Perspective Is Everything

The Predictive Validity of Six Working Alliance Instruments

Lisa R. Fenton, Psy.D.
John J. Cecero, Ph.D.
Charla Nich, M.S.
Tami L. Frankforter, B.S.
Kathleen M. Carroll, Ph.D.

The predictive validity of instruments commonly used to measure the therapeutic alliance was evaluated, using 46 sessions drawn from a clinical trial comparing manual-guided therapies for substance use. The California Psychotherapy Alliance Scale, Penn Helping Alliance Rating Scale, Vanderbilt Therapeutic Alliance Scale, and Working Alliance Inventory (Observer, Therapist, and Client versions) were rated for participants receiving either cognitive-behavioral therapy or twelve-step facilitation. All observer-rated instruments were significantly correlated with outcome; however, therapist-rated and client-rated instruments did not predict outcome. Findings suggest that the different observer-rated instruments are minimally different with respect to predictive validity, whereas patient- and therapist-rated measures may have a weaker relationship to outcome when highly objective outcome measures are used.

(The Journal of Psychotherapy Practice and Research 2001; 10:262–268)

The relationship between therapeutic alliance and clinical outcome has been demonstrated across varied types of treatment.[1] Clinical trials employing cognitive, interpersonal, behavioral, and psychodynamic therapies have demonstrated the robust nature of this finding.[2,3] A strong alliance has been associated with an improved outcome in the treatment of a variety of psychological problems, including depression,[4] personality disorders,[5] alcohol dependence,[6] and cocaine dependence.[7]

Although the alliance–outcome relationship is well established, there is little agreement on the best way to measure the therapeutic alliance. Currently, there are at least 11 commonly used instruments available that vary in perspective (observer, client, or therapist) and in theoretical orientation.[8] Although each of these appears to tap into several common core aspects of the construct,[1] including client–therapist agreement on the

goals of treatment, strategies to achieve those goals, and the quality of shared affective bond, few studies have directly compared the psychometric properties of these instruments.

In one of the first studies that did address this issue, Tichenor and Hill[9] compared six working alliance instruments: the California Psychotherapy Alliance Scales (CALPAS), the Penn Helping Alliance Rating Scale (Penn), the Vanderbilt Therapeutic Alliance Scale (VTAS), and the Working Alliance Inventory–Observer, Therapist, and Client forms (WAI-O, WAI-T, WAI-C), using a sample of depressed outpatients. Results showed that all instruments had high internal consistency and good interrater reliability. In addition, the CALPAS, VTAS, and WAI-O were highly correlated, and the Penn was correlated to the WAI-O. Interestingly, client- and therapist-rated versions of the WAI were not significantly related to each other or to any other observer-rated instrument.

Cecero et al.[10] expanded upon the Tichenor and Hill study by evaluating the psychometric properties of the same instruments with a larger, substance-dependent sample. This study was one of the few in which data were drawn from a controlled clinical trial and involved discriminable, manual-guided treatments. Results suggested that all instruments were internally consistent and had acceptable levels of interrater reliability.

Although these data suggest that these commonly used instruments are psychometrically sound and comparable, their relative predictive validity is still unknown. In one of the few studies addressing this issue, Safran and Wallner[11] compared the predictive validity of client-rated versions of the WAI and CALPAS. Results indicated that both were predictive of some outcome measures in a short-term cognitive therapy for depression. Although the study provides valuable information about the predictability of outcome from the client's perspective, results may not generalize to observer-rated instruments or to other types of therapy. To date, the comparative predictive validity of the more commonly used observer-rated alliance instruments has not been addressed.

In the current study, we evaluated the predictive validity of six instruments commonly used to measure the therapeutic alliance: the WAI-O, WAI-C, and WAI-T;[12] the Penn;[13] the VTAS;[14] and the CALPAS (C.R. Marmar & L. Gaston, 1988, unpublished manual), using data drawn from a randomized clinical trial of manu-

alized psychotherapies for cocaine dependence. The following research questions were addressed: First, are the instruments comparable with respect to correlations with outcome? Second, does predictive validity of the instruments vary across two types of treatment, particularly types that differ greatly with respect to theoretical foundation and technique? Third, does predictive validity vary across rating perspective (observer-, therapist- and patient-rated versions) of alliance instruments?

