Within Treatment Therapeutic Alliance Ratings Profiles Predict Posttreatment Frequency of Alcohol Use

Mark A. Prince and Gerard J. Connors
University at Buffalo

Stephen A. Maisto
Syracuse University

Ronda L. Dearing
University at Buffalo

Although past research has demonstrated a positive relationship between the therapeutic alliance (TA) and improved drinking outcomes, specific aspects of the alliance have received less attention. In this study, we examined the association between alliance characteristics during treatment and 4-month follow-up drinking reports. Sixty-five treatment-seeking alcohol dependent clients who participated in 12 weeks of individual outpatient treatment provided weekly TA ratings during treatment and reported on pretreatment, during treatment, and posttreatment alcohol use. Latent profile analysis was conducted to discern distinct profiles of client and therapist ratings of therapeutic alliance with similar alliance characteristics. TA profiles were based on clients' and therapists' mean alliance rating, minimum alliance rating, maximum alliance rating, the range of alliance ratings, and the difference in session number between maximum and minimum alliance ratings. One- through 4-class models were fit to the data. Model fit was judged by comparative fit indices, substantive interpretability, and parsimony. Wald tests of mean equality determined whether classes differed on follow-up percentage of days abstinent (PDA) at 4-months posttreatment. Three-profile solutions provided the best fit for both client and therapist ratings of the therapeutic alliance. Client alliance rating profiles predicted drinking in the follow-up period, but therapist rating profiles did not. These results suggest that distinct profiles of the therapeutic alliance can be identified and that client alliance rating profiles are associated with frequency of alcohol use following outpatient treatment.

Keywords: therapeutic alliance, common factors in psychotherapy, alcohol use disorder, latent profile analysis

There has been considerable interest over the past several decades in the role of the therapeutic alliance in the treatment of alcohol and other substance use disorders. In a comprehensive review of this literature, Meier, Barrowclough, and Donmall (2005) concluded that the therapeutic alliance, particularly when assessed early in treatment, is a consistent predictor of treatment engagement and retention. The therapeutic alliance also was found

to predict early improvements in treatment but was less consistent in predicting posttreatment outcomes. In subsequent studies, each specific to alcohol use disorders, Hartzler and colleagues (2011) and Maisto and colleagues (2015) both found a modest relationship between the therapeutic alliance and AUD outcomes that was mediated by client self-efficacy for abstinence. In another study including consideration of self-efficacy, Ilgen, Tiet, Finney, and Moos (2006) found that a positive therapeutic alliance can mitigate the negative impact of low pretreatment self-efficacy for abstinence. Ilgen and colleagues found that a positive therapeutic alliance may be especially important for clients with low motivation to change. Taken together, the available research demonstrates a modest and generally consistent positive relationship between the therapeutic alliance and treatment outcomes. Further, a positive therapeutic alliance may counteract certain factors negatively associated with treatment outcome, such as low self-efficacy for abstinence and low motivation to change.

Although the contribution of the therapeutic alliance to treatment process and in many cases outcome is now widely acknowledged, there continue to be calls for additional research. Castonguay, Constantino, and Holtforth (2006) and Meier et al. (2005) identified the need for research elucidating the nature of the therapeutic alliance over the full course of treatment. Indeed, the majority of research in this area has been focused on assessments of the therapeutic alliance at predetermined session points (e.g.,

following the second treatment session). Dependence on such snapshot approaches precludes evaluation of the therapeutic alliance across time. If instead the therapeutic alliance was evaluated after each treatment session, a much larger array of alliance assessments would be available for study. For example, it would be possible to evaluate such dimensions as the mean of therapeutic alliance ratings across the course of treatment, the highest rating provided, and the lowest rating provided. In addition, it would be possible to calculate the difference between the highest and lowest rated sessions, determine the time (in days or sessions) between those ratings, and calculate the temporal patterns of those ratings (i.e., Does the highest session rating of the therapeutic alliance come before or after the lowest session rating?). Taken together, the assessment of the therapeutic alliance following each treatment session affords the opportunity to assess a range of variables potentially relevant to furthering our understanding of the treatment process in alcohol and other substance use disorders treatment.

A second question warranting further consideration is the perspective through which the rating of the therapeutic alliance is provided. Most of the research in this area has drawn upon the client's ratings of the therapeutic alliance. In fewer studies, therapist ratings of the therapeutic alliance also have been gathered. In the broader arena of psychotherapy research, client ratings of the therapeutic alliance have been better predictors of treatment process and outcome than therapist ratings (Horvath & Symonds, 1991). However, among outpatients receiving treatment for an alcohol use disorder, Connors et al. (1997) found that both client and therapist ratings of the therapeutic alliance predicted treatment participation and drinking behavior during and after alcohol treatment. In contrast, Fenton et al. (2001) found that neither client nor therapist alliance ratings predicted retention (although observer ratings measured after the second session did so). For the present, there appears to be value in continuing to investigate the alliance from various perspectives (e.g., client, therapist, and observer).

The purpose of this study was to address questions about the therapeutic alliance identified earlier in the context of outpatient treatment for an alcohol use disorder. Ratings of the therapeutic alliance were provided by the client and therapist following each treatment session for the duration of a treatment episode. Latent profile analysis (LPA) was used to discern distinct profiles of alliance ratings with similar characteristics, based on five unique aspects of client and therapist therapeutic alliance ratings (mean rating, minimum rating, maximum rating, range of alliance ratings (i.e., maximum–minimum), and latency between the maximum and minimum alliance ratings (i.e., session number for maximum rating–session number for minimum rating). There were two hypotheses proposed. First, it was hypothesized that latent profiles of the therapeutic alliance can be identified that distinguish the characteristics of the therapeutic alliance throughout the course of treatment. Second, it was hypothesized that latent profiles of the therapeutic alliance predict drinking frequency following a treatment episode.

