This paper investigates ways in which concepts of model building and cost-effectiveness can be applied to the health planning process. The problem is described and methodological implications drawn that recommend large-scale digital-computer simulation techniques. A study of effectiveness measures suggests that the system's basic purpose is to reduce the likelihood that individuals will be in a state of ill health. The proposed measure of ineffectiveness is defined as the expectation of the weighted sum of the population's expected duration of stay in each of m disability states. The applicability of this and other measures to specific health problems is also discussed. The paper deals with the problem of estimating costs. Distinctions are made among the various kinds of costs and among the various bearers of the costs.

The objective of this paper is to investigate some ways in which the basic concepts of model building and cost-effectiveness can be applied to the health planning process.

The first two sections describe the problem and the approach taken, and define the concepts involved.

The third section outlines the methodological implications that can be drawn and recommends large-scale digital computer simulation techniques as a potentially effective means of applying cost-effectiveness concepts to community health system planning.

The next section, a study of effectiveness measures, suggests that the system's basic purpose is to reduce the likelihood that individuals will be in a state of ill health. The proposed measure of ineffectiveness is defined as the expectation of the weighted sum of the population's expected duration of stay in each of m disability states. The applicability of this and other measures to specific health programs is also discussed.

The fifth section deals with the problem of estimating costs. Distinctions are made among the various kinds of costs, such as: fixed and variable;

*This research was performed under PHS Contract No. 108-26-269. Other related tasks under this contract are: (1) Describe the total community health service system; (2) Identify subsystems which can be profitably modeled; (3) Develop and test a subsystem model (the Maternal and Infant Care model was selected); and (4) Recommend methods of incorporating additional subsystems into the framework of the total system.
sunk (historic) and future; actual (direct) and implied (indirect). Distinctions are also made among the various bearers of the costs: the government, the community, and the individual.

The final section is a summary.

**Scope of the Problem**

Expenditures for health and medical care services in Fiscal 1965 totaled $38.4 billion, or 5.9 per cent of the Gross National Product (GNP); governmental expenditure amounted to almost $10 billion, or 26 per cent of this amount. Since 1955 the annual expenditure has increased 112 per cent; this represents a 25-per cent increase in the percentage of GNP expended for health and medical care.

The increased expenditure has not been uniformly borne. Insurance benefits accounted for 19 per cent of private personal health care expenditures in Fiscal 1955; in Fiscal 1965 the proportion had risen to 31.6 per cent. The government's expenditures have risen more rapidly than the total; its share has increased from 24.4 to 25.9 per cent for all types over the stated period. Also, the federal share of public expenditure has increased, from 45.2 to 51.2 per cent in ten years.\(^1\)

An examination of these trends and public statements by national political leaders suggests that health and medical care expenditures will be subjected to greater influences, both directly and indirectly, from various government sources.

It is imperative under these circumstances that health services be rationalized, that the limited available resources be allocated to maximize benefits to society, and that the principles of cost-effectiveness be applied at all decision-making levels within the system.

Some educational institutions and research centers, certain specialized hospitals, and some portions of the pharmaceutical and medical insurance industries serve the entire nation or large regions; however, the great mass of health resources such as hospitals, nursing homes, public health organizations, clinics, treatment and diagnostic centers, private physicians and dentists, nurses, etc., serve individual communities. Since resources developed in national institutions are finally implemented at the community level, it is most appropriate to apply cost-effectiveness considerations at the individual community level.

The need for coordinated planning of community health resources has been recognized.\(^2, 3\) This concept of planning is written as law under the Hill-Burton Act. The Bureau of State Services (Community Health) accepted as its mission "the advancement of a concept of community health in which public and private resources together deliver to all segments of the population a comprehensive range of preventive, curative, and restorative
health services in the most efficient and effective manner feasible, with full utilization of relevant scientific advances. [4]

**Difficulties in Planning**

Recognition of the need for comprehensive planning is a necessary, but not sufficient, condition to its satisfaction. The nature of health services makes effective planning difficult. First, there is the problem to which this paper is addressed—adequate, operational measures of effectiveness are elusive and appropriate data are frequently unavailable. Second, some of the most important resources—doctors, hospitals, nursing homes, etc.—typically operate in an independent fashion. Third, many aspects of these resources (e.g., the hospital's location and basic physical layout, or the location, pattern, etc., of the private physician's practice) are fixed for long periods during which the various resources are continually subject to rapidly changing patterns of demand and utilization owing to factors such as advances of medical science, rapid growth of health insurance coverage, improvements in the standards of living, changes in population characteristics, and a greater degree of urbanization. [5] A fourth system characteristic that contributes to the difficulty of planning is the system's stochastic nature. [6] Fifth, because interaction between the health resources and the population occurs at the community level, effective planning must be adapted to the needs and resources of specific communities. Recognition of this fact is implicit in recent legislation (Public Law 89–749). One objective of this Act is the substitution of statewide planning and funding of health services for the current practice of categorical grants (i.e., TB, venereal disease, maternal and child care, etc.).

**A General Approach**

The five problems listed above are only a brief introduction to the conceptual difficulties involved in applying analytical techniques to interdisciplinary problems of socioeconomic planning; other difficulties will be described later in the paper. Viewed pessimistically the problem may appear to be impossible to solve. However, in spite of difficulties, planning decisions have been and will continue to be made. The realistic goal is progress towards improving these decisions and not an 'ultimate' solution. This goal will not be achieved if the analytical technique is so simplified and rigid that it provides conclusions contrary to good judgment. For example, a measure that weights the effectiveness of health programs so heavily in terms of current earnings that programs directed towards the control of children's diseases are neglected is worse than useless.

The recommended approach is to begin the analysis with the problem defined in all its complexity and to proceed by abstracting from this com-
plexity to a model of the problem that is simple enough to be conducive to analysis. During the abstracting process, simplifying assumptions should be made explicit so that it will be apparent whether the conclusions of the analysis should be supplemented or replaced by judgment. As progress is made in developing and using the model, it should be possible to add back those important aspects of reality that were initially ‘assumed away.’ Thus the analytical process and the degree to which complexity is recognized, is divided into two stages as illustrated in Fig. 1. The first is the conceptual stage in which the problem is taken in its full complexity and explicitly simplified until a manageable model is obtained. The second is the operational stage in which the simplified model is developed, validated, and made more realistic.

