Measurement of costs and benefits in health and health services

K. G. WRIGHT

From the Institute of Social and Economic Research, University of York

This paper is about the principles and practice of the measurement of costs and benefits. It assumes that efficiency in resource allocation (that is, obtaining the maximum possible benefit from the resources available) is a generally accepted national objective, and that measurement of costs and benefits is required to assess its achievement. The paper also takes for granted that the case for the cost-benefit approach has already been accepted; therefore it focuses on the measurement problems. In other words, it is to be seen as an analytical rather than a proselytising exercise.

Two major sections follow immediately after this introduction. The first deals with the measurement of benefits. This logically precedes the discussion of costs which, as will be explained, are simply benefits foregone. In this section there are some preliminary discussions about the purposes and criteria of benefit measures before the actual measurement problem is tackled. The discussion on costs follows and this concentrates on the need for a specific context in which to carry out costing exercises. The last section looks towards future developments and the potential provided by the cost-benefit approach for fruitful interdisciplinary collaboration.

The Measurement of Benefits

Definition

Much time could be spent discussing the different meanings given to the terms benefits, output, and effectiveness in economic and other literature on evaluation, planning, and control of public services. Here, however, these terms are treated as being synonymous. The benefits of activities or policies are taken to be their effectiveness in meeting stated objectives.

Purposes

Measures of benefit serve two principal purposes: (1) to help evaluate alternative courses of action in planning policies and (2) to set targets for the monitoring and control of performance.

Planning and control take place at several different levels in health care policies. The consultative document on priorities (Department of Health and Social Security, 1976) illustrated this point. It was concerned with the objectives of general health policy, with the objectives of major choices in the type of care to be given (notably in terms of prevention versus treatment), with policies for specific age groups (for example, children, the elderly) or for specific disease groups (for example, the mentally ill, the mentally handicapped), or with major groups of services (for example, primary care, acute hospitals). There were related sub-objectives for each of these major groups (for example, to maintain the independence of the chronically disabled, to reduce hospital waiting lists). It is easy to continue this process by identifying objectives below this level (for example, to determine the optimum mix of domiciliary services for the chronically disabled, to reduce length of stay in hospitals), but the major implication is that benefit measurement can also be subdivided in a similar way; that simple measures can be used at the subordinate level but that the degree of complexity of measurement will increase through the hierarchy of objectives.

However, it is tempting, but misleading, to use measures developed for the assessment of achieving objectives at one level to measure the benefits at another. It is easy, for example, for workload figures such as nurse/patient ratios or social casework loads to be used to assess the quality of work achieved, and for some magic number to become a policy objective, although it is doubtful whether workload statistics in health services have reached the hallowed heights achieved by the student/teacher ratio in education. Similarly, incidence and prevalence statistics for various diseases can be used to measure the effectiveness of some policies, for example, immunisation programmes. But, setting aside the difficulties of collecting morbidity data, these would not help in measuring the effectiveness of alternative programmes which produced a range of changes in the incidence of different diseases. This is because the collection of incidence and prevalence statistics is a matter of counting phenomena or events, and decisions on the allocation of resources to the prevention of disease...
x as against disease y require measurement or judgement about the relative severity of these two conditions. The rest of this section deals with the problems of measuring severity.

Criteria

Measures of benefit have to meet the same criteria as any other measure. That is, they must be:

(i) **Valid**, in that they measure the properties which they purport to measure.

(ii) **Reliable**, in the sense that different people measuring the same condition reach the same answer and that different subjects who are in the same state give equivalent measures.

(iii) **Sensitive**, in that the measure shows change when the condition being measured changes.

Validating measures is always a tricky exercise, but it is especially complex in the measurement of the benefits of health care because of the problems of defining the properties to be included. Much of the detailed discussion of properties of benefit measures is dealt with in the subsequent section on the dimensions of benefit measures, but it is useful here to touch on some general problems.

One of these is to keep measurement simple enough to allow information to be collected routinely, because most of the forces are pulling in the opposite direction. Most health care activities serve several objectives, and, unless the achievement of all these objectives is tested, the benefit measure will be incomplete and therefore not valid. However, the efforts to improve the coverage of a measure serve only to store up more problems, because the next stage in benefit measurement will be to aggregate all the different aspects identified into a single, comprehensive measure. The more aspects or dimensions the benefit measure has to contain, the greater will be the problem of aggregation.

Simplicity of measurement is important not only for the collection of information but also for its subsequent intelligibility. Because these measures will be used for general decision-making they must be intelligible to the layman. This does not mean that sophisticated methods cannot be used in developing the measure, but it does mean that at the end it must be possible to give some general intuitive explanation of the results.

With these basic warnings in mind, the next step is to see how measures of benefit can be developed. This development is tackled in three main stages: (1) identifying the basic dimensions of health care benefits; (2) measuring along a single dimension; and (3) combining separate dimensions into a composite measure.

