

# Relation of the Therapeutic Alliance With Outcome and Other Variables: A Meta-Analytic Review

Daniel J. Martin, John P. Garske, and M. Katherine Davis  
Ohio University

To identify underlying patterns in the alliance literature, an empirical review of the many existing studies that relate alliance to outcome was conducted. After an exhaustive literature review, the data from 79 studies (58 published, 21 unpublished) were aggregated using meta-analytic procedures. The results of the meta-analysis indicate that the overall relation of therapeutic alliance with outcome is moderate, but consistent, regardless of many of the variables that have been posited to influence this relationship. For patient, therapist, and observer ratings, the various alliance scales have adequate reliability. Across most alliance scales, there seems to be no difference in the ability of raters to predict outcome. Moreover, the relation of alliance and outcome does not appear to be influenced by other moderator variables, such as the type of outcome measure used in the study, the type of outcome rater, the time of alliance assessment, the type of alliance rater, the type of treatment provided, or the publication status of the study.

In the past two decades, psychotherapy researchers and practitioners have postulated that the therapeutic alliance—defined broadly as the collaborative and affective bond between therapist and patient—is an essential element of the therapeutic process. Although the alliance concept originated in early psychoanalytic theories (e.g., Freud, 1912/1958, 1913/1958; Greenson, 1967; Zetzel, 1956), it has become increasingly common in recent conceptualizations of the therapeutic process generally. The primary reason the alliance has grown in significance is the consistent finding that the quality of the alliance is related to subsequent therapeutic outcome. Indeed, although researchers from several theoretical orientations have assessed the alliance in a variety of ways using an assortment of measures, they have consistently found that the alliance is related to therapeutic outcome (Horvath & Symonds, 1991).

Another reason interest in the alliance has increased in the past 20 years is the inability of researchers to find a consistent difference in the effectiveness of psychotherapy across orientations. As a result of the conclusion that psychotherapies are generally found to be effective (e.g., Lambert & Bergin, 1994; Smith, Glass, & Miller, 1980; Stiles, Shapiro, & Elliot, 1986), many researchers have looked for common factors across therapies that can explain therapeutic outcomes. Of these researchers, several have begun to conceptualize the alliance as a common factor

across therapeutic disciplines; some have even begun to argue that the quality of the alliance is more important than the type of treatment in predicting positive therapeutic outcomes (e.g., Safran & Muran, 1995). Consequently, many contemporary theories of psychotherapeutic change now emphasize the importance of the alliance, so much so that some theorists have referred to the alliance as the “quintessential integrative variable” (Wolfe & Goldfried, 1988, p. 449) of therapy.

It seems worthwhile to empirically review the many existing alliance studies to identify underlying patterns in the literature. Although a meta-analysis of the alliance literature has already been conducted (Horvath & Symonds, 1991), the number of new studies that have investigated the relation of the alliance and outcome justified an updated meta-analytic review of the data.

Although the therapeutic alliance is now commonly referred to as a single construct, it has actually developed from various understandings of the relationship between therapist and patient. Indeed, whereas some have conceptualized the alliance as representing only one construct (e.g., Sterba, 1934; Zetzel, 1956), others have hypothesized that the alliance consists of several independent dimensions (e.g., Bordin, 1979; Luborsky, 1976). In addition, theorists and practitioners have used various terms to describe different aspects of the relationship between a therapist and a patient, such as *therapeutic alliance*, *working alliance*, *therapeutic bond*, and *helping alliance*. Unless otherwise specified, we have referred to the general construct under discussion as the *alliance*. Although there are differences among the many alliance conceptualizations, most theoretical definitions of the alliance have three themes in common: (a) the collaborative nature of the relationship, (b) the affective bond between patient and therapist, and (c) the patient’s and therapist’s ability to agree on treatment goals and tasks (Bordin, 1979; Gaston, 1990; Horvath & Symonds, 1991; Saunders, Howard, & Orlinsky, 1989).

## Alliance Measures

Researchers have attempted to create scales that measure the alliance based on these various theoretical understandings of the

---

Daniel J. Martin, John P. Garske, and M. Katherine Davis, Department of Psychology, Ohio University.

Daniel J. Martin is now at the Substance Abuse Treatment Unit, Department of Psychiatry, Yale University School of Medicine.

M. Katherine Davis is now at the Division of Medical Psychology, Duke University Medical Center.

This article is based on Daniel J. Martin’s doctoral dissertation. John P. Garske chaired the dissertation. Scott Baker, M. E. Crowley, Carla Kmett, Nichole Marioni, and Jennifer Zerbst Martin assisted in the coding process.

Correspondence concerning this article should be addressed to Daniel J. Martin, Substance Abuse Treatment Unit, Department of Psychiatry, Yale University School of Medicine, 1 Long Wharf Drive, Box 18, New Haven, Connecticut 06511. Electronic mail may be sent to Daniel.Martin@Yale.Edu.

concept. Although some scales were created for specific research projects, others were intended to be used more generally. Researchers created the early alliance scales so that they could measure the alliance in individual, adult psychotherapy; these scales were rated by trained clinical judges or by independent observers. To avoid the limitations of relying on alliance measures assessed from only an observer's perspective, researchers adapted most of the early scales so that patients and therapists could also rate the alliance. Although the alliance scales were developed by several independent research groups, the scales have been shown to be highly correlated (e.g., Hatcher & Barends, 1996; Safran & Wallner, 1991; Tichenor & Hill, 1989). Because the various alliance scales are based on somewhat different theoretical understandings of the alliance and rely on different methodologies for measuring the relationship, the origins and development of the most common alliance scales are presented below.

### *The Pennsylvania Scales*

Of the early alliance scales, one of the most commonly used sets of scales was the measures that Luborsky and his colleagues developed at the University of Pennsylvania (referred to collectively as the Penn scales). The Penn scales were created to empirically test Luborsky's (1984) psychodynamic conceptualization of the helping alliance, measuring both Type 1 signs (the patient's experience of the therapist as providing the help that is needed) and Type 2 signs (the patient's experience of treatment as a process of working together with the therapist toward the goals of treatment). Over the years, Luborsky and his colleagues have constructed several scales to assess these two dimensions of the alliance.

At first, the Pennsylvania researchers relied on the Helping Alliance Counting Signs method (HACs; Luborsky, 1976). Within the HACs, there are four subscales for Type 1 signs of the alliance and three subscales for Type 2 signs. To assess the alliance using the HACs, a clinical observer locates and rates all relevant patient statements that appear in a therapy session transcript. After locating a statement, the observer determines the subscale to which the statement belongs, decides whether the statement is positive or negative, and rates the statement on a 5-point scale (1 = *very low*, 5 = *very high*). The final HACs score is the sum of the positive and negative signs for each subscale. (See Alexander & Luborsky, 1986, for a more complete description of this procedure.)

Because rating the alliance with the HACs was a time-consuming procedure, the Pennsylvania researchers created the Penn Helping Alliance Rating Method (HAR; Luborsky, Crits-Christoph, Alexander, Margolis, & Cohen, 1983; Morgan, Luborsky, Crits-Christoph, Curtis, & Solomon, 1982). The HAR has several advantages over the HACs: The HAR is less time-consuming, is not dependent on clinical transcripts, and is used to rate briefer samples of therapy. They developed this scale by converting each of the subscales from the HACs into 10-point rating scales (1 = *very little or none*, 5 = *moderate amount*, 10 = *very much*). In addition, the researchers added three new subscales to the HAR. The HAR is assessed by clinical observers; the patient's score is equal to the sum of the subscale ratings. In a study that measured both the HACs and the HAR (Luborsky et al., 1983), the two scales were found to be highly correlated, with correlations for the positive subscales ranging from .57 to .86. Because the two

scales seemed to be measuring similar constructs, the more user-friendly HAR has become the preferred scale for researchers attempting to assess Luborsky's conceptualization of the alliance from an observer's perspective.

These researchers later developed versions of the HACs and the HAR that were to be rated by therapists. These scales are virtually identical to the observer-rated versions, except that they are reworded so that they are from the therapist's point of view. These therapist-rated versions of the Penn scales correlated with the observer-rated versions, with a correlation range of .76 to .85 (Alexander & Luborsky, 1986).

Luborsky and his colleagues next constructed the Helping Alliance Questionnaire Method (HAQ; Luborsky, McLellan, Woody, O'Brien, & Auerbach, 1985). This alliance scale allows the psychotherapy patient to rate the alliance on 11 items that are similar to those included in both the HACs and the HAR; 8 items refer to Type 1 signs of the alliance, whereas 3 items refer to Type 2 signs. The patient rates each item on a 6-point scale (-3 = *No, I strongly feel that it is not true*, 3 = *Yes, I strongly feel that it is true*); the alliance score is the sum of the subscale ratings. Luborsky et al. stated that the HAQ was highly correlated with the therapist-rated version of the HAR, but they failed to report the exact correlation. In a later study (Gerstley et al., 1989), a therapist-rated parallel form of the HAQ was created. Thus, researchers could use the Penn scales to assess the alliance from patients', therapists', and observers' perspectives.

