CENTRE FOR HEALTH PROGRAM EVALUATION

Critique and Some Recent Contributions to the Theory of Cost Utility Analysis

Professor Jeff Richardson

Director, Health Economics Unit
Centre for Health Program Evaluation
Monash University

CENTRE PROFILE

The Centre for Health Program Evaluation (CHPE) is a research and teaching organisation established in 1990 to:

- undertake academic and applied research into health programs, health systems and current policy issues;
- develop appropriate evaluation methodologies; and
- promote the teaching of health economics and health program evaluation, in order to increase the supply of trained specialists and to improve the level of understanding in the health community.

The Centre comprises two independent research units, the Health Economics Unit (HEU) which is part of the Faculty of Business and Economics at Monash University, and the Program Evaluation Unit (PEU) which is part of the Department of Public Health and Community Medicine at The University of Melbourne. The two units undertake their own individual work programs as well as collaborative research and teaching activities.

PUBLICATIONS

The views expressed in Centre publications are those of the author(s) and do not necessarily reflect the views of the Centre or its sponsors. Readers of publications are encouraged to contact the author(s) with comments, criticisms and suggestions.

A list of the Centre's papers is provided inside the back cover. Further information and copies of the papers may be obtained by contacting:

The Co-ordinator
Centre for Health Program Evaluation
PO Box 477
West Heidelberg Vic 3081, Australia
Telephone + 61 3 9496 4433/4434   **Facsimile** + 61 3 9496 4424
E-mail CHPE@BusEco.monash.edu.au

ACKNOWLEDGMENTS

AUTHOR(S) ACKNOWLEDGMENTS

ABSTRACT

The theme of this paper is that the objectives of cost utility analysis as it is currently practiced have been driven by economic conventions which, in the case of utility measurement are misleading. Adherence to these conventions has meant that social objectives in the health sector have been poorly investigated. In particular, the role of population values–ethical beliefs–has been neglected. This theme will be illustrated with two main arguments: (1) that the use of the standard gamble to measure 'utility' (preferences under risk) has been subject to two related technical errors and has presupposed what may prove to be an incorrect social objective and (2) that the ethical values embodied in the simple QALY are far more restrictive than those that operate in the health sector.

Table of Contents

List of Tables

List of Figures

1   Introduction

The theme of this paper is that the objectives of cost utility analysis as it is currently practiced have been driven by economic conventions which, in the case of utility measurement are misleading. Adherence to these conventions has meant that social objectives in the health sector have been poorly investigated. In particular, the role of population values–ethical beliefs–has been neglected. This theme will be illustrated with two main arguments: (1) that the use of the standard gamble to measure 'utility' (preferences under risk) has been subject to two related technical errors and has presupposed what may prove to be an incorrect social objective and (2) that the ethical values embodied in the simple QALY are far more restrictive than those that operate in the health sector.

2 Measurement Under Risk

The standard gamble has commonly been accepted as a gold standard for measurement because, if the von Neumann-Morgenstern (NM) axioms are true, then the values it produces represent 'utility' or, more precisely, 'NM utility'. It is argued that these capture true preferences as preferences are being measured 'under risk' as occurs in real decision making[1].

However the relationship between the standard gamble, NM utility and observed behaviour is commonly forgotten. Consider the following example adapted from Luce and Raiffa (1957). There are three events: A, B and C which represent the receipt of \$20, \$9 and \$1 respectively. A choice in the form of a standard gamble is offered; viz:

$$(\$9,1) \quad \text{versus} \quad (\$20, p) \quad (\$1, 1\text{-}p)$$
$$B \quad = \quad p. A \quad + \quad (1\text{-}p) C$$

The certainty of \$9 is compared with a gamble the outcomes of which are \$20 with a probability p and \$1 with a probability (1-p). A risk averse individual may select, for example, p = 2/3, implying:

$$B \quad = \quad 2/3\ A \quad + \quad 1/3\ C$$

---

[1] It is often asserted that the standard gamble produces utilities that are consistent with, or are implied by, the NM axioms. This is untrue. *If* the axioms are true *and* the standard gamble is used *then* the NM utilities will predict preference behaviour.

