

An analysis of the basic principles of evaluation as they apply to current public health practice is developed. The relations between assumptions and validated knowledge are discussed, and the steps to program activity are presented.

EVALUATION IN PUBLIC HEALTH PRACTICE

George James, M.D., F.A.P.H.A.

ACCORDING to the American Public Health Association's definition, evaluation is the "process of determining the value or amount of success in achieving a predetermined objective. It includes at least the following steps: Formulation of the objective, identification of the proper criteria to be used in measuring success, determination and explanation of the degree of success, recommendations for further program activity."¹

It differs from research primarily in that it does not seek new knowledge, but attempts to mark progress toward a pre-stated objective. While research can end with the presentation of results, evaluation is viewed as part of a circular process. Its findings are reincorporated into the specific program from which they came. Evaluation utilizes the same general statistical, epidemiological, and technical methods as research.

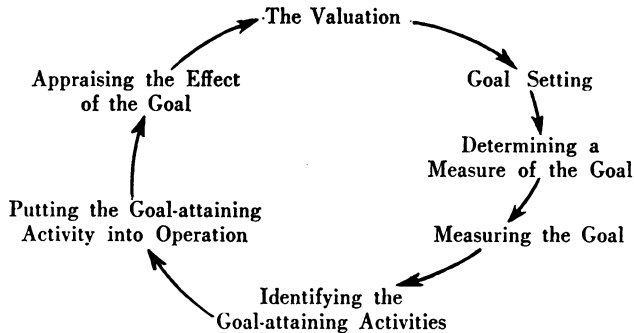
Although evaluation has always been an important concern of health workers, it is no secret that its priority ran well behind that given to the administration of the program itself. So great was our faith in service technics that we begrudged any diversion of effort from them. During the 1930's, when handi-

capped children's programs achieved nation-wide scope, the cry was for more clinics, more children brought to care, more programs offering corrective services. Not one carefully planned, controlled, prospective evaluation study of the long-range restorative power of the program was begun. Tuberculosis case-finding efforts, child health clinics, and child guidance clinics have increased to meet "obvious" demands for such services without the development of methodical attempts to evaluate how well the underlying health needs were being met.

Recently, it began to appear that public health workers intend to pay much more serious attention to evaluation. They are becoming increasingly concerned because their best efforts are not effective against the major health problems of the present era. Of the 20 leading causes of death, only two are now capable of being controlled. We are fighting current health problems fully equipped to win the struggle against those of 1920. The hope for solution must rest in a greater emphasis upon research and demonstration.^{2,3} This costs money, money which must come at least partially from the budgets of some of the traditional but no longer essential programs. It is not

easy to convince both the public and many public health workers that certain services long provided are no longer required.

The evaluation process is a circular one, stemming from and returning to our value system. The steps in the process are outlined below⁴:



A discussion of the practical application of this evaluative process to public health services must begin with a brief comment about the three keystones of program planning: needs, resources, attitudes. A public health need is the lack of a service or program required to protect the public health. To be directly useful as a goal of public health effort it must be rendered in practical terms.

It is not very productive for the health officer to dwell upon the fact that coronary disease is the leading cause of death among males in the United States. Not until he can find a facet of the problem that can be solved by the application of specific resources can he hope to have an impact upon it. In tuberculosis control, for example, the over-all need to reduce morbidity and mortality from the disease must be expressed by certain practical objectives such as (a) the examination of all familial contacts of a case of tuberculosis; (b) the routine x-ray examination of admissions to general hospitals

as a good source of new cases of tuberculosis. These are practical expressions of a need for tuberculosis control in the community, and can lead directly to the formulation of specific objectives and action programs aimed at meeting these practical needs.

In order to describe a need in prac-

tical terms we must possess an available and useful resource. Until such a resource has been developed and its effectiveness proved, research and demonstration constitute our most useful program, unless we wish to proceed by faith alone.

Concern for community attitudes is equally important. In setting objectives we are quite dependent upon our cultural value systems. Often we are forced to yield to pressures exerted by a vocal but uninformed minority, illogical or unscientific as it may be.

