The reciprocal relationship between alliance and early treatment symptoms: A two-stage individual participant data meta-analysis

Christoph Flückiger
Department of Psychology, University of Zurich, Switzerland

Julian Rubel
Department of Psychology, Justus Liebig University Giessen, Germany

AC Del Re
Naval Health Research Center, San Diego, USA

Adam O. Horvath
Faculty of Education, Simon Fraser University, Vancouver, Canada

Bruce E. Wampold
Modum Bad Psychiatric Center, Norway and University of Wisconsin – Madison

Paul Crits-Christoph
Department of Psychiatry, University of Pennsylvania, USA

Dana Atzil-Slonim
Department of Psychology, Bar-Ilan University, Ramat Gan, Israel

Angelo Compare
Department of Human and Social Science, University of Bergamo, Italy

Fredrik Falkenström and Annika Ekeblad
Department of Behavioural Sciences and Learning, Linköping University, Sweden

Paula Errázuriz
Department of Psychology, Pontificia Universidad Católica de Chile, Santiago, Chile

Hadar Fisher
Department of Psychology, Bar Ilan University, Ramat Gan, Israel

Asle Hoffart
Modum Bad Clinic, Norway

Jonathan Huppert and Yogev Kivity
Department of Psychology, Hebrew University of Jerusalem, Israel

Manasi Kumar
Department of Psychiatry, University of Nairobi, Kenya

Wolfgang Lutz
Department of Psychology, University of Trier, Germany

John Christopher Muran
Gordon F. Derner School of Psychology, Adelphi University, New York, USA

Daniel Strunk
Department of Psychology, Ohio State University, Columbus, USA

Giorgio A. Tasca
School of Psychology, University of Ottawa, Canada

Andreea Vîslă
Department of Psychology, University of Zürich, Switzerland

Ulrich Voderholzer
Department of Psychiatry, University of Munich, Germany

Christian A. Webb
Department of Psychiatry, Harvard Medical School & McLean Hospital, Boston

Hui Xu
School of education, Loyola University Chicago, USA

Sigal Zilcha-Mano
Department of Psychology, University of Haifa, Israel

Jacques P. Barber
Gordon F. Derner School of Psychology, Adelphi University, New York, USA

June 2020, APA Journal of Consulting and Clinical Psychology,

10.1037/ccp0000594

Corresponding author: Christoph Flückiger, PhD, Department of Psychology, University of Zürich, Binzmühlestrasse 14/04, 8050 Zürich, Switzerland, christoph.flueckiger@uzh.ch, skype: christoph Flückiger

Authors Note

We are very thankful for the investigators and authors of the primary studies for their contribution to this meta-analysis. The authors list and its alphabetic order do not represent the sequence of engagement of the primary study investigators to collect their datasets, and it does not represent the conceptual and methodological innovations of all the authors engaged in the primary manuscripts. More explicit standards are necessary on how to conduct meta-analysis that investigates individual participant datasets, particularly when integrating future open source datasets.

Data transparency statement

There is no prior manuscript that analyzed this set of data. Furthermore, there is no manuscript submitted or in pipeline that is based on the present dataset. We do not intent to further publish any manuscript with the same set of data.

Founding

This study was supported by the Swiss National Science Foundation (Grant PP00P1_163702) and the Foundation for Research in Science UZH (Grant: STWF-19-018) (PI: Christoph Flückiger).

Abstract

Objective: Even though the early alliance has been shown to robustly predict posttreatment outcomes, the question whether alliance leads to symptom reduction or symptom reduction leads to a better alliance remains unresolved. To better understand the relation between alliance and symptoms early in therapy, we meta-analyzed the lagged session-by-session within-patient effects of alliance and symptoms from sessions 1 to 7. **Methods:** We applied a two-stage individual participant data meta-analytic approach. Based on the datasets of 17 primary studies from nine countries that comprised 5350 participants, we first calculated standardized session-by-session within-patient coefficients. Second, we meta-analyzed these coefficients by using random-effects models to calculate omnibus effects across the studies. **Results:** In line with previous meta-analyses we found that early alliance predicted posttreatment outcome. We identified significant reciprocal within-patient effects between alliance and symptoms within the first seven sessions. Cross-level interactions indicated that higher alliances and lower symptoms positively impacted the relation between alliance and symptoms in the subsequent session. **Conclusion:** The findings provide empirical evidence that in the early phase of therapy, symptoms and alliance were reciprocally related to one other, often resulting in a positive upward spiral of higher alliance/lower symptoms that predicted higher alliances/lower symptoms in the subsequent sessions early in therapy. Two-stage individual participant data meta-analyses have the potential to move the field forward by generating and interlinking well-replicable clinically useful knowledge for psychological interventions.

Keywords: working alliance; early response, process-based therapy; within-patient effects; individual participant data meta-analysis

Public Health Significance Statements

Improvements in the quality of the patient-rated alliance are associated with subsequent symptom reduction early in psychotherapy, and symptom reduction is associated with further improvement in the subsequent alliance. This meta-analysis provides empirical evidence for good clinical wisdom that collaborative qualities within the therapist–patient relationship and early distress remediation go “hand-in-hand.” These results underscore the relevance of respectful, collaborative, and ethically sound care for mental health patients to positively impact therapy outcomes.

The reciprocal relationship between alliance and early treatment symptoms: A two-stage individual participant data meta-analysis

The alliance has been one of the most frequently investigated therapeutic factors associated with psychotherapy success (Norcross & Lambert, 2019). The most recent meta-analytic synthesis of the overall association between the alliance assessed early in treatment and treatment outcome indicated that the early alliance predicted on average 5% of the variance in therapy outcomes including dropout rates (k = 295; Flückiger, Del Re, Wampold, & Horvath, 2018). Flückiger and colleagues found no evidence suggesting that the early alliance and outcome relation was substantially impacted by the patient's pretreatment severity and/or particular psychotherapy orientation (see also Flückiger, Del Re, Wlodasch, Horvath & Wampold, 2020), two factors that have been raised as potential confounds for the alliance–outcome relation (e.g., Barber, 2009; Crits-Christoph, Connolly Gibbons, Mukherjee, 2013; DeRubeis, Brotman, & Gibbons, 2005). However, it is less clear whether a stronger alliance relates to subsequent symptom reduction or vice-versa (e.g., Barber, Connolly, Crits-Christoph, Gladis, & Siqueland, 2000, Wampold & Imel, 2015; Zilcha-Mano, 2017). Several studies have examined the unfolding of the alliance and symptoms over the course of therapy, but no meta-analytic synthesis of this relation has been conducted.

The early alliance and treatment outcome relation suggests that the first phase of therapy is critical to the success of therapy, a conjecture that is widely accepted across theoretical orientations (e.g., Barber, Zilcha-Mano, Gallop, Barret, McCarthy, & Dinger, 2014; Beard & Delgadillo, 2019; Goldfried & Norcross, 2019; Gmeinwieser, Hoagmayer, Pieh, & Probst, 2019; Kivity, Levy, Kolly, & Kramer, 2019; Spencer, Goode, Penix, Trusty, & Swift, 2019). Some authors have underscored the relevance of particular tasks and related interventions within well-specified disorder-specific treatment approaches to promote treatment reduction (e.g., by

behavioral activation or homework; Sasso, Strunk, Braun, DeRubeis, & Brotman, 2016, Strunk, Brotman, & DeRubeis, 2010). While, other authors have conceptualized this early treatment phase in a broader context of patients' re-moralization, generally characterized by an alleviation of hopelessness and the promotion of (subtle) optimistic expectations about the treatment and the development of the trust in the therapeutic relationship (e.g., Frank & Frank, 1991; Wampold & Imel, 2015). Research across a variety of disorders and orientations has found that early treatment response is predictive of posttreatment outcome (e.g., Delgadillo, McMillan, Lucock, Leach, Ali, & Gilbody, 2014; Linardon, Brennan, & de la Piedad, 2016; Lutz et al., 2014; Lutz, Stulz & Köck, 2009; Wucherpfennig, Rubel, Hofmann, & Lutz , 2017; Shalom et al., 2018; Nazar et al., 2017; Rubel, Lutz, & Schulte, 2015). Moreover, a recent meta-analysis found large effect size differences in posttreatment outcomes between patients who showed early treatment response and participants without early improvements (r = .40; g = 0.8; Beard & Delgadillo, 2019). Thus, we have evidences from parallel lines of research of two potential early-therapy indicators of therapy outcome. In this study we used a meta-analytic approach focused on the session-by-session relation between alliance and symptoms early in therapy to better understand the relation between early symptom reduction and alliance.

The alliance is conceptualized as a primarily pan-theoretical construct (Bordin, 1979; Horvath, 2018). Alliance in the early phase of therapy includes collaboration between therapists and clients in coordinated planning of distress reduction, emphasis of the potential relevance of the patient's belief in the therapist as a potent source of help and a warm, supporting, and caring relationship (e.g., Goldfried & Norcross, 2019; Horvath, 2018; Luborsky, 1976; Norcross & Lambert, 2018). However, theoretical positions on the role of alliance and its relation to early symptoms vary across researchers; e.g., in cognitive behavioral therapy (e.g., Coyne, Constantino, Westra, & Antony, 2019; Sasso, et al., 2016; Strunk et al., 2010) as well as in

psychodynamic-oriented therapies (e.g., Barber, 2009; Zilcha-Mano, 2017). Thus, more data-based approaches are called for to improve the understanding of how early alliance and early symptoms are connected to each other on a session-by-session basis across theoretical perspectives (e.g., APA, 2006; Horvath, 2018).

Most meta-analytic research on the alliance–outcome relation has investigated the impact of alliance on outcome from a between-patient (BP) perspective. Most frequently, one alliance assessment (usually in the early phase of therapy) across many patients predicts between-patient mid or posttreatment outcomes (e.g., Flückiger et al., 2018). The BP effects addresses the question "Do patients with better early alliance ratings have better posttreatment outcomes than patients with lower early alliance ratings?" By contrast, within-patient (WP) effects provide information about the temporal relationship between alliance and outcomes within patients (e.g., Barber, et al., 2000; Hawley, Ringo Ho, Zuroff & Blatt, 2006; Klein, Schwartz, Santiago, Vivian et al. 2003; Zilcha-Mano, 2017). This session-by-session or WP effect addresses the question, "Is it the case for a particular patient, that sessions with better-than-usual alliance ratings are followed by lower-than-usual symptom ratings, or vice-versa?" Generalizing results from BP effects to WP effects is not warranted because WP effects may be independent of the respective BP effects and different therapeutic factors can play a role on these different levels (e.g., Beltz, Wright, Sprague, & Molenaar, 2016; Curran & Bauer, 2011; Hamaker, Kuiper & Grasman, 2015).

