Economic Assessment of Health Care : Theory and Practice

J. Richardson
National Centre for Health Program Evaluation

CONFPAP.NOV

Contents

1. Introduction*

Health systems throughout the developed world have passed through two distinct phases and are presently entering a third. From the 1950s to the late 1970s virtually all of these countries experienced an increase in health expenditures so far in excess of GDP that the sectors share of the national product more than doubled. For the OECD as a whole, health expenditures rose from 3.8 to 7.0 per cent of the GDP between 1960 and 1980. By the end of this period there was virtually unanimous agreement that outlays were not being matched, on the margin, by improvements in health outcome. Following the economic downturn of the 1970s the second phase commenced. This was characterised by a significant restraint on health expenditures and, in the following seven years, the health cost to GDP ratio in the OECD rose by 0.3 percentage points (Schieber and Poulier, 1989). By the middle of the 1980s concern was being expressed in a number of countries that indiscriminate fiscal measures were jeopardising the quality of the health service. Increasingly, there has been a demand for resources within the health sector to be rationalised so that high priority areas may be separated from those where benefits are more problematical; that is, for resources to be allocated on the basis of costs and benefits. This has lead to the belief that within the health sector the 1990s will be characterised as a period of economic evaluation.

While there can be little doubt that this would be a beneficial development, the techniques for the evaluation of health care have not fully evolved. This is partly a reflection of the imperfect state of evaluation techniques generally but, more particularly, it is associated with problems that arise from the nature of the output. The benefits of health care include life and an improved quality of life. The measurement of these and their satisfactory integration with economic theory has not proved to be easy. The objective of the present paper is to outline the historical development of economic assessment in the health sector and, more importantly, to review a number of the unresolved questions. The review does not purport to be exhaustive and, in particular, it will emphasise two sets of issues which have received little attention in the literature namely, the value systems which underly the evaluation techniques and the measurement and inclusion of the "indirect" production benefits which result from the preservation of life.

* The author would like to thank Brian Parmenter of the IAESR at the University of Melbourne for valuable comments and suggestions.

The need for explicit evaluation of health care is particularly acute in a health system where services are provided by the Government or where their physical supply is directly or indirectly regulated. The absence of fully developed evaluation techniques may therefore suggest the desirability of a free market approach to the allocation of resources. In this context it is worth noting that the exception to the historical pattern described above has been the USA where the market based system has been perceived as being in a cost crisis since the 1950s. Phases have been characterised only by the intensity of the crisis. For example, in the 1980s while other western countries constrained expenditures, the so-called "competition revolution" in the USA was associated with an acceleration in the rate at which health costs have been increasing. It has been a combination of this US experience and a number of persuasive theoretical arguments that has led the majority of health economists and governments to doubt the efficacy of the market "solution" and to regard it as being about as attractive as nature's "solution" to the problem of old age. Apart from its generally unacceptable distributional effects, the unregulated market does not appear to allocate health care resources efficiently. Market based systems may prove to be viable if an appropriate regulatory framework can be devised[1] but any system must face the apparently inescapable fact that consumers will not be the agents that decide between the technical alternatives (although the assessment may be based upon their preferences - see below). The agents making choices will, consequently, need an acceptable framework for measuring and comparing costs and benefits.

2. The Departure from Conventional Welfare Theory

Explicit project evaluation was a late arrival in economics[2]. In its modern form it dates back to 1936 when a US law directed the army to investigate the costs and benefits of river and harbour projects for which there was no market. It took ten years for cost benefit analysis (CBA) to evolve into its present form. When compared with later forms of project evaluation, the defining characteristic of CBA is its attempt to reduce all benefits and costs to a single measurement unit, the dollar. While this has the major advantage of allowing a

---

*

[1] See, for example, Enthoven, 1988 and Scotton (1990).

[2] For a review see Warner and Luce (1982).

direct comparison of cost and benefits there has never been a satisfactory method for quantifying the value of life itself. Two quite different techniques were employed to achieve this. First, and in the tradition of the material welfare school of Marshall and Pigou, the human capital approach envisaged the value of life as "external" to the individual and to be measured by the individual's contribution to subsequent output. The approach was first used in the health sector by Mushkin (1962) and subsequently developed by Rice and Cooper (1967) and Cooper and Rice (1976). Because of its simplicity it has been the most commonly used measure in the cost benefit literature and was the method recommended by the US Public Health Service (Hodgson and Meiners, 1982). Two defects of the technique have proved to be fatal. First, it has strong and, to most, unacceptable distributional implications. Cooper and Brady (1976) for example, found the present value of a 25-29 year old male to be $US475,000 for a white college graduate and $US165,000 for a black high school dropout; a person approaching retirement aged 65 had a present value of $US41,000[3]. In all such studies women are valued less than men and the retired elderly have a present value of zero! Secondly, the dominant view amongst (non-health) economists has been that the appropriate measure of value is the intensity of individual preferences (preference utilitarianism) and that this is not reflected by future productivity.

The latter belief forms the theoretical basis for the alternative, willingness to pay (WTP), technique. Intensity of preference is believed to be reflected in the revealed expenditure of an individual. The problem of interpersonal comparison of welfare is resolved by appeal to the Calder-Hicks doctrine of potential Pareto improvement. In practice, a persons willingness to pay has been quantified using interviews, by extrapolation from the willingness to pay for an incremental improvement in safety and, most commonly, from the wage increments associated with hazardous jobs.

Like the human capital approach, WTP has been criticised in both its application and on its theoretical foundations. In their review, Landefeld and Seskin (1982) note that, to date, the results obtained from the use of the technique have varied from the improbably high to the improbably low and, depending upon the results adopted, virtually any project could be accepted or rejected. As with the human capital approach, the technique has strong distributional implications that are likely to be widely rejected. The wealthy can reveal a

[3] Quoted from Robinson, 1986.

higher value of life than the poor and the outcome of WTP based cost benefit analysis will favour this social group. The theoretical objection to WTP is discussed below.

In practical terms, the consequence of these attempts to derive a dollar equivalent to the value of human life was a widespread perception amongst non economists that CBA was not a very useful technique to assist with the allocation of health care resources (see, for example, Abel Smith, 1985). The next methodological development occurred in the 1950s. It reflected a more general belief that in a number of areas, and not simply health care, program benefits could not sensibly be converted into dollars. Once again it was the US military that took the initiative. With the development of cost effectiveness analysis (CEA) projects were ranked according to the criterion of cost per unit of effectiveness. For the military, projects would be preferred which had, for example, a lower cost per death inflicted (on the enemy!) or in "macro" analysis, cost per mega-death. As noted, similar problems of evaluation arise in the health sector and the techniques developed by the military were adapted - with, of course, some modification to the unit of effectiveness. In health evaluation this has sometimes been measured by the number of morbid days or episodes of illness. On occasions, an intermediate measure has been used such as the number of cases of a disease detected by a diagnostic test. Most commonly, however, output has been equated with the number of lives or life years saved. Thus a redistribution of funds to projects with a low cost per life year will increase the total number of life years that may be gained.[4]

The major weakness of CEA is that it treats all life years as having equal value no matter what the quality of the life. However, alternative interventions may have different side effects and a longer life may, rationally, be traded off against comfort and functional ability. In principle, the willingness to pay approach could include these issues but with its rejection both CBA and CEA could simply note, but not quantify, these "intangible" factors. With the development of cost utility analysis, (CUA) in the last twenty years, techniques have evolved which permit the quantification of these intangibles and their inclusion in the analysis by the calculation and comparison of costs per Quality Adjusted Life Year.

WTP versus an External Objective

[4] The major CEA studies in the literature are reviewed by Drummond (1981).