We may find ourselves deeply involved in programs of minor value to disease control or prevented from performing an effective health service control because of forceful community attitudes.

Utilizing our knowledge of needs, resources, and attitudes, we then proceed to establish program objectives. The first and most all-embracing of these is the ideal objective or statement of purpose. This defines our ultimate goal

in disease control such as "the elimination of all tuberculosis cases and deaths." Although such a formulation suggests neither a specific set of activities nor a timetable for effort, it does provide us with (a) a reason for our program, (b) a specific end point which defines what we consider to be ultimate success, and (c) a set of mortality and morbidity rate yardsticks against which all other measures of success must some day be validated.

Nevertheless, despite the above defense of the ideal objectives, it is the practical objectives which are the translation from purpose to program, and which make health services possible. Program evaluation consists essentially of the measurement of our success in reaching the practical objectives.

Some students of administration speak of "objectives," "steps," and "activities" as a descending order, with each of the latter terms used to denote action taken to implement a former one. Others, including myself, prefer to use the single term "objective." These objectives are then considered as making up an ordered series, each of which is dependent for its existence upon an objective at the next higher level, and each in turn is implemented by means of lower level objectives. In this framework there is a descending order of objectives beginning at the ideal objective and ending at the lowest level at which the task is to be subdivided. One can consider that, in general, the line officer of highest rank in a health department is responsible for the highest order objective, with each of the succession of lower ranked workers charged with one of the objectives of the descending scale.

Let us use a county dental health program as an illustration. The ideal objective is complete dental care for all children in the county. A high order practical objective for the health officer might be the complete dental care of

school children through a combination of private dental care, school dental corrections, and topical fluoride. His dental director may adopt as his 1961 practical objective the complete dental care of all first-grade children. The school dentist implements this, and his practical objective might be to achieve complete dental care for all first-grade children at the Central Avenue school. His dental hygienist has the task of applying topical fluoride to the first-grade children's teeth after the dentist completes the operative work. Her assistant is responsible for the objective of obtaining parental consent to all dental procedures.

Some may feel that the lower levels in this example should not be dignified by the term "objective." Significant evaluations have been performed at lower levels than these, i.e., how to write letters that bring consent, or how to educate the parent so that he demands dental care for his child.

Program evaluation should be applied to the lowest levels first and then successively to each objective up the scale. After we know the degree to which we have met an objective, this finding then becomes a step toward the next highest objective. If each of the dental hygienists in our illustration does her task satisfactorily, the resulting progress then becomes a part of the program for each school dentist. If each dentist carries out his full responsibility (part direct service and part administration) the combined result satisfies the dental director's objective for the year.

Most of the difficulty in communication about evaluation has occurred because of the confusion between these levels of objectives. Some have felt it sufficient to evaluate a training program by noting that the student has learned his lesson well. Others insist we must first prove that his learning has actually resulted in his doing better work, before

we can state the program has been a success. According to the framework described here, they are both right—they are merely talking about objectives at different levels.

The cement which holds our hierarchy of objectives together, and which is the cause of so much poor communication and argument in the entire field of evaluation, rests on the assumptions we must make whenever we create a new objective. These assumptions are of two essential types, value assumptions and validity assumptions. We will say little here about the value assumptions, although they are a rich field for discussion,⁵ particularly, but by no means exclusively, as we deal with foreign lands and subcultures within our own. Such value assumptions are: the value of saving a human life, our complete lack of interest in the survival of such species as the rat or *M. tuberculosis* compared to our great cultural concern for the dog. They include the value people put on health in relation to other human needs. They are highly significant, but not as treacherous to our practical understanding of evaluation as are the validity assumptions.

An assumption of validity must be made whenever we move from a higher order objective to a lower one. Hence, every lower level objective must assume all of the assumptions we have made for all of the objectives above it in the scale. Any program which is based upon a set of false major assumptions cannot be rescued by its lower level objectives, although quite valid evaluations might still be made for each of them individually. It is possible to evaluate the ways of making a pamphlet more readable, even if the public health facts in the reading matter are false. It was quite feasible to show that mothers could be motivated to feed their babies on a rigid time schedule, even though today we believe that this principle is wrong.

