American Handbook of Psychiatry

Evaluation of Behavior Change and Therapeutic Effectiveness

EBERHARD H. UHLENHUTH

Evaluation of Behavior Change and Therapeutic Effectiveness

Eberhard H. Uhlenhuth

e-Book 2015 International Psychotherapy Institute

From *American Handbook of Psychiatry: Volume 5* edited by Silvano Arieti, Daniel X. Freedman, Jarl E. Dyrud

Table of Contents

Evaluation Of Behavior Change And Therapeutic Effectiveness

Introduction

Evaluation is the new bon mot in mental health. We share this preoccupation with social programmers (Rossi, 1972) on the one hand and with physicians in general (Brook, 1973; Schulberg, 1969) on the other hand. The status of the issue in our field is reflected by the recent appearance of several monumental overviews that are rich sources for the serious student (Bergin, 1971; Franks, 1969; Levine, 1971; Meltzoff, 1970).

The growing interest in evaluating treatments in psychiatry signals a developing consensus that we now have real therapeutic alternatives requiring choices among a variety of effective procedures, probably even differentially effective in different problem situations. So the focus is shifting gradually from a reductionists search for the common element in psychiatric therapy to ferreting out clinically significant differences between treatment tactics (Frank, 1973; Klett, 1965; Strupp, 969).

Among the burgeoning variety of new pharmacological agents and psychological techniques, presumably only a few will offer significant advances. This provides another motive for careful evaluation, which the public deems so pressing that it finds institutional representation in such

governmental agencies as the Food and Drug Administration. The high cost of health care, including psychiatric care, is another public concern generating pressure to justify therapeutic interventions on the basis of relative benefit and expense. The increasing support of psychiatric care through public (governmental) and private (insurance) organizations accelerates this trend. Current social values and economic conditions seem to promote an emphasis on accountability throughout the social structure.

Under these pressures it is important to distinguish clearly, as Suchman does, between ". . . *evaluation* as the general process of judging the worthwhileness of some activity regardless of the method employed, and *evaluative research* as the specific use of the scientific method for the purpose of making an evaluation" (1967). In the past, we often have taken the position that only the clinician's intuitive, impressionistic evaluation of the effects of his interventions can do justice to the richness and subtlety of human function in the social context. In such highly complex situations, unfortunately, even professional observation and judgment, unaided by the signposts of a systematic plan, all too frequently seem to lose their way (May, 1971).

A pertinent example concerns our persistent tendency to confuse treatment effects and patients' inherent potential for change: we continue to select candidates for *treatment* on the basis of characteristics that predict

spontaneous improvement. From the start of our clinical training we learn to select patients with an eye to their probable (gratifying) response, which powerfully reinforces us in our daily practice. Unfortunately, this process takes place under complex circumstances that consistently and without our awareness prevent us from sorting out specific *effects* that contribute to a patient's *response.*

The rest of this discussion deals with evaluative research: substantive knowledge, principles, and procedures useful in minimizing such unwitting errors of observation and judgment in evaluating treatment effects. The distinction between evaluating the effects of a program and evaluating the effects of a variable is an important issue (Suchman, 1967). The results of program evaluation may be valid only for the special circumstances of that program. Properly designed evaluations of the effect of a clearly defined variable often have more general validity and contribute understanding of pathological and restorative processes. The unique potential of this entry point to basic knowledge about human function is itself a justification for evaluative research (Klein, 1969).

Discussions of evaluative research usually list several principal areas of interest (Donabedian, 1966; Zusman, 1969): (1) effort—the accounting of services rendered and resources employed; (2) outcome —the results of the effort expended; (3) process—the mechanism by which the effort produces

the outcome; and (4) cost—the expense entailed. Meaningful evaluation of treatments probably requires some attention to each of these closely interwoven topics. The present discussion emphasizes especially outcome as the central point of departure for the evaluation of treatments.

Criteria of Outcome, Change, and Effectiveness

Over the past several decades, criteria of outcome, change and therapeutic effectiveness have received the enormous amount of attention they deserve (Bergin, 1971; DiMascio, 1972; Federal Drug Administration, 1974; Group for the Advancement of Psychiatry, 1959; Hogarty, Mimeo; McNair, unpublished; May, 1968; Meltzoff, 1970; Waskow, forthcoming; Weissman, 1972). Whereas investigators often must concern themselves especially with precision of measurement (reliability), clinicians' interest centers on adequately representing the range of human functions that are the object of therapeutic influence (validity). The measurement of "intrapsychic processes" or "character structure" is a particular issue. The Menninger Foundation's Psychotherapy Research Project (Kernberg, 1972; Sargent, 1961) well illustrates the current potentials and limitations in this area. Clearly, highly skilled observers can make direct, quantitative estimates of such global, highly conceptualized dimensions as ego strength.

Psychometricians approach the same problem in a more empirical,

atomistic fashion with formal test measures of personal "traits" (Cattell, 1965) and pathological personality trends (Dahlstrom, 1972). Unfortunately, these measures often seem rather insensitive to differential treatment effects.

A less ambitious framework for criteria grows rather pragmatically from roots in the medical model. The final criteria of therapeutic effectiveness become reductions in discomfort and ineffective function (Park, 1965). Other criteria that are often proposed, such as continuation in treatment or development of insight, take their places as possible mediating variables in the process of treatment. This framework underscores the relationship of psychiatric treatment to general health care, spans the pharmacotherapeutic agents and the "proprietary" psychotherapies with an atheoretical bridge (Strupp, 1969), and points toward specific measures to make the criteria operational. Finally, many of the measures derived from this framework do indeed discriminate treatment effects.

Discomfort, in particular the states of anxiety and depression, is essentially a subjective phenomenon. The patient's verbal reports constitute the principal source of data, although the skillful observer integrates these with the patient's concomitant nonverbal behavior. Measurement of this domain merits particular emphasis in patients with neurotic and affective disturbances. In addition measurement of the broader symptom picture is essential in any evaluation of treatment.

Ineffective function refers to social-role performances that, at least in principle, are forms of behavior open to direct observation by others. These include occupational productivity, effectiveness as a spouse and parent, adequacy in social relations, and such basics as hospitalization vs. community tenure, self-care in the hospital, etc. Although some of these functions may be measurably impaired by any psychiatric disturbance requiring treatment, role-performance criteria are central in therapeutic studies of severe, especially psychotic, conditions.

Both comfort and effectiveness may be viewed by various observers, and each has had advocates: patient, therapist, nurse or other professional, relatives. The therapist's assessment of the patient's clinical condition is the major factor in evaluation by reason of his skill and his knowledge of the patient. The patient's assessment is an important addition in work with adult neurotics, who have predominantly subjective complaints and are relatively reliable observers. Nurses, relatives, and other outside observers can contribute valuable assessments in studies of young, old, psychotic, addicted, and psychopathic individuals.

The "independent research assessor" (Guy, 1967) represents a viewpoint of special interest in certain situations. He can bring to the assessment professional skills without therapeutic investment; and he may be kept blind with regard to the details of treatment. The cost is less

familiarity with the patient.

Curiously enough, assessments by different observers (e.g., therapist and patient) using the same measuring instrument differ more than assessments by the same observer using different instruments (Park, 1965). Assessments by multiple observers in an evaluative study provide broader coverage of relevant changes and tend to compensate for the unique biases of each.

A plethora of instruments have been developed for all types of observers, rating every facet of comfort and effectiveness. Comprehensive efforts to sort out this bewildering array of outcome measures are now under way by the Clinical Research Branch, NIMH, for application to psychotherapy (Hogarty, Mimeo; Waskow, forthcoming), and by the Bureau of Drugs, FDA, for application to pharmacotherapy (DiMascio, 1972; Federal Drug Administration, 1974; McNair, unpublished; Weissman, 1972). We may anticipate substantial overlap between the efforts of the two groups. Also available and under continuing development by the Psychopharmacology Research Branch, NIMH, are comprehensive batteries of well-established instruments linked to a complete data-processing system (Guy, 1970; McGlashan, 1973).