**IDENTIFYING THE RELEVANT DIMENSIONS**

There are two important, related problems in identifying the major dimensions of a measure of benefit for health care policies. One is principally academic and is concerned with the concept of health which is to form the main basis for benefit measurement. The other is mainly political and is concerned with the identification of the major objectives of health policy. These two problems are related because different concepts of health imply different objectives, and therefore have a fundamental effect on benefit measurement and the allocation of resources. Unfortunately, while there has been much useful debate in academic circles on the various concepts of health, the political debate on objectives has been shallow and sporadic. Thus, it is impossible to use stated objectives to plumb into the identification of the major dimensions of a benefit measurement. Instead, the following pattern has been imposed on this section of the paper. Firstly, the different types of concepts of health are baldly summarised to show the general problems which benefit measurement must take into account if it is to be comprehensive. Secondly, the different concepts are set against the other criteria for benefit measurement and an attempt is then made to show why particular concepts have been more widely used than others.

Most of the debate on concepts of health is to be found in the medical sociology literature. This has recently been surveyed and summarised (Stacey, 1977) so a brief summary of the conclusions is all that is needed here. Stacey identified several ways of looking at health:

(i) **Health as an ideal or absolute concept.** An optimal level of health for which it is possible to strive but impossible in practice to attain. The relevant optima could be defined in several ways—biological/physiological functioning of the body or the social functioning of the individual.

(ii) **Health as a relative concept.** Health can be defined only in relation to illness and disease.

(iii) **Health as a capacity.** Health is an important determinant in a person’s ability to function socially. This capacity to function is important for maintaining individual life styles as well as for the equilibrium of the social system.

(iv) **Health as welfare.** Health care is for the incurable as well as the curable.

The economic contribution to the debate occurs mainly in the literature on human capital. The most naïve formulation is that the benefits of health care can be seen mostly in terms of the maintenance of
wage-earning production. A much wider implication, that investment in health enables the individual to produce many things other than wage-related production, such as the benefits from leisure-time interests, has been set out in a theoretical and empirical investigation into the demand for health (Grossman, 1972). This notion of the maintenance of health as an investment in human resources is clearly related to the capacity concept of health and the ability to function in a variety of social roles.

Many of the implications of these different concepts for priorities and the allocation of resources have also been examined in discussions of the definition of need and the allocation of health care (Williams, 1974; Mooney, 1977a). These have yielded some very interesting points about the definitions of health and ill-health and about who is to decide which treatment, if any, should be given to which individuals, but they are largely peripheral to this paper, although some of the ground will be covered in a later section.

Objectives and measures of benefit differ according to the concept of health. Concepts based on biological or physiological functioning, whether in absolute or relative terms, imply different types of measures from those based on health as welfare or health as a capacity for social functioning. Measures of a very scientific kind will create difficulty in meeting the criteria set out earlier. Firstly, these measures are intelligible only to the specifically trained, and will therefore tend to handicap both the technical interdisciplinary discussions and the political debate on objectives and policy implications. Secondly, such measures would be difficult and expensive to use as a matter of routine. Thirdly, there would be many different measures to collect for any individual and the subsequent aggregation problem would be horrendous. Fourthly, they will tend to concentrate on the physiological and neglect the social and emotional factors related to health. Fifthly, there will be problems related to the variety of the measures, in so far as it will be hard at times to decide what is normal or abnormal functioning, and the effect of this on the individual.

The capacity concept will also present several difficulties. Again, the focus on social functioning may neglect the emotional and social factors which do not interfere with performance. There will be a problem of validity, too, unless it is clear whether the capacity concept is related to ability, to function, or to actual functioning. A capacity notion suggests the idea of ability, but there are likely to be differing views of how a person should function and what is managed in practice. It has been shown that performance can vary according to environment and observer (Kelman and Willner, 1962) and this poses serious problems for reliability tests.

However, the capacity concept based on actual social functioning has some basic advantages over the physiological approaches. Firstly, it is intelligible to all the people who are likely to be involved in decision-making in health policy. The measurement discussed below is in terms of how people manage their daily routines and it concentrates on the ways in which health or ill-health affects them. Thus, although many people would not comprehend the significance of measurements of their blood pressure, they would understand that an abnormally high reading could interfere with their mobility by preventing them from walking far without feeling giddy or out of breath. Secondly, the capacity concept covers many different diseases or biological conditions and it therefore presents fewer problems of aggregation, although it does not avoid them.

A step can be taken towards resolving arguments about the relevance to benefit measurement of the different concepts of health if it can be shown that there is a relationship between them. It seems intuitively plausible that there is a correlation between physiological functioning, the presence and progress of disease, and the ability to carry out the usual daily activities. This relationship has been empirically established in some studies (Katz et al., 1963). In other cases it would be interesting to speculate on the lack of such relationships: for example, when social functioning might deteriorate although an individual might appear very fit and no illness could be diagnosed. The rest of this section concentrates on two conclusions that can be drawn from the discussion on concepts of health. These are that (1) the social functioning approach is most likely to meet the criteria set out earlier and (2) the concept of health as welfare will be important for certain groups of patients who are chronically or terminally ill. Are these conclusions reinforced by the debate on policy objectives and the literature on benefit measurement?

