IN LIMINE SAPIENTIÆ

UNIVERSITY
OF YORK

CENTRE FOR HEALTH ECONOMICS

Health Service Efficiency — Appraising the Appraisers

by

TONY CULYER, *University of York*

DISCUSSION PAPER 10

UNIVERSITY OF YORK
CENTRE FOR HEALTH ECONOMICS

HEALTH SERVICE EFFICIENCY - APPRAISING THE APPRAISERS

by

A.J. Culyer - University of York

June 1985

Paper prepared for the meeting of the Verein fur Sozialpolitik in Saarbruecken, September 1985.

The Author

Tony Culyer is a Professor of Economics in the Economics Department, University of York.

Acknowledgements

The paper draws on Blades, Culyer and Walker (1985), which is a comprehensive critique of current practice in economic appraisals of health programmes. Thanks for comments on this paper (and the usual absolution) are due to Ron Akehurst, Chris Blades, Mike Drummond, Jeremy Hurst and Arthur Walker.

Further Copies

Further Copies of this document are available (at price £3.00 to cover the costs of publication, postage and packing) from:

The Secretary,
Centre for Health Economics,
IRISS,
University of York,
Heslington,
York, YO1 5DD.

Please make cheques payable to the University of York. Details of our Discussion Papers can be obtained from the same address, or telephone York (0904) 59861, extension 5752.

The Centre for Health Economics is a Designated Research Centre of the Economic and Social Research Council and the Department of Health and Social Security.

I INTRODUCTION

In 1971 I concluded an article on the welfare economics of health care (Culyer 1971) with a plea for less *a priori* argument and more applied welfare economics in the form of cost-effectiveness and cost-benefit analyses. In the intervening 15 years there has been a refreshing redirection of effort - particularly in Britain and North America, so it is appropriate now to look back and appraise the appraisers. In particular, I want to concentrate on how well we have been practising economic principles as we would want to judge practice ourselves - rather than as others may judge us.

We have all at various times come across requests - sometimes rather petulant ones - for evidence of the efficacy of economic appraisal. It is observed that there have been several large and well-publicised economic appraisals of this or that which seem to have had little impact on public decisions - as evidenced by decisions that seem to have been contrary to the recommendations of the appraisal. Accordingly, it would seem that the scope of economic appraisal has become overblown; the claims made for it by its proponents have been exaggerated; its usefulness is limited; it is possibly of no use at all.

At one level this type of charge is easy to rebut, for the fact that what seemed to be a preferred option in a study turned out not to be such can in no way be held to imply that the study itself was of no use, that its scope was inappropriate or that it was limited in any inherently arbitrary or unhelpful way. After all, the decision may have been wrong. Or it may have been based on other considerations as well.

More fundamentally, the charge is to be answered by the assertion that economic appraisal is not appropriately to be judged in this fashion for it is not a method of, so to speak, *dictating* public choices or of obviating the need for public decision makers to exercise their judgement. Rather, it is a means by which the important issues about which judgement has to be exercised can be identified, relevant aspects quantified, and the appropriate judgemental skills (according as they are value judgements, technical judgements, methodological judgements, factual judgements, and so on) exploited.

This is the case regardless of the position you take about the way in which economic appraisal relates to a social welfare function. To use the distinction made by Sugden and Williams (1978), those adopting the *Paretian* approach (e.g. Mishan, 1971, 1974) seek to identify potential Pareto improvements, and that is all. On this basis it is perfectly clear, since economic efficiency is but one consideration which social decision makers may have in mind, and that the claims made for economic appraisal are relatively modest, that appraisal affords at best a partial structuring and solution of social choice problems, and its conclusions may well be properly overturned with the introduction of other considerations that transcend efficiency.[1]

Those adopting the decision-making approach (e.g. Dasgupta, Marglin and Sen, 1972; Little and Mirrlees, 1974; Williams, 1972) take as authoritativ a judgement about the arguments (and their various weights) in a SWF made by an individual or group who is, or whose members are, legitimately empowered by a social process to make such judgements. On this view, potential Pareto improvements are but one of a range of issues for exploration, as are the range of and weights attached to arguments in the SWF - usually after lengthy and detailed discussions between analysts and research customers. In this approach one's expectations of the authority and social acceptability[2] of an appraisal's conclusions depend very much upon the degree of on-going collaboration between analysts and customers. To the extent that it is small, or not on-going, the greater the probability that trade-offs and other types of judgement will still remain to be tackled after the formal appraisal is completed. To the extent that customers are brought more or less continuously into the analysis itself, the greater the probability that relevant points of controversy (that may otherwise overturn an appraisal's conclusions) will have already been fully aired, with the decision-maker's judgements already incorporated. In reality, the likely outcome is that a fully annotated short list of options will be the outcome of the study, presented in such a fashion that all the relevant components for a final decision will have been laid bare.

My prejudice is that this latter approach is in general the most helpful to decision-makers. Though it is not always feasible (for example, because the study may have no specific social decision-maker in mind), it remains a useful target for analysts to attempt as far as they are able.[3]

In this paper, I shall take a restricted view of the scope and limits of health economics, concentrating on that area which has, in Europe at least, been expanding fastest in the last decade or so. I am going to focus, as implied in all I have said so far, on economic appraisal in health care. I want to argue that - on the decision-making approach - there are, to all intents and purposes, no limits and that the scope is, or can be, as wide as may be appropriate. Limitations of a sort there are, however. They are not limitations of feasibility or practicability; nor are they limitations of quantification or measurement. They are limitations mostly of imagination and of competence on the part of the analysts.

What I propose to do is to review some of the empirical work that has been conducted in recent years (most of it British, for that is what I know best) in order to demonstrate bad and good practice, and how it is only bad practice and failures of imagination that limit the potential scope of economic appraisal - and, of course, its usefulness. The scope for the practice of appraisal is, of course, limited by the demand generated by our health care systems and, though I shall not say much about that here, I shall return to it at the end.

For the purposes of this paper, appraisal includes studies that may use terms in their titles like 'investment appraisal', 'option appraisal', 'cost-effectiveness analysis', 'cost-utility analysis', or 'cost-benefit analysis'. It excludes what in commercial circles is called 'financial appraisal'.

The paper is organised into four principal sections: II deals with treatments of cost, III deals with treatments of benefit, IV deals with design and presentational issues, and section V provides a checklist of factors to bear in mind in designing, conducting and appraising appraisals, and which unashamedly sets no limits to the scope of this branch of health economics.

II COST

Breadth and coverage

The narrowest approach focuses usually on the costs incurred by a particular agency or even department of a decision made therein. Wider would be an extension to the consequences for, say, a hospital or the hospital sector, family doctor services, the health care system as a whole, other public agencies' services or, eventually, the knock-on consequences for the public sector as a whole may be encompassed. Further widening of the scope of cost consequences would bring in the voluntary and private sectors (including households). Often only at this stage do the costs to patients and their families receive any reckoning.

Costs are not necessarily only financial of course. The time of patients and family carers may not be directly paid for by anyone, but it is rarely costless and in principle ought to be considered if efficiency is being sought. Similarly, the value of resources like land and buildings already owned may not be revealed in any cash transactions but these resources too have valuable alternative uses - that is to say they have costs - which should again be considered. Transfer payments made to, say, social security beneficiaries do, by contrast, feature in cash outlays and appear as costs to the agencies in question. But, as is well known, they do not represent the alternative value of any resources used and so are not social costs. Here is another source of difference in scope depending on the perspective.

Clearly, the appropriate breadth of the cost concept employed will depend upon the question that a study is addressing. There is little point in the assiduous tracking down of all the minutiae of the cost consequences of a particular decision, especially if there are some fairly major approximations used in the estimation of the major consequences.

One study that quite properly took a very comprehensive view of the breadth of coverage was that by Hagard, Carter and Milne (1976). This study sought to identify the costs and 'economic' benefits of a programme for the mass screening of pregnant women for early detection and abortion of fetuses with neural tube defects. The costs of the programme were seen as falling primarily on the health service (capital and revenue) but a significant portion of costs also fell on patients (even ignoring the emotional and ethical traumas associated with screening and abortion). Patient costs were about 28 per cent of recurrent health service costs and included lost earnings and travel expenses associated with genetic counselling, amniocentesis, ultrasonography, and extra clinical attendances for second blood samples to be taken.

Another study taking an appropriately wide notion of cost was that of Russell, Devlin, Fell, Glass and Newell (1977). This study compared the costs and effectiveness of day case and inpatient surgical treatment of hernias and haemorrhoids. The costs included embraced hospital resources, district nursing and general practitioner costs, and patients' and their families' expenditure. They also provided some (non monetary) estimates of the costs of patient time.

The scope of costs can make a major difference to the overall results of a study. A spectacular example of this is Culyer and Maynard (1981). This was a cost-effectiveness study of medical and surgical alternatives in the treatment of duodenal ulcers. Inclusion of the cost of surgical fatalities (which does not fall on hospitals) produced a range of cost per case for the surgical option of £1,180 - £16,370, relative to a range for the medical option of £1,020 - £1,240. Excluding these costs, however, brought the cost per surgical case down to the range £950 - £1,370. Thus, the more comprehensive scope suggested quite strongly that the medical option was the cost-effective option whereas the narrower scope produces a far more ambiguous result, including the possibility that the surgical option would be the more cost-effective.

