THE BEHAVIORAL AND BRAIN SCIENCES (1983) 6, 275–310
*Printed in the United States of America*

# An analysis of psychotherapy versus placebo studies

**Leslie Prioleau**
*Department of Psychology, Wesleyan University, Middletown, Conn., 06457*

**Martha Murdock**
*Department of Psychology, Wesleyan University, Middletown, Conn., 06457*

**Nathan Brody**
*Department of Psychology, Wesleyan University, Middletown, Conn., 06457*

**Abstract:** Smith, Glass, and Miller (1980) have reported a meta-analysis of over 500 studies comparing some form of psychological therapy with a control condition. They report that when averaged over all dependent measures of outcome, psychological therapy is .85 standard deviations better than the control treatment. We examined the subset of studies included in the Smith et al. meta-analysis that contained a psychotherapy and a placebo treatment. The median of the mean effect sizes for these 32 studies was .15. There was a nonsignificant inverse relationship between mean outcome and the following: sample size, duration of therapy, use of measures of outcome other than undisguised self-report, measurement of outcome at follow-up, and use of real patients rather than subjects solicited for the purposes of participation in a research study. A qualitative analysis of the studies in terms of the type of patient involved indicates that those using psychiatric outpatients had essentially zero effect sizes and that none using psychiatric inpatients provide convincing evidence for psychotherapeutic effectiveness. The only studies clearly demonstrating significant effects of psychotherapy were the ones that did not use real patients. For the most part, these studies involved small samples of subjects and brief treatments, occasionally described in quasibehavioristic language. It was concluded that for real patients there is no evidence that the benefits of psychotherapy are greater than those of placebo treatment.

**Keywords:** meta-analysis; methodology; outcome research; placebo; psychotherapy

Eysenck's well-known (1952) paper is the first of a long series of studies dealing with the question of the effectiveness of psychotherapy. Eysenck argued that many patients recover spontaneously and that the changes following psychotherapy do not exceed the spontaneous recovery rate. Eysenck has reviewed literature on psychotherapy outcome on other occasions and has continued to argue that the studies suggest that psychotherapy is an ineffective treatment (see, e.g., Eysenck 1966). Other reviewers, more favorably disposed to psychotherapy, have argued that Eysenck distorted the data and dealt with a biased sample. Meltzoff and Kornreich (1970), for example, reviewed a larger body of work dealing with psychotherapy outcomes and argued that the better-designed studies tended to provide stronger evidence for the benefits of psychotherapy and that there was an ample body of convincing evidence suggesting that psychotherapy was an effective treatment.

Smith, Glass, and Miller (1980) have attempted to resolve the controversy surrounding the effectiveness of psychotherapy by using the statistical procedure of meta-analysis as a technique for reviewing systematically a substantial – and, they claim, unbiased – portion of the literature dealing with the effectiveness of psychotherapy. They analyzed all the data they could find comparing psychotherapy or behavior therapy and a control group. For each dependent variable included in each of the studies surveyed they computed a measure of effect size defined as the difference between the mean of the therapy group and the mean of the control group, divided by the standard deviation of the control group. They conclude that the mean effect size of psychological therapies is .85, indicating that when averaged over all measures in all studies, the outcome of psychological therapy is superior to that of nontreatment in control groups.

Smith et al.'s (1980) analyses appear to provide definitive evidence in favor of the effectiveness of psychological therapies. However, we felt that the research analyzed by Smith et al. should be subjected to further analyses. We had some reservations about the use of meta-analytic procedures for a body of literature as diverse as that summarized by Smith et al. (1980) (see Eysenck 1978; Strahan 1978). [See also Rosenthal & Rubin: "Interpersonal Expectancy Effects" *BBS* 1 (3) 1978.] While meta-analysis may be appropriate for summarizing the results of investigations using the same dependent variable with similar subject populations, it is questionable whether the method should be extended to the analysis of research using grossly different patient populations being subjected to grossly different methods of therapy where the outcomes are assessed using different dependent variables. Accordingly, we have tried to look in somewhat greater detail at a subset of the studies used by Smith et

 0140-525X/83/020275-36/$06.00 

al. (1980) and we have tried to supplement a meta-analysis by a more traditional examination of individual studies.

The procedures used by Smith et al. (1980) in their meta-analysis may not have been ideal. In particular, these researchers performed a meta-analysis using dependent variables as their unit of measure. This procedure, in effect, weights a study by the number of dependent variables included in the analysis. Given the degree of variability across studies, we feel that it is more appropriate to use the study itself as a unit of analysis. Accordingly, we present, separately, effect size measures for each dependent variable included within a study and we obtain a mean effect size for each one (see Landman & Dawes 1982 for a comparable reanalysis of a subset of the studies used by Smith et al. 1980).

In order to permit us to examine this body of literature in greater depth we have focused on the subset of studies reported by Smith et al. (1980) using psychotherapy rather than behavior therapy. Although psychotherapy and behavior therapy may no longer be as theoretically distinct as they once were, the techniques, patients, and methods of assessment of the outcomes of therapy used in the research literature for these two broad classes of therapeutic treatments are still somewhat different. Our decision to limit the scope of our analyses to research on psychotherapy was done in part for theoretical reasons and in part in order to permit us to examine a subset of studies in somewhat greater detail.

Finally, we restricted our analysis to those studies that included a placebo treatment. We believe that placebo treatments provide a more appropriate control group for assessing psychotherapeutic outcome than the more usual wait-list controls. Wait-list controls may lead to outcomes that are more negative than would have occurred merely through the passage of time. Individuals who seek therapeutic services and who are placed in a wait-list control group may be disappointed. In addition, such individuals may be experiencing an unintended reverse placebo effect. In being told they are being placed on a wait list, they are in effect told that they should not expect to improve since no therapeutic intervention will be provided for them. Since there is no appropriate control for a wait-list control group, there is no way of testing this notion. Whether wait-list controls are appropriate or not, it is relevant to inquire whether the benefits of psychotherapy exceed changes attributable to placebo expectations. Smith et al. (1980) deal with this issue only in passing. They indicate that the majority of placebo-controlled studies are in the behavior-therapy rather than the psychotherapy outcome literature. They assert that psychological treatments are approximately twice as effective as placebo treatments, and accordingly they expect that a comparison of psychotherapy against placebo would yield an effect size of approximately .42 standard deviations. We will focus on this comparison in our analysis.

We had several reasons for focusing on the comparison of psychotherapy treatments and placebo treatments. First, as a general principle, the comparison of treatment with some type of placebo control is a standard research design. The comparison is justified, since there is abundant evidence that individuals who believe they are receiving a treatment will improve as a result of the belief they are in treatment, even if there is no other theoretical reason for the treatment to be efficacious (see Shapiro & Morris 1978).

Second, we were aware of a study by Brill, Koegler, Epstein & Forgy (1964) which provided evidence that the psychotherapy effect was equivalent to the placebo effect. Brill et al. (1964) randomly assigned psychiatric outpatients to one of several groups: a psychotherapy group that received 20 sessions of psychoanalytically oriented psychotherapy administered by psychiatric residents; a wait-list control group; a pill-placebo group that received chemically inert pills combined with occasional brief visits to psychiatrists (primarily to check on their response to medication, which was administered in a double blind design); and groups that received psychoactive drugs. Several outcome measures were used to assess therapeutic effects including the MMPI (Minnesota Multiphasic Personality Inventory), therapist and patient ratings, independent reports by a social worker, and a rating from a relative, spouse, or friend. Brill et al. (1964) report that for all measures the patients who received treatment were improved relative to the patients who were assigned to the wait-list control group. However, there were no significant differences among the various forms of treatment, including the placebo treatment. Brill et al. (1964) examine the effectiveness of psychoanalytically oriented psychotherapy of somewhat longer duration than is characteristic of many outcome studies, have a sample size exceeding that which is typical in outcome research (30 patients in each of several conditions), and use real patients. They provide evidence for the proposition that the effects of psychotherapy are equivalent to the effects of a relatively minimal placebo, which is essentially equivalent to knowledge that one is in treatment.

There are several limitations to the Brill et al. (1964) study. There was a high dropout rate; although the range of dependent variables used to assess outcomes was moderately varied, there were no behavioral measures used; and the therapy was administered by relatively inexperienced therapists. We wanted to see whether the corpus of about 500 studies included in the Smith et al. (1980) reviews would yield data that contradicted or supported the results obtained by Brill et al. (1964).

Third, an analysis of the differences between psychotherapy and placebo treatments may have implications for the provision of psychological treatment. If, for example, psychotherapy is no more effective than pill placebos, it may be cheaper and simpler to provide patients with pill placebos administered by general practitioners rather than long and relatively expensive treatment by trained psychotherapists. Our example should not be construed as advocacy of any form of treatment. Rather, we use the example as an illustration of the proposition that the results of an analysis of outcome research comparing psychotherapy to placebo treatments may have consequences for the design of treatment programs.

Fourth, there are results in the literature which are at least suggestive of the possibility that the outcomes of psychotherapy may, in part, be attributable to the influence of placebos. These include the following:

A. Duration of treatment is unrelated to the magnitude of therapeutic effect, according to Smith et al. (1980). If quite brief treatment and extended treatment produce outcomes of similar magnitude, one is led to believe that the activities engaged in by the therapist are irrelevant to the outcomes of treatment.

B. There is a body of research which suggests that experience and training in psychotherapy are unrelated to the magnitude of the psychotherapeutic effect (see Durlak 1979). Presumably, professional training would lead a therapist to engage in therapeutically relevant activities in a more accomplished manner. If the competence acquired by the therapist is unrelated to therapeutic outcome then it is possible that the activities of the therapist are not the cause of the changes in the patient.

C. Strupp (1973) and Bergin and Lambert (1978), among others, have suggested that some of the variance in outcome of psychotherapy is attributable to the characteristics of the patient. Patients who are articulate and intelligent and inclined to interpret their problems as being of psychological origin are said to have a higher probability of favorable outcome than patients without these characteristics. If there are patient characteristics that can predict the outcomes of psychotherapy, it is at least possible that these effects occur autonomously. That is, certain patients have the ability to change either as a result of their personality characteristics or the environmental circumstances in which they find themselves, or both, and a sufficient condition for the change is the knowledge that one is receiving some form of therapy. Thus it is possible that the activities of the therapist are irrelevant to the actual changes that occur.

D. Strupp (1977) and Bergin and Lambert (1978) have suggested that in a small minority of patients psychotherapy may produce adverse outcomes. While the general effect of placebo treatment is beneficial, placebo effects have also been found on occasion to produce reverse effects. Duncan and Laird (1980), for example, have suggested that individuals who are self-attentive are more likely to experience reverse placebo effects. Such individuals may become aware of the fact that a placebo has not dramatically altered their condition even though they were told it would be beneficial, and as a result they infer that their problems are more severe than they had thought. Thus placebo effects could, in principle, account for a possible deterioration effect in a minority of patients as a result of psychotherapy.

We do not mean to imply that we accept any of these conclusions about the results of outcome research on psychotherapy. We merely wish to indicate that a number of the conclusions which responsible reviewers have drawn from their examination of outcome studies are at least compatible with the assertion that part or all of the psychotherapy effect is attributable to the effects of placebos.

We have used the term "placebo" without an explicit definition. We consider a treatment a placebo treatment if the patient is led to believe that the treatment is efficacious and the treatment does not contain any other therapeutic components. Such a treatment defines an ideal case of a placebo. An operational procedure that comes close to meeting this idealized definition is the provision of a chemically inert pill to a patient combined with the assertion by a professional therapist that the pill will be an effective treatment. There are several variables that can, at least on speculative grounds, influence the effectiveness of this type of treatment. The patient must believe the assertion made by the therapist, and if the therapist harbors doubt about the effectiveness of chemically inert medications (or, in the case of a double blind study, of psychoactive drugs) for the treatment of psychological problems, the therapist may in subtle ways communicate these doubts to the patient and the effectiveness of the placebo may be mitigated. Moreover, the beliefs of the patient with respect to the potential efficacy of pill medications for the treatment of psychological problems may influence the effectiveness of such placebo treatments. Finally, in an actual experimental situation pill placebos may be accompanied by several other quasitherapeutic elements. For example, in the Brill et al. (1964) study, patients who received drug treatments including placebo were seen by resident psychiatrists for 15 minutes or less weekly, biweekly, or monthly. The sessions were brief in order to decrease the probability of psychotherapeutically relevant exchanges, and the residents were instructed to focus on the drug reactions of the patients. Despite these strictures, there is no way of knowing the extent to which these conditions were adhered to in a large-scale study with several therapists. It is conceivable, as Brill and his colleagues note, that some brief psychotherapeutically relevant interchanges may have occurred on occasion.

The problem of the comparison of psychotherapy with placebo treatments is complicated by the fact that many placebo treatments include a variety of elements in addition to an attempt to manipulate the belief that one is receiving an efficacious treatment. Some studies have used placebo treatments that include discussion groups in which a therapist explicitly attempts to steer the groups' conversation toward topics that are assumed to be irrelevant to the psychological problems which led the patients to be selected for psychotherapy. Such a placebo treatment attempts to control for such features of psychotherapy as duration of treatment, meeting with fellow-patients, and having an opportunity to engage in conversation in a quasitherapeutic setting. Presumably the treatment is construed as a placebo in the belief that discussions that do not focus on specific problems are not therapeutically efficacious. But this belief may be no more than an act of faith. It may well be that the essential features of psychotherapy which account for its therapeutic effectiveness are well reproduced by this type of placebo treatment. Grünbaum (1981) has stressed the importance of theoretical assertions as a basis for ascertaining the placebogenic status of a particular form of treatment. However, since our understanding of the nature of the processes that induce change is imperfect (if not downright lacking), our attempts to distinguish between placebo treatments and treatments that reproduce several features of psychotherapy may involve imprecise theoretical and empirical distinctions. Although we prefer minimal placebo treatments we are restricted to the available research literature. We will accordingly consider as a placebo treatment any procedure so described by an investigator where we are informed or able to infer that the possibility of therapeutic benefits is conveyed to the patient. In addition, we require either implicit or explicit evidence that the authors have a theoretical rationale for the assertion that the placebo treatment omits features of psychotherapy which are essential for therapeutic effectiveness.

While the reader may feel that the variability of placebo treatments renders a comparison between placebo and psychotherapeutic treatments vexed, it should be noted that there may be considerable variability among wait-list control treatments. Duration of wait-list assignment may be a critical variable. The extent to which the assignment to a wait-list control is accompanied by preliminary assessment procedures, the kinds of information conveyed to the patient about the potential harm involved in the wait period, and the availability of psychotherapeutic services during the wait period in the event of an emergency may have a considerable influence on the changes that may occur among patients during their time on the wait list. For example, Sloan, Staples, Cristol, Yorkston, and Whipple (1975) explain their failure to find therapeutic improvement relative to their wait-list control group on follow-up as follows: The psychological assessment that preceded assignment to the wait list group, combined with the information conveyed by the researchers that wait-listed patients could receive therapeutic treatment in an emergency were sufficient to create therapeutic benefits in the wait-list control subjects that eventually matched the alleged benefits of psychotherapy and behavior therapy. Thus wait-list control conditions may vary considerably, and the comparison of such controls with psychotherapeutic treatments may involve the comparison of two forms of treatment, each variable across different investigations. Thus the comparison between psychotherapy and wait-list control treatments may raise as many theoretical problems as the comparison between psychotherapy and placebo treatments.

## Method

We were able to locate and read in either original or abstract form 513 of the 520 studies included in the psychotherapy and drug meta-analyses included in Smith et al. (1980). Of the 513, we selected for analysis only the small subset that included both a psychotherapy treatment and a placebo treatment. The distinction between psychotherapy and behavior therapy was not problematic for the great majority of studies analyzed (for a more comprehensive treatment of this issue see Murdock 1982; Prioleau 1982). We classified as psychotherapy a number of the studies using such techniques as rational emotive therapy and even social learning therapy described in rather traditional behavioristic language, whenever we were able to infer from the description of the therapist's activities that he was required to engage in a process of exploration and clarification of the emotional experiences of the patient. In addition to classifying a study as containing a psychotherapy treatment, we looked for evidence that there was an attempt to foster and develop an emotional relationship between the therapist and the patient.

We judged that 40 of the 513 studies contained both a psychotherapy and a placebo treatment. From these 40 we discarded 8, either on the grounds that they were so seriously flawed as to render any comparison unjustified or because even with a number of ad hoc assumptions it was not possible to compute measures of effect size.

We computed a measure of effect size for each of the dependent variables included in the remaining 32 studies. Effect size was defined as the difference between the psychotherapy group mean and the placebo group mean divided by the pooled standard deviation for psychotherapy and placebo groups. If pretreatment scores existed for a measure, we obtained the difference between the pre- and posttreatment score, subtracted the comparable difference score for placebo treatment groups, and divided the difference between the change scores by the pooled standard deviation of the change scores. Where such standard deviations were not available or could not be calculated or inferred from the statistics presented, we assumed that the correlation between pre- and posttreatment scores was .5, and we adjusted the posttreatment standard deviation to obtain an estimate of the standard deviation of the change scores. If a variable was presented in terms of nominal scale measurement (e.g., percentage of improvement), we dummy coded the variables, calculated $r$, and converted $r$ into a measure expressed in terms of standard deviation units. If standard deviations were not reported, we attempted to derive them from available statistics. Where the author indicated that there were no differences between therapy and placebo treatments and data were not presented, we assumed that the effect size was zero. (For a general discussion of obtaining measures of effect size, see Cohen 1969; and for a discussion of inferring effect size measures from limited data presentations, see Smith et al. 1980.) We adjusted our effect size indices such that positive scores indicated that the psychotherapy group did better on a particular measure than the placebo group. We obtained a mean effect size for each study by taking the mean of the separate effect sizes.

We have departed from the procedures used by Smith et al. (1980) in the calculation of effect sizes in three respects. First, the numerator of the fraction we use to define effect size is defined as the difference between psychotherapy and placebo treatments rather than as the difference between psychotherapy and a control group. Second, Smith et al. (1980) use as a denominator of the fraction the standard deviation of the control group, whereas we use a pooled standard deviation. The use of the standard deviation of the control group yields the advantage of permitting one to define mean differences between two or more therapy treatments in a single study against a common base line. We chose to use a pooled standard deviation because we found that for a majority of studies included here, separate standard deviations were not available and accordingly we were forced to estimate standard deviations from such statistics as $t$ and $F$. Such a procedure permits one to obtain only a pooled standard deviation rather than a separate estimate for the control and treatment group. Since we were forced to use pooled estimates for some of our calculations, it seemed to us to be more consistent to use pooled estimates for all our calculations. Smith et al. (1980) report that they found no difference in variability of control group and therapy outcomes. If a comparable result holds for the studies we examined, the decision to use pooled estimates should not appreciably influence our estimates of effect size, although it might influence an estimate in a particular set of data. Third, for the subset of studies in which data on the standard deviation of change scores were not provided, we arbitrarily assumed that the correlation between pretest and posttest scores was .5. Smith et al.

(1980) used a variable value to estimate corrections depending on the nature of the outcome variables used and the duration of time intervening between pretest and posttest. We felt that for many of the measures used we were not in a position to make an informed guess about the value of the test–retest correlation. The value of .5 was at the upper end of the range of correlations used by Smith et al. (1980). It seemed to us to be somewhat less arbitrary to use a standard value rather than a variable one. In any case, this arbitrary correction was used in fewer than one-third of our calculations and is probably not a major source of influence in the magnitude of effect sizes.

## Results

Table 1 presents a description of the 32 studies included in our analysis and includes the calculations of mean effect sizes for each. The mean values are corrected for the sampling bias of effect sizes (see Hedges 1981). The corrected distribution of mean effect sizes is skewed to the right. The modal category of a grouped frequency distribution with an interval of .20 occurs at an effect size value of .00. The median effect size is .15 and the mean is .42. The distribution includes one extreme value. This study, in which virtually all of the children given psychotherapy and none of the children in the placebo treatment are reported as having improved, has an effect size 1.54 standard deviations higher than any other.

It is apparent that there is considerable diversity in magnitude of outcome in these studies. We attempted to define characteristics that might be related to the measure of effect size. We analyzed the following variables: (1) duration of treatment, (2) sample size, (3) the use of real patients as opposed to subjects solicited by the investigators, and (4) the nature of the outcome measure used.

Duration of therapy was correlated −.24 (n.s.) with effect size. Thus the nonsignificant trend is for the benefits of therapy relative to placebo to decrease as the duration of treatment increases. The correlation between sample size and effect size was −.21 (n.s.). Sample size was inversely related to therapeutic effects. The mean effect size for the eight studies using patients was .35 (median = .08) and the mean effect size for the studies using solicited subjects was .44 (median = .16). A $t$ test comparing these means yields a value of .36 (n.s.).

We assigned each dependent variable to one of six categories (see Table 1). We obtained a mean effect size for the category of undisguised self-report measures and a mean effect size measure for all other measures used in each study. We compared these means in a subset of 19 studies, which permitted us to obtain within the same study a mean effect size for one or more outcomes of undisguised self-reports and one or more measures of any other type. The mean of the means for undisguised self-reports was .43 and the mean of the means for all other measures was .17. A matched group $t$ test had a value of 1.58 (n.s.).

Six of the studies included in Table 1 reported follow-up data. We calculated that there were minimal changes during follow-up in mean effect sizes for three of these studies: Schwartz and Dubitzky (1967), Gillan and Rachman (1974), and DiLoreto (1971). Two showed declines of .23 and .61 in mean effect size (Paul 1964 and Hedquist & Weinhold 1970, respectively), and one (Jarmon 1972) showed a gain in effect size of .22. These data indicate that for this small subset of studies there is no tendency for the benefits of psychotherapy relative to placebo to increase during a follow-up period.

## Discussion

Our estimate of a mean effect size of .42 is exactly in agreement with the magnitude of the psychotherapy effect size relative to placebo treatments estimated by Smith et al. (1980). Quite apart from the central tendency of effect size in these studies, the trends in the characteristics of studies that are related to measures of effect size, albeit weakly, do not support an intuitive model, which suggests that the effects of therapy are more powerful than the effects of placebo treatments. Assume that the effects of psychotherapy are strong and that the effects of placebo are weak and emphemeral. One might argue on intuitive grounds that placebo effects would decline more than therapy through time and hence would be less likely to be equivalent to therapy where outcomes are assessed after long durations of therapy and at follow-up investigation. Moreover, one might expect ephemeral and perhaps misguided beliefs in the benefits of placebo to be most strongly present for undisguised self-report measures. In addition, one would expect real patients to be more likely to benefit from the effects of a powerful treatment than solicited subjects, on the assumption that the former group are more disturbed than the latter. Our data are consistent in their bearing on this set of crude intuitions – in all respects our findings contradict these expectations. Our data suggest that as we examine these studies in a more critical manner and examine their implications for the benefits of psychotherapy, we are led to assume that the benefits of therapy relative to placebo treatments become vanishingly small.

