

# Comparative Effects of Short-Term Psychodynamic Psychotherapy: A Meta-Analysis

Martin Svartberg and Tore C. Stiles  
Department of Psychiatry and Behavioural Medicine  
University of Trondheim, Trondheim, Norway

In a review of 19 clinically relevant comparative outcome studies published 1978-1988, short-term psychodynamic psychotherapy (STPP) was evaluated as to overall effects, differential effects, and moderating effects vis-à-vis no-treatment controls (NT) and alternative psychotherapies (AP), respectively. Overall, STPP was superior to NT at posttreatment, inferior to AP at posttreatment, and even more so at 1-year follow-up. STPP was inferior to AP in treating depression and, in particular, to cognitive-behavioral therapy for major depression. STPP was equally successful with mixed neurotics. As research quality increased, STPP grew less superior to NT. Furthermore, STPP decreased its overall superiority over NT and increased its overall inferiority to AT on a series of clinically relevant variables. Improvement in research quality from 1978 to 1988 was noted. Evidence, although limited, supported the view that STPP approaches do seem to differ along a few major dimensions.

Gradually, short-term psychotherapy has come to be the most popular form of psychotherapy (Garfield, 1989; Koss & Butcher, 1986). Among the specific models of short-term psychotherapy, the short-term psychodynamic psychotherapy (STPP) approaches are the most numerous. According to Koss and Butcher (1986), more than 20 STPP variants exist that conceptually originate from psychodynamic or psychoanalytic theory. Over the last 10-15 years, the field of STPP has witnessed the proliferation of new variants that have attempted to enlarge the scope of STPP beyond healthier patients with circumscribed neurotic difficulties. With more or less success, the techniques of STPP have been used with major depressives (Thompson, Gallagher, & Breckenridge, 1987), addicts (Woody et al., 1983), patients with mild personality disorders (Davanloo, 1978; Winston et al., in press), and even patients with severe personality disorders (Lazarus, 1982; Leibovich, 1981).

The efficacy of STPP has been claimed from clinicians (e.g., Davanloo, 1978; Sifneos, 1987), individual outcome studies (e.g., Meyer, 1981; Strupp & Hadley, 1979), and narrative reviews (Koss & Butcher, 1986; Ursano & Hales, 1986), but no comprehensive and quantitative review summarizing the STPP research evidence has yet appeared. However, several quantitative reviews of the general and the more specific psychotherapy literature have reported separate analyses for subsamples of psychodynamic and verbal psychotherapies of short-term duration

(Andrews & Harvey, 1981; Dobson, 1989; Miller & Berman, 1983; Robinson, Berman, & Neimeyer, 1990; Smith & Glass, 1977; Smith, Glass, & Miller, 1980; Shapiro & Shapiro, 1982). All of these analyses have only limited relevance to the field of STPP as a result of either small subsample sizes; atypical patients, therapists, and treatment settings; broad and nonspecific STPP categories; or lack of intentionally planned short-term duration of treatment. Despite very small subsamples, the Nicholson and Berman (1983) meta-analysis represents an exception in that patient samples are clinically relevant and the STPP category relatively pure. They found that dynamic psychotherapy was equally as effective as no treatment (NT) and significantly inferior to alternative psychotherapies (AP) on a sample of individually treated adult neurotics (three studies before 1979). For relatively pure dynamic subsets of studies, Smith and Glass (1977) and Smith et al. (1980) found that dynamic psychotherapy was clearly superior to NT.

Clinically, the STPP approaches vary from the more directive and event-centered variants (e.g., Bellak & Small, 1978; Horowitz, Marmar, et al., 1984) to the more typically interpretive and personality-centered variants (e.g., Davanloo, 1978; Sifneos, 1987). As yet, it is unknown whether diversities of focus, therapeutic characteristics, and time limitation differentially affect the comparative outcome of STPP.

Complementing the narrative reviews of the field, the major aims of the present meta-analysis were (a) to examine the overall effects of STPP at two levels of research design—compared with NT and compared with AP; (b) to examine the differential effects of STPP for certain patient, therapist, and treatment characteristics; and (c) to uncover variables conducive to an increase or decrease in the relative effectiveness of STPP.

## Method

### Studies

In order to be in line with previously published and influential meta-analyses of psychotherapy outcome research (Berman, Miller & Mass-

---

Part of this article was presented at the World Psychiatric Association Regional Symposium on Etiology of Mental Disorder, Oslo, Norway, August 1990.

This research was supported in part by the Norwegian Research Board for Science and the Humanities (Project Number 13.39.66-155).

We wish to thank Finn L. Dybvik, who assisted in the data computations, and J. Öystein Gjörv for secretarial assistance.

Correspondence concerning this article should be addressed to Martin Svartberg, Department of Psychiatry and Behavioural Medicine, University of Trondheim, Östmarka Hospital, Post Office Box 3008, Lade, N-7002 Trondheim, Norway.

man, 1985; Miller & Berman, 1983; Nicholson & Berman, 1983; Shapiro & Shapiro 1982), Martin Svartberg conducted a manual search restricted to studies published in journals and listed in *Psychological Abstracts* and *Index Medicus*, July 1978–July 1988. A total of 19 studies were identified that met the following inclusion criteria: (a) Group designs should include short-term psychodynamic psychotherapy (STPP) as one of the treatment conditions and either an NT group or an AP group or both as the other treatment condition. (b) At least two of the following definitional criteria of psychodynamic psychotherapy should be stated in the report itself or in related references mentioned in the report: The underlying theory of the approach should be psychodynamic or psychoanalytic, the stated goal of the treatment should be the acquisition of insight or the achievement of personality change, and the specific techniques applied should have emphasis on interpretation and transference work. Other criteria were (c) conceptually planned brief duration of treatment ( $\leq 40$  sessions), (d) individually conducted STPP, and (e) nonpsychotic outpatients.<sup>1</sup>

Some descriptive characteristics of the constituent studies are shown in Table 1. To reduce variability of length of the follow-up phase, two follow-up (FU) points, FU1 (*Median* post-FU1 interval 6 months, range 3–9 months) and FU2 (all intervals = 1 year), were established.

The no-treatment condition included patients awaiting (four studies) and not awaiting psychotherapy (five studies). Among the alternative treatments the great majority were psychotherapies of a specific theory-driven nature (cognitive-behavioral, behavioral, experiential). In four studies, however, the approach was not related to any specific orientation and was labeled *nonspecific psychotherapy* (e.g., systematic treatments given by nurses).

In all studies but one (Strupp & Hadley, 1979 university setting), the treatments were conducted in clinical settings. Fifty-five percent of the studies used patients referred from other health service agencies, 18% used patients that presented themselves to the clinic, and in the remaining 27% patients were solicited through advertisement. Among the solicited patients all met clinical diagnostic criteria like those of the *Diagnostic and Statistical Manual of Mental Disorders* (3rd ed., American Psychiatric Association, 1980) or analogous systems. The clinical relevance of the studies is further underscored by the patient, therapist, and treatment characteristics displayed in Table 2. Therapist average educational level for the NT comparisons was at the PhD/psychiatrist level with less than 7 years of postgraduate experience, whereas for the AP comparisons it was at the advanced PhD candidate/psychiatric resident level.

## Classification and Coding

**Methodological quality.** To quantitatively assess the methodological rigor of the studies Martin Svartberg developed the Rating Form for the Quality Assessment of Comparative Outcome Studies (RFQACOS).<sup>2</sup> The RFQACOS consists of five scales entitled Internal Validity, Quality of Therapists, Quality of Treatments, Quality of Outcome Measures, and Statistical Conclusion Validity. The latter also includes an estimate of the adequacy of the sample size used. All but one of the scales are provided with multiple subscales yielding a total of 15 subscales.

