Evaluation of Program Effectiveness

O. L. DENISTON, M.P.H., I. M. ROSENSTOCK, Ph.D., and V. A. GETTING, M.D., Dr.P.H.

A SYSTEMATIC, comprehensive approach is needed to evaluate the effectiveness of programs in public health. Our approach is based on the assumption that all programs in public health can be viewed as consisting of a combination of resources, activities, and objectives of several kinds. We maintain that each program is characterized by one or more program "objectives," which represent the desired end result of program activities, and that each objective implies one or more necessary conditions, termed "sub-objectives," which must be accomplished in order that the program objective may be accomplished. "Activities" are performed to achieve each sub-objective and consequently the program objectives. "Resources" are expended to support the performance of activities. A sharp distinction is made between activities, which imply the performance of work, and objectives, which refer to conditions of people or of the environment deemed desirable. Every program plan, whether written or not, makes three kinds of assumptions: (*a*) the expenditure of resources as planned will result in the performance of planned activity, (*b*) each activity, if properly performed, will result in the attainment of the sub-objective with which it is linked, and (*c*) each sub-objective must necessarily be accomplished before the next one can be achieved and, if all sub-objectives are attained, the program objective will be attained.

In evaluating the effectiveness of programs, specific measures of accomplishment of each sub-objective and the program objectives are set up, and data on attainment of each are collected systematically, following accepted principles of research design. In addition, data are collected on the extent to which each activity is performed as planned and on the extent to which resources are used as planned. Findings from the several sets of data are used to strengthen subsequent program planning.

A second paper will deal with program efficiency, defined as the cost in resources of attaining objectives.

Our logic and methods of evaluating program performance have been and are being applied successfully. An account of a field application, "Report of Evaluation of Agricultural Labor Camp Program, 1966," an unpublished report, is available on request to the Michigan Department of Public Health.

Kinds of Evaluative Questions

The evaluative questions that program directors ask most frequently can be grouped into four categories.

Appropriateness

Questions on appropriateness concern the importance of the specific problems selected for programing and the relative emphasis or prior-

Mr. Deniston is a research associate and lecturer in public health administration, and Dr. Rosenstock and Dr. Getting are professors of public health administration. All three authors are members of the department of community health services, University of Michigan School of Public Health, Ann Arbor.
The investigation was supported by Public Health Service research grant No. CH 00044 from the Division of Community Health Services.

ity accorded to each. Program directors are concerned with appropriateness when they ask, "Are our program objectives worthwhile and do they have a higher priority than other possible objectives of this or other programs?"

Adequacy

Ideally, objectives are oriented toward elimination of the problem which gave rise to the program, but various constraints may necessitate reducing the scope of an objective from focus on complete solution of a problem to the more modest scope of reducing a problem by a specific amount or limiting an objective to a portion of a population experiencing the problem rather than trying to reach all those at risk. Questions concerning how much of the entire problem the program is directed toward overcoming refer to the adequacy of program objectives.

Effectiveness

Programs may differ in their effectiveness; that is, in the extent to which pre-established objectives are attained as a result of activity. Effectiveness in attaining objectives is distinct from program appropriateness and adequacy.

Efficiency

Program efficiency is defined as the cost in resources of attaining objectives. The efficiency of a program may be unrelated to its effectiveness, adequacy, and appropriateness.

These four kinds of evaluative questions may be asked before a program begins or at some point after it has been in operation. Applied beforehand, the questions are an evaluation of the planning process. They can then be phrased as asking whether the proposed program has important objectives, whether it is aimed at overcoming a large proportion of the problem, whether the activities proposed are likely to attain the objectives, and whether the unit cost of attaining objectives is likely to be acceptably low.

When these four questions are asked about an operating program, they constitute an evaluation of performance. The questions then focus on (*a*) whether the program has in fact been directed toward important problems, (*b*) how much of the total problem has been controlled, (*c*) the extent to which the predetermined program objectives have been attained, and (*d*) the actual costs of attaining objectives.

Our proposed model for evaluation is applicable only to assessing performance of a program and not the planning of a program. Furthermore, the model does not deal with appropriateness. Our model is applicable to adequacy only when all the dimensions of a problem can be specified. We can determine the extent to which a home health care program solved a specified set of problems in a specified sample of a population. But we cannot determine how adequately it solved the entire range of health problems in the whole population unless we are able to identify in advance the total range of health problems in the total population affected. If such information is available, the measurement of program adequacy can be computed simply from the measure of program effectiveness.

Our model is intended to answer two questions: (*a*) to what extent were objectives attained as the result of activities (program effectiveness) and (*b*) at what cost (program efficiency)? The model builds upon a number of contributions to program evaluation, especially those of Paul (*1*), MacMahon and co-workers (*2*), Hutchinson (*3*), Freeman (*4*), and James (*5*). If there is anything unique in our model, it is the attempt to be comprehensive, uniform, and consistent in our definitions and logic and in the application of the definitions and logic to health programs.

Model to Evaluate Program Effectiveness

Our model for evaluating effectiveness requires systematic description and measurement of each variable of a program, that is, resources, activities, and objectives. If a variable or portion of a variable cannot be measured, the model cannot be fully applied. However, even partial application of the model will provide information useful to subsequent planning; in addition, it will show the evaluator precisely where additional measurements are needed.

The model is intended primarily for use by program personnel to evaluate certain aspects of their own performance. However, an outside evaluator can also use it. Regardless of who performs the evaluation, it should be remembered that the purpose of evaluation is improve-

ment. Therefore, the evaluation should be endorsed, if not performed, by those who have the authority to make changes.

As a final constraint, the model does not offer a systematic procedure for assessing any unplanned impact of activities, although activities performed for a specific purpose may indeed have side effects. We recommend that program personnel attempt to assess side effects on a subjective, impressionistic basis until a more systematic way of measuring them is developed.

Definition of Terms

The model uses terms that are familiar but have not always been used consistently. First, therefore, these terms will be defined.

