Check for updates

Received: 12 September 2023 | Revised: 22 April 2024 | Accepted: 27 May 2024

DOI: 10.1002/mhs2.72

METHODOLOGICAL ARTICLE

WILEY

Warning signals for mental health problems in at-risk young adults may be informed by momentary emotions reported by the general population: A novel application of the principles of statistical process control

Marieke J. Schreuder[1] | Peter Kuppens[1] | Evelien Schat[1] | Peter de Jonge[2] | Catharina A. Hartman[3] | Eva Ceulemans[1]

[1]Quantitave Psychology and Individual Differences, KU Leuven, Leuven, Belgium

[2]Department of Developmental Psychology, University of Groningen, Groningen, Netherlands

[3]Interdisciplinary Center Psychopathology and Emotion Regulation, University of Groningen, University Medical Center Groningen, Groningen, Netherlands

Correspondence

Marieke J. Schreuder, Quantitave Psychology and Individual Differences, KU Leuven, Leuven, Belgium.
Email: marieke.schreuder@kuleuven.be

Funding information

Research Council KU Leuven, Grant/Award Number: C14/19; FWO, Grant/Award Number: 12AVE24N

Abstract

Statistical process control (SPC) was recently introduced as a method for detecting person-specific warning signals for mental ill-health. Such warning signals occur when a person's repeatedly assessed emotions exceed a control limit. This control limit should in principle be based on the same person's emotions in a healthy period. As such data are often unavailable, this preregistered study investigated whether general population data can be used instead to estimate control limits. We used data from the HowNutsAreTheDutch study, in which adults from the general population ($N = 746$) rated their emotions three times a day for 1 month. Based on these data, we computed control limits according to the exponentially weighted moving average (EWMA) and Shewhart SPC methods. Next, we investigated how often young adults with versus without persistent mental health problems from the TRAILS TRANS-ID study ($N = 100$)–who rated their emotions daily for 6 months–reported scores beyond these general population-based control limits. Generally, warning signals occurred more often in young adults with persistent mental health problems compared to healthy young adults ($p < 0.05$). The predictive performance of SPC did not consistently improve when control limits were conditioned on individuals' age, sex, and depressive symptoms, nor differ between methods (EWMA vs. Shewhart). The different emotions that were monitored, however, affected SPC performance, so that for most settings, warning signs in feeling tired and upset were worse for detecting mental-ill health compared to warning signs in other emotions (e.g., feeling nervous, relaxed, etc.). It follows that warning signs in individual's emotions can perhaps be monitored using relatively generic norms, derived from the general population, opening up new avenues for research and low-threshold clinical application.

KEYWORDS

at-risk mental state, ecological momentary assessments, experience sampling method, psychopathology, statistical process control, warning signals

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2024 The Author(s). *Mental Health Science* published by Wiley Periodicals LLC.

1 | INTRODUCTION

Mental disorders typically first emerge during childhood and adolescence, and thereafter remit and often recur (Caspi et al., 2020; Kessler et al., 2007). To illustrate, 45% of individuals meet criteria for a mental disorder at some point during childhood and adolescence, and the majority of them will re-experience mental health problems as adults (Caspi et al., 2020; Ormel et al., 2015). As such problems cause enormous burden to individuals and society at large, much might be gained from early intervention and prevention. Such efforts are likely most effective when implemented for the right person, at the right moment in time (Nahum-Shani et al., 2018). Thus, to prevent the onset and progression of mental disorders, it is necessary to know *for whom* and *when* mental health problems are impending.

The question *who* is at increased risk for mental disorders has received much attention, and led to the specification of risk profiles (Fusar-Poli et al., 2017). For instance, a meta-analysis showed that children with externalizing behavioral problems are 1.5 times more likely to develop depression as adults compared to children without such problems (Loth et al., 2014). Similarly, adulthood anxiety disorder, ADHD, and substance use disorder have been linked to externalizing problems during childhood (Copeland et al., 2009; Reef et al., 2011). Although such profiles always carry uncertainty, in the sense that they are not deterministic, they do provide some guidance with regard to the target population that could benefit most from preventive interventions. Knowing *when* these individuals are at increased risk for mental health problems has, however, only recently become within reach. By monitoring day-to-day fluctuations in individuals' moods through ecological momentary assessments (EMA), it is possible to detect abnormal moods or maladaptive behavior in real-time (Fisher & Soyster, 2021; Thompson et al., 2014; Yuen et al., 2018). Thus, EMA data might eventually allow for flagging at what moment in time a particular individual is vulnerable for mental health problems.

To derive warning signs that predict upcoming mental health problems from EMA data, it is necessary to adopt prospective methods. Such methods yield a new (updated) risk assessment each time new data comes in (i.e., with each new EMA observation). This makes it possible to monitor how risk for mental health problems changes over time, and potentially, to intervene when risk exceeds a certain threshold. Statistical process control (SPC) is a method that enables such prospective prediction (Montgomery, 2009; Roberts, 1959). SPC is a stepwise method that was developed for continuously monitoring production processes, but recently found its way to clinical psychology, where it can be used to monitor the ebb and flow of emotions (Schat, Tuerlinckx, De Ketelaere, et al., 2023; Smit & Snippe, 2022; Smit et al., 2019; Snippe et al., 2023). During phase I, a person's normal mood level and variability are determined based on EMA data collected during a healthy, normative period that should ideally last at least 50 days (Figure 1a,b; Schat, Tuerlinckx, Smit, et al., 2023).

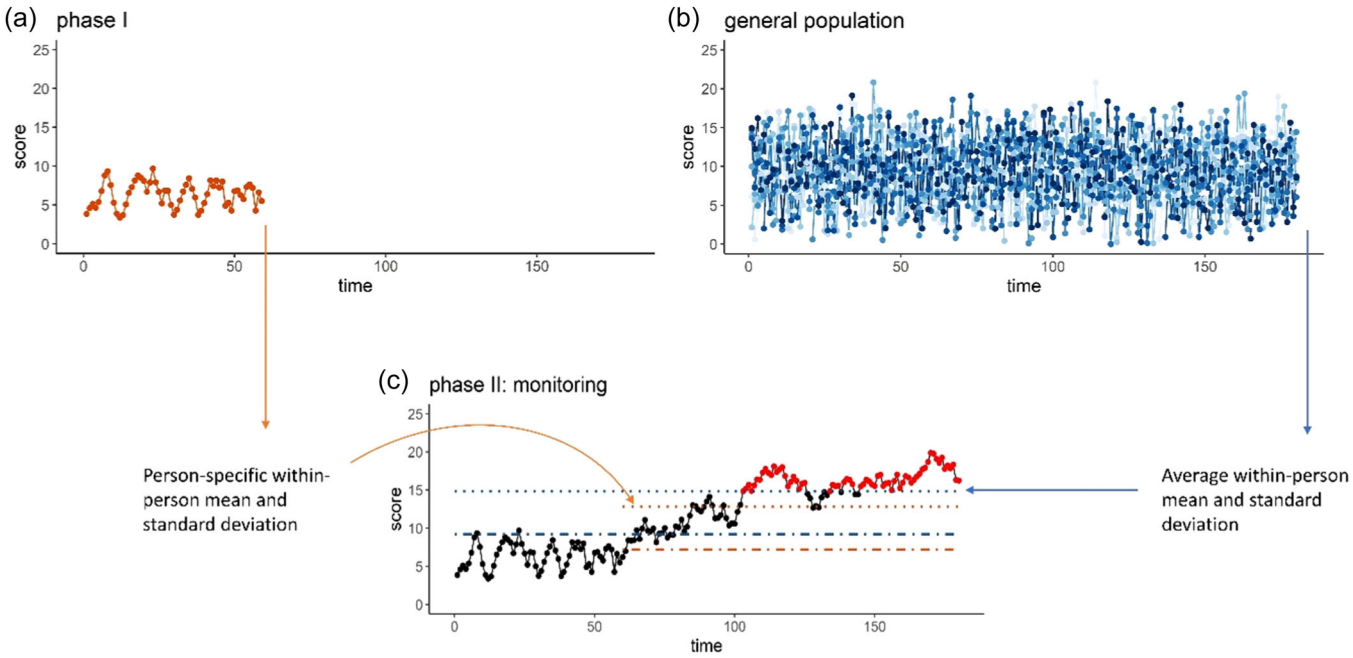


FIGURE 1 The original, one-sided statistical process control (SPC) procedure (a, c) and the one-sided statistical process control procedure proposed in the current manuscript (b, c). (a) Simulated phase I EMA data of a single individual (e.g., Alice). These data are often unavailable, but are illustrated here to clarify the SPC procedure. (b) Simulated general population EMA data, with a different color for each individual. (c) Originally, the control limit (orange dotted line) used for monitoring Alice's mental health is based on the person-specific mean (orange dot-dashed line) and standard deviation observed in phase I. In this manuscript, the control limit (blue dotted line) will be based on the average within-person mean and standard deviation form general population EMA data, yielding warning signs that are colored in red. This will in principle enable monitoring in absence of phase I EMA data.

26423588, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mhs2.72 by Ku Leuven, Wiley Online Library on [10/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

This yields so-called control limits. During phase II, which is the monitoring phase, it is determined whether incoming EMA observations are beyond those control limits – which implies that they deviate from what is normal for a specific person and emotion. Such out-of-control observations can be considered warning signs for upcoming changes in mental health. In two-sided applications of SPC, a monitored score is considered out of control if they are too high (>upper control limit) or too low (<lower control limit). In one-sided scenarios, which are relevant to the current manuscript, only a single control limit is used to determine whether an incoming observation is out of control. This implies, for instance, that feeling down only yields a warning signal if it is too high (and not when it is "too low") or that feeling enthusiastic only yields a warning signal if it is too low (and not when it is "too high"). Indeed, a case study and a replication thereof in a larger sample showed that depressive episodes can be forecasted by applying SPC to EMA data (Schat, Tuerlinckx, De Ketelaere, et al., 2023; Smit et al., 2019; Snippe et al., 2023). While a more transdiagnostic application of SPC is still lacking, this suggests that SPC might facilitate targeted and timely preventive interventions.

Although SPC could be a promising technique for deriving personalized warning signs for mental ill-health, it carries an important limitation: phase I is often insufficiently long or missing altogether (Schat et al., 2024). Indeed, if a clinician would wish to monitor warning signs in a vulnerable patient, he/she cannot go back in time to collect phase I EMA data for the respective patient in healthy times. Additionally, even if it would be possible to reach at-risk individuals when they are still healthy, collecting sufficient phase I EMA data comes with significant burden. Thus, establishing reliable and accurate person-specific control limits based on phase I EMA data will–in practical settings–often be impracticable or simply impossible. The current, preregistered study therefore set out to investigate whether these person-specific control limits based on phase I EMA data can be replaced by control limits based on EMA data from the general population. If possible, this greatly broadens the utility of SPC in clinical settings, as at-risk individuals could directly enter the monitoring phase (without necessitating phase I EMA data; Figure 1c,d).

The aim of this study was to investigate whether control limits derived from an EMA study in the general population (HowNutsAreTheDutch; Van der Krieke et al., 2016) can be used to distinguish at-risk individuals with distinct clinical profiles who completed a 6-month daily diary study (TRAILS TRANS-ID; Schreuder et al., 2020, 2023). Specifically, we investigated how often daily self-reported emotions of at-risk young adults with versus without persistent mental health problems exceeded control limits derived from the general population. It is noteworthy that this implies that the normative range of emotions is no longer considered fully person-specific (as in previous SPC applications), but rather, more or less generic. To explore the impact of imposing generality, we derived control limits with various degrees of personalization by conditioning on gender, age, and/or depressive symptomatology. We additionally compared the frequency of out-of-control emotions between two SPC procedures (exponentially weighted moving averages, or EWMA vs. Shewhart) and emotions. In doing so, we hope to investigate if and for which methodological variations SPC can be used to detect mental health problems in settings where person-specific phase I EMA data are not available. If phase I can indeed be circumvented, detecting worsening mental-health in time will be much more feasible.

