

EVALUATING AN ALLIANCE-FOCUSED TREATMENT FOR PERSONALITY DISORDERS

J. CHRISTOPHER MURAN

*Beth Israel Medical Center and
Albert Einstein College of Medicine*

LISA WALLNER SAMSTAG

*Long Island University and
Beth Israel Medical Center*

JEREMY D. SAFRAN

*New School University and
Beth Israel Medical Center*

ARNOLD WINSTON

*Beth Israel Medical Center and
Albert Einstein College of Medicine*

This paper presents a study with the aim of evaluating the relative efficacy of an alliance-focused treatment, brief relational therapy, in comparison to a short-term dynamic therapy and a cognitive-behavioral therapy on a sample of highly comorbid personality disordered patients. Results indicated that the three treatments were equally effective on standard statistical analyses of change, including those conducted on repeated measures and residual gain scores. Some significant differences

were indicated regarding clinically significant change and reliable change, favoring the brief relational and cognitive-behavioral models. There was also a significant difference regarding dropout rates, favoring brief relational therapy.

Keywords: therapeutic alliance, personality disorders, brief psychotherapies, treatment outcome, clinical significance, reliable change, dropout status

J. Christopher Muran, Department of Psychiatry, Beth Israel Medical Center and Albert Einstein College of Medicine; Jeremy D. Safran, Department of Psychology, New School University and Department of Psychiatry, Beth Israel Medical Center; Lisa Wallner Samstag, Department of Psychology, Long Island University at Brooklyn and Department of Psychiatry, Beth Israel Medical Center; and Arnold Winston, Department of Psychiatry, Beth Israel Medical Center and Albert Einstein College of Medicine.

The research was supported in part by a grant from the National Institute of Mental Health MH50246. It was presented at the annual meeting of the Society for Psychotherapy Research, Santa Barbara, California, June 2002.

We wish to acknowledge the contributions of Bernard S. Gorman and all our research assistants. In addition, we would also like to thank our cognitive-behavioral supervisors Arlinza Turner, Elizabeth Ochoa, and Polly Scarvalone, our short-term dynamic supervisors Jerome Pollack, Leslie Baliff, and Michael Laikin, as well as all the therapists and patients who participated.

Correspondence concerning this article should be addressed to J. Christopher Muran, Beth Israel Medical Center, First Avenue at 16th Street, New York, NY 10003. E-mail: jcmuran@bethisraelny.org

Much has been written about the limited ecological validity of the research on many empirically supported treatments that have been tested largely on patients with very circumscribed diagnostic profiles and primarily Axis I conditions (cf. Elliott, 1998). In their critical review of randomized clinical trial studies, Westen and Morrison (2001) found that the exclusion rates of patients who presented with multiple diagnoses typically ranged from 60% to 70%. The vast majority of patients that seek treatment in our mental health clinics or private practices, however, do present with multiple diagnoses with estimates of comorbidity ranging from 40% to 70% (see Kessler et al., 1994). Thus, the extent to which the great body of psychotherapy research conducted on singly diagnosed patients is relevant to practice in real-world settings is arguable.

In contrast, the empirical literature on the treatment of personality disorders (Bateman & Fonagy, 2000; Crits-Christoph, 1998; Perry, Lingiardi, & Ianni, 1999; Piper & Joyce, 2001; Shea, 1993), which is growing but remains limited in

comparison to the research on the treatment of various Axis I disorders, may be less susceptible to such a criticism regarding ecological validity. In addition to the considerable comorbidity found among Axis I conditions (e.g., Rapaport, 2001; Regier, Rae, Narrow, Kaelber, & Schatzberg, 1998; Zimmerman, McDermt, & Mattia, 2000), the literature indicates especially high rates of comorbidity between Axis I and II conditions. In their recent review, for example, Dolan-Sewell, Krueger, and Shea (2001) estimated that the percentage of personality-disordered patients who also meet for an Axis I disorder ranges from 66% to 97%.

As would be expected, several studies have found patients comorbid with a personality disorder to be the most treatment resistant (Chambless, Renneberg, Gracely, Goldstein, & Fydrich, 2000; Mavissakalian & Hamman, 1987; Persons, Burns, & Perloff, 1988; Shea, Pilkonis, Beckham, & Collins, 1990; Turner, 1987). These patients present with longstanding and inflexible patterns of emotional and interpersonal difficulties (Benjamin, 1993; Livesley, 2001; Millon & Davis, 1996), which invariably pose great challenges to psychotherapists, especially with regard to the negotiation of the therapeutic alliance (Benjamin & Karpiak, 2002; Muran, Segal, Samstag, & Crawford, 1994). Therapists are more likely to encounter ruptures in the therapeutic alliance with personality-disordered patients, given the emotional lability or constriction of these patients that makes empathy difficult, as well as their restricted range of interpersonal behavior that elicits certain behavioral responses from therapists, which in turn confirm and perpetuate patients' pathogenic beliefs.

The higher risk for alliance ruptures inherent in the treatment of personality-disordered patients takes on added significance when one considers several well-established findings from the psychotherapy research literature. One such finding is the strong predictive relationship of the therapeutic alliance to treatment outcome (including dropout status), which has been consistently demonstrated (Horvath & Symonds, 1991; Martin, Garske, & Davis, 2000; Samstag, Batchelder, Muran, Safran, & Winston, 1998; Tryon & Kane, 1990, 1993, 1995). Another is that poor outcome cases show greater negative interpersonal process (i.e., hostile and complex interactions) than good outcome cases (e.g., Coady, 1991; Henry, Schacht, & Strupp, 1986; Samstag, 1999;

Weinstock-Savoy, 1986). There is also the indication that differences in individual therapist ability may be critical to treatment success (e.g., Crits-Christoph & Mintz, 1991; Shapiro, Firth-Cozens, & Stiles, 1989) and that the more successful therapists are better able to facilitate the development of a therapeutic alliance (Najavits & Strupp, 1994). These findings are consistent with the growing evidence that the process of recognizing and addressing weakness or ruptures in the therapeutic alliance may play an important role in successful treatment (Foreman & Marmar, 1985; Lansford, 1986; Rhodes, Hill, Thompson, & Elliott, 1994; Safran & Muran, 1996).

The aim of our research program has been to move beyond establishing the predictive validity of the alliance and toward elucidating the processes involved in resolving ruptures in the alliance (Safran & Muran, 1996; Safran, Muran, Samstag, & Stevens, 2002). Our efforts over the years have been geared toward (a) the development and evaluation of a model of rupture resolution through a series of intensive single-cases studies and (b) the development and evaluation of a treatment incorporating interventions informed by the research regarding rupture resolution (see Muran, 2002, for a review). This treatment, referred to as brief relational therapy (BRT), is designed specifically to focus on ruptures in the therapeutic alliance and to treat them as opportunities for change. Since the development of BRT has been informed by our process research on therapeutic alliance ruptures, research regarding its efficacy provides indirect evidence regarding the validity of the rupture resolution model emerging from our process research (e.g., Safran & Muran, 1996). The evaluation of this treatment can thus be understood both as a treatment outcome study and as a verification study of our rupture resolution model.

In this article, we describe the results of a new comparative outcome study from the Beth Israel Brief Psychotherapy Research Program, which was conducted with a sample of personality-disordered patients. This population was selected to increase both the ecological validity of the research and the likelihood of therapeutic alliance ruptures emerging. The patients who did not meet inclusion criteria for our small-scale study of patients at risk for treatment failure, previously described in Safran, Muran, Samstag, and Winston (2005), were admitted into a larger-scale study, described herein. The purpose of this study

is to evaluate the relative efficacy of BRT to a short-term dynamic and a short-term cognitive-behavioral therapy in the treatment of personality disorder.

