THE JOURNAL OF CHILD PSYCHOLOGY AND PSYCHIATRY
ACAMH THE ASSOCIATION FOR CHILD AND ADOLESCENT MENTAL HEALTH

Therapist variability, patient-reported alliance, and clinical outcomes in adolescents undergoing mental health treatment - a systematic review and meta-analysis

Journal:	*Journal of Child Psychology and Psychiatry*
Manuscript ID	JCPP-RR-2016-00431.R3
Manuscript Type:	Practitioner Review
Date Submitted by the Author:	n/a
Complete List of Authors:	Murphy, Regina; NHS Orkney, Child and Adolescent Mental Health Service Hutton, Paul; Edinburgh Napier University, School of Health & Social Care
Key Words:	Adolescence, Meta-analysis, Outcome, Psychotherapy

Therapist variability, patient-reported therapeutic alliance, and clinical outcomes in adolescents undergoing mental health treatment - a systematic review and meta-analysis

Regina Murphy[1]
Paul Hutton[2]

[1]Child & Adolescent Mental Health Service
Kirkwall Health Centre
New Scapa Road
Kirkwall
Orkney
KW15 1BX

[2]School of Health & Social Care
Edinburgh Napier University
9 Sighthill Court
Edinburgh
EH11 4BN

Abstract

Background: Previous meta-analyses have only found small correlations (r=0.10 to r=0.19) between therapeutic alliance and clinical outcomes in samples of adolescents receiving psychological therapy. Although study-level variables have been found to moderate this, little is known about the impact of therapist variability. The present meta-analysis aimed to address this gap by using patient-therapist ratio as a moderator variable.

Methods: Contrary to previous reviews of adolescent alliance, individual effect sizes were extracted using a pre-registered conceptual hierarchy. Controlling for treatment-level confounds, a random effects meta-analysis assessed the moderating effect of patient-therapist ratio on the alliance-outcome relationship in pre-defined single-predictor and multi-predictor meta-regressions.

Results: The alliance-outcome relationship was found to be larger than previously thought (k=28, N=2911, r=0.29, 95% Confidence Interval 0.21, 0.37; p<.0001, I^2 = 80%). When study samples exceeding the adolescent 12-19 age range were removed, the correlation rose (k=15, N=1797, r=0.34, 95% Confidence Interval 0.23, 0.45; p<.0001, I^2 = 83%). In contrast to research with adults, patient-therapist ratio did not moderate this relationship in either single-predictor (p=.26) or multi-predictor (p=.22) models.

Conclusions: The alliance-outcome relationship for adolescents was larger than previously thought, and comparable to estimates in adult samples. The failure of patient-therapist ratio to moderate its strength however, challenges the hypothesis that variability in therapist characteristics is an important determinant of the alliance-outcome effect in this age group.

Keywords: alliance; alliance-outcome relationship; adolescent therapy; therapist variability; meta-analysis

1. Introduction

1.1 Alliance Conceptualisation & Rater Perspective

Alliance is an umbrella term used to denote relational aspects of psychological treatment. It was first spoken of within the context of psychodynamic therapy and had close conceptual links to constructs such as positive transference (Freud, 1958). Due to increased scientific interest, our understanding of the alliance concept has evolved, giving rise to a range of definitions and associated labels. Drawing on models first proposed by Luborsky (1976) and Bordin (1979), alliance is now commonly described as a multi-dimensional construct. As such, it incorporates elements relating to reciprocal affect, patient-therapist agreement, and collaboration.

While differences in treatment contexts may be substantial across young patients, alliance is undoubtedly influenced by how adolescents experience service involvement. This is also evident in well-documented links between adolescent alliance ratings and factors such as retention, engagement, and service satisfaction (Florsheim, Shotorbani, Guest-Warnick, Barratt, & Hwang, 2000; Shirk & Saiz, 1992; Hawley & Weisz, 2005). Examining adolescent perceptions of alliance can therefore provide valuable insights into patient experiences of participating in psychological treatment and inform routine practice in mental health services. With the importance of this in mind, the present meta-analysis specifically focuses on alliance as assessed through adolescent self-report.

1.2 Alliance-Outcome Relationship

The link between alliance and treatment outcome has been quantified over decades of empirical research and yields consistently replicable positive correlations. With effect sizes ranging from $r = 0.22$ to $r = 0.28$, meta-analytic estimates of the alliance-outcome

relationship appear to be fairly robust and nearly reach moderate magnitudes for adult mental health patients (Martin, Garske, & Davis, 2000; Horvath, Del Re, Flückiger, & Symonds, 2011). At present, differences in adult alliance can therefore be linked to approximately 5-8% of variability in treatment outcomes. While this might only represent a fraction among other variables, alliance has become a commonly referenced predictor of treatment success and is presumed to contribute to symptomatic improvements in adult patients (e.g. Falkenström, Granström, & Holmqvist, 2013). As recent reviews indicate, the alliance-outcome relationship in adolescent mental health patients is somewhat weaker than its adult counterpart. With current effect sizes ranging from $r = 0.10$ to $r = 0.19$, meta-analytic estimates for this age group are clearly small in size (McLeod, 2011; Shirk et al., 2011). Differences in adolescent alliance can thus only be linked to approximately 1-4% of variability in treatment outcomes, indicating little prognostic value.

1.3 Moderator Variables in Adolescent Patients

While effect size differences between adults and adolescents are now well documented, the focus has gradually shifted towards examining effect size variation within clinical populations. In recent meta-analyses, numerous variables have been shown to moderate the alliance-outcome relationship in adolescent patient samples. Problem type, referral source, and timing of alliance assessment, for instance, are among a range of identified moderators and can be linked to some of the variance observed in alliance-outcome correlations (Shirk & Karver, 2003; McLeod, 2011). On a theoretical level, these findings indicate that estimates of the alliance-outcome relationship are not uniform across adolescent patients. As a number of primary studies confirm, alliance may not always be linked to treatment outcome to the same degree, with some correlations not

reaching statistical significance (Kendall, 1994; Kendall et al., 1997). This suggests that even positive alliance may not consistently predict treatment success in the presence of certain variables.

1.4 Role of Therapist Variability

While alliance-outcome moderators have been linked to a number of factors within the therapeutic process, comparatively little is known about the impact therapist characteristics have on the alliance-outcome relationship. As highlighted in relevant systematic reviews, therapist variables have been widely neglected in empirical research and received unexpectedly little attention in clinical discourse (e.g. Ackerman & Hilsenroth, 2003). With the exception of individual studies in the adult field, the impact of therapist variability within psychological treatment is still underexplored in the adolescent population (e.g. Jung, Wiesjahn, Rief, & Lincoln, 2015; Heinonen et al., 2014). Considering the pivotal role therapists play in treatment delivery and patient discharge, this knowledge gap is more than surprising. Previous meta-analyses addressed this issue to varying degrees, but largely failed to incorporate therapist variables altogether (e.g. Shirk & Karver, 2003; Shirk, Karver, & Brown, 2011). As an appraisal of eligible literature suggests, this relates to a lack of relevant data collection across primary studies and can often be linked to inadequate reporting of therapist-related information (Karver, Handelsman, Fields, & Bickman, 2005).

1.5 Rationale

Considering the paucity of available therapist data across adolescent literature, informative conclusions about the role of therapist variability could not be drawn so far. Accordingly, it remains unclear to what extent differences between therapists actually

affect the alliance-outcome relationship. Using a comparatively novel methodological approach, the present meta-analysis aims to address this gap. Following a recent review of adult literature, the impact of therapist variability on the alliance-outcome relationship was examined using patient-therapist ratio as a moderator (Del Re, Flückiger, Horvath, Symonds, & Wampold, 2012). Also referred to as PTR, this variable represents the quotient of patients to therapists within a given treatment sample and can be used to index the degree to which within-group variability may affect measured variables. When PTR is very large, e.g. with all patients seeing the same therapist, therapist variability is non-existent and does therefore not affect alliance or outcome. In such a scenario, variance in the alliance-outcome relationship should primarily be influenced by variability in patient characteristics. This changes, however, as PTR values decrease.

1.6 Hypothesis

If differences between therapists have an impact on the link between alliance and outcome, shifts in therapist variability should influence the strength of the alliance-outcome relationship. Accordingly, we should expect patient-therapist ratio to moderate effect sizes in the present meta-analysis. This would be in line with findings reported by Del Re et al. (2012).

2. Method

2.1 Review Registration

This meta-analysis was conducted in accordance with AMSTAR criteria for systematic reviews (Shea et al., 2007). The corresponding protocol was pre-registered on PROSPERO in June 2015 and can be found under registration number 42015023275. Use of pre-registration is strongly recommended in PRISMA guidelines and marks an advance

on previous meta-analyses in the alliance field (Liberati et al., 2009). As with clinical trials, it is a widely accepted mark of quality and minimises the risk of methodological decisions being unduly influenced by findings (Quintana, 2015).

2.2 Literature Search

This meta-analysis synthesised data on the alliance-outcome relationship in adolescent study cohorts. To identify relevant literature, the following search methods were deployed: (1) Database searches of articles published before June 2015 were conducted using PsycINFO, PsycARTICLES, EMBASE, Medline, CINAHL, ERIC, PubPsych, PsyJournals, and The Cochrane Library. Unpublished literature was searched for using EThOS, OpenGrey, and Dissertation Abstracts International. To restrict the search to studies with predominantly adolescent samples, the terms *alliance*, *therapeutic relationship*, and *bond* were crossed with *youth*, *adolescent*, *child,* and *family*. (2) Reviews of alliance-related topics were screened to identify reference trails.

Study titles and abstracts were screened for eligibilty according to below selection criteria. Duplicates and clearly irrelevant papers were removed. The remaining studies were then examined by full text and a decision on final eligibility was made. Corresponding authors were contacted to resolve uncertainty over eligibility and were asked to provide additional information for effect size computations where required.

2.3 Selection Criteria

While the present meta-analysis aimed to provide an updated effect size estimate of the alliance-outcome relationship, it was not designed to replicate prior publications. To attain a reasonable degree of comparability however, selection criteria were modelled on

those described in McLeod (2011) and read as follows: (1) the study needed to be clinical; (2) the study sample needed to comprise at least five participants; (3) interventions administered to participants needed to be psychological and target mental health difficulties; (4) studies involving participants with medical diagnoses or neurodevelopmental conditions were excluded; (5) studies involving participants with significantly impaired cognitive functioning were excluded; (6) the mean age of each study sample needed to fall within the 12-19 age range; these restrictions differ from criteria presented in previous meta-analyses to exclude samples primarily comprised of children; (7) the study needed to use formal measures of alliance and outcome; (8) outcome measures used in the study needed to assess symptom severity, global functioning, or emotional distress; (9) alliance needed to be assessed via self-report; (10) the relationship between alliance and outcome needed to be quantifiable, i.e. correlation coefficients had to be obtainable; (11) the study needed to be written in English or German; these language criteria differ from previous meta-analyses to facilitate a more comprehensive screening of literature.