In the present meta-analysis, the RFQACOS was used by us in reviewing the 19 studies included. Interrater reliabilities (product-moment correlation coefficient) on the scales and subscales to be used in the subsequent computations of the meta-analysis were mean overall RFQACOS score,  $r = .89$  ( $n = 30$ ); mean Internal Validity score,  $r = .85$  ( $n = 30$ ); mean Quality of the Outcome Measures score,  $r = .92$  ( $n = 30$ ); mean Quality of Therapists score,  $r = .84$  ( $n = 22$ ); mean Quality of Treatments score,  $r = .92$  ( $n = 23$ ); and for two internal validity subscales worth examining separately, Diffusion of Treatment subscale,  $r = .85$  ( $n = 30$ ) and Treatments Equally Valued subscale,  $r = .65$  ( $n = 22$ ).

**Diagnostic categories.** Given the great diversity of patient problems (see Table 1) and the small sample of studies, the number of diagnostic categories was reduced to five by using a modification of Shapiro and Shapiro's (1982) classification scheme. The categories were (a) depression (minor [ $n = 1$ ] and major [ $n = 4$ ];  $n = 1$  for the NT comparisons and  $n = 5$  for the AP comparisons); (b) mixed neuroses ( $n = 5$  for the NT comparisons and  $n = 4$  for the AP comparisons); (c) anxiety ( $n = 1$  for the AP comparisons, only); (d) habit/physical problems (e.g., bulimia, opiate addicts, pain;  $n = 2$  for the NT comparisons and  $n = 4$  for the AT comparisons); and (e) bereavement reactions ( $n = 1$  for both types of comparisons). Categorizations were performed by the authors independently of each other. Percentage agreement was 100.

**Treatment categories.** The alternative treatments were classified into subcategories to allow for comparisons between STPP and more well-defined approaches. The four subcategories were cognitive-behavioral (number of comparisons possible,  $n = 7$ ), behavioral ( $n = 5$ ), experiential ( $n = 2$ ), and nonspecific psychotherapy ( $n = 5$ ). The categorical labels used by the original authors were as far as possible taken as the guiding principle of classification. The classifications into subcategories were performed by the authors independently of each other. Percentage agreement was 100.

**Patient, therapist, and treatment characteristics.** Some selected items from the Smith et al. (1980) coding form (pp. 191–194) were used to collect patient and treatment data. These were IQ, age, sex, marital status, social class, client-therapist similarity, treatment setting, and treatment duration. These codings were performed by Svartberg. Therapist characteristics (i.e., length of clinical experience, length of specific training, and educational level) were collected using the RFQACOS. Scores were consensus scores reached after reviewing the ratings on the Quality of Therapists scale of the RFQACOS.

**STPP characteristics.** Because of restricted diversity of treatment structure, therapeutic focus, and duration among the studies, the four-dimensional classification system for short-term treatments by Burlingame and Fuhrman (1987) was condensed into three major dimensions: focus (event and personality-centered subdimensions), duration (fixed and flexible time limit subdimensions) or therapist style/activity (interpretive of transference or directive and interpretive subdimensions; see Table 1). The coding of the STPP approaches of the constituent studies performed by Svartberg was facilitated by the fact that all but six studies used well-known pioneer STPP models (e.g., Bellak & Small, 1978; Davanloo, 1978; Horowitz, Marmar, et al., 1984; Malan, 1976; Sifneos, 1987) already categorized by Burlingame and Fuhrman (1987).

## Computational Procedures

The effect size estimate used in our meta-analysis was the product-moment correlation coefficient,  $r$ . Through the binomial effect size display (BESD; Rosenthal, 1984),  $r$  can indicate the increase in clinical improvement rates from the poorer to the better outcome groups. As most of the constituent studies did not explicitly report effect size estimates, they had to be computed from the provided tests of signifi-

<sup>1</sup> In retrospect, we conducted a computer search of the Medline and Psych data bases to check the exhaustiveness of the manual search. No additional studies were located that met the inclusion criteria. Studies that systematically combined psychotherapy with psychoactive drug treatments were not included unless the drug regimens were administered on identical terms to the STPP and to the comparative approaches. Three of the studies combined psychotherapy with placebo pills (Study 7) or medical pharmacotherapy (Study 19, methadone to opiate addicts; Study 16, antiulcer medication).

<sup>2</sup> The RFQACOS is available from Martin Svartberg.

Table 1  
*Surveys of Comparative Outcome Studies and Their Characteristics*

Author	Patient population <sup>a</sup>	STPP characteristics			Comparison group <sup>a</sup>	Methodological quality <sup>b</sup>
		Focus	Time limit	Therapist style/activity		
1. Basset & Pilowsky (1985) <sup>c</sup>	Pain	Event-centered	Fixed	Interpretive of transference	Cognitive-supportive	4.35
2. Beutler & Mitchell (1981)	Impulsives and depressives	Personality-centered	Flexible	Interpretive of transference	Experiential	5.70
3. Brockman, Poynton, Ryle, & Watson (1987)	Minor depression	Personality-centered	Fixed	Spontaneous and interactive	Cognitive	5.25
4. Brodaty & Andrews (1983) <sup>d</sup>	Neurotics	Personality-centered	Fixed	Interpretive of transference	Family doctor therapy	5.65
5. Fairburn, Kirk, O'Connor, & Cooper (1986) <sup>cd</sup>	Bulimia nervosa	Event-centered	Flexible	Directive and interpretive	No treatment	4.60
6. Gallagher & Thompson (1982) <sup>cd</sup>	Major depression	Event-centered	Fixed	Directive and interpretive	Cognitive-behavioral	6.25
7. Hersen, Bellack, Himmelhoch, & Thase (1984)	Major depression	Event-centered	Flexible	Directive and interpretive	Social skill plus placebo	6.10
8. Horowitz, Weiss, et al. (1984) <sup>e</sup>	Bereavement reactions	Event-centered	Fixed	Directive and interpretive	No treatment	6.20
9. Manos & Vasilopoulou (1984)	Personality and anxiety disorders	Personality-centered	Flexible	Interpretive of transference	No treatment	4.45
10. Marmar, Horowitz, Weiss, Wilner, & Kaltreider (1988) <sup>cd</sup>	Bereavement reactions	Event-centered	Fixed	Directive and interpretive	Mutual help group treatment	3.45
11. McLean & Hakstian (1979) <sup>e</sup>	Severe depression	Personality-centered	Flexible	Interpretive of transference	No treatment	4.65
12. Meyer (1981) <sup>c</sup>	Mixed neurotics	Personality-centered	Flexible	Interpretive of transference	Cognitive-behavioral	5.70
13. Pierloot & Vinck (1978) <sup>c</sup>	Anxiety	Personality-centered	Fixed	Interpretive of transference	Relaxation	5.25
14. Rosser et al. (1983) <sup>e</sup>	Chronic bronchitis and emphysema	Personality-centered	Fixed	Interpretive of transference	Experiential	4.95
15. Sifneos, Apfel, Bassuk, Fishman, & Gill (1980)	Anxiety and adjustment disorders	Personality-centered	Flexible	Interpretive of transference	No treatment	5.40
16. Sjödén, Svedlund, Ottoson, & Dotevall (1986) <sup>d</sup>	Peptic ulcer disease	Event-centered	Flexible	Directive and interpretive	Systematic desensitization	5.95
17. Strupp & Hadley (1979) <sup>c</sup>	Neurotic anxiety and depression	Personality-centered	Flexible	Interpretive of transference	Supportive	5.60
18. Thompson, Gallagher, & Breckenridge (1987)	Major depression	Event-centered	Flexible	Directive and interpretive	Nurse therapy	6.45
19. Woody et al. (1983) <sup>e</sup>	Opiate addicts	Personality-centered	Flexible	Spontaneous and interactive	No treatment	3.30

Note. STPP = short-term psychodynamic psychotherapy.

