THE THERAPEUTIC ALLIANCE OVER 10 SESSIONS OF THERAPY FOR BORDERLINE PERSONALITY DISORDER: AGREEMENT AND CONGRUENCE ANALYSIS AND RELATION TO OUTCOME

Yogev Kivity, PhD, Kenneth N. Levy, PhD, Stéphane Kolly, MD, and Ueli Kramer, PhD

The authors examined whether alliance dynamics are affected by tailoring the therapeutic relationship to the individual patient in brief psychotherapy of borderline personality disorder. Sixty patients were randomized to 10-session Good Psychiatric Management (GPM-BV) or GPM combined with Motive-Oriented Therapeutic Relationship techniques (MOTR+GPM-BV). Patient- and therapist-rated alliance was assessed weekly. Self-reported symptomatic distress was assessed pre-, mid-, and posttreatment. In MOTR+GPM-BV, stronger therapist-rated alliance predicted lower symptomatic distress in the same timepoint, but not in a lag, whereas symptomatic distress predicted therapist-rated alliance in a lag. Therapist-rated alliance was lower than patient-rated alliance in GPM-BV but not in MOTR+GPM-BV. In MOTR+GPM-BV, higher agreement on strong alliance tended to predict lower symptomatic distress. Patient- and therapist-rated alliances were temporally congruent, but congruence did not predict outcome. Addressing the relationship needs of patients may partly exert its salutary effect by increasing agreement between patients' and therapists' experience of the alliance.

Keywords: therapeutic alliance, borderline personality disorder, good psychiatric management, motive oriented therapeutic relationship, plan analysis

Borderline personality disorder (BPD) is a severe and prevalent disorder characterized by instability in interpersonal relationships as well as emotion dysregulation and impulsivity (Skodol et al., 2002). A number of efficacious treatments for BPD exist (e.g., Bateman & Fonagy, 1999; Clarkin, Levy, Lenzenweger, & Kernberg, 2007; Gunderson & Links, 2008; Linehan, Armstrong, Suarez, Allmon, & Heard, 1991), but their mechanisms of change remain poorly understood.

One of the hallmark features of BPD is unstable and tumultuous relationships that tend to vacillate between idealization and devaluation (American Psychiatric Association, 2013). As clinicians and researchers have noted, the difficulties in interpersonal relations of these patients pose challenges in building a strong therapeutic alliance with BPD patients (e.g., Gabbard et al., 1988; Yeomans et al., 1994). These difficulties may explain the fluctuations in the alliance with BPD patients (Levy, Beeney, Wasserman, & Clarkin, 2010) and may undermine the utility of the therapeutic alliance, which is known to be a robust predictor of good treatment outcome (Horvath, Del Re, Flückiger, & Symonds, 2011). Indeed, a recent meta-analysis (Levy, Scala, & Ellison, 2017) found that the alliance–outcome correlation in treatments of BPD was weak ($r = .12$), and significantly smaller than the moderate correlation reported in the most up-to-date meta-analysis of the alliance–outcome correlation (Horvath et al., 2011).

One possible way to mitigate the difficulties of forming a therapeutic alliance with patients with BPD is by increasing the responsiveness to their relationship needs using individualized treatment plans. Indeed, a recent study (Kramer, Kolly, et al., 2014) examined the utility of using an ingredient of treatment individualization, called the motive-oriented therapeutic relationship (MOTR) method (Caspar, 2007), for treatment outcome. MOTR is based on the principles of Plan Analysis (PA), an integrative case conceptualization that includes hierarchical structure of inferred goals and means that underlie the patient's interpersonal or intrapsychic behaviors and experiences (Caspar, 2007; Grawe, 1980). MOTR assumes that if a therapist addresses the patient's underlying goals within the limits of the therapeutic relationship, it is no longer necessary for the patient to use problematic means to attain personal goals (Caspar, 2007). Because each patient is assumed to be guided by different hypothetical Plans, the therapeutic relationship offer must be constructed differently for each patient, in a fundamentally idiographic and thus responsive manner (Caspar, 2007; Kramer & Stiles, 2015). Theoretically, these adaptations could help overcome some of the difficulties in alliance in the treatment of BPD by creating a more stable alliance. In addition, negotiating and addressing the therapeutic alliance should result in increased agreement between therapists and patients on the alliance and in better ability to capitalize on the alliance to promote better outcomes.

Kramer and colleagues (Kramer, Flückiger, et al., 2014; Kramer, Kolly, et al., 2014) examined the effects of adding MOTR to a 10-session, brief version of Good Psychiatric Management for BPD (GPM-BV; Gunderson & Links, 2008). GPM is a psychiatric and psychotherapeutic approach based on an attachment-informed etiological model of BPD that includes components such as establishment of psychiatric diagnosis, goal setting, motiva-

tion enhancement, and relational interpretations of core conflictual themes. Kramer and colleagues (Kramer, Flückiger, et al., 2014; Kramer, Kolly, et al., 2014) found that MOTR+GPM-BV resulted in greater changes in symptomatic distress, the primary outcome measure in the study. In addition, both treatments were efficacious in general and led to similar changes in BPD-specific symptoms.

Given the presumably central role of the alliance in MOTR, Kramer and colleagues (Kramer, Flückiger, et al., 2014; Kramer, Kolly, et al., 2014) also conducted preliminary examinations of alliance ratings and their relations to outcome. It was found that therapist-rated (but not patient-rated) alliance tended to increase more rapidly in MOTR+GPM-BV compared to GPM-BV, while there were no differences between treatments in average levels (across sessions) of the alliance (Kramer, Kolly, et al., 2014). Kramer, Flückiger, et al. (2014) also examined the correlations between alliance ratings in each of the 10 sessions of treatment and overall change in symptomatic distress. On average across the 10 sessions, stronger patient-rated alliance predicted better outcome in MOTR+GPM-BV but not in GPM-BV, while, on the contrary, weaker therapist-rated alliance predicted better outcome in MOTR+GPM-BV but not in GPM-BV (Kramer, Flückiger, et al., 2014).

Thus, these preliminary results show that MOTR+GPM-BV is related to better outcome, but it is still unclear what drives this effect and how it is related to the therapeutic alliance. In addition, these studies examined the associations between alliance and outcome using zero-order correlations on all available data, without taking into consideration the multilevel structure of the data and without handling missing data. Thus, the present study aims to examine the dynamics of the therapeutic alliance in the Kramer and colleagues studies (Kramer, Flückiger, et al., 2014; Kramer, Kolly, et al., 2014). In addition, by utilizing multilevel modeling on imputed data, the present study is expected to increase the power of the analyses and result in more accurate understanding of the dynamic and longitudinal association between alliance and outcome (Enders, Mistler, & Keller, 2016). These data-analytic approaches may provide a more nuanced understanding of the role of the alliance in MOTR+GPM-BV, and in early phases of treatments of BPD in general.

As mentioned above, Kramer, Flückiger, et al. (2014) found that MOTR+GPM-BV resulted in more rapid increases in therapist-rated alliance with no differences in average levels of the alliance. This suggests that *changes* in therapist-rated alliance may be more strongly related to outcome in MOTR+GPM-BV than the average strength of the alliance per se. Some studies used repeated measurements of the alliance to disentangle the within- and between-patient variability in the alliance and it relation to outcome (reviewed in Zilcha-Mano, 2017). The within-patient variability in alliance ratings (henceforth session-level alliance) is a "state-like" measure that captures the fluctuations of the alliance for a given patient during therapy, whereas the between-patient variability in alliance ratings (henceforth patient-level alliance) is a "trait-like" measure that captures the average level of the alliance for a given patient during therapy. These studies found that session-level alliance is generally related to outcome (reviewed in Zilcha-Mano, 2017),

which is consistent with the findings in the studies by Kramer and colleagues (Kramer, Flückiger, et al., 2014; Kramer, Kolly, et al., 2014), although a direct examination of this possibility in the Kramer and colleagues' (Kramer, Flückiger, et al., 2014; Kramer, Kolly, et al., 2014) dataset is still needed.