METHODS

The data for this study were drawn from psychotherapy sessions that were part of a randomized clinical trial evaluating treatment for comorbid cocaine and alcohol dependence.[15] Cognitive-behavioral therapy (CBT) and twelve-step facilitation (TSF) were compared with a psychotherapy control condition (clinical management). The treatments were delivered over 12 weeks in individual sessions, and all sessions were videotaped for assessment of therapist adherence and competence. Independent evaluators who were blind to treatment assignment rated 90% of all sessions and found that treatments were highly discriminable and therapists were highly adherent to manual guidelines.[16] In addition, the psychometric properties of the alliance instruments had been established in a previous study using these data.[10]

For the present study, alliance–outcome relationships were evaluated by using only CBT and TSF sessions. This choice was made because 1) we wished to evaluate the "active" rather than the supportive control condition, which was found to be less effective than CBT and TSF (and which would have therefore complicated evaluation of alliance–outcome relationships across instruments); 2) focusing on treatments widely used in clinical practice would maximize the utility of results; and 3) the smaller sample size of the psychotherapy control condition ($n = 14$) would have precluded clear evaluation of the predictive validity of the instruments across treatments.

Clients and therapists were asked to complete their respective versions of the WAI after the third session. We included only those sessions where both the client and therapist completed the form on the same day. This was done so that all perspectives could be compared for the same session. Early session ratings were selected both to avoid bias that might be associated with attrition

and because early measures of alliance have been found to be predictive of outcome.[8]

Of the 90 clients assigned to CBT and TSF, 44 were excluded because of data missing for the following reasons: 1) client dropped out prior to completing the second week of treatment ($n=19$); 2) therapist and/or client failed to complete the WAI on the same day ($n=16$); 3) videotaping equipment had failed ($n=6$); or 4) staff inadvertently did not give instruments to the therapist or client ($n=3$).

Of the 46 participants in the subset, 25 received TSF and 21 received CBT.

Subjects

All clients participating in this study were treated at the Clinical Research Unit of the APT Foundation, a nonprofit substance abuse treatment center affiliated with Yale University School of Medicine in New Haven, CT, and met DSM-III-R criteria for cocaine and alcohol dependence or abuse. Exclusion criteria included a history of psychotic or manic episodes, or current suicidal or homicidal plans.

Therapists

The six therapists who delivered study treatments were experienced in and committed to the type of treatment they delivered in the trial. Level of training varied by type of treatment: four Ph.D. psychologists conducted CBT and two M.A.-level clinicians conducted TSF.

Raters and Training

Six clinicians (one male, five female; three at Ph.D., two at M.A., and one at B.A. level) served as raters for the study. Raters were divided into three pairs and each pair was assigned to a different instrument (VTAS, CALPAS, or Penn) to prevent contamination of ratings across instruments. In addition, all raters were trained to use the WAI to allow for a direct comparison of reliability across the three rater pairs. The WAI was chosen as the "standard" because it was considered the most commonly used instrument in this population and previous reports indicated that relatively little training was required to achieve interrater reliability on the instrument.[9]

Rater training was administered by the first two authors and required approximately 16 hours for each rater pair. After attending a didactic seminar that included an introduction to the instruments and an item-by-item review, the raters rated and discussed a single session tape, using both the WAI and the measure assigned to their pair. Raters were then given two session tapes to rate on their own. These ratings were used to provide an informal estimate of interrater reliability. This procedure was repeated until the raters achieved satisfactory interrater reliability (intraclass correlation coefficients of 0.60 or above). Recalibration sessions were provided twice during the study to prevent and correct rater drift.