Method

Participants

This study included 65 participants from the community who were recruited using local newspaper and radio advertisements and

who were currently seeking outpatient treatment for an alcohol use disorder. To be included in the study, participants were required to (a) be between 18 and 65 years old, (b) meet *DSM–IV* criteria for an alcohol use disorder (American Psychiatric Association, 1994) based on the Mini International Neuropsychiatric Interview (MINI, Sheehan et al., 1997, 1998), and (c) live within commuting distance from the program site. Participants were excluded if they (a) currently met criteria for an organic mental disorder, schizophrenia, delusional (paranoid) disorder, or any other psychotic disorders; (b) had poor performance on the structured mental status examination (MMSE; Folstein, Folstein, & McHugh, 1975) indicating gross neurocognitive impairment; or (c) had been in substance abuse treatment in the past 12 months (excluding self-help groups). Figure 1 presents the participant recruitment flowchart. Of the 81 potential participants who were scheduled for a baseline assessment 65 were eligible and participated in at least one treatment session. The final sample consisted of 62 participants who provided both alliance ratings and follow-up alcohol use data. Participants included in the final sample attended an average of 9.26 out of 12 (*SD* = 3.65) treatment sessions. The median number of sessions attended was 11, and the modal number of sessions attended was 12.

Participants were predominately male (68%) and Caucasian (75%) and had a mean age of 48.27 years (*SD* = 10.64). Approximately half (53%) of the participants reported full-time employment, 22% reported being unemployed, 10% identified as disabled, and 14% reported being retired. Roughly a third of the sample reported being currently married (35%), and a similar number reported previous (prior to the past 12 months) outpatient treatment for an alcohol use disorder (37%).

Measures

Demographics. The baseline assessment included questions about marital status, employment, and substance abuse treatment history.

Timeline follow-back. The Timeline follow-back (TLFB; Sobell & Sobell, 1992) is an interviewer administered assessment of daily alcohol use that uses a calendar-based retrospective recall approach to recording alcohol use. The TLFB was used to assess alcohol use during the baseline period, during the 12-week course of treatment, and during the follow-up period. The TLFB has consistently demonstrated reliability and accuracy for both alcohol and other substance use reporting among clients with alcohol dependence (e.g., Ehrman & Robbins, 1994; Sobell & Sobell, 1992, 1996). In the present study, we focus on percent days abstinent, given abstinence was the goal of the treatment intervention.

Working Alliance Inventory. The present study used the client and therapist versions of the 12-item (Working Alliance Inventory-Short Version (WAI-S; Horvath & Greenberg, 1986; Tracey & Kokotovic, 1989). The WAI-S is a comparable to the 36-item WAI (Busseri & Tyler, 2003) and maintains three subscales assessing the client/therapist agreement regarding the goals of therapy, the tasks of therapy, and the client/therapist bond. In the current sample these three subscales were highly correlated and as such we used summary statistics related to the session by session WAI-S total score in our analyses. The WAI-S has satisfactory reliability and validity (Tracey & Kokotovic, 1989).
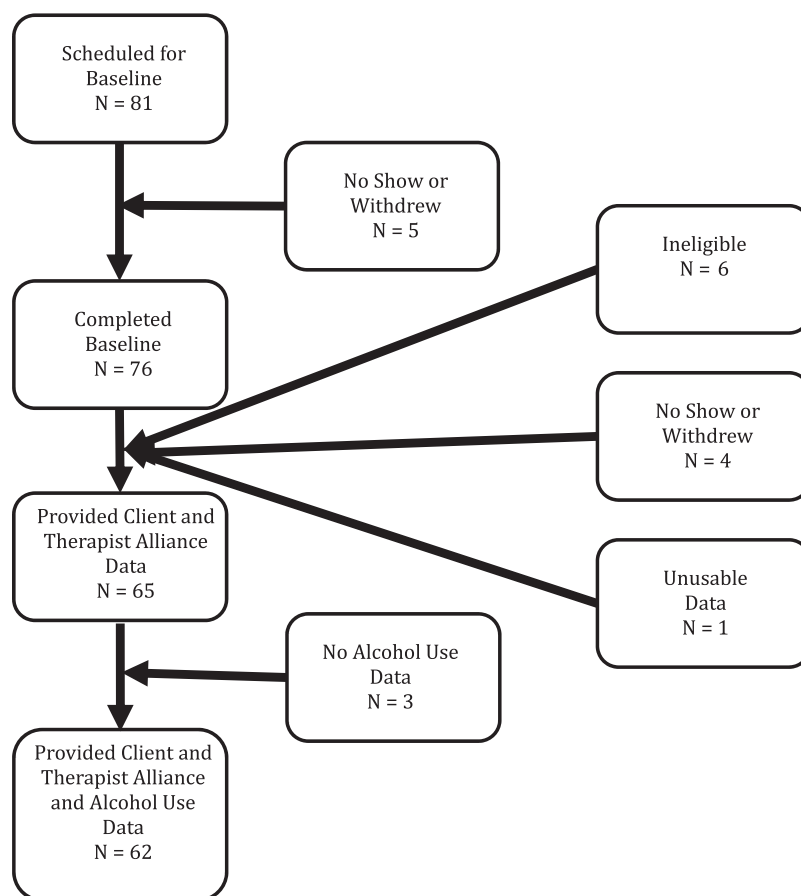
Figure 1.   Participant recruitment flow chart.

Procedure

Participants called the project phone number in response to radio and newspaper advertisements and were screened for preliminary inclusion criteria (i.e., age, commuting distance, past year treatment) as well as provided with a description of the treatment program. Potentially eligible participants were invited to participate in a 90 min baseline/intake interview. After verification of study eligibility based on a current diagnosis of alcohol use disorder (Mini International Neuropsychiatric Interview; Sheehan et al., 1997, 1998) and MMSE performance, informed consent, and baseline survey measures were administered during the baseline appointment. The University at Buffalo Social and Behavioral Sciences Institutional Review Board approved the current research project.

The research site's outpatient clinical research center served as the study treatment facility and provided all participants with 12 weeks of standard cognitive behavioral therapy (CBT; Kadden et al., 1992) for alcohol use disorder. Each session was 60-min long with the exception of the first session, which was 90 min. Sessions were conducted with one of three clinic therapists (all female) who had at least 5 years of experience treating clients with alcohol use disorder. Therapists were blind to the study hypotheses. Therapists were trained on the study protocol and supervised weekly by the fourth author. Supervision included periodic review of audiotaped

sessions to assess compliance with the treatment protocol. Therapists and participants each provided WAI-S ratings at the end of each treatment session based on their perceptions of that day's session.