In addition, the finding that patient- and therapist-rated alliances were differentially associated with outcome emphasizes the importance of examining the degree of correspondence between these two perspectives, and the way it differs by treatment. Recent studies have begun to examine this correspondence by focusing on two different aspects of correspondence: the temporal covariation of patient- and therapist-rated alliance ("temporal congruence") and the absolute differences between these ratings within specific sessions ("disagreement") and the way they predict outcome (Atzil-Slonim et al., 2015; Compare, Tasca, Lo Coco, & Kivlighan, 2016; Marmarosh & Kivlighan, 2012; Zilcha-Mano, Snyder, & Silberschatz, 2017). In these studies, therapists tended to rate the alliance as weaker than their patients, while conflicting findings exist regarding the degree of temporal congruence between patient and therapist with some studies reporting a positive correlation between patient- and therapist-rated alliance over time and some studies failing to find such a correlation (also reviewed in Tryon, Blackwell, & Hammel, 2007). Also, it is still unclear whether agreement and congruence predict outcome, and if so, whether the session-level or patient-level component of these measures is more predictive. Existing studies so far have not examined differences in agreement and congruence between different types of treatments, which could possibly explain their conflicting findings.

In sum, the present study examines session- and patient-level alliance and the correspondence between patient- and therapist-rated alliance in brief treatments of BPD, and the way these are related to outcome. We had the following hypotheses. (a) Given previous findings (Kramer, Kolly, et al., 2014), we expected session-level alliance to predict outcome in MOTR+GPM-BV but not in GPM-BV. To better control for temporality, we examined prediction of outcome at the same, as well as subsequent, timepoints. We did not expect patient-level alliance to predict outcome. (b) Based on previous findings (Atzil-Slonim et al., 2015; Compare et al., 2016), we expected therapists to rate the alliance as weaker than patients ("directional disagreement"). Given the explicit focus on responsiveness in MOTR, we expected the degree of agreement to be higher in MOTR+GPM-BV compared to GPM-BV. We did not have a strong hypothesis regarding the mean temporal congruence of patient and therapist ratings across treatment types because of the conflicting findings in the literature. However, we did expect the degree of temporal congruence to be higher in MOTR+GPM-BV compared to GPM-BV because of the focus on responsiveness in this treatment. (c) We examined the association of agreement and congruence with treatment outcome without strong hypotheses regarding the nature of these associations because of the mixed findings in the literature (Atzil-Slonim et al., 2015; Compare et al., 2016; Marmarosh & Kivlighan, 2012; Zilcha-Mano et al., 2017).

METHOD
PARTICIPANTS

Participants were outpatients at a European French-speaking university psychiatry clinic who met *DSM-IV-TR* criteria for BPD (American Psychiatric Association, 2000). In the original study, 85 patients were randomized but 11 patients discontinued after the first treatment session (before MOTR was introduced), and therefore were not included in analyses. Out of the 74 patients who were included in the original intent-to-treat analyses, 60 patients had at least one report of the alliance and were included in the present study. Out of these, two MOTR+GPM-BV and two GPM-BV patients dropped out during treatment but were still included in analyses. The final sample consisted of 32 MOTR+GPM-BV patients (17 females, age 20–55, M = 35.31, SD = 9.89), and 28 GPM-BV patients (23 females, age 20–55, M = 31.43, SD = 9.19). The distribution of family status was: single (MOTR+GPM-BV = 25%, GPM-BV = 54%), married (MOTR+GPM-BV = 47%, GPM-BV = 21%), divorced/separated (MOTR+GPM-BV = 28%, GPM-BV = 25%). No group differences emerged in the rate of dropout and in age, but the GPM-BV group had significantly more females and single patients, and fewer married patients.

TREATMENTS AND THERAPISTS

Twenty-two therapists were randomized to treatment condition (ITT sample: GPM-BV, n = 13; MOTR+GPM-BV, n = 9). GPM-BV therapists treated 1–11 patients (M = 2.9), and MOTR+GPM-BV therapists treated 1–14 patients (M = 4). Therapists were psychiatrists/psychologists (n = 19) and nurses (n = 3) with at least a basic psychodynamic training and an average of 2.5 years of residency. Therapists were trained in GPM-BV at the outset of the study and were supervised during the entire study. MOTR+GPM-BV therapists were trained and supervised by the model developer and an expert in this approach. Therapists received individual supervision twice over the course of each treatment (after Session 1 and at midtreatment).

GPM-BV. This brief version of GPM included a psychiatric and psychotherapeutic approach based on an attachment-informed etiological model of BPD (see Gunderson & Links, 2008). A specific manual was elaborated to adapt the treatment principles to 10 weekly sessions (Kolly et al., 2010) that included: (a) establishment of psychiatric diagnoses and communication of them to the patient, (b) psychiatric history taking, (c) identification of main problems and establishment of treatment focus, (d) definition of short-term objectives and general enhancement of motivation, (e) dealing with treatment-interfering problems, and (f) formulation of relational interpretations of core conflictual themes.

MOTR+GPM-BV. This treatment included, in addition to the GPM-BV intervention, formulating a PA and use of MOTR techniques when indicated (Caspar, 2007). PA is an integrative case conceptualization method with ensuing relational techniques (MOTR). In PA, the therapist makes inferences about the patient's implied underlying needs based on his or her verbal and nonverbal behavior. Based on PA, the therapist defines and implements relational techniques (MOTR) in order to address the patient's needs and motives within the limits of the therapeutic relationship, without reinforcing problematic behaviors or experiences. The intake session provided the therapist with information for the establishment of the PA and the ensuing MOTR, which were implemented in Sessions 2–10.

As reported by Kramer, Kolly, et al. (2014), there was excellent adherence to GPM-BV principles for both GPM-BV (GPM adherence scale: $M = 4.32$; $SD = .37$) and MOTR+GPM-BV ($M = 4.37$; $SD = .26$), which did not differ by condition ($p = .57$). Adherence to MOTR was higher in MOTR+GPM-BV (PA and MOTR ratings; $M = 1.55$; $SD = .44$) than in GPM-BV ($M = .48$; $SD = .39$; $p < .01$).

MEASURES

Working Alliance Inventory–Short Form. The French version (Corbière, Bisson, Lauzon, & Ricard, 2006) of this 12-item questionnaire (Tracey & Kokotovic, 1989) was used to assess patient- (WAI-P) and therapist-rated (WAI-T) alliance on a 1 (*never*) to 7 (*always*) scale (e.g., "I feel that my therapist appreciates me"). Questionnaires were completed after each session, and total scores were used for analyses. Multilevel internal consistency coefficients (Shrout & Lane, 2012) in the current study were $\alpha = .99$ for both WAI-P and WAI-T.

Outcome Questionnaire-45.2 (OQ-45). This 45-item self-report questionnaire (Lambert et al., 2004) was used to assess symptomatic distress on a 0 (*never*) to 4 (*always*) scale (e.g., "I feel stressed at work/school") during the past week, including the day of the assessment. Questionnaires were completed at pre-, mid-, and posttreatment, and total scores were used for analyses. Multilevel internal consistency in the current study was $\alpha = .98$. We focused on this measure of all outcome measures collected in the study because this was the primary outcome measure in the original study.

