MONASH University
Business and Economics

Centre for Health Economics

Research Paper 36 2009

The invalidity of Expected Utility Theory and its misuse in the economic evaluation of health and safety

Forthcoming in M Schlander (ed) *Economic Evaluation of Health Care Programs: Current Concepts, Controversies and International Experience,* Springer: New York

Professor Jeff Richardson
Foundation Director, Centre for Health Economics
Monash University, Australia

Robin Pope
Senior Fellow
Center for European Integration Studies and Experimental Economics Laboratory
Bonn University, Germany

May, 2009
Centre for Health Economics
ISSN 1833-1173
ISBN 1 921187 35 2

ACKNOWLEDGEMENTS

For comments we thank Professor Mike Jones-Lee of the Newcastle University Business School, UK; and Professor José Luís Pinto Prades of the Universidad de Murcia, Spain and most especially Professor Reinhard Selten of the Experimental Economics Laboratory Bonn University, Germany.

Correspondence:

Centre for Health Economics
Faculty of Business and Economics
Building 75
Monash University
Clayton
Melbourne, Victoria 3800
Australia

Phone: +61 (0)3 9905 0754, Fax: +61 (0)3 9905 8344
jJeff.richardson@buseco.monash.edu.au

ABSTRACT

Evaluating the quality and length of life has proved to be problematical. Economic evaluations of health and safety programs have used methods which draw upon the authority of Welfare Economics and, in particular, Expected Utility Theory (EUT). The methods – use of the Standard Gamble (SG) for measuring quality of life and the Value of a Statistical Life (VSL) for measuring the economic value of life itself – are independently flawed. However the focus in this chapter is upon the relationship between the methods and EUT. As a general theory EUT is invalid. Its defects flow through to the logic supporting the use of the SG and VSL and detract from, rather than add to, confidence that what they measure is useful in economic evaluation studies.

TABLE OF CONTENTS

List of Tables

List of Figures

The invalidity of expected utility theory and its misuse in the economic evaluation of health and safety

1 Introduction

Expected Utility Theory (EUT) has been one of the most influential theories in economics and has been a pivotal element in much of the analysis in finance and portfolio theory, insurance and welfare economics. More generally, it has been the analytical "workhorse" in most of the analyses involving risk and uncertainty. In the health economics literature it led to the widespread adoption of the Standard Gamble (SG) as a measure of utility and to the advocacy of the "Value of a Statistical Life" (VSL) as a method for placing a dollar value upon "statistical lives", ie the unknown lives lost or gained as a result of life-threatening risks. As elsewhere, EUT theory is part of the standard analytical apparatus in health economics, and consistency with EUT is explicitly held up as a criterion for optimal decision-making as, for example, with Garber and Phelps' influential work comparing the consistency of CEA with "optimal decisions" (Garber and Phelps 1997; Garber 2000).

It is argued here that EUT is invalid as a general decision theory and that analysis based upon EUT is consequently problematical. Two applications of EUT are assessed, namely its use to justify a special status for the SG in the measurement of the Quality Adjusted Life Year (QALY), and to calculate the Value of a Statistical Life (VSL) in order to monetise QALYs.

Since the development of the concepts of a Quality Adjusted Life Year (QALY) and Cost Utility Analysis (CUA) there has been debate about the appropriate instrument for measuring the health related Quality of Life (QoL). While seldom acknowledged, there is a near consensus that the QoL, does not, in fact, refer to "quality" in a quasi-objective sense. Rather, QoL refers to subjective perceptions that are usually equated with individual "utilities" ie a person's preferences for his or her own health state – regardless of whether or not others might agree with the person's assessment. Less commonly, "quality" refers to "social utility" or "value" which is a perception or preference on behalf of the community as a whole. The individual makes an assessment in his or her capacity as a citizen. This latter concept is not of primary concern here and it is the former concept that is adopted by welfare theory.

Following Torrance's influential article on measurement (Torrance 1986), the Standard Gamble (SG) has been widely accepted as the gold standard for estimating individual utility (Gold, Siegel et al. 1996; Drummond, Sculpher et al. 2005). This is because the technique derives from axiomatised Expected Utility Theory (EUT). The resulting utilities are also known as "N-M Utilities" in honour of von Neumann and Morgenstern who constructed the SG technique. If an individual reaped utility in the way that EUT assumes, then setting aside problems of measurement, the understanding of probabilities and other sources of bias, the SG technique would measure utility. Other methods for constructing QALYs have been used – the Time Trade-off (TTO), Rating Scale (RS) and, from a social perspective the Person Trade-off (PTO) and a fifth metric the Relative Social Willingness to Pay (RS-WTP) is under development (Richardson and McKie 2009). However these are often considered second best or proxies for the theoretically "correct" SG.

Table 1 Alphabetical list of abbreviations used

AQoL (4D, 6D, 7D, 8D)	Assessment of Quality of Life instruments: recent MAU instruments with adaptations for vision and mental health which use the TTO technique
CEA	Cost Effectiveness Analysis
CUA	Cost Utility Analysis
EQ5D	(EuroQoL) a five dimensional QoL instrument
EUT	axiomatised Expected Utility Theory: Expected Utility Theory as justified by axioms of individual choice eg those in Von Neumann and Morgenstern
HUI-3	Health Utilities Index, a MAU measure of QoL which uses the SG technique
DALY	Disability Adjusted Life Years, an "instrument" promoted by the World Health Organisation based upon the PTO
MAU	Multi-Attribute Utility
NHS	National Health Scheme
N-M Utilities	Utilities of a person who conforms to EUT, so named since the first full axiomatisation of EUT was that of von Neumann and Morgenstern
NPV	Net Present Value
QALY	Quality Adjusted Life Year
QWB	Quality of Wellbeing, an early MAU measure of QoL
PTO	Person Trade-Off, the number of people who would need to be saved from death and returned to full health to make a hypothetical programme of equal value to a second programme that returns a stated number (usually 10 or 100) people from certain death to the health state of interest, or a variant of this technique
QoL	Quality of Life
RS-WTP	Relative Social WTP, Richardson et al (2008)
SF6D	A MAU measure of QoL from 36 characteristics, collapsed to six on the basis of similarities, which uses the SG technique
SG	Standard Gamble, a gamble/certainty equivalent trade-off devised by Ramsey and von Neumann and Morgenstern to compute utilities of those who conform to EUT
RS	Rating Scale: a scale with clearly defined end points. In the QALY context these are often "full health and death"
TTO	Time Trade-Off between longevity and QoL
VSL	Value of a Statistical Life, that as typically constructed employs EUT assumptions
WTP	Willingness To Pay
15D	A Finnish MAU measure of QoL that incorporates 15 dimensions (health attributes) of QoL

The main alternative to the QALY technique is the determination of the Value of a Statistical Life (VSL) which can be compared with other dollar costs or benefits in an economic evaluation. Data are obtained from either people's Willingness To Pay (WTP) to reduce risk or their Willingness to Accept compensation (WTA) for increasing risk. These data are typically combined with EUT to infer the value of the health state or, more commonly, of life per se and to obtain a dollar value for the health state, or for the value of a statistical life. If the assumptions used to justify the two techniques – SG and VSL – and particularly EUT were correct, then they would produce consistent results (Section 4) and they might produce estimates no worse than other measurement techniques that, like these, are were only subject to framing, endowment and path dependence effects and errors arising from cognitive complexity. However people do not generally (and should not) obey EUT. This introduces a new and potentially dominating source of error, which is the main theme here.

It is argued below that neither the SG nor VSL measure what they purport to measure and that the principal (but not exclusive) reason for this is the normative and descriptive invalidity of EUT. To anticipate a common response, this means that neither technique can be defended on the following grounds. *All measurement is subject to framing, endowment and path dependence effects and cognitive limitations, so that EUT-based measurement is no different from other measurement techniques in having defects.* The problem is not attributable to measurement error but to what is measured by EUT under idealised conditions. Since it measures something other than what it purports to measure, reliance upon the authority of EUT detracts from, rather than enhances, the credibility of measurement, and both the SG and VSL must be justified by some other means.