In the United Kingdom, the defining of objectives for health policy began only recently with the publication of the ‘priorities documents’ (Department of Health and Social Security, 1976; 1977). So it is hardly surprising that it has proved difficult to provide major objectives for health policy other than ‘to meet the health needs of the community’, a meaningless phrase as a guide to policy or benefit measurement. As stated earlier, a perusal of the ‘priorities documents’ at least reveals some objectives for particular aspects of policy and for specific groups and services. There is in practice a varied set of concepts in operation. Some are based on the disease or medical model
(for example, prevention and treatment of illness), some on the capacity concept (for example, maintaining or improving the ability of the mentally ill, mentally handicapped, chronically sick and elderly to live within the community), and some on welfare (for example, provision of care for the elderly severely mentally infirm). However, it would not be true to say that one dominant concept was being used in this process.

The literature on benefit measurement reveals two major dimensions. One is concerned with the performance of daily activities in terms of mobility, ability for self-care, and ability to work and pursue usual leisure interests. The second dimension has several aspects which might be summarised as distress; these also include such properties as pain and suffering. Where there has been more detailed discussion of policy objectives it is possible to see that the number of dimensions increases, and work on the care of the elderly (Wright, 1974) has shown that there are important dimensions such as the relief of loneliness and stress to add to the two major dimensions.

Most psychological aspects of health care such as the relief of suffering have proved very difficult to measure, although work on hospital outputs by Rosser and Watts (1972) did attempt to order and scale pain and distress as well as severity of physical disability. The rest of this section is concentrated on physical disability because data are available to illustrate relevant points.

Measuring along a single dimension

In creating a scale to measure even a single property such as general mobility, one of the first problems is to fix its two end points. In existing studies, one end is frequently fixed at 'death', but current discussions on the effectiveness of life-support systems seem to encourage the view, by no means novel, that there are states worse than death. The dilemma at the 'good' end of the scale is whether 'good' functioning should be defined in relation to the best levels of functioning of appropriate peer groups or in relation to an individual's usual level of functioning.

The points at issue can be illustrated by using the conceptual framework for the measurement of social indicators developed at York (Culyer et al., 1971). Discussion of the benefits of health was based on Fig. 1, where the intensity of health is plotted against its duration. The benefits of a health treatment were taken to be the reductions it achieved in intensity and/or duration. Thus in Fig. 1, the end point O is fixed with reference to best levels of functioning in the community, and the shaded area represents the benefit to a person whose state of health without treatment would have followed the time path ABCD and whose state of health with treatment followed the time path ABCE. Someone who was in a better state of health at O at time OH, who has the same prognosis for the 'without treatment' course of action, and who is actually restored to normal functioning at time OG, receives the additional benefit represented by the cross-hatched area below EF. But if the end point had been at A on the intensity axis, this additional benefit would not have been recorded. Similarly, no benefits at all would be assigned to a treatment which kept in the best state of health a person who, without that treatment, might have followed the time path HBEG.

![Fig. 1 Changes in health status over time.](image_url)

The propensity to underestimate benefits in these examples arises from equivalencing several different states between OA into the one state represented by A on the intensity axis. However, if the end point were to be fixed by reference to an individual's normal level of functioning, which has the same effect as equivalencing all the points between O and A on the intensity axis, different measures of benefit would be derived. Using the time paths of Fig. 1 again, point O is described as 'carrying out usual work and recreational activities'. Two people with different occupations, say a clerk and a professional footballer, both suffer a leg injury at time OH. The clerk returns to his normal social functioning along the path HBCEG after suffering some period of disability. However, the professional footballer's career is ruined and he has to take up a career as a clerk. He is no longer achieving his usual social functioning and he returns along the path HBCEF. Thus, the benefit to the footballer is less than to the clerk. Benefit maximisers will prefer to treat clerks rather than footballers, restoring the relatively unfit to their usual state of unfitness rather than the usually super-fit to average levels of fitness.

The beginning of a scale is a special part of the sensitivity argument. The extra benefit represented...
by the area below EF in Fig. 1 may not be important when large groups of the population are considered, but may be important when looking at specific conditions. For example, the treatment of low back pain may have little effect on the ability of an office worker to remain at work but will cause some manual workers to lose time. The more points a scale contains the greater is its sensitivity to change. There are times when it will be important to pick up fairly small changes in functioning. For example, if two surgical treatments are being compared, although both treatments restore patients to the same level of functioning at the end of a short time period, changes of functioning within that period will affect the choice of the more effective treatment. In other words, the sensitivity of a measure will be related to the purpose it will serve and to how seriously small fluctuations in the measure will affect the decision to be made.

Restriction of activity in terms of general mobility and self-care has some advantages for measurement in that there is an apparent intuitive progression in severity. Not being able to get out of bed is worse than not being able to walk round a room which in turn is worse than not being able to walk far out of doors. Health care personnel who are used to assessing people’s ability to cope in various situations can order the severity of restriction of activity with some common agreement (for example, Wright, 1974; 1978).

This professional ranking of severity can be tested in two ways. One is by the study of the loss and regain of functions during illnesses to determine whether a particular order is established, for example, on the lines of the work of Katz et al. (1963) on biological and social functioning. The second method is to use scaling techniques, the most common of which has been the Guttman scaling approach (Williams et al., 1976).