Working Alliance Measures

The Working Alliance Inventory (WAI) is designed to capture Bordin's[17] pantheoretical perspective of the working alliance. The 36-item instrument consists of three subscales: the Goal subscale addresses the extent to which therapy goals are important, mutual, and capable of being accomplished; the Task subscale focuses on the participant's agreement about the steps taken to help improve the client's situation; and the Bond subscale measures mutual liking and attachment by focusing on tone of voice, empathy, and comfort in exploring intimate issues.

The California Psychotherapy Alliance Scale (CALPAS) draws from a variety of perspectives including Freud's[18] concept of the patient's affective bond with the therapist, the patient's ego capacity for a working alliance,[19] mutual agreement on tasks and goals,[17] and the therapist's role as an empathic listener.[20] The CALPAS has 24 items grouped into four subscales: patient working capacity, patient commitment, patient–therapist agreement on goals and strategies, and therapist understanding and involvement.

The Penn Helping Alliance Rating Scale (Penn) focuses on two types of alliance. Type I reflects a psychoanalytic focus on the client's affective bond with the therapist,[18] and Type II appears more closely related to Bordin's concept of a mutual agreement on tasks and goals.[17] The Penn, with 10 items, is the shortest of the scales.

The Vanderbilt Therapeutic Alliance Scale (VTAS) represents a theoretical blend of dynamic and eclectic frameworks.[8] The instrument attributes a successful alliance to the presence or absence of six factors: positive climate, therapist intrusiveness, client resistance or

anxiety,[21] client motivation,[22] and client responsibility.[17] The VTAS consists of 44 items within three subscales: therapist contribution to the alliance, client contribution to the alliance, and client–therapist interactions.

Outcome Assessment

Clients were assessed before treatment, weekly during treatment, and at termination by an independent clinical evaluator. The primary outcome measure was maximum consecutive days abstinent from cocaine while in treatment. This measure was selected because it serves as a good compound measure of both retention and substance use, and has shown to be highly predictive of long-term outcome.[23]

Procedure

Following rater training, the study was conducted in two stages. First, the three rater-pairs rated the same 8 sessions using both their assigned instrument and the WAI-O. These data were used to establish reliability. Second, the remaining 38 sessions were divided between members of each pair so that for each of these sessions, the CALPAS, VTAS, and Penn were completed one time and the WAI-O was completed three times. Thus, each member of a pair rated a total of 27 sessions (8 reliability tapes and half of the total remaining tapes) by using their assigned instrument and the WAI-O.

RESULTS AND DISCUSSION

Psychometric Properties

Intraclass correlation coefficients were used to assess interrater reliability for each instrument. Estimates were based on the mean of the ratings for each rater pair. For the 8 sessions that were rated by all raters, random-effect intraclass correlation estimates were as follows: WAI-O = 0.70; CALPAS = 0.76; VTAS = 0.60; and Penn = 0.71.

Means and standard deviations for alliance instruments were as follows: Penn: 3.93 ± 1.16; CALPAS: 5.20 ± 0.89; VTAS: 4.39 ± 0.78; WAI-O: 4.38 ± 0.099; WAI-T: 5.11 ± 0.60; WAI-C: 5.76 ± 0.88.

Pearson correlations of the six alliance instruments

are presented in Table 1. Higher correlations were generally found among observer measures; client and therapist measures were less consistently intercorrelated at a significant level.

Predictive Validity

As seen in Table 2, Pearson correlations between all four observer instruments and outcome were statistically significant, a stronger alliance being associated with longer periods of abstinence during treatment. However, correlations between the WAI-C and WAI-T and outcome were not significant.

Predictive Validity by Treatment

Table 2 shows some variability among relationships between outcome and alliance across scales when cor-

TABLE 1. Intercorrelations of six therapeutic alliance instruments

	Penn	VTAS	WAI-O	WAI-C	WAI-T
CALPAS	0.62**	0.38*	0.37	0.31	0.51**
Penn		0.49**	0.53**	0.36	0.44*
VTAS			0.87**	0.02	0.36
WAI-O				0.09	0.36
WAI-C					0.43*

➡ *Note:* Pearson correlations. $n = 47$ for all scales, except $n = 46$ for WAI-C and WAI-T. CALPAS = California Psychotherapy Alliance Scale; Penn = Pennsylvania Helping Alliance Rating Scale; VTAS = Vanderbilt Therapeutic Alliance Scale; WAI = Working Alliance Inventory; O = Observer, C = Client, T = Therapist.
*$P<0.01$; **$P<0.001$.

