<http://www.diva-portal.org>

Postprint

This is the accepted version of a paper published in *Journal of counseling psychology*. This paper has been peer-reviewed but does not include the final publisher proof-corrections or journal pagination.

Citation for the original published paper (version of record):

Sun, Q., Holmqvist Larsson, M., Falkenström, F. (2021)
Separating the effects of alliance ruptures and improvements on outcome session-by-session using the asymmetric fixed-effect model
Journal of counseling psychology, 68(6): 696-704
<https://doi.org/10.1037/cou0000569>

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:lnu:diva-102471>

©American Psychological Association, 2021. This paper is not the copy of record and may not exactly replicate the authoritative document published in the APA journal. Please do not copy or cite without author's permission. The final article is available, upon publication, at: [DOI:10.1037/cou0000569].

**Separating the effects of improvements and deteriorations in mechanisms on outcome using the
asymmetric effects model**

Qiwu Sun¹

Mattias Holmqvist Larsson²

Fredrik Falkenström^{2, 3}

¹ School of Psychology, Central China Normal University

² Department of Behavioral Sciences and Learning, Linköping University, Sweden

³ Department of Psychology, Linnaeus University, Sweden

Author Note

The project was supported by the National Residents' Knowledge of Mental Health and Promotion Strategies (16ZDA232), and Open Research Program from Key Laboratory of Human Development and Mental Health of Hubei Province, People's Republic of China.

Corresponding author: Fredrik Falkenström, Department of Psychology, Linnaeus University, SE-351 95 Växjö, Email: Fredrik.Falkenstrom@lnu.se

Abstract

Objective: The aim of this paper was to introduce the Asymmetric Fixed Effects model to psychotherapy mechanisms of change researchers as a novel way of studying the effects of improvements and deteriorations in the candidate mechanism(s) separately. Alliance-outcome research was used to illustrate the possibility of estimating separate effects of improvements and deteriorations in the alliance.

Method: Two archival datasets were used. One was from community-based primary care services in Sweden using the Clinical Outcomes in Routine Evaluation — Outcome Measure (CORE-OM) and the Working Alliance Inventory—Short form (WAI-S, therapist form) each session with 1096 patients. The other dataset was from a university counseling center in China using the Session Rating Scale (SRS) and the Outcome Rating Scale (ORS) each session with 292 patients. Data was analyzed using the asymmetric fixed-effects model.

Results: The findings indicated that with raw scores, improvements in alliance from one session to the next were followed by lower symptoms/distress scores by the next session, but alliance deteriorations had no effect on next-session symptoms/distress. With alliance deteriorations and improvements defined relative to the sample's average linear change over time, improvements and deteriorations had equal but opposite effects on next session symptom level.

Conclusion: Findings confirm the utility of the Asymmetric Fixed Effect model across two cross-national samples in showing that alliance deteriorations and improvements can predict next session symptoms separately at the within-person level. Findings raise new questions regarding the use of detrending in within-patient mechanism of change studies.

Keywords: Working Alliance, Asymmetric Effects, Detrending, Mechanisms of Change, Within-Person

Public Significance Statement: Previous mechanisms of change research has assumed that improvements and deteriorations in the candidate mechanism affect outcome equally but in opposite directions. Using the newly developed asymmetric fixed effects model, we show how to separate improvements from deteriorations in working alliance to relax and test the assumption of symmetric effects on outcome. This method generates new possibilities for studying fluctuations in potential mechanisms of change at the within-person level.

Separating the effects of improvements and deteriorations in mechanisms on outcome using the asymmetric effects model

In recent years, psychotherapy mechanisms of change, or process-outcome, research has focused on within-patient fluctuations in mechanisms and outcome across sessions using panel data. These within-patient fluctuations are sometimes referred to as “state-like” processes (Zilcha-Mano, 2017), as opposed to “trait-like” differences between patients. On the within-patient level, fluctuations over time in a candidate mechanism are usually used to predict change in outcome at a later time-point, often on a session-by-session basis. There are several advantages of these kinds of analyses, e.g., by allowing causal direction to be explicitly modeled and tested, often yielding findings that are perceived as clinically meaningful (Falkenström, Finkel, Sandell, Rubel, & Holmqvist, 2017).

For instance, using session-by-session data of alliance and treatment outcome, a recent line of research focuses on within-patient alliance-outcome relations across sessions (Flückiger et al., 2020). These studies indicate that fluctuations in alliance levels throughout treatment predict symptom/distress change to the following sessions. This is often explained by rupture/repair theory, in that ruptures are followed by symptom deterioration and repairs by improvement. This interpretation rests on the assumption that improvements and deteriorations of alliance affect next-session symptom level equally, but in opposite ways, i.e., a symmetry assumption. However, it is a common phenomenon observed in the social sciences that effects are *asymmetric* (Allison, 2019), meaning that the drops and improvements in an independent variable do not have equally strong effects on a dependent variable (e.g., the happiness of earning \$100 is less than the pain of losing \$100). In psychotherapy research, asymmetric effects may be often be plausible, e.g., increases in insight have been shown to lead to symptom relief (Johansson et al., 2010), but decreases in insight may not have the same but opposite

effect on symptoms; increases in behavioral activation has been shown to be beneficial for depressed patients (Hopko, Lejuez, Ruggiero, & Eifert, 2003), but decreases in behavioral activation may not increase depression to the same extent. Thus, studies based on traditional regression models do not tell us whether improvements or deteriorations in the independent variable are most responsible for associations between a mechanism and outcome.