As noted above, the WTP technique for the evaluation of human life has considerable appeal and, *prima facie* might appear to be the theoretically "correct" procedure. It is based upon the widely accepted principles of welfare economics, and, in particular, upon an acceptance of revealed preference as the criterion for value in social decision making. Even in theory, however, WTP does not measure the value of life or the value of a health state but rather the *ex ante* value of the perceived risk of such a health state or the loss of life. The compensation required to satisfy the potential Pareto criterion is similarly based upon the value of perceived risk. It is possible to infer the individual's value of life from the value of this risk only if the perception of risk is correct and if the von Neumann-Morgenstern axioms (or an acceptable alternative set of axioms) are believed to describe individual behaviour. However there are now compelling theoretical and empirical reasons for rejecting these axioms and with them the validity of the extrapolation from risk to the value of life[5].

Even though the WTP technique may not reveal the value of life (or a health state) it may still be the appropriate measure. Thus it has been argued that there is no acceptable alternative; that value cannot be revealed *ex post* (dead men may tell no lies but they also reveal very little!); that the value of a foregone life `ex post' has little meaning and that, consequently, the appropriate magnitude to be measured is, in fact, the *ex ante* value of risk and not the *ex post* value of a life[6]. In his original statement of this position Mishan (1971) concludes that despite its practical difficulties WTP is the preferred measurement technique because "there is more to be said for rough estimates of the precise concept than precise estimates of economically irrelevant concepts" (p.705).

Mishan's statement reflects a surprising but widespread misunderstanding of the role of social values in economic evaluation. It may be true that WTP is the basis for evaluation in welfare economics generally. It is not, however, the only possible basis and the relevant issue is whether or not it is appropriate in the health sector for the achievement of social objectives.

---

[5] For a theoretical and empirical review see Pope (1989) and Schoemaker (1982) respectively.

[6] For an interesting debate on this issue see the exchange between Broome (1978, 1979) and Buchanan and Faith (1979) and Williams (1979).

Two questions have been asked in this context. The first is whether or not WTP *should* (in the normative sense) be the basis of decision making in the health sector. The second is whether or not WTP *is* (in the positive sense) accepted as the basis for value. The first question has resulted in a large an inconclusive *a priori* debate about the presumed attributed of health, health care and the health care market. (For a review see Richardson, 1988). Of the many characteristics which have been claimed to differentiate health care from normal economic commodities the most plausible are its presumed meritorious status - the fact that there may be a socially desired level and distribution of particular services - and the fact that in health care there is a pervasive lack of reliable information about the impact of all but the most trivial services. As a consequence consumers and even health professionals are unable to accurately evaluate either the probabilities of events or the likely outcome of alternative treatments in individual cases. In these circumstances it should not, perhaps, be surprising that preferences based upon individual *ex ante* expectations might be rejected as the basis for resource allocation.

The second, and arguably more relevant question than the determination of what society's preferences *should be*, is the empirical question of whether or not society does or does not accept WTP in the health sector as the basis for value and resource allocation. The evidence here is unequivocal. In every country consumer sovereignty has been replaced to a greater or lesser extent by either the direct provision of health care or through the provision of social insurance. The rhetoric of health care systems invariably presumes that the objective of the system is the improvement of mortality and morbidity and this is the primary objective of governments as reflected by the information collected and the research carried out by the various health authorities.

The replacement of consumer sovereignty with some other criterion of value is not necessarily "theoretically incorrect" or evidence of regulatory failure, but may simply be the result of a social choice between alternative value systems. While there has been a strong liberal tradition in economics and western society in which individuals are responsible for both choices and consequences there has also been a more paternalistic tradition which imposes choices when the consequences of consumer sovereignty are deemed to be undesirable. The social concern is neither surprising nor inconsistent with the emphasis on freedom of choice elsewhere. In most choice contexts the consequences of *ex post* "wrong"

decisions are not catastrophic and error learning is possible. Error learning however, is seriously inhibited by death. While no fully consistent and theoretically satisfying method has been devised for determining when there should be interference with individual choice, there is no doubt about either the existence of such a social value system or the ability of society to distinguish between different choice contexts. Historically one of the clearest examples of such a social objective has been the minimisation of *ex post* and objectively measured mortality and morbidity. There has been little public concern with *ex ante* risks per se, nor, for that matter, with the value of life or the utility derived from it as envisaged by economists. This implies that, to the extent that CBA is specifically designed to assist with *social* decision making (and only to this extent)[7] the *ex ante* concept referred to by Mishan and other defenders of conventional welfare economics is an inappropriate basis for measurement. In Robinson's words:

> "It is clear that the subjective orientation of the willingness to pay approach could lead to an allocation of public funds in a manner inconsistent with the principles of cost effectiveness. These principles maintain first and foremost that governmental energies should be devoted to those areas where the potential improvements in health status and longevity are greatest. There is no reason to assume that the most cost effective programs, where benefits are measured in terms of mortality and morbidity statistics, would, in every case, be those most appreciated by the citizenry. Indeed, it is precisely in impatience with the allocation of public funds according to the subjective preferences of the politically powerful sectors of the population, rather than according to objectively measurable standards of maximum effectiveness, that lies at the root of economists interest in cost benefit and cost effectiveness analysis."
>
> Robinson, 1986 (p.148)

The Ethical Basis of CUA

The principle ethical question to be resolved in cost utility analysis is the way in which individual evaluations should be aggregated and, therefore, the way in which interpersonal comparisons are to be made. As noted by Torrance (1986) the basic assumption in CUA is

[7] In the private sub sector of the market or in some public projects the explicit objective may be the satisfaction of private, consumer demand. In these circumstances WTP would clearly be appropriate, but such projects are unusual in the health evaluation literature.

that "the difference in utility between being dead and being healthy is set equal across people. In this way the method is egalitarian ... each individual's health is counted equally" (p.17). As a result the QALYs or HYEs gained by different individuals may simply be summed. In Mooney and Olsen's (1989) phrase this implies a form of "quasi utilitarianism" in which the maximand is not total utility but a weighted average of individuals utilities where the weights are designed to treat individuals equally irrespective of the absolute intensity of their preferences.[8]

Such quasi utilitarianism necessarily conflicts with other ethical bases. Libertarians would reject an aggregation rule which constrained an individual's right to reveal their own preferences, that is, through their willingness to pay. More generally, those subscribing to a deontological view of ethics would argue that resource allocation should be determined, not only by consequences but by "ethical rules" and "human rights". Thus, for example, in his critique of QALYs Harris (1987) argues that the only priority in health care should be the preservation of life and that all have an equal right to life no matter what its length or quality. In his critique of this paper Williams (1987) notes the incompatibility of the ethical bases and simply argues that "at the end of the day we simply have to stand up and be counted as to which set of principles we wish to have underpin the way the health care system works" (p.123). Proponents of CUA must commence with the view that there will be widespread acceptance of the ethical principle that, all else equal, more healthy year equivalents should be preferred to less.

3. Cost Utility Analysis

The Measurement of Quality[9]

It has been recognised for a long time that evaluation studies should include the measurement of the quality of life and a very large number of health status measures have been developed.[10] However, the usefulness of these instruments has been variable. Many

---

[8] This `quasi-utilitarianism' is, in fact, the same as Jeremy Bentham's original statement of utilitarianism.

[9] The most comprehensive statement on utility measurement is Torrance (1986).

[10] Some of the most well known are the Sickness Impact Profile (SIP) (Bergner et al 1976), the Spitzer QL index (Spitzer et al 1981), the

have had poor if any evidence of validity or reliability and the purpose for which they were developed has varied. While these instruments may have contributed in a general sense to the "evaluation of health outcome" they have often been unsuited to the specific question addressed in economic evaluation, namely, whether or not they indicate a treatment which should be chosen in preference to some other treatment for the same or for some other disease.

