



HHS Public Access

Author manuscript

Psychother Res. Author manuscript; available in PMC 2017 July 01.

Published in final edited form as:

Psychother Res. 2016 July ; 26(4): 387–398. doi:10.1080/10503307.2015.1026423.

A Re-Examination of Process-Outcome Relations in Cognitive Therapy for Depression: Disaggregating Within-Patient and Between-Patient Effects

Katherine E. Sasso^{a,*}, Daniel R. Strunk^a, Justin D. Braun^a, Robert J. DeRubeis^b, and Melissa A. Brotman^c

Robert J. DeRubeis: derubeis@psych.upenn.edu; Melissa A. Brotman: brotmanm@mail.nih.gov

^aDepartment of Psychology, The Ohio State University, 1835 Neil Avenue, Columbus, OH 43210

^bDepartment of Psychology, University of Pennsylvania, 3720 Walnut Street, Philadelphia, PA 19104; Tel: +1 215 573 5804

^cMood and Anxiety Disorders Program, National Institute of Mental Health, 15K North Drive, Room 208, Bethesda, MD 20892

Abstract

Objective—We previously examined alliance and therapist adherence as predictors of symptom change. Applying a new analytic strategy, we can ensure that any relations identified were not attributable to stable patient characteristics.

Method—Participants were 57 depressed Cognitive Therapy patients. We disaggregated within- and between-patient variation in process measures.

Results—Between-patients, variability in adherence to Cognitive Methods and Negotiating / Structuring predicted patients' symptom change. Within-patients, only variability in ratings of adherence to Cognitive Methods predicted next-session symptom change.

Conclusions—Relations involving between-patient process variables are potentially attributable to stable patient characteristics. However, the relation of within-patient Cognitive Methods and session-to-session symptom change cannot be attributed to stable characteristics and is consistent with a causal relationship.

Keywords

cognitive behavior therapy; depression; alliance; process research; statistical methods

In psychotherapy research, the therapeutic alliance and therapist adherence to manual-specified techniques are two of the most widely studied therapy process variables. If these process variables have a causal effect on patient outcomes, then one would expect that within-patient variability in these variables would predict symptom change (Curran & Bauer, 2011). Nonetheless, the vast majority of analytic methods employed in process research

*Corresponding author. Tel: +1 440 452 6423; fax: +1 614 688 8261; sasso.9@buckeyemail.osu.edu.

No authors have relevant conflicts of interest.

assess process predictors at a single time point (where the data are inherently between patients) or collect repeated measures of process scores, but fail to disaggregate within- and between-patient variability in these predictors. This is problematic because a relation of between-patient variability in process scores and subsequent outcome could be attributable to the confounding influence of one or more stable patient characteristics. Suppose therapists utilize different treatment procedures depending on a patient's stable characteristics and that such characteristics also influence outcome, then one could find a process-outcome relation in the absence of a true causal effect of the process variable on outcome. In contrast, a relation of within-patient variability in a predictor and subsequent outcome could not be accounted for by any between-patient differences in stable characteristics (Allison, 2005).

Curran and Bauer (2011) recently argued that a focus on the hypothesized effects of within-participant variability would allow for a more meaningful evaluation of many psychological theories. To our knowledge, no adherence studies and very few alliance studies have employed the strategies necessary to disentangle the within- and between-patient effects of these process variables (Crits-Christoph, Gibbons, Hamilton, Ring-Kurtz, & Gallop, 2011; Tasca & Lampard, 2012; Falkenström, Granström, & Holmqvist, 2013). Insofar as a process variable has a causal influence on outcome, one would expect that within-patient variability in that process variable would predict changes in outcome. With this in mind, we aim to examine process-outcome relations while disaggregating within- and between-patient variation in adherence and alliance scores using a residualized group-mean centering procedure suggested by Curran and Bauer.

Existing research does include a number of observational studies of the relationship between these process measures and outcome, with meta-analyses of the relatively large alliance-outcome literature showing a small, but reliable relation (r s ranging from .22–.28; Horvath, Del Re, Flückiger, & Symonds, 2011; Martin, Garske, & Davis, 2000) and a recent first meta-analysis of the smaller adherence-outcome literature showing a near zero and non-significant relation ($r = .02$; Webb, DeRubeis, & Barber, 2010). To be informative, meta-analytic estimates must be derived from studies featuring sound research methods. As Feeley, DeRubeis, and Gelfand (1999) have argued, a process-outcome relation can be accounted for by any of three possibilities: (1) the process variable causes change in outcome; (2) changes in outcome impact the process variable (i.e., reverse causality); or (3) a third variable exerts a causal impact on both the psychotherapy process and outcome variables (i.e., spuriousness).

In observational studies, reverse causality can be ruled out when a process variable is assessed prior to the period of time over which change in the outcome variable is observed. Although the potential for the influence of spurious effects cannot be eliminated entirely in observational studies, careful consideration of the analytic strategy one employs can reduce the potential for spurious effects to bias estimates of effects of interest. In most process-outcome studies investigators have examined a process variable (typically assessed at a single time-point) as a predictor of change in outcome over a relatively large time period, such as from an early session to the end of treatment. This often covers a timespan of many weeks or months (see Webb et al., 2012), allowing for a host of other influences on the course of symptom change to obscure any causal effect of the measured process variable.

Predictions of change from one session to the next involve a much shorter span of time, which may increase the likelihood that the causal effects of interest in psychotherapeutic interactions will be captured. Additionally, the multiple waves of process data, from each patient, required when predicting session-to-session change (e.g., Strunk, Brotman, & DeRubeis, 2010) open up the possibility of disentangling within- and between-patient variability in process measures so that the relation between symptom change and within-patient variability in a process measure can be examined. By examining within-patient variability specifically, one potential source of spuriousness, namely between-patient differences in stable traits, can be ruled out.

Distinguishing Within- and Between-Patient Variability

Recently, several quantitative experts have described the inferential benefits of disaggregating within- and between-participant effects (Allison, 2005; Curran & Bauer, 2011). In process-outcome research, there are several potentially important stable characteristics that might affect both process variables and outcome, thereby biasing the process-outcome relation of interest. If these stable characteristics are not taken into account in the research design and analytic plan, their influence will go undetected, such that the model that is tested could be mis-specified (by failing to account for confounding variables). For example, personality characteristics and personality disorders have been found to predict differential response to Cognitive Therapy CT (Sasso & Strunk, 2013; Fournier et al., 2008) and in other studies personality characteristics have been found to predict the formation of the therapeutic alliance (Zuroff et al., 2000). Thus, one could find a process-outcome relation even if no true causal relation existed, as could be the case if between-patient differences in personality characteristics introduced a spurious process-outcome relation. While one could statistically control for personality variables, there are a number of other potentially important stable between-patient characteristics that might also introduce a spurious relation. By examining within-patient variability, one can rule out not only these characteristics, but any stable patient characteristics as a rival explanation for process-outcome relationships (Allison, 2005). Moreover, if a true causal process-outcome effect exists, and its effects occur over short session-to-session intervals, a statistical relation between within-patient variability in the process variable and session-to-session outcomes should be evident.