One advantage of additionally investigating WP effects in parallel to BP effects is the potential to test the temporal relation between two simultaneously occurring factors (e.g., alliance and symptoms) at a fine-grained, session-by-session level. Given the potential benefits of investigating of WP effects, an increasing number of recent studies have examined the alliance-symptoms link on a session-by-session basis (e.g., Wampold & Imel, 2015; Zilcha-Mano, 2017).

However, to summarize or generalize such effects (across studies) is difficult because these studies have tended to use diverse statistical approaches. Furthermore, they have often reported unstandardized coefficients, preventing systematic comparisons across studies.

WP and BP effects of early alliance or symptoms may not be unrelated to each other, that is, BP differences in the alliance or symptoms may impact the session-by-session WP alliance-symptom effects (i.e., cross-level interactions). Preliminary results indicated that patients who rate alliance higher (i.e., BP effects) also report higher alliance-symptom effects (WP effects) compared with patients with lower average alliance ratings (e.g., Hoffart, Øktedalen, Langkaas, & Wampold, 2013; Rubel et al., 2019). Overall, whether BP alliance and symptom scores moderate the WP alliance-symptom relation early in treatment remains unknown.

The primary aim of this study was to examine the WP effects of early alliance and early symptoms on a session-by-session basis. Figure 1 illustrates four different WP effects of alliance and early treatment symptoms. We, a priori, specified a definition of early therapy as comprising sessions 1 to 7[1]. In our analyses, we used a two-stage individual participant data meta-analysis approach to apply a uniform analysis to all primary datasets and compute standardized WP coefficients to enable the investigation of meta-analytic omnibus effects across particular research questions and study conditions. The first stage comprised WP calculations of alliance-symptoms relations within each study; the second stage comprised integrating these results across

[1] There is no uniform definition of early treatment phase in the literature across countries, treatment orientations, and patient populations (e.g., Lutz, Stulz & Köck, 2009; Flückiger, Wampold, Degadillo, Rubel & Lutz, 2020). In the alliance literature early phase is often defined as before session 6 (e.g., Horvath et al., 2011). In our primary data 14 out of the 17 datasets, treatment duration was more than 14 sessions. For the current analyses, we attempted to balance the statistical requirements (not too few sessions) with definitions of "early phase" used in the literature (i.e., not too many sessions).

patients using standard meta-analytical statistical methods. Based on available prior literature, we derived the following research questions:

Research Question 1 (RQ_1): We investigated the WP effect of the early alliance and early symptoms on a session-by-session basis by examining the relation of alliance at time t (A_t) on symptoms at time t+1 ($A_t \rightarrow S_{t+1}$) and the relation of symptoms at time t (S_t) on alliance at time t ($S_t \rightarrow A_t$). We also examined the relation of alliance on the alliance at the next session ($A_t \rightarrow A_{t+1}$) and the relation of symptoms on symptoms at the next session ($S_t \rightarrow S_{t+1}$). We hypothesized that alliance and symptoms were negatively associated with each other (i.e., higher alliance was associated with lower symptoms) in these lagged analyses.

Research Question 2 (RQ_2): Based on our interest in the WP alliance-symptom coefficient ($A_t \rightarrow S_{t+1}$), we investigated cross-level interactions. We hypothesized that high BP alliance and low BP symptoms will *positively* affect the WP alliance-symptom coefficient.

Methods

Selection of Data

The present study (data selection, inclusion/exclusion criterion, methods, etc.,) was preregistered at PROSPERO (CRD42019133312). A stepwise strategy was used to select the primary studies: First, we identified studies that had investigated the relation between alliance and symptoms on a session-by-session level. Second, the corresponding authors of these studies were contacted to provide more information whether their data fulfill the inclusion and exclusion criteria. The (Figure 2) provides an overview of the data collection procedure.

Systematic search. To locate studies on the relation between alliance and symptoms on a session-by-session basis the we replicated and updated the searches for the last two meta analyses on alliance outcome relation (Flückiger et al., 2018; Horvath, Del Re, Flückiger, & Symonds, 2011) on EBSCO for PsycINFO database and PSYNDEX (for German-language articles) in

December, 2018. The results of these searches were screened for inclusion in the present study using the following criteria: (1) the article referred to the therapy process variable as *helping alliance*, *working alliance*, or *therapeutic alliance;* (2) the articles examined the data of session-by-session alliance and session-by-session symptom improvement; (3) the manuscript reported estimates of the relation session-by-session between alliance and symptoms; (4) the patients were adults (mean age > 18 years); and (5) reports were written in English, Italian, German, or French. The exclusion criteria were as follows: (1) use of nonclinical samples (e.g., career counseling), and (2) use of fewer than four patients' self-reported measures at the first seven sessions on the alliance and/or symptoms assessments.

We identified 500 manuscripts (140 manuscripts from Horvath et al., 2011; 201 additional manuscripts from Flückiger et al., 2018, and 159 manuscripts from the updated search). Forty-one out of 500 manuscripts fitted the inclusion criteria based on screening the abstract. These 41 manuscripts were carefully read in full by two of the authors (C.F., A.D). This screening process identified 22 studies potentially fulfilling the inclusion and exclusion criteria. The two authors screened the same number of cases, and marginal cases/decisions were resolved by consensus.

Identifying and contacting corresponding authors. We contacted each corresponding author from the 22 studies to solicit collaboration for the present project. The purpose of the invitation was to obtain standardized coefficients based on a common statistical analysis across all included datasets. The corresponding study authors were provided three options to calculate the standardized coefficients from their primary datasets: (1) the primary authors to ran a R statistical software command created by our research team on their dataset, (2) the primary authors to provide their datasets in a structured form to our research team and we calculate the coefficients ,

and (3) the primary study authors to provide any relevant (unstructured) datasets, which were then restructured analyzed appropriately by us.[2]

Of the 22 corresponding authors contacted through email, 19 authors responded and agreed to contribute coefficients from their datasets (13 authors selected the aforementioned options 2 or 3). Of these 19 potential datasets, 15 datasets met all inclusion/exclusion criteria (three datasets were excluded based on having fewer than four alliance/symptom assessments within the first seven sessions, and one did not provide patients' self-reported alliance ratings). Furthermore, each author was asked to provide additional datasets mentioned in submitted papers but not yet published. Three authors provided an additional three datasets (Falkenström et al., 2019; Huppert et al., 2018; Rubel et al., 2019). Based on preliminary meta-analytic diagnostics, one dataset (Tasca et al., 2016) was deemed to be an extreme outlier in the session-by-session estimates (up to 8 SD from the mean omnibus tests; Viechtbauer & Cheung, 2010) and was therefore excluded from further analyses, resulting in a $k = 17$ included datasets.[3]

Statistical Analysis and Meta-Analytic Procedure

This analysis followed a two-stage individual participant data meta-analysis approach (e.g., Stewart et al., 2012; Tierney et al., 2015). Accordingly, the coefficients reported in the primary studies were recalculated for this meta-analytic synthesis by using identical statistical models, as described later. The synthesis involved two stages: First, standardized beta-coefficients were generated from the primary study's raw data of each individual patient by

[2] For the invitation letter please see "For_Authors.pdf" at https://osf.io https://osf.io/xstz2/?view_only=0c9051510b3242dca9e548f9da4fd4b9). The uniform R code is available in the supplemental materials. This procedure allowed us to integrate datasets where the policies only allowed for inhouse analyses.

[3] Descriptive characteristics are based on the primary study reports and checked by the corresponding authors. There was no traditional meta-analytic data extraction where the coefficients and study characteristics are indirectly extracted from manuscripts.

applying identical statistical models, and second, standard meta-analytic methods were used to calculate the overall meta-analytic estimates. This approach was used: First, to control heterogeneity that could stem from the use of diverse statistical approaches between studies (e.g., SEM with multiple control variables or a longitudinal MLM with nested random effects). Second, to generate standardized coefficients for the meta-analytic synthesis, whereas the primary studies' reports have usually provided *unstandardized* coefficients (e.g., the default of statistical software such as HLM is an unstandardized coefficient; Bryk & Raudenbush, 2002). Third, the two-stage approach integrated standardized coefficients from datasets in cases where the authors were not permitted to share raw data and chose Option 1 (i.e., they ran the R statistical software command by themselves).

Calculation of standardized coefficients.

For *RQ1*, we calculated lagged WP beta-coefficients from one time period to another. Alliance was typically measured after a session, and symptoms were usually measured at the beginning of a session (except for Webb et al., 2014, who analyzed day intervals). Therefore, when predicting symptoms from the alliance ($A_t \rightarrow S_{t+1}$), the lagged WP coefficient mostly reflected the association of the alliance at the end of a session (post-session) with the symptoms measured at the beginning of the next session (pre-session). However, when predicting alliance from symptoms ($S_t \rightarrow A_t$), the lagged coefficient reflected the association of the symptoms measured at the beginning of a session with alliance measured at the end of the session (Figure 1).

As over 80% of the primary studies used a multilevel framework, we a priori defined parsimonious multilevel models that can be applied within an open source software (for studies using a structural equation framework see e.g., Falkenström et al., 2017; Rubel et al., 2019; Xu et al., 2015). This approach allowed us to estimate the effects of interest with an identical syntax

across all datasets. The estimation of a lagged coefficient requires the existence of at least two time points (i.e., t and t+1). Consequently, the WP analyses are based only on those *patients* that provided at least two time points. WP coefficients were standardized within persons to estimate the strength of the lagged coefficients. WP standardization has recently been illustrated as a better method for enabling meaningful interpretations of lagged coefficients than use of group-based standardization (Schuurman, Ferrer, Boer-Sonnenschein, & Hamaker, 2016; Wang, Zhang, Maxwell, & Bergeman, 2019). Specifically, before inclusion in the analysis, all variables were centered at the respective person-specific means and standard deviations. To disentangle WP from BP variation, raw scores were person-mean centered based on the recent recommendations of Wang and Maxwell (2015) to obtain a parsimonious model applicable across the included longitudinal datasets. Furthermore, BP estimates were standardized at the overall mean and SD to obtain generalizable coefficients across studies. Equation 1 exemplifies the adjusted session-by-session alliance-symptoms model ($A_t \rightarrow S_{t+1)}$):

$$WP\_S_{t+1\,i} = \beta_0 + \beta_1(WP\_A_{t\,i}) + \beta_2(WP\_S_{t\,i}) + [u_{0\,i} + e_{t\,i}] \quad (1);$$

where WP_S_{t+1} is a given patient's (i) standardized WP symptom score in session t+1; β_0 is the average intercept, which is allowed to vary between patients ($u_{0\,i}$); β_1 is the standardized WP effect of the alliance in session at time t on next session symptoms and is considered fixed between patients (no random effect); and β_2 is the average autoregressive effect of patients' symptom scores at a given session (at time t) on their symptom score at the next session (at time t+1) and is considered fixed between patients (no random effect). We tested if the models improved when this WP effect was allowed to vary between patients (i.e., a random term was included in the model; u_{2i}). Finally, $e_{t\,i}$ reflected the session-specific error term. These residuals on Level 1 were modeled with a first-order autoregressive (AR[1]) covariance structure, considering that sessions closer together should be more highly correlated than sessions farther

apart. We examined models where the autoregressive effects β_2 were included (*adjusted* models) and models where the autoregressive effects ß2 were excluded (*unadjusted* models). Both models provided conservative estimates of the fixed autoregressive effects because the residuals at Level 1 were modeled as an AR(1) function for both models and thus accounted for parts of the autoregressive effects (e.g., Hoffman, 2015). To predict the alliance, we used a similar analytic strategy. Models were estimated by software R package "nlme" (Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2019) to generate a transparent statistical approach.

