Lowson, K. V., Drummond, M. F., and Bishop, J. M. (1981). Costing new services: long-term domiciliary oxygen therapy. *Lancet*, ii, 1146–9.
Ludbrook, A. (1981). A cost-effectiveness analysis of the treatment of chronic renal failure. *Appl. Economics*, 13, 337–50.
Mehrez, A., Gafni, A. (1989). Quality-adjusted life-years, utility theory and health years equivalents. *Medical Decision Making*, 9, 142–9.
Neuhauser, D. and Lewicki, A. M. (1975). What do we gain from the sixth stool guaiac? *N. Engl. J. Med.*, 293, (5), 226–8.
Neumann, P. and Johannesson, M. (1994). The willingness to pay for in vitro fertilization: a pilot study using contingent valuation. *Medical Care*, 32, 686–699.
O'Brien, B. J., Novosel, S., Torrance, G. and Streiner, D. (1995). Assessing the economic value of a new antidepressant: a willingness-to-pay approach. *PharmacoEconomics*, 8(1), 34–5.
Oldridge, N., Furlong, W., Feeny, D., Torrance, G., Guyatt, G., Crowe, J. *et al.* (1993). Economic evaluation of cardiac rehabilitation soon after acute myocardial infarction. *Amer. J. Cardiol.*, 72, 154–61.
Reynell, P. C. and Reynell, M. C. (1972). The cost-benefit analysis of a coronary care unit. *Br. Heart J.*, 34, 897–900.
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.
Sackett, D. L. (1980). Evaluation of health services. In *Health and preventive medicine* (ed. J. M. Last), pp. 1800–23. Appleton-Century Crofts, New York.
Stason, W. B. and Weinstein, M. C. (1977). Allocation of resources to manage hypertension. *N. Engl. J. Med.*, 296, 732–7.
Waller, J., Adler, M., Creese, A., and Thorne, S. (1977). Early discharge from hospital for patients with hernia or varicose veins. *Department of Health and Social Security (UK)*, HMSO, London.
Weisbrod, B. A., Test, M. A., and Stein, L. I. (1980). Alternatives to mental hospital treatment: economic cost-benefit analysis. *Arch. General Psychiatry*, 37, 400–5.

3

Critical assessment of economic evaluation

Those who receive or read an economic evaluation are often faced with the difficult task of assessing study results. The question that readers of evaluations are most likely to ask themselves is: '*Are these results useful to me in my setting?*' The answer to this question is determined by the answers to the following specific questions.

1. Is the methodology employed in the study appropriate and are the results valid?
2. If the results are valid, would they apply to my setting?

This chapter concentrates on question (1), and is designed to assist users of economic evaluation in assessing the validity of the results they encounter.

When assessing the validity of evidence, whether pertaining to efficacy, effectiveness, availability, or efficiency, we normally proceed by examining closely the methods employed to produce the evidence. Often it is helpful to separate the various elements of a methodology so that each can be scrutinized more closely. In this chapter we identify the key elements of any economic evaluation and discuss methodological characteristics which users may expect to find in well-executed studies. A brief summary of relevant questions to ask about an economic evaluation is provided in Box 3.1, and this critical appraisal check-list is then applied to a published article.

Of course, it is unrealistic to expect every study to satisfy all of the points; however, the systematic application of these points will allow readers to identify and assess the strengths and weaknesses of individual studies.

Box 3.1. A check-list for assessing economic evaluations

1. Was a well-defined question posed in answerable form?
- 1.1 Did the study examine both costs and effects of the service(s) or programme(s)?
- 1.2 Did the study involve a comparison of alternatives?
- 1.3 Was a viewpoint for the analysis stated and was the study placed in any particular decision-making context?

2. Was a comprehensive description of the competing alternatives given (i.e. can you tell who did what to whom, where, and how often)?
- 2.1 Were any important alternatives omitted?
- 2.2 Was (Should) a *do-nothing* alternative (be) considered?

3. Was the effectiveness of the programmes or services established?
- 3.1 Was this done through a randomized, controlled clinical trial? If so, did the trial protocol reflect what would happen in regular practice?
- 3.2 Was effectiveness established through an overview of clinical studies?
- 3.3 Were observational data or assumptions used to established effectiveness? If so, what are the potential biases in results?

4. Were all the important and relevant costs and consequences for each alternative identified?
- 4.1 Was the range wide enough for the research question at hand?
- 4.2 Did it cover all relevant viewpoints? (Possible viewpoints include the community or social viewpoint, and those of patients and third-party payers. Other viewpoints may also be relevant depending upon the particular analysis.)
- 4.3 Were capital costs, as well as operating costs, included?

5. Were costs and consequences measured accurately in appropriate physical units (e.g. hours of nursing time, number of physician visits, lost work-days, gained life-years)
- 5.1 Were any of the identified items omitted from measurement? If so, does this mean that they carried no weight in the subsequent analysis?
- 5.2 Were there any special circumstances (e.g. joint use of resources) that made measurement difficult? Were these circumstances handled appropriately?

6. Were costs and consequences valued credibly?
- 6.1 Were the sources of all values clearly identified? (Possible sources include market values, patient or client preferences and views, policy-makers' views and health professionals' judgements.)
- 6.2 Were market values employed for changes involving resources gained or depleted?
- 6.3 Where market values were absent (e.g. volunteer labour), or market values did not reflect actual values (such as clinic space donated at a reduced rate), were adjustments made to approximate market values?
- 6.4 Was the valuation of consequences appropriate for the question posed (i.e. has the appropriate type or types of analysis – cost-effectiveness, cost-benefit, cost-utility – been selected)?

7. Were costs and consequences adjusted for differential timing?
- 7.1 Were costs and consequences which occur in the future 'discounted' to their present values?
- 7.2 Was any justification given for the discount rate used?

8. Was an incremental analysis of costs and consequences of alternatives performed?
- 8.1 Were the additional (incremental) costs generated by one alternative over another compared to the additional effects, benefits, or utilities generated?

9. Was allowance made for uncertainty in the estimates of costs and consequences?
- 9.1 If data on costs or consequences were stochastic, were appropriate statistical analyses performed?
- 9.2 If a sensitivity analysis was employed, was justification provided for the ranges of values (for key study parameters)?
- 9.3 Were study results sensitive to changes in the values (within the assumed range for sensitivity analysis, or within the confidence interval around the ratio of costs to consequences)?

10. Did the presentation and discussion of study results include all issues of concern to users?
- 10.1 Were the conclusions of the analysis based on some overall index or ratio of costs to consequences (e.g. cost-effectiveness ratio)? If so, was the index interpreted intelligently or in a mechanistic fashion?
- 10.2 Were the results compared with those of others who have investigated the same question? If so, were allowances made for potential differences in study methodology?
- 10.3 Did the study discuss the generalizability of the results to other settings and patient/client groups?
- 10.4 Did the study allude to, or take account of, other important factors in the choice or decision under consideration (e.g. distribution of costs and consequences, or relevant ethical issues)?
- 10.5 Did the study discuss issues of implementation, such as the feasibility of adopting the 'preferred' programme given existing financial or other constraints, and whether any freed resources could be redeployed to other worthwhile programmes?