An NM utility or index number may be derived by giving events A and C any arbitrarily selected values and with these anchor points calculating an index number for B. For example if A = 21 and C = 15 then B = 2/3 (21) + 1/3 (15) = 19. Alternatively, if A = 1 and C = 0 then B = 2/3 (1) + 1/3 (0) = 2/3. Each set of numbers (21, 19, 15), (1, 2/3, 0) or any linear transformation of these give numbers whose relative values can represent the person's preferences as shown in the gamble. And this was the purpose of von Neumann and Morgenstern; viz, to represent various choices numerically in the context of games theory. Their axioms allowed the rules of probability to be applied to NM utilities in such a way that preferences for different risky outcomes in the context of a particular game could be predicted. Significantly, they explicitly denied that they had created a general theory of utility as illustrated in the following two passages.

> 'The conceptual and practical difficulties of the notion of utility and particularly of the attempts to describe it as a number, are well known and their treatment is not amongst the primary objectives of this work... Let it be said at once that the standpoint of the present book on this very important and very interesting question will be mainly opportunistic.'
>
> <div align="right">von Neumann & Morgenstern 1944 p 28</div>

> 'I want to make it absolutely clear that I believe–as von Neumann did–that there may be a pleasure of gambling, of taking chances, a love of assuming risk, etc. But what we did say and what I do feel I have to repeat even today after so many efforts have been made by so many learned men, is that the matter is still very elusive. I know of no axiomatic system worth its name that specifically incorporates a specific pleasure or utility of gambling together with a general theory of utility.... I am not saying that it is impossible to achieve it in a scientifically rigorous manner. I am only saying (as we did in 1944) that this is a very deep matter.'
>
> <div align="right">Morgenstern 1979 p 181</div>

Technical Error 1: NM Utilities Are Not Generally Applicable

NM utilities do not represent preferences for particular states. Rather, they represent the preferences revealed in particular gambles. Luce and Raiffa make this distinction explicit using an example in which, as above, the difference between the numbers representing B and C is always twice the difference between the numbers representing A and B. They comment:

> 'Does this permit us to say that going from B to C is twice as (or even just more) desirable than going from A to B? We think not! The number 2/3 was determined by choices amongst risky alternatives, and it reflects attitudes towards gambling, not toward the two intervals. Suppose, for example, that, because of his aversion to gambling, our subject reported he would be indifferent between paying out $9 and having a 50/50 chance of paying out $10 or nothing. His response could then be summarised by saying that his utilities for $0, -$9, and -$10 are 1, 1/2 and 0. We would be unwilling, however, to say that going from -$10 to -$9 is "just as enjoyable" as going from -$9 to $0 (p 22).'

The NM utility values in this and all other cases reflect three quite separate magnitudes; viz, the preference for the riskless states, the probabilities of these states occurring and the 'specific utility of the gamble'; the hope, fear, excitement, exhilaration, anticipation or anticipated regret which are experienced prior to–and therefore separate from–the outcome of the gamble (Richardson 1994). As noted above, von Neumann and Morgenstern did not believe they had created a general axiomatic model representing this latter aspect of risk. More specifically, while a set of numbers, U_i, derived from the context of a particular gamble–for example, the life-death gamble of cost utility analysis–may represent preferences of subjects facing a life-death gamble, the magnitudes U_i, U_j would not necessarily represent preferences if the gamble involved only the health states from which U_i U_j were derived. The specific utility of risk is unlikely to be the same when there is a risk of death and when there is only a risk of mild discomfort.

