

# Therapeutic Alliance Predicts Symptomatic Improvement Session by Session

Fredrik Falkenström, Fredrik Granstrom and Rolf Holmqvist

**Linköping University Post Print**



N.B.: When citing this work, cite the original article.

Original Publication:

Fredrik Falkenström, Fredrik Granstrom and Rolf Holmqvist, Therapeutic Alliance Predicts Symptomatic Improvement Session by Session, 2013, Journal of counseling psychology, (60), 3, 317-328.

<http://dx.doi.org/10.1037/a0032258>

This article may not exactly replicate the final version published in the APA journal. It is not the copy of record.

Copyright: American Psychological Association

<http://www.apa.org/>

Postprint available at: Linköping University Electronic Press

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-96424>

Running head: Alliance Predicts Symptomatic Improvement

Therapeutic Alliance Predicts Symptomatic Improvement Session  
by Session

Fredrik Falkenström<sup>1,2</sup>, Fredrik Granström<sup>2</sup> & Rolf Holmqvist<sup>1</sup>

<sup>1</sup>Department of Behavioural Sciences and Learning  
Linköping University  
Sweden

<sup>2</sup>Center for Clinical Research Sörmland  
Uppsala University  
Sweden

Acknowledgements:

The authors would like to thank Bruce Wampold and Scott Baldwin for help with this study.

Corresponding author address: Fredrik Falkenström, Lustigkullevägen 17, SE 616 33 Åby,  
Sweden, phone: +46 736 814567, fax: +46 155 247931, email: [Fredrik.Falkenstrom@liu.se](mailto:Fredrik.Falkenstrom@liu.se)

### Abstract

**Background:** The therapeutic alliance has been found to predict psychotherapy outcome in numerous studies. However, critics maintain that the therapeutic alliance is a by-product of prior symptomatic improvements. Moreover, almost all alliance research to date has used differences between patients in alliance as predictor of outcome, and results of such analyses does not necessarily mean that improving the alliance with a given patient will improve outcome (i.e. a within-patient effect).

**Method:** In a sample of 646 patients (76 % women, 24 % men) in primary care psychotherapy, the effect of working alliance on next session symptom level was analyzed using multi-level models. The Clinical Outcomes in Routine Evaluation – Outcome Measure was used to measure symptom level and the patient version of the Working Alliance Inventory – Short form revised was used to measure alliance.

**Results:** There was evidence for a reciprocal causal model, in which the alliance predicted subsequent change in symptoms while prior symptom change also affected the alliance. The alliance effect varied considerable between patients. This variation was partially explained by patients with personality problems showing stronger alliance effect.

**Conclusions:** These results indicate that the alliance is not just a by-product of prior symptomatic improvements, even though improvement in symptoms is likely to enhance the alliance. They also point to the importance of therapists paying attention to ruptures and repair of the therapy alliance. Generalization of results may be limited to relatively brief primary care psychotherapy.

*Keywords:* therapeutic alliance, psychotherapy outcome, therapy process, multi-level models, prediction.

## **Therapeutic Alliance Predicts Symptomatic Improvement**

### **Session by Session**

The concept of the therapeutic alliance stems from psychoanalytic theory (e.g. Greenson, 1965), in which the therapeutic alliance is distinguished from the transference aspects of the therapy relationship (i.e. aspects of the patient's relationship to the therapist that are distorted by past experiences). Later definitions have been more pan-theoretic, with Bordin's definition (Bordin, 1979) being one of the most influential. In Bordin's view the alliance is defined as agreement on the goals and tasks of therapy in the context of a positive affective bond between patient and therapist. In contrast to early psychoanalytic authors, Bordin thought the alliance to be curative in itself rather than just a precondition for therapeutic work. Empirical research has found alliance ratings to be related to treatment outcome. A recent meta-analysis based on more than two hundred studies (Horvath, Del Re, Fluckiger, & Symonds, 2011) reported a mean effect of alliance on outcome corresponding to a correlation of  $r = .275$  (95 % CI between .25 and .30). This effect did not depend on therapy orientation, alliance measure, rating perspective (self report, therapist report, observer measure), or time of assessment. The empirical finding of a correlation between alliance ratings early in treatment and final outcome is often interpreted as supporting the theoretical idea that a good alliance is causing better outcome, and that it is important to work directly with the alliance – especially if the alliance is poor (e.g. Safran & Muran, 2000).

However, the causal interpretation of these correlational findings has been questioned (e.g. Crits-Christoph, Gibbons, & Hearon, 2006; DeRubeis, Brotman, & Gibbons, 2005; DeRubeis & Feeley, 1990; Kazdin, 2005; Strunk, Cooper, Ryan, DeRubeis, & Hollon, 2012; Tang & DeRubeis, 1999). Alliance ratings could be influenced by confounders, the most likely

being prior symptom improvements and pre-treatment patient characteristics. In the present study we are particularly interested in the first of these, the possible influence of prior improvement on the alliance-outcome relationship. If the patient reports reduced symptoms between session one and session three, this symptom reduction might contribute to a positive alliance rating in that session and to positive outcome at the end of treatment. The association between early positive alliance ratings and favorable final outcome would thus be a pseudo-association and the real association would be between improvement prior to the alliance measurement and subsequent improvement. Moreover, many alliance studies have tested the relationship between alliance and outcome over the full course of treatment, thus including within the dependent variable such change that occurred before the alliance was measured (Strunk, Brotman, & DeRubeis, 2010).

For the relationship between alliance and symptom change to be credibly interpreted as causal, it is necessary to show that symptom improvement follows after measurement of the alliance and to control for the possibility of reverse causation (i.e. prior symptom improvement explaining association between alliance and subsequent outcome). We have been able to find twelve studies fulfilling these criteria. The results of these studies give a mixed picture of the robustness of the alliance as a predictor of outcome when that relationship is adjusted for prior symptom change. Seven studies (Barber, Connolly, Crits-Christoph, Gladis, & Siqueland, 2000; Crits-Christoph, Gibbons, Hamilton, Ring-Kurtz, & Gallop, 2011; De Bolle, Johnson, & De Fruyt, 2010; Gaston, Marmar, Gallagher, & Thompson, 1991; Klein et al., 2003; Tasca & Lampard, 2012; Zuroff & Blatt, 2006) report significant associations between therapeutic alliance and subsequent symptom change, controlling for prior symptom change. Five studies (DeRubeis & Feeley, 1990; Feeley, DeRubeis, & Gelfand, 1999; Puschner, Wolf, & Kraft, 2008;

Strunk et al., 2010; Strunk et al., 2012) report no association between alliance and subsequent symptom change when prior change was controlled.

Among the five studies showing no significant relationship between alliance and outcome, two (DeRubeis & Feeley, 1990; Feeley et al., 1999) had very small samples ( $N = 25$  and  $32$ , respectively). These studies were probably underpowered to find an effect even if it existed in the population. A third study (Strunk et al., 2010) was somewhat larger ( $N = 60$ ), and there were also two (Puschner et al., 2008; Strunk et al., 2012) that can be considered large ( $N = 259$  and  $176$ , respectively). The lack of a significant effect of the alliance can thus not be completely explained by lack of power.

Four out of five studies showing no relationship between alliance and outcome seem to have been done by the same research group (DeRubeis & Feeley, 1990; Feeley et al., 1999; Strunk et al., 2010; Strunk et al., 2012), studying the relationship between alliance and symptom change in cognitive therapy for depression. It also seems that in all these studies the alliance was measured by observer measures rather than patient or therapist report. The fifth study (Puschner et al., 2008) is difficult to compare to the other studies mentioned, mainly because it was done on comparatively long treatments (the average length of treatment was 61 sessions). In this study early symptom change consisted of change during the 15 first sessions, and relationships between alliance and symptom change was studied between four phases of treatment.

