AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Consulting and Clinical Psychology

© 2024 American Psychological Association
ISSN: 0022-006X

2024, Vol. 92, No. 8, 505–516
https://doi.org/10.1037/ccp0000871

Monitoring Emotional Intensity and Variability to Forecast Depression Recurrence in Real Time in Remitted Adults

Marieke J. Schreuder[1], Evelien Schat[1], Arnout C. Smit[2], Evelien Snippe[3], and Eva Ceulemans[1]

[1] Quantitative Psychology and Individual Differences, Faculty of Psychology and Educational Sciences, Katholieke Universiteit Leuven
[2] Clinical Psychology, Faculty of Behavioral and Movement Sciences, Vrije Universiteit Amsterdam
[3] Interdisciplinary Center Psychopathology and Emotion Regulation, University Center for Psychiatry, University Medical Center Groningen, University of Groningen

Objective: Recurrent depressive episodes are preceded by changing mean levels of repeatedly assessed emotions (e.g., feeling restless), which can be detected in real time using statistical process control (SPC). This study investigated whether monitoring changes in the standard deviation (*SD*) of emotions and negative thinking improves the early detection of recurrent depression. ***Method:*** Formerly depressed adults ($N = 41$) monitored their emotions five times a day for 4 consecutive months. During the study, 22 individuals experienced recurrent depression. We used SPC to detect warning signs (i.e., changing means and *SD*s) of four emotions (positive and negative affect with high or low arousal) and negative thinking. ***Results:*** *SD*-based warning signs only preceded 23%–36% of recurrences, but almost never reflected a false alarm (0%–16%). Correspondingly, *SD*-based warnings had a high specificity (at the cost of sensitivity), while mean-based warnings had a higher sensitivity (but lower specificity). There was little overlap in mean- and *SD*-based warning signs. For the majority of emotions, monitoring for high *SD*s alongside monitoring changes in mean levels improved the detection of depression ($p < .015$) compared to when only monitoring for changing mean levels. ***Conclusions:*** Warning signs for depression manifest not only in changing mean levels of emotions and cognitions but also in increasing *SD*s. These warnings could eventually be used to detect not just who is at increased risk for depression but also when risk is rising. Further research is needed to evaluate the clinical utility of depression SPC.

What is the public health significance of this article?
Changes in the mean levels of self-reported emotions precede recurrent depressive episodes and are therefore considered warning signals. This study shows that the early detection of recurrent depression improves when warning signals are not only tracked in the mean but also in the variability of emotions. Thus, monitoring for both changes in emotional intensity and variability may aid the early detection, and possibly the prevention, of depression.

Keywords: recurrent depression, ecological momentary assessments, statistical process control

Supplemental materials: https://doi.org/10.1037/ccp0000871.supp

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Major depressive disorder is among the leading causes of disease burden worldwide, accounting for 7.5% of all years lived with disability (World Health Organization, 2017). This burden follows not only from the high prevalence of depression but also from the fact that depressive episodes tend to recur. To illustrate, individuals with depression experience on average eight episodes in their lifetime, each lasting approximately 6 months (Andrews, 2008). An important goal of psychological treatment is therefore to prevent recurrence by recognizing impending episodes as early as possible.

A method that may aid in this pursuit is ecological momentary assessment (EMA), in which participants repeatedly register their mental states (e.g., emotions, cognitions). The idea here is that carefully monitoring the extent to which individuals, for instance, feel restless or worry, may provide an indication of their risk for

This article was published Online First March 21, 2024.
Marieke J. Schreuder https://orcid.org/0000-0003-4901-2697
This project has received funding from the Netherlands Organisation for Health Research and Development (ZonMw Off Road; project 451001029 [Grant 40-08125-98-18041] awarded to Evelien Snippe), the HORIZON EUROPE European Research Council (ERC) under the European Union's Horizon 2020 research and innovative program (ERC-CoG-2015; Grant 681466 awarded to M. Wichers), and the Research Council Onderzoeksraad, KU Leuven (project C14/19 awarded to Eva Ceulemans and P. Kuppens). The funders had no role in the design of this study, nor in the collection, management, analysis, and interpretation of data, nor in the decision to submit findings for publication.

The authors thank the participants for their time and effort, P. Harder for assistance with the recruitment of participants, M. Messchendorp and R. de Vries for data collection, and Marieke Wichers, Harriëtte Riese, Marieke A. Helmich, and Yoram K. Kunkels for codesigning the transitions in depression tapering study. The authors have no conflicts of interest to disclose.

Data are not freely available due to restrictions related to the European law regarding general data protection regulation. A detailed protocol of the transitions in depression tapering study is available on the open science framework at https://osf.io/zbwkp. Further, a preregistration of the present study can be found online at https://osf.io/4rcn3/. Collaboration on

continued

recurrent depression. Indeed, several studies suggested that EMA may facilitate in the early detection of depressive episodes (Groot, 2010; Palmier-Claus et al., 2011; Snippe et al., 2023; Torous et al., 2015; van Os et al., 2017; Wichers & Groot, 2016).

An often-cited example of the potential clinical utility of EMA is a case study, which showed that a relapse in depression was anticipated by changing dynamics in momentary, self-reported emotions (Groot, 2010; Wichers & Groot, 2016). This case study was followed up by a larger study in which individuals with a history of major depressive disorder gradually discontinued their antidepressant medication and simultaneously reported their mental states five times a day for 4 consecutive months (Smit et al., 2020). Due to the tapering of medication, individuals were at increased risk for a recurrence of depression, which indeed occurred in almost 60% of the participants (Smit, Snippe, et al., 2023). These recurrences of depression were often preceded by changes in the mean levels of mental states (Snippe et al., 2023). These so-called warning signs occurred approximately 1 month prior to the recurrence and were only seldomly found in individuals without a recurrence. This showed that monitoring changing mean levels of momentary mental states in real time may aid the early detection of depression.

The innovative aspect of the study by Snippe et al. (2023) is that a real-time method, namely statistical process control (SPC), was used to detect change in the intensity of emotions (Box 1). Although the clinical utility of such a real-time method is promising (see Box 1), the sensitivity and specificity of this approach leave room for improvement. One way to achieve this may be to monitor not only mean levels but also the variability (*SD*) of mental states. Indeed, several studies have shown higher variability in the emotions of people with depression compared to controls, suggesting that heightened variability may be a sign of maladaptive emotion regulation that is characteristic of depression (E. H. Bos et al., 2019; Dejonckheere et al., 2019; Houben et al., 2015; Houben & Kuppens, 2020; Koval et al., 2013; Sperry et al., 2020; Sperry & Kwapil, 2022; Trull et al., 2015; van de Leemput et al., 2014). More cognitive characteristics of depression, such as ruminating or having negative thoughts about oneself, were also found to be more variable in persons with depression than in controls (Bean, 2023; Crowe et al., 2019; Franck & De Raedt, 2007). The latter could signal a fragile or vulnerable self-esteem, which can both precipitate and maintain depression. The idea that variability in mental states may convey risk for upcoming depression was further promoted by complex systems approaches to depression (Cramer et al., 2016; van de Leemput et al., 2014; Wichers & Groot, 2016). Briefly, these approaches state that recurrent depression may be anticipated by critical slowing down. Critical slowing down describes a process of slowing recovery in complex systems prior to critical transitions. It involves increasingly persistent fluctuations away from baseline, which translates to rising variability (Scheffer, 2009). Complex systems approaches to depression thus suggest that rising variability of mental states may forewarn imminent transitions in depression (TRANS-ID; recurrence). Following this reasoning, monitoring the variability of mental states may aid the early detection of depression.

Although rising trends in the variability of mental states were indeed associated with worsening depression (Curtiss et al., 2023; Smit, Helmich, et al., 2023), they were only found in a minority of participants. Thus, if useful at all, warning signs in the variability of mental states should probably be considered complementary rather than stand-alone indices of risk for depression. Furthermore, it is unclear to what extent previously detected trends in variability can be attributed to changes in mean levels of mental states (Mestdagh et al., 2018), implying that rises in variability may not convey unique information about one's vulnerability for depression. Finally, rises in variability reported in former studies were not detected in real time (i.e., continuously updating the risk evaluation with new data), but retrospectively (i.e., making a single prediction after all data have been collected; Box 1). As a result, it is still unknown whether monitoring warning signs in the variability of mental states could eventually be useful in (clinical) practice.