TABLE 2. Correlations between alliance and outcome by instrument

	All Treatments	By Treatment	
		CBT	TSF
Instrument	($n = 46$)	($n = 21$)	($n = 25$)
Penn	0.50**	0.63**	0.42*
CALPAS	0.37**	0.56**	0.28
VTAS	0.49**	0.46**	0.55**
WAI-O	0.39**	0.32	0.48*
WAI-C[a]	0.03	0.07	0.00
WAI-T[a]	0.27	0.27	0.30

➡ *Note:* Outcome = maximum consecutive days abstinent from cocaine while in treatment. CBT = cognitive-behavioral therapy; TSF = twelve-step facilitation. For other abbreviations, see Table 1.
[a]$n = 45$ for all treatments, $n = 20$ for CBT, and $n = 25$ for TSF because of partially completed forms.
*$P<0.01$; **$P<0.001$.

relations are evaluated by treatment. In the CBT condition, all observer instruments with the exception of the WAI-O showed significant correlations between alliance and outcome. In the TSF condition, all observer instruments except the CALPAS showed significant correlations between alliance and outcome. Again, neither the WAI-C nor the WAI-T was significantly correlated with outcome for either CBT or TSF.

This evaluation of the predictive validity of six therapeutic alliance instruments across two types of manual-guided treatments indicated that there were significant correlations between alliance and outcome for all observer-rated instruments. Thus, our results were reminiscent of those of Luborsky et al.[24] when they cited the "dodo bird" verdict suggesting that "everyone has won and all must have prizes." In other words, these findings suggest that investigators are free to select among observer-rated alliance instruments on the basis of other considerations, such as continuity with previous research, theoretical consistency with treatment orientation, ease of administration, or the availability of parallel patient, therapist, and observer versions when needed.

We also evaluated the predictive validity of each instrument across type of treatment and found some variability in the results. The Penn and the VTAS predicted outcome across treatments; however, the CALPAS predicted outcome only for CBT and the WAI predicted outcome only for TSF. The CALPAS and the WAI are both designed to be transtheoretical instruments and have been found to be highly intercorrelated in other studies, although they appear to measure slightly different aspects of the alliance.[11] Consequently, there is no theoretical reason to conclude that either is a superior measure of alliance in one particular type of therapy. It should be noted that the smaller sample sizes for the within-treatment comparison limit the strength of these findings and suggest that the results should be interpreted with caution.

The comparatively poor predictive validity of therapist- and client-rated measures was an unexpected result of the study. The bulk of the literature in this area suggests client ratings of the alliance tend to be strongly related to outcome and therapist ratings are weaker predictors of outcome.[1] There are several possible explanations for this unexpected effect.

First, the outcome measure used was of a different type from those in most prior studies. Much of the literature on alliance–outcome relationships has been in the field of depression and anxiety, where outcomes tend to be subjective assessments evaluated from the patient perspective (e.g., Beck Depression Inventory scores). In the present study, the outcome measure was highly objective (e.g., retention and substance use confirmed by urinalysis), and outcome was measured not by the clients or therapists, but by a clinical evaluator who was blind to treatment assignment. When assessments are more subjective as in the earlier studies, it is possible that alliance–outcome relationships may be influenced by a "halo" effect where the alliance and outcome are rated by the same source. (That is, if the patient believes she or he has improved, it is also likely she or he will rate the alliance as positive.) Indeed, Horvath and Symonds[1] noted that effect sizes were larger when the data were derived from homogeneous sources (e.g., patient ratings as predictors of outcome measured from the patient's perspective, versus from the therapist's or an independent evaluator's perspective). These authors suggest that the difference between heterogeneous and homogeneous sources "was probably not systematically related to the 'source' factor." However, Safran and Wallner[11] found the strongest significant correlations in their sample between client-rated alliance and client-rated global success, with smaller correlations between client-rated alliance and therapist-rated outcomes.