The latter question is explicitly addressed by cost utility analysis (CUA). Projects or options are ordered according to the cost per Quality Adjusted Life Year (QALY) attributable to the project. All else equal, the most desirable options are taken to be those which result in the cheapest QALYS. That is, QALYS are the criterion of value in the sense that more are better and, all else equal, projects with more QALYS should be preferred. Despite the recognition of numerous practical problems there appears to be a fairly widespread acceptance of the steps involved in the calculation of QALYS. They are estimated as expected life years times an index of "utility", where this is measured on a 0-1 scale and is taken as quantifying that aspect of the quality of life upon which decisions should be made.[11]

Five techniques have been commonly used which purport to measure utility directly. These involve the use of category rating or a rating scale (RS), the standard gamble (SG), the time trade off (TTO), equivalence techniques (ET) and magnitude estimation (ME). (These are described in Appendix 1). Each technique involves the presentation of a health state description to interviewees and the eliciting of their preferences for the health state relative to some reference states, usually full health and death. The scale is calibrated by setting the value of these reference states equal to unity and zero respectively. The utility revealed by these techniques is taken as having an interval property. Thus, for example, the difference between utility values of 0.2 and 0.4 is treated as being quantitatively equivalent to the difference between 0.6 and 0.8. The property is required for the valid summation of utilities.

A second, indirect, approach to measurement requires the prior establishment of a "multi

---

Quality of Wellbeing (Kaplan et al 1976) and the Nottingham Health Profile (Hunt et al 1986). For review of the major scales see McDowell and Newell (1987).

[11] For a discussion of these steps see Gudex and Kind (1988).

attribute" utility (MAU) scale which may be applied to any health state. Three commonly used scales have been devised by Rosser and Kind (1978), Torrance (1982) and Kaplan et al (1976, 1982). With each of these, a health state is broken down into different "attributes" or "dimensions" such as "physical functioning", "socio-emotional function" and "health problem". Each attribute has a separate scale with scores initially determined from interviews using one of the techniques listed above. A particular health state may then be measured and scored on each scale and the scores combined with a predetermined formula.

Using both the multi attribute and the holistic approach to measurement a large number of health states have been assessed and "league tables" published to show the utility of different states. One of these is reproduced in Table 1. Despite the temptation to use such tables to produce a mechanistic ranking of project priorities they can only serve as one input into decision making. The reason for this is illustrated by the existence of negative utilities - health states where death would be preferred. There is clearly no policy imperative to administer euthanasia as this may (or may not) be overridden by some other ethical, political or distributional consideration. Despite this and other qualifications discussed below, it is commonly accepted that "CUA can no longer be considered as being in the experimental stage but is now at the point where it merits serious consideration by health care decision makers" (Drummond, 1987). At least one regional authority in the UK has explicitly employed it as an aid to decision making (Gudex, 1986).

Practical Issues

While not necessarily disagreeing with the conclusion above, it should be recognised that there are still a number of very fundamental conceptual and practical questions which are unresolved. Tables 2-4 list the chief issues noted in the literature and classifies them (sometimes a little arbitrarily) as dealing with measurement, interview technique or theory.

The first of these groups of issues is summarised in Table 2. A basic requirements of measurement is that it should be valid (measure what it purports to measure), internally reliable (repeated measurements should produce the same result) and have an acceptable level of precision (standard deviation of individual measurements). These are empirical issues and there has been little such developmental research reported in the literature. Validity is commonly established by a comparison of results against a gold standard. The

conventional - but not unanimous - view is that this is provided by the standard gamble, the technique derived from the von Neumann-Morgernstern axioms of consumer choice. However most of the techniques have not been validated against this yardstick. In some instances other approaches to validation have been used. The rating scale and magnitude estimation techniques are derived from the psychometric literature and proponents point to this in support of the procedures. Few studies have been undertaken to determine internal reliability. While the results to date appear "fairly satisfactory" (Torrance, 1986) they do not represent a very secure basis for a new branch of health measurement.

A second and related issue to these questions is whether or not measurement techniques produce the same result. If so, then confidence in each is increased as it would support the view that they are measuring a common quantity. Once again, however, evidence is scarce. A reasonable correspondence has been found between the time trade-off and standard gamble. While Richardson et al (1990) also find a close correlation between these, they were also able to reject the hypothesis that the two measures, or the rating scale gave the same result for the health state being measured. In the same study it was found that the two multi attribute scales investigated seriously underestimated utility - a result predicted by the authors as the health state - breast cancer - involved significant psycho-social costs and these are not adequately measured by the two MAU scales. A reasonable inference from this latter result is that, while quicker and cheaper to use, the all-purpose MAU scale is likely to give a less accurate result than the health state specific vignette approach to measurement. Finally, Torrance (1976) has suggested that power function transformations may improve the relationship between the Rating Scale and Time Trade Off. The function reported by Torrance could not be reproduced by Richardson, et al (1990) and it did not improve the correlation between the results of the RS and TTO.

An important measurement issue which has received some recent attention is whether or not a single index of utility is applicable to a health state over an extended period of time. QALYS are calculated by multiplying life years by an index of utility. The question is whether or not the value of the index varies with the stage of a person's life and with medical prognosis. More generally, the functional relationship between a person's utility and time may be incompatible with a single yearly index. An additional problem is that future benefits - QALYS - are normally discounted in CUA studies in an orderly, exponential, way using an estimate of the social rate of time preference. However, there is

little evidence to date that individuals behave in this way when evaluating future benefits.

Health states are unlikely to remain unchanged until death. The usual approach to this problem is to evaluate each of the health states which will be encountered and then to sum the discounted QALYS experienced in each health state on the assumption that the health states are independent. If the assumptions discussed above are invalid then this procedure is also invalid. Mehrez and Gafni (1989a) have suggested that, for this reason, the usual composite approach to the calculation of QALYS should be replaced by a holistic measure of the utility of an entire multi-period scenario. They suggest that the resultant utility be converted into Healthy Year Equivalents (HYE's). The suggested change in title - HYE to replace QALY -would emphasise the new methodological basis for the calculation. Richardson, et al (1989) have investigated this issue empirically. They found that the holistic and composite approaches to a multi-state scenario give very different results, thereby supporting Mehrez and Gafni's suggestion.

As reported in Table 3 there has been a fairly substantial investigation of the importance of the interview. This draws upon an extensive literature on the subject outside the context of CUA. As elsewhere, it has been found that results are sensitive to the way in which health states are described and the way in which scales are presented. Two additional questions are relevant. The first is the choice of subject for interview. Some have argued for the use of patients or health professionals who have experienced or observed the health state and can appreciate and evaluate it more accurately. Others have argued that a random cross-section of the society should be interviewed as it is society's resources that are allocated to health programs. A further view is that if CUA is to reflect consumer sovereignty, interviewees should be potential patients as it is the values of this group that would be relevant if individual decision making was possible. The second question is whether stated preferences in the context of an interview correspond with the preferences individuals reveal when faced with a real choice. There has been no evaluation of this difficult issue in the CUA literature. However, results from transport economics suggests that "stated preferences" closely correspond to "revealed preferences" when interviewing is carried out in an appropriate way.

Theoretical Issues : The Unit of Measurement

The theoretical issues discussed in the literature are summarised in Table 4. Only three of these will be considered below. The first returns to the question of the unit of measurement and the value basis of CUA. The second discussed in section 4 reconsiders an old issue in the CBA literature namely, whether or not the "indirect" production benefits resulting from averted death should be included using the gross value of production or the value net of consumption. Thirdly, there is a consideration of how these "gross or net" benefits should be included in the analysis.

In Section 2 of the paper it was noted there have been two traditions in economics reflecting two sets of social values. In the first (material welfare) tradition value is "external"; in the second, which commenced with Robbins, a value is internal. It is equated with the strength of an individuals preferences and normally revealed through the individual's willingness to pay. Cost utility analysis represents an interesting fusion of the two traditions. On the one hand, the QALY is directly related to an external measure - the life year - and one which combines both dimensions of public policy, namely, morbidity and mortality. On the other hand, the evaluation of healthy year equivalents is based upon individual preferences. The issue which has not been resolved in the literature is which of the available measurement techniques reflects, most accurately, the intensity of an individual's preferences. Most of the theoretical discussion in this respect has been concerned with the adequacy of the von Neumann-Morgernstern axioms. These must describe individual behaviour if the standard gamble is to be a satisfactory gold standard for the measurement of preferences. The theoretical issue is whether or not the axioms allow for the "specific utility of risk" - something which von Neumann and Morgernstern did not claim (and explicitly disclaimed) but which subsequent writers accepted. There has now been a significant body of empirical research to determine whether the axioms are supported by observed behaviour. As noted earlier in the paper the outcome of this debate is that the axioms cannot be accepted as generally true in all contexts and that, consequently, the standard gamble cannot be accepted as a gold standard for measurement.