For *RQ2*, we investigated models that considered cross-level interactions of the average alliance or symptoms scores on the $A_t \rightarrow S_{t+1}$ coefficient. Equation 2 represents one of these adjusted cross-level interaction models:

$$WP\_S_{t+1\,i} = \beta_0 + \beta_1(WP\_A_{t\,i}) + \beta_2(WP\_S_{t\,i}) + \beta_3(BP\_A_i) + \beta_4(BP\_A_i * WP\_A_{t\,i}) + [u_{0i} + e_{t\,i}] \quad (2);$$

where the dependent variable and standardized coefficients β_0 to β_2 have the same meaning as in equation 2. β_3 is the standardized association between the person-specific average alliance score over the first seven sessions and symptoms. β_4 represents the cross-level interaction effect between the average alliance level over the first seven sessions and the WP alliance on outcome. As such, β_4 indicates whether the WP_Alliance effect is moderated by the average alliance quality over the first seven sessions.

Meta-Analytic procedures. To meta-analyze the standardized beta-coefficients we used random effects meta-analysis applying inverse variance method for pooling (e.g., Borenstein, Hedges, Higgins & Rothstein, 2009). A few studies have reported two measures of primary symptoms (e.g., two measures to assess depression). To avoid favoring these studies (and thus violated the assumption of independent samples), we aggregated within-study coefficients (Del Re & Hoyt, 2010) such that each study finally contributed one coefficient to each model in the meta-analysis. Next, to estimate the overall effects across the study-level coefficients, we calculated overall

omnibus tests for each coefficient. To meta-analyze the standardized beta-coefficients from the lag-models (*RQ 1, 2*), each coefficient was weighted in inverse proportion to its variance (inverse-variance weighting; e.g., Borenstein, et al., 2009; Marin-Martinez & Sanchez-Meca, 2010). We used a random-effects model estimator (REML), assuming that the studies in this meta-analysis were sampled from a population of studies. All analyses were conducted using the R software packages for meta-analysis "MAc" (Del Re & Hoyt, 2010) and "metafor" (Viechtbauer, 2017).

Heterogeneity was assessed using the *Q* and I^2 statistics (Higgins & Thompson, 2002). If the *Q* statistic was significant, we assumed that the effects aggregated in the analysis were heterogeneous and a moderator analysis might be justified. The statistic I^2 is an index of the degree of heterogeneity computed as a percentage of the observed variability among. In addition, we calculated credibility intervals as a further indicator of heterogeneity (e.g., Wiernik, Kostal, Wilmot & Dichert, 2017). We also examined the hypotheses that our search may be biased because we excluded unpublished studies with potentially low or nonsignificant results; we used funnel plots, rank correlation (Begg & Mazumdar, 1994), and regression tests (Egger, Smith, Schneider, & Minder, 1997).

Results

Preliminary analyses

A description of the included datasets and their characteristics are presented in Table 1. The number of patients per dataset ranged from 29 to 1550 patients; the number of sessions ranged from 2 to 7, *M* = 5.07 per patient. The gender distribution was 27% to 100% female (*M* = 62% female), and the mean patients age ranged from 26.1 to 46.7 years (*M* = 37.3). There were 10 disorder-specific datasets (four depression, four anxiety, one posttraumatic stress disorder, and

one eating disorder samples) and seven datasets with mixed diagnoses (usually in the depression-anxiety cluster). Treatment orientations included cognitive-behavioral therapy (k = 8), psychodynamic therapies (k = 2), and eclectic/various orientations (k = 6). One study provided a contrast between cognitive-behavioral therapy and alliance-focused therapy. Seven datasets were gathered under randomized controlled trial conditions, and 10 datasets were sourced from routine clinical practice. Geographical distribution of the data: United States (k = 4), Israel (k = 3) Germany, Sweden, Switzerland (each k = 2), Canada, Chile, Kenya, and Norway (each k = 1). Patient-rated alliance was primarily assessed by a short form of the Working Alliance Inventory (WAI, k = 11), Bern Post-Session Report (BPSR, k = 3), California Psychotherapy Alliance Scale (k = 2), and Session Report Scale (SRS, k = 1). In nine datasets, patients' self-reported session symptoms and posttreatment outcome were assessed by a disorder-specific measure, and eight datasets provided a more general distress measure. Two studies had reported positive mental health measures and were reversely coded for this study. The original manuscripts had been published in the *Journal of Consulting and Clinical Psychology* (k = 7), *Journal of Counseling Psychology* (k = 5), *Behavior Research and Therapy* (k = 2), *Psychotherapy Research*, *Journal of Anxiety Disorders*, and *Psychotherapy* (each k = 1).

To test whether the prior large-scale meta-analytic findings of the positive relation between early alliance and outcome can be replicated within the present set of data, we run several BP alliance-outcome correlations. The overall BP effect (k = 17) of the correlation between early alliance$_{session1\text{-}7}$ and posttreatment symptoms was $r_{unadjusted}$ = -.274 (95% CI [-.212, -.327]; p < .0001; $Q_{unadjusted}$ = 27.1, p < .03; I^2 = 43%). When adjusted for baseline symptoms the correlations were $r_{adjusted}$ = -.219 (95% CI [-.160, -.279]; p < .0001, $Q_{adjusted}$ = 40.2 , p < .0004; I^2 = 64%). The BP effect between session alliance$_{session1\text{-}7}$ and dropout rates (k = 11) was $r_{unadjusted}$ = .244 (95% CI [.152, .337]; p < .0001, $Q_{unadjusted}$ =17.5 , p < .06; I^2 = 0%) and $r_{adjusted}$ = .232 (95%

CI [.136, .329]; $p < .0001$, $Q_{adjusted} = 14.3$, $p < .16$; $I^2 = 0\%$). Overall, these findings were in line with the estimates reported in Flückiger et al., (2018).

Session-by-Session Lag-Models *(RQ_1)*

A summary of the session-by-session lag models is presented in Table 2. The standardized beta-coefficients were interpreted as the magnitude of relations between two sequential observations. For example, for the $A_t \rightarrow S_{t+1}$ beta-coefficient: for every 1-point SD increase in WP alliance, there was a corresponding WP 1-point SD decrease in symptoms at the next time point.

Alliance$_t$ → Alliance$_{t+1}$. The overall WP effect ($k = 17$) of the lag-models between A_t and A_{t+1} was $\beta = .044$ (95% CI [-.002, .089]; $p < .06$), indicating a statistical trend in the direction that high alliance at time t was positively related to high alliance at time t+1; however, this trend did not reach statistical significance. Note; we observed significant heterogeneity in these effects ($Q = 112.5$, $p < .0001$; $I^2 = 89\%$). Inspection of the funnel plots indicated no substantial asymmetry (rank correlations and regression tests, $p > .52$), suggesting that publication bias was likely absent.

Alliance$_t$ → Symptoms$_{t+1}$. The overall WP effect ($k = 17$) of the relation between A_t and S_{t+1} was $\beta_{unadjusted} = -.072$ (95% CI [-.101, -.042]; $p < .0001$) and $\beta_{adjusted} = -.065$ (95% CI [-.092, -.038]; $p < .0001$), indicating that high alliance at time t was related to low symptoms at time t+1. We observed significant heterogeneity in these effects ($Q_{unadjusted} = 52.2$, $p < .0001$; $I^2 = 71\%$; $Q_{adjusted} = 41.4$, $p < .001$; $I^2 = 63\%$). The funnel plots indicated no substantial asymmetry (rank correlations and regression tests, $p > .18$).

Symptoms$_t$ → Symptoms$_{t+1}$. The overall WP effect ($k = 17$) of the lag-models between S_t and S_{t+1} was $\beta = .082$ (95% CI [.034, .130]; $p < .0001$) in the direction high symptoms at time t were positively related to high symptoms at time t+1. There was significant heterogeneity in

these effects (Q = 15.4, p < .0001; I^2 = 90%). The funnel plots indicated no substantial asymmetry (rank correlations and regression tests, p > .79).

Symptoms$_t$ → Alliance$_t$. The overall WP effect (k = 17) of the relation between S_t and A_t was $\beta_{unadjusted}$ = -.194 (95% CI [-.260, -.127]; p < .0001) and $\beta_{adjusted}$ = -.148 (95% CI [-.215, -.081]; p < .0001) in direction that high pre-session symptoms were related with low post-session alliance. We observed significant heterogeneity in these effects ($Q_{unadjusted}$ = 329.7, p < .0001; I^2 = 95%; $Q_{adjusted}$ = 41.4, p < .001; I^2 = 63%). The funnel plots indicated no substantial asymmetry (rank correlations and regression tests, p > .54).

Cross-level Interactions on A_t → S_{t+1} ***(RQ$_2$)***

Cross-Level Interaction Alliance$_{1-7}$. The overall effect of the cross-level interaction of the BP alliance$_{1-7}$ on the A_t → S_{t+1} WP coefficient (k = 17) was $\beta_{unadjusted}$ = -.028 (95% CI [-.042, -.014]; p < .0001) and $\beta_{adjusted}$ = -.025 (95% CI [-.038, -.012]; p < .0002), indicating that the association A_t → S_{t+1} was stronger in individuals with generally high alliances (see Table 3). We observed low heterogeneity in these effects ($Q_{unadjusted}$ = 20.7, p > .19, I^2 = 4%; $Q_{adjusted}$ = 18.7, p > .28, I^2 = 1%). The funnel plots indicated no substantial asymmetries (rank correlations and regression tests, p > .26).

Cross-Level Interaction Symptoms$_{1-7}$. The overall effect of the cross-level interaction of the BP symptoms$_{1-7}$ on the A_t → S_{t+1} WP coefficient (k = 17) was $\beta_{unadusted}$ = .030 (95% CI [.008, .051]; p < .007) and $\beta_{adjusted}$ = .027 (95% CI [.004, .051]; p < .02), indicating that the A_t → S_{t+1} coefficient was weaker for individuals with high symptoms compared to individuals with low symptoms (see Table 3). We observed considerable heterogeneity in these effects ($Q_{unadjusted}$ = 31.6, p > .01, I^2 = 36%; $Q_{adjusted}$ = 33.1, p > .007, I^2 = 46%). The funnel plots indicated no

substantial asymmetries (rank correlations and regression tests, $p > .32$).[4]

Discussion

Although the relation between the alliance and outcome has been a primary interest of psychotherapy researchers for several decades, the underlying dynamics of this association can be difficult to interpret (Goldfried & Norcross, 2019; Hofmann & Hayes, 2019; Muran & Barber, 2010; Norcross & Lambert, 2019; Wampold & Imel, 2015). One major unresolved question is how symptoms and the alliance interact over the course of treatment, particularly in the critical early portion of therapy (e.g., DeRubeis et al., 2005). The primary purpose of this meta-analysis was to estimate the session-by-session effects of alliance on symptoms and the impact of symptoms on the subsequent alliance in the early phase of treatment. Using datasets from 17 independent studies from nine countries, we first calculated the alliance–outcome relation and the standardized session-by-session lagged WP estimates for each of the first 7 sessions. Second, we meta-analyzed these estimates using random-effects models to calculate omnibus effects across the studies.