3.1. ELEMENTS OF A SOUND ECONOMIC EVALUATION

1. Was a well-defined question posed in answerable form?

Such a question will clearly identify the alternatives being compared and the viewpoint(s) from which the comparison is to be made. Questions such as, '*Is*

a chronic home care programme worth it?' and, '*Will a community hypertension screening programme do any good?*' solicit the issues of *to whom and compared to what*. Similarly, questions such as, '*How much does it cost to run our intensive care unit?*', and '*What are the costs and outcomes of adolescent counselling by social workers?*' are not efficiency questions because they fail to specify the alternatives for comparison. (See Chapter 2 for a review of the basic types of economic evaluation.) This is not to say that the questions do not provide important accounting or management information; they may do so, but the answers to them do not by themselves qualify as efficiency statements.

A well-specified question, for example, might look as follows: '*From the viewpoint of (a) both the Ministry of Health and the Ministry of Community and Social Services budgets, and (b) patients incurring out-of-pocket costs, is a chronic home care programme preferable to the existing programme of institutionalized, extended care in designated wards of general hospitals?*'. Note that the viewpoint for an analysis may be that of a specific provider or providing institution, the patient or groups of patients, a third-party payer (public or private), or society (i.e. all costs and consequences to whomsoever they accrue). Obviously it is difficult for the analyst to consider every single cost and consequence of a health care programme to all members of society. Indeed, the 'ripple effects' of some programmes may be far-reaching and consideration of some items may have to be excluded for practical reasons. However, it is important to recognize that in considering the use of the community's scarce resources, the viewpoint of the providing institution may often be too restrictive and a broader viewpoint should also be considered. For example, it may be that a programme is preferable from the societal viewpoint, but not from the viewpoint of the providing institution. In such a case the Ministry of Health may wish to consider giving an incentive to the providing institution to ensure that the socially preferred programme goes ahead. The existence of different viewpoints was highlighted by Weisbrod *et al.* (1980) in their study of community-oriented and hospital-based treatments for mental illness.

2. Was a comprehensive description of the competing alternatives given?

A clear and specific statement of the primary objective of each alternative programme, treatment, or service is critical in selecting among cost-effectiveness, cost–utility, and cost–benefit analysis as the type of evaluation to be undertaken. A full description of the competing alternatives is essential for three further reasons:

1. Readers must be able to judge the applicability of the programmes to their own settings.
2. Readers should be able to assess for themselves whether any costs or consequences may have been omitted in the analysis.

3. Readers may wish to replicate the programme procedures being described.

Therefore, readers should be provided with information allowing an identification of costs:

- who does what to whom,
- where, and
- how often,

and consequences:

- what are the results?

3. Was there evidence that the programmes' effectiveness had been established?

We are not interested in the efficient provision of ineffective services, i.e. those services which have been shown to do no more good than harm (by themselves, or compared with no treatment). In fact, we are not interested in the provision of such services under any conditions, efficient or otherwise. *If something is not worth doing, it is not worth doing well!* Therefore, if the economic evaluation assumes effectiveness, some indication of the prior validation of effectiveness should be given. It is also possible that the efficiency evaluation may have been conducted simultaneously with the evaluation of efficacy or effectiveness. This is the case in many randomized trials of therapies which also include a comparison of the costs of the experimental programme and a control, which may be a placebo or the currently existing programme. Note, however, that efficiency evaluations, by themselves, are incapable of establishing effectiveness. Precedent or simultaneous evidence of effectiveness is required. After all, there are efficient methods of worsening the quality of life as well as improving it! [Those wishing to know more about the methods of establishing whether a therapy does more good than harm should consult Sackett *et al.* (1991).]

4. Were all the important and relevant costs and consequences for each alternative identified?

Even though it may not be possible or necessary to measure and value all o the costs and consequences of the alternatives under comparison, a ful identification of the important and relevant ones should be provided. Th combination of information contained in the viewpoint statement and pro gramme description should allow judgement of what specific costs an consequences or outcomes it is appropriate to include in the analysis.

An overview of the categories of costs and consequences that are relevant t an economic evaluation of health services and programmes was given in Fi 2.2. (This is presented again here as Fig 3.1.)

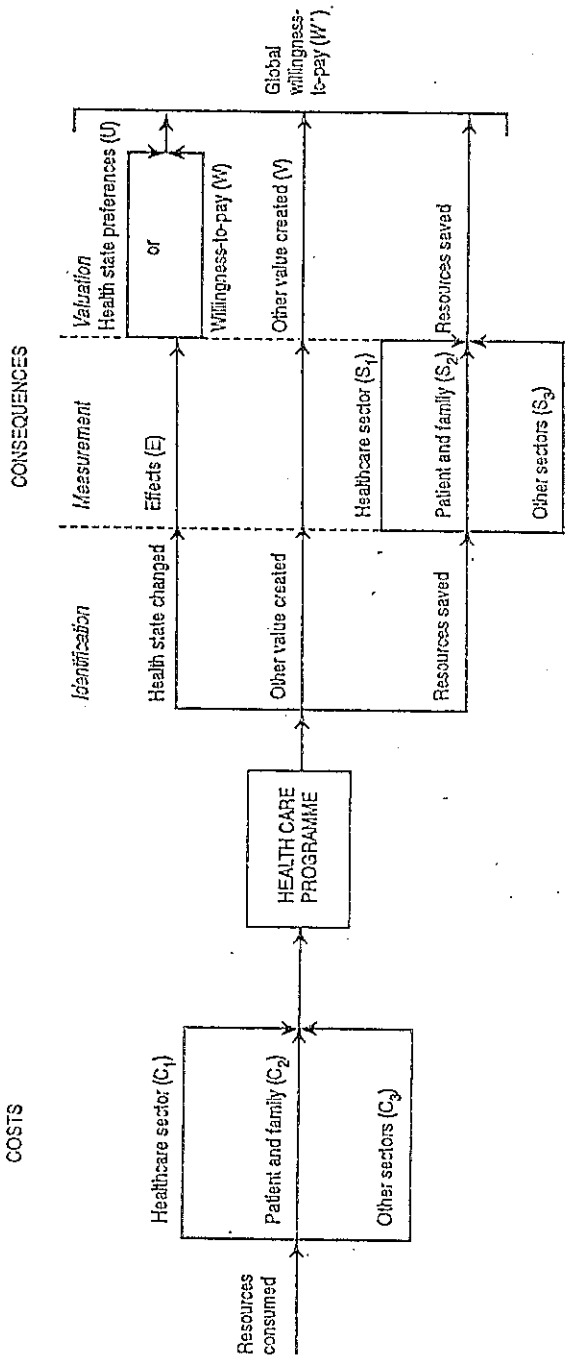


Figure 3.1. Components of economic evaluation in health care

Three categories of cost are identified. The health care resources consumed consist of the costs of organizing and operating the programme, including dealing with the adverse events caused by the programme. The identification of these costs often amounts to listing the *ingredients* of the programme – both variable costs (such as the time of health professionals or supplies) and fixed or overhead costs (such as light, heat, rent or capital costs). The ways of measuring and valuing these items are discussed in more detail in Chapter 4.