Despite the effort that has been expended in developing highly specific

measuring instruments, the simple seven-point global rating of status or change by the therapist remains one of the most sensitive methods of discriminating treatment effects (Lipman, 1965). Unfortunately, it offers no information as to the nature of the therapeutic effect. In a strict sense, this measure may not provide comparable information about different patients, as the rater is free to accent the most relevant features in each individual.

Highly structured measures of clinical condition can be criticized on the grounds that they may be irrelevant to many patients. An approach to this problem is to rate "target symptoms" defined individually with each patient (Battle, 1966). This concept applies quite naturally to behavior modification with its emphasis on changing specific behavior.

The timing of assessments is as critical as the instrument and the observer in discerning certain treatment differences. Chlorpromazine is preferable to lithium in treating acute, highly active manics partly because of its rapid action (Prien, 1972): assessments after the first week of treatment do not reveal this difference in the time of onset of action. Assessments repeated at regular intervals have the particular advantages of permitting trend fitting to study the course of change over time (Dixon, 1969) and to provide a more stable estimate of the individual's response based on all this data. Repeated measurements over time also offer an approach to assessing character traits, defined as characteristic tendencies to exhibit recurrent

states or behavior.

On the other hand, assessment over extended periods, as in follow-up studies, presents special problems in interpretation since treatment during the interim cannot be controlled (May, 1965; Stone, 1961). In such cases only criteria applied before uncontrolled treatment are useful, typically the occurrence of relapse or the decision to reinstitute treatment.

The Patient

The patient's own characteristics remain perhaps the most potent determinants of outcome or prognosis (Luborsky, 1972). In the comparative evaluation of treatments it is crucial that patients in different groups be equivalent so that effects due to treatment and to prognostic factors can be separated. Equivalence with respect to unknown prognostic factors—and this is still an important issue—can be assured only by assigning patients to treatments at random (possibly with some restrictions to allow for matching on certain characteristics).

Prognostic factors may influence outcome independently of treatment. Even more significant are patient characteristics that affect outcome differentially in relation to the treatment employed (interaction). These patterns provide the differential indications for treatment that clinicians seek (Uhlenhuth, 1968).

Details of the patient's clinical condition contain basic prognostic information. A high level of current disturbance, particularly affective disturbance, seems to be a favorable sign (Group for the Advancement of Psychiatry, 1969; GAP, 1975; Kernberg, 1972; Strupp, 1969). The favorable outlook for episodes of the major affective disorders and acute schizophrenia is well known (Stephens, 1965).

On the other hand, a healthy characteristic level of function (traits, character style) seems to be prognostically favorable (Evans, 1973; Garfield, 1971; Gelder, 1967; Kernberg, 1972; Meltzoff, 1970; Vaillant, 1964). Competent function relates to such variables as high intelligence, educational and occupational levels, verbal ability, ego strength, low neuroticism, and satisfactory interpersonal relations, including marriage.

Direct studies of these as prognostic variables in psychotherapy, however, have produced surprisingly conflicting results (Garfield, 1971; Meltzoff, 1970). The same is true for age, sex, race, and prior treatment. There is more general agreement that the patient's expectations upon entering treatment influence the outcome (Fiske, 1970; Frank, 1973; Frank, 1959; Friedman, 1963; Goldstein, 1962; Hoehn-Saric, 1964; Rickels, 1971; Uhlenhuth, 1968).

The situation is more promising with pharmacotherapy, where

prognostic variables of both general significance and specific significance in the choice of drug (interaction) are beginning to emerge. Psychiatric diagnosis and other features of the clinical picture offer useful indications for the choice of medication (Klein, 1969), and diagnostic refinements are proceeding apace (Woodruff, 1974). Among patients with major affective disorders, differential indications for treatment are relatively clear: in bipolar patients lithium prevents affective episodes but in unipolar patients Imipramine prevents affective episodes at least as effectively as lithium (Prien, 1973). Among schizophrenics, patients with good premorbid adjustment and no paranoid symptoms carry the best prognosis (Evans, 1973) and benefit least from phenothiazines (Evans, 1972; Judd, 1973). Among anxious psychoneurotic patients, those with higher anxiety levels, more chronic disturbance, and better response to previous drug treatment benefit most from medication (Rickels, 1971).

Consensus on the prognostic significance of other variables is building gradually as the data accumulate (Raskin, 1972; Rickels, 1971; Uhlenhuth, 1969). For example, neurotic patients with higher indices of social advantage seem to benefit more from antidepressants (Downing, 1973) or anti-anxiety agents (Rickels, 1971). Older depressed patients respond better than younger ones to antidepressants (Raskin, 1972).

What accounts for the apparently more rapid advance of prognostic

discrimination related to pharmacotherapy than to psychotherapy? The positive indicators for response to psychotherapy, for spontaneous remission (Gottschalk, 1967) and for response to placebo are similar in many respects. Curiously, many of the positive indicators for pharmacotherapy seem to stand at the opposite pole: beneficial medication effects are manifest especially in patients who would be less likely to improve otherwise.

Much of the information on prognostic indicators for psychotherapy derives from studies of a single group of subjects, all treated with the therapy of interest. In such a situation the observed responses include some unspecifiable mixture of nonspecific effects and effects attributable specifically to therapy. Differential indications for different treatments, including "no treatment," emerge clearly only from comparative studies designed to separate treatment effects from other effects on outcome. A survey of therapeutic outcomes by Saenger (1970) illustrates that even relatively crude approximations to this design highlight differential prognostic effects. His results also suggest that the specific benefits of treatment, contrary to popular opinion, emerge in patients who have a relatively poor prognosis without treatment.

The Treatment

Psychotherapeutic procedures generally have not been well specified.

This tendency, surprisingly, is perhaps stronger among investigators than among clinicians and teachers. In the psychoanalytic setting considerable attention has been paid to technical detail as illustrated, for example, by Bibring's (1954) classification of interventions. There is growing recognition that "psychotherapy" is not a unitary process, so that questions about "the effects of psychotherapy" carry meaning only on a very gross level." Many investigators are joining Paul (1969) in his call for specification: "What treatment by whom, is most effective for this individual with that specific problem, under which set of circumstances, and how does it come about?" Note also, however, May's (1971) argument for our need to be concerned with the effectiveness of "average psychotherapy."

Pharmacotherapy can serve as a model for the specification of treatment in psychiatry. The precision of chemical specification doubtless helps to account for the rapid development of evaluative research in pharmacotherapy during the past two decades. The ideal presented by this model probably is unattainable in studies of psychotherapy, even in the area of behavior modification, since the inherent difficulties are substantially greater.

The "central therapeutic ingredients" (Truax, 1971) of psychotherapy have not yet been identified clearly. Indeed, the analogy to medication, despite its attractiveness, eventually may prove inadequate for

conceptualizing psychotherapeutic influence. For the time being the notion is useful, partly because it highlights the still unresolved issue between specific procedures (skills) and less tangible personal qualities as the principal therapeutic tools (Strupp, 1969). The final common pathway of both, of course, must be therapist's specific behavior.

Granted the importance and the difficulty of specifying psychotherapeutic treatment, then what significant dimensions can be specified at our present level of understanding? Some studies have shown differential effects related to *general treatment technique*—psychoanalysis, expressive therapy, supportive therapy, behavior therapy—when patient characteristics also are considered. For example, desensitization seems to be more effective and more rapidly effective than other psychotherapies for specific phobias of moderate or less severity (Luborsky, 1972). The relative effects on general adjustment, however, are less clear. The most comprehensive comparative study of psychoanalysis and related treatments (Kernberg, 1972) suggests (1) that psychoanalysis is more effective than supportive or expressive therapies with persons who have high ego strength, and (2) that an intermediate supportive-expressive therapy is more effective than psychoanalysis or primarily supportive therapy with persons who have low ego strength.

