
4.

**“Literature on the Effectiveness
of Psychotherapy**

Literature on the Effectiveness of Psychotherapy

Undoubtedly, there is confusion about the effectiveness of psychotherapy; both in the professional and popular literature, a variety of claims have been made about the worth of psychotherapeutic treatments. Although it is not possible here to resolve fully the differences between conflicting viewpoints, the present chap-

ter briefly reviews the major areas of dispute and attempts to apply the methodological principles described in the previous chapter to assess the various claims about psychotherapy. The legitimate disagreements among reviewers are acknowledged, and an effort is made to assess the differences objectively.

SCOPE OF REVIEW

Although no one has classified all the research conducted on psychotherapy, it is clear from even a cursory review of the literature that there is a plethora of research (see, e.g., 208,276). The research ranges from very specific analyses of psychotherapeutic procedures to large-scale program evaluations. Clearly, there are differences in the usefulness of particular types of research for the general question of psychotherapy's efficacy, but it is difficult to differentiate research according to a relevance criterion.

While some commentators have made a distinction between basic and evaluative research (e.g., 66, 268), the utility of any research to answer efficacy questions is usually a matter of degree. Basic research on psychotherapy does not indicate whether or not treatments are effective as practiced, but does provide some understandings about the nature of the intervention. For those concerned with understanding the processes underlying psychotherapy, such data are very important. Conversely, evaluative studies, although they provide direct information about the actual effectiveness of a treatment, often yield equivocal analyses of causality and must be interpreted in conjunction with basic research data. In the present chapter, categorizations are avoided; the analysis includes a variety of research studies which have been considered, by at least some reviewers, to be rele-

vant to understanding the efficacy problem. Several additional caveats should also be noted.

One important issue relating to how one assesses the efficacy literature is the nature of the problems for which psychotherapy treatments are offered. Thus, a great deal of research examines the treatment of very specific nondisabling mental/behavioral problems, and it is difficult to assess the generalizability of this literature to understanding the efficacy of treatments for very severe and disabling conditions. Unfortunately, the problem is even more complex, for much of the psychotherapy literature deals with conditions that are difficult to classify. In part, this problem has to do with the fact that definitions of severity are relative and dependent on the context in which the problem exists. The present chapter attempts to clarify, wherever possible, the degree to which findings about particular problems are generalizable to other mental health problems.

In terms of the scope of the present chapter, the main focus is on a description and analysis of previous reviews of the literature. As discussed in chapter 3, no matter how well conceived and executed, single studies have limited use within a scientific framework. A number of very extensive reviews of the literature (i.e., discussions of multiple studies) have already been conducted, and an effort is made in the

present chapter to describe these reviews as fully as possible. Where appropriate, as part of this analysis, the data and methods of particular studies discussed by these reviews are presented. OTA's decision to focus on reviews has to do both with a goal to best represent the available literature and with the fact that these reviews comprise a focal point for much of the current debate about psychotherapy's effectiveness.

One omission from the literature reviews in the present chapter are popularized summaries of the psychotherapy literature. For example, Gross' recent book *The Psychological Society* has not been included. Gross' treatise is an interesting (although critical) report on psychotherapy, but is primarily a secondary source review of the literature (see also 279,281). Instead of discussing such work directly, the chapter describes the scholarly reviews on which Gross based his book. Not included here either are

some of the scholarly literature reviews which have as their primary focus the contrast of one therapy or approach with others. We were centrally interested in reviews that dealt, most generally, with the problem of whether psychotherapy is efficacious under what conditions. The central focus of a review on this question guided OTA's selections.

The reviews discussed below are presented in approximate chronological order according to their appearance in the literature. It is hoped that this ordering will give readers a sense of how research and thinking about effectiveness of psychotherapy have evolved during the past several decades. The chapter also describes the findings of a number of the most important studies on outcome. These descriptions are presented, where necessary, to facilitate understanding of the issues raised by reviewers.

EFFICACY REVIEWS

In the reviews described below, the data reported are reviewed first; then, some of the commentary that has been stimulated by the review is discussed. The methods used by the reviewer to select studies and analyze them are compared against the methodological criteria described in chapter 3. As noted above, the goal is to describe the research that is most relevant to the efficacy question and to assess what appear to be the most reasonable implications of this research literature.

