Behavior Research Methods (2026) 58:12
https://doi.org/10.3758/s13428-025-02870-5

ORIGINAL MANUSCRIPT

Triggering just-in-time adaptive interventions based on real-time detection of daily-life stress: Methodological development and longitudinal multicenter evaluation

S. A. Bögemann[1] · F. Krause[1] · A. van Kraaij[2] · M. A. Marciniak[3,4,5] · J. M. van Leeuwen[1] · J. Weermeijer[6] · J. Mituniewicz[7] · L. M. C. Puhlmann[8,9] · M. Zerban[10] · Z. C. Reppmann[11] · D. Kobylińska[7] · K. S. L. Yuen[8,10] · B. Kleim[3,4] · H. Walter[11,12] · I. Myin-Germeys[6] · R. Kalisch[8,10] · I. M. Veer[13] · K. Roelofs[14,15] · E. J. Hermans[1]

Received: 4 November 2024 / Accepted: 6 October 2025

Abstract

Stress-related disorders present a significant global burden, highlighting the need for effective, preventive measures. Mobile just-in-time adaptive interventions (JITAI) can be applied in real time and context-specifically, precisely when individuals need them most. Yet, they are rarely applied in stress research. This study introduces a novel approach by performing real-time analysis of both psychological and physiological data to trigger interventions during moments of high stress. We evaluated the feasibility of this JITAI algorithm, which integrates ecological momentary assessments (EMA) and ecological physiological assessments (EPA) to generate a stress score that triggers interventions in real time by relating the score to a personalized stress threshold. The feasibility of the technical implementation, participant adherence, and user experience were assessed within a multicenter study with 215 participants conducted across five research sites. The JITAI algorithm successfully processed EMA and EPA data to trigger real-time interventions. A total of 68% (standard deviation [SD] = 29%) of EMA beeps contained extracted EPA features, demonstrating technical feasibility. The algorithm triggered 1.61 (SD = 1.26) interventions per day, with 43% (SD = 27%) of EMA beeps per week leading to triggered interventions. Compliance rates of 43% (SD = 22%) for EMA and 43% (SD = 30%) for the JITAI were achieved, with feedback indicating areas for improvement, particularly for daily-life integration. Our findings provide preliminary support for the feasibility of the developed JITAI algorithm, demonstrating effective data processing and intervention triggering in real time, while also highlighting areas for improvement. Future research should focus on minimizing participant burden, including the intensity of EMA protocols, to improve participant adherence and acceptability while maintaining the benefits of real-time intervention delivery.

Keywords Ecological momentary assessment · Ecological physiological assessment · Ecological momentary intervention · Just-in-time adaptive intervention · Resilience · Stress · Mental health

Introduction

Background

Stress-related disorders, such as anxiety and depression, rank among the highest contributors to the global burden of disease worldwide (Baxter et al., 2014; James et al., 2018; Liu et al., 2020; Vos et al., 2012). Despite the availability of a range of pharmacological and psychological treatments, the prevalence of these disorders remains high. This underlines both the presence of a treatment gap and the substantial burden these conditions place on individuals and society at large (Jorm et al., 2017). Consequently, there has been a shift in the field away from solely investigating mental health disorders toward understanding factors and mechanisms underlying "*the maintenance or quick recovery of mental health despite adversity*," a concept known as resilience (Bonanno et al., 2011; Kalisch et al., 2017). This paradigm shift places a stronger focus on prevention and the promotion of mental health maintenance, in recognition of the limits of focusing on treatment approaches only. Resilience research emphasizes the importance of developing adaptive skills and strategies in advance to buffer against the potentially harmful effects of stress, thereby offering an approach to intervene

Extended author information available on the last page of the article

Published online: 01 December 2025

in individuals at risk for the development of stress-related disorders.

Together with these research developments, the evolution of mobile technologies has unlocked a promising new avenue for interventions (Marciniak et al., 2020; Riley et al., 2011). The emergence of smartphone-based ecological momentary interventions (EMI) has transformed the way therapy and prevention can be delivered, providing real-time and context-specific support in naturalistic settings (Balaskas et al., 2021; Heron & Smyth, 2010). Recent innovations have further enabled the delivery of EMIs at precisely those moments when individuals need it most, for example, during moments of high stress (Nahum-Shani et al., 2018; Wang & Miller, 2020). The delivery of such just-in-time adaptive interventions (JITAI) is assumed to significantly enhance the efficacy of EMIs (Marciniak et al., 2024; Xu & Smit, 2023).

