

50685

CHAPTER 21

OUTCOME STUDIES IN MENTAL HEALTH PROGRAM EVALUATION¹

Wm. A. Hargreaves, C. Clifford Attkisson, University of California, San Francisco;
and Frank M. Ochberg, National Institute of Mental Health, Rockville, Maryland

Few community mental health centers systematically examine client outcomes (Hargreaves et al. 1975). The first program evaluation concern is usually a broad monitoring of the process of the center. This includes assessing community needs, describing the utilization of various services by different types of clients, meeting funder reporting requirements, and integrating client and staff statistics with the accounting system in order to carry out cost finding and rate setting. Usually it is only after these tasks are begun that attention shifts to examining outcomes for the individual clients served by the center programs. We have suggested elsewhere that this developmental sequence seems appropriate (McIntyre, Attkisson, and Keller 1974; Attkisson et al. 1974; Hargreaves et al. 1975). Regular, effective use of outcome studies is an advanced stage in the maturation of program evaluation capability in the management of a center. This general rule should not blind the reader, however, to the many possibilities for creative use of outcome information in program management and planning. The six examples which follow illustrate a few of these possibilities, ranging from simple monitoring to advanced program evaluation.

Some Examples

Monitoring Clients' Level of Functioning

A county mental health facility records a global rating of each client's level of functioning at admission and at each subsequent contact. Changes in this rating from beginning to end of service are tabulated in relation to the type and cost of treatment received. This provides a periodic overview of the performance of the center so that changes from year to year can be documented. Client subgroups that tend to show worsening have been identified

1. Preparation of this chapter was supported by Research Scientist Development Award MH-36809 (Dr. Hargreaves) and contract NO1-MH-2-0105 (OP) with the National Institute of Mental Health, Office of Program Planning and Evaluation.

and steps taken to improve the quality and continuity of their care. The breakdown of costs and treatment type according to intake level of functioning helps in budgeting for new services, such as an increased load of former State hospital patients. The availability of this very simple and inexpensive outcome measure will also allow tentative cost-outcome comparisons of different treatment approaches. Such comparisons may suggest opportunities for program improvement, or may raise questions about the allocation of program resources which can be answered by a planned, time-limited study using more adequate measures of outcome (Rotenberg, Gordon, and Underhill 1974).

Followup Using Volunteers

A group of volunteers who worked with the inpatient unit of a community mental health center were concerned about the lack of aftercare facilities for patients who had been hospitalized. The center's program evaluator encouraged this citizen group to carry out a followup study, which they did. The evaluator helped them develop a simple questionnaire, and all patients discharged during a 3-month period were included. Members of the citizens' group visited as many as possible of these former inpatients 6 months after their discharge. They were appalled by the lonely, isolated life situation of many of the people they visited. The citizen group was motivated by this experience to develop a proposal for a new aftercare program and effectively demanded that the center shift its priorities in this area. In this example, the information-gathering function of the outcome study was secondary to its function as a community organization effort that developed a vocal constituency for a particular program change (Landsberg and Hammer 1975).

Second Chance for Daycare Failures

In a large daycare program, 250 chronic schizophrenic patients were seen twice weekly in groups of 25 to 30. The staff had identified 24 patients whom they considered "treatment refractory." It was decided to explore two new approaches to

these patients. The 24 patients were randomly divided into two groups of 12. Group A was offered a program which featured socialization and refreshments within a group setting. Group B was offered closeness with the therapist through individual supportive psychotherapy but permitted emotional distance from fellow patients. Patient responses during an 18-week period were compared by clinic attendance, rated degree of increased socialization, and cost efficiency, all of which favored group A. Both groups did somewhat better in the new programs than during a prior 18-week baseline, but attendance progressively fell in group B. Group A patients were no longer considered "refractory" at the end of the study. The approach used with group A therefore seemed promising for future use with "refractory" patients in the daycare program. This is an example of a very simple outcome experiment aimed at program improvement (Donlon, Rada, and Knight 1973).

Feedback to Therapists

Applicants are assessed in an outpatient clinic using Goal Attainment Scaling (described in chapter 23 of this volume). The method makes it possible to assess the degree to which each client has attained an individually defined set of treatment goals. The attainment of the goals was judged by an independent interviewer who saw the client several months after the beginning of treatment. At one point the clinic therapists received feedback comparing the outcome of their own clients with the distribution of outcomes for the clinic as a whole. It was found that the therapists receiving this feedback showed a subsequent improvement in their average outcomes. This suggests that the feedback of client outcomes in this setting may have been a useful part of the continuing inservice training and supervision of outpatient therapists (Kiresuk 1975).