The preliminary analyses indicated that better average early alliance scores had a significant positive association with patients' posttreatment outcome and dropout rate, confirming that these studies are representative, given the well-established relationship between early alliance and outcome (Flückiger et al., 2018). Of note, that the positive associations with posttreatment outcome indicated some heterogeneity (within the positive associations). The primary studies included in the meta-analysis reported a variety of outcome measures across

[4] We further explored the moderating effects of psychotherapy orientations (i. e., Cognitive Behavioral Therapy, Psychodynamic Therapy and others) for all WP coefficients (*RQ 1, 2*). The results of meta-analytic moderator analyses indicated no moderating effect of the treatment orientation at study level (for all moderator analyses: $Q < 2.99$, $p > .23$, no correction for multiple testing). These preliminary analyses were exploratory in nature.

different disorders and populations. This lack of uniformity likely contributed to the observed heterogeneity in some of the analyses. The observed associations were independent of the severity of symptoms at baseline, in line with previous results of a meta-analysis on the alliance–outcome partial correlations that adjusted for a broad range of pretreatment characteristics (k = 60; Flückiger, Del Re et al., 2020; Sharf, Primavera, & Diener, 2010).

In respect to the session-by-session lag-models, higher-than-usual alliance scores in one session were followed by lower-than-usual symptoms in the following session, adjusting for previous session symptoms. Notably, however, the reverse effects ($S_t \rightarrow A_t$) were significant as well. Namely, higher-than-usual symptoms ratings reported at the beginning of a session were followed by lower-than-usual alliance ratings in that session, adjusting for previous session alliance. Several considerations may have affected the reciprocal but also heterogeneous relation between time-specific changes in alliance and symptoms early in therapy (in these lag-models).

First, the associations observed suggest that the assessment of alliance and symptom improvement go hand-in-hand (e.g., Hatcher, 2010; Huppert, Fabbro, & Barlow, 2006). These early pre and post-session evaluations may go along with many further evaluations that must be monitored and coordinated simultaneously, such as gaining a comprehensive overview of a patient's distress and overall situation, setting an overall psychotherapy schedule, creating positive expectations for change, eliciting remoralization, and detecting potential fluctuations in early progress, for example, in symptom severity, well-being, or psychosocial functioning (e.g., Luborsky, 1976; Wampold & Imel, 2015; Wucherpfennig et al., 2017; Flückiger, Grosse Holtforth, Del Re, & Lutz, 2013). Significant reciprocal relations between the alliance and symptoms were found from session to session. Notably, these associations are characterized by different time intervals (Figure 2): The symptom-alliance coefficient typically indicated a time period of approximately 50 minutes pre to post-session, and the alliance-symptom coefficient

usually covered a time period of 1 week between the alliance at time t and subsequent symptoms at time t+1. Thus, differences in the magnitude of various coefficients must be interpreted cautiously (even adjusted for prior assessments) and might be a consequence of the differences in assessment time intervals and particular assessment times. Weiss, Kivity, and Huppert (2014) made a critical innovation when they assessed alliance pre and post-session in a cognitive-behavioral therapy for panic disorder. Using this assessment plan, the authors observed a constant pattern of within-session alliance improvements followed by decreases between sessions (i.e., a "sawtooth pattern" of the alliance across assessments). Clearly, more research is necessary to improve the understanding of the potential impact of particular assessment times (i.e., at pre and post-session).

A second consideration of the reciprocal relation between alliance and symptoms may regard the overlap between symptom and alliance *self-reports* since both measures are evaluated from the patient perspective. Such a monomethod assessment is limited, and it would have been interesting to investigate other perspectives such as observer- and/or therapist rating for both concepts, alliance and symptoms (e.g., Horvath et al, 2011; for a notable inclusion of a therapist alliance rating see e.g., Falkenström et al., 2016 and alliance observer rating see Strunk et al., 2010).

Third, the findings of pre-session and post-session assessments may represent a generalized overall–evaluation of the entire session-by-session process rather than an assessment to specific interventions and behaviors (e.g., Ogles, 2013). Consequently, attention should be exercised regarding conclusion of the pre-session and post-session evaluations to in-session behaviors. For example, patients who perceive themselves as "too friendly" may report socially desirable strong alliances with their therapists (e.g., Coyne, et. al. 2019; Dinger, Strack, Sachsse, & Schauenburg, 2009; Gómez Penedo, Babl, Krieger, Heinonen, Flückiger & grosse Holtforth,

2020) and they may be more cautious in reporting negative events in the relationship, such as immediate ruptures during the sessions (e.g., Eubanks, Muran, & Safran, 2018; Rubel, Zilcha-Mano, Feils-Klaus, & Lutz, 2018). There is growing evidence that ruptures (i. e., negative disruptions in alliance levels) are common occurrences in treatments and if these negative events go unresolved, they may be followed by less positive outcomes (Eubanks et al., 2018). Consequently, pre and post-session evaluations may provide clinically relevant information of how patients *process* the many tasks related to the early phase of therapy.

Compared with the considerable heterogeneous findings observed with respect to RQ_1 (i.e., relatively large heterogeneity of the overall effects), the results of the cross-level interactions (RQ_2) were much more homogeneous: For patients who generally reported better alliances, the fluctuations in the alliance significantly related to a higher subsequent decrease in symptoms compared to lower BP alliance scores. Similarly, for patients who were less severely distressed during the first seven sessions, larger improvements in the alliance was related significantly to larger subsequent symptom reduction compared to patients with higher symptom severity scores. This pattern was similar to what Hoffart and colleagues (2013) reported for the task component of the WAI in individuals with a posttraumatic stress disorder and Rubel and colleagues (2019) report in a generalized anxiety disorder population. Thus, the results of this study (across $RQs_{1, 2}$) support a positive upward spiral of higher alliance/lower symptoms that facilitates higher alliances and lower symptoms in the subsequent sessions early in therapy (Grawe, 2004; Wucherpfennig et al., 2017; Flückiger et al., 2013). Moreover, the directions of the WP effects were consistent with those of the BP effects.

Although the moderator analyses of this meta-analysis were exploratory in nature given the relatively small number of primary studies, our results are in line with previous meta-analytic findings that have suggested non-significant differences between treatment orientations ($p > .23$

for all analyses). Overall, these results may best be understood as evidence that the relation between alliance and outcome, and the interactive dynamics between alliance and session level outcome are reasonably uniform across treatment methods and diagnostic classifications in early phases of therapy: The growth in the alliance and decreases in symptoms are imbedded in the patients' engagement, treatment acceptance and related remoralization early in treatment (Frank & Frank, 1991; Horward, Moras, Brill, Martinovich, & Lutz, 1996; Luborsky, 1976; Wampold & Imel, 2015). Clearly, further research should focus on improved detection and understanding of (at risk) patients with high symptoms and low alliances that do not show a positive upward spiral early in treatment (e.g., Brattland et al., 2016; Lambert, Whipple & Kleinstäuber, 2018; Wucherpfennig et al., 2017).

The results of this meta-analysis have notable clinical implications. Past literature has tended to contrast early symptoms and early alliances, emphasizing a priori either early symptoms (e.g., DeRubeis et al., 2005) or alliance (e.g., Barber et al., 2000). In contrast, this meta-analysis highlights the reciprocal, dialectical nature of both early symptoms and early alliance, and does not support the of an "either-or" hypothesis. Moreover, our results support "hand-in-hand" processes that might affect, for example, decisions regarding the assessment of routine outcome monitoring, clinical supervision, and daily practice. More specifically, this meta-analysis provides further empirical evidence for good clinical wisdom, that is, collaborative qualities within the therapist–patient relationship and early distress remediation do not contradict one other, but rather, are synergistic in with one another early in therapy (e.g., APA, 2006; Ribeiro, Ribeiro, Gonçalves, Horvath, & Stiles, 2013). Moreover, the results support the importance of active patient involvement in therapist–patient collaboration early in therapy (e.g., Ryan, Lynch, Vansteenkiste & Deci, 2011; Scheel, 2011; Pope & Vasquez, 2016). This meta-analysis integrated samples from nine countries including a sample from Sub-Saharan Africa,

representing a broad range of professional training and mental health contexts. Global mental health researchers may be encouraged by the present results to generate patient-centered public health awareness that highlights collaborative qualities between the health providers and their patients (e.g., Stewart, 2001; Ogden, Barr, & Greenfield, 2017).

Limitations, Further Directions, and Summary

Overall, our findings provide additional support for the growing body of research showing temporal reciprocity of alliance and symptoms on a session-by-session basis. By focusing on WP associations, potentially confounding effects of more stable person-specific characteristics were controlled for, which may allow a firmer conclusion regarding the bidirectional or interactive nature of these variables (e.g., Falkenström, Finkel, Sandell, Rubel, & Holmqvist, 2017).

There are several questions raised, but not yet answered, by this investigation: A limitation of this meta-analysis is the relatively limited size of primary studies (k = 17) which may not accurately reflect the diverse universe of psychotherapy approaches; thus, an investigation of potential differential effects across particular treatment orientations was limited. Time-specific and outside-therapy confounds, for example, intersession processes, could have impact the results. This is a concern as the primary studies, as most studies of the alliance, had usually not assessed, systematized, and reported intersession processes (for exceptions see e.g., Hartmann, Orlinsky, & Zeeck, 2011; Kaiser & Laireiter, 2019; Strunk et al., 2010; Quirk, Smith & Owen, 2018).

Based on our decision to include datasets where the alliance and symptoms were measured from session 1 to 7 to predict posttreatment outcome, we cannot draw conclusions regarding later phases of therapy. Notably, this restriction has some advantages because research has shown that the time of the alliance assessment affects the overall alliance–outcome association (e.g., Flückiger et al., 2018) and that later alliance may partly reflect the general

growth of changes in treatment perceived by each patient (Horvath et al., 2011). Furthermore, the alliance assessed early in treatment may provide the best opportunity for making adjustments in treatment, if necessary. Additionally, from a meta-analytic perspective, concentrating on the early phase allowed us to investigate the omnibus effects across short-term, long-term, and time-unlimited treatments.