2.4 Data Extraction

2.4.1 Strategy

Data extraction was carried out using predefined templates. Due to resource limitations, this process was managed by a single reviewer. Since this deviates from existing AMSTAR guidelines, it may be associated with some methodological concerns, e.g. risk of error and bias (Shea et al., 2007). We therefore advise readers to exercise caution when interpreting our findings. To increase the robustness of our extraction procedure however, all data were double-extracted and checked for errors. Where there was ambiguity with regard to selection decisions, discussions were held between both researchers.

2.4.2 Sample Characteristics

Information on the following variables was extracted for every study: (a) Sample Size; (b) Sample Mean Age; (c) Sample Age Range; (d) Gender Distribution Percentage; (e) Study Setting; (f) Target Problem; (g) Treatment Reception. For studies with more than one participant sample, information was extracted for each treatment group.

2.4.3 Patient-Therapist Ratio

Patient-therapist ratio was computed by dividing participant N by therapist N. Due to participant attrition, some sample sizes decreased throughout the course of administered treatment. To compute an adequately representative PTR value, the following criteria were established: (1) If sample size decreased throughout the first half of treatment, post-attrition participant numbers were used to calculate PTR; (2) For studies were information on attrition was not reported, intention-to-treat participant numbers were used to calculate PTR.

2.4.4 Treatment Variables

The impact of individual differences between therapists may differ depending on the nature of provided interventions. Psychological treatment characterised by a more structured session delivery, for instance, can be expected to decrease variation among those who administer it. It was therefore important to determine whether PTR moderates the strength of the alliance-outcome relationship when variables pertaining to treatment structure are controlled for. To assess this, two categorical variables were extracted for subsequent analysis: (a) CBT Elements (yes, no); (b) Manual Use (yes, no).

2.4.5 Measure Variables

Data on measure-related variables were extracted for each study: (a) Alliance Measure; (b) Outcome Measure, (c) Outcome Domain; (d) Timing of Alliance Assessment; (e) Timing of Outcome Assessment. The latter two variables were coded as early-treatment, mid-treatment, late-treatment, end-of-treatment, or treatment follow-up.

2.5 Data Analysis

2.5.1 Quality Appraisal

Assessment of study-level bias was conducted using the Quality Assessment Tool for Quantitative Studies (Effective Public Health Practice Project, 1998). Details of quality appraisal ratings were used to interpret study findings and are presented in *Table 3*.

2.5.2 Publication Bias

Presence of publication bias was initially assessed via examination of a funnel plot. Plot asymmetry was then determined statistically using the Egger test (Egger, Smith, Schneider, & Minder, 1997).

2.5.3 Effect Size Extraction & Computation

Effect sizes for alliance-outcome relationships were extracted as Pearson's *r*. Where correlation coefficients were not reported, techniques described by Rosenthal (1994) were used to compute *r* values from alternative information. As observed in clinical practice, outcome is typically assessed against treatment goals established in collaboration with each patient. As seen in routinely used screening tools such as CORE, these typically fall within three broad domains (Evans et al., 2002). To reflect this in the selection of effect sizes, the following conceptual hierarchy was established to guide the extraction of correlational data: (1) Symptom Severity, (2) Global Functioning, and (3)

Emotional Distress. Hierarchical approaches to outcome selection have been used in several recent meta-analyses of psychiatric treatments and are preferable to computing an aggregate effect size for each study, i.e. an average of all reported outcomes (e.g. Leucht et al., 2013; Hutton & Taylor, 2014; Stovell, Morrison, Panayiotou, & Hutton, 2016). As highlighted in recent publications, it allows inclusion of a larger number of eligible studies and yields a much more interpretable effect size estimate. It also reduces the risk of meta-analytic findings being influenced by decisions made by original authors, especially in relation to the number and type of outcomes they decided to measure, or indeed the number and type of outcomes they decided to report.

Where studies used multiple measures to assess alliance-outcome correlations within the same outcome domain, the following a priori criteria were used to determine effect size selection: (1) Where multiple measures assessed the same construct, the most frequently used measure was prioritised; (2) Where different rater versions of the same outcome measure were available, participant self-report was prioritised; (3) Where outcome was measured at multiple time points, end-of-treatment assessment was prioritised; where clinical outcome was only measured at post-treatment follow-up, the first follow-up assessment was prioritised; (4) Where alliance was measured at multiple time points, early-treatment alliance ratings were prioritised.

The meta-analytic estimate of the alliance-outcome relationship, i.e. the mean weighted effect size across included studies, was computed using a random effects model (Raudenbush, 1994). Keeping in line with the methodology used by Del Re et al. (2012), a restricted maximum-likelihood estimator was used to approximate the model. Effect sizes contributing to this cross-domain average were selected using the hierarchy

described above. Applying the same model, pooled domain-specific effect size estimates were also computed. All analyses were performed using the '*metafor*' package for meta-analyses in R (Viechtbauer, 2010). To adjust for bias in the *r* distribution, extracted coefficients were converted into Fisher's *Z* prior to analysis (Hedges & Olkin, 1985).

2.5.4 Heterogeneity

The I^2 statistic was computed to index the proportion of effect size variability attributable to between-study heterogeneity. Since methodological differences are common in the alliance field, heterogeneity was expected to be high.

2.5.5 Moderator Analysis

2.5.5.1 Single-Predictor Meta-Regression

To assess whether patient-therapist ratio moderates the strength of the alliance-outcome relationship, PTR was entered into a single-predictor meta-regression model as a moderator variable. This generated the following regression model:

$$\boldsymbol{ES = \beta_0 + \beta_1\,(PTR) + v;}$$

with *ES* designating the effect size estimate of the alliance-outcome relationship, and *v* designating within-study and between-study error variance.

2.5.5.2 Multi-Predictor Meta-Regression

To assess whether PTR remains a statistically significant moderator in the presence of potential confounds, CBT Elements and Manual Use were entered into a multi-predictor meta-regression model as additional moderator variables. This yielded the following regression model:

$$ES = \beta_0 + \beta_1 (\text{PTR}) + \beta_2 (\text{CBT Elements}) + \beta_3 (\text{Manual Use}) + v;$$

with *ES* designating the effect size estimate of the alliance-outcome relationship, and *v* designating within-study and between-study error variance.

3. Results

3.1 Literature

Literature search produced an initial pool of 14, 492 electronic records. After removal of duplicate entries, 5784 records were screened by title and abstract. Throughout this process, a substantial number of previously unidentified duplicates were excluded. The remaining 178 studies were examined by full text, with 62 studies meeting initial eligibility criteria (*Figure 1*). For 35 of these studies, information was insufficient to compute usable effect sizes. The final literature sample comprised 27 studies and a total of 28 independent effect sizes. 10 of these were derived from unpublished doctoral dissertations. Information on the 151 studies excluded following full-text screening is provided as supplementary material.

All included studies were completed between 1992 and 2015, with 12 published after 2010. Despite efforts to identify relevant non-English literature, only one study in the present sample was written in German (Cropp et al., 2008).

3.2 Sample Characteristics

Table 1 displays sample characteristics and other relevant variables. Overall, included studies comprised a total of 2911 participants (M=103.96, SD=101.89, range 20-470). The mean age reported for participant samples ranged from 13.10 to 18.40 years

(M=15.18, SD=1.28), with a total cross-study age range of 7 to 25 years. For one study, sample mean age was not obtainable. On average, samples were comprised of approximately 51% female participants (M=50.89, SD=23.95, range 15-100%). 21 of the 27 studies included in this sample (78%) were conducted in the USA. Substance abuse, behavioural difficulties, and mixed mental health presentations were addressed in 7 studies (26%) each. Two studies focused on eating disorders (7%). PTSD, OCD, anxiety and mood disorders were addressed in one study (4%) each.

3.3 Patient-Therapist Ratio

Extracted PTR values ranged from 0.96 to 102 (M=14.62, SD=22.26) and displayed considerable variability across study samples. For 7 studies, PTR values could not be computed and were not available from the authors.

3.4 Treatment Variables

The treatment approaches used in 19 study samples (68%) drew on treatment elements akin to CBT. Treatment approaches used in the remaining samples were mainly categorised as psychotherapeutic or counselling-based. The use of treatment manuals was confirmed for 10 study samples (36%). The majority of studies did not report information on this variable however.

3.5 Measure Variables

Table 2 presents an overview of measure-related variables. Overall, 12 different measures of therapeutic alliance were used across included studies. The Working Alliance Inventory (Horvath & Greenberg, 1989) was used in 10 distinct studies (37%) and stood out as the most frequently applied tool. The Therapeutic Alliance Scale for

Children/Adolescents (Shirk & Saiz, 1992), the Therapeutic Alliance Quality Scale (Bickman et al., 2010), and the Penn Helping Alliance Questionnaire (Luborsky, 1976) were used in three studies each and thus covered one third of the meta-analytic sample in total (33%). The remaining eight alliance measures were used in one study respectively. In 13 studies, just less than half of the present literature sample, therapeutic alliance was assessed early in treatment (48%).

Overall, 20 distinct measures were used to assess clinical outcome. With a total of 20 studies (74%), outcome assessment was most frequently conducted at the end of treatment, i.e. the ultimate or penultimate treatment session. Four studies assessed clinical outcome at post-treatment follow-up (15%). Exact follow-up timing ranged from 3 months to 6 months post-treatment. Early, mixed, and late outcome assessments were conducted in one study each. With a total of 20 studies (74%), symptom severity was the most commonly encountered outcome domain. Global functioning was assessed in seven studies (26%). Outcome measures evaluating emotional distress were only used in two studies (7%). In total, two studies assessed clinical treatment outcome for multiple domains.

3.6 Quality Assessment

Table 3 presents ratings on global study quality and five tool subdomains for the present literature sample. Since the subdomain covering confounding variables was only relevant to studies conducting group comparisons, it was excluded in the present quality assessment. Most studies recruited participants through referral from mental health professionals, inpatient settings, or schools and had participation rates above 60% after initial eligibility screenings. Approximately 67% of studies had moderate evidence of

selection bias and recruited participant samples which were adequately representative of the target population. Using the EPHPP tool definition, most studies were categorised as cohort studies and comprised only one participant sample (78%). Six studies were identified as Randomised Controlled Trials or Controlled Clinical Trials and were thus rated as strong on study design. Where studies used secondary data analysis, the design of the original study was assessed.

Most studies did not report enough information to draw definitive conclusions about the blinding of participants and outcome assessors. Following tool guidelines, these studies were rated as moderate in this subdomain. Since studies assessed therapeutic alliance through self-report questionnaires, participants were inadvertently alerted to the relevance of their patient-therapist relationship. Five studies, however, made conscious efforts to reduce possible reporting bias, e.g. by assuring participants that completed questionnaires would not be viewed by therapists, or by designating questionnaire collection to another researcher. For a number of studies, it was difficult to establish whether participating therapists were aware of the research questions and the participant status of their patients. Overall, however, it appeared that therapists had knowledge of administered measures.