To investigate whether real-time monitoring of within-person changes in the variability of mental states may aid the early detection of recurrent depression, we will analyze the TRANS-ID tapering data (Smit et al., 2020). We will build on earlier studies on the same data, which revealed that recurrent depression can be anticipated by (a) real-time changes in mean levels of mental states (Snippe et al., 2023) and (b) retrospectively detected trends in the variability of mental states (Smit, Helmich, et al., 2023). To extend these earlier findings, we will use a method that detects warnings in the variability of mental states in real time (Box 1) and will compare these warnings to the changes in the mean that were reported before. We hypothesize that (1) warning signs in the variability of four types of emotions, namely positive affect (PA) high arousal, PA low arousal, negative affect (NA) high arousal and NA low arousal, as well as negative thinking (neg. thinking), occur more often in individuals with recurrent depression than in individuals who remain in remission and that (2) monitoring warning signs in both

transitions in depression tapering data is possible and can be requested via h.riese@umcg.nl. Data are available upon request.

The articles (Smit, Snippe, et al., 2023; Smit, Helmich, et al., 2023; Smit & Snippe, 2023; Snippe et al., 2023) used the same data as the current article: Together, these studies show that recurrence of depression may happen suddenly (Smit, Snippe, et al., 2023), that rising trends in emotional variability retrospectively predict recurrent depression (Smit, Helmich, et al., 2023), and that real-time trends in emotional intensity predict recurrent depression (Smit & Snippe, 2023; Snippe et al., 2023). The present study adds to this by showing that emotional variability can also be monitored in real time, and further, that monitoring emotional variability together with emotional intensity improves the detection of recurrent depression. Throughout the article, the authors emphasized how their study relates to these earlier studies.

Marieke J. Schreuder and Evelien Schat contributed equally to this work.

Marieke J. Schreuder played a lead role in writing–original draft and an equal role in conceptualization and methodology. Evelien Schat played a lead role in formal analysis and software and an equal role in conceptualization, methodology, and writing–original draft. Arnout C. Smit played a supporting role in writing–review and editing and an equal role in conceptualization, data curation, investigation, methodology, and software. Evelien Snippe played an equal role in conceptualization, data curation, investigation, methodology, software, and writing–review and editing. Eva Ceulemans played a lead role in supervision and an equal role in conceptualization, methodology, and writing–review and editing.

Correspondence concerning this article should be addressed to Marieke J. Schreuder, Quantitative Psychology and Individual Differences, Faculty of Psychology and Educational Sciences, Katholieke Universiteit Leuven, Tiensestraat 102 (Box 3713), 3000 Leuven, Belgium. Email: mariekeschreuder@gmail.com

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Box 1

Detecting Depression in Time Through Statistical Process Control

Preventive interventions are likely most effective when provided at the right moment in time. To pinpoint this moment, we need methods that can detect warning signs for depression in real time. Such real-time methods contrast with retrospective approaches, which are applied only once, after an outcome (in this case, recurrent depression) has occurred. Such approaches can tell that, in hindsight, an outcome was preceded by a certain trend (e.g., a rising trend in the *SD*). A crucial shortcoming of retrospective approaches is that the resulting warnings are—by definition—situated in the past, precluding timely intervention. Instead, real-time methods are repeatedly conducted while new data are being gathered and can thus in principle generate a warning sign each time new data come in.[1] SPC (Montgomery, 2009) accommodates this real-time detection of warning signs (Schat et al., 2021; Smit et al., 2019; Smit, Schat, & Ceulemans, 2023; Smit & Snippe, 2023; Snippe et al., 2023). Although SPC is well-established in other fields, it was only recently introduced in psychology. Figure 1 illustrates the application of SPC to openly available EMA data from a single person with a history of depression (Kossakowski et al., 2017). Warning signs in both the weighted averages of within-day means and within-day *SDs* were detected from Day 50 onward (indicated by the red dots), thus preceding a recurrence of depression (blue vertical line). Of note, Figure 1 illustrates a one-sided SPC application, which is most relevant for the current article and involves a single control limit (in this case, upper). It is, however, also possible to have two-sided applications, yielding an upper and lower control limit (LCL), and thus a range in which repeatedly assessed emotions are expected to fall.

the variability and the mean of mental states improves the early detection of recurrent depression when compared to monitoring the mean only. Related to the latter hypothesis, we will also address the overlap between warning signs in the *SD* and the mean and compare the timing of both warning signs. Finally, we will exploratively evaluate whether the predictive performance of *SD*-based warning signs improves when monitoring not only increases but also decreases in the *SD*. Whether this could improve SPC is arguable. On the one hand, monitoring decreasing *SD*s would accommodate more complex trends in the *SD* prior to a recurrence of depression—for example, a dip followed by a rise. On the other hand, an earlier study on these data showed that declining trends in the *SD* were more pronounced in individuals who remained in remission compared to individuals with recurrent depressive symptoms (Smit, Helmich, et al., 2023), implying that reductions in the *SD* may be a marker for remission rather than recurrence. Altogether, the present study aims to investigate if warning signs in mean levels and the variability of momentary mental states detected in real time can predict whether (but not when) recurrent depression occurs.

Method

Procedure and Participants

We analyzed data from the TRANS-ID tapering study, which has been described in detail on open science framework (Smit et al., 2020). The TRANS-ID tapering study was designed to investigate whether recurrence of depression can be foreseen by within-person changes in emotion dynamics. Briefly, the study consisted of a 4-month EMA study, which was complemented by weekly assessments of depressive symptoms and regular phone calls that lasted 2 months longer (i.e., 6 months). Prior to the EMA study, a semistructured diagnostic interview (short Schedules for Clinical Assessment in Neuropsychiatry, or mini-SCAN, Nienhuis et al., 2010) was administered. Participants additionally completed two semistructured evaluation interviews, scheduled after the EMA study and after 6 months, in which qualitative data about the course of depressive symptoms were collected. For instance, participants who reported a recurrence of depressive symptoms were asked whether they could retrospectively date this recurrence. The TRANS-ID tapering study also included heart rate and actigraphy measurements, but these are not relevant for the present study and will not be described further.