It is possible to supplement the quantitative analysis of these studies by a somewhat more traditional analysis of their descriptive properties. And given the diversity of measures, therapies, and subjects included in these investigations, there is some question whether attempts to relate quantitative indices of effect size to other variables is entirely legitimate. After examining this set of studies we have concluded that the most informative classification is derived from an analysis of the types of patients included. It is possible to organize these studies into four subclasses as defined by the type of patient receiving therapy. The first sub-class, which we consider the most crucial for an evaluation of psychotherapy outcome research, is defined as studies of outpatients, in which the patient population seeks psychological services and is not institutionalized. The three studies included in this group all deal with patients who may be described as neurotics. In addition to the Brill et al. (1964) study, which we have already described in the introduction and which we calculate to have an effect size of .07, Lorr, McNair, and Weinstein (1963) reported a similar investigation using a pill placebo treatment for psychiatric outpatients. The duration of treatment was brief (four sessions), and only two measures were used to assess outcome (global

Table 1. *Description of design, placebo, dependent measures, and effect sizes of studies in the analysis*

| Author and year | Type of therapy | Subjects | N[a] | Contact hours | Therapists | Placebo description | Dependent measures[b] | Estimated effect sizes[c] | Remarks |
|---|---|---|---|---|---|---|---|---|---|
| Winkler et al. (1965) | Rogerian | Underachieving elementary school students | 60 | 7 | M.A. level | Listened to records and stories | (2) CA test of personality<br>(6) GPA | −.55<br>.11<br>−.22 | [d] |
| Bruyere (1975) | Group client-centered | Disruptive junior high school students | 48 | 16 | School counselors | Group problem solving | (1) Self-concept<br>(5) Conduct GPA<br>Behavior scale<br>Ratings of disruptive classroom behavior | −.64<br>−.09<br>.00<br>+<br>−.18 | Data for effect size computation n/a for last measure |
| Schwartz and Dubitzky (1967) | Group therapy | Moderate smokers | 72 | 12 | Psychologists | Pill placebo | (6) Cessation of smoking | −.13 | |
| Orlov (1972) | Group Rogerian | Maladjusted middle school students | 40 | 15 | Psychologists | Reading and discussion Books | (5) Sociometric test<br>Rating scale<br>Classroom behavior<br>Improvement rating by teacher<br>(6) GPA | .00<br>.00<br>00<br>−.30<br>−.21<br>−.10 | |
| Bruce (1971) | Group client-centered | Vocational rehabilitation clients | 20 | 12 | Psychology graduate student | Social group functions | (1) Index of adjustment<br>(5) Job performance evaluation | .89<br>−1.06<br>− .09 | |
| Gillan and Rachman (1974) | Group insight rational therapy | Multiphobic outpatients | 24 | 15 | Psychiatrists Psychologist | Muscle relaxation training, phobic hierarchy no relaxation training | (1) Anxiety scale<br>(2) EPI<br>(4) Therapist rating phobia & depression<br>(5) Rating phobia & depression, external rater<br>(6) Behavioral avoidance test<br>Skin conductance | −.46<br>00<br>.19<br>−.24<br>.00<br>.00<br>−.09(−.07) | [c,f] |
| Matthews (1972) | Group reality therapy | Maladjusted elementary school children | 221 | 15 | Elementary school teachers | Language arts classes | (2) CPI<br>(5) Problem behavior rating<br>(6) Reading | −.11<br>.18<br>−.18<br>−.04 | |
| Jarmon (1972) | Group rational emotive therapy | Speech-anxious college students | 41 | 3 | Psychology graduate students | Group discussion of neutral topics; reading RET book | (1) Fear survey schedule<br>Confidence as a speaker<br>Social anxiety<br>Fear of evaluation<br>Irrational ideas<br>Fear rating<br>(5) Speech anxiety rating<br>Speech disruptions | −.31<br>−.16<br>.34<br>.07<br>−.18<br>.25<br>−.24<br>−.23<br>−.01 | [e] |
| Shapiro and Knapp (1971) | Group ego therapy | High-anxious college students | 42 | 14 | No information | Group discussion of general topics | (1) Personal integration<br>Response bias<br>Anxiety level | .00<br>.00<br>.00<br>.00 | |
| Coche and Douglas (1977) | Group problem-solving training | Adult psychiatric inpatients | 46 | 8 | No information | Group play reading | (1) Adjective checklist<br>(2) Minimult<br>Means-ends problem-solving | .26<br>.01<br>−.11<br>.05 | |
| Desrats (1975) | Group Rogerian | Institutionalized adolescent orphans | 39 | 25 | Counselors with "limited counseling experience" | Viewed films and study lessons | (1) Self-esteem<br>(2) CPI<br>(5) Behavior rating<br>(6) GPA<br>Adjustment tally | .00<br>00<br>.00<br>.28<br>.00<br>.06 | [f] |
| Brill et al. (1964) | Psychoanalytic | Adult outpatients | 60 | 20+ | Psychiatric residents | Pill placebo | (1) Patient rating<br>(2) MMPI<br>(4) Therapist rating<br>Global evaluation | .05<br>−.12<br>.10<br>.00 | [f]<br>S.D. estimate ⅕ of range |

Table 1. *(cont.)*

| Author and year | Type of therapy | Subjects | N[a] | Contact hours | Therapists | Placebo description | Dependent measures[b] | Estimated effect sizes[c] | Remarks |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | (5) Relative rating<br>Social worker rating | −.12<br>.38<br>.07 | |
| Herman (1972) | Humanistic counseling (group and individual) | High-anxious junior high school students with reading problems | 40 | 4.5 | Junior high school counselors | "Bull" session | (1) MAS<br>(5) Anxiety rating<br>(6) Reading test | .74<br>−.08<br>− 42<br>.08 | d |
| Lorr et al. (1963) | Individual therapy | Adult outpatients | 50 | 4 | Psychiatrists, psychologists, and social workers | Pill placebo | (1) Global improvement<br>(3) Global improvement | 19<br>.00<br>.10 | |
| Rosentover (1974) | Group counseling | Underachieving high school students | 63 | 7 | Education graduate student | Heard speakers, viewed films, limited discussion | (2) Minnesota Counseling Inventory<br>(6) GPA | .00<br>.23<br>.12 | |
| Warner (1969) | Group verbal and model reinforcement | Alienated junior high school students | 102 | 4.5 | Three high school counselors | Group discussion | (1) Alienation<br>Anxiety<br>Self-concept<br>(6) GPA | .56<br>−.06<br>.11<br>−.15<br>.12 | |
| Paul (1964) | Individual insight therapy | College students with speech performance anxiety | 30 | 5 | Psychologists | Pill placebo and boring task | (1) Anxiety<br>(3) Global rating<br>(5) Anxiety<br>(6) Pulse rate<br>Palmar sweat | .22<br>.40<br>.69<br>−.03<br>−.37<br>.18( 17) | |
| West (1969) | Group client-centered | Disruptive elementary school children with learning difficulty | 16 | 10 | Psychology graduate students | Read, played with puzzles, or sat quietly under supervision of counselor. No verbal interchange | (1) Self-esteem<br>(2) Draw-a-Person Apperception Test<br>(5) Sociometric<br>(6) WISC | .29<br>− 01<br>−.27<br>.69<br>16<br>.17(.16) | |
| Rehm and Marston (1968) | Nonspecific group | High-anxious college students | 16 | 2.5 | Psychology graduate students | General group discussion | (1) Fear of opposite sex<br>Situation test<br>Fear survey<br>Anxiety scale<br>Adjective list<br>(5) Situation anxiety | 1.13<br>.50<br>−.07<br>−.54<br>.16<br>.07<br>.21(.19) | |
| Paykel et al. (1975) | Individual supportive | Clinically depressed inpatients | 34 | 36 | Psychiatric social workers | Pill placebo | (1) Psychic and somatic complaints<br>(4) Psychiatric rating<br>Interview for depression<br>Depression scale<br>(3) Psychiatric evaluation<br>(3) Relapse rate<br>(5) Social adjustment | .00<br>.00<br>.00<br>.00<br>.79<br>.32<br>.78<br>.27(.26) | f |
| Alper and Kranzler (1970) | Individual client-centered therapy | Disruptive elementary school children | 18 | 10 | Graduate students in counseling | Read and discuss stories | (2) Self social symbols<br>(5) "Out of seat" behavior<br>Sociometric test | −.10<br>1.26<br>−.12<br>+.34(.32) | |
| Trexler and Karst (1972) | Group rational emotive therapy | Speech-anxious college students | 22 | 4 | Psychology graduate student | Relaxation training | (1) Irrational beliefs test<br>Anxiety scale<br>Confidence as a speaker<br>(5) Behavior checklist<br>Overall estimate of anxiety<br>(6) Finger sweat print | 1.37<br>−1.03<br>1.10<br>41<br>.25<br>.26<br>.39(.36) | |

*(continued)*

Table 1. *(cont.)*

| Author and year | Type of therapy | Subjects | N[a] | Contact hours | Therapists | Placebo description | Dependent measures[b] | Estimated effect sizes[c] | Remarks |
|---|---|---|---|---|---|---|---|---|---|
| Lester (1973) | Individual relationship counseling | High-anxious junior high school students | 13 | 3 | Graduate student in counseling | Guided discussion of current events | (1) Test anxiety<br>Global self-report | .00<br>1.18<br>.59(.55) | |
| Coche and Flick (1975) | Group problem-solving training | Psychiatric inpatients | 64 | 8 | No information | Group play reading with discussion | (2) Means-ends problem-solving procedure | .69<br>.69 | f |
| Hogan and Kirchner (1968) | Individual eclectic therapy | Snake-phobic college students | 20 | .75 | No information | Bibliotherapy | (6) Ability to lift snake | .74<br>.74(.71) | |
| DiLoreto (1971) | Group client-centered and rational emotive therapy | High-anxious college students | 60 | 9 | Psychology graduate students | Group discussion over lunch focused on academic topics | (1) Interpersonal anxiety<br>S-R inventory of anxiousness<br>Trait anxiety<br>Social desirability<br>(5) Checklist of interpersonal anxiety | .37<br>.86<br>1.17<br>.50<br>1.25<br>.82 | d |
| Grande (1975) | Group rational emotive therapy | High-anxious college students | 34 | 4.5 | Three graduate students and one under-graduate | Relaxation training via tape | (1) Interpersonal anxiety<br>MAS<br>Fear survey<br>(5) Interpersonal anxiety<br>Behavior rating | 1.32<br>1.21<br>1.09<br>.50<br>.74<br>.97(.95) | |
| Meichenbaum et al. (1972) | Individual and group rational emotive therapy | High-anxious college students | 24 | 8 | Psychologists | Group discussion | (1) Checklist for anxiety<br>Anxiety differential<br>(5) Performance anxiety<br>(6) Duration of silences<br>Number of "ah" statements | 1.34<br>.42<br>.57<br>2.00<br>1.02<br>.98(.95) | d |
| Hedquist and Weinhold (1970) | Group social learning | Anxious college students | 30 | 6 | Psychology graduate students | Group discussion | (1) Assertive behaviors | 1.22<br>1.22(1.19) | |
| House (1970) | Group non-directive play therapy | Unpopular elementary school children | 24 | 10 | Education graduate student | Reading group | (1) Self-concept and motivation inventory<br>(5) Sociometric test | 2.50<br>.70<br>1.60(1.55) | |
| Roessler et al. (1977) | Group counseling | Physical rehabilitation clients | 43 | 10 | Rehabilitation counselor | Personal hygiene training | (1) Self-scale<br>Facility outcome measure<br>Goal attainment scale | .80<br>.84<br>3.40<br>1.68(1.65) | f |
| Platt (1970) | Group Adlerian | Disruptive elementary school children | 24 | 5 | No information | Listened to records and studied | (5) Rating by parents<br>Rating by teachers | 3.80<br>2.80<br>3.30(3.19) | Ratings were not "blind" |

[a]Only the number of subjects in the psychotherapy and placebo treatments are included. [b]The number in parentheses refers to type of dependent measure according to the following code: 1 = undisguised self-report; 2 = disguised self-report; 3 = global report by therapist; 4 = rating of independent behavior by therapist; 5 = rating of independent behavior by others; 6 = independent behavior. [c]Numbers in parentheses indicate mean effect size corrected for sampling bias. [d]Study used two types of psychotherapy and one placebo group. Information on number of subjects and effect size is pooled across both comparisons. [e]Study used one psychotherapy group and two types of placebo. Information on number of subjects and effect size is pooled across both comparisons. [f]Study used real patients.

ratings by therapists and patients). However, unlike in the Brill et al. (1964) study, the therapists were described as being experienced. We calculate the effect size of Lorr et al. (1963) to be .10. The last study in this group, Gillan and Rachman (1974), provides a significant amount of therapy to a small number of patients, all multiphobic. Psychotherapy was administered by experienced therapists who are described as believing that psychotherapy is the treatment of choice for this condition. Two placebo treatments were used: relaxation training without an attempt to pair the relaxation response with the phobic stimulus and a placebo condition in which the phobic hierarchy was presented without relaxation training. Several outcome measures were used, including behavioral and psychophysiological measures. We calculate the effect size of this study to be negative, −.07.

An additional study reported by McLean and Hakstian (1979) was not included in our formal analysis since it was published too late to appear in the Smith et al. (1980) analysis. However, it is relatively well designed and does buttress the conclusions suggested by the three studies we review in the category of psychiatric outpatients. McLean and Hakstian report data for 37 randomly assigned clinically depressed outpatients who received 8–12 hours of insight-oriented psychotherapy from licensed psychologists and psychiatrists; one group of therapists had at least 2–4 years of experience and the other 5 years or more. Psychotherapy was the treatment of choice for these professionals. The experience of the therapist was unrelated to outcome and accordingly this variable was dropped from the analysis. Outcome was assessed by the use of 10 self-report measures derived from analyses of questionnaire data. A number of outcome measures were adjusted for relevant covariates. The placebo treatment was administered to 38 randomly assigned patients who received 10 hours of muscle relaxation therapy. Subjects were informed that the muscle relaxation treatment was therapeutically relevant, although McLean and Hakstian assert that they consider the variable as a placebo since there is no compelling rationale for the view that treatment of depression by muscle relaxation is therapeutically efficacious. At the end of therapy, in a comparison of the patients assigned to the psychotherapy and the placebo treatment group there was no appreciable difference on any outcome measure, although a group of subjects randomly assigned to a behavior therapy treatment condition were discernibly improved relative to the relaxation therapy patients. The relaxation therapy patients had slightly better outcomes than the psychotherapy patients. We calculate the effect size to be negative (−.11). Three months after the termination of psychotherapy there was again no discernible difference in outcome. The effect size was negative (−.08).

In a number of ways, McLean and Hakstian's is a well-designed study. It uses well-trained psychotherapists with a commitment to the virtues of psychotherapy; it includes a well-defined patient group and a follow-up. It has perhaps two limitations. There is exclusive reliance on questionnaire and self-report data (although arguably the instruments used are well standardized and have some validity for the particular target population). And it is conceivable that the placebo treatment might include a number of complex aspects (e.g., following instructions, etc.) that extend beyond the pure case of an expectancy manipulation. However, if considered in conjunction with the results of the three other studies of outpatients, this study appears to buttress the view that psychotherapy does not lead to outcomes that are more favorable than those attained by placebo treatment for outpatients.

A second subset of studies deals with institutionalized patients. There are four in this group. Desrats (1975) deals with institutionalized adolescents given counseling by relatively untrained counselors and has an effect size of .06. Paykel, DiMascio, Haskell, and Prusoff (1975) deal with clinically depressed inpatients and use a pill placebo treatment. We calculate an effect size of .26 for this study. In this group, only Coche and Flick (1975) report a substantial positive effect size. It is a study of outcome among psychiatric inpatients who are given group problem-solving training. For a single measure of group problem solving, which may or may not be contaminated by the therapeutic procedures followed, we calculate an effect size of .69. However, in a study designed in part to be a replication of Coche and Flick's (1975) work, Coche and Douglas (1977) report a failure to replicate their earlier findings. For a somewhat more extended set of outcome measures used in this study, we calculate an effect size of .05. Thus the results of the only study using psychiatric inpatients with a substantial positive effect size are nonreplicable.

The third class of studies deals with students in school. One of these, Bruyere (1975), deals with students attending a special school for disruptive junior high school pupils. Bruyere's study has a negative effect size. School counseling did not lead to positive outcomes on a variety of measures. There are 21 additional studies in this group, all dealing with therapy provided to solicited subjects: groups of students who are nominated or selected for therapy because they have an extreme score on some measure. Clearly, such subjects are not representative of the patients who seek therapeutic services. Six of these studies have relatively positive effect sizes ranging from .74 to 3.19. This subset, with relatively large positive effect sizes, may be characterized collectively (with exceptions that can be noted by an examination of Table 1) as studies in which relatively brief therapy is provided to relatively small groups of subjects. Three of the six involve rational emotive therapy and one involves group social learning therapy. Therefore, these studies deal with quasibehavioristic forms of treatment.

Our last group of studies is somewhat heterogeneous, consisting of those that do not fit into the preceding three groups. One study in this group, Schwartz and Dubitzky (1967), which has a negative effect size, deals with group therapy as a treatment for smokers who wish to stop smoking. Another with a negative effect size involves vocational rehabilitation clients. The final study in this group, Roessler, Cook, and Lillard (1977), is of the effectiveness of group counseling for physical rehabilitation clients. The effect size is 1.65 and it is the study we calculate to have the largest effect size among those that do not deal with solicited subjects. The outcome measures consist solely of undisguised self-reports, and the authors note that they do not know whether the optimism of their clients will be reflected in their actual behavior.

## Concluding Speculations

One can distinguish between two propositions: (1) In the research literature surveyed, psychotherapy has been found to produce changes in real patients that are equivalent to those produced by placebo treatments. (2) In general, psychotherapy produces changes in patients that are equivalent to those produced by placebo treatments. Proposition 1 is limited in its range of application to the studies we review and does not attempt to imply anything about what is true of psychotherapy in general. Yet, clearly, the goal of outcome research in psychotherapy is to discover what is true about psychotherapy and not to discover what is true about a particular body of research, which may be flawed. Accordingly, in what follows we shall speculate about the possible truth of proposition 2. Any psychotherapy outcome study has as its variables, therapists, patients, types of therapy, and measures of outcomes. Let us consider the potential limitations of the studies we have reviewed from the perspective of the several variables that jointly define an outcome study.

Some therapists may be able to produce beneficial effects through treatment. For a variety of reasons we are not sanguine about the possibility that variations among therapists is a major or important source of variance in outcome research. If the true size of the psychotherapy effect relative to placebo treatment is .0 then any allocation of main effect variance to individual differences among therapists would imply that some therapists consistently make their patients worse. The existence of individual differences among therapists as a major source of variance in outcome for diverse patients is of little practical relevance unless some way can be found of communicating to potential patients information about the competence of therapists. Since competence to produce changes among patients appears to be unrelated to professional training, it is hard to see how it could be practically determined. In addition, it is likely that some (and perhaps a major part) of the variance in outcome is not associated with a consistent effect of therapists but is best described as interaction variance in which certain therapists may have a higher likelihood of success with some types of patients. Finally, it should be noted that the hypothesis of individual differences in therapists' ability consistently to produce beneficial changes in their patients is testable in any study in which outcomes are obtained for several therapists each treating several patients. With such a design one could obtain a measure of the consistency of therapeutic outcome as a function of individual differences among therapists.

Certain types of therapy may have effect sizes that exceed those of placebo treatments. Clearly, the 32 studies we reviewed do not contain an exhaustive range of therapies. For example, rational emotive therapy appears to produce positive effect sizes in some of the work we review. However, none of the data we examined involved this type of therapy with true patients: nor was family therapy examined. It could be that some forms of therapy are more effective than placebo treatments and, similarly, there may exist certain types of patients who consistently benefit from psychotherapy. Obviously, we would be rash to deny such a possibility. However, it should be noted that such reviewers as Bergin and Lambert (1978) and Smith et al. (1980) have suggested that differences in outcome among types of therapy are minimal.

The weakest aspect of outcome research involves the measures used to assess change. It is possible that psychotherapy produces beneficial changes which exceed those attributable to minimal placebo treatments, but that the available outcome measures are too crude to detect such differences. Perhaps the major respect in which outcome research could be improved would be through the use of a wider range of dependent variables with more power to detect (possibly subtle) differences. None of the studies we examined used individually tailored measures of outcome in which one specifies at the start of treatment the kinds of changes that would indicate therapeutic progress for each patient. While it is certainly true that researchers could use a more extended and imaginative set of outcome measures than are characteristically used, there is no guarantee that such measures would increase the likelihood of demonstrating positive effects of psychotherapy. Recall that in the studies we examined the largest effects of psychotherapy relative to placebo were obtained for measures we would consider as biased and theoretically primitive: undisguised self-reports. One possible explanation of this result is that such measures are construed in part as a representation of the believability of the placebogenic component of therapy. If some patients prefer psychotherapy to, say, relaxation training as a method of treatment, then they may come to believe psychotherapy to be the more powerful treatment; they would accordingly report greater benefits from psychotherapy than from relaxation treatment. Undisguised self-report measures suggesting that psychotherapy is more beneficial than placebo may be attributable to psychotherapy's being a more believable placebo treatment than most actual placebo treatments. It would be useful to include a measure of the expectations for therapeutic improvement at the start of therapy for subjects assigned to various placebo and therapy treatments. Such a measure might relate to outcome measures.

Apart from the occasional use of psychophysiological measures, the studies we have reviewed have rarely relied on laboratory procedures to define outcomes. Although laboratory measures might yield positive results, there could be questions as to their ecological validity and their relation to significant actions and judgments in the everyday life of the patient.

We recognize that our speculations about the potential for demonstrating significant effects of psychotherapy relative to placebos are just that – speculations. On the basis of the available data we see no reason to believe that subsequent research using better research procedures and investigating other types of therapy administered to other types of patients will yield clear-cut indications that psychotherapy is more beneficial than placebo treatment. Thirty years after Eysenck (1952) first raised the issue of the effectiveness of psychotherapy, twenty-eight years after Meehl (1955) called for the use of placebo controls in psychotherapy, eighteen years after Brill et al. (1964) demonstrated in a reasonably well-done study that the psychotherapy effect may be equivalent to the placebo effect, and after about 500 outcome studies have been reviewed – we are still not aware of a single convincing demonstration that the benefits of psychotherapy exceed those of placebos for real patients. Such a study would have to show that psychotherapy administered to real patients yields improvements relative to placebo on a variety of measures and that these improvements endure over time. We believe that securing such data should be viewed as an urgent task by those who practice or advocate the use of psychotherapy. Given the absence of convincing contradictory data, and considering the partial support (at least) that the available research literature provides, we regard it as likely that the benefits of psychotherapy do not exceed those of placebo in real patients. That is, our conclusion may not only be valid when its range of application is restricted to a limited set of studies but may also be true of psychotherapy in general.

# Open Peer Commentary

*Commentaries submitted by the qualified professional readership of this journal will be considered for publication in a later issue as Continuing Commentary on this article. Integrative overviews and syntheses are especially encouraged.*

## Psychotherapy outcome: A wider view leads to different conclusions

Gavin Andrews
*School of Psychiatry, University of New South Wales, Sydney, Australia 2036*

Whether the psychological therapies benefit patients with neuroses is an important health care question because neurosis is a principal cause of invalidism. Whether the psychological therapies benefit troublesome schoolchildren, uncooperative felons, well people seeking self-actualization, or even schizophrenic and depressive psychotics are very much subsidiary problems. Last, whether such therapies benefit well college students – the naive sophomore subjects beloved of the dissertation industry – is of no practical and very little theoretical value in health-care decision making.

Smith, Glass, and Miller (1980) conducted a meta-analysis of some 500 studies of the psychological therapies applied to a broad spectrum of individuals and claimed to find positive benefits. There have been three reexaminations of different aspects of their data. In the first reexamination (Andrews & Harvey 1981a) we selected from their actual data set only those 81 studies coded as "neurotics who had sought or been referred for treatment," discarding all nonpatient studies. The mean study effect size (ES) superiority of treated over control groups was 0.75. The behavioral psychotherapies (mean ES = 0.98) were significantly better than the verbal dynamic psychotherapies (ES = 0.74) and these in turn were significantly better than counseling (ES = 0.37) or the four trials of placebo treatments (ES = 0.55). The benefits of dynamic psychotherapy and counseling, if grouped together in the fashion of Prioleau, Murdock, and Brody's paper, would not have proved significantly superior to placebo, however. The benefits of both behavioral psychotherapy and verbal psychotherapy seemed as evident in inpatients as outpatients and were stable during the first year after treatment. We noted that the 81 research studies were a very poor mirror of practice, and that more appropriate research was needed; research directed to the question of which treatment is indicated for a particular disorder rather than the general question of whether a particular therapy works for all patients.

In the second reexamination, Landman and Dawes (1982) rescored a one-in-seven sample of the original articles and then selected those in which assignment to control group had been random. The mean ES superiority over control group for all types of psychotherapy was 0.78, with treatment showing a 0.38 superiority over placebo in the four treatment placebo comparisons that were included. These results are congruent with our reanalysis.

The present target article is the third reanalysis. The authors rescored 32 studies of dynamic psychotherapy and counseling (but not of behavior therapy) that provided a treatment-placebo comparison and named these treatments "psychotherapy." They report that the mean ES superiority of the treated group over the placebo group is 0.42, median 0.15. The mean ES for the eight studies using patients was 0.35 (median 0.09) but, as we had noted earlier, few of the treatments or patient groups are representative of practice. All this is relatively straightforward until Prioleau et al. generalize from this small group of eight studies about the comparative value of "psychotherapy" and placebo treatments in the health sciences. Kazrin, Durac, and Agteros (1979) speculated that in naive hands meta-analysis could lead to ridiculous results. I think their fear has been realized, for the eight studies cited just do not represent dynamic psychotherapy and thus such a generalization is unwarranted. I think the trouble is methodological formalism. I know this is heresy, but even though Gordon Paul's (1964) study of speech anxiety in college students is elegant, neither the design nor the result is relevant to the delivery of health care.

The issue of whether a placebo-controlled trial is the only valid method of assessing psychotherapy outcome needs to be considered. Placebo treatments that look like dynamic psychotherapy and yet confer no benefit are virtually impossible to organize, and if the patient is in any sense sick they are probably unethical. Other research paradigms can, and have, produced important results.

Symptomatic treatment for stutterers was eclipsed in the 1930s by the rise of the mental hygiene movement. In the 1960s the behaviorists rediscovered the effective symptomatic treatments, but as attitudes of professionals are hard to change, acceptance was slow. We conducted a meta-analysis of 42 pre–post treatment studies of adult stutterers. The two most successful therapies have mean ES's of 1.6, the results last over time, measurement is reliable and valid, and the techniques have been demonstrated in different countries and on large groups of patients. But shouldn't we have had placebo control groups? From seven studies of untreated stutterers waiting for treatment we calculated that regression to the mean contributes about 0.2 S.D. in the first three months and nothing thereafter, and that placebo effect, measurement habituation, and spontaneous remission are not issues. Clearly, an apparently less than optimal research design can and did produce valid data (Andrews, Guitar & Howie 1980; Andrews & Harvey 1981b). In contrast, a methodologically purer meta-analysis confined to controlled studies found only four papers on outmoded treatments and produced an ES of −0.1 (Shapiro & Shapiro 1981). Similarly, meta-analyses of pre–post studies of agoraphobia and of baseline pre–post studies of treatments for hypertension have proven to be a satisfactory way to estimate the benefits of the psychological therapies for these disorders despite the absence of control groups (Andrews, MacMahon, Austin & Byrne 1982; Quality Assurance Project 1982).

We are at present working on a meta-analysis of controlled trials of the treatment of depression. Here the controlled trial, whether using no-treatment, wait-list, placebo, or comparison treatment is important, for both spontaneous remission and regression to the mean are confounding variables. In the treatment of neurotic depression we found four placebo-controlled trials of behavior therapy (mean ES = 0.64) and six of wait-list or comparison treatment-controlled trials (mean ES = 1.41). Thus the placebo-controlled trial may seriously underestimate improvement.