2 | METHOD

We used EMA data and cross-sectional data collected in the HowNutsAreTheDutch study (Van Der Krieke et al., 2016) as well as diagnostic interview and EMA data collected in the TRAILS TRANS-ID study (Schreuder et al., 2020). Both studies are briefly described below, and sample characteristics are presented in Table 1.

TABLE 1 Sample characteristics.

	HND		TT					
			Full sample		Persistently healthy		Persistently ill	
N	746		122		74		26	
Age (mean ± SD)	39.41	±13.31	23.64*	±0.67	23.61	±0.67	23.76	±0.68
Male (%)	15.42		56.56*		55.41		50.00	
QIDS (mean ± SD)	6.75	±4.65						
Depressive disorder (%)			16.39		0		53.84	
Anxiety disorder (%)			7.38		0		26.92	

Note: The QIDS was not assessed in the TT study. For HND participants, diagnostic status was unknown.

Abbreviations: HND, HowNutsAreTheDutch; QIDS, quick inventory of depressive symptoms; SD, standard deviation; TT, TRAILS Trans-ID.

*Significant difference between the HND and TT sample ($p < 0.001$).

26423588, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mhs2.72 by Ku Leuven, Wiley Online Library on [10/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2.1 | How nuts are the Dutch

The HND study is a general population crowdsourcing study into mental health ($N = 12503$) and has been described elsewhere (Van Der Krieke et al., 2016). Briefly summarized, the study invited all Dutch individuals to take part in an EMA study. Individuals could take part in the HND study by registering themselves online, and thereby also confirmed that they were older than 18 years old, owned a smartphone, and consented to their data being used for research. In total, 1302 individuals expressed their interest in the EMA study. This included many individuals who dropped out relatively early and who possibly did not really have the intention to participate (a consequence of the crowdsourcing methods in this study[1]). We therefore selected those participants who completed at least 60% of the momentary assessments ($N = 746$). This preregistered criterion ensured that all participants had completed at least 54 assessments, which acknowledges recommendations concerning the optimal number of assessments to be included in phase I (Schat, Tuerlinckx, De Ketelaere, et al., 2023) without excluding too many participants.

At enrollment, participants completed questionnaires that assessed symptoms of depression and anxiety (see Measures). Subsequently, they entered the EMA study, which involved completing three daily questionnaires for 30 consecutive days (i.e., max. 90 questionnaires/person) on their smartphone (Van Der Krieke et al., 2017). Questionnaires were sent at fixed, equidistant moments in time (with 6-h intervals), with the exact time points depending on participant's sleep-wake schedule. Questionnaires were sent via Roqua, and contained questions on participants' mood, mindfulness, current activity and context. Participants completed these questionnaires on a visual analog scale (VAS) ranging from 0 (not at all) to 100 (very much). After completing the EMA study, participants received an overview of their own data, including a personalized network showing dynamic associations between items. The HND study was exempted from review by the Medical Research Involving Human Subjects Act (in Dutch: WMO) by the local Medical Ethical Committee (ref. M13.147422).

2.2 | TRAILS Trans-ID (TT)

The TT study was designed to monitor the progression of mental health problems in daily life, and is described in detail elsewhere (Schreuder et al., 2020). In brief, the TT study consists of a 6-month EMA study in young adults from a large, ongoing cohort study named TRAILS (TRacking Adolescents' Daily Lives Survey; Huisman et al., 2008; Oldehinkel et al., 2015). TRAILS includes a general population cohort and a clinical cohort, which are both assessed every two to 3 years. Individuals in the clinical cohort had been in contact with mental health care services before the age of 11 years, and are therefore considered vulnerable for mental health problems later in life. At the age of 22, individuals enrolled in the clinical cohort from TRAILS were invited to participate in the TT study. In total, 134 individuals entered the study, of whom 122 completed the EMA study (see below). Immediately before and after the EMA study, a semi-structured diagnostic interview was conducted (see Section Measures). At the start of the EMA study, 40 TT participants (32.78%) met criteria for a mental disorder. Major depressive disorders ($N = 25$) and anxiety disorders ($N = 21$) were most common, followed by attention-deficit hyperactivity disorder ($N = 8$). Of those who met diagnostic criteria at the start of the EMA study, some remitted during the EMA study, while others either developed additional problems or experienced no change in their mental health. Similarly, of those individuals who did not meet criteria for a mental disorder at the start of the EMA study ($N = 82$), some developed mental health problems while others remained healthy. This illustrates four distinct clinical profiles, namely: individuals who developed mental health problems during the EMA study ($N = 8$); individuals whose problems remitted during the EMA study ($N = 14$); individuals with persistent mental health problems (hereafter labeled persistently ill; $N = 26$); and individuals who were persistently healthy ($N = 74$). For the purposes of the current study, only persistently healthy and persistently ill individuals were selected ($N = 100$; Supporting Information S1: Supplement 2. Inclusion of participants). These participants were younger (Welch's t (771) = 32.03, $p < 0.001$), and more often male ($\chi^2(1) = 79.73$, $p < 0.001$) when compared to the HND participants (Table 1).

Every evening, for a period of six consecutive months, participants received a link to a questionnaire on their mobile phone. The questionnaires were sent via Roqua, and included questions on participants' mood during the past day (e.g. to what extent did you feel anxious today?). Participants could answer on a visual analog scale (VAS) ranging from 0 (not at all) to 100 (very much). After completing the EMA study, participants received a financial compensation and an overview of their own data. The TT study was approved by the local Medical Ethical Committee (ref. 207/103) and all participants provided written informed consent.

2.3 | Measures

2.3.1 | Baseline assessment (HND)

Before the EMA study, HND participants completed the Quick Inventory of Depressive Symptoms (QIDS; Rush et al., 2003). The QIDS is a 16-item questionnaire designed to assess the severity of depressive symptoms with acceptable to good psychometric properties (Reilly et al., 2015). Each item describes a symptom of depression that can be rated on a four-point scale (e.g., ranging from "I do not feel sad" to "I feel sad nearly all of the time"). Based on the total score, which can range from 0 to 27, depressive symptoms can be considered nonexistent (<5), mild (6–10), moderate (11–15), severe (16–20), or very severe (>21) (Rush et al., 2003).

[1]For further elaboration on the large number of interested, but noncomplying, individuals, please see Supporting Information S1: Supplement 2. Inclusion of participants.

26423588, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mhs2.72 by Ku Leuven, Wiley Online Library on [10/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2.3.2 | Diagnostic interview (TT)

Before and immediately after the EMA study, TT participants took part in a semi-structured diagnostic interview. The short version of the Schedules for Clinical Assessment in Neuropsychiatry (mini-SCAN) (Nienhuis et al., 2010) was administered by trained interviewers, and served to evaluate whether participants had met the diagnostic criteria for a mental disorder (as listed in the fifth version of the Diagnostic and Statistical Manual; American Psychiatric Association, 2013) in the month before the EMA study or during the EMA study.

2.3.3 | Ecological momentary assessments

Analyses were restricted to the 11 items that were assessed in both HND and TT. These items were: feeling energetic, feeling relaxed, feeling appreciated, feeling that one's life is worth living, being physically active, feeling down, feeling restless, feeling lonely, feeling tired, feeling easily upset, sleeping well. A list of all items that were assessed in HND and TT is provided online (www.osf.io/hvx2r). Although not all assessed items can be considered emotions, we will refer to them as such throughout this manuscript for readability.

2.4 | Analyses

We focused on two distinct SPC methods, which were both applied to psychological EMA data before (Schat, Tuerlinckx, De Ketelaere, et al., 2023; Schat, Tuerlinckx, Smit, et al., 2023; Smit & Schat, & Ceulemans, 2022; Snippe et al., 2023), namely the Shewhart method (Shewhart, 1931) and the exponentially weighted moving average (EWMA; Roberts, 1959) method. These methods follow the same general principles, but differ in the type of score that is monitored and in their assumptions. Briefly summarized, the Shewhart method monitors raw observations and therefore requires less tuning than the EWMA method, which monitors smoothed observations. A related benefit of the Shewhart method is that out-of-controls are more intuitive to interpret, as they indicate a significant deviation in one's current mood (as opposed to the exponentially weighted moving average of one's current mood). However, the Shewhart method performs poorly when observations are not normally distributed (which is often the case in EMA data) and is relatively insensitive to smaller shifts (Jones, 2002; Montgomery, 2009). Consequentially, small but meaningful changes in mood (e.g., mild dysthymia) may go undetected. In contrast, the EWMA method is relatively robust against deviations from normality and is suitable for detecting smaller shifts in mood (Montgomery, 2009; Schat, Tuerlinckx, De Ketelaere, et al., 2023). This method tracks an estimate that is based on both the current observation and previous observations, thus taking into account the history of a process. The weight given to the most recent score is given by a constant (λ), which can range between 0 and 1. With smaller values of λ, smaller shifts can be detected. In simulation studies, the EWMA method generally outperforms the Shewhart method (Roberts, 1959; Schat, Tuerlinckx, De Ketelaere, et al., 2023), and is therefore more frequently used. For the purpose of detecting mental ill-health, however, it is still largely unknown which method is most reliable, and we therefore compared the performance of both methods.

TABLE 2 Overview of control limits.

	EWMA (Roberts, 1959)	Shewhart (Shewhart, 1931)
Monitored score (z) at time t	$z_{iet} = \lambda x_{iet} + (1 - \lambda) z_{iet-1}$	$z_{iet} = x_{iet}$
Control limit (CL)	$CL_{ie} = \hat{\mu}_{ie} \pm L\hat{\sigma}_{ie}\sqrt{\frac{\lambda}{2-\lambda}}$	$CL_{ie} = \hat{\mu}_{ie} \pm L\hat{\sigma}_{ie}$
Settings	λ = 0.05, L = 2.42	L = 2.78

Note: Monitored scores and control limits were computed for each person (i) and each emotion (e) separately. Here, x_{iet} reflects person i's rating of emotion e at time t, $\hat{\mu}_{ie}$ reflects person i's mean rating of emotion e, and $\hat{\sigma}_{ie}$ reflects person i's standard deviation in emotion e.

Analyses consisted of the following steps, which are described in detail below: (1) compute one-sided control limits for each emotion with varying degrees of personalization using the Shewhart and EWMA methods based on the HND data, (2) examine how frequently monitored scores of at-risk young adults with versus without persistent mental health problems yielded warning signals when applying these control limits to the TT data, and (3) assess the performance of SPC by examining how well young adults with versus without persistent mental health problems can be distinguished based on their warning signals.

2.4.1 | Computation of control limits

We computed univariate, one-sided control limits. Generally, these control limits are based on an estimate of the mean ($\hat{\mu}_{ie}$) and standard deviation ($\hat{\sigma}_{ie}$) of daily emotion *e* of person *i*–which are estimated based on phase I EMA data of this person. Constant L is multiplied with $\hat{\sigma}_{ie}$, and thus defines the width of the control limit (Montgomery, 2009; Schat, Tuerlinckx, De Ketelaere, et al., 2023). As shown in Table 2, the Shewhart method and the EWMA method differ in the score that is being monitored. Specifically, the Shewhart method tracks raw observations while the EWMA method monitors exponentially weighted moving averages (first row of Table 2), which also impacts how control limits should be computed (middle row of Table 2). Based on earlier recommendations (Schat, Tuerlinckx, De Ketelaere, et al., 2023), we will set λ to 0.05 and L to 2.14 (EWMA method) and 2.78 (Shewhart method[2]; final row of Table 2). With these settings, a false warning signal (i.e., an out-of-control emotion in a healthy individual) should only occur after on average 370 observations (Montgomery, 2009; Schat, Tuerlinckx, De Ketelaere, et al., 2023).