Both the standard gamble and the measurement techniques observed from the psychometric literature have a common feature. The measurement requires information on two separate relationships. The first is between the health state described to individuals and the scale used by the technique (centimetres on the rating scale; probabilities with the standard gamble); the second is between the scale and the intensity of individual choice. In

each of these cases the second relationship is unknown. In the psychometrics literature it is simply assumed that there is a one to one relationship between the scale and preferences and that the scale has an acceptable interval property. As there is no linear relationship between the results of the different psychometric scales this latter conclusion must be wrong. In the case of the standard gamble the second relationship is confounded by the specific utility of risk.

For this reason, Richardson (1990) has tentatively suggested that the preferred measurement technique should be the time trade-off. With this, individuals indicate directly the number of healthy years which are considered to be equivalent to a given number of years in the health state being evaluated. Preferences are directly measured by the healthy year equivalents without the need for a confounding, intermediate, scale. The healthy year equivalent (HYE) revealed by subjects is the unit of output.

4. Indirect Benefits

What to include

The direct effect of a successful medical intervention may be an increased number of life years or QALYS. However, if the beneficiaries are in the workforce there will also be an increase in the value of future production. This has been treated two ways in the literature. In some UK studies it has been ignored, the argument being that the purpose of the UK National Health Scheme is to maximise health and that it is not concerned with other objectives (see Williams (1985)). While resolving very decisively the issues raised below this is an anomalous position for an economic analysis to adopt. As Klarman (1965) notes in his discussion of the issue "there is a distinction between a health program that saves people from death for useful labour and one that saves people from death to pursue an unproductive life". (p.380) The more common approach has been to subtract the present value of indirect production benefits from the direct cost of the procedure to obtain a net resource cost per QALY.

At the end of the 1960's there was some discussion of whether or not the appropriate quantity to be included in an economic evaluation was the full value of production or the value net of consumption. In the words of Prest and Turvey (1965)[12] "if society loses the production of the decadent, does it not also gain by not having to supply his own consumption" (p.722-723).

The answer given to this question depended upon the viewpoint explicitly or implicitly adopted on the composition of "society". Dowie (1970) notes that it is the "slipshod use of "society", "community", "nation" [that] is responsible for the bulk of the confusion in the health literature" (p.27). He quotes approvingly Weisbrod's resolution of the issue.

> "The choice between the two measures of the economic value of a person - present value of gross or net future earnings - rests upon the viewpoint taken ... If "society" is defined to include everyone, including the individual whose value is being considered, then his contribution to the group is the total value of his output, and

[12] Quoted from Dowie (1970). Also see Dowie for a review of this debate.

> his "value" is the present value of his *gross* future earnings. But if "society" is so defined as to exclude the individual whose life is being valued (for example, as all those who would be left were he to die), then his contribution to "society" consists only of any excess of what he adds to total output over what he subtracts from it, his consumption; and his economic worth is the present value of his *net* future earnings". (p.35-36)

Others did not adopt such a neutral position as Weisbrod. Mishan (1971), for example, simply asserts that the net production method "is not satisfactory for the simple reason that it has no regard for the feelings of the potential decadents. It restricts itself to the interest only of the surviving members of society: it ignores society *ex ante* and concentrates wholly on society *ex post*" (p.690). There are, as Dowie (1970) notes, a large number of potential social groupings whose members' feelings may or may not be included. A more persuasive argument for the *ex ante* definition of society is provided by Prest and Turvey:

> "... the society whose representatives decide whether or not to undertake a measure which would save lives includes those people who may lose their life if the proposed measure is not undertaken. Hence, so the argument might run, society is relevantly defined as including the prospective decadent and his consumption is part of the social loss contingent upon his death" (p.723 quoted from Dowie (1970)).

These arguments appear to have been decisive and the overwhelming majority of the evaluation and cost of illness studies in the health economics literature have used the gross value of production as the preferred measure of material benefits. The arrival of CEA and CUA alters the logic of the previous arguments. However this does not appear to have been recognised in the literature and the "material benefits" of a person's life continues to be included using the gross measure. Previously, and in the tradition of the material welfare school, the value of life was equated with the value of the consumption benefits. Some of these were enjoyed by the individual producing the income (measured consumption) and some were enjoyed by unidentified persons who were the beneficiaries of the individual's savings and tax payments. Both CEA and CUA *replaced* this method of valuing life with an analysis in which the life year or QALY was itself the unit of output. From biological necessity a life year must involve consumption and unless an explicit reference is made to the level of disposable income it is likely that respondents to a CUA interview would

assume this to be unchanged in the different health states. This implies that the inclusion of life years as a benefit and future consumption as a negative cost involves double counting of these benefits.[13]

Despite the prima facie appearance of double counting, the gross measure of output might be defended on the grounds that for "conceptual clarity" material benefits should be separated from the non-material benefits of a life per se. The former may be precisely quantified whereas the latter involves ethical and emotional issues that confound measurement. Following this line of argument, the life year does not measure the intensity of the material benefits experienced so that these may be separately measured. Even if it is the individual gaining the life years who also receives the consumption benefits this is an issue of distribution, not production and, in principle, could be subjected to the potential compensation principle.

The argument is not compelling. Even accepting the ex ante view of society and the concept of a life year uncoupled in some sense from material benefits (and it is not obvious that such an uncoupling results in "conceptual clarity") a closer examination of the distributional consequences of the argument is likely to result in its widespread rejection. Suppose that two individuals, A and B consume all of their income, C_A, C_B and these individuals require interventions which would result in the same number of additional life years for each. According to the previous argument, priority would be determined by comparing the "net resource cost" (T_i - C_i) where T and C represent the treatment cost and individual's consumption respectively. The two individuals would be equally likely to receive the treatment when T_A - C_A = T_B - C_B. By assumption, neither A nor B contributes to the collective provision of health care and, consequently, T_A and T_B represent the amount the remainder of the society is prepared to pay for the person's treatment. The equation indicates that the higher the value of consumption the higher is the value of T that

[13] This issue has not been discussed in the literature or subject to empirical testing. It is not necessarily a simple matter. It is certainly likely that respondents to some health states scenarios would assume that the *pattern* of consumption would change. For example, the physically disabled would have to substitute passive for active forms of recreation. In cases of extreme disability it may be physically impossible to consume at the same level. In the case of permanent institutionalisation the entire cost of living might be included in the measurement of direct costs. While such cases may be exceptional they indicate that the extent of the double counting may vary from case to case.

the society is prepared to pay. In general, this rule requires non recipients of the benefits to pay more, and implicitly value more highly, an individual when that individual consumes more. While conceivable, this situation appears to be an improbable description of social values and precisely the situation that the introduction of CEA and CUA was intended to avoid. The rhetoric and provision of health care in most national health schemes is more compatible with the view that the willingness to pay by non recipients is determined by the presumed impact on life and the quality of life, that is, by the units adopted directly in CUA. This implies that the relevant benefits are fully included in the measurement of QALYs and that a further adjustment for consumption will indeed lead to double counting.

The assumption in the previous example that consumption is equal to the full value of income is not generally true. Normally the present value of an individual's consumption is less than the present value of income because of taxation and savings, the cumulative value of which is passed on as an inheritance to children and other heirs. For the recipient of state funded pensions this situation might be reversed and the value of consumption could exceed the value of an individual's earned income. Consequently, there are a number of alternative "net resource costs" associated with an intervention. Using the notation: T = direct treatment costs; Y = earned income; C = consumption; X = tax; S = savings; P = pensions; and where all the values represent the cumulative lifelong present value of the magnitudes these are:

(i) (T - Y): This is the usual measure discussed above. It implies a higher willingness to pay by the rest of society as a person's own consumption rises.