Among the studies showing a significant relationship between alliance and outcome after control for prior symptom change, three (Barber et al., 2000; Crits-Christoph et al., 2011; Gaston et al., 1991) were small to medium sized ( $N = 86$ ,  $45$ , and  $54$ , respectively), while the remaining four (De Bolle et al., 2010; Klein et al., 2003; Tasca & Lampard, 2012; Zuroff & Blatt, 2006) were large ( $N = 567$ ,  $367$ ,  $238$ , and  $191$ , respectively). Patient samples were more mixed than in

the studies showing no alliance effect, although depression was the most common diagnosis. Treatment methods were also more mixed, including both CBT and PDT treatments.

Most of the above mentioned studies used the alliance measured at a single time point early in treatment as predictor of subsequent outcome. The alliance variable in these cases is a between-subjects variable (i.e. there is only variation between patients, no variation within subjects because the variable is only measured once). As Tang and DeRubeis (1999) point out, a more interesting way of studying psychotherapy process data is to study associations between variables within individual time-courses. Moreover (and more importantly), it is a logical fallacy to generalize a between-person effect to an individual (Curran & Bauer, 2011). The fact that there is a between-subjects relationship between alliance and outcome cannot be interpreted as evidence that improving the alliance for a given patient (a within-patient effect) improves outcome. In fact, it is quite possible for opposite relationships to be true for within- and between-person effects. An example used by Curran & Bauer is that for a given person, there is an increased risk of experiencing a heart attack while exercising (positive within-person effect). Still, people who in general exercise more have a lower risk of heart attack (negative between-person effect). Similarly, in alliance research it is possible that the between-patient relationship between the alliance at a given time point and subsequent outcome is a proxy for some patient variable (e.g. temperament, diagnosis, etc) that is related to outcome and that also make some patients experience the alliance as better than others. In this scenario, working to enhance the alliance may not improve outcome, because the alliance is just a proxy for something else.

An alternative to a between-patient design is to use a time-varying predictor variable. This enables the researcher to study relationships between variables from session to session. The time-varying predictor may be lagged, i.e. shifted one or more steps forward in time, to make it

possible to predict future values of the dependent variable. The use of time-varying covariates, however, raises several methodological complications. First, when a time-varying predictor is used, the within- and between-patient parts of variation in the predictor need to be disaggregated (Curran & Bauer, 2011), otherwise the resulting regression coefficient will represent an aggregation of between- and within-patient effects. Methods for disaggregating a time-varying predictor include using some form of person-mean centering or separating the between- and within-person effects within the statistical model using latent variables (Curran, Lee, Howard, Lane, & MacCallum, 2012).

In addition to the necessity of disaggregating between- and within person variation in the time-varying predictor, statistical models for testing time-series relationships rest on the assumption of stationarity, meaning that there can be no significant changes in the mean level or variance over time (e.g. King, King, McArdle, Shalev, & Doron-LaMarca, 2009). Because in a treatment study both process and outcome variables are likely to change with time, variables either need to be transformed to become stationary or the statistical model needs to separate the session-to-session effects from the growth curves or trajectories across the whole of treatment.

Among the above mentioned alliance studies, Crits-Christoph et al. (2011) and Tasca and Lampard (2012) tested within-patient effects of repeated measures of the alliance on symptom change to the next session. Crits-Christoph et al. (2011) explored temporal relations between symptom and alliance change from session to session. They found that increase in alliance scores from the previous session was related to symptomatic improvement to the next session. They found symptom change to predict later alliance change late, but not early, in treatment. These authors used first-differences (i.e. subtracting each score from the previous score on the same variable). This is a method commonly used in time-series analysis to achieve stationarity. This



method is also likely to reduce between-patient variability, thus isolating the within-patient effect, although we are unaware of any research directly testing this. Tasca and Lampard (2012), using Latent Change Score Modeling, found evidence for reciprocal causal effects; that is, change in alliance predicted change in symptoms to the next session, but symptom change also predicted alliance change to the next session. The latent change score model achieves stationarity by estimating a latent random time slope for each variable to capture the mean trajectory across treatment for each patient, while simultaneously estimating separate latent change scores to capture change from session to session. Similarly, the between-patient effect is separated from the within-patient effect by the estimation of a random intercept for each variable.

Two more studies (Strunk et al., 2010; Strunk et al., 2012) analyzed relationships between repeated measures of both alliance and symptoms in the early phase of cognitive therapy for depression. Neither of these studies found any significant effect of the alliance on later symptom change. However, none of them seem to have implemented adequate methods to distinguish within- and between patient levels of effects. The results of these studies are thus hard to interpret since they are likely to reflect a combination of within- and between-person effects.

We predict that in our data, high alliance scores after a given session will predict lower symptom scores immediately before the next session, when taking a possible influence of symptoms on alliance into account.

## **Method**

### **Participants**

The data collection was conducted at primary care services in two regions in Sweden, with a total population of about 600 000 people. Psychologists, social workers, and other staff

providing counseling or psychological help at health care services and psychiatric youth clinics, were asked to participate. A total of 83 therapists took part in the study. Of these, 62% were social workers and 28% were psychologists. The mean age was 48.5 (range 29 to 64) and 14.4 years (range 1 to 39) was the average time working within psychological therapy. Ninety-four percent were women. The mean number of patients treated by each therapist was 15 ( $SD = 12$ , range 1-50).

The therapists were asked to invite all patients who started treatment within a 6 month period (November 2009 to April 2010) to take part in the study. Of 1431 patients, 1096 delivered at least one self-report questionnaire. Demographic information was available for between 75 to 80 percent of all patients. This information is summarized in Table 1. As Table 1 shows, the mean age was 37.3 years (median 35,  $SD = 14.3$ , range 14-88), 74 % were women, and 92 % were born in Sweden. 61 % were employed, 14 % were students, 13 % were unemployed and 3 % were on parental leave. 56 % were living with a partner while 35 % lived alone. The most common problems were anxiety (47 %), relationship problems (35 %), depression (34 %), grief (20 %), work related problems (19 %) and somatic problems (13 %). The majority of patients had more than one problem (92 %). All participating patients gave informed consent. The study was approved by the Regional Ethical Review Board of Linköping (nr M72-09).

### **Treatments**

Treatments were delivered as usual in each primary care setting, and most were brief. The mean length of treatment was 4.6 ( $SD = 4.0$ ; Median = 4) sessions, with a range between 1 and 37. This is very similar to what has been reported from primary care routine practice in the US (Hansen, Lambert, & Forman, 2002) and UK (Stiles, Barkham, Mellor-Clark, & Connell, 2008). Information on treatment type was available for 62 % of treatments. 34 % of these were ticked

with more than one orientation. The most common therapy types were supportive (30 %), psychodynamic (24 %), CBT (18 %), crisis intervention (15 %), cognitive (15 %), behavioral (9 %), relational (9 %), existential (7 %), systemic (7 %), and interpersonal (6 %). It was not uncommon that more than one type was marked as true, therefore the percentages add up to more than one hundred.

## Measures

### **Clinical Outcomes in Routine Evaluation – Outcome Measure** (Evans et al., 2002).

The CORE-OM is a patient self-report measure consisting of 34 items measuring psychological distress experienced during the preceding week, on a five-point scale ranging from “Not at all” to “Most or all the time”. The items cover four major problem areas; subjective wellbeing, problems/symptoms, life functioning and risk (to self or others). The scoring is problem-oriented in that higher scores indicate greater distress. The instrument has shown good internal- and test-retest- reliability (0.75-0.95), convergent and discriminant validity and sensitivity to change. In the present study, the CORE-OM was completed immediately before each session and only the total score, which has a possible range between 0 and 40, was used. This total score was calculated according to standard CORE procedure as the mean of all 34 items multiplied by 10.