Eysenck's Reviews

The earliest and probably the most controversial review of the psychotherapy outcome literature was conducted by Eysenck almost 30 years ago (70; see also 71,72). Eysenck, in his early review, considered 24 research articles which included 8,053 cases of psychotherapy conducted with neurotic patients. He divided the studies, on the basis of his assessment of their therapeutic approach, into two groups: 1) psychoanalytic therapy and 2) eclectic therapy. The principal criterion for assessing effectiveness

was a rating Eysenck developed of improvement following therapy.

The most controversial aspect of Eysenck's report was his comparison of improvement rates in therapy against an improvement rate that he calculated for two no-treatment groups. In the eclectic therapy studies, Eysenck calculated an improvement rate for therapy of 64 percent within 2 years. He compared this rate with what was achieved in comparison groups of patients who did not have therapy. Eysenck found that the no-treatment improvement rate was approximately the same as that achieved in the eclectic therapy studies. The improvement rate calculated for psychoanalysis was lower, approximately 44 percent, and was below that of the no-treatment groups. One source of Eysenck's data on no-therapy improvement was a study by Denker (61) of patients with emotional problems who were treated by general medical practitioners (another source was Landis (153)). Denker's data consisted of insurance company records on 500 individuals who had submitted mental disability claims. He had found that

within 1 year, without receiving any specific psychotherapy, 44 percent had returned to work, and that by the end of 2 years, an additional 27 percent had returned to work.

Eysenck's work unleashed an extraordinary reaction, some of which still affects thinking about psychotherapy research (17, 18, 20, 160, 181, 242, 268). Much of the reaction has been negative, and, in response, Eysenck and his colleagues have tried to refute claims of his critics and to provide additional data (e. g., 70, 71). The debate between Eysenck and other researchers is important to unravel, both because of its centrality in discussions about psychotherapy, and because of its implications for the most important questions about psychotherapy. In terms of Eysenck's original report, three types of problems seem to be important in interpreting his analysis: 1) the nature of the data that Eysenck reviewed, 2) the utility of Denker's comparison data, and 3) Eysenck's interpretation of the data.

The first problem, that of specificity of the data, has been raised by a number of Eysenck's critics (e. g., 160). In essence, these reviewers have claimed that it was inappropriate to generalize very widely about psychotherapy on the basis of Eysenck's data. The most important difficulty is that it is not clear what was done in each of the studies. One cannot determine how much therapy was actually received by patients and what the quality of this treatment was (19). It also seems important to note that the research reported by Eysenck was conducted prior to 1950, at a time when nonpsychodynamically based therapies were only beginning to be used. The generalizability of this work to the types of nonpsychodynamic therapies carried out today, as well as to currently practiced psychoanalytic treatments, is not at all clear (see, e.g., 269).

The second problem has to do with the "control" group data used by Eysenck in his analysis. Eysenck's control data for the effects of treatment were drawn from a nonrandomly selected control group. It is difficult to determine how this nontreated group might have differed from individuals who received therapy. They may have been more or less troubled and dysfunctional than those who sought treatments.

Not only are the differences between the control and treatment group subjects not clear, but as noted by Meltzoff and Kornreich (181), neither Denker's (61) nor Landis' (153) data were representative (i. e., the data were not derived from a survey).

Meltzoff and Kornreich (181) also note that the control group data did not really represent no-treatment data. Especially in Denker's study, where patients were diagnosed by general medical practitioners, a form of treatment was provided. Patients were provided "sedatives, tonics, as well as reassurance, and a placebo type of treatment." Furthermore, since none of these patients was severely disturbed to begin with, the return-to-work measure may not indicate an actual remission of the symptoms that first brought the individual to the practitioner.

Finally, a number of commentators have pointed to errors in Eysenck's categorizing of studies and errors in the way he handled data. Bergin (19) reanalyzed the data used in Eysenck's original report and finds a different rate of recovery for psychoanalytic treatment. What Bergin did, among other procedures, was to eliminate cases where individuals left therapy; Eysenck had, instead, counted these cases as failures. Bergin also claims to have found coding errors in Eysenck's original report (e. g., outcome data incorrectly represented). Bergin's reanalysis had yielded an improvement rate of 65 percent for the eclectic therapies (similar to Eysenck's). He notes that improvement rates for all types of treatment may be deceptive, because there was a great deal of systematic variation across personality types, therapists, and clinics. Bergin concludes that global questions of psychotherapeutic efficacy are "silly" and that one must analyze specific therapies for specific problems.