DATA PREPARATION AND DATA-ANALYTIC APPROACH

All data analyses were implemented in R version 3.3.2 (R Core Team, 2016).

Missing Data Imputation. Out of the possible 600 session-by-session timepoints in the study (60 patients × 10 sessions), 195 timepoints had missing data in at least one of the study variables (149–188 within measures). Out of the possible 180 OQ-45 measurement timepoints in the study (60 patients

× 3 measurements), 33 timepoints had missing data. Because these data cannot be assumed to be missing completely at random, we employed multilevel multiple imputation by chained equations (van Buuren & Groothuis-Oudshoorn, 2011) to arrive at a complete dataset and reduce possible biases in random slopes models due to nonresponse (Enders et al., 2016). Data were imputed using package "mice" (van Buuren & Groothuis-Oudshoorn, 2011) with 20 imputations and 20 iterations for each imputation. Patient-level predictors included clinical and demographic variables at intake, dropout status, and treatment type. Time-level predictors included session number and all outcome and process variables in the study. Postimputation diagnostics suggested that the imputation solution achieved good results. Thus, 600 timepoints were included in analyses of session-by-session variables, and 180 timepoints were included in analyses involving the OQ-45. Subsequent analyses were run repeatedly on all copies of the data and then pooled to arrive at the final estimates.

Changes in Alliance During Treatment. We first attempted to replicate the analyses in the Kramer, Kolly, et al. study (2014) regarding differences between treatments in change in alliance. Given the hierarchical nature of our data (assessments repeated within patients), data were analyzed using linear multilevel models in package "nlme" (Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2016). Analyses were adjusted for repeated measures with restricted maximum likelihood estimation method and random intercepts and slopes at the patient level. In addition, a first-order autoregressive covariance structure [AR(1)] was modeled at Level 1. AR(1) was chosen in order to account for auto-regression in the dependent variable and at the same time avoid a possible bias due to inclusion of the lagged dependent variable as a predictor in the model (see Falkenström, Finkel, Sandell, Rubel, & Holmqvist, 2017, for a detailed discussion). Session number was included as a Level 1 predictor while treatment type was included as a Level 2 predictor. Session number was centered around the first session, so intercepts represent the alliance score at the beginning of treatment, while slopes represent the average change in alliance between sessions.[1]

Disaggregation of Session- and Patient-Level Alliance. We decomposed session- and patient-level alliance following procedures recommended by Wang and Maxwell (2015). This approach allows us to disentangle trait-like from state-like components of the alliance by looking at the variability in alliance scores at the within- and between-patient levels. Mean alliance scores for a given patient served as patient-level scores (centered around the grand

---

1. We also utilized longitudinal cluster analysis to examine how patients cluster in their patterns of alliance development. We found two different clusters of linear development that differed only in the baseline level and rate of growth, but not in the shape of the trajectory of change. These findings are not reported here due to space limitations and because they did not contribute to the understanding of the development of the alliance beyond what was already revealed by the multilevel models. Full details are available upon request from the first author.

mean), whereas session-by-session scores served as session-level scores (centered around the individual patient mean).[2]

To better understand the temporal sequence of these relationships, we examined how alliance predicts symptomatic distress in the same as well as in subsequent timepoints (with a lag of one timepoint) and also how outcome predicts alliance in subsequent timepoints (with a lag of one timepoint).[3] The session-level effect represents the session-by-session association between the alliance and outcome, while the patient-level effect represents the association between the global strength of the alliance (i.e., across sessions) and mean levels of outcome. Consistent with the Wang and Maxwell approach (2015), we did not control for time effects in our data because time effects are considered an inherent part of the alliance–outcome correlation (i.e., alliance and outcome simultaneously improve during treatment; for a similar argument see Falkenström et al., 2017). We also included treatment type as a Level 2 predictor and its interactions with the within- and between-patient effects in order to examine group differences in these effects. Separate models were fitted to patient- and therapist-rated alliance scores.

Agreement and Congruence in Alliance Scores. To examine whether therapists' alliance shows agreement and congruence with patients' ratings, we adopted West and Kenny's (2011) Truth & Bias Model. We first centered *both* patient- and therapist-rated alliance scores around the individual *patient-rated alliance*. Then, we fitted a two-level model (sessions repeated within patients) in which therapist-rated alliance was predicted from patient-rated alliance. The estimated intercept represents the directional disagreement, with negative values suggesting lower therapists' ratings compared to patients' ratings. The estimated slope coefficient of patient-rated alliance represents the degree of temporal congruence between patient and therapist ratings over time. To examine differences between treatments in agreement and congruence, we also included treatment type as a Level 2 predictor and its interactions with agreement/congruence scores.

To examine session- and patient-level agreement as predictors of outcome, we fitted multilevel polynomial regression models (Kivlighan, Kline, Gelso, & Hill, 2017; Shanock, Baran, Gentry, Pattison, & Heggestad, 2010)

---

2. The OQ-45 measures symptoms during the past week, including the day of the assessment, while the WAI measures current alliance. Thus, it would have been preferable to predict symptoms from alliance in the previous session to test for "simultaneous prediction." However, the OQ-45 was administered only at Sessions 1, 5, and 10, which means that the Session 1 OQ-45 would not have a WAI to match, thus removing a third of the observations (and even two thirds in cross-lagged models) and would have significantly underpowered the analyses. Given that the OQ-45 also takes into account symptom severity during the day of the assessment, we felt that using the WAI from the same session was preferable in the current study.

3. We also examined whether alliance ratings predict weekly self-esteem as measured by the self-esteem subscale of the Bern Post-Session Report. We focused on this specific subscale because it was the only subscale to show significant change during treatment and it was the focus of the original study (Kramer et al., 2014a). The results showed that therapist- and patient-rated alliance predicted same-session self-esteem in both groups. No cross-lagged effects emerged. These findings are not reported in full here due to self-esteem being of secondary interest and due to space limitations. Full details about these analyses are available upon request from the first author.
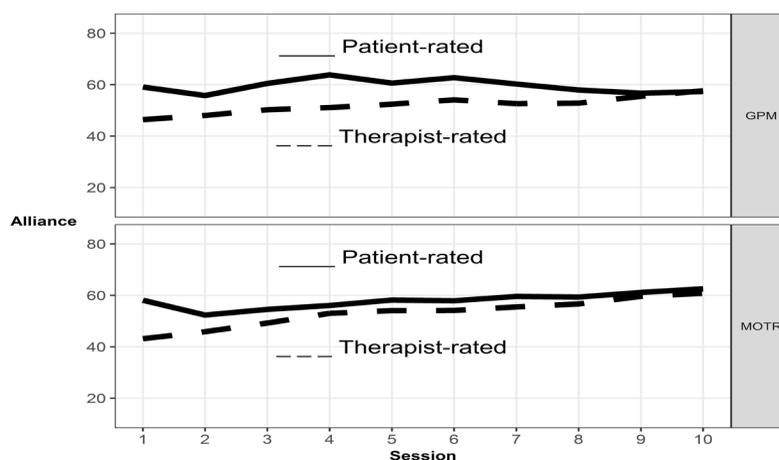
FIGURE 1. Change in alliance ratings during Good Psychiatric Management, brief-version (GPM-BV; top panel) and GPM combined with Motive-Oriented Therapeutic Relationship techniques (MOTR+GPM-BV; bottom panel). Solid lines represent patient-rated alliance scores, and dashed lines represent therapist-rated alliance scores. A significant increase in patient-rated alliance was observed in MOTR+GPM-BV but not in GPM-BV. A significant increase in therapist-rated alliance was observed in both treatments.

separately for each group, with OQ-45 as the dependent variable and the following predictors at each level (session and patients): WAI-P, WAI-T, WAI-P^2, WAI-P×WAI-T, and WAI-T^2. All WAI variables were centered to the scale midpoint before inclusion in the model. We then used the estimates to calculate slopes and curvatures along the lines of agreement and disagreement and to plot response surfaces. The following estimates were examined: The slope along the line of agreement tests whether symptom severity is lower when therapist and patient agree that the alliance is strong compared to when they agree that it is weak. The curvature along the line of agreement tests whether the effect of agreement changes as a function of alliance strength (e.g., if agreement is more important for weak alliances than for strong ones). The curvature along the line of disagreement tests whether disagreement between patient and therapist predicts higher symptom severity. The slope along the line of disagreement tests the direction of the effect of disagreement (e.g., whether symptom severity is lower when patient-rated alliance is higher than therapist-rated alliance, or the other way around).