The chapter reviews the role and methods of QoL measurement in economic evaluation (Sections 2 and 3); the formal equivalence of the SG and VSL (Section 4); the fundamental theoretical flaw in EUT and the arguments used to defend its role (Section 5; a detailed consideration of the consequences of this flaw for the logic of the VSL (Section 6). In the concluding section we briefly address the problem of measuring of QoL and the dollar value of life in the absence of a gold standard. Appendix 1 notes an additional conceptual perversity in the use of either the SG or VSL in the context of an NHS. This is not, however, the result of EUT.

2 Cost Benefit and Cost Utility Analysis

The framework of economic evaluation is simple and powerful. Benefits of an activity should, in principle, encompass anything of value to human wellbeing. Costs, or opportunity costs, are the loss of benefits as a result of the activity, and most obviously, from the resources that are used. If the world were certain, the best activity would maximise the difference between the future discounted benefits gained and lost, ie the net present value (NPV).

Even under certainty, NPV would be difficult to operationalise because of a conceptual issue that may be, in principle, intractable (Richardson and McKie 2009). What, for example, is the criterion of value? In a country with a national health system, should value be measured from the perspective of an aggregation of individual utilities or from that of the society? The two will in general yield distinct values (Bergson 1938; Samuelson 1947; Richardson 1994). In either case, how should interpersonal comparisons be made? In the context of health, the most obvious conceptual problems arise from the measurement of QoL and the inclusion of human life in economic evaluation. Cost Effectiveness Analysis (CEA) deals with the issue by ranking projects by cost per life year saved. The lower the ratio, the more life years may be gained from given resources. Cost Utility Analysis (CUA) is a modification of CEA that takes account of differences in QoL (the “health related qualify of life”).

Health quality (utility) is usually measured on a 0-1 scale and used to convert life years into QALYs (Quality Adjusted Life Years). For example 10 life years with a QoL (utility) score of 0.8 will produce 8 QALYs. QALY estimates have been carried out in practice in one of two broad ways. The holistic approach is to obtain a description of an entire health state and encapsulate it in a vignette or written scenario. This may contain anything relevant to a person’s QoL: social relationships, pain, energy, psychological state, etc. Generally, scenarios are obtained by interviewing patients in the health state. The scenario is then scaled “holistically” by asking a second group of respondents (usually a cross-section of the general public) to rate the scenario to obtain a utility index number.

The alternative MAU (multi-attribute utility) approach is to obtain a set of answers to questions on multiple health attributes and to apply a predetermined algorithm in order to aggregate these answers into a single utility score. The earliest examples of MAU instruments were the Quality of Wellbeing (QWB), the 15D (which uses a rating scale to quantify and combine over 15 attributes), and the Rosser-Kind Index (although the latter was primarily experimental and is no longer used). More recent MAU instruments include the Health Utilities Index HUI-3 and SF6D, both of which use the SG technique, and two other instruments that use the TTO technique– namely the 5 item EQ-5D (nee EuroQoL), the Assessment of Quality of Life (AQoL) (and its specialised adaptations for vision and mental health). The World Health Organisation has widely promoted a MAU-like metric, the Disability Adjusted Life Year, the DALY (Murray 1996).

The current state of MAU measurement is problematical. In part, this is due to the differences in the descriptive systems adopted, and in part to the differences in scaling techniques that have been employed to obtain "utility".

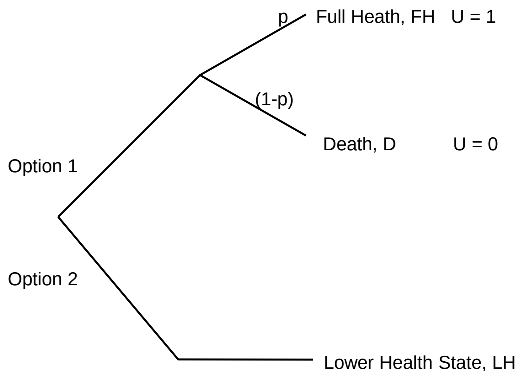
3 Scaling Techniques for QALYs and the VSL

Both the holistic and the MAU instruments require, in principle, two stages: (1) obtaining a description of a health state (the descriptive system); and (2) scaling (calibrating) the health state, ie assigning an index number to quantify the utility of the health state. The validity of the QALY clearly depends upon the method used to construct the index of utility. As described in Richardson (1994), since QALYs (Q) equal the index (U) times the length of life (L) (Q = U.L), the validity of the index requires a very demanding ("strong interval") property namely that a 10 percent increase in the index is equally valued as a 10 percent increase in the quantity of life, ie 1.1Q = (1.1U).L = U(1.1L) where 1.1 indicates a 10 percent increase in the measured amount. In this spirit, economists have favoured trade-off procedures for scaling, even though like all stated preference results, trade-off data are subject to framing, endowment and path dependence effects. The three most commonly used scaling techniques are the Time Trade-off (TTO), the Person Trade-off (PTO), and gamble/certainty equivalent trade-off obtained via the Standard Gamble (SG).

The TTO is the most frequently used of these procedures. It asks an individual to nominate the percentage of their life that they would be willing to sacrifice to achieve full health rather than the health state described by a holistic scenario or a description that uses the attributes of an MAU instrument. The PTO replaces length of life by number of people. Survey respondents are asked the number of people who would need to be saved from death and returned to full health to make a hypothetical program of equal value to a second program which returns a stated number (usually 10 or 100) of people from certain death to the health state of interest. Several variants of the PTO exist. Sacrifice of life years or people's lives allows the calculation of the utility Index.

The SG (Standard Gamble) technique is illustrated in Figure 1. Respondents to a survey are offered two options, either the health state, H, which is of interest (Option 2) or a Standard Gamble (Option 1). This consists of a lottery-type gamble in which there is a probability of p that the person will be returned to full health (utility Index, 1.00) or probability (1-p) that they will die almost immediately (utility Index, 0.0). The probability, p, is varied until the two options are equally attractive. At this point Expected Utility Theory (EUT) is invoked to assert that, as the two options are of equal value and death has a utility value of 0 and full health a value of 1, then the utility of the health state may be represented by the numerical value of the probability, p, as shown in Figure 1. Since EUT can be derived from the sets of axioms which are enshrined in orthodox economic theory the SG technique has been widely accepted as the "gold standard" for measuring "utility". The enshrining of the EUT axioms results in many economists downgrading other scaling techniques and adopting the view that these simply measure "value", which may be related to, but will differ from, true "utility" (Torrance 1986; Von Winterfeldt and Edwards 1986; Gold, Siegel et al. 1996; Drummond, Sculpher et al. 2005).

Figure 1 Standard Gamble to obtain a QALY



U = pU (FH) + (1-p) U(Death)

If U(FH) = 1, U(Death) = 0

Then under EUT if person is indifferent between Options 1 and 2

U(LH) = p(1) + (1-p) (0) = p

CEA, Cost Effectiveness Analysis, and CUA, Cost Utility Analysis, only rank projects by their cost per unit of outcome (the life year and QALY respectively). For a given health budget this may suffice (Pope 1976). But in deciding how large to make the budget, it is desirable to place a monetary value on the life year or QALY and to make a direct comparison with the monetary costs of obtaining them. In this respect some have argued for a Willingness to Pay (WTP), approach based upon stated or revealed preferences for avoiding a risk of (usually) death or a Willingness to Accept (WTA), compensation for a risk of death. The latter – the willingness to accept compensation – has commonly been inferred from econometric studies of wages and their relationship to the risk of death in particular industries. These methods produce the "Value of a Statistical Life" (VSL), so called because the lives saved cannot be identified but only inferred statistically from the change in the risk of death which may be traced to (Fromm 1965). Schelling (1968) introduced the risk-money trade-off methodology in the following way in the context of compensation:

> "(if) the loss of 0.6% of his income (after taxes) in perpetuity is about equivalent to one chance in 1,000 of immediate death (and) similar answers were obtained from 1,000 men of similar income and ages it could be concluded that they would together rather give up the equivalent of 6 discounted life-time incomes than suffer one immediate accidental death".