Program. An organized response to reduce or eliminate one or more problems. This response includes (*a*) specification of one or more objectives, (*b*) selection and performance of one or more activities, and (*c*) acquisition and use of resources. Although the term "program" probably suggests similar concepts to most health workers, two ambiguities are common. First, for many workers, human ailments or hazardous environmental conditions constitute the only legitimate focus for a program. Thus, concern with tuberculosis or water pollution is a program, but disease casefinding or food handler training is not. Such workers frequently classify casefinding, food handler training, and similar concerns as a "subprogram," "component," "project," or "technique."

To simplify terminology and logic, any area or scope of concern may be considered a program for the purpose of evaluation of performance. Thus, disease casefinding, food handler training, and professional education can be evaluated, although their immediate objectives do not have direct impact on a human ailment or environmental hazard.

A second common ambiguity results when the word "program" is further specified by adding a content area, such as "school health program." To some workers, a school health program means correction of defects; to others, measurement of height and weight or immunization; and to yet others, school health may connote certain areas of instruction.

Both sources of ambiguity may be removed by stating the objectives of the program and listing the activities performed and resources used. If this is done, the result is a statement that the program consists of resources a, b, c, used to perform activities d, e, f, which, in turn, are designed to attain objectives g and h. This definition, therefore, permits widely varying scopes of work to be defined properly as programs.

Objective. A situation or condition of people or of the environment which responsible program personnel consider desirable to attain. To permit subsequent evaluation, the statement of an objective must specify (*a*) what—the nature of the situation or condition to be attained, (*b*) extent—the quantity or amount of the situation or condition to be attained, (*c*) who—the particular group of people or portion of the environment in which attainment is desired, (*d*) where—the geographic area of the program, and (*e*) when—the time at or by which the desired situation or condition is intended to exist.

Within the framework of our definition are three kinds of objectives, each meeting the basic definition.

1. Ultimate objective. A condition which is desired in and of itself according to the value system of those responsible for the program. Reductions in morbidity and mortality are examples of conditions that are typically regarded as inherently desirable.

2. Program objective. A statement of that particular situation or condition which is intended to result from the sum of program efforts. It may or may not be considered inherently desirable, that is, an ultimate objective.

3. Sub-objective. A subordinate or sub-objective is an objective which must be attained before the program objective may be obtained. A sub-objective is seldom inherently desirable.

Most programs have several sub-objectives. All sub-objectives are related in time to each other and to the program objective; that is, the program planner believes they must be accomplished in a particular order. Frequently, two or more sub-objectives must be attained simultaneously. In some programs sub-objective 1 must be accomplished before sub-objectives 2, 3, and 4 may be accomplished, and 2, 3, and 4 may have to be accomplished simultaneously in order that sub-objective 5 may be obtained,

and so on. Other writers have used such terms as "intermediate objectives" or "activity goals" to describe sub-objectives.

There is a commonly used distinction between long-range and short-range objectives. The phrases are not recommended because they can be ambiguous, as the following examples illustrate.

In some circumstances long range refers to a program objective and short range to a sub-objective. Thus the long-range (program) objective might be a 90 percent reduction in the prevalence of tuberculosis after 5 years and one short-range (sub-) objective might be that all people with tuberculosis know how to follow a prescribed chemotherapeutic regimen.

In other instances long range and short range refer to amounts of the program objective that can be expected at any given stage. The long-range objective might be a 90 percent reduction in the prevalence of tuberculosis after 5 years and the short-range objective might be its reduction by 20 percent after 1 year.

The meanings of the concepts are different in these two examples. In the first, the short term objective is actually a sub-objective which might be wholly attained and still not imply any attainment of the program objective. In the second, the short-range objective represents partial attainment of the program objective. The distinctions used in this paper make it possible to describe plans and outcomes without differentiating between long-range and short-range objectives.

Activity. Work performed by program personnel and equipment in the service of an objective. Activity as we use it does not imply any fixed amount or scope of work; it may be applied with equal validity to such diverse efforts as writing a letter or providing comprehensive health care. An activity can usually be subdivided into more specific activities. Providing comprehensive health care, for example, could be subdivided into providing curative health care and providing preventive health care; these, in turn, are capable of further subdivision and specification.

Probably the greatest cause of confusion and difficulty in both planning and evaluating health programs is lack of a clear and consistent distinction between an activity and an objective. James (*5a*) has made the distinction in terms of an analogy to a bird—the activity is flapping wings, the objective is being at some desired place. Activities consume program time and resources whereas objectives do not.

The distinction between objectives and activities may be further clarified by an analogy between the logic of an experiment and the logic of a program plan. In an experiment, the investigator asks whether a cause-effect relationship can be demonstrated. He performs some procedure on a group of subjects (cause) and predicts that a specific result will or will not occur (effect). The experimental procedure is linked to the expected outcome by an hypothesis. The hypothesis can be stated in an "if . . . then" form; that it, if treatment A is provided, then effect B will result. Program planning parallels the logic of an experiment. After identification and analysis of needs or problems, a program objective is established and decisions are made about the activities to be undertaken. A program objective is parallel to the experimenter's expected result or effect, and the program activities are parallel to the experimental procedure or cause. The planner hypothesizes that a given method or set of activities will lead to the attainment of the objective; if a certain activity is performed, then the desired objective will be achieved. The hypothesis can be tested only by evaluation.

Resource. Personnel, funds, materials, and facilities available to support the performance of activity. Resources, like activities, may be described with varying levels of specificity.