The patient and family resources consumed include any out-of-pocket expenses incurred by patients and/or family members as well as the value of any resources that they contribute to the treatment process. This would include the value of their time and sometimes patients and/or family members may lose time from work while seeking treatment or participating in a health care programme.

While the above two categories of costs cover most of the cost items relevant to economic evaluations of health services, a third category, resources consumed in other sectors, also warrants mention. As mentioned in Chapter 2, some programmes, such as those for care of the elderly, consume resources from other public agencies or the voluntary sector. Occasionally it may also be the case that the operation of a health service or programme changes the resource use in the broader economy. Examples of situations where these factors may be important are:

(1) an occupational health and safety programme (perhaps legislated by government) that changes the production process in an automobile manufacturing plant, thereby using up more resources, perhaps in a more labour-intensive way. These costs are passed on in the increased price of cars and are borne by the purchasers of cars, who are likely not the workers for whom the programme was initiated;
(2) a 55-mile-per-hour speed limit policy which reduces morbidity and mortality due to accidents, but increases the price of, for example, fruit which now takes longer to arrive (i.e. a higher wage bill for the truck driver).

In principle, these factors should be considered in the economic evaluation, though for many health care programmes they may be insignificant. (Few economic analyses of alternative health care programmes take them into account.)

Three categories of consequences of health services or programmes are also shown in Fig 3.1. The changes in health state relate to changes in the physical, social or emotional functioning of individuals. In principle, such changes can be measured objectively, and refer only to an individuals ability to function and not to the significance, preference or value attached to this ability by the individual, or by others. However, as indicated in Fig 3.1 and discussed later, values can be attached either by health state preference scores or willingness-to-pay.

In addition, other value may be created by the programme (e.g. reduction in anxiety) and resources may be freed. For example, a vaccination programme may free resources if fewer individuals contract the disease and thus require treatment. This was discussed in Chapter 2.

Given the wide range of costs and consequences indicated above, it may be unrealistic to expect all relevant items to be measured and valued in the analysis, due to their small size or influence relative to the effort required to measure or value them accurately; however, it is helpful to users to identify as many relevant items as possible. It is particularly important that the outcomes of interest be identified clearly enough for a reader to judge the appropriateness of the type (or types) of economic evaluation chosen. That is, it should be apparent:

- whether a single outcome is of primary interest as opposed to a set of outcomes;
- whether the outcomes are common to both alternatives under comparison, and;
- to what degree each programme is successful in achieving each outcome of interest.

Similarly, it is important to know whether the consequences of primary interest are the therapeutic effects themselves (thus implying cost-effectiveness analysis if possible), the change in quality of life of patients and their families (cost–utility analysis), or the overall value created (cost–benefit analysis).

5. Were costs and consequences measured accurately in appropriate physical units?

While identification, measurement, and valuation often occur simultaneously in analyses, it is a good practice for users of evaluation results to view each as a separate phase of analysis. Once the important and relevant costs and consequences have been identified, they must be measured in appropriate physical and natural units. For example, measurement of the operating costs of a particular screening programme may yield a partial list of *ingredients* such as 500 physical examinations performed by physicians, 10 weeks of salaried nursing time, 10 weeks of a 1000 square foot clinic, 20 hours of medical research librarian time from an adjoining hospital, etc. Similarly, costs borne by patients may be measured, for instance, by the amount of medication purchased, or the number of times travel was required for treatment, or the time lost from work while being treated.

Notice that situations in which resources are jointly used by one or more programmes present a particular challenge to accurate measurement. How much resource use should be allocated to each programme? And on what basis? A common example of this is found in every hospital, where numerous

clinical services and programmes share common overhead services (e.g. electric power, cleaning, and administration), provided centrally. In general, there is no non-arbitrary solution to the measurement problem; however, users of results should satisfy themselves that *reasonable* criteria (number of square feet, number of employees, number of cases, etc.) have been used to distribute the common costs. Users should definitely ascertain that such shared costs have been allocated, in fact, to participating services or programmes, as this is a common omission in evaluations! Clinical service directors often argue that small changes in the size of their programmes (up or down) do not affect the consumption of central services. Sometimes it is even argued that overhead costs are unaffected by the service itself. However though this argument may be intuitively appealing from the viewpoint of a particular programme or service director, the extension of this method to each service in the hospital would imply that the totality of services could be operated without light, heat, power, and secretaries! (The allocation of overhead costs is discussed further in Chapter 4.)

With respect to the measurement of consequences, if the identification of outcomes of interest has been clearly performed, then selection of appropriate units of measurement for programme effects should be relatively straightforward. For example, effects might relate to mortality and be measured in life years gained or deaths averted; they might relate to morbidity and be measured, for example, in reductions in disability days or improvement on some index of health status measuring physical, social, or emotional functioning; they may be even more specific, depending upon the alternatives under consideration. Thus, percentage increase in weight-bearing ability may be an appropriate natural measurement unit for an evaluation of a physiotherapy programme, while the number of correctly diagnosed cases may be appropriate for a comparison of venography with leg scanning in the diagnosis of deep-vein thrombosis.

Changes in resource use resulting from the effects will be measured in physical units similar to those employed for costs. Thus the changes in utilization resulting from any particular programme will likely be recorded in numbers of procedures, or amounts of time, space, or equipment. Changes in the resource use by patients will continue to be measured in amounts of medication purchased, trips taken for treatment, and so forth.

While the nature of changes in the quality of life may be described, this is one case where measurement in objective, physical, or natural units is difficult, although the consequence of some surgical interventions may be quantified in number of complications. However, the adjustment of effects for quality of life is usually a matter of valuation, although we also discuss the use of quality of life scales in cost-effectiveness analysis in Chapter 5.

6. Were costs and consequences valued credibly?

The sources and methods of valuation of costs, benefits, and utilities should be clearly stated in an economic evaluation. Costs are normally valued in units of local currency, based on prevailing *prices* of, for example, personnel, commodities, and services, and can often be taken directly from programme budgets. All current and future programme costs are normally valued in constant dollars of some base year (usually the present), in order to remove the effects of inflation from the analysis.

It should be remembered that the objective in valuing costs is to obtain an estimate of the worth of resources depleted by the programme. This may necessitate adjustments to some apparent programme costs (e.g. the case of subsidized services or volunteer labour received by one programme instead of another). In addition, valuation of the cost of a day of institutional care for a specific condition is particularly troublesome in that the use of an average cost per day (the widely quoted *per diem*), calculated on the basis of the institution's entire annual case-load, is almost certainly an overestimate or underestimate of the actual cost for any specific condition, sometimes by quite a large amount.

In principle and (with great effort) in practice, it is possible to identify, measure and value each depleted resource (e.g. drugs, nursing time, light, food, etc.) in treating a specific patient or group of patients. While this yields a relatively accurate cost estimate, the detailed monitoring and data collection are usually prohibitively expensive. The other broad alternative costing strategy is to start with the institution's total costs for a particular period and then to improve upon the method of simply dividing by the total patient-days to produce an average cost-per-day. Quite sophisticated methods of cost allocation to individual hospital departments or wards have been developed, as illustrated by Boyle *et al.* (1982) with respect to neonatal intensive care. An intermediate method involves acceptance of the components of the general *per diem* relating to *hotel* costs (since these are relatively invariant across patients) combined with more precise calculation of the medical treatment costs associated with the specific patients in question. For an example of this intermediate approach see Hull *et al.* (1982). Of course, the effort devoted to accurate *per diem* estimates depends upon their overall importance in the study; however, unthinking use of *per diems* or average costs should be guarded against. (This is discussed further in Chapter 4.)