Implementing JITAIs in the context of resilience crucially relies on the detection of stress in daily life (Smyth et al., 2023). The concept of stress, as originally defined by Hans Selye in 1950, describes it as the "*nonspecific response of the body to any demand*" (Selye & Fortier, 1950). Lazarus and Folkman's transactional theory of stress and coping, proposed in 1987, offers further insight. According to this theory, the concept of stress emerges from an individual's subjective appraisal of a given situation (Lazarus & Folkman, 1987). The subsequent stress response then leads to change across several domains, including physiological, psychological, cognitive, and behavioral domains (Schlotz, 2018).

Perceived stress levels can be monitored throughout the day by means of ecological momentary assessments (EMA) (Smyth & Heron, 2016). The strength of this approach lies in its ability to capture ecologically valid, within-person momentary fluctuations in stress levels, effectively minimizing the influence of recall bias (Myin-Germeys & Kuppens 2022). However, assessing perceived stress faces conceptual limitations. A recent study highlights this by showing that individuals who self-identify as "very stressed" differ significantly in psychological but not physiological stress factors compared to those who consider themselves "not stressed" (Lupien et al., 2022). This suggests that subjective feelings of stress might not always be reflective of the biological stress responses.

In addition to release of the stress hormone cortisol, bodily markers of the stress responses include increases in heart rate (HR) and skin conductance (SC) through activation of the sympathetic nervous system (Wijsman, 2014). Wearable devices offer a non-intrusive, continuous approach to measuring such physiological changes in daily life (Smets et al., 2019). However, relying purely on these ecological physiological assessments (EPA) for the operationalization of daily-life stress is insufficient, as EPA essentially measures arousal, which can signify both stress and positive excitement (Tutunji et al., 2023). Therefore, supplementing it with markers that can differentiate between positive and negative arousal is essential to acquire valid measurements. Combining the psychological (affective) and the physiological domains of the stress response is particularly suitable for this purpose due to the minimal intrusiveness and relative ease of data collection outside controlled environments (Myin-Germeys & Kuppens 2022).

Current study

The Dynamic Modelling of Resilience (DynaMORE) consortium has developed a state-of-the-art algorithm which integrates both ecological psychological and physiological data to provide a multifaceted operationalization of daily-life stress. Crucially, the developed pipeline operates in real time, comparing momentary stress levels against the individual's baseline stress state several times a day. Upon identifying elevated stress, the algorithm triggers a JITAI which is specifically designed to buffer the potentially harmful effects of stress. In addition to these triggered interventions, participants also have the option to self-trigger interventions whenever they feel the need for support. The triggering algorithm was tested in the DynaMORE interventional study (DynaM-INT), a multicenter study designed to investigate the efficacy of two EMIs targeted at resilience (Bögemann et al., 2023). Both interventions were developed by the DynaMORE consortium and were tailored to foster resilience by targeting two distinct resilience factors chosen for their potential to improve mental health despite adversity.

The first intervention, ReApp, targets positive cognitive reappraisal (Marciniak et al., 2023b), an important class of cognitive processes that generate positive appraisals in line with the Positive Appraisal Style Theory of Resilience (Kalisch et al., 2015). Positive appraisal style refers to the tendency of an individual to appraise potential stressors in a positive, yet realistic manner, avoiding delusional positive appraisals and thereby potentially reducing the perceived threat and emotional impact of stressors. The second intervention, Imager, targets reward sensitivity (Marciniak et al., 2023a, c), drawing upon recent findings that suggest its involvement in resilience (Dutcher, 2022; Dutcher & Creswell, 2018; Kalisch et al., 2024). This aligns with evidence that robust reward processing may buffer against the development of stress-related symptoms and may hence contribute to quicker recovery following adversity (Nielson et al., 2021).

In the current paper, we evaluate the feasibility of the developed JITAI algorithm by testing its ability to trigger interventions during times of high stress at a large scale. The JITAI algorithm aims to target the 30% most stressful periods of the day, individually calibrated for each participant. Rather than only detecting extreme stress levels, this approach ensures that interventions are provided consistently. In other words, even if some individuals do not experience extreme stress at all, or not in some intervention weeks, participants would still receive interventions. By maintaining consistent delivery, these interventions help establish a routine, increasing participant engagement and supporting long-term benefits. The initial stress threshold was set to 60% of the individual's baseline distribution, meaning that, retrospectively, 40% of all completed baseline beeps were labeled as being stressful. This threshold was selected based on feasibility considerations and prior experience with EMIs, aiming to deliver up to three interventions per day. As EMA compliance was expected to be around 70%, the threshold was set at 40% of all completed baseline beeps to approximate our goal of successfully triggering interventions during the 30% most stressful moments.