Recently, efforts to distinguish within- and between-patient processes have been made by a number of psychotherapy researchers (see Boswell, Castonguay, & Wasserman, 2010; Crits-Christoph et al., 2011; Falkenström et al., 2013; Tasca & Lampard, 2012). In the paper with the greatest analytic similarity to our own, Falkenström and colleagues (2013) examined patient reported alliance scores at each session as predictors of overall psychological distress at the next session. They studied a heterogeneous sample of 646 patients participating in psychotherapy in a primary care setting. The duration of the psychotherapy varied naturalistically, and it tended to be brief (mean number of sessions attended was 4.6). They found a small ($\beta = -.05$) yet statistically significant effect of the within-patient variation in the alliance on symptom change from session-to-session. Their approach not only disaggregated within- and between-patient variability in process scores, but also focused on an examination of the therapeutic alliance as a predictor of subsequent session-to-session

changes in outcome (i.e., a time lagged approach). In our own previous work on the process of change in CT for depression, we have examined adherence and alliance variables as predictors of session-to-session symptom change (Strunk et al., 2010; Strunk, Cooper, Ryan, DeRubeis, & Hollon, 2012). Whereas with the approach we have employed in previous analyses we could establish the temporal precedence of the process measures relevant to symptom change, we did not examine separately the effects on symptom change of within-patient and between-patient variation in process scores.

Purpose of This Study

Given the importance of disaggregating within-patient and between-patient variation, in this paper, we present a re-analysis of our previous report examining adherence and alliance as predictors of session-to-session symptom change across the first four sessions of CT (Strunk et al., 2010). In that paper, we found two of three adherence scores significantly predicted session-to-session symptom improvements: Cognitive Methods ($r = .44$) and Negotiating/Structuring ($r = .38$). The third adherence score, Behavioral Methods/Homework, failed to predict symptom change ($r = -.04$). In analyses of the alliance, the total alliance score failed to significantly predict symptom change ($r = .15$). By disaggregating the within- and between-patient effects of adherence and alliance variables on session-to-session symptom change, in this paper we aimed to determine if the significant relations obtained by Strunk and colleagues (2010) were driven by within-patient or between-patient variation in process scores. As Falkenström and colleagues (2013) note, such clarification is key for ensuring that any process-outcome relations identified could not be accounted for by stable patient characteristics. Additionally, a key strength of our approach is its modeling of time-lagged session-to-session process-outcome relations while examining the role of both within- and between-patient variability in our predictors. Because of this advantage, we now take the effects reported here to be a more informative test of the relations of interest.

Method

Participants

Patients were 57 of 60 adults who were randomly assigned to the CT condition of a two-site (University of Pennsylvania and Vanderbilt University) clinical trial of CT, pharmacotherapy, and placebo for moderate to severe depression (DeRubeis et al., 2005). The trial was approved by Institutional Review Boards, and all participants provided written informed consent prior to participating. Our analytic procedures to obtain within-patient and between-patient component scores require at least three observations per patient. Three of the sixty patients lacked adequate data for inclusion. For additional details on sample characteristics, full inclusion and exclusion criteria, and overall clinical outcomes, see DeRubeis et al. (2005).

Therapists

CT was provided by six therapists (two of whom were female), with three therapists at each site. All of the therapists were trained in CT. Five were licensed Ph.D. level psychologists and one was a psychiatric nurse practitioner (MSN). All therapists followed the procedures

outlined in standard texts of CT for depression (Beck, Rush, Shaw, & Emery, 1979; Beck, 1995). At each site, CT therapists met weekly for 90 minutes to review ongoing cases.

Measures

Depressive symptoms—The Beck Depression Inventory-II (BDI-II; Beck, Steer, & Brown, 1996) is a widely used and well-validated 21-item self-report measure. Patients completed the BDI-II prior to each session to assess session-to-session changes in depressive symptoms.

Observer-rated process measures

Therapeutic alliance: Observer ratings of the therapeutic alliance were made using the short form of the Working Alliance Inventory (WAI-S; Horvath & Greenberg, 1989; Tracey & Kokotovic, 1989), which consists of 12 items rated on a 7-point Likert-type scale ranging from 0 ('not at all') to 6 ('thoroughly'). Although the total score, a sum of all 12 items, is often used as the primary alliance measure, subscales have been suggested (Andrusyna, Tang, DeRubeis, & Luborsky, 2001; Bordin, 1979). For our primary analyses, we used the two subscales suggested by a factor analysis of observer rated WAI items (Andrusyna et al., 2001). The *Agreement* subscale reflects the sum of 9 items and refers to the consensus between therapist and patient on the tasks and goals of treatment. The *Relationship* subscale reflects the sum of 3 items and refers to the mutual and affective bond between therapist and patient. The mean of the Cronbach's alphas across session 1 – 4 were .96 for the Agreement subscale and .83 for the Relationship subscale. To facilitate comparison of results with those obtained by Strunk et al. (2010), we also examined the total alliance score.

Therapist adherence: The Collaborative Study Psychotherapy Rating Scale (CSPRS; Hollon et al., 1988) was used to measure therapist adherence. Higher scores indicate that a therapist engaged in a manual-specified CT behavior more extensively. We used a subset of the CSPRS items to measure therapist adherence to the three subscales of therapy-specific techniques identified by Strunk and colleagues (2010). The 9-item Cognitive Methods subscale evaluates therapists' efforts to help patients identify and evaluate the accuracy of their automatic thoughts. The 8-item Negotiating/Structuring subscale evaluates therapists' efforts to structure the therapy session, collaborate with patients, and negotiate therapy content. Finally, the 5-item Behavioral Methods/Homework subscale evaluates therapists' use of behavioral strategies and therapists' assigning and reviewing of homework, including but not limited to homework involving behavioral interventions. Each subscale score reflects the average of the constituent items. The mean of the Cronbach's alphas across session 1 – 4 was .67 for Cognitive Methods, .67 for Negotiating/Structuring, and .65 for Behavioral Methods/Homework.

Number of psychotherapy process ratings and BDI-II scores—To examine psychotherapy process scores as predictors of session-to-session symptom change across early sessions, we examined process score predictors at sessions 1 – 4 and BDI-II data at sessions 1 – 5. For the BDI-II data, if all 57 patients had data at all five time points, we would have obtained 285 BDI-II scores. Due to missing data (largely attributable to patient dropout), we examined 271 BDI-II scores. For the psychotherapy process variables, of the

57 patients included, 52 had four sessions rated and five patients had three sessions rated. Thus, we had 223 of a possible 228 ratings for the 57 patients examined.

Procedure—Two advanced graduate students trained in CT independently observed and rated each therapy session tape. The raters, who were blind to outcome, provided ratings for each of the first four sessions of CT, performing these ratings sequentially. Therapy sessions were both audio and video recorded. Audio recordings were used when video recordings were unavailable or damaged (for further details, see Strunk et al., 2010).