Leaving aside the placebo control issue, has it been shown that verbal psychotherapies are of little or no benefit to patients? In the eight studies that Prioleau et al. cite, only four (Brill, Koegler, Epstein & Fargy 1964; Gillan & Rachman 1974; Lorr, McNair & Weinstein 1962; Paykel, DiMascio, Haskell & Prusoff 1975) are studies of dynamic psychotherapy. The Paykel study was a study of maintenance of improvement in depressives who had already responded to drugs, while in the Gillan and Rachman study (as in the vaunted McLean & Hakstian [1979] study) the psychotherapy group was a control or alternate treatment in a study investigating the benefits of behavior therapy. In our study of neurotic depression we have already identified three placebo-controlled trials of dynamic psychotherapy (Paykel et al. 1975; McLean & Hakstian 1979; Weissman, Klerman & Prusoff 1979; Weissman, Prusoff, DiMascio et al. 1981). The first two, which Prioleau et al. cited, had ES's of 0.22 and 0.02; the other, which they chose not to cite, had an ES of 0.88, an effect comparable to the benefits of placebo-controlled trials of behavior therapy (0.64) or tricyclic antidepressant therapy (0.55) for this condition. Clearly, in some depressions, if the drugs haven't worked, psychotherapy or behavior therapy may be lifesaving.

Prioleau et al. found little evidence in four trials of dynamic psychotherapy or in four trials of counseling that patients benefited. However, many of their studies were potentially flawed despite the presence of a placebo control. Obviously, more research in psychotherapy is urgently needed, but even from our limited experience more data exist than they chose to analyze. It seems to me that even Erica Jong's (1977) character Isadora had better data than they. She had been in analysis and had lived with two psychotherapists before deciding she had to save her own life another way. And that is the issue. People seeking therapy are trying to save their own lives, to stop the continual self-defeat in work and love that constitutes neurosis. On balance I think that the verbal psychotherapies have been shown to work. Some behavior therapies work for some varieties of neurosis, and some verbal dynamic psychotherapies do likewise. Counseling seems no different from the nonspecific effects of therapy as represented by a placebo treatment, and sometimes even a placebo treatment can be preferable to no treatment.

## Placebo control conditions: Tests of theory or of effectiveness?

David S. Cordray and Richard R. Bootzin
*Department of Psychology, Northwestern University, Evanston, Ill. 60201*

Prioleau, Murdock, and Brody conclude on the basis of a meta-analysis that there is no evidence that psychotherapy is more effective than placebo controls. They contend that the effectiveness of psychotherapy (T) requires evidence that its effects exceed those demonstrated by placebo conditions (PC). Their analysis raises a number of important issues about the use of quantitative synthesis procedures to evaluate the effectiveness of psychotherapy. We will discuss two issues: (1) the appropriateness of the selection of placebo control conditions to evaluate the effectiveness of psychotherapy, and (2) how treatment conditions and research procedures should be characterized as part of the meta-analysis process.

Prioleau et al. have confused designs appropriate for answering questions about theoretical mechanisms with designs to demonstrate effectiveness (Bootzin & Lick 1979). It does not follow that psychotherapy is not effective from the finding that therapy is only marginally more effective than placebo treatments. Credible placebo manipulations are themselves often effective treatments (Lick & Bootzin 1975). Credible placebos are used to help understand why treatments work, not whether they work. For example, placebos are used in drug research to evaluate the extent to which the effects of a drug are due to its physiological components as compared to the psychological components of the drug-taking process. In systematic desensitization research, placebos are used to evaluate whether systematic desensitization works due to counterconditioning as opposed to more general features common to all therapy.

Even if Prioleau et al. intended to use meta-analysis to test the distinctive theoretical contribution of therapy, and not its "effectiveness," their conclusions are limited by having selected only studies including placebo comparison. First, this results in an atypical sample of studies. Placebo conditions are unlikely to be used in clinical settings. Of the 32 selected studies, 13 had children and 10 had college students as "patients." Second, and most important, there are other designs also assessing the distinctive contributions of therapy that could be evaluated. These include studies that use therapy component control conditions, dose-response evaluations, and within-subject comparisons. Thus, a focus on only placebo comparisons is inappropriately narrow. Despite these limitations, we might still ask whether Prioleau et al.'s conclusions are sound given the manner in which the assessment was conducted.

All things considered, the meta-analytic procedures used by Prioleau et al. are probably *too coarse* to provide a meaningful test of the effects of therapy beyond those demonstrated with placebo manipulations. A strong test of theory requires, *at a minimum*, (1) careful consideration of the quality of research procedures used in each study (i.e., statistical and internal validity), (2) careful examination of the adequacy of the operational realizations of treatments and the placebo conditions (i.e., construct validity of cause), (3) careful choice of measures that are relevant and sensitive to the T versus PC distinction (i.e., construct validity of effects), and (4) consideration of differences (across studies) in context, patient and therapist characteristics, and type of psychotherapy. Prioleau et al. did devote some attention to a few of these issues in the latter point.

As to quality, Prioleau et al.'s narrative description of the studies leads us to believe that they differed in quality. This could certainly have contributed to the sizable variability in effect sizes (T−PC) across measures and studies. Meta-analysts usually treat differential research quality as a covariate or cross-classification variable. Prioleau et al. did not.

Second, we know that treatment realizations do not necessarily match theoretical specifications. If the PC is contaminated with unintended treatment elements and the T is imperfectly implemented, the distinction between T and PC is diminished. Conversely, if PC is a weak realization of the factors it is supposed to control (Lick & Bootzin 1975), the difference between T and PC may be substantially overstated, again, a poor test of the theoretical distinction between T and PC. Note also that in a few studies, Prioleau et al. combined results from more than one placebo condition; the same was true when multiple therapy conditions were reported. Primary researchers usually invest resources in control conditions for good reasons. Combining different groups ignores these reasons and adds to the coarseness of the assessment.

Further, construct validity of effects is central to theory testing. Prioleau et al. made some effort to distinguish between types of measures (e.g., undisguised self-report versus others) but a more important point concerns the relevance of individual measures. If one looks at the measures they extracted from each study, they do not appear to be equally relevant or sensitive to the T versus PC comparison. So, for example, Warner (1969) used four measures to assess whether *alienated students* improved after group therapy based on verbal and model reinforcement. The placebo condition involved a group discussion. The effect size for each outcome was: Alienation (ES = .56); Anxiety (ES = −.06); Self-concept (ES = .11); and GPA (ES = −.15). The average is .12. A number of factors could obviously account for these differences across measures but given the focus of the treatment we would expect changes in alienation to be the most relevant proximal effect. Changes in GPA are not directly linked to T but are probably mediated by changes in level of alienation. In another study (Bruyere 1975), the subjects were *disruptive* students and the "rating of disruptive classroom behavior" was reported as "not available." Prioleau et al. assigned a value of 0.0 to this outcome. The average value for this study (−.18) is almost totally determined by a −.64 ES for self-concept outcome. Should these measures be weighted equally?

We are not sure. The point here is that relevance is important and should be systematically examined. Averaging across measures is not meaningful.

These issues are not unique to this analysis. They raise some serious questions about how quantitative synthesis should be conducted (see also Cordray & Orwin, in press). Our recommendation is to quantify available evidence – as was done by Prioleau et al. and Smith, Glass, and Miller (1980) – *and* also to assign a rating for each variable which characterizes the judged relevance of the measure, the integrity of the treatment, the adequacy of the placebo, the quality of the research, and so forth. These can then be used to classify the studies into strong versus weak tests of theory. If these factors are ignored, effects attributable to a particular theoretical notion will only be detectable if they are robust to the noise produced by differences across studies. We do not need to add additional noise by throwing everything into the "knowledge pot," stirring a bit, and hoping that an answer will bubble to the top.

## Give choice a chance in psychotherapy research

Hartvig Dahl
*Department of Psychiatry, SUNY Downstate Medical Center, Brooklyn, N.Y. 11203*

I want to make two points: (1) Prioleau, Murdock, and Brody's analysis does not reach a conclusion that is different from Smith, Glass, and Miller (1980); and (2) in future controlled comparison studies, patients ought to be given a choice of therapies and therapists.

Prioleau et al.'s conclusion that the benefits of psychotherapy do not exceed those obtained for placebo treatments for real patients should come as no surprise to anyone who has even casually looked at Smith et al., whose Table 5–1 (p. 89) gives .56 as the mean effect size for 200 measures of placebo treatment, compared with a mean of .63 calculated for 700-odd measures in psychotherapies of the type included by Prioleau et al. (Prioleau et al.'s value of .43 for placebo mean effect size is, as they indicate, equal to Smith et al.'s mean for 51 placebo measures with "neurotics/true phobics.") In fact, Smith et al. directly addressed this lack of differences among various psychotherapies:

> Our findings speak of communalities, not of differences: the essential equivalence of benefits from any serious attempt at psychotherapy; the *surprisingly* strong showing of placebo treatments, which by definition may potentiate general factors but not specific ones. The weight of evidence that now rests in the balance so greatly favors the general factors interpretation of therapeutic efficacy that it can no longer be ignored by researchers and theoreticians. . . . More research is needed in which quite different types of therapy are compared; and more such research should be entrusted to neutral third parties not caught up in holy wars. (pp. 186–87; italics added)

I propose that one plausible explanation for no differences is that traditional random assignment to experimental and control groups effectively eliminates variance due to patients' expectations and beliefs about their treatment. Prioleau et al.'s important contribution is to focus on the central role of beliefs in understanding the "surprisingly" strong showing of placebo treatments. Psychotherapy, they suggest, may sometimes be a "more believable placebo treatment than most actual placebo treatments." Thus a measure of expectations for therapeutic improvement might be related to outcome. A patient "may come to believe," for example, that psychotherapy is more powerful than relaxation training and thereby benefit more from psychotherapy. However, it is not that a patient comes to believe, but rather that he arrives in the first place with a host of tacit, unarticulated expectations and beliefs that bear on his encounter with any particular therapy, therapist, or placebo.

If I am educated and active and like to have something to say about my fate, I will surely not take well to a treatment that places a high premium on passive compliance. If I am passive and like things done to and for me I am conversely unlikely to thrive in a treatment that demands active participation. If I am comfortable only with warm, openly friendly people and I end up with a cool, detached therapist, I am not likely to prosper or take kindly to it. If I am uncomfortable with medication I may reject a pill placebo yet do well in a talking treatment. Indeed, such expectations and beliefs seem so compelling that it is startling to realize how little systematic research has been done on the issue beyond including separate patient and therapist variables and making limited attempts to assess the "match" between patient and therapist. Few studies have gotten at the subtle and tacit interactions that common sense says must play a role.

The Penn Psychotherapy Project (Luborsky et al. 1980; Morgan et al. 1982) is an exception, and their recent reports reinforce the hypothesis that such interactions are important for outcome. They found that random assignment was associated with significantly less improvement than assignment with the therapist's participation. Furthermore, measures of a patient's positive attitudes toward the therapist were among the best predictors of outcome. These findings support the idea that there ought to be a radical change in psychotherapy research designs if we are to do anything but add ad infinitum to the conclusions of Smith et al. and Prioleau et al.

The proposed change is simple. Initially, assign patients randomly to one of two groups: (1) an experimental group *within* which patients are offered a choice both among treatments (including placebos) and therapists; and (2) a control group *within* which patients are randomly assigned to different treatments (including placebos) and therapists. Since, as both Smith et al. and Prioleau et al. conclude, the available evidence says that there are no significant differences among the treatments and placebos, there should be no serious ethical question of informed consent. Although there may well be practical problems in designing "believable" placebos, these should not be different in kind from those in the past and should prove no more nor less manageable.

In order to give choice a true chance, after the initial random assignment to either the experimental or control groups, a patient in the experimental group could be given a video presentation of typical examples of, say, two psychotherapies and a placebo treatment, and after choosing among these would then be given a further choice among two or three therapists, again based on sample video presentations. A patient in the control group would be randomly assigned to a treatment or placebo and to a therapist.

The choice approach to psychotherapy research offers a design that more closely approximates the situation faced by prospective private patients, who often already have in mind a general kind of treatment and then may consult with more than one therapist before reaching a decision. Luborsky (personal communication) is currently offering patients the chance for "repairing," that is, for choosing among two therapists, each of whom the patient has seen three times.

It would be a shame if major lessons, learned at such great cost, were ignored and the field were to continue in the same vein, acting as if patients' wishes and choices had to be indefinitely sacrificed in our "holy wars" for turf.

## Trends based on cotton candy correlations

Robyn M. Dawes
*Department of Psychology, University of Oregon, Eugene, Ore. 97403*

Prioleau et al. graciously included the data on which they base their conclusions (see target article Table 1).

Table 1 (Dawes). *Regression analysis: Predicting effect size from sample size, duration, and patient status*

| | | N | $d$ | $p$ | "Validities" ($r$'s) |
|---|---|---|---|---|---|
| | N | — | .15 | −.04 | −.24 |
| Inter- | $d$ | | — | .35 | −.24 |
| correlations | $p$ | | | — | −.08 |
| $b$ Weights | | −.21 | −.21 | −.02 | |

(1) The average effect size is .42, "exactly in agreement with the magnitude of the psychotherapy effect size relative to placebo treatments estimated by Smith et al. (1980)." Twenty-three of the effect sizes are positive, eight negative. (The $z$ testing the null hypothesis of a .50 probability of a positive result is 2.69, $p < .01$; moreover, the $t$ testing a population mean of 0 effect size is 3.35, $p < .01$.) Usually, psychologists have almost a fetish about statistically significant results, but the authors find some *nonsignificant* trends in their data more alluring.

(2) Five such trends "do not support an intuitive model, which suggests that the effects of therapy are more powerful than the effects of placebo treatments." Three of these variables are assessed between studies: sample size (N), duration ($d$), and whether or not the subjects are patients ($p$; yes = 1, no = 0). From Table 1 of the target article it is possible to perform a multiple regression analysis with effect size as the dependent variable. The resulting multiple $R^2$ is .1024. Details are given in Table 1 (Dawes). (Using uncorrected effect estimates and two different calculators, I obtained an $r$, between $d$ and effect size, of −.24, rather than −.21; that would [if in error] only increase $R^2$, minutely.)

The sample size (N) is 32, the number of predictors ($k$) is 3, and the standard Wherry-Lord formula for estimating the "cross-validated" (same coefficients) correlation on a new sample is:

$$1 - (1 - R^2)\frac{(N - 1)}{(N - k - 1)} = 1 - (1 - .1024)\frac{31}{28} = .0062$$

Alternatively, equal (negative) weights correlate $.81^{1/2}$ with the optimal weights of −.21, −.21, and −.02 (for N, $d$, and $p$) so that the expected squared correlation of an unweighted composite is .0825, a larger value than above due to small $R^2$ and N. That might be a better estimate of the strength of these three variables *in combination*, because Prioleau et al. make no claims about differential effect.

(3) Consequently, the three between-study variables that Prioleau et al. stress jointly account for at best 8% of the variance, and we cannot even have the traditional probabilistic assurance ("significance") that they account for anything at all. A change of opinion is not warranted.

## Psychotherapy versus placebo: An end to polemics

Morris N. Eagle

*Department of Psychology, York University, Downsview, Ontario, Canada M3J 1P3*

Much of the debate in the area of psychotherapy outcome research is polemical. Skeptics and critics of psychotherapy present data demonstrating that psychotherapy does not work while advocates of and believers in psychotherapy desperately look for loopholes in the former's methodology and logic and scurry for evidence which proves that psychotherapy does work. My own view is that it is time to put an end to the debate in its present form. Even as well-reasoned and reasonably fair a paper as Prioleau, Murdock, and Brody's does not escape the kinds of problems characterizing work in this area as it is presently constituted. My purpose here is not to argue that psychotherapy does (or does not) work, but to demonstrate the ambiguity of the evidence. (It seems to me that even if the Smith, Glass, and Miller [1980] results were to hold up, the differences between therapy and control groups are too small to provide comfort to anyone espousing the benefits of psychotherapy.)

First and most important, any meta-analysis of already collected data, however sophisticated the statistical procedure, is as good or as poor as the data on which it is based. This is, of course, as true of the current re-analysis as it is of Smith, Glass, and Miller's original analysis. As Prioleau et al. themselves note, "given the diversity of measures, therapies, and subjects included in these investigations, there is some question whether attempts to relate quantitative indices of effect size to other variables is entirely legitimate."

There are, however, even more serious difficulties, having to do with the background and training of the therapists involved in the studies covered. A look at Prioleau et al.'s table 1 reveals that of the 32 studies investigated, there was no information on who the therapists were in 5 studies; in 1 study the therapist was at the M.A. level; in 5 studies they were school counselors; in 4 they were "psychologists," with no indication of the nature of their specialization and their training; the therapists were graduate students in 12 studies; in one study, it was a "teacher"; in 1 it was a psychiatric resident; in 1 a social worker, and in the remaining 2 studies therapists were a mixture of psychiatrists, psychologists, or social workers. Now, the question I put is whether it is really possible to reach serious conclusions on the basis of data from these sorts of studies. Would one ever want to test the efficacy of any intervention when its practitioners represent the kind of conglomeration I have just described? And, as for studies singled out by the authors partly because they involve experienced therapists, in one study (Lorr, McNair & Weinstein 1962) the duration of treatment was *four sessions* (!) and in the other study (McLean & Hakstian 1979) clinically depressed outpatients received 8–12 hours of "insight oriented psychotherapy."

Now, if a study as fair and as careful as the present one cannot escape the blatant deficiencies I have described, (and, I remind the reader, that this is so, not because of Prioleau et al.'s carelessness, but because of the limitations of the approach) then perhaps it is time to call a halt to outcome studies and debates in this current form. There is simply no substitute for controlled studies in which fine-grained microanalysis of process – that is, of what therapists (and clients) specifically *do* – is combined with and related to positive and negative outcome with specified types of clients. (One hopes that the current nationwide NIMH projects represent such an effort.)

I would also underscore Prioleau et al.'s call for individually tailored measures of outcome in "which one specifies at the start of treatment the kinds of changes that would indicate therapeutic progress for each patient." One may learn from such studies that certain types of problems are relatively amenable to psychotherapy and that some are not.

Some additional issues which seem to merit comment: Prioleau et al. tend to dismiss the "possibility that variations among therapists is a major or important source of variance in outcome research" partly because, given the zero or near-zero difference between psychotherapy and placebo groups, this "would imply that some therapists consistently make their patients worse" and partly because such differences would be "of little practical relevance unless some way can be found of communicating to potential patients information about the competence of therapists." With regard to the first point, I certainly would not dismiss the possibility that some therapists consistently make a large number of their patients worse. And as for the second point, whether of practical relevance or not, the existence of individual differences among therapists, if they could be related to what therapists actually do and what cues they provide to patients, would be of great theoretical importance.

There is one issue of practical, certainly of human relevance, that is not mentioned at all in discussions of outcome research. Let us assume that there are indeed no differences between psychotherapy and control, including placebo, groups. It is still possible, however, that while all subjects reach the same general end point, psychotherapy clients do so with a greater sense of support and with less pain and suffering (and possibly with a greater sense of where to turn should they experience similar difficulties in the future). If this were so, it would certainly justify, in human terms, the practice of therapy. However, we will never learn about these sorts of things unless we combine studies of outcome with a focus on ongoing process.

Finally, under the rubric "Fourth" in their introduction, Prioleau et al. seem to be somewhat confused regarding the placebo concept. Of course, some placebo conditions incorporate features (e.g., belief that the treatment is effective, hope, positive expectation) which may be the very same ones that account for positive outcome when psychotherapy is effective. But that is just the point. For in such cases no differences between placebo and therapy groups suggest that the factors which according to theory X are the operative factors in positive outcome (e.g., insight) are not the determinative ones. Rather, the very factors which according to theory X are placebo factors would turn out to be the operative factors. Of course, for a theory such as that proposed by Frank (1974), which says that expectancy, hope, and so forth are the critical factors in positive outcome, these features can no longer be seen as inadvertant placebo factors. Prioleau et al. cite Grünbaum's (1981) lucid article on the placebo concept, but they seem to misunderstand the implications of his basic point that what constitutes a placebo is always relative to a theory specifying factors other than the purported placebo factors as the operative ones.

## Psychotherapy, placebos, and wait-list controls

Edward Erwin

*Philosophy Department, University of Miami, Coral Gables, Fla. 33124*

I am sympathetic to many of Prioleau, Murdock, and Brody's views, and I think their contribution important, but I question some of the inferences that might be drawn from their results.

1. Prioleau et al. distinguish between: "(1) For the research literature surveyed, psychotherapy has been found to produce changes in real patients which are equivalent to the changes produced by placebo treatments, and (2) In general, psychotherapy produces changes in patients that are equivalent to those produced by placebo treatments." I am uncertain whether they intend to argue for (2) or not. In their abstract, they conclude that (2) is true, and later they suggest that it is "likely." Earlier, however, they suggest that they are arguing only for (1), and are merely speculating that (2) is true.

*If* they are arguing in support of (2), then one of their assumptions appears to be that acceptable evidence for psychotherapy being superior to a placebo must come from studies of "real" patients. This assumption requires support if they mean to include as *real* patients only outpatients or institutionalized patients. It may be true that favorable results from studies of this class of subject generally provide more convincing evidence, other things being equal, than evidence from studies of college students or other volunteers. However, what is the warrant for excluding the latter kind of evidence altogether? Suppose that the first well-designed study of systematic desensitization with real patients was not done until after 1970. Should we say, then, that none of the studies reviewed favorably by Paul (1969) provided *any* evidence that the therapy would be useful for treating similar problems with real patients? Why? The anxiety suffered by a volunteer may be just as severe as that of an outpatient; the problems of some outpatients are sometimes just as trivial as those of certain volunteers for experiments. I am not saying, of course, that there are no differences between so-called real and nonreal patients. I am asking: What are the general differences that explain why well-designed studies of real patients but not of nonreal patients are relevant to the assessment of proposition (2)?

Even if evidence from studies of nonreal patients is admissible, it could turn out that in every case where the therapy effect exceeds the placebo effect, the study has fatal design flaws. Demonstrating that this is so, however, would obviously be a formidable task.

2. A defender of psychotherapy might agree to proposition (2), but reply that there is still evidence that some forms of psychotherapy are effective for at least some real patients. The evidence comes from certain relatively well-designed studies of real patients in which the psychotherapy group showed significantly more improvement than the wait-list control group. One major example is the study by Sloane, Staples, Cristol, Yorkston & Whipple (1978). Prioleau et al. anticipate this reply by giving reasons for thinking that "wait-list controls may lead to outcomes that are more negative than would have occurred merely through the passage of time." Their reasons are plausible but not decisive; they may be counterbalanced by pointing out that patients on a wait list may expect to be helped in the not too distant future. This expectation may offset the disappointment at not receiving immediate help. Furthermore, wait-list patients often have some therapist contact, as in the Sloane study, and this may be beneficial. Given these conflicting speculations, empirical evidence is needed to determine the exact effects of being placed on a wait list. Some of the data from the study by Sloane and his colleagues bear on this issue. The improvement rate on target symptoms for the wait-list controls was 48%; and 77% of the control subjects were judged at the end of four months to have improved on overall adjustment. Is it plausible to think that the passage of time alone would have produced significantly more improvement? If not, then the Sloane study provides some evidence that a certain kind of psychotherapy is more effective with some real patients than no treatment. This conclusion does not contradict Prioleau et al.'s proposition (2), and I agree that (2) is important for some of the reasons they cite. However, the aforementioned conclusion is also of interest, and is the conclusion reached by more cautious supporters of psychotherapy. (For example, Vanden Bos and Pino [1980:36] do not deny [2], but conclude instead that: "The empirical literature amply demonstrates that psychotherapy is more effective than no treatment.")

3. Prioleau et al. count as a placebo any treatment so described by an investigator provided that: (a) "the possibility of therapeutic benefits is conveyed to the patient," and (b) "the authors [investigators] have a theoretical rationale for the assertion that omits features of psychotherapy which are essential for therapeutic effectiveness." On this account, a treatment may be a placebo and yet be more effective than no treatment. To illustrate, in the Gillan and Rachman (1974) study, the placebo treatment of phobic patients consisted of relaxation training plus discussion. However, there is evidence that relaxation training contributes to successful treatments of phobias with systematic desensitization; it has also been used by itself in treating anxiety conditions (Rachman & Wilson 1980: 124–28). I do not intend this point to be a criticism. Prioleau et al. concede that what an investigator hypothesizes to be therapeutically inefficacious may be effective, and they make the useful point that

comparisons to wait-list controls rather than placebos also raise problems. My point concerns what may reasonably be inferred from the conclusion that psychotherapy with real patients never produces therapeutic benefits greater than placebo treatments. Even if we discount all analogue studies and all studies using wait-list controls, we are still not entitled to infer that there is no evidence that psychotherapy with real patients is better than no treatment. We would need another premise: that for all well-designed studies in which the psychotherapy and placebo groups improved to the same extent, there is no warrant for thinking the placebo treatment efficacious. This premise, however, is dubious, as is shown by the example of using relaxation training as a placebo.

One theoretical possibility is worth mentioning here. Suppose, as Bandura has argued (1977, 1978), that psychological procedures work through a common cognitive mechanism: when successful, they serve as ways of creating and strengthening expectations of personal effectiveness. It may turn out, then, that some forms of psychotherapy and some so-called placebo treatments have modest beneficial effects for certain kinds of problems, and for the same reason. If this conjecture were established, then it would be less surprising to find psychotherapy groups doing no better than placebo groups, while exceeding the gains of wait-list controls.

