

**Threat alert: The effect of outliers on the alliance-outcome correlation**

Simon B. Goldberg<sup>1,2</sup>, Robbie Babins-Wagner<sup>3</sup>, Zac E. Imel<sup>4</sup>, Derek D. Caperton<sup>3</sup>, Lauren Weitzman<sup>5</sup>, & Bruce E. Wampold<sup>1,6</sup>

<sup>1</sup>Department of Counseling Psychology, University of Wisconsin – Madison, Madison, WI, USA

<sup>2</sup>Center for Healthy Minds, University of Wisconsin – Madison, Madison, WI, USA

<sup>3</sup>Calgary Counselling Centre, Calgary, Alberta, Canada

<sup>4</sup>Department of Educational Psychology, University of Utah, Salt Lake City, UT, USA

<sup>5</sup>University of Utah Counseling Center, Salt Lake City, UT, USA

<sup>6</sup>Skillsetter.com

**Author Note**

The authors have no conflicts of interest to disclose. SBG was supported by grant K23AT010879 from the National Center for Complementary and Integrative Health of the National Institutes of Health. We are unable to share patient data, but code used in analyses is available by request. This study was not preregistered.

Correspondence should be addressed to: Simon B. Goldberg, Department of Counseling Psychology, University of Wisconsin-Madison, 335 Education Building, 1000 Bascom Mall, Madison, Wisconsin, 53706, United States, phone: 608-265-8986, fax: 608-265-4174, [sbgoldberg@wisc.edu](mailto:sbgoldberg@wisc.edu)

Recommended citation: Goldberg, S. B., Babins-Wagner, R., Imel, Z. E., Caperton, D. D., Weitzman, L., & Wampold, B. E. (in press). Threat alert: The effect of outliers on the alliance-outcome correlation. *Journal of Counseling Psychology*.



### Abstract

Meta-analyses have established the alliance as the most robust predictor of outcome in psychotherapy. A growing number of studies have evaluated potential threats to the conclusion that alliance is a *causal* factor in psychotherapy. One potential threat that has not been systematically examined is the possibility that the alliance-outcome association is driven by low alliance outliers. We examined the influence of removing low alliance outliers on the alliance-outcome association using data drawn from two largescale, naturalistic psychotherapy data sets ( $Ns = 1,052; 11,029$ ). These data sets differed in setting (university counseling center, community mental health center), country (United States and Canada), alliance measure (4-item Working Alliance Inventory Short Form Revised, 10-item Session Rating Scale), and outcome measure (Counseling Center Assessment of Psychological Symptoms-34, Outcome Questionnaire – 45). We examined the impact of treating outliers in five different ways: retaining them, removing values three or two standard deviations from the mean, and winsorizing values three or two standard deviations from the mean. We also examined the effect of outliers after disaggregating alliance ratings into within-therapist and between-therapist components. The alliance-outcome correlation and the proportion of variance in post-test outcomes explained by alliance when controlling for pre-test outcomes were similar regardless of how low alliance outliers were treated (change in  $r \leq .04$ , change in  $R^2 \leq 1\%$ ). Results from the disaggregation were similar. Thus, it appears that the alliance-outcome association is not an artifact of the influence of low alliance outliers.

*Keywords:* therapeutic alliance; psychotherapy process; psychotherapy outcome; outliers

**Public Significance Statement**

This study suggests that the alliance-outcome association is not driven by low alliance outliers. Although it remains to be established whether alliance *causes* improvements in outcome, the current study helps rule out low alliance outliers as explanation for why the alliance-outcome association is consistently observed.

**Threat alert: The effect of outliers on the alliance-outcome correlation**

Of all the relationship factors that have been examined, the most robust predictor of outcome is the alliance (Wampold & Flückiger, in press; Wampold & Imel, 2015). The alliance (also known as the therapeutic, working, or helping alliance) can be viewed as the “holistic collaborative aspects of the therapist-client relationship” (Flückiger et al., 2018, p. 317). The alliance includes key relational elements such as therapists’ and patients’ agreement on the tasks and goals of therapy along with the affective bond that develops between the therapist and the patient (Bordin, 1979). Meta-analyses over several decades have shown that the alliance, measured relatively early in psychotherapy, predicts the outcome of psychotherapy (Flückiger et al., 2018; Horvath et al., 2011; Horvath & Symonds, 1991; Martin et al., 2000). The most recent meta-analysis (Flückiger et al., 2018) examined nearly 300 studies containing over 30,000 patients and found the aggregate correlation of the alliance and psychotherapy outcome to be .28, which is a moderate-sized correlation. This correlation between early alliance and final outcomes is one of the most frequently mentioned results in the psychotherapy literature. The estimate can be said to be reliable given that the standard error of this estimate was small (standard error of estimate = 0.011; see Flückiger et al., 2018).

Despite the robust correlation of alliance and final outcome, there are many threats to the conclusion that the alliance is a *causal* factor in psychotherapy. Studies have been conducted to rule out various threats, including meta-analyses of studies designed to examine various aspects of the alliance-outcome association. It appears that the alliance is not simply a result of patient characteristics or early processes (Del Re et al., 2012; Del Re et al., 2021; Flückiger et al., 2012; Flückiger, Del Re, et al., 2020), is not solely a consequence of prior symptom change (Flückiger, Rubel, et al., 2020; Zilcha-Mano et al., 2014), and is not due to methodological features such as

rater perspective (patient, therapist, or observer) of either the alliance or the outcome or the particular alliance measure used (Flückiger et al., 2018). Surprisingly, however, one methodological issue not examined extensively is the influence of outliers on the alliance-outcome correlation. Even though the alliance-outcome association has been observed in meta-analyses of hundreds of studies (e.g., Flückiger et al., 2018), it is entirely possible that each study is similarly impacted by the influence of outliers. Bias due to outliers is therefore not addressed simply by aggregating effects across studies via meta-analysis. Similarly, traditional meta-analytic risk of bias assessments (e.g., tests of funnel plot asymmetry, coding of study features; Borenstein et al., 2009; Higgins & Thomas, 2019) are not designed to evaluate the influence of outliers.

There is reason to be concerned about the potential influence of outliers in alliance ratings. There is now clear meta-analytic evidence that alliance ratings are often not normally distributed, but instead show pervasive ceiling effects (Meier & Feeley, 2021; Tryon, Blackwell, & Hammel, 2008). Ceiling effects refer to instances in which ratings are made disproportionately at the high end of a scale, producing a distribution that does not correspond to the normal (i.e., Gaussian) distribution but rather shows a negative skew. The possibility that this restriction of range may attenuate the alliance-outcome association has been recognized (e.g., Baldwin & Goldberg, 2021). However, this ceiling effect may also make the alliance-outcome association prone to the influence of outliers, specifically relatively rare instances in which patients rate their alliance very low and have poor outcomes, perhaps even deteriorating outcomes over the course of therapy.

Figure 1 provides an illustration of how low alliance outliers could artifactually produce the appearance of an alliance-outcome association when in fact no such relationship exists for the

vast majority of observations.<sup>1</sup> As shown in the figure, there may not be an association between alliance and psychotherapy outcome for either the upper end of alliance ratings (shown as circles) or the low alliance outliers (shown as triangles). The dashed regression lines for each group illustrate the null association in each group when examined separately. However, a linear regression line fit to the full range of alliance scores (solid line) shows the negative slope associated with a Pearson's  $r$  of  $-.28$ , albeit one that is driven by a small number of very low alliance ratings that are linked with high distress ratings. In this figure with 5,000 observations, only 75 outliers (1.5% of points) can produce an alliance-outcome correlation of  $.28$ , despite there being no association between alliance and psychotherapy outcome for 98.5% (4,925) of observations.

Despite the theoretical possibility that the alliance-outcome association is “driven” by a small number of low outlier alliance ratings, no study to our knowledge has examined this possibility directly. If removing outliers dramatically reduces the correlation, it may well be that a few outlying cases are responsible for the observed correlations. It is also well established that restricting the range of variables attenuates the magnitude of Pearson correlation coefficients (Cohen, Cohen, West, & Aiken, 2003). Thus, an investigation of the distribution of alliance scores and the effects of characteristics of the distribution on the correlation is needed.

The purpose of the present study was to examine the distribution of alliance scores and to determine the effect of removing low alliance outliers in two large, naturalistic psychotherapy

---

<sup>1</sup> Of note, we illustrate this relationship using the hopefully easily interpreted correlation between alliance and psychotherapy outcome, which is the relationship emphasized in widely cited meta-analyses of the alliance-outcome association (e.g., Flückiger et al., 2018). However, it is important to recognize that this association represents the total correlation, which is a mixture of the influence of both within-therapist and between-therapist contributions to the alliance (Baldwin et al., 2007). Moreover, a simple correlation that ignores the nesting of patients within therapists may inflate Type I error rates (Baldwin et al., 2005). Thus, we are not recommending the simple correlation as the preferred method for characterizing the alliance-outcome relationship. We address concerns related to nesting of observations within therapists and the disaggregation of alliance ratings into within- and between-therapist components in further analyses in the current study.

data sets to determine the effect on the alliance-outcome correlation. Data were drawn from two distinct contexts - a university counseling center in the United States and a community-based mental health clinic in Canada. Each data set included a different measure of alliance and psychotherapy outcome. Thus, we were able to examine the degree to which the pattern of findings was similar across samples. We had no *a priori* hypothesis regarding the extent to which the alliance-outcome association would be influenced by the removal of low alliance outliers. We focus our examination on the simple alliance-outcome correlation, as this is the correlation often referenced as evidence for the importance of the alliance and the correlation estimated in the widely cited meta-analyses of the alliance-outcome association (*viz.*, [Flückiger et al., 2018](#)). Nonetheless, there are important limitations associated with this analytic approach, as it ignores the nesting of patients within therapists (which can produce biased estimates of standard errors; Baldwin et al., 2005) as well as the use of alliance scores that are not disaggregated into within- and between-therapist components (Baldwin et al., 2007). Thus, we also include analyses that investigate the influence of removing outliers in multilevel models that account for the nesting of patients within therapists as well as in models in which alliance scores are disaggregated into within- and between-therapist components. The details of these analytic approaches are discussed below.