As is often the case, phase I EMA data were not available in the TT study. We therefore estimated $\hat{\mu}_{ie}$ and $\hat{\sigma}_{ie}$ based on the HND data

[2]See Supporting Information S1: Supplement 1, deviation from preregistration.

26423588, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mhs2.72 by Ku Leuven, Wiley Online Library on [10/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 3 Overview of different degrees of personalization.

		$\hat{\mu}_{ie}$–center line around which emotions are expected to fluctuate (within-person mean)		$\hat{\sigma}_{ie}$–fluctuations in emotions (within-person standard deviation)	
Setting	**Degree of personalization**	***Predictors***	***Sample***	***Predictors***	***Sample***
1	HND	–	HND	–	HND
2	HND sex-specific	Sex	HND	Sex	HND
3	HND age-specific	Age	HND	Age	HND
4	HND sex- and age-specific	Sex, age	HND	Sex, age	HND
5	HND person-specific (QIDS = 3)	Sex, age, $QIDS_3$	HND	Sex, age, $QIDS_3$	HND
6	HND person-specific (QIDS = 6)	Sex, age, $QIDS_6$	HND	Sex, age, $QIDS_6$	HND
7	TT	Sex, age	TT[a]	Sex, age	TT[a]

Note: This table shows the characteristics and the sample used to predict $\hat{\mu}_{ie}$ and $\hat{\sigma}_{ie}$, with setting 1 reflecting the most generic setting and setting 7 reflecting the most personalized setting. $QIDS_3$ and $QIDS_6$ denote that we assumed QIDS scores of 3 (few depressive symptoms) and 6 (mild depressive symptoms) for TT participants–which was necessary because the QIDS was not administered in the TT study.

Abbreviations: HND, HowNutsAreTheDutch; QIDS, quick inventory of depressive symptoms; TT, TRAILS Trans-ID.

[a]TT participants ($N = 74$) who were persistently healthy.

using different degrees of personalization.[3] As shown in Table 3 (setting 1), the most generic approach we took involved applying the same, HND-based control limits for every person in the TT sample. To this end, we first computed within-day averages for each HND participant and emotion, and subsequently computed the mean and standard deviation of these within-day averages. This was done to match the sampling frequency (and hence, the dynamics) of the HND EMA study (3x/day) to the sampling frequency used in the TT EMA study (1x/day). Next, we explored several ways of personalizing these generic control limits. Specifically, we fitted robust linear regression models to the HND data where we predicted within-person means ($\hat{\mu}_{ie}$) and standard deviations ($\hat{\sigma}_{ie}$) by participants' sex (setting 2), age (setting 3), sex and age (setting 4), or sex, age, and depressive symptoms as indexed by their QIDS score (settings 5 and 6) in the HND data. These regression models were fitted using the robustbase package (version 0.99-2) in R (version 4.2) (Maechler et al., 2022; R Core Team, 2020), using settings recommended by the package authors (i.e., MM-type estimation with the defaults suggested by Koller and Stahel (Koller & Stahel, 2011)). Next, we used the predicted $\hat{\mu}_{ie}$ and $\hat{\sigma}_{ie}$ for TT participants, based on their sex and age, as input for the control limits. Since the QIDS was not administered in the TT study, we assumed a QIDS score of 3 (setting 5) and 6 (setting 6) for all participants. The thus-obtained control limits reflect the boundaries of in-control emotions, taking into account individuals' sex and age, and assuming that they have few depressive symptoms (QIDS = 3, setting 5) or relatively mild depressive symptoms (QIDS = 6, setting 6) (Rush et al., 2003). Even after taking into account sex, age, and depressive symptom severity, HND-based control limits may not accurately capture in-control emotions for the TT sample. Hence, in setting 7, we aimed to further personalize control limits by defining $\hat{\mu}_{ie}$ and $\hat{\sigma}_{ie}$ as the average within-person mean and standard deviation of persistently healthy TT participants[4] (N = 74), taking into account sex- and age-related differences in these estimates. The reasoning here was that control limits should resemble the range of scores that characterizes healthy individuals, so that individuals with worsening mental health can easily be detected. An illustrative example of the different control limits that follow from different settings is provided in Figure 2 and in Supporting Information S1: Supplement Figure S4.

2.4.2 | Frequency of warning signals

We applied the different control limits to the TT participants who were persistently healthy or had persistent mental health problems. We next examined the occurrence of warning signals, defined as monitored scores beyond the control limit. The frequency of warning signals was computed as the number of warning signals relative to the total number of assessments for a specific item and individual. Warning signals were expected to be more frequent in individuals with persistent mental health problems compared to persistently healthy individuals.

2.4.3 | Performance of SPC

If SPC would consistently show more warning signals in those with persistent mental health problems compared to those without

[3]See Supporting Information S1: Supplement 1, deviation from preregistration.

[4]To prevent overlap between the sample used for estimating $\hat{\mu}_{ie}$ and $\hat{\sigma}_{ie}$ (and thereby, the control limits) and the sample used for monitoring warning signs, we applied a leave-one-out procedure when monitoring healthy TT participants. Thus, for each healthy TT participant, we estimated control limits by regressing the within-person mean and standard deviation of *all other healthy TT participants* ($N = 73$) on their sex and age, and subsequently used the resulting model to estimate $\hat{\mu}_{ie}$ and $\hat{\sigma}_{ie}$.

26423588, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mhs2.72 by Ku Leuven, Wiley Online Library on [10/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

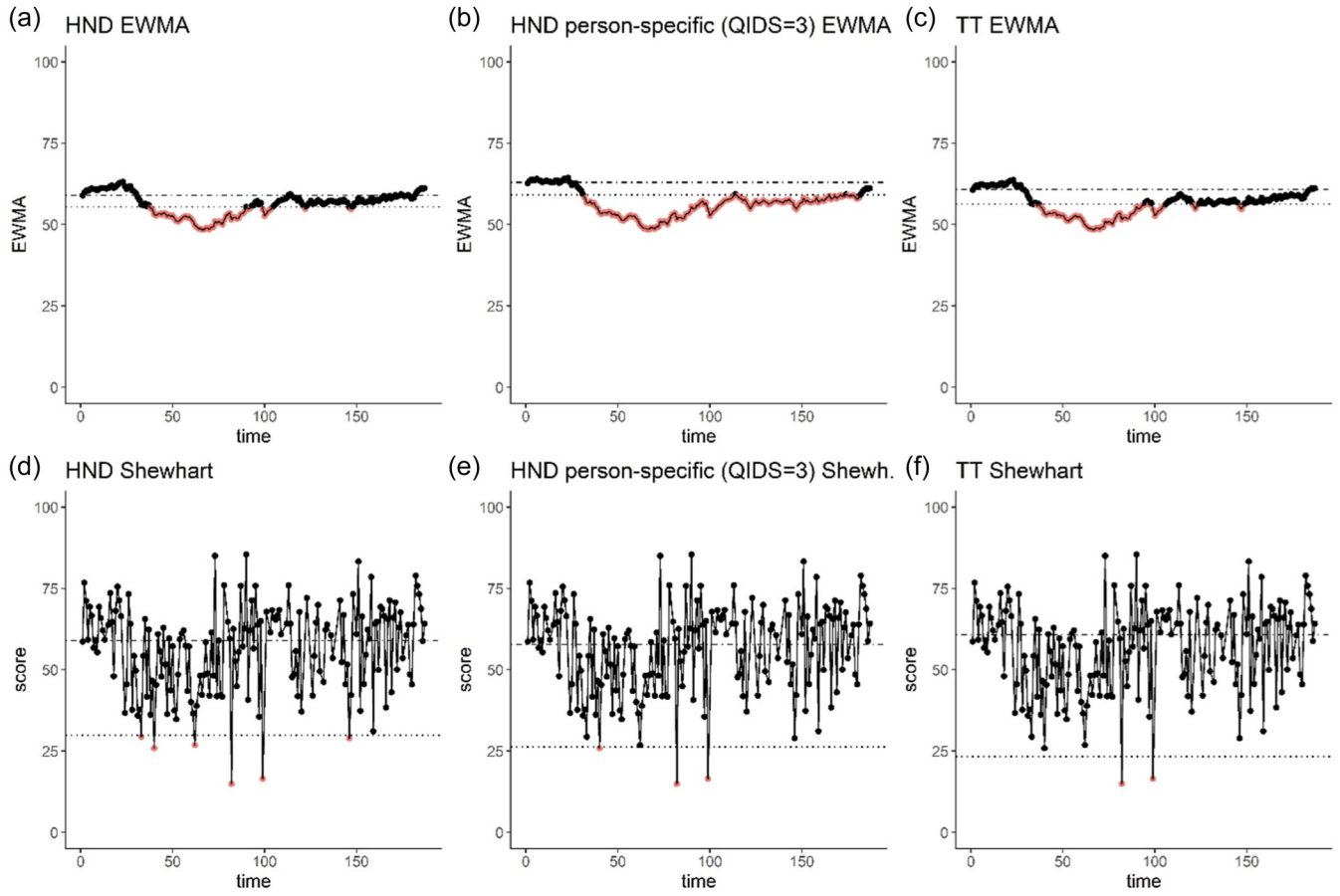


FIGURE 2 Illustration of three (out of seven) different personalization settings, based on ratings on feeling appreciated from a persistently healthy individual. Dotted lines reflect the control limit, while the dashed line reflects the estimated within-person mean. The upper three plots show the EWMA charts with different personalization levels, namely: (a) none at all, meaning that the control limit (lower horizontal line) is based on the average within-person mean and standard deviation in the HND sample, (b) moderate personalization, meaning that the control limit is based on the HND sample while taking into account sex-, age- and depression-related differences in control limits, and (c) high personalization, meaning that the control limit is based on TT participants who were persistently healthy and taking into account their sex and age. The lower three plots show the Shewhart charts with (d) no personalization, (e) moderate personalization, and (f) high personalization. Supporting Information S1: Supplement Figure S4 illustrates how control limits (and thereby, warning signals) vary across all seven personalization settings.

problems, it should be possible to predict individuals' mental health status (i.e., group membership) based on the frequency of warning signals. As an example, imagine that we consider a person ill if EWMA-based warning signals occur at least 50% of the time (i.e., at least 50% of the monitored scores are beyond the control limit). With this (arbitrary) decision rule, the control charts illustrated in Figure 2a,c, which show warning signals 33.16% and 38.50% of the time, would be considered reflective of a healthy person. In contrast, the control chart in Figure 2b would signify persistent illness, as 79.14% of monitored scores are out of control. It follows that the decision rule–that is, the minimal frequency of warning signals considered indicative of persistent illness–will impact the prediction of individuals' mental health status, and hence, the performance of SPC. We therefore systematically varied the decision rule from very liberal (threshold at 5%) to very conservative (threshold at 95%) in steps of 5%, and explored the impact thereof on the performance of SPC. In line with earlier SPC studies (Snippe et al., 2023), we quantified the performance of SPC by computing Matthews correlation coefficient (MCC, also known as phi coefficient) (Matthews, 1975). Calculating the MCC boils down to computing the correlation between two binary variables (prediction: ill/healthy vs. reality: ill/healthy), which ranges between −1 (implying that the prediction of mental health status is always wrong) and 1 (implying perfect prediction). An MCC of 0 reflects chance-level prediction. We used permutation tests to determine whether the MCC significantly differed from 0. We additionally computed the accuracy of our prediction (i.e., percentage of individuals who were correctly classified), the sensitivity/specificity, and positive/negative predictive values (for further details on the latter parameters, as well as the permutation procedure, see Supporting Information S1: Supplement 8. Performance of SPC: sensitivity, specificity, predictive values, and accuracy).

