Detecting impending symptom transitions using early warning signals in individuals receiving treatment for depression

Marieke A. Helmich1*, Arnout C. Smit1, Laura F Bringmann1,2, Marieke J. Schreuder1, Albertine J. Oldehinkel1, Marieke Wichers1, Evelien Snippe1

1 University of Groningen, University Medical Center Groningen, Department of Psychiatry, Interdisciplinary Center Psychopathology and Emotion Regulation, P.O. Box 30.001 (CC72), 9700 RB Groningen, the Netherlands

2 University of Groningen, Faculty of Behavioural and Social Sciences, Department of Psychometrics and Statistics, Grote Kruisstraat 2/1, 9712 TS, Groningen, The Netherlands

*Corresponding author. Email address: m.a.helmich@umcg.nl
Abstract

**Background:** The path to depressive symptom improvement during therapy is often complex, as many individuals experience periods of instability and discontinuous symptom change. If the process of remission follows complex dynamic systems principles, early warning signals (EWS) may precede such depressive symptom transitions.

**Aims:** We aimed to test whether EWS, in the form of rises in lag-1 autocorrelation and variance, occur in momentary affect time series preceding transitions towards lower levels of depressive symptoms during therapy. We also investigated the presence of EWS in patients without symptom transitions.

**Methods:** In a sample of 41 depressed individuals who were starting psychological treatment, positive affect and negative affect (high and low arousal) were measured five times a day using ecological momentary assessments (EMA) for four months (521 observations per individual on average; yielding 25,197 observations in total), and depressive symptoms were assessed weekly over six months. We used a moving window method and time-varying autoregressive generalized additive modeling (TV-AR GAM) to determine whether EWS occurred in these momentary affect measures, within-persons.

**Results:** For the moving-window autocorrelation, 89% of individuals with transitions showed at least one EWS in one of the variables (versus 62.5% in the no-transition group), and the proportion of EWS in the separate variables was consistently higher (~44% across affect measures) than for individuals without transitions (~27%). Rising variance was found for few individuals, both preceding transitions (~11%) and for individuals without a transition (~12%).

**Conclusions:** The process of symptom remission showed critical slowing down in at least part of our sample. Our findings indicate that EWS are not generic across all affect measures and may have limited value as a personalized prediction method.

**Keywords:** complexity; destabilization; symptom remission; sudden improvement; personalized models; idiographic change; replicated single-subject design; experience sampling method; critical slowing down
**Background**

Researchers have started drawing on dynamical systems theory for methods to improve their understanding of the processes of change in psychopathology [1–3]. In part, this renewed interest in dynamical systems is due the rise in studies revealing that changes in psychopathology are often nonlinear [4–10]. In the dynamical systems framework, people are viewed as complex systems of interacting components (e.g., behaviors, cognitions and emotions) in which the dynamics of those components can be studied to anticipate impending changes [11–13]. Particularly, the phenomenon of ‘critical slowing down’ may lead to early warning signals (EWS) in the temporal dynamics of the system variables before a critical transition from one dynamically stable state to another (e.g., from a depressed to non-depressed state). When a system critically slows, it destabilizes and gradually becomes less resilient to external shocks (e.g., daily events), which results in more extreme fluctuations (rising variance) and a longer time to return to the equilibrium position (rising autocorrelation). EWS have been described as generic indicators that a system may be losing resilience and nearing a tipping point where the likelihood of a critical transition is higher [14–17]. EWS are generic in the sense that they are expected to occur in a wide range of complex dynamical systems [16,18], and are beginning to be explored in the field of psychiatry [19–28].

An important reason that EWS have attracted such attention from psychopathology researchers is the promise that they may serve as a clinical tool to detect symptom shifts ahead of time. Descriptively, they fit with clinical observations: the process of change during psychotherapy is frequently characterized by destabilization and nonlinear patterns [29–34], such as symptom shifts between therapy sessions known as sudden gains and sudden losses [8,19,35–38], or rapid early response curves over the first few weeks of treatment [39–43]. These nonlinear patterns are often linked to better outcomes (e.g., in Refs. [8,29,33,37,39,40,44]), so using EWS to improve early detection of destabilization and
potential shifts could provide therapists clues about their patients’ response potential, and allow them to personalize the treatment and timing of interventions accordingly. Thus, examining the promise of EWS is clinically important, as it may give patient and therapist valuable insight into whether the system is sensitive to therapeutic input (is destabilizing), and significant improvement could be imminent [2,34,45,46].

Moreover, studying the occurrence of EWS during therapy is theoretically relevant, as it takes the next step from describing the process of symptom remission from a complex systems angle, to actually testing whether dynamical systems phenomena such as critical slowing down apply to transitions in psychological systems [47]. EWS-like dynamics have shown promise in anticipating shifts in depressive symptoms in a few (group-level) studies [20,25,27,29,48–50], which correlated individuals’ average autocorrelation or variance in mental states to their average symptom severity. However, to truly test the idea that EWS precede symptom transitions in psychopathology, it is imperative to investigate within-person rises in autocorrelation or variance prior to transitions [47]. To that end, Wichers et al. [21] first showed that EWS could be used as a personalized predictor of depressive relapse in a single patient, a result they replicated in another patient recently [22]. In the context of therapy, Olthof et al. [19] identified transitions in time series of daily self-ratings of problem intensity for a large sample of mood disorder patients during treatment. They found that EWS in the form of increased dynamic complexity predicted an increased probability of having a transition in symptom severity in the next four days. These results are very promising, as they show that individually calculated destabilization and within-person defined transitions were positively associated. However, the study by Olthof et al. [19] used a multi-level model to examine the likelihood of transitions based on between-person differences in the level destabilization, and thus does not provide a full within-person test of rising EWS prior to transitions. Therefore,
this study aims to test whether impending transitions toward lower levels of depressive symptoms are consistently preceded by within-person rises in EWS.

Theoretically, we would expect EWS to occur in variables that are central to the system under consideration. A state of depression can be said to be composed of a persistent collection of individual depressive symptoms (e.g., anhedonia, sad mood, lack of energy), which in turn are expressed in shorter-lived feelings [51,52]. In the current study, we therefore look for EWS in the smallest building blocks that contribute to the overall course of a depressive disorder [51–54]: the moment-to-moment fluctuations in affect measured by ecological momentary assessments (EMA). Notably, the presence and strength of EWS may differ between variables [55–58]. Therefore, an important secondary goal in the investigation of EWS in the context of psychiatry is to examine which affective states most reliably anticipate transitions towards reduced symptoms of depression.

Gathering appropriate data is challenging, and only two single-subject studies of EWS prior to depressive relapse have been published [21,22]. To be able to capture and calculate EWS in affect observations before shifts in depressive symptoms, we need frequent momentary assessments over a period in which clinical change is likely. High resolution time series data [24,59–61] that captures moment-to-moment variations at a time scale that is short enough to cover the full range of fluctuations in the state of the system [12,62–65], and over a long enough period to capture the entire state change, is required to detect change in the system dynamics. Therefore, we collected intensive longitudinal data from 41 individuals who were starting treatment for depressive complaints as they entered the study, and were thus likely to show symptom improvement. We gathered four months of 5-times daily EMA affect observations to be able to detect EWS, and six months of weekly depressive symptom assessments for the identification of symptom shifts. As EWS are studied within-system (i.e., within each
individual), these 41 individuals each represent a single-subject study and replicated test of our hypothesis.

The current study was uniquely designed to test whether EWS consistently occur prior to transitions toward depressive symptom improvement, and is the first to test this hypothesis empirically, at the within-person level, for multiple individuals. We aimed to establish whether destabilization in the form of EWS such as increasing trends in lag-1 autocorrelation and variance can be detected in momentary-measured affect before transitions towards improvement in depressive symptoms. We hypothesized that EWS would be most prevalent in variables congruent with the direction of the transition, that is, positive affect variables would be most likely to show changing dynamics leading up to a clinical improvement [21,22,25,66].

To provide the clearest picture of how well EWS work as personalized indicators of imminent change, we will also examine the rates of EWS in individuals without transitions. We hypothesized that individuals with clear symptom shifts would have relatively more and stronger EWS than individuals without transitions.

**Method**

**Sample description**

Individuals with current depressive symptoms who were about to start psychological treatment were recruited for this study. This population was chosen with the expectation that these patients would be likely to show symptom improvements during the study period. To be eligible for participation, individuals needed to be 18 years or older, have current depressive symptoms (a score ≥ 14 on the Inventory of Depressive Symptomatology Self Report (IDS-SR)[67], and be bound for psychological treatment for depression within one month from the (intended) participation start date. Participants also needed to be capable of following the study procedures and operating a smartphone, have sufficient command of the Dutch language, and
be willing and able to give informed consent. Exclusion criteria were chronicity of depressive complaints (persistence ≥ 2 years), current manic or psychotic symptoms, and a primary diagnosis for personality disorder.