Advocates of the standard gamble must therefore either assume that the specific utility of the gamble is (1) non existent or quantitatively trivial; (2) described by or accommodated for by the axioms; or (3) identical in all cases of risk. There is no evidence for the first two alternatives and compelling reasons for doubting their validity. Defenders of the 'risk versus utility' dichotomy appear to select the third option when they argue that utility may only be measured 'under risk'. But this position is both logically and empirically indefensible. The NM axioms explicitly permit multiple lotteries to be collapsed into a single lottery–and, as argued by Richardson (1994) in the context of the two stage HYE procedure–this is inconsistent with a coherent notion of the specific utility of the gamble. The specific utility enters the compound lottery at least twice. When it is collapsed into a single lottery it occurs only once and yet the specific utilities associated with the compound and simple lotteries are considered to be identical. Secondly, the 'sure thing' principle, explicit or implicit in the different versions of the axioms, permit the introduction or elimination of risk into the comparison of outcomes, as in the Allais paradox. The introduction or elimination of the associated specific disutility of risk is inconsistent with the axioms if it has a positive value. As a consequence NM utilities are very limited in the range of preferences that they can represent and predict.

Technical Error 2: The Standard Gamble Does Not Measure the Medical Risk

The argument for measurement 'under risk' is based upon the fact that medical outcomes are subject to risk and the claim that the standard gamble represents measurement 'under risk'. However the risk in the standard gamble is altogether different from the risk associated with the outcome unless there is a fortuitous coincidence of the two. In the standard gamble risk is an instrument to measure a person's preference for the *quality of life* of a health state; it is not the risk of entering that state. It is possible that the QoL of a health state is such that it implies an NM utility near 0.5 (high risk in the SG). But the heath state may be very unlikely to occur (low risk). Conversely, the health state may be near full health (low risk in the SG) but its likelihood of occurring may be very uncertain. In sum, the standard gamble measures the wrong risk to model people's uncertainty and is only defensible on the intuitively unlikely and–to my knowledge– empirically unsupported hypothesis that the specific disutility of risk is always the same.

A final defence of the standard gamble is that the use of NM utilities will impose consistency upon decision making, where consistency is defined by the NM axioms. Consistency is argued to be 'rational' and, consequently, the purpose of utility measurement is redefined from a positive role (predicting individual preferences) to a normative role (imposing rational behaviour). The defence is, however, indefensible! First, and as noted above, the NM axioms do not satisfactorily account for the specific utility of gambling and this is true whether the axioms are used in a positive or in a normative analysis. Consistent behaviour as defined is not necessarily rational. It ensures that the NM utilities used in an economic evaluation consistently reflect the specific utility of the gamble used to quantify the QoL. Generally this is a life-death gamble. Hence the specific utility arising from a life-death gamble will be embodied in the NM utilities used to evaluate non life threatening medical outcomes. Put another way, the consistent use of wrong data is not, in the usual sense, rational.

Ethical Values

There is an implicit objective in the use of the SG that utility measurement should take risk into account. This is consistent with the ethical basis of 'welfarism' in which social welfare is a function of individual utilities. However there are two interpretations of 'welfarism'. With 'ex ante' or libertarian welfarism, individuals make unrestricted decisions about their treatment options. These are necessarily made before the outcome of the treatment is known and the choice, therefore, includes all of the pre outcome emotions that are associated with the risk.

By contrast with this, 'ex post welfarism' measures the preference for (certain) health states after the outcome is known. (There is, of course, a semantic issue here. If the distinction between 'value and utility' is denied then the terminology here is unobjectionable. If utility is *defined* as being something measured under risk then the label 'ex post welfarism' could be replaced by 'extra welfarism' although in this case the social objective remains the satisfaction of preferences, albeit preference weighted health outcomes).

It is entirely a matter of social values which of these options should be incorporated in cost utility analysis. Economic theory–but not economic convention–is silent on this issue. It is, however, possible to empirically investigate the opinion of the population on the presumption that their values should normally be respected.