In valuation of preferences or utilities, we are basically attempting to ascertain how much better the quality of life is in one health situation or 'state' compared with another (e.g. dialysis at home with help from a spouse or friend versus dialysis in hospital). Several techniques are available for making the comparison; the important thing to note is that each will produce an adjustment factor with which to increase or decrease the value of time spent in health situations or 'states', resulting from the alternative in question relative to some baseline. The results of these analyses are usually expressed in *healthy years* or *quality-adjusted life-years* gained, as a result of the programmes being evaluated.

Two broad approaches to estimating health state preferences can be found in the literature. The first approach, outlined by Torrance (1986), emphasizes the development of measurement methods and empirical testing on different populations. The other approach, outlined by Weinstein (1981), places emphasis on the estimation of preference scores by a quick (and inexpensive) consensus-forming exercise, and then undertaking extensive sensitivity analysis on the chosen scores to see whether study results change if the chosen estimates are varied. We see a role for both approaches. The latter approach is useful in getting decision-makers to think about resource allocation problems and is, in fact, relatively quick and inexpensive. The measurement approach is useful in highlighting the fact that different actors (doctors, policy-makers, patients, and the general public as taxpayers) may have different preferences, and is clearly crucial in situations where the study result is sensitive to the preference scores assigned. (An example of such a case arose in the study by Stason and Weinstein (1977) on the economics of hypertension therapy. The study result was sensitive to whether it was assumed that the side effects of antihypertensive drugs constituted a one or two per cent reduction in health status.)

Since the measurement of preferences in health is still a relatively new field, there are many unresolved issues which readers of cost–utility analyses should note. Users of such analyses will probably want to know, at minimum, *whose* preferences were used to construct the adjustment factor – the patient's, the provider's, the taxpayer's or the bureaucrat's? If patients' preferences have not been employed, we may want to assure ourselves further that the persons whose preferences did count clearly understood the characteristics of the health state, either through personal experience or through a description of the state presented to them. These issues are taken up in Chapter 6.

Many of the same issues arise when estimating willingness-to-pay, either for a change in health state or for the overall impact of the programme in question. These issues are taken up in Chapter 7.

One of the important consequences of health care programmes is the creation of healthy time. The valuation of this item poses difficulties. Indeed, its categorization, under changes in health state or patient and family resources freed is uncertain. The reasons for this are as follows.

The value of healthy time can manifest itself in a number of ways. First, living in a better health state has a value to the individual in its own right (e.g. less pain, better quality of life). Secondly, healthy time can be used in leisure. Thirdly, healthy time can be used for work, which generates income for the individual and productive output for society.

In a CEA, the measurement of the effects (E) does not capture the value of healthy time. Therefore, if it is to be included it would have to be estimated

separately, as an element of the patient and family resources freed. In a CUA or CBA, where we are attempting to *value* the consequences of the programme, we might expect that the value of living in a better health state is captured in the preference score (U), or the willingness-to-pay (W). The value of healthy time in leisure is probably also captured in (U) or (W), as it is closely linked with improved quality of life.

Whether or not the value of using healthy time for work is also included in (U) or (W) probably depends on how the scenario (used for valuation) is written. It may be possible, for example, to ask individuals to imagine that their income would not be affected by their health state, since this is covered by unemployment insurance. In such a case, an analyst undertaking an evaluation from a societal viewpoint may wish to include a separate estimate of productivity gains to society if a persons health state is improved. If so, it would be important to ensure that the person did *not* include this in their own valuation, so as to avoid double counting.

Of course, healthy time can also be consumed by a programme if it requires the individual to spend time seeking or undergoing treatment, perhaps in hospital. In a CEA, the value of healthy time lost would have to be estimated separately.

In a CUA or CBA, the way forward would again depend on how the health state scenario is described. If the description contained elements of the process of undergoing care (e.g. you will be admitted to hospital for seven days for treatment), this may be reflected in the preference score (U) or willingness-to-pay (W). Otherwise, the value of healthy time lost in therapy may have to be estimated separately.

Therefore, we have another example of where the assembly of components (building blocks) in the analysis is partly dependent on how they are measured and valued. In conducting an economic evaluation from a societal perspective, it is important to avoid both zero-counting and double-counting. Of course, the estimation of global willingness-to-pay (W') avoids the problem of categorizing the value of healthy time, either as a component of the change in health state or as a component of patient and family resources freed. Rather, the challenge of this approach is to ensure that individuals appreciate *all* the elements of value created and resources consumed by a health care programme, so that this is reflected in their valuation (W').

7. Were costs and consequences adjusted for differential timing?

Since comparison of programmes or services must be made at one point in time (usually the present), the timing of programme costs and consequences which do not occur entirely in the present must be taken into account. Different programmes may have different time profiles of costs and consequences. For example, the primary benefits of an influenza immunization programme are immediate while those of hypertension screening occur well into the future. The time profile of costs and consequences may also differ within a single programme; the costs of the hypertension screening programme would be incurred in the present. Therefore, future dollar cost and benefit streams are reduced or 'discounted' to reflect the fact that dollars spent or saved in the future should not weigh as heavily in programme decisions as dollars spent or saved today. This is primarily due to the existence of *time preference*. That is, individually and as a society we prefer to have dollars or resources now as opposed to later because we can benefit from them in the interim. This is evidenced by the existence of interest rates (as well as the popular wisdom about 'a bird in the hand'). Moreover since *time preference* is not exclusively a financial concept, discounting of outcomes should also be undertaken in cost-effectiveness and cost–utility studies (Weinstein and Stason 1977). The mechanics of discounting and the choice of discount rate are discussed in Chapter 4.

8. Was an incremental analysis of costs and consequences of alternatives performed?

For meaningful comparison, it is necessary to examine the additional costs that one service or programme imposes over another, compared with the additional effects, benefits, or utilities it delivers. This *incremental* approach to analysis of costs and consequences can be illustrated by reference to one of the examples cited in Chapter 2, namely strategies for the diagnosis of deep-vein thrombosis (DVT) (Hull *et al.* 1981).

Table 3.1 shows the costs and outcomes (in terms of correct diagnoses) generated by two alternative strategies: impedance plethysmography (IPG) alone versus IPG plus out-patient venography if impedance plethysmography is negative. (IPG is a non-invasive strategy, whereas venography, the diagnostic 'gold standard' for DVT, can cause pain and other unpleasant side effects.) Although one could compare the simple ratios of costs to outcomes for the two alternatives, the correct comparison is the one of incremental costs over incremental outcomes, since this tells us how much we are paying (for each extra correct diagnosis) in adding the extra diagnostic test. In this case the relevant figure is therefore $4781 per correct diagnosis, not the average figure for the second programme, $3003 per correct diagnosis. It may be decided that $4781 is still a price worth paying; however, it is important to be clear on the principle since earlier (in Chapter 2) we pointed out that, in the case of screening for cancer of the colon, there was a big difference between the average cost (per case detected) of a protocol of six sequential tests and the incremental cost of performing a sixth test, having already done five (Neuhauser and Lewicki 1975).

