The temporal order of emotional, cognitive, and behavioral gains in daily life during treatment of depression

Snippe, Evelien; Elmer, Timon; Ceulemans, Eva; Smit, Arnout C; Lutz, Wolfgang; Helmich, Marieke A

Published in:
Journal of Consulting and Clinical Psychology

DOI:
10.1037/ccp0000890

The Temporal Order of Emotional, Cognitive, and Behavioral Gains in Daily Life during Treatment of Depression

Evelien Snippe[1], Timon Elmer[2], Eva Ceulemans[3], Arnout C. Smit[1], Wolfgang Lutz[4], Marieke A. Helmich[5]

[1]University of Groningen, University Medical Center Groningen, Department of Psychiatry, Interdisciplinary Center Psychopathology and Emotion regulation, Groningen, The Netherlands

[2]University of Zurich, Department of Psychology, Zurich, Switzerland

[3]KU Leuven, Faculty of Psychology and Educational Sciences, Quantitative Psychology and Individual Differences, Leuven, Belgium

[4]University of Trier, Department of Psychology, Trier, Germany

[5]University of Oslo, Department of Psychology, Oslo, Norway

Author note

The authors are grateful for the time and effort put forth by the participants. We thank Elly Sol, Margriet Brandsma, Abel Busz, René deVries, and Anneke Hol-Steegstra for their help with the data collection and Harriette Riese, Marieke Wichers, and Yoram K. Kunkels for co-designing the TRANS-ID Recovery study. This work was supported by a Swiss National Science Foundation (SNSF) grant (no. PZ00P1_208742) awarded to Timon Elmer, research grants from the Fund for Scientific Research-Flanders (FWO; Project No. G0C9821N) and from the Research Council of KU Leuven (C14/19/054; iBOF/21/090) awarded to E. Ceulemans, and a grant of the European Research Council (ERC) under the European Union's Horizon 2020 research and innovative programme (ERC-CoG-2015; No 681466) awarded to M. Wichers. The analysis plan of the current study was preregistered on OSF (https://osf.io/ch4gr) and a detailed study protocol of the TRANS-ID Recovery study can be found as well on OSF (Helmich et al., 2020a). The data are not stored in a public repository due to restrictions related to the European Law regarding General Data Protection Regulation (GDPR), the sensitivity of the data, and the restrictions in the informed consent. The project data and documentation is stored at a repository of the University Medical Center Groningen, which ensures the security of the data and back-up. Collaboration on TRANS-ID Recovery data is possible and can be requested via h.riese@umcg.nl. The study analysis code will be made available upon request (eveliensnippe@gmail.com). The data on emotions have been previously used in a study by Helmich et al. (2022) to examine early warning signals of transitions in depression. Correspondence concerning this article should be addressed to Evelien Snippe, University Medical Center Groningen, Interdisciplinary Center Psychopathology and Emotion regulation (ICPE), PO Box 30.001, 9700 RB Groningen, the Netherlands. Email: eveliensnippe@gmail.com

Abstract

Objective: Despite the importance for understanding mechanisms of change, little is known about the order of change in daily life emotions, cognitions, and behaviors during treatment of depression. This study examined the within-person temporal order of emotional, cognitive, and behavioral improvements using Ecological Momentary Assessment (EMA) data.

Method: Thirty-two individuals with diagnosed depression completed EMA questions on emotions (sad mood, happy mood), behaviors (social interaction, number of activities), and cognitive variables (worrying, negative self-thoughts) five times a day during a four-month period in which they underwent psychotherapy for depression. Non-parametric change-point analyses were used to determine the timing of gains (i.e., improvements in the mean of each variable) for each individual. We then established whether the first (i.e., earliest) gains in emotions preceded, followed, or occurred in the same week as cognitive and behavioral gains.

Results: Contrary to our hypotheses, first gains in behaviors did not precede first emotional gains (3 times, 8%) more often than they followed them (26 times, 70%). Cognitive gains often occurred in the same week as first emotional gains (43 times, 58%), and less often preceded (13 times; 18%) or followed emotional gains (18 times; 24%). **Conclusion:** The first improvements in behaviors did not tend to precede the first improvements in emotions, likely because fewer behavioral gains were found. The finding that cognitive variables tend to improve around the same time as sad mood may explain why many studies failed to find that cognitive change predicts later change in depressive symptoms.

Keywords: Experience sampling method, Major Depressive Disorder, time-series analysis, temporal precedence, process of change, sudden gains.

Public health significance statement: This study suggests that happy mood, sad mood, and negative ways of thinking often start to improve around the same time during treatment of depression. Fewer improvements were found in activities and interactions, which explains why activities and interactions often improved after individuals started to feel better.

Research has shown that various types of psychotherapy are effective for treating depression (Cuijpers et al., 2008), but less is known about the change processes taking place within individuals during treatment. Most psychological treatments aim to alleviate sad mood and increase positive emotions by targeting both behaviors and cognitions, albeit through different techniques. For example, cognitive therapy (CT) focuses on behavioral activation and challenging unhelpful cognitions to improve a depressed state (Beck et al., 1979), whereas mindfulness-based treatments encourage activation through meditation and yoga practice, and trains individuals to disengage from their thoughts and feelings to reduce rumination (Segal et al., 2002). Thus, different forms of psychotherapy rely on the premise that patients must exhibit some form of behavioral or cognitive change in order to improve a depressed mood.

Yet, few studies have been able to establish that cognitions and behaviors indeed change before an individual's mood improves. Demonstrating temporal precedence, such as showing that cognitive and behavioral changes precede emotional change, is important to rule out reverse causation (Kazdin, 2007; Zilcha-Mano, 2019). When only concurrent changes are studied, we do not know whether change in the emotions was driving change cognitions and behaviors or whether change in cognitions and behaviors was responsible for change in emotions. To infer causality, it must be demonstrated that a cause has happened before its effect (Hill, 1965). Showing that the assumed process variables (e.g., behaviors and cognitions) have changed before the outcome (e.g., emotions) is therefore one of the criteria to validate that a causal process has been identified (Kazdin, 2007), although showing temporal precedence alone is not sufficient for establishing causality.

Several recommendations have been made to demonstrate that change in an assumed process variable temporally precedes change in an outcome (see Falkenstrom et al., 2017; Kazdin, 2007; Zilcha-Mano, 2019). It is recommended to perform multiple assessments of process and outcome variables, for example at a weekly or daily frequency (Kazdin, 2007;

Lemmens et al., 2016; Zilcha-Mano, 2019). With multiple assessments, it can be determined whether change in a process variable precedes the outcome, or if there is evidence for reverse associations with the outcome preceding the process variable (Crits-Christoph & Gibbons, 2021; Falkenstrom et al., 2022). Another recommendation is to use analyses that isolate within-person associations from stable between-person differences (Falkenstrom et al., 2017; Falkenstrom et al., 2022; Zilcha-Mano & Webb, 2021). Focusing on within-person changes is vital as theories on processes of change mostly include hypotheses concerning processes occurring within individuals over time (Hoffart, 2016).

There are two fields of research that have focused on the temporal order of within-person changes using multiple assessment during treatment: research using dynamic models and research on sudden gains (Lutz, 2021). It has been argued that dynamic models applied to longitudinal data, such as multilevel vector autoregressive (VAR) models (Bringmann et al., 2013), dynamic Structural Equation Modeling (SEM) (Asparouhov et al., 2018), or random intercept cross-lagged panel models (Hamaker et al., 2015), are particularly useful for analyzing psychotherapy process data as they allow for the examination of bidirectional time-lagged within-person associations (Falkenstrom et al., 2017; Falkenstrom et al., 2022). Yet, the results of studies using such dynamic models to investigate the within-person temporal relationships between cognitive or behavioral variables and symptoms of depression during treatment vary widely (Fitzpatrick et al., 2020; Kertz et al., 2015; Rubel et al., 2017; van Genugten et al., 2021; Webb et al., 2019).

Moreover, whereas studies using such dynamic models offer insight into micro-dynamics during treatment (Falkenstrom et al., 2017; Falkenstrom et al., 2022), they do not address the question of whether an overall improvement in cognitions and behaviors sets in prior to an overall and persisting improvement in depressed mood. Instead, the described dynamic models include fluctuations around a person's mean level or trend over time and examine if these fluctuations in variables correlate from one assessment to the next. By not investigating change in the mean over time, these models are moving away from the process-research question of how individuals improve over time (Crits-Christoph & Gibbons, 2021). Moreover, dynamic models such as the multilevel VAR model, dynamic SEM models, and the random intercept cross-lagged panel model, have been specifically designed to study stationary processes in which the mean, temporal dependency, and variance are stable over time. These models are therefore less suitable to study non-stationary processes of change (Bringmann et al., 2017), such as symptom improvement over time.