To examine patient-level temporal congruence as a predictor of outcome, we computed a correlation between patient- and therapist-rated alliance across sessions separately for each patient. We then used those correlations (Fisher's z-transformed and grand-mean centered) as Level 2 predictors of OQ-45. To examine differences between treatments in the effect of congruence on symptom severity, we also included treatment type as a Level 2 predictor and its interactions with congruence scores.

Throughout the article, effect sizes are calculated from t values and dfs from multilevel models (cf. Crits-Christoph et al., 2009) and are presented in Cohen's d metric.

RESULTS
CHANGES IN ALLIANCE DURING TREATMENT

Figure 1 presents changes in patient- and therapist-rated alliance during MOTR+GPM-BV and GPM-BV. As the figure shows, therapist-rated alliance at baseline could be characterized as weak-moderate (estimated means: MOTR+GPM-BV = 45.57; GPM-BV = 47.25) and did not differ by group (t_{225} = .90, p = .35, d = .12). In addition, therapist-rated alliance increased during both MOTR+GPM-BV (b = 1.88, t_{73} = 4.61, p < .01, d = 1.08) and GPM-BV (b = 1.04, t_{180} = 2.78, p < .01, d = .41), without group differences in rates of increase (t_{99} = 1.50, p = .14, d = .30).

As Figure 1 shows, patient-rated alliance at baseline could be characterized as moderate (estimated means: MOTR+GPM-BV = 53.38; GPM-BV = 58.75) and did not differ by group (t_{452} = 1.54, p = .12, d = .14). In addition, while patient-rated alliance increased during MOTR+GPM-BV (b = .76, t_{88} = 2.09, p = .04, d = .45), it did not change during GPM-BV (b = .24, t_{110} = .65, p = .52, d = .12), although the group difference did not reach significance (t_{94} = .99, p = .33, d = .20).

SESSION- AND PATIENT-LEVEL ALLIANCE TO PREDICT TREATMENT OUTCOME (HYPOTHESIS 1)

Estimated multilevel coefficients for the prediction of symptomatic distress from alliance ratings are presented in Table 1.

Prediction of Outcome in the Same Timepoint From Therapist-Rated Alliance. As Table 1 shows, we found that, consistent with hypotheses, session-level alliance predicted symptomatic distress in the same timepoint in MOTR+GPM-BV (t_{70} = –2.81, p < .01, d = .67), but not in GPM-BV (t_{49} = –1.41, p = .16, d = .40), although the difference between groups did not reach significance (t_{58} = –.36, p = .72, d = .09). Thus, stronger therapist-rated alliance in a given timepoint predicted lower symptomatic distress in the same timepoint among patients receiving MOTR+GPM-BV but not among patients receiving GPM-BV. As we hypothesized, patient-level alliance did not predict symptomatic distress in MOTR+GPM-BV (t_{83} = –1.20, p = .23, d = .26) or in GPM-BV (t_{100} = –.75, p = .45, d = .15). Therefore, the mean therapist-rated alliance for a given patient was not associated with symptomatic distress.

Prediction of Outcome in the Same Timepoint From Patient-Rated Alliance. As Table 1 shows, contrary to our hypotheses, session-level alliance did not predict symptomatic distress in the same timepoint in MOTR+GPM-BV (t_{44} = –1.15, p = .25, d = .35) or in GPM-BV (t_{56} = –.50, p = .62, d = .13). In addition, as we hypothesized, patient-level alliance did not predict symptomatic distress in MOTR+GPM-BV (t_{107} = –1.27, p = .25, d = .24) or in GPM-BV (t_{90} = –1.18, p = .27, d = .25).

TABLE 1. Estimated Multilevel Regression Coefficients and 95% Confidence Intervals for the Simultaneous and Cross-Lagged Prediction of Symptoms From Alliance and Cross-Lagged Prediction of Alliance From Symptoms

Model	Therapist-Rated Alliance (Model a)		Patient-Rated Alliance (Model b)	
	MOTR+GPM-BV	GPM-BV	MOTR+GPM-BV	GPM-BV
Alliance predicting symptoms (simultaneous)				
Session-level alliance	−.67 [−1.15, −.20]**	−.52 [−1.25, .22]	−.50 [−1.37, .37]	−.20 [−1.01, .60]
Patient-level alliance	−.50 [−1.33, .33]	−.35 [−1.27, .57]	−.42 [−1.07, .24]	−.39 [−1.03, .26]
Alliance predicting symptoms (cross-lagged)				
Session-level alliance	−.47 [−1.26, .32]	.02 [−1.11, 1.16]	−.01 [−1.31, 1.30]	.02 [−1.31, 1.35]
Patient-level alliance	−.42 [−1.34, .49]	−.31 [−1.37, .75]	−.46 [−1.18, .26]	−.56 [−1.31, .19]
Symptoms predicting alliance (cross-lagged)				
Session-level symptoms	−.28 [−.53, −.04]*	−.07 [−.27, .13]	−.02 [−.23, .20]	.00 [−.22, .23]
Patient-level symptoms	−.13 [−.32, .06]	−.12 [−.33, .09]	−.10 [−.33, .12]	−.10 [−.34, .14]

Note. MOTR = Motive-Oriented Therapeutic Relationship; GPM-BV = Good Psychiatric Management, brief-version. Symptomatic distress was measured by the Outcome Questionnaire–45.2; Therapist- and patient-rated alliance ratings were analyzed in separate models. */** significantly different from zero at the $p < .05/.01$ level.

Cross-Lagged Relationships Between Outcome and Therapist-Rated Alliance. Contrary to our hypotheses, as shown in Table 1, session-level alliance did not predict symptomatic distress in subsequent timepoints in MOTR+GPM-BV (t_{30} = −1.21, p = .24, d = .44) or in GPM-BV (t_{25} = .04, p = .97, d = .02). However, session-level symptomatic distress predicted therapist-rated alliance in subsequent timepoints in MOTR+GPM-BV (t_{22} = −2.41, p = .02, d = 1.03) but not in GPM-BV (t_{34} = −.67, p = .51, d = .23). Thus, only in MOTR+GPM-BV did lower symptomatic distress in a given timepoint predict stronger therapist-rated alliance at subsequent timepoints.