To do this, a postal questionnaire was sent to a group of Melbourne residents selected randomly from the telephone directory. With each of eight questions respondents were told that two alternative treatments were being considered for inclusion in Medicare. These were known by experts to be equally effective but the population did not necessarily believe this information which was given to them. In each case, the first option was cheaper but in each case the second option was preferred by the population and people were prepared to pay for the extra cost of the second option.

Three basic scenarios were presented each of several ways to explain the population's preferences. Each emphasised a pre-outcome factor which could influence preferences–factors

that may influence the specific utility of risk. First, (for reasons given) the second service reduced anxiety (reassurance); secondly, (for reasons given) people receiving the first service anticipated regretting their choice if there was an adverse outcome. Thirdly, the population based first option engendered uncertainty which the second option did not.

In each case respondents were asked to consider two arguments: (1) the second option should be selected because this is what people wanted and were prepared to pay for; (2) the first option should be selected because the purpose of a public health service is to provide health care and should not be influenced by people's other concerns.

Results are summarised in Table 1. Overall 68 percent of answers favoured Option 1 thereby supporting the second argument. When the extra cost of Option 2 was to be paid for by increased taxation rather than by a reduction in other public health services support for the second option rose from 23 to 37 percent. These results suggest that the population does not accept the ethical basis of ex ante welfarism and, consequently, the need to include risk preferences in the measurement of health state utilities.

Table 1: Preference for Ex Post versus Ex Post Valuation[1]

Number of answers implying a willingness to pay for the	(i) Reassurance	(ii) Regret	(iii) Risk	Total
A Ex Ante factor	52	46	60	158
B Ex Post benefit only	117	123	102	342
% Ex Post	69%	72.3%	63%	68%

Notes: (1)   Each of 64 respondents were asked 8 questions
               500 answers were useable

3 Other Dimensions of Social Value

In the economic tradition of ex ante, individualistic, libertarian, consequentialism (to use a pithy phrase!) economists have been primarily concerned with the impact of health services on life expectancy and the quality of life as perceived by individuals. Evidence from several studies is summarised below to illustrate the breadth of the unresolved issues associated with social (as distinct from individualistic) objectives: objectives which people, acting as citizens, as distinct from self interested individuals, might seek.

Severity of Illness (Need)

A series of recent studies have suggested that need per sé may be of independent importance in social decision making (Nord 1995, Nord & Richardson 1993, Haddorn 1991). The clearest evidence is from the US guidelines for the prioritisation of organ transplants. Treatment of less severe cases would increase the benefit received but priority is based primarily upon severity and consciously not upon potential benefit. From the perspective of the individual–and in economic theory to date–the severity of an illness is only important as the baseline for health improvement and this will be taken into account in the individual's assessment of the utility of the baseline health state. From a social perspective there are at least two possible reasons why additional priority might be given to severe illnesses and which might explain the empirical evidence cited above. First, treatment of more severe cases reduces health inequalities. Secondly, society collectively may apply what Haddorn (1991) described as the 'rule of rescue', the belief that some states are so severe that irrespective of individual behaviour or cost benefit calculations, they must be treated.

A direct test of the severity hypothesis was carried out on a convenience sample of 78 students. They were, initially, presented with Table 2 and told that from the viewpoint of patients the two programs were of equal value. They were then presented with the QoL rating scale in Figure 1 which indicated that Illness A was more severe than Illness B. The improvement AA^1 is the same as BB^1. Students were asked firstly, to decide which illnesses, if any, should receive priority. 57 percent nominated A; 16 percent B and 28 percent nominated equal priority. All respondents were asked what number of patients receiving Treatment B that would be needed to make the program equally valuable as Program A. The mean and median responses were 303 and 208 respectively.

These results lend significant weight to the 'severity hypothesis'.