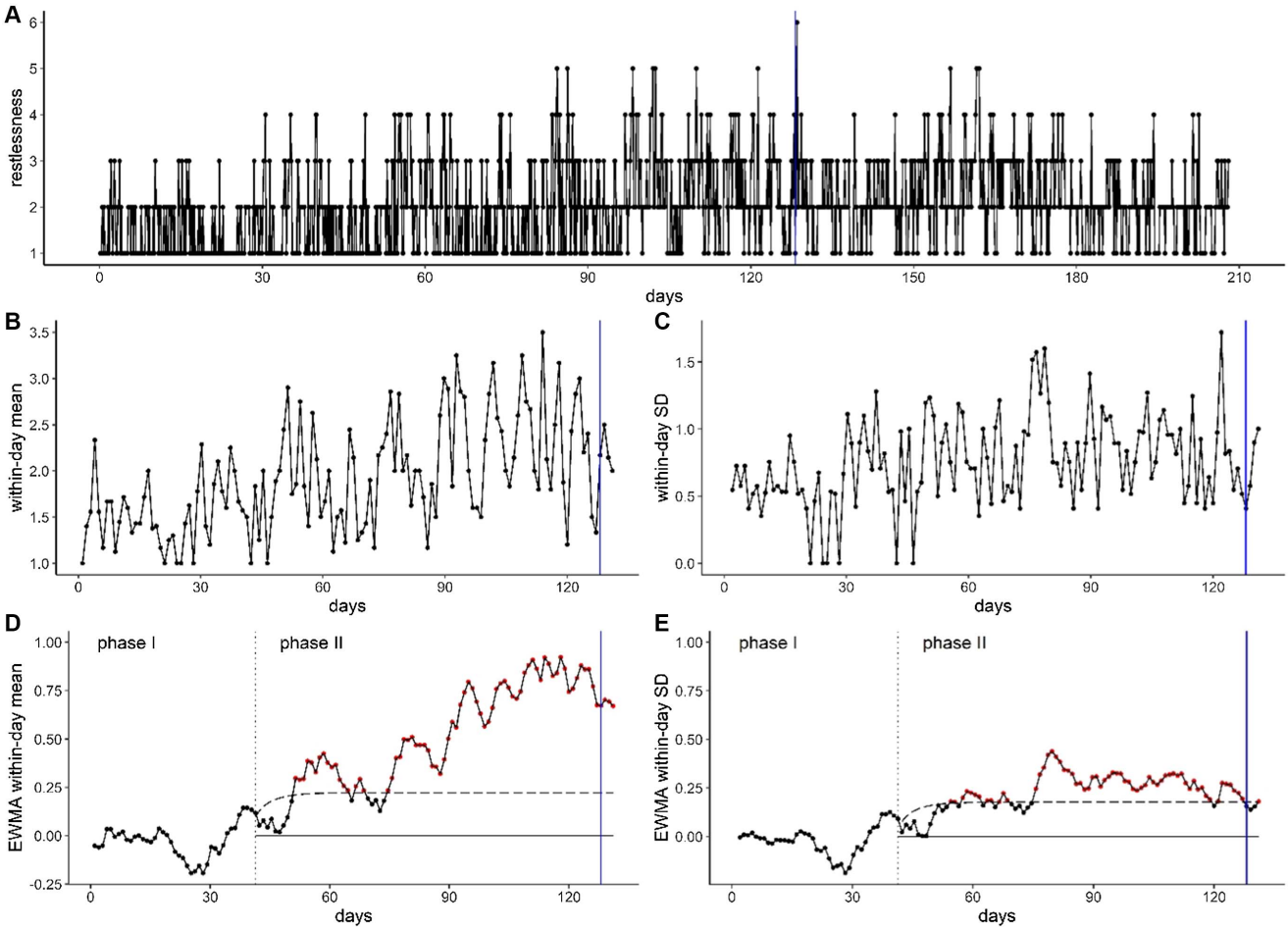
Participants were recruited through advertisements in newspapers, social media, and a Dutch pharmacy (Regenboog Apotheek) that offers tapering strips (Groot & van Os, 2018). Participants were included if they (a) fulfilled the *Diagnostic and Statistical Manual of Mental Disorders, fourth edition* criteria of a past depressive episode, (b) were older than 18 years, (c) currently used antidepressant medication, and (d) had decided to taper and discontinue their medication in consultation with their doctor. Exclusion criteria were (a) currently meeting diagnostic criteria for a depressive episode, bipolar disorder, or a psychotic disorder, (b) reporting a personality disorder, (c) planning to start another type of antidepressant medication. Out of the 299 individuals who initially expressed their interest in participation, 69 were included (Smit & Snippe, 2023; Smit, Snippe, et al., 2023; Snippe et al., 2023). In line with earlier studies on the TRANS-ID tapering data (Smit & Snippe, 2023; Snippe et al., 2023), we excluded individuals for whom (a) less than 1 month of >80% EMA data were available, (b) presence, absence, or timing of recurrence of depressive symptoms could not be determined, (c) less than 6 weeks of EMA data prior to recurrence were available. This resulted in a final sample of 41 participants, of whom 22 experienced a recurrence of depression (see section Recurrence of Depression) and 19 remained in remission. Both groups did not significantly differ in gender (overall 83% females), age (overall 46.5 ± 12.9 years old), or baseline symptoms of

[1] The distinction between retrospective versus real-time methods could perhaps also be understood through a navigation analogy. Real-time location tracking continuously shows you where you are, meaning that the dot on your screen is updated when you move. Retrospective tracking shows you the route you took when you reached the destination. From a methodological viewpoint, the advantages of a real-time method like SPC are that the probability of false positives (Type I errors) can be set by the researcher (or clinician) and that it does not necessitate enormous amounts of data (nor does it require to have "future" data, e.g., data collected after depression recurrence). Such advantages put SPC ahead of other potential approaches (e.g., repeatedly applying retrospective methods).

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Figure 1

Statistical Process Control Applied to a Person's Ecological Momentary Assessments of the Item "Restless"



Note. (A) Raw data, (B) time series of the within-day means, (C) within-day *SDs*, and (D)–(E) statistical process control charts depicting warning signs in the (D) within-day means and (E) *SDs*. As illustrated, SPC consists of two phases. Phase I (Days 1–41) is used to establish a control limit. During Phase II (from Day 43 onward), weighted incoming ratings are compared against this control limit. Warning signs (red dots) occur when scores exceed the control limit. EWMA = exponentially weighted moving average; SPC = statistical process control. See the online article for the color version of this figure.

depression assessed by the Symptom Checklist-90 (SCL-90) Depression subscale (overall 28.5 ± 9.3). Most participants (88%) completed higher education or university. More detailed participant characteristics have been reported elsewhere (Smit & Snippe, 2023; Smit, Snippe, et al., 2023; Snippe et al., 2023). All participants provided written informed consent. The TRANS-ID tapering study was approved by the medical ethical committee of the University Medical Center Groningen. The data are not freely available due to restrictions related to the European law regarding general data protection regulation but can be accessed through a collaboration with researchers who curate the data.

Ecological Momentary Assessment

The EMA study started 1 month prior to full discontinuation of antidepressants (i.e., reaching a dose of 0 mg). Participants were asked to complete a questionnaire five times a day for 4 consecutive months, yielding a maximum of 610 assessments per individual. Questionnaires were sent with fixed 3-hr time intervals via text messages to participants' mobile phones and contained 27 items on participants' current mood (e.g., "I feel tired"), activity, and company. These items were rated on a Visual Analog scale ranging from 0 (*not at all*) to 100 (*very much*). Additionally, the questionnaire contained an open text field where participants could report remarkable experiences. In line with earlier studies using the TRANS-ID tapering data (Smit, Helmich, et al., 2023; Snippe et al., 2023), we analyzed the following five constructs (EMA items between brackets): high arousal PA (cheerful, energetic), low arousal PA (content, at ease), high arousal NA (restless, stressed), low arousal NA (down, listless, tired), and neg. think.[2] (worrying, negative

[2] Note that, in earlier articles, this construct was termed repetitive neg. think. (e.g., Snippe et al., 2023). This is not an emotion, but in the interest of readability, we will henceforward use the term emotions to refer to the constructs in which warning signs were monitored (i.e., trends in mean levels and standard deviations).

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

thoughts about self). The median individual-level internal consistency of these constructs, quantified by Crohnbach's α, was .74 (high arousal NA), .64 (low arousal NA), .70 (high arousal PA), .67 (low arousal PA), and .52 (neg. think.; Snippe et al., 2023). Participants received a financial compensation if they completed at least 80% of the EMA assessments. Median compliance of participants included in the present study was 87.6%, corresponding to 540 assessments (range 220–609).

Weekly Depressive Symptoms

To monitor recurrence of depression, participants were asked to complete the SCL-90 Depression subscale (Derogatis, 1977; Derogatis & Cleary, 1977) every week for a period of 6 consecutive months (i.e., 2 months longer than the EMA study). The SCL-90 Depression subscale consists of 16 items, of which two items on suicidal ideation and self-esteem were omitted because evaluating these constructs on a weekly basis was considered too burdensome. Participants were additionally given the opportunity to denote meaningful changes in their mood or impactful events in open text fields.

Recurrence of Depression

The presence and timing of depression recurrences were established in a previous study (Smit, Helmich, et al., 2023). Criteria for recurrence were (a) having a reliable increase in depressive symptoms, defined as an increase of at least 8.5 points on the SCL-90 subscale (Derogatis & Cleary, 1977) compared to the first 2 weeks of the study based on the reliable change index (Jacobson & Truax, 1991), which (b) persisted for at least 3 consecutive weeks, (or) coincided with the start or increase of treatment, (or) to the interruption of tapering. Additionally, three independent raters with clinical training (two psychologists, one psychiatrist) judged the recurrence based on participants' self-reports regarding meaningful increases in depressive symptoms (e.g., self-reported discomfort due to worsening depressive symptoms). Such self-reported, qualitative information was obtained by means of open text fields in the daily and weekly questionnaires, emails and telephone calls, and semistructured evaluation interviews (for details, see Smit, Snippe, et al., 2023).