Similar incremental analyses could be performed if the effects were in years of life or healthy years. These can be illustrated graphically on a four quadrant diagram known as the *cost-effectiveness plane* (Black 1990). (See Box 3.2.)

Table 3.1. Economic evaluation of alternative diagnostic strategies for 516 patients with clinically suspected deep-vein thrombosis*

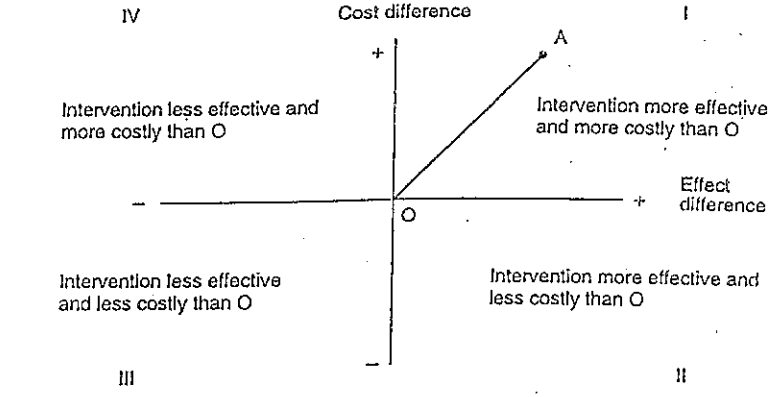
Programme	Costs ($ US)	Outcomes (No. of correct diagnoses)	Ratio of cost to outcome ($ per correct diagnoses)
1. IPG (alone)	321 488	142	2264
2. IPG plus out-patient venography if IPG negative	603 552	201	3003
3. Increment (of Programme 2 over Programme 1)	282 064	59	4781

*Data drawn from Table 1, Hull *et al.* (1981), by permission.

Box 3.2. The cost-effectiveness plane

In the diagram the horizontal axis represents the difference in effect between the intervention of interest (A) and the relevant alternative (O), and the vertical axis represents the difference in cost. The alternative (O) could be the status quo or a competing programme.

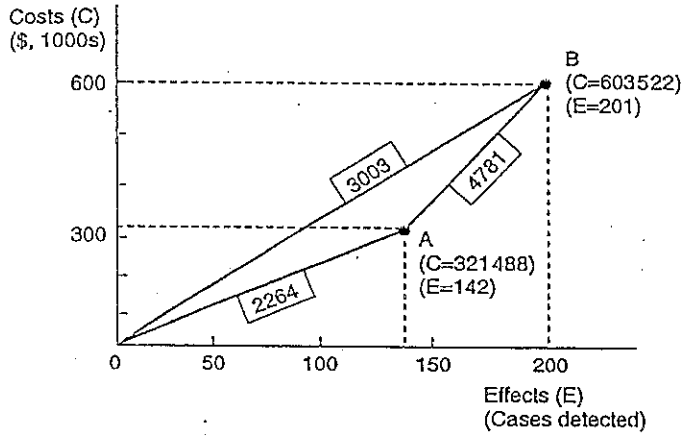
If point A is in quadrants II or IV the choice between the programmes is clear. In quadrant II the intervention of interest is both more effective and less costly than the alternative. That is, it *dominates* the alternative. In quadrant IV the opposite is true. In quadrants I and III the choice depends on the maximum cost-effectiveness ratio one is willing to accept. The slope of the line OA gives the cost-effectiveness ratio.



Adapted from Black (1990)

In practice the impact of most interventions falls in quadrant I. That is, they add to cost but increase effectiveness, certainly when compared with no intervention. Let us therefore plot the data given in Table 3.1 (see Fig. 3.2). Here we only show quadrant I and the slopes of the lines from the origin give the average cost-effectiveness ratios for the two programmes, which are $2264 and $3003 per case detected for A and B respectively. The incremental cost-effectiveness ratio ($4781 per case detected) is given by the slope of the line joining points A and B.

Figure 3.2. Average and incremental cost-effectiveness ratios



9. Was allowance made for uncertainty in the estimates of costs and consequences?

Every evaluation will contain some degree of uncertainty, imprecision, or methodological controversy. What if the compliance rate for influenza vaccination was 10 per cent higher than considered for the analysis? What if the *per diem* hospital cost still understated the true resource cost of a treatment programme by $100? What if a discount rate of six per cent had been used instead of two per cent? Or what if indirect costs and benefits had been excluded from the analysis? Users of efficiency studies will often ask these and similar questions; therefore, careful analysts will identify critical methodological assumptions or areas of uncertainty. Furthermore, they will often attempt to rework the analysis (qualitatively if not quantitatively), employing different assumptions or estimates in order to test the sensitivity of the results and conclusions to such changes. If large variations in the assumptions or estimates underlying an analysis do not produce significant alterations in the results then one would tend to have more confidence in the original results. If the converse occurs, more effort is then required to reduce the uncertainty and/or improve the accuracy of the critical variables. In either

case, this *sensitivity analysis* is an important element of a sound economic evaluation. Sensitivity analysis is discussed further in Chapter 5.

Up until recently most of the data presented in economic evaluations have been deterministic (i.e. given as point estimates). Therefore, sensitivity analysis has been the primary method for allowing for uncertainty. However, today more economic evaluations are being conducted alongside clinical trials. Here the data are typically stochastic (i.e. having a mean and variance). For the stochastic clinical data it is customary to perform tests of statistical significance, or to present confidence intervals around the estimates of clinical effect. This would also be an option for the resource use or cost data; for example, statistical tests could be performed to determine whether or not the mean lengths of stay or total costs for two treatments were indeed different from one another.

Therefore, in principle uncertainty in the estimates of costs and consequences can be allowed for by sensitivity analysis, statistical inference, or some combination of the two approaches. This issue is explored in more detail in Chapter 8.

10. Did the presentation and discussion of study results include all issues of concern to users?

It will be clear from the foregoing discussion that the economic analyst has to make many methodological judgements when undertaking a study. Faced with users who may be mainly interested in the 'bottom line' – e.g. 'Should we buy a CT scanner' – how should the analyst present the results?

Decision indices such as cost-effectiveness and cost–benefit ratios are a useful way of summarizing study results. However, they should be used with care for, in interpreting them, the user may not be completely clear on what has gone into their construction. Some analysts give a range of results. For example, in their economic evaluation of neonatal intensive care for very-low-birth-weight infants, Boyle *et al.* (1983) compare the results for infants below 1000 g and from 1000-1500 g in terms of costs to hospital discharge, costs and consequences to age 15, and costs and consequences for lifetime (Table 3.2). They leave it to the users of the study to decide on which index (or indices) neonatal intensive care should be judged, since the different measures incorporate different value judgements and varying amounts of precision. (For example, the index of 'net economic benefit' includes production gains/losses, and the index of 'cost per quality-adjusted life-year' incorporates the preferences for health states of a sample of the local population.) This leads to another general point, namely it is important for analysts to be as explicit as possible about the various judgements they have made in carrying out the study. A good study should leave the user more (rather than less!) aware of the various technical and value judgements necessary to arrive at resource allocation decisions in health care.