Prediction of Outcome in Subsequent Timepoints From Patient-Rated Alliance. As shown in Table 1, contrary to our hypotheses, session-level alliance did not predict symptomatic distress in subsequent timepoints in MOTR+GPM-BV (t_{27} = −.01, p = .99, d = .00) or in GPM-BV (t_{26} = .03, p = .97, d = .01). In addition, session-level symptomatic distress did not predict patient-rated alliance in subsequent timepoints in MOTR+GPM-BV (t_{29} = −.16, p = .88, d = .06) or in GPM-BV (t_{26} = .02, p = .99, d = .01)

AGREEMENT AND CONGRUENCE IN ALLIANCE SCORES AND THEIR ASSOCIATION WITH TREATMENT OUTCOME (HYPOTHESES 2 AND 3)

Agreement. Estimated mean directional disagreement in MOTR+GPM-BV and GPM-BV and its association with treatment outcome are presented in Table 2. As the table shows, consistent with our hypotheses, we found that GPM-BV therapists had a negative disagreement with their patients (t_{394} = −2.92, p < .01, d = .29), while MOTR+GPM-BV therapists did not (t_{455} = −1.56, p = .12, d = .15), even though the difference in agreement between treatments did not reach significance (t_{395} = 1.08, p = .28, d = .11). Thus, GPM-BV therapists rated the alliance as consistently lower than their patients, while MOTR+GPM-BV therapists did not.

TABLE 2. Estimated Mean Levels of Disagreement and Temporal Congruence Scores by Treatment Type From Truth and Bias Analysis (Top Panel) and Multilevel Polynomial Regression Coefficients for the Association of Session- and Patient-Level Disagreement Symptomatic Distress (Bottom Panel)

	MOTR+GPM-BV	GPM-BV
Truth and bias model		
Mean disagreement	−3.96 [−8.95, 1.03]	−8.06 [−13.49, −2.63]**
Mean level of temporal congruence	.25 [.04, .45]*	.26 [.03, .49]*
Multilevel polynomial regression analysis		
Session-level estimates		
Slope along line of agreement	−3.39 [−6.92, .14]†	−1.29 [−3.40, .81]
Curvature along line agreement	.10 [−.03, .23]	.05 [−.05, .15]
Slope along line of disagreement	−.30 [−3.03, 2.42]	.61 [−1.39, 2.61]
Curvature along line disagreement	.10 [−.05, .25]	.05 [−.06, .16]
Patient-level estimates		
Slope along line of agreement	1.33 [−3.45, 6.11]	.30. [−2.33, 2.93]
Curvature along line of agreement	−.02 [−.24, .20]	.13 [−.03, 0.29]
Slope along line of disagreement	.05 [−3.02, 3.12]	−.73 [−3.73, 2.26]
Curvature along line disagreement	−.03 [−.27, .20]	.14 [−.09, .36]

Note. MOTR = Motive-Oriented Therapeutic Relationship; GPM-BV = Good Psychiatric Management, brief-version. Numbers in brackets represent 95% confidence intervals. Negative disagreement scores indicate that therapist-rated alliance ratings were lower than those of their patients. Positive temporal congruence indicates positive association between therapist- and patient-rated alliance over time. †/*/** significantly different from zero at the $p < .06/.05/.01$ level.

Results of the multilevel polynomial regression analysis to examine the association between agreement and symptomatic distress are presented in Table 2, and the respective response surfaces are plotted in Figure 2. The results showed a trend for session-level agreement to predict symptomatic distress in MOTR+GPM-BV ($t_{90} = -1.91$, $p = .059$) but not in GPM-BV ($t_{78} = -1.22$, $p = .22$). Thus, only in MOTR+GPM-BV, in sessions in which therapists and patients agreed more on a strong alliance, did patients also tend to report lower symptomatic distress. The curvilinear effect of agreement was not significant in either MOTR+GPM-BV ($t_{90} = 1.55$, $p = .12$) or GPM-BV ($t_{78} = .96$, $p = .34$), which means that the effect of agreement was the same regardless of how strong the alliance was. This effect is also evident in Figure 2, which shows that in MOTR+GPM-BV at the session level, the slope above the line of perfect agreement (WAI-T = WAI-P) is linear and negative. In other words, in MOTR+GPM-BV, greater session-level agreement on strong alliance tends to predict lower session-level symptomatic distress regardless of whether the alliance is high or low in that session. None of the other measures of agreement and disagreement had significant effects on symptomatic distress (ps = .11–.97).

Temporal Congruence. Estimated mean temporal congruence in MOTR+GPM-BV and GPM-BV is presented in Table 2. As the table shows, we found that positive congruence existed in both MOTR+GPM-BV ($t_{205} = 2.35$, $p = .02$, $d = .33$) and GPM-BV ($t_{218} = 2.23$, $p = .03$, $d = .30$). However, contrary to our hypotheses, there were no differences between treatments in the degree of congruence ($t_{232} = -.09$, $p = .93$, $d = .01$). Therefore, in sessions

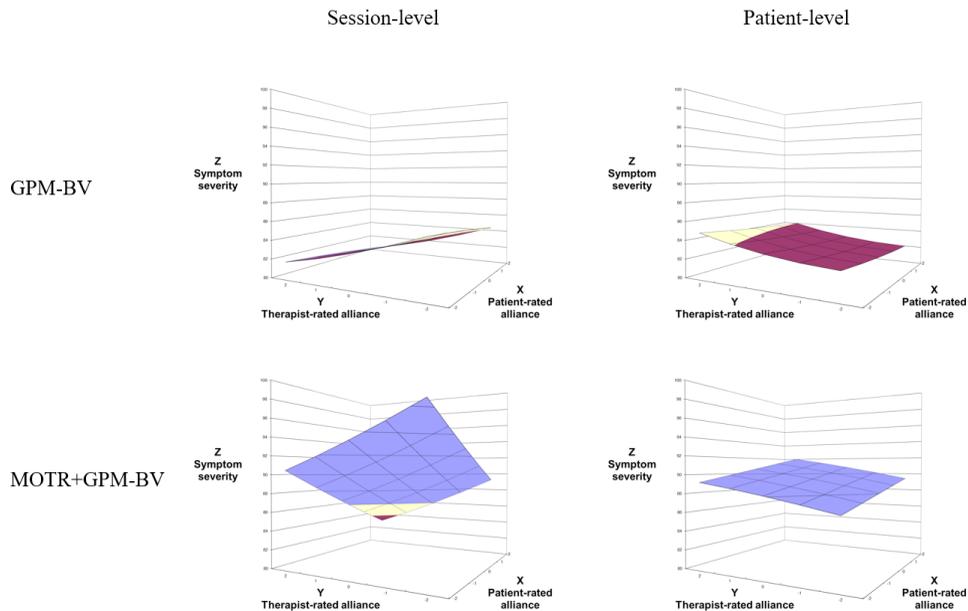Session-level   Patient-level

FIGURE 2. Symptom severity as a function of therapist- and patient-rated alliance during Good Psychiatric Management, brief-version (GPM-BV; top panels) and GPM combined with Motive-Oriented Therapeutic Relationship techniques (MOTR+GPM-BV; bottom panels) at the session- (left panels) and patient-levels (right panels).

in which therapists rated the alliance as relatively strong, their patients were also likely to rate the alliance as strong, regardless of treatment type.

Looking at how patient-level congruence predicts symptomatic distress, we found that congruence did not predict symptomatic distress in MOTR+GPM-BV (b = .30, t_{108} = .12, p = .91, d = .02) or in GPM-BV (b = 1.44, t_{109} = .89, p = .38, d = .17), with no differences between treatments (b = –1.14, t_{109} = -.38, p = .70, d = .07). Therefore, patient-level temporal covariation in alliance between therapist and patient did not predict symptomatic distress.

DISCUSSION

The present study examined the role that the dynamics and synchrony of patient and therapist ratings of the alliance play in the effect of adding MOTR to GPM-BV for BPD. We found that increases in patient-rated alliance occurred in MOTR+GPM-BV but not in GPM-BV, while therapist-rated alliance increased in both groups. Furthermore, stronger therapist-rated alliance predicted lower symptomatic distress in the same timepoint in MOTR+GPM-BV but not GPM-BV. In addition, only in MOTR+GPM-BV did lower symptomatic distress predict subsequent stronger therapist-rated alliance, but not the other way around.