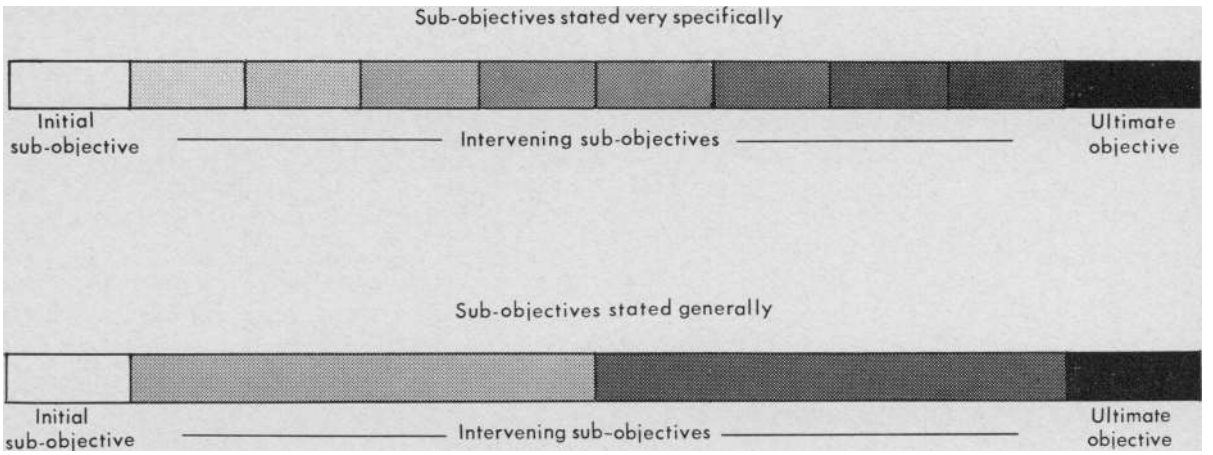
Program assumption. An hypothesis concerning the nature of relationships among the various aspects of a program. Every program plan includes three major kinds of assumptions.

1. The assumption that use of resources as planned will result in the performance of planned activity.

2. The assumption that performing planned activity will result in the attainment of the desired objectives. Similar assumptions link subdivisions of program resources to the subset of activities they support and, in turn, to the program sub-objective they are intended to establish.

3. The assumption that each sub-objective must necessarily be attained before the program

Figure 1. Time sequence of objectives



objective can be attained and that attainment of all sub-objectives will result in attainment of the program objective.

A Program Overview

It is helpful at this point to describe the logical planning of a program if such limiting factors as financial and technical constraints could be ignored. Assume that a program objective as we have defined it has been established; that is, a statement has been formulated that the program is intended to attain a given situation or condition in a particular group of people or portion of the environment, in a given geographic area, by a particular time, and to a particular extent. Ideally, the planners, having specified the objective of a program, would then specify the conditions that would have to occur before the objective could be attained. Each of these necessary conditions is a sub-objective.

The planner then identifies alternative activities which might be effective in attaining the objectives. He considers the anticipated costs and effectiveness of each alternative. Finally, the planner selects the best alternatives in terms of his assessments of program appropriateness, adequacy, effectiveness, and efficiency. Current approaches to selecting objectives and activities include planning-programing-budgeting, cost-benefit analysis, systems analysis, and operations research.

The final phase of planning is assignment of resources to support the activity selected.

It has already been indicated that the total program plan contains many assumptions about the relationships among resources, activities, and objectives. In a very real sense, evaluation of effectiveness is the determination of the extent to which these assumptions are true; evaluation assesses (*a*) whether the expenditure of resources did lead to the performance of planned activity, (*b*) whether each activity did attain its intended outcome or sub-objective, (*c*) whether each sub-objective was necessary to attain the next higher sub-objective, and (*d*) whether attainment of all sub-objectives was sufficient to accomplish the program objective.

Application of the Model

To conduct an evaluation of program effectiveness using the model proposed involves a series of actions. The process is divided arbitrarily into three steps.

Step 1. Describing the Program

The program description consists of naming the program to be evaluated and specifying the program objective or objectives, sub-objectives, activities, and resources. If these things have already been done in the planning phase, this step in evaluation is relatively simple and may require only copying them from the program plan. However, it is rare in current health practice to find written program plans with objectives spelled out in sufficient detail and precision to permit evaluation of effectiveness.

Specification of objectives. Specification of the program objective or objectives and sub-objectives may prove especially troublesome if

these concepts are new to health workers. Drawing up a sequence of objectives may improve understanding.

The time sequence of objectives may be placed on a horizontal line, with an ultimate objective at the extreme right. At left is the initial condition or sub-objective that, in the opinion of program planners, must exist if the ultimate objective is to be attained. Other planners might formulate a still earlier sub-objective. The initial sub-objective is arbitrarily chosen to represent the first new condition that the planner believes must be attained before the succeeding conditions can occur. Everything to the left of the initial sub-objective is taken as a given, that is, it is assumed to take place without program intervention.

Between the initial sub-objective and the ultimate objective are the intervening sub-objectives or necessary conditions. Many sub-objectives are possible if each is stated specifically, or all sub-objectives can be grouped under two or three general headings (fig. 1). There are disadvantages, however, in specifying either a very small or very large number of sub-objectives.

The first task in describing a program to be evaluated is to state its objective. The program objective may or may not be an ultimate objective from a health professional's point of view. A program may encompass an entire line, or any portion of such a line. Nevertheless, the program objective is an arbitrary point on a line that is expected to culminate in an ultimate health objective. Thus one program might include only the portion of the line that includes the first three intervening sub-objectives (fig. 1).