However, a recent methodological development makes it possible to test the effects of improvements and deteriorations in a candidate mechanism separately. Allison (2019), elaborating on prior developments by York and Light (2017), described the how to test what he called an Asymmetric Fixed Effects model (henceforth referred to as AFE model). Basically, two variables are created, one that contains all increases in the candidate mechanism from the previous session and is set to zero when the mechanism has decreased, and another that contains all decreases from the prior session and is set to zero when the mechanism has increased (both variables are zero when the mechanism has not changed). This model thus yields two coefficients for the $X \rightarrow Y$ prediction, one for increases and one for decreases in X. This makes it possible to test the effects of improvement and deterioration in the mechanism separately. If this model fits the data better than a model with only a single variable containing both improvements and deteriorations, this is evidence against the symmetry assumption.

Asymmetric Effects: The Case of the Working Alliance

There are potential asymmetric effects in the case of the alliance-outcome relation, as indicated by working alliance theory. There are two major theoretical bases for believing that improvements in the alliance would predict symptom improvement; either by providing a beneficial context for the technical aspects of treatment or by providing a corrective emotional experience (Christian, Safran, & Muran, 2012; Falkenström, Granström, & Holmqvist, 2013; Marmarosh, 2017; Zilcha-Mano, 2017).

Deteriorations in the alliance quality, from major breakdowns to minor tensions, of which one or both of the participants may be only vaguely aware, is a common definition of an alliance *rupture* (Eubanks, Muran, & Safran, 2018; Safran, Crocker, McMain, & Murray, 1990). Whether a rupture would be expected to lead to symptom deterioration already by the following session is not clear from theory. If the alliance is viewed as the necessary but not sufficient background for therapy techniques to have an effect (the “beneficial context” theory), a rupture would be expected to lead to a therapy stalemate in the form of reduced or no therapeutic effects until the deterioration is resolved, but probably not to symptomatic deterioration. On the other hand, if the alliance is viewed from a relational perspective (the “relational” theory), a rupture may be expected to be the opposite of a corrective emotional experience, i.e., something that reinforces maladaptive relational schemas. If a corrective emotional experience leads to symptomatic improvement, we can expect its opposite to lead to symptomatic deterioration. Thus, from this perspective we might expect more or less symmetric effects.

The Present Study

The overarching aim of the current study was to introduce the AFE model (Allison, 2019) to psychotherapy researchers. To achieve this aim, we used alliance research as an example of how the asymmetric effects model can be used to separate the effects on outcome of deteriorations and improvements in session-by-session alliance scores. Two specific hypotheses were posed: 1) alliance deteriorations predict next-session symptom deterioration; 2) alliance improvements predict next-session improvements in symptoms. Stated differently, we believed that the symmetry assumption of traditional cross-lagged panel models would hold, at least approximately (i.e., we did not expect the effects to be exactly equal). A secondary aim was to compare alliance deteriorations as well as alliance improvements defined as deviations from a general trend vs defined as deviations from the prior level

(Eubanks-Carter, Gorman, & Muran, 2012; Falkenström et al., 2017). This is important because the trend issue for repeated measures in naturalistic settings may contaminate true alliance effects (Whelen, Murphy, & Strunk, 2021).

Methods

Participants

Patients from two archival datasets were used, one was from community-based primary care services collected in Sweden (Holmqvist, Ström, & Foldemo, 2014) and the other was from a university counseling center in China (She et al., 2018). All available cases were included, which has the advantage of increased statistical power due to more patients being included in analyses. In addition, the less restrictive Missing At Random assumption is needed, compared to listwise deletion of cases with missing data, which assumes the more restrictive Missing Completely At Random assumption (Rubin, 1976).

The original primary care sample consisted of 1096 patients, including 614 female and 197 males (the remaining did not respond to the question). The average age was 36.81 years (range 16 to 88) with 16.1% missing data. Their primary presenting concerns included anxiety, relationship problems, depression, grief, work related problems, and somatic problems. About one third of the therapists used more than one therapeutic approach; including supportive, psychodynamic, CBT, crisis intervention, cognitive, behavioral, relational, existential, systemic, and interpersonal.

The university counseling center sample consisted of 292 patients, including 63 men and 199 women. The average age was 21.44 years (range 16 to 32). Their primary presenting concerns included interpersonal problems, family problems, emotional problems, and decreases in academic performance. Among them, 8.6% patients reported that they had a mental health diagnosis, and 7%

patients reported that they had a history of medical treatment for mental disorders. Using a 5-point scale ranging from *low* (1) to *high* (5), therapists rated the extent to which they adhered to the following theoretical approaches in their therapy work and showed the following results: humanistic/client-centered therapy, $M = 3.72$; $SD = 1.43$; cognitive/cognitive-behavioral, $M = 2.25$; $SD = 1.40$; and psychoanalytic/psychodynamic, $M = 2.63$; $SD = 0.86$.

Therapists

In the primary care sample, there were 64 female and 4 male therapists. Among them, forty-two were social workers, nineteen were psychologists, seven were psychiatric nurses, occupational therapists, or psychiatrists. Therapist age ranged from 29 to 64 years ($M = 48.5$, $SD = 9.6$) and clinical experience for 1 to 39 years ($M = 14.4$; $SD = 10.4$).

In the university counseling center sample, there were 33 female and 10 male therapists. Among them, nineteen were student-psychotherapists completing their internship (13 in a master's program and six in doctoral programs in counseling and clinical psychology), six were full-time staff therapists, and eighteen were part-time therapists. Therapist age ranged from 24 to 52 years ($M = 32.0$; $SD = 6.7$) and clinical experience from 0.5 to 16 years ($M = 4.9$; $SD = 3.4$).

Measures

The primary care sample used the Clinical Outcomes in Routine Evaluation — Outcome Measure (CORE-OM; Evans et al., 2002) as outcome variable, and the Working Alliance Inventory — Short form (WAI-S; Hatcher, Lindqvist, & Falkenström, 2020) for measuring therapist-reported working alliance. The university counseling sample used the Outcome Rating Scale (ORS; Miller et al., 2003) as outcome variable, and the Session Rating Scale (SRS; Duncan et al., 2003) for patient-rated working alliance.