The CBT protocol consisted of 7 core sessions that included (a) introduction to coping skills training, (b) coping with craving and urges to drink, (c) managing thoughts about alcohol and drinking, (d) problem solving, (e) drink refusal skills, (f) planning for emergencies and coping with a lapse, and (g) seemingly irrelevant decisions. Four of the remaining sessions were tailored to meet the client's clinical needs through collaboration between the therapist and the client. Some session topics included: starting conversations, assertiveness, anger management, managing negative thinking, and enhancing social support networks. The final session focused on reviewing treatment gains and termination of treatment.

If a client missed a treatment session or cancelled an appointment he or she was contacted via telephone or mail to reschedule. Study therapists made weekly outreach attempts to reengage the client in treatment. If a participant had 4 consecutive weeks of noncontact with the clinic, then his or her chart was placed on inactive status.

Analysis Plan

The first goal of the analyses was to discern distinct groups of individuals with similar client and therapist therapeutic alliance

(TA) rating characteristics. Alliance profiles were derived using LPA from five core client and therapist TA rating characteristics, that is, mean TA rating, minimum TA rating, maximum TA rating, range of TA ratings (maximum–minimum), and latency between maximum and minimum (session number for maximum rating–session number for minimum rating). Pretreatment and during treatment percent days abstinent were controlled.

For each set of TA ratings (i.e., client and therapist), one-through four-class models were run with variance parameters constrained to be equal across classes and then run with variance parameters freely estimated across classes. Thus, four series of LPAs were run that included (a) client TA rating variances constrained, (b) client TA rating variances freely estimated, (c) therapist TA rating variances constrained, and (d) therapist TA rating variances freely estimated. In this case the models with constrained variances are nested within the models with freely estimated variances and models with different numbers of classes are considered nested as well (Nylund et al., 2007). This nesting allows for overall model fit comparisons to be conducted within each of the four series listed earlier (i.e., comparing one-, two-, three-, and four-class models to one another within each series) and within client and therapist series across constraint types (e.g., client TA constrained models can be compared to client TA freely estimated models). Overall model fit of the client and therapist ratings is not directly comparable, because the models are not nested; however, the substantive interpretability of the best fitting model for each set will be discussed.

Analyses were conducted using Mplus version 7.2 (Muthén & Muthén, 1998–2012). Missing data were handled using the maximum likelihood method, which is a best practice strategy for managing missing data (Schafer & Graham, 2002). We chose model fit statistics based on recommendations from a Monte Carlo study that determined the most appropriate fit indices for LPA (Nylund et al., 2007) and four recommended criteria (Muthén & Muthén, 2000).

The first criterion was the Lo-Mendell-Rubin likelihood ratio test of model fit (LMR; Lo, Mendell, & Rubin, 2001). The LMR compares a model with k classes to a model with k-1 classes. The LMR statistically tests the probability that the data have been generated by the model with k-1 classes (i.e., a significant p value indicates that the k-class model is an improvement over the k-1-class model).

The second criterion was the sample size adjusted Bayesian information criterion (saBIC; Sclove, 1987). The saBIC maximizes the likelihood ratio statistic while rewarding parsimony. Lower values indicate better model fit, and the model with the lowest saBIC is generally preferred (Muthén & Muthén, 2000). Third, entropy values provided an index of model classification quality. Values range from 0 to 1; higher values indicate better classification quality (Celeux & Soromenho, 1996). Values greater than 0.80 are generally considered to have adequate classification quality (Jung & Wickrama, 2008). The fourth criterion was the average latent class probabilities for the most likely latent class membership by latent class discrimination. Values close to 1 in the primary diagonal and values close to 0 in off-diagonal represent good fit. Values close to 0.50 indicate that individuals in a particular group would fit equally well in another group. These values provide an index of how likely the individuals within a latent class

belong in that class. The usefulness of the LPA classes to differentiate participants on variables of interest was also considered. We were interested in a model that could be used to differentiate profiles of client and therapist TA rating characteristics during treatment that were associated with subsequent differences in follow-up alcohol use frequency. Final model selection was based on goodness of model fit indices, parsimony, and substantive interpretability of the model.

The second goal of the analyses was to determine whether the latent profiles identified were associated with differences in PDA during the follow-up period. Wald tests of mean equality determined whether classes had different levels of PDA in the follow-up period as well as whether classes differed on number of sessions attended (Asparouhov, 2007). Wald tests use chi-square to compare latent groups with a posterior probability-based multiple imputation strategy (Clark & Muthén, 2009). These analyses are conducted simultaneously with LPAs and allow consideration of the probabilistic class membership of participants to control error. This method of comparing latent classes on distal auxiliary variables has been referred to as "the pseudo class method" and has been compared to other multistep approaches to comparing latent classes on distal auxiliary variables (for a detailed overview of these approaches, see Asparouhov & Muthén, 2014). Simulation studies have shown that the pseudo class method works well when entropy is high (Clark & Muthén, 2009). Equality tests of means across classes have 2 degrees of freedom for each overall test and 1 degree of freedom for each pairwise test. To provide an index of effect size for the Wald tests we calculated Cramer's V, which is well established for tests of independence and has cutoffs for small, medium, and large effects based on the degrees of freedom for a chi-squared test. Cohen (1988) recommended cutoffs of .07 for small effects, .21 for medium effects, and .35 for large effects, when the chi-squared test has 2 degrees of freedom.

Results

Bivariate Correlations

Table 1 provides all bivariate correlations among client and therapist WAI-S total score ratings across all 12 sessions. As would be expected, alliance ratings that were near in time were highly correlated (e.g., Session 3 with Session 4) for both client and therapist alliance ratings, and alliance ratings that were separated in time (e.g., Session 2 with Session 10) were less strongly associated. Similarly, the associations between clients' and therapists' ratings of the same session (e.g., client rating of Session 1 and therapist rating of Session 1) were moderately correlated, and in general associations between clients' and therapists' ratings of sessions more distant in time (e.g., client rating of Session 1 with therapist rating of Session 10) were uncorrelated.

