PARADOXES IN EVALUATING MENTAL HEALTH PROGRAMMES

HENRY J. MEYER

The University of Michigan

AND

EDGAR F. BORGATTA

Russell Sage Foundation

IN recent years, attention to mental illness has extended beyond the problem of custody, and a variety of services and programmes have developed external to the mental hospital. Partly in response to the cost of custodial care, prevention on the one hand and rehabilitation on the other have become important goals. These new dimensions have now increased in commitment of effort and money to the point where they constitute a major enterprise. They have also become interwoven with existing social welfare programmes.

Thus, social agencies originally intending to serve families with domestic problems, or children in need, or special disadvantaged groups, may now concern themselves with the mental health of their clients from a prophylactic viewpoint, or because mental illness has directly or indirectly affected those with whom they deal. Likewise, educational and recreational programmes relate themselves to mental health, sometimes through deliberate efforts, and more often indirectly, in the belief that their existing activities are relevant. Indeed, the mental hygiene movement that started as an attack on problems directly associated with mental hospitals has become so diffuse and pervasive that it is impossible to draw its boundaries and define its purposes, agencies and activities in any simple way.

In this context of interest, the importance of evaluation is broadly recognized. The meaning given to evaluation, however, is varied; it is sometimes confused, and frequently oversimplified. If persons are making a devoted effort, it is hard for them and for those sympathetic to their objectives not to feel that a contribution of some sort is being made. Therefore, industrious and sincere application to the problem is often taken as the equivalent of achieving successful results as a consequence of the effort expended. But the magnitude and importance of programmes in mental health as well as in other social services require that evaluation go beyond appraisal of effort in this sense.

Although need for evaluation is recognized, it is rarely taken as an occasion to examine the efficacy of programmes or techniques against objective tests of successful results. Evaluation in the mental health field occurs at different levels, implicating different sets of values. At one level, evaluation is equated with

satisfaction of the recipients or the purveyors of the service. If they like it, it is good; if they are dissatisfied with it, it is not. At a second level, evaluation may mean a comparative but subjective appraisal of one service or activity among several that are offered, often by "experts" whose criteria are presumed to have wider relevance than those directly involved in the activity. It has been pointed out in one comprehensive review of evaluation in mental health that, at this level:

> ". . . No precise yardstick is used to make the evaluation. Although there is danger that hunches, hypotheses, personal bias, and faulty judgment may enter into the evaluation process, this method is recognized and supported by the community and government."

A third level of evaluation is that of scientific measurement involving standardized and logical methods of assessment, widely recognized in other areas, but seldom applied in mental health or social welfare. Evaluation in this sense is often tedious, difficult and undramatic. Nevertheless, it is only through the rigorous and painstaking effort required for scientific evaluation that information necessary for sound appraisal of even the most conscientious effort can be made.

Evaluation at this third level requires more stability and security for the agencies involved than evaluation at the other two levels. It is unlikely that those offering services, or communities supporting them, will tolerate scientific evaluation unless they feel fairly secure about their achievements. Otherwise, rigorous and detailed evaluation of certain aspects of a total programme may constitute an intolerable threat. It may, indeed, be harmful rather than constructive if it fails to acknowledge the commitment of the agency to its objectives. However, as pointed out in a U.S. Public Health Service publication, "If used properly and interpreted correctly, this type of evaluation is of great help in improving the overall programme, and it is also useful in gaining public support for expansion or continuation of the programme."

It may be assumed, in view of the magnitude of effort and investment made in mental health programmes and activities, that scientific evaluation is now overdue and can be used constructively. It is with this sort of evaluation that the following discussion is concerned.

SCIENTIFIC EVALUATION IN MENTAL HEALTH AND SOCIAL WELFARE

There is nothing especially magical or esoteric about evaluative research. It occurs in many different fields, and the principles that are involved are common to all. In essence, evaluative research requires only the simple notion of a control group design, where subjects who are exposed to treatment are compared to subjects who receive no treatment (or a known, different treatment), and where the subjects for both groups are selected on a random basis, and presumably do not differ on any essential criteria prior to assignment as treatment or comparison cases. Evaluation is simply an attempt to give a logical answer to the question: Has a given effort, and that alone, resulted in an observable and desirable outcome?

Sometimes the investment of effort is of such dramatic return that evaluative research is superfluous. This is not the case in the field of mental illness,

where there is almost the opposite situation. The necessity of giving attention to the patient is itself often considered of prime significance, so that no concern is given to finding out whether success did indeed result from the effort. Efficacy of effort in social welfare programmes, however, is not demonstrable in any clear-cut manner by obvious effects. Enthusiasm and satisfaction cannot be substituted for evaluation.