The **CORE-OM** comprises 34 items measuring psychological distress rated on a five-point scale ranging from “not at all” to “most of or all the time”. The items cover four major problem areas: well-being, problems/symptoms, functioning, and risk (to self or others). The CORE-OM has shown good internal consistency and test–retest reliability, convergent and discriminant validity, and sensitivity to change (Mellor-Clark, Barkham, Connell, & Evans, 1999). Research indicated that the Swedish version of the CORE-OM showed high internal consistency ($\alpha = .93$) and test–retest reliability ($r = .85$), and acceptable convergent validity (the correlation with HAD depression is .66, and the correlation with HAD anxiety is .68) (Elfström et al., 2013). Higher scores indicate more distress. The internal consistency of CORE-OM for the first session in the current sample was .92

WAI-S (Tracey & Kokotovic, 1989) consists of 12 items reflecting Bordin's (1979) three alliance components; agreement on tasks and goals, and a positive emotional bond. The original seven-point rating scale ranging from “never” to “always” was used, with higher scores indicating better quality of the working alliance. The factor structure of the therapist-reported WAI-S was recently found adequate (Hatcher et al., 2020). The internal consistency of WAI-S for the first session in the current sample was .93.

The **ORS** is a visual analog scale consisting of four 10-cm lines, corresponding to four domains (personal well-being, family and close relations, social function, and general sense of well-being). A centimeter ruler is used to measure the distance to the nearest millimeter from the left end of the scale to the client’s mark on each of the four lines. The scores range from 0 to 40, with lower scores signaling higher distress. The present study used the Chinese version of the ORS. Previous validation research (Qing & Hu, 2008; She, Sun, & Jiang, 2017) demonstrated good internal consistency ($\alpha = .84$) and moderate concurrent validity ($r = .60$) for the Chinese version of the ORS. The internal consistency of

the ORS for the first session in the current sample was $\alpha = .83$.

The **SRS** is a four-item visual analog scale based on Bordin's (1979) conceptualization of the alliance. Similar to the ORS, patients are directed to rate their session on four 10-cm lines, corresponding the following items: relationship with the therapist, goals and topics, the approach used in therapy, and the overall rating of the session. The English SRS also has evidence of generating reliable and valid scores in psychometric investigations (Duncan & Reese, 2015). Psychometric investigations of SRS in the current sample showed acceptable internal consistency ($\alpha = .93$) (She et al., 2018). Higher scores indicate better quality of the working alliance.

Procedure

For the primary care sample, all patients who were referred for psychological treatment at primary care centers in two Swedish regions (Östergötland and Jönköping) were asked to participate in the study between November 2009 to April 2010. Data collection was approved by the Regional Ethical Review Board of Linköping (nr M72– 09). After receiving the informed consent, the patient completed the CORE-OM before each session and the therapist completed WAI-S after each session.

For the counseling center sample, all students who registered at a university counseling center of a large public university located in central China were asked to participate in the study between October 2015 and June 2016. This university counseling center has a policy that each student can only receive up to eight therapy sessions per academic year. Severely distressed clients who needed medication or hospitalization were referred to local hospitals. The study was approved by the Research Ethics Committee of the psychology center at the university. After receiving the informed consent, clients were asked to fill out the ORS before each session and the SRS after each session.

Statistical Analysis

The AFE model for panel data was proposed by Allison (2019). A fixed effects model, or within estimator, is a regression model for panel data in which subject-specific means are removed in order to estimate a model on purely within-person fluctuations over time. This is desirable in, for instance, cross-lagged panel models, when researchers want to separate changes over time from stable individual differences (see Falkenström et al., 2017). There are several ways of estimating a fixed-effect model; one can include a dummy variable for each individual in the dataset (except one, which will be the reference category), which will eliminate the person means. Alternatively, the fixed effects model can be estimated by person-mean centering all included variables, i.e., manually removing the mean of each person's scores from each observation, or by a first-difference model which is estimated on the difference scores between each observation and its immediately prior observation. All of these methods achieve the same aim, i.e., removing subject-specific means, so that only within-subject fluctuations remain. The dummy variable and the person-mean centering approaches will yield the same estimates, while the first-difference approach will not yield the same results when the number of waves is larger than two. Initially, two regular Fixed-effect Models were estimated, as follows:

Model 1

$$OU_{i,t} = \mu + \beta WAI_{i,t-1} + \alpha_i + \varepsilon_{i,t}$$

Model 2

$$OU_{i,t} = \mu + \beta WAI_{i,t-1} + \text{Session}_t + \alpha_i + \varepsilon_{i,t}$$

In these models, μ is the model intercept (representing OU when there is no change in WAI from the previous session), α_i is the person-mean term, which is estimated using the mean-centering

approach¹, and $\varepsilon_{i,t}$ is the model error term specific to each session and assumed to be independent of anything else on the right-hand side of the equation. β represents the change in OU (outcome) for a one-unit change in (person-mean centered) WAI, either that improvements in WAI predict improvements in outcome, or that deteriorations in WAI lead to deterioration in outcome, or both. These models were used as baselines against which to compare the AFE models.

The same raw score may have the opposite meaning depending on deteriorations defined in level or in trend (as an illustration, see Figure 1). To compare the possible impact of different definitions, two models were estimated. Session is not included in model (1), which means it identifies deteriorations in level (raw alliance scores across sessions for a given patient). Session is included in model (2) as a continuous variable; thus, it identifies deteriorations compared to the trend.