The data collection took place between June 2017 and May 2020. The intended sample size was 45-50 participants, to arrive at a sufficient number of patients with a depressive symptom transition who could serve as replications of one another. The final sample consisted of 41 participants who completed the six-month study period (for an overview, see Figure 1). The participants were mostly female \( (n = 35, 85\%) \), with a mean age of 40.1 years \( (SD = 14.4, \text{ min } = 19, \text{ max } = 70) \), at the time of starting the study. Written informed consent was obtained from all subjects. All procedures were approved on December 12, 2016 by the Medical Ethical Committee of the University Medical Center Groningen (reg. number: NL58848.04.16). Detailed procedures for this research project are described in Ref. [68].

**Figure 1**

*Flowchart of the Transitions in Depression (TRANS-ID) Recovery participant inclusion steps.*

- Assessed for eligibility \( (n = 263) \)
  - Excluded \( (n = 204) \)
    - Did not meet inclusion criteria \( (n = 120)^* \)
    - Declined to participate \( (n = 14) \)
    - Unreachable \( (n = 62) \)
    - Exclusion during baseline interview \( (n = 8) \)
- Included \( (n = 59) \)
- Completed the full 4 months of ambulatory assessment and the evaluation interview \( (n = 41) \)
- Drop out \( (n = 18) \)
  - During the ambulatory assessment period
    - Participation is too burdensome or did not fit in daily schedule \( (n = 12) \)
    - No reason specified \( (n = 1) \)
    - Personal circumstances \( (n = 2) \)
    - Decided against sharing data \( (n = 1) \)
    - Unreachable \( (n = 2) \)
- Completed 2 months of weekly follow-up symptom assessments \( (n = 41) \)

*Note: If potential participants did not meet inclusion criteria and did not want to participate, they were counted only as part of the group that did not meet inclusion criteria.
Materials

Ecological momentary assessment (EMA): momentary mood

Over a period of four months, participants filled in a 27-item EMA questionnaire about their current feelings, activities and surroundings (available online, see https://osf.io/a8572/) five times a day at set times, with fixed intervals of three hours. Affect was measured with items like “I feel cheerful”, and “I feel down”, which were answered on a visual analogue scale ranging from “Not at all” to “Very much” (underlying scale ranged from 0 to 100).

With a compliance of 85%, this resulted in \( M = 522 \) (SD = 42, min = 423, max = 591) momentary mood measurements per individual, on which the EWS could be calculated (25,197 EMA observations in total). To that end, individual items were standardized, person-mean centered and averaged to create four variables based on the affect circumplex [69,70]: Positive Affect (PA) high arousal (“cheerful”, “energetic”), PA low arousal (“content”, “at ease”), Negative Affect (NA) high arousal (“restless”, “stressed”, “irritated”), and NA low arousal (“down”, “listless”, “tired”).

Depressive symptoms

Depressive symptoms were measured weekly with fourteen items of the Symptom Checklist-90 (SCL-90) Depression subscale [71–73]. Every weekend, over a period of six months, participants were asked to rate how much in the past week they were bothered by depressive symptoms on a five-point scale from “not at all”, to “extremely”. With a compliance of 98%, this gave \( M = 29 \) symptom observations per person (SD = 1.4, min = 23, max = 30), and 1,211 assessments in total. The full sixteen-item (Dutch) subscale has shown high internal consistency (\( \alpha = .91 \)) and reliability (\( \omega_k = .93 \)), as well as a strong relationship to the general factor psychological distress [71,74]. For this study, two questions on suicidal ideations were removed from the questionnaire to reduce mental burden for participants. The repeated
measurements of the SCL-90 depression subscale were used to identify depressive symptom transitions (see below).

**Psychological treatment**

To optimize the chance of capturing transitions in symptoms during the study period, participation was started approximately three weeks before the start of a psychological treatment aimed at reducing depressive symptoms. Participants self-enrolled into a psychological treatment as part of care as usual, completely independent of the study.

**Analysis**

This analysis has been preregistered, for additional detail see [https://osf.io/xftuq/](https://osf.io/xftuq/). Briefly, we first identified transitions in depressive symptoms (based on weekly symptom data) and then tested (i) whether EWS (based on momentary affect assessments) preceded those transitions and (ii) how often EWS occurred in individuals without transitions. EWS were analyzed for each individual and each affective variable (PA high arousal, PA low arousal, NA high arousal, and NA low arousal) separately using two different methods, namely a moving-window approach (rises in autocorrelation and variance) and Time-Varying Autoregressive Generalized Additive Modeling (rises in autocorrelation). We elaborate on each analytical step below.

**Transition identification**

In line with descriptions of critical transitions, we aimed to identify symptom reductions that appeared relatively abrupt, were reliable, and remained stable for two weeks [15,17,21]. We did not know the time period over which a critical transition toward improvement would take place, but symptom reductions occurring over time periods of a week to a month have
been described as ‘rapid’ in the clinical literature [75,76]. Therefore, we used an adapted version of the Reliable Change Index (RCI) [77], the Duration-adjusted RCI (DaRCI) [78], to examine whether symptoms improved reliably over one to maximum four weeks.

The RCI determines the threshold at which the difference between two points is larger than can be expected from measurement error [77], and is defined as: \( RCI = SE_{\text{diff}} \times Z \), with \( SE_{\text{diff}} = \sqrt{2 \times (SE_m)^2} \). Drawing on psychiatric outpatient norm group scores for the SCL-90 depression subscale [71], we used a standard error of measurement \( SE_m \) of 4.37 for this study\(^1\). With a \( Z \)-score of 1.96 to set a confidence level of 95%, this yielded \( RCI_{95} = 6.18 \times 1.96 = 12.11 \approx 13 \), indicating that a reliable change between two observations needed to be 13 points or larger (rounded up to preserve the confidence level).

Next, to calculate the thresholds for reliable change over two, three and four weeks, we used the DaRCI. Essentially, this method calculates the critical threshold of change from a starting point \( t_{\text{start}} \) to the last observation in that period \( t_n \) while accounting for the longer period between observations. This is done by reducing the RCI to a confidence range around one point, which is then proportionally extended over \( n \) points: \( DaRCI = \left( \frac{Z + SE_{\text{diff}}}{2} \right) \times n \). This method allowed us to examine change that is reliable (maintains the confidence threshold) over various durations (for detail see Ref. [78]). Thus, transitions were identified as reductions of 13, 19, 25 or 31 points over one, two, three or four weeks, respectively [71]. To summarize, if the time between observations is larger, the size of the change must also be larger to be equally reliable.

To ensure that the identified transitions were not just temporary fluctuations in scores, we added a stability criterion based on the RCI\(_{95}\) that ensured that the mean of the two weeks prior to the first point of a transition \( (t_{\text{start}-1}, t_{\text{start}-2}) \), and the two weeks after a transition \( (t_{n+1}, t_{n+2}) \),

---

\(^1\) Note that this \( SE_m \) is based on all sixteen depression subscale items of the SCL-90, while we only used fourteen items in this study. This meant the threshold we calculated is probably more conservative in identifying transitions.
was at least different by 8.6 points. This threshold accounts for the additional power gained from applying it to a change between two means rather than two single assessments by dividing it by the square root of $n$, such that: $\frac{RCl_{95}}{\sqrt{2}} = 8.6$. For each individual, the earliest transition that met the DaRCI threshold and remained stable the two weeks after the transition was identified. This stability criterion was not preregistered, but it was considered prior to the analyses.

Finally, as this method for identifying transitions has not been tested previously, we examined the (Da)RCI at different confidence levels. We took 95% confidence as a default, and additionally explored transitions identified at 90% and 99% confidence.

**Moving window method**

In line with the two studies that examined EWS within-persons prior to recurrence of depressive symptoms [21,22], we used a metric-based moving window approach to examine rises in autocorrelation and variance in the momentary affect measures (cf. [61]). Inside a linearly detrended window of 70 observations (two weeks), with outliers winsorized (i.e., extreme values were set to the 5th and 95th percentile), we calculated the values of the lag-1 autocorrelation and the standard deviation of the affect variables within that window. The window was then moved forward one observation at a time, and the calculation of the autocorrelation and variance repeated, until the end of the time series (no transition group) or until the transition was reached. The Mann-Kendall test (M-K test, coefficient $\tau$) was used to test for a monotonic trend in the entire EWS time series of window estimates of autocorrelation and the variance. To improve the specificity of our test, we maintained a minimally relevant effect size of the trend in the EWS time series of $\tau \geq 0.1$. Specifically, we hypothesized that EWS would be stronger for individuals with compared to individuals without transitions, and therefore, that the $\tau$-coefficients for individuals with a transition would be distributed above zero, and more often have a significant positive trend of $\tau \geq 0.1$ in the EWS time series.
participants without a transition, we would expect the τ-coefficient be distributed around 0, and therefore to show a lower percentage of significant positive trends.

