Article:

Merzhvynska, M., Simmonds-Buckley, M., Delgadillo, J. et al. (1 more author) (2021) Trajectories of change in the therapeutic alliance during Cognitive Analytic Therapy for depression. Psychology and Psychotherapy: Theory, Research and Practice, 94 (3). pp. 464-480. ISSN 1476-0835

https://doi.org/10.1111/papt.12322

Trajectories of change in the therapeutic alliance during Cognitive Analytic Therapy for depression

[Authors list masked for peer review]

Running Head: Alliance trajectories during CAT

Word count (excl. figures and tables): 4,594

Declaration of interest: None.

Abstract [word count: 190]

Background: Managing the alliance is considered to be a core competency and central therapeutic change process during cognitive analytic therapy (CAT). This study examined latent trajectories of change in the alliance and their relationship to depression treatment outcomes.

Design: Secondary analysis of a randomised controlled trial.

Methods: A sample of N=79 depressed participants completed standardised alliance (WAI-SF) and depression symptom measures (PHQ-9) every session during an 8-session CAT intervention. Growth mixture modelling was applied to model alliance trajectories and to classify cases into different latent classes. Associations between alliance class and post-treatment PHQ-9 scores were examined using hierarchical linear regression, controlling for confounders.

Results: There were two classes of alliance trajectories. The majority class (91%) displayed stable alliance trajectories, whilst a minority class (9%) had initially poor alliance ratings that significantly improved during treatment. Baseline severity and early change in depression symptoms significantly predicted treatment outcomes, but early alliance and longitudinal alliance change did not.

Conclusions: Alliance trajectories did not significantly predict depression treatment outcomes after controlling for initial symptomatic severity and early change. An important limitation concerns the small sample size, so future replication in larger samples is necessary.

Key words: Cognitive Analytic Therapy; depression; alliance; early response

Historically, the concept of the therapeutic alliance emerged from psychoanalytic theory, and was differentiated from the transference aspects of the analytic relationship (Greenson, 1965). Although a number of definitions have been proposed (e.g., see review by Elvins & Green, 2008), the conceptualisation of the alliance proposed by Bordin (1979) has been particularly influential in the field. According to Bordin's concept, the alliance is a multidimensional construct which reflects the therapist and patient's agreement on the goals and tasks of the therapy, in the context of an affective bond. A number of measures have been developed and widely used to study the alliance in psychotherapy, such as the California Psychotherapy Alliance Scale (CALPAS, Marmar, Horowitz, Weiss, & Marziali, 1986), Helping Alliance Questionnaire (HAQ, Alexander & Luborsky, 1986), Vanderbilt Psychotherapy Process Scale (VPPS, Suh, Strupp, & O'Malley, 1986), and the Working Alliance Inventory (WAI, Horvath & Greenberg, 1989). A vast number of empirical studies on this topic have consistently found that alliance ratings, mostly measured during the early sessions of therapy, are associated with post-treatment outcomes. For example, a recent meta-analysis (Flückiger, Del Re, Wampold & Horvath, 2018) of k=295 independent samples (N>30,000 patients) reported a significant alliance-outcome association of $r = .278$, $p < .0001$; equivalent of $d = .579$). This type of evidence is often interpreted as supporting the notion that positive alliance formation results in better treatment outcomes, and therefore therapists should work directly on developing and maintaining the alliance, particularly when it appears to be fragile or poor (e.g., Safran & Muran, 2000).

Despite the large number of investigations that underline the alliance as a key process in psychotherapy, there are also controversies surrounding this topic. Most studies in this field measure the alliance at a single time-point, which means that its development over time and related temporal confounds are rarely studied (Barber, 2009). Importantly, the idea that within-subject change (e.g., deterioration or improvement) in the alliance may cause symptomatic change does not logically follow from findings of studies that only assess between-subject variability in early alliance ratings. DeRubeis, Brotman, and Gibbons (2005) argued that the alliance could simply be a proxy for patient characteristics (e.g., presenting problem, personality, baseline severity, expectations, etc.) or therapist characteristics (e.g., competence in the delivery of therapy) which may influence treatment outcomes. This possibility cannot be ruled out in studies which measure the alliance at a single time-point, as it is unclear whether alliance ratings were stable over time or if they changed within subjects. Furthermore, there are studies that indicate that alliance ratings may vary as a result of symptomatic improvements, and therefore the quality of the alliance may be a consequence rather than a cause of therapeutic change (e.g., DeRubeis & Feeley, 1990; Feeley et al., 1999; Tang & DeRubeis, 1999).

More recently, studies that applied session-by-session measurement have enabled a more detailed understanding of temporal changes in the alliance (e.g., Falkenström, Granström, & Holmqvist, 2013; Hoffart, Øktedalen, Langkaas, & Wampold, 2013; Tasca & Lampard, 2012; Zilcha-Mano, Dinger, McCarthy, & Barber, 2014). This emerging literature supports the notion of a reciprocal causal model, where the alliance predicts change in symptoms and prior symptom-change also predicts subsequent alliance ratings. The alliance-outcome association also appears to be more important in some cases compared to others, such as in patients with personality and relational dysfunction (Falkenström et al., 2013; Zilcha-Mano, McCarthy, Dinger, & Barber, 2014). These insights about the alliance-outcome association have been enabled by intensive longitudinal measurements during therapy and the application of analytic tools capable of handling complex time-series data (Zilcha Mano, 2016). These have included methods such longitudinal multilevel modelling (Flückiger, Del Re, Wampold, Symonds, & Horvath, 2012) or autoregressive cross-lagged modelling (Zilcha-Mano, Dinger, McCarthy, & Barber, 2014).