We relied on computing standardized WP beta-coefficients across datasets based on one statistical method using an uniform syntax. There are emerging alternate approaches for cross-lagged models (e.g., Hamaker et al., 2015; Asparouhov, Hamaker & Muthén, 2018). Each of these approaches have some benefits and costs and no consensus has been reach as to the best approach. In the future, alternative statistical methods should be explored to find the most useful methodological routes (e.g., Falkenström et al., 2017).

We investigated the *average* WP alliance-outcome association across patients. Consequently, our overall beta-coefficients of the WP alliance-outcome association may obscure potentially important heterogeneity that might be present between subgroups of patients (e.g., Lorenzo-Luaces, DeRubeis, & Webb, 2014; Zilcha-Mano & Errázuriz, 2015). This variation could have been separately modeled in the first stage of this two-stage analysis by including a random effect for the WP alliance-outcome estimate. This would have allowed the WP alliance-outcome association to vary between patients (i.e., different alliance-outcome associations per patient) and would have produced an estimate of this variability. We decided to not explicitly model this potential variability for several reasons. Most importantly, models with an additional random slope for the WP alliance-outcome association have a higher probability for non-converge in small samples. In order to be able to include as many studies in our analysis as possible, we chose a parsimonious model that estimates our effect of interest (i.e., the beta-coefficient of the WP alliance-outcome association) and has a high likelihood to converge in a

range of different circumstances. Future research may usefully investigate the questions if between-patient variability in WP alliance-outcome effects is significantly different from zero when pooled across studies and which study characteristics moderate the amount of variability.

Possible therapist effects were not considered in this meta-analysis (e.g., Del Re, Flückiger, Horvath, Symonds, & Wampold, 2012). This decision was primarily based on the methodological consideration that for WP effects, the exclusion of higher-order effects would generally not result in substantial changes in the lower-level estimates (e.g., Van Landeghem, De Fraine, & Van Damme, 2005; Falkenström, Solomonov & Rubel, 2020). Additionally, not all datasets had information on the treating therapists, which would have resulted in an exclusion of several studies. Nonetheless, particular therapists may be more sensitive to early changes and fluctuations in alliance and symptoms compared to other therapists, which may potentially impact the overall unfolding of the early phase of therapy (e.g., Eubanks et al., 2018; Safran & Muran, 2000).

We observed a general lack of assessing and reporting various symptoms and outcome measures simultaneously, somewhat neglecting further outcome components within the broad definition of the World Health Organization such as well-being or psychosocial functioning (WHO, 2018; Howard, Lueger, Maling, & Martinovich, 1993; for exceptions see Huppert et al., 2018 and Weiss et al., 2014). Further research is necessary to better understand to what extent the outcome definition and assessment method may affect the association between process-based psychotherapy factors such as the alliance and outcome (e.g., Flückiger et al., 2019).

Although this meta-analysis integrated studies conducted across nine countries, our results may primarily summarize investigations from Westernized contexts in line with other meta-analyses in the field (e.g., Norcross & Lambert, 2019). Clearly, further research is necessary to better understand the potential generalizability of effects across (sub-) cultural contexts (e.g.,

Kumar, Kuria, Othieno & Falkenström, 2018; Errázuriz & Zilcha-Mano, 2018; Flückiger, Del Re, Horvath, Symonds, Ackert & Wampold, 2013; Vasquez, 2007).

To conclude, this meta-analysis is a step in developing a rigorous empirical foundation of process-based psychotherapy across many theoretical considerations (e.g., Crits-Christoph, Gallop, Gaines, Rieger, & Connolly Gibbons, 2018; Goldfried & Norcross, 2019; Hofmann & Hayes, 2019; Muran & Barber, 2010; Norcross & Lambert, 2019; Wampold & Imel, 2015), particularly, early in therapy (e.g., Beard & Delgadillo, 2019; Howard, et al., 1993; Spencer et al., 2019; Wucherpfennig et al., 2017). More specifically, this meta-analysis is the first to investigate WP effects on a session-by-session basis suggesting that the alliance and symptoms influence each other early in therapy (Wampold & Imel, 2015, Zilcha-Mano, 2017). Moreover, there is no evidence of a replication problem in the alliance–outcome literature, in contrast, this rigorous meta-analysis provides a practical example of how to investigate the robustness of psychological effects across populations and treatments by taking coordinated advantage of advanced statistical models (e.g., Muthukrashna & Henrich, 2019). This meta-analysis is a first step toward exploring the practicability and applicability of two-stage individual participant data meta-analysis in process-based psychotherapy summarizing WP effects.

Literature

American Psychological Association, Presidential Task Force on Evidence-Based Practice (2006). Evidence-based practice in psychology. *American Psychologist, 61*, 271-285. https://doi.org/10.1037/0003-066X.61.4.271

Asparouhov, T., Hamaker, E. L. & Muthén (2018). Dynamic structural equitation models. *Structural Equitation Modeling: A Multidisciplinary Journal*, 25, 359-388. doi: https://doi.org/10.1080/10705511.2017.1406803

Barber, J.P. (2009). Toward a working through of some core conflicts in psychotherapy research. *Psychotherapy Research. 19,* 1-12. http://doi.org/10.1080/10503300802609680

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,* 1027-1032. https://doi.org/10.1037//0022-006x.68.6.1027

Barber, J. P., Zilcha-Mano, S., Gallop, R., Barret, M., McCarthy, K. S., & Dinger, U. (2014). The associations among improvement and alliance expectations, alliance during treatment, and treatment tutcome for major depressive disorder. *Psychotherapy Research, 24*(3), 257-268. https://doi.org/10.1080/10503307.2013.871080

Beard, J. & Delgadillo, J. (2019). Early response to psychological therapy as a predictor of depression and anxiety treatment outcomes: A systematic review and meta-analysis. *Depression and Anxiety, 36*(9), 866-878. https://doi.org/10.1002/da.22931

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics, 50*(4), 1088-1101. https://doi.org/10.2307/2533446

Beltz, A. M., Wright, A. G. C., Sprague, B. N., & Molenaar, P. C. M. (2016). Bridging the nomothetic and idiographic approaches to the analysis of clinical data. *Assessment, 23*(4), 447-458. https://doi.org/10.1177/1073191116648209

Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, Research & Practice, 16*(3), 252-260. https://doi.org/10.1037/h0085885

Borenstein, M., Hedges L.V., Higgins, J.P., Rothstein, H.R. (2009): *Introduction to meta-analysis.* Chichester: John Wiley & Sons

Brattland, H., Høiseth, J. R., Burkeland, O., Tryggve S., Inderhaug, T.S., Binder, P. E., Valentina C., & Iversen, V. C. (2018). Learning from clients: A qualitative investigation of

psychotherapists' reactions to negative verbal feedback. *Psychotherapy Research*, *28*(4), 545-559. https://doi.org/10.1080/10503307.2016.1246768

Bryk, A. S., & Raudenbush, S. W. (2002). *Hierarchical linear models: Applications and data analysis methods* [2nd ed.]. Thousand Oaks, CA: Sage.

Coyne, A. E., Constantino, M. J., Westra, H. A., & Antony, M. M. (2019). Interpersonal change as a mediator of the within- and between-patient alliance-outcome association in two treatments for generalized anxiety disorder. *Journal of Consulting and Clinical Psychology*, *87*(5), 472-483. http://dx.doi.org/10.1037/ccp0000394

Crits-Christoph, P., Connolly Gibbons, Mukherjee, D. (2013). Psychotherapy Process-Outcome Research. In M. J. Lambert (Ed.). *Bergin and Garfields Handbook of Psychotherapy and Behavior Change (6th ed., pp. 298-340).* New York, NY: Wiley.

*Crits-Christoph, P., Connolly Gibbons, M. B., Crits-Christoph, K., Narducci, J., Schamberger, M., & Gallop, R. (2006). Can therapists be trained to improve their alliances? A pilot study of Alliance-Fostering Therapy. *Psychotherapy Research, 16*(3), 268-281. https://doi.org/10.1080/10503300500268557

Crits-Christoph, P., Gallop, R., Gaines, A., Rieger, A., & Connolly Gibbons, M. B. (2018). Instrumental variable analyses for causal inference: Application to multilevel analyses of the alliance–outcome relation. *Psychotherapy Research, 30*(1), 53-67. https://doi.org/10.1080/10503307.2018.1544724

*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. https://doi.org/10.1037/a0023668

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. https://doi.org/10.1146/annurev.psych.093008.100356

Del Re, A. C., & Hoyt, W. T. (2010). MAc: Meta-analysis with correlations. R package version 1.1.1.

Del Re, A.C., Flückiger, C., Horvath, A. O., Symonds, D. & Wampold, B. E. (2012). Therapist effects in the therapeutic alliance–outcome relationship: A restricted-maximum likelihood meta-analysis. *Clinical Psychology Review*, *32*(7), 642-649, https://doi.org/10.1016/j.cpr.2012.07.002

Delgadillo, J., McMillan, D., Lucock, M., Leach, C., Ali, S., & Gilbody, S. (2014). Early changes, attrition, and dose-response in low intensity psychological interventions. *British Journal of Clinical Psychology*, *53*(1), 114-130. https://doi.org/10.1111/bjc.12031

DeRubeis, R. J., Brotman, M. A. & Gibbons, C. J. (2005). A conceptual and methodological analysis of the nonspecifics argument. *Journal of Clinical Psychology, Science and Practice. 12*(2), 174-183. https://doi.org/10.1093/clipsy/bpi022

Dinger, U., Strack, M., Sachsse, T., & Schauenburg, H. (2009). Therapists' attachment, patients' interpersonal problems and alliance development over time in inpatient psychotherapy. *Psychotherapy, 46*, 277-290. https://doi.org/10.1037/a0016913.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*(7109), 629-634. https://doi.org/10.1136/bmj.315.7109.629

Eubanks, C. F., Muran, J. C., & Safran, J. D. (2018). Alliance rupture repair: A meta-analysis. *Psychotherapy, 55*(4), 508-519. https://doi.org/10.1037/pst0000185

Errázuriz, P., & Zilcha-Mano, S. (2018). In psychotherapy with severe patients discouraging news may be worse than no news: The impact of providing feedback to therapists on psychotherapy

outcome, session attendance, and the alliance. *Journal of Consulting and Clinical Psychology, 86*(2), 125-139. http://dx.doi.org/10.1037/ccp0000277

*Falkenström, F., Ekeblad, A., & Holmqvist, R. (2016). Improvement of the working alliance in one treatment session predicts improvement of depressive symptoms by the next session. *Journal of Consulting and Clinical Psychology*, *84*(8), 738-751. https://doi.org/10.1037/ccp0000119

*Falkenström, F., Granström, F., & Holmqvist, R. (2013). Therapeutic alliance predicts symptomatic improvement session by session. *Journal of Counseling Psychology*, *60*(3), 317-328. https://doi.org/10.1037/a0032258

*Falkenström, F., Kuria, M., Othieno, C., & Kumar, M. (2019). Working alliance predicts symptomatic improvement in public hospital-delivered psychotherapy in Nairobi, Kenya. *Journal of Consulting and Clinical Psychology*, *87*(1), 46-55. https://doi.org/10.1037/ccp0000363

Falkenström, F., Finel, S. Sandell, R., Rubel, J.A. & Homqvist, R. (2017). Dynamic models of individual change in psychotherapy process research. *Journal of Consulting and Clinical Psychology, 85*(6), 537-549. https://doi.org/10.1037/ccp0000203.

