Received: 25 October 2016 | Revised: 12 December 2016 | Accepted: 12 December 2016

DOI: 10.1002/eat.22672

REVIEW

International Journal of EATING DISORDERS

A meta-analysis of the relation between therapeutic alliance and treatment outcome in eating disorders

Tiffany A. Graves, MA[1] | Nassim Tabri, PhD[1] | Heather Thompson-Brenner, PhD[1] | Debra L. Franko, PhD[1] | Kamryn T. Eddy, PhD[1] | Stephanie Bourion-Bedes, MD[2] | Amy Brown, PhD[3] | Michael J. Constantino, PhD[4] | Christoph Flückiger, PhD[5] | Sarah Forsberg, PsyD[6] | Leanna Isserlin, MD[7] | Jennifer Couturier, MD[8] | Gunilla Paulson Karlsson, MD[9] | Johannes Mander, PhD[10] | Martin Teufel, MD[11] | James E. Mitchell, MD[12] | Ross D. Crosby, PhD[12] | Claudia Prestano, PhD[13] | Dana A. Satir, PhD[14] | Susan Simpson, DClinPsych[15] | Richard Sly, PhD[16] | J. Hubert Lacey, MD[17] | Colleen Stiles-Shields[18] | Giorgio A. Tasca, PhD[19] | Glenn Waller, DPhil[20] | Shannon L. Zaitsoff, PhD[21] | Renee Rienecke, PhD[22] | Daniel Le Grange, PhD[23] | Jennifer J. Thomas, PhD[1]

[1]Eating Disorders Clinical and Research Program (EDCRP), Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, United States

[2]Centre Hospitalier Lorquin, Lorquin, France

[3]Eating Disorder Service, South London and Maudsley NHS Foundation Trust, London, England, United Kingdom

[4]Department of Psychological and Brain Sciences, University of Massachusetts-Amherst, Amherst, Massachusetts, United States

[5]Department of Psychology, University of Wisconsin-Madison, Madison, Wisconsin, United States and Institute fur Psychologie, University of Bern, Bern, Switzerland

[6]Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Palo Alto, California, United States

[7]Division of Child & Adolescent Psychiatry, University of Ottawa, Ottawa, Ontario, Canada

[8]Pediatric Eating Disorders Program, McMaster Children's Hospital, Hamilton, Ontario, Canada

[9]Anorexia-Bulimia Eating Disorder Unit, The Sahlgrenska University Hospital, Gothenburg, Sweden

[10]ZZP Center for Psychological Psychotherapy, University of Heidelberg, Heidelberg, Germany

[11]Department of Psychosomatic Medicine and Psychotherapy, University Hospital Tuebingen, Tuebingen, Germany and Clinic for Psychosomatic Medicine and Psychotherapy, University of Duisburg-Essen, Essen, Germany

[12]Neuropsychiatric Research Institute, Fargo, North Dakota, United States and University of North Dakota School of Medicine and Health Sciences, Grand Forks, North Dakota, United States

[13]Niccolò Cusano University, Roma, Italy

[14]Department of Psychiatry, University of Colorado Denver, Denver, Colorado, United States

[15]School of Psychology, Social Work, and Social Policy, University of South Australia, South Australia, Australia

[16]School of Nursing Science, University of East Anglia, Norwich, Norfolk, United Kingdom

[17]Department of Health and Social Care Sciences, St. George's, University of London, London, England, United Kingdom

[18]Department of Psychiatry and Behavioral Sciences, Northwestern University Feinberg School of Medicine and The School of Social Service Administration, University of Chicago, Chicago, Illinois, United States

[19]Brain and Mind Research Institute, University of Ottawa and the Ottawa Hospital, Ottawa, Ontario, Canada

[20]Department of Psychology, University of Sheffield, Sheffield, United Kingdom

[21]Department of Psychology, Simon Fraser University, Burnaby, British Columbia, Canada

[22]The University of Michigan Comprehensive Eating Disorders Program, Ann Arbor, Michigan, United States

[23]Department of Psychiatry, University of California, San Francisco, California, United States

 © 2017 Wiley Periodicals, Inc.

Correspondence
Jennifer J. Thomas, Ph.D., Eating Disorders Clinical and Research Program, Massachusetts General Hospital, 2 Longfellow Place, Suite 200, Boston, MA 02114.
Email: jjthomas@mgh.harvard.edu

Abstract

The therapeutic alliance has demonstrated an association with favorable psychotherapeutic outcomes in the treatment of eating disorders (EDs). However, questions remain about the interrelationships between early alliance, early symptom improvement, and treatment outcome. We conducted a meta-analysis on the relations among these constructs, and possible moderators of these relations, in psychosocial treatments for EDs. Twenty studies met inclusion criteria and supplied sufficient supplementary data. Results revealed *small-to-moderate* effect sizes, βs = 0.13 to 0.22 ($p < .05$), indicating that early symptom improvement was related to subsequent alliance quality and that alliance ratings also were related to subsequent symptom reduction. The relationship between early alliance and treatment outcome was partially accounted for by early symptom improvement. With regard to moderators, early alliance showed weaker associations with outcome in therapies with a strong behavioral component relative to nonbehavioral therapies. However, alliance showed stronger relations to outcome for younger (vs. older) patients, over and above the variance shared with early symptom improvement. In sum, early symptom reduction enhances therapeutic alliance and treatment outcome in EDs, but early alliance may require specific attention for younger patients and for those receiving nonbehaviorally oriented treatments.

Resumen: Objetivo: La alianza terapéutica entre paciente y terapeuta ha demostrado ser una relación con resultados psicoterapéuticos favorables en el tratamiento de los trastornos de la conducta alimentaria (TCA). Sin embargo, quedan preguntas acerca de la inter-relación entre alianza temprana, mejoría temprana de síntomas y resultados del tratamiento. Hicimos un meta-análisis de la relación entre estos constructos y los posibles moderadores de estas relaciones en los tratamientos psicosociales para TCA. Método: Veintiún estudios reunieron los criterios de inclusión y aportaron suficientes datos suplementarios. Resultados: los resultados revelaron un efecto de la talla pequeño a moderado, β = 0.13 a 0.22 ($p < .05$), encontrando que la mejoría temprana de los síntomas estuvo relacionada con la subsecuente calidad de la alianza y las calificaciones de la alianza también estuvieron relacionadas con la subsecuente reducción de los síntomas. La relación entre alianza temprana y resultados de tratamiento fue parcialmente explicada por la temprana mejoría de los síntomas. Con relación a los moderadores, la alianza temprana mostró débiles asociaciones con el resultado en terapias con un fuerte componente conductual relativo a terapias no conductuales. Sin embargo, la alianza mostró más fuerte relación con los resultados para pacientes más jóvenes (versus mayores), por encima y sobre la varianza compartida con la temprana mejoría de síntomas. Discusión: En resumen, la reducción temprana de los síntomas refuerza la alianza terapéutica y los resultados del tratamiento en TCA, pero la alianza temprana puede requerir atención específica para pacientes jóvenes y para aquellos que no reciben tratamientos basados en una orientación conductual.

KEYWORDS
eating disorder, meta-analysis, therapeutic alliance, treatment outcome

1 | INTRODUCTION

Therapeutic alliance, defined as the collaborative working relationship between patient and therapist, is one of the most frequently investigated common factors associated with psychotherapy outcome (Horvath, Del Re, Flückiger, & Symonds, 2011; Karver, Handelsman, Fields, & Bickman, 2006; Shirk, Karver, & Brown, 2011). In a meta-analysis of 190 studies of adult patients with various psychiatric diagnoses, alliance correlated moderately with outcome at $r = 0.28$ (95% confidence interval 0.25 to 0.30) (Horvath et al., 2011). A meta-analysis of child and youth psychotherapy had similar findings, r_w[1] = 0.22 (95% confidence interval 0.16 to 0.28) (Shirk et al., 2011). Given the robust association between therapeutic alliance and outcome, researchers have concluded that alliance is a critical component of effective psychotherapies (Horvath et al., 2011; Miller & Mizes, 2000; Shirk et al., 2011).

Substantial debate surrounds the importance of therapeutic alliance in eating disorders (EDs). Although qualitative research has consistently indicated that individuals with EDs find their relationship with

[1] r_w = weighted mean correlation

the therapist to be important to their well-being, recovery, and treatment satisfaction (e.g., Escobar-Koch, Mandlich, & Urzua, 2012), quantitative research on the relationship between the alliance and outcome in ED treatment has yielded mixed results. Multiple studies have shown that therapeutic alliance predicts outcome (e.g., Bourion-Bedes et al., 2013; Constantino, Arnow, Blasey, & Agras, 2005; Zeeck & Hartmann, 2005); yet, other studies have found little or no association (e.g., Waller, Evans, & Stringer, 2012; Zaitsoff, Doyle, Hoste, & Le Grange, 2008). Discrepant results across studies may be due to study-level differences in therapeutic approach, ED diagnosis, patient age, or drop-out.

The importance of early alliance relative to that of early symptom change in ED treatment is also unclear. A number of studies have observed strong associations between symptom change and therapeutic alliance in the first few weeks of treatment (Brown, Mountford, & Waller, 2013b; Constantino et al., 2005), as well as early symptom change and later outcomes (Le Grange, Accurso, Lock, Agras, & Bryson, 2014; Raykos, Watson, Fursland, Byrne, & Nathan, 2013). Thus, it could be argued that the alliance is simply a by-product of early symptom change, and that alliance-outcome associations that do not account for the role of early symptom change may be spurious (DeRubeis, Brotman, & Gibbons, 2005). To the extent that a quality alliance may result from versus promote change, some have questioned whether alliance is overvalued, and whether its importance may vary by treatment type (Brown, Mountford, & Waller, 2013a).

1.1 | Possible moderators of the relation between therapeutic alliance and outcome

The strength of the relation between therapeutic alliance and outcome reported in prior studies may depend on a number of study-level characteristics, including therapy type, mean patient age, patient diagnosis, alliance rater, and dropout rate.

1.1.1 | Therapy type

Findings regarding differences in the relationship between alliance and treatment outcome for different types of therapy have been inconclusive. In the non-ED literature, a study investigating two treatments for borderline personality disorder indicated that alliance was more important for outcome in patients receiving behavioral (i.e., dialectical behavioral therapy) versus nonbehavioral (i.e., community care by experts) treatment (Bedics, Atkins, Harned, & Linehan, 2015). Conversely, one meta-analysis found that alliance was relevant to the outcome of therapy *only* when that therapy was relatively unstructured (i.e., nonbehavioral) (Crits-Christoph et al., 1991); though other meta-analyses have not replicated this distinction (Horvath et al., 2011). In EDs specifically, CBT researchers have questioned the relationship between alliance and outcome, with certain studies finding no relationship between alliance and outcome in CBT for anorexia nervosa (AN; e.g., Waller et al., 2012) and bulimia nervosa (BN; e.g., Raykos et al., 2013).

1.1.2 | Patient age

The development of therapeutic alliance may differ in younger versus older patients. Specifically, child and adolescent patients may have limited abstract reasoning skills (Bravender et al., 2007), minimize or deny symptoms, or feel pressure from caregivers to enter treatment involuntarily (Sperry, Roehrig, & Thompson, 2009). Thus, some have argued that clinicians should pay extra attention to establishing a strong alliance relative to other goals early in youth treatment (Sperry et al., 2009). In line with these suggestions, in studies of child and adolescent therapy in general, Shirk et al. (2011) found a trend for stronger alliance-outcome associations among younger patients. In contrast, there is also reason to believe that alliance-outcome associations might be less important to outcome in youth with EDs. For example, family-based treatment (FBT), which empowers parents to take charge of their child's eating, emphasizes a strong alliance with caregivers early in treatment, which may alter the nature of the relationship between patient-rated alliance and outcome. In a meta-analysis of youth treatment studies for a variety of psychiatric disorders, the alliance-outcome association was weaker for family versus individual therapies (McLeod, 2011). While prior ED studies have separately focused on patients of different ages, none have examined patient age as a moderator of the association between alliance and outcome.

1.1.3 | Patient diagnosis

Clinicians have posited differences in the overall quality of the alliance based on ED diagnosis and have speculated that treatment resistance among patients with AN may hinder the development of a positive alliance (Strober, 2004). However, multiple studies have shown alliance to be relatively strong among patients with AN (Sly, Morgan, Mountford, & Lacey, 2013; Waller, et al., 2012). In fact, Antoniou and Cooper's qualitative review of the relationship between alliance and outcome in EDs (2013) suggested that the alliance strongly predicted outcome for patients with AN, whereas findings for BN, binge eating disorder (BED), and subthreshold eating disorders were mixed.