A study that legitimately took a very restrictive notion of cost was that by Coverdale, Gibbs and Nurse (1980). This sought to identify econometrically the NHS financial costs of treatment and hospital 'hotel' services per patient day by specialty and is used (with suitable adjustment for inflation) officially in determining budgets in the NHS and for assessing the cost consequences of cross-boundary flows between English districts and regions. The limited notion of cost employed was quite clearly justified by the purpose for which the results have been used.

Health Notice HN(81)30 of the DHSS in England recommended yet another way of identifying the scope of costs. This was to group together all effects, positive or negative, to which money values had been assigned, as costs with other (net) effects described as benefits. This is an entirely arbitrary distinction, implying an arbitrary scope of cost, having no basis in theory, for which the Department has rightly been taken to task by Akehurst and Buxton (1985). The method proposed is likely to be particularly misleading if it is used in conjunction with benefit-cost ratios but the fundamental objection to it is that it detracts attention from what should be the focus of the costing part of all appraisals - how best to identify the alternative use value of resources - and creates confusion about the differences between opportunity cost, expenditure and 'disadvantages' in place or more precision and care in the application of often elusive concepts.

As we all know, the essential feature of opportunity cost is that it represents the highest alternative use value of resources - it is what is necessarily foregone as the result of a decision to commit resources. Unfortunately, even eminent economists sometimes give the impression that 'cost' can be used loosely to cover 'harm' and the disadvantages of options (e.g. Coase in his 1960 classic) and this can cause great confusion - as well as outright error - in studies by more workaday economists.

Opportunity cost is <u>not</u> the same as the "disadvantages" or "blood, sweat and tears". Consider the homely example in Table 1. It is about

Table 1. Advantages, disadvantages and opportunity cost

	Reading this paper	Taking the dog for a walk	Cooking dinner
Advantages	20 utils	8 utils	15 utils
Disadvantages	18 utils	3 utils	11 utils
Net advantage	2 utils	5 utils	4 utils

6.00 p.m., and three effective choices confront you: read this paper, take the dog for a walk or cook dinner. The pain, sweat etc. of reading this paper is valued, say, at 18 utils (one is imagining that each activity can be rated on a ratio scale) which is far higher than the disutility of walking the dog (3 utils) or preparing dinner (11 utils). But this is not the opportunity of reading this paper. To discover that one needs to know the advantages of the three options. As can be seen, the net utilities from each are 2, 5 and 4 utils. So, if you read the paper, you are actually forgoing a net utility gain of 5 utils (not 4, of course, since if you do not read this paper you will take the dog our rather than cook). This is the highest alternative use value of your time (which is the resource being valued in this case). So, being a utility maximiser, you put the paper down and take the dog out.

Note that here one is dealing explicitly with subjective opportunity cost. Sometimes these costs get revealed in market prices by what people are prepared to pay for things, or need to receive in compensation for parting with them. Sometimes, however, market prices need to be replaced by shadow prices, and sometimes one has no relevant market prices at all, so the shadow pricing starts from scratch.

What the appropriate scope of the cost concept should be depends at root on two factors: the nature of the decision' and the perspective of the study. Costs are not immutable 'facts' lying ripe in the field waiting merely to be garnered, or even selectively winnowed, by diligent clerical officers. Nor are they simply to be lifted from the ledgers. The nature of the decision for which a study is a managerial input will dictate the likely pattern of costs: what expenditures are 'overhead' or 'fixed' with respect to the decision and what will likely vary; what is capital and what revenue; what is a relevant margin and what a relevant average. A good discussion of the appropriate inclusion of overheads in average rather than marginal costs is in the Buxton et al (1985) study of the UK heart transplant programme, where the distinction between a new service development and an extension of an existing programme is carefully drawn.

Likewise the perspective is a no less important determinant. Is the analysis to be undertaken from a public sector point of view, or a society-wide point of view, etc.? There is clearly no "correct" answer to this question but clarity of perspective is important, for what is a cost from one perspective may not be one from another; what is a relevant (and helpful) perspective for one research client may not be so for another. An important phase in every appraisal is therefore the initial one that explores the needs of the research customer (actual or targeted) and then tailors the scope of the study accordingly.

Estimating "indirect" costs

The 'economic' benefits attributed in many evaluative studies often take the form of averted costs. For example, one benefit of prevention is the present value of future costs of health care avoided by either the individual or society in general. Averted costs are likely to figure particularly prominently in evaluations of preventive programmes of care. For example, Dowell (1976) made estimates of the reduction in dental caries that could be expected over time from the fluoridation of water supplies. He found that the present value (at 10 per cent) of the averted costs per person over a 30 year period, of a fluoridation programme beginning in year 1, were about £2.50 per person in 1974 prices. In a different field, Bartlett, Neil-Dwyer, Banham and Cruikshank (1978) estimated the cost reductions to be expected from fewer angiograms, air studies and neurosurgery/neurology, and shorter inpatient stays, as a result of installing a brain scanner.

A difficulty with the 'averted costs' approach to benefits is well-illustrated by Hagard, Carter and Milne (1976). In this study the 'economic' benefits were defined to be the excess individual, educational, social and personal resources normally used by surviving babies with spina bifida cystica that would be avoided by screening and abortion. Various estimates of life expectancy and the distribution of handicap were made.

These data represented the additional costs of a surviving handicapped person above those of providing for a healthy normal person. They included, for example, the difference between the earnings of mothers with handicapped children and mothers with normal children.

A particular difficulty with this approach can be highlighted by considering the following:

> 'Since geniuses are more costly to educate than normal people, an imaginary technology that identified very high IQ in fetuses would enable the termination of such pregnancies and hence avert substantial extra costs'.

and:

> 'The more society cares about the disabled and, hence, the more it is willing to spend on services for them, the greater the averted costs by aborting damaged fetuses. Conversely,

the less it cares and the less it spends, the less the averted costs. Therefore the more it cares the more it terminates and the less it cares the more people with handicap it has to care for'.

What seems to be going wrong here is that the benefits of care, education and the like are completely ignored in the analysis. Taking a highly simplified numerical example, the procedure adopted by Hagard, Carter and Milne is shown in Table 2.

In Table 2 the total costs of care for a handicapped baby are 500 compared with 300 for a healthy baby. The averted costs, assuming the fetus is replaced, are thus 200. This is therefore to be counted as a benefit of the screening/abortion programme. If the earnings differential for a handicapped and normal person is 100, then this is an additional averted cost, bringing the total to 300. The averted costs in the case of the genius, again assuming replacement, are 100 (all educational) ignoring earnings effects. If geniuses earn more than normal healthy people, this benefit must be reduced by the differential, or increased if they earn less (assuming that the social contribution is measured entirely by earnings).

What this analysis entirely fails to capture, however, are the benefits of the health and educational programmes provided. Assume that no programmes are offered whose benefits are less than costs and that in all cases the benefits exceed the costs. A fuller accounting framework then becomes that shown in Table 3. Here it is assumed that the intrinsic benefits of health care and education are the same per individual (measured, say, in terms of 'value' added) - save in the case of geniuses, for whom there is a greater benefit from education (which is also, however, costlier than average), so there is no discrimination against the handicapped relative to the average on this account, nor on grounds of educational cost compared to the healthy genius.

In this case, the differential costs averted by replacing a damaged fetus by a healthy normal person are, as before, 500 - 300 = 200. There is likewise a gain to be hand, on the cost side considered in isolation, from a programme to screen and abort geniuses (of 100 attributable to educational costs) though this becomes zero when the assumed additional earnings of the healthy genius are taken into account. Note that no advantage or disadvantage is held in this analysis to attach to abortion *per se* (replacement of a healthy normal fetus with another is a matter of

Table 2 Averted costs if fetus is aborted

	Normal healthy baby costs	Handicapped baby costs	Genius healthy baby costs
Health care	100	200	100
Education	200	300	300
Total	300	500	400

Table 3 Net benefits if fetus is aborted

	Healthy normal			Handicapped			Healthy genius		
	Costs	Benefits	Net	Costs	Benefits	Net	Costs	Benefits	Net
Health care	100	250	150	200	250	50	200	250	150
Education	200	350	150	300	350	50	300	400	100
Earnings		200	200		100	100		300	300
Total	300	800	500	500	700	200	400	950	550

indifference, as is replacement of a healthy normal person by a healthy genius, or vice versa).

However, once the benefits are considered, the picture changes. Replacement of a handicapped by a normal healthy fetus involves a loss of 200 and a gain of 500 - a net gain of 300 (exactly the same as before). But the anomaly regarding the genius has now vanished. Replacement now entails a loss of 550 for a gain of 500 - a net loss of 50.

The correct technical procedure is thus to use genuine opportunity costs. Hagard, Carter and Milne obtained the correct answer but only under the implicit assumption that there was no differential in the benefits of programmes. In this kind of problem, relevant opportunity costs are not the opportunity costs of the resources used in health and educational programmes; they are the net payoffs to the relevant alternative forgone. The opportunity cost of an unreplaced damaged fetus is 300 because this is the difference between the prospective net benefit of a handicapped and a normal healthy person, not because it is the difference between the value of resources each uses up.