## Method

### Participants and Procedures

Archival data were drawn from two naturalistic psychotherapy settings. We are unable to share patient data, but code used in analyses is available by request. This study was not preregistered. Study procedures were approved by the Institutional Review Board at the

University of Wisconsin - Madison (Naturalistic Psychotherapy Project) and the University of Utah (Digital Exploration of Psychotherapy Project).

The first setting was a university counseling center at a large, Western university in the United States. Patients at the counseling center completed the 34-item Counseling Center Assessment of Psychological Symptoms (CCAPS; Center for Collegiate Mental Health [CCMH], 2014; Locke et al., 2012) prior to each session of individual psychotherapy. Data were collected between September 2007 and January 2018. This university counseling center provides approximately 10,000 sessions per year. Patients present with mental health concerns common among undergraduate and graduate students (e.g., depression, anxiety, substance use; Benton et al., 2003). Treatment is provided by both licensed mental health professionals as well as trainees pursuing graduate mental health degrees (e.g., masters of social work, doctorate in counseling or clinical psychology).

The second setting was a community mental health center in a large, Canadian city. Patients at the community health center completed Outcome Questionnaire-45 (OQ-45; Lambert et al., 2004) prior to each session of individual psychotherapy. Internal consistency across all OQ-45 assessments was high ( $\alpha = .95$ ). Data were collected between January 2015 and December 2019. The community mental health center provides approximately 45,000 sessions per year. The most common reasons patients seek counseling at this center are anxiety, depression, relationship challenges, and stress. Similar to the university counseling center, treatment is provided by both licensed mental health professionals as well as trainees pursuing graduate mental health degrees and/or post-graduate supervised training.

Both the CCAPS and the OQ-45 are widely used, are designed to assess change occurring over the course of psychotherapy, and provide total scores of general distress (Boswell et al.,

2013; Center for Collegiate Mental Health, 2014; Lambert et al., 1996, 2004; Locke et al., 2012). Although both measures have demonstrated some desirable psychometric properties (e.g., high internal consistency reliability, convergent and discriminant validity, sensitivity to change, measurement invariance across subsamples; Boswell et al., 2013; Locke et al., 2013; Sherman et al., 2021; Yoon et al., 2022), there has been more recent debate regarding the factor structure of these measures, including when disaggregated into within- and between-person levels (Kim et al., 2010; McAleavey et al., 2020; Tabet et al., 2020). Nonetheless, for the purposes of the current study, we adopted the commonly used practice of computing a total score across all items as an indicator of general distress.

The two samples were filtered, as is the case with naturalistic data sets. First, for patients who saw multiple therapists or completed multiple episodes of treatment, only the first episode of individual psychotherapy with a single therapist was included. We adopted definitions of an episode of care used by the two respective clinics. At the university counseling center, an episode of care was defined as sessions occurring without a gap of >90 days between sessions. At the community mental health center, an episode of care was defined as sessions occurring without a gap of >44 days between sessions. Second, both samples were restricted to patients who completed outcome measures (CCAPS or OQ-45) at their last session of psychotherapy and a rating of alliance associated with their second, third, or fourth session of psychotherapy. Third, both samples were restricted to patients in the clinical range of the respective outcome measures at baseline (distress index raw score >1.21 on CCAPS,  $\geq 63$  on OQ-45). These restrictions resulted in retaining 20.7% of the counseling center sample and 45.0% of the community mental health clinic sample.

Demographics of the counseling center sample ( $N = 1,052$ ) were as follows: predominantly female (57.8%), with 39.6% identifying as male, 2.1% identifying as another gender identity (e.g., queer, non-binary), and 0.5% not reporting their gender; predominantly non-Hispanic White (73.0%), with 9.0% identifying as Hispanic/Latinx, 8.2% identifying as Asian/Asian American, 5.7% identifying as multi-racial, 1.5% identifying as African American/Black, 0.6% identifying as Pacific Islander, 0.3% identifying as Middle Eastern/North African, 0.2% identifying as American Indian/Native American/Alaskan Native, and 1.5% not reporting their race/ethnicity; on average 24.34 years old (standard deviation [SD] = 5.55). Patients in the counseling center sample completed at average of 7.13 sessions of psychotherapy (SD = 5.34, range = 3 to 63). These patients were seen by 93 therapists who had on average 11.31 patients in the data set (SD = 11.98, range = 1 to 55).

Demographics of the community mental health center sample ( $N = 11,029$ ) were as follows: predominantly female (59.8%), with 37.7% identifying as male, and 2.5% not reporting; predominantly non-Hispanic White (66.4%), with 7.6% identifying as Asian, 3.0% as First Nations/Metis/Inuit, 2.2% as Hispanic/Latinx, 1.7% as African/Caribbean/Black, 1.1% as Arab/Middle Eastern, and 18.1% not reporting; on average 33.05 years old (SD = 11.10). Patients in the community mental health center sample completed at average of 5.95 sessions of psychotherapy (SD = 5.18, range = 2 to 86). These patients were seen by 314 therapists who had on average 35.12 patients in the data set (SD = 40.05, range = 1 to 283).

### **Alliance Measures**

Patients at the counseling center completed a previously validated (Imel, Hubbard, Rutter, & Simon, 2013) four-item version of the Working Alliance Inventory – Short Form Revised (WAI-SR; Hatcher & Gillaspay, 2006). Items assessed the goal (“\_\_\_\_\_ and I are

working towards mutually agreed upon goals”), task (“I believe the way that we are working on my problem is correct”, and bond (“I feel that \_\_\_\_\_ appreciates me”) dimensions of therapeutic alliance. Items were rated on a 1 (Never) to 7 (Always) Likert-type scale, with a total score calculated by averaging across items. Internal consistency across assessments was high ( $\alpha = .88$ ). To reduce patient burden, alliance ratings were completed prior to each session when outcome measures were also completed (with the exception of the first session), with patients reflecting back on their experience of the alliance in the previous session.

Patients at the community mental health center completed the 10-item Session Rating Scale (SRS 2.1; Johnson, 1995). This measure asks patients to rate ten dimensions of therapeutic alliance (e.g., goal agreement, therapist understanding, patient feeling of acceptance) on a 5-point Likert scale with symmetrical anchors indicating a low (e.g., marking *Agree with this Side* for “I feel criticized or judged”) or high (marking *Agree with this Side* for “I felt accepted”) alliance. The 10-item SRS has been linked to clinical outcomes (Shaw et al., 2019) and influenced development of the four-item SRS 3.0 (Duncan et al., 2003) which has been widely used (Murphy et al., 2020) and with scores shown to relate to outcome in the context of psychotherapy (Sun et al., 2020). The community mental health setting preferred the 10-item version because it was determined to have more substantive process content to discuss with clients. Patients completed the SRS at the conclusion of each psychotherapy session. Internal consistency was high ( $\alpha = .95$ ).

## **Data Analysis**

### ***Assessing Alliance-Outcome Association***

Two methods were used to characterize the alliance-outcome association. First, Pearson correlations were examined between alliance and pre-post change scores. This method was

employed due to its ease of interpretation as well as the fact that this is the association that has been aggregated in widely cited meta-analyses focused on the alliance-outcome association (e.g., Flückiger et al., 2018). Thus, it seemed important to evaluate the influence of outliers on this association specifically. Change scores were calculated as post-treatment minus pre-treatment distress, where lower score indicates greater decreases in distress over time. Pre-treatment reflected baseline levels and post-treatment was defined as the score provided at patients' last psychotherapy session. Correlations and their 95% confidence intervals (CIs) were calculated using the 'psych' (Revelle, 2020) package in R (R Core Team, 2021).

It has been increasingly recognized that alliance ratings include contributions from both the patient and the therapist (e.g., Baldwin et al., 2007). The correlation typically meta-analyzed to represent the alliance-outcome association (e.g., Flückiger et al., 2018) is the total correlation, which is a combination of both within-therapist (i.e., patient contributions to alliance) and between-therapist (i.e., therapist contributions to alliance) components. It is theoretically possible that outliers occur at either the within- or between-therapist level. Thus, we also conducted an analysis that disaggregated alliance ratings into these two components.

To calculate the within-therapist component, we centered patients' alliance ratings around their therapists' mean alliance rating:

$$Y_{ij} = x_{ij} - \bar{x}_j,$$

where  $Y_{ij}$  is the within-therapist alliance score of a given patient ( $i$ ) seen by a given therapist ( $j$ ). This value is computed as a given patients raw alliance score ( $x_{ij}$ ) with that patient's therapist's mean alliance score ( $\bar{x}_j$ ) (i.e., averaged across all that therapists' patients) subtracted. Therapists' mean alliance score ( $\bar{x}_j$ ) represented the between-therapist component. All models conducted for the non-disaggregated alliance ratings (including the multilevel models that

account for the nesting of observations within therapists) were also conducted using the alliance ratings disaggregated into within- and between-therapist components. For identifying between-therapist alliance outliers, we calculated a pooled SD that was pooled across therapists weighted based on their number of patients in the data set.