1.1.4 | Therapeutic alliance rater

Studies have shown differential effects depending on whether therapeutic alliance was rated by the patient, the therapist, or an independent observer. In some studies, patient and independent observer ratings of alliance have shown stronger relationships to treatment outcome than therapist ratings (Bachelor & Horvath, 1999). In the case of FBT for EDs, the alliance rating is also complicated by the presence of not only the patient but also the parents, who are expected to implement important treatment interventions. Differences between mother-rated, father-rated, and observer-rated alliance and outcome were noted in a study of FBT for AN (Ellison et al., 2012), with mother-rated alliance showing the strongest relationship to weight gain. Two different studies analyzing data from a large randomized controlled trial comparing CBT and IPT for BN (Constantino et al., 2005; Loeb et al., 2005) found that patient-rated alliance predicted outcome, whereas observer-rated alliance did not.

1.1.5 | Drop-out

Drop-out is a substantial problem in ED treatment studies, with attrition rates ranging from 20% to 73% in inpatient and outpatient settings (Fassino, Pierò, Tomba, & Abbate-Daga, 2009). ED research reflects consistent findings from the wider alliance literature, observing that poor alliance predicts drop-out (Morlino et al., 2007; Sly et al., 2014). Given that variability in therapeutic alliance is associated with drop-out, it is possible that studies with high drop-out would show different alliance-outcome associations versus those with low drop-out.

1.1.6 | Other variables

Other variables, including how therapeutic alliance is measured, and how treatment outcome is defined, could also impact the relation between alliance and outcome.

1.2 | The current meta-analysis

1.2.1 | Primary questions

To better understand the relationship between therapeutic alliance and treatment outcome in EDs, we conducted the first meta-analysis on this topic. Specifically, we evaluated the aggregated strength of the relationship between alliance and outcome by conducting temporal analyses of symptom change. Thus, change in ED symptoms (i.e., weight, ED behaviors, and ED cognitions) over the course of treatment was our definition of outcome in the current meta-analysis. A significant correlation between therapeutic alliance measured at some point in treatment and a treatment outcome, with no covariates in the model, does not demonstrate that the alliance is a causal mechanism of symptom change. In this scenario, there is no control over (1) temporal precedence (i.e., that alliance promotes change measured *after* alliance measurement) and (2) the potential role of change occurring prior to alliance measurement (i.e., the notion that the alliance may be epiphenomenal to symptom reduction that has already occurred). To better assess whether alliance changes independently from, or in interaction with, symptom change, we needed to analyze the alliance-outcome association across multiple points in treatment (with time lags to address temporal sequencing) and account for the role of prior symptom change (Brown et al., 2013a). Because the data required to perform temporal analyses were not included in published articles, our team contacted the corresponding authors of all studies meeting inclusion criteria to acquire the necessary data. Studies whose author(s) responded to our request and were able to retrieve the needed data were included in our meta-analysis (see Method).

Our analyses addressed four questions. The first three concerned the relationship between symptom change and later therapeutic alliance at different points in the treatment, and the fourth addressed the relationship between early alliance and later symptom improvement: (1) Does early change in symptoms (i.e., early improvement) predict early/mid alliance? (2) Does mid-to-end of treatment change in symptoms predict alliance at the end of treatment? (3) Does change in symptoms across the entirety of treatment predict alliance at the end of treatment? (4a) Does early/mid alliance predict subsequent change in symptoms? And (4b) Do early/mid alliance and early symptom change each predict unique variance in subsequent change in symptoms (i.e., Question 4b is an extension of Question 4a, but controlling for symptom change)?

1.2.2 | Potential moderators

In addition to evaluating the strength of the relationship between therapeutic alliance and symptom change, we explored potential moderators (i.e., study-level characteristics that could explain variance in effect sizes). Based on prior literature, we hypothesized that study-level characteristics including therapy type, patient age, patient diagnosis, alliance rater, and study drop-out rate would contribute to differences in effect size.

2 | METHODS

2.1 | Inclusion criteria

We set the following inclusion criteria for studies in our meta-analysis: (a) comprised a sample of patients diagnosed with one or more ED(s), including AN, BN, BED, EDNOS, or subthreshold diagnoses; (b) included a measure of therapeutic alliance at one or more time points to one or more sample groups during the study (e.g., Working Alliance Inventory, Helping Alliance Questionnaire, Helping Relationship Questionnaire, or California Psychotherapy Alliance Scales); (c) conducted and reported at least one statistical analysis of the relationship between alliance and a primary treatment outcome variable (e.g., weight, binge/purge frequency, self-report or interview measure of ED psychopathology); (d) was not a case report; (e) was published between the dates of January 1978 (i.e., the date of the first ED treatment study to report alliance-outcome data) and January 2014: (f) was published in English; and (g) did not utilize data already reported in another study included in the meta-analysis. All studies meeting each of these requirements were retained for further inspection, while the remaining studies were assigned reasons for exclusion.

2.2 | Selection of studies

To identify relevant studies, we conducted a computer-based search using PsycINFO, PubMed, and Academic Search Premier. We also searched ProQuest Dissertations and Theses specifically to locate unpublished studies. We identified search terms for alliance and EDs in the controlled vocabulary of each database. For example, in PsychINFO, the terms for therapeutic alliance were *alliance*, *therapeutic alliance*, *treatment alliance*, *helping alliance*, *working alliance*, *psychotherapy relationship*, *therapeutic relationship*, *therapeutic bond*, *helping relationship*, and *patient therapist relationship*. The PsychINFO terms for EDs were *eating disorders*, *anorexia*, *bulimia*, *binge eating disorder*, *EDNOS*, and *eating disorder not otherwise specified*. We then searched each database for studies that were tagged with both alliance and EDs controlled-vocabulary terms. Lastly, we mined the reference section of eight review articles relevant to alliance in EDs which we identified via the initial electronic database search (Fassino & Abbate-Daga, 2013;

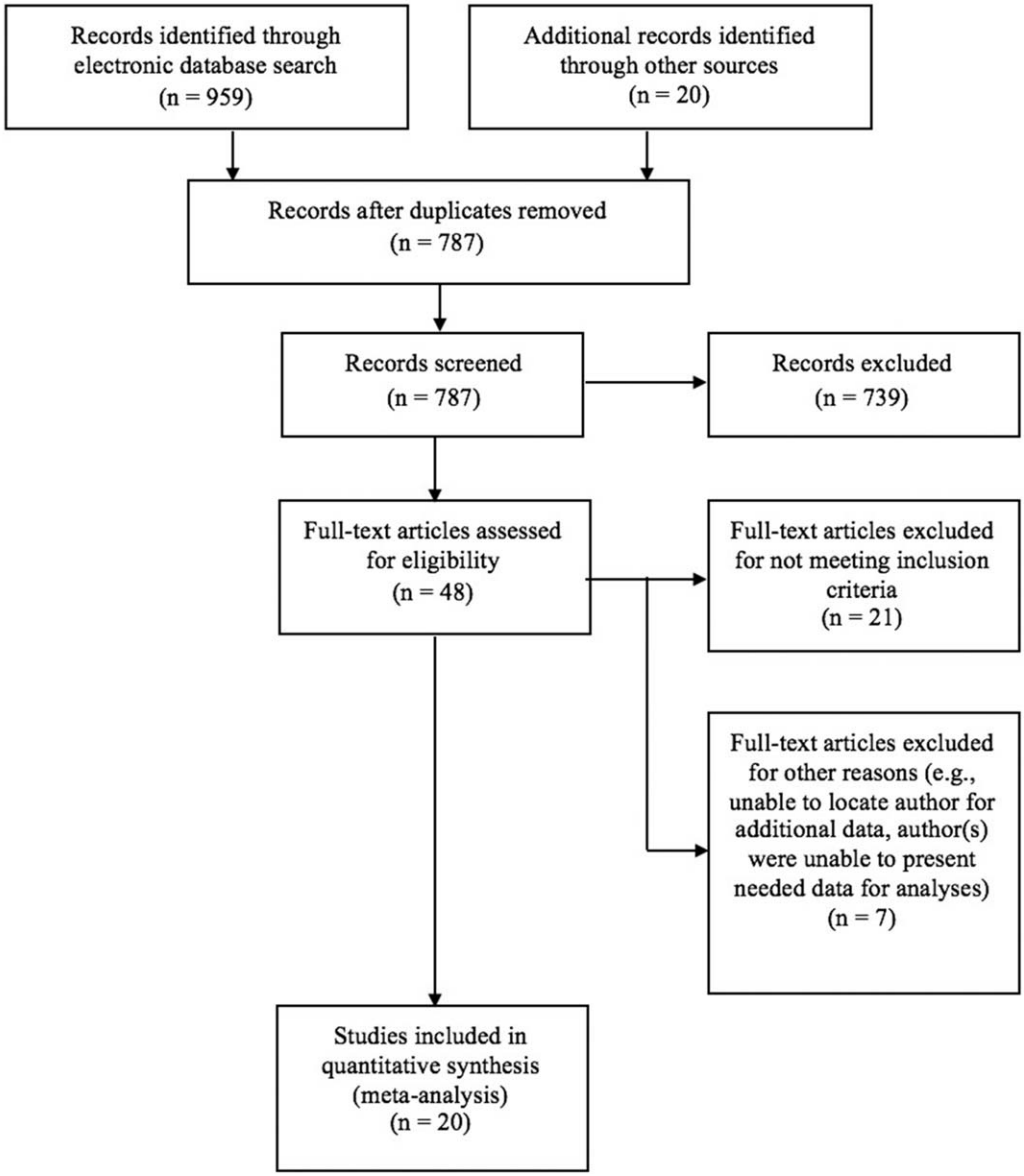

FIGURE 1 PRISMA flow diagram of study selection

Manlick, Cochran, & Koon, 2013; Martin et al., 2011; Shirk et al., 2011; Vitousek & Watson, 1998; Vocks et al., 2010; Westwood & Kendal, 2012; Wilson, 2011) for any additional relevant studies that may have been missed.

The electronic database search combined with the hand search of the review articles resulted in an initial candidate study pool of 787 studies. These studies were then reduced in a stepwise fashion by two independent coders (the first and second authors), as described in the PRISMA diagram (Figure 1). The two coders first screened each abstract, applying the a priori inclusion and exclusion criteria. Of the initial pool of abstracts, 48 studies were retained for full-text screening. The inter-rater reliability between the two coders for abstract screening was acceptable with a kappa = 0.67, $p < .01$. When the coders' ratings diverged, they were discussed until consensus was achieved. The two coders then independently screened the full text of the 48 retained studies. This process resulted in a reduced pool of 27 eligible studies. The inter-rater reliability between the two coders for the full-text screening was substantial with a kappa = 0.76, $p < .01$. These studies were then back-searched using Google Scholar to locate any additional studies referencing those already included in the pool. None of the new studies located during this final step met inclusion criteria.

2.3 | Requests for additional data

To perform the temporal analyses necessitated by our research questions, our team contacted the corresponding authors of all 27 eligible studies to request additional—typically unpublished—data that would be required. We formulated individualized email requests for each author(s) based on data available from the published report. We

received data from the participating studies between May and October of 2014.

Of the 27 authors who received email requests, 20 responded positively, and were able to forward all necessary data in a usable format for the proposed temporal analyses. Only six authors responded negatively to our request, citing that they either (1) did not wish to participate (Ellison et al., 2012; Hildebrandt, Loeb, Troupe, & Delinsky, 2012; Hoffman, 2006); (2) were not able to provide the requested data because it was inaccessible (Treasure et al., 1999; Wilson et al., 1999); or (3) did not collect data from the needed time points (Hartmann, Orlinsky, Weber, Sandholz, & Zeeck, 2010). Finally, one dissertation (Leonard, 2007) could not be included because we could not locate contact information for the corresponding author.

2.4 | Measures of outcome (i.e., ED symptoms) and therapeutic alliance

2.4.1 | Outcome (i.e., ED symptoms)

In the 20 studies included in our meta-analysis, investigators measured improvement in ED symptoms with several relevant measures including body mass index (BMI), weight, percent ideal body weight, binge/purge frequency, vomiting frequency, body checking frequency, Eating Disorder Examination-Questionnaire (EDE-Q), Outcome Questionnaire-45.2, and urge to restrict.