We investigated the performance of SPC–reflected in the MCC–across different methods (EWMA vs. Shewhart), personalization settings, emotions, and decision rules. Next, for the optimal combination of aforementioned methodological variations, we

26423588, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mhs2.72 by Ku Leuven, Wiley Online Library on [10/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

investigated how quickly SPC yields warning signals in persistently healthy and ill young adults. To this end, we monitored the frequency of warning signals in windows of 2 weeks (i.e., max. 14 observations), which matches the window size used in earlier studies into warning signs for changes in mental health (although these studies used other methods than SPC; Bos et al., 2022; Curtiss et al., 2021; Helmich et al., 2022; Ludwig et al., 2024; Schreuder et al., 2022). We then identified the earliest occasion at which the given decision rule was met, and expected that this timing would, on average, be earlier for persistently ill young adults compared to healthy young adults. However, as monitoring percentages is not straightforward–for example, dependent on an arbitrarily chosen window size – this comparison should be cautiously interpreted, and merely serves to illustrate how our results could also be applied prospectively.

3 | RESULTS

3.1 | EMA descriptives

Out of 90 assessments, HND participants ($N = 746$) completed between 54 and 87 (mean = 71, SD = 8) assessments. Persistently healthy ($N = 74$) and ill TT participants (N = 26) were prompted daily for six consecutive months (i.e., max. 183 assessments), and completed around 159 assessments on average (range = 113–183, SD = 18). Generally speaking, HND participants reported similar, or somewhat higher levels on positively valenced EMA items (e.g., feeling appreciated) compared to persistently healthy TT participants, while persistently ill TT participants reported lower levels on these items (Table 4). With regard to negatively valenced EMA items (e.g, feeling lonely), in contrast, HND participants were more similar to persistently ill TT participants, and reported higher levels compared to persistently healthy TT participants (Table 4, Supporting Information S1: Supplement Figure S3).

3.2 | Computation of control limits

As illustrated in Figure 3, the Shewhart method yields more extreme control limits (i.e., lower for positive emotions and higher for negative emotions) compared to the EWMA method. This is due to a difference in the type of score that both methods monitor (Shewhart: raw data, EWMA: weighted averages). The effect of taking into account sex-related differences in control limits varies across emotions: while control limits of feeling energetic and feeling tired clearly differ between males and females, control limits of other emotions are more similar. This means that, for instance, the same high rating of feeling tired (e.g. around 50 for the EWMA and 80 for the Shewhart method) is considered a warning signals for males, but not for females. In a similar vein, accounting for age did not

TABLE 4 EMA descriptives.

	HowNutsAreTheDutch (HND)		TRAILS Trans-ID (TT)				
			Persistently healthy		Persistently ill		
	N = 746		*N = 74*		*N = 26*		
EMA compliance (%)	78.90	±9.13	87.60	±9.81	86.60	±11.90	[ab]
EMA ratings							
Relaxed	60.13	±11.70	54.76	±12.91	40.42	±17.47	[abc]
Energetic	49.44	±12.57	51.65	±14.02	38.70	±14.71	[bc]
Appreciated	59.00	±14.00	61.33	±14.32	45.31	±19.91	[bc]
Life worth living	70.54	±18.32	65.58	±11.89	51.08	±17.96	[abc]
Phys. active	35.47	±12.28	49.19	±15.03	39.03	±14.00	[ac]
Slept well	62.99	±11.77	58.11	±11.48	46.92	±15.21	[abc]
Tired	45.79	±16.14	31.31	±14.93	39.69	±14.86	[a]
Nervous	30.89	±16.66	15.32	±11.19	26.55	±14.69	[ac]
Down	23.27	±15.32	11.75	±10.85	23.75	±15.98	[ac]
Upset	21.60	±15.48	11.02	±10.31	16.27	±11.52	[a]
Lonely	22.97	±17.55	11.10	±11.12	22.74	±15.00	[ac]

Note: Numbers reflect mean ± standard deviation unless otherwise noted. EMA ratings reflect within-person means and the standard deviation thereof (for a bar plot of these statistics, please see Supporting Information S1: Supplement Figure S3). For HND participants, we first averaged within days, to match the sampling frequency of the TT EMA study. [a]Significant difference between HND sample and persistently healthy young adults ($p < 0.05$). [b]Significant difference between HND sample and persistently ill young adults ($p < 0.05$). [c]Significant difference between persistently healthy and ill young adults ($p < 0.05$).

Abbreviations: EMA, ecological momentary assessments; phys., physically.

26423588, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mhs2.72 by Ku Leuven, Wiley Online Library on [10/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

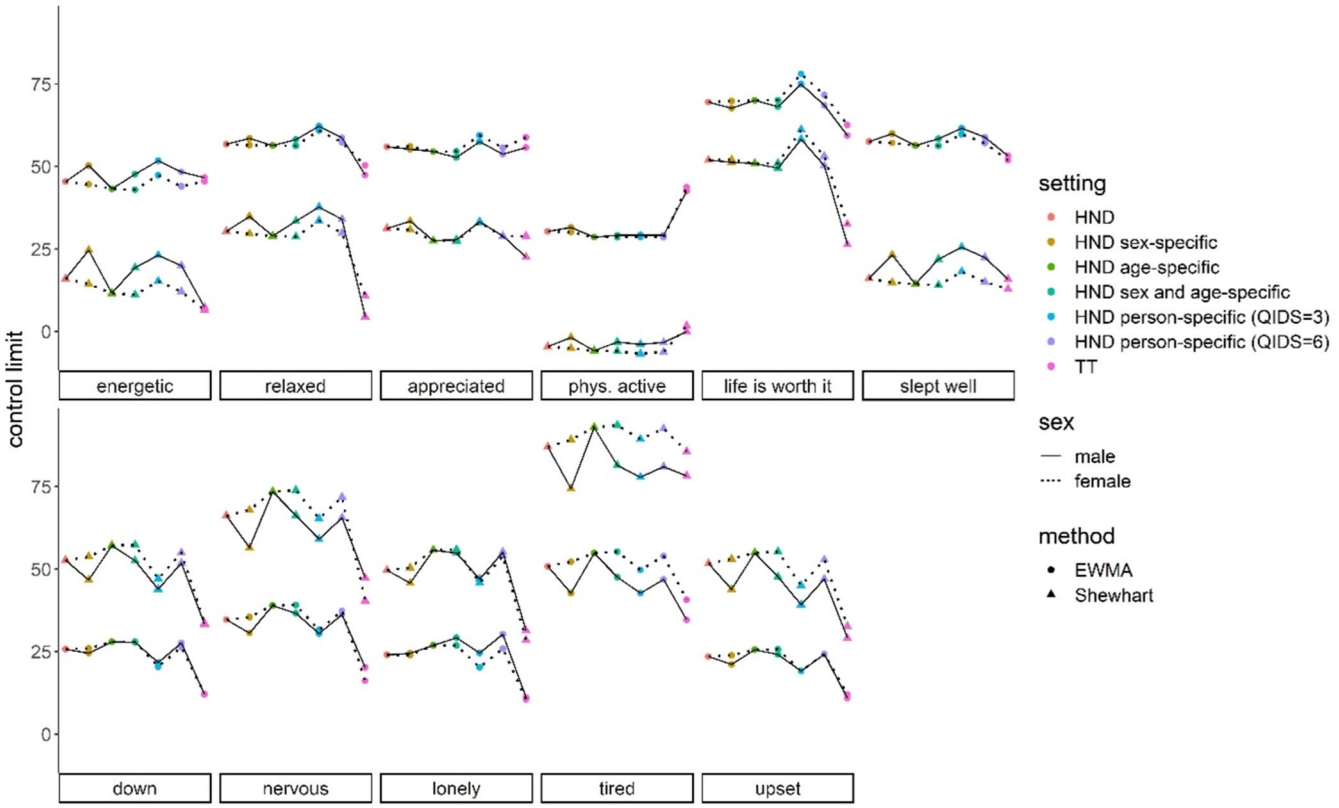


FIGURE 3 Control limits for all seven personalization settings ranging from fully generic (left, colored in red) to more person-specific (right, colored in pink) for positive and negative emotions (upper and lower row). For positive emotions, lower control limits are plotted: monitored scores below this value reflect warning signals. For negative emotions, upper control limits are plotted: monitored scores above this value reflect warning signals. When sex is not taken into account (as in settings 1, HND, and 3, HND age-specific), control limits for males (solid line) and females (dotted line) are similar.

substantially affect the control limits of settings in which the HND sample was used to derive control limits. Instead, in the TT setting, accounting for age had a larger effect (see Supporting Information S1: Supplement 5. Effect of age on control limits, for details). Taken together, increasing the degree of personalization did affect control limits, but not in a consistent way (i.e., there was no systematic in- or decrease of control limits with increasing personalization).

3.3 | Frequency of warning signals

We obtained warning signals for each individual and emotion, using the control limits that were based on either EWMA or Shewhart method and one of the seven personalization settings. For readability, the current section only describes the frequency of warning signals obtained when control limits were based on the HND sample (setting 5 in Table 3, which takes into account sex and age and assumes QIDS = 3; findings for other personalization settings are provided in Supporting Information S1: Supplement 7. Frequency of warning signals across personalization settings). To illustrate the results on the level of specific individuals, Figure 4 shows warning signals in feeling appreciated in an individual who met criteria for a depressive disorder both before and during the EMA study, and was hence considered persistently ill (Figure 4a,b), and in a persistently healthy individual (Figure 4c,d). Here, such warning signals indicate that individuals feel less appreciated than what would be considered "normal". In line with our hypothesis, the monitored scores of the persistently ill individual quickly and often cross the lower control limit compared to the monitored scores of the persistently healthy individual (EWMA: 96.8% vs. 0%, Shewhart: 32.6% vs. 0%).

Across individuals, the frequency of warning signals varied depending on the emotion that was monitored and the SPC method that was used (Figure 5; see Supporting Information S1: Supplement 7 for variation across personalization settings). As hypothesized, individuals with persistent mental health problems had more warning signals compared to individuals who were persistently healthy (Figure 5 and Supporting Information S1: Supplement Table S2). For some emotions, however, the difference between groups was not statistically significant (e.g. feeling upset, feeling tired). Further, some emotions showed many warning signals for persistently healthy individuals (e.g. feeling life is worth it; Shewhart method: 29.1%, EWMA method: 76.0%). This means that, on average, individuals who are healthy report scores that are outside the normative range–as defined by the HND sample–approximately 29.1% of the time.