Table 3.2. Measures of economic evaluation of neonatal intensive care according to birth-weight class (5 per cent discount rate)[a]

Period	Birth-weight class	
	1000–1499 g	500–999 g
	$	
To hospital discharge[b]		
Cost/additional survivor at hospital discharge	59 500	102 500
To age 15 (projected)		
Cost/life-year gained	6 100	12 200
Cost/QALY[c] gained	7 700	40 100
To death (projected)		
Cost/life-year gained	2 900	9 300
Cost/QALY[c] gained	3 200	22 400
Net economic benefit (loss)/live birth	(2 600)	(16 100)
Net economic cost/life-year gained	900	7 300
Net economic cost/QALY[c] gained	1 000	17 500

[a]Values are expressed in 1978 Canadian dollars. Multiply by 0.877 to calculate equivalent 1978 US dollars.
[b]All costs and effects occurred in year one.
[c]QALY denotes quality-adjusted life-year.
(From Boyle *et al.* (1983), by permission.)

Finally, a good study should begin to help the user interpret the results in the context of his or her own particular situation. This can be done by being explicit about the viewpoint for the analysis (an earlier point) and by indicating how particular costs and benefits might vary by location. For example, the costs of instituting day-care surgery may vary, depending on whether a purpose-built day-care unit already exists or whether wards have to be converted. Similarly, the benefits of day-care surgery may vary depending on whether, in a particular location, there is pressure on beds and whether beds will be closed or left empty (Russell *et al.* 1977). Obviously it is impossible for the analyst to anticipate every possibility in every location, but one limitation of economic evaluation techniques (discussed in Section 3.2) is that they assume that freed resources will be put to other beneficial uses. Evans and Robinson (1980) argue that in the case of day-care surgery the full economic pay-off may not have been obtained in at least one Canadian hospital.

The presentation, interpretation, and use of economic evaluation results raise a number of practical issues. For example, can the results (e.g. cost-effectiveness ratios) from different studies be meaningfully compared; can results of studies be generalized from one setting, or country, to another; can

guidelines for good practice in the presentation of results be specified? These issues will be explored in Chapter 9.

3.2. LIMITATIONS OF ECONOMIC EVALUATION TECHNIQUES

Our main purpose in this chapter is to make the user of economic evaluation results more aware of the methodological judgements involved in undertaking an economic evaluation in the health care field. In Box 3.1 we have consolidated the points made in this chapter into a suggested check-list of questions to ask when critically assessing economic evaluation results. Some of these questions signal limitations of economic evaluation techniques. For example, economic evaluation techniques rely on the establishment of programme effectiveness. In addition, there are several other limitations of which users should be aware.

Of primary concern from a policy viewpoint is the fact that economic evaluations do not usually incorporate the importance of the distribution, of costs and consequences, among different patient or population groups, into the analysis. Yet, in some cases, the identity of the recipient group (e.g. the poor, the elderly, working mothers, or a geographically remote community) may be an important factor in assessing the social desirability of a service or programme. Indeed, it may be the motivation for the programme in the first place. Although it is sometimes suggested that differential weights be attached to the value of outcomes accruing to special recipient groups, this is not normally done within an economic evaluation. Rather, an equitable distribution of costs and consequences across socioeconomic or other defined groups in society is viewed as a competing dimension upon which decisions are made, in addition to that of efficient deployment of resources.

A more subtle, yet important, point is that the various forms of analysis discussed above embody different equity criteria. For example, cost–benefit analysis values health outcomes in terms of individuals' willingness-to-pay. Sometimes willingness-to-pay may be constrained by ability to pay and therefore valuations are dependent on the existing income distribution. On the other hand the simple aggregation of QALYs in a cost–utility analysis implies that QALY is being valued the same no matter to whom it accrues. Therefore, in reality it is difficult to divorce equity considerations from the economic evaluation and analysts should be aware of this when selecting a particular analytic technique.

It should also be noted that economic evaluation techniques assume that resources freed or saved by preferred programmes will not in fact be wasted but will be employed in alternative worthwhile programmes. This assumption warrants careful scrutiny, for if the freed resources are consumed by other ineffective or unevaluated programmes, then not only is there no saving, but overall health system costs will actually increase without any assurance of additional improvements in the health status of the population.

Finally, evaluation of any sort is in itself a costly activity. Bearing in mind that *even a cost–benefit analysis should be subject to a cost–benefit analysis*, it seems reasonable to suggest that economic evaluation techniques will prove most useful in situations where programme objectives require clarification, the competing alternatives are significantly different in nature, or large resource commitments are under consideration.

3.3. CONCLUSIONS

In these introductory chapters, we have tried to assist users of economic evaluations in interpreting evaluation studies and assessing their usefulness for health care decisions, or for planning further analyses. The rationale for economic evaluation, its fundamental characteristics, and the basic types of economic evaluation were described in Chapter 2. In Chapter 3 we have identified and discussed 10 questions which readers of economic evaluations can ask in order to critically assess a particular study; a check-list of these questions is given in Box 3.1.

Our intent in offering a check-list is not to create hypercritical users who will be satisfied only by superlative studies. It is important to realize, as emphasized at the outset, that for a variety of reasons it is unlikely that every study will satisfy all criteria. However, the use of these criteria as screening devices should help users of economic evaluations to identify quickly the strengths and weaknesses of studies. Moreover, in assessing any particular study, users should ask themselves one final question, '*How does this evaluation compare with our normal basis for decision-making?*' They may find that the method of organizing thoughts embodied in the evaluation compares well with alternative approaches, even bearing in mind the possible deficiencies in the study.

3.4. CRITICAL APPRAISAL OF A PUBLISHED ARTICLE

MARK, D. B. *et al.* **Cost-effectiveness of thrombolytic therapy with tissue plasminogen activator (t-PA) as compared with streptokinase (SK) for acute myocardial infarction (AMI)**

New England Journal of Medicine 1995; 332: 1418–24

1. Was a well-defined question posed in an answerable form?

__X__ YES _____ NO _____ CAN'T TELL

The authors explain the context of their study, which is that some observers have questioned whether the improved survival rates observed in the GUSTO study (a multicentre, randomized clinical trial; comparing t-PA with SK) are worth the substantial additional cost (p.1418).

The authors state that they conducted a cost-effectiveness analysis to compare the value of t-PA treatment with that of streptokinase (p.1418).

They state that they used a social perspective to identify relevant costs, although indirect costs (e.g. time lost from work) and non-medical costs were not included (p. 1419). The reasons for exclusion of some relevant costs are not given. The reader would have to assess whether the exclusion of some costs biases the results.

2. Was a comprehensive description of the competing alternatives given (i.e. can you tell who, did what, to whom, where, and how often)?

__X__ YES _____ NO _____ CAN'T TELL

Details are given of the GUSTO study which compared four different regimens: accelerated tissue plasminogen activator (t-PA), streptokinase with intravenous heparin, SK with subcutaneous heparin and a combination of t-PA and SK. Accelerated is defined as the administration of t-PA over a period of 1.5 hours, rather than the conventional period of three hours (p.1418). The alternatives for the economic study are accelerated t-PA and SK.