Eysenck acknowledges many of the methodological problems with his data and has argued that the burden of proof to provide evidence should be on those who seek to promote psychotherapy (see 223). He has also updated his original research, and, in later reviews (e. g., 69), cites an additional 11 studies. Although his initial conclusion about the lack of effectiveness of psychotherapy remains, he finds supportive

evidence for at least one type of psychotherapy, a behavioral approach based on the work of Wolpe. It should also be noted that a number of commentators (e.g., 181) have criticized Eysenck for selectively reviewing the literature. By the time of his 1965 paper (which was published along with 17 critiques by prominent researchers), there were at least 70 control group studies of psychotherapy. Most of these were not included in Eysenck's review, although his reasons for not selecting these studies are not clear.

Although Eysenck's work is often interpreted as an indictment of all psychotherapy, as noted above, Eysenck (69) cites evidence as to the efficacy of therapies based on behavior theory. Thus, he describes a study conducted by Lazarus (155) of phobic patients who were treated either by group desensitization (a behavior therapy), group interpretation (a psychodynamic therapy), and a combination of nonbehavioral methods. The results showed a 72-percent recovery rate for the behavior therapy versus a 12-percent recovery rate for the verbal therapy. Eysenck also relies on an unpublished report by Wolpe (the developer of desensitization therapy) in which the results of Wolpe's investigations (122 cases) were compared with reported results from two large psychoanalytic institutions (approximately 400 cases). Wolpe had reported that 90 percent of his cases were cured or much improved versus 60 percent of the psychoanalytically treated patients.

In trying to make sense of the controversy over Eysenck's research, a number of aspects of his work should be highlighted. First, whether one accepts his findings or not, the primary purpose of the original review was to assess the data for treatment of neurotics. Severe mental disability was considered only secondarily in later reports. Second, while Eysenck's critique has been interpreted as a critique of all psychotherapy (and psychotherapy research), his comments are most critical of psychoanalytically derived therapies. He is supportive, as described above, of therapies based on behavior theory. Finally, it should be noted that Eysenck did not find psychotherapy to be harmful and, in fact, found that eclectic therapies were associated with fairly high rates of improvement. The criti-

cal question—whether improvement was caused by psychotherapy or resulted from other factors (e.g., spontaneous remission, placebo effects)—is not resolved by Eysenck's work.

Meltzoff and Kornreich's Analysis

According to Meltzoff and Kornreich (181), their work was stimulated by a belief that there had been advances in the development of research on the effectiveness of psychotherapy since Eysenck (who initially published in the early 1950's), but that this work was underrepresented in the literature. Meltzoff and Kornreich, in their book-length discussion of the psychotherapy research literature, made an important departure from earlier reviews and classified studies by methodological adequacy. Groups of studies were formed according to their methodological quality and compared in terms of the positiveness of findings. Meltzoff and Kornreich wanted to come up with a "fair" assessment of what was contained in the available literature.

In their category of "adequate" research designs, they included studies which used a control group condition and adequate outcome measures. In their "questionable" research design category, they included studies with control groups that may not have been comparable and/or used poor outcome measures or analysis procedures. They also included in this latter category studies that used analog designs. The results of the studies they assessed were categorized as either "positive," "negative," or null. Since most studies included multiple outcome measures, their subdivision of outcomes was a judgment of the balance of statistically significant findings. Their analysis included approximately 100 control group studies.

According to Smith, et al. (268), who tabulated the results of Meltzoff and Kornreich's review, 80 percent of the controlled studies they reported yielded positive results, while the other 20 percent had null or negative results. There was also a positive relationship between research quality and positive findings. Thus, 84 percent of Meltzoff and Kornreich's adequately designed studies yielded positive results about

the effectiveness of psychotherapy, while only 33 percent of the studies of questionable design yielded positive results.

Meltzoff and Kornreich (181) also assessed the degree to which results were “major” or “minor.” The studies categorized as “adequately designed, with major outcomes” varied on a number of dimensions. These studies included research such as Grace, Pinski, and Wolff’s (107) assessment of group therapy for individuals with ulcerative colitis. In this study, patients in the treatment group received psychotherapy to alleviate stress. Over a 2-year period, it was found across a variety of “hard” criteria (e.g., operations required) that those who received group therapy were less likely to require medical treatment than those who did not (there was a lower morbidity rate for therapy patients, as well).