### **Working Alliance Inventory – Short form revised** (Hatcher & Gillaspay, 2006). The Working Alliance Inventory is based on Bordin’s pan-theoretical definition of the alliance, consisting of three components: bond, tasks and goals. The original instrument (Horvath & Greenberg, 1989) consists of 36 items scored on a 7-point Likert scale. The items are made up from the three alliance components bond, tasks and goals. The instrument has shown adequate reliability and validity, although the three subscales are strongly intercorrelated which might question the distinctness of the three components. In the current study, the revised short form

(Hatcher & Gillaspay, 2006), consisting of 12 items, was scored by the patient immediately after each session. This version has also shown good reliability and validity. The possible range of this variable is between 1 and 7.

### **Statistical analyses**

Longitudinal multilevel models (e.g. Raudenbusch & Bryk, 2002) were used, where repeated measurements (level 1) were nested within patients (level 2). When testing within-person effects, predictor variables need to be centered around each patient's mean in order to isolate the within-person effect (Raudenbusch & Bryk, 2002). Otherwise the parameter estimate will be an amalgam of between- and within-person effects. In addition, the statistical models for time-series relationships need to control for non-stationarity, because in treatment studies the requirement for no significant change of means over time is clearly unrealistic. Moreover, autoregressive effects (the effect on a variable of previous measurements of the same variable) need to be accounted for (Curran & Bollen, 2001).

In the Structural Equations Modeling literature there are several models suitable for these purposes (Bollen & Curran, 2004; Curran et al., 2012; McArdle, 2009). However, for our purposes a multilevel approach was deemed more suitable, since multilevel models offer a better way of handling unbalanced designs with a different number of measurements per person (Raudenbush, 2001; Singer & Willett, 2003). We therefore chose to adjust for non-stationarity using a form of residualized group mean centering proposed by Curran and Bauer (2011). Specifically, we estimated a separate least squares (OLS) linear regression for each patient, in which each variable (CORE-OM and WAI-S) was separately regressed on time, saving the residuals as new variables. The residuals were used as "detrended" estimates of the within-patient variation on WAI-S and CORE-OM, from which the linear time trend was removed in

order to achieve stationarity. Group mean centering was accomplished simultaneously, since residuals are parameterized to have a mean of zero. This transformation method requires at least three observations per individual, because with fewer observations prediction will be perfect and there will be no residuals.

To control for autoregression, a lagged version of the dependent variable was included as a predictor in all models. This has the added advantage that because the model is adjusted for the effect of the previous session value of the dependent variable, what is left to predict is change in the dependent variable from session to session (Curran & Bollen, 2001).

We then estimated three models: In Model 1, WAI-S was used as a time-varying covariate predicting change in CORE-OM to the next session. This was accomplished by lagging the residualized group mean centered WAI-S variable. The base model was a single level equation:

$$CORE - OM_{ti} = \beta_1(CORE - OM)_{t-1i} + \beta_2(WAI - S)_{t-1i} + e_{ti}. \quad (1)$$

where  $\beta_1$  is the autoregression coefficient for CORE-OM,  $\beta_2$  is the cross-lagged coefficient for the effect of WAI-S, and  $e_{ti}$  is the residual term. Because of the centering of the dependent variable, the intercept has been constrained to zero in this model.

Model 2 tested if change in WAI-S from session t-1 to session t could be predicted by the symptom level from the previous week CORE-OM. Because the CORE-OM was filled out immediately before each session, and asked for symptoms as remembered from the past week, and the WAI-S was filled out immediately after the session and asked for alliance during the session, these two measures were non-overlapping in time with CORE-OM preceding WAI-S.

The base model was:

$$WAI - S_{ti} = \beta_1(WAI - S)_{t-1i} + \beta_2(CORE - OM)_{ti} + e_{ti}. \quad (2)$$

This model is similar to the previous one, with the exception that WAI-S is the dependent variable and the CORE-OM predictor is not lagged. Finally, these two models were put together into Model 3, a bivariate model where autoregression and cross-lagged effects were modeled for both WAI-S and CORE-OM simultaneously.

$$\begin{aligned} CORE - OM_{ti} &= \beta_1(CORE - OM)_{t-1i} + \beta_2(WAI - S)_{t-1i} + e_{1ti} \\ WAI - S_{ti} &= \beta_3(WAI - S)_{t-1i} + \beta_4(CORE - OM)_{ti} + e_{2ti}. \end{aligned} \quad (3)$$

Note that this model has two dependent variables, in contrast to previous models that were univariate. The two dependent variables are estimated simultaneously as one model using Maximum Likelihood Estimation. Here  $\beta_1$  is the autoregression coefficient for CORE-OM,  $\beta_2$  is the cross-lagged coefficient for the effect of WAI-S on next session CORE-OM, and  $e_{1ti}$  is the residual term for the CORE-OM equation.  $\beta_3$  is the autoregression for WAI-S,  $\beta_4$  is the effect of CORE-OM for the preceding week on the current session WAI-S, and  $e_{2ti}$  is the residual term for WAI-S.

Mplus 7.0 (L. K. Muthén & Muthén, 1998-2012) was used for these analyses, mainly because of the flexibility and ease with which this program can handle models with more than one dependent variable. Base models were tested using standard model fit criteria (Model Chi-square test, RMSEA, CFI, SRMR). We then extended the base model by testing whether adding a level 2 random slope for any of the coefficients improved model fit. These models are examples of multilevel path analysis (B. Muthén & Asparouhov, 2011). Absolute model fit criteria are not available for multilevel models, so these extended models were instead compared to the base model using relative model fit criteria (AIC and BIC).

## Missing data

Although multilevel modeling accommodates unbalanced designs with comparable ease, the models are based on the assumption that observations are missing at random (MAR). This means that missing data is allowed to be related to covariates and to the dependent variable at other occasions, but not to the dependent variable at the dropout occasion (e.g. Enders, 2011; Gallop & Tasca, 2009). In a naturalistic dataset such as the present one, this assumption is unlikely to be met. For example, Baldwin, Berkeljon, Atkins, Olsen, and Nielsen (2009) found that random coefficients for time were linearly related to length of treatment in an unbalanced naturalistic dataset comparable to the present one. Differences in length of treatment are not mainly due to treatment dropout, but rather reflect the fact that patients reach a “good enough level of functioning” after a different number of sessions. However, the statistical models will still treat all observations after treatment termination for patients with shorter treatments than the longest one as missing data.

Missing-not-at-random (MNAR) modeling can be used to test if there are MNAR mechanisms in the data. If there is evidence for MNAR it is possible to perform sensitivity analyses to evaluate if that biases results. Because the value of the dependent variable at the dropout occasion is by definition unknown, there is no direct way of testing the MAR assumption. However, there are ways of indirectly testing it. Two classes of models are most common; pattern-mixture (Hedeker & Gibbons, 1997; Little, 1993) and selection models (Diggle & Kenward, 1994; Wu & Carroll, 1988).

The pattern-mixture approach tests if the parameter estimates depend on missing data by estimating the model separately in subgroups with different missing data patterns. A weighted average across subgroups can be calculated to get an overall estimate that can be

compared to the MAR based results as a sensitivity analysis (Enders, 2011; Hedeker & Gibbons, 1997). If the weighted average from the pattern-mixture model differs from the MAR analysis, then the MAR results are likely to be biased. We used the Hedeker and Gibbons (1997) approach to estimate the pattern-mixture model. This means that we included  $k-1$  dummy-coded pattern coefficients as level 2 predictors of the random coefficients. Because few patients attended more than 10 sessions, treatments between 11 and 15 sessions long were combined into one group and treatments 16 sessions or longer were combined into one. Because only patients attending at least three sessions were included in the analyses (due to the transformations described previously), this means that there were 10 patterns in the pattern-mixture analysis (3-10 sessions separately, 11-15 sessions in one group, and 16-37 sessions in the last group). Nine dummy-coded pattern variables were created, with the tenth group (treatments 16 sessions or longer) was used as reference category. We were unable to test the impact of the pattern coefficients on the autoregressions because of non-positive definiteness of the first-order product matrix when all random effects were regressed on all pattern coefficients simultaneously. We therefore chose to simplify by only regressing the cross-lagged coefficients (which are the theoretically most important ones anyway) on the pattern variables. To test if the pattern-mixture model significantly improved model fit we performed a likelihood ratio test.