For further details the reader would have to consult the GUSTO clinical study. However, one can infer a hospital setting as SK can only be administered in hospital. One important factor is the period of time, post-AMI, before patients receive thrombolytic therapy. This has been shown to be important in previous studies, but no details are given here.

In addition, some subgroup analyses are performed, by age of patient and location of infarction. Implicitly these represent additional alternatives for economic analysis, although they are not the major alternatives being examined.

3. Was there evidence that the programmes effectiveness had been established?

__X__ YES _____ NO _____ CAN'T TELL

The primary clinical endpoint, survival to one year, was estimated in the randomized controlled trial, representing the strongest form of evidence. One year after enrolment, patients who received t-PA had a higher survival rate (an

increase of 1.1 per cent, or 11 per 1000 patients treated) than SK-treated patients.

However, in order to calculate the life-years gained, the denominator in the cost-effectiveness ratio, it was necessary to extrapolate beyond one year. This was done by; (i) a Cox proportional-hazards model based on the experience of 4379 patients in the Duke Cardiovascular Disease Database (giving an extrapolation from one to 15 years) and; (ii) a statistical extrapolation for the tail of the survival curve (beyond 15 years) (p.1419).

A number of assumptions were necessary for the extrapolation, the main one being that the hazard of death after one year did not depend on the thrombolytic agent received (i.e. that the survival curves of the two treatment groups were parallel).

4. Were all the important and relevant costs and consequences for each alternative identified?

_____ YES _____ NO __X__ CAN'T TELL

As mentioned above, changes in productive output and non-medical costs were excluded. This probably makes no substantial difference to the cost-effectiveness ratio, but it is hard to tell.

One area where these costs could be important is in disabling, non-fatal stroke. In the first 30 days after treatment in the GUSTO study, t-PA produced a net increase of one disabling non-fatal stroke per 1000 patients treated, as compared with the rate with SK. The authors investigate the impact of stroke on medical costs and overall reduction in the life expectancy for the t-PA group in a sensitivity analysis (see below). This has a modest impact on the cost-effectiveness ratio for t-PA. The impact could be greater if the excluded costs were considered.

5. Were costs and consequences measured accurately in appropriate physical units?

__X__ YES _____ NO _____ CAN'T TELL

The use of medical resources during the initial hospitalization was measured for all the US subjects (23,105) in the GUSTO study. Resource use to one year was estimated for a random sample (2600) of the surviving members of the cohort. This was done by telephone survey. (These data are given in Table 2, p.1420.)

Many of the data presented in Table 2 are medians. This is appropriate for testing statistically, as the data are negatively skewed. However, for descri

e purposes, means are more informative, as in planning (say) hospital bed ovision, it is important to know about the tail of the distribution. However, e resource consumption in the first year was generally similar for the two oups, so did not enter in the incremental cost-effectiveness ratio.
Resource use beyond one year was assumed to be the same for both groups. ithough telephone surveys may not be the ideal method of determining source use, practical considerations would seriously limit other methods. verall, the authors provide an accurate measurement of resource use.
The main consquence of therapy (life years gained) was estimated by the ethods described above.

Were costs and consequences valued credibly?

(for consequenses) (for costs)
__X__ YES _____ NO __X__ CAN'T TELL

Health state preference values were measured in structured telephone interiews one year after treatment by the time trade-off method (p.1419). Thus, the uality of life impact of some morbidity, particularly that caused by strokes, may ave been missed if these patients were unable to complete the interview. lowever, it would be unlikely to have a major impact on the result of the study.
The unit costs (prices) used in the analysis are given in Table 1 of the paper. These were a mixture of costs (from Duke University Hospital) and charges Medicare DRG reimbursement rates). Two approaches were used to stimate the costs of the thrombolytic drugs, since these had a major impact on the results.

7. Were costs and consequences adjusted for differential timing?

__X__ YES _____ NO _____ CAN'T TELL

Both survival and costs were discounted at an annual rate of 5 per cent (p.1419). The authors state that this is consistent with conventional practice.

8. Was an incremental analysis of costs and consequences of alternatives performed?

__X__ YES _____ NO _____ CAN'T TELL

All the cost-effectiveness ratios presented in the paper are incremental of t-PA relative to SK. The primary analysis gave an incremental cost per year of life saved of $32,678.

9. Was allowance made for uncertainty in the estimates of costs and consequences?

__X__ YES _____ NO _____ CAN'T TELL

A number of one-way sensitivity analyses were performed. These included varying survival and costs in both the short and the long-term for the t-PA group, such as:

- taking the 95% confidence interval for the 1.1 per cent increase in one year survival, as estimated in the clinical trial;
- reducing life expectancy or assuming that the survival curves converge after one year;
- assuming that the (non-significant) increase in one year costs of t-PA (excluding the cost of the drug) did truly exist;
- assuming that the non-significant increase in costs was maintained beyond one year;
- assuming that the price for t-PA was much lower (e.g. the same as a typical European price).

In addition, sensitivity analyses were conducted on the impact of disabling strokes, and leaving costs and consequences undiscounted.
Finally, the quality-of-life adjustment is presented as a sensitivity analysis on the cost-effectiveness ratio. The incremental cost-effectiveness ratio changes from $32,678 per life-year gained to $36,402 per QALY.
In general, the authors highlight the changes in assumptions that would cause the incremental cost-effectiveness ratio to rise above a threshold of $50,000 per life-year gained.

10. Did the presentation and discussion of study results include all issues of concern to users?

__X__ YES _____ NO _____ CAN'T TELL

The authors include a fairly full discussion of the results. They introduce the notion that the upper limit for an acceptable cost-effectiveness ratio remains controversial, but values of more than $100,000 per year of life saved are generally considered too high (p.1422). Most estimates of the incremental ratio are less than $100,000.
The authors state that their subgroup analyses should be interpreted cautiously. A major finding, which perhaps may seem counter-intuitive, is that the incremental cost-effectiveness ratios for treatment by t-PA are more favourable for the older patients (especially those above 60 years old). This is

because the younger patients have the lowest one-year mortality rates and the smallest increases in survival due to treatment with t-PA. The clinical and social implications of this finding are not explored.

Another issue, not explored in the paper, is whether the findings are generalizable beyond the trial and transferable to other settings. For example, the accelerated use of t-PA may not be achievable in some hospitals. In addition, other papers commenting on the GUSTO study have pointed out that the US patients (56% of total recruitment) were managed differently from the non-US patients in a number of ways, including greater use of invasive revascularization such as PTCA and CABG, and greater use of non-protocol medications (van der Verf *et al. 1995)*.

This could lead to differences in costs and, in addition, the mortality reduction with t-PA was greater in the US (1.2% absolute decrease versus 0.7% elsewhere). However, the test for treatment-by-country interaction was not significant.

Finally, the costs of resources could vary from place to place, so overall, it may not be wise to assume that similar results would be found in other locations.

REFERENCES

Black, W. C. (1990). The cost-effectiveness plane: a graphic representation of cost-effectiveness. *Medical Decision Making*, 10(3), 212–15.

