i.e. a reduction in risk corresponding to the prevention of one fatality.

WTA = Willingness-to-accept, i.e. the smallest sum of money needed to compensate for a loss, for example an increase in the risk of a fatal injury.

WTP = Willingness-to-pay, i.e. the largest sum of money a person is willing to pay for a good, for example a reduction in the risk of a fatal injury.

Summary

The Value of Life – The Rise and Fall of a Scientific Research Programme

Research designed to obtain a monetary valuation of life and limb has long traditions. The value of saving a life was originally estimated as the capitalised value of output lost as a result of a premature death. This approach was abandoned around 1970 and a new theoretical foundation for the monetary valuation of life and limb was proposed: the willingness-to-pay (WTP) approach. Since that time, a large number of studies of willingness-to-pay for reduced risk of death have been made. This report reconstructs the history of this research by applying the methodology of scientific research programmes, developed by Imre Lakatos. The methodology of scientific research programmes can explain why a field of research can continue to exist despite that fact that many of its findings are difficult to make sense of and diverge enormously. Monetary valuation of life and limb fits this description. Estimates of the value of preventing a death, often referred to as the value of a statistical life, vary enormously. Some of this variation can be explained according to economic theory, but quite a lot cannot. One response to this state of affairs has been a reformulation of relevant parts of economic theory which allows for re-interpreting findings that were initially regarded as anomalous, but become theoretically plausible when re-interpreted. The reformulation of theory has almost gone to the point of making any finding theoretically plausible. It seems clear the valuation research based on willingness-to-pay originally had a goal of finding a single, uniform vale of a statistical life. Today, this objective has been given up and valuation research appears to live happily with the enormous diversity in estimates of the value of a statistical life.

Starting point: The values are all over the place

The starting point for the study is the observation that estimates of the value of a statistical life are all over the place. The value of a statistical life is the monetary value of a reduction in the risk of death, which statistically corresponds to the prevention of one death. Many studies have been made to estimate the value of a statistical life. The results of these studies vary enormously, from less than 5,000 US dollars to close to 200 million US dollars.

One might think that a field of research producing such diverse estimates, all of which are intended to measure the same thing, would be abandoned. However, valuation research continues to prosper and new studies are published quite frequently. This forms the background of the first main research problem of this study:

How can one explain that a field of research, producing enormously diverse results, some of which appear to contradict the theoretical foundations of the research, continues almost as if the contradictory results did not exist?

There exists a theory of science which seems well-suited to explaining the apparent paradox of valuation research. That is the methodology of scientific research programmes, developed by philosopher Imre Lakatos. Hence, the second main research problem is:

Can the methodology of scientific research programmes help to better understand, and possibly explain, the historic development of research on the monetary valuation of life and limb, in particular valuation of improving road safety?

Before summarising the answers to these questions, the context of the study will be briefly explained.

Do we need a monetary valuation of life and limb?

Many people find the very idea of assigning a monetary value to life or health objectionable, or at least strange. Many will ask: Do we really need these monetary valuations and what are their principal uses?

In this report, it is regarded as a basic and self-evident fact that trade-offs between different goods and objectives are made, and have to be made. Whenever an individual or government makes a decision about how much to spend on activities or measures that reduce the risk of death or injury, a trade-off is made between this good and other goods. It is impossible to avoid making trade-offs, simply because the resources at our disposal are limited. Therefore, the idea that life has an infinite value makes no sense.

However, it does not follow from this that a monetary value must by necessity be assigned to human life and health. It is entirely possible to avoid doing so and still make intelligent trade-offs between human life and health and other goods. Priorities between measures intended to reduce mortality or improve health can be set according to cost-effectiveness. The less a measure costs per fatality prevented, the higher should be its priority. The applicability of cost-effectiveness analysis is highly limited. Cost-effectiveness is undefined if a measure has an effect both on fatalities and injuries. Moreover, cost-effectiveness does not tell us when a measure is too expensive.

By explicitly assigning a monetary value to life and health, one may in principle:

1. Make trade-offs between safety and other policy objectives, like mobility or environmental protection – provided these objectives are also stated in monetary terms.
2. Determine if a measure improves societal welfare, which means that its benefits are greater than the costs, so that compensation preventing anyone from a net loss is in principle possible. It is assumed that the monetary valuation of non-market goods reflects their impacts on welfare.
3. Find the optimal use of policy instruments, for example road safety measures. The use of a set of measures is optimal when net benefits are maximised. In principle, this guideline can also be used to determine the size of a budget (it should be exactly large enough to cover the costs of all measures whose marginal benefits are equal to or greater than marginal costs).

These points indicate ways in which policy making can be informed by monetary valuation, but not without monetary valuation.

The definition, measurement and valuation of risk

Monetary valuation of life and limb refers to the valuation of changes in the risk of dying or of sustaining an injury. Most valuation studies state the risk of dying as a population mean fatality rate for a specific cause of death, for example that the current mortality rate in road accidents in Norway is close to 3 per 100,000 inhabitants per year.

This is the current population average for Norway. Risk varies in the population, by gender, by age groups, by place of residence and according to how much one travels by road. A population average may therefore be misleading as an estimate of the risk referring to a certain group of the population or a specific individual. Some valuation studies have therefore asked respondents about their subjective estimates of risk, i.e. what they think their own risk is. The majority of valuation studies has relied on statistical estimates of risk based on official statistics, applying either to the entire population of a country or a group of the population.

The good being valued in a valuation study is a change in risk, for example a reduction of fatality rate by 2 in 100,000. If this reduction is valued at 500 NOK, the value of a statistical life is estimated as follows:

Value of a statistical life (VSL) $= \dfrac{500}{\left(\frac{2}{100000}\right)} = 25,000,000$ NOK

The value of a statistical life is the value of a risk reduction which statistically corresponds to reduction of the number of fatalities by one.

The methodology of scientific research programmes

The first estimates of the value of preventing a road accident fatality were made in the 1950s in Great Britain, Sweden and the United States. These estimates were based on the so called human capital approach. The basic idea of the human capital approach was to estimate the capital value of a human being. This value was usually estimated as the present value of future earnings.

The method had some glaring deficiencies. Children and the retired had no value, since they did not earn anything. Housewives doing unpaid household work also had a value of zero, or, in some studies, a negative value since their consumption had to be supported by others. Besides, the method had no theoretical foundation.

Valuation of life and health as a scientific research programme started by defining the theoretical basis of valuation research. This was done around 1970 by economists Schelling, Mishan and Jones-Lee. They all argued that the only theoretically meaningful approach to the valuation of non-market goods, like life and limb, is the willingness-to-pay approach. The value of any good is indicated by the amount an individual is willing to pay for the good. The more valuable we think something is, the more we are willing to pay for it. Valuation in terms of willingness-to-pay is individual and subjective. No objectively "correct" valuation exists.

But how can we find out whether or not we can trust the results of valuation studies if there is no correct answer to the question asked in these studies? Do we simply have to accept any amount people state? No, as explained later, there are many ways of assessing whether the results of valuation studies can be taken seriously or have to be rejected.

Schelling, Mishan and Jones-Lee placed the valuation of life and limb squarely within consumer theory. This is well-developed branch of economic theory with a long research history. Consumer theory basically seeks to explain how people allocate their spending between different commodities and services. It does so by relying on the assumption that people choose the pattern of consumption that gives them the greatest overall satisfaction. Another way of saying this is that consumers are assumed to maximise utility (the term utility denotes the satisfaction of preferences; to put it colloquially: to maximise utility is to do what you like best).

The assumption of utility maximisation is the core of consumer theory. This theory has the same characteristics as a scientific research programme as defined by Imre Lakatos. Lakatos developed the methodology of scientific research programmes principally as a descriptive theory of how research actually takes place, intended as a conceptual framework for what he called the "rational reconstruction" of the history of science. By rational reconstruction he meant a historical reconstruction of science as the product of rational choices made by researchers working in a field of knowledge. He proposed that the rationality of the choices made by researchers should be judged according to standards defined by the researchers themselves, in particular the standards he labelled "positive heuristic" and "negative heuristic" (see explanation of these terms below).

According to Lakatos, a scientific research programme consists of a hard core, a protective belt, a positive heuristic and a negative heuristic. A programme may, at any point in time, be in a progressive phase or a degenerative phase. The hard core of a scientific research programme consists of basic assumptions made by all researchers working within the programme. The basic assumptions are taken for granted; it is forbidden to raise doubts about them. The basic assumptions are not tested empirically. The hard core is surrounded by a protective belt. The protective belt consists of hypotheses that are derived from the hard core. Hypotheses in the protective belt are tested empirically and can be rejected. Rejection of a hypothesis in the protective belt is in general regarded as undesirable, since any hypothesis in the protective belt is formulated by means of deductive reasoning based on hard core assumptions. Logically speaking, therefore, rejection of a hypothesis in the protective belt casts doubt on the validity of the assumptions forming the hard core. Any finding that, taken at face value, casts doubt on the hard core is called an anomaly. Anomalies, Lakatos argues, do not normally lead to the rejection of the hard core of a research programme. On the contrary, the hard core is normally left intact and research continues as if the anomalies did not exist.

There are usually many interpretations of an anomaly. The positive heuristic of a research programme calls for researchers to increase the empirical content of the programme and, in particular, to develop hypotheses or research methods that will eliminate anomalies. The empirical content of a research programme consists of all observations implied by the hypotheses forming the protective belt, both observations that have been confirmed and observations not yet made. A programme is in a progressive phase when its empirical content increases. A programme enters a degenerative phase when its empirical content no longer increases and when anomalies come to be the normal finding of empirical research. One sign of a degenerative phase is that the anomalies are explained by means of ad hoc hypotheses only, i.e. hypotheses that explain a single anomalous finding, but have no implications predicting novel findings. The negative heuristic calls on researchers not to question the hard core and not to develop ad hoc hypotheses to explain anomalies.

What happens when anomalies become very many? At some point a research programme may be abandoned, but according to Lakatos this does not take the form of a scientific revolution as suggested by Thomas Kuhn. He argues that findings contradicting a theory are, by themselves, not enough to reject the theory. A theory is only rejected when a new

and better theory has been developed; better, in the sense that it explains all verified content of a research programme as well as (at least most of) the anomalous findings of that programme. In other words: anomalies cease to be anomalies if a theory is developed that re-interprets them as normal findings.

Application of the methodology of scientific research programmes to valuation research

The methodology of scientific research programmes is highly applicable when trying to reconstruct the history of valuation research.

The hard core of this research is the assumption made in consumer theory that consumers are rational utility maximisers. This assumptions is purely formal; it has no empirical content and merely states that consumer choices can always be modelled as a utility function which is maximised. All one needs to assume to apply this basic postulate, is that preferences can be represented by means of a utility function which has the mathematical properties necessary for a maximum to exist. These properties are very weak and innocent-sounding, boiling down essentially to the requirements that preferences should be transitive and complete. Researchers are at great liberty to make further assumptions about individual utility functions. Thus, as an example, Jones-Lee (1974) made the following assumptions:

1. The individual maximises expected utility (which is a probability weighted utility of a lottery with life and death as potential outcomes).
2. The individual prefers more wealth to less and is financially risk averse (prefers an income received with certainty to an uncertain income).
3. The individual does not want descendants to be exposed to a greater financial risk than himself or herself.
4. At a given level of wealth, the individual prefers to be alive rather than dead.
5. The marginal utility of wealth is greater when the individual is alive than when the individual is dead.

Based on these assumptions, Jones-Lee could deduce that a positive willingness-to-pay for reduced risk of death will exist. He further deduced that willingness-to-pay will be positively related to income and positively related to the level of risk. This example shows how one can use theoretical predictions to assess whether empirical results make sense or not. If you find that willingness-to-pay varies systematically as predicted by theory, results make sense. If you do not find the predicted pattern, interpretation becomes more complicated. It could be that your theory is wrong, but it could also be that the methods were not good enough to uncover the expected pattern. At any rate, this example shows the essential function of theory in willingness-to-pay research: It is to predict a systematic pattern of variation in willingness-to-pay that may serve as reference in assessing whether the results of empirical studies make sense or not.

An important positive heuristic in the early days of valuation research was to develop theory to help identify meaningful patterns of variation in willingness-to-pay. Important theoretical contribution were made around 1980. However, empirical research did not always give results that supported the hypotheses. This was widely interpreted as a problem of research method. Another important positive heuristic in the early days of valuation research was therefore to continuously develop and improve methods for valuation studies.

The progressive phase (1970-1995)

Following the definition of the theoretical foundation of valuation research around 1970, some years passed before the first empirical studies got started. From about 1980, empirical studies were made both in Europe and the United States. Different research traditions developed on the two continents. In Europe, the contingent valuation method was used in all studies until the late 1990s. This method elicits willingness-to-pay by asking respondents to state how much they are willing to pay for a certain change in risk. Several versions of the method have been developed. In the United States, the dominant method was studies of compensating wage differentials. These studies estimated the additional wages paid to compensate for work-place risks.

The first major valuation study in Europe was performed by Jones-Lee and others in Great Britain. The study was originally reported in 1983, but has been published a number of times, including as a chapter in a book by Jones-Lee published in 1989. When the study was in progress, Jones-Lee convened an international research conference in Geneva in 1981. The conference was attended by the leading researchers at the time and several major theoretical contributions were presented during the conference. These theoretical contributions increased the empirical content of valuation research (i.e. they predicted findings of empirical studies).

The empirical content of valuation research increased further as a result of a number of empirical studies. Replications, or near-replications, of the British study were made in Austria, Sweden, New Zealand, Denmark, Switzerland and France. All these studies were reported before 1995. Their findings were, mostly, not very different from the original British study.

The beginning of the end of the progressive phase was an increasing number of anomalous results. These results raised doubts about the validity of contingent valuation estimates of the value of a statistical life. In the United States, there was always a greater scepticism to contingent valuation than in Europe, and in 1993 a critical assessment of the method was published, providing methodological guidelines on how to conduct good contingent valuation studies. A common problem was insensitivity to scope. Insensitivity to scope means that willingness-to-pay does not increase in proportion, or near proportion, to the size of the risk reduction. Thus, it was typically found that people were not willing to pay twice as much for a risk reduction of 4 in 100,000 as for a risk reduction of 2 in 100,000. It was believed that one source of the problem was that people had difficulty in understanding small changes in low levels of risk. The numbers 4 in 100,000 and 2 in 100,000 do not seem to be very different – both are very low numbers.

A French study tried to get around this problem by asking for willingness-to-pay for reductions in the number of traffic fatalities in France ranging from 50 to 5000. The idea was that people would more easily notice the difference between these numbers than the differences between low levels of risk. However, the French were not willing to pay 100 times more for reducing the number of traffic fatalities by 5000 than for reducing them by 50. In fact, they were, on the average, only willing to pay slightly more than 4 times more for reducing fatalities by 5000 than for reducing them by 50. Thus, the problem remained unsolved.

Moreover, it was found that different methods for eliciting willingness-to-pay in contingent valuation studies produced different results, although the methods were equivalent according to economic theory. Things that ought not to make a difference according to economic theory did in fact make a difference. Studies of compensating wage differentials gradually improved, as both new sources of data became available and statistical modelling became more advanced. Yet this progress was not associated with more consistent

findings. On the contrary, the range of estimates became bigger, despite progress with respect to data quality and statistical analysis.

By the end of the 1990s, the progressive phase was over. Valuation research entered a new phase.

The struggle between progressive and degenerative tendencies (1995-2005)

The results of contingent valuation studies were, by the end of the 1990s, full of anomalies. Some researchers went so far as to say that all results of such studies were anomalies. Jones-Lee, who had championed contingent valuation in the early phase of valuation research, rejected the method in 1998 and proposed a new method bypassing the need to ask people about changes in low levels of risk.

From about 2000, a number of valuation studies started to use the stated choice design. Respondents were asked to choose between two alternatives, in most cases two roads. The roads differed with respect to, for example, travel time, number of accidents and toll charges. Respondents were asked to choose which road to take; the attributes of the two roads were then modified and respondents asked to choose once more. Each respondent would typically make 5-10 choices. The valuation of the non-monetary attributes was obtained by analysing the choices made, usually relying on random-utility functions (i.e. functions respondents were assumed to maximise, but that would have a residual terms since analysts did not know all factors influencing utility).

In the United States, Peter Dorman launched a strong criticism of studies of compensating wage differentials in 1996. He argued that all these studies were flawed and should be rejected. He concluded that one should abandon monetary valuation of life and limb entirely. The leading proponent of wage-risk studies in the United States, Kip Viscusi, dismissed Dorman's criticism. Many studies of compensating wage differentials have been published since 1996. Dorman was not successful in his attempt to bring this research to an end. He may well have been right in much of his criticism, but, as Lakatos pointed out; criticism per se is rarely enough to overturn a scientific research programme. You have to offer something better. Dorman did not offer any alternative. He simply said: "Stop doing this", without saying what one should rather do.

Stated choices were initially thought to be superior to contingent valuation, for example because people were not asked about changes in low levels of risk. However, anomalies soon turned up in stated choice experiments. Lexicographic choices were common. An individual chooses lexicographically if he or she always prefers the alternative that is best with respect to one of the attributes, and ignores the others (for example, always chooses the safest road). Many respondents made inconsistent choices. This means that at stage N in a sequence of choices, they preferred an alternative implying a valuation which was inconsistent with a choice made at stage N – 1 of the sequence. Only about 10-20 percent of respondents made choices that were fully consistent with economic theory.

Research was a struggle between progressive and degenerative tendencies. The launching of new methods represented the progressive element; the repeated finding of anomalies represented the degenerative element. The anomalies did not go away, they merely took new forms.

Meanwhile, the theoretical foundation of valuation research was undergoing a rapid transformation.

The protective belt becomes almost all-inclusive (2000-2010)

The early theoretical contributions to valuation research made clear predictions about empirical results. As an example, Jones-Lee predicted that willingness-to-pay would increase as the level of risk increased. Finding the opposite would falsify his theory. It did not take long, however, before more complex models were developed and predictions became ambiguous. The extremely complex model proposed by Dehez and Drèze is an example. Here are the predictions of this model:

1. If an individual does not have life insurance or an annuity, and if the marginal utility of money is greater when alive than when dead, willingness to pay will increase when risk level increases.
2. If the individual has optimal life insurance and annuity at actuarially fair rates, willingness to pay is independent of the level of risk.
3. If the individual holds life insurance and annuity at less than actuarially fair rates, willingness to pay will increase as the level of risk goes down.
4. If the individual holds life insurance and annuity at more than actuarially fair rates, willingness to pay will increase as risk level increases.
5. If the individual has life insurance and annuity and the terms of the contracts are adjusted as risk level changes, willingness to pay will increase as risk level decreases.

One could say that they hedge their bets. Everything is possible; that willingness to pay does not depend on risk level, that it increases with risk level, or that it decreases with risk level. None of these findings is ruled out theoretically. No matter what you find, it has theoretical support – unless, that is, that you can collect detailed data on the insurance coverage of respondents. But even if such data are available, it may be difficult to determine if insurance is actuarially fair or not. One would normally expect insurance to be less than actuarially fair, but in some countries tax rebates for life insurance may make insurance contracts close to actuarially fair.

In short: It is in practice almost impossible to falsify the hypotheses proposed by Dehez and Drèze. Their contribution was the start of a series of theoretical contributions that have reached the point where almost no finding contradicts theory. Perhaps the most consequential contribution is the directionally bounded utility function introduced by Amiran and Hagen (2003, 2010). This utility function predicts insensitivity to scope, turning what was long regarded as a major anomaly in valuation research into a theoretically expected finding, perfectly consistent with rational utility maximisation. This has revolutionary implications. If insensitivity to scope is to be expected, all the efforts that have made to develop methods to increase sensitivity to scope look like a complete waste of time. Almost any finding must be taken seriously if one takes directionally bounded utility functions seriously.

One could, to be sure, try to ascertain whether respondents do indeed have a directionally bounded utility function or a utility function which is not directionally bounded. However, it is by no means clear how to do this, and a wide range of choices can be consistent with a wide range of mathematical forms of a utility function. It seems quite likely that the data that could realistically be obtained would be inconclusive.

Today, theory has come close to an immunising stratagem. An immunising stratagem, a concept introduced by Karl Popper, is a reformulation of a scientific theory so as to make it immune to falsification. Popper would say that such a reformulation makes the theory unscientific, since he regarded only theories that could be falsified as truly scientific. What seems clear, is that bringing the theory close to an immunising stratagem undermines its function in research. If no result can be ruled out on theoretical grounds, it is no longer

possible to refer to theory to support or reject an empirical funding. Theory no longer discriminates between meaningful and meaningless findings.

A hard core in dissolution? (2005-2015)

There is little doubt that the original ambition of valuation research was to find a single value of a statistical life that could be applied uniformly. The many references to allocative efficiency made in early contributions attest to the importance given to this objective. Allocative efficiency can only be attained if the same value of a statistical life is used in all sectors of society. If one allows much more to be spent on saving life in one sector than in another, one may in principle save a larger number of lives by transferring spending from the "expensive" sector to the "cheap" sector.

As research has produces an ever widening gap in estimates of the value of a statistical life, the objective of finding a single value that can be applied universally has been toned down. Many researchers have argued that a single value does not exist. We know, for example, that age and income are likely to influence willingness to pay. In recent years, some researchers have started to ask whether one should allow the value of a statistical life to vary, rather than using a single, uniform value. The issue then becomes how much variation to allow for and what sources of variation would be legitimate. Both Jones-Lee and Viscusi have argued that income is legitimate: it is entirely appropriate to treat rich people's lives as more valuable than poor people's lives. They both argue that doing so is more consistent with the theoretical foundation of cost-benefit analysis than using a single, uniform value of a statistical life.

At this point a full circle has been travelled. Research set out to find a single, universally applicable value of a statistical life. It was quickly realised that such a value does not exist. Hypotheses were developed to predict systematic variation in valuation. The underlying idea was not necessarily that the value of a statistical life should also vary; it was rather that if one found variation making sense according to economic theory, more trust could be placed in findings than if one did not find such a variation.

The expected variation was only partly found. Besides, a lot of variation attributable to sources that ought to be irrelevant according to economic theory was found. As this continued in study after study, researchers turned their attention to theory once more and embarked on reformulating it to enable the anomalies to be interpreted as normal findings. This has been so successful that probably quite few findings would now be regarded as anomalous.

The huge variation in estimates of the value of a statistical life was thus transformed from a problem into something which is to be expected. Doubts have been raised about the prescriptive ideal of applying a uniform value of a statistical life in cost-benefit analyses.

Today, therefore, little remains of the original research programme that inspired valuation research. That programme has travelled full circle and come to an end in the sense that the hope of finding a single value of a statistical life has been given up and even the ideal of allocative efficiency, as traditionally understood, is being questioned.

We will nevertheless ask: Where do we go from here? What are the prospects for valuation research?

Can meta-analysis create order in chaos? (2000-2015)

A number of meta-analyses of value of life studies have been reported during the last fifteen years. The report presents these analyses and discusses whether they have been able to make sense of the huge diversity of estimates of the value of a statistical life.

Most of the meta-analyses are somewhat simplistic and do not fulfil the methodological standards for high-quality meta-analyses. The analyses do, for example, not apply optimal statistical weights, nor do most of them test for the possible presence of publication bias. Some of the meta-analyses have tried to sort primary studies according to study quality.

Norwegian researchers Ståle Navrud, Henrik Lindhjem and Nils Axel Bråthen, co-operating with the French economist Vincent Biausque, have made the most comprehensive meta-analysis reported so far. It has been published in several rounds; the most recent included 931 estimates of the value of a statistical life. Meta-regression analyses were performed in order to explain variation in estimates; these analyses were made both for the entire sample and for a subsample of the studies that were classified as methodologically best.

The factors that were found to have the largest influence on the value of a statistical life were income (higher income, higher value), size of the risk change (larger change, smaller value) and whether safety was provided by means of a public good or a private good (lower value for public goods). It may strike readers as surprising that the value of a statistical life was lower when the change in risk was large than when it was small. This is attributable to insensitivity to scope, as shown by the numerical example given below:

Risk reduction	Willingness-to-pay	Value of a statistical life
$1 \cdot 10^{-5}$	400	40,000,000
$2 \cdot 10^{-5}$	500	25,000,000
$5 \cdot 10^{-5}$	600	12,000,000
$10 \cdot 10^{-5}$	800	8,000,000

Considerable publication bias has been found in the value of life literature. Studies that have adjusted for publication bias indicate that by doing so, the value of a statistical life drops to about one third of the un-adjusted value (e.g. from 9 million to 3 million).

Although some of the meta-analyses explain most of the variation in the value of a statistical life, it is apparent that residual terms are very large, in particular at the ends of the distribution. Moreover, most meta-analyses do not recommend a best estimate of the value of a statistical life. The analyses also remain silent on the topic of whether a variable value of a statistical life should be applied.

Can more promising methods be found?

Valuation research has so far not been able to produce very precise estimates of the value of a statistical life. Even within the same study, estimates often vary be a factor of 10 or more. Can we think of other methods that might produce more precise estimates?

The report reviews a few options. One of them is to use Quality Adjusted Life Years (QALYs) as a starting point. A QALY is a numerical scale for quality of life related to health state. By convention, death is given the value of 0 and perfect health the value of 1. States of reduced health as assigned values between 0 and 1, closer to 0 the worse they are. If a health state involving a loss of, for example 0.02 QALYs is valued at 50,000, the idea is that one may scale this up to a value of a statistical life year of $1/0.02 \cdot 50,000 = 2,500,000$.

There are two big problems with this approach. In the first place, there is no universally accepted method for obtaining QALYs. On the contrary, quite many scales can be found in the literature and they do not agree. A specific health state will therefore not have the same QALY score according to all scales. This means that estimates of the value of a statistical life may differ substantially, depending on which QALY scale is used to scale up the value applying a slight reduction of health to the value of a statistical life. In the second place, QALYs rely on utility functions with very restrictive properties; indeed much more restrictive than the utility functions normally assumed in conventional value of life studies. The QALY approach is therefore not very promising.

Another approach discussed is the capability approach introduced by Amartya Sen. The core of this approach is that to obtain an acceptable quality of life and standard of living, an individual must possess certain capabilities, such as access to clean water, literacy, housing, and so on. Sen has proposed this approach as an alternative to subjective well-being, noting that even people who live in great hardship often report a high level of subjective well-being. The capability approach resembles the "social indicators movement", which proposed to measure how well a society is taking care of its citizens by using social indicators like literacy rates, access to water and toilet facilities, and so on. Indicators like these are probably best suited to low-income countries. They are completely inconsistent with the theoretical foundation of valuation research by being fully paternalistic. The capability approach is hence judged as irrelevant to valuation research.

The third approach discussed is to develop empirical utility functions based on surveys of subjective well-being (happiness surveys). Such surveys have been made in many countries. By studying the relationship between reported subjective well-being and, for example, income, functions can often be fitted that have the same characteristics as utility functions as usually defined in economic theory. Interpreting such empirical functions as empirical utility functions is, however, still controversial among economists, although the idea seems to be gaining increasing support. Another problem is that different data sets give rise to different functions, that may not imply the same valuation of a statistical life.

None of these three approaches would therefore seem to be clearly superior to the conventional designs used in valuation research so far.

Sammendrag

Verdien av liv - et forskningsprograms vekst og fall

Forskning som har til hensikt å finne en økonomisk verdi av liv og helse har lange tradisjoner. I begynnelsen ble et liv verdsatt på grunnlag av verdien av det en person kunne produsere. Omkring 1970 ble denne metoden forlatt og et nytt grunnlag for å verdsette liv og helse økonomisk foreslått: Betalings-villighetsmetoden. I tiden etterpå er det utført mange undersøkelser om betalingsvillighet for redusert risiko for å dø eller pådra seg helseskader. I denne rapporten rekonstrueres historien til betalingsvillighetsstudier ved å bygge på filosofen Imre Lakatos' teori om vitenskapelige forskningsprogrammer. Denne teorien forklarer hvordan en forskningstradisjon kan leve med et vell av uforklarlige resultater og resultater som tilsynelatende strider mot forskningens teorigrunnlag. Betalingsvillighetsforskningen svarer godt til en slik karakteristikk. Den har produsert enormt sprikende resultater. Forsøk på å forklare det enorme spriket viser at det bare delvis har sitt opphav i faktorer som ifølge økonomisk teori kan forventes å føre til variasjon i betalingsvillighet. Ett av svarene på dette har vært at deler av betalingsvillighetsteorien er reformulert, slik at resultater som opprinnelig stred mot den ikke lenger nødvendigvis gjør det. Reformulering av teorien har i dag gått så langt at nesten ethvert resultat er teoretisk mulig. Dermed er det store spriket i resultater av empiriske undersøkelser definert bort som et problem. Dette har ført til at betalingsvillighetsteorien i dag nærmest kan forklare ethvert resultat. På den måten har betalings-villighetsforskningen kommet til veis ende: Den begynte med et klart mål om å finne en bestemt verdi av et menneskeliv. Det viste seg etter hvert at det var umulig å finne en slik verdi. Teorien ble da reformulert slik at den i dag sier at en bestemt verdi av et menneskeliv ikke finnes. Det finnes uendelig mange forskjellige verdier og disse kan variere enormt.

Denne rapporten dokumenterer resultatene av prosjektet «En historisk og vitenskaps-filosofisk rekonstruksjon av forskning om økonomisk verdsetting av trafikksikkerhet ved hjelp av Imre Lakatos' teori om vitenskapelige forskningsprogrammer». Prosjektet er finansiert av Forskningsrådets TRANSIKK-program.

Utgangspunkt: Verdsettingene spriker voldsomt

Utgangspunktet for studien er at resultatene av undersøkelser om den økonomiske verdien av å forebygge et dødsfall, i litteraturen vanligvis omtalt som verdien av et statistisk liv, spriker voldsomt. Betegnelsen verdien av et statistisk liv henviser til en nedgang i risiko som statistisk sett tilsvarer at man unngår ett dødsfall. Det er utført mange undersøkelser med sikte på å finne verdien av et statistisk liv. Resultatene av disse undersøkelsene viser et enormt sprik, fra verdier på mindre enn 50.000 kroner til verdier opp mot 1 milliard kroner.

Man skulle kanskje tro at forskning med sikte på å anslå verdien av et statistisk liv ble gitt opp som et håpløst prosjekt når resultatene varierer så mye og ingen har klart å gi en særlig god forklaring på variasjonen. Men verdsettingsforskningen lever tvert om i beste velgående og resultater av nye studier publiseres hyppig. Dette danner bakgrunnen for den første hovedproblemstillingen i undersøkelsen:

Hvordan kan man forklare at forskning som gir voldsomt sprikende resultater, som til dels strider mot teorigrunnlaget for forskningen ikke oppgis, men blir videreført som om alle de unormale resultatene ikke fantes?

Det er utviklet en vitenskapsteori som kan bidra til å forklare at forskning som gir resultater som strider mot forskningens teorigrunnlag blir videreført som om disse resultatene ikke fantes. Det er teorien om vitenskapelige forskningsprogrammer, som er utviklet av Imre Lakatos. Studiens andre hovedproblemstilling er derfor:

Kan teorien om vitenskapelige forskningsprogrammer bidra til økt forståelse, og forklaring, av den historiske utviklingen av forskning om økonomisk verdsetting av liv og helse, spesielt knyttet til risikoen for trafikkulykker?

Før de to hovedproblemstillingene besvares, defineres studiens kontekst.

Trenger vi økonomisk verdsetting av liv og helse?

Mange finner selve tanken om økonomisk verdsetting av liv og helse forkastelig eller underlig og vil spørre om vi trenger en slik verdsetting og hva den i så fall skal brukes til.

Denne rapporten tar utgangspunktet i at avveininger mellom ulike mål og ønsker alltid må gjøres og blir gjort. Når individer eller myndigheter bestemmer hvor mye de vil bruke på aktiviteter eller tiltak som kan redusere risikoen for å dø eller bli skadet, gjør de en avveining mellom dette godet og andre goder. Det er ganske enkelt umulig å unngå å gjøre en slik avveining, rett og slett fordi de ressurser vi kan bruke på å fremme liv og helse er begrensede.

Av dette følger imidlertid ikke at liv og helse må verdsettes eksplisitt i økonomiske termer. Vi kan unngå å fastsette en økonomisk verdi på liv og helse. Tiltak kan prioriteres på grunnlag av kostnadseffektivitet, det vil si hvor mye de koster per liv de redder. Kostnads-effektiviteten til et tiltak blir imidlertid udefinert dersom tiltaket ikke bare virker på antall drepte, men også på skader av ulik alvorlighetsgrad. Videre sier ikke kostnadseffektivitet noe om hvor grensen går for når et tiltak blir for dyrt.

Med en eksplisitt økonomisk verdsetting av liv og helse kan man i prinsippet:

1. Veie sikkerhet for liv og helse mot andre goder, som kortere reisetid eller bedre miljø – forutsatt at også disse er verdsatt økonomisk.
2. Finne ut om bestemte tiltak bedrer velferden i samfunnet, det vil si gir en nytte som er større enn kostnadene, slik at alle i prinsipper kommer bedre ut. Det er her forutsatt at den økonomiske verdsettingen primært er et uttrykk for velferdsgevinster ved lavere risiko for å dø eller bli skadet.
3. Finne en optimal bruk av, for eksempel, trafikksikkerhetstiltak; det vil si den bruk av tiltakene som gir størst nytte for en gitt kostnad. I prinsippet kan en analyse med sikte på å finne optimal bruk av tiltak også benyttes som grunnlag for å bestemme budsjettets størrelse.

Disse tre punktene peker på hva man kan oppnå med en økonomisk verdsetting av liv og helse, men ikke oppnå uten en slik verdsetting.

Definisjon, måling og verdsetting av risiko

Det som verdsettes økonomisk i studier av verdsetting av liv og helse er endringer i risikoen for å dø i løpet av en gitt periode, eller endringer i risikoen for å pådra seg nærmere definerte skader. I de fleste verdsettingsstudier er risikoen for å dø angitt på

grunnlag av offentlig statistikk, for eksempel at det i gjennomsnitt dør om lag 3 mennesker per 100.000 innbyggere i trafikkulykker i Norge hvert år.

Dette er et gjennomsnittstall for alle innbyggere i Norge. Risikoen varierer mye, blant annet avhengig av kjønn og alder og hvordan og hvor mye man ferdes i vegtrafikk. En gjennomsnittlig risiko for hele befolkningen kan følgelig være misvisende for en bestemt gruppe eller et enkelt individ. I noen verdsettingsstudier har man derfor bedt respondentene om å oppgi hvilken risiko de tror de har. Det store flertallet av verdsettingsstudier bygger imidlertid på en statistisk beregnet risiko for enten hele befolkningen eller en bestemt gruppe i befolkningen.

Det som blir verdsatt er en endring i risiko, for eksempel en reduksjon på 2 per 100.000. Dersom en slik reduksjon er verdsatt til, for eksempel, 500 kroner, fremkommer verdien av et statistisk liv på følgende måte:

$$\text{Verdi av et statistisk liv} = \frac{500}{\left(\frac{2}{100000}\right)} = 25.000.000 \text{ kroner}$$

Verdien av et statistisk liv er verdien av en endring i risiko som statistisk sett tilsvarer en forventet nedgang i antall drepte på 1.

Teorien om vitenskapelige forskningsprogrammer

De første forsøk på å beregne verdien av å unngå dødsfall ved trafikkulykker ble gjort i 1950-årene i Storbritannia, Sverige og USA. Den gang benyttet man den såkalte «humankapitalmetoden». Den gikk i korthet ut på at man beregnet «kapitalverdien» av et menneske, som regel angitt ved verdien av den produksjon mennesket utførte i løpet av sitt (gjenstående) liv. Verdien av produksjonen ble målt med lønnsinntekt.

Denne metoden hadde åpenbare svakheter. Barn og pensjonister hadde ingen verdi; heller ikke hjemmeværende husmødre hadde noen positiv økonomisk verdi. Metoden manglet dessuten et teoretisk grunnlag.

Verdsetting av liv og helse som et vitenskapelig forskningsprogram startet med at teorigrunnlaget ble utviklet. Det skjedde omkring 1970 med bidrag fra økonomene Schelling, Mishan og Jones-Lee. Alle disse argumenterte for at moderne velferdsøkonomi, som danner det teoretiske grunnlaget for nyttekostnadsanalyser, tilsier at det eneste økonomisk meningsfulle og riktige grunnlaget for å verdsette liv og helse er betalingsvillighet for endringer i risiko. Betalingsvillighet er et mål på hvor stor verdi den enkelte tillegger et gode. Jo mer man er villig til å betale for godet, desto mer verdifullt er det. Verdien er individuell og subjektiv; det finnes ingen «objektivt» riktig verdi.

En tanke som melder seg umiddelbart, er hvordan man kan vite om resultatene av verdsettingsstudier er til å stole på når det ikke finnes noen «riktig» verdsetting. Må vi bare godta at folk oppgir hva som helst og ta ethvert beløp alvorlig? Nei, det er, som vi kommer tilbake til senere, mange måter å undersøke verdsettingers sannferdighet på.

Den gren av økonomisk teori Schelling, Mishan og Jones-Lee plasserte verdsettingsstudier innenfor er, i videste forstand, konsumentteorien, eller teorien om hvordan mennesker velger det forbruksmønster som gir dem størst nytte for pengene. Dette er en svært høyt utviklet gren av økonomisk teori med århundrelange tradisjoner. Den bygger på en grunnsetning om at mennesker velger å bruke sine penger på ulike goder på en slik måte at det gir dem mest tilfredshet. En annen måte å si det samme på er at forbrukere antas å være rasjonelt nyttemaksimerende.

Forskning som bygger på en grunnsetning om rasjonelt nyttemaksimerende individer har de kjennetegn Imre Lakatos tillegger vitenskapelige forskningsprogrammer. Lakatos utviklet denne teorien som en (primært) beskrivende teori av hvordan forskere som arbeider innenfor et fagfelt faktisk oppfører seg. Teorien er ment som et begrepsskjema til hjelp ved det Lakatos kalte «rasjonell rekonstruksjon» av forskningshistorien på et fagfelt. Med rasjonell rekonstruksjon mener han en beskrivelse av forskningshistorien som viser at de valg forskere har gjort – med hensyn til hvilke temaer det var lovende å studere og hvordan man best kunne studere dem – var rasjonelle ut fra faginterne normer for rasjonalitet.

Ifølge Lakatos består ethvert forskningsprogram av en hard kjerne, et beskyttende belte, en positiv retningslinje og en negativ retningslinje (begrepene defineres nedenfor). Videre kan et program enten være i en progressiv fase eller i en degenererende fase. Den harde kjernen i et forskningsprogram er grunnsetninger som aldri trekkes i tvil og som det hersker sterke normer mot å trekke i tvil. Disse grunnsetningene tas for gitt av alle som jobber i et program og testes ikke empirisk, det vil si de kan ikke avsannes.

Den harde kjernen er omgitt av et beskyttende belte. Det beskyttende beltet består av hypoteser som kan testes empirisk og dermed kan falsifiseres (avsannes). Dersom en hypotese i det beskyttende beltet blir falsifisert, betegnes dette som en anomali, det vil si som noe unormalt, som tilsynelatende strider mot den harde kjernen og gir grunnlag for å trekke dens grunnsetninger i tvil. I praksis, hevder Lakatos, skjer dette sjelden. Et vitenskapelig forskningsprogram kan leve med mange anomalier. Man bare går videre som om anomaliene ikke fantes, eller prøver å reformulere hypotesene i det beskyttende beltet slik at anomaliene blir å betrakte som normale funn. Et forskningsprogram befinner seg i en progressiv fase dersom dets empiriske innhold øker. Med empirisk innhold mener Lakatos alle observasjoner som predikeres av hypotesene i det beskyttende beltet, både de som er verifisert og de som foreløpig ikke er testet. Det (potensielle) empiriske innholdet kan med andre ord økes ikke bare ved å teste hypoteser empirisk, men også ved å formulere nye, ennå ikke testede hypoteser som har klare empiriske implikasjoner.

Et forskningsprogram befinner seg i en degenererende fase når anomaliene er blitt så mange at de utgjør dominerende funn. Den positive retningslinjen (positive heuristic) i et forskningsprogram går ut på at man skal foreslå hypoteser som utvider det empiriske innholdet, eller som kan forklare anomalier og dermed omgjøre dem til normale funn. Den negative retningslinjen går ut på at man skal unngå å formulere ad hoc hypoteser, det vil si hypoteser som bare forklarer en bestemt anomali, men ikke predikerer nye resultater.

Når det er blitt mange nok anomalier, kan et forskningsprogram bli oppgitt og erstattet av et nytt program. Dette skjer ifølge Lakatos imidlertid ikke i form av en såkalt «vitenskapelig revolusjon» slik Thomas Kuhn hevder. Det skjer snarere ved at et degenererende forskningsprogram assimileres i et nytt forskningsprogram, som aksepterer alt verifisert innhold i det forlatte programmet, men i tillegg inneholder nytt verifisert innhold, og da spesielt innhold som eliminerer anomaliene i det forlatte programmet.

Lakatos spør: Hvis et funn tyder på at en teori er feil, vil da forskerne forkaste teorien? Han svarer nei. Ett, eller for den saks skyld mange, funn som tilsynelatende avsanner en teori er ikke nok til at teorien forkastes. En teori forkastes først når det er utviklet en ny teori som både forklarer de funn som tilsier å forkaste en teori og de funn som støtter teorien. Kort sagt: Man må ha en god forklaring på en anomali før man forkaster den teorien som tolker et funn som en anomali.

Anvendelse av Lakatos' teori på verdsettingsforskningen

Lakatos' teori om vitenskapelige forskningsprogrammer er meget anvendelig for å rekonstruere forskningen om økonomisk verdsetting av liv og helse.

Den harde kjernen utgjøres av grunnsetningen i konsumentteorien om at forbrukere er rasjonelt nyttemaksimerende (det vil si har preferanser som oppfyller visse krav og som de søker å maksimere). Dette er en rent formell, empirisk tom grunnsetning som kan fylles med et høyst varierende empirisk innhold. Ved å forutsette nyttemaksimering (som et rent abstrakt, matematisk postulat), kan man ved å gjøre supplerende antakelser utlede nyttefunksjoner som har bestemte kjennetegn og gir empiriske prediksjoner. Et konkret eksempel på hva dette betyr kan hentes fra Jones-Lee (1974):

1. Individet maksimerer forventet nytte (det er en sannsynlighetsvektet nytte med «levende» og «død» som mulige utfall).
2. Individet foretrekker høy inntekt fremfor lav og er finansielt risikounnvikende (det vil si foretrekker en sikker inntekt fremfor en usikker).
3. Individet ønsker ikke at de etterlatte (arvingene) skal være utsatt for større finansiell risiko enn ham eller henne selv.
4. Individet foretrekker ved en gitt inntekt å være i live fremfor å være død.
5. Grensenytten av inntekt er større når individet er i live enn når individet er avgått ved døden (nytten av inntekt når man er død antas å være knyttet til et ønske om å etterlate en arv).

Gitt disse forutsetningene kan Jones-Lee utlede at det vil eksistere en positiv betalingsvillighet for redusert risiko for å dø. Denne betalingsvilligheten vil ha en positiv sammenheng med inntekt (rike vil betale mer for å redusere en gitt risiko enn fattige). Den vil også ha en positiv sammenheng med risikonivået (jo høyere risiko, desto høyere betalingsvillighet).

Dette eksemplet viser kjernen i logikken ved å utlede hypoteser om betalingsvillighet for redusert risiko på grunnlag av nyttemaksimering. Det viser også hvordan man forholdsvis lett kan finne holdepunkter for å bedømme om resultatene av empiriske undersøkelser gir mening. De gir mening hvis man finner at rike betaler mer enn fattige og hvis man finner at folk betaler mer for reduksjon av en høy risiko enn for reduksjon av en lav risiko. Hypotesene predikerer med andre ord *et systematisk mønster i variasjon i betalingsvillighet*, og man kan fastslå empirisk om resultatene av en undersøkelse viser et slikt mønster eller ikke. Finner man det ventede mønsteret, støttes teorien. Finner man det ikke, svekkes (eller i ytterste konsekvens falsifiseres) teorien.

Den viktigste positive retningslinjen for verdsettingsforskningen var å utvikle hypoteser om systematisk variasjon i betalingsvillighet. Disse hypotesene dannet forskningsprogrammets beskyttende belte. I en nokså tidlig fase av verdsettingsforskningen, omkring 1980, ble det gitt en rekke vesentlige teoribidrag. Det ble tidlig klart at man ikke alltid fant den teoretisk forventede variasjon i betalingsvillighet. Dette ble i første omgang forklart med at metodene som ble brukt i undersøkelsene var for dårlige. En annen viktig positiv retningslinje i verdsettingsforskningen ble derfor å arbeide for en kontinuerlig metodeutvikling for å kunne gjøre bedre studier.

Den progressive fasen (1970-1995)

Etter at teorigrunnlaget for verdsettingsforskningen ble etablert omkring 1970, tok det noen år før empirisk forskning kom i gang for alvor. Men fra ca. 1980 begynte empirisk forskning både i USA og Europa. To ulike forskningstradisjoner utviklet seg. I USA ble den dominerende metoden studier av faktisk atferd (revealed preferences) der sikkerhet veies mot andre goder, i første rekke studier av lønnskompensasjon for høy risiko i yrker. I Europa ble den dominerende metoden betinget verdsetting (contingent valuation), der man stiller direkte spørsmål om betalingsvilligheten for en reduksjon i risiko, eventuelt hva man vil kreve i kompensasjon for en høyere risiko.

Den første store undersøkelsen i Europa ble utført av Jones-Lee med flere i Storbritannia. Resultatene av undersøkelsen forelå i 1983, men er publisert flere ganger både i form av artikler i vitenskapelige tidsskrifter og i form av en bok Jones-Lee utga i 1989. Samtidig med at denne undersøkelsen pågikk, tok Jones-Lee initiativ til å organisere en internasjonal forskerkonferanse som ble holdt i Geneve i 1981. Flere av bidragene på denne konferansen var teoretiske og førte til at en rekke nye hypoteser om variasjon i betalingsvillighet kunne formuleres. Slik økte det empiriske innholdet i verdsettingsforskningen.

Det empiriske innholdet økte også ved at den britiske verdsettingsstudien tjente som forbilde for tilsvarende studier i Østerrike, Sverige, New Zealand, Danmark, Sveits og Frankrike. Alle disse undersøkelsene var tilnærmet kopier av, eller hadde vesentlige fellestrekk med den opprinnelige britiske studien. Alle undersøkelsene ble utført før 1995.

Denne perioden må betegnes som en progressiv fase for verdsettingsforskningen som et forskningsprogram. Så vel teorigrunnlag som empiriske resultater økte raskt i omfang. I den samme perioden ble det også utført en rekke lønnskompensasjonsstudier i USA. Noen av disse avdekket feilkilder i eldre studier, noe som kan betegnes som et progressivt element siden det innebar utvikling av bedre metoder for slike studier.

Det som markerte at den progressive fasen gikk mot slutten var et økende innslag av anomalier i resultatene og en økende tvil om metodene var gode nok. Et vanlig funn i betingede verdsettingsstudier var manglende følsomhet for størrelsen på risikoreduksjonen (insensitivity to scope). Et typisk mønster var at folk ikke ville betale dobbelt så mye for en dobbelt så stor nedgang i risiko, men bare litt mer, eksempelvis 20-30 prosent. Det ble antatt at dette kunne skyldes at folk ikke oppfattet forskjeller i endringer i lave risikonivåer korrekt. En risikonedgang på 4 per 100.000 er det dobbelte av 2 per 100.000, men begge reduksjoner er mikroskopiske tall som kanskje ikke oppfattes som så forskjellige. I en fransk undersøkelse forsøkte man å unngå dette problemet ved å spørre om ulike reduksjoner i antall drepte, fra 50 til 5000. Sistnevnte tall innebar en reduksjon av antall drepte i trafikken i Frankrike på mer enn 50 prosent. Man skulle tro at folk lettere ville oppfatte forskjellen mellom 50 og 5000 enn forskjellen mellom 2 per 100.000 og 4 per 100.000. Men mønsteret gjentok seg. Franskmennene var ikke villige til å betale 100 ganger så mye for å redusere antall drepte med 5000 som for å redusere antall drepte med 50 – bare drøyt 4 ganger så mye (de samme personene ble spurt om 50 og 5000).

I tillegg viste det seg at ulike måter å få folk til å svare på spørsmål om betalingsvillighet ga ulike resultater, selv om de ulike svarformatene ifølge økonomisk teori ikke burde bety noe. I lønnskompensasjonsstudiene viste det seg at resultatene spriket mer og mer jo flere studier som ble gjort, selv om det var rimelig å hevde at de nyeste studiene både bygde på bedre data og benyttet bedre analysemetoder enn eldre studier. Med andre ord: bedre data og bedre metoder førte ikke til mer ensartede resultater, men til det motsatte.

Mot slutten av 1990-årene var den progressive fasen slutt. Verdsettingsforskningen gikk over i en ny fase.

Kampen mellom progressive og degenererende tendenser (1995-2005)

Resultatene av betingede verdsettingsstudier var mot slutten av 1990-årene sterkt preget av anomalier. Noen gikk så langt som til å hevde at resultatene utelukkende var anomalier. Jones-Lee, som hadde spilt en nøkkelrolle i den progressive fasen, forkastet i 1998 betingede verdsettingsstudier i den form han til da hadde utført dem og lanserte en ny metode der man unngikk å spørre folk om endringer i lave risikonivåer.

Fra omkring 2000 ble det utført verdsettingsstudier der man i stedet for å spørre direkte om betalingsvillighet ga folk valgoppgaver (stated choice). Den vanligste formen for valgeksperiment var valg mellom to veger, som hadde ulike kjennetegn med hensyn til for eksempel reisetid, bompenger og antall ulykker. Respondentene ble bedt om å velge den ene eller andre vegen; kjennetegnene ble variert og folk bedt om å gjøre et nytt valg. Som regel gjorde hver respondent 5-10 valg. Respondentenes verdsetting av de ulike kjennetegnene ble så utledet under analysen av valgene.

I USA rettet Peter Dorman i 1996 et kraftig angrep mot lønnskompensasjonsstudiene. Han hevdet at alle disse studiene hadde avgjørende metodefeil og at metoden derfor burde oppgis. Den ledende forskeren innenfor lønnskompensasjonstradisjonen i USA, Viscusi, forkastet Dormans kritikk og har senere utført en rekke nye studier. Dormans kritikk vant med andre ord ikke fram. En mulig grunn til det, med henvisning til at Lakatos sier at funn som strider mot en teori i seg selv ikke er nok til at forskere forkaster teorien, er at Dorman ikke hadde noe bedre å komme med. Hans argument var bare at: «Dette er ikke godt nok» uten at han forklarte hva som ville være godt nok.

Valgeksperimenter ble lansert som en bedre metode enn betinget verdsetting, blant annet fordi man unngår å spørre om endringer i mikroskopiske risikonivåer. Det viste seg imidlertid raskt at valgeksperimentene bød på nye problemer. Mange valgte leksikografisk, det vil si de valgte alltid det som var best ut fra ett av kjennetegnene og så bort fra de andre. Mange valgte inkonsistent, det vil at de på trinn N i en valgsekvens valgte i strid med den verdsetting de avslørte på, for eksempel, trinn N – 1 i valgsekvensen. Bare 10-20 prosent av respondentene valgte fullkomment rasjonelt i henhold til økonomisk teori.

Dermed artet forskningen seg som en tvekamp mellom progressive og degenererende tendenser. Nye verdsettingsmetoder representerte de progressive tendensene. Svakheter som ble funnet med de nye verdsettingsmetodene representerte de degenererende tendensene. Anomaliene forsvant ikke, men antok bare nye former.

Parallelt med disse motstridende tendensene i empirisk forskning skjedde en vesentlig reformulering av verdsettingsteorien.

Det beskyttende beltet blir nesten altomfattende (2000-2010)

De tidlige teoribidragene i verdsettingsforskningen ga entydige prediksjoner av resultater. Eksempelvis predikerte Jones-Lee (se tidligere i sammendraget) at jo høyere risikonivået var, desto høyere ville betalingsvilligheten for å redusere det være (alt annet likt). Det tok ikke lang tid før mer kompliserte teorier ble utviklet. Disse teoriene ga ikke like klare prediksjoner. Et godt eksempel er den usedvanlig kompliserte teorien Dehez and Drèze (1982) utviklet. Den predikerer følgende:

1. Dersom individet ikke har livsforsikring eller inntektsforsikring og grensenytten av penger er større i live enn ved død, vil betalingsvilligheten synke når risikonivået synker.

2. Dersom individet har optimal livsforsikring og inntektsforsikring til aktuarmessig rettferdige betingelser (det vil si premien som betales er identisk med forventet tap), er betalingsvilligheten uavhengig av risikonivået. Forklaringen på dette er at grensenytten av penger da er den samme ved liv som ved død.
3. Dersom individet har livsforsikring og inntektsforsikring som er mindre enn aktuarmessig rettferdig, vil betalingsvilligheten øke med synkende risikonivå.
4. Dersom individet har livsforsikring og inntektsforsikring som er mer enn aktuarmessig rettferdig, vil betalingsvilligheten øke med økende risikonivå.
5. Dersom individet har livsforsikring og inntektsforsikring der premien justeres i takt med risikonivået for å opprettholde aktuarmessig rettferdighet, vil betalingsvilligheten øke med synkende risikonivå.

Dette kan man trygt kalle helgardering. Alt er mulig, både at betalingsvilligheten øker når risikoen blir høy, at den er uavhengig av risikonivået og at den synker når risikoen øker. Ingen av disse resultatene er teoretisk utelukket. Uansett hva man finner i en empirisk undersøkelse, har det teoretisk støtte – vel og merke så lenge man ikke innhenter detaljerte opplysninger om hvordan folk er forsikret. Selv med slike opplysninger kan det by på problemer å avgjøre om forsikringen er aktuarmessig rettferdig eller ikke. Normalt vil en forsikringspremie overstige forventet tap, siden forsikringsselskapet har administrasjonskostnader og ønsker å tjene penger. Men i noen land gis skattefradrag for livsforsikring. Kanskje gjør det at premien likevel kan betraktes som aktuarmessig rettferdig?

Kort sagt: Det er i praksis bortimot umulig å falsifisere de hypoteser Dehez og Drèze foreslo. Deres bidrag var begynnelsen til flere lignende teoribidrag som i dag har kommet så langt at nesten ethvert resultat har teoretisk støtte. Ett av de mest konsekvensrike bidragene til teoriutviklingen er Amiran og Hagens (2003, 2010) teori om retningsbegrensede nyttefunksjoner (directionally bounded utility functions). Teorien sier, noe som umiddelbart høres ganske rimelig ut, at det er grenser for hvor mye av inntekten folk er villige til å bruke på ikke-markedsgoder. De vil ha igjen noe, som regel det aller meste, av inntekten til å bruke på løpende forbruk. Selv om noen sparer en del av inntekten, betyr ikke det at de kan bruke sparepengene på hva som helst. Sparingen har gjerne et formål, og det oppnår man ikke hvis sparepengene brukes opp på å støtte alskens gode formål man blir spurt om å støtte.

Plausibel eller ikke, teorien om retningsbegrensede nyttefunksjoner har revolusjonerende implikasjoner. Den predikerer nemlig at betalingsvilligheten vil være tilnærmet uavhengig av mengden av godet man blir bedt om å betale for. Så da er det likevel ikke så rart, eller uventet, at folk sier at de vil betale nesten det samme uansett hvor stor nedgangen i risiko er. Det er akkurat slik vi skulle vente.

Imre Lakatos' kollega ved London School of Economics, Karl Popper, lanserte begrepet immuniserende reformulering om reformuleringer av en teori for å gjøre den immun mot falsifisering. Betalingsvillighetsteorien er i dag reformulert så grundig at de fleste resultater som før ble regnet som anomalier i dag betraktes som teoretisk plausible.

Men dermed har man ødelagt teoriens viktigste støttefunksjon i verdsettingsforskningen. Når teorien ikke lenger sier at bare et bestemt mønster i resultater gir mening, men sier at ethvert mønster, selv et salig rot av resultater som spriker i alle retninger gir mening, da kan man ikke lenger henvise til teori for å forkaste resultater av empirisk forskning. Ethvert resultat har teoretisk støtte.

Går den harde kjerne i oppløsning (2005-2015)?

Det er liten tvil om at ambisjonen for verdsettingsforskningen til å begynne med var å finne en bestemt verdi av et statistisk liv. Argumentasjonen fra økonomene for nytten av en økonomisk verdsetting av risiko understreket sterkt et ideal om effektiv ressursallokering. Effektiv ressursallokering betyr at man bruker like mye på å redde et liv på alle sektorer i samfunnet. Dette sikrer at man redder flest liv for en gitt sum penger. Det er lett å vise at dersom man bruker, for eksempel, 100 millioner kroner per reddet liv i en sektor, men bare 10 millioner kroner i en annen sektor, kan man (forutsatt at man ikke har nådd en absolutt grense for reduksjon av risiko) redde flere liv totalt sett ved å bruke, for eksempel, 25 millioner kroner per reddet liv i alle sektorer.

Etter hvert som empirisk forskning har vist mer og mer sprikende resultater har noen begynt å gi opp tanken om å finne en bestemt verdi av et statistisk liv. Flere har argumentert for at en slik verdi ikke finnes. Vi vet fra teorien at betalingsvillighet varierer etter, for eksempel, alder og inntekt. De siste årene har mange spurt om man bør begynne å bruke en varierende verdi av et statistisk liv og hvilke kilder til variasjon man i så fall bør tolerere. Både Jones-Lee og Viscusi har argumentert for at verdien av et statistisk liv bør variere etter inntekt. Det er mest i samsvar med teorigrunnlaget for nyttekostnadsanalyser, hevder de.

Dermed kan man vel på en måte si at sirkelen er sluttet. Forskningen startet med en ambisjon om å finne en universelt gyldig verdi av et statistisk liv. Man innså fort at en slik verdi ikke fantes. Hypoteser ble utviklet om systematisk variasjon i betalingsvillighet. Tanken bak disse hypotesene var ikke nødvendigvis at man skulle la verdien av et statistisk liv variere tilsvarende. De var snarere ment som holdepunkter for å vurdere validiteten i empiriske resultater.

Man fant bare delvis den systematiske variasjonen som var teoretisk forventet. I tillegg fant man masse variasjon som teoretisk ikke var forventet og som hadde sitt opphav i kilder som ifølge økonomisk teori er irrelevante. Etter hvert som dette vedvarte, til tross for metodeutvikling, utviklet det seg en tendens til å reformulere betalingsvillighetsteorien. Flere resultater som i utgangspunktet ble regnet som anomalier kunne dermed omtolkes som normale resultater.

Det store spriket i resultater ble på denne måten omgjort fra et problem til noe som var teoretisk forventet. I lys av dette ble også idealet om å benytte en fast, enhetlig verdi av et statistisk liv trukket i tvil.

I dag er følgelig lite igjen av det opprinnelige innholdet i verdsettingsforskningen som et vitenskapelig forskningsprogram. Programmet kan sies å ha kommet til veis ende i den forstand at man både har oppgitt tanken om å finne en bestemt verdi av et statistisk liv og langt på veg også er i ferd med å oppgi idealet om å bygge på en slik verdi for å oppnå effektiv ressursallokering.

Vi spør likevel: Hvor går veien videre? Hva er utsiktene for verdsettingsforskningen?

Kan meta-analyser rydde opp? (2000-2015)

Det er de siste ca. 15 årene gjort flere meta-analyser av litteraturen om verdien av et statistisk liv. Rapporten gjennomgår disse analysene og vurderer om de har bidratt til å forklare den store variasjonen i verdier som finnes i studiene.

De fleste meta-analyser er relativt enkle og oppfyller ikke de krav som bør stilles til gode meta-analyser. Blant annet anvender de fleste analyser ikke optimale statistiske vekter av resultatene, og de færreste tester for publikasjonsskjevhet. Noen meta-analyser forsøker å skille mellom gode og dårlige verdsettingsstudier.

Den mest omfattende meta-analysen, publisert i flere omganger, er gjort av de norske forskerne Ståle Navrud, Henrik Lindhjem og Nils-Axel Bråthen, med bistand fra den franske økonomen Vincent Biausque. I siste oppdatering bygger denne analysen på 931 anslag på verdien av et statistisk liv. De sorterer undersøkelser i to grupper etter kvalitet og har gjort omfattende meta-regresjonsanalyser for å finne faktorer som påvirker verdien av et statistisk liv. De har ikke testet for publikasjonsskjevhet.

Analysene tyder på at det som i sterkest grad påvirker verdien av et statistisk liv er inntekt (jo høyere inntekt, desto høyere verdi), størrelsen på risikoendringen (jo større endring, desto lavere verdi) og om godet er kollektivt eller ikke (lavere verdi for kollektive goder). Noen vil muligens stusse over at verdien av et statistisk liv er lavere jo større endringen i risiko er. Dette forklares ved at betalingsvilligheten ikke øker proporsjonalt med endringen i risiko, men langt mindre. Et regne-eksempel kan illustrere dette:

Risikoreduksjon	Betalingsvillighet	Verdi av et statistisk liv
$1 \cdot 10^{-5}$	400	40.000.000
$2 \cdot 10^{-5}$	500	25.000.000
$5 \cdot 10^{-5}$	600	12.000.000
$10 \cdot 10^{-5}$	800	8.000.000

Det er påvist betydelig publikasjonsskjevhet i litteraturen om verdien av et statistisk liv. Dersom man korrigerer for denne, tyder analyser på at verdien av et statistisk liv er overvurdert med en faktor på 3 (det vil si at verdien korrigert for publikasjonsskjevhet skal være, for eksempel, 3 millioner kroner, mot publisert 9 millioner kroner).

Noen av meta-analysene forklarer en høy andel av variasjonen i et statistisk liv. Det er likevel åpenbart at den faktiske variasjonen er langt større enn den man kan reprodusere ved hjelp av en meta-regresjon. Det er store restledd i begge ender av fordelingen (ved de laveste og høyeste verdier). Meta-analysene er ikke i stand til å peke på en bestemt verdi av et statistisk liv som det, alt i alt, beste anslaget. Analysene gir heller ikke svar på det normative spørsmålet om man bør la verdien av et statistisk liv variere, og i så fall hvor mye og ut fra hvilke kriterier.

Kan nye metoder gi mer presise svar?

Det er åpenbart at de metoder som hittil har vært brukt i verdsettingsforskningen ikke har ført til at man har oppnådd særlige presise anslag på verdien av et statistisk liv. Selv i hver enkelt undersøkelse kan ulike anslag ofte variere med en faktor på minst 10. Spørsmålet er om andre metoder kan tenkes brukt som kan gi mer presise resultater.

Noen slike metoder blir vurdert. Den første metoden er å bygge på kvalitetsjusterte leveår. Man kan da verdsette, for eksempel, en kjent helsetilstand som mange har erfaring med økonomisk. Denne tilstanden forutsettes å ha en kjent verdi i form av kvalitetsjusterte leveår. Et kvalitetsjustert leveår er et leveår som tilordnes en verdi mellom 0 (død) og 1 (fullkommen helse) avhengig av hvor mye helsetilstanden er redusert. Dersom en reduksjon av helsetilstanden som fører til at man tilbringer ett år i en helsetilstand med verdien 0,98 verdsettes til 50.000 kroner, blir verdien av et statistisk liv (leveår) $1/0,02 \cdot 50.000 = 2.500.000$ kroner.

Det er to hovedproblemer med en slik metode. For det første finner det mange måter å måle kvalitetsjusterte leve år på og de gir ikke samme resultater. En gitt helsetilstand kan derfor tilordnes mange forskjellige verdier for kvalitetsjusterte leveår. En oppskalering til verdien av et statistisk liv (eller leveår) kan dermed gi høyst sprikende resultater og vil ikke løse problemet med at verdien av et statistisk liv spriker så mye. For det andre bygger beregning av kvalitetsjusterte leveår på mange restriktive forutsetninger om individers nyttefunksjon knyttet til helse, langt mer restriktive enn de forutsetninger som vanligvis gjøres i betalingsvillighetsstudier. Det anbefales derfor ikke å bruke kvalitetsjusterte leveår som et utgangspunkt for å finne verdien av et statistisk liv.

Den andre metoden som nevnes er «the capability approach» som er utviklet av Amartya Sen. Han er opptatt av å finne objektive mål på livskvalitet som blant annet kan brukes til å begrunne kamp mot fattigdom og sosial nød. Denne tilnærmingsmåten er beslektet med forskning om sosiale indikatorer, som andelen av befolkningen som har innlagt bad og WC, andelen som kan lese og skrive, osv. I våre dager er denne typen indikatorer trolig mest aktuelle i utviklingsland, som indikatorer på hvordan de utvikler seg. Metoden synes ikke egnet til økonomisk verdsetting av liv og helse og er dessuten grunnleggende paternalistisk (i den forstand at den forteller folk hva som er bra for dem, i motsetning til at de selv gir uttrykk for hva som er bra for dem).

Den tredje metoden er å beregne verdien av et statistisk liv på grunnlag av empiriske nyttefunksjoner. Med en empirisk nyttefunksjon menes en funksjon som viser hvordan subjektiv velvære (lykke) avhenger av inntekt og andre ting som påvirker det. Det finnes en rekke slike funksjoner i litteraturen og de kan brukes til å anslå verdien av et statistisk liv. Problemet er at det finnes ulike funksjoner som gir ulike resultater og at ikke alle er enig i at denne typen funksjoner kan tolkes som nyttefunksjoner i den økonomisk betydningen av begrepet.

1   Background and research problem

1.1   Background

Research for the purpose of assigning a monetary value to the saving of human life in the transport sector, often referred to as the cost of road accidents, has a history going back more than 60 years. The first studies were published in the nineteen fifties. Thus, Dawson (1967) quotes a study by Reynolds (1956), published in the Journal of the Royal Statistical Society in 1956. The first estimate of road accident costs for Great Britain, also briefly described by Dawson, dates to 1938.

All the early studies of the monetary value of life saving were based on the human capital approach (Becker 1964). According to this approach, the monetary value of saving a life was equal to the human capital that life represented. Human capital was estimated in terms of the discounted value of the future earnings of an accident victim. In some studies, the value of the accident victim's future consumption was subtracted, in order to gain a measure of the surplus of value an individual generated beyond what he or she needed to support himself or herself. This was referred to as the net lost output method. To obtain the total cost of accidents or injuries, direct costs, such as costs of medical treatment, property damage or costs of police investigations were added to the value of lost earnings.

In most estimates of the costs of road accidents made by means of the human capital approach, the value of lost output made up most of the costs. An example of the results obtained when using the net lost output approach is given in Table 1.1, which is taken from the report by Dawson (1967).

Table 1.1. Total cost of loss of output due to fatalities and average cost per fatality. Taken from Dawson, 1967, Table 3

	Value of net lost output for road accident fatalities in 1963 (GB pounds)					
	Urban areas		Rural areas		All areas	
Gender	Per victim	Total	Per victim	Total	Per victim	Total
Male	3720	10670000	5220	11360000	4360	22030000
Female	-1530	-2040000	-110	-60000	-1120	-2100000
Both	2040	8630000	4150	11300000	2880	19930000

It is seen that the value of net lost output is negative for females. Dawson remarks the following about this:

"A negative loss implies that from a strictly material point of view the community gains from a person's death: however, when the subjective factors are taken into account (see chapter 7) the losses became positive in all cases."

It is obviously somewhat embarrassing when an estimate of the benefits to society of preventing road accident fatalities ends up by showing that society would be better off by simply killing some of the road accident victims. Indeed, according to the net lost output approach, all those who did not earn enough to contribute to supporting others, had negative values. This included children, the retired, and housewives not belonging to the market labour force.

Dawson specified the "subjective factors" as follows: (1) Pain, suffering and shock, (2) Loss of amenities of life, (3) Loss of expectation of life, (4) Inconvenience and discomfort, (5) Exemplary damages. These items were not further explained and would seem to involve some double counting. What, for example, is the precise difference between pain and suffering on the one hand and inconvenience and discomfort on the other?

Dawson made use of a study by Thedié and Abraham (1961) in order to estimate the value of the "subjective factors". He apparently had some misgivings (or at least un-answered questions) about that study and suggested the following method for estimating the value to society of preventing road accident fatalities:

"A possible way of arriving at an estimate of the amount that the community is prepared to pay to save life is to examine what, in effect, is paid in a number of different circumstances. Costs, direct and indirect, are incurred in making trains, ships and aircraft safer, in providing firefighting and lifeboat facilities, and in a number of ways in the field of medicine. It is possible that, by examining a number of such cases, a value would be arrived at which provides a consensus of opinion. It is, however, possible that the scatter of values would be so wide that no useful result will emerge. In the meantime it is suggested that the following rather arbitrary, average values should be used …."

Thus, arbitrariness was regarded as the lesser evil when compared to the embarrassment of assigning a negative value to life saving. A very comprehensive study along the lines suggested by Dawson was reported by Tengs et al. in 1995 (Tengs et al. 1995). Tengs et al. studied the cost-effectiveness of 587 lifesaving interventions. Cost-effectiveness was stated as the cost of the intervention per life year saved. Costs per life year saved ranged from negative to more than 10 billion US dollars, confirming that (slightly paraphrasing Dawson) "the scatter of values is so wide that no useful result emerges".

The scientific approach to estimating the value of preventing human death has changed fundamentally since the days of Dawson. Prominent economists called for switching to a different method not so long after Dawson published his report. The new approach, the willingness-to-pay approach, was firmly anchored in modern welfare economics. It was argued (Schelling 1968, Mishan 1971) that the only theoretically correct measure of the value of preventing a fatality for use in cost-benefit analysis was one based on the willingness-to-pay for the safety improvement of those who would benefit from it (assuming potential beneficiaries can be identified in advance). They also argued that the safety improvement should be stated as a reduction in the risk of death. The papers arguing for adopting the willingness-to-pay approach are discussed more in detail in Chapter 3 of the report.

Figure 1.1 gives an overview of the methods that have been used to obtain a monetary valuation of reduced risk of death. All these methods are still used, although historically there is a clear trend towards using methods based on the willingness-to-pay approach. Even if nearly all economists today would recommend the willingness-to-pay approach, papers based on the net lost output approach continue to be published (Pukalskas et al. 2015). This report will focus on studies based on the willingness-to-pay approach.

There are two main methods for eliciting willingness-to-pay: stated preference methods and revealed preference methods. Most studies of the valuation of road safety have employed stated preference methods. There are two main versions of stated preference methods: the contingent valuation method and the stated choice method.
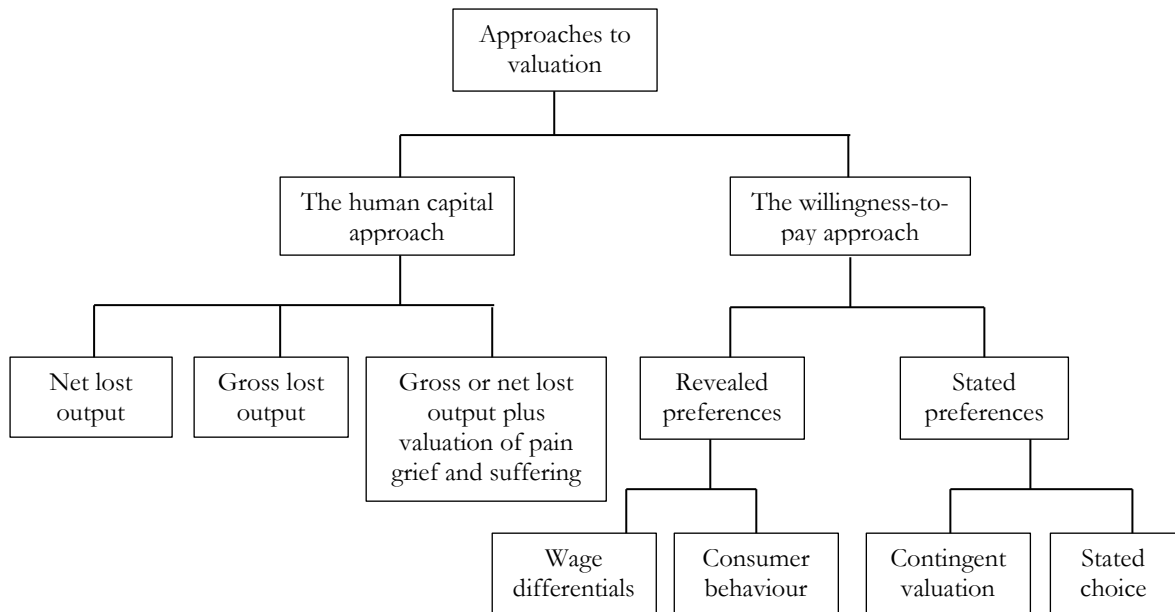
Figure 1.1: Approaches to the economic valuation of reduced risk of accidental death

In the contingent valuation method, a sample of the population is asked direct questions about how much they are willing to pay for a certain reduction of the risk of dying or getting injured in a road accident (or another source of risk, such as the risk of contracting a certain disease). There are many versions of the method. The simplest version is to ask directly about willingness-to-pay, without indicating any answer (open ended). Another version is to provide a so called "payment card", indicating different amounts and asking people to select one of these amounts. A third version is called "iterative bidding". Respondents are offered a bid (price) and asked to take it or not. If the first bid is rejected, a lower bid is offered. If that bid is accepted, iteration ends; otherwise it continues until the bid is accepted. Conversely, if the first bid is accepted, higher bids are offered until the last bid is rejected. A fourth version of the method, the "referendum method", involves stating a bid and asking people if they take it or not, by voting yes or no to it. This version of the method is perhaps the one that most closely resembles a real market in which consumers decide whether or not to buy a good based on its price.

The stated choice method asks people to make a choice between two options. The options are characterised by certain attributes, one of which is safety. Respondents do not state an amount they are willing to pay. They simply choose an option, and the valuation implicit in that choice is estimated by the analyst. The choices presented would typically be between two roads, two residential areas or two modes of transport.

Revealed preference studies examine actual choices in real markets. As far as road safety is concerned, such a choice might be the purchase of a new car. Cars differ with respect to safety features; if the relative importance of the factors that influence the choice of car, such as price, size, motor power, safety features, etc. can be determined, the implicit value placed on various safety features can be estimated. Studies of so called compensating wage differentials, i.e. extra payment for taking on risky jobs have been very common in the United States, but less common in Europe.

There is a distinct difference in the approaches taken in North America and Europe regarding the monetary valuation of safety. In North America, almost all studies are based on revealed preferences. In Europe, by contrast, most studies are based on stated preferences.

This report reconstructs the history of studies of willingness-to-pay for transport safety. It will not include studies relying on different approaches, as there is almost unanimity among economists that willingness-to-pay is the only meaningful approach. Several hundred studies have been made to estimate the willingness-to-pay for improved transport safety. These studies have produced a very wide range of estimates of the monetary value of transport safety. One critical observer, Ezra Hauer (2011A), notes that the values produced by studies of willingness-to-pay are all over the place. This is true.

A typical willingness-to-pay study deals with the monetary valuation of small changes in risk. These small changes are then aggregated into the value of preventing one fatality, often referred to as the value of a statistical life (abbreviated VSL). If, for example, the mean willingness-to-pay (arbitrary monetary units) for a risk reduction of 2 in 100,000 is 500, the value of a statistical life is:

$$\text{Value of a statistical life} = \frac{500}{\left(\frac{2}{100000}\right)} = 25{,}000{,}000 \text{ in any currency}$$

An equivalent measure of valuation is called willingness-to-accept (WTA). The problem then is how much an individual needs to be compensated in order to accept a certain risk. Willingness-to-pay is often abbreviated to WTP and willingness-to-accept abbreviated to WTA. The literature on the monetary valuation of reduced risk of death now contains more than one thousand estimates of the value of a statistical life. These estimates vary enormously. A recent meta-analysis by Bellavance et al. (2009) illustrates this. The meta-analysis dealt with studies of compensating wage differentials. Figure 1.2 is based on the study.

Studies have been listed chronologically. As can be seen from the figure, the diversity of the estimates has grown over time. Bellavance et al. remark (2009:453):

"After 30 years of research and publication on the topic, we might expect a certain convergence in the values obtained. When we examine Figure 1, we note quite the contrary. The most recent studies seem to diverge instead. And it is also interesting to observe a positive relation between the values of a statistical life and the year of publication."

A more recent meta-analysis (Lindhjem et al. 2011) included a total of 856 estimates of the value of a statistical life based on stated preference studies, by far the largest number of estimates included in any meta-analysis so far. The estimates ranged from 4,450 US-dollars (2005-prices) to 197 million US dollars, a ratio of more than 44,000. The range of values was smaller when only the studies that were classified as "best" were included, but still substantial.

Is it possible to account for this huge range in estimates of the value of a statistical life? Can the sources of diversity be identified? If a single value is to be extracted from the literature for use in cost-benefit analysis, how can it best be done? Are all estimates found in the literature to be trusted, or should some of them be rejected? If so, on what grounds?

Estimates of the value of a statistical life by year of study - compensating wage differentials - all studies - Bellavance et al. 2009
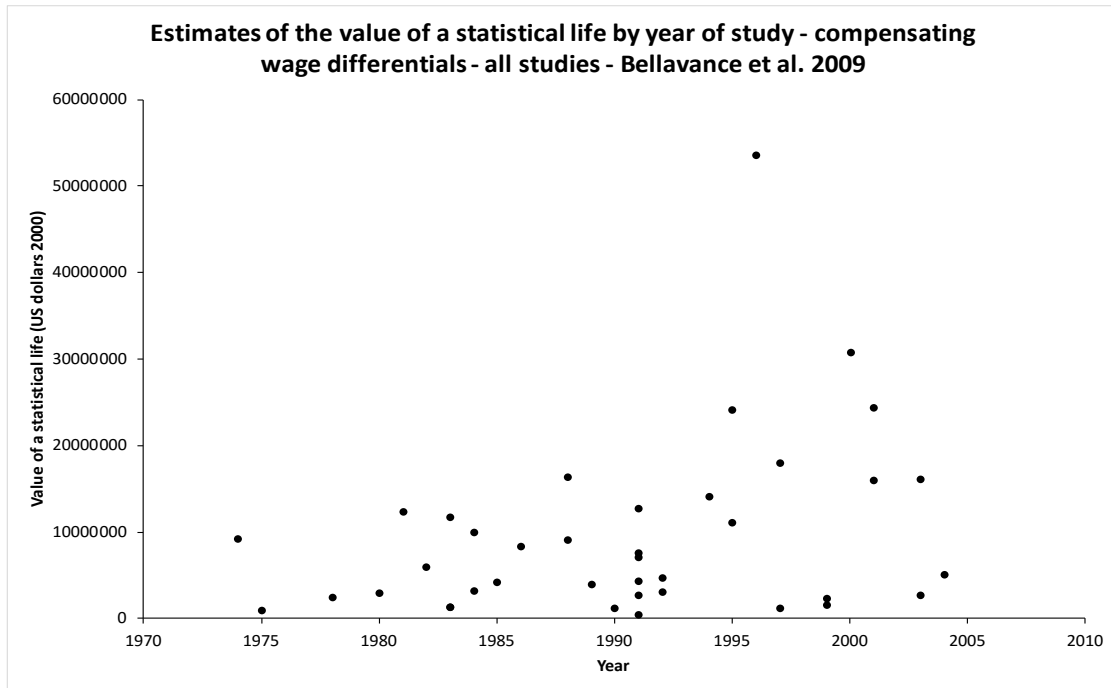
Figure 1.2: Estimates of the value of a statistical life by year of study. Taken from Bellavance et al. 2009

These are just a few of the questions that need to be asked in view of the huge variation in estimates of the value of preventing a fatality found in the literature. Hauer (2011A) notes:

"Variability of VSL estimates has several causes. First, what VSL is, is a matter of definition. The 'human capital' definition considers VSL to be based on a person's future earnings; the willingness-to-pay (WTP) definition is based on how much money people are willing to part with for a certain reduction in the risk to die. Different definitions lead to different estimates. Second, VSL estimates are extracted from imperfect data by methods based on assorted unsupported assumptions with all the attendant inadequacies. Third, VSL is not like a physical constant that has the same value for everyone, everywhere and always. For a homo economicus the VSL depends on that person's traits. As all VSL estimates are averages for a specific group of people at a particular time, they reflect the traits (age, wealth, norms, etc.) of those people at that time. Fourth, for a homo economicus the VSL depends on the specifics of the intervention options. To what extent these four reasons explain (the) very wide range of VSL estimates is not clear. The fifth reason for the diversity of VSL estimates is more basic. The Homo sapiens is cognitively badly equipped to contemplate small changes in small future risks. … It is almost as if one surveyed the customers at a gas station about what they think is the molecular weight of unleaded gasoline."

Hauer is by no means alone in voicing these concerns. Dorman (1996) argues that all studies relying on the compensating wage differentials model are methodologically flawed and should be rejected. The compensating wage differentials model is based on the assumption that workers are compensated for occupational risks by means of higher wages. Kahneman et al. (1999) argue that the results of valuation studies, in particular those relying on the contingent valuation method – by which people are asked directly how much they are willing to pay for a certain good – are expressions of attitudes towards the provision of the goods ("road safety is a good thing") rather than of decisions about how much money to spend on providing the goods. Loomes (2006) notes that the assumptions that underpin the conventional economic model of 'rational agents' tend to be substantially violated in studies designed to obtain valuations of health, safety and environmental goods. Hausman (2012) concludes that the contingent valuation method has gone from bad to hopeless and suggests that it should no longer be used.

In short, the current state of knowledge about the value of preventing a fatality based on studies of willingness-to-pay can be characterised as follows:

1. Research has not produced a firm estimate of the value of preventing a fatality. On the contrary, estimates vary enormously, by a factor of more than 44,000.
2. The huge variation in estimates of the value of preventing a fatality has not diminished over time. There is rather a tendency for estimates to become more diverse over time.
3. Nobody can account very well for the huge variation in estimates of the value of preventing a fatality, but it is clear that part of the variation can be attributed to factors that, according to economic theory, should not produce the variation found (such as anchoring effects in iterative bidding studies).
4. There is no consensus among economists about the best method for studying willingness-to-pay to prevent a fatality. Some economists reject methods that have been widely used.
5. There is no consensus about the interpretation of the results of studies designed to elicit willingness-to-pay. Some argue that these studies do not actually measure what they are intended to measure, but rather measure attitudes.

All these points could have been made with equal force 20 years ago. In the meantime, valuation research has continued unperturbed, almost as if the points of criticism listed above did not exist. One wonders why a field of research which does not produce meaningful results, and in which there is no consensus about research methods, continues to exist and, indeed, flourish. This forms the background of the research problems to be studied in this report.

1.2   Research problems

The following main research problems will be studied in this report:

1. What is the rationale for studying the monetary valuation of preventing fatalities and injuries in transport? Can effective and rational transport safety policies be developed without applying a monetary valuation of transport safety?
2. Can changes in the risk of dying be treated as a homogeneous commodity to which it makes sense to attach a fixed value, or is risk and changes in it a multidimensional concept for which the various dimensions cannot be reduced to a single monetary value? How do different academic disciplines conceive of risk and the possibility of assigning a monetary value to changes in it?
3. What is the appropriate theoretical foundation according to economic theory for studying the monetary valuation of transport safety? How did economists justify the need for, and the basic approach to this field of research?
4. How can one explain that a field of research producing so diverse findings as studies of the monetary valuation of transport safety, in which there is no agreement on the best method, continues to exist despite the diversity of findings and methods? Are there theories of science that may help in understanding and explaining the continuation of research in a field characterised by enormously varying findings that are difficult to explain?
5. One theory of science that tries to explain the continuation of research in a field characterised by anomalous (i.e. unexpected and difficult to explain) findings: the methodology of scientific research programmes, proposed by Imre Lakatos. Can be the methodology of scientific research programmes be applied to reconstruct the history of research on the monetary valuation of transport safety? Can the concepts

of this theory of science be used to identify phases in the history of valuation research? Does the methodology of scientific research programmes help in understanding the development of theories and methods in the study of willingness-to-pay for transport safety?

6. What are the principal sources of variation in willingness-to-pay for transport safety from a theoretical point of view? How have researchers developed hypotheses about this? Do the hypotheses make predictions that can be tested empirically?

7. One commonly applied method to try to summarise a large body of research and look for systematic patterns in results is meta-analysis. Can meta-analysis make sense of the widely diverging estimates of the value of a statistical life? Can meta-analysis identify sources of this huge variation and help in selecting studies of high methodological quality?

8. There is a growing understanding of the fact that trying to find a single monetary value of transport safety that can be applied to any decision influencing transport safety is doomed to failure. It is argued that, in theory, there is no uniform monetary valuation of transport safety; rather the value depends on the context. What are the implications of adopting a variable monetary valuation of transport safety? Which sources of variation are legitimate and which are not? How should the range of values be determined?

9. Viewed as a whole, can the results of studies of the monetary value of transport safety be trusted? Do the results of these studies show true valuations of transport safety, or do they mostly or fully reflect methodological shortcomings of the studies, or, more fundamentally, that the phenomenon these studies aim to study does not exist?

10. Given the fact that the studies reported so far on the monetary valuation of transport safety have produced an extremely wide range of estimates, one must ask: Are there alternative approaches to valuation that are likely to produce less divergent estimates? Which alternative approaches can be applied? What are the strengths and weaknesses of these approaches?

The first point on this list will be discussed in Chapter 2. It will be argued that although it is possible to develop effective transport safety policies without resorting to a monetary valuation of transport safety, an implicit monetary valuation is inevitably made when developing policy. Rather than leaving this valuation implicit and unspoken of, making it explicit can help in developing more effective policies than those that are not based on an explicit monetary valuation. The ways in which an explicit monetary valuation can inform policy making are described.

The second point, dealing with the concept of risk, its dimensions and its measurement is discussed in the Chapter 3. Studies of the monetary valuation of changes in risk asks people to assign a value to such changes, thereby treating changes in risk as a commodity to which standard demand theory can be applied. What reasons have people got for treating changes in risk, in particular reductions of it, as something they ought to spend money on? Different academic disciplines have developed quite different perspectives on risk. Some of these perspectives argue that risk is difficult, if not impossible, to meaningfully quantify at the individual level. If one accepts this point of view, changes in risk cannot easily be quantified the way most valuation research assumes.

Together, Chapters 2 and 3 define and discuss the societal and epistemic context within which valuation research has taken place. This context has clearly influenced the course of this research. However, in order to explain why valuation research has continued despite its many problems, it is not sufficient to describe the societal context. Quite the opposite, many, perhaps most, people who are not themselves engaged in valuation research regard

this type of research as meaningless. Had their opinion prevailed, valuation research might never have started or been given up long ago. Yet, it continues. Research is often strongly influenced by norms that are internal to the scientific community, i.e. by what researchers who are active in a field regard as appropriate topics for study and appropriate methods for studying these topics.

Points 3, 4 and 5 on the list above are dealt with in Chapter 4. That chapter both introduces a theory of science that may help explain the history of valuation research, and the formulation of the theoretical foundation for valuation research by some prominent economists. The theory of science which is introduced is the methodology of scientific research programmes, proposed by Imre Lakatos (1968, 1970, 1971, 1978). This is a theory of science intended to help in a rational reconstruction of its history. The methodology of scientific research programmes is unique by explaining how a field of study can proceed despite many results that apparently contradict the theoretical foundations of research.

Chapter 5 – point 6 on the list above – shows how hypotheses about systematic variation in willingness-to-pay for changes in fatality risk can be interpreted as forming a "protective belt" for this research. A protective belt is a key concept in the methodology of scientific research programmes, explained in Chapter 4.

In subsequent chapters, the methodology of scientific research programmes will be applied as a frame of reference for interpreting and structuring the history of research on the monetary valuation of transport safety. Chapter 6 describes the progressive phase of valuation research. This was the period roughly from 1980 to 1995 when the research programme was launched, attracted researchers and produced results that were, at the time, regarded as encouraging. The next chapters, 7 and 8, describe the increasing problems faced by valuation research and the attempts to solve them.

Chapter 9 discusses attempts to make sense of the results of valuation research by performing meta-analyses of the results of this research – point 7 on the list above. It is concluded that meta-analysis is only partly able to explain the huge variation in estimates of the value of a statistical life. Chapter 10 (points 8 and 9 on the list) discusses whether valuation as a scientific research programme has come to an end, or during the course of its development undergone changes that have changed its basic objective and intended application.

Chapter 11 discusses alternative approaches to the monetary valuation of transport safety (point 10 on the list). Finally, Chapter 12 summarises the main conclusions of the study.

2   The societal context

2.1   The inevitability of trade-offs and the impossibility of infinite values

If one asks a person how much money he or she would demand in order to give up his or her life, the person will most likely react by taking the question as somewhat insulting, but then say: No amount of money could make me give up my life. In that sense the value of life is infinite. We cannot, except perhaps for those who are suicidal, terminally ill, or live in extreme poverty, be bribed to die voluntarily.

The matter is somewhat different when it comes to saving life. How much are you willing to pay for a life-saving operation? Well, essentially as much as you possibly can without having to live in great poverty and discomfort after the operation. You may certainly be willing to pay more than your annual income; any amount up to the maximum size of a loan you could service after the operation could be acceptable. Only if the operation cost more than the maximum amount of money you could bring forward would you have to forgo it.

Thus, maximum willingness-to-pay is constrained by the ability to pay. This is no different at the societal level than it is for an individual. Even if the entire gross national product was spent to save a single life, it would still be a finite amount. In that sense, life does have a finite value. Indeed, the idea of infinite values cannot make sense as long as the resources available to protect these values are limited. This, of course, does not mean that all trade-offs are allowed or possible to make. However, prohibiting certain trade-offs does not imply that values are infinite or resources unlimited.

It is, for example, illegal to trade your right to vote in a public election. You cannot sell the right to vote to your underage daughter because she takes a keen interest in politics and you do not care about voting. The trade is not allowed. But does the right to vote therefore have an infinite value? No, it does not. Like any human right, upholding it comes at a cost and there are probably limits to how much of its resources society can commit to upholding the right to vote.

The purpose of assigning a monetary value to human life is not to engage in trading in the usual sense of that term. It is simply to provide a guideline with respect to the amount of resources we would like to spend on the prevention of accidents or injuries, given the fact that not all of our resources can be spent for this purpose. Some form of economic reasoning – that is some form of thinking that recognises the fact that resources are limited and can be put to very many alternative uses – is simply inevitable, given the following basic facts (Elvik 2012):

1. A limited amount of resources is at our disposal for the prevention of accidents or injuries, or indeed for catering to any human need.

2. Human needs and value systems are complex and multi-dimensional. While safety is certainly one of the more basic human needs, it is not the only one, and no society would ever be able to spend more than a fraction of disposable resources on the prevention of accidents or injuries.

3. How much to spend on the prevention of accidents or injuries will depend, and ought to depend, on how important people think this good is, seen in relation to all other goods they would like to see produced.

4. It is, in principle, possible both to provide too little safety and to provide too much of it. The objective of monetary valuation and cost-benefit analysis is to help us find the right balance between safety and other goods.

If these observations are accepted as a fair description of the choices we are facing, then some kind of cost-benefit reasoning, although not necessarily formalised, is simply inevitable: We engage in this sort of thinking whether we are conscious of it or not. In short: Trade-offs are inevitable; resources are limited; the number of uses resources can be put to virtually unlimited; and different values are compared to each other all the time.

It does not follow from these observations that trade-offs have to be made in monetary terms or that everything can be meaningfully converted to a monetary scale. Thus, one can adopt, for example, an air pollution standard stating the maximum permitted concentration of certain pollutants in air. The lower the limits, the higher is the priority given to clean air. Yet, any limit implies a trade-off. By the same token, one may set a certain target for the maximum number of traffic fatalities. Reaching the target has a certain cost, which indicates the priority given to reducing traffic fatalities. One does not have to convert the reduction in the number of traffic fatalities to a monetary value, although such a value will be implied by the ratio of the cost of reaching the target to the number of fatalities prevented (i.e. the benefit of preventing a fatality must be valued at least as high as the cost of doing so).

The question of whether a monetary valuation of transport safety is needed in order to develop effective policy is discussed in the next section. Following that, the arguments economists have made in favour of an explicit monetary valuation of safety are presented. An example of inefficient priorities is then given. Finally, it is noted that economic theory actually speaks with more than one voice as far as standards of consistency and efficiency in priority setting are concerned, and that the efficiency argument in the form it was originally put by economists to justify the monetary valuation of safety represents just one of several norms of consistency and efficiency proposed in economic theory. Herein lies the germ of contradictions that lay dormant for a long time, but in the end surfaced and lead some researchers to propose a reformulation of the chief objective of valuation research.

2.2   Is monetary valuation needed for making trade-offs?

While making trade-offs, in the sense of choices about how much to spend on, for example, road safety, health care, primary school, national defence, etc. is an inevitable part of public policy, it does not follow that these trade-offs have to be made by relying on an explicit monetary valuation of the different objectives. Indeed, no meaningful monetary valuation exists, or is relevant, for deciding how much to spend on primary education. In modern, western societies, it is regarded as a human right not to be illiterate. The question is never asked whether the benefits of learning children to read and write exceed the costs of doing so. So why should cost-benefit analysis be used to set priorities for road safety? Can we manage without it?

One option is to use of cost-effectiveness analysis to help set priorities between road safety measures. In cost-effectiveness analysis, no monetary value is assigned to safety effects. These effects are stated in "natural units", i.e. the number of accidents, fatalities and

injuries prevented by a road safety measure or set of measures. The less a road safety measure costs per fatality or injury prevented, the more cost-effective it is.

Cost-effectiveness gives sufficient information for setting priorities between road safety measures when the following two conditions are fulfilled (Hauer 2011B):

1. Either two road safety measures, A and B, are expected to prevent accidents of the same severity, or one of the measures dominates the other.
2. The question of when a road safety measure becomes "too expensive" does not arise.

Suppose that measures A and B cost the same. If A (as a long-term statistical average) prevents 5 injuries and 1 fatality, it will dominate B if B only prevents 3 injuries and 0 fatalities. If, on the other hand, B prevents 4 injuries and 2 fatalities, the choice is no longer obvious. It depends on what we think is most important (or "valuable" an economist might say): Preventing fatalities or preventing injuries. There is probably consensus that it is more important to prevent fatalities than to prevent injuries. But how much more important? To help answer this question, a widely applied weighting scheme in the United States is the EPDO, or Equivalent Property Damage Only weight. A case of property damage is given the weight of 1. Larger weights are given to injury accidents and fatal accidents, reflecting how much more important it is to prevent these accidents than to prevent a property-damage-only accident. Table 2.1 shows the weights assigned in some American States, as well as the weights resulting from a monetary valuation of injuries of different severities (based on Miller 1993 and Hauer 2011B).