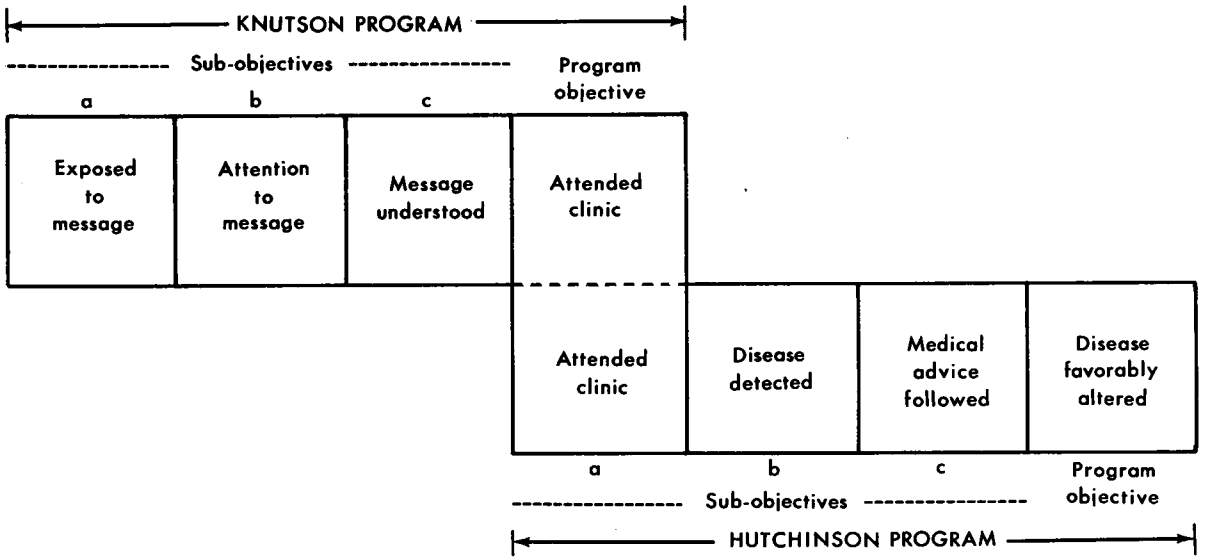
An evaluation of program effectiveness must include measurement of the condition that is specified in the program objective. In addition, it should include measurement of as many sub-objectives as available time and resources permit. In general, we recommend that several sub-objectives be measured in order to locate the source of trouble if a program is less effective than desired. Measurement of a large number of sub-objectives can consume great quantities of time and possibly of other resources. Should an administrator wish to evaluate the effectiveness of several programs and have limited resources for evaluation, he may prefer to measure attainment of program objectives only for all programs, returning to measure sub-objectives for those programs manifesting lowest effectiveness.

No dictum can yet be given as to the optimal number of sub-objectives since an infinite number of previous conditions (sub-objectives) are necessary for a given condition (objective) to occur. Suppose sanitarians are attempting to increase restaurant operators' knowledge of defects in their operations. One necessary condition (sub-objective) for acquisition of information is that the operator understand the sanitarian's vocabulary. But a necessary condition for the operator to understand is that he pay some attention to what is being said, and a necessary condition for his paying attention is that he be physically exposed to the message (he be physically present and capable of hearing, seeing, and thinking). The ability to perceive and think, in turn, is contingent upon the functioning of nervous tissue which, in turn, is contingent upon more basic biochemical balances. Biochemical function is contingent upon atomic motion which is dependent on subatomic motion and so on.

Although this is reduction to absurdity, it is clear that a somewhat arbitrary division will be needed to determine the number of sub-objectives to be measured. The cutoff point would seem to be the point where the apparent disadvantages of expending further resources on measurements would about equal the apparent disadvantages of assuming that doubtful conditions will in fact be realized. Specifying too many sub-objectives may make the evaluation too costly and detailed; specifying too few may yield insufficient information about weak aspects of the program. Most administrators assume that restaurant operators are not deaf, blind, and mentally defective, but many will be unwilling to assume that operators will automatically pay attention to what the sanitarian says. In that instance, the adequate functioning of sense organs would be accepted as a given rather than as a sub-objective and would be to the left of the initial sub-objective, but the operator's attentiveness would be a sub-objective.

The nature of ultimate objectives, program objectives, and sub-objectives is illustrated by combining examples from Hutchinson (*3*) and Knutson (*6*).

Figure 2. Relationship of Knutson and Hutchinson program lines



In a discussion of programs of early casefinding, Hutchinson cites a program with an ultimate objective. He considers that alteration of the natural course of disease in a favorable direction is intrinsically valuable from the point of view of the medical profession—although it may be only an intermediate or sub-objective for such professions as theology and philosophy. On the other hand, a number of sub-objectives (which he terms intermediate) are crucial to program effectiveness but are not in and of themselves intrinsically valuable, such as (*a*) that people come for screening, (*b*) that cases of illness are detected, and (*c*) that persons with the disease follow prescribed treatment. Sub-objectives a, b, and c are desired not because of their inherent value but because the ultimate objective cannot be attained unless each of them is attained.

Knutson refers to a hypothetical health education program, whose objective is some desired behavior of people to whom the program is directed. He lists a number of sub-objectives for his program which he, like Hutchinson, terms intermediate objectives: (*a*) the people must be exposed to the material, (*b*) they must give the material their attention, and (*c*) they must understand the words and concepts. If the objective (behavior of target audience) referred to by Knutson were that people come for screening, that objective would be identical with initial sub-objective specified by Hutchinson. The sub-objectives specified by Knutson, then, occupy a portion of the program line to the left of Hutchinson's initial objective (fig. 2).

The program objective for Knutson's health education program is the initial sub-objective of Hutchinson's broader disease control program. Knutson's sub-objectives are taken as givens in the Hutchinson example.

The importance of correctly stating the program objective may be illustrated by another example. Assume that the health education program within a larger disease control program is to be evaluated. The true objective is that all members of a particular group residing in a given area come to a clinic for a particular screening test on a specified date. Figure 2 shows that the adjacent objectives, in elaborated form, are that all members of the group understand the words and concepts of the educational material and that all positive cases of disease in the group are detected.

If either adjacent objective were mistakenly stated as the program objective, the results of subsequent evaluation would be misleading. In the first instance, the program would be judged as more effective than it actually was, since many people may indeed have understood the words and concepts, but nevertheless failed to attend the clinic—failed, that is, to take the desired action. In the second instance the pro-

gram would be judged as less effective than it actually was, since failures of diagnosis would incorrectly be attributed to the health education program. Therefore, it is essential to state as the objective of the program the precise outcome that is desired and expected to result from the activities to be evaluated.