Methods

Participants

The data analyzed in this study were collected over an 8-year period from 1992 to 2000 and did not include 18 patients (13 in the short-term dynamic therapy, 5 in the cognitive-behavioral therapy) who were determined to be at risk for treatment failure (based on postsession questionnaire ratings completed by patients and therapists on a number of dimensions) and offered the opportunity to be reassigned to another treatment condition in the Safran et al., (2005) study.

Patients. The patients included 60 men and 68 women ($N = 128$), ranging in age from 21 to 65 ($M = 41.33$, $SD = 10.52$). Fifty percent of the patients were single, never married, 28% married or remarried, 21% divorced or separated, and 1% widowed. Nineteen percent were high school graduates, 50% college graduates, and 31% had graduate degrees. Eighty-one percent were employed. Ninety percent were white, 1% black, 5% Latino, and 4% other. Thirty-seven percent were Jewish, 34% Christian, and 29% other. The patients presented with a myriad of longstanding difficulties related to depression, anxiety, and interpersonal functioning. Fifty-five percent met criteria for a current primary diagnosis of Mood Disorder, 28% Anxiety Disorder, 13% V-Code, and 4% Adjustment Disorder on Axis I of *DSM-IV* (APA, 1994); and 35% met for multiple Axis I diagnoses. The principal inclusion criterion was that they met for a diagnosis of Personality Disorder (PD) Cluster C or Not Otherwise Specified (NOS) on Axis II. We focused on Cluster C patients rather than Cluster A or B personality disordered patients because of a concern that patients meeting criteria for A or B diagnoses would be more likely to need a longer term therapeutic approach. Sixty-six percent met criteria for a diagnosis of PD NOS, 22% Avoidant PD, 10% Obsessive-Compulsive PD, and 2% Dependent PD; and 19% met for multiple Axis II diagnoses (see below for how DSM diagnoses were established). Exclusion criteria included evidence of organicity, psychosis, mania, impulse control and compulsive disorder (e.g., current

eating and obsessive-compulsive disorder),¹ and any current substance use disorder; patients on psychotropic medication or in another psychosocial treatment have also been excluded to better isolate the effect of psychotherapy. Patients provided informed consent with respect to the parameters of the research protocol. They also paid a nominal fee per session based on an income-sensitive sliding scale in order to approximate a naturalistic treatment setting, but there was no fee for the intake evaluation.

Therapists. The 128 cases included 70 therapists (36% men and 64% women), ranging in age from 25 to 65 ($M = 38.14$, $SD = 8.89$), who were clinical psychologists (41%), psychiatry attendings (12%) and residents (7%), social workers (11%), and psychology interns (17%) and externs (12%) in a psychiatry department at a major metropolitan medical center. Ninety-six percent of the therapists were White, 2% Asian, and 2% other (Black or Latino). Forty-five percent were single, never married, 50% married or remarried, and 5% divorced or separated. Fifty-nine percent were Jewish, 18% Christian, 8% other, and 15% reported no religious affiliation. Clinical experience averaged 4.77 years ($SD = 7.50$). Ninety-two percent of the therapists had been in personal therapy. Therapists also provided informed consent with respect to the parameters of the research protocol.

Treatment Models & Training Procedure

The three treatment models were manualized and designed to treat personality-disordered patients in a fixed 30-session, one-session-per-week format. Therapists were only permitted to participate in one treatment condition. All the psychotherapy sessions were videotaped. All sessions were conducted at an outpatient mental health clinic in a general hospital setting. Below are brief descriptions of the treatments; more detail is provided in the preceding article which describes the Safran et al. study (2005; see also Muran, 2002).

Short-Term Dynamic Psychotherapy. The

¹ Patients with compulsive disorders, such as eating disorders and OCD, were excluded because it seemed inappropriate to submit such patients to a design where they could be assigned to an analytically informed treatment when the empirical literature so strongly supports cognitive-behavioral approaches to these disorders (e.g., Lambert, 2004).

short-term dynamic psychotherapy (STDP: Pollack, Flegenheimer, & Kaufman, 1992), is based on a standard psychodynamic principles (see Luborsky, 1984; Strupp & Binder, 1984, for similar models). It begins with the establishment of a case formulation in which a major maladaptive pattern is identified. Therapists then help patients gain insight into this pattern through persistent interpretation of transference relationships and resolve the conflict inherent in the pattern by making the unconscious, maladaptive material conscious and thus available to the adaptive orientation of the ego. The therapeutic relationship and ruptures in the therapeutic alliance are construed in terms of transference.

Cognitive-Behavioral Therapy. The cognitive-behavioral therapy (CBT: Turner & Muran, 1992) is a schema-focused model (Beck, Freeman, & Associates, 1990) that also begins with the establishment of a case formulation, in which a core belief system is defined (Persons, 1989). The treatment protocol then entails two intervention phases: (a) Symptom Reduction, in which the Axis I conditions are addressed, and (b) Schema Change, in which core beliefs are modified or restructured. Both phases include traditional cognitive-behavioral strategies, including self-monitoring, cognitive restructuring, behavioral exercises and experimentation. The therapeutic relationship is founded on the principle of "collaborative empiricism" (Beck, Rush, Shaw, & Emery, 1979), and alliance ruptures are typically approached as examples of patients' presenting problems (Newman, 1998).

Brief Relational Therapy. Brief relational therapy (BRT: Muran & Safran, 2002; Safran, 2002; Safran & Muran, 2000) is a model that integrates principles primarily from relational psychoanalysis and humanistic psychotherapy (see Safran et al., 2005, for more detail). It has also been greatly informed by our research on rupture resolution (Safran & Muran, 1996). BRT places greater emphasis on process than the other two models and eschews establishing a case formulation early in treatment. It is oriented toward cultivating awareness of self in relation to other rather than resolving a central conflict or correcting an irrational belief. Its primary task is to track alliance ruptures and engage the patient in a process of metacommunication (e.g., establish a communication about the communication process) in order to bring awareness to bear on what is going on unwittingly in the therapeutic rela-

tionship (in other words, on the implicit negotiation between patient and therapist with regard to their respective needs). BRT is essentially based on a social constructionist model of the therapeutic relationship, whereby ruptures and their resolution are understood as coparticipatory processes involving both patient and therapist.

Training Procedure. All therapists attended a 90-minute weekly case seminar throughout their participation in the study. Each seminar was coded by two senior supervisors. (The first two authors supervised the BRT condition; the supervisors for the other treatment conditions are acknowledged in the author's note.) Therapists also received 1-hour weekly individual supervision by a senior supervisor for their first case. Those who were unlicensed continued in individual supervision. Both supervision formats made extensive use of the videotaped psychotherapy sessions. Training in both STDP and CBT was primarily didactic in orientation. STDP supervision focused on helping therapists to develop dynamic formulations and identify interactions of transference and countertransference, while CBT supervision instructed therapists on establishing a case formulation and developing various cognitive-behavioral change strategies. In contrast, the supervision process in BRT had an experiential orientation involving intensive exploration of the therapist internal experience through the use of various awareness exercises in order to facilitate metacommunication.

Measures & Assessment Procedure

Diagnostic Status. The Structured Interview for DSM-IV-Axis I & II (SCID: First, Spitzer, Gibbon, & Williams, 1995) was used to establish the diagnostic status of each patient. The SCID is a semistructured interview used to determine diagnoses on Axis I and Axis II of DSM-IV (APA, 1994). The interview was administered by research assistants who were first- to fourth-year graduate students in clinical psychology. Training for these students involved viewing a demonstration video, role-playing, observation of a live administration, and completion of an interrater reliability test that consisted of rating various videotaped samples of previous interviews conducted by trained interviewers. These samples included patients presenting with anxiety and mood disorders, as well as various personality disorders. The standard for completing training

was an intraclass correlation $\geq .90$ on both sections (Axis I & II) of a reliability test.