In the context of alliance research, we use time (e.g., sessions) as a continuous variable in contrast to Allison (2019) who used it as a categorical variable. Time as a categorical variable means we are detrending against non-linear average change patterns in the group. In economics and political science, it makes sense to use categorical time, since the studies are usually done in real time – e.g., all participants (or companies, countries etc.) are measured at the same time (e.g., years). That makes sense if there was, for instance, economic recession in one particular year, or to control for a “Trump effect”. Since those are effects that happen in particular years, we can reasonably assume they will affect everyone in the sample equally. In psychotherapy research, for instance session 4 is usually not at the same point in time for all participants, so it seems unlikely that a dummy variable for session 4 will control for anything that affects all participants equally. In addition, some recent papers question the use of this so-called two-way fixed effects model, due to problems in interpretation and in its

---

¹ Most psychotherapy researchers are familiar with MLM, in which this term is estimated using a random intercept.

statistical identification (Imai & Kim, in press; Kropko & Kubinec, 2020). Therefore, we used Session as a continuous variable in Model 2. All analyzes were conducted using Stata15 (StataCorp, 2017).

To enable the estimation of separate effects for improvements and deteriorations, Allison (2019) started with a first-difference model. The asymmetric fixed-effects model can be estimated by (1) calculating first differences session-by-session for the alliance scores; (2) separating the difference scores into two variables; WAI^+ containing positive changes of the working alliance and WAI^- containing negative changes of the working alliance from the prior session; (3) replacing the missing values for the first session with zero. Finally, 4) WAI^+ and WAI^- are used as independent variables in an OLS regression model with first-differences of the outcome as dependent variable.

However, the first-difference model is not the most common way of estimating the fixed effects model, and its results are not identical to other fixed effects models. Allison (2019) showed that by using cumulative WAI^+ and cumulative WAI^- , i.e., summing up all consecutive positive/negative changes up to each session, and adding these as predictors in a mean-deviation model is mathematically equivalent to using WAI^+ and WAI^- in a model using difference scores. Missing values were generated and replaced with zero in step (3) if there was a missing value for the first session because WAI^+/WAI^- cannot be calculated then. In case of missing data between sessions, we kept the value of the cumulative WAI^+ and cumulative WAI^- missing for this particular session but calculated the cumulative WAI^+ and cumulative WAI^- after that session to keep more information. Two models were tested, as follows:

Model 3

$$OU_{i,t} = \mu + \beta^+ WAI_{i,t-1}^+ + \beta^- WAI_{i,t-1}^- + \alpha_i + \varepsilon_{i,t}$$

Model 4

$$OU_{i,t} = \mu + \beta^+ WAI_{i,t-1}^+ + \beta^- WAI_{i,t-1}^- + \text{Session}_t + \alpha_i + \varepsilon_{i,t}$$

In these models, μ is the model intercept (representing OU when both WAI^+ and WAI^- are zero, that is when there is no change in WAI from the previous session), WAI^+ and WAI^- are (person-mean centered) cumulative difference scores as described above, α_i is the person-mean term, which is estimated using the mean-centering approach¹, and $\varepsilon_{i,t}$ is the model error term specific to each session and assumed to be independent of anything else on the right-hand side of the equation. β^+ represents the change in OU (outcome) for a one-unit increase in WAI^+ , and β^- represents the change in OU for a one-unit decrease in WAI^- .

To test whether β^+ and β^- differ significantly, we can use a Wald z-test, which directly tests the difference in size of coefficients (in this case testing whether $\beta^+ = -1 \times \beta^-$), or a likelihood ratio χ^2 test comparing the fit of the AFE model with the fit of a regular fixed effects model (i.e., Model 1/2). Following Allison (2019), we used the Wald test.

Results

Descriptive Statistics

Table 1 presents the descriptive statistics of WAI^+ and WAI^- session-by-session. Among the total 5060 sessions (including the 1096 first sessions) for the primary care sample, 16.54% of them were deteriorated and 24.62% of them were improved in contrast to the previous session. For the university counseling center sample, 21.84% among the 1305 sessions (including the 291 first sessions) were deteriorated and 30.42% of them were improved in contrast to the previous session. If the first session was excluded, the proportion of alliance deteriorations changes to 21.11% for the primary care

sample (alliance improvements 31.43%), and to 28.10% for the university counseling center sample (alliance improvements 39.15%).

Testing the Effect of Alliance Change on Next-session Symptoms

Table 2 presents the effects of alliance change on next-session symptoms for the ordinary fixed effects models (Models 1 and 2, assuming symmetry of effects). The results showed that the effects of alliance change on next-session treatment outcome were stronger than the effects of alliance change when Session was not included in the model, as has been shown in previous research (e.g., Falkenström et al., 2017). This was the case for both the primary care sample and the university counseling center sample. The coefficients for WAI in these models were all statistically significant at $p < .001$, and in the expected directions. However, the interpretation of the estimated coefficient is limited in the sense that we do not know whether 1) improvements in the alliance are followed by improvements in symptom distress, 2) deteriorations in alliance are followed by deterioration in symptom distress, or 3) both. Usually we interpret results such as these as 3), i.e., that the effect of the alliance on symptoms is symmetric. However, this is an assumption that cannot be tested when using this model.

Testing the Symmetry of the Effects of Alliance Deteriorations and Improvements on Next-session Symptoms

As a next step, we estimated the AFE (Models 3 and 4) instead of the regular fixed effects models. Table 3 presents the effects of alliance deteriorations and improvements on next-session symptoms. The results showed that when Session was not included in the model (Model 3), the effects of alliance improvements on next-session treatment outcome were stronger than the effects of alliance deteriorations. That is, evidence of asymmetric effects was found using Model 3 in both samples (Wald test for primary care sample: -1.27 , $se = 0.36$, $z = -3.56$, $p < .001$, 95% CI $[-1.97, -0.57]$; University

counseling center sample: 0.41, $se = 0.06$, $z = 6.46$, $p < .001$, 95% CI [0.28, 0.53]). However, results did not show such evidence if Session was included in the equation (Model 4). As shown in Table 3, the Wald test of the difference between WAI^+ and WAI^- was statistically non-significant (primary care sample: $p = .96$; University counseling center sample: $p = .35$) once Session number was used as a covariate. Stated differently, the alliance deteriorations and improvements have equal effects on next-session symptoms but in opposite ways, thus the symmetry assumption held after detrending against average linear rate of change in the sample.