![Fig. 1. The analytic process.](image)

**MODEL BUILDING FOR COST-EFFECTIVENESS ANALYSIS**

**Some Definitions**

It may be helpful to define some of the terms used to describe the systems being modeled. The community, together with human and physical resources which serve its health needs, comprises a complex and dynamic system. The system's state is defined by the value of certain variables which describe the state of the individual system elements, e.g., the prevalence of particular diseases, the available hospital beds, etc.

The set of variables and the set of parameters can each be divided into two subsets. One subset contains variables and parameters which are generally controllable by the policymakers, while the complementary subset contains those that are relatively uncontrollable. Examples of each type are presented in Table I. It should be noted that “controllable” and “uncontrollable” are defined only in relative terms and depend on the time, the authority, and the resources available to the policymaker.

The policymakers are vitally interested in the values, or outcome, of one particular subset of the variables. This is the subset by which the cost and effectiveness of the system is measured. Any planning process
requires a prediction of the changes to be anticipated in these crucial variables as a result of alternative actions.

Past model-building efforts have had, among others, the following three objectives:

1. To train people in the performance of complex tasks, including management activities.
2. To provide a research tool to aid in the study of complex systems.
3. To act as an aid in the decision-making process.\[^{36}\]

This study is principally concerned with the third objective and, thus, the model must explicitly recognize the policy variables and provide for their convenient manipulation. Since the three objectives are not mutually exclusive, the suggested model should also be useful in meeting the remaining two objectives.

The decision-making objective can be subclassified for those models that support planning:

1. To study complex systems for the purpose of system design, where the interactions are too complex to be understood and simulation can provide insight.
2. To test proposed solutions or hypotheses.
3. To quantify the improvement anticipated from the use of a proposed analytically derived solution.\[^{37}\]
4. To prove the value of proposed changes and to present them in a manner that enhances acceptance of changes and facilitates implementation.\[^{37}\]

For a model to be useful as an aid in decision-making, it should be of the following form:

\[
V = f(a_i, X_i, b_j, Y_j), \quad (i = 1, 2, \ldots, n; j = 1, 2, \ldots, m) \tag{1}
\]
where

\[ V \] is a measure of the system's value,

\[ X_i \] are the controllable elements or decision variables,

\[ a_i \] are the controllable relations or parameters,

\[ Y_j \] are the uncontrollable variables or elements, and

\[ b_j \] are the uncontrollable relations or parameters.

If the relations can be defined and values assigned to the parameters and variables for two alternative system configurations, the model can be used to compare the proposed configurations.

**Cost-Effectiveness**

The planner's objective is to identify for the policymaker the state of the controllable variables and parameters that will maximize the value of the system subject to an assumed set of conditions. The assumed set of conditions is defined by the state of the uncontrollable variables and parameters. Budget constraints may or may not belong to this uncontrollable set. If the budget is in fact fixed, it is an uncontrollable variable and the objective is to maximize effectiveness under the budget (and other) constraints.

Some of the controllable variables and parameters can be varied, within limits, without incurring any appreciable cost; this subset of controllable variables generally includes allocation of a specified resource, establishment of the sequence or the priority of treatment, etc. On the other hand, there is a complementary subset of controllable variables and parameters that require significant expenditures to effect a change; increasing the total facility or personnel resources generally falls into this category.

For a specified expenditure level, or budget, and a specified set of values for uncontrollable variables, there are one or more optimum settings (i.e., optimum system configurations) for controllable variables that maximize the system's effectiveness. At a different expenditure level, a new configuration may maximize effectiveness. The decision as to whether the additional effectiveness is worth the additional cost is properly left to the policymaker. The analysis can provide cost-effectiveness ratios, indicate the marginal cost-effectiveness of additional investment, and eliminate from consideration any inferior alternatives—i.e., system configurations dominated by alternatives that are either more effective and no more costly, or less expensive and no less effective.

**Uncertainty**

The planning problem is complicated by the uncertainty associated with estimates of costs and effectiveness levels. Uncertainty results from the inability accurately to predict the variable and parameter values for
the time period relevant to the alternative plans being considered. Prediction difficulties arise from two sources: first, many relations are only imperfectly known, and second, it is impossible to know whether the relations that currently prevail are stable and will persist. The imperfectly known relations between air pollution and lung cancer or between ‘quality’ of medical care and patient prognosis are examples of the first source of uncertainty. An example of the second type is a significant breakthrough in cancer prevention or treatment that makes existing or planned treatment facilities obsolete and thus invalidates conclusions drawn from models reflecting the current state of medical science.

Another planning difficulty is evaluating future expenditures or benefits. Most economic analyses discount future expenditures and income by applying an interest rate or cost to capital; a variant of this technique may prove useful in evaluating future benefits. Discounting future benefits raises all the conceptual difficulties of interpersonal welfare comparisons if the benefits are to be foregone by one generation so they may accrue to another. An example is the allocation of limited financial resources to a research effort for finding a cure rather than for extending current treatment facilities.

Figure 2 illustrates the above concepts. The cross-hatched areas \((A, B, C)\) indicate the cost-effectiveness regions for three alternative configurations. The points \((a, b, c)\) within the areas represent the expected value of cost and effectiveness. System \(A\) need not be considered further since it is dominated, under all circumstances, by system \(B\). (The entire region \(A\) is less effective and more costly than \(B\).) The choice between systems \(B\) and \(C\) is left to the policymaker.

The planning objective, illustrated in Fig. 2, is to provide the optimum system configuration (setting of the controllable elements and relations)
for a number of budget levels.\textsuperscript{10} Better information will reduce uncertainty and diminish the areas ($A$, $B$, $C$) surrounding the expected values.

**METHODOLOGICAL REQUIREMENTS AND AN APPROACH**

**Requirements**

This section discusses some of the methodological implications of the problem of planning the community health service system and attempts to show why a simulation approach may be appropriate. This should not be taken to mean that analytical models should not be used.* These latter models are generally to be preferred and should be used where possible and efficient. Frequently analytical submodels can be incorporated into a computer simulation program of larger scope.