A review of the literature of evaluation in the fields of mental health and social welfare would prove disappointing. There are but few studies that have given attention to rigorous research design, and to adequate samples. A committee of the National Advisory Mental Health Council made an extensive survey of published and current research and reported its findings in a recent book entitled *Evaluation in Mental Health*. Among almost 1,000 titles, very few represent scientific evaluative studies. The committee makes the following comment:

> "In reviewing the literature the committee has been impressed with the quantity of research studies. . . . However, the numbers of studies embodying adequate methodology for scientific validation of the efficacy of treatment procedures are few. Absence of control groups in studies on the results of treatment seems to be the rule. . . ."

The paucity of scientific evaluation in mental health is not a result of ignorance or stupidity of researchers nor of stubborn refusal to accept the canons of science. It probably results, first of all, from the enthusiasm of effort that cannot wait for calm, dispassionate assessment. In the second place, it reflects the fairly undeveloped state of knowledge about mental illness in general. Conceptualization of mental health and illness, diagnosis, theories of ætiology, methods of treatment and control, and, indeed, the very description of disease processes, are far from systematic. Recently a number of efforts have been made to place knowledge about mental health in the perspective required, if services and programmes are to be based on more than limited and partial conceptions of the problem. This is particularly the case with respect to social, as compared to medical, attacks on mental illness.

There have been many publications representing polemic exhortations, case-materials indicating the interplay of patient and therapist, descriptions of how a given psychotherapeutic procedure is carried out, subjective appraisals of the presumed consequences of therapeutic intervention, and general expositions of theories about personality, social conditions, situational factors and mental health. Study of concomitants of mental illness and recovery has taken place in scattered and disjointed fashion rather than cumulatively. Few systematic reviews of these scattered reports have appeared but many pieces of prognostic information have been published. A comprehensive study of the literature has recently been undertaken by Joseph Zubin and his associates at the New York State Psychiatric Institute. Such reviews should provide a fund of knowledge in organized form that may be suggestive of many hypotheses for understanding factors associated with mental illness, and those associated with treatment.

A third reason why scientific evaluation both in mental health and in social welfare has been rare arises from some difficulties inherent in the character of programmes in these fields, and these are now discussed.

BARRIERS TO EVALUATIVE RESEARCH IN SOCIAL WELFARE AND MENTAL HEALTH

If the principle of control group design which lies at the base of evaluative research is easy to understand, its application in the practical situation is a much more difficult matter. Many factors serve to hinder the implementation of such research in practising social agencies. Among these are the following:

(1) In order to be effective, practitioners often feel that they must maintain confidence in their own procedures. An attitude of confidence when dealing with the patient or client is often viewed as a prerequisite to good practice, and this attitude involves belief in the efficacy of treatment procedures used. In the light of this attitude, professional defensiveness against having the efficacy of treatment brought under question is understandable. When several professions are involved, as is usually the case, the problem becomes even more delicate and complex.

(2) Agencies providing services must depend on public support and are almost always on limited budgets. In part, the reputation of an agency may depend on its appearance of smooth and efficient operation. Evaluation research design may require an interruption in the flow of cases into the agency and hence represent a visible threat to the agency's operations. Thus, such research may become not only a burden of unanticipated expenses, but it may also appear as a diversion of resources within the agency. In addition, it may place a strain on the ordinary administrative and operative procedures. Thus, it may appear to undermine the basis upon which the agency seeks support from the public.

(3) Evaluative research may expose the realistic operations of the agency in contrast to its public appearance. At the manifest level, the work of an agency may appear focused on its public objectives. However, a considerable amount of effort may go into incidental services for a wide variety of persons not conceived to be the principal objects of interest. Because evaluative research forces attention to what is actually done rather than what is intended, it may force the agency to reveal and to recognize its multiple functions. Any organization that has existed for some time will ordinarily acquire many secondary functions beyond those for which it has been established. To make a public acknowledgment of these activities may well be upsetting, and resistance to evaluative research may arise from this basis.

(4) Other barriers to evaluative research in social welfare agencies arise from some of the specific demands that the research design itself may create. Difficulties from this source tend to centre on a number of focal issues.

Evaluative research requires that specific control or comparison groups be designated and utilized as standards of success. This involves the entire question of defining what is meant by subjects suitable for agency attention and then designating a segment of them for service, and another segment for no service. Thus, the practitioner and the agency who assert the efficacy of their efforts may well decry the fact that useful services are being withheld from some.