Second, this was one of the few studies evaluating alliance–outcome relationship from the client, therapist, and observer perspectives, so there were limited prior data suggesting how alliance ratings of the three were likely to compare. It should be noted that the level of the alliance differed by perspective. The WAI indicated that alliance was rated most highly by the clients in our sample (mean $\pm$ SD = 5.76 $\pm$ 0.88), followed closely by therapists (5.11 $\pm$ 0.60), while the lowest ratings came from the observers (4.38 $\pm$ 0.99). Yet only observer ratings predicted outcome. There are several reasons client ratings of the alliance may have been relatively higher in this sample. In particular, clients with substance abuse problems often have financial constraints, and our treatment was provided free of charge in a market where treatment availability is often limited. Consequently, a sense of indebtedness or fear of expressing negative feelings about the therapist may have elevated client ratings. In addition, many clients have little basis for comparison in rating the alliance (most of our clients had never been in psychotherapy before), in contrast to

Fenton *et al.*

observers who view many dyads and might be able to provide a more balanced assessment of the relationship.

On the other hand, the higher predictive validity associated with observer-rated alliance measures may simply suggest that observers are less susceptible to situational demands or transference and countertransference issues that may influence an evaluation of the alliance. Such influences on clients and therapists may help to explain why, in the addiction literature where objective measures of outcome are available (e.g., biochemical measures), relationships of client and therapist ratings of the alliance to outcome have been weaker in comparison to the general psychotherapy literature, which often evaluates outcome based on more subjective assessments from the client or therapist perspective. For example, in the Project MATCH alcoholism study, therapist ratings of the alliance were more strongly related to outcome than client ratings; however, the overall level of variance accounted for by either perspective was quite modest[6] and the observer's perspective was not rated. Similarly, in the National Institute on Drug Abuse Cocaine Collaborative Study, where alliance was measured only from the client perspective, no significant alliance–outcome relationships were found.[25]

Several limitations of the present study should be noted. First, interrater reliability estimates were based on a sample of eight sessions. Although each of these instruments have been shown to have good interrater reliability[1,9] and recalibration meetings were held throughout the study to correct rater drift, nevertheless it is possible that drift may have occurred. Second, the small sample size for the within-treatment comparisons limits confidence in the finding that there was some variation in alliance–outcome relationships across instruments. Finally, these findings should be replicated with other treatments, instruments, and populations to evaluate the extent to which they can be generalized. Nevertheless, this remains one of the few studies that has examined process–outcome relationships across treatment conditions and the only one that has done so using several measures of the alliance.

Despite these limitations, this study contributes several important findings regarding the therapeutic alliance and the instruments with which it is measured. First, researchers should be aware that when outcome measures are highly objective, therapist- and client-rated measures of the alliance may not be strong predictors of outcome, in contrast to studies that rely on more subjective measures or in which process–outcome relationships are evaluated using only homogeneous perspectives (e.g., patient–patient, therapist–therapist). Second, regarding the commonly used observer-rated instruments, these data add to the growing body of literature showing relative psychometric comparability of the scales[9] by suggesting that different instruments used to measure working alliance are minimally different with respect to predicting outcome. With greater confidence in the comparative psychometric properties of these instruments, we can continue to explore how alliance affects treatment, how alliance interacts with other key factors that may affect treatment, and how therapists can influence some of these factors to improve the overall outcome of psychotherapy.