3.7 Publication Bias

Visual examination of the funnel plot displayed in *Figure 2* indicated that extracted effect sizes were not symmetrically distributed, with several data points located outside the right-hand side of the funnel shape. The Egger test supported this impression and revealed significant plot asymmetry ($z = 2.1058$, $p = .04$). Assuming that plot asymmetry arose from the presence of publication bias, the trim and fill method (Duval & Tweedie, 2000) was used as a statistical correction procedure. As displayed in *Figure 3*, the

corresponding analysis estimated two missing studies on the left-hand side of the funnel plot and added these for subsequent analysis.

3.8 Effect Size Computation

Across the identified sample, all extracted correlations indicated a positive relationship between quality of patient-therapist alliance and quality of clinical outcome. This is in line with findings reported in previous meta-analyses (Shirk, Karver, & Brown, 2011; McLeod, 2011; Shirk & Karver, 2003). Extracted alliance-outcome correlations ranged from $r = 0.03$ to $r = 0.71$. *Figure 4* presents a forest plot of all extracted effect sizes and filled studies. The mean weighted effect size of the alliance-outcome relationship, computed drawing on 28 extracted correlations, was $r = 0.29$ (95% CI: 0.21, 0.37; $p<.0001$). Based on the guidelines established by Cohen (1988), this effect size was interpreted as small to moderate in magnitude. As expected, heterogeneity among effect sizes was significant ($Q = 135.11$, $p<.0001$, $I^2 = 80.43\%$), with approximately 80% of effect size variability being attributable to between-study differences. Overall, these results indicate that differences in alliance were associated with approximately 8% of the outcome variance across included studies.

Applying the same meta-analytic model, domain-specific effect size estimates were computed in a separate analysis. The mean weighted effect size for studies assessing symptom severity, computed drawing on 21 pooled correlations, was $r = 0.29$ (95% CI: 0.20, 0.38; $p<.0001$). As described above, this estimate was also small to moderate in magnitude. Again, heterogeneity among effect sizes was high ($Q = 109.24$, $p<.0001$, $I^2 = 82.87\%$), with approximately 83% of effect size variability being attributable to between-study differences. According to pooled data from eight independent samples, the mean weighted effect size for studies assessing global functioning was $r = 0.33$ (95% CI: 0.23,

0.42; p<.0001). Effect size heterogeneity within this outcome domain was not significant (Q = 9.70, p=.21, I^2 = 32.41%). Since only two extracted correlations assessed emotional distress, computation of a domain-specific effect size estimate did not yield statistically significant results (r = 0.40, 95% CI: -0.19, 0.77; p=.18).

3.8.1 Sensitivity Analyses

3.8.1.1 Age Range

The mean age of study samples needed to fall within the 12-19 range for this meta-analysis. These restrictions were established to exclude samples primarily comprised of non-adolescent participants. While all reported age means were in line with these specifications, only 56% of extracted effect sizes had participant samples with age ranges falling within the 12-19 bracket. To examine whether this had an impact on obtained meta-analytic results, relevant effect sizes were removed for analysis. The resulting pooled effect size estimate of the alliance-outcome relationship was still small to moderate, but somewhat larger in magnitude (r = 0.34; 95% CI: 0.23, 0.45; p<.0001). Heterogeneity remained high (Q = 82.93, p<.0001, I^2 = 83.11%).

3.8.1.2 Common Source Bias

The inclusion of study-level measures derived from self-report can inflate the magnitude of effect sizes through common source bias, particularly when estimates are computed as correlation coefficients. If present effect size estimates were affected by such inflation, excluding studies with outcomes derived from self-report should yield an average effect size that is markedly smaller than overall correlations (r = 0.29 to r = 0.34). The results of a post-hoc sensitivity analysis did not support this hypothesis however, revealing an effect

size estimate very much in line with original findings (r = 0.33, 95% CI 0.20, 0.46; p<.0001, I^2 = 76%).[1]

3.9 Meta-Regression

The presence of significant heterogeneity indicates that study-level variables might moderate the magnitude of effect sizes across included studies. However, PTR was not found to be a significant moderator of effect sizes within a single-predictor meta-regression analysis (β = −0.0022, 95% CI: -0.01, 0.00; p=.26). Consequently, effect size heterogeneity remained high within this model (Q = 84.51, p<.0001, I^2 = 74.12%). When CBT Elements and Manual Use were controlled for in a multi-predictor meta-regression analysis, these findings did not change (β = −.0029, 95% CI: -0.01, 0.00; p=.22). CBT Elements (β = 0.1498, 95% CI: -0.06, 0.36; p=.16) and Manual Use (β = −.1291, 95% CI: -0.38, 0.12; p=0.31) were not identified as significant moderators either (Q= 75.60, p<.0001, I^2 = 73.50%).[2]

4. Discussion

4.1 Moderator Analysis

Contrary to expectations, patient-therapist ratio was not identified as a significant moderator in this meta-analysis, even after treatment-level confounds were controlled for. The absence of a moderating effect does not only deviate from findings reported by Del Re et al. (2012) but also challenges the hypothesis that shifts in therapist variability influence the strength of the alliance-outcome relationship. More specifically, it seems to indicate that variance in the alliance-outcome relationship can be equally linked to

[1] When studies with outcomes derived from other sources were excluded in a different sensitivity analysis, the resulting effect size estimate was comparable (r = 0.31, 95% CI 0.23, 0.37; p<.0001, I^2 = 37%).
[2] The use of manuals has previously been suspected to affect alliance quality. While the meta-regression results are thus somewhat surprising, they are in line with findings reported by Langer, McLeod and Weisz (2011).

patients and therapists. Interestingly, R^2 values obtained in both meta-regression analyses were below 1% and suggested that PTR did not account for any of the variability observed in synthesised effect sizes. This was also reflected by the substantial effect size heterogeneity (I^2 > 70%) remaining in both meta-regression models. Overall, these findings suggest that other, not presently assessed, factors might account for the significant variability observed across alliance-outcome correlations.

4.2 Effect Size

Despite deploying more stringent selection criteria than previous publications, the present meta-analysis identified the largest pool of alliance-outcome correlations derived from predominantly adolescent samples. The average effect size was equivalent to 0.29 and thus noticeably larger than the estimate of the most recent similarly sized meta-analysis (r= 0.10; McLeod, 2011). When study samples breaching the 12-19 age range were removed in a sensitivity analysis, the effect size even rose to 0.34. As these findings indicate, the strength of the adolescent alliance-outcome relationship is markedly stronger than previously assumed and can be compared to estimates reported for adult mental health patients (r = 0.28; Horvath et al. 2011). More specifically, these findings also suggest that alliance may account for an unprecedented 8-12% of variability in treatment outcomes for adolescent patients. This does not only challenge conclusions drawn in previous reviews of youth alliance, but also refocuses attention on the importance of relationship variables in psychological treatment.

4.3 Strengths & Limitations

While a number of methodological differences distinguish the present meta-analysis from recent predecessors, effect size extraction is perhaps the most significant point of

deviation. Since many primary studies report multiple alliance-outcome correlations for the same treatment group, reviewers are tasked with distilling this information and extracting one effect size for analysis. Previous meta-analyses have attempted to aggregate relevant data and averaged all reported correlations to obtain a mean (e.g. Shirk, Karver, & Brown, 2011). Unfortunately, estimates derived through aggregation are at high risk of being adversely influenced by decisions made in the original studies and often incorporate measures of unclear clinical relevance, e.g. attendance rates (McLeod, 2011). Perhaps most importantly however, estimates based on an aggregate of multiple outcomes are also difficult for practitioners to interpret and may not provide clinically meaningful information. This makes drawing informative conclusions with regard to clinical practice very challenging. To avoid such pitfalls, the present meta-analysis used a pre-determined hierarchy-based approach and extracted single unaveraged correlations for three relevant outcome domains. While this may have introduced significant heterogeneity ($I^2 \geq 70\%$) across effect sizes, it also provided transparent information on the clinical relevance of extracted correlations and hopefully facilitated more meaningful conclusions about the alliance-outcome relationship.

Despite other efforts to conduct a methodologically robust meta-analysis, a number of noteworthy shortcomings still exist. The extraction of study-level PTR values, for instance, was more problematic than originally anticipated and resulted in missing values for 7 included studies. In all cases, this was due to unavailable information on therapist numbers and could not be resolved through initiated correspondence. To avoid reductions in PTR sample size, missing values were initially replaced by the PTR mean at analysis stage. When mean substitution is used to address missing data, researchers act on the assumption that the variable mean is sufficiently representative of missing observations.

As Pigott (1994) highlighted, however, this can result in an artificial deflation of variability in the affected variable and skew further statistical analysis. When applied to a moderator, mean substitution can thus affect observed variance and artefactually mask significant moderator effects. To assess this in our present findings, both meta-regression models were repeated using complete case analysis, i.e. pairwise removal of missing values. Since obtained results did not change upon re-analysis, both missing data approaches seemed to be interchangeable in this instance. Nevertheless, these findings do not confirm optimal suitability of either strategy. While the present meta-analysis did not assess this in further detail, alternative approaches such as regression imputation (Buck, 1960) may have been more appropriate.

Some individual studies (e.g. Ormhaug, Shirk, & Wentzel-Larsen, 2015) suggest that the strength of the alliance-outcome correlation differs across individual rater perspectives, i.e. patient, therapist, and observer. The meta-analysis conducted by McLeod (2011) however, does not support this hypothesis: Follow-up contrasts of data obtained from 38 studies found no difference between these alliance rater groups in relation to the strength of the alliance-outcome correlation. Focusing on patient ratings of alliance should therefore not, by itself, be expected to skew meta-analytic findings. Nevertheless, it is important to highlight that the present findings may still not be entirely free from bias: Since adolescents were asked to rate the patient-therapist alliance, attention was inadvertently drawn to the importance of this variable and may have elicited unrepresentative alliance reports (Furnham, 1986). As established through quality assessments, only five studies seemed to have taken additional measures to reduce the probability of such response bias. Consequently, it is not possible to rule out that primary

alliance ratings potentially contributed to an inflated effect size estimate of the alliance-outcome relationship.

4.4 Practice Implications

While the results of previous meta-analyses have questioned the role of therapeutic alliance in youth treatment (McLeod, 2011), the present meta-analysis highlighted its clinical relevance for adolescent patient groups. As present effect sizes indicate, the meta-analytic estimate of the alliance-outcome relationship is markedly larger than previously reported. Accordingly, alliance has a moderate link to subsequent outcome and may be a comparatively reliable predictor of clinical progress.

Since alliance was assessed through patient self-report across included studies, the present findings offer insights into the effect of the patient perspective and permit comparatively specific conclusions regarding the alliance-outcome relationship. This meta-analysis is thus the first to highlight a concrete link between adolescent perceptions of alliance and treatment outcomes in three distinct domains, i.e. symptom severity, global functioning, and emotional distress. Applied to routine mental health practice, these findings emphasise the informational value that regular monitoring of patient-perceived alliance can carry in treatment delivery. This could be particularly beneficial when alliance is flagged up as poor, and may enable clinicians to address relational barriers to positive treatment outcomes. As highlighted in a recent RCT, assessment of patient-reported alliance can be integrated into CBT delivery within research settings and even contribute to a positive stabilisation of the patient-therapist relationship (Flückiger et al., 2011).