Psychometric Properties of the Scales—Random effects intraclass correlation coefficients (ICCs) were calculated to evaluate the inter-rater reliabilities of the scales as rated independently by the two raters. ICC estimates were calculated using only the 57 patients examined in this report and corrected for the number of raters. ICCs for the three adherence scores were: .60 for Cognitive Methods, .71 for Negotiating/Structuring, and .70 for Behavioral Methods/Homework. ICCs for the alliance scores were: .79 for the WAI total score, .79 for the Agreement subscale, and .76 for the Relationship subscale. For subsequent analyses, we used the average of the ratings made at each session.

Analytic Approach—For our primary analyses, we decomposed within-patient and between-patient variation in each process measure following the procedures recommended by Curran and Bauer (2011). For a given process variable, this involved utilizing a separate ordinary least squares (OLS) regression for each patient in which session (mean centered) was entered as a predictor of the process variable. We then examined the deviations in each session-specific process score from the model-implied values of the process score for that session in order to obtain the within-patient process score at each session (i.e., we retained the session-specific residuals from a regression in which time served as a predictor of the process variable, with regressions computed separately for each patient). To obtain the between-patient score for a given process variable, we retained the patient specific intercepts from these models. By mean centering time, these intercepts reflect the value of the process variable at the mid-point of the sessions examined. We can present this procedure formally as a one-process variable predictor regression equation estimated separately (case by case) for each patient in the sample. For a given process variable (e.g., Cognitive Methods), this would be represented as:

$$\text{Cognitive_Methods}_{ti} = b_{0i} + b_{1i}\text{Session}_{ti} + e_{ti} \quad (1)$$

where *Cognitive_Methods_{ti}* is the session-specific value of the process variable, or the patient's Cognitive Methods score at a given session. *Session_{ti}* is the measure of time, *b_{0i}* represents the between-patient process score (i.e., the intercept), *b_{1i}* is the slope of Cognitive Methods across the sessions. Finally, *e_{ti}* reflects within-patient variation in Cognitive Methods, or the session-specific residual from the model. As Curran and Bauer (2011) have argued, this approach results in de-trended estimates of within-patient variation from which the time trend underlying one's time-varying predictor has been taken into account. Thus, de-trending avoids violating the assumption of stationarity (i.e., the assumption of no change in the mean level of a repeated measures predictor across time; Falkenström et al., 2013).

After decomposing within- and between-patient variation in each of the process variables, we then examined the within- and between-patient scores (e_{ti} and b_{0i} respectively, both obtained from equation 1). For a given process variable, we examined these scores simultaneously as predictors of session-to-session symptom change in a repeated measures regression model using SAS proc mixed (i.e., the within- and between-patient Cognitive Methods scores were examined as predictors in one model, within- and between-patient Negotiating/Structuring scores examined in another model, etc.). The model can be represented by a single-level equation as illustrated in equation 2¹

$$BDI_{t+1i} = \beta_0 + \beta_1(BDI_{ti}) + \beta_2(\text{Site}_i) + \beta_3(b_{0i}) + \beta_4(e_{ti}) + \varepsilon_{ti} \quad (2)$$

where BDI_{t+1i} reflects a given patient's BDI-II scores at sessions 2 – 5. β_0 represents the intercept of BDI-II scores at sessions 2 – 5. β_1 reflects the effect of patients' BDI-II score at a given session (t) on their BDI-II score at the next-session (BDI_{t+1i}). β_2 reflects the effect of site on patients' BDI-II scores across sessions 2 – 5. β_3 reflects the effect of between-patient variability in patients' process score (i.e., patients' person-specific intercepts from equation 1, or b_{0i}) on patients' BDI-II scores across sessions 2 – 5. β_4 represents the effect of within-patient variability in patients' process score at a given session (i.e., patients' session-specific residual from equation 1, or e_{ti}) on patients' BDI-II score at the next-session. Finally, ε_{ti} represents the residual term or error from the model. Using the repeated line in SAS, we specified session as the repeated variable and also specified subject = patient (see Kiernan, Tao, & Gibbs, 2012 for a discussion of how this is similar to specification of patients' random intercept).

Except for the fact that these models disaggregate within-patient and between-patient variance in the process predictors, they are similar to those reported by Strunk et al. (2010). No random effects were specified other than a random therapist effect in the model in which we tested that effect specifically. To facilitate interpretation, signs have been adjusted so that a positive relationship always indicates that higher process scores (within- or between-patient) are associated with more positive outcomes (i.e., lower symptom scores as indicated by the BDI-II). For our primary repeated measures regression models (i.e., the model depicted in equation 2), we examined several different covariance structures (*viz.* autoregressive (AR), heterogeneous AR, compound symmetry, toeplitz, and unstructured) and selected the structure that provided the best fit. Compound symmetry was determined to have the best fit across all information criteria indices. For all models, we specified maximum likelihood (ML) as the estimation method for the covariance parameters and between-within as the method for computing the denominator degrees of freedom. As we did in Strunk et al. (2010), we included site as a covariate in all of our session-to-session repeated measures regression models (as illustrated in equation 2). Although site was not a significant predictor in any of our session-to-session models (all $ps > .06$), we retained it because outcomes varied markedly by site in the primary paper reporting on this trial (see DeRubeis et al., 2005). To test site by process variable interactions, we examined our primary model (equation 2), with the following additional predictors: $b_{0i} * \text{site}$ and $e_{ti} * \text{site}$.

¹We also examined within-patient and between-patient scores individually as predictors in the same repeated measures regression models. All significant predictors in the combined models were also significant when examined individually.

We found no evidence for such interactions for any of the psychotherapy process scores we examine (all $ps > .24$).

Results

Prior to examining process-outcome relations, we examined the M and SD of patients' BDI-II and raw process variable scores at each session (see Table 1). For patients' raw psychotherapy process scores and their BDI-II scores, we calculated ICCs to estimate the proportion of the variation in the scores that was accounted for by patients (i.e., between-patient variability). ICCs for the alliance total score and two subscales ranged from .67–.69, indicating that the majority of the variance in alliance scores (i.e., 67–69%) was between-patient variation. The remaining variation (31–33%) was within-patient. For the adherence subscales, ICCs estimating between-patient variance were: .28 for Cognitive Methods, .52 for Negotiating/Structuring, and .37 for Behavioral Methods/Homework. These ICCs indicate that for two of the three adherence subscales (i.e., Cognitive Methods and Behavioral Methods/Homework), the majority of variation was within-patient variability. For Negotiating/Structuring, the percent of the variance that was between-patient was roughly comparable to the percent of variance that was within-patient (i.e., .52 vs. .48, respectively). The ICC for patients' BDI-II scores (sessions 1 – 5) was .78, indicating that the majority of variation (78%) was between-patient variation.