Overall Model Fit

In all cases models with freely estimated variances provided comparatively better model fit than the models with constrained variances. In both the client and therapist TA rating profiles the three-class solutions provided the best overall model fit to the data (see Tables 1 and 2). Specifically, in the client models the three-class solution had the lowest saBIC, the highest entropy (not

Table 1
Bivariate Correlations of TA and CA Total Scores

	CA total score correlations										
Week	1	2	3	4	5	6	7	8	9	10	11
2	.76**										
3	.59**	.76**									
4	.64**	.82**	.80**								
5	.66**	.79**	.73**	.91**							
6	.56**	.74**	.72**	.91**	.92**						
7	.47**	.64**	.61**	.82**	.87**	.90**					
8	.62**	.62**	.64**	.78**	.89**	.88**	.90**				
9	.64**	.60**	.59**	.77**	.89**	.87**	.92**	.92**			
10	.59**	.50**	.69**	.66**	.73**	.73**	.79**	.83**	.85**		
11	.49**	.37**	.49**	.70**	.78**	.79**	.86**	.91**	.87**	.89**	
12	.44*	.37*	.52**	.71**	.66**	.64**	.75**	.76**	.77**	.84**	.88**
	TA total score correlations										
2	.72**										
3	.57**	.75**									
4	.58**	.69**	.88**								
5	.52**	.66**	.78**	.86**							
6	.57**	.74**	.76**	.80**	.87**						
7	.44**	.62**	.65**	.70**	.74**	.89**					
8	.46**	.67**	.65**	.68**	.80**	.91**	.84**				
9	.38**	.56**	.98**	.73**	.71**	.72**	.64**	.70**			
10	.22	.43**	.47**	.49**	.48**	.53**	.47**	.57**	.85**		
11	.11	.17	.31	.41**	.46**	.41*	.46**	.49**	.80**	.83**	
12	.26	.27	.32	.46**	.50**	.53**	.63**	.57**	.79**	.76**	.87**

	CA with TA total score correlations											
	TA1	TA2	TA3	TA4	TA5	TA6	TA7	TA8	TA9	TA10	TA11	TA12
CA1	.38**	.31*	.14	.20	.28*	.22	.13	.36*	.19	.12	.13	.14
CA2	.48**	.48**	.36**	.44**	.52**	.48**	.41**	.49**	.24	.09	.07	.16
CA3	.24	.22	.23	.29**	.29*	.25	.24	.27	.14	.09	.06	.03
CA4	.33*	.29*	.34**	.43**	.43**	.32*	.35*	.37*	.24	.18	.12	.05
CA5	.31*	.31*	.39**	.37**	.45**	.38**	.40**	.45**	.33*	.31*	.28	.21
CA6	.22	.36**	.44**	.46**	.52**	.49**	.51**	.53**	.41**	.37*	.28	.25
CA7	.12	.28	.38**	.32*	.38**	.41**	.55**	.47**	.37*	.41**	.40*	.39*
CA8	.18	.24	.27	.28	.32*	.30*	.42**	.36*	.28	.35*	.33*	.27
CA9	.14	.23	.27	.25	.35*	.32*	.43**	.40**	.34*	.43**	.47**	.39*
CA10	.04	.19	.21	.17	.18	.23	.25	.26	.31*	.39*	.40*	.30
CA11	−.08	.06	.07	.06	.07	.11	.19	.19	.21	.40*	.44**	.33
CA12	−.17	−.11	−.07	−.23	−.22	−.18	−.10	−.17	−.08	.17	.26	.18

Note. CA = Working Alliance Inventory (WAI-S) client alliance total score ratings; TA = WAI-S therapist alliance total score ratings.
* $p < .05$.   ** $p < .01$.

counting the one-class model which will always have an entropy of 1), and the highest average latent class probabilities. Further, examination of the LMR tests for the client models indicated that the two-class model was an improvement over the one-class model, the three-class model was an improvement over the two-class model, but the four-class model was not an improvement over the three-class model. In addition, the class breakdown of the three-class model was reasonable, with 44% of the sample most likely in the largest class, 34% in the middle class, and 23% in the smallest class.

In the therapist models, the three-class solution had lower saBIC than the one- and two-class models but slightly higher saBIC than the four-class model. The three-class therapist model had comparable entropy and average latent class probabilities to the two- and four-class models. Despite these similarities, the LMR tests re-

vealed that the two-class therapist model was an improvement over the one-class model and the three-class model was an improvement over the two-class model, but the four-class model was not an improvement over the three-class model. Finally, the three-class therapist model demonstrated reasonable class breakdown with 39% of the sample most likely in the largest class, 32% in the middle class, and 29% in the smallest class.

Description of the Best Fitting Models

Tables 3 and 4 present the means and standard deviations estimated for each latent class on each of the TA rating characteristics included in deriving the latent profiles. These tables also present mean pretreatment, during treatment, and posttreatment percent days abstinent values for each latent class; however, it is

Table 2

Comparative Latent Profile Analysis Overall Model Fit Statistics for Client Reports on the Therapeutic Alliance

	Overall model fit			
	1-class	2-class	3-class	4-class
saBIC	1048.07	795.55	404.27	665.85
Entropy	1.00	.936	.971	.967
ALC-Prob.	—	.958–.987	.98–1.0	.92–1.0
% sample/class				
1	65 (100%)	22 (35%)	21 (34%)	9 (15%)
2		40 (65%)	27 (44%)	13 (21%)
3			14 (23%)	26 (42%)
4				14 (23%)
LMR (P value)	—	196.29 (.00)	102.12 (.00)	50.23 (.54)

Note. saBIC = sample size adjusted Bayesian information criterion; ALC-Prob. = average latent class probability for most likely latent class membership; LRM = Lo-Mendell-Rubin adjusted likelihood ratio test. Wald tests of equality were used to compare profiles in models with more than one class on follow-up percent days abstinent. Thus, in the 1-class model the total N is 65, as there were no classes to compare; the total N is 62 for the two- through four-class models due to the loss of three participants who failed to provide follow-up data.

important to note that pretreatment and during treatment PDA were controlled for by specifying those variables as predicting the categorical latent class variable and follow-up PDA was estimated as a distal auxiliary variable. Thus, PDA values were not used to derive the alliance profiles.