26423588, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mhs2.72 by Ku Leuven, Wiley Online Library on [10/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

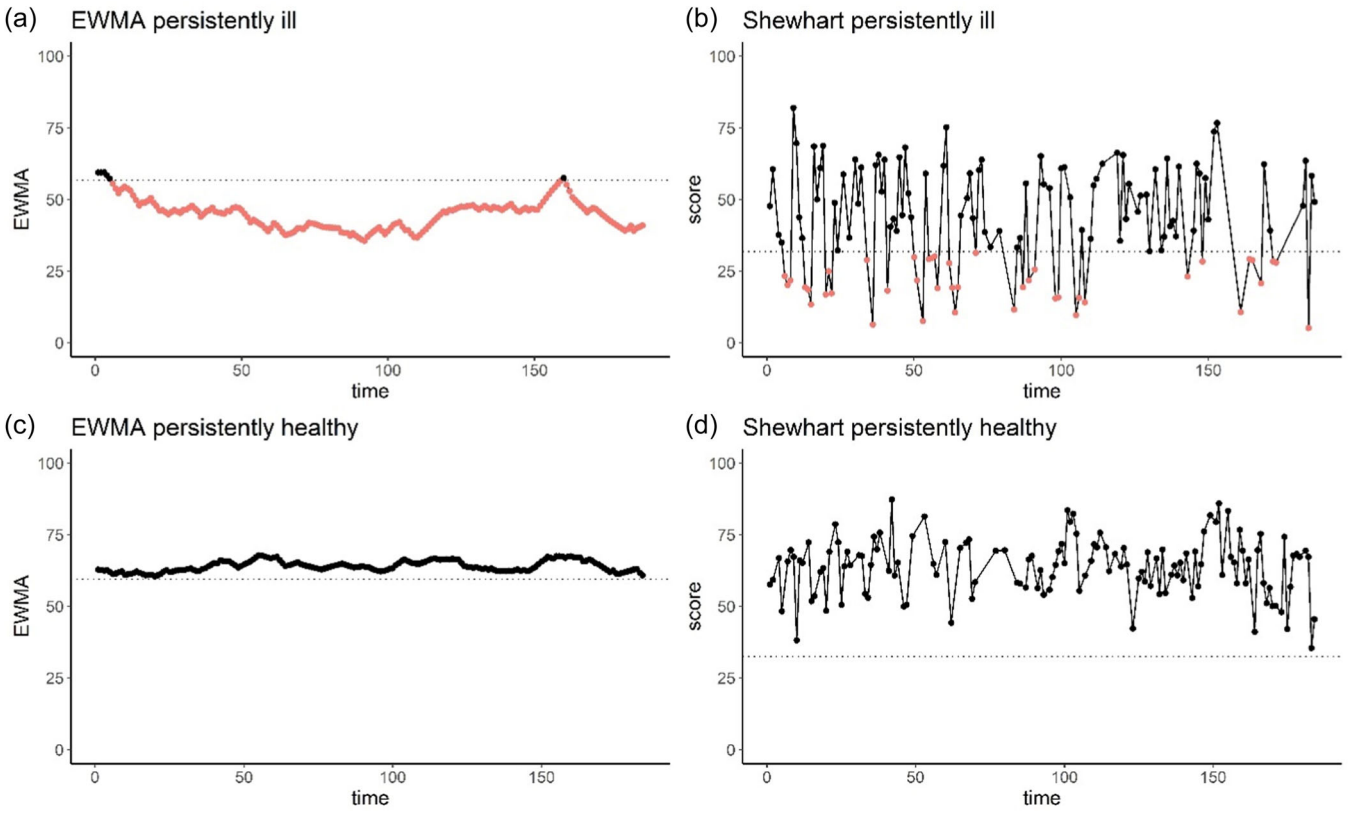


FIGURE 4 Example of SPC-derived warning signals in feeling appreciated in a persistently ill individual (a, b) and in a persistently healthy individual (c, d). Horizontal lines reflect the lower control limits, which were based on the HND study (setting 5 in Table 3). While the monitored scores of the persistently healthy individual continuously remains within the normative range (above the lower control limit), the monitored scores of the persistently ill individual are often below the lower control limit, which can be considered a warning signal (colored in red).

Therefore, warning signals in feeling that life is worth it have low specificity. There were also emotions that showed few warning signals (e.g., feeling nervous; Shewhart method: 1.8%, EWMA method: 10.1%) or no warning signals at all (being physically active; Shewhart method: 0%) in persistently healthy individuals. The latter occurred because, given the relatively low within-person mean in this item in HND participants, the lower control limit was below zero, and hence, outside the response scale. Put differently, for some emotions, it was impossible for TT participants to score outside the normative range.

3.4 | Performance of SPC

Figure 6 displays the optimal (highest across the evaluated decision rules) MCC per emotion, SPC method, and personalization setting (see Supporting Information S1: Supplement Table S3 for accompanying statistics, including the corresponding sensitivity, specificity, and predictive values). It thus shows how often monitored scores should provide a warning signal (i.e., the percentage of warning signs) for optimally distinguishing between persistently healthy and persistently ill individuals. For instance, detecting persistently ill individuals based on how tired they feel is relatively difficult (with MCC values around 0.21 for most methodological variations), but is most accurate if persistent illness requires warning signs in tiredness at least 5%–10% of the time (depending on the method and personalization setting). As illustrated, almost all MCC values were significantly larger than 0, implying that the prediction of clinical profiles (healthy vs. ill) through SPC-derived warning signs was better than chance-level prediction.[5] In what follows, we will briefly discuss the effect of distinct emotions, personalization settings, and SPC methods on the performance of SPC.

Regarding the effect of the monitored emotions on the performance of SPC, there were few significant differences in MCC values retrieved from warning signals in different emotions (Supporting Information S1: Supplement 9). First, warning signs in feeling tired (Shewhart method; MCC = 0.20, Sens. = 8%, Spec. = 99%) performed worse than warning signs in most other emotions (MCC = 0.31–0.45, Sens. = 27%–53%, Spec = 86%–99%, $p < 0.05$, Supplement 9).

[5]Following an anonymous reviewer's suggestion, we conducted logistic regression analyses were we predicted individuals' clinical status (healthy vs. ill) based on (1) mean levels and (2) the amount of warning signs detected through SPC for different emotions, personalization settings, and SPC methods separately. In Supporting Information S1: Supplement 8, we indicated those combinations of emotions, personalization settings, and SPC methods for which warning signs did not significantly predict clinical status after accounting for mean levels.

26423588, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mhs2.72 by Ku Leuven, Wiley Online Library on [10/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

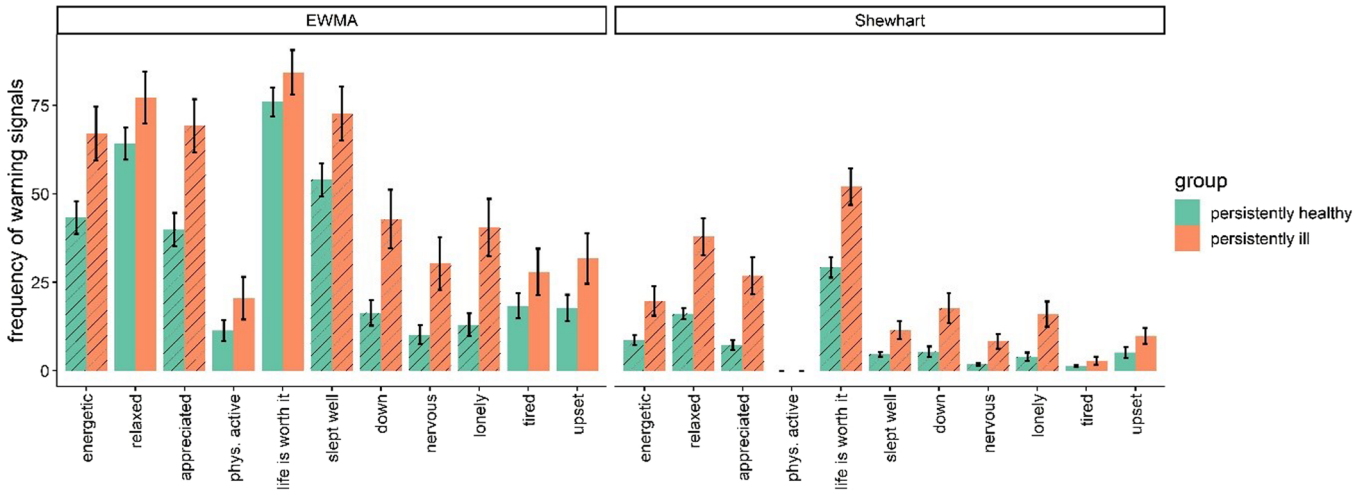


FIGURE 5 Average frequency (in %) of warning signals derived from SPC, with control limits based on the HND study (setting 5 in Table 3; see Supporting Information S1: Supplement 7 for other personalization settings). Error bars reflect the standard error of the mean. Shaded bars denote statistically significant differences between groups, which was evaluated through Welch two-sample t tests ($\alpha = 0.05$).

Secondly, for specific personalization settings,[6] warning signs in feeling upset (EWMA method; MCC = 0.15, Sens. = 35%, Spec. = 80%) yielded MCC values that were significantly lower than MCC values retrieved from other emotions (MCC = 0.18–0.40, Sens. = 31%–65%, Spec. = 78%–93%; $p < 0.05$, Supporting Information S1: Supplement 9). Physically active (Shewhart method) also proved a poor warning signal (MCC = 0.13, Sens. = 22%, Spec. = 90%), which was due to the fact that warning signals were impossible for the majority of settings. Finally, a recurring pattern across SPC settings and methods was that positively valenced emotions required far more warning signs to differentiate between persistently healthy and ill young adults compared to negatively valenced emotions (positive emotions: ~50.9%, negative emotions: ~16.5%).

Regarding the effect of personalization settings on the performance of SPC, all settings performed roughly similar. As an exception, MCC values of life is worth it (EWMA method) were significantly lower with personalized control limits (age-specific control limits, person-specific control limits with QIDS = 3) compared to generic control limits ($p = 0.044$ and $p = 0.049$; Figure 6). Generally speaking, when control limits were based on the HND sample–either with or without taking into account age, sex, and depressive symptoms (cf. settings 1-6 in Table 3)–MCC values ranged between 0.31 and 0.34 (averaged across 11 items and 2 SPC methods; Sens. 37%–51%, Spec. 80%–91%). Similar performance was obtained when control limits were based on TT participants who did not meet diagnostic criteria at the start of the EMA study (cf. setting 7 in Table 3, MCC = 0.31, Sens. = 42%, Spec. = 85%). The lack of pronounced differences across personalization settings likely occurred because, contrary to our expectations, increasing levels of personalization did not yield increasingly accurate estimates of within-person means ($\hat{\mu}_{ie}$ in Table 2). Put differently, regardless of the personalization setting, within-person means for almost all emotions were consistently overestimated (settings 1–6) or over- and underestimated (setting 7; see Supporting Information S1: Supplement Figure S7 for details).

Finally, regarding the effect of different methods (EWMA vs. Shewhart) on the performance of SPC, the Shewhart and EWMA method did not yield significantly different performances (averaged across items; EWMA: MCC = 0.31, Sens. = 50%, Spec. = 81%; Shewhart: MCC = 0.34, Sens. = 33%, Spec. = 93%). As an exception, the Shewhart method outperformed the EWMA method when monitoring feeling upset using age-specific control limits (MCC EWMA = 0.15, Shewhart = 0.41, $p = 0.0373$; Figure 6). The Shewhart method also required more fewer signals (~17.1%) than the EWMA method (~52.4%) to optimally distinguish between clinical profiles–which was most evident in emotions of positive valence.

To illustrate how current findings could be used prospectively, we inspected the timing of warning signals in feeling down and lonely, which were selected because they yielded relatively high MCC values on average (i.e., >0.40) without necessitating a high percentage of warning signs (i.e., <10% for most methodological variations). Figure 7 illustrates the distribution of the timing of warning signals in persistently healthy and ill young adults for different personalization settings and SPC methods (which each had their own optimal decision rule). Most warning signs occurred relatively early on, which has been reported before in SPC literature and is an inevitable feature (Gan, 1993; Schat et al., 2024). A second, more conceptual explanation for the early occurrence of warnings is that individuals' mental health did not substantially change throughout the EMA study (i.e., they had persistent problems or were persistently healthy), meaning that any time-varying feature (such as a warning sign) describing these individuals should occur early on (or not at all). As hypothesized, warning signals generally occurred earlier for young adults with persistent mental health problems compared to healthy young adults, although the magnitude of this difference differed across personalization settings and SPC methods.

