Meta-Analysis of Psychotherapy Outcome Studies

2 authors:

Mary Lee Smith
78 PUBLICATIONS   4,445 CITATIONS

SEE PROFILE

Gene V Glass
Arizona State University
185 PUBLICATIONS   14,262 CITATIONS

SEE PROFILE

Meta-Analysis of Psychotherapy Outcome Studies

MARY LEE SMITH   *University of Colorado—Boulder*
GENE V GLASS   *University of Colorado—Boulder*

ABSTRACT: *Results of nearly 400 controlled evaluations of psychotherapy and counseling were coded and integrated statistically. The findings provide convincing evidence of the efficacy of psychotherapy. On the average, the typical therapy client is better off than 75% of untreated individuals. Few important differences in effectiveness could be established among many quite different types of psychotherapy. More generally, virtually no difference in effectiveness was observed between the class of all behavioral therapies (systematic desensitization, behavior modification) and the nonbehavioral therapies (Rogerian, psychodynamic, rational-emotive, transactional analysis, etc.).*

Scholars and clinicians have argued bitterly for decades about the efficacy of psychotherapy and counseling. Michael Scriven proposed to the American Psychological Association's Ethics Committee that APA-member clinicians be required to present a card to prospective clients on which it would be explained that the procedure they were about to undergo had never been proven superior to a placebo ("Psychotherapy Caveat," 1974). Most academics have read little more than Eysenck's (1952, 1965) tendentious diatribes in which he claimed to prove that 75% of neurotics got better regardless of whether or not they were in therapy—a conclusion based on the interpretation of six controlled studies. The perception that research shows the inefficacy of psychotherapy has become part of conventional wisdom even within the profession. The following testimony was recently presented before the Colorado State Legislature:

Are they [the legislators] also aware of the relatively primitive state of the art of treatment outcome evaluation which is still, after fifty years, in kind of a virginal state? About all we've been able to prove is that a third of the people get better, a third of the people stay the same, and a third of the people get worse, irregardless of the treatment to which they are subjected. (Quoted by Ellis, 1977, p. 3)

Only close followers of the issue have read Bergin's (1971) astute dismantling of the Eysenck myth in his review of the findings of 23 controlled evaluations of therapy. Bergin found evidence that therapy is effective. Emrick (1975) reviewed 72 studies of the psychological and psychopharmacological treatment of alcoholism and concluded that evidence existed for the efficacy of therapy. Luborsky, Singer, and Luborsky (1975) reviewed about 40 controlled studies and found more evidence. Although these reviews were reassuring, two sources of doubt remained. First, the number of studies in which the effects of counseling and psychotherapy have been tested is closer to 400 than to 40. How representative the 40 are of the 400 is unknown. Second, in these reviews, the "voting method" was used; that is, the number of studies with statistically significant results in favor of one treatment or another was tallied. This method is too weak to answer many important questions and is biased in favor of large-sample studies.

The purpose of the present research has three parts: (1) to identify and collect all studies that tested the effects of counseling and psychotherapy; (2) to determine the magnitude of effect of the therapy in each study; and (3) to compare the effects of different types of therapy and relate the size of effect to the characteristics of the therapy (e.g., diagnosis of patient, training of therapist) and of the study. Meta-analysis, the integration of research through statistical analysis of the analyses of individual studies (Glass, 1976), was used to investigate the problem.

Procedures

Standard search procedures were used to identify 1,000 documents: *Psychological Abstracts, Dissertation Abstracts*, and branching off of bibliographies of the documents themselves. Of those documents located, approximately 500 were selected for inclusion in the study, and 375 were fully analyzed. To be selected, a study had to have at least one ther-

apy treatment group compared to an untreated group or to a different therapy group. The rigor of the research design was not a selection criterion but was one of several features of the individual study to be related to the effect of the treatment in that study. The definition of psychotherapy used to select the studies was presented by Meltzoff and Kornreich (1970):

Psychotherapy is taken to mean the informed and planful application of techniques derived from established psychological principles, by persons qualified through training and experience to understand these principles and to apply these techniques with the intention of assisting individuals to modify such personal characteristics as feelings, values, attitudes, and behaviors which are judged by the therapist to be maladaptive or maladjustive. (p. 6)