REFERENCES

1. Horvath AO, Symonds BD: Relation between working alliance and outcome in psychotherapy: a meta-analysis. Journal of Counseling Psychology 1991; 38:139–149
2. Krupnick JL, Sotsky SM, Simmens S, et al: 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. J Consult Clin Psychol 1996; 64:532–539
3. Gaston L, Thompson L, Gallagher D, et al: Alliance, technique, and their interactions in predicting outcome of behavioral, cognitive, and brief dynamic therapy. Psychotherapy Research 1998; 8:190–209
4. Castonguay LG, Goldfried MR, Wiser S, et al: Predicting the effect of cognitive therapy for depression: a study of unique and common factors. J Consult Clin Psychol 1996; 64:497–504
5. Hellerstein DJ, Rosenthal RN, Pinsker H, et al: A randomized prospective study comparing supportive and dynamic therapies: outcome and alliance. J Psychother Pract Res 1998; 7:261–271
6. Conners GJ, Carroll KM, DiClemente CC, et al: The therapeutic alliance and its relationship to alcoholism treatment participation and outcome. J Consult Clin Psychol 1997; 65:588–598
7. Carroll KM, Nich C, Rounsaville BJ: Contribution of the therapeutic alliance to outcome in active versus control psychotherapies. J Consult Clin Psychol 1997; 65:510–514

8. Horvath AO, Luborsky L: The role of the therapeutic alliance in psychotherapy. J Consult Clin Psychol 1993; 61:561–573

9. Tichenor V, Hill CE: A comparison of six measures of working alliance. Psychotherapy 1989; 26:195–199

10. Cecero JJ, Fenton LR, Nich C, et al: The therapeutic alliance in focus: the psychometric properties of six measures across three treatments. Psychotherapy 2001; 38:1–11

11. Safran JD, Wallner LK: The relative predictive validity of two therapeutic alliance measures in cognitive therapy. Psychol Assess 1991; 3:188–195

12. Horvath AO, Greenberg LS: The development of the Working Alliance Inventory, in The Psychotherapeutic Process: A Research Handbook, edited by Greenberg LS, Pinsof WM. New York, Guilford, 1986, pp 527–556

13. Luborsky L, Crits-Christoph P, Alexander L, et al: Two helping alliance methods for predicting outcome of psychotherapy: a counting signs vs. a global rating method. J Nerv Ment Dis 1983; 171:480–492

14. Hartley D, Strupp HH: The therapeutic alliance: its relationship to outcome in brief psychotherapy, in Empirical Studies of Psychoanalytic Theories, edited by Masling J. Hillsdale, NJ, Lawrence Erlbaum, 1983, pp 1–37

15. Carroll KM, Nich C, Ball SA, et al: Treatment of cocaine and alcohol dependence with psychotherapy and disulfiram. Addiction 1998; 93:713–728

16. Carroll KM, Nich C, Sifry RL, et al: A general system for evaluating therapist adherence and competence in psychotherapy research in the addictions. Drug and Alcohol Dependence 2000; 57:225–238

17. Bordin ES: The generalizability of the psychoanalytic concept of the working alliance. Psychotherapy: Theory, Research and Practice 1979; 61:252–260

18. Freud S: On beginning the treatment: further recommendations on the technique of psychoanalysis (1913), in The Standard Edition of the Complete Psychological Works of Sigmund Freud, vol 12, edited and translated by Strachey J. London, Hogarth Press, 1958, pp 121–144

19. Greenson RR: The working alliance and the transference neuroses. Psychoanal Q 1965; 34:155–181

20. Bowlby J: A Secure Base: Clinical Applications of Attachment Theory. London, Routledge, 1988

21. Langs R: The Therapeutic Interaction, vol 2. New York, Jason Aronson, 1976

22. Greenson RR: Technique and Practice of Psychoanalysis. New York, International University Press, 1967

23. Carroll KM, Nich C, Ball SA, et al: One year follow-up of disulfiram and psychotherapy for cocaine-alcohol abusers: sustained effects of treatment. Addiction 2000; 95:1335–1349

24. Luborsky L, Singer B, Luborsky L: Comparative studies of psychotherapies: is it true that "everyone has won and all must have prizes?" Arch Gen Psychiatry 1975; 32:995–1008

25. Barber JP, Luborsky L, Crits-Christoph P, et al: Therapeutic alliance as a predictor of outcome in treatment of cocaine dependence. Psychother Res 1999; 9:54–73