Table 2.1: Relative weights assigned to preventing accidents of different severity in some states of the United States. Based on Miller 1993 and Hauer 2011B

State	Property damage	Injury accident	Fatal accident
Colorado	1	5.4	154.0
Massachusetts	1	5.0	10.0
North Carolina	1	8.4	76.8
Ohio	1	6.9	292.8
All states (Miller 1993)	1	13.7	1053.9

It is seen that the weights vary considerably. Hauer (2011B:3) remarks:

"It is hard to believe that road users in Ohio would value fatalities 29 times more than in Massachusetts. Whether one fatal accident is equivalent to 10 PDO accidents or to 292.8 accidents will determine which of several alternative actions appears to be more cost-effective."

To this can be added that none of the weights applied by the states listed in Table 2.1 are anywhere close to the weights based on the monetary valuation of preventing fatalities and injuries based on willingness-to-pay (Miller 1993, Tables 6 and 8).

Hauer (2011B) notes that cost-effectiveness analysis can never determine whether spending public money can be justified. It does not define a "cost limit" beyond which a safety measure is regarded as too expensive. In practice, as shown by the study of Tengs et al. (1995), see further details in section 2.4, such a limit does not exist. The amounts spent per fatality prevented vary enormously and erratically.

While it is certainly possible to develop public policy without resorting to any monetary valuation of human life and limb, such a valuation can support policy in three ways that are not possible without a monetary valuation:

1. Monetary valuation of public policy objectives makes it easier to compare different objectives to each other and find solutions that maximise the overall realisation of the objectives when they are partly or fully conflicting. When all objectives are stated in the same metric (money), they are made comparable.
2. A uniform monetary valuation of life and limb makes it easier to set policy priorities that maximise the number of fatalities or injuries prevented with a given budget.
3. Monetary valuation of life and limb makes it possible to determine how much to spend in total on the prevention of fatalities or injuries.

The reader of this study is not asked to agree with these arguments or find them persuasive. The objective of this study is not to persuade readers about the blessings of a monetary valuation of life and limb. It is rather – given the fact that many economists have argued in favour of a monetary valuation of life and limb, and tried to obtain this valuation – to try to explain why a field of research many observers would say has failed utterly to realise its purpose has continued to thrive and grow despite the apparent lack of success.

2.3 The consistency argument in favour of a uniform value of a statistical life

The consistency argument in favour of a uniform monetary valuation of life and health was forcefully put by Hills and Jones-Lee (1983). Their argument is worth quoting at length. They use examples of policy choices to illustrate their points. The first example is given in Table 2.2.

Table 2.2: Impacts of two transport projects. Only one can be chosen. Based on Hills and Jones-Lee, Table 1

	Investment cost	Annual savings in vehicle operating costs	Annual reduction of fatalities
Project A	5,000,000	450,000	1
Project B	5,000,000	150,000	4

A policy maker choosing project A reveals that his or her valuation of saving a life must be less than 100,000 – otherwise project B would be better. Conversely, a policy maker choosing project B reveals that his or her valuation of saving a life must be at least 100,000. Unless the monetary valuation of saving a life is made explicit, choices between options such as A and B in Table 2.1 are likely to be inconsistent. On one occasion, A may be chosen. On a different occasion, B may be chosen. The study by Tengs et al. (1995), quoted in Chapter 1, shows that this is indeed the case. In general, that means that society does not get as large safety benefits from spending a given amount of money as it could by spending the money efficiently. If a monetary valuation of, for example, 125,000 was adopted for saving a life, cost-benefit analysis would always find that project B is better than project A.

To maximise benefits, it is important that the value of saving life is uniform, i.e. only a single, constant value is applied. The next example shows this. It is given in Table 2.3.

Table 2.3: Comparison of safety investment in rail transport and bus transport. Based on Hills and Jones-Lee, table 2

Rail transport			Bus transport		
Fatalities per million personkm	Expected number of fatalities	Annualised cost (1,000) of reducing risk	Fatalities per million personkm	Expected number of fatalities	Annualised cost (1,000) of reducing risk
1.0	10	0	4.0	40	0
0.9	9	15	3.0	30	80
0.8	8	40	2.0	20	250
0.7	7	65	1.5	15	420
0.6	6	100	1.2	12	590
0.5	5	160	1.0	10	750
0.4	4	220	0.9	9	860
0.3	3	350	0.8	8	1000
			0.7	7	1180
			0.6	6	1420

It is assumed that the two modes transport the same number of people. Rail is much safer than bus (lower number of fatalities per million personkm). Suppose, first that a safety standard has been set allowing a fatality rate of not more than 0.9 for each mode. Achieving this level would cost 15 for rail and 860 for bus, for a total of 875. It would prevent 1 fatality in rail and 31 for the bus, in total 32. The mean cost per prevented fatality would be 875/32 = 27.3. The marginal cost (the cost of the last fatality prevented) is (860-750)/(10 – 9) = 110/1 = 110. Closer inspection of the data suggest that this would not be an efficient use of money. The marginal cost of preventing one fatality is 15 (the difference between 15 and 0) for rail and 110 (860 – 750) for bus. This suggests that one could prevent more fatalities per unit of money spent (and possibly more in total) by shifting spending from bus to rail.

If an equal maximum marginal cost of 60 per fatality prevented is assumed (equivalent to a uniform monetary valuation of preventing a fatality of 60), one should aim for 5 fatalities in rail (marginal cost 60 = 160 – 100) and 12 fatalities for the bus [(590 – 420)/3 = 56.7)]. The number of fatalities prevented would be 5 in rail and 28 in bus, in total 33. Total cost would be 750, which is less than if the safety standard was introduced. Moreover, if the maximum marginal cost of 60 is interpreted as the monetary value of preventing a fatality, net benefit for rail would be 140 (300 – 160) and for bus 1090 (1680 – 590).

Suppose next that rail is safer because safety has been valued more highly there than for the bus, i.e. differing monetary valuations have been applied. For simplicity, suppose the relative valuation of safety is inversely proportional to risk. This means that preventing a fatality is valued four times higher for rail than for the bus. If, say, the valuation is 100 for rail and 25 for the bus, rail should reduce the number of fatalities to 4 (going further down to 3 has a marginal cost of 130, exceeding the value of the benefits). The bus should aim for 20 fatalities. The total number of fatalities prevented would be 26 for a total cost of 470.

Finally suppose that a uniform monetary valuation of safety of 100 is used for both modes of transport. The optimal levels of safety would be at 4 fatalities in rail and 10 for the bus. Total number of fatalities prevented would be 36 for a total cost of 970.

The following lessons can be learnt from this simple numerical example. First, setting a safety standard, or a quantified target for that matter, without considering what it costs to meet the standard is likely to generate an inefficient use of resources, since the marginal

costs of achieving the safety standard are likely to vary between different organisations or types of activity subject to the standard. Whenever marginal costs vary, one may in principle increase efficiency by shifting spending to equalise marginal costs.

Second, setting priorities according to a uniform valuation gives a more cost-effective solution than adopting the safety standard. The cost is 22.7 per fatality prevented (750/33) when a uniform valuation of 60 is adopted versus 27.3 (875/32) when the safety standard is adopted.

Third, if a uniform monetary valuation of life-saving is adopted, it will be most efficient to prevent fatalities when the cost of doing so is low; hence, mean expenditure per prevented fatality will be minimised. In short, if improving safety is very costly (costs more than the valuation of safety), one should shift resources to areas where the prevention of a fatality costs less.

Thus, a uniform monetary value of saving a life supports an analysis designed to find the least costly way of preventing fatalities. This set of priorities will maximise the number of fatalities that can be prevented within a given budget. Departure from a uniform value of saving a life will, all else equal, result in a lower number of lives saved. It seems clear from the examples discussed by Hills and Jones-Lee (1983) that they regarded obtaining a uniform monetary valuation of saving a life as the primary objective of research on the benefits of preventing accidents.

2.4   The efficiency of priorities matters in practice

Inefficient priorities have real costs in terms of saving fewer lives than one would otherwise do. Tengs and Graham (1996) estimated the opportunity cost of haphazard investments in life-saving for 185 life-saving interventions in the United States for which data could be obtained at the national level of their cost of implementation and the number of lives and life-years each intervention would save. For each intervention, a variable called degree of implementation was defined. This variable was defined as the percent of people in the target group for the intervention for whom the safety intervention had been implemented as of 1992. Figure 2.1 shows the relationship between the degree of implementation and the cost of the intervention.
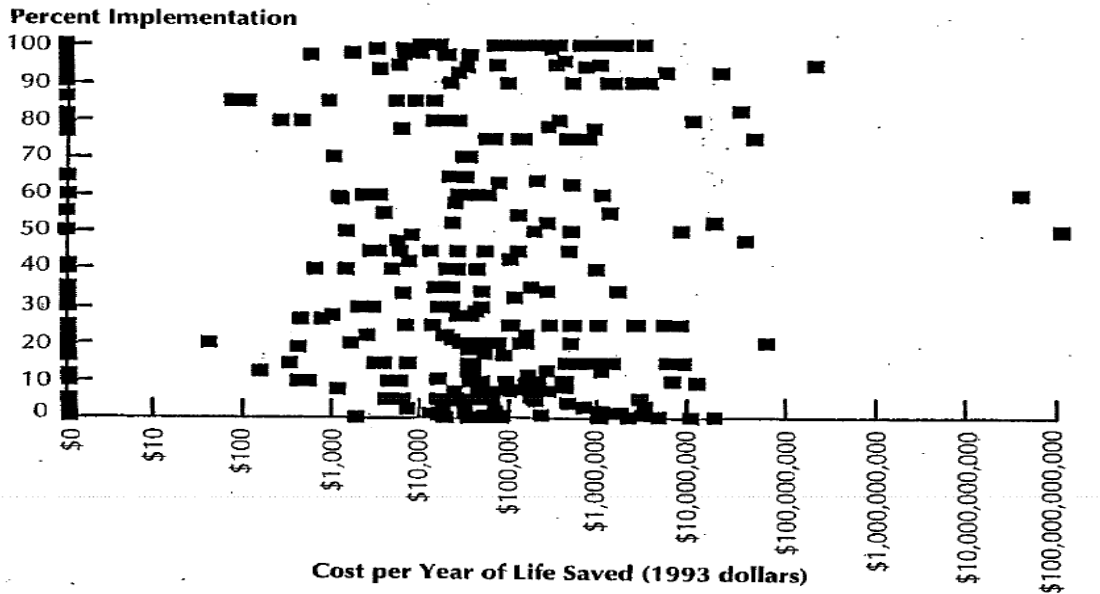
Figure 2.1: Degree of implementation and cost of 185 life-saving interventions. Taken from Tengs and Graham (1996), Figure 8.1

It can be seen that the cost-effectiveness of the life-saving interventions varied enormously, from a negative cost to a cost of 100 billion US dollars per life-year saved. Costs were negative when an intervention had other benefits than life-saving which were greater than the cost of implementing the intervention. The data points are scattered all over the diagram. There is no relationship between cost-effectiveness and degree of implementation. The total annual cost of all safety interventions listed in Figure 2.1 was estimated to 21.4 billion US dollars. Annual effects were estimated to the saving of 56,700 lives and 592,000 life-years. Tengs and Graham (1996) conducted two analyses in order to estimate the opportunity cost of these haphazard priorities. The first of these sought to answer the question: How many lives or life-years can be saved annually if we spend the same total amount of money as today (21.4 billion), but spend it exclusively on the most cost-effective interventions? The second analysis sought to answer the question: How little could we spend, while still saving the same number of lives and life-years as today?

The first analysis found that by setting priorities strictly according to cost-effectiveness, one could, for an annual budget of 21.4 billion US dollars save 117,000 lives per year – more than twice as many as saved by the actual priorities – and 1,230,000 life-years – again more than twice as many as by the current inefficient priorities.

The results of the second analysis were even more staggering. It was found that one could save the same number of lives and life-years as current policies at a net annual cost of close to minus 10 billion US dollars per year. The negative cost was the result of the fact that many of the life-saving interventions had other favourable impacts in addition to life-saving. When these other favourable impacts were valued and taken into account, net cost became negative. Figure 2.2 shows costs and effects according to maximally efficient priorities.
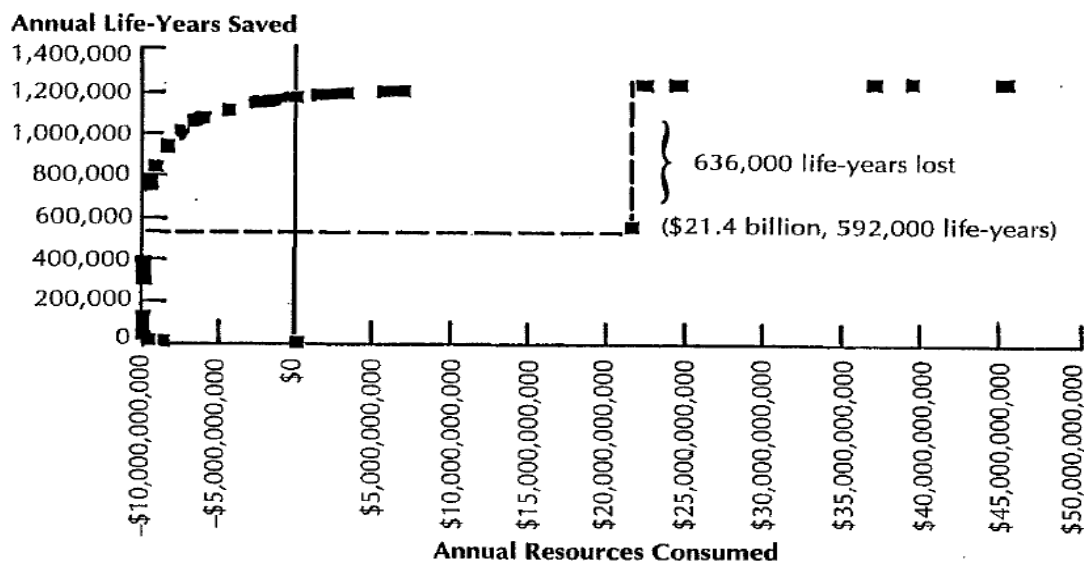
Figure 2.2: Maximally efficient priorities for life-saving interventions. Based on Tengs and Graham (1996), Figure 8.2

Elvik (2003) reported analyses along similar lines assessing the efficiency of road safety policies in Norway and Sweden. For both countries, it was found that if priorities were set strictly according to cost-benefit analyses, considerably greater reductions in the number of road accident fatalities would be achieved than if current inefficient priorities were continued.

To analyse the efficiency of policies is therefore not merely an academic exercise with no practical implications. Basically, inefficient priorities lead to unnecessary loss of life. A uniform monetary valuation of life and health makes the problem of finding efficient priorities analytically tractable and provides a basis for assessing what the costs, in terms of lives lost, of inefficient policy priorities are.

2.5   Multiple standards of consistency in economic theory

While efficiency in priority setting historically has been a main argument for the monetary valuation of transport safety, it is worth noting that multiple standards of consistency and efficiency in the monetary valuation of transport safety can be derived from economic theory and from the theory of willingness-to-pay for safety. The relationship between these many standards of consistency will be discussed more in detail in Chapter 10 of the report. At this point, the standards of consistency will merely be noted for the sake of completeness.

1.   Consistency (maximum efficiency) in priority setting

Consistency in priority setting means that policy priorities are based on a uniform monetary valuation of safety ensuring full comparability between alternative projects and that priorities are set so as to equalise marginal benefits with marginal costs.

2.   Consistency with demand

Consistency with demand means that the provision of safety should match the aggregate demand for it. The demand for road safety can be derived from willingness-to-pay by estimating the value of a statistical life. Consistency with demand means that if the value of a statistical life is found not to be the same for all types of risk, the provision of safety should be differentiated accordingly.

3. Consistency with the Pareto criterion

The most widely applied criterion of welfare improvement in modern economic welfare theory is a potential Pareto improvement. This criterion states that those who gain from a project should be able to compensate those who lose from it and still retain a net benefit. However, examples can be given of situations in which a potential Pareto improvement cannot be realised and consistency with the Pareto criterion thus not be attained. Such an example is discussed in Chapter 7 of the report.

4. Consistency with majority preferences

To achieve democratic legitimacy, policy priorities should be supported by a majority. Willingness-to-pay typically has a skew distribution, which means that mean willingness-to-pay may be higher than median willingness-to-pay. A safety budget based on market demand (mean willingness-to-pay multiplied by the number of people) may then be voted down by the majority as being too large.

5. Consistency with individual preferences

Consistency with individual preferences means that one seeks to match the provision of safety exactly to individual demand for it, in analogy to the provision of market goods. Since individual preferences and willingness-to-pay vary, matching individual demand for safety will not result in a uniform monetary valuation of it.

6. Consistency between ex ante and ex post

Safety project are normally evaluated ex ante, i.e. before they are implemented. Consistency between ex ante and ex post means that one should accurately predict ex post evaluation of the project. If such a consistency is not present, one may perpetually regret choices and want to reverse them.

7. Consistency with individual welfare

Consistency with individual welfare means that welfare (utility) should be improved according to the compensation test. Welfare is normally defined in utility terms, which means that in order to assess utility, one needs to know the marginal utility of money. A measure only improves welfare if a hypothetical evaluation of a change in welfare coincides with the actual change in welfare once real (as opposed to hypothetical) payments have been made for the measure.

The potential for conflicts between these standards of consistency will be discussed in greater detail in Chapter 10 of the report.

2.6 Conclusions

The main conclusions that can be drawn from the discussion in this Chapter are:

1. Making trade-offs between improving transport safety and other policy objectives is inevitable.
2. Assigning a uniform monetary value to human life can help policy makers compare conflicting policy objectives, set priorities so as to maximise the number of fatalities prevented within a given budget, and determine the optimal size of the budget.
3. Policy priorities that are not informed by a monetary valuation of human life have been found to be haphazard and wasteful.

3   Perspectives on risk and the valuation of it

The economists who launched research on the monetary valuation of road safety as a research programme, briefly mentioned in Chapter 1, defined changes in the risk of dying as the good to be valued. Risk, however, is not a concept with a standard definition and even a specific risk, such as the risk of dying in a road accident, can be operationally defined in many ways. This chapter briefly reviews different perspectives on risk and the monetary valuation of changes in risk. The perspectives discussed will be drawn from the main academic disciplines that have studied risk.

3.1   Economic perspectives

The economists calling for research in order to assign a monetary value to life-saving were careful to point out that they were not talking about a particular life, but about reducing the risk of death. Thus, Schelling wrote in the first paragraph of his 1968-paper (Schelling 1968:127):

"It is not the worth of human life that I shall discuss, but of "life-saving", of preventing death. And it is not a particular death, but a statistical death. What is it worth to reduce the probability of death – the statistical frequency of death – within some identifiable group of people …?"

In economic theory, the modern analysis of risk goes back to the seminal work of Von Neumann and Morgenstern (1953). Based on their theory, a distinction has been made between four levels of knowledge regarding the consequences of decisions:

1. Decision making under certainty: All consequences (potential outcomes) of a decision are known with certainty.
2. Decision making under risk: The probability of all potential outcomes of a decision are known.
3. Decision making under uncertainty: The potential outcomes of a decision are known, but their probabilities are unknown.
4. Decision making under ignorance: Neither all potential outcomes of a decision nor their probabilities are known.

Decision making under ambiguity (Ellsberg 1961) is a hybrid category: Under ambiguity, some potential outcomes of a decision have known probabilities, other potential outcomes have no known probabilities. Thus, the concept of risk in economic theory denotes the probability of occurrence of a potential outcome of a decision. It should be noted that this is just one definition of risk found in the scientific literature. Definitions of risk in other academic disciplines are discussed in subsequent sections.

The distinction between risk and uncertainty is not always clear. It may be the case that probabilities are imprecisely known, for example, estimated to be in the range between 0.1 and 0.3. Are we then in the domain of risk or in the domain of uncertainty? Valuation research aims to estimate monetary values of changes in risk. This is only possible if risk can be estimated numerically. However, is it reasonable to think that, for example, a car

driver knows his or her risk in numerical terms or is at all able to provide a numerical estimate of risk? It is unlikely that any driver will have a numerical notion of his or her risk. The task posed to a driver in order to obtain a monetary valuation of a change in risk will therefore be both unfamiliar and abstract.

Today, the most widely accepted and parsimonious theory of rational choice under risk in economics is the subjective expected utility model (Raiffa 1968, Simon 1983). According to this model, an individual is rational if he or she acts so as to maximise his or her expected utility. Utility is usually defined as preference satisfaction. Expected utility denotes the expected value of a lottery, i.e. the probability-weighted mean value of a set of potential outcomes to each of which a utility value is attached. According to the theory of subjective expected utility, probabilities are defined as degrees of belief, i.e. the probabilities are subjective and do not reflect any "objective" reality.

The (subjective) expected utility framework is readily applicable to studies designed to obtain monetary valuations of changes in the risk of dying. The word subjective was put in parenthesis, as statistical (frequentist; see section 3.4 for a discussion of the difference between frequentist and subjective estimates of risk) estimates of the risk of death are widely available and may serve as the basis for valuation studies.

The expected utility model is eminently suited to form the hard core for developing a theory about willingness-to-pay for reduced risk of death. Utility is usually modelled as a function of income or wealth. A distinction is commonly made between two types of risk (Jones-Lee 1989): Financial risk and physical risk. Financial risk refers to the possibility of economic losses, such as loss of income or property. It is often possible to insure against financial losses, for example by insuring your car or house. In principle, even income can be insured by means of a so called annuity, which is a fixed amount paid annually for an agreed period.

Physical risk is the risk of death or loss of health. It is often possible to insure against financial losses that are associated with death, illness or injury, but the loss of welfare (utility) associated with health risks cannot be eliminated by means of insurance. In subjective utility theory, the loss of welfare associated with death or reduced health is usually modelled by making the utility function state-dependent, i.e. dependent on the outcome. There is one utility function conditional on survival, another utility function conditional on death or on a reduced state of health. Economic theory assumes that the primary motivation for wanting to pay for a reduced risk of death or ill health is to avoid the loss of welfare associated with this risk. Thus, Schelling (1968:159-160) wrote:

"The difficult part of the problem is not evaluating the worth of a man's livelihood to the different people who have an interest in it, but the worth of his life to himself or to whoever will pay to prolong it. This is what is not insurable in terms that permit replacement. This is the consumer interest in a unique and irreplaceable good. His livelihood he can usually insure, not exactly but approximately, sharing the loss and making it a matter of diffuse economic interest; it is valuing his life that poses the problem."

In some utility models, a positive utility is assumed conditional on death. It is important to be clear about the interpretation of a positive utility conditional on death. First, it refers to the utility of financial assets only, not to any subjective well-being an individual might experience once dead. Second, it refers to bequest motives only, i.e. to the well-being an individual, while alive, gets from knowing that he or she will leave behind an estate to his or her descendants. It is important to note these points, because, in principle, it is possible to purchase life insurance up to the point when utility conditional on death equals utility conditional on being alive. From a strictly formal point of view, an individual would then be indifferent between life and death. But this is an anomaly of utility theory as a formal tool of analysis. Gary Fromm (1968:167-176) has shown how a closed-form expression for the monetary value of reducing the risk of death can be derived. The following

presentation draws both on Fromm and on Bellavance et al. (2009). The starting point is that expected utility is:

Expected utility = EU(w) = $(1 - p)U_a(w) + pU_d(w)$ (1)

Here, p denotes the probability of dying, w is wealth (or income; for the moment the two are treated as interchangeable), subscript a denotes that the individual is alive and subscript d denotes that the individual is dead. It is normally assumed that the individual prefers life to death; hence his or her utility from wealth will be greater when alive than when dead:

$$U_a(w) > U_d(w)$$ (2)

Wealth can be assumed to be the same both when the individual is alive and when he or she is dead, provided the individual can buy insurance that covers all financial losses. It is often assumed that the marginal utility of wealth is greater when alive than when dead:

$$U'_a(w) \geq U'_d(w) > 0$$ (3)

Here the prime (') denotes the first derivative. Another property of utility functions commonly assumed in economic theory is risk aversion. Risk aversion means that marginal utility is strictly decreasing both in case of life and in case of death:

$$U''_a(w), U''_d(w) < 0$$ (4)

The double prime (") denotes the second derivative. A utility function for which the first derivative is positive and the second derivative is negative is called strictly increasing and concave. It rises throughout the range, but at a slower and slower rate (becomes flatter). In the following, it is assumed that w is the same in life and death. This need not be the case if an individual has life insurance for a different amount than w.

The optimal amount to pay for a reduced risk of death is the amount x an individual would give up to reduce p to p* while keeping expected utility constant. This is the amount x that satisfies the following equality:

EU(w) = $(1 - p)U_a(w) + pU_d(w) = (1 - p^*)U_a(,w - x) + p^*U_d(w - x)$ (5)

If x is paid, the loss of utility of income resulting from the fact that it is reduced by x is exactly offset by the gain in expected utility from the fact that p is reduced to p*. The optimal amount to pay for a reduction in risk is the marginal rate of substitution between wealth (income) and risk of death. This marginal rate of substitution, which is identical to the value of a statistical life, is found by finding the derivative of the left side of equation 5 (i.e. expected utility in the initial situation) with respect to both w and p, while holding expected utility constant. This yields (Bellavance et al. 2009:446):

VSL = marginal rate of substitution = $\dfrac{dw}{dp} = \dfrac{U_a(w) - U_d(w)}{(1-p)U'_a(w) + pU'_d(w)}$ (6)

The numerator represents the difference in the utility of wealth between life and death. If the utility of wealth in death is zero, the numerator becomes identical to the utility of wealth in life. The denominator represents the marginal expected utility of wealth. The prime denotes the first derivative.

The idea underlying the equality in equation 5 is that the individual maximises expected utility. Hence, the left hand side represents the maximum of expected utility in the current situation. A change involving wealth and the probability of death will therefore only be attractive if it maintains the utility maximum. This condition is fulfilled when the loss of income incurred in paying for a reduction of risk is exactly offset by the gain in expected utility attributable to the fact that death becomes a less likely outcome.

We can see that the marginal rate of substitution given in equation 6 is identical to the value of a statistical life by noting that dw (the numerator) is the (small) amount you pay for a risk reduction and dp (the denominator) is the change in risk. Thus, we have, as shown in Chapter 1 (the numbers are arbitrary):

$$\text{VSL} = \frac{500}{\left(\frac{2}{100000}\right)} = 25{,}000{,}000 \text{ in any currency}$$

The shape of a utility function is often described in terms of its degree of risk aversion. There are two standard measures of risk aversion (Pratt 1964, Arrow 1965): the coefficient of absolute risk aversion and the coefficient of relative risk aversion. These measures are defined as follows:

$$\text{Coefficient of absolute risk aversion} = \frac{-u''(w)}{u'(w)} \tag{7}$$

The double prime denotes the second derivative of the utility function with respect to wealth, the single prime denotes the first derivative. The coefficient of relative risk aversion is defined as follows:

$$\text{Coefficient of relative risk aversion} = w \cdot \left(\frac{-u''(w)}{u'(w)}\right) \tag{8}$$

A simpler indicator of the degree of risk aversion is the ratio of the certainty equivalent of a utility function to the expected value of a lottery. The lower the certainty equivalent, the stronger the risk aversion. To help clarify these ideas, a numerical example will be given. Figure 3.1 shows two utility functions for income displaying different degrees of risk aversion.

To see the difference between the two utility functions with respect to risk aversion, consider a lottery offering an annual income of 100 with probability 0.5 and an income of 900 with probability 0.5. The expected value of annual income is 500. A risk-neutral individual would be indifferent between this income and the lottery. However, since both utility functions display risk aversion, the certainty equivalent for income is lower than the expected value of the lottery. For the lower utility function, it can be found that an annual income of 399 would make an individual indifferent between the lottery and an income of 399 with certainty. For the upper utility function, the corresponding amount is 300. An income of 300 with certainty is equivalent to the lottery between 100 and 900. The lower the certainty equivalent, i.e. the sum obtained for certain which is equivalent to the expected utility of the lottery, the stronger is risk aversion. Chapter 5 of the report reviews theoretical contributions that make use of the concepts introduced here.

Two utility functions for income displaying different degrees of risk aversion
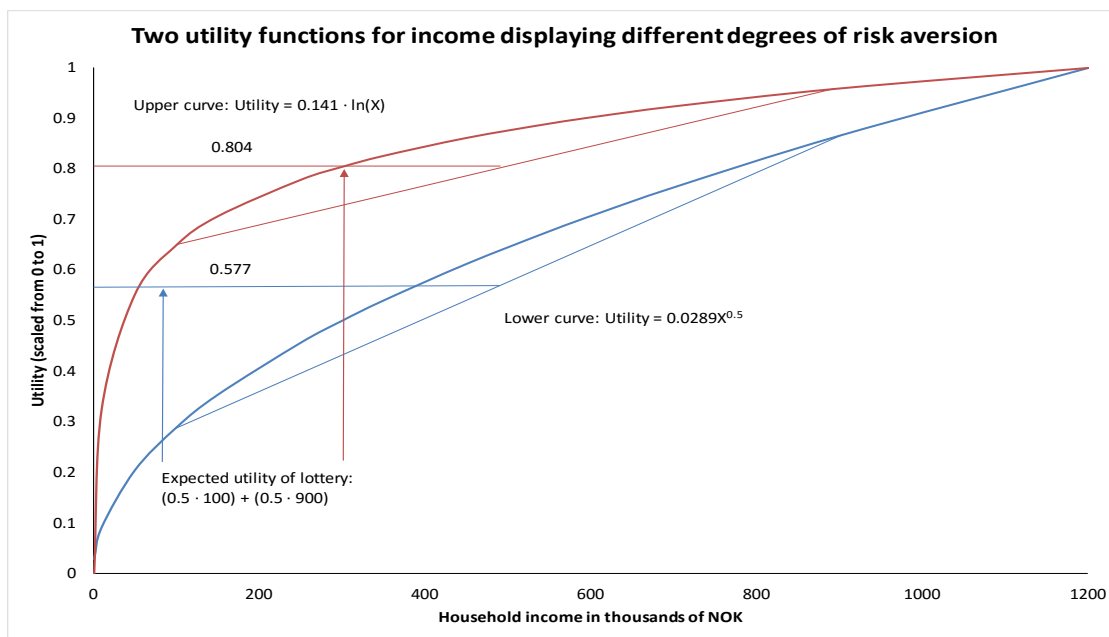
Figure 3.1: Two utility functions for income displaying different degrees of risk aversion.

3.2   Philosophical perspectives

Within the discipline of philosophy, risk has mainly been investigated in the fields of epistemology and moral philosophy. In this section, some philosophical contributions with respect to how to make defensible decisions about risk will be reviewed. More specifically, alternatives to the economic perspective on risk will be discussed. Hammond (1982:97), in discussing the distinction between the evaluation of risk ex-ante and ex-post, remarks the following:

"These are matters which are obviously far from exclusively economic questions, yet in deciding how much to spend on road safety measures, for example, economic and social consequences become closely linked. And although the utilitarian welfare economist's approach to such questions may seem far from ideal, it is not at all obvious how the approach can be improved upon.

The contrast between ex-ante and ex-post is stark in the case of such choices because, ex-post, one is trading off lives against ordinary economic resources whereas, ex-ante, one is merely trading off probabilities of death against economic resources. The latter seems much more comfortable …

However, I have argued that the ex-ante approach to utilitarianism under uncertainty is only appropriate when it coincides with the ex-post approach, and that this coincidence is rather unlikely. Thus I am arguing that one should use the ex-post approach consistently, even in matters of life and death. … There remains the issue of whether and how individual attitudes to risk are to be allowed for in the ex-post welfare function, short of making it coincide with the ex-ante welfare function. One might, for example, include ex-ante utility in each individual's ex-post utility function."

Hammond does not discuss the implications of these points of view with respect to the monetary valuation of changes in risk. He does not altogether reject a monetary valuation of changes in risk. As far as the risk of dying is concerned, ex-ante refers, as Hammond notes, to the evaluation of risks; i.e. probabilities of dying. Ex-post, on the other hand, can only refer to a situation in which the risk has materialised in the form of a certain number of deaths. As usually conceived of in economic theory, utility conditional on death refers to bequest motives. Standard utility models (see section 3.1) incorporate risks to health by making the utility function state-dependent, i.e. by defining one utility function for perfect

health, one (or more) for reduced health and one for death. If there are no bequest motives, utility in case of death may be zero, and the ex-post valuation of life, i.e. the compensation needed to restore ex-ante utility will be infinite. Ex-post compensation can only be finite if there is a positive marginal utility of wealth.

Hammond suggests applying the ex-ante utility function. He presumably refers to ex-ante expected utility, i.e. the probability-weighted average of utility conditional on life and utility conditional on death. It is difficult to make sense of the statement about including the ex-ante utility function in the ex-post utility function. Ex-post, the individual is dead and no utility function can be determined. The only way of determining a utility function having death as a possible outcome is to do it ex-ante. Thus, the meaning of the ex-post approach as advocated by Hammond is far from clear.

Hansson (2007) discusses three approaches to risk and ethics. He states that an ethical analysis of risk can be performed by answering the following seven questions (2007:28):

1. *To what extent do the risk-exposed benefit from the risk exposure?*
2. *Is the distribution of risks and benefits fair?*
3. *Can the distribution of risks and benefits be made less unfair by redistribution or by compensation?*
4. *To what extent is the risk exposure decided by those who run the risk?*
5. *Do the risk-exposed have access to all relevant information about the risk?*
6. *Are there risk-exposed persons who cannot be informed or included in the decision process?*
7. *Does the decision-maker benefit from other people's risk exposure?*

These questions identify some of the dimensions of risk that have been found to influence attitudes to it, such as personal benefit (1), fairness in distribution (2, 3), voluntariness of exposure (4), knowledge about the risk (5), possibility of influencing the risk (6) and external effects of the risk (7). It is noted in passing that fairness is a complicated concept with no standard definition. Hansson rejects the use of expected utility theory to evaluate risks. His main argument is that an ex-ante calculation of expected utility may well find that running a risk is worthwhile, but that any such calculation loses legitimacy once an accident happens. One never hears a company defend itself after a major accident by arguing that the benefits of running the risk were greater than the costs. Hansson argues for trying to develop a mode of thinking he calls hypothetical retrospection, i.e. we must try to imagine ex-ante how we would justify decisions about risk after an accident has occurred. This, he argues, is not the same as trying to anticipate regret. It is to think in a way that ensures that, whatever happens, the decision one makes will be morally acceptable (permissible) from the perspective of actual retrospection.

This suggestion comes close to requiring perfect foresight, much in the same manner as the ethical principles of Vision Zero, discussed in Chapter 2. One may inadvertently run a risk which is subsequently found to be unacceptable, simply because the risk was unknown at the time of exposure. Thus, when the flu pandemic threatened in 2009, mass vaccination seemed like a good precaution. It was, however, not known that the vaccine might cause serious side-effects, like narcolepsy in children. Moreover, this side-effect appeared to occur randomly, in the sense that only a few unlucky children among those who were vaccinated developed narcolepsy. Is it reasonable to require that one should have been able to foresee this side-effect? In case the side-effect had been foreseeable, should one then have refrained from mass vaccination? What would have caused the greater harm – allowing the flu pandemic to run its course without mass vaccination, thereby avoiding the side-effects, or doing mass vaccination, thereby preventing potential deaths from the flu, but generating serious side-effects?

No attempt will be made to answer these questions; they are just examples of the difficulties one may run into in trying to practice hypothetical retrospection. It is the nature of many risks that they are not known well enough to know what an "appropriate" level of precaution against them would be. Actual retrospection, call it hindsight bias, learning-by-doing, post-hoc rationalisation or whatever, will always differ from hypothetical retrospection when risks are poorly known. Hypothetical retrospection therefore at best provides vague support and guidance in making decisions about risk control.

Wolff (2007) discusses the monetary valuation of preventing a fatality from a philosophical perspective. He does not reject the idea of a monetary valuation, but points out some problems in obtaining credible estimates of the value of preventing a fatality. Actual behaviour may not reliably "reveal" preferences, if, for example, people buy a hazardous product they erroneously believe is safe. Asking people directly about willingness-to-pay, as is done in contingent valuation studies may also be dubious as: "There is a legitimate worry that some people are simply plucking numbers out of the air, rather than revealing willingness-to-pay" (Wolff 2007:59).

Morton (2007) tries to identify the virtues an individual basing his choices on expected utility would need to have. His point is that any choice between options where different outcomes may occur with a certain probability, always entails the risk that the outcome will be bad. To consistently rely on expected utility as a principle of choice, an individual should prepare for the occasional bad outcome. Morton (2007:96) states it as follows:

"I have mentioned a number of virtues that expected-value choosers should have: the ability to think of one's preferences and one's degree of confidence in numerical terms, the ability to make contingency plans for the inevitable times when a gamble with a high expected value has a low actual one, and the ability to schedule and gather together one's choices for the best overall outcome. If you don't have these virtues, you should stay away from expectational thinking."

This is a criticism of relying on expected value or expected utility as a principle of rationality often made by philosophers. Its relevance to actual choices is limited, first, by the fact that at least common financial risks can be spread by means of insurance and thus no longer involve a potentially ruinous outcome for an individual. Second, some risks that cannot be insured are nevertheless impossible to avoid. Women have to run the risks involved in child-bearing for the human race to survive. Many other everyday risks (such as accidental food poisoning) are also, to all intents and purposes, impossible to entirely avoid. Any choice involving such risks will take the form of a gamble. To adopt a more conservative principle of choice than expected value would sometimes lead to inaction (by rejecting the gamble) whose long term outcome would be worse than the worst outcome of gamble. By not eating, death by starvation is a certain outcome, whereas by eating, the occasional case of food poisoning will rarely be fatal, but involve a less bad worst outcome.

3.3 Psychological perspectives

Psychological research on risk has a long tradition. During the period after the monetary valuation of safety by means of the willingness-to-pay approach was launched by economists (roughly after 1970), the dominant approach within psychology has become known as "the psychometric paradigm" (Slovic 2000). A persistent theme in psychological research on risk has been that the expected utility model of economic theory does not describe how people actually make choices involving risk. Very many experiments have been conducted by psychologists showing the shortcomings of the expected utility model. No attempt will be made in this report to review all these studies, but a few key findings deserve to be mentioned.

In an early paper, Slovic, Kunreuther and White (1974), reviewed the studies available at that time. Most of the studies included in their review were laboratory studies in which subjects were given "artificial" choice tasks intended to test whether they chose rationally. Aware of the objections economists might make to these experiments, they remarked (Slovic 2000:23):

"Finally, the laboratory conclusions are congruent with many observations of non-optimal decision-making outside the laboratory – in business, governmental policy setting and adjustment to natural hazards. The belief that people can behave optimally when it is worthwhile for them to do so gains little support from these studies. The sources of judgemental bias appear to be cognitive, not motivational. They have a persistent quality not unlike that of perceptual illusions."

Fischhoff et al. (1978), in a widely quoted study, asked "How safe is safe enough?" – virtually the same question economists try to answer when asking people about willingness-to-pay for improving safety (an economist would say an activity is safe enough when additional expenditures for making it safer exceed the amounts people are willing to pay for increased safety). In the study 30 different activities were compared with respect to their perceived benefits, their perceived risks, the acceptability of its current level of risk and its position on each of nine dimensions of risk.

Motor vehicles (the context makes it clear that this term refers to cars) scored fourth highest according to perceived benefit and second highest according to perceived risk. The use of motor vehicles was thus perceived to be both highly beneficial and highly risky. Motorcycles scored low for benefits, high for risks. Bicycles scored comparatively low both for benefits and risks.

Nine dimensions of risk were defined:

1. Voluntariness: Do people voluntarily expose themselves to a risk?
2. Immediacy of effect: Does the risk kill instantly or only after a latency period?
3. Knowledge about risk: Do those who are exposed to the risk know it well?
4. Scientific knowledge: To what extent is the risk known by science?
5. Control over risk: The extent to which those exposed to a risk can influence it by their own actions
6. Newness: Is the risk new or old?
7. Chronic-catastrophic: Does the risk kill people one at a time (chronic) of in large numbers at once (catastrophic)?
8. Common-dread: Are people calm about the risk or does it induce fear?
9. Severity of consequences: Will an adverse event always result in death or will most events have less serious consequences?

Each activity was rated on each dimension by means of a seven point scale. There was a weak negative correlation between perceived benefit and perceived risk: The larger the perceived benefit, the lower the perceived risk. Motor vehicles was an outlier in having both high perceived benefit and high perceived risk. Motor vehicles scored close to the midpoint of the scale (score = 4) on all nine dimensions of risk.

The acceptability of current risk increased when perceived benefit increased and was higher for voluntary risks than for risks rated as involuntary. An interesting question is whether these tendencies make sense according to utility theory. Should an individual seeking to maximise expected utility consider all the nine dimensions of risk? In case one answers yes to this question, it follows that risk and changes in it cannot be treated as a homogeneous commodity. Changes in a voluntary risk may be valued differently from changes in an involuntary risk. A risk with a latency may be valued differently from an immediate risk, and so on. If a rational utility maximiser varies his or her valuation of changes in risk according to the dimensions of risk influencing its acceptability, willingness-to-pay will

differ between different types of risk, which may result in multiple values of a statistical life for a given individual.

Slovic (2000:xxxvi) summarises three decades of research on the perception and acceptability of risks in the following terms:

"One of the most important conclusions … is that risk is inherently subjective. In this view, risk does not exist 'out there', independent of our minds and cultures, waiting to be measured. Instead, human beings have invented the concept of risk to help them to understand and cope with the dangers and uncertainties of life. Although these dangers are real, there is no such thing as real risk or objective risk. Even the simplest, most straightforward risk assessments are based on theoretical models, whose structure is subjective and assumption-laden and whose inputs are dependent upon judgment. … Whoever controls the definition of risk controls the rational solution to the problem at hand."

It is true that even comparatively well-defined risks, like the risk of a fatal road accident, can be quantified in many ways that are likely to influence views about the desirability of reducing the risk, and are thus not neutral, since one way of presenting the risk may clearly imply that it ought to be reduced, while another may suggest it is too low to worry about. The context into which a specific risk is put is highly relevant. Road safety advocates often point to the fact that, per hour spent in an activity, travel by road is, for most people, the most dangerous thing they do (Elvik 2005). A different perspective may lead people to think differently. Thus, fewer people die in road accidents each year in Norway than from drug overdoses, suicides or accidents in the home. Preventing road deaths may appear less urgent when compared to more frequent deaths than when compared to activities involving less risk per hour spent performing them. The risk of a road accident fatality per trip made is extremely low; lifetime risk is vastly higher (Slovic, Fischhoff and Lichtenstein 1978).

Slovic is thus right in saying that there are many ways of presenting information about risks and that these different ways are not neutral, in the sense that some of them will induce people to think that reducing the risk is urgent, while others may induce them to think that the risk is acceptable. It is obvious that different ways of presenting information about risk may influence the results of studies designed to elicit willingness-to-pay for reducing risks.

It does, however, not follow from this that there is no such thing as real risk or objective risk. It can, in principle, be ascertained objectively (i.e. by intersubjectively valid observations) how many trips are made by car in Norway in a given year. A true number of trips exists and is in principle countable. There is also no doubt about how many of the trips ended in a fatal accident. Hence, an objective indicator of risk can be defined – objective in the sense that everybody who computes risk as the number of fatalities divided by the number of trips should get the same answer. Such an indicator is, however, only one among very many indicators of traffic risks. While all these indicators may be objective in the sense that they are based on publicly available data with known accuracy, they are not neutral with respect to the impression they make on people when presented to them.

Slovic et al. (2004) suggest that informal judgements about risk are made by relying on an "affect heuristic". If people have positive feelings toward an activity, they will judge its benefits as large and its risks as low. If feelings are negative, benefits will be assessed as low and risks as high. It is not the case that people rely on a kind of utility calculus in which probability and value are independent terms. On the contrary, the affect heuristic instantaneously combines probability and value into an overall judgement.

One final issue arising from psychological research on road accident risk that deserves to be discussed is the so called "optimism bias" in traffic accident risk perception. Studies of this phenomenon started with a much quoted paper by Svenson: *"Are we all less risky and more skillful than our fellow drivers?"* (Svenson 1981). In the paper, he asked students in the

United States and Sweden to place themselves in percentiles with respect to driving skill and safety. Ten percentiles (0-10, 11-20, etc.) were listed. The percentiles were ranked so that the first (0-10) indicated the bottom ten percent with respect to skill and safety and the last (91-100) the upper ten percent with respect to skill and safety. If students had a realistic perception of their skill and safety, then, by definition, each percentile should contain ten percent of the students. However, Svenson found that 87.5 percent of US students and 77.1 percent of Swedish students rated their safety in the upper five percentiles, i.e. safer than the median (50th percentile) driver.

Some researchers seem to assume that it is mathematically impossible for more than half of drivers to be safer than average. Thus, Svenson, Fischhoff and MacGregor state (1985:119): *"Of course, it is no more possible for most people to be safer than average than it is for most to have above average intelligence".* Hence, when more than 50 percent of drivers state that they are safer than the average driver this is interpreted as showing a biased perception of driver safety.

Elvik (2013A) shows that this reasoning is faulty. It is entirely possible that most drivers could actually be safer than the average driver. In several data sets surveyed by Elvik, the percentage of drivers who were safer than the average driver varied between 60 and 90 percent. Thus, if a large majority of drivers state that they are safer than the average driver, they may in principle be right about this, although no study has compared stated and actual level of safety within the same group of drivers.

It should be noted that although a majority of drivers usually are safer than the average driver, they are only a little safer. A minority of drivers, typically 10-20 percent, have a considerably higher accident rate than the average driver. The notion of subjective risk and potential bias in it is relevant for the monetary valuation of transport safety, as it can be argued that the value a driver puts on safety depends on the risk the driver believes he or she is exposed to. A driver who think his risk is very low, may not be willing to pay much to reduce it. As an example, Andersson and Lundborg (2007) found that male drivers aged 25-34 on average (geometric mean) stated that their fatality risk was 3.6 per 100,000. Actual risk, as computed on the basis of the number of traffic fatalities per 100,000 males aged 25-34 was 10.8. Perhaps somewhat counterintuitively, underestimating risk is likely to be associated with an overestimation of the value of improved safety. Suppose a driver believing his risk to be 3.6 in 100,000 is willing to pay 500 (arbitrary monetary units) to reduce the risk by 50 percent. The value of a statistical life then is:

$$\text{Value of a statistical life} = \frac{500}{\left(\frac{1.8}{100,000}\right)} = 27,777,778$$

Had the driver stated the same amount with respect to actual risk, the value of a statistical life would only have been 9,259,259. Of course, the driver might have indicated a higher willingness-to-pay for a higher risk. However, as will become clear in the following chapters of this report, willingness-to-pay tends not to be proportional to the size of the change in risk.

3.4   Statistical perspectives

It is instructive to introduce statistical perspectives on risk by quoting Terje Aven's book "Misconceptions of risk" (Aven 2010). In the book, Aven discusses various definitions of risk that have been proposed, and argues that none of them capture all elements of the concept. The first definition he discusses is risk as expectation, most often defined as expected loss:

Risk = Probability of an unwanted event · Consequences of the event

According to this definition, 0.1 · -10, 0.01 · -100 and 0.001 · -1000 are identical risks, since the expected loss is the same in all cases (-1). The three risks differ with respect to their variance and the size of the loss. An individual might be able to bear the first risk, entailing a potential loss of 10, but might find a loss of 1000 ruinous. Therefore, to compare risks, one needs to consider both probability, maximum loss and variance.

He next discusses definitions of risk as a probability or probability distribution, as quantiles of a probability distribution, as uncertainty, as an event, as expected disutility, as referring to statistical estimates of probability only, as based on historical statistics only and a few other dimensions, finding them all to miss an important aspect. Thus, the statistical definition and analysis of risk identifies it as a multidimensional concept, not unlike the psychometric research identifying various dimensions of risk.

The dimensions are, however, not the same. The dimensions psychologist highlight as important, see section 3.3, are voluntariness, degree of control, catastrophic potential and dread. Dread is the key emotional dimension of risk: Is it feared or is it regarded as more pleasurable?

From a statistical perspective, some relevant characteristics of risk are:

1. Whether well-defined frequentist estimates of probability can be made or not.
2. Whether frequentist estimates of risk need to be updated frequently or remain stable over time.
3. The possibility of completely specifying the set of potential outcomes of unwanted events associated with a risk.

There is a long-standing controversy in statistics about the definition of a probability. The frequentist school defines a probability as the long-term relative frequency of an event when an activity is repeated a large number of times (Aven 2014). The repetitions to which the relative frequency refers to have to be, if not identical down to the smallest detail, at least reproducible under sufficiently similar conditions that the chance of the event is nearly the same on each repetition.

The other school of thought in statistics is the Bayesian school. Bayesians argue that a long-term relative frequency is not always well-defined or may not make sense at all (Aven 2014). If, for example, one wants to estimate the risk of terrorist attacks, a historical count of such attacks may be of limited value. Terrorists know that to be successful, their attack must come as a surprise. This often means that it will employ a different mode of operation from previous attacks, attack at a different location, and so on. In short, each new act of terrorism will in important ways be different from the previous ones, making a simple count of such events largely uninformative. In addition, the exposure class – the denominator in the estimate of probability – is also very difficult to define. Bayesians argue that probabilities often cannot be defined as relative frequencies, but have to be defined as degrees of belief. Thus, our estimate of the risk of a terrorist attack reflects our belief about how likely we think it is to occur, not some historical relative frequency.

As far as risks to life and health are concerned, including risks in transport, good statistical estimates can be developed based on easily available sources of data. All individuals in a country who were alive on January 1 of a given year as well as all those who are born during the year are at risk of dying during the same calendar year. Hence, the population exposed to the risk of dying is known with great accuracy. The probability of dying can be estimated, again with great precision, as the number of deaths during one year divided by the size of the population exposed to risk during the same year. A mortality rate estimated this way is a good estimate of the objective probability of dying and can be broken down according to age and gender into groups within which there is little within-group variation

in mortality rates. Moreover, mortality rates normally change slowly over time. The rate of change is usually quite well known. Hence, historical mortality rates provide a good basis for predicting future rates, at least for a reasonable period of time (say less than 25 years).

Thus, as far as the risk of dying in traffic accidents is concerned, the first two points listed above are essentially moot. The concept of probability is well-defined both theoretically and operationally and none of the arguments commonly made by Bayesians against a frequentist definition of risk would seem to be relevant. With respect to the (mean or group-specific) risk of dying in a traffic accident, there is an objectively correct estimate of probability. It therefore makes sense to ask whether people know the risk or not, or are able to correctly estimate it without being given any clues about the right answer. If, however, one wants an estimate of risk applying to a single individual, a subjective estimate will be the only one that makes sense.

This issue is lucidly discussed by O'Hagan et al. (2006:2-3):

"A more complex example is the probability that a specific person is killed in a road accident in the next 12 months. If we know nothing about the person except that he/she lives in England, then we could assess that probability as about one in 20,000. … A person's chance of being killed on the road varies with their age and gender, where they live in England, their occupation, whether they are married, and so on. Pursuing this example further, what is the probability that I will be killed in a road accident in the next 12 months? If we consider all the relevant conditioning factors – my age, gender, location, marital status, the model of car I drive, the number of miles that I drive each year, and so on – there is nobody else in England (and never has been) with exactly the same characteristics. There will therefore be no data on which to assess that probability, and it is even questionable how to define it."

Estimates of probability must by necessity rely on aggregated data (repeated trials in which a certain outcome has a certain relative frequency of occurrence). The estimates apply to groups and may be interpreted as the mean value for a certain group. Statistics is about averages and groups and do not in any straightforward manner apply to single individuals.

Is it correct to apply population or group mean estimates of the probability of dying when trying to elicit individual preferences for reducing this probability? An individual could argue that the stated probability does not apply to him or her and give reasons why this might be the case. O'Hagan et al. (2006:14) state:

"It may be entirely natural to ask about the probability that I will be killed on the road in the next 12 months, but it is not possible to give a frequency interpretation to such a probability. The only sense in which we can discuss it meaningfully is within the personal probability framework."

According to O'Hagan et al., the frequentist notion of probability becomes meaningless if applied to a specific individual, whereas a subjective notion of probability remains meaningful even at the individual level. This point of view has been adopted in some, but not all, studies of the monetary valuation of road safety. The practice of valuation studies is, in other words, inconsistent as far as the definition of risk, more specifically the probability of dying in a traffic accident, is concerned. Does this inconsistency make the studies incomparable? Which approach, if any, is the correct one, relying on statistical estimates of risk or eliciting personal (subjective) estimates of risk?

Valuation research is, as explained in Chapters 4 and 5, explicitly based on the assumption that individuals are rational utility maximisers, or make rational choices. As pointed out by Elster (2007:209), contemporary theory of rational choice is "subjective through and through", i.e. it refers only to what individuals believe and prefer and not to some external standard. Elster adds that: *"One might, to be sure, take the word "rational" in an objective sense, implying that a rational agent is one who makes decisions that make his life go better as judged by objective criteria such as health, longevity, or income. Used in this way, however, the idea would not have any*

explanatory power". This point of view is obviously correct as far as explaining choices by showing that they were (subjectively) rational is concerned.

It seems altogether more plausible to assume that individuals as road users relate to the risk they think there is, not to some statistical estimate they most likely do not know or have never heard about. It is, however, doubtful that subjective estimates of risk on the road are mentally represented in terms of a number or a set of numbers. The perception of risk in road traffic takes the form of "gut reactions", or immediate emotional reactions to situations in which risk materialises in the form of an unambiguous material threat of injury (Vaa 2007, 2013). Clearly, individuals can reflect upon such experiences in the calm atmosphere of their living rooms and try to articulate how they rate their risk in traffic. Still, it seems likely that many individuals will think that a numerical scale makes little sense and is fundamentally arbitrary.

Therefore, if a valuation study attempts to elicit subjective, numerical estimates of individual risk, it is necessary to ensure that:

1. The reference value provided to help respondents develop a numerical estimate of their own risk (the estimate has to be numerical to support a monetary valuation) should be as representative as possible. The reason for this is that any such value provided will serve as an "anchor" that will influence individual estimates of risk.
2. It must be possible to check if the sum of individual estimates of risk makes sense. As noted in the section about psychological perspectives on risk, it is not mathematically impossible for a large majority of drivers to be safer than the average. If, however, individual estimates of risk add up (for a set of groups) to an overall risk which is lower than the frequentist risk, the estimates must be rejected as biased. It is possible for a part of the population to be safer than the whole population, but it is not possible for everybody to be safer than average.

The subjective estimates of risk presented by Andersson and Lundborg (2007) fail according to the second criterion. For both genders, overall statistical fatality risk was 6.68 in 100,000. Overall subjective risk was 3.40 in 100,000. Females were better at estimating risk than males. For females, mean statistical risk was 3.08 in 100,000. Mean subjective risk was 3.37 in 100,000, which is actually slightly above the mean statistical risk. Among males, mean statistical risk was 10.24 in 100,000. Mean subjective risk was 3.42 in 100,000. Men of all ages underestimated their risk considerably. Their willingness-to-pay will not refer to their actual risks, but to risks that, on the average, are merely a third of the actual risks.

These results indicate a true optimism bias, as opposed to the mostly spurious optimism bias found by Svenson (1981) and others.

3.5   Conclusions

Valuation research has adopted the economic perspective on risk. According to this perspective, risk is the probability of an unwanted event, such as premature death. Risks are in most valuation studies defined as an objective, statistical estimate of mortality rate, usually stated as the annual number of traffic fatalities per 100,000 inhabitants of a jurisdiction. According to the economic perspective, risks can be made acceptable if they can be insured against. Despite the fact that life insurance is possible, it does not make sense to say that life insurance protects against the risk of losing one's life. It does not reduce or eliminate the risk, it only provides a financial protection to heirs. Life itself, i.e. the enjoyment of being alive cannot be insured. It is this fact that can motivate individuals to want to pay for a reduced risk of death. A reduced risk of death is not sought for the

material gains it may bring, but for the gain in welfare, i.e. the gain in subjective well-being and the joy of life.

Economic theory does not necessarily assume that reduced risk of death is a homogeneous good that will always and everywhere have the same value to everybody. The valuation of reduced risk of death is expected to vary depending on the circumstances.

Philosophers have discussed how to manage risks from an ethical perspective. When do risks pose moral problems and when do they not? This problem is obviously different from the valuation problem that has been the main interest of economists; yet economic and philosophical perspectives on risk partly intersect. If A performs an activity that imposes a risk on B, while B does not get any benefits of the activity, this will likely be treated as a problem in both moral philosophy and economics. In the latter discipline, it will be treated as an externality and externalities are prime examples of market failure. Some philosophers have argued that expected utility does not form an acceptable basis for making decisions about risk. While some of the criticism is plausible, the alternatives reviewed in section 3.2 (ex ante retrospection; virtuous properties required to base decisions on expected outcomes) are so far not operational and go too far in assuming perfect foresight.

Psychologists have mainly been interested in how individuals perceive risk and how they react to it emotionally. A research tradition, often referred to as the psychometric paradigm, has dominated psychological research on risk. This research tradition has identified many dimensions of risk that influence views about its acceptability. These characteristics include voluntariness of exposure, degree of control of the risk, the benefits produced by a risky activity, fairness in the distribution of risk, whether a risk is old or new, how well it is known to those who are exposed to it, how well it is known to science, and how feared (or dreaded) the risk is. Thus, individuals regard risks as multidimensional; risks are not simply the number of fatalities per some unit of exposure. They are a lot more.

The characteristics of risk that psychologists have found to influence the strength of the desire to reduce risk are clearly candidates for a list of explanatory variables in valuation studies. On the other hand, these are characteristics that differentiate different types of risk from one another; they are perhaps less relevant when valuations are sought for a single type of risk, like fatality risk in traffic.

According to psychometric research, risks are judged instantaneously by means of the affect heuristic. The core of this heuristic is whether people like, or see any benefit, of the activity that generates risk. If an activity is liked, it is rated as high in benefit and low in risk. If an activity is disliked, it is rated as low in benefit and high in risk. The affect heuristic operates by assuming that if something is positive, it must be positive in all respects, including having a low risk. Thus, the affect heuristic bypasses the need for making trade-offs, i.e. the need for balancing positive and negative aspects of an activity. Automobile travel, by the way, is quite an outlier in this respect. Unlike nearly all other activities, it is judged as having both high benefits and an unacceptably high risk. Psychometric research therefore strongly suggests that a positive willingness-to-pay to reduce traffic risk should exist.

Statistics, like psychology, treats risk as a multidimensional concept. While expected loss (probability times consequence) is often used as an indicator of risk, most statisticians would probably agree that it is an imperfect indicator. There are two fundamentally different views on probability and risk in statistics: the frequentist school and the Bayesian school. Bayesians argue that frequentist estimates of probability cannot always be made, and, in particular that a subjective estimate of probability defined as degree of belief makes sense at an individual level, whereas frequentist estimates of probability cannot be fully individualised. Some valuation studies have relied on subjective estimates of risk, but most have not.

4   The methodology of scientific research programmes

4.1   Description of the methodology of scientific research programmes

Imre Lakatos produced several versions of his theory of the methodology of scientific research programmes. The description given here is based on the 1970-version (Lakatos 1970), reprinted without changes in 1978 (Lakatos 1978 in Worrall and Currie (eds) 1978). This is the most comprehensive description Lakatos produced of the methodology of scientific research programmes. Shorter presentations are found in Lakatos (1968) and Lakatos (1971).

Textbox 4.1: Biographical data for Imre Lakatos

Imre Lakatos, c. 1960s

Imre Lakatos was born in Debrecen, Hungary on November 9, 1922 and died in London on February 2, 1974. He studied mathematics at the University of Debrecen and obtained a PhD in mathematics in 1948. He was imprisoned in Hungary for political reasons from 1950 to 1953. He fled from Hungary to England during the revolt in 1956 and studied philosophy. He obtained a doctoral degree in philosophy at Cambridge University in 1961. From 1960 until his death, he worked at the London School of Economics, where another faculty member was Karl Popper. He edited the British Journal for the Philosophy of Science from 1971 until his death.

As a background to the methodology of scientific research programmes, Lakatos discussed various versions of falsificationism, a theory of science proposed by Karl Popper, and the theory of scientific revolutions, proposed by Thomas Kuhn. He argued that neither of these theories give a satisfactory account of the history of science and cannot serve as a basis for what he terms its "rational reconstruction". The concept of rational reconstruction of the history of science will be further discussed later in this section.

Lakatos argued that science does not proceed by successive falsifications of theories, but that empirical testing of theories is more complex (1970:115):

"Indeed, it is not difficult to see at least two crucial characteristics common to both dogmatic and … methodological falsificationism which are clearly dissonant with the actual history of science: that (1) a test is – or must be made – a two-cornered fight between theory and experiment so that in the final confrontation

only these two face each other; and (2) the only interesting outcome of such a confrontation is (conclusive) falsification: the only discoveries are refutations of scientific hypotheses. However, the history of science suggests that (1') tests are – at least – three-cornered fights between rival theories and experiment and (2') some of the most interesting experiments result, prima facie, in confirmation rather than falsification."

Lakatos defined dogmatic falsificationism as the belief that all scientific theories can be falsified without qualification. Methodological falsificationism, on the other hand, recognises that any empirical test of a theory must rely on certain assumptions whose validity is taken for granted and which are therefore not subject to falsification. These remarks establish the foundation of two key propositions of the methodology of scientific research programmes:

1. A scientific theory is not rejected simply as the result of an empirical observation that seems to falsify the theory. A theory is not rejected until a new theory has been developed which can explain the apparently falsifying observation.
2. As a consequence of the fact that scientists evaluate theories not only by comparing them to the empirical facts, but also by comparing different theories to each other, findings that support a theory are regarded as just as, and possibly more, valuable than findings that contradict a theory.

Lakatos introduced a new interpretation of falsificationism, labelled sophisticated falsificationism, which he characterised in the following terms (1970:116):

"For the sophisticated falsificationist a theory is 'acceptable' or 'scientific' only if it has corroborated excess empirical content over its predecessor (or rival), that is, only if it leads to the discovery of new facts. This condition can be analysed into two clauses: that the new theory has excess empirical content ('acceptability$_1$') and that some of this excess content is verified ('acceptability$_2$')."

Lakatos clearly interprets sophisticated falsificationism within the general framework of the hypothetico-deductive method. Thus, he remarks that whether a theory has excess empirical content compared to another theory can be determined by logical analysis alone: one merely deduces the observations predicted by the hypotheses. A theory predicting a larger number (variety) of observations, including observations so far not made, has excess empirical content compared to a theory predicting fewer observations. Theories should, obviously, be compared not just in terms of the number of observations predicted; different theories can predict entirely different observations that are not consistent with each other. How should scientists choose between theories that offer inconsistent empirical predictions? Lakatos provides the following answer (1970:116):

"For the sophisticated falsificationist a scientific theory T is falsified if and only if another theory T' has been proposed with the following characteristics: (1) T' has excess empirical content over T: that is, it predicts novel facts, that is facts improbable in the light of, or even forbidden, by T; (2) T' explains the previous success of T, that is, all the unrefuted content of T is included (within the limits of observational error) in the content of T'; and (3) some of the excess content of T' is corroborated."

Thus, apparent contradictions between theories, or observations refuting a theory, are resolved by developing a new theory which is consistent not just with the new observations, but also with previous observations that were interpreted as supporting the earlier theory. This leads to the following propositions, which form part of the methodology of scientific research programmes:

3. The development of scientific theories is influenced by the results of empirical studies. Observations that are inconsistent with a theory may lead to revisions of the theory in order to account for the observations (theories are developed so that their contents are consistent with the facts).
4. Revisions of scientific theories are guided by a norm of formulating theories so as to be as consistent as possible with all known empirical facts.
5. Successive revisions of theories will therefore result in ever more general theories, seeking to account for a growing and diverse body of empirical findings.

Following the discussion of the logic of how theories are refuted, Lakatos introduces the notions of progressive and degenerative problemshifts, explaining them in the following terms (1970:118):

"Let us take a series of theories, T1, T2, T3, … where each subsequent theory results from adding auxiliary clauses to (or from semantical re-interpretations of) the previous theory in order to accommodate some anomaly, each theory having at least as much content as the unrefuted content of its predecessor. Let us say that such a series of theories is theoretically progressive (or constitutes a theoretically progressive problemshift) if each new theory has some excess empirical content over its predecessor, that is, if it predicts some novel, hitherto unexpected fact. Let us say that a theoretically progressive series of theories is also empirically progressive (or constitutes an empirically progressive problemshift) if some of this excessive empirical content is also corroborated, that is, if each new theory leads us to the actual discovery of some new fact. Finally, let us call a problemshift progressive if it is both theoretically and empirically progressive, and degenerating if it is not."

Lakatos remarks in a footnote that the term "problemshift" may not be ideal, but that it is intended to denote a shift of theories, which Lakatos argues will typically not take the dramatic form suggested by the notion of "scientific revolutions" as proposed by Kuhn (1962), but rather by successively and gradually revising a theory so as to account for anomalies.

An anomaly is a result which, if taken at face value, contradicts the hard core of a scientific research programme. The concepts of hard core and scientific research programme are explained below. A key objective when developing a scientific theory is to account for, i.e. eliminate, anomalies by making them consistent with the empirical content of a theory. Remember that the term "empirical content" denotes all observational predictions that can be made deductively from the hypotheses of a theory; the actual observations need not have been made. Indeed, Birkeland (Jago 2002) proposed a theory of the Northern Lights that had implications that were not empirically testable at the time and were only confirmed more than 60 years later when satellite observations from space became possible. The predictions made by Birkeland did, however, increase the empirical content of the theory of Northern Lights by identifying hitherto unmade observations that would support (or not support) the theory.

A progressive problemshift thus denotes the enrichment of a theory by developing hypotheses that predict new empirical findings. The opposite of a progressive problemshift is a degenerative problemshift. A degenerative problemshift will often involve developing an ad hoc hypothesis. An ad hoc hypothesis is always proposed after the act, i.e. after an observation has been made, and is formulated so that it explains that single fact only and has no other observational implications. An ad hoc hypothesis will therefore not be able to guide future research, since it does not identify any potential observations so far not made. This suggests that the following guidelines are part of the methodology of scientific research programmes:

6. The development of scientific hypotheses is guided by a norm of developing hypotheses from which many observational predictions can be made.
7. Ad hoc hypotheses should be avoided if possible, but may sometimes be needed to account for unexpected findings. It is an aim to reformulate ad hoc hypotheses to make them more general.

How to test scientific hypotheses empirically, i.e. how to design an experiment or empirical study is a vastly complex topic which as traditionally not been the main focus of the philosophy of science. Lakatos does, however, briefly dwell on this topic in discussing how to test probabilistic hypotheses, which most likely make up the vast majority of modern scientific hypotheses, at least in the social sciences. A probabilistic hypothesis does not propose a universal scientific law from which no exceptions are imaginable. It rather identifies imperfect correlations, statistical relationships which, although containing an element of regularity, are always influenced by factors that are not observed, or even observable, by the scientist. Although statistical criteria have been developed to minimise the probability or erroneous conclusions (wrongly concluding that there is no relationship when in fact there is one; or vice versa), Lakatos argues that these criteria are fundamentally insufficient since they do not address the ceteris paribus assumption always made when testing probabilistic hypotheses. Ceteris paribus means "all else equal". The problem is that the statistical criteria (of significance and power) do not tell researchers when all potentially confounding factors have been eliminated. Lakatos comments on the ceteris paribus assumption in the following terms (1970:110):

"How can one test a ceteris paribus clause severely? By assuming that there are other influencing factors, by specifying such factors, and by testing these specific assumptions. If many of them are refuted, the ceteris paribus clause will be regarded as well-corroborated. Yet the decision to accept a ceteris paribus clause is a very risky one because of the grave consequences it implies."

To accept a ceteris paribus clause is to conclude that everything else was equal; hence, the observed relationship was not disturbed by any unknown or unmeasured factors that might influence it. This, as Lakatos points out, is obviously a very strong claim, which is difficult, perhaps even impossible, to support with empirical data.

Textbooks in research method provide two answers to the question of how to deal with the ceteris paribus problem: (1) To perform randomised controlled trials, i.e. experiments involving random assignment, which will ensure that there are no systematic differences between groups of subjects assigned to different experimental conditions. It is often possible to obtain data on various characteristics of the groups and thus verify that they were identical, or nearly so, with respect to these characteristics. (2) To perform multivariate statistical analyses, trying to include as many potential confounding factors as possible and estimate their effects statistically. One must rely on theory to identify potential confounding factors.

Studies of willingness-to-pay for non-market goods may be set up in ways that resemble an experiment. Yet, these studies are not really experiments in the ordinary sense of that term. Subjects are not exposed to any treatments that are designed to produce certain effects; rather they are – as it were – asked to provide the "treatments" themselves by indicating how much they are willing to pay for certain amounts of the good being studied. All factors that can influence willingness-to-pay then become relevant as potential explanatory factors for the results of the study. Thus, if two studies of willingness-to-pay get different results, there is no way of knowing why this is the case unless one can perform a comprehensive analysis of all factors that can influence willingness-to-pay. If studies of willingness-to-pay are regarded as a scientific research programme, it is plausible to conjecture that the following guideline is a key element of the research programme:

8. Methodological studies designed to enable more rigorous tests of the empirical content of the key hypotheses of a scientific research programme are part of its positive heuristic (see below) and are associated with rewards and high status in the scientific community.

Having provided a context, Lakatos is quite concise in describing the methodology of scientific research programmes. The description is short and dense and key portions of it are quoted below (1970:132):

"The programme consists of methodological rules: some tell us what paths of research to avoid (negative heuristics), and others what paths to pursue (positive heuristic)."

A scientific research programme is, in other words, a set of methodological rules. These rules prescribe, in general terms, the topics that are regarded as worth pursuing and those that are regarded as dead-ends. Lakatos remarks that "even science as a whole can be regarded as a huge research programme". The concept is, however, applicable to more specific topics, possibly including studies of the monetary valuation of non-market goods, such as improved transport safety. Lakatos continues by describing the negative heuristic of a research programme (1970:133):

"All scientific research programmes may be characterised by their 'hard core'. The negative heuristic of the programme forbids us to direct the modus tollens at this 'hard core'. Instead, we must use our ingenuity to articulate or even invent 'auxiliary hypotheses', which form a protective belt around this core. And we must redirect the modus tollens to these. It is this protective belt of auxiliary hypotheses which has to bear the brunt of tests and get adjusted and re-adjusted, or even completely replaced, to defend the thus-hardened core. A research programme is successful is all this leads to a progressive problemshift; unsuccessful if it leads to a degenerating problemshift."

Several new and important concepts are introduced in this quote. These concepts are elaborated below.

The hard core of a research programme may be defined as a set of basic axioms or commonly accepted assumptions made in all studies within a the research programme, irrespective of whether these studies are theoretical or empirical. Thus, historically, Newton launched a research programme based on very basic insights about gravity; these insights were never questioned by researchers working within the programme until Einstein proposed relativity theory. Within this research programme it would, for example, be regarded as absurd to suggest that gravity did not exist, that it could somehow be repealed or drastically modified, or that hypotheses about any phenomenon influenced by gravity could be meaningfully derived without assuming that gravity existed and exerted its effects the way Newton had described them.

The modus tollens is an elementary syllogism with the following form:

If P (hard core), then Q

Not Q

Therefore, not P

Such a use of modus tollens is not allowed in a scientific research programme (given that P denotes the hard core). One may not deduce an implication which, if refuted by the data, would lead one to conclude that the hard core must be rejected. In this sense, the hard core of a scientific research programme is protected from empirical testing; it is never subjected to such testing and researchers are discouraged from even thinking about such an idea.

The hard core is surrounded by a protective belt. The protective belt consists of auxiliary hypotheses that may be subjected to vigorous empirical testing. The hypotheses in the protective belt are not immutable, but continuously revised and refined. The auxiliary hypotheses need to be consistent with the hard core; otherwise they would not function as a protective belt. This requirement, however, is not necessarily very restrictive.

To illustrate these ideas, consider the case of neoclassical micro-economic theory, which forms the hard core of almost all empirical research in economics, including research on the willingness-to-pay for non-market goods. (The reason why the qualifier "almost" was inserted is that a few valuation studies rely on other foundations, such as prospect theory; see Chapter 5). It basically consists of a single postulate: Individuals are rational utility maximisers. This means that they make choices they believe will maximise the satisfaction of their preferences. If one accepts this idea, there are virtually no limits to the hypotheses that can be derived from it. One only needs to make a few additional assumptions about the utility function (i.e. about individual preferences), suitably adapted to the problem at hand. The rest is a matter of deductive logic. One may deduce hypotheses proposing, among other things, that both criminal acts and the abstention from them are consistent with the hard core (i.e. both actions can be modelled as rational); addictions as well as self-restraint are consistent with the hard core; paying, as well as not paying, for a non-market public good are consistent with the hard core, and so on.

In the limit, a protective belt may develop to the point of becoming an immunising stratagem, meaning that the auxiliary hypotheses are formulated so that a joint falsification of all them is impossible. In other words, if one of the auxiliary hypotheses appears to be rejected, another will be supported. One may therefore always refer to an empirically supported auxiliary hypothesis to defend the hard core from criticism, although there may be many other auxiliary hypotheses that are not supported. It is, however, difficult to imagine a situation like this unless the auxiliary hypotheses are inconsistent (i.e. one or more of them have implications contradicting one or more of the other). Consistency in the set of auxiliary hypotheses is likely to be valued.

Turning to the positive heuristic of a research programme, Lakatos describes it in the following terms (1970:135;137):

"The positive heuristic sets out a programme which lists a chain of ever more complicated models simulating reality: the scientist's attention is riveted on building models following instructions which are laid down in the positive part of the programme." … "Which problems scientists working in powerful research programmes rationally choose, is determined by the positive heuristic of the programme rather than by psychologically worrying anomalies. The anomalies are listed but shoved aside in the hope that they will turn, in due course, into corroborations of the programme."

The positive heuristic of a scientific research programme can thus be interpreted as pointing to methodologically oriented research that holds the prospect of eliminating or explaining the anomalies existing at any time. The existence of anomalies is, by itself, not regarded as a sufficient reason for rejecting a scientific research programme. The term methodologically oriented research refers to research designed to develop, for example, new and more powerful statistical techniques for analysing data or offering new interpretations of apparently anomalous findings, intended to resolve the anomaly.

Earlier in this chapter (second paragraph at the start), the term rational reconstruction of the history of science was used. An important objective for Lakatos when developing the methodology of scientific research programmes was to provide a framework for what he called the rational reconstruction of the history of science. A rational reconstruction of history is a reconstruction which shows that the choices made by scientists were rational according to the standards of rationality embodied in the research programme within which

the researchers were working. Thus, for physicists, accepting Newton's theory of gravity was clearly a better (i.e. more rational) choice than rejecting it.

For an economist embarking on a study designed to obtain monetary valuations of non-market goods, accepting the hard core of neo-classical demand theory clearly provides a better and more meaningful foundation for research than rejecting this hard core. Indeed, if one rejects the hard core, it is not clear that any, so to speak, "alternative hard core" even exists that can provide a meaningful starting point for research. Psychology, for example, is inclined to claim that the well-ordered preferences assumed by economic theory do not exist (see e.g. Fischhoff 1991). Or, to put it in no uncertain terms: the phenomenon economists are studying in willingness-to-pay studies does not exist; hence studies make no sense. One cannot really be surprised by the fact that economists have adopted a perspective which makes their field of study meaningful and worthwhile pursuing.

Based on this presentation of the methodology of scientific research programmes, it is proposed that the following guidelines, in addition to those already listed, describe its main content:

9. The main elements of a scientific research programme are the hard core, the protective belt, the negative heuristic and the positive heuristic. The hard core consists of basic assumptions that are made when developing empirically testable hypothesis (auxiliary hypotheses) forming the protective belt. Rejection of hypotheses constituting the protective belt will not lead to rejection of the hard core, but to reformulation of the hypotheses forming the protective belt.

10. The hard core of a scientific research programme is never subjected to empirical testing involving the risk of falsification. If a scientific research programme degenerates, the hard core of the degenerating programme will in the end be incorporated into the hard core of a new programme, with its content modified so as to be consistent with the hard core of the new programme. All empirically verified content of a degenerating programme is assimilated into a new programme, i.e. a new programme is not established by rejecting the confirmed empirical content of a degenerating programme.

11. A scientific research programme can survive and thrive even if there are many anomalous results. Anomalous results are not sufficient to overturn a scientific research programme until and unless a new theory is developed to account for them. However, even an overwhelming number of anomalous results are likely to be dismissed, at least as long as the positive heuristic of a scientific research programme continues to produce innovations in research methods that hold the promise of avoiding the anomalies in future research.

For ease of reference, the key concepts of the methodology of scientific research programmes have been listed and defined in textbox 4.2.

Textbox 4.2: Key concepts of the methodology of scientific research programmes.

Scientific research programme
A set of methodological guidelines or heuristics identifying promising and unpromising topics for research
Hard core
The basic assumptions or postulates forming the basis for deducing hypotheses for empirical research. The hard core is not tested empirically and is accepted by all researchers working in a scientific research programme
Protective belt
Hypotheses developed in order to support the hard core and protect it from criticism. Hypotheses in the protective belt are tested empirically and can be falsified
Anomalies
Results that, if taken at face value, falsify the hard core of a research programme. Anomalies are never interpreted as rejecting the hard core and a research programme continues despite the anomalies in the hope of explaining them
Positive heuristic
A guideline identifying promising topics for research, in particular those that will increase the empirical content of a programme
Negative heuristic
A guideline identifying topics for research that should be avoided, in particular a prohibition against exposing the hard core to the risk of falsification

The key guidelines of scientific research programmes, stated in short form and presented hierarchically, are summarised in textbox 4.3. These guidelines are an interpretation of what Lakatos wrote and an attempt to put the essence of the methodology of scientific research programmes in a concise form. The hierarchy is intended to show the guidelines/heuristics in order of importance. Each of the key concepts listed in textbox 4.1 can be elaborated further and for some of the concepts, this will be done in later chapters.

Textbox 4.3: Key guidelines/heuristics of scientific research programmes.

Develop a hard core
The hard core of a scientific research programme should ideally speaking be axiomatic and proven, or so stated that any objection to it is self-contradictory or does not provide a foundation for research based on the hypothetico-deductive method
Develop a protective belt
The protective belt consists of hypotheses developed deductively on the basis of the hard core of the scientific research programme. Hypotheses in the protective belt are tested empirically and can be falsified
Theories reject theories
The falsification of a hypothesis should not lead to rejection of the theory it is based on until a better theory has been developed that can explain both the falsification and all the verified content of the original theory
Confirmations are sought
Science does not proceed by successive falsifications of theories. Confirmations of hypotheses are sought and are regarded as at least as valuable as rejections
Theories are developed after the facts are known
Hypotheses may be proposed after relevant facts are known in order to explain or rationalise these facts. These hypotheses may be accepted as true although they have not been tested empirically
Make theories consistent with the facts
Not only single hypotheses may be proposed after the facts are known. The whole set of hypotheses forming a theory may be reformulated to be as consistent as possible with all relevant findings of empirical research
Revisions of theories make them more abstract and general
To be able to account for a growing number of empirical findings, not all of which may be consistent with the hard core of a research programme, revisions make theories more abstract and general
Develop hypotheses with many implications
Hypotheses developed deductively are more valuable the more testable implications they have
Proceed as if anomalies did not exist
Anomalous research findings do not lead to the abandoning of a research programme. The programme proceeds as if the anomalies did not exist
Research resolving anomalies has high status
Research that develops new methods or models of analyses that can explain or resolve anomalies has high status in a scientific research programme
Never develop ad hoc hypotheses
Ad hoc hypotheses, explaining a single anomalous finding but having no other testable implications, should never be developed

In subsequent chapters, this interpretation of the methodology of scientific research programmes will serve as a frame of reference for interpreting and explaining the history of research designed to obtain monetary valuation of improving transport safety. First, however, the application of the methodology of scientific research programmes to valuation research will be discussed.

4.2 Application of the methodology of scientific research programmes to valuation research

Lakatos did not discuss the applicability of the methodology of scientific research programmes to specific scientific disciplines; the examples he gave to illustrate the key concepts of the methodology of scientific research programmes were all taken from astronomy or physics. It did not take long, however, before economists took an interest in the ideas proposed by Lakatos and their applicability to economic research. The collection of papers edited by Latsis (1976) illustrates this interest and generally take a favourable view of the applicability of the methodology of scientific research programmes to economics.

To be able to describe a field of research as a scientific research programme, one must be able to identify the elements of such a programme within the field in which one wants to apply the methodology of scientific research programmes. Questions that must be answered include:

- Is it possible to identify the hard core of the research programme?
- Has a protective belt been developed and can the hypotheses constituting the protective belt be identified?
- To what extent have the hypotheses forming the protective belt been tested empirically?
- Have anomalies been found? In case they have, have they been interpreted as (a) evidence that the hard core must be rejected, (b) the result of poor data or poor research or (c) suggesting a revision of the hypotheses forming the protective belt?

These and other questions are discussed at length in subsequent chapters. For the moment, only the hard core of valuation research will be identified. The historical development of the other elements making up a scientific research programme will be traced in subsequent chapters.

Two highly quoted papers, one by Schelling (1968), the other by Mishan (1971), are widely regarded as establishing the theoretical foundation for research based on the willingness-to-pay approach for valuing transport safety. These papers, in other words, defined the hard core forming the basis for empirical research in the field of monetary valuation of transport safety (as already noted, other approaches to valuation lack a theoretical foundation and are considered further).

It should be noted in passing that French economists (Abraham and Thedié 1960, Drèze 1962) published papers in the early nineteen-sixties, making the same arguments as Schelling and Mishan did a few years later. These paper were published in French and were largely ignored by the much larger community of researchers publishing in English. It may even be noted that a Norwegian economist made essentially the same argument as Schelling and Mishan in 1970 (Østre 1970), quoting among others Schelling (1968). Mishan had at that time still not published his paper. Østre published his report in Norwegian, and however persuasive and well-argued his report may have been, he obviously never got an international audience for it.

These mostly unknown studies are still worth mentioning because they make the same argument as Schelling and Mishan did and therefore underline the consensus among economists with respect to the theoretical foundations of valuation research. The French economists obviously did not know the papers published later and could not have been influenced by them. Schelling did not quote the French economists and may not have been aware of their papers. Østre quoted Schelling and a paper in English by Thedié and Abraham in Traffic Engineering and Control. Mishan quoted Schelling but not the French economists.

Schelling (1968:142) opens by asking the following question:

"Suppose a program to save lives has been identified and we want to know its worth. …Surely, it is sensible to ask the question, What is it worth to the people who stand to benefit from it? If a scheme can be devised for collecting the cost from them, perhaps in a manner reflecting their relative gains if their benefits are dissimilar, it surely should be their privilege to have the program if they are collectively willing to bear the cost."

Schelling goes on by discussing how one may find out whether those who benefit from a safety programme are willing to pay for it:

"There are two main ways of finding out whether some economic benefits are worth the costs. One is to use the price system as a test of what something is worth to the people who have to pay for it. … Another way of discovering what the benefits are worth is by asking people. This can be done by election, interview, or questionnaire…"

Schelling was clearly aware of the fact that asking people about willingness-to-pay created a hypothetical situation that might tempt people not to take the task too seriously and give dishonest answers. He did not regard this as a decisive objection, remarking:

"Unexpected death has a hypothetical quality whether it is merely being talked about or money is being spent to prevent it."

Schelling was not very optimistic regarding how precise estimates of willingness-to-pay for reduced risk of death might be:

"What results should we anticipate if we engage in the kind of inquiry I have described, or if we survey the market evidence of what people will pay to avoid their own deaths or the deaths of the people who matter to them? … At the outset, we can conjecture that any estimate based on market evidence will at best let us know to within a factor of 2 or 3 (perhaps only 5 or 10) what the reflective individual would decide after thoughtful, intensive inquiry and good professional advice."

Schelling's most pessimistic guess, a factor of 10, would cast serious doubt on the applicability of the estimates in cost-benefit analyses. As shown in the examples given in Chapter 1, estimates of the value of preventing a fatality range by considerably more than a factor of 10. Supposing all estimates were equally well-founded from a methodological point of view (i.e. they were all based on "high-quality" research), even the much smaller range by a factor of 10 would leave considerable room for choice. The choice of a specific value within the range of values could have decisive influence on the results of a cost-benefit analysis.

Schelling expected the valuation of changes in risk to be proportional to the size of the risk reduction:

"There are good reasons for considering the worth of risk-reduction to be proportionate to the absolute reduction of risk, for considering a reduction from 10 percent to 9 percent to be equivalent to a reduction from 5 percent to 4 percent."

Theoretical models developed later do not support Schelling on this point; see the discussion of the paper by Weinstein et al. (1980) in the next chapter. Schelling rounded off his paper by re-asserting the respect for consumer sovereignty which is fundamental in economic theory:

"The gravity of decisions about lifesaving can be dispelled by letting the consumer (taxpayer, lobbyist, questionnaire respondent) express himself on the comparatively unexciting subject of small increments in small risks, acting as though he has preferences even if in fact he does not. People do it for life insurance; they could do it for lifesaving. The fact that they may not do it well, or may not quite know what they are doing as they make the decision may not bother them and need not disenfranchise them in the exercise of consumer-taxpayer sovereignty."

The interpretation of these points of view within the framework of the methodology of scientific research programmes is discussed below. Before doing so, the paper by Mishan (1971) will be presented.

Mishan started the paper by reviewing some traditional approaches for assigning monetary values to lifesaving, such as the human capital approach (see Chapter 1). He argued that all these approaches were inconsistent with the principles of cost-benefit analysis. He further argued:

"Consistency with the criterion of a potential Pareto improvement (i.e. a change resulting in benefits that are large enough to compensate any losses; my remark) and, therefore, consistency with the principle of evaluation in cost-benefit analyses would require that the loss of a person's life be valued by reference to his CV (Compensating Variation); by reference, that is, to the minimum sum he is prepared to accept in exchange for its surrender. For unless a project that is held to be responsible for, say, an additional 1000 deaths annually can show an excess of benefits over costs after meeting the compensatory sums necessary to restore the welfare of these 1000 victims, it is not possible to make all members of the community better off by a redistribution of the net gains."

Mishan's choice of words gives the impression that he is talking about the certain death of an individual or the certain death of 1000 individuals. However, the context provided earlier in his paper makes it clear that he is talking about risks only. A risk may entail, as an average, 1000 deaths. If this sounds dramatic, remember that in the order of 7000 people were killed each year in road accidents in Great Britain at the time Mishan wrote the paper. Thus, on the next page, Mishan remarks:

"It is never the case, however, that a specific person, or a number of specific persons, can be designated in advance as being those who are certain to be killed if a particular project is undertaken. … And it is this fact of complete ignorance of the identity of each of the potential victims that transforms the calculation. Assuming universal risk aversion, the relevant sums to be subtracted from the benefit side are no longer those which compensate a specific number of persons for their certain death but are those sums which compensate each person in the community for the additional risk to which he is exposed."

Throughout his discussion Mishan talks about projects involving additional deaths, that is projects increasing risk. Nearly all empirical studies that have been made to obtain monetary values of changes in risk have dealt with reductions in risk. The core of Mishan's argument is, however, not affected by the sign of the change in risk. His main point is to establish consumer sovereignty and willingness-to-pay (or accept) as the only theoretically meaningful basis for monetary valuations of changes in risk. He states:

"All the reader has to accept is the proposition that people's subjective preferences of the worth of a thing must be counted. … People's imperfect knowledge of economic opportunities, their imprudence and unworldliness, have never prevented economists from accepting as basic data the amounts people freely choose at given prices. Such imperfections cannot, therefore, consistently be invoked to qualify people's choices when, instead, their preferences are exercised in placing a price on some increment of a good or 'bad'".

These remarks almost echo those made by Schelling to the effect that even if people do not fully know what they are doing when they try to place a value on changes in risk, this is nevertheless what we should ask them to do, and we should accept the results without question. Consumer sovereignty should be respected (although as later contributions by Mishan show, see Chapter 10, he found lack of consumer rationality troublesome). Mishan states this in very clear terms in the following paragraph:

"Economists are generally agreed – either as a canon of faith, as a political tenet, or as an act of expediency – to accept the dictum that each person knows best his own interest. If, therefore, the economist is told that a person, A, is indifferent regarding not assuming a particular risk or assuming it along with a sum of money, V, then, on the Pareto principle, the sum V has to be accepted as the relevant cost of his being exposed to that risk."

Mishan rounded off his discussion by the following remarks:

"Before concluding, however, it should be emphasised that the basic concept introduced in this paper is not simply an alternative to, or an auxiliary to, any existing methods that have been proposed for measuring the loss or saving of life. It is the only economically justifiable concept. … One may be forgiven for asserting that there is more to be said for rough estimates of the precise concept than precise estimates of economically irrelevant concepts."

Although the writing style is different, the basic message in the papers by Schelling and Mishan is the same. It is a forceful statement of what they both regard as so basic principles in economic theory that they are prepared to defend them even if empirical research may suggest that consumers are less than perfectly informed and less than perfectly rational when trying to find out how much they are willing to pay for a reduced risk of death. It thus seems appropriate to interpret their papers as statements of the hard core of a scientific research programme intended to obtain monetary valuations of changes in the risk of death (possibly also changes in other health risks, although such risks were not explicitly discussed by Schelling or Mishan).

4.3   The hard core of valuation research as a scientific research programme

The key elements of the hard core of valuation research as a scientific research programme are stated in textbox 4.4. This statement of the hard core can no doubt be elaborated considerably. Suffice it at this stage to briefly comment on some elements of the hard core.

The first element is a normative guideline for researchers. Assuming rational utility maximisation is a regulative idea of economic theory. It gives the theory unity and it has the huge advantage of permitting hypotheses to be stated in mathematical terms. In most contexts utility maximisation is a well-defined concept.

It is normative in the sense that economists rely on it as an assumption when developing hypotheses; it is not necessarily intended as a descriptive concept. An obvious objection is that it makes little sense to base scientific hypotheses on assumptions that are not realistic. Milton Friedman, in 1953 (Friedman 1953, quoted from the reprint in Hausman 2008), gave an answer to this objection that many mainstream economists would endorse:

"The only relevant test of the validity of a hypothesis is comparison of its predictions with experience. The hypothesis is rejected if its predictions are contradicted; it is accepted if its predictions are not contradicted."

Textbox 4.4: The hard core of the scientific research programme for monetary valuation of changes in health risks.

Utility maximisation
Hypotheses about valuation of non-market goods should be developed by assuming that individuals are perfectly rational utility maximisers.
Consumer sovereignty
Consumer preferences for the provision of any good, including a non-market good, should always be respected.
Willingness-to-pay
Consumer preferences for the provision of a good are expressed in terms of the maximum amount a consumer is willing to pay for the good rather than go without it.
Potential Pareto improvement
Valuations of non-market goods are obtained for use in cost-benefit analyses designed to determine if a project is a potential Pareto improvement. There is a potential Pareto improvement whenever those who gain from a project can compensate those who lose from and still retain a net gain.
Decision rule
A potential Pareto improvement exists if the benefits of a project, in monetary terms, exceed its cost, in monetary terms. Projects not satisfying this condition should be rejected.
Maximising efficiency
The principal application of valuations of changes in health risks is in cost-benefit analyses designed to maximise efficiency in the use of scarce public funds.

Friedman went on to say:

"Truly important and significant hypotheses will be found to have "assumptions" that are wildly inaccurate descriptive representations of reality, and, in general, the more significant the theory, the more unrealistic the assumptions (in this sense). … The relevant question to ask about the "assumptions" of a theory is not whether they are descriptively "realistic", for they never are, but whether they are sufficiently good approximations for the purpose in hand. And this question can be answered only by seeing whether the theory works, which means whether it yields sufficiently accurate predictions."

Friedman used the example of a free-falling body to discuss whether assumptions are sufficiently accurate for the purpose at hand. He noted that if a ball is dropped, the distance it falls in any specified time can be calculated with reasonable accuracy by assuming a vacuum. In other words, air resistance does not slow down the ball sufficiently to introduce a major inaccuracy in the prediction based on the assumption of a vacuum. He contrasted this with a feather. The feather is sufficiently slowed down by air resistance for the vacuum assumption to lead to erroneous predictions.

In contrasting the ball and the feather, Friedman comes close to contradicting himself, by suggesting that the assumptions made need to be quite realistic after all. Assuming a vacuum is not too unrealistic for the ball, but totally unrealistic for the feather. However, the rest of his paper suggests that such an interpretation would be wrong. Friedman asked whether a hypothesis can be tested empirically by assessing the realism of its assumptions and argued that such an approach to testing hypotheses is entirely misguided. For example, he noted:

"… Under a wide range of circumstances individual firms behave as if they were seeking rationally to maximize their expected returns and had full knowledge of the data needed to succeed in this attempt …"

In other words, it does not matter whether behaviour actually is rational or not if it produces the same outcomes as rational behaviour would do. The quotes from Schelling and Mishan above make it clear that they agree with Friedman about this. Both Schelling and Mishan argue for accepting the monetary valuations of reduced risk of death people state, although as Schelling suggests: *"They may not do it well or not quite know what they are doing"*, or, as Mishan remarks: *"People's imperfect knowledge of economic opportunities, their*

imprudence and unworldliness, have never prevented economists from accepting as basic data the amounts people freely choose at given prices."

Friedman's points of view regarding the role of assumptions in developing scientific hypotheses are supported by many economists, but not all. Keen (2011:158-174) argues against basing hypotheses on unrealistic assumptions. He makes a distinction between three types of assumptions that have different roles in science:

1. Negligibility assumptions: These state that some aspect of reality is unimportant for the phenomenon under investigation. Whether a theory conforms with this particular aspect of reality or not will not make a difference, because the aspect is irrelevant.