(ii) (T - (Y - C)) = T - (S + X): This is the net measure also discussed above. It assumes a greater willingness to pay as the individual contributes more to the society via taxation. It also assumes a greater willingness to pay as the individual accumulates greater wealth for his or her heirs.

(iii) T - (Y - C - S) = T - X: Treatment costs net of tax excludes the benefits received by heirs and assumes that the willingness to pay of society rises with a person's contribution to the general social wealth.

(iv) T + P: This is the special case of (ii) in which Y = 0 and C = P. For consistency, if the general contribution of individual's to the society via taxation is considered relevant to the resource cost then a negative contribution via pension payments may also be considered relevant.

The four options do not indicate a different concept of cost *per se* but rather a different view of the social groups whose benefits should be included in the analysis. The need to specify this arises both from the fact that in CEA and CUA the benefits to an individual of extending their life are measured separately from other material benefits and from the fact that the distributional affects of these benefits are very precisely focused. The argument here may be summarised by saying that under these circumstances the cost benefit question should not be posed in terms of abstract "net social costs" since, as Dowie notes, the term "society" is too hazy. Rather it should specify the group which bares the costs and the group whose material benefits are considered to be relevant. The latter decision is unavoidably value laden. The fourth measure above makes this most obvious. All else equal the inclusion of pensions as a cost to the remainder of society will reduce the likelihood of their provision. It is, however, an inescapable fact that pensions involve a cost to the remainder of society. There is, no "correct" measure or "preferred" measure as the choice between these options depends upon distributional judgments. As noted some English writers disregard all of the indirect benefits. It seems likely, however, that most would wish to include at least the value of positive tax payments in the analysis. The treatment of savings and pensions is more problematical.

Combining Direct and Indirect Benefits

Following a decision about the appropriate definition of material benefits there remains the question of how these should be combined with other relevant information on costs and benefits in order to evaluate a project. With the notable exception of Linard (1990) the complexities of this task do not appear to have been fully appreciated in the health literature. Linard's conclusion is that there is no satisfactory means of combining costs and benefits when these are measured in different units. This implies that CUA and CEA have been unsuccessful in their objective of avoiding the need to evaluate life years in dollars.

The problem does not arise because `there is no dollar equivalent to a life year'. If the ethical basis for valuation is some variant of consequentialism where outcomes have finite values then in principle life years must have a dollar equivalent. The problem is that within the very broad scope of consequentialism (which includes both of the traditions in the economics literature discussed earlier) there is no concensus about the appropriate ethical

basis and no means for *quantifying* value the (finite) dollar equivalent of a life year or QALY. Their value or net present value is therefore unknown and cannot be combined with costs or benefits denominated in dollars. The dilemma is presented in Table 5. In the first row the "net cost to benefit ratio" will product a correct ranking of projects when there is no budget constraint. However, with a fixed budget an organisation will not maximise either life years or (the unknown) net present value (of the QALYs plus other net benefits) by the use of this ratio to prioritise its projects. This self evident result follows from the fact that the budget may be quickly exhausted upon projects with a large budgetary cost but a relatively low net cost while a superior outcome (however this is measured) might be achieved through expenditures on projects with small budgetary, but relatively larger net costs. In this context only the budgetary costs should be in the denominator as this will indicate where the greatest benefit is achieved per unit of the constrained resource. In principle, social benefits will be maximised with a fixed budget by ranking projects according to the ratio of the net social benefit to budgetary costs (row two). In practice this encounters the very problem that CEA and CUA were designed to avoid, namely, the need to combine benefits and costs measured in incompatible units. This is the major conclusion of Linard (1990) who notes that

> "In a situation of sectoral budget constraints the non-budget "dollars" should be aggregated with the effectiveness or utility "benefits" for comparison with the budget outlay. This, of course, destroys the simplicity of the ratio, creating major problems for comparison or interpretation. However having a "simple" ratio is of no benefit if it is irrelevant to the objective function". (p.12)

It is important to note in this quotation that there is only a dilemma for project evaluation when there is an effective budget constraint. The importance of this fact depends upon the purpose of CBA and that there is disagreement about this. One view is that its objective is simply the determination of the actual projects to be undertaken by a department or enterprise. It is from this perspective that the usual evaluation ratio used in CEA and CUA is incorrect or misleading. Linard (1990), for example, argues that:

> "an evaluation of an isolated proposal which reports that "... benefits are greater than costs therefore the project is worthwhile ...", or that the "... benefit/cost ratio is greater than 1.0 therefore the project is worthwhile ...", is implicitly assuming that

> unlimited resources are available, that there are no budget constraints. Using Mishan's terminology, such studies are "*economically irrelevant*". Worse, they are misleading to the public and the decision makers".

The contrary view is put forward by Keeler and Cretin (1987) who argue that:

> "In the health field, cost-benefit and cost effectiveness analysis have been carried out for two purposes. The first and more frequent application found in the literature uses cost-benefit or cost effectiveness analysis as a tool to describe and evaluate new and promising health care programs. ... the goal is to improve the policy debate rather than to specify the correct policy. The second, less frequently reported use addresses the problem of how an agency with a fixed budget should spend its funds to maximise net social benefit" (p.276).

With this more commonly accepted perspective on the purpose of CBA, the ratio of net cost to life years is of value as it indicates whether or not health program resources are being allocated efficiently by a society and whether or not there is a strong prima facie case for the consideration of a particular project.

From the perspective of a particular department responsible for prioritising health projects this latter conclusion is unhelpful, and at present there would appear to be only *ad hoc* solutions to Linard's dilemma. For example, making the heroic assumption that investment within a department is based upon rational planning, the value of life within that agency could be set equal to the maximum expenditure currently undertaken to save the life of a person plus the estimated value of the indirect benefits. This would provide a minimum estimate of the value of life for investment purposes within that agency.[14] If the assumption of rational planning is too fanciful - and in many institutions it may appear like oxymoron - then a figure could be selected for planning purposes, possibly based upon observed

---

[14] For the most marginal project let
$$(Q + OB)/C = \lambda$$
where Q is the value of life;
OB is the dollar value of other indirect benefits, and
C is the budgetary cost,
λ is the "shadow price of capital" in this context. From this:
$$Q = \lambda C + OB$$
The minimum value of life, Q, is obtained when $\lambda = 1$.

willingness to pay elsewhere. The adoption of such a heuristic would not be equivalent to a reversion to a pre CEA/CUA methodology. The importance of the estimate would be limited by the relative importance of the indirect benefits and, on many occasions, these would be small. In general, they would fall with the increasing age of beneficiaries and as the relative importance of the life years increased. Similarly, the magnitude of any error in the procedure would fall if indirect benefits were measured by the value of future taxation or net savings and not by the full value of production. While it would make no difference to the ranking of health projects if lives were converted into dollars or vice versa, in practice it would appear preferable to convert dollars into life years especially when the indirect benefits were relatively unimportant. This would discourage the inappropriate comparison of health related and financial projects.

5. Conclusions

The chief conclusion of this paper is that there are still significant unresolved issues in the theory and application of cost utility analysis. Most economists have uncritically assumed that the theoretical apparatus of conventional welfare theory is the appropriate gold standard in all decision contexts. It has been argued in this paper that this is not necessarily an appropriate assumption and that economic evaluation in the health sector has, in fact, been based upon a different value system. The question is not which is right and which is wrong; rather, it is an issue of which *do we wish* to have as the basis for our resource allocation. At present, QALYs or HYEs imply utilitarian values to the extent that value - as reflected in the determination of equivalent health years - is defined by people's preferences. However as each persons HYE is given an equal vote in the allocation of resources, CUA is more accurately described as quasi -utilitarian.