## The effectiveness of psychotherapy: The specter at the feast

H. J. Eysenck

*Department of Psychology, Institute of Psychiatry, University of London, London SE5 8AF, England*

It is one of the ironies of life that the Smith, Glass, and Miller (1980) book on the benefits of psychotherapy, employing the method of meta-analysis, came to such very positive conclusions, which, as they stated, contradicted my own rather more pessimistic reviews. Their conclusion is premised on an almost incredible error of judgment, incorporated in their Table 5-1, which contrasts the effectiveness of different types of therapy, comparing each with a no-therapy control group. One of the types of therapy, oddly enough, is called "placebo treatment." Thus, Smith et al. seem to regard placebo treatment as a genuine type of psychotherapy, whereas Prioleau, Murdock, and Brody used placebo treatment as a control. It seems to me that there is no question whatsoever that Smith et al. are wrong and Prioleau et al. are right in this. Even using the definition of psychotherapy given by Smith et al., there is no doubt that placebo treatment does not in any way agree with this definition. The impression given to the reader is that Smith et al. failed to adopt the correct procedure of contrasting psychotherapy treatment effects with placebo treatment because to do so, would, as Prioleau et al. have shown, give entirely negative results, and would support Eysenck's (1952) original evaluation.

The effectiveness of placebo treatment, as compared with no treatment, is 0.56, on Smith et al.'s showing. The effectiveness of psychodynamic therapy is 0.69, Adlerian therapy 0.62, client-centered therapy 0.62, Gestalt 0.64, rational-emotive therapy 0.68, transactional analysis 0.67, reality therapy 0.14, etc.; it is clear that testimony from their own table implicates Smith et al. of inconsistency, and of drawing entirely false conclusions from their own data. Prioleau et al., in their own way, lend emphasis to this conclusion, which the present writer has elaborated in two recent critiques of the Smith et al. book (Eysenck in press a; in press b).

The effectiveness of psychotherapy has always been the specer at the wedding feast; where thousands of psychiatrists, psychoanalysts, clinical psychologists, social workers, and others celebrate the happy event and pay no heed to the need of evidence for the premature crystallization of their spurious orthodoxies, the need to do so, emphasized by experimentalists and other critical spirits, has always threatened to upset the happy union. Rachman and Wilson (1980), in their monumental review, might have been thought to have finally shown how substantial the specter was, and how premature the rejoicing; the publication of the Smith et al. book seemed to have revived the corpse, but as Prioleau et al. have shown, the glad tidings are spurious. The scholarly review of Rachman and Wilson should serve to convince even the most biased of the emptiness of the rhetoric which inspires the conclusions in Smith et al.'s book.

My only criticism of the Prioleau et al. contribution would be that, very much like Smith et al., they do not take very seriously the fact that data can only support or invalidate particular hypotheses. In their discussion of placebo treatment they go out of their way to suggest ways in which such treatment might contain substantive effects, effects which might justify Smith et al. in their inclusion of it as a proper method of treatment. However, none of these possible contaminants of a "pure" placebo treatment would justify us in regarding it as a substantive form of treatment when we judge it in terms of the psychoanalytic, cognitive, or other theories which have been advanced to legitimize the many different forms of psychotherapy. All the suppositions that Prioleau et al. examine are entirely ad hoc and irrelevant to the concept of "psychotherapy" as understood by psychotherapists. Presumably they are leaning over backwards in order to convince the faithful, but in doing so they depart from what I would consider the only proper position, namely one that examines the theories in question on their own terms. None of the theories would have predicted placebo treatment to be just about as effective as psychotherapy; it is hardly necessary to go beyond that. When we add that even Smith et al. found that duration of treatment was uncorrelated with effectiveness, and that the training of the therapist was also uncorrelated with effectiveness, we find that their major conclusions, too, are directly in contradiction with their own data. I have no doubt, therefore, that Prioleau et al. are right in their evaluation, and that they carry forward the meta-analysis of Smith et al. in a very valuable fashion.

It is unfortunate for the well-being of psychology as a science that, however clear-cut the results of this study may be, the great majority of psychologists, who after all are practising clinicians, will pay no attention whatsoever to the negative outcome of all the studies carried on over the past thirty years, but will continue to use methods which have by now not only failed to find evidence in support of their effectiveness, but for which there is now ample evidence that they are no better than placebo treatments. This is not only sad from the point of view of psychology as a science; it is difficult to see how, from the ethical point of view, we can reconcile this refusal to face facts with the social duties imposed on the applied scientist. Do we really have the right to impose a lengthy training on medical doctors and psychologists in order to enable them to practise a skill which has no practical relevance to the curing of neurotic disorders? Do we have the right to charge patients fees, or get the State to pay us for a treatment which is no better than a placebo? Do we have the right to continue to teach students general psychology theories, such as the Freudian, for which there does not exist any experimental evidence, and which have failed in their application to psychiatric treatment? These are serious questions raised by the Prioleau et al.'s target paper; they admit, in my view, of only one answer. Psychology will have to disown methods and theories which fail to provide sound positive evidence for their value, and will have to begin to imitate the hard sciences, where, as T. H. Huxley once said, we encounter the true tragedy of science – the slayings of a beautiful theory by a hard fact. Freudian and other psychological theories have hitherto been under a preservation order which has made them immune to the killing effect of adverse facts; it is time this preservation order was rescinded.

## Enhancing the therapeutic respectability of placebos

Jefferson M. Fish
*Department of Psychology, St. John's University, Jamaica, N.Y. 11439*

Placebos, even in the article by Prioleau, Murdock, and Brody, seem continually to be viewed in a negative light. Frank (1973) and Fish (1973) have argued that the placebo effect is a potent one, and that psychotherapy (as the term is used by Prioleau et al.) may well be effective because of it. If this is so, the appropriate course of action would seem to be one of maximizing the social influences referred to as placebos by developing ways of tailoring them to individual clients and their problems. For some reason, though, the alternative course of viewing placebos as a control treatment to be overwhelmed by the "real thing" seems to be much more popular. Prioleau et al. present interesting evidence that this line of investigation has not lived up to expectations.

In truth, many therapists, for a variety of what they take to be theoretical or ethical reasons, do not define goals clearly with clients so that the latter can know whether they are improving, they do not communicate in ways believable to clients that they are likely to improve, and they imply that therapy is a long, drawn-out process with an uncertain outcome. This may be the reason for the (nonsignificant) negative correlation between effect size and duration of therapy: The longer it went on, the less reason the clients may have had for believing the therapy would do them any good. Judging by the limitations in how psychotherapy functions as a placebo – though the target article gives no way of evaluating this in the studies cited – it is impressive that therapy functioned as well as placebo treatment. One possible explanation is that the placebos themselves may not have been that potent.

Advocates of many of the traditional forms of psychotherapy have argued that the therapeutic relationship is so important that it is the principal effective ingredient in change. This may be so, but for a different reason. It may be that the relationship functions essentially as a placebo component that can carry along any other elements which possess additional validity (e.g., techniques of behavioral or family therapy). Prioleau et al.'s tentative suggestion of giving pill placebos instead of psychotherapy may be premature, since two conditions would have to be met, neither of which is addressed by their paper. These are (1) that there be no interpersonal placebos which are more effective than pill placebos, and (2) that there be no therapeutic techniques which have validity independent of the placebo effect. While space limitations make it impossible to address these issues in detail, they can at least be commented on briefly.

The work of T. X. Barber and associates on hypnosis – much of which has been summarized in Barber (1969) and Barber, Spanos, and Chaves (1974) – contains control groups who receive task-motivational instructions. These control groups can be seen as being in many ways analogous to the placebo controls in the studies discussed. Despite the fact that they do not receive a hypnotic induction, subjects' behavior alters within minutes in dramatic ways, so that they produce the various classical hypnotic phenomena, including age regression, amnesia, auditory and visual hallucinations, deafness, blindness, and so forth. While this is an extremely complex subject, I am referring to it merely to point to the possibility that the outer limits of interpersonal placebos have not yet been explored. Many other areas of social psychological research could be cited to make essentially the same point.

As to the second issue, the authors' decision to limit their review to psychotherapy, along with certain of their remarks, suggests that they are aware that there may be therapeutic techniques which have validity independent of the placebo effect, though they may be enhanced by it.

One final methodological point deserves mention. Despite their attempt to limit the focus of their review, Prioleau et al. are aware that the studies listed in their Table 1 are quite diverse as regards type of therapy, subjects, contact hours and the other dimensions along which they are classified. One may question the appropriateness of performing a meta-analysis on such a varied group of studies. That is, looking for the average effect of psychotherapy versus placebo on personal problems may be like looking for the average effect of drugs versus placebo on illness. Thus, their contribution would seem to be a conditional one: If it is meaningful to evaluate the overall effectiveness of psychotherapy, and if this evaluation can be accomplished by a meta-analysis of a diverse group of studies, then psychotherapy is no more effective than placebo treatment.

## The placebo is psychotherapy

Jerome D. Frank
*Department of Psychiatry, Johns Hopkins Hospital, Baltimore, Md. 21205*

The finding that in a subset of the studies analyzed by Smith, Glass, and Miller (1980) psychotherapy proved to be no more effective than placebo provides food for thought.

Like the original study from which this sample is drawn, the findings are far from conclusive because of many deficiencies pointed out by others (Parloff 1980), but they have been replicated so often that they must be taken seriously. They are confirmed by our research showing that symptom relief with placebo is on the average identical with that from four months of psychotherapy (Frank 1978), and indirectly by the Sloane, Staples, Cristol, Yorkston, and Whipple (1970) study, which found that wait-list patients showed a 75% improvement rate, the wait-list patients having first had a long intake interview with a psychiatrist. Presumably patients perceived this intake interview as therapeutic. In our placebo studies many patients experienced a marked drop in symptomatic distress following the extensive initial work-up and before they received the pill, apparently for the same reason (Frank 1978). Along the same lines, in the Smith et al. (1980) study success was unrelated to duration of therapy.

With many patients the placebo may be as effective as psychotherapy because the placebo condition contains the necessary, and possibly the sufficient, ingredient for much of the beneficial effect of all forms of psychotherapy. This is a helping person who listens to the patient's complaints and offers a procedure to relieve them, thereby inspiring the patient's hopes and combatting demoralization. The help may be immediate in the form of medication or it may be anticipatory, as with patients on a wait list.

In contrast to the finding of Prioleau et al. is the widespread clinical impression that a few patients who have had symptoms for years and have not responded to therapy previously do respond to the therapy being given by their last therapist. Although this impression is based largely on self-reports, with all their sources of contamination, it is too widespread to be dismissed.

It may be possible to reconcile these two sets of findings by remembering that in all the reported studies, about 30% of patients are not improved by any of the therapies listed. These unimproved, as well as those included among the improved who improved only a little bit, might be the ones who would have responded differentially to therapies other than those included in the sample. Omitted therapies include, for example, cognitive therapy for depression, which some studies have found to be at least as effective as medication (Rush, Beck, Kovacs & Hollon 1977).

Rational-emotive therapies may also be inadequately represented. Only five are reviewed. Four obtained a positive if weak effect size as compared with placebo in anxious college students

(DiLoreto 1971; Grande 1975; Meichenbaum, Gilmore & Fedoravicious 1972; Trexler & Karst 1972). This finding is at least consistent with the hypothesis that emotional arousal may be relevant to therapeutic change.

In this connection, the most serious omission may be the absence of all therapies focused on producing intense emotional arousal and/or altered states of consciousness, such as primal therapy (Janov 1970; Rosen 1977), est (Baer & Stolz 1978), the procedures of Milton Erickson (Haley 1977), and many others. Most of these therapies are not considered respectable in research circles, and the practitioners of all are not temperamentally equipped to do controlled research, nor are the therapies temperamentally congenial to most researchers.

Thus it may be that the approximately three-fourths of patients who are relieved equally by psychotherapy and placebo are suffering primarily from demoralization. Symptoms found by Dohrenwend, Oksenberg, Shrout, Dohrenwend & Cook (1982) to characterize this condition are poor self-esteem, hopelessness-helplessness, dread, confused thinking, sadness, anxiety, psychophysiological symptoms, physical ill health. In the Dohrenwend studies, the remaining 30% who do not respond to either psychotherapy or placebo suffer from other symptoms such as obsessions, addictions, and symptoms of psychoses. In these patients, although the function of the symptoms in the person's adjustment to life may be determined by psychological factors, their *form* may be primarily determined by pathological processes in the central nervous system. As such, they can be modified only by procedures powerful enough to affect the central nervous system, directly or indirectly.

Medication is obviously the most direct way of relieving symptoms through influencing the central nervous system (Klein 1981; Marks 1981), but it is also possible that indirect influences on the central nervous system may be exerted by psychotherapies that produce altered states of consciousness or strong emotional arousal. These must have physiological concomitants, even though we have not yet identified them. The bridge between psychotherapy and changes in the central nervous system is being built by researchers in neurophysiology, who are beginning to specify the precise relationships between certain psychological processes like intentions (Mountcastle 1975) and learning (Kandel 1979) and specific electrical or biochemical changes in the central nervous system.

In short, the similar effectiveness of placebo and conventional psychotherapy in relieving symptoms accompanying demoralization could be explained by the hypothesis that central nervous system processes underlying demoralization are easily reversible by any intervention that arouses the patient's hopes and similar emotions. Symptoms not produced by demoralization, such as obsessions, involve more refractory neuropathological processes that could be relieved only by more powerful interventions such as medications or psychological procedures producing altered states of consciousness or very strong emotions. Since these procedures are not included in the studies reported in Prioleau et al., the possibility that they might be more effective than placebo must remain open.

## Does psychotherapy work? Yes, no, maybe

Sol L. Garfield

*Department of Psychology, Washington University, St. Louis, Mo. 63130*

The analysis of psychotherapy versus placebo studies by Prioleau, Murdock, and Brody adds another touch to the perennial controversy over the effectiveness of psychotherapy. It is indeed a rather depressing state of affairs when what appear to be diametrically opposed conclusions are drawn from the same source of research studies by different reviewers.

Reviews of the literature by different reviewers in the past have led to differing conclusions concerning the relative effectiveness of psychotherapy (Garfield 1980; Garfield in press). In some cases, the opposing views were due to different interpretations of results or to more critical evaluations of certain studies. In other instances, differing conclusions were considered to result in part from selectivity in the studies reviewed (Meltzoff & Kornreich 1970; Rachman & Wilson 1980; Smith, Glass & Miller 1980). Thus, disagreements in evaluation could be ascribed to the particular preferences or potential biases of the individual reviewers. The application of meta-analysis to psychotherapy outcome data, however, appeared to usher in an era of possible objectivity and systematic analysis in this confused and controversial field (Smith & Glass 1977). Surprisingly, the use of meta-analysis has appeared to raise the level of controversy to new heights.

The current report by Prioleau et al. is simply the latest in a series of individual papers on this topic. In the present instance, the findings and interpretations of Smith et al. (1980), based on their meta-analysis of 475 studies, are challenged on the basis of an analysis of a selected group of 32 studies in which psychotherapy was compared to a placebo control condition. Although this is outwardly defensible, there are evident problems in the meaning and comparability of the placebo conditions used in the different studies and in the conclusions which can be drawn. An examination of the different placebo conditions listed in Table 1 in Prioleau et al.'s target article clearly demonstrates this. Although time does not allow for a detailed appraisal of each of the studies included, an examination of the table raises questions about the suitability of the placebo used. This is the case in the use of a phobic hierarchy as a placebo in the study of phobic patients by Gillan and Rachman (1974) and the use of an RET (rational-emotive therapy) book as a placebo condition in evaluating RET therapy (Jarmon 1972). I would in fact be loath to base any conclusions on the latter study, which had three contact hours provided by psychology graduate students.

It is also pertinent to point out that in two other meta-analytic studies, the general conclusions of Smith and her collaborators have received support. Landman and Dawes (1982), for example, took a sample of studies from the total group analyzed by Smith and Glass (1977) in which each study had a randomly selected no-treatment or placebo control group. They also determined the magnitude of the placebo effect and "the influence of statistical non-independence of data points on the results obtained by Smith and Glass" (Landman & Dawes 1982:508). In brief, their findings, based on 42 studies, were in essential agreement with those reported by Smith and Glass (1977). Shapiro and Shapiro (in press), in a meta-analysis of a different sample of 143 studies which only overlapped slightly those of Smith et al. (1980), also secured results which were congruent with those of the latter.

We are thus presented with different conclusions, and the matter of interpretation becomes critical. For example, in the analysis by Prioleau et al., the mean effect size for psychotherapy compared with placebo treatments is .42, the same estimate offered by Smith et al. The comparable effect size secured by Landman and Dawes is .38 when individual outcome measures are the unit of analysis and .58 when each individual study is the unit of analysis. Generally, these findings would be viewed as relatively consistent. However, instead of agreement, we find very different conclusions drawn by the different reviewers. "The results of the present study . . . uphold the positive conclusions regarding psychotherapeutic efficacy originally drawn by Smith and Glass" (Landman & Dawes 1982:504). In contrast, Prioleau et al. state: "It was concluded that for real patients there is no evidence that the benefits of psychotherapy are greater than those of placebo treatment." Why this difference in the conclusions reached? The answer seems to be that individual reviewers with different value systems interpret the same data differently.

As long as reviewers evaluate and make judgments on individual studies, there are going to be differences in the conclusions drawn, as is evident in the present instance. To the present writer this is reminiscent of the nature-nurture controversy, which has been around much longer, and I'm not hopeful of a speedy resolution of the current controversy. For example, Prioleau et al. emphasize certain findings in a manner which appears suggestive of a particular attitude toward psychotherapy. Their conclusions concerning sample size, duration of therapy, and the use of real patients are based on nonsignificant findings. Such a practice is hard to condone and they can be rightly criticized for their overemphasis on nonsignificant results. Other remarks made in the target article also seem to be highly critical, even when they are mainly speculative.

There is little question that much of the research on psychotherapy is not representative of the psychotherapy carried out in existing clinical settings. Much of the research has been of an "analogue" nature (Shapiro & Shapiro in press), and many forms of therapy have received little research study. Many if not most studies have rather serious defects, and it is not difficult to evaluate them critically. The variability of findings among studies has also been noted in past reviews. Consequently, we need to be cautious in the conclusions we draw. High quality clinical research is expensive and time consuming, but such research is necessary before definitive conclusions concerning the efficacy of psychotherapy are clearly supported. Clinicians should be sensitive to this problem. At the same time, reviewers of studies have a responsibility to use wisdom and caution and to avoid what may be viewed as sensational conclusions.

## Placebo effects in psychotherapy outcome research

Gene V Glass;[a] Mary Lee Smith;[a] and Thomas I. Miller[b]
*[a]University of Colorado, Boulder, Colo. 80309; [b]City of Boulder, Boulder, Colo. 80302*

Prioleau, Murdock, and Brody analyzed the same set of studies we analyzed (Smith, Glass & Miller 1980) and confirmed the fact that the average person receiving verbal psychotherapy is advanced to a position about .40 standard deviation units above the average person receiving placebo treatment on a variety of outcome measures. An effect of .40 sigma units is roughly the size of the gain that students in a graduate seminar in psychology would show from pretest to posttest on a comprehensive written exam over the contents of the course – and no one we know suggests that students cease to study or professors to teach. Yet, Prioleau et al. move in 12 paragraphs from "Results" to such "Concluding Speculations" (their word, not ours): "On the basis of the available data we see no reason to believe that subsequent research using better research procedures and investigating other types of therapy administered to other types of patients will yield clear-cut indications that psychotherapy is more beneficial than placebo treatment." Tracing (in fewer than 1,000 words) the unlikely path from the statistical results to the concluding speculation is difficult, but we shall try.

Prioleau et al. base their speculation on "an intuitive model" of how effective verbal psychotherapy should perform, which they contend is not supported by the data; they fail to consult data from studies published in the last five years; and they engage in the ad hoc impeaching of evidence that runs counter to their conclusions in a manner that has long been common to those who have attempted to discredit the value of verbal psychotherapies.

Prioleau et al. argue that if verbal psychotherapy is truly more effective than placebo treatment, then its superiority should (1) grow the longer it is applied, (2) grow the longer one waits to "follow up" the effects, (3) be relatively greater with more objective (less transparent) measures of outcome, and (4) be greater for "real patients" than for "solicited subjects." Prioleau et al. claim that each prediction of this intuitive model is contradicted in the data from the verbal psychotherapy versus placebo experiments and, hence, that verbal psychotherapy is not superior to a placebo, even though the aggregate superiority in the literature is +.40 sigma units.

Their "intuitive model" is ad hoc and unconvincing. One need only read its rationale to see what arbitrary patchwork it truly is. But the error in their reasoning is more serious than that. Prioleau et al. fail to distinguish what is true about psychotherapy from what is true about the *literature of psychotherapy research*. (One of us has examined this confusion at some length and shown how it has led to some misunderstandings of our work; see Glass & Kliegl 1982). The relationship between outcomes and such *study characteristics* as therapy duration, type of client, reactivity to outcome measure, follow-up interval, and the like will depend on how *studies* are designed and may not reflect directly on how psychotherapy processes themselves operate. One example should suffice: If studies that use psychotics as subjects are designed so that more hours of therapy are given than in studies using mildly phobic subjects, then at the level of meta-analysis of studies there will appear little or no correlation between duration of therapy and effect size because the more disturbed patients will improve less than the phobic patients despite receiving longer therapies. Whether longer therapies are superior to shorter ones must be examined *within* studies with proper controls exercised.

Our work was published in 1980; the manuscript went to the publisher in 1979; we wrote it in 1973, so our literature search covered studies published before then. We cannot imagine why Prioleau et al. confined their analysis to our data when they could have searched at least four more years of material. In less than an afternoon of library work, we found a dozen comparisons of verbal psychotherapy and placebo published since 1978 in the *Journal of Consulting and Clinical Psychology* and the *Journal of Counseling Psychology*. The studies show consistently positive and large effects of verbal psychotherapy in comparison to placebo treatment.

Prioleau et al. aver that the most important distinction to draw among the verbal psychotherapy versus placebo experiments involves the "types of patients" studied. They distinguish four types: outpatients, inpatients, students, and "others." They classify as "outpatient" three of the experiments we included in our analysis, and they consider them as having failed to show superiority of verbal psychotherapy to placebo. We read the findings of these studies as being more equivocal and mixed. Prioleau et al. devote two paragraphs to a description of McLean and Hakstian's (1979) study, although it appeared too late for our book and was not included in Prioleau et al.'s data either, because it was well designed (an indirect compliment to one of us, who served as the doctoral adviser of one of them) and because it showed no superiority of verbal psychotherapy over placebo for depressed outpatients (although it did show superiority of behavioral therapy to placebo). Prioleau et al. have more faith in one study than we do; we offer Comas-Díaz (1981) as an outcome study on depressed outpatients with findings that counterbalance McLean and Hakstian (1979) – but we hasten to add that advancing findings of "favorite studies" is not the way in which knowledge about psychotherapy will grow.

Of four studies with inpatients that Prioleau et al. single out, they discount a large superiority of verbal psychotherapy over placebo (Coche & Flick 1975) because it was "nonreplicable" (Coche & Douglas 1977). That logic is assymmetrical, to put the matter kindly. What is really at issue is Prioleau et al.'s expectations for empirical regularity in psychotherapy research. (Why, for example, isn't it concluded that the zero effect "failed to replicate"?)

The logic with which they impeach 21 studies in the "students" category makes the mind whimper. The studies are irrelevant, they maintain, because the psychotherapy is brief, the number of subjects are small, the subjects were solicited (hence, not "true" or "real," as they say), and because rational emotive and social learning therapy are "quasibehavioristic." Questions of the external validity of psychotherapy research have occupied methodologists for years, and thinking has progressed considerably beyond calling some subjects "true patients" and others "untrue." In scores of summaries of the drug and psychotherapy literature, we found authors ignoring hundreds of studies because they failed to meet idiosyncratically defined standards of acceptable evidence. Perhaps the cat is let fully out of the bag by Prioleau et al.'s final quibble. When a verbal psychotherapy shows superiority to placebo, they call it a "quasibehavioristic" therapy and thus suggest that it is incidentally successful or that its success may be credited as they see fit. "An Analysis of Psychotherapy versus Placebo Studies" looks to be just another skirmish in the antagonistic battle between verbal and behavioral psychotherapy.

## Revisiting psychotherapy outcome: Promise and problems

Roger P. Greenberg

*Department of Psychiatry, State University of New York, Upstate Medical Center, Syracuse, N.Y. 13210*

The article by Prioleau, Murdock, and Brody represents the third attempt of which I am aware to reanalyze the psychotherapy outcome data reviewed by Smith, Glass, and Miller (1980). Each previous reanalysis of this work came to conclusions that are at striking variance with Prioleau et al. Landman and Dawes (1982), looking only at studies with no-treatment or placebo comparison groups, upheld the positive conclusions regarding psychotherapeutic efficacy drawn from the original analysis. Psychotherapy was found to be more potent than either placebos or no-treatment in a sample of the original set of outcome studies. Similarly, Andrews and Harvey (1981) reexamined the data only for studies of persons who would normally seek psychotherapy. Unlike Prioleau et al., they focused their analysis exclusively on the reports of treatment for neurosis, depressive disorders, or emotional-somatic disorders. Patients had to have sought treatment themselves or been referred for treatment. No analogue studies (such as those with recruited student subjects) were included. Again, the reanalysis proved to be strongly supportive of positive psychotherapy outcome. Furthermore, both the verbal and the behavioral psychotherapies were deemed superior to placebo treatment.