<sup>a</sup> Diagnostic and therapy labels reflect those used by the authors. <sup>b</sup> Scores are mean total score on the Rating Form for the Quality Assessment of Comparative Outcome Studies, (range 1-8, 8 = optimal). <sup>c</sup> Study design included follow-up assessment < 1 year (*Mdn* = 6 months) after posttreatment. <sup>d</sup> Study design included follow-up assessment 1-year after posttreatment. <sup>e</sup> Study design included only follow-up assessment.

Table 2  
*Patient, Therapist, and Treatment Characteristics of the Studies*

Characteristic	Comparison			
	STPP versus no treatment		STPP versus alternative treatment	
	<i>M</i>	Range	<i>M</i>	Range
<b>Patient characteristic</b>				
Age (years)	40.8	20-67	40.3	20-67
Female (%)	59.0	47-77	65.0	0-100
Married (%)	51.2	22-70	58.0	20-100
Pretest BDI score <sup>a</sup>			24 <sup>b</sup>	14-28
Duration of complaint before therapy <sup>ac</sup>			50.1	3-132
<b>Therapist characteristic</b>				
Clinical experience (years)	10.4	4-23	5.6	1.5-10
Specific training in the modality (years)	5.6	0-20	3.0	0-11.5
Treatment duration (sessions) <sup>d</sup>	14.1	6-26	15.2	6-28

*Note.* STPP = short-term psychodynamic psychotherapy. BDI = Beck Depression Inventory.  
<sup>a</sup> Too few observations for the no-treatment comparison to allow for computation of means. <sup>b</sup> Indicates moderate to severe depression. <sup>c</sup> In months. <sup>d</sup> Defined as 1 hr long unless less than 40 min. If less, two sessions made up 1 hr.

cance, *t*, *z*, *F*(1,-), or more frequently from the means, standard deviations and sample sizes reported (for details see Rosenthal, 1984, p. 25). Sometimes only a *p*-value and *df* were reported, calling for the use of Formula 2.18 (Rosenthal, 1984). If the report stated only that the finding was significant, a .05 significance level was assumed. In the case of a nonsignificant finding, *Z* and consequently *r* were assumed to equal zero. Effect sizes were computed by Svartberg. However, a sample of 25 computational tasks were randomly drawn to be independently computed by an undergraduate psychology student. Percentage agreement on exact estimates was 84.

All but two studies used multiple outcome measures. To obtain a single summary effect size per treatment comparison per study from the multiple effect sizes generated by these outcome measures, we adopted the procedures proposed by Rosenthal and Rubin (1986).<sup>3</sup>

For each of the two types of treatment comparison (STPP vs. NT and STPP vs. AP), the summary effect sizes of each study were combined into an overall mean effect size weighted by methodological quality (mean RFQACOS scores, Table 1; Rosenthal, 1984, p. 91). The method of adding weighted *Z*s (Rosenthal, 1984, p. 97) was used to estimate significance levels of the overall mean effect size. Table 3 shows all the study summary effect sizes going into the overall mean effect sizes and their corresponding overall *Z*s for both types of treatment comparison at three different test points: posttreatment, 6-month follow-up (FU1), and 12-month follow-up (FU2). The test points performed as unique data points in the analyses. By exception, FU1 and FU2 data were pooled in an additional analysis to yield one single follow-up overall effect size for the STPP versus NT comparisons. When STPP was to be compared with more than one AP (in four studies), means and standard deviations for the APs were averaged and then compared with STPP.

As shown in Table 3, summary effect sizes from multiple studies were summarized by only one parameter estimate, the weighted average of the summary effect size estimates. Whether all the effect sizes in a set were drawn from the same population was examined through a homogeneity test of the effect sizes (Hedges, 1982).<sup>4</sup>

Our meta-analysis was also concerned with the uncovering of potential moderators of overall treatment effectiveness. Traditionally, moder-

ators are of two kinds, substantive (e.g., patient or treatment characteristics) and methodological (e.g., design features). In the search for substantive moderators, methodological moderators were removed first by applying the procedure proposed by Strube (1988) to remove methodology-related variability. The method of linear contrast analysis (Rosenthal, 1984, Formula 4.27) on mean RFQACOS scores was used to compute amount of methodology-related variability. For the NT comparison at posttreatment, the variability was found statistically signifi-

<sup>3</sup> The technical procedure requires that some judgment calls be made. First, the problem of unequal sample sizes was resolved by taking the harmonic mean. Second, because of the acceptable clinical relevance of most outcome measures used in a given study, equal importance was assigned to the measures. Third, the average intercorrelation among outcome measures was set at .50 in studies with only a single informant (e.g., patient report) and at .40 in studies with two types of informants, given that the contribution of the additional informant (e.g., observer ratings) constituted more than one fourth of the total body of information.

<sup>4</sup> Technically, the statistical significance of the homogeneity of the various sets of effect sizes shown in Table 3 was obtained from chi-square tests (Rosenthal, 1984 pp. 77-78). In the case of a significant chi-square test (heterogeneous sample), the set of effect sizes was made homogeneous by excluding one or more effect sizes (studies) from the set (Hedges, 1983). Given the contribution of each study to the chi-square value (Rosenthal, 1984, p. 78), studies to be excluded were easily identified as long as the two following criteria were followed: The least number of studies was excluded, and there was a barely nonsignificant chi-square test. Table 3 shows that two sets of effect sizes at posttreatment and two sets at FU2 were found heterogeneous. Consequently, one study (Study 10, the only study meeting both criteria of exclusion) was excluded from the STPP versus AP comparison at posttreatment, two studies (4 and 15) from the NT comparison at posttreatment, and one study from the FU2 AT comparison set (Study 10) to obtain homogeneous sets of effect sizes. For the NT FU2, no heterogeneity correction was performed because only two studies were involved.

Table 3  
*Effect Sizes and Significance Levels of Treatment Comparisons at Various Time Points*

Study	Posttreatment			STPP versus alternative treatment			6-month follow-up			STPP versus alternative treatment			12-month follow-up			STPP versus alternative treatment		
	STPP versus no treatment			<i>r</i>	<i>Z</i>	<i>N</i>	<i>r</i>	<i>Z</i>	<i>N</i>	<i>r</i>	<i>Z</i>	<i>N</i>	<i>r</i>	<i>Z</i>	<i>N</i>	<i>r</i>	<i>Z</i>	<i>N</i>
1				+0.12	0.34	8				+0.24	0.76	10						
2				-0.40	2.53	40												
3				-0.46	3.19	48												
4	-0.33	2.03	38	0.00	0.00	36							-0.34	2.07	37	-0.33	1.87	32
5				-0.29	1.36	22				-0.30	1.41	22				-0.26	1.22	22
6				-0.30	1.34	20				-0.52	2.33	20				-0.42	1.78	18
7				-0.16	0.96	36												
8							-0.27	2.11	61									
9	+0.34	2.70	66															
10				+0.11	0.86	61				+0.07	0.55	61				+0.16	1.25	61
11				-0.06	0.52	76				-0.04	0.35	75						
12	+0.09	0.61	46	-0.05	0.33	43				-0.01	0.06	35						
13				+0.01	0.05	22				+0.01	0.05	22						
14							+0.13	0.75	33				+0.19	1.06	31			
15	+0.86	4.63	29															
16	+0.12	1.22	103										+0.20	2.01	101			
17	-0.08	0.44	30	+0.30	1.67	31	+0.12	0.71	35	+0.10	0.56	31						
18	+0.09	0.60	43	-0.01	0.08	60												
19										+0.17	1.43	71						
Uncorrected $\bar{r}^a$	+0.16			-0.11			+0.01			-0.03			-0.07			-0.24		
Corrected $\bar{r}^b$	+0.10			-0.12												-0.34		
Overall $Z^c$		1.78			2.50			0.12			0.10			0.06			3.32	
Combined $p^d$		0.04			0.006			0.45			0.46			0.48			0.0005	