Those studies in which the treatment was labeled "counseling" but whose methods fit the above definition were included. Drug therapies, hypnotherapy, bibliotherapy, occupational therapy, milieu therapy, and peer counseling were excluded. Sensitivity training, marathon encounter groups, consciousness-raising groups, and psychodrama were also excluded. Those studies that Bergin and Luborsky eliminated because they used "analogue" therapy were retained for the present research. Such studies have been designated analogue studies because therapy lasted only a few hours or the therapists were relatively untrained. Rather than arbitrarily eliminating large numbers of studies and losing potentially valuable information, it was deemed preferable to retain these studies and investigate the relationship between length of therapy, training of therapists, and other characteristics of the study and their measured effects. The arbitrary elimination of such analogue studies was based on an implicit assumption that they differ not only in their methods but also in their effects and how those effects are achieved. Considering methods, analogue studies fade imperceptibly into "real" therapy, since the latter is often short term, or practiced by relative novices, etc. Furthermore, the magnitude of effects and their relationships with other variables are empirical questions, not to be assumed out of existence. Dissertations and fugitive documents were likewise retained, and the

measured effects of the studies compared according to the source of the studies.

The most important feature of an outcome study was the magnitude of the effect of therapy. The definition of the magnitude of effect—or "effect size"—was the *mean difference between the treated and control subjects divided by the standard deviation of the control group,* that is, $ES = (\bar{X}_T - \bar{X}_C)/s_C$. Thus, an "effect size" of $+1$ indicates that a person at the mean of the control group would be expected to rise to the 84th percentile of the control group after treatment.

The effect size was calculated on any outcome variable the researcher chose to measure. In many cases, one study yielded more than one effect size, since effects might be measured at more than one time after treatment or on more than one different type of outcome variable. The effect-size measures represent different types of outcomes: self-esteem, anxiety, work/school achievement, physiological stress, etc. Mixing different outcomes together is defensible. First, it is clear that all outcome measures are more or less related to "well-being" and so at a general level are comparable. Second, it is easy to imagine a Senator conducting hearings on the NIMH appropriations or a college president deciding whether to continue funding the counseling center asking, "What kind of effect does therapy produce—on anything?" Third, each primary researcher made value judgments concerning the definition and direction of positive therapeutic effects for the particular clients he or she studied. It is reasonable to adopt these value judgments and aggregate them in the present study. Fourth, since all effect sizes are identified by type of outcome, the magnitude of effect can be compared across type of outcome to determine whether therapy has greater effect on anxiety, for example, than it does on self-esteem.

Calculating effect sizes was straightforward when means and standard deviations were reported. Although this information is thought to be fundamental in reporting research, it was often overlooked by authors and editors. When means and standard deviations were not reported, effect sizes were obtained by the solution of equations from t and F ratios or other inferential test statistics. Probit transformations were used to convert to effect sizes the percentages of patients who improved (Glass, in press). Original data were requested from several authors when effect sizes could not be derived from any reported information. In two instances, effect sizes were impossible to recon-

struct: (a) nonparametric statistics irretrievably disguise effect sizes, and (b) the reporting of no data except the alpha level at which a mean difference was significant gives no clue other than that the standardized mean difference must exceed some known value.

Eight hundred thirty-three effect sizes were computed from 375 studies, several studies yielding effects on more than one type of outcome or at more than one time after therapy. Including more than one effect size for each study perhaps introduces dependence in the errors and violates some assumptions of inferential statistics. However, the loss of information that would have resulted from averaging effects across types of outcome or at different follow-up points was too great a price to pay for statistical purity.

The effect sizes of the separate studies became the "dependent variable" in the meta-analysis. The "independent variables" were 16 features of the study described or measured in the following ways:

1. The type of therapy employed, for example, psychodynamic, client centered, rational-emotive, behavior modification, etc. There were 10 types in all; each will be mentioned in the Results section.
2. The duration of therapy in hours.
3. Whether it was group or individual therapy.
4. The number of years' experience of the therapist.
5. Whether clients were neurotics or psychotics.
6. The age of the clients.
7. The IQ of the clients.
8. The source of the subjects—whether solicited for the study, committed to an institution, or sought treatment themselves.
9. Whether the therapists were trained in education, psychology, or psychiatry.
10. The social and ethnic similarity of therapists and clients.
11. The type of outcome measure taken.
12. The number of months after therapy that the outcomes were measured.
13. The reactivity or "fakeability" of the outcome measure.
14. The date of publication of the study.
15. The form of publication.
16. The internal validity of the research design.

