

Routine monitoring of therapeutic alliance to predict treatment engagement in a Veteran Affairs
substance use disorders clinic

Simon B. Goldberg

University of Wisconsin – Madison

Gail Rowe, Carol A. Malte, Hang Ruan

VA Puget Sound Health Care System – Seattle Division

Jesse J. Owen

University of Denver

Scott D. Miller

International Center for Clinical Excellence

Simon B. Goldberg, Department of Counseling Psychology, University of Wisconsin – Madison; Gail Rowe and Hang Ruan, Addiction Treatment Center, VA Puget Sound Health Care System – Seattle Division, Seattle, WA; Carol A. Malte, Center for Excellence in Substance Abuse Treatment and Education, VA Puget Sound Health Care System – Seattle Division, Seattle, WA; Jesse J. Owen, Department of Counseling Psychology, University of Denver; Scott D. Miller, International Center for Clinical Excellence, Chicago, IL.

Correspondence concerning this article should be addressed to Simon B. Goldberg, Department of Counseling Psychology, University of Wisconsin – Madison, Madison, WI, 53706. Email: sbgoldberg@wisc.edu

Acknowledgments: The views expressed in this article are solely those of the authors and do not reflect an endorsement or the official policy or position of the Department of Veterans Affairs. Dr. Miller owns the copyright of the Group Session Rating Scale.

Recommended citation: Goldberg, S. B., Rowe, G., Malte, C. A., Ruan, H., Owen, J. J., & Miller, S. D. (in press). Routine monitoring of therapeutic alliance to predict treatment engagement in a Veteran Affairs substance use disorders clinic. *Psychological Services*.

Abstract

Measurement-based care (MBC) can improve mental health treatment outcomes and is a priority within the Department of Veterans Affairs (VA). However, to date MBC efforts within the VA have focused on assessment of psychological symptoms, to the exclusion of psychotherapy process variables such as the therapeutic alliance that may predict treatment response. This quality improvement project involved the implementation of routine monitoring of alliance within a VA substance use disorder (SUD) clinic predominantly serving veterans with serious mental illness. Alliance ratings were provided by $n = 98$ veterans following group therapy sessions. Low alliance ratings were used by the clinicians ($n = 4$) leading the groups ($n = 9$) as opportunities to discuss veterans' treatment experience and increase engagement. Using multilevel models that accounted for the nested nature of the data and veteran demographics, alliance ratings showed a small increase over time ($B = 0.075, p < .001, f^2 = 0.033$). In addition, maximum alliance rating (i.e., patients' highest rating of alliance across all observations) was significantly but modestly associated with attendance at both MBC group sessions and all SUD-related visits in the three months following the initial alliance rating ($Bs = 0.96$ and $1.79; ps = .006$ and $.004; f^2s = 0.079$ and 0.088 , respectively). Average alliance rating, however, was not associated with treatment attendance ($ps > .050$). Findings suggest that assessment of alliance is feasible within a VA SUD clinic and may provide information signaling risk for disengagement that could be used for increasing treatment engagement.

Keywords: measurement-based care, therapeutic alliance, veterans, substance use disorders, serious mental illness

Routine monitoring of therapeutic alliance to predict treatment engagement in a Veteran Affairs
substance use disorders clinic

Measurement-based care (MBC) involves the systematic administration of rating scales in the context of health care and the use of results to inform clinical decision-making (Fortney et al., 2017; Goldberg, Buck, Raphaely, & Fortney, 2018). Various terms have been used to refer to MBC in the context of mental health treatment, including routine outcome monitoring, feedback-informed treatment, and the use of practice-based evidence (Prescott, Maeschalck, & Miller, 2017; Wampold, 2015). Interest in MBC in mental health treatment has increased in recent decades, in part due to a growing body of empirical research indicating MBC can improve treatment outcomes. Meta-analyses of randomized clinical trials (RCTs) testing MBC within heterogeneous psychiatric samples have demonstrated that timely feedback of patients' progress to providers can make treatment more efficacious (Knaup, Koesters, Schoefer, Becker, & Puschner, 2009; Shimokawa, Lambert, & Smart, 2010). MBC paired with feedback-informed supervision (Miller & Bertolino, 2012) has also been associated with improved therapist performance within an outpatient community mental health agency (Goldberg et al., 2016). Efforts to integrate MBC into routine practice have begun within several large-scale health care systems, including the Department of Veterans Affairs (VA; Fortney et al., 2017), the United Kingdom's National Health Service (Clark et al., 2018), and Group Health / Kaiser Permanente (Steinfeld, Franklin, Mercer, Fraynt, & Simon, 2016).

MBC is typically focused on the assessment of patients' symptoms (Harding et al., 2011). There are several advantages of this emphasis, including the ability to detect deterioration and stagnation, monitoring clinic performance on a highly relevant metric (i.e., symptoms), and demonstrating effectiveness to payers (Fortney et al., 2017). While symptom monitoring is

crucial, it is not the only source of information on treatment progress. Decades of research on psychosocial interventions have highlighted the relevance of studying not only the outcome but also the processes by which change occurs within treatments (Lambert, 2013) and numerous treatment processes have been shown to predict outcome (Norcross & Lambert, in press).

The assessment of treatment process may be particularly relevant in the context of long-term treatment for chronic mental health conditions, including substance use disorders (SUD) and serious mental illness (SMI). In contrast to estimated trajectories for mood and anxiety symptoms, which are expected to diminish over the course of therapy (Cuijpers, van Straten, Andersson, & van Oppen, 2008; Westen & Morrison, 2001), symptoms for SUD and SMI may not follow this pattern (Hayes, Laurenceau, Feldman, Strauss, & Cardaciotto, 2007; Warner, 2009). Patterns of substance use have been shown to include sudden and discontinuous change that is better predicted by nonlinear than linear change models (e.g., catastrophe modeling based on dynamical systems theory; Witkiewitz & Marlatt, 2007). Symptoms of SMIs with relatively modest full remission rates (e.g., schizophrenia; Warner, 2009) may likewise not be expected to diminish in a linear fashion over the course of treatment. Therefore, process variables may be more sensitive to the progress of therapy and factors that predict key outcomes for chronic mental illness, such as treatment retention.

Of the various psychotherapy processes that have been linked to treatment outcome (e.g., therapist empathy, patient emotional experiencing; Norcross & Lambert, in press), therapeutic alliance has been the most robust (Flückiger, Del Re, Wampold, & Horvath, 2018). Therapeutic alliance has been defined as the personal bond between therapist and patient coupled with agreement on the tasks and goals of therapy (Horvath et al., 2011). In a recent meta-analysis of $k = 295$ studies representing over 30,000 patients, alliance was moderately linked with treatment

outcome ($r = .28$; Flückiger et al., 2018), with effects largely consistent across various psychological treatments (e.g., cognitive behavioral therapy, psychodynamic psychotherapy) and alliance raters (i.e., patient-, observer-, or therapist-rated). Some variation in the strength of the alliance-outcome association was found across diagnoses, with effects smaller for samples with substance use disorders ($r = .14$) and when treatment drop-out was used as the outcome (Flückiger et al., 2018). Alliance has been positively linked with treatment outcomes in group therapy contexts as well (e.g., Joyce, Piper, & Ogrodniczuk, 2007; Tasca, Compare, Zarbo, & Brugnera, 2016). In the context of group treatment, alliance has been shown to correlate highly with other key therapeutic relationship factors including group cohesion (Johnson, Burlingame, Olsen, Davies, & Gleave, 2005), which itself has been robustly linked with treatment outcome in meta-analysis ($r = .26$, $k = 55$ studies; Burlingame, McClendon, & Yang, 2018).