Treatment Fidelity. The extent to which therapists conducted the treatments according to the respective manuals was assessed by the Beth Israel Fidelity Scale (BIFS), a 44-item measure scaled in a Likert-type format and comprising four scales: (a) 12 items were developed to assess interventions associated with BRT, (b) 12 items to assess CBT interventions, (c) 12 items to assess STDP interventions, and (d) 8 items to assess aspects of treatment that cut across theoretical orientations identified as common or nonspecific factors. The instrument has evidenced in two studies sound psychometric properties, including adequate internal consistency, interrater reliability, and discriminant validity (Patton et al., 1998; Santangelo et al., 1994). Research assistants were trained to reliable standards (i.e., intraclass correlation $\geq .90$) to conduct this assessment.

Treatment Outcome. Treatment outcome was assessed on multiple dimensions, including measures of subjective distress or symptomatology, level of adaptive functioning, and interpersonal or personality style, which is consistent with recommendations regarding the assessment of change in personality disorders (Shea, 1997). These outcome measures included the following (see Safran et al., 2005, for a more elaborate description):

Symptom Checklist-90-Revised (SCL-90R: Derogatis, 1983) is a 90-item self-report inventory developed to assess general psychiatric symptomatology. In this study, the Global Severity Index (GSI), which is an overall mean score, was used.

Target Complaints (TC: Battle et al., 1966) is an idiographic self-report instrument developed to assess the particular presenting problems of the patients. Both patients (PTC) and therapists (TTC) independently rate the severity of the problems. In this study, the ratings of the problems were averaged for an overall index.

Global Assessment Scale (GAS: Endicott, Spitzer, Fleiss, & Cohen, 1976) is a clinician-rated scale for evaluating overall mental health. It involves a single rating on a continuum ranging from 1 (*sickest*) to 100 (*healthiest*). All therapists were trained fol-

lowing the authors' protocol to reliable standards (i.e., intraclass correlation $\geq .90$).

Inventory of Interpersonal Problems - 64 (IIP: Horowitz, Alden, Wiggins, & Pincus, 2000) is a 64-item self-report inventory developed to assess patient social adjustment and interpersonal difficulties. In this study, the overall mean score was used to determine outcome.

Wisconsin Personality Inventory (WISPI: Klein, Benjamin, Treece, Rosenfeld, & Greist, 1993) is a 214-item self-report questionnaire scaled in a Likert-type format and derived from an interpersonal perspective on the *DSM-III-R* model of personality disorders (PDs). In this study, an overall mean score was used to determine outcome.

In addition to determining outcome based on repeated ratings of the measures described above, treatment outcome was alternatively determined by treatment completion (i.e., completion of all 30 sessions of the treatment protocol) or patient compliance to protocol. In this regard, cases were categorized as completed or dropped out. Drop-out status was defined as termination before the contracted 30 sessions and unilaterally determined by the patient; premature termination resulting from a change in location of residence were not considered dropout (see Samstag et al., 1998; Wierzbicki & Pekarik, 1993 for rationale regarding definitional parameters).

Assessment Procedure. Patients underwent an intake process including the administration of the SCID interview, the completion of a demographic questionnaire, and a battery of self-report measures described above (SCL-90R, PTC, IIP, and WISPI), which they were asked to complete again at termination and at 6-month follow-up. They were also asked to complete a brief questionnaire after every session regarding session impact and the therapeutic relationship. Data from the postsession questionnaire were not the focus of this particular study (see Muran, 2002, for a review of the research on this questionnaire). Therapists completed a number of measures described above (TTC and GAS) after the third session of treatment for intake and then at termination. As previously mentioned, all psychotherapy sessions were videotaped, which permitted the assessment of treatment fidelity with the BIFS.

TABLE 1. Treatment Fidelity: Means, Standard Deviations, and Results from Four One-Way ANOVAs

	<i>M</i> (<i>SD</i>)	<i>F</i> (2,53)	α	Post hoc analysis
				<i>M</i> difference (<i>SE</i>)
STDP Cases (<i>N</i> = 16)				
STDP Scale	2.05 (0.47)	26.20	.0000	.09 (.10) BRT vs. CBT
CBT Scale	1.33 (0.25)			.72 (.11) STDP vs. CBT*
BRT Scale	1.42 (0.21)			.63 (.11) STDP vs. BRT*
CBT Cases (<i>N</i> = 19)				
STDP Scale	1.15 (0.11)	13.72	.0000	.43 (.10) CBT vs. STDP*
CBT Scale	1.58 (0.47)			.01 (.10) STDP vs. BRT
BRT Scale	1.14 (0.13)			.44 (.09) CBT vs. BRT*
BRT Cases (<i>N</i> = 21)				
STDP Scale	1.52 (0.36)	11.22	.0001	.54 (.14) BRT vs. STDP*
CBT Scale	1.50 (0.44)			.56 (.13) BRT vs. CBT*
BRT Scale	2.05 (0.44)			.02 (.14) STDP vs. CBT
Common Factor Scale				
STDP cases	3.49 (0.18)	.31	.7363	.03 (.18) STDP vs. CBT
CBT cases	3.51 (0.20)			.09 (.15) CBT vs. BRT
BRT cases	3.60 (0.49)			.11 (.17) BRT vs. STDP

Note. STDP = Short-Term Dynamic Psychotherapy; CBT = Cognitive-Behavioral Therapy; BRT = Brief Relational Therapy.

* $p < .05$ (Scheffé Test).

Results

Preliminary Analyses

Patient and Therapist Demographics. No differences were found among the three treatment conditions with regard to patient sex, marital status, education, employment, race, and religion. Likewise, no differences were found with regard to therapist sex, marital status, race, and religion. In addition, there were no differences regarding clinical experience and personal therapy. We did, however, find a statistically significant difference among the training degrees in the treatment conditions, $\chi^2(8, N = 70) = 37.54, p = .001$, with more psychiatry residents in STDP ($n = 8$) and more MA psychology trainees ($n = 17$) in BRT; however, there was no significant relation between training degree and treatment outcome at termination, $F(10, 108) = 1.03, p = .425$, and follow-up, $F(10, 82) = 1.18, p = .323$, or between training degree and dropout, $\chi^2(5, N = 128) = 8.43, p = .134$.²

Treatment Fidelity. Fifty-six of the 128 cases (44%) in this study were randomly sampled to evaluate treatment fidelity (16 STDP, 19 CBT, and 21 BRT). Three sessions were randomly selected from across the three treatment thirds to assess early (Sessions 1–10), middle (11–20), and late (21–30) treatment fidelity on the BIFS. A series of one-way ANOVAs that examined be-

tween treatment differences on the scale scores (which were averaged across the three treatment thirds) indicated treatment fidelity as would be predicted. Table 1 presents the results from these analyses, including post hoc tests. In sum, therapists evidenced higher ratings on the scales designed to measure the treatment model that they were supposed to conduct. There were no differences among the treatment conditions with regard to the common factor scale.

Therapist Effects. In preliminary analyses, we examined therapist effect on treatment outcome. We followed Crits-Christoph and Mintz's (1991) recommendation to adjust p values to greater than .2 or .3 to evaluate the statistical significance of any therapist effect. With regard to cases that completed treatment protocol, we conducted a series of MANOVAs for each treatment condition to test for therapist main effect on two composite outcome indexes (see below for a description). In the STDP condition ($n = 23$), case assignments per therapist averaged 1.78 and ranged from 1–6, and there was no indication of a significant therapist effect, $F(16, 12) = .93, p =$

² When the cell frequencies were small, sparse, or skewed, we computed exact p -values by permutational methods (CYTEL, 1989), we eschewed traditional asymptotic methods of computing statistical significance, which would probably have been invalid (Bishop, Feinberg, & Holland, 1975).