Definitions and conventions were developed to increase the reliability of measurement of the features of the studies and to assist the authors in estimating the data when they were not reported. The more important conventions appear in Table 1. Variables not mentioned in Table 1 were measured in fairly obvious ways. The reliability of measurement was determined by comparing the codings of 20 studies by the two authors and four assistants. Agreement exceeded 90% across all categories.[1]

---

[1] The values assigned to the features of the studies, the effect sizes, and all procedures are available in Glass, Smith, and Miller (Note 1).

Analysis of the data comprised four parts: (1) descriptive statistics for the body of data as a whole; (2) descriptive statistics for the comparison of therapy types and outcome types; (3) descriptive statistics for a subset of studies in which behavioral and nonbehavioral therapies were compared *in the same study;* and (4) regression analyses in which effect sizes were regressed onto variables descriptive of the study.

Findings

DATA FROM ALL EXPERIMENTS

Figure 1 contains the findings at the highest level of aggregation. The two curves depict the average treated and untreated groups of clients across 375 studies, 833 effect-size measures, representing an evaluation of approximately 25,000 control and experimental subjects each. On the average, clients 22 years of age received 17 hours of therapy from therapists with about $3\frac{1}{2}$ years of experience and were measured on the outcome variables about $3\frac{3}{4}$ months after the therapy.

For ease of representation, the figure is drawn in the form of two normal distributions. No conclusion about the distributions of the scores within studies is intended. In most studies, no information was given about the shape of an individual's scores within treated and untreated groups. We suspect that normality has as much justification as any other form.

The average study showed a .68 standard deviation superiority of the treated group over the control group. Thus, the average client receiving therapy was better off than 75% of the untreated controls. Ironically, the 75% figure that Eysenck used repeatedly to embarrass psychotherapy ap-
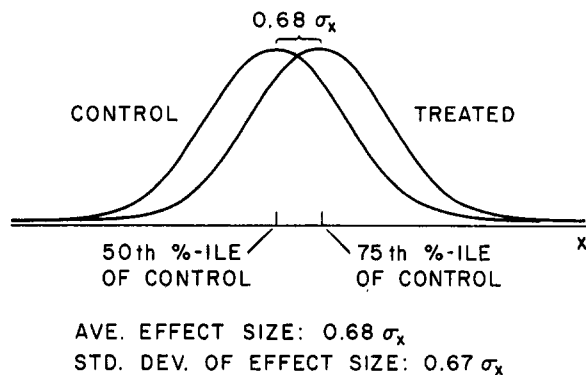
Figure 1. Effect of therapy on any outcome. (Data based on 375 studies; 833 data points.)

TABLE 1: *Conventions for Measurement of the Features of Studies*

Study feature	Value	Study feature	Value
Experience of therapist (when not given)	Lay counselor (0 years) MA candidate (1 year) MA counselor (2 years) PhD candidate or psychiatric resident (3 years) PhD therapist (4 years) Well-known PhD or psychiatrist (5 years)	Type of outcome measure (*continued*)	Work/school achievement: grade point average, job supervisor ratings, promotions. Personality traits: MMPI or other trait inventories, projective test results. Social behavior: dating, classroom discipline, public speaking, information-seeking behavior, sociometrics. Emotional-somatic disorder: frigidity, impotence. Physiological stress: galvanic skin response, Palmer Sweat Index, blood pressure, heart rate.
Diagnosis of client (neurotic or psychotic)	Neurotic unless symptoms or labels clearly indicate otherwise.		
IQ of client (low, average, high)	Average unless identified as otherwise by diagnostic labels (e.g., mentally retarded) or institutional affiliation (college attendance).		
Source of subjects	Clients solicited for purpose of the study. Clients committed to institution, hence to therapy. Clients recognized existence of problem and sought treatment.	Reactivity of measurement	1 (low): Physiological measures; grade point average 2 Projective device (blind); discharge from hospital (blind) 3 Standardized measures of traits (MMPI, Rotter) 4 Experimenter-constructed questionnaires; client's self-report to experimenter; discharge (nonblind); behavior in presence of therapist 5 (high): Therapist rating; projective device (nonblind)
Similarity of therapist and client ("very similar" to "very dissimilar")	College students: very similar Neurotic adults: moderately similar Juveniles, minorities: moderately dissimilar Hospitalized, chronic adults, disturbed children, prisoners: very dissimilar		
Type of outcome measure	Fear, anxiety: Spielberger & Cattell anxiety measures, behavioral approach tests. Self-esteem: inventories, self-ideal correlations, ratings by self and others. Adjustment: adjustment scales, improvement ratings, rehospitalization, time out of hospital, sobriety, symptomatic complaints, disruptive behavior.	Form of publication	Journal Book Thesis Unpublished document
		Internal validity (high, medium, low)	High: Randomization, low mortality Medium: More than one threat to internal validity Low: No matching of pretest information to equate groups

pears in a slightly different context as the most defensible figure on the efficacy of therapy: The therapies represented by the available outcome evaluations move the average client from the 50th to the 75th percentile.