### *The Vanderbilt Scales*

Strupp and his colleagues, at Vanderbilt University, developed scales that reflect Strupp's combination of dynamic and integrative conceptualizations of the alliance (e.g., Strupp & Binder, 1984), as well as the theories of Bordin (1979), Greenson (1965), and Luborsky (1976). The Vanderbilt researchers first developed and refined the Vanderbilt Psychotherapy Process Scale (VPPS; Gomes-Schwartz, 1978; O'Malley, Suh, & Strupp, 1983; Suh, Strupp, & O'Malley, 1986), an 80-item descriptive measure of the therapist-patient relationship and the psychotherapy process. The scale was "designed to assess both positive and negative aspects of the patient's and therapist's behavior and attitudes that are expected to facilitate or impede progress in therapy" (Suh et al., 1986, p. 287). The VPPS was created so that clinical observers could rate a given segment of therapy on a 5-point scale (1 = *not at all*, 5 = *a great deal*) for each of the 80 items. The scale's first 3 items assess global impressions of the quality of the patient-therapist relationship, the productivity of the session, and the patient's current level of functioning. Of the remaining items, 40 pertain to the patient and 37 pertain to the therapist; these items rate the patient and therapist on both their behavior and their demeanor during the therapy segment (Suh et al., 1986).

Although the VPPS measured features of the therapist-patient relationship, it was not specifically designed to measure the alliance; thus, Strupp and his colleagues developed the Vanderbilt Therapeutic Alliance Scale (VTAS; Hartley & Strupp, 1983). They designed this scale so that observers could rate the alliance for a given segment of a therapy session; each item is rated on a 6-point scale (0 = *none at all*, 5 = *a great deal*). Of the 44 items on the VTAS, 14 pertain to the patient, 18 pertain to the therapist, and 12 pertain to the patient-therapist interaction. A factor analysis con-

ducted on the two scales found that the VPPS and the VTAS had similar factor structures (Hartley & Strupp, 1983); therefore, the VTAS has become their preferred global alliance scale.

More recently, the Vanderbilt researchers (e.g., Henry, Schacht, & Strupp, 1986) have advocated the use of Benjamin's Structural Analysis of Social Behavior System (SASB; Benjamin, 1974) as an alliance measure. This scale measures small units of interpersonal transactions, usually only a sentence or less; thus, there are many ways in which different versions of the SASB can be used to assess the therapist-patient relationship. Although Strupp and his colleagues agree that global ratings of the alliance are useful indices, they have become interested in obtaining more refined measures of the interpersonal process that exists within the therapeutic setting (Henry & Strupp, 1994). Their use of the SASB is indicative of this interest.

### *The Toronto Scales*

Created by Marziali and her colleagues at the University of Toronto, along with Marmar and his colleagues at the Langley Porter Psychiatric Institute in San Francisco, these scales measure classic psychodynamic conceptualizations of the alliance as well as Bordin's (1979) integrative model. After combining items they generated with items selected from other scales (e.g., the VPPS [Gomes-Schwartz, 1978], the VTAS [Hartley & Strupp, 1983], and the HAcS [Luborsky, 1976]), they created the Therapeutic Alliance Rating Scale (TARS; Marziali, Marmar, & Krupnick, 1981). While developing this scale, Marziali et al. focused on affective aspects of the alliance rather than on therapist interventions or specific responses. The TARS allowed nonparticipant observers to evaluate the positive and negative aspects of both therapists and patients. After revisions, the scale contained 42 items, with 21 pertaining to the patient and 21 pertaining to the therapist. Each item is rated on a 6-point scale (0 = *not present*, 5 = *intensely present*). In their preliminary study, Marziali et al. found that the TARS items that assessed patient contributions to the alliance predicted subsequent therapy outcome, but the items that assessed therapist contributions did not discriminate between "good" and "bad" outcomes.

Later, Marziali (1984) created therapist- and patient-rated versions of the TARS by rewording the original observer-rated version of the scale. The 42 items are again rated on a 6-point scale (0 = *not present*, 5 = *intensely present*). Marziali found that the three versions of the TARS were correlated on the patients' positive and negative contributions to the therapy and the therapists' positive contributions but not on the therapists' negative contributions. Nevertheless, all three versions of the TARS were related to therapeutic outcome, with the patient- and therapist-rated versions being better predictors than the observer-rated version.

### *Working Alliance Inventory (WAI)*

In an attempt to measure Bordin's (1979) integrative model of the alliance, Horvath and his colleagues developed the WAI (Horvath, 1981; Horvath & Greenberg, 1986, 1989). Their goal in creating a new alliance scale was threefold: They wanted to measure alliance factors in all types of therapy, to document the relation between the alliance measure and the theoretical constructs underlying the measure, and to connect the alliance mea-

sure to a general theory of therapeutic change (Horvath, 1994). Using techniques that focused on content validity, they developed the WAI so that it would measure Bordin's three aspects of the alliance: the bond, the agreement on goals, and the agreement on tasks. To allow measurement of the alliance from different perspectives, Horvath and his colleagues developed patient-, therapist-, and independent observer-rated versions of the WAI; each item is rated on a 7-point scale (1 = *never*, 7 = *always*). Recently, others (e.g., Tracey & Kokotovic, 1989) have developed a shortened version of these scales. Research has shown strong support for the reliability of the WAI scales and some support for its validity (see Horvath, 1994, for a review). For example, Horvath and Greenberg (1989) found that the WAI's reliability ranged from  $r = .85$  to  $r = .93$  and that the scale correlated with a variety of outcome indices.

### *The California Scales*

The California alliance scales were based on Marziali and Marmar's TARS (Marziali, 1984; Marziali et al., 1981). Although the earlier scales were commonly referred to as the TARS, the California researchers sometimes refer to them as the California TARS (CALTARS). Marmar, Gaston, Gallagher, and Thompson, (1989) revised the TARS (or CALTARS) and developed the California Psychotherapy Alliance Scales (CALPAS). The CALPAS was also based on theoretical writings (e.g., Bordin, 1979; Freud 1913/1958; Greenson, 1967; Luborsky, 1976) and on the empirical results of other alliance measures (e.g., Gomes-Schwartz, 1978; Hartley & Strupp, 1983; Marmar, Gaston, et al., 1989). The current CALPAS consists of 24 items that were created to assess Gaston's (1990) four aspects of the alliance: (a) The therapeutic alliance is measured by the Patient Working Capacity scale, (b) the working alliance is assessed by the Patient Commitment scale, (c) the therapist's contribution to the alliance is measured by the Therapist Understanding and Involvement scale, and (d) the patient-therapist agreement on treatment goals and tasks is assessed by the Working Strategy Consensus scale. Patient-, therapist-, and independent observer-rated versions of the CALPAS exist; each item is rated on a 7-point scale (1 = *not at all*, 7 = *very much so*). Factor analytic studies have shown some confirmation for these four factors (e.g., Gaston, 1991; Marmar, Gaston, et al., 1989; Marmar, Weiss, & Gaston, 1989). Because the CALPAS offers versions that are rated by patients, therapists, and independent observers, and because of the integrative nature of the instrument, the CALPAS has been used extensively by researchers who have wanted to assess the alliance in psychotherapy outcome studies.

### *Therapeutic Bond Scales (TBS)*

Saunders and his colleagues developed the TBS (Saunders et al., 1989) based on the generic model of psychotherapy (Orlinsky & Howard, 1986). These scales measure the three dimensions of Orlinsky and Howard's therapeutic bond: working alliance, empathic resonance, and mutual affirmation. In addition, the TBS provides a Global Bond Scale, which is a composite of the three subscales. Saunders et al. created the TBS by extracting items from the Therapy Session Report (Orlinsky & Howard, 1966), a 145-item general survey of patients' experiences during therapy; they

chose items from this scale that they thought represented an aspect of the therapeutic bond. After revisions due to psychometric considerations, the 50-item TBS consists of the following dimensions: 15 items compose the Working Alliance scale, 17 items compose the Empathic Resonance scale, and 18 items compose the Mutual Affirmation scale. When using the TBS, patients' rate their experiences during therapy on a 21-point scale (0 = *no experience*, 10 = *some experience*, 20 = *a lot of experience*).

Research has provided some support for the reliability and validity of the TBS (e.g., Saunders et al., 1989). For example, the internal reliabilities of each subscale were adequate ( $r = .72$  to  $.87$ ), as was the internal reliability of the Global Bond scale ( $r = .62$ ). In addition, two of the scales failed to correlate (Working Alliance and Empathic Resonance), indicating that the scales were measuring different aspects of the alliance. All three scales and the Global Bond scale were related to patient ratings of session quality ( $r = .34$  to  $.60$ ), whereas only the Global Bond scale was related to observer ratings of termination outcome ( $r = .19$ ).

In summary of the alliance measures reviewed, a few conclusions are apparent. First, some scales attempt to measure specific theoretical conceptualizations of the alliance (the Penn scales, the WAI, the CALPAS, and the TBS), whereas other scales attempt to measure more eclectic blends of alliance constructs (the VPPS, the VTAS, and the TARS). Second, the scales use different rating systems that vary considerably to assess the alliance. Indeed, the counting signs method of the HACs is much different than the 5-point rating system of the VPPS and the 6-point rating systems of the VTAS and the TARS, which again are much different than the 7-point system of the WAI, the 10-point system of the HAR, or the 21-point system of the TBS. Third, the number of items vary across alliance measures: from the 7 items of the HACs to the 80 items of the VPPS. Fourth, the scales purport to measure different numbers of alliance dimensions; for example, there are two dimensions of the Penn scales, three dimensions of the WAI and the TBS, and four dimensions of the CALPAS. In their review of the alliance scales, Horvath and Luborsky (1993) concluded that there are two core aspects of the alliance measured by most scales: (a) therapist-patient affective attachments and (b) collaboration or willingness to invest in the therapy process. The following meta-analysis attempted to integrate 20 years of research using these various alliance scales.