A different type of study included by Meltzoff and Kornreich (181) was Morton’s (190) study of college students who were referred to psychotherapy for severe personal problems. A randomized control group design was implemented by having some subjects placed on a waiting list. Raters, who were blind to information about the students’ treatment (they were the counselors who made the actual referral), talked with students after 3 months. On a number of measures, including those that assessed cognitive and behavioral variables, 93 percent of the treatment group students and only 47 percent of the control group students showed improvement. The students who received psychotherapy were considered to have received a major benefit.

Although it is difficult to systematically analyze the types of studies included in Meltzoff and Kornreich’s (181) review, most of the studies they report appear not to be of severely disabled patients. In fact, almost all the studies in which there were adequate designs and major positive benefits were studies in which the problems were relatively specific and the patients were not institutionalized. The reason for this may be connected with the problems of doing control group research, as well as with the nature of psychotherapy. As noted in chapter 3, it is easier to conduct well-designed control group research when problems are specific and the

potential harm of withholding treatment is not severe. The hypothesis must be left open that the more general and unspecified the problem, the less efficacious is psychotherapy.

This point has been made by Malan (168) about Meltzoff and Kornreich’s work (181). Malan noted that, although he agreed with the Meltzoff and Kornreich conclusion about the beneficial effects of psychotherapy, its generalizability for a variety of patient groups was not clear. There is also a problem, in generalizing from Meltzoff and Kornreich, of not knowing what specific techniques work with what specific problems. Although their methodological focus is useful for assessing hypotheses about the relation between research quality and outcome, it obscures the synthesis of findings about specific treatments. From a more optimistic perspective, it is comforting that their positive findings rate is a bit higher than that of Eysenck, and further, that these positive results were obtained in well-designed studies.

Bergin’s Reviews

Bergin’s 1971 review (19) of psychotherapeutic effectiveness was described briefly above in reference to his critique of Eysenck’s work. Bergin’s analysis is important (see also 18,21) because, unlike earlier critics of Eysenck, he actually examined the studies used in the original report (72) and recomputed Eysenck’s treatment remission rates. Bergin used different assumptions and procedures for establishing improvement rates, and, as noted above, he found a discrepancy in Eysenck’s analysis of psychodynamic therapy. In addition, Bergin (19) calculated a different remission rate for no-treatment (based on data not discussed by Eysenck), and found that only about 30 percent of the patients would have recovered had there not been psychotherapy. Within the framework of Eysenck’s nonexperimental data base, there are limits to one’s certainty about any conclusions.

Bergin (19), in part to remedy this problem, reviewed 52 studies of psychotherapy outcome. He claimed that these studies represented a “cross-section” of the available literature (see 276). Approximately half of the studies had control groups, and Bergin categorized the studies on the basis of a number of variables, including

the adequacy of the research design, duration of therapy, type of therapy, and therapist experience. Studies were judged and cross-tabulated in terms of the results of the therapy: “positive,” “negative,” or “in doubt.”

Of the 52 studies, Bergin (19) judged 22 to be positive (indicating psychotherapy was effective), he judged 15 as having negative results, and 15 were judged “in doubt.” Bergin concluded that “psychotherapy, on the average, has moderate positive results.” Typical of the studies Bergin classified as positive was Gottschalk, Mayerson, and Gottlieb’s (106) study of brief therapy. Using a no-treatment control group design, Gottschalk, et al. studied neurotics who were given emergency outpatient therapy. The treated patients improved significantly more than the controls on a variety of behavioral and cognitive measures. Another positive study cited by Bergin was Vorster’s (288) investigation of neurotics treated by eclectic therapies. Vorster found the treatment group significantly improved, although this study’s design did not employ a randomized control group.

In a more recent extension of Bergin’s (19) review, Bergin and Lambert (21) discuss additional evidence to support their hypothesis that psychotherapy is efficacious. They also reply to the criticism (e.g., 222) of Bergin’s original work. Bergin and Lambert cite as one of the best examples of more recent research a study by Sloane, Staple, Christol, Yorkston, and Whipple (266). That study used 90 outpatients, of whom two-thirds were neurotic and one-third had personality conduct disorders. The patients were randomly assigned (they were also matched with respect to sex and severity of symptoms) to short-term analytically oriented therapy, behavior therapy, or a minimal treatment wait-list group. Treatment outcome was assessed after 4 months. On average, all groups improved significantly across a set of target symptoms, but the two therapy groups (analytically-oriented and behavior) improved significantly more than the minimal treatment wait-list group. No significant differences were found in outcomes between the actual therapy groups.