An implication of the fuller analysis is that (for equal benefit) if educational and health care costs were the same for handicapped and healthy babies, there would be no opportunity costs saved by screening and abortion (as also implied by the Hagard, Carter and Milne methodology), and that if resource costs were the same, but health and educational benefits deemed higher for the healthy babies, then opportunity costs saved by screening and abortion would be positive (as is not implied by the Hagard et al. methodology).

The appraisal of day-case surgery by Russell et al. (1977) affords a good example of an attempt to identify the true opportunity costs within an organisation. If the increased use of day-case surgery would enable the closure of a small ward, each inpatient stay averted would save £24 on their estimates. Alternatively if it averted the necessity for building a small ward extension, the equivalent cost saving was estimated to be £33. These are the bounds, so they argued, of the costs (or part of them) from not moving into more intensive use of day-surgery - not involving any intensification of hospital activity but the useful redeployment of hospital resources.

Marginal costs

The stool guiac is a relatively cheap way of detecting occult blood in stools. In 1976 the American Cancer Society was recommending that six sequential stool guiacs be performed on all people over the age of 40 in order to maximise the chances of identifying asymptomatic cancer of the colon. Six was recommended in order to minimise the number of false negatives.

In the U.S.A. around 1968, the average cost per cancer diagnosed if one test were done was $1,175, and about 6 cancers would be missed out of 10,000 screened patients. If 6 tests were done the average cost was only about twice this ($2,451) and the number of cancers missed fell to a tiny fraction. The average was, however, a very misleading number to use.

Neuhauser and Lewicki (1976) computed the marginal costs per true cancer detected (not treated) with dramatic results, largely due to the low marginal yield of true positives from additional tests. The remarkable result was that the marginal cost per cancer found of the sixth stool guiac was more than $47 million. This is a spectacular example of how averages can mislead in what is essentially a marginal choice (how many more tests to perform).

Relevant marginal costs are often exceedingly elusive and even sophisticated practitioners have often failed to spot them. The work of Hagard, Carter and Milne (1976) contains two instructive examples of such a failure (though the consequences for the conclusions of the study cannot be assayed). One was in their assumption that 90 per cent of women would attend ante-natal clinic between the 15th and 20th week of gestation. This degree of attendance was estimated to cost £20,000 in publicity costs, and would identify an expected 92 pregnancies affected by 'open' spina bifida in a population of 43,000 pregnant women. The incidence of spina bifida in such a population was estimated, however, to be 120. The marginal cost of procuring additional attenders through publicity is likely to increase quite sharply. Suppose, for example, that doubling the publicity effort would increase attendance from 90 per cent to 95 per cent. This would add an expected 5 pregnancies affected by spina bifida: a marginal cost of £4,000 per affected pregnancy, or £5,000 per screened and aborted fetus. This marginal cost is not relevant in a decision to establish a programme (or not) with a 90 per cent take-up, but is highly relevant in a decision to go for a programme with 80%, or 90%, or 95%, etc. take-ups.

A further relevant margin was in the rate of use of ultrasonography - particularly to establish gestational age (96 per cent of cases referred for ultrasonography). This procedure is estimated to add only about seven terminatable cases. Even ignoring the equipment costs, the present value of recurrent costs of ultrasound (at 10%) was £161,500, or £23,100 per additional termination.

These examples illustrate the importance of identifying the question to be addressed (e.g. to establish, extend or reduce) and point up the important possibility that the cost of finding and treating an additional case may be sufficiently high for the question of the scale of programme (or of some part of it) to be important research questions.

A similar marginal consideration arises in a paper by Rich, Glass and Selkon (1976) in which they investigated the cost-effectiveness of two methods of screening schoolgirls for asymptomatic bacteriuria (a supervised collection of midstream specimens of urine - MSU - and unsupervised, self-administered dipslides). With the MSU procedure, an 85 per cent success rate would have been achieved in a programme in which each school class was tested and second specimens only for those having positive tests on the first round. A 96 per cent response was generated, however, by the actual procedures used in the study, which involved extensive 'chasing'. The authors estimated that 850 children out of 1,000 could be screened at an average (MSU) cost of 55p per child. The next 110 would cost £2.68 per child. Per active case of significant bacteriuria, the costs were £22,70 for the first 85% and £110.60 for the next 11%.

Discounting

Discounting is likely to make a difference in the relative attractiveness of alternatives when the time profile of costs (or benefits, or net benefits) is substantially different as between the options.

Failure to use discounting, though much frowned upon by DHSS in Britain, is often less important in practice than is emphasised by writers on principle. Nonetheless, because conclusions can be sensitive to discounting, it is good practice to use either present values of cost, or annual equivalent charges[4] that correspond to capital costs.

An example of failure to discount comes from the DHSS's own journal Health Trends (Thomson, 1977). This was a study of the cost-effectiveness of a brain scanner. The purchase price of the machine was £140,000 in 1974/75. It was assumed to have a useful life of 10 years. The annualised capital cost was taken to be £14,000 by straight line depreciation. Using a public sector test discount rate of 5 per cent, the correct annual equivalent charge should have been £18,130, however (or £19,936 at a 7 per cent discount rate). The annual cost flows equivalent to capital expenditures are also sensitive to changes in the expected life of capital equipment. In fast developing imaging technologies 10 years of probably biases costs downwards. For example, had the expected life of the scanner been 5 instead of 10 years, the annual equivalent charge would have been £32,340 at 5 per cent, or £34,159 at 7 per cent. Replacing the £14,000 cost attributed in the study with that latter figure would have raised the overall annual cost of a programme of 2,000 scans per year by 68 per cent.

Another case when failure to discount produced highly misleading information is in the paper on the value of preventing deaths from nephritis and nephrosis by Longmore and Rehahn (1975). Over a ten year period their estimate of the cost of renal deaths was £120 million. Discounted at 5%, this cost falls by 26 per cent to £89 million.

III BENEFIT

Economic appraisals can be usefully subdivided into cost-effectiveness analyses, cost-utility analyses, and cost-benefit analyses. While each of these three in principle takes a comprehensive view of opportunity costs, as discussed in section II, they vary in the detail in which benefit questions are addressed. Cost-effectiveness analysis (CEA) seeks to identify either the least-cost method of achieving a given objective which is usually treated as single-dimensional (e.g. per case, per life-year gained) or the maximum output (usually single dimensional) attainable for a given cost. Such studies may, as suggested above, emphasise unit costs (or output per cost unit), or the minimisation of programme costs (or maximisation of programme output for a given cost outlay). It is thus a characteristic of CEA that the benefit side is restricted to a fairly elementary and restrictive notion of output.

The next stage in sophistication arises when there are reasons for believing that the restricted notion of "output" in a CEA seriously fails to capture some important dimensions of benefit that ought to be taken into account. This involves cost-utility analysis (CUA) in which the approach is essentially the same as in CEA save that a more comprehensive set of outcomes is considered, raising additional questions as to how they are to be measured, how far superior performance in one outcome dimension may compensate for less good performance in another in a comparison between options, and other related issues. I use the shorthand "output" and "outcome" to describe the dimensions in terms of which effectiveness is measured in CEA and CUA.

At best, CEA and CUA are able to offer guidance as to the least cost way of achieving a stated goal or set of goals. Questions that entail asking whether an activity is worth continuing, extending, or commencing, require cost-benefit analysis (CBA) since they all involve some assessment of the worthwhileness of the benefits relative to the costs of any activity. This terrain, more than any other, illustrates the limitations of analysts' imaginations in not attempting to measure what is not routinely available. It also, however, illustrates how easy it is to be so bemused by quantification that outright error crops up.

Output

The cost study by Coverdale, Gibbs and Nurse (1980) is a good example of the relatively simple approach to output (cost per inpatient day) that may be entirely appropriate in its context. In their study of duodenal ulcer, Culyer and Maynard (1981) used cost per case. Here the differential fatality rate for surgical and medical procedures was treated as a net cost of the surgical procedure rather than a benefit of the medical procedure which raises questions about how the quality difference between the two procedures is best measured and valued.[5]

The study by Lowson, Drummond and Bishop (1981) of oxygen supplies for chronic bronchitics, is a good example of what seems to have been an ideal case for using CEA rather than CUA or CBA. Here the outcomes (or benefits) of the alternative ways of providing oxygen to chronic bronchitics were judged to be identical for all alternatives. The only question, given a target number of patients to be treated, was which is the least cost method. The appropriate output dimension was cost per patient.[6]

Outcomes

Sophisticated application of outcome measures as a species of utility measurement (hence the term cost-utility analysis) has proceeded further and faster in North American studies than in British. Four recent

European examples are by Wright, Cairns and Snell (1981), Fordyce, Mooney and Russell (1981), Leu (1984) and Williams (1985). (For a review of practice see Culyer, 1983.)

Wright and his colleagues compared the cost-effectiveness of alternative modes of care for the elderly (hospital, residential homes and domiciliary services) for varying degrees of dependency of elderly clients. Their measure of dependency was a multi-dimensional Guttmann scale (Guttman, 1944) with seven dimensions (ability in bathing, ambulation, dressing, getting out of bed, sitting and standing, washing and feeding). This produces an <u>ordinal</u> overall classification of dependence based on the cumulative nature of dependencies of this sort. On this basis, a seven dimensional scale produces eight overall levels of dependency (for the majority of patients whose dependence is cumulative in this fashion).

Their results indicated that there appear to be cost advantages, for the lower dependency groups, from community care.