### The Present Study

In an earlier attempt to empirically review the relation between alliance and outcome, Horvath and Symonds (1991) found an average effect size of .26 between quality of alliance and outcome. Researchers, theorists, and clinicians throughout the psychotherapy literature have referred to this alliance-outcome effect size as evidence that the alliance is related to therapeutic outcome. Since the time of Horvath and Symonds's review, researchers have published many empirical studies that have included alliance and outcome measures; these studies needed to be integrated using meta-analytic techniques. For example, Horvath and Symonds found 24 studies that met their criteria for inclusion. In our review of the literature, we located over 60 additional studies that seemed to meet their inclusion criteria. Because the number of empirical studies we identified is more than three times greater than the number in Horvath and Symonds's review, it seemed necessary to

conduct another meta-analytic review of this area of inquiry. In the present study, we attempted to reanalyze Horvath and Symonds's previous comparisons, such as the relation of the alliance-outcome correlation with type of treatment, type of rater, and time of alliance rating. In addition, we analyzed the relation of the alliance-outcome correlation with type of alliance measure. Furthermore, we assessed whether a publication bias or a file drawer problem existed in the literature. Finally, we assessed the homogeneity of each effect size to ascertain the uniformity of the effect.

## Method

### Selection of Studies

In the present meta-analysis, 79 studies were analyzed, including 58 published studies and 21 unpublished doctoral dissertations or master's theses. The following are the procedures that we followed to arrive at these numbers.

**Inclusion criteria.** Based on the criteria used by Horvath and Symonds (1991), the following guidelines were used as the inclusion criteria for the present meta-analysis: (a) The alliance construct had to be referred to as *therapeutic alliance*, *working alliance*, *helping alliance*, *therapeutic bond*, or simply *alliance*; (b) the study had to include a quantifiable measure of the relationship between the alliance and some assessment of outcome, measured subsequently; (c) the study had to be clinical, not analogue; (d) the study must have included at least 5 patients; (e) the therapeutic treatment must have been individual, as opposed to group, family, or couples treatment; (f) the study needed to be presented in English; and (g) the study had to be available (published in journals or in *Dissertation Abstracts* or presented at a professional conference) between January 1, 1977, and January 1, 1997. Overall, the most common reason for the elimination of a study was the failure to include a quantifiable measure of the relationship between alliance and outcome.

**Literature review.** In an attempt to find all relevant published articles, we conducted a literature review using the following databases: PsycINFO, PsycLIT, ERIC, and Medline. When we entered the search words *alliance* and *therapeutic bond*, 1,405 studies were found. We carefully read the abstracts of all 1,405 studies and identified 111 studies that seemed to match the inclusion criteria. Next, we manually searched the abstracts from the following journals for the previous 2 years (1995 and 1996 inclusive): *American Journal of Psychiatry*, *Archives of General Psychiatry*, *Counseling Psychologist*, *Journal of Consulting and Clinical Psychology*, *Journal of Counseling Psychology*, *Journal of Nervous and Mental Disease*, *Psychotherapy*, *Psychotherapy Research*, and *Professional Psychology: Research and Practice*. An additional 24 studies were identified from this procedure that had not been identified by the database reviews. As a last precaution, we examined the references of the selected studies to identify any overlooked studies; 5 studies were identified by this method that had failed to be identified by the other two search procedures.

As a result of all three search procedures, 140 published studies were identified as possibly meeting the inclusion criteria. We then read each study, which resulted in the removal of 69 studies from further consideration; these studies failed to meet at least one of the inclusion criteria. At this time, the rating procedure began with 71 published studies still under consideration. Because of data that were unanalyzable (e.g., no comparison of alliance with outcome or no statistics were reported), another 13 studies were eliminated, which resulted in 58 published studies remaining in the meta-analysis.

Similar, but not as extensive, procedures were used to identify unpublished presentations, doctoral dissertations, or master's theses. Primarily, we relied on searching the *Dissertation Abstracts* database, but we also examined the references of the selected studies to identify any overlooked unpublished studies. When the words *alliance* and *therapeutic bond* were searched using *Dissertation Abstracts*, 3,040 studies were found. We then

limited our search by including the word *outcome*. From this search, 271 doctoral dissertations or master's theses were identified. We then read the abstracts of all 271 studies and found 37 that seemed to meet the inclusion criteria. Using various methods, we attempted to obtain these 37 unpublished studies; we ultimately obtained 35 dissertations or master's theses. After reviewing these 35 studies, 11 were eliminated from further consideration because they failed to meet the inclusion criteria. Therefore, 24 unpublished studies were included when the rating procedure began. One study was later eliminated from further analysis because it had been published, and 2 studies were eliminated because they reported only the results of factor analysis or multiple regression analyses. Thus, 21 unpublished studies were included in the meta-analysis at the end of the selection process.

### Coding Procedure

After selecting the studies that were included in the present meta-analysis, five graduate students and two advanced undergraduates coded and recorded several variables from each study. Coding materials were created to translate the features of each alliance study into usable quantitative data. The following variables were coded: publication year, type of publication (e.g., journal article or dissertation), methodological quality (e.g., retrospective or prospective, whether random assignment was present, and whether an experimental manipulation was present), research institute that conducted the study (e.g., University of Pennsylvania or Vanderbilt University), number of patients per study, gender of patients, presenting problem (e.g., depression or personality disorder), treatment setting (e.g., outpatient, partial hospital, or inpatient), number of total therapists, gender of therapists, therapist years of experience, therapist professional affiliation (e.g., psychologist, psychiatrist, or social worker), type of treatment (e.g., behavioral, cognitive, or psychodynamic), type of alliance scale (e.g., the WAI or the CALPAS), type of alliance rater (e.g., patient, therapist, or observer), time of alliance rating (e.g., early, middle, late, averaged), reliability information of alliance scales, type of outcome measure (e.g., the Global Assessment Scale [GAS], the Target Complaints Scale, or the Beck Depression Inventory [BDI]), type of outcome rater (e.g., patient, therapist, or observer), and overall correlation of alliance and outcome for each alliance measure paired with each outcome measure. The research assistants were trained to use the coding materials by reading selected articles on the alliance and by completing four practice codings of studies that were not included in this meta-analysis. After completing the practice codings, the research assistants met with the principal investigator to discuss discrepancies. When the research assistants reported that they were comfortable and confident in their coding abilities, the coding of the studies included in the meta-analysis began.

To ensure thoroughness, each study was assessed by two coders: one of the six research assistants and the principal investigator. If any discrepant ratings were found between the original two coders, a third coder rated the study. The coders then discussed the discrepancy until a unanimous verdict was reached. (On several occasions, it was necessary to reread the primary study to solve the discrepancy.) Discrepancies usually resulted from research assistants being uncertain whether a specific variable was an outcome measure. In addition, discrepancies were sometimes noted when a primary study reported several statistical indices to explain a single comparison; coders were uncertain which index to code. Discrepancies also occurred because of coder transcriptional errors.

As a check on the accuracy of the coding procedures, the variables coded for methodological quality were compared before consensus was reached. There was 87% agreement on whether the study was retrospective or prospective, 87% agreement on whether random assignment to conditions occurred, and 92% agreement on whether an experimental manipulation was included. The kappas for the agreement between the two raters were .76, .68, and .85, respectively. These findings indicate that there was a high level of agreement between raters before consensus was obtained.

### Estimation of Effect Size

Because analyses in this area of research are usually reported as correlations between alliance and outcome, we used a variation of the product-moment correlation as the effect size estimate (Hedges & Olkin, 1985; Hunter & Schmidt, 1990; Rosenthal, 1991). The best weighted estimate of the population correlation is

$$\bar{r} = \frac{\sum [N_i r_i]}{\sum N_i},$$

where  $r_i$  is the correlation in study  $i$  and  $N_i$  is the number of individuals in study  $i$ . The corresponding weighted variance is

$$s_r^2 = \frac{\sum [N_i (r_i - \bar{r})^2]}{\sum N_i}.$$

If the effect size from a study was not reported as a correlation, we converted it into a correlation using procedures recommended by Hedges and Olkin, Hunter and Schmidt, or Rosenthal. Because Hunter and Schmidt argued that converting correlations into Z scores never results in a more accurate meta-analysis, we did not rely on this correction. As Horvath and Symonds (1991) had done, we considered results reported as nonsignificant without the actual value of the statistic being equal to zero. However, as an estimate of the upper boundary of the possible distribution of the overall alliance–outcome correlation, we also recalculated the overall correlation with the results that were set equal to zero removed. Presenting both a conservative and a nonconservative estimate of the overall alliance–outcome correlation allowed for a confidence interval in which the “true” correlation was highly likely to be found.

### Combination of Effect Sizes

Before this meta-analysis could be completed, a few issues needed to be resolved a priori on how to deal with certain situations. The following is the list of guidelines we used when conducting several controversial steps in the meta-analytic process.

1. In this meta-analysis, we computed weighted effect sizes rather than unweighted effect sizes; we weighted the effect sizes by the sample size of each study. In their review of meta-analytic techniques, Hunter and Schmidt (1990) could not identify a situation in which unweighted effect sizes would provide more information than weighted effect sizes. Thus, we did not compute separate unweighted effect sizes.
2. Data reported in separate studies but based on the same patient sample were averaged into an overall effect size that was entered into the analysis once. If the studies used the exact same sample, the overall effect size was simply an average of the effect sizes across studies; if the studies had different sample sizes (e.g., one study included only a subsample of the data), the overall effect size for that set of studies was weighted by the sample size of each study.
3. When data from a study were reported in more than one source (e.g., doctoral dissertation, conference presentation, journal article, or book chapter), we used the most accessible source (usually the journal article). If possible, we located all sources of the study and combined the results as described above. However, if this was not possible, we only reported data from the source we had obtained.
4. When studies reported more than one outcome measure, we examined the effect sizes from that study separately and as an aggregate effect size for that study. The outcome measures for each study were averaged together into an overall effect size per study and were entered into an overall alliance–outcome analysis. In addition, the overall alliance–outcome relation was disaggregated by type of outcome and was reanalyzed separately.
5. Throughout the meta-analysis, we corrected the variance for sam-

pling error. The following section presents the logic behind this methodology, as well as the statistical formulas used to correct for sampling error.