Bergin and Lambert (21) conclude that, “Psychoanalytic/insight therapies, humanistic or pa-

tient-centered psychotherapy, many behavioral techniques and, to a lesser degree, cognitive therapies rest on a reasonable empirical base.” They contend that these therapies achieve results that are superior to no-treatment and to various placebo treatment procedures. Bergin and Lambert also assert that even if one accepts Eysenck’s “two-thirds remission within 2 years” formula, there is still positive evidence for psychotherapy. This is because treatment effects of the same magnitude are frequently obtained in 6 months or less in formal psychotherapy, “. . . a considerable evidence of therapy’s efficiency/efficacy over no treatment.” In effect, psychotherapy hastens the recovery period and reduces both suffering and the effects of disability.

Rachman’s Critique

As can be seen from the above two reviews, the early 1970’s was a fertile time for psychotherapy assessments. Rachman (222), a frequent collaborator of Eysenck’s, provided another review which served, in part, to refute Eysenck’s critics. Rachman, in particular, responded to Bergin’s (19) reanalysis of Eysenck’s data and to the dispute about what studies should be included in a review of psychotherapy efficacy. Rachman, in analyzing Bergin’s work, disallowed a number of studies (5 out of 14), because the subjects were delinquents or had psychosomatic complaints, rather than being neurotic. If Rachman’s analysis had been extended to Meltzoff and Kornreich’s (181) review, the same problem would have existed.

Rachman (222) analyzed 23 studies which he considered to be relevant for assessing the efficacy of “verbal” psychotherapy (i.e., therapies based on conversation between therapist and patient—primarily, humanistic and psychodynamically based). Of these studies, Rachman found that only one provided tentative evidence of the effectiveness of psychotherapy. Five studies produced negative effects, where treatment group results failed to exceed control groups or base-line remission rates. It is important to note that Rachman excluded 17 studies for a variety of reasons, only 2 of which showed negative effects. The reasons for excluding the 15 studies that showed positive results varied, and in-

cluded the use of unacceptable outcome measures (e. g., a projective test), the exclusion of subjects who left therapy before treatment was completed (which Rachman said inflated improvement rates), and inconsistency across outcome measures (i. e., some measures showed positive effects, while one or more measures showed negative effects). The two negative studies were excluded because randomized control group conditions were not employed.

A number of commentators (e.g., 257,268) have reviewed Rachman's (222) critique and analysis, and they claim that he used inconsistent standards for evaluating evidence. In particular, they assert that Rachman used criteria for assessing the verbal psychotherapies that were different from the criteria he used for the behavioral therapy experiments. Smith, et al. (268) have criticized Rachman because he selectively chose studies to review. They indicate that many more studies were available to Rachman than he included, and add that it cannot be inferred from his discussion why particular studies were not referenced. Smith, et al. also criticize Rachman for "ex post facto" exclusion of studies based on methodological criteria. They argue that he should have compared the good and poor designs to determine whether or not they yielded different kinds of conclusions, as Meltzoff and Kornreich (181) did in their analysis.

Luborsky, Singer, and Luborsky's Review

In a more current assessment, Luborsky, Singer, and Luborsky (162) tried to evaluate all reasonably controlled studies of psychotherapy on "real" patients (i. e., analog studies were excluded). Luborsky, et al. examined available studies that had assessed therapy for treating recognized problems of individuals who sought psychological treatment. Many of these studies were ones that had been used by Bergin (19) and by Meltzoff and Kornreich (181). Although Luborsky, et al. 's definition excluded some behavioral research (primarily because it was not conducted in actual clinical settings) and some patient populations, their scope was wider than that of other reviewers.

Each of the studies Luborsky, et al. (162) assessed was categorized on a number of dimensions and then summarized in a "box-score" analysis (see ch. 3). One central dimension was research quality, which was determined on the basis of 12 criteria. The criteria included the study's method for assigning subjects to comparison groups, the procedures for dealing with premature therapy termination, experience of therapists, tailoring of outcome measures to therapeutic goals, and the adequacy of the sample size. Luborsky, et al. summarized the codings by "grading" each study on a 5-point scale of research quality. They then categorized results in terms of whether there were significant differences showing better effects for the treatment group (+), the comparison group (-), or no significance between the groups (0).