The standard deviation of the effect sizes is *.67*. Their skewness is *+.99*. Only 12% of the 833 effect-size measures from the 375 studies were negative. If therapies of any type were ineffective and design and measurement flaws were immaterial, one would expect half the effect-size measures to be negative.

The 833 effect-size measures were classified into 10 categories descriptive of the type of outcome

being assessed, for example, fear and anxiety reduction, self-esteem, adjustment (freedom from debilitating symptoms), achievement in school or on the job, social relations, emotional-somatic problems, physiological stress measures, etc. Effect-size measures for four outcome categories are presented in Table 2.

Two hundred sixty-one effect sizes from over 100 studies average about 1 standard deviation on measures of fear and anxiety reduction. Thus, the average treated client is better off than 83% of those untreated with respect to the alleviation of fear and anxiety. The improvement in self-esteem is nearly as large. The effect sizes average .9 of a standard deviation. Improvement on variables in the "adjustment" outcome class averages considerably less, roughly .6 of a standard deviation. These outcome variables are measures of personal functioning and frequently involve indices of hospitalization or incarceration for psychotic, alcoholic, or criminal episodes. The average effect size for school or work achievement—most frequently "grade point average"—is smallest of the four outcome classes.

The studies in the four outcome measure categories are not comparable in terms of type of therapy, duration, experience of therapists, number of months posttherapy at which outcomes were measured, etc. Nonetheless, the findings in Table 2 are fairly consistent with expectations and give the credible impression that fear and self-esteem are more susceptible to change in therapy than are the relatively more serious behaviors grouped under the categories "adjustment" and "achievement."

TABLE 2: *Effects of Therapy on Four Types of Outcome Measure*

Type of outcome	Average effect size	No. of effect sizes	Standard error of mean effect size[a]	Mdn treated person's percentile status in control group
Fear-anxiety reduction	.97	261	.15	83
Self-esteem	.90	53	.13	82
Adjustment	.56	229	.05	71
School/work achievement	.31	145	.03	62

[a] The standard errors of the mean are calculated by dividing the standard deviation of the effect sizes (not reported) by the square root of the number of them. This method, based on the assumption of independence known to be false, gives a lower bound to the standard errors (Tukey, Note 2). Inferential techniques employing Tukey's jackknife method which take the nonindependence into account are examined in Glass (in press).

TABLE 3: *Effects of Ten Types of Therapy on Any Outcome Measure*

Type of therapy	Average effect size	No. of effect sizes	Standard error of mean effect size	Mdn treated person's percentile status in control group
Psychodynamic	.59	96	.05	72
Adlerian	.71	16	.19	76
Eclectic	.48	70	.07	68
Transactional analysis	.58	25	.19	72
Rational-emotive	.77	35	.13	78
Gestalt	.26	8	.09	60
Client-centered	.63	94	.08	74
Systematic desensitization	.91	223	.05	82
Implosion	.64	45	.09	74
Behavior modification	.76	132	.06	78

Table 3 presents the average effect sizes for 10 types of therapy. Nearly 100 effect-size measures arising from evaluations of psychodynamic therapy, that is, Freudianlike therapy but *not* psychoanalysis, average approximately .6 of a standard deviation. Studies of Adlerian therapy show an average of .7 sigma, but only 16 effect sizes were found. Eclectic therapies, that is, verbal, cognitive, nonbehavioral therapies more similar to psychodynamic therapies than any other type, gave a mean effect size of about .5 of a standard deviation. Although the number of controlled evaluations of Berne's transactional analysis was rather small, it gave a respectable average effect size of .6 sigma, the same as psychodynamic therapies. Albert Ellis's rational-emotive therapy, with a mean effect size of nearly .8 of a standard deviation, finished second among all 10 therapy types. The Gestalt therapies were relatively untested, but 8 studies showed 16 effect sizes averaging only .25 of a standard deviation. Rogerian client-centered therapy showed a .6 sigma effect size averaged across about 60 studies. The average of over 200 effect-size measures from approximately 100 studies of systematic desensitization therapy was .9 sigma, the largest average effect size of all therapy types. Implosive therapy showed a mean effect size of .64 of a standard deviation, about equal to that for Rogerian and psychodynamic therapies. Significantly, the average effect size for implosive therapy is markedly lower than that for systematic desensitization, which was usually evaluated in studies using similar kinds of clients with similar problems—principally,

simple phobias. The final therapy depicted in Table 3 is Skinnerian behavior modification, which showed a .75 sigma effect size.