Selection models offer a method for bias-correction by means of adding a statistical model for the missing data process to the substantive model being tested. For our purposes, the selection model developed by Diggle and Kenward (1994) seemed especially well-suited. This model predicts the probability of missingness from the repeated measures of the outcome variable at the previous occasion and at the same occasion. While the pattern-mixture model estimates the impact of missing data on a between-patient level, the selection model estimates



impact of missingness on level 1. Because the value of the outcome variable at the dropout occasion is unknown, this model utilizes the assumption of multivariate normality for the repeated measures variables. This assumption makes it possible to estimate the probability of missingness depending on the value of the outcome variable at the dropout occasion even though that value is unknown (Enders, 2011).

In order to test the impact of the values of the repeated measurements of CORE-OM and WAI-S on data missingness, we first added a missing observation at the end of each patient's time-series. We then created two dummy coded variables for data missingness, one for CORE-OM and one for WAI-S. Observed values were coded "0" and missing observations were coded "1". Finally, we estimated Model 3 with a submodel for data missingness included on the within-patient level. This model predicted data missingness for CORE-OM by the values of CORE-OM and CORE-OM lag1, using logistic regression. The same was done for WAI-S. Significant effects of the lagged variables on data missingness would indicate MAR processes, while a significant effect of the same session variable would indicate MNAR.

## **Results**

### **Preliminary analyses and descriptive statistics**

Table 2 shows the number of observations, means, standard deviations, and ranges for the CORE-OM and WAI-S. As can be read from Table 2, most patients had very brief treatments. The mean initial values of the CORE-OM were about one and a half points lower than the mixed clinical sample described by Evans et al. (2002). The WAI-S values were compared to descriptive statistics provided by Busseri and Tyler (2003). The session 4 means and standard deviations in our study were almost identical to those reported by the authors above.

The intraclass correlation for the CORE-OM was .55, showing that 55 % of variation in this measure was at the between-patient level and the rest (45 %) was within-patient variation.

For WAI-S, the intraclass correlation was .54.

### **Alliance predicting symptom change to the next session**

Table 3 shows parameter estimates for Model 1. The single level model showed acceptable model fit, as indicated by a non-significant  $\chi^2$  test despite a relatively large sample. The model  $\chi^2$  test is the only way of testing model fit, and is generally regarded as more important than approximate fit indices such as the RMSEA, CFI, and SRMR (e.g. Barrett, 2007; Hayduk, Cummings, Boadu, Pazderka-Robinson, & Boulianne, 2007). In this case all approximate fit indices except for the CFI showed excellent model fit. Adding random effects for the between-patient differences in within-patient effects significantly improved model fit, as shown by reductions in the values of the information criteria AIC and BIC.

Parameter estimates showed that there was a statistically significant effect of WAI-S on the following session CORE-OM. Higher therapeutic alliance scores at the end of each session predicted lowered symptom levels at the start of the next session. On average, an increase of one point on the WAI-S after a given session lead to a decrease of .32 points on the CORE-OM to the following session. This is a small effect, corresponding to a standardized effect of  $\beta = .05$  (standardized estimates are only available for single level models, but since unstandardized estimates were very similar between single- and two-level models we can assume that standardized estimates also were very similar). However, there was considerable variation between patients in this effect. The variance estimate was 0.99, indicating that there were subgroups in which the effect of the alliance was considerably stronger.

### **Symptom level predicting alliance**

Model fit information and parameter estimates for Model 2 are shown in Table 4. The model Chi<sup>2</sup> test was significant ( $p = .04$ ), indicating ill fit. All approximate fit indices were acceptable. Generally, parameter estimates for a model showing significant ill fit according to the model Chi<sup>2</sup> test should only be interpreted with great caution (Kline, 2011). However, as with Model 1, adding random effects for between-patient variance in regression coefficients significantly improved model fit as shown by large reductions in AIC and BIC. It is thus likely that the significant Chi<sup>2</sup> test for the single level model was due to the assumption of no between-patient variation in regression coefficients. Parameter estimates showed that symptom level reported before a given session predicted the alliance in the same session significantly. Higher symptom level at the start of the session predicted worse alliance during the session and vice versa.

### **Combined model: Alliance predicting symptom change to next session while controlling for symptom change from previous session**

Table 5 shows model fit information and parameter estimates for the combined model (Model 3). As with Model 2, the Chi<sup>2</sup> test for the Model 3 was significant ( $p = .01$ ), but adding random effects improved model fit significantly. As can be read from the table, therapeutic alliance predicted symptom change to the next session even when the effect of change in symptoms from the previous week on alliance was controlled for. In fact, all parameter estimates were essentially identical to the ones obtained in Models 1 and 2, indicating that the effects of WAI-S on next session CORE-OM and the effect of CORE-OM on WAI-S are essentially independent. As in Model 1, the mean effect of the alliance was small, but there was a large variation between patients in this effect.

### Missing data analyses

We only tested the impact of missing data on Model 3, since this was the most inclusive model. We first examined influence statistics to see if there were cases that were overly influential. Two multivariate outliers with level 2 Cook's distance values larger than 1 were identified (Snijders & Bosker, 1999). As a precaution we re-estimated all models without these two patients, but there were no differences in results. There was a slight tendency for patients with longer treatments to be more influential, but this is to be expected since they contribute more data.

**Pattern-mixture model.** The -2 log likelihood for Model 3 was 18355.3 and for the pattern-mixture model 18329.9. The reduction in -2 log likelihood was thus 25.4, which with 18 degrees of freedom (nine pattern coefficients for each of the two cross-lagged slopes) is not significant. However, we still calculated the weighted average for the effect of WAI-S on subsequent CORE-OM change across subgroups based on missing data patterns. The weighted average was  $b = .32$ ,  $t = 2.46$ ,  $p = .01$ , which is almost exactly the same as the one found for the MAR based analysis. When inspecting the individual coefficients of the pattern-mixture model, it seemed that there was a tendency for the alliance effect to be smaller for the shortest treatments. No other pattern was apparent.

**Stratified model.** We also tested a model in which a continuous treatment length variable was entered as level 2 moderator of all the random coefficients, similar to what Baldwin et al. (2009) called a stratified model. This model resembles the pattern-mixture model, but instead of estimating the model separately for different lengths of treatment, treatment length is used as a continuous variable that is assumed to interact linearly with the random effects. Treatment length

significantly moderated all the random effects except for the cross-lagged effect of WAI-S on CORE-OM, indicating that this particular effect did not depend linearly on treatment length.

**Selection model.** Results showed significant effects of CORE-OM ( $b = .07, t = 2.42, p = .02$ ), WAI-S ( $b = -0.51, t = 2.92, p = .004$ ), and WAI-S lag1 ( $b = -0.24, t = 2.66, p = .008$ ) on data missingness, while the effect of CORE-OM lag1 was not significant ( $b = 0.01, t = 0.37, p = .71$ ). This is evidence for significant MAR processes for both CORE-OM and WAI-S, and MNAR for WAI-S. However, the estimated effect of WAI-S lag1 on subsequent CORE-OM change for the selection model was very similar to Model 3 without the selection model addition ( $b = 0.35, t = 2.92, p = .003$ ). The negative coefficients for WAI-S and WAI-S lag1 means that a better alliance leads to lower probability of missing data on the WAI-S at the same and next session, respectively, and the positive coefficient for CORE-OM means that higher symptom level leads to higher probability of missing data on the CORE-OM at the same session. Although this seems plausible, Enders (2011) cautions against substantive interpretation of the MNAR coefficients (i.e. the same session coefficients), because they are based on untestable distributional assumptions.