**Sensitivity checks.** We explored the effect of the chosen model settings on our findings by rerunning the analysis with the following parameters altered: window sizes of 35, 105 and 140 observations (instead of 70), global detrending (instead of within-window), not removing outliers (instead of winsorizing). We also checked the effect of using only the M-K significance test at $\alpha < .05$ to detect positive trends in EWS, and explored the results when the dependency between window estimates (due to overlapping data) was corrected for, with the modified M-K test $\alpha < .05$, as proposed by Hamed and Rao [79]. In accordance with our preregistration, the single items “I feel cheerful” and “I feel down” were also analyzed. The results for those variables and the sensitivity checks are reported in the detailed Supplementary Materials.

**Time-Varying Autoregressive Generalized Additive Modeling (TV-AR GAM)**

To examine whether the autocorrelation-coefficient rises before a transition, we explored a recently developed method: TV-AR GAM [80]. This method accounts for non-stationarity in the data by simultaneously modeling the trend (rather than removing it in preprocessing), allowing both the intercept and autoregressive coefficient to vary over time. The TV-AR GAM uses thin-plate splines to fit smooth curves to the time series, and provides $p$-values for the intercept and autocorrelation parameter, and estimated 95% credible intervals for the fitted curve. A time-varying trend in the autocorrelation was indicated for curves with (a) a significant autocorrelation parameter, and (b) where a horizontal line could not be fitted within the credible intervals (i.e., the maximum value of the lower limit > the minimum value of the upper limit). To determine if there was truly an overall rise in the trend, three raters (authors M.A.H., L.F.B., and E.S) then visually inspected the diagnostic plots and the modeled smooth curves and decided by consensus on the direction of change – an increase, decrease, or
a nonlinear trajectory (e.g., quadratic, cubic, or more complex nonlinear trajectories, apart from those that were very convincingly rising or declining).

**Sensitivity checks.** As the time series for the non-transitioning group had a higher maximum number of observations \((M = 616)\) than the transition group, we performed a sensitivity analysis for the non-transitioning group with the time series shortened to the average length of time series in the transition-group: 334, 441, and 458 observations, for transitions identified at 90%, 95% and 99% confidence levels, respectively.

**Results**

**Identified transitions**

For nine out of 41 participants (22%), we identified reliable symptom reductions with the DaRCI at 95% confidence. The duration of these symptom transitions varied, with five transitions occurring over one-week, three transitions over two weeks, and one transition over three weeks. On average, transitions took place after 82 days \((SD = 29, \text{range} = 51–165)\). With the DaRCI set to 90% confidence, fourteen individuals (34%) showed transitions, and at 99% confidence, six cases (15%) remained (for the analysis results for the transition groups at 90% and 99% confidence, see the Supplementary Material, Tables S2 and S3).

**EWS with the moving window method**

**Autocorrelation.** The results of the moving window method indicated that, of the nine individuals with a transition, 8 (89%) showed EWS in the autocorrelation of at least one affective state \((M = 1.78 \text{ EWS per individual}, SD = 1.2)\). By comparison, of the 32 individuals without transitions, 20 (62.5%) showed at least one EWS in autocorrelation \((M = 1.09 \text{ EWS per individual}, SD = 1.17)\).
When investigating EWS in the four variables separately (PA high and low arousal, NA high and low arousal), increases in autocorrelation over time occurred in a higher percentage of individuals with a transition than in individuals without a transition (see Table 1). For individuals with a transition, EWS occurred most often in NA high arousal (55.6%), and least often in PA low arousal (33.3%), and for the no-transition group EWS were most often indicated in NA high and low arousal (31.2% in both), and least often in PA high arousal (18.8%). Contrary to our expectations, PA high arousal did not show a higher number of true positive EWS than the other variables, instead showing the lowest sensitivity.

**Variance.** Of all nine individuals with a transition, 4 (44%) showed EWS in the variance of at least one affective state ($M = 0.44$ EWS per individual, $SD = 0.53$), as did 8 (25%) of the 32 individuals without transitions ($M = 0.47$ EWS per individual, $SD = 0.95$).

The moving window analysis showed lower proportions of EWS in the variance of the affect measures than the autocorrelation (see Table 1). For some variables, the percentage of EWS appeared to be higher in the transition group than in the non-transition groups (e.g., for PA low there were 3 (33%) EWS in the transition group and 2 (6.2%) in the non-transition group). However, averaging across the four variables revealed no discernible difference between the groups, with 11.1% true positive EWS in the transition group and 11.7% false positive EWS in the non-transition group. Again, PA high was not the best performing measure, showing 0% true positive EWS and 15.6% false positives in the variance.

**Distribution of the τ-coefficients.** We expected the Mann-Kendall τ-coefficients to be distributed above zero for individuals with transitions, and around zero in the group without transitions. However, the average τ-coefficients were lower than expected. As can be seen in Table 1, the average τ-coefficients for the autocorrelation of the different affect measures appeared to be distributed only marginally above zero in the transition group, and below zero for the no-transition group. Only the τ-coefficients of the autocorrelations of PA low and NA
Using early warning signals to detect depressive improvement

high had average values above the $\tau \geq 0.1$ threshold, and showed mostly slightly negative values in the group without transitions. For the variance, the $\tau$-coefficients of all affect measures seemed to be more strongly negative on average, with similar values in both groups (for the distributions of the $\tau$-values, see Figures S1a and S1b in the Supplementary Materials).

Table 1
Results of the moving window and TV-AR GAM analyses: average $\tau$-values and proportion of early warning signals in the autocorrelation and variance for individuals with and without transitions.

<table>
<thead>
<tr>
<th></th>
<th>Autocorrelation</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TV-AR GAM</td>
<td>Moving Window method</td>
</tr>
<tr>
<td></td>
<td>EWS % (N)</td>
<td>EWS % (N)</td>
</tr>
<tr>
<td>Transition (N = 9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PA high</td>
<td>0% (0)</td>
<td>33.3% (3)</td>
</tr>
<tr>
<td>PA low</td>
<td>11.1% (1)</td>
<td>44.4% (4)</td>
</tr>
<tr>
<td>NA high</td>
<td>11.1% (1)</td>
<td>55.6% (5)</td>
</tr>
<tr>
<td>NA low</td>
<td>0% (0)</td>
<td>44.4% (4)</td>
</tr>
<tr>
<td>Overall mean</td>
<td>5.6%</td>
<td>44.4%</td>
</tr>
<tr>
<td>No transition (N = 32)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PA high</td>
<td>3.1% (1)</td>
<td>18.8% (6)</td>
</tr>
<tr>
<td>PA low</td>
<td>3.1% (1)</td>
<td>28.1% (9)</td>
</tr>
<tr>
<td>NA high</td>
<td>3.1% (1)</td>
<td>31.2% (10)</td>
</tr>
<tr>
<td>NA low</td>
<td>3.1% (1)</td>
<td>31.2% (10)</td>
</tr>
<tr>
<td>Overall mean</td>
<td>3.1%</td>
<td>27.3%</td>
</tr>
</tbody>
</table>

Note. EWS = early warning signals; PA = Positive Affect; NA = Negative Affect; high = high arousal; low = low arousal; TV-AR GAM = time-varying autoregressive generalized additive model; SD = standard deviation; $\tau$ = coefficient of the Mann-Kendall trend test.

The results for the TV-AR GAM represent cases where the modeled smooth curve was significantly increasing. The results for the Moving Window method are based on the analysis of transitions at 95% confidence, where individual time series were winsorized, locally detrended within windows of 70 observations, and the number of early warning signals based on cases with $\tau \geq 0.1$.

**Sensitivity checks.** For the autocorrelation, the finding that proportionally more EWS were found in the transition group was robust across all the altered model settings for the moving window analysis. The changes in the number of EWS for the autocorrelation found with these altered model settings were similar for people with a transition (difference in $N$ EWS compared to the main analysis: Median = 0, first quartile (Q1) = -1, third quartile (Q3) = 0.75) and for people without a transition (Median = 0, Q1 = -1, Q3 = 1). Increasing the window size
to 105 or 140 observations yielded an improvement of the accuracy for this sample, at 50% and 58% true positives and 33 and 31% false positives, respectively.