Hay's ω^2, which relates the categorical variable "type of therapy" to the quantitative variable "effect size," has the value of .10 for the data in Table 3. Thus, these 10 therapy types account for 10% of the variance in the effect size that studies produce.

The types of therapy depicted in Table 3 were clearly not equated for duration, severity of problem, type of outcome, etc. Nonetheless, the differences in average effect sizes are interesting and interpretable. There is probably a tendency for researchers to evaluate the therapy they like best and to pick clients, circumstances, and outcome measures which show that therapy in the best light. Even so, major differences among the therapies appear. Implosive therapy is demonstrably inferior to systematic desensitization. Behavior modification shows the same mean effect size as rational-emotive therapy.

EFFECTS OF CLASSES OF THERAPY

To compare the effect of therapy type after equating for duration of therapy, diagnosis of client, type of outcome, etc., it was necessary to move to a coarser level of analysis in which data could be grouped into more stable composites. The problem was to group the 10 types of therapy into classes, so that effect sizes could be compared among more general types of therapy. Methods of multidimensional scaling were used to derive a structure from the perceptions of similarities among the 10 therapies by a group of 25 clinicians and counselors. All of the judges in this scaling study were enrolled in a graduate-level seminar. For five weeks, the theory and techniques of the 10 therapies were studied and discussed. Then, each judge performed a multidimensional rank ordering of the therapies, judging similarity among them on whatever basis he or she chose, articulated or unarticulated, conscious or unconscious. The results of the Shepard-Kruskal multidimensional scaling analysis appear as Figure 2.

In Figure 2 one clearly sees four classes of therapies: the ego therapies (transactional analysis and rational-emotive therapy) in front; the three dynamic therapies low, in the background; the behavioral triad, upper right; and the pair of "humanistic" therapies, Gestalt and Rogerian. The average effect sizes among the four classes of
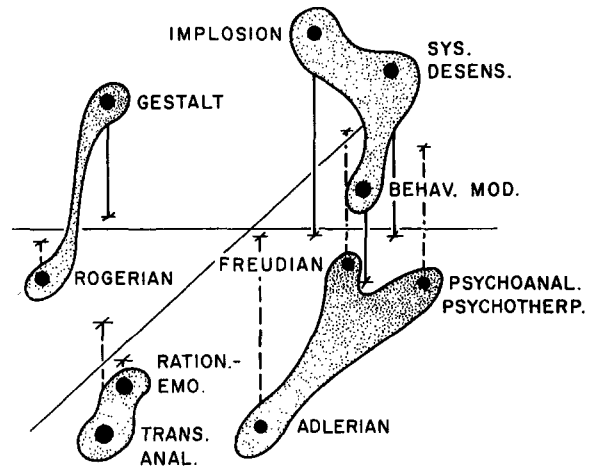

Figure 2. Multidimensional scaling of 10 therapies by 25 clinicians and counselors.

therapies have been compared, but the findings are not reported here. Instead, a higher level of aggregation of the therapies, called "superclasses," was studied. The first superclass was formed from those therapies above the horizontal plane in Figure 2, with the exception of Gestalt therapy for which there was an inadequate number of studies. This superclass was then identical with the group of behavioral therapies: implosion, systematic desensitization, and behavior modification. The second superclass comprises the six therapies below the horizontal plane in Figure 2 and is termed the *nonbehavioral superclass*, a composite of psychoanalytic psychotherapy, Adlerian, Rogerian, rational-emotive, eclectic therapy, and transactional analysis.