As described in the Analytic Approach section, we obtained within-patient and between-patient scores from a series of regression models in which session number served as a predictor of a given process variable for each patient². Ms and SDs for the within-patient and between-patient scores of each process variable are provided in Table 2. For between-patient scores, calculating Ms and SDs was straightforward as each patient had one between-patient score for each process variable. For within-patient scores, each patient had four scores (a score for each of the first four sessions) for each process variable. To provide an index of variability in these scores, we obtained patient-specific SDs of the within-patient scores for each process variable (i.e., for a given process variable and patient, a single value representing the SD of his/her within-patient scores across sessions). The average and range of these patient-specific SDs are reported in Table 2. Because residuals are parameterized to sum to zero, the mean for all within-patient scores was zero. At the suggestion of a reviewer, we also examined our primary models while excluding patients with very low levels of within-patient variability. In analyses excluding patients with a within-patient SD in the lowest 5th percentile, the significance of all predictors was unchanged from the results reported below.

²In order to examine the average relationship between time and each process variable over the first 4 sessions, we used Hierarchical Linear Models to test for the fixed effect of time (i.e., session) on each raw psychotherapy process variable. We modeled random intercepts and slopes. The primary test of interest in these models was the fixed effect of session. Results showed that the slope of Cognitive Methods was significantly positive ($b = .12$, $SE = .03$, $t(56) = 3.75$, $p = .0004$) and the slope of Negotiating/Structuring was significantly negative ($b = -.08$, $SE = .03$, $t(56) = -2.96$, $p = .005$). For three variables (viz., Behavioral Methods / Homework, Agreement, and Relationship), the slope did not differ significantly from zero (all $ps > .4$). For these three variables, we also examined the between and within-patient scores derived from a person mean-centering approach (i.e., equation 1 without session included as a predictor) as predictors of session-to-session symptom change. The significance of predictors in these models did not differ from that obtained in our primary models.

Prior to examining the within- and between-patient scores as predictors of symptom change, we also examined the relationships among these scores. As shown in Table 3, correlations among the between-patient scores (shown below the diagonal) were significantly correlated for all but one of the between-patient process variables. The correlation of Behavioral Methods/Homework and Cognitive Methods was the one non-significant relation. For within-patient scores, we examined the correlations among these scores separately at each of the first four sessions. We then converted these correlation coefficients to z-scores using Fisher's r to z transformations. We averaged these z-scores and converted those means back to r -values. These values, which reflect the average correlation of within-patient process scores over the first four sessions, are provided above the diagonal in Table 3. Compared with the between-patient scores, the within-patient process scores were substantially less related, with only 3 of 10 correlations achieving significance at one or more sessions. Two of these three relations were significant at each of the first four sessions: (1) the relation of Cognitive Methods and Negotiating/Structuring; and (2) the relation of the Agreement and Relationship alliance subscales.

Within-Patient and Between-Patient Process Scores Predicting Session-to-Session Symptom Change

Prior to examining the within-patient and between-patient process-outcome relations, we considered the role of therapists in our models. We calculated the ICC for therapist in order to determine the variability in patients' BDI-II scores that was attributable to between-therapist differences. The ICC was .12, indicating that 12% of variation in patients' BDI-II scores is attributable to between-therapist variability. Then, we followed the procedures outlined by West, Welch, and Galecki (2007) to test the fit of models in which therapist was and was not specified as a random effect. Modeling therapist as a random effect failed to improve model fit (difference in likelihood ratio statistic = 0.00, $df = 1$, $p = .50$, $-2 \log = 1298.3$) and also failed to change the statistical significance of any predictors in the models reported below. Thus, therapist was not included as a random effect in subsequent analyses.

For each of the three adherence scales, we conducted a repeated measures regression model (see equation 2) in which both the within-patient and between-patient components for a given adherence scale were examined as predictors (i.e., one model for each adherence scale). The within-patient and between-patient Cognitive Methods scores were the numerically strongest predictors of next-session symptom improvement, with higher values of both Cognitive Methods scores significantly predicting greater next-session symptom improvements (Between: $b = 1.58$, $SE = .51$, $t(54) = 3.14$, $p = .003$; Within: $b = 3.58$, $SE = 1.14$, $t(156) = 3.15$, $p = .002$). For within-patient Cognitive Methods scores, our finding shows that on average, when patients' Cognitive Methods score at a given session was one point higher than expected, the model predicted improvement in their next-session BDI-II score was 3.58 points greater than expected (after controlling for their current session BDI-II score). To make the points more meaningful, we considered the predicted BDI improvement for a 2 point change on one item included in the Cognitive Methods scale. For example, a therapist who did not inquire about the evidence relevant to evaluating any automatic thought would obtain a "0, not at all" score on the examining evidence item. In another session, this therapist might engage in a clear instance of asking about such evidence and be

rated a “2, some.” Such a 2 point change in a within-patient Cognitive Methods item would predict a .8 point improvement in next-session BDI-II scores. This finding for the between-patient Cognitive Methods scores shows that for every one point higher a patient’s average Cognitive Methods score, patients reported a 1.58 point improvement in BDI-II scores on average (alternatively, a 2 point difference in any of the constituent Cognitive Methods items predicted an average .4 point BDI-II improvement at the next session).

For Negotiating/Structuring, the between-patient component emerged as a significant predictor of symptom change, whereas there was a non-significant trend for the within-patient component to predict symptom change (Between: $b = 1.02$, $SE = .46$, $t(54) = 2.21$, $p = .03$; Within: $b = 2.23$, $SE = 1.28$, $t(156) = 1.74$, $p = .08$). As the between-patient finding indicates, for every one point higher a patient’s Negotiating/Structuring score, patients reported a 1.02 point improvement in BDI-II scores on average (alternatively, a 2 point difference in any of the constituent Negotiating/Structuring items predicted an average .3 point BDI-II improvement at the next session). Neither the between-patient nor the within-patient components of the Behavioral Methods/Homework scale significantly predicted symptom change (Between: $b = -.05$, $SE = .46$, $t(54) = -.11$, $p = .92$; Within: $b = .49$, $SE = .87$, $t(156) = .56$, $p = .58$). In Figure 1, we compare standardized beta weights for within- and between-patient adherence scores to those obtained when Strunk and colleagues (2010) examined the raw scores as predictors of session-to-session symptom change in the same repeated measures regression model.

As we utilized different alliance subscales than those examined by Strunk et al. (2010), we first used Strunk and colleague’s analytic approach to examine patients’ raw Agreement and Relationship scores individually as predictors of session-to-session symptom change in our repeated measures regression model³. Neither subscale emerged as a significant predictor (Agreement: $b = .05$, $SE = .04$, $t(157) = 1.34$, $p = 0.18$; Relationship: $b = .02$, $SE = .12$, $t(157) = .19$, $p = .85$).

Using the model described in equation 2 separately for each alliance variable, we found that neither the between-patient Relationship nor the within-patient Relationship scores significantly predicted symptom change (Between: $b = .05$, $SE = .12$, $t(54) = .43$, $p = .67$; Within: $b = -.54$, $SE = .52$, $t(156) = -1.03$, $p = .31$). For the Agreement scores, only the between-patient score emerged as a trend predictor (Between: $b = .07$, $SE = .04$, $t(54) = 1.99$, $p = .05$; Within: $b = -.23$, $SE = .16$, $t(156) = -1.48$, $p = .14$). Specifically, a one point higher between-patient Agreement score was associated with .07 points of BDI-II improvement on average.⁴ In Figure 2, we compare the standardized beta weights for within- and between-patient alliance scores to those obtained when Strunk et al. (2010) examined the raw scores as predictors of session-to-session symptom change in the same repeated measures regression model.