Who Accepts Group Therapy?

A large urban outpatient clinic had innovated a "contact group" to which clients could come without an appointment, as frequently as every day if they wished, to talk about their problems or simply to listen to others. Some therapists saw the contact group as the treatment of choice for certain clients, but most staff saw it as a holding strategy, an "active waiting list" prior to treatment assignment, and a dumping ground for poorly motivated applicants. Intake interviewers also expressed a more general concern that they lacked clear criteria for assigning applicants to individual therapy, group therapy, or contact group. To explore these issues, they developed a comparative experiment in collaboration with an experienced evaluator. Applicants

who preferred not to participate, or from whom random assignment was judged to be unsafe, were excluded and were generally seen in individual therapy. The remaining applicants were randomly assigned to individual therapy, weekly group therapy, or contact group. The acceptance of this referral was judged by whether the applicant kept his first appointment and, if he did, by client and therapist ratings of satisfaction following this first appointment. Results suggested that the acceptance of treatment could be improved by not initially assigning poorly motivated applicants to either type of group therapy. More surprising, however, was the finding that shy, distressed clients accepted contact group best, while the more articulate and outgoing clients more readily accepted the traditional ongoing weekly group. Clients assigned to the "wrong" type of group therapy showed poor acceptance of the referral, while the "correctly" referred group showed initial acceptance as great as comparable clients assigned to individual therapy. While the results said nothing about ultimate treatment outcomes, it did identify a specific group which might benefit from a contact group. It also told the intake interviewers that any group assignment for a "poorly motivated" client had almost the same effect as refusing treatment. This is an example of a relatively complex experiment aimed at improving the match of clients to services (Hargreaves et al. 1974).

Does Community Treatment Cost Less and Achieve More for Clients?

The community mental health center movement is based in part on the premise that severely disturbed patients can be treated more effectively through early crisis intervention in or near their home community than in a State hospital. The staff of the Singer Zone Center in Rockford, Illinois, found an opportunity to test the validity of this premise. During its first years of operation, it served only 40 percent of the population of its region, while the remainder continued to be served by three traditional State hospitals. All first admission inpatients in the region admitted during a 6-month period were screened, and a matched pair of samples of State hospital and zone center patients was selected for study. All patients were interviewed every 6 months during the subsequent 4 years. The center group had fewer days hospitalized, less disability, less total cost per patient and per capita, as well as a lower cost-outcome ratio than the State hospital group. This study is reported in another published paper in which the authors discuss a number of program changes at the center which were based on specific findings. An ambitious study such as this

can only be undertaken by a center with excellent program evaluation capability (Smith, Kaplan, and Siker 1974).

Uses of Outcome Evaluation

These examples convey the variety of functions that outcome evaluation may serve. In some cases the issue to be studied requires a carefully designed controlled experiment, but more often a simpler approach is better. Some programs gather simple outcome information routinely on a selected sample of their clients. More frequently, outcome information is collected to answer a specific, time-limited question that has arisen in the management of the program. Funding agencies have also undertaken outcome studies or required routine reporting of outcome data. This has most often been true in the drug abuse and alcoholism areas, but there are examples in general mental health as well, usually undertaken by some statewide mental health authority. In this chapter, however, we will restrict our attention to the use of outcome evaluation within the management of community-based programs.

In the above examples the reader will have noticed that a wide range of outcomes was examined. Improvements in the psychological, interpersonal, and vocational functioning of the clients are among the ultimate outcome goals of nearly every mental health effort. Yet, often, a specific study will focus on intermediate outcomes which are easier to measure, such as client satisfaction, rehospitalization, or acceptance of a recommended treatment. These intermediate goals may be valued in themselves, but in part their value derives from assumed relationships to ultimate improvement in client functioning. One may choose to focus on intermediate or long-term goals depending on the management issue that motivates the study. If an important decision must be made soon, it may be better to have *some* information about the intermediate effects of a treatment than *no* information about its effects. If the issue is a recurring one, it may be better to invest in a study which can obtain a balanced picture of both intermediate and long-term client outcomes, since the intermediate outcomes alone may be misleading.

Program goals include more than individual client goals, however. For example, a community mental health center attempts to be accessible to all residents of its catchment area without regard to ethnic background or ability to pay for services. Centers that are located in poverty areas often have goals of employing catchment area residents and providing career ladders for indigenous staff. Some management issues may motivate a study of the program

outcomes which indicate the attainment of these goals. While the comments which follow focus mostly on the study of individual client outcome, the program evaluator will want to attend to the full range of program goals.