For the variance, the lack of a proportional difference in EWS between the transition and non-transition group was consistent across the sensitivity checks. For the altered analysis settings, similar changes in the number of indicated EWS were found for the transition group (difference in $N$ EWS compared to the main analysis: Median = 0, $Q1 = -0.75$, $Q3 = 0$), and the people with no transition ($Median = 0, Q1 = 0, Q3 = 1$; see also Tables S1a and S1b in the Supplementary Materials). None of the altered model settings for the variance yielded notably improved accuracy.

**EWS with the TV-AR GAM**

The trajectories fitted by the TV-AR GAM indicated that, across the four affect variables, the proportion of significant increases in the autocorrelation was comparable between the transition group ($M = 5.6\%$) and the non-transition group ($M = 3.1\%$; see Table 1). Significant decreases and nonlinear trajectories were indicated more often than increases in both groups, with respective means of 8.3% and 27.8% in the transition group, and $M = 9.4\%$ and 25.8% in the non-transition group.

**Sensitivity checks.** Reducing the time series length for the non-transitioning individuals resulted in fewer cases with a significantly time-varying smooth curve (approximately 30% compared to 40% at full length), although the proportion of significant increases in autocorrelation remained similar (2.3%, compared to 3.1% at full length). With transitions identified at 90% or 99% confidence, the number of true-positive significant rises in autocorrelation was comparable to the results at 95% confidence (although at 99% confidence, no EWS remained in the transition group), as did the number of false positive EWS (see Table S3 in the Supplementary Materials).
Exploratory findings

In addition to answering the main research question, we also aimed to investigate ways to potentially improve the balance of true and false positives of our results. We explored whether we could improve the sensitivity of the moving window analysis by separating participants according to a minimum cumulative number of EWS. Combining the results of the variance and autocorrelation for PA high, PA low, NA high and NA low showed that selecting individuals with a minimum of two EWS (out of a possible eight) improved the sensitivity, with 77.8% true positives compared to 44.4% in the main analysis, but also resulted in an increase in false positives from 27.3% in the main results, to 40.6% (for more detail, see Figures S2 and S3 in the Supplementary Materials).

Moreover, we explored different minimally relevant effect sizes by varying the threshold $\tau$-value for the moving window analysis between 0 and 0.5. Our results did not reveal a clear cut-off value that improved the accuracy for all variables (see Figure S4 in the Supplementary Material). Overall, lowering or raising the $\tau$-threshold led to a change the number of true positives that was approximately matched by the change in false positives.

Discussion

This is the first paper to provide an empirical investigation of whether EWS in the form of increasing trends in lag-1 autocorrelation and variance could be detected in individual affect time series in the period before a transition toward depressive symptom improvement. We examined the prevalence of EWS prior to the symptom transitions identified for nine individuals, as well as the rate of false positive EWS in the time series of 32 individuals without transitions. We found that rising autocorrelation was present in at least one of the affect measures for eight out of nine (89%) individuals with a transition, and for twenty out of the 32 (62.5%) individuals without transitions. When examining the affect measures separately in the
moving window analysis, we found that EWS in the form of rising autocorrelation were present in proportionally more individuals with a transition (~44%) than without a transition (~27%). A rise in the variance was found in at least one variable for four out of nine (44%) individuals with a transition, and for eight of the 32 (25%) individuals without transitions. The proportion of EWS in the variance for the separate affect measures showed similar rates for individuals with a transition (~11%) and without a transition (~12%) and thus did not show the same differentiation between the groups. Furthermore, the TV-AR GAM analysis indicated at most one individual with a significant increase in autocorrelation for each of the four affect measures, in both the group with transitions (~5.6%) and in the no-transition group (~3.1%). In our extensive sensitivity checks, the findings for our analyses were robust. This strengthens our conclusion that the moving-window detected autocorrelation results were in the hypothesized direction, and the process of remission from depression showed signs of critical slowing down in part of the individuals in our sample.

Looking at the broad pattern of our results, our findings correspond reasonably well to earlier studies of EWS in depression [19,21,22]. Like the single-subject studies by Wichers et al. [21,22], our results showed that within-person rises in autocorrelation occurred for some of the participants before a transitions in depressive symptoms – this also occurred relatively more often for people with a transition than for people with no transition. Unlike those studies by Wichers et al. [21,22], however, we did not find rising variance as an EWS. Furthermore, our finding that EWS occurred before symptom transitions during therapy is in line with the findings by Olthof et al. [19], who found that critical fluctuations had a higher probability of occurring in the four days before a sudden gain or loss. However, critical fluctuations are a measure of the spread of scores, like the variance, which did not function as an EWS in our study. The difference in findings may be explained by the fact that we looked at a different population than Wichers et al. [21,22](individuals showing depressive symptom remission,
rather than recurrence), and a much longer time period than Olthof et al. [19] for the rising fluctuations (over weeks or months, as opposed to days). Yet, taking in the broader evidence of all these studies, including the current study: significant destabilization in the form of rising autocorrelation could be identified prior to transitions in symptoms in the affect measures of depressed patients. In our sample, we found signs of critical slowing down, i.e., a rise in autocorrelation in at least one of the affect measures, for eight of the nine individual systems, which was in line with our hypothesis. This suggests that for those individuals the process of depressive remission showed the nonlinearity one would expect from a dynamic system, particularly in the context of therapy [2,30,33,34,50,81].

Although almost all individuals with transitions showed EWS in at least one affect measure, looking at the separate affect measures presents a less consistent picture of critical slowing down in our study. Each variable showed EWS for only a few people in the transition group, and, contrary to what we hypothesized, PA high arousal did not turn out to be the most sensitive affect measure to predict of symptom improvements. No single affect variable showed EWS for more than five people with transitions. Thus, EWS were not ‘generic’ indicators of impending symptom change when considering the variables separately. Theoretical works in ecology have found that variables can be ‘silent’ and lack EWS when they do not change in the same direction as the destabilization of the main system [55–58]. That is, perhaps the heterogeneity of which variables showed EWS was due to the fact that the destabilization occurred in different depressive symptoms for different people. By that reasoning, PA high arousal may have shown fewer EWS if destabilization occurred more strongly in depressive symptoms that do not correlate highly with positive experiences (cheerfulness, feeling energetic)(cf. [20,66,82]). False negatives may also have occurred due to the nature of transitions in our study. Based on theory, critical slowing down and its corresponding EWS would be expected to occur primarily before critical transitions or shifts between states in
systems that are dynamically stable around a single point and are slowly losing resilience [15,17]. It is possible that the transitions we detected did not follow this pattern, for example, because the states prior to and after the transition were not sufficiently stable or because the transition was not preceded by a gradual accumulation of instability [83]. The latter may happen if a transition was triggered by a large push (e.g., a psychologically impactful event) rather than gradual destabilization of the system [62,84–86]. Thus, it could be that only a subset of individuals in our sample met the theoretical requirements for detecting EWS, or that only some of the affect measurements picked up on the changing dynamics, which may have contributed to a lower sensitivity of EWS.

Furthermore, there were six to ten individuals without transitions who showed false positive EWS in our study. A potential explanation is that those individuals without transitions still experienced a destabilizing influence from entering therapy, and thus also showed changes in the temporal dynamics of their affect during the observation period, even if this did not precede a clear discontinuous symptom transition (or end in a change, at all). This interpretation is supported by the fact that, when we lowered our DaRCI confidence level to 90%, the number of true positive EWS increased (from 3–5, to 5–7 EWS per affect measure), while the false positives decreased (from 6–10, to 5–9 EWS per affect measure; see Supplementary Table S2). This suggests that some false positive EWS in the main analysis were likely due to individuals who did experience a relevant change, but it did not meet the transition criteria at the 95% confidence level. This is key to interpreting our findings, as our transition criterion may have been too conservative and selected only some of the relevant transitions, while categorizing others as belonging to the no-transition group, thereby potentially inflating the rate of false positives, and underestimating the number of true positives. Further explanations are derived from other fields, where critical slowing down has also been shown to occur before gradual transitions [18,87], as well as in periods with no significant system change [85,87]. However,
most of these EWS studies have been conducted in an ecological context, with very different data, and it is hard to tell which of the processes that influence the finding of EWS in other fields translate to our ability to detect EWS in psychological processes.