To illustrate further the usefulness of measuring attainment of sub-objectives, consider a program similar to the health education program discussed previously. The true objective was that stated in the previous example, and 50 percent of the specified group attended the clinic. Although that finding is important, it does not provide a clue as to why the program failed with half its intended audience. Such knowledge can be acquired, however, by measuring attainment of the program's sub-objectives. In addition to the program objective, the following four sub-objectives may have been specified: (*a*) all eligibles are exposed to the educational material, (*b*) all eligibles attend to (read, listen, and so forth) the material, (*c*) all eligibles understand the point of the communication, and (*d*) all eligibles be interested in early detection of the disease in question.

Recalling that 50 percent of the eligibles came in for screening, evaluative results can be arbitrarily assigned to each sub-objective. A sample survey might show that for each 100 eligibles: 95 were exposed to the material; of those, 90 paid attention to it; of those, 65 adequately understood the point of the communication; and of those, 35 were interested in detecting disease early. Finally, all 35 satisfying all four sub-objectives came in for screening. Thus, the first two sub-objectives were attained with a total loss of 10 percent of the eligibles. Some attention might be given to reducing this loss. However, more important is the additional loss of 55 percent (25 and 30 percent respectively) that occurred in attaining the third and fourth sub-objectives. Thus, of all 90 people attending to the program material, more than 60 percent (55 of 90) failed to understand the message and to become interested in early detection. Clearly, activities to accomplish these two sub-objectives need to be strengthened.

It may be noted that 50 percent came in for screening, but only 35 percent were interested in early detection. This suggests that personal interest is important but not absolutely necessary to obtain participation. Perhaps further study would show that some people came in because of the influence of relatives or friends. The planners might wish to build on such a finding in subsequent programs.

The preceding discussion and examples have implied an objective that is identical for each person in the program population. However, for some programs, the objective for each member of the target group may be different as those in mental health and home care programs.

In these, it would be more appropriate to state a separate objective (and sub-objectives) for each person to be served. In such programs the attending physician may establish a unique objective for each patient; for example, by the end of some time period, Mr. A will return to work, Mr. B will bathe and dress himself. The program objective can then be summarized as all, or some proportion of, program clientele will attain their unique objectives within specified periods.

Thus far, the discussion of objectives has focused on only one kind of content objective associated with programs. However, two other kinds of objectives need to be recognized.

Each health worker in a program will have personal objectives, such as advancement in rank or title, a higher salary, respect of his peers, popularity, and so forth. These may or may not be consistent with program objectives.

In addition to these personal objectives, every agency or organization has what may be called survival programs—a set of activities undertaken to insure the stability and continued existence of the agency. Certain public relations and public service activities are examples of survival programs.

Concern with personal satisfaction and organizational survival will act as constraints in planning programs, in the setting of priorities, and in the selection of objectives and activities. In this sense, they do not interfere with evaluation of program performance although knowledge of constraints may be useful in interpreting evaluation results. If desired, however, our model could be applied to the evaluation of employee morale or organization survival "programs."

The attention given to program objectives is deemed necessary because rarely have objectives been stated clearly when evaluation is desired and because it is difficult but extremely important to distinguish between objectives, sub-objectives, and activities.

Some reasons for lack of predetermined objectives have been described by Selznick (*7*).

> Once an organization becomes a "going concern," with many forces working to keep it alive, the people who run it can readily escape the task of defining its purposes. This evasion stems partly from the hard intellectual labor involved, a labor that often seems but to increase the burden of already onerous daily operations. In part, there is the wish to avoid conflicts with those in and out of the organization who would be threatened by a sharp definition of purpose, with its attendant claims and responsibilities.

The threat engendered by making program objectives explicit becomes intensified when one seriously proposes measuring attainment (*8–10*). We do not see any ready way to eliminate all threatening aspects of evaluating program effectiveness, but we do believe that the threat can often be overcome if the benefits can be perceived as outweighing the costs.

Specification of activities. When the program objectives and sub-objectives have been stated, the next task in step 1 is to specify all program activities, linking each to the objective or sub-objective it is intended to accomplish. There are two reasons to do this. First, making activities explicit can serve as a check on the adequacy and completeness of stated objectives. If a planned or continuing activity cannot be linked to any stated objective or sub-objective, either a necessary objective or sub-objective has been omitted, or the activity is unnecessary. Conversely, if a stated objective or sub-objective has no activity linked to it, either an essential activity is not being planned or performed or the stated objective or sub-objective is not necessary to the program.

The second reason for including activities in the program description is to determine the extent to which they were performed as intended. For this purpose, activity must be carefully specified—what is to be done, by whom it is to be done, and when and where it is to be done.

If an objective or sub-objective is not attained, either an activity was not performed as planned or the assumption linking the activity and the objective or sub-objective was not valid. Of course, if the activity was not performed or not performed properly, the linking assumption must remain untested.

Specification of program resources. The final task in step 1, specification of program resources, makes it possible to determine if resources were used as planned. If planned activities were not performed, knowledge of whether resources were used will allow determination of the validity of assumptions linking resources and activities made in the planning process.

In summary then, the first step in evaluating program effectiveness requires a clear statement of the program objectives, the specification of a reasonable number of sub-objectives, specification of program activities, and a description of program resources.

Step 2. Measurement

A complete treatment of steps 2 and 3 is not possible in this paper. The interested reader is referred to standard texts which cover the material in detail (*11–14*). In addition, consultation from experts such as statisticians and behavioral scientists will often be helpful in completing steps 2 and 3.

Step 1 outlined a method for describing programs to permit evaluation of program effectiveness. Step 2 requires identification of the kinds of evidence needed to determine that an objective or sub-objective has or has not been achieved.

In general, valid and reliable measures of program accomplishment are needed. Briefly, validity of a measure is the extent that an obtained score measures the characteristic that it is intended to measure. The terms "sensitivity" and "specificity" applied to diagnostic tests are components of validity. Reliability of evidence is the consistency or repeatability of a score.