2.4.2 | Therapeutic alliance

In the 20 studies included in our meta-analysis, investigators measured therapeutic alliance with nine different scales: Agnew Relationship Measure; Bern Post-Session Reports for Patients; California Psychotherapy Alliance Scales; Helping Alliance Questionnaire; Helping Relationship Questionnaire; System for Observing Family Alliances; Scale for the Multiperspective Assessment of General Change Mechanisms in Psychotherapy; Treatment Satisfaction Scale; and Working Alliance Inventory. We broadly defined early alliance as the point in treatment when alliance was first measured. For most studies, this point in treatment was between sessions 1 and 5 with the exception of one naturalistic longitudinal study that first measured the alliance at 6 months of treatment, which was approximately mid-way through therapy (average length of treatment, $M = 18 \pm 19$ months; Paulson Karlsson, Clinton, & Nevonen, 2013). We defined mid alliance as the point at which the alliance rating occurring closest to the midpoint of treatment. For most studies, mid alliance was measured between session 6 and 12. We defined late alliance as the alliance rating at the end of treatment. This was always the last alliance measurement taken; timing varied across studies.

2.5 | Levels of each moderator variable

We examined the following variables as possible moderators of the alliance-outcome effect size, classifying each study as falling into one of the following levels on each moderator.

2.5.1 | Therapy type

We coded therapy type as a categorical variable with five categories: behavioral weight-loss therapy (BWLT), CBT, FBT, individual-focused therapy, or multiple therapies. The BWLT category comprised a manualized behavioral treatment following the tetrahydrolipstatin-based weight loss manual that focuses on balanced nutrition and physical activity to promote weight loss (Margraf, 2000; Munsch, Biedert, & Keller, 2003). The CBT category included manualized treatments that employ both cognitive and behavioral strategies to promote eating-disorder symptom change. The FBT category included a manualized treatment that empowers parents to effect change in their child's eating-disorder symptoms (Lock, Le Grange, Agras, & Dare, 2001). The individual-focused therapy category included therapies that fostered the development of insight in related areas, but did directly encourage change in eating-disorder symptoms, including adolescent-focused therapy (AFT), IPT, and supportive psychotherapy (SPT). The multiple therapies category included studies in which participants received two or more different types/modes of therapy either simultaneously or consecutively (e.g., a mixture of inpatient, day-patient, outpatient, individual, group, and/or family therapies [Paulson Karlsson et al., 2013], a treatment combining individual therapy and a supportive program aimed at improving weight and eating behaviors [Bourion-Bedes et al., 2013]).

2.5.2 | Mean age

We recorded the mean age of each study sample as a continuous variable. When there was more than one sample in a study (e.g., a randomized controlled trial), we recorded the mean age for each subsample. However, when the mean age for each subsample was not reported and there was no statistically significant difference in mean age between the subsamples, the mean age for the total sample was used for both subsamples.

2.5.3 | ED diagnosis

We coded ED diagnosis as a categorical variable with four categories: AN, BN, BED, or multiple. A sample was coded as multiple when the sample was composed of people with different ED diagnoses.

2.5.4 | Therapeutic alliance rater

We coded therapeutic alliance rater as a categorical variable with two categories: patient-rated or independent observer-rated. In instances where data from more than one rater of the alliance was included in a study (e.g., patient-rated and parent-rated alliance or patient-rated and therapist-rated alliance), we chose to use the patient-rated alliance or the independent observer-rated alliance (if patient ratings were not provided) because patient ratings and independent observer ratings have shown stronger associations to treatment outcome than parent or therapist ratings of the alliance (Bachelor & Horvath, 1999; Horvath et al., 2011). We did not have any studies in our pool that only collected therapist- or parent-rated alliance data.

2.5.5 | Study drop-out rate

We recorded study drop-out rate as a continuous variable. When there was more than one sample in a study (e.g., a randomized controlled trial), each subsample was assigned the same study drop-out rate, unless drop-out was reported individually for each subsample.

2.6 | Effect size information

We used the standardized regression coefficient (β) to evaluate effect size for each of our four meta-analytic research questions. Rather than extracting effect-size data from the original papers, we obtained more detailed information (i.e., descriptive statistics and correlations) directly from the authors of each study—this was necessary because many studies did not report the information needed to calculate temporal effect sizes. When there was missing data in the summary statistics, we used pairwise deletion in the analyses required to obtain the effect sizes of interest to increase sample size and thus power. When there was no missing data reported in the summary statistics, analyses required to obtain the effect sizes of interest were based on the total sample.

To facilitate comparison across therapy types within each study, we calculated separate effect sizes of the alliance-outcome relation for each treatment arm. We then calculated our own effect sizes, standard errors (*SE*), 95% confidence intervals (CI), and *p*-values using individual multiple linear regression analyses in SPSS. In all analyses, we coded treatment outcome variables such that higher positive scores indicated greater symptom change (e.g., increased BMI in AN trials, decreased binge/purge frequency in BN trials) and stronger alliance. Some studies utilized multiple measures of the same construct—either alliance or ED symptoms. When a particular measure included more than one subscale that could be combined into a global score (e.g., EDE-Q), we used the global score to calculate the effect size for the relation between alliance and outcome. When a measure contained subscales that could not be combined to achieve a total score (e.g., Working Alliance Inventory), we averaged the effect sizes for the outcome-alliance relation for each subscale, to obtain an average effect size for that study (as in Thomas, Vartanian, & Brownell, 2009).

For the first question (i.e., Does early symptom change predict early/middle alliance?), the regression analysis included (1) ED symptoms at baseline, and (2) change in symptoms from baseline to when alliance was first measured, as predictors of the first measure of alliance. Thus, the standardized regression coefficient indexed the relationship between early symptom change and alliance, controlling for baseline symptom level. For the second question (i.e., Does middle to end of treatment symptom change predict later alliance?), the regression analysis included (1) ED symptoms when alliance was first measured and (2) change in symptoms from when alliance was first measured to the end of treatment as predictors of alliance at the end of treatment. Therefore, the standardized regression coefficient indexed the relationship between change in treatment outcome and alliance at the end of treatment, controlling for symptom level at the time of first alliance measurement. For the third question (i.e., Does symptom change across treatment predict later alliance?), the regression analysis included (1) ED symptoms at baseline and (2) change in symptoms from baseline to end of treatment as predictors of alliance at the end of treatment. Thus, the standardized regression coefficient indexed the relationship between change in treatment outcome from baseline to end of treatment and alliance at the end of treatment, controlling for baseline symptom level. For the fourth question (i.e., Does early/mid alliance predict subsequent symptom change?), the regression analysis included (1) alliance and (2) ED symptoms when the alliance was first measured as predictors of change in treatment outcome from when the alliance was first measured to the end of treatment. Therefore, the standardized regression coefficient indexed the relationship between early/mid alliance and subsequent symptom change, controlling for symptom level at the time of alliance measurement. We also conducted a second regression analysis to examine whether early/mid alliance predicts subsequent symptom change *above and beyond early symptom change*. Thus, the regression analysis included (1) the first measure of alliance and (2) change in symptoms from baseline to when alliance was first measured as predictors of change in treatment outcome from when the alliance was first measured to the end of treatment. The standardized regression coefficient indexed the relationship between early/mid alliance and subsequent symptom change while statistically controlling for early symptom change.

2.7 | Meta-analytic procedures

For each research question, we pooled relevant effect sizes, weighted by their inverse-variance ($1/SE^2$). The *SE* of each effect size (β) was calculated using the formula provided by Cohen, Cohen, West, & Aiken (2003; also see Card, 2013):

$$SE_{\beta_i} = \sqrt{\frac{1 - R_Y^2}{n-k-1}} \sqrt{\frac{1}{1-R_i^2}}$$

where R_Y^2 is the variance explained in the dependent variable by the independent variables in the regression model, R_i^2 is the variance explained in the independent variable of interest by the remaining independent variables in the regression model, n is the sample size, and k is the number of independent variables in the regression model. We interpreted the magnitude of each effect size according to Cohen's (1988) conventions for correlation coefficients, where 0.10 is small, 0.30 is moderate, and 0.50 is large.

To allow us to generalize our results beyond the current sample, we used a random-effects model. We assessed publication bias using Egger's test (Egger, Davey, Schneider, & Minder, 1997), which examines the presence of asymmetry in a funnel plot of effect sizes. To examine the impact of each individual effect size on the overall mean effect size, we also conducted a one study removed sensitivity analysis for each meta-analytic research question. Furthermore, we assessed whether the effect sizes were more heterogeneous than expected by sampling variability alone using the test of heterogeneity (*Q*-statistic). When there was evidence of heterogeneity, we used the I^2 statistic to quantify the extent of heterogeneity. We then conducted follow-up moderator analyses using random-effects analogue to ANOVA for categorical moderators, and random-effects meta-regression for continuous moderators. When one or more statistically significant moderators

were at least moderately correlated, we conducted a meta-regression analysis in which we controlled for their shared association. We conducted all analyses using Comprehensive Meta-Analysis Version 2.0 software program (Borenstein, Hedges, Higgins, & Rothstein, 2005) except for the meta-regressions which we conducted using SPSS macros (Lipsey & Wilson, 2001).

3 | RESULTS

3.1 | Does early symptom change predict early/mid alliance?

3.1.1 | Omnibus test

A total of 18 independent effect sizes from 14 different reports evaluated the relationship between early symptom change and early/mid therapeutic alliance (Table 1, Figure 2). For most reports, early/mid alliance was measured between sessions 3 and 10 of treatment; one naturalistic longitudinal study first measured the alliance after 6 months of treatment, which was approximately mid-way through therapy in this particular design (average length of treatment, $M = 18 \pm 19$ months) (Paulson Karlsson et al., 2013). As expected, greater positive change in symptoms (i.e., greater improvement) from baseline to when the alliance was first measured predicted stronger early/mid alliance, $\beta = 0.19$, 95% CI [0.11, 0.28], $z = 4.38$, $p < .0001$. The magnitude of the mean effect size was *small-to-moderate* and there was no evidence of publication bias, Egger's regression intercept = 0.02, t (16) = 0.06, $p = .95$. The mean effect size was stable in our one study removed sensitivity analysis, ranging from 0.17 to 0.22.

3.1.2 | Moderator analyses

In addition, the effect sizes were heterogeneous, Q (17) = 28.41 $p = .04$, but the extent of heterogeneity was low, $I^2 = 40.16$. In follow-up moderator analyses, study drop-out rate was associated with effect sizes at trend-level, Q (1) 3.65, $p = .06$. Specifically, studies with higher drop-out rates had larger effect sizes, slope = 0.01, 95% CI [−0.0002, 0.014], $z = 1.91$, $p = .06$. To further evaluate this finding, we examined the mean effect size at high (+1 *SD*) and low (−1 *SD*) levels of study drop-out rate (weighted $M = 14.60\%$ drop-out rate, $SD = 5.19$). At 1 *SD* above the mean of study drop-out rate, the effect size was *small-to-moderate* and statistically significant, $\beta = 0.21$, 95% CI [0.13, 0.30], $z = 4.91$, $p < .001$. Likewise, at 1 *SD* below the mean of study drop-out rate, the effect size was *small* and statistically significant, $\beta = 0.14$, 95% CI [0.04, 0.24], $z = 2.84$, $p = .004$. Taken together, these findings indicate a positive linear relationship between the magnitude of the effect sizes and study drop-out rate. None of the remaining moderators were statistically significant (Table 2).

3.2 | Does mid-to-end of treatment change in symptoms predict later alliance?

3.2.1 | Omnibus test

A total of ten independent effect sizes from eight different reports evaluated the relationship between mid-to-end of treatment symptom change and later therapeutic alliance (see Table 3, Figure 2). Alliance in all reports was measured at the end of treatment (i.e., at the final treatment session). Results for the overall mean effect size indicated that change in symptoms (i.e., improvement) from when early/mid alliance was measured until the end of treatment was not related to later alliance, $\beta = 0.10$, 95% CI [−0.04, 0.24], $z = 1.46$, $p = .15$. The mean effect size was not statistically significant and there was no evidence of publication bias, Egger's regression intercept = 0.02, t (8) = 0.02, $p = .97$. The mean effect size was stable in our one study removed sensitivity analysis, ranging from 0.07 to 0.12.

3.2.2 | Moderator analyses

Because the effect sizes were homogenous, Q (9) = 2.79, $p = .97$, we did not evaluate potential moderators.[2]

3.3 | Does change in symptoms across treatment predict later alliance?