The issues of costs and relative cost-outcome ratios are implicitly present in every outcome evaluation. The last of the six examples illustrates the way one program evaluator made an explicit cost-outcome comparison. Clinicians, and program evaluators who trained originally as clinicians, sometimes feel reluctant to consider cost comparisons because in their experience cost-conscious administrators have seemed too ready to reduce service quality in order to save money. But the program manager has to face the task of integrating quality and economy, and this trade-off process must be recognized in planning outcome studies if these studies are to be relevant to realistic program alternatives.

We see three aspects of program management which may motivate an outcome study. These are: ongoing monitoring of program quality; demonstrating program effectiveness; and making program modifications aimed at program improvement. While some of the above examples show how a study may contribute to more than one of these areas, each motive leads to a different allocation of program evaluation effort.

Monitoring Client Outcome

In routine monitoring, one examines indicators which describe the program's functioning and compares these indicators to some previously formulated objective, expectation, or norm in order to detect problems needing remedy or situations requiring closer examination. Every manager and every clinical supervisor utilizes a variety of sources of information for monitoring the operation of his or her program. When a program grows beyond a small face-to-face group, skilled managers find that periodic statistical summaries are needed to keep the program activities visible and manageable. Data on the outcomes of the program are a useful part of this monitoring. Management information systems should be designed to provide continuing information on the attainment of program objectives. This is not difficult for many program objectives and intermediate client outcomes, but data on ultimate client outcomes are relatively expensive to gather and are usually obtained on only a sample of clients (Cooper 1973; Elpers and Chapman 1973).

Only the most basic measures of client functioning are appropriate for routine collection and statistical summary. The Global Assessment Scale (see following chapter) is a good example. It is a single,

global rating of client functioning and might be thought of as a modest step beyond the "improved/unimproved/worse" ratings traditionally required in a clinical discharge note. It can be used economically by clinical staff at the first and at subsequent treatment contacts for all clients. This not only provides information about change during treatment but also a comparison of the level of functioning of clients entering each of the center's programs. This is valuable for planning purposes, but also allows the selection of comparison samples matched for initial level of functioning. For children we can as yet recommend no such global rating scale, but some usable parents' and teachers' rating scales are available (Conners 1973). Other adult and child rating instruments are discussed in the next chapter in this volume. Finally, client satisfaction ratings can be economically collected in a variety of ways and are also suitable as a program monitoring technique.

A number of mental health, drug abuse, and alcoholism programs have attempted routine monitoring of client outcomes using some form of individual goal attainment assessment (Davis 1973). The best known of these methods is Goal Attainment Scaling, described in Kiresuk and Sherman (1968) as well as chapter 23 of this volume. The basic idea is to formulate one or more treatment goals, unique to each client. These goals are scaled in terms of specific outcome events which will define the extent to which each goal has been reached. The attainment of each goal then can be scored quite economically and reliably, and the results combined into an indicator of the extent to which a given client has attained his or her own goals. Different clients, or groups of clients, can be compared in terms of this goal attainment score. As a monitoring technique, Goal Attainment Scaling is of special interest because it can be integrated effectively into treatment planning and clinical supervision and can be a part of negotiating an initial treatment "contract" with the client. Houts and Scott (1973) report a series of controlled studies in which the use of clear, time-limited treatment objectives in the supervision of treatment staff sometimes improved the attainment of these specific treatment objectives. We know of at least one instance in which initiation of the goal-setting procedures resulted in a considerable reduction in case conference time (Shaw, personal communication). Goal Attainment Scaling manuals are now beginning to appear for use in specific treatment programs, such as methadone maintenance (Putnam et al. 1973) and mental health services for children (Ricks et al. 1973). In spite of these promising advantages, it represents a major investment in staff orientation and continuing training to

install Goal Attainment Scaling as a routine procedure. Centers should be very cautious about goal attainment methods as a monitoring approach unless the responsible clinical staff are ready to incorporate the goal scaling into their everyday operation. Several centers that use an individual goal attainment method report enthusiasm about its value as a routine part of a clinical program. However, if one's task is to install a routine monitoring procedure across a heterogeneous set of decentralized treatment programs, caution is recommended in beginning with Goal Attainment Scaling. It requires a level of enthusiastic support from treatment staff that will usually not be possible to enlist under such circumstances.