We are concerned with validity and reliability because test scores do not always measure consistently what they are intended to measure. Suppose a series of measures are obtained on a group of persons or restaurants or on samples of water. A range of scores will be obtained. Differences in scores may reflect true differences

in the characteristic being measured, but different scores may reflect other factors. If the measure is of people, responses may not only reflect the item being measured but also such transitory factors as mood or fatigue. In measures of the physical environment, factors such as variations in the administration of a test and the care with which instruments are read will also affect scores.

Another possible source of variation in scores exists in measurements of complex concepts such as health status, morbidity, or cleanliness. Such concepts are composed of many specific subconcepts. For example, good health might include almost an infinite number of measurements of the functions of various organ systems. It is unlikely that any one test will measure all functions. A test of two or three functions applied to a group of people might show that some are healthier than others without giving recognition to the fact that had tests been made of other functions, results might have been different.

Because test scores are determined not only by true differences in what is being measured but also by other causes, it is never completely safe to accept a test score at face value. When possible, evidence should be obtained that the test is valid (it measures what it is intended to measure) and that it is reliable (successive administrations of the test or administration by different persons yield similar scores).

In selecting a measure of accomplishment, the evaluator may know of valid and reliable measures or he may search the literature for relevant measures that others have used. If he fails to locate an acceptable measure, he may have to develop a unique measure which satisfies the basic criteria of measuring instruments. However, if his resources do not permit the development of a measure for certain objectives, he may be forced to omit some measures from the evaluation, thus reducing the resultant amount of information bearing on the success of a program.

When to measure. The program objectives and sub-objectives state the time period in which the measures are to be applied. Although one tends to think of evaluation as being conducted over a relatively short period, evaluation will be most valuable if it is conceived as a more nearly continuous process. Since attainment of sub-objectives occurs in a time sequence, attainment of each should be measured soon after attainment is expected.

How to measure. In deciding how to make the needed measurements, two problems are particularly important: how to avoid bias and the problem of sampling.

The possibility of bias is great if one evaluates his own work. This possibility is especially great if observational rather than physical measures are used, such as reporting the cleanliness of an object or the satisfaction of a patient. Bias can be reduced by using physical measures when possible, or if observation or judgment is necessary, by having more than one person judge.

Sampling procedures are often used in evaluating a program since it is rarely feasible to measure the attainment of objectives in every person in the target group or at every location. In such cases, a probability sample which accurately represents the total population must be selected.

The size of the sample required depends on such technical considerations as variance in the distribution of the quality being measured, the amount of change expected as a result of program activity, and the level of certainty desired when inferring that what is true of the sample is also true of the population from which it was drawn.

Collecting the data. Data on the attainment of objectives and sub-objectives as well as on the performance of activities and use of resources must be collected at times indicated in the program description.

Step 3. Determining Effectiveness

In evaluating effectiveness the question is not merely were the program objectives accomplished but to what extent can achievement of the objective be attributed to the activities of the program?

Analysis of program effectiveness can be simplified by using a set of ratios involving the three program variables: resources, activities, and objectives.

Simplest is the ratio of actual resources to

planned use of resources, $\frac{AR}{PR}$.

Slightly more complicated is the ratio of actual program activities performed to planned activities, $\frac{AA}{PA}$.

The ratio that indicates attainment of objectives is still more complex. We denote this ratio as $\frac{AO}{PO}$. AO is the net attainment of the objective attributable to program activity and PO is the attainment desired less the status that would have existed in the absence of the program. It might be imagined that the proper comparison would be between actual status of the objective when evaluation is performed and the status of the objective that had been planned. However, such a comparison is not valid since it does not take account of effects on the program of activities and events outside it. Evaluation should assess the extent to which achievement of the objective can be attributed to activities performed in the program.

Therefore, it is necessary to find a way of comparing the net accomplishment attributable to the program with the accomplishment intended for the program. One way of doing this is to determine the status of the objective at the time of evaluation and then to subtract from it an estimate of what the status would have been had the program not been undertaken. For example, if a program operator finds that 90 percent of a group of clients are immune to a disease following the conduct of a program, he cannot properly take credit for all 90 percent, but only for those who would not be immune had his program not been undertaken.

What is true for the actual status of the objective, the numerator, is also true for planned attainment, the denominator. One must subtract from planned attainment that portion of the desired status that would have occurred in the absence of the program. For example, suppose it was desired that 90 percent of a population be immune to a disease. Evaluation shows that 80 percent actually became immune but that half, 40 percent, became immune through activity outside the program (visits to physicians and so forth). Program effectiveness would then be $\frac{80-40}{90-40}=\frac{40}{50}=80$ percent.

Another example based on actual data will show how the ratio may be computed. In a food service sanitation program consisting of inspections, a rating system was used as the measure both of the problem and the objective. The objective was that the average sanitation rating of food establishments in the county will be at least as high as 90 by July 1, 1966. On July 1, 1966, the average rating was 85.7.

Additional data showed that the average rating in an uninspected section of the county was 81 on July 1, 1966. If we use the rating of 81 as an estimate of what the countywide rating would have been without the program, the program effectiveness ratio becomes

$$\frac{85.7-81}{90-81}=\frac{4.7}{9}=52.2 \text{ percent.}$$

How is it possible to estimate the status of the objective in the absence of the program? The most certain way is to use a control or comparison group similar to the one exposed to the program. The control group procedure maximizes confidence in judging the results that may be properly attributed to the program.

Control groups are not always feasible in evaluations of health programs, but they could be employed more often than they currently are. For example, when a new program cannot be initiated throughout a jurisdiction, it may be possible to begin it in several places selected at random and to use the remaining areas as controls. Or alternate procedures to accomplish objectives might be applied systematically in different parts of the jurisdiction, as is done in clinical field trials to test whether one procedure is superior to another.