Given that ignoring the nesting of patients within therapists can provide biased estimates of standard errors (Baldwin et al., 2005), multilevel models were constructed to examine the influence of outliers while account for this nesting. The formula for this model was:

$$Y_{ij} = \beta_{00} + \beta_{10}(\text{pre} - \text{posttest distress}) + \beta_{20}(\text{alliance}) + [U_{0j} + e_{ij}],$$

where  $Y_{ij}$  is the outcome (post-test distress) of a given patient ( $i$ ) seen by a given therapist ( $j$ ). Post-test distress is predicted by a fixed intercept ( $\beta_{00}$ ), a given patient's pre-test distress ( $\beta_{10}$ ), a given patients' alliance rating ( $\beta_{20}$ ), along with a random intercept unique to each therapist ( $U_{0j}$ ) and a residual error term ( $e_{ij}$ ).

To derive an effect size metric to allow comparisons across methods of handling outliers, we calculated a semi-partial  $R^2$  value using the 'r2beta' function in the 'r2glmm' package (Jaeger, 2017). This value represents the proportion of variance in the outcome variable explained by each fixed effect in a multilevel model. We focus on the proportion of variance associated with alliance ratings. Models were constructed examining these metrics of the alliance-outcome association for alliance rated at sessions two, three, and four of psychotherapy.

### ***Treatment of Alliance Outliers***

To examine the influence of outliers on the alliance-outcome association, we treated outliers five different ways. First, models were run using all the available data with no modification to alliance outliers (i.e., raw data). Next, two models were run with outliers excluded, with outliers defined as values greater than either three or two SDs from the mean

value. Although outliers could in theory be high or low values, all outliers were low values (i.e., low alliance ratings), given the negative skew. Lastly, two models were run with outliers winsorized, with outliers again defined as values further than either three or two SDs from the mean value. Winsorizing involves assigning outliers the value just outside of the range used to define an outlier (Tukey, 1962). For example, values more than three SDs below the mean alliance rating (e.g., <23.06 on SRS at session 2) were assigned the alliance value three SDs below the mean (23.06).

### Results

Descriptive statistics and effect sizes for the alliance-outcome association in both samples are reported in Tables 1 and 2. Mean raw alliance ratings (i.e., outliers included) were towards the maximum value for both samples (e.g., 5.88 out of 7.00, 38.04 out of 40.00, for session two alliance in the counseling center and community mental health center, respectively). Skewness and kurtosis were within the recommended range (skew < 2, kurtosis < 7; Curran, West, & Finch, 1996) in the counseling center data. There was, however, evidence of marked deviations from normality in the community mental health data (e.g., skewness = -6.75 and kurtosis = 50.98 for session four alliance ratings). As expected, outlier treatments reduced both skewness and kurtosis. Also as expected, mean values increased and SDs decreased as low outliers were removed, and of course, the sample size decreased as low outliers were removed and did not decrease when outliers were winsorized. Correlation coefficients were somewhat smaller than that found in meta-analyses of the alliance-outcome association (i.e.,  $r = .28$ ; Flückiger et al., 2018).

The major finding for both samples in the current study was correlations and proportion of variance in post-test distress explained by alliance remained essentially constant across

alliance outlier treatments. For example, the correlation between outcome and session two alliance using all counseling center data (i.e., raw alliance ratings) was  $-.15$ , with this association ranging from  $-.15$  to  $-.12$  across outlier treatments. In the counseling center data, correlations with adjusted alliance ratings were equivalent to or slightly smaller than correlations using raw alliance ratings. In contrast, correlations with adjusted alliance ratings were slightly larger than correlations using raw alliance ratings in the community mental health center data. All correlations remained statistically significant, with two exceptions in the counseling center data: when removing outliers two SDs from the mean and examining associations with session three alliance ( $r = -.06$ , 95% CI  $[-.13, .01]$ ) and when removing outliers three SDs from the mean and examining associations with session four alliance ( $r = -.08$ , 95% CI  $[-.15, .00]$ ). However, when examined in multilevel models, regression coefficients for both alliance variables were statistically significant negative predictors of post-test distress, controlling for pre-test distress ( $B = -0.078$  and  $-0.074$ ,  $p = .013$  and  $.022$ , for session three and session four alliance, respectively).  $R^2$  values were almost identical across outlier treatments and in all instances were within 1% of the  $R^2$  value using raw alliance ratings with the exception of the session four alliance in the community mental health center where removing outliers increased the  $R^2$  value by  $> 1\%$ .

Figure 2 illustrates the alliance-outcome association at session two in the counseling center data. As can be seen, the non-parametric regression lines (i.e., loess or local regression curve which does not assume a linear or other specific shape of association; Jacoby, 2000) indicate these associations were not heavily influenced by low alliance outliers. Rather, figures show a largely linear association, regardless of how outliers are treated.

Figure 3 illustrates the alliance-outcome association at session two in the community mental health center data. Non-parametric regression lines showed less linearity in these data.

However, in all cases, there was no evidence that the alliance-outcome association was being driven by low alliance outliers. In fact, if anything, low alliance outliers appeared to attenuate rather than inflate the alliance-outcome association (consistent with the increasing correlation coefficients when outliers were removed).

### **Disaggregation Analyses**

Models were rerun removing outliers based on alliance disaggregated into within- and between-therapist components. Results were very similar to those observed using the non-disaggregated alliance ratings. For within-therapist alliance, the alliance-outcome correlations were similar in magnitude with and without alliance outliers excluded (i.e., change in  $r_s \leq .04$ ). Changes in  $R^2$  values again remained  $\leq 1\%$ . In only one instance (session three alliance in the counseling center data) was an association significant with raw within-therapist alliance ratings but no longer significant following the treatment of outliers (removing values two SDs from the mean resulted in  $p = .077$ ).

Very few outliers were detected at the between-therapist level. The treatment of these outliers did not substantially modify results. Alliance-outcome correlations changed  $\leq .01$  and  $R^2$  values changed  $\leq 0.50\%$ . Significance tests for the fixed effect of alliance predicting post-test distress were unchanged across outlier treatments.

### **Discussion**

Science progresses by repeated attempts to discredit a conclusion by examining various threats to validity (Popper, 1963). Although there is a robust alliance-outcome correlation (Flückiger et al., 2018), establishing the importance of the alliance as a therapeutic factor depends on ruling out alternative explanations. The alliance as a therapeutic factor has survived many falsification attempts (Wampold & Imel, 2015; Wampold & Fluckiger, in press). To our

knowledge, the present study is the first study to examine outliers as a threat to validity by evaluating the influence of removing low alliance outliers on the alliance-outcome association. Low alliance outliers are an important potential methodological threat to the alliance-outcome association, particularly given known ceiling effects on these measures (Meier & Feeley, 2021) may make this association more vulnerable to the influence of outliers.

Across two large, naturalistic psychotherapy data sets ( $Ns = 1,052$  and  $11,029$ ), collected in two different clinical settings (university counseling center, community-based mental health clinic) in two different countries (US and Canada), and using two different sets of alliance (4-item WAI-SR and 10-item SRS) and outcome measures (CCAPS and OQ-45), the alliance-outcome association was largely unchanged with and without low alliance outliers. This was true across four different outlier treatments (defining outliers as 3 SDs or 2 SDs, dropping outliers or winsorizing) and whether the alliance-outcome association was examined as a correlation or as an  $R^2$  value within a multilevel model framework. Results were very similar when disaggregating alliance ratings into within- and between-therapist components. Thus, it appears that the alliance-outcome association is robust to this potential threat.

One small potential difference emerged in the pattern of results across the two data sets. In particular, the counseling center data showed evidence of slightly *decreasing* effect sizes as outliers were removed or winsorized while the community mental health data showed evidence of slightly *increasing* effect sizes as outliers were removed or winsorized. Slightly decreasing effect sizes were to be expected, given range restriction attenuates correlations (Cohen et al., 2003). However, the slightly increasing effect size is intriguing. Examination of the non-parametric regression line in the community mental health center data showed an unexpected positive association at the low end of alliance scores. One possibility that may be worth

exploring in a future study is whether some participants misinterpret the symmetrical SRS scale anchors and provide low alliance ratings when meaning to provide high ratings. It is possible removing these points may have reduced measurement error variance and thereby strengthened the observed correlation.

This study has several important limitations. Notably, the samples were drawn from non-specialty mental health treatment settings and may therefore not necessarily generalize to the treatment of a specific disorder or the use of a specific form of psychotherapy. It is possible that low alliance ratings are more impactful for certain diagnoses. The interpersonal reactivity characteristic of borderline personality disorder (American Psychiatric Association, 2013), for example, may produce a combination of low alliance ratings and low ratings of distress improvement which could inflate the alliance-outcome association. Indeed, borderline personality disorder showed a particularly large alliance-outcome correlation in Flückiger et al. (2018),  $r = .32$ . Similarly, low alliance ratings may be more common in some forms of psychotherapy, which could make the alliance-outcome association more prone to influence by outliers. In addition, although we examined two alliance ratings and two outcome measure types, there are many other widely used measures of alliance and other outcome measures which in theory may be more vulnerable to the influence of outliers. The retrospective assessment of alliance (i.e., ratings made about the previous session at the beginning of the subsequent session) in the counseling center data may have reduced the validity or reliability of the alliance ratings. As we were interested in investigating the influence on alliance outliers on the alliance-outcome association primarily using methods that have been used to establish this relationship within the literature (i.e., correlations between alliance ratings and pre-post change scores), we employed fairly simple modeling strategies (i.e., correlations between alliance ratings with change scores,

multilevel models predicting post-treatment distress from pre-treatment distress and alliance ratings). However, it may be useful in a future study to investigate the influence of alliance outliers using more recently employed modeling strategies such as varieties of longitudinal structural equation modeling that model repeated assessment of both alliance and psychotherapy outcome (e.g., dynamic structural equation modeling, random intercept cross-lagged panel models; Flückiger et al., 2022; Sun et al., 2020). It is possible that low alliance outliers do exert an important influence when examined using a different modeling approach. Information was lacking regarding the therapists in both settings, which limits our ability to evaluate the degree to which these particular samples of therapists are similar to the general population of therapists. Moreover, the majority of the patients in both naturalistic data sets were excluded from analyses due to insufficient outcome/alliance data or due to baseline distress scores below the clinical cutoff. This may limit generalizability of the current findings to the general patient population seen in these settings. Lastly, both patient samples were predominantly White and female, which limits generalizability to other demographic groups.