[6]Namely, HND age-specific and HND sex- and age-specific. Reported performance metrics (MCC, sensitivity, specificity) refer to these specific settings.

26423588, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mhs2.72 by Ku Leuven, Wiley Online Library on [10/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

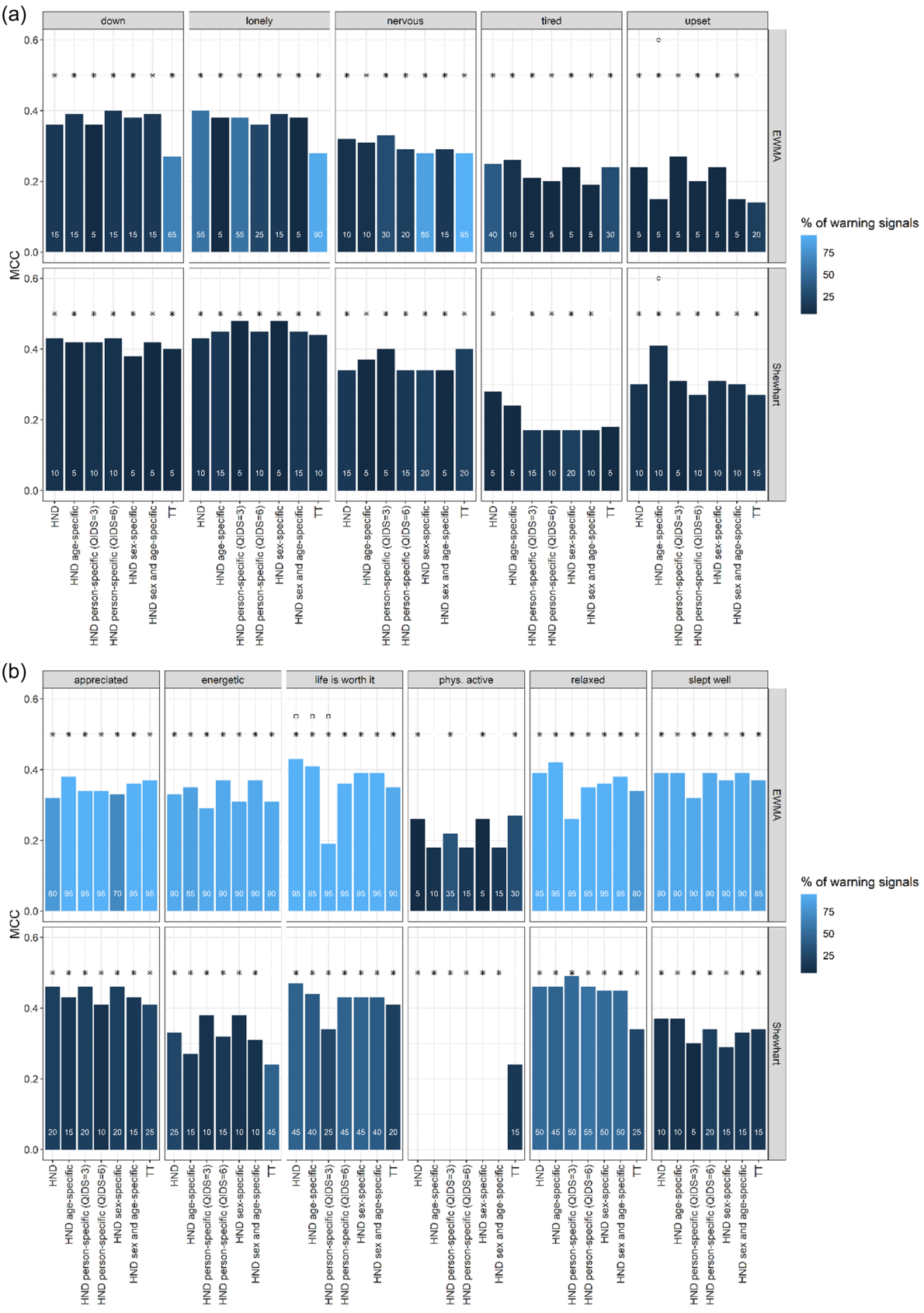


FIGURE 6 (See caption on next page).

26423588, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mhs2.72 by Ku Leuven, Wiley Online Library on [10/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

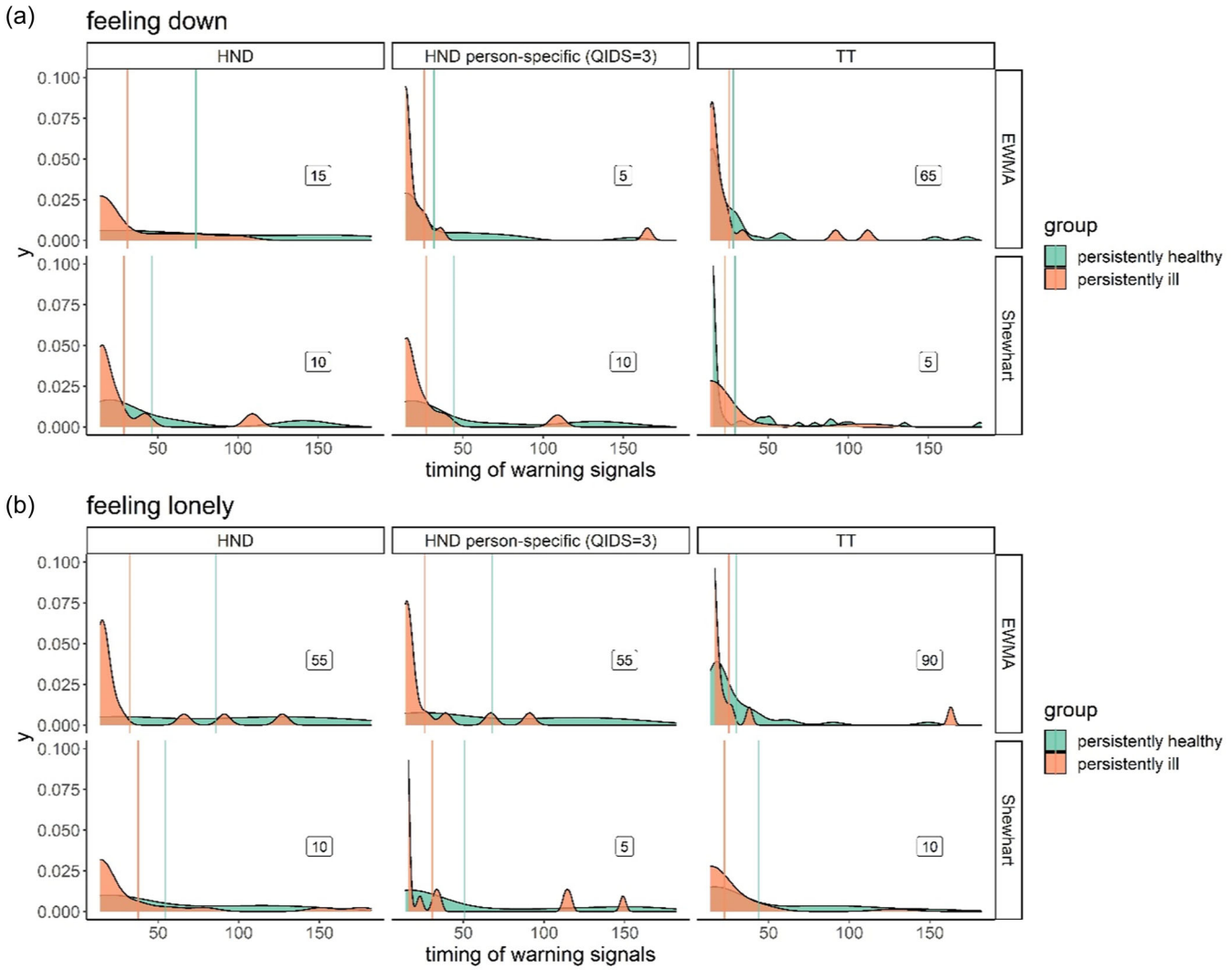


FIGURE 7 Distribution of the timing of warning signals in feeling down (a) and feeling lonely (b) across three different personalization settings (ranging from fully generic, or "HND", to more personalized, or "TT"), for different SPC methods (EWMA, Shewhart). Vertical lines reflect the average timing of warning signals in each group; numbers denote the decision rule that yielded the highest MCC (i.e., the amount of warning signals necessary to optimally distinguish between healthy and ill young adults).

4 | DISCUSSION

Statistical process control (SPC) enables the real-time detection of upcoming mental health problems by establishing whether the emotions of a specific individual, on a specific day, are within the normal range or not (Schat, Tuerlinckx, De Ketelaere, et al., 2023; Smit et al., 2019, 2022; Smit & Snippe, 2022; Snippe et al., 2023). Evidently, a key question is how we define the normal range–or, in SPC terminology, the control limits. The current study found that relatively generic control limits, which were based on general population EMA data, more often yield warning signals in young adults with persistent mental health problems than in healthy young adults. We next investigated whether the distinction of clinical profiles would improve when increasing the extent to which control

FIGURE 6 Matthews correlation coefficient (MCC) across different personalization settings (ranging from fully generic, or "HND", to more personalized, or "TT"), for different SPC methods (EWMA, Shewhart), for emotions of negative valence (a) and positive valence (b). MCC values for being physically active (Shewhart method) could not be computed because warning signals were impossible (see Frequency of warning signals). The number in each bar reflects the optimal decision rule (indicating the percentage of warning signs required to optimally distinguish between healthy and ill individuals). * Signs denote that MCC values were significantly larger than 0 (α = .05). ° Signs denote that MCC values significantly differed across SPC methods (EWMA vs. Shewhart). □ Signs denote that MCC values significantly differed across personalization settings. A comparison of differences across different items is provided in Supporting Information S1: Supplement 9. Comparison of items.

26423588, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mhs2.72 by Ku Leuven, Wiley Online Library on [10/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

limits were personalized, and found that this was not the case. Similarly, different SPC methods (EWMA vs. Shewhart) hardly yielded different MCC values. By contrast, which emotions were being monitored consistently impacted the performance of SPC. Specifically, emotions on which healthy and ill youth reported comparable within-person means (upset, tired) also yielded relatively poor discriminative ability when applying SPC, a point that we will further discuss below. Taken together, this work illustrates that warning signals based on population-derived control limits can distinguish between young adults with versus without persistent mental health problems, implying that vulnerability for mental health-problems may manifest not only as a deviation from one's own, personal norm (as shown in earlier SPC studies), but also as a deviation from a more generic norm. Eventually, adopting such generic norms could considerably broaden the potential utility of SPC.

In its current application, SPC essentially leverages on between-person differences in momentary emotions, and it is therefore not surprising that emotions that "work well" in terms of SPC (i.e., have a higher MCC value) are those with relatively high mean-level differences between healthy versus ill youth. Vice versa, the two emotions that yielded a relatively poor SPC performance (upset, tired) also did not differ significantly in their mean levels when comparing healthy and ill youth. At the same time, mean-level differences could not fully explain the performance of SPC across different emotions. Moreover, the simple existence of mean-level differences does not by itself provide insights that may be used for monitoring purposes. It follows that the merits of SPC not necessarily lie in informing about the emotional processes that may underlie psychopathology, but rather, in providing a formal, statistically solid method for monitoring such processes in real-time. Before moving towards such real-time implementation, however, it is crucial to further improve the feasibility and performance of SPC. With regard to the latter, we found that the best-performing methodological variations (emotion, personalization setting, SPC method) had MCC values around 0.50, corresponding to a predictive accuracy of 80%. Although our approach – examining the percentage of warning signs across the entire monitoring period – differs from earlier SPC studies – which applied person-specific control limits and monitored the occurrence of single warning signs – this predictive performance is in line with what has been reported earlier (Snippe et al., 2023). Thus, a tentative conclusion might be that deriving control limits from general population EMA data – as opposed to person-specific phase I EMA data – does not substantially lower the performance of SPC.