Another method capable of analysing patterns in longitudinal change is growth mixture modelling (GMM), which is used to identify latent classes of cases who show similar trajectories of change in the construct being measured over time (Ram & Grimm, 2009). In this way, GMM enables the modelling of between-subject variability (latent classes) and within-patient variability over time. Given that the alliance seems more important in some cases compared to others (between-subject variability) and also appears to change over time (within-subject variability), GMM could be a useful analytical tool in this area of research. Zilcha-Mano and Errázuriz (2017) used this method to identify distinctive patterns of alliance development during the early phase of treatment (first four sessions) using data from 166 patients that received integrative psychotherapy, most of whom (73.5%) were diagnosed with depression. The analysis classified cases into three patterns of early alliance development referred to as (1) early gradual strengthening, (2) early repaired rupture, and (3) early unrepaired rupture. The first two classes continued to show increases in the alliance over time, whereas alliance ratings in the third class remained stable over time. Furthermore, interactions between latent class and initial interpersonal functioning were found. Patients who had better initial interpersonal functioning tended to experience greater symptomatic improvements if early alliance gradually improved (1) or if early ruptured were repaired (2). However, patients with poorer initial interpersonal functioning had greater symptomatic improvement if their alliance ratings followed the early unrepaired rupture pattern (3). To date, few studies have examined alliance development patterns using repeated longitudinal measurements and GMM methods, so the extent to which the above alliance development patterns are generalizable is unknown. Furthermore, recent studies examining longitudinal changes in the alliance fail to control for important confounders, such as the absolute level of the early alliance (which could be a proxy for unmeasured patient-traits as explained above) and early symptomatic improvement during the initial treatment sessions – which is a well-established predictor of psychotherapy outcomes (see review by Beard & Delgadillo, 2019). Further replications are necessary to better understand how different alliance development patterns may influence symptomatic changes in psychotherapy.

Building on the above literature, the present study applied GMM to investigate the development of the therapeutic alliance during cognitive analytic therapy (CAT) for depression. This form of integrative psychotherapy was particularly suitable to study alliance change, as the treatment model (Ryle & Kellett, 2018), change process (Sandhu, Kellett, & Hardy, 2017), competency model (Parry, Roth, Bennnett, & Kellett, 2020) and associated competency measure (Bennett & Parry, 2004) all emphasise using the therapeutic relationship as a specific vehicle for change. The primary methods by which this is achieved during CAT is through analysis of 'reciprocal role enactments' in the therapeutic relationship that mirror the early development trauma of the patient (Ryle & Kellett, 2018), application of a CAT-specific therapeutic rupture-repair model (Bennett, Parry, & Ryle, 2006) and through the tools of CAT (i.e. the psychotherapy file, narrative reformulation, sequential diagrammatic reformulation, self-states description process and goodbye letters) all making reference to the manner in which the patient's history shapes the manner in which the alliance and therapeutic relationship is experienced (Ryle & Kerr, 2002). The primary objective in this study was to identify latent trajectories in alliance development and a secondary objective was to assess if alliance development latent trajectories were associated with post-treatment depression outcomes. Given the paucity of research on longitudinal alliance development classes, we took an exploratory approach to data analysis and did not specify a priori hypotheses.

Methods

Participants

This study was based on data collected from 79 Primary Care patients (73% female), aged 19-65 (*M* = 42.2, SD = 11.26), with clinical case-level depression symptoms measured using the Patient Health Questionnaire-9 (PHQ-9; Kroenke, Spitzer & Williams, 2001) and a DSM structured diagnostic interview. These participants received cognitive analytic therapy (CAT) for depression during a randomized dismantling trial. Full details of the trial, recruitment process and sample characteristics are reported by Kellett et al. (2018). A randomised dismantling trial allows researchers to identify which specific elements of a larger treatment package are causative of change. Participants are randomly assigned to either receiving the full treatment package (in the Kellett trial, full 8-session CAT) or only some components of the treatment package (in the Kellett trial, 8-sessions but with the narrative reformulation component removed). This design controls for threats to internal validity (e.g., maturation, repeated testing, regression to the mean) and amount of time spent in therapy (Behar & Fortune, 2017). The Kellett study allocated N=95 participants therefore to either full-CAT (full-CAT N=52) or CAT minus the narrative reformulation (CAT-NR N= 43). Multilevel modelling showed no differences in depression outcomes following narrative reformulation, at post-treatment or at 8-weeks follow-up and the same results were obtained in a sensitivity analysis controlling for treatment duration. There were large within-group effect sizes (d+ > 1.5), as full-CAT or CAT-NR both produced significant reductions in depression ($p < .01$).

Sampling strategy

The current study included data from participants in both arms of the trial (i.e. full-CAT versus CAT-NR groups), as no differences in outcome between the arms were observed at either mid-treatment, end of treatment or at follow-up (Kellett et al., 2018). Additional inclusion criteria for the present analysis specified that the participants had to have attended all eight treatment sessions and have a depression outcome score and alliance score for each of those sessions. The overall sample size therefore included 632 data-points, clustered in time-series of eight consecutive sessions, within 79 cases (i.e. 83.15% of those included in the original study).

Treatment

Treatment comprised of eight (50-minute) outpatient, one-to-one therapy sessions following the three-phase theoretical structure of CAT: reformulation, recognition and revision (Ryle & Kerr, 2002). A more detailed description of the theory and treatment procedures is available elsewhere (Kellett et al., 2018). The intervention was highly standardised and followed a treatment manual (Stockton, 2012). It was delivered by seven (five females and two males) clinical psychologists and clinical psychology trainees in their final year of training. Therapists attended weekly supervision groups facilitated by a Consultant Clinical Psychologist and accredited CAT psychotherapist, supervisor and trainer. Session audio-tapes were reviewed to ensure treatment fidelity and therapist competency (assessed using the Competency in Cognitive Analytic Therapy measure [CCAT]; Bennett & Parry, 2004). Kellett et al. (2018) concluded that the CCAT evidence indicated that the CAT was competently delivered.