Client three-class model. The three classes of the best fitting client model can be described as "high client alliance" (high-CA), "medium client alliance" (medium-CA), and "low client alliance" (low-CA). (See Table 5.) The high-CA class was associated with the highest average client TA rating, highest minimum client TA rating, and highest maximum client TA rating. Further, the high-CA class was associated with the smallest range of client ratings and the middle score for min-max latency. Note that the standard deviation of the min-max latency score in the high-CA is nearly equal to the mean score. Because this score was calculated as max–min session number a positive value indicates that the minimum rating occurs before the maximum rating in the course of treatment. Given the magnitude of the standard deviation relative to the magnitude of the mean for this variable one can infer that reporting decreasing alliance throughout treatment was rarely associated with the high-CA class. The high-CA class was associated with the highest PDA rating on all indices.

The medium-CA class was associated with mean values lower than the high-CA class and higher than the low-CA class on all indicators except the min-max latency variable. On this characteristic the medium-CA class was associated with the smallest latency, but the largest standard deviation. This suggests that, although the medium-CA class was associated with relatively strong alliance ratings on other indices, it was not uncommon for the minimum rating to occur later in treatment compared to the maximum rating, suggesting some deterioration in alliance associated with this profile. The medium-CA class was also associated with the middle PDA rating on all indices.

The low-CA class was associated with the lowest mean ratings of average client TA rating, minimum client TA rating, and maximum client TA rating. The low-CA class was also associated with

the lowest PDA ratings on all indices. The low-CA class was associated with the highest ratings on client TA range and min–max latency, suggesting the most variation in client alliance ratings was associated with this profile.

Client Wald tests of equality. The omnibus test was significant for client follow-up PDA ratings, and medium to large in effect size, $\chi^2 = 10.56$, $p < .01$, Cramer's V = .29. Pairwise comparisons revealed that the high-CA group had significantly more PDA at follow-up compared to the low-CA group in the medium to large effect size range, $\chi^2 = 10.14$, $p = .001$, Cramer's V = .29. The pairwise comparisons between low-CA and medium-CA and medium-CA and high-CA revealed trends toward significance such that medium-CA had more PDA compared to low-CA and less PDA than high-CA (low-CA vs. medium-CA: $\chi^2 = 3.23$, $p = .07$, Cramer's V = .16; medium-CA vs. high-CA: $\chi^2 = 3.41$, $p = .07$, Cramer's V = .17). The effect size estimates for these comparisons were in the small to medium range.

The omnibus test was significant for session attendance among the client rating profiles, and medium to large in effect size, $\chi^2 = 9.56$, $p < .01$, Cramer's V = .28. Pairwise comparisons revealed that the high-CA group attended significantly more sessions, on average, compared to both the medium-CA and low-CA groups (high-CA vs. medium-CA: $\chi^2 = 5.16$, $p = .02$, Cramer's V = .20; high-CA vs. low-CA: $\chi^2 = 7.92$, $p < .01$, Cramer's V = .25). Effects were in the small to medium and medium to large effect size range, respectively. The pairwise comparisons between low-CA and medium-CA groups was not significant for session attendance, $\chi^2 = .61$, $p = .44$, Cramer's V = .07, and the effect size was in the small range.

Therapist 3-class model. The three classes of the best fitting therapist model can be described as "high therapist alliance" (high-TA), "medium therapist alliance" (medium-TA), and "low therapist alliance" (low-TA). The high-TA class was associated with the highest mean therapist alliance rating, highest minimum alliance rating, highest maximum alliance rating, and highest min-max latency in therapist alliance ratings. The high-TA class was

Table 3

Comparative Latent Profile Analysis Overall Model Fit Statistics for Therapist Reports on the Therapeutic Alliance

	Overall model fit			
	1-class	2-class	3-class	4-class
saBIC	968.079	774.26	698.48	636.45
Entropy	1.00	.95	.95	.97
ALC-Prob. range	—	.98–1.0	.97–1.0	.97–1.0
% sample/class				
1	65 (100%)	25 (40%)	18 (29%)	19 (31%)
2		37 (60%)	24 (39%)	20 (32%)
3			20 (32%)	18 (29%)
4				5 (8%)
LMR (P value)	—	139.36 (.00)	86.91 (.03)	78.85 (.91)

Note. saBIC = Sample size adjusted Bayesian information criterion; ALC-Prob. = average latent class probability for most likely latent class membership; LRM = Lo-Mendell-Rubin adjusted likelihood ratio test. Wald tests of equality were used to compare profiles in models with more than one class on follow-up percent days abstinent. Thus, in the one-class model the total N is 65, as there were no classes to compare; the total N is 62 for the two- through four-class models due to the loss of three participants who failed to provide follow-up data.

Table 4
Descriptive Information for 3-Profile Model of Client Ratings of Therapeutic Alliance

	Low-CA, $n = 21$ (34%) M (SD)	Medium-CA, $n = 27$ (44%) M (SD)	High-CA, $n = 14$ (22%) M (SD)
Variables defining latent profiles			
Client alliance mean	4.98 (.65)	6.18 (.24)	6.77 (.09)
Client alliance min	4.05 (.60)	5.63 (.40)	6.13 (.21)
Client alliance max	5.58 (.81)	6.34 (.29)	6.98 (.03)
Client alliance difference (range)	1.53 (1.04)	1.01 (.57)	.85 (.21)
Client session number difference (min-max latency)	3.27 (4.73)	2.90 (5.03)	2.63 (2.98)
Control variables			
PDA Pre-Tx	.30 (.25)	.38 (.27)	.43 (.24)
PDA Tx	.66 (.10)	.80 (.24)	.89 (.16)
Auxiliary outcome variable	M (SE)	M (SE)	M (SE)
PDA 4m FU	.59 (.09)	.77 (.05)	.91 (.05)
Session attendance	8.23 (.87)	9.11 (.70)	11.14 (.56)

Note. low-CA = weak client alliance profile; medium-CA = moderate client alliance profile; high-CA = strong client alliance profile; PDA = percent days abstinent; Pre-Tx = pretreatment; Tx = treatment; 4m FU = 4-month follow-up.

also associated with the smallest range of therapist alliance ratings. Further, the high-TA class was associated with the highest PDA ratings at pretreatment and the middle PDA ratings both during and posttreatment.