3.3.1 | Omnibus test

A total of 18 independent effect sizes from 12 different reports evaluated the relationship between change in symptoms across treatment and later alliance (Table 4, Figure 2). In almost all reports, later alliance was measured at the end of treatment. As expected, greater positive change in symptoms (i.e., improvement) across each study's duration predicted greater subsequent alliance, $\beta = 0.17$, 95% CI [0.06, 0.29], $z = 2.96$, $p = .003$. The mean effect size was *small-to-moderate* and there was no evidence of publication bias, Egger's regression intercept = 0.81, t (16) = 0.94, $p = .36$. The mean effect size was stable in our one study removed sensitivity analysis, ranging from 0.16 to 0.20.

3.3.2 | Moderator analyses

Effect sizes were homogenous, Q (17) = 24.17, $p = .12$, so we did not evaluate moderators.

3.4 | Does early/mid alliance predict subsequent symptom change?

3.4.1 | Omnibus test

A total of 19 independent effect sizes from 15 different reports evaluated the relationship between early/mid therapeutic alliance and change in symptoms from when early/mid alliance was measured to the last time-point of data on symptoms available in each report (see Table 5, Figure 2). For almost all reports, the last time-point of data on symptoms was the end of treatment; in one naturalistic longitudinal study, alliance was assessed after 6 months of treatment[3] (Paulson Karlsson et al., 2013). As expected, greater early/mid alliance

[2]Although moderator analyses are often underpowered in meta-analyses comprising a relatively small number of studies, we chose to remain conservative by following the recommendations of Cooper, Hedges, & Valentine (2009) and Lipsey & Wilson (2001) not to evaluate moderators following a nonsignificant omnibus test.

[3]While Paulson Karlsson et al. (2013) also measured alliance 36 months after the start of treatment, we did not include those data in the current meta-analysis because the length of follow-up at the final time point of this longitudinal study differed so greatly from the other included studies, which were primarily much briefer randomized controlled trials.

TABLE 1 Question 1: Does early symptom change predict early/mid alliance?

	Moderator variables								Effect size information		
Report	*N*	Mean age	DX	Therapy setting	Therapy type/ mode	Alliancerater	Session/ of Total	Study drop-out (%)	Alliance rating measure	$\beta(SE)$	95% CI	*p*
Bourion-Bedes et al. (2013)(A)	66	15.30	AN	Inpatient	Multi/IND	PA	3/VAR	0	HAQ	−.01(0.13)[a]	−0.26, 0.23	.92
Bourion-Bedes et al. (2013)(B)	42	15.30	AN	Outpatient	Multi/IND	PA	3/VAR	0	HAQ	0.19(0.17)[a]	−0.15, 0.53	.28
Brown et al. (2013b)	35	25.70	AN	Outpatient	CBT/IND	PA	6/30–40	32.31	WAI	0.46(0.19)[b]	0.10, 0.83	.01
Constantino et al. (2005)(A)	75	28.10	BN	Outpatient	CBT/IND	PA	12/19	25.91	HAQ	0.54(0.19)[b]	0.17, 0.90	< .01
Constantino et al. (2005)(B)	82	28.10	BN	Outpatient	IPT/IND	PA	12/19	25.91	HAQ	0.40(0.12)[b]	0.16, 0.63	< .01
Forsberg et al. (2013)(A)	40	14.80	AN	Outpatient	AFT/IND	IO	3–5/32	17.36	WAI	−0.05(0.17)[a]	−0.38, 0.29	.78
Forsberg et al. (2013)(B)	38	14.00	AN	Outpatient	FBT/IND	IO	3–5/20	17.36	WAI	0.35(0.17)[a]	0.02, 0.67	.04
Isserlin & Couturier (2012)	13	14.00	AN	Outpatient	FBT/IND	IO	3/MDN=12	42.86	SOFTA	0.55(0.39)	−0.22, 1.32	.16
Paulson Karlsson et al. (2013)	41	23.90	AN	MIX	Multi/MIX	PA	MO6/MO18 ± 19	38.00	TSS	0.26(0.15)[b]	−0.02, 0.55	.07
Prestano et al. (2008)	6	16.00	MUL	Outpatient	Other/GRP	PA	WK4/WK104	25.00	CAPAS	0.07(0.81)[a,c]	−1.52, 1.65	.93
Satir (2012)	6	26.90	AN	Outpatient	Multi/IND	PA	5/24	14.29	WAI	−0.20(0.50)[a]	−1.17, 0.78	.69
Simpson et al. (2005)	6	32.00	BN	Outpatient	CBT/IND	PA	4/17	0	ARM	−0.38(0.77)[a]	−1.88, 1.13	.63
Sly et al. (2013)	78	27.73	AN	Inpatient	Other/IND	PA	WK4/VAR	0	WAI	0.23(0.11)[b]	0.02, 0.45	.03
Tasca & Lampard (2012)	127	26.11	MUL	Outpatient	Multi/GRP	PA	WK4/WK12	28.00	CAPAS	0.16(0.08)[b]	−0.01, 0.32	.06
Thompson-Brenner et al. (2015)	36	25.63	BN	Outpatient	CBT/IND	PA	3–5/20	24.00	WAI	−0.52(1.02)[b]	−2.52, 1.48	.61
Waller et al. (2012)	44	27.20	MUL	Outpatient	CBT/IND	PA	6/6	14.00	WAI	0.18(0.02)[b]	0.15, 0.21	< .01
Zaitsoff et al. (2008)(A)	33	11.25	BN	Outpatient	FBT/IND	PA	10/20	11.25	HRQ	−0.39(0.18)[b,c]	−0.74, −0.04	.03
Zaitsoff et al. (2008)(B)	29	11.25	BN	Outpatient	SPT/IND	PA	10/20	11.25	HRQ	0.35(0.19)[b,c]	−0.02, 0.72	.07

Note. $\beta = 0.19$, 95% CI [0.11, 0.28], $z = 4.38$, $p < .0001$; A = treatment arm A; B = treatment arm B; DX = sample diagnosis; AN = anorexia nervosa; BN = bulimia nervosa; MUL = multiple eating disorder diagnoses; MIX = multiple settings; CBT = cognitive behavioral therapy; IPT = interpersonal psychotherapy; AFT = adolescent-focused therapy; FBT = family-based therapy; SPT = supportive psychotherapy; IND = individual; GRP = group; PA = patient; IO = independent observer; VAR = varied; MDN = median; MO = months; WK = week; HAQ = Helping Alliance Questionnaire; WAI = Working Alliance Inventory; SOFTA = System for Observing Family Alliances; TSS = Treatment Satisfaction Scale; CAPAS = The California Psychotherapy Alliance Scales; ARM = Agnew Relationship Measure; HRQ = Helping Relationship Questionnaire; CI = confidence interval.

[a]The study effect size was based on total sample analyses.

[b]The study effect size was based on pairwise regression analyses.

[c]The baseline measure of the outcome was not included in the regression analysis due to multicolinearity.

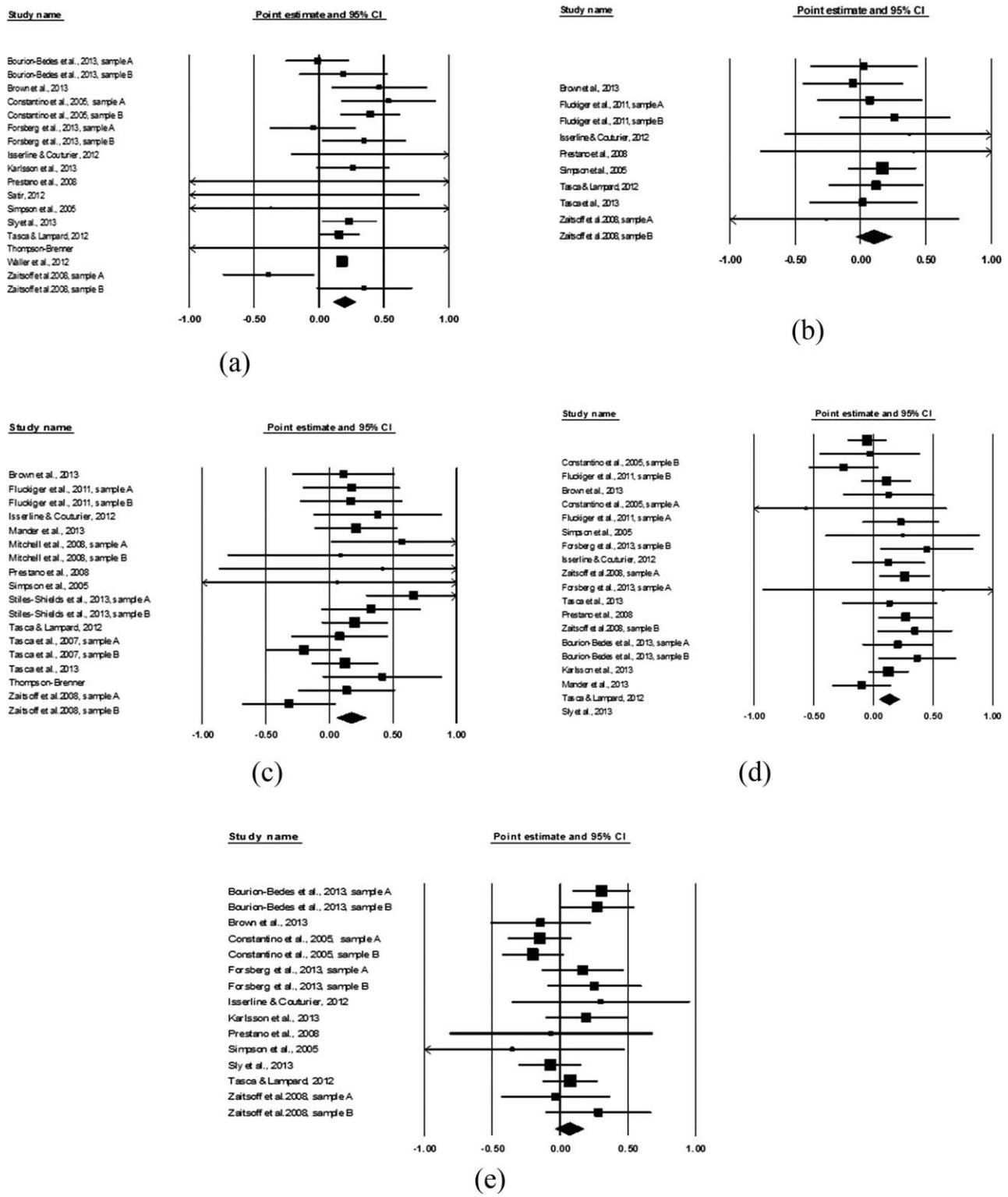

FIGURE 2 Forest plots for all meta-analytic research questions

predicted greater subsequent symptom change, $\beta = 0.13$, 95% CI [0.05, 0.22], $z = 3.10$, $p = .002$. The magnitude of the mean effect size was *small* and there was no evidence of publication bias, Egger's regression intercept = 0.48, t (17) = 0.67, $p = .51$. The mean effect size was stable in our one study removed sensitivity analysis, ranging from 0.11 to 0.15.

3.4.2 | Moderator analyses

The effect sizes were heterogeneous, Q (18) = 26.55, $p = .09$, and the extent of heterogeneity was low, $I^2 = 32.20$. As such, we evaluated potential moderators. Therapy type was related to effect size, Q (4) = 10.61, $p = 0.03$. Specifically, greater early/mid alliance predicted greater subsequent positive change in treatment outcome for studies

TABLE 2 Results of random-effects moderator analyses

	Does early change in treatment outcome predict early/mid alliance?			Does early/mid alliance predict subsequent symptom change?			Does early/mid alliance predict subsequent symptom change above and beyond early symptom change?		
Moderator variable	Q	*df*	*p*	Q	*df*	*p*	Q	*df*	*p*
Therapy type	1.56	3	.67	10.61	4	.03	5.89	3	.12
Mean age	2.77	1	.10	1.03	1	.31	16.20	1	<.01
Eating disorder diagnosis	0.07	2	.97	0.60	3	.90	6.10	2	.047
Alliance rater	0.01	1	.93	0.25	1	.62	1.53	1	.22
Study drop-out rate	3.65	1	.06	0.63	1	.43	0.95	1	.33

TABLE 3 Question 2: Does mid-to-end of treatment change in symptoms predict later alliance?