Table 2: Health Improvement

	Benefits (Patient Perspective)	
	Project A	**Project B**
• Cost	Same as B	Same as A
• Program improves	QoL	QoL
• People are willing to personally pay	$30,000	$30,000
• It is rated as equally valuable as gaining	1 year of life	1 year of life

Figure 1: Quality of Life Scale
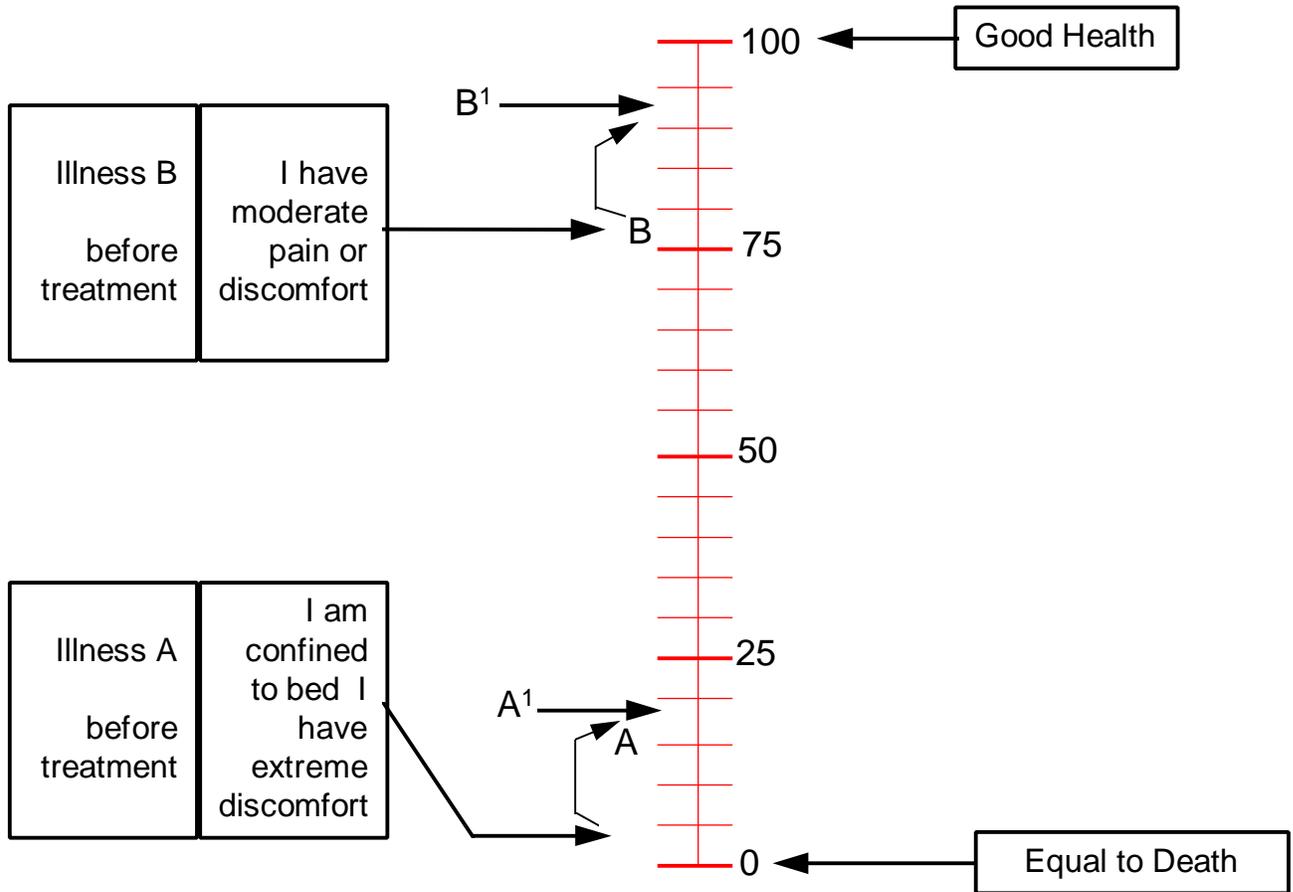
Distributive Neutrality

In simple cost utility analysis the reason for a decrease in the cost to QALY ratio is unimportant and it is unaffected by the distribution of costs and benefits. On the benefit side, 'a QALY is a QALY is a QALY': the QALY score rises *directly* with (1) the present value of the life years gained (the 'duration effect'); (2) the quality of life; and (3) the number receiving quality or quantity benefits; (4) other social characteristics are treated as being unimportant and (5) the cost to QALY ratio rises directly with cost.