³As these raw Agreement and Relationship scores were not partitioned into within- and between-patient components, there was no minimum requirement for the number of observations per patient. Given this, and in order to keep consistency with Strunk et al. (2010), for these analyses only we used the complete sample ($N = 60$).

⁴We also examined the within-patient and between-patient alliance total scores simultaneously as predictors in this model. Neither the within-patient nor the between-patient alliance total scores predicted session-to-session symptom change ($b = -.19$, $SE = .13$, $t(156) = -1.46$, $p = .15$; $b = .05$, $SE = .03$, $t(54) = 1.65$, $p = .10$).

Examining Predictors in a Combined Model Simultaneously

In an effort to examine whether the relations we obtained above may be partly accounted for by other process scores that emerged as significant predictors of symptom change, we tested one model that included all significant and trend-level predictors ($p < .10$) from our combined models and examined them simultaneously as predictors of session-to-session symptom change (viz., Cognitive Methods –within and between, Negotiating/Structuring–within and between, and Agreement–between), while controlling for the main effect of patient’s current BDI-II scores and site. When all five scores were entered as predictors, within-patient Cognitive Methods remained a significant predictor of session-to-session symptom change ($b = 3.38$, $SE = 1.30$, $t(155) = 2.60$, $p = .01$) and between-patient Cognitive Methods was reduced to the level of a non-significant trend ($b = 1.40$, $SE = .72$, $t(52) = 1.95$, $p = .06$). Neither the within-patient Negotiating/Structuring, the between-patient Negotiating/Structuring, nor the between-patient Agreement scores remained significant predictors ($ps > .75$).

Discussion

To our knowledge, this is the first effort to examine the relation of within-patient variability in adherence and outcome in any psychotherapy and the first such examination of within-patient variability in the alliance in CT for depression specifically. In our results, therapist adherence to Cognitive Methods was the sole process variable for which within-patient variation significantly predicted next-session symptom change. As we noted, when patients’ Cognitive Methods scores at a current session were one point higher than expected, patients experienced an average improvement in their next-session BDI-II score 3.58 points greater than expected (after controlling for their current session BDI-II score). These results clarify that the results reported by Strunk et al. (2010) for Cognitive Methods were not driven by between-patient differences in stable characteristics alone.

For between-patient process variables, we identified Cognitive Methods and Negotiating/Structuring as the only predictors of symptom change in our sole process variable predictor models (per equation 2). However, neither of these between-patient variables remained significant when examined simultaneously in our combined model. As these two between-patient process variables were moderately correlated ($r = .61$), the relations of the variables with outcome identified in our sole process variable predictor models were plausibly attributable to shared variance. Even had these relations remained significant, as between-patient scores, such relations could potentially be attributable to stable patient characteristics rather than a causal relation of the process variable and outcome.

To understand our findings in context, it is important to consider the degree of within- and between-patient variability we observed in patients’ raw process scores (as indicated by the relevant ICCs). We found striking differences in the amount of between-patient variability in the process variables we examined. Across the alliance scales, the variation observed was overwhelmingly between-patient. In contrast, for two of the three therapist adherence scales (i.e., Cognitive Methods and Behavioral Methods/Homework), the majority of variability was within-patient. For the third adherence scale (i.e., Negotiating/Structuring), within- and between-patient variability were roughly comparable. While we are not aware of such

estimates for adherence in the literature, our estimates of within vs. between-patient variability in the alliance scores are higher than those reported in previous studies (ICCs of .39 and .54 in Weiss, Kivity, & Huppert, 2014 and Falkenström et al., 2013, respectively).

Insofar as there was limited within-patient variability in raw process variable scores, our power to detect a within-patient process-outcome relation was reduced. From session-to-session, if a process variable shows limited within-patient variation, it is unlikely to predict variation in symptoms. The small amount of within-patient variation in the alliance scores could be a function of either a true lack of within-patient variability in the alliance or the insensitivity of the observer-rated WAI to true within-patient alliance fluctuations. It is also possible that our estimates differ from previous studies due to the nature of the treatment provided, the patients in the sample, or the time period over which variability was assessed. Compared to treatments in which the alliance is hypothesized to be a primary mechanism of therapeutic gains, it may not be unreasonable to expect less within-patient variability in alliance scores in CT for depression. In this study, we examined within-patient variability over session-to-session intervals. While we have no reason to suspect this, it remains possible that within-patient variation as assessed over other intervals would exhibit stronger relations with outcome. Weiss and colleagues (2014) have recently shown that systematic variability in the alliance can be detected within a single session.

Although it has not been a major focus in our analyses, it is important to consider therapist effects and how they might ultimately relate to analyses such as those presented in this paper. In our analyses, modeling therapist as a random variable failed to improve model fit. Nonetheless, the ICC for therapist indicated that 12% of the variation in early session-to-session outcomes was accounted for by therapist differences. Such therapist effects merit future investigation, but ideally in a larger sample optimized for the study of such effects. In addition, we suspect therapist effects may be larger outside the context of a trial in which therapists were selected for their ability to provide CT. In fact, in a naturalistic sample of 80 therapists, Baldwin, Wampold, and Imel (2007) found therapists with higher average alliance scores achieved superior outcomes. This between-therapist difference in outcomes could be accounted for by two possibilities. Either the alliance is causally related to outcome or therapist differences in outcome arise for another reason (e.g., therapists who utilize more of the true active ingredients of treatment also tend to form better alliances). If a true causal alliance-outcome relation exists, one would expect within-patient variability in the alliance to predict outcome. Repeated measures allowing for the disaggregation of within- and between-patient scores are needed to test this possibility. While our analyses failed to find such a within-patient alliance-outcome relation, we look forward to future studies with larger samples of therapists in which the researchers could both examine therapist effects and parse within- and between-patient variability.

Limitations

We note some key limitations of this study. First, in the absence of an experimental manipulation of the adherence or alliance, we cannot establish a definitive causal relationship between these variables and outcome. Nonetheless, in our analysis of within-patient variation in process variables, we were able to rule out any stable third variables as

alternative explanations for the relations we identified. What we still cannot rule out is the possibility that an unanticipated *time-varying* characteristic accounted for the relation of within-patient Cognitive Methods and next-session symptom change.

Second, as our analyses focus on early sessions of CT for depression specifically, the process-outcome relations we obtained may not generalize to later sessions or to other forms of treatment. While our models were well-suited to examining the presumed immediate effects of process variables, we did not examine the possibility that “delayed effects” of therapeutic processes influenced symptom change multiple sessions later (i.e., as these effects do not conform to our session-to-session model). While our analytic strategy does not test such effects, they could be of interest insofar as early therapist behaviors set the stage for maximizing the benefits patients derive from therapeutic techniques used later in the course of treatment. Third, our results may be limited to CT of depression. Process-outcome relations may vary across psychotherapies.