Due to the link between alliance and outcome, some MBC systems¹ have begun incorporating assessment of alliance (e.g., Partners for Change Outcome Management System [PCOMS], Miller, Hubble, Chow, & Seidel, 2015; Outcome Questionnaire – Analyst, Lambert, 2015). However, to our knowledge, MBC efforts within the VA have thus far focused solely on the assessment of psychiatric symptoms (Oslin et al., 2006; Pomerantz, Kearney, Wray, Post, & McCarthy, 2014).

In addition to assessing alliance, MBC systems may be helpful in providing clinicians with a “signal” of difficulties in alliance formation. The use of symptom-focused treatment signals has been widely implemented in the MBC literature, and it appears that providing

¹ An MBC system can be defined as one or more measures (e.g., symptom measures) paired with software capable of comparing patients’ trajectories of change with normative data for patient progress (Wampold, 2015). Typically, measures are completed at each clinical encounter and the MBC system may or may not provide additional computer-generated feedback (e.g., indication of suicide risk, recommendations for ways to modify treatment, suggestions for additional dimensions to assess).

clinicians with a signal that treatment is not progressing as expected is part of what MBC effective (Delgadillo et al., 2018; Lambert, Hansen, & Finch, 2001; Lambert, 2015). The pattern of scores that might constitute a signal of treatment difficulties in the domain of alliance remains less clear. Meta-analytic evidence suggests that alliance assessed at various points in treatment (e.g., early, middle, late) is associated with therapy outcome (Flückiger et al., 2018). Further, it appears that aggregating alliance scores across multiple assessment points provides a more dependable metric of alliance than single assessments within outpatient psychotherapy for depression (Crits-Christoph, Connolly Gibbons, Hamilton, Ring-Kurtz, & Gallop, 2011). Thus, an aggregated alliance rating may provide a global signal of the alliance strength over a course of treatment.

Another potential summary of alliance that may be a relevant signal in the context of MBC is the maximum alliance score. In contrast to aggregated scores, maximum alliance ratings indicate whether or not an individual ever reported high levels of alliance. Given a known negative skew to ratings of alliance (Tryon, Blackwell, & Hammel, 2008) and evidence that alliance fluctuates over the course of treatment (Stiles et al., 2004), a patient's failure to ever provide high alliance ratings could also serve as a signal of difficulties in alliance formation. Maximum alliance scores could therefore be used as a method for detecting individuals who have been consistently dissatisfied with their treatment (i.e., always providing lower ratings of alliance) and therefore at risk for drop-out.

Present Evaluation

This evaluation of a quality improvement project involved the implementation of ongoing, brief alliance-focused assessment within the context of a VA SUD clinic predominantly serving veterans with SMI. As treatment engagement is an ongoing concern within the veteran

population (Hoge et al., 2014), as well as among individuals with SMI receiving SUD treatment (Brown, Bennett, Li, & Bellack, 2011; Drake, Mueser, Brunette, & McHugo, 2004), we were interested in whether alliance ratings could be used to predict treatment engagement. Veterans attending SUD groups provided ratings of alliance that were reviewed by group leaders and used to inform treatment (e.g., initiate a discussion with individual group members or the group as a whole regarding aspects of treatment). We examined both average and maximum alliance ratings as predictors of treatment engagement, given that each provide potentially unique information about the strength of the therapeutic relationship.

Our evaluation addressed two empirical hypotheses. First, based on the recommendation to the participating clinicians that they modify their interventions based on alliance feedback, we hypothesized that ratings of alliance would increase over the course of treatment. Second, we hypothesized that alliance ratings would demonstrate clinical utility through predicting treatment engagement. Specifically, we predicted that higher alliance would be associated with greater attendance at SUD groups that implemented the alliance measure as well as at SUD visits in general. We expected that both average and maximum alliance ratings would be associated with attendance.

Method

Patients

Our sample included $n = 98$ veterans receiving SUD treatment within a VA hospital located in a large, metropolitan city. The XXXXX [omitted for blind review] VA is a Level 1 facility – the complexity category of VA facilities with the highest patient volume, level of patient risk, and degree of teaching and research activity. The SUD clinic is one of the largest

substance use treatment programs in the VA system nationally. Patients were seen in one of $n = 9$ SUD groups engaging in alliance-focused MBC.

Sample demographics are reported in Table 1. The average age within the sample was 55.72 years ($SD = 10.66$) and the vast majority of patients were male ($n = 97, 99.0\%$) due in part to the availability of women-only SUD programming at the facility. The sample was majority White ($n = 57, 58.2\%$), followed by Black ($n = 31, 31.6\%$) and other race ($n = 10, 10.2\%$). A minority ($n = 12, 12.2\%$) were Operation Enduring Freedom (OEF) or Operation Iraqi Freedom (OIF) veterans, with the remaining veterans drawn from other eras of service. Eight out of the nine treatment groups were drawn from the co-occurring disorders clinic serving veterans with comorbid SUD (alcohol- and/or drug-use disorders²) and SMI, defined as disorders with high severity. Of note, these comorbidities included psychotic disorders, bipolar disorders, as well as high severity anxiety, PTSD, and depressive disorders (Table 1). The remaining treatment group was drawn from a clinic serving veterans with SUD only or SUD with mild to moderate non-SUD psychiatric comorbidity. The majority of patients ($n = 63, 64.3\%$) had a drug use disorder diagnosis. This project was approved as a Quality Improvement (QI) evaluation by the relevant governing review boards. Data were collected between August 2017 and June 2018. As the evaluation involved analysis of data collected as part of a quality improvement project, informed consent was not deemed necessary by the relevant governing review boards.

Intervention

All patients were enrolled in SUD treatment that included a combination of weekly group sessions and periodic individual sessions. Group and individual treatment was based on relapse prevention and harm reduction principles and included elements of cognitive-behavioral therapy,

² No patients had only a tobacco use disorder. All drug-use disorders included other substances (e.g., cocaine, methamphetamine).

Acceptance and Commitment Therapy, and mindfulness. Patients also received pharmacotherapy from medical providers located in the clinic. The nine treatment groups included in the current project were run by four VA providers (two psychologists, two social workers). Treatment groups were open (i.e., new patients were added to the group on an ongoing basis).

Patients provided alliance ratings at the end of each group meeting on paper forms distributed to the group. Instructions were given to group members to provide their honest responses. Patients were informed that group leaders were interested in learning about their experience in the group and would be reviewing their responses.