2. Domain assumptions: These specify the conditions under which a particular theory will apply. If the conditions do not apply, then neither does the theory. If, for example, the assumption is made that individuals are perfectly rational (the characteristics of perfect rationality being spelt out to whatever degree of detail an author chooses), but they are not, a hypothesis based on this assumption is wrong even if its predictions get empirical support. That support is then either co-incidental or needs a different explanation.

3. Heuristic assumptions: These are assumptions that are known to be false, but they are made as a first step towards a more general theory. Heuristic assumptions will often be stated as "for the moment, let us assume …", but will be dropped at a later stage in reasoning.

The assumptions about rationality often made by economists are domain assumptions. Economic theory is a theory about rational behaviour. If people are not rational, economic theory does not apply.

The next chapter will review theoretical contributions to the willingness-to-pay literature. All these theoretical contributions have been developed by assuming that individuals are utility maximisers. The theoretical contributions are consistent with all elements of the hard core listed in textbox 4.3. They propose hypotheses which make predictions about systematic variation in willingness-to-pay for reduced risk of death. They also identify some factors that, if the assumption of rationality is valid, should not influence willingness-to-pay. If the predicted variations are found, the hypotheses are supported; if it is not found, the hypotheses are rejected.

Thus, it is an empirical question whether individuals are rational or not. If they are rational, in the sense that the assumptions made when developing hypotheses about willingness-to-pay are descriptively accurate, one would expect the predictions made by these hypotheses to be, ceteris paribus, supported. If, on the other hand, the assumptions made are not descriptively accurate, or are in an important way incomplete (i.e. important assumptions have not been stated explicitly), the predictions made by the hypotheses based on these assumptions may not be supported.

Respect for consumer sovereignty is also part of the hard core. Indeed, it would not make sense to ask people about willingness-to-pay if one did not respect their right to choose. The alternative to consumer sovereignty is some kind of paternalism. Paternalism means that someone else, say, an expert of some sort, makes choices on your behalf because you are judged incapable of making these choices yourself. Economists tend to be sceptical to paternalism.

In the marketplace, the intensity of preferences is revealed in consumer behaviour. A person who really likes chocolate will buy a lot of it. If behaviour is taken to reveal preferences, this is true by definition; it is a tautology. Much of consumer theory is tautological in this sense. This does not make it worthless. It does, however, once again

raise the issue of empirical testing. It is not always the case that the market choices made represent the highest preference of a consumer (Sen 1973). In particular, if choices can be modelled as a Prisoners' dilemma, consumers are likely to make choices that do not result in Pareto-optimal outcomes (Tay 2002). Tay (2002), for example, shows that everybody would be better off by choosing small cars, but end up choosing big cars because they give better crash protection. But if all cars were small, the superior crash protection offered by big cars would not be needed.

Conversely, when preferences are merely stated, as in a questionnaire, their influence on actual choices remains unknown. The relationship between stated and actual willingness-to-pay must somehow be established; otherwise one does not know if the intentions stated in the context of a purely hypothetical market will translate into behaviour in a real market.

The principal application of a monetary valuation of transport safety is in cost-benefit analyses of transport safety projects. The objective of cost-benefit analyses is to develop solutions that are collectively rational, i.e. that improve welfare for everybody. A solution is normally regarded as improving everybody's welfare if it is Pareto-optimal, i.e. if it improves the welfare of at least one person while not reducing welfare for anyone. Few solutions are Pareto-optimal. Hence, a weaker criterion is applied, a potential Pareto improvement. A solution is a potential Pareto improvement if those who gain from it can compensate those who lose from it and retain a net benefit after compensation. The compensation test is generally regarded as passed if benefits are greater than costs. This is a simplification and need not be correct. Moreover, actual compensation almost never takes place.

It is important to note that compensation of losers should restore their original welfare level. Welfare is normally defined in utility terms. Hence, the compensation test of cost-benefit analyses is in utility terms. To determine whether welfare has been restored in utility terms, it is necessary to know the marginal utility of money. In most cost-benefit analyses, no attempt is made to estimate the marginal utility of money or to present the results of analysis in utility terms.

Finally, cost-benefit analyses are intended to maximise efficiency. Roughly speaking, this means that it endorses a use of measures that maximises the surplus of benefits over costs. Cost-benefit analysis is not concerned with fairness. Measures that are cost-effectiveness according to cost-benefit analyses may therefore not always be regarded as fair or promoting social justice.

4.4   Conclusions

The main conclusions of the discussion in this chapter can be summarised as follows:

1. The methodology of scientific research programmes is a theory of science which is judged to be applicable to the historic reconstruction of research on the monetary valuation of transport safety.
2. According to the methodology of scientific research programmes, every field of study has a hard core which consists of assumptions that are made by all researchers working in the field.
3. The willingness-to-pay approach to the monetary valuation of transport safety, as advocated by Schelling and Mishan, places this field of study within consumer theory and demand theory.
4. In neo-classic economic theory, the hard core of consumer theory is the assumption that individuals are rational utility maximisers.

5. This assumption is a purely formal requirement with no empirical content. It therefore leaves researchers virtually limitless opportunities for developing empirical hypotheses based on the hard core (i.e. the restrictions imposed by the hard core are few).

6. The hard core of valuation research as a scientific research programme also consists of the basic principles of cost-benefit analysis. These include: respect for consumer sovereignty, efficient resource allocation as the primary objective of cost-benefit analysis and acceptance of a potential Pareto-improvement as the criterion of a welfare improvement.

5   Developing a protective belt

The papers by Schelling (1968) and Mishan (1971) were very clear about defining the hard core of valuation research. Both papers placed this field of research squarely within standard consumer theory. The chief task in developing a protective belt was therefore to develop more specific hypotheses about willingness-to-pay for transport safety, or, in a wider context, willingness-to-pay for a reduced risk of dying or sustaining injury. This chapter will describe and assess the development of a protective belt for research on the monetary valuation of reduced risk of death, in terms of specific hypotheses that have been proposed regarding factors that influence willingness-to-pay and create systematic variation in it.

Neither Schelling nor Mishan proposed any specific hypotheses. Jones-Lee (1974) was the first to do so. In this chapter, an attempt will be made to give an overview of at least the principal empirically testable hypotheses that have been developed about willingness-to-pay for reduced mortality risk. The hypotheses have been placed in groups with respect to their main topic or the specification of the context in which they were developed. Thus, the following main topics have been defined:

1.   General characteristics of the individual valuation function for changes in risk,
2.   The relationship between the level of the risk and willingness-to-pay,
3.   The relationship between the size of the change in risk and willingness-to-pay,
4.   The relationship between the direction of changes in risk and valuation of the changes,
5.   The nature of the good producing changes in risk (private or public),
6.   The relationship between individual characteristics and willingness-to-pay,
7.   The effect of experiencing injury or a life-threatening event on willingness-to-pay,
8.   The effects of income and insurance coverage on willingness-to-pay,
9.   The relationship between human capital and willingness-to-pay,
10. The effects of the distribution of risk and wealth on willingness-to-pay,
11. Benevolence and altruism,
12. The degree of financial risk aversion,
13. The existence of background risks,
14. The existence of a maximum limit on effective life-saving expenditures at the societal level.

These topics are to some extent, but not fully, chronological, in that early theoretical contributions focussed on the topics on top of the list, whereas more recent contributions have focussed more on topics closer to the bottom of the list. Therefore, the presentation in this chapter will not be strictly chronological. A summary of the hypotheses that have been developed is given at the end of the chapter. Based on that summary, the functions of the hypotheses as a protective belt of the hard core of valuation research will be discussed. All the topics listed except for topic 14 refer to individual valuations of changes in risk.

As noted in Chapter 4, nearly all theoretical contributions to the study of willingness-to-pay for reduced risk of death are based on hard core assumptions. That means that all these theoretical contributions have been developed by assuming a certain individual preference structure, which can be represented by means of an individual utility function that has the

characteristics normally assumed for individual utility functions, i.e. it is strictly increasing and concave with respect to its main argument.

Textbox 5.1 contains definitions of some key concepts that are used in the hypotheses constituting the theory of willingness-to-pay or safety.

Textbox 5.1: Key concepts in hypotheses about willingness-to-pay

Concept	Definition
Income	Earnings (wages etc.) per time period
Annuity	A fixed annual income, eliminating variation in income by means of insurance
Wealth	The value of everything a person owns (house, car, financial assets etc.)
Assets	The same as wealth (see above)
Human capital	The capitalised (i.e. present) value of potential future earnings
Ex ante	Before the event; i.e. risks are valued in terms of probabilities of stated outcomes only
Ex post	After the event; i.e. risks are valued after resolution, after one of the outcomes occurred

5.1 The theory of willingness-to-pay for increased safety

5.1.1 The valuation function

Jones-Lee (1974) was one of the first researchers to propose hypotheses about how utility maximising individuals are likely to value prospects involving changes in the risk of dying. The term "prospect" denotes an option with a defined set of possible outcomes, one of which is that the individual dies. The shape of the individual valuation function for changes in mortality risk as proposed by Jones-Lee is shown in Figure 5.1.

The region for a positive willingness-to-pay is located to the left on the initial level of risk. Jones-Lee assumed that no individual would be willing to go bankrupt in order to obtain perfect safety; hence, the maximum willingness-to-pay for reduced risk is less that the wealth of an individual. In the region of increases in risk, Jones-Lee proposed that there would be a maximum level of acceptable risk for which an individual could be compensated.

These hypotheses suggest two limits to the trade-offs individuals are willing to make between their wealth and safety. On the one hand, individuals will not be willing to spend their entire wealth in order to eliminate a very low risk. Thus, when risk becomes very low, there will exist an upper bound on willingness-to-pay for further reductions. On the other hand, when risk becomes very high, an individual may be unwilling to assume the risk for any amount of money. Jones-Lee uses the game of Russian roulette as an example. Few people would be willing to play Russian roulette even if paid handsomely for doing it.
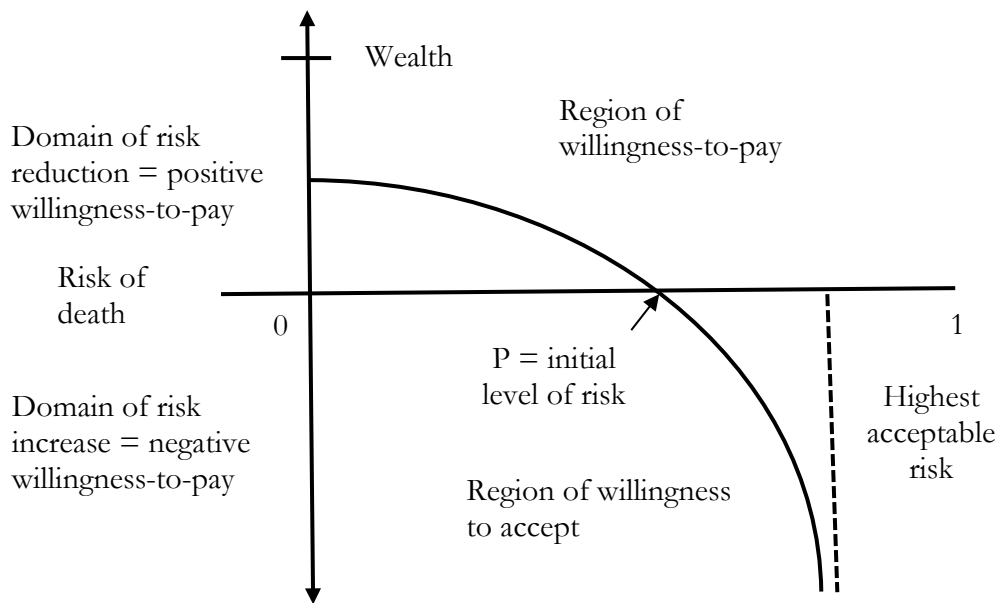
Figure 5.1: Shape of function for valuation of changes in risk. Based on Jones-Lee 1974

5.1.2 Level of risk and experience of a life-threatening event

Weinstein, Shepard and Pliskin (1980) were concerned about how the economic valuation of changes in mortality risk depends on the level of risk and on whether a person has experienced a life-threatening event, such as an accident or serious illness. They were also the first to make a distinction between the buying price of safety (how much an individual will pay for improving safety = willingness-to-pay (WTP)) and the selling price (how much an individual will ask in compensation for abstaining from an improvement in safety or accepting an increase in risk = willingness-to-accept (WTA)).

Their analysis relied only on two simple assumptions that have great face validity (something has face validity if it immediately strikes us as reasonable). The first was that life is preferred to death in a given asset position (i.e. for a given level of wealth). The second was that individuals prefer to get increases in assets when alive rather than dead. In addition, they made the assumption that individual preferences could be represented by means of a cardinal utility function satisfying the Von Neumann-Morgenstern axioms. A cardinal utility functions is a function measured at the interval level of measurement, permitting statements to be made about the size of differences in utility. Economics usually assumes that a utility function is ordinal only, i.e. it is merely a ranking and numerical values attached to the ranks need only fulfil the condition that higher ranks are assigned higher numbers than lower ranks.

By assuming utility maximisation, they found that the higher the base level of risk, the higher would be both the buying price and selling price of a given change in risk. Thus, willingness-to-pay for a reduction of risk from 0.4 to 0.3 would be greater than for a risk reduction from 0.2 to 0.1. However, this applies only to an individual who has no insurance. In case optimal life insurance and annuities (an annuity is an insurance against financial risk, replacing a variable income with a fixed annual income) are available, the marginal utility of assets becomes equal in life and death. For an optimally insured individual, the marginal utility of assets would be the same in life or death. Willingness-to-pay for reduced risk of death would then be independent of the level of risk.

This result introduces a recurring theme in the theoretical literature. Hypotheses about willingness-to-pay depend on the properties of the underlying utility functions (the term utility function in economic theory denotes a mathematical representation of individual preferences). Thus, Weinstein, Shepard and Pliskin actually put forward two hypotheses with contradictory implications:

1. For the uninsured: Willingness-to-pay is positively related to initial level of risk.
2. For the optimally insured: Willingness-to-pay is unrelated to the initial level of risk.

How to interpret an empirical study therefore depends on whether individuals are insured or not. Unless data are collected on this, including data on the actuarial fairness of insurance, there is no way of knowing whether an absence of a relationship between the level of risk and willingness-to-pay is evidence of insensitivity to scope (individuals will pay the same at any level of risk no matter how much risk changes) or rational behaviour by a fully insured individual.

Which is the most reasonable assumption to make: That individuals are uninsured or that they hold optimal life insurance and annuities? Insurance as offered by insurance companies is always less than actuarially fair. The premium always exceeds the expected loss. Life insurance may, however, be viewed differently. It is not bought principally to protect the insurance holder from risk, but to protect his or her heirs from financial losses. It is thus more akin to an annuity than to insurance against material losses from fire, floods or illness.

As far as financial risks are concerned, it is difficult to imagine how one can fully insure against them. The extent to which the welfare state protects against financial risk varies a lot between countries. In Scandinavia, there is still a strong welfare state. Hospital treatment is free, unemployment benefits are generous, and economic support is available for many other contingencies. The welfare state does not eliminate financial risk, but it reduces it considerably. Hence, probably the most realistic assumption is that most people are partly insured and that the terms of insurance contracts are less than actuarially fair. The implications of this for willingness-to-pay is discussed later.

The other topic Weinstein et al. studied was how the valuation of safety is influenced by experiencing a life-threatening event. The results were again conditional. The compensation required to accept an increase in risk is greater ex post (after the event) than ex ante for an individual who is financially risk averse or risk neutral. The willingness-to-pay for reduced risk is greater ex post than ex ante for individuals who are financially risk neutral or risk seeking, but not necessarily for individuals who are risk averse. This complex result again means that interpreting empirical studies becomes difficult unless one knows whether an individual is financially risk averse or not. Nevertheless, Weinstein et al. (page 389) offer the following guidelines:

"At least in terms of the selling price (WTA, compensation asked) of life-and-death gambles, and for risk-averse or risk-neutral individuals, the ex post value of statistical life is greater than the ex ante value. For buying prices, the result is ambiguous, and empirical assessment of the magnitudes of risk aversion over the relevant range are needed. In any case, the tendency, caused by the behavioural assumption that lifetime assets are more important than the legacy, is for the ex post value per life to exceed the ex ante value."

In conclusion Weinstein, Shepard and Pliskin (1980:393) remark:

"It has been shown that the notion of a unique willingness-to-pay value per expected life saved is inconsistent with the utility theory of the individual. The value per life saved depends on the level of the mortality probability being changed, and not just on the increment: the higher the level, the higher the value. Moreover, the value obtained ex ante will differ from the value obtained ex post, the ex post value being generally the greater of the two."

Thus, as early as 1980, Weinstein et al. did not expect empirical studies to find a single unique value of a statistical life. From a logical point of view, the next step in developing theory would be to propose more specific hypotheses and predictions regarding the range of values one might expect to find. If a unique value of a statistical life does not exist, what range of values would be theoretically plausible?

5.1.3   The size of changes in risk

Weinstein et al. (1980) found that WTP for a given change in risk would be positively related to the level of risk. They did not address the related issue of varying sizes of changes in risk from a given initial level. Most of the utility functions used in the early theoretical contributions predicted that willingness-to-pay would be proportional or nearly proportional to the size of the change in risk. As an example, Kornhauser (2001) applied the following utility functions to illustrate this:

Utility of survival = 5 + 5 · ln(W + 1)

Utility of death = ln(W + 1) (i.e. utility of bequests)

W represents income. If one assumes an annual income of, for example, NOK 600,000 per year (roughly equal to the current GDP (gross domestic product) per capita in Norway), a rational utility maximiser would be willing to pay NOK 42 for reducing risk by $6.25 \cdot 10^{-6}$ (6.25 per million = about 25 percent reduction of the current traffic fatality rate in Norway). The corresponding value of a statistical life is NOK 6,720,000. If a risk reduction of $12.5 \cdot 10^{-6}$ is assumed (twice as large), WTP becomes NOK 87.5 and the value of a statistical life NOK 7,000,000. In this example, doubling the size of the risk reduction was associated with a little more than a doubling of WTP. The reason for this is that in the utility functions assumed, the income elasticity of WTP is greater than 1.

Based on such utility functions, it was predicted that WTP would be proportional or nearly proportional to the size of the risk change. However, this was not found in empirical studies. On the contrary, many studies (some of which are reviewed in Chapter 6, 7 and 8) found that WTP was insensitive to scope, i.e. it increased far less than in proportion to the size of the risk change (although it did increase).

Amiran and Hagen (2003, 2010) proposed a directionally bounded utility function to account for this. They introduce this type of utility function by first comparing it to a standard Cobb-Douglas utility function:

$U(q, x) = q^{\alpha}x^{1-\alpha}$

They assume that x is a bundle of commodities and q is a non-market good. The Cobb-Douglas utility function imposes no limits on the amount of one good that can be given up in order to increase the amount of the other good. Amiran and Hagen refer to this property as hypersubstitutability. They argue that the existence of hypersubstitutability is highly implausible (2010:294):

"It implies, for example, that consumers who prefer more birds to fewer birds would be willing to give up nearly all of their housing (and other material goods) in exchange for a sufficient number of additional Purple Martins. Consumers are unlikely to commit themselves to extreme material poverty in order to achieve an incremental gain in any environmental amenity."

To avoid this implication, Amiran and Hagen introduced what they called a directionally bounded utility function. According to such a function, utility is monotonically increasing with respect to all arguments of the utility function, but as utility increases (with respect to one of its arguments) it approaches an asymptotic value which represents the boundary beyond which the consumer is no longer willing to give up further quantities of a certain

good in order to get more of another good. They propose the following modified Cobb-Douglas function as one example of a directionally bounded utility function:

$$U(x_1, x_2) = \left(\frac{x_1}{x_1 + a_1}\right)^{\alpha_1} \left(\frac{x_2}{x_2 + \alpha_2}\right)^{\alpha_2}$$

Amiran and Hagen point out that this utility function satisfies all neoclassical axioms of consumer theory including strict convexity of preferences and weak (as well as strong) non-satiation (i.e. the function is consistent with the hard core).

Applying this function, Amiran and Hagen argue that WTP can be highly insensitive to scope (the amount of the good being valued). Figure 5.2 shows the difference between a directionally bounded utility function and a utility which is not directionally bounded.
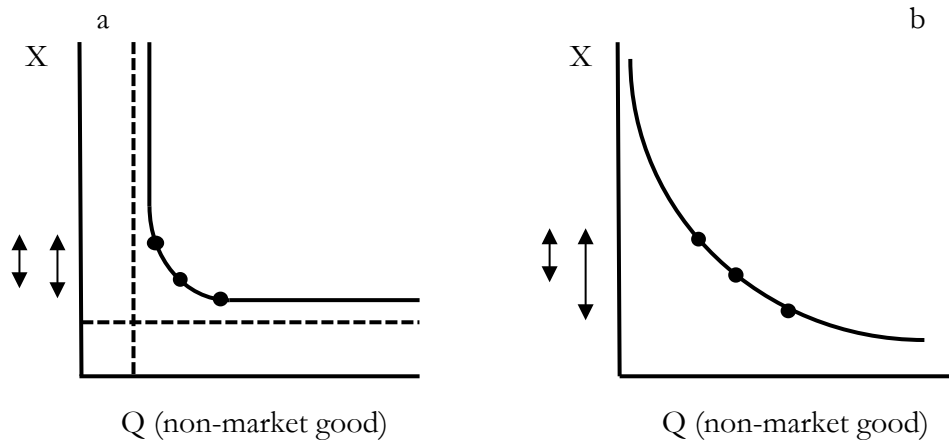
Figure 5.2: Implications for sensitivity to scope of a directionally bounded (panel a) and not directionally bounded (panel b) utility function. Based on Amiran and Hagen 2010

They conclude as follows:

"Unlike other explanations for low sensitivity to scope, our results are shown to be consistent with a rational, self-interested consumer, whose preferences are consistent with all of the neoclassical axioms. Given the plausibility of directionally bounded utility functions, there is no simple a priori basis for the notion that the degree of sensitivity to scope should be large."

In other words, findings (insensitivity to scope) that were for a long time regarded as anomalous were not necessarily so. Again, however, it must be added that to show that low sensitivity to scope is the result of a directionally bounded utility function, and not of something else, there must be independent evidence that individual preferences are in fact better described by means of a directionally bounded utility function than by means of a utility function which is not directionally bounded.

Depending on the utility function assumed, there are thus two contradictory hypotheses in the theory of willingness-to-pay: (1) Willingness-to-pay will be nearly proportional to the size if the change in risk, (2) Willingness-to-pay will be insensitive to the size of the change in risk, i.e. far less than proportional to the size of the change in risk. Results in-between these outcomes are of course also possible. It should be noted that standard utility functions (standard demand functions) do not necessarily imply that willingness-to-pay is strictly proportional to the amount of a good.

5.1.4   The direction of changes in risk

The valuation function proposed by Weinstein et al. (1980) implies that, from a given point on the function, the compensation required for an increase in risk (WTA) would be higher than the willingness-to-pay (WTP) for a reduction in risk. The function has a steeper slope in the direction of increase in risk than in the direction of reduction of risk. Still, the differences between WTA and WTP would be expected to be small.

Contrary to this expectation, empirical studies have found large differences between WTA and WTP. This was regarded as an anomaly for some years, until first Hanemann (1991) and then Amiran and Hagen (2003) showed theoretically that in some cases one would expect WTA to be considerably greater than WTP.

Hanemann (1991) argued that the near-equality between WTA and WTP traditionally assumed in economic theory applies to price changes and does not necessarily extend to changes in the quantity of commodities, in particular not when one of the commodities is a public good. He stated:

"I show that, holding income effects constant, the smaller the substitution effect (i.e. the fewer substitutes available for the public good) the greater the disparity between WTP and WTA. This surely coincides with common intuition. If there are private goods that are readily substitutable for the public good, there ought to be little differences between an individual's WTP and WTA for a change in the public good. However, if the public good has almost no substitutes, there is no reason why WTP and WTA could not differ vastly: in the limit, WTP could equal the individual's entire (finite) income, while WTA could be infinite."

Amiran and Hagen (2003) confirmed this result by relying on a directionally bounded utility function (see section 4.1.3 above). A substitute in economic theory is a replacement. Good B is a substitute for good A if one can replace a by B and get the same benefit.

To determine whether an empirically observed discrepancy between WTA and WTP makes sense or not from a theoretical point of view, it is thus necessary to find out: (1) Whether the valuation applies to a public or private good, and, in case of a public good, whether private goods are good substitutes for the public good (i.e. can provide the same benefits as the public good), or (2) The nature of the utility function that best describes individual preferences, in particular if these preferences are best described in terms of a directionally bounded utility function.

It must not be forgotten that theory predicts a small differences between WTP and WTA if the good subject to monetary valuation is a private good and if the individual does not have a directionally bounded utility function.

5.1.5   The nature of the good producing changes in risk

Measures that improve safety can either be ordinary market goods, such as safer cars, or public goods, such as road lighting. The market will normally not provide public goods in an optimal amount (Olson 1965). The reason for this is that there are incentives to free-ride: Since a public good is available to everybody once it is offered, there is no reason to pay for it if you believe you can get it for free. The traditional assumption in economic theory has therefore been that willingness-to-pay for a public good will be smaller than willingness-to-pay for a market good.

Johannesson, Johansson and O'Conor (1996) argue that it is not always the case that an individual will pay less for a public good than for a private good. They propose that altruistic motives can lead an individual to value a public good more highly than a private good. In keeping with Jones-Lee (1991, 1992; see below) they make a distinction between two types of altruism:

1. Pure altruism: A pure altruist cares about the overall welfare of another individual and is willing to pay for goods that would enhance the welfare of the other person. A pure altruist respects the preferences of the other person, i.e. does not dispute what the other person states increases his or her welfare.
2. Paternalistic altruism: A paternalistic altruist cares only about the safety of another person and believes the other person does not value his or her safety sufficiently.

Johannesson et al. (1996) argue that a pure altruist may be willing to pay more for a project that improves safety for everybody than for a project that only improves his or her own safety, depending on whether he or she believes others will pay less than, the same, or more than himself or herself. In other words, making contributions to a public goods is effective when it is believed to be effective, i.e. when everybody believes everybody else will contribute to the good. They argue:

"Let us assume that he is willing to pay $t for a ceteris paribus increase in his own safety. His total WTP for a uniform public risk reduction of the same magnitude will fall short of $t if he believes that others are willing to pay less than $t but will still be forced to pay that amount ($t) for the project. This is because those other individuals, for whom he cares will then experience a lower utility if the program is implemented."

Johannesson et al. go on to compare willingness-to-pay for a private safety measure and a public safety measure. A pure altruist will pay more for the public good than for the private good if he believes the public good improves the welfare of others. A paternalistic altruist will also be willing to pay more for a public safety measure than for a private safety measure, because he or she thinks improving the safety of others is good, even if it does not improve their overall welfare. However, a pure altruist would pay less for a public safety measure than for a private safety measure if he believed that it did not improve the overall welfare of others.

In short, beliefs about what will or will not improve the welfare of others, combined with beliefs about the contributions of others to the provision of public goods, and the presence of various forms of altruism may cause WTP for safety as a public good to be either higher than, equal to, or lower than WTP for safety as a private good. Any of these outcomes would be consistent with theory. The underlying motives would, however, be difficult to observe or ascertain.

5.1.6 Individual characteristics

Age and income are the principal individual characteristics that have attracted attention from a theoretical perspective. This section deals with age. Arthur (1981) argues that the value of a statistical life should decline monotonically with age. Shepard and Zeckhauser (1982), relying on a highly complex model of life-cycle consumption, agree that the value of a statistical life should decline monotonically for an optimally insured individual. For an uninsured individual, they conclude that the value of a statistical life should first increase, then decline, much like the typical pattern for the variation of human capital according to age.

Johansson (2002) argues that the common definition of the value of a statistical life, WTP divided by the change in risk, is valid only when the change in risk lasts for a very short period of time (a "blip") or if optimal consumption is constant throughout the duration of life. If WTP is a once-and-for-all payment to be made for a lasting change in risk, estimation of the value of a statistical life must account for how this lasting change influences optimal lifetime consumption. In general, the optimal path of lifetime consumption depends on time preferences, the market interest rate and the instantaneous risk of death (referred to as the hazard rate by Johansson).

Johansson (2002) examined (theoretically) the age-dependency of the value of a statistical life (VSL) for two cases: (1) An uninsured individual (uninsured = no protection against an uncertain income), (2) An individual insured by an actuarially fair life-long annuity contract. For the first case he found that VSL is independent of age if optimal consumption is constant throughout life. VSL declines with age if optimal consumption decreases with age, and vice versa. Thus, VSL can be both constant, increasing and decreasing with age depending on the age-pattern of optimal consumption. It may also have a more complex relationship to age, for example, first increasing then decreasing. These conclusions were found to be the same when the case of an actuarially fairly insured individual was examined.

This analysis shows that no specific pattern in the relationship between age and the value of a statistical life can be ruled out on theoretical grounds. Therefore, any empirical result would prima facie be consistent with theory. One might, to be sure, try to estimate an optimal life-cycle consumption pattern. Were it possible to do so, one might be able to propose, for example, that VSL is more likely to decrease with age than to increase with age. Unfortunately, optimal life-cycle consumption patterns can never become more than theoretical postulates. Any such pattern depends on contingencies beyond the control of the individual. The optimal consumption path may change abruptly if, for example, the individual becomes unemployed, inherits a fortune, becomes permanently disabled or experiences any other major life event that changes both tastes and opportunities. To this should be added the well-known facts of hedonic adaptation (Menzel et al. 2002) and the poor ability humans have for predicting future utility (Loewenstein et al. 2003). These factors conspire to make any estimate of optimal life-cycle consumption speculative at best.

5.1.7 Income and insurance coverage

There is universal agreement among economists that willingness-to-pay for safety is positively related to (individual) income. However, there can be fairly complex interactions between insurance and willingness-to-pay and between current and permanent income and willingness-to-pay.

Some of these interactions are discussed in a highly technical paper by Dehez and Drèze (1982), which has been summarised in an accessible manner by Jones-Lee (1985). Their starting point was the relationship between willingness-to-pay (WTP) and the level of risk. They found that the insurance coverage of an individual influences this relationship.

If the individual does not hold life insurance or annuity contracts, then WTP will decrease as initial risk becomes lower. This agrees with the theoretical results of Weinstein et al. (1980). If, on the other hand, the individual does hold life insurance or annuity contracts, but the terms of these contracts do not vary with changes in risk level, the relationship between WTP and risk level depends on whether the insurance contracts are actuarially fair or not. If the contracts are initially fair, WTP will be independent of the level of risk, because for an optimally insured individual, the marginal utility of wealth is the same in life as in death. Consequently, the marginal utility of wealth is independent of the level of risk. Optimal insurance essentially makes the individual indifferent about the level of risk, since, by definition, utility no longer depends on survival. For an optimally insured individual, one might expect not only that WTP is independent of the level of risk, but that it is zero (since equalisation of utility in life and death by means of insurance means that nothing is at stake; the psychological plausibility of this assumption is by no means obvious).

If the terms of insurance are less than actuarially fair, which would normally be the case, WTP increases as the level of risk decreases. If the terms of insurance are more than actuarially fair, WTP decreases as the level of risk decreases. Finally, Dehez and Drèze

studied what happens when the terms of insurance are adjusted to changes in the level of risk. They then found, counterintuitively, that WTP increases when the level of risk decreases (the lower the risk, the more you pay for its further reduction). They briefly remark that this result is counterintuitive, but do not delve deeper into why this might be so.

This analysis shows once more that the relationship between risk level and WTP depends on contingencies that are likely to be unknown or only partially known in any empirical study. It requires a major data collection effort to find out whether the subjects in a study have insurance coverage or not, the completeness of this coverage and the actuarial fairness of it. However, unless data of this sort are collected, one cannot know whether finding that WTP increases as risk becomes lower is an anomaly or reflect rational behaviour, given a certain insurance coverage. What remains clear is that any shape of the relationship between risk level and WTP would be consistent with theory: positive, negative, no relationship, or even zero WTP – all of them are consistent with at least one of the many utility models that have been proposed in the theoretical literature. No finding can be ruled out on theoretical grounds. Anything goes.

In most of the theoretical literature, no sharp distinction is made between income and wealth; WTP is assumed to be, ceteris paribus, positively related both to income and wealth. However, both Jones-Lee (1974) and Amiran and Hagen (2010) develop models in which individuals are presumed not to be willing to spend their entire wealth on safety. They want to leave part of their wealth behind to be able to enjoy life even after making a major investment in a life-saving measure.

Both income and wealth tend to change during life. In childhood and adolescence, income is zero or low. It then typically starts at a comparatively low level when an individual starts working and grows, perhaps almost until retirement, after which it drops. Wealth is accumulated gradually and may grow until retirement, after which it would normally stop growing and perhaps start to decline.

Robles-Zurita (2015) introduces a model making a distinction between current income and permanent income. Current income is your income right now. Permanent income is your expected mean income throughout life. He applies prospect theory, proposed by Kahneman and Tversky (1979), to develop a model predicting how the valuation of safety depends on current income and permanent income. According to prospect theory, valuations depend on a reference point and take different shapes according to whether a person is in the domain of gains or the domain of losses. It should be noted that prospect theory is not regarded as consistent with standard hard core assumptions in economic theory.

Robles-Zurita proposes that permanent income is the reference point. A person whose current income is greater than permanent income is in the domain of gains. A person whose current income is smaller than the permanent income is in the domain of losses. A person whose current and permanent income are equal is at a neutral position. He then proposes that those who are in the gain domain are willing to pay more for a given reduction of risk, at a given current income, than those who are in the neutral or loss domains. A questionnaire survey in Spain found support for this hypothesis.

If one interprets permanent income as an indicator of wealth, the model proposed by Robles-Zurita, and supported by empirical research, implies that WTP is negatively related to wealth, which is the opposite of what other theorists have proposed. It therefore seems clear that a utility model based on prospect theory can have different implications from one based on standard neoclassical utility theory (according to which reference points do not matter).

5.1.8 Willingness-to-pay and human capital

As noted in Chapter 1, the human capital method was the usual method for valuing changes in mortality risk before the willingness-to-pay approach was launched. One of the arguments proponents of the willingness-to-pay approach made, was that the valuations were likely to be higher than those obtained by means of the human capital method. Whether one can expect monetary valuations based on willingness-to-pay to be higher than those based on human capital was investigated from a theoretical perspective by Bergstrom (1982).

Bergstrom asked: When is a man's life worth more than his human capital? To answer the question, he developed several utility models. The models indicated that the conditions under which WTP will exceed human capital are weak and likely to be fulfilled in the vast majority of cases. Jones-Lee (1985) illustrates this for two cases, in which an individual has an annuity ensuring a constant consumption per year. In the first case, it is found that WTP will exceed human capital if the individual prefers survival with zero consumption to death. This case is not very realistic. In the second case, it is assumed that there exists a very low level of consumption below which death is preferred to survival. However, if consumption is above this level, WTP will exceed human capital.

Bergstrom introduced the idea of compensating an individual for a decrease in consumption by increasing the probability of survival. He argued that few people are likely to accept an actuarially fair offer, i.e. a trade in which the expected value of consumption (expected value of consumption = consumption in case of survival multiplied by the probability of survival) remains unchanged. If an individual rejects such an offer, his or her WTP must be greater than his or her human capital.

5.1.9 The distribution of risk and wealth

Pratt and Zeckhauser (1996) give a lucid analysis of how willingness-to-pay for reduced risk of death depends on the distribution of risk and on wealth. For illustration, they use the utility functions shown in Figure 5.3. There are two utility functions: one conditional on survival, one conditional on death. The marginal utility of wealth is always greater in survival than in death. Utility in the present context denotes preferring to have more wealth rather than less and preferring to be alive rather than dead. Representing utility as a mathematical function is an analytic device; it permits the use of mathematical analysis to derive prediction.
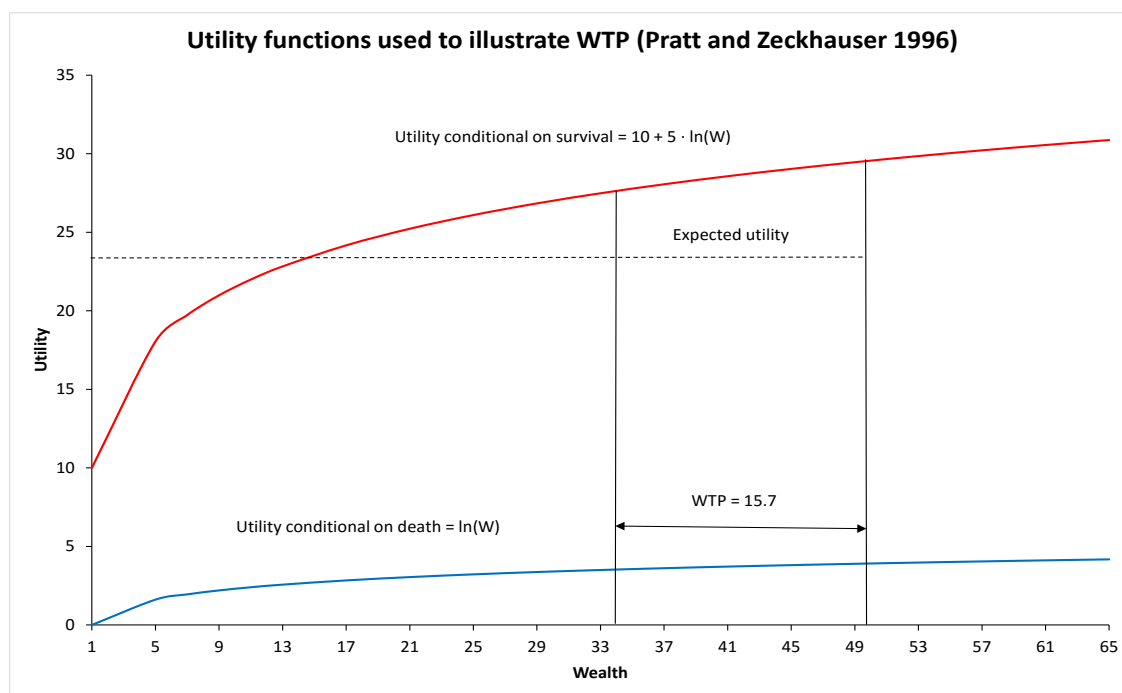
Figure 5.3: Willingness-to-pay for reducing the risk of death from 0.2500 to 0.1875. Based on Pratt and Zeckhauser 1996

Applying the utility functions shown in Figure 5.3, Pratt and Zeckhauser show that willingness-to-pay for reducing the risk of death from 0.2500 to 0.1875 is 15.7. To help readers understand how this was obtained, note that expected utility in the initial situation is:

Expected utility in initial situation = [0.75 · 10 + 5 · ln(50)] + [0.25 · ln(50)] = 23.15

Now, as explained in Chapter 3, the optimal amount to pay is the amount that keeps expected utility constant. Therefore, expected utility after payment for the risk reduction should equal 23.15. This condition obtains if:

Expected utility after payment = [0.8125 · 10 + 5 · ln(34.3)] + [0.1875 · ln(34.3)] = 23.15

As can be seen from Figure 5.3, the amount paid for the risk reduction represents a sizable proportion of wealth (initial wealth was assumed to be 50; hence WTP consumes 31.4 percent of wealth).

Pratt and Zeckhauser then investigate how WTP is influenced by inequality in the distribution of risk and wealth. Two limiting conditions are defined:

1. The dead anyway effect (explained below): This drives up WTP when initial risk is (very) high. An individual needing an expensive operation to survive, may be willing to spend very much of his or her wealth on the operation.
2. The high payment effect: This effect works in the opposite direction of the dead anyway effect, because by spending very much on reducing risk, an individual becomes poorer and may be less able to have an acceptable material standard of living after having spent a lot on risk reduction.

Pratt and Zeckhauser explain the dead anyway effect as follows: This effect pushes payment up when risk is very high, because the money spent on reducing risk has a greater probability of coming from the low-utility state, which is dead (the utility function conditional on death in figure 5.3 is flatter than the utility function conditional on being alive; hence the marginal value of money is smaller in the dead state than in the alive state, making it easier to give up a lot of money without reducing utility very much). Based on

these limiting conditions, Pratt and Zeckhauser defined distributions of the risk of one death between 100 individuals ranging from perfectly concentrated (1 individual has a 100 percent risk of dying) to perfectly egalitarian (each of 100 individuals has a 1 percent risk of dying). Three risk reductions were compared: 10 percent, 17 percent and 25 percent. Initial wealth was assumed to be 50.

For the 10 percent risk reduction, WTP declined monotonically the more dispersed the risk became. For the 25 percent risk reduction, WTP increased monotonically, although at a gradually declining rate the more dispersed the risk became. For 17 percent risk reduction, WTP initially increased, but then declined when more than 1.37 percent of the population was bearing the risk.

Thus, a large risk reduction will be associated with an increasing mean WTP the more equally it is distributed in a population benefitting from it. For a small risk reduction, the opposite tendency prevails (intuitively this makes sense: reducing a 1 percent risk by 10 percent would be worth less than reducing it by 25 percent).

Pratt and Zeckhauser next asked whether WTP is an appropriate guideline for public policy designed to reduce risk. They state it may not be an appropriate guideline. They remark the following about using WTP as a basis for public policy:

"Ex ante, the members would maximize expected utility by using a guideline that invoked not WTP, but WTP corrected for the expected marginal utility of a dollar in a manner described below, that is, to eliminate the dead-anyway effect."

The point Pratt and Zeckhauser are making, is that by paying for reducing risk, individuals increase the marginal utility of money, and society should rely on a measure of WTP which has compensated for this effect. Return for a moment to Figure 5.3. Before paying for reducing risk, the expected marginal utility of money (wealth) is:

Expected marginal utility before WTP = $(0.75 \cdot 0.10) + (0.25 \cdot 0.02) = 0.08$.

Here, 0.75 is initial probability of survival. 0.10 is the first derivative of the utility function conditional on survival at the initial wealth of 50 $(dU/dW) = 5/X = 5/50 = 0.10$; 0,25 is initial probability of death and 0.02 is the first derivative of the utility function conditional on death. Expected marginal utility is 0.08. After paying for a risk reduction to 0.1875 (leaving remaining wealth at 34.3), expected marginal utility becomes:

Expected marginal utility after WTP = $(0.8125 \cdot 0.146) + (0.1875 \cdot 0.029) = 0.124$.

Given the shape of the utility functions in Figure 5.3, the effect of a given WTP on the marginal utility of money will be greater the lower initial wealth is. Pratt and Zeckhauser propose correcting for this by equalising WTP according to the expected marginal utility of wealth assessed after payment has been made for risk reduction.

These points of view are interesting, but would not seem to have the support of most economists, who argue that utility functions are, in general, not sufficiently well known to make the adjustments Pratt and Zeckhauser call for.

5.1.10 Benevolence and altruism

The status given to benevolence and altruism in theory of willingness-to-pay for safety depends, as discussed above, on the nature of benevolence and altruism (Bergstrom 1982, Jones-Lee 1991, 1992, Lindberg 2006). Bergstrom (1982) and Jones-Lee (1991) make a distinction between pure altruism and paternalistic altruism. In pure altruism, an individual cares about the overall welfare of another individual and is willing to pay for increasing it. However, assuming the altruist respects the preferences of the other individual, it would be double counting to add the WTP of the altruist to that of the other person, since, by definition, the other person has already made the right trade-off between safety and other

goods. To believe otherwise would be to overrule, i.e. not respect, the preferences of the other person. His choices would therefore be counted twice if the valuations of a pure altruist were added to his own (i.e. the same valuations would be entered twice).

If, on the other hand, the altruist does not think the other person has made the right trade-off, but is paying too little for safety, a case can be made for adding an altruistic WTP to that of the other person. Jones-Lee (1992) extended the analysis by modelling altruism as a continuum ranging between pure and paternalistic. Based on the model, Jones-Lee concludes that for a society where altruism is recognised, the value of a statistical life should be between 10 and 40 percent higher than in a society consisting only of purely self-interested individuals.

There is a hint of inconsistency in the analyses granting legitimacy to paternalistic altruism, since in nearly all other contexts economists tend to reject any form of paternalism. Why should this particular form of paternalism be respected? This point may, however, be moot, if one re-interprets what passes for altruism as really being a self-regarding preference concerning the external effects of a death. If, for example, a mother states that she wants to pay more for her son's safety than the son himself, it might be because she realises that his death would be devastating to her. This clearly is a self-regarding preference, but it is legitimate to add it to the son's WTP if there is evidence that his WTP is purely self-regarding and that it does not include any valuation of the loss of welfare other family members would suffer in case of his death.

Lindberg (2006) analysed willingness-to-pay for the safety of family members and other (anonymous) citizens along these lines. He found that the predominant type of altruism was paternalistic, meaning that it should be added to a purely selfish valuation in order to obtain the total value to society of improving road safety.

5.1.11 The degree of financial risk aversion

Utility functions, as typically defined in economic theory, display aversion to financial risk. This means that receiving a given amount of money for sure is preferred to a lottery having the same amount of money as its expected value. The larger the difference in utility between the expected value of a lottery and an income earned with certainty, the stronger is the aversion to financial risk. Arrow (1965) and Pratt (1964) have suggested to measure the degree of risk aversion by the ratio:

Coefficient of absolute risk aversion $= \dfrac{-u''(w)}{u'(w)}$

The double prime denotes the second derivative of the utility function with respect to wealth, the single prime denotes the first derivative.

Is willingness-to-pay for reduction of physical risk related to the degree of financial risk aversion? Eeckhoudt and Hammitt (2004) have investigated this question. They studied the relationship between financial risk aversion and the valuation of physical safety under a number of different conditions. The first analysis considered the case in which there was no bequest motive (i.e. utility conditional on death = zero) and the ransom value of life was constant. The ransom value of life is the WTP for changing risk of death from 1 to 0. For this case, Eeckhoudt and Hammitt found that increasing financial risk aversion is associated with an increase in the value of a statistical life (VSL), an increase in WTP to eliminate mortality risk (from an initial level less than 1), and an ambiguous effect on WTP for a partial risk reduction (the term partial risk reduction means that mortality risk is reduced from some positive value to a lower, but still positive value).

The second analysis considered the case where there was a positive bequest motive, but the ransom value of life was kept constant. For this case, it was found that increasing financial

risk aversion increased WTP to eliminate mortality risk, but had an ambiguous effect on VSL and on WTP for a partial risk reduction.

The third analysis relaxed the assumption that the ransom value of life was constant. It was then found that differences in VSL can arise even if financial risk aversion is held constant. Increasing financial risk aversion was associated with a decrease in the ransom value of life. Holding VSL constant, increasing financial risk aversion was associated with a decrease in WTP both to eliminate risk and for a partial reduction of risk. Holding WTP to eliminate initial risk constant, it was found that VSL increases as financial risk aversion increases, but that the results for WTP to eliminate risk or partially reduce it were ambiguous.

The fourth analysis considered the implications of a non-standard utility model, more specifically a utility function consistent with cumulative prospect theory (Tversky and Kahneman 1992). Cumulative prospect theory is a version of prospect theory, the key features of which are that: (1) The utility function depends on a reference point, and (2) It is risk averse in the domain of gains and risk seeking in the domain of losses. The relationship between risk aversion and VSL and between risk aversion and WTP for a partial risk reduction were then ambiguous. WTP to eliminate mortality risk was found to increase with increasing financial risk aversion.

The relationship between financial risk aversion and the monetary valuation of reduced mortality risk is therefore complex and many results were ambiguous. It would therefore seem that it is too simple to assume that risk aversion is a general trait of individuals with clear implications for the valuation of risk. On the contrary, the ambiguity of many of the results, in particular for partial risk reductions, show that no findings can be ruled out on theoretical grounds and any relationship between financial risk aversion and WTP for safety can be interpreted as being consistent with a neoclassical conception of utility. A further somewhat unorthodox aspect of the models of Eeckhoudt and Hammitt (2004) is that they use both WTP and VSL as dependent variables. When summarising their results later in this chapter, the results referring to WTP will be used as they are most comparable to other theoretical contributions.

Kaplow (2005) discusses the relationship between the coefficient of relative risk aversion and the value of a statistical life. The starting point for his study is that recent research has shown that the coefficient of relative risk aversion is considerably greater than one. This implies that the income elasticity of the value of a statistical life should also be greater than 1. However, empirical research suggests that the income elasticity is only about 0.5 to 0.6. Kaplow observes:

"Taken together, the results indicate that estimates of the income elasticity of VSL in the empirical literature, which are about half the lowest value apparently obtainable from utility maximization, seem difficult to reconcile with rational behaviour."

The coefficient of relative risk aversion is defined as follows (w denotes wealth or income):

$$\text{Coefficient of relative risk aversion} = w \cdot \left(\frac{-u''(w)}{u'(w)} \right)$$

For the utility function conditional on survival in Figure 5.3, the first derivative is 5/W. The second derivative is $-5/W^2$. Wealth varies between 1 and 65, and the coefficient of relative risk aversion has the constant value of 1, as it should for a logarithmic utility function (Arrow 1965).

Kaplow analysed utility functions to try to resolve the inconsistency. However, it remained and his conclusion was that individuals may behave inconsistently in different contexts or that either the empirical estimates of the coefficient of relative risk aversion, the empirical estimates of the income elasticity of VSL, or both, are wrong.

63

5.1.12 The existence of background risks

Eeckhoudt and Hammitt (2001) investigated how background risks influence the value of a statistical life. The background risks were of two types: Mortality risks and financial risks.

They noted that under reasonable assumptions, VSL increases both as income (wealth) increases and as mortality risk increases. Eeckhoudt and Hammitt based their analysis on a standard model, in which the utility of wealth is greater for survival than for death and the marginal utility of wealth is positive both for survival and death. Risk aversion was assumed (the second derivative is negative). The principal results can be summarised as follows:

1. A competing mortality risk decreases willingness-to-pay (WTP) for reducing a specific mortality risk if the marginal utility of a bequest is positive.
2. A competing mortality risk has no effect on WTP for reducing a specific risk if the marginal utility of a bequest is zero.
3. A background financial risk (with a negative expected value) decreases WTP for reducing mortality risk if the financial risk is independent of mortality risk.
4. A less desirable financial risk (higher probability of loss, greater potential loss) decreases WTP to reduce mortality risk.
5. Positive correlation between financial and mortality risks is preferred to independence of risks, but has an ambiguous effect on VSL (a positive correlation enhances welfare by reducing the probability that any of the risks materialises).
6. Positive correlation between financial and mortality risks increases WTP for reducing mortality risk if risk aversion with respect to bequests is equal to risk aversion conditional on survival, or if risk aversion is zero.

These results are quite complex and conditional on the assumptions made about utility functions and about correlation between financial risks and mortality risks. Since empirical studies rarely collect data on the background conditions, resolving apparently inconsistent findings may be difficult or impossible, since the source of the inconsistency is unknown, or, even worse, close to unknowable. For example, obtaining good empirical estimates of the utility of bequests is difficult. Also, knowing which financial risks an individual faces is very difficult. Some of these risks may have a large random component beyond the control of the individual. Eeckhoudt and Hammitt note that the effects of large financial and mortality risks can be substantial, but the effects of small background risks are negligible. They conclude that the failure to account for background risks in nearly all empirical studies designed to value changes in specific mortality risks is unlikely to produce substantial bias.

5.1.13 The existence of an upper bound on safety spending

Is it possible to spend too much on saving a life? Does such a question make sense at all? If so, what does spending "too much" mean and how can we determine if too much is being spent on saving a life?

Around 1990 a rapidly growing literature addressed these questions. The main argument made in these studies (Keeney 1990, 1994, 1997, Lutter and Morrall 1994, Viscusi and Zeckhauser 1994) was that there is a negative relationship between income and mortality. Hence, if so much is spent on a risk-reducing programme that income per capita goes down, mortality may increase. If the increase in mortality attributable to reduced income is greater than the reduction in mortality attributable to a safety programme, the safety programme is too expensive and does not reduce overall mortality. In a paper analysing potential implications of Vision Zero for traffic fatalities, Elvik (1999A) presented some examples of data sets showing the relationship between income and mortality. Figure 5.4 gives an example of such a data set (collected after the 1999-paper was published).

There is a negative relationship between the mean household income in a municipality and standardised mortality. Standardised mortality means that crude mortality rates have been adjusted for age and gender, so that differences between municipalities with respect to the distribution of the population by age and gender have been controlled for.
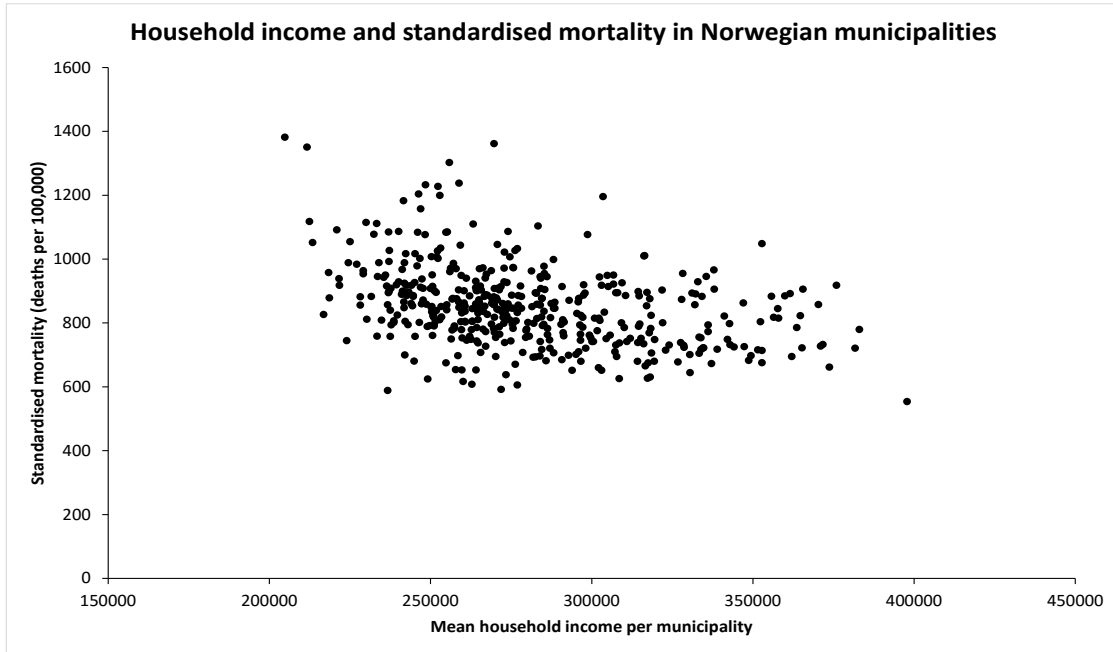

Figure 5.4: Household income and standardised mortality in Norwegian municipalities

Statistical analysis shows that an inverse function describes the relationship between income and mortality. The decrease in income associated with one additional death can be estimated to NOK 23,490,156 – a surprisingly small amount. It indicates that if more than about 23 million is spent to save a life, the expenditure is counterproductive, because a drop in income of this size – equivalent to only 12 NOK per household – is enough to statistically increase overall mortality so that there is one additional death. Viscusi (1994:95) remarks the following:

"This negative relationship between income and health creates a new kind of tradeoff for government policy. Expenditures on safety may lead to a direct reduction of risk levels, but making society poorer through the opportunity cost associated with these efforts will cause some associated increase in risk."

Viscusi subsequently (Viscusi 1998) identified several problems with studies of the simple bivariate relationship between income and mortality, such as the one shown in Figure 5.4 (which, strictly speaking, is not a simple bivariate relationship, but controls for age and gender, since the mortality rates were standardised by age and gender). In the first place, causal direction is ambiguous, in the sense that good health may be a cause of high income, rather than the other way around. Good health, as measured by indicators such as blood pressure and body weight, is in turn a strong predictor of (low) mortality.

In the second place, Viscusi questioned the plausibility of the results of many studies of the loss in income associated with one additional death. Some of these studies suggest that spending more than about 5 million US dollars is counterproductive by inducing additional deaths. This low estimate is implausible, since it implies that a lot of safety programmes associated with clear reductions in mortality are counterproductive. The offsetting increase in mortality lacks plausibility since overall mortality has been decreasing consistently for a long time. One would not observe such a decrease in overall mortality if it was the case that

spending as little as 5 million US dollars to prevent a fatality generated an additional fatality from a different cause. Thus, Tengs and Graham (1996) report an annual total expenditure of 21.4 billion US dollars on safety programmes in the United States. Presumably, this would cause nearly 4,300 additional deaths annually if the 5 million US dollar cut-off value is taken seriously.

Viscusi (1994) developed a general model intended to capture the two-way causal directions between individual health, income and mortality. The model identified two sources of individual mortality risk: exogenous risk and endogenous risk. The latter is directly influenced by an individual's health investments. These may take several forms, including buying safer products, adopting a healthier lifestyle, moving to a less polluted area, and so on. Investing in one's own health makes sense, because an individual gets a higher utility of income when healthy than when sick. This assumption is reasonable: if you have any kind of functional limitation due to poor health, it limits what you can do and most likely makes whatever you can do less enjoyable, if there is pain or discomfort associated with ill health.

Relying on utility functions with standard properties (a positive first derivative, a negative second derivative), Viscusi (1994, 1998) finds that the critical amount of expenditure for a safety programme is:

$$\text{Maximum amount of expenditure} = \frac{Value\ of\ a\ statistical\ life}{Marginal\ propensity\ to\ spend\ on\ health}$$

Viscusi states that the marginal propensity to spend on health is typically around 0.1 (the marginal propensity to spend on health is the share of a marginal (small) increase in income spent on health). This means that if the value of a statistical life has been estimated as 5 million US dollars, an expenditure of 50 million US dollars would be needed before it became counterproductive in terms of increasing overall mortality.

5.2 Assessing the theory of willingness-to-pay for reduced risk of death as a protective belt

A protective belt, it will be recalled, consists of a set of empirically testable hypotheses which form a system designed to protect the hard core of a scientific research programme, i.e. the outcome of tests of the hypotheses forming a protective belt do not lead to the rejection of the hard core, but to the revision and refinement of the hypotheses constituting the protective belt. The question to be discussed in this section, is whether the hypotheses about willingness-to-pay for reduced risk of death that have been presented above can reasonably be interpreted as forming a protective belt around the hard core of valuation research as a scientific research programme.

The hard core consists of basic postulates and assumptions that are, from a logical point of view, treated as axioms when the hypotheses in the protective belt are developed. The most basic and general of the axiomatic statements of the hard core is that individuals are utility maximisers. This is a very general statement that can be elaborated in many ways. When developing hypotheses, it serves as an assumption from which more specific hypotheses are deduced, usually by specifying some further assumptions, regarding, for example, aversion to financial risk, the availability of insurance, or more specific properties of the utility function (the structure of preferences).

The theoretical contributions reviewed above are summarised in Table 5.1. Table 5.1 lists the background conditions assumed when developing the hypotheses, the aspect of risk covered by the hypotheses and the predictions made by the hypotheses regarding willingness-to-pay for changes in the risk of death. Fifteen topics are listed in Table 5.1.

The question of whether the hypotheses about willingness-to-pay for changes in the risk of death listed in Table 5.1 form a protective belt can only be assessed by considering all the hypotheses as a system. It is their joint observational implications which identifies a pattern of results that would support them or not. It is therefore important to assess whether the hypotheses make sufficiently precise or determinate predictions to allow for an empirical testing of them – or more precisely whether the hypotheses identify findings that would lead to a rejection of them. If, as a system, the hypotheses do not identify any findings that would contradict them, they form what is referred as an immunising stratagem (Popper 1979), i.e. a theory making predictions that cannot be falsified.

The form of the valuation function (topic 1) suggested by Jones-Lee (1974) is consistent with a mainstream neoclassic demand function. It is bounded at both ends, in that Jones-Lee hypothesised that nobody would spend all of their wealth (go bankrupt, in his terms) on reducing a very low risk and that no finite amount of money could compensate for the certainty of death, implying that there will exist a maximum level of risk (less than certainty of death) at which an individual is willing to make trade-offs. By and large, these assumptions have not been challenged in subsequent contributions. The shape of the function is consistent with a standard utility function, i.e, a strictly increasing concave function (first derivative is positive, second derivative is negative). In principle, the function can be falsified, i.e., individual valuations could display a pattern not consistent with the function sketched in Figure 5.1.

It is notable that Eeckhoudt and Hammitt (2004) accept the idea of a compensation value of life, which is the amount an individual must be paid to compensate for an increase in the risk of death from 0 to 1. It should be added, however, that theirs is the only contribution of those reviewed here that makes use of such a concept.

The mainstream hypothesis is that willingness-to-pay (and willingness to accept) is positively related to the level of risk (topic 2). But this holds only for uninsured individuals. Depending on the form and terms of insurance coverage (see also topic 8 in Table 5.1), the relationship between level of risk and willingness-to-pay (or accept) can have any shape: Positive, negative or no relationship. Any of these outcomes is consistent with a model of a utility maximising individual and does therefore not, by itself, justify rejecting such a model. Thus, any relationship, or absence of a relationship, between level of risk and willingness-to-pay is consistent with theory, i.e. with at least one of the utility functions that have been assumed in developing the hypotheses. One would need to know in fairly great detail the insurance coverage of an individual to determine if a finding was consistent with theoretical predictions or not. If only the relationship between level of risk and valuation (valuation will be used as a common term for willingness-to-pay or willingness to accept) is observed, it can have any (monotonic) form. In that sense, no observation can be interpreted as refuting the underlying theory.

Table 5.1: Summary of theoretical contributions regarding willingness-to-pay for changes in mortality risk

Background conditions and key assumptions	Aspect of risk or its valuation addressed	Predictions made
The individual does not hold life insurance; more wealth is preferred to less; life is preferred to death	1 Shape of valuation function (confer Figure 5.1)	1 A positive WTP for reduced risk will exist
		2 Compensation will be demanded for an increase in risk (WTA)
		3 Maximum WTP will be less than disposable wealth
		4 There will exist a maximum acceptable level of risk; above this no compensation is acceptable
The individual holds no insurance	2 Level of risk	1 WTP for reduced risk is higher the higher the level of risk
The individual is optimally insured (income and life)		2 WTP for reduced risk is independent of the level of risk
The individual is financially risk averse or risk neutral	3 Experience of a life-threatening event	1 WTA ex post is higher than WTA ex ante
The individual is financially risk seeking or risk neutral		2 WTP ex post is higher than WTP ex ante
The individual is risk averse		3 It is indeterminate whether WTP ex post is higher than WTP ex ante
The individual has a standard utility function	4 Size of change in risk	1 WTP will be proportional or nearly proportional to the size of the change in risk (sensitivity to scope)
The individual has a directionally bounded utility function		2 WTP may be the same or nearly the same irrespective of the size of the change in risk (insensitivity to scope)
The individual has a standard utility function	5 Direction of change in risk	1 WTA for an increase in risk will slightly exceed WTP for a reduction in risk
The individual has a directionally bounded utility function		2 WTA for an increase in risk may be very much higher than WTP for a reduction in risk
The individual has no altruistic motives	6 Nature of good producing changes in risk	1 WTP for safety as a public good will be smaller than WTP for safety as a private good
The individual has altruistic motives and believes others will contribute to the provision of a public good		WTP for safety as a public good may be equal to or higher than WTP for safety as a private good
Optimal consumption is constant throughout life	7 Individual characteristics (age)	1 WTP (VSL) is independent of age
Optimal consumption declines with age		2 WTP (VSL) declines with age
Optimal consumption increases with age		3 WTP (VSL) increases with age
Optimal consumption first increases, the declines		4 WTP (VSL) first increases with age, then declines
The individual holds actuarially fair insurance; terms are not adjusted to changes in risk	8 Insurance coverage and level of risk	1 WTP is independent of level of risk
The individual holds insurance which is less than actuarially fair; terns are not adjusted to changes in risk		2 WTP increases as level of risk decreases
The individual holds insurance which is more than actuarially fair; terms are not adjusted to changes in risk		3 WTP decreases as level of risk decreases
The term of insurance contracts are adjusted to changes in risk		4 WTP increases as level of risk decreases

Background conditions and key assumptions	Aspect of risk or its valuation addressed	Predictions made
The individual has a standard utility function	9 Income and wealth	1 WTP for a given change in risk increases with income
The individual has a prospect theory utility function using wealth (lifetime income) as reference point		2 WTP for a given change in risk is negatively related to wealth (permanent income)
The individual is unwilling to accept a lottery offering reduced consumption combined with increased survival at a constant expected level of consumption	10 Human capital	1 VSL as estimated from WTP will be higher than human capital (the present value of future earnings)
The individual has a standard utility function	11 Inequality in distribution of risk	1 WTP for a small risk reduction declines as the distribution of risk becomes more egalitarian
		2 WTP for a medium large risk reduction first increases, then declines as the distribution of risk becomes more egalitarian
		3 WTP for large risk reduction increases as the distribution of risk becomes more egalitarian
The individual has a utility function characterised by pure altruism	12 Benevolence and altruism	1 Inclusion of WTP based on pure altruism is self-contradictory since a pure altruist by definition respects the choices made by others
The individual has a utility function characterised by paternalistic altruism		2 WTP including paternalistic altruism would normally be higher than WTP not including an altruistic component
The individual has a utility function with no bequest motives and the ransom value of life is constant	13 Degree of financial risk aversion	1 WTP to eliminate risk increases as financial risk aversion increases
		2 WTP to reduce (but not eliminate) risk has an ambiguous relationship to the degree of financial risk aversion
The individual has a utility function with a positive bequest motive and the ransom value of life is held constant		3 WTP to eliminate risk increases as financial risk aversion increases
		4 WTP to reduce (but not eliminate) risk has an ambiguous relationship to the degree of financial risk aversion
Willingness-to-pay to eliminate initial risk is held constant		5 WTP to eliminate risk has an ambiguous relationship to the degree of financial risk aversion
		6 WTP to reduce (but not eliminate) risk has an ambiguous relationship to the degree of financial risk aversion
The individual has a prospect theory utility function and the ransom value of life is held constant		7 WTP to eliminate risk increases as the degree of financial risk aversion increases
		8 WTP to reduce (but not eliminate) risk has an ambiguous relationship to the degree of financial risk aversion
The individual has a standard utility function with a positive bequest motive	14 Existence of background risks	1 A competing mortality risk decreases WTP for the target risk
The individual has a standard utility function with no bequest motive		2 A competing mortality risk has no effect on WTP for the target risk
The individual has a standard utility function		3 An independent financial risk decreases WTP for the target risk

Background conditions and key assumptions	Aspect of risk or its valuation addressed	Predictions made
		4 A less desirable independent financial risk decreases WTP for the target risk
		5 Positive correlation between financial and mortality risks is preferred to independence
		6 A positive correlation between financial and mortality risks increases WTP to reduce target risk if risk aversion with respect to bequests is equal to risk aversion conditional on survival or equal to zero
Spending very much to save a life can be counterproductive by increasing overall mortality	15 The existence of a maximum limit for VSL	1 There exists, in theory, an upper limit for spending on life-saving beyond which overall mortality may increase

Experience of a life-threatening event (topic 3) is more likely to be associated with a higher valuation than the opposite. Again, however, predictions depend the characteristics of the utility function. In particular results are ambiguous for a risk averse individual. A standard utility function (as defined above) is risk averse.

As far as the size of the change in risk is concerned (topic 4), the mainstream hypothesis is that valuation should be nearly proportional to it. This is usually referred to as sensitivity to scope (i.e. the more you buy, the more you pay). Again, however, an alternative model, based on directionally bounded utility functions (Amiran and Hagen 2010) has been proposed. The alternative model is consistent with the hard core, but predicts that valuation could be only weakly related to the amount of the good offered, i.e. a high degree of insensitivity to scope can be consistent with utility maximisation.

Insensitivity to scope was long regarded as one of the great blemishes of valuation studies. It was interpreted, at least by some, as casting serious doubts on the theoretical foundations of this research. If one accepts the model proposed by Amiran and Hagen, there is no longer any reason to worry about insensitivity to scope. On the contrary, it is to be expected. Thus, both sensitivity and insensitivity to scope can be interpreted as supporting the theoretical foundations of valuation research. Both outcomes are consistent with hypotheses in the protective belt and thus protect the hard core.

With respect to the direction of change in risk (topic 5), it has long been accepted that WTA for an increase in risk should exceed WTP for a reduction in risk, but not necessarily by very much, given the fact that the only difference between the two measures of value is that they involve movements in different directions (but presumably of the same magnitude) along the same demand curve. The conventional wisdom about the relationship between WTA and WTP was challenged many years ago by Hanemann (1991), who subsequently got vigorous support from Amiran and Hagen (2003). According to the models proposed by these authors, both of them consistent with standard hard core assumptions, the difference between WTA and WTP could be infinitely large. The two measures need not be close in value at all. Once more, therefore, any result of an empirical study, possibly except for finding that WTA is smaller than WTP, would be consistent with the theoretical hard core.

Mainstream models tend to assume that, all else equal, the valuation of safety in the form of a public good will be lower than the valuation of it in the form of a market good (topic 6), principally as a result of the free-rider problem involved in the provision of public goods. However, it has been proposed that if the valuation of a public good is motivated by altruism, and if everybody thinks that everybody else (or at least a "sufficient" number) are also altruists, the valuation of a public good could be higher than the valuation of a private good, which could substitute the public good. While one may regard such idealistic motivations as rare, they cannot be ruled out. Therefore, on theoretical grounds, one cannot rule out any of the following: private goods are valued more highly than public goods, public goods are valued more highly than private goods, or the two types of goods are valued the same. Any of these results is consistent with the theoretical hard core.

As far as individual characteristics are concerned (topic 7), the only ones that have received much attention are age and income. At this point, it is perhaps no surprise that hypotheses, derived from hard core assumptions, have been put forward claiming that valuation of safety declines monotonically with age, that valuation first increases with age, then declines, or that valuation of safety is independent of age. Even more complex patterns than those mentioned here cannot be ruled out. Thus, no pattern in the relationship between age and the valuation of safety is inconsistent with theory.

The effects of insurance coverage (topic 8) on the valuation of safety have been extensively studied. Willingness-to-pay for reduced risk of death is influenced by fairly complex interactions between insurance coverage and level of risk. The relationship between level of risk and willingness-to-pay depends on the terms and fairness of insurance and could be both negative and positive. One would need to know in great detail what sort of insurance individuals have in order to know whether a negative or positive relationship between level of risk and willingness-to-pay is consistent with theory or not. If one merely observes the relationship between level of risk and willingness-to-pay, not knowing anything about insurance, the relationship is theoretically indeterminate and can be both negative, positive or flat (independent). This again means that, prima facie, no outcome would falsify the underlying theory. To really test theory, one would have to determine the characteristics of the underlying utility functions. This would, in a sense, be to test a theory by judging the realism of its assumptions.

Unlike the other topics discussed so far, there seems to be unanimous agreement among economists that the valuation of safety is positively related to income (topic 9). Were one to find the opposite, it would challenge widely accepted hard core assumptions. One escape route might be to claim that safety is an inferior good, but that does not sound very plausible. Few models make a distinction between income and wealth. One model making such a distinction (Robles-Zurita 2015) found that willingness-to-pay was negatively related to wealth. Even in this model, however, willingness-to-pay remained positively related to income.

The willingness-to-pay approach was introduced as a better way of valuing safety than the human capital approach (topic 10). Theoretical analyses suggest that valuations based on willingness-to-pay will normally be higher than valuations based on human capital. Again, exceptions are possible, but only if an individual is very poor and on a steep part of the utility function, where getting more money matters more than being exposed to a higher fatality risk. Nevertheless, it is fair to say if a study made in a rich country found that valuations based on willingness-to-pay were lower than those based on human capital, the result would at least be quite surprising.

Inequality in the distribution of risk influences the valuation of safety (topic 11). However, the relationship between the degree of inequality in the distribution of risk and the valuation of safety depends on the size of the change in risk the valuation applies to. Depending on the size of the risk reduction, one can find both that valuation increases as the distribution of risk becomes more egalitarian, that it first increases, then decreases, or that it decreases monotonically as the distribution of risk becomes more egalitarian. It must therefore once more be concluded that no result can be ruled out in theory, and that no result can therefore be interpreted as falsifying theory.

Altruism and benevolence (topic 12) have already been mentioned in connection with topic 6, whether safety is provided as a public good or private good. A distinction is made between pure and paternalistic altruism. The former type of altruism has been shown to lead to double counting (since, by definition, it respects the preferences of others, and these, therefore, cannot in any meaningful way be "wrong", i.e. there is nothing to add to them or subtract from them) whereas a role is granted for paternalistic altruism. As has been remarked already, this is a quite surprising conclusion in view of the general scepticism to paternalism in economic theory. Granting legitimacy to paternalistic preferences is to allow someone else to overrule the preferences of an individual ("I do not think you value your own safety enough; therefore I am putting some of my money into it"). On the other hand, a paternalist will spend his or her own money on behalf of somebody else, and not force the other person to spend more on safety. In that sense, it does not interfere with the preferences of the beneficiary. One may nevertheless doubt

whether the theoretical distinctions between the different types of altruism can be reliably implemented in empirical research. Moreover, the distinction between paternalistic altruism ("I will pay for your safety ….") and self-regarding preferences concerning the effect of the death of a family member (" …. because your death would devastate me") is razor thin. Motives may parade as altruistic when in fact they are self-regarding.

The effects of financial risk aversion (topic 13) on the valuation of safety are complex and many theoretical results are ambiguous. It therefore seems fair to say that any result would be consistent with theory and would not lead researchers to conclude that the theory has been falsified.

The relationship between background risks (financial or physical) (topic 14) and the valuation of safety with respect to a target (selected) source of risk is complex. However, in the majority of the cases that have been analysed in theory, a competing risk decreases the valuation of reducing a target risk. Exceptions are possible, but under conditions that seem somewhat implausible. If a competing risk is higher than the target risk, and if both risks are equally amenable to reduction, it is clearly rational to pay for reducing the higher of the risks, and this might imply a zero valuation of the target risk. In most studies of willingness-to-pay, subjects have only been presented with options for reducing a single risk and competing risks have not been considered. Focusing on a certain risk can make it more salient than it really is; the risk of dying in a road accident might not even be mentioned if subjects were asked to list, say, the five most important health risks facing them. Indeed, it probably ought not to be mentioned, since it now represents only about 0.25 percent of the annual number of deaths in Norway. Surely, there must be more important sources of risk to pay for having reduced than one that merely represents a fraction of a percent of all deaths.

The final topic of those discussed above, whether it is possible to make sense of the idea of paying too much to reduce a target risk (topic 15), is of a somewhat different nature than the other topics that have been discussed. Still, if one accepts the model proposed by Viscusi (1994), it is in principle possible to think of a lower and upper bound for the value of a statistical life. The lower bound would be the human capital value. The upper bound would be an expenditure whose net effect on income was sufficiently large to induce one additional statistical death. Theoretical models suggest that the lower and upper bounds may differ by a factor of less than 20, which is a considerably narrower range than the one found in empirical studies estimating the value of a statistical life (see the review of meta-analyses in Chapter 9). According to a theoretical model developed by Viscusi (1994) it is in principle possible to determine empirically whether an estimate of the value of a statistical life is inside or outside the range implied by the model.

To conclude, the theoretical models that have been developed to deal with various topics related to the monetary valuation of safety have come to resemble what Karl Popper (1979:30) referred to as an immunising stratagem. An immunising stratagem is a reformulation of a theory to make it immune to falsification. With respect to valuation theory, this means that the set of hypotheses in the protective belt viewed as a whole rule out any results that could be interpreted as falsifying the basic postulates of the hard core. Such an interpretation is perhaps too categorical, since many of the hypotheses make predictions that are derived from specific antecedent conditions. If these conditions are absent, the results predicted by a hypothesis would (most likely) not be found and the hypothesis regarded as falsified. The problem with this argument is that the antecedent conditions represent the assumptions made when developing a hypothesis and most economists subscribe to the methodological guideline proposed by Friedman, discussed in Chapter 4, stating that a hypothesis should never be tested by assessing the realism of its assumptions, only by assessing the accuracy of its predictions. If one adheres strictly to this

guideline, it becomes a problem when the protective belt contains hypotheses making contradictory predictions. On the one hand, a standard utility function predicts at least some sensitivity to scope, if not strict proportionality. On the other hand, a directionally bounded utility function predicts weak or no sensitivity to scope.

When is the sensitivity to scope "too weak" for a standard utility function, and when is it "too strong" for a directionally bounded utility function? Surely, there must be a region of overlap or doubt as to which of the underlying utility models is most consistent with empirical results. The predictions are typically qualitative only; they predict the direction of an effect, but not its strength. Interpretation thus becomes difficult when hypotheses make contradictory predictions. If empirical findings are in a region of doubt, i.e. they can be consistent with more than one underlying utility model, attempting to ascertain which utility model best explains the results is likely to be inconclusive. This means that few results can be ruled out on theoretical grounds.

Falsification has thus been rendered almost impossible. Table 5.2 tries to summarise for each of the topics discussed above whether outcomes that falsify the hypotheses forming the protective belt are possible or not. It is seen that for most topics, falsification of the hypotheses by reference to the predictions they make is not possible, since different hypotheses make different and often contradictory predictions. To get further, one would then have to examine the validity of the assumptions made when developing the hypotheses, which would lead research in a completely different direction than what the hypotheses were developed for.

The theoretical models have thus been extremely successful in developing a protective belt. The protective belt is nearly perfect in the sense that the predictions of the hypotheses hardly rule out any finding.

One can still find corners in this theoretical structure in which predictions are sufficiently definite to rule out certain empirical results. For example, a specific form has been suggested for the valuation function; finding a different shape would falsify this hypothesis. Willingness-to-pay is predicted to be positively related to income; were one not to find this, the hypothesis would be rejected. This, however, is a very weak hypothesis. The consumption of nearly all goods is positively related to income. It would be quite extraordinary if safety should be an exception to this rule. Valuation based on willingness-to-pay is predicted to exceed valuation based on a human capital estimate. This is a testable hypothesis; it is falsified if a WTP-based estimate of the value of a statistical life is lower than a human capital based estimate. Again, however, very weak assumptions need to be made in order to predict that a WTP-based estimate of the value of a statistical life will exceed the human capital based estimate. Hence, the prediction is not bold in the sense that there is a high probability of falsification. Quite the opposite, the probability of falsification is almost zero. Finally, it is predicted that a valuation that includes a paternalistic altruistic component will be higher than a valuation not including such a component. This prediction may in principle be falsified, but a non-negligible uncertainty about whether the true motives for willingness-to-pay are altruistic or self-regarding is likely to remain.

Apart from these cases, any finding regarding the other characteristics of risks, individuals or background conditions can be interpreted as supporting theory by reference to one or more of the utility models that have been used as a basis for developing hypotheses. Keep in mind that all these utility models, except perhaps those based on prospect theory, are based on hard core assumptions, i.e. that individuals maximise (subjective) expected utility. They all assume that individuals are perfectly rational. Even prospect theory in a sense conforms to this idea. It does not posit that individuals make choices at random or reject

options that are better than the one chosen; in that sense prospect theory is also consistent with the very general idea of utility maximisation.

Indeed, utility maximisation is such a general idea that it can be elaborated in almost any number of ways. A minimalist definition of subjective utility maximisation is that an individual does what he or she thinks is best. According to Thaler (2015:161), Kenneth Arrow remarked during a seminar on behavioural economics that "rationality alone does not get you very much."

That is of course true in the sense that the rationality concept of hard core economic theory is purely formal; it is a consistency requirement only and has no empirical content. The minimalist definition, doing what you think is best, can of course be filled what almost any content. The versatility of the utility models developed in valuation research attests to this fact.

For theories about willingness-to-pay for reduced risk of death to remain empirically testable, it must be possible to falsify them. This means that the theories should predict a pattern of findings in valuation studies which, if found, will confirm the theories or, if not found, will reject them. The essential function of theory in science is to guide empirical research and the interpretation of the results of empirical research by distinguishing between findings that make sense from a theoretical point of view and findings that do not make sense from a theoretical point of view.

By and large, the theories reviewed in this chapter can no longer serve this function. Viewed as a whole, these theories no longer predict a specific pattern in the results of empirical studies. Thus, both finding and not finding sensitivity to scope is consistent with theory. Finding that WTA and WTP are close as well as finding them to be far apart is consistent with theory. Finding that people value safety more after a life-threatening event and finding that they do not value it more are both consistent with theory. And so on.

Table 5.2: Empirical testability of hypotheses about willingness-to-pay for reduced risk of death

Topic addressed	Hypotheses proposed	Falsification possible
Shape of valuation function	Demand curve bounded at both ends	Yes, a curve with a different shape may be found
Level of risk	Relationship can be positive, negative or independent	No, any relationship found, or absence of a relationship, is consistent with at least one theoretical model
Life-threatening event	Relationship is indeterminate if individual is risk averse	No, any relationship found is consistent with at least one utility model
Size of change in risk	There can be sensitivity or insensitivity to the size of change in risk	No, both finding sensitivity and not finding it is consistent with a utility model based on hard core assumptions
Direction of change in risk	There can be both a small and a very large difference between compensation demanded for an increase in risk and willingness-to-pay for a reduction in risk	No, both finding a small difference between WTA and WTP and finding a large difference between WTA and WTP is consistent with utility models based on hard core assumptions
Nature of good	Safety as a public good can be valued both lower than, equal to or higher than safety as a private good	No, any differences between valuations of safety as a public or private good are consistent with some utility model based on hard core assumptions
Age of individual	The valuation of safety can have any relationship or no relationship to age	No, any finding is consistent with some model of the optimal path of lifetime consumption
Insurance coverage interacting with level of risk	Depending on the type of insurance coverage, valuation may be negatively, positively or unrelated to the level of risk	No, any sign and strength of the relationship between level of risk and willingness-to-pay is consistent with some utility model based on hard core assumptions
Income and wealth	A standard utility model predicts that willingness-to-pay is positively related both to income and wealth; a model based on prospect theory shows that a negative relationship to wealth is possible	Yes, for income there is a clear prediction of a positive relationship; the hypothesis is falsified if this is not found
		No, for wealth models predict both a positive and a negative relationship, making falsification impossible
Human capital	Under weak assumptions, VSL as estimated from willingness-to-pay will exceed human capital	Yes, finding that a WTP value is lower than a human capital estimate of VSL would falsify the hypothesis
Unequal distribution of risk	Willingness-to-pay may increase, decrease or first increase, then decrease when the distribution of risk becomes more egalitarian	No, the relationship can have different directions depending on the size of the risk reduction – the theory does not address the issue of promoting an egalitarian distribution of risk by reducing high risks more than low risks
Altruism	Willingness-to-pay should be higher for a paternalistic altruist than for an egoist	Yes, provided the motivations for willingness-to-pay can be truthfully revealed; this, however, is highly doubtful
Financial risk aversion	Different versions of utility models cannot predict the relationship between financial risk aversion and willingness-to-pay	No, the theoretical results are ambiguous – any relationship would therefore be consistent with theory
Background risks	A competing mortality risk may decrease willingness-to-pay for a target risk or have no effect on it	Partly, if a competing risk is found to increase willingness-to-pay for the target risk that would be inconsistent with theory
	A background financial may both decrease and increase willingness-to-pay for a target risk	No; while the conditions for a decrease or increase in willingness-to-pay are different, they are in practice unobservable
Maximum valuation	Very high expenditures on saving a life can increase overall mortality	Yes, in principle it is possible to estimate empirically the maximum valuation that is consistent with a net reduction of mortality

In trying to make sense of the wide dispersion of estimates of the value of life, the theoretical models have overstretched themselves to the point of accepting almost any finding as theoretically plausible. Hence, the guidance theory can give in interpreting empirical results is very limited. It cannot help very much in sorting out those empirical estimates that make sense and those that do not. The wide dispersion observed, and findings that were for a long time regarded as anomalous, are now to a large degree regarded as normal and something to be expected.

The source of mischief is the proliferation of utility models. To get past the impasse created by this, it is necessary to test the various utility models empirically and try to determine whether some of them get more empirical support than others. Is there more empirical support for a directionally bounded utility function, which suggests that there may be insensitivity to scope, than for a standard utility function, which suggests that there should be at least some sensitivity to scope? Which of the many versions of insurance coverage discussed by Dehez and Drèze (1982) is the more common? Surely, it cannot be the model that relies on insurance which is more than actuarially fair? Is there more empirical support for a utility function based on prospect theory than for a standard utility function? Which utility model accounts best for the WTA/WTP discrepancy – a directionally bounded utility function or a utility function consistent with prospect theory?

It is questions like these that must be answered in order to eliminate some of the many utility models that have been proposed and retain only those that make unique predictions, not predictions of the form that both A and not-A are outcomes that support the theory.

The use of utility functions to support the valuation of transport safety is further discussed in Chapter 11 of the report.

5.3   Conclusions

The following main conclusions can be drawn with respect to the development of a protective belt for research on willingness-to-pay for reduced risk of death or loss of health:

1. A large number of empirically testable hypotheses have been developed regarding how characteristics of risk, of individuals, of societal institutions (in particular insurance) and other contextual variables influence willingness-to-pay for changes in the risk of death or losses of health.
2. Nearly all the hypotheses are based on hard core assumptions, i.e. that individuals are rational utility maximisers.
3. A number of different utility functions have formed the basis for developing hypotheses. Over time, the proliferation of utility functions has resulted in a set of hypotheses that make contradictory predictions.
4. Today, the set of hypotheses forming the protective belt come close to representing an immunising stratagem, meaning that falsification is impossible. If, for example, one hypotheses predicts a certain result (say, sensitivity to scope), another hypotheses will predict a different result (say, insensitivity to scope), meaning that any empirical result can find support in one of the hypotheses constituting the protective belt.
5. Many results that were initially regarded as anomalous are no longer necessarily interpreted as anomalous, due to the reformulation of hypotheses in the protective belt.

6   The progressive phase

The publication of the papers by Schelling in 1968 and Mishan in 1971, followed shortly after by the paper by Jones-Lee in 1974 did not immediately lead to empirical studies of the value of safety. In 1976, Jones-Lee published a book called "The Value of Life: An economic analysis" (Jones-Lee 1976). The book reviewed the few studies available at that time, none of which adhered completely to the willingness-to-pay approach as developed in the theoretical papers. In the book, Jones-Lee further developed the theoretical analysis first presented in his 1974-paper. He also performed an empirical study designed to value flight safety. The study was conducted in a small convenience sample, i.e. a sample of individuals who could be reached easily, but were not sampled from a sampling frame. 31 responses were received.

The study showed that respondents were able and willing to answer hypothetical questions about the valuation of flight safety. The answers indicated very high valuations; these might not apply outside the sample, but at least the study showed that it was possible get people to answer questions about the valuation of changes in safety.

With a few exceptions, it was not until 1980 that empirical research on the monetary valuation of reduced risk of death got underway. The research took different directions in Europe and New Zealand on the one hand, and North America on the other hand. In Europe and New Zealand, the stated preference approach, in particular the contingent valuation approach, became dominant. In North America, the revealed preference approach, in particular studies of compensating wage differentials, became dominant. Since these approaches differ in many respects, the history of research will be told separately for each approach.

6.1   Stated preference studies – the innovative phase

6.1.1   The pioneering study of Jones-Lee et al (1980-1983)

The year after Jones-Lee published his book on the value of life, a public commission in Great Britain published a report in which the approach to the valuation of transport safety was discussed. The commission (the Leitch Committee) remarked the following (1977:104):

"Indeed, we are not entirely convinced that the line of argument used by the Department is consistent with the general principles of cost-benefit analysis. That would suggest that the Department should aim to find the amount that an average individual would be willing to pay (or would require in compensation for) for a reduction (increase) of (correctly perceived) risk of sustaining an accident."

The Department of Transport did not immediately follow up this recommendation. It was only after Jones-Lee lobbied the department for some time that he was able to convince department officials that a willingness-to-pay survey should be conducted. In the report presenting this survey (Jones-Lee, Hammerton and Abbott 1983), its origin is described as follows:

"Following discussions and correspondence over a number of years with one of the authors of the report, the Department of Transport decided that, in view of its flexibility and potential for providing answers to a

wide variety of pertinent questions, the questionnaire approach warranted more careful evaluation and testing than it had hitherto received. The Department of Transport therefore commissioned a two-phase programme of research, commencing in January, 1980, to be undertaken jointly by the Departments of Economics and Psychology of the University of Newcastle upon Tyne."

The first phase of the study tested the feasibility of the questionnaire approach. It was concluded that the approach functioned well enough to proceed to the main study (Hammerton, Jones-Lee and Abbott 1982). The main study was reported in 1983 (Jones-Lee et al. 1983), but has also been presented by Jones-Lee in 1985 (Jones-Lee et al. 1985) and 1989 (Jones-Lee 1989). The presentation here is based on Jones-Lee (1989).

Valuations were elicited by asking direct questions about willingness-to-pay. As an example, one of the questions asked had the following wording (Jones-Lee 1989:211):

"As we said earlier, the risk of a car driver being killed in an accident is 10 in 100,000. You could choose to have a safety feature fitted to your car which will halve the risk of the car driver being killed, down to 5 in 100,000. Taking into account how much you can personally afford, what is the most that you would be prepared to pay to have this safety feature fitted to the car?"

If respondents were unable to answer, interviewers read different amounts of money (0, 1, 2, 5 etc.) until the respondent asked them to stop. Respondents were then asked to state their maximum willingness-to-pay. The study produced several estimates of the value of a statistical life, ranging from 200,000 pounds to 20,340,000 pounds (1982-value). The results contained both patterns that were expected and some anomalous findings.

To help assess the extent to which answers were consistent with the axioms of utility theory, a consistency score ranging from 0 to 8 was developed. Nearly 50 percent of respondent scored at least 5.5 points on this scale. Regression models were developed to identify factors associated with willingness-to-pay. The authors conclude that (1989:186):

"We believe that, taken together, these regression results are not consistent with the hypothesis that respondents were generally either guessing or systematically misrepresenting their true willingness-to-pay."

Willingness-to-pay was highly positively skewed, meaning that the mean value was higher than the median value. The authors commented on this difference in the following terms:

"If values of safety are to be defined in strict accordance with the principles of conventional social cost-benefit analysis without distributional weights,… the value of a statistical life should be based on the mean of individual marginal rates of substitution. If, by contrast, one wished to base decisions about transport safety on values that would command majority support in relation to any proposed alternative, then the median would seem to be the more appropriate statistic."

In the end, the authors recommended assigning a value of at least 600,000 pounds to a statistical life, adding that this was very conservative and that values well in excess of 1 million pounds were supported by study results.

The Department of Transport was not immediately convinced by the results of the study. Only in 1988, after many years of hesitation (remember that the study was published in 1983), did the Department of Transport decide to give up the human capital approach to valuation and start relying on the willingness-to-pay approach. The value of a statistical life was set to 500,000 pounds, a very conservative estimate close to the bottom of the range of values found at that time in the literature.

6.1.2 Creating support for a research programme

While working on the first valuation study in Great Britain, Jones-Lee took the initiative to convene an international conference of the value of life and safety. The conference,

attended by prominent economists from Europe and North America with an interest in the valuation of safety, was held in Geneva on March 30 and 31, and April 1, 1981. It resulted in the book "The Value of Life and Safety", edited by Jones-Lee and published by North-Holland (now Elsevier Science) in 1982.

In retrospect, this conference must be regarded as a key step in creating support for a research programme on the monetary valuation of safety, and in building a network of researchers who were to communicate and co-operate about research for many years to come. The conference was attended both by proponents of the willingness-to-pay approach and by critics of it.

Many of the presentations at the conference dealt with various aspects of theory. Some of these presentations were quoted in Chapter 5. It is clear that the conference made a major contribution to developing theory about willingness-to-pay for safety.

In the preface to the book, Jones-Lee wrote:

"It is my impression that a narrow majority of participants maintained the view (with varying degrees of conviction) that the willingness-to-pay methodology has in principle much to commend it as a basis for the definition and estimation of values of life and safety, at least in public sector allocative decision making. Nevertheless a significant dissenting minority argued that the willingness-to-pay approach is for various reasons, inappropriate or at least seriously flawed."

It took some years before anyone tried to replicate the valuation study made by Jones-Lee and others in Great Britain. But when the first replication was successfully made, it set in motion a rapid succession of studies, nearly all of them relying on highly similar designs.

6.1.3 Replication in Austria

The first study valuing changes in road safety by means of the contingent valuation method, and explicitly stating that it was intended as a replication of the British study, was reported in Austria in 1989 (Maier, Gerking and Weiss 1989). It was a small pilot study made in a convenience sample of 98 respondents. Eight questions were asked about willingness-to-pay (WTP) for changes in road accident fatality risk. Two of the questions asked for compensation for an increase in risk, the other six asked for willingness-to-pay for reduced risk.

Estimated values of a statistical life ranged between 16.9 and 328.4 million Austrian schilling. Six of the questions gave values mean between 16.9 and 46.6 million schilling. Nearly all respondents were willing to answer the questions about compensation for an increase in risk. The mean amount of compensation required was higher than willingness-to-pay for a reduction in risk, but not very much higher.

The authors studied the relationship between WTP and background characteristics such as age and income. By and large the results were consistent with theoretical expectations. The authors state that only 3 out of 98 respondents gave inconsistent answers, but do not explain what an "inconsistent" answer is. However, given the fact that respondents were asked about willingness-to-pay for different sizes of risk reduction (or increase), inconsistency probably means that respondents stated a lower willingness-to-pay for large reduction in risk than for a small reduction. The authors conclude that: *"Despite the small sample size in this pilot study the contingent valuation method seems to be quite promising in the context of road safety."*

6.1.4 Replication in Sweden

Persson and Cedervall (1991) reported on a Swedish valuation study, conducted in 1986 and 1987. It was a questionnaire survey and the questionnaire was very similar to the one used by Jones-Lee et al. in Great Britain.

One important difference between the two studies is that Persson and Cedervall relied on respondents' subjective estimates of their own risk as a basis for valuing changes in risk. Theoretically speaking, this is plausible. Unless people are informed about a certain risk, they are unlikely to have a very precise quantitative notion of it. When asked to value changes in a risk, people will anchor their answers on what they think the risk is, unless specifically instructed to rely on a stated risk.

At the time Persson and Cedervall conducted their study, the mean fatality risk for a car driver in Sweden was about 10 in 100,000. Respondents were told that the fatality risk of a motorcyclist was 130 in 100,000. They were then asked to estimate: (a) The mean risk for a car driver in Sweden, and (b) Their own risk as a car driver. The mean value stated for the risk to an average car driver was 84 in 100,000. The median value was 50 in 100,000. Both these estimates show that risks were considerably overestimated, possibly (indeed likely) as a result of anchoring estimates to the risk stated for a motorcyclist. The mean value of the estimate of own risk as a car driver was 51 in 100,000; the median value was 10 in 100,000. The median subjective personal risk as a car driver happened to be identical to the mean objective risk to car drivers.