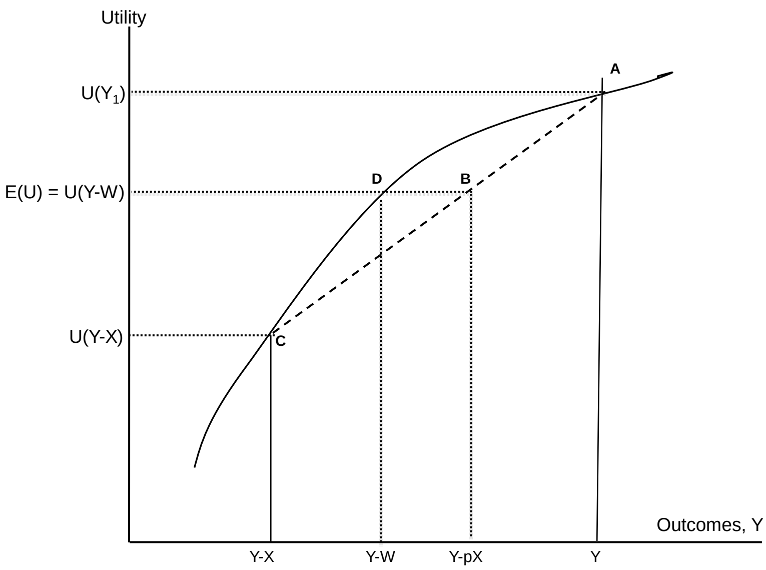
For example, if a 1/500 risk of death is associated with compensation of $10,000 then, multiplying by 500 (ie the combined effect of 500 such people), the expectation of one death is associated with $5 million of compensation, which is therefore the VSL. This type of calculation has been used extensively in transport economics as well as the health sector when interventions result in life saving (Jones-Lee 1969; Viscusi 1993; Viscusi 1998; Viscusi and Aldy 2003; Baker, Chilton et al. 2008). Developmental work on the methodology of the VSL is ongoing.

This deceptively simple procedure stems from multiple unacceptable assumptions, including, as Schelling himself admitted, that its basis in EUT unrealistically abstracts from the anxiety of death, which he deemed to be as important for utility as death itself. We return to this fundamental problem with EUT, and its consequences for the VSL methodology in Sections 5 and 6.

4 The formal equivalence of the SG and WTP for risk reduction

Figure 2 is one of the first applications of EUT's SG technique and was used in Friedman and Savage's famous paper (1948) in the context of insurance. It demonstrated how EUT may be used to estimate the WTP to reduce a risk. Faced with a probability, p, of a loss X, an individual would be willing to purchase insurance up to a cost of W, or accept the gamble G in which there is an outcome Y with probability 1–p and an income Y-X with probability, p. EUT states that individuals will select the option with the greatest expected (probability weighted) utility, and be willing to pay up W for insurance if the expected utility of a gamble is no more than the certain utility after paying W from full insurance against the loss, X.

Figure 2 Illustration of EUT: The WTP for insurance against the risk p of a loss x



In Figure 2, the expected utility of the gamble E(U), is the utility on the line CA where the ratio CB/BA equals the ratio p/(1-p). Point D indicates utility after purchasing insurance and having a certain income of Y–W. Because of the diminishing marginal utility, the utility at D is no greater than the utility at B from being uninsured and WTP for insurance is quantified by the distance between these points.

The equivalence of this WTP for insurance example and the Standard Gamble is clear. The SG offers two options: firstly a certain outcome Y-W or, secondly, a gamble with probability p of outcome Y-X and a probability (1-p) of outcome Y. According to EUT, the value of the gamble is equal to the sum of the probability weighed utilities of these two possible outcomes, ie p U(Y-X) + (1-p) U(Y). Therefore the value of the certain outcome, Y-W, may be found experimentally by varying p until the values of the two options are considered to be the same. When this occurs at

E(U), the equilibrating value of p is an index of the value – utility – of the outcome Y-W. In the context of WTP, W is the WTP for reducing risk by p. Repeated experiments would allow the mapping of the "utility income" function. In the context of CUA, Y-W is replaced by a poor health state, Y and Y-X by full health and death respectively and the equilibrating, p, is therefore the utility of the poor health state on a life-death (0-1) scale.

In the 1948 paper Friedman and Savage note that a person who behaves consistently with EUT obtains no utility or disutility from the outcome being risky and that the EUT axioms treat the utility from each outcome as if it were certain (even when the outcome itself is risky). If this were not so then the outcome – utility function depicted in Figure 2 would not be unique. While anchored at C and A where outcomes are certain the shape of the frontier between these points would need to differ with each level of risk as determined by individual risk attitudes and the relationship between risk and WTP. W would vary with the convexity/concavity of the function, a result confirmed empirically by Hershey et al. (1982). This problem is expanded below.

5 The invalidity of EUT

As noted, the argument that the SG and VSL are theoretically superior to other metrics depends upon the assumption that EUT correctly represents people's decision-making in the face of risk or that, as a normative proposition, it describes what people should do. From the time of publication of the 1944 edition of von Neumann and Morgenstern's book, the axioms of EUT have been subject to intensive empirical and theoretical examination, and found to be consistently defective empirically. There has also been a substantial literature from 1944 on its defects as a normative theory. There have also been a number of subsequent attempts to modify ETU and retain the axiomatic approach to the analysis of risk.

In a series of articles, Pope has focused attention, not upon the axioms, but upon a more fundamental theoretical criticism that, inter alia, explains the empirical and normative invalidity of the axioms. A risky choice necessarily involves time: an earlier period when outcomes are unknown and the risk remains unresolved and a later period when the outcome will become known but there is no longer any risk. Yet the elements of EUT depicted in all of the figures in this chapter ignore this fundamental fact. The outcomes shown are just that – ex-post outcomes, which exclude any emotions or behaviours which arise before one of the outcomes becomes certain. However, it has been known for a long time that, in addition to outcomes, the experience of risk is an ex-ante element that has been variously referred to as risk attitude, the disutility of risk per se/process utility/utility of gambling/love of gambling secondary satisfactions. The existence of risk per se is directly or indirectly recognised by Plato (wonder); Adam Smith (surprise wonder); Pascal (boredom); Ramsey (love of chance, excitement); von Neumann Morgenstern (a positive or negative utility of the mere act of taking a chance).

With the widespread adoption of EUT between 1948 and 1952, the existence and significance of this ex-ante element was disregarded. When re-introduced into theory, it was subsequently and incorrectly assumed by many to be included in EUT and wrongly termed the classical interpretation of EUT, eg. Schoemaker (1982). This black comedy of errors is traced in Pope (1983; Pope 1996/97).

The fundamental error of EUT is an error of omission – namely, the time before the risky outcome will become known – and the resulting error of many economists is to misinterpret EUT as including risk attitude – "the risk per se". But an a-temporal theory cannot take account of these intrinsically temporal phenomena and no amount of manipulating a-temporal elements can model time-dependent behaviours. Logically and historically, the error is extraordinary. Logically, risk cannot occur without a time dimension. Historically, von Neumann and Morgenstern added an appendix to the 1947 edition of their book explicitly explaining the difficulties of the “deep matter” they had encountered when seeking to go beyond EUT to include risk attitude into a general theory of risk behaviour. Their view of the matter could not have been more explicit.