Group leaders reviewed all patient responses at the end of each group. Group leaders were encouraged to use their clinical judgment to determine when responses indicated concerns regarding a patient's alliance, in which case this feedback was discussed individually by the group leader with the patient. For example, when a patient indicated concerns regarding the groups "goals and topics," group leaders discussed the patient's understanding of the purpose of the group and the topics that the patient would like to discuss that were not currently being addressed. Group leaders' interventions in response to low alliance scores were not constrained by a particular treatment approach. However, group leaders tended to use motivational interviewing and cognitive-behavioral therapy (e.g., problem-solving) approaches in response to low ratings. Given most ratings were quite high, group leaders tended to discuss feedback that did not follow this pattern (e.g., ratings of 8 or lower).

Measures

Demographic and clinical characteristics. Patients' alliance ratings were linked with demographic and clinical characteristics drawn from their electronic medical record and clinic

intake assessments. Data extracted included patients' age, race/ethnicity, Operation Enduring Freedom (OEF) or Operation Iraqi Freedom (OIF) service, legal involvement (probation or parole), and substance use and mental health diagnoses in the year prior to the initial alliance rating.

Treatment engagement. Treatment engagement was assessed through the electronic medical record and represented as a continuous variable. Engagement was operationalized in two ways: 1) the number of group sessions attended for the group in which alliance was being monitored (ascertained via SUD-related clinic stop code and clinic name), 2) the total number of SUD visit days attended (ascertained via SUD-related clinic stop code). In order to allow comparison of attendance rates across all patients, attendance was tracked for a period of three months prior to and following the initial alliance assessment. The three-month observation period was used regardless of whether or not an individual remained engaged in treatment during that time.

Group Session Rating Scale (GSRS; Duncan & Miller, 2007). Alliance was assessed using the GSRS. The GSRS is a four-item measure designed to measure alliance in the context of group psychotherapy. The GSRS items reflect the theory of alliance in psychotherapy groups, which includes elements of both group cohesion and climate (Burlingame, McClendon, & Alonso, 2011). Respondents provide ratings on a visual analog scale for the following areas: Relationship, Goals and Topics, Approach or Method, and Overall. Descriptive anchors are provided for high and low alliance ratings for each item (e.g., the low Relationship anchor is "I did not feel understood, respected, and/or accepted by the leader and/or the group," the high Approach or Method anchor is "The leader and group's approach is a good fit for me"). The GSRS has been previously used in SUD groups and has shown desirable psychometric properties

including: adequate internal consistency and test-retest reliability; correlation with measures of working alliance, group cohesion, and group climate; and prediction of early treatment change (Quirk, Miller, Duncan, & Owen, 2012). Visual analog responses were measured using a ruler and converted to a 0 to 10-point scale. As recommended previously, the four items were summed to create a total GSRS alliance score (Quirk et al., 2012) which were used in all analyses. The four items showed high internal consistency reliability in the current sample ($\alpha = .92$).

Alliance summary scores for models using alliance as a predictor of treatment engagement were computed in two ways. Alliance was calculated as either the aggregate (i.e., average) of all alliance assessments provided by an individual (Average Alliance) or as the maximum of all alliance ratings provided by an individual (Maximum Alliance) over the three months following the initial alliance rating. For example, if an individual provided the following alliance scores (7, 9, 8, 6) over the 3 months of group participation, the Average Alliance score would be 7.5 and the Maximum Alliance score would be 9.

Although related, these two metrics of alliance provide theoretically distinct information. Average Alliance was intended to provide the most dependable global estimates of an individuals' alliance with the group and group leader, based on prior research indicating that dependability increased with aggregation across multiple observations (Crits-Christoph, Connolly Gibbons, Hamilton, Ring-Kurtz, & Gallop, 2011). In contrast, Maximum Alliance assesses whether or not an individual ever reported high levels of alliance. As low alliance ratings were being discussed with group members and we hypothesized that alliance ratings would therefore increase over time, persistently low ratings may serve as a "signal" of difficulties in the alliance.

Data Analysis

Separate multilevel models (Snijders & Bosker, 2012) were used to test our two hypotheses. To examine changes in GSRS total scores over time, we fit longitudinal multilevel models with GSRS observations nested (Level 1) within 98 patients over time (Level 2).

Subsequent models examined the impact of nesting of patients within nine treatment groups (Level 3) and four providers (Level 4) using χ^2 log-likelihood tests. Demographic covariates (race/ethnicity, OEF/OIF veteran status, age, drug use diagnosis, legal involvement) were also included. The equation for the simpler two-level models was:

$$Y_{ij} = \gamma_{00} + \gamma_{01}(\text{Observation number})_j + [u_{0j} + e_{ij}] \quad (\text{Equation 1})$$

where Y_{ij} is the GSRS score for patient (j) at observation number (i). GSRS scores were predicted by a fixed intercept (γ_{00}) and slope (γ_{01}) that were equivalent across all individuals and a random intercept (u_{0j}) that could vary across individuals, and an error term (e_{ij}). A subsequent model assessed changes in model fit with the addition of a random slope coefficient for observation number.

Next, multilevel models were constructed predicting treatment engagement (i.e., attendance) from either Average Alliance or Maximum Alliance. Again, demographic covariates were also included. Three-level models were tested accounting for nesting of patients within treatment groups and providers:

$$Y_{ijk} = \gamma_{000} + \beta_{1jk}(\text{Alliance})_{ijk} + [v_{00k} + u_{0jk} + e_{ijk}] \quad (\text{Equation 2})$$

where Y_{ijk} is the number of sessions attended by patient (i) in treatment group (j) led by provider (k). Attendance was predicted by a fixed intercept (γ_{000}) and slope (β_{1jk}) that were equivalent across all individuals and random intercepts that could vary across groups (v_{00k}) and providers (u_{0jk}), and an error term (e_{ijk}).

Sensitivity analyses were run with GSRS or attendance outliers excluded. Outliers were defined as values three or more standard deviations from the mean.

In order to increase ease of interpretation of our results, we included Cohen's (1988) f^2 , a standardized effect size that has been recommended for use in the context of multilevel models (Selya et al., 2012). Like more commonly used effect sizes (e.g., Cohen's [1988] d), f^2 also includes guidelines for interpretation of the magnitude of effects (i.e., small, medium, large).

Results

Patients had an average of 10.85 ($SD = 7.07$) SUD visit days and attended 6.37 ($SD = 3.89$) visits of the group where MBC was being implemented (see Supplemental Materials Tables 1 and 2 for the distribution of attendance pre- and post-initial GSRS administration). Patients completed a total of 498 GSRS ratings. GSRS scores were generally high with an overall mean of 8.96 ($SD = 1.36$). The sample mean Average Alliance was 8.82 ($SD = 1.24$) and the sample mean Maximum Alliance was 9.35 ($SD = 1.06$). A large and statistically significant correlation was observed between these two metrics of alliance ($r = .85, p < .001$).

Several outliers (≥ 3 standard deviations from the mean) were identified in the data set including 13 low outlier individual GSRS ratings (out of $n = 498$ total ratings), two low outlier Average Alliance scores (out of $n = 98$ patients), three low outlier Maximum Alliance scores (out of $n = 98$ patients), and one high outlier SUD visits attendance value (out of $n = 98$ patients). As described below, sensitivity analyses were conducted with these values excluded.