Statistical Analyses

Analyses were preregistered (see Supplemental Material 7) and included two preprocessing steps, each accompanied by sensitivity analyses[3] (for a schematic overview, see Supplemental Figure S1). First, we aggregated the EMA items into the emotions listed above by taking the average of the corresponding items per measurement occasion. Building on the recommendations of Schat et al. (2021), we computed within-day means and *SD*s for each person and emotion. This timescale of 1 day was chosen to strike a balance between accuracy of the estimates—which increases with the number of observations, and thus would call for aggregating across a longer period—and the accuracy of control limits—which likely decreases when aggregating across multiple days. The latter occurs because serial dependence seriously impairs the performance of SPC, and therefore, the days across which we aggregate should be nonoverlapping. This effectively means that aggregating across, for instance, 2 days (instead of 1) halves the number of estimates available for computing control limits. The consequence of a 1-day timescale was, however, that *SD*s could not be estimated for days on which an individual completed fewer than three out of five EMA assessments.[4] On average, 5.10 days ($SD = 7.06$, range = 0–37) per person were left out in the analyses. Thus, an average of 113.07 days were available for analyzing *SD*-based warning signs. In sensitivity analyses, we considered a 2-day timescale, where mean levels and *SD*s were computed across 2 days.

A second preprocessing step in our analyses concerned the winsorization of the estimates obtained during the first 28 days (i.e., Phase I). For each participant, emotion, and monitored statistic (day mean and day *SD*) separately, we replaced day means and day *SD*s that deviated more than 3 *SD*s from their respective means during the first 28 days of the study with their respective extreme value (i.e., $M \pm 2$ *SD*s). This is in line with earlier SPC studies using the TRANS-ID tapering data (Smit & Snippe, 2023; Snippe et al., 2023) and was done so that Phase I can be considered reflective of "normal" emotions (and thus, forms a solid basis for control limits). In sensitivity analyses, however, we did not winsorize the Phase I data.

Statistical Process Control

Based on the data collected during Phase I, which spans the first 28 days of the study, we computed control limits that reflect the "normal range" of a certain emotion, for a certain person. The upper control limit (UCL) and LCL were defined as follows:

$$\mathrm{UCL} = \hat{\mu} + L\hat{\sigma}\sqrt{\frac{\lambda}{2-\lambda}\left[1-(1-\lambda)^{2t}\right]}, \tag{1}$$

and

$$\mathrm{LCL} = \hat{\mu} - L\hat{\sigma}\sqrt{\frac{\lambda}{2-\lambda}\left[1-(1-\lambda)^{2t}\right]}. \tag{2}$$

Depending on the statistic that is monitored, $\hat{\mu}$ is the mean of the day means ($\hat{\mu}_M$) or day *SD*s ($\hat{\mu}_{SD}$), and $\hat{\sigma}$ is the *SD* of the day means ($\hat{\sigma}_M$) or day *SD*s ($\hat{\sigma}_{SD}$). The parameters λ and L together determine the sensitivity (i.e., proportion of true alarms) and specificity of SPC (i.e., proportion of false alarms). Similar to earlier SPC studies (Schat et al., 2021; Smit & Snippe, 2023; Snippe et al., 2023), we set λ to 0.10. The parameter L reflects the width of control limits and depends on (a) the average run length (ARL_0), or the average time it should take before a false alarm occurs, and (b) whether one- or two-sided tests are performed. In line with earlier studies, we set L to 2.62 in case of one-sided tests—where we monitored warning signs in a single direction (i.e., too high mean negative emotion, too low positive emotion, too high *SD*s)—and 2.70 in case of two-sided tests—where we examined both too high and too low *SD*s (Schat et al., 2021, 2023; Snippe et al., 2023).

[3] Sensitivity analyses were not preregistered.

[4] Even though it is technically possible to compute a standard deviation based on two assessments only, the resulting estimate will be extremely uncertain. We therefore disregarded days with fewer than three completed assessments.

This means that, in case an individual remains in remission, a false alarm is expected to occur after on average 370 days. In sensitivity analyses, we varied λ between 0.01, 0.05, and 0.15 (see also Snippe et al., 2023). Keeping the average run length constant, this corresponds to L values of 1.77, 2.42, and 2.71 (one-sided tests) or 1.82, 2.49, and 2.80 (two-sided tests). For more information regarding the impact of λ and L, we refer the reader to Smit, Schat, and Ceulemans (2023).

During Phase II, which spans the final 3 months of the EMA study, we monitored—for each participant and emotion separately—whether the day means and day *SD*s exceed the relevant control limit(s). First, at each day t, we compute the exponentially weighted moving average, EWMA_t:

$$\text{EWMA}_t = \lambda x_t + (1-\lambda)\text{EWMA}_{t-1}. \quad (3)$$

Here, x_t denotes the observed statistic at day t (i.e., within-day mean or within-day *SD*). We set EWMA_0 to the estimated Phase I mean of the corresponding statistic (*i.e.*, $\hat{\mu}_M$ or $\hat{\mu}_{SD}$). Next, for the day means and day *SD*s separately, we examined whether and at what point in time at least one EWMA score exceeded the relevant control limit(s). For more detailed information regarding the EWMA procedure in detecting changes in day means and day *SD*s, we refer the reader to Schat et al. (2021, 2023).

Warning Signs

Warning signs in the mean were operationalized as the presence of at least one EWMA score that was higher than the UCL (NA, neg. think.) or lower than the LCL (PA, cf. Snippe et al., 2023). Warning signs in the *SD* were operationalized as the presence of at least one EWMA score that exceeds the UCL (one-sided main analysis) or either the UCL or LCL (two-sided explorative analysis). Note that for the individuals who experienced recurrence of depression, we only considered the days before the transition to determine the presence of warning signs. For individuals who stayed in remission, we looked at the entire EMA period.

For mean- and (one-sided as well as two-sided) *SD*-based warning signs separately, we computed the sensitivity, specificity, predictive values, and accuracy with which they predicted recurrent depressive symptoms. Sensitivity reflects the proportion of individuals with recurrent depression for whom warning signs were detected, and specificity reflects the proportion of individuals without recurrent depression for whom warning signs were not detected (expressed as percentages). Predictive values reflect the probability of recurrent depression, given that warning signs were observed (positive predictive value), or the probability of remaining in remission, given that warning signs were not observed (negative predictive value). Accuracy reflects the total proportion of correct predictions (i.e., true positives and true negatives), and higher values indicate a better (more accurate) prediction of recurrent depression. Matthews correlation coefficient (MCC; Matthews, 1975) takes into account all aforementioned indicators and essentially reflects the correlation between two binary variables (e.g., predicted recurrence vs. remission and observed recurrence vs. remission). An MCC value of 1 indicates perfect prediction, while an MCC of 0 suggests that warning signs occur equally frequently for individuals with recurrent depression and for individuals who remained in remission. We used permutation testing to test whether the MCC was significantly larger than 0. Specifically, we randomly shuffled the predicted recurrence/remission values and compared these with the observed recurrence/remission, leading to a new MCC value (Supplemental Figure S2). The p value equals the number of times the permutation MCC values were larger than the observed MCC value, divided by the total number of permutations (i.e., 10,000).

To evaluate the overlap of the *SD* and mean warning signs, we again computed the MCC. Here, an MCC of 1 indicates perfect overlap between both indicators, whereas an MCC of 0 would mean that the overlap between indicators does not differ from what would be expected by chance. Again, we used permutation tests to examine whether the MCC was significantly larger than 0. Here, we randomly shuffled the mean warnings signs and compared them to the observed *SD* warning signs. The p value was defined similarly as described above.

We also investigated whether predictive performance improves when combining warning signs. This was achieved by examining whether MCC values increase when both warning signs are considered (compared to MCC values that are based on single statistics, i.e., either the mean or *SD*). Here, we considered warning signs that occurred in the mean or the *SD*, or in both. Permutation tests were done to examine whether the MCC values from combined warning signs were significantly higher than the MCC values obtained when looking at mean-based warning signs only[5] (Supplemental Figure S2). To generate permuted data according to the null hypothesis (i.e., warning signs in the *SD* do not improve the prediction of depressive episodes beyond the mean), we first calculated the number of warning signs that were added by monitoring the *SD* on top of the mean. These warning signs were then randomly allocated to individuals without warning signs in the mean. We next compared the MCC values obtained through these permutations to the MCC value we observed in original data. The p value equals the number of times the permuted MCC value was larger than the observed MCC value, divided by the number of permutations (i.e., 10,000). Additionally, for individuals in whom we found both mean- and *SD*-based warning signs, we descriptively compared the timing of warning signs.[6] Specifically, we looked at when the first EWMA score exceeded the relevant control limit.