There is now evidence of variable quality to suggest that :

1   People may wish the present value of life years gained to be further discounted as it may be perceived to be 'unfair' to discriminate against those whose life expectancy is only extended a little (Nord and Richardson 1994).

2   Increasing quality may also be discounted for analogous reason and because, as quality rises, severity falls (Nord and Richardson 1995).

3   There may be a non linear relationship between social value and the percentage of the population eligible for benefits. The possibility of treatment allows for the hope of a cure even when very few are treated (eg liver transplantation). Initial cures are therefore disproportionately valued. The extension of programs from the majority to all of the population may be of particular value because of the elimination of perceived discrimination thereby, once again, resulting in the disproportionate valuation of these treatments.

4   By contrast, discrimination against particular groups may be perceived as socially desirable. For example, in addition to the lower life expectancy of the aged, additional age discrimination may be considered desirable as a result of the 'fair go' or 'equal share' argument (Nord 1995, Murray 1997). There is evidence that Australians would discriminate against smokers with smoke related illness (Nord & Richardson 1995).

5   There is also evidence from Australian surveys that people do not wish the value of health programs to fall directly with their cost as it is believed that this is unfair to people who happen to have a high cost illness (Nord & Richardson 1995, 1996). Respondents to a two part survey persisted with this opinion through three levels of argument and finally allocated a budget in a way that explicitly did not maximise health benefits in order to prevent what was perceived as cost based discrimination.

The common feature of all of these distributional issues is that they (necessarily) draw upon 'social values' as distinct from the individual values which are the basis of most economic inquiry.

Breadth of Objectives

The issues above all concern benefits related to health states or the distribution of health improvements where these states and improvements are perceived in conventional terms. There has been, to the author's knowledge, no empirical inquiry by economists to determine whether these issues exhaust social objectives. One exception is a recent Decision Analytic study of the objectives of health personnel working in the South Australian community health sector (Peacock & Richardson, 1998). In this, an open forum was used to determine the dimensions of social benefit that were considered of importance to health personnel. Dimensions were precisely defined by a series of response levels from best to worst. The resulting multi-dimensional (or

multi-attribute) instrument was scaled using a modified rating scale procedure in which no benefits could occur in the absence of individual health benefits. Item responses and dimensions were assigned importance weights and South Australian health programs were evaluated.

The scale values obtained from this exercise were unexpected and, consequently, a mail validation exercise was carried out in which the instrument was re-scaled using a modified person trade-off technique. Results confirmed the initial findings that on a 100 point scale of value health workers believed that health gain should score about 30: equity about 40 and 'community health'– the extent of community empowerment, participation and decision making, about 30. As this latter objective was explicitly not a means to a health related end, economic theory does not recognise its existence.

Willingness to Pay

By contrast with the lack of inquiry into the issues discussed above, there has been a re-emergence of interest in the willingness to pay criterion for health benefits and the development of the techniques of contingent valuation for applying this criterion. The trend appears to be more a result of enthusiasm for the new techniques than the outcome of an ethical debate or empirical evidence regarding the appropriateness of the criterion. Indeed, it is seldom acknowledged that the issue of the evaluative criterion is ethical in nature; rather, economists typically assert or quote authorities who assert that willingness to pay is the correct criterion 'according to economic theory'.

There is no definitive ethical or empirical test for the validity of a criterion. As noted earlier, however, in the absence of compelling ethical considerations to the contrary, there is a prima facia case for accepting the view of the majority of the population.