.560. In the CBT condition ($n = 21$), case assignments per therapist averaged 2.19 and ranged from 1–8, and there was no indication of a therapist effect, $F(22, 18) = 1.16, p = .378$. In the BRT condition ($n = 26$), case assignments per therapist averaged 2.19 and ranged from 1–5, and there was no indication of a therapist effect, $F(32, 10) = 1.17, p = .416$. With regard to cases that dropped out before completing treatment protocol, we conducted a series of chi-square analyses, and found no significant effect in the STDP condition $\chi^2(22, n = 41) = 18.55, p = .673$, in the CBT condition $\chi^2(20, n = 46) = 18.46, p = .557$, or in the BRT condition $\chi^2(25, n = 41) = 22.22, p = .623$.

Comparative Analyses of Treatment Outcome

We approached the measurement of change and the question of treatment outcome for the three treatment conditions from several perspectives.

Statistically Significant Change. First, we conducted two between-groups, repeated measures multivariate analyses of variance (MANOVAs),

based on multiple dependent measures from patient and therapist perspectives, and focused on data regarding the patients who completed treatment. Table 2 presents the means and standard deviations on the patient-report and therapist-report measures at intake, termination, and follow-up for each of the three treatment conditions. In one analysis, we examined differences from intake to termination and found a statistically significant main effect for time, $F(6, 53) = 42.24, p = .000$, Effect Size (r) = .91. We found no differences between treatment conditions, although there was an interaction effect that approached statistical significance with a large effect size, $F(12, 106) = 1.77, p = .063$, ES (r) = .55. In a second analysis, we examined differences across intake and termination to follow-up. These included only patient-report measures. The percentage of cases completing follow-up ratings was 61% overall with no significant differences between conditions (56% STDP, 59% CBT, and 65% BRT). We found again a significant main effect for time, $F(8, 130) = 10.12, p = .000$, ES

TABLE 2. Means and Standard Deviations of Raw Scores on Standard Measures of Change

Report	STDP ($n = 22$)		CBT ($n = 29$)		BRT ($n = 33$)		Total ($n = 84$)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Patient Report								
Target Complaints (Mean)								
Intake	10.44	0.97	10.22	1.48	9.29	1.56	9.93	1.47
Termination	7.16	2.06	5.66	2.33	6.44	2.47	6.35	2.36
Follow-up	6.33	2.87	6.67	2.88	5.18	1.62	5.97	2.46
SCL-90R (Global Severity Index)								
Intake	0.83	0.39	1.03	0.45	0.78	0.42	0.88	0.43
Termination	0.76	0.43	0.63	0.43	0.56	0.30	0.64	0.39
Follow-up	0.93	0.75	0.77	0.69	0.64	0.55	0.75	0.65
Inventory of Interpersonal Problems (Mean)								
Intake	1.47	0.49	1.58	0.59	1.37	0.48	1.47	0.52
Termination	1.38	0.60	1.34	0.63	1.21	0.47	1.30	0.56
Follow-up	1.27	0.80	1.39	0.57	1.17	0.53	1.27	0.61
Wisconsin Personality Inventory (Mean)								
Intake	-0.41	0.67	0.02	0.95	-0.26	0.69	-0.20	0.80
Termination	-0.17	0.85	-0.30	1.00	-0.47	0.76	-0.33	0.87
Follow-up	-0.50	1.04	-0.53	0.75	-0.54	0.65	-0.52	0.77
Therapist Report								
Target Complaints (Mean)								
Intake	9.99	1.55	9.51	1.44	10.18	1.30	9.89	1.42
Termination	6.48	1.54	5.47	1.71	6.55	1.90	6.14	1.79
Global Assessment Scale								
Intake	63.31	5.84	64.73	6.70	62.35	5.40	63.46	6.00
Termination	68.19	8.54	72.55	9.24	66.74	6.34	69.21	8.34

Note. STDP = Short-Term Dynamic Psychotherapy, CBT = Cognitive-Behavioral Therapy, BRT = Brief Relational Therapy.

(r) = .79, and a nonsignificant interaction effect of medium magnitude for Time x Condition, $F(16, 199.22) = .71$, $p = .79$, ES (r) = .39.³

Second, in order to increase the precision and reliability of our analysis (see Huck & McClean, 1975), we calculated standardized residual gain scores for all available dependent measures regarding the following intervals: (a) intake to termination, (b) intake to follow-up, and (c) termination to follow-up. We then conducted three separate principal components analyses with varimax rotations of these scores for each interval, with the rationale of reducing the number of dependent measures and thus increasing power. We extracted components with eigenvalues exceeding 1.00 and with a substantial percentage of the variance accounted for.⁴ From each of these analyses, two components emerged: one that could be interpreted as measuring Axis I (symptomatology), the other as measuring Axis II (interpersonal functioning) of *DSM-IV*. We subsequently calculated two outcome composites for each interval (a) first by selecting all gain scores that loaded $\geq .45$ on the two components, (b) then by applying yielded component scores as weights to these gain scores (e.g., multiplying component scores by corresponding gain scores), and (c) finally by averaging the gain scores of the measures for each component. Table 3 presents the means, standard deviations, and results from MANOVAs of these composite indexes. There was no indication of between condition differences, although a difference of a medium effect approached significance with regard to change from intake to termination.

Clinically Significant Change. Third, we addressed the question of clinical significance, or to what extent change indicated a shift to normal functioning. In this regard, we applied the following two of Jacobson and Truax's (1991) formulas to determine cutoff scores. For the two measures for which we had normative data on functional population, the SCL-90R and the IIP, we used the following formula: Patients were considered to have achieved clinically significant change when their level of functioning subsequent to therapy placed them closer to the mean of the functional population than it did to the mean of the dysfunctional population. We used Derogatis' (1983) reported mean score of .31 ($SD = .31$) on the GSI of the SCL-90R for a functional population ($N = 974$), and the intake mean score of .93 ($SD = .49$) of our own per-

sonality disordered sample ($N = 128$) for a dysfunctional population, which yielded a cutoff score of .55.⁵ We used Horowitz et al.'s (2000) reported mean score of .80 ($SD = .54$) on the overall mean of the IIP (64-item version) for a functional population ($N = 800$), and the intake mean score of 1.44 ($SD = .55$) from our own sample ($N = 128$) for a dysfunctional population, which yielded a cutoff score of 1.12. A percentage of patients were in the functional range on both measures at the start of treatment: 26% on the GSI and 29% on the IIP; they were equally distributed among the three treatment conditions. These patients were excluded from subsequent analyses.

Table 4 presents the percentage of cases demonstrating clinically significant change, moving from the dysfunctional to the functional range, on these measures at termination and at follow-up. When we conducted a chi-square analysis of these results, we found no significant difference among the treatment conditions at termination on the SCL-90R, $\chi^2(2, n = 58) = 1.72$, $p = .287$, ES (r) = .17, or on the IIP, $\chi^2(2, n = 56) = .313$, $p = .904$, ES (r) = .08. We also found no significant difference at follow-up on the SCL-90R, $\chi^2(2, n = 44) = .528$, $p = .797$, ES (r) = .11, or on the IIP, $\chi^2(2, n = 40) = 1.13$, $p = .667$, ES (r) = .17.