A promising area of research that does focus on large symptom improvements over time instead of symptom fluctuations, is the study of sudden gains (i.e., sudden symptom improvements from one week or treatment session to the next). Various studies have investigated cognitions during the session immediately preceding a sudden gain in depressive symptoms. Some studies have demonstrated that there is increased within-session cognitive change during the pre-gain session compared to a control session (Tang & DeRubeis, 1999; Tang et al., 2005) whereas others did not find this (Andrusyna et al., 2006; Lemmens et al., 2021). Other studies reported no significant difference in cognitions in the pre-gain session compared to a control session (Bohn et al., 2013; Kelly et al., 2007; Kelly et al., 2005). The relation between behavioral change and sudden gains in depression has only been examined in one study, which found a trend towards more within-session behavioral change in the pre-gain session during CT for depression (Lemmens et al., 2021).

Although the sudden gain approach is an elegant method for examining what precedes a major symptom improvement, a key drawback is that sudden gain studies have not assessed the relative timing of gains in behaviors, cognitions, or emotions. As a result, these studies were not able to determine the order of change in these variables, and thereby the directionality. Similarly, many of the studies using dynamic models did not investigate the directionality of the associations and thus did not rule out reverse causation. Yet, untangling the relative timing of gains is relevant, as studies have demonstrated that important changes may occur following gains, e.g., in coping skills (Wucherpfennig et al., 2017) and cognitions (Vincent & Norton, 2019), or may occur in the same week (Bohn et al., 2013; Lemmens et al., 2021; van Genugten et al., 2021).

In addition to the aforementioned drawbacks of these methods, most research so far focused on session-to-session change and has not examined processes occurring in daily life. More fine-grained data, such as collected with Ecological Momentary Assessment (EMA; Delespaul, 1995; Shiffman et al., 2008), may provide more detailed insight into between-session changes occurring within individuals' daily life (Lemmens et al., 2021). The high resolution at which EMA is collected, with multiple assessments throughout the day, may allow the detection of small shifts over time in daily life experiences. Furthermore, most previous studies focused on predicting change in depressive symptoms, which encompasses more than a depressed mood, and may have blurred the findings as different depressive symptoms may remit at different moments in time (Snippe et al., 2021). Considering these methodological challenges, a method is required that permits comparison of the person-specific timing of structural changes in process and outcome variables. It is essential to establish the timing of improvement for each variable and for each individual, as the timing of gains may differ among individuals (Lutz et al., 2013).

In the current paper, we demonstrate how to test assumptions on the order of improvements in emotions, behaviors, and cognitions during therapy using an idiographic method to establish the within-person temporal order of improvement (i.e., change in the mean) in daily life experiences. We show how change-point analyses applied to EMA data gathered over the course of treatment can be used to identify the starting point of an overall improvement, i.e., the first gain. Based on the timing of the first gains in behaviors, cognitions and emotions, their temporal order can be determined.

In this study, we use retrospective change point analyses (see e.g., Page, 1955; Truong et al., 2020). Change point analyses can identify the timing of abrupt distributional changes in time-series data (Cabrieto et al., 2017), such as a change in the mean, and therefore are suitable for determining the timing of sudden jumps or turning points in long-term EMA data. The difference between retrospective change point analyses and prospective models to detect changes in the mean (see e.g. Snippe et al., 2023), such as control charts (Montgomery, 2012), is that control charts are designed for anomaly detection, while retrospective change point analyses are designed to detect segments with different distributional features (Truong et al., 2020). Change point analyses are distinct from most other methods to study incremental change processes, such as (spline-based) growth models and catastrophe cusp models (see Ram & Grimm, 2015), which can reveal the shape of change over time, but cannot determine the timing of changes in the mean. An advantage of change point analyses compared to other time-series models to detect change in the mean such as (hidden-Markov based) regime-switching models (Hamaker et al., 2010) is that change point analyses do not need to specify and fit a full time-series model (Ariens et al., 2020). For our purposes, non-parametric change point analyses are advantageous as they do not rely on any assumptions about the distribution of the data (Hamaker, et al., 2010) as the distribution of EMA data may be non-normal (Haslbeck et al., 2023). In addition, regime-switching models often detect frequent changes between regime states, whereas change-point analysis is more suited to detect longer-lasting changes, with many adjacent time points belonging to the same phase (e.g., Cabrieto et al., 2018).

Although change point analyses have been successfully applied in other fields of research (Chen & Gupta, 2012), and the relevance of change-point analyses for psychotherapy research has been acknowledged (Eubanks-Carter et al., 2012), only a few studies applied change-point detection methods in the field of clinical psychology. For example, studies examined whether change points in cardio-respiratory preceded panic attacks (Rosenfield et al., 2010), whether change points in sleep preceded the recurrence of depression (Minaeva et al., 2024), and whether therapeutic change is gradual or abrupt (Hehlmann et al., 2021; Helmich et al., 2020b). Our study is the first to introduce change-point analyses as a method to gain insight into the temporal order of within-person improvements in daily life experiences during treatment.

The empirical aim of the study is to examine whether behaviors and cognitions start to improve before the core emotions of depression start to improve during treatment for depression. We hypothesize that the first gains in emotions tend to be preceded by gains in behaviors, i.e., number of activities and social interaction. This hypothesis is based on Lewinsohn's (1974) behavioral model of depression which poses that depression can be explained by low levels of behaviors that result in positive reinforcement. Based on this model, it was theorized that engaging more in positively reinforcing behavior, such as social and other activities, may alleviate depression (Dimidjian et al., 2011; Lewinsohn, 1974). Similarly, we hypothesize that first gains in emotions tend to be preceded by gains in cognitions. i.e., worrying and habitual negative thoughts about the self. This hypothesis is based on Beck's et al. (1979) cognitive theory of depression which assumes that maladaptive ways of thinking, such as negative thoughts about the self and the future, may cause depression. Beck and colleagues further postulate that adapting unhelpful patterns of thinking would result in relieving a depressed mood. Therefore, the current study focuses on individuals who received psychotherapeutic treatment with a focus on behavioral, cognitive and emotional change for Major Depressive Disorder (MDD). The study participants completed EMA five times a day over a period of four months during which they underwent treatment. The study focuses on two core emotional features of depression, namely sad mood (feeling down and listless) and the absence of happy mood (feeling cheerful and content).

Methods

Participants

Participants were individuals with depression included in the TRANS-ID Recovery study (Helmich et al., 2020a). Participants were recruited both through online advertisements and at mental health institutions. All participants provided written informed consent. Inclusion criteria of the TRANS-ID Recovery study were: age ≥ 18, Inventory of Depressive Symptomatology (IDS-SR) score ≥ 14, diagnosis of a Major Depressive Disorder assessed with the mini-SCAN (Nienhuis et al., 2010), and being set to start psychotherapy for depression. Exclusion criteria were: presence of a current manic episode or current psychotic symptoms, chronic depressive symptoms (≥ 2 years), reported primary diagnosis of a personality disorder, and inability to work with a smartphone. In total, 59 participants met these criteria. Additional exclusion criteria for the present study were not having completed the study period of four months (n = 15), not having received a minimum of four sessions of psychotherapy during the study period (n = 8), and starting the study more than a week after starting psychotherapy (n = 4). This resulted in a final sample of 32 participants. Table 1 summarizes the participant characteristics. Depression at baseline was moderate to severe for most participants (according to the cut-off scores of the IDS-SR).

Study procedures

A detailed study protocol of the TRANS-ID Recovery study can be found on OSF (Helmich et al., 2020a). All study procedures and the study protocol were approved by the Medical Ethical Committee of the University Medical Center Groningen (reg. number:

NL58848.04.16). Data were collected between June 2017 and May 2020. EMA and questionnaire data was collected by sending an SMS with a link to the secure online questionnaire application Roqua (www.Roqua.nl). RoQua is hosted in data centers from the University of Groningen and complies with the EU General Data Protection Regulation (GDPR). Informed consent was obtained from all participants prior to the baseline interview, during which study procedures were explained and a trial round of EMA was performed. Over a four-month period, participants monitored their daily experiences with EMA on their smartphone. EMA prompts were scheduled at fixed intervals of three hours, with a total of five assessments conducted each day. Participants had thirty minutes to complete an EMA prompt. The full list of EMA questions can be found in the study protocol (Helmich et al., 2020a). Over a six month period, participants completed the Symptom checklist-90 (SCL-90) Depression Scale (Arrindell, 2003; Derogatis & Cleary, 1977) on a weekly basis using their smartphones. Our team contacted the participants regularly to encourage adherence to the EMA questionnaires. Participants were financially rewarded with 50 € for each month that they completed at least 80% of the EMA questionnaires and 50 € for completing the weekly questionnaires. The mean compliance rate to the EMA questionnaires was 86% (with a range of 76%-96%).