The designation of a control or comparison group not only requires the identification of the relevant subject population, but it requires that the control group be selected by the same criteria used to select the treatment group. This,

in effect, means that clients are arbitrarily assigned to these groups and the privilege of choosing those whom the practitioner believes "would best respond to treatment" is lost. Once more, this challenges the judgment of the practitioner and places on him the burden of proof as to his success. Random assignment to treatment and comparison groups is necessary because, without it, it is impossible to attribute the differences that may occur between the groups to the treatment procedures rather than to possible selective factors. The effects of selection cannot be handled on a *post hoc* basis without the very knowledge that evaluation research is intended to yield.

The control group design requires that patients assigned to the experimental group actually receive treatment. If the entire group does not receive treatment, the question arises as to whether those who do were selected on some unknown but relevant basis and hence were more (or less) likely to respond to the treatment. While it is not necessary that *every* patient receives the treatment, the preponderant group must; otherwise the generalization possible is restricted. Thus, an agency must often take an aggressive approach to its clients. This may seem inconsistent with the viewpoint that clients must voluntarily want services for the treatment to be effective. The only alternative to aggressive attraction of experimental subjects would be denial of services to some of those voluntarily seeking them. This would be even more at variance with other values of the helping professions.

The design of evaluative research may embrace the totality of an agency's activities rather than the specific therapeutic procedures the agency sees itself as practising. In general, practitioners tend to think they have specific skills, and that these produce the results they believe they achieve. When one comes to measure service or treatment in an agency setting, however, the procedures may be so interwoven with administrative, custodial, and other operations that it is impossible to determine whether results are attributable to the treatment procedures or to the complex within which they occur. Hence, agencies may feel that evaluation misses its mark when, in fact, effects result from activities not at all conceived to be a part of treatment efforts.

Finally, evaluative research requires that practitioners articulate what they mean by successful results of treatment. The criteria by which success is to be judged need not be of any specific type so far as evaluation is concerned, but they must be identifiable, so that they can be applied in a standardized way to the relevant cases in both the experimental and comparison groups. Because success is always a quantitative assertion, evaluation of success requires that research indicate how many, how much, or how much more, of some given change has occurred in the experimental group when compared with the control group. What are "good" or desired changes, on the other hand, depend on value-judgments that designate a specified criterion or set of criteria of success. If these are not supplied by the practitioner in his own terms, they must be devised by the researcher. If this occurs, the practitioner may feel that the meaning of success and failure is imposed on him.

This can become a serious difficulty when certain viewpoints towards therapeutic success are taken by practitioners. Sometimes, particularly in psychotherapy, the practitioner asserts that each case is unique, and hence no criteria

of success can be generally applied. This position denies that general knowledge is used in treatment. If this were the case, evaluation would indeed be impossible, but, likewise, it would be impossible to develop any principles of practice or treatment. On other occasions, the practitioner may assert that only he can have the intimate knowledge of his patients or clients necessary to permit meaningful evaluation. This is tantamount to claiming private, perhaps omnipotent, judgment as a requisite for evaluation, and runs counter to the scientific canon of communicable, public knowledge. Finally, practitioners sometimes feel that success is the equivalent of what has happened to their clients; in other words, they take a retrospective view and identify perceived change with desired change. With such a viewpoint, scientific evaluation becomes impossible and meaningless except in the unlikely event that no changes at all take place.

The barriers to evaluative research that have been mentioned are not offered as criticisms of agencies carrying on programmes in mental health or social welfare so much as descriptions of some of the problems that must be faced. Basically, these problems reflect conflicting values of practitioners on the one hand and scientists on the other. But improvement in practice—whether in engineering or medicine—has been made when scientific methodology is employed in pursuit of the goals of practitioners. The so-called helping professions—social work, psychotherapy, rehabilitation, etc.—must ultimately follow the model of the older professions, such as medicine, and encourage, rather than resist, objective assessments of their activities. Those who engage in scientific evaluation are, in turn, obligated to accept the valued goals of the agencies and programmes they evaluate, and to acknowledge the good faith and sincerity of practitioners.

One obvious cause of difficulty in carrying out scientific evaluation in the social welfare and mental health fields is unfamiliarity of the agencies and practising professions with research requirements. Therefore, early evaluation studies must be interpreted cautiously for a number of reasons. In the first place, agencies may make a special effort when evaluation is taking place, and therefore what is evaluated may be unusual rather than typical. Contrariwise, the work of the agency may be somewhat disrupted by evaluative research, and therefore the results may reflect less than usual achievement. Furthermore, evaluation seldom encompasses the total activities of an agency, and extreme care must be taken to specify exactly what it applies to. Negative findings about a part of a programme or an aspect of an agency's efforts should not be taken as an indication that other aspects are not effective. Certainly, the whole programme should not be brought into question when only one part of it is under evaluation.

Despite the problems that beset scientific evaluation, there can be no alternative to facing them, and attempting, in the best possible spirit of mutual understanding, to conduct such research. The difficulties should be recognized in order that they can be overcome.