With an array of client outcome indicators available, a sensible approach to treatment monitoring will match the indicators to the functional components of the treatment process. To illustrate this idea, consider a simplified model of the methadone maintenance treatment of heroin addicts. Four components or stages of this treatment might be called induction, maintenance, growth, and voluntary withdrawal. Each individual client could be classified at any particular time into one of these components. Goals during *induction* might be regular attendance and a steady reduction in the incidence of urine tests positive for morphine and other illicit drugs. The *maintenance* goal might be continued program participation without positive urine tests and a stable social adjustment. *Growth*, by contrast, would aim for specific improvements in vocational and interpersonal adjustment. Finally, *withdrawal* would focus on maintaining a stable social adjustment and continued abstinence from illicit drugs, while achieving a decreasing dose of the maintenance drug. These outcomes are routinely monitored in methadone maintenance programs, though sometimes quite superficially. Organizing the program into functional components linked to specific treatment interventions, emphasizing different outcome indicators in different components, then provides a rational framework for allocating both treatment and program evaluation resources. In this example, induction and maintenance may require only the routine program monitoring procedures, while both an increased treatment effort and a more detailed outcome measurement effort, such as Goal Attainment Scaling (see chapter 23) or the Social Adjustment Scale (see chapter 22), may be focused on the growth and withdrawal components. The key to this approach is to identify client groups with relatively homogeneous goals and identify the corresponding functional components of the service delivery system, whether or not these correspond to existing organizational components.

Demonstrating Program Quality

This topic is close to the hearts of program directors. With the inadequate evaluation capability of most treatment programs today, the director must deal with advisory boards, county government, funding agencies, and third-party payors with no straightforward, uniformly accepted way of portraying the quality of the program. This is becoming an increasingly difficult issue as these agencies raise their expectations for program accountability. The technical problem in demonstrating program quality is the usual lack of a relevant comparison against which to judge the program's effects. When an ongoing program is not changing and must serve everyone who applies, it is usually impossible to rigorously demonstrate that it has any effect. Yet there are some practical ways to approach this accountability issue.

Our favorite definition of program evaluation is this: Program evaluation is a process of making reasonable judgments about program effectiveness and appropriateness based on systematically gathered information. Scientific proof, in its most rigorous sense, is one very special basis for judgment but is a very expensive and cumbersome standard to be applied to program evaluation, except by gradual approximation. A first rudimentary step in this approximation is to identify and operationalize the goals and objectives of each component of the service system. For publicly supported mental health, drug abuse, and alcoholism programs, many general goals have been embodied in legislation and funding agency regulations. For community mental health centers, Spaner and Ellsworth (NIMH 1973) have compiled these public goals and developed a proposed set of indicators. Other groups are also compiling such sets of overall goals. Since there is usually good consensus about service goals at this most general level, these existing efforts should be reviewed before expending staff or advisory board time in formulating goal statements *de novo*.

From general goals one needs to formulate specific objectives. These state that the program aims to accomplish a particular improvement by a specific date (Mager 1972). Attainment of this improvement will be assessed by one or more measurable indicators specified in the objective. The set of current program objectives then defines an evaluation task to assure that the relevant indicator data are available as needed to monitor the attainment of the objectives.

A baseline of previous performance provides a comparison that allows more precision in setting objectives. For example, one goal of a publicly supported facility is to be accessible to the relevant

potential client groups in the community. In monitoring the characteristics of people who are utilizing services and comparing this to, say, census data, a particularly underserved group may be identified (e.g., older people, poor people, or residents in a particular geographic area). Further investigation suggests that there are specific barriers preventing the underserved group from utilizing needed services. Therefore an objective is set to increase the utilization of services by this group. The effectiveness of subsequent efforts to remove barriers to service will then be evaluated by comparing later client utilization statistics (and census data) to the baseline information.

The above example concerns a program outcome rather than a level of client outcome. Objectives and their related measures have often been restricted entirely to this type of program outcome, and adequate attention has not been given to direct measures of client outcomes. Even very simple data, such as the Global Assessment Scale ratings mentioned in the discussion of monitoring, would enable a program to gather baseline data on client change and help to focus some program objectives on improving client outcomes.

Comparisons between programs, or with normative groups of programs, would also be helpful in setting realistic program objectives and evaluating their attainment. A few descriptive program characteristics gathered by Federal agencies such as the NIMH Biometry Branch are available, but client outcome norms do not exist. The comparison of client outcomes from one program to the next involves some problems, but it would be instructive for a few groups of collaborating programs to experiment with it. One interprogram cost-outcome comparison method is being tested by the State of California Department of Health (Hanson 1975). The extensive outcome data collected by federally funded alcoholism and drug abuse programs may also provide useful pilot experience. As a practical tool for routine use in demonstrating program effectiveness, however, interprogram outcome comparisons are still in the future.