### Correcting the Variance for Sampling Error

Because the observed variance ( $s_r^2$ ) confounds two things—variation in population correlations and variation in sample correlations produced by sampling error—an estimate of the variance in population correlations is obtained by correcting the observed variance for sampling error (Hunter & Schmidt, 1990). For correlational meta-analysis, the best estimate of sampling error is

$$\sigma_e^2 = \frac{(1 - \bar{r}_i^2)^2}{N - 1},$$

where  $\bar{N}$  is the average number of patients per study. The replication of sampling error across studies allows the use of averaging to reduce the impact of that sampling error. The larger the number of studies in the meta-analysis, the more the impact of sampling error can be eliminated. Taking this information into account, the population variance ( $\sigma_p^2$ ) is best estimated by subtracting the sampling error variance ( $\sigma_e^2$ ) from the observed weighted variance ( $s_r^2$ ). That is,

$$\sigma_p^2 = s_r^2 - \sigma_e^2.$$

Thus,  $\sigma_p^2$  is the variance of the population correlation corrected for sampling error.

### Tests of Homogeneity

To ensure that effect sizes were measuring a single population of effects, we conducted tests of homogeneity. This involved calculating  $\bar{r}$ ,  $s_r^2$ , and  $\sigma_e^2$  for each effect size that was found. Homogeneity was then examined using Hedges and Olkin's (1985)  $Q$  statistic, which has approximately a chi-square distribution with  $K - 1$  degrees of freedom, where  $K$  equals the number of effect sizes being examined. The formula for the  $Q$  statistics is as follows:

$$Q = Ks_r^2/\sigma_e^2.$$

If the statistic is significant, it is assumed that the variation of the category under review is not due to sampling error: The category is representing more than one population of effects. When a category is found to have a nonsignificant  $Q$ , the researcher can assume that the category represents a single population of effects.

Although tests of homogeneity were conducted throughout this meta-analysis, many cautions exist against strict reliance on the use of significance tests. As Hunter and Schmidt stated (1990), "significant variation may be trivial in magnitude, and even nontrivial variation may still be due to research artifacts" (p. 110). Researchers, therefore, must not be overly confident in conclusions based only on tests for homogeneity. What can be reasonably concluded from the  $Q$  statistic is that if the chi-square is not significant, there is probably no true variation across studies.

### Significance Tests

To determine whether two effect sizes were significantly different from one another, the following formula from Hunter and Schmidt (1990) was used:

$$z = \frac{\bar{r}_1 - \bar{r}_2}{\sqrt{s_1^2 + s_2^2}},$$

where  $\bar{r}_1$  is the first average effect size,  $\bar{r}_2$  is the second average effect size,  $s_1^2$  is the variance of the first effect size, and  $s_2^2$  is the variance of the second effect size. When using this formula, if  $z$  was greater than 1.96, the effect

sizes were considered to be significantly different at the .05 level. Although significance tests were conducted during this meta-analysis, caution must be taken when interpreting them; there are many logical and statistical problems with relying on significance tests when conducting a meta-analysis.

## Results

### Sample Characteristics

The 79 studies that met the inclusion criteria were conducted over an 18-year span, with 30 studies available before 1990 and 49 studies available between 1990 and 1996. Of these studies, 58 were from published sources and 21 were unpublished doctoral dissertations or master's theses. The mean sample size was 60.39 patients ( $SD = 64.64$ ), and the average length of treatment was 22.18 sessions ( $SD = 18.76$ ). Approximately two thirds of the patients were female. The mean number of therapists per study was 20.22 ( $SD = 19.99$ ), and the average amount of therapist experience was 8.10 years ( $SD = 5.23$ ).

*Type of patient.* The majority of patients who participated in this sample of studies were from an outpatient population ( $n = 59$ ), but a minority of studies ( $n = 18$ ) included patients who were more severely mentally disordered (psychotic disturbances or severe personality disorders). Although most studies included patients with various diagnoses ( $n = 59$ ), some studies included only patients with a specific diagnosis, such as depression ( $n = 8$ ), substance abuse ( $n = 2$ ), bereavement ( $n = 2$ ), or eating disorders ( $n = 1$ ). Because most of the studies contained a heterogeneous sample of patients, meta-analytic techniques were unable to address whether type of patient diagnosis affects the alliance–outcome relation. In addition, most studies included both male and female patients, but the studies failed to break down the alliance–outcome correlation by gender. Thus, meta-analytic techniques were unable to address whether patient gender affects the relation of alliance and outcome.

*Alliance scales.* The sample of studies used several alliance scales to measure the collaborative and affective bond between therapist and patient. Most studies relied on a single family of alliance scales ( $n = 72$ ), but some studies contained more than one type of alliance measure ( $n = 7$ ). Because some studies used more than one alliance scale, the following categories are not mutually exclusive. The WAI was used most often in the sample of studies ( $n = 22$ ), followed by the CALPAS ( $n = 16$ ), the Penn scales ( $n = 12$ ), the Vanderbilt scales ( $n = 9$ ), the TARS ( $n = 5$ ), and the TBS ( $n = 3$ ). In addition, 10 studies relied on scales that were developed to be used with severely mentally disordered patients, and 8 studies used alliance scales that are not commonly referred to in the literature. These scales were either created by the researchers for the specific study or were combinations of the other more common alliance scales.

The alliance measures were rated by either patients, therapists, or observers. Some studies used more than one type of rater; thus, the following are not mutually exclusive. In the sample of studies, patients were the most common rater of the alliance ( $n = 37$ ), followed by therapists ( $n = 26$ ), and observers ( $n = 25$ ). The large number of patient ratings could possibly be a result of Horvath and Symonds's (1991) conclusion that patients' ratings of the alliance were more correlated with outcome than were therapists' ratings. In addition, the increase in patient and therapist ratings could be

due to the ease with which these ratings are obtained: The training of raters and the reliance on audio- or videotapes are eliminated when using patient and therapist ratings.

**Outcome measures.** Many diverse outcome measures were represented in the present sample of studies. Because over 60 different outcome measures were used, the measures were categorized into five broad types of outcome measures: (a) mood scales, such as the BDI and the Hamilton Rating Scale for Depression; (b) symptom scales, such as the Symptom Checklist—90 and the State-Trait Anxiety Inventory; (c) global scales, such as the GAS and overall assessments of change; (d) specific outcome scales, such as the Target Complaints scale and ratings of drug use; and (e) termination status, which is a measure of whether or not the patient remained in therapy until the patient and therapist agreed it was time to terminate the relationship. Because many of the studies used more than one outcome measure, the following are not mutually exclusive. Global scales were the most common type of outcome measure in the sample of studies ( $n = 38$ ), followed by specific scales ( $n = 27$ ), symptom scales ( $n = 24$ ), termination status ( $n = 13$ ), and mood scales ( $n = 9$ ).

The outcome measures were rated by patients, therapists, or observers. Some studies used more than one type of rater; thus, the following are not mutually exclusive. In the sample of studies, patients were the most common rater of outcome ( $n = 51$ ), followed by observers ( $n = 20$ ) and therapists ( $n = 18$ ). This overreliance on patient-rated outcomes is expected on the basis of past reviews of methodological issues (e.g., Lambert & Hill, 1994).

### Reliability Information

The sample of studies reported a total of 93 reliability indices for the alliance scales. Table 1 presents the aggregated alliance reliability information from these studies. The overall average reliability of the alliance scales based on various estimation methods was  $.79$  ( $n = 93$ ,  $SD = .16$ ). This overall index was then separated by type of estimation method. When Cronbach's alpha was used, the average alliance scale reliability was  $.87$  ( $n = 44$ ,  $SD = .10$ ); when interrater reliability was reported, the average reliability was  $.77$  ( $n = 33$ ,  $SD = .15$ ); and when a test-retest correlation was used, the average correlation was  $.63$  ( $n = 16$ ,

$SD = .19$ ). These numbers are slightly lower than those reported by Horvath and Symonds (1991), but they still indicate that the various alliance measures have acceptable reliability.

To determine the average reliabilities of individual alliance measures, the overall index was broken down by type of alliance scale. Table 1 presents this reliability information. If a type of reliability is not listed for a specific scale, it indicates that no information was available for that scale. As shown in Table 1, all of the alliance scales had overall average reliability indices above  $.70$ . This finding indicates that each of the alliance scales had adequate reliability. However, some of the average indices were based on a small number of reliability correlations. For example, the average reliability index for the TARS was a combination of six indices, and the average reliability index for the TBS was based on only two correlations. Because the overall reliability indices of the TARS and the TBS are based on such small numbers of correlations, their reliability is questionable.

The overall reliability index was also separated by type of rater: patient, therapist, or observer. Table 2 presents this reliability information. When patients rated the alliance, the average reliability index was  $.82$  ( $n = 28$ ,  $SD = .11$ ); when therapists rated it, the average reliability index was  $.72$  ( $n = 16$ ,  $SD = .18$ ); and when observers rated it, the average reliability index was  $.80$  ( $n = 46$ ,  $SD = .18$ ). These results indicate that reliability indices based on patients', therapists', and observers' views of the alliance all have adequate reliability. Because of the large number of reliability indices included in this analysis, these findings probably will not vary substantially in future meta-analytic reviews. Although the ratings of therapists seem to be slightly less reliable than those of patients and observers, therapists' ratings of the alliance are still within the acceptable range. The most noticeable difference among the raters was their test-retest correlations: Patient ratings had adequate reliability, whereas therapist and observer ratings were low. However, this finding should be considered tentative; the difference among the reliabilities was not found to be significant ( $z = 1.70$ ,  $p > .05$ ). In addition, the comparison was based on only 16 reliability indices. Taken together, the results indicate that reliabilities based on patient, therapist, and observer ratings are acceptable.