When examining agreement and congruence in alliance ratings, we found that in GPM-BV therapists' ratings of the alliance were lower than pa-

tients' ratings, while in MOTR+GPM-BV this disagreement was mitigated. Importantly, higher agreement predicted lower symptomatic distress in the same timepoint in MOTR+GPM-BV but not in GPM-BV. Congruence over time between patient- and therapist-rated alliance existed in both groups, but was not related to outcome.

CHANGE IN ALLIANCE DURING THERAPY

Across treatments and raters, we found that the alliance tended to be weak-moderate at baseline, which is consistent with clinical writings on difficulties in forming a strong alliance with BPD patients (Gabbard et al., 1988). However, we also found that patient- and therapist-rated alliance tended to increase during MOTR+GPM-BV, while only therapist-rated alliance increased during GPM-BV. This suggests that despite the difficulties, it is possible to strengthen the alliance with BPD patients, even in short-term treatments. This also suggests that individualizing treatment to the relationship needs of the patient results in a stronger alliance on the part of the patient. The fact that our analyses resulted in different results from those reported by Kramer, Kolly, et al. (2014) could be attributed to the high percentage of missing alliance data, especially patient-rated alliance data (188 missing patient-rated observations vs. 149 missing therapist-rated observations). This finding emphasizes the importance of utilizing imputation methods when appropriate (Enders et al., 2016).

In terms of outcome prediction, consistent with the existing literature (reviewed in Zilcha-Mano, 2017), we found that session-level changes in the alliance were predictive of outcome in MOTR+GPM-BV but not in GPM-BV. These findings suggest that the alliance is a dynamic construct that evolves during therapy, and that these dynamics are especially predictive of outcome. In addition, the stronger prediction found in MOTR+GPM-BV emphasizes the important role that the alliance may play in individualized and relationship-based psychotherapies, and it may suggest a specific mediation that occurs in that group. Thus, consistent with other approaches to training therapists in strengthening the alliance (Crits-Christoph et al., 2006; Eubanks-Carter, Muran, & Safran, 2015), our findings suggest that tailoring the therapeutic relationship to the individual patient mitigates the attenuation of the alliance–outcome correlation that has been observed in psychotherapies for BPD (Levy et al., 2017). However, it should be noted that in contrast to some other studies (reviewed in Zilcha-Mano, 2017), the alliance did not predict outcome in a lag, but rather the opposite: In MOTR+GPM-BV, better outcome predicted stronger alliance in a lag. Thus, although something about the additional component of MOTR caused therapist-rated alliance to be correlated with outcome, it seems like changes in outcome precede changes in alliance and not the other way around. However, the short duration of the treatments described here does not allow us to arrive at conclusions regarding the longer treatment durations that are more typical for patients with BPD. For example, it could still be that early in therapy, changes in alliance follow symptom change, and only in later phases, when changes in interpersonal difficulties occur, does the alliance begin to drive symptom

change (cf. Zilcha-Mano, 2017). Thus, these possible explanations should be further examined in long-term treatments in order to fully understand the temporal unfolding of the alliance–outcome correlation in this specific population.

AGREEMENT AND CONGRUENCE IN ALLIANCE SCORES

We found that therapist-rated alliance was lower than patient-rated alliance in GPM-BV, which is consistent with other published studies on the topic (Atzil-Slonim et al., 2015; Compare et al., 2016; Zilcha-Mano et al., 2017). The lower therapist- compared to patient-rated alliance is also consistent with a meta-analysis of cross-sectional studies that reported that the difference between patient- and therapist-rated alliance tended to be higher in studies of severely disturbed patients, including patients with BPD (Tryon et al., 2007). However, the addition of MOTR eliminated this disagreement, which suggests that individualizing the therapeutic relationship for the needs of the patient decreases disagreement between the patient's and the therapist's experience of the therapeutic alliance. Importantly, in MOTR+GPM-BV (but not in GPM-BV), symptomatic distress tended to be lower in timepoints in which patient and therapist agreed that the alliance is strong. This finding adds to the existing findings about other potential mediators of the effect of MOTR+GPM-BV (i.e., change in behavioral coping; Kramer et al., 2017) and suggests that MOTR+GPM-BV exerts its effects partly by increasing agreement. Importantly, to our knowledge, the current study is the first to decompose session- and patient-level contributions of agreement to the prediction of symptomatic distress. The fact that only session-level, but not patient-level, agreement tended to predict lower symptomatic distress speaks to the importance of examining the dynamics of the alliance, rather than viewing it as a stable patient characteristic. However, the effect was marginally significant and the actual temporal order of it is still unknown, so more research is needed before conclusions can be reached.

We found a significant temporal congruence between patient- and therapist-rated alliance in both treatments. This means that therapists' and patients' ratings tended to covary together over time, such that therapists and patients showed similar fluctuations in alliance throughout treatment. This finding is consistent with the findings of Atzil-Slonim et al. (2015) but not Compare et al. (2016). Given that Compare et al. (2016) examined group psychotherapies, it is possible that the alliance to multiple group members makes it difficult for therapists to reach congruence with the alliance of a specific patient. However, in contrast to Atzil-Slonim et al. (2015), in our study congruence did not predict outcome, which may suggest that, at least with BPD patients, agreement (lower mean differences) is more important than temporal congruence.

STUDY LIMITATIONS

One of the strengths of the present study is the inclusion of both patient- and therapist-rated measures of the alliance. However, including an observer-

rated measure of the alliance could have complemented our findings as well as allowing for a more objective evaluation of the alliance. Related to that, the outcome measures in the study included only patient reports, and not therapist- or clinician-rated measures. Therapist-rated measures of symptomatic distress would have enabled us to perform a full two-person perspective analysis examining the independent and interactive effects of both raters (Zilcha-Mano et al., 2016).

Another limitation to consider is the relatively short duration of the psychotherapies used in the study (10 sessions). The length of treatment prevented us from examining the unfolding of the alliance during a longer period of time, although studies do show that the alliance predicts outcome, irrespective of when it is assessed (Horvath et al., 2011).

Last, the moderate sample size is another limitation that has likely underpowered some of our analyses. This is especially true regarding group-level comparisons, in which our analyses were mostly powered to detect large effects, and regarding analyses involving the OQ-45, which was measured in only three timepoints.

CLINICAL AND RESEARCH IMPLICATIONS AND CONCLUDING COMMENTS

Clinically, our findings point to the potential benefits of individualizing the therapeutic relationship to match the specific needs of the patient. In addition, as more similar findings accrue, the findings may also suggest the importance of monitoring patient and therapist alliance and examining possible sources of discrepancy between the two.

In terms of research implications, the study demonstrates the importance of studying the dynamics of the alliance in a nuanced way, including session-level alliance as well as agreement and congruence between patient- and therapist-reports. Special attention should be given to cross-lagged effects that may shed light on the causal direction between alliance and outcome, as well as the appropriate time lag of these associations. These cutting-edge examinations require more intensive data collection than is typically done in psychotherapy research, as well as collection of alliance and outcome measures from multiple raters (e.g., patient and therapist).

In conclusion, our study examined the effect of adding an ingredient of alliance individualization to a brief -treatment of patients with BPD, a severe, difficult-to-treat population. The findings show that such individualization does not necessarily result in stronger patient-rated alliance, but rather in higher agreement between patients' and therapists' experience of the alliance and in stronger ties with outcome. Thus, individualizing psychotherapy may exert part of its salutary effect by increasing the agreement between patients and therapists.