Note. STPP = short-term psychodynamic psychotherapy. Signs designate higher (+) and lower (-) effectiveness of STPP group.

<sup>a</sup> Weighted by methodological quality but not corrected for heterogeneity of effect sizes. <sup>b</sup> Weighted by methodological quality and corrected for heterogeneity of effect sizes. <sup>c</sup> Weighted by methodological quality on homogeneous sets of effect sizes. <sup>d</sup> The  $p$ s are one-tailed for the STPP versus no-treatment comparisons and two-tailed for the STPP versus alternative treatment comparisons.

cant ( $Z = 5.05, p = .20 \times 10^{-6}$ ) and nonsignificant for the AP comparison ( $Z = 1.25, p = .10$ ). After the removal of the methodology-related variabilities, the remaining variabilities of the effect sizes for the STPP versus NT comparison at posttreatment and the STPP versus AP comparison at posttreatment were still larger than expected by chance,  $\chi^2(5, N = 355) = 20.25, p = .001, \chi^2(11, N = 503) = 19.7, p = .045$ , respectively. Hence, a search for potential substantive moderators for these two comparisons was warranted. Because methodology-related variability was found statistically significant at least for the STPP versus NT comparison, contrast weights of the linear contrast analyses throughout the substantive moderator examination were determined by multiplying the actual substantive feature score by the methodological quality weight (mean RFQACOS score) of the corresponding study. The sum of the contrast weights in any given analysis, however, was always zero, making it a real orthogonal contrast analysis.<sup>5</sup> An omnibus test like the chi-square test with  $df > 1$  can only tell whether effect sizes differ significantly among themselves. In uncovering substantive moderators, however, it was vital to examine whether effect sizes varied in a predictable or meaningful way with certain patient, therapist, or treatment characteristics. Linear contrast analysis was the procedure chosen to address such focused examinations (Rosenthal, 1984; especially Formula 4.27). As the various potential moderators did not have equally spaced quantitative steps, Robson's (1959) procedure for determining orthogonal polynomial-based contrast weights was chosen. The contrasts were performed on five separate families of hypotheses: (a) methodological features, (b) patient characteristics, (c) duration of treatment, (d) therapist characteristics, and (e) client-therapist similar-

ity. The guidelines given by Grove and Andreasen (1982) for defining a family of hypotheses were followed. As multiple hypotheses within these families were tested simultaneously, Bonferroni correction of alpha level was done separately on these families. Given the exploratory nature of these contrast analyses and the small sample sizes involved ( $n = 3-13$ ), the overall alpha was set at  $p = .10$ , thereby balancing the threats of Type I and Type II errors.

The procedures described by Rosenthal (1984, pp. 107-110) on the basis of adding  $Z$ s were followed in computing a fail-safe  $N$ . The fail-safe  $N$  yields the number of unretrieved null-summing studies that would have to exist to raise the combined probability (the overall  $Z$  values in Table 3) to above  $p > .05$ . Consequently, the fail-safe  $N$  provides a measure of the robustness of the findings in relation to past and future unretrieved studies.

<sup>5</sup> Mean RFQACOS scores were also used to examine, through a linear contrast, whether overall methodological quality of studies had gradually improved from the early to the recently published studies. Contrast weights were determined on the basis of publication year, following Robson's (1959) procedures for unequal spacing of levels of the time variable. For studies that did not include posttreatment assessment, RFQACOS scores at FUI were used instead. Significant trends were demonstrated both for STPP versus NT studies ( $Z = 1.64, p = .05, n = 9$ ) and for STPP versus AP studies ( $Z = 2.21, p = .01, n = 15$ ). This suggests that a significant improvement of overall psychotherapy outcome research quality has taken place over the last 10 years.

## Results

### Overall Effectiveness

Mean overall effect sizes for the comparisons of STPP versus NT and STPP versus AP at posttreatment, 6-month follow-up (FU1), and 1-year follow-up (FU2) are presented in Table 3.

*STPP versus no treatment.* When compared with the NT condition, STPP showed a small-sized ( $\bar{r} = .10$ , heterogeneity-corrected) but significant superiority ( $Z = 1.78$ ,  $p = .04$ , one-tailed) assessed at posttreatment. At FU1 (median length 6 months<sup>6</sup>) and FU2, however, the two conditions were approximately equal ( $\bar{r} = .01$ ,  $Z = .12$ ,  $p = .45$  and  $\bar{r} = -.07$ ,  $Z = .03$ ,  $p = .48$ , respectively). The FU2 mean effect size, however, was not representative of the two study effect sizes; but heterogeneity correction was rendered meaningless given only two effect sizes, and therefore it was not performed. In an additional analysis the effect sizes at FU1 and FU2 were pooled and yielded  $\bar{r} = .01$ . To evaluate the meaning of the posttreatment mean effect size, it was first transformed into Cohen's effect size estimate  $d$  (Cohen, 1988) yielding  $d = .20$  (small). Second, by means of the binomial effect size display (BESD; Rosenthal, 1984),  $\bar{r} = .10$  is associated with an increase in clinical improvement or success rates from 45% (for the NT controls) to 55% (for the STPP patients). Third, the fail-safe  $N$  for the posttreatment comparison was 13 (assumed tolerance level = 59; Rosenthal, 1984), suggesting low robustness to the effects of unretrieved null-effect studies. Furthermore, in a series of  $t$  tests on several patient, therapist, treatment, and methodological characteristics, the FU2 effect sizes were examined as to representativeness of those in the posttreatment sample. The two samples were found identical on all characteristics except for duration of treatment,  $t(7) = 2.74$ ,  $p = .03$ , marginally significant as a result of multiple tests. Mean number of sessions for the FU2 studies was 6.8 (range 6–7.6) and 15.3 (range 6–26) for the posttreatment sample. It may be argued, then, that STPP in the FU2 studies did not have enough time to show its benefits and therefore came out approximately equal to NT.