Determining costs and effectiveness for existing programs, though a difficult task, is relatively simple compared to projecting accurately the future state of a rapidly changing environment. However, the prediction difficulty is alleviated somewhat because comparing alternative programs requires estimates of relative outcomes of alternative actions rather than accurate predictions. As the number of alternative programs to be considered increases, it is no longer appropriate to compare two or three programs at a time. What is necessary is a constant set of assumptions—or an unbiased framework—for measuring alternatives; if a new set of assumptions is to be considered, a means is needed to reevaluate all affected alternatives under the new conditions. In other words, the requirement is for a function that maps alternative decisions into their resulting costs and effectiveness. The only consequence of a prediction error, as long as it is unbiased, would be to shift the areas shown in Fig. 2. Of course the uncertainty that underlies the error will also generally increase the areas ($A$, $B$, $C$) surrounding the expected values and will, if large enough, make it impossible to discriminate among alternatives.

The elements comprising the community health system are synergistic. That is, the joint effect of changes in individual elements of the system will, in general, not be equal to the sum of the individual effects of the changes in each element. Thus, alternative combinations of programs will have to be evaluated jointly. Obviously, double counting on the cost and effectiveness side of the ledger must be avoided; this is an especially important consideration in situations where multiple health problems are prevalent.

* The dividing line between analytical and simulation models is rather vague. For the remainder of the paper, if a mathematical technique or algorithm (such as linear programming) can be used to find an optimum solution, the model is of the analytical type. If heuristic techniques are required to search for ‘good’ (in place of optimum) solutions, the model is of the simulation type.
A methodological consideration discussed in the section dealing with measures of effectiveness is that no single utility function that is appropriate for all programs is known. Even if a single measure of effectiveness is adopted for all programs, the present inability to quantify the relation between, for example, pollution levels or doctors’ visits and this single measure, suggests the use of a multidimensional scale for measuring results. This typically precludes the use of an optimizing algorithm.

Intercommunity difference in population composition, climate, topography, economic conditions, existing health resources, political organizations, and other health-influencing factors demand that the methodology developed should be of a general nature applicable to this diversity. When applied to the relevant data of individual communities, however, the methodology must produce specific results for the individual situation.

**Recommended Approach**

There are alternative approaches that can produce the function required to relate decisions to estimates of their resulting costs and effectiveness. Obviously, the choice of approach should be dictated by the nature of the problem and decisions to be evaluated. One choice is between analytical and simulation models. (As noted previously, the line dividing these two classes of models is quite vague.) An analytical model is generally more efficient if the problem can be stated in terms such that available mathematical techniques can be used to ‘solve’ for an optimum solution. If heuristic methods of ‘searching’ for the optimum are required, or if Monte Carlo techniques are necessary, then simulation is usually more efficient. Another methodological question concerns the use of a computer; this must be answered in terms of the complexity of the model, the number of alternatives to be considered, the number of sets of assumptions under which the alternatives will be evaluated, etc.

For many of the more significant problems, the methodological requirements noted above can frequently be met best by a large-scale digital computer simulation model. This model or program should attempt to accomplish the following, insofar as is practical:

1. Tie together the analytical solutions or subprograms that describe parts of the system not only to provide a common data base and a set of assumptions but also to ensure the logical consistency of these various parts.
2. Provide a means to investigate, through heuristic procedures, those parts of the system that are too complex to be solved by explicit mathematical formulas.
3. Permit flexibility of output in terms of the various measures of effectiveness to be employed.
4. Be able to accept data describing the significant variables of a specific community; that is, become a model of the community specified by the input data.
5. Provide an analysis of the time path of the impact of the particular system configuration being evaluated. For example, if a new organization of hospital facilities is being considered, what will the transition period between the current organization and that proposed look like?

6. Provide a convenient device for testing the sensitivity of the projected results to errors in the data, relations, or assumptions.

If the final requirement for sensitivity testing is met, it should estimate (1) the expected error of the conclusions drawn from the simulation and (2) the relative significance to the error term of current gaps in knowledge of the system.

Because what will happen to a community and the health of its population is so much a function of uncertain internal as well as external factors, it is unlikely that any simulation model will accurately predict it. As noted previously, scientific breakthroughs, changes in national or community economic situations, changes in social and educational structure, etc., while generally excluded from the model, play important roles. The model should, however, be able to forecast what would happen, given a set of assumptions, in the absence of the exogenous occurrences noted above. Of more value, the model should provide estimates of the impact of changes in the controllable parameters or decision variables under various sets of assumptions.

The research strategy should be based upon a building-block approach. As research is performed, for example, which relates intensive preventive medical care to general health or sociodemographic characteristics to the hospital experience of mental patients, it should be possible to include these relations in the model. Also, the larger model should be capable of utilizing analytical or simulation models of subsystems as they are developed. These smaller subsystem models are the building blocks with which the total system model will be constructed. The advantages of developing and testing each building block before its inclusion in a larger model include the following:

1. Testing and verifying the larger model is simpler.
2. The experience gained in developing the earlier models is increased to the benefit of subsequent efforts.
3. Perhaps most important, the completed subsystems can provide intermediate and worthwhile payoffs during the long, expensive modeling of a complex total system.

MEASURES OF SYSTEM EFFECTIVENESS

Basic Objective

Implicit throughout the previous discussion is the need for an operationally useful measure of effectiveness of the alternative systems (or subsystems) to be modeled.

A necessary precondition to the development of measures of effec-
tiveness is an acceptable, perhaps qualitative, statement of the system's basic goals. Alternative programs, with specific aims, which are often strikingly diverse, must be evaluated against these basic goals. For example, certain programs such as those dealing with the control of communicable disease, vector control, and accident prevention, intend to reduce the incidence of specific health problems. Other resources such as hospitals, diagnostic and treatment centers, ambulances and other emergency equipment, and chronic disease and rehabilitation programs attempt to ameliorate the effects of specific diseases. Health education activities, environmental control efforts (e.g., sanitation; noise abatement; and air, water, and land pollution control), and certain preventive health programs are often related to a broad spectrum of health problems. The inclusion of mental and emotional health activities often requires a complete restatement of basic objectives. Objectives of other social welfare programs in economic development, education, health insurance, housing, poverty alleviation, etc., must also be evaluated. The yardstick (whether it is a general statement of goals or a precise measure of effectiveness), if it is to be universally applied, must be able to cope with this diversity.