Falkenström, F., Solomonov, N. & Rubel, J. (2020). *Do therapist effects really impact estimates of within-patient mechanisms of change? A Monte Carlo simulation study.* Manuscript in revision for publication.

*Fisher, H., Atzil-Slonim, D., Bar-Kalifa, E., Rafaeli, E., & Peri, T. (2016). Emotional experience and alliance contribute to therapeutic change in psychodynamic therapy. *Psychotherapy*, *53*(1), 105-116. https://doi.org/10.1037/pst0000041

Flückiger, C., Del Re, A. C., Horvath, A. O., Symonds, D., Ackert, M., & Wampold, B. E. (2013). Substance use disorders and racial/ethnic minorities matter: A meta-analytic examination of the

relation between alliance and outcome. *Journal of Counseling Psychology, 60*, 610-616. http://dx.doi.org/10.1037/a0033161

Flückiger, C., Del Re, A. C., Wampold, B. E., & Horvath, A. O. (2018). The alliance in adult psychotherapy: A meta-analytic synthesis. *Psychotherapy, 55*(4), 316-340. https://doi.org/10.1037/pst0000172

Flückiger, C., Del Re, A. C., Wlodasch, D., Horvath, A. O. & Wampold, B. E. (2020). Assessing the alliance-outcome association adjusted for patient characteristics and treatment processes: A meta-analytic summary of direct comparisons. *Journal of Counseling Psychology.* https://doi.org/10.1037/cou0000424

*Flückiger, C., Hilpert, P., Goldberg, S., Caspar, F., Wolfer, C., Held, J. & Vîslă, A. (2019). Investigating the impact of early alliance scores on predicting subjective change at post treatment. *Journal of Counseling Psychology.* https://doi.org/10.1037/cou0000336

Flückiger, C., Grosse Holtforth, M., Del Re, A. C., Lutz, W. (2013). Working along sudden gains - responsiveness on small and subtle early changes and exceptions. *Psychotherapy, 50*(3), 292-297. https://doi.org/10.1037/a0031940

Frank, J. D. & Frank, J. B. (1991). *Persuasion and Healing.* Baltimore (MD): John Hopkins.

Gmeinwieser, S., Hoagmayer, Y., Pieh, C. & Probst, T. (2019). General change mechanisms in the early treatment phase and their associations with the outcome of cognitive behavioral therapy on patients with different levels of motivational incongruence. *Clinical Psychology & Psychotherapy, 26*(5),550-561. https://doi.org/10.1002/cpp.2381

Goldfried, M. R. & Norcross, J. C. (2019). *Handbook of Psychotherapy Integration.* New York, NY: Oxford University Press

Gómez Penedo, J. M., Babl, A., Krieger, T., Heinonen, E., O., Flückiger, C. & grosse Holtforth, M. (2020). Interpersonal Agency as Predictor of the Within-Patient Alliance Effects on

Depression Severity. *Journal of Consulting and Clinical Psychology,* https://doi.org/10.1037/ccp0000475

Grawe, K. (2004). *Psychological Therapy.* Göttingen: Hogrefe.

Hawley, L. L., Ho, M.-H. R, Zuroff, D. C., & Blatt, S. J. (2006). The relationship of perfectionism, depression and therapeutic alliance during treatment for depression: Latent difference score analysis. *Journal of Consulting and Clinical Psychology, 74*, 930-942. http://doi.org/10.1037/0022-006X.74.5.930

Hamaker, E. L., Kuiper, R. M. & Grasman, R. P. P. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods, 20*(1), 102-116. http://doi.org/10.1037/a0038889

Hartmann, A., Orlinsky, D., & Zeeck, A. (2011). The structure of intersession experience in psychotherapy and its relation to the therapeutic alliance. *Journal of Clinical Psychology, 67*(10), 1044-1063. http://doi.org/10.1002/jclp.20826

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21*(11), 1539-1558. https://doi.org/10.1002/sim.1186

*Hoffart, A., Øktedalen, T., Langkaas, T. F., & Wampold, B. E. (2013). Alliance and outcome in varying imagery procedures for PTSD: A study of within-person processes. *Journal of Counseling Psychology*, *60*(4), 471-482. https://doi.org/10.1037/a0033604

Hoffman, L. (2015). *Longitudinal Analysis – Modeling within-person fluctuation and change*. London: Routledge.

Hofmann, S.G. & Hayes, S. C. (2019). The future of intervention science: Process-based therapy. *Clinical Psychological Science, 7*(1), 37-50. https://doi.org/10.1177/2167702618772296

Horvath, A. O. (2018). Research on the alliance: Knowledge in search of a theory. *Psychotherapy Research*. *28*(4), 499-516. https://doi.org/10.1080/10503307.2017.1373204

Horvath, A. O., Del Re, A. C., Flückiger, C., & Symonds, D. (2011). Alliance in individual psychotherapy. *Psychotherapy*, *48*(1), 9-16. https://doi.org/10.1037/a0022186

Howard, K. I., Lueger, R. J., Maling, M. S., & Martinovich, Z. (1993). A phase model of psychotherapy outcome: Causal mediation of change. *Journal of Consulting and Clinical Psychology, 61*(4), 678-685. https://doi.org/10.1037//0022-006X.61.4.678

*Huppert, J., D., Kivity, Y., Cohen, L., Strauss, A. Y., Elizur, Y., & Weiss, M. (2018). A pilot randomized clinical trial of cognitive behavioral therapy versus attentional bias modification for social anxiety disorder: An examination of outcomes and theory-based mechanisms. *Journal of Anxiety Disorders, 59*, 1-9. https://doi.org/10.1016/j.janxdis.2018.08.002

Huppert, J. D., Fabbro, A., & Barlow, D. H. (2006). Evidence-based practice and psychological treatments. In C. D. Goodhart, A. E. Kazdin, & R. J. Sternberg (Eds.), *Evidence-based psychotherapy: Where practice and research meet* (pp. 131–152). Washington, DC: American Psychological Association.

Klein, D. N, Schwartz, J. E., Santiago, N. J., Vivian, D., Vocisano, C, Castonguay, …, Keller, M. B. (2003). Therapeutic alliance in depression treatment: Controlling for prior change and patient characteristics. *Journal of Consulting and Clinical Psychology, 71,* 997-1006. https://doi.org/10.1037/0022-006X.71.6.997

Kumar, M., Kuria, M. W., Othieno, C. J. & Falkenström, F. (2018). Improving psychotherapies offered in public hospitals in Nairobi, Kenya. Extending practice-based research model for LMICs. *International Journal for Mental Health Systems, 12*, 76. http://www.doi.org/10.1186/s13033-018-0254-7

Kaiser, T. & Laireiter, A.-R. (2019). Daily dynamic assessment and modelling of intersession processes in ambulatory psychotherapy: A proof of concept study. *Psychotherapy Research, 29*(8), 1062-1073. https://doi.org/10.1080/10503307.2018.1497213

Kivity, Y., Levy, K. N., Kolly, S. & Kramer, U. (2019). The therapeutic alliance over 10 sessions of therapy for borderline personality disorder: Agreement and congruence analysis and relation to outcome. *Journal of Personality Disorder*, 28, 1-21, http://www.doi.org/10.1521/pedi_2019_33_376

Lambert, M.J., Whipple, J.L., Kleinstäuber M. (2018). Collecting and delivering progress feedback. A meta-analysis of routine outcome monitoring. *Psychotherapy, 55*(4), 520-537. https://doi.org/10.1037/pst0000167

Linardon, J., Brennan, L., & de la Piedad Garcia, X. (2016). Rapid response to eating disorder treatment: A systematic review and meta-analysis. *International Journal of Eating Disorder, 49*(10), 905-919. https://doi.org/10.1002/eat.22595

Lorenzo-Luaces, L., DeRubeis, R. J., & Webb, C. A. (2014). Client characteristics as moderators of the relation between the therapeutic alliance and outcome in cognitive therapy for depression. Journal of Consulting and Clinical Psychology, 82(2), 368–373. https://doi.org/10.1037/a0035994

Luborsky, L. (1976). Helping alliances in psychotherapy. In J. L. Cleghhorn (Ed.), *Successful psychotherapy* (pp. 92–116). New York, NY: Brunner/Mazel.

Lutz, W., Hofmann, S. G., Rubel, J., Boswell, J. F., Shear, M. K., Gorman, J. M., & Barlow, D. H. (2014). Patterns of early change and their relationship to outcome and early treatment termination in patients with panic disorder. *Journal of Consulting and Clinical Psychology, 82*(2), 287-297. https://doi.org/10.1037/a0035535

Lutz, W., Ehrlich, T., Rubel, J., Hallwachs, N., Röttger, M.A., Jorasz, C., Mocanu, S., Vocks, S., Schulte, D. and Tschitsaz-Stucki, A. (2013). The ups and downs of psychotherapy: Sudden gains and sudden losses identified with session reports. *Psychotherapy Research, 23*(1), 14-24. https://doi.org/10.1080/10503307.2012.693837

Lutz, W., Stulz, N., & Köck, K. (2009). Patterns of early change and their relationship to outcome and follow-up among patients with major depressive disorders. *Journal of Affective Disorders*, *118*(1-3), 60–68. https://doi.org/10.1016/j.jad.2009.01.019

Ogden, K., Barr, J. & Greenfield, D. (2017). Determining requirements for patient-centered care: A participatory concept mapping study. *BMC Health Services Research, 17*, 780, https://doi.org/0.1186/s12913-017-2741-y

Marin-Martinez, F. & Sachez-Meca, J. (2010). Weighting by inverse variance or by sample size in random-effects meta-analysis. *Educational and Psychological Measurement, 70*(1), 56-73. https://doi.org/10.1177/0013164409344534

Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement: Interdisciplinary Research & Perspective*, *2*(4), 201-218. https://doi.org/10.1207/s15366359mea0204_1

Muran, J. C. & Barber, J. P. (2010). *The therapeutic alliance: An evidence-based guide to practice*. New York, NY: Guilfors Press.

Muthukrashna, M. & Henrich, J. (2019). A problem in theory. *Nature Human Behavior*, *3*(3), 221-229. https://doi.org/10.1038/s41562-018-0522-1

Nazar, B., Gregor, K. L., Albano, G., Marchica, A., Lo Coco, G., Cardi, V. & Treasure, J. L. (2017). Early response to treatment in eating disorders: A systematic review and a diagnostic test accuracy meta-analysis. *European Eating Disorders Review, 25*(2), 67-79. https://doi.org/10.1002/erv.2495

Norcross, J. C. & Goldfried, M. R. (2019). *Handbook of Psychotherapy Integration (3th Ed.).* New York, NY: Oxford University Press.