Measures

The PHQ-9 is a valid and reliable 9-item self-report measure of depression symptoms designed for use in Primary Care (Kroenke, Spitzer, & Williams, 2001). Every item is rated on a Likert scale ranging between 0-3, yielding a total depression severity score between 0-27. A cut-off of ≥10 has been recommended as providing the best trade-off between sensitivity (88%) and specificity (88%) for a diagnosis of major depressive disorder (Kroenke, et al., 2001), and a change of 6 or more points is recommended to determine reliable change (Richards & Borglin, 2011). The PHQ-9 was used as the primary measure for evaluating treatment outcome and was completed at screening and before the start of every weekly CAT session.

The Working Alliance Inventory-Short Form *(*WAI-SF) is a self-report measure of the strength of the therapeutic alliance from the patient's perspective (Tracey & Kokotovic, 1989), with demonstrated internal consistency (Smits, Luyckx, Smits, Stinckens, & Claes, 2015) and predictive validity (Busseri & Tyler, 2003). Following Bordin's (1979) conceptualisation, the WAI-SF comprises three subscales: agreement on goals, agreement on tasks and quality of the therapeutic bond. This is a 12-item measure; each item is rated on a seven-point scale (1 = never to 7 = always), yielding a total score between 12 to 84. The WAI-S was used to evaluate therapeutic alliance over the course of treatment and was completed at screening and after the end of every CAT session.

Data analysis

Growth mixture modelling (GMM) was employed to investigate patterns of change in the therapeutic alliance (WAI-SF scores) over the course of treatment. GMM reveals latent subgroups of individuals that have similar trajectories of change over time. Wickrama, Lee, O'Neal and Lorenz (2016) defined GMM as a latent class extension of latent growth curve modelling (GCM). Whereas conventional GCM is based on standard normal distributional assumptions and assumes a single population with common parameters, GMM assumes a mixture of normal distributions and can accommodate the existence of sub-populations captured by latent trajectory classes (Feldman, Masyn, & Conger, 2009; Wickrama et al., 2016).

Therefore, GMM is more appropriate than GCM for the investigation of individual trajectories when heterogeneity in the trajectories of interest is assumed (Muthén & Muthén, 2000). Additional advantages of GMM is that it can model latent trajectories over missing data points and this procedure is robust in cases of non-normality.

GMM was implemented using a conventional step-by-step procedure as recommended by Wickrama et al. (2016), without a need to impute missing data-points (<5%). The procedure included following steps: (1) specification of a traditional growth curve model; (2) specification of a latent class growth analysis (LCGA); (3) specification of a growth mixture model; (4) selection of the optimal unconditional model.

First, session-by-session WAI-SF scores (coded t = 0 to 7) were used to specify and compare a series of growth curve (linear, loglinear, quadratic, cubic and latent basis) models using maximum likelihood estimation with robust standard errors (MLR). Models were compared using conventional fit indices (listed in Tables 1) and Chi-Square difference tests adjusted using the Satorra-Bentler scaling correction (Satorra & Bentler, 2010). The single-group model with the best fit indices was selected as the basis for the LCGA and GMM. Next, a series of LCGA and GMM models were specified using maximum likelihood estimation with robust standard errors (MLR). LCGA is a restricted version of a GMM, which estimates means, but fixes within-class variances to zero and thus assumes that all individual trajectories within a class are homogeneous (Nagin, 1999; Wickrama et al., 2016). LCGA often serves as a starting point for conducting GMM. It is recommended as an initial exploratory option prior to specifying a GMM model, because fixing within-class variances to zero allow for faster model convergence (Jung & Wickrama, 2008).

Due to limited previous research on latent sub-group trajectories in the therapeutic alliance, an exploratory approach was adopted, whereby a one-class model was initially specified and the number of classes was increased until there were no convergence problems, or no further statistical model improvement was present (Ram & Grimm, 2009).

Next, the optimal unconditional model was selected by examining several information criteria (listed in Tables 2-3) and also the entropy index that represented model-based classification accuracy based on posterior probabilities. Entropy values of .40, .60, and .80 represent low, medium, and high class separation (Clark & Muthén, 2009). Adjacent class models (k-1 class model and k-class model, where k refers to number of classes; Nylund, Asparouhov, & Muthén, 2007) were compared using the Lo-Mendell-Rubin (LMR) adjusted likelihood ratio tests and bootstrapped likelihood ratio tests (BLRT). Final model selection was guided by identification of the best fitting and most parsimonious model alongside interpretability of the classes (Wickrama et al., 2016). In the case of controversy between fit indices, BLRT, which outperforms the LMR-LRT, and BIC, that is considered as superior to all other ICs (Nylund et al., 2007), were preferred.

After the final GMM model was selected, we applied a stepwise hierarchical regression analysis to investigate associations between alliance development classes (independent variable) and depression treatment outcomes (dependent variable = PHQ-9 score at the last treatment session), controlling for potential confounders: baseline depression severity, early response, and early alliance level. Predictors were entered in the regression model in four blocks: (B1) baseline PHQ-9; (B2) early alliance measured at session 3; (B3) early symptomatic response measured at session 4[*]; (B4) alliance latent class membership. This modelling approach enabled us to establish the amount of variance explained by each variable. Approximately 5.7% of data-points were missing in key variables for the intended analysis, and these were imputed using a Monte Carlo Markov Chain model that generated five imputed datasets (Gilks, Richardson, & Spiegelhalter, 1995) prior to running the above hierarchical regressions. In a secondary analysis, we ran the same hierarchical regression strategy including early response measured at session 2 in the third block. In this way, we could examine the temporal precedence of early alliance before early response and vice-versa. GCM, LCGA and GMM analyses were conducted using MPlus version 8.2 (Muthén & Muthén, 1998-2017) and stepwise hierarchical regression analysis using SPSS v.26.