These limitations notwithstanding, the current study provides evidence consistent with the view that the alliance-outcome association is not a mere artifact of low alliance outliers. This association remained with and without low alliance outliers and appeared robust across two large, naturalistic psychotherapy samples that differed in several ways. Thus, results add to the growing literature highlighting the therapeutic alliance as an important and potentially causal mechanism in the context of psychotherapy.

### **References**

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

- Baldwin, S., & Goldberg, S. B. (2021). Methodological foundations and innovations in psychotherapy research. In M. Barkham, W. Lutz, and L. Castonguay (Eds.), *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change (7th ed.)* Hoboken, NJ: Wiley & Sons.
- Baldwin, S.A., Murray, D.M., & Shadish, W.R. (2005). Empirically supported treatments or Type I errors? Problems with the analysis of data from group-administered treatments. *Journal of Consulting and Clinical Psychology, 73*(5), 924-935.  
doi: 10.1037/0022-006X.73.5.924
- Baldwin, S.A., Wampold, B.E., & Imel, Z.E. (2007). Untangling the alliance-outcome correlation: Exploring the relative importance of therapist and patient variability in the alliance. *Journal of Consulting and Clinical Psychology, 75*(6), 842-852. 852.  
doi:10.1037/0022-006X.75.6.842
- Barton, K. (2020). *MuMIn: Multi-model inference*. R package version 1.43.17, <https://CRAN.R-project.org/package=MuMIn>
- Benton, S.A., Robertson, J.M., Tseng, W.C., Newton, F.B., & Benton, S.L. (2003). Changes in counseling center client problems across 13 years. *Professional Psychology: Research and Practice, 34*(1), 66-72. doi: 10.1037/0735-7028.34.1.66
- Bordin, E.S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, Research and Practice, 16*(3), 252-260.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. New York: Wiley.
- Boswell, D. L., White, J. K., Sims, W. D., Harrist, R. S., & Romans, J. S. (2013). Reliability and validity of the Outcome Questionnaire–45.2. *Psychological Reports, 112*(3), 689-693.

- Center for Collegiate Mental Health. (2014). *CCAPS 2012 Technical Manual*. University Park, PA.
- 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.
- Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods, 1*(1), 16-29.
- Del Re, A. C., Flückiger, C., Horvath, A. O., Symonds, D., & Wampold, B. E. (2012). Therapist effects in the therapeutic alliance–outcome relationship: A restricted-maximum likelihood meta-analysis. *Clinical Psychology Review, 32*(7), 642-649.  
<https://doi.org/10.1016/j.cpr.2012.07.002>
- Del Re, A. C., Flückiger, C., Horvath, A. O., & Wampold, B. E. (2021). Examining therapist effects in the alliance–outcome relationship: A multilevel meta-analysis. *Journal of Consulting and Clinical Psychology*. <https://doi.org/10.1037/ccp0000637>
- Duncan, B.L., Miller, S.D., Sparks, J.A., Claud, D.A., Reynolds, L.R.,...& Johnson, L.D. (2003). The Session Rating Scale: Preliminary psychometric properties of a "working" alliance measure. *Journal of Brief Therapy, 3*(1), 3-12.
- 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.  
<https://doi.org/http://dx.doi.org/10.1037/pst0000172>
- Flückiger, C., Del Re, A. C., Wampold, B. E., Symonds, D., & Horvath, A. O. (2012). How central is the alliance in psychotherapy? A multilevel longitudinal meta-analysis. *Journal*

- of Counseling Psychology*, 59(1), 10-17. <https://doi.org/10.1037/a0025749>
- Flückiger, C., Del Re, A. C., Wlodasch, D., Horvath, A. O., Solomonov, N., & Wampold, B. E. (2020). Assessing the alliance–outcome association adjusted for patient characteristics and treatment processes: A meta-analytic summary of direct comparisons. *Journal of Counseling Psychology*. <https://doi.org/10.1037/cou0000424>
- Flückiger, C., Horvath, A. O., & Brandt, H. (2022). The evolution of patients’ concept of the alliance and its relation to outcome: A dynamic latent-class structural equation modeling approach. *Journal of Counseling Psychology*, 69(1), 51–62.
- Flückiger, C., Rubel, J., Del Re, A. C., Horvath, A. O., Wampold, B. E., Crits-Christoph, P., Atzil-Slonim, D., Compare, A., Falkenström, F., Ekeblad, A., Errázuriz, P., Fisher, H., Hoffart, A., Huppert, J. D., Kivity, Y., Kumar, M., Lutz, W., Muran, J. C., Strunk, D. R., ...Barber, J. P. (2020). The reciprocal relationship between alliance and early treatment symptoms: A two-stage individual participant data meta-analysis. *Journal of Consulting and Clinical Psychology*, 88(9), 829-843. <https://doi.org/10.1037/ccp0000594>
- Hatcher, R. L., & Gillaspay, J. A. (2006). Development and validation of a revised short version of the Working Alliance Inventory. *Psychotherapy Research*, 16(1), 12-25.  
doi: 10.1080/10503300500352500
- Higgins, J.P.T., & Thomas, J. (2019). *Cochrane Handbooks for Systematic reviews of Interventions (Second ed.)*. Oxford: Cochrane Collaboration and John Wiley & Sons.
- Horvath, A. O., Del Re, A. C., Flückiger, C., & Symonds, D. (2011). Alliance in individual psychotherapy. *Psychotherapy*, 48(1), 9-16. <https://doi.org/10.1037/a0022186>
- Horvath, A. O., & Symonds, B. D. (1991). Relation between working alliance and outcome in psychotherapy: A meta-analysis. *Journal of Counseling Psychology*, 38, 139-149.

- Imel, Z. E., Hubbard, R. A., Rutter, C. M., & Simon, G. (2013). Patient-rated alliance as a measure of therapist performance in two clinical settings. *Journal of Consulting and Clinical Psychology, 81*(1), 154-165. doi: 10.1037/a0030903
- Jacoby, W. G. (2000). Loess:: a nonparametric, graphical tool for depicting relationships between variables. *Electoral Studies, 19*(4), 577-613.
- Jaeger, B. (2017). *r2glmm: Computes R Squared for Mixed (Multilevel) Models*. R package version 0.1.2, <https://CRAN.R-project.org/package=r2glmm>
- Johnson, L. D. (1995). *Psychotherapy in the age of accountability*. New York: Norton.
- Kim, S. H., Beretvas, S. N., & Sherry, A. R. (2010). A validation of the factor structure of OQ-45 scores using factor mixture modeling. *Measurement and Evaluation in Counseling and Development, 42*(4), 275-295.
- Lambert, M.J., Burlingame, G.M., Umphress, V., Hansen, N.B., Vermeersch, D.A., Clouse, G.C., & Yanchar, S.C. (1996). The reliability and validity of the Outcome Questionnaire. *Clinical Psychology and Psychotherapy, 3*(4), 249-258.
- Lambert, M. J., Morton, J. J., Hatfield, D., Harmon, C., Hamilton, S., Reid, R. C.,...Burlingame, G. B. (2004). *Administration and scoring manual for the Outcome Questionnaire-45*. Orem, UT: American Professional Credentialing Services.
- Locke, B. D., McAleavey, A. A., Zhao, Y., Lei, P. W., Hayes, J. A., Castonguay, L. G., ... & Lin, Y. C. (2012). Development and initial validation of the Counseling Center Assessment of Psychological Symptoms–34. *Measurement and Evaluation in Counseling and Development, 45*(3), 151-169.
- Martin, D. J., Garske, J. P., & Davis, M. K. (2000). Relation of the therapeutic alliance with outcome and other variables: A meta-analytic review. *Journal of Consulting and Clinical*