To assess the feasibility of the developed JITAI setup, we will focus on three key areas: technical implementation, participant adherence, and user experience. For technical implementation, we will evaluate the reliability and performance of the real-time decision pipeline, focusing on the stability of data uploads and the effectiveness of the threshold adjustment algorithm. These technical aspects are crucial to ensure the system operates reliably in real-world settings. Next, participant adherence will be assessed through the compliance rates of EMA questionnaires, both algorithm-triggered and self-triggered interventions, and the overall time participants spent using the application. High adherence rates will indicate that participants are engaging with the interventions as intended, which is crucial for the JITAI algorithm to function effectively, Lastly, user experience will be assessed to understand how participants perceive the app's usability and its impact on their mood and behavior during stressful periods. Positive user experience is vital for the long-term adoption and effectiveness of the intervention. These analyses were outlined in our protocol paper (Bögemann et al., 2023) and aimed to provide a comprehensive understanding of the feasibility of our JITAI algorithm. This feasibility study is an essential first step toward assessing the intervention's efficacy, which will be thoroughly evaluated in a subsequent paper. The significance of the current paper extends beyond technical validation as we attempt to make a methodological contribution to the field. By demonstrating the feasibility and scalability of our JITAI algorithm, we seek to establish a foundation for more effective, personalized health interventions, addressing a critical need in our healthcare systems.

Methods

Participants

The DynaM-INT study was conducted at five study sites: Charité—Universitätsmedizin Berlin, Department of Psychiatry and Psychotherapy in Berlin, Germany; Universitätsmedizin Mainz, Neuroimaging Center in Mainz, Germany; Donders Centre for Cognitive Neuroimaging and Radboudumc in Nijmegen, the Netherlands; Sagol Brain Institute, Tel Aviv University and Tel Aviv Soursaky Medical Center, Tel Aviv, Israel; University of Warsaw, Faculty of Psychology in Warsaw, Poland. The study was approved by the local ethics committees and was conducted in accordance with the Declaration of Helsinki. Written informed consent was obtained from all participants.

Participants were recruited between April 2022 and April 2023, using various outreach methods, including e-mail distribution lists, social media advertisements, flyers, digital blackboards, and word of mouth. The study specifically targeted students, as this group is known to be particularly vulnerable to stress-related psychopathology, with several mental disorders often first emerging in this life phase (Reavley & Jorm, 2010). Determining eligibility involved an initial online pre-screening and subsequent phone screening. Main inclusion criteria were age between 18 and 27, heightened internalizing problems (a score of ≥ 20, assessed with the General Health Questionnaire, 28-item version; Goldberg et al., 1997), and no recent (within 9 months before inclusion) diagnosis of any mental disorder other than a mild depressive episode, tobacco abuse/dependence, or substance abuse, as assessed by trained staff using the Mini-International Neuropsychiatric Interview (Sheehan et al., 1998). In total, 215 participants were included.

For the current analysis, however, we included a subset of 203 participants, as 12 participants dropped out before they had provided any daily-life data. As our main analysis centered around feasibility of the JITAI setup, we excluded participants who dropped out from the study before the start of the first JITAI week ($N=23$) due to an initial technical issue during threshold calculation ($N=2$) or where threshold calculation failed due to human error ($N=1$). As the implementation of the JITAI setup relied on timely threshold calculation, we lastly excluded weeks that mistakenly occurred before the participant's threshold was set (8 weeks, leading to the exclusion of one more participant). Therefore, the final sample that was analyzed in this paper consisted of 176 participants (Berlin: $N=19$; Mainz: $N=28$; Nijmegen: $N=33$; Tel Aviv: $N=29$; Warsaw: $N=67$).

Procedure

Upon inclusion, participants entered the baseline characterization phase in which a dense battery of different data types was collected. Measures included functional magnetic resonance imaging and blood and stool samples, as well as self-report questionnaires to assess baseline demographics and psychological traits. Part of the baseline characterization phase was 1 week of EMA and EPA. Following this initial "calibration week," participants were randomly assigned to one of two interventions for the 4-month EMI phase: ReApp, targeting positive cognitive reappraisal, or Imager, targeting reward sensitivity (Marciniak et al., 2023b, c).

The EMI phase consisted of a 2-week training period in which participants got acquainted with the assigned intervention through a fixed protocol. After a short break, participants engaged in three "booster weeks" in which the interventions were triggered during stressful moments, detected in real-time. These booster weeks were scheduled once per month to lower the participant burden in such an extensive trial. In weeks in between, participants were encouraged to continue practicing the intervention on their own phones. This approach has been shown to decrease the time cost of study participation without having a negative impact on the EMI's effectiveness and adherence to the interventions (Marciniak et al., 2024). The EMI phase is visualized in Fig. 1.