Norcross, J. C. & Lambert, M. J. (2019). *Psychotherapy relationships that work (Third Edition, Vol. I).* New York, NY: Oxford University Press.

Ogles, B. M. (2013). Measuring change in psychotherapy research. In M. Lambert (Ed.). *Bergin and Garfield's handbook of psychotherapy and behavior change* (pp. 134 – 166). New York, NY: Wiley.

Quirk, K., Smith, A., & Owen, J. (2018). In here and out there: Systemic alliance and intersession processes in psychotherapy. *Professional Psychology: Research and Practice, 49*(1), 31-38. http://dx.doi.org/10.1037/pro0000174

Ribero, E., Ribero, A. P., Gonçales, M. M., Horvath, A. O., Stiles, W. B. (2012). Howe collaboration in therapy becomes therapeutic: The therapeutic collaboration coding system. *Psychology and Psychotherapy, 86*(3), 294-314. https://doi.org/10.1111/j.2044-8341.2012.02066.x

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team (2019). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-140, https://CRAN.R-project.org/package=nlme.

*Rubel, J. A., Hilpert, P., Wolfer, C., Held, J., Vîslă, A., & Flückiger, C. (2019). The working alliance in manualized CBT for generalized anxiety disorder: Does it lead to change and does the effect vary depending on manual implementation flexibility? Journal of Consulting and Clinical Psychology, 87(11), 989–1002. https://doi.org/10.1037/ccp0000433

*Rubel, J. A., Rosenbaum, D., & Lutz, W. (2017). Patients' in-session experiences and symptom change: Session-to-session effects on a within- and between-patient level. *Behaviour Research and Therapy, 90*, 58-66. https://doi.org/10.1016/j.brat.2016.12.007

Rubel, J. A., Zilcha-Mano, S., Feils-Klaus, V., & Lutz, W. (2018). Session-to-session effects of alliance ruptures in outpatient CBT: Within- and between-patient associations. *Journal of Consulting and Clinical Psychology, 86*(4), 354-366. https://doi.org/10.1037/ccp0000286

Ryan, R. M., Lynch, M. F., Vansteenkiste, M., & Deci, E. L. (2011). Motivation and autonomy in

counseling, psychotherapy, and behavior change: A look at theory and practice. *The Counseling Psychologist, 39*, 193–260. https://doi.org/10.1177/0011000009359313

Safran, J. D. & Muran, J. C. (2000). *Negotiating the Therapeutic Alliance: A Relational Treatment Guide.* New York, NY: Guilford Press.

Sasso, K. E., Strunk, D. R., Braun, J. D., DeRubeis, R. J., & Brotman, M. A. (2016). A re-examination of process-outcome relations in cognitive therapy for depression: Disaggregating within-patient and between-patient effects. *Psychotherapy Research*, *26*(4), 387-398. https://doi.org/10.1080/10503307.2015.1026423

Scheel, M. J. (2011). Client common factors represented by client motivation and autonomy. The *Counseling Psychologist, 39*, 276-285. https://doi.org/10.1177/0011000010375309

Schuurman, N. K., Ferrer, E., Boer-Sonnenschein, M. de, & Hamaker, E. L. (2016). How to compare cross-lagged associations in a multilevel autoregressive model. *Psychological Methods*, *21*(2), 206-221. https://doi.org/10.1037/met0000062

*Schwartz, C., Hilbert, S., Schlegl, S., Diedrich, A., & Voderholzer, U. (2018). Common change factors and mediation of the alliance-outcome link during treatment of depression. *Journal of Consulting and Clinical Psychology*, *86*(7), 584-592. https://doi.org/10.1037/ccp0000302

Shalom, J. G., Gilboa-Schechtman, E., Atzil-Slonim, D., Bar-Kalifa, E., Hasson-Ohayon, I., van Oppen, P., . . . Aderka, I. M. (2018). Intraindividual variability in symptoms consistently predicts sudden gains: An examination of three independent datasets. *Journal of Consulting and Clinical Psychology*, *86*(11), 892-902. https://doi.org/10.1037/ccp0000344

Sharf, J., Primavera, L. H., & Diener, M. J. (2010). Dropout and therapeutic alliance: A meta-analysis of adult individual psychotherapy. *Psychotherapy, 47*(4), 637– 645. https://doi.org/10.1037/a0021175

Spencer, J., Goode, J., Penix, E., Trusty, W. & Swift, J. K. (2019). Developing a collaborative

relationship with clients during the initial sessions of psychotherapy. *Psychotherapy, 56*(1),7-10, https://doi.org/10.1037/pst0000208

Stewart, M. (2001). Toward a global definition of patient centered care. *British Medical Journal,* 322, 444-445. https://doi.org/10.1136/bmj.322.7284.444

Stewart, G.B., Altman, D.G., Askie, L.M., Duley, L., Simmonds. M.C., & Stewart, L.A. (2012). Statistical analysis of individual participant data meta-analyses: a comparison of methods and recommendations for practice. *PloS One, 7*(10):e46042 https://doi.org/10.1371/journal.pone.0046042

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. https://doi.org/10.1016/j.brat.2010.03.011

Tang, T. Z., & DeRubeis, R. J. (1999). Sudden gains and critical sessions in cognitive-behavioral therapy for depression. *Journal of Consulting and Clinical Psychology, 67*(6), 894-904. https://doi.org/10.1037//0022-006x.67.6.894

*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, 59*(4), 507-517. https://doi.org/10.1037/a0029947

*Tasca, G. A., Compare, A., Zarbo, C., & Brugnera, A. (2016). Therapeutic alliance and binge-eating outcomes in a group therapy context. *Journal of Counseling Psychology, 63*(4), 443-451. https://doi.org/10.1037/cou0000159

Tierney, J.F., Vale, C., Riley, R., Smith, C.T., Stewart, L., Clarke, M. & Rovers (2015) M. Individual participant data (IPD) meta-analyses of randomised controlled trials: guidance on their use. *PLoS Medicine, 12*(7):e1001855. https://doi.org/10.1371/journal.pmed.1001855

Van Landeghem, G., De Fraine, B., & Van Damme, J. (2005). The consequence of ignoring a level

of nesting in multilevel analysis: A comment. *Multivariate Behavioral Research, 40*(4), 423-434. https://doi.org/10.1207/s15327906mbr4004_2

Vasquez, M. J. T. (2007). Cultural difference and the therapeutic alliance: An evidence-based analysis. *American Psychologist, 62*, 878-885. https://doi.org/10.1037/0003-066X.62.8.878

Viechtbauer, W. (2017). metafor: Meta-analysis package for R. R package version 2.0-0.

Viechtbauer, W. & Cheung, M. W. L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods 1*(2), 112–25. https://doi.org/10.1002/jrsm.11

Wampold, B. E., & Imel, Z. E. (2015). *The great psychotherapy debate: The evidence for what makes psychotherapy work (2nd ed.).* New York, NY, US: Routledge.

Wang, L. P. & Maxwell, S. E. (2015). On disaggregating between-person and within-person effects with longitudinal data using multilevel models. *Psychological Methods., 20*(1), 63 – 83. http://www.doi.org/ 10.1037/met0000030.

Wang, L., Zhang, Q., Maxwell, S. E., & Bergeman, C. S. (2019). On standardizing within-person effects: Potential problems of global standardization. *Multivariate Behavioral Research, 54*(3), 382-403. https://doi.org/10.1080/00273171.2018.1532280

*Webb, C. A., Beard, C., Auerbach, R. P., Menninger, E., & Björgvinsson, T. (2014). The therapeutic alliance in a naturalistic psychiatric setting: Temporal relations with depressive symptom change. *Behaviour Research and Therapy, 61*, 70-77. https://doi.org/10.1016/j.brat.2014.07.015

*Weiss, M., Kivity, Y., & Huppert, J. D. (2014). How does the therapeutic alliance develop throughout cognitive behavioral therapy for panic disorder? Sawtooth patterns, sudden gains, and stabilization. *Psychotherapy Research, 24*(3), 407-418. https://doi.org/10.1080/10503307.2013.868947

WHO - World Health Organization (2019). *WHO – Constitution.* https://www.who.int/about/who-

we-are/constitution.

Wiernik, B. M., Kostal, J. W., Wilmot, M. P., Dichert, S. (2017). Empirical benchmarks for interpreting effect size variability in meta-analysis. *Industrial and Organizational Psychology*, 10(3), 472-479, https://doi.org/10.1017/iop.2017.44

Wucherpfennig, F., Rubel, J. A., Hofmann, S. G., & Lutz, W. (2017). Processes of change after a sudden gain and relation to treatment outcome—Evidence for an upward spiral. *Journal of Consulting and Clinical Psychology*, *85*(12), 1199–1210. https://doi.org/10.1037/ccp0000263

*Xu, H., & Tracey, T. J. G. (2015). Reciprocal influence model of working alliance and therapeutic outcome over individual therapy course. *Journal of Counseling Psychology*, *62*(3), 351–359. https://doi.org/10.1037/cou0000089

Zilcha-Mano, S. (2017). Is the alliance really therapeutic? Revisiting this question in light of recent methodological advances. *The American Psychologist*, *72*(4), 311–325. https://doi.org/10.1037/a0040435

*Zilcha-Mano, S., & Errázuriz, P. (2017). Early development of mechanisms of change as a predictor of subsequent change and treatment outcome: The case of working alliance. *Journal of Consulting and Clinical Psychology*, *85*(5), 508-520. https://doi.org/10.1037/ccp0000192

*Zilcha-Mano, S., Errázuriz, P. (2015). One size does not fit all: Examining heterogeneity and identifying moderators of the alliance-outcome association. *Journal of Counseling Psychology, 62*(4), 579-591. https://doi.org/10.1037/cou0000103

*Zilcha-Mano, S., Muran, J. C., Hungr, C., Eubanks, C. F., Safran, J. D., Winston, A. (2016). The relationship between alliance and outcome: Analysis of a two-person perspective on alliance and session outcome. *Journal of Consulting and Clinical Psychology, 84*(6), 484-496. https://doi.org/10.1037/ccp0000058

Figure 1

Figure 1. Investigating the impact of alliance in one session on symptoms in a subsequent session and vice versa. Four within-patient effects of early symptoms and early alliance using session-by-session lag-models from sessions 1 to 7.