To determine this, a convenient sample of 67 students were questioned. The concept of a criterion of value was introduced and the use of the willingness to pay as a criterion illustrated in the context of a free market. Respondents were asked if they believed it was ethically acceptable or objectionable to use an individual willingness to pay criterion of value in the health sector. Eighty four percent replied that it was objectionable. In response to a follow-up question, 97% (two not sure) disagreed with the proposition that the government should allocate resources in this way in the health sector.

The result is highly unsurprising and there can be little doubt that a similar response would be given by a more representative cross-section of the community. For well known reasons there is almost universal agreement that the value of health services should be assessed with different criteria than apply elsewhere. The uncritical re-emergence of the market criterion in economic analysis is not a reflection of social values.

Conclusions

The common theme is each of the disparate arguments presented in this paper is that economists have been remarkably unadventurous in their investigation of social objectives. Inquiry has been largely based upon *a priori* theorising rather than empirical observation, and, not surprisingly, it has focussed upon issues arising from and consistent with conventional theory.

It has been suggested here that there is a large agenda of unresolved issues requiring urgent theoretical and empirical attention. By contrast, debate in the peak health economic journals over issues such as the Healthy Year Equivalent, the QALY and their correspondence with the axioms of economic theory are a long way from the type of analysis needed to clarify real world social objectives. Modern economic theory often appears to be methodologically closer to Plato than to the modern physical sciences. Repeated statements of the form 'according to economic theory...' seems to imply a belief in some metaphysical world of Platonic forms relating to economic behaviour and that it is the economist's task to apply ever more demanding analytical techniques to unlock the secrets of this world. This is a far cry from what, in my view, is the role of a social science; viz, to determine by conjecture and empirical investigation the constraints and trade-offs in the achievement of goals and–repeating the main theme of this paper–to determine, empirically what goals are valued by the population. Whether and to what extent policy makers should take any notice of these goals is another issue which occupies remarkably little shelf space in the economist's library.

References

Haddorn, DC 1991, 'Setting health care priorities in Oregon', *Journal of the American Medical Association*, 265, 2218-2225.

Luce, RD & Raiffa, H 1957, *Games and Decisions: Introduction and Critical Survey*, John Wiley, London.

Morgenstern, O 1979, Some reflections on utility', in Allais, M *Progress in Utility and Risk Theory*, Hypothesis and the Allais Paradox, Oslo Conference.

Murray, C 1996, 'Rethinking DALYs' in Murray, C & Lopez A *The Global Burden of Disease*, WHO, Harvard School of Public Health, Harvard University Press.

Nord, E, Richardson, J & Macarounas-Kirchman, K 1993, 'Social evaluation of health care versus personal evaluation of health states', *International Journal of Technology Assessment in Health Care'*, 9:4 463-478

Nord, E 1995 'The person trade-off approach to valuing health care programs', *Medical Decision Making*, 15, 3, 201-208.

Nord, E, Richardson, J, Street, A, Kuhse, H & Singer, P 1995, 'Who cares about cost? Does economic analysis impose or reflect social values', *Health Policy,* 34, 79-94.

Nord, E, Richardson, J et al 1995, 'Maximising health benefits versus egalitarianism: An Australian survey of health issues, *Social Science & Medicine*, 41, 10, 1429-1437.

Nord, E, Richardson, J et al 1996, 'The significance of age and duration of effect in social evaluation of health care', *Health Care Analysis*, 4, 103-111.

Peacock, S & Richardson, J (forthcoming), 'Setting priorities and community health: Experiences with PBMA in South Australia', in Harris, A *Economics in Health 1997, Australian Studies in Health Service Administration*, University of New South Wales.

Richardson, J 1994, 'Cost utility analysis: What should be measured?', *Social Science & Medicine*, 39, 1, 7-21.

von Neumann, J & Morgenstern, O 1944, *Theory of Games and Economic Behaviour*, Princeton University Press, Princeton.