* *Note*: Early response was conceptualised as a statistically reliable improvement o ≥ 6 points between session 1 and session 4, based on the recommended reliable change index for the PHQ-9 (Richards & Borglin, 2011). This was a categorical variable coded 1 = early response; 0 = no evidence of early response.

Results

Examination of WAI-SF distributions

Session-by-session density plots and skewness and kurtosis values (available in supplementary materials) were used to initially assess heterogeneity in alliance distributions, representing possible distinct latent subgroups (Wickrama et al., 2016). Visual analysis of the density plots and Z-scores of skewness and kurtosis showed non-normal distributions (presence of Z-scores > ±1.96), as well as change of the distributions over time. This provided an initial indication of potential heterogeneity in the development of alliance scores over treatment time.

Growth mixture modelling

Latent growth curve modelling (LGCM)

Single-group growth trajectories of WAI-SF scores over treatment time were estimated using intercept-only, linear, log-linear, quadratic, cubic and latent basis functions models. All but the cubic LGCM achieved convergence, so the cubic model was excluded from further model comparisons. Table 1 presents the information criteria and model fit indices for the remaining models. Overall, and taking into account all the indices, the quadratic model was chosen as offering the best goodness-of-fit. Chi-square comparisons of nested models supported the selection of the quadratic model, demonstrating a better model fit compared to the linear model (Satorra-Bentler scaled χ^2 diff (4) = 160.33, p <.001). Therefore, the quadratic growth curve model was taken as the basis for subsequent LCGA and GMM models. Despite selection of the quadratic model as the best fitting growth curve model, fit statistics demonstrated overall suboptimal fit of the data (Hu & Bentler, 1998; Tucker & Lewis, 1973). The suboptimal model fit and significant variance in the growth factor (I_V = 185.694, p < .001; S_V = 22.620, p =.016; Q_V = 0.230, p =.010) indicated the potential existence of sub-populations with distinct alliance trajectories over treatment time.

[Table 1]

Latent class growth analysis (LCGA)

Latent trajectories of alliance scores over time were first explored by estimating LCGA with a quadratic growth curve model, producing four models extracting one, two, three and four-class solutions respectively. A comparison of the LCGA models is presented in Table 2. All model fit indices except for the LRT-LMR test suggested better model fit with a three-class solution, compared to the one or two-class solutions. The results of the three and four class models were conflicting. The model with a four-class solution showed a better model fit than the three-class solution, however it revealed a class with a very small sample size (4 participants, 5%). Given that LCGA often leads to over-extraction of classes (Jung & Wickrama, 2008; Infurna & Grimm, 2018) and guided by the principle of parsimony, the four-class solution model was discarded, leaving the model with the three-class solution as the best fitting model. However, the comparison of information criteria of a quadratic single-group growth model (from step 3.2.1 above) and a three-class latent class growth model showed that a model with a three-class solution did not improve on the conventional quadratic GCM.

[Table 2]

Growth mixture modelling (GMM)

Latent trajectories of WAI-SF scores over treatment time were explored by estimating a series of quadratic GMMs (GMM), producing two GMM models; a one-class and a two-class solution. Estimation of a three-class model had convergence problems and was therefore excluded from further model comparisons and the number of classes was not increased any further. Table 3 presents the GMM models comparison. Inspection of the fit indices supported the selection of the two-class model, which had the best goodness-of-fit.

Further comparison with the quadratic single-group growth model (from step 3.2.1 above) and a three-class LCGA model (step 3.2.2 above) also indicated that the two-class GMM model provided additional model improvement and it was therefore selected as the final optimal model. The two-class solution GMM model indicated the existence of a larger class of individuals (Class 1 N = 72; 91%) with a stable level of therapeutic alliance (I_M = 66.6, p < .001; S_M = 0.56, p =.145; Q_M = 0.064, p =.190) and a smaller class of individuals (Class 2 N = 7; 9%) with low initial levels of therapeutic alliance that show a significant improvement over time (I_M = 27.138, p <.001; S_M = 17.02, p <.001; Q_M = -1.397, p <.001). Class 1 had a mean initial WAI-SF score of 66.03 (SD = 11.31) and Class 2 had a mean initial WAI-SF score of 34.60 (SD = 10.78). Figure 1 displays a visual plot of mean and standard deviation alliance scores for each latent class over the 8-sessions of CAT. This visually displays the class 1 alliance stability pattern and the class 2 alliance improvement pattern over sessional time.

[Table 3]

[Figure 1]

Association between therapeutic alliance latent class and outcome

Table 4 presents the coefficients for each of four blocks in the hierarchical linear regression model. After controlling for baseline depression severity which explained 10% of variance in treatment outcomes, adding early alliance in Block 2 (B = -0.11, SE = 0.06, p = .08) did not significantly improve the model's predictive value. Adding early symptom response (by session 4) into Block 3 (B = -4.01, SE = 1.18, p < .001) significantly improved the model's predictive value, explaining approximately 13% of variance. Once all of the above variables were controlled for, the therapeutic alliance class was not statistically significant (B = 1.28, SE = 1.97, p = .52) and did not significantly improve the model. Overall, the final model explained 28% of variance in depression treatment outcomes.

A secondary analysis (full output available in supplementary materials) was conducted in which early symptomatic response by session 2 was entered into the hierarchical regression model. The fully adjusted model (Block 4) revealed that baseline severity (B = 0.45, SE = 0.11, p <.001) and early response (B = -3.78, SE = 1.29, p < .01) were significantly associated with treatment outcomes. Early alliance (B = -0.11, SE = 0.06, p = .05) and therapeutic alliance class (B = 1.57, SE = 2.01, p = .44) were not significant, although the probability value for early alliance was borderline at three decimal places (p = .045). Early alliance was marginally significant (p = .044) in the more parsimonious model in Block 3, which excluded alliance class membership.