Luborsky, et al. (162) found 33 studies in which psychotherapy treatment groups were compared with no-treatment control groups. Of these, 20 studies yielded psychotherapy treatment groups which were significantly better off than control groups, and 13 showed no difference. Luborsky, et al. did not find any instance of a control group being better than a psychotherapy treatment group. They found 19 studies in which schizophrenic populations (i. e., severely disabled individuals) were studied. Of these, 11 yielded results in favor of the psychotherapy condition, and 8 yielded no differences. Luborsky, et al. also found that in a majority of comparisons (13 out of 19), there were no significant differences in outcomes to patients between behavior therapy and other psychotherapies. While the reviewers note the positive research findings on behavior therapies, they suggest the need for more studies in which behavior therapies are applied to patients who have generalized maladjustments.

Luborsky, et al. (162) conclude that control studies find that patients who go through psychotherapy do, in fact, gain. Because they used "box scores" where effect sizes were not estimated (see 221), it is not possible to determine how strong these effects will be. However, according to Smith, et al. (268), there has been no published substantive criticism of Luborsky, et al. and (as is described below) Luborsky, et

al. 's general finding has been substantiated by other reviews.

NIMH Synthesis

Perhaps the most comprehensive review of psychotherapy outcome research has been conducted at the National Institute of Mental Health by Parloff and his colleagues (208). This review provided an assessment of psychosocial treatments for mental disorders and was prepared for the Institute of Medicine (National Academy of Sciences) as part of that Institute's work for the President's Commission on Mental Health (219). The Parloff, et al. review differs from earlier works in that it uses a more narrowly defined treatment and is organized primarily according to disabling conditions. Thus, for each of a variety of mental disorders, the available evidence as to the effectiveness of psychotherapy was assembled and analyzed. The reviewers tried to make some general statements about psychosocial therapies (psychotherapies that do not use drug treatments) and how these therapies are affected by other variables, such as therapist and patient characteristics. Parloff, et al. 's work is very extensive, so a detailed summary is not attempted here; only the work's general conclusions are described below.

Parloff, et al. 's (208) general finding (see also 207) was that "patients treated by psychosocial therapies show significantly more improvement in thought, mood, personality, and behavior than do comparable samples of untreated patients." These reviewers found that spontaneous remission rates developed from separate samples provide evidence that psychosocial treatments seem to result in greater improvement than would be expected without psychotherapeutic treatment. Their finding is supported most clearly for disorders such as "anxiety states, fears, and phobias." Parloff, et al. 's relatively positive assessment, however, was accompanied by a number of caveats. For many disorders, psychotherapy alone (i.e., without other treatments such as drugs) has not been demonstrated to be effective; nor does it appear to be effective for particular populations (e.g., children). Finally, the review notes, although effectiveness evidence exists for a number of dis-

orders, only some types of therapies (with particular therapists) may be effective.

The central aspect of Parloff, et al.'s (208) review was a summary, by each psychopathological condition, of the available treatment research evidence. To appreciate the complexity of this task, consider their discussion of severe mental disorders such as schizophrenia. For these disorders, Parloff, et al. found that individual and group psychotherapy provide an ambiguous amount of improvement for institutionalized patients; however, in conjunction with drug therapies and other psychological treatments, they appear to have important effects. The authors note that ". . . drugs do not teach individual social and interpersonal skills." For such hospitalized populations, however, Parloff, et al. found considerable evidence that a specific type of therapy (behavior based) improves social adjustment (on a variety of social and interpersonal variables). They also found that the return of severely disturbed patients to their community had positive effects on treatment outcome, although this finding was limited to patients with certain interaction skills, and under the condition that the patient returns to a "good" family situation.

One important feature of the Parloff, et al. work, both as it was presented in the original form (208) and as it was summarized by the President's Commission on Mental Health (219), is that this report examines the evidence for alternative hypotheses about the effects of psychotherapy. In particular, Parloff, et al. found a variety of reported spontaneous remission rates; that is, different improvement rates are obtained for disturbed patients who do not receive therapy (see, e.g., 280). Despite the fact that patients improve to some extent without therapy, however, the hypothesis that such spontaneous remission effects account for changes in treated patients cannot be validated. Parloff, et al. (208) report that studies which have been controlled for placebo effects find that changes associated with treatment are greater than those associated with the placebo. Unfortunately, there is an inherent problem in theoretically identifying the nature of a psychotherapy placebo (see ch. 3) and separating it

from the effects of the treatment, so these results must be regarded as tentative.

