Journal of Public Health

The importance of evaluation in public health action cannot be overstated. This is, if possible, even truer in the new fields such as mental health. The need for a clear statement of the principles that must guide evaluation of community mental health programs is exceedingly well served in the analysis offered here.

PRINCIPLES IN THE EVALUATION OF COMMUNITY MENTAL HEALTH PROGRAMS

Brian MacMahon, M.D., F.A.P.H.A.; Thomas F. Pugh, M.D., M.P.H.; and George B. Hutchison, M.D., M.P.H., F.A.P.H.A.

UNFORTUNATELY, there is no accumulated body of experience in the evaluation of community mental health programs from which governing or guiding principles can be drawn. Recently, there have appeared detailed accounts of two plans for the scientific evaluation of a community mental health program. However, one of these—the Milbank Memorial Fund's model for Syracuse[1]—was never intended to be put into effect, and although the other—the St. Louis School program[2]—has reached the operative stage, no results of the evaluation have yet been reported. Therefore, attempts at defining what constitute principles in the evaluation of community mental health programs must depend on knowledge gained from other areas, and discussion of the applicability of such principles in the evaluation of community mental health programs will perforce be theoretical.

Fortunately, we were asked to deal with principles, and not with practical problems. By comparison with the practical problems the principles involved are relatively simple, and briefly stated. They may be classed under one of three general headings:

1. The determination of what type of evaluation is required before designing the evaluatory plan.
2. The definition of the program, the population to be served, and the effects desired.
3. The choice of comparison groups which will permit the inferences required by the type of evaluation selected.

Types of Evaluatory Studies

According to the underlying purpose, evaluatory studies fall into one of two

major categories. The first general category of evaluatory studies consists of those designed to test the hypothesis that a certain practice, if successfully carried out within specified limits, has a measurable beneficial outcome in the group on whom it is practiced; for example, to test the idea that surgical removal of the affected breast leads to a lengthening of life among patients with breast cancer. This process we will refer to as evaluation of accomplishment. Effective studies of this variety are by far the less common of the two types of evaluation. In the field of community mental health they are conspicuous by their absence. Meyer and Borgatta[3] have suggested many reasons for this lack, but the most important appear to stem from problems encountered in studying all cause-effect relationships, exemplified in this instance by the desired cause-effect relationship between the treatment and decreased disease prevalence. Unless the beneficial effects of a therapy or preventive are startlingly obvious, a situation which does not seem to be the case in most mental disorders, the evaluation of accomplishment requires formal experimental methods, including random or systematic allocation to treatment and comparison groups, objective assessment of results, and so forth.

The second category comprises studies designed to find out whether a supposedly therapeutic or preventive practice is in fact being carried out within specified limits—for example, are cancerous breasts being removed in accordance with criteria established as "good surgical practice." This process we refer to as "evaluation of technic." In evaluation of technic, cause and effect are not at issue—the procedure is concerned merely with the description of the quality of the events of which the technic is comprised. Compared with evaluation of accomplishment, evaluation of technic is relatively easy, and much has been done in this area. In the field of surgery, for example, where controlled evaluation of accomplishment is so rare, evaluation of technic is highly developed. On the other hand there are still large areas in which much remains to be done, as witness the singular scarcity of studies such as that of Peterson[4] on the evaluation of the technic of general medical practice.

Hutchison,[5] has previously drawn attention to the distinction between these two types of evaluation, using a different terminology. He distinguished "evaluation of intermediate objectives" (the technic) from "evaluation of ultimate objectives" (the health benefits derived from the procedure). However, as he pointed out, while physicians tend to regard health as an ultimate objective, not everybody shares this point of view; economists, for example, may regard health (including their own) as an objective intermediate to the economic well being of a country.

It is imperative to distinguish the purposes of these two types of evaluatory study. Without any doubt, the first category—the controlled evaluation of accomplishment—is the more vital. Unless it has been shown that the use of a certain technic is followed by beneficial results, what is the use of making sure that the technic is being followed? Once a technic has been shown to be of value, however, continuing studies designed to evaluate the application of the technic become crucial. The personnel and other facilities necessary to the controlled evaluation of accomplishment in a community program are so considerable that it is obviously impractical for them to be built into more than a few pioneering studies. Furthermore, once a technic has been clearly shown to be of value, further critical evaluations may be contra-indicated because of the undesirability of retaining untreated groups as controls. The importance of the initial controlled evaluations is evident since their results will be assumed to be

applicable to much wider populations than those on which they were carried out. The subsequent evaluatory process is simply one to insure that in this wider application the technic as applied in a particular program is sufficiently similar to that applied in the original evaluations to justify the assumption that the same beneficial effects will occur as a result. Since the evaluation of technic does not require the identification of a cause-effect relationship, but merely the definition of the technic, and the observation, recording, and cataloguing of events of which it consists, it is a process that can be, and usually is, built into each individual program.