As the present results indicate, alliance has a significant link to treatment success and may account for approximately 8-12% of variability in adolescent outcomes. Even though a causal relationship has not been confirmed so far, alliance is a widely cited component of psychological therapy and may be an active agent in therapeutic change (Shirk & Saiz, 1992). Clinicians should therefore not only be aware of relational processes, but also be able to respond to variations in alliance quality appropriately. As numerous studies indicate, this is applicable to all professionals working in a therapeutic capacity and requires skills in establishing, maintaining, and repairing alliances (Horvath & Marx, 1991). Since adolescents constitute a particularly complex patient group, this may call for specialised professional training in relational skills. Alliance-fostering training programmes, for instance, have already been trialled in adult mental health services and yielded promising findings for a diverse group of clinicians (Crits-Christoph et al., 2006).

Despite the recommendations highlighted above, it remains unclear how regular assessment of alliance can be integrated into routine mental health practice, particularly when ratings are requested from vulnerable patient groups such as looked-after adolescents. Being asked to rate alliance may well be anxiety-provoking for young service users and have an adverse effect on other therapeutically relevant factors, e.g. patient self-disclosure. There is also very little clarity regarding the type of measure most suitable for alliance assessment in relevant service settings. To date, most measures used to assess adolescent alliance are top-down adaptations of existing adult tools and echo their respective originals quite closely. Measures specifically developed for adolescent patients, e.g. the Therapeutic Alliance Scale for Children/Adolescents (Shirk & Saiz, 1992), seem to be comparatively rare.

4.5 Further Research Directions

The strength of the alliance-outcome relationship can vary in the presence of certain variables. In the most extreme of cases, alliance may simply be a passive treatment feature and bear no relation to clinical outcome. While the present meta-analysis questions the impact of therapist variability on the alliance-outcome association, it is clear that more primary research is necessary to assess the role of other therapist variables. The present meta-analysis also highlighted a relative lack of primary studies with predominantly adolescent participant samples. Despite comparatively stringent selection criteria with regard to mean age, almost half of included studies had participant samples breaching the 12-19 age range. The present target population was therefore not exclusively adolescent and encompassed a noteworthy number of children under the age of 12. It is therefore advisable to repeat all present analyses once a more age-homogeneous sample of studies can be identified.

5. Conclusion

In contrast to previous research with adult samples, patient-therapist ratio did not emerge as a significant moderator of the alliance-outcome relationship for adolescents undergoing mental health treatment. This unexpected finding challenges the hypothesis that therapist variability has an important influence on the alliance-outcome relationship. However, present effect size estimates for adolescent patients are larger than previously reported and suggest that alliance may account for 8-12% of variability in treatment outcomes across assessed domains. Despite a number of methodological limitations, these findings indicate that alliance is a notably more reliable predictor of adolescent treatment outcomes than previously estimated.

Acknowledgements

We wish to thank all authors who took the time to assist us with additional queries. In particular, we want to give thanks to Nicholas Allen, Karine Bertrand, Poornima Bhola, Stéphanie Bourion-Bédès, Brian Chu, Carola Cropp, Gary Diamond, Guy Diamond, Krister Fjermestad, Sharon Flicker, Gary Geffken, Eva Gilboa-Schechtman, Mark Godley, Michele Guzmán, Aaron Hogue, Lee Johnson, Jeffrey Kahn, Marc Karver, Noah Kaufman, Claudia Kinnen, Annie Lemieux, Silje Ormhaug, Julie Owens, Joan Simon, Stephen Shirk, and Susan Spence.

Correspondence

Correspondence should be addressed to Regina Murphy, Kirkwall Health Centre, New Scapa Road, Kirkwall, Orkney, KW15 1BH, (+44) 01856 888 280

regina.murphy.email@gmail.com

Key Practitioner Messages

- The alliance-outcome relationship for adolescent mental health patients is markedly larger than previously assumed, with average effect sizes presently ranging from $r = 0.29$ to $r = 0.34$.
- Alliance as perceived by adolescents accounts for 8-12% of variability in treatment outcomes; these are specifically linked to domains of symptoms severity, global functioning, and emotional distress.
- These findings exceed estimates previously reported for adult mental health patients and highlight the prognostic value of patient-reported alliance in psychological treatment for adolescents.
- Monitoring patient perceptions of alliance can provide valuable information in routine mental health practice and enable practitioners to address early relational barriers to positive treatment outcomes.
- Patient-therapist ratio did not moderate the strength of the alliance-outcome relationship; the idea that therapist variability affects the alliance-outcome relationship can therefore not be supported.

Areas for Future Research

- Conduct more primary research on the alliance-outcome relationship in adolescent patient groups.
- Repeat meta-analysis on adolescent alliance-outcome link with a more age-homogeneous sample.
- Identify moderators of adolescent alliance-outcome relationship linked to treatment variables.
- Explore how regular alliance assessment can be implemented in routine mental health practice.
- Assess suitability of alliance measures for use in Child and Adolescent Mental Health Services.

References

Ackerman, S. J., & Hilsenroth, M. J. (2003). A review of the therapist characteristics and techniques positively impacting the therapeutic alliance. *Clinical Psychology Review, 23*(1), 1–33.

*Anderson, R.E.E., Spence, S.H., Donovan, C.L., March, S., Prosser, S., & Kenardy, J. (2012). Working Alliance in Online Cognitive Behavior Therapy for Anxiety Disorders in Youth: Comparison With Clinic Delivery and its Role in Predicting Outcome. *Journal of Medical Internet Research, 14*(3), e88.

*Auerbach, S.M., May, J.C., Stevens, M., & Kiesler, D.J. (2008). The interactive role of working alliance and counselor-client interpersonal behaviors in adolescent substance abuse treatment. *International Journal of Clinical and Health Psychology, 8*(3), 617-629.

*Bertrand, K., Brunelle, N., Richer, I., Beaudoin, I., Lemieux, A., & Menard, J.M. (2013). Assessing Covariates of Drug Use Trajectories Among Adolescents Admitted to a Drug Addiction Center: Mental Health Problems, Therapeutic Alliance, and Treatment Persistence. *Substance Use & Misuse, 48*, 117–128.

*Bhola, P., & Kapur, M. (2013). The Development and Role of the Therapeutic Alliance in Supportive Psychotherapy with Adolescents. *Psychological Studies, 58*(3), 207–215.

Bickman, L., Athay, M.M., Riemer, M., Lambert, E.W., Kelley, S.D., Breda, C., & Andrade, A.R. (Eds.). (2010). *Manual of the Peabody Treatment Progress Battery* (2nd ed.). Nashville, TN: Vanderbilt University.

*Bickman, L., Vides de Andrade, A.R., Athay, M.M., Chen, J.I., De Nadai, A.S., Jordan-Arthur, B.L., & Karver, M.S. (2012). The Relationship Between Change in Therapeutic Alliance Ratings and Improvement in Youth Symptom Severity: Whose Ratings Matter the Most? *Administration and Policy in Mental Health, 39*, 78–89.

Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, Research & Practice, 16*, 252–260.

*Bourion-Bedes, S., Baumann, C., Kermarrec, S., Ligier, F., Feillet, F., Bonnemains, C., Guillemin, F., & Kabuth, B. (2013). Prognostic Value of Early Therapeutic Alliance in

Weight Recovery: A Prospective Cohort of 108 Adolescents With Anorexia Nervosa. *Journal of Adolescent Health, 52*, 344–350.

Buck, S.F. (1960). A method of estimation of missing values in multivariate data suitable for use with an electronic computer. *Journal of the Royal Statistical Society, Series B, 22,* 302-303.

*Cenerelli, A. (2013). Working Alliance with Adolescents who Receive Mandated School-based Counseling Services. Unpublished doctoral dissertation, City University of New York, New York, New York.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

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 preliminary study of alliance-fostering psychotherapy. *Psychotherapy Research, 16,* 268–281.

*Cropp, C., Streeck-Fischer, A., Jaeger, U., Masuhr, O., Schroder, A., & Leichsenring, F. (2008). Der Zusammenhang zwischen Behandlungserleben und Behandlungserfolg bei der stationaren Psychotherapie mit Kindern und Jugendlichen. *Zeitschrift fur Kinder- und Jugendpsychiatrie und Psychotherapie, 36*(3), 205-213.

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*, 642–649.

*Diamond, G.S., Liddle, H.A., Wintersteen, M.B., Dennis, M.L., Godley, S.H., Tims, F. (2006). Early Therapeutic Alliance as a Predictor of Treatment Outcome for Adolescent Cannabis Users in Outpatient Treatment. *The American Journal on Addictions, 15*, 26–33.

*Darchuk, A. J. (2007). The role of the therapeutic alliance and its relationship to treatment outcome and client motivation in an adolescent substance abuse treatment setting. Unpublished doctoral dissertation, Ohio University, Athens, Ohio.

Duval, S., & Tweedie, R. (2000). Trim and fill: a simple funnel plot based method of adjusting for publication bias in meta-analysis. *Biometrics, 56*, 455-463.

Effective Public Health Practice Project (1998). *Quality Assessment Tool For Quantitative Studies*. Hamilton, ON: Effective Public Health Practice Project.

Egger, M., Smith, G.D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315,* 629-634.

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. *The British Journal of Psychiatry, 180*(1) 51-60.

Falkenström, F., Granström, F., & Holmqvist, R. (2013). Therapeutic alliance predicts symptomatic improvement session by session. *Journal of Counseling Psychology, 60*(3), 317–328.

Florsheim, P., Shotorbani, S., Guest-Warnick, G., Barratt, T., & Hwang, W. (2000). Role of the working alliance in the treatment of delinquent boys in community-based programs. *Journal of Clinical Child Psychology, 29*, 94–107.

Flückiger, C., Del Re, A., Wampold, B.E., Znoj, H., Caspar, F., & Jörg, U. (2011). Valuing clients' perspective and the effects on the therapeutic alliance: A randomized controlled study of an adjunctive instruction. *Journal of Counseling Psychology, 59*(1), 18–26.

Freud, S. (1958). On the beginning of treatment: Further recommendations on the technique of psychoanalysis. In J. Starchey (Ed. & Trans.), *The standard edition of the complete psychological works of Sigmund Freud* (vol. 12, pp. 122-144). London: Hogarth Press.

Furnham, A. (1986). Response bias, social desirability and dissimulation. *Personality and Individual Differences, 7*, 385-400.

*Handwerk, M.L., Huefner, J.C., Ringle, J.L., Howard, B.K., Soper, S.H., Almquist, J.K., Chmelka, M. B., & Father Flanagan's Boys' Home (2008). The Role of Therapeutic Alliance in Therapy Outcomes for Youth in Residential Care. *Residential Treatment for Children & Youth, 25*(2), 145-165.