The study concluded with a follow-up phase in which part of the baseline battery was repeated to capture intervention effects. User experience was assessed with an adapted version of the user version of the Mobile Application Rating Scale (uMARS) questionnaire (Stoyanov et al., 2016). In the following sections, we will elaborate on elements of the design that are integral to assess the feasibility of our JITAI setup: the calibration and booster weeks, as well as the uMARS.

Calibration week

During the calibration week (6 days), participants were equipped with a study smartphone (Motorola Moto E6 Play in Berlin, Mainz, Nijmegen, and Warsaw; Xiaomi Redmi 7/7A in Tel Aviv) and the Chill +, a wrist-worn physiological and actigraphy wearable developed by consortium partner imec (https://www.imec-int.com/). EMA data was collected through the RADAR-aRMT application (Ranjan et al., 2019), which was further extended for the use in the DynaMINT study by software developers from The Hyve (https://www.thehyve.nl/). To ensure that the calibration week represented a typical baseline of daily-life stress, participants were instructed to select a week that did not include major planned stressors, such as exams or deadlines.

Ecological momentary assessments (EMA) Each day, participants received 10 EMA notifications following a fixed notification schedule with semi-randomly scheduled questionnaires (or "beeps") sent out in 90-min blocks (Table 1). After 5 min, a reminder notification was sent. Participants were instructed to answer the beep as soon as it arrived; each beep remained visible in the application for 10 min. Each beep contained the same short questionnaire assessing the participant's mood and context.

Upon completion, each beep was uploaded to the Donders Centre for Cognitive Neuroimaging at the Nijmegen site. A subset of EMA questions was processed in real time to quantify the participant's current affective state. For negative affect (NA), the score was calculated by averaging the responses to the following statements: *"I feel irritated," "I feel anxious," "I feel insecure," and "I feel sad."* Similarly, positive affect (PA) was calculated by averaging scores for the statements *"I feel happy," "I feel satisfied," and "I feel relaxed."* Finally, PA scores were reversed (rev-PA).

Ecological physiological assessments (EPA) Participants wore the Chill + wearable during daytime (approximately 16 h per day). The wearable measures galvanic skin responses, heart rate (photoplethysmography), movement (accelerometer), and skin temperature. To ensure reliable measurements, participants were instructed to wear the wristband securely to maintain continuous skin contact and to put it on immediately upon waking, removing it only during sleep. They were advised not to frequently remove and reapply the device, as this could impact data quality, and not

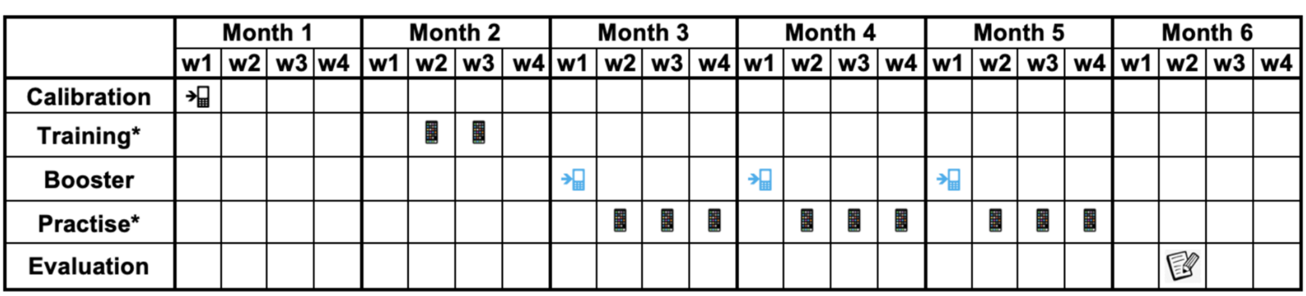

Fig. 1 Ecological momentary intervention (EMI) phase. The EMI phase contained one calibration week (to obtain observational baseline data from each participant), two training weeks (in which participants got familiar with the assigned intervention), three booster weeks (with JITAI), and nine encouraged practice weeks. To address the feasibility of the JITAI algorithm, only data collected during the calibration and booster weeks are analyzed. * Indicates elements that were part of the broader study design but were not analyzed in the current paper