[Table 4]

Discussion

Main findings

We examined therapeutic alliance trajectories during a brief 8-session CAT intervention for depression delivered in primary care. Overall, the best-fitting model revealed two latent classes with distinctive longitudinal trajectories of change in the alliance. The majority class (91%) started therapy with above-average alliance scores (M = 66.03) and showed fairly stable scores over the duration of treatment. The minority class (9%) started therapy with below-average scores (M = 34.60) which significantly improved over the course of treatment. However, these alliance development patterns were not significantly associated with depression treatment outcomes. This evidence suggests that whilst it was possible to profile two differing classes of alliance patterns, those with longitudinal improvements in the alliance (i.e. within-subject change), were no more likely than those with stable alliance patterns to show improvements in depression at the end of treatment. Early alliance quality (a between-subjects measurement) was marginally associated with symptomatic improvement when controlling for baseline symptom severity and early response by session 2. However, it was no longer significant when early symptomatic response was measured at session 4. These findings indicate that early response is a more important prognostic indicator, explaining around 13% of variance in depression treatment outcomes in this sample, which is consistent with the moderate effect size attributable to early response in a meta-analysis of depression samples (Beard & Delgadillo, 2019). Furthermore, the proportion of variance attributable to early response in the present sample is nearly twice the magnitude of variance explained by early alliance reported in a meta-analysis of previous studies which was approximately 7% (Flückiger et al., 2018). The results shared similarities with the study by Zilcha-Mano and Errázuriz (2017) examining alliance trajectories during integrative psychotherapy for depression. The smaller class in the current study with improvements in the alliance over time, seems to share trajectory profiles with the early gradual strengthening cluster and also the early repaired rupture clusters (Zilcha-Mano & Errázuriz, 2017).

Strengths and limitations

This study is one of very few longitudinal investigations of the alliance in psychotherapy, adding new insights to an underdeveloped literature on patterns of alliance development and change. To our knowledge, this is the first investigation of patterns of alliance development in CAT, addressing an important gap in the literature for this treatment model (Calvert & Kellett, 2012). A particular strength is that the treatment was highly standardised, conducted in a routine service setting, guided by a treatment manual, closely supervised and with formal competency and fidelity checks within the context of a controlled trial (Kellett et al., 2018). We applied a robust data analytic approach to investigate patterns of change in the alliance, following well-established guidelines on longitudinal model fitting and interpretation (Wickrama et al., 2016). To our knowledge, only one previous study applied a similar analytic approach to study alliance patterns (Zilcha-Mano & Errázuriz, 2017), albeit in a sample where treatments were eclectic in nature and less well-defined or standardised.

Important limitations concern the overall sample size in the present study, which is relatively small with reference to guidelines for latent class modelling (Wickrama et al., 2016). Nevertheless, we obtained models that converged, showing adequate goodness-of-fit indices. Future replications of this analysis are necessary in larger samples to draw firmer conclusions about the generalizability of the latent classes observed in this sample. A further limitation is that we had very limited information about the patient sample. Future studies could collect pre-treatment measures of personality traits and interpersonal problems to examine if these may be correlated with early alliance ratings and distinctive trajectories of change. Furthermore, results were entirely based on patient-reported alinace and depression outcomes, and specifically confined to a sample that met diagnostic criteria for major depressive disorder.

Implications for theory and future research

Like previous researchers (e.g., Falkenström, Granström, & Holmqvist, 2013; Hoffart et al., 2013; Tasca & Lampard, 2012; Zilcha-Mano et al., 2014; Zilcha-Mano & Errázuriz, 2017), we found evidence that some patients show longitudinal changes in the alliance. This evidence fits with the notion that the alliance is a process which unfolds over time within a responsive relationship (Kramer & Stiles, 2015), and suggests that changes in the quality, meaning and experience of the alliance can shift over the course of therapy (Luborsky, 1976). In the present sample, however, a minority (~9%) of patients showed a longitudinal change in their alliance over time, and most patients had a stable level. This study analysed alliance trajectories for participants that had attended all sessions and future research should compare alliance trajectories of differing clinical groups such as dropouts versus completers. Further research should be conducted analysing trends in the alliance with more complex presentations, as CAT therapists tend to work with complex client groups (Ryle, Kellett, Hepple, & Calvert, 2014). Similarly, this research was restricted to analysing the relationship over sessional time between the alliance and depression outcome. Further studies could intensively sample and rate other aspects of the in-session behaviours of therapists over time (e.g. empathy, non-judgemental approach, genuineness) and apply GMM methods to assess if therapeutic clusters are apparent over time. For example, therapists' rated facilitative interpersonal skills have been shown using multi-level modelling to significantly predict client symptom change during brief therapies (Anderson, McClintock, Himawan, Song, & Patterson, 2016).

Whilst the alliance is considered relevant across all psychotherapy models, it has been proposed that some forms of psychotherapy place greater emphasis on the alliance compared to others (Horvath, 2018). Within CBT, the working alliance is operationally defined as consisting of the three essential features of the bond, the goals and the tasks of the therapy (Bordin, 1979), with the relationship being ‘necessary but not sufficient’ in terms of therapeutic change. Within humanistic therapy the alliance is often referred to as the ‘authentic humanness’ of the relationship (Hill & Knox, 2009) and deemed necessary and sufficient. During psychoanalytic therapy, the relationship is the vehicle through which therapy is conducted (Marmarosh, 2012). The alliance in integrative psychotherapy is therefore a framework for developing an integration of these theories and techniques, with the alliance defined as ‘a complex, reciprocal and multidimensional entity’ (Norcross, 2002).