Studies To Aid Decisions

It is the authors' impression, after reviewing program evaluation efforts in many mental health centers, that the best payoff of good program evaluation capability derives from the evaluator's contribution to specific decisions. These decisions may be internal to the program, or involve external decisions, e.g., a funding agency's decision about whether to fund a new program component. The payoff is possible because the evaluator has relevant

information or the skill, tools, and staff to quickly gather or review such information, and a good understanding of the decision process, through his or her direct participation in the management of the program. The evaluator and other program leaders are then able to respond to emerging decision situations with more information than would otherwise be possible. This means that the evaluator must be ready to respond in a timely way by mounting simple "studies" or retrieving and effectively "packaging" existing data. These special studies usually consume only a small part (say 10 percent) of the center's program evaluation resources. Most of the effort is devoted to planned ongoing data collection and report generation, but one function of these ongoing activities is to provide the resources to respond to emerging decision issues.

These ad hoc studies are not just a matter of the evaluator responding to requests for information. It is our impression that program management rarely uses even available information to good advantage unless the evaluator is struggling with these emerging decision issues along with the other program leaders. When all are actively looking for ways in which new information would be useful, the probability of practical payoff from program evaluation is greatly increased.

The studies which emerge from this type of evaluation capability will at first be restricted to simple, brief undertakings. Some of these will include the examination of client outcomes, but many will not. As the evaluator and the other center leaders gain skill in these brief studies, they may be able to consider a few substantial comparative studies of client outcomes, studies focused on major program decisions. The remainder of this chapter deals with planning such studies.

Two types of program decision seem most likely to be aided by a comparative outcome study. The first is a decision to offer a new or changed service, and the second is the ongoing decision process of assigning clients to the best currently available treatment. In the former case, when a new procedure is instituted to improve results with a particular problem or client group, one may choose to run both the new and the old procedure for a time, assigning clients to both procedures and comparing treatment acceptance and outcome in each group. The second type of question arises when alternate treatments are available and the choice between treatments is currently made without any clear or convincing rationale, at least for some portion of the clients. Here a simple comparison of two treatments will not do, since one treatment may be better for some clients, while the opposite may be true for other clients. To examine this question, one

or more "predictor variables" are tested to see whether they can be used in selecting treatment assignment so that client acceptance and outcome are improved.

A useful step in ascertaining the appropriateness of an outcome study is to outline the flow of clients through the functional components of the center's service programs and identify the major points at which choices are made about what services to provide. Selecting several of these choice points, the evaluation planner should discuss the actual decision process with the personnel involved, attempting to understand what questions are currently most important to these decision makers. Are there client groups who seem poorly served no matter what service approach is used? Is a new approach being considered for them? If so, what is the readiness of program staff, center management, and funders to attempt a new approach in a context in which its effectiveness can be compared to the existing approach?

It is also fruitful to inquire about the process of matching clients to the best service approach among those already available. Are there types of clients for whom the staff say they are unsure about the best approach or often disagree with each other about treatment assignment? Are any client groups essentially being assigned at random to different treatments? By exploring these issues the evaluator can judge when there may be an opportunity to improve the effectiveness of the program with the aid of a comparative outcome study.

If an outcome study is to influence a specific decision, it is essential that the decision makers participate in the design of the study. The decision maker may be a member of the county board of supervisors, the mental health center director, the head of a particular program, or a group of front-line staff workers. The decision maker must be considering two questions: "What are the actual options or alternatives which are possible to consider? What information would be sufficient to justify choosing this alternative over that one?" (Harper and Babigian 1971). In planning the study, one test of its potential usefulness is to consider the possible study results and ask the decision maker(s) what course of action they would take in each case. If the actions are the same regardless of the study results, then obviously the study should be redesigned or abandoned.

Planning a clinical experiment will naturally also involve the program staff who will participate in the study or otherwise be affected by it. Experienced evaluators have noted that successful clinical experiments often have a stormy organizational course during their early phases (Glaser and Taylor 1973; Weiss 1973). It seems that this vigorous interaction

is necessary to develop collaboration and avoid later disastrous misunderstandings.

Ten Issues in Planning an Outcome Study

There are 10 issues which must be resolved when planning a study. If the clinical participants and the decision-making users of the study also review and critique these specific details, the study will have a better chance to run smoothly and have maximum impact. These 10 issues are:

1. Selection of the treatment comparison
2. Assignment to treatment
3. Selection of subjects
4. Informed consent procedure
5. Description of the sample
6. Description of treatment received
7. Selection of predictor variables
8. Selection of outcome measures
9. Analysis of data
10. Feedback of results into the decision process

These topics are discussed in a number of texts, such as Riecken and Boruch (1974). However, the basic ideas are straightforward and are summarized below.