There are only two ways one can move up the scale of objectives in an evaluation: (a) by proving the intervening assumptions through research effort, i.e., changing an assumption to a fact, or (b) by assuming their validity without full research proof. When the former is possible, we can then interpret our success in meeting a lower level objective as automatic progress toward a higher one. Knowing the high potency of tetanus toxoid, we can equate a certain program of immunization to a given level of community immunity. Similarly, we can feel fairly sure that a 1 ppm sodium fluoride concentration in our water supply is a valid expression for a 60 per cent caries reduction among the children drinking it since birth.

When an assumption cannot be proved, we still must attempt to progress upward in interpreting an evaluation, since, as program administrators, none of us wishes to defend low order objectives for their own sake. But we go upward at our peril.

Public health as well as all other community services would be impossible unless validity assumptions were made. It is part of our value system that the population will not forgive our failure to make validity assumptions—in the absence of fact they expect us to use our expert opinion. During a polio outbreak 15 years ago some health officers were severely criticized for not closing schools and swimming pools while others who took these epidemiologically unproved steps received high praise. Perhaps the real objective involved was to allay fear and insecurity which is satisfied by forthright expert action no matter how unproved its effectiveness. Our task today, however, is not to examine the reasons why validity assumptions are made but to emphasize that they are made liberally in every one of our health programs.

Despite such apologies for freely

made validity assumptions, it has become apparent that their cost can come too high. Newer health programs, particularly those in the chronic disease field, could bankrupt any community which wished to proceed too vigorously along the path of current assumptions. Could we afford the cost of all-out obesity control, a continuous program to promote the unsaturated-fat diet, a public program of annual physical examinations? The health officer's answer to this general problem is the demonstration. He tries out such programs on a small scale while awaiting and hopefully contributing to the research solutions required to prove validity.

The only sure answer to this dilemma for the evaluation is to identify each validity-assumption clearly and meticulously both at the time it is made and again at the time of evaluation. Only in this way can we keep clear the lines of communication between the professionals in our field and openly invite research investigation.

A group of people belonging to the same profession tend to make the same assumptions. This situation generally relegates evaluation to a low priority, since a low level evaluation may be too readily accepted as proof of attainment of a higher level objective. Charles Ascher compares this to the story of the emperor's new clothes. Not until a stranger from a far away empire remarked out loud about the emperor's robeless costume did the natives dare believe what their eyes revealed. Public health workers need consultants from fields such as social science to help them become alert to the many assumptions implied in health programs. We must learn other ways of building more dissatisfaction into our programs, to keep us alert, critical, and flexible instead of smug, self-satisfied, and rigid. One such way is to provide for a periodic review by a critical advisory committee with a changing membership.

The establishment of evaluation standards is a task familiar to health workers. A standard is a practical objective, and once established serves as a measure of progress. There is much which is dangerous in this practice. It is difficult to find the source for the fact that 2,400 *Escherichia coli* per ml is the upper limit of safety for beach water, let alone the proof we have of its validity. We must expect that every standard has in it a number of validity assumptions which should be methodically identified, listed, and described at the time the standard is established. Each of the assumptions of the objectives must be identified at each step in ascending order up to and including the ideal objectives dealing with the elimination of disease and the development of optimal health. If we do this we need not argue too vehemently over exactly what the standard should be. We will have provided the challenge for its own eventual refinement. If we do this we will avoid being classified with the members of those learned boards of expert nutritionists during the 1930's who, as McCollum once remarked, spent so much of their time "solemnly passing biologic laws."

Let us use the familiar program of tuberculosis control as an example of how to handle assumptions. The ideal objective is "the elimination of all morbidity and mortality from tuberculosis." Its chief assumptions are as follows:

a. Man's life is worth prolonging. His productivity should be kept high as long as possible, and disease and suffering are to be avoided. This is a value assumption which requires no further justification in our culture.

b. The continued biologic existence of the tubercle bacillus is unnecessary and undesirable. Although partly a value assumption, this is also partly a validity assumption. It is possible that the eradication of tuberculosis could lead to circumstances which would be even more unfavorable for mankind. However, in the absence of such evidence, we must make the assumption.

c. The total physical, social, and emotional

cost of tuberculosis control will be less than that of the disease. This is another part value, part validity assumption, which we can restudy from time to time as our program proceeds.