Table 1 Semi-random beep schedule

Day 1	Day 2	Day 3	Day 4	Day 5	Day 6
7:54	8:02	7:41	7:57	8:00	7:31
10:03	9:02	9:32	9:05	9:17	9:33
10:41	11:13	11:12	11:24	11:27	10:32
12:46	13:26	13:04	12:28	12:59	12:33
13:56	13:53	13:47	13:52	13:32	14:34
15:13	16:21	16:28	16:07	15:08	15:12
16:32	17:21	17:06	16:33	17:16	17:37
18:56	18:13	18:54	19:24	18:15	18:56
20:14	20:17	20:49	20:11	20:00	20:25
22:21	21:05	22:29	21:59	22:09	22:20

During the initial calibration week, the beep schedule was fixed. During the booster weeks, the same schedule was used, however, with a variable starting day. Participants started the schedule on any day between day 1 and day 6, depending on the start date of their calibration week. The beep schedule operated on a continuous loop in the background.

to lend it to others. The wristband was not water-resistant and therefore could not be worn while taking a shower.

The wristband further featured a "stress" button, which participants were instructed to press whenever they felt stressed. EPA data consisted of 10-min segments, recorded immediately before each EMA notification. These data were, along with the completed EMA beep, uploaded to the Donders Centre for Cognitive Neuroimaging via a manual upload process, which required a single button press from the participant at the end of every EMA. The upload was enabled through a Bluetooth connection with the phone via the DynaMORE Chill + app, developed by consortium partner imec for use in the DynaM-INT study.

Similarly to the EMA data, the EPA data were processed in real time, using an in-house feature extraction algorithm developed by consortium partner imec. The uploaded segments were analyzed in 1-min bins. For each bin, the number and magnitude of spontaneous skin conductance (SC) responses (based on the methods described by Healey & Picard, 2000), the maximum and mean heart rate (HR) (based on a combination of a frequency-based and a time spectrum analysis, developed by consortium partner imec), and the total magnitude of acceleration (detected in the x, y, and z directions) were computed.

Finally, the number of button presses (indicating the subjective stress experience) was processed alongside the EPA data, with the total count recorded for each 10-min window.

The real-time feature extraction algorithm evaluated the quality of incoming data, using modality-specific signal quality indicators (SQI). Notably, these SQIs did not assess the reliability of stress measures themselves but ensured that only high-quality physiological data were included in the real-time analyses. For SC, the SQI was calculated based on methods described by Kocielnik and colleagues using an SQI threshold of 0.4 (Kocielnik et al., 2013). For HR, the quality of the photoplethysmography (PPG) signal was ascertained through an SQI derived from a comparison between the employed frequency-based and time spectrum analyses, along with a machine-learning algorithm developed by imec, trained on annotated PPG signals. Here, the SQI translated to a binary score, indicating whether the data was of sufficient quality or not. For both modalities, bins with insufficient quality were treated as missing, while the remaining values were combined to obtain one average value per feature. Additionally, the percentage of bins with sufficient quality was returned by the algorithm.

Threshold calculation

Following the calibration week, the real-time calculated EMA and EPA features (i.e., negative affect [NA], reversed positive affect [rev-PA]), the number and magnitude of spontaneous skin conductance [SC] responses, and the maximum and mean heart rate [HR]) were further analyzed to calculate individual stress thresholds and distribution parameters to be used during the subsequent booster weeks. Using a custom MATLAB [version 9.13.0 (R2022b), The MathWorks Inc., 2022] script, we calculated mean and standard deviations for each EMA and EPA feature. To ensure comparability across modalities, we applied z-scoring to standardize the collected features during the calibration week. These z-scores were then also stored for real-time standardization during subsequent booster weeks. Upon this standardization, z-scored features were averaged within modality, resulting in two separate composite scores. Specifically, for EMA, the z-scored NA and rev-PA were averaged, while for EPA, the z-scored SC-number, SC-magnitude, HR-mean, and HR-maximum were combined into a single EPA score per beep.

Next, a linear regression was fitted between the averaged EPA score and the total magnitude of motion, defined as total acceleration detected in the x, y, and z directions, measured by the accelerometer. This regression model was implemented to account for the fact that physiological arousal naturally increases with physical activity. Without adjusting for movement, the algorithm might therefore misinterpret elevated physiological arousal as stress-related. For each beep, a motion-corrected EPA score was calculated as the residual of the respective EPA score on the individual's regression line between all averaged EPA and motion features obtained during the calibration week. Again, the slope and intercept of this regression line were stored to repeat the motion correction in real time during the booster weeks.