Treatment

The study included individuals who started a type of outpatient psychotherapy for depression during the study period as part of treatment as usual. Therapy was not provided as part of the study. The majority of the sample (63%) received CBT for depression as part of their treatment (see Table 1). The remaining participants received eclectic therapy, mindfulness-based treatment, supportive counseling, psychodynamic psychotherapy, systemic therapy, schema therapy, or interpersonal therapy (IPT). The minimum dosage was one session a week. In addition to psychotherapy, 44% of the sample was treated with antidepressants.

EMA Measures

Emotion variables

We focused on momentary emotions that reflect the core symptoms of depression: sad mood and loss of pleasure (i.e., the inverse of a happy mood). Participants were instructed to rate how they felt at the moment of filling out the questionnaire. Sad mood was measured by taking the mean of the items “I feel down” and “I feel listless”. Happy mood was measured by taking the mean of the items “I feel content” and “I feel cheerful”. The items were rated on a Visual Analogue Scale (VAS). The chosen position of the VAS slider was converted into a score ranging from 0 (*not at all*) to 100 (*very much*).

Behavior variables

Measured behavior variables were social interaction and number of activities. Social interaction was measured with the item “In the past three hours I have spoken with others” rated on a VAS scale ranging from 0 (*not at all*) to 100 (*very much*). Participants were instructed to rate the amount of time spent on all the forms of interactions in which they used their voice, such as face-to-face conservations, phone-calling, and video-calling. Number of activities was measured by taking the sum of the performed activities during the past three hours: eating, household tasks or groceries or administration, self-care (e.g., shower, brushing teeth, clothes), working or studying, taking care of (grand)children, sports or walking or cycling, something calm (e.g., watching TV, internet), hobby (e.g. making music), outing (e.g., visit to town, a museum), something with another person/others, contacting someone, be on the way, sleeping. The range of the sum score was 0-12, with higher scores indicating a higher diversity of activities. The activity categories were based on a previous study (Van Der Krieke et al., 2016) and a pilot EMA study with six formerly depressed individuals who were interviewed on the type of activities that increased when they recovered from their depression.

Cognition variables

Measured momentary cognition variables were worrying and habitual self-thoughts, which are two forms of repetitive negative thinking (Mahoney et al., 2012). Worrying was measured momentarily with “I am worrying” rated on a VAS scale ranging from 0 (*not at all*) to 100 (*very much*). Habitual self-thoughts were measured with “my thoughts about myself are” rated on a VAS scale ranging from 0 (*very negative*) to 10 (*very positive*).

Statistical analysis

Change-point analyses

Timing of change in the mean of each of the six EMA variables was assessed for each individual with change-point analysis using the E.divisive function of the R-package *ECP* (James & Matteson, 2014). E-divisive applies a non-parametric change-point analysis directly to the data by dissecting the time series in two at any possible location and optimizing the dissection so that the distribution of scores of the two segments differ maximally from one another (James & Matteson, 2014). Significance of this difference in distribution of scores of before and after the change point is tested using permutation testing. Change points in different distributional features (e.g., change in only the mean or any distributional change including variance, skewness, and the mean) can be tested by tuning the parameter alpha. E-divisive yields a nested solution, meaning that after the determination of the first change point, the two segments are split again using the same procedure to look for additional change points. Using similar settings as in the current study, a simulation study showed that E-divisive is relatively accurate in detecting change of one or two standard deviations (Cabrieto et al., 2017), with more false positives occurring when autocorrelation was high (i.e., .70)[1].

E-divisive was set to detect change points in the mean rather than any distributional change by setting alpha to 2. The minimum number of observations between potential change

---

[1] One key difference between the Cabrieto et al. (2017) study and ours is that they used multivariate change-point analysis, while we analyzed the trajectories of each variable separately (i.e., univariately). Further simulation studies would be needed to determine in what ways multivariate change-point analyses are superior to univariate ones.

points was set to 30 (min.size = 30), which reflects a time period of roughly a week (i.e., mean compliance (0.86) times 35 observations in a week). We did so because change points closer to one another than a week may reflect daily fluctuations rather than a consistent improvement. The number of random permutations to test significance was set to 1000 (R = 1000). Considering that change points were examined in six variables, a Bonferroni-correction was applied, and the significance level was set to $\alpha = .0083$ (.05/6). The E-divisive function of the *ECP* package cannot handle missing data. As mean compliance was high in this sample (86%) we decided not to impute the missing data but instead to run the analyses on the available observations.

Criteria for gains

A change point was considered a gain if it satisfied two criteria. First, the change needed to represent an improvement. Improvements were determined by comparing the direction of the difference in the mean of the observations in the week before the change point and the week after the change point, in line with the criteria for 'sudden gains'(Tang & DeRubeis, 1999). regardless of the size of the difference. Increases in happy mood, social interaction, and number of activities, as well as decreases in sad mood, worry, and negative self-thinking were regarded as improvements. We did not include criteria on the relative and absolute size of the difference between the data in the week before and after the change point, as is done in the sudden gain literature, as the change point analysis already tests whether the mean of the data before the change point significantly differs from the mean after the change point. Post-hoc analyses showed that the mean effect size of the difference between the data in the week before and after the change point were medium to large (see Table 2). Since the change points detected in the data do not necessarily reflect a sudden shift from one week to the next, we call the improvements found in the current study 'gains' instead of 'sudden gains'.

The second criterion was that the change point needed to occur within an overall trajectory of improvement in order to exclude gains occurring in patients who deteriorate over time. An overall trajectory of improvement defined as a significant change point towards improvement when the maximum number of change points was constrained to 1, reflecting the largest change point. As we are interested in the start of an overall improvement, the main analyses focus on comparing the timing of the first (i.e., earliest) gain in each of the variables.

Descriptive statistics

The timing of all gains in each of the EMA variables are visualized in a plot for one exemplary patient. The frequency of gains is described across participants. For individuals who experienced emotional gains, it is described how often they experienced at least one gain in behaviors and cognitions. Additionally, we describe how often first gains in behavioral and cognitive variables occur before, during (i.e., within the same 30-observation time window), or after first gains in the emotion variables.

Inferential statistics

To test our first hypothesis, which posits that gains in behaviors tend to precede gains in emotions rather than follow them, we compared the timing of the first gain in each behavior variable to the timing of the first gain in each emotion variable. We determined which variable improved first in each pair of behaviors and emotions, resulting in four pairs (two behaviors and two emotions). A gain in a behavior variable (B) is considered to precede a gain in an emotion variable (E) when the first gain in B occurred at least one week before the first gain in E (i.e., the sequence BE). The same procedure was used to examine whether a gain in a behavioral variable (B) occurred after a gain in an emotional variable (E; i.e., the sequence EB). The temporal order of a certain pair could thus only be determined for an individual when both variables of the pair showed at least one gain, and when the first gains of the two variables were separated by at least 30 observations (~ one week). We then counted the number of times that the sequence BE and EB occurred across all pairs of behaviors and emotions for all individuals. Using this information, we determined the proportion of times that first behavioral gains occurred preceding first emotional gains (i.e., BE total / (BE total + EB total)). The null hypothesis was that this proportion would not significantly differ from the proportion to be expected if all change points were random (50%). The alternative hypothesis was that first gains in behaviors preceded first gains in emotions more often than 50% (i.e., H_a: BE > .50). To test this, we performed a binomial test in R using the function “binom.test(x = BE, n = (BE + EB), p = 0.5, alternative = "greater")”, where BE is the number of times the sequence BE was observed across all observed variable pairs across all individuals, and EB is the number of times the sequence EB was observed across all observed variable pairs across all individuals. The same procedure was repeated to test our second hypothesis that first gains in cognitions tend to precede gains in emotions.

Pre-registered robustness checks

We conducted two pre-registered robustness checks. First, we investigated whether incorporating a theoretically informed stability criterion affected the results. To determine stability, we adopted Tang and DeRubeis’ (1999) theoretical criteria for sudden gains and defined a gain as stable if it lasted for a period of three weeks. The additional stability criterion therefore entailed that the mean of the observations three weeks after the change point was an improvement compared to the mean of the three weeks before the change point. If there was insufficient data available prior to the initial change point, only the available data was included. Second, we examined whether using a different time-window affected the results, as the choice of excluding first gains that happened within a window of one week may be arbitrary. We repeated the analyses regarding first gains occurring within a window of 4 or 10 days as occurring at the same time.

Sample size calculation

The power to detect change points in the time series of a participant depends on the number of observations per participants rather than the number of individuals. A previous simulation study showed that E-divisive performed well in detecting a change of one or two SD's when applied to time-series of 300 data points that included one change point (Cabrieto et al., 2017). The 32 included participants completed between 461 and 591 EMA observations.

The power to test whether first gains in behaviors precede rather than follow first gains in emotions depends on the number of times the four pairs of behavior and emotion variables show a gain. In the present study, the maximum number of pairs is 128 (32 participants times four pairs). If first gains in behaviors precede first gains in emotions rather than follow in 70% of the pairs, a power analysis indicates a power of 81% to find that this proportion is significantly higher than the chance proportion of 50% when including 40 pairs of variables to perform a one-sided binomial test. The same number of pairs is required to have enough power to find that first gains in cognitions tend to precede first gains in emotions.