Changes in Alliance Over Time

An initial longitudinal multilevel model including only a fixed slope for observation number and a patient-level random intercept showed a significant increase in GSRS scores over time ($B = 0.079, p < .001$). The addition of random intercepts at the group- and provider-level

did not improve model fit ($p > .050$). Examination of the intraclass correlation coefficients indicated that the bulk of the variance occurred at the patient-level ($ICC = .57$) with very little at the group- ($ICC = .0076$) or provider-level ($ICC = .0081$), with the remaining variance representing error variance (i.e., within-patient variance). The significant increase over time observed in the two-level model was essentially unchanged when adding random intercepts for both group and provider ($B = 0.078, p < .001$) and when controlling for patient demographics ($B = 0.075, p < .001$). Model fit was not improved by adding a random slope term ($\chi^2[2] = 4.99, p = .083$), indicating that increases in GSRS scores were relatively uniform across the sample (i.e., slopes did not significantly vary; Snijders & Bosker, 2012). Results from the final four-level model including random intercepts at the patient-, group-, and provider-level along with demographic covariates are presented in Table 2. Based on an f^2 effect size value of 0.033, this model shows a small magnitude increase in GSRS scores over time. Increases in GSRS scores over time remained significant in a sensitivity analysis that excluded low outlier GSRS values ($B = 0.077, p < .001$).

Alliance Predicting Treatment Engagement

Three-level models examined whether alliance ratings were associated with treatment engagement. In models with no additional predictors, Average Alliance was not associated with attendance at either SUD visits in general ($B = 0.76, p = .175$) or MBC group visits ($B = 0.39, p = .212$). Results were unchanged when modeling demographic covariates and when excluding either low Average Alliance outliers or the high SUD visit attendance outlier (Table 3).

Associated standardized effect sizes were also very small (f^2 s = 0.021 and 0.009, for SUD visits in general and MBC group visits, respectively).

In contrast, higher Maximum Alliance scores were associated with greater attendance for both SUD visits in general ($B = 1.86, p = .003$) and MBC group visits ($B = 1.06, p = .003$). Results were unchanged when modeling demographic covariates as well as when excluding low Maximum Alliance outliers and the high SUD visits attendance outlier (Table 3). Associated standardized effect sizes were in the small range (i.e., $0.02 \leq f^2 \leq 0.15$, Cohen [1992]; f^2 s = 0.088 and 0.079, for SUD visits in general and MBC group visits, respectively).

Discussion

MBC has been identified as a priority treatment innovation for improving mental health outcomes with the VA system (Oslin et al., 2006; Pomerantz et al., 2014). The current evaluation examined the clinical utility of routinely monitoring therapeutic alliance in SUD groups for veterans, predominantly those with comorbid SUD and SMI. At the most basic level, our results suggest that alliance-focused MBC is feasible within a VA SUD clinic, notably within a predominantly SMI context where patients frequently present with significant psychiatric and/or cognitive impairments. Patients and providers appeared willing and able to complete a very brief (four item) measure of alliance at the end of each group meeting. Indeed, with MBC presented by clinical staff as a routine part of care, no patients were reported refusing to complete the GSRS measure at the conclusion of group meetings. This is consistent with a previous study showing the feasibility of routine alliance assessment in the context of SUD treatment (Forman et al., 2007). Our results also indicated that patient-rated alliance scores increased over the course of time, with a small magnitude increase with each additional session attended ($f^2 = 0.033$). This was anticipated, as group leaders were explicitly encouraged to discuss alliance ratings with group members, in particular when ratings were low and may therefore indicate difficulties in alliance formation.

Most importantly, it appeared that patient-rated alliance was prognostic of treatment engagement. Interestingly, this was the case only for the maximum alliance rating, and not for the average rating of alliance. Patients whose highest alliance rating was low (e.g., due to dropping out of treatment prior to showing the sample's typical increase over time, failing to ever show the high levels of alliance reported by the modal patient) attended the fewest SUD visits and MBC group visits over the three-month observation period. Based on standardized effect sizes, this relationship was in the small range (f^2 s = 0.088 and 0.079, for SUD visits in general and MBC group visits, respectively), supporting the validity of maximum alliance as a predictor of treatment engagement. This finding specifically suggests that maximum alliance scores may be a valuable signal of difficulties in the alliance. This possibility is consistent with a recent multisite, randomized controlled trial showing the MBC was more effective than treatment-as-usual only for individuals who were not on track for symptom improvement (Delgadillo et al., 2018). It has long been recognized that alliance scores change over the course of psychotherapy (Stiles et al., 2004), and it may be that individuals who never report high ratings of alliance are particularly at risk for treatment drop-out. This possibility should be examined further in larger datasets and those collected in other treatment contexts (e.g., non-SUD treatment, treatments with explicit start and end points).

The lack of relationship between average patient-rated alliance scores and treatment engagement is consistent with the relatively weaker association between alliance and treatment drop-out in general and between alliance and treatment outcome generally in SUD treatment (Flückiger et al., 2018). Of note, the effect for average alliance was in the expected direction, although not statistically significant in our sample of 98 and of small magnitude (f^2 s = 0.021 and 0.009, for SUD visits in general and MBC group visits, respectively). It is possible that a

significant relationship would be observed if examined in a larger sample with greater power to detect small effects. In contrast to maximum alliance ratings, which may function as a signal for alliance difficulties, average alliance ratings may provide a more reliable yet diffuse assessment of alliance that obscures difficulties in alliance formation. Of course, while the two metrics are correlated in the current sample ($r = .85$) and would be mathematically equivalent for individuals who report the same score at all assessments, maximum alliance may be more sensitive in the current context.

Our evaluation project, although with numerous limitations (discussed below), may have implications for psychologists in public sector settings serving in various clinical, administrative, and research capacities. Those in clinical roles could consider collecting a brief assessment of alliance as a method for detecting risk for treatment drop-out. Based on our findings, clinicians would be encouraged to attend to instances in which low alliance ratings are consistently provided, potentially signaling disruption in the therapeutic relationship. Even in the absence of formal quantitative alliance assessment, clinicians can attend to qualitative indications of low alliance (i.e., low agreement on the tasks and goals of therapy, lack of an emotional bond) and attempt to improve these elements of treatment (e.g., through alliance repair; Eubanks, Muran, & Safran, 2018). Administrators may be encouraged by the relative ease of including alliance-focused MBC within treatment, along with initial evidence that this measure includes clinically meaningful information. In tandem with other large-scale MBC efforts, some of which have included brief alliance assessment (e.g., Group Health; Steinfeld et al., 2016), the current evaluation adds to a growing body of literature supporting the feasibility and utility of MBC. These findings may be helpful for increasing stakeholder buy-in, particularly among providers who may be skeptical of MBC (Boswell, Kraus, Miller, & Lambert, 2015; Dowrick et al., 2009).

Based on the current study, alliance may be a candidate construct to include along with symptom measures within existing MBC implementation efforts (Pomerantz et al., 2014). Lastly, for public sector researchers, the current evaluation study may serve as an example of ways in which clinical data generated as a routine part of MBC may be useful to address open research questions. As MBC is implemented at scale, the massive amounts of treatment outcome and process data generated could be used to identify actionable predictors of treatment effectiveness (e.g., clinic-level variables; Clark et al., 2018) as targets for health service innovations.