Selection of the Treatment Comparison

The most important issue is the relevance of the proposed treatment comparison to future management decisions. There are some practical limitations which should be kept in mind, however, in selecting the service strategies to be compared.

A common error is to compare too many different treatments. Two or three treatments will usually be enough to handle in one study. Beyond this one usually must continue the study too long in order to attain sufficient numbers of subjects in each group. How many subjects are enough? It is a waste of time to carry out a study which has little statistical power to detect a difference between treatment groups. The size of sample needed is related to the size of effect which would be clinically important and to the variability of outcomes *within* the treatment groups being compared. Derogatis, Bonato, and Yang (1968) discuss sample size issues in relation to drug studies with schizophrenics, and the reader may find their article a useful introduction to this topic (see also Overall and Dalal 1968; Cohen 1969). Rarely will it be worth doing a comparative experiment with fewer than 20 subjects per treatment group. In some situations much larger sample sizes are needed. In estimating sample sizes, one must also realistically anticipate some subject attrition during the study.

These sample size considerations limit the feasible complexity of the study.

Assignment to Treatment

The basic task is to assign clients to treatment groups so that differences in outcome can be attributed to the treatment assignment rather than to preexisting differences in the client groups. Random assignment is a traditional, and conceptually simple, way to achieve this but gives rise to apprehension among clinicians, community groups, and clients. Random assignment should be carried out only with adequate safeguards through careful subject screening and informed, uncoerced consent. Nevertheless, many evaluators prefer to avoid randomized trials in favor of quasi-experimental designs (Campbell and Stanley 1969) even in situations where randomization is feasible. The choice should be based on a judgment about the trade-off between the risk to the study subjects and the benefit to future clients that the study may bring about. It is our view that avoidance of the issue of random assignment often means that some basic issues about the value and the design of the study are not being faced by the study planners. For example, if it seems unethical to randomly assign a particular group of clients to two alternative treatments, it may mean that insufficient thought has been given to the criteria for screening out clients for whom the study is inappropriate. If all concerned are in agreement about the proper treatment assignment for all clients, there is probably no need for the study. If clinicians are not clear about treatment assignment for some subgroup of clients, this subgroup may already be assigned randomly, or at least capriciously, to treatments.

There are many situations where randomization is not possible, and other situations where quasi-experimental comparisons can easily be made and will provide useful information. In these cases, one can attempt a partial control of preexisting subject differences through matching and covariance techniques. A quasi-experimental approach is often taken when evaluating a major change in a service program. Outcomes are measured for a group of clients before the change and again after the new program has begun. We have seen instances when a mental health center was able to institute a new program on a trial basis and successfully justify its continued funding on the basis of this type of outcome comparison. Various approaches to such quasi-experimental studies have been discussed by Campbell and Stanley (1969).

Selection of Subjects

Criteria for exclusion of subjects should be agreed upon in advance as much as possible, and each

potential subject should be screened for appropriateness prior to being invited to participate. Reasons for exclusion should be recorded in each case.

Informed Consent Procedure

This is actually a larger issue regarding the protection of subjects, of which informed consent is a key component. Study planners and program directors should be familiar with *The Institutional Guide to DHEW Policy on Protection of Human Subjects* (DHEW 1971). Biomedical and social research practices have recently come under sharp scrutiny by public interest groups. It may be wise to collaborate with members of the center's community advisory board or other representatives of the larger community in developing specific informed consent procedures. Informed consent is appropriate whenever a client's activities in the program are affected by research or evaluation beyond those procedures which are undertaken as a routine and necessary part of his or her treatment. Random assignment is always such an occasion, but so are many other research and evaluation activities. For example, it is now generally considered inappropriate to undertake followup evaluations after treatment unless clients give their consent for later followup while they are in treatment and also consent at the time the followup is carried out. They must be free to refuse without prejudicing their further treatment (American Psychological Association 1973).

Description of the Sample

It is usually helpful to record some demographic and prognostic characteristics of subjects in the study and a limited amount of information about clients considered but screened out of the study, clients refusing to participate, and subjects leaving the study before its completion. This allows one to assess the similarity of the different treatment groups, which is important even if they were randomly assigned to treatments. It also allows a description of the ways in which study subjects differ from other clients in the program.

Description of Treatment Received

To judge the relevance of a treatment study to other settings, a clear description of the treatments being compared is essential. This may be an unnecessary frill in simple studies that are entirely for local consumption, but, if a major investment has been made in a study, it should be published so that other programs have a chance to benefit as well. More important than a global verbal description of treatment philosophy are records of the specific service activities carried out and the extent

to which clients actually accepted and participated in them.