Figure 3 represents the mean effect sizes for studies classified by the two superclasses. On the average, approximately 200 evaluations of behavioral therapies showed a mean effect of about $.8\sigma_x$, standard error of .03, over the control group. Approximately 170 evaluations of nonbehavioral studies gave a mean effect size of $.6\sigma_x$, standard error of .04. This small difference ($.2\sigma_x$) between the outcomes of behavioral and nonbehavioral therapies must be considered in light of the circumstances under which these studies were conducted. The evaluators of behavioral superclass therapies waited an average of 2 months after the therapy to measure its effects, whereas the postassessment of the nonbehavioral therapies was made in the vicinity of 5 months, on the average. Furthermore, the reactivity or susceptibility to bias of the outcome measures was higher for the behavioral super-
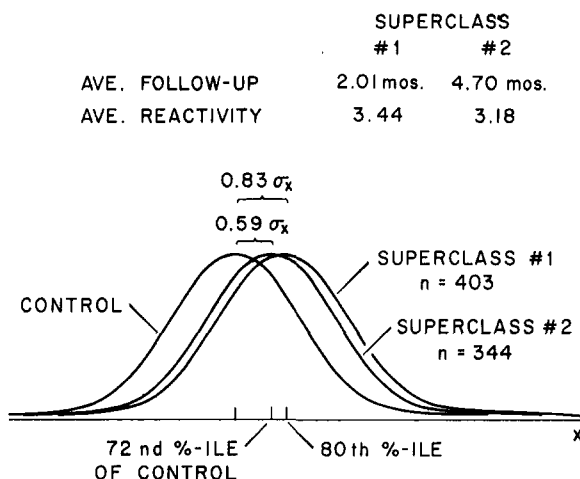
TREATMENT DESCRIPTION

	SUPERCLASS	
	#1	#2
AVE. FOLLOW-UP	2.01 mos.	4.70 mos.
AVE. REACTIVITY	3.44	3.18

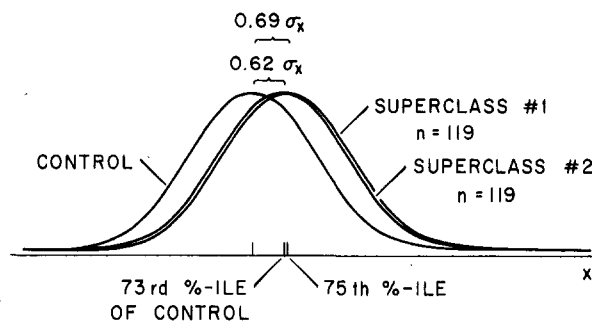Figure 3. Effect of Superclass #1 (behavioral) and Superclass #2 (nonbehavioral).

Figure 4. Effect of Superclass #1 (behavioral) and Superclass #2 (nonbehavioral). (Data drawn only from experiments in which Superclass #1 and Superclass #2 were simultaneously compared with control.)

class than for the nonbehavioral superclass; that is, the behavioral researchers showed a slightly greater tendency to rely on more subjective outcome measures. These differences lead one to suspect that the $.2\sigma_x$ difference between the behavioral and nonbehavioral superclasses is somewhat exaggerated in favor of the behavioral superclass. Exactly how much the difference ought to be reduced is a question that can be approached in at least two ways: (a) examine the behavioral versus nonbehavioral difference for only those studies in which one therapy from each superclass was represented, since for those studies the experimental circumstances will be equivalent; (2) regress "effect size" onto variables descriptive of the study and correct statistically for differences in circumstances between behavioral and nonbehavioral studies.

Figure 4 represents 120 effect-size measures derived from those studies, approximately 50 in number, in which a behavioral therapy and nonbehavioral therapy were compared simultaneously with an untreated control. Hence, for these studies, the collective behavioral and nonbehavioral therapies are equivalent with respect to all important features of the experimental setting, namely, experience of the therapists, nature of the clients' problems, duration of therapy, type of outcome measure, months after therapy for measuring the outcomes, etc.

The results are provocative. The $.2\sigma_x$ "uncontrolled" difference in Figure 3 has shrunk to a $.07\sigma_x$ difference in average effect size. The standard error of the mean of the 119 different scores (behavioral effect size minus nonbehavioral effect size

in each study) is $.66/\sqrt{119} = .06$. The behavioral and nonbehavioral therapies show about the same average effect.

The second approach to correcting for measurable differences between behavioral and nonbehavioral therapies is statistical adjustment by regression analysis. By this method, it is possible to quantify and study the natural covariation among the principal outcome variable of studies and the many variables descriptive of the context of the studies.