Transparency and Openness

Due to restrictions related to the European law regarding general data protection regulation, data are not freely available. However, data can be accessed by contacting the curating researcher (H. Riese). The preregistration and R code pertaining to this study are provided in Supplemental Material and posted on the open science framework at https://osf.io/4rcn3/. Analyses were performed in R (R Core Team, 2020), using the spc package (Knoth, 2020) to compute the L value of the control limit(s) and the qcc package (Strucca, 2004) to apply the EWMA procedure.

[5] Note that we made two comparisons: MCC both versus MCC mean and MCC both versus MCC *SD*. The permutations we applied for both comparisons follow an identical procedure, and we therefore only describe the permutations relevant for the first comparison.

[6] We preregistered a Mann–Whitney U test to compare the timing of mean- and *SD*-based warning signs, but deviated from this because of the small number of individuals for whom we detected both types of warning signs.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Results

We first provide a summary of the mean- and *SD*-based warning signs. Next, we zoom in on the added value of monitoring the *SD* on top of the mean, as indicated by the overlap between the two indices and by increases in predictive performance when monitoring the *SD* on top of the mean. Last, we look at the timing of the warning signs.

Mean- and *SD*-Based Warning Signs

Figure 2 illustrates the sensitivity, specificity, and MCC of the mean- and *SD*-based warning signs (see also Supplemental Table S1). For the latter indicator, the results are shown for both the one-sided tests (where warning signs reflect too high *SD*s) and two-sided tests (where warning signs reflect too high and too low *SD*s). For monitoring the *SD*, the MCC values of all emotions are higher for the one-sided test than for two-sided test. The one-sided *SD* test predicts the recurrence of depression with higher specificity, whereas the two-sided *SD* test leads to higher sensitivity. Moreover, the MCC values of the two-sided tests for all emotions were not significant. These results suggest that it is better to monitor increases in *SD*, rather than simultaneously monitoring increases and decreases. Therefore, in the remainder of this section we will only focus on the one-sided *SD* tests (i.e., too high *SD*s).

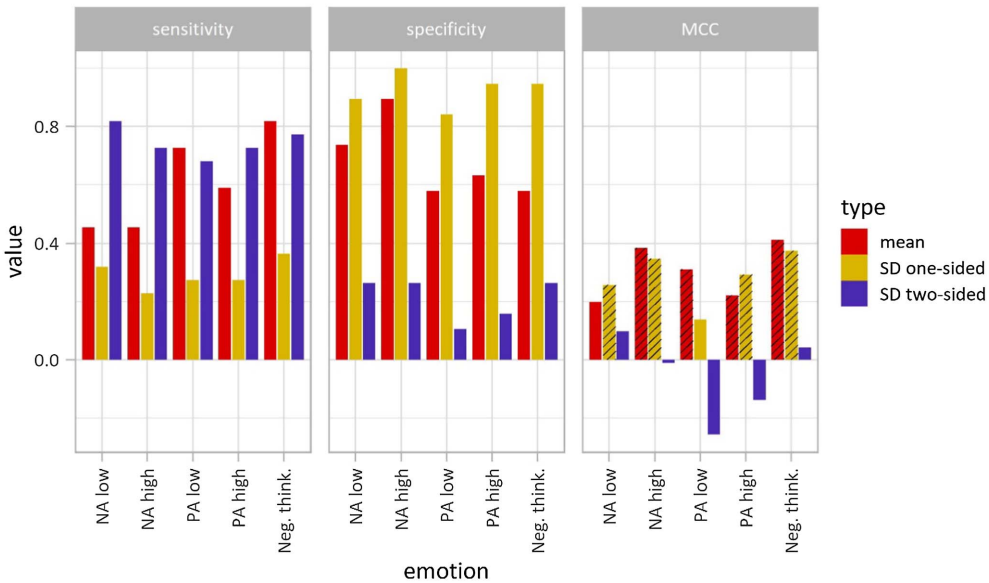
As Snippe et al. (2023) reported, recurrence of depression could be relatively accurately foreseen by mean-based warning signs in each emotion, with the best predictive performance (i.e., highest MCC values) for neg. think. and NA high arousal. Specifically, with the preregistered SPC settings, neg. think. had a high sensitivity and a moderate specificity, whereas NA high arousal had a high specificity and a moderate sensitivity. The MCC values of warning signs in the mean were significant for all emotions, except for NA low arousal. Similar to the warnings based on the mean, *SD*-based warnings in neg. think. and NA high arousal were the most accurate predictors with the highest MCC values. Both emotions had a high specificity and a low sensitivity. Again, the MCC values of warning signs in the *SD* were significantly different from 0, except for PA low arousal. When comparing the warning signs in the mean and *SD* that were obtained with the preregistered SPC procedure, we see that monitoring mean-based warning signs leads to higher sensitivity, whereas monitoring for warning signs in the *SD* leads to higher specificity. For example, a mean-based warning sign in neg. think. was detected in 18 of the 22 participants with recurrence (sensitivity 0.82) and in eight of 19 participants who remained in remission (specificity 0.58). For the *SD*, on the other hand, a warning sign in neg. think. was only detected in eight of 18 participants with recurrence (sensitivity 0.36) and in one of 19 participants who remained in remission (specificity 0.95). As a consequence of these different trade-offs between sensitivity and specificity, mean-based warning signs generally had a higher negative predictive value whereas *SD*-based warning signs had a higher positive predictive value.

Added Value of Monitoring the *SD*

Monitoring the *SD* is only worthwhile if *SD*-based warnings convey unique information concerning individuals' risk for depression, which we checked in three ways. First, we computed the within-person correlation between day means and day *SD*s. These correlations strongly varied between persons and were on

Figure 2

Sensitivity, Specificity, and MCC Values for Warning Signs in Mean Levels, too High SDs (SD One-Sided), and too High/Low SDs (SD Two-Sided)



Note. For MCC values, shaded bars denote that values were significantly larger than 0. MCC = Matthews correlation coefficient; NA = negative affect; PA = positive affect; Neg. think. = negative thinking. See the online article for the color version of this figure.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

average higher for negative compared to positive emotions, NA high = 0.564 (*SD* = 0.234); NA low = 0.357 (*SD* = 0.269); PA high = −0.086 (*SD* = 0.302); PA low = −0.263 (*SD* = 0.345); Neg. thinking = 0.380 (*SD* = 0.324); see Supplemental Figure S3 for a visualization. Second, we checked whether there was overlap in the warning signs in mean and *SD* by inspecting the MCC. The MCC value of NA low arousal was significant (0.209, $p = .045$), indicating that there was significant overlap between the indicators. The MCC values for the other emotions were not significantly different from 0, suggesting that there is unique information in the *SD*-based warning signs as compared to the mean-based warning signs. Third, and finally, we investigated whether predictive performance of warning signs improves when monitoring the *SD* on top of the mean. A warning sign was now defined as a detected change in at least one of the two indicators. The MCC values of all emotions were significant (Figure 3; Supplemental Table S1). Further, combining mean- and *SD*-based warning signs increased the sensitivity without affecting the specificity for NA high arousal and neg. think. For other emotions (NA low arousal, PA high arousal, PA low arousal), specificity decreased. We tested whether this trade-off was worth it by comparing MCC values and found that for all emotions except for PA low arousal, MCC values of the combined warning signs were higher than MCC values of warnings in either the mean or the *SD*. Specifically, for NA low arousal, the MCC value of the combined warning signs was significantly higher than the MCC value of the mean ($p = .035$), but not significantly higher than the MCC value of the *SD* ($p = .085$). For NA high arousal, PA high arousal and neg. think., the MCC value of the combined warning signs was significantly higher than the MCC values of the mean and *SD* (all $p < .015$). For PA low arousal, the MCC value was highest for the mean-based warning sign only, as compared to the combined warning signs and the *SD*-based warning sign only.