Transparency and openness

This study's hypotheses, methodological approach, and analyses were preregistered on the Open Science Framework (https://osf.io/ch4gr). In this article, we report all measures, statistical analyses, and the sample size calculation. Data were analyzed using R, version 4.2.0 (R Core Team, 2022) and the R-package *ECP* (James & Matteson, 2014). For data visualization, we used *ggplot2* (Wickham, 2016). The data are not stored in a public repository due to restrictions related to the European Law regarding General Data Protection Regulation (GDPR), the sensitivity of the data, and the restrictions in the informed consent. The study analysis code will be made available upon request.

Results

Single-case example of sequences of gains

Figure 1 depicts one individual's raw scores (dots), the running mean (non-linear line), and the timing of the identified gains (vertical lines) of the separate EMA variables. Figures of the other participants can be found in the supplement. This individual had multiple gains in sad mood, happy mood, worrying, and self-thoughts and only one gain in number of activities and social interaction. Notable is that gains in cognitions (worrying and self-thoughts) and emotions (sad mood and happy mood) often occurred in the same week. The first gains in emotions and cognitions were observed more than two weeks before the first gain in social interaction and more than four weeks before the first gain in number of activities. Within this individual, the sequences EC (i.e., emotion preceding cognitions) or CE (i.e., cognitions preceding emotion) are not present as the first gains in the cognition variables and emotion variables occurred during the same time span of a week. The sequence EB (i.e., emotion preceding behavior) is observed four times as first gains in both sad mood and happy mood were observed before the first gains in both social interaction and number of activities (i.e., two times two). Notable is that after the gains in behaviors, there were further gains in emotions and cognitions. The timing of some gains clearly indicated a sudden gain from one level to another level (e.g., the first gain in worrying), whereas other gains seem to occur in the middle of a linear trajectory of change (e.g., the third gain in worrying).

Description of gains

Table 2 shows the number of individuals with an overall trajectory of improvement in the EMA variables, the total number of gains per EMA variable, and the number of gains per individual. Sad mood and worrying improved most often over time (see Table 2, Column 1). Happy mood, negative self-thoughts, and number of activities showed an overall trajectory of improvement in about half of the sample whereas social interaction showed an overall trajectory of improvement only in a quarter of the individuals.

Of the 32 participants, 27 individuals (84%) showed an overall trajectory of improvement in at least one of the emotion variables. All these 27 participants (100%) also showed an overall trajectory of improvement in at least one of the cognition variables. Only 19 (70%) out of these 27 individuals showed an overall trajectory of improvement in at least one behavior variable. The total number of gains across individuals was much higher in emotions and cognitions than in behaviors (see Table 2, Column 2). Corresponding with that, the median number of gains per individual was higher in emotions and cognitions than in behaviors (see Table 2, Column 3).

The order of first gains in behavior and emotions

To evaluate our hypothesis of first gains in behaviors preceding first gains in emotions, we compared the order of first gains in all four pairs of behaviors and emotions (i.e., two behaviors times two emotions). Contrary to our pre-registered hypotheses, first gains in behaviors followed first gains in emotions more often (EB, 26 times) than that they preceded first gains in emotions (BE, 3 times, see Table 3). Despite the low number of pairs, a binomial test showed that the proportion of gains in behaviors preceding gains in emotions (BE; 10%; Confidence Interval: 0%-25%) was significantly lower than the proportion of 50% that would be expected if all change points were random ($p < .001$)[2]. When looking more closely at the separate behavior and emotion variables, we also observed that first gains in number of activities and social interaction occurred most often after first gains in sad mood and happy mood (see Table 3).

The order of first gains in cognitions and emotions

To test our hypothesis that first gains in cognitions tend to precede first gains in emotions, we compared the order of first gains in all the four pairs of cognitions and emotions. First gains in cognitions preceded first gains in emotions 13 times (CE) and they

[2] First gains in behaviors that occurred in the same week as first gains in emotions were disregarded in the binomial test.

followed first gains in emotions 18 times (EC). A binomial test showed that the 42% of gains in cognitions preceding gains in emotions did not differ significantly from the proportion of 50% (Confidence Interval: 29% – 100%, p = .86). Moreover, we found in this sample that gains in cognitions occurred most often in the same week as first gains in emotions (43 times, see Table 3). Visual inspection of the timing of gains for each individual revealed that this was often not only the case for first gains, but also for later gains in emotions and cognitions, as is also illustrated in Figure 1. When looking more closely at the separate cognition and emotion variables (see Table 3), it can be observed that especially first gains in sad mood tended to occur often in the same weeks as first gains in cognitions.

Robustness checks

We preregistered two robustness analyses. First, we excluded gains that did not meet the stability criterion of three weeks (see methods). The results were almost identical when gains were excluded that did not persist three weeks (see Supplementary Table 1), which is not surprising as only 2% of the 285 found gains did not meet the stability criterion.

Second, we examined whether the results differed if we regarded gains occurring in a window of 4 days or 10 days as occurring at the same time (see Supplementary Tables 2 and 3). Logically, a time window of 10 days (i.e. 50 beeps) led to a few more gains occurring within the same time window whereas a time window of 4 days (i.e., 20 beeps) led to fewer gains occurring within the same time window. However, the main conclusions of these two additional analyses remained identical to those of the original analyses (see Supplementary Tables 2 and 3).

Post-hoc analyses

Upon inspection of each participant's graphs, we identified seven individuals who displayed a general downward trend not only in sad mood and worrying but also in happy mood and positive self-thoughts. These opposing trends could be attributed to an “initial elevation bias”

(Shrout et al., 2018), resulting in a decrease in all variables. Therefore, we deemed the gains of these individuals less reliable and repeated the analyses excluding the seven participants who displayed the aforementioned downward trends. The results were almost identical to the main findings (see Supplementary Table 4).

Upon further inspection of participants' graphs, we observed that the variance of behavior variables appeared to be higher (see e.g., Figure 1). Post-hoc analyses revealed that the median within-person variance of social interaction (*SD* = 21 on scale ranging from 0 till 100) was indeed slightly higher than that of the other variables measured on the same scale (range *SD* = 12-16 on a scale ranging from 0 till 100). As for the variance of the number of activities, it was difficult to make comparisons as the scale of this variable is different (0-12). We also repeated the analyses using Z-scores, i.e., each variable was standardized within-person. The results of these analyses were identical to the original analyses.

Discussion

Many forms of psychotherapy make assumptions about the order of improvements in emotions, behaviors, and cognitions during therapy for depression. This study demonstrated how to test these assumptions using an idiographic method to establish the within-person temporal order of gains in daily life experiences over the course of treatment. The findings show that change-point analyses applied to longitudinal EMA data can reveal whether cognitions and behaviors start to improve before, after, or at the same time as emotions. We aimed to answer the question whether individuals with MDD indeed start to feel better after they start to think and behave differently during a period that they received psychotherapy for depression. We did not find support for this hypothesis. Fewer change points were detected in behaviors than in emotions, which reduced the chance of finding that behavioral improvements occurred earlier than emotional. This may explain why we found that the earliest behavioral gains preceded the earliest gains in emotions less often than followed them. Whether cognitive gains tended to precede emotional gains more often than followed them could not be disentangled as cognitions and emotions started to improve in the same week more than half of the times. The results were similar when repeating the analyses using Z-scores, when using a time-window of 4 or 10 days instead of 7 days for concurrent gains, when including only those gains that persisted for three weeks, or when excluding participants with an 'initial elevation bias'.

The finding that the process of improvement starts with emotional gains is not in line with our hypotheses, but is consistent with previous research. In a prior study, we found that weekly assessed sad mood and loss of pleasure improved more often before than after symptoms of anxiety, energy, and negative cognitions (Snippe et al., 2021). Other studies have found that, among remitters, sad mood shows the largest early improvement in treatment for depression (Sakurai et al., 2013; Tokuoka et al., 2016). Together, these results suggest that improvements in sad mood and positive mood may mark the start of remission of depression.

The current study showed that the first improvements in emotions were often accompanied by gains in worrying and habitual self-thoughts in the same week. Because of that, we did not have enough power to test whether cognitions and emotions tended to precede or follow one another. Cognitive gains most often occurred in the same week as sad mood, whereas this was less pronounced for happy mood. The findings may either imply that emotions and cognitions structurally improve very quickly after one another (i.e., within seven days) or that they are part of the same underlying state. Appraisal theories view cognitive appraisals and feelings as part of the same emotional episode (Moors et al., 2013; Russell, 2003), in which changes in emotions and cognitive appraisal will immediately feed back to one another. Neurophysiological studies also support the theory that thoughts and feelings are outputs of the same neural network (Salzman & Fusi, 2010). Others have posed a process of an upward spiral, in which cognitive changes are followed by sudden symptom gains, which may again trigger cognitive change (Tang & DeRubeis, 1999). If cognitive processes and feelings are indeed so intertwined and if they indeed affect one another quickly, this may explain why studies failed to find altered cognitions preceding sudden symptom gains (Bohn et al., 2013; Kelly et al., 2007; Kelly et al., 2005). Studies applying time-varying vector autoregressive models (Bringmann et al., 2017) to intensive data of cognitions and emotions, and especially studies incorporating experimental manipulation of cognitions, may reveal how quickly fluctuations in cognitions and emotions follow one another.