Table 1  
Reliability of the Alliance Scales

Type	Cronbach's $\alpha$			Interrater			Test-retest			Overall		
	Reliability	$n$	$SD$	Reliability	$n$	$SD$	Reliability	$n$	$SD$	Reliability	$n$	$SD$
Penn scales	.91	5	.04	.68	6	.07	.55	3	.02	.74	14	.15
Vanderbilt scales	.91	4	.05	.83	9	.10				.86	13	.09
WAI	.90	8	.05	.92	2	.01	.73	5	.11	.84	15	.11
CALPAS	.85	14	.12	.66	6	.24	.65	6	.21	.76	26	.20
TARS	.91	3	.03	.73	3	.04				.82	6	.09
TBS	.71	2	.09							.71	2	.09
SMD scales	.76	3	.12	.84	6	.07	.48	2	.24	.75	11	.19
Other scales	.90	5	.03	.92	1	.00				.90	6	.03
Overall	.87	44	.10	.77	33	.15	.63	16	.19	.79	93	.16

Note. Missing data indicate that there was no information available for that scale. WAI = Working Alliance Inventory; CALPAS = California Psychotherapy Alliance Scale; TARS = Therapeutic Alliance Rating Scale; TBS = Therapeutic Bond Scale; SMD = severely mentally disordered.

Table 2  
Reliability of Alliance Raters

Type	Cronbach's $\alpha$			Interrater			Test-retest			Overall		
	Reliability	<i>n</i>	<i>SD</i>	Reliability	<i>n</i>	<i>SD</i>	Reliability	<i>n</i>	<i>SD</i>	Reliability	<i>n</i>	<i>SD</i>
Patient	.84	19	.12	.72	1	.00	.78	8	.09	.82	28	.11
Therapist	.81	7	.10	.77	5	.07	.49	4	.18	.72	16	.18
Observer	.91	16	.05	.78	26	.17	.49	4	.11	.80	46	.18

### Overall Relation of Alliance and Outcome

The overall weighted alliance–outcome correlation was .22 ( $n = 68$ ,  $SD = .12$ ), with an estimated population variance ( $\sigma_p^2$ ) of zero. The test of homogeneity conducted on this overall correlation resulted in a  $Q$  value of 57.89 ( $df = 67$ ,  $p > .05$ ), indicating that the correlation represented a homogeneous population. This finding suggests that there are no moderator variables within the sample of alliance–outcome correlations. According to the logic of homogeneity testing, further analyses were unnecessary; there was not enough variance remaining to allow the breakdown of the overall correlation by the proposed moderator variables. In sum, the overall weighted alliance–outcome correlation represents a single population of effects, indicating that further analyses could not lead to a more explanatory model of the relation of the alliance and outcome.

It could be argued that this overall alliance–outcome correlation of .22 is a conservative estimate because effect sizes that were not reported or that were listed simply as nonsignificant were coded as being equal to zero. As an estimate of the upper boundary of the possible distribution of the overall alliance–outcome correlation, the overall correlation was recalculated with the results that were set equal to zero removed. When only nonzero correlations remained in the meta-analysis, the overall weighted alliance–outcome correlation was .23 ( $n = 67$ ,  $SD = .10$ ), with an estimated population variance ( $\sigma_p^2$ ) of zero. This finding indicates that the overall alliance–outcome correlation of .22 is not an overly conservative estimate and therefore adequately depicts the relation of the alliance and outcome.

Finding that the overall alliance–outcome correlation is homogeneous nullified many of the proposed analyses. To allow other analyses to be conducted, alternative meta-analytic techniques needed to be used in which variance remained within the overall alliance–outcome correlation. By scanning the data, we determined that the averaging of correlations within studies was significantly reducing the amount of variance available in the overall alliance–outcome correlation. Because many of the proposed analyses needed to be conducted on all of the effect sizes, not just the averaged effect size per study, we conducted a meta-analysis of each weighted alliance–outcome correlation without averaging effect sizes within each study. This meta-analysis should not be thought of as a negation of the original weighted alliance–outcome correlation. Rather, it was conducted as an exploratory analysis of the many moderator variables that were coded in preparation for the present research project.

When all alliance–outcome correlations were entered into the meta-analysis, an overall weighted effect size of .23 ( $n = 261$ ,  $SD = .20$ ) was obtained, with an estimated population variance

( $\sigma_p^2$ ) of .02. The test of homogeneity conducted on this overall correlation resulted in a  $Q$  value of 628.49 ( $df = 260$ ), indicating that the correlation represented a heterogeneous population ( $p < .05$ ). Although the alliance–outcome correlations were no longer averaged within studies, the change in the overall alliance–outcome relation was negligible (.22 to .23). This finding allowed a search for moderator variables in an attempt to separate the sample of correlations into homogeneous groupings. Nonetheless, when the overall alliance–outcome correlation was disaggregated by type of outcome measure, type of alliance rater (patient, therapist, or observer), type of outcome rater (patient, therapist, or observer), time of alliance rating (early middle, late, or averaged across sessions), methodological quality, or type of psychotherapy, the model failed to account for additional variance. Only when the overall correlation was disaggregated by type of alliance scale was additional variance successfully explained. When disaggregated by alliance scale, the Penn scales ( $\bar{r} = .29$ ,  $n = 24$ ), the Vanderbilt scales ( $\bar{r} = .25$ ,  $n = 35$ ), the WAI ( $\bar{r} = .24$ ,  $n = 80$ ), and the CALPAS ( $\bar{r} = .17$ ,  $n = 49$ ) were correlated with outcome. The TARS, on the other hand, did not appear to be related to outcome ( $\bar{r} = .07$ ,  $n = 11$ ). The results of the limited amount of research that has been conducted on the TBS seems promising ( $\bar{r} = .37$ ,  $n = 4$ ), but this result was based on only four alliance–outcome correlations. Other miscellaneous alliance scales ( $\bar{r} = .21$ ,  $n = 23$ ) seem to be moderately related to outcome.

### Publication Status

To determine whether publication status influenced the relation of alliance and outcome, we separated the overall alliance–outcome correlation into two groups: published studies and unpublished dissertations or master's theses. This analysis is presented in Table 3. The average correlation for published studies was .25, whereas the average correlation for unpublished studies was .19. Neither of these groupings, however, were homogeneous. Although the correlation for published studies was larger than that for unpublished studies, this difference was not reliable and failed

Table 3  
Influence of Publication Status on the Alliance–Outcome Relation

Publication status	$\bar{r}$	$\sigma_p^2$	$Q$	<i>df</i>	<i>p</i>
Published	.25	.02	410.01 <sup>a</sup>	189	.05
Unpublished	.19	.03	207.39 <sup>a</sup>	70	.05

<sup>a</sup> Indicates rejection of the null hypothesis of homogeneity.



to reach statistical significance ( $z = 1.66, p > .05$ ); thus, a publication bias does not appear to exist in the studies that have been conducted assessing the alliance–outcome relation.

### File Drawer Problem

A file drawer analysis for effect sizes was conducted on the overall average alliance–outcome correlation (Hunter & Schmidt, 1990; Rosenthal, 1991). This analysis estimates the number of unlocated studies with null results that would need to exist to reduce the significance level of the effect size by a certain amount. To complete this analysis, an unweighted overall average correlation was found using the following formula:

$$\bar{r}_K = \sum r_K/K,$$

where  $K$  is the number of studies included in the effect size. An unweighted effect size was necessary because it is impossible to determine the number of patients in the unlocated studies. Different values of  $\bar{r}$  were chosen to demonstrate how many studies with null results would be necessary to lower the overall unweighted average correlation to a given level. The file drawer analysis was completed by determining  $x$ , which represents the number of missing studies with null results that would be necessary to lower the average correlation to the critical value. The formula for this analysis is as follows:

$$x = K(\bar{r}_K/r_c - 1).$$

The unweighted average alliance–outcome correlation was .26. To reduce this effect size to .15, 58 undiscovered studies with null findings would need to exist. To lower the average correlation to .10, 126 null studies would need to be located. To lower the average correlation still further to .05, 331 studies with null results would be needed. Because only 78 studies were located for this meta-analysis and because few of these studies had null results, it is highly unlikely that enough studies exist to reduce the average correlation to .15, and the chances are virtually nonexistent that enough studies exist to reduce it below .10. Therefore, it seems that the file drawer problem is unable to render the overall alliance–outcome correlation insignificant.

### Discussion

Using various techniques, this review indicates that alliance is moderately related to outcome ( $\bar{r} = .22$ ). The average alliance–outcome correlation is within the range of many other effect sizes that are associated with psychotherapy outcome (see Matt & Navarro, 1997, for a review). In addition, the relation of alliance and outcome appears to be consistent, regardless of many of the variables that have been posited to influence this relationship. Indeed, the test of homogeneity suggests that the correlation represents a homogeneous population. In sum, the present meta-analysis indicates that the overall alliance–outcome correlation represents a single population of effects that cannot be reduced by a moderator variable into a more explanatory model of the relation of the alliance and outcome.