For the other measures for which we did not have normative data on a functional population, we applied the following formula: Patients were considered to have achieved clinically significant

³ In ancillary analyses, we compared the results regarding the STDP cases conducted in this study ($n=23$) to those conducted in a previous study ($n=24$), which demonstrated its efficacy in comparison to a waitlist control condition ($n=26$; Winston et al., 1994). We conducted repeated measures MANOVAs on the measures administered in both studies and found no significant difference on the SCL-90R and Target Complaints (patient-rated) at termination, $F(2, 40) = .849$, $p = .435$, ES (r) = .20, and on the Target Complaints at followup, $F(1, 30) = .141$, $p = .710$, ES (r) = .07. Because dropout data was not available from the previous study, we could not conduct a comparative analysis of dropout rates. The results from these analyses thus suggest no difference in the effective application of STDP.

⁴ Readers can contact the first author for more details regarding the results of the principal components analyses.

⁵ Given the difference in variance between the functional and dysfunctional populations, the formula $(SD1)(M2) + (SD2)(M1)/SD1 + SD2$ was applied to determine the cutoff score (Jacobson, Follette, & Revenstorf, 1984).

TABLE 3. Means, Standard Deviations, and Results from MANOVAs of Standardized Residual Gain Scores for Two Component Model of Change

	STDP (n = 22)		CBT (n = 29)		BRT (n = 33)		Univariate analyses		Multivariate analyses	
	M	SD	M	SD	M	SD	F (df)	α	F (df)	α
Intake-to-Termination										
Component 1	0.19	0.91	-0.45	1.08	0.09	0.84	2.98(2, 59)	.059	2.38(4, 114)	.056
Component 2	0.51	0.82	-0.26	1.39	-0.17	0.77	1.58(2, 68)	.213		
Intake-to-Follow-up										
Component 1	0.24	1.33	0.13	1.28	-0.34	0.59	0.76(2, 34)	.475	1.16(2, 37)	.336
Component 2	0.31	1.28	-0.30	1.02	-0.12	0.96	0.44(2, 40)	.649		
Termination-to-Follow-up										
Component 1	.014	1.23	0.26	1.23	-0.20	0.59	1.19(2, 38)	.315	0.45(4, 66)	.774
Component 2	-0.11	0.93	0.19	0.88	-0.08	1.38	0.92(2, 45)	.405		

Note. STDP = Short-Term Dynamic Psychotherapy; CBT = Cognitive-Behavioral Therapy; BRT = Brief Relational Therapy; ES = Effect Size.

change when their level of functioning subsequent to treatment fell outside the range of the dysfunctional population, where range was defined as extending two standard deviations beyond (in the direction of functionality) the mean for that population. Again the dysfunctional population was defined by the intake scores of our sample with the average scores for the overall WISPI mean equaling $-.20$ ($SD = .98$), for the overall PTC mean 9.88 ($SD = .149$), for the overall TTC mean 9.69 ($SD = 1.64$), and for the GAS 63.18 ($SD = 6.51$). The derived cutoff scores were -2.16 for the WISPI, 6.90 for the PTC, 6.41 for the TTC, and 76.20 for the GAS.

Table 4 also presents the percentage of cases demonstrating clinically significant change on these measures. Chi-square analyses indicated no difference on the WISPI at termination χ^2 incalculable ($n = 72$) or follow-up χ^2 incalculable ($n = 51$). There was a difference on the PTC at termination, $\chi^2(2, n = 72) = 9.937, p = .007, ES(r) = .37$. Post hoc analyses indicated that this difference was accounted for by significant differences between STDP and CBT, $\chi^2(1, n = 45) = 8.091, p = .007, ES(r) = .42$, as well as between STDP and BRT, $\chi^2(1, n = 45) = 14.40, p = .000, ES(r) = .60$. There was no difference,

TABLE 4. Percentage of Cases with Clinically Significant Change

Measurement	STDP	CBT	BRT
SCL-90R (GSI)			
Termination	27	50	52
Follow-up	31	40	44
IIP (Mean)			
Termination	31	40	44
Follow-up	18	7	7
WISPI (Mean)			
Termination	0	0	0
Follow-up	0	0	0
PTC (Mean)			
Termination	26	69	63
Follow-up	57	63	86
TTC (Mean)			
Termination	52	60	66
GAS			
Termination	24	31	10

Note. STDP = Short-Term Dynamic Psychotherapy; CBT = Cognitive-Behavioral Therapy; BRT = Brief Relational Therapy; SCL-90R (GSI) = Symptom Checklist-90 Revised (Global Severity Index); IIP = Inventory of Interpersonal Problems; WISPI = Wisconsin Personality Inventory; PTC = Patient Target Complaints; TTC = Therapist Target Complaints; GAS = Global Assessment Scale.

however, on the PTC at follow-up, $\chi^2(2, n = 51) = 4.037, p = .149, ES(r) = .28$. Finally, no differences were found at termination on the TTC, $\chi^2(2, n = 80) = .876, p = .691, ES(r) = .11$, and the GAS, $\chi^2(2, n = 80) = 3.764, p = .170, ES(r) = .22$.

Reliable Change. Fourth, we also calculated Reliable Change Indexes (RCIs; Jacobson & Truax, 1991) in order to establish statistically reliable criteria accounting for measurement error and indicating how much change has occurred. The RC coefficient equals the difference between two test scores divided by the standard error of the difference between the scores, which is derived from test-retest reliability of a measure and standard deviation of pretreatment scores on the measure. An RC coefficient greater than 1.96 is considered unlikely to occur ($p < .05$) without actual change. We assessed reliable change on the GSI, the overall IIP mean, and the overall WISPI mean because of the availability of test-retest reliability to calculate RCIs. The standard error of the difference (S_{diff}) scores⁶ used in the present study were .28 ($M = .93; SD = .49, N = 128$) for the GSI, .37 ($M = 1.45; SD = .55, N = 128$) for the IIP, and .48 ($M = -.20; SD = .98, N = 128$) for the WISPI. These were derived from the intake scores of our 128 cases and test-retest reliability coefficients reported by Derogatis (.84, $N = 94$),⁷ by Horowitz et al. (.78, $N = 60$), and by Klein et al. (.88, $N = 80$).

Table 5 presents the percentage of cases with reliable change on these measures at termination and at follow-up. When we conducted a chi-square analysis, we found no significant difference among the three treatment conditions at termination on the SCL-90R, $\chi^2(2, n = 72) = 3.291, p = .211, ES(r) = .21$, or on the WISPI, $\chi^2(2, n = 71) = 2.259, p = .445, ES(r) = .18$, but did so with regard to the IIP, $\chi^2(2, n = 72) = 6.932, p = .027, ES(r) = .31$. Post hoc analyses indicated this difference was accounted for by a significant difference between STDP and CBT, $\chi^2(1, n = 45) = 6.058, p = .031, ES(r) = .37$. When we conducted an analysis of reliable change at follow-up, we found no significant difference among the three treatment conditions on the SCL-90R, $\chi^2(2, n = 51) = 3.100, p = .230, ES(r) = .25$, on the IIP, $\chi^2(2, n = 52) = .184, p = .999, ES(r) = .06$, or on the WISPI, $\chi^2(2, n = 49) = 1.768, p = .482, ES(r) = .19$.