However, if a strict control group is not feasible, a control group can be approximated by comparing community status before and after the program with information about nearby communities not exposed to the program. While this is not an ideal procedure, it may provide guidance as to the impact of the program.

A major danger in using natural groups as comparisons or controls is that an available group, within or outside the community, may not be similar to the study group in crucial re-

spects. The laboratory practice for minimizing this danger is to assign subjects randomly to treatment and control procedures. Sometimes this practice can be used in evaluating health programs, but often it will be impossible because treatment must be given to or withheld from whole groups.

Baseline measures are helpful when random assignments to experimental (program) and control groups cannot be made. If baseline measures show that the program and comparison groups were similar at the beginning of the program, one may be more confident that the status of the comparison group at evaluation represents what would probably have occurred in the program group without the program. If the groups differ at the beginning, one should be much less confident.

Where no comparison group can be devised, it may still be possible to obtain information on the probable impact of program activity on the objective. One can, for example, formulate alternative explanations for the outcome of the program and see whether available facts support the alternative explanations. Suppose, for example, one wishes to determine whether a decline in the incidence of tuberculosis in a community can properly be attributed to an ongoing tuberculosis control program, but the community in question cannot be compared with another.

The operator might examine other possible explanations for the falling incidence. He might consider improved nutrition and improved housing as two possible alternative explanations and investigate whether nutrition and housing indeed improved over the period being considered. If neither improved substantially, he could with greater confidence attribute the reduced incidence of tuberculosis to his program. If one or both alternative hypotheses were borne out by evidence, he could not attribute the outcome to his program. At that point he could, however, use the analysis of cross-tabulations to study the interrelationships among the alternative explanations and thus throw more light on the relative contribution of each explanation to the program objectives that were attained (*14*).

The conclusion that program activities caused program outcomes requires a judgment that can never be made with absolute certainty. After using control groups or testing alternative hypotheses, however, one can make a more confident judgment than would be legitimate without the use of such procedures.

Use of Findings

Most evaluations of programs will reveal imperfect success in attaining objectives and sub-objectives. Evaluation does more, however, than demonstrate degree of attainment. It also pinpoints where problems exist. Our model for evaluation of programs assumes that programs have been planned to expend resources to enable activities to be performed and that the activities are intended to cause the attainment of sub-objectives and the program objective.

A program may be less effective than planned for several reasons.

1. Resources were not used as planned.
2. The assumptions linking resources to activities were invalid.
3. Activities were not performed as planned.
4. The assumptions linking activities to sub-objectives or objectives were invalid.
5. The assumptions linking sub-objectives to the program objective were invalid.

Locating program difficulties requires measuring each of three program variables: resources, activities, and objectives and sub-objectives. If evaluation can pinpoint the problems, subsequent program planning should proceed more effectively than it could in the absence of evaluation. Thus, in the hands of the thoughtful administrator, evaluation of program effectiveness can improve planning of programs and thereby increase program effectiveness.

REFERENCES

(*1*) Paul, B. D.: Social science in public health. Amer J Public Health 46: 1390–1396, November 1956.

(*2*) MacMahon, B., et al.: Principles in the evaluation of community mental health programs. Amer J Public Health 51: 963, July 1961.

(*3*) Hutchinson, G. B.: Evaluation of preventive services. J Chronic Dis 11: 497 (1960).

(*4*) Freeman, R. B.: Public health nursing practice. W. B. Saunders Co., Philadelphia, 1963, p. 289.

(*5*) James, G.: Administration of community health services. International City Managers Association, Chicago, 1961, ch. 6, pp. 114–134; (*a*) p. 118.

(*6*) Knutson, A.: Pretesting: a positive approach to

evaluation. Public Health Rep 67: 699–703, July 1952.
(7) Selznick, P.: Leadership in administration, Row, Peterson and Co., New York, 1957.
(8) Bergen, B. J.: Professional communities and the evaluation of demonstration projects in community mental health. Amer J Public Health 55: 1057–1066, July 1965.
(9) Stanley, D. T.: Excellence in public service—how do you really know. Public Admin Rev 24: 170–174, September 1964.
(10) Herzog, E.: Some guidelines for evaluative research. Children's Bureau Publication, No. 378, U.S. Department of Health, Education, and Welfare, 1959.
(11) Selltiz, C., Jahoda, M., Deutsch, M., and Cook, S.: Research methods in social relations. Henry Holt and Co., New York, 1960.
(12) Cronbach, L. J.: Essentials of psychological testing. Ed. 2. Harper & Row, Publishers, New York, 1960.
(13) Kerlinger, F. N.: Foundations of behavioral research, Holt, Rinehart and Winston, Inc., New York, 1966.
(14) Hyman, H.: Survey design and analysis. Free Press, Glencoe, Ill., 1955.

PUBLICATION ANNOUNCEMENTS

Address inquiries to publisher or sponsoring agency.

Standard Methods for the Examination of Dairy Products. 1967; 304 pages; $10. American Public Health Association, 1740 Broadway, New York, N.Y. 10019.

Issues in Community Action Research. Report of the spring research forum on evaluation efforts in three New York City community action programs held on May 4, 1966. 1967; 72 pages; $2. Research Department, Community Council of Greater New York, 225 Park Ave. South, New York, N.Y. 10003.

Principles of Radiation Protection. A textbook of health physics. Edited by K. Z. Morgan and J. E. Turner. 1967; 622 pages; $13.95. John Wiley and Sons, Inc., Publishers, 605 Third Ave., New York, N.Y. 10016.

Programmed Instruction and the Hospital. A report on the use of programmed instruction in health care institutions. 1967; 155 pages; $6. Hospital Research and Educational Trust, 840 North Lake Shore Dr., Chicago, Ill. 60611.

Health Planning: Its past and potential. Health Administration Perspectives No. A5. 1967; 76 pages; $1.50. University of Chicago, Center for Health Administration Studies, 5720 Woodlawn Ave., Chicago, Ill. 60637.