In terms of reciprocity in the alliance, the CAT competency model (Parry, Roth, Bennett, & Kellett, 2020) and associated competency measure (Bennett & Parry, 2004) both emphasise the analysis and associated management of *enactments* in the therapeutic relationship (i.e., the therapist interprets reciprocal roles active in the therapeutic relationship that reflect the early life of the patient to facilitate insight and change) as a key skill. For more complex presentations, CAT therapists use the multiple self-states model (MSSM) to map the various (often contrasting) key states that the client can express and are a reflection of childhood trauma (Ryle & Fawkes, 2007). Therefore, the manner in which the alliance is experienced by both CAT therapist and client can different in each of the states and so places even greater emphasis on reciprocal role analysis to manage ruptures in the alliance. Despite the emphasis that CAT places on the management and analysis of the alliance, only a minority of cases showed improvements in the alliance in the current study, and those who showed improvement were no more likely to attain symptomatic improvements compared to those with stable alliance patterns. However, it is worth noting that the interpretive and relational work concerning the alliance that the CAT model emphasises (Parry et al., 2020) may have facilitated continued engagement and attendance in the therapy in patients with initially poor alliance ratings.

To conclude, it is plausible that early alliance quality contributes to early symptomatic improvements, as suggested by previous research (Zilcha-Mano & Errázuriz, 2017), by enabling the agreement and adherence to therapeutic goals and tasks. However, we found little evidence that changes in the alliance that occur after the early phase contribute to treatment outcomes, particularly after early symptomatic response has been assessed at session 4. This challenges the theoretical assumption that direct change in the alliance influences change in depression symptoms, at least after the earliest sessions of therapy. The study of alliance trajectories needs to be expanded beyond the realm of integrative psychotherapies to assess whether modalities that place more or less emphasis in the therapeutic relationship differ in terms of alliance patterns and symptom changes over time.

References

Anderson, T., McClintock, A.S., Himawan, L., Song, X., & Patterson, C.L. (2016). A prospective study of therapist facilitative interpersonal skills as a predictor of treatment outcome. *Journal of Consulting and Clinical Psychology, 84*, 57-66. https://doi.org/10.1037/ccp0000060

Alexander, L.B. & Luborsky, L. (1986) The Penn Helping Alliance Scales. In: Greenberg, L.S. and Pinsoff, W.M., Eds., *The psychotherapeutic process: a research handbook.* Guilford Press, New York, (pp 325-366).

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

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

Behar, E. & Fortune, M. (2017). Dismantling research design. In A. Wenzel (Ed.), *The sage encyclopedia of abnormal and clinical psychology* (Vol. 1, pp. 1124-1124). Thousand Oaks, CA: SAGE Publications. doi: 10.4135/9781483365817.n446

Bennett, D., & Parry, G. (2004). A measure of psychotherapeutic competence derived from cognitive analytic therapy. *Psychotherapy Research, 14*, 176-192. https://doi.org/10.1093/ptr/kph016

Bennett, D., Parry, G., & Ryle, A. (2006). Resolving threats to the therapeutic alliance in cognitive analytic therapy. *Psychology and Psychotherapy, 79*, 395–418. https://doi: 10.1348/147608305X58355

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

Busseri, M.A., & Tyler, J.D. (2003). Interchangeability of the Working Alliance Inventory and Working Alliance Inventory, short form. *Psychological Assessment*, *15*, 193–197. https://doi.org/10.1037/1040-3590.15.2.193

Calvert, R. & Kellett, S. (2014). Cognitive analytic therapy: A review of the outcome evidence base for treatment. *Psychology and Psychotherapy: Theory, Research and Practice, 87,* 253–277.

Castro-Schilo, L., & Grimm, K.J. (2018). Using residualized change versus difference scores for longitudinal research. *Journal of Social and Personal Relationships, 35*(1), 32-58. https://doi.org/10.1177%2F0265407517718387

Clark, S.L., & Muthén, B.O. (2009). Relating latent class analysis results to variables not included in the analysis. Los Angeles: University of California. Available from: https://www.statmodel.com/download/ relatinglca.pdf.

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

Elvins, R., & Green, J. (2008). The conceptualization and measurement of therapeutic alliance: An empirical review. Clinical Psychology Review, 28(7), 1167-1187. https://doi.org/10.1016/j.cpr.2008.04.002

Falkenström, F., Granström, F, & Holmqvist, R. (2013). Therapeutic alliance predicts symptomatic improvement session by session. *Journal of Counselling Psychology, 60,* 317-28. doi: 10.1037/a0032258.

Flückiger, C., Del Re, A.C., Wampold, B.E., Symonds, D., & Horvath, A.O. (2012). How central is the alliance in psychotherapy? A multilevel longitudinal meta-analysis. *Journal of Counselling Psychology*, *59*,10-17. https://doi.org/10.1037/a0025749

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

Gilks, W.R., Richardson, S., Spiegelhalter, D. (1995). *Markov Chain Monte Carlo in practice*. New York: Chapman and Hall/CRC. https://doi.org/10.1201/b14835

Greenson, R. R. (1965). The working alliance and the transference neurosis. *The Psychoanalytic Quarterly, 34*, 155–179.

Hill, C.E., & Knox, S. (2009). Processing the therapeutic relationship. *Psychotherapy Research*, 19(1), 13-29. https://doi.org/10.1080/10503300802621206

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

Horvath, A.O., & Greenberg, L. S. (1989). Development and validation of the Working Alliance Inventory. *Journal of Counseling Psychology, 36*, 223-233. https://doi.org/10.1037/0022-0167.36.2.223

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

Hu, L.T., & Bentler, P.M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*, 424–453. https://doi.org/10.1037/1082-989X.3.4.424

Infurna, F.J., & Grimm, K.J. (2018). The use of Growth Mixture Modeling for studying resilience to major life stressors in adulthood and old age: Lessons for class size and identification and model selection. *The Journals of Gerontology: Series B*, *73*, 148–159. https://doi.org/10.1093/geronb/gbx019

Jung, T., & Wickrama, K.A.S. (2008). An introduction to Latent Class Growth Analysis and Growth Mixture Modeling. *Social and Personality Psychology Compass*, *2*, 302–317. https://doi.org/10.1111/j.1751-9004.2007.00054.x