*STPP versus alternative psychotherapy.* Compared with AP approaches, STPP was inferior at posttreatment ( $\bar{r} = -.12$ ,  $Z = 2.50$ ,  $p = .012$ , two-tailed); not so at 6-month follow-up ( $\bar{r} = -.03$ ,  $Z = .10$ ,  $p = .92$ , two-tailed); but considerably so at 12-month follow-up ( $\bar{r} = -.34$ ,  $Z = 3.32$ ,  $p = .001$ , two-tailed). However, the three studies involved at the 12-month follow-up comparison showed  $\bar{r} = -.21$  at posttreatment, favoring AP more than the entire posttreatment sample did. On the other hand, the FU2 and the posttreatment samples did not differ as to any of the patient, therapist, treatment, and methodology characteristics assessed. To assess the meaning of the effect sizes, Cohen's  $d$  was small (.24) at posttreatment, improvement rates on BESD were from 44% (STPP patients) to 56% (AP patients), and the fail-safe  $N$  was 7 (tolerance level = 75). At 12-month follow-up, Cohen's  $d$  was .72 (close to large), the BESD showed a considerable increase in improvement rates from 33% (STPP patients) to 67% (AP patients), and the fail-safe  $N$  was 6 (tolerance level = 25). For the purpose of assessing the maintenance of therapy gains, a subsample of eight studies with both post and FU1 assessments showed a mean effect size of  $\bar{r} = -.04$  ( $Z = .50$ ,  $p = .60$ ) at posttreatment and  $\bar{r} = -.09$  ( $Z = 1.05$ ,  $p =$

.30) at FU1. At FU2 the three studies involved after heterogeneity correction also included posttreatment assessment. Among these three studies mean effect sizes were  $\bar{r} = -.21$  at posttreatment and  $\bar{r} = -.34$  at FU2. Additionally, there were three studies that provided posttreatment, FU1, and FU2 assessments (Fairburn, Kirk, O'Connor, & Cooper, 1986; Gallagher & Thompson, 1982; and Marmar, Horowitz, Weiss, Wilner, & Kaltreider, 1988). For this subsample, results were  $\bar{r} = -.19$  ( $Z = 1.28$ ,  $p = .20$ ) at posttreatment,  $\bar{r} = -.29$  ( $Z = 2.04$ ,  $p = .04$ ) at FU1, and  $\bar{r} = -.21$  ( $Z = 1.25$ ,  $p = .21$ ) at FU2. All in all this suggests that the comparative treatment gains of AP were well maintained throughout the follow-up period.

### Differential Effectiveness

*By alternative psychotherapy method.* The results are shown in Table 4.<sup>7</sup> The effect of cognitive-behavioral treatment vis-à-vis STPP is striking by its statistical and clinical significance (BESD % 38.5–61.5 at posttreatment and 39–61 at 6-month follow-up), its magnitude ( $d = .47$  and  $.45$ , medium-sized), and its stability over time. (Posttreatment mean effect size for the four follow-up studies was  $\bar{r} = -.22$ .) In additional analyses on the most numerous diagnostic categories only the STPP versus cognitive-behavioral comparison for major depression showed significance ( $\bar{r} = -.22$ ,  $Z = 1.87$ ,  $p < .05$ ,  $n = 3$ , at posttreatment, and  $\bar{r} = -.33$ ,  $Z = 1.88$ ,  $p < .05$ ,  $n = 2$ , at 6-month follow-up; Bonferroni-corrected  $p$ s). For the significant findings in Table 4, the fail-safe  $N$ s were well below the tolerance level, implying sensitivity to the retrieval of null-result studies. Twelve-month follow-up assessment included too few observations to justify computation.

*By diagnostic category.* Four diagnostic categories provided a sufficient number of effect sizes to meaningfully compute mean effect sizes. These categories were depression (minor and major), major depression, mixed neuroses, and habit disorders/physical complaint. The only significant finding that emerged was for the depression category where AP was superior to STPP ( $\bar{r} = -.20$ ,  $Z = 2.59$ ,  $p < .05$ , two-tailed and Bonferroni-corrected). For the other categories, mean effect sizes were very close to the overall estimates at posttreatment and FU1 for both comparison types.

*By STPP characteristics.* Except for fixed time limit for STPP versus NT ( $\bar{r} = -.33$ ,  $n = 1$ ), mean effect sizes for neither STPP versus NT nor STPP versus AP showed statistical differences on any STPP characteristics (see Method). To examine whether one subdimension was superior to the other within the same dimension or whether any combination of characteristics would prove particularly effective, two 2-way analyses of variance (ANOVAs) were performed on study effect sizes of the

<sup>6</sup> FU1 encompassed follow-up phases of varying length. To test the hypotheses that the longer the follow-up phase, the smaller the effect sizes to be expected, a linear contrast analysis was performed yielding  $Z = .53$ ,  $p = .30$ . Thus, length of follow-up phase (for FU1) did not seem to affect the effect sizes significantly.

<sup>7</sup> All the comparisons were performed on independent samples of studies. Hence, Bonferroni correction of alpha levels was not applied.

Table 4  
Comparison of STPP With Different Types  
of Alternative Treatment

Comparison	Posttreatment		6-month follow-up	
	$\bar{r}$	<i>n</i>	$\bar{r}$	<i>n</i>
STPP versus				
cognitive-behavioral	-.23**	6	-.22*	4
STPP versus behavioral	-.05	5	-.40	1
STPP versus experiential	-.26*	2	-.01	1
STPP versus nonspecific	+.14	3	+.14	3

Note. STPP = short-term psychodynamic psychotherapy. When indicated,  $\bar{r}$ s are heterogeneity-corrected. Signs designate higher (+) and lower (-) effectiveness of STPP group.

\*  $p < .05$ , two-tailed. \*\*  $p = .0004$ , two-tailed.

STPP versus AP comparison.<sup>8</sup> The first ANOVA (Type of Focus  $\times$  Type of Time Limit) yielded a nonsignificant main effect for focus,  $F(1, 8) = 3.39$ ,  $p > .05$ ,  $r = .54$ , and for time limit,  $F(1, 8) = 3.78$ ,  $p > .05$ ,  $r = .57$ . However, personality-centered focus was clinically more effective than event-centered (BESD % 23–77) and fixed time limit clinically more effective than flexible (BESD % 21.5–78.5). The interaction effect was negligible,  $F(1, 8) = .06$ ,  $p > .05$ . The second ANOVA (Type of Time Limit  $\times$  Type of Technique) revealed a significant main effect for technique,  $F(1, 8) = 8.63$ ,  $p = .004$ ,  $r = .72$ , BESD % 14–86, with interpretive of transference technique being more effective than directive and interpretive technique. The main effect for time limit was nonsignificant,  $F(1, 8) = 2.05$ ,  $p > .05$ ,  $r = .45$ , BESD % 27.5–72.5, fixed time limit clinically better than flexible time limit, as was the interaction effect,  $F(1, 8) = .68$ ,  $p > .05$ ,  $r = .28$ , BESD % 36–64, Fixed Time Limit  $\times$  Interpretive of Transference clinically better than Flexible Time Limit  $\times$  Directive and Interpretive. All  $p$ s were Bonferroni-corrected. The number of effect sizes was too small to perform a Type of Focus  $\times$  Type of Technique ANOVA and to perform any ANOVAs on the STPP versus NT comparison effect sizes.

### Moderators of Overall Effectiveness

**Methodological moderators.** The contrast procedure described by Rosenthal (1984, p. 84) was used to analyze whether the relative effectiveness of STPP increased or decreased linearly with the methodological features listed to the left in Table 5. As Table 5 reveals, all the methodological moderators tested in the NT comparison showed highly significant linear trends in disfavor of STPP in the entire sample of effect sizes and in the mixed neurotic (MN) subsample.<sup>9</sup> This implies that as the methodological quality of studies increased, the superiority of STPP over NT decreased.

For the STPP versus AP comparison, only the linear trend for the Treatments Equally Valued subscale scores in the entire sample came out statistically significant at  $p < .10$ . The direction was in disfavor of STPP. This implies that as therapist and patient confidence (taken together) in the two approaches grew equal, the inferiority of STPP to AP increased.