Hawley, K.M., & Weisz, J.R. (2005). Youth Versus Parent Working Alliance in Usual Clinical Care: Distinctive Associations With Retention, Satisfaction, and Treatment Outcome. *Journal of Clinical Child and Adolescent Psychology, 34*(1), 117–128.

Hedges, L. V., & Olkin, I. (1985). *Statistical Methods for Meta-Analysis*. New York, NY: Academic Press.

Heinonen, E., Lindfors, O., Härkänen, T., Virtala, E., Jääskeläinen, T., & Knekt, P. (2014). Therapists' Professional and Personal Characteristics as Predictors of Working Alliance in Short-Term and Long-Term Psychotherapies. *Clinical Psychology and Psychotherapy, 21*, 475-494.

Horvath, A., Del Re, A. C., Flückiger, C., & Symonds, D. (2011). The alliance. In J. C. Norcross (Ed.), *Relationships that work* (pp. 25–69). New York, NY: Oxford University Press.

Horvath, A.O., & Greenberg, L.S. (1989). Development and validation of the Working Alliance Inventory. *Journal of Counseling Psychology, 36*, 223-233.

Horvath, A.O., & Marx, R.W. (1991). The development and decay of the working alliance during time-limited counseling. *Canadian Journal of Counselling, 24*, 240-259.

*Hurley, K.D., Van Ryzin, M.J., Lambert, M., & Stevens, A.L. (2015). Examining Change in Therapeutic Alliance to Predict Youth Mental Health Outcomes. *Journal of Emotional and Behavioral Disorders*, 1-11.

Hutton, P., & Taylor, P.J. (2014). Cognitive behavioural therapy for psychosis prevention: a systematic review and meta-analysis. Psychological Medicine, 44(3), 449-468.

*Johnson, L.N., Wright, D.W., & Ketring, S.A. (2002). The therapeutic alliance in home-based family therapy: is it predictive of outcome? *Journal of Marital and Family Therapy, 28*(1), 93-102.

Jung, E., Wiesjahn, M., Rief, W., & Lincoln, T.M. (2015). Perceived therapist genuineness predicts therapeutic alliance in cognitive behavioural therapy for psychosis. *British Journal of Clinical Psychology, 54*, 34-48.

*Karpenko, V. (2010). Clinically Significant Symptom Change in Adolescents Receiving Outpatient Community Mental Health Services: Does it Relate to Satisfaction, Perceived Change, Therapeutic Alliance, and Improvement in Presenting Problems? Unpublished doctoral dissertation, Ohio University, Athens, Ohio.

Karver, M.S., Handelsman, J.B., Fields, S., & Bickman, L. (2005). A Theoretical Model of Common Process Factors in Youth and Family Therapy. *Mental Health Services Research, 7*(1), 35-51.

*Keeley, M.L., Geffken, G.R., Ricketts, E., McNamara, J.P.H., Storch, E.A. (2015). The therapeutic alliance in the cognitive behavioral treatment of pediatric obsessive–compulsive disorder. *Journal of Anxiety Disorders, 25*, 855– 863.

Kendall, P.C. (1994). Treating anxiety disorders in children: results of a randomized clinical trial. *Journal of Consulting and Clinical Psychology, 62,* 100-110.

Kendall, P.C., Flannery-Schroeder, E., Panichelli-Mindel, S.M., Southam-Gerow, M., Henin, A., & Warman, M. (1997). Therapy for youths with anxiety disorders: a second randomized clinical trial. *Journal of Consulting and Clinical Psychology, 65*, 366-380.

*Kim, H. (2007). Client growth and alliance development in solution-focused brief family therapy. Unpublished doctoral dissertation, State University of New York at Buffalo, Buffalo, New York.

*Korchmaros, J.D., & Stevens, S.J. (2014). Examination of the Role of Therapeutic Alliance, Treatment Dose, and Treatment Completion in the Effectiveness of The Seven Challenges. *Child and Adolescent Social Work Journal, 31*, 1–24.

Langer, D.A., McLeod, B.D., & Weisz, J.R. (2011). Do Treatment Manuals Undermine Youth–Therapist Alliance in Community Clinical Practice? *Journal of Consulting and Clinical Psychology, 79*(4), 427-432.

Liberati, A., Altman, D.G., Tetzlaff, J., Mulrow, C., Gøtzsche, P.C., Ioannidis, J.P.A., … Moher, D. (2009). The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration. *PLoS Medicine, 6*(7): e1000100.

Leucht, S., Cipriani, A., Spineli, L., Mavridis, D., Orey, D., Richter, F., … Davis, M.J. (2013). Comparative efficacy and tolerability of 15 antipsychotic drugs in schizophrenia: a multiple-treatments meta-analysis. *The Lancet, 382*(9896), 951-962.

*Litter, M. (2004). Relationship-based psychotherapy with court-involved youth: the therapy relationship's effect on outcome. Unpublished doctoral dissertation, Wright Institute, Berkeley, California.

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

Martin, D. J., Garske, J. P., & Davis, M. K. (2000). Relation of the therapeutic alliance with outcome and other variables: a meta-analysis review. *Journal of Consulting and Clinical Psychology, 68*(3), 438–450.

McLeod, B. D. (2011). Relation of the alliance with outcomes in youth psychotherapy: A meta-analysis. *Clinical Psychology Review, 31*, 603–616.

*Ormhaug, S.M., Jensen, T.K., Wentzel-Larsen, T., & Shirk, S.R. (2014). The Therapeutic Alliance in Treatment of Traumatized Youths: Relation to Outcome in a Randomized Clinical Trial. *Journal of Consulting and Clinical Psychology, 82*(1), 52-64.

Ormhaug, S., Shirk, S., & Wentzel-Larsen, T. (2015). Therapist and client perspectives on the alliance in the treatment of traumatized adolescents. *European Journal of Psychotraumatology, 6*.

Pigott, T.D. (1994). Methods for Handling Missing Data in Research Synthesis. In H. E. Cooper, & L. V. Hedges (Eds.), *Handbook of Research Synthesis* (pp. 163−175). New York, NY: Russell Sage Foundation.

Quintana, D.S. (2015). From pre-registration to publication: a non-technical primer for conducting a meta-analysis to synthesize correlational data. *Frontiers in Psychology, 6*, 1549.

Raudenbush, S.W. (1994). Random Effects Models. In H. E. Cooper, & L. V. Hedges (Eds.), *Handbook of Research Synthesis* (pp. 301−321). New York, NY: Russell Sage Foundation.

Rosenthal, R. (1994). Parametric measures of effect size. In H. E. Cooper, & L. V. Hedges (Eds.), *Handbook of Research Synthesis* (pp. 231−244). New York, NY: Russell Sage Foundation.

*Sarlin, N. S. (1992). Working relationships in the treatment of adolescent inpatients: Early treatment predictors and associations with outcome. Unpublished doctoral dissertation, University of Denver, Denver, Colorado.

*Savicki, K.L. (2007). Working Alliance, Readiness for Change, and Theory of Change as Predictors of Treatment Success Among Incarcerated Adolescents. Unpublished doctoral dissertation, University of Texas at Austin, Texas.

Shea, B.J., Grimshaw, J.M., Wells, G.A., Boers, M., Andersson, N., Hamel, C., … Bouter, L.M. (2007). Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Medical Research Methodology, 15*(7), 10.

*Shelef, K., Diamond, G.M., Diamond, G.S., & Liddle, H.A. (2005). Adolescent and Parent Alliance and Treatment Outcome in Multidimensional Family Therapy. *Journal of Consulting and Clinical Psychology, 73*(4), 689-698.

*Shirk, S.R., Gudmundsen, G., Kaplinski, H.C., & McMakin, D.L. (2008). Alliance and Outcome in Cognitive-Behavioral Therapy for Adolescent Depression. *Journal of Clinical Child & Adolescent Psychology, 37*(3), 631-639.

Shirk, S. R., & Karver, M. (2003). Prediction of treatment outcome from relationship variables in child and adolescent therapy: A meta-analytic review. *Journal of Consulting and Clinical Psychology, 71*(3), 452−464.

Shirk, S. R., Karver, M. S., & Brown, R. (2011). The alliance in child and adolescent psychotherapy. *Psychotherapy, 48*(1), 17–24.

Shirk, S.R., & Saiz, C.C. (1992). Clinical, empirical, and developmental perspectives on the therapeutic relationship in child psychotherapy. *Development and Psychopathology, 4*(4), 713–28.

*Smith, R. D. (1999). Using object relations to predict outcome for adolescents in residential treatment. Unpublished doctoral dissertation, University of Denver, Denver, Colorado.

Stovell, D., Morrison, A. P., Panayiotou, M., Hutton, P. (2016). The effect of shared decision-making on empowerment-related outcomes in psychosis: a systematic review and meta-analysis. *British Journal of Psychiatry, 209*(1), 23-28.

*Sullivan, T.C. (2012). Validity of session rating scale scores with adolescent clients in a school setting. Unpublished doctoral dissertation, University of Central Arkansas, Conway, Arkansas.

*Tetzlaff, B.T, Kahn, J.H., Godley, S.H., Godley, M.D., Diamond, G.S., & Funk, R.R. (2005). Working Alliance, Treatment Satisfaction, and Patterns of Posttreatment Use Among Adolescent Substance Users. *Psychology of Addictive Behaviors, 19*(2), 199-207.

*Ventura, D. (2010). Outcome and Therapeutic Alliance in Solution-Focused Brief Family Therapy. Unpublished doctoral dissertation, State University of New York at Buffalo, New York.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), 1–48.

*Zack, S.E., Castonguay, L.G., Boswell, J.F., McAleavey, A.A., Adelman, R., & Kraus, D.R. (2015). Attachment History as a Moderator of the Alliance Outcome Relationship in Adolescents. *Psychotherapy, 52*(2), 258-267.

*Zaitsoff, S.L., Doyle, A.C., Hoste, R.R., & le Grange, D. (2008). How do adolescents with bulimia nervosa rate the acceptability and therapeutic relationship in family-based treatment? *International Journal of Eating Disorders, 41*, 390−398.

*Zandberg, L.J., Skriner, L.C., & Chu, B.C. (2015). Client-Therapist Alliance Discrepancies and Outcome in Cognitive-Behavioral Therapy for Youth Anxiety. *Journal of Clinical Psychology, 71*(4), 313–322.