Finally, the motion-corrected EPA score was averaged with the EMA score, to obtain a final composite "stress score." By combining affective ratings with physiological indices of sympathetic arousal, the composite EMA/EPA stress score primarily captures high-arousal negative affective states characteristic of acute stress responses, rather than lower-arousal states such as sadness or boredom. Together, all these stress scores (one per beep) from the calibration week formed the individual's distribution of stress levels during a "normal week" (i.e., without particular stressful events such as an important exam). During the calibration week, 60 beeps were triggered, meaning that these individual distributions consisted of up to 60 stress scores (depending on the participant's compliance). The initial stress threshold was set to 60% of this individual distribution, meaning that, in retrospect, 40% of all completed beeps were labelled as stressful for a particular participant. The choice of this specific threshold was based on feasibility considerations and experience in application of the EMIs in earlier studies (Marciniak et al., 2023b, c). We aimed to deliver up to three interventions per day and anticipated a compliance rate of 70%. Therefore, from the seven beeps that we expected to be answered, a 40% threshold would ensure that the top three beeps would trigger an intervention.

Booster weeks

During three monthly booster weeks (3 × 6 days), participants collected EMA and EPA data similarly to the calibration week. This time, however, the calculated features were standardized using the individual's distribution parameters from the calibration week in real time. Upon motion correction of the EPA data (described above), the EMA and EPA features (i.e., negative affect [NA], reversed positive affect [rev-PA]), the number and magnitude of spontaneous skin conductance [SC] responses, and the maximum and mean heart rate [HR]) were combined into a stress score. If the real-time stress score exceeded the participant's set threshold, the assigned intervention was triggered by sending a push notification through the RADAR-aRMT application on the study phone (Fig. 2). The real-time decision pipeline also triggered an intervention when the participant had pressed the stress button in the 10 min before the EMA notification, regardless of their stress level. This specific 10-min window was chosen to ensure alignment with the temporal resolution of the EPA data. To avoid overburdening participants, the algorithm never triggered more than four interventions per day. In case of missing EPA, the triggering decision relied on the EMA data.

Ecological momentary interventions (EMI) Each time an intervention was triggered, participants received their assigned digital interventions via a push notification through the RADAR-aRMT application, the same app used for EMA. These interventions consisted of short, text-based, self-guided exercises designed to be completed in approximately 2–3 min. The intervention was delivered