> “I want to make it absolutely clear, that I believe, as Von Neumann, that there may be a pleasure of gambling, of taking chances, a love of assuming risk, etc. … I know of no axiomatic system worth its name that specifically incorporates a specific pleasure of utility of gambling together with the general theory of utility... I am not saying that this is impossible to achieve in a scientifically rigorous manner. I’m only saying, (as we did in 1944) that this is a very deep matter.” (Morgenstern 1974).

The very deep matter, that eluded von Neumann and Morgenstern, is that risk cannot be consistently modelled atemporally. This is because risk involves a succession of stages, each ending with a change in knowledge ahead, ie with new knowledge about the future. Pope et al. (2007) summarise the temporal dimension of risk concerning time-dependent learning in each new epistemic stage as follows:

> “He does not (initially) know his alternatives. Later he has a change in knowledge ahead, having discovered at least two alternative acts...still he does not know what act he will choose. Later he has a change in knowledge ahead, having chosen an act. Still, if the act is risky he still does not know the outcome of his act. He is in his pre-outcome period. Later he has a change in his knowledge ahead, having learned the outcome of his chosen act. He is in his post-outcome period. This division of the future by evolving stages of knowledge ahead allows the chooser to anticipate his secondary satisfactions... and have those affect his choice of an act.” (Pope, Leitner et al. 2007p 203).

During the pre-outcome period the existence of uncertainty (ie the limited knowledge of what is to occur subsequently – “knowledge ahead”) may generate risk-based emotions such as hope and fear. The limited knowledge ahead may also change behaviour because of the inability to make forward commitments (obtain loans, undertake ventures). A person facing the possibility of almost immediate death is likely to increase their rate of time preference: reduce savings, bring forward holidays and adopt a more indulgent lifestyle. None of these positive and negative satisfactions arising from uncertainty – from a limited degree of knowledge ahead – can be consistently included under EUT or other a-temporal axiomatisations of risk behaviour.

The omission of the pre-outcome period is evident in the standard compensation for risk analysis (illustrated later in Figure 3b). At low levels of risk, the required compensation may be small, possibly even negative when context and a small element of risk generates enjoyable excitement (eg. white water rafting, mountain climbing, etc.). At very high levels of risk this may be replaced by paralysing fear requiring compensation which asymptotes to infinity.

A variety of escape strategies and adaptations of a-temporal EUT have been attempted, none of which can describe time dependant behaviour and emotions. Many assumed that, since the SG technique involves probabilities, it captures risk attitude: both enjoyable excitement and terror. This argument originated with Marschak’s redefining risk aversion in terms of a concave utility function and this inappropriate and confusing change in terminology became ensconced in the Arrow-Pratt measures of risk aversion.

Secondly it has been argued that EUT should be treated as a normative theory of rational behaviour (where “rational” is sometimes tautologically defined by the axioms! See Marschak (1950) and Savage (1954 and 1972). To help overcome the tendency to label whatever EUT excludes as irrational or frivolous, Pope (2001) introduced the more neutral term secondary satisfactions for the specific utility of gambling that is derived from a person’s degree of knowledge ahead. The more neutral term also avoids the impression that EUT only omits emotional satisfactions and dissatisfactions.

There is no rational defence for ignoring the experiences of the pre-outcome period, whether behaviour change or emotional responses as occurs under a-temporal EUT. A decision which includes additional relevant considerations will not, in general, give the same result as a decision excluding these considerations. Contrary to the normative re-interpretation, ignoring elements which alter behaviour and wellbeing would, by most accounts, be considered irrational.

Another defence of EUT is the “elaboration of expected outcomes”. If the emotions and behaviours in the pre-outcome period are defined as “outcomes”, then it is argued that EUT may be applied by elaborating the outcomes to include these “secondary satisfactions”. On inspection, however, elaborating the outcomes in this manner destroys EUT’s axiomatic basis. To see this, suppose a person fears an operation where the outcome will be full health or death. With a positive probability of death, the person will experience fear, which according to this argument is therefore an outcome that occurs before the success of the operation is known. However, the axioms require any probability whatsoever between zero and one to be applicable to any outcome. Take the case where the probability of full health is one. The axioms would require a person to be simultaneously fearful and certain of full health; that is, a contradiction. In short, EUT cannot have elaborated outcomes concerning the pre-outcome period which are consistent with a-temporal axioms. Pope provides a formal argument of the incompatibility of elaborated outcomes and axiomatisation.

The absurdity of abstracting, *inter alia,* from the fear of death is nicely summarised by comedian Jack Benny:

> “Robber: ‘This is a stick-up. Your money or your life.
> (pause)
> Robber: ‘Look, bud. I said your money or your life.’
> Jack Benny: ‘I’m thinking it over’.”

“The Jack Benny Radio Program”, March 8, 1948 quoted from Ashenfelter (2006 p 2).

The point, of course, is that unlike Jack Benny, we would willingly give up our wealth to avoid the certain loss of our own life.

6 The invalidity of VSL

Before turning to EUT/VSL logic the implausibility of the VSL procedure and the magnitude of the potential error it introduces are illustrated. Figure 3a represents a person’s WTP for the elimination of the risk of death as the risk rises. As the ability, and therefore the willingness, to pay is constrained by a person’s income, the curve has been constructed to be concave from below (the real relationship could be more complex).[1] The slope of the line from the origin represents the average willingness to pay at a particular level of risk and the dotted lines indicate the result of linear extrapolation of the WTP from these levels of risk.

[1] As risk rises the marginal utility of risk reduction would also be expected to rise with the increasing fear of death. However as the (premium) cost of eliminating risk increases the marginal utility of residual income would normally rise. These two factors might, in principle, off-set each other and result in a linear or even convex function. A concave-convex function is likewise possible and if, the WTP for very small risk (a thrill) was negative the function might initially fall below the horizontal axis. It is illegitimate, however, to make the unstated assumption that the function is linear without evidence of this coincidental result.

Figure 3a WTP and the VSL

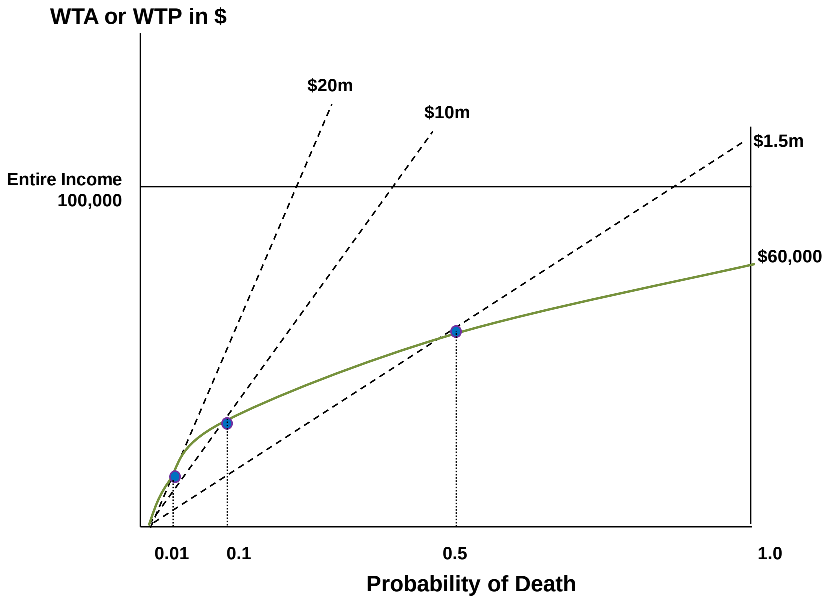
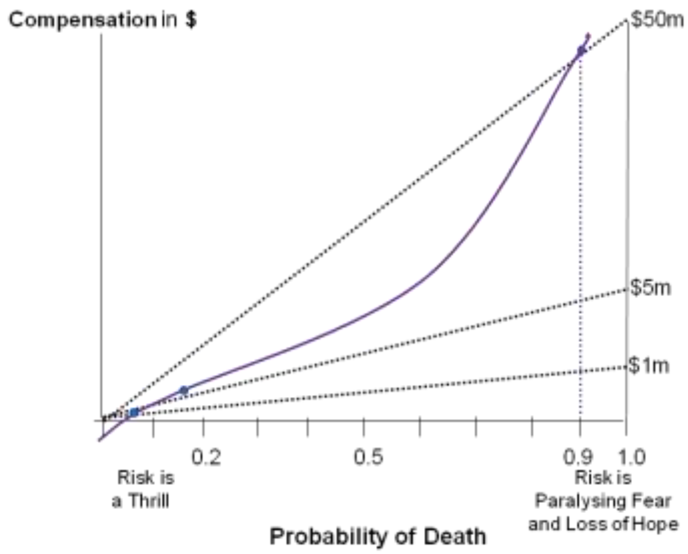