Because behavioral gains occurred much less often than gains in emotions, possibly because of the higher variance and smaller sizes of the gains, our study may not have had a fair chance to find support for the hypothesis that behavioral gains tend to precede emotional gains. Because behavioral gains occurred less often, behavioral gains had a larger chance to be observed later than gains in emotions, which we also found in the present study. Increases in longer effortful activities may not have been captured in our measure of number of activities as we do not differentiate how long and effortful the reported activities are. This limitation may explain why we observed a lower number of change points in the number of activities. Future studies may include a measure of the time spent on activities.

Another explanation of the findings is that there may be other factors than behavioral change responsible for emotional gains early in therapy, such as the start of medication (Katz et al., 2004), hope, and emotional processing (Abel et al., 2016). However, we had expected that gains in activities would have preceded or co-occurred with emotional gains more often, since previous research indicated that performance of more activities on a day is associated with higher levels of positive affect and lower levels of negative affect (Mausbach et al., 2008) and that lower end of day depressive symptoms were predicted by more talking with others during the day (Snippe et al., 2016).

In addition to the findings, the current study showcases an idiographic method that may be used as a tool to give personalized insight into which factors mark the start of a process of change during treatment. This approach could aid patients in comprehending their process of remission and identifying the factors that led to improvement. The study findings underline the significance of utilizing an idiographic method as we revealed individual differences in the sequences of gains and in what improved first.

We chose to use the *ECP* R package to detect change points due to its sensitivity in picking up change in the mean (Cabrieto et al., 2017). However, this same sensitivity to change might be a disadvantage if larger changes are of interest. Given the moderate levels of autocorrelation in the emotion and cognition variables, some of the found change points in the emotion and cognitions variables may in fact have been false positives. Another drawback of change point analyses is that they assume that there is a specific moment of change. Yet, visual inspection of the data (see supplements) revealed that some change points occurred in the middle of a gradual process of change, indicating that there is sometimes indeed no specific point of change and that it is difficult to identify the exact beginning of improvement. Consequently, the timing of the gains may not always have been entirely accurate, although the change points will still reflect a mean improvement in the data after the change point compared to the data before the change point. Future research may consider using the other types of change point analyses that do take autocorrelation and gradual change into account, such as the time-varying change point autoregressive model (Albers & Bringmann, 2020). Integrating spline-based and change-point analyses in such models could prove fruitful in further distinguishing between gradual and sudden gains.

In addition to the limitations that were already mentioned, we would like to point out a few other limitations. First, the study lacked sufficient power for the binomial tests because of the relatively low number of behavioral gains and because cognitive gains often occurred in the same week as emotional gains. Therefore, we cannot rule out type II errors. Second, it must be noted that pairs of variables were nested within individuals and therefore the assumption of independence of the binomial tests was violated. However, it is unlikely that these limitations affect the conclusions of the study as the results were not only insignificant, but the temporal order of improvement was found in the hypothesized direction only a minority of the times. Additionally, as we aimed to study the start of an overall trajectory of improvement and therefore only focused on the first gains, there was substantial data reduction. Third, the included VAS scales might have resulted in relatively high measurement error, picking up numerical differences that may not reflect a meaningful choice by the participant. Simultaneously, the high resolution of the VAS scales, and the frequency of our measurements also allowed us to detect smaller changes. Whereas the disparity in granularity between the 0-100 VAS scale and the 0-12 scale used for the behavioral activity item may account for the limited precision in measuring this behavioral item, potentially contributing to the scarcity of identified change points in this domain. Fourth, the mental and practical burden of the study was a reason for dropout for many of the 25% of participants who did not complete the study. However, the sampling schedule of five assessments a day was feasible for the participants who finished the study, as evidenced by the high compliance rates. The high sampling density of five assessments a day was needed to arrive at a sufficient sample size to detect small shifts in the data. A disadvantage of sampling multiple times a day is the increased level of autocorrelation, which may induce false positive change points (Cabrieto et al., 2017). Finally, the participants received varied types of psychological and pharmacological treatments. We therefore cannot determine if the type of therapy caused specific change sequences. The type of treatment may have had an effect on the findings in two ways. It may be that the provided treatments did not focus specifically on increasing the number of activities or the amount of social interaction, explaining why we found fewer change points in behaviors. Second, the fact that a part of the participants received antidepressant treatment in addition to psychotherapy may explain why emotions often improved first, given the quick effects antidepressants and placebo pills may have on depressed mood (Sakurai et al., 2013; Tokuoka et al., 2016). Future research could examine how different types of psychotherapeutic treatments are associated with differences in the order of improvements of emotions, cognitions, and behaviors.

In conclusion, the current study provides a deeper understanding of the process of emotional, cognitive, and behavioral change during psychotherapy for patients with MDD. By utilizing longitudinal EMA data, the study sheds light on the order of improvement occurring within individuals in a naturalistic daily life setting, resulting in ecologically valid results. We introduced an idiographic method that combines change point analysis and EMA to facilitate the identification of a sudden or gradual beginning of an overall improvement. Our study demonstrates the relevance of examining the presence of gains in and the relative order of both the outcome (emotions) and process variables (behaviors and cognitions).

References

Abel, A., Hayes, A. M., Henley, W., & Kuyken, W. (2016). Sudden Gains in Cognitive-Behavior Therapy for Treatment-Resistant Depression: Processes of Change. *Journal of Consulting and Clinical Psychology*, *84*(8), 726-737. https://doi.org/10.1037/ccp0000101

Albers, C. J., & Bringmann, L. F. (2020). Inspecting Gradual and Abrupt Changes in Emotion Dynamics With the Time-Varying Change Point Autoregressive Model. *European Journal of Psychological Assessment*, *36*(3), 492-499. https://doi.org/10.1027/1015-5759/a000589

Andrusyna, T. P., Luborsky, L., Pham, T., & Tang, T. Z. (2006). The mechanisms of sudden gains in Supportive-Expressive therapy for depression. *Psychotherapy Research*, *16*(5), 526-535. https://doi.org/10.1080/10503300600591379

Ariens, S., Ceulemans, E., Adolf, J.K. (2020). Time series analysis of intensive longitudinal data in psychosomatic research: A methodological overview. *Journal of Psychosomatic Research. 137*,110191. https://doi: 10.1016/j.jpsychores.2020.110191

Arrindell, W. A., Ettema, J.H.M. (2003). *SCL-90: Handleidingbij een multidimensionele psychopathologie-indicator* [SCL-90: Manual to a multidimensional psychopathologyindicator] (2nd ed.). Swets & Zeitlinger.

Asparouhov, T., Hamaker, E. L., & Muthen, B. (2018). Dynamic Structural Equation Models. *Structural Equation Modeling-a Multidisciplinary Journal*, *25*(3), 359-388. https://doi.org/10.1080/10705511.2017.1406803

Beck, A. T., Rush, A. J., & Emery, G. (1979). *Cognitive therapy of depression*. Wiley.

Bohn, C., Aderka, I. M., Schreiber, F., Stangier, U., & Hofmann, S. G. (2013). Sudden Gains in Cognitive Therapy and Interpersonal Therapy for Social Anxiety Disorder. *Journal of Consulting and Clinical Psychology*, *81*(1), 177-182. https://doi.org/10.1037/a0031198

Bringmann, L. F., Hamaker, E. L., Vigo, D. E., Aubert, A., Borsboom, D., & Tuerlinckx, F. (2017). Changing Dynamics: Time-Varying Autoregressive Models Using Generalized Additive Modeling. *Psychological Methods*, *22*(3), 409-425. https://doi.org/10.1037/met0000085

Bringmann, L. F., Vissers, N., Wichers, M., Geschwind, N., Kuppens, P., Peeters, F., Borsboom, D., & Tuerlinckx, F. (2013). A Network Approach to Psychopathology: New Insights into Clinical Longitudinal Data. *Plos One*, *8*(4). https://doi.org/10.1371/journal.pone.0060188

Cabrieto, J., Tuerlinckx, F., Kuppens, P., Grassmann, M., & Ceulemans, E. (2017). Detecting correlation changes in multivariate time series: A comparison of four non-parametric change point detection methods. *Behav Res Methods*, *49*(3), 988-1005. https://doi.org/10.3758/s13428-016-0754-9

Cabrieto, J., Adolf, J. K., Tuerlinckx, F., Kuppens, P., & Ceulemans, E. (2018). Detecting long-lived autodependency changes in a multivariate system via change point detection and regime switching models. *Scientific Reports*, *8*, 15637, 1-15. https://doi.org/10.1038/s41598-018-33819-8

Chen, J., & Gupta, A. (2012). *Parametric statistical change point analysis with applications to genetics, medicine and finance* (2nd ed.). Springer. https://doi.org/10.1007/978-0-8176-4801-5

Crits-Christoph, P., & Gibbons, M. B. C. (2021). Psychotherapy process–outcome research: Advances in understanding causal connections. In M. Barkham, W. Lutz, & L. G. Castonguay (Eds.), *Bergin and Garfield's handbook of psychotherapy and behavior change: 50th anniversary edition* (pp. 263-295). John Wiley & Sons.