One approach has been directed towards the development of a health index. This suggests that an index of community health could measure the effectiveness of alternative programs. An acceptable index that (1) is sensitive to the effectiveness of health programs and (2) is composed of measurable components has proved elusive. Indices of crude and age-adjusted death rates have previously been used. Since 1956, however, the death rates in the United States have been relatively constant; thus greater attention has been paid to morbidity. Since increased morbidity is often the result of reduced mortality (e.g., the longevity of diabetics has increased morbidity while decreasing mortality), the evaluation of mortality and morbidity rates is subject to much interpretation.

Sullivan suggests classifying morbidity by four categories of disability: confined, limited mobility, limited activity, and restricted activity. Chiang proposes an index composed of age-specific components derived from the death rate and the incidence and duration of illness. Saunders suggests measuring health rather than illness and uses a concept of the individual's functional adequacy to fulfill a social role. In no case has the development progressed to the point of producing an operationally useful index.

Requirements of the Measures of Effectiveness and Alternative Approaches

Measures of effectiveness are functions of observable variables by which the degree of attainment of the basic goal (e.g., community health) can
be measured. For instance, the basic goal of defense expenditures is the maintenance of peace; a secondary goal derived from the primary objective is to maintain an effective deterrent. To obtain a measurable variable, however, defense expenditure decisions attempt to maximize the number of kills of one sort or another. The measure of effectiveness suggested in the next section will, for many health programs, be measurable in a primary sense. In certain instances, however, it will be necessary to use variables related in a secondary or tertiary manner to the basic goal.

Assume that a finite list of variables related to effectiveness can be agreed on; these might include such things as age-specific death rates, incidence and prevalence of disease, indications of mental health, etc. If the variables in the list are numbered 1, 2, ..., n, the list can be represented by an n-dimensional effectiveness vector, \((X_1, X_2, \ldots, X_n)\), where the \(X\)'s take on different values for alternative system configurations. If each element in one vector takes a value preferable to that of the corresponding element in the alternative vector, there is no difficulty in selecting the superior or more effective system. The real problem arises because this unidirectional relation does not generally exist. According to the utility concept of value, or effectiveness, there exists a way to obtain a utility function to evaluate the alternative vectors: \(\phi = \phi(X_1, X_2, \ldots, X_n)\).

At least three alternative approaches for measuring effectiveness can be identified. The first evaluation alternative, in accord with cardinal utility theory, is to derive a von Neumann-Morgenstern utility function by which a numerical value can be assigned to \(\phi\) for alternative effectiveness vectors. This approach is implicit in placing a numerical scale on the abscissa of a cost-effectiveness graph such as that shown in Fig. 2. The second alternative is to rely on ordinal utility theory. Ordinal utility theorists would not assume that a numerical value can be assigned to each effectiveness vector, but would state that such vectors can be ordered—i.e., one vector is better, worse, or equal to another. The scale of the abscissa of Fig. 2 would then become indeterminate and a statement could be made about relative, but not absolute, effectiveness. For example, the system configuration represented by point \(c\) is more effective than point \(b\), and point \(b\) is more effective than point \(a\); however, there is no way to measure or even compare these differences. The decision as to whether the additional effectiveness is worth the additional cost must be left to the policymaker's value judgment. Accepting the premises of ordinal utility theory, the concepts of marginal effectiveness or cost-effectiveness ratios lose their meaning. The third alternative is to present the policymaker with the complete effectiveness vectors corresponding
to the alternative system configurations and allow him to make his own choice. This has been the approach for most, if not all, evaluations that have been made of complex systems.

The three approaches are not necessarily mutually exclusive. A numerical-valued (cardinal) utility function can be both a useful conceptual device and a goal to be sought even if it cannot be achieved in an operational sense. Using this approach, an attempt can be made to combine as many elements as possible in an effectiveness vector. That is, if a number (say, \( m \)) of utility functions can be defined by combining sets of the \( n \) variables, then the policymaker can be presented with a shorter and more meaningful effectiveness vector of the form \( \phi = \phi_1, \phi_2, \ldots, \phi_m \), where \( m \) is less than \( n \). He must then reconcile these ultimate effectiveness measures.* Analogies can be observed in industrial situations: Ideally, the basic objective may be to maximize the vaguely defined 'long-run profit'; in fact, the policymaker must reconcile an effectiveness vector written in terms of current profit, market share, liquidity, public image, etc.

**The Objective Function: A Case-Month Approach**

The basic goal, or objective, of the system may be stated in a negative manner as the absence of ill health. The measure discussed below, then, is a measure of the ineffectiveness of the system; the measure of effectiveness is complementary to that of ineffectiveness.

First, consider the system as viewed by an individual. The system's ineffectiveness is some weighted sum of his chances of being in any state of ill health for some finite time period.† If \( m \) states of impairment are defined to represent various disability levels (and ultimately death), then from the individual's point of view the system's ineffectiveness is given by:

\[
I^p = \sum_{i=1}^{i=m} c_i \phi_i, \tag{3}
\]

* By placing each of the \( n \) components of the original effectiveness vector into one or more (*not* mutually exclusive) sets \( A, B, \ldots, m \), we can define \( m \) utility functions of the form

\[
\phi_1 = \phi_A(X_i); \text{ an element of } A, \nonumber
\]

\[
\phi_2 = \phi_B(X_i); \text{ an element of } B. \nonumber
\]

In this way the basic effectiveness vector \((X_1, X_2, \ldots, X_n)\) can be transformed into \((\phi_1, \phi_2, \ldots, \phi_m)\). If, in general, \( \phi_1 \) has more intuitive meaning in a policy sense than does \( x_1 \), and \( m \) is less than \( n \), then the transformation is useful.