Taken together, the MNAR sensitivity analyses showed some evidence for MNAR processes, but there were no indications that missing data had biased the estimates for the effect of the alliance on subsequent CORE-OM change.

### **Therapist effects**

There has been considerable interest in therapist effects in recent years, and in alliance research researchers have found the mean alliance level for therapists to be a stronger predictor of outcome than differences in alliance between patients (Baldwin, Wampold, & Imel, 2007; Crits-Christoph et al., 2009; Zuroff, Kelly, Leybman, Blatt, & Wampold, 2010). The within-

patient effect of alliance on subsequent symptom change is not theoretically expected to vary between therapists, but we nevertheless wanted to check the robustness of our results in the face of therapist variation. We ran a three-level model in which the four regression coefficients of Model 3 were allowed to vary not just between patients but also between therapists. A likelihood ratio test comparing the three-level with the two-level model was borderline significant ( $\text{Chi}^2 = 9.46(4), p = .05$ ), although none of the random coefficients for level three were significant (all  $p$ -values  $> .13$ ). The average effect of the alliance on subsequent symptom change was unchanged by the inclusion of random therapist effects ( $b = .33, t = 2.24, p = .02$ ).

### **Subscale analyses**

The subscales of WAI-S turned out to be strongly intercorrelated in this sample. The Goal and Task subscales were particularly strongly correlated ( $r = .93$ ), while Task and Bond and Goal and Bond were slightly less strongly correlated ( $r = .73$  and  $r = .70$ , respectively). The correlations among transformed scales were somewhat smaller; Task and Goal correlated  $r = .82$ , Task and Bond  $r = .63$ , and Goal and Bond  $r = .58$ . A principal component analysis produced only one component with eigenvalue  $> 1$ , so it did not appear possible to create orthogonal subscales on the basis of principal components. We therefore found no better alternative than to conduct separate analyses for the three subscales, and to compare the effects of WAI-S subscales on subsequent CORE-OM change across analyses. We only tested subscale effects on Model 3. When tested separately, both Task and Bond subscales significantly predicted subsequent CORE-OM scores (Task:  $b = -.25, t = 2.25, p = .02$ ; Bond:  $b = -.32, t = 2.59, p = .01$ ), while Goal was non-significant ( $b = -.16, t = 1.41, p = .16$ ). However, regression coefficients were not significantly different from each other, as indicated by largely overlapping confidence intervals.

### **Moderator analyses**

Because of the large between-patient variation in the within-patient effect of the working alliance on subsequent change in CORE-OM, we explored the impact of several patient-level moderator variables. As with the missing data, therapist effects, and subscale analyses, moderator analyses were conducted on Model 3 only. Continuous variables were grand mean centered before included as level 2 predictors of all four random coefficients (i.e. two autoregression coefficients and two cross-lagged coefficients). Dichotomous variables were dummy coded. We first tested demographic variables sex and age. None of these significantly moderated any of the random effects.

Next, initial levels on the CORE-OM and WAI-S were tested. None of these significantly moderated the effect of WAI-S on CORE-OM. We also tested the diagnostic variables depression, anxiety, and personality problems as moderators. Depression was chosen because it seems to be the most common diagnosis on which alliance – outcome studies have been done, and anxiety because it was the most common problem reported in this study. Personality problems were reported for only a small subgroup of patients ( $N = 68$ ), but this variable was included because previous research has shown that the alliance is particularly important when treating patients with personality problems (De Bolle et al., 2010; Muran et al., 2009). It turned out that neither depression nor anxiety moderated the alliance effect, but personality problems did ( $b = -1.59, t = 2.67, p = .008$ ) despite the relatively small number of patients in this group. To explore this further, we tested a single-level multigroup analysis, with one group consisting of patients diagnosed with personality problems and the other of all other patients. We first estimated a model in which the effect of the alliance was forced to be equal across the two groups, then a second model in which the alliance effect was estimated separately in the two

groups. The constrained model showed poor model fit ( $\text{Chi}^2 = 34.55 (10), p < .001$ ), while the model with separately estimated alliance effects showed better fit ( $\text{Chi}^2 = 25.65 (9), p = .002$ ). A  $\text{Chi}^2$  difference test was highly significant ( $\Delta \text{Chi}^2 = 8.9 (1), p < .01$ ). This means that there was a statistically significant difference between the regression coefficients for change in WAI-S on subsequent CORE-OM change between the groups. In the group with personality problems, the standardized effect of WAI-S on subsequent CORE-OM change was six times larger than in the group without personality problems ( $\beta = -.30, t = 3.66, p < .001$ ).

Finally, we tested if the impact of the alliance on symptom reduction differed between treatment types. The three most commonly used treatment types, as reported by therapists, were supportive, psychodynamic, and cognitive or cognitive-behavioral. In order to enhance statistical power we combined the cognitive, behavioral, and cognitive-behavioral treatments into one group, and psychodynamic and relational treatments into one. We created one dummy variable for each of the three treatment types (Supportive, CBT, and PDT). A fourth group composed of all treatments that were none of these three (e.g. client centered, interpersonal therapy, crisis interventions, gestalt therapy, family therapy) was used as reference group with which these three treatments were compared. None of the treatments differed significantly from the reference group or from each other (confidence intervals were overlapping for all three treatments).

## Discussion

Results from the present study showed a statistically significant effect of the within-patient variation in therapeutic alliance on symptom change from session to session. While between-patient relationships between alliance and outcome are likely to be significantly influenced by patient and/or therapist characteristics, analyzing the impact of fluctuations in the



alliance from session to session comes closer to the theoretical view of the alliance as a process that is ongoing throughout treatment, reflecting the interaction between patient and therapist. Most previous studies on the alliance-outcome relationship have used overall treatment outcome as the dependent variable, and the between-patient variation in alliance scores in a given session as the predictor. In the present study we were also able to show that the within-patient effect of the working alliance on symptom change to the next session held even if we controlled for the “reverse causation” effect of immediately preceding symptom change on therapeutic alliance. We believe that these analyses give much stronger support for the alliance as a causal mechanism of change in psychotherapy than most previous research. Although the between-patient relationship between alliance and outcome may be important, it cannot be used as evidence that working to improve the alliance with a given patient will improve outcome.

Our results qualify the conclusions of Baldwin et al. (2007) which showed that only the mean level of alliance for each therapist was important for outcome. The within-patient effect of alliance on symptom level varied significantly between patients, but not between therapists. It seems important for therapists (at least in brief primary care psychotherapy) to monitor and work with the alliance not only in the beginning of treatment but in each session, perhaps especially if that patient has personality problems. The session-to-session effect of the alliance on symptom level points to the importance of continually monitoring the alliance throughout treatment. Our findings indicate that when the alliance is worse than usual for a given patient, symptoms are likely to get worse to the next session. Although the concept of alliance ruptures has been defined differently in previous research (e.g. Stevens, Muran, Safran, Gorman, & Winston, 2007; Stiles et al., 2004), an alliance that is worse than usual for a particular patient could be seen as one way of defining a rupture.

The framework of Safran and Muran (2000) may be important for learning to better recognize and address ruptures in the alliance. Safran and Muran have shown that therapists can learn to recognize ruptures in the therapeutic alliance and they also provide a treatment protocol for the resolution of ruptures. If a therapist is able to recognize an alliance rupture in time, he/she may be able to utilize the session-to-session effect of the alliance on outcome more efficiently. This does not mean that therapists should necessarily talk with their patients about the alliance in each session, only that therapists should try to be sensitive to fluctuations in the alliance and, if necessary, apply interventions for addressing ruptures that are relevant within the framework of their treatment model.

Tasca and Lampard (2012) have proposed a reciprocal influence model for the relationship between alliance and symptom change, in which alliance and symptoms affect each other throughout treatment. Our results support this model. Not only did the alliance predict subsequent symptom change, but symptom change predicted subsequent alliance change. It seems natural for a patient who experiences a worsening of symptoms during treatment to start doubting the effectiveness of the treatment he or she is in, thus weakening the therapeutic alliance, and vice versa. Thus, it may be important for therapists to be aware that an increased symptom level for a patient in a given session is a predictor for poor therapeutic alliance in that session. When a patient reports high levels of symptoms in a session, therapists should be alert for signs of alliance ruptures.