- Psychology*, 68, 438-450.
- McAleavey, A. A., Castonguay, L. G., Hayes, J. A., & Locke, B. D. (2020). Multilevel versus single-level factor analysis: Differentiating within-person and between-person variability using the CCAPS-34. *Journal of Consulting and Clinical Psychology*, 88(10), 907-922.
- Meier, S. T., & Feeley, T. H. (2021). Ceiling effects indicate a possible threshold structure for working alliance. *Journal of Counseling Psychology*. <https://doi.org/10.1037/cou0000564>
- Murphy, M. G., Rakes, S., & Harris, R. M. (2020). The psychometric properties of the Session Rating Scale: A narrative review. *Journal of Evidence-Based Social Work*, 17(3), 279-299.
- Popper, K. R. (1963). *Conjectures and refutations: The growth of scientific knowledge*. New York: Basic Books.
- R Core Team (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Revelle, W. (2020). *psych: Procedures for psychological, psychometric, and personality research*. R package version 2.0.12, <http://CRAN.R-project.org/package=psych>
- Shaw, S. L., Lombardero, A., Babins-Wagner, R., & Sommers-Flanagan, J. (2019). Counseling Canadian Indigenous Peoples: The therapeutic alliance and outcome. *Journal of Multicultural Counseling and Development*, 47, 49-69.  
<https://doi.org/10.1002/jmcd.12120>
- Sherman, M. F., Sriken, J., Erford, B. T., Smith, H. L., MacInerney, E., Niarhos, F., & Kipper-Smith, A. (2021). Psychometric analysis of CCAPS-34 scores with a large university sample. *Measurement and Evaluation in Counseling and Development*, 54(4), 219-232.
- Sun, Q., Wu, C., Wang, C. D., Yu, L., She, Z., & Falkenström, F. (2020). Alliance-outcome

- relation and progress feedback: Secondary data analyses of a randomized clinical trial study in China. *Psychotherapy Research*, 1-12.
- Tabet, S. M., Lambie, G. W., Jahani, S., & Rasoolimanesh, S. M. (2020). The factor structure of outcome questionnaire–45.2 scores using confirmatory tetrad analysis–partial least squares. *Journal of Psychoeducational Assessment*, 38(3), 350-368.
- 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
- Tukey, J.W. (1962). The future of data analysis. *The Annals of Mathematical Statistics*, 33(1), 1-67.
- Wampold, B. E., & Flückiger, C. (in press). The alliance in mental health care: Conceptualization, evidence and clinical applications. *World Psychiatry*.
- Wampold, B., & Imel, Z.E. (2015). *The great psychotherapy debate: The evidence for what makes psychotherapy work (2nd ed.)*. New York: Routledge.
- Wampold, B. E., & Owen, J. (2021). Therapist Effects: History, methods, magnitude, and characteristics of effective therapists. In L. G. Castonguay, M. Barkham, & W. Lutz (Eds.), *Bergin and Garfield's Handbook of psychotherapy and behavior change (7th ed., pp. 301-330)*. Wiley.
- Yoon, S., Nevins, C. M., Peckham, A. D., & Pinder-Amaker, S. L. (2022). Factor structure of the counseling center assessment of psychological symptoms (CCAPS): Comparisons across college students in inpatient and outpatient settings. *Psychiatry Research*, 310, 114464.

**Table 1***Descriptive Statistics and Effect Sizes for Alliance-Outcome Association in University Counseling Center Sample*

Alliance Treatment	<i>n</i>	Mean	<i>SD</i>	Min	Max	Skew	Kurtosis	<i>r</i>	<i>r</i> <sub>lb</sub>	<i>r</i> <sub>ub</sub>	<i>R</i> <sup>2</sup>	<i>p</i>
Sess #2 raw	1,005	5.88	0.93	1.00	7.00	-1.01	1.44	-0.15	-0.21	-0.09	3.03	<.001
Sess #2 3 SD removed	991	5.93	0.84	3.25	7.00	-0.59	-0.34	-0.13	-0.19	-0.07	2.63	<.001
Sess #2 2 SD removed	959	6.00	0.76	4.25	7.00	-0.39	-0.83	-0.12	-0.18	-0.05	2.02	<.001
Sess #2 3 SD winsorized	1,005	5.89	0.90	3.10	7.00	-0.78	0.23	-0.15	-0.21	-0.09	3.05	<.001
Sess #2 2 SD winsorized	1,005	5.91	0.85	4.03	7.00	-0.52	-0.64	-0.15	-0.21	-0.09	3.09	<.001
Sess #3 raw	822	6.00	0.83	1.75	7.00	-0.89	0.96	-0.10	-0.16	-0.03	1.59	<.001
Sess #3 3 SD removed	812	6.04	0.77	3.75	7.00	-0.57	-0.36	-0.09	-0.15	-0.02	1.37	.001
Sess #3 2 SD removed	789	6.10	0.70	4.50	7.00	-0.37	-0.85	-0.06	-0.13	0.01	0.83	.013
Sess #3 3 SD winsorized	822	6.01	0.81	3.51	7.00	-0.73	0.10	-0.09	-0.16	-0.03	1.57	<.001
Sess #3 2 SD winsorized	822	6.03	0.77	4.34	7.00	-0.47	-0.71	-0.09	-0.16	-0.02	1.49	.001
Sess #4 raw	653	6.10	0.81	2.25	7.00	-1.09	1.69	-0.09	-0.17	-0.01	1.24	.006
Sess #4 3 SD removed	647	6.14	0.75	4.00	7.00	-0.67	-0.33	-0.08	-0.15	0.00	0.86	.022
Sess #4 2 SD removed	636	6.17	0.70	4.50	7.00	-0.55	-0.67	-0.09	-0.17	-0.01	1.09	.010
Sess #4 3 SD winsorized	653	6.11	0.78	3.68	7.00	-0.79	0.05	-0.09	-0.16	-0.01	1.14	.008
Sess #4 2 SD winsorized	653	6.13	0.74	4.49	7.00	-0.58	-0.66	-0.09	-0.16	-0.01	1.14	.008

*Note:* Min = minimum value; Max = maximum value; *r* = correlation coefficient; *r*<sub>lb</sub> = lower bound of 95% confidence interval (CI); *r*<sub>ub</sub> = upper bound of 95% CI; *R*<sup>2</sup> = proportion of variance in post-test distress explained by alliance fixed effects in multilevel models accounting for nesting within therapists; *p* = *p*-value for alliance fixed effect. Sess # = session number; raw = outliers included without adjustment; 3 SD removed = alliance scores three standard deviations (SDs) from mean excluded; 2 SD removed = alliance scores two SDs from mean excluded; 3 SD winsorized = alliance scores three SDs from mean are assigned value of alliance three SDs from mean; 2 SD winsorized = alliance scores two SDs from mean are assigned value of alliance two SDs from mean. The maximum sample size at a specific time point is lower than the total sample size as patients could have provided alliance ratings at some but not all sessions.

**Table 2**

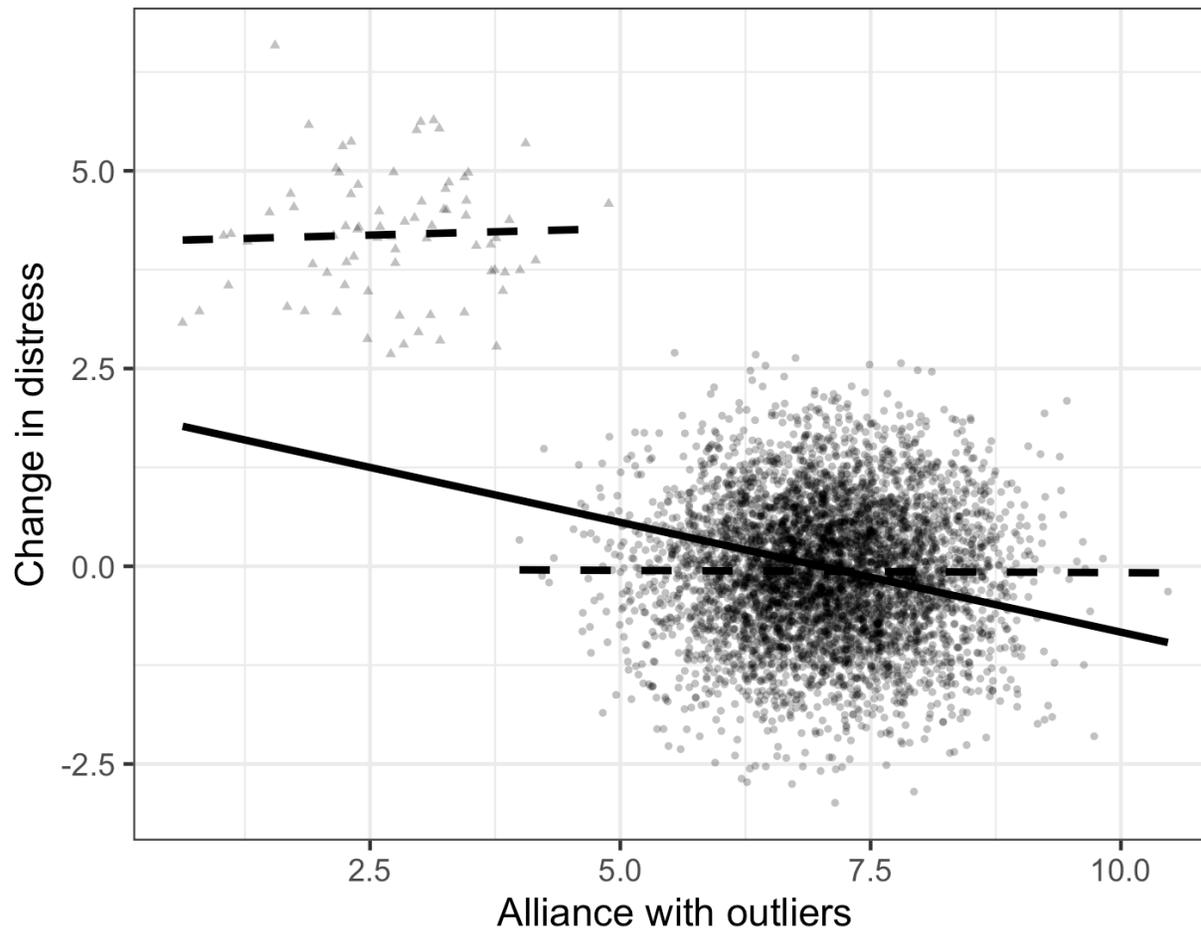
*Descriptive Statistics and Effect Sizes for Alliance-Outcome Association in Community Mental Health Center Sample*