There are differences among the reviews with respect to the studies selected for inclusion and the statistical treatment of the data. All of the reanalyses use different subsets of the original group of studies reviewed by Smith et al. It is interesting that Prioleau et al. managed to somehow include in their "reanalysis" studies that were not even included in the original psychotherapy meta-analysis. For example, the authors indicate that there were only three studies in the grouping they considered "most crucial for an evaluation of psychotherapy outcome." Two of these three studies were not listed in the psychotherapy meta-analysis performed by Smith et al. (Brill, Koegler, Epstein & Forgy 1964; Lorr, McNair & Weinstein 1962).

The reviews also differ on what they use as the unit of analysis. The original work, and the review by Andrews and Harvey, use each outcome measure as an independent unit of effect size, while the Prioleau et al. review, as well as the work of Landman and Dawes, pool effect sizes within each study before averaging across studies. Each method has its own disadvantages. Using each measure independently weights the overall findings disproportionately with studies that contribute more measures. Pooling the effects of various measures within any study (as Prioleau et al. have done) obscures results in studies showing improvements on some measures and none on others. Many studies do indeed show variability in effect size across measures.

Yet, despite differences in the studies included and methods of analysis, all three previous reports are consistently positive in their conclusion that psychotherapy treatments produce results that are superior to either placebos or no treatment. Are Prioleau et al. justified in coming to a different conclusion? I think not.

Their own analysis indicated an overall effect size of .42. This can be interpreted to mean that the average patient treated with psychotherapy was better off than 66% of the patients receiving placebo treatments. Moreover, 72% of the articles reviewed showed that psychotherapy achieved results superior to placebos. Admittedly, not all effects were strong, but the gross results are hardly an indictment of psychotherapy.

More importantly, one must ask whether Prioleau et al. really have the data base from which to draw general conclusions about individual psychotherapy outcome for "real" adult patients. An examination of the studies reviewed indicates 81% used inexperienced or relatively untrained therapists. In addition, 75% were based on group procedures, *not* individual treatment. The duration of therapy was 10 or fewer sessions in 69% of the sample and more than 20 sessions only in 9%. Only 8 studies looked at individual treatment and at least half of these were based on fewer than 10 sessions, used inexperienced therapists or dealt with children or recruited subjects. The diversity of measures, therapies, and subjects led even Prioleau et al. to raise questions about whether effect size indices could be legitimately related to other variables. The class of studies considered "most informative" included only three reports, and these were variously burdened with such problems as inexperienced therapists, few sessions of treatment, the inclusion of psychotic patients, lack of clearly defined treatments, and high patient dropout rates. Overall, the characteristics and heterogeneity of the few studies selected for review suggest that they are unsuitable as a basis for drawing sweeping inferences about outpatient psychotherapy.

Prioleau et al. seem very concerned that they found no relationship between therapy effects and duration. In view of the relative brevity and restricted range of "treatment" durations sampled, this does not seem surprising. Many investigators have found a relationship between duration and outcome (Luborsky, Chandler, Auerbach, Cohen & Bachrach 1971 cited 20 supporting studies). Moreover, a comprehensive review of the empirical evidence bearing on Freud's ideas (Fisher & Greenberg 1977) suggested that dynamic therapy studies require a treatment duration of at least six months to a year to evaluate optimal effects.

As highlighted in our review of research on psychoanalytic propositions (Fisher & Greenberg 1977), the literature is filled with attempts to test the efficacy of amorphous meetings between individuals that are deceptively given a dynamic psychotherapy label. Clearly, more specificity and measurement are needed in defining such treatments. Terms (such as "insight") are bandied about too loosely in the literature. Many reports in the Prioleau et al. review exemplify this problem well. Without specificity, differences between psychotherapy and placebo are indeed blurred.

The authors' lament that competence cannot be practically determined seems overly pessimistic. Actually, exciting debates, based on evidence, are being waged about therapist factors necessary for good outcome, and future measurement may be a good deal more sophisticated (e.g. Parloff, Waskow & Wolfe 1978). The measurement of therapist incompetence has led to clearer results. Researchers have repeatedly found that therapist psychopathology and conflicts are harmful to patients (Greenberg & Staller 1981). Unfortunately, there is little indication that concerns about incompetence have as yet had much impact on the selection or training of future therapists.

In sum, three previous reviews, from different perspectives, all offer positive conclusions about the advantages of psychotherapy relative to placebos. Prioleau et al. offer a contrasting viewpoint. While the positive interpretations are appealing, deficiencies in previous research leave room for debate. Old data do not improve with age, and definitive answers await more careful specification and measurement of treatments.

## Statistical summaries in research integration

Larry V. Hedges
*Department of Education, University of Chicago, Chicago, Ill. 60637*

The method of meta-analysis has frequently been criticized for oversimplifying the results of a series of studies. Many of the criticisms stem from concern about summarizing effect sizes from many studies by a few parameter estimates. In the simplest case a reviewer summarizes the results of a series of studies by a single estimate of effect size, usually the average of all the effect size estimates. Is this an oversimplification of the results of the studies? One suspects that it often is, but until recently methods to address this question directly were not available. Some authors have used the strategy of examining subgroups or of fitting regression equations to vectors of study characteristics, but the results are often inconclusive. Presby (1978) pointed out that reviewers could reasonably disagree on the characteristics of studies that might be related to effect size. Failure to find a systematic relationship between characteristics of studies and effect sizes might therefore stem from a poor choice of study characteristics, and not from a lack of systematic variation in effect sizes.

Recent progress in theory for the statistical analysis of effect size data provides better alternatives. A statistical test of homogeneity of effect sizes across $k$ experiments (Hedges 1982a; Rosenthal & Rubin 1982) provides a method for testing the agreement of effect size estimates. This test provides a method for determining whether the variation in the effect size estimates is greater than would be expected if all of the studies shared the same underlying (population) effect size.

Given that the assumptions for the $t$ test are reasonably well met in each study and that the treatment and placebo groups are approximately equal in size, we can apply these methods to the data presented by Prioleau, Murdock, and Brody. Here we also simplify the analysis by treating the average effect size estimate for each study as if it were the only effect size estimate for the study. The test for homogeneity of effect sizes from $k$ studies yields a statistic $H_T$ that is distributed as a chi-square on $(k - 1)$ degrees of freedom when all of the studies share a common effect size. The value of this homogeneity statistic for the 32 studies reported is $H_T = 82.94$, which is significant beyond the .001 level. That is, it would be very unlikely to obtain the 32 effect size estimates reported if all the studies really have the same underlying effect size.

It is therefore dubious to proclaim any single value as representative of the common underlying effect size for all of the studies. The studies do *not* all give the same answer in the sense of estimating the same underlying effect size. In particular, we should be skeptical of the overall average effect size as describing anything particularly meaningful. Looking more closely at the studies, there is one study (Platt 1970) whose effect size is very much larger than the rest. This study was identified by Prioleau et al. and would surely be rejected as different from the other 31 studies by a statistical test such as that developed by Hedges (1982b). Omitting this study as a potential outlier, the 31 remaining studies still do not have homogeneous effect sizes ($H_T = 60.46$, $p < .01$). The authors suggest that one way to understand their results is to group the studies into four subclasses according to the type of patient receiving therapy (e.g., outpatients, institutionalized patients, students, and others). Does this model adequately explain the effect sizes of the 31 studies (note the exclusion of the obviously outlying study)? I answer this question by using the analogue to the analysis of variance for effect sizes proposed by Hedges (1982b). This technique partitions the overall fit statistic $H_T$ into between subclass $H_B$ and within subclass $H_W$ fit components, and permits rigorously defensible tests of the relationship between subclass and effect size as well as tests of the overall fit of the subclass model. Table 1 (Hedges) provides a summary of this analysis, showing that the overall model does not fit well ($H_W = 54.12$, $p < .01$), which suggests that these categories do not explain the variation in effect sizes. We see that the between-class variation is not significant ($H_B = 6.29$, $.05 \le p \le .10$). This is suggested by the substantial overlap of confidence intervals for the average effect sizes of the four subclasses. Thus the proposed model neither fits the data well nor explains a significant amount of variation in the effect sizes.

If Prioleau et al.'s model is not very enlightening, how should we look for a better explanation of the variation in the data? One possibility is to exclude a small portion of the data for separate consideration and to model the remaining data. There is a simple alternative to the authors' model that is reasonably consistent with over 90% of the data. If we exclude two additional studies (House 1970; Roessler et al. 1977) for separate consideration on empirical grounds, then the effect sizes of the remaining 29 studies are reasonably consistent ($H_T = 35.58$, $.10 \le p \le .20$). That is, the variation exhibited by the remaining 29 studies does not seem to contradict strongly the model of a common effect size. (This assertion should not be taken as a rigorous significance test, since the excluded studies were chosen on the basis of effect size. A rule of thumb, however, is to decide that the model fits well if a small proportion of values are excluded and the fit statistic is reasonably close to $k - 1$, the number of degrees of freedom for the reference distribution.) The weighted average effect size is .15 and a 95% confidence interval for the effect size (using theory given in Hedges 1981 or 1982a) is .04 to .26. One might conclude, therefore, that the bulk of the data is consistent with a small but positive effect size. If pushed for a statistical significance test, the conclusion would be that this effect size is significantly different from zero at the

Table 1 (Hedges). *Summary of statistical analyses*

| Subclass/ source | Effect size estimate[a] | 95% confidence limits[b] Lower | Upper | Fit statistic[c] | Degrees of freedom |
|---|---|---|---|---|---|
| Outpatients | .08 | −.29 | .45 | .01 | 1 |
| Institutionalized patients | .45 | .19 | .70 | 17.16** | 5 |
| Students | .19 | .06 | .32 | 36.99* | 20 |
| Others | 0.12 | −.52 | .29 | .01 | 1 |
| Between classes | — | — | — | 6.29 | 3 |
| Total | .21 | .10 | .32 | 60.46** | 30 |

[a]Weighted average effect size. [b]Confidence intervals using methods given in Hedges (1982a). [c]Within class, between class, or total fit statistics as defined in Hedges (1982b).
$*p < .05$. $**p < .01$.

5% level. One would also have to add that about 10% of the studies showed different and substantially larger effect sizes.

## Meta-analysis of psychotherapy: Criteria for selecting investigations

Alan E. Kazdin
*Western Psychiatric Institute and Clinic, University of Pittsburgh School of Medicine, Pittsburgh, Pa. 15213*

Since the publication of the meta-analyses by Smith and Glass (1977; Smith, Glass & Miller 1980), several other analyses and reanalyses of psychotherapy outcome have appeared. The reanalysis of Prioleau, Murdock, and Brody addresses a central question, namely, Have the effects of therapy been shown to surpass those achieved by "placebo control" procedures? Their analysis, involving 32 investigations, is thoughtful and appropriately cautious. And of course the conclusions are highly provocative.

Several issues might be challenged regarding the particular details of this meta-analysis, such as the failure to distinguish placebo controls (where the procedure is known to be inert) from nonspecific treatment control procedures (where the active treatment components may be unknown but nonneutral), the fact that most evaluation strategies to investigate treatment outcome do not require the use of a placebo or nonspecific treatment control group (Kazdin 1980), the problems with combining studies across different types of patients and clinical problems, and others. My commentary focuses on one of the more general problems of meta-analysis, which is illustrated by this particular analysis.

The major problems with meta-analysis are not with the analysis itself, that is, the quantitative procedures (although issues are far from resolved here), but rather with the implicit and explicit decisions regarding what to include in the analysis. The authors began with 40 investigations for the analysis. Eight were discarded because effect sizes could not be estimated or because they were "so seriously flawed." The latter reason raises a critical question about meta-analysis, namely, what studies should be entered? For a moment, I would grant, sight unseen, that the studies excluded for the reasons stated were sufficiently flawed and did not warrant inclusion. In fact, I would argue further that probably very many more investigations of the original 40 should be excluded on methodological grounds. The criteria that the authors invoked for deciding what was seriously flawed are important. The criteria, if applied more stringently, would no doubt lead to exclusion of more studies. Which studies should be excluded, and on what grounds, is difficult to decide. Methodological criteria that should be invoked to decide whether a study is well controlled and provides a fair test of treatment are by no means resolved. Yet the implications for the varying criteria on the results of a meta-analysis are enormous.

An assumption of meta-analysis is that multiple studies may yield important information, even though the studies themselves are individually flawed. The rationale is that the different problems probably "cancel each other out," so to speak. Consistencies in the findings that emerge from the diversity of the problems of the individual studies probably reflect veridical and perhaps even robust conclusions. However, this rationale can be challenged. Many outcome studies suffer similar problems or different problems that operate in the same direction, that is, showing little treatment effects. Common problems include relatively small doses of treatment, little or no attempt to assess or ensure the integrity of treatment, little statistical power, the lack of well-validated measures or of long-term follow-up, and so on. Combining studies with these characteristics in a meta-analysis is mildly interesting to be sure but provides a base of unclear value in reaching substantive conclusions about treatment effects.

We do not need a meta-analysis of large numbers of studies with flaws that are undisclosed or buried by the analysis. We might profit greatly from a meta-analysis of investigations that were shown in advance to meet the highest standards of rigor and clinical relevance, but there would probably be little need for a meta-analysis of such investigations, given the small number in the sample. Yet the conclusions might be very different from what is obtained in the large unselected analysis.

Interest in the effects of psychotherapy is great and can be attributed to several converging influences and pressures. It is natural that new methodologies such as meta-analysis should be called upon to clarify the enormous amount of data and to resolve the surplus ambiguity that has characterized attempts to reach conclusions about treatment outcome. Meta-analysis provides a quantitative method of analyzing masses of data and reaching conclusions that would be difficult if not impossible to reach in a replicable fashion otherwise. Unfortunately, meta-analysis has in no way solved the basic problems that have faced psychotherapy researchers and the evaluation of the outcome literature. Fundamental issues entail such points as what standards need to be met before an investigation should be taken seriously, what outcome measures should be given greater credence than others, at what point in time (posttreatment or later follow-up) should the ultimate effects of treatment be evaluated, and so on. Answers to these questions are still undergoing debate. Because meta-analysis circumvents these issues, it is not likely to provide answers to the questions about treatment effects. The analysis can raise provocative questions in the scholarly way that Prioleau et al. have in their evaluation. Yet, whether the answers can be achieved through meta-analysis is not yet clear.

## Meta-analysis, measurement, and methodological problems in the study of psychotherapy

Paul Kline
*Department of Psychology, University of Exeter, Exeter EX4 4QG, Devon, England*

Prioleau, Murdock, and Brody have attempted to demonstrate by a meta-analysis of a subset of studies examined by Smith, Glass, and Miller (1980) and a qualitative analysis that placebo treatments of real patients are as effective as psychotherapy. There are certain points, however, which in my view, do much to render that conclusion dubious.

***Problems of meta-analysis.*** Meta-analysis does not seem a powerful procedure for the following reasons, some of which have been discussed at great length by numerous critics (e.g., Wilson 1982; Shapiro & Shapiro 1982).

a) No matter how elegant the statistical procedures of meta-analysis, poor research is still poor research. All depends, therefore, on the quality of the researchers in the meta-analysis.

b) A claimed advantage of meta-analysis is its objectivity (Smith et al. 1980). This, however, taken with point (a) above, is its fatal flaw. One properly conducted study cannot be outweighed as it must be in meta-analysis by errorful researches. A subjective element is essential.

c) As Cattell and Kline (1977) have argued in their study of factor-analytic procedures, it is essential to examine the results of studies that have no artefactual content. Meta-analysis and subjective analysis based upon appropriate technical methods gave quite different results.

Although, as Shapiro and Shapiro (1982) argue, there are other difficulties in meta-analysis, such as selection of studies to be included, and problems of the statistical significance of meta-statistics, the importance of these is small relative to the basic difficulty which I have discussed above.

Thus, to conclude this point, I should like to ignore the summary statistics and outcome measures and concentrate on the quality of the studies as it bears upon the conclusions drawn by the authors.

***Problems of measures of therapeutic outcome.*** This is another much discussed but nonetheless important point. In many studies the measures of outcome are inappropriate for the aims of the therapy, and certainly not comparable with other different outcome measures. I am not arguing that psychotherapeutic effects are too subtle to measure but that at present there are few satisfactory measuring instruments. The problem is practical, not logical. This simple practical difficulty renders many studies of little worth. Indeed I have argued (Kline 1981) that for psychodynamic techniques, outcome studies with our present level of measurement expertise are not useful and that process studies of the therapy are superior for evaluation.

For example, use of the MMPI (Minnesota Multiphasic Personality Inventory) and its derivatives is dubious. This scale measures resemblance to certain nosological groups. The psychological nature of the variables is unknown by virtue of its constructional method. The test/retest reliabilities are poor; thus the meaning of change scores is dubious. Furthermore, are the aims of techniques likely to be reflected by MMPI scale score changes? Similar objections can be raised about the use of projective tests (as regards validity and reliability) e.g. the Draw a Person and the apperception test. Again use of the Eysenck Personality Inventory which is claimed by its authors to reflect fundamental personality traits, seems a peculiarly unfortunate choice for a study of therapeutic change. Rating scales, too, are crude measures, of generally little validity as almost all elementary psychometric textbooks aver. All these relatively poor devices are cited in the paper by Prioleau et al. Little weight can be attached to any findings, positive or negative, with measures this crude. Contrast these with the subtlety of the evaluations of therapy by Malan and his colleagues (Malan 1959).

***Designs of experiments.*** Kiesler (1966) has pointed out the problems of designing experiments that adequately test the efficacy of therapy, stressing the need for a proper sample of therapists for each school and for a variety of patients, as well as the need for valid outcome measures. For practical reasons (including the unethical necessity to refuse treatment to patients for reasons of experimental design), few experiments reach these criteria. Consequently, again, few experiments can be taken seriously.

All this leads one to the inevitable conclusions that there are few researches sufficiently well executed and designed to be able to answer adequately the questions of therapeutic efficacy. Only those that do meet such criteria deserve consideration. Hence meta-analysis, unless it confines itself to studies that have been subjectively selected in the first place, has little part to play in the evaluation of psychotherapy.

There is a further problem with outcome studies, however, especially negative ones. Even if a therapy is shown to be ineffectual, we have no means of knowing that the therapy was well done, even when a large sample of therapists is used. Logically, failure of a therapy can be imputed to poor therapists.

This discussion merely underlines a tacit but fundamental difficulty in the study of psychotherapy. It is easy to demonstrate logically that there is no effect or that any effects could be attributable to other uncontrolled factors, but it is difficult to show positive effects. That is why experimental designs have to be so complex. This is essentially the burden of the paper by Prioleau et al. While I agree with the findings, I attach, by reason of the difficulties mentioned, little psychological significance to the results. Good experiments are awaited.

## Meta-analysis: We need better analysis

Brendan Maher

*Department of Psychology and Social Relations, Harvard University, Cambridge, Mass. 02138*

In the United States, psychotherapy is now a major source of income for a wide spectrum of persons described as mental health professionals. Psychiatrists, psychologists, social workers, counselors, paraprofessionals, graduate holders of the M.D., Ph.D., Ed.D., M.S.W., and most recently the Psy.D. have moved in increasing numbers into the practice of psychotherapy. Doctoral programs in psychology report a decline in the number of Ph.D. candidates in the research and academic fields of psychology – an understandable response to the tight job market in teaching and research – but a rise in the numbers seeking the Ph.D. in clinical psychology.

Payment for psychotherapy is not likely to be forthcoming from comparable increments in the budgets of mental health agencies staffed with salaried therapists. In fact, all of the evidence is that tax limitation statutes have operated to reduce the budgets of state agencies that normally provide such services, and that salaried mental health staff resources are being reduced rather than expanded. For many psychotherapists the economic key to this situation is the possibility of private practice, which, in turn, depends heavily upon third-party payments from insurance carriers.

This chain of events has led to several predictable consequences. One is intensive lobbying to mandate mental health coverage in medical insurance policies (see Maher 1981). Another is the development of aggressive "merchandising" of mental health services through media advertising, publicly visible *pro bono* programs, and the like (see, for example, Gist & Stolz 1982). Yet a third is the increase in the sometimes bitter fighting between holders of one credential versus the holders of others to see which shall have the major share of a limited therapy-income pie. Into this already tense situation there entered in 1981 the decision by the federal government that future reimbursements from federal funds for psychotherapy costs would be dependent upon hard evidence that psychotherapy is effective. Under the circumstances, any analysis that suggests that the costs of psychotherapy may not be justified is as welcome as a tax auditor at a business lunch. Prioleau, Murdock, and Brody must expect to be attacked with a vigor that will be relatively independent of the rationality of their conclusions.

Their conclusions are quite justified: If there are criticisms to be made of their paper it may be that they are too willing to accept some questionable assumptions made by Smith, Glass, and Miller (1980). The first of these, which they mention in passing, concerns the legitimacy of the meta-analysis. Taken as a body of subjects and data, the material that was meta-analyzed by Smith et al. is so diverse in every respect (subject age and other demographic variables, diagnosis, nature of treatment, dependent variables, and so forth) that the tactic of combining probabilities for analysis directed toward single conclusions seems to be most dubious. As the present authors fail to find positive results from their own meta-analysis, we may be inclined to overlook the tenuousness of the method that was adopted originally. Their doubts about meta-analysis might perhaps have been more forcefully expressed.

A second problem that arises with the original report by Smith et al. (1980) resides in the use of effect size as the criterion of interest. Surely the important clinical question is whether or not a patient has returned to adequate functioning. The knowledge that the mean of a group of patients now exceeds that of another group by some proportion of a standard deviation tells us very little unless we know whether these values have now moved into a range that has implications for actual behavioral effectiveness.

Yet a third problem with this kind of meta-analysis is that it attempts to answer a question so broad as to be practically meaningless. Were we to pose the question "Do physicians cure sickness?" we would want to know with some precision which doctors, doing what to which kinds of sickness. The range of psychotherapies covered in even the limited number of studies that met Prioleau et al.'s criteria is heterogeneous as to techniques, the class of clients served, the kind of therapist employed, and so forth. The fault lies with the Smith et al. (1980) analysis in the first place, of course, and with their assumption that such an approach could answer meaningful questions about psychotherapy.

Finally, we should note that only some of the studies cited seem to have employed pre- to posttreatment change scores as the dependent variable. Where this is not the case, there seems to be no way to tell whether the treatment and placebo groups improved together, remained essentially unchanged, or deteriorated together. All that we know is that whatever happened, happened in about the same degree to both. It would have been helpful to have indicated in Table 1 which studies did employ change scores, and the mean direction of the changes.

But surely the message conveyed by the work of Prioleau et al. must be that we need more than meta-analytic attempts to salvage meaning from a plethora of published studies that have neither techniques, clients, nor therapists in common. The matter has been well put by London and Klerman (1982): "For a psychotherapy to meet the efficacy criterion, it must be measurably equal to or better than other treatments and better than non-treatment by standards that permit independent observers, using the same methods, to disconfirm the results of the original investigator. When others have tried and failed to disconfirm, then the efficacy of the treatment is established. Until then, it is not." (p. 715)

Prioleau et al. deserve our gratitude for their work. It should put to rest recent sanguine claims that the benefits of psychotherapy have already been demonstrated (e.g., Parloff 1982) and should direct us to the increasingly urgent task of establishing reliable and replicated findings obtained under conditions of scientific objectivity.

## Improving meta-analytic procedures for assessing the effects of psychotherapy versus placebo

Robert Rosenthal

*Department of Psychology and Social Relations, Harvard University, Cambridge, Mass. 02138*

The meta-analytic enterprise is neither mechanical nor perfectly reliable. The process is made up of a series of judgments just as is the case in any other data analysis. For that reason meta-analyses are just as much in need of replication as are the individual studies that make up the meta-analysis. The valuable meta-analysis by Smith, Glass, and Miller (1980) has now been replicated at least twice; earlier by Landman and Dawes (1982) and now by Prioleau, Murdock, and Brody.

In various places in their paper, most notably in their abstract, Prioleau et al. give the impression of having obtained results that were inconsistent with those of Smith et al. (1980). That impression is misleading. In fact, their estimate of the effect of psychotherapy relative to placebo treatments is admitted to be in exact agreement with that reported by Smith et al. (1980), $d = .42$ in both cases. What then, does the meta-analysis by Prioleau et al. add to our understanding of the effects of psychotherapy versus placebo treatment? In principle, it might have added considerably to our understanding; in fact it did not. It did not because it employed meta-analytic procedures that used far too little of the information contained in the 32 studies summarized. No conclusion can reasonably be drawn about the moderating effects of any variable examined by Prioleau et al.

The rest of this commentary reexamines the data examined by Prioleau et al. employing some useful procedures to help us understand the difference between psychotherapy effects and placebo effects. There is no sense in which the following analysis is intended to be definitive. Other meta-analysts interested in different issues would be expected to address different questions employing different procedures. However, use of the procedures summarized here will ordinarily yield more infor-

Table 1 (Rosenthal). *Summary of statistics employed in meta-analysis of psychotherapy effects*

| | Students | | | Patients | | | |
|---|---|---|---|---|---|---|---|
| | Elementary school | Secondary school | College level | Psychological placebo | Medical placebo | All results combined | Combined results reported in target article |
| 1. Number of studies | 6 | 6 | 10 | 6 | 4 | 32 | 32 |
| 2. Total number of persons | 363 | 306 | 319 | 236 | 216 | 1440 | Not reported |
| 3. Weighted mean $d$ ($\bar{d}$) | .17 | .06 | .53 | .44 | .02 | .24[a] | 42 (unweighted) |
| 4. Z for mean $d$ | 1.58 | 0.52 | 4.56 | 3.26 | 0.15 | 4.50[b] | Not tested |
| 5. $p$ for Z above[c] | .06 | .30 | .001 | .001 | .44 | .001[d] | Not reported |
| 6. $\chi^2$ for heterogeneity of $d$'s | 44.4 | 1.94 | 13.9 | 19.4 | 1.02 | 80.66 | Not tested |
| 7. d.f. for $\chi^2$ above | 5 | 5 | 9 | 5 | 3 | 27 | Not reported |
| 8. $p$ for $\chi^2$ above | .001 | .90 | .15 | .002 | .80 | .001 | Not reported |
| 9. Z for linear contrast | −4.54 | −0.57 | −0.13 | 2.35 | −0.91 | −1.70 | Not tested by contrasts |
| 10. $p$ for Z above[c] | .001[d] | .28 | .45 | .009 | .18 | .04 | Reported as n.s. |
| 11. $r$ based on linear contrast | −.24 | −.03 | −.01 | .16 | −.06 | −.05[e] | −.21 |

[a]The five $d$'s upon which this weighted mean is based differ significantly among themselves; $\chi^2(4) \cong 13.87$, $p < .01$. [b]Computed as $\Sigma Z/5$ (Rosenthal 1978). [c]One tailed. [d]More precisely, $p < .000005$. [e]The five $r$'s upon which this weighted mean $r$ is based differ significantly among themselves; $\chi^2(4) \cong 22.75$, $p < .001$.

mative results than will those employed by Prioleau et al. A general framework for the conceptualization of these procedures has been provided elsewhere (Rosenthal 1983).