Future studies could expand on this work. An ideal study would include both a control condition and randomization of patients to alliance MBC and control (e.g., treatment-as-usual) conditions in order to evaluate whether the inclusion of alliance assessment impacts treatment outcomes and engagement. Such a study design would follow numerous other studies conducted on symptom-focused MBC (see Delgadillo et al., 2018; Knaup et al., 2009; Shimokawa et al., 2010), including work examining MBC in military samples with SUD (Schuman, Slone, Reese, & Duncan, 2015). It may be that being able to associate alliance ratings to a specific patient (as was done in the current evaluation) is important for responding to feedback and may explain a lack of outcome improvement associated with anonymous ratings of alliance in the context of SUD treatment (Crits-Christoph et al., 2010). It would also be valuable to see whether routinely monitored alliance predicts treatment engagement in non-SUD treatment contexts, especially those with high risk for treatment drop-out (e.g., post-traumatic stress disorder outpatient clinics within the VA; Imel, Laska, Jakupcak, & Simpson, 2013; Seal et al., 2010). Another avenue to explore is the possibility of assessing alliance ratings outside of a clinical encounter. A growing body of literature documents the use and potential clinical relevance of remote MBC for mental health treatment (Goldberg et al., 2018), and increased sophistication and ubiquity of mobile

technologies may make these methodologies increasingly feasible. Assessing alliance outside of the context of a clinical encounter could alert providers to patients who are at risk for treatment drop-out due to alliance-related concerns.

Future studies could also continue to investigate areas of active debate within the alliance-outcome literature, such as questions of potential patient-level factors that might explain the alliance-outcome relationship (e.g., high problem severity, unstable emotional states; Flückiger et al., 2018). The current evaluation examined some patient-level variables that may theoretically be linked to both ratings of alliance and attendance (e.g., legal involvement, drug use disorders), and found the relationship between maximum alliance and attendance to be robust to these factors. However, future studies could examine more proximal patient-level factors that could explain variance in alliance as a relational variable (e.g., attachment style; Diener, & Monroe, 2011).

The current evaluation has several limitations. First, patients were included at varying points in their treatment. Meta-analytic evidence suggests that the timing of alliance assessment impacts the strength of the alliance-outcome association, with a weaker relationship between early alliance and outcome (Flückiger et al., 2018). Substantively, the meaning of alliance assessed mid-treatment may differ from alliance assessed at the beginning of a course of treatment, as may the trajectory of changes in alliance ratings. Nonetheless, assessing alliance on individuals at varying points in treatment arguably accurately reflects the reality of clinical practice and may therefore be an ecologically-valid representation of alliance-focused MBC in a VA SUD clinic. Including patients at varying points in treatment and outside of a clinical trial context or time-limited groups (in which the beginning and end of treatment are clearly defined) also complicates the assessment of specific features of attendance, such as treatment drop-out. A

future study could examine routinely assessed alliance as a predictor of drop-out within time-limited groups. A second limitation includes the relatively short three-month observation period. A future study could examine the degree to which alliance ratings predict treatment engagement (or other important outcomes such as life functioning and quality of life) at longer follow-up periods. Relatedly, although patients and providers were willing to complete measures during the nine months of this quality improvement project, it is unclear from the current evaluation whether long-term implementation of alliance assessment would have been feasible. A third limitation was the use of a very brief measure of alliance. Brief measures are potentially less reliable than longer assessments (Crocker, & Algina, 2008), which may have introduced additional measurement error and attenuated our statistical power. A fourth limitation was the homogeneity of the current sample in terms of gender and era of service. Lacking non-male identified patients and a larger sample of OEF/OIF veterans may limit the generalizability of our results to female and gender non-binary populations as well as OEF/OIF veterans. A fifth limitation was a lack of information regarding whether a given course of treatment was court mandated. Mandated treatment theoretically may attenuate the alliance-attendance correlation given that individuals are legally required to attend. Although our analyses did model legal involvement, this variable likely overlaps imperfectly with mandated treatment (e.g., individuals may have legal involvement but not be mandated to attend). A future study could examine whether relationships between alliance and engagement differ for those with and without mandated treatment. A sixth limitation was a negative skew in our alliance variable. Although expected (Tryon et al., 2008), this may have attenuated the relationship between alliance and engagement (Cohen, Cohen, West, & Aiken, 2003). A final limitation collapsing across all alliance observations when examining the relationship between alliance and treatment

engagement. This approach necessarily sacrificed potentially meaningful session-by-session fluctuations in alliance ratings. Previous studies have shown session-by-session relationships between alliance and outcome, typically within settings with meaningful session-level outcomes (e.g., symptoms within primary care psychotherapy; Falkenström, Granström, & Holmqvist, 2013). As our outcome of interest was treatment engagement over the course of three months (rather than symptoms at a given session), a session-by-session approach was deemed less informative.

These limitations notwithstanding, the current evaluation is the first to our knowledge to demonstrate feasibility and suggest potential clinical utility of routine assessment of alliance in the context of a VA SUD clinic. Assessment of alliance, a key psychotherapy process variable, may be particularly valuable for populations at risk for treatment drop-out, such as veterans with dually-diagnosed SUD and SMI. It appears that a very brief measure of alliance may be able to detect signals of treatment engagement and may therefore be useful in the context of MBC efforts within the VA and other mental health treatment settings.

References

- Boswell, J. F., Kraus, D. R., Miller, S. D., & Lambert, M. J. (2015). Implementing routine outcome monitoring in clinical practice: Benefits, challenges, and solutions. *Psychotherapy Research, 25*(1), 6-19. doi:10.1080/10503307.2013.817696
- Brown, C. H., Bennett, M. E., Li, L., & Bellack, A. S. (2011). Predictors of initiation and engagement in substance abuse treatment among individuals with co-occurring serious mental illness and substance use disorders. *Addictive behaviors, 36*(5), 439-447. doi:10.1016/j.addbeh.2010.12.001
- Burlingame, G. M., McClendon, D. T., & Alonso, J. (2011). Cohesion in group therapy. *Psychotherapy, 48*(1), 34-42. doi: 10.1037/a0022063
- Burlingame, G. M., McClendon, D. T., & Yang, C. (2018). Cohesion in group therapy: A meta-analysis. *Psychotherapy, 55*(4), 384-398. doi: 10.1037/pst0000173
- Clark, D. M., Canvin, L., Green, J., Layard, R., Pilling, S., & Janecka, M. (2018). Transparency about the outcomes of mental health services (IAPT approach): an analysis of public data. *Lancet, 391*(10121), 679-686.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.)*. Hillsdale, NJ: Erlbaum.
- Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155-159.
- Cohen, J., Cohen, P., West, S.G., & Aiken, L.S. (2003). *Applied multiple regression / correlation analysis for the behavioral sciences (3rd ed.)*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.
- Crits-Christoph, P., Connolly Gibbons, M.B., Hamilton, J., Ring-Kurtz, S., & Gallop, R. (2011). The dependability of alliance assessments: The alliance-outcome correlation is larger

than you think. *Journal of Consulting and Clinical Psychology*, 79(3), 267-278.

doi: 10.1037/a0023668

Crits-Christoph, P., Ring-Kurtz, S., McClure, B., Temes, C., Kulaga, A., Gallop, R.,...& Rotrosen, J. (2010). A randomized controlled study of a web-based performance improvement system for substance abuse treatment providers. *Journal of Substance Abuse Treatment*, 38, 251-262. doi:10.1016/j.jsat.2010.01.001

Crocker, L., & Algina, J. (2008). *Introduction to classical and modern test theory*. Mason, Ohio: Cengage Learning.