Cuijpers, P., van Straten, A., Andersson, G., & van Oppen, P. (2008). Psychotherapy for depression in adults: a meta-analysis of comparative outcome studies. *J Consult Clin Psychol*, *76*(6), 909-922. https://doi.org/10.1037/a0013075

Delespaul, P. (1995). *Assessing Schizophrenia in Daily Life* Maastricht.

Derogatis, L. R., & Cleary, P. A. (1977). Confirmation of the dimensional structure of the SCL-90: A study in construct validation. *Journal of Clinical Psychology*, *33*(4), 981-989. https://doi.org/10.1002/1097-4679(197710)33:4<981::Aid-jclp2270330412>3.0.Co;2-0

Dimidjian, S., Barrera, M., Martell, C., Munoz, R. F., & Lewinsohn, P. M. (2011). The Origins and Current Status of Behavioral Activation Treatments for Depression.

Annual Review of Clinical Psychology, *7*, 1-38. https://doi.org/10.1146/annurev-clinpsy-032210-104535

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

Falkenstrom, 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

Falkenstrom, F., Solomonov, N., & Rubel, J. A. (2022). How to model and interpret cross-lagged effects in psychotherapy mechanisms of change research: A comparison of multilevel and structural equation models. *Journal of Consulting and Clinical Psychology*, *90*(5), 446-458. https://doi.org/10.1037/ccp0000727

Fitzpatrick, O. M., Whelen, M. L., Falkenstrom, F., & Strunk, D. R. (2020). Who Benefits the Most From Cognitive Change in Cognitive Therapy of Depression? A Study of Interpersonal Factors. *Journal of Consulting and Clinical Psychology*, *88*(2), 128-136. https://doi.org/10.1037/ccp0000463

Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. (2015). A Critique of the Cross-Lagged Panel Model. *Psychological Methods*, *20*(1), 102-116. https://doi.org/10.1037/a0038889

Hamaker, E. L., Grasman, R. P., & Kamphuis, J. H. (2010). Regime-switching models to study psychological process. In P. C. M. Molenaar & K. M. Newell (Eds.), *Individual pathways of change: Statistical models for analyzing learning and development* (p. 155–168). American Psychological Association.

Haslbeck, J., Ryan, O., & Dablander, F. (2023). Multimodality and Skewness in Emotion Time Series. *Emotion*. https://doi.org/10.1037/emo0001218

Hehlmann, M. I., Schwartz, B., Lutz, T., Penedo, J. M. G., Rubel, J. A., & Lutz, W. (2021). The Use of Digitally Assessed Stress Levels to Model Change Processes in CBT-A Feasibility Study on Seven Case Examples. *Frontiers in Psychiatry*, *12*. https://doi.org/10.3389/fpsyt.2021.613085

Helmich, M. A., Smit, A. C., Bringmann, L. F., Schreuder, M. J., Oldehinkel, A. J., Wichers, M., & Snippe, E. (2022). Detecting Impending Symptom Transitions Using Early-Warning Signals in Individuals Receiving Treatment for Depression. *Clinical Psychological Science*. https://doi.org/10.1177/21677026221137006

Helmich, M. A., Snippe, E., Kunkels, Y. K., Riese, H., Smit, A. C., & Wichers, M. (2020a). Transitions in Depression (TRANS-ID) Recovery: Study protocol for a repeated intensive longitudinal n=1 study design to search for personalized early warning signals of critical transitions towards improvement in depression. *Open Science Framework*. https://doi.org/10.31234/osf.io/fertq

Helmich, M. A., Wichers, M., Olthof, M., Strunk, G., Aas, B., Aichhorn, W., Schiepek, G., & Snippe, E. (2020b). Sudden gains in day-to-day change: Revealing nonlinear patterns of individual improvement in depression. *Journal of Consulting and Clinical Psychology*, *88*(2), 119-127. https://doi.org/10.1037/ccp0000469

Hill, A. B. (1965). Environment and Disease - Association or Causation. *Proceedings of the Royal Society of Medicine-London*, *58*(5), 295-300. https://doi.org/10.1177/003591576505800503

Hoffart, A. (2016). Cognitive Models for Panic Disorder With Agoraphobia: A Study of Disaggregated Within-Person Effects. *Journal of Consulting and Clinical Psychology, 84*(9), 839-844. https://doi.org/10.1037/ccp0000114

James, N. A., & Matteson, D. S. (2014). ecp: An R Package for Nonparametric Multiple Change Point Analysis of Multivariate Data. *Journal of Statistical Software*, *62*(7), 1-25. https://doi.org/10.18637/jss.v062.i07

Katz, M. M., Tekell, J. L., Bowden, C. L., Brannan, S., Houston, J. P., Berman, N., & Frazer, A. (2004). Onset and early behavioral effects of pharmacologically different antidepressants and placebo in depression. *Neuropsychopharmacology*, *29*(3), 566-579. https://doi.org/10.1038/sj.npp.1300341

Kazdin, A. E. (2007). Mediators and mechanisms of change in psychotherapy research. *Annu Rev Clin Psychol*, *3*, 1-27. https://doi.org/10.1146/annurev.clinpsy.3.022806.091432

Kelly, M. A. R., Roberts, J. E., & Bottonari, K. A. (2007). Non-treatment-related sudden gains in depression: The role of self-evaluation. *Behaviour Research and Therapy*, *45*(4), 737-747. https://doi.org/10.1016/j.brat.2006.06.008

Kelly, M. A. R., Roberts, J. E., & Ciesla, J. A. (2005). Sudden gains in cognitive behavioral treatment for depression: when do they occur and do they matter? *Behaviour Research and Therapy*, *43*(6), 703-714. https://doi.org/10.1016/j.brat.2004.06.002

Kertz, S. J., Koran, J., Stevens, K. T., & Bjorgvinsson, T. (2015). Repetitive negative thinking predicts depression and anxiety symptom improvement during brief cognitive behavioral therapy. *Behaviour Research and Therapy*, *68*, 54-63. https://doi.org/10.1016/j.brat.2015.03.006

Lemmens, L. H. J. M., DeRubeis, R. J., Tang, T. Z., Schulte-Strathaus, J. C. C., & Huibers, M. J. H. (2021). Therapy Processes Associated With Sudden Gains in Cognitive Therapy for Depression: Exploring Therapeutic Changes in the Sessions Surrounding the Gains. *Frontiers in Psychiatry*, *12*. https://doi.org/10.3389/fpsyt.2021.576432

Lemmens, L. H. J. M., Muller, V. N. L. S., Arntz, A., & Huibers, M. J. H. (2016). Mechanisms of change in psychotherapy for depression: An empirical update and evaluation of research aimed at identifying psychological mediators. *Clinical Psychology Review*, *50*, 95-107. https://doi.org/10.1016/j.cpr.2016.09.004

Lewinsohn, P. M. (1974). A behavioral approach to depression. In R. J. Friedman & M. M. Katz (Eds.), *The Psychology of Depression: Contemporary Theory and Research* (pp. 157-174). Wiley.

Lutz, W., de Jong, K., Rubel., J., Delgadillo, J. (2021). Measuring, Predicting and Tracking Change in Psychotherapy. In M. Barkham, Lutz, W., Castonguay, L.G. (Ed.), *Bergin and Garfield`s Handbook of Psychotherapy and Behavior Change* (7th ed., pp. 89-133). Wiley.

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

Mahoney, A. E. J., McEvoy, P. M., & Moulds, M. L. (2012). Psychometric properties of the Repetitive Thinking Questionnaire in a clinical sample. *Journal of Anxiety Disorders*, *26*(2), 359-367. https://doi.org/10.1016/j.janxdis.2011.12.003

Mausbach, B. T., Coon, D. W., Patterson, T. L., & Grant, I. (2008). Engagement in Activities Is Associated With Affective Arousal in Alzheimer's Caregivers: A Preliminary Examination of the Temporal Relations Between Activity and Affect. *Behavior Therapy*, *39*(4), 366-374. https://doi.org/10.1016/j.beth.2007.10.002

Minaeva, O., Schat, E., kunkels, Y. K., van Roon, A. M., Ceulemans, E., & Riese, H. (2024). Individual-specific change points in circadian rest-activity rhythm and sleep in patients tapering their antidepressant medication: An actigraphy study. *Scientific Reports, 14,* 855. https://doi.org/10.1038/s41598-023-50960-1

Montgomery, D. C. (2012). *Statistical quality control: A modern introduction* (7 ed.). John Wiley & Sons Inc.