This meta-analysis supports the belief that the relation of the therapeutic alliance with outcome is consistent within the psychotherapy literature. With the improved quality of recent investiga-

tions of this relationship (e.g., Gaston, Piper, Debbane, Bienvenu, & Garant, 1994; Krupnick et al., 1996), there is increased confidence that this finding is not a result of confounds in the literature. The direct association between the alliance and outcome identified in this empirical review is supportive of the hypothesis that the alliance may be therapeutic in and of itself (Henry et al., 1994). In other words, if a proper alliance is established between a patient and therapist, the patient will experience the relationship as therapeutic, regardless of other psychological interventions. However, alternative explanations for the relation of the alliance and outcome (e.g., the alliance may have an indirect effect on outcome or the alliance may interact with other interventions) cannot yet be ruled out (Gaston, 1990). What is evident from this review is that the strength of the alliance is predictive of outcome, whatever the mechanism underlying the relation.

From the empirical review of the reliabilities of the various alliance scales, it seems clear that all the alliance measures have adequate reliability. Although the overall reliability index for the various scales was somewhat lower than that found in the previous meta-analysis (.79 vs. .86), the present index still reaches an acceptable standard of consistency. Moreover, when the overall alliance index was separated by individual alliance scales, every alliance measure had an overall reliability index above .70. Surprisingly, even the scales that are not well established as measures of the alliance had adequate reliabilities. Given these results, all the alliance scales seem to have acceptable reliability. These findings are consistent with those reported by other studies (e.g., Hatcher, Barends, Hansell, & Gutfreund, 1995) that found no differences when direct comparisons of the reliabilities of the various alliance scales were conducted.

This meta-analysis did not implicate a specific alliance scale as being more reliable than the others, but it also failed to eliminate a scale from further consideration as a research tool because of its psychometric properties. These results suggest that researchers cannot base their choice of an alliance scale on its reliability indices; the scales all tend to receive strong support.

The overall reliability index was obviously reduced from the previous meta-analysis by the inclusion of test–retest information. Because the alliance is expected to vary somewhat throughout treatment, it was anticipated that the test–retest correlations for all the measures would be lower than the other reliability indices. Indeed, the alliance scales in which there was test–retest information available tended to have lower overall reliabilities than did the scales in which this information was unavailable. It could be argued that alliance measures compared over time should not be referred to as test–retest and that this information is inappropriate as a measure of reliability. That is, because the alliance could conceivably change from week to week, measures of the alliance taken after different sessions are not expected to remain consistent. However, because some of the alliance studies reported these correlations as test–retest information, they were included in the present empirical review as such.

The alliance ratings of patients, therapists, and observers all tended to have adequate reliability. Although the ratings of therapists seemed to be slightly less consistent than those of patients and observers, therapists' ratings of the alliance were still within the acceptable range. Across therapy sessions, patients tended to rate the alliance more consistently than did therapists or observers. This finding was unexpected but helped clarify the contradictory

findings associated with the stability of the alliance (e.g., Frank & Gunderson, 1990; Hartley & Strupp, 1983; Klee, Abeles, & Muller, 1990; O'Malley et al., 1983; Tunis, Delucchi, Schwartz, Banys, & Sees, 1995). On the basis of the present meta-analysis, it seems that patients tend to view the alliance as stable, whereas therapists and observers tend to indicate more change over time in their alliance ratings. The implications of this finding are clear: Because patients tend to view the alliance consistently throughout treatment, they are more likely to view the alliance as positive at termination if their initial assessment was positive. Thus, therapists must be effective at establishing positive alliances with their patients early in the therapy process. However, because of the small sample size of this comparison, the greater consistency of patient ratings across alliance sessions should be considered a tentative finding. Future research directly assessing this comparison is necessary.

Most of the alliance scales have been shown to be related to outcome. The Penn scales, the Vanderbilt scales, the WAI, and the CALPAS were moderately correlated with outcome, but the TARS failed to receive support. Because the TARS has not been associated with outcome, researchers should avoid this scale and choose among the other established alliance scales in their attempts to associate the alliance with outcome. In addition, the Penn scales, the Vanderbilt scales, the WAI, and the CALPAS have received far more empirical scrutiny than any of the other alliance scales and therefore should be used in future research studies unless persuasive reasons exist to develop or use a different measure. As Henry, Strupp, Schacht, and Gaston (1994) noted, "given the diversity and multiplicity of alliance measures already available, it may be time for a halt in the production of new alliance scales" (p. 482).

Of these measures, the WAI is likely to be appropriate for most research projects. The scale was designed to measure alliance factors in all types of therapy and to measure the theoretical constructs underlying the alliance. The scale provides both an overall alliance score and also an assessment of Bordin's (1979) three aspects of the alliance: the bond, the agreement on goals, and the agreement on tasks. The WAI also provides an assessment of Horvath and Luborsky's (1993) two core aspects of the alliance measured by most scales: (a) therapist-patient affective attachments and (b) collaboration or willingness to invest in the therapy process. In addition, patient-, therapist-, and independent observer-rated versions of the scale are available, as are shortened versions of these scales.

The overall correlation of alliance and outcome did not seem to be influenced by publication status. Although the unpublished studies included in the meta-analysis had a slightly lower average correlation than did the published studies, the difference was not significant. This finding supports the tentative conclusion by Horvath and Symonds (1991) that the publication status of the alliance literature fails to influence the relation of the alliance and outcome. Similarly, it is highly unlikely that enough unlocated studies with null results exist in file drawers to reduce the overall alliance-outcome correlation to a level of nonsignificance. Indeed, it would take 331 studies averaging null results to reduce the correlation of the alliance and outcome to .05. This finding should assure the reader that even if the present meta-analysis failed to locate some relevant studies, the results still represent relations that exist within the population of alliance studies.

## References

References marked with an asterisk indicate studies included in the meta-analysis.

- Alexander, L. B., & Luborsky, L. (1986). The Penn Helping Alliance Scales. In L. S. Greenberg & W. M. Pinsof (Eds.), *The psychotherapeutic process: A research handbook* (pp. 325-366). New York: Guilford Press.
- \*Allen, J. G., Tarnoff, G., & Coyne, L. (1985). Therapeutic alliance and long-term hospital treatment outcome. *Comprehensive Psychiatry*, *26*, 187-194.
- \*Allen, J. G., Tarnoff, G., Coyne, L., Spohn, H. E., Buskirk, J. R., & Keller, M. W. (1986). An innovative approach to assessing outcome of long-term psychiatric hospitalization. *Hospital and Community Psychiatry*, *37*, 376-380.
- \*Andreoli, A., Frances, A. J., Gex-Fabry, M., Aapro, N., Gerin, P., & Dazord, A. (1993). Crisis intervention in depressed patients with and without *DSM-III-R* personality disorders. *Journal of Nervous and Mental Disease*, *181*, 732-737.
- \*Andreoli, A. V., Muehlebach, A., Gognalons, M., Abensur, J., Grimm, S., & Frances, A. J. (1992). Crisis intervention response and long-term outcome: A pilot study. *Comprehensive Psychiatry*, *33*, 388-396.
- \*Ankuta, G. Y. (1992). *Initial and later therapeutic alliance and psychotherapy outcome*. Unpublished doctoral dissertation, Michigan State University, East Lansing.
- \*Barkham, M., Agnew, R. M., & Culverwell, A. (1993). The California Psychotherapy Alliance Scales: A pilot study of dimensions and elements. *British Journal of Medical Psychology*, *66*, 157-165.
- Benjamin, L. S. (1974). Structural analysis of social behavior. *Psychological Review*, *81*, 392-425.
- \*Bieschke, K. J., Bowman, G. D., Hopkins, M., Levine, H., & McFadden, K. (1995). Improvement and satisfaction with short-term therapy at a university counseling center. *Journal of College Student Development*, *36*, 553-559.
- Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, Research and Practice*, *16*, 252-260.
- \*Broome, K. M. (1996). *Antisocial personality and drug abuse treatment process*. Unpublished doctoral dissertation, Texas Christian University, Fort Worth.
- \*Card, C. A. L. (1991). *Therapeutic alliance, complementarity, and distress during stress prevention training after HIV testing*. Unpublished doctoral dissertation, New School for Social Research, New York, NY.
- \*Castonguay, L. G., Goldfried, M. R., Wiser, S., Raue, P. J., & Hayes, A. M. (1996). Predicting the effect of cognitive therapy for depression: A study of unique and common factors. *Journal of Consulting and Clinical Psychology*, *64*, 497-504.
- \*Cislo, D. A. (1988). *Client perceptions of the therapeutic alliance and session outcome*. Unpublished doctoral dissertation, University of Toledo, Toledo, OH.
- \*Clarkin, J. F., & Crilly, J. L. (1987). Therapeutic alliance and hospital treatment outcome. *Hospital and Community Psychiatry*, *38*, 871-875.
- \*Crits-Christoph, P., Cooper, A., & Luborsky, L. (1988). The accuracy of therapists' interpretations and the outcome of dynamic psychotherapy. *Archives of General Psychiatry*, *56*, 490-495.
- \*Eaton, T. T., Abeles, N., & Gutfreund, M. J. (1988). Therapeutic alliance and outcome: Impact of treatment length and pretreatment symptomatology. *Psychotherapy*, *25*, 536-542.
- \*Feeley, W. M. (1993). *Treatment components of cognitive therapy for major depression: The good, the bad, and the inert*. Unpublished doctoral dissertation, University of Pennsylvania, Philadelphia.
- \*Ferleger, N. A. (1993). *The relationship of dependency and self-criticism to alliance, complementarity and outcome in short-term dynamic psy-*