Compared with the results shown in Table 2, the results of Model 3 showed that the deterioration of WAI could not significantly predict next-session symptoms (primary care sample: $p = .40$; University counseling center sample: $p = .92$). This cannot be determined from the results of Model 1, i.e., the results of Model 1 could mean that improvements in WAI predict improvements in outcome, or deteriorations in WAI lead to deterioration in outcome, or both. Moreover, the results of Model 4 indicated that *both* improvements and deteriorations in WAI symmetrically predicted next-session symptom levels. Again, this could not have been deducted from Model 2.

Discussion

Psychotherapy mechanisms of change research is to a great extent based on regression models, which yield a single coefficient for the effect of each variable on outcome. An underlying assumption is that effects are symmetric, that is, the effects on outcome of increases in the mechanism are the same (but opposite) as the effects of decreases in the mechanism. Up until recently, there has been no way of testing this assumption, but the development of the AFE model has changed this. Since asymmetric effects have been shown in other social sciences and may be plausible in many psychotherapy research

applications as well, we believe this model is important for psychotherapy researchers to incorporate into their modeling ‘toolbox’.

A recent line of research has shown effects of fluctuations in alliance across sessions on change in symptom distress to the next session, and this result is often interpreted as evidence of rupture-repair processes. However, up to now it has not been possible to separate effects of deteriorations, i.e., downward shifts in the quality of the alliance from one session to the next, from effects of improvements in the alliance from one session to the next. We solved this issue using the AFE and tested the effect of session-by-session alliance deteriorations and improvements on next-session symptom distress. Strengthened by samples from two different cultures as well as two different rater perspectives, we believe the findings highlight some important points in the alliance-outcome literature. At the same time, the findings raise additional questions.

First, our findings showed that increases in raw alliance scores (increases in alliance levels) predicted decreases in symptom scores, just as one would assume based on working alliance theory. However, decreases in alliance levels did not predict increases in symptom scores. This would be in line with the “beneficial context” theory of alliance, since in this framework the absence of an alliance only means that therapy will not work (or be less efficient) but there is no reason to expect a deterioration in alliance to lead to deterioration in symptoms. However, the “relational” theory of alliance would predict alliance ruptures to cause symptom deterioration, which was not the case for this model.

On the other hand, if defined against the sample trend, our findings support symmetric effects of alliance deteriorations and improvements on next-session symptoms. That is, both WAI^+ and WAI^- significantly predicted next-session symptoms in opposite directions (see Table 3), in accordance with

our interpretation of the “relational” theory of alliance (but not the “beneficial context” theory). In this theory, fluctuations in the quality of the psychotherapy relationship can directly cause improvements and deteriorations in the patient’s symptoms. Sessions in which the alliance was especially improved in relation to the previous session, were followed by symptomatic improvement. However, since the average trends for the alliance in these studies were positive, i.e., the alliance generally increased slightly over sessions, the significant effect of WAI¹ on outcome should be interpreted not necessarily as alliance/symptom deteriorations but as less improvement than expected.

Compared to cross-sectional data, panel data comes with some issues that make them more complex, but at the same time opens for new possibilities. The estimation of asymmetric effects is one such possibility, and there are also the possibilities of controlling unobserved time-invariant cofounders as well as unobserved confounders related to time-series trends (Brossart, Willson, Patton, Kivlighan Jr., & Multon, 1998; Curran, Howard, Bainter, Lane, & McGinley, 2014; Falkenström et al., 2017; Wang & Maxwell, 2015). Including Session in the equation, Model 2 and Model 4 control the time-series trend of the data. Results showed that Session number significantly predicted symptom distress in the two samples, which is consistent with the literature (e.g., Baldwin, Berkeljon, Atkins, Olsen, & Nielsen, 2009; Falkenström, Josefsson, Berggren, & Holmqvist, 2016).

Complicating the answer to our research question, this aspect of our results puts more emphasis on the question of whether detrending should be done in the research on psychotherapeutic change (Falkenström et al., 2017). Arguments for and against detrending have been put forth in the methods literature (e.g., Curran & Bauer, 2011; Wang & Maxwell, 2015), but the issue is hard to settle without knowledge of the underlying causal structure of the data (Falkenström et al., 2017), and this is obviously not known, or there would be little need for doing the research. If the trend represents the

effect of some external confounder, detrending should be done or results will be biased. On the other hand, if trends appear in the data due to the causal effect of interest, detrending will be overly conservative. Moreover, detrending can be done in several ways; in our study we used the sample trend, but it is also possible to detrend against person-specific trends (Curran & Bauer, 2010). One possibility is that modern causal methods may solve the issue, because if causality can be ascertained by other means, detrending may not be necessary (Falkenström, Macintosh, & Park, in press).

There are several limitations of this study. First, the trend in Model 2 and Model 4 was simply session number included as a fixed effect in the model. Further studies may consider more complex models that efficiently deal with time-series data, e.g., stationary assumptions, the issue of patient-specific trend/detrending, autoregression etc. (cf. Falkenström et al., 2017). The limitations of the AFE model are the same as the ones for the regular fixed effects model. The primary limitation is the inability to adjust for outcome at the prior session, since outcome at the prior session is a highly likely confounder of the $X_{i,t-1} \rightarrow Y_{i,t}$ regression. Future research should explore ways of testing the asymmetric effects model using Structural Equation Modeling, in which adjusting for the lagged outcome variable is straightforward. Another problem with the fixed effects model is that between-patient differences are not possible to model, since they disappear completely due to the model setup. The remedy for this is, however, simple, we just estimate α_i using a random intercept rather than using person-mean centering (Falkenström et al., 2017). Moreover, fixed effects models cannot estimate bidirectionality, i.e., including the possibility of reverse causality from symptoms to alliance in the same model. Again, this is straightforward using Structural Equation Modeling.