Eleven features of each study were correlated with the effect size the study produced (Table 4). For example, the correlation between the duration

TABLE 4: *Correlations of Several Descriptive Variables with Effect Size*

Variable	Correlation with effect size
Organization (1 = individual; 2 = group)	−.07
Duration of therapy (in hours)	−.02
Years' experience of therapists	−.01
Diagnosis of clients	
(1 = psychotic; 2 = neurotic)	.02
IQ of clients	
(1 = low; 2 = medium; 3 = high)	.15**
Age of clients	.02
Similarity of therapists and clients	
(1 = very similar; ... ; 4 = very dissimilar)	−.19**
Internal validity of study	
(1 = high; 2 = medium; 3 = low)	−.09*
Date of publication	.09*
"Reactivity" of outcome measure	
(1 = low; ... ; 5 = high)	.30**
No. of months posttherapy for follow-up	−.10*

* $p \le .05$.
** $p \le .01$.

Independent variable	Unstandardized regression coefficients		
	Psychodynamic $(n = 94)$	Systematic desensitization $(n = 212)$	Behavior modification $(n = 129)$
Diagnosis (1 = psychotic; 2 = neurotic)	.174	−.193	.041
Intelligence (1 = low; ... ; 3 = high)	−.114	.201	.201
Transformed age[a]	.002	−.002	.002
Experience of Therapist × Neurotic	−.011	−.034	−.018
Experience of Therapist × Psychotic	−.015	.004	−.033
Clients self-presented	−.111	.287	−.015
Clients solicited	.182	.088	−.163
Organization (1 = individual; 2 = group)	.108	−.086	−.276
Transformed months posttherapy[b]	−.031	−.047	.007
Transformed reactivity of measure[c]	.003	.025	.021
Additive constant	.757	.489	.453
Multiple R	.423	.512	.509
σ_0	.173	.386	.340

[a] Transformed age $= (\text{Age} - 25)(|\text{Age} - 25|)^{\frac{1}{2}}$.
[b] Transformed months posttherapy $= (\text{No. months})^{\frac{1}{2}}$.
[c] Transformed reactivity of measure $= (\text{Reactivity})^{2.25}$.

of the therapy in hours and the effect size of the study is nearly zero, −.02. The correlations are generally low, although several are reliably non-zero. Some of the more interesting correlations show a positive relationship between an estimate of the intelligence of the group of clients and the effect of therapy, and a somewhat larger correlation indicating that therapists who resemble their clients in ethnic group, age, and social level get better results. The effect sizes diminish across time after therapy as shown by the last correlation in Table 4, a correlation of −.10 which is closer to −.20 when the curvilinearity of the relationship is taken into account. The largest correlation is with the "reactivity" or subjectivity of the outcome measure.

The multiple correlation of these variables with effect size is about .50. Thus, 25% of the variance in the results of studies can be reduced by specification of independent variable values. In several important subsets of the data not reported here, the multiple correlations are over .70, which indicates that in some instances it is possible to reduce more than half of the variability in study findings by regressing the outcome effect onto contextual variables of the study.

The results of three separate multiple regression analyses appear in Table 5. Multiple regressions were performed within each of three types of therapy: psychodynamic, systematic desensitization, and behavior modification. Relatively complex forms of the independent variables were used to account for interactions and nonlinear relationships.

For example, years' experience of the therapist bore a slight curvilinear relationship with outcome, probably because more experienced therapists worked with more seriously ill clients. This situation was accommodated by entering, as an independent variable, "therapist experience" in interaction with "diagnosis of the client." Age of client and follow-up date were slightly curvilinearly related to outcome in ways most directly handled by changing exponents. These regression equations allow estimation of the effect size a study shows when undertaken with a certain type of client, with a therapist of a certain level of experience, etc. By setting the independent variables at a particular set of values, one can estimate what a study of that type would reveal under each of the three types of therapy. Thus, a statistically controlled comparison of the effects of psychodynamic, systematic desensitization, and behavior modification therapies can be obtained in this case. The three regression equations are clearly not homogeneous; hence, one therapy might be superior under one set of circumstances and a different therapy superior under others. A full description of the nature of this interaction is elusive, though one can illustrate it at various particularly interesting points.

In Figure 5, estimates are made of the effect sizes that would be shown for studies in which simple phobias of high-intelligence subjects, 20 years of age, are treated by a therapist with 2 years' experience and evaluated immediately after therapy with highly subjective outcome measures.