Figure 3b Compensation and the VSL



The first difficulty illustrated is that, following Schelling's procedure, the linearly extrapolated value may exceed a person's income (or wealth) and cannot therefore represent a person's WTP for the avoidance of the certainty of death. The numerical extrapolation is justified by aggregating the WTP of different people. But the collective WTP is not equivalent to an individual's WTP which is determined by the individual's equating both costs and benefits according to a common scale, namely their own values. The collective WTP is the summation of the WTP of different individuals, with different values. No individual considers the certainty of death, but only its risk.

Schelling and other writers, of course, recognise that the final VSL is an aggregation of different individual preferences, but fail to note that the final magnitude cannot, therefore, claim to be a coherent concept based upon the authority of orthodox welfare theory. Unless the social welfare function is simply the summation of dollar values, the concept corresponding with the magnitude is problematical. But further ambiguity exists.

This is that the value of a SLY changes with the level of risk used in the initial measurement. This is shown by the three dotted lines in Figure 3a. Unless the function is (improbably) linear (see last footnote) as the slope of the function changes so does the extrapolated value of the WTP to avoid risk. Figure 2 assumes an individual has only their own income (wealth) from which to insure against risk of death. A further complexity is that the function would rotate counter-clockwise if additional resources (wealth or relatives' assistance) could be obtained. As the function rotates, so the extrapolated VSL also changes.

Figure 3b represents the compensation needed to keep utility constant as a function of the risk of death. At low levels, risk creates little concern and may even be a source of pleasure (the thrill of danger). Negative compensation is implied. Close to death, fear increases and the compensation that must be paid for the risk of death rises exponentially and, in this illustration, is asymptotic to the certainty of death (trivial caveats might apply if the person wished to pass on an inheritance to their heirs). As with the WTP depicted in the previous figure, the extrapolated VSL varies with the level of risk. In this case, however, it rises, not falls, with risk.

Standard Gamble: Analogous problems to those discussed above, arise with the standard gamble. An individual's response to a risky choice will similarly depend upon risk attitude in a particular context and the anticipated affect of risk upon future behaviour. Parallel to the "which risk" problem for the VSL there is a "which method" problem for the use of the EUT's Standard Gamble technique. The SG is usually implemented by varying the probability in the choice which offers a gamble, as described earlier. This "probability equivalence" methodology should give the same result as a "certainty equivalent" methodology in which the probability in the gamble is fixed and the "certain outcome" of the alternative choice is varied. However as pointed out by Pope 2004 the two methods produce significantly different results with predictable bias. Hershey et al. (1982) showed that with the "probability equivalence method" the utility of income generally twists from concave to convex. With the probability equivalents version, the opposite result is obtained. Similar results were found by Hershey and Shoemaker (1985). The results demonstrate a fundamental problem with either the theory underlying SG or its application.

Returning to the logic of the EUT/VSL, the procedure described earlier is deceptively plausible. If a person is willing to pay an amount W to reduce the risk of death by 1 in 1,000, then it is intuitively appealing to conclude that the amount that 1,000 such people would spend is related to the value of the life which will be gained. In the context of accident prevention the logic appears almost trivial. However while it is certainly true that in these circumstances particular expenditures will have occurred and a life saved and that these events will be causally related, the less obvious question is whether or not the amounts spent represent the valuation of some coherent notion of life which may be used elsewhere, or even if the amounts spent represent any, non circular, notion of value.

The relationship between the VSL procedure and what, if anything, it values can be analysed by dividing the procedure into four steps:

Step 1 estimates W, the amount in dollars that a person is willing to pay for a reduced risk of death;
Step 2 uses EUT to replace the value of a probability of life, p, (risk of death (1-p)) with a probability times the utility of life;
Step 3 determines a relationship between the utility of life and dollars; and
Step 4 extrapolates the individual result to obtain the VSL.

As noted earlier a common approach to Step 1, estimating the WTP for risk reduction or compensation, W, is to calculate the revealed compensation for the risk of death from the regression of wages upon various attributes of jobs including the risk of death. The coefficient upon the latter variable is taken to be W, the dollar compensation required for accepting the corresponding risk. An alternative methodology is to use stated preferences and to ask a person (or a representative set of people) what dollar value equates the utility of a risky and riskless situation.

The relationships between the WTP, W, and the value of eliminating the risk of death are shown in Figures 4a and 4b using the decision tree format of the standard gamble to emphasise the analytical similarity between the two cases we are considering. If the person is indifferent between the upper branch of the tree (with the probability (1-p) of death and a higher wealth level of Y_1) and the lower branch of the tree (with guaranteed life but reduced wealth (Y_1-W)), then the indifference shown in Figure 4a can be expressed by the equation:

$$U[\Pr(\text{life with } Y1) = p;\ \Pr(\text{death with } Y1 = (1-p)] = U[\Pr(\text{life with } (Y1-W)) = 1] \qquad \text{eq (1)}$$

In Step 2 EUT is assumed which (crucially) allows the risk, p, to be detached – disembodied – from the utility and placed outside the utility function.

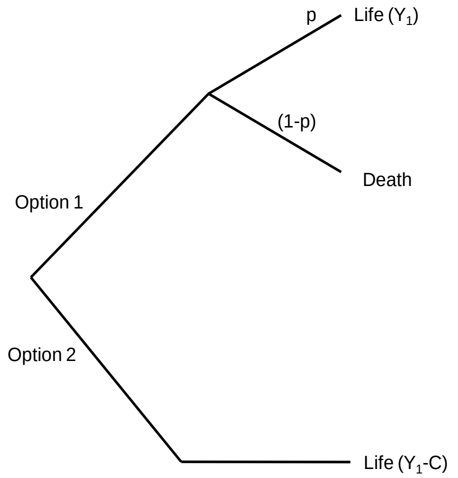
$$p.U(\text{life}, Y_1) + (1-p)\ U(\text{death}, Y_1) = U[\text{life}, (Y_1-W)] \qquad \text{eq (2)}$$

Simplifying, so that U [(death, Y_1)] = 0 (ie there are no benefits to the individual from bequeathing income to others) and abbreviating notation,

$$p.U(Y_1) = U(Y_1-W) \qquad \text{eq (3)}$$

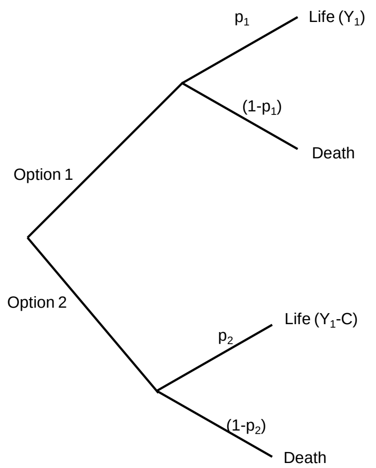
This is the result obtained in the orthodox standard gamble used in CUA but with U(Y_1-W) replacing an intermediate health state.