Boyle, M. H., Torrance, G. W., Horwood, S. P., and Sinclair, J. C. (1982). *A cost analysis of providing neonatal intensive care to 500-1499-gram birth-weight infants*, Research Report No. 51, Programme for Quantitative Studies in Economics and Population. McMaster University, Hamilton, Ontario.

—, Torrance, G. W., Sinclair, J. C., and Horwood, S. P. (1983). Economic evaluation of neonatal intensive care of very-low-birth-weight infants. *N. Engl. J. Med.* 308, 1330–7.

Drummond, M. F. (1981). Welfare economics and cost–benefit analysis in health care. *Scottish Political Economy*, 28, 125–45.

Evans, R. G. and Robinson, G. C. (1980). Surgical day care: measurements of the economic pay-off. *Can. Med. Assoc. J.*, 123, 873–80.

Hull, R., Hirsh, J., Sackett, D. L., and Stoddart, G. L. (1981). Cost-effectiveness of clinical diagnosis, venography and non-invasive testing in patients with symptomatic deep-vein thrombosis. *N. Engl. J. Med.*, 304, 1561–7.

—, —, —, — (1982). Cost-effectiveness of primary and secondary prevention of pulmonary embolism in high-risk surgical patients. *Can. Med. Assoc. J.* 127, 990–5.

Neuhauser, D. and Lewicki, A. M. (1975). What do we gain from the sixth stool guaiac? *N. Engl. J. Med.*, 293 (5), 226–8.

Russell, I. T., Devlin, H. B., Fell, M., Glass, N. J., and Newell, D. T. (1977). Day-case surgery for hernias and haemorrhoids: a clinical, social and economic evaluation. *Lancet*, i 844–7.

Sackett, D. L., Haynes, R. B., Guyatt, G. H., and Tugwell, P. (1991). *Clinical epidemiology: a basic science for clinical medicine*. (2nd edn). Little Brown and

Stason, W. B. and Weinstein, M. D. (1977). Allocation of resources to manage hypertension. *N. Engl. J. Med.*, 296, 732–9.

Stoddart, G. L. (1982). Economic evaluation methods of health policy. *Evaluation and the Health Professions*, 5, (4), 393-414.

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

van der Werf, F., Topol, E. J., Lee, K. L., Woodlief, L. H., Granger, C. B. *et al.* (1995). Variations in patient management and outcomes for acute myocardial infarction in the United States and other countries. *J. Am. Med. Assoc.*, 273, 1586–91.

Weinstein, M. C. (1981). Economic assessments of medical practices and technologies. *Medical Decision-Making*, 1 (4), 309–30.

and Stason, W. B. (1977). Foundations of cost-effectiveness analysis for health and medical practices. *N. Engl. J. Med.*, 296 716–21.

Weisbrod, B. A., Test, M. A., and Stein, L. I. (1980). Alternatives to mental hospital treatment, II. Economic benefit-cost analysis. *Arch. General Psychiatry*, 37, 400–5.

Williams, A. H. (1981). Welfare economics and health status measurement. In *Health, economics and health economics* (ed. J. van der Gaag and M. Perlman), pp. 271–81. North-Holland, Amsterdam.

because the younger patients have the lowest one-year mortality rates and the smallest increases in survival due to treatment with t-PA. The clinical and social implications of this finding are not explored.

Another issue, not explored in the paper, is whether the findings are generalizable beyond the trial and transferable to other settings. For example, the accelerated use of t-PA may not be achievable in some hospitals. In addition, other papers commenting on the GUSTO study have pointed out that the US patients (56% of total recruitment) were managed differently from the non-US patients in a number of ways, including greater use of invasive revascularization such as PTCA and CABG, and greater use of non-protocol medications (van der Verf *et al. 1995)*.

This could lead to differences in costs and, in addition, the mortality reduction with t-PA was greater in the US (1.2% absolute decrease versus 0.7% elsewhere). However, the test for treatment-by-country interaction was not significant.

Finally, the costs of resources could vary from place to place, so overall, it may not be wise to assume that similar results would be found in other locations.

REFERENCES

Black, W. C. (1990). The cost-effectiveness plane: a graphic representation of cost-effectiveness. *Medical Decision Making*, 10(3), 212–15.

Boyle, M. H., Torrance, G. W., Horwood, S. P., and Sinclair, J. C. (1982). *A cost analysis of providing neonatal intensive care to 500-1499-gram birth-weight infants*, Research Report No. 51, Programme for Quantitative Studies in Economics and Population. McMaster University, Hamilton, Ontario.

—, Torrance, G. W., Sinclair, J. C., and Horwood, S. P. (1983). Economic evaluation of neonatal intensive care of very-low-birth-weight infants. *N. Engl. J. Med.* 308, 1330–7.

Drummond, M. F. (1981). Welfare economics and cost–benefit analysis in health care. *Scottish Political Economy*, 28, 125–45.

Evans, R. G. and Robinson, G. C. (1980). Surgical day care: measurements of the economic pay-off. *Can. Med. Assoc. J.*, 123, 873–80.

Hull, R., Hirsh, J., Sackett, D. L., and Stoddart, G. L. (1981). Cost-effectiveness of clinical diagnosis, venography and non-invasive testing in patients with symptomatic deep-vein thrombosis. *N. Engl. J. Med.*, 304, 1561–7.

—, —, —, — (1982). Cost-effectiveness of primary and secondary prevention of pulmonary embolism in high-risk surgical patients. *Can. Med. Assoc. J.* 127, 990–5.

Neuhauser, D. and Lewicki, A. M. (1975). What do we gain from the sixth stool guaiac? *N. Engl. J. Med.*, 293 (5), 226–8.

Russell, I. T., Devlin, H. B., Fell, M., Glass, N. J., and Newell, D. T. (1977). Day-case surgery for hernias and haemorrhoids: a clinical, social and economic evaluation. *Lancet*, i 844–7.

Sackett, D. L., Haynes, R. B., Guyatt, G. H., and Tugwell, P. (1991). *Clinical epidemiology: a basic science for clinical medicine*. (2nd edn). Little Brown and

Stason, W. B. and Weinstein, M. D. (1977). Allocation of resources to manage hypertension. *N. Engl. J. Med.*, 296, 732–9.

Stoddart, G. L. (1982). Economic evaluation methods of health policy. *Evaluation and the Health Professions*, 5, (4), 393-414.

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

van der Werf, F., Topol, E. J., Lee, K. L., Woodlief, L. H., Granger, C. B. *et al.* (1995). Variations in patient management and outcomes for acute myocardial infarction in the United States and other countries. *J. Am. Med. Assoc.*, 273, 1586–91.

Weinstein, M. C. (1981). Economic assessments of medical practices and technologies. *Medical Decision-Making*, 1 (4), 309–30.

and Stason, W. B. (1977). Foundations of cost-effectiveness analysis for health and medical practices. *N. Engl. J. Med.*, 296 716–21.

Weisbrod, B. A., Test, M. A., and Stein, L. I. (1980). Alternatives to mental hospital treatment, II. Economic benefit-cost analysis. *Arch. General Psychiatry*, 37, 400–5.

Williams, A. H. (1981). Welfare economics and health status measurement. In *Health, economics and health economics* (ed. J. van der Gaag and M. Perlman), pp. 271–81. North-Holland, Amsterdam.