Several questions were asked to elicit willingness-to-pay. Willingness-to-pay was elicited for three levels of change in fatality risk for car drivers: 10 percent reduction, 25 percent reduction and 50 percent reduction. All these changes used the subjectively stated personal risk as baseline. Subjective risk concentrated to seven levels: 1 in 100,000; 2 in 100,000; 5 in 100,000; 10 in 100,000; 20 in 100,000, 50 in 100,000 and 100 in 100,000 The first four of these levels were lower than or identical to objective risk. 72 percent of respondents stating one of the seven listed levels stated the four lower levels of risk, thus indicating that they regarded themselves as safer than an average driver in Sweden.

A very interesting table in the report shows the value of a statistical life estimated for risk reductions of 10, 25 or 50 percent for respondents rating their initial risk as one of the seven levels listed above. This gives a total of $7 \cdot 3 = 21$ data points. Figure 6.1 shows these data points.

There is a strong negative relationship between the size of the risk reduction and the value of a statistical life (VSL). Moreover, the range for the value of a statistical life is enormous, from 971 million SEK to 4 million SEK, a very much wider range than Jones-Lee et al. found in their pioneering study.

Remember from Chapter 1 that:

$$\text{VSL} = \frac{WTP}{\Delta R}$$

Here ΔR denotes the change in risk. It follows that when VSL and ΔR are known, WTP can be estimated as:

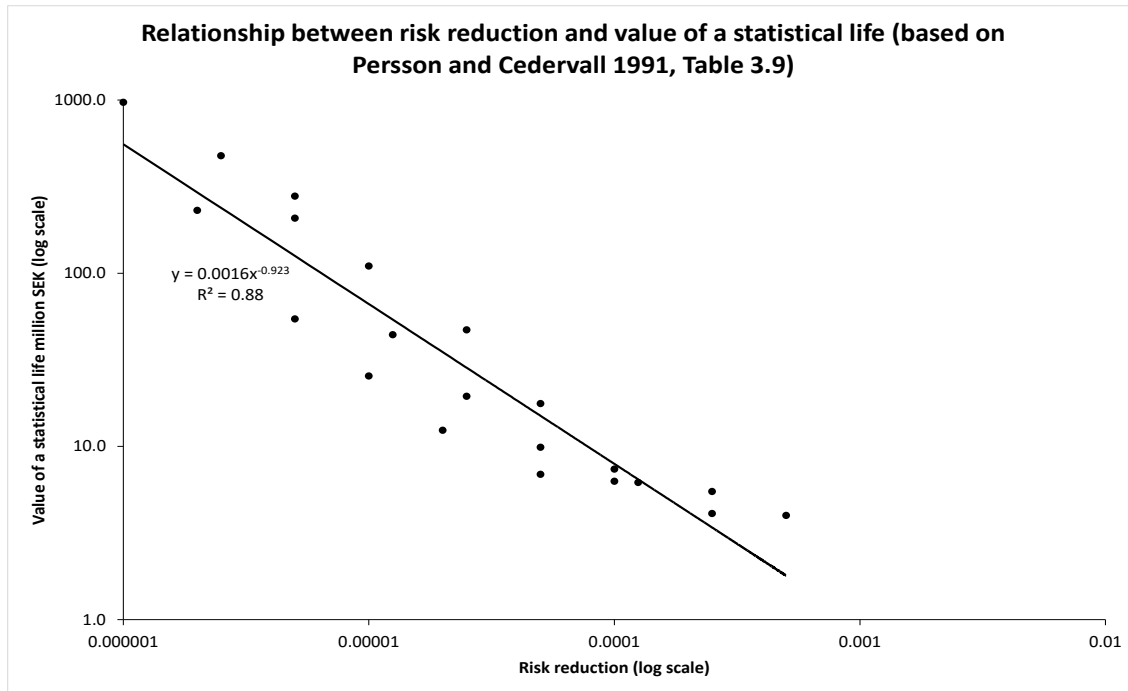$$\text{WTP} = \text{VSL} \cdot \Delta R$$

Figure 6.1: Relationship between size of risk reduction and value of a statistical life. Based on Persson and Cedervall 1991

One may therefore test whether willingness-to-pay varies in proportion, or nearly in proportion, to the size of the change in risk as theory predicted it would at the time Persson and Cedervall published their study. Figure 6.2 shows the relationship between risk reduction and willingness-to-pay.

Except for the data point to the far right, there is no tendency for willingness-to-pay to increase as the size of the risk reduction increases. There is strong insensitivity to scope. Persson and Cedervall (1991) discussed the issue, but argued that answers to questions referring to different sizes of changes in risk were given by different individuals and therefore did not necessarily show that individuals were insensitive to changes in risk.

Median values of a statistical life were considerably lower than mean values; indeed the mean values were about ten times higher than median values. On the whole, however, Persson and Cedervall were fairly optimistic in interpreting their findings. At one point, their study was more successful than the study made by Jones-Lee et al. In both studies, the following question was asked in order to probe how well respondents understand low probabilities and changes in them:

"Imagine that you, at the same time, have to face two different risks of being killed:

- *In the 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 one would you prefer:

- *The risk of 2 in 100,000 is reduced to 1 in 100,000*
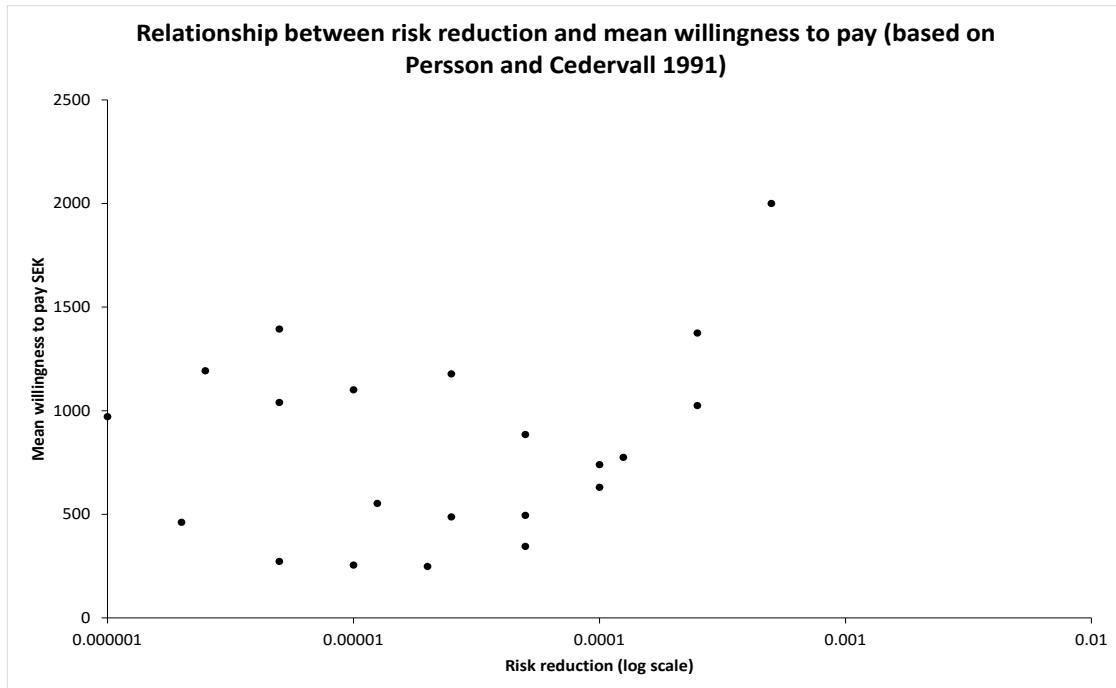- *The risk of 20 in 100,000 is reduced to 15 in 100,000"*

Figure 6.2: Relationship between size of risk reduction and mean willingness-to-pay. Based on Persson and Cedervall 1991

Given the conditions specified, the correct answer is to reduce the higher risk. In the study by Jones-Lee et al. 48 percent chose the higher risk, 47 percent the lower risk and 5 percent did not know. However, as asked by Jones-Lee et al., the question could be (mis)interpreted as a choice between facing a risk of 1 in 100,000 or a risk of 15 in 100,000, in which case choosing the lower risk would be rational. Persson and Cedervall emphasised that individuals were exposed to both risks, but had to choose which of them to reduce. In their study, 62 percent chose to reduce the higher risk, 20 percent chose to reduce the lower risk and 18 percent did not have an opinion.

In their discussion, Persson and Cedervall listed arguments for and against believing in the results of their study. Although they recognised that there were anomalies in the findings, the overall pattern was interpreted as sufficiently systematic to conclude that the results made sense and could be applied as a basis for valuation of road safety. Persson and Cedervall added that the results of their study were very similar to the results of the pioneering study in Great Britain.

6.1.5   Replication in New Zealand

The next replication of a transport safety valuation study employing the contingent valuation method was in New Zealand (Miller and Guria 1991). The study took place in 1989-1990 and was co-ordinated with the national household travel behaviour survey. Several questions were asked about willingness-to-pay for improved road safety. Before presenting the results of the study, the authors offered the following guidelines on the design and interpretation of contingent valuation studies:

1. The safety measure must be realistic: *"People should be asked the price they would pay for familiar goods and services, ones they believe work."*
2. Risk levels must be comprehensible: *"We believe the limit of understanding probably is around 1 in 10,000."*
3. Risks must be realistic: *"People get confused when risks or risk reductions are unrealistic."*

4. Do not trust zero bids: *"Some people legitimately will not pay for risk reduction. … Most zero bids, however, are protests … As the literature suggested, we probed why people bid zero. The probes let us discard protest bids."*
5. Do not trust very high bids: *"Bids so large they defy rationality are another form of protest … These bids also should be discarded."*
6. Choose payment methods carefully: *"Payment methods affect bids. People hate taxes …"*

While the first three of these points represent good advice in designing a contingent valuation survey, points 4 and 5 comes close to rejecting results one does not like. Rather than rejecting results because they are zero or very high, it would be more pertinent to test how well respondents have understood the valuation task. Miller and Guria (1991) ended up by rejecting a very high share of answers by reference to points 4 and 5, without explaining more about these decisions than the short notes quoted above.

The questions intended to elicit willingness-to-pay dealt with: (a) paying to use a safer toll road compared to a less safe road without a toll; (b)a safety training course for the family; (c) extra safety feature on a car; (d) living in a safer neighbourhood, (e) extra taxes (in violation of point 6 above) for a series of road safety measures.

51 percent of answers to the safer toll road question were discarded. The mean value of a statistical life was 1,188,000 NZ dollars for answers that were accepted, 2,026,000 NZ dollars for answers that were rejected. 52 percent of answers to the safety training course question were discarded. Mean valuations (all amounts are in NZ dollars) were 803,000 for accepted answers and 427,000 for rejected answers. 52 percent of answers were discarded for the safer car question. Mean valuations were 1,064,000 for accepted answers and 975,000 for rejected answers. 8 percent of answers to the safer neighbourhood question were discarded. Mean valuations were 1,871,000 for accepted answers and 21,920,000 for rejected answers. Finally, 72 percent of answers to the questions about taxes for safer roads were discarded. Mean valuations were 1,323,000 for accepted answers and 1,722,000 for rejected answers.

These comparisons show that the range of valuations was greatly reduced by discarding answers that were classified as protest answers of zero, as offering too high valuations, or as being inconsistent with an answer given to a different valuation question. The range of mean values based on accepted answers was from 803,000 to 1,871,000 NZ dollars. The range of mean values based on all answers was from 427,000 to 21,920,000NZ dollars.

On the average, about half of the answers were rejected. The half that was retained made sense according to economic theory. The authors recommended adopting a value of 2 million NZ dollars for saving a life. According to the foreword to the report, the New Zealand Minister for Transport accepted this recommendation and decided that a value of 2 million NZ dollars per statistical life should be used in all evaluations of transport projects.

6.1.6 Replication in Denmark

The next replication of a valuation survey using the contingent valuation method was in Denmark, in the form of a PhD dissertation by Kristian Kidholm (Kidholm 1995). The survey was conducted in February 1993. It included a valuation of the prevention of serious and slight injury in addition to the valuation of preventing traffic fatalities.

Kidholm discussed various sources of error and bias in contingent valuation surveys. The first of these was strategic answers, which he, based on previous studies, regarded as unlikely. The second was hypothetical bias, i.e. since no real payment is made in a contingent valuation survey, respondents may be tempted to inflate their willingness-to-pay. Kidholm accepted that this could be a source of bias. The third source of bias was

embedding effects. An embedding effect may occur when valuations are sought for a more than one non-market good in the same survey. It has then been found that the valuation of a given good A is smaller when it is part of a package consisting of goods A, B and C, than when it is considered on its own. The fourth source of bias was called the purchase of moral satisfaction (a name introduced by Kahneman). The idea is that people will state a positive willingness-to-pay simply to be politically correct or support a good cause. If valuations are mainly expressions of moral satisfaction, one would not expect them to vary according to the amount of the good on offer, i.e. moral satisfaction as an underlying motive is consistent with insensitivity to scope. The fifth source of bias discussed by Kidholm concerns the possibility that preferences do not really exist, but are simply made up when answering a valuation survey. He does not reject this possibility and states that the answer to it must be to retest a sample a second time in order to test the consistency of answers over time.

Despite the extensive discussion of the problems a contingent valuation study may encounter, Kidholm concluded that it was a feasible approach for studying how Danes value improving road safety.

Like Persson and Cedervall (1991), Kidholm (1995) asked respondents about their subjective risk of dying in traffic. Before being asked about this, respondents were informed about the mean traffic fatality risk in Denmark at the time of the survey, 11 in 100,000. Unsurprisingly, nearly half the respondents answered that their own risk was equal to the mean objective risk. Kidholm also tested the understanding of changes in low probabilities by asking the same 2 in 100,000 versus 20 in 100,000 question that was asked by Jones-Lee et al. and Persson and Cedervall. 61 percent preferred to reduce the higher risk, 31 percent preferred to reduce the lower risk, 8 percent did not make a choice.

Willingness-to-pay for reducing fatality risk was elicited by means of three questions. Based on subjective risk, estimated mean values of a statistical life ranged between 70.1 and 245.4 million DKK. Based on objective risk, the corresponding range was from 32.5 to 69.4 million DKK. The higher values obtained when relying on subjective estimates of risk show that respondents on the average rated their personal risks as lower than the objective mean value (since VSL is obtained by dividing WTP by change in risk, the lower the denominator, the higher will be the value of VSL). Median values were very much lower than mean values, ranging between 0.2 and 25.1 million DKK per statistical life based on subjective risk and between 0.1 and 18.4 million DKK based on objective risk.

In discussing the validity of the results, Kidholm gave six arguments for regarding them as valid, four arguments to the contrary. He performed a retest by conducting a second interview with 200 respondents (the total in the first survey was 945). Although positive correlations were found, Kidholm concluded that the test was fundamentally ambiguous, since one cannot rule out that some respondents have changed their valuations between the first and the second interview. He concluded by recommending a value of a statistical life based on median willingness-to-pay.

The current official value of a statistical life in Denmark is close to the recommendation by Kidholm, but it has not been possible to ascertain whether the value originates in his study or has another basis.

6.1.7 Replication in Switzerland

Schwab Christe (1995) reported on a replication of a valuation of safety study by means of the contingent valuation method made in Switzerland. The study was modelled on the previous studies in Great Britain (Jones-Lee et al. 1983) and Sweden (Persson and Cedervall 1991).

To test the understanding of changes in low levels of risk, reductions of risk of 20 in 50,000 and 5 in 50,000 were compared. 66 percent of respondents chose the lower risk, 32 percent the higher risk and 2 percent did not make a choice. Thus, unlike what was found by both Persson and Cedervall and Kidholm, the majority chose to reduce the smaller of the risks.

Willingness-to-pay was elicited for a 50 percent reduction in fatality risk and the risk of four types of injury. Mean willingness-to-pay was highest for reducing the risk of fatal injury. As found in other studies, mean willingness-to-pay was higher than median willingness-to-pay.

The study was an open ended contingent valuation study, i.e. no answer was suggested unless the respondent indicated that he or she wanted a starting bid. Willingness-to-pay was found to be higher among those who answered the questions spontaneously than among those who were given a starting bid.

The study was a pilot survey only using a small sample. A study was made in a larger sample (496) and published in 1996 (Schwab Christe and Soguel 1996). Some changes in the method were made based on the pilot study, but the main elements were the same. The mean value of a statistical life was estimated to be 4.1 million Swiss francs and the median value 1.7 million Swiss francs.

6.1.8   Replication in France

Desaigues and Rabl (1995) presented a contingent valuation study of road safety made in France. The study was performed in 1994. Sample size was 1000. To avoid the problem of asking respondents about changes in low levels of risk, the study adopted a different approach.

Respondents were first asked about how many traffic fatalities they believed there was during one year in France. The mean number stated was 28,345; the median was 8000. The actual number in 1993 was 9568. On the average, therefore, the number of fatalities was considerably overestimated. A similar question about the annual number of injuries found that it was underestimated.

Respondents were informed about the actual annual numbers of traffic fatalities and injuries in France and then asked how much they were willing to pay for reducing the number of fatalities by 50, 100, 500, 1000, 2000 or 5000. The largest of these reductions represents a 52 percent reduction of the number of fatalities, using the recorded number in 1993 as basis. It should be noted that these reductions refer to the total number of fatalities and not to the risk facing each road user.

The results were striking. Figure 6.3 shows the relationship found in the study between the number of lives saved and mean willingness-to-pay.
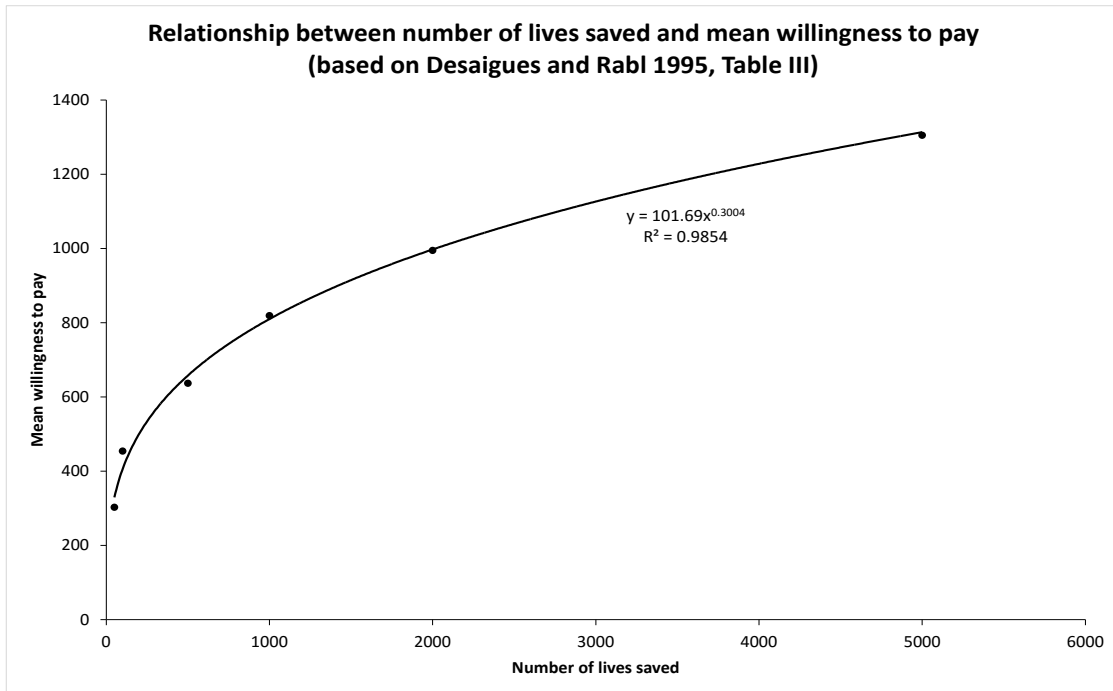
**Relationship between number of lives saved and mean willingness to pay
(based on Desaigues and Rabl 1995, Table III)**

$y = 101.69x^{0.3004}$
$R^2 = 0.9854$

Figure 6.3: Relationship between number of lives saved and mean willingness-to-pay. Based on Desaigues and Rabl 1995

It is seen that willingness-to-pay does increase as the number of lives saved increases. The increase is, however, far less than proportional to the increase in the number of lives saved. Saving 5000 lives – a hundred times more than saving 50 lives – is only valued 4.3 times higher. When the value of a statistical life is estimated, it declines sharply as the number of lives saved increases. Thus, the total benefit of saving 5000 lives is only 4.3 times as high as saving 50 lives. Moreover, answers were highly skewed and median values were only about 30 percent of mean values.

In other words, the attempt to make the risk reduction more comprehensible by stating it in terms of the number of lives saved, rather than in terms of reductions in low levels of risk, did not lead to a greater sensitivity to scope. It is perhaps not so surprising that many people do not notice, or are sensitive to, the difference between 2 in 100,000 and 6 in 100,000. The difference between 50 and 5000, on the other hand, is readily apparent and one might expect it to be associated with larger differences in WTP than what was found. It is possible that respondents had a safety budget and were not willing to spend more than the amount stated for the largest reduction of the number of fatalities. Still, it is difficult to explain why respondents would want to spend so much of their budget on saving a small number of lives.

6.1.9 Extension to non-fatal injury

Some the studies quoted above, tried to use the contingent valuation method to value changes in the risk of non-fatal injury. Persson and Cedervall (1991) concluded that the method seem to work, Kidholm (1995) was more negative. The largest and most systematic attempt to value changes in the risk of non-fatal injury by means of the contingent valuation method was made in Great Britain.

In 1989, short time after the Department of Transport in Great Britain decided to adopt the willingness-to-pay approach to the valuation of transport safety, the department decided to start a large research programme to obtain monetary valuations of non-fatal

injury. The contingent valuation approach was applied, but only as one of three approaches (O'Reilly et al. 1994). The other two approaches were:

1. The standard gamble approach, in which respondents were given a choice between two treatments for a given injury: (a) a standard treatment with an outcome that was certain (a state with some health impairment), and (b) a new treatment with normal health or a worse health state than the standard treatment as possible outcomes. The task was to determine the probability of the worse health state that would make respondents indifferent between the treatments.
2. The relative loss of utility approach, in which health state indexes were reviewed in order to quantify the quality of life associated with certain injuries.

In both the standard gamble approach and the relative loss of utility approach, one health state was valued in monetary terms and then used as a reference for valuing the other health states. The contingent valuation approach was found not to function very well. The valuations suggested that the prevention of even rather slight injuries was valued much higher than suggested by the standard gamble approach and the relative loss of utility approach. It was therefore concluded that the valuation of the prevention of injuries should be based on these approaches rather than the contingent valuation approach.

6.1.10 The end of the progressive phase

By reference to the methodology of scientific research programmes, it is fair to say that the period from about 1980 until about 1995 was a progressive phase in the use of the stated preference method for valuing transport safety, in particular the contingent valuation method. The reasons for regarding this period as a progressive phase include:

1. The contingent valuation method, a method most economists would discount as useless, proved to be workable. It was certainly not perfect, but it at least initially produced results that could reasonably be interpreted as showing mainly systematic (as opposed to nonsensical) patterns.
2. The contingent valuation method was quickly adopted by a number of researchers, leading to replication of the original British study in Austria, Sweden, New Zealand, Denmark, Switzerland and France. All replications relied exclusively on the contingent valuation method and most of them framed questions about valuation the same way as the original study.
3. In the early part of this phase, important theoretical contributions were made with respect to a number of issues. The theoretical results predicted patterns in empirical results that could be checked to assess whether these results made sense from a theoretical point of view.

A research programme is progressive as long as its empirical content increases, which was certainly the case for contingent valuation studies in the early phase, both because the refinement of theory predicted new empirical results and because at least some of these predictions were confirmed by empirical studies.

Yet, both unresolved theoretical problems and empirical anomalies existed from the very first day. Some of these problems will be discussed in the next chapter. It will then be shown that over time, anomalies became so numerous and severe as to tip the balance of power between those who defended the contingent valuation programme and those who criticised it. With very few exceptions, no valuation study made after about 1995 has relied exclusively on the contingent valuation method. Today it is common to combine at least two methods in the same valuation study.

What was regarded as a severe anomaly, insensitivity to scope, started to appear in more and more pronounced form towards the end of the progressive phase. It was very clearly evident in the Swedish study (Persson and Cedervall 1991) and in the French study (Desaigues and Rabl 1995), although in the latter safety benefits were presented in terms of the number of lives saved rather than changes in very low risk levels, which are more difficult to understand. In the end, as will be shown in the next chapter, this anomaly was regarded as so severe that it lead prominent researchers in the field to abandon the contingent valuation method in its classic form. This definitely marked the end of the progressive phase.

It is ironic that today, two plausible theoretical models – which in a sense posit the same mechanism – have been proposed to account for insensitivity to scope. One of the theories is the mental accounting model of behavioural economics, introduced by Richard Thaler (who, incidentally, was also one of the pioneers of valuation research; see next section). Briefly stated, this model states that money is not fungible in the sense usually assumed in economic theory, i.e. an individual or household does not think of money as something that can freely be spent on whatever they wish. On the contrary, money is compartmentalised. Some money is set aside on a savings account, for example. The savings account is treated as untouchable as far as daily expenses are concerned. You simply do not withdraw money from it to cover daily expenses. If you do, you are in trouble. As long as you know that you did not touch the savings account, your conscience is good and you confirm to yourself that you manage your private economy well.

In much the same way, you may have a budget for certain items. As long as your spending stays within or close to budget limits, you will not see strong reasons for changing it. You might well be willing to indicate your support for a good cause like road safety by saying that you are willing to pay a little for improving it. But only a little. You only have a small road safety budget, not a large one. After all, you probably do not want to cut back too much on other items in your daily consumption.

The other model is the directionally bounded utility function model introduced by Amiran and Hagen (2010). This model says pretty much the same as the mental accounting model. Very simply put, it says that there are limits to your trade-offs. Nobody is willing to give up more than a fairly small fraction of his or her income to pay for improved road safety.

However, even if these theories explain or predict insensitivity to scope, that does not mean that insensitivity to scope is not a problem. It can create severe problems in the application of monetary valuations, leading to highly counterintuitive, if not outright paradoxical choices. If valuations that are insensitive to scope are accepted as a basis for priority setting, one can forget about achieving the consistency in priority setting that was one of the main arguments economists made for the need to value safety in monetary terms.

6.2 The revealed preference approach – the discovery of new complexities

The starting point of the revealed preference approach to the valuation of safety is very different from the starting point of the stated preference approach. The theory of revealed preferences was originally developed by Paul Samuelson (1947) to enable more stringent mathematical analysis of market demand data. The key element of revealed preference theory is the so called weak axiom of revealed preferences. This states that if a consumer chooses A when A and B are both available and attainable within a given budget constraint, the consumer reveals that he or she prefers A to B.

In the strong axiom of revealed preference, two more conditions are added: (1) The consumer is never indifferent between A and B, but always prefers (read: chooses) one to the other; (2) Preferences are transitive. If these assumptions are made, consumer behaviour can be modelled by means of a utility function which the consumer maximises.

Thus, revealed preference studies involve trying to reconstruct the utility function of an individual based on market choices made by the individual. Revealed preference studies therefore presume the existence of a market, whereas one of the main justifications given for stated preference studies is that no market exists. These perspectives appear to be contradictory. How can proponents of stated preference studies claim that no market exists, whereas proponents of revealed preference studies claim that studying real choices made in real markets is the best way to obtain monetary valuations of safety?

There is obviously no market for safety where you can buy safety the same way you buy commodities like groceries or clothes. There are, however, markets in goods and services that have safety as one of their characteristics. When you buy a house, for example, you will consider several aspects in addition to the price: number of rooms, age, condition, and perhaps also exposure to natural hazards like floods or landslides. These are factors that influence the price of the house. Revealed preference studies try to reconstruct "what counts", or what influences an observed market price. Revealed preference studies can be used to obtain monetary valuations of safety whenever safety is one of the factors that influence price.

6.2.1   Compensating wage differentials

By far the largest number of revealed preference studies designed to estimate the value of a statistical life have been based on the theory of compensating wage differentials. The essential elements of this theory will therefore be briefly presented. The presentation is based on Viscusi (1993), Viscusi and Aldy (2003) and Kniesner, Viscusi and Ziliak (2014).

The theory goes back to Adam Smith, who proposed that workers are compensated for unpleasant aspects of their work by means of higher wages. One of these unpleasant aspects could be a high level of risk. A wage compensation for a high level of risk will exist under mild conditions. First, safety is costly. Employers must decide how much to spend on safety. In making this choice, they will look for the cheapest solution. At a point, improving safety becomes more costly than offering employees a higher wage to make them accept the risk. There will thus exist firms with varying levels of safety and varying levels of wages associated with these levels of safety. As far as workers are concerned, they will demand wage compensation for risk if they prefer to be healthy rather than injured (or dead) and the marginal utility of income is positive. The choices facing firms and workers can be shown in a diagram.
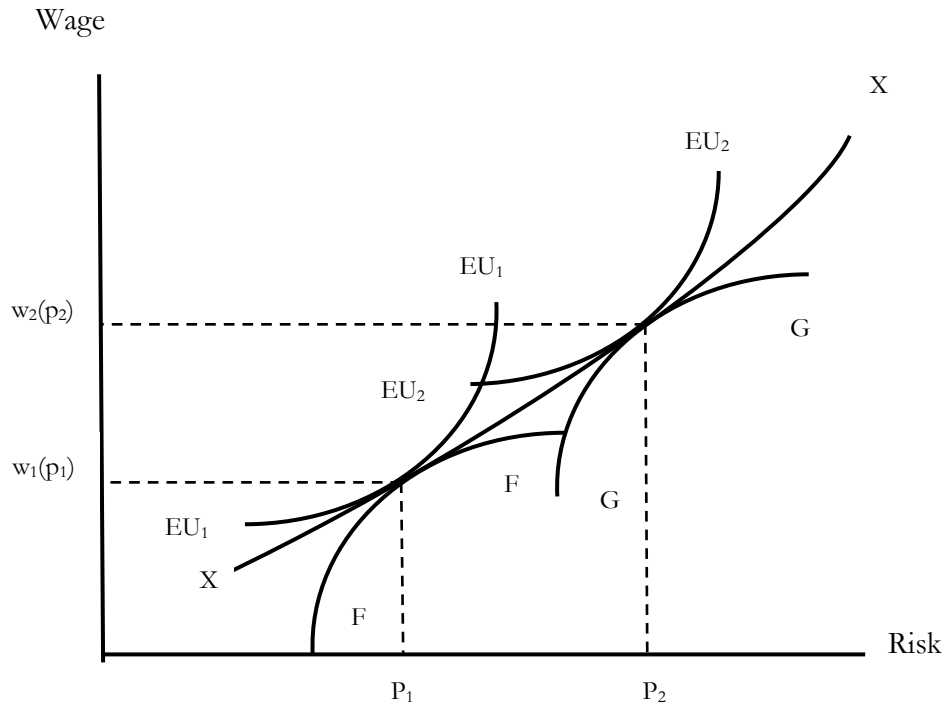
Wage

Figure 6.4: Illustration of the theory of compensating wage differentials. Based on Viscusi 1993

In Figure 6.4, FF and GG are the wage offer curves of two firms. EU_1 and EU_2 are the utility functions of worker 1 and worker 2. It is seen that the worker differ with respect to their preferences between risk and wages. Workers will choose the combination of risk and wages that give the highest utility. In Figure 6.4, the utility function of worker 1 intersects the wage offer curve at the point (p_1, w_1). For worker 2, the corresponding point of intersection is (p_2, w_2).

These data points are those that are observable in market data. The task of the analyst is to identify these data points by means of econometric analysis. The most common model is to specify a wage equation of the following form (often using the natural logarithm of wages as the dependent variable):

$$w_i = \alpha + \sum_{m=1}^{M} \psi_m \chi_{im} + \gamma_0 \rho_i + \gamma_1 q_i + \gamma_2 q_i WC_i + u_1$$

W_i is the wage of worker i, α is the constant term, the x_{im} are characteristics of the worker and of the job (age, experience, union membership, industry, etc.), ρ_i is the fatality rate of worker i, q_i is the injury rate of worker i, WC is the compensation paid in case of an injury (i.e. the economic support a worker gets while injured; in most of Europe part of the public sector social security system, in the United States more often part of private insurance systems) and u_i is the residual term.

The data used in studies of compensating wage differentials are typically combined from several sources. One main data source would contain data about wages and various characteristics of workers. The other main data source would contain data about risk. Risk data are then matched to wage data by applying codes for industry and occupation. The early studies often used quite crude risk data. Over time, more detailed data sources on risk have become available, in particular in the United States.

Ideally speaking, the wage equation should include everything that explains variation in wages. This is of course a very difficult requirement to fulfil. Thus, briefly stated, the history of research on compensating wage differentials has been about the discovery of ever more complexities or new sources of data.

In his first review of these studies, Viscusi (1993) stressed the importance of including two variables in addition to fatality risk:

1. Non-fatal risk
2. Workers compensation

With respect to the risk of non-fatal injury, Viscusi (1993) remarked (page 1919):

"Inclusion of this variable is sometimes difficult either because of the correlation between the death risk variable and the nonfatal risk measures or because differences in the data sources and the reference populations for which these data have been gathered may make it difficult to include both variables simultaneously. … The exclusion of the nonfatal injury variable may lead to an upward bias in the estimated coefficient for the fatality risks if the death risk variable's coefficient captures the omitted influence of the premiums for nonfatal risks, which should be positively correlated with fatality risks."

Viscusi remarked that most early studies of compensating wage differentials did not include a workers compensation variable. Including workers compensation is important, because workers compensation (i.e. income support given while a worker is on sick leave) reduces the difference in utility between a healthy state and an injured state. Viscusi and Evans (1990) fitted utility functions that depend on health status and found that the marginal utility of money is lower when health is impaired than in full health (one gets less enjoyment from something in reduced health than in full health).

The values of a statistical life estimated in compensating wage studies show the willingness-to-accept values associated with the mean risk in the sample (Kniesner, Viscusi and Ziliak 2014), i.e. the coefficient for risk refers to the level of risk and thus implicitly compares actual wages to those workers would get in a risk-free job. Unlike stated preference studies, compensating wage studies do not normally compare different changes in risk.

6.2.2 The progressive phase in studies of compensating wage differentials

As indicated in the title of this section, the history of the study of compensating wage differentials is the history about how economists have gradually discovered the complexity of estimating the wage premia associated with risky work. Against this background, the progressive phase of this research will be defined as follows:

Research on compensating wage differentials is progressive when improved control for potentially confounding variables is sustained in subsequent studies, i.e. these studies do not revert to simpler models of analysis that involve a poorer control for potentially confounding variables.

To assess whether studies of compensating wage differentials were progressive in this sense, 50 studies have been reviewed. Many of these studies were reviewed by Bellavance et al. (2009) – see Chapter 9. Others were listed by Viscusi (2014). The list includes those studies for which both the mean risk facing workers and their mean wage rate were known. For each of the 50 studies, control for the following potentially confounding variables was checked:

1. Risk of non-fatal injury, which is often highly correlated with the risk of fatal injury.
2. Workers compensation, i.e. income support given to injured workers.
3. Union membership; unions are often believed to have power to negotiate higher risk premia than those offered to non-unionised workers.

4. Endogeneity of risks, which refers to the fact that workplace risks are often to some degree under worker control and are thus not fixed once and for all.
5. Dimensions of risk variation, which refers to how detailed estimates of risk it is possible to develop based on a classification of workers by industry and occupation.

Figure 6.5 shows the proportion of studies controlling for these confounding factors by decade. There were too few studies in the 1970s to produce statistics.

The picture is mixed. As far as the risk of nonfatal injury is concerned, a majority of the most recent studies have controlled for it, which is an improvement compared to studies made thirty years ago. A minority of studies control for workers compensation, and recent studies are no better in this respect than studies made in the 1990s. Control for union membership has improved over time, although the share of workers in the United States who are union members has declined over time. The potential endogeneity of risk, an idea first introduced by Garen (1988), does not seem to have taken off. Only a minority of studies control for it or consider the issue at all.

One aspect that has improved greatly over time is the description of the risk facing workers. In the United States, four sources of data have been used. The oldest is a mortality table produced by the Society of Actuaries. It was used in some of the oldest studies. It shows overall mortality, not deaths in accidents. The next source of data was statistics produced by the Bureau of Labor Statistics. Although it identified major industries and occupations, many researchers regarded it as too crude. The third data source used in the United States was the National Traumatic Occupational Fatalities register developed by the National Institute of Occupational Safety and Health. Finally, after 2000, most studies have relied the Census of Fatal Occupational Injury, which is the most detailed data sources that has been developed in the United States on occupational fatalities.

It is worth noting that no study has controlled for all the five confounding factors listed in Figure 5.5.
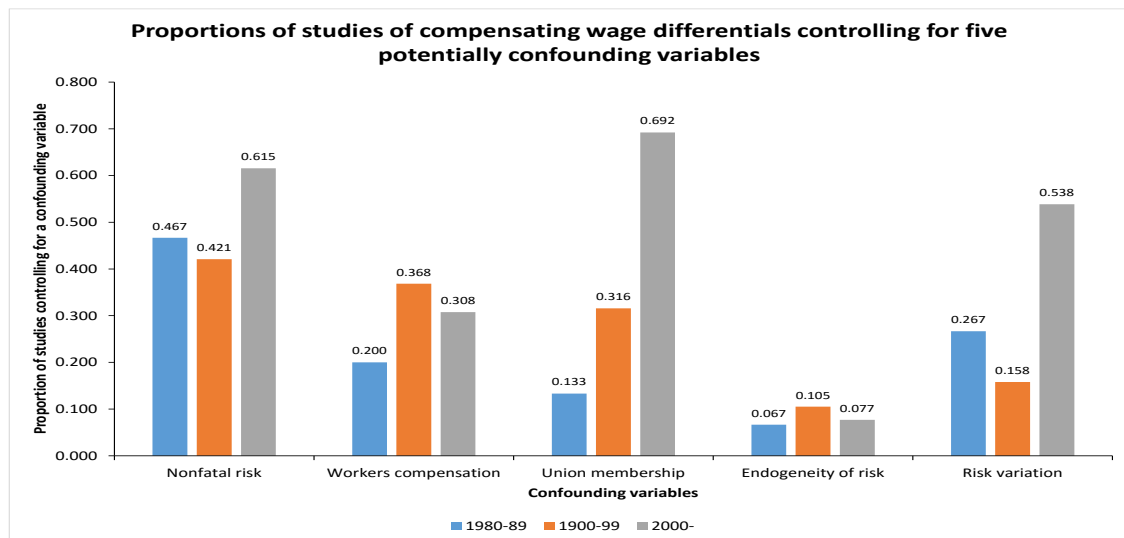
Figure 6.5: Proportions of studies of compensating wage differentials controlling for various confounding factors

According to the criterion proposed above, it is difficult to identify a progressive phase in studies of compensating wage differentials. Different econometric models have been used, depending on the characteristics and complexity of the data. Yet, even recent and comparatively advanced models fail to explain most of the variation in wages. Models

developed in the 1980s on average included 17.8 explanatory variables and had a mean R-squared value of 57.1 percent. In the 1990s, these numbers were 17.7 variables and 48.9 percent explained variance. This further declined to 12.8 variables and 35.8 percent explained variance after 2000. From the point of view of explaining wages, therefore, models do not seem to have improved their performance over time.

If, despite this somewhat untidy picture, one were to identify the end of the progressive phase of revealed preference studies based on compensating wage differentials, a suitable point in time might be 1996, when Peter Dorman published the book "Markets and mortality" (Dorman 1996). He was the first researcher who completely rejected the approach. When researchers who are active in a field come to reject the commonly applied method of research in the field, that is a sign that anomalies have become so widespread that at least some researchers conclude that the research programme no longer contains any positive heuristics worth pursuing. A similar breaking point occurred around 1997-98 for the contingent valuation method, when leading researchers, like Jones-Lee, rejected their own previous studies and decided to adopt a new approach.

Studies of compensating wage differentials have, however, continued more or less along the same lines as before 1996. Not everybody was convinced by the criticism put forward by Dorman. He was not influential enough to change the course of the research programme. Nevertheless, his criticism did not go entirely unnoticed and has been discussed in many of the contributions made to the compensating wage literature after he made it. Chapter 7 will further discuss the ongoing debate about the interpretation of studies of compensating wage differentials.

6.2.3   Other revealed preference approaches

While the compensating wage differentials framework has been the dominant revealed preferences approach to the valuation of safety, there have been a few studies of other market choices or behavioural choices. A few studies, starting with Atkinson and Halvorsen (1991) have studied car purchases. Some studies, including Blomquist (1979) and Blomquist, Miller and Levy (1996) have studied choices of behaviour, like wearing seat belts, crash helmets and child restraints.

These studies, and other similar studies, have been too few and far between to create their own research programme. They have been embedded in the broader research programme based on the revealed preference approach. The main objective of this study is not to give a detailed presentation of the results of each study that has sought to obtain a monetary valuation of transport safety. It is to understand the endurance of a scientific research programme that has been riddled with apparently anomalous results. To this end, the anomalous results are the most interesting. Nobody is surprised, and no explanation is required, if a successful research programme continues to prosper. Nothing succeeds like success, as the saying goes.

Therefore, the next chapter will highlight a number of anomalies and unresolved theoretical problems that have characterised studies of the monetary valuation of reduced risk of death.

6.3   Conclusions

The following main conclusions can be drawn based on the review of studies presented in this chapter:

1. It is possible to identify a progressive phase in valuation research based on the contingent valuation approach. This phase fulfils the main criterion of a progressive phase according to the methodology of scientific research programmes in that the empirical content of research increased, both (a) As a result of theoretical work and (b) As a result of many empirical studies that were close replications of each other in terms of research method. The findings of these studies were initially interpreted as showing a mostly meaningful pattern, although anomalous findings were present all the time.

2. After about ten years, anomalous findings in the contingent valuation approach became more dominant and the method, as applied to valuation of environmental goods, was strongly criticised in the United States. During the 1990s, some prominent researchers developed stronger misgivings about the contingent valuation approach and came to reject it. These researchers did, however, not reject valuation research as a research programme nor did they question the hard core assumptions underlying the research. They only rejected a particular research method and sought to develop other methods that might function better.

3. Research based on compensating wage differentials dominated in the United States. This research started in the 1970s, but the number of studies increased rapidly during the 1980s. In that sense, research was in a progressive phase. The approach was accepted and many replications were published.

4. From the late 1980s, there was increasing controversy between researchers about how best to implement compensating wage differentials studies. These controversies concerned the sources of risk data, the endogeneity of risk, exactly which confounding variables to control for when estimating the risk premium and how far to go in dividing the labour market into segments between which compensating wage differentials might differ.

5. Some researchers rejected the compensating wage differentials approach during the 1990s, but others continued to use it. The controversy thus remained unresolved, but critics of compensating wage differentials did not succeed in their calls for abandoning the approach altogether.

7 Anomalies and hard core complexities

7.1 Anomalies of the contingent valuation approach

By 1995 insensitivity to scope had been established as a major anomaly in the contingent valuation approach. It was not the only one.

A paper by Dubourg, Jones-Lee and Loomes (1997) presents a number of anomalies of the contingent valuation method. The authors warn readers early in the paper about what is coming (page 682):

"… The results we present may have even more radical implications, raising the possibility that, for many non-marketed goods, individual preferences may be so imperfectly formed, and conform so poorly with certain axioms of standard economic theory, that the viability of the whole CV approach is called into question."

The first question dealt with reduction in the risk of sustaining five different types of injury. The question applied the so called iterative bidding (IB) procedure, in which respondents were first shown an amount and asked if they were willing to pay that amount. If they answered yes, a higher amount was shown, and so on, until the respondent indicated that he or she was not willing to pay more. If a respondent answered no to the initial bid, a lower amount was suggested in successive iterations until the respondent accepted the bid.

Strong starting point bias was found. Half the sample were given an initial bid of 75 pounds. The other half were given an initial bid of 25 pounds. The mean willingness-to-pay – for identical risk reductions of identical injuries – was from 1.89 times to 2.87 times higher for the higher initial bid than for the lower. The results show that analysts can manipulate such surveys and get the results they want by stating a suitable starting bid.

Respondents were also asked about how sure they were about what they would pay. Three levels were defined. The lower level was the largest amount an individual definitely would pay. The upper level was the smallest amount an individual definitely would not pay. These two amounts can be interpreted as a personal confidence interval on the amount and individual might be willing to pay. Respondents were then asked which amount in-between their lower and upper limits was their "best estimate" of their willingness-to-pay. Denote by P^* the best estimate, by P_L the lower limit and by P_U the upper limit. One can then form the statistic: $(P^* - P_L)/(P_U - P_L)$, which can take on values between 0 and 1. The values of the statistic found for the 75 pounds starting point were in the range $0.17 - 0.41$, with most values close to 0.25. The values found for the 25 pounds starting point were in the range $0.22 - 0.54$, with most values close to 0.30. This shows that when the higher starting point was used, values tended to cluster closer to the bottom of the personal confidence interval than when the lower starting point was used.

The width of the interval $(P_U - P_L)$ was also found to display starting point bias, being consistently higher for the 75 pounds starting point than for the 25 pounds starting point. Despite this, the ratio P_{L75}/P_{U25} exceeded 1 in all cases, i.e. the lower limit for the 75 pounds starting point was always above the upper limit for the 25 pounds starting point. Dubourg et al. discuss their findings in view of the recommendation of the NOAA panel

on contingent valuation (Arrow et al. 1993) to use a direct choice (DC) method for eliciting willingness-to-pay in contingent valuation studies. The argument given for using this method was that in all common market transactions, an individual decides whether or not to buy a good at a stated price; he or she never decides what the price ought to be. Dubourg et al. remark (page 688):

"We note that this evidence of considerable degree of imprecision and powerful starting-point effects does not constitute an argument for rejecting the IB approach in favour of DC designs: on the contrary, the fact that the first value presented to respondents can have such a dramatic effect upon whole intervals elicited by an iterative process raises the serious possibility that the particular set of prices chosen by the survey designer to be presented to the various subsamples of respondents in a DC study may exert a major influence upon the shape and location of the demand schedule derived, and hence upon the mean WTP inferred from the survey."

Respondents in the survey conducted by Dubourg, Jones-Lee and Loomes were initially asked to rank the following five injuries according to severity (the injuries are listed here in the order that would seem to be the most reasonable, with 1 the most severe, 5 the least severe):

1. Death (labelled K in the survey)
2. Serious permanent disability (R)
3. Slight permanent disability (S)
4. Recovery in 1-3 years (X)
5. Recovery in 3-4 months (W)

Note that these injuries had different initial levels of risk. The survey elicited WTP for reducing each risk by 50 percent, which amounted to 4, 6, 8, 12, 18 and 10 in 100,000 for K, R, S, X and W. To compare willingness-to-pay for reducing the various injuries, WTP was stated as a proportion of the willingness-to-pay for reducing the risk of death. At first sight, results then seemed to make quite good sense: WTP for reducing the risk for each of the injuries was uniformly lower than for reducing the risk of death, and these differences were not influenced by starting point bias. However, another anomaly was found (Dubourg et al. 1997:690):

"… Despite the fact that the overwhelming majority of respondents rated injury W as clearly less bad than X in the ranking and scaling exercises, comparisons of m_w/m_k with m_x/m_k suggest that respondents are willing to pay between 30 % and 70 % more to reduce the risk of W by 1 in 100,000 per annum than to reduce the risk of X by the same amount."

They continue (page 691) by explaining:

"Inspection of the data strongly suggests that what is driving these results is a widespread lack of sensitivity among respondents, not only to reductions in the severity of injuries as they move from X to W, but, perhaps much more importantly, to differences in the magnitude of risk reduction. Thus, although the ranking/scaling exercise indicates that W is regarded by most people as clearly less bad than X, and although the risk reduction associated with W is only just over half of that associated with X, mean WTP in the question concerning W is not even 30 % less than the mean WTP in the question featuring X for the most sensitive subsample of respondents."

Thus, the whole edifice of contingent valuation studies is shattering in its foundations and in the process of collapsing completely. In stage 2 of the study, the main objective was to test sensitivity to the size of the reduction in risk. To reduce starting point bias, the iterative bidding approach was replaced by the use of payment cards. A payment card is a card where various amounts are printed and respondents select one of these. Two versions of the payment card were used. In one, the amounts varied between 0 and 500 pounds. In the other, amounts varied between 0 and 1500 pounds.

At this stage, it should perhaps not come as a surprise that mean WTP was higher for the 0-1500 payment card than for the 0-500 payment card. In fact, there was no overlap in WTP between the payment cards, the lower 0-1500 bound being higher than the upper 0-500 bound. WTP for reducing injury S by 12 in 100,000 (50 percent) was compared to reducing it by 4 in 100,000 (17 percent). Since the first risk reduction is three times greater than the second, one would expect WTP for the largest risk reduction to be, if not exactly three times as large, at least considerably larger than WTP for the smallest risk reduction. What was found? In a small sample of 33 respondents, the ratio of WTP for the larger risk reduction to the smaller was >2 for 4 respondents, exactly 2 for 4 respondents, between 1 and 2 for 14 respondents, exactly 1 for 9 respondents and <1 for 2 respondents. Depending on whether the group with a WTP-ratio between 1 and 2 are given the benefit of doubt or not, this suggests that between 33 percent and 76 percent of respondents are insensitive to scope. Dubourg et al. (page 697) comment:

"What appears to be important to respondents in these CV questions is that the safety feature is viewed as being a good thing, with the exact degree of goodness – especially since it involves a difficult conflation of very small probability changes with the implications of quite unfamiliar states of health – receiving only secondary attention."

The concerns raised in the paper by Dubourg et al. (1997) were echoed in a paper in 1998 by Beattie et al. (Beattie et al. 1998). Indeed, the authors stated in the abstract of the paper that the results cast serious doubt on the reliability and validity of willingness-to-pay based monetary values of safety estimated using the conventional contingent valuation procedure.

In phase 1 of the study, four injuries and their annual risk of occurrence were identified: fatal injury (6 in 100,000), serious permanent injury (20 in 100,000), serious temporary injury (50 in 100,000) and minor injury (500 in 100,000). Next, the following levels of risk reduction were defined: 1 in 100,000 (F1) and 3 in 100,000 (F3) for fatal injury, 10 in 100,000 (P) for serious permanent injury, 25 in 100,000 for serious temporary injury (T), and 250 in 100,000 for minor injury (M). Respondents were then asked about their willingness-to-pay for increasingly comprehensive packages of these risk reductions, viz.: [F1], [F3], [F3 + P], [F3 + P + T], [F3 + P + T + M]. The risk reductions associated with these packages were, respectively, 1 in 100,000, 3 in 100,000, 13 in 100,000, 38 in 100,000 and 288 in 100,000. The first two of these refer to different sizes of risk reductions for fatal injury. The other packages add successively less serious injuries. Although one would expect the prevention of a less serious injury to be valued less than the prevention of a fatality, these injuries are more numerous. On the whole, therefore, one would expect WTP to increase as the risk reductions become more comprehensive. Figure 7.1 shows the relationship that was found.
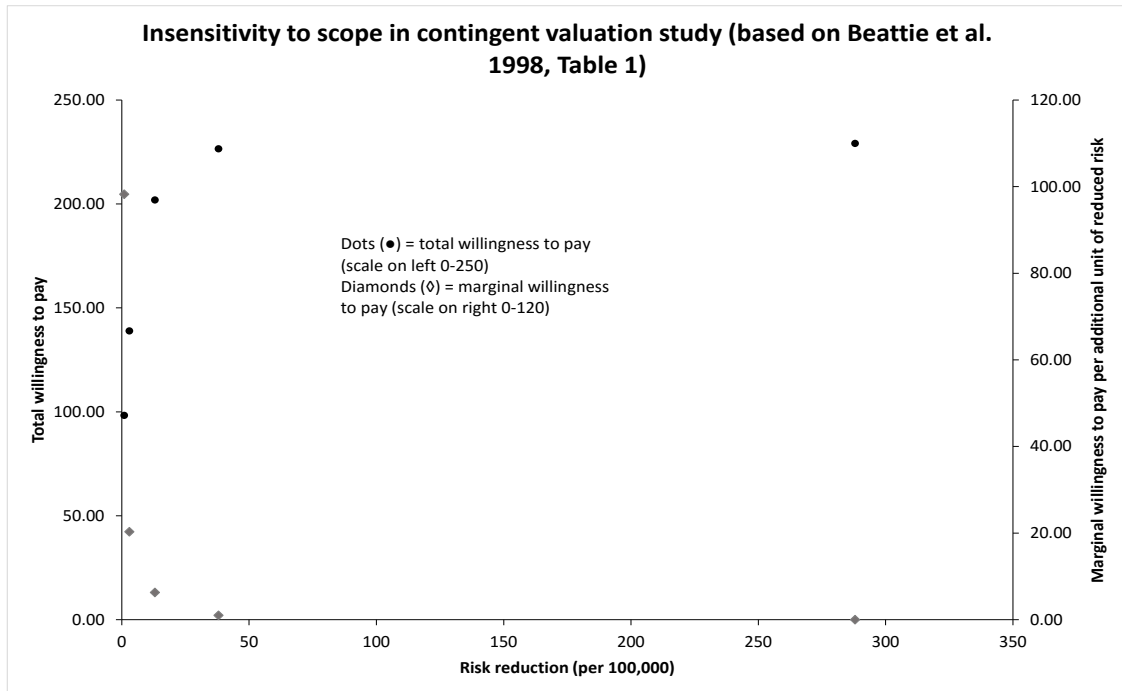
Figure 7.1: Insensitivity to scale and scope in willingness-to-pay study. Based on Beattie et al. 1998

The risk reduction F3 was three times larger than the risk reduction F1. Willingness-to-pay for F3 was, however, only 41 percent greater than willingness-to-pay for F1. The additional (marginal) willingness-to-pay for a more comprehensive package of risk reduction rapidly became smaller, with virtually no extra willingness-to-pay for reducing the risk of minor injury by 250 in 100,000.

It was suspected that insensitivity to scope may have been caused by difficulties in understanding changes in low levels of risk. In phase 2 of the study, risk reductions were therefore stated as the number of fatalities prevented in a community of 1 million people. Figure 7.2 shows the results of the study.

Stating safety benefits in terms of the number of fatalities prevented did not help. There was still insufficient sensitivity to scope. When the number of fatalities prevented increased from 5 to 75 (a factor of 15), mean willingness-to-pay increased only from 79.30 pounds to 196.14 pounds (a factor of 2.47).

Based on these findings, the authors concluded that the conventional contingent valuation approach to the valuation of safety had to be rejected. What they proposed to replace this method will be discussed in Chapter 8.

Two other studies published around 2000 confirmed the collapse of the contingent valuation method in its classic form. The first of these studies was reported in New Zealand (Guria et al. 1999). It confirmed the existence of starting point bias, although the authors conclude that no clear pattern was found. The tendency was, however, weaker than the similar tendency in the studies by Dubourg et al. (1997) and Beattie et al. (1998).
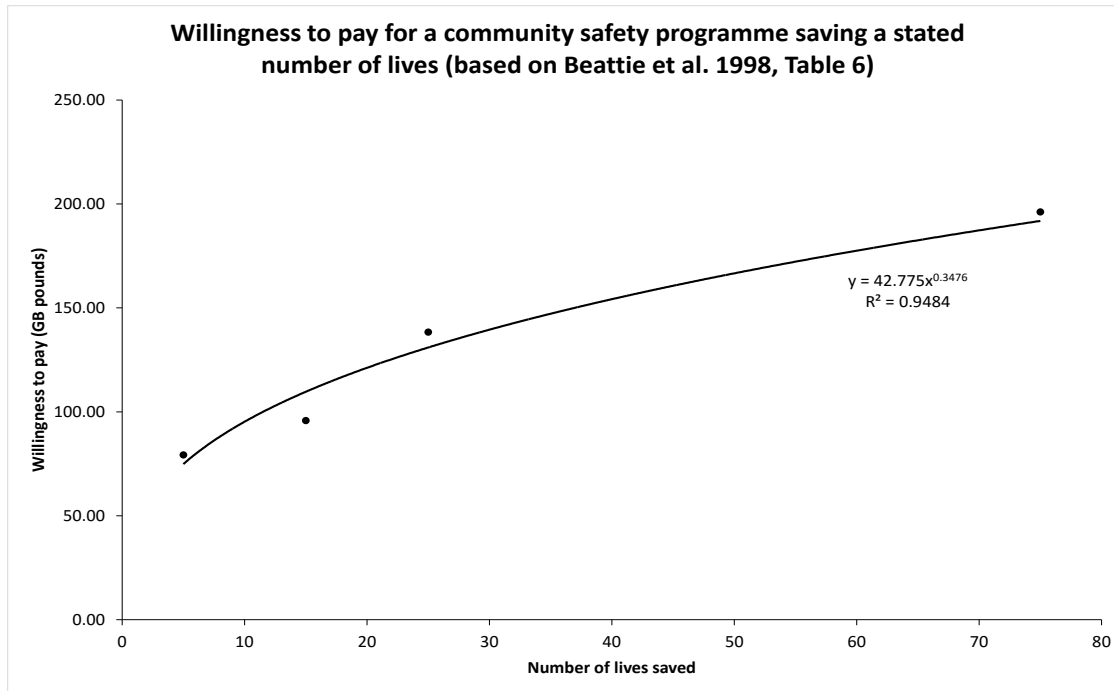
Figure 7.2: Willingness-to-pay to reduce the number of fatalities by a stated number. Based on Beattie et al. 1998

Finally, Persson et al. (2000) reported a contingent valuation study in Sweden. Figure 7.3 is taken from the study and shows the relationship between the size of the reduction in fatality risk and willingness-to-pay.

Note that WTP is plotted on a log scale. Each data point is for one individual. There was no relationship whatsoever between the size of the risk reduction and willingness-to-pay. Persson et al. tried to fit a function to the data points by means of minimum absolute deviation. The fitted function had an R-squared value of 0.003, i.e. it explained 0.3 percent of the variation in willingness-to-pay. Despite this very poor fit, Persson et al. used the fitted curve to derive values of a statistical life.

Based on the studies quoted above, it is clear that by the late 1990s, the contingent valuation method was not able to give reliable estimates of the monetary value of improving transport safety. Whether it ever had been able to do so is not entirely obvious, since even the earliest studies contained anomalies. These, however, were not judged to be dominant enough to reject the method, since there was – at least on a generous interpretation – a somewhat systematic pattern in results on top of the anomalies. For the pioneers of the approach, it may have been difficult to concede defeat from the outset. Having advocated the method, the pioneers had invested sufficient prestige in it to be inclined to defend it as best they could. Besides, there was no obvious alternative.

By the time the major anomalies – starting point bias and insensitivity to scope – came to be the normal findings of contingent valuation studies, it became clear that the method was vulnerable to manipulation, in particular with respect to the choice of starting point in iterative bidding, the choice of amounts listed on payment cards or the vehicle for payment (taxes or other forms of payment). By judiciously choosing values for these parameters, analysts could greatly influence the results of studies.
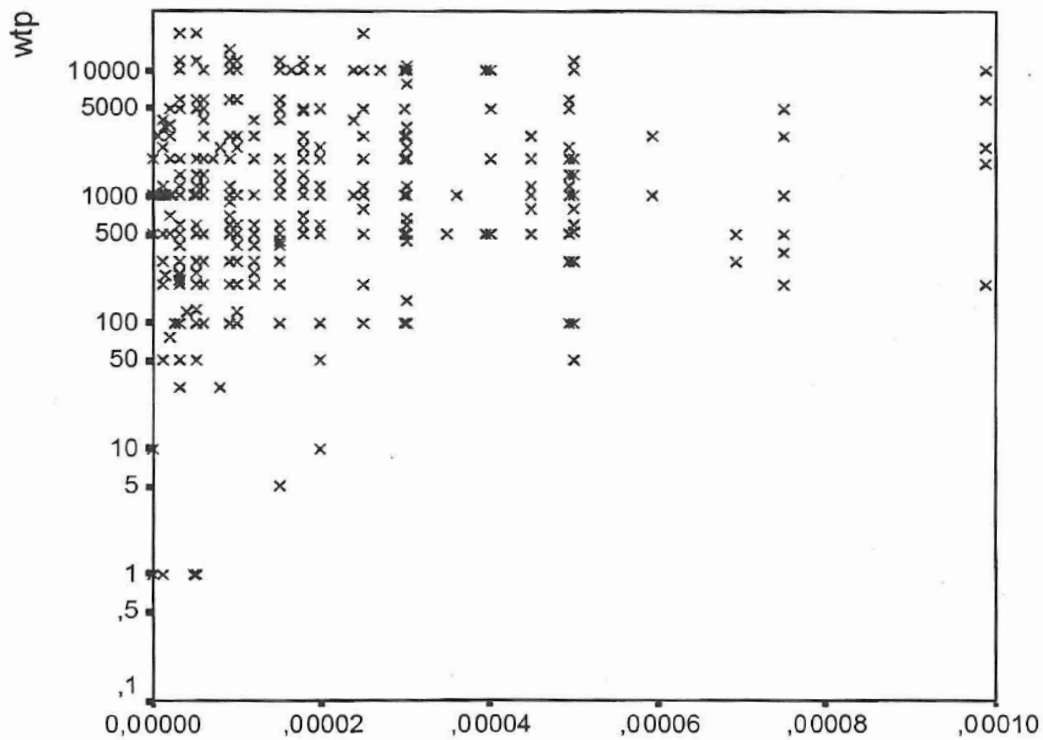
Figure 7.3: Relationship between size of reduction in fatality risk (e.g. 0.00006 = 6 in 100,000) and willingness-to-pay in Swedish contingent valuation study. Based on Persson et al. 2000

Insensitivity to scope, on the other hand, seemed to be more difficult to get rid of. It seemed clear that asking about changes in very low levels of risk made it more likely that there would be insensitivity to scope. However, asking about changes in the number of traffic fatalities directly, without introducing the concept of risk, did not seem to work.

Around the year 2000, there was thus a great need for either innovations in contingent valuation studies or for adopting different stated preference methods in valuation studies. Both these things happened and in the next chapter their impact on valuation studies will be discussed.

7.2 Problems associated with the non-linearity of demand

Given the fact that directionally bounded utility functions (see Chapter 5) predict insensitivity to scope, one might be tempted to conclude that this really is no anomaly and nothing to worry about. However, if one accepts the results of valuation studies displaying great insensitivity to scope, and rely on these studies as a basis for decisions on the provision of safety, choices that are inconsistent with basic criteria of rationality may result. This has been known for a long time and was one of the first points of criticism against the monetary valuation of safety (Broome 1982).

7.2.1 Choices that depend on irrelevant alternatives

The following discussion of how choices that are inconsistent with criteria of rationality may occur as a result of insensitivity to the size of a risk reduction draws heavily on Elvik (2013B). Suppose that a number of valuation studies have been made. Suppose further that the results of these studies can be summarised in terms of a demand function for transport safety. According to Lindhjem et al. (2011; see also Chapter 9), the value of a statistical life (i.e. a risk reduction which is expected to reduce the number of deaths by one, abbreviated VSL) can be modelled in terms of the following function:

$Ln(VSL) = 7.451 - 0.761 \cdot ln(\text{change in risk})$

For a change in risk of 1 in 1,000,000 (0.000001) this becomes:

$Ln(VSL) = 7.451 - 0.761 \cdot ln(0.000001) = 7.451 - 0.761 \cdot (-13.8155) = 17.9646$

By taking the exponential function of this, the estimated value of a statistical life becomes 63,376,490 US dollars (2005). Since VSL is obtained as the marginal rate of substitution between income and risk, mean willingness-to-pay for a risk reduction of 1 in 1,000,000 can be estimated as:

$WTP = VSL \cdot \text{risk change} = 63,376,490 \cdot 0.000001 = 63.38$.

The demand function is:

$WTP = 63.376 \cdot X^{0.239}$

In this function, X denotes the size of the change in risk, which is usually stated per 100,000 or per 1,000,000. Marginal willingness-to-pay is the first derivative of the demand function, which is:

$\text{Marginal WTP} = 15.147 \cdot X^{-0.761}$

The resulting values for WTP and VSL are shown in Table 7.1.

Table 7.1: Willingness-to-pay (WTP) for risk reductions and the value of a statistical life (VSL). Derived from Lindhjem et al. 2011

Risk reduction (per million)	Willingness-to-pay (US dollars 2005)	Marginal willingness-to-pay	Value of a statistical life (US dollars 2005)
1	63.38	15.15	63,376,490
5	93.11	4.45	18,621,386
10	109.88	2.63	10,988,241
15	121.06	1.93	8,070,914
20	129.68	1.55	6,484,020
50	161.43	0.77	3,228,583
100	190.51	0.46	1,905,146
200	224.84	0.27	1,124,202

It is seen that willingness-to-pay increases as the size of the risk reduction increases but not in proportion to the size of the risk reduction. Marginal willingness-to-pay shows the additional amount paid per additional unit of risk reduction. The value of a statistical life is obtained by dividing willingness-to-pay by the risk reduction, for example 109.88/0.00001 = 10,988,241. It can be seen that while willingness-to-pay increases as a function of the size of the risk reduction, the value of a statistical life declines as a function of the size of the

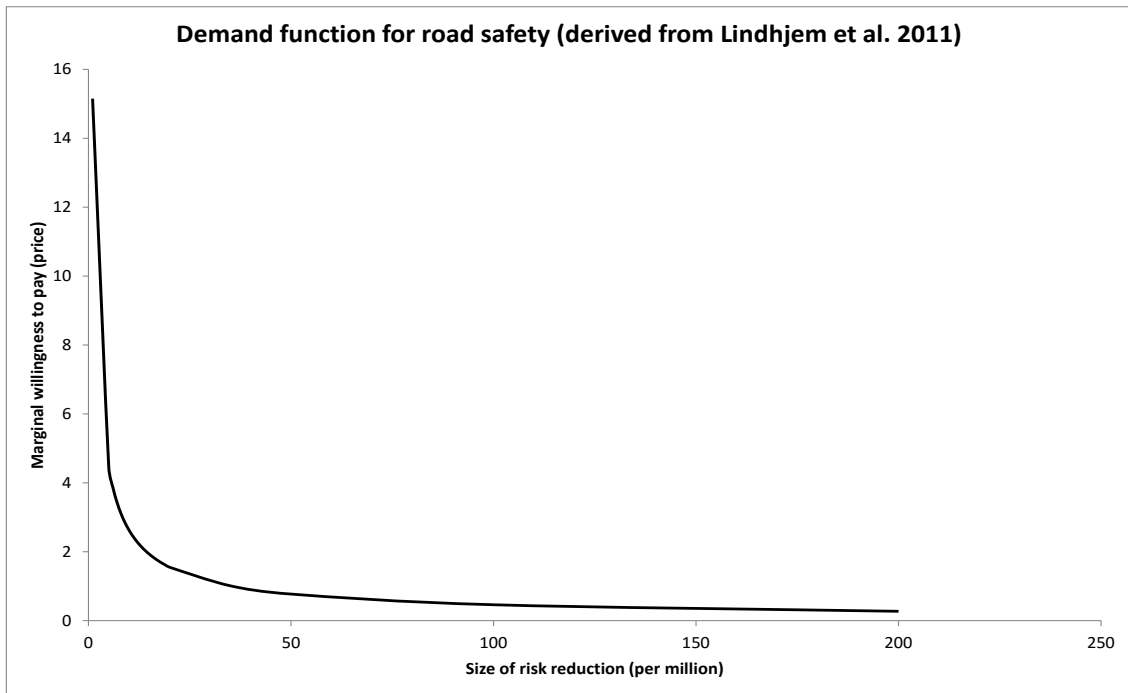risk reduction. The function assumed for willingness-to-pay implies the demand function shown in Figure 7.4.

Figure 7.4: Demand function for fatality risk reduction. Derived from Lindhjem et al. 2011

Suppose that a government agency tries to provide safety so that it exactly matches the demand for it, i.e. it strictly applies the demand function in Figure 7.4 (as far as is known, no government actually tries to do this, but the implications of doing so consistently are nevertheless interesting). For any project involving a change in risk, the government agency will estimate the size of the change in risk and apply the willingness-to-pay for a change of that size. Consider, as an example, what this implies for the choice between options A and B as shown in Table 7.2.

Table 7.2: Choice between options involving different changes in risk. Taken from Elvik 2013B

Characteristics	Option A	Option B
Initial risk	20 in 1.000.000	20 in 1.000.000
Risk reduction	2 in 1.000.000	10 in 1.000.000
Size of population	10,000,000	2,000,000
Fatalities prevented	20	20
Benefit in monetary terms (US dollars)	748 million	220 million

In both options an initial risk of 20 per million is reduced. In both cases the risk reduction results in an expected reduction of 20 fatalities. Thus, the options are identical with respect both to initial risk and the number of fatalities prevented and no basis for preferring one option to the other exists in terms of these characteristics.

If choice between these options is to be based on monetary benefit, option A will be chosen. The monetary benefit of saving 20 lives in option A is more than three times

greater than in option B. The reason for this is that the non-linearity of willingness-to-pay for safety with respect to the size of the risk reduction means that the value of a statistical life in option B is lower than in option A.

This result was discovered long ago. The first one to point it out was John Broome (1982), who argued that preferring one option to another when both options saved the same number of lives was a violation of the axiom of independence of irrelevant alternatives, which is one of the axioms of rational choice proposed by Von Neumann and Morgenstern.

It seems clear that Broome is right about this. One can imagine any number of combinations of background characteristics like initial risk, the size of the risk reduction, the size of the population benefitting from the risk reduction, the mean income of that population, the shape of the demand function, and so on, that would result in options that are:

1. Identical with respect to the safety benefits stated in natural units (lives saved, injuries prevented), and
2. Different in terms of the monetary valuation of the safety benefits.

If faced by a string of such choices, a decision maker adopting monetary benefits as the only criterion would in effect make the choice dependent on arbitrary factors influencing willingness-to-pay. It is fair to label these factors as arbitrary, since they are not subject to control by the decision maker and may vary randomly from one choice to another. In responding to Broome, Jones-Lee (1989:20) states that:

"... It is clear that under certain circumstances the dictates of coherence and consistency in government decision making will inevitably conflict with considerations of democracy (widely construed to include a requirement that government decisions should take account of individual wishes and attitudes to risk). In such conflicts, Broome appears to favour coherence whereas for advocates of the willingness-to-pay approach democracy is of primary importance."

This reply, although reasonable, does not really refute the argument made by Broome. Indeed, consistency in priority setting has been one of the main arguments economists have put forward to justify why a monetary valuation of life and limb is needed. It is therefore ironic when monetary valuations that are based on individual preferences do not ensure consistency in public policy based on these valuations.

7.2.2   Preference reversal as a result of preference aggregation

In Table 7.1, the column labelled willingness-to-pay shows individual preferences with respect to the provision of risk reductions of differing magnitudes. As can be seen, the largest risk reduction is the most preferred, the smallest risk reduction is the least preferred. These preferences are aggregated to form the value of a statistical life. As can be seen from Table 7.1, the value of a statistical life is highest for the smallest risk reduction, lowest for the largest risk reduction – exactly the opposite pattern of that found for individual willingness-to-pay. This may generate highly counterintuitive choices between options that involve a different number of lives saved. An example of such a choice is given in Table 7.3.

Table 7.3: Choice between options involving a different number of lives saved. Taken from Elvik (2013B)

Characteristics	Option A	Option B
Initial risk	20 in 1.000.000	20 in 1.000.000
Risk reduction	1 in 1.000.000	12 in 1.000.000
Size of population	10,000,000	2,000,000
Fatalities prevented	10	24
Benefit in monetary terms (US dollars)	634 million	230 million

Initial risk is identical in the two options, but option B reduces risk much more than option A. Option B saves 2.4 times as many lives as option A. Nevertheless, if monetary benefit is used as the criterion of choice, option A will be chosen. This is problematic for several reasons:

1. Option B reduces risk by 12 in 1 million, whereas option A only reduces risk by 1 in 1 million.
2. Option B results in a final level of risk (8 in 1 million) which is lower than that attained by option A (19 in 1 million).
3. Option B saves 2.4 times as many lives as option A (24 versus 10).
4. Willingness-to-pay for option B is almost twice as high (114.78 versus 63.38; estimated by applying the demand function shown above) as willingness-to-pay for option A.
5. If the options cost the same, preferring option A to option B can be considered as wasting money, since more lives could be saved by preferring option B.

Choices that are consistent with the monetary valuations in Table 7.3 may not necessarily arise as a result of a direct comparison of options, but may take place sequentially and be consistent with movements along a single demand function.

This example goes straight to the core of the argument made by economists for basing priorities for safety measures on cost-benefit analyses, rather than setting priorities informally. It has been argued (see, for example Hills and Jones-Lee 1983) that setting priorities informally entails the risk of using public funds inefficiently, thereby saving fewer lives than if priorities were set according to an economic criterion ensuring consistency. The choice of option A in the above example – which on the surface might appear suboptimal if one assumes that the two options cost the same – is however perfectly consistent with the monetary valuation of the lives saved. The problem is that this valuation is not the same for the two options. In general, one would not expect the monetary valuation of lives saved to be invariant with respect to background characteristics. In practice a common value of life which is invariant with respect to background characteristics is normally used. In that case, option B would be preferred.

7.3 Consistency between ex ante and ex post

The monetary valuation of changes in risk is obtained ex ante, i.e. it refers to risk stated as probabilities of certain outcomes. It does not refer to the valuation of specific events after they have occurred. Is this a problem? Ulph (1982) has pointed out that it could be a problem.

Suppose that willingness-to-pay for reduced risk of injury in road traffic is estimated. Most of these injuries are slight and do not result in permanent impairment. Let utility in the un-injured state be represented by the function $5 + 5 \cdot \ln(w)$, where w is annual income. In the numerical example, an annual income of NOK 600,000 will be assumed; this is close to the current GDP (gross domestic product) per capita in Norway. Let (just as a numerical example) the utility in the injured state be $5 + 4.5 \cdot \ln(w)$, meaning that the marginal utility of income is reduced by 10 percent.

Taking into account incomplete reporting of injuries in official statistics, the current risk of sustaining a traffic injury in Norway is about 8 in 1000. Given the utility functions, a rational utility maximiser with an annual income of 600,000 would be willing to pay NOK 799 to reduce the risk of injury by 1 in 1000 to 7 in 1000. The value of a prevented injury would be 799,000 (799/0.001).

Now suppose an individual sustained the injury. What is the compensation that individual would need to be offered in order to restore utility to the same level as the initial expected utility? By applying the utility function for the injured state, one can work out that the required compensation would be 2,000,351 NOK. This is considerably more than the ex ante willingness-to-pay for reducing the injury. Thus, if the money paid for reducing the risk was put aside on an account reserved for compensating the victims of injury, there would not be enough money to pay compensation to even a single individual, not to mention those seven individuals who, in a group of 1,000 would still be expected to be injured even after the reduction of risk.

There are two sources of the difference between the ex ante evaluation and the ex post evaluation. First, ex ante, risk exists only as a possible outcome. It is treated probabilistically and has a minimal impact on utility. Thus, for an annual income of 600,000, utility is 71.523 in the un-injured state, 64.871 in the injured state. Expected utility ex ante, using probabilities of 0.992 and 0.008 as weights, is 71.470, which is only 0.07 percent lower than utility in the un-injured state. Reducing the probability of injury to 0.007 slightly increases expected utility; a utility maximiser could sacrifice a little more than 0.13 percent of income to pay for this reduction while maintaining ex ante expected utility.

Second, when risk is resolved, seven individuals among one thousand will sustain the injury. Their utility is reduced to 64.781. To bring it back up to the ex ante expected level, their income would need to increase to 2,600,351. Thus, risk carries a much higher price tag once it is resolved than when it only exists as a possible outcome with a small probability. This is of course no surprise; indeed it only describes in simple terms the economic rationale for insurance.

Nevertheless, the mechanism operating here may perhaps be part of the explanation of a slightly paradoxical development over time in many highly motorised countries: as the number of fatalities has declined, the monetary valuation of preventing them has gone up. From an individual perspective, one would expect the opposite pattern: the lower a risk becomes, the less it is worth spending to further reduce it.

Figure 7.5 shows the number of traffic fatalities and the official monetary value of preventing a fatality in Sweden for selected years after 1965 (Persson 2003). The value of preventing a traffic fatality is stated in 2001-prices. It is seen that the value of preventing a traffic fatality has increased, while the number of traffic fatalities has decreased. If, in 1965, all traffic fatalities had been prevented, the total benefit would have been $1.8 \cdot 1313 = 2,363$ million SEK. In 2010, the benefit of eliminating traffic fatalities was $20.9 \cdot 266 = 5,559$ million SEK.

Similar changes over time can be found in many highly motorised countries. The changes probably reflect the combined effects of three trends: (1) Countries have become more

wealthy and are thus, all else equal, able to afford to spend more to prevent traffic fatalities. (2) Once it is seen that the number of traffic fatalities can be reduced, this may generate a new aspiration level for reducing them; additional efforts can only be justified if benefits are valued more highly. (3) The realisation that there are irreversible losses of welfare may grow; ex post does not equal ex ante.
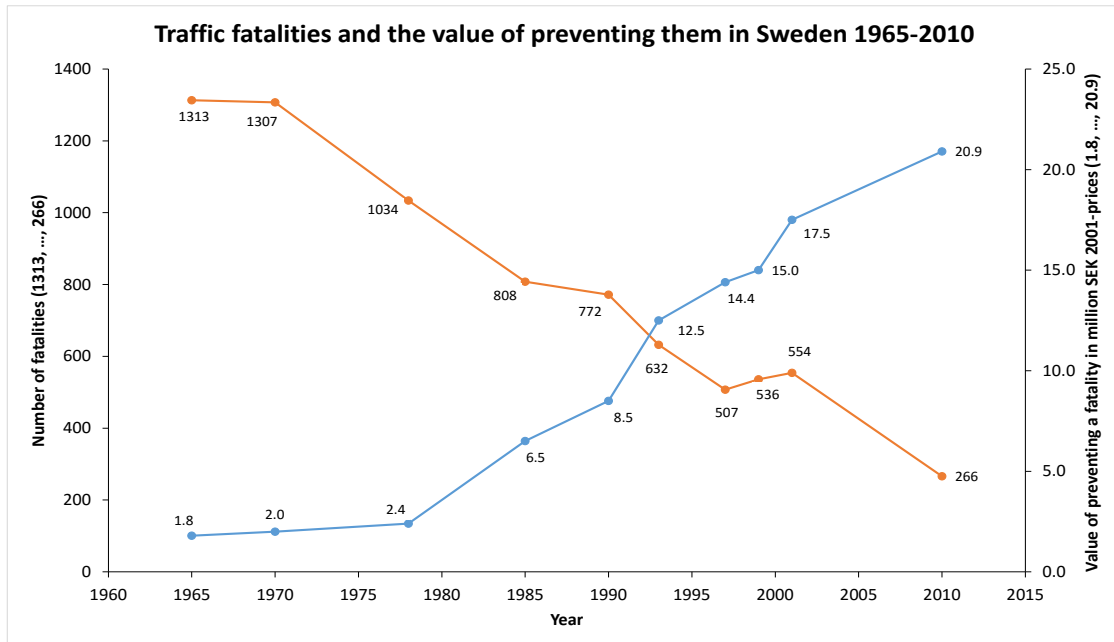

Figure 7.5: Traffic fatalities and the value of preventing them in Sweden 1965-2010

7.4   The possible non-existence of potential Pareto-improvements

Blackorby and Donaldson (1986) asked: Can risk-benefit analysis provide consistent policy evaluations of projects involving loss of life? Their answer was no. What they did, was to identify a version of a paradox originally discovered by Tibor Scitovsky and bearing his name: The Scitovsky paradox. In the original formulation, the paradox shows that if in state A, it may be a Pareto improvement to move to state B. If in state B, all else equal, it may be a Pareto improvement to move to state A. In other words, the transitivity of ranking states by preference breaks down: A is better than B and B is better than A.

Blackorby and Donaldson build their case around a numerical example. They assume that there are two individuals with initial utility levels conditional on survival of 10,000 (person 1) and 10,000 (person 2). Their probabilities of survival are 0.75 (person 1) and 0.50 (person 2). Utility conditional on death equals zero for both individuals. Their expected utilities in the initial situation are then 7500 (0.75 · 10,000) for person 1 and 5000 (0.50 · 10,000) for person 2.

Now suppose that person 2 wants to start a business that will increase his utility conditional on survival to 14,000. It will not influence his probability of surviving. For person 1, however, the probability of survival drops to 0.50 as a result of a negative external effect of the business set up by person 2. Expected utility now drops from 7500 to 5000 for person 1. For person 2, it increases from 5000 to 7000. Will person 2 be able to compensate person 1 for his loss of utility while still making a net gain by starting the business?

The answer is no. The remain at the initial level of utility, person 1 would need 5000 in compensation, since his initial expected utility was 7500 and, with a survival probability of 0.50, an income of 10,000 + 5000 = 15,000 is needed to stay at the initial level of utility. But the additional income for person 2 is only 4000. He therefore does not earn enough to compensate person 1 for the reduction in survival probability. Therefore, the initial state (A) is to be preferred to the new state (B) in which person 2 has started his business.

What if B is the initial state? Will person 1 then be able to compensate person 2 for the loss in utility he suffers by giving up his business? Again, the answer is no. Person 2 needs to be compensated by 4000 to remain at the initial level of utility. However, person 1 gains only 2500 if person 2 closes down his business. Therefore, it is better to remain in state B than move to state A.

In discussing this paradox, Jones-Lee (1989) noted that probabilities of death assumed by Blackorby and Donaldson are vastly higher than normal all-cause mortality levels and traffic risks. He repeated the example, assuming that both individuals have initial utility levels of 10,000 conditional on survival and that their initial probabilities of survival are 0.999 for person 1 and 0.998 for person 2. He then asked what might happen if person 2 starts an activity that raises his utility in survival from 10,000 to 10,000 + X, while reducing the survival probability of person 1 to 0.998.

He then shows that the paradox identified by Blackorby and Donaldson can only arise if X (the gain in utility to person 2) is between 10.01 and 10.02, which must be regarded as very unlikely. In other words, as the probability of death goes to zero, so that the differences in utility associated with changes in the probability of death become smaller, the likelihood that the paradox will occur becomes smaller.

To this can be added that the paradox cannot occur at all if both individuals hold optimal insurance. This is shown by Elvik (1993). The example is worth discussing, since it shows in a very simple way how the existence of insurance can make an activity which increases risk Pareto-optimal. Arrange the data for the original situation as shown below:

	State A		State B	
	Survival probability	Utility in survival	Survival probability	Utility in survival
Person 1	0.75	10,000	0.50	10,000
Person 2	0.50	10,000	0.50	14,000

These data can be rearranged to form a variable called "societal risk". In rearranged form, the data are:

	State A		State B	
Outcomes	Probability	Loss of utility	Probability	Loss of utility
Both survive	0.375	0	0.250	0
Person 1 dies	0.125	10,000	0.250	10,000
Person 2 dies	0.375	10,000	0.250	14,000
Both die	0.125	20,000	0.250	24,000

Suppose both individuals buy optimal insurance. Optimal insurance makes utility independent of the outcome. Thus, for person 1 in state A, it is optimal to insure for 8000 by paying a premium of 2000. For person 1, expected utility then becomes:

Expected utility = $(0.75 \cdot 8000) + (0.25 \cdot 8000) = 8000$

Similarly, for person 2 in state A it is optimal to insure for 6667 for a premium of 3333.

In state B, it is optimal for person 2 to insure for 9333 by paying a premium of 4667. Expected utility for person 2 then is 9333. By paying person 1 a compensation of 1333, his initial utility is kept at 8000. Person 2 still makes a net gain from 6667 in state A to 8000 in state B.

Thus, sharing risk makes the move from state A to state B Pareto-optimal. However, the probability that both individuals die has increased from 0.125 to 0.25. The example is only valid if a positive utility of wealth conditional on death makes sense. If one thinks that this does not make sense, the paradox cannot be resolved by a risk-sharing scheme.

7.5   Conclusions

This chapter has described a new phase of valuation research, which may not be labelled progressive as that concept is used in the methodology of scientific research programmes. It should rather be seen as period characterised by a proliferation of anomalies and a realisation of the existence of unresolved problems in the theoretical foundations of valuation studies. More specifically, the following conclusions can be drawn:

1. It was increasingly found that the results of contingent valuation studies were influenced by anomalies, such as starting point bias, bidding range bias, hypothetical bias and insensitivity to scope. This lead to an increasing criticism of the method and scepticism towards it.

2. A number of complexities in the theoretical foundation for the monetary valuation of changes in risk were discovered and discussed. First, insensitivity to scope can lead to preference reversals when individual willingness-to-pay is aggregated to form an estimate of the value of a statistical life. WTP will be highest for largest changes in risk, VSL will be lowest for the largest changes in risk. This can lead to choices that are highly counterintuitive, almost paradoxical. Second, valuation ex ante may be inconsistent with valuation ex post. One risk resolves in terms of injuries, a fund based on ex ante willingness-to-pay is unlikely to be sufficient to compensate injury victims for their loss of utility. Third, there is a possibility that Pareto-improvements may be undefined: when in A, B is better and when in B, A is better. In practice, however, the probability of this happening is very small.

8   The struggle between progressive and degenerative tendencies

When anomalies start to become common in a scientific research programme, the positive heuristic may guide research in two directions. One direction is to try to develop new research methods in the hope of avoiding the anomalies by using these methods. The other direction is to reformulate theory so that the anomalies can be reinterpreted as normal findings. Eventually, both these guidelines were followed in valuation research. However, in the late 1990s, the focus was on research methods, not on reformulating theory. To preserve at least a broadly chronological presentation, this chapter will review the methodological innovations made around 2000 before discussing theoretical innovations. The title of the chapter was chosen because methodological innovations always hold the promise of bringing a degenerative research programme back on a progressive track. However, if the methodological innovations are unsuccessful, degenerative tendencies may gain the upper hand.

8.1   Methodological innovations in stated preference methods

The virtual collapse of the contingent valuation method in the late 1990s lead to a search for new methods relying on the stated preference approach. Three of these methodological innovations will be discussed in this section:

1. The standard gamble chained contingent valuation (SG-CV) approach developed by Jones-Lee et al.
2. The stated choice approach, which started to be used simultaneously by Persson, de Blaeij and Rizzi and Ortúzar, but is often credited to Rizzi and Ortúzar.
3. The addition of various mechanisms to the contingent valuation approach in order to reduce hypothetical bias.

The discussion will focus on whether these methodological innovations succeeded in removing the anomalies they were intended to address.

8.1.1   The standard gamble chained contingent valuation approach

The starting point for developing the standard gamble chained contingent valuation approach, as noted by Carthy et al. (1999:188), was that *"… it may be over-optimistic to expect people to be able to give considered and accurate answers to hypothetical questions which involve a direct trade-off between money and small changes in already small risks of death."* The new approach obtained monetary valuations in a four stage process.

In the first stage, respondents were asked about their willingness-to-pay for a quick and complete cure for a slight injury (notably an injury not leading to any permanent impairment), or for the compensation they would require for sustaining the injury. The idea was, first, to avoid asking people about changes in low levels of risk, since the injury was merely described in general terms without stating its probability or frequency of occurrence. Secondly, many respondents would have personal experience with slight

injuries and would therefore be able to relate more directly to the question than to a question involving the risk of death.

In the second stage, the answers given to the question about WTP or WTA for a slight injury were converted to marginal rates of substitution between wealth and the risk of a non-fatal injury by relying on minimal assumptions about individual rationality. The marginal rate of substitution, it may be recalled, is the amount you pay for a small change in risk, divided by the change in risk. The result of the division is either the value of a statistical life, or, in the present case, the value of a statistical injury.

In the third stage, the standard gamble was introduced. Respondents were asked to choose between two treatments for an injury. One treatment would either give a standard outcome (described as a certain health state) or death with a probability θ ($\theta > 0$). The other treatment would either result in a return to full health within 3-4 days or death with a probability π ($\pi > \theta$). The task for the respondent was to find the level of π where the he or she would be indifferent between the two treatments at a stated value for θ. This would then give the ratio of the rates of marginal substitution (WTP) between death and the slight injury m_D/m_I.

In the fourth stage the m_D/m_I ratio was "chained" to the estimate of WTP for the slight injury from stage two in order to obtain the value of a statistical life.

In the first study relying on this approach, respondents were presented with these descriptions of injuries:

Injury X: In hospital for 2 weeks, full recovery after 18 months.

Injury W: In hospital for 2-3 days, full recovery after 3-4 months.

For each of these injuries, respondents were asked for their WTP to avoid it or their WTA to sustain it. Next a standard gamble was introduced for injury X in which failure would result in death, followed by a standard gamble for injury W in which failure would result in a prognosis identical to injury X (i.e. 2 weeks, rather than 2-3 days in hospital, and 18 months, rather than 3-4 for full recovery).

Mean WTA was 6.9 times higher than mean WTP for injury W and 6.4 times higher for injury X. Injury X was rated as worse than injury W. Injury X was rated as 0.041 times as bad as death, i.e. about 25 such injuries were judged as equivalent to one death. Mean estimates of the value of a statistical life ranged between 2.62 and 3.41 million pounds; median values were considerable lower, ranging between 0.31 and 0.55 million pounds.

Although the SG-CV approach was feasible and researchers regarded the results as credible and reasonable, it has not found wide application. Indeed, the UK study in 1997-1998 appears to be the only major example of use of the method.

8.1.2 Stated choices

While avoiding to ask people about changes in small levels of risk, the SG-CV method still asked direct questions about willingness-to-pay. This task is bound to be unfamiliar to most people. We are not normally asked to state the price or value of something, in particular not something that clearly does not have a market value, like an injury. Proponents of the stated choice approach argued that this approach avoided both: (a) the need to ask about changes in low levels of risk, and (b) the need to ask people to state explicitly their willingness-to-pay.

In a much quoted paper, Rizzi and Ortúzar (2003) introduced the stated choice approach to the valuation of road safety. Respondents were asked about their choice of route when driving between Santiago and Valparaiso, two major cities in Chile located 120 km apart. There is a toll road between the two cities, offering a fast connection on a motorway. In

the stated choice task, respondents were asked to choose between A and B, in which A initially had the current toll rate, the current number of accidents per year and the current travel time. These three attributes were then varied systematically and respondents made repeated choices. Each respondent was asked to make nine choices. Respondents were informed about the annual number of fatalities on the route, which was 27 during 1996 and 1997.

342 responses that were suitable for analysis were received. 150 respondents answered lexicographically, i.e. they always chose the option that was best with respect to one of the attributes, regardless of how the option scored with respect to the other attributes. These respondents did not make trade-offs between the attributes, but assigned a privileged role to one of the attributes at the expense of the other two. Four logit models were develop to analyse choices and estimate the value of a statistical life implied by these choices. In the two models that included lexicographic choices, the value of a statistical life was 772,271 US dollars in model 1 and 1,286,064 US dollars in model 2. In the two models that excluded lexicographic choices, the value of a statistical life was 392,817 US dollars in model 1 and 381,473 US dollars in model 2. The decision about whether or not to include lexicographic choices thus had a major influence on the estimated value of a statistical life.

The choices involved both reductions and increases in the number of fatalities compared to the reference level. According to prospect theory (Kahneman and Tversky 1979) one would expect the value of a statistical life to be higher in choices involving an increase in the number of fatalities than in choices involving a reduction of the number of fatalities. This was indeed found. When lexicographic choices were included, the WTA value (increase in fatalities) was 1.68 times higher than the WTP value (reduction in fatalities). When lexicographic choices were excluded, WTA exceeded WTP by a factor of 1.92. The WTP value of a statistical life, excluding lexicographic choices, was 268,344 US dollars.

Thus, a fairly broad range of values were estimated, the highest (1,286,064) being 4.79 times higher than the lowest (268,344). Rizzi and Ortúzar state that if forced to propose a recommended value based on their study, they would propose a VSL of 285,000 US dollars. This is close to the lower end of the range of values estimated, but Rizzi and Ortúzar justify a conservative interpretation of their study in view of the fact that 44 percent of respondents chose lexicographically and did therefore not display the pattern of preferences expected by economic theory. An alternative explanation is that these respondents did not really have lexicographic preferences, but that the differences in the values of the attributes between the alternatives were always in a range that made one of the attributes dominant. 72 of the 150 respondents who chose lexicographically always did so with respect to the safety attribute. To induce these respondents to choose differently, one might have reduced the differences in safety (i.e. made them less important; safety is almost the same whatever I choose; I must therefore look for other differences) between the alternatives presented.

Rizzi and Ortúzar were actually not the first researchers to apply the stated choice method to value transport safety. Trawén, Hjalte, Norinder and Persson (1999) applied the method in 1999 and framed it as a choice between residential areas. An example of the alternative respondents were asked to choose between is given below:

ALTERNATIVE A	ALTERNATIVE B
Annual house-related costs 48,000	Annual house-related costs 50,000
Travel time to work or school 30 minutes	Travel time to work or school 20 minutes
Traffic fatality rate 13 per 100,000	Traffic fatality rate 7 per 100,000

Note that Trawén et al. (1999) chose to present safety in terms of a microscopic level of risk. The possibility of avoiding this was precisely one of the reasons Rizzi and Ortúzar gave for preferring the stated choice approach to the contingent valuation approach. Trawén et al. (1999) noted how the range of values provided in the alternatives determine the range of values of a statistical life that may emerge from a study. Thus, in the residential area choice tasks, the lowest possible value of a statistical life implied by the range provided was 9.1 million SEK; the highest possible value was 2460 million SEK.

The residential area choice task gave estimates of the value of a statistical life between 113.7 and 242 million SEK, depending on how family members were included. The range of these estimates was much smaller than the theoretically possible range of values. The stated choice values were considerably higher than the values estimated on the basis of a contingent valuation study made at the same time. However, one must wonder whether these very high valuations really reflect preferences or are merely artefacts of the design of the stated choice task. Some of the differences in the assumed house-related costs were quite large, which would imply a high value of a statistical life. Trawén et al. (1999) did not discuss lexicographic or inconsistent choices, which are the major anomalies of stated choice studies. These anomalies will now be discussed.