Fordyce, Mooney and Russell (1981) used a similar dependency measure to analyse the balance of care for the elderly and related it to the marginal costs of care. Those patients were identified, in terms of their dependency, who were on the margin of transfer from one mode of care to another. Given the incremental costs of changing the mode of a person's care (which depended in part on the numbers being transferred from one mode to another), it was then possible to identify the appropriate stock of caring modes, given the population's characteristics, that would minimise overall costs.

There really is no intellectual limit on the use of quantification of this sort to extend the informational base of decision making. The limits are those of <u>imagination</u> and <u>willingness to collaborate</u> with specialists in psychometric techniques, and appropriate professional sources for assessing and monitoring outcomes.

The measures of dependence in studies just described were ordinal, in which only the ranking of dependence was indicated. A much more ambitious type of outcome measurement is the ratio scale - the strongest form of cardinal measurement. Kind, Rosser and Williams (1982) developed a scale that combined disability and distress (the former having 8 degrees of severity and the latter 4) enabling patients to be classified into 32 possible states and changes in their health status to be monitored over time. Using a sample of respondents (including doctors) cardinal numbers were assigned to these states on a scale with two fixed points 1, 0, with the former corresponding to good health and the latter to death. The resulting (averaged) scores are shown in Table 4. This yielded the not surprising result that some states were regarded as worse than death (negative values in the cells of the table).

Williams (1985) has gone on to use these scales in order to compute QALYs (quality adjusted life years) as a more sensitive measure of the outcome of various procedures that prolong life but with varying degrees of quality (hence transforming what is essentially an output measure into an outcome measure) with results reported later in Table 8. The procedure here is to weight each expected life year gained by the quality index appropriate for that year and discounting to preserve symmetry with the treatment of costs.

Cardinal scoring for outcomes (or degree of importance of different objectives) has been advocated (and practised) by Akehurst and Buxton (1985) as a useful method of identifying important trade-offs in option appraisals of major capital investments. An actual study using this method is described in Akehurst and Holtermann (1985) for an English District (which is anonymised). In this study (of mental illness service development options), the objectives to which numbers were assigned included: ease of contact with therapists, quality of action, staff morale, travel time, stigma, ease of liaison with other specialties and flexibility. This illustrates the great variety of outcomes, states or characteristics of options to which numbers <u>can</u>, with imagination, be assigned. The art lies, of course, in their intelligent assigning so that the process <u>highlights</u> important choice problems rather than obfuscating them - the numbers must be servants, not masters. It is, of course, in work like this that the difference between the decision-making and the Paretian approaches are most marked.

All these applications require a clear understanding of the meaning of quantitative measures - and the interpretations that may and may not be put upon numbers - on the part both of researchers and research customers. The principal distinction to be made is between ordinal and cardinal measurement and, within cardinal measurement, between interval and ratio scales. Table 5 contains 4 health states. These may be simply coded A, B, C, D, indicating the state in shorthand and, if significance is to be attached to the alphabetical order, also the rank order of the states (descending). The next three columns give numerical scores corresponding

Table 4. Health status indicator (Kind, Rosser and Williams)

Disability rating	Distress rating			
	1 None	2 Mild	3 Moderate	4 Severe
1. None	1.000	0.995	0.990	0.967
2. Slight social	0.990	0.986	0.973	0.932
3. Severe social, difficulty with heavy tasks	0.980	0.972	0.956	0.912
4. Choice much restricted, light tasks only	0.964	0.956	0.942	0.870
5. Unable to take paid employment, full time education, mostly confined to home	0.946	0.935	0.900	0.700
6. Confined to chair or wheelchair	0.895	0.845	0.680	0.00
7. Confined to bed	0.677	0.564	0.000	-1.486
8. Unconscious	-1.028	n.a.	n.a.	n.a.

Table 5. The meaning of numerical measurement

STATE	CODE	NUMBERS						
		Ordinal			Cardinal			
					Interval (Linear)		Ratio	
		(a)	(b)	(c)	(a)	(b)	(a)	(b)
Healthy	A	4	217	2 million	100	212	5.0	8.05
Angina case: some difficulty in bending and stretching	B	3	15	0	60	140	3.0	4.83
Paralysed right side, can cook and wash but needs some help (e.g. shopping)	C	2	10	-1	50	122	2.0	3.22
Housebound, wheel-chair, needs feeding, etc.	D	1	0.5	-16	0	32	1.0	1.61

to the order. Any one is equally as good as any other to indicate order. 4 does not mean "twice as good as" 2 (evidently 217 is not twice 10).

The first columns in the cardinal set have been assigned to the states according to a different rule. One measurement technique that yields this form of scaling is the standard gamble (see Culyer, 1978, for an introduction and Torrance, 1970, for several North American applications). The interesting characteristic of the two sets of numbers in these columns is that if the intervals between numbers are rising or falling according to one set, they also rise or fall according to the other (hence the term "interval" scale). They are related by linear equations of the form (b) = A ± B(a) where A and B are non-zero constants and the equation for the scales in the table is (b) = 32 + 1.8(a).

The numbers selected here are, in fact, those for ^{o}Celsius and ^{o}Fahrenheit, so this form of measurement is akin to that used for temperature measurement. Again the ratios mean nothing. If 100^{o}C is "hot" and 50^{o}C is "cool", one cannot say that "hot" is twice as hot as "cool" (on Fahrenheit, evidently 212 is not double 122).

The final pair of scales illustrates ratio scaling. Here the link between the scales is proportionate: (b) = A(c) and one can speak of "twice as much", "half as bad", etc. This is the strongest form of measurement and it is all too frequently supposed to be the only form. It is of crucial importance that the nature of the measurement exercise be clear to those who are assigning numbers to the entities in question. If in their assigning they implicitly only rank entities, then it is quite illegitimate to interpret the resulting scales as cardinal. Here great care must always be taken and even the sophisticated can be led astray. Thus Wolfson (1974), who was a pioneer of the standard gamble approach in Canada, computed severity weights for 224 diagnostic categories and used these with incidence data based on hospital discharges and insurance claims data to compute a health status index for each of the 54 counties in Ontario. A conclusion, among others, was that Frontenac county was three times sicker than Peel county. The linear form of measurement of ill-health, however, did not permit so strong a comparison for Wolfson's index was derived by using the standard gamble on subjects (doctors) who assigned severity ratings to diagnoses. As is well-known, the standard gamble (Neumann and Morgenstern, 1953) yields only an interval scale. At best, therefore, Wolfson could have used his numbers as a kind of "temperature chart" of the health state of patients, counties, etc.

Interval scales are increasingly used in multiple regression analyses where they appear variously as dependent or independent variables (e.g. Leu, 1984). Here great care must be taken in interpreting the resultant elasticities, whose signs are uniquely determined, whose absolute values are, however, not uniquely determined, and whose ratios may or may not be independent of the transform used.[7]

If economic sophisticates can be led astray by numbers, how much more less statistical clinicians. Grogono and Woodgate (1971) constructed a scale in 10 dimensions, on each of which a patient could be scored 0, ½ or 1. This was unambiguously treated in ratio fashion and was proposed for use as a weighting system to be applied to "health-years". The question of the relative weights of the 10 dimensions was not even raised, however, nor any possibility that the marginal rate of substitution at various levels of intensity may vary.

Valuing outputs and outcomes in money terms

A common form of monetary assessment of benefit relates to the speed with which patients are returned to productive work. Sometimes this element of benefit is presented in a non-monetary form (e.g. in Russell, Devlin, Fell, Glass and Newell, 1977). In other examples estimates are made of the value of output lost or gained (e.g. Beresford, Chant, Jones, Piachaud and Weddell, 1978).

Buxton and West (1975) used the present value of gross wages as a measure of this benefit (which is clearly dependent on factors like the speed of rehabilitation, probability of unemployment, etc.) and found that the present value of additional output for an initial cohort of 1,000 patients treated by hospital dialysis was £3.6 million compared with £6.3 million for home dialysis.

Earlier studies often deducted the present value lifetime consumption from the present value of earnings (e.g. Weisbrod, 1971). However, this procedure is rarely found in the literature now - it clearly takes an inappropriately instrumental view of the individual as a producer and ignores the individual's own consumption benefits (Dowie, 1970).

Longmore and Rehahn (1975) used average GNP as their estimate of the contribution of workers to output in order to calculate the potential benefits from reduced death from nephritis and nephrosis. They also failed to discount. The procedure was illegitimate for several reasons: GNP includes other sources of income than earnings from work (interest, dividends, rents, profits and net property income from abroad), it is gross of depreciation and it is variously measured at market prices and factor cost (which, is not clear in this study). The use of an average figure assumes (implausibly) that the marginal contribution of each worker (let alone possibly disabled ones) is the same as the average. Failure to discount leads to gross exaggeration of the benefits no matter how calculated.

A general difficulty with this "production" approach to benefit is the potential it affords for discrimination between low and high earners, men and women, children/retired and those who work, etc. These distributional issues should invariably be identified and discussed in studies making use of this approach.

The value of output gained, or "human capital" approach (in which people are treated rather as though they were carthorses) is not however, the only way in which attempts have been made to value lives saved. Mooney (1984) has reviewed some of the approaches, which are indicated in Table 5. The first four methods <u>infer</u> variously minimum or maximum values based on decisions taken or not taken. The rationale here is that if a decision was taken <u>not</u> to institute an arrangement that was predicted to save life, then the lives expected to have been saved cannot have been worth the expenditure. Conversely, if a decision is taken to incur a certain cost, then the expected lives saved must have been worth <u>at least</u> that much.