Family-Centered Maternity Nursing. By Ernestine Wiedenbach, R.N., M.A., C.N.M., and Caroline E. Falls, R.N., B.S. 1967; 384 pages; $7. G. P. Putnam's Sons, 200 Madison Ave., New York, N.Y. 10016.

Comprehensive Medical Care and Teaching. A report on the New York Hospital-Cornell Medical Center program. Edited by George G. Reader, M.D., and Mary E. W. Goss, Ph.D., assisted by Barbara Korsch, M.D., with a foreword by David P. Barr, M.D., 1967; 391 pages; $11. Cornell University Press, Ithaca, N.Y.

The Patient, The Hospital, and the Patient in England. Rights and responsibilities under National Health Service. Michigan International Business Studies No. 6. By Arthur F. Southwick, Jr. 1967; 376 pages; $7.50. Bureau of Business Research, Graduate School of Business Administration, University of Michigan, Ann Arbor.

Poverty and Mental Health. Psychiatric Research Report 21. Edited by Milton Greenblatt, M.D., Paul E. Emery, M.D., and Bernard C. Blueck, Jr., M.D. 1967; 174 pages; $4. American Psychiatric Association, 1700 Eighteenth St., NW., Washington, D.C. 20009.

Sewers for Growing America. By Morris M. Cohn, Sc.D. 1966; 202 pages; $10. Certain-teed Products Corp., Pipe Division, Ambler, Pa.

Education and Manpower for Community Health. By Hilary G. Fry in collaboration with William P. Shepard and Ray H. Elling. 1967; 128 pages; $5.95, cloth; $2.50 paper. University of Pittsburgh Press, Pittsburgh, Pa. 15213.

Public Health Nursing. A book of readings. By Dorothy M. Steward, R.N., M.S., and Pauline A. Vincent, R.N., M.P.H. 1968; 504 pages; $5.75. Wm. C. Brown Company Publishers, 135 Locust St., Dubuque, Iowa.

Physician Service Patterns and Illness Rates. A research report on medical data retrieved from insurance records. By Helen Hershfield Avnet. 1967; 452 pages. Group Health Insurance, Inc., 221 Park Ave. South, New York, N.Y. 10003.

A Decade Later: A follow-up of social class and mental illness. By Jerome K. Meyers and Lee L. Bean in cooperation with Max P. Pepper. 1968; 250 pages; $7.95. John Wiley & Sons, Inc., Publishers, 605 Third Ave., New York, N.Y. 10016.

Team Nursing. A programmed learning experience. 4 vols. *Plan of Care: Goals of team nursing; Team leadership; Differential of function; Rationale for assignments;* and *Philosophy of team nursing.* By Russell C. Swanburg, R.N., M.A. 1968; 640 pages; $9 per set. G. P. Putnam's Sons, 200 Madison Ave., New York, N.Y. 10016.

Public Health Service Staff Appointments

Edward J. McVeigh has been appointed assistant to the Surgeon General and director of information of the Public Health Service. He succeeds J. Stewart Hunter, who is now the associate director for public services of the Department of Health, Education, and Welfare Office of Information.

Mr. McVeigh, a native of Jackson, Mich., and a graduate of Michigan State University with a degree in journalism, has been a consultant to the Office of the Surgeon General, and made a survey of information activities of the Public Health Service.

Before he became consultant to the Surgeon General, Mr. McVeigh was manager of the public relations services department of American Cyanamid Company, Wayne, N.J. From 1957 to 1963, he was with Dow Metal Products Company Division of Dow Chemical Company, first as a merchandising manager and then head of the chemical section of its public relations department. Mr. McVeigh has also worked as managing editor on two newspapers—the *Flint News-Advertiser* and the *Dowagiac Daily News*.

James David Isbister, formerly executive officer of the National Library of Medicine, has been appointed executive officer of the National Institute of Mental Health.

Mr. Isbister began his government career in 1960 as a management intern at the National Institutes of Health. After serving a year in the Air Force, he became a management analyst with NIH. From 1963 to 1965 he was assistant to the Assistant Secretary for Administration in the Department of Health, Education, and Welfare.

Mr. Isbister received his B.A. degree with honors from the University of Michigan in 1958 and his M.A. degree from George Washington University in 1966.

A former Woodrow Wilson Fellow, Mr. Isbister is a member of Pi Sigma Alpha, the American Society for Public Administration, and the American Political Science Association. He has also received the Regents Alumni Honor Award from the University of Michigan, the Kappa Sigma National Scholarship Leadership Award, and the William A. Jump Foundation Meritorious Award.

Dr. E. James Lieberman has been appointed chief of the Center for the Studies of Child and Family Mental Health, National Institute of Mental Health.

Dr. Lieberman, a child psychiatrist, came to the Institute in 1963 as a consultant. He has been acting chief of the Center since its creation in early 1967. The Center is the focal coordinating point for Institute activities in studies of child and family mental health.

Dr. Lieberman received his A.B. degree from the University of California at Berkeley in 1955 and his M.D. degree from the University of California School of Medicine in San Francisco in 1958. In 1963 he received an M.P.H. degree from the Harvard School of Public Health while engaged in part-time psychiatric practice. He is a diplomate in psychiatry, American Board of Psychiatry and Neurology, and he received the certificate of training from the American Association of Psychiatric Clinics for Children in 1966.

Dr. Lieberman interned at the U.S. Public Health Service Hospital, Staten Island, N.Y., and was a resident at Massachusetts Mental Health Center and Putnam Children's Center in Boston, and at Children's Hospital of Washington, D.C.

He is a fellow of the American Public Health Association and a member of the American Psychiatric Association, American Association for the Advancement of Science, Group for the Advancement of Psychiatry, and the National Council on Family Relations. He serves on the Board of Directors of the Sex Information and Education Council of the U.S., and the National Child Research Center, Washington, D.C. He is clinical assistant professor, division of psychiatry, Howard University Medical School, Washington, D.C.