8.1.3 Anomalies in stated choice methods

In a series of papers, Sælensminde (2002, 2003, 2006) has discussed various anomalies in stated choice valuation studies. The principal anomalies are lexicographic choices, inconsistent choices and embedding effects.

Sælensminde (2006) notes that choices that appear to be lexicographic need not really be so. Apparently lexicographic choices can be the result of a simplification of the choice task, by which a respondent chooses to focus on only one attribute to reduce the mental effort. Apparently lexicographic choices may also be a result of too large differences in attribute levels, which could make one of the attributes dominate the others if a respondent regarded a particular attribute as more important than the other attributes. Finally, even choices made at random could look like they were lexicographic, in particular if there are few attributes and each respondent makes few choices.

It is not always possible to determine whether lexicographic choices are only apparently so, or show that respondents who refuse to trade off goods against each other, have misunderstood the task or have an extremely strong preference for one of the attributes. In the Norwegian value of travel time study in 1997, each respondent made nine choices between trips that differed in terms of travel time, travel cost, frequency of service (public transport) or presence of speed cameras (car trips). The share of lexicographic choices varied between 25.1 and 43.9 percent. Lexicographic choices were defined as those who consistently, in all nine choices, chose the alternative that was best with respect to a single attribute. Unsurprisingly, the estimated value of travel time for those who chose lexicographically with respect to it (i.e. always chose the shortest travel time) was considerably higher than for those who did not choose lexicographically. Thus, the presence of lexicographic choices does influence the estimated value of non-market goods. In stated choice studies of the value of safety, lexicographic choices based on the safety attribute have been found to be quite common.

The value of travel time study used as an example by Sælensminde (2006) also contained a contingent valuation study. This study made it possible to assess whether those who answered lexicographically with respect to travel time in the stated choice task really did have a higher valuation of travel time than those who did not answer lexicographically. This was found to be the case. Lexicographic choices do therefore, to some extent, reflect

real preferences. However, to find out whether this is the case, one in general needs two sources of information about preferences, which would not be the case if a study included only a stated choice task and no contingent valuation survey. Therefore, combining different stated preference methods in the same study enables a more systematic testing for anomalies than if only a single method is used.

Inconsistent choices is a second anomaly in stated choice studies. In most stated choice valuation tasks, respondents are asked to make a sequence of choices. In making a sequence of choices, one or more of the choices can imply a valuation that is inconsistent with previous or subsequent choices in the sequence. Sælensminde (2002) explains this by means of a simple example based in the Norwegian value of travel time study in 1997. The example is reproduced below as Figure 8.1, based on Figure 1 and 2 in Sælensminde (2002). The ray diagram used in the figure is explained in the text.

The upper panel (Figure 1) shows four stated choice tasks. In each task, a choice must be made between left (LHS) and right (RHS). There are three attributes: Travel cost, travel time and headway. Headway denotes the gap between successive departures. The choices refer to long trips by train; hence, the frequency of service is quite low with 4, 5 or 6 hours between departures (there is normally only some 3-6 trains per day per direction between the major cities in Norway). Assume for the moment that in the first task, a respondent cares only about price and travel time. If the respondent chooses LHS, he or she reveals a valuation of at least NOK 100 for saving one hour. This gives the intersect 0,100 on the abscissa for ray number 1 in the lower part of the figure (Figure 2). Now suppose the respondent does not care about travel time, only about price and headway. If the respondent chooses LHS in task 1, he or she reveals that each hour of waiting time saved is worth at least NOK 50. This gives the intersect 0,50 on the ordinate for ray number 1 in Figure 2. All points located on ray number 1 represent combined valuations of travel time and headway that are consistent with the continuity axiom of rational choice theory. The continuity axiom states that goods are tradable (i.e. trade-offs can be made between them) in arbitrarily small amounts. The rays 2, 3 and 4 in Figure 2 (lower part of Figure 8.1) represent choice tasks 2, 3 and 4 in the upper part of the diagram.

	Choice no. 1		Choice no. 2		Choice no. 3		Choice no. 4	
	LHS	RHS	LHS	RHS	LHS	RHS	LHS	RHS
Price (NOK)	300	200	300	200	300	225	300	225
Travel time (hour)	4	5	4	6	4	5	4	4,5
Headway (hour)	4	6	4	5	4	5	4	5

Figure 1. Four different choices between journeys described by the three attributes, price, travel time and headway. In each choice the respondent chooses whether he/she prefers the journey presented on the left hand side (LHS) or the journey presented on the right hand side (RHS).
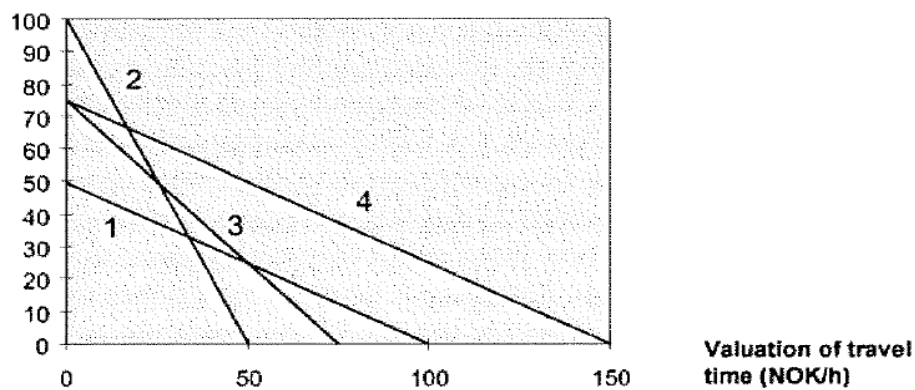

Figure 2. "Ray-diagram" based on the four choices in Figure 1.

Figure 8.1: How to test for inconsistent choices in stated choice valuation tasks. Based on Sælensminde (2002)

Successive choices made in a sequence are consistent if the valuation area (area below any of the rays) determined by one choice is inside the valuation area determined by another choice. If the valuation areas do not overlap, choices are inconsistent. Thus, if in choice 4, a respondent chose RHS, that choice would be located above ray number 4 and would thus be inconsistent with choosing LHS in choice 1, which would be located along ray number 1. Consistent choices throughout the entire sequence would imply valuations located inside the area delimited by 0-50 on the horizontal axis and 0-50 on the vertical axis. It is immediately seen that inconsistent choices are associated with higher valuations than consistent choices.

Based on this logic, two tests of consistency for a sequence of choices can be developed. The least restrictive test compares only pairs of choices, i.e. choice 2 versus 1, 3 versus 2, 4 versus 3, and so on. A more restrictive test compares a given choice, say 4, to the sequence of all previous choices 1, 2 and 3 and judges choices as inconsistent unless the full series of choices is consistent. Sælensminde (2002:411) noted that the number of inconsistent choices in a sequence can be counted in two ways. Suppose that choices 7 and 9 in a sequence are not consistent with choices 1, 2, 3, 4, 5, and 6. One may then count choices 7 and 9 as inconsistent with the previous choices, making for 2 inconsistent choices in a sequence of 9. However, one may also count the six first choices as inconsistent with the

subsequent choices 7 and 9, making for 6 inconsistent choices in a sequence of 9. Sælensminde opted for the lowest number.

Based on the stated choice tasks given in the Norwegian value of travel time study, Sælensminde (2002) found that between 60.3 and 75.5 percent of respondents made at least one inconsistent choice. Respondents who chose lexicographically had then been omitted. Respondents who chose inconsistently were found to have a higher valuation of travel time than those who chose consistently. This is shown in Figure 8.2.

It is seen that the mean value of travel time is considerably higher among those who chose inconsistently than among those who chose consistently. This means that the results of stated choice studies depend on whether or not lexicographic choices are included and on whether or not inconsistent choices are included. In the Norwegian value of travel time study, only a minority of respondents made choices that conformed to economic theory, i.e. compensatory (not lexicographic) and consistent. This is indicated by Figure 8.3.
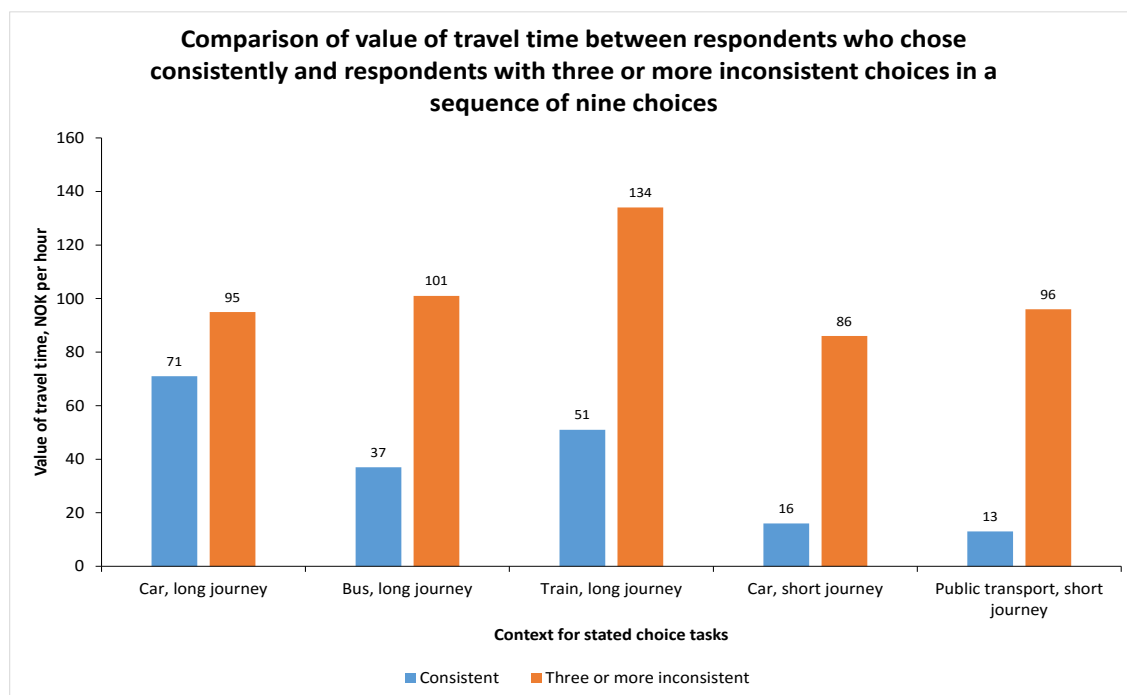

Figure 8.2: Comparison of value of travel time between respondents who chose consistently and respondents with inconsistent choices. Based on Sælensminde 2002

Figure 8.3 identifies three groups of respondents in stated choice tasks: (1) Those who choose lexicographically; (2) Those who choose inconsistently; (3) Economic man, i.e. those who choose in accordance with the rationality axioms of economic theory. It is seen that group 3 makes up only about 20 percent of all respondents.
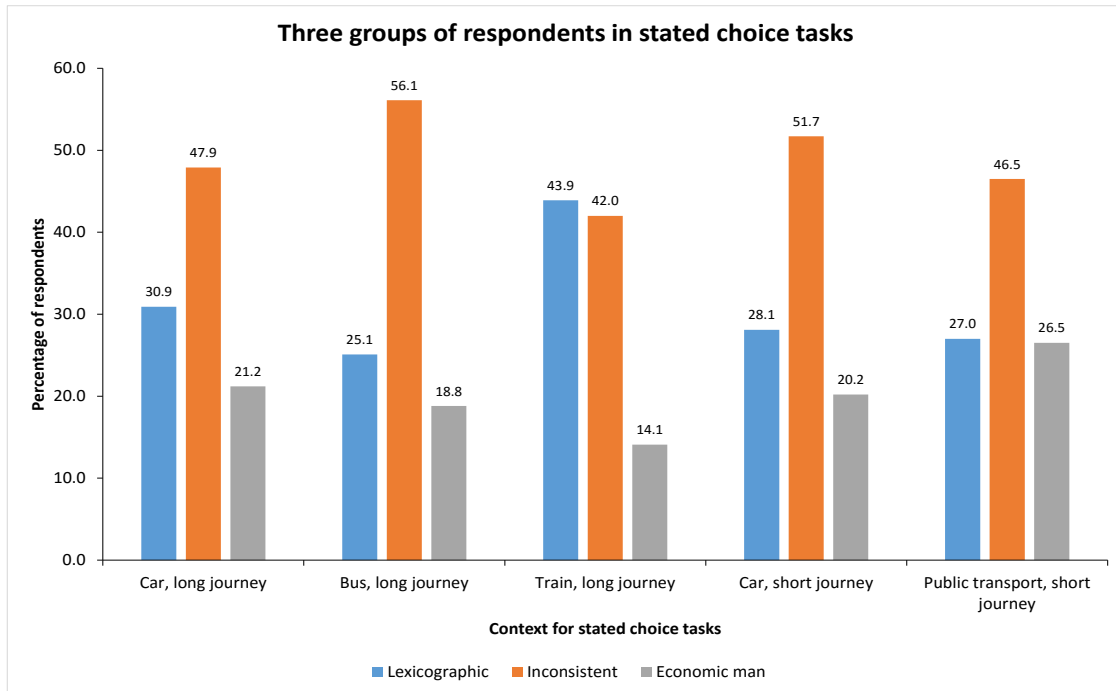
Figure 8.3: Three groups of respondents in stated choice tasks. Based on Sælensminde 2002, 2006

A third anomaly in stated preference studies discussed by Sælensminde (2003) was embedding effects. There is an embedding effect when the monetary valuation of a given good is lower when the good is valued as part of a package consisting of more than one non-market good than when it is valued alone. Sælensminde found embedding effects in a contingent valuation study for the monetary valuation of travel time, safety and environmental quality. When these goods were valued as a package, the mean total valuation was NOK 6144. When the goods were valued separately, the total mean valuation was NOK 8540, which is 39 percent higher than the valuation of the goods as a package.

It is tempting to think that embedding effects can be counteracted by adopting a stated choice approach, rather than a contingent valuation approach. However, as shown above, a sizable proportion of respondents do not trade off goods in stated choice tasks the way theory predicts. Moreover, the complexity of a stated choice task grows the more attributes one wants to include, increasing the probability of inconsistent choices. It is therefore unlikely that embedding effects can be avoided by relying on a stated choice design.

Randall and Hoehn (1996) point out that embedding can be expected according to standard economic theory and should therefore not necessarily be regarded as an anomaly if found in stated choice studies. At any rate, the fact that embedding effects are found when many attributes are included in stated choice tasks is no objection to including many attributes, but may, on the contrary, make the valuation task more realistic.

An attractive feature of stated choice studies is that they permit informative tests of respondent rationality, as shown above. Furthermore, the valuations obtained can be compared between respondents exhibiting varying degrees of rationality. In brief the methodological studies of Sælensminde and others show that:

1. Many respondents, often 20-40 percent choose lexicographically and do not make trade-offs between the attributes characterising the options. These respondents will have a higher valuation of the good they prefer lexically than respondents making trade-offs.

2. Many respondents make inconsistent choices, i.e. the valuation implied by one choice in a sequence of choices is not consistent with the valuation implied by a different choice. Respondents choosing inconsistently have a higher valuation of a good than respondents choosing consistently.
3. Valuing one non-market good in isolation results in higher values than valuing the good as part of a package consisting of more than one non-market good.
4. Analysts may to some extent influence the results of stated choice studies by their choice of attribute levels.

In sum, these lessons show that the problems encountered by the contingent valuation method are not solved by the stated choice method. Anomalies continue to occur, tending, as in the contingent valuation method, to produce too high valuations of the goods. However, by using both approaches in the same study they may to some extent cross-fertilise each other and act as a check on each other, permitting at least a limited assessment of convergent validity, i.e. whether they produce the same results in the same sample of respondents.

8.1.4 Controlling hypothetical bias in contingent valuation studies

In some contingent valuation studies, tests have been made to determine whether respondents are really willing to pay the amounts they state when asked about willingness-to-pay for something. An early meta-analysis of such studies, based on 29 studies providing 174 estimates of willingness-to-pay (List and Gallet 2001) found that stated WTP on the average overstates actual WTP by a factor of about 3. Thus, contingent valuation studies may greatly overestimate the value of non-market goods.

The meta-analysis tried to identify study characteristics that were associated with the ratio of stated to actual WTP. It found, among other things, that the disparity between hypothetical and actual WTP was greater for public (collective) goods than for private goods. It was greater when WTA was elicited than when WTP was elicited. Dichotomous choice (saying yes or no to a stated price) was associated with less hypothetical bias than open ended contingent valuation studies. However, none of the methodological aspects examined were able to eliminate hypothetical bias. The question therefore remains whether contingent valuation studies can be designed so as to entirely avoid hypothetical bias.

Dubourg (1995) introduced a response certainty scale to control for hypothetical bias. Respondents in contingent valuation studies are asked to indicate how certain they are about their answers. Certainty scales have evolved over time, and permit respondents to assign a certainty value to their answers on a scale ranging from 0 to 10. It has been found, see more details in Chapter 10, that respondents who indicate that there are highly certain about their answers have lower valuations than respondents who indicate that they are less certain about their answers. By relying on answers only from those respondents who are most certain, one may to a large extent reduce hypothetical bias in contingent valuation studies.

Veisten and Navrud (2006) tested a truth-telling mechanism and an elicitation format in a contingent valuation study of the passive-use value of preserving protected forest areas in Nordmarka close to Oslo. To ensure the protection of these areas, forest owners could lease them to the World Wildlife Fund (WWF). This would give them a compensation for not being able to exploit these forest areas commercially, as well as indicating the societal value of preserving the forest areas. Moreover, it provided an opportunity for testing actual willingness-to-pay by asking respondents in a contingent valuation study to contribute to the WWF forest fund. The contingent valuation study was conducted as a mailed questionnaire. Actual willingness-to-pay was tested in two ways: (1) One group received an

invoice from WWF one week after they answered the questionnaire; (2) Another group received the WWF invoice at the same time as the questionnaire. The idea was that the second group would feel a stronger motivation to pay and would thus be induced to give more truthful answers about their willingness-to-pay.

In the first group (bill one week later) less than 10 percent of respondents made an actual payment. On the average, actual payment was less than 10 percent of hypothetical payment. In the second group (bill and questionnaire at the same time), a little more than 10 percent of respondents made an actual payment. Although the payments made in this group were considerably higher than those made in the first group, they were not sufficient to eliminate hypothetical bias. In short, hypothetical bias was reduced, but not eliminated.

Morrison and Brown (2009) tested three instruments designed to reduce hypothetical bias in contingent valuation studies: (1) Cheap talk scripts, which are texts reminding respondents about budget limits and calling on them to consider the fact that by paying for a non-market good, they will have less money left for ordinary consumption; (2) Certainty scales, which allow respondents to indicate how certain they are about their stated willingness-to-pay; (3) Dissonance minimising, which is based on the dichotomous choice approach to contingent valuation. The idea is that many respondents feel that they must answer "yes" to a stated amount, even if the amount is more than they are actually willing to pay, because they do not want to come across as miserly or politically incorrect. However, when offered follow-up options, respondents may reveal that they do indeed have some positive WTP, but smaller than the amount stated in the dichotomous choice.

All the tested instruments were found to reduce hypothetical bias, with certainty scales and cognitive dissonance minimising being the most effective. If none of the instruments were used, hypothetical bias was found to be present.

Fifer, Rose and Greaves (2014) show that hypothetical bias can be found in stated choice studies. This has received less attention than hypothetical bias in contingent valuation studies, perhaps because it has been believed that trading off various attributes against each other will tend to reduce hypothetical bias. Their study was mainly exploratory, but it indicated that both cheap talk scripts and certainty scales can reduce hypothetical bias in stated choice studies. It may be noted that estimates of willingness-to-pay have often been found to be higher in stated choice studies than in contingent valuation studies. For an example, see Veisten et al. (2010).

To conclude, any good study employing the contingent valuation method or the stated choice method should test for the presence of hypothetical bias and try to reduce such bias if it is found. Unless this is done, there is no way of knowing whether or not a study is affected by hypothetical bias and how large any such bias is. However, given the fact that hypothetical bias has been found in very many valuation studies, it is not correct to assume that there is no such bias in a study that did not test for it. It is more correct to assume that the size of the bias is unknown.

8.2 Methodological innovations in statistical analysis

In nearly all valuation studies, there is a large variation in preferences, as indicated by the distribution of estimates of willingness-to-pay. In contingent valuation studies, the distribution will typically be skewed to the right. Mean willingness-to-pay may exceed median willingness-to-pay considerably. This, in turn, creates a dilemma with respect to which estimate to use in cost-benefit analysis. Economic theory is clear about this: The mean value should be used, because only the mean, multiplied by the number of respondents, will accurately reproduce the area under the demand curve. On the other

hand, if the mean value is larger than the amount the majority is willing to pay, a safety budget based on the mean value would be voted down in a referendum as being too expensive. Therefore, by invoking the median voter theorem of public choice theory (see, for example, Downs 1957), it has been argued that the median value is the one that would have the largest support and should therefore be used.

In stated choice studies, the task of the analyst is to describe as accurately as possible the factors influencing individual choices, i.e. to accurately model the utility function underlying these choices. In most stated choice studies, respondents are offered a choice between two options, a so called binary choice. This choice has traditionally been modelled by means of a random utility function (Ben-Akiva and Lerman 1985). The name "random" is used because the utility function, as specified by the analyst, contains a residual term:

$U_{in} = V_{in} + \varepsilon_{in}$

Here V is the systematic part of the utility function and ε is the residual term. The by far most widely applied model for analysing binary choices is the logit model. It predicts the probability that alternative i will be chosen over alternative j as follows:

$$P_n(i) = \frac{e^{\mu V_{in}}}{e^{\mu V_{in}} + e^{\mu V_{jn}}} = \frac{1}{1 + e^{-\mu(V_{in} - V_{jn})}}$$

This is the standard logit model for binary choice. Parameters describing the utility function are normally estimated by maximum likelihood methods. Around 2000, the mixed logit model was developed. This chief difference between the mixed logit model and the standard logit model is that it allows the marginal utility coefficients to vary randomly between respondents according to a specified distribution, for example the normal distribution (Hess 2010). This allows for a more accurate modelling of the heterogeneity of preferences.

Nowadays, the mixed logit model has become the standard approach to the analysis of stated choice data. Does it make a difference to the estimates? A recent Norwegian valuation study (Veisten et al. 2013) can be used to shed light on this question. Analyses were run using both a standard logit model and a mixed logit model. There were very small differences in the results. Mixed logit models must therefore be viewed mainly as a descriptive tool that permits a more precise analysis of the variation in preferences and may therefore explain more of the variation in willingness-to-pay than standard logit models.

8.3   New theory – progressive or ad hoc?

As noted in Chapter 5, new theories have been proposed after 2000 that seem to account for some of the anomalies of valuation studies, in particular insensitivity to scope, which is consistent with directionally bounded utility functions (Amiran and Hagen 2010) and therefore not necessarily an anomaly. Does the introduction of directionally bounded utility functions represent a progressive problem shift or a degenerative problem shift in the sense explained by Lakatos?

Remember that to be regarded as a progressive problem shift, a new theory should have a larger empirical content than an older theory, i.e. it should explain all empirical results that are consistent with the old theory as well as predict new empirical results. The opposite of a progressive problem shift is a degenerative problem shift. A new theory will then be an ad hoc theory, i.e. it explains a particular anomaly, but does not predict any new empirical findings in addition to the particular anomaly.

The case for directionally bounded utility functions, as made by Amiran and Hagen (2010), is quite plausible. In most situations, it makes sense that people have certain limits to what

they are willing to spend on a certain item. Once WTP gets close to that limit, it will not increase further even if more of the good is offered. The result will be an insensitivity to scope. Directionally bounded utility functions also predict that WTA can be much greater than WTP, indeed infinite.

Would finding sensitivity to scope be consistent with directionally bounded utility functions? Yes, it would, if the trade-offs are located far away from the directional boundaries. In that sense, the theory of directionally bounded utility functions would be able to accommodate results that are consistent with standard neoclassical utility theory.

There are, however, several problems with the theory. In the first place, if faced by a very high risk, such as premature death unless a life-saving operation is performed, an individual may be willing to spend as much as he or she can, the only constraint being that he or she wants to have an acceptable standard of living conditional on survival. This means that spending more than your annual income cannot be ruled out, provided the money came from a loan you would be able to make down payments on after the operation. This was predicted already by Jones-Lee (1974) within a neoclassical framework; his only condition was to assume that nobody was willing to go bankrupt to reduce a risk to their life. It is not uncommon to have a mortgage on your house that may exceed your annual income by a factor of, say, 2-3. Spending in the same order of magnitude on a life-saving operation cannot be ruled out. In these circumstances, therefore, the boundaries postulated by the theory of directionally bounded utility functions would be at least temporarily suspended.

In the second place, if insensitivity to scope is found, does that confirm the theory of directionally bounded utility functions? Obviously not. It is an elementary logical error, confirming the consequent, to conclude that a theory is confirmed if its predictions are supported empirically. The profound implications of this logical error was one of the reasons why Popper proposed to make falsifiability the key criterion of a truly scientific theory, as opposed to pseudo-science. A scientific theory contains hypotheses that can be falsified, i.e. whose empirical predictions can be contradicted by the facts.

While insensitivity to scope would be consistent with directionally bounded utility functions, it would of course also be consistent with many other theories. The theory of mental accounting, proposed by Thaler (1994), can account for insensitivity to scope. The theory of bounded rationality, more specifically its notion of aspiration levels (Simon 1982), would account for insensitivity to scope: Once safety exceeded the aspiration level, it would be "good enough" and spending more to improve it further would not make sense. Finally, attitude theory, as explained by Kahneman, Ritov and Schkade (1999), would account for insensitivity to scope.

If one adopts a restrictive interpretation of the theory of directionally bounded utility functions, regarding it as a theory that only predicts the WTA/WTP disparity and insensitivity to scope, it is clearly an ad hoc theory. Moreover, it is superfluous, as other theories, both within the field of economics and in other disciplines, also predict insensitivity to scope. If one adopts a more generous interpretation, according to which both sensitivity to scope and insensitivity to scope are regarded as consistent with directionally bounded utility functions, the theory becomes entirely vacuous since any outcome would be consistent with it. In general, once a theory ceases to be falsifiable, meaning that any observation would be consistent with it, it also ceases to have empirical content. Theories seeking to explain everything – in this case both sensitivity and insensitivity to scope – actually explain nothing.

To remain fruitful as a basis for empirical research, theories must therefore be falsifiable, i.e. there must at least in principle exist observations that contradict a theory and that, once made, would lead researchers to reject the theory. It is a key feature of a scientific research programme as characterised by Lakatos that it does not accept this logic. On the contrary,

apparent falsifications are not interpreted as real falsifications, and a theory is dogmatically upheld in the face of extensive evidence of its falsity until a better theory is developed. A better theory, in the Lakatosian sense, is a theory that both: (1) explains empirical findings that were consistent with the established theory; (2) in addition explains empirical findings that apparently contradicted the established theory; and (3) predicts new findings that were not predicted by the established theory.

It is a tall order indeed to establish a better theory in this sense, but this has clearly been the ambition of many researchers who have made theoretical contributions to the study of the monetary valuation of non-market goods. Thus, following Dehez and Drèze (1982) WTP may be (but does not have to be) unrelated to the level of risk and may even be (but does not have to be) zero (if insurance coverage is generous enough). Following Hanemann (1991), and later Amiran and Hagen (2003), WTA may be much larger (indeed infinitely larger) than WTP, but it does not have to be like that (i.e. for ordinary market goods, WTA and WTP should be close). Then, according to Johansson (2002) the relationship between age and WTP could have "any shape", of course including the inverted U-shape proposed by earlier theorists. Finally, insensitivity to scope, as well as sensitivity to scope, are both perfectly rational and consistent with utility maximising behaviour (Amiran and Hagen 2010). It all depends on the assumptions made about characteristics of the utility function.

The sum of all these additions to the original hard core theory is to make the theory about willingness-to-pay for non-market goods immune to falsification by not ruling out certain findings as being inconsistent with the theoretical models. Rather than interpreting anomalous findings as evidence that the theory is false, new twists and turns have been added to the theory to make sense of the anomalous findings.

Neoclassic economists claim that no equally comprehensive and well-established alternative theory exists. They are right. Although behavioural economics is ascendant, it still only consists of bits and pieces that do not form the same kind of hard core as the basic postulates of neoclassic economic theory. So, from the Lakatosian point of view, there can be as many anomalies as there are stars in the sky; they amount to nothing as long as no superior theory has been developed to account for them.

This suggests that the methodology of scientific research programmes, as proposed by Lakatos, has great descriptive accuracy. Valuation research continues as if most of the anomalies did not exist. Methodological research intended to develop more refined methods for eliciting willingness-to-pay seems to have slowed down, perhaps because the need for it seems less pressing when some of the apparently anomalous findings may perhaps not be anomalous after all, as some recent theoretical contributions suggest.

Immunising a theory from falsification comes at a price, however. From a Popperian perspective, one would conclude that a theory is no longer scientific if it cannot be falsified. Such an argument is unlikely to impress those who adhere to a Lakatosian research programme very much; their main positive heuristic is to protect the hard core as best they can; a research programme is successful if it protects the hard core.

Yet, in a wider social context, valuation research loses its credibility by turning inwards and focusing on developing theory to account for all its findings. Methodological innovations are obviously still possible. Yet, the wide dispersion of estimates of the value of a statistical life suggests that the real problem is more fundamental: The preferences studies aim to elicit simply do not exist. If they did, and were as well-ordered as economic theory assumes, one would not find all the methodological artefacts that have been found (starting point bias, payment card bias, and so on). Stable, well-ordered preferences should not be so easily influenced by framing as many studies have found them to be. People can be framed into making choices that contradict the most elementary property of a preference, namely

that it is asymmetric. You cannot, without self-contradiction, at the same time both prefer A to B and B to A. But such a pattern is exactly what framing brings out.

If the very wide dispersion of estimates of the value of a statistical life found in empirical studies is to be expected from a theoretical point of view, what remains of the consistency argument that was made to justify this research in the first place? Most lay observers, who may not be familiar with the theoretical models in the willingness-to-pay literature, will probably conclude that the wide dispersion of estimates of the value of a statistical life simply shows that this line of research is nonsense and has not produced any meaningful results. That, indeed, is what one prominent economist who contributed to valuation research concluded. The next section summarises his reflections.

8.4 A prominent economist bids farewell to valuation research

Graham Loomes is one of the founders of behavioural economics and a pioneer in developing alternatives to the classic model of utility maximisation. Together with Robert Sugden he introduced "regret theory" in 1982 (Loomes and Sugden 1982), a theory that brings back psychological insights to economic theory by suggesting that people choose so as to minimise regret. To minimise regret is not necessarily the same as to maximise utility, but may, for example, involve sequences of choices that would be regarded as inconsistent within the conventional utility framework (such as being influenced by sunk costs as a way of trying to reduce regret).

Graham Loomes joined Michael Jones-Lee's group around 1990 and co-authored several valuation studies with him over the next 15 years. Some of these papers were quite critical of the contingent valuation method, but did not reject valuation research as a scientific research programme. However, in 2006, Loomes reached a different conclusion (Loomes 2006).

In a paper entitled "(How) Can we value health, safety and the environment?", Loomes asked if it was possible at all to obtain valid and reliable monetary valuations of safety and environmental goods. He remarked (page 715):

"The essence of the problem appears to be that although the model individual is assumed to have a complete set of values and preferences which she can access and process quite readily, the typical member of the population is not like that. Rather, most people have only somewhat imprecise and partly-formed values for such goods, so that when confronted with questions of the kind indicated above, they cannot simply pull the answer 'off the shelf'."

The questions Loomes referred were, for example, questions about the willingness-to-pay for small reductions in low levels of risk, such as 1 in 100,000. Loomes went on to discuss how the results of valuation studies are influenced by theoretically irrelevant factors, like starting point bias, bias due to the range of amounts displayed on payment cards, and the very large differences between willingness-to-accept and willingness-to-pay found in some studies.

Note that Loomes criticised valuation studies because their findings did not make sense from a theoretical point of view. The target of his criticism was therefore not the hard core of economic theory; quite to the contrary, the inconsistency of valuation studies with hard core economic theory was used to argue for rejecting the findings of valuation studies.

Loomes went on to criticise valuation studies because their findings were not sufficiently sensitive to factors that ought to make a difference according to economic theory. Examples included embedding effects and insensitivity to scope. Loomes went on to

question the belief that people have well thought-out values and preferences for everything; the only task for the economist is to find the best way of describing these preferences. He argued (page 719):

"For the economic model to stand a chance of working in practice, two components would seem to be essential. First, for any good or benefit, the utility an individual would experience would have to be accurately anticipated. Second, the individual would have to be able to translate such unbiased estimates into expressions of preference or value which would give the same comparisons between goods irrespective of the particular elicitation procedure employed to elicit them. … There is a body of psychological research which suggests that neither of these conditions are likely to hold to the extent required for the standard economic model to work."

When individuals take part in valuation studies, they are asked to make a decision, either in the form of stating an amount they are willing to pay, or by choosing between options that differ with respect to the good which is to be valued as well as other attributes. Standard utility theory applies to such decisions; it may thus be referred to as decision utility. To be able to successfully maximise utility, an individual must predict what the actual utility will be when making the decision. It is very difficult to believe that this can really be the case as far as valuing risks to life and health are concerned.

As far as fatality risk is concerned, the task in a valuation study refers to changes in an already low risk. The individual will basically not experience any difference in utility at all associated with a small change in the low risk of dying in a road accident. Indeed, it is only a slight exaggeration to say: What would happen if nobody died in traffic? Nobody would notice the difference. As long as the individual survives, no noticeable change in utility is associated with a reduced fatality risk in traffic. You essentially pay for an abstract and invisible good, producing no noticeable changes in your experienced utility. The only difference you may possibly notice, is that you have a little less money to spend on other things. Yet, even this is not the case in contingent valuation studies in which no real payment is made. It is an entirely hypothetical exercise which has no effect at all on experienced utility – except, perhaps, that some people transiently may feel a trifle happier after having expressed their support for a good cause like road safety.

Things are different as far as injury risk is concerned. An individual sustaining an injury will experience a change in utility. However, as long as the injury does not materialise, the argument made for fatality risk applies with full force: There is no noticeable change in utility. The individual will have paid for, so to speak, an invisible good whose consumption cannot be perceived by any of the senses.

Literature that will be reviewed in greater detail in Chapter 11 shows that individuals are notoriously bad at predicting the general quality of life (read: utility) they would experience in certain health states. Thus, healthy individuals believe the reductions in quality of life associated with many health states are greater than the reductions reported by those who have experienced the health states. Even wheelchair users report almost the same level of happiness, or subjective well-being – at least after a period of adaptation – as healthy individuals. To a healthy individual, this may seem strange and almost unbelievable. There are just so many nice things a person in a wheelchair cannot do. However, the ability of humans to adapt to adverse events and circumstances is a strong survival mechanism, probably deeply encoded in our genes. The person in the wheelchair will obviously be very well aware of his or her limitations; there is simply no way of forgetting about them. Yet, he or she will also understand – if perhaps only subconsciously – that life is less miserable is you focus on the things you can do, if you take up new activities that bring you pleasure, if you seek out new challenges – rather than if you engulf yourself in blame, guilt, bitterness, complaints, and so on.

Therefore, predicted utility (decision utility) is rarely the same as experienced utility. Decisions based on predicted utility may by a stroke of luck happen to be right on target, but in general, this will not be the case. Loomes asked (page 732):

"Is there scope for closing the gap between experienced utility and decision utility and delivering measures of value which meet some basic requirements of coherent deliberative judgment and which, although unlikely to ever be demonstrably optimal, can be defended as boundedly rational and as likely to advance the wellbeing of the population? There is clearly no consensus about the answer to this question …"

Chapter 11 will discuss some alternatives to the conventional approaches to the valuation of transport safety.

8.5 In what sense are preferences revealed?

In view of the pervasive violations of the axioms of rational choice found in stated choice experiments, it is pertinent to ask in what sense preferences are revealed in revealed preference studies, in particular in studies of compensating wage differentials. These studies give valid estimates of the value of life only if workers made free, informed and rational choices between occupations and if economists correctly modelled those choices. As the following discussion will show, these assumptions are quite restrictive and unlikely to be fulfilled in practice. The discussion will start with issues related to econometric modelling and continue with issues related to worker rationality and freedom of choice.

8.5.1 Issues of econometric modelling in studies of compensating wage differentials

Viscusi and Aldy (2003) discuss extensively a number of issues that arise when estimating the wage equations in studies of compensating wage differentials. The first issue concerns the quality of risk data. They note that the choices of both workers and firms are likely to be based on their perceptions of risk; rarely, if indeed ever, will a job applicant scrutinise statistics about the risks associated with various occupations. Firms, in particular small firms, may also lack statistical information about risks. This lack of information is replaced by subjective estimates of risk that may be right or wrong.

Very few studies have utilised data on worker perceptions of risk and no study has relied on data about how firms perceive risk. Virtually all studies of compensating wage differentials have relied primarily on official statistics about risks. These statistics are subject to incomplete reporting, at least with respect to non-fatal injuries, and may have fairly crude classifications by industry and occupation. Viscusi and Aldy (page 14) point out that before 1992, official US data sources were incomplete even with respect to fatal injuries.

Viscusi and Aldy note that: *"Failing to capture all of the determinants of a worker's wage in a hedonic wage equation may result in biased results if the unobserved variables are correlated with observed variables."* As noted in Chapter 4, no study of compensating wage differentials has controlled statistically for all variables that have been shown to be relevant. One may therefore not rule out omitted variable bias in any of these studies.

Dorman (1996) argues that all models of compensating wage differentials are likely to be flawed and that the evidence for a wage compensation for risk evaporates if industrial dummies (i.e. variables identifying industries, like food processing, car manufacturing, transport and so on) are included in the wage equation. Viscusi and Aldy (2003) dispute this and argue that Dorman fails to refer to studies that included industrial dummies and still found a wage premium for risk.

It is beyond the scope of this study to try to settle this disagreement; suffice it to note that it shows that not all economists are convinced about the existence of compensating wage differentials.

8.5.2   Rational choice and the structure of the environment

At a more fundamental level, one may ask to what extent workers really make free, informed and rational choices of occupation. Anderson (1993), for one, discusses this question. Modern labour markets are highly segmented and specialised. Segmentation means that there are "layers" in the labour market requiring different levels of education or skill. Unskilled work, or work that can be learnt with brief instruction, is at the bottom. Examples would be simple service occupations like doing room service in hotels, waiters in restaurants or serving the cash register in grocery stores. This type of work requires no formal education and an individual may change between these types of jobs without having to invest in an education or, in many cases, without having to move to a different city, commute a longer distance, or, indeed, earn much more or much less than in any other low-level service job. Switching between these types of work is simple and there are very often jobs on offer.

The next layer consists of jobs requiring some practical skills. Examples are bus or truck drivers, operators of machines in factories, most types of farm work, fisheries, mining, constructing power lines, road construction, and similar types of work. In many of these types of work, workers may get on-the-job training; in other cases the required skills must be acquired before applying for a job – an individual without a driving licence will not get a job as a truck driver.

One can imagine successive layers; each would require a larger dose of talent and investment in education. Scientists at the Nobel laureate level have a global labour market to choose from. At the lower levels, labour markets are mostly local and the jobs on offer therefore depend on local factors beyond individual control. Anderson (page 197) remarks that:

1. *"First, workers must be free to choose without duress. This requires that workers are mobile and see themselves as having a significant range of worthwhile alternatives to the choice they actually make.*
2. *Second, workers' choices must reflect deliberation upon full information about the risks they encounter. This requires not only that information be available to workers, but that they fulfil the internal conditions of autonomy necessary for them to make good use of this information.*
3. *Third, workers' choices must express their own valuations, not the valuations others make of their lives. …*
4. *Fourth, they must choose egoistically, with concern only for their own welfare, when they make the wage/risk tradeoff. …*
5. *Fifth, workers must care only about the relative magnitudes of risk/money tradeoffs in evaluating the acceptability of risks. This requires that they find the same risk/money tradeoffs in different social contexts equally acceptable."*

It is worth noting that no study of compensating wage differentials has reconstructed the actual choices made by workers. The data typically used in these studies do not show choices between alternative occupations; they merely show the mean wages of workers in occupations that differ in fatality rates, as well as some additional characteristics of workers and of their workplaces. No actual choices are studied; the set of opportunities from which workers made a choice is not reconstructed. The studies are, in other words, correlational only. The claim that the values of a statistical life estimated reflect worker preferences is

essentially only an article of faith, since no data on the options facing workers or on the rationality of their choices between these options are presented.

The title of this section is borrowed from a classic paper by Herbert Simon (1956). In that paper, Simon shows how characteristics of the environment determines the options available to an individual and generates cues that may favour the choice of some alternatives over others. It is not the case that the environment dictates individual choice by structuring the options in such a way that only one option remains for choice; rather, it limits the number of options and creates cues that aids an individual in choosing a good option, even if the individual is not optimising in the strict sense of economic theory. There are at least morphological analogies between the choice of work and the choices facing the imaginary organism discussed by Simon. In particular, the choice of work may have a large element of randomness, in the sense that the individual does not generate the options, may not know them all and may thus only make quite limited comparisons between the options.

8.5.3 Choices may not reveal preferences

Amartya Sen (1973) shows that choices do not always reveal preferences and do not always result in outcomes that are Pareto-optimal. This raises a fundamental objection to the theory of revealed preferences. Sen notes (page 60):

"If a person chose x when y was available, it would seem reasonable to argue that he did not really regard y to be better than x."

He then goes on to show, using the Prisoners' dilemma as an example that observed choices do not reveal preferences in this game. Robert Frank (2000) shows the validity of this argument using occupational choice as an example. He presents the case as follows:

"Each must choose between two jobs – a safe job that pays $300 per week and a risky job that pays $350 per week. The value of safety to each is $100 per week, and each evaluates relative income as follows: Having more income than his neighbour provides the equivalent of $100 per week of additional satisfaction; having less income than his neighbour means the equivalent of a $100 per week reduction in satisfaction; and having the same income as his neighbour means no change in the underlying level of satisfaction."

Frank models occupational choice for two workers, calling them Gary and Sherwin. The payoffs for each of them are shown in Table 8.1. In each cell, the payoff to Gary is shown in the lower left corner; the payoff to Sherwin is shown in the upper right corner.

Table 8.1: Choice between a safe job and an unsafe job as a Prisoners' dilemma. Based on Frank 2000

		Sherwin	
		Safe job $300/week	Unsafe job $350/week
Gary	Safe job $300/week	400 400	450 300
	Unsafe job $350/week	300 450	*350* *350*

If Gary chooses the safe job, the best choice for Sherwin is the unsafe job. He then gets paid $350 per week and enjoys the additional $100 satisfaction of knowing that he earns more ten Gary. The reasoning for Gary is exactly analogous. Therefore, the best choice for

both Gary and Sherwin is the unsafe job. Yet, both Gary and Sherwin would be better off by choosing the safe job.

This example is of course highly unrealistic in that, in practice, it is not the case that everybody ends up in an unsafe job. Yet, the motivational structure proposed by Frank is probably not unrealistic. Most people care about their social status, i.e. they care about whether they earn more than their friends and neighbours and compare themselves to others all the time. Indeed, this mechanism is so powerful that it is likely that it accounts for the so called Easterlin paradox in happiness research, to be reviewed more in detail in Chapter 11 of this report. The Easterlin paradox refers to the phenomenon that, while subjective well-being (happiness) is at any time in any country positively related to income (those who earn more report a higher level of subjective well-being), increasing incomes over time are not associated with an increase in the mean level of subjective well-being. People try to keep up with the Joneses all time, perpetually running on a hedonic treadmill leading nowhere.

Another important lesson from the example given by Frank is that workers may value safety even if their choices, prima facie, indicate otherwise. Safety often has the characteristics of a collective good or of a commons. It is also very often the result of a process of learning. From the point of view of econometric modelling, this means that safety is often endogenous, i.e. in predicting how workplace risks influence wages, one cannot assume that the workplace risks are not influenced by other independent variables in the model, or perhaps indeed by wages themselves if a system of bonus payments for a good safety record has been introduced. This introduces a profound circularity into any modelling exercise, in which independent and dependent variables may switch places in endless cycles.

By trying to obtain, ideally speaking, customised estimates of risk for each worker, the compensating wage differentials literature is treating safety as an individual good. In many occupations, it is not. If a railway introduces automatic train control, safety is improved for all train drivers, although their wages depend principally on seniority and their willingness to work inconvenient hours, and maybe not at all on the risk they are facing, which would, as a reasonable approximation, be the same for all train drivers. Hence, whenever safety is a collective good, (1) individual choices do not necessarily reveal preferences, and (2) the relationship between wages and risk is severed, because everybody faces more or less an identical level of risk, although their wages could differ for other reasons.

It is concluded that choices do not always reveal preferences and that the context in which choice of occupation is made is very different from the context in which choices influencing transport safety, in particular road safety, are made. However, the few studies of the monetary value of road safety that have combined the stated and revealed preference approaches will be reviewed in the next section.

8.6   Recent studies – the current state-of-the-art

If one were to summarise very briefly the methodological lessons that have been learnt in valuation research, it is that each of the three main methods that have been used – contingent valuation, stated choice, and revealed preferences – is insufficient on its own. Contingent valuation normally focuses on a single good; merely by doing so it makes that good appear more important than it would be if attention was not drawn to it. Hypothetical bias and insensitivity to scope are major problems in contingent valuation studies. Although various mechanisms have been proposed to counteract these tendencies,

they are not one hundred percent effective. It is therefore highly doubtful if contingent valuation studies produce unbiased estimates of the real willingness-to-pay for safety.

In theory, the stated choice approach ought to avoid some of the problems of the contingent valuation method. However, it has turned out that lexicographic and inconsistent choices are common, and that analysts can, as in the contingent valuation method, influence the results by choosing which attributes to include and the values of those attributes. A strong feature of the stated choice approach is that it allows testing the rationality of choices.

Finally, the revealed preferences approach, at least as implemented in the study of compensating wage differentials, is problematic for a number of reasons. In the first place, it is highly likely that workers self-select into high-risk occupations, and that these self-selected workers may not be representative of the population in general. More specifically, workers choosing high risk occupations are likely to be less risk averse than other people. In the second place, the risks of fatal and non-fatal injury are likely to be highly correlated and one may doubt if studies have been able to fully control for this. In the third place, when safety is a collective good, individual choices may not reveal preferences. In the fourth place, most of the factors that influence wages do not reflect individual preferences for the provision of safety, but quite different things like education, experience, union membership, and so on. It would be more informative to study choices in which safety is explicitly compared, although it may be just one of many attributes influencing choice.

Rationality is a relationship between beliefs, preferences and actions (Elvik 2016A). More specifically, an action is rational if it is believed to be the one (in a set of alternative actions) that realises preferred outcomes to the largest extent. To assign a monetary value to a non-market good is an act of choice. That choice can be given a normative (prescriptive) status if it is rational. If it is not rational, it has no normative significance. Nobody suggests that public policy and priorities should be based on inconsistent preferences, wishful thinking about risks, lack of understanding of changes in risk, or suboptimal choices. Any study of willingness-to-pay for a non-market good must therefore be designed so that it permits an assessment of rationality.

The extent to which people are rational is an empirical matter. When making familiar routine choices, people are probably quite rational. When making unfamiliar choices, people are more likely to pay attention to cues that are irrelevant, but that may influence their choices. Given the fact that all the three main approaches that have been used in valuation research have both strengths and weaknesses, one may perhaps get the best results by combining elements from them by means of a sort of methodological triangulation.

Thus, one may use an initial questionnaire to test the understanding of concepts like risk and changes in risk and to ask people about factors that influence safety-related choices, like the choice made when buying a new car or the choice between routes when driving. It is important to ask about choices that can be studied not only in a hypothetical setting, but also in real life. It is also important that these choices concern the particular risk one wants to value, i.e. that the choices are made in the relevant context.

Since framing effects are notorious in valuation studies, one should give as few cues as possible. One should see how far one gets by simply asking people what influences their choice, for example, when they buy a car. Do not give them a long list of attributes. People will just say that all of them are important – in particular safety of course. But if people do not mention safety spontaneously in answering an open question about, say, the five most important factors they think of when buying a car – when it does not come to their mind – there is simply no reason to believe that it counts that much. If people say something

different if safety is given as a cue, it is only because they want to be politically correct or please the interviewer.

A good and successful valuation study, therefore, is one that:

1. Combines elements from contingent valuation, stated choice and actual choice (revealed preferences).
2. The valuation should not be based on contingent valuation; rather a set of initial questions, asked with as few cues as possible, should probe the understanding of choices involving risk and factors influencing these choices.
3. The results of the initial questions should be used to develop a stated choice task. Respondents should then perform this task. One should test for lexicographic and inconsistent choices.
4. A real choice as similar as possible to the stated choice should then be studied. It is important that the stated and actual choices are highly similar.
5. The valuation elicited from the stated and actual choices, as well the weights of the factors influencing these choices, should be compared. The valuations should be close to each other.

A study is successful if it conforms to these guidelines and produce findings with convergent validity, i.e. stated and actual choices produce the same or nearly same valuations. If there is large proportion of lexicographic or inconsistent choices, or if the valuations are clearly different, the study is unsuccessful and the conclusion should be that there is not a sufficient degree of rationality to make valuation studies meaningful.

A few recent studies are briefly discussed below in view of these criteria.

8.6.1 Blaeij 2003

Arianne de Blaeij conducted a study valuing transport safety in the Netherlands. The study was her PhD dissertation. In-depth interviews with a small sample of 29 respondents were made to test understanding of risk concepts. Respondents were asked which of two risks they would prefer to reduce, one from 7 to 4 in 100,000, the other from 20 to 15 in 100,000. Individuals faced both risks, i.e. their total risk was the sum of the two risks presented. 73 percent chose to reduce the higher risk, which was regarded as the right answer.

Respondents were asked about willingness-to-pay for identical risk reductions from different initial levels. The risk reductions were in all cases 3 in 100,000. The initial levels of risk varied between 4 in 100,000 and 20 in 100,000. The hypothesis was that WTP would increase as initial risk increased. However, only 16 percent of respondents gave answers that were consistent with this hypothesis.

In a hypothetical route choice exercise, respondents were asked which of two reductions, from 28 to 24 fatalities per year, or from 18 to 14 fatalities per year, is the bigger. In both cases 4 fatalities are prevented. The reductions are therefore identical. However, only 38 percent of respondents rated the two reductions as identical. The results of the in-depth pilot survey were thus highly mixed.

The main study consisted of two parts. The first was a route choice (stated choice) task. Respondents were asked to choose between two routes. Both were toll roads and the toll varied between 2.50 and 12.50 guilders. Travel time varied between 50 minutes and 1 hour. Road safety varied between 12 and 36 fatalities per year. There was 1055 respondents in total.

284 respondents chose lexicographically (27 percent). 159 respondents chose inconsistently (15 percent). In total, 42 percent of respondents did not make rational choices according to

the rationality criteria of economic theory. Estimates of the value of a statistical life varied and were 5.3 million guilders for those who answered consistently (remember that lexicographic answers are consistent and therefore included) and 4.1 million for those who did not answer lexicographically.

The second part of the main study was a combined stated choice/contingent valuation study. In the stated choice part of the study, respondents were asked to choose between three car models that differed in safety and price. They were informed that the three car models were identical except for the differences in safety and price. Respondents were asked how many kilometres they drive each year. A fatality risk, stated as the expected number of fatalities per 100,000 (ranging between 3 in 100,000 and 12 in 100,000 for the least safe car model), was then presented and initial bids were presented to elicit WTP for choosing car B or C (car A was least safe, car B safer, car C the safest). This design is clearly prone to starting point bias.

Not surprisingly, the estimate of the value of a statistical life was 11.2 million guilders for the full sample – more than twice the value found in the route choice study. Following the choice task, respondents were asked an open-ended question about their maximum WTP for a safer car. The mean estimate of the value of a statistical life based on this question was 11 million guilders.

The different approaches to valuation did thus not produce the same values of a statistical life. The same sample of people took part in the whole study, and if they have the stable underlying preferences assumed by economic theory, there is no reason why one method for eliciting valuations should produce values that were, on the average, more than twice as high as the other method. It is altogether more reasonable to believe that these differences are artefacts of the methods, which in turn arise because the underlying preferences may not exist.

8.6.2 Andersson 2005

Henrik Andersson (2005) presents a study that compared a revealed preference study and a contingent valuation study. It is his PhD dissertation. The revealed preference study concerned the choice of car. It was based on data from 502 respondents in a contingent value survey. Respondents provided detailed data on the car they owned. By combining these data with other data sources, it was possible to develop a database containing data on the expected fatality rate of each car model, its price, and many other characteristics. It was then possible to estimate, the price/safety trade-off controlling for a host of other factors.

Estimates of the value of a statistical life based on this approach were all between 7.5 million SEK and 12.2 million SEK, with three estimates close to the upper estimate. The database was cross-sectional only, containing data on characteristics of the cars and their owners. The estimates were based on observed between-car variation in price, safety and many other characteristics. It was not possible to reconstruct the actual choices made when the car was purchased.

The second study was a contingent valuation study. This study was discussed already in section 7.1. The results were extremely noisy and estimates of the value of a statistical life, as presented by Andersson, were between 25.5 and 129.3 million SEK. These estimates are considerably higher than the estimate based on the car ownership study and show once again that the two methods for obtaining monetary valuations did not produce the same results.

8.6.3 Brabander 2006

Bram de Brabander's PhD dissertation (Brabander 2006) is a study of the valuation of road safety in Flanders, Belgium. It consisted of a contingent valuation study employing payment cards and a stated choice study involving route choice. The value of a statistical life based on the contingent valuation study was between 3.7 and 14.7 million Euro. There was insensitivity to scope. Mean WTP for the largest risk reduction (11 in 100,000) was 475 Euro, compared to 372 Euro for the smallest risk reduction (3 in 100,000). The ratio of the risk reductions is $11/3 = 3.67$. The ratio of mean WTP is only $475/372 = 1.28$. As a result, the value of a statistical life was highest for the smallest risk reduction.

Two payment cards were compared. On one card, the highest amount shown was > 3,000 Euro. On the other, the highest amount shown was > 525 Euro. There was a tendency, albeit not entirely consistent, for the estimated value of a statistical life to be lower based on the 525 Euro payment card than based on the 3,000 Euro payment card.

In the stated choice study, respondents were asked to make a route choice 12 times. The routes differed in terms of travel time, risk and cost. For all respondents, the value of a statistical life was estimated as 7.3 million Euro. However, when only those who did not answer lexicographically and who indicated that they were certain about their choices were included, the value of a statistical life was only 3.1 million Euro. Thus, the presence of lexicographic choices and/or preference uncertainty greatly inflated the value of a statistical life. The share of respondents who answered lexicographically is not stated.

The value of a statistical life was found to decline as the assumed length of the trip in the route choice task increased. In other words, a higher exposure to traffic risk was associated a lower valuation of safety, which is surprising. For those who did not answer lexicographically and had a trip time of 2 hours, the value of a statistical life was just 1.5 million Euro.

The study did not eliminate insensitivity to scope and lexicographic choices. Moreover, it did not treat these problems as serious enough to undermine the credibility of the results. Finally, it is noteworthy that estimates of the value of a statistical life within the same study varied by a factor of almost 10 (from 1.5 to 14.7 million Euro).

8.6.4 Svensson 2007

In his PhD dissertation, Mikael Svensson (Svensson 2007) asked "What is a life worth?", adding as a subtitle methodological issues in estimating the value of a statistical life. The main focus of the dissertation was methodological.

In the first study, it was found that insensitivity to scope is related to cognitive abilities. Those with higher cognitive abilities displayed a greater sensitivity to scope. This finding adds to the many explanations for insensitivity to scope: Directionally bounded utility functions, answers express attitudes only, answers reflect mental accounting, or answers reflect the fact that once an aspiration level has been reached, further gains have no value.

Whatever the reason, insensitivity to scope is a problem when trying to apply the results of valuation studies in practice, as discussed in Chapter 7. A different perspective on insensitivity to scope will be offered in Chapter 10.

The second study was a contingent valuation study conducted as a mailed questionnaire in the cities of Örebro and Karlstad. Alternative bids were stated and respondents stated their WTP by checking "yes" or "no" to the various bids. Respondents then indicated how certain they were of their answers, using a scale from 1 to 10, where 10 was absolutely certain. The mean value of a statistical life was 29.4 million SEK for all respondents in Örebro. For those who checked 10 on the certainty scale, the mean value of a statistical life

was 21.1 million SEK. The corresponding values in Karlstad were, respectively 50.0 million SEK (all) and 19.7 million SEK (10 on the certainty scale). Those who indicated that they were certain, had lower valuations than those who are more uncertain. Therefore, merely asking people to reflect on how certain they are about their answers lower valuations markedly.

A third study compared the valuation of safety as a public good to the valuation of safety as a private good and found that the value was lower for the public good.

The fourth study compared estimates of the value of a statistical life based on a contingent valuation study to the value revealed by the use of seat belts and bicycle helmets. The mean values of a statistical life were estimated to be 77 million SEK based on the contingent valuation study, 44.9 million SEK based on seat belt wearing, and 38.5 million SEK based on bicycle helmet wearing. As in the study by Andersson (2005) the revealed preference values were lower than the contingent valuation values, again suggesting that the latter are affected by hypothetical bias.

8.6.5   Veisten, Flügel and Elvik 2010

The fifth and final study to be reviewed here is the Norwegian valuation study that was published in 2010 (Veisten, Flügel and Elvik 2010). The study was based on a stated preference survey relying both on stated choice tasks and on contingent valuation. It also collected data intended for use in a revealed preference study, but these data were, regrettably, never analysed. The study was part of a larger research programme on valuation and was conducted in a sample who had already taken part in a value of travel time study. Values of safety were obtained for car drivers, cyclists and bus passengers. The stated choice valuation task for car drivers was implemented as a route choice that was pivoted on the last completed journey made by each respondent. The idea behind this design was to make the choice as realistic as possible, by referring to a real trip on a real road. The design made it necessary to provide safety data relevant for daily trips, most of them quite short, on different types of road. A table of estimates of the expected annual number of killed or seriously injured road users, adjusted for incomplete reporting in official statistics was developed.

It turned out that this table re-introduced the issue of very low numbers. Even though the numbers were not stated as risks, they were quite low for the shortest trips and had to be rounded up to avoid presenting zero as the expected number of killed or seriously injured road users. Many respondents indicated that they thought the numbers were too high, and they were right in thinking so.

Different scenarios were used for car drivers, cyclists and users of public transport. In addition to the stated choice study, a contingent valuation study was conducted. For this study, four regions of Norway were defined, each with about 1 million inhabitants. The questions asked about willingness-to-pay for reducing the number of killed or seriously injured road users in each of the regions. The study used the certainty scale as a tool for controlling for hypothetical bias.

The study has been reanalysed a number of times, most recently by Veisten (2016). Figure 8.4 shows 66 estimates of the value of a statistical life extracted from the study. Estimates have been sorted from the lowest to the highest. The lowest estimate is 15.8 million NOK, the highest is 362.7 million NOK. This is a huge range, but extracting all these estimates of the value of a statistical life gives an excellent opportunity to analyse the effects on the estimates of various methodological factors.

Such analyses have been done and the main results are presented below. However, before presenting these results, another Figure showing all 66 estimates is presented. Figure 8.5

shows the relationship between the estimates of the value of a statistical life and the variance of the estimates. There is a positive relationship. The higher the estimate of the value of a statistical life (VSL), the larger the variance associated with that estimate. Thus, if one prefers precise estimates, those are to be found among the lower of the 66 estimates that were developed.

In meta-analyses based on inverse-variance statistical weights, a subsample of the 22 best estimates was created. Based on the stated choice experiments, these were estimates omitting respondents who answered lexicographically and models including an alternative specific constant. With respect to the contingent valuation survey, the subsample included only respondents who indicated a high degree of certainty about their answers.
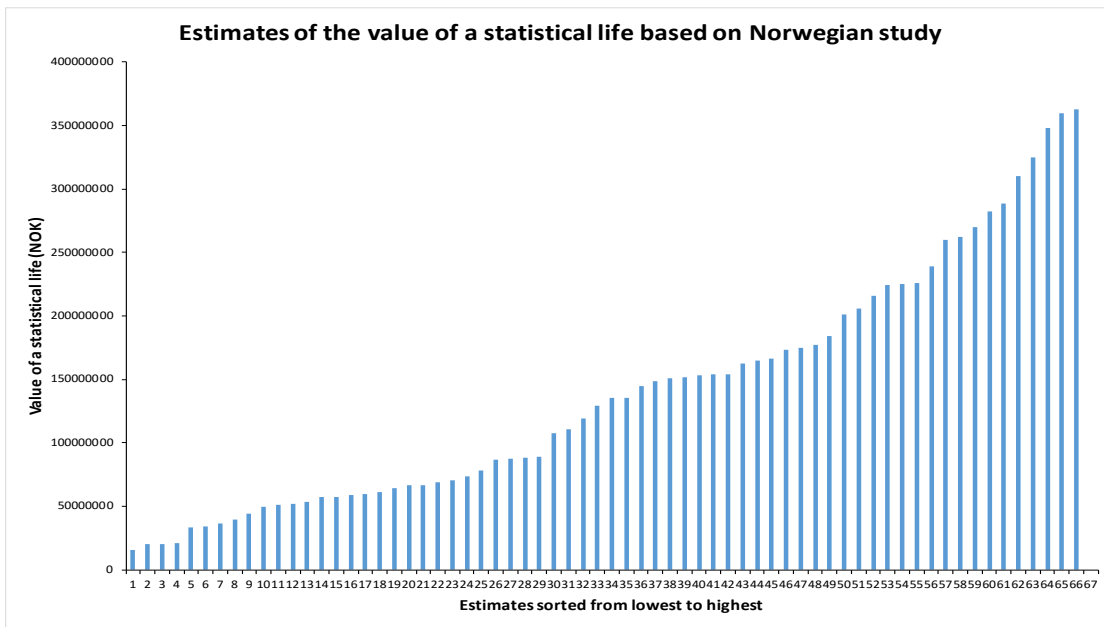

Figure 8.4: 66 estimates of the value of a statistical life extracted from the Norwegian valuation study
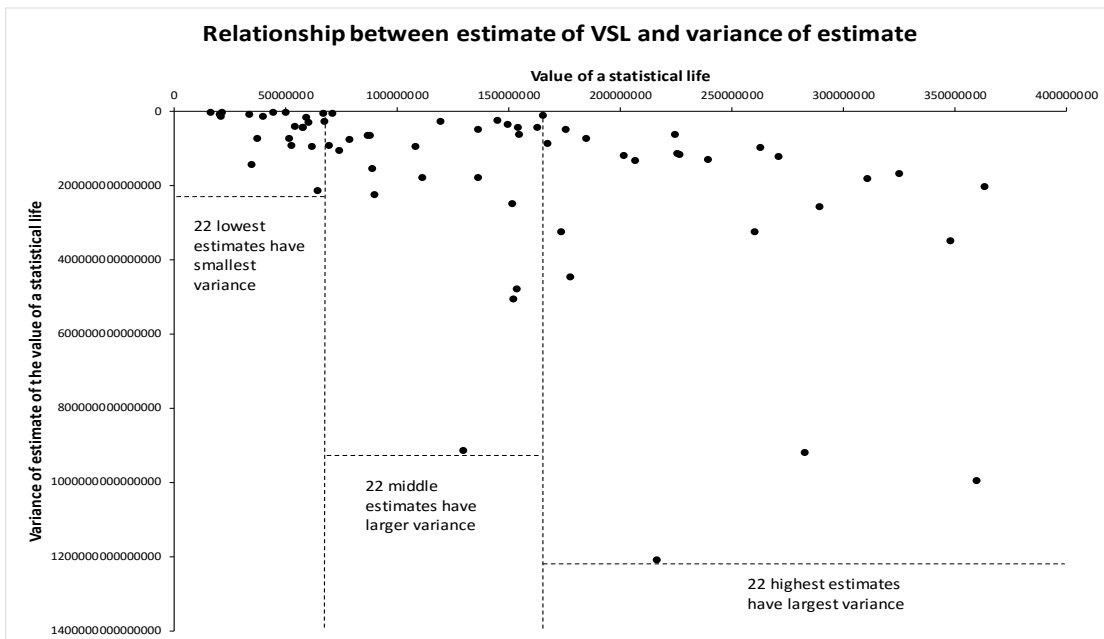

Figure 8.5: Relationship between estimates of VSL and their variance

The weighted mean value of a statistical life was 135.9 million NOK according to a random-effects model when all 66 estimates were included. A fixed-effects summary estimate based on 66 estimates was 47.3 million NOK. Based on the 22 presumably best estimates, the summary estimate of the value of a statistical life was 52.8 million NOK based on a random-effects model of meta-analysis and 30.6 million NOK based on a fixed-effects model of meta-analysis. The lower summary estimates emerging from the fixed-effects analyses reflect the tendency, shown in Figure 8.5, for the lowest estimates to have the smallest variance.

8.7   Concluding reflections

Until the last half of the 1990s, the contingent valuation approach was the only one that was used to value transport safety in Europe and New Zealand. The contingent valuation method was regarded with considerable scepticism in the United States, where most studies valuing safety were revealed preference studies, mostly relying on the compensating wage differentials approach. Thus, two quite distinct traditions developed in valuation studies.

By the late 1990s it was clear that many anomalies – hypothetical bias, insensitivity to scope, starting point bias, range bias, and others – were associated with the contingent valuation approach. In some studies, notably the study presented by Dubourg et al. (1997), the findings mainly reflected these anomalies. There was only noise, no signal. Even long-time proponents of the method realised that the anomalies were so massive that the method had to somehow be improved or abandoned.

Meanwhile, the history of compensating wage differentials studies in the United States was a history about the discovery of new sources of data on risk and new, previously not-controlled for, sources of confounding in econometric wage models. One may clearly read this history as a history of progress: researchers have found better sources of data and better methods for estimating wage equations. But it is progress at a price. Strictly speaking, it implies that all older studies, i.e. those made before the new sources of data were used or the new techniques for econometric estimation were introduced, should be rejected because they are not up to the current state-of-the-art. Research really never becomes truly cumulative if one ever so often must conclude that nearly all older studies are substandard and must be rejected. That way, one is at any point in time only left with the most recent few studies that can be trusted – so long as that lasts.

Besides, as will be shown in the next chapter, studies of compensating wage differentials have not converged on a single estimate of the value of a statistical life. On the contrary, the dispersion in values has increased over time, although it is much smaller than the dispersion in estimates based on stated preference studies. At any rate, compensating wage differentials studies were strongly criticised by Dorman (1996). Today it is probably fair to say that there are two camps among economists: Those who continue to believe in compensating wage differentials studies and those who, like Dorman, have rejected the approach.

The widespread acceptance of the anomalies in the contingent valuation method around 2000 lead to a number of innovations, like use of certainty scales and truth-telling mechanisms. Did these innovations help? They did, but only to some extent. Insensitivity to scope has proved to be a very difficult anomaly to avoid. Stated choice emerged as an alternative, or supplement, to the contingent valuation method, but it was soon found that it had its own set of anomalies. Some of these anomalies, in particular lexicographic and inconsistent choices, have deeper implications than some of the anomalies in the

contingent valuation method and may be interpreted as indicating that well-ordered preferences underlying rational choices do not exist.

One might still claim that the choice of occupation is based on rational trade-offs, although no direct evidence of this exists. Yet, even if the choice of occupation is, if not perfectly rational, then at least the result of a choice containing a large element of rationality, there are reasons to doubt both that estimates of the value of a statistical life are valid and that they can be transferred to transport safety.

As for the validity of the estimates based on compensating wage differentials, there are three unresolved issues. The first concerns the data on risk and the way these data are used in compensating wage differentials studies. Estimates of risk appear to be very crude. Ideally speaking, one should estimate the risk to each worker, relying, on statistical modelling and perhaps the empirical Bayes method (Hauer 1997, 2015). In road safety studies, methods have been developed for estimating the risk of individual study units, whether they are drivers (Elvik 2013), junctions (Elvik 2014) or road sections (Høye 2014). Compared to these methods, whose accuracy is well-established, the approach taken to the estimation of risk in compensating wage studies seems primitive. It is very likely that there are errors of unknown direction and magnitude in the risk data used in compensating wage studies.

The second reason concerns the potential endogeneity of risk. Few studies have tried to control for this, but it is quite likely that occupational risks are often endogenous, in that both each worker improves his or her own safety over time as a result of learning, and because – in particular in small firms – any serious accident is a shock that may profoundly change safety practices in the firm.

The third reason concerns the treatment of industrial dummies in estimating wage equations. Dorman (1996) regards failure to include such dummies as a fatal flaw; Vicusi and Aldy (2003), on the other hand, dismiss the point made by Dorman as essentially erroneous. The matter can only be resolved by comparing models including or excluding the industrial dummies to see if it makes a difference.

With respect to the transferability to transport safety of estimates of the value of a statistical life based on studies of compensating wage differentials, there are also a number of issues. First, wage compensation should be interpreted as a willingness-to-accept value, whereas a willingness-to-pay value is most often sought for transport safety programmes. Second, workers are likely to self-select into high risk occupations and their valuations of safety may not be representative of the general population. Third, risk levels are considerably higher than current risk levels in road transport in the safest countries of the world.

Based on this discussion, it is concluded that estimates of the value of a statistical life based on compensating wage differentials are not applicable to transport safety.

9 Can meta-analysis create order in chaos?

Once research designed to estimate the monetary value of improving transport safety had been successfully launched as a scientific research programme, the results of empirical studies started to accumulate, slowly at first, then at an accelerating rate. It did not take long before it became clear that the results of empirical studies varied considerably. It also did not take long before the first summaries of the results of empirical studies were made. The first summaries simply listed the results of each study and commented on the results. As the number of empirical studies kept growing, formal syntheses of results by means of meta-analysis was applied. By now, more than ten meta-analyses the monetary valuation of reduced risk of death have been reported. These analyses have become more sophisticated over time. An increasingly important objective of the analyses has been to explain the huge variation in estimates of the value of a statistical life. One may thus view the systematic literature review and meta-analyses as attempts to create order in the bewildering dispersion of estimates of the value of a statistical life.

This chapter will critically discuss the systematic literature reviews and meta-analyses of the literature on the value of a statistical life. First, essential elements of meta-analysis will be presented. Then, the meta-analyses that have been published will be critically reviewed according to current methodological guidelines for meta-analyses. It is recognised that methods of meta-analysis have developed over time. It may perhaps seem unfair to compare meta-analyses made many years ago to the current state-of-the-art. However, doing so creates a basis for giving advice on how to improve meta-analyses and assess the prospects for future meta-analyses to more successfully explain the enormous variation in estimates of the value of a statistical life than these analyses have so far been able to.

9.1 Elements of meta-analysis

A meta-analysis is a statistical analysis of results of empirical research dealing with a certain topic for the purpose of estimating one or more weighted mean results and explain why the results of different studies vary. There are many techniques of analysis in meta-analysis, but the most widely applied is the inverse-variance technique (Elvik 2016B). Each results is assigned a statistical weight which is inversely proportional to its sampling variance. These statistical weights minimise the variance of the weighted summary result.

9.1.1 Systematic literature survey

A meta-analysis should always be embedded in a systematic literature survey. A systematic literature survey is a survey guided by an explicit protocol intended to make it as comprehensive and replicable as possible. This includes the following elements:

1. A systematic search for relevant literature is made in relevant bibliographic databases. The databases that were searched and the search terms used should be stated explicitly.

2. A key objective of most systematic literature surveys is to be comprehensive, i.e. identify and include all studies that have been made about a topic, both published and unpublished studies.
3. Criteria for selecting the studies included in the review should be stated explicitly.
4. Each study included in a systematic literature review should be reviewed according to a standard protocol. To assist this, it is useful to code key characteristics of each study.

Once a systematic literature survey has been made, a list of studies coded according to their key characteristics exists. This list of studies may then be considered for inclusion in a meta-analysis. If the meta-analysis relies on the inverse-variance method, two items of information must be known to include a study in the meta-analysis:

1. One or more estimates of the result of interest, e.g. one or more estimates of the value of a statistical life.
2. The standard error of each result of interest, i.e. the standard error of each estimate of the value of a statistical life.

The standard error of an estimate will not always be reported. It may still be possible to perform some kind of meta-analysis, although any statistical technique other than the inverse-variance method will be less efficient, i.e. associated with a larger variance of the pooled estimate. There are three possibilities:

1. Assign the same weight to all estimates, i.e. compute a simple arithmetic mean. This will be statistically inefficient, i.e. have a larger standard error than a weighted mean.
2. Assign weights based on sample size. Sample size may be stated in a study, even if standard errors for the value of a statistical life are not.
3. Assign weights that are inversely proportional to residual variance (Elvik 2013C). i.e. assign low weight to studies whose estimates are far away from the arithmetic mean. An analysis based on residual variance would be made in two stages: first a simple arithmetic mean would be estimated, then an adjusted mean based on residual variance weights.

These procedures are not ideal and it is not well known how closely they approximate an inverse-variance meta-analysis.

In order to identify sources of variation in estimates of the value of a statistical life, it is useful to code several characteristics of each study, including:

1. Publication year
2. Country of origin
3. General approach (revealed or stated preferences)
4. Description of elicitation method
5. Tests made of the validity of the assumption of rationality
6. Confounders controlled for
7. Sensitivity analyses reported

A full meta-analysis has three main stages: Exploratory analysis, main analysis and sensitivity analysis. The purpose of exploratory analysis is to decide whether it makes sense to proceed to a main analysis. This depends on how "well-behaved" the distribution of estimates in primary studies are. One can think of a normal distribution as a model of a well-behaved distribution. It has a clearly defined peak at the mean, it is single modal, it is symmetrical, and it is not influenced by outlying data points.

9.1.2 Exploratory analysis

The key element of exploratory analysis is therefore to examine characteristics of the distribution of estimates in primary studies. How widely dispersed are estimates? Do most estimates cluster close to the mean? Does the distribution have a single peak or is it bimodal? Is the distribution symmetrical around the mean? Is it influenced by outlying data points? A widely used tool for investigating these questions in meta-analysis is the funnel plot. A funnel plot can only be developed if the standard error (or some other measure of precision) of each estimate is known. Yet, even if standard errors are not known, one may gain an impression of the shape of the distribution by using, for example, a stem-and-leaf plot (Elvik and Ramjerdi 2014). The stem, plotted vertically, show the first decimal of estimates, the leaves, plotted horizontally, shows the second decimal. Visual inspection of the plot gives an impression of the distribution of estimates, both its width and symmetry.

It has already been noted many times that estimates of the value of a statistical life are very widely dispersed. The distribution of these estimates may therefore not be well-behaved in the usual sense of that term in meta-analysis. This does not necessarily mean that a meta-analysis will not make sense. Rather than summarising studies by means of a weighted mean, one may look for patterns that may not be apparent in a funnel plot or stem-and-leaf plot. It could, for example, be the case that the value of a statistical life depends strongly on a variable which varies so much that it can produce large variation in the estimated value of a statistical life. If this variable can be identified and its relationship to the value of a statistical life modelled statistically, it may still be possible to develop a summary of the results of primary studies in terms of a function relating the value of a statistical life to its principal determinant.

One potential source of bias in meta-analysis which can be assessed by means of a funnel plot is publication bias. If a funnel plot cannot be developed, it is difficult to assess the potential presence of publication bias. A stem-and-leaf plot may give some indication, but does not lend itself to the formal analyses, such as trim-and-fill (Duval and Tweedie 2000A, 2000B, Duval 2005), that can be applied to a funnel plot.

A full exploratory analysis is therefore difficult if estimates cannot be represented by means of a funnel plot. An exploratory meta-analysis should always report:

1. How relevant studies were identified.
2. If any studies were omitted from a review and/or meta-analysis and the reason(s) for omitting them.
3. The information recorded for each study, including whether standard errors were reported or could be estimated.
4. If a funnel plot could be developed, and, if so, how it was analysed.
5. If a funnel plot could not be developed, if other tools for exploratory analysis were used.
6. Key characteristics of the distribution of estimates of the value of a statistical life, such as mean, median, modality, range and skewness.
7. Any indication of the presence of publication bias (see also sensitivity analysis below).
8. Whether estimates are best summarised in terms of measures of central tendency (mean, median, mode) or in terms of a function or set of functions describing systematic variation in estimates.

9.1.3 Main analysis

The final point on this list refers to the transition from an exploratory analysis to a main analysis. There are two approaches to main analysis. One approach is to identify subgroups within which estimates of the value of a statistical life are more homogeneous (vary less) than in the entire data set. Subgroup analysis is probably likely to be most informative with respect to methodological aspects of studies, as these can often be defined by forming groups of studies. The other approach is meta-regression analysis. Meta-regression is a weighted regression analysis, using the weights assigned to each estimate. It is well suited for continuous variables that may influence valuation, such as income or age.

The simplest form of meta-analysis consists of developing one or more summary estimates of the value of a statistical life, based on the studies included. However, as estimates of the value of a statistical life are known to vary enormously, a meta-analysis should always contain an analysis intended to identify sources of this variation. These sources are of two types: methodological and substantive. Methodological sources are aspects of the method used in the studies. If aspects of the method are found to be the predominant source of variation in estimates of the value of a statistical life, that suggests either that: (A) Not all studies have used the methods that are regarded as best, or (B) There is no agreement about the methods that are best suited to elicit valuations, or (C) The preferences valuation studies seek to elicit do not exist, or are at least less well-ordered than assumed in the hard core of the research programme (rational utility maximisation; see chapter 4).

Substantive sources are all variables that can be expected to be related to valuations according to the hypotheses forming the protective belt of valuation studies as a scientific research programme (see chapter 5). These include (but are not limited to): The level of risk, the size of the change in risk, the direction of the change in risk (increase or reduction), the severity of the outcome (fatal or non-fatal), income, age, the presence of competing risks, insurance coverage, experience of a life-threatening event, and so on. The more of these factors a meta-analysis tests for, the better becomes the basis for concluding whether variation in estimates of the value of a statistical life are mainly attributable to methodological factors or mainly attributable to substantive factors.

Based on this discussion, the key aspects of the main analysis stage of a meta-analysis are:

1. The analysis should report both measures of central tendency and measures of dispersion to indicate how widely spread estimates of the value of a statistical life are.
2. The analysis should contain at least one variable describing the method used by each study to obtain monetary valuations.
3. An attempt should made to formally rate studies by quality.
4. The analysis should test for at least one variable identifying a source of systematic variation in monetary valuations according to economic theory.
5. The results of analysis should indicate which sources of variation in estimates of the value of a statistical life are the most important.

9.1.4 Sensitivity analysis

The final stage of meta-analysis is sensitivity analysis. In any analysis based on a fairly large number of estimates of the value of a statistical life (e.g. more than 30), a sensitivity analysis should be feasible with respect to at least: outlying data points, publication bias, quality of primary studies and various aspects of the meta-analysis, in particular the meta-regression analysis.

In meta-analysis, an outlying data point is defined as any data point that exerts a decisive influence on the summary estimate. If, by omitting an estimate, the weighted mean value of

a statistical life changes significantly, then that estimate is outlying. Testing for outlying data points in this sense should only be done once, by successively omitting one data points and re-estimating the summary mean based on the remaining N – 1 data points. If the analysis is repeated after one or more outlying data points have been omitted, one or more of the remaining data points, not classified as outlying in the first analysis, risk being classified as outlying, simply because by omitting outlying data points, both the mean and the spread of the distribution are changed.

Publication bias is always a concern in meta-analysis. One can easily imagine that it would occur in studies of the monetary valuation of transport safety. If a positive valuation is not found, researchers may not trust the finding and choose not to publish it because they are unable to explain it convincingly. Economic theory would, at least in what one might refer to as "normal conditions", seem to rule out a zero valuation or a negative valuation. Still, in certain conditions, perhaps unlikely to occur in practice, one might expect to find a valuation at least close to zero. This would apply to, for example, very poor individuals valuing changes in very low levels of risk. There may, however, be publication bias even if all estimated values are clearly positive. An example of this will be examined in detail later in this chapter.

The quality of primary studies is also always a concern in meta-analysis. As the discussion in the previous chapters has shown, this is certainly the case with respect to valuation studies. Indeed, attempts have been made to legislate minimum standards for the quality of such studies (Arrow et al. 1993), but many studies had been made before these standards were developed. Moreover, the standards applied only to contingent valuation studies, not to other methods used in valuation studies.

One very important aspect of the quality of primary studies is the extent to which they have tested for rationality. Unless it can be shown that the rationality assumptions forming the hard core of the research programme are supported in practice, there is no reason to believe that valuations reflect rational trade-offs.

As far as the meta-analysis itself is concerned, it is very important that it succeeds in identifying variables that may account for the variation in the estimates of the value of a statistical life. In view of the enormous dispersion of estimates of the value of a statistical life, no meta-analysis could possibly have a more important objective than to try to account for this diversity. If a meta-analysis does not succeed in doing this, or if it indicates that the principal sources of the diversity are various methodological aspects of the studies, the only possible conclusion is that the values are artefacts of the methodology only and do not reflect any stable, underlying preferences. Thus, a checklist for the sensitivity analysis of a meta-analysis is:

1. A test for the presence of outlying data points should be made and the results reported.
2. A test for the presence of publication bias should be made and a summary estimate adjusted for publication bias be developed, if possible.
3. The relationship between study quality and estimates of the value of a statistical life should be investigated.
4. The sensitivity of the summary estimate of the value of a statistical life to methodological and substantive variables should be tested.

With these criteria as a basis, studies that have systematically reviewed estimates of the value of a statistical life will be reviewed in chronological order.

9.2   A survey of systematic reviews and meta-analyses

9.2.1   Blomquist 1982

Glenn Blomquist (1982) reviewed 12 studies estimating the value of a statistical life. He presented 15 estimates based on these studies. The estimates of the value of a statistical life ranged from 57,000 US dollars (1980-prices) to 10,120,000 US dollars. His review is the first published review of studies designed to estimate the value of a statistical life.

Blomquist did not report how the studies were obtained. He probably included all studies he knew about. Some of the studies were unpublished. Blomquist did not perform a meta-analysis, but briefly commented each study. In addition to presenting the value of a statistical life estimated in each study, he also presented the size of the risk reduction each estimate applied to. The size of the risk reduction was only stated to the nearest order of magnitude and four orders of magnitude were represented among the studies: 10^{-3}, 10^{-4}, 10^{-5} and 10^{-6} (i.e. risk reductions of 1 per 1000 to 1 per 1,000,000).

In all the following presentations of meta-analyses of studies of the value of a statistical life, the following summary statistics will be presented:

1. Arithmetic mean: The mean of the values of statistical life
2. Median: The median value of a statistical life
3. The standard deviation of the distribution of values of a statistical life
4. Maximum: The highest value of a statistical life
5. Minimum: The lowest value of a statistical life
6. The dispersion index

The dispersion index is defined as follows:

$$\text{Dispersion index} = \frac{(Maximum\ VSL - minimum\ VSL)}{Median\ VSL}$$

VSL is an abbreviation for the value of a statistical life. If all estimates were identical, the numerator would be zero and the dispersion index would have the value of zero. The numerator was defined as the difference between the highest and lowest estimate, rather than the ratio, because – as will be shown later – some estimates of VSL are zero. Finally, nearly all reviews of the VSL literature show that estimates are highly skewed, and that mean estimates can be greatly influenced by a few extremely high estimates. Therefore, the median was regarded as more representative than the mean. For the estimates presented by Blomquist, these statistics are (US dollars 1980):

Mean	1,704,000
Median	428,000
Standard deviation	2,670,000
Maximum	10,120,000
Minimum	57,000
Dispersion index	23.51

There was a clear negative relationship between the size of the risk reduction and the value of a statistical life. This is shown in Figure 9.1. An inverse function fits the data quite well, with $R^2 = 0.608$.
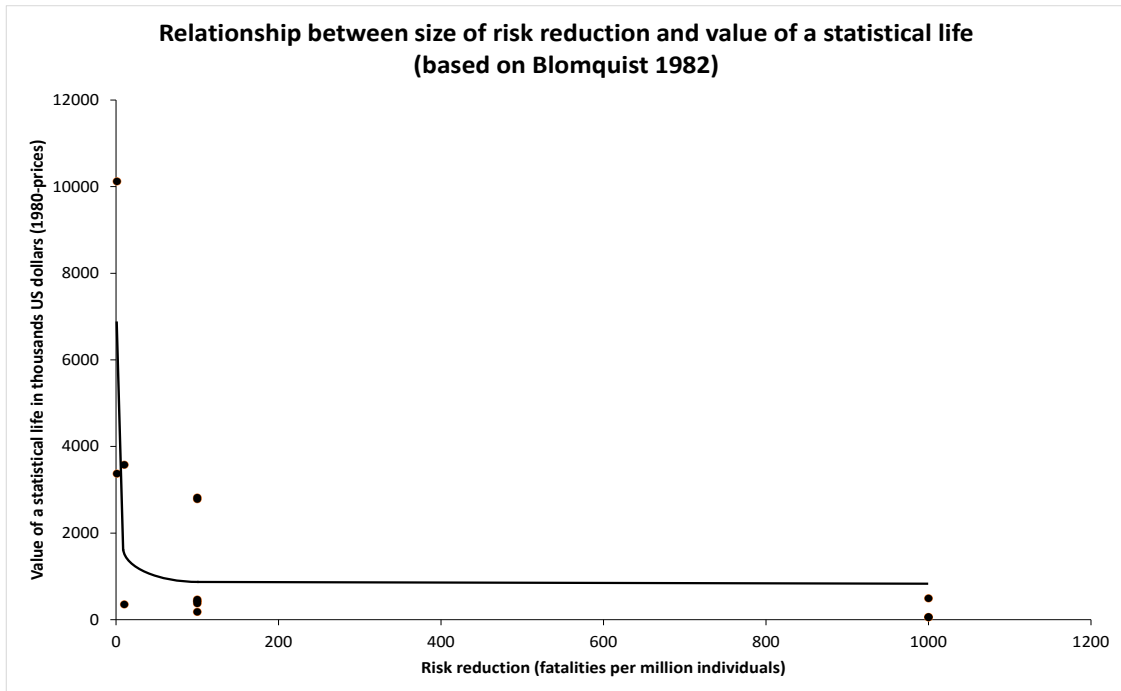
Figure 9.1: Relationship between the size of the risk reduction and the value of a statistical life. Based on Blomquist 1982

Given this highly negative relationship, one may ask what the relationship between willingness-to-pay and the size of the risk reduction is. Remember that (Chapter 1):

$$\text{VSL} = \frac{WTP}{\Delta R}$$

Here ΔR denotes the change in risk, in this case reductions of between 1/1000 and 1/1,000,000. It follows that when VSL and ΔR are known, WTP can be estimated as:

$$\text{WTP} = \text{VSL} \cdot \Delta R$$

To estimate the relationship between WTP and the size of the risk reduction, averages were taken of estimates of WTP that referred to identical risk reductions. There were 2 estimates for 10^{-3}, 2 estimates for 10^{-4}, 7 estimates for 10^{-5} and 4 estimates for 10^{-6}. The relationship between WTP and ΔR is shown in Figure 9.2.

It is seen that willingness-to-pay increases as the size of the risk reduction increases. This is an expected pattern. A logarithmic function describes the relationship very well.

A particularly interesting aspect of Blomquist's review, is that he discusses how misperception of risk may influence estimates of the value of a statistical life. It is reasonable to think that how much people are willing to pay to reduce a certain risk, depends on how large they think the risk is. There is, for example, strong demand for air travel safety, because many people overestimate the risk involved in flying. On the other hand, the risk involved in car driving may be underestimated and the willingness-to-pay for increased safety correspondingly reduced.

To account for misperceptions of risk, Blomquist relied on a seminal study by Lichtenstein et al. (1978). In that study, students and staff at the University of Oregon were informed about the number of people dying each year of a specific cause and asked to state the number of people they believed died of 40 other causes.
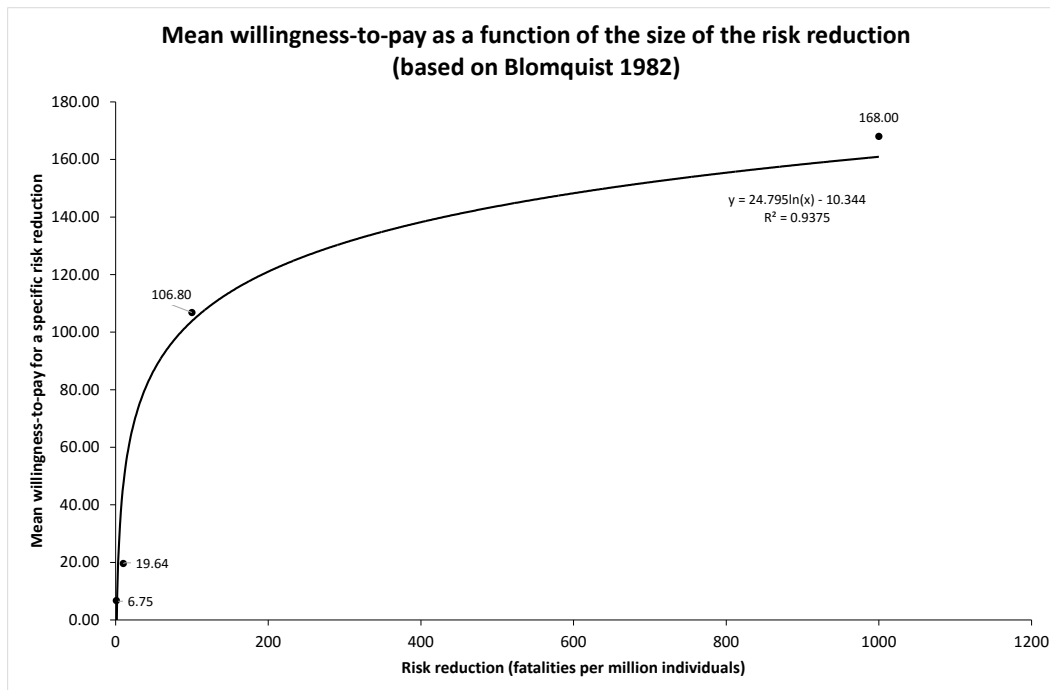
Figure 9.2: Relationship between willingness-to-pay and size of risk reduction. Based on Blomquist 1982

One group was told that 1,000 people die of electrocutions (i.e, accidental exposure to high-voltage electricity) in the United States each year. They then stated the number of people they believed died of 40 other causes each year. Rare causes of death were found to be overestimated (believed to cause more deaths than they actually do). Common causes of death, in particular common diseases, were considerably underestimated.

Blomquist estimated adjusted values of a statistical life intended to account for misperceptions of risk. As an example, in a study he made himself (Blomquist 1979), the value of a statistical life revealed by wearing seat belts was estimated as 466,000 US dollars (1980). According to the study of Lichtenstein et al. (1978), the geometric mean estimate of the number of road accident fatalities in the United States was 33,884. The actual number was 55,350. In other words, the real risk was underestimated. To adjust for this, Blomquist multiplies his estimate of VSL (466,000) by the ratio of the actual to the perceived number of traffic fatalities (55,350/33,884 = 1.634) to get an adjusted estimate of 761,000.

This adjustment is consistent with the assumption that, all else equal, an underestimation of risk will inflate the value of a statistical life, since the denominator of the marginal rate of substation, ΔR, becomes smaller. On the other hand, if a risk is erroneously believed to be smaller than it really is, the numerator of the marginal rate of substitution may also be affected, i.e. individuals will want to pay less to reduce the risk than if they knew it correctly. A lower WTP will, all else equal, be associated with a lower VSL. To adjust for misperceptions of risk, one needs to estimate how much individuals would be willing to pay to reduce the risk as they perceive it. To this end, one may apply the function fitted in Figure 9.2. Blomquist (1982) stated the risk reduction for seat belt wearing as 1/10,000. If the results of Lichtenstein et al. (1978) are used, the perceived risk reduction is only about 0.6/10,000. Applying the logarithmic function in Figure 9.2, WTP for such a risk reduction is about 92 US dollars.

Blomquist relied only on the study of Lichtenstein et al. (1978). He was unable to obtain adjusted estimates of VSL for air travel, heart attack prevention and nuclear power. For heart attack prevention, Lichtenstein et al. did present results suggesting that the number of deaths caused by heart disease (which admittedly is a broader category than heart attacks) is

144

considerably underestimated (geometric mean of stated number = 21,503; actual number = 738,000). For air travel and nuclear power, one can use the results presented in a follow-up study by Slovic et al. (1979), not quoted by Blomquist. Slovic et al. (1979) find that the risks of air travel are overestimated. Their results for nuclear power are somewhat difficult to interpret. In the study quoted by Blomquist (1982), three levels of risk reduction associated with nuclear power are stated: 10^{-3}, 10^{-4} and 10^{-5}. Presumably, these refer to events mapped onto an FN-curve (a diagram in which the frequency of events (F) is shown vertically and the number of fatalities in each event (N) is shown horizontally; a curve fitted to such a diagram with a log scale for both axes will usually slope down to the right, i.e. events with many fatalities (high N) are expected to occur less frequently (low F) than events with few fatalities). The most frequent event (10^{-3}) is associated with a low number of fatalities, and the more infrequent events associated with a higher number of fatalities. Based on this interpretation, the results of Slovic et al. (1979) suggest that the risk of the frequent event is underestimated and the risks of the less frequent events overestimated.

When WTP is estimated by relying on the function in Figure 9.2 and the perceived levels of risk, the following summary statistics are obtained:

Mean	2,530,000
Median	1,221,000
Standard deviation	2,798,000
Maximum	7,637,000
Minimum	54,000
Dispersion index	6.21

Both mean VSL and median VSL increase. The difference between mean VSL and median VSL is reduced, suggesting that the distribution is less skewed. The dispersion index is reduced from 23.51 to 6.21.

In other words, misperceptions of risk appear to explain some of the wide dispersion in estimates of VSL. This analysis is very simple and should be regarded only as an example. It nevertheless shows that misperceptions of risk may influence estimates of VSL and that failure to try to account for this is a potential source of systematic error of an unknown magnitude in estimates of VSL.

9.2.2 Jones-Lee 1989

Michael Jones-Lee (1989; reprinted in Layard and Glaister (eds) 1994) reviewed 21 estimates of the value of a statistical life based on revealed preference studies and 8 estimates of the value of a statistical life based on stated preference studies employing the contingent valuation method. He extracted only a single estimate of VSL from each study. For both groups, he estimated the mean and median VSL.