Kellett, S., Stockton, C., Marshall, H., Hall, J., Jennings, C., & Delgadillo, J. (2018). Efficacy of narrative reformulation during cognitive analytic therapy for depression: Randomized dismantling trial. *Journal of Affective Disorders*, *239*, 37–47. https://doi.org/10.1016/j.jad.2018.05.070

Kramer, U., & Stiles, W.B. (2015). The responsiveness problem in psychotherapy: A review of proposed solutions. *Clinical Psychology: Science and Practice, 22*, 277-295. https://doi.org/10.1111/cpsp.12107

Kroenke, K., Spitzer, R.L., & Williams, J.B. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine*, *16*, 606–613. https://doi.org/10.1046/J.1525-1497.2001.016009606.X

Lambert, M. J. (1992). Psychotherapy outcome research: Implications for integrative and eclectic therapists. *Handbook of Psychotherapy Integration*, 34(1), 94-129.

Marmar, C.R., Horowitz, M.J., Weiss, D. S., & Marziali, E. (1986). The development of the Therapeutic Alliance Rating System. In L. S. Greenberg & W. M. Pinsof (Eds.), *Guilford clinical psychology and psychotherapy series. The psychotherapeutic process: A research handbook* (p. 367-390). Guilford Press.

Marmarosh, C. L. (2012). Empirically supported perspectives on transference. *Psychotherapy*, 49(3), 364-369.

Meissner, W.W. (2007). Therapeutic alliance: Theme and variations. *Psychoanalytic Psychology, 24*(2), 231–254. https://doi.org/10.1037/0736-9735.24.2.231

Muthén, L.K., & Muthén, B.O. (1998-2017). *Mplus User's Guide*. Eighth Edition. Los Angeles, CA: Muthén & Muthén.

Nagin, D.S. (1999). Analyzing developmental trajectories: A semiparametric, group-based approach. *Psychological Methods*, *4*, 139–157. https://doi.org/10.1037/1082-989X.4.2.139

Norcross, J. C. (2002). *Psychotherapy relationships that work: Therapist contributions and responsiveness to patients.* New York, NY: Oxford University Press.

Nylund, K.L., Asparouhov, T., & Muthén, B.O. (2007). Deciding on the number of classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo simulation study. *Structual Equation Modeling*, *14*, 535–569.

Ram, N., & Grimm, K. J. (2009). Methods and measures: Growth mixture modeling: A method for identifying differences in longitudinal change among unobserved groups. *International Journal of Behavioral Development, 33*(6), 565-576. https://doi.org/10.1177%2F0165025409343765

Richards, D.A., & Borglin, G. (2011). Implementation of psychological therapies for anxiety and depression in routine practice: Two year prospective cohort study. *Journal of Affective Disorders, 133*(1–2), 51–60. https://doi.org/10.1016/J.JAD.2011.03.024

Ryle, A., & Fawkes, L. (2007). Multiplicity of selves and others: cognitive analytic therapy. *Journal of Clinical Psychology, 63*, 165-174. doi: 10.1002/jclp.20339

Parry, G., Roth, A., Bennnett, D., & Kellett, S. (2020). The development of a competency model of cognitive analytic therapy. *Psychology and Psychotherpay, Therory, Research and Practice* (submitted paper).

Ryle, A., & Kerr, I.B. (2002). *Introducing Cognitive Analytic Therapy: Principles and Practice*. *Introducing Cognitive Analytic Therapy: Principles and Practice*. Chichester, UK: John Wiley & Sons. https://doi.org/10.1002/9780470713587

Ryle, A., & Kellett, S. (2018). Cognitive analytic therapy. In *Handbook of Personality Disorders: Theory, Research, and Treatment* (2nd Edition). Edited by W. John Livesley and Roseann Larstone. Guilford Press; NY.

Safran, J.D., & Muran, J.C. (2000). *Negotiating the therapeutic alliance: A relational treatment guide.* Guilford Press.

Satorra, A., & Bentler, P.M. (2010). Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika*, *75*, 243–248. https://doi.org/10.1007/s11336-009-9135-y

Smits, D., Luyckx, K., Smits, D., Stinckens, N., & Claes, L. (2015). Structural characteristics and external correlates of the working alliance inventory-short form. *Psychological Assessment*, *27*, 545–551. https://doi.org/10.1037/pas0000066

Stockton, C. (2012). Treatment manuals for depression dismantling study. University of Sheffield, UK.

Suh, C. S., Strupp, H.H., & O'Malley, S.S. (1986). The Vanderbilt process measures: The Psychotherapy Process Scale (VPPS) and the Negative Indicators Scale (VNIS). In L.S. Greenberg & W.M. Pinsof (Eds.), *Guilford clinical psychology and psychotherapy series. The psychotherapeutic process: A research handbook* (pp. 285-323). Guilford Press.

Tasca, G. A., & Lampard, A.M. (2012). Reciprocal influence of alliance to the group and outcome in day treatment for eating disorders. *Journal of Counseling Psychology, 59*(4), 507. https://psycnet.apa.org/doi/10.1037/a0029947

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

https://doi.org/10.1037/0022-006X.67.6.894

Tracey, T.J., & Kokotovic, A.M. (1989). Factor structure of the Working Alliance Inventory. *Psychological Assessment*, *1*, 207–210. https://doi.org/10.1037/1040-3590.1.3.207

Tucker, L.R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, *38*, 1–10. https://doi.org/10.1007/BF02291170

Wickrama, K., Lee, T.K., O'Neal, C.W., & Lorenz, F.O. (2016). *Higher-order growth curves and mixture modeling with Mplus: A practical guide*. New York, NY: Routledge. https://doi.org/10.4324/9781315642741

Zilcha-Mano, S., McCarthy, K.S., Dinger, U., & Barber, J.P. (2014). To what extent is alliance affected by transference? An empirical exploration. Psychotherapy, 51(3), 424–433. https://doi.org/10.1037/a0036566