Figure 4a EUT and the Willingness to Pay for eliminating the risk of death



C = WTP to eliminate risk of death, (1-p)
If U(death) = 0, EUT states $pU(Y_1) = U(Y_1-C)$

Figure 4b EUT and the Willingness to Pay for life risk reduction



C = WTP to reduce risk of death, (p_2-p_1)
If U(death) = 0, EUT states $p_1 U(Y_1) = p_2 U(Y_1-C)$

As noted in the previous section, this second step commits an error of omission. Pre-outcome emotions and potential behaviour changes are simply eliminated. To the extent that people have emotional and other reactions to the probability of death (and from Plato onwards their importance has been widely recognised), the analysis is invalidated as WTP for risk reduction reflects these secondary satisfactions, not just the risk-free value of life and income.

The task in step 3 –obtaining a relationship between utility and dollars – is illustrated in Figure 5 which repeats Figure 2, except that the size of the deduction, W, needed to achieve certainty is reduced. Payment of W reduces income from Y_1 to Y_1-W, and utility from $U(Y_1)$ to $U(Y_1-W)$.

$$AB = U(Y1) – U(Y1\text{-}W)$$

From equation 3 (which is a result of EUT) $U(Y_1-W) = p.\ U(Y_1)$

$$AB = U(Y1) – pU(Y1)$$

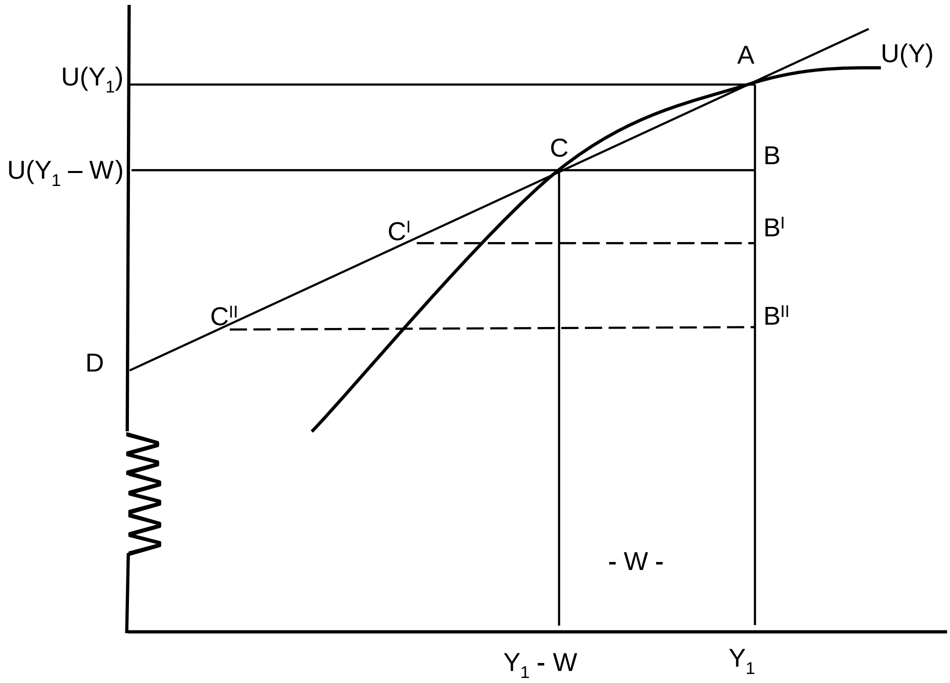
$$AB = (1\text{-}p)\ U(Y1) \qquad \text{eq (4)}$$

Since a person equates the value of W dollars with the utility AB and the value of the dollars depends upon the marginal utility of dollars, MU, AB = W.MU

Substituting in (4)

$$W.MU = (1\text{-}p)\ U(Y1) \qquad \text{eq (5)}$$

Figure 5 The local linearity assumption



Proceeding from this point is problematical and there are at least three options:

Option 1: Assume marginal utility (MU) is equal to average utility. Without any behavioural consequences this allows the simplifying calibration that a unit of utility may be assigned an index number 1.00. Equation (5) reduces to (6).

$$W = (1\text{-}p)\ U(Y_1) \qquad \text{eq (6)}$$

Stage 4 may be achieved by extrapolation of an individual's WTP. Multiplying both sides of equation 6 by 1/(1-p) obtains the formula for the VSL, viz VSL = W/(1-p).

The assumption that average and marginal utilities are equal is, however, implausible for the usual reasons. The extrapolated values may also exceed a person's income, as illustrated above. Possibly for this reason the assumption does not appear to be made explicitly in the literature and, as noted, the VSL has not been equated with an individual's utility, $U(Y_1)$.

Option 2: Assume marginal utility differs from average utility but is the same for similar individuals. In Figure 5 AB/W = MU and the triangle ABC represents the MU-dollar relationship for one person. A second, identical person would double the dimensions of the triangle to $AB^{I}C^{I}$, a third person to $AB^{II}C^{II}$ and so on. The line $ACC^{I}C^{II}D$ therefore represents the relationship between cumulative dollar compensation (WTP) and utility loss. This clearly does not follow the individual's utility function, (unless the individual is assumed to have a linear function, ie the MU is equal to average utility as above).

This option encounters two problems. First, when sufficient individuals are included to give the expectation of a single life saved the cumulative utility effect – the distance from $U(Y_1)$ to D – is different from the utility of a single person with income Y_1, (ie $U(Y_1)$) as it differs by the amount D

– the procedure does not measure individual utility and loses the authority of welfare theory. This is recognised in the literature). More fundamentally, the assumption of similar individual utilities – or marginal utilities – violates one of the key assumptions of welfare theory, viz, the incommensurability of individual utilities. The doctrine dates back to Robbins's classic 1932 article and was one of the reasons for the creation of the "new" welfare theory in its present form (Robbins 1932). Irrespective of the usefulness of Robbins' assumption, its violation indicates the need for a coherent theoretical basis for the VSL.

Option 3: Finally, no assumption may be made about the marginal utility of different individuals and, falling back upon the initial intuition, the VSL may simply be equated with the cumulative dollar WTP which corresponds with an expectation of one less death. Dollars, not utilities, are summed and the VSL may be defined as a new concept. However in terms of welfare, the concept is incoherent. Adding different marginal utilities is like adding expenditures in different currencies (yen, sheckles, etc). The latter provides no information about social or individual expenditure and adding marginal utilities measured in different units, provides no information about social or individual welfare.

Coherence of the concept: If one million individuals each pay $10 for a one-in-one-million chance of winning a car, then cumulatively they will have paid $10 million and one car will be won. While it is possible to define the "value of a statistical car" as $10 million, this concept is not useful and should not impact upon the allocation of resources. If one thousand people were prepared to pay a maximum of $700 to avoid a one-in-one-thousand chance of the loss of the contents of their home, it would be possible to define the "value of the statistical contents" as $700,000. However, this would have only a limited relationship to the real value of the contents and the concept is unlikely to have ever been suggested as a basis for resource allocation decisions.

If an individual will accept nothing less than the payment of $250 million to his heirs in exchange for his life, then any project which results in a death and a benefit less than $250 million cannot be (potentially) Pareto efficient. This inconvenient but self-evident conclusion has been papered over with the use of the concept of the VSL and EUT is, figuratively, the paper used to achieve this. The mathematical analysis facilitated by EUT deeply obscures the omissions and errors noted here and bestows authority to the analysis. The value of life cannot, however, sensibly be determined in a way which is consistent with welfare theory: there is no market for lives and compensation to the individual involved is virtually meaningless.

Defence of the Concept: Jones-Lee (1989) attempts to defend the concept of the VSL (as distinct from deriving its numerical value) by arguing that the concept's coherence only requires the assumption of a smooth (coherent?) marginal rate of substitution, MRS, between wealth (income) and the probability of death. This MRS is denoted m, and the marginal probability of death, dp. For this marginal risk, the individual will therefore pay m.dp and, as the number of identical individuals increases to n, the cumulative payment, V, is

$$V = \sum_{i=1}^{n} m_i dp \qquad \text{eq (7)}$$

If dp.n = 1 (ie n is the number of such risks needed to create the expectation of one death), then dp = 1/n. Substituting (7), cumulative payments,

$$V = (1/n)\sum_i m_i \qquad \text{eq (8)}$$

Jones-Lee defines this as the VSL (1989 p 12-13).