Fourth, variability described as within-patient was calculated as any variability that was not between-patients; thus, measurement error may well have contributed disproportionately to the within-patient components. Future replication of our finding of within-patient process-outcome relations is needed to provide evidence of the validity of these scores as indicators of the underlying constructs. Fifth, while the ICCs and alphas for our process measures were comparable to previous research, some of these values raise the possibility that measurement error prevented us from detecting effects of interest. Sixth, we have not used simulated datasets to evaluate the accuracy or precision of estimates derived from our approach in a sample of this size. Future work is needed to examine the power of models such as those we employed. Seventh, while modeling therapist as a random effect did not improve model fit in this study, examining therapist effects may well be important in contexts in which therapist effects are deemed likely or are evident. Because we did not include therapist in our primary models, therapist differences in process variables could have accounted for some portion of the variability in between-patient process variables. Finally, our approach did not evaluate possible interactions of therapist and patient characteristics. While testing these interactions may not be feasible in many research contexts, we would be interested to see such tests in large datasets with adequate variability in therapist and patient characteristics.

Conclusion

In this study, we re-examined early session process-outcome relations in the CT condition of a clinical trial of treatments for moderate to severe depression (DeRubeis et al., 2005). In this re-examination, we disaggregated within-patient and between-patient variation in therapist adherence and the therapeutic alliance. Cognitive Methods emerged as the sole within-patient predictor of next-session symptom change. As this relation was obtained specifically with a measure of within-patient variation, we can be confident that it was not due to any stable patient characteristics. The only other significant process-outcome relations identified in our primary models involved between-patient scores; had we not disaggregated within-patient and between-patient variability, we would have failed to recognize that the relation of these variables and outcome was due to mean differences across patients. If process variables have a true causal impact on outcome, one would expect

a within-patient process-outcome relation. We encourage future psychotherapy process researchers to collect and analyze repeated measures data across multiple sessions so that we might better account for confounding patient-characteristics when examining the relation of process variables and therapeutic outcomes.

Acknowledgments

The data presented in this article were collected as part of a clinical trial conducted between 1998 and 2003, which was supported by the National Institute of Mental Health under Grants MH55877 (R10), MH55875 (R10), MH01697 (K02), and MH01741 (K24).

References

- Allison, PD. Fixed effects regression methods for longitudinal data using SAS. Cary, NC: SAS Institute, Inc.; 2005.
- Andrusyna TP, Tang TZ, DeRubeis RJ, Luborsky L. The factor structure of the Working Alliance Inventory in cognitive-behavioral therapy. *Journal of Psychotherapy Practice and Research*. 2001; 10:173–178. [PubMed: 11402080]
- Baldwin SA, Wampold BE, Imel ZE. Untangling the alliance-outcome correlation: Exploring the relative importance of therapist and patient variability in the alliance. *Journal of Consulting and Clinical Psychology*. 2007; 75:842. [PubMed: 18085902]
- Beck, JS. *Cognitive therapy: Basics and Beyond*. New York, NY: Guilford Press; 1995.
- Beck, AT.; Rush, AJ.; Shaw, BF.; Emery, G. *Cognitive Therapy of Depression*. New York, NY: Guilford Press; 1979.
- Beck, AT.; Steer, RA.; Brown, GK. *Manual for the Beck Depression Inventory*. 2nd ed. San Antonio, TX: The Psychological Corporation; 1996.
- Bordin ES. The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, Research, and Practice*. 1979; 16:252–260.
- Boswell JF, Castonguay LG, Wasserman RH. Effects of psychotherapy training and intervention use on session outcome. *Journal of Consulting and Clinical Psychology*. 2010; 78:717–723. [PubMed: 20873906]
- Crits-Christoph P, Gibbons MBC, Hamilton J, Ring-Kurtz S, Gallop R. The dependability of alliance assessments: The alliance-outcome correlation is larger than you might think. *Journal of Consulting and Clinical Psychology*. 2011; 79:267–278. doi:<http://dx.doi.org/10.1037/a0023668>. [PubMed: 21639607]
- Curran PJ, Bauer DJ. The Disaggregation of Within-Person and Between-Person Effects in Longitudinal Models of Change. *Annual Review of Psychology*. 2011; 62:583–619.
- DeRubeis RJ, Hollon SD, Amsterdam JD, Shelton RC, Young PR, Salomon RM, Gallop R. Cognitive therapy vs medications in the treatment of moderate to severe depression. *Archives of General Psychiatry*. 2005; 62:409–416. [PubMed: 15809408]
- Falkenström F, Granström F, Holmqvist R. Therapeutic alliance predicts symptomatic improvement session by session. *Journal of Counseling Psychology*. 2013; 60:317–328. [PubMed: 23506511]
- Feeley M, DeRubeis RJ, Gelfand LA. The temporal relation of adherence and alliance to symptom change in cognitive therapy for depression. *Journal of Consulting and Clinical Psychology*. 1999; 67:578–582. [PubMed: 10450629]
- Fournier JC, DeRubeis RJ, Shelton RC, Gallop R, Amsterdam JD, Hollon SD. Antidepressant medications v. cognitive therapy in people with depression with or without personality disorder. *British Journal of Psychiatry*. 2008; 192:124–129. [PubMed: 18245030]
- Hollon, SD.; Evans, MD.; Auerbach, A.; DeRubeis, RJ.; Elkin, I.; Lowery, A.; Piasecki, JM. Development of a system for rating therapies for depression: Differentiating cognitive therapy, inter-personal psychotherapy and clinical management pharmacotherapy. Nashville, TN: Vanderbilt University; 1988. Unpublished manuscript