The second issue of conceptual confusion discussed here has been the inclusion of indirect production benefits when the benefits of life are *separately* measured by HYEs. The current convention in welfare economics and cost benefit analysis is to separate issues of production and distribution. It has been argued here that in the health sector this is not possible at least with respect to indirect benefits. The chief beneficiaries of increased output from a life extending intervention are the individuals whose lives have been separately included as a benefit. Double counting can only be avoided by netting out consumption from the value of production. However, even the remaining net benefits have

very precise distributional consequences. The major beneficiaries are the heirs of the individual whose life is saved. For this reason it has been suggested that the rather vague concept of "social benefits" and "social costs" be replaced with an analysis which compares the costs to be borne by an identified group with the benefits to be gained by an identified group. The relevant policy decision is whether or not the former group will accept these costs in order to bestow the benefits on the latter group.

Even with a satisfactory resolution of these conceptual issues the present paper has documented a formidable list of unresolved questions. These are primarily associated with the measurement of utility. In a recent critique, Carr-Hill (1989) concludes by "doubting the utility of a global index given the wide variety of assumptions involved". The contrary view put forward by advocates of CUA is that the underlying assumptions are implicit in any decisions that are made and that the major strength of CUA, as with CBA more generally, is that the assumptions are made explicit; that they may be located and subjected to sensitivity analysis.

The unresolved questions outlined in this paper should be kept in perspective. While their resolution will improve and refine the methodology and provide a more satisfactory theoretical basis for an evaluation, in many cases this is unlikely to result in a substantial, quantitative, change in the measurement and the ranking of projects. Even the unresolved theoretical issue of combining QALYs with other benefits will be comparatively unimportant in the large number of projects where the ratio of indirect (other) net benefits to QALYs is small. The available studies show differences between the major quality of life techniques but a surprising similarity in the order of magnitude of the results. Other issues, such as the distribution of benefits and the choice of a discount rate are not unique to CUA or to the economic evaluation of health care. They are issues which must be faced in any evaluative methodology. Perhaps the most fundamental defence of CUA is that at present there is no alternative methodology which avoids the deficiencies of the willingness to pay approach to valuing human life but which simultaneously takes into account the strength of preferences for a health state. The cost utility ratio does not summarise all of the factors relevant to the assessment of health projects but it does embody a sufficiently important component of the relevant information and in a sufficiently comprehensible format that it should narrow the breadth of disagreement and limit the scope for arbitrary decision making.

References

Abel-Smith, B. (1985), "Global perspective on health service financing", *Social Science and Medicine*, 21, 9, 957-963.

Birch, S. and Donaldson, C. (1987), "Application of cost benefit analysis to health care: departures from welfare economic theory", *Journal of Health Economics*, 6, 3, 221-226.

Boyd, N.F. et al (1982), "A comparison of methods of assessing value quality in laryngeal cancer", in Deber, R.B. and Thompson, (eds.) *Cancer in Health Care: Decision Making and Effectiveness*, Department of Health Administration, University of Toronto, 141-144.

Broome, J. (1978), "Trying to value a life", *Journal of Public Economics*, 9, 91-100.

Broome, J. (1979), "Trying to value a life: a reply", *Journal of Public Economics*, 12, 259-262.

Brooks, R.G. (1988), *Scaling in health status measurement: outline guide and commentary*, Swedish Institute of Health Economics.

Buchanan, J.M. and Faith, R. (1979), "Trying again to value a life", *Journal of Public Economics*, 12, 245-248.

Buxton, M.. Ashby, J. and O'Hanlon, M. (1986), *Valuation of health states using the time trade-off approach*, HERG Discussion Paper No. 2, Health Economics Research Group, Brunel University, Middlesex.

Carr-Hill, R.A. (1989), "Assumptions of the QALY procedure", *Social Science and Medicine*, 29, 3, 469-477.

Churchill et al (1984), "Cost effectiveness analysis comparing continuous ambulatory peritoneal dialysis to hospital haemodialysis", *Medical Decision Making*, 4, 20-23.

Cooper, B. and Brody, W. (1976), *1972 lifetime earnings by age, sex, race and education level*, Research and Statistics Note No. 12, U.S. Department of Health, Education and Welfare, Washington.

Cooper, B. and Rice, D. (1976), "The economic cost of illness revisited", *Social Security Bulletin*, 39, 2, 21-36.

Dowie, J.A. (1970), "Valuing the benefits of health improvement", *Australian Economic Papers*, 9, 14, 21-41.

Drummond, M. (1981), *Studies in Economic Appraisal in Health Care*, Oxford University Press, Oxford.

Drummond, M. (1987), "Discussion: Torrance's approach to measuring health related quality of life", *Journal of Chronic Disease*, 40, 6, 601-603.

Enthoven, A.C. (1988), "Managed competition: An agenda for action", *Health Affairs*, 7, 3, 25-47.

Epstein, A.M., Hall, J.A. et al (1989), "Using proxies to evaluate quality of life", *Medical Care*, 27, 3, 91-98.

Evans, R.W. et al (1985), "The quality of life of patients with end stage renal disease", *New England Journal of Medicine*, 312, 533-539.

Evans, R.W. et al (1987), "The quality of life of kidney and heart transplant recipients", *Transplant Procedure*, 17, 1579-82.

Evans, D.B. (1990), "The discount rate for use in public sector projects", in Scotton, R.B. and Goss, J., *Economic Evaluation of Health Services*, Australian Institute of Health, Canberra.

Gafni, A. and Torrance, G.W. (1984), "Risk attitude and time preference in health", *Management Science*, 30, 4, 440-451.

Gudex, C. (1986), *QALYS and their use by the health service*, Discussion Paper 20, Centre for Health Economics, University of York.

Harris, J. (1987), "QALYfying the value of life", *Journal of Medical Ethics*, 13, 117-123.

Hellinger, F.J. (1989), "Expected utility theory and risky choices with health outcomes", *Medical Care*, 27, 3, 273-279.

Hershey, J.C., Kunreuther, J.C. and Schoemaker, P.J. (1982), "Sources of bias in assessment procedures for utility functions", *Management Science*, 29, 8, 936-954.

Hershey, J.C. and Schoemaker, J.H. (1985), "Probability versus certainty equivalence methods in utility measurement: are they equivalent", *Management Science*, 31, 10, 12, 13-1229.

Hodgson, T. and Meiners, M. (1979), *Guidelines for cost illness studies in the public health service*,
Public Health Service Task Force on Cost of Illness Studies, Washington.

Jones-Lee, M.J. (1979), "Trying to value a life", *Journal of Public Economics*, 12, 249-256.

Kaplan, R.M., Bush, J. and Berry, C. (1976), "Health states: types of validity and the index of well being", *Health Services Research*, 11, 4, 478-507.

Kaplan, R.M. and Bush, J.W. (1982), "Health related quality of life measurement for evaluation research and policy analysis", *Health Psychology*, 1, 61-8.

Klarman, H.E. (1965), "Syphilis control programs", in Dorfman, R. (ed.), *Measuring benefits of government investments*, Brookings Institution, Washington D.C., 367-414.

Kuhse, H. and Singer, P. (1988), "Age and the allocation of medical resources", *The Journal of Medicine and Philosophy*, 13, 101-116.

Landefeld, S. and Seskin, E. (1982), "The economic value of life: linking theory to practice", *American Journal of Public Health*, 72, 6, 555-566.

Linard, K. (1990), "Decision contexts and decision rules for economic evaluation of health services", in Scotton, R.B. and Goss, J., *Economic Evaluation of Health Services*, Australian Institute of Health, Canberra.

Lipscomb, J. (1989), "Time preference for cost effectiveness analysis of medical care", *Inquiry* 27, 3, 233-253.

Llewellyn-Thomas, H., Sutherland, H.J. et al (1984), "Describing health states: methodological issues in obtaining values for health states", *Medical Care*, 22, 6, 543-552.