The accuracy of our moving-window detected autocorrelation analyses was not high enough to warrant any claims about clinical utility at this stage. Our exploratory analyses indicated that there was room to improve the method’s performance, as we found improved accuracy for larger window sizes and when looking at the cumulative number of EWS within individuals (a minimum number of EWS threshold also worked well in two studies of EWS before depressive symptom recurrence, see Refs. [22,88]). Moreover, single predictors are rarely 100% accurate in isolation, and EWS may provide complementary information to early response indicators such as severity of depression [89–91]. Gaining a better understanding of which participants are likely to show detectable EWS could also help improve the sensitivity of this method.

The results for the variance indicated no proportional differences in EWS between individuals with and without transitions. In fact, most people in our study (with and without transitions) showed a negative trend in the variance over time. A clinical explanation for these downward trends in variance relates to our study population: given that our participants were entering treatment for depressive symptoms, they may have shown higher instability at the start of the observation period due to the destabilizing influence of entering therapy (and because high levels of variability are associated with higher levels of depression [92–94]). As they improved under the influence of therapy, their emotional responses may have become less volatile, and thus an overall downward trend in variance may have emerged in their time series. Another, practical reason for the declines in the variance is that participants may have gotten more habituated to filling out the EMA questionnaires, and thus used less and less of the available scale. Finally, previous EWS studies have also indicated that the variance may show
decreases when less data is available and is a less robust as an indicator of impending change compared to autocorrelation [22,25,95].

Even though the autocorrelation worked reasonably well as an EWS with the moving window method, the more conservative TV-AR GAM did not show the same results. Possibly, this approach was not optimally suited for the purposes of testing for an overall rise in autocorrelation. Instead, the TV-AR GAM results highlighted the high degree of nonlinearity in the autocorrelation of affect time series in both groups. While the M-K test may detect an overall significant trend over time even if the rise in autocorrelation or variance starts later in the time series, the GAM method is aimed at testing the overall shape of the data, and may instead identify similarly late rises as a curving, nonlinear pattern. We do not know over what time period a rise in EWS would be expected in our sample, but a method that allows significant rises to be detected over shorter periods (e.g., a few weeks prior to a transition), may provide an improved rate of EWS detection.

Another reason the TV-AR GAM may have resulted in lower numbers of EWS, is that this method requires a large amount of data to fit credible intervals that were narrow enough to meet our strict change criterion (the maximum value of the lower bound being larger than the lowest value of the upper bound). Furthermore, the TV-AR GAM is generally conservative and biased to detect zero change if little data is available (cf. Refs. [80,96,97]). While determining the direction of change through visual inspection, we noted that many individuals who showed linear change had confidence bounds that were too wide to identify an increase with certainty. The effect of (in)sufficient data on the TV-AR GAM findings was evident from the sensitivity check in the no-transition group: when the time series were shortened (e.g., from, on average, 616 to 411 observations in the main analysis), fewer significant smooth curves were found, and particularly the number of decreases were reduced. It thus seems that the available time series length may have played a limiting role in our ability to detect EWS with this method.
Because there were too few individuals with a transition to reliably test whether the proportional differences between the group with a transition and the group without a transition were significant, it is difficult to determine to what extent these findings generalize beyond our sample. However, this study was foremost a proof of principle to see if EWS could be found in individual time series of people receiving treatment for depressive symptoms. Our study was designed with the specific intent to detect EWS in individual time series, with several months intensive longitudinal within-person affect and symptom measurements (totaling 26,408 observations) gathered during a period where symptom improvement was likely. Given that our population differed from the previous studies of within-person EWS that looked at relapse in depression [21,22], and taking into account the new insights from recent studies [20,23,24], we preregistered and conducted a very thorough analysis with a range of sensitivity checks. Because of this, we were able to discern the broad pattern of how EWS occurred in our dataset and can more confidently state that our findings were robust.

Conclusions

EWS in the form of a rise in the autocorrelation in NA high arousal, NA low arousal, PA high arousal, and PA low arousal preceded symptom transitions toward depressive symptom improvement in part of the participants, and occurred more often for individuals with transitions than for individuals without transitions. Almost all individuals (89%) with a transition had EWS in autocorrelation for at least one the affect measures, compared to about two-thirds of the participants without transitions. EWS based on the variance of affect did not systematically precede symptom transitions. It is too early to claim any clinical utility of EWS in the context of psychotherapy. However, this study provides much-anticipated empirical evidence that the process of symptom remission can follow dynamical systems principles, beyond a post hoc explanation of nonlinear patterns in psychopathology.
Declarations

Preprint link
This manuscript has been uploaded as a preprint to PsyArXiv: https://psyarxiv.com/vf86s/

List of abbreviations
DaRCI: Duration-adjusted Reliable Change Index
EMA: Ecological Momentary Assessment
EWS: Early warning signals
M-K test: Mann-Kendall test
NA: Negative Affect
PA: Positive Affect
SCL-90: Symptom Checklist-90
TV-AR GAM: Time-Varying Autoregressive Generalized Additive Model

Ethics approval and consent to participate
Written informed consent was obtained from all subjects. All procedures were approved on December 12, 2016 by the Medical Ethical Committee of the University Medical Center Groningen (reg. number: NL58848.04.16).

Availability of data and materials
The datasets analyzed in the current study are not publicly available due to the sensitive nature of the data (a small sample of patient data with many observations within patients that cannot be fully anonymized), but are available from the corresponding author on reasonable request.

Funding
This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovative program (ERC-CoG-2015; No 681466 to M. Wichers) and the Netherlands Organisation for Health Research and Development (ZonMw Off Road; project 451001029 to E. Snippe). The funding sources were not involved in the preparation of this manuscript.

Author Contributions

Acknowledgements
We would like to thank Elly Sol, Margriet Brandsma, Abel Busz, René de Vries and Anneke Hol-Stieegstra for their invaluable help with the data collection, and all the mental health professionals and clinics for their assistance with participant recruitment. We also wish to acknowledge the insightful input from members of the larger TRANS-ID team, including Yoram K. Kunkels, Robin N. Groen, Harriette Riese, Sandip V. George and Fionneke M. Bos. Finally, we are especially thankful to all who individuals participated in the study and contributed their data.
References


Supplementary Material to
‘Detecting impending symptom transitions using early warning signals in individuals receiving treatment for depression’

Marieke A. Helmich, Arnout C. Smit, Laura F. Bringmann, Marieke J. Schreuder, Albertine J. Oldehinkel, Marieke Wichers, Evelien Snippe

0. Brief introduction

In this Supplementary Material, we show various additional plots and tables to provide further insight into the effects of different model settings (sensitivity checks) on our results, using different confidence levels for the transition identification, and the findings of our further pre-registered exploratory work.

With the low number of transition cases, it is prudent to interpret the proportion of EWS in the groups and for individual variables with caution, and we recommend looking mainly at overall patterns.

For the comparison of the analyses at different confidence levels, it is also worth noting that there is relevant overlap between the groups at different thresholds, which may affect the number of false positive and false negative EWS. For those who change transition status, the analysis is conducted as if they have no transition – over their entire time series.

1. Moving Window method

1.1 Main analysis

The main moving window analysis was conducted with transitions identified at 95% confidence level (CL), a minimally relevant effect size of $\tau \geq .1$, a window size of 70 observations, outliers winsorized and within-window detrending.
Figure S1a

Density plot of the $\tau$-values found in the moving window analysis of the autocorrelation.

Note. PA = Positive Affect; NA = Negative Affect; high = high arousal; low = low arousal.

Note that group sizes were unequal: $N = 9$ individuals with a transition, and $N = 32$ without a transition, at a 95% confidence threshold. For this analysis, individual time series were winsorized, detrended within windows of 70 observations. Number of early warning signals indicated in the upper right corner of each plot for both groups, meet the criterion of $\tau \geq .1$ The dotted vertical line indicates 0, and the orange vertical line indicates the $\tau \geq .1$ minimally relevant effect size threshold.

Distribution of $\tau$ for the autocorrelation

In Figure S1a, we can see that the distributions of the $\tau$ showed different spreads and peaks in the two groups - with the modes in the transition group appearing relatively more on the right (higher, more positive values of the $\tau$), and the modes of the no-transition group appearing generally closer to zero or below.

Furthermore, it is noteworthy that many people showed negative values of $\tau$, indicating significant decreases in autocorrelation over time (both for individuals with and without transitions).
**Figure S1b**

*Density plot of the $\tau$-values found in the moving window analysis of the variance.*

Note. PA = Positive Affect; NA = Negative Affect; high = high arousal; low = low arousal.

Note that group sizes were unequal: $N = 9$ individuals with a transition, and $N = 32$ without a transition, at a 95% confidence threshold. For this analysis, individual time series were winsorized, detrended within windows of 70 observations. Number of early warning signals indicated in the upper right corner of each plot for both groups, meet the criterion of $\tau \geq .1$. The dotted vertical line indicates 0, and the orange vertical line indicates the $\tau \geq .1$ minimally relevant effect size threshold.