Cuijpers, P., van Straten, A., Andersson, G., & van Oppen, P. (2008). Psychotherapy for depression in adults: a meta-analysis of comparative outcome studies. *Journal of Consulting and Clinical Psychology*, 76(6), 909-922.

Delgadillo, J., de Jong, K., Lucock, M., Lutz, W., Rubel, J., Gilbody, S., ... & O'Hayon, H. (2018). Feedback-informed treatment versus usual psychological treatment for depression and anxiety: a multisite, open-label, cluster randomised controlled trial. *Lancet Psychiatry*, 5(7), 564-572. doi: 10.1016/S2215-0366(18)30162-7

Diener, M. J., & Monroe, J. M. (2011). The relationship between adult attachment style and therapeutic alliance in individual psychotherapy: A meta-analytic review. *Psychotherapy*, 48(3), 237-248. doi: 10.1037/a0022425

Dowrick, C., Leydon, G. M., McBride, A., Howe, A., Burgess, H., Clarke, P.,...& Kendrick, T. (2009). Patients' and doctors' views on depression severity questionnaires incentivised in UK quality and outcomes framework: qualitative study. *BMJ*, 338, b663. doi:10.1136/bmj.b663

Duncan, B. L., & Miller, S. D. (2007). *The Group Session Rating Scale*. Jensen Beach, FL:

Author.

Eubanks, C. F., Muran, J. C., & Safran, J. D. (2018). Alliance rupture repair: A meta-analysis.

Psychotherapy, 55(4), 508-519. doi: 10.1037/pst0000185

Falkenström, F., Granström, F., & Holmqvist, R. (2013). Therapeutic alliance predicts

symptomatic improvement session by session. *Journal of Counseling Psychology*, 60(3), 317-328. doi: 10.1037/a0032258

Forman, R., Crits-Christoph, P., Kaynak, Ö., Worley, M., Hantula, D. A., Kulaga, A., ... &

Muchowski, P. (2007). A feasibility study of a web-based performance improvement system for substance abuse treatment providers. *Journal of Substance Abuse Treatment*, 33(4), 363-371.

Fortney, J. C., Unützer, J., Wrenn, G., Pyne, J. M., Smith, G. R., Schoenbaum, M., & Harbin, H.

T. (2017). A Tipping Point for Measurement-Based Care. *Psychiatric Services*, 68(2), 179-188. doi: 10.1176/appi.ps.201500439

Flückiger, C., Del Re, A.C., Wampold, B.E., & Horvath, A.O. (2018). The alliance in adult psychotherapy: A meta-analytic synthesis. *Psychotherapy*, 55(4), 316-340.

doi:10.1037/pst0000172

Goldberg, S.B., Babins-Wagner, R., Rousmaniere, T., Berzins, S., Hoyt, W.T., Whipple, J.L.,

Miller, S.D., & Wampold, B.E. (2016). Creating a climate for therapist improvement: A case study of an agency focused on outcomes and deliberate practice. *Psychotherapy*, 53(3), 367-375. doi: 10.1037/pst0000060

Goldberg, S. B., Buck, B., Raphaely, S., & Fortney, J. C. (2018). Measuring psychiatric

symptoms remotely: A systematic review of remote measurement-based care. *Current Psychiatry Reports*, 20(10). doi: 10.1007/s11920-018-0958-z

- Harding, K.J.K., Rush, A.J., Arbuckle, M., Trivedi, M.H., & Pincus, H.A. (2011). Measurement-based care in psychiatric practice: A policy framework for implementation. *Journal of Clinical Psychiatry, 72*(8), 1136-1143.
- Hayes, A.M., Laurenceau, J., Feldman, G., Strauss, J.L., & Cardaciotto, L. (2007). Change is not always linear: The study of nonlinear and discontinuous patterns of change in psychotherapy. *Clinical Psychology Review, 27*, 715-723.
- Hoge, C. W., Grossman, S. H., Auchterlonie, J. L., Riviere, L. A., Milliken, C. S., & Wilk, J. E. (2014). PTSD treatment for soldiers after combat deployment: low utilization of mental health care and reasons for dropout. *Psychiatric Services, 65*(8), 997-1004.
- Imel, Z.E., Laska, K., Jakupcak, M., & Simpson, T.L. (2013). Meta-analysis of dropout in treatments for posttraumatic stress disorder. *Journal of Consulting and Clinical Psychology, 81*(3), 394-404. doi: 10.1037/a0031474
- Johnson, J. E., Burlingame, G. M., Olsen, J. A., Davies, D. R., & Gleave, R. L. (2005). Group climate, cohesion, alliance, and empathy in group psychotherapy: Multilevel structural equation models. *Journal of Counseling Psychology, 52*(3), 310-321.
doi: 10.1037/0022-0167.52.3.310
- Joyce, A. S., Piper, W. E., & Ogrodniczuk, J. S. (2007). Therapeutic alliance and cohesion variables as predictors of outcome in short-term group psychotherapy. *International Journal of Group Psychotherapy, 57*(3), 269-296.
- Knaup, C., Koesters, M., Schoefer, D., Becker, T., & Puschner, B. (2009). Effect of feedback of treatment outcome in specialist mental healthcare: A meta-analysis. *British Journal of Psychiatry, 195*, 15-22. doi: 10.1192/bjp.bp.108.053967
- Lambert, M. J., Hansen, N. B., & Finch, A. E. (2001). Patient-focused research: Using patient

- outcome data to enhance treatment effects. *Journal of Consulting and Clinical Psychology*, 69, 159–172. doi: 10.1037/0022-006X.69.2.159
- Lambert, M. J. (Ed.). (2013). *Bergin and Garfield's handbook of psychotherapy and behavior change*. New York, NY: John Wiley & Sons.
- Lambert, M.J. (2015). Progress feedback and the OQ-system: The past and future. *Psychotherapy*, 52(4), 381-390. doi: 10.1037/pst0000027
- Miller, S. D., & Bertolino, B. (Eds.). (2012). *ICCE feedback informed treatment training and treatment manuals*. Chicago: ICCE.
- Miller, S.D., Hubble, M.A., Chow, D., & Seidel, J. (2015). Beyond measures and monitoring: Realizing the potential of feedback-informed treatment. *Psychotherapy*, 52(4), 449-457. doi: 10.1037/pst0000031
- Norcross, J.C. & Lambert, M.J. (Eds.). (in press). *Psychotherapy relationships that work (3rd ed.)*. New York: Oxford University Press.
- Oslin, D. W., Ross, J., Sayers, S., Murphy, J., Kane, V., & Katz, I. R. (2006). Screening, assessment, and management of depression in VA primary care clinics: the Behavioral Health Laboratory. *Journal of General Internal Medicine*, 21(1), 46-50.
- Pomerantz, A. S., Kearney, L. K., Wray, L. O., Post, E. P., & McCarthy, J. F. (2014). Mental health services in the medical home in the Department of Veterans Affairs: Factors for successful integration. *Psychological Services*, 11(3), 243-253. doi: 10.1037/a0036638
- Prescott, D. S., Maeschalck, C. L., & Miller, S. D. (Eds.). (2017). *Feedback-informed treatment in clinical practice: Reaching for excellence*. Washington, DC, US: American Psychological Association. doi: 10.1037/0000039-000
- Quirk, K., Miller, S., Duncan, B., & Owen, J. (2013). Group Session Rating Scale: Preliminary