REFERENCES

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

Atzil-Slonim, D., Bar-Kalifa, E., Rafaeli, E., Lutz, W., Rubel, J., Schiefele, A.-K., & Peri, T. (2015). Therapeutic bond judgments: Congruence and incongruence. *Journal of Consulting and Clinical Psychology*, *83*, 773–784.

Bateman, A., & Fonagy, P. (1999). Effectiveness of partial hospitalization in the treatment of borderline personality disorder: A randomized controlled trial. *American Journal of Psychiatry*, *156*, 1563–1569.

Caspar, F. (2007). Plan analysis. In T. D. Eels (Ed.), *Handbook of psychotherapy case formulation* (2nd ed., pp. 251–289). New York, NY: Guilford Press.

Clarkin, J. F., Levy, K. N., Lenzenweger, M. F., & Kernberg, O. F. (2007). Evaluating three treatments for borderline personality disorder: A multiwave study. *American Journal of Psychiatry*, *164*, 922–928.

Compare, A., Tasca, G. A., Lo Coco, G., & Kivlighan, D. M., Jr. (2016). Congruence of group therapist and group member alliance judgments in emotionally focused group therapy for binge eating disorder. *Psychotherapy*, *53*, 163–173.

Corbière, M., Bisson, J., Lauzon, S., & Ricard, N. (2006). Factorial validation of a French short-form of the Working Alliance Inventory. *International Journal of Methods in Psychiatric Research*, *15*(1), 36–45.

Crits-Christoph, P., Gallop, R., Temes, C. M., Woody, G., Ball, S. A., Martino, S., & Carroll, K. M. (2009). The alliance in motivational enhancement therapy and counseling as usual for substance use problems. *Journal of Consulting and Clinical Psychology*, *77*, 1125–1135.

Crits-Christoph, P., Gibbons, M. B. C., Crits-Christoph, K., Narducci, J., Schamberger, M., & Gallop, R. (2006). Can therapists be trained to improve their alliances? A preliminary study of alliance-fostering psychotherapy. *Psychotherapy Research*, *16*, 268–281.

Enders, C. K., Mistler, S. A., & Keller, B. T. (2016). Multilevel multiple imputation: A review and evaluation of joint modeling and chained equations imputation. *Psychological Methods*, *21*, 222–240.

Eubanks-Carter, C., Muran, J. C., & Safran, J. D. (2015). Alliance-focused training. *Psychotherapy*, *52*, 169–173.

Falkenström, F., Finkel, S., Sandell, R., Rubel, J. A., & Holmqvist, R. (2017). Dynamic models of individual change in psychotherapy process research. *Journal of Consulting and Clinical Psychology*, *85*, 537–549.

Gabbard, G. O., Horwitz, L., Frieswyk, S., Allen, J. G., Colson, D. B., Newsom, G., & Coyne, L. (1988). The effect of therapist interventions on the therapeutic alliance with borderline patients. *Journal of the American Psychoanalytic Association*, *36*, 697–727.

Grawe, K. (1980). Die diagnostisch-therapeutische Funktion der Gruppeninteraktion in verhaltenstherapeutischen Gruppen [The diagnostic-therapeutic function of group interaction in behavior therapy groups]. In K. Grawe (Ed.), *Verhaltenstherapie in Gruppen* (pp. 88–232). Munich, Germany: Urban & Schwarzenberg.

Gunderson, J. G., & Links, P. S. (2008). *Borderline personality disorder: A clinical guide*. Washington, DC: American Psychiatric Publishing.

Horvath, A. O., Del Re, A. C., Flückiger, C., & Symonds, D. (2011). Alliance in individual psychotherapy. *Psychotherapy*, *48*(1), 9–16.

Kivlighan, D. M., Jr., Kline, K., Gelso, C. J., & Hill, C. E. (2017). Congruence and discrepancy between working alliance and real relationship: Variance decomposition and response surface analyses. *Journal of Counseling Psychology, 64*, 394–409.

Kolly, S., Kramer, U., Herrera, F., Follonier, G., Maksutaj, R., Schopfer, S., … Preisig, M. (2010). *Manuel du programme trouble de la personnalité: investigation psychiatrique et psychodynamique* [Manual for the personality disorder program: Psychiatric and psychodynamic investigation]. Unpublished manuscript, University of Lausanne, Lausanne, Switzerland.

Kramer, U., Flückiger, C., Kolly, S., Caspar, F., Marquet, P., Despland, J.-N., & De Roten, Y. (2014). Unpacking the effects of therapist responsiveness in borderline personality disorder: Motive-oriented therapeutic relationship, patient in-session experience, and the therapeutic alliance. *Psychotherapy and Psychosomatics*, *83*, 386–387.

Kramer, U., Keller, S., Caspar, F., de Roten, Y., Despland, J.-N., & Kolly, S. (2017). Early change in coping strategies in responsive treatments for borderline personality disorder: A mediation analysis. *Journal of Consulting and Clinical Psychology, 85*, 530–535.

Kramer, U., Kolly, S., Berthoud, L., Keller, S., Preisig, M., Caspar, F.,...Despland, J.-N. (2014). Effects of motive-oriented therapeutic relationship in a ten-session general

psychiatric treatment of borderline personality disorder: A randomized controlled trial. *Psychotherapy and Psychosomatics*, *83*, 176–186.

Kramer, U., & Stiles, W. B. (2015). The responsiveness problem in psychotherapy: A review of proposed solutions. *Clinical Psychology: Science and Practice, 22*, 277–295.

Lambert, M. J., Morton, J. J., Hatfield, D., Harmon, C., Hamilton, S., Reid, R. C.,...Burlingame, G. M. (2004). *Administration and scoring manual for the Outcome Questionnaire-45*. Orem, UT: American Professional Credentialing Services.

Levy, K. N., Beeney, J. E., Wasserman, R. H., & Clarkin, J. F. (2010). Conflict begets conflict: Executive control, mental state vacillations, and the therapeutic alliance in treatment of borderline personality disorder. *Psychotherapy Research*, *20*, 413–422.

Levy, K. N., Scala, J. W., & Ellison, W. D. (2017). *The therapeutic alliance in the treatment of borderline personality disorder: A meta-analysis*. Manuscript in preparation.

Linehan, M. M., Armstrong, H. E., Suarez, A., Allmon, D., & Heard, H. L. (1991). Cognitive-behavioral treatment of chronically parasuicidal borderline patients. *Archives of General Psychiatry*, *48*, 1060–1064.

Marmarosh, C. L., & Kivlighan, D. M. (2012). Relationships among client and counselor agreement about the working alliance, session evaluations, and change in client symptoms using response surface analysis. *Journal of Counseling Psychology, 59*, 352–367.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2016). nlme: Linear and Nonlinear Mixed Effects Models, R package version 3.1-128 [Computer software]. Retrieved from http://CRAN.R-project.org/package=nlme

R Core Team. (2016). R: A language and environment for statistical computing [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Shanock, L. R., Baran, B. E., Gentry, W. A., Pattison, S. C., & Heggestad, E. D. (2010). Polynomial regression with response surface analysis: A powerful approach for examining moderation and overcoming limitations of difference scores. *Journal of Business and Psychology, 25*, 543–554.

Shrout, P. E., & Lane, S. P. (2012). Psychometrics. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 302–320). New York, NY: Guilford Press.

Skodol, A. E., Gunderson, J. G., Pfohl, B., Widiger, T. A., Livesley, W. J., & Siever, L. J. (2002). The borderline diagnosis I: Psychopathology, comorbidity, and personality structure. *Biological Psychiatry*, *51*, 936–950.

Tracey, T. J., & Kokotovic, A. M. (1989). Factor structure of the Working Alliance Inventory. *Psychological Assessment*, *1*, 207–210.