The medium-TA class was associated with alliance levels that were lower than high-TA and higher than low-TA for average therapist alliance rating, minimum therapist alliance rating, maximum alliance rating, and range. However, the medium-TA class was associated with the lowest min-max latency value and had the largest standard deviation of the min–max latency. This suggests that therapists' alliance ratings in this class were most likely to include negative values (i.e., most likely to have provided their minimum rating in a session subsequent to the session in which they provided their maxi-

mum rating). It should be noted that the standard deviation relative to the mean was high in all three therapist alliance classes, indicating that therapists reported declining alliance ratings more overall compared to clients. Further, although the medium-TA class was associated with the highest PDA in the during- and posttreatment periods it was associated with the middle PDA rating in the pretreatment period.

The low-TA class was associated with the lowest mean therapist alliance rating, lowest minimum therapist alliance rating, and lowest maximum therapist alliance rating. Further, the low-TA class was associated with the largest therapist alliance range and the middle value for min-max latency. The low-TA class was associated with the lowest PDA at pretreatment, during treatment, and posttreatment.

Table 5
Descriptive Information for Three-Profile Model of Therapist Ratings of Therapeutic Alliance

	Low-TA, $n = 18$ (29%) M (SD)	Medium-TA, $n = 20$ (32%) M (SD)	High-TA, $n = 24$ (39%) M (SD)
Variables defining latent profiles			
Therapist alliance mean	4.82 (.50)	5.65 (.26)	6.11 (.17)
Therapist alliance min	3.94 (.86)	4.71 (.26)	5.70 (.25)
Therapist alliance max	5.56 (.57)	6.19 (.24)	6.36 (.25)
Therapist alliance difference (range)	1.62 (1.05)	1.48 (.26)	.66 (.34)
Therapist session number difference (min-max latency)		2.12 (5.98)	2.74 (5.72)
Control variables			
PDA Pre-Tx	.26 (.23)	.30 (.24)	.49 (.25)
PDA Tx	.64 (.11)	.84 (.22)	.81 (.18)
Auxiliary outcome variable	M (SE)	M (SE)	M (SE)
PDA 4m FU	.61 (.10)	.82 (.06)	.78 (.07)
Session attendance	7.24 (1.05)	10.51 (.62)	9.71 (.64)

Note. low-TA = weak therapist alliance profile; medium-TA = moderate therapist alliance profile; high-TA = strong therapist alliance profile; PDA = percent days abstinent; Pre-Tx = pretreatment; Tx = treatment; 4m FU = 4-month follow-up.

Therapist Wald tests of equality. Neither the omnibus test nor any of the pairwise tests of mean equality were significant (omnibus: $\chi^2 = 3.00$, $p = .22$, Cramer's V = .16; low-TA vs. high-TA: $\chi^2 = 2.20$, $p = .14$, Cramer's V = .13; low-TA vs. medium-TA: $\chi^2 = 3.21$, $p = .07$, Cramer's V = .16; medium-TA vs. high-TA: $\chi^2 = .24$, $p = .63$, Cramer's V = .04). However, the medium-TA class showed trend level more PDA compared to the low-TA class. The effect size estimates revealed that the omnibus test and the tests comparing low-TA versus high-TA, low-TA versus medium-TA, and medium-TA versus high-TA were in the small to medium range. The final comparison between medium-TA versus high-TA had an effect size that did not meet the minimum cutoff to be considered a small effect.

The omnibus test was not significant for session attendance among the therapist rating profiles, and was in the small to medium effect size range, $\chi^2 = 5.17$, $p = .08$, Cramer's V = .20. Pairwise comparisons revealed that the low-TA group attended significantly less sessions, on average, compared to both the medium-TA and high-TA groups (low-TA vs. medium-TA: $\chi^2 = 6.80$, $p < .01$, Cramer's V = .23; low-TA vs. high-TA: $\chi^2 = 4.04$, $p = .04$, Cramer's V = .18). Effects were in the medium to large and small to medium effect size range, respectively. The pairwise comparisons between medium-TA and high-TA groups was not significant for session attendance, $\chi^2 = .80$, $p = .37$, Cramer's V = .08, and the effect size was in the small range.

Discussion

This study was designed to identify latent profiles of therapeutic alliance ratings that distinguish the characteristics of therapeutic alliance across a course of treatment. We also planned to test whether different therapeutic alliance ratings profiles predicted abstinence from alcohol during the follow-up period. Our findings supported three profiles for both client and therapist ratings based on five therapeutic alliance indices: mean alliance, minimum alliance, maximum alliance, range of alliance, and the latency from the maximum to the minimum alliance rating. For both client and therapist alliance ratings, the three profiles were labeled low, medium, and high alliance profiles, respectively. The client alliance profiles predicted percent days abstinent during the follow-up period, but the therapist alliance profiles did not. This result is similar to recent findings by Cook, Heather, and McCambridge, and the United Kingdom Alcohol Treatment Trial Research Team (2015), who showed that client but not therapist alliance ratings predicted alcohol treatment outcome, were associated with drinking during treatment, and had effects on posttreatment motivation to change. In addition, our analyses revealed differences in profiles based on session attendance. Specifically, the high-CA profile was associated with greater session attendance, on average, than either the medium-CA and low-CA profiles, and the low-TA profile was associated with significantly less sessions attendance, on average, compared to the medium-TA and high-TA profiles.

Examination of Tables 3 and 4 reveal some interesting findings. First, even among the weak alliance profiles, both therapists and clients were providing mean therapeutic alliance ratings near 5 on a 7-point scale and minimum ratings near 4. In contrast, the mean ratings for the strong alliance profiles were 6.77 for client ratings and 6.11 for therapist ratings out of 7, and minimum ratings in the strong alliance profiles were 6.13 for clients' and 5.70 for thera-

pists' ratings. Despite the potential for a ceiling effect for the WAI-S, these data suggest that even small deviations from the top of the scale may be cause for concern.