Loomes, G. and McKenzie, L. (1989), "The use of QALYS in health care decision making", *Social Science in Medicine*, 28, 4, 299-308.

Mehrez, A. and Gafni, A. (1989a), "Quality adjusted life years, utility theory and healthy year equivalents", *Medical Decision Making*, 9, 2, 142-149.

Mehrez, A. and Gafni, A. (1989b), *Evaluation health related quality of life: interpretation for the time trade off technique*, Working Paper, Centre for Health Economics and Policy Analysis, McMaster University.

Mishan, E.J. (1971), "Evaluation of life and limb: a theoretical approach", *Journal of Political Economy*, 687-705.

McDowell, I. and Newell, C. (1987), *Measuring health: a guide to rating scales and questionnaires*, Oxford University Press, New York.

Mooney, G. and Olsen, J.A. (1989), *QALYS: where next?*, mimeo, Institute of Social Medicine, University of Copenhagen.

Mushkin, S. (1962), "Health as an investment", *Journal of Political Economy*, 70, 5, 129-157.

Pope, R.E. (1989), "Additional Perspectives on Modelling Health Insurance Decisions", in Selby Smith, C. (ed), *1989 Economics and Health*, Public Sector Management Institute, Monash University, Melbourne.

Read, J.L., Quinn, R.J., Berwick, D.M., Fineberg, H.V. and Weinstein, M.C. (1984), "Preferences for health outcomes: comparison of assessment methods", *Medical Decision Making*, 4, 3, 315-329.

Rice, D. and Cooper, B. (1967), "The economic value of human life", *American Journal of Public Health*, 57, 11, 1954-1966.

Richardson, J. (1987), "Ownership and Regulation in the Health Care Sector", in Abelson, P. *Privatisation: An Australian Perspective*, Australian Professional Publications, Sydney.

Richardson, J. (1990), "Cost utility analysis: what should be measured - utility, value or healthy
year equivalents", paper presented to the Second World Congress on Health Economics, University of Zurich, Switzerland, September 10-14, 1990.

Richardson, J., Hall, J. and Salkeld, G. (1990), "Cost utility analysis: the compatibility of measurement techniques in the measurement of utility through time" in Selby Smith, C. (ed.), *Economics and Health 1989*, PSMI, Monash University, 31-60.

Robinson, J.C. (1986), "Philosophical origins of the economic valuation of life", *The Millbank Quarterly*, 64, 1, 133-155.

Rosser, R. and Kind, P. (1978), "A scale of valuations of states of illness: is there a social consensus?", *International Journal of Epidemiology*, 7, 4, 347-358.

Sackett, D.L. and Torrance, G.W. (1978), "The utility of different health states as perceived by the general public", *Journal of Chronic Diseases*, 31, 697-704.

Schieber, G.J. and Poullier, J.P. (1989), "International health care expenditure trends: 1987", *Health Affairs*, 8, 3, 169-177.

Schoemaker, P. (1982), "The Expected Utility Model: its variants, purposes, evidence and limitations", *Journal of Economic Literature*, 20, 529-563.

Scotton, R.B. (1990), "Integrating Medicare with private health insurance: the best of both worlds", in Selby Smith, C. (ed.), *Economics and Health 1989*, Proceedings of the Eleventh Australian Conference of Health Economists, PSMI, Monash, 219-238.

Sutherland, H.J., Dunn, V. and Boyd, N.F. (1983), "Measurement of values for states of health with linear analog scales", *Medical Decision Making*, 3, 4, 477-487.

Torrance, G.W. (1976), "Social preferences for health states: an empirical evaluation of three measurement techniques:, *Socio-Economic Planning Sciences*, 10, 3, 129-136.

Torrance, G.W. (1982), "Multiattribute utility theory as a method of measuring preferences for health", in Kane, R. and Kane R. (eds.), *Values and long term care*, Lexington.

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

Torrance, G.W. (1987), "Utility approach to measuring health-related quality of life", *Journal of Chronic Diseases*, 40, 6, 593-600.

Torrance, G.W. and Feeny, D. (1989), "Utilities and Quality Adjusted Life Years", *International Journal of Technology Assessment in Health Care* (forthcoming).

Warner, K.E. and Luce, B.R. (1982), *Cost benefit and cost effectiveness analysis in health care: Principles, practice and potential*, Health Administration Press, Ann Arbor, Michigan.

Weinstein, M.C. and Stason, W.B. (1976), *Hypertension: A Policy Perspective*, Harvard University
Press, (Mass.), Cambridge.

Williams, A. (1979), A note on "Trying to value a life", *Journal of Public Economics*, 12, 257-258.

Williams, A. (1985), "Economics of coronary artery bypass grafting", *British Medical Journal*, 291, 326-329.

Williams, A. (1987), "Response: QALYfying the value of life", *Journal of Medical Ethics*, 13, 123.

Wilson, D.K., Kaplan, R.M. and Schneiderman, L.J. (1987), "Framing of decisions and selections of alternatives in health care", *Social Behaviour*, 2, 57-59.

Wright, S.J. (1986), *Age sex and health: a summary of findings from the York health evaluation survey*, Discussion Paper 15, Centre for Health Economics, University of York.

APPENDIX I*

UTILITY MEASUREMENT TECHNIQUES

1. Rating Scales

A typical rating scale consists of a line on a page with clearly defined endpoints. The most preferred health state is placed at one end of the line and the least preferred at the other end. The remaining health states are placed on the line between these two, in order of their preference, and such that the intervals between the placements correspond to the differences in preference as perceived by the subject. Variations on this procedure include category scaling in which a specified number of categories (for example 11 categories) are used. The most preferred health state is placed in category 1, the least preferred in category 11, and the others are placed in the category that seems most appropriate to the subject in order to reflect the strength of the subject's preference, assuming equal change in preference between adjacent categories. Other variations include visual aids such as a thermometer with a scale from 0 to 100 on a felt background with foam sticks labelled with the health states.

2. Standard Gamble

The subject is offered two alternatives. Alternative 1 is a treatment with two possible outcomes: either the patient is returned to normal health and lives for an additional t years (probability p), or the patient dies immediately (probability 1 - p). Alternative 2 has the certain outcome of chronic state i for life (t years). Probability p is varied until the respondent is indifferent between the two alternatives, at which point the required preference value for state i is simply p; that is, $h_i = p$.

3. Time Trade Off

The subject is offered two alternatives - alternative 1 : state *i* for time *t* (life expectancy of an individual with the chronic condition) followed by death and alternative 2 : healthy for time x < t followed by death. Time *x* is varied until the respondent is indifferent between the two alternatives, at which point the required preference value for state *i* is given by $h_i = x/t$.

4. Equivalence Technique

The subject is asked the following kind of question: "If there are *x* people in adverse health situation *A* and *y* people in adverse health situation *B,* and if you can only help (cure) one group (for example, due to limited time or limited resources), which group would you choose to help?' One of the numbers *x* or *y* can then be varied until the subject finds the two groups equivalent in terms of needing or deserving help. If *x* and *y* are the equivalent numbers as judged by the subject, the undesirability (disutility) of condition *B* is *x/y* times as great as that of condition *A*. By asking a series of such questions all conditions can be related to each other on the undesirability scale.

5. Ratio Scale (Magnitude Estimation Technique)

The subjects are asked to provide the ratio of undesirability of pairs of health states - for example, is one state two time worse, three time worse, etc. compared to the other state? Then, if state *B* is judged to be *x* times worse than state *A*, the undesirability (disutility) of state *B* is *x* times as great as that of state *A*. By asking a series of questions all states can be related to each other on the undesirability scale.

* Descriptions in this Appendix are reproduced directly from a more detailed description

in Torrance, (1986). Also see Brooks (1986).