The studies included in this review appear to be those the author knew about. At the time of the review, it was probably fairly complete with respect to published studies. Unlike Blomquist (1982), however, Jones-Lee did not include any unpublished studies. In principle, publication bias could be a source of error in the review.

Jones-Lee did not state levels of risk or the change in risk. It is therefore not possible to use his review to estimate how WTP varies as a function of the size of the change in risk.

The summary statistics for the studies reviewed by Jones-Lee is as follows (Pounds 1987-prices):

Mean	2,097,000
Median	600,000
Standard deviation	2,887,000
Maximum	11,700,000
Minimum	50,000
Dispersion index	19.42

The mean value of VSL was higher in contingent valuation studies (3,090,000) than in revealed preference studies (1,719,000). Blomquist (1982) found the same.

For the contingent valuation studies, Jones-Lee stated sample size. This makes it possible to assess whether there is a "small study effect", which is an indicator of the possible presence of publication bias. Figure 9.3 shows the relationship between estimated VSL and sample size.

When estimates are weighted by sample size, the mean estimates VSL is 1,487,000. The simple mean (not weighted) is 3,090,000. This indicates a considerable small study bias. It is surprising that Jones-Lee did not note this and that he did not estimate a mean VSL weighted by sample size.
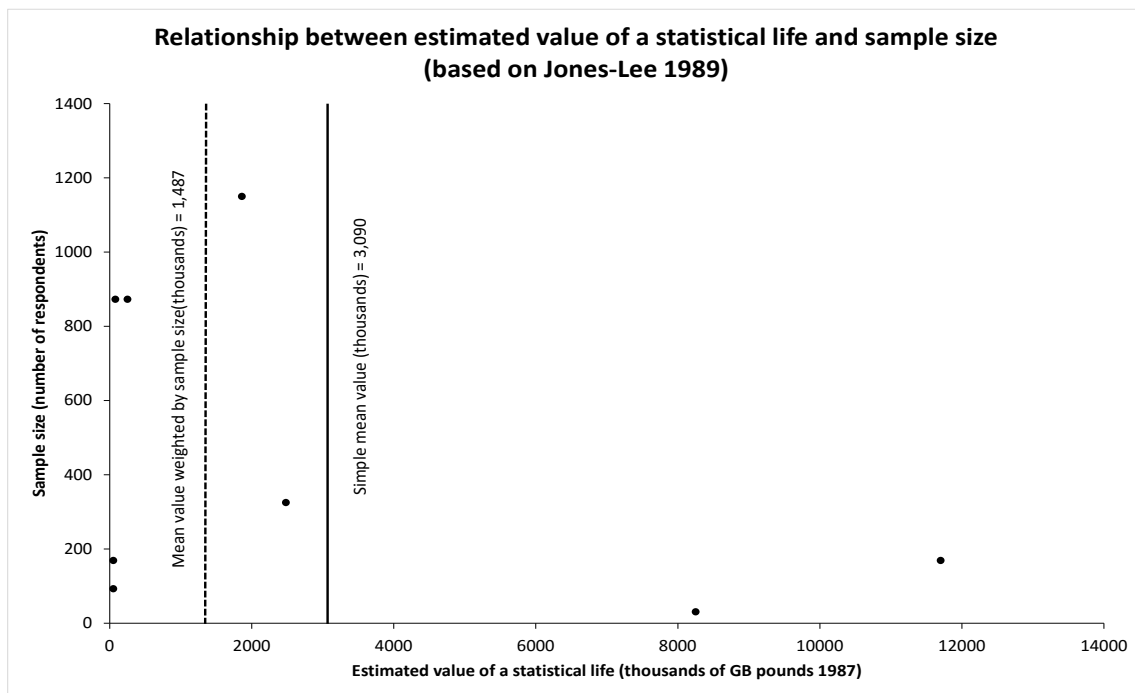

Figure 9.3: Relationship between estimated value of a statistical life and sample size. Based on Jones-Lee 1989

A formal assessment of publication bias has been made, relying on the trim-and-fill technique (Duval and Tweedie 2000A, 2000B, Duval 2005). Sample size has been used as an indicator of standard error, meaning that estimates based on large samples are more precise than those based on small samples. The analyses deleted four data points, resulting in a trimmed mean estimate of 773,000. This indicates the presence of publication bias.

Jones-Lee argued that some of the estimates were more reliable than others. When only the studies he identified as "best" were included, mean VSL was 2,830,000 and median VSL was 2,470,000. These estimates were based on 11 studies, having eliminated as unreliable 18 of the estimates.

Unfortunately, the criteria by which Jones-Lee identified the best studies are described only in fairly general terms and no evidence is provided to show that the studies Jones-Lee decided to disregard are biased. This part of his review must therefore be regarded as somewhat subjective and not very well validated.

9.2.3 Miller 1990

Ted Miller (1990) reviewed 67 estimates of the value of a statistical life. Most of the studies were made in the United States. Miller was concerned about the quality of data and analysis in these studies and decided to omit 20 of the estimates. He adjusted all the 47 estimates he retained. Some of the adjustments appear to be slightly arbitrary. Moreover, all the adjustments reduced the variation of the estimates. In this presentation, the results as originally reported have been included as this gives the most correct impression of the variability of estimates.

For most studies, Miller presented a range of estimates. In compensating wage differentials studies, for example, several econometric models will normally be developed to estimate the value of a statistical life. These models will often produce varying estimates of the value of a statistical life. For each study, there will therefore be a lowest and highest estimate, as well as a best estimate. The best estimate will normally be based on the preferred econometric model. The statistics describing variation in the estimates of the value of a statistical life can therefore be based either on the range in each study or on the best estimate in each study. To represent the ranges, the maximum estimate was the highest estimate reported in any study and the minimum estimate was the lowest estimate reported in any study. Results based on these two options are shown below (US dollars, 1988-prices):

	Based on ranges	Based on best estimates
Mean	3,753,000	3,327,000
Median	2,452,000	2,635,000
Standard deviation	3,436,000	2,646,000
Maximum	16,172,000	9,365,000
Minimum	0	0
Dispersion index	6.60	3.55

The difference between the mean and median is smaller than reported by Blomquist (1982) and Jones-Lee (1989). The value of the dispersion index is also considerably smaller than in the literature reviews reported by Blomquist and Jones-Lee. In particular, when only a single best estimate is extracted from each study, the dispersion index is markedly reduced.

The initial level of risk was reported in a majority of cases. In the compensating wage differentials studies, this is identical to the change in risk to which the estimate of the value of a statistical life applies. In the other studies represented in the review, which were studies of consumer behaviour and contingent valuation studies, the initial risk level will usually overstate the change in risk, since the safety devices bought by consumers, or the changes in risk presented in contingent valuation surveys, will rarely eliminate risk. Thus, willingness-to-pay refers to the elimination of risk in studies of compensating wage differentials, but to a reduction, not resulting in an elimination of risk in the studies of consumer behaviour and the contingent valuation studies. It is nevertheless of interest to show how mean willingness-to-pay was related to the initial level of risk. Figure 9.4 presents the relationship.

It is seen that willingness-to-pay increases almost in proportion to the level of risk. When the two data points on top are omitted, the relationship is even closer to linear, with an exponent of 0.8574 in the power function.
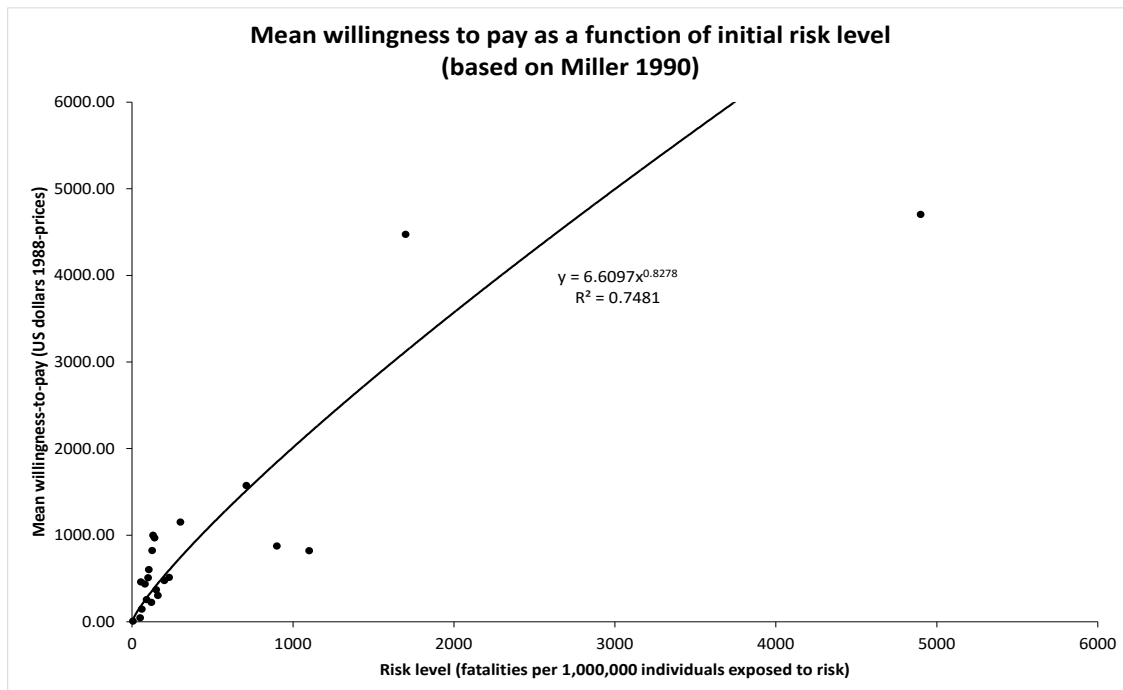

Figure 9.4: Relationship between initial level of risk and willingness-to-pay. Based on Miller 1990

Miller's review can be interpreted as a summary of the progressive phase of the research programme (see Chapter 6). The subsequent literature reviews and meta-analyses (see the following sections) were to paint an ever darker picture of the results of studies estimating the value of a statistical life.

9.2.4 Miller 2000

Miller (2000) studied variation between countries in the value of a statistical life. The study is an expanded version of the study reported by Miller in 1990 (see section 9.2.3). To the studies made in the United States, Miller added 30 estimates based on studies made in other countries. When these estimates were combined with the US studies, a total of 68 estimates were available for analysis.

Details of the estimates of the value of a statistical life were presented only for the non-US studies. For each study, a maximum and minimum estimate was given. In addition a best

estimate was extracted from each study. In the subsequent analysis, only the best estimate, corrected to after-tax dollars was used. Based on these estimates, the summary statistics presented below can be produced (US dollars 1995-prices):

	Based on ranges	Based on best estimates
Mean	3,778,000	3,220,000
Median	2,566,000	3,108,000
Standard deviation	4,129,000	2,191,000
Maximum	21,562,000	10,829,000
Minimum	0	678,000
Dispersion index	8.40	3.27

These statistics are based on the 30 non-US studies only. To explain variation between countries in the value of a statistical life, Miller fitted the following regression model to the 68 estimates:

$$Ln(VSL) = \alpha + \beta \ln(Y) + \gamma Z$$

The natural logarithm of the value of a statistical life (VSL) is explained by the natural logarithm of real income in a country (Y) and a set, Z, of other influencing variables. This type of model has become the standard model to explain variation in the value of a statistical life by means of regression analysis. In the models developed by Miller, GDP per capita was used as a measure of income. In one model, GDP per capita was adjusted to purchasing power parity. It should be noted that GDP per capita may be a misleading indicator of typical individual income if a high share of GDP is spent on purposes that do not directly benefit individuals, like military spending, or if the distribution of income is highly unequal, meaning that he majority of people earn less than the mean income.

Two sets of models were developed. One was based on the 68 individual studies. The other was based on the mean value of a statistical life in 13 countries. Six models were developed based on the 68 individual estimates. These models explained between 58 and 71 percent of the variation in the value of a statistical life. In addition to GDP per capita, the models included various variables describing study method (compensating wage differentials, other revealed preference method or contingent valuation) and the source of risk data employed in a study. In general, these variables were statistically significant.

Five models were developed based on the mean values for the 13 countries. These models included only GDP per capita and one or two other explanatory variables. The models explained between 84 and 92 percent of the variation in the value of a statistical life.

Apparently, these models explained most of the variation in the value of a statistical life. Yet a summary index of goodness-of-fit, like R-squared does not tell the whole story (Szklo and Nieto 2014). In Figure 9.5 the mean values of a statistical life estimated in studies made in each country have been plotted against the model-predicted values. Ideally speaking, if model predictions were perfect and unbiased, they would be located on top of the diagonal straight line drawn in the figure, or oscillate randomly around that line.

Figure 9.5 shows that this is not the case. Model predictions are systematically wrong. In most countries where national estimates are below about 3,500,000, the model systematically predicts higher values. In most countries where national estimates are about 3,500,000 or more, the model systematically predicts lower values. The best fitting line connecting actual and model-predicted values is a logarithmic function that clearly and systematically departs from the line of ideal predictions.

A sample size of 13, or even 68, is probably too small to give a basis for good models intended to explain the variation in the value of a statistical life. With only 13 data points, almost any model would have good fit, unless the data were very noisy. But the starting point for developing models to explain the variation in the value of a statistical life is exactly the opposite – that the data exhibit systematic variation, not simply random noise.
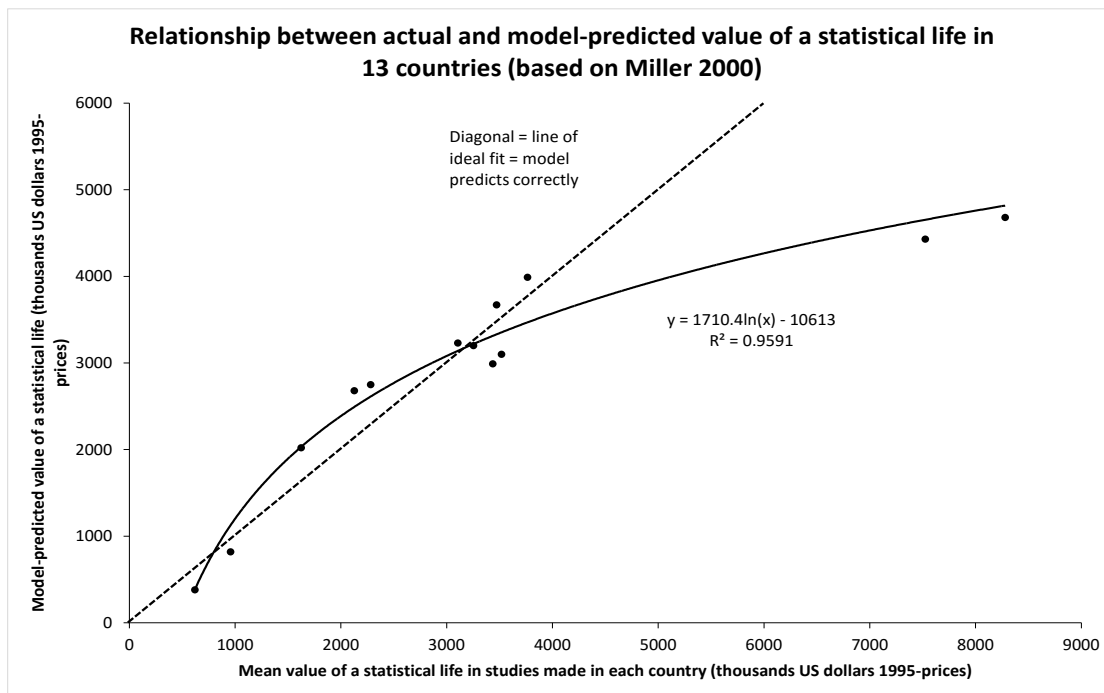

Figure 9.5: Relationship between national estimates and model-predicted estimates of the value of a statistical life. Based on Miller 2000

9.2.5 Mrozek and Taylor 2002

Janusz Mrozek and Laura Taylor (2002) published the first meta-analysis of estimates of the value of a statistical life. The studies presented so far were, possibly except for Miller (2000), not really meta-analyses. They were systematic literature surveys in which a simple mean of the results was estimated.

Mrozek and Taylor (2002) compiled a database of 203 estimates of the value of a statistical life based on 33 studies. All these studies were revealed preference studies based on the compensating wage differentials approach. The meta-analysis relied on the 203 estimates; however if a study contained, for example 8 estimates, each of them was assigned a weight of 1/8. Summary statistics regarding variation in estimates of the value of a statistical life (US dollars 1998-prices) are reported below.

The dispersion index has a quite low value, although the range of estimates goes from a low of 16,000 to a high of 30,700,000 (a factor of almost 2000, i.e. the highest estimate is almost 2000 times higher than the lowest). Estimates are slightly skewed, as indicated by the fact that the mean value is higher than the median value.

	Based on ranges	Based on best estimates
Mean	6,175,000	6,109,000
Median	5,150,000	5,230,000
Standard deviation	6,106,000	4,529,000
Maximum	30,700,000	16,550,000
Minimum	16,000	240,000
Dispersion index	5.96	3.12

Mrozek and Taylor (2002) developed four models to explain variation in the value of a statistical life. Before presented the results of their analysis, some characteristics of the data will be described.

Studies of compensating wage differentials are based on the theory that any undesirable characteristic of work, like a high risk of accident, will be compensated by higher wages. Figure 9.6 shows the simple bivariate relationship between hourly wages and risk (fatalities per million workers).
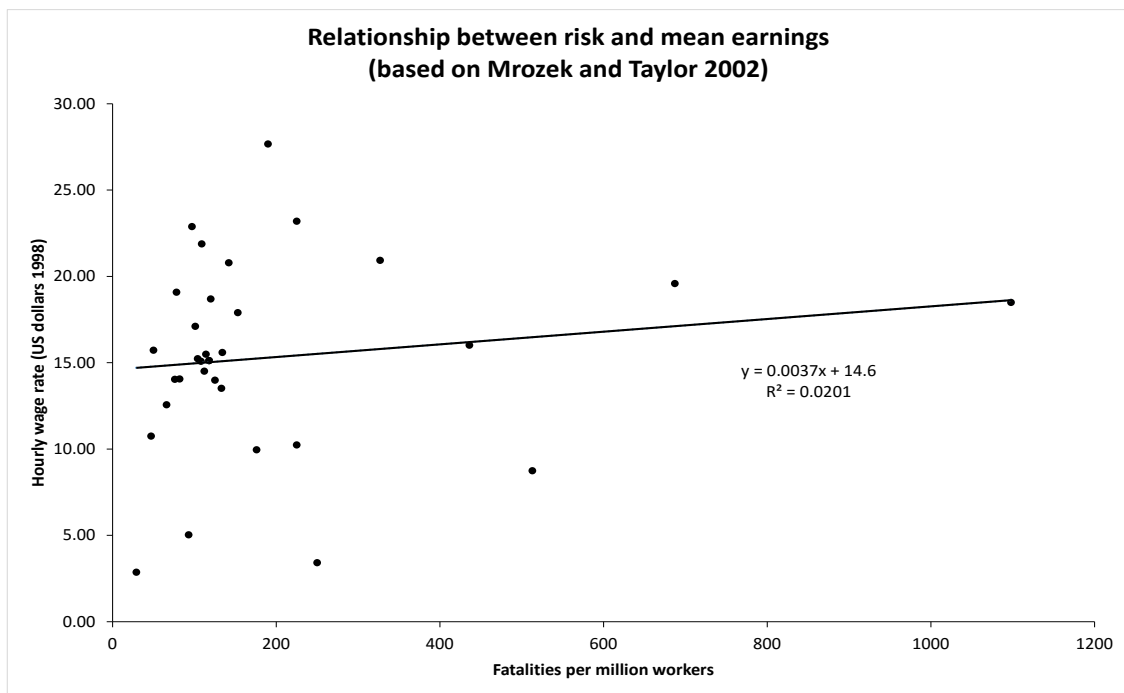

Figure 9.6: Relationship between hourly wages and occupational fatality risk. Based on Mrozek and Taylor 2002

In Figure 9.6 there is essentially no correlation between wages and risk. Obviously, this does not rule out that a stronger relationship will emerge in multivariate analyses. Since many factors influence wages, it is important to control for as many of these as possible in a multivariate analysis. Another problem, discussed by Mrozek and Taylor, concerns the quality of risk data. For the moment, it will be noted that both the quality of risk data and

the completeness of model specification are very important for the results of studies of compensating wage differentials. Any meta-analysis should therefore probe for the effects of these factors.

Figure 9.7 shows the relationship between the estimated value of a statistical life and risk. There is a strong negative relationship between the variables.
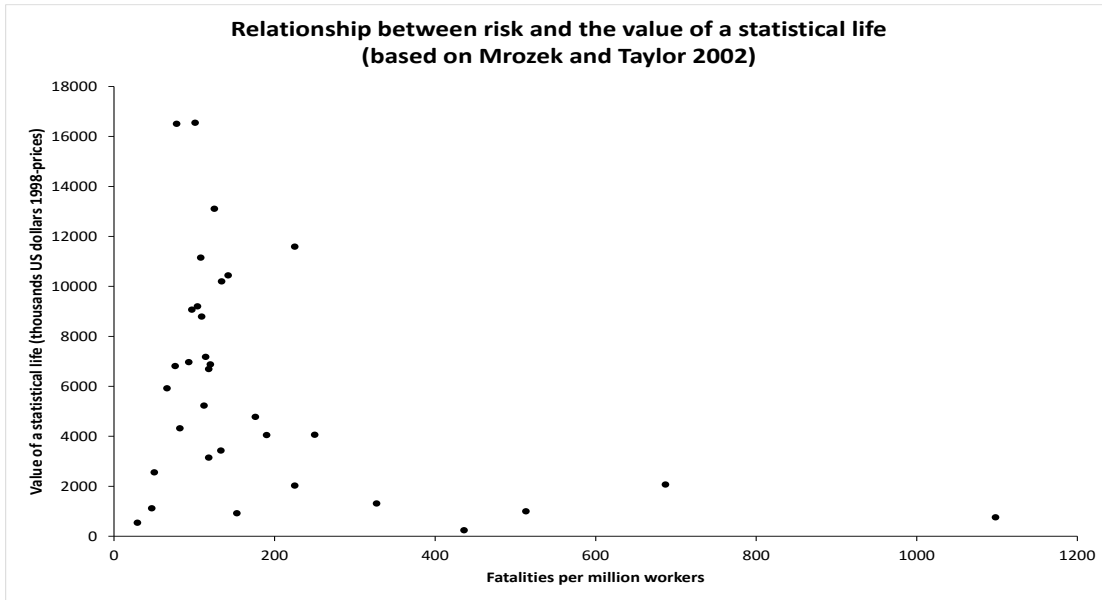

Figure 9.7: Relationship between estimated value of a statistical life and risk. Based on Mrozek and Taylor 2002

The negative relationship found in Figure 9.7 calls for a test of the relationship between the level of risk and willingness-to-pay for reducing a specific level of risk. Figure 9.8 shows this relationship.
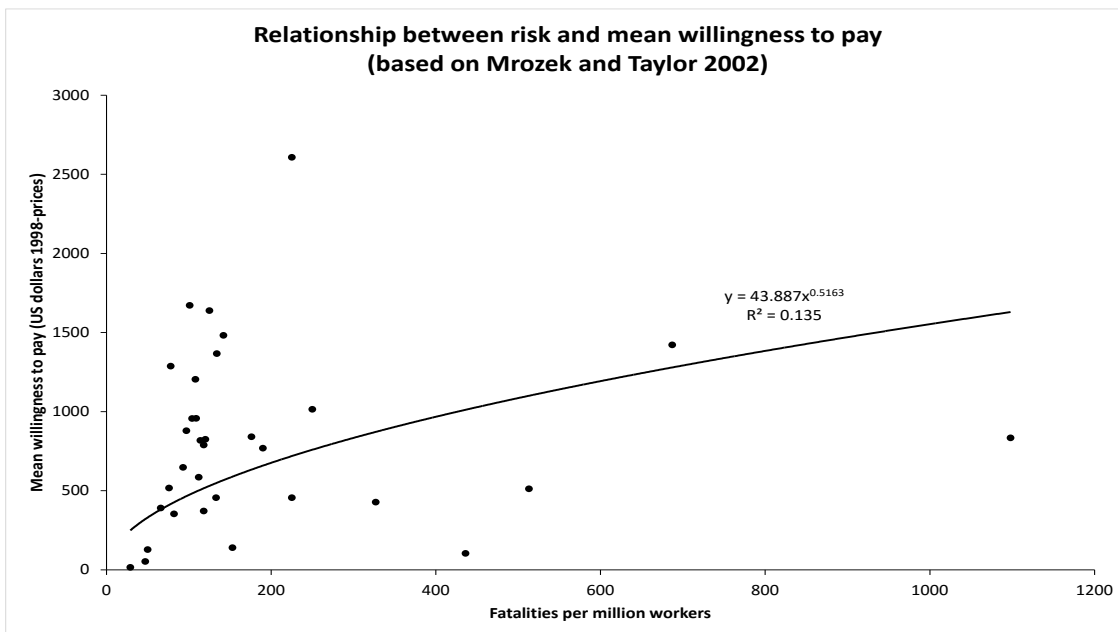

Figure 9.8: Relationship between risk and mean willingness-to-pay. Based on Mrozek and Taylor 2002

The relationship is quite noisy, but weakly positive. Taken together, Figures 9.6, 9.7 and 9.8 show data containing considerable noise and only weak tendencies towards systematic

relationships. How did Mrozek and Taylor approach the task of performing a meta-analysis of these data?

They developed four different models. A total of 28 explanatory variables were defined. The models explained between 71 and 77 percent of the variation in the value of a statistical life. Unfortunately, the analyses cannot be replicated, as the paper does to provide data for all the variables included in the models. It is clear, however, that Mrozek and Taylor were not very satisfied with any of the models, despite the fact that R-squared had acceptable values. They comment one of the findings of model 1, which included all estimates (203), like this (page 264):

"Model 1 indicates that the value of statistical life estimates begin to decline when the mean risk of a sample of workers becomes greater than approximately 1.2 deaths per 10,000. … These results may indicate that selection effects among workers with heterogeneous risk-preferences may dominate over some range of risks. In other words, those with lower risk aversion may be self-selecting into higher risk jobs and require less compensation, all else equal."

Mrozek and Taylor were concerned about study quality. For each of the explanatory variables, they therefore defined a "best practice". They then estimated "adjusted" values of a statistical life for a hypothetical high-quality study conforming to best practice with respect to all variables. Based on this analysis, they conclude that the best estimate of the value of a statistical life is probably in the range between 1,500,000 and 2,500,000 – a considerably lower value than the sample mean of about 6,000,000 US dollars (1998-prices).

In other words, studies of low quality tend to overestimate the value of a statistical life, a finding confirming the "Stainless steel law" of evaluation studies proposed by Peter Rossi (1987).

9.2.6   Blaeij et al. 2003

Arianne de Blaeij, Raymond Florax, Piet Rietveld and Erik Verhoef (2003) performed a meta-analysis of studies estimating the value of a statistical life in road safety. The meta-analysis included 30 studies providing a total of 95 estimates of the value of a statistical life. Summary statistics regarding the variation in estimates is provided below (US dollars 1997-prices; computed from Table 1 of the paper):

Mean	4,646,000
Median	1,903,000
Standard deviation	6,329,000
Maximum	29,933,000
Minimum	143,000
Dispersion index	15.65

There was wide dispersion in estimates of the value of a statistical life (note that the mean and median listed above were estimated on the basis of Table 1 in the paper by Blaeij et al. and are not identical to the values stated in the paper as they did not list all 95 estimates).

Two analyses were reported in the paper. The first was a subgroup analysis, in which the mean value of a statistical life was estimated for various groups of studies. As an example, the mean value of a statistical life was estimated to be 3,932,000 when safety was a private good, but only 2,124,000 when safety was a public good. Mean value based on willingness-

to-pay was 3,298,000; based on willingness to accept it was 4,368,000. The groups may not have been entirely homogeneous with respect to all factors that may influence valuation.

A meta-regression analysis was performed, using sample size as statistical weight. The best-fitting model explained 63 percent of the variation in estimated values of a statistical life. This model was based on stated preference studies only, containing a total of 54 estimates. Of particular interest are the results pertaining to the initial level of risk and the size of the change in risk. The coefficient for initial risk level is positive, showing that all else equal, the value of a statistical life increases as risk increases. The coefficient for risk reduction is negative, suggesting that the larger the risk reduction, the lower is the value of a statistical life. For a plausible range of values for initial risk and risk reduction, the value of a statistical life varied by a factor of about two. Based on these findings, Blaeij et al. (2003, page 984) conclude that:

"Our results suggest that VOSLs cannot be viewed independent of the prevailing level of risk and the hypothesized change in risk levels. The assumption that "life" can be summarized in a single numerical value ("the" VOSL), as is often suggested by scholars as well as policy makers, is neither sound from a theoretical perspective, nor warranted on the basis of empirical analysis."

In other words, the search for a value of a statistical life lead to the conclusion that the concept makes no sense.

9.2.7 Viscusi and Aldy 2003

William Kip Viscusi and Joseph Aldy (2003) reviewed revealed preference estimates of the value of a statistical life from all countries where estimates have been developed. Nearly all the estimates included in the review were based on studies of compensating wage differentials. A majority of the studies had been made in the United States of America. There was huge dispersion in estimates of the value of a statistical life, as indicated below (US dollars 2000 prices):

Mean	8,010,000
Median	5,250,000
Standard deviation	11,000,000
Maximum	74,100,000
Minimum	200,000
Dispersion index	14.08

There was, as found in previous reviews, a negative relationship between initial risk level and the value of a statistical life. This is shown in Figure 9.9. No meaningful relationship could be found between the level of risk and willingness-to-pay for reduction of a specific level of risk.

Viscusi and Aldy performed a meta-analysis based on the compensating wage differentials studies. The best-fitting model was based on 46 estimates of the value of a statistical life and explained 72 percent of the variation. The dependent variable was the natural logarithm of the value of a statistical life. The most influential independent variables were income (included as natural logarithm) and risk level. The mean estimated value of a statistical life for the full sample was 5.9 million US dollars (2000-prices) with a 95 % confidence interval from 2.7 to 13.9 million US dollars.
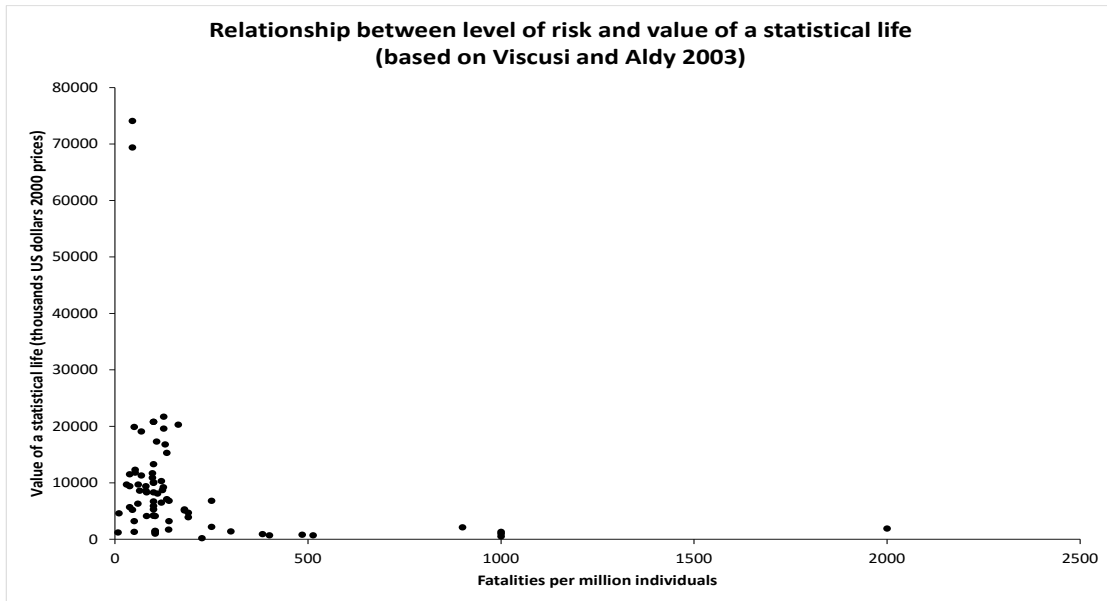
Figure 9.9: Relationship between risk level and value of a statistical life. Based on Viscusi and Aldy 2003

To give an impression of the range of values predicted by the model, estimates have been developed for: (1) Income of 26,279 (mean for sample), 50,000 and 12,500; (2) Risk of 197 per million (mean for sample), risk of 985 per million and risk of 39 per million. All these values are well within the range of values found in the data. When applying the model, the dummies for various characteristics listed by Viscusi and Aldy were not included. Only income, risk and risk squared were included. Income was found to have a great influence on estimates of the value of a statistical life. The risk variables had only a minor influence. The estimated value of a statistical life was 4,156,000 for the lowest income (12,500), 6,491,000 for the sample mean income (26,279) and 9,548,000 for the high income (50,000).

Thus, the meta-analysis confirmed what Blaeij et al. found, that a single value for a statistical life makes little sense.

9.2.8 Dionne and Lanoie 2004

Georges Dionne and Paul Lanoie (2004) reviewed studies estimating the value of a statistical life for the purpose of finding a best value for use in cost-benefit analyses of transport projects in Canada. The survey did not include a meta-analysis. Studies were grouped according to design and in each group the mean and median values were estimated. Although this was not a meta-analysis, it is nevertheless of some interest to present the main findings of the review. It included a total of 86 estimates of the value of a statistical life, and only one estimate was extracted from each study. 42 estimates were based on the compensating wage differentials method, 15 estimates were based on consumer behaviour and 29 estimates were based on the contingent valuation method. The number of estimates in each group is sufficient to compare the three methods to see if they give the same results. This comparison is shown below (VSL, Canadian dollars, 2000-prices).

	Wage differentials	Consumer behaviour	Contingent valuation	All methods
Mean	7,919,000	3,171,000	6,600,000	6,646,000
Median	6,304,000	1,553,000	3,716,000	4,707,000
Standard deviation	6,863,000	2,660,000	8,257,000	7,020,000
Maximum	25,679,000	8,811,000	33,000,000	33,000,000
Minimum	312,000	506,000	159,000	159,000
Dispersion index	4.02	5.34	8.84	6.98
Number of estimates	42	15	29	86

The compensating wage differentials method is associated with the highest mean estimate of the value of a statistical life. Revealed preference studies based on consumer behaviour is associated with the lowest mean estimate of the value of a statistical life. Within group variation in estimates, as indicated by the dispersion index, is smaller for the compensating wage differentials method and the consumer behaviour method than for the contingent valuation method. The skewness in the distribution of estimates is smallest for compensating wage differentials, as indicated by the fact that the median in this group of studies is closer to the mean than in the other two groups of studies.

Although the mean estimates are within the same order of magnitude, they still differ considerably. The value of a statistical life based on compensating wage differentials is more than twice the value based on studies of consumer behaviour. The main conclusion is therefore that method matters. Different methods for estimating the value of a statistical life do not produce identical estimates. The question is whether any of the methods can be identified as better than the other.

9.2.9 Kochi, Hubbell and Kramer 2006

Ikuho Kochi, Bryan Hubbell and Randall Kramer (2006) introduced an empirical Bayes approach to combining and comparing estimates of the value of a statistical life. Their analysis was based on 40 studies containing a total of 197 estimates of the value of a statistical life.

Unfortunately, Kochi et al. do not list the studies that were included in their analysis. It is therefore not possible to reconstruct the distribution of estimates of the value of a statistical life. The main point of the study was to show how to use the empirical Bayes method when estimating a weighted mean estimate of the value of a statistical life.

The empirical Bayes method is widely used in road safety evaluation studies (Hauer 1997). It is a widely applicable general approach to the estimation of unbiased long-term mean values in distributions that are characterised by a mixture of random and systematic variation. To apply the method, it must be possibly to identify and estimate the relative contributions of the random and systematic components to the observed variance in a distribution of observations. Empirical Bayes estimation involves a linear shrinkage of a distribution of values (Junghard 1992). Thus, in a sample of 239 marked pedestrian crossings in Oslo and its suburbs (Elvik 2016C), the count of accidents during a period of five years varied between 0 and 11. The empirical Bayes estimate of the long-term expected number of accidents varied between 0.28 and 10.37. In this case, the shrinkage was small,

because the model fitted to identify sources of systematic variation in the number of accidents explained only a small part of that variation.

The meaning of concepts like random and systematic variation is less clear when applied to a sample of estimates of the value of a statistical life. Kochi et al. assign a statistical weight to each estimate of the value of a statistical life based on what they call the sample variance. Apparently, by this term they refer to the squared standard error of the estimate of the value of a statistical life in each study. They go on to estimate a weighted mean value of a statistical life and test for heterogeneity in the distribution of estimates. Finally, an empirical Bayes estimate is developed for each estimate of the value of a statistical life.

The whole procedure hinges on accepting the standard error of the estimate of the value of a statistical life as a valid indicator of sampling variance. In nearly all studies, the value of a statistical life is estimated as a function of various independent variables. In that case, the standard error will depend both on which variables a model includes and on details of the model specification. If estimates of the standard error are highly sensitive to model specification, it reflects not just (random) sampling variance but also correlations with other variables included in a model.

For the full sample of 197 estimates of the value of a statistical life, Kochi et al. find that the mean value is 10.8 million US dollars (2000-prices). The range is from 0.1 to 95.5 million. The median estimate is not stated, but the dispersion index based on the mean value is $(95.5 - 0.1)/10.8 = 8.83$. It would most likely have been higher had the median been stated.

The empirical Bayes mean estimate of the value of a statistical life is 5.4 million US dollars (2000-prices). There was, in other words, a considerable shrinkage towards zero. Interestingly, the empirical Bayes mean value of a statistical life was 9.6 million for compensating wage differentials studies and 2.8 million for contingent valuation studies, confirming that different methods produce different values.

9.2.10 Kluve and Schaffner 2008

Jochen Kluve and Sandra Schaffner (2008) performed a meta-analysis of estimates of the value of a statistical life based on studies made in Europe. The analysis was based on 37 studies containing a total of 94 estimates of the value of a statistical life. Summary statistics, based on Table 1 of the paper, are given below. The statistics based on ranges are based on the ranges of values within each study. The statistics based on best estimates are based on extracting a single best estimate from each study.

There is enormous variation. The dispersion index based on ranges has the highest value found in any of the systematic reviews or meta-analyses that have been presented so far. Six models were developed to explain the variation in the value of a statistical life. The best fitting model explained 59 percent of the variation.

	Based on ranges	Based on best estimates
Mean	7,907,000	7,265,000
Median	3,337,000	3,848,000
Standard deviation	17,733,000	15,025,000
Maximum	119,308,000	89,377,000
Minimum	0	0
Dispersion index	35.76	23.23

According to the best fitting model, the method used was the strongest predictor. In studies based on consumer behaviour and contingent valuation studies, the mean value of a statistical life is almost 90 percent lower than in compensating wage differentials studies. Level of risk was not found to influence the value of a statistical life.

9.2.11 Bellavance, Dionne and Lebeau 2009

François Bellavance, Georges Dionne and Martin Lebeau (2009) report a meta-analysis of 39 estimates of the value of a statistical life extracted from 37 studies relying on the compensating wage differentials method. Three estimates for different countries were extracted from one study; otherwise only a single estimate was extracted from each study.

Despite the limited number of studies, the meta-analysis is very comprehensive and in many ways a pioneering work. It will therefore be discussed in some detail. It was, perhaps except for Kochi et al. (2006), the first meta-analysis of the value of a statistical life employing the inverse variance technique of meta-analysis. The authors estimated the standard error of the estimate of the value of a statistical life for each study. It is important to note that the appropriate statistical weight in meta-analysis is the inverse of the squared standard error of the estimate of the value of a statistical life, not sample size. The standard summary statistics showing the mean, median, etc. are presented below (US dollars, 2000-prices).

Mean	9,481,000
Median	6,049,000
Standard deviation	10,313,000
Maximum	53,627,000
Minimum	462,000
Dispersion index	8.79

Bellavance et al. (2009) noted that the dispersion in estimates of the value of a statistical life has increased over time. This was shown in Chapter 1, but the figure is repeated here for convenience (Figure 9.10). Figure 9.11 presents the relationship between risk level and the value of a statistical life. As found in previous reviews, the relationship is negative.
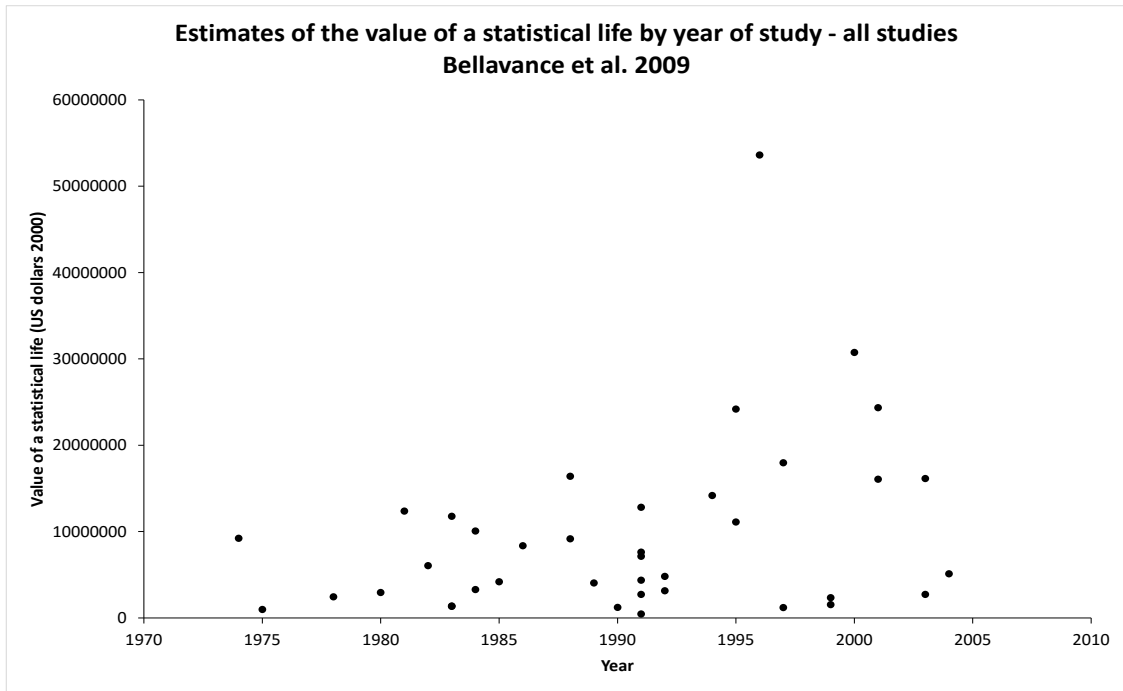
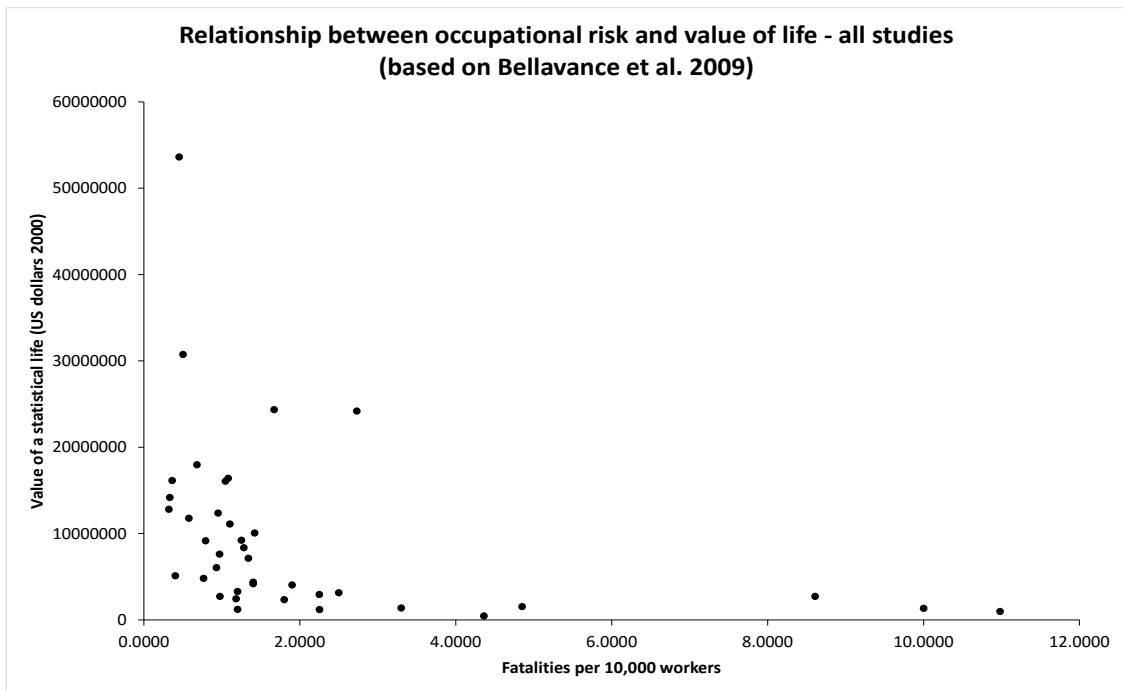*Figure 9.10: Estimated values of a statistical life by year of study. Based on Bellavance et al. 2009*

Figure 9.11: Relationship between fatality risk and value of a statistical life. Based on Bellavance et al. 2009

A random-effects weighted meta-regression analysis was performed, using the estimated value of a statistical life as dependent variable (unlike most other analyses, which have used the natural logarithm of the value of a statistical life as dependent variable). A total of nine different models were developed. The analysis confirmed that the value of a statistical life is positively related to income and negatively related to risk. Bellavance et al. interpret the latter finding as a result of selective recruitment of less risk averse workers into high-risk occupations. The values of the coefficients varied substantially between the models, meaning that estimates of the value of a statistical life are sensitive to the choice of model.

9.2.12 Doucouliagos, Stanley and Giles 2012

Chris Doucouliagos, Tom Stanley and Margaret Giles (2012) re-analysed the meta-analysis reported by Bellavance et al. (2009) and found that it was strongly influenced by publication bias. Their study is the first to document the presence of publication bias in studies estimating the value of a statistical life.

A common tool for exploratory analysis in order to detect the possible presence of publication bias is the funnel plot. Figure 9.12 shows a funnel plot of the estimates of the value of a statistical life included in the meta-analysis reported by Bellavance et al. (2009).
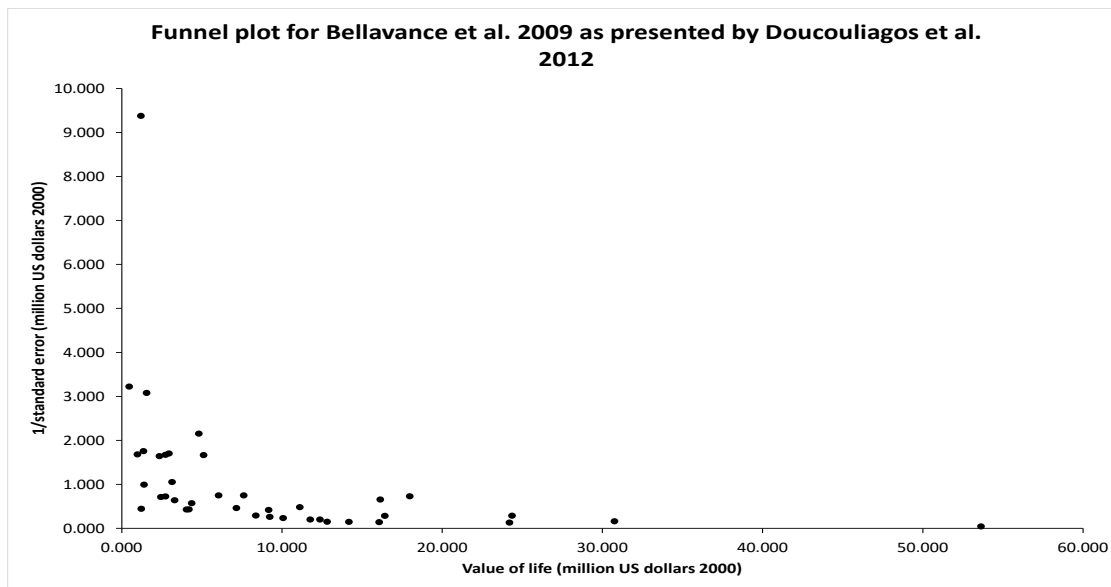

Figure 9.12: Funnel plot of estimates of the value of a statistical life included in meta-analysis reported by Bellavance et al. 2009

The idea underlying the funnel plot, is that if all studies dealing with a topic have been published, the data points should distribute like a funnel turned upside down. Precise estimates, located at the top of the diagram, should be less dispersed than imprecise estimates, located at the bottom of the diagram. If one of the tails of the distribution is missing, this indicates publication bias. Doucouliagos et al. (2012) applied a meta-regression technique to adjust for publication bias (details of this method are explained in section 9.2.14). The best estimate of the value of a statistical life then dropped from 9.48 million to 1.66 million, indicating a massive publication bias.

The meta-regression technique used by Doucouliagos et al. (2012) is not the most common method for assessing and adjusting for publication bias in meta-analysis. Is it possible to detect and adjust for publication bias by means of, for example, the trim-and-fill technique?

To test for this, a conventional inverse-variance weighted mean estimate of the value of a statistical life was estimated, applying a fixed-effects model and including all 39 estimates reported by Bellavance et al. (2009). The summary estimate of the value of a statistical life was 1.76 million – without adjusting for publication bias. A single study contributed 64 percent of the statistical weights. This summary estimate is remarkably close to the estimate reported by Doucouliagos et al. (1.66 million), but very far from the mean value of 9.48 million. The reason for this is no mystery. The mean value of 9.48 million is simply the arithmetic mean, giving all studies the same weight. This underscores the importance of applying appropriate statistical weights in meta-analysis.

Given the fact that the (weighted) summary estimate was just 1.76 million, is there still any evidence of publication bias? To test for this, the funnel plot shown in Figure 9.13 was developed.
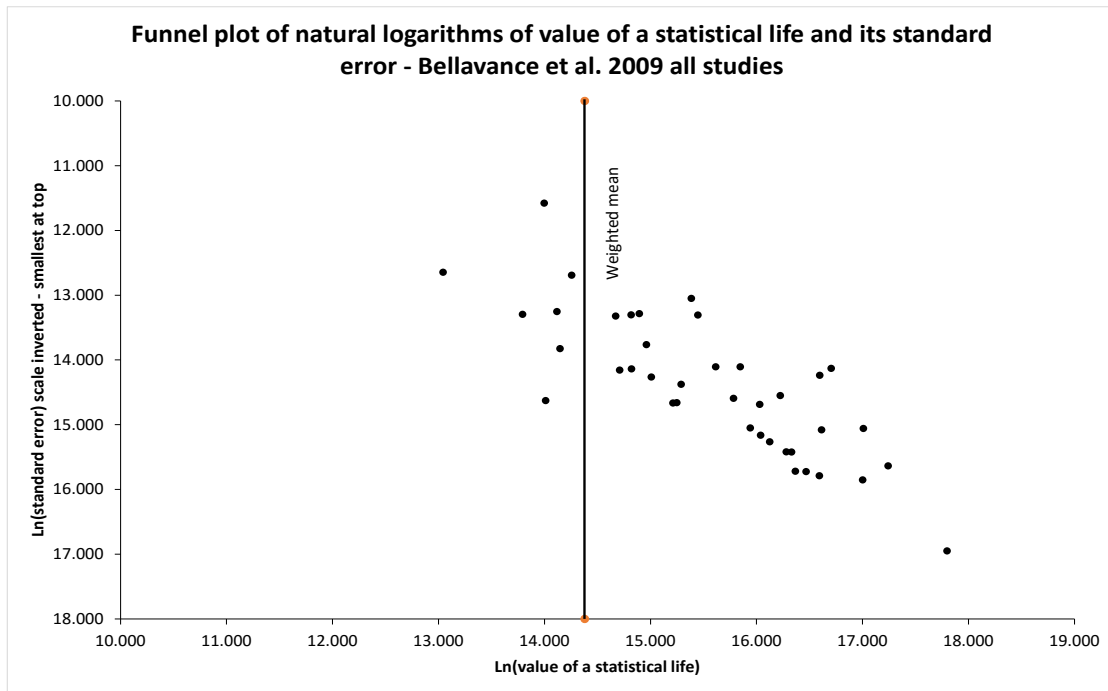

Figure 9.13: Funnel plot of estimates of the value of a statistical life included in meta-analysis by Bellavance et al. 2009. Logarithmic scale

The weighted mean (1.76 million in original units) is indicated by the straight line. The data points clearly have a skew distribution, with most of them located to the right of the summary estimate. The asymmetric shape of the funnel plot indicates publication bias.

A formal test was performed by means of the trim-and-fill method (Duval and Tweedie 2000A, 2000B, Duval 2005). The analysis indicated 20 missing data points. Figure 9.14 shows these.

The trimmed mean estimate of the value of a statistical life was 1.51 million. Thus, although 20 out of 39 data points were deleted, the trimmed mean was still not very different from the overall mean (1.76 million). There was clearly publication bias, but not of the magnitude suggested by Doucouliagos et al (2012). What they found was basically only that when individual estimates are appropriately weighted by their inverse variances, you get a different summary estimate than if you simply assign the same weight to all studies. The meta-regression method introduced by Stanley and Doucouliagos may, however, nevertheless be useful as an indicator of publication bias, as will be shown later (see section 9.2.14) (Stanley 2008, Stanley and Doucouliagos 2013, 2014).
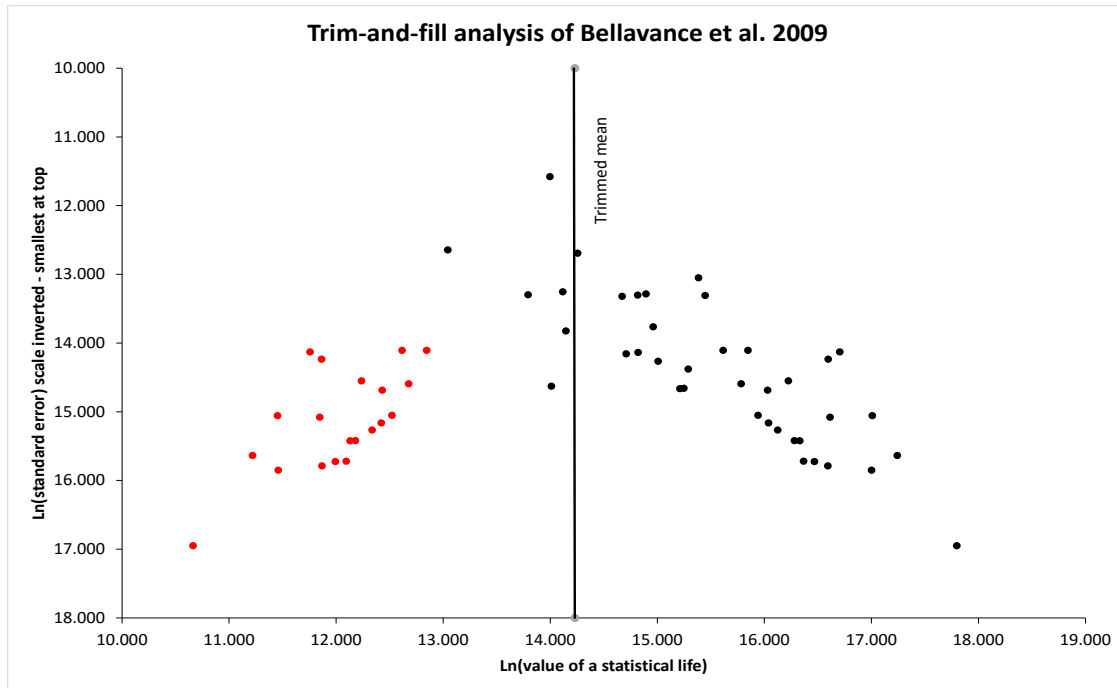
Figure 9.14: Results of trim-and-fill analysis of estimates of the value of a statistical life included in meta-analysis by Bellavance et al. 2009

Unfortunately, the standard error of an estimate of the value of a statistical life is not always reported in primary studies. The papers of Kochi et al. (2006) and Bellavance et al. (2009) show that estimating the standard error when it is not presented in a study is quite complex and data demanding. It may therefore not be feasible to perform these estimates. An item which is reported far more often – indeed in almost all studies – is sample size. Can sample size serve as an indicator of standard error?

To probe this question, the unique data set compiled by Bellavance et al. (2009) was once again used. Sample size was stated for all studies. A funnel plot was developed. The trim-and-fill method was then applied to this funnel plot. The mean estimate of VSL, weighted by sample size was 8.45 million, considerably higher than the mean weighted by inverse variance (1.76 million). The analysis indicated publication bias. 16 data points were deleted (20 were deleted in the inverse-variance analysis). The trimmed mean estimate of VSL was 4.37 million. Thus, relying on sample size did reveal publication bias, showing that it may not be altogether misguided to use it as a proxy for standard error.

9.2.13 Lindhjem, Navrud, Braathen and Biausque 2011

Henrik Lindhjem, Ståle Navrud, Nils Axel Braathen and Vincent Biausque have performed by far the largest and most comprehensive meta-analysis of studies of the value of a statistical life. Their study has been reported a number of times (Lindhjem et al. 2010, 2011, 2012A, 2012B). Here two of the reports will be used. The first is a paper published in Risk Analysis in 2011 (Lindhjem et al. 2011). The meta-analysis presented in that paper was based on 856 estimates of the value of a statistical life extracted from 74 studies. All studies were stated preference studies. Revealed preference studies were not included. Summary statistics based on all 856 estimates of the value of a statistical life are shown below(US dollars 2005-prices, adjusted to purchasing power parity).

Mean	6,065,000
Median	2,378,000
Standard deviation	14,365,000
Maximum	197,000,000
Minimum	4,450
Dispersion index	82.86

There was an enormous variation in estimates, ranging from nearly 200 million to just 4000 US dollars. Skewness in the distribution is apparent from the fact that the median is considerably lower than the mean. The dispersion index had a much higher value than in any of the other meta-analyses. Trying to account for this huge variation was therefore a daunting task.

Nils Axel Braathen has kindly sent a copy of the very extensive database created for this meta-analysis. This allows some interesting statistics to be produced. Figure 9.15 shows the development over time of estimates of the value of a statistical life.

The development is irregular. Until the early 1980s the dispersion increased. Then it was reduced until about 1995, after which the dispersion of estimates increased dramatically. In recent years, however, the dispersion of the estimates has again been reduced.

Figure 9.16 shows the relationship between the change in risk and the value of a statistical life.

There is, as other meta-analyses have found, a negative relationship between the change in fatality risk and the value of a statistical life. The dispersion of estimates of the value of a statistical life is greater when small changes in risk are studied than when larger changes (to the right on the horizontal axis) are studied. One possible reason for this is that respondents may have difficulty in understanding small changes in low levels of risk and answer somewhat arbitrarily.
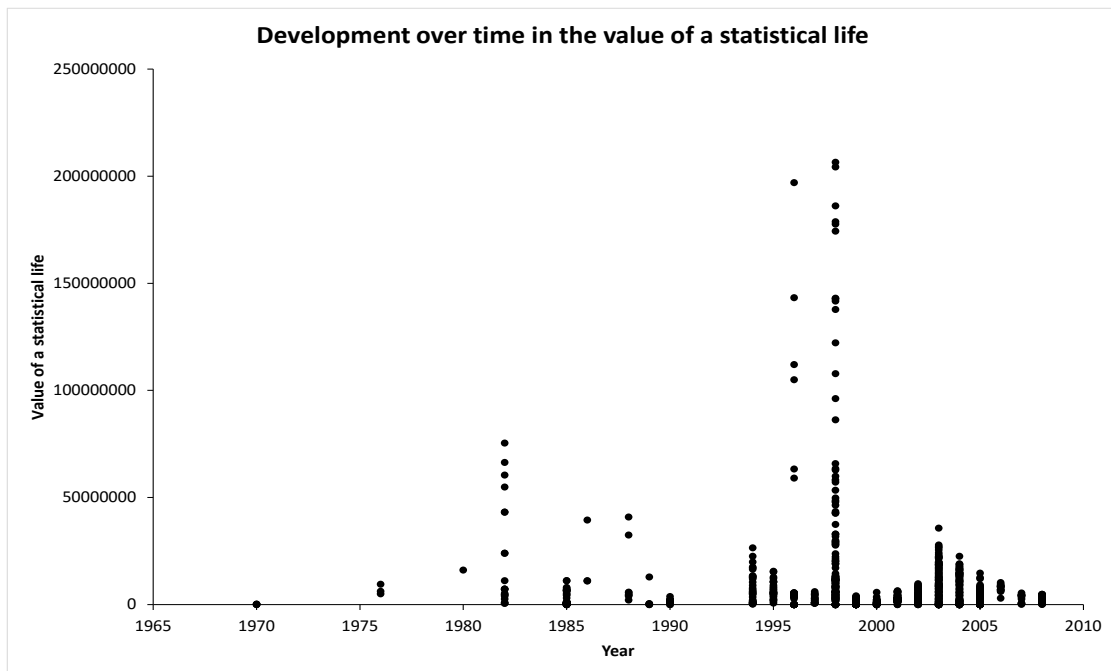

Figure 9.15: Development over time in the value of a statistical life. Based on database for Lindhjem et al. 2012
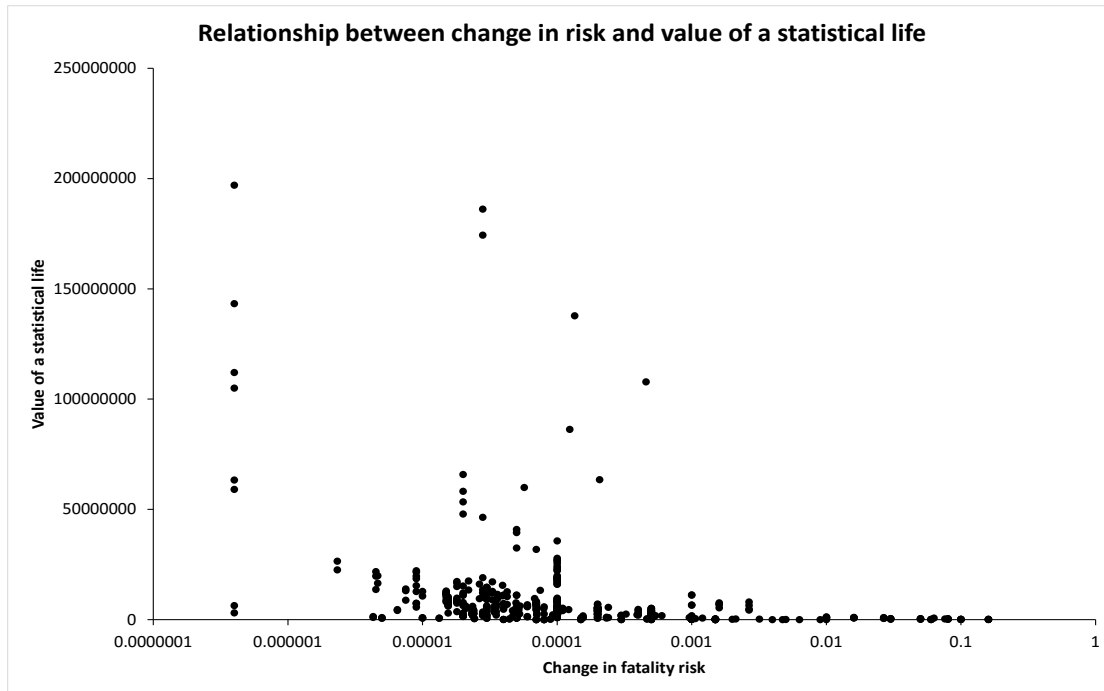
Figure 9.16: Relationship between change in risk (in nearly all cases a reduction) and value of a statistical life. Based on database for Lindhjem et al. 2012

Lindhjem et al. (2011) developed several models to explain variation in the value of a statistical life. Models were fitted both to the full dataset and to subsets of it. Perhaps the most informative model as far as traffic risk is concerned, is Model V fitted to the subset that contained data about change in risk. The model explained 83 percent of the variation in the value of a statistical life. The strongest predictors of the value of a statistical life were GDP per capita and the size of the change in risk. To illustrate the range of predicted values, estimates have been developed for the following combinations of values:

GDP per capita: 40,000; 60,000; 80,000

Risk change: $5 \cdot 10^{-6}$; $10 \cdot 10^{-6}$; $15 \cdot 10^{-6}$

Type of good: private or public

GDP per capita is in the range of current GDP per capita in Norway (given in US dollars adjusted to purchasing power parity). The change in risk represents realistic levels of change in the risk of a road accident fatality in Norway (fatality rate in 2015 was about 24 per million inhabitants; hence a reduction of 15 per million would imply reducing traffic fatalities by about 63 percent). Measures taken to reduce risk can either be private goods (buying a safer car) or public goods (installing road lighting). Results are given below. The values are in US dollars 2005-prices, adjusted to purchasing power parity.

Risk reduction (per million)	GDP per capita (US dollars, 2005)		
	40,000	60,000	80,000
	Safety as a private good		
5	29,094,768	39,966,337	50,063,520
10	19,503,861	26,791,686	33,560,396
15	15,435,330	21,202,906	26,559,654
	Safety as a public good		
5	11,676,267	16,039,229	20,091,415
10	7,827,259	10,751,998	13,468,406
15	6,194,483	8,509,118	10,658,880

It is seen that the estimated value of a statistical life varies by a factor of about 8 even for this limited range of combinations of values on the explanatory variables. All three variables included are seen to have a major effect on estimates of the value of a statistical life.

9.2.14 Lindhjem, Navrud, Braathen and Biausque 2012

Lindhjem et al. (2012B) is a so far unpublished update of the meta-analysis discussed above. It is presented in a manuscript dated January 30, 2012. It is based on an updated database, now containing 937 estimates of the value of a statistical life. The results are very similar to those reported above. A detailed presentation of the results will therefore not be given.

An issue not dealt with in the meta-analyses of Lindhjem et al. (2012B) is publication bias. As noted in the discussion of the meta-analysis by Bellavance et al. (2009) and the subsequent re-analysis of this by Doucouliagos et al. (2012), publication bias is an issue in meta-analyses of the value of a statistical life.

Testing for publication bias by means of, for example, the trim-and-fill technique is difficult as not all studies state the standard error of the estimate of the value of a statistical life. By examining the database provided by Nils Axel Braathen, 514 estimates of the value of a statistical life and its standard error were identified. A funnel plot was developed on the basis of these 514 estimates and is shown in Figure 9.17.
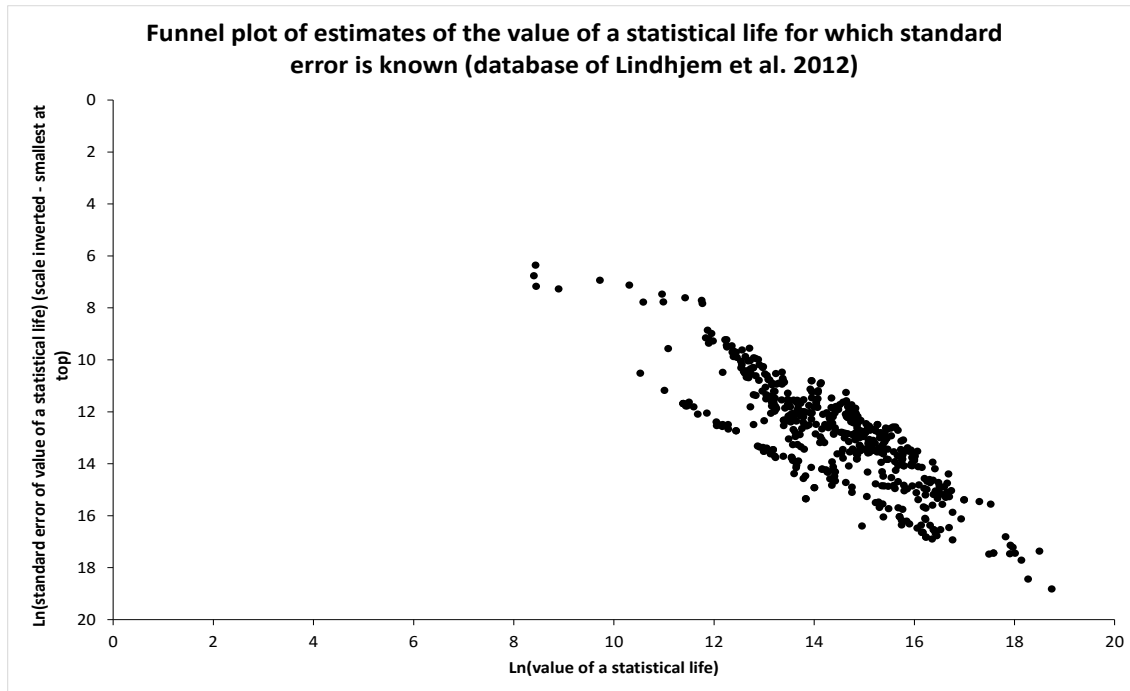
Figure 9.17: Funnel plot of estimates of the value of a statistical life for which the standard error is known. Based on database of Lindhjem et al. 2012

There is a remarkable skewness in this funnel plot, suggesting massive publication bias. When estimates are weighted by their inverse variance, the summary estimate of the value of a statistical life is close to the data points located on top and on the far left. This is to the left of nearly all data points and suggests that the inverse-variance technique cannot be used, because there is a high correlation between the estimates of the value of a statistical life and the standard errors of those estimates.

The simple mean of the estimates of the value of a statistical life represented in Figure 9.17 is 5,328,478 – not very different from the mean value for the entire data set (6,065,000). The inverse-variance weighted mean is just 25,000 – clearly a misleading estimate which is smaller than 509 of the 514 estimates.

Doucouliagos et al. (2012) present the following formula for estimating a summary estimate correcting for publication bias:

$$\overline{VSL} = \alpha_0 + \alpha_1 SE_i + \varepsilon_i$$
(1)

Here, the relationship between the value of a statistical life (VSL) and its standard error (SE) is described by a linear function. If the two variables are unrelated, the coefficient for standard error (α_1) will be zero (i.e., there is no slope in the relationship between VSL and SE) and the constant term (α_0) will equal the mean value of a statistical life. If the variables are related, as shown in Figure 9.17, the constant term shows the value of a statistical life corrected for publication bias and the slope shows the degree of publication bias. Doucouliagos et al. note that equation 1 is generally not estimated because of its heteroskedasticity. Heteroskedasticity means unequal variance and is readily apparent from Figure 9.12, where estimates of VSL located at the bottom of the diagram are much more widely spread than those located at the top of the diagram. This is what a funnel plot would normally show.

The funnel plot in Figure 9.17 does, however, not show a similar heteroskedasticity. The bandwidth of the spread of the data points appears to be fairly constant across the range of estimates of VSL. Equation 1 was therefore applied. The resulting estimate of the value of a statistical life was 2,411,759. The arithmetic mean was 5,328,478. There is, in other words, considerable publication bias.

It should be noted that one possible explanation for the distribution of the data points in Figure 9.17 is that many estimates of VSL have been extracted from the same study (close to 11.6 on the average). Estimates from the same study tend to be correlated. This reduces variance and makes the estimates cluster closer together.

9.3   Meta-analyses of special topics

In addition to the meta-analyses of studies estimating the value of a statistical life, there have been a few meta-analyses of special topics. Three of these topics have been selected (these three are those that have been most frequently studied):

1. Publication bias
2. The discrepancy between estimates based on willingness-to-accept and estimates based on willingness-to-pay
3. Hypothetical bias in contingent valuation studies

9.3.1   Publication bias

It has been noted for several of the reviews presented above that one may suspect the presence of publication bias. Viscusi (2015) has tested this a little more systematically, although only for revealed preference studies using the compensating wage differentials approach.

He compiled two data sets. The first contained 550 estimates of the value of a statistical life based on 17 studies. All these studies employed risk data from the Census of Fatal Occupational Injuries (CFOI), which is currently the best source of data about occupational accidents in the United States. The mean value of a statistical life based on this data source was estimated as 14.035 million US dollars (2013-prices). Several models were developed to test for publication bias; all of them were meta-regression precision estimates of the form developed by Stanley and Doucouliagos. The lowest of the estimates was 8.145 million. All models indicated publication bias, but Viscusi (2015) judged it as being of moderate magnitude.

The other data set consisted of the studies included in the meta-analysis of Bellavance et al. (2009), augmented by studies included in some older meta-analyses, such as Viscusi and Aldy (2003). The mean estimates of the value of a statistical life in four samples of studies varied between 10.212 and 12.883 million US dollars (2013). Again a number of different models (all of the Stanley and Doucouliagos type) were developed to test for the presence of publication bias. In the most comprehensive of these models, the value of a statistical life was estimated to be between 3.503 and 4.268 million US dollars, indicating considerable publication bias.

The review of systematic literature surveys meta-analyses presented in section 9.2 provided several indications of publication bias in some of the surveys and analyses. It was first noted that the arithmetic mean estimate of the value of a statistical life based on the contingent valuation studies reviewed by Jones-Lee (1989) was 3,090,000 pounds. Adjusted for publication bias, the mean value was only 773,000 pounds.

Next Doucouliagos et al. (2012) found publication bias in the meta-analysis of Bellavance et al. (2009), a finding that was replicated by means of a trim-and-fill analysis. Applying the meta-regression technique of Doucouliagos et al. (2012), publication bias was found in the studies reviewed by Lindhjem et al. (2012) for which the standard error was stated. Finally, Viscusi (2015) found publication bias in five data sets he examined for the purpose of testing for the presence of publication bias.

The evidence from these reviews is summarised in Figure 9.18, which compares crude mean estimates of the value of a statistical life to estimates that have adjusted for publication bias.
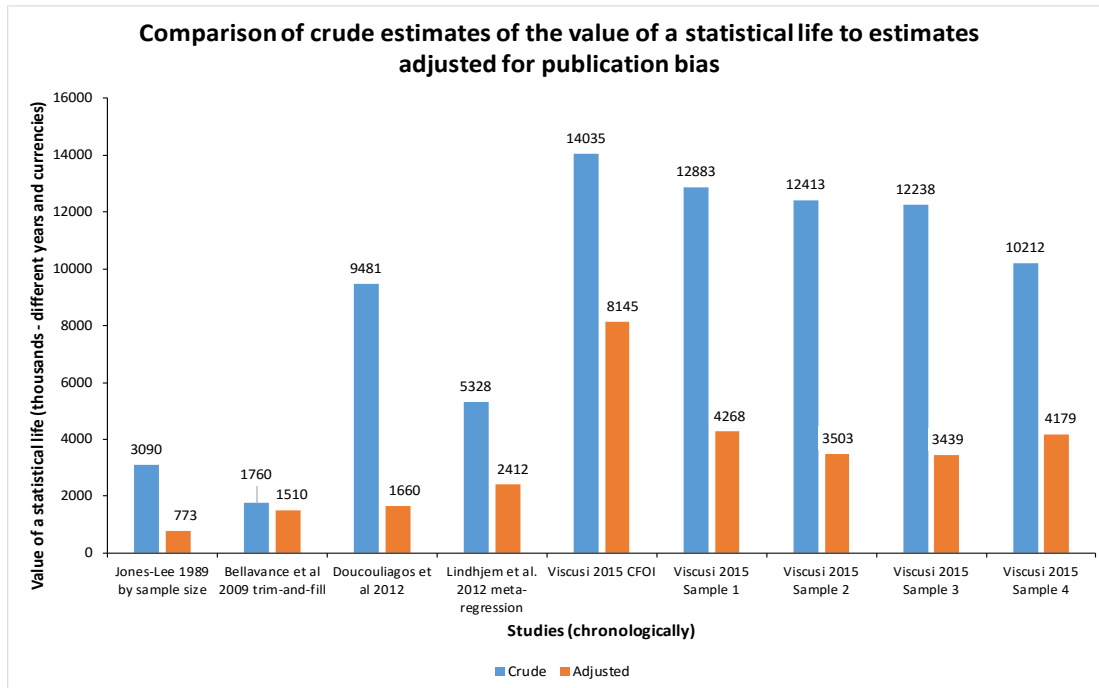

Figure 9.18: Comparison of crude estimates of the value of a statistical life to estimates that have adjusted for publication bias

It is seen that evidence suggesting publication bias has been found in all studies that have tested for it. It is obviously very difficult to demonstrate publication bias positively, by locating unpublished studies and comparing their estimates to those of published studies. In that sense, all statistical techniques designed to test for the presence of publication bias rely on assumptions whose validity cannot be tested in a straightforward manner, but must simply be accepted as reasonable when using the techniques. Nevertheless, if one accepts the analyses, they indicate a very large publication bias in studies eliciting the value of a statistical life. The mean of the crude estimates in Figure 9.18 is 9049 (9.049 million). The mean of the adjusted estimates is 3321 (3.321 million). This shows that published studies may overestimate the value of a statistical life by a factor of almost three.

Are there reasons to believe that there is publication bias in studies of the value of a statistical life? Unfortunately, there are several reasons. First, theory predicts that the value of a statistical life will almost always be positive. If a positive value is not found, researchers may be reluctant to publish their findings, since it would often be difficult to give a convincing interpretation of the finding. It is indeed very unreasonable to expect the value of a statistical life to be zero or even negative. On the other hand, some positive values may seem implausibly large. If these values are accepted, it may induce publication bias.

Second, estimating the value of a statistical life in revealed preference studies involves the use of complex econometric methods, which give analysts ample opportunities for testing various models and refining the models until they are satisfied with the results. Anyone who has tried to develop a somewhat complex multivariate model knows that this is to a large extent a process of trial and error. Indeed, some phenomena are so complex that one cannot hope to find the best model on the first attempt. So when do analysts stop looking for a better model? One might surmise that they stop when they have found a model that confirms their theoretical predictions.

Third, in contingent valuation surveys, it is not uncommon that a sizable proportion of respondents state a zero willingness-to-pay. How to interpret zero responses has been a matter of controversy, but some researchers (Miller and Guria 1991) argue that zero bids are really protest answers and should be disregarded. If one accepts this interpretation, and goes on to estimate the value of a statistical life based only on those who state a positive willingness-to-pay, it is clear that the estimated value will be biased upwards.

Fourth, in some of the early contingent valuation studies, various analyses were made to evaluate whether respondents satisfied criteria of rationality when stating willingness-to-pay. This practice was continued in stated choice studies, in which one could, for example, test consistency in a sequence of choices. In most cases, it was found that the more rational respondents were found to be, the lower were their valuations. In recent studies, testing rationality has gone out of favour and is rarely done, or at least rarely reported. Analysts are perhaps reluctant to apply what might be viewed as "censorship" or, worse still, paternalism by rejecting some answers in contingent valuation studies or some choices in stated choice studies. The result is likely to be that published estimates of the value of a statistical life are higher than they would be if only estimates based on respondents rated as highly rational were published.

Taken together, these reasons conspire to make it quite likely that there will be publication bias in value of life studies.

9.3.2 Willingness-to-accept versus willingness-to-pay

As noted in Chapter 5, there are two ways to estimate the value of a non-market good:

1. By finding out how much people are willing to pay to obtain the good (buying price = willingness-to-pay).
2. By finding out how much people need to be paid to abstain from the good (selling price = willingness-to-accept).

In equilibrium in a perfect market, buying price of course equals selling price. There is just one price, the market equilibrium price, and a distinction between buying price and selling price does not have to be made. However, when trying to value non-market goods, the distinction between willingness-to-pay and willingness-to-accept has turned out to be very important.

Horowitz and McConnell (2002) performed a meta-analysis of the disparity between WTA and WTP found in 45 studies containing a total of 208 estimates. On the average, the ratio between WTA and WTP (WTA/WTP) was 7.17. For non-market goods, the ratio was 10.41. WTA was higher than WTP even for market goods. This could be interpreted as a straightforward falsification of the hard core of neoclassical economic theory. Interestingly, Horowitz and McConnell are reluctant to put forward such an interpretation, stating that (page 427):

"This paper does not take up the issue of whether the WTA/WTP findings provide evidence for or against the neoclassical paradigm, even though that potential has been the theme of much of the literature we

reviewed. Rather, our goal is to draw broad-based results from this long and rich experimental track record. We are not concerned with whether the observed ratios are consistent with the standard neoclassical model. It is possible to investigate these results without addressing the neoclassical question."

It is obviously possible to review a body of empirical research without discussing its theoretical implications. Perhaps the authors did not want to discuss theoretical implications because, as Lakatos suggested, a finding that prima facie refutes a theory does not lead to the rejection of that theory until a better theory is proposed. One may for sure propose a psychological theory to explain the WTA/WTP disparity, but that theory would, from the perspective of neoclassical economic theory probably be regarded as an ad hoc theory, a hypothesis proposed to explain an anomaly, with no general implications (i.e. capable of explaining the anomaly only, nothing else).

Horowitz and McConnell end their paper (page 442) by asking the following questions, to which they do not offer any answer:

"Second, does the WTA/WTP disparity provide sufficiently broad and deep evidence against the neoclassical model? Does that evidence warrant substantially modifying that model, at least in some situations for which economists' expertise might be called upon?

Is the WTA/WTP disparity relevant in studies of the value of a statistical life? It is highly relevant. Nearly all estimates of the value of a statistical life based on stated preference studies are WTP-estimates. However, some of the revealed preference studies, in particular studies of compensating wage differentials, are best interpreted as WTA-estimates. Kniesner, Viscusi and Ziliak (2014:188) write the following about the interpretation of these studies:

"Labour market estimates capture the compensating differential that workers require to incur job risks as compared to a risk-free job. Consequently, from the vantage point of a model in which workers are comparing a hypothetical baseline risk-free job with a risky job, the estimated wage-risk tradeoffs are not estimates of willingness-to-pay (WTP) for a decrease in risk, but rather are measures of willingness to accept (WTA) for the increase in risk associated with taking the hazardous job compared to the safe alternative."

Given this interpretation, one would expect estimates of the value of a statistical life based on compensating wage differentials approach to be higher than, for example, estimates based on contingent valuation surveys. At least two of the systematic reviews and meta-analyses reviewed above confirm this. Dionne and Lanoie (2004) found a mean VSL of 7,919,000 in hedonic wage studies, versus 6,600,000 in contingent valuation studies. Kochi et al. (2006) estimated the mean value of a statistical life (VSL) to 9.6 million in hedonic wage studies and 2.8 million in contingent valuation studies. There is thus evidence of a WTA/WTP disparity with respect to the value of a statistical life.

9.3.3 Hypothetical bias

One of the oldest objections to the contingent valuation method is that the answers are hypothetical only. Respondents are normally not asked to pay anything. While reminders of budget constraints are given in some studies, these may not be sufficient to prevent respondents from overstating their actual willingness-to-pay in order to express support for a good cause or give an impression of selfless idealism.

James Murphy, Geoffrey Allen, Thomas Stevens and Darryl Weatherhead (2005) reported a meta-analysis of hypothetical bias in stated preference studies. The analysis was based on 28 studies with a total of 83 estimates of the relationship between hypothetical and actual willingness-to-pay. The mean ratio between hypothetical and actual willingness-to-pay was 2.60, i.e. hypothetical willingness-to-pay overstated actual willingness-to-pay by 160 percent.

Meta-regression found that hypothetical bias was reduced when valuation took the form of stated choice.

Harrison and Rutström (2008) also reported on hypothetical bias in valuation surveys. They did not perform a meta-analysis, but listed studies that had found hypothetical bias. In nearly all cases, the bias was positive and the maximum bias was 2600 percent.

Few of the stated preference studies of willingness-to-pay for reduced risk of death have tried to assess whether there is hypothetical bias, and, in case such a bias is found, how to correct for it. Based on the reviews quoted above, it must be concluded that there is a high probability of hypothetical bias in stated preference surveys, but that – failing systematic attempts to ascertain this bias – its magnitude is unknown.

Carson and Groves (2007) are admirably clear about when the results of stated preference surveys can be taken seriously. They ask (2007:183):

"Does a survey question need to meet certain conditions before it can be expected to produce useful information about an agent's preferences? This question is easy to address. First, the agent answering a preference survey question must view their responses as potentially influencing the agency's actions. Second, the agent needs to care about what the outcomes of those actions might be."

They make a distinction between consequential questions and inconsequential questions, and make it clear that economic theory applies only to the former. A consequential question must have an impact on the respondents welfare. It can only have such an effect by being more than merely an expression of an opinion. It should predict actual behaviour. If the answers to preference questions are not validated by studying how well they predict behaviour, they should be treated as inconsequential, and thus not really being possible to interpret by means of economic theory.

9.4 Summary and discussion of meta-analyses

The main problem this chapter aims to answer is whether summarising and analysing the results of valuation studies by means of meta-analysis has succeeded in creating order in chaos, that is: (1) Succeeded in explaining why estimates of the value of a statistical life vary so enormously and (2) Showing that the primary sources of this variation are factors that according to the hypotheses forming the protective belt are expected to influence willingness-to-pay for reduced risk of death. Both points must be satisfied in order to conclude that the enormous variation in estimates has been successfully explained. Were it to be found that the principal explanation of the variation is methodological artefacts, or factors that ideally speaking should not be associated with variation in estimates, such as hypothetical bias in contingent valuation studies or a preponderance of lexicographic and inconsistent choices in stated choice studies, it must be concluded either that the methods used in valuation studies are inadequate, or that sufficiently well-ordered preferences for the provision of safety do not exist. Such a conclusion would confirm that chaos is to be expected, not that what looks like chaos actually hides a systematic pattern confirming the predictions of economic theory.

To be conclusive, meta-analyses must satisfy certain minimum standards. Some of these were discussed at the beginning of this chapter. A few more will be added now. The standards are based on Elvik (2016B), the QUOROM statement (Moher et al. 1999) and Nelson and Kennedy (2009). The standards are (keywords in parentheses):

1. The meta-analysis should be based on a systematic literature search aiming to include all studies about the topic (literature search).

2. The meta-analysis should include all study designs that have been employed to obtain a monetary valuation of safety (all designs included).
3. The meta-analysis should rely on optimal statistical weighting of each estimate, preferably inverse variance weighting (correct weighting).
4. The meta-analysis should include an exploratory analysis of the distribution of estimates to assess whether a main analysis would make sense (funnel plot).
5. The meta-analysis should decide how best to extract data from studies that produce multiple estimates of valuation (dependence of estimates).
6. The meta-analysis should test for the possible presence of publication bias and adjust for such bias if it is indicated (publication bias).
7. The meta-analysis should try to assess the quality of primary studies, preferably be means of a numerical quality score (quality assessment).
8. The meta-analysis should include a meta-regression analysis of factors influencing variation in estimated monetary valuations. Hypotheses about the sign of coefficients should be stated explicitly (meta-regression).
9. Several versions of the meta-regression model should be tested in a sensitivity analysis (model exploration).
10. The sensitivity of results to the exclusion of low-quality studies should be tested (quality screening).

Table 9.1 shows how the systematic reviews and meta-analyses that have been reviewed in this chapter score with respect to the ten points above. It may be noted that the early reviews were not based on a systematic literature search, did not apply statistical weights and did not include meta-regression analysis. These reviews satisfied only a few of the requirements listed above. More recent analyses have a better quality. In particular, the meta-analysis of Lindhjem et al. is very comprehensive and satisfies most of the requirements for a good meta-analysis. It did not test for publication bias, and the analysis reported in this chapter indicates that it may be influenced by publication bias.

The task facing those who want to explain variation in estimates of the value of a statistical life has grown more difficult over time. Figure 9.19 shows how the dispersion index for studies estimating the value of a statistical life has changed over time.
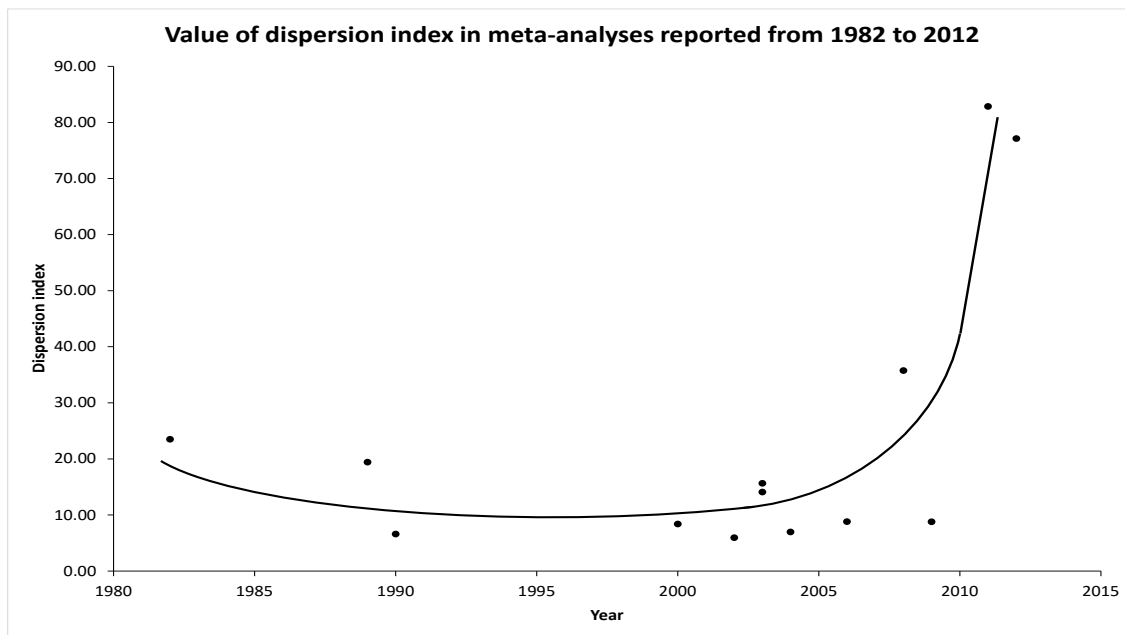

Value of dispersion index in meta-analyses reported from 1982 to 2012

Figure 9.19: Changes over in the dispersion index for studies of the value of a statistical life

From 1982 until early after 2000, the dispersion index tended to become smaller. In recent years, however, it has grown dramatically. One may perhaps interpret Figure 9.19 as a graphical illustration of the phases of a scientific research programme. The progressive phase was characterised by studies showing smaller-and-smaller dispersion in the values of a statistical life. But then, the trend turned around. The research programme entered a degenerative phase characterised by numerous anomalous findings and an increasing dispersion in estimates of the value of a statistical life. At the same time, methodological innovations took place. These innovations lead to a critical re-examination of older studies, made before the methodological innovations were made. Reviewers of the literature were thus faced with a choice about whether to reject older studies or keep them in the base of studies included in a review.

Table 8.1: An overview of systematic reviews and meta-analyses of studies of the value of a statistical life

Study	Literature search	Designs included	Statistical weighting	Funnel plot analysed	Dependence of estimates	Publication bias	Quality assessment	Meta-regression	Model comparisons	Quality screening
Blomquist 1982	No	All	None	No	Not relevant	Not tested	No	No	No	No
Jones-Lee 1989	No	All	None	No	Not relevant	Not tested	Crude	No	No	Crude
Miller 1990	No	All	None	No	Not relevant	Not tested	No (§)	No	No	No
Miller 2000	No	All	None	No	Not relevant	Not tested	No	Yes	Yes	No
Mrozek and Taylor 2002	No	Only one	None	No	1/N weight	Not tested	Yes	Yes	Yes	Yes
Blaeij et al. 2003	No	All (traffic)	Sample size	No	Discussed	Not tested	Yes	Yes	Yes	No
Viscusi and Aldy 2003	No	Only one	None	No	Not relevant	Not tested	No	Yes	Yes	Indirectly ($)
Dionne and Lanoie 2004	No	All	None	No	Not relevant	Not tested	Crude	No	No	Crude
Kochi et al. 2006	No	Two	Yes	No	By subsets	Not tested	No	No (#)	No	No
Kluve and Schaffner 2008	Yes	All	No	No	Not relevant	Not tested	No	Yes	Yes	Indirectly ($)
Bellavance et al. 2009	Yes	Only one	Yes	No	Not relevant	Not tested	No	Yes	Yes	Indirectly ($)
Doucouliagos et al. 2012	No	Only one	Yes	Yes	Not relevant	Tested	No	Yes	Yes	No
Lindhjem et al. 2011	Yes	Two	None	No (£)	1/N weight	Not tested	Yes	Yes	Yes	Yes

(§) Many estimates were adjusted to account for unreliable data sources

(#) Subgroup analysis was performed

(£) A distribution of estimates was shown, both in natural units and on a log scale

($) By means of variables included in meta-regression, intended to capture quality aspects

A valuable element of the meta-analysis made by Lindhjem et al. (2011) is that they stated explicit hypotheses about the expected sign of the coefficients for a number of variables. With respect to the change in risk (converted to ln(change in risk) in the analyses), their hypothesis was a coefficient of zero, i.e. the variable should not influence the value of a statistical life.

This means that the value of a statistical life should be constant, i.e. independent of the size of the change in risk. As is easily seen, this implies that WTP for specific changes in risk should be proportional to the size of those changes:

Risk change: $10 \cdot 10^{-6}$	WTP = 50	VSL = 5,000,000
Risk change: $5 \cdot 10^{-6}$	WTP = 25	VSL = 5,000,000
Risk change: $1 \cdot 10^{-6}$	WTP = 5	VSL = 5,000,000

As for other factors influencing the value of a statistical life, the sign was indeterminate for all variables except for a dummy for cancer risk (positive coefficient expected), whether willingness-to-pay referred to a household or an individual (positive coefficient), and income (positive coefficient expected).

The indeterminacy of the expect effects is not surprising in view of the completeness of the protective belt developed for valuation research, as discussed in Chapter 5. Lindhjem et al. (2011) worry about the low sensitivity to scope (i.e. the fact that willingness-to-pay does not increase in proportion to the amount of the good) indicated by the negative sign of the coefficient for risk change. They need not worry. Such a relationship is entirely consistent with one of the utility functions that have been proposed to account for the findings of valuation studies (Amiran and Hagen 2010). As noted in Chapter 5, the proliferation of utility functions seeking to explain or "normalise" every anomaly is a problem.