	Moderator variables								Effect size information		
Report	*N*	Mean age	DX	Therapy setting	Therapy type/mode	Alliance rater/ session	Study drop-out (%)	Alliance rating measure	β(*SE*)	95% CI	*P*
Brown et al. (2013b)	33	25.70	AN	Outpatient	CBT/IND	PA/EOT	32.31	WAI	0.03(0.21)[a]	−0.39, 0.44	.90
Flückiger et al. (2011)(A)	29	45.93	BED	Outpatient	CBT/IND	PA/EOT	29.00	BPSRP	-0.06(0.20)[b]	−0.44, 0.33	.77
Flückiger et al. (2011)(B)	26	45.93	BED	Outpatient	BWLT/IND	PA/EOT	29.00	BPSRP	0.07(0.21)[b]	−0.33, 0.48	.73
Isserlin & Couturier (2012)	14	14.00	AN	Outpatient	FBT/IND	IO/EOT	42.86	SOFTA	0.26(0.22)[a]	−0.17, 0.69	.23
Prestano et al. (2008)	6	16.00	MUL	Outpatient	Other/GRP	PA/EOT	25.16	CAPAS	0.37(0.49)[b]	−0.59, 1.33	.45
Simpson et al. (2005)	6	32.00	BN	Outpatient	CBT/IND	PA/EOT	0	ARM	0.41(0.60)[b]	−0.77, 1.59	.50
Tasca & Lampard (2012)	65	26.11	MUL	Outpatient	Multi/GRP	PA/EOT	28.00	CAPAS	0.17(0.13)[a]	−0.10, 0.43	.22
Tasca et al. (2013)	49	44.30	BED	Outpatient	IPT/GRP	PA/EOT	18.00	CAPAS	0.12(0.19)[a]	−0.24, 0.48	.52
Zaitsoff et al.(2008)(A)	28	16.10	BN	Outpatient	FBT/IND	PA/EOT	11.25	HRQ	0.02(0.21)[a]	−0.40, 0.43	.93
Zaitsoff et al.(2008)(B)	24	16.10	BN	Outpatient	SPT/IND	PA/EOT	11.25	HRQ	−0.26(0.52)[a]	−1.28, 0.76	.62

Note. $\beta = 0.10$, 95% CI [-0.04, 0.24], $z = 1.46$, $p = .15$.; A = treatment arm A; B = treatment arm B; DX = sample diagnosis; AN = anorexia nervosa; BN = bulimia nervosa; BED = binge eating disorder; MUL = multiple eating disorder diagnoses; CBT = cognitive behavioral therapy; IPT = interpersonal psychotherapy; BWLT = behavioral weight loss treatment; AFT = adolescent-focused therapy; FBT = family-based therapy; SPT = supportive psychotherapy; IND = individual; GRP = group; PA = patient; IO = independent observer; EOT = end of treatment; WAI = Working Alliance Inventory; BPSRP = Bern Post-Session Reports for Patients; SOFTA = System for Observing Family Therapy Alliances; CAPAS = The California Psychotherapy Alliance Scales; ARM = Agnew Relationship Measure; HRQ = Helping Relationship Questionnaire; CI = confidence interval.

[a]The study effect size was based on pairwise regression analyses.

[b]The study effect size was based on total sample analyses.

involving multiple therapies ($\beta = 0.18$, 95% CI [0.05 0.32], $z = 2.70$, $p = .007$, $k = 6$); individual-focused therapies ($\beta = 0.21$, 95% CI [0.05, 0.37], $z = 2.56$, $p = 0.01$ $k = 4$); and FBT ($\beta = 0.31$, 95% CI [0.08, 0.54], $z = 2.62$, $p = .009$, $k = 3$). In contrast, early/mid alliance was not related to subsequent positive change in treatment outcome for studies involving BWLT ($\beta = -0.05$, 95% CI [−0.20, 0.11], $z = -0.59$, $p = .56$, $k = 2$), and CBT ($\beta = -0.02$, 95% CI [−0.25, 0.21], $z = -0.19$, $p = .85$, $k = 4$). A follow-up meta-regression analysis evaluated mean differences in effect sizes as a function of therapy type. In the meta-regression, we used CBT as the reference group for therapy type (BWLT was not suitable to serve as the reference group because there were only two studies). Therapy type accounted for 43% of the variance in the effect sizes, $R^2 = 0.43$, Q (4) = 10.35, $p = .04$. The mean effect size for studies involving FBT were larger than the mean effect size for studies involving CBT, $B = 0.31$, $z = 2.12$, $p = .03$. Likewise, there was a nonstatistically significant trend indicating that the mean effect for studies involving multiple therapies tended to be larger than the mean effect size for studies involving CBT, $B = 0.18$, $z = 1.78$, $p = .07$. Similarly, there was also a nonstatistically significant trend indicating that the mean effect size for studies involving individual-focused therapies tended to be larger than the mean effect size for studies involving CBT, $B = 0.21$, $z = 1.76$, $p = .07$. Also, the mean effect size for studies involving BWLT did not differ from the mean effect size for studies involving CBT, $B = -0.04$, $z = -0.37$, $p = .71$, although in this case the size and direction of the effect did not reflect a similar pattern to the other variables, i.e., it was more similar to the results for CBT. In sum, greater early/mid alliance predicted greater subsequent positive change in treatment outcome for studies involving FBT, multiple therapies, and individual-focused therapies relative to studies involving CBT where there was no such effect. None of the remaining moderators were related to variability in the effect sizes (Table 2).

TABLE 4 Question 3: Does change in symptoms across treatment predict later alliance?

	Moderator variables								Effect size information		
Report	*N*	Mean age	DX	Therapy setting	Therapy type/mode	Alliance rater/ session	Study drop-out (%)	Alliance rating measure	β(*SE*)	95% CI	*p*
Brown et al. (2013b)	31	25.70	AN	Outpatient	CBT/IND	PA/EOT	32.31	WAI	0.11(0.21)[a]	−0.29, 0.51	.58
Flückiger et al. (2011)(A)	29	45.93	BED	Outpatient	CBT/IND	PA/EOT	29.00	BPSRP	0.17(0.20)[c]	−0.21, 0.56	.37
Flückiger et al. (2011)(B)	26	45.93	BED	Outpatient	BWLT/IND	PA/EOT	29.00	BPSRP	0.17(0.21)[c]	−0.23, 0.58	.41
Isserlin & Couturier (2012)	14	14.00	AN	Outpatient	FBT/IND	IO/EOT	42.86	SOFTA	0.38(0.26)[a]	−0.12, 0.89	.14
Mander et al. (2013)	39	27.70	AN	Inpatient	Multi/MIX	PA/EOT	28.00	SACiP	0.21(0.17)[c]	−0.12, 0.54	.21
Mitchell et al. (2008)(A)	35	29.60	BN	Outpatient	CBT/IND	PA/EOT	37.50	WAI	0.57(0.29)[a]	0.01, 1.13	.05
Mitchell et al. (2008)(B)	36	28.40	BN	Outpatient	CBT/IND	PA/EOT	37.50	WAI	0.09(0.45)[a,b]	−0.80, 0.98	.85
Prestano et al. (2008)	6	16.00	MUL	Outpatient	Other/GRP	PA/EOT	25.00	CAPAS	0.42(0.66)[b,c]	−0.87, 1.71	.52
Simpson et al. (2005)	6	32.00	BN	Outpatient	CBT/IND	PA/EOT	0	ARM	0.06(0.55)[c]	−1.01, 1.13	.91
Stiles-Shields et al. (2013)(A)	24	33.40	AN	Outpatient	CBT/IND	PA/EOT	22.58	HAQ	0.66(0.19)[a]	0.29, 1.04	< .01
Stiles-Shields et al. (2013)(B)	28	33.40	AN	Outpatient	SSCT/IND	PA/EOT	12.50	HAQ	0.33(0.20)[a]	−0.06, 0.72	.10
Tasca & Lampard (2012)	65	26.11	MUL	Outpatient	Multi/GRP	PA/EOT	28.00	CAPAS	0.20(0.13)[a]	−0.06, 0.46	.13
Tasca et al.(2007)(A)	38	43.86	BED	Outpatient	CBT/GRP	PA/EOT	22.73	CAPAS	0.08(0.20)[a]	−0.30, 0.47	.68
Tasca et al.(2007)(B)	52	43.86	BED	Outpatient	IPT/GRP	PA/EOT	22.73	CAPAS	−0.20(0.15)[a]	−0.50, 0.10	.19
Tasca et al. (2013)	72	44.30	BED	Outpatient	IPT/GRP	PA/WK16	18.00	CAPAS	0.12(0.13)[a]	−0.14, 0.39	.35
Thompson-Brenner et al. (2015)	37	25.63	BN	Outpatient	CBT/IND	PA/14-EOT	24.00	WAI	0.42(0.24)[a]	−0.05, 0.89	.08
Zaitsoff et al.(2008)(A)	29	16.10	BN	Outpatient	FBT/IND	PA/EOT	11.25	HRQ	0.14(0.20)[a,b]	−0.24, 0.52	.49
Zaitsoff et al.(2008)(B)	31	16.10	BN	Outpatient	SPT/IND	PA/EOT	11.25	HRQ	−0.32(0.19)[a,b]	−0.69, 0.06	.10

Note. $\beta = 0.17$, 95% CI [0.06, 0.29], $z = 2.96$, $p = .003$; A = treatment arm A; B = treatment arm B; DX = sample diagnosis; AN = anorexia nervosa; BN = bulimia nervosa; BED = binge eating disorder; MUL = multiple eating disorder diagnoses; CBT = cognitive behavioral therapy; IPT = interpersonal psychotherapy; BWLT = behavioral weight loss treatment; AFT = adolescent-focused therapy; FBT = family-based therapy; SPT = supportive psychotherapy; SSCT= specialist supportive clinical management; IND = individual; GRP = group; PA = patient; IO = independent observer; WK = week; EOT = end of treatment; WAI = Working Alliance Inventory; BPSRP = Bern Post-Session Reports for Patients; SOFTA = System for Observing Family Therapy Alliances; SACiP = Scale for the Multiperspective Assessment of General Change Mechanisms in Psychotherapy; CAPAS = The California Psychotherapy Alliance Scales; ARM = Agnew Relationship Measure; HAQ = Helping Alliance Questionnaire; HRQ = Helping Relationship Questionnaire; CI = confidence interval.

[a]The study effect size was based on pairwise regression analyses.

[b]The baseline measure of the outcome was not included in the regression analysis due to multicolinearity.

[c]The study effect size was based on total sample analyses.

3.5 | Does early/mid alliance predict subsequent symptom change above and beyond early symptom change?

3.5.1 | Omnibus test

A total of 15 independent effect sizes from 11 different reports allowed us to evaluate the relationship between early/mid therapeutic alliance and change in symptoms from when early/mid alliance was measured to the last time-point of data on symptoms available in each report while statistically controlling for early symptom change (Table 6, Figure 2). For almost all reports, the last time-point of data on symptoms was at end of treatment; however, for one study, the last time-point of data on symptoms was at 6-month follow-up (Paulson Karlsson et al., 2013). The mean effect size was not statistically significant, $\beta = 0.07$, 95% CI [−0.04, 0.17], $z = 1.26$, $p = .21$, and there was no evidence of publication bias, Egger's regression intercept = 0.09, t (13) = 0.09, $p = .93$. The mean effect size was stable in our one study removed sensitivity analysis, ranging from 0.03 to 0.09.

3.5.2 | Moderator analyses

The effect sizes were heterogeneous, Q (14) = 23.15, $p = .058$, and the extent of heterogeneity was low, $I^2 = 39.52\%$. Thus, we evaluated potential moderators. Sample mean age was related to effect size, Q (1) = 16.20, $p < .01$. Specifically, studies with older samples had smaller effect sizes relative to studies with younger samples, $B = 0.03$, $z = -4.03$, $p < .001$. To further evaluate this finding, we examined the mean effect size at high (+1 *SD*) and low (−1 *SD*) levels of sample mean age (weighted $M = 22.08$ years old,

TABLE 5 Question 4a: Does early/mid alliance predict subsequent change in symptoms?