† It is not logically necessary that conclusions obtained using this definition will be contradictory to those arrived at using the definition of the World Health Organization: “Health is a state of complete physical, mental, and social well-being, not merely the absence of disease or infirmity.” All that is required is that any perceived deviation from the ideal state be described as one of the disability states.
where $c_i$ is the weighting factor for disability state $i$, and $t_i$ is the duration of stay in state $i$. The $m$ states are states of disability or restricted activity.

A reasonable starting point might be to use states corresponding to the four morbidity levels suggested by Sullivan and to add an initial state (controlled disease—minor disability) and a terminal state (premature death). The duration of stay in this terminal state is defined as the actuarially estimated reduction in life span that is due to bad health. Note that this measure aggregates all diseases and accidents to which the individual is vulnerable. Thus the value of $I^p$ must be built from an analysis of each potential hazard. One of the $m$ impairment states must be assigned to each definable stage of the disease or accident and then the time in each state estimated. The appropriate aggregation can then be performed. To avoid double counting it is necessary to assign a single disability state when multiple diseases and/or accidents are involved.

$I^p$ is an individual's measure of the system ineffectiveness. By focusing on the future and examining the probability of being in each state, one can avoid in a reasonable way the problem of the 'infinite' value of life. That is, we are all willing to incur a finite probability of a premature death to avoid a (larger) finite probability of ill health—e.g., to undergo an operation to avoid a serious disability.

Given $I^p$, the individual's measure of the system's ineffectiveness, a function is needed to relate the community's measure of ineffectiveness to the number of people in each state of ill health. That is, how can the individual measures be aggregated to obtain a measure for the community? Any attempt to aggregate over an individual utility function, such as that given in equation (3), encounters the difficulties of interpersonal welfare comparisons. Certainly the values imputed to each $c_i$ may not only vary among individuals at any point in time but also vary for a single individual among different points in time. This is the essence of the 'index number problem' familiar to economic analysis.

The problem is not generally soluble because the solution requires a social indifference curve or a preference function for making interpersonal comparisons of welfare.[23] The general insolubility should not be taken to mean that important particular solutions are not available or that reasonable approximations cannot be made. In fact, for many of the real decisions that have to be made, this measure will provide an unambiguous choice. This results if the situation is analogous to Pareto optimality. In this instance, any possible alternative to the 'optimal' decision will create additional case-months in some disability state without any offsetting reductions in some other state. If it can be assumed that the disability states can be strictly ordered (e.g., limited mobility is
strictly preferred to confinement), the area of unambiguous choice can be further extended. In this instance any possible alternative to the ‘optimal’ decision will create additional case-months in some less preferred state at least equal to the offsetting number of cases in the more preferred state.

These concepts may be illustrated by a simple example. Assume that five alternative disease control programs (A, B, C, D, E) are available and that the estimated case-months in each disability state associated with each are as given in Table II.

As a first guess, assume that Program C is the optimum program. Program C is unambiguously superior to A on the basis of the analog with Pareto optimality; for each disability state the number of case-months associated with Program A equals or exceeds that for C. Program C is unambiguously superior to B by the second criterion. The relative reduction in the number of case-months in (the more preferred) state 4 associated with program B is at least offset by the increase in case-months in (the less preferred) state 5. A comparison between programs C and D is only a more complicated case of that between programs C and B. Again C is unambiguously superior, since reductions in case-months in more preferred states are at least offset by increases in case-months in less preferred states. A comparison between programs C and E illustrates the point previously made that no general solution can be found. The relative evaluation of 1,000 case-months of restricted activity against one additional month of premature death is ultimately a subjective judgment.

The remainder of this section is an attempt to derive possible utility (objective) functions that can map the number of case-months in each disability state into a single numerical measure of effectiveness and thus

<table>
<thead>
<tr>
<th>Disability state</th>
<th>Disease control programs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
</tr>
<tr>
<td>1—Minor disability</td>
<td>1000</td>
</tr>
<tr>
<td>2—Restricted activity</td>
<td>1100</td>
</tr>
<tr>
<td>3—Limited activity</td>
<td>1200</td>
</tr>
<tr>
<td>4—Limited mobility</td>
<td>1000</td>
</tr>
<tr>
<td>5—Confined</td>
<td>1000</td>
</tr>
<tr>
<td>6—Death</td>
<td>1000</td>
</tr>
</tbody>
</table>
reduce the area of ambiguity. The particular functional forms presented below only illustrate the kind of development necessary to obtain a single-valued function.

The objective function should deal explicitly with the future and thus must be a function of the probabilities of future occurrences. The function should also distinguish between the individual’s and the community’s point of view and thus must aggregate the individual disutilities as given in equation (3).

Admittedly, choosing appropriate values for $c_i$ (the disutility weights assigned to each disability state) is a task of subjective value judgments. It seems impossible to avoid this problem. Initial insights may, however, be obtained by determining the weights implicitly assigned by some current health programs. (It would be interesting to examine the consistency of these weights among various programs. For example: Is the implicit value of an extra year's life implied by current kidney disease or cancer programs equivalent to that implied by current accident prevention or preventive health programs?) Also, individuals reveal preferences by their expenditure decisions. Demand (income-expenditure) studies might indicate the relative value or disutility of certain health problems implicitly assigned by individuals in the marketplace.

Certainly the selected function should be monotonic and should increase with the number of ill persons. For each disability state, every additional individual in that state (1) will create an additional burden upon the community by demanding certain health resources and (2) will, because of his own reduced productivity, diminish the community’s capacity to cope with the burden. Thus, in general the ‘disutility’ function will have an increasing derivative; that is, the disutility to the community of each additional person’s ill health will be greater than that of those preceding him.

One possible aggregation scheme that satisfies this requirement would be to multiply the individual measures of ineffectiveness by a ratio of the form:

\[
\frac{\text{number of people in disability state } i}{\text{total population} - \text{number in disability state } i}
\]

In this scheme, each individual entering the state by both adding to the numerator and decreasing the denominator, would make more than a proportionate increase in the measure.

A second, more complex scheme is to concentrate on rate of change in health and to use a dynamic measure that concentrates on the time path of an individual’s health state rather than on static values.