- psychometrics in substance abuse group interventions. *Counselling and Psychotherapy Research*, 13(3), 194-200. doi:10.1080/14733145.2012.744425
- Schuman, D. L., Slone, N. C., Reese, R. J., & Duncan, B. (2015). Efficacy of client feedback in group psychotherapy with soldiers referred for substance abuse treatment. *Psychotherapy Research*, 25(4), 396-407.
- Seal, K. H., Maguen, S., Cohen, B., Gima, K. S., Metzler, T. J., Ren, L., ... & Marmar, C. R. (2010). VA mental health services utilization in Iraq and Afghanistan veterans in the first year of receiving new mental health diagnoses. *Journal of Traumatic Stress*, 23(1), 5-16. doi:10.1002/jts.20493
- Selya, A. S., Rose, J. S., Dierker, L. C., Hedeker, D., & Mermelstein, R. J. (2012). A practical guide to calculating Cohen's f^2 , a measure of local effect size, from PROC MIXED. *Frontiers in Psychology*, 3, 111. doi: 10.3389/fpsyg.2012.00111
- Shimokawa, K., Lambert, M.J., & Smart, D.W. (2010). Enhancing treatment outcome of patients at risk of treatment failure: Meta-analytic and mega-analytic review of a psychotherapy quality assurance system. *Journal of Consulting and Clinical Psychology*, 78(3), 298-311. doi: 10.1037/a0019247
- Snijders, T.A. & Bosker, R.J. (2012). *Multilevel analysis: An introduction to basic and advance multilevel modeling (2nd ed.)*. London: Sage.
- Steinfeld, B., Franklin, A., Mercer, B., Fraynt, R., & Simon, G. (2016). Progress monitoring in an integrated health care system: Tracking behavioral health vital signs. *Administration and Policy in Mental Health and Mental Health Services Research*, 43(3), 369-378. doi: 10.1007/s10488-015-0648-7
- Stiles, W.B., Glick, M., Osatuke, K., Hardy, G.E., Shapiro, D.A., Agnew-Davies, R.,....&

- Barkham, M. (2004). Patterns of alliance development and the rupture-repair hypothesis: Are productive relationships U-shaped or V-shaped? *Journal of Counseling Psychology, 51*(1), 81-92. doi: 10.1037/0022-0167.51.1.81
- Tasca, G. A., Compare, A., Zarbo, C., & Brugnera, A. (2016). Therapeutic alliance and binge-eating outcomes in a group therapy context. *Journal of Counseling Psychology, 63*(4), 443-451. doi: 10.1037/cou0000159
- Tryon, G. S., Blackwell, S. C., & Hammel, E. F. (2008). The magnitude of client and therapist working alliance ratings. *Psychotherapy: Theory, Research, Practice, Training, 45*(4), 546-551. doi: 10.1037/a0014338
- Wampold, B.E. (2015). Routine outcome monitoring: Coming of age - with the usual developmental challenges. *Psychotherapy, 52*(4), 458-462. doi: 10.1037/pst0000037
- Warner, R. (2009). Recovery from schizophrenia and the recovery model. *Current Opinion in Psychiatry, 22*(4), 374-380. doi:10.1097/YCO.0b013e32832c920b
- Westen, D., & Morrison, K. (2001). A multidimensional meta-analysis of treatments for depression, panic, and generalized anxiety disorder: an empirical examination of the status of empirically supported therapies. *Journal of Consulting and Clinical Psychology, 69*(6), 875-899.
- Witkiewitz, K., & Marlatt, G. A. (2007). Modeling the complexity of post-treatment drinking: It's a rocky road to relapse. *Clinical Psychology Review, 27*(6), 724-738.

Table 1. Sample demographics

Variable	<i>n</i>	%	Mean	<i>SD</i>	Min	Max
Race/ethnicity						
White	57	58.2%				
Black	31	31.6%				
Other race	10	10.2%				
Male gender	97	99.0%				
Age			55.72	10.66	27.10	71.80
Substance use disorders						
Alcohol	87	88.8%				
Stimulant	52	53.1%				
Cannabis	30	30.6%				
Opioid	11	11.2%				
Other substance use	9	9.2%				
Any drug disorder	63	64.3%				
Mental health disorders						
Depression	54	55.1%				
Bipolar	22	22.4%				
Psychotic	18	18.4%				
PTSD	50	51.0%				
Anxiety	35	35.7%				
Legal Involvement	37	37.8%				
OEF/OIF era	12	12.2%				
GSRs Total			8.96	1.36	0.75	10
Average Alliance			8.82	1.24	4.67	10
Maximum Alliance			9.35	1.06	5	10
SUD visit days baseline			9.86	6.04	1	25
SUD visit days index			10.85	7.07	1	41
MBC group attendance baseline			3.54	4.17	0	18
MBC group attendance index			6.37	3.89	1	14

Note: GSRs = Group Session Rating Scale; SUD visit days = all substance use disorder-related visit codes from medical record in three months prior to initial GSRs rating (i.e., baseline) or following initial GSRs rating (i.e., index); MBC (measurement-based care) group visits = all visits to group where GSRs was assessed in three months prior to initial GSRs rating (i.e., baseline) or following initial GSRs rating (i.e., index). Possible range of all alliance scores (GSRs Total, Average Alliance, Maximum Alliance) was from 0 to 10.

Table 2. Changes in Group Session Rating Scale scores over time

Predictor	<i>B</i>	95% CI	<i>p</i>	<i>f</i> ² *
Session number	0.075	0.037 – 0.114	< .001	0.033
Race				
White	ref			
Black	-0.048	-0.548 – 0.453	.852	
Other race	-0.055	-0.774 – 0.663	.880	
Male gender	-0.196	-2.413 – 2.022	.863	
OEF/OIF	0.289	-0.579 – 1.157	.514	
Age	-0.001	-0.028 – 0.025	.926	
Any drug use diagnosis	-0.404	-0.876 – 0.069	.094	
Legal involvement	-0.051	-0.492 – 0.390	.820	
Intercept	9.140	6.400 – 11.880	< .001	

Note: OEF/OIF = Operation Enduring Freedom / Operation Iraqi Freedom.