- Horvath AO, Del Re AC, Flückiger C, Symonds D. Alliance in individual psychotherapy. *Psychotherapy relationships that work: Evidence-based responsiveness*. 2011; 2:25–69.
- Horvath AO, Greenberg LS. Development and validation of the working alliance inventory. *Journal of Counseling Psychology*. 1989; 2:223–233.
- Kiernan, K.; Tao, J.; Gibbs, P. SAS Global Forum. Cary, NC: SAS Institute Inc.; 2012. Tips and strategies for mixed modeling with SAS/STAT® procedures; p. 332-2012. Retrieved from <http://support.sas.com/resources/papers/proceedings12/332-2012.pdf>
- Martin DJ, Garske JP, Davis MK. Relation of the therapeutic alliance with outcome and other variables: A meta-analytic review. *Journal of Consulting and Clinical Psychology*. 2000; 68:438–450. [PubMed: 10883561]
- Sasso KE, Strunk DR. Thin slice ratings of client characteristics in intake assessments: Predicting symptom change and dropout in cognitive therapy for depression. *Behaviour Research and Therapy*. 2013; 51:443–450. [PubMed: 23747583]
- Strunk DR, Brotman MA, DeRubeis RJ. The process of change in cognitive therapy for depression: Predictors of early inter-session symptom gains. *Behaviour Research and Therapy*. 2010; 48:599–606. [PubMed: 20362978]
- Strunk DR, Cooper AA, Ryan ET, DeRubeis RJ, Hollon SD. The process of change in cognitive therapy for depression when combined with antidepressant medication: Predictors of early inter-session symptom gains. *Journal of Consulting and Clinical Psychology*. 2012; 80:730–738. [PubMed: 22774791]
- Tasca GA, Lampard AM. Reciprocal influence of alliance to the group and outcome in day treatment for eating disorders. *Journal of Counseling Psychology*. 2012; 59:507–517. [PubMed: 22963649]
- Tracey TJ, Kokotovic AM. Factor structure of the working alliance inventory. *Psychological Assessment*. 1989; 1:207–210.
- Webb CA, DeRubeis RJ, Barber JP. Therapist adherence/competence and treatment outcome: A meta-analytic review. *Journal of Consulting and Clinical Psychology*. 2010; 78:200–211. [PubMed: 20350031]
- Webb CA, DeRubeis RJ, Dimidjian S, Hollon SD, Amsterdam JD, Shelton RC. Predictors of patient cognitive therapy skills and symptom change in two randomized clinical trials: The role of therapist adherence and the therapeutic alliance. *Journal of Consulting and Clinical Psychology*. 2012; 80:373–381. [PubMed: 22468907]
- Weiss M, Kivity Y, Huppert JD. How does the therapeutic alliance develop throughout cognitive behavioral therapy for panic disorder? Sawtooth patterns, sudden gains, and stabilization. *Psychotherapy Research*. 2014; 24(3):407–418. [PubMed: 24320241]
- West, BT.; Welch, KB.; Galecki, AT. *Linear mixed models: A practical guide using statistical software*. Boca Raton, FL: Chapman Hall/CRC Press; 2007.
- Zuroff DC, Blatt SJ, Sotsky SM, Krupnick JL, Martin DJ, Sanislow CA, Simmens S. Relation of therapeutic alliance and perfectionism to outcome in brief outpatient treatment of depression. *Journal of Consulting and Clinical Psychology*. 2000; 68:114–124. [PubMed: 10710846]

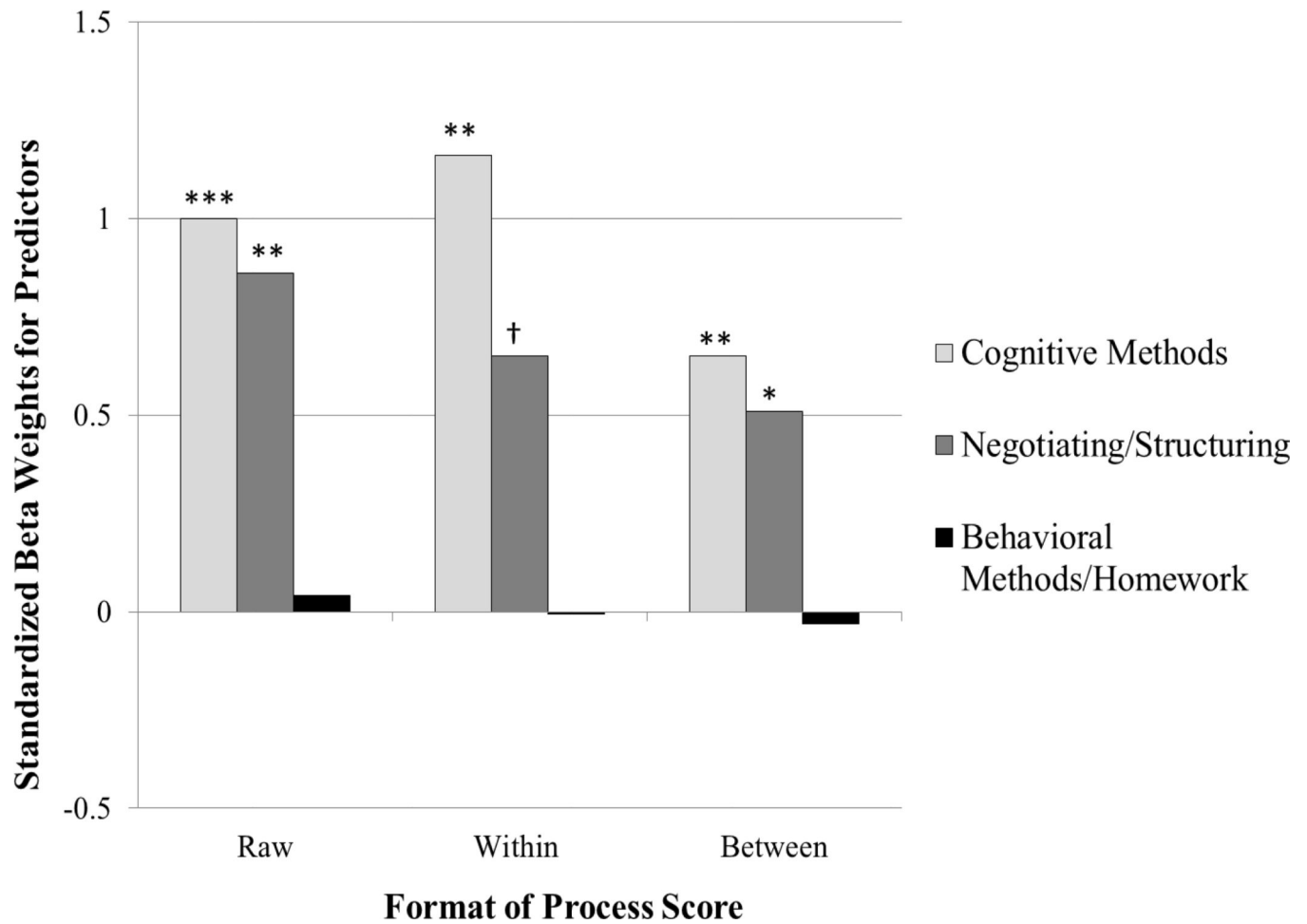


Figure 1. Beta Weights for Standardized Adherence Scores as Predictors of Session-to-Session BDI-II Scores

Note: Relation of three types of therapist adherence and session-to-session symptom change in CT for depression. For each type of adherence, two separate repeated measures regression models were implemented: one in which the raw score was examined as a predictor (as reported by Strunk et al., 2010), and another in which within- and between-patient process scores were examined simultaneously as predictors. For figures 1 and 2 only, we standardized raw, within-, and between-patient process scores to a $M = 0$ and $SD = 1$.

† $p < .1$, * $p < .05$, ** $p < .01$, *** $p < .001$.