Timing of Warning Signs

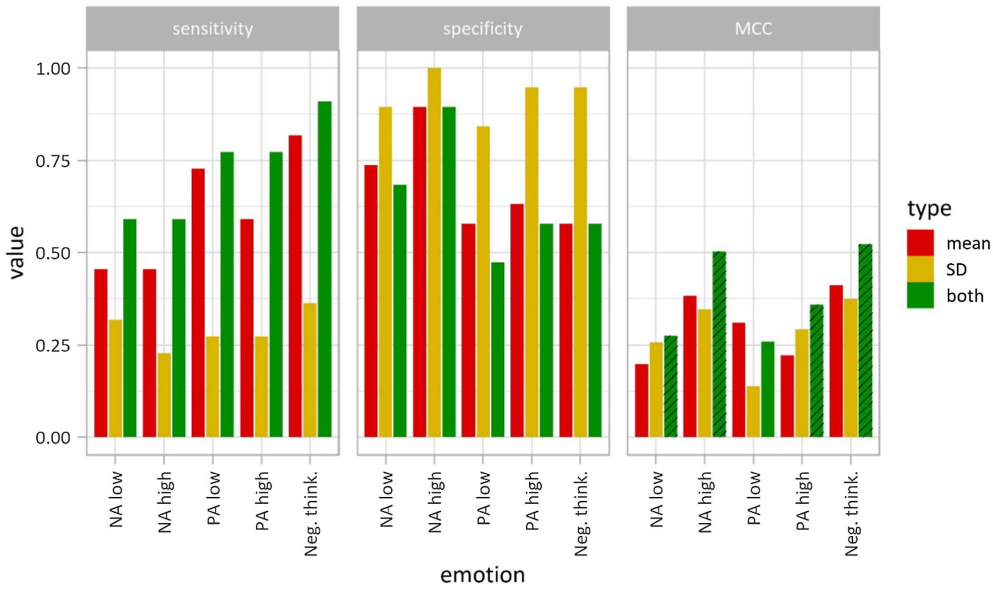
To compare the timing of the mean- and *SD*-based warning signs, we considered the individuals for whom depressive symptoms recurred. For each emotion, there was only a small number of individuals for whom both types of warning signs occurred (range 2–6). Within these samples, we found no consistent pattern regarding the timing of the warning signs. For some participants a mean-based warning sign occurred prior to a *SD*-based warning sign, and for other participants vice versa (see Supplement Material 5 for the results).

Sensitivity Analyses

Three types of sensitivity analyses (see Supplemental Figure S1 and methods) yielded largely similar findings (Supplement Material 4), with a few notable exceptions. Specifically, warning signs in mean levels of NA low arousal became significant when (a) statistics were computed across 2 days, (b) Phase I data were not winsorized, and (c) applying alternative SPC settings (namely, $\lambda = 0.05$ and 0.15). Instead, warning signs in mean levels of PA high arousal were no longer significant in the second and some of the third sensitivity analyses ($\lambda = 0.15$). For *SD*-based warnings, the only discrepancy with our main findings was that PA low arousal became a significant indicator of future depression in almost all sensitivity analyses,

Figure 3

Sensitivity, Specificity, and MCC Values for Warning Signs in Mean Levels, SDs (One-Sided), and a Combination of Mean Levels and SDs (Both)



Note. For MCC values, shaded bars denote that including warnings in the *SD* (on top of those in the mean) significantly improved the predictive performance of warnings. MCC = Matthews correlation coefficient; NA = negative affect; PA = positive affect; neg. think. = negative thinking. See the online article for the color version of this figure.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

which only strengthens our main findings (for further details, see Supplement Table S2).

Discussion

One of the primary causes of the considerable burden inflicted by depression is that episodes often recur. It is therefore crucial to prevent recurrence, and to this end, we need to foresee depressive episodes before they arise. Warning signs detected in repeatedly assessed emotions may aid this pursuit. Specifically, earlier studies reported that increasing risk of depressive symptom recurrence is anticipated by changes in mean levels of emotions (Smit et al., 2019; Smit & Snippe, 2023; Snippe et al., 2023). Recurrence of depression may additionally coincide with changes in the variability of emotions, and it was recently shown that such changes can be monitored in real time (Schat et al., 2023). The present study therefore aimed to explore whether real-time monitoring of emotional variability further improves the detection of nearby depression. We found that SPC can provide *SD*-based warning signs that can be used to foresee the recurrence of depression with low sensitivity and high specificity. Even though within-person means and *SD*s were correlated, warning signs in the *SD* did not show significant overlap with mean-based warning signs, with the exception of NA low arousal. Further, monitoring warning signs in the *SD* alongside the mean improved the early detection of recurrent depression for almost every emotion, except for PA low arousal. This was largely robust against differences in preprocessing or parameter settings. Thus, for the purpose of forecasting depression, it is useful to monitor not just emotional intensity but also variability. Particularly accurate warning signs occurred in neg. think. and NA high arousal, which is in line with an earlier study that used the same data, but only examined changing mean levels (Snippe et al., 2023). If future studies (with other data) replicate this, it could be that sampling a limited subset of items may already suffice for monitoring warning signs for depression through SPC. Whether mean- and *SD*-based warning signs differ in their timing remains an open question: The concurrence of warning signs was too seldom (and our sample size was too small) to statistically compare the timing of different warning signs. Finally, when monitoring the *SD*, only relatively high—not low—values should be considered warning signs. Taken together, real-time monitoring whether and when emotional variability reaches too high values may contribute to the early detection of recurrent depression.

We found that monitoring the *SD* on top of the mean is worthwhile, as it improves the discrimination between individuals with versus without recurrent depression. To illustrate, for all emotions except for PA low arousal, monitoring warning signs in both the *SD* and the mean (instead of the mean only) increased accuracy by 3%–7%, effectively meaning that 1–3 out of 41 individuals are better classified when warning signs in the *SD* are taken into consideration. Given that monitoring *SD*s on top of mean levels does not come at any additional costs in terms of data collection, it presents an efficient way of improving the sensitivity of mean-based warning signs. When using SPC to monitor indices in isolation, mean levels may be favored over *SD*s. This is because, in the current sample, the cost of a missed is likely higher than the cost of a false alarm, implying that high sensitivity (of mean-based warnings) may be more important than high specificity (of *SD*-based warnings). Importantly, however, the favorable trade-off between sensitivity and specificity of mean-based warnings can generally be further improved when additionally considering *SD*-based warning signs. Further, it should be noted that other methods (e.g., an alternative monitoring techniques or different SPC settings) may yield contrasting findings (e.g., favoring isolated warnings in the *SD* or mean above combined warnings).

For most emotions, we found little overlap between mean- and *SD*-based warning signs, with the exception of NA low arousal. Put differently, warning signs in the *SD* often occurred independently from those in the mean. A similarly low agreement between warning signs has been reported in earlier studies (F. M. Bos et al., 2022; Schreuder et al., 2023). Thus, although the mean and the *SD* are often correlated in EMA data (Mestdagh et al., 2018), which was also the case for the current data, warning signs in the mean and the *SD* may still convey unique information concerning an individual's vulnerability for recurrent depression. As a result, concurrent warning signs in the mean and the *SD* are relatively seldom, and it is therefore recommended to take appropriate action (e.g., implement a benign preventive intervention) when either metric (instead of both metrics) shows a warning sign.

Of note, the added utility of monitoring changes in the *SD* only appeared when we considered too high—but not too low—*SD*s as warning signs. This is in line with an earlier study using the same data, which reported an overall decline in *SD*s that was steeper in individuals who remained in remission than in individuals with recurrent depressive symptoms (Smit, Helmich, et al., 2023). Whether this decline reflects a shifting response tendency (e.g., becoming habituated to completing momentary questionnaires) or a phenomenon of substantive interest remains to be investigated. The selective focus on too high *SD*s as warning signs also fits well with the literature on critical slowing down and critical fluctuations in relation to depression (Cramer et al., 2016; Olthof et al., 2020; van de Leemput et al., 2014; Wichers & Groot, 2016). Briefly, this literature considers increasing *SD*s as a sign that a system is increasingly affected by external perturbations, and hence becomes increasingly vulnerable for regime shifts (Scheffer, 2009). In the context of psychology, rising trends in the *SD*s may thus mean that daily events have an increasing impact on momentary emotions, which in turn makes an individual vulnerable for worsening mental health (e.g., recurrence of depression). Empirical support for such rising trends in the *SD* as warning signs for depressive relapses has to date remained ambiguous (Curtiss et al., 2023; Kunkels et al., 2023; Olthof et al., 2020; Smit, Helmich, et al., 2023) and exclusively retrospective (i.e., yielding a single risk evaluation). By proposing a method for detecting real-time changes in the *SD*, with which each new observation (mood assessment) may yield a new risk evaluation (i.e., present or absent warning sign), the present study fills this gap.