The defence conceals rather than avoids the problems outlined earlier. The “MRS” will not be smooth or coherent. The WTP for risk will depend upon context-specific and individual-specific “secondary satisfactions”. However the coherence of m requires the Stage 2 assumption above that risk may be disembodied from utility. That is, the utility of a risky option depends upon outcomes (incomes) and objective probabilities and is uninfluenced by pre-outcome secondary satisfactions.

A far-fetched defence (analogous to the elaboration of outcome defence of EUT) is to argue that if the increment of risk is sufficiently small for one individual, all risk related behaviour may be taken into account. It may be possible to conceptualise, for example, a smooth transition into terror, or from a stable life into a final pre death fling. It is not worth commenting on the likelihood of these marginal propensities being the same for all of the individuals whose preferences are summed by the VSL procedure. More fundamentally, the Jones-Lee defence is circular since n individuals each have dp = 1/n, W, the total payment in equation 7 is $\sum m(1/n) = nm(1/n) = m$ and the argument has simply defined the amount, W, which individuals will pay for life as the VSL which is option 3 above.

An Empirical Test: The purpose of an NHS is to satisfy social preferences. This suggests the possibility of a simple question to the public to test the suitability of the VSL for quantifying the value of life in this context. The question asks people to select one of two options: Option 1 is to pay 2,000 workers full compensation of $2,000 each p.a. for taking a 1 in 1000 risk of death at a cost to the taxpayer of $4m. Two people die annually. Option 2 is more expensive. For $4.4 million per annum, safety equipment will reduce the death rate to zero.

Respondents to a pilot application were told that the Australian Department of Roads accepted the argument underlying the use of the VSL, which was explained, and would select Option 1. They were told that Option 2 is considered fair by those affected and that taxes should be minimised.

From a total of 30 respondents, 4 selected policy 1 and 26 selected policy 2. Respondents were asked to give reason or reasons for their choice. Of the 45 reasons/comments, only 2 were consistent with the view that compensation satisfactorily offset the undesirability of death. 39 gave what might be described as deontological-moral rule-based-reasons, implying the need for a social, not an economic, resolution to the problem.

The test is subject to a variety of criticisms, but very strongly supports the hypothesis that social preferences will vary from private preferences and that the VSL does not represent an acceptable method for calculating the social value of a life year.

7 Conclusion

Despite voluminous contrary evidence and argument, EUT, maintains a position in economics that cannot be properly justified by rational argument. Its retention as the basis for ascertaining the VSL and using the SG to calculate QALYs results in systematic errors which, in turn, result in biased decisions with respect to treatment or non-treatment, and therefore the life and death of patients favoured or disadvantaged by the errors.

The assumption of EUT has had a deadening effect upon theoretical progress into important questions in economics and specifically health economics. Its attraction has been that it purports to provide definitive answers and, as a secondary bonus (but possibly a primary motivation), it allows rigorous analyses with orthodox mathematics. However, these lures have resulted in a fifty-year postponement in the construction of a theoretical basis for valid measurement of the value of health outcomes: ie measurement of that which we want to measure. This question subsumes the prior question of what it is that we want to measure discussed below.

Rather than investigate this question, economists have accepted EUT and overlooked implications which, to a non-initiate to its beguiling derivation, would probably appear to be very odd.

In the case of the SG, the value of health states such as intermittent mild depression, a headache, or poor sleep are evaluated using a choice which involves the possibility of instant death. At best, if respondents to the SG take the possibility of immediate death seriously then the equilibrating probability will probably be set too high. The fear of death is likely to obliterate or seriously distort all other considerations. Following the 9/11 terrorist attacks, for example, air traffic in USA reported to have dropped by over 30 percent, revealing an extremely strong reaction to an extremely low probability of death. If a money cost was placed upon the inconvenience endured voluntarily for safety, it is likely that the implied VSL would have been immense. Clearly the effect of fear is real and potentially dominating. For others, the planning problems associated with possible death will be dominating. Yet EUT's SG derives utility numbers, assuming that neither fear nor planning problems matter in evaluating an outcome.

In the case of the VSL, the objection is even more obvious. In the example used earlier, an individual who was relatively cheerful about a 1/500 risk of death in exchange for $10,000 is likely to be less cheerful about the certainty of death with or without compensation of $5,000,000. The extrapolation is self evidently invalid in the case of one individual. Because this objection is so obvious, the goal posts have been shifted and a new concept created to correspond with what can be calculated. The problem was acknowledged by the concept's initiator, but subsequently ignored. But there is no intrinsic meaning to this mix of context specific risk-based emotions, disutilities of anticipated behaviour change and dollar evaluation of life. Its use in the literature is, we suspect, attributable to the authority derived from its elaborate algebraic derivation, no understanding of the implausibility of the concealed assumptions and the objective overtones of the term "Value of a Statistical Life" or, more recently, "the Value of Preventing a Statistical Fatality". It is possibly assisted by the desire for a technical solution to a social problem, namely how much should the community pay for a social service.

This critical conclusion is directed primarily against the theoretical foundations of measurement in economic evaluation and not the broader framework or pragmatic use of some of the metrics. The systematic evaluation of interventions using consistent and sensible, if imperfect, criteria almost certainly results in a better outcome than anarchic decision-making by uncoordinated and self-interested groups uninformed by any economic evaluation. The SG, TTO and PTO all trade off life and death, albeit imperfectly, and with different framing effects and perspectives. The results are sufficiently similar to suggest that they are measuring something akin to preferences.

We suggest that a totally different approach is needed to establish a theoretical basis for measurement. It should not commence with axioms. After 2,500 years the physical sciences accepted that there were no knowable general rules which are universally true irrespective of context. (Einstein's displacement of Newton's laws of motion ended this belief). The discipline of ethics has likewise been unable to reach agreement on universal normative principles. The conclusion appears inescapable that, except for tautologies, we live in a world of best tentative hypotheses where unusual predictions or results which extrapolate beyond known data need to be carefully checked.

Measurement requires, firstly, information concerning the objectives of measurement, and secondly, measurement instruments which can be tested for validity: whether they measure what they purport to measure. The first question is social and the second, technical.

Despite its apparent simplicity, the question of what to measure in the absence of an assumed gold standard is deceptively difficult. In the context of an NHS, “health” is self-evidently important, but this broad term subsumes a myriad of unanswered questions. Should the unit of measurement be lives or life years; should it be measured using the median or mean; should quality be assessed from the perspective of patients or taxpayer-citizens; should the temporal perspective be ex-ante (“decision utility”), concurrent utility (“affect”) or informed ex-post assessment; should utility/preferences be the weighting factor or should a system of weights be adopted reflecting capacity (Sen 1997); should weights be adopted for other attributes – age, severity, health potential, social contribution, self-inflicted illness. Finally, what is the role (if any) of the TTO, PTO and SG? Each of these has defects but, with the exception of Richardson et al. (2008), there has been no experimentation with new metrics.

Answers to these questions might be country-specific and reaching agreement might be difficult. Each decision has a distributional effect, and compensation for losers is not possible. For this reason Richardson and McKie (2009) have argued that it is impossible to determine an optimal metric from purely technical analyses. Social choices are involved and this takes the analyses beyond the technical and into the realm of the political where economists have no mandate.

In sum, the misuse of EUT has provided wrong answers in health economics and elsewhere for almost half a century. It has held out the false promise of a technical solution to a social problem and has distracted economists from focusing upon where they might make the greatest contribution: explicitly investigating health-related social goals, proposing and experimenting with different ways of measuring these and inputting this information into a social decision-making process in which they would play an important role: possibly as co-convenors, certainly as analysts, but not as final arbiters of social values.