	Moderator variables								Effect size information			
Report	*N*	Mean age	DX	Therapy setting	Therapy type/ mode	Alliance rater	Session/ of total	Study drop-out (%)	Alliance Rating measure	β(*SE*)	95% CI	*p*
Bourion-Bedes et al. (2013)(A)	66	15.30	AN	Inpatient	Multi/IND	PA	3/VAR	0	HAQ	0.27(0.12)[a]	0.04, 0.51	.02
Bourion-Bedes et al. (2013)(B)	42	15.30	AN	Outpatient	Multi/IND	PA	3/VAR	0	HAQ	0.35(0.16)[a]	0.04, 0.66	.03
Brown et al. (2013b)	33	25.70	AN	Outpatient	CBT/IND	PA	6/30–40	32.31	WAI	−0.25(0.15)[b]	−0.54, 0.05	.10
Constantino et al. (2005)(A)	72	28.10	BN	Outpatient	CBT/IND	PA	12/19	25.91	HAQ	0.11(0.11)[b]	−0.10, 0.32	.31
Constantino et al. (2005)(B)	76	28.10	BN	Outpatient	IPT/IND	PA	12/19	25.91	HAQ	−0.05(0.08)[b]	−0.21, 0.12	.56
Flückiger et al. (2011)(A)	29	45.93	BED	Outpatient	CBT/IND	PA	6/22	29.00	BPSRP	0.13(0.16)[a]	−0.25, 0.51	.51
Flückiger et al. (2011)(B)	26	45.93	BED	Outpatient	BWLT/IND	PA	6/22	29.00	BPSRP	−0.03(0.22)[a]	−0.45, 0.40	.90
Forsberg et al. (2013)(A)	40	14.80	AN	Outpatient	AFT/IND	IO	3–5/20	17.36	WAI	0.13(0.16)[a]	−0.18, 0.44	.41
Forsberg et al. (2013)(B)	38	14.00	AN	Outpatient	FBT/IND	IO	3–5/20	17.36	WAI	0.23(0.16)[a]	−0.09, .055	.16
Isserlin & Couturier (2012)	13	14.00	AN	Outpatient	FBT/IND	IO	3/MDN=12	42.86	SOFTA	0.25(0.33)[b]	−0.40, 0.90	.46
Paulson Karlsson et al. (2013)	47	23.90	AN	MIX	Multi/MIX	PA	MO6/ MO18 ± 19	38.00	TSS	0.21(0.15)[b]	−0.09, 0.50	.18
Mander et al. (2013)	39	27.70	AN	Inpatient	Multi/MIX	PA	DAY1/M= DAY48.8	28.00	SACiP	0.37(0.17)[a]	0.04, 0.70	.03
Prestano et al. (2008)	6	16.00	MUL	Outpatient	Other/GRP	PA	WK4/WK104	25.00	CAPAS	0.59(0.77)[a]	−0.93, 2.01	.45
Simpson et al. (2005)	6	32.00	BN	Outpatient	CBT/IND	PA	4/17	0	ARM	−0.56(0.60)[a]	−1.74, 0.62	.35
Sly et al. (2013)	78	27.73	AN	Inpatient	Other/IND	PA	WK4/VAR	0	WAI	−0.10(0.13)[b]	−0.35, 0.15	.44
Tasca & Lampard (2012)	89	26.11	MUL	Outpatient	Multi/GRP	PA	WK4/WK12	28.00	CAPAS	0.13(0.09)[b]	−0.04, 0.30	.14
Tasca et al. (2013)	50	44.30	BED	Outpatient	IPT/GRP	PA	WK4/WK16	18.00	CAPAS	0.26(0.11)[b]	0.05, 0.48	.02
Zaitsoff et al. (2008)(A)	28	16.10	BN	Outpatient	FBT/IND	PA	10/20	11.25	HRQ	0.45(0.20)[b]	0.06, 0.84	.03
Zaitsoff et al. (2008)(B)	26	16.10	BN	Outpatient	SPT/IND	PA	10/20	11.25	HRQ	0.14(0.20)[b]	−0.26, 0.54	.50

Note. $\beta = 0.13$, 95% CI [0.05, 0.22], $z = 3.10$, $p = .002$; A = treatment arm A; B = treatment arm B; DX = sample diagnosis; AN = anorexia nervosa; BN = bulimia nervosa; BED = binge eating disorder; MUL = multiple eating disorder diagnoses; CBT = cognitive behavioral therapy; IPT = interpersonal psychotherapy; BWLT = behavioral weight loss treatment; AFT = adolescent-focused therapy; FBT = family-based therapy; SPT = supportive psychotherapy; IND = individual; GRP = group; PA = patient; IO = independent observer; VAR = varied; MDN = median; MO = month; DAYS = days; WK = week; HAQ = Helping Alliance Questionnaire; WAI = Working Alliance Inventory; BPSRP = Bern Post-Session Reports for Patients; SOFTA = System for Observing Family Therapy Alliances; TSS = Treatment Satisfaction Scale; SACiP = Scale for the Multiperspective Assessment of General Change Mechanisms in Psychotherapy; CAPAS = The California Psychotherapy Alliance Scales; ARM = Agnew Relationship Measure; HRQ = Helping Relationship Questionnaire; CI = confidence interval.

[a]The study effect size was based on total sample analyses.

[b]The study effect size was based on pairwise regression analyses.

SD = 5.94). At 1 *SD* above sample mean age, the effect size was *small* and was not statistically significant, $\beta = -0.10$, 95% CI [−0.20, 0.01], $z = -1.71$, $p = .09$. This finding indicates that early/mid alliance did not predict change in symptoms from when the early/mid alliance was measured to the end of treatment above and beyond early symptom change in studies with older patients. However, at 1 *SD* below the sample mean age, early/mid alliance predicted greater improvement in symptoms from when the early/mid alliance was measured to the end of treatment above and beyond early symptom change in studies with younger patients, $\beta = 0.22$, 95% CI [0.11, 0.33], $z = 3.98$, $p = .0001$. The magnitude of the effect size was *small-to-moderate*.

ED diagnosis was also related to the effect sizes, $Q(2) = 6.10$, $p = .04$. Specifically, the mean effect size was statistically significant and *small* for studies with samples of AN ($\beta = 0.16$, 95% CI [0.04, 0.27], $z = 2.57$, $p = 0.01$), but was not significant for studies with samples of BN ($\beta = -0.10$, 95% CI [−0.26, 0.06], $z = -1.17$, $p = .24$) and studies with mixed ED samples ($\beta = 0.06$, 95% CI [−0.18, 0.30], $z = 0.49$, $p = .62$). In short, early/mid alliance predicted greater improvement in symptoms from when the early/mid alliance was measured to the end of treatment above and beyond early symptom change in studies with samples of AN. None of the remaining moderators were statistically significant (Table 2).

TABLE 6 Question 4b: Does early/mid alliance predict subsequent change in symptoms above and beyond early change in symptoms?

	Moderator variables								Effect size information			
Report	*N*	Mean age	DX	Therapy setting	Therapy type/mode	Alliance rater	Session/ of Total	Study drop-out (%)	Alliance rating measure	$\beta(SE)$	95% CI	*p*
Bourion-Bedes et al. (2013)(A)	66	15.30	AN	Inpatient	Multi/IND	PA	3/VAR	0	HAQ	0.31(0.11)[a]	0.09, 0.52	.01
Bourion-Bedes et al. (2013)(B)	42	15.30	AN	Outpatient	Multi/IND	PA	3/VAR	0	HAQ	0.27(0.14)[a]	−0.01, 0.55	.05
Brown et al. (2013b)	33	25.70	AN	Outpatient	CBT/IND	PA	6/30–40	32.31	WAI	−0.14(0.19)[b]	−0.52, 0.23	.46
Constantino et al. (2005)(A)	72	28.10	BN	Outpatient	CBT/IND	PA	12/19	25.91	HAQ	−0.15(0.12)[b]	−0.39, 0.09	.21
Constantino et al. (2005)(B)	76	28.10	BN	Outpatient	IPT/IND	PA	12/19	25.91	HAQ	−0.20(0.12)[b]	−0.43, 0.03	.09
Forsberg et al. (2013)(A)	40	14.80	AN	Outpatient	AFT/IND	IO	3–5/32	17.36	WAI	0.17(0.16)[a]	−0.14, 0.47	.28
Forsberg et al. (2013)(B)	38	14.00	AN	Outpatient	FBT/IND	IO	3–5/20	17.36	WAI	0.25(0.18)[a]	−0.10, 0.60	.16
Isserlin & Couturier (2012)	13	14.00	AN	Outpatient	FBT/IND	IO	3/MDN=12	42.86	SOFTA	0.30(0.34)[b]	−0.36, 0.96	.37
Paulson Karlsson et al. (2013)	47	23.90	AN	MIX	Multi/MIX	PA	MO6/ MO18 ± 19	38.00	TSS	0.19(0.15)[b]	−0.11, 0.50	.21
Prestano et al. (2008)	6	16.00	MUL	Outpatient	Other/GRP	PA	WK4/WK104	25.00	CAPAS	−0.07(0.77)[a]	−0.82, 0.69	.86
Simpson et al. (2005)	6	32.00	BN	Outpatient	CBT/IND	PA	4/17	0	ARM	−0.35(0.42)[a]	−1.18, 0.48	.41
Sly et al. (2013)	78	27.73	AN	Inpatient	Other/IND	PA	4/VAR	0	WAI	−0.07(0.12)[b]	−0.31, 0.16	.53
Tasca & Lampard (2012)	89	26.11	MUL	Outpatient	Multi/GRP	PA	WK4/WK12	28.00	CAPAS	0.07(0.11)[b]	−0.13, 0.28	.48
Zaitsoff et al.(2008)(A)	28	16.10	BN	Outpatient	FBT/IND	PA	10/20	11.25	HRQ	−0.03(0.21)[b]	−0.43, 0.37	.88
Zaitsoff et al.(2008)(B)	26	16.10	BN	Outpatient	SPT/IND	PA	10/20	11.25	HRQ	0.28(0.20)[b]	−0.11, 0.68	.16

Note. $\beta = 0.07$, 95% CI [−0.04, 0.17], $z = 1.26$, $p = .21$; A = treatment arm A; B = treatment arm B; DX = sample diagnosis; AN = anorexia nervosa; BN = bulimia nervosa; BED = binge eating disorder; MUL = multiple eating disorder diagnoses; CBT = cognitive behavioral therapy; IPT = interpersonal psychotherapy; BWLT = behavioral weight loss treatment; AFT = adolescent-focused therapy; FBT = family-based therapy; SPT = supportive psychotherapy; IND = individual; GRP = group; PA = patient; IO = independent observer; VAR = varied; MDN = median; MO = month; WK = week; HAQ = Helping Alliance Questionnaire; WAI = Working Alliance Inventory; HAQ = Helping Alliance Questionnaire; SOFTA = System for Observing Family Therapy Alliances; TSS = Treatment Satisfaction Scale; CAPAS = The California Psychotherapy Alliance Scales; ARM = Agnew Relationship Measure; HRQ = Helping Relationship Questionnaire; CI = confidence interval.

[a]The study effect size was based on total sample analyses.

[b]The study effect size was based on pairwise regression analyses.

Sample mean age ($R^2 = .53$, $p = .04$) and ED diagnosis (Cramer's $V = .52$, $p < .001$) were both moderately associated with therapy type. However, sample mean age and ED diagnosis were not related ($R^2 = .14$, $p = .33$). Thus, we conducted a follow-up meta-regression to examine whether sample mean age and ED diagnosis remained statistically significant predictors of effect size while controlling for shared variance with therapy type. In the meta-regression, we used AN as the reference group for ED diagnosis and individual-focused therapies as the reference group for therapy type. Mean age, ED diagnosis, and therapy type together accounted for 88% of the variance in the effect sizes, $R^2 = 0.88$, $Q\ (6) = 20.42$, $p = .002$. Sample mean age accounted for unique variance in the effect sizes above and beyond ED diagnosis and therapy type, $B = -0.03$, $z = -3.\ 01$, $p = .003$. In contrast, differences in ED diagnosis did not account for unique variance in the effect sizes above and beyond mean age and therapy type. Differences in therapy type did not account for variance in the effect sizes. In sum, findings indicated that early/mid alliance predicted greater subsequent improvement in symptoms above and beyond early symptom change in studies with younger patients, regardless of their ED diagnosis and therapy type.

4 | DISCUSSION

Although ED clinicians have long stressed the role of therapeutic alliance in facilitating symptom change, ED researchers studying behavioral treatments have instead stressed the importance of early symptom change for promoting therapeutic alliance and have debated the relative and temporal influences of these two factors on each other and outcome. Our meta-analysis of 20 ED treatment studies, examining the relations between symptom change and alliance across time and samples, supports a reciprocal relationship between symptom change and alliance. In addition, our analyses are unique in that they are the first in ED treatments to identify that the relative importance of therapeutic alliance for treatment outcome may differ across treatment type, patient age, and patient diagnosis. Interestingly, alliance

rater (independent rater vs. patient) did not impact effect sizes. Further, the current study succeeded in connecting multiple well-known research groups in the field of EDs from across the globe, representing data from nine different countries. We evaluated four distinct research questions, finding statistically significant results for three of the four, with all effect sizes being in the hypothesized direction.