At present, there is no convincing evidence of differences in

effectiveness among practitioners of different *theoretical orientations*—psychoanalytic, Rogerian, behavioral—except for those already mentioned in regard to the behavioral approach (Luborsky, 1972; Meltzoff, 1970). This area, however, has received relatively meager attention.

From the patient's viewpoint, the *persons included in treatment*—patient, family, group of unrelated persons—are part of the treatment setting (see below). Currently available reports provide little evidence for differential effectiveness associated with this aspect of the setting (Luborsky, 1972; Meltzoff, 1970; Reid, 1969), although the issue certainly is not closed, especially with regard to certain forms of group therapy with schizophrenic patients (Meltzoff, 1970; O'Brien, 1972).

Differences among *settings*—office, clinic, day hospital, full-time hospital—and between different settings in the same class are observed commonly in controlled treatment studies (Rickels, 1971). Some schizophrenics respond better to day-hospital than to outpatient treatment (Guy, 1969) and also respond differently to differences in milieu among hospital wards (Kellam, 1967). Instruments for measuring dimensions of ward atmosphere are available (Kellam, 1967; Moos, 1971).

Since random assignment of patients to settings generally is not feasible, the confounding of patient characteristics, which often differ among

settings, and intrinsic differences among settings cannot be undone. A recent study of pharmacotherapy and group psychotherapy with ambulatory depressed patients addressed itself specifically to this issue by employing the same study team in two clinics serving patients of differing social class status (Covi, 1973). The highly significant response contrasts in favor of Imipramine and against diazepam (compared with placebo) were contributed primarily by the clinic with patients of higher social class.

The *therapist's style* can be viewed as some combination of his personality, orientation, and procedures. For present purposes this variable is of interest insofar as it can be specified. The Rogerian school has developed the most comprehensive body of work in this direction, well summarized by Truax and Carkhuff (1967) and Truax and Mitchell (1971). They report that different therapists characteristically provide different "levels of core interpersonal skills" along scaled dimensions of accurate empathy, non possessive warmth, and genuineness.

Another series of investigations center around the therapist's type as originally defined by Whitehorn and Betz (1954) according to success (Type A) or failure (Type B) in treating schizophrenic patients. These types showed different patterns on the Strong Vocational Interest Blank, and subsequent work was based on the types as differentiated by this test index. Despite fifteen years of research, the value of this discrimination remains

controversial (Meltzoff, 1970).

The *therapist's experience* is another factor that intuitively seems important in therapeutic results. Although the weight of the evidence favors this idea (Meltzoff, 1970), it is much less conclusive than might be expected (May, 1971).

It is equally surprising to find a lack of definitive studies and conclusive results relating treatment outcome to *amount of therapeutic contact*—duration, frequency, number, and regularity of sessions (Meltzoff, 1970). Most of the available evidence suggests that therapeutic gains increase with the number of sessions up to some limit probably determined by other factors, but lower than generally supposed, say twenty.

A major study on the amount of treatment indicates that minimal contact (thirty minutes every two weeks) provides about the same relief as weekly individual (one hour) or group (one and one-half hour) sessions. Improvement in social function, however, is more rapid in the individual or group modalities (Fox, 1968).

Most of these studies suffer from the difficulty of disentangling the several time-related dimensions. Another problem is the possible confounding effect of both patient and therapist expectations (Truax, 1971).

The behavioral approaches offer a model for the specification of *particular technical maneuvers.* Although behavioral techniques as a class already have compiled an encouraging record of effectiveness (Paul, 1969; Luborsky, 1972), the promise of differentiation among technical variations remains largely unfulfilled in this young field.

Investigations of differential effects due to other specific tactics also show considerable promise. Therapists who lose lower-class patients and therapists who retain them in treatment differ markedly on such maneuvers as addressing their patients by name (Howard, 1970). Patients prepared for psychotherapy with detailed information about the procedures and their anticipated results benefit more than unprepared patients (Hoehn-Saric, 1964).

As noted earlier, medication lends itself to detailed specification more readily than other aspects of treatment. Nevertheless, even in studies of pharmacotherapy, problems regarding the therapeutic agent do arise, including the regimen followed by the patient as distinct from the regimen prescribed by the investigator. These issues are discussed comprehensively in Levine et al. (1971) The critical point here is that many studies of psychotherapy fail to control or even to take account of medications used by the patients. In an era when, each year, 22 percent of the population ingest psychotropic prescription drugs of established potency (Mellinger, 1973), this

oversight can have disastrous consequences for a study (May, 1971).

The patient's environment beyond the treatment setting is another greatly neglected area that bears on outcome. External events, for example, can diminish drug-placebo contrasts, since unfavorable events differentially reduce drug response whereas favorable events differentially increase placebo response (Rickels, 1965). In the only controlled study of psychotherapy that paid explicit attention to environmental variables, no association with outcome was noted (Kernberg, 1972).

Cost

The cost of health care is a subject of such concern that it finds expression in the public press almost daily. Cost-benefit concepts and their importance have been discussed in the psychiatric literature for some time (Fox, 1968). Nevertheless, controlled studies on the effectiveness of treatment leave the reader to draw his own inferences on the issue. There are no data, with the striking exception of May's report on the treatment of schizophrenia (1968). He shows that ataraxic drugs markedly reduce and psychotherapy markedly increases the cost of treatment, absolutely and in relation to effectiveness.

The cost of treatment usually is tallied under direct and indirect expenses. Indirect expenses generally include items such as maintenance of

the facility, room and board, basic nursing care, and clerical functions. These are charged in direct proportion to the length of hospital stay or the number of outpatient visits. Direct expenses usually include procedures that are costly and individually ordered, such as tests, psychotherapy, ECT, and medications. The cost of outpatient care is largely attributable to the special procedures that the therapist personally prescribes and performs and so should be easily within reach of the clinical investigator in many outcome studies.

The cost of psychiatric illness itself can be modified by treatment, but its estimation presents some almost insurmountable difficulties. It includes items such as welfare; pensions or disability compensation; loss of productivity, income and taxes; and human losses in function as a spouse or parent. Even long-delayed and tenuously related consequences like increased susceptibility among the children of ill parents could involve significant costs. Despite the acknowledged problems, partial accounting of the costs of illness should be possible now with data from individual patients on days of work lost, income, and its sources, and certain criteria of ineffective functioning.

Design

For present purposes the design of a study is a plan of procedure that will allow the investigator to make causal inferences from his data by separating the sources of variation in outcome among patients. In its broadest

sense, this plan specifies all details of the research, including criteria for selecting patients and therapists, treatments to be compared and their mode of delivery, allocation of patients to treatments and therapists, criteria of treatment effect and the timing of their application, management of deviations from protocol, analysis and interpretation of the data. In a more limited sense, design refers to the manner in which patients enter and pass through the experimental treatment framework.

Specialists in design with an extensive statistical background have developed in the field of psychology (Campbell, 1966). An alternative pattern of close collaboration between a clinical investigator and a biomedical statistician is emerging in the field of psychiatry (Dixon, 1969; Gurland, 1972; Uhlenhuth, 1969).

The importance of statistical skills in design now seems to be generally recognized. Since adequate design depends primarily on keen logical analysis of the real treatment situation, the very active participation of a person with substantive and logistic knowledge of the field, i.e., a clinician, is equally critical. In the absence of a strong empirical influence from the clinician-investigator, a tendency may develop to deal with issues by carrying deduction well beyond the point of support by data. A pertinent example in psychopharmacologic evaluation is the current rather uncritical rejection of patient-own-control designs, specifically “intensive design” (Chassan, 1967),

based on the *possibility* that a treatment effect may carry over into a subsequent phase of the study. The current press for ever larger sample sizes — with their increasing possibilities for error-—as a solution for the problem of error variation within treatment groups may prove to be another example (Derogatis, 1968; Overall, 1967; Rickels, 1968).