The discrepancy between WTA and WTP, the small sensitivity to scope, the indeterminate relationship between age and willingness-to-pay – all these findings and many more are consistent with theory.

Thus, trying to make sense of the wide dispersion of empirical estimates of the value of a statistical life is very difficult, since any pattern one might find in these estimates by means of meta-analysis would be consistent with one or more contributions to the theory of willingness-to-pay. Willingness-to-pay seems to be "insufficiently" sensitive to scope (whatever that may mean). But, if individuals have directionally bounded utility functions, as suggested by Amiran and Hagen (2010), this finding is to be expected and does not necessarily indicate that individuals have misunderstood the valuation task or are behaving irrationally.

For sure, both interpretations are possible. However, to sort them out, one would need to find out whether individual preferences can be represented by means of a directionally bounded utility function. If there is support for this, one could argue that insensitivity to scope is to be expected and is not an anomalous finding. On the other hand, individual preferences may be broadly consistent with many utility functions, not just a directionally bounded utility function. In other words, several different utility functions might describe individual preferences almost equally well, making it difficult to identify one of them as the best. There is thus no way of resolving the issue. Sceptics will argue that individuals have simply not understood the valuation task and are pulling numbers from the air. Defenders of the hard core will, on the other hand, invoke one of the theoretical contributions to argue that: no, there is no anomaly here. This is entirely as expected.

10 A dissolution of the hard core?

10.1 A research programme in trouble

As evidenced in Chapters 5, 7, 8 and 9, valuation research as a scientific research programme is in trouble. There are many troublesome aspects of the research programme:

1. The development of theory intended to protect the hard core of the programme has, in a manner of speaking, been entirely too successful. It has now come close to representing not just a protective belt, but an immunising stratagem (Popper 1979). This means that theoretical propositions in willingness-to-pay theory are immune to falsification, i.e. any result is consistent with theory. This is a problem, because it ruins the essential function of theory in any progressive scientific research programme, which is to help in identifying positive heuristics, i.e. results that make sense and from which new implications can be deduced. However, if any result makes sense, theory ceases to give support in interpreting empirical research, in particular in guiding research intended to explain why estimates of the value of life vary so enormously. All you can say is that the wide dispersion of estimates makes sense and is not surprising from a theoretical point of view.

2. This in turn leads to further questions: Has the hard core been dissolved? Or are the various theoretical propositions self-contradictory? These questions will be discussed in this chapter, but a preview of the conclusions can already be given: No, it is not correct to claim that the hard core has been dissolved, since all theoretical models are based on the axioms of rationality and utility maximisation that form the hard core. There is, in this sense, no essential differences between the various theoretical models of willingness-to-pay; they differ only in terms of the assumptions made about specific characteristics of individual utility functions. By the same token, the theoretical propositions are not self-contradictory. While, on the surface, empirical predictions of insensitivity to scope are inconsistent with predictions of sensitivity to scope, these predictions are based on different models, both of which can be true.

3. There is, however, another sense in which the hard core can be said to have dissolved. If widely differing estimates of the value of a statistical life are tolerated, one may ask what remains of the consistency argument economists made in favour of using a uniform value of a statistical life in all sectors of society, because it is only by doing so that saving a life is treated as equally valuable in all sectors, so that the number of lives saved is maximised. Has the hard core, thus interpreted, been dissolved? Have economists stopped arguing for a uniform value of a statistical life? If so, how much and by what criteria should the value of a statistical life be allowed to vary? These questions will be discussed in this chapter.

4. As briefly noted in Chapter 2, there are actually multiple standards of consistency within the theory of valuation of reduced risk of death. An important issue is to clarify whether these standards of consistency are contradictory or not. If some of the standards contradict one or more of the others, there is a need to discuss whether: (a) The contradictions can be resolved, or (b) A lexical priority established between the various standards of consistency, or (c) One should simply live with

the inconsistency. This chapter will discuss possible contradictions between standards of consistency.

5. During the past 30 years, behavioural economics has grown rapidly and attracted an increasing number of adherents. Various theories of choice have been proposed within behavioural economics, such as prospect theory (Kahneman and Tversky 1979) and regret theory (Loomes and Sugden 1982). These theories are intended to be descriptive, but in a number of recent contributions, which will be reviewed in this chapter, the question is asked whether normative guidance can be extracted from behavioural economics. To once more preview conclusions, the neoclassical model remains unrivalled as a normative, or prescriptive model. It is difficult to derive policy prescriptions from models that are based on systematic violations of rationality. Doing so would at least require replacing rationality and the efficient use of resources with different normative ideals for public policy, and so far none exist.

6. It has been noted in the previous chapters that insensitivity to scope generates a tendency for the value of a statistical life to be inversely related both to the initial level of risk and the size of the risk reduction. The lower the risk, and the smaller its reduction, the higher the value of a statistical life. This tendency has mostly been viewed as anomalous, resulting from weaknesses of the studies, and not as a valid expression of individual preferences. However, the tendency has been found both in stated and revealed preference studies. Moreover, reviews of how much society spends to reduce risk have also found a similar inverse relationship: the smaller the risk, the larger the expenditure per life saved. The possibility therefore exists that what looks like grossly inefficient priorities actually reflect individual preferences. In this chapter, the value of a statistical life estimated for various risk reductions based on the meta-analysis of Lindhjem et al. (2012) will be compared to the values implied by various regulatory decisions in order to assess if the two sets of values are consistent or not.

The issues raised above are discussed in the sections that follow.

10.2 From a uniform to an individualised value of life?

There is little doubt that the initial objective of valuation research was to obtain a uniform value of a statistical life for application in all sectors of society. However, even some early papers warned that this might be unrealistic. Thus, Weinstein, Shepard and Pliskin (1980:393), quoted in Chapter 5, remarked:

"It has been shown that the notion of a unique willingness-to-pay value per expected life saved is inconsistent with the utility theory of the individual. The value per life saved depends on the level of the mortality probability being changed, and not just on the increment: the higher the level, the higher the value. Moreover, the value obtained ex ante will differ from the value obtained ex post, the ex post value being generally the greater of the two."

It did not take many years before the great disparity in empirical estimates of the value of a statistical life started to be noticed. Thus, Mishan wrote in 1985 (1985:160):

"The values attributed to a statistical life by economists basing their calculations either on revealed or expressed preferences are so various as to excite mirth even among professionals. To my knowledge, the lowest of these calculated values of life (in 1980 dollars) is about $ 15,000 in Needleman (1980) for roofing workers. The figure produced by the Thaler and Rosen (1976) study, based on 1967 data, average about $ 200,000, or more than twice that amount in 1980 dollars. Of the several other calculations made, the highest mean valuation of life, about $ 10 million, was derived by the questionnaire method from a rather limited sample by Jones-Lee (1976)."

Mishan noted that there was insensitivity to scope: large risk reductions were not valued more highly than small risk reductions. This had the somewhat problematic consequence that the smaller the risk reduction, the larger was the estimated value of a statistical life. Mishan labelled this irrational. His conclusion was quite radical (1985:167):

"For these reasons, I propose that economists give up the search to discover a value for human life, or for a statistical life in any specific circumstances, and, in order to avoid error, to restrict their investigations of those projects or policies that, inter alia, affect life and limb to deriving consistent estimates only of compensating variations for the relevant changes in the specific risks associated with the particular project in question, these alone being the pertinent benefits or losses."

In short, for every project that affects risks to life and health, there will exist a monetary valuation which is unique for that project, and one cannot hope to obtain any monetary valuations of more general validity.

Did Mishan reject the research programme he helped to initiate merely some 15 years earlier? Yes and no. Yes, in the sense that the studies available by 1985 seem to have convinced him that individuals may have great trouble in understanding small changes in low levels of risk and showing adequate sensitivity to such changes. He urged that testing for consistency should be part of any study. Yes, also in the sense that he did not believe that a uniform value of life existed or could be found by research. No, in the sense that he had not abandoned using willingness-to-pay as the criterion for assessing projects. However, willingness-to-pay had to be estimated in each project and one should not expect the values to be the same from project to project.

It was noted in Chapter 9, that de Blaeij et al. (2003), in their meta-analysis remarked that: The assumption that "life" can be summarized in a single numerical value ("the" VSL), as is often suggested by scholars as well as policy makers, is neither sound from a theoretical perspective, nor warranted on the basis of empirical analysis.

No scholar has made more sweeping statements in favour of disaggregating the value of statistical lives than Cass Sunstein (2004). In a paper dated February 2004, freely accessible on the web, but not known to have been published in a scientific journal, he makes the following statements (pages 3 and 4):

"The value of a statistical life (VSL) should vary along two dimensions. First, it should vary across risks. For example, there is reason to think that VSL is higher for cancer deaths than for sudden, unanticipated deaths; Deaths that produce unusual fear, or that are accompanied by high levels of pain and suffering, should be expected to produce a higher VSL. … Second, VSL should vary across individuals. People who are risk averse will show a higher VSL than people who are risk-seeking. People who are thirty will show a higher VSL than people who are sixty. Those who are rich will show a higher VSL than those who are poor. It follows that different demographic groups will show diversity in their VSL as well."

He goes on to lament that a fully individuated VSL is currently not possible. He continues to argue that:

"… We can see that there is not one VSL, but an exceptionally large number of VSLs. In fact each of us has not one VSL but a number of them, targeted to each risk that each of us faces. A policy that truly tracks WTP would seek to provide each person with the level of protection for which he is willing to pay to reduce risk. Tracking WTP is the goal that underlies current practice; and apart from questions of administrability, it calls for a maximum level of individuation."

While one could reasonably argue that this is a correct interpretation of the theory of willingness-to-pay, viz. the many models introduced in Chapter 5 identifying factors influencing WTP, it also makes eminently clear the self-defeating, if not self-contradictory nature of that theory. To repeat the question asked in the introduction to this chapter: What is left of the consistency argument of economists if one allows VSL to vary along the dimensions suggested by Sunstein? How fine-graded should such variation be? Should it be based both on characteristics of risk and individual characteristics? How can one know that an appropriate range has been determined? Should the highest value be ten times the lowest, or a thousand times greater than the lowest? These are just a few of the questions that come to mind when reading Sunstein's interpretation of the theory of willingness-to-pay for changes in risk.

Sunstein's call for greater differentiation in the value of a statistical life has not been implemented anywhere, at least to the extent he called for. However, the complete absence of any tendency towards a convergence in empirical estimates of the value of a statistical life even in recent, and comparatively rigorous studies, has prompted other scholars to ask the same question as Sunstein. Has the time arrived for allowing the value of a statistical life to vary?

Hoel (2003) asked: Is life more valuable to the rich and healthy than to the poor and sick? He states: *"Almost everybody would answer "no" to this question. Economists are the only exception."* He went on to develop utility models supporting a positive answer to the question he asked in the title of his paper. Yes, life is more valuable to the rich than to the poor and more valuable to those in perfect health than to those in a reduced health state.

In two papers, Baker et al. (2008, 2009) discussed whether the widespread practice of using a uniform value of a statistical life, i.e. the arithmetic mean of willingness-to-pay (WTP) in a population, is consistent with the theoretical foundations of cost-benefit analysis, and, if not, if an acceptable normative foundation can be defined for applying a single value of a statistical life. They note that (2009:814, 815):

"In spite of the tendency to apply a uniform VSL in any given context, so far as we are aware of no satisfactory theoretical foundation has so far been provided that justifies the application of a common WTP-based VSL equal to, say, the overall population arithmetic mean of marginal rates of substitution (MRS) of wealth for risk of death by a given cause, other than under conditions which from a practical point of view appear somewhat implausible. … To the extent that the marginal rates of substitution will typically depend upon the income, age and other personal characteristics of those affected by the safety improvement, the logic underpinning standard cost-benefit analysis would seem to require that the VSL employed in the evaluation of a safety improvement that affects a poorer (or older) group in society should be smaller than the value applied to a wealthier (or younger) group. … Clearly, though, if a normative rationale is to be provided for this "uniform valuation of safety" approach then this will have to rely upon value judgments that differ somewhat from those underpinning conventional cost-benefit analysis."

In short, the principles of cost-benefit analysis, as presented here by Baker et al. imply that VSL should vary according to the variation in WTP between different groups in society. A uniform value is inconsistent with cost-benefit analysis. These points of view are clearly at odds with the efficiency arguments made in favour of applying a uniform VSL. As was shown by means of simple numerical examples in Chapter 2, it is only by applying a uniform VSL that the number of lives saved for a given public expenditure can be maximised.

Baker et al. (2008, 2009) show that a uniform VSL is justified if, rather than equalising marginal welfare with respect to changes in risk (i.e. equalising the change in utility associated with a change in risk, which depends on the utility of money), a cost-benefit analysis aims to equalise the marginal social welfare of survival probability. The uniform value to be applied for equalising marginal social welfare with respect to survival

probability is the harmonic mean of the marginal rates of substitution between wealth and risk. Since wealth tends to have an approximately lognormal distribution, the harmonic mean will be lower than the arithmetic mean. The harmonic mean is defined as the value c that satisfies the following equation:

Harmonic mean $= \frac{2}{c} = \frac{1}{a} + \frac{1}{b}$

Thus, the arithmetic mean of 3 and 6 equals $9/2 = 4.5$. The harmonic mean of 3 and 6 is 4 $(1/3 + 1/6 = \frac{1}{2} = 2/c; c = 4)$.

Baker et al. (2008) were clearly not convinced that trying to obtain a uniform marginal social welfare with respect to survival probability was a very practical criterion, stating (2008:137) that it is "to say the least – fairly restrictive". They note that many people think that one should allow the value of a life to vary depending on, for example, age.

Viscusi (2010) has also discussed the policy challenges associated with the huge variation in estimates of the value of a statistical life. He states that whether heterogeneity in estimates of VSL should be incorporated in policy evaluations (cost-benefit analyses) depends in part on the source of heterogeneity. However, as far as age and income are concerned, he clearly recommends taking account of these factors by applying VSL-estimates that vary by age and income. Regarding age, he states (2010:121):

"In many contexts, such as those involving regulations that affect people with very short remaining life expectancy, it is not appropriate to use the standard VSL measure. Rather, taking into account the difference in longevity often leads to the use of the VSLY (value of a statistical life year)."

He goes on to show that the value of a statistical life year depends on age. It has an inverted U-shape and reaches maximum at an age of about 50 years (based on empirical studies). Viscusi sums up whether income effects should matter in the following terms:

"The proper benefits measure should be grounded in the WTP of the beneficiaries of the policy. Whether these individual preferences indicate a positive income elasticity or other types of heterogeneity in preferences does not invalidate the importance of adhering to reliance on individuals' WTP for guidance in setting benefit levels. To impose constraints on income adjustments or any other aspects of the benefit assessment process in effect overrides individual preferences and the pivotal economic role of consumer sovereignty."

Thus, Viscusi and Baker et al. agree that strict adherence to the principles of cost-benefit analysis implies that differences in income should count. Standard cost-benefit analysis assumes that the marginal utility of income is the same for everybody and does not apply welfare weights to account for the fact that the marginal utility on income is smaller for a rich individual than a poor individual. Nyborg (2012) regards this as a serious weakness of cost-benefit analysis, since it will systematically favour the rich, because their willingness-to-pay will almost always be greater than willingness-to-pay among the poor. She also argues that there is no generally accepted method in economics for measuring the marginal utility of income. In chapter 11, a different point of view regarding how one can measure utility will be introduced. For the moment, suffice it to note that although there appears to be consensus among economists that willingness-to-pay is the only legitimate basis for cost-benefit analysis, there is no consensus about whether observed willingness-to-pay can always be regarded as a valid measure of welfare.

Although Viscusi clearly argues in favour of varying the value of a statistical life according to age and income, it is not obvious that he would accept any source of variation. He shows, for example, that recent Mexican immigrants to the United States run a higher fatality risk in their work than native Americans, but do not get any compensating wages, unlike native American workers. Although he refrains from using the word "discrimination" to describe this finding, it seems clear from the overall pattern of results

of studies of compensating wage differentials that the variation reflects both gender and racial discrimination.

To the extent that differences in income is the result of discriminatory practices on the labour market, it would seem ethically dubious to tolerate variation in VSL as a result of income differences. In short, whether or not one should allow the value of a statistical life to vary is an ethical judgment that cannot be fully justified simply by referring to the principles of welfare economics.

In summary, four points of view – four different recommendations – can be identified regarding the use of estimates of the value of a statistical life in cost-benefit analysis:

1. Use a uniform arithmetic mean value. The chief argument for doing so is that only a uniform value ensures an efficient allocation of resources between competing safety programmes (efficient with respect to safety effects).
2. Use a uniform harmonic mean value. This will ensure that one equalises marginal utility with respect to changes in fatality risk. If income distribution is positively skewed, the harmonic mean will be lower than the arithmetic mean.
3. Use a uniform median value. This is based on the median voter theorem of public choice theory and is the value that will have maximum public support. The median value will usually be considerably lower than the arithmetic mean.
4. Use values of a statistical life that vary according to, for example, age and income. These values will be closest to individual willingness-to-pay and will thus respect and reproduce individual preferences better than any mean value.

These recommendations are not consistent and will not produce identical results.

10.3 Multiple and inconsistent standards of consistency

In Chapter 2, it was briefly mentioned that the consistency argument economists made for the need for monetary valuation of safety takes many forms. It is therefore of interest to assess whether the different standards of consistency are internally consistent, or if some consistency standards contradict others. The consistency standards are:

1. Consistent (efficient) priority setting among safety measures
2. Consistency with market demand
3. Consistency with the Pareto principle
4. Consistency with majority preferences
5. Consistency with individual preferences
6. Consistency between ex ante and ex post valuations
7. Consistency with individual welfare

10.3.1 Consistent (efficient) priority setting

Consistency in policy priorities (1), i.e. efficient priority setting, was the main argument put forward to justify a monetary valuation of safety. It was argued that if priorities are set without the guidance of a uniform monetary valuation, they are likely to be inefficient, i.e. resources will not be spent in a way that maximises the number of fatalities prevented.

It is easy to show this. The only way to ensure efficient priorities is to apply a uniform value of preventing a fatality. In practice, much more is being spent to prevent fatalities in some sectors of society than in other sectors.

10.3.2 Consistency with market demand

Consistency with market demand (2) requires that safety should be provided up to the point where marginal benefits (in terms of WTP) equal marginal costs. Consistency with market demand can be obtained both by relying on the arithmetic mean value of a statistical life, or, if one wishes to segment the market, by relying on a set of arithmetic mean values of a statistical life customised to each market segment. It is essential to note that it is the arithmetic mean value that should be used, not any other estimator of the mean.

Consistency principles 1 and 2 are not necessarily in conflict with one another. They will not be contradictory as long as only a single uniform value of a statistical life, identical to the arithmetic mean, is used. However, once the value of a statistical life is allowed to vary, in order to match demand, no set of priorities can fulfil both principles of consistency. Elvik (2006) illustrates this in the context of policies designed to promote three different types of safety objectives, that require different approaches to the monetary valuation of safety. The three policy objectives were:

1. Seeking the maximum reduction of the total number of accident fatalities.

2. Reducing disparities in fatality rate between modes of transportation or groups of travellers.

3. Preventing disasters (reducing the frequency of accidents with multiple fatalities).

These policy objectives imply contradictory schemes for the monetary valuation of safety impacts. This can easily be shown by means of numerical examples.

Suppose that 100 people are each exposed to an identical fatality risk of 0.2. The expected number of fatalities is then $100 \cdot 0.2 = 20$. A program that reduces the expected number of fatalities by 50 percent, will prevent 10 fatalities. If each fatality prevented is assumed to be valued at 1 million dollars, the benefits are 10 million dollars.

Next suppose that among 100 individuals, 10 are exposed to a fatality risk of 0.8, 10 face a fatality risk of 0.4 and 80 face a fatality risk of 0.1. The expected number of fatalities is again 20 $[(10 \cdot 0.8) + (10 \cdot 0.4) + (80 \cdot 0.1)]$. If the objective is to reduce disparities in risk, a program that benefits the group exposed to the highest risk is more valuable than a program that benefits the group exposed to the lowest risk. To model this, suppose that reduction of the highest risk is valued at 3 million dollars per fatality prevented, reduction of the next-to-highest risk is valued at 2 million dollars per fatality prevented, and reduction of the lowest risk is valued at 1 million dollars per fatality prevented. A program that reduces the highest risk by 40 percent and the next-to-highest risk by 5 percent will then give a benefit of 10 million dollars, despite the fact that only 3.4 fatalities are now prevented $[(3.2 \cdot 3 = 9.6) + (0.2 \cdot 2 = 0.4)]$.

Finally, suppose that there is a risk of 0.1 for an accident involving 50 fatalities, a risk of 0.2 for an accident involving 20 fatalities, a risk of 0.5 for an accident involving 10 fatalities, and a risk of 0.3 for 20 accidents in each of which 1 person is killed. These risks are independent of each other. The expected number of fatalities is again 20 $[(0.1 \cdot 50) + (0.2 \cdot 20) + (0.5 \cdot 10) + (0.3 \cdot 1 \cdot 20)]$. An objective of preventing major accidents can be represented by applying a higher monetary valuation to the prevention of each fatality in major accidents than to the prevention of each fatality in minor accidents. Suppose, as an example that the prevention of each fatality in the largest accident is valued at 7 million dollars, the prevention of each fatality in the next-to-largest accident is valued at 4.5 million dollars, the prevention of each fatality in the third largest accident is valued at 3 million dollars, and the prevention of each fatality in each of the minor accidents is valued at 1 million dollars. Reducing the risk of the largest accident by 28 percent will reduce the

expected number of fatalities by 1.4. This gives the same benefit (10 million dollars) as in the first example, in which everybody faced the same risk and no major accidents were expected to occur.

In principle, one can reconcile the various policy objectives in monetary terms by adopting a complex set of monetary valuations. A baseline valuation would then represent the objective of reducing the total number of accident fatalities, and higher valuations would be introduced to reflect the additional policy objectives of minimising differences in risk and preventing major accidents. Adopting such a scheme of monetary valuation would make consistency principles 1 and 2 above contradictory and their simultaneous fulfilment impossible.

10.3.3 Consistency with the Pareto principle

Consistency with the Pareto principle (3) requires that those who gain from a measure can compensate those who lose and still have a net benefit. An example of a case where this is impossible, due to Blackorby and Donaldson (1986), was discussed in Chapter 7. It was concluded that the problem identified by Blackorby and Donaldson was unlikely to occur as far as the monetary valuation of safety is concerned, because the relevant level of risk is much lower than in the example given by Blackorby and Donaldson and because insurance can facilitate compensation that would be impossible if risks cannot be pooled. The situation discussed by Blackorby and Donaldson is therefore regarded as so improbable to occur that it is mainly of theoretical interest. A considerably greater difficulty can arise if the compensation criterion is interpreted in utility terms, and not simply in monetary terms, which is the common interpretation of the criterion today, i.e. a potential Pareto improvement is generally believed to exist when the monetary benefits of a measure exceed the monetary costs (Nyborg 2012). However, if account is taken of the marginal utility of money, it may no longer be the case that compensation in utility terms is possible even if monetary benefits exceed monetary costs. Elvik (2013B) gives an example of a case, based on real data, where the priority between two projects is altered depending on whether willingness-to-pay is applied in crude form, or adjusted to account for the difference in the marginal utility of money between a rich and a poor area of a town.

10.3.4 Consistency with majority preferences

Setting policy priorities according to willingness-to-pay (WTP) is sometimes presented as a democratic procedure. Indeed, one version of the contingent valuation method, direct choice, asks respondents to accept or reject bids that are offered, and thus, in a sense, mimics a referendum (or decisions made to buy or not buy something depending on its price). However, in stated preference surveys, it is nearly always found that mean WTP exceeds median WTP, often by a considerable amount. Hence, setting priorities according to mean WTP would result in expenditures that the majority of people regard as too large (principle 4, consistency with majority preferences). Median WTP will be supported by exactly half the people, opposed by the other half. This is the maximum support any amount can possibly enjoy. Any lower amount would be rejected by a majority as too low; any higher amount would be rejected by a majority as too high. Small wonder that cost-benefit analyses are often controversial and their results disputed. Adopting median WTP, rather than the arithmetic mean, is inconsistent with consistency principle 2, of matching the provision of a good to the demand for it.

10.3.5 Consistency with individual preferences

It is tempting to think that controversies about which value of WTP to use in cost-benefit analyses can be avoided by adopting, along the lines proposed by Sunstein (2004), a fully individuated WTP (consistency principle 5). However, this is impossible in practice unless there is a market for safety. The provision of safety by government cannot possibly be matched exactly to individual demand for safety, as expressed in individual willingness-to-pay. The principal reason for this is that many safety measures provided by government are public goods. It is not feasible to provide road lighting on a public road only to those who indicated that they are willing to pay enough for it to cover the cost of providing it. One cannot dim the lights whenever someone who did not indicate a willingness-to-pay travels on the road. Moreover, how would one know if that individual did not want to pay? It is only by hard-wiring the brains of everybody to some kind of lie detector, whose readings would be available to an omnipotent dictator, that one could have a remote hope of finding out whether people were really willing to pay for road lighting, or whether they just pretended to be willing to do so, but all the time hoped to be free-riders.

In principle, adopting an advanced system of road pricing can help in providing safety that matches the demand for it almost as closely as supply and demand are matched on perfect markets in equilibrium. Elvik (2010A) discussed whether road pricing could give government stronger incentives to set efficient priorities, i.e. priorities that provide what road users want at a minimum cost. He concluded that, in principle, it is technically feasible to introduce quite advanced systems of road pricing, but that even such systems would encounter problems in revealing the demand for public goods, like road lighting. He proposed that referenda (electronically) be held among users of a given road regarding their WTP for public goods and that the goods should be provided if a majority voted for providing them. Drivers voting against would then also be obliged to pay. This, for sure, is not undemocratic. Even those, in a municipal council, who voted against a speed limit of 30 km/h in residential areas must comply with the speed limit if it is introduced. Smokers must abstain from smoking in restaurants, and so on and so forth. It is by no means uncommon that minorities in a democracy must comply with laws they are opposed to.

For the moment, it is concluded that safety cannot be provided in a way that is consistent with individual preferences. Matching provision to individual preferences is impossible. One may at best determine a level of safety provision based on some kind of mean or median WTP, which would obviously not be identical to individual WTP except for those few individuals whose WTP happened to be close to the mean or median value.

10.3.6 Consistency between ex ante and ex post utility

Studies of willingness-to-pay are always based on an ex ante perspective, i.e. they deal with changes in risk before the risk materialises in terms of the occurrence of one of the events it refers to (consistency principle 6). As far as the risk of fatality is concerned, the ex ante perspective is the only one that is feasible. With respect to survivable injuries, however, it is legitimate to ask about the compensation that would be required to restore an individual to the initial level of utility, if nothing else to probe whether the ex post perspective gives the same answer as the ex ante perspective.

It has long been known that ex ante and ex post analyses do not give the same answers (Ulph 1982). In fact, these two perspectives must by necessity give different answers and there will, in that sense, never be consistency between ex ante and ex post estimates of WTP (ex post is perhaps better labelled WTA). The reason for this is very simple. Ex ante, risk exists only as a set of possible outcomes, each of which may occur with a certain probability. If the probability of an outcome is low, expected utility ex ante will be close to

the utility associated with the most probable outcome. This is particularly true for the risk of a fatal injury, which is much lower than the risk of a non-fatal injury. Hence, when a perfectly rational utility maximiser decides how much he or she is maximally willing to pay for reduced risk of a fatal injury, it is expected utility that guides the choice. Expected utility is the probability-weighted average of utility when alive and utility when dead. Utility when alive is likely to contribute well in excess of 99 percent to expected utility. Indeed, if utility when dead is zero (there are no bequest motives), utility when alive contributes 100 percent to expected utility.

If utility when dead is zero, ex post compensation is impossible. It is only with respect to a utility function that takes on positive values that one can estimate the compensation needed to restore pre-event utility. Since even slight injuries are likely to reduce the level of utility (even a minor injury, like a broken wrist, makes, for example, taking a shower a little more difficult, since you must cover the cast to prevent it from getting wet), it follows that ex post compensation will always be greater than ex ante WTP.

If one accepts the idea that utility depends on health state (Viscusi and Evans 1990), utility is, probably almost without exception, likely, ceteris paribus (i.e. for a given income and keeping all other factors influencing utility constant), to be lower when an individual has a health problem than when the individual does not have such a problem. This implies that ex post compensation for the health problem will always exceed ex ante WTP to eliminate the problem.

A choice must therefore be made between the ex ante and ex post perspectives. Utility theory, as developed within neoclassical economic theory, refers to ex ante utility, i.e. it is a decision utility which determines the choices made between options with different consequences that have not yet occurred at the time of decision (Loewenstein and Ubel 2008).

10.3.7 Consistency with individual welfare

The final criterion of consistency to be discussed has been labelled consistency with individual welfare. This criterion might strike some readers as a superfluous addition to the criteria already discussed. Surely, one could argue, if the provision of safety is based on the (aggregate) willingness-to-pay for it, it will by definition be consistent with individual welfare. Within the framework of neoclassical economic theory, this is true by definition. If 100,000 people have stated that, collectively, they are willing to pay 8 million dollars to reduce the number of deaths by one, society will improve the (collective) welfare of these individuals by spending up to that amount on safety measures that will reduce the number of deaths by one. Welfare will then have been improved by reducing fatality risk, at no net expense (i.e. monetary benefits are at least as great as expenditures).

This reasoning is, however, entirely too simplistic. If one assumes that individuals initially maximise utility, how can one then account for the fact that they are willing to spend anything at all on safety? This may not be as mysterious as it perhaps sounds. Since, at least for now, there is no market for safety, individuals are not able to buy safety unless someone asks them to imagine a hypothetical market that permits them to buy safety. However, imagine the following scenario, which is fictitious but still not implausible:

Interviewer: You have just told me that you are willing to spend 2000 on this safety device. Please tell me what you would now spend less on. Where would you take the money?

Respondent: I do not understand. Spend less on? Surely, I can go on spending just as much as I do today on all the things I enjoy.

Interviewer: No, you cannot. You do not have those 2000 any longer. You just spent them on safety. So, you must cut back on other things. Please tell me what you would cut back on.

Respondent: Well, the hell no! If it is like that, then I do not want to pay anything for that stupid safety device.

Other respondents might be a bit more polite and forthcoming and offer a noncommittal answer like this:

Respondent: Before getting specific about that, I would have to review my household budget to see where I can most easily make cuts. I guess things like holidays can be made cheaper, so that might be where I would cut back.

In short, as long as payment remains hypothetical, it remains unreal. Even after it has been made, it is by no means certain that it will actually improve welfare. The experience could turn out to be different from what you thought. You may conclude that it was not worth the money after all. Your welfare has then not been improved.

It is completely unrealistic to think that people can predict perfectly the utility they actually will experience. Therefore, to ensure that individual welfare has actually been improved, actual and hypothetical monetary valuations of identical goods should coincide both before and after the goods have been purchased. Only if there is such a consistency can it legitimately be claimed that individual welfare has been improved.

10.3.8 A self-contradictory hard core?

An ideal of helping to develop more rational public policy for improving safety underlies studies of willingness-to-pay. It is therefore useful to try to clarify as far as possible the exact meaning of rationality as far as the provision of safety by means of public policy is concerned. It turns out that many dimensions of rationality can be regarded as relevant. Each of these dimensions can be stated in terms of a principle of consistency. The question is whether these principles of consistency are consistent among themselves, i.e. can all be satisfied, or contradictory, i.e. an action satisfying one of the principles will violate another.

It has been found that the principles of consistency are to some extent contradictory. Table 10.1 summarises the main findings of the discussion in the preceding sections.

Table 10.1: Principles of consistency for safety policy and the relationship between them

Principle of consistency	Main implication	Practicality
1: Consistency in priority setting	Use a uniform (arithmetic mean) value of a statistical life	Easy to implement; inconsistent with principle 2
2: Provide safety according to the demand for it	Adopt different values of a statistical life according to variation in willingness-to-pay	Complicated to implement; which sources of variation in valuations are legitimate, which are not?
3: Promote potential Pareto-improvements	Equate marginal benefits to marginal costs; the net surplus of benefits can compensate losers	Compensation in monetary terms may not be identical to compensation in utility terms
4: Be consistent with majority preferences	Median willingness-to-pay should be adopted	Easy to implement; inconsistent with principles 1 and 2
5: Be consistent with individual preferences	Adopt a fully individuated value of life saving	Cannot be implemented in practice
6: Ensure consistency between ex ante and ex post valuations	Consistency prevents reversing decisions as a result of regrets generated by actual outcomes	Ex ante and ex post valuations are never consistent
7: Promote individual welfare	Intentions and actions should be consistent	Experienced utility should not deviate much from decision utility

If the provision of safety is based on a uniform value of a statistical life, as is by far the most common practice today, it will in most cases: (1) Not account for variation in the demand for safety according to, for example, age and income; (2) Not necessarily lead to potential Pareto-improvements assessed in utility terms; (3) Not necessarily be consistent with majority preferences; (4) Not at all be consistent with individual preferences on a one-to-one bias (i.e., each individual gets exactly the amount of safety he or she is willing to pay for); (5) Not lead to outcomes for which losses in utility can be fully compensated by money set aside in a fund based on ex ante willingness-to-pay; on the contrary restoring initial utility will in most cases cost more than an ex ante fund of money can pay for; (6) Not necessarily promote individual welfare, since stated willingness-to-pay may deviate from actual willingness-to-pay, and actual payments may not bring the satisfaction one thought when the payments were made.

It seems clear that no schedule of monetary valuation can satisfy all the aspects of policy priorities that are regarded as desirable. Such is the current state of economic theory as far as valuation of safety is concerned.

10.4 Is an alternative hard core emerging?

The hard core of valuation research as a scientific research programme consists of a set of assumptions economists make when developing theory, of which the most essential is that individuals are rational utility maximisers. However, as shown in previous chapters, utility is an elusive concept. Utility functions can be specified in many ways that have quite different implications. Ultimately, therefore, the value of assuming utility maximising as a positive heuristic is limited. Rationality does not get you very far, as Kenneth Arrow (quoted in Thaler 2015) once remarked. Today, as shown in Chapter 5, there are so many utility models about willingness-to-pay that almost any result can be explained in terms of one or more of these models. Falsification has been rendered almost impossible. Almost any finding can be accounted for theoretically.

More than seventy years ago, Herbert Simon (1943) pointed out a number of basic limitations of the neoclassic model. This became the beginning of the field of behavioural economics, which has grown rapidly in the past 25-30 years (Thaler 1994, Kahneman and Tversky 2000, Camerer 2004, Munro 2008, Thaler and Sunstein 2008, Cartwright 2011, Thaler 2015). Behavioural economics brings insights from psychology into economics and rejects the classic model of a perfectly rational utility maximiser. Has behavioural economics developed to the point where it can form an alternative hard core for valuation research?

A number of papers have dealt with this question, and in this section, the key findings of these papers will be reviewed. The first paper (Berg 2003) takes as its starting point that behavioural economics is descriptive and that it does not challenge the classic model as the proper normative ideal. It then goes on to ask whether this purely descriptive interpretation of behavioural economics is correct.

Six techniques of investigation that are used in behavioural economics are discussed, and for each of these the paper shows that it leads to a policy recommendation that is ruled out by the standard neoclassic technique (i.e. postulating a utility function with certain properties and then deriving its empirical implications). It is argued that techniques of analysis are idiosyncratic, suggesting that recommendations based on them may not have general validity. It is concluded that policy recommendations should be developed by combining several methods of analysis. Apart from this very general conclusion, the paper

is extremely vague and no specific guidelines can be extracted from it about a normative application of behavioural economics.

Two papers by Robinson and Hammitt (2011A, 2011B) are more concrete. They note that (2011A:7):

"If individuals' behaviour suggests preferences that appear irrational, unstable, or contrary to their self-interest, choosing a policy to satisfy those preferences may not improve social well being."

They nevertheless think the temptation to disregard preferences and adopt paternalism should be resisted (2011A:8):

"While perfection in decision making may be impossible, our hope is to at least attempt to avoid paternalistic views of what individuals "should" prefer, deferring to the preferences that individuals express when provided opportunities for contemplation and learning."

Robinson and Hammitt (2011A) discuss the monetary valuation of changes in risk fairly extensively and note the widespread finding that individuals are insensitive to the size of changes in risk. They label this "troubling". They argue that valuation studies should be well-designed to help separate values that reflect misinformation or misunderstanding from values that reflect more stable and carefully considered preferences. This line of reasoning assumes that the latter type of preferences actually exist, which is not clear as far as changes in low levels of risk are concerned. They conclude that values should not be rejected unless the study is of poor quality. Yet, as shown in Chapter 8, even studies that are regarded as methodologically rigorous fail to eliminate such anomalies as insensitivity to scope and lexicographic and inconsistent choices. This suggests that any preferences, if they exist, do not fulfil minimal standards of consistency, and that what people express when answering valuation surveys are not preferences, but attitudes at a highly general level ("Road safety is a good thing, I cannot come across as not supporting it").

Robinson and Hammitt conclude that when values are uncertain (which is always), sensitivity analyses should be performed. They also refer to the ongoing discussion about how best to design stated preference studies. Their conclusions regarding this are quite vague, merely stating that there are "abundant lessons for researchers", without describing what these lessons are. Their conclusion (2011A:36) is that:

"Because behavioural economics is a large and rapidly growing field, the significance or pervasiveness of many of its findings are not yet clear and these findings have not been combined into a widely-accepted model that supplements or supplants the standard economic framework."

This is obviously correct. The neoclassic model remains unchallenged as a normative framework. It is difficult to see how behavioural economics could replace neoclassic economics as the foundation for prescriptions (policy advice), since a major topic of behavioural economics has been systematic deviations from rationality. It is difficult to argue that preferences or choices that systematically violate even the most basic standards of rationality can be granted a normative status. Perhaps the most striking illustration of the complete lack of rationality found in this literature was presented by Tversky and Kahneman in 1986 (Tversky and Kahneman 1986). It is worth quoting at length.

A contagious disease (say, a bad flu) is about to hit the country. 600 people are expected to die if no preventive action is taken. A vaccine which is known to protect people (Vaccine A) exists, but there is only enough in store to vaccinate one third of the population. There is not enough time to produce more of vaccine A. There is another vaccine (Vaccine B) which has more uncertain effects, but is estimated to protect one third of those who are vaccinated. There is enough of this vaccine in store to vaccinate the entire population.

Having been given this background information, subjects were asked to choose between A and B:

188

A: If A is chosen, 200 people will be saved for certain (Vaccine A is known to be 100 percent protective).

B: If B is chosen, there is a probability of 1/3 that 600 people will be saved (The expected number of people saved is 200).

When given this choice, most subjects chose A. Please note that the number of people expected to be saved is the same in A and B (200); the alternatives differ only with respect to the risk involved. Following this, the same subjects were asked to choose between A and B:

A: If A is chosen, 400 people will die for certain (Those who do not get vaccine A).

B: If B is chosen, there is a probability of 2/3 that 600 people will die (Vaccine B protects only 1/3 of those who get it).

Please note that these alternatives are identical to those presented in the first choice task. It is only the description of the alternatives that has been changed. However, in the second round, a majority chose B.

This pattern of choices cannot be rationalised by referring to differences in the expected number of people saved. That number is the same, 200, in alternatives A and B, irrespective of how they are described. Moreover, the pattern of choices cannot be rationalised by pointing to the fact that in A the outcome is known for certain, whereas in B the outcome is probabilistic. If an individual is risk-averse, he or she should go for alternative A in both rounds of choices.

In short, the dominant pattern of choices seems to be the result of the differences in the description of the alternatives only, which is blatantly irrational by any standard of rationality. If a majority of people make blatantly irrational choices between these very simple options, one shudders at the thought of how rational they can be expected to be when faced with more complex choices.

Paternalism nevertheless remains deeply unattractive. Before discussing whether a way can be found to maintain consumer sovereignty and incorporate individual preferences and values as the basis of policy priorities, the second paper by Robinson and Hammitt (2011B) will be discussed.

There is some overlap between the two papers. Robinson and Hammitt (2011B:1412) state that:

"We generally take the perspective that analysts should avoid making judgments about whether values are "rational" or "irrational", but should make every effort to ensure that studies are designed to elicit well-informed, thoughtful preferences."

One wonders how to design a study suitable for studying something that does not exist. The most reasonable conclusion based on the results of valuation studies is that the preferences these studies were designed to elicit simply do not exist. The results of these studies must therefore be rejected, since nobody has argued that policy priorities ought to be based on "preferences" that are the result of framing, starting point bias, range bias, hypothetical bias, or, if inferred from choices, are lexicographic or inconsistent. If, by contrast, one assumes that "well-informed" and "thoughtful" preferences nevertheless do exist, but have not been successfully elicited by the studies made so far, the conclusion is the same: These studies must be rejected because they have not been able to properly uncover and describe the phenomenon they were intended to describe.

Elvebakk (2015) discusses the role of paternalism in road safety policy. She notes, correctly, that in many cases, the state does not trust individuals to make the right decisions to keep themselves safe. Thus, laws requiring the use of seat belts and crash helmets are clearly paternalistic. Even speed limits could be regarded as paternalistic, but the case for speed

limits can be made by showing that a free choice of speed, i.e. having no speed limits, would not lead to Pareto-optimal outcomes at a societal level (Elvik 2010B). Thus, one could well argue that the speed choices made by each driver is subjectively rational (i.e. each driver chooses the speed he or she thinks is best), but even such an assumption is dubious. In dense traffic, many drivers will be "forced" to adopt a different speed from the one they would have chosen freely. Moreover, the choice of speed is likely to have external effects that drivers do not consider, such as effects on traffic noise and air pollution. There is, in a sense, a "market failure" that justifies the use of speed limits to guide drivers towards choosing speeds that will minimise the negative externalities.

Yet, the tradition of paternalism in road safety policy is so strongly entrenched that introducing a different approach, based on respect for citizen sovereignty is widely viewed as alien. The concept of willingness-to-pay is routinely misunderstood (see Elvik 2016D for an example of common misunderstandings and an attempt to sort them out). It is routinely argued that it is unethical to assign a monetary value to the prevention of road accident fatalities.

Obviously, these remarks are in no way intended to suggest that an economic approach is without problems. It is rife with them.

10.5 Do revealed and stated preferences agree?

It was noted in Chapter 7 that the insensitivity to the size of changes in risk found in stated preference studies results in an inverse relationship between the size of a risk reduction and the value of a statistical life: the smaller the risk reduction, the larger the value of a statistical life. Can a similar pattern be found in government decisions about regulation of risk? Does government spend disproportionately to reduce low risks compared to expenditures on reducing high risks?

Morrall (2003) has reviewed regulatory decisions by the federal government in the United States. For some of these regulations, both the initial level of risk, the number of fatalities expected to be prevented and the societal cost of the regulation are known. For these regulations, all of which have been implemented, it is possible to study the relationship between the level of risk and the cost of the regulation. Cost is stated as cost per life saved.

Figure 10.1 shows the relationship between the initial level of risk and the cost per fatality prevented. A curve has been fitted to the data points, but it fits quite badly. Most data points in the left half of the diagram are located below the fitted curve; most data points in the right half of the diagram are located above the fitted curve. Data points display wide dispersion and both axes are plotted on a log scale to better indicate the spread of the data points.

A generous interpretation is that there is a positive relationship: The higher the initial level of risk, the higher the expenditure to reduce the risk. The main impression is, however, that there is wide dispersion and nothing resembling a uniform limit on expenditures to prevent fatalities. Thus, the figure confirms the conjecture made by many economists that if no uniform value of a statistical life is applied, the amount of resources used to prevent unintentional deaths is likely to vary from case to case. These variations are not likely to reflect variations in preferences for risk regulation.
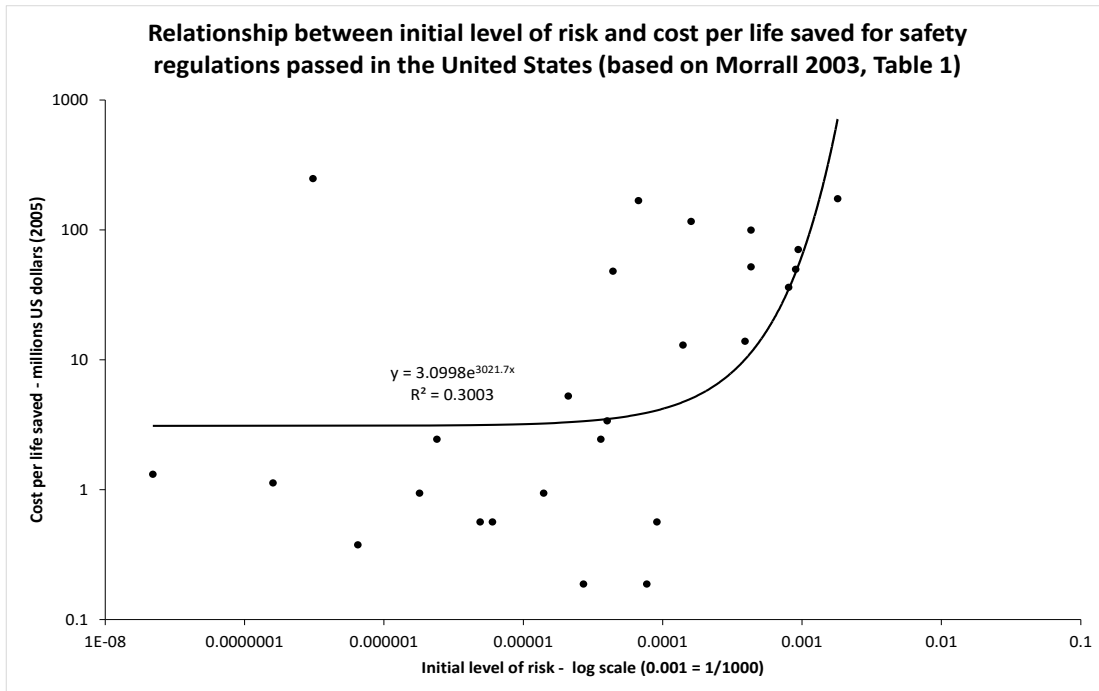
Figure 10.1: Relationship between initial level of risk and cost per fatality prevented for regulations implemented in the United States. Based on Morrall 2003

10.6 Conclusions

The discussion in this chapter leads to the following main conclusions:

1. The realisation among many researchers that the wide dispersion in estimates of the value of a statistical life is unlikely to be greatly reduced by continuing to do valuation studies, has lead to a discussion about differentiating, and, in the limit, fully individuating the value of a statistical life.

2. Some prominent economists working in the field have started to discuss differentiating the value of a statistical life. Some leading economists argue that the value of a statistical life should vary according to income. This means that the lives of rich people will be valued more highly than the lives of poor people.

3. Attempts have been made to develop an alternative normative foundation for welfare economics based on behavioural economics. These attempts have so far not been successful. The hard core of neo-classical economic theory remains the dominant foundation for welfare economics.

4. There are multiple standards of consistency in economic theory. No schedule of economic valuation of changes in risk can satisfy all the consistency standards. These standards are therefore, to some extent, self-contradictory.

11 Exploring alternative approaches

11.1 Assessing the current status of valuation research as a scientific research programme

Except for a few studies made in the last half of the 1970s, all empirical studies designed to obtain monetary valuations of transport safety have been made after 1980. By now, these studies have produced hundreds of estimates of the value of a statistical life. Not all of the studies have been reviewed in this report, but the sample of studies discussed gives a fair impression both of the range of approaches used and the variation of the findings.

When the valuation of safety was launched as a scientific research programme, the first question researchers needed to answer was how to obtain the valuations. Was it feasible at all, given the fact that no market for safety exists? Two approaches emerged: The contingent valuation method and the compensating wage differentials approach. These approaches were dominant until about 2000. The contingent valuation approach was first used by Jones-Lee from about 1980. Following his pioneering study, a number of replications quickly followed. This period was characterised by optimism and the feeling that progress was made. By 1987, Hoehn and Randall (1987) concluded that the results of contingent valuation studies could be applied in cost-benefit analyses, if the valuations had been obtained in a way that encouraged respondents to tell the truth. Not long after, Mitchell and Carson (1989) published a comprehensive textbook on the contingent valuation method.

These researchers worked in the United States, where the contingent valuation method had been controversial from the start. In the United States, the contingent valuation method was predominantly used to value environmental goods, not safety. In Europe and New Zealand, on the other hand, the contingent valuation method was applied to value transport safety. Researchers in the United States relied, almost exclusively, on the study of compensating wage differentials to value safety.

From around the middle of the 1990s, criticism of the contingent valuation method started to grow. In 1993, Hausman (1993) edited a book criticising the contingent valuation method. By the late 1990s, several weaknesses associated with the method had been uncovered and found to be widespread in empirical studies, including:

1. Hypothetical bias (values were overstated because no real payment was involved),
2. Strategic response bias (free riding leading to understating the value of public goods),
3. Starting point bias (an initial bid exerting undue influence on responses),
4. Payment range bias (the range of amounts shown on a payment card influencing responses),
5. Payment vehicle bias (resistance to increased taxes leading to lower valuations),
6. Embedding effects (a good valued lower when part of package of goods than when valued alone),
7. Insensitivity to scope (valuations not responsive to different amounts of a good),
8. Loss aversion (willingness to accept often much higher than willingness-to-pay).

Some practitioners of the method, in particular Jones-Lee, came to reject their own previous research and tested new approaches, although these were still, in a wide sense of the term, contingent valuation. However, after about 2000, a growing share of stated preference studies have relied on the stated choice approach. In this approach, respondents are not asked to state their willingness-to-pay, but are asked to make choices between alternatives described in terms of various attributes, such as cost, safety and time spent. It quickly became clear that this approach to valuation had its own problems. It was subject to framing effects, meaning that different ways of presenting choices influenced them even if there were no real differences (only the descriptions were different). It was, in a sense, subject to manipulation in that researchers designed both the options and their attributes and thus, effectively, determined the range of values that could be obtained. There was, in many studies, a high share of lexicographic and inconsistent choices.

Stated preferences studies have, as noted in the meta-analyses of Lindhjem et al., produced an extremely wide range of values of a statistical life.

Meanwhile, Dorman (1996) launched a broad attack on studies of compensating wage differentials in a book published in 1996. His aim was clearly to persuade researchers to stop doing these studies. He did not succeed and in the years that followed, studies of compensating wage differentials continued more or less unaffected by his strong criticism. All studies of compensating wage differentials rely on the assumption that worker are perfectly rational in their choice of work. Unless this assumption is valid, there is no reason to believe in the studies. Many years ago, a seminar on the monetary valuation of road safety was held in Denmark (Christensen 1988). As part of the seminar, Danish health economist Kjeld Møller Petersen presented an assessment of the compensating wage differentials approach. He noted (1988:21; my translation):

"A problem arises because, although the mathematical model of wage formation is elegant, it does not identify the functional form of the wage equation. As an example, the theoretical model does not state whether one should estimate the natural logarithm of wages or wages in natural units. Since the theory does not tell the analyst which functional form to use, it is tempting to engage in data mining and model testing until one gets plausible findings. The question is whether this method can be defended statistically."

He goes on to describe a Danish wage-risk study based on data for the city of Odense. The study relied on quite detailed data about the risk facing workers and 244 different estimates of risk could be assigned to workers. In addition to risk, data were available on age, gender, health state, educational level, vocational training, and so on. He continued (page 21; my translation):

"Now, different regression models were tested, and it was found that only a double logarithmic specification (i.e. transforming both the left (dependent variable) and right hand (independent variables) to natural logarithms) produced coefficients with values between 0 and 1, which is the range where coefficients should be to make sense from a theoretical point of view. It was found that different specifications of the wage equation produced very different results. It is, accordingly, very dubious that the researcher has such a great freedom of choice.

The "most satisfactory" equation found that the value of reducing the number of fatal workplace accidents by 1 per 1,000 workers is 314,000 Danish crowns. Had another regression equation been preferred, the value might just as well have been zero. … Model estimation is easy. Model selection is difficult."

After his presentation, Møller Petersen was interviewed. Here is an excerpt from the interview (page 24; my translation):

"Interviewer: Are you able to formally make the distinction between – on the one hand – trying different regression models and selecting the one you like best, and – on the other hand – pure fraud?

Møller Petersen: No. I have the impression that it is a matter of psychology: You stop testing equations when you have found one that is consistent with what you expected to find."

It is perhaps too harsh to label compensating wage differentials studies as fraud. It may be fairer to say that they are strongly influenced by theory and that its proponents believe it is fair to assume that workers make rational choices between occupations involving different levels of risk. The vast psychological literature on biases and heuristics and on how easily choices can be framed casts doubt on the rationality assumption. Most workers face very low levels of risk in their jobs (Elvik 2005), so low, in fact, that one may doubt that considerations of risk enter their minds at all when they choose occupation. Be that as it may; to be valid, studies of compensating wage differentials should provide evidence that the worker actually did consider risk as a relevant aspect when choosing occupation. Only by providing such evidence can the argument made by Møller Petersen about data mining and confirmation bias (i.e. preference for the model that supports theory) be effectively refuted. It is noteworthy that almost no study of compensation wage differentials provides such evidence and in the few studies providing some data (e.g. Hersch and Viscusi 1990), the data are crude and not always validated (by showing that perceived risk is close to actual risk).

Some economists tend to dismiss the psychometric literature since it is based on experiments only. This point of view is remarkable and almost unique in empirical science. In most other fields of empirical research, randomised controlled trials are regarded as the gold standard for research, something to strive for, not something to be dismissed as an inferior form of research. Economics is the only empirical science that regards evidence from experiments as worth less than evidence from more or less poorly controlled observational studies of market behaviour. These are just traps set up by psychologists, it is argued. They do not apply to real market behaviour. The market will punish those who deviate from rationality in a way the artificial trials created by psychologists will not do, it is argued.

It would be a digression from the main topic of this report to discuss these points of view at length. Suffice it to note that consumers may persist in suboptimal behaviour simply out of ignorance or habit, or because they settle for what is "good enough" rather than what gets the best value for money. There have been experiments in real markets, with money at stake, showing that consumer behaviour is influenced by arbitrary anchoring effects (Ariely, Loewenstein and Prelec 2003). Ariely et al. conclude (page 73):

1. The market behaviour induced by arbitrary anchors cannot be interpreted as a rational response to information (normatively, the two last digits of your social security number should not influence your willingness-to-pay for a bottle of good wine; but it does influence willingness-to-pay).
2. The behaviour (the anchoring effect) does not decrease as a result of experience with a good.
3. The behaviour is not necessarily reduced by market forces.
4. The behaviour is not unique to cash prices.

Ariely et al. conclude that market data need not reveal true consumer preferences in any normatively significant sense of the term. This conclusion is likely to hold with respect to the choice of work as well. In particular:

1. The choice of occupation is made very rarely, often only once in life, and the opportunities for learning from repeated choices are highly limited.
2. The choice of occupation is often highly constrained, by, for example, education and geography. The labour market is highly segmented.
3. There is no way an individual can have perfect foresight about how well he or she will like a job; decision utility may not be the same as experienced utility (Loewenstein, O'Donoghue and Rabin 2003).

4. Occupational risks may be partly endogenous, meaning that workers may to some extent control them; partly actions to mitigate risks may be collective goods, the demand for which cannot be revealed by individual behaviour (Sen 1973).
5. Workers who deliberately choose high-risk occupations are likely to have different attitudes to risk than the rest of the population and the valuations revealed by their choices cannot be generalised to the general population.

In summary: The results of stated preference studies contain a number of anomalies showing that respondents do not respond to the valuation tasks with the degree of rationality assumed by economic theory. The fact that the values are all over the place suggests that there are no preferences to elicit. People simply pick a number from thin air or are unduly influenced by initial bids or other elements of the study design that give hints about what a reasonable valuation would be. Since people do not know what a reasonable valuation would be, it is no surprise that they look for cues and are heavily influenced by them. Indeed, in the direct choice version of contingent valuation, the analyst offers an answer and respondents only have to say yes or no to it. Thus, analysts exert a very large influence on the results of valuation studies.

Revealed preference studies, on the other hand, rely on an assumption of rationality which is normally not tested as part of the study. Indeed, a wage equation might not have a solution, or have many solutions, if utility maximisation is not assumed. It is mathematically necessary to assume utility maximisation to be able to (uniquely) solve the equation. However, the widespread violations of rationality found in stated preference studies cast doubt on the validity of the rationality assumption made in revealed preference studies. It is unconvincing to invoke the argument that the market punishes irrational behaviour. A worker may well come to regret his or her choice of occupation; in that sense a kind of "punishment" is imposed. But that only shows that a choice based on decision utility may not be identical to one based on experienced utility. A perfectly rational individual is assumed to accurately predict experienced utility at the time the decision is made. Unless this assumption is made, preferences become endogenous (i.e. influenced by the choice made), meaning that they change once the outcome of the decision becomes known, leading to an instability of preferences which undermines the notion of maximising utility.

If people cannot behave rationally when faced by the highly simplified choices offered them in stated preference studies, there is even less reason to believe that the vastly more complex choices made in "real life" are rational. Thus, a stated choice task between two alternatives, each with three attributes, induces lexicographic and inconsistent choices on a massive scale. What then, about the choice of, first, an education, then an occupation suitable for the education, involving maybe, say, the comparison of three occupations with respect to ten attributes?

When anomalies develop in a scientific research programme, there are two directions research can take to deal with the anomalies. The first is to pursue methodological research, to try to develop better methods for studying the phenomenon of interest. This line of research is based on the assumption that the phenomenon studied does exist and has the characteristics suggested by the hard core of the programme; it is only the measuring instruments that are not good enough. This line of research has been vigorously pursued in the valuation research programme. It has reduced some of the anomalies, but not quite eliminated them. As an example, hypothetical bias in contingent valuation studies can be reduced, but not fully removed.

The other line of attack on anomalies consists of developing new theory aiming to show that the anomalies are not really anomalies after all, but are consistent with predictions derived from the assumptions forming the hard core. This line of research has also been

pursued in the valuation research programme. Thus, utility functions have now been proposed that are capable of accounting for almost all apparent anomalies in the research programme. These functions have been reviewed earlier in the report.

The versatility of utility functions that may now be invoked to explain a finding basically means that almost all findings make sense from a theoretical point of view. Although the researchers proposing the various utility functions may have had the noblest intentions, such as eliminating an anomaly by showing that it is actually consistent with the hard core of the research programme (neoclassical utility maximisation), the proliferation of utility models has had the unintended consequence that the current set of hypotheses about willingness-to-pay is so versatile that it almost forms an immunising stratagem. This means that whatever you find, for example that: (1) WTP is unrelated to the size of risk change; (2) WTP increases in proportion to the size of the risk change; or (3) WTP increases, but much less that proportionally to the size of the risk change – all these results are consistent with at least one of the utility models constituting the theory of WTP. In short: No result can be interpreted as falsifying a hypotheses in the protective belt. Thus, the hard core remains immune to criticism.

One might think that it is every researcher's dream to develop a theory which cannot be falsified. It is, however, more like a curse than a blessing. When a theory becomes immune to falsification, it ceases to have the essential function of a theory in empirical science: to identify those results that make sense (support the theory) and those that do not (refute the theory). If every result supports the theory, it no longer makes sense to ask whether a certain result was produced by a bad research method or a faulty measuring instrument. If, say, insensitivity to scope, which was long regarded as an anomaly, really is to be expected, there is no reason to try to develop better methods for eliciting preferences so as to avoid insensitivity to scope. Thus, developing an immunising stratagem does not strengthen a theory, but puts it on a path towards its own destruction.

This was clearly seen and vividly described by Percival (1997). He noted (page 126):

"Drastic revisions of a theory through the use of an immunizing stratagem are rare, for they are too obvious and unconvincing. The revisions are more often of a marginal nature."

This description fits well to the history of valuation research. The various utility functions have been introduced slowly, one at a time, in seemingly unrelated papers by different authors. Only when the implications of all the different utility models in WTP-theory are put together does one realise that they come close to forming an immunising stratagem. Hands (1993) makes related observations. In discussing whether Popper's falsificationism or Lakatos' methodology of scientific research programmes best fits economic methodology, he notes (quoted from reprint in Hausman (Ed) 2008:190, 192, 193):

"Falsificationism is seldom if ever practised in economics. … The qualitative comparative statics technique used in economics makes severe testing very difficult and cheap corroborational success "too easy". In economics it is very often the case that the strongest available prediction is a qualitative comparative statics result which only specifies that the variable in question increases, decreases or remains the same. Since having the correct sign is much easier than having both the correct sign and magnitude, an emphasis on such qualitative prediction generates theories which are low in empirical content, have few potential falsifiers, and are difficult if not impossible to test severely. …Strict adherence to falsificationist norms would virtually destroy all existing economic theory …"

He goes on to add that (2008:195, 196):

"For Lakatos progress comes from corroboration not falsification of novel facts. …A philosophical programme such as Popperian falsificationism which requires practitioners to be willing to give up almost any part of their research programme at any time will not provide as adequate a guide for economists as Lakatos' methodology which allows for such pervasive hard cores. This economic preference for Lakatos over

The circle has now been closed. The empirical study of the monetary valuation of safety was motivated by the observation that if one did not explicitly put a monetary value on human life, decisions about safety policies would be inconsistent and wasteful. One might, in one case, spend 100,000 to save a life, in another case spend 145,000,000 to save a life. It was obvious that if a uniform monetary value was applied, one could find the allocation of expenditures that would maximise the number of lives saved.

It did not take many years before this idea started to unravel both as a result of refinement of theory – principally in the form of the many individual utility functions proposed – and as a result of the enormous variation in empirical estimates of the value of a statistical life, which called for an explanation. Quite a lot of the variation in empirical estimates of the value of a statistical life was attributable to anomalies, i.e. factors that in theory ought not to influence the results, or at least have a considerably smaller impact than they did. Various methodological innovations were developed in order to eliminate the anomalies, but none of them were entirely successful. At best, the anomalies were reduced, but not eliminated. The rescue came in the form of further developing theory that made it possible to re-interpret the anomalies as being quite normal findings after all. WTA much greater than WTP? No problem, this is exactly as we would expect. WTP insensitive to the size of the risk reduction? No problem, there are utility functions – consistent with the hard core – that would predict exactly that. WTP lower for safety as a public good than as a private good? Makes perfect sense, you only need to remember what Samuelson (1954) taught us about it. WTP higher for safety as a public good than as a private good? That makes perfect sense if the utility function is strongly altruistic.

Plausible or not, the various utility functions nowadays predict exactly what we observe: An enormous variation in the value of a statistical life. The search for a single uniform value seems to have been given up. The key issue now is what amount of variation one should tolerate, and by what criteria (age, income, and so on) the value of a statistical life should be allowed to vary.

A major problem with the utility functions that have been proposed by theorists is that they cannot be observed. It is not straightforward to uncover whether any of the utility functions exist, at the very least in the minds of people. Most likely they do not. The vast body of psychological research summarised by, for example, Connolly, Arkes and Hammond (2000), Kahneman and Tversky (2000), Gilovich, Griffin and Kahneman (2002), suggests that the well-ordered preferences which can be modelled by means of utility functions simply do not exist. This also explains why the anomalies persist despite attempts to device methods that were expected to eliminate them. It is, so to speak, difficult to device a method which is suitable for studying a phenomenon that does not exist.

Can the values of a statistical life that have been estimated in empirical studies be taken seriously? Probably not. There several reasons for rejecting these values.

In the first place, all values based on stated preference studies are to some extent influenced by known sources of bias, for example hypothetical bias or bias attributable to inconsistent choices. Not a single study has been able to entirely avoid these biases. Moreover, the biases appear to inflate the values. Faced with this, one might ask: How about selecting estimates based on respondents who can be shown to have solved the valuation tasks in a manner that satisfies some minimal requirements of rationality? In some studies, such a selection is feasible, but in many studies it is not, because the study did not include any test of respondent rationality. Selecting answers this way would involve

discarding the bulk of the data in stated preference studies (in most cases probably in the order of 50-80 percent). Doing so sits uneasily with the strong calls made by both Schelling (1968) and Mishan (1971) in favour of respecting consumer sovereignty even when consumers are not perfectly informed or perfectly rational.

In the second place, very many values based on stated preference studies are likely to be exaggerated and not representative of typical (median voter) preferences. Mean values, which are the appropriate ones to use according to economic theory, are often considerably higher than median values, see the survey of meta-analyses in Chapter 9. Clearly, the majority of the population would regard these values as too high and not representative of their willingness-to-pay.

In the third place, the exaggeration of values is reinforced by publication bias. The analyses made in Chapter 9 indicated massive publication bias. These analyses indicated that the published values should be reduced by about 60-70 percent to adjust for publication bias. It is worth noting that publication bias affects both the mean and median values of a statistical life and therefore indicate that even the median values, which tend to be much lower than the mean values, should be drastically reduced.

In the fourth place, revealed preference studies, predominantly studies of compensating wage differentials, fail to show that the choices studied were rational and thus reveal preferences. Two characteristics of risk make it unlikely that, for example, choice of occupation can be interpreted as revealing preferences regarding the reduction of risk. The first characteristic is that many risks are endogenous, i.e. they are partly under worker control and therefore not a fixed parameter (a given value) that can be inserted into a wage equation. The second characteristic is that measures taken by the employer to reduce risk may be collective (public) goods. Preferences for the provision of public goods cannot be revealed by market behaviour, since markets do not normally provide public goods in sufficient amounts to match the (latent) demand for them. Yet, even if these complications are disregarded, and behaviour is assumed to be rational, it is highly doubtful that the emerging values are representative of valuations of the general population. Workers self-select into high-risk occupations, and accept compensation – a willingness-to-accept value – for exposing themselves to risks that perhaps the majority of people would not run for any finite amount of compensation. Moreover, the revealed preference studies are, just as the stated preference studies, characterised by a very wide range of values and the presence of strong publication bias.

In the fifth place, if one decided to trust all published values of a statistical life, the huge range of such values would make it very difficult to select one or a few of them for use in cost-benefit analyses. If the assumption is made that all values can be trusted, one cannot choose one or a few of them for use in cost-benefit analysis by arguing that these values are somehow "more trustworthy" than other values. You either trust this body of research, or you do not. An intermediate position might be to trust some studies, and reject others. It would, however, be somewhat arbitrary exactly where and how to draw the dividing line between studies of acceptable quality and studies of unacceptable quality. The problem is compounded by the fact that standards of quality have evolved over time. Perhaps the most striking illustration of this is the research by Jones-Lee and his team, which he defended in some detail in his book in 1989, but rejected less than ten years later (1998) under the pressure of a growing number of anomalies in the findings. By the same token, the history of studies of compensating wage differentials is characterised by a tendency to reject older studies whenever a new potentially confounding factor has been discovered or when better sources of data regarding risk have become available.

There have, to be sure, been attempts to identify the "best" studies, notably in the meta-analyses of Lindhjem et al. (2011, 2012A, 2012B). Experience shows that it does not solve

the problem. Even among the best studies, there is a huge range of values – smaller than if all studies are included – but still too large for the results to be meaningfully applied in policy analyses, such as cost-benefit analysis.

Finally, in the sixth place, current theory predicts that the value of a statistical life will vary. Unfortunately, the theory does not predict how large the variation will be; it only identifies sources of variation and the direction of their impact. Hence, it is impossible to say how much of the huge variation in estimates of the value of a statistical life is theoretically plausible, and how much of it is "excess" variation attributable to sources that, according to theory, should not produce variation in the value of a statistical life. Therefore, one cannot identify what Miller (1990) termed "the plausible range for the value of a statistical life" by critically examining empirical estimates from a theoretical perspective. Simple estimates, made in this report, suggest that most of the variation in estimates of the value of a statistical life is artefactual, and only a small part of it can be regarded as theoretically plausible.

It is concluded that empirical studies of the value of saving a life have not produced estimates that can be trusted and applied in cost-benefit analysis. It is judged as unlikely that future research applying the methods that have been used so far will produce trustworthy and sufficiently precise values.

11.2 Alternatives: The valuation of quality of life

In Chapter 2 it was argued that some form of economic valuation of safety is inevitable. Valuation research as conducted so far has not produced credible estimates of the value of safety. Given the fact that an economic valuation will always be made, if only implicitly, it is of some interest to investigate whether an alternative approach to the monetary valuation of safety can be imagined that may produce more precise estimates than standard valuation studies have done so far.

One of the problems encountered in contingent valuation studies was poor understanding of changes in low levels of risk. To get around this problem, Carthy et al. (1999) developed the chained contingent valuation/standard gamble approach. Respondents were first asked about their willingness-to-pay (WTP) for a certain cure of a slight injury that would not result in any lasting impairment or the amount they would need to get in compensation after having sustained the injury (WTA). Once these amounts had been obtained, respondents were then given a standard gamble task involving two different treatments for injuries of different severities. The idea was to elicit the relative level of utility loss associated with a given injury, compared to death. Once this loss of utility has been obtained, it was used as a basis for estimating the value of a statistical life by scaling up the monetary valuations of the cure for the slight injury.

To give a (somewhat simplified) example. Suppose WTP for curing the injury is 10,000. Suppose further that its relative utility weight (compared to death) is 0.04. The value of a statistical life is then 10,000/0.04 = 250,000. This is actually a bit simplified, as Carthy et al. (1999) formally specify a number of utility functions and insert the values into these functions.

This approach involves trying to value life by assigning utility values to health states. There exists a huge literature on how to assess the quality of life associated with specific health states. For an overview of various instruments for measuring health, see Bowling (2005). The most common measure of quality of life related to health state is the Quality Adjusted Life Year, commonly abbreviated QALY. QALYs are obtained by means of a multidimensional scaling of health states. For each dimension, a number of levels of

functioning are defined, ranging from normal function to no function (e.g. normal eyesight (including corrected to normal), slightly reduced vision, several reduced vision, total blindness). Common dimension of health include: ambulatory function (ability to walk), cognitive function, self-care function (ability to dress, go to the toilet, wash oneself, etc.), sensory function (hearing, seeing, smell, taste, sense of touch), presence of pain and changes in appearance. An early example of use of a QALY scale to estimate road accident costs can be found in a paper by Miller (1993). Table 11.1. reproduces table 2 in that paper.

For each dimension, five levels of functioning are defined, ranging from no loss of function (coded as 0) to complete loss of function (code as 4). The loss in quality of life associated with loss of function varies between the dimensions of health. It is seen that complete loss of cognitive function is rated as the worst, having a value of 95, meaning that on the 0 to 1 scale usually applied for QALY, a state of complete loss of cognitive function during one year would be rated as 0.05.

Table 11.1: Loss of function related to seven dimensions of health. Copied from Miller 1993

Table 2. Percentage loss in overall functional capacity resulting from loss on a single dimension of functioning

Severity	Dimension of functioning						
	Mobility	Cognitive	Self care	Sensory	Cosmetic	Pain	Ability to work
4-Total	68	95	71	37	10	60	33
3-Severe	55	90	36	24	6	10	NA
2-Moderate	28	20	33	15	3	3	NA
1-Mild	13	5	8	5	1	1	8
0-None	0	0	0	0	0	0	0

Note that the time dimension is not used in assigning the scores for functional losses. Thus, a person suffering a mild loss of mobility for a limited period, as a result of, for example, an uncomplicated leg fracture might regain normal function after two months. If the only additional loss of function was moderate pain during the two months, the overall loss of quality of life for the person would be (assuming that the dimensions can be treated as independent):

QALY = 1 - [0.87 (mobility) · 1.00 (cognitive) · 1.00 (self-care) · 1.00 (sensory) · 1.00 (cosmetic) · 0.97 (pain) · 1.00 (ability to work)] · part of year in that health state =

1 - 0.844 · (2/12) = 0.026, equivalent to a QALY-score of 0.974, i.e. nearly perfect health during one year.

Several instruments for estimating QALYs exist. Elvik (1995) compared four of them and showed that they do not give identical results for a given health state, described in terms of levels of functioning and the presence of pain and discomfort. Data from a comprehensive Norwegian study of the long term consequences of traffic injury (Haukeland 1991) were mapped onto four QALY-instruments (health state indexes):

1. The Quality of Well-Being scale (QWB-scale)
2. The McMaster Health Classification System (McMaster)
3. The Rosser and Kind Index (Rosser and Kind)
4. The EuroQol Instrument (transformed) (EuroQol)

There are several attractions to using QALY values referring to specific types of injury or health states as a basis for obtaining monetary valuations of safety. In the first place, some of the instruments permit a quite detailed description of health states. As an example, the functional capacity index developed in the United States (MacKenzie et al. 1996), identifies ten areas of functioning and up to seven levels of functioning in each of these areas. There

are hundreds of possible combinations of values for the ten areas of functioning, permitting a concise description of very many health states.

Each area of functioning is assigned a score and an overall score, intended to be interpreted as a utility value, can be estimated. In the second place, monetary values can be elicited without referring to low levels of risk or without offering respondents hypothetical choices. One only needs to describe the health states in sufficient detail. These descriptions will often be easier to understand than hypothetical changes in very low levels of risk.

In the third place, for many slight injuries or losses of function, many respondents will have experienced them personally and thus be quite familiar with the topic of a valuation study.

On the other hand, there are problems in using QALYs as a basis for monetary valuations. A given health state, described in terms of levels of functioning in specific areas (eating, dressing, going to the toilet, walking, feeling distressed, etc.) is not necessarily assigned the same QALY value by all QALY-instruments that have been developed. One might try to circumvent this problem by not stating QALY-values, for example by asking respondents in a valuation study to assign these values, or by asking them directly (i.e. without using QALY-values at all) to state their willingness-to-pay for avoiding a certain health state or the compensation they would need if having to live in the health state permanently. However, bypassing a formal assignment of QALY-values is not feasible if one wants to value lifesaving, since the QALY values are needed as "building blocks" in obtaining a monetary valuation of saving a life.

The basic idea is very simple. A QALY scale is standardised so that death has the value of 0 and perfect health a value of 1. States of reduced health, for example as a result of a traffic injury, have values between 0 and 1. A value of 0.9 would indicate a loss of 0.1 years of living in perfect health. If a health state with a QALY value of 0.9 is valued at 250,000, the implicit value if a statistical life is 250,000/0.1 = 2,500,000.

Using QALY values as a scaling factor this way becomes problematic if different ways of eliciting the QALY values associated with specific health states produce different values. There are four ways of eliciting QALY values (Nord 1992, 1999):

1. Visual analogue scale: This is a very simple instrument. Respondents are asked to assign a numerical value to a given health state by placing it on a numerical scale, ranging from, for example 0 to 100 (values are subsequently rescaled to the 0 to 1 interval).
2. Time trade-off: Individuals are asked how much life expectancy they are willing to give up in return for avoiding a certain heath impairment. Results show not only the value of the health state, but also discounting. Moreover, results may be sensitive to whether one has experienced the health state or not. There is strong adaptation to changes in health, and an individual who has adapted well to a certain health impairment may not be willing to sacrifice any length of life in order to be restored to perfect health.
3. Standard gamble: Respondents are offered a choice between a standard treatment that with certainty will result in some lasting impairment, and a new treatment that with a certain probability will result in a perfect cure and with a complementary probability result in immediate death. The sum of probabilities for the two outcomes – perfect health and death – equals one. The task is to determine the probability of death that makes the individual indifferent between the standard treatment and the new treatment. Results will reflect not just the valuation of the health states, but also attitudes to risk.

4. The person trade-off or equivalence method: Respondents are asked to compare two health states, one of which is associated with a minor reduction in health, the other with a major reduction in health. They are then asked to state how many more people that would need to be cured of the minor health impairment to be equivalent to one person cured of the major health impairment.

Unsurprisingly, these methods for eliciting QALY values do not produce identical results. The four QALY-instruments used by Elvik (1995) to describe the long term impacts of traffic injury in Norway did not produce identical results, and would yield very different scaling factors for estimating the value of a statistical life. Figure 11.1. shows QALY values for these instruments for adults who were either slightly injured (AIS 1) or seriously injured (AIS 3-5). The dashed lines indicate how the QALY values change as time passes since the injury, showing that a process of slow recovery is taking place (the values are getting closer to 1, which is perfect health). The period covered goes up to 4.5 years after injury (0-0.5 is the first half year, 0.5-1.0 is the second half year, and so on). The lower dashed line is close to the health state values obtained by the Quality of Well-being scale (QWB). The upper dashed line is close to the health state values obtained by the EuroQol instrument (yellow bars). Values estimated according to the McMaster Health Classification System are located between the dashed lines. The Rosser and Kind index stands apart from the other three. It has the value of 0.99, with a single exception. It is shown by the grey bars rising above the others in the diagram.
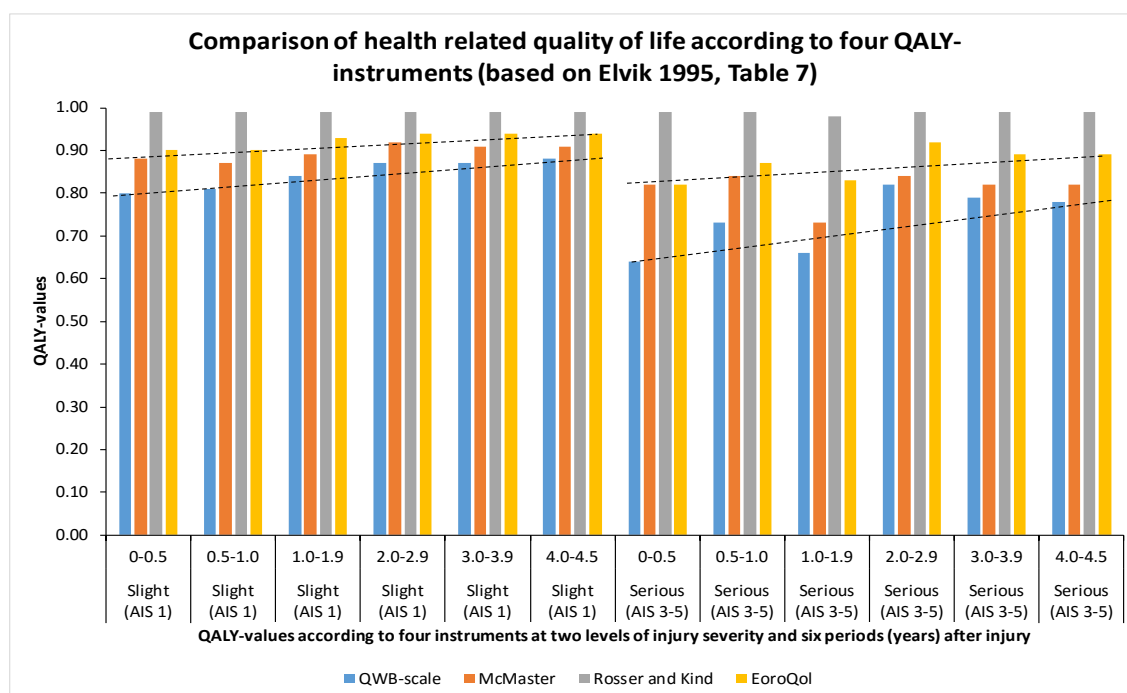

Figure 11.1: Comparison of health-related quality of life among traffic injury victims according to four QALY-instruments. Based on Elvik 1995

Based on the QALY-values several scaling up factors can be estimated as a basis for estimating the value of a statistical life based. Suppose, at one extreme, that the Rosser and Kind index for slight injury is used. Since, according to this index, a slight injury involves the loss of only 0.01 year of living with perfect health, the scaling factor is 100, i.e. the value of a statistical life is 100 times larger than the value of the loss of 0.01 years of perfect health. At the other extreme, the mean QALY score for a slight injury according to the QWB-scale is 0.845, implying a scaling-up factor of only 6.45. Moreover, the QWB-scale

indicates that QALY-values change over the course of the 4.5 years covered by the data, starting at a low of 0.80 during the first half year after injury and rising to 0.88 when at least four years have passed since the injury. To use a monetary value for a QALY as a scaling factor to estimate the value of a statistical life, one ought to account for this trend. These simple examples show just a few of the complications one will encounter when trying to use QALY-values as a basis for valuing a statistical life.

There are, however, more fundamental problems. Hammitt (2002) discusses the relationship between QALYs and willingness-to-pay. He notes that although both these modes of valuation are based on individual preferences, they make very different assumptions about the utility function underlying these preferences. In particular, since QALYs are intended to be additive, cardinal (measured at the interval level of measurement) and comparable across individuals, they must satisfy a number of quite severe restrictions on the utility function. Hammitt (2002:987-988) lists four restrictions:

1. Mutual utility independence. This restriction means that preferences between lotteries on health states do not depend on remaining life span, and that preferences between lotteries on life span do not depend on health state. These restrictions are needed in order to represent health state as a product between its level (say 0.9) and its duration (say 5 years). The assumption is very unrealistic. It implies, for example, that preferences for extending life span are the same for a person in perfect health as for a person in a very reduced state of health, involving restrictions on functioning that prohibit all activities people normally enjoy.
2. Constant proportional trade-off of longevity for health. This restriction means that the QALY score assigned to a given state of health is independent of the duration of that state. Again, this is very unrealistic. Thus, as already mentioned, there are many studies showing considerable hedonic adaptation to changes in health state. A state which is awful in the beginning normally gets less awful as time goes on.
3. Risk neutrality over life span. This restriction means that holding health state constant, the individual prefers whichever lottery on life span that provides the greatest life expectancy.
4. Additive independence across periods. This means that the individual's preferences between lotteries on health in any period, do not depend on health in other periods. Again, one can easily imagine that this restriction may not be very realistic. Thus, an individual who already has a permanent health impairment, may, even if well-adapted to that impairment, be less inclined to accept a gamble involving the potential of a further impairment than a healthy individual.

Hammitt (2002:988) notes that empirical research shows that individual preferences for health often violate the restrictions that are necessary for QALYS to be a valid measure of utility associated with health. He goes on to remark that estimates of QALYs may be less variable between people and studies than estimates of WTP, because the framework needed to elicit QALYs imposes stronger restrictions than the framework needed to elicit WTP. Yet, as shown in the numerical example above, even QALYs may differ substantially and lead to estimates of the value of a statistical life that differ by a factor of at least 10.

In a subsequent paper, Haninger and Hammitt (2006) show that willingness-to-pay is not proportional to the number of QALYs, as it should be for QALYs to a valid measure of utility. There is, as has been found for changes in risk, strong subproportionality (insensitivity to the number of QALYs).

To conclude: There are many ways of eliciting QALYS. These do not produce identical results. Willingness-to-pay for changes in health states described in terms of QALYs is insensitive to the number of QALYs. To use QALYs as a basis for valuing lifesaving if therefore likely to reproduce the same wide dispersion of values as the traditional valuation

studies have done and would therefore not solve the problem of the wide dispersion in values.

11.3 Alternatives: The capability approach

As already mentioned, several studies have found considerable hedonic adaptation to health impairments. Hedonic adaptation means that subjective well-being recovers toward a normal level after an initial reduction as a result of a loss of health. An early, and widely quoted study showing this was reported by Brickman, Coates and Janoff-Bulman (1978). They compared self-reported happiness among lottery winners, paraplegic accident victims and a control group from the normal population. The lottery winners are of no interest in the present context, but key findings for the other two groups are reported in Figure 11.2.
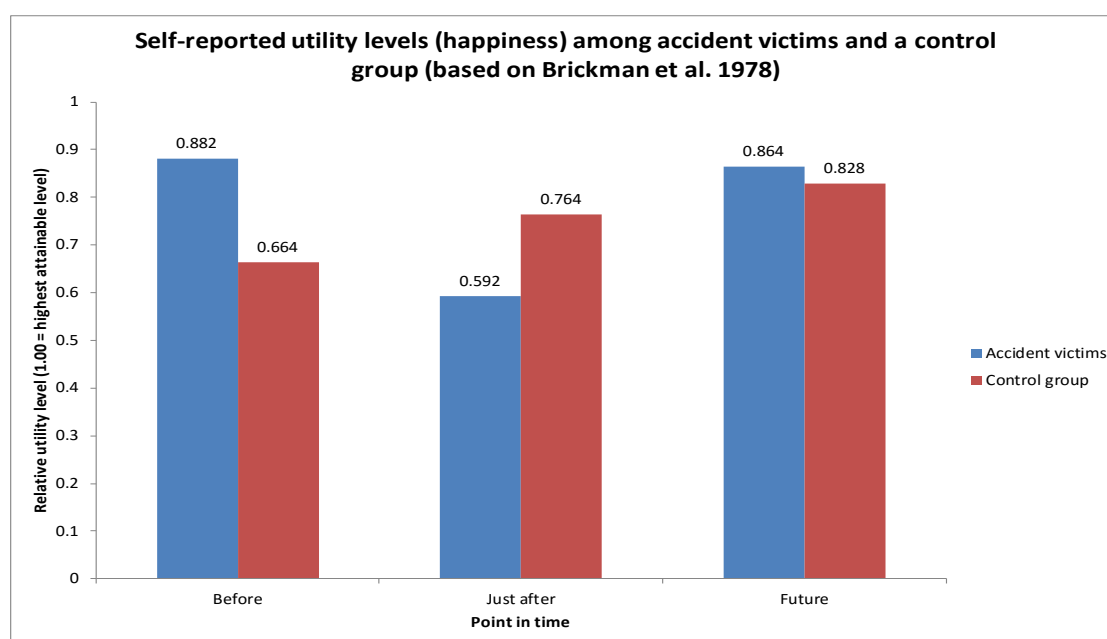

Figure 11.2: Self-reported happiness (converted to a scale with range between 0 and 1) among healthy subjects and accident victims who recently became paraplegic. Based on Brickman et al. 1978

The reported levels of happiness have been converted to a scale in which the maximum value is 1.00. It can be seen that the accident victims reported a significant drop in their happiness just after the accident, but that they expect happiness to recover to almost the level before the accident. The study has been criticised, among other things because the sample was small. Nevertheless, its main finding is not unique. Similar findings have been reported by Nord (1999).

Sen (1987:45) is critical of allowing for hedonic adaptation when trying to measure utility, stating that:

"The hopeless beggar, the precarious landless labourer, the dominated housewife, the hardened unemployed or the over-exhausted coolie may all take pleasures in small mercies, and manage to suppress intense suffering for the necessity of continuing survival, but it would be ethically deeply mistaken to attach a correspondingly small value to the loss of their well-being because of this survival strategy."

The fact that a wheelchair user reports almost the same level of happiness as before the injury does not mean that the loss of functioning – the loss of capabilities – is insignificant. Menzel et al. (2002) comment upon the view of Sen in the following terms:

"Undoubtedly, these are important points to which utilitarianism as a moral philosophy must respond. The question in our context, however, is whether the deprivation factor throws the ratings of HRQoL (Health Related Quality of Life) that are carefully procured from adapted patients into the same kind of doubt that deprivation generally throws a utilitarian metric of desire fulfilment. Adaptation often involves genuinely successful achievements and shrewd control over the trajectory of a person's inner life. In these cases, the adapted person is anything but "broken", and hardly "subdued". … Thus, while Sen's argument from entrenched deprivation should give us pause about too readily or generally using adapted patients' quality-of-life ratings, it does not justify an across-the-board rejection of values shaped by adaptation."

Sen has proposed the capability approach to human welfare as an alternative to the traditional utilitarian measure based on preferences and self-reported quality-of-life. He has described this approach in many publications, but perhaps at greatest length in "Inequality reexamined" (Sen 1992). The basic idea of the capability approach is that what an individual has the capability to do determines how much freedom that individual has. Capabilities can be determined objectively and can be compared between individuals, unlike utility as traditionally conceived in economic theory.

Could the capability approach serve as a basis for a monetary valuation of life and health? While Sen clearly is sceptical to conventional valuation studies, he adopts a broad perspective on cost-benefit analysis (Sen 2000:939), stating that:

"It is compatible, for example, with weights based on willingness-to-pay as well as some quite different ways of valuation (for example, through questionnaires), which may supplement or supplant the willingness-to-pay framework."

Apart from this remark, Sen did not mention the capability approach in his discussion of cost-benefit analysis. In a later contribution (Sen 2008), he argues that the capability approach is more suitable for assessing quality of life than subjective measures like happiness or utility, because the subjective measures are vulnerable to hedonic adaptation, i.e. people reduce their ambitions and expectations to adapt to adverse circumstances. He further remarks (page 19):

"The subject of welfare economics suffered a big blow in the 1930s when economists came to be persuaded by arguments presented by Lionel Robbins and others (influenced by "logical positivist" philosophy) that interpersonal comparisons of utility have no scientific basis and cannot sensibly be made. One person's happiness, it was argued, could not be compared, in any way, with the happiness of another."

He later remarked that if happiness is to be used as a criterion for social evaluation, it should be used in an interpersonally comparable form.

The capability approach was developed in response to the widespread finding that even people living in extreme poverty, deprived of a decent life in very many ways, reported quite high levels of subjective well-being. Sen feared that these reports might be used as an excuse for not doing anything about poverty or deprivation. The link between the capability approach and valuation of life and health is not clear. On the one hand, one might think that people who report a high level of subjective well-being would have a high willingness-to-pay for reducing the risk of premature death. But, they would not have the ability to pay.

The other touching point concerns people who have sustained serious and permanent injury, but after an period of adjustment nevertheless report a high level of subjective well-being. Again, the concern can be raised that since these people remain happy, preventing their injuries is not so important. This is a misinterpretation, and as will be shown in section 11.5, even slight reductions in utility are compatible with very high monetary valuations of the prospect of a complete recovery from the injury.

Sen has never developed the capability approach to the point of offering a list of essential capabilities, akin to the list of primary goods provided by John Rawls (1971) in his theory of justice (primary goods are goods anyone would want). Sugden (2008) argues that the capability approach is paternalistic, an interpretation Sen strongly disagrees with.

Although contexts can easily be imagined in which an objective measure of welfare – based, for example, on the quality of housing, access to clean water, having a regular job, and so on – would be more relevant for policy priorities than reports of subjective well-being, it is not clear how trade-offs between different indicators of objective welfare can be established so as to most effectively promote welfare. Which is the most important: clean water or a regular job? Which is the most important: learning to read or not living close to a very polluted street?

It seems difficult to answer such questions without using information about what those who are concerned think themselves. Setting priorities based on so called objective indicators of welfare without even asking those who are intended to benefit from the policies is thus very clearly paternalistic, although it is, on any reasonable interpretation, well-meaning. It should remain an objective to develop a method for valuing life and health that is non-paternalistic.

11.4 Alternatives: Utility functions based on happiness studies

The histories of economics and psychology have an interesting parallel. Originally, economists treated well-being as subjective; it was a state of mind, which was observable by way of individual reports of it. As noted in the quote from Sen above, this conception of welfare was rejected in the 1930s. You cannot observe what goes on inside somebody's head, it was argued. Concepts based on mental states are therefore unscientific. The only thing we can observe is behaviour. This line of thinking lead to revealed preference theory and to the reduction of utility to an ordinal measure that could not be compared between people. These points of view have characterised mainstream economic theory until now.

Meanwhile, in psychology something very similar happened. It was argued that we cannot observe how people feel or think and we cannot trust what they tell us about it. We can only observe behaviour, and we can infer what shapes behaviour by manipulating the consequences of behaviour to determine so called "contingencies of reinforcement", i.e. factors that reinforce behaviour and maintains it in the long run. This line of thinking was the behaviourist approach, whose most prominent representative was the hugely influential psychologist B. F. Skinner. For a long time, revealed preferences were the only kind of preferences economists would take seriously, and observed behaviour, together with its contingencies of reinforcement, was the only thing psychology could study scientifically.

This started to change in the late 1960s and cognitive psychology was reborn and revolutionised under the leadership of Amos Tversky and Daniel Kahneman. They soon extended their research to testing key propositions of the modern version of utility theory developed by Von Neumann and Morgenstern (1953). A Von Neumann-Morgenstern (VNM) utility function, unlike the utility functions of standard economic theory, is cardinal (measured at the interval level of measurement). In a VNM utility function, it makes sense to compare levels of utility and speak about the size of differences in utility. Moreover, a VNM utility function can be derived from statements about preferences, in particular preferences between lotteries. These properties make VNM utility functions eminently suited to the study of changes in health risks, since these risks can be represented as lotteries and the degree of risk aversion readily determined.

The revival of interest in cognitive psychology found an expression in economics in a revival of interest in the study of subjective well-being, a field of research currently known by the name of happiness research. This field of research has grown very fast in the past 20 years, and many books and papers have been published (Kahneman, Diener and Schwartz 1999, Frey and Stutzer 2002A, 2002B, Layard 2005, Kahneman and Krueger 2006, Adler and Posner 2008, Bruni, Comim and Pugno 2008, Clark, Frijters and Shields 2008, Dolan and Kahneman 2008, Dolan, Peasgood and White 2008, Loewenstein and Ubel 2008, Oswald and Powdthavee 2008, Van Praag and Ferrer-i-Carbonell 2008, Diener et al. 2010, Easterlin 2010, Finkelstein et al. 2013, Stone et al. 2013, Viscusi 2013, Weimann, Knabe and Schöb 2015). The results of happiness studies have already been applied to value non-market goods (Welsch 2006, Ferreira and Moro 2010). It is relevant to ask if a similar application to the valuation of changes in the risk of death and health impairments is possible.

At this point, it is relevant to revisit the paper by Fischhoff (1991) on value elicitation. Fischhoff asked "Is there anything in there?", suggesting that the well-structured preferences often assumed in economic theory simply do not exist. He made a distinction between what he called the philosophy of articulated values and the philosophy of basic values. According to the former philosophy people have well-thought-out preferences and are able to articulate these preferences when asked to do so. According to the latter theory, people lack well-differentiated values for all but the most familiar of evaluation questions, about which they have had the chance, by trial, error and rumination, to settle on stable values.

The study of subjective well-being asks people very simple questions about how happy they are. While this might not be an elicitation of values in the economic sense of the term, it is an elicitation of basic feelings. There are certainly sources of error these surveys. If one asks at a time when a person is in a bad mood, he or she is likely to rate subjective well-being lower than when in a good mood. There are, as in all questionnaires, scale effects. If a scale has just three levels, many people will settle for the middle level. In general, few people use the extreme values. It would suggest that, for example, that they are as happy as they could possibly be. Few people would say that. When thinking about it, most people, while having a high level of subjective well-being, can think of a few things in life they would like to improve, and therefore not select the top score on a scale, ranging from, say, 0 to 10.

Despite these reservations, it has been proposed that surveys of subjective well-being can be interpreted as empirical utility functions. Interpreting such surveys as indicators of utility is controversial; for a critical perspective, see e.g. Clark et al. (2008). However, results of happiness studies have been used to estimate the utility of money (Layard et al. 2008).