TABLE 1 SOME UTILITIES FOR HEALTH STATES

Health State	Utility
Healthy (reference state)	1.00
Life with menopausal symptoms (judgment)	0.99
Side effects of hypertension treatment (judgment)	0.95-0.99
Mild angina (judgment)	0.90
Kidney transplant (TTO, Hamilton, patients with transplants)	0.84
Moderate angina (judgment)	0.70
Some physical and role limitation with occasional pain (TTO)	0.67
Hospital dialysis (TTO, Hamilton, dialysis patients)	0.59
Hospital dialysis (TTO, St John's, dialysis patients)	0.57
Hospital dialysis (TTO, general public)	0.56
Severe angina (judgment)	0.50
Anxious / depressed and lonely much of the time (TTO)	0.45
Being blind or deaf or dumb (TTO)	0.39
Hospital confinement (TTO)	0.33
Mechanical aids to walk and learning disabled (TTO)	0.31
Dead (reference state)	0.00
Quadriplegic, blind and depressed (TTO)	<0.00
Confined to bed with severe pain (ratio)	<0.00
Unconscious (ratio)	<0.00

Source: Torrance (1987).

TABLE 2 MEASUREMENT ISSUES IN CUA

Issue	Authors	Comments
1. Are utility measures reliable?	Torrance, 1976, 1982 Churchill, 1984 Buxton, 1988 Richardson et al, 1990a	Internal, test and re-test reliability "satisfactory". However correlation between tests is not high.
2. Are utility measures valid?	Torrance, 1986 Churchill et al, 1984, 1985 Evans, 1985, 1987	Validity is often defined using the SG as the gold standard. See Section 4.
3. Do different utility measures give the same result?	Torrance et al, 1976 Read et al, 1984 Buxton et al, 1986 Richardson et al, 1990.	. Patient and clinician's evaluations correspond. Comparison suggests: TTO < SG. . RS systematically differs. . Orders of magnitude of TTO, and SG very similar, but statistically significant differences exist.
4. Do MAU measures give the same results as direct measurement?	Buxton et al, 1986 Richardson et al, 1990	. Rosser scale must be transformed. . Similar order of magnitude of results. . Present MAU scales appear unable to measure disutility of psycho-social distress. . No comparisons of Quality of Well Being (QWB) and MAU scales.
5. Can results from one scale be to obtain results with another?	Torrance, 1976 Richardson et al, 1990	. Torrance improves transformed compatibility of RS and compatible TTO using a power function. . Richardson et al cannot replicate result.
6. Can utility results be applied	Loomes & McKenzie, 1989	. Standard deviation of

to individuals or only to populations?	Torrance & Feeny, 1989	individual measures is high. . For the mean value of populations it is not.

Issue	Authors	Comments
7. Will utility vary with Socio-Economic Status?	Sackett & Torrance, 1978	This single study found only a weak relationship.
8. Is the utility of a health state constant through time?	Sackett & Torrance, 1978 Mehrez & Gafni, 1989 Loomes & McKenzie, 1989 Richardson et al, 1990	. Daily utility falls with duration of condition. . Utility may vary with prognosis, stage of life, idiosyncratic time preferences. . Rate may differ when future generations' welfare is affected. . Suggests the need to measure an entire health scenario until death or return of full health, not an annual index of health state utility.

TABLE 3 INTERVIEWING TECHNIQUES

Issue	Authors	Comments
1. Who should be interviewed? . patients . cross section of population . health professionals	Torrance, 1986 Loomes & McKenzie, 1989 Epstein et al, 1989 Carr-Hill, 1989	Patients often give higher utilities than non patients for a given health state. . Claimed that patients and professionals have a greater understanding of the health state. . The counterclaim is that societal values including those (i.e. of non patients) are desired, not those of patients.
2. Are patient responses the overall same as non patients?	Epstein et al 1989	. High correlation for health, functional, social and emotional status.
3. How adequate are health state descriptions?	Boyd et al, 1982 Llewellyn Thomas et al, 1984 Torrance, 1986 Carr-Hill, 1989	. General agreement that good descriptions are a critical factor in accurate measurement but level of detail in practice varies greatly. . Sequence and method of presenting descriptions alters results.
4. Do utilities depend upon the `context' of the health state?	Brooks, 1988 Hellinger, 1989 Sutherland et al, 1983	Results differ significantly with context which casts doubt upon conventional theory.
5. Will utilities vary with the reference points used for a scale?	Hershey et al, 1982 Sutherland et al, 1983 Llewellyn Thomas et al 1982	Yes
6. Are utilities sensitive to found the `framing' of descriptions `framing' and labels?	Brooks, 1988 Hershey et al, 1982, 1985 Sutherland et al, 1983 O'Connor et al, 1987 Wilson et al, 1987 Sackett & Torrance, 1978	. Outcomes normally to vary with and use of labels. . O'Connor found no `framing' effect. . `Positive' language increases utility values.
7. Will stated preference (in an from interview) correspond with	Brooks, 1988 Carr-Hill, 1989	No reliable results but those transport economics supports

the
revealed preferences?

validity of stated preferences,
correctly obtained.

TABLE 4 THEORETICAL ISSUES IN THE MEASUREMENT OF UTILITY AND CUA

Issue	Authors	Comments
1. Is the theoretical basis of the SG acceptable: Is it a `gold standard'?	Torrance & Feeny 1989 Schoemaker, 1982 Hershey et al, 1982	. See section 4 of this paper. . The Von Neuman-Morgenstern axioms are often contradicted. . The axioms have been defended as being normative, not positive.
2. Does SG produce consistent results?	Hershey et al, 1985 Llewellyn Thomas et al 1982	. Achieving equivalence in the SG by varying probabilities gives a different result from varying certainty quantities. . Results depend upon outcome.
3. What is the theoretical basis of the TTO technique?	Torrance, 1986 . Mehrez & Gafni, 1989b	Initially proposed as an approximation to the SG revealing `value' not `utility'. . M&G argue that the TTO identifies points on an indifference curve between quality and quantity.
4. Should only the patients utility be considered?	Loomes & McKenzie, 1989 Carr-Hill, 1989	L&M argue that a wider group, namely those chiefly affected, should be included.
5. Should health costs of a normal life be considered as one of the costs of saving a life?	Weinstein & Stason, 1976	In principle, yes; in practice this may not make a great difference.
6. How should medical risk and uncertainty be included?	Gafni & Torrance, 1984 Mehrez & Gafni, 1989b	. M&G argue that this necessitates the use of the . G&T argue risk is the sum of gambling, quantity and time preference effects. . H's empirical results suggest no pure risk behaviour but context specific behaviour.
7. Is total cost / QALY the appropriate ratio for choice?	Linard, 1990 Birch & Donaldson, 1987	. Numerator should have budgetary costs only if budgets are limited. . Indirect benefits cannot therefore be in the denominator: they cannot be combined with QALYS in the

marginal changes where possible.

Issue	Authors	Comments
8. Is it possible to aggregate utility across individuals?	Torrance & Feeny, 1989 Torrance, 1986 Carr-Hill, 1989	. Generally agreed this must be and is done in practice. . Limited enquiry into the assumptions which permit this and little discussion of the relation between assumptions and underlying value systems.
9. How should QALYS be distributed?	Kuhse & Singer, 1988 Richardson & Hall, 1990 Loomes & McKenzie, 1989 Wright, 1986	QALYS may be valued differently at different ages or subject to some other distributional criteria.
10. Which discount rate should be applied for future values?	Torrance & Feeny, 1989 Richardson et al, 1990 Evans, 1990 Lipscombe, 1989 Carr-Hill, 1989	Dispute between the use of the social opportunity cost and the social rate of time preference.

TABLE 5 THE CHOICE OF THE COST-BENEFIT RATIO

Ratio	Fixed Budget	No Constraint
(C - OB)/Q	INVALID: Does not maximise QALYS	Correct ranking to produce life years most efficiently
(Q + OB - C)/C	Maximise net benefits but unmeasurable	Serves no purpose

Key:
C = Budgetary cost (dollars)
OB = Indirect (other) benefits (dollars)
Q = Number of QALYs
Q* = Value of QALYs