Designs incorporate controls for a variety of errors, both random and systematic (bias [Group for the Advancement of Psychiatry, 1959]). These controls include clear conceptualization, specification of all aspects of procedure, appropriate sampling techniques, treatment-comparison groups, and the structure of the informational context within which the study takes place. The most familiar example of the last is the "double blinding" of patient and investigator as to the specific treatment received by an individual patient at a particular time. Limitations in double blinding present a special problem in evaluating non-medicinal treatments in psychiatry that can be resolved only partly by using an "independent research assessor" (Guy, 1967).

In any event, the double blind is too narrow a concept. All of the information transmitted in connection with a study, either verbally or nonverbally, must be carefully considered and structured in view of its probable effects on patients and therapists (Orne, 1962; Rosenthal, 1966). These effects probably are mediated by activation of patients' and therapists' expectations (see pp. 941 and 944). The expectations activated, however,

often bear no simple relation to the information supplied by the investigator (Park, 1965).

Designs in the strictest sense are for experiments in which the investigator manipulates the variable of causal interest, the treatment. The basic strategy is to distribute unintended and especially unidentifiable effects on outcome equally among treatment groups in order to disassociate such effects from the treatment effects under study. The usual tactic employed is to assign patients to treatments at random, sometimes with certain systematic restrictions (stratification, matching) to assure balance with respect to variables of known importance.

There are, however, many situations where random assignment is not ethically justifiable. The issue arises when one of the experimental treatments is inferior or essentially a dummy. Complete assurance that a particular study is sensitive to treatment differences unfortunately requires the inclusion of a standard treatment and a dummy treatment in addition to the treatment being evaluated. This conflict between experimental and ethical requirements is likely to increase as more effective treatments are discovered.

Many investigators of psychotherapy advocate “naturalistic” studies (Butler, 1962). The basic strategy here is the careful, systematic observation of existing situations with correlational analysis of the resulting data. This

approach can comprehend many of the variables found in such complex situations as psychotherapy and their relationships (Kiesler, 1971). It does not encounter the ethical problems mentioned above. Because it requires minimal interference with the way treatment is prescribed and delivered, the naturalistic approach may offer the only practical hope of evaluating the effectiveness of ongoing treatment programs of immediate public interest. The large data banks that are gradually accumulating by virtue of computer technology require similar approaches for their full utilization.

Statisticians for some time have considered the problems of non-experimental research in general (Blalock, 1964; Cochran, 1965). Survey research is based on non-experimental techniques that, in part, are applicable to outcome studies (Schwartz, 1973). The growing interest in this area among psychopharmacologists was manifested by a workshop in January 1973 to discuss "Approaches to the Use of Observational Data in Psychopharmacology."

Critical consideration of the non-experimental approach in the light of some of these materials reveals some problems that are shared by the experimental approach, but are often overlooked. The major issues concern confounding the effects of the treatment under study and the effects of other variables. Co-variance techniques can be used to account for identified sources of confounding if they are not too severe. Unidentified sources of

confounding that are intrinsically related to treatment (conceivably some effects of treatment settings) are clearly unmanageable in either experimental or non-experimental research. This technical problem can be reformulated as a substantive question about the mechanism of treatment effects.

Unidentified sources of confounding that are not intrinsically related to treatment (conceivably some prognostic characteristics of patients who tend to be selected for some treatment) can be controlled by the random assignment of patients to treatments. The particular weakness of non-experimental research lies in its inability to account for this type of confounding. The consensus nevertheless seems to regard observational research as a fruitful source of hypotheses that should be and usually can be confirmed by supplementary experiments (Jick, 1970).

An important problem encountered with increasing frequency in treatment evaluations stems partly from growing public sophistication about psychiatry. More patients come with more definite wishes for a specific form of treatment. In samples where most patients favor one of the experimental treatments, random assignment to treatments is likely to create rather than control bias since the evenly allocated expectations are likely to enhance response in the favored treatment group and inhibit response in the other groups. Additional complications may ensue if patients selectively drop from the treatments that are not favored. If these trends gain strength, it may be

useful to consider studies not concerned with the direct comparison of treatment effects, but rather with the factors prognostic of benefit among patients who presented themselves for a particular treatment.

Analysis

The analysis of results from evaluative research, of course, is closely related to the design employed. Analyses of variance, co-variance, and regression are the procedures commonly applied in comparing independent treatment groups. Extensions of the basic methods to include multiple independent variables, as in multiple regression (Cochrane, 1972) and discriminant function analyses, and to include also multiple dependent variables, as in multivariate analyses of variance and covariance offer elegant approaches to understanding complex sets of correlated data (Bock, 1967; Morrison, 1967). Several well-documented, flexible computer programs are now available to implement these procedures at different levels of complexity.1 2 21 38 106 135 148 At least one data processing system complete from measuring instruments through analyses is operational (McGlashan, 1973).

The application of these methods in experiments, where increased precision is the main objective, is relatively straightforward. In non-experimental studies, where the removal of bias due to imbalance of

confounding factors among treatments is a prime objective, the use of these methods remains controversial (Cochran, 1965; Johnson, 1972; Schwartz, 1973).

As interest grows in processes of change, the time dimension gains importance in evaluative research. The analysis of repeated measures for trends is a useful device in this connection. Such analyses provide a dynamic picture of treatment effects, increase precision by employing all the data from every patient, and may be less affected by an occasional missing value (Dixon, 1969).

Analyses of data from patient-own-control designs, especially those extending over multiple time periods, also deserve special mention. They encounter the problems of internal correlation common to all time series, but generally do not provide enough data to take advantage of such special methods as auto- and cross-correlation. Chassan (1967) offers a simple, but controversial, approach.

With the vast power of modern electronic computers so readily available, it is easy to act as if sufficiently sophisticated statistical manipulations could solve most of the problems of evaluative research. It is tempting to carry out massive searches through the data to uncover effects that may confound, interact with, or be even more striking than the treatment

effects under study. The amounts of data required for such efforts to produce replicable results are difficult to appreciate (Boston Collaborative Drug Surveillance Program, 1973; Jick, 1970). It is also tempting to try endless variations on the same analytic theme to wring one more sufficient F ratio ($p < 0.05$) from the data. Unfortunately, close practical (computational) similarities sometimes lurk under hotly argued theoretical differences separating some of these analytic approaches. There is evidently no substitute for a carefully conceived, designed and executed study, with results that are clear on inspection of the raw data and confirmable with simple statistics.

Concluding Remarks

Evaluative research in psychiatry has had an interesting history—even exciting, especially in psychopharmacology, where experimental methods could be developed, applied, and refined most readily. With interest growing in the question of what works, the future promises further development of evaluative research. Hopefully, the pressures to evaluate particular programs will not compromise the design of studies to provide information about specific variables valid beyond the immediate context of the particular programs.

More detailed specification of psychological variables surely is a

keystone in the future development of evaluative research in psychiatry. Pertinent specification in turn depends partly on the emergence of additional substantive knowledge about the very effects we wish to study, including the effects of psychological interventions. Promising new chemical entities with distinctly different structures would serve as a similar stimulus for research on the effects of psychopharmacological treatments.

Advances in methods also are needed. Some basic statistical developments seem indicated, for instance, in reference to interpreting results from procedures that "search" stepwise a pool of independent variables for relations to a dependent variable.

Improved measurement and vastly increased sample size have received much emphasis as a means of coping with variation within treatment groups ("error"). Although further improvement is always desirable, great strides in measurement have been made. Logistic problems in assembling large samples probably disproportionately escalate error. Statistically significant differences between treatments in large samples also may not be large enough to be clinically meaningful.