Table 1 (Rosenthal) summarizes the results of the present meta-analysis. The 32 studies were divided into five groups. The first three groups were entirely comprised of students divided on the basis of age level into elementary, secondary, and college level. The last two groups were entirely comprised of patients divided on the basis of the type of placebo employed – psychological vs. medical. The psychological placebo patients (as well as all the student groups) were those who received some form of placebo that could have been viewed by patients as in some sense psychological. The medical placebo patients were those who received only a pill placebo, that is, they received only a "medical" placebo treatment.

The first two rows of Table 1 (Rosenthal) show the number of studies summarized and the total number of persons whose data entered into a determination of the average size of the effect ($d$). The third row gives the mean $d$ for each group, the fourth and fifth rows give the standard normal deviate (Z) and the $p$ level associated with each mean $d$. The college students and the patients who (like the college students) were given psychological placebos both showed substantial benefits of psychotherapy relative to placebo controls and these differences were significant at $p$ well less than .001. The grand mean effect size of .24 ($p < .000005$, one tailed) was smaller than that obtained by Prioleau et al. and by Smith et al. because it was computed with weighting inversely as the variance of $d$. (All required computational formulas are given in Rosenthal & Rubin 1978; 1982a; 1982b and the interpretation of effect sizes is discussed in 1982c.)

Rows 6, 7, and 8 of Table 1 (Rosenthal) give the results of tests of heterogeneity of effect sizes, that is, tests of whether the $d$'s in each set of studies differ significantly among themselves. Studies of elementary school children and of patients receiving psychological placebo yielded $d$'s that were significantly heterogeneous (see Hedges 1982 and Rosenthal & Rubin 1982a; 1982b for discussions of such tests).

Lines 9, 10, and 11 address the question of the relationship between the size of the study and the size of the $d$. The Z's and $p$'s for linear contrasts show that among elementary school children larger $d$'s were found in smaller studies ($p < .000005$; $r = -.24$) but among patients receiving psychological placebos larger $d$'s were found in larger studies ($p = .009$; $r = .16$). Thus, although for all studies combined, larger $d$'s are associated with smaller studies, there are statistically significant reversals of this overall relationship.

Table 2 (Rosenthal) shows that the mean $d$'s of the five groups examined can be compared meaningfully within the framework of a set of four contrasts. The first contrast shows that there is little difference between students and patients in the degree to which psychotherapy is more effective than placebo. The second contrast shows that with increasing age of students greater $d$'s are obtained. The third contrast shows that the average of the elementary and college student groups yields a larger $d$ than does the group of secondary students. In interpreting these contrasts in age we should note that age is likely to be confounded here with such variables as IQ, type of treatment, type of placebo control, and so forth. The fourth contrast shows that psychotherapy is more effective relative to psychological than to medical placebo controls. Perhaps pill placebos are so effective that it is difficult for psychotherapy to be superior to them.

Table 2 (Rosenthal). *Contrasts among five groups of studies*

| Contrast | Z | $p$ (one-tailed) |
|---|---|---|
| 1. Students vs. patients | .20 | .42 |
| 2. Linear trend in age of students | 2.27 | .012 |
| 3. Quadratic trend in age of students | 2.08 | .019 |
| 4. Psychological vs. medical placebo given to patients | 2.18 | .015 |

To address this last question, to help understand the significant linear and quadratic contrasts in age, the meaning of the sometimes positive, sometimes negative correlation between $d$ and N, and the significant heterogeneity of $d$'s found among studies of elementary school children and studies of patients given psychological placebo, additional studies will be required.

This commentary has tried to illustrate how the systematic application of improved meta-analytic procedures can lead to firmer inferences about a domain of research. At the same time, however, it should be clear that meta-analyses need not close off further research in an area; indeed they may be employed to help us formulate more clearly just what that research should be.

ACKNOWLEDGMENT

Preparation of this paper was facilitated by a grant from the National Science Foundation.

## Outcome research: Isn't sauce for the goose sauce for the gander?

Ted L. Rosenthal

*Department of Psychiatry, University of Tennessee College of Medicine and Memphis Mental Health Institute, Memphis, Tenn. 38105*

Since I share Prioleau, Murdock, and Brody's concerns both about lack of evidence confirming evocative psychotherapy's value, and about the conclusions and procedural hygiene of the Smith, Glass, and Miller (1980) "omnibus gristmill" analysis of heterogeneous treatment outcomes, let us move promptly to our areas of mild-to-severe divergence.

At the "mild" end of dissent are a few questions of methodology. Since Prioleau et al.'s pooling procedure departed from that of Smith et al., should not an estimate of effect size based on the original methodology be included for comparative purposes? Likewise, they cite A. Bergin's view (Bergin & Lambert 1978) that evocative psychotherapy helps some patients but hampers others. For the relevant studies, Bergin's argument could be expressed as a significant increase in variance between the pre- and posttreatment scores and hence could be addressed empirically; this was not done. Understandably, they make some assumptions open to debate, such as assigning .5 as the value of correlation between pre- and posttreatment scores when standard deviations could not be extracted from the published data. They defend their "arbitrary correction" because: (a) The .5 value was in the upper range of correlations used by Smith et al. This in effect is to adopt as a yardstick a number drawn from a study whose procedures they directly question. (b) They also assert that a correction was used "in less than one-third of our calculations." Yet one-third of some 32 studies should be enough to test if the ten or so "corrected" studies differed notably from the noncorrected remainder. Further, it would be of heuristic worth in such a comparison to assess *several* arbitrary values, for example, .3, .5, and .7, to determine the impact of varying the .5 "standard" arbitrarily imposed. Each foregoing point was explicitly raised in the reviewer/editorial feedback sent to Prioleau et al. after their initial submission. Hence, one feels some chagrin that none of these refinements was applied to the data in their revised manuscript.

Moderate divergence stems from their assignment to the "placebo" category (see Table 1) of such techniques as those used by Gillan and Rachman (1974) involving muscle relaxation and hierarchy construction and by Grande (1975), who used taped relaxation training. Such methods have been shown clinically useful in much past research. Prioleau et al.'s accepting any techniques given to clients "as a placebo treatment [if] so described by an investigator" appears an unwarranted abdication of scholarly judgment in what is – at base – a *methodological* analysis.

Severe divergence begins with their passively accepting *any* theoretical rationale which makes a case that some technique is a "placebo," that is, has nonspecific impact, and with their failure to recognize or discuss that the same complexities exist in the behavior therapy literature. This in effect creates a scholarly double standard between their (and my) preferred treatment approaches – which receive privileged status by omission – versus evocative psychotherapies to which their analysis is confined. Despite heroic rhetoric to the contrary, the facts disclose that "nonspecific" effects, alien to most behavioristic formulations, are rampant throughout the treatment literature. Thus, in a review of over 100 (largely behavioral) therapy studies, it emerged that *nonspecific* variants – which the investigators expressly intended as inert, placebo control conditions – quite often yielded outcomes as good as, or trivially weaker than, the putatively "real" or "active" treatment (see Rosenthal 1980). Moreover, hypotheses drawn from "conditioning" and other behavioristic theories were repeatedly refuted. In some cases, the response-reinforcement contingency was destroyed, abridged, or inverted, yet outcome results were much the same. In some cases, to create "placebo" conditions, severe deformations of structural, sequential, or hierarchical organization – usually construed as essential desiderata – were introduced, but such permutations often led to equivalent outcomes. There were even some cases in which methods that on classical theoretical grounds should have raised the severity of the clinical problem, but these variations, designed to contrast with the presumptive treatment of choice, worked just as well. Here are a few concrete examples. Diverse conditioning, "non-specific," and nonconditioning techniques led to comparable reductions in smoking (Russell, Armstrong & Patel 1976). Traditional desensitization fared no better to relieve speech anxiety than did two "placebo" variants (equally plausible to clients); in one, clients were shocked while imagining hierarchy scenes; in the other, a "free association" method brought equivalent success (Kirsch & Henry 1977; and see McGlynn, Kinjo & Doherty 1978 for similar results). In an analogue study with dysphoric clients, guidance to "punish" blue moods by performing versus imagining tedious events, versus performing versus imagining positive events when a blue mood struck, all created comparable progress. Further, a distraction condition in which matched clients were taught to rehearse neutral content at fixed (noncontingent) times daily led to substantial progress no weaker than from the contingent punishment and reward plans (above) also studied, even though this distraction method was judged less credible by its recipients both before and after treatment (Catanese, Rosenthal & Kelley 1979). Perhaps the best illustration of the points at issue emerges from a study to reduce dysphoria conducted by Zeiss, Lewinsohn, and Munoz (1979). They made yeomanlike efforts to separate treatment techniques and the dependent measures expected to reflect the differential impact of each therapy. They compared guidance to increase positive activities, versus cognitive self-management (e.g. favorable self-talk and thought stopping), versus training in such interpersonal skills as assertiveness. Yet each "distinctive" treatment yielded equivalent gains *across* the range of specific dependent measures targeted for the respective interventions to best modify.

In light of (much) such data, one can conclude that behavioral therapies often have positive impact but their impact violates the theoretical premises of most behaviorist theories, as shown by the frequent success of treatments intended to be "inert" or "placebo" conditions. The foregoing issues are nowhere addressed in the Prioleau et al. target article. Hence, while agreeing with what they conclude, I am dismayed by what they leave unsaid. If an analogy between science and carpentry is permissible, their article is not a well-crafted cabinet.

## Psychotherapy and placebo: "Sticks and stones will break my bones, but can words never harm me?"

Thomas A. Sebeok

*Research Center for Language and Semiotic Studies, Indiana University, Bloomington, Ind. 47402*

Since what is conventionally called "healing" has been throughout history, and is in all cultures, accompanied by a stream of verbal patter and/or the giving, and the giving off, of strings of nonverbal signs, it is intuitively clear – as well as being borne out by limited empirical data – that *all* acts of induced medical adjustment depend in some degree on the placebo effect for their success. Psychotherapy – like all other varieties of faith-healing – crucially builds upon the so-called patient/doctor relationship, a subtle two-way semiotic transaction the outcome of which is determined, in Miller's (1978:80) unaffected phrasing, "not simply by what the doctor hands out, but by the confidence and respect the patient gives him in return."

The very word *psychotherapy* is misleading in its dualistic implications and opposition to the concept of *somatotherapy* (for which an equivalent term exists in, for instance, German clinical practice). The administration of a drug or, say, surgical intervention cannot but be addressed to the patient-as-a-whole, regardless of whether the doctor's implicit or explicit philosophical outlook is monistic or dualistic.

Consequently, the questions considered in Prioleau et al.'s target article can have but one answer, which the authors put forward with such timidity. They are also overly modest: Their two propositions surely have more than speculative status; they are of heuristic value.

The fundamental problems begin, however, where they leave off: precisely how does the placebo effect work (not only in human clinical contexts but also – perhaps surprisingly – in veterinary medicine; see, for example, Chertok & Fontaine 1963)? For answers, one must look to the rapidly developing field of brain electrochemistry, as I have set forth elsewhere (Sebeok 1981: Chap. 10).

## Refinement, precision, and representativeness in meta-analysis

David A. Shapiro

*MRC/SSRC Social and Applied Psychology Unit, The University, Sheffield S10 2TN England*

Prioleau, Murdock, and Brody are not alone in their concern with the pitfalls of excessive generalization in meta-analysis (Cook & Leviton 1980; Shapiro & Shapiro 1982a, 1982d; Strube & Hartmann 1982; Wilson 1982). Their analysis shows that Smith, Glass, and Miller's (1980) average effect size of .85 is not itself sufficient evidence to refute Eysenck's (1966) negative view of verbal psychotherapy. On the other hand, the target article's own conclusions are unduly pessimistic, since its own analysis is incomplete.

Comparable data were obtained by Shapiro and Shapiro (1982c). In a meta-analysis of 143 studies comparing two or more treatments with a control group, only 21 of which were also included by Smith et al. (1980), we found an effect size of only .40 for 16 studies of verbal therapies, considerably below the overall mean of .98. Only three studies compared verbal therapies with a placebo condition, yielding effect sizes of −.26, .51, and 1.15. Despite this convergence of findings, several questions remain.

First, does the target article's selection of 32 studies involving placebo control groups in fact yield a more dependable estimate of the effects of verbal psychotherapy than Smith et al.'s (1980) mean of .85 obtained over 597 effects? Whilst the placebo control group does in principle yield a more conservative test than does the no-treatment control group, it is noteworthy that Shapiro and Shapiro (1982c) found no difference between 95 groups evaluated with reference to a placebo control (mean effect size = .96) and 319 groups evaluated with reference to a no-treatment condition (mean effect size = .99). Furthermore, comparative analysis of treatment effect sizes obtained with reference to placebo control conditions of different types showed little evidence that more rigorous controls attenuated the apparent effects of treatment relative to less rigorous controls (Shapiro & Shapiro 1982b). In these circumstances, the loss of precision resulting from the reduced data set in the target article may not be outweighed by any reduction of bias resulting from excluding the large proportion of the Smith et al. (1980) data without placebo control groups.

Second, to what extent are verbal therapies introduced as "straw men" or quasicontrol conditions not expected to yield strong effects (Smith et al. 1980: 119f)? This question gains force from Smith et al.'s obtained association between the apparent allegiance of the researcher and effect size, and the evidence implicating reactivity of measurement and the researcher's knowledge of treatment assignment as influences upon obtained effect size (Smigh et al. 1980; Shapiro & Shapiro 1982b).

Third, to what extent might the conclusions reached by the target article be undermined by thorough consideration of covariation among the several correlates of effect size? The clearest comparison between treatments comes from studies in which the treatments in question appear side by side in the same experiment, so that such concomitant variations are minimised (Smith et al. 1980). The modest differences between treatment types reported by both meta-analytic and traditional reviewers may be due at least in part to the confounding influence of other variables. Consistent with this, Shapiro and Shapiro (1982c) found, for example, that verbal therapies yielded a mean outcome .53 of a standard deviation inferior to other (behavioral and cognitive) treatments over 13 such comparisons. The negative correlation between treatment duration and effect size reported in the target article could be due to other factors, such as problem severity, systematically related to treatment duration via clinically appropriate research designs. Thus Shapiro and Shapiro (1982c) found a negative correlation between therapist experience and effect size, which was abolished by statistical control for the nature of the target problem via multiple regression analysis.

Fourth, does the target article overplay the apparent superiority of "quasibehavioristic" methods in its data? For example, effect sizes of −.01 and .39 for rational-emotive therapy go unmentioned in the text, and the fact that the DiLoreto (1971) effect size is averaged over rational-emotive and client-centered therapy is overlooked.

Fifth, is the size and quality of the data base available for verbal therapies adequate to justify any clear conclusion as to their effectiveness? Shapiro and Shapiro (1982b) concur with Prioleau et al. in bemoaning unrepresentative samples and inadequate attention to the influence of nonspecific and demand effects. However, the target article's discussion of credibility overlooks the possibility that some placebo conditions, such as relaxation, may be surprisingly credible (Shapiro 1981). The 32 studies reviewed in the target article are too few, too heterogeneous, and too unrepresentative to permit more than the most tenuously speculative evaluation of the article's "proposition 2."

To raise these questions is not to deny, however, the force of Prioleau et al.'s demonstration that a plausibly selected, albeit small, subset of the Smith et al. (1980) data does not support the latter authors' finding of a .85 effect size for verbal therapies. Taken together with Shapiro and Shapiro's (1982c) findings, these data suggest that the case for verbal therapies from controlled outcome research remains quite weak, although any apparent superiority of behavioral and cognitive methods is largely confined to unrepresentative studies of students. The target article thus provides a timely reminder of the pressing need for clinically realistic comparative evaluation of verbal and behavioral therapies.

## Psychotherapy outcome research and Parloff's pony

Michael Shepherd
*Institute of Psychiatry, London SE5 8AF, England*

The long honeymoon between psychotherapeutics and American public opinion appears to have ended at last (*Time* 1979). In both professional and nonprofessional circles, an evident failure to justify therapeutic claims, the arrival of other forms of treatment, and the growing insistence on cost-benefit analysis have all contributed to a state of disillusionment and a demand for long-overdue studies of evaluation. Unfortunately, as Morris Parloff has pointed out, psychotherapy research "has not yet been designed or conducted in a manner that can provide truly responsive answers . . . The best I can say after years of sniffing about in the morass of outcome research literature is that in my optimistic moods I am confident that there's a pony in there somewhere" (Parloff 1979).

Parloff's pony has eluded its pursuers for so long that it is surely time to wonder whether the quest should not be conducted with a little horse sense. A host of studies have now been conducted which, with all their imperfections, have made it clear that (a) any advantage accruing from psychotherapy is small at best; (2) the differences between the effects of different forms of psychotherapy are negligible; and (3) psychotherapeutic intervention is capable of doing harm.

It is apparent, therefore, that the continued interest in psychotherapeutics can hardly be due to its efficacy. To what, then, can it be attributed? A satisfactory answer to this question demands a broader conceptual framework of inquiry than that of the medico-scientific investigator (Shepherd 1979). Indeed it can be maintained that a stubborn preoccupation with the rigours of the "hard" methodology and statistical analysis may prove counterproductive, partly because the material does not lend itself readily to such examination and partly because such activity tends to deflect attention from the "softer" issues which are assuming importance. It is not always recognized that most of the procedures labeled "psychotherapeutic" represent merely the professional end of a pastoral and thaumaturgic spectrum, on which status, power, and financial reward are all represented. Since psychotherapeutic theory and practice are heavily culture-dependent, an example from another country may be taken to illustrate the matter.

The case in question comes from the United Kingdom and concerns what its founder has called "the first thoroughly validated psychotherapy," namely scientology or dianetics. Validated or not, the activities of some scientologists in Britain were dubious enough to have prompted an official inquiry by a distinguished lawyer, Sir John Foster (1971). In his report, a model of impartiality, Foster was moved to consider some of the legal and social implications of psychotherapeutic practice and went so far as to recommend that "psychotherapy (in the general sense of the treatment, for fee or reward, of illnesses, complaints or problems by psychological means) should be organised as a restricted profession open only to those who undergo an appropriate training and are willing to adhere to a . . . proper code of ethics, and that the necessary legislation should be drafted and presented to Parliament as soon as possible."

A decade after the publication of the Foster report there is still no consensus on the registration of psychotherapy in Britain (Shepherd 1980). In the United States Hogan has devoted four scholarly volumes to the problems of regulating a range of activities which, he argues, legitimately includes such exotic variants as bioenergetics, est, encounter groups, psychodrama, and life-planning laboratories (Hogan 1980). The bioethics of many of these procedures and their place in health-care systems supported by the taxpayer remain controversial and unresolved topics.

In this disputed area carefully conducted studies like that of Priodeau, Murdock, and Brody are indispensable but insufficient. As the Foster report observes, a sizable population is at risk – "the weak, the insecure, the nervous, the lonely, the inadequate and the depressed whose desperation so often is such that they are willing to do and pay anything for some improvement of their condition." For such people some form of placebo may be indispensable. Perhaps the objective should be to make it as simple, as cheap, and as innocuous as possible. Parloff's pony could turn out to be grazing in the submarine meadows of Achaea, with Poseidon's horses.

## Therapeutic effectiveness: What domain is being studied?

Donald P. Spence
*UMDNJ–Rutgers Medical School, Piscataway, N.J. 08854*

Although the title of Prioleau, Murdock, and Brody's report promises information about the relative effectiveness of psychotherapy as a treatment modality, a closer examination of the studies represented raises questions about just what is being studied. One-half of the 32 studies appear to represent psychotherapy as the term is generally understood; the others represent a range of treatments from problem-solving training to social learning to rational emotive therapy. The first two categories can hardly be called psychotherapy by even the loosest definition, and the third represents a deviant subcategory at best. A second cause for concern results from a closer look at contact hours. Not one of the 32 studies lasted more than 36 sessions, all but three lasted no more than 20 sessions, and the average length of treatment was 9.8 sessions. (One "psychotherapy" lasted .75 sessions.) Whatever is being studied, it is clearly short-term and time-oriented, and some reference to this fact belongs in the title.

A second problem has to do with the use of the placebo control. Prioleau et al. admit the obvious variability of placebo treatments, and this defect is hardly contradicted by their statement that variability may also exist in wait-list controls. Not only is one defect not canceled by avoiding a second, but it could be argued that wait-list variability may well be the lesser of the two evils, not the greater. Because of the variety of experimental contexts being studied, the reader is never clear how persuasive the placebo control was; some procedures, in the hands of some experimenters, may have more face validity than others, and the effectiveness of the control obviously varies with its persuasiveness. Nor is any attention paid to the distribution of placebo-prone subjects in the various studies, an important contributor to any placebo effect.

In choosing to average across all dependent variables in each study, Prioleau et al. are taking the position that different response measures are equally significant. Yet this is clearly not the case. The measure of self-esteem of West (1969) is probably more relevant to assessing therapeutic effect than the WISC; the self-report of anxiety in Paul (1964) is more clinically relevant than pulse rate. Furthermore, in each of these examples we find that the clinically relevant measures showed higher effect sizes than the less relevant, suggesting that some way of weighting the measures by relevance might have significantly affected the outcome of the study. It also follows that a single response measure with low or negative effect size will depress the average of all measures, and if the low scores are not particularly relevant to studying therapeutic effect, they introduce spurious noise into the analysis.

Finally, it is important to call attention to Prioleau et al.'s conclusion that "On the basis of the available data we see no reason to believe that subsequent research using better research procedures and investigating other types of therapy administered to other types of patients will yield clear-cut indications that psychotherapy is more beneficial than placebo treatment." On the contrary, there would be *every* reason to expect different findings if a more representative set of studies were examined; the studies under discussion can hardly be called a sample of the domain of the clinical activity currently called psychotherapy. A new study is urgently needed to set the record straight.

## Limitations of meta-analysis and the lack of evidence that psychotherapy works

G. Terence Wilson
*Graduate School of Applied and Professional Psychology, Rutgers University, Piscataway, N.J. 08854*

Prioleau, Murdock, and Brody are to be commended for evaluating the effects of psychotherapy compared to placebo treatments. Without acceptable controls for "nonspecific" determinants of change (something that is *presumably* achieved using a placebo treatment), we cannot know whether specific treatment methods contribute to observed therapeutic change. There are problems, however, with the way in which Prioleau et al. chose to make this comparison.

By resorting to the technique of meta-analysis, and, more particularly, the meta-analysis conducted by Smith, Glass, and Miller (1980), Prioleau et al. immediately encounter a number of serious problems that detract from their evaluation. Meta-analysis, it is argued, is not only a quantitative method for summarizing and evaluating a large mass of independent studies, but also more comprehensive than other methods of reviewing diverse studies. Unfortunately, Prioleau et al. fail to mention, let alone seriously discuss, detailed critiques of meta-analysis as a means of evaluating the effects of psychological therapies in general, and the Smith et al. study in particular (Kazdin & Wilson 1978; Rachman & Wilson 1980; Wilson & Rachman 1982).

There is broad agreement that the question that should guide outcome research on the psychological therapies is "What method is most effective for what problems in which patients, on which measures, at what cost?" By indiscriminately lumping together different therapies, problems, patient populations, and measures, the Smith et al. meta-analysis fails even to begin to answer the real question. Prioleau et al. voice some concern about this most basic problem in their target article, but they fail to take seriously its full implications. Their solution is to use only a subset of the Smith et al. data base. Does this circumvent the problems with the latter, such as the massive omission of relevant studies, the inclusion of any study regardless of quality, and the assignment of equal weight to methodologically strong and weak studies?

By restricting themselves to psychotherapy, Prioleau et al.

avoid most of the distortions due to omission of studies in the Smith et al. study. It was particularly behavior therapy that was so misleadingly represented there. By selecting studies that contained a placebo control treatment Prioleau et al. eliminate many of the uninterpretable studies that obscured Smith et al.'s analysis. Still, it is noteworthy that they discarded fully 25% of the 40 studies containing both psychotherapy and placebo treatments because they were so seriously flawed. Exactly what criteria Prioleau et al. used in making this cut are unknown, although most of the remaining studies can still be criticized on methodological grounds. As in previous meta-analyses of the psychotherapy literature, Prioleau et al. do not address the consequences of attaching equal weight to good and bad studies. Unlike Smith et al., who used individual dependent variables as their unit of analysis, thereby disproportionately weighting different studies, Prioleau et al. used each study as the unit of analysis by deriving a mean effect size for each study. A substantial body of theory and research exists demonstrating the importance of analyzing the effects of comparative treatments across multiple individual measures of outcome (e.g., Rachman & Wilson 1980). Different treatments may have different effects on different measures. In the Paul (1964) study, for example, estimating a single index of outcome from the specific behavioral observations, physiological measures and questionnaires assessing state and trait anxiety obliterates meaningful patterns of outcome.