Another aggregation scheme with the above-mentioned characteristics is the exponential form. In this case, the cost to the community of
having \( x \) individuals in disease state \( i \) is expressed as
\[
k_i = c_i t_i e^{\alpha x_i},
\]
where \( k_i \) is the measure of the system's ineffectiveness from the community's point of view, \( x_i \) is the number of individuals in state \( i \) for time \( t_i \), \( \alpha \) is some appropriate constant (perhaps \( \alpha \) should be replaced by \( \alpha_i \) permitting the assignment of a second constant to each disability state), and \( t_i \) and \( c_i \) are as previously defined.

The expected value of the measure of ineffectiveness is then the sum of these weighted by the probability of their occurrence. That is,
\[
E[k_i] = \bar{k}_i = \int_0^\infty k_i f(k_i) \, dk_i,
\]
where \( f(k_i) \), the probability distribution of \( k_i \), is a function of the joint probability distribution of \( x_i \) and \( t_i \). If it can be assumed that these two variables—i.e., the number of individuals in the disability state \( i \) and the duration of stay in that state—are independent, then the expected measure of ineffectiveness is given by:
\[
\bar{k}_i = c_i E[t_i] E[e^{\alpha x_i}]
= \left[ c_i \int_0^\infty t_i f(t_i) \, dt_i \right] \cdot \left[ \int_0^N e^{\alpha x_i} p(x_i) \, dx_i \right],
\]
where \( f(t_i) \) and \( p(x_i) \) are the probability distributions associated with \( t_i \) and \( x_i \) respectively, and \( N \) is the total number in the population.

The variance of the independence of \( x \) and \( t \) is not always valid. For example, if the disabilities in question are the result of communicable diseases, then the number of carriers—and thus the number of secondary infections—is a function of the time each individual spends in the infective state. Even in this case, the independence assumption may provide a good first approximation if the disease prevalence is small relative to the total population.*

The variance of the effectiveness measure is significant in choosing between alternatives; that is, a planner might prefer a program with an estimated value of 10±1 to an alternative with an estimate of 11±10. Under the same assumption of the independence of \( x \) and \( t \), the variance of \( k_i \) is given by:
\[
V[k_i] = E[t_i^2] \cdot E[e^{\alpha x_i^2}] - E^2[k_i]
= \left[ \int_0^\infty t_i^2 f(t_i) \, dt_i \right] \cdot \left[ \int_0^N e^{2\alpha x_i} p(x_i) \, dx_i \right] - (\bar{k}_i)^2.
\]

*It is possible to test this assumption for certain diseases or groups of diseases. Also note that the final integral implies that sufficient numbers of persons are in each state so that this discrete variable can be treated as a continuous one. Integrating the first bracketed term to infinity is a mathematical convenience since the maximum duration of stay in any disability state is limited to a lifetime.
Equations (6) and (7) evaluate the expected value and variance of the system ineffectiveness for a single disability state. If independence is assumed among the states,* then the total system's mean effectiveness is given by

$$\bar{K} = \sum_{i=1}^{i_{\text{max}}} k_i,$$

and the variance by

$$V[K] = \sum_{i=1}^{i_{\text{max}}} V[k_i].$$

(8) (9)

If the constants $c_i$ and $\alpha$ are chosen so that $\bar{K}$ is restricted to values between 0 and 1, then the expected value of the effectiveness of the community health system is given by

$$\bar{E} = 1 - \bar{K} = 1 - \sum_{i=1}^{i_{\text{max}}} c_i E[t_i] E[e^{\alpha x_i}],$$

and the variance by

$$V[E] = V[K].$$

(10) (11)

If the probability distributions of the duration of disability states and the number of individuals in each state ($t_i$ and $x_i$) are known, then $\bar{E}$ and $V[E]$ may be determined by performing the mathematical operations shown in equations (6) through (11). For some health problems, sufficient data are available to establish these functions and their parameters. The measure of effectiveness of a program is the net increase in $\bar{E}$ anticipated over the program's life, i.e., the weighted sum of the probable net decrease in case-months in each disability state. Future benefits may be discounted. Once the admittedly difficult task of establishing the $c_i$'s has been accomplished, a means will have been found to compare accident prevention programs with programs designed to facilitate emergency treatment, or to compare additional diagnostic centers with additional hospital beds. These comparisons can be made if the appropriate probabilities can be projected for each alternative plan considered.

**Program-Specific Measures**

In order to apply any of the case-month approaches suggested above, it is necessary to be able to relate the effects of alternative actions (e.g., prevention, treatment) to changes in the number of case-months. In

* This assumption is obviously not perfectly valid. If the number of persons suffering from an illness is fixed, the size of the group in any disability category is a function of the number of persons in the remaining categories. Thus the off-diagonal elements in the variance-covariance matrix are, in general, different from zero, and equations (8) through (11) become more complex. The assumption may more reasonably be stated thus: Assume that the total number of disabled persons is small relative to the total population and that most entries into any disability category come from the well population, and that the off-diagonal elements in the variance-covariance matrix are small.
some instances, the relations are well known from laboratory experiments or from many case histories, and the case-month method may be applied directly. For example, the effectiveness of immunization programs designed to control communicable diseases such as smallpox, diphtheria, influenza, etc., may be measured by the distribution of secondary cases following one or more initial cases. The effectiveness of detection and treatment programs intended to control communicable diseases such as venereal disease, tuberculosis, etc., could also be directly measured in terms of the probability distribution of the case-months in each disease (i.e., disability) state.

There are some programs for which the functional relations between disability state changes and controllable variables or parameters are not known with a high degree of certainty. Then it is often necessary to rely on expert judgment to obtain the required relations between prevention or treatment and illness. For programs in which these relations are not well known, it may be advisable to use more easily obtainable measures of effectiveness to make intraprogram decisions. This may be the case for some programs concerned with the utilization of health resources, namely: certain aspects of physician care and nursing services, and certain aspects of hospitals or other facility services. For example, while the beneficial emotional and physical effects of comfort care are recognized, it is difficult to quantify the relations or to establish the point at which further comfort no longer serves a useful purpose.