The next lower level of objectives would include: "The earliest possible detection and isolation of all cases of reinfection tuberculosis." Let us examine the assumptions associated with this objective:

a. The disease spreads from infectious persons to others, hence prompt detection and isolation of these cases will reduce tuberculosis incidence. This is a validity assumption which we will probably never test under controlled-study conditions. We must confess that tuberculosis case and death rates have fallen as rapidly in many areas which have not had specific case finding and case control programs. Moreover, despite all of the years of intensive community x-ray effort aimed at early detection, we still seem unable to increase appreciably the proportion of early stage cases among the total cases reported.

b. Infectious cases may be discovered by modern technics. This validity assumption has been proved as essentially correct as long as one is willing to note the problems of reliability which exist to a predictable degree if only one person reads the chest x-ray.

c. A chest x-ray is a relatively harmless procedure.

A next lower order of objectives of the objective discussed above would be: "The examination by x-ray of all contacts to known cases of tuberculosis." The chief assumption implied here is that this group not only has a higher incidence of infection than average but one sufficiently higher to justify its being singled out for special follow-up. The truth of this assumption may vary from area to area and from time to time. The problem becomes more complicated if we try to compute the changing relative priority of contact examinations as a method of case finding with, for example, the examination of old cured cases, inmates of nursing homes, jails, or general hospitals.

The next level of case-finding ob-

jectives brings us finally to practical goals: "At least one x-ray examination on all (or 80 per cent if one wishes to be even more practical) of the contacts to cases of reportable tuberculosis, and one such examination per year thereafter for those remaining in contact with active cases." Now we have finally arrived at what we commonly recognize as a "standard of recommended practice." If we wish this standard to be considered as a true expression of the ideal objective "the elimination of tuberculosis" we must remember the crucial effect of each of the assumptions we have made so far. In addition, we have made a new assumption in establishing the present objective or standard; namely, that this specific procedure is the most valid, reliable, efficient, and adequate method for detecting these particular tuberculosis cases.*

When one considers all of the assumptions which have been made, it is not surprising that we feel dissatisfied with our final standard. This is all very well, and it is far better that we, as professionals, should be the first and not the last to experience these doubts! Such insecurities lead to a frequent restudy of the problem, a healthy sharing of experience with our colleagues and further refinements of the standard. The alternatives are dogmatism, lack of flexibility and an outdated concern with an armamentarium incapable of solving the health problems of our times.

Let us here emphasize again that if one of the key assumptions of our higher objectives is proved wrong, the standard must inevitably collapse. If, for example, the chest x-ray were found to be a significantly harmful procedure, or if tuberculous disease in a future era

* There is also a special assumption involved whenever we pick a practical goal which is less than 100 per cent of approved practice. We, of course, are implying that the group not reached is substantially similar to the group reached.

became largely endobronchial in site and could not be detected by x-ray, then our illustrative standard would be meaningless. These are not all far-fetched possibilities. Concern with ionizing radiation has already had an impact upon the frequency of the chest x-ray examination of certain groups in the community.

Evaluation in public health can be performed under several categories: effort, performance, adequacy of performance, and efficiency. Effort evaluations are those whose yardsticks and objectives are based either upon the capacity for effort or on the actual effort itself. It is obvious that such objectives and evaluations assume that the specific effort involved is a valid means of attaining a higher order objective of health accomplishment. Effort evaluation is the easiest to perform in public health. As a matter of fact, the best currently available measure of the adequacy of local health services in this country is a very general capacity-to-serve measure: the presence of a sufficient number of specific kinds of qualified health personnel.

The next category of evaluation is performance, which is an end result of effort. Performance occurs at several levels. It could be, for example, the number of cases of tuberculosis found after x-ray, the number of these cases hospitalized or the number cured after hospitalization. It is not generally recognized, but there are a number of key validity assumptions involved in many evaluations of performance. The fact that a number of children are reported as having received complete dental care does not insure that all of this care was done properly and was truly completed. Problems of reliability are common in performance standards, and must be taken into consideration whenever they might be of major significance. For example, the proportion of diabetics diagnosed in a case-finding

program will vary with the blood sugar test used, the age groups tested, the time elapsed after the last meal and the follow-up procedures used. Hence every evaluation standard devised must specify all modifying conditions of significance.