This method typically produces a bewilderingly large range of values (see Table 6) which are in any case evidently contaminated by other factors (for example, political pressures, uncertainty about the number of lives to be saved, other benefits not related to life years gained).[8]

Jones-Lee (1976) has developed a method of computing the value placed <u>by those at risk</u> (and their relatives) on reductions in the risk of death.[9] These typically yield values substantially in excess of those derived by the human capital approach though not usually so large as that shown in Table 6.

Table 6 . Alternative estimates of the value of life

	SOURCE	COST PER LIFE	COMMENTS
INFERRED VALUES FROM OTHER CONSIDERATIONS	Government decision not to introduce child-proof containers for drugs	£1,000	In 1971 Government refused to introduce child-proof containers for drugs on grounds of expense; net cost per child's life estimated at £1,000.
	Motorway driving	£94,000	Given optimum motorway speed, the price of petrol, and the value of time then the implied value of life was estimated to be £94,000.
	Proposals for improved safety of trawlers	£1 million	Estimate that cost per death averted would be about £1m.
	Change in building regulations following collapse of Ronan Point high-rise flats	£20 million	From the report of enquiry following the collapse of Ronan Point, a high-rise block of flats in London, the cost of raising safety standards and the resultant fall in the risk of such occurrences was estimated. From these figures the cost per life saved can be inferred to be in excess of £20m.
DIRECT APPROACH	Quesionnaire on valuation of mortality risk reduction	£3 million	Jones-Lee (1976) used a questionnaire to elicit from individuals how much they would be prepared to pay to reduce their risk of death.
HUMAN CAPITAL APPROACH	Department of Transport	£151,000	Based on the present value of life-time earnings.

The appropriate approach to adopt in CBA remains an unsettled question. There is a consensus that to ignore this benefit is unjustifiable (effectively taking the value of life to be zero), as there is that to treat life as infinitely valuable is no less absurd (since it would justify infinite expenditures on life-saving resources and imply that society was infinitely risk averse to life-threatening circumstances). It is also quite clear that it is false to claim (like Muir-Gray, 1979) that monetary values cannot be placed on human life, and no less unsatisfactory simply to wave one's hands helplessly in confronting the issue (Reynell and Reynell, 1972). Since many decisions inevitably involve the comparison of programmes that save lives with those that do not, or of programmes that save different numbers of lives, there is no ducking the issue of how such benefits are to be treated - and I would judge that it is always better to be explicit about it (so as to enable informed dissent and discussion of the sources of disagreement) and also to be eclectic so a that one can see the differences (if any) that alternative approaches make to the conclusions, explore the procedural validity of the alternatives in a given context, and their potential for discrimination (and its acceptability or otherwise) between different kinds of lives.

IV DESIGN AND PRESENTATION

There has, over the years, been a marked improvement in the way that published results are presented though there are still some major deficiencies that are commonly met. The essential requirement is, of course, that the problem investigated by the study should be stated clearly. This obviously has important implications for the nature of all subsequent reporting of the research that has been undertaken as it will define the alternatives considered and the scope of costs and benefits considered as well as identify the target readership.

Alternatives

A major weakness of many studies consists in their failure to consider alternative options. Much of the clinical, and some of the epidemiological, literature, is of limited usefulness to economists because it fails to consider alternative ways of accomplishing objectives - and often leaves objectives only implicit. Implicit in much of this literature - and some of the appraisal literature too - is that the procedure or programme being evaluated is being compared with a "do nothing" option. While this is often an important option to consider - and should then be fully appraised in its own right - it is also often not a relevant option (the decision, for example, is often not whether to treat a patient group, but how to do so). A defect of Hagard, Carter and Milne (1976), for example, was that it implicitly compared a screening programme for neural defects with "do nothing". Another important alternative to that under active investigation is "current practice", which should normally feature as one of the options to be compared, if only because the status quo is a kind of benchmark.

In England and Wales the Department of Health requires major capital proposals to be considered by local Health Authorities in terms of explicitly stated alternative options. In this type of appraisal (and in many others too) it is usually worth getting right down to fundamentals by conducting a brainstorming session among the relevant parties in order to liberate the imagination and elicit the objectives that are being sought - usually defined in terms of effects on patients (outcomes) rather than resources used (inputs). This will lead to the identification of a - probably large - set of options, having different outcome characteristics in terms of who benefits, the nature of the benefits, their expected duration, and so on. The process often leads to the discovery of at least one new, previously unconsidered, option.

There is no official guidance as to how a 'short list' of options is to be derived from this 'long list' but some helpful guidelines have been given by Akehurst and Buxton (1985) which are indicated in Table 7.

In general, then, the first presentational requirement is for the alternatives to be clearly stated and justified in the context of the objectives of the service or project in question. This is always important, but particularly so for policy purposes. Lowson, Drummond and Bishop (1981) was a good example of a variety of alternatives being compared (5 in this case). Thomson (1977) is a good example of a study that failed to make the alternatives clear: implicitly the brain scanner was being compared with the status quo but the comparison was neither systematic nor complete.

The study of alternatives is often hampered by inadequate epidemiological information and by poor research design. Some studies begin life as efficacy studies (does the treatment have the effects claimed?) and gradually transform themselves into effectiveness studies (is the treatment likely to be effective in its impact on natural history as it is

Table 7. Criteria for selecting options from the 'long list'

1. Include any option having powerful political support from any quarter (no matter how implausible it may appear). The case for this is that, if it really is a poor option relative to others, this will need to be documented carefully.

2. Exclude any that are ruled out by any truly binding constraints.

3. Exclude any that are completely dominated by others on an initial rough appraisal of costs and benefits.

4. Options that are similar can often be represented by just one option of its kind. If the representative option turns out to be a good one it can be subsequently refined into some of its variations for more detailed appraisal.

likely to be practised?) and thence into cost-effectiveness studies (is it less costly per unit of outcome than alternatives?) during which process it becomes progressively more difficult to mount a satisfactory design - the first stage will often have looked at an unrepresentative situation and used service-irrelevant 'controls' or alternatives; the second stage may do its best to rectify these omissions but will fail to collect relevant cost and patient outcome data that are essential for the economic appraisal stage.[6]

Once again, clarity about the objectives of the study, as well as the objectives of the procedures, helps to avoid these problems.

Sensitivity and robustness

It is all too easy to give a false impression of exactitude in quantitative studies. Where studies are based on sampled data, confidence intervals should always be presented so that a judgement can be made about the statistical significance of any quantitative results (the variance can often be as important a statistic as the mean for policy purposes). This practice is well-nigh universal, however. Less common are tests of sensitivity and robustness.

Sensitivity refers to the extent to which the ranking of alternatives depends upon the assumptions made. A good study would seek to highlight those assumptions, or ranges of possible error, to which the order is particularly sensitive. Culyer and Maynard (1981) has already been cited as an example of a study in which the ranking was sensitive to the inclusion or exclusion of some cost categories. Buxton and West (1975) subjected their study of haemodialysis to a comprehensive set of sensitivity tests related to long-term survival of patients, about which there was considerable uncertainty, using a "best estimate", a "low survival" estimate and a "high survival" estimate; discount rates (6%, 10% and 14%); cost estimates ("best estimate", "best" doubled, and "best" halved); and assumptions about rehabilitation rates ("low" "best estimate" and "high"). Although these variations were substantial, they found that, nonetheless, hospital dialysis was consistently more costly than home dialysis, so this exemplary and comprehensive sensitivity analysis confirmed and reinforced confidence in the superiority of the home dialysis programme relative to hospital. Sensitivity analysis is especially important when the specificity and sensitivity of screening and diagnostic tests are uncertain and where patient compliance is also largely unknown or variable. It is valuable when there is methodological uncertainty whether or not a particular cost category should be included (for example, Buxton et al. 1985 investigated the effects of including or excluding pupil/student nurses in their study of the UK heart transplant programme). It also helps to highlight areas where further research may be important (e.g. Roberts et al., 1983). It can also be used to highlight the effects of different institutional settings in which services are provided (Buxton et al. 1985).

Another trick that can be usefully employed as a form of sensitivity analysis is to introduce systematic bias into cost and benefit estimation. This enables the analysts to make a fortiori claims for the results. For example, it the costs of option A have been systematically biassed upwards and the benefits downwards, relative to option B, then if A turns out to be superior to B, one's confidence that it is indeed so is reinforced. Conversely, if costs are systematically biassed downwards and benefits upwards in A relative to B, then if A is not superior, one is again reinforced in one's belief that this is indeed so. Of course, in the former cases, had B been found superior to A one should not infer that this is most likely actually to be so; nor in the latter case, had A been found superior to B, should one infer that A really was superior. In such cases there is no alternative to a full sensitivity analysis along the lines of Buxton and West (1975). A study that used the method of systematic bias as well as ranges based on "low" and "high" cost estimates was Culyer and Maynard (1981).

Robustness refers to the extent to which a preferred option is capable of subsequent adjustment as the world changes. There is always the danger that, though a particular capital programme or service development may appear "best" at one date, subsequent changes in the "external" world occur that may make it seem less preferred in that an alternative earlier option may have been more capable of adjustment to the changed circumstances.