Moors, A., Ellsworth, P. C., Scherer, K. R., & Frijda, N. H. (2013). Appraisal Theories of Emotion: State of the Art and Future Development. *Emotion Review*, *5*(2), 119-124.

Nienhuis, F. J., van de Willige, G., Rijnders, C. A., de Jonge, P., & Wiersma, D. (2010). Validity of a short clinical interview for psychiatric diagnosis: the mini-SCAN. *British Journal of Psychiatry*, *196*(1), 64-68. https://doi.org/10.1192/bjp.bp.109.066563

Page, E. S. (1955). A Test for a Change in a Parameter Occurring at an Unknown Point. *Biometrika*, *42*(3-4), 523-527. https://doi.org/10.2307/2333401

Ram, N., & Grimm, K. J. (2015). Growth curve modeling and longitudinal factor analysis. In W. F. Overton, Molenaar, P.C.M., Lerner, R.M. (Ed.), *Handbook of child psychology and developmental science: Theory and method.* (7th ed., Vol. 1, pp. 758–788). Wiley & Sons. Inc. https://doi.org/10.1002/9781118963418.childpsy120

R Core Team (2022). *R: A language and environment for statistical computing*. R foundation for statistical computing. https://www.R-project.org/

Rosenfield, D., Zhou, E. L., Wilhelm, F. H., Conrad, A., Roth, W. T., & Meuret, A. E. (2010). Change point analysis for longitudinal physiological data: Detection of cardio-respiratory changes preceding panic attacks. *Biological Psychology*, *84*(1), 112-120. https://doi.org/10.1016/j.biopsycho.2010.01.020

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

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*(1), 145-172. https://doi.org/10.1037/0033-295x.110.1.145

Sakurai, H., Uchida, H., Abe, T., Nakajima, S., Suzuki, T., Pollock, B. G., Sato, Y., & Mimura, M. (2013). Trajectories of individual symptoms in remitters versus non-remitters with depression. *Journal of Affective Disorders*, *151*(2), 506-513. https://doi.org/10.1016/j.jad.2013.06.035

Salzman, C. D., & Fusi, S. (2010). Emotion, Cognition, and Mental State Representation in Amygdala and Prefrontal Cortex. *Annual Review of Neuroscience, Vol 33*, *33*, 173-202. https://doi.org/10.1146/annurev.neuro.051508.135256

Segal, Z. V., Williams, J. M. G., & Teasdale, J. D. (2002). *Mindfulness-based cognitive therapy for depression*. The Guilford Press.

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology*, *4*, 1-32. https://doi.org/10.1146/annurev.clinpsy.3.022806.091415

Shrout, P. E., Stadler, G., Lane, S. P., McClure, M. J., Jackson, G. L., Clavel, F. D., Iida, M., Gleason, M. E. J., Xu, J. H., & Bolger, N. (2018). Initial elevation bias in subjective reports. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(1), E15-E23. https://doi.org/10.1073/pnas.1712277115

Snippe, E., Smit, A.C., Kuppens, P., Burger,. H., Ceulemans, E. (2023). Recurrence of Depression Can Be Foreseen by Monitoring Mental States With Statistical Process Control. *Journal of Psychopathology and Clinical Science. 132(*2*)*, 145-155. https://doi.org/10.1037/abn0000812

Snippe, E., Doornbos, B., Schoevers, R. A., Wardenaar, K. J., & Wichers, M. (2021). Individual and common patterns in the order of symptom improvement during outpatient treatment for major depression. *Journal of Affective Disorders*, *290*, 81-88. https://doi.org/10.1016/j.jad.2021.04.097

Snippe, E., Simons, C. J. P., Hartmann, J. A., Menne-Lothmann, C., Kramer, I., Booij, S. H., Viechtbauer, W., Delespaul, P., Myin-Germeys, I., & Wichers, M. (2016). Change in Daily Life Behaviors and Depression: Within-Person and Between-Person Associations. *Health Psychology*, *35*(5), 433-441. https://doi.org/10.1037/hea0000312

Snippe, E., Elmer, T., Smit. A.C., Ceulemans, E., Lutz, W., Helmich, M.A. (2022). Introducing an idiographic way to examine the temporal order of gains in daily life emotions, behaviors and negative thinking during psychological treatment. *Open Science Framework.* https://osf.io/ch4gr

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

Tang, T. Z., DeRubeis, R. J., Beberman, R., & Pham, T. (2005). Cognitive changes, critical sessions, and sudden gains in cognitive-behavioral therapy for depression. *Journal of Consulting and Clinical Psychology*, *73*(1), 168-172. https://doi.org/10.1037/0022-006x.73.1.168

Tokuoka, H., Takahashi, H., Ozeki, A., Kuga, A., Yoshikawa, A., Tsuji, T., & Wohlreich, M. M. (2016). Trajectories of depression symptom improvement and associated predictor analysis: An analysis of duloxetine in double-blind placebo-controlled trials. *Journal of Affective Disorders*, *196*, 171-180. https://doi.org/10.1016/j.jad.2016.02.039

Truong, C., Oudre, L., & Vayatis, N. (2020). Selective review of offline change point detection methods. *Signal Processing*, *167*. https://doi.org/10.1016/j.sigpro.2019.107299

Van Der Krieke, L., Jeronimus, B. F., Blaauw, F. J., Wanders, R. B. K., Emerencia, A. C., Schenk, H. M., De Vos, S., Snippe, E., Wichers, M., Wigman, J. T. W., Bos, E. H., Wardenaar, K. J., & De Jonge, P. (2016). HowNutsAreTheDutch (HoeGekIsNL): A crowdsourcing study of mental symptoms and strengths. *International Journal of Methods in Psychiatric Research*, *25*(2), 123-144. https://doi.org/10.1002/mpr.1495

van Genugten, C. R., Schuurmans, J., Hoogendoorn, A. W., Araya, R., Andersson, G., Banos, R., Botella, C., Cerga Pashoja, A., Cieslak, R., Ebert, D. D., Garcia-Palacios, A., Hazo, J. B., Herrero, R., Holtzmann, J., Kemmeren, L., Kleiboer, A., Krieger, T., Smoktunowicz, E., Titzler, I.,… & Riper, H. (2021). Examining the Theoretical Framework of Behavioral Activation for Major Depressive Disorder: Smartphone-Based Ecological Momentary Assessment Study. *JMIR Mental Health*, *8*(12), e32007. https://doi.org/10.2196/32007

Vincent, L., & Norton, P. J. (2019). Predictors and consequences of sudden gains in transdiagnostic cognitive-behavioural therapy for anxiety disorders. *Cognitive Behaviour Therapy*, *48*(4), 265-284. https://doi.org/10.1080/16506073.2018.1513557

Webb, C. A., Stanton, C. H., Bondy, E., Singleton, P., Pizzagalli, D. A., & Auerbach, R. P. (2019). Cognitive Versus Behavioral Skills in CBT for Depressed Adolescents: Disaggregating Within-Patient Versus Between-Patient Effects on Symptom Change. *Journal of Consulting and Clinical Psychology*, *87*(5), 484-490. https://doi.org/10.1037/ccp0000393

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. Springer. ISBN 978-3-319-24277-4. https://ggplot2.tidyverse.org

Wucherpfennig, F., Rubel, J. A., Hofmann, S. G., & Lutz, W. (2017). Processes of Change After a Sudden Gain and Relation to Treatment Outcome-Evidence for an Upward Spiral. *Journal of Consulting and Clinical Psychology*, *85*(12), 1199-1210. https://doi.org/10.1037/ccp0000263

Zilcha-Mano, S. (2019). Major developments in methods addressing for whom psychotherapy may work and why. *Psychotherapy Research*, *29*(6), 693-708. https://doi.org/10.1080/10503307.2018.1429691

Zilcha-Mano, S., & Webb, C. A. (2021). Disentangling Trait-Like Between-Individual vs. State-Like Within-Individual Effects in Studying the Mechanisms of Change in CBT. *Frontiers in Psychiatry*, *11*. https://doi.org/10.3389/fpsyt.2020.609585

Table 1

Participant characteristics

	% (*n*) or *M* (*SD*)
Age	39.4 (14.5)
Female	84% (27)
In a romantic relationship	69% (22)
Education	
Primary/lower vocational school	12% (4)
Secondary /advanced vocational school	38% (12)
Higher education/ University	50% (16)
Has a paid job	50% (16)
Type of psychotherapy	
Cognitive Behavioral Therapy (CBT)	31% (10)
CBT plus other types of treatment	19% (6)
4-days a week intensive group-program including IPT and CBT	13% (4)
Eclectic therapy	6% (2)
Mindfulness-based treatment	6% (2)
Systemic therapy	6% (2)
Supportive counseling	6% (2)
Psychodynamic psychotherapy	6% (2)
Schema therapy	3% (1)
Interpersonal therapy (IPT)	3% (1)
Used antidepressant medication	44% (14)
IDS-SR at baseline	40.1 (7.6)
SCL-90 depression at baseline	43.0 (8.8)
SCL-90 depression at 4 months	34.4 (13.3)

Note: M = mean, *SD* = standard deviation, EMA = ecological momentary assessment, IDS-SR = Inventory of Depressive Symptomatology-Self Report (range 0-84), SCL-90 depression = mean of 14 items of the Symptom Checklist-90 depression subscale (range 0- 56) instead of the original 16 items. SCL-90 depression at baseline = mean of the pre-treatment baseline assessment and the first weekly assessment of the EMA period. SCL-90 depression at 4 months = mean of the last weekly assessment in the four-month study period and the first post-EMA weekly score.