Alliance Treatment	<i>n</i>	Mean	<i>SD</i>	Min	Max	Skew	Kurtosis	<i>r</i>	<i>r</i> <sub>lb</sub>	<i>r</i> <sub>ub</sub>	<i>R</i> <sup>2</sup>	<i>p</i>
Sess #2 raw	10,399	38.04	5.11	0.00	40.00	-5.40	35.00	-0.09	-0.11	-0.07	1.02	<.001
Sess #2 3 SD removed	10,245	38.56	2.73	23.00	40.00	-2.51	6.74	-0.12	-0.14	-0.10	2.01	<.001
Sess #2 2 SD removed	10,146	38.68	2.42	28.00	40.00	-2.25	4.79	-0.12	-0.14	-0.10	1.78	<.001
Sess #2 3 SD winsorized	10,399	38.32	3.31	22.71	40.00	-2.78	8.30	-0.11	-0.13	-0.09	1.75	<.001
Sess #2 2 SD winsorized	10,399	38.42	2.91	27.82	40.00	-2.24	4.42	-0.12	-0.14	-0.10	1.95	<.001
Sess #3 raw	7,968	38.43	5.28	0.00	40.00	-5.98	39.15	-0.09	-0.11	-0.06	0.87	<.001
Sess #3 3 SD removed	7,829	39.06	2.19	23.00	40.00	-3.33	12.92	-0.11	-0.13	-0.08	1.45	<.001
Sess #3 2 SD removed	7,788	39.13	1.96	28.00	40.00	-2.98	9.51	-0.11	-0.13	-0.09	1.61	<.001
Sess #3 3 SD winsorized	7,968	38.77	3.06	22.58	40.00	-3.64	14.34	-0.11	-0.13	-0.09	1.49	<.001
Sess #3 2 SD winsorized	7,968	38.88	2.56	27.86	40.00	-2.93	8.45	-0.12	-0.14	-0.10	1.69	<.001
Sess #4 raw	6,166	38.76	4.64	0.00	40.00	-6.75	50.98	-0.09	-0.12	-0.07	1.10	<.001
Sess #4 3 SD removed	6,082	39.24	2.00	25.00	40.00	-3.57	14.04	-0.14	-0.16	-0.11	2.52	<.001
Sess #4 2 SD removed	6,028	39.34	1.69	30.00	40.00	-3.30	11.45	-0.13	-0.16	-0.11	2.40	<.001
Sess #4 3 SD winsorized	6,166	39.04	2.59	24.83	40.00	-3.72	14.73	-0.13	-0.16	-0.11	2.27	<.001
Sess #4 2 SD winsorized	6,166	39.12	2.22	29.47	40.00	-3.10	9.25	-0.14	-0.16	-0.12	2.62	<.001

*Note:* Min = minimum value; Max = maximum value; *r* = correlation coefficient; *r*<sub>lb</sub> = lower bound of 95% confidence interval (CI); *r*<sub>ub</sub> = upper bound of 95% CI; *R*<sup>2</sup> = proportion of variance in post-test distress explained by alliance fixed effects in multilevel models accounting for nesting within therapists; *p* = *p*-value for alliance fixed effect. Sess # = session number; raw = outliers included without adjustment; 3 SD removed = alliance scores three standard deviations (SDs) from mean excluded; 2 SD removed = alliance scores two SDs from mean excluded; 3 SD winsorized = alliance scores three SDs from mean are assigned value of alliance three SDs from mean; 2 SD winsorized = alliance scores two SDs from mean are assigned value of alliance two SDs from mean. The maximum sample size at a specific time point is lower than the total sample size as patients could have provided alliance ratings at some but not all sessions.

**Figure 1**

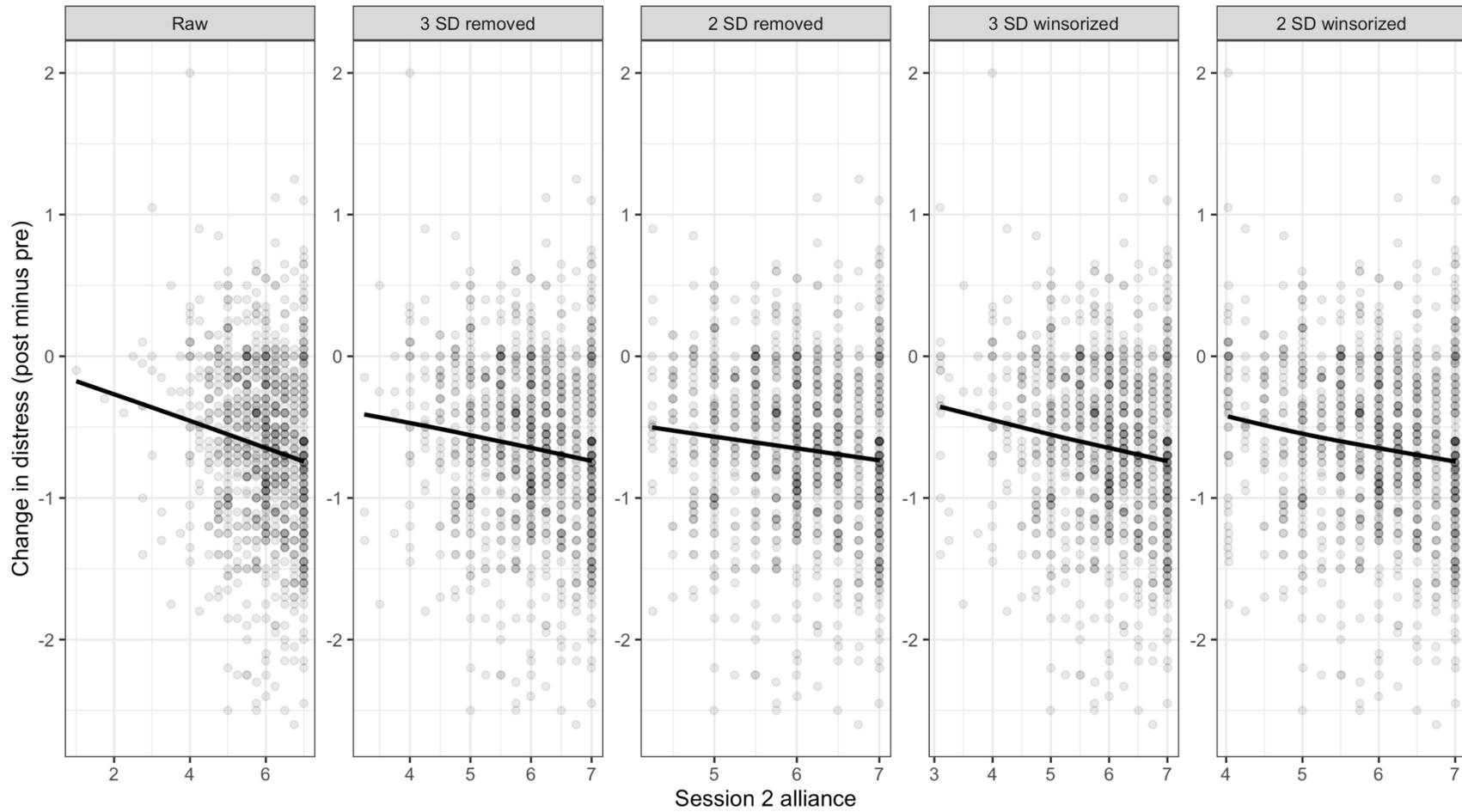
*Scatterplot of Hypothetical Alliance-Outcome Association with Outliers Present*



*Note:* Figures displays hypothetical data to illustrate the potential threat of low alliance and poor outcome on the alliance-outcome association. Linear regression line for the full sample (solid line) shows the expected alliance-outcome association (Pearson's  $r = .28$ ; Flückiger et al., 2018). Separate linear regression lines for the low alliance outliers (triangles) and non-outliers (circles) shows an absence of association when these two regions are examined separately. Change in distress is calculated as post-treatment outcome minus pre-treatment outcome, where a lower score on the outcome is better (i.e., a more negative change score indicates improvement in distress). Alliance mean = 7 and SD = 1.  $N = 5,000$  observations with 75 (1.5%) being low alliance outliers with alliance mean = 2.71, SD = 0.86, minimum = 0.63, maximum = 4.88.