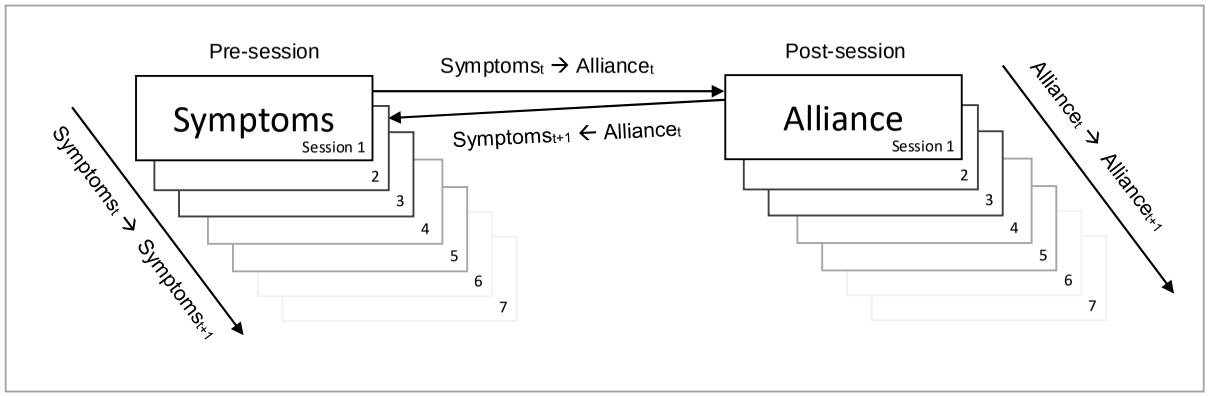

Figure 2: Flowchart of the included and excluded manuscripts

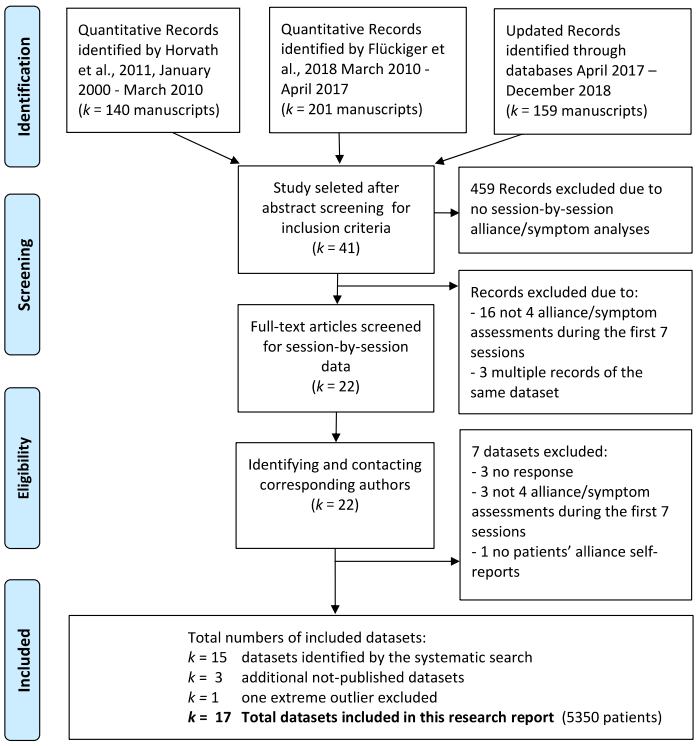

Table 1: Study characteristics of the included datasets

First-Author, (Publication year)	*N*	% Females	Age	Primary Diagnosis	SUD	Type	Duration	Design	Country	Alliance Measure	Symptom Measure	Disorder - specific	Period of time	$Alliance_t \rightarrow Symptoms_{t+1}$ ß (SE)	$Symptoms_t \rightarrow Alliance_t$ ß (SE)
Crits-Christoph (2011)	45	0.56	42.8	MDD	exluded	PsyDyn	16 w	RCT	USA	CALPAS	HAM-D*, BDI-II	yes	1-7 w	-0.190 (0.056)	-0.127 (0.057)
Falkenström (2013)	645	0.74	37.3	Transd.	not ex.	various	1-37 s	other	Sweden	WAI	CORE-OM	no	1-7 s	-0.148 (0.019)	-0.343 (0.017)
Falkenström (2016)	96	0.69	33.8	MDD	excluded	various	14 s	RCT	Sweden	WAI	BDI-II	yes	1-7 s	-0.067 (0.043)	-0.157 (0.042)
Falkenström (2019)	345	0.27	28.9	Transd.	not ex.	various	1-8 s	other	Kenya	WAI	CORE-OM	no	1-7 s	-0.007 (0.058)	-0.121 (0.058)
Fisher (2016)	101	0.62	39.8	Transd.	not ex.	PsyDyn	4-49 s	other	Israel	SRS	ORS	no	2-7 s	-0.001 (0.050)	-0.163 (0.044)
Flückiger (2013)	430	0.62	37	Transd.	excluded	CBT	1-169 s	other	Switzerland	BPSR	OC*, Improv.	no	1-7 s	-0.022 (0.022)	-0.076 (0.219)
Hoffart (2013)	65	0.58	45.2	PTSD	not ex.	CBT	10 s	RCT	Norway	WAI	PSS-SR*, GSI	yes/no	1-7 s	-0.171 (0.054)	-0.014 (0.053)
Huppert (2018)	29	0.44	29	SAD	excluded	CBT	20 s	RCT	Israel	WAI	SPIN	yes	1-7 s	-0.113 (0.070)	-0.124 (0.070)
Rubel (2017)	1550	0.63	36	Transd.	not ex.	CBT	1-109 s	other	Germany	BPSR	GSI	no	1-7 s	-0.039 (0.010)	-0.084 (0.010)
Rubel (2019)	55	0.75	43.9	GAD	not ex.	CBT	14 s	RCT	Switzerland	WAI	BAI	yes	1-7 s	-0.116 (0.055)	-0.132 (0.053)
Schwartz (2018)	193	0.53	46.7	MDD	not ex.	CBT	22-111d	other	Germany	BPSR	BDI-II	yes	1-7 w	-0.057 (0.054)	-0.490 (0.045)
Tasca (2012)	238	1.00	26.1	ED	not ex.	various	14 w	other	Canada	CALPAS	UtR	yes	1-7 w	0.015 (0.036)	-0.100 (0.035)
Webb (2014)	103	0.64	36	MDD	not ex.	CBT	2-26 d	other	USA	WAI	CESD-10	yes	1-4 d	-0.004 (0.086)	-0.423 (0.075)
Weiss (2014)	29	0.58	31	Panic	exluded	CBT	12 s	other	Israel	WAI	PDSS-SR	yes	1-7 s	-0.208 (0.068)	-0.213 (0.066)
Xu (2015)	638	0.46	-	Transd.	not ex.	various	3-14 s	other	USA	WAI	OQ	no	3-7 s	-0.081 (0.035)	-0.119 (0.034)
Zilcha-Mano (2015)	547	0.74	41.3	Transd.	not ex.	various	1-55 s	RCT	Chile	WAI	OQ, Life satis.	no	1-7 s	-0.063 (0.025)	-0.123 (0.023)
Zilcha Mano (2016)	241	0.65	42	MDD	excluded	CBT/AFT	30 s	RCT	USA	WAI	GSI*, Probl.solv.	no	1-7 s	-0.054 (0.028)	-0.398 (0.026)

Note. * = only at post treatment assessment, *N* = patient sample size of the session-by-session analyses, MDD = Major Depressive Episode, Transd.= Transdiagnostic samples, PTSD = Post Traumatic Stress Disorder, ED = Eating Disorder, GAD = Generalized Anxiety Disorder, SAD = Social Anxiety Disorder, SUD = Substance Use Disorder, excluded = SUD excluded from the primary study, not ex. = SUD not excluded from the primary study, PsyDyn = Psychodynamic Therapy, CBT = Cognitive Behavioral Therapy, AFT = Alliance Focused Therapy, various = multiple therapy approaches, Duration: Prescriptive duration of a therapy; w =

weeks, s = sessions, d = days, RCT = Randomized Clinical Trial, other = other design than RCT under naturalistic conditions, CALPAS = California Psychotherapy Alliance Scale, WAI = Working Alliance Inventory - Short Version, BPSR = Bern Post-Session Report, SRS = Session Report Scale, HAM-D = Hamilton Depression Rating Scale (patient selfreport), BDI-II = Beck Depression Inventory II, CORE-OM = Clinical Outcomes in Routine Evaluation – Outcome Measure, ORS = Reliable Change Chart, PSS-SR = PTSD Symptom Schale – Selfreport, GSI = Global Severity Index, UtR = Urge to Restrict, CESD-10 = Center of Epidemiologic Studies Depression Scale 10, SPDSS-SR = Panic Disorder Severity Scale – Selfreport, OQ = Outcome Questionnaire, Life satis. = Life Satisfaction Scale, OC = Outcome Composite of seven outcome measures, Improv. = Weekly Symptom Improvement Items, BAI = Beck Anxiety Inventory, Social Phobia Inventory, SPIN = Social Phobia Inventory, Period of time = Analyzed session range (assessed in the primary studies)

Table 2
Omnibus test results of the session-by-session lag-models (k = 17)

	$\beta_{unadjusted}$ *[95% CI]*	τ^2	*80% CrIn*	$\beta_{adjusted}$ *[95% CI]*	τ^2	*80% CrIn*
$Alliance_t$ → $Alliance_{t+1}$	.044 [-.002, .130]$^{+}$	.007	-.064, .152			
$Alliance_t$ → $Symptoms_{t+1}$	-.072 [-.101, -.042]***	.002	-.131, -.034	-.065 [-.092, -.038]***	.002	-.115, -.015
$Symptoms_t$ → $Symptoms_{t+1}$	.082 [.034, .130]***	.008	-.034, .198			
$Symptoms_t$ → $Alliance_t$	-.194 [-.260, -.127]***	.017	-.359, -.029	-.148[-.215, -.081]***	.017	-.313, .017

Note. β = meta-analytic estimates of the standardized WP beta-coefficients, adjusted beta-coefficients are adjusted for prior symptoms (at t) or alliance (at t-1) respectively, *CI* = confidence interval, τ^2 = absolute value of the true variance (heterogeneity). *CrIn* = credibility interval, *** $p < .0001$, $^{+}$ $p < .010$;

Table 3

Omnibus test results of the cross-level interaction Alliance$_t$ → Symptoms$_{t+1}$

	$\beta_{unadjusted}$ *[95% CI]*	τ^2	*80% CrIn*	$\beta_{adjusted}$ *[95% CI]*	τ^2	*80% CrIn*
Alliance$_{1\text{-}7}$	-.028 [-.042, -.014]***	<.0001	-.036, -.020	-.025 [-.038, -.012]***	<.0001	-.029, -.022
Symptoms$_{1\text{-}7}$	.030 [.008, .051]**	.0005	.000, .059	.027 [.004, .051]*	.0008	-.010, .065

Note. β = meta-analytic estimates of the standardized WP beta-coefficients, adjusted beta-coefficients are adjusted for prior symptoms (at t), *CI* = confidence interval, τ^2 = absolute value of the true variance (heterogeneity). *CrIn* = credibility interval, * $p < .01$, ** $p < .001$, *** $p < .0001$