There is also the possibility that other methods would be equally or better suited to studying asymmetric effects. One possibility might be to use Polynomial Regression with Response Surface

Analysis (RSA; Shanock, Baran, Gentry, Pattison, & Heggestad, 2010)². This model has been put forth as a way of overcoming problems with difference score models, which the AFE model is an example of. It is usually applied in studies of congruence (e.g., between two raters), but can be applied also to study the direction of discrepancy of two predictors on an outcome. Although we have not seen any such studies, it should be possible to use $X_{i,t-1}$ and $X_{i,t-2}$ as predictors in an RSA to study asymmetric effects. We recommend for future research to compare this possibility with the AFE. RSA would probably have some advantages to the AFE, although it might be more difficult to integrate with Structural Equation Modeling to enable adjustment for lagged Y.

In addition, since the two samples used were from naturalistic settings, we did not have further information on whether the treatments terminated normally, nor specific information on therapists' intervention or characteristics of the therapists, the diagnosis or other characteristics of the patients, effects of medication, etc. More research should relate the specific condition (treatment approach, characteristics of the patients, diagnosis, etc.), to better understand the effect of alliance improvements and deteriorations on treatment outcomes. Moreover, it is worth noting that the SRS was not validated in the Chinese sample which potentially make the conclusions questionable, though our study showed predictive validity and acceptable internal consistency of the SRS.

Conclusion

The current study introduces the asymmetric fixed-effect model to psychotherapy researchers by applying it to alliance-outcome research. Findings confirm the utility of this novel approach across two cross-national samples in showing that alliance deteriorations and improvements can predict next session symptoms separately at the within-person level.

---

² This possibility was suggested to us during the review process by the Journal Editor.

References

- Allison, P. D. (2019). Asymmetric fixed-effects models for panel data. *Socius: Sociological Research for a Dynamic World*, 5, 1–12. <https://doi.org/10.1177/2378023119826441>
- Baldwin, S. A., Berkeljon, A., Atkins, D. C., Olsen, J. A., & Nielsen, S. L. (2009). Rates of change in naturalistic psychotherapy: Contrasting dose–effect and good-enough level models of change. *Journal of Consulting and Clinical Psychology*. American Psychological Association. <https://doi.org/10.1037/a0015235>
- 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>
- Brossart, D. F., Willson, V. L., Patton, M. J., Kivlighan Jr., D. M., & Multon, K. D. (1998). A time series model of the working alliance: A key process in short-term psychoanalytic counseling. *Psychotherapy: Theory, Research, Practice, Training*, 35(2), 197–205. <https://doi.org/10.1037/h0087645>
- Christian, C., Safran, J. D., & Muran, J. C. (2012). The corrective emotional experience: A relational perspective and critique. In L. G. Castonguay & C. E. Hill (Eds.), *Transformation in psychotherapy: Corrective experiences across cognitive behavioral, humanistic, and psychodynamic approaches* (pp. 51–67). Washington, DC, US: American Psychological Association. <https://doi.org/https://doi.org/10.1037/13747-004>
- Curran, P. J., & Bauer, D. J. (2010). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual Review of Psychology*, 62(1), 583–619. <https://doi.org/10.1146/annurev.psych.093008.100356>
- Curran, P. J., Howard, A. L., Bainter, S. A., Lane, S. T., & McGinley, J. S. (2014). The separation of

between-person and within-person components of individual change over time: a latent curve model with structured residuals. *Journal of Consulting and Clinical Psychology*, 82(5), 879–894. <https://doi.org/10.1037/a0035297>

Duncan, B. L., Miller, S. D., Sparks, J., Claud, D. A., Reynolds, L. R., Brown, J., & Johnson, L. D. (2003). The Session Rating Scale: Preliminary psychometric properties of a “working” alliance measure. *Journal of Brief Therapy*. <https://doi.org/10.1.1.463.8178>

Duncan, B. L., & Reese, R. (2015). The Partners for Change Outcome Management System (PCOMS) revisiting the client’s frame of reference. *Psychotherapy*, 52(4), 391–401. <https://doi.org/10.1037/pst0000026>

Elfström, M. L., Evans, C., Lundgren, J., Johansson, B., Hakeberg, M., & Carlsson, S. G. (2013). Validation of the Swedish version of the Clinical Outcomes in Routine Evaluation Outcome Measure (CORE-OM). *Clinical Psychology & Psychotherapy*, 20(5), 447–455. <https://doi.org/10.1002/cpp.1788>

Eubanks-Carter, C., Gorman, B. S., & Muran, J. C. (2012). Quantitative naturalistic methods for detecting change points in psychotherapy research: An illustration with alliance ruptures. *Psychotherapy Research*, 22(6), 621–637. <https://doi.org/10.1080/10503307.2012.693772>

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

Evans, C., Connell, J., Barkham, M., Margison, F., McGrath, G., Mellor-Clark, J., & Audin, K. (2002). Towards a standardised brief outcome measure: Psychometric properties and utility of the CORE-OM. *British Journal of Psychiatry*, 180(1), 51–60. <https://doi.org/10.1192/bjp.180.1.51>