Another dimension of the alliance that warrants attention is the session number difference index. This metric is an index of the number of sessions between the minimum and maximum rating, independent of the magnitude of the rating. Across the latent profiles the data showed that the typical values were positive (improvement in the alliance) for this index, but among the client alliance rating profiles there were larger standard deviations for the weak and moderate alliance profiles, indicating more negative values (decline in the alliance) associated with those profiles relative to the strong alliance profile. The variability in client perceptions of changes in their perceptions of the alliance (as indicated by a positive or negative score on the variable defined as "session number with maximum rating minus session number with minimum rating") likely helped to define client alliance rating profiles. In contrast, the therapist alliance profiles showed similar levels and standard deviations across profiles for that index, suggesting that therapists' perceptions of whether the therapeutic alliance was improving or deteriorating likely did not provide much weight in the differentiation of the profiles.

Finally, of note across both client and therapist alliance profiles are the high maximum values and small range in ratings seen in the strong alliance profiles. This finding suggests that small variations from the top of the scale are indicative of a strong alliance profile. Further, this profile was associated with the most abstinence based on client ratings, but not based on therapist ratings. Moreover, the strong client alliance rating profile was associated with 13% more days abstinent than the strong therapist alliance rating profile (i.e., client strong alliance profile 91% days abstinent vs. therapist strong alliance profile 78% days abstinent). Thus, a client consistently rating alliance at or near the top of the scale predicts a high proportion of abstinent days in the follow-up period.

Given the subtleties of the differences among the client and therapist therapeutic alliance ratings profiles, their potential clinical utility should be examined. First of all, clinicians using the WAI-S should recognize that even small deviations from the top of the scale may signal poorer treatment outcomes. Second, both clients and therapists rarely used ratings below the midpoint of the scale, suggesting that ratings near or below the midpoint in the scale may indicate that the alliance is in jeopardy, has been ruptured, or was not well established. In any case, ratings near or below the midpoint in the scale perhaps should be considered and/or discussed in subsequent therapy sessions. A third potential issue that can be inferred from the present study is that a positive bias exists for both clients and therapists when rating therapeutic alliance. To the extent that clients or therapists are not comfortable rating their true feelings regarding the therapeutic alliance, then changing the scale to a relative rating scale; for example, *"We agree on what is important for me to work on."* Response options range from 1 (*more than*), −1 (*less than*), or 0 (*the same as last session*) could also help reduce this bias.

Results from the current study are consistent with the broader literature on therapeutic alliance. We found that pattern of client therapeutic alliance ratings profiles over a course of treatment predicts abstinence in the follow-up period. These results extend findings from the Meier et al. (2005) review, which concluded that therapeutic alliance early in treatment are associated with improve-

ments early in treatment. Previous research has also demonstrated the relative benefit for client ratings over therapist ratings of alliance in predicting outcomes (Horvath & Symonds, 1991), as was the case in this study.

Our results extend the current literature. The approach used in the current study extends research examining a single time point or focusing on alliance ratings early in treatment by characterizing patterns of therapeutic alliance ratings across the full course of treatment. Our results revealed a more complex picture with three distinct profiles of alliance for both client and therapist ratings.

This study has a number of strengths. Principally, we used statistical analyses to identify latent profiles of therapeutic alliance ratings for both client and therapist ratings across a course of treatment. In addition, we were able to compare therapist and client ratings across treatment. This study also should be considered in light of its limitations. First, the sample size was small to moderate, and thus future research should replicate and extend these findings with a larger sample. A second limitation is that the data to create latent profiles are not fully available until the end of a course of treatment. Nevertheless, clinicians can use these profiles as a way to be aware of warning signs of poorer treatment outcome, as ratings are collected after each session. A third limitation is that the current study included only three therapists, precluding the possibility of examining therapist characteristics that may influence therapeutic alliance ratings. In addition, independent observer ratings of the alliance were not available for evaluation. The extent to which findings similar to those in the current study would emerge using that third rating perspective is not known.

Building on the present study, we can identify four areas for future research. First, although our results indicate that the identified profiles predict abstinence in the follow-up period, we did not assess for client factors that might predict class membership. Future studies should examine client factors (e.g., alcohol dependence severity, coping skills, alcohol craving, self-efficacy, and readiness to change) that may identify indicators of protective or risk factors that therapists can utilize to tailor treatment. Second, there are many ways to examine the longitudinal relationship between alcohol use and therapeutic alliance. Future research could examine the dynamic relationship between these constructs using difference score models, latent growth models, or cross-lagged regression models. Each of these approaches would provide unique and interesting information about the interplay between therapeutic alliance and alcohol use over time. Third, one limitation of the current study is that complete profile information is not available until the 12th session. Future research could examine the consistency of the present findings with data from fewer total sessions. Finally, we found that even small deviations from the top of the working alliance scale are indicative of worse outcomes. Future treatment studies could be designed that include rapport-building strategies based on client ratings of alliance in an effort to improve treatment engagement (e.g., session attendance) and client outcomes.

In conclusion, examining therapeutic alliance ratings across the full course of treatment revealed that three distinct latent profiles exist among both client and therapist ratings. Client but not therapist ratings profiles predicted drinking in the follow-up period. Most ratings were near the top of the scale, and even small deviations or scores near the midpoint in the scale may be an indicator of poorer treatment outcome. This study chiefly provides information about "the quality" of therapeutic alliance, defined here as a description of alliance characteristics over a course of treatment. The quality of therapeutic alliance has been identified as an important predictor of outcome (Henry, Strupp, Schacht, & Gaston, 1994), and the current findings provide a novel way to evaluate the quality of the therapeutic alliance in predicting abstinence from alcohol use in the months following treatment.

References

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Asparouhov, T. (2007). *Wald test of mean equality for potential latent class predictors in mixture modeling: Technical appendix*. Los Angeles, CA: Muthén & Muthén.