When replacing person-specific phase I EMA data with EMA data from the general population, a crucial question that arises is to what extent the reduced personalization of control limits affects the performance of SPC. Based on the current findings, it seems that this effect is surprisingly small, and almost never statistically significant.[7] This could mean that our approach to personalization was less effective than anticipated. Indeed, an evaluation of personalization settings showed that our estimation of within-person means – which are key to setting control limits – did not become more accurate as we took into account additional variables when estimating these means (i.e., age, sex, and depressive symptoms). The reason for this might be that the effects of age, sex, and depressive symptoms on within-person means are small, meaning that predicted within-person means might have remained similar across regression models that included different combinations of age, sex, and depressive symptoms as predictors (*i.e.*, settings 2-6). This is supported by the relatively small values of R^2 in models that predicted within-person means based on age ($R^2 = 0.01$) or sex ($R^2 < 0.01$) alone. Depressive symptoms, as assessed by the QIDS, had a larger effect on within person means ($R^2 = 0.21$). However, the QIDS was not assessed in the TT study, and we could therefore not inform the estimation of within-person means by TT participants' QIDS score. It follows that the extent to which we could personalize the control limits was constrained by the limited overlap between the HND and TT study designs, and it requires further research to explore whether taking into account other potentially influential characteristics, such as personality or symptoms of mental disorders other than depression, may lead to more accurate predictions of within-person means (and hence, more personalized control limits).

Besides different emotions and personalization settings, a final factor that we expected to influence the performance of SPC concerned the method that was used (Shewhart vs. EWMA). We found that the most accurate distinctions between persistently ill versus healthy individuals were obtained using the Shewhart method. However, significant differences between the Shewhart and EWMA method were virtually absent[8] and the average performance of both methods across all methodological variations was similar (EWMA: MCC = 0.31, Shewhart: MCC = 0.34). This is somewhat in contrast with earlier work, which reported that the EWMA method yields more reliable warning signals than the Shewhart method (Schat, Tuerlinckx, De Ketelaere, et al., 2023). An explanation for this discrepancy could be that, in contrast to earlier SPC applications, we did not predict mood transitions *during* the EMA study, but rather predicted persistent mental health problems. It is possible that the EWMA method picked up on small, irrelevant mood shifts in the persistently healthy group (Montgomery, 2009; Roberts, 1959)–thus producing false alarms (or type I errors) that would lower its overall predictive performance. In a similar vein, EWMA-based warning signals may have been absent in the persistently ill group in case of small (clinically irrelevant) mood improvements (type II error). If such EWMA-based type I/II errors occur, it is more likely that the next score will also yield a type I/II error. This is because, in contrast to the Shewhart method, the EWMA method yields autocorrelated warning signals (Figure 2). Thus, the high sensitivity to small changes combined with the autocorrelated nature of warning signals might have put the EWMA method at a disadvantage in the current study.

Taken together, we found few consistent, statistically significant differences in SPC performance across personalization settings and

[7]Exception: feeling life is worth it, EWMA method, HND person specific (QIDS = 3) versus HND age-specific and HND (Figure 6).

[8]Exception: feeling upset, age-specific control limits (Figure 6).

26423588, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mhs2.72 by Ku Leuven, Wiley Online Library on [10/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

SPC methods. This could mean that such differences are simply absent or that the current design obscured them, as explained above. Alternatively, it is possible that the methodological factors we examined only impact the performance of SPC in interaction, such that for instance personalization of control limits is beneficial when monitoring one emotion, but not when monitoring another emotion. This would imply that vulnerability to mental health problems may be characterized by certain emotions exceeding one's *personal* norm, and by other emotions that exceed a more *generic* norm. Indirect support for this reasoning comes from studies that reported the intra-class correlation coefficient (ICC) of distinct emotions, and found that some emotions may be particularly informative on a within-person level, while others may be more informative on a between-person level (Erbas et al., 2019; Goetz et al., 2015). It follows that the extent to which control limits should be personalized may differ across emotions. In a similar vein, it is possible that monitoring raw scores using the Shewhart method is particularly valuable for specific emotions or psychopathological symptoms (e.g. those for which an absolute value, such as a value above zero, may have a salient clinical meaning, such as momentary suicidal ideation). Inspecting such interaction effects between emotions and personalization settings (or SPC methods) on the performance of SPC was beyond the scope of this paper, and thus calls for further research. Likewise, it requires follow-up research to examine whether optimal methodological choices (e.g., emotions, personalization settings, and SPC methods) vary depending on the outcome that is being predicted (e.g., onset or recurrence of depression, binge eating episode, etc.).

4.1 | Limitations and future directions

By estimating control limits based on general population EMA data as opposed to person-specific phase I EMA data–and inspecting the ramifications thereof for the performance of SPC–this study aimed to make SPC more feasible. Our design also came with several limitations, all of which point to future research directions. First, the EMA study that was used to establish control limits (i.e., the HND study) included individuals from the general population, but these individuals were not representative of the general population. Specifically, as a consequence of self-selection, highly educated women were overrepresented in the HND EMA study (Van Der Krieke et al., 2016). While we corrected the sex imbalance by taking into account sex in the computation of control limits, we could not correct for educational attainment. This might have biased the control limits (although the direction of such bias remains unclear), and follow-up studies should thus take care to base control limits on EMA data in a sample representative of the general population. With a larger general population sample, and a richer set of person characteristics to be matched to the monitored individuals (here, TT participants), we may derive more accurate approximations of people's control limits. Such person characteristics need not be limited to socio-demographic information, but could also cover personality and mental health issues beyond depressive symptoms.

As a second limitation, evaluating relative frequencies of warning signals (as in the current study) is more complicated than monitoring the occurrence of single warning signals (as in conventional SPC applications), and requires the choice of a certain timeframe within which frequencies are evaluated. Earlier studies have often chosen a timeframe of 14 days, and the current study followed this convention. However, if motivated at all, this timeframe has often followed from pragmatic reasons (e.g., data availability, Curtiss et al., 2021) rather than theoretical or conceptual ones (Bos et al., 2022). Hence, more research is needed with respect to the time window within which warning signals are expected to occur. This is not only relevant to the current SPC study, but to the broader field of research into warning signals for mental health problems, in which the timescale at which emotions unfold and warning signals (should) arise is still largely unknown.

Third, we could not investigate the performance of SPC in TT participants who either developed mental health problems during the diary period or experienced remission of such problems. The reason for this was that both the timing of onset and remission of mental health problems could not be inferred from the retrospective diagnostic interviews, and therefore, we were unable to evaluate the performance of SPC for individuals with worsening or remitting mental health problems. An important next step for further research is therefore to extend current findings to populations who, during an ecological momentary assessment study, experience changes in specific mental health problems that can be pinpointed in time (cf. Schreuder et al., 2024; Smit & Snippe, 2022; Snippe et al., 2023).

Fourth, we could not compare our approach–that is, deriving phase I data from the general population–to the conventional SPC approach–that is, deriving phase I data from a person's own EMA data (benchmark). This was because it was impossible to delineate phases I and II in the TT data. Follow-up research is thus needed to evaluate whether the advantages of circumventing phase I by deriving control limits from the general population (i.e., lower burden for individuals who are delivering EMA data) weigh up to its costs (e.g. potential loss of sensitivity/specificity).

4.2 | Conclusions

There is an increasing interest in methods to detect person-specific warning signals for mental ill-health. An important merit of SPC is that it allows for detecting such warning signs in real-time, which is crucial for intervening in time (i.e., before mental health problems have worsened). While SPC has proven promising (Schreuder et al., 2024; Snippe et al., 2023), the necessity of person-specific EMA data collected during a stable period of at least 50 days (i.e., phase I) considerably compromises the feasibility of the method (Schat, Tuerlinckx, De Ketelaere, et al., 2023; Schat et al., 2024). We therefore investigated whether phase I can be replaced by population-based EMA data. Specifically, we derived control limits from general population EMA data and subsequently used those control limits to detect warning signs in EMA data from at-risk young

26423588, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mhs2.72 by Ku Leuven, Wiley Online Library on [10/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

adults. We found that warning signals inferred from SPC could differentiate persistently healthy and ill individuals. Further, it seemed that the performance of SPC was mostly affected by the emotion that was monitored, and less by the method used (EWMA vs. Shewhart) or the extent to which control limits took into account person characteristics (personalization settings). Thus, it might be that prospective monitoring techniques, such as SPC, could be applied without necessitating phase I data. As this would substantially improve the feasibility of SPC, follow-up research into the utility of more or less person-specific control limits for forecasting mental health problems is encouraged.

ACKNOWLEDGMENTS

We thank all participants in the HowNutsAreTheDutch study and the TRAILS TRANS-ID study for their time and effort. This research was funded by the Research Council KU Leuven (project C14/19 to Eva Ceulemans and Peter Kuppens) and by the Fund of Scientific Research Flanders (FWO, grant 12AVE24N).

CONFLICT OF INTEREST STATEMENT

There are no conflicts of interest.

DATA AVAILABILITY STATEMENT

The R code for the statistical analyses pertaining to this study will be made available on the Open Science Framework (www.osf.io/hvx2r). Data cannot be made openly available due to privacy regulations, but are available upon request (for HowNutsAreTheDutch, contact Peter de Jonge; for TRAILS TRANS-ID, contact Catharina A. Hartman).

ETHICS STATEMENT

All participants provided their written informed consent. All procedures outlined in this study were approved by ethical committees (HowNutsAreTheDutch: ref. M13.147422, TRAILS TRANS-ID: ref. 207/103) and are in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki.

ORCID

Marieke J. Schreuder http://orcid.org/0000-0003-4901-2697
Peter Kuppens http://orcid.org/0000-0002-2363-2356
Evelien Schat http://orcid.org/0000-0003-1169-3984
Peter de Jonge http://orcid.org/0000-0002-0866-6929
Catharina A. Hartman http://orcid.org/0000-0002-8094-8859
Eva Ceulemans http://orcid.org/0000-0002-7611-4683

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1002/mhs2.72.

REFERENCES

American Psychiatric Association. (2013). Diagnostic and Statistical Manual of Mental Disorders (5th ed.).

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), 12. https://doi.org/10.1186/s40345-022-00258-4

Caspi, A., Houts, R. M., Ambler, A., Danese, A., Elliott, M. L., Hariri, A., Harrington, H., Hogan, S., Poulton, R., Ramrakha, S., Rasmussen, L. J. H., Reuben, A., Richmond-Rakerd, L., Sugden, K., Wertz, J., Williams, B. S., & Moffitt, T. E. (2020). Longitudinal assessment of mental health disorders and comorbidities across 4 decades among participants in the Dunedin birth cohort study. *JAMA Network Open*, *3*(4), e203221. https://doi.org/10.1001/jamanetworkopen.2020.3221

Copeland, W. E., Shanahan, L., Costello, E. J., & Angold, A. (2009). Childhood and adolescent psychiatric disorders as predictors of young adult disorders. *Archives of General Psychiatry*, *66*(7), 764–772. https://doi.org/10.1001/archgenpsychiatry.2009.85

Curtiss, J. E., Mischoulon, D., Fisher, L. B., Cusin, C., Fedor, S., Picard, R. W., & Pedrelli, P. (2021). Rising early warning signals in affect associated with future changes in depression: a dynamical systems approach. *Psychological Medicine*, *53*, 3124–3132. https://doi.org/10.1017/S0033291721005183

Erbas, Y., Ceulemans, E., Blanke, E. S., Sels, L., Fischer, A., & Kuppens, P. (2019). Emotion differentiation dissected: between-category, within-category, and integral emotion differentiation, and their relation to well-being. *Cognition and Emotion*, *33*(2), 258–271. https://doi.org/10.1080/02699931.2018.1465894

Fisher, A. J., & Soyster, P. D. (2021). Generating accurate personalized predictions of future behavior: A smoking exemplar. *Preprint*.