We identified the strongest association between symptom change and subsequent alliance, specifically a *small-to-moderate* sized relationship between early symptom change and early/mid alliance (Question 1), as well as a *small-to-moderate* relationship across-treatment symptom change and subsequent alliance (Question 3). This relationship between symptom change and alliance early in therapy was not moderated by treatment type, ED diagnosis, or other factors, and therefore should be assumed to hold across all levels of these moderator variables. The finding that positive symptom change strengthens therapeutic alliance is consistent with evidence from other psychological disorders, including depression (Tang & DeRubeis, 1999).

However, additional analyses also supported the temporal role of the early alliance in facilitating later symptom change. Results for Question 4a (Does early/mid alliance predict subsequent symptom change?) indicated that early/mid alliance ratings also predicted subsequent changes in outcome. Although differences were noted in the relationship between early alliance and later symptom change between different types of treatment, these results were only significant at the trend level, and should therefore be interpreted with caution. The results of moderator analyses supported the role of early alliance in predicting later symptom-change for individual-focused therapies (e.g., IPT, AFT, and SPT), FBT, and multiple therapies; but not for CBT or BWLT. Further, meta-regression to explore individual comparisons indicated that the differences between CBT and other treatments, excepting BWLT, were particularly strong. These results are very interesting in light of the importance of early symptom change to outcome in CBT (Brown et al., 2013b). It is possible that the alliance is particularly critical in therapies where it is viewed and cultivated as an agent of change; however, further research is needed to confirm this.

Unique analyses compared the relative strength of the associations between treatment outcome and (a) early therapeutic alliance, and (b) early symptom change, including moderator analyses to explore potential differences according to patient age and patient diagnosis. The results indicated that the early alliance was significantly related to subsequent symptom change for younger patients and for patients with AN, but not for older patients or those with other ED diagnoses. Further analyses controlling for the correlations between patient age, ED diagnosis, and treatment type, found that patient age produced the only statistically significant effect after controlling for ED diagnosis and therapy type, indicating that it was a particularly important predictor of a stronger association between the early alliance and outcome. These findings reflect the observations of some alliance researchers outside the field of EDs (Shirk et al., 2011), as well as clinicians who treat adolescents. Importantly, age was a significant moderator even after controlling for treatment type (i.e., individual versus family-based), suggesting that extra attention may need to be paid to the alliance relative to other goals early in treatment for younger patients with EDs (Sperry et al., 2009), regardless of theoretical orientation. Of course, given that age was examined as a study-level (rather than individual-level) moderator in this meta-analysis, we can only draw conclusions about studies that recruited younger patients, rather than any specific youth patient, or youth patients in general.

Our findings also suggested that drop-out rate should be considered when interpreting the size and significance of the relation between symptom change and early alliance ratings. Results indicated that when drop-out was low, symptom change showed a smaller relationship to early/mid alliance ratings, whereas when it was high, early improvement more strongly predicted early/mid alliance ratings. Patients drop out of studies for a wide range of reasons and at various points in treatment (both early and late). Studies that retain patients who are otherwise likely to drop out may include patients with a variety of factors influencing both alliance and symptom change, introducing other sources of variance and error into the symptom change/alliance relationship. It is also possible that patients who drop out of treatment tend to have lower levels of the alliance at the outset or are initially less symptomatic. Thus, it could be argued that drop-out, outcome, and therapeutic alliance are confounded. This possibility should temper the interpretation that early symptom change predicts later alliance as well. A more nuanced study of the drop-out/alliance/outcome relationships in ED samples would help to clarify these questions.

Within the 20 studies included in this meta-analysis, nine different measures of therapeutic alliance were used. Due to the diverse range of alliance measures, it was not possible to include this variable in our moderator analyses. Research indicates that the shared variance among the numerous measures of therapeutic alliance is $<50\%$, even among the four so-called "core" measures (i.e., Working Alliance Inventory, Helping Alliance Questionnaire, California Psychotherapy Alliance Scale, and Vanderbilt Psychotherapy Process Scale) (Horvath, 2009). This suggests that these scales may all be measuring slightly different constructs. Future research should be designed to investigate if and how the type of alliance rating measure used may affect the resulting alliance-outcome associations.

This study has limitations that should be noted. First, our sample of included studies was relatively small. Although there has been an increased focus on therapeutic alliance in recent years, there are still relatively few treatment studies within the ED field that have collected both alliance and outcome data. Moreover, of the studies that have collected such data, most only assess these variables a few times across treatment. To truly begin to untangle this issue, alliance and outcomes should be measured repeatedly, from session 1 to end of treatment. Our findings, combined with others from the ED field (i.e., Tasca & Lampard, 2012), suggest that alliance and outcomes are not static constructs. They change over time and it is quite possible that it is the change in these constructs that is key. Moreover, although our meta-analysis provided the unique opportunity to evaluate changes in both alliance and symptoms over time, the temporal precedence of one over the other does not necessarily imply causality.

Further, despite our best attempts to locate all relevant studies and contact all corresponding authors, there remained a number of applicable studies that were excluded from our meta-analyses because of (1) difficulties contacting the corresponding author(s), or (2) the inability of corresponding author(s) to retrieve the needed data. The inclusion of these missing data could have yielded different results. Second, with regard to our moderator analyses, it is important to note that study sample sizes (*k*) for many of these analyses were quite small, and therefore, results from these analyses should be interpreted with caution. This is particularly true of the moderator analyses involving therapy type. Third, the majority of the included studies were composed of Caucasian females (~90%), which greatly reduces the generalizability of our results to only one subset of the population receiving ED treatment. It is not yet known whether these results would apply to males and/or patients from ethnically diverse backgrounds. In fact, one meta-analysis investigating the moderating effects of the presence of racial/ethnic minorities on the strength of the alliance-outcome association, found that the percentage of overall minorities (particularly African Americans) attenuated the alliance-outcome association (Flückiger et al., 2013). Unfortunately, due to largely homogenous study samples in terms of race and gender and a lack of data regarding patient comorbidities (e.g., substance use disorders), we were unable to investigate the moderating impact of these variables. Fourth, other patient variables (e.g., personality characteristics, attachment style) and therapist characteristics (e.g., gender, experience level) that may impact both alliance and outcome were not measured in a sufficient number of studies to be included as potential moderators. A final limitation of the current meta-analysis is that it was impossible to exclude all third-variable confounds. For instance, it is plausible that patient characteristics not accounted for in our analyses, such as high interpersonal functioning, patient level of insight, or patient motivation or expectancies for change, are associated with both greater alliance and outcome (Jones, Lindekilde, Lübeck, & Clausen, 2015).

5 | CONCLUSIONS

Overall, the bidirectional relationship between therapeutic alliance and outcome found in our meta-analysis strongly suggests the critical value of both early and sustained symptom change, *as well as* the patient-therapist relationship in this clinically challenging population. Symptom improvement was shown to predict subsequent alliance both early in and across the span of treatment, irrespective of treatment type, patient age, or ED diagnosis. Differences in the strength of the relationship between the early alliance and treatment outcome were observed for different treatments, with CBT and BWLT showing weaker associations than other treatment types. Multivariate analyses examining the relative strength of associations between early alliance and later outcome controlling for early symptom change, and examining differences in these relationships according to patient age and patient diagnosis, found that early symptom change accounts for a moderate portion of observed associations between the early alliance and outcome. Analyses indicated that for older patients and those with BN, BED, and mixed/subclinical diagnoses, attention to early symptom change may yield the most benefit for both the early alliance and eventual treatment outcome. However, results of these analyses indicated that younger patients may show specific benefit from additional attention to the early alliance, which showed associations with outcome even when early symptom change was taken into account. These results support a more fine-grained and complex approach to research concerning the inter-relationships between symptom improvement, alliance, and treatment outcome, with attention paid to possible differences in these relationships according to treatment approach and patient factors. Further research is needed to determine the extent to which the bidirectional relationship between therapeutic alliance and symptom change and its attendant moderators is unique to EDs, or more broadly applicable across psychiatric disorders.

ACKNOWLEDGMENTS

This meta-analysis was truly a multisite study. Most of the included papers did not include all of the necessary information in the original published reports to enable our Boston-based research team to answer each meta-analytic research question. For example, some papers reported alliance at the beginning of treatment and its relation to symptom change at the end of treatment, but not to symptom change early in treatment. Therefore, to make this study possible, we reached out to each of the authors of the papers that met inclusion criteria. We asked if these authors could provide their raw data, so that we could recalculate effect sizes for each study and combine them together for the current meta-analysis. To acknowledge the important contribution of these raw data, we invited the first (or, in some cases, corresponding) author on each of the included studies to be a coauthor on the current meta-analysis. For 5 of the 20 studies, we included more than one coauthor, in recognition that both coauthors assisted in preparing raw data for inclusion.

REFERENCES

References marked with an asterisk indicate studies included in the meta-analysis.

Antoniou, P., & Cooper, M. (2013). Psychological treatments for eating disorders: What is the importance of the quality of the therapeutic alliance for outcomes? *Counselling Psychology Review*, *28*, 34–46.

Bachelor, A., & Horvath, A. (1999). The therapeutic relationship. In M. A. Hubble, B. L. Duncan, & S. D. Miller (Eds.), *The heart and soul of change: What works in therapy* (pp. 133–178). Washington, DC: American Psychological Association.

Bedics, J. D., Atkins, D. C., Harned, M. S., & Linehan, M. M. (2015). The therapeutic alliance as a predictor of outcome in dialectical behavior therapy versus non-behavioral psychotherapy by experts for borderline personality disorder. *Psychotherapy*, *52*, 67–77.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). *Comprehensive meta-analysis: Version 2.0*. Englewood, NJ: Biostat.

*Bourion-Bedes, S., Baumann, C., Kermarrec, S., Ligier, F., Feillet, F., Bonnemains, C., ... 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.

Bravender, T., Bryant-Waugh, R., Herzog, D., Katzman, D., Kreipe, R. D., Lask, B., ... Nicholls, D. (2007). Classification of child and adolescent eating disturbances. Workgroup for Classification of Eating Disorders in Children and Adolescents (WCEDCA). *The International Journal of Eating Disorders*, *40*, S117–S122.

Brown, A., Mountford, V. A., & Waller, G. (2013a). Is the therapeutic alliance overvalued in the treatment of eating disorders? *International Journal of Eating Disorders*, *46*, 779–782.

*Brown, A., Mountford, V. A., & Waller, G. (2013b). Therapeutic alliance and weight gain during cognitive behavioural therapy for anorexia nervosa. *Behavior Research and Therapy*, *51*, 216–220.

Card, N. A. (2013). *Applied meta-analysis for social science research*. New York, NY: Guilford Press.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. New York, NY: Routledge Academic.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.

*Constantino, M. J., Arnow, B. A., Blasey, C., & Agras, S. W. (2005). The association between patient characteristics and the therapeutic alliance in cognitive-behavioral and interpersonal therapy for bulimia nervosa. *Journal of Consulting and Clinical Psychology*, *73*, 203–211.

Cooper, H. M., Hedges, L. V., & Valentine, J. C. (2009). *The handbook of research synthesis and meta-analysis*. New York: Russell Sage Foundation.

Crits-Christoph, P., Baranackie, K., Kurcias, J. S., Beck, A. T., Carroll, K., Perry, K., ... Zitrin, C. (1991). Meta-analysis of therapist effects in psychotherapy outcome studies. *Psychotherapy Research*, *1*, 81–91.

DeRubeis, R. J., Brotman, M. A., & Gibbons, C. J. (2005). A conceptual and methodological analysis of the nonspecific argument. *Clinical Psychology: Science and Practice*, *12*, 174–183.

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

Ellison, R., Rhodes, P., Madden, S., Miskovic, J., Wallis, A., Baillie, A., ... Touyz, S. (2012). Do the components of family based treatment for anorexia nervosa predict weight gain? *International Journal of Eating Disorders*, *45*, 609–614.