Meta-analyses have fallen afoul of the therapy uniformity myth (Kiesler 1966) in combining different treatments. Prioleau et al. are aware of the dangers inherent in evaluating so general a category as "psychotherapy," and issue the appropriate caveats. Similarly, they do a workmanlike job with their analysis of "placebo treatments." Devising and interpreting placebo treatments in therapy outcome research is difficult (Kazdin 1980). In all the placebo treatments in Table 1 we have to *assume* that they were successful in equating for the common "nonspecific" influences that can critically affect the interpretation of a technique's specific therapeutic impact. In superior research strategies that are now used in behavior therapy, independent evaluations of the adequacy of the placebo treatment, such as the extent to which it engenders expectations of improvement comparable to those of the active treatment, enhance internal validity and allow more refined judgements of a treatment's efficacy. The issue comes to the fore only when the active treatment is superior to the placebo treatment. As their study and other evaluations (Rachman & Wilson 1980) show, as yet there has been no acceptable demonstration of an advantage in nonbehavioral psychotherapy over placebo treatment. If psychotherapy is to be shown to be more effective than placebo treatment, the latter will have to meet the stringent standards of methodological control that have already been incorporated in behavior therapy.

Given the sorry state of research on psychotherapy outcome as a whole, the rare study that is well-controlled assumes added significance. Such is the recent study by Strupp and Hadley (1979), which neither Smith et al. nor Prioleau et al. consider. Patients suffering primarily from depression and anxiety reactions (with borderline personalities common) were treated either by highly experienced psychoanalytically oriented psychotherapists or seen for the same period of time by college professors chosen for their ability to form understanding relationships with students (the control condition). There were no differences between the two treatments across a variety of different outcome measures.

Whereas I take issue with Prioleau et al.'s reanalysis of Smith et al.'s data, their conclusion is consistent with Rachman and Wilson's review, which, I submit, provides a more discriminating evaluation than available meta-analyses. In the ultimate analysis, the questions of whether psychotherapy is effective, and whether alternative therapies have differential effects will be answered not on the basis of rehashes of flawed research or from attempts at statistical alchemy of existing inadequacies as exemplified by Smith et al.'s analysis. Well-controlled studies that assess the specific effects of well-defined treatments on multiple measures are what the field needs.

# Author's Response

## Where are the emperor's clothes?

Nathan Brody
*Department of Psychology, Wesleyan University, Middletown, Conn. 06457*

In this reply to our critics I will discuss four different issues. The issues are: (1) the adequacy of the existing data base for inferences about the effectiveness of psychotherapy relative to placebo treatment; (2) the use of placebos in psychotherapy outcome research; (3) the place of meta-analysis in psychotherapy outcome research and the particular methods used by us in our target article; (4) some new theoretical speculations about psychotherapy.

**Glass, Smith & Miller** fault us for failing to include newer studies that would have provided us with a more complete sample of studies on which to base our conclusions. We wanted to reanalyze the data reported by Smith, Glass, and Miller (1980) on which their conclusions rested. Smith et al's assertion seems to imply that there are several studies extant in the more recent literature that would contradict our analysis. As we made clear in the introduction to our paper, we were led to reanalyze the studies reported by Smith et al. in an effort to discover whether their analyses included studies that contradicted the findings reported by Brill et al. (1964), who found that the benefits of psychotherapy did not exceed those of a minimal placebo treatment (a pill placebo) for neurotic patients who seek therapeutic services. Contrary to the claims of Glass et al. in their critique of our study, studies comparing psychotherapy to placebo with neurotic outpatients are not abundant. I, too, spent a few hours reading the titles of all of the articles published since 1978 in the *Journal of Counseling Psychology* and the *Journal of Consulting and Clinical Psychology*. I do not claim that this haphazard procedure provides a perfectly reliable literature search. However, I can report that I found no relevant studies to add to the crucial category of psychotherapy–placebo comparisons for neurotic outpatients. We may be faulted for exceeding the boundary of our literature by including one study that we considered relevant to our most crucial category of studies.

The Comas-Diáz (1981) study that is cited in opposition to McLean & Hakstian (1979) is not relevant to our discussion since the control group is not a placebo control but rather a wait-list control. **Andrews** too implies that our conclusions are flawed because we missed relevant studies. In this connection he cites the two reports of the study by Weissman et al. (1979, 1981). Patients assigned to their wait-list control were told that they could obtain treatment if they believed that it was necessary. However, they were not led to believe that they were the recipients of an active treatment program. The Weissman et al. study is *not* a placebo-controlled study. Andrews also suggests that the comparisons between the psychotherapy and placebo-controlled conditions in the Gillan and Rachman (1974) and the McLean and Hakstian (1979) articles are somehow not legitimate because psychotherapy was used as a control for behavior-therapy treatments. I find this assertion puzzling. In these studies neurotic outpatients were randomly assigned to different treatment conditions, including a condition in which they received treatment from experienced professionals who were described as believing in both instances that psychotherapy was the treatment of choice for depression and multiphobia, respectively. Does Andrews mean to discard all studies of psychotherapy in which it is compared with any other form of treatment, since in such instances the psychotherapy treatment is a control condition?

**Spence** also argues that our conclusions are flawed by our selection of a body of studies that are of dubious clinical relevance. We agree that most of these studies are not relevant. That is why we described our studies and indicated that the class that we relied on in forming our conclusion consisted of those that appeared to us to be clinically relevant – namely the studies of psychotherapy provided to neurotic outpatients – a procedure also followed by **Andrews** in his reanalysis of the studies published by Smith et al. (1980). The comments of Spence, Andrews, and **Glass et al.** suggest that there is a body of literature demonstrating that psychotherapy for neurotic outpatients is more effective than placebo treatments. Moreover, Glass et al. assert the studies we cite as relevant to this issue are, for reasons that they do not indicate, equivocal and mixed. I invite the reader to examine the results of the Brill et al. (1964), Gillan and Rachman (1974), and McLean and Hakstian (1979) studies. I find no mixed or equivocal evidence in favor of psychotherapy in these studies – only negative evidence.

In fact, I would like to make a public offer to these psychologists or to any readers of this *BBS* treatment. I will publicly retract, with the editor's permission, any conclusions reached in our target article, if anyone can provide me with evidence of a study finding that psychotherapy leads to benefits that exceed those obtained for placebo treatments for neurotic outpatients. Ideally, I would like evidence that the benefits are maintained through time. However, failing evidence of maintained benefits, I will be pleased to issue a partial retraction covering the effects of psychotherapy relative to placebos at the end of treatment.

Several of our critics question the cogency of our conclusions given alternative interpretations of the set of studies analyzed originally by Smith et al. (1980). For example, **Garfield** implies that other analyses of the same set of studies have reached alternative conclusions. The differences are, in my judgment, more apparent than real. The Andrews and Harvey (1981) reanalysis excludes analogue studies that do not use neurotic patients. They report a mean effect size for verbal psychotherapies of .74 and a mean effect size of .35 for the class of developmental therapies encompassing client-centered and counseling therapies. The weighted average effect size for all forms of psychotherapy for neurotic patients is .595 as against a mean placebo effect size of .55. Thus, Andrews and Harvey find that, averaged over all studies and measures dealing with neurotic subjects, psychotherapy is equivalent to placebo treatments. Of course, one could split the psychotherapy studies and make a separate comparison for more traditional psychotherapies as against the placebo treatment. The difference in effect size between placebo and psychotherapy is then .19. However, it should be noted that this comparison is not based on the same studies. Such comparisons across different sets of studies may not be legitimate. Landman and Dawes (1982) report the results of a random sample of psychotherapy and behavior therapy studies and find that therapies are superior to placebos. We do not disagree with this assessment. We do believe, however, that for the clinically relevant sample of neurotic patients – the same group addressed in the Andrews and Harvey (1981a) analysis – there are no studies demonstrating that psychotherapy is superior to placebo controls.

Shapiro and Shapiro (1982) used a different data set in which two or more methods of therapy were compared in the same study. Their data suggested that verbal psychotherapies were inferior to other forms of therapy in the data set, and, as **Shapiro** notes in his critique of our article, their data did not permit them to infer that verbal psychotherapy was an effective form of treatment. Shapiro also notes that their data set did not permit them to infer anything meaningful about the effectiveness of verbal psychotherapies in comparison to placebo treatments.

Thus, when other comparable analyses available in the literature are looked at critically there is more agreement than is apparent from **Garfield**'s remark that other critics have arrived at other conclusions. Of course, there is an important issue of interpretation and of the probity of the weight to be assigned to different classes of investigations. And, contrary to the views of **Dahl,** we do disagree with Smith et al. (1980) about the proper interpretation of these data. Dahl's quotation from Smith et al's book is misleading. Their discussion of the data in Table 5–1 of their book on pages 89–91 clearly indicates that Smith et al. believe that psychotherapy has an effect size that is approximately twice that of placebo treatments. We do not disagree; we do question the relevance of that number to understanding issues of the effectiveness of psychotherapy.

Are there other studies not dealing with the psychotherapy-placebo comparison that bear crucially on issues of psychotherapeutic effectiveness? **Erwin** cites the well-known study of Sloane et al. (1975) as providing critical evidence for the effectiveness of psychotherapy against a wait-list control in a group of neurotic patients. The Sloane et al. study does not provide clear-cut evidence for the benefits of psychotherapy against their wait-list control. The follow-up data indicate that the benefits of therapy did *not* exceed those of patients in the wait-list control condition. Thus, their data suggest that the benefits of psychotherapy do not endure over time. **Wilson** cites Strupp and Hadley (1979) as a definitive study demonstrating the ineffectiveness of psychotherapy. Our target article does deal, although only in passing, with the issue raised by the Strupp and Hadley study, namely, the question of the professional competence of the therapist;

and we do cite Durlac (1979) in this connection who, in turn, does cite data from the Vanderbilt studies. However, strictly speaking, the Strupp and Hadley study does not deal with the issue of the effectiveness of psychotherapy but rather with the issue of whether a psychotherapist must be professionally trained in order to obtain positive outcomes. I believe that the study is important but tangential to the issues addressed in our paper.

**T. L. Rosenthal** criticizes us for failing to discuss the issue of the role of placebo treatments in behavior therapy, and he indicates (as do **Cordray & Bootzin**) that such effects occur in behavior therapy. I have no quarrel with this assertion. I merely wish to reiterate that our paper did not deal with behavior therapy, and we make no assertions about its effectiveness. Several of our critics imply that we are somehow secret supporters of behavior therapy, and that this commitment lurks in the background of our analysis. This is an unwarranted inference. **Shapiro** notes that we may overgeneralize the benefits of quasi-behavioristic treatment, and **Glass et al.** and **Garfield** imply that we are somehow allied with those behavior therapists who are opposed to psychotherapy on ideological grounds. Our assertion that the class of studies we analyzed in which psychotherapy was superior to placebo controls included several studies that used quasi-behavioristic psychotherapeutic treatment is merely to be regarded as a factual description. I do not consider such studies to be particularly relevant to the evaluation of behavior therapy since they are at best analogue studies and do not deal with relevant clinical samples.

We analyzed two studies that do provide evidence relevant to the effectiveness of behavior therapy: McLean and Hakstian (1979) and Gillan and Rachman (1974). And I would argue that neither study provides particularly convincing data in support of behavior therapy treatment. Both find that behavior therapy, unlike psychotherapy, produced improvement at the end of therapy relative to the placebo controls, and both studies reported that the benefits of behavior therapy faded on follow-up. I do not wish to draw any inference from these two studies; I do wish to reserve our right to restrict the scope of our article to the issues we dealt with. It seems unreasonable to criticize us for hidden biases we are presumed to have or for merely descriptive statements we make about studies.

What may be inferred from the existing literature about the effectiveness of psychotherapy? I certainly agree with the comments of **Greenberg, Kline, Spence, Garfield,** and **Shapiro** that we need better data. And, I certainly do not believe that meta-analysis is a substitute for a better data base. Indeed, we described the studies we analyzed in order to establish that many of the studies were irrelevant. Without this textured detail the presentations of quantitative indices of effect size are uninformative, and the attempt to relate effect sizes to features of the studies is a poor substitute for an adequate data base. Given the acknowledged irrelevance of much of the data cited in our article, why did we assert that it was unlikely that better designed research would lead to different conclusions? We made this assertion not solely from a sense of exasperation that 30 years of outcome research had failed, in our judgment, to produce convincing evidence of the efficacy of psychotherapy, but we believed that the available literature, while very far from permitting a definitive conclusion, did, nevertheless, provide us with some support for the view that in the target article we called proposition 2.

I take Brill et al. (1964) to be as close to a definitive study as we have in the literature. For a representative sample of neurotic outpatients, Brill was able to show, using clinically relevant judgments, that on a variety of outcome measures the effects of 20 weeks of psychotherapy sessions were equivalent to a *minimal* pill placebo treatment. What are the prospects that the findings of Brill et al. will be impeached by other studies? Perhaps therapy of longer duration would produce significant effects. Although Brill et al. found that 20 sessions of therapy was virtually equivalent to no therapy at all, that finding does not logically imply that the equivalence would be maintained for still longer therapeutic treatment. However, as I am unaware of any convincing evidence that duration of treatment affects outcome, it seems unlikely, although not impossible, that long-term therapy would lead to significant effects. The therapists in the Brill et al. study were inexperienced. Perhaps if Brill and his colleagues had used experienced therapists they would have obtained different results. Again, we think this unlikely. As we have noted, Strupp and Hadley (1979) as well as other research summarized by Durlac (1979), fail to provide evidence that therapeutic experience is related to therapeutic outcomes. Further, McLean and Hakstian (1979) and Gillan and Rachman (1974) used highly experienced psychotherapists committed to the virtues of psychotherapy as a preferred treatment modality and found that psychotherapy was not superior to placebo treatments – although they did not use minimal placebos. Therefore, I think it unlikely that a study using more experienced psychotherapists would report significantly different results.

A number of other suggestions have been made in the criticisms of our target article. **Dahl** suggests that we should permit the patient choice with respect to the preferred treatment and the preferred therapist. This seems like a reasonable innovation in therapy research and might lead to different results. **Frank** suggests that we should investigate the efficacy of modes of therapy that produce intense emotional arousal and altered states of consciousness. Our article indicates that the available literature provides evidence on only a small subset of therapies. It is entirely possible that there exist a subset of therapies – or as **Eagle** has indicated, therapists – that produce consistently positive results or consistently positive results for a certain class of patients.

**Placebos.** Placebos are appropriate for determining the reasons why psychotherapy is effective, argue **Cordray & Bootzin**, but not for determining the effectiveness of psychotherapy per se. I disagree. If it were definitively established that the necessary and sufficient condition for psychotherapeutic effectiveness were as **Frank** asserts, namely, that "a helping person listens to a patient's complaints and offers a procedure to relieve them, thereby inspiring the patient's hopes and combatting demoralization," then a number of practical implications would follow. It would, for example, be unnecessary to provide more than simple, inexpensive, and innocuous therapeutic interventions, as suggested by **Shepherd.** The rationale for the extensive training of professionals in the development of psychotherapeutic skills would evaporate. Thus, the issue of the relative effectiveness of psychotherapy and what we called minimal placebos in our target article is directly relevant to the evaluation of the effectiveness of psychotherapy, because it bears, in ways that seem essential, on the rationale for the provision of psychotherapeutic services, and on the evolution of the professional role of the therapist.

Quite apart from the issue of the use of placebos as a basis to evaluate the effectiveness of therapy, I am in complete sympathy with **Cordray & Bootzin** and **Sebeok's** concerns for a better theoretical understanding of the role of placebos. What is the minimal necessary and sufficient condition for an effective placebo treatment? Did Brill and his colleagues obtain the results they did because of the high credibility associated with chemical interventions? Would a trip to Baden Baden to luxuriate in the baths be equally effective? Does "bibliotherapy" (i.e., reading a book providing psychological guidance) act as an effective placebo? **Fish** raises the possibility that a variety of other placebo techniques may be more effective than pill placebos. Clearly, we need to learn a good deal more about the effects of placebos, and we need to explore possible techniques to enhance their effectiveness.

I agree with **Eysenck**'s assertion that our current theories of the processes that are alleged to cause change in therapy should at least be discomfited by the finding that therapy is no more effective than placebo treatment for neurotic patients. And I agree with **Eagle** that Grünbaum's (1981) analysis points to the need to understand more complex placebos by reference to a theory of the psychotherapy effect, although I am somewhat at a loss to understand his assertion that we did not understand Grünbaum. I would think that a careful reading of our introductory remarks on placebo treatments and our discussion of minimal placebos would have clarified this point.

**Meta-analysis.** Several of our critics (e.g., **Kazdin, Maher,** and **Wilson**) appear to suggest that we have erred by placing too much faith in meta-analysis; other critics (e.g., **R. Rosenthal**) assert that we did not carry our analysis far enough. I believe that we failed to make clear to those critics who, in effect, assert that we have slavishly adhered to meta-analytic procedures, that we were not advocating them as a way of evaluating psychotherapy. I would have thought this obvious from the fact that we reached conclusions diametrically opposed to those of Smith et al. although we obtained an identical estimate of the overall effectiveness of psychotherapy relative to placebos. We presented the details of the studies in Table 1 in order to demonstrate to the reader the irrelevance of most of the studies to serious attempts to assess the effectiveness of psychotherapy. Our meta-analysis might be construed more as a reductio ad absurdum. When one looks at the kinds of studies that go into an estimate of the psychotherapy versus placebo comparison included in the Smith et al. corpus, the mean value .42 (or the median, .15) is simply not credible. It is for this reason that we made clear at the beginning of our paper our reliance on the Brill et al. (1964) study and that our attempt was to study the literature cited in Smith et al.'s analysis to discover whether the results obtained by Brill et al. were anomalous. I do not believe that the existing literature is sufficient to permit one to rescue by statistical manipulation a credible analysis of psychotherapy outcome. In this connection I am sympathetic to **Eysenck's** *critiques of meta-analysis (Eysenck, in press) and to the* detailed discussion of the literature on psychotherapy reported by Rachman and Wilson (1980). I do not think that Wilson and I are in substantial disagreement about the usefulness of meta-analysis for an evaluation of psychotherapy, although I do think that the recently published paper by Shapiro and Shapiro (1982) does provide us with some useful information about behavior therapy.

Were we in error in our meta-analytic procedures, and did we fail to carry out an appropriate analysis as is implied by **Dawes, Hedges, R. Rosenthal, T. L. Rosenthal,** and **Glass et al.**? T. L. Rosenthal criticizes us for a failure to take account of his suggestions with respect to using a pooled variance rather than the variance of the control group and for the use of the arbitrary correlation value of .5 in correcting some of our data. We did, in point of fact, modify our original article to explain why we used the pooled variance, noting that in the majority of studies standard deviations for controls and therapy treatments were not available and hence a pooled standard deviation had to be used. Since we used a pooled standard deviation for the majority of our calculations we felt it appropriate to use this procedure for all of our calculations. Also, Smith et al. report that there were no consistent differences in the variance of therapy and wait-list controls. As to the decision to use the arbitrary value of .5 to correct for the assumed correlation between pretest and posttest values, where the correlation was not given in the article, we also added to our paper a rationale for this procedure. I regret that T. L. Rosenthal did not find this convincing. In any case, the use of estimated correlation values is only one of the many rather arbitrary decisions that are involved in calculating effect sizes in psychotherapy for limited data presentations. One has to decide how to summarize data where many subscores are reported. The fact that the calculation of an effect size for a study is not an automatic and mechanical process should be apparent from the presentation in Smith et al.'s book of data on the agreement between two different individuals calculating effect size values for the same study. In any case, our quantitative results are in good agreement with those obtained by Smith et al., and whatever slight variations in procedure we introduced for reasons that seemed sufficient to us did not result in sharp divergences in our results.

**Dawes** argues that our study is flawed because the model we use does not account for variance in the outcomes. This is quite irrelevant, however. We were not trying to find the variables that accounted for variance in effect sizes – we were not on a fishing expedition. We examined the studies to see whether there was any suggestion across studies that the effect sizes varied in a way that would suggest that the effects of placebos were ephemeral and weak relative to those of therapy. The data did not fit the theoretical model we tested. I leave it to the reader to decide whether the questions addressed are meaningful or not. Evidently, **Glass et al.** do not find them meaningful.

**Hedges**'s reanalysis of our data raises a similar issue. Hedges argues that our attempt to distinguish among classes of studies by an analysis of the type of subject population is flawed because our model does not fit the observed variations in effect sizes among studies. I find this criticism irrelevant. The qualitative analysis of studies by type of subject population was a rational division and not an attempt to fit an empirically derived model to the data. The distinction between analogue studies with students and true neurotic populations and inpatients is well developed in the literature on psychotherapy. For example, it figures prominently in Andrews and Harvey's (1981a) reanalysis of the Smith et al. data as well as in **Shapiro**'s discussion of the meaning of his meta-analysis. Sophisticated clinicians such as **Spence** and **Eagle** clearly point to the irrelevance of many of the studies we analyzed for an analysis of the effectiveness of psychotherapy. Hedges's reanalysis of these data, removing a small number of outliers and finding a weighted average effect size of .15 as a consistent estimate that fits the bulk of these data, is, from my perspective meaningless, since the studies are so diverse that a calculation that fails to consider the significant differences among them with respect to their relevance for the evaluation of psychotherapy in a clinically relevant sense is simply absurd.

**R. Rosenthal**'s reanalysis of our data raises similar issues. Rosenthal performs a number of analyses of the studies that we present in Table 1 and concludes that there are several meaningful relationships that we failed to discover because our meta-analysis was insufficiently searching. I do not find the meta-analysis performed by R. Rosenthal theoretically well motivated. The rationale for the division of student populations by age is not compelling. The quadratic relationships between age and effect size are, as Rosenthal notes, so confounded with other variables that they are not likely to be meaningful. Rosenthal concludes that these studies imply that psychotherapy is more powerful than psychological placebos but not more powerful than pill placebos. I found this argument unconvincing. Rosenthal omits the McLean and Hakstian (1979) study that contains a negative effect size for psychotherapy relative to psychological placebos. Adding this study to the list one finds negative effect sizes for two of the studies comparing psychological placebos to psychotherapy. Moreover, the relatively large effect size reported by Rosenthal for patients receiving psychological placebos derives principally from Roessler et al.'s (1977) study using self-report measures of outcome with physical rehabilitation clients and Coche and Flick's (1975) study of psychiatric inpatients of group problem solving training. It seems to me that neither of these is particularly relevant to the issue of outcome research in psychotherapy. I leave it to the reader to judge whether such a heterogeneous category is more meaningful than our grouping of neurotic studies, including four showing close to zero effect sizes for all outcome measures. Of this group of four, two used pill placebos and two used psychological placebos. R. Rosenthal's analyses appear to me to demonstrate the pitfalls of the application of meta-analysis to a diverse group of studies without an attempt to investigate the nature of the differences among the studies.

**Theoretical speculations.** The truth of a proposition is never considered in isolation – it is always relative to a body of theory that renders the particular finding or hypothesis sensible. In this concluding section I want to sketch some theoretical speculations derived from other areas of research that may help to explain why psychotherapy may not be a potent treatment.

First, there may be considerable long-term stability to neurotic behaviors. My colleague, James Conley (1982), has recently completed an analysis of longitudinal data on the consistency of neurotic tendencies over a 45-year period. His analysis suggests a surprising degree of consistency in neurotic tendencies over the adult life span.

Second, I believe that clinicians may have an inadequate understanding of the etiology of the conditions they treat. I have several reasons for this assertion:

(A) Kahneman, Slovic, and Tversky (1982) have presented data dealing with the tendency of clinicians to fail to deal properly with base-rate problems. Clinicians frequently treat conditions that have a relatively low base rate of occurrence in the population. They often attempt to explain these conditions with reference to events that have a high base rate of occurrence in the population. The conditional probability of the relevant alleged etiological occurrences given the clinical condition may be high, but the more relevant conditional probability of the clinical condition given the alleged etiological event may be low, thereby rendering the alleged aetiological events nugatory.

(B) Recent behavior-genetic research (Lykken 1982) strongly implies that there is a genetic basis for neuroticism and that the between-family environmental influence on personality traits is close to zero. Findings that monozygotic twins reared apart are as similar as those reared together and that dyzygotic twins reared together are hardly similar at all suggest that many of our theories of socialization and of the impact of the family environment on personality will have to be revised.

(C) Nisbett and Wilson's (1977) analysis of the validity of verbal reports about psychological processes suggests that such reports are very likely to be flawed. Of course, this has always been a cardinal tenet of psychoanalytic theorists. However, such theorists have assumed that the use of clinical methods of inference permits the therapist to reconstruct those events that are the true cause of the person's behavior. The available literature on the adequacy of clinical predictions of behavior suggests that this belief may not be well founded. Recall Milgram's finding that psychiatrists were quite wrong in their predictions of the behavior of subjects in his obedience studies (Milgram 1974). It may be that clinicians are not in general able to infer the psychological processes that govern a patient's actions and this inability might render successful therapeutic intervention difficult.

(D) Psychotherapy might influence a person's thoughts and emotions. However, such alterations of thoughts and emotions might not invariably influence human actions (see Brody 1983, chap. 5, for a discussion of this issue).