Some countries have a long tradition of conducting polls in which a representative sample of the population are asked about how happy they are, all things considered. As an example, the General Social Survey in the United States, contains the question:

All things considered, how happy would you say you are these days? Answers can be given as "not too happy" (score 1), "pretty happy" (score 2) and "very happy" (score 3).

This survey has been repeated many times. Frey and Stutzer (2002A) present the results of two of the surveys, reproduced below in Table 11.1. The term equivalence income is household income divided by the square root of the number of people belonging to a household (an arbitrary rule for adjusting for household size).

Table 11.1: Summary data on happiness and equivalence income (1996 values) in two rounds of the General Social Survey in the United States. Based on Frey and Stutzer 2002A.

Income (deciles)	Mean equivalence income		Mean happiness rating	
	1972-74	1994-96	1972-74	1994-96
1	2522	2586	1.92	1.94
2	5777	5867	2.09	2.03
3	8694	8634	2.17	2.07
4	11114	11533	2.22	2.15
5	13517	14763	2.19	2.19
6	15979	17666	2.29	2.29
7	18713	21128	2.24	2.20
8	22343	25745	2.31	2.20
9	28473	34688	2.26	2.30
10	46338	61836	2.36	2.36
Full sample	17434	20767	2.21	2.17

Several observations can be made. First, although mean income increased in real terms (income is stated in 1996-values both for 1972-74 and for 1994-96), mean happiness score did not. In fact, it declined ever so slightly. Second, inequality in income increased. The growth in real income was very small or non-existent in the four lowest deciles of the income distribution. In the tenth decile, however, income increased by 33 percent. Third, mean happiness scores are positively related to income, although the increase in happiness associated with higher income is quite small. This might of course be a result of the scale used, which had a range from 1 to 3, meaning that the happiest conceivable group, according to this scale, could not score more than three times the unhappiest conceivable group, whereas income in the last period varied by a factor of almost 24.

Figure 11.3. shows the relationship between income and happiness score. Logarithmic functions describe the relationship very well, although a power function fitted marginally better to the data points for the most recent survey (1994-96).
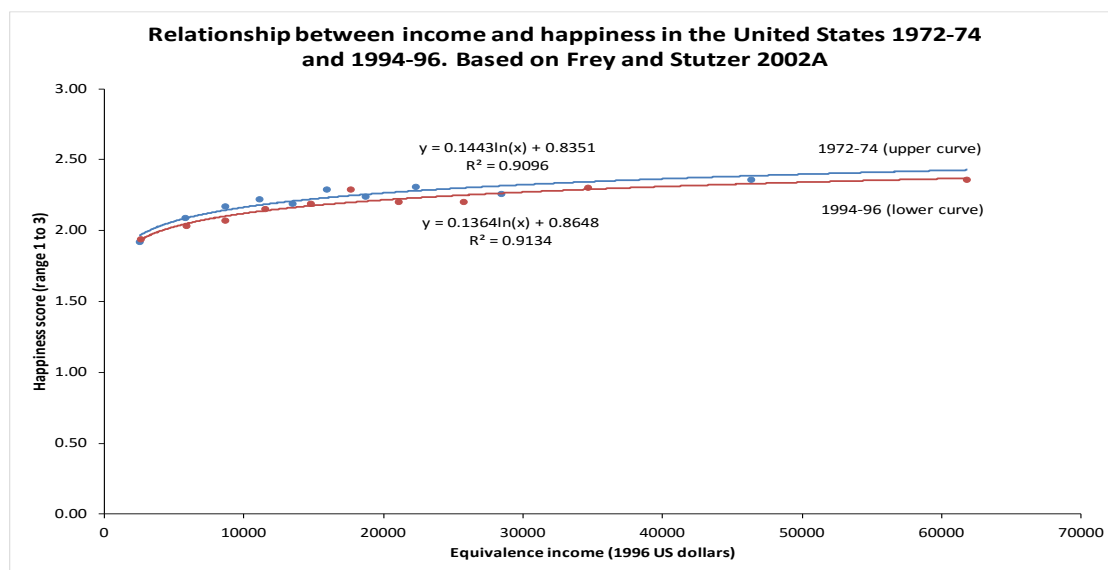

Figure 11.3: Happiness as a function of income. Based on Frey and Stutzer 2002A

It is seen that the curve fitted to the most recent study is located below the curve fitted to the oldest study, although mean income increased from the first to the second survey.

Can curves like those shown in Figure 11.3, fitted to data on income and subjective well-being (happiness) be interpreted as utility functions? If they can, what are the implications for the monetary valuation of life and health?

Opinions differ regarding this issue. Viscusi (2013) is sceptical. He offers the following assessment of happiness scores (pages 1736-1737):

"Although happiness scores elicited in surveys are not tantamount to utility levels, many researchers have advocated them as measures of well-being. However, unlike the VSL formulation, well-being measures have no explicit economic content and no cardinal significance. A representative well-being survey question asks the respondent to rate his or her happiness or satisfaction with life on a numerical scale such as 0 to 10, 1 to 10, or 1 to 7. At a most fundamental level, how should a person even think about such a question? What is the reference point for such an assessment? … If you have a permanent disability, then you may nevertheless feel pretty good about how your life is going on a particular day, but you might be much happier if you were not disabled – and you would give a different happiness score if the no-disability state were in the reference set."

He goes on to remark that happiness scores share the inherent inadequacies of ordinal measures. He adds that for the scales to have meaning, the intervals for different respondents (e.g. going from 7 to 8, or from 8 to 10 on a 10-point scale) must represent identical welfare effects across people. This is unlikely to be the case. Different people will, for example, interpret a score of 7 on a scale from 0 to 10 differently. One person might regard this a very high level of happiness, another might rate a score of 7 as signifying a population average level of happiness (nearly all studies find that average scores are in the upper half of a scale).

These points of view are clearly reasonable and suggest that interpreting functions like those shown in Figure 11.3 as utility functions is indefensible. Weimann, Knabe and Schöb (2013:146) share Viscusi's points of view and state:

"One cannot say that two people are equally satisfied or equally happy just because they mark the same score (say, a 7) on a scale of 0 to 10, since each of them defines for himself what a 7 on the scale means for him. … In much the same way that a 7 marked by Ms A may mean something different from a 7 marked by Mr B, Ms A's 7 in the year 2000 does not mean the same thing as her 7 in the years 2013. If the reference point has changed in the thirteen years between the surveys, the scale Ms A uses in 2013 is entirely different from the one she used in 2000."

Weimann, Knabe and Schöb are also sceptical about studies showing that disabled people adapt to their disability, and suggest that it might be the case that they reinterpret the happiness scale – in other words shift the reference point.

The views of Viscusi and Weimann et al. have great psychological plausibility. Frey and Stutzer (2002A) on the other hand state that (page 426):

"Happiness is not identical to the traditional concept of utility in economics. It is, however, closely related. … Subjective well-being can be considered a useful approximation to utility, which economists have avoided measuring explicitly. This allows us to empirically study problems that previously were analysed only on an abstract theoretical level."

An important issue is whether answers to happiness surveys can be interpreted as a cardinal scale. Van Praag and Ferrer-i-Carbonell (2008) discuss this issue in some detail. Using the German scale, which ranges from 0 to 10 as the starting point, they note that this scale is ordinal. However, using an ordered probit analysis, they argue that the scale can be treated as approximately cardinal, and show that the results obtained when treating the scale as cardinal are almost identical to those obtained when treating it as ordinal.

Clark, Frijters and Shields (2008) argue that utility, as normally defined in economics, is expected utility from a choice, i.e. decision utility. Happiness, on the other hand, is an evaluation of what has occurred, an evaluation of experienced utility that may not be identical to decision utility. This point is discussed at greater length by Loewenstein and Ubel (2008). They argue that experienced utility may not be suitable as guideline for public policy due to the emotional adaptation to adverse events. They state (page 1799):

"If we based public policy on experience utility, we might avoid spending scarce public resources on measures to prevent adversities like leg amputations, spinal cord injuries, and kidney failure which most people would be very adverse to experiencing, but which lead, for most people, to significant emotional adaptation. … Not only do such policy implications conflict with common intuitions and values, but, despite reporting levels of mood and well-being that are similar to healthy persons, people experiencing these conditions report a willingness-to-pay large sums of money or make other major sacrifices to restore their lost function. In our own research examining different measures of utility for different medical conditions, we have repeatedly found striking divergences between measures based on experience utility and those based on decision utility."

The divergence Loewenstein and Ubel mention is actually not surprising and is no paradox, as will be shown in the next section. Loewenstein and Ubel think that an ideal measure of utility for public policy should reflect both decision utility and experience utility. Can the functions shown in Figure 11.3 serve as a basis both for estimating experienced (the past tense is arguably more correct than the term "experience" as used by Loewenstein and Ubel) utility and decision utility? Yes, they can and precisely for the type of problem Loewenstein and Ubel discuss.

As mentioned, results of happiness studies have been interpreted as utility functions by some researchers. Most notably, Layard, Mayraz and Nickell (2008) estimate the marginal utility of money based on the results of eight different surveys of happiness. These eight surveys used different rating scales and were made in different countries. The results were nevertheless very consistent. The marginal utility of money, holding other variables influencing happiness scores constant, was best described by a function of the following form:

Marginal utility of money $= \dfrac{y_{it}^{1-\rho} - 1}{1-\rho}$

This is a Box-Cox transformation. The parameter p was estimated to a value of 1.26, which means that the marginal utility of money decreases by 1.26 percent for each percent increase in income, i.e. marginal utility declines more than implied by a logarithmic utility function, in which p would be 1.00. The utility function fitted by Layard et al. fitted the data very well. The subscript it represents country and year.

Researchers who have used the results of happiness surveys to obtain monetary valuations have an altogether more optimistic view regarding the utility interpretation of functions fitted to happiness data. Consider first the paper by Welsch (2006). He used a study of life satisfaction (the terms happiness, life satisfaction and subjective well-being are all used in the literature, but all refer to the same types of study) to value the elimination of various types of air pollution. Having first discussed the problems of standard approaches to valuation, he explained the life satisfaction approach in the following terms (page 802):

"The life satisfaction approach avoids some of these difficulties. This technique does not rely on asking people how they value environmental conditions. Instead, individuals are asked in surveys how satisfied they are with life, and econometric analysis is used to identify if and how their en masse answers move with environmental conditions. Thus, the approach does not require awareness of cause-effect relationships on the part of the individual. … For these reasons the life satisfaction approach is cognitively less demanding than contingent valuation and does not evoke answers considered desired by the respondents."

Welsch obtains monetary valuations of nitrogen, particles and lead by estimating a double log equation, having the log of happiness scores (ranging from 1 to 4) as dependent variable, and the log of income, and the logs of the concentrations of the pollutants as independent variables. Ferreira and Moro (2008) note that the use of data on subjective well-being in valuation studies is based on the assumption that these data are a good approximation to a utility function, which is controversial. They discuss whether decision utility and experienced utility are always the same, as assumed in mainstream economic theory. It is well established that they are not, and that experienced utility, which incorporates learning from past choices, is the best measure of utility. Or, put somewhat differently: An individual who repeatedly makes choices he or she regrets, is likely to report a lower level of subjective well-being than an individual who makes choices which have more desirable consequences.

The next section explores the use of data on subjective well-being to obtain values of a statistical life. It relies on the closed-form expression for the value of a statistical life, shown in Chapter 3 based on the original derivation by Fromm.

11.5 Implications of selected utility functions

There is a standard formula for the value of a statistical life, based on an assumption of utility maximisation, which gives a closed-form solution to the valuation problem if estimates of the utility of income or wealth can be obtained. See chapter 3 for the derivation of the formula. The standard formula is:

$$\text{Value of a statistical life} = \frac{U_a(w) - U_d(w)}{(1-p)U_a'(w) + pU_d'(w)}$$

$U_a(w)$ is utility of wealth conditional on survival (alive), $U_d(w)$ is utility of wealth conditional on death. The prime denotes the first derivative. P is the probability of dying and $1 - p$ is the probability of survival.

If the utility of death is zero, the equation can be simplified as follows:

$$\text{Value of a statistical life} = \frac{U_a(w)}{(1-p)U_a'(w)}$$

Functions fitted to happiness surveys and interpreted as utility functions obviously represent utility when alive and say nothing about the utility of death. Below, therefore the value of a statistical life will be estimated by applying the simplified version of the formula.

The estimations are exploratory only. It is, however, important to determine how the value of a statistical life, as estimated by relying on utility functions based on subjective well-being depends on:

1. The functional form of the utility function (logarithmic, exponential, power, quadratic, etc.)
2. The level of income
3. The level of risk

These issues will be explored for a number of utility functions that can be fitted to data or found in the literature.

11.5.1 A quadratic utility function (Hellevik 2008)

The first utility function to be examined was fitted to the data in Hellevik (2008). Hellevik scored happiness by taking the difference between the percentage who reported they were very happy and the percentage who reported they were not so happy. This indicator can vary between –100 and +100. In the data he presented, values between –9 and +36 were reported. To make all values positive, 100 was added them, producing a range from 91 to 136. When these values are plotted against household income, Figure 11.4 emerges. Household income was stated as intervals and the midpoint of each interval was used, except the uppermost interval, which was set equal to 1,200,000.

A second degree polynomial is seen to fit the data almost perfectly. The shape of the polynomial is very similar to the shape of a typical utility function as presented in economics textbooks. According to Arrow (1965) fitting utility functions to data by means of second degree polynomials is very common in economics. He argues against this practice. While the polynomial may be well-behaved in the range of the data it has been fitted to, as the one in Figure 11.4, extrapolating a second degree polynomial will almost always lead to nonsensical results. At some point, the quadratic term becomes dominant, and marginal utility become negative. This is not plausible. Arrow argues for using logarithmic utility functions. These have nice mathematical properties making them convenient to analyse, and their shape can easily be changed by changing parameters.
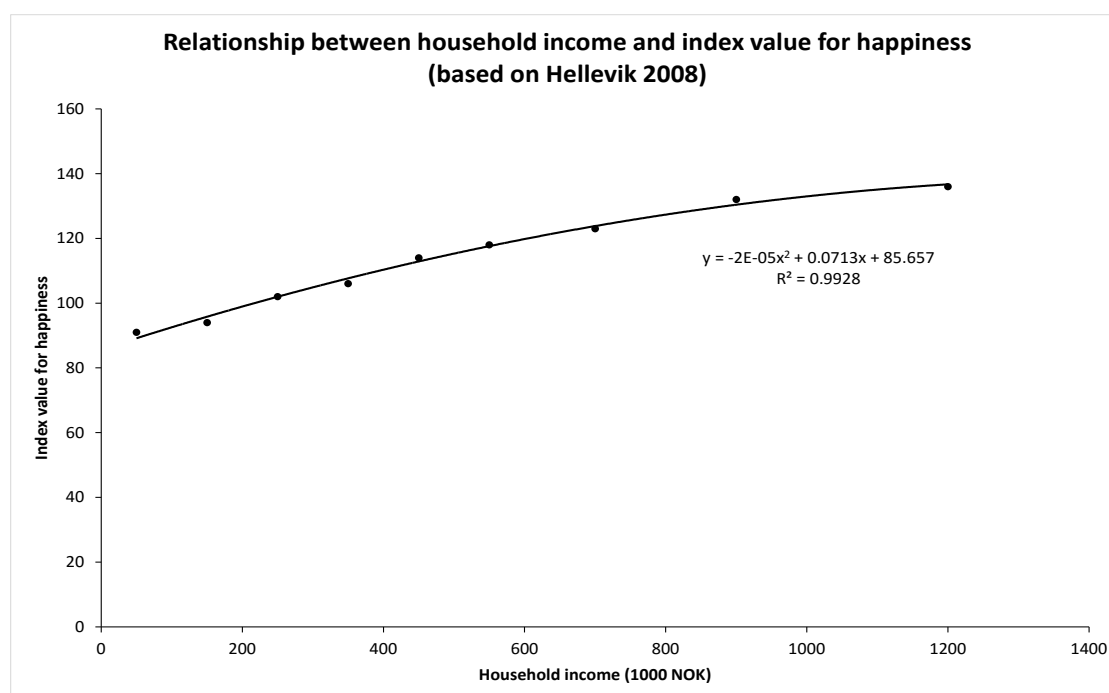

Relationship between household income and index value for happiness (based on Hellevik 2008)

$y = -2E\text{-}05x^2 + 0.0713x + 85.657$
$R^2 = 0.9928$

Figure 11.4: Function fitted to Norwegian data on income and happiness. Based on Hellevik 2008

Nevertheless, for the purpose of the exploratory analyses reported here, the quadratic function will be used as it represents a mathematical form that has sometimes been used for utility functions (see Arrow 1965). However, the data have been rescaled so that income starts at the value of 1 and utility takes on values between 0 and 1. It was checked that this rescaling did not change the functional form – it just made the coefficients easier to work with. The function was:

Utility of income $= 0.4283 + 0.0178x - 0.0003x^2$

The linear term is positive and the quadratic term negative, implying that the function increases as a slowing rate, reaches a maximum and then starts declining. The maximum is located far outside the range of the data used to fit the function. The first derivative of the function is 0.0178 – 0.0006x.

Three levels of risk were used: 10 percent (extremely high for most applications), 1 percent and 1 in a million. The latter risk level is very low and represents about 5 deaths per year in Norway (population slightly more than 5 million). There was 117 traffic fatalities in 2015 in Norway.

The values estimated will be shown for three levels of income: 150,000 (low), 550,000 (which is close to GDP per capita) and 1,200,000 (high). Results are shown in Table 11.2.

The estimated value of a statistical life at an income close to the mean income is only between 2.5 and 3 million NOK, considerably lower than the values indicated by valuation studies. However, as noted above, these values are very likely to have a considerable upward bias, since almost all inconsistent patterns of answers, or lexicographic answers tend to inflate the values, and since there is evidence of a large publication bias in the literature.

Table 11.2: Estimated values of a statistical life based on quadratic utility function fitted to Norwegian data Hellevik 2008

	Survival probability		
Income (NOK)	0.900000	0.990000	0.99999
150,000	1,663,000	1,512,000	1,497,000
550,000	2,915,500	2,651,000	2,624,000
1,200,000	11,155,000	10,140,000	10,039,500

The values nevertheless seem quite low and other utility functions have therefore been investigated to see what their implications are regarding the value of a statistical life.

11.5.2 Logarithmic utility functions based on US data 1972-74 and 1994-96

The logarithmic utility functions presented in Figure 11.3 have been applied to estimate the value of a statistical life in Norway. The function fitted to the data for 1972-74 implies that the value of a statistical life is about 15 times annual income at lower levels of income, rising to a factor of about 17 at higher levels of income. Applying multiplicators in this range to the income levels in Table 11.2 produces the estimated values shown in Table 11.3.

These values are higher than those presented in Table 11.2, but, perhaps except for the values referring to the highest income, still lower than most estimates found in the valuation literature. For the middle income level, the values are about four times higher than those presented in Table 11.2.

Table 11.3: Estimated values of a statistical life based on logarithmic utility function fitted to US data 1972-74

	Survival probability		
Income (NOK)	0.900000	0.990000	0.99999
150,000	2,722,800	2,250,000	2,227,500
550,000	9,680,000	8,800,000	8,712,000
1,200,000	22,968,000	20,880,000	20,671,000

There are two principal explanations for the differences. First, the logarithmic utility function is more risk averse than the second degree polynomial. This is most easily understood by comparing the certainty equivalents for the two functions.

A 50-50 lottery between the lowest and highest income in the data provided by Hellevik (2008) gives an expected income of 625,000 NOK per year with an expected utility of 0.565 (on the 0 to 1 scale for utility). An income for certain (the certainty equivalent) giving the same utility is 450,000. For a risk averse utility function the certainty equivalent will by definition have lower value than the expected value of a lottery on income. In this case, the ratio of the certainty equivalent to the expected value is 450,000/625,000 = 0.72.

For the utility function fitted to US data for 1972-74, expected income in a 50-50 lottery between the highest and lowest values is 24,430 US dollars (1996-prices). The certainty equivalent (i.e. the income received for certain that has the same utility as the expected utility of the lottery) is 10,800. The ratio of the certainty equivalent to the expected value is 10,800/24,430 = 0.44. The lower the ratio, the more risk averse the utility function.

The second explanation for the difference in estimates, is that the utility levels reported in the United States, when converted to a 0 to 1 scale are higher than in Norway. It follows from the definition of the value of a statistical life that the higher the utility when alive, the higher will, all else equal, the value of a statistical life be.

Hellevik (2008) presented his data in a form that was not readily usable as an indicator of utility. It could be that the interpretation of these data when converting them to a utility measure was not correct. At any rate, the resulting utility values, ranging from 0.46 to 0.68 when converted to a 0 to 1 scale are surprisingly low. By contrast, the World Happiness Report 2013 (Helliwell, Layard and Sachs 2013), reports a mean population happiness score for Norway of 7.655 according to the Cantril ladder, which ranges from 0 to 10. Converted to a utility scale bounded by 0 and 1, this corresponds to a value of 0.7655, which is higher than any of the values computed from Hellevik's data and more consistent with the utility values reported by the US sample in 1972-74, which ranged between 0.64 and 0.79.

It is therefore likely that the utility values derived from Hellevik are systematically too low for Norway. If the utility values derived from Hellevik are adjusted, so that they range between 0.65 and 0.88, the estimated value of a statistical life increases by 30-45 percent, but remains lower than the values based on US data.

A logarithmic utility function fitted the US data for 1994-96 very well. It implies broadly speaking the same values of a statistical life as the utility function fitted to the 1972-74 data. However, the value of a statistical life was higher in 1994-96 in the upper income deciles than in 1972-74, as it was only in these deciles of the income distribution there was a growth in real income from 1972-74 to 1994-96. A power function fitted the data marginally better than the logarithmic function; hence its implications will be examined.

11.5.3 Power utility function based on US data 1994-96

The power function fitted to the US data for 1994-96 implies that the value of a statistical life is proportional to a multiplicator of income. The value of this multiplicator is the same for all levels of income. The implied values, in Norwegian monetary units, are shown in table 11.4.

Table 11.4: Estimated values of a statistical life based on power utility function fitted to US data 1994-96

	Survival probability		
Income (NOK)	**0.900000**	**0.990000**	**0.99999**
150,000	2,620,000	2,382,000	2,358,000
550,000	9,609,000	8,735,000	8,648,000
1,200,000	20,964,000	19,058,000	18,868,000

The estimated values are close to those obtained by the logarithmic utility function. In this case, therefore, the choice between a logarithmic utility function and a power function did not matter.

11.5.4 A multivariate function for the United States (Graham 2009)

All the utility functions compared so far are simple bivariate functions. They relate income and happiness scores. However, happiness or utility is influenced by many other factors than just income. Graham (2009:119) presents coefficient estimates for factors influencing happiness in the United States between 1972 and 1998. The estimated function included variables such as age, gender, marital status, education, ethnicity, health status and others in addition to income. Income was entered as the natural logarithm of income. The estimated coefficient was 0.163. This is close to the crude estimates reported in Figure 11.3, which were 0.1443 for 1972-74 and 0.1364 for 1994-96, suggesting that potentially confounding factors do not influence the relationship between income and happiness scores very much.

If the estimated coefficient of 0.163 is applied to the 1994-96 US data, appropriately scaled to the data by a constant term, it suggests a little steeper relationship than the crude functions. This means that the value of a statistical life increases more rapidly as income increases. Estimated values of a statistical life, converted to Norwegian currency, are very similar to those reported in Tables 11.3 and 11.4, the principal difference being slightly higher values at the highest income.

This analysis suggests that the relationship between income and happiness is robust to confounding.

11.5.5 A double logarithmic function (Welsch 2006)

The examples of valuation derived from utility functions so far have been based on data from Norway or the United States. The valuation study reported by Welsch (2006) relied on the Eurobarometer life satisfaction data, ranging between 1 and 4. He developed a double logarithmic model in which the (presumably) natural logarithm of life satisfaction was modelled as a function of the natural logarithms of the independent variables (presumably was inserted because he uses the abbreviation log, whereas ln is normally used for the natural logarithm).

Unfortunately, Welsch (2006) does not present all details about his regression model. It contains a constant term and coefficients for countries; these are never presented. The model was fitted to the mean values for all variables, and an ad hoc constant term was added so that the model correctly predicted the mean happiness score. It is, of course, by no means certain that the mean values of all independent variables accurately predict the mean value of the dependent variable, but it was necessary to make this assumption to apply the model. Income was then varied from the lowest to the highest value found and happiness score estimated, keeping all other variables constant. The estimated happiness scores were converted to a 0 to 1 scale. It turned out that the values fitted perfectly to a

power function with constant term 0.1179 and exponent 0.1909. Based on this function, values of a statistical life were derived.

The value of a statistical life was found to be between 5.2 times and 5.8 times income, depending on survival probability (the highest multiplicator for the lowest probability of survival). For an income of NOK 550,000, this corresponds to values between 2,860,000 and 3,190,000 NOK. These values are quite low – much lower than the values reported in studies of willingness-to-pay.

11.5.6 Analysis of a global data set

A global data set has been compiled by merging data from the World Bank global governance data base (Kaufmann, Kraay and Mastruzzi 2010), The world happiness report (Helliwell, Layard and Sachs 2013) and World Health Statistics 2014 (World Health Organization 2014) containing data on several variables that may influence subjective well-being (happiness). After editing to ensure complete data coverage, 124 countries were left. Based on these 124 countries, Figure 11.5 shows the simple bivariate relationship between mean income per capita (US dollars at purchasing power parity) and mean happiness score (Cantril ladder, range from 0 to 10).
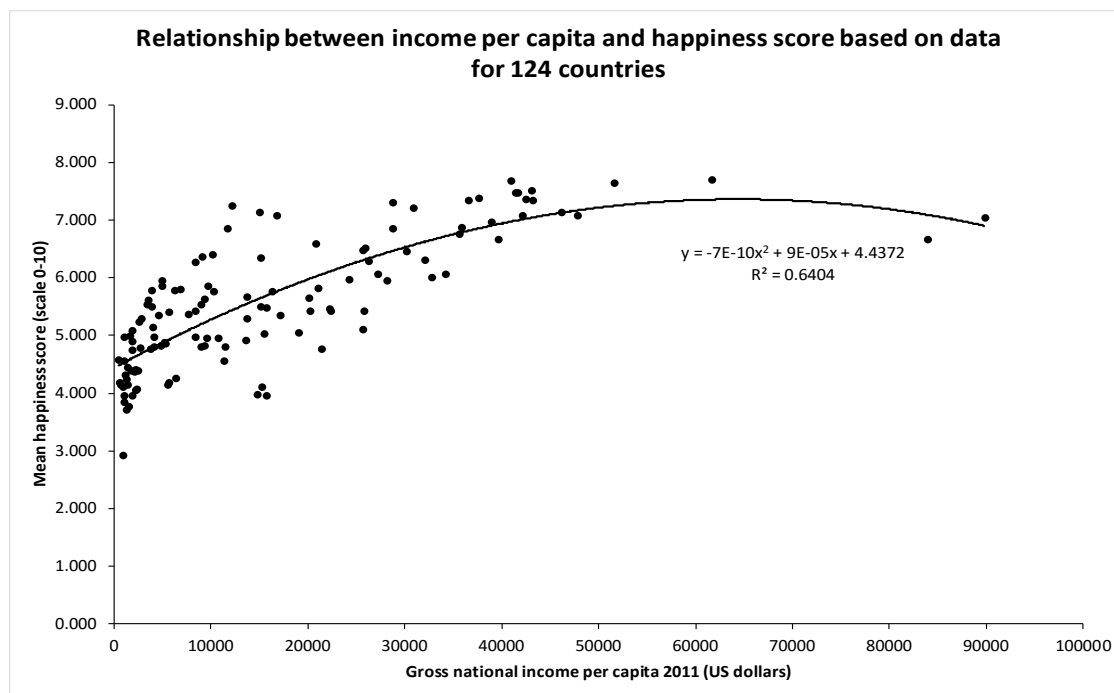
Figure 11.5: Relationship between mean income per capita and happiness score in 124 countries. Income and happiness scores refer to 2010-2012

A second degree polynomial best fits the data. It has a turning point and indicates that happiness starts to decline once income is more than 70,000 dollars per capita. This would imply a negative value of life, which, to say the least, is highly implausible. A multivariate analysis has therefore been made in order to identify more precisely the relationship between income and happiness score. Six different models were developed. In all these models, happiness score was the dependent variable. The independent variables were:

1. Percent rank score for political stability (range 0 to 100; higher scores = higher political stability),
2. Percent rank score for control of corruption (range 0 to 100; higher scores = better control of corruption),
3. Health care expenditures as percentage of GDP in 2011,
4. Marginal propensity to spend on health care assessed in terms of changes from, 2000 to 2011 (this variable is explained in detail below),
5. Gross domestic product per capita in 2011,
6. Gross domestic product per capita in 2001 squared (in models 2 and 5),
7. The natural logarithm of gross domestic product per capita in 2011 (in models 3 and 6),

Models 1-3 included all countries. Models 4-6 only included high-income countries, defined as countries with GDP per capita of at least 10,000 US dollars in 2011.

The marginal propensity to spend on health care was defined as the share of real increase in GDP per capita from 2000 to 2011 that was spent on health care. Consider as an example Norway. GDP per capita increased from 36,363 in 2000 to 61,677 in 2011, an increase (in real terms) of 25,314. Spending on health care increased from 3309 to 6106, an increase of 2797. This increase (2797) makes up 11 percent of the total increase in income (2797/25,314 = 0.110). Marginal propensity to spend on health care is therefore 0.11.

A linear regression model was fitted. To probe for non-linearity in the relationship between income and happiness score, the income squared or log of income variables were included. Results are shown in Table 11.5.

Table 11.5: Regression models of the relationship between income and happiness. Regression coefficients – standard errors in parentheses

Terms	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Constant	4.417 (0.192)	4.386 (0.178)	-0.096 (0.634)	4.501 (0.389)	4.014 (0.438)	-4.250 (2.371)
Political stability	-0.003 (0.004)	-0.004 (0.004)	-0.001 (0.004)	-0.010 (0.006)	-0.011 (0.006)	-0.012 (0.006)
Control of corruption	0.011 (0.005)	0.004 (0.005)	0.007 (0.004)	0.018 (0.008)	0.015 (0.008)	0.016 (0.008)
Health as % of GDP	0.017 (0.028)	0.016 (0.026)	0.038 (0.027)	0.005 (0.076)	0.030 (0.075)	0.022 (0.075)
Marginal propensity	-0.509 (0.516)	-0.526 (0.478)	-0.274 (0.482)	3.437 (4.231)	0.049 (4.383)	1.563 (4.206)
GDP/capita 2011	0.0000354 (0.000)	0.0000865 (0.000)		0.0000261 (0.000)	0.0000788 (0.000)	
GDP/capita squared		-6.597(E-10) (0.000)			-5.727 (E-10) (0.000)	
Ln(GDP/capita)			0.560 (0.079)			0.960 (0.258)
R-squared	0.587	0.649	0.637	0.482	0.522	0.501
N	124	124	124	64	64	64

The coefficients for political stability and control of corruption have the same sign and roughly the same magnitude in all models. These coefficients are not statistically significant. In increasing share of GDP spent on health contributes positively to happiness, whereas the effects of increasing spending on health care appear to be positive for happiness only in high-income countries. GDP per capita is positively associated with income, but the

quadratic term is negative, indicating that the curve becomes flatter at higher income. The natural logarithm of income is also positively related to happiness. Model 2 best fits the data and explains a higher share of variance than any of the other models.

Application of this model indicates a turning point in happiness at high incomes, which implies negative values of life. The same is found when using the results for high-income countries only. This does not make sense and suggests that Arrow's warnings about the use of quadratic utility functions should be heeded.

The logarithmic functions have a marginally poorer fit than the second degree polynomials. Application of the logarithmic function based on data for all countries to Norway produces estimates of the value of a statistical life shown in table 11.6.

Table 11.6: Estimates of the value of a statistical life in Norway based on logarithmic function fitted to data for 124 countries

	Survival probability		
Income (NOK)	**0.900000**	**0.990000**	**0.99999**
150,000	1,689,000	1,536,000	1,520,000
550,000	6,988,000	6,353,000	6,290,000
1,200,000	16,289,000	14,808,000	14,660,000

Application of model 6, based on data for high-income countries only, gives very similar results to those presented in Table 11.6.

These examples show that, in principle, it is possible to specify empirical utility functions and use these functions to estimate the value of a statistical life. However, different data sets and different functions give different estimates of the value of a statistical life. Nevertheless, the differences are considerably smaller than those found in valuation studies relying on established methods. Analysing empirical utility functions may therefore form part of a broad approach employing different methods to try to value safety.

11.5.7 Hedonic adaptation and compensation needed to restore utility

Loewenstein and Ubel (2008) hint that it may a paradox that permanently impaired individuals state happiness, or utility, values that are almost as high as those of healthy individuals. Yet, the permanently impaired are willing to pay substantial amounts to cure the impairment. Perhaps one should not take their apparently successful hedonic adaptation too seriously?

Consider the two utility functions in Figure 11.3 once more. At an income of 11,000 US dollars, the lower function indicates a 2 percent lower level of utility than the upper curve. However, to compensate for this quite small loss, and restore utility to its initial level, income would need to increase from 11,000 to 15,167 US dollars, an increase of almost 38 percent. It is thus not at all paradoxical that even small losses in utility require a disproportionate increase in income to restore utility to its original level. Figure 11.6 indicates how the required compensation can be estimated.
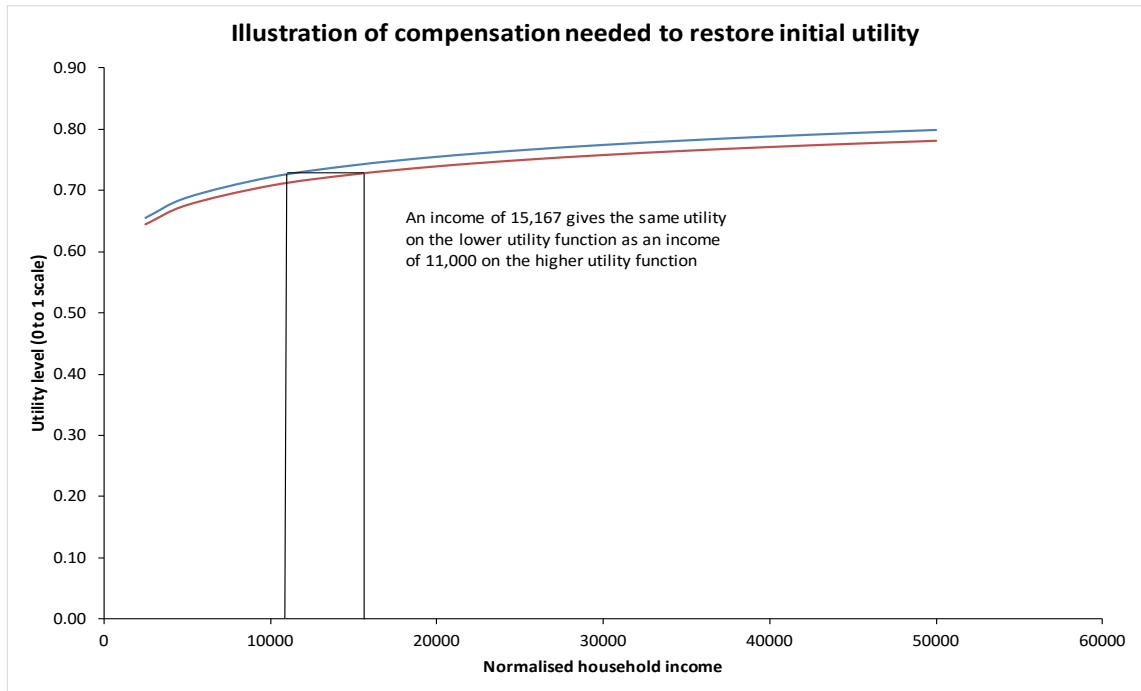
Figure 11.6: Estimation of monetary compensation restoring original level of utility

Move horizontally from the point at 11,000 on the higher utility function until you intersect the lower utility function. Read off the income at that point.

12 Discussion and conclusions

12.1 Discussion

What is a scientific research programme? How can it be identified? How can we know if it is progressive, i.e. produces new knowledge, or degenerative, i.e. tries to account for failures to produce new knowledge? How can we know when a scientific research programme starts and when it ends? How can we determine when a scientific research programme is succeeded by a new programme?

Imre Lakatos (1970, 1978) asked all these questions and indicated how to answer them, mainly be giving examples, mostly from the natural sciences. It did, however, not take long before his ideas were applied to the social sciences, mainly economics, see for example the papers in Latsis (1976). How well do the main ideas of the methodology of scientific research programmes fit to the history of research designed to obtain monetary valuations of life and limb? Is it fruitful to apply elements of the methodology of scientific research programmes to reconstruct and interpret the history of the monetary valuation of life and limb?

Two observations formed the main motivation for trying to apply the methodology of scientific research programmes to research on the monetary valuation of safety. First, as noted in Chapter 1, research on the monetary valuation of safety has produced an enormous diversity of values. Many of these values are anomalous, in the sense that they are partly or fully determined by factors that, according to theory, ought not to influence the values. Still, these values get published in the scientific literature and are taken seriously by many governments, although it must be added that most governments have opted for a conservative interpretation of the literature on the value of life, selecting estimates close to the lower end of the range. Yet, the range of values is enormous and no clear signs can be seen for it to become smaller. Many observers would say that this research has been unsuccessful and wonder why it still goes on.

This leads to the second observation. The methodology of scientific research programmes states that a research programme will continue even if it contains many anomalous results, i.e. results that prima facie reject the hard core of the programme. The hard core consists of the basic assumptions made when developing the key theoretical propositions of the programme. Valuation research is based on mainstream neoclassic economic theory. The hard core of this theory is that hypotheses should be derived by assuming that individuals are subjectively rational utility maximisers. This core assumption is formal only; it is empirically empty and only intended to guide researchers to state their hypotheses in a form that lends itself to mathematical analysis. To fill it with content, researchers have to make more specific assumptions about individual preferences and the utility functions representing these preferences.

Researchers obviously have considerable freedom to specify preferences and utility functions. The hard core imposes few restrictions on the content of the hypotheses. This means that it will sometimes be possible to reinterpret an apparently anomalous finding by proposing a new utility function, according to which the anomalous finding no longer is anomalous. Anomalies that cannot be made to conform with the hard core can be recast as positive heuristics calling for development of research methods or techniques for data

analysis. If methods can be developed to design studies that do not produce anomalous findings, again the anomalies will go away.

Still, one wonders why the existence of a lot of anomalies does not make researchers abandon a research programme. The explanation, according to Lakatos, is that empirical falsifications by themselves are never enough to reject a theory. A theory is upheld even in the case of massive anomalies, unless a new theory is developed which accounts for all the anomalies as well as all the empirically verified content of the old theory.

There is no theory which has replaced neoclassic economic theory. Behavioural economics continues to grow, but its main focus has been on describing and explaining violations of the neoclassic model of utility maximisation, not on developing an alternative normative of prescriptive framework. Nobody argues that choices that violate rationality as blatantly as the choices framed by Tversky and Kahneman (1986) can be given any normative status.

Indeed, the real issue is whether all results of valuation research are anomalous or whether this body of research contains any results at all that make sense. The list of anomalies that have been found again and again is awesome. For stated preference studies the main anomalies include (only those discussed extensively in the literature are listed):

1. Insensitivity to scope (willingness-to-pay does not vary "sufficiently" with the size of the change in risk).
2. Disparity, often very large, between willingness-to-accept and willingness-to-pay.
3. Inconsistent relationship between respondent age and willingness-to-pay (different studies find different shapes of the relationship).
4. Hypothetical bias, since no real payment is involved, amounts can be greatly overstated.
5. Starting point bias in studies relying on iterative bidding.
6. Payment range bias in studies relying on payment cards.
7. Payment vehicle bias in studies using taxes as the means of paying.
8. Strategic answers when safety is provided as a public good (free-riding).
9. Lexicographic choices in stated choice tasks.
10. Inconsistent choices in stated choice tasks.
11. Discrepancy between actual and perceived risk in studies that have investigated perceived risk.

It is not possible to find a single stated preference valuation study which is not affected by one or more of these problems. Each of them represents a sufficient reason for rejecting the findings of the study.

Within the valuation research programme, the first three anomalies listed above are no longer necessarily regarded as anomalies, if one accepts the directionally bounded utility functions proposed by Amiran and Hagen (2003, 2010) and the theory of the relationship between age and willingness-to-pay proposed by Johansson (2002). Moreover, the relationship between willingness-to-pay and the nature of the good, whether it is private or public (point 8 on the list above), is theoretically indeterminate (Johannesson et al. 1996). Thus, four of the anomalies have been reinterpreted as normal findings by reformulating theory.

Anomalies 4 through 7 on the list above are closely related to the design of contingent valuation studies. Although, as one would expect according to the methodology of scientific research programmes, methodological studies have been made in order to overcome or at least reduce these anomalies, this research as only been partly successful. For example, although one can design studies so as to reduce hypothetical bias, it is not entirely eliminated. It is likely that most contingent valuation studies to some extent reflect hypothetical bias and overstate true willingness-to-pay.

Stated choices are somewhat less hypothetical. After all, respondents are asked to make a choice, ideally speaking one that they should be familiar with. Yet, the choices do not have any real economic consequences. Respondents do not have to pay anything. It is tempting to make "idealistic" choices, for example by always choosing the safer alternative. One wonders whether respondents really are so idealistic in the real world. Do they always choose the safer road? Most probably not. In many cases the safer road will be a motorway where a toll must be paid. In such cases, some road users will be tempted to take another and in most cases less safe route where there is no payment. This was found in studies many years ago (Kristiansen 1978). In other cases, road users simply lack the information to choose the safest road. They cannot be expected to know the accident rate of every road or to trust official statistics about accident rates. There is, after all, incomplete accident reporting in official statistics and the level of reporting cannot be assumed to be the same for every road. Hence, official estimates of accident rate will have an unknown and variable degree of bias.

Lexicographic choices have the same effect as hypothetical bias. They lead to an overstatement of real willingness-to-pay. It is perhaps a little more surprising that even inconsistent choices are associated with inflated valuations. Lexicographic and inconsistent choices are widespread in stated choice valuation studies and cast serious doubt on the ability of respondents to make trade-offs. Both anomalies appear to lead to inflated valuations.

In view of these problems, it is tempting to argue that the revealed preference approach gives more reliable estimates of the value of a statistical life. Yet, even this approach is associated with huge problems. The choices studied rarely involve the purchase of a safety product in a direct sense; rather individuals are choosing between complex options having safety as one of their attributes. The number of relevant attributes, as well as the number of relevant options, would normally be very large, which would make a perfectly rational choice difficult (Bruni and Sugden 2007). Moreover, choices are likely to reflect self-selection, meaning that those who choose, for example, a high-risk occupation are not representative of the general population. More specifically, the following problems are associated with using compensating wage differentials to estimates the value of a statistical life:

1. Workers self-select into high risk occupations. These workers probably have a different attitude to risk than the general population.
2. No study of compensating wage differentials has controlled for all potentially confounding variables about which there is agreement that controlling for them is necessary. In particular, the potential for self-selection bias has not been controlled for by any study.
3. Estimates of the risk faced by each worker are crude in many studies and based on mean values only. It is very likely that more precise estimates could be developed by applying, for example, the empirical Bayes method.
4. Risks are in many cases likely to be partly endogenous, i.e. workers do not face a given level of risk that they cannot influence, but have some degree of control over the risk.
5. Measures influencing risks may be public (collective) goods, the preferences for which cannot be revealed by individual behaviour. An individual may prefer the public good risk reduction to be provided, but there is no opportunity to express this preference through market behaviour.
6. There is evidence that wage formation reflects discriminatory practices, such as paying recent immigrants less than national citizens or paying women less than men.

7. The labour market is highly segmented and reflects social inequalities. Thus, even workers in prestigious low-risk occupations earn a risk-premium, which, although it may be lower than the risk-premium in high-risk occupations, still leads to an inverse relationship between risk and the value of a statistical life: the lower the risk, the higher the value of a statistical life (Viscusi 2010).

For these reasons, it is highly problematic to use compensating wage differentials to value risk reductions. Moreover, the values show great dispersion, although not as great as the stated preference studies. Finally, there is evidence of publication bias. There are, to be sure, a few studies of more directly safety-related behaviour, such as wearing seat belts, wearing bicycle helmets and buying safer cars. Since the behaviour is in all cases voluntary, self-selection bias is very likely to occur. Interestingly, the values of a statistical life derived from studies of safety-related consumer behaviour (Blomquist 1979, Blomquist, Miller and Levy 1996, Andersson 2005) tend to be lower than those derived from compensating wage differentials and from stated preference studies.

As noted in Chapter 9, studies estimating the value of a statistical life are likely to be affected by considerable publication bias. This bias comes on top of other sources of bias in primary studies leading to inflated estimates. It is therefore likely that published values of a statistical life are greatly inflated, perhaps by a factor of 3-10. The tentative estimates based on utility functions in Chapter 11 are much lower than most estimated based on willingness-to-pay studies, lending support to the conjecture that published values are inflated.

12.2 Conclusions

To summarise, the main results and conclusions from the research presented in this report can be stated as follows:

1. The monetary valuation of life and limb was launched as a scientific research programme around 1970. Launching research on this topic as a scientific research programme was accomplished by incorporating it into the hard core of neoclassic economic theory, i.e., by arguing that the valuation of safety should be treated as the valuation of any other consumer good. This implied that the valuation of life and limb should be based on the willingness-to-pay for reducing risks to life and health.

2. In the initial phase of the programme, research focused on finding methods for determining willingness-to-pay and on developing theory about factors influencing willingness-to-pay. Two main classes of methods were developed: stated preference methods and revealed preference methods. Over time, two distinct research traditions developed, with most European studies relying on stated preference methods and most North American studies relying on revealed preference methods.

3. Both theoretical and empirical research had a slow start, but around 1980 quick progress was made both in terms of theoretical developments and empirical studies. The period from about 1980 until around 1990 has the characteristics of a progressive phase of the research programme.

4. From around 1990, results that were regarded as anomalies turned up with increasing frequency. The initial reaction to the anomalous findings was to develop new methods for eliciting preferences. These methodological innovations were only partly successful and some widely discussed anomalies, such as insensitivity to scope, persisted in study after study. By the late 1990s, the burden of anomalies

associated with the contingent valuation method had become so great that many researchers switched to another stated preference method, the stated choice method. Meanwhile, the revealed preference method, in particular the study of compensating wage differentials, was strongly criticised in 1996, but the criticism did not convince the leading proponents of the method to abandon it.

5. As the number of empirical estimates of the value of a statistical life increased, it became clear that these estimates varied enormously. Starting from about 2000, several meta-analyses of estimates of the value of a statistical life have been made. These meta-analyses have had a primary focus on identifying sources of the enormous variation in estimates of the value of a statistical life. The most comprehensive of the meta-analyses, based on 856 estimates of the value of a statistical life, identified income and the size of the change in risk as the principal determinants of the value of a statistical life. The coefficient for the size of the change in risk indicated insensitivity to scope, i.e. when the change in risk doubled, the valuation of it increased by much less. Meta-analyses that have tested for publication bias indicate that the published estimates of the value of a statistical life may be subject to massive publication bias.

6. Insensitivity to scope is a major problem when applying the estimates in cost-benefit analyses. It implies that the value of a statistical life becomes greater the smaller the risk reduction, implying that money is best spent on small reductions from low initial levels of risk, rather than on major reductions from high initial levels of risk. Besides, even for a given risk reduction, the values found in the literature vary considerably, as a result of, for example, differences in income.

7. There has been a noticeable change of focus as research on the monetary valuation of life and limb has progressed. In the early phase, the objective was clearly to obtain a single estimate of the value of saving a life, so that this uniform value could be applied in all sectors of society, to support an efficient allocation of public money. However, as the great dispersion in values became more and more apparent, focus shifted to discussing whether there are legitimate sources for varying the value of saving a life and how large such a variation might be. Thus, prominent economists working in the field have argued that the value of saving a life ought to vary with income.

8. Some results that were for a long time regarded as anomalous have been re-interpreted as normal results by developing new models of the utility functions that form the basis of stated preferences. In particular, according to directionally bounded utility functions, insensitivity to score is not an anomaly, but something to be expected. However, as noted above, insensitivity to scope makes it very difficult to apply the values in cost-benefit analyses.

9. There is currently an extremely large dispersion of values of saving a life in the literature. It is fair to say that all studies are affected by problems that are severe enough to reject their findings. In this report, therefore, it is concluded that the research performed until now for the purpose of estimating the value of saving a life must, as a whole, be rejected. Were one to conclude the opposite, that these studies should be taken seriously, one would face the almost hopeless task of selecting the best values from among the many hundreds of estimates that range from almost zero to more than 100 million US dollars. One might try to reduce the scale of this problem by identifying the "best" studies. Yet, even among the best studies there is a large dispersion in estimates of the value of saving a life.

10. If one accepts the necessity and inevitability of valuing lives in monetary terms, as argued in Chapter 2, and if current valuation studies must be rejected, the problem arises of how best to obtain monetary valuations of life and limb. Some approaches

were discussed in Chapter 11. The first of these, relying on quality adjusted life years, sounds attractive, but turns out, on closer examination, to raise serious problems. It is therefore not recommended. The second approach, the capability approach introduced by Amartya Sen, is not really applicable to the valuation of non-market goods, and necessarily involves paternalism in any practical application. A third approach is to fit utility functions to life-satisfaction data, which are widely available in many countries. There is not full agreement among economists regarding the interpretation of life-satisfaction data as an approximation to a utility function. If, however, such an interpretation is accepted, obtaining the value of a statistical life has a closed-form solution, based on the standard formula for the value of a statistical life of a rational utility maximiser. A sample of utility functions were applied. They did not give identical estimates of the value of a statistical life, but the range was very much smaller than in the valuation literature.

References

Abraham, C., Thedié, J. 1960. Le prix d'une vie humaine dans les decisions économique. Revue Française de Recherche Opérationelle, 20, 157-167.

Adler, M. D., Posner, E. A. Happiness research and cost-benefit analysis. Journal of Legal Studies, 37, 253-292.

Amiran, E. Y., Hagen, D. A. 2003. Willingness-to-pay and willingness to accept : How much can they differ? Comment. American Economic Review, 93, 458-463.

Amiran, E. Y., Hagen, D. A. 2010. The scope trials: Variation in sensitivity to scope and WTP with directionally bounded utility functions. Journal of Environmental Economics and Management, 59, 293-301.

Anderson, E. 1993. Value in ethics and economics. Cambridge, Harvard University Press.

Andersson, H. 2005. Willingness-to-pay for a reduction in road mortality risk. Experience from Sweden. PhD dissertation. Lund Economic Studies 126. Lund, Lund University.

Andersson, H., Lundborg, P. 2007. Perception of own death risk. An analysis of road-traffic and overall mortality risks. Journal of Risk and Uncertainty, 34, 67-84.

Ariely, D., Loewenstein, G., Prelec, D. 2003. "Coherent arbitrariness": stable demand curves without stable preferences. The Quarterly Journal of Economics, 118, 73-105.

Arrow, K. J. 1965. Aspects of the theory of risk bearing. Helsinki, The Yrjö Jahnsson Foundation.

Arrow, K. J., Solow, R., Leamer, E., Portney, P., Radner, R., Schuman, H. 1993. Report of the NOAA panel on contingent valuation. Federal Register, 58, 4601-4614.

Arthur, W. B. 1981. The economics of risks to life. American Economic Review, 71, 54-64.

Atkinson, S. E., Halvorsen, R. 1991. The valuation of risks to life: evidence from the market for automobiles. Review of Economics and Statistics, 73, 133-136.

Aven, T. 2010. Misconceptions of risk. Chichester, John Wiley and Sons.

Aven, T. 2014. Risk, surprises and black swans. Fundamental ideas and concepts in risk assessment and risk management. London, Routledge.

Baker, R., Chilton, S., Jones-Lee, M., Metcalf, H. 2008. Valuing lives equally: Defensible premise or unwarranted compromise? Journal of Risk and Uncertainty, 36, 125-138.

Baker, R., Chilton, S. M., Jones-Lee, M. W., Metcalf, H. R. T. 2009. Valuing lives equally in a benefit cost analysis of safety projects: A method to reconcile theory and practice. Safety Science, 47, 813-816.

Beattie, J. Covey, J., Dolan, P., Hopkins, L., Jones-Lee, M., Loomes, G., Pidgeon, N., Robinson, A., Spencer, A. 1998. On the contingent valuation of safety and the safety of contingent valuation: Part 1 – Caveat Investigator. Journal of Risk and Uncertainty, 17, 5-25.

Becker, G. S. 1964. Human capital. A theoretical and empirical analysis, with special reference to education. Chicago, The University of Chicago Press.

Bellavance, F., Dionne, G., Lebeau, M. 2009. The value of a statistical life: A meta-analysis with a mixed effects regression model. Journal of Health Economics, 28, 444-464.

Ben-Akiva, M., Lerman, S. R. 1985. Discrete choice analysis. Theory and application to travel demand. Cambridge, MIT Press.

Berg, N. 2003. Normative behavioural economics. Journal of Socio-Economics, 32, 411-427.

Bergstrom, T. C. 1982. When is a man's life worth more than his human capital? In Jones-Lee, M. W. (Ed): The value of life and safety, 3-26. Amsterdam, North-Holland Publishing Company.

Blackorby, C., Donaldson, D. 1986. Can risk-benefit analysis provide consistent policy evaluations of projects involving loss of life? Economic Journal, 96, 758-773.

Blaeij, A. de 2003. The value of a statistical life in road safety. Stated preferences methodologies and empirical estimates for the Netherlands. PhD dissertation. Amsterdam, Tinbergen institute and Vrije Universiteit Amsterdam.

Blaeij, A. de, Florax, R. J. G. M., Rietveld, P., Verhoef, E. 2003. The value of statistical life in road safety: a meta-analysis. Accident Analysis and Prevention, 35, 973-986.

Blomquist. G. C. 1979. Value of life saving. Implications of consumption activity. Journal of Political Economy, 87, 540-558.

Blomquist. G. C. 1982. Estimating the value of life and safety: recent developments. In Jones-Lee, M. W. (Ed): The value of life and safety, 27-40. Amsterdam, North-Holland Publishing Company.

Blomquist, G. C., Miller, T. R., Levy, D. T. 1996. Values of risk reduction implied by motorist use of protection equipment. Journal of Transport Economics and Policy, 30, 55-66.

Bowling, A. 2005. Measuring health, A review of quality of life measurement scales. Third edition. Maidenhead, Berkshire, Open University Press.

Brabander, B. de 2006. Valuing the reduced risk of road accidents. Empirical estimates for Flanders based on stated preference methods. PhD dissertation. Universiteit Hasselt, Faculteit Toegepaste Economische Wetenschappen.

Brickman, P., Coates, D., Janoff-Bulman, R. 1978. Lottery winners and accident victims: Is happiness relative? Journal of Personality and Social Psychology, 36, 917-927.

Broome, J. 1982. Uncertainty in welfare economics, and the value of life. In Jones-Lee, M. W. (Ed): The value of life and safety, 201-216. Amsterdam, North-Holland Publishing Company.

Bruni, L., Comim, F., Pugno, M. (Eds). 2008. Capabilities and Happiness. Oxford, Oxford University Press.

Bruni, L., Sugden, R. 2007. The road not taken: how psychology was removed from economics, and how it might be brought back. The Economic Journal, 117, 146-173.

Camerer, C., Loewenstein, G., Rabin, M. (Eds) 2004. Advances in behavioural economics. New York, Russell Sage Foundation.

Carson, R. T., Groves, T. 2007. Incentive and informational properties of preference questions. Environmental and Resource Economics, 37, 181-210.

Cartwright, E. 2011. Behavioural economics. New York, Routledge.

Carthy, T., Chilton, S., Covey, J., Hopkins, L., Jones-Lee, M., Loomes, G., Pidgeon, N., Spencer, A. 1999. On 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.

Christensen, J. (Red). 1988. Nordisk seminar om ulykkesomkostninger 18.-20. maj 1988. Proceedings publisert av Akademiet for de tekniske videnskaber, Risikoutvalget, Lyngby.

Clark, A. E., Frijters, P., Shields, M. A. 2008. Relative income, happiness, and utility: An explanation for the Easterlin paradox and other puzzles. Journal of Economic Literature, 46, 95-144.

Connolly, T., Arkes, H. R., Hammond, K. R. (Eds). 2000. Judgment and Decision Making. An interdisciplinary reader. Cambridge, Cambridge University Press.

Dawson, R. F. F. 1967. Cost of road accidents in Great Britain. RRL Report LR 79. Crowthorne, Berkshire, Road Research Laboratory.

Dehez, P., Drèze, J. H. 1982. State-dependent utility, the demand for insurance and the value of safety. In Jones-Lee, M. W. (Ed): The value of life and safety, 41-65. Amsterdam, North-Holland Publishing Company.

Desaigues, B., Rabl, A. 1995. Reference values for human life: An econometric analysis of contingent valuation in France. In Schwab Christe, N. G and Soguel, N. C. (Eds), Contingent valuation, transport safety and the value of life, 85-112. Boston, Kluwer Academic Publishers.

Diener, E., Ng, W., Harter, J., Arora, R. 2010. Wealth and happiness across the world: Material prosperity predicts life evaluation, whereas psychosocial prosperity predicts positive feeling. Journal of Personality and Social Psychology, 99, 52-61.

Dionne, G., Lanoie, P. 2004. Public choice about the value of a statistical life for cost-benefit analyses. Journal of Transport Economics and Policy, 38, 247-274.

Dolan, P., Kahneman, D. 2008. Interpretations of utility and their implications for the valuation of health. The Economic Journal, 118, 215-234.

Dolan, P., Peasgood, T., White, M. 2008. Do we really know what makes us happy? A review of the economic literature on the factors associated with subjective well-being. Journal of Economic Psychology, 29, 94-122.

Dorman, P. 1996. Markets and mortality. Economics, dangerous work, and the value of human life. New York, Cambridge University Press.

Doucouliagos, C., Stanley, T. D., Giles, M. 2012. Are estimates of the value of a statistical life exaggerated ? Journal of Health Economics, 31, 197-206.

Downs, A. 1957. An economic theory of democracy. New York, Harper and Row.

Drèze, J. 1962. L'utilité sociale d-une vie humaine. Revue Française de Recherche Opérationelle, 22, 139-155.

Dubourg, W. R. 1995. Are preferences for safety too imprecise for contingent valuation? In Schwab Christe, N. G. and Soguel, N. C. (Eds) : Contingent valuation, transport safety and the value of life, 137-156. Boston, Kluwer Academic Publishers.

Dubourg, W. R., Jones-Lee, M. W., Loomes, G. 1997. Imprecise preferences and survey design in contingent valuation. Economica, 64, 681-702.

Duval, S. 2005. The trim and fill method. In Rothstein, H., Sutton, A. J., Borenstein, M. (Eds): Publication bias in meta-analysis: Prevention, assessment and adjustments, 127-144. Chichester: John Wiley and Sons.

Duval, S., Tweedie, R. 2000A. Trim and fill: a simple funnel plot based method of testing and adjusting for publication bias in meta-analysis. Journal of the American Statistical Association, 95, 89-98.

Duval, S., Tweedie, R. 2000B. A non-parametric trim and fill method of assessing publication bias in meta-analysis. Biometrics, 56, 455-463.

Easterlin, R. A. 2010. Happiness, growth, and the life cycle. Oxford, Oxford University Press.

Eeckhoudt, L. R., Hammitt, J. K. 2001. Background risks and the value of a statistical life. Journal of Risk and Uncertainty, 23, 261-279.

Eeckhoudt, L. R., Hammitt, J. K. 2004. Does risk aversion increase the value of mortality risk ? Journal of Environmental Economics and Management, 47, 13-29.

Ellsberg, D. 1961. Risk, ambiguity and the Savage axioms. The Quarterly Journal of Economics, 75, 643-669.

Elster, J. 2007. Explaining social behavior. More nuts and bolts for the social sciences. Cambridge, Cambridge University Press.

Elvebakk, B. 2015. Paternalism and acceptability in road safety work. Safety Science, 79, 298-304.

Elvik, R. 1993. Økonomisk verdsetting av velferdstap ved trafikkulykker. Dokumentasjonsrapport. Rapport 203. Oslo, Transportøkonomisk institutt.

Elvik, R. 1995. The validity of using health state indexes in measuring the consequences of traffic injury for public health. Social Science and Medicine, 40, 1385-1398.

Elvik, R. 2003. How would setting policy priorities according to cost-benefit analysis affect the provision of road safety? Accident Analysis and Prevention, 35, 557-570.

Elvik, R. 2005. A catalogue of risks of accidental death in various activities. Working paper 1661/2005. Oslo, Institute of Transport Economics.

Elvik, R. 2006. Major transportation accidents in Norway: assessing long-term frequency and priorities for prevention. Transportation Research Record, 1969, 101-106.

Elvik, R. 2010A. Strengthening incentives for efficient road safety policy priorities: The roles of cost-benefit analysis and road pricing. Safety Science, 48, 1189-1196.

Elvik, R. 2010B. A restatement of the case for speed limits. Transport Policy, 17, 196-204.

Elvik, R. 2012. Cost-benefit analysis. Text written for European Road Safety Observatory (ERSO) website.

Elvik, R. 2013A. Can it be true that most drivers are safer than the average driver? Accident Analysis and Prevention, 59, 301-308.

Elvik, R. 2013B. Paradoxes of rationality in road safety policy. Research in Transportation Economics, 43, 62-70.

Elvik, R. 2013C. International transferability of accident modification functions for horizontal curves. Accident Analysis and Prevention, 59, 487-496.

Elvik, R. 2014. Problems in determining the optimal use of road safety measures. Research in Transportation Economics, 47, 27-36.

Elvik, R. 2016A. A theoretical perspective on road safety communication campaigns. Accident Analysis and Prevention (forthcoming).

Elvik, R. 2016B. Meta-analytic methods. Chapter 19 in Lord, D., Washington, S. (Eds) Safe mobility: challenges, methodology and solutions (Volume VIII) (forthcoming).

Elvik, R. 2016C. Safety-in-numbers: estimates based on a sample of pedestrian crossings in Norway. Accident Analysis and Prevention, 91, 175-182.

Elvik, R. 2016D. Færre trafikkulykker betyr ikke sparte milliarder. Kommentar i Samferdsel (nettutgaven), publisert 4.2.2016.

Ferreira, S., Moro, M. 2010. On the use of subjective well-being data for environmental valuation. Environmental and Resource Economics, 46, 249-273.

Fifer, S., Rose, J., Greaves, S. 2014. Hypothetical bias in Stated Choice Experiments: Is it a problem? And if so, how do we deal with it? Transportation Research Part A, 61, 164-177.

Finkelstein, A., Luttmer, E. F. P., Notowidigdo, M. J. 2013. What good is wealth without health? The effect of health on the marginal utility of consumption. Journal of the European Economic Association, 11, 221-258.

Fischhoff, B. 1991.Value elicitation. Is there anything in there? American Psychologist, 46, 835-847.

Fischhoff, B., Slovic, P., Lichtenstein, S., Read, S., Combs, B. 1978. How safe is safe enough? A psychometric study of attitudes toward technological risks and benefits. Policy Sciences, 9, 127-152.

Frank, R. H. 2000. Why is cost-benefit analysis so controversial? Journal of Legal Studies, 29, 913-930. Reprinted in Hausman, D. M. (Ed): The philosophy of economics: an anthology, 251-269. New York, Cambridge University Press, 2008.

Frey, B., Stutzer, A. What can economists learn from happiness research? Journal of Economic Literature, XL, 402-435, 2002A.

Frey, B., Stutzer, A. Happiness and economics. How the economy and institutions affect human well-being. Princeton, Princeton University Press, 2002B.

Friedman, M. 1953. The methodology of positive economics. In: Essays in positive economics. Chicago, The University of Chicago Press.

Fromm, G. 1968. Comment to "The life you save may be your own" In Chase, S. B. (Ed). Problems in public expenditure analysis, 166-176. Washington D. C., the Brookings Institution.

Garen, J. 1988. Compensating wage differentials and the endogeneity of job riskiness. Review of Economics and Statistics, 70, 9-16.

Gilovich, T., Griffin, D., Kahneman, D. (Eds). 2002. Heuristics and biases. The psychology of intuitive judgment. Cambridge, Cambridge University Press.

Graham, C. 2009. Happiness around the world. The paradox of happy peasants and miserable millionaires. Oxford, Oxford University Press.

Guria, J., Jones, W., Jones-Lee, M., Keall, M., Leung, J., Loomes, G. 1999. The values of statistical life and prevention of injuries in New Zealand. Wellington, Land Transport Safety Authority.

Hammerton, M., Jones-Lee, M. W., Abbott, V. 1982. The consistency and coherence of attitudes to physical risk. Some empirical evidence. Journal of Transport Economics and Policy, 16, 181-199.

Hammond, P. J. 1982. Utilitarianism, uncertainty and information. In Sen A., Williams, B. (Eds), Utilitarianism and beyond, 85-102. Cambridge. Cambridge University Press.

Hammitt, J. K. 2002. QALYs versus WTP. Risk Analysis, 22, 985-1001.

Hands, D. W. 1993. Popper and Lakatos in economic methodology. In Mäki, U., Gustafsson, B., Knudsen, C. (Eds), Rationality, institutions and economic methodology, 61-75. Reprinted in Hausman, D. M. (Ed), The philosophy of economics: an anthology, 188-203. Cambridge, Cambridge University Press, 2008.

Hanemann, W. M. 1991. Willingness-to-pay and willingness to accept: How much can they differ? American Economic Review, 81, 635-647.

Haninger, K., Hammitt, J. K. 2006. Willingness-to-pay for quality-adjusted life years: Empirical inconsistency between cost-effectiveness analysis and economic welfare

theory. Unpublished manuscript dated February 2006. Center for Risk Analysis, Harvard University.

Hansson, S. O. 2007. Risk and ethics. Three approaches. In Lewens, T. (Ed): Risk. Philosophical Perspectives, 21-35. London, Routledge.

Harrison, G. W., Rutström, E. 2008. Experimental evidence on the existence of hypothetical bias in value elicitation methods. Chapter 81, 753-766, Handbook of Experimental Economics Results, Volume 1.

Hauer, E. 1997. Observational before-after studies in road safety. Oxford, Elsevier science.

Hauer, E. 2011A. Computing what the public wants: some issues in road safety cost-benefit analysis. Accident Analysis and Prevention, 43, 151-164.

Hauer, E. 2011B. Cost-effectiveness in road safety – when is it a way out. Unpublished manuscript prepared for the Transportation Board Annual Meeting.

Hauer E. 2015. The art of regression modelling in road safety. Heidelberg, Springer.

Haukeland, J. V. 1991. Velferdstap ved trafikkulykker. Rapport 92. Oslo, Transportøkonomisk institutt.

Hausman, D. M. (Ed) 2008. The philosophy of economics. An anthology. Third edition. New York, Cambridge University Press.

Hausman, J. (Ed). 1993. Contingent valuation. A critical assessment. Amsterdam, Elsevier Science.

Hausman, J. 2012. Contingent valuation: From dubious to hopeless. Journal of Economic Perspectives, 26, 43-56.

Hellevik, O. 2008. Jakten på den norske lykken. Norsk monitor 1985-2005. Oslo, Universitetsforlaget.

Helliwell, J., Layard, R., Sachs, J. 2013. World Happiness Report 2013. New York, UN Sustainable Development Solutions Network.

Hersch, J., Viscusi, W. K. 1990. Cigarette smoking, seatbelt use, and differences in wage-risk tradeoffs. The Journal of Human Resources, 25, 202-227.

Hess, S. 2010. Conditional parameter estimates from Mixed Logit models: distributional assumptions and a free software tool. Journal of Choice Modelling, 3, 134-152.

Hills, P. J., Jones-Lee, M. W. 1983. The role of safety in highway investment appraisal for developing countries. Accident Analysis and Prevention, 15, 355-369.

Hoehn, J. P., Randall, A. 1987. A satisfactory benefit cost indicator from contingent valuation. Journal of Environmental Economics and Management, 14, 226-247.

Hoel, M. 2003. Er verdien av liv høyere for de rike og friske enn for de fattige og syke? Norsk Økonomisk Tidsskrift, 117, 77-88.

Horowitz, J. K., McConnell, K. E. 2001. A review of WTA/WTP studies. Journal of Environmental Economics and Management, 44, 426-447.

Høye, A. 2014. Utvikling av ulykkesmodeller for ulykker på riks- og fylkesvegnettet i Norge. Rapport 1323. Oslo, Transportøkonomisk institutt.

Jago, L. 2002. The Northern Lights. New York, Vintage books.

Johannesson, M., Johansson, P-O., O'Conor, R. M. 1996. The value of private safety versus the value of public safety. Journal of Risk and Uncertainty, 13, 263-275.

Johansson, P-O. 2002. On the definition and age-dependency of the value of a statistical life. Journal of Risk and Uncertainty, 25, 251-263.

Jones-Lee, M. W. 1974. The value of changes in the probability of death or injury. Journal of Political Economy, 82, 835-849.

Jones-Lee, M. W. 1976. The value of life: an economic analysis. London, Martin Robertson.

Jones-Lee, M. W. (Editor) 1982. The value of life and safety. Proceedings of a conference held by the Geneva Association. Amsterdam, North-Holland Publishing Company.

Jones-Lee, M. W. 1985. The value of life and safety: A survey of recent developments. The Geneva Papers on Risk and Insurance, 10, 141-173.

Jones-Lee, M. W. 1989. The economics of safety and physical risk. Oxford, Basil Blackwell.

Jones-Lee, M. W. 1991. Altruism and the value of other people's safety. Journal of Risk and Uncertainty, 4, 213-219.

Jones-Lee, M. W. 1992. Paternalistic altruism and the value of statistical life. The Economic Journal, 102, 80-90.

Jones-Lee, M. W., Hammerton, M., Abbott. V. 1983. The value of transport safety: Results of a national sample survey. Report to the Department of Transport. Newcastle, University of Newcastle.

Jones-Lee, M. W., Hammerton, M., Philips, P. R. 1985. The value of safety : results of a national sample survey. Economic Journal, 95, 49-72.

Junghard, O. Linjær krympning i trafikolycksmodeller. VTI-meddelande 662. Linköping, Väg- och Trafikinstitutet.

Kahneman, D., Diener, E., Schwartz, N. (Eds). 1999. Well-being: The foundations of hedonic psychology. New York, Russell Sage Foundation.

Kahneman, D., Krueger, A. B. 2006. Developments in the measurement of subjective well-being. Journal of Economic Perspectives, 20, 3-24.

Kahneman, D., Ritov, I., Schkade, D. 1999. Economic preferences or attitude expressions?: An analysis of dollar responses to public issues. Journal of Risk and Uncertainty, 19, 203-235.

Kahneman, D., Tversky, A., 1979. Prospect Theory: An Analysis of Decision under Risk. Econometrica, 47, 263-292.

Kahneman, D., Tversky, A. (Eds). 2000. Choices, values and frames. Cambridge, Cambridge University Press.

Kaplow, L. 2005. The value of a statistical life and the coefficient of relative risk aversion. Journal of Risk and Uncertainty, 31, 23-34.

Kaufmann, D., Kraay, A., Mastruzzi, M. 2010. The worldwide governance indicators. Methodology and analytical issues. Policy Research Working Paper 5430. Washington D.C., The World Bank, Development Research Group, Macroeconomics and Growth Team.

Keen, S. 2011. Debunking economics. The naked emperor dethroned? London, Zed books.

Keeney, R. L. 1990. Mortality Risks Induced by Economic Expenditures. Risk Analysis, 10, 147-159.

Keeney, R. L. 1994. Mortality Risks Induced by the Costs of Regulations. Journal of Risk and Uncertainty, 8, 95-110.

Keeney, R. L. 1997. Estimating Fatalities Induced by the Economic Costs of Regulations. Journal of Risk and Uncertainty, 14, 5-23.

Kidholm, K. 1995. Betalingsviljen for øget trafiksikkerhed: resultater af en dansk interviewundersøgelse. PhD afhandling. Odense, Odense Universitet.

Kluve, J., Schaffner, S. 2008. The value of life in Europe – a meta-analysis. Sozialer Fortschritt, 57, 279-287.

Kniesner, T. J., Viscusi, W. K., Ziliak, J. P. 2014. Willingness to accept equals willlingness to pay for labour market estimates of the value of a statistical life. Journal of Risk and Uncertainty, 48, 187-205.

Kochi, I., Hubbell, B., Kramer, R. 2006. An empirical Bayes approach to combining and comparing estimates of the value of a statistical life for environmental policy analysis. Environmental and Resource Economics, 34, 385-406.

Kornhauser, L. A. 2001. On justifying cost-benefit analysis. In Adler, M. D. and Posner, E. A. (Eds), Cost-benefit analysis. Legal, economic, and philosophical perspectives, 201-221. Chicago, The University of Chicago Press.

Kristiansen, K. S. 1978. Bompengefinansiering ved alternative vegvalg. Erfaring fra motorvegen gjennom Drammen. Rapport. Oslo, Transportøkonomisk institutt.

Kuhn, T. S. 1962. The structure of scientific revolutions. Chicago, The University of Chicago Press.

Lakatos, I. 1968. Criticism and the methodology of scientific research programmes. Proceedings of the Aristotelian Society, 69, 149-186.

Lakatos, I. 1970. Falsification and the methodology of scientific research programmes. In Lakatos, I., Musgrave, A. (Eds): Criticism and the growth of knowledge, 91-196. London, Cambridge University Press.

Lakatos, I. 1971. History of science and its rational reconstruction. Boston Studies in the Philosophy of Science, VIII, 91-136.

Lakatos, I. 1978. The methodology of scientific research programmes. Philosophical Papers Volume 1. Cambridge, Cambridge University Press.

Latsis, S. (Ed). Method and appraisal in economics. Cambridge, Cambridge University Press, 1976.

Layard, R. 2005. Happiness. Lessons from a new science. New York, The Penguin Press.

Layard, R., Glaister, S. (Editors). 1994. Cost-benefit analysis. Cambridge, Cambridge University Press.

Layard, R., Mayraz, G., Nickell, S. 2008. The marginal utility of income. Journal of Public Economics, 92, 1846-1857.

Lichtenstein, S., Slovic, P., Fischhoff, B., Layman, M., Combs, B. 1978. Judged frequency of lethal events. Journal of Experimental Psychology: Human Learning and Memory, 17, 551-578.

Lindberg, G. 2006. Valuation and pricing of traffic safety. Örebro studies in economics 13. PhD dissertation. Örebro University.

Lindhjem, H., Navrud, S., Braathen, N. A. 2010. Valuing lives saved from environmental, transport and health policies: a meta-analysis of stated preference studies. ENV/EPOC/WPNEP(2008)10/FINAL. Paris, OECD.

Lindhjem, H., Navrud, S., Braathen, N. A., Biausque, V. 2011. Valuing mortality risk reductions from environmental, transport and health policies: a global meta-analysis of stated preference studies. Risk Analysis, 31, 1381-1407.

Lindhjem, H., Navrud, S., Biausque, V., Braathen, N. A. 2012A. Mortality risk valuation in environment, health and transport policies. Report. Paris, OECD Publishing.

Lindhjem, H., Navrid, S., Biausque, V., Braathen, N. A. 2012B. Meta-analysis of stated preference VSL studies: Further model sensitivity and benefit transfer issues. ENV/EPOC/WPNEP(2010)10/FINAL. Paris, OECD.

List, J. A., Gallet, C. A. 2001. What experimental protocol influence disparities between actual and hypothetical stated values? Environmental and Resource Economics, 20, 241-254.

Loewenstein, G., O'Donoghue, T., Rabin, M. 2003. Projection bias in predicting future utility. The Quarterly Journal of Economics, 118, 1209-1248.

Loewenstein, G., Ubel, P. A. 2008. Hedonic adaptation and the role of decision and experience utility in public policy. Journal of Public Economics, 92, 1795-1810.

Loomes, G. 2006. (How) Can we value health, safety and the environment? Journal of Economic Psychology, 27, 713-736.

Loomes, G., Sugden, R. 1982. Regret theory: An alternative theory of rational choice under uncertainty. Economic Journal, 92, 805-824.

Lutter, R.; Morrall, J. F. 1994. Health-Health Analysis: A New Way to Evaluate Health and Safety Regulation. Journal of Risk and Uncertainty, 8, 43-66.

MacKenzie, E. J., Damiano, A., Miller T., Luchter, S. 1996. The development of the Functional Capacity Index. The Journal of Trauma: Injury, Infection, and Critical Care, 41, 799-807.

Maier, G., Gerking, S., Weiss, P. 1989. The economics of traffic accidents on Austrian roads: risk lovers or policy deficit? Empirica – Austrian Economic Papers, 16, 177-192.

Menzel, P., Dolan, P., Richardson, J. Olsen, J. A. 2002. The role of adaptation to disability and disease in health state valuation: a preliminary normative analysis. Social Science and Medicine, 55, 2149-2158.

Miller, T. R. 1990. The plausible range for the value of life – some red herring among the mackerel. Journal of Forensic Economics, 3, 17-39.

Miller, T. R. 1993. Costs and functional consequences of U.S. roadway crashes. Accident Analysis and Prevention, 25, 593-607.

Miller, T. R. 2000. Variations between countries in values of statistical life. Journal of Transport Economics and Policy, 34, 169-188.

Miller, T. R., Guria. J. C. 1991. The value of statistical life in New Zealand. Market research on road safety. Wellington, Land Transport Division, Ministry of Transport.

Mishan, E. J. 1971. Evaluation of life and limb: a theoretical approach. Journal of Political Economy, 79, 687-705.

Mishan, E. J. 1985. Consistency in the valuation of life: a wild goose chase? Social Philosophy and Policy, 2, 152-167.

Mitchell, R. C., Carson, R. T. 1989. Using surveys to value public goods: the contingent valuation method. Washington D. C., Resources for the Future.

Mizuno, K., Kajzer, J. 1999. Compatibility problems in frontal, side, single car collisions and car-to-pedestrian accidents in Japan. Accident Analysis and Prevention, 31, 381-391.

Moher, D., Cook, D. J., Eastwood, S., Olkin, I., Rennie, D., Stroup, D. F. 1999. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. The Lancet, 354, 1896-1900.

Morrall, J. F. 2003. Saving lives: a review of the record. Journal of Risk and Uncertainty, 27, 221-237.

Morrison, M, Brown, T. C. 2009. Testing the effectiveness of certainty scales, cheap talk, and dissonance-minimization in reducing hypothetical bias in contingent valuation studies. Environmental and Resource Economics, 44, 307-326.

Morton, A. 2007. Great expectations. In Lewens, T. (Ed): Risk. Philosophical Perspectives, 84-98. London, Routledge.

Mrozek, J. R., Taylor, L. O. 2002. What determines the value of life? A meta-analysis. Journal of Policy Analysis and Management, 21, 253-270.

Munro, A. 2008. Bounded rationality and public policy. A perspective from behavioural economics. Heidelberg, Springer.

Murphy, J. J., Allen, P. G., Stevens, T. H., Weatherhead, D. 2005. A meta-analysis of hypothetical bias in stated preference valuation. Environmental and Resources Economics, 30, 313-325.

Nelson, J. P., Kennedy, P. E. 2009. The use (and abuse) of meta-analysis in environmental and natural resource economics: an assessment. Environmental and Resource Economics, 42, 345-377.

Nord. E. 1992. Efficiency and priority setting. Some problems in cost-effectiveness analysis of health care. PhD dissertation. Oslo, University of Oslo.

Nord E. 1999. Cost-value analysis in health care. Making sense of QALYs. Cambridge, Cambridge University Press.

Nyborg, K. 2012. The ethics and politics of environmental cost-benefit analysis. London, Routledge.

O'Hagan, A., Buck, C. E., Daneskhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J. Oakley, J. E., Rakow, T. 2006.Uncertain judgements. Eliciting experts' probabilities. Chichester, John Wiley and Sons.

Olson, M. 1965. The logic of collective action. Cambridge, Harvard University Press.

O'Reilly, D., Hopkin, J., Loomes, G., Jones-Lee, M., Philips, P., McMahon, K., Ives, D., Soby, B., Ball, D., Kemp, R. 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.

Oswald, A. J., Powdthavee, N. 2008. Does happiness adapt? A longitudinal study of disability with implications for economists and judges. Journal of Public Economics, 92, 1061-1077.

Percival, R. S. 1997. Campbell's blind variation in the evolution of an ideology and Popper's world 3. Philosophica, 60, 113-154.

Persson, U. 2003. Economic valuation of traffic safety. The development of methods for costing accidents in Sweden. Manuscript dated June 11, 2003. Prepared for the workshop Economic Valuation of Health Effects due to Transport, Stockholm, June 12-13, 2003.

Persson, U., Cedervall, M. 1991. The value of risk reduction: results of a Swedish sample survey. IHE working paper 1991:6. Lund, the Swedish Institute for Health Economics.

Persson, U., Hjalte, K., Nilsson, K., Norinder, A. 2000. Värdet av att minska risken för vägtrafikskador. Bulletin 183. Lunds Tekniska Högskole, Institutionen för teknik och samhälle.

Popper, K. R. 1979. Objective knowledge. An evolutionary approach. Revised edition. Oxford, Oxford University Press.

Pratt, J. W. 1964. Risk aversion in the small and in the large. Econometrica, 32, 122-136.

Pratt, J. W., Zeckhauser, R. J. 1996. Willingness-to-pay and the distribution of risk and wealth. Journal of Political Economy, 104, 747-763.

Pukalskas, S., Peceliunas, R., Sadauskas, V., Kilikeviciene, K., Bogdevicius, M. 2015. The methodology for calculation of road accident costs. Transport, 30, 33-42.

Raiffa, H. 1968. Decision analysis. Introductory lectures on choices under uncertainty. Reading, Mass. Addison-Wesley.

Randall, A., Hoehn, J. P. 1996. Embedding in market demand systems. Journal of Environmental Economics and Management, 30, 369-380.

Rawls, J. 1971. A theory of justice. Cambridge, Harvard University Press.

Reynolds, D. J. 1956. The cost of road accidents. Journal of the Royal Statistical Society, Series A, 119, 393-408.

Rizzi, L. I., Ortuzar, J. de D. 2003. Stated preference in the valuation of interurban road safety. Accident Analysis and Prevention, 35, 9-22.

Robinson, L. A., Hammitt, J. K. 2011A. Behavioural economics and the conduct of benefit-cost analysis: towards principles and standards. Journal of Benefit-Cost Analysis, 2, Article 5.

Robinson, L. A., Hammitt, J. K. 2011B. Behavioural economics and regulatory analysis. Risk Analysis, 31, 1408-1422.

Robles-Zurita, J. A. 2015. Valuation of safety under reference-dependent evaluation of income. Accident Analysis and Prevention, 79, 70-79.

Rossi, P. 1987. The Iron Law of evaluation and other metallic rules. Research in Social Problems and Public Policy, 4, 3-20.

Samuelson, P. A. 1947. Foundations of economic analysis. Revised and enlarged edition. Cambridge, Harvard University Press.

Samuelson, P. A. 1954. The pure theory of public expenditure. Review of Economics and Statistics, 36, 387-389.

Schelling, T. C. 1968. The life you save may be your own. In Chase, S. B. (Ed), Problems in public expenditure analysis, 127-176. Washington D. C., The Brookings Institution.

Schelling, T. C. 1984. Choice and consequence. Perspectives of an errant economist. Cambridge, Harvard University Press.

Schwab Christe, N. G. 1995. The valuation of human costs by the contingent method: the Swiss experience. In Schwab Christe, N. G and Soguel, N. C. (Eds), Contingent valuation, transport safety and the value of life, 19-43. Boston, Kluwer Academic Publishers.

Schwab Christe, N. G., Soguel, N. C. 1996. The pain of road accident victims and the bereavement of their relatives: A contingent-valuation experiment. Journal of Risk and Uncertainty, 13, 277-291.

Sen, A. K. 1973. Behaviour and the concept of preference. Economica, 40, 241-259.

Sen, A. K. 1987. On ethics and economics. Oxford, Basil Blackwell.

Sen, A. K. 1992. Inequality reexamined. Cambridge, Harvard University Press.

Sen, A. K. 2000. The discipline of cost-benefit analysis. The Journal of Legal Studies, 29, 931-952.

Sen, A. K. 2008. The economics of happiness and capability. In Bruni, L., Comim, F., Pugno, M. (Eds) Capabilities and Happiness, 16-27. Oxford, Oxford University Press.

Sen, A., Williams, B. (Eds). 1982. Utilitarianism and beyond. Cambridge, Cambridge University Press.

Shepard, D. S., Zeckhauser, R. J. 1982. Life-cycle consumption and willingness-to-pay for increased survival. In Jones-Lee, M. W. (Ed): The value of life and safety, 95-141. Amsterdam, North-Holland Publishing Company.

Simon, H. A. 1943. Administrative behaviour. A study of the decision making processes in administrative organizations. New York, The Free Press.

Simon, H. A. 1956. Rational choice and the structure of the environment. Psychological Review, 63, 129-138.

Simon, H. A. 1982. Models of bounded rationality. Volume 2. Behavioural economics and business organization. Cambridge, MIT Press.

Simon, H. A. 1983. Reason in human affairs. Oxford, Basil Blackwell.

Slovic, P. 2000. (Ed). The perception of risk. London, Earthscan Publications.

Slovic, P., Finucane, M. L., Peters, E., MacGregor, D. G. 2004. Risk as analysis and risk as feelings: Some thoughts about affect, reason, risk and rationality. Risk Analysis, 24, 311-322.

Slovic, P., Fischhoff, B., Lichtenstein, S. 1978. Accident probabilities and seat belt usage: A psychological perspective. Accident Analysis and Prevention, 10, 281-285.

Slovic, P., Fischhoff, B., Lichtenstein, S. 1979. Rating the risks. Environment, 21, 14-20, 36-39.

Slovic, P., Kunreuther, H., White, G. F. 1974. Decision processes, rationality and adjustment to natural hazards. Originally printed in G. F. White (Ed): Natural hazards: Local, national, global. Reprinted in Slovic, P. 2000. (Ed). The perception of risk, 1-31. London, Earthscan Publications.

Stanley, T. D. 2008. Meta-regression methods for detecting and estimating empirical effects in the presence of publication selection. Oxford Bulletin of Economics and Statistics, 70, 103-127.

Stanley, T. D., Doucouliagos, H. 2013. Neither fixed nor random: weighted least squares meta-analysis. Statistics in Medicine, 34, 2116-2127.

Stanley, T. D., Doucouliagos, H. 2014. Meta-regression approximations to reduce publication selection bias. Research Synthesis Methods, 5, 60-78.

Stone, A. A. et al. 2013. Subjective well-being. Measuring happiness, suffering and other dimensions of experience.

Sugden, R. 2008. Capability, happiness, and opportunity. In Bruni, L., Comim, F., Pugno, M. (Eds) Capabilities and Happiness, 299-322. Oxford, Oxford University Press.

Sunstein, C. R. 2004. Are poor people worth less than rich people? Disaggregating the value of statistical lives. Working paper 207. The University of Chicago Law School, John M. Olin Law and Economics Working Paper Series.

Svenson, O. 1981. Are we all less risky and more skillful than our fellow drivers? Acta Psychologica, 47, 143-148.

Svenson, O. Fischhoff, B., MacGregor, D. 1985. Perceived driving safety and seatbelt usage. Accident Analysis and Prevention, 17, 119-133.

Svensson, M. 2007. What is a life worth? PhD dissertation. Örebro studies in economics 14. Örebro university.

Szklo, M., Nieto, F. J. 2014. Epidemiology. Beyond the basics. Third edition. Burlington (MA), Jones and Bartlett Learning.

Sælensminde, K. The impact of choice inconsistencies in stated choice studies. Environmental and Resource Economics, 23, 403-420.

Sælensminde, K. 2003. Embedding effects in valuation of non-market goods. Transport Policy, 10, 59-72.

Sælensminde, K. 2006. Causes and consequences of lexicographic choices in stated choice studies. Ecological Economics, 331-340.

Tay, R. 2002. The Prisoner's Dilemma and Vehicle Safety. Some Policy Implications. Journal of Transport Economics and Policy, 36, 491-495.

Tengs, T. O., Adams, M. E., Pliskin, J. E., Safran, D. G., Siegel, J. E., Weinstein, M. C., Graham, J. D. 1995. Five-hundred life-saving interventions and their cost-effectiveness. Risk Analysis, 15, 369-390.

Tengs, T. O., Graham, J. D. 1996. The opportunity cost of haphazard social investments in life-saving. In Hahn, R. W. (Ed), Risks, costs, and lives saved, 167-182. New York, Oxford University Press.

Thaler, R. H. 1994. Quasi rational economics. New York, Russell Sage Foundation.

Thaler, R. H. 2015. Misbehaving. The making of behavioural economics. New York, W. W. Norton & Company.

Thaler, R. H., Sunstein, C. R. 2008. Nudge: Improving decisions about health, wealth and happiness. London, Penguin Books.

Thedié, J., Abraham, C. 1961. Economic aspect of road accidents. Traffic Engineering and Control, 10, 589-595.

Trawén, A., Hjalte, K., Norinder, A., Persson, U. 1999. Skattning av riskvärden och dödsfallsekvivalenter i vägtrafiken med hjälp av conjoint analysis, risk-risk trade off och standard gamble metoderna. Bulletin 182. Lunds Tekniska Högskola, Institutionen för teknik och samhälle.

Tversky, A., Kahneman, D. 1986. The framing of decisions and the psychology of choice. In Elster, J. (Ed). Rational Choice, 123-142. Oxford, Basil Blackwell.

Tversky, A., Kahneman. D. 1992. Advances in Prospect Theory: Cumulative representation of uncertainty. Journal of Risk and Uncertainty, 5, 297-323.

Ulph, A. 1982. The role of ex ante and ex post decisions in the valuation of life. Journal of Public Economics, 18, 256-276.

Vaa, T. 1997. Null-visjonen. En drøfting av forutsetninger og konsekvenser. Arbeidsdokument SM/0879-B/1997. Oslo, Transportøkonomisk institutt.

Vaa, T. 2007. Modelling Driver Behaviour on Basis of Emotions and Feelings: Intelligent transport systems and behavioural adaptations. In Cacciabue, P. C. (Ed): Modelling Driver Behaviour in Automotive Systems – Critical Issues in Driver Interactions with Intelligent Transport Systems, 208-232. London, Springer.

Vaa, T. 2013. Proposing a driver behaviour model based on emotions and feelings: Exploring the boundaries of perception and learning. In M.A. Regan, John D. Lee and Trent W. Victor (Eds) Driver distraction and inattention: Advances in research and countermeasures, 103-119. Surrey, Ashgate.

Van Praag, B. M. S., Ferrer-i-Carbonell, A. 2008. Happiness quantified. A satisfaction calculus approach. Revised edition. Oxford, Oxford University Press.

Veisten, K. 2016. Et konkret forslag til en liten nedjustering av ex-ante verdsetting av "lettere skade" basert på Verdsettingsstudien. Arbeidsdokument 51015. Oslo, Transportøkonomisk institutt.

Veisten, K., Flügel, S. Elvik, R. 2010. Den norske verdsettingsstudien. Ulykker – verdien av statistiske liv og beregning av ulykkenes samfunnskostnader. Rapport 1053C. Oslo, Transportøkonomisk institutt.

Veisten, K. Flügel, S. Rizzi, L. I., Ortuzar, J. de D., Elvik, R. 2013. Valuing casualty risk reductions from estimated baseline risk. Research in Transportation Economics, 43, 50-61.

Veisten, K., Navrud, S. 2006. Contingent valuation and actual payment for voluntarily provided passive-use values: Assessing the effect of an induced truth-telling mechanism and elicitation formats. Applied Economics, 38, 735-756.

Viscusi, W. K. 1993. The value of risks to life and health. Journal of Economic Literature, 31, 1912-1946.

Viscusi, W. K. 1994. Mortality effects of regulatory costs and policy evaluation criteria. Rand Journal of Economics, 25, 94-109.

Viscusi, W. K. 1998. Rational risk policy. The Arne Ryde memorial lectures. Oxford, Clarendon Press.

Viscusi, W. K. 2010. Policy challenges of the heterogeneity of the value of statistical life. Foundations and Trends in Microeconomics, 6, 99-172.

Viscusi, W. K. 2013. The benefits of mortality risk reduction: happiness surveys vs. The value of a statistical life. Duke Law Journal, 62, 1735-1745.

Viscusi, W. K. 2014. The value of individual and societal risks to life and health. Chapter 7, 385-452, Handbook of the Economics of Risk and Uncertainty, Volume 1.

Viscusi, W. K. 2015. The role of publication selection bias in estimates of the value of a statistical life. American Journal of Health Economics, 1, 27-52.

Viscusi, W. K., Aldy, J. E. 2003. The value of a statistical life: a critical review of market estimates throughout the world. Journal of Risk and Uncertainty, 27, 5-76.

Viscusi, W. K., Evans, W. N. 1990. Utility functions that depend on health status: estimates and economic implications. American Economic Review, 80, 353-374.

Viscusi, W. K.; Zeckhauser, R. J. 1994. The Fatality and Injury Costs of Expenditures. Journal of Risk and Uncertainty, 8, 19-41.

Von Neumann, J., Morgenstern, O. 1953. Theory of games and economic behavior. Third edition. Princeton, Princeton University Press.

Weimann, J., Knabe, A., Schöb, R. 2015. Measuring happiness. The economics of well-being. Cambridge, The MIT Press.

Weinstein, M. C., Shepard, D. S., Pliskin, J. S. 1980. The economic value of changing mortality probabilities: A decision-theoretic approach. Quarterly Journal of Economics, 94, 373-396.

Welsch, H. 2006. Environment and happiness: Valuation of air pollution using life satisfaction data. Ecological Economics, 58, 801-813.

Wolff, J. 2007. What is the value of preventing a fatality? In Lewens, T. (Ed): Risk. Philosophical Perspectives, 54-67. London, Routledge.

World Health Organization. 2014. World health statistics 2014. Geneva, World Health Organization.

Østre, S. 1970. Økonomisk vurdering av trafikkulykker og trafikksikkerhet. Rapport 14. Oslo, Utvalg for trafikksikkerhetsforskning.