**Figure 2**

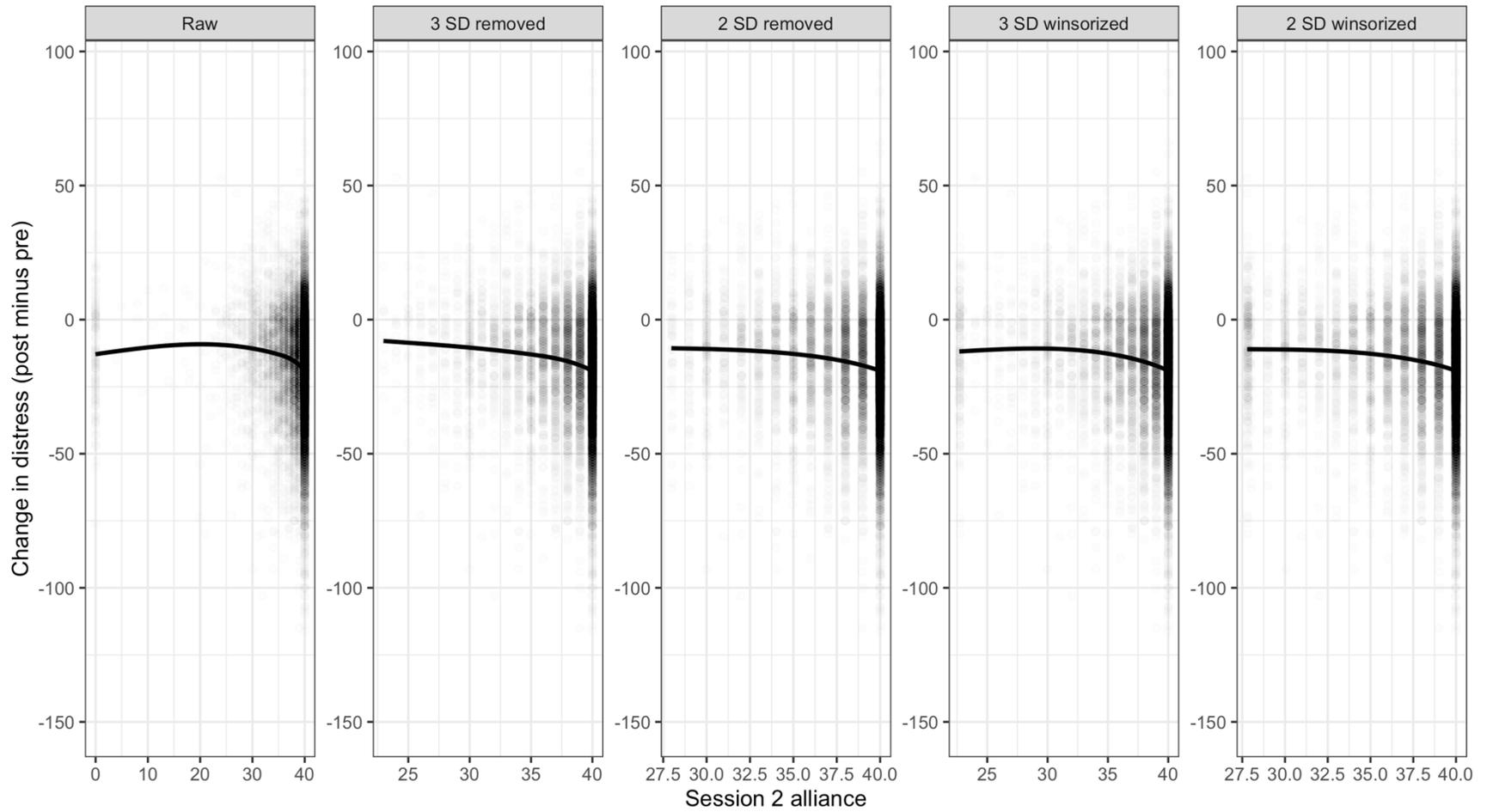
*Scatterplots of Alliance-Outcome Association with Five Treatments of Outliers in Counseling Center Data*



*Note:* Figures display the alliance-outcome association across five treatments of outliers. Data are drawn from the university counseling center sample ( $N = 1,052$ ). Regression line displays non-parametric loess curve that does not assume a linear association. Raw = outliers included without adjustment; 3 SD removed = alliance scores three standard deviations (SDs) from mean excluded; 2 SD removed = alliance scores two SDs from mean excluded; 3 SD winsorized = alliance scores three SDs from mean are assigned value of alliance three SDs from mean; 2 SD winsorized = alliance scores two SDs from mean are assigned value of alliance two SDs from mean.

**Figure 3**

*Scatterplots of Alliance-Outcome Association with Five Treatments of Outliers in Community Mental Health Center Data*



*Note:* Figures display the alliance-outcome association across five treatments of outliers. Data are drawn from the community mental

health center sample ( $N = 11,029$ ). Regression line displays non-parametric loess curve that does not assume a linear association. Raw = outliers included without adjustment; 3 SD removed = alliance scores three standard deviations (SDs) from mean excluded; 2 SD removed = alliance scores two SDs from mean excluded; 3 SD winsorized = alliance scores three SDs from mean are assigned value of alliance three SDs from mean; 2 SD winsorized = alliance scores two SDs from mean are assigned value of alliance two SDs from mean.

**Supplemental Materials Table 1**

*Descriptive Statistics and Effect Sizes for Within-Therapist Alliance-Outcome Associations in University Counseling Center Sample*

Alliance Treatment	<i>n</i>	Mean	<i>SD</i>	Min	Max	Skew	Kurtosis	<i>r</i>	<i>r</i> <sub>lb</sub>	<i>r</i> <sub>ub</sub>	<i>R</i> <sup>2</sup>	<i>p</i>
Sess #2 raw	1,005	0.00	0.87	-4.71	1.67	-0.93	1.51	-0.13	-0.19	-0.06	2.22	<.001
Sess #2 3 SD removed	991	0.04	0.79	-2.55	1.67	-0.52	-0.12	-0.10	-0.16	-0.04	1.59	<.001
Sess #2 2 SD removed	969	0.09	0.73	-1.73	1.67	-0.30	-0.61	-0.09	-0.15	-0.03	1.29	.001
Sess #2 3 SD winsorized	1,005	0.01	0.85	-2.61	1.67	-0.72	0.39	-0.12	-0.18	-0.06	2.18	<.001
Sess #2 2 SD winsorized	1,005	0.03	0.80	-1.74	1.67	-0.41	-0.52	-0.12	-0.18	-0.06	2.11	<.001
Sess #3 raw	822	0.00	0.78	-3.97	1.56	-0.79	0.90	-0.08	-0.15	-0.01	1.15	.003
Sess #3 3 SD removed	816	0.02	0.74	-2.32	1.56	-0.57	-0.06	-0.08	-0.15	-0.01	1.05	.004
Sess #3 2 SD removed	786	0.09	0.66	-1.53	1.56	-0.27	-0.67	-0.05	-0.12	0.02	0.42	.077
Sess #3 3 SD winsorized	822	0.00	0.77	-2.33	1.56	-0.66	0.16	-0.08	-0.15	-0.01	1.13	.003
Sess #3 2 SD winsorized	822	0.02	0.73	-1.56	1.56	-0.40	-0.56	-0.08	-0.14	-0.01	1.02	.005
Sess #4 raw	653	0.00	0.75	-3.46	1.50	-0.93	1.60	-0.06	-0.14	0.01	0.61	.052
Sess #4 3 SD removed	647	0.03	0.69	-2.13	1.50	-0.52	-0.20	-0.05	-0.13	0.03	0.34	.150
Sess #4 2 SD removed	632	0.07	0.64	-1.45	1.50	-0.35	-0.56	-0.06	-0.14	0.02	0.37	.135
Sess #4 3 SD winsorized	653	0.01	0.72	-2.25	1.50	-0.65	0.16	-0.06	-0.14	0.02	0.53	.069
Sess #4 2 SD winsorized	653	0.02	0.69	-1.50	1.50	-0.43	-0.53	-0.06	-0.13	0.02	0.50	.080

*Note:* Min = minimum value; Max = maximum value; *r* = correlation coefficient; *r*<sub>lb</sub> = lower bound of 95% confidence interval (CI); *r*<sub>ub</sub> = upper bound of 95% CI; *R*<sup>2</sup> = proportion of variance in post-test distress explained by alliance fixed effects in multilevel models accounting for nesting within therapists; *p* = *p*-value for alliance fixed effect. Sess # = session number; raw = outliers included without adjustment; 3 SD removed = alliance scores three standard deviations (SDs) from mean excluded; 2 SD removed = alliance scores two SDs from mean excluded; 3 SD winsorized = alliance scores three SDs from mean are assigned value of alliance three SDs from mean; 2 SD winsorized = alliance scores two SDs from mean are assigned value of alliance two SDs from mean. The maximum sample size at a specific time point is lower than the total sample size as patients could have provided alliance ratings at some but not all sessions.

**Supplemental Materials Table 2**

*Descriptive Statistics and Effect Sizes for Between-Therapist Alliance-Outcome Associations in University Counseling Center Sample*

Alliance Treatment	<i>n</i>	Mean	<i>SD</i>	Min	Max	Skew	Kurtosis	<i>r</i>	<i>r</i> <sub>lb</sub>	<i>r</i> <sub>ub</sub>	<i>R</i> <sup>2</sup>	<i>p</i>
Sess #2 raw	1,005	5.88	0.32	4.00	7.00	-0.22	2.40	-0.10	-0.16	-0.03	1.17	.002
Sess #2 3 SD removed	1,005	5.88	0.32	4.00	7.00	-0.22	2.40	-0.10	-0.16	-0.03	1.17	.002
Sess #2 2 SD removed	1,003	5.88	0.31	4.75	7.00	0.18	0.34	-0.11	-0.17	-0.04	1.46	.001
Sess #2 3 SD winsorized	1,005	5.88	0.32	4.00	7.00	-0.22	2.40	-0.10	-0.16	-0.03	1.17	.002
Sess #2 2 SD winsorized	1,005	5.88	0.31	4.11	7.00	-0.16	1.94	-0.10	-0.16	-0.04	1.19	.002
Sess #3 raw	822	6.00	0.29	4.50	7.00	-0.36	1.99	-0.06	-0.12	0.01	0.57	.043
Sess #3 3 SD removed	822	6.00	0.29	4.50	7.00	-0.36	1.99	-0.06	-0.12	0.01	0.57	.043
Sess #3 2 SD removed	822	6.00	0.29	4.50	7.00	-0.36	1.99	-0.06	-0.12	0.01	0.57	.043
Sess #3 3 SD winsorized	822	6.00	0.29	4.50	7.00	-0.36	1.99	-0.06	-0.12	0.01	0.57	.043
Sess #3 2 SD winsorized	822	6.00	0.29	4.50	7.00	-0.36	1.99	-0.06	-0.12	0.01	0.57	.043
Sess #4 raw	653	6.10	0.31	4.50	7.00	-0.52	1.52	-0.08	-0.16	-0.01	1.10	.009
Sess #4 3 SD removed	653	6.10	0.31	4.50	7.00	-0.52	1.52	-0.08	-0.16	-0.01	1.10	.009
Sess #4 2 SD removed	652	6.11	0.30	5.00	7.00	-0.34	0.70	-0.09	-0.16	-0.01	1.19	.007
Sess #4 3 SD winsorized	653	6.10	0.31	4.50	7.00	-0.52	1.52	-0.08	-0.16	-0.01	1.10	.009
Sess #4 2 SD winsorized	653	6.10	0.31	4.51	7.00	-0.51	1.49	-0.08	-0.16	-0.01	1.10	.009