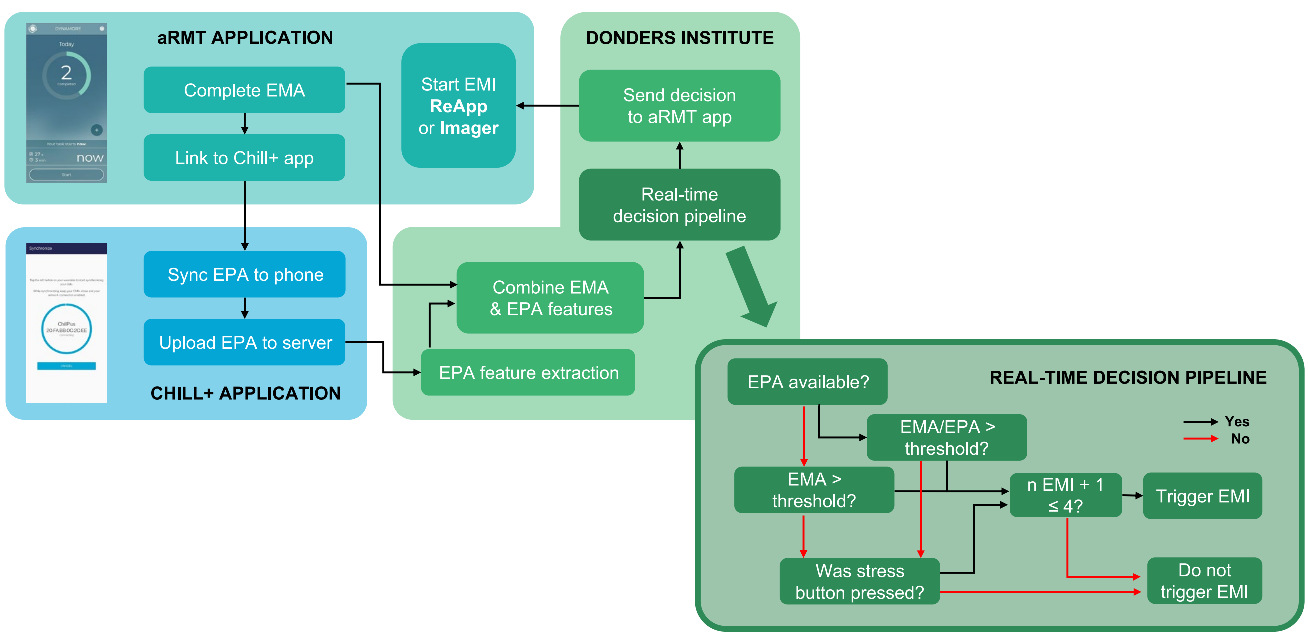

Fig. 2 Dataflow during booster weeks. This figure illustrates the data-processing steps across both applications. The RADAR-aRMT application is used for ecological momentary assessments (EMA) and ecological momentary interventions (EMI), while the DynaMORE Chill + application uploads ecological physiological assessments (EPA) to the servers at the Donders Institute, where real-time feature extraction occurs. In the real-time decision pipeline (pop-out window in the bottom right), a decision is made on whether it is an appropriate moment to trigger an EMI

approximately 20 min after the start of the EMA questionnaire and remained available for 1 h, allowing participants some flexibility in when they engaged with it.

At the end of each booster week day, the number of triggered interventions was compared to the desired number (three). If less than three interventions were triggered on a particular day, the threshold would be decreased by 0.01 for the next day. This small adjustment lowers the stress level required to trigger an intervention, increasing the likelihood of interventions being triggered. Conversely, if more than three interventions were triggered on any given day, the threshold would be increased by 0.01, raising the required stress level and reducing the likelihood of triggering interventions. The increments of 0.01 were determined through in-house simulations, aimed at identifying an optimal number to keep triggering three interventions per day throughout the study.

ReApp intervention At the start of each ReApp EMI, participants were asked to describe the most negative event they had experienced since the last app contact. If they had not encountered a recent stressful event, they could instead choose a stressful future event that was on their mind. Participants were prompted to generate three different reappraisals for the event by considering possible unexpected positive aspects, identifying lessons learned, and reflecting on the advice they might give to a friend in a similar situation or the perspective a trusted person might offer. They typed their responses directly into the app, encouraging engagement with the exercise.

Imager intervention At the start of each Imager EMI, participants were asked to think about a pleasant event that was going to happen within the next few hours and to describe it in the first-person perspective. If they were unsure what to answer, participants saw an example list of other people's events (e.g., I will listen to music; I will see or show photos). They were then guided through a mental imagery exercise, in which they visualized themselves actively engaging in the event. To facilitate vivid imagery, they were advised to close their eyes and take their time in constructing the scene. After this visualization, participants typed the three most important aspects of their imagined experience into the app, describing, among others, what they saw, felt, and heard.

During the booster week, participants were encouraged to complete additional interventions whenever they felt like they could benefit from it. This encouragement was framed as a valuable opportunity to improve their skills trained by the assigned intervention. Finally, participants were asked to complete at least one intervention per day, always right before going to bed. These interventions were self-triggered by the participant.

Data analysis

Data analysis was performed in R (version 4.2.1, R Core Team, 2021). To assess the feasibility of the developed JITAI setup, we focused on three feasibility questions (fQ), assessing the technical implementation of the real-time decision pipeline (fQ1), participant adherence (fQ2), and user experience (fQ3). These analyses were outlined in the protocol paper (Bögemann et al., 2023).

fQ1: Technical implementation

To assess the technical implementation of the real-time decision pipeline, we employed six different implementation metrics, each focusing on different aspects of the pipeline's performance and reliability.

Decision pipeline First, we assessed the percentage of EMA beeps that yielded successful EPA uploads and feature extractions per booster week. This measure indicates the reliability of our system in uploading and processing real-time data. Second, we determined the number of minutes per EPA upload in those weeks, indicating the stability of the Bluetooth connection between the study devices. Third, we determined the number of triggered interventions per day and the percentage of triggered interventions per week in each booster week. These measures are crucial for the technical implementation, as they describe the outcome of our real-time decision pipeline, as well as its ability to trigger EMIs successfully based on daily-life data.

Finally, we evaluated the effectiveness of the threshold adjustment algorithm by comparing the number/percentage of interventions triggered against a hypothetical scenario where interventions would be triggered based on a constant, unadjusted threshold. This comparison involved simulating the decision pipeline as if it only operated with the initial threshold (i.e., the stress threshold derived from the individual distribution of stress scores collected during the calibration week) throughout the entire study, allowing us to evaluate the effectiveness of the threshold adjustment algorithm. We will consider the JITAI algorithm to outperform the unadjusted threshold if it identifies extra moments for triggering interventions or maintains the desired number of three triggered interventions more accurately per day. Specifically, we expected this algorithm to maintain the desired number of three triggered interventions more accurately per day, compared to a scenario where the threshold remained at its initial value throughout the study. Alternatively, it is also possible that participant's stress levels remain stable with no significant drift, in which case the need for threshold adjustment may be minimal due to the stability of the signals.