Zilcha-Mano, S., Dinger, U., McCarthy, K.S., & Barber, J.P. (2014). Does alliance predict symptoms throughout treatment, or is it the other way around? Journal of Consulting and Clinical Psychology, 82(6), 931–935. https://doi.org/10.1037/a0035141

Zilcha-Mano, S. (2016). New analytic strategies help answer the controversial question of whether alliance is therapeutic in itself. *World Psychiatry, 15*(1), 84. https://dx.doi.org/10.1002%2Fwps.20305

Table 1. Fit statistics for single-group Latent Growth Curve Modelling (LGCM) model comparisons (n=79)

LGCM model	CFI	TLI	RMSEA	X^2	*df*	AIC	BIC	SS-BIC	SRMR
Intercept only	0.072	0.236	0.309	291.127	34	4117.610	4141.304	4109.774	0.566
Linear	0.541	0.585	0.228	158.135	31	3997.968	4028.77	3987.781	0.229
Log linear	0.657	0.69	0.197	126.063	31	3988.515	4019.318	3978.328	0.411
Quadratic	0.783	0.775	0.168	86.978***	27	3959.859	4000.14	3946.538	0.185
Cubic[+]	-	-	-	-	-	-	-	-	-
Latent Basis	0.624	0.579	0.23	129.12	25	3988.288	4033.307	3973.399	0.435

* *p*<.05, ** *p*<.01, ****p*<.001

[+]Cubic model did not converge. Abbreviations: CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; RMSEA = Root Mean Square Error Approximation; X^2 = Chi-Square test of model fit; *df* = degrees of freedom; AIC = Akaike's Information Criterion; BIC = Bayesian Information Criterion; SS-BIC = sample size adjusted Bayesian Information Criterion; SRMR = Standardised Root Mean Squared Residual.

Table 2. Fit statistics for Latent Class Growth Analysis (LCGA) models (n=79)

LCGA Model		1 class	2 classes	3 classes	4 classes
LL (No. of parameters)		-2207.586 (11)	-2075.016 (15)	-2018.184 (19)	-1974.114 (23)
AIC		4437.172	4180.032	4074.367	3994.228
BIC		4463.236	4215.573	4119.387	4048.726
SS-BIC		4428.553	4168.277	4059.479	3976.206
Entropy		1	0.919	0.936	0.961
adj. LRT-LMR test		n/a	250.792*	107.513	83.369
BLRT test		n/a	265.141***	113.664***	-
Group-size (%)	C1	79 (100)	37 (47)	25 (32)	25 (32)
	C2		42 (53)	18 (23)	35 (44)
	C3			36 (45)	4 (5)
	C4				15 (19)

* *p*<.05, ** *p*<.01, ****p*<.001

Abbreviations: LL = Log-likelihood; AIC = Akaike's Information Criterion; BIC = Bayesian Information Criterion; SS-BIC = sample size adjusted Bayesian Information Criterion adj.; LRT-LMR = Lo-Mendell-Rubin adjusted likelihood ratio test; BLRT = bootstrapped likelihood ratio test; C = class.

Table 3. Fit statistics for Growth Mixture Modelling (GMM) models (n=79)

GMM-CI model		1 class	2 classes
LL (No. of parameters)		-1962.930(17)	-1937.984 (21)
AIC		3959.859	3917.968
BIC		4000.14	3967.727
SSABIC		3946.538	3901.513
Entropy		1	0.987
adj. LRT-LMR test		n/a	47.191**
BLRT test		n/a	49.891***
Group-size (%)	C1	79 (100)	72 (91)
	C2		7 (9)

* *p*<.05, ** *p*<.01, ****p*<.001
Abbreviations: LL = Log-likelihood; AIC = Akaike's Information Criterion; BIC = Bayesian Information Criterion; SS-BIC = sample size adjusted Bayesian Information Criterion adj.; LRT-LMR = Lo-Mendell-Rubin adjusted likelihood ratio test; BLRT = bootstrapped likelihood ratio test; C = class.

Table 4. Hierarchical regression model testing associations between therapeutic alliance class membership and treatment outcome, controlling for baseline severity, early change (at session 4) and early alliance (at session 3)

	B	SE	β	*p*
Step 1 (R^2=0.10)				
Constant	*1.81*	*1.70*		
Baseline PHQ-9	0.32	0.11	0.32	**.004**
Step 2 (R^2=0.14)				
Constant	*8.35*	*4.02*		
Baseline PHQ-9	0.39	0.12	0.38	**.001**
Alliance at session 3	-0.11	0.06	-0.22	.075
Step 3 (R^2=0.27)				
Constant	*5.58*	*3.88*		
Baseline PHQ-9	0.45	0.11	0.45	**<.001**
Alliance at session 3	-0.06	0.06	-0.12	.308
Early PHQ-9 change at session 4	-4.01	1.18	-0.38	**.001**
Step 4 (R^2=0.28)				
Constant	*5.665*	*3.89*		
Baseline PHQ-9	0.44	0.11	0.44	**<.001**
Alliance at session 3	-0.06	0.06	-0.12	.295
Early PHQ-9 change at session 4	-3.86	1.21	-0.37	**.002**
Alliance class membership	1.28	1.97	0.07	.515

Step 2: ΔR^2 = 0.04; Step 3: ΔR^2 = 0.13**; Step 3: ΔR^2 = 0.01.

* *p*<.05; ** *p*<.01; ****p*<.001; β = standardised beta coefficients, averaged across all imputed datasets.

Abbreviations: WAI = Working Alliance Inventory; PHQ-9 = Patient Health Questionnaire – depression symptom severity.

Figure 1. Mean session-by-session alliance (WAI-SF) scores by latent therapeutic alliance class [error bars indicate the SD at each session]