- Falkenström, F., Finkel, S., Sandell, R., Rubel, J. A., & Holmqvist, R. (2017). Dynamic models of individual change in psychotherapy process research. *Journal of Consulting and Clinical Psychology, 85*(6), 537–549. <https://doi.org/10.1037/ccp0000203>
- Falkenström, F., Granström, F., & Holmqvist, R. (2013). Therapeutic alliance predicts symptomatic improvement session by session. *Journal of Counseling Psychology, 60*(3), 317–328. <https://doi.org/10.1037/a0032258>
- Falkenström, F., Josefsson, A., Berggren, T., & Holmqvist, R. (2016). How much therapy is enough? Comparing dose-effect and good-enough models in two different settings. *Psychotherapy*. <https://doi.org/10.1037/pst0000039>
- Falkenström, F., Macintosh, C., & Park, S. (in press). Using copulas to enable causal inference from non-experimental data: Tutorial and simulation studies. *Psychological Methods*.
- Flückiger, C., Rubel, J., Del Re, A. C., Horvath, A. O., Wampold, B. E., Crits-Christoph, P., ... Ekeblad, A. (2020). The reciprocal relationship between alliance and early treatment symptoms: A two-stage individual participant data meta-analysis. *Journal of Consulting and Clinical Psychology, 88*(9), 829–843. <https://doi.org/10.1037/ccp0000594>
- Hatcher, R. L., Lindqvist, K., & Falkenström, F. (2020). Psychometric evaluation of the Working Alliance Inventory—Therapist version: Current and new short forms. *Psychotherapy Research, 30*(6), 706–717. <https://doi.org/10.1080/10503307.2019.1677964>
- Holmqvist, R., Ström, T., & Foldemo, A. (2014). The effects of psychological treatment in primary care in Sweden—A practice-based study. *Nordic Journal of Psychiatry, 68*(3), 204–212. <https://doi.org/10.3109/08039488.2013.797023>
- Hopko, D. R., Lejuez, C. W., Ruggiero, K. J., & Eifert, G. H. (2003). Contemporary behavioral

- activation treatments for depression: Procedures, principles, and progress. *Clinical Psychology Review*, 23(5), 699–717. [https://doi.org/10.1016/S0272-7358\(03\)00070-9](https://doi.org/10.1016/S0272-7358(03)00070-9)
- Imai, K., & Kim, I. S. (2020). On the use of two-way fixed effects regression models for causal inference with panel data. *Political Analysis*, 1–11. <https://doi.org/10.1017/pan.2020.33>
- Johansson, P., Høglend, P., Ulberg, R., Amlo, S., Marble, A., Bøgwald, K.-P., ... Heyerdahl, O. (2010). The mediating role of insight for long-term improvements in psychodynamic therapy. *Journal of Consulting and Clinical Psychology*. American Psychological Association. <https://doi.org/10.1037/a0019245>
- Kropko, J., & Kubinec, R. (2020). Interpretation and identification of within-unit and cross-sectional variation in panel data models. *PLOS ONE*, 15(4), e0231349. <https://doi.org/10.1371/journal.pone.0231349>
- Marmarosh, C. L. (2017). Fostering engagement during termination: Applying attachment theory and research. *Psychotherapy*, 54(1), 4–9. <https://doi.org/10.1037/pst0000087>
- Mellor-Clark, J., Barkham, M., Connell, J., & Evans, C. (1999). Practice-based evidence and standardized evaluation: Informing the design of the CORE system. *The European Journal of Psychotherapy, Counselling & Health*, 2(3), 357–374. <https://doi.org/10.1080/13642539908400818>
- Miller, S. D., Duncan, B. L., Brown, J., Sparks, J. A., & Claud, D. A. (2003). The Outcome Rating Scale: A preliminary study of the reliability, validity, and feasibility of a brief visual analog measure. *Journal of Brief Therapy*, 2(2), 91–100.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3), 581–592. <https://doi.org/10.2307/2335739>

- Safran, J. D., Crocker, P., McMain, S., & Murray, P. (1990). Therapeutic alliance rupture as a therapy event for empirical investigation. *Psychotherapy: Theory, Research, Practice, Training*, 27(2), 154–165. <https://doi.org/10.1037/0033-3204.27.2.154>
- Shanock, L. R., Baran, B. E., Gentry, W. A., Pattison, S. C., & Heggstad, E. D. (2010). Polynomial regression with response surface analysis: A powerful approach for examining moderation and overcoming limitations of difference scores. *Journal of Business and Psychology*, 25(4), 543–554. <https://doi.org/10.1007/s10869-010-9183-4>
- She, Z., Duncan, B. L., Reese, R. J., Sun, Q., Shi, Y., Jiang, G., ... Clements, A. L. (2018). Client feedback in China: A randomized clinical trial in a college counseling center. *Journal of Counseling Psychology*, 65(6), 727–737. <https://doi.org/10.1037/cou0000300>
- She, Z., Sun, Q., & Jiang, G. (2017). Reliability and validity of Chinese version of Outcome rating scale. *Chinese Journal of Clinical Psychology*, 25(2), 272–275.
- StataCorp. (2017). *Stata Statistical Software: Release 15*. College Station, TX: StataCorp LLC.
- Tracey, T. J., & Kokotovic, A. M. (1989). Factor Structure of the Working Alliance Inventory. *Psychological Assessment*, 1(3), 207–210. <https://doi.org/10.1037/1040-3590.1.3.207>
- Wang, L. (Peggy), & Maxwell, S. E. (2015). On disaggregating between-person and within-person effects with longitudinal data using multilevel models. *Psychological Methods*, 20(1), 63–83. <https://doi.org/10.1037/met0000030>
- Whelen, M. L., Murphy, S. T., & Strunk, D. R. (2021). Reevaluating the alliance–outcome relationship in the early sessions of Cognitive Behavioral Therapy of depression. *Clinical Psychological Science*, 2167702620959352. <https://doi.org/10.1177/2167702620959352>
- York, R., & Light, R. (2017). Directional asymmetry in sociological analyses. *Socius*, 3.