**Substantive moderators.** The moderators examined are

listed to the left in Table 6. Like the analysis of methodological moderators, linear contrast procedures were performed, except for social class, IQ, and client-therapist similarity where contrasts between groups were designed (e.g., those scoring 3 vs. those scoring 2).

A striking feature of Table 6 is the dearth of findings that increase the relative STPP effectiveness. Client-therapist similarity (socioeconomic and ethnic) in the entire sample for the NT comparison is the only significant pro-STPP moderator. As Table 6 shows, STPP's overall superiority over NT ( $\bar{r} = .10$ ) is reduced in samples of female patients, or with clinically experienced, or specifically trained, or well-educated therapists. Furthermore, STPP's inferiority to AP ( $\bar{r} = -.12$ ) is increased in samples of young neurotics or with clinically experienced therapists treating mixed neurotic patients.

As to duration of treatment, STPP decreased in effectiveness as number of sessions grew linearly larger for the AP comparison (MN subsample). As some evidence exists indicating that treatment effects fall slightly beyond 12 sessions (Glass & Kiegl, 1983), an additional analysis was performed contrasting effect sizes of briefer variants (<12 sessions) with those of longer variants (>12). It was found that the longer variants significantly increased the superiority of STPP over NT (sample  $Z = 2.97$ ,  $p = .0015$ ,  $ns = 4$  vs. 3; MN subsample  $Z = 3.49$ ,  $p = .0002$ ,  $ns = 4$  vs. 1). For the AP comparison, this analysis was nonsignificant.

### Discussion

Overall, STPP demonstrated a small but significant superiority to waiting-list patients at posttreatment, but a significant and small-sized inferiority to AP at posttreatment and close to a large-sized inferiority at 1-year follow-up. Furthermore, AP showed a clinically significant surge in relative effectiveness beyond posttesting. Thus, the popular incubation theory, which asserts that psychodynamic psychotherapy shows its gains way beyond posttesting, did not receive any support. Whether this also applies to long-term psychodynamic psychotherapy remains an open question. Also, there is evidence that other psychotherapies than STPP are to be preferred in treating depressed patients and that cognitive-behavioral therapy, in particular, is effective with major depression. On the other hand, STPP seems to rival AP when treating mixed neurotic patients unless patients are young or therapists are clinically experienced. Likewise, the case for STPP vis-à-vis NT is weakened with female patients, or experienced, or specifically trained, or well-educated therapists and, vis-à-vis AP, with young neurotics or experienced therapists treating neurotics. Furthermore, the STPP technique of transference interpretation is clearly more effective than the directive and interpretive tech-

<sup>8</sup> Three-way analysis of variance was precluded because of the small number of effect sizes.

<sup>9</sup> The mixed neurotic subsample for both types of treatment comparison and the major depression subsample for the STPP versus AT comparison were the only subsamples large enough to warrant a meaningful examination. However, in the major depression subsample, none of the linear trends examined reached significance.

Table 5  
Methodological Moderators of Overall Effectiveness

Moderator	Comparison			
	STPP versus no treatment		STPP versus alternative treatment	
	Z	n	Z	n
Overall quality				
Sample	-5.05**	7	-1.25	13
MN subsample	-5.06**	5	-2.0	4
Internal validity				
Sample	-4.49**	4	-0.59	13
MN subsample	-4.35**	4	-1.19	4
Treatments equally valued <sup>a</sup>				
Sample			-3.36*	13
MN subsample			-2.15	4
Diffusion of treatments <sup>a</sup>				
Sample			+1.97	13
MN subsample			+1.58	4
Outcome measure quality				
Sample	-6.20**	7	-0.19	13
MN subsample	-6.44**	5	+0.64	4
Treatment quality <sup>a</sup>				
Sample			-1.42	13
MN subsample			-2.13	4
Therapist quality <sup>a</sup>				
Sample			+2.62	13
MN subsample			+2.55	4

Note. STPP = short-term psychodynamic psychotherapy. MN = mixed neurotic. Signs designate pro STPP(+) and con STPP(-) trends.

<sup>a</sup> Effect sizes were too few to compute meaningful trends for the STPP versus no-treatment comparison.

\*  $p < .05$ . \*\*  $p < .001$ , one-tailed for the no-treatment and two-tailed for the alternative treatment comparison and Bonferroni-corrected.

nique. Moreover, the combination of transference interpretation and fixed time limit seems to be associated with clinical benefits surpassing those of the combination of directive and interpretive technique and flexible time limit. However, this is an exploratory finding and should be treated with caution.

Except for the Nicholson and Berman meta-analysis (1983; dynamic vs. NT  $\bar{r} = -.02$ ), STPP's superiority over NT controls is surprisingly small compared with the findings of meta-analyses of the general psychotherapy literature (Andrews & Harvey, 1981; Shapiro & Shapiro, 1982; Smith et al., 1980; Smith & Glass, 1977), whose effect sizes for verbal psychotherapies at posttreatment fall within the range ( $\bar{r}$ ) of .20 to .35. The most obvious reasons why the present finding differs from these meta-analyses may have to do with different therapy categories (broad verbal vs. STPP) or different patients, therapists, and treatment settings. Another reason may pertain to the heterogeneity correction of summary mean effect sizes performed. Unlike these meta-analyses, our study did examine whether a given single overall effect size found was representative of a common underlying effect size for all studies and, furthermore, took measures to correct the overall effect size in case of heterogeneity. As a matter of fact, such a correction reduced the STPP versus NT overall effect size from a nonrepresentative  $\bar{r} = .16$  to a representative  $\bar{r} = .10$ .

Relative to NT controls, STPP's superiority tended to decrease as overall methodological quality, internal validity, and

quality of outcome measures increased. The internal validity finding is consistent with Shapiro's (1985) conclusion summarizing the available psychotherapy data. It is interesting, indeed, that STPP shows its superiority predominantly in methodologically poor studies. This implies that our present knowledge of STPP effectiveness over NT controls seems to be based on rather shaky evidence.

Given the view that outcomes of different psychotherapies seem to be equivalent (Stiles, Shapiro, & Elliott, 1986), the post-treatment and, in particular, the 1-year follow-up findings for the STPP versus AP comparisons are noteworthy. Clinically, patients will increase their chance of improvement (assessed 1-year posttesting) from 33% (STPP) to 67% by having undergone other psychotherapies than STPP. These findings pointing to the STPP inferiority vis-à-vis AP at posttreatment as well as at follow-up are consistent with those reported by Nicholson and Berman (1983). Additionally, STPP seems to be particularly inferior to cognitive-behavioral therapy (CBT) in general as well as in a subsample of major depressive patients. For the 6-month follow-up finding, the success rate increase was from 33.5 (STPP) to 66.5 (CBT), implying that for every STPP patient (major depression) who improved, two CBT patients did so. This superiority of CBT adds to Dobson's (1989) meta-analytic findings for cognitive therapy with depression.