Considerations and Future Directions

Even though monitoring *SD*-based warning signs on top of mean-based warning signs improves the predictive performance of SPC, the absolute sensitivity, specificity, and predictive values of warning signs still leave room for improvement. For instance, the overall accuracy of warning signs in the mean or the *SD*, or both, ranged between 63% and 76%. It follows that at least 24% of individuals were "falsely" classified using combined warnings (i.e., with

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

remission even though they experienced a recurrence, or with a recurrence even though they remained in remission). To rectify this, further tailoring SPC toward clinical psychology is needed. As an example, it could be that multivariate SPC methods, or monitoring alternative dynamics (e.g., autocorrelation), further improve the early detection of recurrent depression. Further, it could be that the type of emotion that is informative for future depression differs across individuals. It requires both simulation studies and large empirical studies to account for such possibilities.

Besides the imperfect accuracy of warning signs, there are several other pressing issues that should be resolved before SPC can be used as a clinical tool. First, the time interval at which warning signs should be considered is still unknown. For instance, it is currently undecided whether extreme values in the mean or *SD* that occur 3 months prior to a recurrent depression should be considered warning signs or instead false alarms. This not only complicates the evaluation of warning signs but also impacts eventual follow-up clinical decisions (e.g., with respect to the type of intervention that is appropriate). Second, monitoring several momentary emotions is more burdensome than, for instance weekly symptom assessments, and it should therefore be investigated whether—and for whom—the additional burden is worth it in terms of eventual treatment outcomes. It could be, for instance, that those who are tapering their medication have more to gain from monitoring warning signs in their emotions compared to other at-risk populations. Third, and finally, it is unknown whether interventions informed by SPC-derived warning signs indeed lower the incidence of recurrent depression. In conclusion, further research is needed in order to move from the current proof-of-concept to using SPC-derived warnings in clinical practice.

Limitations

A first limitation of the present study is that our sample size was relatively small ($N = 41$) and our Phase I period was relatively short (4 weeks), potentially threatening the generalizability and robustness of our findings. For instance, we found that applying alternative preprocessing steps (not winsorizing/computing statistics across 2 days) and alternative SPC parameters slightly affected our inferences (see Supplement Material 4), which likely results from the fact that our small sample and short Phase I period amplified the inferential consequences of minor analytical changes. Hence, in order to verify the robustness of our findings, replication in a larger, longer EMA study is necessary. At the same time, it should be noted that the data used in the present study were already unique in terms of sample size and duration, and it is therefore worthwhile to explore other methods for reaching robust results. For instance, alternative approaches for estimating control limits could overcome the necessity of collecting long-term Phase I EMA.

A related concern is that, so far, SPC has only been applied in the current data set. This means that we are uncertain whether this method for detecting real-time warning signs also works in other at-risk populations than patients who are tapering their antidepressants. Finally, due to the rare concurrence of mean- and *SD*-based warning signs and our sample size, we could not investigate differences in the timing of mean- versus *SD*-based warning signs. This topic thus warrants further investigation.

Conclusion

Depressive episodes tend to recur, and each new episode makes a future episode more likely. Preventing depressive episodes before they manifest is therefore crucial to reducing the burden of depression. A recent advancement in this direction is the introduction of SPC, which can yield real-time warning signs for impending depressive symptoms based on repeatedly assessed emotions and cognitions (Schat et al., 2021; Smit & Snippe, 2023; Snippe et al., 2023). The present study shows that such warning signs manifest not only in the mean of these emotions and cognitions but also in the *SD*. Monitoring the *SD* along with the mean may thus aid the early detection and prevention of depression. If replicated, and compared against alternative methods for detecting impending depression in real time, this supports the idea that SPC could aid the prevention of recurrent depression.

References

Andrews, G. (2008). Reducing the burden of depression. *Canadian Journal of Psychiatry*, *53*(7), 420–427. https://doi.org/10.1177/070674370805300703

Bean, C. (2023). *An examination of ruminative inertia and variability as predictors of change in depression and anxiety*. Kent State University.

Bos, E. H., de Jonge, P., & Cox, R. F. A. (2019). Affective variability in depression: Revisiting the inertia-instability paradox. *British Journal of Psychology*, *110*(4), 814–827. https://doi.org/10.1111/bjop.12372

Bos, F. M., Schreuder, M. J., George, S. V., Doornbos, B., Bruggeman, R., van der Krieke, L., Haarman, B. C. M., Wichers, M., & Snippe, E. (2022). Anticipating manic and depressive transitions in patients with bipolar disorder using early warning signals. *International Journal of Bipolar Disorders*, *10*(1), Article 12. https://doi.org/10.1186/s40345-022-00258-4

Cramer, A. O. J., van Borkulo, C. D., Giltay, E. J., van der Maas, H. L. J., Kendler, K. S., Scheffer, M., & Borsboom, D. (2016). Major depression as a complex dynamic system. *PLOS ONE*, *11*(12), Article e0167490. https://doi.org/10.1371/journal.pone.0167490

Crowe, E., Daly, M., Delaney, L., Carroll, S., & Malone, K. M. (2019). The intra-day dynamics of affect, self-esteem, tiredness, and suicidality in major depression. *Psychiatry Research*, *279*, 98–108. https://doi.org/10.1016/j.psychres.2018.02.032

Curtiss, J. E., Mischoulon, D., Fisher, L. B., Cusin, C., Fedor, S., Picard, R. W., & Pedrelli, P. (2023). Rising early warning signals in affect associated with future changes in depression: A dynamical systems approach. *Psychological Medicine*, *53*(7), 3124–3132. https://doi.org/10.1017/S0033291721005183

Dejonckheere, E., Mestdagh, M., Houben, M., Rutten, I., Sels, L., Kuppens, P., & Tuerlinckx, F. (2019). Complex affect dynamics add limited information to the prediction of psychological well-being. *Nature Human Behaviour*, *3*(5), 478–491. https://doi.org/10.1038/s41562-019-0555-0

Derogatis, L. R. (1977). *SCL-90-R Symptom Checklist-90-R administration, scoring, and procedures manual*. PsychCorp.

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

Franck, E., & De Raedt, R. (2007). Self-esteem reconsidered: Unstable self-esteem outperforms level of self-esteem as vulnerability marker for depression. *Behaviour Research and Therapy*, *45*(7), 1531–1541. https://doi.org/10.1016/j.brat.2007.01.003

Groot, P. C. (2010). Patients can diagnose too: How continuous self-assessment aids diagnosis of, and recovery from, depression. *Journal of Mental Health*, *19*(4), 352–362. https://doi.org/10.3109/09638237.2010.494188

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Groot, P. C., & van Os, J. (2018). Antidepressant tapering strips to help people come off medication more safely. *Psychosis*, *10*(2), 142–145. https://doi.org/10.1080/17522439.2018.1469163

Houben, M., & Kuppens, P. (2020). Emotion dynamics and the association with depressive features and borderline personality disorder traits: Unique, specific, and prospective relationships. *Clinical Psychological Science*, *8*(2), 226–239. https://doi.org/10.1177/2167702619871962

Houben, M., Van Den Noortgate, W., & Kuppens, P. (2015). The relation between short-term emotion dynamics and psychological well-being: A meta-analysis. *Psychological Bulletin*, *141*(4), 901–930. https://doi.org/10.1037/a0038822

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *59*(1), 12–19. https://doi.org/10.1037/0022-006X.59.1.12

Knoth, S. (2020). *spc: Statistical process control—Calculation of ARL and other control chart performance measures* (Version 0.6.4). https://cran.r-project.org/package=spc

Kossakowski, J. J., Groot, P. C., Haslbeck, J. M. B., Borsboom, D., & Wichers, M. (2017). Data from 'critical slowing down as a personalized early warning signal for depression.' *Journal of Open Psychology Data*, *5*(1), Article 1. https://doi.org/10.5334/jopd.29

Koval, P., Pe, M. L., Meers, K., & Kuppens, P. (2013). Affect dynamics in relation to depressive symptoms: Variable, unstable or inert? *Emotion*, *13*(6), 1132–1141. https://doi.org/10.1037/a0033579

Kunkels, Y. K., Smit, A. C., Minaeva, O., Snippe, E., George, S. V., van Roon, A. M., Wichers, M., & Riese, H. (2023). Risk ahead: Actigraphy-based early-warning signals of increases in depressive symptoms during antidepressant discontinuation. *Clinical Psychological Science*, *11*(5), 942–953. https://doi.org/10.1177/21677026221148101

Matthews, B. W. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta*, *405*(2), 442–451. https://doi.org/10.1016/0005-2795(75)90109-9

Mestdagh, M., Pe, M., Pestman, W., Verdonck, S., Kuppens, P., & Tuerlinckx, F. (2018). Sidelining the mean: The relative variability index as a generic mean-corrected variability measure for bounded variables. *Psychological Methods*, *23*(4), 690–707. https://doi.org/10.1037/met0000153

Montgomery, D. C. (2009). *Introduction to statistical quality control* (6th ed.). Wiley.