Improved design has received less attention in the search for greater experimental precision. Investigators tend to fall into the routine of comparing treatments in parallel independent groups while focusing debate

on whether six or eight weeks duration is more appropriate. The imaginative development and application of more fundamental modifications in design now seem to hold promise for increasing the sensitivity of evaluative studies in psychiatry and also meeting ethical requirements. One possible direction is variation on the patient-own-control theme.

There is a notable tendency to employ deductive debate, sometimes at astonishing length, rather than experiment, to resolve methodologic uncertainties. For example, only one empirical study—and this is not comparative—has seriously addressed the question of the value of employing an initial placebo washout period in clinical psychopharmacologic research (Jones, 1966). This stance among evaluators seems incongruous, to say the least, since empirical evaluative research contains within itself the means for its own improvement, should we choose to employ them.

Bibliography

Andrews, F. M., J. M. Morgan, and J. A. Sonquist. *Multiple Classification Analysis: A Report on a Program for Multiple Regression Using Categorical Predictors.* Survey Research Center, Institute for Social Research. Ann Arbor, Mich.: University of Michigan Press, 1967.

Armour, D. J. and A. S. Couch. *The Data-Text Primer: An Introduction to Computerized Social Data Analysis Using the Data-Text System*. New York: Free Press, 1972.

Arnold, M. F. "Evaluation: A Parallel Process to Planning," in M. F. Arnold, L. V. Blankenship, and J. M. Hess, eds., *Administering Health Systems*, pp. 263-282. Chicago: Aldine-Atherton,

1971.

Astrup, C., A. Fossum, and R. Holmboe. *Prognosis In Functional Psychoses.* Springfield, Ill.: Charles C. Thomas, 1963.

Bahn, A. K. and C. A. Thrall. "Impairment of Psychiatric Outpatients and Change with Treatment," *Ment. Hyg.*, 48 (1964), 217-242.

Battle, C. C., S. D. Imber, R. Hoehn-Saric et al. "Target Complaints as Criteria of Improvement," *Am. J. Psychother.,* 20 (1966), 184-192.

Bergin, A. E. "The Evaluation of Therapeutic Outcomes," in A. E. Bergin and S. L. Garfield, eds., *Handbook of Psychotherapy and Behavior Change: An Empirical Analysis,* pp. 217-270. New York: Wiley, 1971.

Bergin, A. E. and S. L. Garfield, eds. *Handbook of Psychotherapy and Behavior Change: An Empirical Analysis.* New York: Wiley, 1971.

Bibring, E. "Psychoanalysis and the Dynamic Psychotherapies," *J. Am. Psychoanal. Assoc.,* 2 (1954), 745-770.

Binner, P. R. "Operational Problems of Program Evaluation." Presented at S. Reg. Conf. Ment. Health Statis. Workshop, Norman, Okla., August 1968.

Blalock, H. M., Jr. *Causal Inferences in Non-experiment al Research*. Chapel Hill: University of North Carolina Press, 1964.

Bock, R. D. "Multivariate Analysis of Variance of Repeated Measures," in C. W. Harris, ed., *Problems in Measuring Change,* pp. 85-103. Madison: University of Wisconsin Press, 1967.

Boston Collaborative Drug Surveillance Program. "Clinical Depression of the Central Nervous System Due to Diazepam and Chlordiazepoxide in Relation to Cigarette Smoking and Age," *N. Engl. J. Med.,* 288 (1973), 277-280.

Brook, R. H. and F. A. Appel. "Quality-of-Care Assessment: Choosing a Method for Peer Review," *N.*

Engl. J. Med., 288 (1973). 1323-1329.

Butler, J. M., L. N. Rice, and A. K. Wagstaff. "On the Naturalistic Definition of Variables: An Analogue of Clinical Analysis," in H. H. Strupp and L. Luborsky, eds., *Research In Psychotherapy,* Vol. 2, pp. 178-205. Washington: Am. Psychological Assoc., 1962.

Campbell, D. T. "From Description to Experimentation: Interpreting Trends as Quasi-Experiments," in C. W. Harris, ed., *Problems In Measuring Change,* pp. 212-242. Madison: University of Wisconsin Press, 1967.

Campbell, D. T. and J. C. Stanley. *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally, 1966.

Caro, F., ed. *Readings in Evaluation Research*. New York: Basic Books, 1971.

Cattell, R. B. *The Scientific Analysis of Personality*. Chicago: Aldine, 1965.

Chassan, J. B. *Research Design in Clinical Psychology and Psychiatry*. New York: Appleton-Century-Crofts, 1967.

Clyde, D. J. *Manova: Multivariate Analysis of Variance on Large Computers.* Miami: Clyde Comput. Serv., 1969.

Cochrane, A. L. *Effectiveness and Efficiency: Random Reflections on Health Services.* London: Nuffield Provincial Hospitals Trust, 1972.

Cochran, W. G. "The Planning of Observational Studies of Human Populations," *J. R. Statis. Soc. Ser.* A, 128 (1965), 234-265.

Cohen, J. "Multiple Regression as a General Data-Analytic System," *Psychol. Bull.,* 70 (1968), 426-443.

Covi, L., R. Lipman, and L. R. Derogatis. "Drugs and Group Therapy in Neurotic Depression." Presented at 126th Annu. Meet. Am. Psychiatr. Assoc., Honolulu, May 8, 1973.

Dahlstrom, G., G. S. Welsh, and L. E. Dahlstrom. An MMPI Handbook. Vol. 1. *Clinical Interpretation.*

Minneapolis: University of Minnesota Press, 1972.

Dent, J. K. A *Bibliographic Index of Evaluation in Mental Health*. U.S. Public Health Service Pub. no. 1545. Washington: U.S. Govt. Print. Off., 1966.

Derogatis, L. R., R. R. Bonato, and K. C. Yang. "The Power of IMPS in Psychiatric Drug Research: As a Function of Sample Size, Number of Raters, and Choice of Treatment Comparison," *Arch. Gen. Psychiatry,* 19 (1968), 689-699.

DiMascio, A., B. Warshaver, and A. Raskin. "Ward Behavior and Psychiatric Rating Scales." *Mimeo.,* 1972.

Dixon, W. J. and P. R. A. May. "Process Analysis in the Assessment and Prediction of Psychiatric Outcome," in P. R. A. May and J. R. Wittenborn, eds., *Psychotropic Drug Response: Advances In Prediction,* pp. 228-267. Springfield, Ill.: Charles C. Thomas, 1969.

Donabedian, A. "Evaluating the Quality of Medical Care," *Milbank Mem. Fund* Q., 44 (1966), 166-206.

Downing, R. W. and K. Rickels. "Predictors of Response to Amitriptyline and Placebo in Three Outpatient Treatment Settings," *J. Nerv. Ment. Dis.,* 156 (1973), 109-129.

Endicott, J. and R. L. Spitzer. "Current and Past Psychopathology Scales (CAPPS)," *Arch. Gen. Psychiatry,* 27 (1972), 678-687.

Engelhardt, D. M., B. Rosen, N. Freedman et al. "Phenothiazines in Prevention of Psychiatric Hospitalization," *Arch Gen. Psychiatry,* 16 (1967), 98-101.

Evans, J. R., M. J. Goldstein, and E. H. Rodnick. "Premorbid Adjustment, Paranoid Diagnosis, and Remission," *Arch. Gen. Psychiatry,* 28 (1973), 666-672.

Evans, J. R., E. H. Rodnick, M. J. Goldstein et al. "Premorbid Adjustment, Phenothiazine Treatment, and Remission in Acute Schizophrenics," *Arch. Gen. Psychiatry,* 27 (1972), 486-490.

Federal Drug Administration. "FDA Guidelines for Psychotropic Drugs (Draft —June 1974)," *Psychopharmacol. Bull.,* 10 (1974), 70-91.