Fusar-Poli, P., Rutigliano, G., Stahl, D., Davies, C., Bonoldi, I., Reilly, T., & McGuire, P. (2017). Development and validation of a clinically based risk calculator for the transdiagnostic prediction of psychosis. *JAMA Psychiatry*, *74*(5), 493–500. https://doi.org/10.1001/jamapsychiatry.2017.0284

Gan, F. F. (1993). An optimal design of EWMA control charts based on median run length. *Journal of Statistical Computation and Simulation*, *45*, 169–184.

Goetz, T., Becker, E. S., Bieg, M., Keller, M. M., Frenzel, A. C., & Hall, N. C. (2015). The glass half empty: How emotional exhaustion affects the state-trait discrepancy in self-reports of teaching emotions. *PLoS One*, *10*(9), e0137441. https://doi.org/10.1371/journal.pone.0137441

Helmich, M. A., Smit, A. C., Bringmann, L. F., Schreuder, M. J., Oldehinkel, A. J., Wichers, M., & Snippe, E. (2022). Detecting impending symptom transitions using early warning signals in individuals receiving treatment for depression. *Clinical Psychological Science*, *11*(6), 994–1010.

Huisman, M., Oldehinkel, A. J., de Winter, A., Minderaa, R. B., de Bildt, A., Huizink, A. C., Verhulst, F. C., & Ormel, J. (2008). Cohort profile: the Dutch "TRacking adolescents" individual lives' survey'; TRAILS. *International Journal of Epidemiology*, *37*(6), 1227–1235. https://doi.org/10.1093/ije/dym273

Jones, L. A. (2002). The statistical design of EWMA control charts with estimated parameters. *Journal of Quality Technology*, *34*(3), 277–288. https://doi.org/10.1080/00224065.2002.11980158

Kessler, R. C., Angermeyer, M., Anthony, J. C., de Graaf, R., Demyttenaere, K., Gasquet, I., de Girolamo, G., Gluzman, S., Gureje, O., Haro, J. M., Kawakami, N., Karam, A., Levinson, D., Medina Mora, M. E., Oakley Browne, M. A., Posada-Villa, J., Stein, D. J., Adley Tsang, C. H., Aguilar-Gaxiola, S., … Ustün, T. B. (2007). Lifetime prevalence and age-of-onset distributions of mental disorders in the world health organization's world mental health survey initiative. *World Psychiatry: Official Journal of the World Psychiatric Association (WPA)*, *6*, 168–176.

Koller, M., & Stahel, W. (2011). Sharpening Wald-type inference in robust regression for small samples. *Computational statistics & data analysis*, *55*, 2504–2515.

26423588, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mhs2.72 by Ku Leuven, Wiley Online Library on [10/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Loth, A. K., Drabick, D. A. G., Leibenluft, E., & Hulvershorn, L. A. (2014). Do childhood externalizing disorders predict adult depression? A meta-analysis. *Journal of Abnormal Child Psychology*, *42*(7), 1103–1113. https://doi.org/10.1007/s10802-014-9867-8

Ludwig, V. M., Reinhard, I., Mühlbauer, E., Hill, H., Severus, W. E., Bauer, M., Ritter, P., & Ebner-Priemer, U. W. (2024). Limited evidence of autocorrelation signaling upcoming affective episodes: a 12-month e-diary study in patients with bipolar disorder. *Psychological Medicine*, *54*, 1844–1852.

Maechler, M., Rousseeuw, P., Croux, C., Todorov, V., Ruckstuhl, A., Salibian-Barrera, M., Verbeke, T., Koller, M., Conceicao, E., & Anna di Palma, M. (2022). *robustbase: Basic Robust Statistics* (R package version 0.95-0). http://robustbase.r-forge.r-project.org/

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

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

Nahum-Shani, I., Smith, S. N., Spring, B. J., Collins, L. M., Witkiewitz, K., Tewari, A., & Murphy, S. A. (2018). Just-in-time adaptive interventions (JITAIs) in mobile health: Key components and design principles for ongoing health behavior support. *Annals of Behavioral Medicine*, *52*, 446–462. https://doi.org/10.1007/s12160-016-9830-8

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. *British Journal of Psychiatry*, *196*(1), 64–68. https://doi.org/10.1192/bjp.bp.109.066563

Oldehinkel, A. J., Rosmalen, J. G., Buitelaar, J. K., Hoek, H. W., Ormel, J., Raven, D., Reijneveld, S. A., Veenstra, R., Verhulst, F. C., Vollebergh, W. A., & Hartman, C. A. (2015). Cohort profile update: the tracking adolescents' individual lives survey (TRAILS). *International Journal of Epidemiology*, *44*(1), 76–76n. https://doi.org/10.1093/ije/dyu225

Ormel, J., Raven, D., van Oort, F., Hartman, C. A., Reijneveld, S. A., Veenstra, R., Vollebergh, W. A. M., Buitelaar, J., Verhulst, F. C., & Oldehinkel, A. J. (2015). Mental health in Dutch adolescents: A TRAILS report on prevalence, severity, age of onset, continuity and co-morbidity of DSM disorders. *Psychological Medicine*, *45*(2), 345–360. https://doi.org/10.1017/S0033291714001469

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

Reef, J., Diamantopoulou, S., Van Meurs, I., Verhulst, F. C., & Van Der Ende, J. (2011). Developmental trajectories of child to adolescent externalizing behavior and adult DSM-IV disorder: Results of a 24-year longitudinal study. *Social Psychiatry and Psychiatric Epidemiology*, *46*(12), 1233–1241. https://doi.org/10.1007/s00127-010-0297-9

Reilly, T. J., MacGillivray, S. A., Reid, I. C., & Cameron, I. M. (2015). Psychometric properties of the 16-item Quick Inventory of Depressive Symptomatology: A systematic review and meta-analysis, In *Journal of Psychiatric Research* (60, pp. 132–140). Elsevier Ltd. https://doi.org/10.1016/j.jpsychires.2014.09.008

Roberts, S. W. (1959). Control chart tests based on geometric moving averages. *Technometrics*, *1*, 239–250.

Rush, A. J., Trivedi, M. H., Ibrahim, H. M., Carmody, T. J., Arnow, B., Klein, D. N., Markowitz, J. C., Ninan, P. T., Kornstein, S., Manber, R., Thase, M. E., Kocsis, J. H., & Keller, M. B. (2003). The 16-Item quick inventory of depressive symptomatology (QIDS), clinician rating (QIDS-C), and Self-Report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. *Biological Psychiatry*, *54*, 573–583. https://doi.org/10.1016/S0006-3223(03)01866-8

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*, *56*, 1459–1475. https://doi.org/10.3758/s13428-023-02103-7

Schat, E., Tuerlinckx, F., Smit, A. C., De Ketelaere, B., & Ceulemans, E. (2023). Detecting mean changes in experience sampling data in real-time: A comparison of univariate and multivariate statistical process control methods. *Psychological Methods*, *28*(6), 1335–1357. https://doi.org/10.1037/met0000447

Schat, E., Tuerlinckx, F., Schreuder, M. J., De Ketelaere, B., & Ceulemans, E. (2024). Forecasting the onset of depression with limited baseline data only: A comparison of a person-specific and a multilevel modeling based exponentially weighted moving average approach. *Psychological Assessment*, *36*(6–7), 379–394. https://doi.org/10.1037/pas0001314

Schreuder, M. J., Groen, R. N., Wigman, J. T. W., Hartman, C. A., & Wichers, M. (2020). Measuring psychopathology as it unfolds in daily life: addressing key assumptions of intensive longitudinal methods in the TRAILS TRANS-ID study. *BMC Psychiatry*, *20*(1), 351. https://doi.org/10.1186/s12888-020-02674-1

Schreuder, M. J., Groen, R. N., Wigman, J. T. W., Wichers, M., & Hartman, C. A. (2023). Participation and compliance in a 6-Month daily diary study among individuals at risk for mental health problems. *Psychological Assessment*, *35*(2), 115–126. https://doi.org/10.1037/pas0001197

Schreuder, M. J., Schat, E., Smit, A. C., Snippe, E., & Ceulemans, E. (2024). Monitoring emotional intensity and variability in real-time to forecast depression recurrence in remitted adults. *Journal of Consulting and Clinical Psychology*. Advance online publication. https://doi.org/10.1037/ccp0000871

Schreuder, M. J., Hartman, C. A., Groen, R. N., Smit, A. C., Wichers, M., & Wigman, J. T. W. (2022). Anticipating transitions in mental health in at-risk youths: A 6-month daily diary study into early-warning signals. *Clinical Psychological Science*, *11*, 1026–1043. https://doi.org/10.1177/21677026221103138

Shewhart, W. A. (1931). *Economic control of quality of manufactured product*. *Braunworth & co*.

Smit, A. C., & Snippe, E. (2022). Real-time monitoring of increases in restlessness to assess idiographic risk of recurrence of depressive symptoms. *Psychological Medicine*, *53*, 5060–5069. https://doi.org/10.1017/S0033291722002069

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

Smit, A. C., Schat, E., & Ceulemans, E. (2022). 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*, 1354–1368. https://doi.org/10.1177/10731911221086985

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.

Thompson, W. K., Gershon, A., O'Hara, R., Bernert, R. A., & Depp, C. A. (2014). The prediction of study-emergent suicidal ideation in bipolar disorder: a pilot study using ecological momentary assessment data. *Bipolar Disorders*, *16*(7), 669–677. https://doi.org/10.1111/bdi.12218

Van Der Krieke, L., Blaauw, F. J., Emerencia, A. C., Schenk, H. M., Slaets, J. P. J., Bos, E. H., De Jonge, P., & Jeronimus, B. F. (2017). Temporal dynamics of health and well-being: A crowdsourcing approach to momentary assessments and automated generation of personalized feedback. *Psychosomatic Medicine*, *79*(2), 213–223. https://doi.org/10.1097/PSY.0000000000000378

26423588, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mhs2.72 by Ku Leuven, Wiley Online Library on [10/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

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

Yuen, H. P., Mackinnon, A., Hartmann, J., Amminger, G. P., Markulev, C., Lavoie, S., Schäfer, M. R., Polari, A., Mossaheb, N., Schlögelhofer, M., Smesny, S., Hickie, I. B., Berger, G., Chen, E. Y. H., de Haan, L., Nieman, D. H., Nordentoft, M., Riecher-Rössler, A., Verma, S., … Nelson, B. (2018). Dynamic prediction of transition to psychosis using joint modelling. *Schizophrenia Research*, *202*, 333–340. https://doi.org/10.1016/j.schres.2018.07.002

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Schreuder, M. J., Kuppens, P., Schat, E., de Jonge, P., Hartman, C. A., & Ceulemans, E. (2024). Warning signals for mental health problems in at-risk young adults may be informed by momentary emotions reported by the general population: A novel application of the principles of statistical process control. *Mental Health Science*, *2*, e72. https://doi.org/10.1002/mhs2.72

26423588, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/mhs2.72 by Ku Leuven, Wiley Online Library on [10/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License