Tryon, G. S., Blackwell, S. C., & Hammel, E. F. (2007). A meta-analytic examination of client–therapist perspectives of the working alliance. *Psychotherapy Research, 17*, 629–642.

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*(3), 1–67.

Wang, L. P., & Maxwell, S. E. (2015). On disaggregating between-person and within-person effects with longitudinal data using multilevel models. *Psychological Methods*, *20*, 63–83.

West, T. V., & Kenny, D. A. (2011). The truth and bias model of judgment. *Psychological Review*, *118*, 357–378.

Yeomans, F. E., Gutfreund, J., Selzer, M. A., Clarkin, J. F., Hull, J. W., & Smith, T. E. (1994). Factors related to drop-outs by borderline patients: Treatment contract and therapeutic alliance. *Journal of Psychotherapy Practice and Research*, *3*(1), 16–24.

Zilcha-Mano, S. (2017). Is alliance really therapeutic? Revisiting this question in light of recent methodological advances. *American Psychologist*, *72*, 311–325.

Zilcha-Mano, S., Muran, J. C., Hungr, C., Eubanks, C. F., Safran, J. D., & Winston, A. (2016). The relationship between alliance and outcome: Analysis of a two-person perspective on alliance and session outcome. *Journal of Consulting and Clinical Psychology*, *84*, 484–496.

Zilcha-Mano, S., Snyder, J., & Silberschatz, G. (2017). The effect of congruence in patient and therapist alliance on patient's symptomatic levels. *Psychotherapy Research, 27*, 371–380.

APPENDIX: MODELS FITTED IN THE CURRENT STUDY

CHANGES IN ALLIANCE DURING TREATMENT

Level-1 Model:

$$WAI_{ti} = \pi_{0i} + \pi_{1i} \times Session\_Number_{ti} + e_{ti}$$

$$e_{ti} \sim N(0,\sigma^2)$$
$$e_{tt'i} \sim N(0,\sigma^2 \times \rho^{|t-t'|})$$

Level-2 Model:

$$\pi_{0i} = \beta_{00} + \beta_{01} \times Group_i + r_{0i}$$

$$\pi_{1i} = \beta_{10} + \beta_{11} \times Group_i + r_{1i}$$

$$\begin{pmatrix} r_{0i} \\ r_{1i} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & 0 \\ 0 & \tau_{11} \end{pmatrix} \right]$$

Where WAI = Working Alliance Inventory–Short Form, Therapist/Patient version; Session_number = session number in therapy, centered around the first session; $Group$ = GPM-BV/MOTR+GPM-BV.

PATIENT- AND SESSION-LEVEL ALLIANCE AS PREDICTORS OF SYMPTOMATIC DISTRESS

Level-1 Model:

$$OQ-45_{ti} = \pi_{0i} + \pi_{1i} \times (WAI_{ti} - \overline{WAI}_i) + e_{ti}$$

$$e_{ti} \sim N(0,\sigma^2)$$
$$e_{tt'i} \sim N(0,\sigma^2 \times \rho^{|t-t'|})$$

Level-2 Model:

$$\pi_{0i} = \beta_{00} + \beta_{01} \times Group_i + \beta_{02} \times (\overline{WAI}_i - \overline{WAI}_{..}) + \beta_{03} \times Group_i \times (\overline{WAI}_i - \overline{WAI}_{..}) + r_{0i}$$

$$\pi_{1i} = \beta_{10} + \beta_{11} \times Group_i + r_{1i}$$

$$\begin{pmatrix} r_{0i} \\ r_{1i} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & 0 \\ 0 & \tau_{11} \end{pmatrix} \right]$$

Where $OQ\text{-}45$ = Outcome Questionnaire–45.2; WAI = Working Alliance Inventory–Short Form, Therapist/Patient version; $Group$ = GPM-BV/MOTR+GPM-BV.

TRUTH AND BIAS MODEL OF ALLIANCE RATINGS

Level-1 Model:

$$(WAI\text{-}T_{ti} - \overline{WAI\text{-}P_i}) = \pi_{0i} + \pi_{1i} \times (WAI\text{-}P_{ti} - \overline{WAI\text{-}P_i}) + e_{ti}$$

$$e_{ti} \sim N(0, \sigma^2)$$
$$e_{tt'i} \sim N(0, \sigma^2 \times \rho^{|t-t'|})$$

Level-2 Model:

$$\pi_{0i} = \beta_{00} + \beta_{01} \times Group_i + r_{0i}$$

$$\pi_{1i} = \beta_{10} + \beta_{11} \times Group_i + r_{1i}$$

$$\begin{pmatrix} r_{0i} \\ r_{1i} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & 0 \\ 0 & \tau_{11} \end{pmatrix} \right]$$

Where *WAI-T* = Working Alliance Inventory–Short Form, Therapist version; *WAI-P* = Working Alliance Inventory–Short Form, Patient version; *Group* = GPM-BV/MOTR+GPM-BV.

MULTILEVEL POLYNOMIAL REGRESSION (SEPARATE MODELS WERE FITTED FOR EACH GROUP)

Level-1 Model:

$$OQ-45_{ti} = \pi_{0i} + \pi_{1i} \times WAI\text{-}P_{ti} + \pi_{2i} \times WAI\text{-}T_{ti} + \pi_{3i} \times (WAI\text{-}P_{ti})^2 + \pi_{4i} \times WAI\text{-}P_{ti} \times WAI\text{-}T_{ti} + \pi_{5i} \times (WAI\text{-}T_{ti})^2 + e_{ti}$$

$$e_{ti} \sim N(0, \sigma^2)$$
$$e_{tt'i} \sim N(0, \sigma^2 \times \rho^{|t-t'|})$$

Level-2 Model:

$$\pi_{0i} = \beta_{00} + \beta_{01} \times \overline{WAI\text{-}P_i} + \beta_{02} \times \overline{WAI\text{-}T_i} + \beta_{03} \times (\overline{WAI\text{-}P_i})^2 + \beta_{04} \times \overline{WAI\text{-}P_i} \times \overline{WAI\text{-}T_i} + \beta_{05} \times (\overline{WAI\text{-}T_i})^2 + r_{0i}$$
$$\pi_{1i} = \beta_{10}$$

$$\pi_{2i} = \beta_{20}$$

$$\pi_{3i} = \beta_{30}$$

$$\pi_{4i} = \beta_{40}$$

$$\pi_{5i} = \beta_{50}$$

$$r_{0i} \sim N(0, \tau_{00})$$

Where *OQ-45* = Outcome Questionnaire–45.2; *WAI-T* = Working Alliance Inventory–Short Form, Therapist version, centered to scale midpoint; *WAI-P* = Working Alliance Inventory–Short Form, Patient version, centered to scale midpoint.

PATIENT-LEVEL TEMPORAL CONGRUENCE AS A PREDICTOR OF SYMPTOMATIC DISTRESS

Level-1 Model:

$$OQ-45_{ti} = \pi_{0i} + e_{ti}$$

$$e_{ti} \sim N(0, \sigma^2)$$
$$e_{tt'i} \sim N(0, \sigma^2 \times \rho^{|t-t'|})$$

Level-2 Model:

$$\pi_{0i} = \beta_{00} + \beta_{01} \times Group_i + \beta_{02} \times (Congruence_i - \overline{Congruence}) + \beta_{03} \times Group_i \times (Congruence_i - \overline{Congruence}) + r_{0i}$$

$$r_{0i} \sim N(0, \tau_{00})$$

Where *OQ-45* = Outcome Questionnaire–45.2; *Congruence* = Person-specific correlation between therapist- and patient-rated alliance across sessions, Fisher's *z* transformed; *Group* = GPM-BV/MOTR+GPM-BV.