**Distribution of $\tau$ for the variance**

In Figure S1b, we can see that the differences between groups on the six variables is negligible, with the distributions largely overlapping. However, it is noteworthy is that in both groups, many people actually showed negative values of $\tau$: significant decreases in variance over time (prior to a transition, or during the entire study period). For the people with transitions in particular, this is contrary to what we expected from the theory of early warning signals.
1.2 Sensitivity checks

Sensitivity checks were run to check the results of the reference moving window analysis (i.e., window size \( n=70 \), outliers winsorized and within-window detrend, minimally relevant effect size of \( \tau \geq .1 \), transition identified at 95% confidence) against:

- Using significance of the Mann-Kendall test only (no minimally relevant effect size)
- Hamed-Rao window-dependency correction for the Mann-Kendall \( p \)-value.
- Global detrend
- No winsorizing (outliers unchanged)
- Alternate window sizes: 35 obs, 105 obs, 140 obs (1, 3, and 4 weeks, respectively)
- Different transition confidence levels: 90% and 99%

Altered model settings for the autocorrelation

Table S1a reveals that the number of individuals with early warning signals in the autocorrelation tended to be higher in the group without transitions, but proportionally (in terms of percentage), the group with transitions consistently showed more EWS across the different sensitivity checks.

Of all the checks, the Hamed-Rao correction for the M-K \( p \)-value the most conservative results: true positives reduced to 25% across the four circumplex variables (i.e., PA high/low – NA high/low), and the false positives to 12.5%. Conversely, increasing the window size to 140 observations seemed to improve the rate of true positives (from 44.4% to 58.3% in the circumplex variables), along with a small increase in false positives (from 27.3% to 31.2% in the circumplex variables). However, most model setting changes led to a slight increase or decrease (ca. 2–5%) in both the true and false positive rates.

Shortening the time series in the non-transition group also appeared to result in an increase in false positive EWS for some variables, particularly for cheerful and the PA variables.
For individual variables, this proportional difference – more true positives than false positives – was almost always maintained in the various analyses (some exceptions occurred in the global detrend analysis in particular), though the percentual differences were not always large.

Table S1a

Summary of the main findings and sensitivity checks: number and proportion of early warning signals in the autocorrelation indicated for the various altered model settings.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Reference results</th>
<th>M-K trend p &lt;0.05</th>
<th>M-K trend HR p &lt;0.05</th>
<th>Global detrend</th>
<th>Not winsorized</th>
<th>Window size 35</th>
<th>Window size 105</th>
<th>Window size 140</th>
<th>Shortened time series</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Transition (N = 9)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PA high</td>
<td>33.3% (3)</td>
<td>33.3% (3)</td>
<td>11.1% (1)</td>
<td>22.2% (2)</td>
<td>33.3% (3)</td>
<td>22.2% (2)</td>
<td>33.3% (3)</td>
<td>33.3% (3)</td>
<td>–</td>
</tr>
<tr>
<td>PA low</td>
<td>44.4% (4)</td>
<td>55.6% (5)</td>
<td>22.2% (2)</td>
<td>44.4% (4)</td>
<td>33.3% (3)</td>
<td>44.4% (4)</td>
<td>55.6% (5)</td>
<td>77.8% (7)</td>
<td>–</td>
</tr>
<tr>
<td>NA high</td>
<td>55.6% (5)</td>
<td>66.7% (6)</td>
<td>33.3% (3)</td>
<td>77.8% (7)</td>
<td>55.6% (5)</td>
<td>44.4% (4)</td>
<td>66.7% (6)</td>
<td>77.8% (7)</td>
<td>–</td>
</tr>
<tr>
<td>NA low</td>
<td>44.4% (4)</td>
<td>44.4% (4)</td>
<td>33.3% (3)</td>
<td>22.2% (2)</td>
<td>44.4% (4)</td>
<td>33.3% (3)</td>
<td>44.4% (4)</td>
<td>44.4% (4)</td>
<td>–</td>
</tr>
<tr>
<td>cheerful</td>
<td>44.4%</td>
<td>50%</td>
<td>25%</td>
<td>41.7%</td>
<td>41.7%</td>
<td>36.1%</td>
<td>50%</td>
<td>58.3%</td>
<td>–</td>
</tr>
<tr>
<td>down</td>
<td>66.7% (6)</td>
<td>66.7% (6)</td>
<td>22.2% (2)</td>
<td>44.4% (4)</td>
<td>66.7% (6)</td>
<td>33.3% (3)</td>
<td>44.4% (4)</td>
<td>44.4% (4)</td>
<td>–</td>
</tr>
<tr>
<td><strong>No transition (N = 32)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PA high</td>
<td>18.8% (6)</td>
<td>28.1% (9)</td>
<td>6.2% (2)</td>
<td>31.2% (10)</td>
<td>18.8% (6)</td>
<td>18.8% (6)</td>
<td>25% (8)</td>
<td>25% (8)</td>
<td>25% (8)</td>
</tr>
<tr>
<td>PA low</td>
<td>28.1% (9)</td>
<td>31.2% (10)</td>
<td>15.6% (5)</td>
<td>25% (8)</td>
<td>25% (8)</td>
<td>25% (8)</td>
<td>31.2% (10)</td>
<td>31.2% (10)</td>
<td>37.5% (12)</td>
</tr>
<tr>
<td>NA high</td>
<td>31.2% (10)</td>
<td>37.5% (12)</td>
<td>15.6% (5)</td>
<td>28.1% (9)</td>
<td>28.1% (9)</td>
<td>21.9% (7)</td>
<td>37.5% (12)</td>
<td>37.5% (12)</td>
<td>31.2% (10)</td>
</tr>
<tr>
<td>NA low</td>
<td>31.2% (10)</td>
<td>31.2% (10)</td>
<td>12.5% (4)</td>
<td>34.4% (11)</td>
<td>31.2% (10)</td>
<td>21.9% (7)</td>
<td>37.5% (12)</td>
<td>31.2% (10)</td>
<td>31.2% (10)</td>
</tr>
<tr>
<td>cheerful</td>
<td>27.3%</td>
<td>32%</td>
<td>12.5%</td>
<td>29.7%</td>
<td>25.8%</td>
<td>21.9%</td>
<td>32.8%</td>
<td>31.2%</td>
<td>31.2%</td>
</tr>
<tr>
<td>down</td>
<td>12.5% (4)</td>
<td>15.6% (5)</td>
<td>6.2% (2)</td>
<td>21.9% (7)</td>
<td>12.5% (4)</td>
<td>9.4% (3)</td>
<td>15.6% (5)</td>
<td>12.5% (4)</td>
<td>25% (8)</td>
</tr>
</tbody>
</table>
| **Note.** PA = Positive Affect; NA = Negative Affect; high = high arousal; low = low arousal; M-K = Mann-Kendall rank correlation; HR = Hamed-Rao correction. Reference results are based on the analysis of transitions identified at 95% confidence level, where individual time series were winsorized, locally detrended within windows of 70 observations. Number of early warning signals based on cases with $\tau \geq 0.1$.

Altered model settings for the variance

Across the sensitivity checks shown in Table S1b, the variance showed EWS in only a few variables for the individuals with transitions (range = 0 – 33.3%), while those without transitions showed at least one EWS in all variables (range = 3.1 – 21.9%). The lack of a proportional difference in EWS between individuals with and without a transition was consistent across the sensitivity checks. Averaged across the affect circumplex variables, true


positives ranged from 8.3–16.7% in the various analyses, and false positives ranged from 10.9–
16.4%. None of the altered model settings led to a relevant change in the results, all showing
only a slight increase or decrease (ca. 2–5%) in both the true and false positive rates.