ESTIMATED EFFECT SIZES

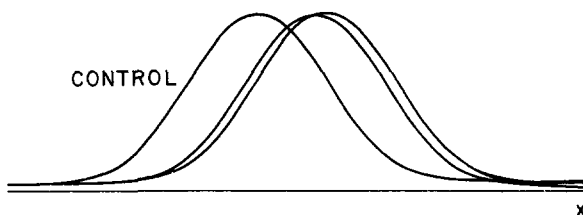|	
PSYCHODYNAMIC	0.919
SYSTEMATIC DESENSITIZATION	1.049
BEHAVIORAL MODIFICATION	1.119

Figure 5. Three within-therapy regression equations set to describe a prototypic therapy client (phobic) and therapy situation.

This verbal description of circumstances can be translated into quantitative values for the independent variables in Table 5 and substituted into each of the three regression equations. In this instance, the two behavioral therapies show effects superior to the psychodynamic therapy.

In Figure 6, a second prototypical psychotherapy client and situation are captured in the independent variable values, and the effects of the three types of therapy are estimated. For the typical 30-year-old neurotic of average IQ seen in circumstances like those that prevail in mental health clinics (individual therapy by a therapist with 5 years' experience), behavior modification is estimated to be superior to psychodynamic therapy, which is in turn superior to systematic desensitization at the 6-month follow-up point.

Besides illuminating the relationships in the data, the quantitative techniques described here can give direction to future research. By fitting regression equations to the relationship between effect size and the independent variables descriptive of the studies

ESTIMATED EFFECT SIZES

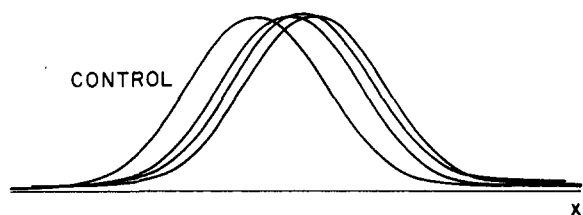|	
PSYCHODYNAMIC	0.643
SYSTEMATIC DESENSITIZATION	0.516
BEHAVIORAL MODIFICATION	0.847

Figure 6. Three within-therapy regression equations set to describe a prototypic therapy client (neurotic) and therapy situation.

and then by placing confidence regions around these hyperplanes, the regions where the input–output relationships are most poorly determined can be identified. By concentrating new studies in these regions, one can avoid the accumulation of redundant studies of convenience that overelaborate small areas.

Conclusions

The results of research demonstrate the beneficial effects of counseling and psychotherapy. Despite volumes devoted to the theoretical differences among different schools of psychotherapy, the results of research demonstrate negligible differences in the effects produced by different therapy types. Unconditional judgments of superiority of one type or another of psychotherapy, and all that these claims imply about treatment and training policy, are unjustified. Scholars and clinicians are in the rather embarrassing position of knowing less than has been proven, because knowledge, atomized and sprayed across a vast landscape of journals, books, and reports, has not been accessible. Extracting knowledge from accumulated studies is a complex and important methodological problem which deserves further attention.

REFERENCE NOTES

1. Glass, G. V, Smith, M. L., & Miller, T. I. *The benefits of psychotherapy*. Book in preparation, 1977.
2. Tukey, J. W. Personal communication, November 15, 1976.

REFERENCES

Bergin, A. E. The evaluation of therapeutic outcomes. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change*. New York: Wiley, 1971.

Ellis, R. H. Letters. *Colorado Psychological Association Newsletter*, April 1977, p. 3.

Emrick, C. D. A review of psychologically oriented treatment of alcoholism. *Journal of Studies on Alcohol*, 1975, *36*, 88–108.

Eysenck, H. J. The effects of psychotherapy: An evaluation. *Journal of Consulting Psychology*, 1952, *16*, 319–324.

Eysenck, H. J. The effects of psychotherapy. *Journal of Psychology*, 1965, *1*, 97–118.

Glass, G. V. Primary, secondary, and meta-analysis of research. *The Educational Researcher*, 1976, *10*, 3–8.

Glass, G. V. Integrating findings: The meta-analysis of research. *Review of Research in Education*, in press.

Luborsky, L., Singer, B., & Luborsky, L. Comparative studies of psychotherapies. *Archives of General Psychiatry*, 1975, *32*, 995–1008.

Meltzoff, J., & Kornreich, M. *Research in psychotherapy*. New York: Atherton, 1970.

Psychotherapy caveat. *APA Monitor*, December 1974, p. 7.