The within-patient effect of the alliance on subsequent symptom change was, on average, small, but there was much variation between patients. For some patients the effect of the alliance was likely much larger. We explored a number of possible moderators of the alliance effect, but found only one significant. Patients reported by therapists to have personality problems showed a

much larger alliance effect than other patients (about six times larger). Still, there was much unexplained variance left after personality problems had been taken into account, so there were probably other moderators of the alliance effect that we were unable to find.

We should acknowledge some limitations of the current study. Most of the treatments were very brief, so generalization of results may only apply to quite brief primary care psychotherapy. It is possible that the relationships look different in longer treatments, because specific therapy techniques become more important in later phases of treatment. The fact that the pattern-mixture model indicated that the alliance effect was smallest in the shortest treatments would seem to indicate the opposite, namely a larger alliance effect in longer treatments. However, this finding may be due to the reliability of the OLS regressions used to transform the variables being low for very short time-series (Curran & Bauer, 2011). We also had relatively few treatments longer than 10 sessions, so results are probably less reliable for longer treatments. The relationships between alliance and outcome may not be the same in psychotherapies in specialized psychiatric units, perhaps because patients have more severe problems or because therapists have more training. Finally, we explored several moderators of the alliance effect, increasing the risk that the one we found significant was a Type I error.

Needless to say, the present results need to be replicated in other samples. In the meantime, it would be safe to say that the present results build upon previous research on the therapeutic alliance and strengthen the idea that the therapeutic alliance is an important part of the process leading to therapeutic change.

### References

- Baldwin, S. A., Berkeljon, A., Atkins, D. C., Olsen, J. A., & Nielsen, S. L. (2009). Rates of change in naturalistic psychotherapy: Contrasting dose-effect and good-enough level models of change. *Journal of Consulting and Clinical Psychology, 77*(2), 203-211.  
doi: <http://dx.doi.org/10.1037/a0015235>
- 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.  
doi: <http://dx.doi.org/10.1037/0022-006X.75.6.842>
- Barber, J. P., Connolly, M. B., Crits-Christoph, P., Gladis, L., & Siqueland, L. (2000). Alliance predicts patients' outcome beyond in-treatment change in symptoms. *Journal of Consulting and Clinical Psychology, 68*(6), 1027-1032.  
doi: <http://dx.doi.org/10.1037/0022-006X.68.6.1027>
- Barrett, P. (2007). Structural equation modelling: Adjudging model fit. *Personality and Individual Differences, 42*(5), 815-824. doi: 10.1016/j.paid.2006.09.018
- Bollen, K. A., & Curran, P. J. (2004). Autoregressive Latent Trajectory (ALT) Models: A Synthesis of Two Traditions. *Sociological Methods & Research, 32*(3), 336-383. doi: 10.1177/0049124103260222
- Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, Research & Practice, 16*(3), 252-260.
- Busseri, M. A., & Tyler, J. D. (2003). Interchangeability of the Working Alliance Inventory and Working Alliance Inventory, Short Form. *Psychol Assess, 15*(2), 193-197.  
doi: <http://dx.doi.org/10.1037/1040-3590.15.2.193>

- Crits-Christoph, P., Gallop, R., Temes, C. M., Woody, G., Ball, S. A., Martino, S., & Carroll, K. M. (2009). The alliance in motivational enhancement therapy and counseling as usual for substance use problems. *Journal of Consulting and Clinical Psychology, 77*(6), 1125-1135. doi: <http://dx.doi.org/10.1037/a0017045>
- Crits-Christoph, P., Gibbons, M. B. C., Hamilton, J., Ring-Kurtz, S., & Gallop, R. (2011). The dependability of alliance assessments: The alliance-outcome correlation is larger than you might think. *Journal of Consulting and Clinical Psychology, 79*(3), 267-278. doi: <http://dx.doi.org/10.1037/a0023668>
- Crits-Christoph, P., Gibbons, M. B. C., & Hearon, B. (2006). Does the alliance cause good outcome? Recommendations for future research on the alliance. *Psychotherapy: Theory, Research, Practice, Training, 43*(3), 280-285. doi: 10.1037/0033-3204.43.3.280
- Curran, P. J., & Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual Review of Psychology, 62*(1), 583-619. doi: doi:10.1146/annurev.psych.093008.100356
- Curran, P. J., & Bollen, K. A. (2001). The best of both worlds: Combining autoregressive and latent curve models. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change*. (pp. 107-135). Washington, DC US: American Psychological Association.
- Curran, P. J., Lee, T., Howard, A. L., Lane, S., & MacCallum, R. (2012). Disaggregating within-person and between-person effects in multilevel and structural equation growth models. In J. R. Harring & G. R. Hancock (Eds.), *Advances in Longitudinal Methods in the Social and Behavioral Sciences*. Charlotte, NC: Information Age Publishing.

- De Bolle, M., Johnson, J. G., & De Fruyt, F. (2010). Patient and clinician perceptions of therapeutic alliance as predictors of improvement in depression. *Psychotherapy and Psychosomatics*, 79(6), 378-385. doi: <http://dx.doi.org/10.1159/000320895>
- DeRubeis, R. J., Brotman, M. A., & Gibbons, C. J. (2005). A conceptual and methodological analysis of the nonspecifics argument. *Clinical Psychology: Science and Practice*, 12(2), 174-183. doi: <http://dx.doi.org/10.1093/clipsy/bpi022>
- DeRubeis, R. J., & Feeley, M. (1990). Determinants of change in cognitive therapy for depression. *Cognitive Therapy and Research*, 14(5), 469-482.  
doi: <http://dx.doi.org/10.1007/BF01172968>
- Diggle, P., & Kenward, M. G. (1994). Informative Drop-out in Longitudinal Data Analysis. [Article]. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 43(1), 49.
- Enders, C. K. (2011). Missing not at random models for latent growth curve analyses. *Psychological Methods*, 16(1), 1-16. doi: 10.1037/a0022640  
10.1037/a0022640.supp (Supplemental)
- Evans, C., Connell, J., Barkham, M., Margison, F., McGrath, G., Mellor-Clark, J., & Audin, K. (2002). Towards a standardised brief outcome measure: Psychometric properties and utility of the CORE--OM. *British Journal of Psychiatry*, 180(1), 51-60.  
doi: <http://dx.doi.org/10.1192/bjp.180.1.51>
- Feeley, M., DeRubeis, R. J., & Gelfand, L. A. (1999). The temporal relation of adherence and alliance to symptom change in cognitive therapy for depression. *Journal of Consulting and Clinical Psychology*, 67(4), 578-582. doi: <http://dx.doi.org/10.1037/0022-006X.67.4.578>

- Gallop, R., & Tasca, G. A. (2009). Multilevel modeling of longitudinal data for psychotherapy researchers: II. The complexities. *Psychotherapy Research, 19*(4-5), 438-452.  
doi: <http://dx.doi.org/10.1080/10503300902849475>
- Gaston, L., Marmar, C. R., Gallagher, D., & Thompson, L. W. (1991). Alliance prediction of outcome beyond in-treatment symptomatic change as psychotherapy processes. *Psychotherapy Research, 1*(2), 104-112.
- Greenson, R. R. (1965). The working alliance and the transference neurosis. *The Psychoanalytic Quarterly, 34*(2), 155-179.
- Hansen, N. B., Lambert, M. J., & Forman, E. M. (2002). The psychotherapy dose-response effect and its implications for treatment delivery services. *Clinical Psychology: Science and Practice, 9*(3), 329-343. doi: 10.1093/clipsy/9.3.329
- 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
- Hayduk, L., Cummings, G., Boadu, K., Pazderka-Robinson, H., & Boulianne, S. (2007). Testing! testing! one, two, three – Testing the theory in structural equation models! *Personality and Individual Differences, 42*(5), 841-850. doi: 10.1016/j.paid.2006.10.001
- Hedeker, D., & Gibbons, R. D. (1997). Application of random-effects pattern-mixture models for missing data in longitudinal studies. *Psychological Methods, 2*(1), 64-78. doi: 10.1037/1082-989x.2.1.64
- Horvath, A. O., Del Re, A., Fluckiger, C., & Symonds, D. (2011). Alliance in individual psychotherapy. *Psychotherapy, 48*(1), 9-16. doi: <http://dx.doi.org/10.1037/a0022186>