Before leaving the subject of categories of evaluation and their purposes, there is one concept that should be mentioned, if only to be dismissed. This is the idea that programs gaining "community acceptance" are, ipso facto, beneficial. That this reasoning is a non sequitur should be evident. It is demonstrated currently by the acceptance of well advertised nostrums, and historically by waves of enthusiasm that have been generated by a variety of medical cults.

Principle of Definition

Turning now to the second area of principle—definition—we have no intention of attempting to define community, mental, health or program. No such definition is needed here. What must be susceptible to definition, however, both as to what it is and as to what its intentions are, is the individual program that is to be evaluated. It is immaterial whether or not it falls into one's preconceptions of what the ingredients of a community mental health program ought to be.

A statement of the essence of a mental health demonstration used by Ernest Gruenberg, however, is particularly useful in identifying the type of program that seems capable of being evaluated.[6] From the context of the statement it does not appear that any distinction is intended between a demonstration and a program; indeed the statement might well be descriptive of a variety of service programs in the health field:

> "The essence of a mental health demonstration is that extra resources, extra interagency cooperation, extra consultative services from experts, etc., are made available to an entire population on a scale large enough that observable changes in the population's mental health can be anticipated."

We note from Gruenberg's statement that a community mental health program is made available to an entire population, and that changes are anticipated in the population's mental health. This population need not be defined as a whole population in the demographic sense, but may be restricted in terms of age, sex, occupation, club or school membership, or in some other way. Glidewell and his colleagues,[2] for example, are concerned with the evaluation of a program designed for school children in St. Louis. The Milbank Memorial Fund's model for Syracuse[1] was based on a program for recipients of Old Age Assistance. Depending on the nature of the program to be offered, there may be further specification of the group within the population that is expected to be the particular target of the preventive or therapeutic measures. This specification will usually be in terms of a particular illness or group of illnesses in the case of a treatment program, or of a particularly susceptible group of persons in the case of a preventive program.

Definition of the nature of the program is an area that may be more difficult in community mental health than in some other subject areas. This is unfortunate, because even if some intuitive procedure on the part of an individual or group is evaluated and found to be effective, there is little pros-

pect for the more general application of the procedure until it can be defined in operational terms that are intelligible to a large body of professional workers. Also unfortunate is the fact that many community mental health programs employ a shot-gun approach in which a great number and variety of services are offered at the same time. If such a program is found effective, those who wish to emulate it are required to duplicate each of the component elements, unless they wish to undertake new evaluations involving particular components of the program. A procedure that seems more logical, although requiring greater patience on the part of the investigator, is the separate evaluation of the individual components of the larger program, if indeed these can be separated.

Next, in the context of definition, we come to definitions involved in the assessment of the effects of the program. To return, once again, to Gruenberg's statement, we note that observable changes can be anticipated. In order to evaluate a program it is necessary to state ahead of time the results which are anticipated, and to compare them with the results actually observed. Observation which makes note of all the observable changes which follow the introduction of a program, while useful as a method of formulating hypotheses, does not evaluate the program. Further studies will be required to test any hypotheses developed as a result of such a survey.

Measures of effectiveness may include such objective measures as mental hospital admission or discharge rates, length of stay, rates of truancy, delinquency, divorce, separation or suicide, and incidence of physical illnesses, such as cerebral syphilis and cerebral palsy. They may also include more complex, and at times subjective, psychologic and psychiatric assessments, provided that due precautions are taken to insure assessments that are replicable and are made independently of knowledge of the treatment group to which the patient belonged. Wherever opinion, whether of the patient or of a professional observer, forms part of the basis for evaluating the program, it is necessary for that opinion to be validated. It is not enough, for example, to take the patient's statement that he feels better at its face value. It is necessary to demonstrate that the holding of such an opinion is correlated with some change in the patient that can be objectively assessed as harmful or beneficial. The same is true of the feeling of a professional worker as to the benefit derived by the recipients of his service.

Gruenberg's concept that the results should be observable seems a very reasonable one, and yet one may wonder how far this principle has really been accepted in the community mental field. We cannot escape the impression that many of the practical problems in evaluating mental health activities, that one hears so much about, originate in the difficulty of identifying a characteristic (whether of people or of communities) that can be shown to change under the influence of a mental health program. From this difficulty it is inferred that our problem is one of identifying a "satisfactory" method of evaluation. We should not, however, overlook the alternative explanation—that the evaluation is in fact satisfactory but that there is no appreciable change to identify.

Certain concepts that are heard of with some frequency in the field of community mental health, although definable in the abstract, cannot be measured even in such simple terms as bigger or smaller, better or worse. Here also, difficulty in evaluation can be anticipated. Programs having as their objective a higher level of "positive mental health," for example, stand no more chance of being successfully evaluated than would a program of venereal disease control that was based on the ob-

jective of fostering positive genital health, however admirable such an objective may seem.

Principle of Comparison

An understanding of the principles underlying the selection of comparison groups is particularly important in studies that purport to evaluate accomplishment.