*Note:* Min = minimum value; Max = maximum value; *r* = correlation coefficient; *r*<sub>lb</sub> = lower bound of 95% confidence interval (CI); *r*<sub>ub</sub> = upper bound of 95% CI; *R*<sup>2</sup> = proportion of variance in post-test distress explained by alliance fixed effects in multilevel models accounting for nesting within therapists; *p* = *p*-value for alliance fixed effect. Sess # = session number; raw = outliers included without adjustment; 3 SD removed = alliance scores three standard deviations (SDs) from mean excluded; 2 SD removed = alliance scores two SDs from mean excluded; 3 SD winsorized = alliance scores three SDs from mean are assigned value of alliance three SDs from mean; 2 SD winsorized = alliance scores two SDs from mean are assigned value of alliance two SDs from mean. The maximum sample size at a specific time point is lower than the total sample size as patients could have provided alliance ratings at some but not all sessions.

**Supplemental Materials Table 3**

*Descriptive Statistics and Effect Sizes for Within-Therapist Alliance-Outcome Associations in Community Mental Health Center*

*Sample*

Alliance Treatment	<i>n</i>	Mean	<i>SD</i>	Min	Max	Skew	Kurtosis	<i>r</i>	<i>r</i> <sub>lb</sub>	<i>r</i> <sub>ub</sub>	<i>R</i> <sup>2</sup>	<i>p</i>
Sess #2 raw	10,399	0.00	4.80	-38.85	16.39	-4.38	28.11	-0.07	-0.09	-0.05	0.85	<.001
Sess #2 3 SD removed	10,234	0.47	2.90	-14.33	16.39	-0.96	5.92	-0.09	-0.11	-0.07	1.63	<.001
Sess #2 2 SD removed	10,116	0.61	2.62	-9.57	16.39	-0.38	5.53	-0.08	-0.10	-0.06	1.49	<.001
Sess #2 3 SD winsorized	10,399	0.23	3.42	-14.39	14.39	-1.73	6.50	-0.09	-0.11	-0.07	1.33	<.001
Sess #2 2 SD winsorized	10,399	0.31	2.99	-9.60	9.60	-1.34	3.35	-0.09	-0.11	-0.07	1.53	<.001
Sess #3 raw	7,968	0.00	4.88	-38.6	20.00	-4.60	30.05	-0.07	-0.10	-0.05	0.72	<.001
Sess #3 3 SD removed	7,829	0.54	2.66	-14.61	20.00	-0.25	9.38	-0.06	-0.08	-0.04	1.06	<.001
Sess #3 2 SD removed	7,772	0.63	2.45	-9.73	20.00	0.54	9.20	-0.06	-0.08	-0.04	1.17	<.001
Sess #3 3 SD winsorized	7,968	0.27	3.29	-14.65	14.65	-1.66	9.15	-0.08	-0.10	-0.06	1.06	<.001
Sess #3 2 SD winsorized	7,968	0.34	2.78	-9.77	9.77	-1.17	5.19	-0.08	-0.11	-0.06	1.25	<.001
Sess #4 raw	6,166	0.00	4.35	-38.90	20.00	-5.46	40.80	-0.08	-0.10	-0.05	0.83	<.001
Sess #4 3 SD removed	6,077	0.43	2.30	-12.92	20.00	-0.62	10.71	-0.08	-0.11	-0.06	1.55	<.001
Sess #4 2 SD removed	6,020	0.53	2.07	-8.62	20.00	0.29	11.37	-0.07	-0.10	-0.04	1.32	<.001
Sess #4 3 SD winsorized	6,166	0.23	2.77	-13.06	13.06	-1.99	9.83	-0.10	-0.12	-0.07	1.46	<.001
Sess #4 2 SD winsorized	6,166	0.30	2.41	-8.71	8.71	-1.35	5.85	-0.10	-0.12	-0.08	1.72	<.001

*Note:* Min = minimum value; Max = maximum value; *r* = correlation coefficient; *r*<sub>lb</sub> = lower bound of 95% confidence interval (CI); *r*<sub>ub</sub> = upper bound of 95% CI; *R*<sup>2</sup> = proportion of variance in post-test distress explained by alliance fixed effects in multilevel models accounting for nesting within therapists; *p* = *p*-value for alliance fixed effect. Sess # = session number; raw = outliers included without adjustment; 3 SD removed = alliance scores three standard deviations (SDs) from mean excluded; 2 SD removed = alliance scores two SDs from mean excluded; 3 SD winsorized = alliance scores three SDs from mean are assigned value of alliance three SDs from mean; 2 SD winsorized = alliance scores two SDs from mean are assigned value of alliance two SDs from mean. The maximum sample size at a specific time point is lower than the total sample size as patients could have provided alliance ratings at some but not all sessions.

**Supplemental Materials Table 4**

*Descriptive Statistics and Effect Sizes for Between-Therapist Alliance-Outcome Associations in Community Mental Health Center*

*Sample*

Alliance Treatment	<i>n</i>	Mean	<i>SD</i>	Min	Max	Skew	Kurtosis	<i>r</i>	<i>r</i> <sub>lb</sub>	<i>r</i> <sub>ub</sub>	<i>R</i> <sup>2</sup>	<i>p</i>
Sess #2 raw	10,399	38.04	1.75	23.61	40.00	-4.09	23.73	-0.05	-0.07	-0.03	0.24	<.001
Sess #2 3 SD removed	10,399	38.04	1.75	23.61	40.00	-4.09	23.73	-0.05	-0.07	-0.03	0.24	<.001
Sess #2 2 SD removed	10,293	38.16	1.28	28.50	40.00	-2.32	7.64	-0.05	-0.07	-0.03	0.16	.003
Sess #2 3 SD winsorized	10,399	38.04	1.75	23.61	40.00	-4.09	23.73	-0.05	-0.07	-0.03	0.24	<.001
Sess #2 2 SD winsorized	10,399	38.06	1.60	28.42	40.00	-3.26	14.24	-0.05	-0.07	-0.03	0.22	.001
Sess #3 raw	7,968	38.43	2.02	20.00	40.00	-4.28	22.43	-0.05	-0.07	-0.03	0.17	.007
Sess #3 3 SD removed	7,966	38.44	2.00	23.78	40.00	-4.23	21.74	-0.05	-0.07	-0.03	0.18	.006
Sess #3 2 SD removed	7,837	38.63	1.29	31.71	40.00	-2.91	10.86	-0.05	-0.07	-0.03	0.13	.020
Sess #3 3 SD winsorized	7,968	38.43	2.01	23.57	40.00	-4.24	21.79	-0.05	-0.07	-0.03	0.17	.007
Sess #3 2 SD winsorized	7,968	38.47	1.81	28.54	40.00	-3.59	14.90	-0.05	-0.07	-0.03	0.16	.010
Sess #4 raw	6,166	38.76	1.62	20.00	40.00	-4.76	32.23	-0.06	-0.08	-0.03	0.41	<.001
Sess #4 3 SD removed	6,143	38.82	1.33	31.28	40.00	-3.27	12.36	-0.05	-0.07	-0.02	0.22	.003
Sess #4 2 SD removed	6,143	38.82	1.33	31.28	40.00	-3.27	12.36	-0.05	-0.07	-0.02	0.22	.003
Sess #4 3 SD winsorized	6,166	38.76	1.56	25.39	40.00	-4.29	24.91	-0.06	-0.08	-0.03	0.41	<.001
Sess #4 2 SD winsorized	6,166	38.78	1.44	29.84	40.00	-3.42	13.60	-0.06	-0.08	-0.03	0.38	<.001

*Note:* Min = minimum value; Max = maximum value; *r* = correlation coefficient; *r*<sub>lb</sub> = lower bound of 95% confidence interval (CI); *r*<sub>ub</sub> = upper bound of 95% CI; *R*<sup>2</sup> = proportion of variance in post-test distress explained by alliance fixed effects in multilevel models accounting for nesting within therapists; *p* = *p*-value for alliance fixed effect. Sess # = session number; raw = outliers included without adjustment; 3 SD removed = alliance scores three standard deviations (SDs) from mean excluded; 2 SD removed = alliance scores two SDs from mean excluded; 3 SD winsorized = alliance scores three SDs from mean are assigned value of alliance three SDs from mean; 2 SD winsorized = alliance scores two SDs from mean are assigned value of alliance two SDs from mean. The maximum sample size at a specific time point is lower than the total sample size as patients could have provided alliance ratings at some but not all sessions.