Although the findings from the combination-of-STPP-characteristics contrasts are sparse and inconclusive, the overall ex-



Table 6  
*Treatment-Related Moderators of Overall Effectiveness*

Moderator	Comparison			
	STPP versus no treatment		STPP versus alternative treatment	
	Z	n	Z	n
Patient-related moderator				
Age				
Sample	-0.18	5	+1.11	12
MN subsample	-0.47	3	-2.83*	3
Female %				
Sample	-2.71**	4	+0.21	10
MN subsample	-1.29	3	—	—
Social class				
Sample	-1.89	4 versus 2 <sup>a</sup>	-1.48	—
MN subsample	-2.13	4 versus 1	+2.29	3 versus 1
IQ				
Sample	-1.70	3 versus 4	+0.35	6 versus 7
Severity of depression				
Sample	—	—	+1.58	5
Married %				
Sample	-1.47	3	+1.75	5
Therapist-related moderator				
Clinical experience				
Sample	-3.41**	5	+1.02	12
MN subsample	-3.73***	4	-2.51*	4
Specific training				
Sample	-2.26*	5	0	9
MN subsample	-2.59**	4	-0.50	4
Educational level				
Sample	-6.03***	6	+0.52	13
MN subsample	-6.08***	4	-0.30	4
Client-therapist similarity				
Sample	+1.72*	4 versus 3	-0.36	6 versus 7
MN subsample	+1.16	5 versus 1	—	—
Treatment duration				
Sample	-0.18	7	-1.23	13
MN subsample	+0.22	5	-1.67*	4

Note. STPP = short-term psychodynamic psychotherapy. Dashes imply too small samples to justify computation.

<sup>a</sup> Refers to number of *ns* contrasted.

\*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < .001$ , one-tailed for the no-treatment, two-tailed for the alternative treatment comparison and Bonferroni-corrected.

amination of the STPP characteristics seems sufficiently interesting to justify consideration when designing future comparative outcome studies. Moreover, theoretically posited differences among the major STPP characteristics have been shown to lead to different treatment outcomes. This is a validation of the discriminant properties of the modified version of Burlingame and Fuhriman's (1987) classification system, at least as far as the separate major characteristics are concerned. To test the discriminant validity of the system in terms of the combinations of characteristics will require a larger sample of studies.

Several previous psychotherapy meta-analyses have failed to establish a relationship between duration and effectiveness (Berman et al., 1985; Miller & Berman, 1983; Robinson et al., 1990; Shapiro & Shapiro, 1982). However, in the present meta-

analysis it was found that for treatments longer than 12 sessions, STPP (relative to NT) grew more effective as duration increased, and AP (in MN subsample) increased its relative effectiveness linearly with increasing numbers of sessions. This adds to the meta-analytic finding reported by Bowers and Clum (1988), who found behavior therapy to be more effective relative to placebo treatments as duration increased. Furthermore, across comparison types, length of therapist general clinical experience was negatively correlated to the relative effectiveness of STPP. This was also found by Shapiro and Shapiro (1982) for the more general psychotherapy literature and by Dush, Hirt, and Schroeder (1983) for the more specific literature, whereas no correlation was found by Miller and Berman (1983). On the other hand, Barker, Funk, and Houston (1988) found a positive

correlation between therapist experience and effectiveness of psychotherapy (vis-à-vis NT).

A few limitations of the present meta-analysis should be borne in mind when evaluating the various findings and trends. First, the small number of studies included makes some of the findings tentative. Furthermore, for many of the tests performed, power was quite low, making a small sample size an alternative explanation for nonsignificant findings. This also applies to the heterogeneity corrections performed on sets of effect sizes. Moreover, small sample sizes precluded the use of multivariate analysis to examine the relative contribution of STPP techniques in explaining the overall outcome variance over and above methodological and patient factors. Second, one may question whether the overall effectiveness findings can be generalized from the published studies included in this meta-analysis to the total population of STPP comparative outcome studies. In other words, sampling bias (i.e., publication bias) may represent an alternative explanation for the significant overall results found. To examine such a possibility, we applied the calculation of a fail-safe  $N$  (Rosenthal, 1984) in the present study. The fail-safe  $N$  expresses the number of null-effect studies that would have to exist (in the file drawer) to overturn a significant combined probability value. The robustness or tolerance level to the file drawer threat may vary from one research area to another (Rosenthal, 1984). Comparative outcome research on STPP is an area of research characterized by a very low annual output rate of studies. Consequently, a relatively small number of null-effect studies will probably make us feel confident that the alternative explanation is ruled out. In our study, the range of fail-safe  $N$ s for the significant overall effectiveness findings was from 6 to 13. It would of course have been more reassuring with higher numbers (e.g., 20–30). However, because firm guidelines are lacking, some uncertainty as to the robustness of the findings seems unavoidable. Despite these limitations we hope that the findings of this meta-analysis will stimulate further research experimentation on which combination of STPP characteristics is to be applied by whom to which patient problems.

## References

- American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.
- Andrews, G., & Harvey, R. (1981). Does psychotherapy benefit neurotic patients. *Archives of General Psychiatry*, *38*, 1203–1208.
- Barker, S. L., Funk, S. C., & Houston, B. K. (1988). Psychological treatment versus nonspecific factors: A meta-analysis of conditions that engender comparable expectations for improvement. *Clinical Psychology Review*, *8*, 579–594.
- Basset, D. L., & Pilowsky, I. (1985). A study of brief psychotherapy for chronic pain. *Journal of Psychosomatic Research*, *29*, 259–264.
- Bellack, L., & Small, L. (1978). *Emergency psychotherapy and brief psychotherapy*. New York: Grune & Stratton.
- Berman, J. S., Miller, R. C., & Massman, P. J. (1985). Cognitive therapy versus systematic desensitization: Is one treatment superior? *Psychological Bulletin*, *97*, 451–461.
- Beutler, L. E., & Mitchell, R. (1981). Differential psychotherapy outcome among depressed and impulsive patients as a function of analytic and experiential treatment procedures. *Psychiatry*, *44*, 297–306.
- Bowers, T. G., & Clum, G. A. (1988). Relative contribution of specific and nonspecific treatment effects: Meta-analysis of placebo-controlled behavior therapy research. *Psychological Bulletin*, *103*, 315–323.
- Brockman, B., Poynton, A., Ryle, A., & Watson, J. P. (1987). Effectiveness of time-limited therapy carried out by trainees: Comparison of two methods. *British Journal of Psychiatry*, *151*, 602–610.
- Brodaty, H., & Andrews, G. (1983). Brief psychotherapy in family practice: A controlled prospective intervention trial. *British Journal of Psychiatry*, *143*, 11–19.
- Burlingame, G. M., & Fuhrman, A. (1987). Conceptualizing short-term treatment: A comparative review. *Counseling Psychologist*, *15*, 557–595.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum.
- Davanloo, H. (1978). *Principles and techniques of short-term dynamic psychotherapy*. New York: Spectrum Press.
- Dobson, K. (1989). A meta-analysis of the efficacy of cognitive therapy for depression. *Journal of Consulting and Clinical Psychology*, *57*, 414–429.
- Dush, D. M., Hirt, M. L., & Schroeder, H. (1983). Self-statement modification with adults: A meta-analysis. *Psychological Bulletin*, *94*, 408–422.
- Fairburn, C. G., Kirk, J., O'Connor, M. E., & Cooper, P. J. (1986). A comparison of two psychological treatments for bulimia nervosa. *Behaviour Research and Therapy*, *24*, 629–643.
- Gallagher, D. E., & Thompson, L. W. (1982). Treatment of major depressive disorder in older adult outpatients with brief psychotherapies. *Psychotherapy: Theory, Research and Practice*, *19*, 482–490.
- Garfield, S. L. (1989). *The practice of brief psychotherapy*. New York: Pergamon Press.
- Glass, G., & Kiegl, R. (1983). An apology for research integration in the study of psychotherapy. *Journal of Consulting and Clinical Psychology*, *51*, 28–41.
- Grove, W. M., & Andreasen, N. C. (1982). Simultaneous tests of many hypotheses in exploratory research. *Journal of Nervous and Mental Disease*, *170*, 3–8.
- Hedges, L. V. (1982). Estimation of effect size from a series of independent experiments. *Psychological Bulletin*, *92*, 490–499.
- Hedges, L. V. (1983). Statistical summaries in research integration. *Behavioral and Brain Sciences*, *6*, 295–296.
- Hersen, M., Bellack, A. S., Himmelhoch, J. M., & Thase, M. E. (1984). Effects of social skill training, amitriptyline and psychotherapy in unipolar depressed women. *Behavior Therapy*, *15*, 21–40.
- Horowitz, M. J., Marmar, C., Krupnick, J., Wilner, N., Kaltreider, N., & Wallerstein, R. (1984). *Personality styles and brief psychotherapy*. New York: Basic Books.
- Horowitz, M. J., Weiss, D. S., Kaltreider, N., Krupnick, J., Marmar, C., Wilner, N., & DeWitt, K. (1984). Reactions to the death of a parent: Results from patients and field subjects. *Journal of Nervous and Mental Disease*, *172*, 383–392.
- Koss, M. P., & Butcher, J. N. (1986). Research on brief psychotherapy. In S. L. Garfield & A. E. Bergin (Eds.), *Handbook of psychotherapy and behavior change* (pp. 627–670). New York: Wiley.
- Lazarus, L. W. (1982). Brief psychotherapy of narcissistic disturbances. *Psychotherapy: Theory, Research, Practice*, *19*, 228–236.
- Leibovich, M. A. (1981). Short-term psychotherapy for the borderline personality disorder. *Psychotherapy and Psychosomatics*, *35*, 257–264.
- Malan, D. H. (1976). *The frontier of brief psychotherapy*. New York: Plenum Press.
- Manos, N., & Vasilopoulou, E. (1984). Evaluation of psychoanalytic psychotherapy outcome. *Acta Psychiatrica Scandinavica*, *70*, 28–35.
- Marmar, C., Horowitz, M. J., Weiss, D. S., Wilner, N. R., & Kaltreider, N. B. (1988). A controlled trial of brief psychotherapy and mutual-help group treatment of conjugal bereavement. *American Journal of Psychiatry*, *145*, 203–209.