- Horvath, A. O., & Greenberg, L. S. (1989). Development and validation of the Working Alliance Inventory. *Journal of Counseling Psychology, 36*(2), 223-233.
- Kazdin, A. E. (2005). Treatment outcomes, common factors, and continued neglect of mechanisms of change. [Comment/Reply]. *Clinical Psychology: Science and Practice, 12*(2), 184-188. doi: <http://dx.doi.org/10.1093/clipsy/bpi023>
- King, D. W., King, L. A., McArdle, J. J., Shalev, A. Y., & Doron-LaMarca, S. (2009). Sequential temporal dependencies in associations between symptoms of depression and posttraumatic stress disorder: An application of bivariate latent difference score structural equation modeling. *Multivariate Behavioral Research, 44*(4), 437-464. doi: 10.1080/00273170903103308
- Klein, D. N., Schwartz, J. E., Santiago, N. J., Vivian, D., Vocisano, C., Castonguay, L. G., . . . Keller, M. B. (2003). Therapeutic alliance in depression treatment: controlling for prior change and patient characteristics. *Journal of Consulting and Clinical Psychology, 71*(6), 997-1006. doi: <http://dx.doi.org/10.1037/0022-006X.71.6.997>
- Kline, R. B. (2011). *Principles and practice of structural equation modeling (3rd ed.)*. New York, NY US: Guilford Press.
- Little, R. J. A. (1993). Pattern-Mixture Models for Multivariate Incomplete Data. [Article]. *Journal of the American Statistical Association, 88*(421), 125-134.
- McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology, 60*, 577-605. doi: 10.1146/annurev.psych.60.110707.163612
- Muran, J. C., Safran, J. D., Gorman, B. S., Samstag, L. W., Eubanks-Carter, C., & Winston, A. (2009). The relationship of early alliance ruptures and their resolution to process and



- outcome in three time-limited psychotherapies for personality disorders. *Psychotherapy: Theory, Research, Practice, Training*, 46(2), 233-248. doi: 10.1037/a0016085
- Muthén, B., & Asparouhov, T. (2011). Beyond multilevel regression modeling: Multilevel analysis in a general latent variable framework. In J. J. Hox & J. K. Roberts (Eds.), *Handbook for advanced multilevel analysis*. (pp. 15-40). New York, NY US: Routledge/Taylor & Francis Group.
- Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus user's guide*. (7th ed.). Los Angeles, CA.
- Puschner, B., Wolf, M., & Kraft, S. (2008). Helping alliance and outcome in psychotherapy: What predicts what in routine outpatient treatment? *Psychotherapy Research*, 18(2), 167-178. doi: 10.1080/10503300701367984
- Raudenbusch, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models. Applications and data analysis methods*. (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Raudenbush, S. W. (2001). Toward a coherent framework for comparing trajectories of individual change. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change*. (pp. 35-64). Washington, DC US: American Psychological Association.
- Safran, J. D., & Muran, J. C. (2000). *Negotiating the therapeutic alliance: A relational treatment guide*. New York, NY: Guilford Press; US.
- Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York, NY US: Oxford University Press.
- Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*: SAGE Publications.

- Stevens, C. L., Muran, J., Safran, J. D., Gorman, B. S., & Winston, A. (2007). Levels and patterns of the therapeutic alliance in brief psychotherapy. *American Journal of Psychotherapy, 61*(2), 109-129.
- Stiles, W. B., Barkham, M., Mellor-Clark, J., & Connell, J. (2008). Effectiveness of cognitive-behavioural, person-centred, and psychodynamic therapies in UK primary-care routine practice: Replication in a larger sample. *Psychological Medicine, 38*(5), 677-688. doi: 10.1017/s0033291707001511
- Stiles, W. B., Glick, M. J., Osatuke, K., Hardy, G. E., Shapiro, D. A., Agnew-Davies, R., . . . Barkham, M. (2004). Patterns of alliance development and the rupture-repair hypothesis: are productive relationships U-shaped or V-shaped? *Journal of Counseling Psychology, 51*(1), 81-92. doi: <http://dx.doi.org/10.1037/0022-0167.51.1.81>
- Strunk, D. R., Brotman, M. A., & DeRubeis, R. J. (2010). The process of change in cognitive therapy for depression: Predictors of early inter-session symptom gains. *Behaviour Research and Therapy, 48*(7), 599-606. doi: 10.1016/j.brat.2010.03.011
- Strunk, D. R., Cooper, A. A., Ryan, E. T., DeRubeis, R. J., & Hollon, S. D. (2012). The process of change in cognitive therapy for depression when combined with antidepressant medication: Predictors of early intersession symptom gains. *Journal of Consulting and Clinical Psychology, 80*(5), 730-738. doi: 10.1037/a0029281  
10.1037/a0029281.supp (Supplemental)
- Tang, T. Z., & DeRubeis, R. J. (1999). Reconsidering rapid early response in cognitive behavioral therapy for depression. *Clinical Psychology: Science and Practice, 6*(3), 283-288. doi: 10.1093/clipsy.6.3.283

- Tasca, G. A., & Lampard, A. M. (2012). Reciprocal Influence of Alliance to the Group and Outcome in Day Treatment for Eating Disorders. *Journal of Counseling Psychology*. doi: 10.1037/a0029947
- Wu, M. C., & Carroll, R. J. (1988). Estimation and Comparison of Changes in the Presence of Informative Right Censoring by Modeling the Censoring Process. *Biometrics*, 44(1), 175-188.
- Zuroff, D. C., & Blatt, S. J. (2006). The therapeutic relationship in the brief treatment of depression: Contributions to clinical improvement and enhanced adaptive capacities. *Journal of Consulting and Clinical Psychology*, 74(1), 130-140.  
doi: <http://dx.doi.org/10.1037/0022-006X.74.1.130>
- Zuroff, D. C., Kelly, A. C., Leybman, M. J., Blatt, S. J., & Wampold, B. E. (2010). Between-therapist and within-therapist differences in the quality of the therapeutic relationship: effects on maladjustment and self-critical perfectionism. *Journal of Clinical Psychology*, 66(7), 681-697. doi: 10.1002/jclp.20683

Table 1.

*Demographic information.*

Age	Mean = 37.3, SD = 14.3
Gender	
Female	74%
Male	26%
Occupational status	
Employed	61%
Students	14%
Unemployed	13%
Parental leave	3%
Accommodation	
Living with a partner	56%
Living alone	35%
Most common presenting problems	
Anxiety	47%
Relationship problems	35%
Depression	34%
Grief	20%
Work related problems	19%
Somatic problems	13%

Table 2.