Comfort care is a good example of the 'consumption' characteristics of many aspects of health care. It has been observed that: "... medical services are probably more often demanded because of a desire to reduce the everyday aches and pains of life, which affect earning power little if at all, than for the treatment of the fatal diseases. . . ." 26 Despite conceptual measurement problems associated with consumption benefits, they must be considered to obtain a meaningful analysis.

In cases where definitive relations between treatment and result are not known, effectiveness will necessarily be measured in terms of availability and quality of care. Availability of care can be measured by the degree to which financial, administrative, spatial, or social considerations impede the necessary treatment for perceived problems. Forecast needs can be compared with available or proposed facilities to determine the probability distribution of expected service delays. Quality of care is more difficult to measure but may be estimated by comparing actual treatment with professionally accepted norms for diagnostic tests performed, specialists and other special services utilized, number of examinations, etc. Meaningful conclusions can also be drawn by observing
whether variations in utilization are correlated with medical needs or with
less relevant factors such as income, race, place of residence, etc.

In other instances the inability to quantify relations between suspected
causative factors and incidence of disease precludes the direct application
of the case-month approach. Environmental health programs (e.g.,
pollution control) are important examples of a health subsystem to which
the case-month measure of effectiveness is currently inapplicable. For
these programs, meaningful measures can be derived in terms of average
and maximum concentrations of various pollutants. Also some weighting
scheme could be used to assign values to the probabilities of encountering
various pollutant levels; e.g., a high weight would be assigned to a finite
probability of a 'dangerous' condition and a low weight for an 'uncom-
fortable' situation. Another possibility is to approximate the case-
month approach if disability states are defined for corresponding levels of
discomfort as well as for resulting illnesses.

Finally, mental health is an important subsystem lacking the relation
between health status and both the causes and the treatment of the
disease. Many criteria noted previously are applicable: e.g., available,
appropriate treatment and probable detection at various disease stages.
Also, some objective measures are available, such as the discharge rate
from mental hospitals, recidivism rate, etc. For the foreseeable future,
however, the absence of quantified relations implies that effectiveness will
have to be measured in terms of professionally accepted notions of treat-
ment.

Development of program-specific measures of effectiveness is important
because: (1) except at the highest policy level, planners of health services
are, and will continue to be, concerned primarily with the allocation of
resources within (not among) specific programs and (2) even at the highest
policy level, program-specific cost-effectiveness schedules are necessary if
interprogram allocation decisions are to be made rationally. That is,
decisions allocating funds between, for example, dental health and a
communicable disease program should be made on the basis of the 'best'
(or at least good) programs in each category.

The measures suggested in this section cannot generally and consistently
be related to the over-all (case-month) measure suggested previously.
However, it may prove valuable to attempt to calculate such a measure,
using rough estimates if necessary, as a supplement to more specific pro-
gram measures. This procedure will force a consideration of the over-all
point of view, identify major gaps in information, and encourage develop-
ment of an operationally useful over-all measure of effectiveness that can
be universally applied to all programs.
MEASURING THE COSTS

General

Costs of alternatives, as well as effectiveness, must be estimated in order to rationally allocate scarce resources. Costs are considered easier to handle than effectiveness because dollars are homogeneous and measurable. However, accurately estimating future costs is not simple. Consideration must be given to uncertainty of the estimates, to inflation, to the time-value of money (i.e., the discount rate to be used), to the opportunity costs of alternative uses, etc.

Recent research in health economics has distinguished between actual (direct) and implied (indirect) costs of medical care. Actual costs are true expenditures; implied costs are usually foregone earnings.

Actual Costs

Actual costs include expenditures for hospital and nursing home care, physician and nursing services, drugs, medical research, construction, medical education, training of personnel, and miscellaneous expenditures for administrative costs of programs, information and referral services, public education, etc. The most common conceptual error made with cost data is the failure to distinguish between costs that are fixed and costs that are variable relative to the considered alternatives. Planning decisions should be concerned with the differences in future costs. Thus, in measuring hospital costs, the part of the daily rate for room and board that is applied to pay off the building bonds is irrelevant to most planning decisions; yet reduction of hospital utilization that will postpone or eliminate the need for construction of a new facility is pertinent. Thus, the ‘cost’ of utilization of any facility may vary depending on the community’s rate of population growth. For purposes of control as well as planning, distinction must be made between fixed and variable costs.

What is the relation between resource utilization and the costs of the definable elements? The costs of things bought in very small (inexpensive) quantities vary with utilization in a continuous fashion. These are the variable cost elements of the system. The term semivariable is frequently applied to costs whose relation to utilization is best depicted by a step function. If the ‘steps’ become sufficiently large so that only one level is associated with the range of utilization under consideration, the costs are referred to as fixed.

Interest and obsolescence charges associated with sunk costs, such as past construction expenditures, are the most obvious examples of fixed costs. Other expenditures, such as those for administration, heat and light, etc., are generally fixed for most decision making. Because of their
discrete nature, personnel and equipment costs are generally semivariable; i.e., 1.5 public health nurses or 2.4 X-ray machines are unobtainable. These costs become more nearly continuously variable as the number of personnel or equipment considered grows. For example, expenditures for new facilities can also be viewed as semivariable if the community is large enough. Materials and supplies and externally contracted services are common examples of variable costs.

Figure 3 illustrates these ideas. As an example let the figure represent the relation between costs and utilization of a hospital facility. Let the fixed-cost line represent the unavoidable costs associated with the current facility; let the semivariable line represent equipment, and let the variable cost line represent labor and material costs (this latter straight line is, in fact, a step function with small steps). Assume initially that alternatives A and B are the only ones available. For this analysis both facility and equipment costs are fixed, i.e., uncontrollable. If alternative C is made available, then equipment costs become a control variable that must be considered. If a fourth alternative with a utilization rate high enough to require additional investment in facilities is feasible, then the 'fixed' facility cost becomes a variable.

Planning (or budgeting) and control can be facilitated if costs are accumulated by cost centers or areas of responsibility. That is, the budget procedure should be such that expenditure estimates are initiated by the individual ultimately responsible for those funds and those areas of ac-
Cost-Effectiveness Concepts

Activity; the budgetary and management information systems should be designed to report the costs in the same manner. Thus, areas of authority and responsibility are made coincident and readily identifiable. Attempts to achieve these administrative goals may lead to more effective, functional organizations.