Parloff (207), in a discussion of the review (208), calls for the development of clinical trial research to assess more widely and systematically the effects of psychotherapy. While Parloff adopts a relatively positive view of the effects of psychotherapy, he finds the current research literature limited. He proposes the use of experimental designs to assess the specific conditions (including aspects of the treatment, therapist, and patient) under which therapy procedures yield particular outcomes. Although one view of such clinical trials would be to test different theories of psychotherapy, an implication of Parloff, et al.'s work is that other factors need to be incorporated in the design of these trials.

One additional note about Parloff, et al.'s (208) findings has to do with their review of behavior-based therapies. Parloff, et al., along with Eysenck, Rachman, Bergin, and to some extent, Meltzoff and Kornreich, report clear-cut evidence that behavior-based therapies are effective treatments for specified conditions. Although its generalizability has not been established across the range of disorders, this finding suggests one important focus for future research.

Smith, Glass, and Miller's Meta-Analysis

The final review considered here, that by Smith, Glass, and Miller (268), has the potential to be as controversial as Eysenck's original work, and may stimulate an entire new set of psychotherapy outcome analyses. The meta-analytic methods used by Smith, et al. (see ch. 3) are at the heart of the controversy about this work. A preliminary published report of the meta-analysis of 375 control group studies of psychotherapy (267) has already stimulated a variety of critiques (e.g., 73,90,218). Smith, et al.'s goal was to determine the state of knowledge about the effects of psychotherapy, using systematic scientific procedures. Their meta-analytic procedure was used to integrate the findings of a disparate set of studies on psychotherapy. In addition, each of the studies in the sample was classified to enable determination of the factors that influence outcome findings.

In order to conduct their review, Smith, et al. (268) tried to include all controlled studies of the effectiveness of any form of psychotherapy. Controlled studies were defined as investigations where a treatment group received psychotherapy, and another group which was comparable did not receive treatment. In some cases, receiving treatment meant a placebo or alternate form of psychotherapy. Smith, et al. considered a study to be relevant if the study investigated therapy that involved: 1) patients who were identified as having an emotional or behavioral problem, 2) treatment that was psychological or behavioral, and 3) therapists who were identified professionals. The definition resulted in the analysis' including a variety of investigations of treatments applied to problems of different degrees of severity. Smith, et al. excluded some types of treatment, including those in which psychoactive drugs were used (these studies were analyzed separately), those which were primarily educational, and those which were not, essentially, psychosocial treatments. They surveyed a great number of sources to identify studies, including published journals, dissertations, and clearinghouses that identify professional publications. Their final sample included 475 controlled studies of psychotherapy.

Probably the most important aspect of Smith, et al.'s (268) analysis was the way they classified the research studies. Each study was coded on a number of dimensions related to the characteristics of the researcher, therapist, and the patient (including diagnosis); most importantly, the studies were classified on a series of methodological criteria. The methodological categories included the nature of the assignment to conditions (e. g., random v. matching) and such factors as experimental mortality and internal validity (see 41,165). Smith, et al.'s principal dependent measure was a standardized score for the size of the effect. From the data reported in each study, Smith, et al. calculated scores for the size of the effect; to allow comparison across studies, they computed each as a standard score. Studies often included more than one outcome measure, so Smith, et al.'s analysis treated each variable as a separate case. Thus, from the 475 studies Smith, et al. analyzed, they found 1,766 effect size measures. In addition, they

coded outcome measures in terms of the type of measure, instrument, and its reactivity (i. e., susceptibility to social desirability, faking, etc.).

Smith, et al.'s (268) principal finding was that, on the average, the difference between average scores in groups receiving psychotherapy and untreated control groups was 0.85 standard deviation units (i.e., the effect size difference was 0.85). According to Smith, et al., this average effect size can be translated to indicate that the average person who receives therapy is better off than 80 percent of the persons who do not. They found little evidence for the existence of harmful effects of psychotherapy (i. e., very few cases where the mean of the control group was higher than the treatment group). Smith, et al. found some significant differences across the types of therapies whose effects were studied (the range was 0.14 to 2.38), but these effects are confounded by variables such as patient and therapist characteristics which were distributed unequally among the therapies. Finally, their methodological categories proved not to correlate with effect sizes; thus, for example, the better designed studies did not yield less positive findings.

OTA has had an opportunity to review Smith, et al.'s (268) coded data of the 475 studies. A sample of studies was drawn, and their reliability was checked by comparing our blind codings with their original data. In particular, effect size ratings for a dozen studies were recomputed. This analysis indicated that the Smith, et al. codings were both easily replicable and apparently reliable. Their validity is more difficult to establish.