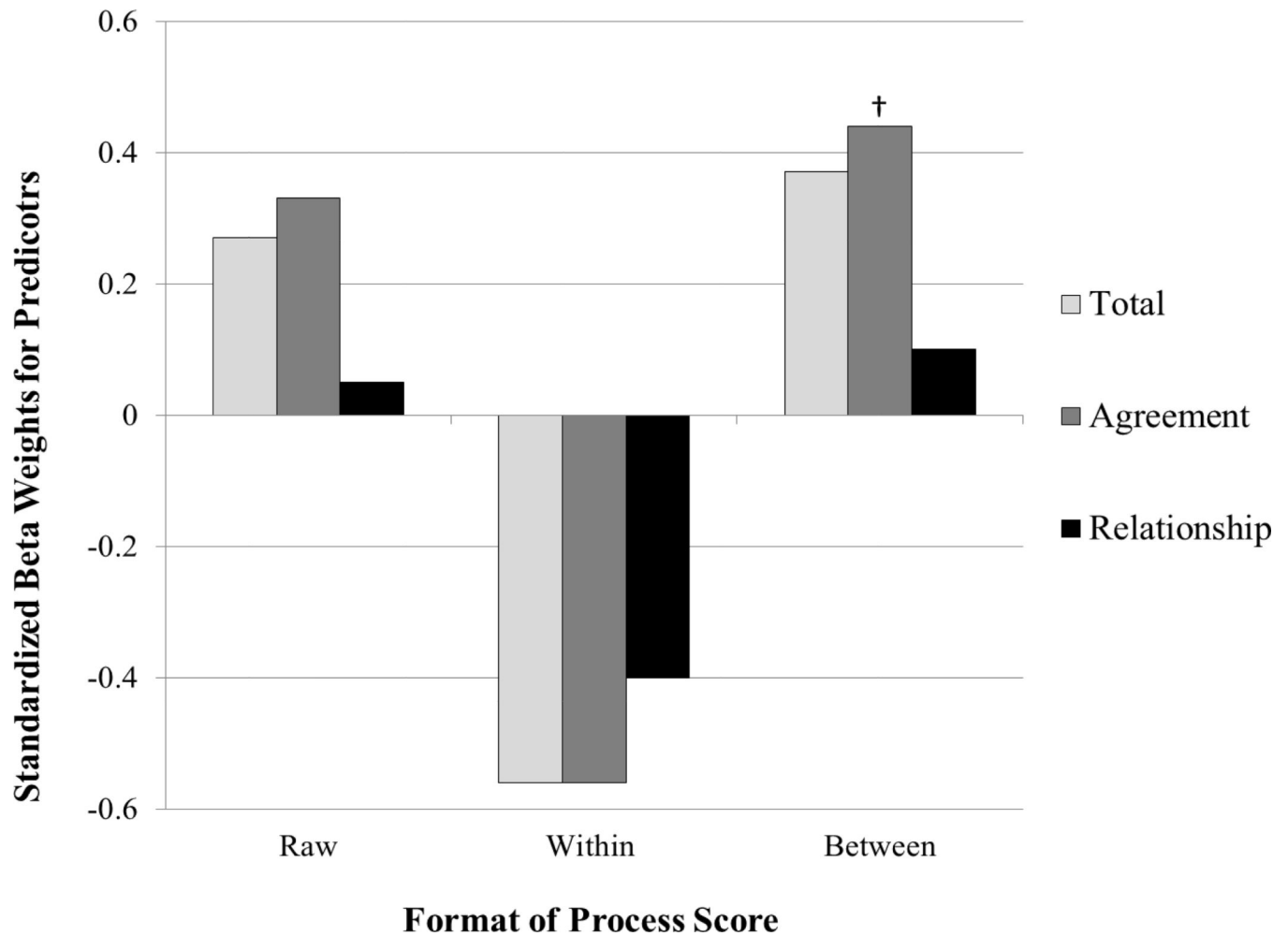


Figure 2. Beta Weights for Standardized Alliance Process Scores as Predictors of Session-to-Session BDI-II Scores

Note: Relation of three aspects of therapeutic alliance (i.e., total score and two subscales) and session-to-session symptom change in CT for depression. For each aspect of the alliance, two separate repeated measures regression models were implemented: one in which the raw score was examined as a predictor, and another in which the within and between scores were examined simultaneously as predictors. For figures 1 and 2 only, we standardized raw, within-, and between-patient process scores to a $M = 0$ and $SD = 1$.

Means and Standard Deviations of Patients' Raw Process Variable and BDI-II Scores at Each Session

Table 1

	Sessions				
	1	2	3	4	5
Cognitive Methods	.87 (.44)	1.00 (.55)	1.19 (.70)	1.19 (.62)	--
Negotiating / Structuring	2.38 (.58)	2.11 (.61)	2.17 (.63)	2.07 (.65)	--
BM/H	1.26 (.82)	1.66 (.96)	1.64 (.78)	1.36 (.64)	--
Agreement	36.54 (6.25)	35.79 (6.91)	37.05 (7.53)	36.82 (7.33)	--
Relationship	12.12 (2.07)	11.63 (2.08)	12.00 (2.44)	12.10 (2.20)	--
BDI-II	28.33 (9.02)	25.82 (9.27)	24.69 (10.30)	22.67 (11.06)	22.86 (10.58)

Note: Means are reported outside of parentheses, standard deviations reported in parentheses.

BM/H = Behavioral Methods / Homework. BDI-II = Beck Depression Inventory II.

Table 2

Descriptive Statistics: Means and Standard Deviations of the Within-and Between-Patient Scores

	Between-Patient Scores	Within-Patient Scores	
	M (SD)	M (average SD)	Range of SD
<u>Adherence</u>			
Cognitive Methods	1.06 (.41)	0 (.33)	.02-.73
Negotiating / Structuring	2.17 (.50)	0 (.30)	.02-.67
Behavioral Methods /Homework	1.50 (.62)	0 (.42)	.03-1.34
<u>Alliance</u>			
Relationship	11.93 (1.91)	0 (.71)	0-1.95
Agreement	36.53 (6.08)	0 (2.35)	0-8.23

Note. For the SD of within-patient scores, we calculated the SD for each patient and averaged those SDs. The range of patient specific SDs is provided in the right most column. As patients only had one between-patient score per process variable, there were no patient specific SDs for between-patient scores.

Correlations among Between-Patient (below diagonal) and Within-Patient (above diagonal) Process Variables across Sessions 1–4

Table 3

	1	2	3	4	5
1. Cognitive Methods	—	.47 ^(4/4)	-.00 ^(0/4)	.07 ^(0/4)	.13 ^(0/4)
2. Negotiating/Structuring	.61*	—	.00 ^(0/4)	.09 ^(0/4)	.18 ^(0/4)
3. BM/H	.03	.43*	—	.12 ^(0/4)	.16 ^(1/4)
4. Relationship	.42*	.55*	.39*	—	.71 ^(4/4)
5. Agreement	.54*	.61*	.41*	.86*	—

Note. Correlations among between-patient process variables are provided below the diagonal.

* $P < .05$.

Above the diagonal, values are the average correlations of within-patient process variables across sessions. Fractions in superscript indicate at how many of the first four sessions the relationship was significant.

BM/H = Behavioral Methods / Homework.