Table S1b
Summary of the main findings and sensitivity checks: number and proportion of early warning signals in the
variance indicated for the various altered model settings

<table>
<thead>
<tr>
<th>Variable</th>
<th>Reference results</th>
<th>M-K trend p &lt;.05</th>
<th>M-K trend HR p &lt;.05</th>
<th>Global detrend</th>
<th>Not winsorized</th>
<th>Window size 35</th>
<th>Window size 105</th>
<th>Window size 140</th>
<th>Shortened time series</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Transition (N = 9)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PA high</td>
<td>0% (0)</td>
<td>11.1% (1)</td>
<td>0% (0)</td>
<td>0% (0)</td>
<td>11.1% (1)</td>
<td>22.2% (2)</td>
<td>0% (0)</td>
<td>11.1% (1)</td>
<td>–</td>
</tr>
<tr>
<td>PA low</td>
<td>33.3% (3)</td>
<td>33.3% (3)</td>
<td>22.2% (2)</td>
<td>33.3% (3)</td>
<td>22.2% (2)</td>
<td>33.3% (3)</td>
<td>22.2% (2)</td>
<td>11.1% (1)</td>
<td>–</td>
</tr>
<tr>
<td>NA high</td>
<td>11.1% (1)</td>
<td>11.1% (1)</td>
<td>11.1% (1)</td>
<td>22.2% (2)</td>
<td>0% (0)</td>
<td>22.2% (2)</td>
<td>11.1% (1)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>NA low</td>
<td>0% (0)</td>
<td>0% (0)</td>
<td>0% (0)</td>
<td>0% (0)</td>
<td>0% (0)</td>
<td>0% (0)</td>
<td>0% (0)</td>
<td>0% (0)</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11.1%</td>
<td>13.9%</td>
<td>8.3%</td>
<td>11.1%</td>
<td>16.7%</td>
<td>11.1%</td>
<td>13.9%</td>
<td>11.1%</td>
</tr>
<tr>
<td>cheerful</td>
<td>33.3% (3)</td>
<td>33.3% (3)</td>
<td>0% (0)</td>
<td>33.3% (3)</td>
<td>33.3% (3)</td>
<td>22.2% (2)</td>
<td>22.2% (2)</td>
<td>22.2% (2)</td>
<td>–</td>
</tr>
<tr>
<td>down</td>
<td>22.2% (2)</td>
<td>33.3% (3)</td>
<td>0% (0)</td>
<td>11.1% (1)</td>
<td>11.1% (1)</td>
<td>33.3% (3)</td>
<td>22.2% (2)</td>
<td>22.2% (2)</td>
<td>–</td>
</tr>
<tr>
<td><strong>No transition (N = 32)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PA high</td>
<td>15.6% (5)</td>
<td>18.8% (6)</td>
<td>12.5% (4)</td>
<td>18.8% (6)</td>
<td>15.6% (5)</td>
<td>15.6% (5)</td>
<td>18.8% (6)</td>
<td>18.8% (6)</td>
<td>21.9% (7)</td>
</tr>
<tr>
<td>PA low</td>
<td>6.2% (2)</td>
<td>9.4% (3)</td>
<td>6.2% (2)</td>
<td>9.4% (3)</td>
<td>6.2% (2)</td>
<td>6.2% (2)</td>
<td>9.4% (3)</td>
<td>12.5% (4)</td>
<td>12.5% (4)</td>
</tr>
<tr>
<td>NA high</td>
<td>9.4% (3)</td>
<td>12.5% (4)</td>
<td>9.4% (3)</td>
<td>12.5% (4)</td>
<td>9.4% (3)</td>
<td>12.5% (4)</td>
<td>15.6% (5)</td>
<td>15.6% (5)</td>
<td>12.5% (4)</td>
</tr>
<tr>
<td>NA low</td>
<td>15.6% (5)</td>
<td>18.8% (6)</td>
<td>15.6% (5)</td>
<td>15.6% (5)</td>
<td>15.6% (5)</td>
<td>15.6% (5)</td>
<td>15.6% (5)</td>
<td>15.6% (5)</td>
<td>18.8% (6)</td>
</tr>
<tr>
<td></td>
<td>11.7%</td>
<td>14.8%</td>
<td>10.9%</td>
<td>13.3%</td>
<td>14.1%</td>
<td>11.7%</td>
<td>13.3%</td>
<td>14.8%</td>
<td>16.4%</td>
</tr>
<tr>
<td>cheerful</td>
<td>12.5% (4)</td>
<td>12.5% (4)</td>
<td>9.4% (3)</td>
<td>15.6% (5)</td>
<td>12.5% (4)</td>
<td>12.5% (4)</td>
<td>15.6% (5)</td>
<td>21.9% (7)</td>
<td>12.5% (4)</td>
</tr>
<tr>
<td>down</td>
<td>9.4% (3)</td>
<td>12.5% (4)</td>
<td>3.1% (1)</td>
<td>9.4% (3)</td>
<td>6.2% (2)</td>
<td>6.2% (2)</td>
<td>6.2% (2)</td>
<td>6.2% (2)</td>
<td>21.9% (7)</td>
</tr>
</tbody>
</table>

Note. PA = Positive Affect; NA = Negative Affect; high = high arousal; low = low arousal; M-K = Mann-Kendall rank correlation;
HR = Hamed-Rao correction. Reference results are based on the analysis of transitions identified at 95% confidence level, where
individual time series were winsorized, locally detrended within windows of 70 observations. Number of early warning signals
based on cases with a τ ≥ .1

Like for the autocorrelation, the Hamed-Rao adjusted M-K p-value resulted in the most
conservative outcomes, especially for the single items. Conversely, the shortened time series
in the non-transition group led to the strongest increase in false positives in the circumplex
variables, however, this was still only a ~5% increase compared to the main analysis. Notably,
NA low arousal never showed EWS in the transition group in any of the analyses, while PA
low showed EWS in 2 or 3 individuals in the transition group (22.2–33.3%), with a similar 2 to 4 false positives in the no-transition group (6.2–12.5%).

**Moving window results at different transition thresholds**

Table S2 details the proportion of EWS found in the moving window analyses for the autocorrelation and variance when transitions were defined at a lower confidence level (90%), and at a higher confidence level (99%), compared to the main analysis (95%). Across the analyses for the autocorrelation, the group with transitions (consisting of 14 individuals at 90% CL, and 6 individuals at 99% CL), the proportion of identified EWS was consistently higher than in the group of individuals without a transition – with the exception of NA high and NA low at 90% CL, and PA low at 99% CL, where the proportion of EWS was similar for both groups. For the variance, the lack of proportional difference in EWS between groups also extended to these alternate analyses.

**Table S2**

*Comparison of the number and proportion of early warning signals in the autocorrelation and variance found in the moving window analysis when different confidence levels were used to identify transitions.*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Autocorrelation</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>90% CL</td>
<td>95% CL</td>
</tr>
<tr>
<td><strong>Transition</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PA high</td>
<td>50% (7)</td>
<td>33.3% (3)</td>
</tr>
<tr>
<td>PA low</td>
<td>50% (7)</td>
<td>44.4% (4)</td>
</tr>
<tr>
<td>NA high</td>
<td>35.7% (5)</td>
<td>55.6% (5)</td>
</tr>
<tr>
<td>NA low</td>
<td>35.7% (5)</td>
<td>44.4% (4)</td>
</tr>
<tr>
<td></td>
<td>46.2%</td>
<td>44.4%</td>
</tr>
<tr>
<td>cheerful</td>
<td>42.9% (6)</td>
<td>33.3% (3)</td>
</tr>
<tr>
<td>down</td>
<td>57.1% (8)</td>
<td>66.7% (6)</td>
</tr>
<tr>
<td><strong>No transition</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PA high</td>
<td>18.5% (5)</td>
<td>18.8% (6)</td>
</tr>
<tr>
<td>PA low</td>
<td>33.3% (9)</td>
<td>28.1% (9)</td>
</tr>
<tr>
<td>NA high</td>
<td>33.3% (9)</td>
<td>31.2% (10)</td>
</tr>
<tr>
<td>NA low</td>
<td>33.3% (9)</td>
<td>31.2% (10)</td>
</tr>
<tr>
<td></td>
<td>22.2%</td>
<td>27.3%</td>
</tr>
<tr>
<td>cheerful</td>
<td>11.1% (3)</td>
<td>12.5% (4)</td>
</tr>
<tr>
<td>down</td>
<td>18.5% (5)</td>
<td>21.9% (7)</td>
</tr>
</tbody>
</table>

*Note.* EWS = Early warning signals; CL = confidence level; PA = Positive Affect, NA = Negative Affect; high = high arousal; low = low arousal. Total $N = 41$. At 90% CL: 13 transitions (one person had insufficient data before the transition to be analysed). At 95% CL: 9 transitions. At 99% CL: 6 transitions. Main analysis at 95% CL in **bold.** For the other CLs, all other analysis settings were equal: individual time series were winsorized, locally detrended within windows of 70 observations. Number of early warning signals based on cases with $a \tau \geq .1$. 


1.3 Further exploration

1.3.1 Combining EWS

The moving window results were further explored to see if the number of EWS per individual could be a simple way to improve the specificity of the results. We would expect that individuals with transitions show EWS in multiple variables in both autocorrelation and variance (4 variables × 2 EWS), whereas individuals without transitions would show no or fewer EWS.