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

Olthof, M., Hasselman, F., Strunk, G., van Rooij, M., Aas, B., Helmich, M. A., Schiepek, G., & Lichtwarck-Aschoff, A. (2020). Critical fluctuations as an early-warning signal for sudden gains and losses in patients receiving psychotherapy for mood disorders. *Clinical Psychological Science*, *8*(1), 25–35. https://doi.org/10.1177/2167702619865969

Palmier-Claus, J. E., Myin-Germeys, I., Barkus, E., Bentley, L., Udachina, A., Delespaul, P. A. E. G., Lewis, S. W., & Dunn, G. (2011). Experience sampling research in individuals with mental illness: Reflections and guidance. *Acta Psychiatrica Scandinavica*, *123*(1), 12–20. https://doi.org/10.1111/j.1600-0447.2010.01596.x

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Schat, E., Tuerlinckx, F., De Ketelaere, B., & Ceulemans, E. (2023). Real-time detection of mean and variance changes in experience sampling data: A comparison of existing and novel statistical process control approaches. *Behavior Research Methods*. Advance online publication. https://doi.org/10.3758/s13428-023-02103-7

Schat, E., Tuerlinckx, F., Smit, A. C., de Ketelaere, B., & Ceulemans, E. (2021). Detecting mean changes in experience sampling data in real-time: A comparison of univariate and multivariate statistical process control methods. *Psychological Methods*. Advance online publication. https://doi.org/10.1037/met0000447

Scheffer, M. (2009). *Critical transitions in nature and society*. Princeton University Press. https://doi.org/10.1515/9781400833276

Schreuder, M. J., Hartman, C. A., Groen, R. N., Smit, A. C., Wichers, M., & Wigman, J. T. W. (2023). Anticipating transitions in mental health in at-risk youths: A 6-month daily diary study into early-warning signals. *Clinical Psychological Science*, *11*(6), 1026–1043. https://doi.org/10.1177/21677026221103138

Smit, A. C., Helmich, M. A., Bringmann, L. F., Oldehinkel, A. J., Wichers, M., & Snippe, E. (2023). *Critical slowing down in momentary affect as early warning signal of impending transitions in depression*. Preprint. https://www.osf.io/preprints/psyarxiv/u6vdg

Smit, A. C., Schat, E., & Ceulemans, E. (2023). The exponentially weighted moving average procedure for detecting changes in intensive longitudinal data in psychological research in real-time: A tutorial showcasing potential applications. *Assessment*, *30*(5), 1354–1368. https://doi.org/10.1177/10731911221086985

Smit, A. C., & Snippe, E. (2023). Real-time monitoring of increases in restlessness to assess idiographic risk of recurrence of depressive symptoms. *Psychological Medicine*, *53*(11), 5060–5069. https://doi.org/10.1017/S0033291722002069

Smit, A. C., Snippe, E., Bringmann, L. F., Hoenders, H. J. R., & Wichers, M. (2023). Transitions in depression: If, how, and when depressive symptoms return during and after discontinuing antidepressants. *Quality of Life Research*, *32*(5), 1295–1306. https://doi.org/10.1007/s11136-022-03301-0

Smit, A. C., Snippe, E., Kunkels, Y. K., Riese, H., Helmich, M. A., & Wichers, M. (2020). *Transitions in depression (TRANS-ID) tapering*. PsyArXiv. https://osf.io/h75p9/

Smit, A. C., Snippe, E., & Wichers, M. (2019). Increasing restlessness signals impending increase in depressive symptoms more than 2 months before it happens in individual patients. *Psychotherapy and Psychosomatics*, *88*(4), 249–251. https://doi.org/10.1159/000500594

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

Sperry, S. H., & Kwapil, T. R. (2022). Bipolar spectrum psychopathology is associated with altered emotion dynamics across multiple timescales. *Emotion*, *22*(4), 627–640. https://doi.org/10.1037/emo0000759

Sperry, S. H., Walsh, M. A., & Kwapil, T. R. (2020). Emotion dynamics concurrently and prospectively predict mood psychopathology. *Journal of Affective Disorders*, *261*, 67–75. https://doi.org/10.1016/j.jad.2019.09.076

Strucca, L. (2004). *qcc: An R package for quality control charting and statistical process control* (Version 2.7). https://cran.r-project.org/package=qcc

Torous, J., Staples, P., Shanahan, M., Lin, C., Peck, P., Keshavan, M., & Onnela, J.-P. (2015). Utilizing a personal smartphone custom app to assess the Patient Health Questionnaire-9 (PHQ-9) depressive symptoms in patients with major depressive disorder. *JMIR Mental Health*, *2*(1), Article e8. https://doi.org/10.2196/mental.3889

Trull, T. J., Lane, S. P., Koval, P., & Ebner-Priemer, U. W. (2015). Affective dynamics in psychopathology. *Emotion Review*, *7*(4), 355–361. https://doi.org/10.1177/1754073915590617

van de Leemput, I. A., Wichers, M., Cramer, A. O. J., Borsboom, D., Tuerlinckx, F., Kuppens, P., van Nes, E. H., Viechtbauer, W., Giltay, E. J., Aggen, S. H., Derom, C., Jacobs, N., Kendler, K. S., van der Maas, H. L. J., Neale, M. C., Peeters, F., Thiery, E., Zachar, P., & Scheffer, M. (2014). Critical slowing down as early warning for the onset and termination of depression. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(1), 87–92. https://doi.org/10.1073/pnas.1312114110

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

van Os, J., Verhagen, S., Marsman, A., Peeters, F., Bak, M., Marcelis, M., Drukker, M., Reininghaus, U., Jacobs, N., Lataster, T., Simons, C., Lousberg, R., Gülöksüz, S., Leue, C., Groot, P. C., Viechtbauer, W., Delespaul, P., & the ESM-MERGE Investigators PhD. (2017). The experience sampling method as an mHealth tool to support self-monitoring, self-insight, and personalized health care in clinical practice. *Depression and Anxiety*, *34*(6), 481–493. https://doi.org/10.1002/da.22647

Wichers, M., Groot, P. C., & the Psychosystems, ESM Group, EWS Group. (2016). Critical slowing down as a personalized early warning signal for depression. *Psychotherapy and Psychosomatics*, *85*(2), 114–116. https://doi.org/10.1159/000441458

World Health Organization. (2017). *Depression and other common mental disorders: Global health estimates*.

Received December 16, 2022
Revision received November 2, 2023
Accepted November 3, 2023 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Members of Underrepresented Groups:
Reviewers for Journal Manuscripts Wanted

If you are interested in reviewing manuscripts for APA journals, the APA Publications and Communications Board would like to invite your participation. Manuscript reviewers are vital to the publications process. As a reviewer, you would gain valuable experience in publishing. The P&C Board is particularly interested in encouraging members of underrepresented groups to participate more in this process.

If you are interested in reviewing manuscripts, please write APA Journals at Reviewers@apa.org. Please note the following important points:

- To be selected as a reviewer, you must have published articles in peer-reviewed journals. The experience of publishing provides a reviewer with the basis for preparing a thorough, objective review.
- To be selected, it is critical to be a regular reader of the five to six empirical journals that are most central to the area or journal for which you would like to review. Current knowledge of recently published research provides a reviewer with the knowledge base to evaluate a new submission within the context of existing research.
- To select the appropriate reviewers for each manuscript, the editor needs detailed information. Please include with your letter your vita. In the letter, please identify which APA journal(s) you are interested in, and describe your area of expertise. Be as specific as possible. For example, "social psychology" is not sufficient—you would need to specify "social cognition" or "attitude change" as well.
- Reviewing a manuscript takes time (1–4 hours per manuscript reviewed). If you are selected to review a manuscript, be prepared to invest the necessary time to evaluate the manuscript thoroughly.

APA now has an online video course that provides guidance in reviewing manuscripts. To learn more about the course and to access the video, visit http://www.apa.org/pubs/journals/resources/review-manuscript-ce-video.aspx.