Finn, J. D. *Multivariance: Univariate and Multivariate Analysis of Variance, Covariance and Regression: A Fortran IV Program,* Version 5. Ann Arbor: Natl. Educational Resources, 1972.

Fiske, D. W., H. F. Hunt, L. Luborsky et al. "Planning Research on Effectiveness of Psychotherapy," *Arch. Gen. Psychiatry,* 22 (1970), 22-32.

Fox, P. D. and J. M. Kuldau. "Expanding the Framework for Mental Health Program Evaluation," *Arch. Gen. Psychiatry,* 19 (1968), 538-544.

Frank, J. D. *Persuasion and Healing: A Comparative Study of Psychotherapy*, rev. Baltimore: The Johns Hopkins University Press, 1973.

Frank, J. D., L. H. Gliedman, S. D. Imber et al. "Patients' Expectancies and Relearning as Factors Determining Improvement in Psychotherapy," *Am. J. Psychiatry,* 115 (1959), 961-968.

Franks, C. M., ed. *Behavior Therapy: Appraisal and Status*. New York: McGraw-Hill, 1969.

Freeman, H. E. "Outcome Measures and Social Action Experiments: An Immodest Proposal for Redirecting Research Efforts," *Am. Sociol.,* 7 (1972), 17-19.

Fried, M. "Evaluation and the Relativity of Reality," in L. M. Roberts, N. S. Greenfield, and M. H. Miller, eds., *Comprehensive Mental Health: The Challenge of Evaluation.* Madison: University of Wisconsin Press, 1968.

Friedman, H. J. "Patient Expectancy and Symptom Reduction," *Arch. Gen. Psychiatry,* 8 (1963), 61-67.

Garfield, S. L. "Research on Client Variables in Psychotherapy," in A. E. Bergin and S. L. Garfield, eds., *Handbook of Psychotherapy and Behavior Change: An Empirical Analysis,* pp. 271-298. New York: Wiley, 1971.

Gelder, M. G., I. M. Marks, and H. H. Wolff. "Desensitization and Psychotherapy in the Treatment of Phobic States: A Controlled Inquiry," *Br. J. Psychiatry,* 113 (1967). 53-73.

Goldstein, A. P. *Therapist-Patient Expectancies in Psychotherapy*. New York: Pergamon, 1962.

Goldstein, A. P. and S. J. Dean, eds. *The Investigation of Psychotherapy: Commentaries and Readings.* New York: Wiley, 1966.

Goldstein, A. P., K. Heller, and L. B. Sechrest. *Psychotherapy and the Psychology of Behavior Change.* New York: Wiley, 1966.

Gottschalk, L. A. "Some Problems in the Evaluation of the Use of Psychoactive Drugs, with or without Psychotherapy, in the Treatment of Non-psychotic Personality Disorders," in D. H. Efron, J. O. Cole, J. Levine et al., eds., *Psychopharmacology: A Review of Progress* 1957-1967, pp. 255-269. U.S. Public Health Service Pub. no. 18,36. Washington: U.S. Govt. Print. Off., 1968.

Gottschalk, L. A. and A. H. Auerbach. *Methods of Research in Psychotherapy.* New York: Appleton-Century-Crofts, 1966.

Gottschalk, L. A., P. Mayerson, and A. A. Gottlieb. "Prediction and Evaluation of Outcome in an Emergency Brief Psychotherapy Clinic," *J. Nerv. Ment. Dis.,* 144 (1967), 77-96.

Group for the Advancement of Psychiatry. *Some Observations on Controls in Psychiatric Research,* Rep. no. 42. New York: GAP, 1959.

----. *Psychiatric Research and the Assessment of Change*, Rep. no. 63. New York: GAP, 1966.

----. *Psychotherapy and the Dual Research Tradition*, Rep. no. 73. New York: GAP, 1969.

----. *Pharmacotherapy and Psychotherapy:Paradoxes, Problems and Progress,* Rep. no. 93. New York: GAP, 1975.

Gurland, B. J., N. J. Yorkston, K. Goldberg et al. "The Structured and Scaled Interview to Assess Maladjustment (SSIAM); II. Factor Analysis, Reliability, and Validity," *Arch. Gen. Psychiatry,* 27 (1972), 264-267.

Gurland, B. J., N. J. Yorkston, A. R. Stone et al. "The Structured and Scaled Interview to Assess Maladjustment (SSIAM); I. Description, Rationale, and Development," *Arch. Gen.*

Psychiatry, 27 (1972). 259-264.

Guy, W. and R. R. Bonato. *Manual for the ECDEU Assessment Battery,* 2nd rev. ed. Washington: NIMH, 1970.

Guy, W., M. Gross, and H. Dennis. "An Alternative to the Double-Blind Procedure," *Am. J. Psychiatry,* 123 (1967), 1505-1512.

Guy, W., M. Gross, G. E. Hogarty et al. "A Controlled Evaluation of Day Hospital Effectiveness," *Arch. Gen. Psychiatry,* 20 (1967), 329-338.

Halpert, H. P., W. J. Horvath, and J. P. Young. *An Administrator s Handbook on the Application of Operations Research to the Management of Mental Health Systems,* Publ. no. 1003. Washington: National Clearinghouse for Mental Health Information, 1970.

Harris, G. W., ed. *Problems In Measuring* Change. Madison: University of Wisconsin Press, 1967.

Herzog, E. *Some Guide Lines for Evaluative Research:Assessing Psycho-Social* Change in Individuals. DHEW, Children's Bureau Pub. no. 379-1959. Washington: U.S. Govt. Print. Off., 1959.

Hoehn-Saric, R., J. D. Frank, S. D. Imber et al. "Systematic Preparation of Patients for Psychotherapy: I. Effects on Therapy Behavior and Outcome," *J. Psychiatr. Res.,* 2 (1964), 267-281.

Hogarty, G. E. "Selected Measures of Community Adjustment following Outpatient Psychotherapy: Ratings from Significant Others." Mimeo.

Hogarty, G. E., S. C. Goldberg, and the Collaborative Study Group. "Drug and Sociotherapy in the Aftercare of Schizophrenic Patients: One Year Relapse Rates," *Arch. Gen. Psychiatry,* 28 (1973), 54-64.

Holt, R. R. "Experimental Methods in Clinical Psychology," in B. B. Wolman, ed., *Handbook of Clinical Psychology,* pp. 40-77. New York: McGraw-Hill, 1965.

Honigfeld, G. "Non-Specific Factors in Treatment," *Dis. Nerv. Syst.,* 25 (1964), M5-239-

Hornstra, R. K. and T. S. McPartland. "Aspects of Psychiatric After-Care," *Int. J. Soc. Psychiatry,* 9 (1963), 135-142.

Howard, K., K. Rickels, J. E. Mock et al. "Therapeutic Style and Attrition Rate from Psychiatric Drug Treatment," *J. Nerv. Ment. Dis.,* (1970), 102-110.

Imber, S. D., E. H. Nash, R. Hoehn-Saric et al. "A Ten-Year Follow-up Study of Treated Psychiatric Outpatients," in S. Lesse, ed., *An Evaluation of the Results of the Psychotherapies,* pp. 1-12. Springfield, Ill: Charles C. Thomas, 1968.

Jick, H., O. S. Miettinen, S. Shapiro et al. "Comprehensive Drug Surveillance," *JAMA,* 213 (1970), 1455-1460.

Johnson, L. C. and A. Lubin. "On Planning Psychophysiological Experiments," in N. S. Greenfield and R. A. Sternbach, eds., *Handbook of Psychophysiology,* pp. 125-158. New York: Holt, Rinehart and Winston, 1972.

Jones, M. B. and J. D. Ainslie. "Value of a Placebo Washout," *Dis. Nerv. Syst.,* 27 (1966), 393-396.