Ideally, we would like to know what would have happened to the population to whom a particular program was offered had there been some other program, or no program, during the same period. This is, of course, impossible to determine, and we must be satisfied with the assessment of the trend of events in some other population that has not been exposed to the program being evaluated but which is otherwise as similar as possible to the exposed group. This similarity, of the compared populations, the crux of the problem, can be assured only by random or systematic allocation of individuals or groups to treatment and comparison samples. Probably the most satisfactory procedure of all is the comparison of treatment and control groups randomly assigned from individuals offering themselves for the program. However, since the services in community mental health are frequently offered to groups rather than to individuals, this course may be impossible to put into practice. Where programs are restricted to small components of a population, such components may be randomized, as for example in the St. Louis study, where 15 schools have been assigned to various treatment and comparison groups in a systematic way. Another method, illustrated by the Milbank Memorial Fund's model for Syracuse, is that of comparing two or more large districts or cities, again selected randomly or systematically. This last method may be forced on us by the community-wide nature of many community mental health services, but it is less satisfactory than the others since there is the danger of changes occurring coincidentally with the program in one area but not in the other. However, as we have seen from the fluoridation experiments, the cumulative evidence from a number of such studies may be quite convincing.

In a large number of evaluatory studies use is made of data collected incidentally during, and without interference in, the course of a program. Comparison groups either are absent or consist of some readily available groups whose features of similarity to or difference from the experimental group are not precisely known. The attempt is to obtain some evaluatory information without the inconvenience of planning special evaluatory studies. To the extent that routine records are used for the purpose of evaluating technic or of describing prognosis for persons availing themselves of the service they may be perfectly satisfactory. In the minds of many investigators, however, it seems that estimates of prognosis among service participants are viewed as substitutes for the evaluation of accomplishment. While one would not discourage periodic analysis of individual program records at the present time, one should stress their limitations, and particularly discourage the idea that such analyses can be used to evaluate the accomplishment of a program.

Conclusion

Lemkau and Pasamanick,[7] preface their discussion of the problem of evaluation of mental health programs with the saying "Any fool can ask a question; the trick is to ask one that can be answered." We do not believe that we have been unrealistic in our statements as to what can and should be done. Rather, it seems to us that, despite the practical difficulties, community mental

health programs are in many ways in unusually advantageous positions to be evaluated, particularly in terms of evaluation of accomplishment.

Of recent years serious doubts have been expressed by competent investigators as to the efficacy of surgical treatment of cancer in prolonging life. The accumulated cost of surgical therapy for cancer in money, personnel, and human misery (not to mention lives) over the past decades is astronomical. Regardless of whether such therapy is or is not effective, it is tragic that the evidence does not exist, and probably cannot now be obtained, to lay these doubts to rest. The organizers of mental health programs should not repeat the mistakes of our surgical predecessors. In the first place, they have before them the experience of the surgeons, and others, from which to learn. Second, mental health programs are being born into an era when critical evaluation is a recognized part of medical thought, and when the professional, technical, and financial help necessary to the operation of effective evaluatory studies is not difficult to obtain. Third, regardless of how firmly one is convinced of the effectiveness of a particular program, the ethics involved in leaving certain groups untreated do not arise, since there are not enough facilities to provide for everybody. Therefore, the major problem in evaluating the accomplishment of any health program—that of providing appropriate comparison groups—can readily be met, if an effort is made to select areas for application of such programs in a systematic and purposive way.

REFERENCES

1. Milbank Memorial Fund. "Model for a Project Which Would Demonstrate the Prevention of Mental Disorder in an Aged Population." In Planning Evaluations of Mental Health Programs. New York, N. Y.: Milbank Memorial Fund, 1958.
2. Glidewell, J. C.; Mensh, I. N.; Domke, H. R.; Gildea, M. C. L.; and Buchmueller, A. D. Methods for Community Mental Health Research. Am. J. Orthopsychiat. 27:38, 1957.
3. Meyer, H. J., and Borgatta, E. F. Paradoxes in Evaluating Mental Health Programs. Internat. J. Soc. Psychiat. 5:136, 1959.
4. Peterson, O. L. An Analytical Study of North Carolina General Practice, 1953-1954. J. M. Educ. 31:1, 1956.
5. Hutchison, G. B. Evaluation of Preventive Services. J. Chronic Dis. 11:497, 1960.
6. Gruenberg, E. M. "Progress Report to the Advisory Council on Mental Health Demonstrations; Development of the Concepts Underlying Mental Health Demonstrations." In Planning Evaluations of Mental Health Programs. New York, N. Y.: Milbank Memorial Fund, 1958.
7. Lemkau, P. V., and Pasamanick, B. Problems in Evaluation of Mental Health Programs. Am. J. Orthopsychiat. 27:55, 1957.

The authors are associated with the Department of Epidemiology, Harvard University School of Public Health, Boston, Mass.

This paper was presented before a Joint Session of the Epidemiology and Mental Health Sections of the American Public Health Association at the Eighty-Eighth Annual Meeting in San Francisco, Calif., November 2, 1960.