- chotherapy. Unpublished doctoral dissertation, Fordham University, Bronx, NY.
- \*Forman, N. W. (1990). *The nature of trait empathy in clients with chronic pain and their counselors and its impact on the development of the working alliance and outcome*. Unpublished doctoral dissertation, Ohio State University, Columbus.
- \*Frank, A. F., & Gunderson, J. G. (1990). The role of the therapeutic alliance in the treatment of schizophrenia: Relationship to course and outcome. *Archives of General Psychiatry*, *47*, 228–236.
- Freud, S. (1958). The dynamics of transference. In J. Starbuck (Ed. & Trans.), *The standard edition of the complete psychological works of Sigmund Freud* (Vol. 12, pp. 99–108). London: Hogarth Press. (Original work published 1912)
- Freud, S. (1958). On the beginning of treatment: Further recommendations on the technique of psychoanalysis. In J. Starbuck (Ed. & Trans.), *The standard edition of the complete psychological works of Sigmund Freud* (Vol. 12, pp. 122–144). London: Hogarth Press. (Original work published 1913)
- \*Gallop, R., Kennedy, S. H., & Stern, S. D. (1994). Therapeutic alliance on an inpatient unit for eating disorders. *International Journal of Eating Disorders*, *16*, 405–410.
- Gaston, L. (1990). The concept of the alliance and its role in psychotherapy: Theoretical and empirical considerations. *Psychotherapy*, *27*, 143–153.
- Gaston, L. (1991). Reliability and criterion-related validity of the California Psychotherapy Alliance Scales—Patient Version. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, *3*, 68–74.
- \*Gaston, L., Marmar, C. R., Gallagher, D., & Thompson, L. W. (1991). Alliance prediction of outcome beyond in-treatment symptomatic change as psychotherapy progresses. *Psychotherapy Research*, *1*, 104–112.
- \*Gaston, L., Piper, W. E., Debbane, E. G., Bienvenu, J. P., & Garant, J. (1994). Alliance and technique for predicting outcome in short- and long-term analytic psychotherapy. *Psychotherapy Research*, *4*, 121–135.
- \*Gerstley, L., McLellan, A. T., Alterman, A. I., Woody, G. E., Luborsky, L., & Prout, M. (1989). Ability to form an alliance with the therapist: A possible marker of prognosis for patients with antisocial personality disorder. *American Journal of Psychiatry*, *146*, 508–512.
- \*Gomes-Schwartz, B. (1978). Effective ingredients in psychotherapy: Prediction of outcome from process variables. *Journal of Consulting and Clinical Psychology*, *46*, 1023–1035.
- \*Greenberg, L. S., & Webster, M. C. (1982). Resolving decisional conflict by Gestalt two-chair dialogue: Relating process to outcome. *Journal of Counseling Psychology*, *29*, 468–477.
- Greenson, R. R. (1965). The working alliance and the transference neuroses. *Psychoanalysis Quarterly*, *34*, 155–181.
- Greenson, R. R. (1967). *Technique and practice of psychoanalysis*. New York: International Universities Press.
- \*Grob, M. C., & Eisen, S. V. (1989). Most likely to succeed: Correlates of good versus poor hospital outcome in young adult inpatients. *Psychiatric Hospital*, *20*, 23–30.
- \*Gunther, G. J. (1992). *Therapeutic alliance, patient object relations and outcome in psychotherapy*. Unpublished doctoral dissertation, Michigan State University, East Lansing.
- \*Gutfreund, M. J. (1992). *Therapist interventions: Their relation to therapeutic alliance and outcome in dynamic psychotherapy*. Unpublished doctoral dissertation, Michigan State University, East Lansing.
- \*Hansson, L., & Berglund, M. (1992). Stability of therapeutic alliance and its relationship to outcome in short-term inpatient psychiatric care. *Scandinavian Journal of Social Medicine*, *20*, 45–50.
- \*Hartley, D., & Strupp, H. (1983). The therapeutic alliance: Its relationship to outcome in brief psychotherapy. In J. Masling (Ed.), *Empirical studies of psychoanalytic theories* (pp. 1–27). Hillsdale, NJ: Erlbaum.
- \*Hatcher, R. L., & Barends, A. (1996). Patients' view of the alliance in psychotherapy: Exploratory factor analysis of three alliance measures. *Journal of Consulting and Clinical Psychology*, *64*, 1326–1336.
- Hatcher, R. L., Barends, A., Hansell, J., & Gutfreund, M. J. (1995). Patients' and therapists' shared and unique views of the therapeutic alliance: An investigation using confirmatory factor analysis in a nested design. *Journal of Consulting and Clinical Psychology*, *63*, 636–643.
- \*Hays, V. L. (1994). *The effects of the therapeutic alliance and social support on therapy outcome and mental health of women*. Unpublished doctoral dissertation, University of Wisconsin, Madison.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego, CA: Academic Press.
- Henry, W. P., Schacht, T. E., & Strupp, H. H. (1986). Structural analysis of social behavior: Application to a study of interpersonal process in differential psychotherapeutic outcome. *Journal of Consulting and Clinical Psychology*, *54*, 27–31.
- Henry, W. P., & Strupp, H. H. (1994). The therapeutic alliance as interpersonal process. In A. O. Horvath & L. S. Greenberg (Eds.), *The working alliance: Theory, research, and practice* (pp. 51–84). New York: Wiley.
- Henry, W. P., Strupp, H. H., Schacht, T. E., & Gaston, L. (1994). Psychodynamic approaches. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (4th ed., pp. 467–508). New York: Wiley.
- \*Hopkins, W. E. (1988). *The effects of conceptual level matching on the working alliance and outcome in time-limited counseling*. Unpublished doctoral dissertation, University of Maryland, College Park.
- \*Horowitz, M. J., Marmar, C., Weiss, D. S., DeWitt, K. N., & Rosenbaum, R. (1984). Brief psychotherapy of bereavement reactions: The relationship of process to outcome. *Archives of General Psychiatry*, *41*, 438–448.
- \*Horvath, A. O. (1981). *An exploratory study of the working alliance: Its measurement and relationship to outcome*. Unpublished doctoral dissertation, University of British Columbia, Vancouver, British Columbia, Canada.
- Horvath, A. O. (1994). Empirical validation of Bordin's pantheoretical model of the alliance: The Working Alliance Inventory perspective. In A. O. Horvath & L. S. Greenberg (Eds.), *The working alliance: Theory, research, and practice* (pp. 109–128). New York: Wiley.
- Horvath, A. O., & Greenberg, L. S. (1986). The development of the Working Alliance Inventory. In L. S. Greenberg & W. M. Pinsof (Eds.), *The psychotherapeutic process: A research handbook* (pp. 529–556). New York: Guilford Press.
- Horvath, A. O., & Greenberg, L. S. (1989). Development and validation of the Working Alliance Inventory. *Journal of Counseling Psychology*, *36*, 223–233.
- Horvath, A. O., & Luborsky, L. (1993). The role of the therapeutic alliance in psychotherapy. *Journal of Consulting and Clinical Psychology*, *61*, 561–573.
- Horvath, A. O., & Symonds, B. D. (1991). Relation between working alliance and outcome in psychotherapy: A meta-analysis. *Journal of Counseling Psychology*, *38*, 139–149.
- Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting errors and bias in research findings*. Newbury Park, CA: Sage.
- \*Jumes, M. T. (1995). *The developmental course of the inpatient working alliance in state psychiatric hospital clients*. Unpublished doctoral dissertation, University of Texas Southwestern Medical Center, Dallas.
- \*Kivlighan, D. M., & Shaughnessy, P. (1995). Analysis of the development of the working alliance using hierarchical linear modeling. *Journal of Counseling Psychology*, *42*, 338–349.
- \*Klee, M. R., Abeles, N., & Muller, R. T. (1990). Therapeutic alliance: Early indicators, course, and outcome. *Psychotherapy*, *27*, 166–174.
- \*Kokotovic, A. M., & Tracey, T. J. (1990). Working alliance in the early phase of counseling. *Journal of Counseling Psychology*, *37*, 16–21.
- \*Kolden, G. G. (1996). Change in early sessions of dynamic therapy:

- Universal processes and the generic model of psychotherapy. *Journal of Consulting and Clinical Psychology*, 64, 489–496.
- \*Krupnick, J. L., Elkin, I., Collins, J., Simmens, S., Sotsky, S. M., Pilkonis, P. A., & Watkins, J. T. (1994). Therapeutic alliance and clinical outcome in the NIMH Treatment of Depression Collaborative Research Program: Preliminary findings. *Psychotherapy*, 31, 28–35.
- \*Krupnick, J. L., Sotsky, S. M., Simmens, S., Moyer, J., Elkin, I., Watkins, J., & Pilkonis, P. A. (1996). The role of the therapeutic alliance in psychotherapy and pharmacotherapy outcome: Findings in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *Journal of Consulting and Clinical Psychology*, 64, 532–539.
- Lambert, M. J., & Bergin, A. E. (1994). The effectiveness of psychotherapy. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (4th ed., pp. 143–189). New York: Wiley.
- Lambert, M. J., & Hill, C. E. (1994). Assessing psychotherapy outcomes and processes. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (4th ed., pp. 72–113). 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.
- \*Lieberman, P. B., von Rehn, S., Dickie, E., Elliot, B., & Egerter, E. (1992). Therapeutic effects of brief hospitalization: The role of a therapeutic alliance. *Journal of Psychotherapy: Practice and Research*, 1, 56–63.
- Luborsky, L. (1976). Helping alliances in psychotherapy. In J. L. Cleghorn (Ed.), *Successful psychotherapy* (pp. 92–116). New York: Brunner/Mazel.
- Luborsky, L. (1984). *Principles of psychodynamic psychotherapy*. New York: Basic Books.
- \*Luborsky, L., Crits-Christoph, P., Alexander, L., Margolis, M., & Cohen, M. (1983). Two helping alliance methods for predicting outcomes of psychotherapy: A counting signs vs. a global rating method. *Journal of Nervous and Mental Disease*, 171, 480–490.
- \*Luborsky, L., McLellan, A. T., Woody, G. E., O'Brien, C. P., & Auerbach, A. (1985). Therapist success and its determinants. *Archives of General Psychiatry*, 42, 602–611.
- \*Mallinckrodt, B. (1993). Session impact, working alliance, and treatment outcome in brief counseling. *Journal of Counseling Psychology*, 40, 25–32.
- \*Mallinckrodt, B. (1996). Change in working alliance, social support, and psychological symptoms in brief therapy. *Journal of Counseling Psychology*, 43, 448–455.
- \*Marmar, C. R., Gaston, L., Gallagher, D., & Thompson, L. W. (1989). Alliance and outcome in late-life depression. *Journal of Nervous and Mental Disease*, 177, 464–472.
- \*Marmar, C. R., Weiss, D. S., & Gaston, L. (1989). Towards the validation of the California Therapeutic Alliance Rating System. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, 1, 46–52.
- \*Marshel, R. L. (1986). *An exploration of the relationship between the therapeutic alliance and premature termination*. Unpublished doctoral dissertation, Northwestern University, Evanston, IL.
- \*Marziali, E. (1984). Three viewpoints on the therapeutic alliance scales similarities, differences and associations with psychotherapy outcome. *Journal of Nervous and Mental Disease*, 172, 417–423.
- \*Marziali, E., Marmar, C., & Krupnick, J. (1981). Therapeutic alliance scales: Development and relationship to psychotherapy outcome. *American Journal of Psychiatry*, 138, 361–364.
- Matt, G. E., & Navarro, A. M. (1997). What meta-analyses have and have not taught us about psychotherapy effects: A review and future directions. *Clinical Psychology Review*, 17, 1–32.
- \*Mohl, P. C., Martinez, D., Ticknor, C., Huang, M., & Cordell, L. (1991). Early dropout from psychotherapy. *Journal of Nervous and Mental Disease*, 179, 478–481.
- \*Morgan, R., Luborsky, L., Crits-Christoph, P., Curtis, H., & Solomon, J. (1982). Predicting the outcome of psychotherapy by the Penn Helping Alliance Rating Method. *Archives of General Psychiatry*, 39, 397–402.
- \*Moseley, D. (1983). *Therapeutic relationship and its association with outcome*. Unpublished master's thesis, University of British Columbia, Vancouver, British Columbia, Canada.
- \*Muran, J. C., Gorman, B. S., Safran, J. D., Twining, L., Samstag, L. W., & Winston, A. (1995). Linking in-session change to overall outcome in short-term cognitive therapy. *Journal of Consulting and Clinical Psychology*, 63, 651–657.
- \*O'Malley, S. S., Suh, C. S., & Strupp, H. H. (1983). The Vanderbilt Psychotherapy Process Scale: A report on the scale development and a process–outcome study. *Journal of Consulting and Clinical Psychology*, 51, 581–586.
- Orlinsky, D. E., & Howard, K. I. (1966). *Psychotherapy Session Report, Form P and Form T*. Chicago: Institute of Juvenile Research.
- Orlinsky, D. E., & Howard, K. I. (1986). Process and outcome in psychotherapy. In S. L. Garfield & A. E. Bergin (Eds.), *Handbook of psychotherapy and behavior change* (3rd ed., pp. 311–381). New York: Wiley.
- \*Piper, W. E., Azim, H. F. A., Joyce, A. S., McCallum, M., Nixon, G. W. H., & Segal, P. S. (1991). Quality of object relations versus interpersonal functioning as predictors of therapeutic alliance and psychotherapy outcome. *Journal of Nervous and Mental Disease*, 179, 432–438.
- \*Piper, W. E., Boroto, D. R., Joyce, A. S., McCallum, M., & Azim, H. F. A. (1995). Pattern of alliance and outcome in short-term individual psychotherapy. *Psychotherapy*, 32, 639–647.
- \*Priebe, S., & Gruyters, T. (1993). The role of the helping alliance in psychiatric community care: A prospective study. *Journal of Nervous and Mental Disease*, 181, 552–557.
- \*Prigatano, G. P., Klonoff, P. S., O'Brien, K. P., Altman, I. M., Amin, K., Chiapello, D., Shepherd, J., Cunningham, M., & Mora, M. (1994). Productivity after neuropsychologically oriented milieu rehabilitation. *Journal of Head Trauma Rehabilitation*, 9, 91–102.
- \*Pugh, M. A. (1991). *The working alliance, pretherapy interpersonal relationships, and therapeutic outcome*. Unpublished doctoral dissertation, Virginia Commonwealth University, Richmond.
- \*Pyne, S. C. (1991). *Therapists' use of countertransference in short-term psychotherapy: A focus on rescuing and counterhostility*. Unpublished doctoral dissertation, University of Colorado, Boulder.
- \*Reiner, P. A. (1987). *The development of the therapeutic alliance*. Unpublished doctoral dissertation, University of North Carolina, Chapel Hill.
- \*Riley, J. D. (1992). *The impact of pre-therapy and extra-therapy role preparation on the development of therapeutic alliance and outcome*. Unpublished doctoral dissertation, Pacific Graduate School of Psychology, Palo Alto, CA.
- Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Newbury Park, CA: Sage.
- \*Rounsaville, B. J., Chevron, E. S., Prusoff, B. A., Elkin, I., Imber, S., Sotsky, S., & Watkins, J. (1987). The relation between specific and general dimensions of the psychotherapy process in interpersonal psychotherapy for depression. *Journal of Consulting and Clinical Psychology*, 55, 379–384.
- Safran, J. D., & Muran, J. C. (Eds.). (1995). The therapeutic alliance [Special issue]. In *Session: Psychotherapy in Practice*, 1(1).
- \*Safran, J. D., & Wallner, L. K. (1991). The relative predictive validity of two therapeutic alliance measures in cognitive therapy. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, 3, 188–195.
- \*Saunders, S. M., Howard, K. I., & Orlinsky, D. E. (1989). The Thera-

- peutic Bond Scales: Psychometric characteristics and relationship to treatment effectiveness. *Psychological Assessment*, 1, 323-330.
- \*Sexton, H. (1996). Process, life events, and symptomatic change in brief eclectic psychotherapy. *Journal of Consulting and Clinical Psychology*, 64, 1358-1365.
- Smith, M. L., Glass, G. V., & Miller, T. I. (1980). *The benefits of psychotherapy*. Baltimore: Johns Hopkins University Press.
- \*Solomon, P., Draine, J., & Delaney, M. A. (1995). The working alliance and consumer case management. *Journal of Mental Health Administration*, 22, 126-134.
- \*Sonnenberg, R. T. (1996). *An examination of the outcome of a brief psychiatric hospitalization with particular reference to the therapeutic alliance and select patient characteristics*. Unpublished doctoral dissertation, University of Nebraska, Lincoln.
- Sterba, R. (1934). The fate of the ego in analytic therapy. *International Journal of Psychoanalysis*, 15, 117-126.
- Stiles, W. B., Shapiro, D., & Elliot, R. (1986). Are all psychotherapies equivalent? *American Psychologist*, 41, 165-180.
- Strupp, H. H., & Binder, J. L. (1984). *Psychotherapy in a new key: A guide to time-limited dynamic psychotherapy*. New York: Basic Books.
- Suh, C. S., Strupp, H. H., & O'Malley, S. S. (1986). The Vanderbilt process measures: The Psychotherapy Process Scale (VPPS) and the Negative Indicators Scale (VNIS). In L. S. Greenberg & W. M. Pinsof (Eds.), *The psychotherapeutic process: A research handbook* (pp. 285-323). New York: Guilford Press.
- \*Svartberg, M., & Stiles, T. C. (1994). Therapeutic alliance, therapist competence, and client change in short-term anxiety-provoking psychotherapy: Annual meeting of the Society for Psychotherapy Research. *Psychotherapy Research*, 4, 20-33.
- \*Tichenor, V. (1989). *Working alliance: A measure of comparison*. Unpublished doctoral dissertation, University of Maryland, College Park.
- Tichenor, V., & Hill, C. E. (1989). A comparison of six measures of the working alliance. *Psychotherapy: Research and Practice*, 26, 195-199.
- Tracey, T. J., & Kokotovic, A. M. (1989). Factor structure of the Working Alliance Inventory. *Psychological Assessment*, 1, 207-210.
- \*Tryon, G. S., & Kane, A. S. (1990). The helping alliance and premature termination. *Counselling 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.
- \*Tunis, S. L., Delucchi, K. L., Schwartz, K., Banys, P., & Sees, K. L. (1995). The relationship of counselor and peer alliance to drug use and HIV risk behaviors in a six-month methadone detoxification program. *Addictive Behaviors*, 20, 395-405.
- \*Windholz, M. J., & Silbershatz, G. (1988). Vanderbilt Psychotherapy Process Scale: A replication with adult outpatients. *Journal of Consulting and Clinical Psychology*, 56, 56-60.
- Wolfe, B. E., & Goldfried, M. R. (1988). Research on psychotherapy integration: Recommendations and conclusions from an NIMH workshop. *Journal of Consulting and Clinical Psychology*, 56, 448-451.
- \*Yeomans, F. E., Gutfreund, J., Selzer, M. A., Clarkin, J. F., Hull, S. W., & Smith, T. E. (1994). Factors related to drop-outs by borderline patients: Treatment contract and therapeutic alliance. *Journal of Psychotherapy: Practice and Research*, 3, 16-24.
- Zetzel, E. R. (1956). Current concepts of transference. *International Journal of Psychoanalysis*, 37, 369-376.

Received June 28, 1999

Revision received October 6, 1999

Accepted October 14, 1999 ■