The setting up of an ordinal scale is important in the planning and control of health care activities. Such scales are useful in evaluative studies to follow the severity of a condition, in planning resource use to give an indication of case severity which is important for staffing policies, and for setting and monitoring the achievement of performance targets. Therefore they are to be welcomed, although their applicability is rather limited. For example, if a measure of benefit has more than one dimension, judgement is required about whether or not benefit has increased when dimensions show contrary movements. Again, in measuring along even a single dimension, ordinal measurement says only that state A is better than B which is better than C, and so on. It says nothing about the relative worth of moving from A to B to C. Unfortunately, this is the type of information required for deciding on priorities for allocating resources. Therefore, it is necessary to see how interval or ratio scales might be developed to meet this requirement.

COMBINING DIMENSIONS

Attention has been drawn to problems of value in the choice of appropriate concepts of health, the articulation of policy objectives, the identification of the dimensions of measures of benefit, and the methods of measuring along one dimension. However, the methods of combining different dimensions pose important questions about what is to be valued, who is to establish the values, and which methods are to be used.

(i) What is to be valued?
The York model used in Fig. 1 (Culyer et al., 1971) was explicit about what was to be valued. ‘... it is important to stress that society’s judgements concerning the relative importance of avoiding one state rather than another are represented by the actual numbers attached to each specifically, for example, state 2 is twice as bad as state 1, and state 10 is ten times as bad. This implication must not be shirked, and must be regarded as a statement about health policy not a technical statement about medical condition’.

This model also contained a ‘duration’ dimension which is of great importance to the benefits expected to be received from following different courses of action. The expected time paths of changes in the state of health gave a probabilistic aspect to the model. ‘For instance recovery from a particular disease will follow one time-path (incorporating both intensity and duration) in 90% of the cases, another in a further 9% and yet another in the remaining 1%’.

The measurement of reductions in the probability of changes in health state, particularly the valuation of small changes in the risk of death, has a special literature of its own (Mishan, 1971; Jones-Lee, 1976; Mooney, 1977b) and it would be impossible justly to summarise this work in this paper. However, many of the basic issues about different valuation methods are taken up in the following sections.

(ii) Who is to establish the values?
The general concern here is to choose between alternative approaches. Is the valuation of changes in states of health, the risk of such changes, or the relative worth of states of health to be undertaken by individuals in the sense that the total value will be the sum of individual values, or should such decisions be left in the hands of ‘experts’ or of the
relevant political representative(s) of the community? The literature on the value of life has enumerated three groups who would be interested in changes in states of health:

(a) The people who are directly affected.
(b) The people who are not directly affected but who suffer some financial loss because of a decrease in the state of health of someone in group (a).
(c) The people who are not directly affected but who suffer some psychic disbenefits because of a decrease in the state of health of someone in group (a). Groups (a) and (b) may coincide.

Thus, the drawback in using individual valuations is that the aspects of (b) and (c) will be forgotten and the sum of individual valuations will understate the total benefits.

In addition, there is a problem of how individual values are to be elicited. In some cases it may be possible to infer these from people's behaviour, and this has been shown to be possible in studies of the value of life, but it is more difficult for changes in state of health which are of less dramatic significance.

One problem in Britain is that many services are provided free of charge, and, although it may be possible to identify some sort of willingness of the consumer to pay in terms of the expenses incurred in time spent travelling and receiving treatment, it is difficult to measure all the costs incurred. Possible solutions would be either to ask individuals how they would react to certain hypothetical situations and infer values from this, or to set people the task of scaling or valuing various changes in state of health.

Consumer ignorance is one of the difficulties of using individual valuations. There can be no doubt that decisions on personal health are difficult for the layman to make. If they are too difficult, valuation passes into the hands of the expert, and leaves not just the technical valuation problems but also the whole problem of social responsibility. Who are these experts and to whom are they answerable? A possibility is that these experts should form a ministerial advisory panel on the valuation of changes in states of health, and leave the ultimate decision with the responsible politicians. Many objections could be raised to this idea. The valuation could become part of the dark area of ministerial and cabinet decision-making on which no light falls until 30 years after the event. Ministers could bend recommendations to allow pet schemes to take priority. Governments are elected on whole programmes; it would be difficult for society to change individual valuations of this type once they had been agreed by the Government.

In recounting all the difficulties in setting up these values of changes in health state, it is easy to forget that valuation is taking place all the time in the health care system. It occurs at all levels in the organisation: at the clinical level of doctor and patient (how much of doctor X's time is patient Y worth? Is it worth spending more resources on keeping patient Z alive?); at the level of the district and area management teams (should we spend more on the acute wards or on the geriatric wards?); and at central government level (we will spend £n more on treating renal failure). Many decisions have consequences which imply relative values, and it would be possible to study decisions in this framework and test the consistency of the values that have been derived (Mooney, 1977a). This approach has important consequences for making explicit the decision-taking and valuation process in this field, and for highlighting absurd inefficiencies in the allocation of resources. But with this approach, too, there is the need to guard against the problems mentioned in the previous paragraph.

(iii) Which methods of valuation?