<https://doi.org/10.1177/2378023117697180>

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

<https://doi.org/10.1037/a0040435>

Figure 1. An illustration for deteriorations defined against the mean level and deteriorations defined against the trend, with 0.5-point downward as the criteria for a deterioration. If deterioration is defined against the mean level, then the alliance score at session 3 (a 0.5-point downward dip) indicates a deterioration, while the alliance scores at session 12 and 14 do not indicate deteriorations. But if the same data are defined against the trend, then the alliance scores at session 12 and 14 are deteriorations, while the alliance score at session 3 is not a deterioration. Mean = person-specific mean. Trend = sample trend.

Table 1*Summary of WAI positive and negative changes by session.*

Session	N	Mean	SD	Skewness	Kurtosis	Mean	SD	Skewness	Kurtosis
Primary Care Sample (Therapist rating form)									
		WAI ⁺				WAI ⁻			
2	507	.41	.51	1.51	5.11	.15	.33	2.76	11.35
3	426	.29	.41	1.69	5.78	.16	.33	2.77	11.28
4	309	.24	.39	2.27	8.98	.19	.32	2.37	10.04
5	246	.22	.33	2.20	8.23	.17	.32	2.24	7.89
6	190	.25	.36	1.85	6.79	.11	.25	4.35	31.18
7	150	.18	.29	2.59	12.36	.19	.32	2.19	7.88
8	116	.18	.28	1.97	6.85	.10	.21	2.58	9.98
9	90	.15	.25	2.05	6.82	.14	.25	3.23	16.29
10	69	.19	.27	1.55	4.58	.13	.32	4.66	29.11
11	57	.19	.27	2.38	10.04	.12	.21	2.37	9.59
12	42	.21	.27	1.61	5.58	.13	.25	2.19	7.12
13	27	.19	.21	.89	2.82	.03	.07	2.12	6.27
14	25	.14	.22	2.04	7.05	.11	.18	1.92	5.87
15	19	.14	.17	.82	2.31	.15	.33	3.18	12.72
University Counseling Center Sample (Patient rating form)									
		SRS ⁺				SRS ⁻			
2	169	1.96	3.01	1.99	7.17	1.32	2.47	2.54	10.11
3	139	1.57	2.37	2.08	8.34	.94	1.69	2.15	7.27
4	108	1.22	1.72	2.19	9.95	.98	1.89	2.88	13.02
5	80	1.24	2.06	2.19	7.45	1.15	2.39	2.49	8.56
6	64	1.88	2.86	1.94	6.07	.84	1.79	2.96	12.46
7	56	1.06	2.32	4.57	27.94	.96	1.75	2.46	9.59
8	44	1.53	1.68	1.28	4.18	.39	1.38	5.21	31.27
9	16	.84	1.29	1.44	3.66	1.02	1.89	2.13	6.69

Note. WAI⁺ represents alliance improvements and WAI⁻ represents alliance deteriorations measured by the Working Alliance Inventory-Short Form (WAI-S; Hatcher, Lindqvist, & Falkenström, 2020); SRS⁺ represents alliance improvements and SRS⁻ represents alliance deteriorations measured by the Session Rating Scale (SRS; Duncan et al., 2003).

Table 2

Results of Fixed-Effect Model for the Primary Care Sample and the University Counseling Center Sample.

Primary Care Sample					
Model 1	<i>b</i>	<i>se</i>	<i>t</i>	<i>p</i>	95% CI
WAI	-1.42	.17	-8.54	<.001	-1.74, -1.09
Model 2					
WAI	-.66	.17	-3.94	<.001	-.99, -.33
Session	-.53	.04	-13.39	<.001	-.61, -.45
University Counseling Center Sample					
Model 1	<i>b</i>	<i>se</i>	<i>t</i>	<i>p</i>	95% CI
SRS	.31	.06	5.16	<.001	.19, .43
Model 2					
SRS	.19	.06	3.39	.001	.08, .31
Session	.82	.09	9.08	<.001	.64, .99

Note. CI = confidence interval

Table 3

Results of Asymmetric Fixed-effect Model for the Primary Care Sample and the University Counseling Center Sample.

Primary Care Sample						Wald test				
Model 3	<i>b</i>	<i>se</i>	<i>t</i>	<i>p</i>	95% CI	<i>b</i>	<i>se</i>	<i>z</i>	<i>p</i>	95% CI
WAI+	-1.46	.16	-8.95	<.001	-1.78, -1.14					
WAI-	-.19	.23	-.83	.404	-.64, .26	-1.27	.36	-3.56	<.001	-1.97, -.57
Model 4										
WAI+	-.68	.19	-3.63	<.001	-1.05, -.31					
WAI-	.69	.25	2.78	.006	.20, 1.18					
session	-.34	.04	-8.04	<.001	-.42, -.25	.01	.27	.05	.96	-.51, .54
University Counseling Center Sample						Wald test				
Model 3	<i>b</i>	<i>se</i>	<i>t</i>	<i>p</i>	95% CI	<i>b</i>	<i>se</i>	<i>z</i>	<i>p</i>	95% CI
SRS+	.41	.06	6.63	<.001	.29, .54					
SRS-	-.01	.08	-.11	.91	-.16, .14	.41	.06	6.46	<.001	.28, .53
Model 4										
SRS+	.15	.07	1.99	.05	.002, .29					
SRS-	-.24	.08	-2.93	.004	-.41, -.08					
session	.95	.16	6.10	<.001	.64, 1.25	-.09	.10	-.93	.35	-.29, .11

Note. CI = confidence interval; the Wald z-test examines whether the coefficients of WAI⁺ and WAI⁻ (or SRS⁺ and SRS⁻) differ significantly.