An evaluation of a community health program can be made more meaningful if, in addition to performance, it is able to give some measure of the extent of the total problem solved. Such an evaluation tells us how adequate the program has been in terms of the denominator of total need. Although accurate data describing the unmet need are not generally available, some estimates have been made which are useful. The National Tuberculosis Association has estimated the total unknown cases of tuberculosis in the United States, and the American Diabetes Association the number of those with unrecognized diabetes. The National Health Survey is expected to supply much valuable data on unmet needs for various disabling illnesses. It can be recognized that the ideal objective always includes an adequacy concept.

It might be useful to present an actual example of a state-wide evaluation utilizing the concept of adequacy of performance. The New York State Health Department knew that fluoridation of the public water supply could reduce dental caries among children by 60 per cent and that topical fluoride applied to the teeth of children who were not served by public water supplies could reduce dental caries by 40 per cent.

At the time the evaluation was made, 20 per cent of the upstate New York population had fluoridated water supplies and 30 per cent of the preschool population were receiving topical applications of fluoride. The effectiveness of these methods of preventing dental caries in children can be seen in Table 1.

Table 1—Caries Prevention in Children in Upstate New York, 1954

Method of Fluoridation	A Per cent Children Affected	B Per cent Effectiveness of Procedure	C Per cent Caries Prevented (AXB)
Water supply	20.0	60.0	12.0
Topical application	30.0	40.0	12.0
Total	50.0	—	24.0

Thus it is seen that 24 per cent of dental caries in children were prevented by the state's program at the time of the evaluation.

The study then considered the potential possibility for the prevention of caries. If all communal water supplies in upstate New York were fluoridated, 80 per cent of the population would be affected. Children in the remaining 20 per cent of the area could be given topical fluoride. Table 2 shows how effective such a program would be in caries prevention.

Potentially 56 per cent of the dental caries in children in upstate New York could be prevented. It is known from local studies that the practicing dentists in New York State are able to take care of almost 40 per cent of the caries load and that after fluoridation they appear to continue to fill the same number of dental caries. Hence, the full application of fluoridation in this state plus the regular dental correction program would meet 56 per cent plus 40 per cent of the prefluoridation unmet need and hence would be 96 per cent adequate. The present program is meeting only 24 per cent plus 40 per cent of the total dental caries need among children and hence has an adequacy rating of 64 per cent.

Despite the great value of the evaluation by adequacy of performance, it is not sufficient for the practitioner of public health. Few programs can be justi-

fied at all cost and a measure of efficiency should be included whenever possible. The emphasis on efficiency is closely related to the health officer's attempt to streamline traditional programs. Can the same end result be achieved at lower cost? Can less skilled personnel substitute adequately for physicians, nurses, and engineers? Can self-inspection programs achieve an effective degree of restaurant sanitation control? Such questions point out that standards of performance will be improved if they can consider the effort-costs involved and arrive at a comparative efficiency rating.

Efficiency concepts do dominate many of the public health decisions made in chronic disease programs. When any new test is considered as a possible addition to the armamentarium for mass screening, careful attention must be paid to the number of false-positives which will occur. The individuals who are screened as positive must be followed by more elaborate and costly examination procedures. A screening program which results in a large number of false-positives could rapidly overwhelm the follow-up mechanisms of a community. Hence, any technic which gives a high proportion of false-positives will be discarded for practical use. A history of chest pain as an indication of coronary heart disease, the measurement of obesity as an index of suspicion of hypertension, the presence of low

gastric acidity as a warning sign for incipient cancer of the stomach are all screening tests which have not been widely accepted as practical for health department detection programs because of their high number of false-positives.