Asparouhov, T., & Muthén, B. (2014). Auxiliary variables in mixture modeling: Three-step approaches using M plus. *Structural Equation Modeling, 21,* 329–341. http://dx.doi.org/10.1080/10705511.2014.915181

Busseri, M. A., & Tyler, J. D. (2003). Interchangeability of the Working Alliance Inventory and Working Alliance Inventory, Short Form. *Psychological Assessment, 15,* 193–197. http://dx.doi.org/10.1037/1040-3590.15.2.193

Castonguay, L. G., Constantino, M. J., & Holtforth, M. G. (2006). The working alliance: Where are we and where should we go? *Psychotherapy: Theory, Research, & Practice, 43,* 271–279. http://dx.doi.org/10.1037/0033-3204.43.3.271

Celeux, G., & Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification, 13,* 195–212. http://dx.doi.org/10.1007/BF01246098

Clark, S., & Muthén, B. (2009). *Relating latent class analysis results to variables not included in the analysis.* https://www.statmodel.com/download/relatinglca.pdf

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Connors, G. J., Carroll, K. M., DiClemente, C. C., Longabaugh, R., & Donovan, D. M. (1997). The therapeutic alliance and its relationship to alcoholism treatment participation and outcome. *Journal of Consulting and Clinical Psychology, 65,* 588–598. http://dx.doi.org/10.1037/0022-006X.65.4.588

Cook, S., Heather, N., & McCambridge, J., & the United Kingdom Alcohol Treatment Trial Research Team. (2015). The role of the working alliance in treatment for alcohol problems. *Psychology of Addictive Behaviors, 29,* 371–381. http://dx.doi.org/10.1037/adb0000058

Ehrman, R. N., & Robbins, S. J. (1994). Reliability and validity of 6-month timeline reports of cocaine and heroin use in a methadone population. *Journal of Consulting and Clinical Psychology, 62,* 843–850. http://dx.doi.org/10.1037/0022-006X.62.4.843

Fenton, L. R., Cecero, J. J., Nich, C., Frankforter, T. L., & Carroll, K. M. (2001). Perspective is everything: The predictive validity of six working alliance instruments. *Journal of Psychotherapy Practice and Research, 10,* 262–268.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12,* 189–198. http://dx.doi.org/10.1016/0022-3956(75)90026-6

Hartzler, B., Witkiewitz, K., Villarroel, N., & Donovan, D. (2011). Self-efficacy change as a mediator of associations between therapeutic bond and one-year outcomes in treatments for alcohol dependence. *Psychology of Addictive Behaviors, 25,* 269–278. http://dx.doi.org/10.1037/a0022869

Henry, W. P., Strupp, H. H., Schacht, T. E., & Gaston, L. (1994). Psy-

chodynamic approaches. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (4th ed., pp. 467–508). Oxford, UK: 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, NY: Guilford Press.

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. http://dx.doi.org/10.1037/0022-0167.38.2.139

Ilgen, M., Tiet, Q., Finney, J., & Moos, R. H. (2006). Self-efficacy, therapeutic alliance, and alcohol-use disorder treatment outcomes. *Journal of Studies on Alcohol, 67,* 465–472. http://dx.doi.org/10.15288/jsa.2006.67.465

Jung, T., & Wickrama, K. A. S. (2008). An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass, 2,* 302–317. http://dx.doi.org/10.1111/j.1751-9004.2007.00054.x

Kadden, R., Carroll, K., Donovan, D., Monti, P., Abrams, D., Litt, M., & Hester, R. (1992). *Cognitive-behavioral coping skills therapy manual*. Rockville, MD: NIAAA.

Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika, 88,* 767–778. http://dx.doi.org/10.1093/biomet/88.3.767

Maisto, S. A., Roos, C. R., O'Sickey, A. J., Kirouac, M., Connors, G. J., Tonigan, J. S., & Witkiewitz, K. (2015). The indirect effect of the therapeutic alliance and alcohol abstinence self-efficacy on alcohol use and alcohol-related problems in Project MATCH. *Alcoholism: Clinical and Experimental Research, 39,* 504–513. http://dx.doi.org/10.1111/acer.12649

Meier, P. S., Barrowclough, C., & Donmall, M. C. (2005). The role of the therapeutic alliance in the treatment of substance misuse: A critical review of the literature. *Addiction, 100,* 304–316. http://dx.doi.org/10.1111/j.1360-0443.2004.00935.x

Muthén, B., & Muthén, L. K. (2000). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research, 24,* 882–891. http://dx.doi.org/10.1111/j.1530-0277.2000.tb02070.x

Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide* (Version 7). Los Angeles, CA: Author.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of latent classes in latent class analysis and growth mixture modeling: A monte carlo simulation study. *Structural Equation Modeling, 14,* 535–569. http://dx.doi.org/10.1080/10705510701575396

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7,* 147–177.

Sclove, L. S. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika, 52,* 333–343. http://dx.doi.org/10.1007/BF02294360

Sheehan, D. V., Lecrubier, Y., Harnett-Sheehan, K., Janavs, J., Weiller, E., Bonara, L. I., . . . Dunbar, G. C. (1997). The validity of the Mini International Neuropsychiatric Interview (MINI) according to the SCID-P and its reliability. *European Psychiatry, 12,* 232–241. http://dx.doi.org/10.1016/S0924-9338(97)83297-X

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., . . . Dunbar, G. C. (1998). The Mini-International Neuropsychiatric Interview (M. I. N. I.): The development and validation of a structured diagnostic psychiatric interview for *DSM–IV* and ICD-10. *The Journal of Clinical Psychiatry, 59*(Suppl 20), 22–33.

Sobell, L. C., & Sobell, M. B. (1992). Timeline follow-back: A technique for assessing self-reported ethanol consumption. In J. Allen & R. Litten (Eds.), *Techniques to assess alcohol consumption* (pp. 41–72). Totowa, NJ: Humana Press. http://dx.doi.org/10.1007/978-1-4612-0357-5_3

Sobell, L. C., & Sobell, M. B. (1996). *Alcohol timeline followback (TLFB) user's manual*. Toronto: Addiction Research Foundation.

Tracey, T. J., & Kokotovic, A. M. (1989). Factor structure of the Working Alliance Inventory. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 1,* 207–210. http://dx.doi.org/10.1037/1040-3590.1.3.207