Selection of Predictor Variables

In a treatment comparison study, predictor variables are ordinarily omitted, although they may be used to increase the sensitivity of the study to treatment differences and to generate retrospective hypotheses about the types of clients who benefit most from each treatment. In treatment *selection* studies, however, predictor variables are the major focus. Variables are examined which are thought to predict *differential* suitability for the specific treatments being compared (as distinct from prognostic variables which predict outcome regardless of treatment). In seeking predictor variables, one approach is to ask clinicians who are involved in making the treatment assignment decision under study to nominate client characteristics which they feel may be relevant.

Selection of Outcome Measures

Often, simple global outcome measures will suffice if they can be rated reliably, and if the expected treatment effects are moderately large and sample sizes are adequate. If the number of available subjects is limited or outcome differences are expected to be subtle, then measurement sensitivity can sometimes be increased by supplementing the global scale with a multiple-item rating scale or with Goal Attainment Scaling and by using more than one independent rater (Derogatis et al. 1968). Be sure the measures include the range of outcomes felt to be important by the decision makers who are the intended consumers of the study findings.

Analysis of Data

The major statistical analyses should be planned before the study is begun, since this may lead to changes in the way that data are collected. When the study objectives have been clarified and the overall study plan developed, it is often helpful to review the plan with a statistical consultant.

If the study utilizes several outcome measuring instruments, it is still useful to include a simple global outcome rating. In the data analysis, this global rating is analyzed first to provide a preliminary overview of results as soon as possible. This will ordinarily maximize the usefulness of the study in the decision process and may engage decision makers in raising further questions to be explored in more detailed analyses. In planning a study, one should anticipate that data analysis will often require an effort equal to the effort invested in data collection. Users of the study results, however, will often assume that findings will be available as soon

as all the data are collected. A complex data analysis may so delay the presentation of results that this factor alone removes any possibility that the results will be utilized. If this seem likely, it may be better to simplify the study, so that it attempts to cover only the most important aspect of the decision issue.

Feedback of Results Into the Decision Process

This is the step at which traditional clinical research has often faltered. In program evaluation, the groundwork for implementation should have been laid during the initial negotiations of the study design. The evaluator should be aware of the crucial decision points for management, when study results are most likely to have an impact. These crucial occasions include budget preparation, submission of annual plans, development of grant proposals, and reports to the board of directors. Data should be analyzed and displayed in the context of such management choice points. Clinicians and other program staff should also be given frequent feedback about the progress of the study and of the data analysis. Finally, integration of study results into program planning will be enhanced if the evaluator is an ongoing advocate for consideration of the study findings in the management of the program.

Conclusion

We began by saying that outcome evaluation is an "advanced stage" of program evaluation. Some experienced evaluators feel that outcome studies are a program evaluation "frill," and will not give a useful return as a direct contribution to program management. In the present state of the art of program evaluation, it is hard to argue against this position. Most centers lack the evaluation and the management skill and experience to apply existing outcome study techniques to program planning issues. As these skills are gradually developed in the field, and relatively simple "standard" methods evolve to deal with common evaluation issues, more centers will be able to utilize outcome evaluation.

Although examples of useful outcome evaluation are not plentiful, there is another point of view which asserts that practical approaches must be found. The public human services resist rational management because the effects of programs are usually unknown. Improving human services requires that program effects be made more visible. This is part of our common task as human service program evaluators. In community mental health, outcome evaluation is an essential component of our efforts to monitor program effects.