Dropout Status. In a final assessment of treatment outcome, we examined dropout rates

TABLE 5. Percentage of Cases with Reliable Change (RCI ≥ 1.96)

	STDP	CBT	BRT
SCL-90R (GSI)			
Termination	16	38	22
Follow-up	14	31	10
IIP (Mean)			
Termination	0	27	11
Follow-up	7	6	10
WISPI (Mean)			
Termination	0	15	12
Follow-up	0	12	5

Note. STDP = Short-Term Dynamic Psychotherapy; CBT = Cognitive-Behavioral Therapy; BRT = Brief Relational Therapy; SCL-90R (GSI) = Symptom Checklist-90 Revised (Global Severity Index); IIP = Inventory of Interpersonal Problems; WISPI = Wisconsin Personality Inventory.

and found a statistically significant difference among the three treatment conditions, $\chi^2(2, N = 128) = 6.75, p = .03, ES(r) = .23$. Forty-six percent of the patients (19/41) dropped from STDP, 37% (17/46) from CBT, and 20% (8/41) from BRT: that is 34% overall (44/128). Post hoc statistical analyses indicated the difference was between STDP and BRT, $\chi^2(1, n = 82) = 6.68, p = .01, ES(r) = .29$, and only approached statistical significance between CBT and BRT, $\chi^2(1, n = 87) = 3.32, p = .07, ES(r) = .20$.

Discussion

In general, the results from traditional statistical strategies employing a repeated measures or residualized gain approach demonstrated that the three treatments, STDP, CBT, and BRT, were equally effective for a sample of patients who were highly comorbid on Axis I and II (e.g., 87% with both Axis I and II disorders, V-codes not inclusive). These results suggest that these treatments can be considered *possibly efficacious* for Cluster C and NOS personality disorders, according to the criteria established for empirically supported psychological therapies (Chambless & Hollon, 1998). They in part replicate findings

⁶ S_{diff} = square root of 2 (standard error of measurement S_e)². $S_e = SD$ (square root of $1 - r_{xx}$).

⁷ This test-retest reliability coefficient represents an average of the coefficients reported for each of the nine SCL-90R subscales, because no such coefficient was reported for the GSI by Derogatis (1983).

from a previous study in our research program (Winston et al., 1994), in which the efficacy of STDP was demonstrated in comparison to a waitlist control condition (although assignment was nonrandom). Accordingly, STDP could be construed in the present study as a criterion condition against which the efficacy of CBT and BRT was measured. Thus the results from the Winston et al. study provide further support for the *possibly efficacious* designation of all three treatments. For an *efficacious* designation, these treatments will have to be evaluated and evidence efficacy in an independent research setting.

Although no significant differences were found among the three treatments on the repeated measures or residualized gain analyses, a number of medium-to-large effect sizes were yielded that approached statistical significance, which may translate into between-condition differences with a larger sample size. As to where the differences may be remains an open question, since lack of significance precluded post hoc analyses in this study. There is also the fact that 18 cases (13 from STDP and 5 from CBT), which were assessed as potential treatment failures in the Safran et al., (2005) study, were excluded from these analyses. At a closer inspection of the Safran et al. data set, 7 of 8 patients who rejected reassignment and remained in STDP eventually dropped out, while 3 patients in CBT reassigned to STDP and 2 patients in STDP reassigned to CBT subsequently dropped out. Thus it is reasonable to assume that the exclusion of these cases probably resulted in an inflated estimate of the efficacy of STDP and CBT, although it is impossible to know to what extent.

With regard to clinically significant change, no significant differences were found among the three treatments on any of the measures, except the PTC, where we found differences favoring BRT and CBT over STDP. Of course, it is important to bear in mind that this was based on a cutoff score indicating only movement out of a dysfunctional range. On the SCL-90R and IIP, for which we could establish a functional range, it is interesting to note that a little more than a quarter of the patients fell within the functional distribution at intake, even though they were diagnosed as personality disordered. This appears consistent with findings from the few other studies that report clinical significance data on personality-disordered patients (e.g., Piper, Joyce, McCallum, & Azim, 1998). The implica-

tions of this finding are not entirely clear, but it does suggest that severity of distress and character pathology are not necessarily highly correlated. It is also interesting to note the percentage of cases that actually resulted in clinically significant change. With regard to some measures, the percentages were consistent with other studies, such as the Piper et al. study in which about 60% of the patients received an Axis II disorder diagnosis (n.b., 100% of the patients in our sample were personality disordered). With regard to other measures, the percentages were low: This was true on the WISPI at termination and the IIP at follow-up. Given the fact that both of these measures assess interpersonal functioning to a greater extent than the other measures, these findings may suggest that this domain is particularly difficult to change in a personality-disordered population in short-term therapy. Another possible factor is that the distribution of scores for our sample at intake (i.e., the extent of variance or standard deviation) resulted in unrealistic cutoff scores for clinical significance.

As for reliable change, although CBT was significantly more effective than STDP on the IIP at termination, for the most part there were no significant differences. The percentage of cases achieving reliable change equaling or exceeding 1.96 was relatively low compared to Piper et al. (1998). Once again, this was especially true for the IIP and the WISPI, and may reflect the difficulty of effecting characterological change with personality-disordered patients in short-term therapy. And once again, these findings may have been influenced by the variance among the scores assessed at intake in our sample, which resulted in substantial standard errors of difference scores and thus fewer instances of reliable change (≥ 1.96). It is also important to note that the measures used to determine reliable change were exclusively patient-report, which is a limitation.

Although there was some suggestion that STDP was less effective than the other two treatments in some analyses and on some measures, it is worth noting the evidence (based on analyses of two measures) that the STDP treatment appeared as effective as it was in the Winston et al., (1994) study. This suggests that any inferiority in the performance of the STDP is not evidently attributable to a deterioration in the quality of the treatment in this particular study. By and large, however, the three treatments should be interpreted as equally effective for patients complet-

ing the protocol. This finding is consistent with the common difficulty in finding differences in treatment efficacy (e.g., Lambert, 2004; Luborsky, Singer, & Luborsky, 1975). It also should be noted that this finding was based on standard measures of change; it remains an open question whether there would be treatment differences with theory-specific change measures.

Finally, there was a significant difference between the conditions on dropout status or patient retention. In this regard, BRT (20%) was superior to STDP (46%). Although the difference between BRT (20%) and CBT (37%) only approached significance, the difference of 17% seems clinically compelling, given the seriousness of premature termination. The dropout rates for STDP and CBT were consistent with the literature. In their meta-analysis of 125 studies, Wierzbicki and Pekarik (1993) reported a mean dropout rate was 46.86% ($SD = 22.25$), with the 95% confidence interval for the estimate of the mean dropout rate at 42.9% to 50.82%. When evaluating this result, one should also bear in mind that of the 18 STDP or CBT cases who were screened out of this study and offered reassignment in the Safran et al. study (2005), the 12 of the 13 that either remained in their original treatment or were reassigned to CBT or STDP ultimately dropped out. The relative success of BRT in retaining patients could be attributed to its intensive focus on alliance ruptures and thus seen as providing some support for the rupture resolution model on which it is largely based.

Even though this study was embedded in an outpatient mental health clinic and focused on a highly comorbid patient population, the trappings of a psychotherapy research program (including the completion of questionnaires, videotaping, supervision, etc.) may diminish the generalizability of our findings to more naturalistic settings. Others factors that may limit generalizability include the lack of racial/ethnic diversity in both our patients and therapists, as well as the focus on only Cluster C and NOS Personality Disorders, although these are the most prevalent of the personality disorder classifications (Mattia & Zimmerman, 2001). Additional limitations include our modest sample size and sporadic missing data that limited the power of our repeated measures analyses. Although our use in this study of standard measures widely used to measure change facilitates comparison with other outcome studies, it may not facilitate the assessment of

between-treatment differences as would theory-specific measures designed to capture change dimensions unique to a particular treatment model (e.g., as certain insight is to a psychoanalytically informed treatment; see Muran, 2002 for elaboration). Future studies should take all these issues into consideration.