Judd, L. L., M. J. Goldstein, E. H. Rodnick et al. "Phenothiazine Effects in Good Premorbid Schizophrenics Divided into Paranoid-Nonparanoid Status," *Arch. Gen. Psychiatry,* 29 (1973), 207-211.

Kellam, S. G., S. C. Goldberg, N. R. Schooler et al. "Ward Atmosphere and Outcome of Treatment of Acute Schizophrenia," *J. Psychiatr. Res.,* 5 (1967), 145-163.

Kellner, R. "The Evidence in Favor of Psychotherapy," *Br. J. Med. Psychol.,* 40 (1967), 341-358.

----. "Part 1. Improvement Criteria in Drug Trials with Neurotic Patients," *Psychol. Med.,* 1 (1971), 416-425.

----. "Part 2. Improvement Criteria in Drug Trials with Neurotic Patients," *Psychol. Med.,* 2 (1972), 73-80.

Kernberg, O. F., E. D. Burstein, L. Coyne et al. "Psychotherapy and Psychoanalysis: Final Report of the Menninger Foundation's Psychotherapy Research Project," *Bull. Menninger*

Clin., 36 (1972), 1-275.

Kiesler, D. J. "Some Myths of Psychotherapy Research and the Search for a Paradigm," *Psychol. Bull.,* 65 (1965), 16-136.

----. "Experimental Designs in Psychotherapy Research," in A. E. Bergin and

S. L. Garfield, eds., *Handbook of Psychotherapy and Behavior Change: An Empirical Analysis,* pp. 36-74. New York: Wiley, 1971.

Kiresuk, T. J. and R. E. Sherman. "Goal Attainment Scaling: A General Method for Evaluating Comprehensive Community Mental Health Programs," *Community Ment. Health J.,* 4 (1968), 443-453.

Klein, D. F. and J. M. Davis. *Diagnosis and Drug Treatment of Psychiatric Disorders,* p. 436. Baltimore: Williams & Wilkins, 1969.

Klein, Z. E. "Research in the Child Psychiatric and Guidance Clinics: A Bibliography (1923-1970)." Mimeo.

Klerman, G. L., A. DiMascio, M. Weissman et al. "Treatment of Depression by Drugs and Psychotherapy." Presented at 126th Annu. Meet. Am. Psychiatr. Assoc. Honolulu, May 8, 1973.

Klett, C. J. and E. C. Moseley. "The Right Drug for the Right Patient," *J. Consult. Psychol,* 29 (1965), 546-557.

Krasner, L. "The Therapist as a Social Reinforcement Machine," in H. H. Strupp and L. Luborsky, eds., *Research In Psychotherapy,* Vol. 2, pp. 61-94. Washington: Am. Psychological Assoc., 1962.

Levine, J., B. C. Schiele, and L. Bouthilet. *Principles and Problems in Establishing the Efficacy of Psychotropic Agents.* U.S. Public Health Service Publ. no. 2138. Washington: U.S. Govt. Print. Off., 1971.

Lipman, R. S., J. O. Cole, L. C. Park et al.

"Sensitivity of Symptom and Nonsymptom-Focused Criteria of Outpatient Drug Efficacy," *Am. J. Psychiatry,* 122 (1965), 24-27.

Luborsky, L. "Comparative Studies of Psychotherapies: Is It True that 'Everybody Has Won and *All* Must Have Prizes'?" Presented at 3rd Annu. Meet. Soc. Psychother. Res. Nashville, June 16, 1972. Unpublished.

Luborsky, L. and H. H. Strupp. "Research Problems in Psychotherapy: A Three-Year Follow-Up," in H. H. Strupp and L. Luborsky, eds., *Research In Psychotherapy,* Vol. 2, pp. 308-329. Washington: Am. Psychological Assoc., 1962.

Lyerly, S. B. and P. S. Abbott. *Handbook of Psychiatric Rating Scales* (1950-1964). U.S. Public Health Service Pub. No. 1495. Washington: U.S. Govt. Print. Off., 1966.

McGlashan, T., A. Campbell, P. Cleary et al. *The Documentation of Clinical Psychotropic Drug Trials.* Washington: NIMH, 1973.

McNair, D. M. "Self-Evaluations of Antidepressants." Unpublished.

May, P. R. A. *Treatment of Schizophrenia: A Comparative Study of Five Treatment Methods.* New York: Science House, 1968.

----. "Psychotherapy and Ataraxic Drugs," in A. E. Bergin and S. L. Garfield, eds., *Handbook of Psychotherapy and Behavior Change: An Empirical Analysis,* pp. 495-540. New York: Wiley, 1971.

May, P. R. A., A. H. Tuma, and W. Kraude. "Community Follow-up of Treatment of Schizophrenia—Issues and Problems," *Am. J. Orthopsychiatry,* 35 (1965), 754-763.

Mellinc.er, G. D., M. B. Balter, H. J. Parry et al. "An Overview of Psychotherapeutic Drug Use in the United States." Presented at Conf. Epidemiology of Drug Use. San Juan, Puerto Rico, Feb. 12-14, 1973. Unpublished.

Meltzoff, J. and M. Kornreich. *Research in Psychotherapy.* New York: Atherton, 1970.

Moos, R. H. *Revision of the Ward Atmosphere Scales (WAS): Technical Report.* Stanford: Stanford

University Press, 1971.

Morrison, D. F. *Multivariate Statistical Methods.* New York: McGraw-Hill, 1967.

Nie, N., D. H. Bent, and C. H. Hull. *Statistical Package for the Social Sciences.* New York: McGraw-Hill, 1970.

O'Brien, C. P., K. B. Hamm, B. A. Ray et al. "Group vs. Individual Psychotherapy with Schizophrenics," *Arch. Gen. Psychiatry,* 27 (1972), 474-478-

Orne, M. T. "On the Social Psychology of the Psychological Experiment: With Particular Reference to Demand Characteristics and Their Implications," *Am. Psychol.,* 17 (1962), 776-783.

Overall, J. E., L. E. Hollister, and S. N. Dalal. "Psychiatric Drug Research: Sample Size Requirements for One vs. Two Raters," *Arch. Gen. Psychiatry,* 16 (1967), 152-161.

Park, L. C .and L. Covi. "Non-Blind Placebo Trial: An Exploration of Neurotic Patients' Responses to Placebo when Its Inert Content Is Disclosed," *Arch. Gen. Psychiatry,* 12 (1965), 336-345-

Park, L. C., E. H. Uhlenhuth, R. S. Lipman et al. "A Comparison of Doctor and Patient Improvement Ratings in a Drug (Meprobamate) Trial," *Br. J. Psychiatry,* 111 (1965), 535-540.

Parloff, M. B., H. G. Kelman, and J. D. Frank. "Comfort, Effectiveness and Self-Awareness as Criteria of Improvement in Psychotherapy," *Am. J. Psychiatry,* 111 (1954), 343-351-

Pasamanick, B., F. R. Scarpiti, and S. Dinitz. *Schizophrenics in the Community: An Experimental Study in the Prevention of Hospitalization.* New York: Appleton-Century-Crofts, 1967.

Paul, G. L. *Insight vs. Desensitization in Psychotherapy: An Experiment in Anxiety Reduction.* Stanford: Stanford University Press, 1966.

----. "Behavior Modification Research: Design and Tactics," in C. M. Franks, ed., *Behavior Therapy: Appraisal and Status,* pp. 29-62. New York: McGraw-Hill, 1969.

----. "Outcomes of Desensitization II: Controlled Investigations of Individual Treatment, Technique Variations, and Current Status," in G. M. Franks, ed., *Behavior Therapy: Appraisal and Status,* pp. 105-159. New York: McGraw-Hill, 1969.

Pollack, E. S. "Monitoring a Comprehensive Mental Health Program: Methodology and Data Requirements," in L. M. Roberts, N. S. Greenfield, and M. H. Miller, eds., *Comprehensive Mental Health: The Challenge of Evaluation*. Madison: University of Wisconsin Press, 1968.