Appendix 1
The doubtful relevance of individual utility valuations in a National Health Scheme

With both the SG and VSL methodology utility is conceptualised and purportedly measured as the utility obtained by an individual. However health sectors are dominated by National Health Schemes (NHS) which have been established to achieve social objectives and it is not self-evident that the objective of such schemes is to maximise a measure based on individual utility. Indeed, in health economic evaluation, the term “extra-welfarism” has been adopted explicitly to indicate that health outcomes and not utilities are the source of social concern: that is, an NHS has been established to protect health, not utility. (If this was not so, then taxpayer-citizens should be happy to allow patients to substitute holidays for medical care so long as the cost was no greater.)

Despite this, even when an extra-welfarist perspective is adopted, both the SG and the VSL are commonly used to evaluate health states on the grounds that the QALY should be interpreted as an individual preferences weighted measure of health outcomes rather than, say, a duration-weighted utility. Clearly, some form of weighting is necessary to take account of the variable dimensions of QoL. However, once an extra-welfarist perspective is adopted, there is no longer a need for the weights to be individual-based utilities. Rather they may be some other expression of social preferences and, in particular, the TTO or PTO. The case for the latter is particularly strong if “social preferences” are desired which take into account an individual’s preferences for how others should be treated. (The PTO typically asks individuals to assume they are in the position of a social decision-maker who makes judgements on behalf of their community, including themselves.) It was for this reason that the World Health Organisation adopted the PTO and not the SG or TTO in the measurement of DALYs (Murray and Lopez 1996).

In the social context of an NHS where costs are partly or fully funded through taxation, the individual perspective embodied in the SG and individual WTP has a bizarre theoretical implication. As an individual patient is willing to pay more for their own treatment, then someone else, the taxpayer, is obliged to pay an increased amount for this person. For each extra dollar one individual is willing to spend on themselves, some other individual has to pay an additional dollar in tax! An analogous but less obvious paradox is implied when individuals nominate their own utility scores. The “efficiency” demonstrated by welfare theory depends upon the same individual incurring costs and benefits and upon their ability to judge these. When a wedge is driven between payer and beneficiary, welfare theory ceases to be relevant and its axioms are of questionable relevance. This point is not new. Jones-Lee (1976) for example, comments on this problem in his usage of the SG and individual WTP:

> “Our [WTP] analysis indicates that a road sweeper will value his own avoidance of certain death by road accident at somewhat less than a millionaire – we doubt whether public valuation should exhibit the same discrepancy” (Jones-Lee 1976 p 334).

References

Ashenfelter, O. (2006). Measuring the Value of a Statistical Life: Problems and Prospects, NBER Working Paper No 11916, National Bureau of Economic Research.

Baker, R., S. Chilton, et al. (2008). "Valuing lives equally: Defensible premise or unwarranted compromise." Journal of Risk and Uncertainty **36**: 125-138.

Bergson, A. (1938). "A reformulation of certain aspects of welfare economics." Quarterly Journal of Economics **52**: 310-334.

Drummond, M., M. Sculpher, et al. (2005). Methods for the Economic Evaluation of Health Law Programs. Oxford, Oxford University Press.

Friedman, M. and L. Savage (1948). "The utility analysis of choices and the measurability of utility." Journal of Political Economy **LVI**(4): 279-304.

Fromm, G. (1965). Civil Aviation Expenditures Measuring Benefits of Government Investment. R. Dorfman. Washington, Brookings Inst.

Garber, A. M. (2000). Advances in CE Analysis. Handbook of Health Economics. A. Culyer and J. Newhouse. Amsterdam, Elsevier. **1A:** 181-219.

Garber, A. M. and C. E. Phelps (1997). "Economic foundations of cost-effectiveness analysis." Journal of Health Economics **16**: 1-31.

Gold, M. R., J. E. Siegel, et al. (1996). Cost Effectiveness in Health and Medicine. Oxford, Oxford University Press.

Hershey, J., H. Kunreuther, et al. (1982). "Sources of bias and assessment procedures for utility functions." Management Science **28**(8): 936-956.

Hershey, J. and P. Shoemaker (1985). "Probability vs certainty equivalents methods in utility measurement: Are they equivalent." Management Science **31**: 1213-1321.

Jones-Lee, M. (1969). "Valuation of reduction in probability of death by road accident." Journal of Transport Economics and Policy **3**: 37-47.

Jones-Lee, M. (1976). The Value of Life: An Economic Analysis. Chicago, University of Chicago Press.

Jones-Lee, M. (1989). The Economics of Safety and Physical Risk. Oxford, Basil Blackwell.

Marschak, J. (1950). "Rational behavior, uncertain prospects, and measurable utility." Econometrica **18**(2): 111-141.

Morgenstern, O. (1974). Some reflections on utility. Expected Utility and the Allais Paradox: Contemporary Discussions of Decisions under Uncertainty with Allais' Rejoinder. M. Allais and O. Hagen. Boston, Dordrecht Reidel.

Murray, C. J. (1996). Rethinking DALYs. The Global Burden of Disease. C. J. Murray and A. D. Lopez. Geneva, World Health Organization.

Murray, C. J. and A. D. Lopez (1996). The Global Burden of Disease. Geneva, World Health Organization.

Pope, R. (1976). "The costing of public funds revisited." Australian Economic Papers **15**: 37-51.

Pope, R. (1983). The pre-outcome period and the utility of gambling. Foundations of Utility and Risk Theory with Applications. B. Stigum and F. Wenstop. Boston, Reidel Dordrecht**:** 137-177.

Pope, R. (1996/97). "Debates on the utility of chance: A look back to move forward." Journal for Science of Research (Zeitschrift für Wissenschaftsforschung) **11/12**: 43-92.

Pope, R. (2001). "Evidence of deliberate violations of dominance due to secondary satisfactions - attractions to chance." Homo Economics **XIV**(2): 47-76.

Pope, R., J. Leitner, et al. (2007). The Knowledge Ahead Approach to Risk: Theory and Experimental Evidence Berlin, Heidelberg, Springer-Verlag.

Richardson, J. (1994). "Cost utility analysis: What should we measure." Social Science & Medicine **39**(1): 7-21.

Richardson, J., A. Iezzi, et al. (2008). The Relative Social Willingness to Pay Instrument, Research Paper 22. Melbourne, Centre for Health Economics, Monash University.

Richardson, J. and J. McKie (2009). The Impossibility of an Ideal Metric for Health Service Benefit Measurement, Research Paper 33. Melbourne, Centre for Health Economics, Monash University.

Robbins, L. (1932). Essays on the Nature and Significance of Economic Science. London, MacMillan.

Samuelson, P. (1947). Foundations of Economic Analysis. Boston, Harvard University Press.

Savage, L. (1954 and 1972). The Foundations of Statistics. New York, John Wiley & Sons.

Schelling, T. (1968). The life you save may be your own. Problems in Public Expenditure Analysis. S. Chase. Washington, The Brookings Institute.

Schoemaker, P. (1982). "The expected utility model: Its variants, purposes, evidence and limitations." Journal of Economic Literature **20**(2): 529-563.

Sen, A. (1997). Individual preference as the basis of social choice. Social Choice Re-Examined. K. J. Arrow, A. Sen and K. Suzumura. London, Macmillan.

Torrance, G. (1986). "Measurement of health state utilities for economic appraisal." Journal of Health Economics **5**: 1-30.

Viscusi, W. (1993). "The value of risks to life and health." Journal of Economic Literature **31**: 1912-1946.

Viscusi, W. (1998). Rational Risk Policy. Oxford, Oxford University Press.

Viscusi, W. and J. Aldy (2003). "The value of a statistical life: A critical review of market estimates throughout the world." Journal of Risk and Uncertainty **27**: 5-76.

Von Winterfeldt, D. and W. Edwards (1986). Decision Analysis and Behavioral Research. Cambridge, Cambridge University Press.