In addition to the four categories, every evaluation exercise should also analyze the processes involved in the program. What is it that has made the program succeed or fail? What changes in technics or methods could have improved program effectiveness? Which recipient of the program benefited the most and which the least? What did the program accomplish in terms of originally unforeseen objectives? Each program is a potential target for numerous research questions. The findings of the evaluation study, aimed though they may be at determining progress toward a specific objective, should be analyzed closely to see which additional questions might also be answered.

A few words should be included here on the desirability of building evaluation into the health program. The health officer sets up controls to keep him posted on results as the program proceeds. This feed-back information is another means of stimulating dissatisfaction and allows him the opportunity to institute prompt corrective measures when necessary. This same information can offer valuable data for the performance of evaluation studies. Recently in New York City the early feed-back data on the

cancer detection clinic program indicated that the particularly susceptible population was not well represented in the patient load. Even after the clinic had been moved to an area of the city where these persons resided, they still did not make use of the clinic. Further study revealed that large numbers of women of the high risk group were patients at the department's social hygiene clinics as well as certain hospitals, and the detection program was instituted there. The first year of operation of this changed program found six times as many patients with cancer of the cervix than had been found previously.

Building evaluation into the program permits data collection to proceed at the time the events occur, a far better method than later retrospective search for data of possible significance. In establishing health programs, the general rules for the development of a longitudinal study should be followed, and the evaluation of the results is similar to the hypothesis-proving analysis of longitudinal data in an experimental design.

In conclusion, the following major rules of evaluation should be observed by the modern public health worker:

1. The practical objectives of each program to be evaluated should be clearly stated.
2. The underlying assumptions of validity associated with each objective should be meticulously identified.

Table 2—Potential Caries Prevention in Children in Upstate New York

Method of Fluoridation	A Per cent Children Affected	B Per cent Effectiveness of Procedure	C Per cent Caries Prevented (AXB)
Water supply	80.0	60.0	48.0
Topical application	20.0	40.0	8.0
Total	100.0	—	56.0

3. Evaluations by effort should always be done. Evaluations by performance, adequacy of performance, and efficiency should be done whenever possible.

4. The entire program should be re-examined in the light of the findings arising from the evaluation exercise.

5. To insure the reliability of any standards developed as aids to evaluation, the status of all significant conditions associated with the use of the standard must be specified.

6. The ultimate value of evaluation to public health programs will depend to a great extent upon research proof of the validity of the assumptions involved in the establishment of key objectives.

7. As in every new field there is a period of time set aside for clarification of terms and construction of conceptual frameworks. Further progress will then occur only from

the performance of actual evaluation projects, carefully designed and analyzed. Public health practitioners today need the stimulation which can be achieved only through the critical appraisal of a large number of such studies. The time for such work is now!

REFERENCES

1. Glossary of Administrative Terms in Public Health. A.J.P.H. 50:225-226 (Feb.), 1960.
2. James, C. Research by Local Health Department: Problems, Methods, Results. *Ibid.* 48:353-361 (Mar.), 1958.
3. ————. Public Health Program Planning in Our Present Era. *Nursing Outlook* 8:334-337 (June), 1960.
4. Elinson, J. Lecture at Columbia University School of Public Health and Administrative Medicine—Course Title: Evaluation in Public Health—1959.
5. Vickers, G. What Sets the Goals of Public Health? *Lancet* (Mar. 22), 1958, pp. 599-604.

Dr. James is first deputy commissioner, New York City Department of Health, New York, N. Y.

Environmental Health: Diversity of Research

“100 Problems in Environmental Health: A Collection of Promising Research Problems” represents the suggestions and cooperative efforts of more than 80 scientists, engineers, and physicians in educational institutions, governmental agencies, professional practice, technical associations, and commercial enterprises.

Edited by Jack F. McKee, Sc.D. (professor of engineering in environmental health, California Institute of Technology), et al., this 185-page paperback was made possible through a Public Health Service grant. It covers such fields as air pollution, food science and technology, occupational health, water supply and pollution control, solid waste disposal, and other related areas. The book is an example of the diversity and scope of environmental health's concerns.

Single copies in limited supply from Dr. William Upholt, chief, Research Grants Branch, Office of Resource Development (Environmental Health), Public Health Service, Washington 25, D. C.