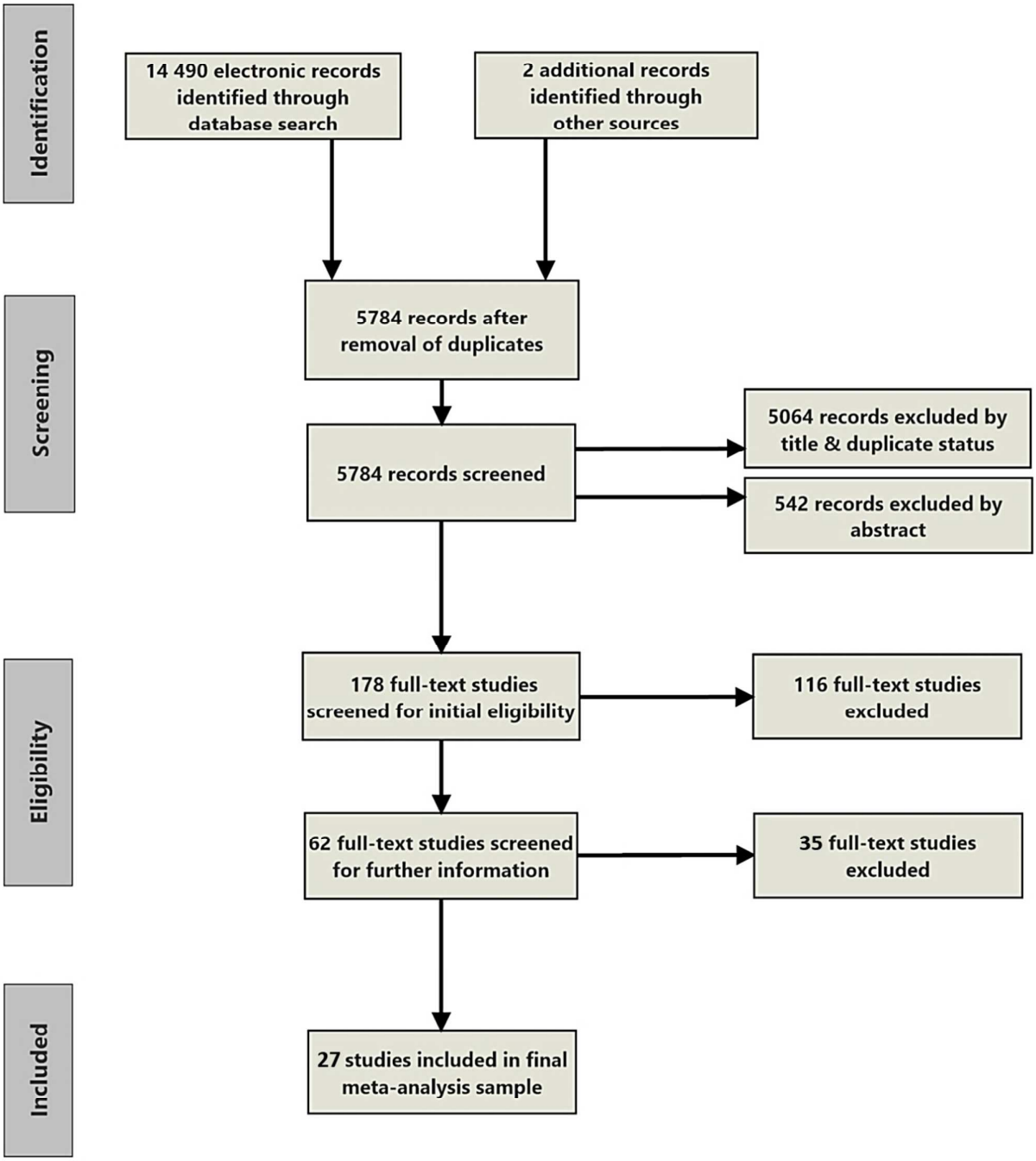

Figure 1. Prisma Flow Chart depicting literature screening process. Two additional records were identified through reference trails in German literature. When records were screened by title, further duplicates were excluded.

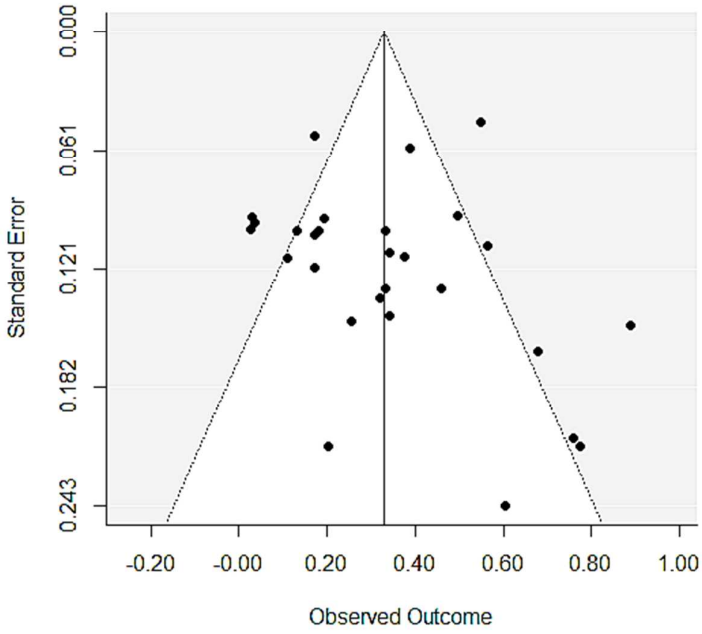

Figure 2. Funnel plot depicting 28 extracted effect sizes plotted against standard errors.

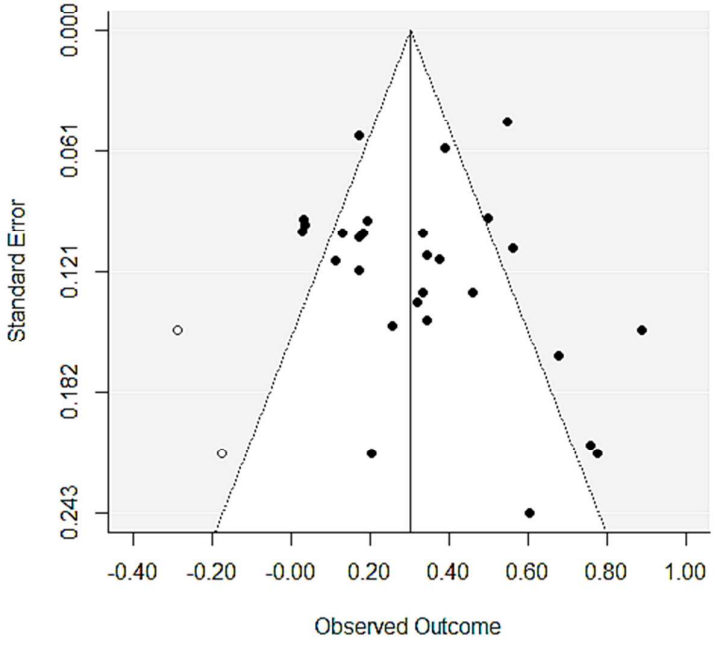

Figure 3. Funnel plot containing two additional effect sizes on the left-hand side.

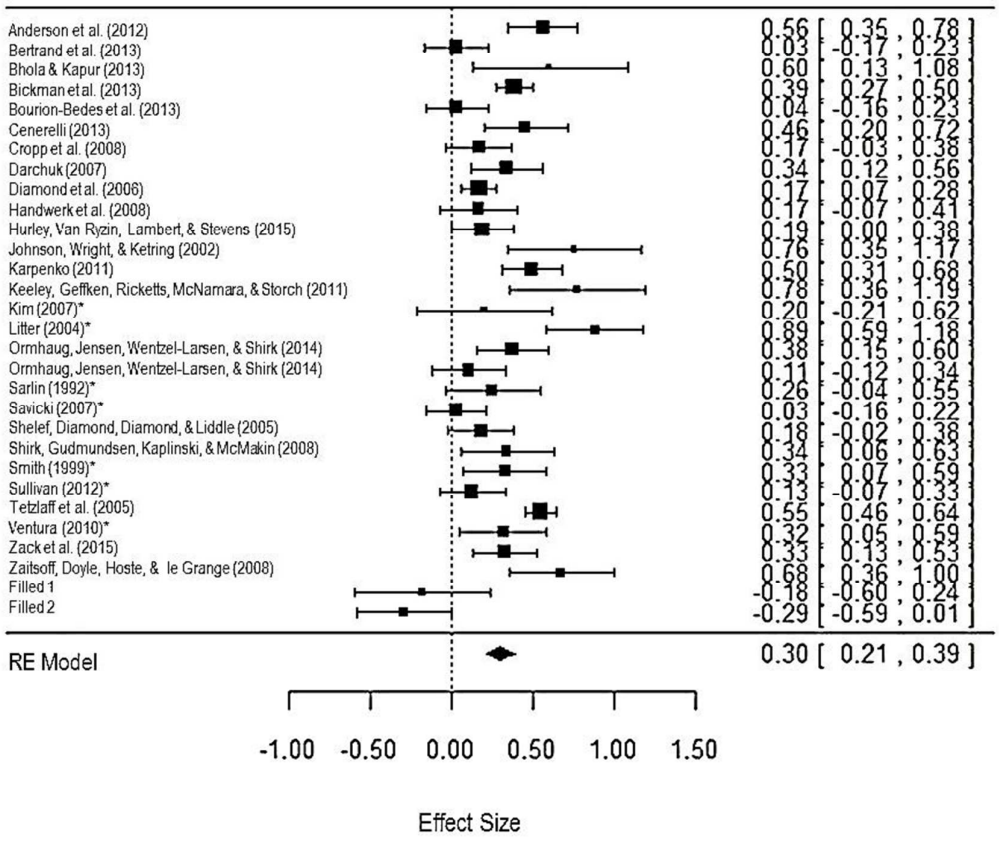

Figure 4. Forest plot depicting effect sizes in Fisher's *Z* with corresponding 95% CI.

Patient-Reported Therapeutic Alliance and Adolescent Outcomes

Table 1. Reviewed studies, sample characteristics, and other variables.