REFERENCES

- American Psychological Association. *Ethical Principles in the Conduct of Research With Human Participants*. Washington, D.C.: American Psychological Association, Inc., 1973.
- Attkisson, C.C., McIntyre, M.H., Hargreaves, W.A., Harris, M.R., and Ochberg, F.M. A working model for mental health program evaluation. *American Journal of Orthopsychiatry*, 44: 741-753, 1974.
- Campbell, D.T., and Stanley, J.C. *Experimental and Quasi-experimental Designs for Research*. Chicago: Rand-McNally, 1969.
- Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*. New York: Academic Press, 1969.
- Conners, C.K. Rating scales for use in drug studies with children. *Psychopharmacology Bulletin*, Special Issue: Psychopharmacology of Children, 24-84, 1973.
- Cooper, E.M. *Guidelines for a Minimum Statistical and Accounting System for Community Mental Health Centers*. DHEW Pub. No. (ADM)74-14. Washington, D.C.: National Institute of Mental Health, 1973.
- Davis, H.R. Four ways to goal attainment. *Evaluation*, Special Monograph Number 1:1-28, 1973.
- Derogatis, L.R., Bonato, R.R., and Yang, K.C. The power of the IMPS in psychiatric drug research: As a function of sample size, number of raters, and choice of treatment comparison. *Archives of General Psychiatry*, 19:689-699, 1968.
- DHEW. *The Institutional Guide to DHEW Policy on Protection of Human Subjects*. DHEW Pub. No. (NIH)72-102. U.S. Government Printing Office (stock No. 1740-032b), 1971.
- Donlon, P.T., Rada, R.T., and Knight, S.W. A therapeutic aftercare setting for "refractory" chronic schizophrenic patients. *American Journal of Psychiatry*, 130:682-684, 1973.
- Elpers, J.R., and Chapman, R.L. Management information for mental health services. *Administration in Mental Health*, 12-25, Fall 1973.
- Glaser, E.M., and Taylor, S.H. Factors influencing the success of applied research. *American Psychologist*, 28:140-146, 1973.
- Hanson, M.R. A California five-county cost-effectiveness study. In: Zusman, J., and Wurster, C. eds. *Program Evaluation: Alcohol, Drug Abuse, and Mental Health Service Programs*. Lexington, Mass.: Lexington Books, 1975.
- Hargreaves, W.A., Attkisson, C.C., McIntyre, M.H., and Siegel, L.M. Current applications of evaluation. In: Zusman, J., and Wurster, C., eds. *Program Evaluation: Alcohol, Drug Abuse, and Mental Health Service Programs*. Lexington, Mass.: Lexington Books, 1975.
- Hargreaves, W.A., Showstack, J., Flohr, R., Brady, C., and Harris, S. Treatment acceptance following intake assignment to individual therapy, group therapy, or contact group. *Archives of General Psychiatry*, 31:343-349, 1974.
- Harper, D., and Babigian, H. Evaluation research: The consequences of program evaluation. *Mental Hygiene*, 55:151-156, 1971.
- Houts, P.S. and Scott, R.A. To evaluate the effectiveness of achievement motivation training for mental patients being rehabilitated to the community. Unpublished project final report, Pennsylvania State University College of Medicine, Hershey, Pa., 1973.
- Kiresuk, T.J., and Lund, S.H. Process and outcome measurement using Goal Attainment Scaling. In: Zusman, J., and Wurster, C. eds. *Program Evaluation: Alcohol, Drug Abuse, and Mental Health Service Programs*. Lexington, Mass.: Lexington Books, 1975.
- Kiresuk, T.J., and Sherman, R.E. Goal attainment scaling: A

- general method for evaluating community mental health programs. *Community Mental Health Journal*, 4:443-453, 1968.
- Landsberg, G., and Hammer, R. Consumer feedback research in mental health: A review of an agency's experience. *Evaluation*, 2(2):19-20, 1975.
- Mager, R.F. *Goal Analysis*. Belmont, Calif.: Lear Siegler, 1972.
- McIntyre, M.H., Attkisson, C.C., and Keller, T.W. Components of program evaluation capability in community mental health centers. (Chapter 2 in this volume.)
- NIMH, "Program Compliance Items for Community Mental Health Centers' Deficiency States, and Measures or Questions." Draft paper. Mental Health Services Support Branch, Rockville, Md., October 1973.
- Overall, J.F., and Dalal, S.N. Empirical formulae for estimation of sample size in clinical research. *Psychometric Laboratory Reports*, No. 1. University of Texas Medical Branch, Galveston, 1968.
- Putnam, D.G., Kiesler, D.J., Bent, R.J., and Stewart, A. Goal attainment scaling training manual. Drug Abuse Services Section, Georgia Department of Human Resources, Unpublished manuscript, 1973.
- Ricks, F., Clarke, F.G., Dlugacz, D., Prato, D., Rothwell, P., Shaw, R.C., and Sowa, E. Manual of case evaluation from the Dellerest Children's Centre. Dellerest Children's Centre, Downsview, Ontario, M3J2E5. Unpublished manuscript, 1973.
- Riecken, H.F., and Boruch, R.F. eds. *Social Experimentation*. New York: Academic Press, 1974.
- Rotenberg, L.A., Gordon, M., and Underhill, W. The day one treatment, client management, and cost/outcome project. The Mental Health Treatment Center, Reading Hospital, Reading, Pa., unpublished, 1974.
- Smith, W.G., Kaplan, J., and Siker, D. Community mental health and the seriously disturbed patient: First admission outcomes. *Archives of General Psychiatry*, 30:693-696, 1974.
- Weiss, C.H. Between the cup and the lip. *Evaluation*, 1:49-55, 1973.

END