In this section valuation is taken to mean the development of interval or ratio scales for both multi- and uni-dimensional scales, and the use of money as a common unit of account for such scales. Likely methods of developing scales were set out nine years ago in one of the first articles on the development of measures of states of health (Fanshel and Bush, 1970). On the whole these methods centred around the use of paired comparisons of various states of health. This could be done in several ways which were subsequently tested in experimental conditions (Patrick et al., 1973):

(a) The category method. Assigning numbers to states such that differences in the number were equal to the subjective difference in the states being compared; that is, if state A were given the number 2 and state B the number 4 and state C the number 6, then the interval between each of them is equal to 2 and they represent equal changes in states of health.

(b) The magnitude method. A ratio measure obtained by giving the judges a standard score for a healthy day and asking them to rate the description of a state of health in terms of whether it was worth a half, a quarter, a tenth, etc., of the standard day.

(c) The equivalence method. In this the judges were asked to consider for any descriptive state of health how many people of the same
Measurement of costs and benefits in health and health services

age were equivalent to a group of 100 in the (standard) maximum state of health.

Although these valuations were obtained from groups of people with health service or medical training, it was concluded that the category and magnitude scales were valid and reliable enough to be tested with other groups in the population. Other experiments of this type were reported by Culyer (1978):

(d) The time trade-off method (Torrance, 1970). Judges were asked to scale states of health in terms of the amount of time in a healthy state considered to be equivalent to some lengthier amount of time in a less desirable state.

(e) The standard gamble method (Woolfson, 1974). It is only natural that economists, who have been discussing the relative merits of ordinal and cardinal measurement of utility for many years, should attempt to apply one of the methods developed for cardinal measurement to the benefit measurement experiments. The method used the following equation

\[ x = py + (1-p)z \]

and each judge was asked to choose between the certainty of obtaining \( x \) against a gamble between \( y \) and \( z \) such that he would get either \( y \) or \( z \) with some probability, but not both. The judge was then asked to choose a probability for \( y \) which makes him indifferent to the choice between the gamble and the certainty. Woolfson used physicians as judges to choose between \( x \), the usual course of a disease with treatment, and \( y \), a remedy which will provide either an instantaneous cure with a probability of \( p \) or death. The judges therefore rank the severity of a condition by the value they give to \( p \) at which they are indifferent in their choice between this relative probability of cure or death and experiencing the usual suffering associated with condition \( x \) and its treatment. The value of \( p \) would be expected to decrease as the severity of \( x \) increases.

Apart from the point that this work again relies on medical expertise, it has two other problems. One is that it may be measuring risk aversion as well as relative severity of states of health. Secondly, it is disease-specific, which makes it much more difficult to use in the allocation of resources to various target groups.

The development of state of health indices is an important advance towards benefit measurement. If it were possible to develop ratio scales, points could be summed and resources could be allocated on the grounds of achieving the most 'health points' per pound spent. Decisions on priorities would then be made in a more informed and explicit manner. However, even this amount of progress would not satisfy those concerned with allocating resources not only within the health and personal social services but also between these services and, for example, education or defence or transport. They would require information about the net benefits (that is, total benefit less total cost) of policies, which means translating benefits into monetary terms to make them commensurate with costs.

If it is true that the last post has sounded for the lost output approach and benefits are not to be cast in terms of lost earnings from employment, there are likely to be few attempts to quantify benefits in terms of money. There has been one notable attempt (Rosser and Watts, 1972) to produce a scale using court awards for injuries as the measure of severity (that is, the judges were judges), but the development of an 'exchange rate' between cash and points on a health index is still awaited.

Summarising progress so far in benefit measurement, it is possible to say that:

(a) Useful progress has been made in developing ordinal scales for disability.

(b) Progress in measuring along a dimension which measures the achievement of objectives to relieve distress, pain, or suffering has not matched the work on disability.

(c) Really encouraging results have been achieved in multidimensional scaling in experimental conditions, but further results are awaited from extending the exercise to more general groups in the community.

(d) The debate on the valuation of life has developed a good methodological framework for the measurement of benefits and has shown the wide range of values of life that can be inferred from either public or individual decision-making.

THE MEASUREMENT OF COSTS

Definition

Economists define costs as foregone benefits. Resources have alternative uses and their cost in use in one activity is the benefit they would have produced in the next most beneficial use. From this simple conceptual notion flows a whole series of practical measurement problems concerned with the identification and valuation of all the resources used in the activities in question.
Purposes
The measurement of costs, like the measurement of benefits, serves important uses both for the planning and evaluation of policies and for the financial control of services. This paper concentrates on the first of these purposes, although most of the cost information that is readily available is designed for the second, so that there are severe difficulties in using it in planning and evaluative studies. Evaluative studies are rarely concerned with the existing cost of an activity. Rather, they pose questions of the following type:

(i) What would be the cost of providing more of x or what costs would be saved if less x were to be provided?
(ii) What is the cost of x now compared with what it used to be or what it will cost in the future?
(iii) What would be the cost of providing more (less) of x while providing less (more) y?
(iv) What is the cost of providing x in location A compared with location B?