Table 2

Improvement in EMA variables and number of gains per person per EMA variable

	Individuals improved	Gains across individuals	Gains per individual	Effect size of gain
	n (%)	*Total*	*Median (Range)*	*Mean Cohen's d*
Emotions				
Sad mood	26 (81%)	74	3 (1-7)	-1.41
Happy mood	16 (50%)	47	3 (1-5)	1.17
Cognitions				
Worrying	27 (84%)	68	2 (1-5)	-1.26
Self-thoughts	19 (59%)	57	3 (1-6)	1.09
Behaviors				
Number of activities	18 (56%)	20	1 (1-2)	0.58
Social interaction	8 (25%)	11	1 (1-3)	0.93

Note: n = number. Total n included in the analyses = 32, EMA = ecological momentary assessment. Improved refers to the number of individuals with an overall trajectory of improvement defined as a significant change point towards improvement when the maximum number of change points was constrained to 1, reflecting the largest change point. Effect size of gain = mean across individuals of (mean week after change point – mean week before change point) / ((SD of week before change point + SD of week after change point)/2). The mean lag-1 autocorrelation in the available data of sad mood was .56, of happy mood .51, of social interaction .18, of number of activities .02, of worrying .47, and of self-thoughts .45.

Table 3

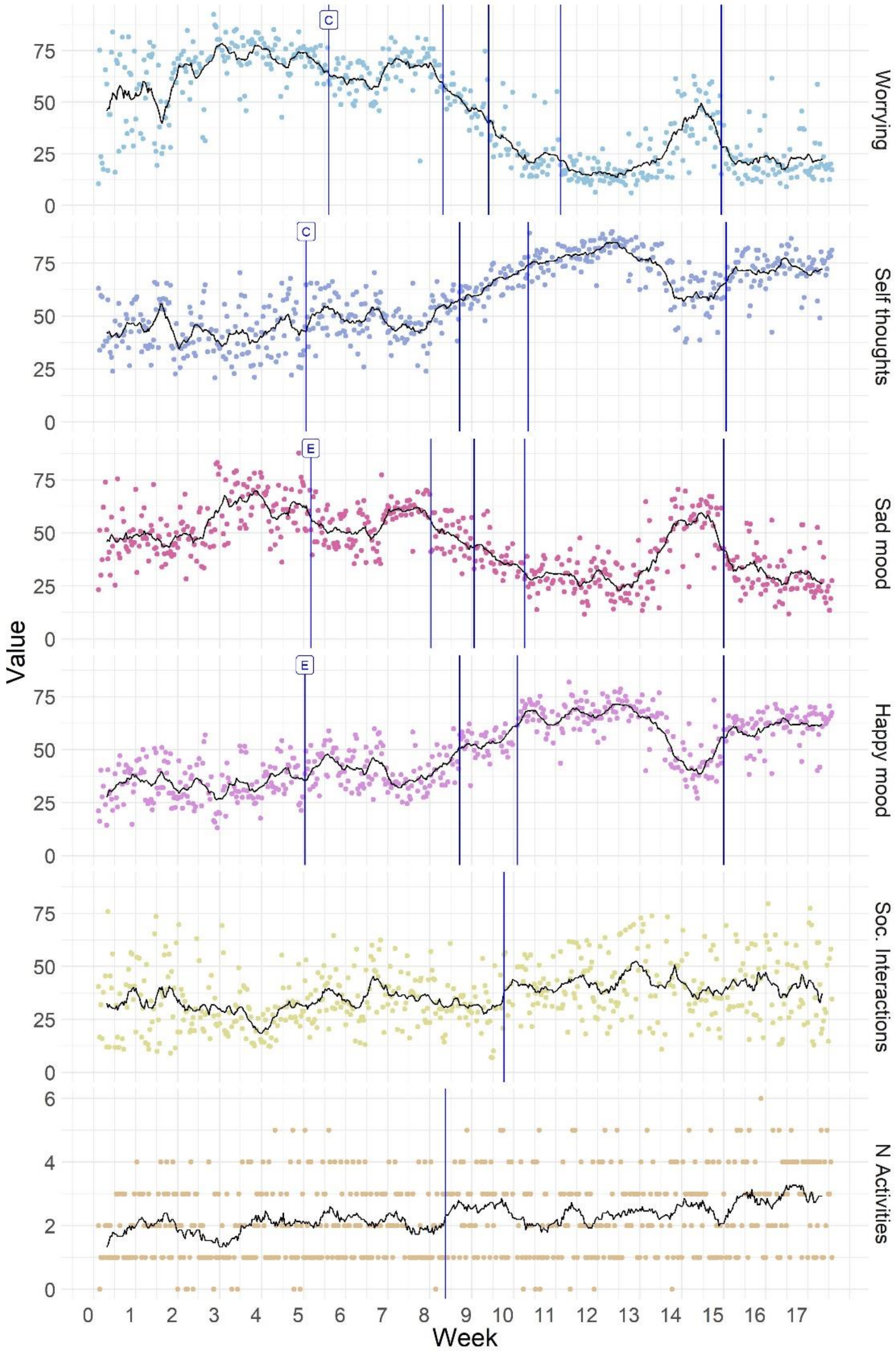
The order of first gains in the separate EMA variables and the behavior, cognition, and emotion categories

		Emotions					
		Sad mood			Happy mood		
		Before	*After*	*Same week*	*Before*	*After*	*Same week*
Behaviors	*n* of activities	1 (7%)	9 (64%)	4 (29%)	0 (0%)	8 (80%)	2 (20%)
	Social interaction	1 (13%)	6 (75%)	1 (13%)	1 (20%)	3 (60%)	1 (20%)
Cognitions	Worrying	5 (19%)	4 (15%)	17 (65%)	3 (19%)	6 (38%)	7 (44%)
	Self-thoughts	3 (18%)	3 (18%)	11 (65%)	2 (13%)	5 (33%)	8 (53%)
		Emotions					
		Before	*After*	*Same week*			
Behaviors		3 (8%)	26 (70%)	8 (22%)			
Cognitions		13 (18%)	18 (24%)	43 (58%)			

Note: n of activities = number of activities, Before = first gains in the variable in the column on the left occurred before first gains in the emotion variables, After = first gains in the variable in the column on the left occurred after first gains in the emotion variables. The lower part of the table presents the totals of all emotion-cognition and behavior-emotion pairs. Same week = first gains occurring within a window of 30 observations, which is equal to ~7 days of data.

Figure 1

Raw scores, running mean, and change points of one example participant



Note: The scales range from 0 (*not at all*) to 100 (*very much*). The scale of self-thoughts (habitual thoughts about the self) ranges from 0 (*very negative*) to 100 (*very positive*). The scale of N Activities represents the number of activities performed since the last beep. Week = time in weeks. Dots represent the raw scores, the black non-linear line represents the running mean (window size = 15), blue vertical lines represent the change point locations. Labels C & E denote that these Cognitions and Emotions respectively improved first.

CRediT author statement:

Evelien Snippe: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing-original draft.

Timon Elmer: Methodology, Software, Formal Analysis, Visualization, Writing-review and editing.

Eva Ceulemans: Methodology, Writing-review and editing.

Arnout C. Smit: Investigation, Methodology, Writing-review and editing.

Wolfgang Lutz: Methodology, Writing-review and editing.

Marieke A. Helmich: Methodology, Data curation, Software, Investigation, Writing-review and editing.

Data transparency

The data reported in this manuscript were collected as part of a larger research study, the TRANS-ID Recovery project (Helmich et al., 2020a). Other findings from this project have been reported in Helmich et al. (2022), a study that focuses on changing dynamics in positive and negative affect, of which four items were also included as the emotion variables in the present study. The study by Helmich et al. (2022) did not focus on identifying change in the mean levels but rather examined whether increasing trends in autocorrelation and variance of positive and negative affect preceded transitions towards lower levels of depressive symptoms.