References

- AMERICAN PSYCHIATRIC ASSOCIATION. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- BATEMAN, A. W., & FONAGY, P. (2000). Effectiveness of psychotherapeutic treatment of personality disorder. *British Journal of Psychiatry*, *177*, 138–143.
- BATTLE, C. C., IMBER, S. D., HOEHN-SARIC, R., STONE, A. R., NASH, E. R., & FRANK, J. D. (1966). Target complaints as criteria of improvement. *American Journal of Psychotherapy*, *20*, 184–192.
- BECK, A. T., FREEMAN, A. A., & ASSOCIATES (1990). *Cognitive therapy of personality disorders*. New York: Guilford Press.
- BECK, A. T., RUSH, J., SHAW, B., & EMERY, G. (1979). *Cognitive therapy of depression*. New York: Guilford Press.
- BENJAMIN, L. S. (1993). *Diagnosis and treatment of personality disorders: A structural approach*. New York: Guilford Press.
- BENJAMIN, L. S., & KARPIAK, C. P. (2002). Personality disorders. *Psychotherapy: Theory, Research, Practice, and Training*, *38*, 487–491.
- BISHOP, Y. M., FEINBERG, S. E., & HOLLAND, P. W. (1975). *Discrete multivariate analysis: Theory and practice*. Cambridge, MA: MIT Press.
- CHAMBLESS, D. L., & HOLLON, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology*, *66*, 7–18.
- CHAMBLESS, D. L., RENNEBERG, B., GRACEY, E. J., GOLDSTEIN, A. J., & FYDRICH, T. (2000). Axis I and II comorbidity in agoraphobia: Prediction of psychotherapy outcome in a clinical setting. *Psychotherapy Research*, *10*, 279–295.
- COADY, N. (1991). The association between client and therapist interpersonal processes and outcomes in psychodynamic psychotherapy. *Research on Social Work Practice*, *1*, 122–138.
- CRITS-CHRISTOPH, P. (1998). Psychosocial treatments for personality disorders. In P. E. Nathan & J. M. Gorman (Eds.), *A guide to treatments that work*. (pp. 544–553). New York: Oxford University Press.
- CRITS-CHRISTOPH, P., & MINTZ, J. (1991). Implications of therapist effects for the design and analysis of comparative studies of psychotherapies. *Journal of Consulting and Clinical Psychology*, *59*, 20–26.
- CYTEL. (1989). *StatXact: A statistical package for non-parametric inference*. Cambridge, MA: Author.
- DEROGATIS, L. R. (1983). *The Symptom Checklist -90 Revised: Administration, scoring and procedures manual II*. Baltimore: Clinical Psychometric Research.
- DOLAN-SEWELL, R. T., KRUEGER, R. F., & SHEA, M. T. (2001). Co-occurrence with syndrome disorders. In

- W. J. Livesley (Ed.), *Handbook of personality disorders: Theory, research, and treatment* (pp. 84–104). New York: Guilford Press.
- ELLIOTT, R. (1998). Ed.'s introduction: A guide to the empirically supported treatments controversy. *Psychotherapy Research*, 8, 115–125.
- ENDICOTT, J., SPITZER, R. L., FLEISS, J. L., & COHEN, J. (1976). The Global Assessment Scale: A procedure for measuring overall severity of psychiatric disturbance. *Archives of General Psychiatry*, 33, 766–771.
- FIRST, M. B., SPITZER, R. L., GIBBON, M., & WILLIAMS, J. B. W. (1995). *Structured Clinical Interview for DSM-IV*. New York: Biometrics Research Department, New York Psychiatric Institute.
- FOREMAN, S. A., & MARMAR, C. R. (1985). Therapist actions that address initially poor therapeutic alliances in psychotherapy. *American Journal of Psychiatry*, 142, 922–926.
- HENRY, W. P., SCHACHT, T. E., & STRUPP, H. H. (1986). Patient and therapist introjects, interpersonal process and differential psychotherapy outcome. *Journal of Consulting and Clinical Psychology*, 58, 768–774.
- HOROWITZ, L. M., ALDEN, L. E., WIGGINS, J. S., & PINCUS, A. L. (2000). *Inventory Of Interpersonal Problems*. New York: The Psychological Corporation.
- HORVATH, A. O., & SYMONDS, B. D. (1991). Relation between working alliance and outcome in psychotherapy: A meta-analysis. *Journal of Consulting and Clinical Psychology*, 38, 139–149.
- HUCK, S. W., & MCCLEAN, R. A. (1975). Using a repeated measures ANOVA to analyze the data from a pretest–posttest design: A potentially confusing task. *Psychological Bulletin*, 82, 511–518.
- JACOBSON, N. S., FOLLETTE, W. C., & REVENSTORF, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behavior Therapy*, 15, 336–352.
- JACOBSON, N. S., & TRUAX, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59, 12–19.
- KESSLER, R. C., MCGONAGLE, K. A., ZHAO, S., NELSON, C. B., HUGHES, M., ESHLEMAN, S., WITTCHEN, H.-U., & KENDLER, K. S. (1994). Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States: Results from the National Comorbidity Study. *Archives of General Psychiatry*, 51, 8–19.
- KLEIN, M. H., BENJAMIN, L. S., TREECE, C., ROSENFELD, R., & GREIST, J. (1993). The Wisconsin Personality Disorder Inventory: Development, reliability and validity. *Journal of Personality Disorders*, 7, 285–303.
- LAMBERT, M. J. (ED.). (2004). *Bergin and Garfield's handbook of psychotherapy and behavior change*. New York: Wiley.
- LANSFORD, E. (1986). Weakenings and repairs of the working alliance in short-term psychotherapy. *Professional Psychology: Research and Practice*, 17, 364–366.
- LIVESLEY, W. J. (ED.). (2001). *Handbook of personality disorders: Theory, research, and treatment*. New York: Guilford Press.
- LUBORSKY, L. (1984). *Principles of psychoanalytic psychotherapy: A manual for supportive-expressive treatment*. New York: Basic Books.
- LUBORSKY, L., SINGER, B., & LUBORSKY, L. (1975). Comparative studies of psychotherapies: Is it true that “everyone has won and all must have prizes”? *Archives of General Psychiatry*, 32, 995–1008.
- MARTIN, D. J., GARSKE, J. P., & DAVIS, M. K. (2000). Relation of the therapeutic alliance with outcome and other variables: A meta-analytic review. *Journal of Counseling and Clinical Psychology*, 68, 438–450.
- MATTIA, J. I., & ZIMMERMAN, M. (2001). Epidemiology. In W. J. Livesley (Ed.), *Handbook of personality disorders* (pp. 107–123). New York: Guilford Press.
- MAVISSAKALIAN, M., & HAMANN, M. S. (1987). DSM-III personality disorder in agoraphobia: II. Changes with treatment. *Comprehensive Psychiatry*, 28, 356–361.
- MILLON, T., & DAVIS, R. O. (1996). *Disorders of personality: DSM-IV and beyond*. (2nd ed.). New York: Wiley.
- MURAN, J. C. (2002). A relational approach to understanding change: Plurality and contextualism in a psychotherapy research program. *Psychotherapy Research*, 12, 113–138.
- MURAN, J. C., & SAFRAN, J. D. (2002). A relational approach to psychotherapy: Resolving ruptures in the therapeutic alliance. In F. W. Kaslow (Ed.), *Comprehensive handbook of psychotherapy* (pp. 253–281). New York: Wiley and Sons.
- MURAN, J. C., SEGAL, Z. V., SAMSTAG, L. W., & CRAWFORD, C. (1994). Patient pretreatment interpersonal problems and therapeutic alliance in short-term cognitive therapy. *Journal of Consulting and Clinical Psychology*, 62, 185–190.
- NAJAVITS, L. M., & STRUPP, H. H. (1994). Differences in the effectiveness of psychodynamic therapists: A process outcome study. *Psychotherapy*, 31, 114–123.
- NEWMAN, C. F. (1998). The therapeutic relationship and alliance in cognitive therapy. In J. D. Safran & J. C. Muran (Eds.), *The therapeutic alliance in brief psychotherapy* (pp. 95–122). Washington, DC: APA Books.
- PATTON, M. J., MURAN, J. C., SAFRAN, J. D., WACHTEL, P., & WINSTON, A. (1998). *Treatment fidelity measure for three brief psychotherapies*. New York: Beth Israel Center.
- PERRY, J. C., LINGIARDI, V., & IANNI, F. (1999). Psychodynamic research can help us to improve diagnosis and therapy for personality disorders: The case of defense mechanisms. In J. Derksen & C. Maffei (Eds.), *Treatment of personality disorders* (pp. 39–52). New York: Kluwer Academic Publishers.
- PERSONS, J. B. (1989). *Cognitive therapy in practice: A case formulation approach*. New York: Norton.
- PERSONS, J. B., BURNS, D. D., & PERLOFF, J. M. (1988). Predictors of dropout and outcome in cognitive therapy for depression in a private practice setting. *Cognitive Therapy and Research*, 9, 539–557.
- PIPER, W. E., & JOYCE, A. S. (2001). Psychosocial treatment outcome. In W. J. Livesley (Ed.), *Handbook of personality disorders: Theory, research, and treatment* (pp. 323–343). New York: Guilford Press.
- PIPER, W. E., JOYCE, A. S., MCCALLUM, M., & AZIM, H. F. (1998). Interpretive and supportive forms of psychotherapy and patient personality variables. *Journal of Consulting and Clinical Psychology*, 66, 558–567.
- POLLACK, J., FLEGENHEIMER, W., KAUFMAN, J., & SADOW, J. (1992). *Brief adaptive psychotherapy for per-*