Differentiating groups by number of EWS per person

Based on Figure S2, we see that the individuals with a transition all \((N = 9)\) showed at least one EWS – out of a possible 8 – in one of the four variables (PA high/low – NA high/low),

**Figure S2**

*Total number of EWS per person in autocorrelation and variance combined, indicated by the moving window analysis*

*Note.* \(N = 9\) individuals with a transition, and \(N = 32\) without a transition.
in either the autocorrelation or variance results of the moving window analysis. In the group without transitions, most people \((N = 19)\) had 0 or only 1 EWS, but there were also 13 individuals with at least two EWS.

Thus, if combining EWS and setting the threshold to a minimum of two EWS per person, this results in 77.8% true positives, and 40.6% false positives. Raising the threshold by one step \((N_{\text{EWS}} \geq 3)\) lowers the sensitivity again, capturing 33.3% true positives, and 18.8% false positives.

**Number of EWS per person in autocorrelation and variance separately**

On the left side of Figure S3, we see that all but one (thus, 88.89%) of the individuals with a transition showed at least one EWS for the autocorrelation (out of a possible 4), compared to 62.5% in the group without transitions. Raising the minimum number of EWS in autocorrelation by one \((N_{\text{EWS}} \geq 2)\) leads to 55.5% true positives from the transition group, and 28.1% false positives. For minimally 3 EWS in autocorrelation, the transition group shows 22.2% true positives, while the no-transition group shows 12.5% false positives.

For the variance, on the right side of Figure S3, 4 out of 9 (44.4%) individuals with a transition had one variable (out of a possible 4) with EWS, compared to 8 (25%) in the group without transitions. The remaining people with transitions had no EWS in the variance at all (i.e., 55.5% false negatives), while 12.5% of non-transitioning individuals was still found with a threshold \(N_{\text{EWS}} \geq 2\).
**Figure S3**

*Number of EWS per person in autocorrelation and variance separately, indicated by the moving window analysis.*

![Figure S3](image)

*Note. N = 9 individuals with a transition, and N = 32 without a transition.*

**1.3.2 MRES: τ exploration**

As indicated in the preregistration, we examined different minimally relevant effect sizes (MRES) than the chosen τ of .1, to compare the balance of true positives (EWS in individuals with a transition) to false positive EWS (EWS in individuals with no transition) at different increments of τ in the moving window results. To that end, we calculated the balanced accuracy (BA; the average of the proportions of correct classifications) for the autocorrelation and variance, as follows:

1. Sensitivity or True positive rate (TPR): N EWS in the transition group/N transition group
2. Specificity or True negative rate (TNR): 1 - (N EWS in the non-transition group/N non-transition group)
3. Balanced accuracy (BA): TPR + TNR / 2
Balanced accuracy for different increments of $\tau$

Figure S4 shows the balanced accuracy for the EWS found in autocorrelation and variance with the moving window method. For the BA, a value of 50% means the rate of true positives to true negatives is equal, no better than chance; values below 50% indicate that there are more false positives or false negatives; values above 50% indicate a higher rate of true positives or negatives. We can see that across the various increments of $\tau$, some variables remain stable in the trade-off between sensitivity and specificity indicated by the BA percentages (e.g., PA high and NA low in autocorrelation, and NA high and NA low in variance), whereas others have a higher BA at some thresholds of $\tau$. For instance, the single item “down” seems to find a higher BA at a $\tau$ between 0 and 0.2, for both the autocorrelation and variance. Note also that for the autocorrelation, NA high appears to work relatively well across all increments. However, overall, this figure shows the BA values for our sample are modest at best, and particularly the variance performs close to chance. Given these low BA values and the inconsistent ‘optimal’ threshold for different variables, it is not possible to make any strong recommendations about appropriate $\tau$ thresholds for future research.
Figure S4

Balanced Accuracy for the autocorrelation and variance results of the moving window analysis

Note. PA = Positive Affect; NA = Negative Affect; high = high arousal; low = low arousal.

N = 9 individuals with a transition, and N = 32 without a transition.

Balanced accuracy (BA): True Positive Rate + True Negative Rate / 2

2. TV-AR GAM method

2.1 Main analysis and sensitivity check

The main TV-AR GAM analysis results show the different time-varying patterns that were found for the transition group, and the group without a transition (columns in bold in the table below). Additional sensitivity checks were conducted to examine the effect of the longer time series in the no-transition group, and the analysis was repeated with the transitions identified at 90% and 99% confidence level.
TV-AR-GAM results for different CL and shortened time series

Table S3 details the proportion of EWS found in the autocorrelation, as modeled by the TV-AR-GAM method, when transitions were defined at a lower confidence level (90%), and at a higher confidence level (99%), compared to the main analysis (95%).

Across these analyses at different confidence levels, the number of cases with a significant increase in autocorrelation is similar in the transition group and non-transition group and remarkably low. At 90%, 95% and 99%, the transition group had 6%, 5.6% and 0% EWS across the four affect circumplex variables, respectively, while the individuals without transitions showed 3.1%, 3.1% and 3.3% significant increases in autocorrelation across those four variables.

Another pattern across analyses, variables and the two groups, was that half the individuals or more (range = 48% to 87%) did not show a significant change over time at all.

Among the significant smooth curves, the proportion of cases with nonlinear changes (i.e., cubic, quadratic, or more complex curves) stands out. For instance, summarized across the four affect circumplex variables, 27.8% of the significant changes were nonlinear in the 95% CL transition group, and 25.8% in the non-transition group. The percentages at the other confidence levels were 10.7% and 25.3% at 90% CL, and 22.2% and 29% at 99% CL for the transition and no-transition groups, respectively.

Of further note are the results of the sensitivity check of the effect of time series length for individuals without transitions. For all confidence levels, the shortened time series led to fewer significant smooth curves, particularly to fewer cases with a decrease. Likely, the loss in significant smooth curves is because TV-AR GAM had less data available in the shortened time series, and could not fit the 95% credible intervals as narrowly as it could for the full-length time series. Wider credible intervals would result in fewer cases where change could be identified with certainty.
Table S3
Comparison of the TV-AR GAM results: direction of the modeled smooth curves and proportion of EWS at different confidence levels and with the shortened time series in the no-transition group.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Prop. EWS</th>
<th>90% CL</th>
<th>Reference: 95% CL</th>
<th>99% CL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(Transition N = 14, No transition N = 27)</td>
<td>(Transition N = 9, No transition N = 32)</td>
<td>(Transition N = 6, No transition N = 35)</td>
</tr>
<tr>
<td>Transition</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PA high</td>
<td>7.1%</td>
<td>0%</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>PA low</td>
<td>7.1%</td>
<td>0%</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>NA high</td>
<td>7.1%</td>
<td>0%</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>NA low</td>
<td>0%</td>
<td>0%</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>cheerful</td>
<td>7.1%</td>
<td>0%</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>down</td>
<td>7.1%</td>
<td>0%</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>No transition</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PA high</td>
<td>0%</td>
<td>0%</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>PA low</td>
<td>0%</td>
<td>0%</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>NA high</td>
<td>0%</td>
<td>0%</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>NA low</td>
<td>0%</td>
<td>0%</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>cheerful</td>
<td>0%</td>
<td>0%</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>down</td>
<td>7.1%</td>
<td>0%</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

Sensitivity check:
No-transition ts set to average transition ts length

<table>
<thead>
<tr>
<th>Variable</th>
<th>Prop. EWS</th>
<th>90% CL</th>
<th>Reference: 95% CL</th>
<th>99% CL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(Transition N = 14, No transition N = 27)</td>
<td>(Transition N = 9, No transition N = 32)</td>
<td>(Transition N = 6, No transition N = 35)</td>
</tr>
<tr>
<td>PA high</td>
<td>0%</td>
<td>0%</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>PA low</td>
<td>0%</td>
<td>0%</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>NA high</td>
<td>0%</td>
<td>0%</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>NA low</td>
<td>0%</td>
<td>0%</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>cheerful</td>
<td>0%</td>
<td>0%</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>down</td>
<td>7.4%</td>
<td>0%</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

Note. ↗ = Increasing trend. ↘ = Decreasing trend. ~ = Nonlinear trend. — = No trend. PA = Positive Affect; NA = Negative Affect; ts = time series; Prop. EWS = Proportion of EWS, which is calculated only over the number of cases with a significant increase in autocorrelation. Average time series length in the transition group = 334, 441 and 458 at 90%, 95% and 99% confidence, and in the no transition group 616, 616 and 615, respectively. Main analysis at 95% CL in bold. For the other CLs, all other analysis settings were equal.