*Descriptive statistics for the Clinical Outcomes in Routine Evaluation – Outcome Measure and Working Alliance Inventory - 12.*

Session	N	CORE-OM				WAI-S				
		Mean	Min	Max	SD	N	Mean	Min	Max	SD
1	925	17.23	2.06	31.76	5.65	857	5.26	1.33	7.00	1.08
2	792	15.34	.59	31.47	5.77	733	5.53	2.00	7.00	1.00
3	658	14.74	.00	30.88	5.75	606	5.68	2.08	7.00	.96
4	538	14.30	.29	32.06	6.03	494	5.84	2.92	7.00	.90
5	435	13.88	.59	35.59	5.87	398	5.86	2.83	7.00	.92
6	357	13.81	.00	32.65	5.84	328	5.86	2.42	7.00	.96
7	287	13.52	.00	35.48	6.13	271	5.91	3.00	7.00	.94
8	226	13.11	.88	31.47	5.86	213	5.97	3.00	7.00	.90
9	179	13.77	.00	32.35	6.03	175	5.98	3.67	7.00	.89
10	144	13.37	.88	31.76	6.27	138	5.94	3.00	7.00	.91
11	106	13.24	.88	30.88	5.91	104	5.96	3.42	7.00	.87
12	83	12.58	.88	23.24	5.55	81	5.89	3.08	7.00	.95
13	66	12.60	1.47	26.47	5.49	63	5.97	3.33	7.00	.93
14	56	12.47	1.18	26.47	5.90	56	5.95	3.44	7.00	.97
15	44	11.32	2.06	23.82	5.29	43	5.82	3.00	7.00	1.12
16	38	10.90	1.47	22.06	5.99	36	5.98	3.67	7.00	1.07
17	28	12.97	2.94	27.06	5.97	28	5.72	3.50	7.00	1.02
18	22	13.70	2.65	24.71	6.23	22	5.69	3.75	7.00	1.07
19	16	14.17	2.94	21.76	5.81	15	5.46	2.67	7.00	1.32
20	13	13.21	3.53	21.47	4.37	12	5.46	2.67	7.00	1.19
21	12	13.16	3.53	21.18	5.06	12	5.83	4.17	7.00	.99
22	11	13.40	6.18	22.06	4.59	11	5.29	3.25	7.00	1.10
23	7	16.53	12.06	26.76	4.91	7	5.88	4.25	7.00	.83
24	8	14.15	3.53	24.41	6.38	8	5.69	4.25	6.42	.74
25	5	15.71	11.76	23.24	4.52	5	5.51	4.25	6.00	.72
26	4	13.82	7.94	18.24	4.30	4	5.54	4.00	6.17	1.03
27	2	12.21	9.41	15.00	3.95	2	4.58	3.92	5.25	.94
28	3	12.55	9.41	18.82	5.43	2	6.00	5.67	6.33	.47
29	3	10.39	7.65	14.71	3.78	3	5.83	4.67	6.83	1.09
30	4	14.04	10.00	20.59	4.74	4	5.90	4.58	6.58	.94
31	3	15.20	8.24	22.94	7.38	3	5.22	2.82	6.58	2.09
32	4	13.38	8.82	15.59	3.14	3	6.14	5.58	6.58	.51
33	1	12.06	12.06	12.06						
34	2	12.35	10.59	14.12	2.50	2	5.29	5.17	5.42	.18

35	1	13.53	13.53	13.53	1	4.92	4.92	4.92
36	1	12.06	12.06	12.06	1	4.33	4.33	4.33
37	1	10.29	10.29	10.29	1	7.00	7.00	7.00

---

Table 3.

Parameter estimates for working alliance predicting symptom change from session to session (Model 1).

	Equation 1, single level				Equation 1, two-level	
Fit indices	$N_{\text{obs}} = 3225, N_{\text{patients}} = 646$				$N_{\text{obs}} = 3225, N_{\text{patients}} = 646$	
$\chi^2$ (df)	2.55 (1), $p = .11$					
RMSEA (90% CI)	0.02 (0.00, 0.06)					
CFI / SRMR	0.88 / 0.01					
Information Criteria	AIC = 15449.4, BIC = 15458.2				AIC = 15292.6	BIC = 15307.1
Fixed effects	<i>b</i>	95% CI	$\beta$	95% CI	<i>b</i>	95% CI
CORE-OM lag1 → CORE-OM	-0.06**	-0.09, -0.02	-0.06**	-0.09, -0.02	-0.15***	-0.20, -0.10
WAI-S lag1 → CORE-OM	-0.30**	-0.52, -0.08	-0.05**	-0.08, -0.01	-0.32*	-0.58, -0.06
Random effects	Estimate		95% CI		Estimate	95% CI
CORE-OM lag1 → CORE-OM					0.07***	0.05, 0.10
WAI-S lag1 → CORE-OM					0.99***	0.24, 1.74
Error (CORE-OM)	7.03 ***		6.58, 7.38		6.19***	5.87, 6.52

Note. \*\*\*  $p < .001$ , \*\*  $p < .01$ , \*  $p < .05$  (two-tailed)

Table 4.

Parameter estimates for symptom level predicting working alliance change from session to session (Model 2).

	Equation 2, single level				Equation 2, two-level	
Fit indices	$N_{\text{obs}} = 3155, N_{\text{patients}} = 645$				$N_{\text{obs}} = 3155, N_{\text{patients}} = 645$	
$\chi^2$ (df)	4.03 (1), $p = .04$					
RMSEA (90% CI)	0.03 (0.00, 0.06)					
CFI / SRMR	0.98 / 0.01					
Information Criteria	AIC = 3221.4, BIC = 3239.6				AIC = 3082.7, BIC = 3113.0	
Fixed effects	<i>b</i>	95% CI	$\beta$	95% CI	<i>b</i>	95% CI
WAI-S lag1 → WAI-S	-0.16***	-0.19, -0.12	-0.16***	-0.19, -0.12	-0.21***	-0.26, -0.17
CORE-OM → WAI-S	-0.03***	-0.03, -0.02	-0.17***	-0.20, -0.13	-0.03***	-0.03, -0.02
Random effects	Estimate		95% CI		Estimate	95% CI
WAI-S lag1 → WAI-S					0.05***	0.03, 0.06
CORE-OM → WAI-S					0.00**	0.00, 0.00
Error (WAI-S)	0.16***		0.15, 0.17		0.14***	0.13, 0.15

Note. \*\*\*  $p < .001$ , \*\*  $p < .01$  (two-tailed)



Table 5.

Parameter estimates for reciprocal relationships between change in therapeutic alliance and symptom change from session to session (Model 3).

	Equation 3, single level				Equation 3, two-level	
Fit indices	$N_{\text{obs}} = 3225, N_{\text{patients}} = 646$				$N_{\text{obs}} = 3225, N_{\text{patients}} = 646$	
$\chi^2$ (df)	11.20 (3), $p = .01$					
RMSEA (90% CI)	0.03 (0.01, 0.05)					
CFI / SRMR	0.95 / 0.02					
Information Criteria	AIC = 18707.4, BIC = 18688.3				AIC = 18375.3, BIC = 18404.3	
Fixed effects	<i>b</i>	95% CI	$\beta$	95% CI	<i>b</i>	95% CI
CORE-OM lag1 → CORE-OM	-0.06**	-0.09, -0.02	-0.06**	-0.09, -0.02	-0.15***	-0.20, -0.10
WAI-S lag1 → CORE-OM	-0.30**	-0.52, -0.08	-0.05**	-0.08, -0.01	-0.32*	-0.58, -0.06
WAI-S lag1 → WAI-S	-0.15***	-0.19, -0.12	-0.16***	-0.19, -0.12	-0.21***	-0.26, -0.17
CORE-OM → WAI-S	-0.03***	-0.03, -0.02	-0.17***	-0.20, -0.13	-0.03***	-0.03, -0.02
Random effects	Estimate		95% CI		Estimate	95% CI
CORE-OM lag1 → CORE-OM					0.07***	0.05, 0.10
WAI-S lag1 → CORE-OM					0.99**	0.24, 1.74
WAI-S lag1 → WAI-S					0.05***	0.03, 0.06
CORE-OM → WAI-S					0.00***	0.00, 0.00
Error (CORE-OM)	7.03***		6.69, 7.38		6.19***	5.87, 6.52
Error (WAI-S)	0.16***		0.15, 0.17		0.14***	0.13, 0.15

Note. \*\*\*  $p < .001$ , \*\*  $p < .01$  (two-tailed)