Escobar-Koch, T., Mandlich, C. C., & Urzua, R. F. (2012). Treatments for eating disorders: The patients' views. In I. J. Lobera (Ed.), *Relevant topics in eating disorders*, Prof. Ignacio Jáuregui Lobera (Ed.), ISBN: 978-953-51-0001-0, InTech. Retrieved from: http://www.intechopen.com/books/relevant-topics-in-eatingdisorders/treatments-for-eating-disorders-the-patients-views

Fassino, S., & Abbate-Daga, G. (2013). Resistance to treatment in eating disorders: A critical challenge. *BMC Psychiatry*, *13*, 1–18.

Fassino, S., Pierò, A., Tomba, E., & Abbate-Daga, G. (2009). Factors associated with drop-out from treatment for eating disorders: A comprehensive literature review. *BMC Psychiatry*, *9*, 1–9.

Flückiger, C., Horvath, A. O., Ackert, M., Del Re, A. C., Symonds, D., & 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.

*Flückiger, C., Meyer, A., Wampold, B. E., Gassmann, D., Messerli-Bürgy, N., & Munsch, S. (2011). Predicting premature termination within a randomized controlled trial for binge-eating patients. *Behavior Therapy*, *42*, 716–725.

*Forsberg, S., LoTempio, E., Bryson, S., Fitzpatrick, K. K., Le Grange, D., & Lock, J. (2013). Therapeutic alliance in two treatments for adolescent anorexia nervosa. *International Journal of Eating Disorders*, *46*, 34–38.

Hartmann, A., Orlinsky, D., Weber, S., Sandholz, A., & Zeeck, A. (2010). Session and intersession experience related to treatment outcome in bulimia nervosa. *Psychotherapy, Theory, Research, Practice, Training*, *47*(3), 355–370.

Hildebrandt, T., Loeb, K., Troupe, S., & Delinsky, S. (2012). Adjunctive mirror exposure for eating disorders: A randomized controlled pilot study. *Behaviour Research and Therapy*, *50*, 797–804.

Hoffman, J. E. (2006). *Alliance as a predictor of outcome in group psychotherapy for binge eating disorder: Making the most predictive model through exploration of mediation and moderation* (Doctoral dissertation). Retrieved from Dissertation Abstracts International: Section B: The Sciences and Engineering. (2009-99040-372). Palo Alto, CA: PGSP - Stanford Psy.D. Consortium.

Horvath, A. O. (2009, October). *Conceptual and methodological challenges in alliance research Is it time for a change?* Paper presented at the Society for Psychotherapy Research European Division, Bolzano, Italy.

Horvath, A. O., Del Re, A. C., Flückiger, C., & Symonds, D. (2011). Alliance in individual psychotherapy. *Psychotherapy*, *48*, 9–16.

*Isserlin, L., & Couturier, J. (2012). Therapeutic alliance and family-based treatment for adolescents with anorexia nervosa. *Psychotherapy*, *49*, 46–51.

Jones, A., Lindekilde, N., Lübeck, M., & Clausen, L. (2015). The association between interpersonal problems and treatment outcome in the eating disorders: A systematic review. *Nordic Journal of Psychiatry*, *13*, 1–11.

Karver, M., Handelsman, J., Fields, S., & Bickman, L. (2006). Meta-analysis of relationship variables in youth and family therapy: Evidence for different relationship variables in the child and adolescent treatment literature. *Clinical Psychology Review*, *26*, 50–65.

Le Grange, D., Accurso, E. C., Lock, J., Agras, S., & Bryson, S. W. (2014). Early weight gain predicts outcome in two treatments for adolescent anorexia nervosa. *International Journal of Eating Disorders*, *47*, 124–129.

Leonard, L. M. (2007). *A time series design evaluating the effectiveness of a residential treatment program for eating disorders*. Retrieved from Dissertation Abstracts International: Section B: The Sciences and Engineering. (2007-99200-1280). Reno, NV: University of Nevada, Reno.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. London, UK: Sage Publications.

Lock, J., Le Grange, D. L., Agras, W. S., & Dare, C. (2001). *Treatment manual for anorexia nervosa: A family-based approach*. New York, NY: Guilford Press.

Loeb, K. L., Pratt, E. M., Walsh, B. T., Wilson, T. G., Labouvie, E., Hayaki, J., ... Fairburn, C. G. (2005). Therapeutic alliance and treatment adherence in two interventions for bulimia nervosa: A study of process and outcome. *Journal of Consulting and Clinical Psychology*, *73*, 1097–1107.

*Mander, J., Teufel, M., Keifenheim, K., Zipfel, S., & Giel, K. E. (2013). Stages of change, treatment outcome and therapeutic alliance in adult in-patients with chronic anorexia nervosa. *BMC Psychiatry*, *13*, 111.

Manlick, C. F., Cochran, S. V., & Koon, J. (2013). Acceptance and commitment therapy for eating disorders: Rationale and literature review. *Journal of Contemporary Psychotherapy*, *43*, 115–122.

Margraf, J. (2000). Aus dick wir nicht dünn. Falsche Erwartungshaltungen. In E. Roche (Ed.), *Hülle und Fülle. Dem Fett auf den Leib gerückt Basel*, Switzerland: Hoffmann-La Roche.

Martin, S., Sutcliffe, P., Griffiths, F., Sturt, J., Powell, J., Adams, A., & Dale, J. (2011). Effectiveness and impact of networked communication interventions in young people with mental health conditions: A systematic review. *Patient Education and Counseling*, *85*, e108–e119.

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

Miller, K. J., & Mizes, J. S. (2000). *Comparative treatments of eating disorders*. New York, NY: Free Association Books.

*Mitchell, J. E., Crosby, R. D., Wonderlich, S. A., Crowe, S., Lancasterb, K., Simonich, H., ... Myers, T. C. (2008). A randomized trial comparing the efficacy of cognitive–behavioral therapy for bulimia nervosa delivered via telemedicine versus face-to-face. *Behaviour Research and Therapy*, *46*, 581–592.

Morlino, M., Di Pietro, G., Tuccillo, R., Galietta, A., Bolzan, M., Senatore, I., ... Valoroso, L. (2007). Drop-out rate in eating disorders: Could it be a function of patient – therapist relationship? *Eating and Weight Disorders*, *12*, 64–67.

Munsch, S., Biedert, E., & Keller, U. (2003). Evaluation of a lifestyle change program for the treatment of obesity in general practice. *Swiss Medical Weekly*, *133*, 148–154.

*Paulson Karlsson, G. P., Clinton, D., & Nevonen, L. (2013). Prediction of weight increase in anorexia nervosa. *Nordic Journal of Psychiatry*, *67*, 424–432.

*Prestano, C., Lo Coco, G., Gullo, S., & Lo Verso, G. (2008). Group analytic therapy for eating disorders: Preliminary results in a single-group study. *European Eating Disorders Review*, *16*, 302–310.

Raykos, B. C., Watson, H. J., Fursland, A., Byrne, S. M., & Nathan, P. (2013). Prognostic value of rapid response to enhanced cognitive behavioral therapy in a routine clinic sample of eating disorder patients. *International Journal of Eating Disorders*, *46*, 764–770.

*Satir, D. (2012). *An experimental analysis of alliance focused treatment for anorexia nervosa* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (Accession Order No. AAT 3500649). Boston, MA: Boston University.

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

*Simpson, S., Bell, L., Knox, J., & Mitchell, D. (2005). Therapy via videoconferencing: A route to client empowerment? *Clinical Psychology and Psychotherapy*, *12*, 156–165.

*Sly, R., Morgan, J. F., Mountford, V. A., & Lacey, J. H. (2013). Predicting premature termination of hospitalised treatment for anorexia nervosa: The roles of therapeutic alliance, motivation, and behaviour change. *Eating Behaviors*, *14*, 119–123.

Sly, R., Morgan, J. F., Mountford, V. A., Sawer, F., Evans, C., & Lacey, J. H. (2014). Rules of engagement: Qualitative experiences of therapeutic alliance when receiving in-patient treatment for anorexia nervosa. *Eating Disorders: The Journal of Treatment & Prevention*, *22*, 233–243.

Sperry, S., Roehrig, M., & Thompson, J. K. (2009). Treatment of eating disorders in childhood and adolescence. In L. Smolak & J. K. Thompson (Eds.), *Body image, eating disorders, and obesity in youth: Assessment, prevention, and treatment* (2nd ed., pp. 261–279). Washington, DC: American Psychological Association.

*Stiles-Shields, C., Touyz, S., Hay, P., Lacey, H., Crosby, R. D., Rieger, E., ... Le Grange, D. (2013). Therapeutic alliance in two treatments for adults with severe and enduring anorexia nervosa. *International Journal of Eating Disorders*, *46*, 783–789.

Strober, M. (2004). Managing the chronic, treatment- resistant patient with anorexia nervosa. *International Journal of Eating Disorders*, *36*, 245–255.

Tang, T. Z., & DeRubeis, R. J. (1999). Sudden gains and critical sessions in cognitive behavioural therapy for depression. *Journal of Consulting and Clinical Psychology*, *67*, 894–904.

*Tasca, G. A., Balfour, L., Ritchie, K., & Bissada, H. (2007). The relationship between attachment scales and group alliance growth differs by treatment type for women with binge-eating disorder. *Group Dynamics: Theory, Research, and Practice*, *11*, 1–14.

*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*, 507–517.

*Tasca, G. A., Ritchie, K., Demidenko, N., Balfour, L., Krysanski, V., Weekes, K., ...Bissada, H. (2013). Matching women with binge eating disorder to group treatment based on attachment anxiety: Outcomes and moderating effects. *Psychotherapy Research*, *23*, 301–314.

Thomas, J. J., Vartanian, L. R., & Brownell, K. D. (2009). The relationship between eating disorder not otherwise specified (EDNOS) and officially recognized eating disorders: Meta-analysis and implications for DSM. *Psychological Bulletin*, *135*(3), 407.

*Thompson-Brenner, H., Shingleton, R. M., Sauer-Zavala, S., Richards, L. K., & Pratt, E. M. (2015). Multiple measures of rapid response as predictors of remission in cognitive behavior therapy for bulimia nervosa. *Behaviour Research and Therapy*, *64*, 9–14.

Treasure, J. L., Katzman, M., Schmidt, U., Troop, N., Todd, G., & de Silva, P. (1999). Engagement and outcome in the treatment of bulimia nervosa: First phase of a sequential design comparing motivation enhancement therapy and cognitive behavioural therapy. *Behavioural Research and Therapy*, *37*, 405–418.

Vitousek, K., & Watson, S. (1998). Enhancing motivation for change in treatment-resistant eating disorders. *Clinical Psychology Review*, *18*, 391–420.

Vocks, S., Tuschen-Caffier, B., Pietrowsky, R., Rustenbach, S. J., Kersting, A., & Herpertz, S. (2010). Meta-analysis of the effectiveness of psychological and pharmacological treatments for binge eating disorder. *International Journal of Eating Disorders*, *43*, 205–217.

*Waller, G., Evans, J., & Stringer, H. (2012). The therapeutic alliance in the early part of cognitive behavioral therapy for the eating disorders. *International Journal of Eating Disorders*, *45*, 63–69.

Westwood, L. M., & Kendal, S. E. (2012). Adolescent client views towards the treatment of anorexia nervosa: A review of the literature. *Journal of Psychiatry and Mental Health Nursing*, *19*, 500–508.

Wilson, G. T. (2011). Treatment of binge eating disorder. *Psychiatric Clinics of North America*, *34*, 773–783.

Wilson, G. T., Loeb, K. L., Walsh, B. T., L., E., Petkova, E., Liu, X., & Waternaux, C. (1999). Psychological versus pharmacological treatments of bulimia nervosa: Predictors and processes of change. *Journal of Consulting and Clinical Psychology*, *67*(4), 451–459.

*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.

Zeeck, A., & Hartmann, A. (2005). Relating therapeutic process to outcome: Are there predictors for the short-term course in anorexic patients? *European Eating Disorders Review*, *13*, 245–254.

How to cite this article: Graves TA, Tabri N, Thompson-Brenner H., Franko DL, Eddy KT, Bourion-Bedes S, Brown A, Constantino MJ, Flückiger C, Forsberg S, Isserlin L, Couturier J, Paulson Karlsson G, Mander J, Teufel M, Mitchell JE, Crosby RD, Prestano C, Satir DA, Simpson S, Sly R, Lacey JH, Stiles-Shields C, Tasca GA, Waller G, Zaitsoff SL, Rienecke R, Le Grange D, Thomas JJ. A meta-analysis of the relation between therapeutic alliance and treatment outcome in eating disorders. *Int J Eat Disord*. 2017;50:323–340. https://doi.org/10.1002/eat.22672