Prien, R. F., E. M. Caffey, Jr., and C. J. Klett. "Comparison of Lithium Carbonate and Chlorpromazine in the Treatment of Mania," *Arch. Gen. Psychiatry,* 26 (1972), 146-153-.

Prien, R. F., C. J. Klett, and E. M. Caffey, Jr. A *Comparison of Lithium Carbonate and Imipramine in the Prevention of Affective Episodes in Recurrent Affective Illness,* Prepubl. Rep. no. 94. Perry Point, Md.: VA Central Neuropsychiatric Research Laboratory, 1973.

Raskin, A. "Factors Affecting Response to Drug Treatment in Depressed Patients." Mimeo., 1972.

Reid, W. J. and A. W. Shyne. *Brief and Extended Casework.* New York: Columbia University Press, 1969.

Rickels, K., R. Cattell, A. MacAfee et al. "Drug Response and Important External Events in the Patient's Life," *Dis. Nerv. Syst.,* 26 (1965), 782-786.

Rickels, K. and R. W. Downing. "Minor Tranquilizers in Anxiety: Some Prognostic Indicators." Presented at Int. Symp. on Relaxation Therapy for Psychosomatic Disorders, St. Moritz, Switzerland, Jan. 13, 1971-Unpublished.

Rickels, K. and B. E. McLaughlin. "Sample Size in Psychiatric Drug Research," *Clin. Pharmacol. Ther.,* 9 (1968), 631-634.

Roberts, L. M., N. S. Greenfield, and M. H. Miller, eds. *Comprehensive Mental Health: The Challenge of Evaluation.* Madison: University of Wisconsin Press, 1968.

Rosenthal, R. *Experimenter Effects in Behavioral Research*. New York: Appleton-Century-Crofts , 1966.

Rossi, P. and W. Williams, eds. *Evaluating Social Programs.* New York: Seminar Press, 1972.

Rubinstein, E. A. and M. B. Parloff, eds. *Research in Psychotherapy.* Washington: Am. Psychological Assoc., 1959.

Saenger, G. "Patterns of Change among 'Treated' and 'Untreated' Patients Seen in Community Mental Health Clinics," *J. Nerv. Ment. Dis.,* 150 (1970), 37-50.

Sargent, H. D. "Intrapsychic Change: Methodological Problems in Psychotherapy Research," *Psychiatry,* 24 (1961), 93-108.

Schulberg, H. C., A. Sheldon, and F. Baker, eds. *Program Evaluation in the Health Fields.* New York: Behavioral Pub., 1969.

Schwartz, C. C., J. K. Myers, and B. M. Astrachan. "The Outcome Study in Psychiatric Evaluation Research: Issues and Methods," *Arch. Gen. Psychiatry,* 29 (1972), 98-102.

Shapiro, S. "End Result Measurements of the Quality of Medical Care," *Milbank Mem. Fund Q.,* 51 (1967), 49-72.

Shlien, J. M., H. F. Hunt, J. D. Matarazzo et al., eds. *Research in Psychotherapy,* Vol. 3. Washington: Am. Psychological Assoc., 1968.

Sonquist, J. A. and N. J. Morgan. *The Detection of Interaction Effects.* Ann Arbor: University of Michigan Press, 1964.

Stephens, J. A. and C. Astrup. "Treatment Outcome in 'Process' and 'Nonprocess' Schizophrenics Treated by 'A' and 'B' Types of Therapists," *J. Nerv. Ment. Dis.,* 140 (1965), 449-456.

Stollak, G. E., B. G. Guerney, and M. Rothberg, eds. *Psychotherapy Research: Selected Readings.* Chicago: Rand McNally, 1966.

Stone, A. R., J. D. Frank, E. H. Nash et al. "An Intensive Five-Year Follow-up Study of Treated Psychiatric Outpatients," *J. Nerv. Ment. Dis.,* 133 (1961), 410-422.

Strupp, H. H. and A. E. Bergin. *Research in Individual Psychotherapy: A Bibliography.* U.S. Public

Health Service Pub. no. 1944. Washington: Govt. Print. Off., 1969-

----. "Some Empirical and Conceptual Bases for Coordinated Research in Psychotherapy: A Critical Review of Issues, Trends, and Evidence," *Int. J. Psychiatry,* 7 (1969), 18-90.

Strupp, H. H. and L. Luborsky, eds. *Research In Psychotherapy,* Vol. 2. Washington: Am. Psychological Assoc., 1962.

Suchman, E. A. *Evaluative Research: Principles and Practice in Public Service and Social Action Programs,* pp. 31; 77. New York: Russell Sage Foundation, 1967.

Thompson, C. P. and N. W. Bell. "Evaluation of a Rural Community Mental Health Program, *Arch. Gen. Psychiatry*, 20 (1969), 448-456.

Truax, C. B. and R. R. Carkhuff. *Toward Effective Counseling and Psychotherapy: Training and Practice*. Chicago: Aldine, 1967.

Truax, C. B. and K. M. Mitchell. "Research on Certain Therapist Interpersonal Skills in Relation to Process and Outcome," in A. E. Bergin, and S. L. Garfield, eds., *Handbook of Psychotherapy and Behavior Change: An Empirical Analysis,* pp. 299-344. New York: Wiley, 1971.

Uhlenhuth, E. H., L. Covi, and R. S. Lipman. "Indications for Minor Tranquilizers in Anxious Outpatients," in P. Black, ed., *Drugs and the Brain: Papers on the Action, Use and Abuse of Psychotropic Agents,* pp. 203-221. Baltimore: The Johns Hopkins Press, 1969.

Uhlenhuth, E. H. and D. B. Duncan. "Subjective Change with Medical Student Therapists: II. Some Determinants of Change in Psychoneurotic Outpatients," *Arch. Gen. Psychiatry,* 18 (1968), 532-540.

Uhlenhuth, E. H., D. B. Duncan, and L. C. Park. "Some Nonpharmacologic Modifiers of the Response to Imipramine in Depressed Psychoneurotic Outpatients: A Confirmatory Study," in P. R. A. May and J. R. Wittenborn, eds., *Psychotropic Drug Response: Advances in Prediction,* pp. 155-197. Springfield, Ill.: Charles C. Thomas, 1969.

Vaillant, G. E. "Prospective Prediction of Schizophrenic Remission," *Arch. Gen. Psychiatry,* 11 (1964), 509-518.

Wartski, S. A. and D. S. Green. "Evaluation in a Home Care Program," *Med. Care,* 9 (1971), 352-364.

Waskow, I., ed. *The Evaluation of Psychotherapy Outcome,* forthcoming.

Weissman, M. M. "The Assessment of Social Adjustment: Measures Suitable for Antidepressant Trials." Mimeo., 1972.

Whitehorn, J. C. and B. J. Betz. "A Study of Psychotherapeutic Relationships between Physicians and Schizophrenic Patients," *Am. J. Psychiatry,* 111 (1954), 321-331.

Wittenborn, J. R. *The Clinical Pharmacology of Anxiety.* Springfield, Ill.: Charles C. Thomas, 1966.

Woodruff, R. A., D. W. Goodwin, and S. B. Guze. *Psychiatric Diagnosis.* New York: Oxford University Press, 1974.

Zolik, E. S., E. M. Lantz, and R. Sommers. "Hospital Return Rates and Prerelease Referrals," *Arch. Gen. Psychiatry,* 18 (1968), 712-717.

Zusman, J., V. Hannon, B. Z. Locke et al. *Bibliography on Epidemiology of Mental Disorders,* Publ. no. 50.30. Washington: National Clearinghouse for Mental Health Information, 1970.

Zusman, J. and E. R. R. Ross. "Evaluation of the Quality of Mental Health Services," *Arch. Gen. Psychiatry,* 20 (1969), 353-357.