Tests for robustness call for a high degree of imagination relating to what might happen. An appraisal of, say, hospital heating systems may at one time indicate a particular dependence on an energy source. An inflexible capital development based on such an appraisal is ill-equipped, however, to coping with, say, a trebling in the price of that energy source, which may subsequently require the entire scrapping of the capital recently built, and its replacement with something else (preferably more flexible).

In high technology medicine, a potent source of error in decision-making lies in the rapid obsolescence of expensive equipment, suggesting that premises, etc. should not be excessively "tailor made", and estimated lives of plant and equipment should be subject to explicit sensitivity analysis.

Robustness does not feature in many economic appraisals at present. It is one of the most challenging aspects of appraisal, it makes heavy imaginative demands, and should feature more prominently than it does.

Decision indices

It is usual for the results of appraisal to be presented as cost per unit of outcome, benefit-cost ratios, and the like. These are often referred to as "decision indices" though it should be emphasised that such appraisals are invariably best seen as guides for decision-makers rather than as substitutes for further thinking. Such indices, while sometimes useful, often hide much important information.

Table 8 contains lists of procedures ranked according to their cost per QALY. The principal use of rankings such as these is that they are suggestive of broad priorities for service developments. They mask, of course, many other important dimensions that may be important for decision-makers. There may, for example, be controversy about the weights to be attached to various dimensions of quality (which should therefore be subject to sensitivity analysis and - in the fulness of time - further research). Even if, however, the weights and scalings are deemed acceptable and comprehensive, there are other important issues.

Of these, that most commonly ignored is the distribution of costs and benefits. Thus, in Table 8, the value judgement is built in that a QALY is of equal social value to whomsoever it accrues (poor or rich, young or old, male or female, etc.). There may be a case for giving differential weights to different groups).

But even if the weights (unitary or otherwise) as between groups are deemed acceptable, there is a further value judgement buried in these indices: they assume that an additional QALY is of equal social value no matter how many QALYs an individual already has. Thus a programme that gives 100 QALYs to 100 people counts the same as a programme costing the same that gives 100 QALYs to two. These distributional features, where they are different as between options, should be fully explored.

Another important category of distributional consideration is often social class, either because the costs and benefits may differ systematically

Table 8. Cost per QALY indices

Torrance and Zipursky	Cost per QALY (1983 U.S. dollars)	Williams	Cost per QALY (1983-84 £)
PKU screening	0	Pacemakers	670
Postpartum anti-D	0	Hip replacement	750
Antepartum anti-D	1,220	Valve replacement	950
Coronary artery bypass surgery with LMD	4,220	Coronary artery bypass surgery with LMD	1,040
Neonatal intensive care	4,500	Coronary artery bypass surgery with 3VD	1,270
T4 (thyroid) screening	6,300	Coronary artery bypass surgery with 2VD	2,280
Treatment of severe hypertension	9,400	Kidney transplantation (cadaver)	3,000
Treatment of mild hypertension	19,100	PTCA with 1VD	3,400
Estrogen therapy	27,000		
School TB testing	43,700		
CAPD	47,100		
Hospital dialysis	54,000		

Sources: Torrance and Zipursky (1984), Williams (1985).

or because social class is associated with compliance and other aspects of patient behaviour (see e.g. Le Grand, 1982).

The study by Rich, Glass and Selkon (1976) was exemplary in its recognition of a social class gradient, where the percentage of girls not returning an unspoilt dipslide was steeply and significantly related to social class. This implies that there is a trade-off between the relatively low cost dipslide method and the high cost (but more comprehensive) MSU method of obtaining urine samples - a trade-off that the study highlighted and that would undoubtedly be an important element in the policy choice between the two screening methods.

Such distributional analyses are, however, all two few.

A commonly presented decision index is the <u>benefit-cost ratio</u>. For example, Buxton and West (1975) computed cost-benefit ratios (the reciprocal of the B/C ratio) for home and hospital dialysis in their sensitivity analysis, coming up with ratios in the range 3.2 to 12.7 for the hospital procedure and 1.6 to 6.3 for the home procedure. Similarly, Hagard, Carter and Milne (1976) calculated B/C ratios for the screening programme at a variety of discount rates (5%, 10% and 15%) yielding ratios of 3.06, 1.86 and 1.39.

It is worth noting that the notion of opportunity cost embodied in C is invariably the opportunity cost of the resources in the most valued use <u>outside</u> the range of options considered in the appraisal (usually outside the health services). While it is possible to treat the net benefits (B-C) of one option as the opportunity cost of adopting another, it is not customary to treat opportunity costs in this way - that is an internal to the decision. Rather, C is taken as representing the opportunity costs of resources in uses that are external to the options under consideration. This may seem, on the face of it, inconsistent with the example of subjective opportunity costs given earlier. However, the point of that example was to demonstrate that cost has the same fundamental grounding as the notion of value in economics. C will always reflect such judgements since they are the basics upon which individuals reckon the compensation they need for parting with any resource they may own. It would be confusing to describe both C as an opportunity cost and the B-C of a rejected option as an opportunity cost, and common usage reserves the label for C only: a usage there seems little point in changing.

The interpretation one is invited to put on these B/C ratios is as follows:

1) for mutually exclusive options, adopt the option for which B/C is highest
2) for mutually non-exclusive options, adopt all for which B/C > 1
3) given a budget, adopt the feasible options for which B/C is highest until you have exhausted your budget ("feasible" options exclude rejected mutually exclusive options)
4) an option for which B/C is higher should be adopted rather than another having a lower ratio.

These "rules" may seem obvious and acceptable (granted the acceptability of the contents of B and C, proper discounting, allowance for uncertainty, etc.) but in fact they are quite dangerous and, in general, the use of such ratios ought to be completely avoided. (They were condemmed quite early in the literature of cost-benefit analysis (see McKean 1958, Hirschleifer _et al_. 1960)).

The first two rules are acceptable provided that important elements in benefit have not been excluded. Since benefit estimation is notoriously hard, and some aspects are invariably omitted or unquantitied, the first rule must be used with great caution and certainly never _slavishly_ followed.

The danger can be illustrated from another territory. It has become fashionable in some circles to use the low social rates of return to higher education as a ground for contracting this sector. Yet "social" rates of return notoriously exclude a wide range of non-financial personal and public benefits of education (which may be significant at this margin). So the correct use of these rates of return - provided that costs are not under-estimated - is as a _minimum_ estimate. If this exceeds alternate rates available elsewhere it can provide a case for _expansion_; but if it is less it cannot provide a case for contraction. For this one requires an estimate of the _maximum_ likely social return to education.

The dangers inherent in the last two rules are more subtle - and hence easily missed. The third rule implies that it may be worth adopting options for which B/C < 1. Even if the benefits are _fully_ captured in B, however, this rule encourages a narrow-mindedness which it is one of the basic purposes of economic appraisal to avoid. The available budget should ideally be set so as to enable the adoption of good options rather than the adoption of as many options as possible. If a budget really does enable the adoption of some options for which B/C < 1, the logic is clear: the budget is too large and should be redeployed to other agencies whose marginal options show B/C > 1. It is in this context that even rough and ready information of the sort developed by Torrance and Zipursky (1984) and Williams (1985) (see Table 8) can help higher tier authorities to make appropriate judgements in budgetary allocations.

The fourth rule is fallacious because it ignores the net benefit of options. Consider two mutually exclusive options, one of which has a cost of £100 and a benefit of £500 and the other a cost of £10,000 and a benefit of £11,000. The first options has a B/C ratio of 5 and the second a B/C ratio of only 1.1. In adopting the first option, however, one would be forgoing a net benefit of £1,000 for the sake of a gain of only £400. The ratio method thus loses any sense of the scale of benefits and costs.

The ratio method can also mislead when it comes to explicit consideration of the appropriate scale of a service or other development. Let costs be £1,000 and benefits £1,500: a B/C ratio of 1.5. Now suppose that by enlarging the scale of this development costs would rise by £100 and benefits by £120. The B/C ratio for the marginal extension of the development is 1.2. It is clearly desirable on grounds of net benefit (yielding £20) but it reduced the overall B/C ratio of the original development (from 1.5 to 1.47) and this appears undesirable - whereas it is actually desirable.

A final difficulty with B/C ratios is that they are sensitive to the classification of cosrs and benefits. For example, if, in the benefits of £1,500, there are £500 averted costs, the ratio rises from 1.5 to 2.0 if averted costs are netted out of costs rather than benefits. On the other hand, the net benefit (B - C) remains the same regardless of how averted costs are treated (i.e. at £500). Similarly, if some benefits are taken as net of costs, with other costs remaining in the denominator, the ratio again changes. For example, if £200 is netted out of benefits, B/C becomes £1,300/£800 = 1.625 rather than 1.5. In general, then, while B/C > 1 is a reasonable rule for identifying options that ought (other considerations apart) to be adopted, the rule "adopt options for which (B - C) > 0" amounts to the same thing and is far less likely to be misinterpreted and misused. In general, ratios are best avoided altogether in favour of benefit-cost differences.