For example, if the state's chief public health officer is responsible for all state expenditures, budgets should be prepared and costs reported on that basis; but fixed costs for which he cannot be held responsible (e.g., for previously built facilities) should be excluded. On the other hand, if the officer in a specific clinic is responsible for only a few classes of personnel and inventory expenditures, these costs must be separately identified and reported. Only in this way can cost-effectiveness be applied at all levels of the organization and the budgeting, planning, and control functions be integrated.

Cost reporting considerations should be taken into account when building simulation models of the system; that is, the output should identify the relevant cost variables, where relevance of a variable is defined by its relation to the appropriate decision-maker. These remarks also apply to selecting the 'measure of effectiveness' variables included in the simulation output.

Implied Costs

Estimating the potential earnings lost because of the incidence of disease is more difficult than estimating actual expenditures made in its prevention and treatment (though the latter is difficult enough, especially for preventive programs not directed to a specific disease—e.g., community health education). Most calculations of implied costs are made in terms of the present value of future losses of output. Reduced efficiency, absenteeism, psychic, and other intangible costs are generally excluded. Values for productive time are calculated for each sex and age class within the labor force using average earnings and adjusting for unemployment rates. Wages for housewives' services are imputed as the mean earnings of domestic servants. In other studies, attempts are made to estimate the earnings lost as a result of underemployment, and in one case a value is imputed to reduced earnings that are due to the social embarrassment of having had syphilis.

Cost-Benefit Analysis

Cost-benefit analysis is an alternative to the cost-effectiveness approach to health planning. Cost-benefit analysis excludes consideration of those factors that cannot be ultimately measured in economic terms. In essence, it proposes that the sum of actual and implied costs be used as
As KLARMAN states:

The total costs of a disease per case serve as the measure of benefits derived from preventing that case. In a cost-benefit calculation the comparison is between contemplated additional expenditures for health and medical services, on the one hand, and the anticipated reduction in costs (direct plus indirect), on the other hand. This is the essential conceptual framework.

In this analysis, economic benefits that accrue to the individual (the implied costs) are lumped with cost-savings that accrue to the government. More useful results would be obtained if the bearer of the costs were identified: the individual, the community, or the government. If this is done, a straightforward cost-revenue evaluation could be made to determine the government’s net cost of proposed programs: i.e., the cost of the program less the anticipated resultant reduction in other programs plus the tax yield from the anticipated additional earnings. Then the decision could be made on the basis of the net economic gain to society (the benefit) that can be obtained from a net public expenditure (the cost). The results of the cost-revenue analysis may also provide a boundary condition for initial program evaluation. (SILBERSTEIN has performed an interesting study along these lines in Israel.)

Cost-revenue and cost-benefit studies deal solely with elements that can be measured in economic terms; a cost-effectiveness study considers other ‘benefits’ in addition to those that are purely economic. However, it is still the net cost to the government of obtaining the economic and noneconomic benefits that is relevant. For instance, changes in productivity and changes in employment in the medical services industry that may result from alternative health programs are important economic factors that must be considered in determining the ‘costs’ of suggested programs.

The appropriate method of handling cost-benefit analyses or the economic aspects of cost-effectiveness analyses may be summarized as follows: For each examined alternative, first determine the resulting changes in income and expenditure streams of each party (public, private, community, etc.); then bring the streams to a single point in time (i.e., the present) by applying discount factors to determine the present value of each stream. The discount factors selected should reflect the opportunity costs, the time-value of money, and perhaps the costs of uncertainty. Concentrating on differences in net financial streams implies that fixed costs will be ignored in the decision-making and that semivariable costs will be considered only if they vary within the range contemplated by the analysis.

Though cost-benefit analysis is simple to handle mathematically, economic considerations cannot be the sole criterion for choosing among
alternatives for an abundant society. (It may be more appropriate in a scarce manpower economy.) The cost-benefit approach distorts the conclusions by overemphasizing morbidity among males in their maximum earning years; by underestimating prevalent problems such as neurotic conditions, common colds, and certain allergic conditions; and by ignoring intangible costs, etc. Actual and implied costs are important dimensions of any resource allocation decision; however, the cost-effectiveness framework permits (and demands) explicit consideration of variables that are measured not in dollars and cents but in terms of basic human values.

**SUMMARY**

The objective of this paper has been to investigate the applicability of current system design and evaluation concepts to planning community health service systems. The concepts include: systems analysis, model building and digital computer simulation, evaluation of system effectiveness, and cost estimating. Each of these topics requires considerably more attention than has been provided in this report. The more salient points and some tentative conclusions are summarized in the following paragraphs.

The costs and effectiveness of alternative system configurations are functions of many variables. System design requires identification of cost and effectiveness variables, specification of the relating functions or parameters, and estimation of their relative controllability.

Models of the total community health service system or significant subsystems within it can provide estimates of costs and effectiveness of alternative system configurations. Large-scale computer simulation models meet the methodological requirements for modeling such complex systems.

The measure of effectiveness of the community health system is essentially multidimensional. One approach has been suggested to reduce the many criteria to a single numerical value. Because of the incomparability of interpersonal utilities, a general solution cannot be achieved; however, it is possible to restrict the area of ambiguity and to reduce the number of indistinguishable alternatives. For a large number of situations, it would appear that a single-valued measure is too restrictive and that the best results can be achieved by reducing the many criteria to a few significant measures to be ultimately reconciled by the policy-maker. The multidimensionality of the effectiveness measure is one of the significant considerations leading to the recommendation of the simulation approach.

* A related, but somewhat different, approach to this problem is to view the decision-making process as an attempt to maximize a committee preference function subject to certain constraints.*[9,30]
Planning is ultimately concerned with future situations. Thus, the relevant costs, like the relevant measures of effectiveness, are those yet to be incurred. This requires that distinctions be made between fixed costs and those that vary owing to the decisions being made, that the degree of uncertainty be estimated, and that the opportunity value of money be considered.

REFERENCES


Copyright 1968, by INFORMS, all rights reserved. Copyright of Operations Research is the property of INFORMS: Institute for Operations Research and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.