As these questions indicate, the costs of an activity will vary with its scale, location, and time.

Costs and output
The valuation of costs with output is a central problem in evaluative work where the questions set refer to the establishment of new programmes or to changes in programme size. Should effective treatment x be introduced? Should disease y be treated at the post-symptomatic stage or should a screening programme be introduced? Should condition z be treated by surgery or by chemotherapy? These are typical evaluative questions which require the study of relative costs and the variation of costs with output. It is intuitively obvious that increases in output require increases in inputs, but the practical measurement problem is to determine the relationship between costs and output.

Economic theory predicts a variation in the rate of change of output with input as illustrated in Fig. 2.

Firstly, it will be noticed that some costs (OC) are incurred before any output is achieved and that these costs will remain constant over a range of output as illustrated by the line CD. Buildings and capital equipment are typical examples of this type of cost. Other costs, typically labour and materials costs, vary with output right from the start of production. The average cost of a unit of output is represented by the tangent between the abscissa and a line drawn to the relevant point on the total cost curve. The average cost of output OA is the tangent of angle $\alpha = \frac{AB}{OA} = \frac{\text{Total cost}}{\text{Total output}}$.

Inspection will show that at outputs below OA average costs are decreasing and after OA they are increasing. It is this variation in average costs which makes it dangerous to generalise programme costs by multiplying estimated output by average cost. Thus the total cost of output OE is EG not EF which is the cost obtained by multiplying the average cost at AB by output OE. Similarly the cost of output OJ is JL and not JK.

This pattern, in which average costs decline until they reach some minimum point and increase thereafter, is caused by the use over a fixed period of time of variable inputs such as labour with fixed inputs such as buildings and land. It is an example of the famous law of diminishing returns; in other words, it becomes increasingly difficult to increase output by the use of variable inputs alone. An increase in the quantity of the fixed input will ease these difficulties. Over longer time periods, when it is possible to vary the input of capital, land, and labour, this pattern of costs may not apply. Over longer periods average costs may be constant (in which case it would be correct to use the average cost to estimate the changes in the cost of output); alternatively they may decrease or increase. This long-run variation of costs has been and continues to be the subject of important econometric studies of health services.

The variation of costs with time and place
Costs vary with time not only because of the greater flexibility available for changing the mix of inputs but also because the demand and supply and, consequently, the relative prices of inputs change. Relative prices of inputs also change from one locality to another. Thus, the most efficient use of resources may cause changes in methods of production and in costs. This, and the need to measure benefits, is why it is difficult to draw conclusions about cross-sectional and time series
analyses of costs. However, useful comparative work has been done in these difficult circumstances. Feldstein (1967) pioneered the study of hospital costs, and the debate on the appropriate funding of English teaching hospitals has been better informed because of the analysis of the differential costs of teaching and general acute hospitals (Culyer et al., 1978). Nevertheless, the authors of these studies would probably be among the first to acknowledge that conclusions from their work should be drawn with caution.

Theory and practice
These short excursions into elementary economics have very important implications for policy-making. One is the provision of facilities of optimum size, where work has been carried out on the size of general hospitals (Hurst, 1978) and local authority homes for the elderly (Wager, 1972; Davies and Knapp, 1978). Another is the costing of alternative ways of achieving objectives, which is the usual task of cost-benefit analysis. The variation of costs with scale of output is particularly relevant to examining the costs of new activities and changes in the scale of existing ones. However, the analyst rarely knows anything except the costs of the existing output. His task is to sketch in the changes in costs as the size of programmes change, but he is forewarned by the simple economic theory that the use of average cost figures may be misleading. Instead, he has to enumerate the changes in resource use that will be brought about by proposed policy changes and attempt to value them.

Identifying changes in resource use is not in itself an easy task. Valuing them is even more difficult. The existing available capacity and the way it is used determines the effect of changes in resource use. For example, if there is a policy to switch the care of a group of people from hospital to the community, the change in resource use depends on how the resources freed in the hospital sector will be used, if at all, and how the community health and social services can deal with the increased load. In some cases it may be difficult to free resources because the change in workload is very small. For example, only a few patients are discharged from each hospital or each ward. The amount of nursing time saved amounts to only an hour or two per week and it is impossible to switch one nurse to other duties. In some cases the resources are shared by many patients and the numbers being cared for have to be substantially reduced before any resources are freed. Heating and lighting costs can be saved only if it is possible to discontinue the use of buildings or parts of buildings. Similar problems occur in the expansion of the community programme. Although increased numbers of personnel can be used to meet the increased demand for community care, at what point will supervisory and administrative capacity have to be increased and new office accommodation obtained?

Noble efforts have been made to identify the resources used by individual patients, but they have been bedevilled by the problems of distinguishing between direct or individual costs and indirect or shared costs—for example, Babson (1973) and Russell (1974). The direct costs are calculated from resource use which can be ascribed to particular individuals, for example, drugs, dressings, and personal nursing care. Indirect costs are those consumed collectively, such as heating and lighting expenses, general cleaning and portering services, and the general supervisory elements of nursing care. Further problems occur in costing some of the items the patients receive, for example, pathology tests or x-ray investigations, since in Britain and other countries with national health services the costs of individual items are rarely known and their measurement other than in broad terms of average cost is a lengthy process in itself.