*Cohen's *f*²: *f*² ≥ 0.02, *f*² ≥ 0.15, and *f*² ≥ 0.35 represent small, medium, and large effect sizes, respectively

Table 3. Average and maximum Group Rating Scale scores predicting treatment engagement

		<u>Primary Analyses</u>					
Model	Outcome	Alliance Predictor	<i>B</i>	95% CI	<i>p</i>	<i>f</i> ^{2*}	
Primary (<i>n</i> = 98)	Total SUD	Average	0.80	-0.29 – 1.89	.150	0.021	
	MBC Group	Average	0.29	-0.31 – 0.90	.341	0.009	
	Total SUD	Maximum	1.79	0.56 – 3.01	.004	0.088	
	MBC Group	Maximum	0.96	0.28 – 1.65	.006	0.079	
		<u>Sensitivity Analyses</u>					
Model	Outcome	Alliance Predictor	<i>B</i>	95% CI	<i>p</i>	<i>f</i> ^{2*}	
Excluding low Average Alliance outliers (<i>n</i> = 96)	Total SUD	Average	0.95	-0.29 – 2.19	.134	0.024	
	MBC Group	Average	0.37	-0.32 – 1.06	.297	0.012	
Excluding low Maximum Alliance outliers (<i>n</i> = 95)	Total SUD	Maximum	2.30	0.68 – 3.91	.005	0.089	
	MBC Group	Maximum	1.42	0.53 – 2.31	.002	0.103	
Excluding high SUD visits outliers (<i>n</i> = 97)	Total SUD	Average	0.51	-0.49 – 1.51	.317	0.01	
	Total SUD	Maximum	1.55	0.44 – 2.67	.006	0.08	

Note: All models include race/ethnicity, age, gender, era of service, and drug use disorder as covariates. Outliers defined as values three or more standard deviations from the mean. SUD = substance use disorder; MBC Group = Group in which measurement-based care was being evaluated; Average = Average Alliance; Maximum = Maximum Alliance; Total SUD = Total SUD visit days; MBC Group = MBC Group visit days

*Cohen’s *f*²: *f*² ≥ 0.02, *f*² ≥ 0.15, and *f*² ≥ 0.35 represent small, medium, and large effect sizes, respectively

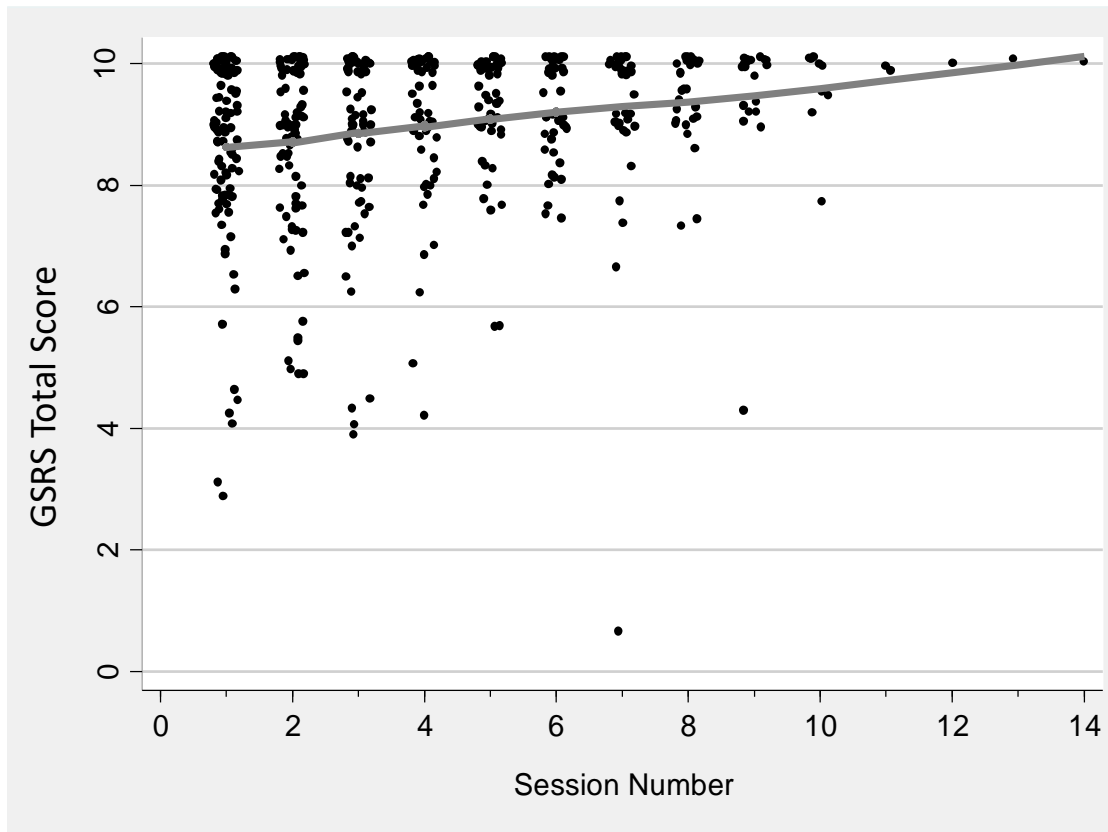


Figure 1. Changes in Group Session Rating Scale (GSR) total score over time (session number). Points jittered for ease of interpretation.

Supplemental Materials Table 1. GSRS group visit days in the three months before and after MBC start

Visit Days	Baseline		Index	
	<i>n</i>	%	<i>n</i>	%
0	36	36.7	0	0.0
1	14	14.3	10	10.2
2	4	4.1	10	10.2
3	7	7.1	12	12.2
4	4	4.1	10	10.2
5	3	3.1	3	3.1
6	3	3.1	8	8.2
7	7	7.1	3	3.1
8	6	6.1	10	10.2
9	1	1.0	4	4.1
10	5	5.1	8	8.2
11	4	4.1	9	9.2
12	2	2.0	5	5.1
13	1	1.0	5	5.1
14	0	0.0	1	1.0
18	1	1.0	0	0.0

Note: GSRS = Group Session Rating Scale; Baseline = visits in the three months prior to initial GSRS administration; Index = visits in the three months following initial GSRS administration.

Supplemental Materials Table 2. All SUD visit days in the three months before and after MBC start

Visit Days	Baseline		Index	
	<i>n</i>	%	<i>n</i>	%
1	3	3.1	2	2.0
2	7	7.1	5	5.1
3	6	6.1	7	7.1
4	11	11.2	1	1.0
5	4	4.1	6	6.1
6	1	1.0	5	5.1
7	6	6.1	10	10.2
8	8	8.2	8	8.2
9	4	4.1	4	4.1
10	5	5.1	9	9.2
11	7	7.1	4	4.1
12	4	4.1	7	7.1
13	8	8.2	4	4.1
14	1	1.0	4	4.1
15	2	2.0	2	2.0
16	4	4.1	2	2.0
17	6	6.1	1	1.0
18	1	1.0	0	0.0
19	3	3.1	4	4.1
20	1	1.0	4	4.1
21	2	2.0	2	2.0
22	2	2.0	2	2.0
23	0	0.0	0	0.0
24	1	1.0	0	0.0
25	1	1.0	2	2.0
27	0	0.0	1	1.0
31	0	0.0	1	1.0
41	0	0.0	1	1.0

Note: GSRS = Group Session Rating Scale; Baseline = visits in the three months prior to initial GSRS administration; Index = visits in the three months following initial GSRS administration.