- McLean, P. D., & Hakstian, A. R. (1979). Clinical depression: Comparative efficacy of outpatient treatments. *Journal of Consulting and Clinical Psychology*, 47, 818-836.
- Meyer, A. E. (Ed.). (1981). The Hamburg short psychotherapy comparison experiment. *Psychotherapy and Psychosomatics*, 35, 81-208.
- Miller, R. C., & Berman, J. S. (1983). The efficacy of cognitive behavior therapies: A quantitative review of the research evidence. *Psychological Bulletin*, 94, 39-53.
- Nicholson, R. A., & Berman, J. S. (1983). Is follow-up necessary in evaluating psychotherapy? *Psychological Bulletin*, 93, 261-278.
- Pierloot, R., & Vinck, J. (1978). Differential outcome of short-term dynamic psychotherapy and systematic desensitization in the treatment of anxious outpatients: A preliminary report. *Psychologica Belgica*, 18, 87-98.
- Robson, D. S. (1959). A simple method for construction of orthogonal polynomials when the independent variable is unequally spaced. *Biometrics*, 15, 187-191.
- Robinson, L. A., Berman, J. S., & Neimeyer, R. A. (1990). Psychotherapy for the treatment of depression: A comprehensive review of controlled outcome research. *Psychological Bulletin*, 108, 30-49.
- Rosenthal, R. (1984). *Meta-analytic procedures for social research*. Beverly Hills, CA: Sage.
- Rosenthal, R., & Rubin, D. B. (1986). Meta-analytic procedures for combining studies with multiple effect sizes. *Psychological Bulletin*, 99, 400-406.
- Rosser, R., Denford, J., Heslop, A., Kinston, W., MacKlin, D., Minty, K., Moynihan, C., Muir, B., Rein, L., & Guz, A. (1983). Breathlessness and psychiatric morbidity in chronic bronchitis and emphysema: A study of psychotherapeutic management. *Psychological Medicine*, 13, 93-110.
- Shapiro, D. A. (1985). Recent applications of meta-analysis in clinical research. *Clinical Psychology Review*, 5, 13-34.
- Shapiro, D. A., & Shapiro, D. (1982). Meta-analysis of comparative therapy outcome studies: A replication and refinement. *Psychological Bulletin*, 92, 581-604.
- Sifneos, P. E. (1987). *Short-term dynamic psychotherapy* (2nd ed.). New York: Plenum Press.
- Sifneos, P. E., Apfel, R. J., Bassuk, E., Fishman, G., & Gill, A. (1980). Ongoing outcome research on short-term dynamic psychotherapy. *Psychotherapy and Psychosomatics*, 33, 233-241.
- Sjödin, I., Svedlund, J., Ottoson, J. O., & Dotevall, G. (1986). Controlled study of psychotherapy in chronic peptic ulcer disease. *Psychosomatics*, 27, 187-200.
- Smith, M. L., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist*, 32, 752-760.
- Smith, M. L., Glass, G. V., & Miller, T. I. (1980). *The benefits of psychotherapy*. Baltimore: Johns Hopkins University Press.
- Stiles, W. B., Shapiro, D. A., & Elliott, R. (1986). Are all psychotherapies equivalent? *American Psychologist*, 41, 165-180.
- Strube, M. J. (1988). Some comments on the use of magnitude-of-effect estimates. *Journal of Counseling Psychology*, 35, 342-345.
- Strupp, H. H., & Hadley, S. W. (1979). Specific vs. nonspecific factors in psychotherapy. *Archives of General Psychiatry*, 36, 1125-1136.
- Thompson, L. W., Gallagher, D., & Breckenridge, J. S. (1987). Comparative effectiveness of psychotherapies for depressed elders. *Journal of Consulting and Clinical Psychology*, 55, 385-390.
- Ursano, R. J., & Hales, R. E. (1986). A review of brief individual psychotherapies. *American Journal of Psychiatry*, 143, 1507-1517.
- Winston, A., Pollack, J., Trujillo, M., McCullough, L., Laikin, M., Flegenheimer, W., & Kastenbaum, R. (in press). Brief dynamic psychotherapy of personality disorders. *Journal of Nervous and Mental Disease*.
- Woody, G. E., Luborsky, L., McLellan, T., O'Brian, C. P., Beck, A. T., Blaine, J., Herman, I., & Hole, A. (1983). Psychotherapy for opiate addicts. *Archives of General Psychiatry*, 40, 639-645.

Received October 1, 1990

Revision received March 11, 1991

Accepted April 16, 1991 ■

### Call for Nominations for the *Journal of Counseling Psychology*

The Publications and Communications (P&C) Board has opened nominations for the editorship of the *Journal of Counseling Psychology* for a 6-year term starting January 1994. Lenore W. Harmon is the incumbent editor.

Candidates must be members of APA and should be available to start receiving manuscripts early in 1993 to prepare for issues published in 1994. Please note that the P&C Board encourages more participation by members of underrepresented groups in the publication process and would particularly welcome such nominees. To nominate candidates, prepare a statement of one page or less in support of each candidate. Submit nominations to

Arthur Bodin  
Mental Research Institute  
555 Middlefield Road  
Palo Alto, California 94301-2124

Other members of the search committee are Nancy Betz, Fred Borgen, Milton Foreman, Lucia Gilbert, and Gail Hackett. First review of nominations will begin January 15, 1992.