Study	Study Setting	N	Mean Age	Age Range	% female	Target Problem	Treatment Reception	CBT Elements	Manual Use	PTR	ES *r*	ES *Z*
Anderson et al. (2012)	Australia	85	13.61	12-18	56.50	Anxiety	individual	yes	yes	.	0.51	0.56
Bertrand et al. (2013)	Canada	102	15.53	14-18	40.20	Substance Abuse	mixed	yes	yes	102.00	0.03	0.03
Bhola & Kapur (2013)	India	20	13.88	13-16	100.00	Mixed	individual	no	no	20.00	0.54	0.60
Bickman et al. (2012)	USA	288	14.80	11-18	48.00	Mixed	individual	no	no	0.96	0.37	0.39
Bourion-Bedes et al. (2013)	France	108	15.30	9-20	94.40	Eating Disorder	mixed	yes	no	4.70	0.04	0.04
Cenerelli (2013)*	USA	60	.	12-20	31.70	Mixed	individual	no	no	2.70	0.43	0.46
Cropp et al. (2008)	Germany	95	16.33	11-21	54.70	Mixed	individual	yes	no	9.50	0.17	0.17
Darchuk (2007)*	USA	81	16.27	13-18	35.80	Substance Abuse	individual	no	no	13.50	0.33	0.34
Diamond et al. (2006)	USA	356	15.70	12-18	19.40	Substance Abuse	mixed	yes	yes	25.40	0.17	0.17
Handwerk et al. (2008)	USA	71	15.90	11-18	48.00	Behaviour	individual	yes	no	10.10	0.17	0.17
Hurley Van Ryzin, Lambert, & Stevens (2014)	USA	112	15.29	10-17	42.90	Behaviour	mixed	yes	no	2.50	0.19	0.19
Johnson, Wright, & Ketring (2002)	USA	26	14.00	11-18	.	Behaviour	systemic	yes	no	.	0.64	0.76
Karpenko (2010)*	USA	116	14.26	12-18	57.00	Mixed	mixed	yes	no	5.50	0.46	0.50
Keeley, Geffken, Ricketts, McNamara, & Storch (2011)	USA	25	13.16	7-17	44.00	OCD	individual	yes	yes	.	0.65	0.78
Kim (2007)*	USA	25	13.10	7-18	52.00	Behaviour	systemic	yes	no	5.10	0.20	0.20
Litter (2004)*	USA	47	16.50	12-17	51.10	Behaviour	mixed	no	no	6.70	0.71	0.89
Ormhaug, Jensen, Wentzel-Larsen, & Shirk (2014)	Norway	79	15.10	10-18	79.50	PTSD	mixed	yes	yes	3.00	0.36	0.38
Ormhaug, Jensen, Wentzel-Larsen, & Shirk (2014)	Norway	77	15.10	10-18	79.50	PTSD	mixed	no	no	1.70	0.11	0.11
Sarlin (1992)*	USA	48	14.11	11-18	62.50	Mixed	individual	yes	no	4.60	0.25	0.26
Savicki (2007)*	USA	114	18.40	.	15.80	Behaviour	mixed	yes	no	.	0.03	0.03
Shelef, Diamond, Diamond, & Liddle (2005)	USA	100	16.00	13-18	15.00	Substance Abuse	systemic	yes	yes	33.30	0.18	0.18
Shirk, Gudmundsen, Kaplinski, & McMakin (2008)	USA	50	15.80	14-18	72.00	Mood	individual	yes	yes	6.30	0.33	0.34
Smith (1999)*	USA	60	13.80	12-16	36.70	Substance Abuse	mixed	no	no	.	0.32	0.33
Sullivan (2012)*	USA	100	13.88	13-17	48.00	Mixed	individual	no	no	9.10	0.13	0.13
Tetzlaff et al. (2005)	USA	470	15.70	13-18	17.00	Substance Abuse	mixed	yes	yes	33.60	0.50	0.55
Ventura (2010)*	USA	56	14.86	13-17	42.90	Behaviour	systemic	yes	yes	.	0.31	0.32

Patient-Reported Therapeutic Alliance and Adolescent Outcomes

Study	Study Setting	N	Mean Age	Age Range	% female	Target Problem	Treatment Reception	CBT Elements	Manual Use	PTR	ES *r*	ES *Z*
Zack et al. (2015)	USA	100	17.39	11-25	32.00	Substance Abuse	mixed	yes	no	6.70	0.32	0.33
Zaitsoff, Doyle, Hoste, & le Grange (2008)	USA	40	16.10	12-19	97.50	Eating Disorder	individual	no	yes	.	0.59	0.68

Note. Studies marked with an asterisk (*) are unpublished doctoral dissertations; [a] is a validated French instrument adapted from the Addiction Severity Index (McLellan, Luborsky, O'Brien, & Woody, 1980); WAI=Working Alliance Inventory; CALPAS=California Psychotherapy Alliance Scales; TAQS=Therapeutic Alliance Quality Scale; PHAQ=Penn Helping Alliance Questionnaire; SEB=Stationserfahrungsbogen, German inpatient measure; WRS=Working Relationship Scale; FTAS=Family Therapy Alliance Scale; TASC/A=Therapeutic Alliance Scale for Children/Adolescents; RRS=Relationship Rating Scale; RCA=Relationship with Counselor Assessment; VPPS=Vanderbilt Psychotherapy Process Scales; SRS=Session Rating Scale; HRQ=Helping Relationship Questionnaire.

Patient-Reported Therapeutic Alliance and Adolescent Outcomes

Table 2. Reviewed studies and measure variables.

Study	Alliance Measure	Outcome Measure	Alliance Timing	Outcome Timing	Outcome Domain
Anderson et al. (2012)	WAI	Children's Global Assessment Scale	early-treatment	6-month follow-up	GF
Bertrand et al. (2013)	CALPAS	IGT-ADO [a]	early-treatment	end-of-treatment	SS
Bhola & Kapur (2013)	WAI	Youth Self Report [b]	mid-treatment	end-of-treatment	SS
Bickman et al. (2012)	TAQS	Symptoms & Functioning Severity Scale [b]	early-treatment	end-of-treatment	SS
Bourion-Bedes et al. (2013)	PHAQ	BMI	early-treatment	end-of-treatment	SS
Cenerelli (2013)*	WAI	Counselling Outcome Measure	mixed	mixed	GF
Cropp et al. (2008)	SEB	Symptom Checklist 90-R [b]	end-of--treatment	end-of-treatment	SS
Darchuk (2007)*	WAI	The Ohio Scales [b]	early-treatment	end-of-treatment	GF; SS
Diamond et al. (2006)	WAI	Global Appraisal of Individual Needs	early-treatment	3-month follow-up	SS
Handwerk et al. (2008)	WRS	Symptom Screener [b]	mixed	end-of-treatment	SS
Hurley et al. (2014)	TAQS	Child Behaviour Checklist	early-treatment	end-of-treatment	SS
Johnson et al. (2002)	FTAS	Outcome Questionnaire OQ-45.2 [b]	end-of-treatment	end-of-treatment	ED
Karpenko (2010)*	WAI	The Ohio Scales [b]	end-of-treatment	late-treatment	SS
Keeley et al. (2011)	TASC/A	Children's Yale-Brown Obsessive-Compulsive Scale	mid-treatment	end-of-treatment	SS
Kim (2007)*	RRS	Progress Rating Scale [b]	early-treatment	end-of-treatment	GF
Litter (2004)*	VPPS	Global Appraisal of Individual Needs	end-of-treatment	end-of-treatment	SS
Ormhaug et al. (2014)	TASC/A	Child PTSD Symptom Scale [b]	mid-treatment	end-of-treatment	GF; SS
Ormhaug et al. (2014)	TASC/A	Child PTSD Symptom Scale [b]	mid-treatment	end-of-treatment	GF; SS
Sarlin (1992)*	PHAQ	Children's Global Assessment Scale	early-treatment	end-of-treatment	GF
Savicki (2007)*	WAI	Count of Rule Violations	mixed	4-month follow-up	SS
Shelef et al. (2005)	WAI	Global Appraisal of Individual Needs	early-treatment	end-of-treatment	SS
Shirk et al. (2008)	TASC/A	Computerized Diagnostic Interview Scale for Children	early-treatment	end-of-treatment	SS
Smith (1999)*	PHAQ	Youth Self Report [b]	end-of-treatment	end-of-treatment	SS
Sullivan (2012)*	TAQS	Outcome Rating Scale [b]	early-treatment	early-treatment	ED
Tetzlaff et al. (2005)	WAI	Global Appraisal of Individual Needs	mixed	3-month follow-up	SS
Ventura (2010)*	SRS	Progress Rating Scale [b]	early-treatment	end-of-treatment	GF

Patient-Reported Therapeutic Alliance and Adolescent Outcomes

Study	Alliance Measure	Outcome Measure	Alliance Timing	Outcome Timing	Outcome Domain
Zack et al. (2015)	WAI	Treatment Outcome Package [b]	mixed	end-of-treatment	SS
Zaitsoff et al. (2008)	HRQ	Eating Disorder Examination	mid-treatment	end-of-treatment	SS

Note. Studies marked with an asterisk (*) are unpublished doctoral dissertations; [a] is a validated French instrument adapted from the Addiction Severity Index (McLellan, Luborsky, O'Brien, & Woody, 1980); [b] marks all outcome measures relying on patient report; WAI=Working Alliance Inventory; CALPAS=California Psychotherapy Alliance Scales; TAQS=Therapeutic Alliance Quality Scale; PHAQ=Penn Helping Alliance Questionnaire; SEB=Stationserfahrungsbogen, German inpatient measure; WRS=Working Relationship Scale; FTAS=Family Therapy Alliance Scale; TASC/A=Therapeutic Alliance Scale for Children/Adolescents; RRS=Relationship Rating Scale; RCA=Relationship with Counselor Assessment; VPPS=Vanderbilt Psychotherapy Process Scales; SRS=Session Rating Scale; HRQ=Helping Relationship Questionnaire.

Table 3. Quality ratings of reviewed studies, conducted using the EPHPP Quality Assessment Tool for Quantitative Studies

Study	Selection Bias	Study Design	Blinding	Data Collection [c] Method	Withdrawals & Dropouts	Global Rating
Anderson et al. (2012)	moderate	strong	moderate	strong	weak	moderate
Bertrand et al. (2013)	moderate	moderate [a]	moderate	strong	moderate	strong
Bhola & Kapur (2013)	strong	moderate	moderate	strong	strong	strong
Bickman et al. (2012)	moderate	moderate [a]	moderate	strong	na	strong
Bourion-Bedes et al. (2013)	moderate	moderate [a]	moderate	strong	weak	moderate
Cenerelli (2013)*	weak	moderate [a]	moderate [b]	strong	strong	moderate
Cropp et al. (2008)	moderate	moderate [a]	moderate	moderate	na	strong
Darchuk (2007)*	moderate	moderate [a]	moderate	strong	na	strong
Diamond et al. (2006)	strong	strong [a]	strong	strong	na	strong
Handwerk et al. (2008)	moderate	moderate [a]	moderate	strong	na	strong
Hurley et al. (2014)	moderate	moderate [a]	moderate	strong	na	strong
Johnson et al. (2002)	moderate	moderate [a]	moderate	strong	moderate	strong
Karpenko (2010)*	moderate	moderate [a]	moderate	strong	weak	moderate
Keeley et al. (2011)	moderate	moderate [a]	moderate [b]	strong	strong	strong
Kim (2007)*	moderate	weak [a]	moderate	strong	strong	moderate
Litter (2004)*	weak	moderate [a]	moderate [b]	strong	moderate	moderate
Ormhaug et al. (2014)	moderate	strong	strong [b]	strong	moderate	strong
Sarlin (1992)*	weak	moderate [a]	moderate	strong	moderate	moderate
Savicki (2007)*	moderate	moderate [a]	moderate	moderate	strong	strong
Shelef et al. (2005)	strong	strong	strong	strong	na	strong
Shirk et al. (2008)	moderate	moderate [a]	moderate [b]	strong	strong	strong
Smith (1999)*	moderate	moderate [a]	moderate	strong	weak	moderate
Sullivan (2012)*	weak	moderate [a]	moderate	moderate	strong	moderate
Tetzlaff et al. (2005)	strong	strong	strong	strong	na	strong

Patient-Reported Therapeutic Alliance and Adolescent Outcomes

Study	Selection Bias	Study Design	Blinding	Data Collection [c] Method	Withdrawals & Dropouts	Global Rating
Ventura (2010)*	moderate	weak [a]	moderate	moderate	moderate	moderate
Zack et al. (2015)	moderate	moderate [a]	moderate	moderate	moderate	strong
Zaitsoff et al. (2008)	strong	strong	moderate	moderate	na	strong

Note. Studies marked with an asterisk (*) are unpublished doctoral dissertations; studies marked [a] used one participant sample only and did not make group comparisons; studies marked [b] took measures to reduce reporting bias; in the subdomain marked [c] ratings focused on the validity and reliability of measures used to quantify the alliance-outcome relationship and do therefore not represent assessments of all data collection measures; studies marked *na* conducted secondary data analysis, quality ratings of participant withdrawals and dropouts were thus not applicable.