- sonality disorders: A treatment manual. San Diego, CA: Social & Behavioral Documents.
- RAPAPORT, M. H. (2001). Prevalence, recognition, and treatment of comorbid depression and anxiety. *Journal of Clinical Psychiatry*, 62(Suppl. 24), 6–10.
- REGIER, D. A., RAE, D. S., NARROW, W. E., KAELEBER, C. T., & SCHATZBERG, A. F. (1998). Prevalence of anxiety disorders and their comorbidity with mood and addictive disorders. *British Journal of Psychiatry*, 173(Suppl. 34), 24–28.
- RHODES, R., HILL, C., THOMPSON, B., & ELLIOTT, R. (1994). Client retrospective recall of resolved and unresolved misunderstanding events. *Counseling Psychology*, 41, 473–483.
- SAFRAN, J. D. (2002). Brief relational psychoanalytic treatment. *Psychoanalytic Dialogues*, 12, 171–195.
- SAFRAN, J. D., & MURAN, J. C. (1996). The resolution of ruptures in the therapeutic alliance. *Journal of Counseling and Clinical Psychology*, 64, 447–458.
- SAFRAN, J. D., & MURAN, J. C. (2000). *Negotiating the therapeutic alliance: A relational treatment guide*. New York: Guilford Press.
- SAFRAN, J. D., MURAN, J. C., SAMSTAG, L. W., & STEVENS, C. (2002). Repairing alliance ruptures. In J. C. Norcross (Ed.), *Psychotherapy relationships that work* (pp. 235–254). New York: Oxford University Press.
- SAFRAN, J. D., MURAN, J. C., SAMSTAG, L. W., & WINSTON, A. (2005). Evaluating an alliance-focused intervention for potential treatment failures. *Psychotherapy*, 42, 512–531.
- SAMSTAG, L. W. (1999). Difficult dyads and unsuccessful treatments: A comparison of dropout, poor, and good outcome groups in brief psychotherapy. *Dissertation Abstracts International*, 59, 5109.
- SAMSTAG, L. W., BATCHELDER, S., MURAN, J. C., SAFRAN, J. D., & WINSTON, A. (1998). Predicting treatment failure from in-session interpersonal variables. *Journal of Psychotherapy Practice and Research*, 5, 126–143.
- SANTANGELO, C., SAFRAN, J. D., MURAN, J. C., & WINSTON, A. (1994,). *Adherence to manualized brief psychotherapy treatments*. Paper presented at the Annual Meeting of the Society for Psychotherapy Research, Pittsburgh, PA.
- SHAPIRO, D. A., FIRTH-COZENS, J. A., & STILES, W. B. (1989). The question of therapists' differential effectiveness: A Sheffield Psychotherapy Project addendum. *British Journal of Clinical Psychology*, 154, 383–385.
- SHEA, M. T. (1993). Psychosocial treatment of personality disorders. *Journal of Personality Disorders*, (Suppl. 1), (Spr), 167–180.
- SHEA, M. T. (1997). Core battery conference: Assessment of change in personality disorders. In H. H. Strupp & L. M. Horowitz & M. J. Lambert (Eds.), *Measuring patient changes in mood, anxiety, & personality disorders* (pp. 89–400). Washington, DC: APA Books.
- SHEA, M. T., PILKONIS, P. A., BECKHAM, E., & COLLINS, J. F. (1990). Personality disorders and treatment outcome in the NIMH Treatment of Depression Collaborative Research Program. *American Journal of Psychiatry*, 147, 711–718.
- STRUPP, H. H., & BINDER, J. L. (1984). *Psychotherapy in a new key: A guide to time-limited dynamic psychotherapy*. New York: Basic Books.
- TRYON, G. S., & KANE, A. S. (1990). The helping alliance and premature termination. *Counseling Psychology Quarterly*, 3, 233–238.
- TRYON, G. S., & KANE, A. S. (1993). Relationship of working alliance to mutual and unilateral termination. *Journal of Counseling Psychology*, 40, 33–36.
- TRYON, G. S., & KANE, A. S. (1995). Client involvement, working alliance, and type of therapy termination. *Psychotherapy Research*, 5, 189–198.
- TURNER, A. E., & MURAN, J. C. (1992). *Cognitive-behavioral therapy for personality disorders: A treatment manual*. San Diego, CA: Social & Behavioral Documents.
- TURNER, R. (1987). The effects of personality disorder diagnosis on the outcome of social anxiety symptom reduction. *Journal of Personality Disorders*, 1, 136–143.
- WEINSTOCK-SAVOY, D. E. (1986). The relationship of therapist and patient interpersonal styles to outcome in brief dynamic psychotherapy. *Dissertation Abstracts International*, 47, 2638.
- WESTEN, D., & MORRISON, K. (2001). A multidimensional meta-analysis of treatments for depression, panic, and generalized anxiety disorder: An empirical examination of the status of empirically supported therapies. *Journal of Counseling and Clinical Psychology*, 69, 875–899.
- WIERZBICKI, M., & PEKARIK, G. (1993). A meta-analysis of psychotherapy dropout. *Professional Psychology: Research and Practice*, 24, 190–195.
- WINSTON, A., POLLACK, J., LAIKIN, M., SAMSTAG, L. W., MCCULLOUGH, L., & MURAN, J. C. (1994). Short-term psychotherapy of personality disorders. *American Journal of Psychiatry*, 151, 191–194.
- ZIMMERMAN, M., MCDERMUT, W., & MATTIA, J. I. (2000). Frequency of anxiety disorders in psychiatric outpatients with major depressive disorder. *American Journal of Psychiatry*, 157, 1337–1340.