(E) I suspect that many of the causes of distress in a person's life derive not from intrapsychic events but from extrapsychic events that the person or therapist may not be able to control. To the extent that this is true we should not expect psychotherapy to be a powerful form of treatment.

I do not maintain that this sketch should be construed as a theoretical explanation of why psychotherapy may not be a powerful form of treatment. Rather, the sketch is offered as an attempt to suggest relationships between outcome research and broader research issues.

## References


Alper, T B. & Kranzler, G. D. (1970) A comparison of the effectiveness of behavioral and client-centered approaches for the behavior problems of elementary school children. *Elementary School Guidance and Counseling* 5.35–43. [taLP]

Andrews, G., Guitar, B. & Howie, P. (1980) Meta-analysis of the effects of stuttering treatment. *Journal of Speech and Hearing Disorders* 45:287–307. [GA]

Andrews, G. & Harvey, R. (1981a) Does psychotherapy benefit neurotic patients? A reanalysis of the Smith, Glass, and Miller data. *Archives of General Psychiatry* 38:1203–8. [GA, NBr, RPG]

(1981b) Regression to the mean in pretreatment measures of stuttering severity. *Journal of Speech and Hearing Disorders* 46:204–7. [GA]

Andrews, G., MacMahon, S. W., Austin, A. & Byrne, D. B. (1982) Hypertension: A comparison of drug and nondrug treatments. *British Medical Journal* 284 1523–26. [GA]

Baer, D. M. & Stolz, S. B. (1978) A description of the Erhard seminars training (est) in the terms of behavior analysis. *Behaviorism* 6:45–70. [JDF]

Bandura, A. (1977) Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review* 84:191–215. [EE]

(1978). The self system in reciprocal determinism. *American Psychologist* 33.344–58 [EE]

Barber, T. X. (1969) *Hypnosis: A scientific approach.* Van Nostrand-Reinhold. [JMF]

Barber, T. X., Spanos, N. P. & Chaves, J. F. (1974) *Hypnotism. Imagination and human potentialities.* Pergamon. [JMF]

Bergin, A. E. & Lambert, M J. (1978) The evaluation of therapeutic outcomes. In: *Handbook of psychotherapy and behavior change: An empirical analysis*, ed. S. L. Garfield & A. E Bergin, pp. 139–89. Wiley. [taLP, TLR]

Bootzin, R. R. & Lick, J. R. (1979) Expectancies in therapy research: Interpretive artifact or mediating mechanism? *Journal of Consulting and Clinical Psychology* 47:852–55. [DSC]

Brill, N. Q., Koegler, R. R., Epstein, L. J. & Forgy, E. W. (1964) Controlled study of psychiatric outpatient treatment *Archives of General Psychiatry* 10:581–95. [GA, RPG, tarLP]

Brody, N. (1983) *Human motivation: Commentary on goal-directed action.* New York. Academic Press. [rNB]

Bruce, J. (1971) The effects of group counseling on selected vocational rehabilitation clients. Ph.D. dissertation, Florida State University. [taLP]

Bruyere, D. H. (1975) The effects of client centered and behavioral group counseling on classroom behavior and self concept of junior high school students who exhibited disruptive classroom behavior. Ph.D. dissertation, University of Oregon. [DSC, taLP]

Catanese, R. A., Rosenthal, T. L. & Kelley, J. E. (1978) Strange bedfellows: Reward, punishment, and impersonal distraction strategies in treating dysphoria. *Cognitive Therapy and Research* 3.229–305. [TLR]

Cattell, R. B. & Kline, P. (1977) *The scientific analysis of personality and motivation.* Academic Press. [PK]

Chertok, L. N. & Fontaine, M. (1963) Psychosomatics in Veterinary Medicine. *Journal of Psychosomatic Research* 7:229–35. [TAS]

Coche, E. & Douglas, A. A. (1977) Therapeutic effects of problem-solving training and play-reading groups. *Journal of Clinical Psychology* 33:820–27. [GVC, taLP]

Coche, E. & Flick, A. (1975) Problem-solving training groups for hospitalized patients. *Journal of Psychology* 91:19–29 [GVC, taLP]

Cohen, J. (1969) *Statistical power analysis for the behavioral sciences.* Academic Press [taLP]

Comas-Diáz, L. (1981) Effects of cognitive and behavioral group treatment on the depressive symptomatology of Puerto Rican women. *Journal of Consulting and Clinical Psychology* 49:627–32. [GVC, tarLP]

Conley, J. J. (1982) Longitudinal consistency of adult personality: Neuroticism and social introversion-extroversion over forty-five years. Unpublished manuscript, Wesleyan University, Department of Psychology. [rNB]

Cook, T. D. & Leviton, L. C. (1980) Reviewing the literature: A comparison of traditional methods with meta-analysis. *Journal of Personality* 48:449–72. [DAS]

Cordray, D. S. & Orwin, R. G. (1983) Facilitating the quality of research: Interconnections among primary research, secondary analysis and quantitative synthesis. In: *Evaluation studies review annual*, ed. R. Light. Sage Publications (in press) [DSC]

Desrats, R. G. (1975) The effects of developmental and modeling group counseling on adolescents in childcare institutions. Ph.D. dissertation, Lehigh University. [taLP]

DiLoreto, A. O. (1971) *Comparative psychology.* Aldine-Atherton. [JDF, taLP, DAS]

Dohrenwend, B. P., Oksenberg, L., Shrout, P. E., Dohrenwend, B. S. & Cook, D. (1982) What brief psychiatric screening scales measure. In: *Proceedings of the Third Biennial Conference on Health Survey Research Methods, May, 1979*, ed. S. Sudman. National Center for Health Services Statistics, in press. [JDF]

Duncan, J. W. & Laird, J. D. (1980) Positive and reverse placebo effects as a function of differences in cues used in self-perception. *Journal of Personality and Social Psychology* 39:1024–36. [taLP]

Durlac, J. A. (1979) Comparative effectiveness of paraprofessional and professional helpers. *Psychological Bulletin* 86.80–92. [tarLP]

Eysenck, H. J. (1952) The effects of psychotherapy. An evaluation. *Journal of Consulting Psychology* 16:319–24. [HJE, taLP]

(1966) *The effects of psychotherapy.* International Science Press. [taLP, DAS]

(1978) An exercise in mega-silliness. *American Psychologist* 33:517. [taLP]

(in press a) Meta-analysis. An abuse of research integration. *Journal of Special Education.* [HJE]

(in press b) Special review: The benefits of psychotherapy. A battlefield revisited. *Behaviour Research and Therapy.* [rNB, HJE]

Fish, J. M. (1973) *Placebo therapy* Jossey-Bass. [JMF]

Fisher, S. & Greenberg, R. P. (1977) *The scientific credibility of Freud's theories and therapy.* Basic Books. [RPG]

Foster, J. G. (1971) *Enquiry into the practice and effects of scientology.* H.M.S O [MS]

Frank, J. D. (1973) *Persuasion and healing. A comparative study of psychotherapy.* Rev. ed. Johns Hopkins Press. [JMF]

(1974) Therapeutic components of psychotherapy. *Journal of Nervous and Mental Disease* 159:325–42. [MNE]

(1978) Expectation and therapeutic outcome – the placebo effect and the role induction interview. In: *Effective ingredients of successful psychotherapy*, ed. J. D. Frank, R. Hoehn-Saric, S. D. Imber, B. L. Liberman & A. R. Stone, pp. 1–34. Brunner/Mazel. [JDF]

Garfield, S. L. (1980) Psychotherapy: A 40-year appraisal. *American Psychologist* 36.174–83. [SLG]

(in press) The effectiveness of psychotherapy: The perennial controversy. *Professional Psychology.* [SLG]

Gillan, P. & Rachman, S. (1974) An experimental investigation of desensitization and phobic patients. *British Journal of Psychology* 124.392–401. [GA, EE, SLG, tarLP, TLR]

Gist, R. & Stolz, S. B. (1982) Mental health promotion and the media: Community response to the Kansas City hotel disaster. *American Psychologist* 37:1136–39. [BM]

Glass, G. V & Kliegl, R. M. (1982) An apology for research integration in the study of psychotherapy *Journal of Consulting and Clinical Psychology* 50 (in press). [GVG]

Grande, L. M. (1975) A comparison of rational-emotive therapy, attention placebo and no-treatment groups in the reduction of interpersonal anxiety. Ph.D. dissertation, Arizona State University. [JDF, taLP, TLR]

Greenberg, R. P. & Staller, J. (1981) On personal therapy for therapists. *American Journal of Psychiatry* 138:1467–71. [RPG]

Grünbaum, A. (1981) The placebo concept. *Behavior Research and Therapy* 19:157–67. [MNE, tarLP]

Haley, J. (1977) *Uncommon therapy: The psychiatric techniques of Milton H. Erickson.* Norton. [taLP]

Hedges, L. V. (1981) Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics* 6.107–28. [LVH, taLP, RR]

(1982a) Estimation of effect size from a series of independent experiments. *Psychological Bulletin* 92:490–99. [LVH, RR]

(1982b) Fitting categorical models to effect sizes from a series of experiments. *Journal of Educational Statistics* 7:119–37. [LVH]

Hedquist, F. S. & Weinhold, B. K. (1970) Behavioral group counseling with socially anxious and unassertive college students. *Journal of Counseling Psychology* 17:237–42. [taLP]

Herman, B. (1972) An investigation to determine the relationship of anxiety and reading disability and to study the effects of group and individual counseling on reading improvement. Ph.D. dissertation, University of New Mexico. [taLP]

Hogan, D. (1980) *The regulation of psychotherapists*, vol. 1. Ballinger. [MS]

Hogan, R. A. & Kirchner, J. H. (1968) Implosive, eclectic, verbal and bibliotherapies in the treatment of fears of snakes. *Behavior Research and Therapy* 6:167–71. [taLP]
House, R. M. (1970) The effects of nondirective group play therapy upon the sociometric status and self-concept of selected second grade children. Ph.D. dissertation, Oregon State University. [LVH, taLP]
Janov, A. (1970) *Primal therapy: The cure for neurosis.* Putnam. [JDF]
Jarmon, D. D. (1972) Differential effectiveness of rational-emotive therapy, bibliotherapy and attention-placebo in the treatment of speech anxiety. Ph.D. dissertation. Southern Illinois University. [SLG, taLP]
Jong, E. (1977) *How to save your own life.* Secker & Warburg. [GA]
Kahneman, D., Slovic, P. & Tversky, A. (1982) *Judgment under uncertainty: Heuristics and biases.* Cambridge University Press. [rNB]
Kandel, E. R. (1979) Psychotherapy and the single synapse: The impact of psychiatric thought on neurobiologic research. *New England Journal of Medicine* 301:1028–37. [JDF]
Kazdin, A. E. (1980) *Research design in clinical psychology.* Harper & Row. [AEK, GTW]
Kazdin, A. E. & Wilson, G. T. (1978) *Evaluation of behavior therapy: Issues, evidence and research strategies.* Ballinger. [GTW]
Kazrin, A., Durac, J. & Agteros, T. (1979) Meta-analysis: a new method for evaluating therapy outcome. *Behaviour Research and Therapy* 17.397–99. [GA]
Kiesler, D. J. (1966) Some myths of psychotherapy research and the search for a paradigm. *Psychological Bulletin* 65:110–36. [PK, GTW]
Kirsch, I. & Henry, D. (1977) Extinction versus credibility in the desensitization of test anxiety. *Journal of Consulting and Clinical Psychology* 45:1052–59. [TLR]
Klein, D. F. (1981) Anxiety misconceptualized. In· *Anxiety, new research and changing concepts,* ed. D. F. Klein & J. G. Rabkin, pp. 235–63. Raven Press. [JDF]
Kline, P. (1981) *Fact and fantasy in Freudian theroy.* 2nd ed Methuen. [PK]
Landman, J. T. & Dawes, R. M. (1982) Psychotherapy outcome. Smith and Glass' conclusions stand up under scrutiny. *American Psychologist* 37:504–16. [GA, RPG, SLG, taLP, RR]
Lester, B. G. (1973) A comparison of relationship counseling and relationship counseling combined with modified systematic desensitization in reducing test anxiety in middle school pupils. Ph.D. dissertation, University of Virginia. [taLP]
Lick, J. R. & Bootzin, R. R. (1975) Expectancy factors in the treatment of fear: Methodological and theoretical issues *Psychological Bulletin* 82.917–31. [DSC]
London, P. & Klerman, G. L. (1982) Evaluating psychotherapy *American Journal of Psychiatry* 139.709–17. [BM]
Lorr, M., McNair, D. M. & Weinstein, G. J. (1963) Early effects of chlordiazepoxide (Librium) used with psychotherapy. *Journal of Psychiatric Research* 1:257–70. [GA, MNE, RPG, taLP]
Luborsky, L., Chandler, M., Auerbach, A. H., Cohen, J. & Bachrach, H. M. Factors influencing the outcome of psychotherapy: A review of quantitative research. *Psychological Bulletin* 75:145–85. [RPG]
Luborsky, L., Mintz, J., Auerbach, A., Christoph, P., Bachrach, H., Todd, T., Johnson, M., Cohen, M. & O'Brien, C. (1980) Predicting the outcome of psychotherapy: Findings of the Penn psychotherapy project. *Archives of General Psychiatry* 37:471–81. [HD]
Lykken, D. T. (1982) Research with twins. The concept of emergenesis *Psychophysiology* 19:361–373. [rNB]
Maher, B. A. (1981) Mandatory insurance coverage for psychotherapy: A tax on the subscriber and a subsidy to the practitioner. *Clinical Psychologist* 1:9–12. [BM]
Malan, D. (1959) On assessing the results in psychotherapy. *British Journal of Medical Psychology* 32:86–105. [PK]
Marks, J. (1981) Behavioral treatment and drugs in anxiety symptoms. In *Anxiety, new research and changing concepts,* ed. D. F. Klein & J. G. Rabkin, pp. 265–89. Raven Press. [JDF]
Matthews, D. B. (1972) The effects of reality therapy on reported self-concept, social adjustment, reading achievement, and discipline of fourth and fifth grades in two elementary schools. Ph.D. dissertation, University of Southern California. [taLP]
McGlynn, F. D., Kinjo, K. & Doherty, G. (1978) Effects of cue-controlled relaxation, a placebo treatment, and no treatment on changes in self-reported test anxiety among college students. *Journal of Clinical Psychology* 34:707–14. [TLR]
McLean, P. D. & Hakstian, A. R. (1979) Clinical depression. Comparative efficacy of outpatient treatments. *Journal of Consulting and Clinical Psychology* 47:818–36. [GA, rNB, MNE, GVG, taLP]
Meehl, P. E. (1955) Psychotherapy. *Annual Review of Psychology* 6:357–79. [taLP]
Meichenbaum, D. H., Gilmore, J. B. & Fedoravicious. (1972) Group insight versus group desensitization in treating speech anxiety. In: *Psychotherapy 1971,* ed. J. D. Matarazzo, pp. 513–23. Aldine-Atherton. [JDF, taLP]
Meltzoff, J. & Kornreich, M (1970) *Research in psychotherapy.* Aldine-Atherton. [SLG, taLP]
Milgram, S. (1974) *Obedience to authority.* Harper and Row. [rNB]
Miller, J. (1978) *The body in question* Random House. [TAS]
Morgan, R., Luborsky, L., Crits-Christoph, P , Curtis, H. & Solomon, J. (1982) Predicting the outcomes of psychotherapy by the Penn helping alliance rating method. *Archives of General Psychiatry* 39:397–402 [HD]
Mountcastle, V. B. (1975) The view from within: Pathways to the study of perception. *Johns Hopkins Medical Journal* 136.109–31. [JDF]
Murdock, M. N. (1982) A meta-analysis of psychotherapy outcome research. "Sentence first – verdict afterwards?" Honors thesis, Wesleyan University. [taLP]
Nisbett, R E. & Wilson, T. DeC. (1977) Telling more than we can know: Verbal reports on mental processes. *Psychological Review* 84.231–259. [rNB]
Orlov, L (1972) An experimental study of the effects of group counseling with behavior problem children at the elementary school level. Ph.D. dissertation, The Catholic University of America. [taLP]
Parloff, M. B. (1979) Can psychotherapy research guide the policymaker? *American Psychologist* 34·296–306. [MS]
(1980) Psychotherapy and research· An anaclitic depression. *Psychiatry* 43.279–93. [JDF]
(1982) Psychotherapy research evidence and reimbursement decisions. Bambi meets Godzilla *American Journal of Psychiatry* 139:718–27 [BM]
Parloff, M. B., Waskow, I. E. & Wolfe, B. E. (1978) Research on therapist variables in relation to process and outcome. In. *Handbook of psychotherapy and behavior change,* ed. S. L. Garfield & A. E. Bergin, pp. 233–82. Wiley. [RPG]
Paul, G. L. (1964) Effects of insight, desensitization and attention-placebo treatment of anxiety: An approach to outcome research in psychotherapy. Ph.D. dissertation, University of Illinois [GA, taLP, DPS, GTW]
Paul, G. (1969) Outcome of systematic desensitization. II. Controlled investigations of individual treatment, technique variations and current status. In: *Behavior therapy: Appraisal and status,* ed. C. M. Franks. McGraw-Hill. [EE]
Paykel, E , DiMascio, A., Haskell, D. & Prusoff, B (1975) Effects of maintenance amitriptyline and psychotherapy on symptoms of depression *Psychological Medicine* 5.67–77. [GA, taLP]
Platt, J. M (1970) Efficacy of the Adlerian model in elementary school counseling. Ph.D. dissertation, University of Arizona. [taLP, LVH]
Presby, S. (1978) Overly broad categories obscure important differences. *American Psychologist* 33.514–15. [LVH]
Prioleau, L A. (1982) A meta-analysis of psychotherapy outcome research: "Sentence first – verdict afterwards?" Master's thesis, Wesleyan University. [taLP]
Quality Assurance Project. (1982) A treatment outline for agoraphobia. *Australian and New Zealand Journal of Psychiatry* 16:25–33 [GA]
Rachman, S. & Wilson, G. T. (1980). *The effects of psychological therapy.* Pergamon Press. [rNB, EE, HJE, SLG, GTW]
Rehm, L. P & Marston, A. R. (1968) Reduction of social anxiety through modification of self-reinforcement: An investigation therapy technique. *Journal of Consulting and Clinical Psychology* 32:5. [taLP]
Roessler, R., Cook, D. & Lillard, D. (1977) Effects of systematic and group counseling on work adjustment clients. *Journal of Consulting Psychology* 24:313–17. [LVH, taLP]
Rosen, D. (1977) A primal primer for psychotherapists. *American Journal of Psychiatry* 134.445–46. [JDF]
Rosenthal, R. (1978) Combining results of independent studies. *Psychological Bulletin* 85:185–93. [RR]
(1983) Assessing the statistical and social importance of the effects of psychotherapy *Journal of Consulting and Clinical Psychology* 51:4–13. [RR]
Rosenthal, R. & Rubin, D. B. (1978) Interpersonal expectancy effects: The first 345 studies. *Behavioral and Brain Sciences* 3:377–415. [RR]
(1982a) Comparing effect sizes of independent studies. *Psychological Bulletin* 92.500–4. [LVH, RR]
(1982b) Further meta-analytic procedures for assessing cognitive gender differences. *Journal of Educational Psychology* 74:708–12. [RR]
(1982c) A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology* 74:166–69. [RR]
Rosenthal, T. L. (1980) Social cueing processes. In: *Progress in behavior modification,* vol. 10, ed. M. Hersen, R. M. Eisler & P. M. Miller, pp. 111–46. Academic Press. [TLR]

Rosentover, I. (1974) Group counseling of the underachieving high school student as related to self-image and academic success. Ph.D. dissertation, Rutgers University. [taLP]

Rush, A. J., Beck, A. T., Kovacs, M. & Hollon, S. (1977) Comparative effects of cognitive therapy and pharmacotherapy in the treatment of depressed outpatients. *Cognitive Therapy and Research* 1.17–37. [JDF]

Russell, M. A. H., Armstrong, E. & Patel, U. A. (1976) Temporal contiguity in electric aversion therapy for smoking. *Behaviour Research and Therapy* 14:103–23. [TLR]

Schwartz, J. L. & Dubitzky, M. (1967) Clinical reduction of smoking. *Addictions* 14:35–44. [taLP]

Sebeok, T. A. (1981) *The play of musement.* Indiana University Press. [TAS]

Shapiro, A. K. & Morris, L. A. (1978) Placebo effects in medical and psychological therapies. In: *Handbook of psychotherapy and behavior change: An empirical analysis*, ed. S. L. Garfield & A. E. Bergin, pp. 396–410. Wiley. [taLP]

Shapiro, D. A. (1981) Comparative credibility of treatment rationales. Three tests of expectancy theory. *British Journal of Clinical Psychology* 21.111–22. [DAS]

Shapiro, D. A. & Shapiro, D. (1981) Meta-analysis of comparative therapy outcome studies. *MRC/SSRC Social and Applied Psychology Memo No. 438*, Department of Psychology, University of Sheffield. [GA]

(1982a) Meta-analysis of comparative therapy outcome research: A critical appraisal. *Behavioural Psychotherapy* 10:4–25. [PK, DAS]

(1982d) Meta-analysis of comparative therapy outcome research: A reply to Wilson. *Behavioural Psychotherapy* 10:307–10. [PK]

(1982c) Meta-analysis of comparative therapy outcome studies. A replication and refinement. *Psychological Bulletin* 92:581–604. [rNB, DAS]

(1982b) Comparative therapy outcome research: Methodological implications of meta-analysis. *Journal of Consulting and Clinical Psychology*, in press. [DAS]

(in press) Meta-analysis of comparative therapy outcome studies: A replication and refinement. *Psychological Bulletin.* [SLG]

Shapiro, S. B. & Knapp, D. M. (1971) The effect of ego therapy on personality integration. *Psychotherapy: Theory, Research and Practice* 8:208–12. [taLP]

Shepherd, M. (1979) Psychoanalysis, psychotherapy, and health services. *British Medical Journal* 2:1557–59. [MS]

(1980) The statutory registration of psychotherapists? *Bulletin of the Royal College of Psychiatrists* November, pp. 166–69. [MS]

Sloane, R. B., Staples, F. R., Cristol, A. H., Yorkston, N. J. & Whipple, K. (1970) *Psychotherapy versus behavior therapy.* Harvard University Press. [EE, JDF, tarLP]

Smith, M. L. & Glass, G. V. (1977) Meta-analysis of psychotherapy outcome studies. *American Psychologist* 32.752–60. [SLG, AEK]

Smith, M. L., Glass, G. V. & Miller, T. I. (1980) *The benefits of psychotherapy.* Johns Hopkins University Press. [GA, rNB, DSC, HD, HJE, MNE, JDF, GVG, RPG, SLG, AEK, PK, BM, trLP, RR, TLR, DAS, GTW]

Strahan, R. F. (1978) Six ways of looking at an elephant. *American Psychologist* 33:693. [taLP]

Strube, M. J. & Hartmann, D. P. (1982) A critical appraisal of meta-analysis. *British Journal of Clinical Psychology* 21:129–39. [DAS]

Strupp, H. (1973) *Psychotherapy: Clinical research and theoretical issues.* Aronson. [taLP]

(1977) *Psychotherapy for better or worse. The problem of negative effects.* Aronson. [taLP]

Strupp, H. H. & Hadley, S. W. (1979) Specific vs. nonspecific factors in psychotherapy. *Archives of General Psychiatry* 36:1125–36. [rNB, GTW]

*Time* (1979) Psychiatry on the couch. 2 April, p. 74. [MS]

Trexler, L. D. & Karst, T. O. (1972) Rational-emotive therapy, placebo, and no-treatment effects on public-speaking anxiety. *Journal of Abnormal Psychology* 79:60–67. [JDF, taLP]

Vanden Bos, G. & Pino, C. (1980) Research on the outcome of psychotherapy. In: *Psychotherapy: Practice, research, policy*, ed. G. Vanden Bos. Sage Publications. [EE]

Warner, R. (1969) An investigation of the effectiveness of verbal reinforcement and model reinforcement counseling on alienated high school students. Ph.D. dissertation, State University of New York – Buffalo. [DSC, taLP]

Weissman, M. M., Klerman, G. O., Prusoff, B. A. et al. (1981) Depressed outpatients. *Archives of General Psychiatry* 38:51–55. [GA, rNB]

Weissman, M. M., Prusoff, B. A., DiMascio, A. et al. (1979) The efficacy of drugs and psychotherapy in the treatment of acute depressive episodes. *American Journal of Psychiatry* 136:555–58. [GA, rNB]

West, W. B. (1969) An investigation of the significance of client-centered play therapy as a counseling technique. Ph.D. dissertation, North Texas State University. [taLP, DPS]

Wilson, G. T. (1982) How useful is meta-analysis in evaluating the effects of different psychological therapies? *Behavioural Psychotherapy* 10:221–31. [PK, DAS]

Wilson, G. T. & Rachman, S. (1982) Meta-analysis and the evaluation of psychotherapy outcome: Limitations and liabilities. *Journal of Consulting and Clinical Psychology*, in press. [GTW]

Winkler, R. C., Teigland, J. J., Munger, P. F. & Kranzler, G. D. (1965) The effects of selected counseling and remedial techniques on underachieving elementary school students. *Journal of Counseling Psychology* 12:384. [taLP]

Zeiss, A. M., Lewinsohn, P. N. & Munoz, R. F. (1979) Nonspecific improvement effects in depression using interpersonal skill training, pleasant activity schedules, or cognitive training. *Journal of Consulting and Clinical Psychology* 47:427–39. [TLR]