Quantophrenia

Quantophrenia is a syndrome in which the quantified drives out the important. Enough has been said in earlier sections to alert both research practitioners and research customers to the dangers of allowing what has been quantified to assume an unwarranted special status relative to the unquantified. While developments in benefit measurement in particular have helped push back the frontiers of what can be helpfully quantified and helpfully picked apart into component issues for further research and policy debate, the pushing back is far from complete. Ironically, however, the more successful analysts become at quantifying the hitherto unquantifiable (and at alerting research customers to the subtle meanings that quantification can have), the greater the risk that the unquantified will be deemed "small" or "unimportant" relative to the quantified. This danger cannot be sufficiently emphasised.

A characteristic of nearly all economic appraisals (whether CEA, CUA or CBA) is that they are task-oriented and that they tend to ignore (usually completely ignore) the process aspects of delivery which are an important aspect of the management of change in the health services. An appraisal may be used as the basis for determining that a particular service development should take place, but whether it will take place and whether to enable it to take place requires the modification of the options depends upon a host of managerial considerations, arbitrary constraints (that may vary from place to place), the placating of vested interest in the status quo, the reassurance of those who feel threatened, and the gaining of their active support. While such considerations may lie beyond economic appraisal - at least as currently practised - they are a set of considerations that can sabotage the implementation of cost-effective procedures unless appropriately dealt with. An ideal appraisal in the decision-making approach would, by virtue of the closeness of the liaison between analysts and research customers, anticipate such problems and incorporate them (or at least strategies for coping with them) in the analysis. Currently, however, this remains a major "unquantified" area in economic appraisals in the health services.

V CHECKLIST OF KEY POINTS IN APPRAISAL

The following checklist of 25 items (based partly on Williams, 1974) is a set of recommendations that is intended to shape the design of studies, their conduct, and their appraisal by research customers or others who may use the results.

STUDY DESIGN

1. The precise question which the study is to address should be clearly specified and kept to the fore throughout.

2. The perspective from which the question is addressed should be clearly specified and kept to the fore throughout.

3. The objectives of the options being appraised should be clearly specified (output, outcome or benefit).

4. The measures by which these objectives are represented (ordinal or cardinal) should be clearly stated.

5. The measures should be a comprehensive set of the relevant characteristics of the benefit/service aspects of the options.

6. They should be measured (forecast) over time where appropriate in constant prices.

7. Relevant value judgements should be identified and justified.

8. Those competent to make the relevant value judgements should be identified and consulted.

9. Data relevant in the exercise of value judgements should be presented.

10. The reliability of the basic clinical, epidemiological and engineering data on which the appraisal is based should be scrutinised carefully.

11. All possible options should be considered; relevant options should be appraised.

12. Exclusion of options that may occur to those to whom an appraisal is directed, as well as inclusions, should be justified.

TREATMENT OF COSTS AND BENEFITS

13. If the scope of the cost concept employed does not transcend the expenditure of the agency concerned, it should nonetheless represent opportunity costs to the agency.

14. If the scope of the cost concept does transcend the expenditure of the agency concerned it should again represent relevant opportunity costs.

15. All potentially relevant marginal costs should be identified and quantified.

16. Costs and benefits should be clearly differentiated.

17. Differential timing of costs and benefits should be allowed for by discounting.

PRESENTATION

18. Quantitatively significant, uncertain, methodologically suspect, and politically contentious assumptions should be subjected to sensitivity analysis.

19. Options should be subjected to analysis of robustness.

20. Important distributional features of the appraisal should be explicitly analysed.

21. Summary indices of the balance of cost and benefit should be carefully chosen and interpreted.

22. Unquantified costs and benefits should not be assumed without discussion to be small relative to those quantified.

23. The summary should highlight the crucial factors that differentiate the options.

24. The summary should highlight those changes that may alter the order of preference of the options.

25. The generalisability of the results should be discussed and cautions given as appropriate.

Careful consideration of these major points would result, I conjecture, in substantial improvements in the conduct, presentation and interpretation of economic appraisals. A final piece of advice, less I be thought to be counselling perfection, is that there is no point in allowing the perfect to become the enemy of the merely good. So the final thing to bear in mind is that no study has to be better than it needs to be for the purposes in hand. Excessive refinement is costly in money and time, and can lead to research atrophy. "Quick and dirty" studies can sometimes be the most illuminating. But even they are best conducted and appraised bearing in mind this checklist.

To return to my starting point: the limits of economic appraisal in the Paretian approach are partly self- and partly institutionally-imposed - and unnecessarily imposed as well. In the decision-making approach, the limitations are largely those of the imagination, together with a possible (and again unnecessary) reluctance to work closely with research customers and other experts. In both approaches, incompetence is a limitation and, by association in customers' minds with the technique of economic appraisal in general, poses a great theat by casting suspicion on the value of such exercises. To be a good health economist you do really need to be quite a good economist before anything else. But to be able to practise as a good health economist, the system itself needs to generate a demand for one's services - and that means that our health systems need built-in incentives or requirements for the undertaking of appraisals. These they mostly lack and so it also behoves health economists to press for environmental and institutional changes. So there are some aspects of the politics of health services from which one cannot draw back. Thus, having advocated in 1971 a withdrawal from the political debate about health services I am now convinced that we cannot and should not impose a complete self-denying ordinance on ourselves. These practical, as distinct from the intellectual, limitations, need an urgent - if divided - attention.

FOOTNOTES

1. A merit of the Paretian approach is that it decisively limits the role of "experts" like economists and protects society against the excesses of appraisers like Turvey: "My feeling is that the value-judgements made by economists are, by and large, better than those made by non-economists ... The point is simply that those who are experienced in systematic thinking about a problem are those who usually make the best judgements about it" (1963, p. 96). Not all economists "systematically think" about value-judgements; some non-economists do; some economists hold distinctly unpleasant values; society often asks particular people to exercise value judgements on its behalf - and professional economists are rarely amongst this privileged set (of politicians, administrators, doctors, nurses, ...).

2. It should not need pointing out that in using terms like social choice and social acceptability I am referring to decisions respectively made by an individual or individuals on behalf of a wider set of individuals (viz. society) and that are more or less acceptable to other individuals who are members of a society.

3. While the force of Mishan's (1974) arguments is recognised, and while it is undoubtedly true that economists are particularly skilled in identifying potential Pareto improvements (viz. sums of compensating variations) it seems silly to restrict the role of economists and the scope of economic appraisals to these efficiency questions alone since economists also have skills in other matters - for example in eliciting the precise values of decision-makers about intangible benefits, and in measuring and presenting distributional questions in an illuminating way. These various types of consideration should be kept distinct (Culyer, 1977) but not so separate that one set becomes altogether banished.

4. Equivalent annual charge = $K \frac{i(1+i)^n}{(1+i)^{n-1}}$ where K is the capital cost and i the rate of discount.

5. Note that, although I have classed the Culyer and Maynard study as a CEA, it has elements of a CUA in it by virtue of its inclusion of this quality differential. In practice there is no hard and fast distinction to be drawn that defines the boundaries between CEA, CUA and CBA.

6. In other studies (e.g. Ludbrook, 1981, in a CEA of renal failure treatments) the limitations of simple output measures (in her case, life years saved) are discussed without attempting any quantification of the quality of life.

7. Suppose a health index, H, is measured on a linear (interval) scale and related linearly to two independent variables thus:

$$H = a + bY + cX.$$

The elasticities are $\varepsilon_{HY} = bY/H$ and $\varepsilon_{HX} = cX/H$. Let H' be a linear transform of H, such that

$$H = d + eH'.$$

Using H' as the dependent variable yields elasticities $\varepsilon_{H'Y} = bY/(H - d)$ and $\varepsilon_{H'X} = cX/(H - d)$. Note, however, that the ratio of elasticities: $\varepsilon_{HY}/\varepsilon_{HX} = \varepsilon_{H'Y}/\varepsilon_{H'X} = bY/cX$.

Now let H be an independent variable in, say, a demand equation $X = a + bY + cH$. Here the elasticities are $\varepsilon_{XY} = bY/X$, $\varepsilon_{XH} = cH/X$ and the ratio $\varepsilon_{XY}/\varepsilon_{XH} = bY/cH$. Using the linear transform $H = d + eH'$ yields elasticities of $\varepsilon_{XY} = bY/X$ and $\varepsilon_{XH'} = c(H - d)/X$ with the ratio $\varepsilon_{XY}/\varepsilon_{XH'} = bY/c(H - d)$, so even the relative sizes of the elasticities are no longer uniquely determined.

8. Moreover, there is a strong element of circularity in a method that first <u>infers</u> values from public decisions and then feeds them back as "authoritative" values into the public decision-making process. A more useful role might be to confront decision-makers with the results so derived, explain that they seem to be implied by past decisions, ask if they are plausible measures in their context, whether they were those "really" felt at the time, and whether they give any leverage on the selection of appropriate values in the context in hand.

9. The rationale here is that if by, say, questionnaire methods, the maximum willingness to pay for a marginal reduction in risk can be ascertained for the individuals in a population at risk, the actual reduction estimated to be achieved by a procedure multiplied by the sum of the marginal willingnesses to pay gives the social value of the risk reduction. When divided by the expected number of lives saved, this yields the average value of a (statistical) life saved. Its use of the compensating variation and the revealed preferences of relevant population groups puts this approach nicely in sympathy with the potential Pareto improvement basis for all economic appraisals.

REFERENCES

Akehurst, R.L. and Buxton, M.J. (1985), Option Appraisal in the NHS: A Guide to Better Decision-Making, Nuffield/York folio 8, London, Nuffield Provincial Hospitals Trust.