This identification of existing and alternative resource use is in line with the doctrine of opportunity cost and it is a necessary prerequisite to the valuation problem. The essence of the resource valuation problem is the distinction between the cash paid for a resource and its opportunity cost. However, before tackling the matter of valuation, it would be advantageous to look at some of the general difficulties in obtaining data on resource use and valuation. Most evaluative and planning studies have to obtain their costing data from the public accounting systems. Unfortunately, these systems are designed more for purposes of financial control than for costing. They concentrate on cash flows and the annual expenditure on large units of administration such as a hospital or its major departments (pharmacy, nursing, pathology, etc.) or the major services provided by local authorities. The low level output indicators such as workload or throughput are often collected for different reasons such as operational control, and refer to different time periods. In the disease costing exercises it has been found difficult to relate hospital costs to individuals, and the same problem occurs in community care where several services might be delivered to an individual. However, the greatest difficulty occurs where treatments involve episodes of hospital inpatient and outpatient care as well as community care, because there are so many different services, with different methods of recording the services delivered to different patients;
it is a Herculean task to link all these records together.

Linkage of records is mainly a practical problem but the reconciliation of cash payments with opportunity costs is an important conceptual problem. By making some simplifications which are not serious in the present context, it can be said that resources will move to their most efficient use and the cash paid for them will be equal to their opportunity costs if their users maximise net benefit and there is free competition between resource owners and producers. In practice, these conditions may not be met because of monopolistic practices by producers and resource owners, and because the government interferes in several ways to distort the free working of markets, for example, by imposing taxation or subsidies on goods and services. Some adjustments of prices may therefore be necessary to equate cash paid to opportunity costs. While this may not be too difficult in deducting some taxes or adding on for some subsidies, there are many instances where it is conceptually advisable to allow for imperfectly functioning markets but in practice it is impossible to do so. Are the wages of nurses artificially low because there is only one major employer and if so what is the correct (opportunity cost) wage rate to use? This is the typical question posed but rarely answered in evaluative studies.

There are two other interesting examples which would tease any cost analyst's brain. One is the use of highly specialised resources the costs of which are sunk once they are committed. Specialised medical equipment and hospital buildings are the typical examples. The use of volunteer labour can be another. The difficulties arise because, although the opportunity cost of these resources to the health care system is zero since they have no alternative use, within the system they can be used to treat or care for different groups. A building could be used as a geriatric or an acute medical ward. A volunteer could be used for child minding or serving meals in the outpatient refreshment room. The use of the resources in one way implies foregone benefits from alternative uses. These foregone benefits provide an alternative, shadow value system for these specialised resources.

At the other extreme there are resources which are often treated as free, although their opportunity cost is probably greater than zero. Patients' time is the frequently quoted example. Time is a scarce resource which has alternative uses and value, although there will be problems of valuing a commodity which is not normally traded except in the sense of choice between earning and the enjoyment of leisure. The expansion of community care will also raise problems of the value of the time of people such as members of the family or neighbours who provide regular support for a patient. Are they to be treated in the same way as the volunteer resources of the previous paragraph, or is this labour given under some form of duress and should its providers be duly compensated? Although there are studies which show how the burden of care increases and the implications of this for service provision (Sainsbury and Grad, 1974; Sanford, 1975), there is a marked absence of the valuation of care which is often given, willingly or perhaps unwillingly, on a continuing basis.

Cost studies therefore require the same degree of intellectual effort as benefit measurement. Although the development of measures of benefit is essential for identifying the alternative values of resources, the allocation and measurement of costs presents some special individualistic problems. These include the allocation of resources to individual patients, the appropriate adjustment of market prices to represent opportunity cost, the valuation of capital, and the valuation of time spent caring for others or receiving treatment.

**PROSPECT AND RETROSPECT**
This paper seems to be replete with difficulties. Problems, pitfalls, worries, dangers, monumental tasks, insuperable hurdles, dire warnings, litter every page. Small wonder that anything is ever attempted, let alone achieved, in the cost-benefit framework. Yet progress has been made. These indications of the problems to be faced warn that solutions cannot be achieved overnight, that specialised research efforts will have to be made to overcome some of the difficulties, and that benefit and cost measurement has to be related to policy and cannot be carried out in splendid academic isolation.

Immediate, useful progress is possible by simplifying the questions to be tackled while recognising the limitations of fairly imprecise answers. A typical hierarchy of work would be:

**Cost-effectiveness analysis**
In cost-effectiveness analysis either costs or benefits are held constant and the choice between alternatives is determined by analysis of the element left variable. If two treatments produce equal benefits, the more efficient is the one with the lower cost. This framework has been applied successfully to a number of decisions in health care, including screening for tuberculosis (Pole, 1971); treatment of varicose veins (Piachaud and Weddell, 1972); treatment of chronic renal failure (Klarman et al., 1968); and hernia repair (Russell et al., 1977).
Measurement of costs and benefits in health and health services

Marginal analysis
Unfortunately many alternatives to be analysed do not fit into the cost-effectiveness framework and produce differential costs and benefits. Given the present state of the art, it may not be possible to make all the benefits commensurable either with each other or with costs, and judgement may be needed to decide whether a more beneficial treatment would be worth the extra costs it entails. This approach allows decision-makers to see the value they are putting on important sets of benefits and how these values could be used in subsequent cost-benefit analyses. Mooney (1977b) has shown how it could be used to obtain implied values of life from public decision-making.