Manuscript JCPP-RR-2016-00431

Narrative Clinical Commentary

While previous reviews questioned the role of alliance in non-adult populations, the present meta-analysis highlights its clinical relevance for adolescents undergoing mental health treatment. Contrary to prior estimates, alliance appears to account for 8-12% of variability in treatment outcomes and may thus be a more reliable predictor of clinical progress than previously assumed. This meta-analysis is also the first to identify concrete links between adolescent alliance perceptions and treatment outcomes in three commonly measured domains, i.e. symptom severity, global functioning, and emotional distress. These findings do not only underscore the importance of the patient perspective, but also point towards potential benefits of monitoring adolescent alliance in routine clinical practice.

Table 4. List of 151 excluded studies with corresponding exclusion reasons.

Study	Reason for Exclusion	Study	Reason for Exclusion
Accurso, Hawley, & Garland (2013)	mean age= 9	Delsignore et al. (2014)	adult sample
Altmann et al. (2014)	adult sample	Dolemeyer, Klinitzke, Steining, Wagner, & Kersting (2013)	adult sample
Andersson et al. (2012)	adult sample	Ellison et al. (2012)	alliance=parent-rated
Auerbach, May, Stevens, & Kiesler (2008)	effect size not obtainable	Eltz, Shirk, & Sarlin (1995)	effect size not obtainable
Ayotte, Lanctot, & Tourigny (2015)	outcome not measured	Elvins (2012)	alliance=observer-rated
Bambling, King, Raue, Schweitzer, & Lambert (2006)	adult sample	Fals-Stewart & Lam (2010)	adult sample
Barber et al. (2014)	adult sample	Faw, Hogue, Johnson, Diamond, & Liddle (2005)	alliance=observer-rated
Barrowclough, Meier, Beardmore, & Emsley (2010)	adult sample	Fitch (2012)	outcome=attrition
Barrowclough et al. (2014)	adult sample	Fjermestad, Wergeland, Havik, & Heiervang (2012)	alliance=observer-rated
Becker, Becker, & Ginsburg (2012)	mean age=11	Flicker (2004)	alliance=observer-rated
Becker et al. (2014)	alliance not measured	Flicker, Turner, Waldron, Brody, & Ozechowski (2008)	alliance=observer-rated
Bedics, Atkins, Comtois, & Linehan (2012)	adult sample	Flückiger et al. (2012)	adult sample
Bedics, Atkins, Harned, & Linehan (2015)	adult sample	Foa, McLean, Capaldi, & Rosenfield (2013)	alliance not measured
Bhati (2014)	adult sample	Forsberg (2011)	alliance=observer-rated
Bovard-Johns, Yoder, & Burton (2015)	outcome not measured	Forsberg et al. (2013)	alliance=observer-rated
Brockmann et al. (2011)	adult sample	Forsberg et al. (2014)	alliance=observer-rated
Cavell, Elledge, Malcolm, Faith, & Hughes (2009)	mean age=8.19	Friedlander, Kivlighan, & Shaffer (2012)	effect size not obtainable
Champion (1998)	mean age=8.50	Gatta et al. (2010)	alliance=observer-rated
Chiu, McLeod, Har, & Wood (2009)	mean age= 9.74	Geffken, Storch, Keeley, & Ricketts (2010)	data published elsewhere
Chu, Skriner, & Zandberg (2013)	effect size not obtainable	Gilbert-Eliot (2013)	effect size not obtainable
Chu, Skriner, & Zandberg (2014)	outcome not measured	Gilboa-Schechtman et al. (2010)	effect size not obtainable
Church, Pina, Reategui, & Brooks (2012)	alliance not measured	Ginsburg, Becker, Drazdowski, & Tein (2012)	mean age=11.12
Cordaro, Tubman, Wagner, & Morris (2012)	outcome=treatment completion	Granic et al. (2012)	alliance=parent-rated
Creed (2007)	mean age=11.20	Green (2001)	mean age= 11.4
Crits-Christoph et al. (2010)	adult sample	Green, Drake, Brunette, & Noordsy (2007)	alliance=therapist-rated
Cummings et al. (2013)	mean age=10.72	Grosse Holtforth et al. (2014)	adult sample
Dakof et al. (2015)	alliance not measured	Hanley (2009)	outcome not measured
Daly, Llewelyn, McDougall, & Chanen (2010)	alliance not measured	Harvey (2008)	mean age=7.95
Daniels & Wearden (2011)	adult sample	Hawley & Weisz (2005)	mean age=11.90
de Haan, Boon, de Jong, Geluk, & Vermeiren (2014)	outcome=treatment dropout	Hauser (2014)	adult sample
de la Pena, Friedlander, Escudero, & Heatherington (2012)	outcome not measured	Henry (2014)	outcome=parenting practices

Study	Reason for Exclusion	Study	Reason for Exclusion
Hentschel, Bijleveld, & Rudolf (1997)	adult sample	Leibert, Smith, & Agaskar (2011)	adult sample
Hersoug, Hoglend, Gabbard, & Lorentzen (2013)	adult sample	Lester (2012)	effect size not obtainable
Hildebrandt, Loeb, Troupe, & Delinsky (2012)	adult sample	Leuchter, Hunter, Tartter, & Cook (2014)	adult sample
Hintikka, Laukkanen, Marttunen, & Lehtonen (2006)	effect size not obtainable	Levin, Henderson, & Ehrenreich-May (2012)	outcome not measured
Hogue, Dauber, Stambaugh, Cecero, & Liddle (2006)	alliance=observer-rated	Levy-Frank, Hasson –Ohayon, Kravetz, & Roe (2011)	adult sample; patient=parents
Holmqvist, Hill, & Lang (2007)	effect size not obtainable	Liber et al. (2010)	mean age=10.22
Huppert et al. (2014)	adult sample	Lindner et al. (2014)	adult sample
Imel et al. (2011)	alliance not measured	LoTempio (2011)	alliance=observer-rated
Isserlin & Couturier (2012)	alliance=observer-rated	LoTempio et al. (2013)	alliance=observer-rated
Johansen, Lumley, & Cano (2011)	adult sample	Lutz et al. (2013)	adult sample
Johnson, Penn, Bauer, Meyer, & Evans (2008)	adult sample	Marcus, Kashy, Wintersteen, & Diamond (2011)	effect size not obtainable
Johnson, Ketring, & Anderson (2013)	outcome not measured	Marker, Comer, Abramova, & Kendall (2013)	mean age=10.19
Jung, Wiesjahn, & Lincoln (2014)	adult sample	McLaughlin, Locke, Aupont, Davis, & Doerffler (2010)	outcome not measured
Kay-Lambkin, Baker, Lewin, & Carr (2011)	adult sample	McLeod & Weisz (2005)	mean age=10.30
Karver et al. (2008)	effect size not obtainable	McLeod et al. (2014)	mean age=9.58
Kaufman, Rohde, Seeley, Clarke, & Stice (2005)	effect size not obtainable	Nirenberg, Baird, Longabaugh, & Mello (2012)	alliance not measured
Kazdin, Marciano, & Whitley (2005)	Mean age=7.20	Nissen-Lie, Havik, Hoglend, Monsen, & Ronnestad (2013)	adult sample
Kazdin, Whitley, & Marciano (2006a)	mean age=9.60	Patterson, Anderson, & Wei (2014)	adult sample
Kazdin & Whitley (2006b)	mean age=7.00	Pereira, Lock, & Oggins (2006)	alliance=observer-rated
Kendall (1994)	mean age=11.00	Pereira (2005)	published elsewhere
Kendall et al. (1997)	mean age=11.00	Pestle (2012)	mean age=10.70
Khanna (2010)	mean age=10.10	Podell et al. (2013)	mean age=10.76
King, Brooner, Peirce, Kolodner, & Kidorf (2014)	adult sample	Raykos et al. (2014)	adult sample
Kinnen & Dopfner (2013)	mean age=9.79	Reininghaus et al. (2013)	adult sample
Köster (2010)	alliance not measured	Reyes (2013)	effect size not obtainable
Korchmaros & Stevens (2014)	effect size not obtainable	Robbins, Turner, Alexander, & Perez (2003)	outcome=dropout
Kramer et al. (2011)	adult sample	Robbins et al. (2006)	outcome=dropout
Kramer et al. (2014)	adult sample	Robbins et al. (2008)	alliance=observer-rated
Lange & Stanton (2011)	outcome not measured	Robbins et al. (2011)	alliance not measured
Langer, McLeod, & Weisz (2011)	mean age=11.27	Rogers, Lubman, & Allen (2008)	effect size not obtainable

Study	Reason for Exclusion	Study	Reason for Exclusion
Ryan, Safran, Doran, & Muran (2012)	adult sample	Swart & Apsche (2014)	alliance not measured
Salzer, Cropp, Jaeger, Masuhr, & Streeck-Fischer (2014)	alliance not measured	ter Huurne, Postel, de Haan, & DeJong (2013)	adult sample
Santisteban et al. (2015)	effect size not obtainable	van Orman (1996)	study not available
Savage (2011)	mean age=11.35	Wagner, Hospital, Graziano, Morris, & Gil (2014)	alliance not measured
Sheehan (2013)	alliance=observer-rated	Weck, Richtberg, Jakob, Neng, & Hofling (2015)	adult sample
Shelef & Diamond (2008)	alliance=observer-rated	Weidel (2013)	outcome=retention
Simpson, Frick, Kahn, & Evans (2013)	outcome not measured	Weitkamp et al. (2011)	outcome not measured
Slesnick, Erdem, Bartle-Haring, & Brigham (2013)	alliance not measured	Weitkamp et al. (2012)	outcome not measured
Sly, Morgan, Mountford, & Lacey (2013)	adult sample	Weitkamp et al. (2014)	alliance not measured
Spirito et al. (2011)	outcome not measured	Westra, Constantino, Arkowitz, & Dozois (2011)	adult sample
Stice, Rohde, Gau, & Wade (2010a)	alliance not measured	Winnett (2008)	mean age=10.83
Stice, Rohde, Seeley, & Gau (2010b)	alliance not measured	Wolfe, Kay-Lambkin, Bowman, & Childs (2013)	adult sample
Stiles-Shields, Kwasny, Cai, & Mohr (2014)	adult sample	Zandberg, Skriner, & Chu (2015)	effect size not obtainable
Stimmel (2013)	adult sample	Zorzella, Muller, & Cribbie (2015)	mean age=9.58
Strunk, Cooper, Ryan, DeRubeis, & Hollon (2012)	adult sample		