Akehurst, R.L. and Holtermann, S. (1985), Provision of Decentralised Mental Illness Services: An Option Appraisal, York, University of York Centre for Health Economics Discussion Paper 5.

Bartlett, J.R., Neil-Dwyer, G., Banham, J.M.M. and Cruickshank, D.G. (1978), 'Evaluating cost-effectiveness of diagnostic equipment: the brain scanner case', Brit. Med. J., 16, September, 815-20.

Beresford, S.A.A., Chant, A.D.B., Jones, H.O., Piachaud, D. and Weddell, J.M. (1978), 'Varicose veins: a comparison of surgery and injection/ compression sclerotherapy, five year follow-up', Lancet, (i), 921-4.

Blades, C.A., Culyer, A.J. and Walker, A. (1985), Some Dos and Don'ts in Economic Appraisal of Health Care, London, Nuffield Provincinal Hospital Trust, forthcoming.

Blades, C.A., Culyer, A.J. and Wiseman, J. and Walker, A. (1985), International Bibliography of Health Economics, Brighton, Harvester.

Buxton, M., Acheson, R., Caine, N., Gibson, S. and O'Brien, B. (1985), Costs and Benefits of the Heart Transplant Programme at Harefield and Papworth Hospitals, London, HMSO.

Buxton, M.J. and West, R.R. (1975), 'Cost-benefit analysis of long-run haemodialysis for chronic renal failure', Brit. Med. J., 17 May, 376-9.

Coase, R.H. (1960), 'The problem of social cost', J. Law Econs., 3, 1-44.

Coverdale, I., Gibbs, R. and Nurse, K. (1980), 'A hospital cost model for policy analysis', J. Opl. Res. Soc., 31, 801-11.

Culyer, A.J. (1971), 'The nature of the commodity "health care" and its efficient allocation', Oxford Ec. Papers, 23, 189-211.

Culyer, A.J. (1977), 'The quality of life and the limits of cost-benefit analysis' in L. Wingo and A. Evans (eds.), Public Economics and the Quality of Life, Baltimore and London, Johns Hopkins Press.

Culyer, A.J. (1978), 'Need, values and health status measurement' in Culyer A.J. and Wright, K.G. (eds.), Economic Aspects of Health Services, London, Martin Robertson.

Culyer, A.J. (ed.) (1983), Health Indicators, Oxford, Martin Robertson.

Culyer, A.J. and Horisberger, B. (eds.) (1983), Economic and Medical Evaluation of Health Care Technologies, Heidelberg, Springer.

Culyer, A.J. and Maynard, A.K. (1981), 'Cost-effectiveness of duodenal ulcer treatment', Soc. Sci. and Med., 15c, 3-11.

Culyer, A.J., Wiseman, J. and Walker, A. (1977), An Annotated Bibliography of Health Economics, London, Martin Robertson.

Dasgupta, P., Marglin, S. and Sen, A.K. (1972), Guidelines for Project Evaluation, New York, United Nations Industrial Development Organization.

DHSS (1981), National Health Service, Review of Capricode: Building Strategy, London, DHSS.

Dowell, T.B. (1976), 'The economics of fluoridation', Brit. Dent. J., February 3, 103-6.

Dowie, J.A. (1970), 'Valuing the benefit of health improvement', Aus. Econ. Papers, June, 21-41.

Drummond, M.F. (1980), Principles of Economic Appraisal in Health Care, Oxford, Oxford University Press.

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

Drummond, M.F. (1985), Cost-effectiveness Analysis in Health Care, Nuffield/York folio 6, London, Nuffield Provincial Hospitals Trust.

Fordyce, J.D., Mooney, G.H. and Russell, E.M. (1981), 'Economic analysis in health care 2, an application to care of the elderly', Hlth. Bull., 39, 1, 29-38.

Geweke, J. and Weisbrod, B. (1981), 'Some economic consequences of technological advance in medical care: the case of a new drug' in Helms, R.B. (ed.), Drugs and Health, Washington, D.C., American Enterprise Institute.

Geweke, J. and Weisbrod, B. (1982), 'Assessing technological change: the case of a new drug', Madison, University of Wisconsin.

Grogono, A.W. and Woodgate, D.J. (1971), 'Index for measuring health', Lancet ii, 1024-6.

Guttman, L. (1944), 'A basis for scaling qualitative data', Am. Soc. Rev., 9, 139-50.

Hagard, S., Carter, F. and Milne, R.G. (1976), 'Screening for Spina Bifida Cystica: a cost-benefit analysis', Brit. J. Prev. Soc. Med., 30, 40-53.

Hirschleifer, J., Dehaven, J.C., Milliman, J.W. (1960), Water Supply: Economics, Technology and Policy, Chicago, University of Chicago Press.

Jones-Lee, M. (1976), The Value of Life: An Economic Analysis, London, Martin Robertson.

Kind, P., Rosser, R. and Williams, A. (1982), 'Valuation of quality of life: some psychometric evidence' in Jones-Lee, M. (ed.), The Value of Life and Safety, Amsterdam, North-Holland.

Le Grand, J. (1982), The Strategy of Equality, London, Allen and Unwin.

Leu, R.E. (1984), 'Medical technology assessment in terms of improved life quality' in Lindgren, B. (ed.), Pharmaceutical Economics, Malmoe, Liber Forlag.

Little, I. M. D. and Mirrlees, J. A. (1974), Project Appraisal and Planning for Developing Countries, London, Heinemann.

Longmore, D. B. and Rehahn, M. (1975), 'The cumulative cost of death', Lancet, (i), 1023-5.

Lowson, K. V., Drummond, M. F. and Bishop, J. M. (1981), 'Costing new services: long-term domiciliary oxygen therapy', Lancet, (i), 1146-9.

Ludbrook, A. (1981), 'A cost-effectiveness analysis of the treatment of chronic renal failure', App. Ecs., 13, 337-50.

McKean, R. N. (1958), Efficiency in Government Through Systems Analysis, New York, Wiley.

Mishan, E. J. (1971), Cost-benefit Analysis, London, Allen and Unwin.

Mishan, E. J. (1974), 'Flexibility and Consistency in project evaluation', Economica, 41, 161, 81-96.

Mooney, G. H. (1984), Valuing Human Life in Health Service Policy, Nuffield/ York folio 3, London, Nuffield Provincial Hospitals Trust.

Muir-Gray, J. A. (1979), 'Choosing priorities', J. Med. Ethics, 5, 73-5.

Neuhauser, D. and Lewicki, A. M. (1976), 'National health insurance and the sixth stool guiac', Pol. Anal., 24, 175-96.

Neumann, J. and Morgenstern, O. (1953), The Theory of Games and Economic Behaviour (3rd ed.), New York, Wiley.

Pace, A. J. (n.d.), Strategic Planning for a Region of the NHS, Wessex RHA.

Reynell, P. C. and Reynell, M. C. (1972), 'The cost-benefit analysis of a coronary care unit', Brit. Heart J., 34, 897-900.

Rich, G., Glass, N. J. and Selkon, J. B. (1976), 'Cost-effectiveness of two methods of screening for asymptomatic bacteriuria', Brit. J. Prev. Soc. Med., 30, 54-9.

Roberts, C. J., Elder, G. H., Laurence, K. M., Woodhead, J. S., et al. (1983), 'The efficacy of a serum screening service for neural-tube defects: the South Wales experience', Lancet, (i), 1315-18.

Russell, I. T., Devlin, H. B., Fell, M., Glass, N. J. and Newell, D. J. (1977), 'Day-case surgery for hernias and haemorrhoids: a clinical social and economic evaluation', Lancet, (i), 844-7.

Sugden, R. and Williams, A. (1978), The Principles of Practical Cost-Benefit Analysis, Oxford, Oxford University Press.

Thomson, J. L. G. (1977), 'Cost-effectiveness of an EMI brain scanner: a review of two year experience', Health Trends, 9, 16-19.

Torrance, G. (1970), A Generalized Cost-Effectiveness Model for the Evaluation of Health Programs, Hamilton (Canada), Faculty of Business, McMaster University.

Torrance, G. and Zipursky, M. D. (1984), 'Cost-effectiveness of antepartum prevention of Rh immunization', Clins. in Perinatology, 11, 2, 267-81.

Turvey, R. (1963), 'Present value versus internal rate of return: an essay in the theory of third best', Economic J., 73, 93-98.

Weisbrod, B. A. (1971), 'Costs and benefits of medical research: a case study of poliomyelitis', J. Pol. Econ., 79, 527-44.

Williams, A. (1972), 'Cost-benefit analysis: bastard science? and/or insidious poison in the body politick?', J. Pub. Econs., 1, 2, 199-226.

Williams, A. (1974), 'The cost-benefit approach', Brit. Med. Bull., 30, 3, 252-6.

Williams, A. (1985), 'Coronary artery bypass grafting: an economic appraisal', Brit. Med. J. (forthcoming).

Wolfson, A. D. (1974), A Health Index for Ontario, 2nd Report, Toronto, Ministry of Treasury, Economics and Intergovernmental Affairs.

Wright, K. G., Cairns, J. A. and Snell, M. C. (1981), Costing Care: the Costs of Alternative Patterns of Care for the Elderly, Sheffield, Sheffield University Joint Unit for Social Services Research.