Cost-benefit analysis
The measurement of all costs and benefits is the summit sought by efficiency studies. In many decision-making areas this summit is a long way off, but in some it has been possible to make progress using the available information, as shown by the work on screening for spina bifida (Hagard and Carter, 1976). In some cases, however, it may well be useful to concentrate on separate, specialised methodological improvements of the measurement of benefits and the measurement of costs, in order to overcome some of the general difficulties which are now frustrating most attempts to carry out ambitious cost-benefit work.

Cost-effectiveness, marginal analysis, and cost-benefit analysis require multidisciplinary work. So does the development of the methodologies of benefit and cost measurement. Although it is obvious that benefit measurement will profit from the input of economists, sociologists, social statisticians, physicians, and psychologists, it is not so obvious that the development of costs is not the sole preserve of the economist. Undoubtedly there are important economic problems to be faced in, for example, the measurement of capital costs, the adjustment of market prices to represent opportunity costs, and the development of ‘shadow’ prices for various resources, but there are some important behavioural aspects to be investigated in the supply and use of voluntary or informal labour in community care programmes which would also benefit from multidisciplinary study.

Many people are daunted by the thought that many years of work will be needed to tackle some of the important but arduous tasks involved in measuring costs and benefits. The failure to cover every aspect of measuring costs and benefits is taken as an excuse for doing nothing. So far, the amount of research devoted to the measurement of costs and benefits has been small in relation to the problems faced, yet a good foundation of an accepted conceptual framework has been laid and small scale studies have shown distinct promise for larger scale work. It is to be hoped that in the future greater priority in research funding will be given to concerted, multidisciplinary work on the measurement of costs and benefits and the evaluation of health care policies.

Comment: JEAN WEDDELL
Allied Medical Group; formerly Senior Lecturer in Community Medicine, St. Thomas's Hospital Medical School, London

Planning is essentially a corporate activity, not only between academic disciplines but also between these disciplines, administrators, and the public. For this collaboration to be fruitful, it is essential not only that the economist defines his terms as clearly as Wright's paper suggests he should but also that the planners agree with his definitions and find them useful in practice. The epidemiologist has little difficulty in grasping the concepts and approaches used, but the lay decision-makers-members of area health authorities and politicians-may find them unfamiliar.

The first difficulty arises in agreeing about the need for measures to be valid, reliable, and sensitive. Here the epidemiologist is on home ground and is in total agreement. Administrators, patients, the general public and the politicians will also concur, in general terms. But because one health activity often has several objectives, to achieve validity of measurement it is necessary for the attainment of each objective to be measured and then for all measures to be aggregated to a single measure. This is the ideal, and it is usually unattainable. At this point the decision-makers and the economists need to agree on the priorities, the relative importance of each measure. The clinician, the patient, the administrator and the politician must decide on the priorities of measurement and identify those that matter most and that will exert the greatest influence on the taking of a decision. Here the good judgement of the decision-maker becomes crucial. The judgement needs to be informed, and based on a thorough understanding of the issues. Probably few decision-makers appreciate the limited ability to measure, and the fact that while it may be possible to establish a rank order of ‘betterness’ it is often impossible to measure by how much A is better than B.

Indirect measures may be the only ones available, but to ask an individual how he would react to a hypothetical situation can be seriously misleading. He may be able to give a limited answer to the question itself, yet be unable to appreciate some consequences that may be more important than those he had considered. Had these been appreciated, he might have totally rejected the course of action determined by indirect measurement.

The stumbling block of consumer ignorance is a very real one and the consumer may be not only the patient but also the clinician. I doubt whether many decisions can be delegated to the expert. Who is this paragon?
But it may be possible to form an expert group (not a group of experts) who can work together to agree on benefits to be measured and on the form of these measures, and then use these in decision-taking.

A switch from hospital to community care usually leads to increased costs. The less seriously ill can often be treated at home, given good organisation of all those caring for them. This means that the case mix in hospitals changes: there is a greater proportion of the seriously ill and this is more expensive.

Experiments are valuable that compare home and hospital treatment, or inpatient and outpatient care. If two treatments produce equal benefits, the more efficient is the one with lower costs. But at what point in the natural history of the disease concerned should the results be accepted? After one, three, or five years? At five years it may be apparent that the benefits of the two treatments are not in fact equal, and that the more expensive treatment provides a better clinical outcome in the long term.

Decision-makers need to decide which factors are of over-riding importance and will play the greatest part in their decision-taking, and then ask for these to be measured as accurately and as fully as possible.

Reprints from K. G. Wright, Esq., Senior Research Fellow, Institute of Social and Economic Research, University of York, Heslington, York Y01 5DD.

References


Measurement of costs and benefits in health and health services