A key question is whether the effect size scores calculated by Smith, et al. (268) are a valid measure. There are reasons to suggest that their average effect size is inflated, but there are also reasons to suggest that it is conservative. Inflation in the effect size measure may have come about because only well-designed studies were included in Smith, et al.'s analysis. Thus, to the extent that the control group studies are completed only for successful psychotherapies or for a limited range of psychotherapy treatments, the sample may be biased positively when compared to psychotherapy research. It is

difficult to determine the exact nature of the bias that may result from this problem. There are also some aspects of Smith, et al.'s procedures which suggest that their average effect size measures may be conservative. Smith, et al. considered placebo treatments, as well as actual forms of therapy and counseling, in their treatment group means. These are all not legitimate therapies, and, in fact, when separately analyzed, showed lower effectiveness as compared to actual psychotherapy.

Most critics of Smith, et al. (268), in particular Eysenck (73), have disputed the meta-analysis approach (actually, the published critiques refer to Smith and Glass (267)), because it lumps together too many things and includes studies of poor design, as well as good design. This criticism is potentially justifiable, but a close review of Smith, et al. indicates that they control for this problem by their classification of studies according to methodological criteria. Not only do they start with a group of relatively well-designed studies (in terms of their definition of controlled psychotherapy research), but they provide analyses of the relationship between effect size and classification variables such as internal validity (a measure of the quality of the research design—see 41). The correlations are close to zero, indicating that studies that use randomized control group designs find the same effect as studies that use poorer designs. There may still be a sampling problem in that the published literature only reports well-controlled studies, but this problem is different from that on which most of the criticism of Smith, et al.'s meta-analytic procedures is based.

At this point, it is difficult to know how to utilize the results of Smith, et al. (268). Their work is certainly more systematic than that of other researchers; however, their methods are not yet widely accepted and their work has not been available long enough for comprehensive reviews to appear. Perhaps, the most important limitation of such meta-analytic research is that it relies on the existing literature. It seems clear that much better research on psychotherapy can be done, and a more definitive meta-analysis may have to await the completion of this new research.

DISCUSSION

If one considers only the trend of findings reported by scholarly reviews and analyses of the psychotherapy outcome literature, it would appear that psychotherapy treatments, under some conditions, have been shown to be efficacious (see, e.g., 287). Although the evidence is not entirely convincing, the currently available literature contains a number of good-quality research studies which find positive outcomes for psychotherapy. There are also a large number of studies which report positive effects, but whose methods or generalizability are difficult to assess. The quality of the evidence varies in terms of the nature of the treatment and the patient's problem. One difficulty with the available literature is that, while a host of factors have been identified as important to psychotherapeutic outcomes, the role of these factors (e.g., characteristics of the patient, therapist, setting) has not been assessed in any definitive way.

Frank (85), for example, in commenting on the state of psychotherapy outcome research, reports a disappointment that research has not produced more specific understandings of what occurs during therapy. While he supports the conclusion that research demonstrates that psychotherapy is somewhat better than no therapy, he also finds that much of the literature indicates equivocal outcomes. One reason for these equivocal findings, according to Frank, is that the personal qualities of the therapist and patient and their interaction may be more important than the therapeutic method. Because these factors vary so widely and are not typically controlled for, variability is built into

psychotherapy studies, making it more difficult to detect significant treatment effects. Frank urges more research that clearly assesses these components of therapeutic treatments.

It is likely, given the theoretical potential to conduct such research on the ingredients of psychotherapy, that a more extensive research literature could be developed (see also 207). It probably means that psychotherapy evaluations, unlike evaluations of new drugs, need to pay equal attention to the conditions of treatment and the treatment per se. Research done on psychotherapy in settings different from actual practice will probably have limited utility and will represent only one stage of a testing program. Although it would be hoped that psychotherapy research could lead to better understandings of the role of therapist characteristics and skills, it might only be possible to assess how these factors interact with treatments as they are actually delivered. Finally, it should also be noted that some forms of potential outcome research, such as program evaluations, do not seem to have made a substantial impact on the literature (cf. 181,212). While program evaluations have been done of a variety of process factors in mental health delivery systems, these studies have less frequently measured outcomes using experimental designs. In part, this reflects the developing nature of the methods for program evaluation (see ch. 3) and the slow shift to the funding of such research. It is possible that this situation will change with increased pressures for accountability and with increased emphasis on understanding complex sets of psychotherapy technologies.