It is important to note that the six implementation metrics have a descriptive nature. They are intended to provide merely a description of the system's performance, offering insights into the operational aspects of the developed JITAI pipeline. Nevertheless, we did investigate whether the computed week-level metrics differed between weeks using linear mixed models (LMMs), employing the lme4 package (version 1.1.31) (Bates et al., 2015). Models included a fixed intercept and independent variable "week" (modelled as factor). To account for individual variability, we included a random intercept, resulting in the following model: $IM_{ij} = \beta_0 + \beta_{1weekj} + u_i + \epsilon_{ij}$, where β_0 represents the intercept, β_{1weekj} the fixed effect of week, u_i the participant-specific random intercept, and ϵ_{ij} the residual error term.

Stressful moments To assess whether we managed to capture the most stressful moments of the day, we compared the real-time EMA and EPA features (i.e., negative affect [NA], reversed positive affect [rev-PA]), the number and magnitude of spontaneous skin conductance [SC] responses, and the maximum and mean heart rate [HR]) of beeps that did and did not trigger an intervention. Using separate LMMs for each feature, we investigated the association between "beep type" (factorized: "baseline," "triggered," and "not triggered"). Again, models included a fixed intercept to represent the overall effect. We adopted a nested random intercept structure to acknowledge the hierarchical nature of the beep-level data (i.e., beeps nested in weeks, nested in participants), resulting in the following: $feature_{ijk} \sim \beta_0 + \beta_{1beep.typej} + u_i + u_{ij} + \epsilon_{ijk}$, where β_0 represents the intercept, $\beta_{1beep.typej}$ the fixed effect of beep type, u_i the participant-specific random intercept, u_{ij} the week-specific random intercept nested within participants, and ϵ_{ijk} the residual error term.

Representativeness In addition to the analyses that were outlined in the protocol paper (Bögemann et al., 2023), we decided to evaluate the representativeness of the calibration week for determining the initial threshold. For this, we computed the intraclass correlation coefficient (ICC) (Hedges et al., 2012; Myin-Germeys & Kuppens 2022). By calculating the ICC, we aimed to determine how closely the features collected during the calibration week reflected those in the subsequent booster weeks. To enhance the accuracy of this assessment, we derived the ICC from the nested random intercept structure described above, as it differentiates between-participant, within-participant (across weeks), and residual (within-week) variances. ICCs were calculated for both the participant and week level to understand where the sources of variability lie. A high participant-level ICC suggests that a large portion of the overall variance in the data is attributable to differences between participants. Similarly, a high week-level ICC suggests large variability across weeks for the same participant (i.e., a weak consistency within participants over time). In the current context, a low week-level ICC would support that the calibration week provides a representative baseline measure for the subsequent booster weeks. However, it is also important to consider that if our intervention is effective, it may reduce stress, which could, in turn, affect the triggering of beeps and influence the ICC values over time.

fQ2: Participant adherence

To assess participant adherence, we again computed various metrics (now: adherence metrics) to gain insights into different aspects of participant interaction with the study protocol. First, we determined the general attrition rate, calculated as the percentage of participants who discontinued during the EMI phase of the DynaM-INT study. This attrition rate is a critical indicator of the overall feasibility and acceptability of our study protocol. Second, we determined the percentage of completed EMA questionnaires (or EMA compliance). Achieving a high level of EMA compliance is a practical necessity for the operational success of our JITAI setup. Without adequate responses, the pipeline lacks the crucial data needed to accurately assess stress levels and trigger interventions. Third, we calculated the percentage of completed triggered interventions (or EMI compliance). Although EMI compliance does not directly impact the setup, it is crucial for evaluating the practicality of assigning interventions during times of stress. Both EMA and EMI compliance were objectively recorded within the app interface, with completion defined as submitting the full EMA questionnaire or progressing through all components of the intervention module, respectively.

Next, we calculated both the number of completed self-triggered interventions and self-triggered evening interventions. Also, the total intervention adherence, which includes all completed triggered and self-triggered interventions, was determined. These numbers are indicative of the participant's engagement with the assigned intervention and focus on the potential difference in engagement between both methods (rather than focusing on the difference in completion per se). Comparing the number of self-triggered interventions to the number of triggered interventions might further inform us about the utility of our JITAI design. It could highlight, for example, that participants preferred to engage with interventions at a later time when stress levels had passed their peak. Finally, we calculated the time participants spent using the RADAR-aRMT application, distinguishing between time spent on EMA, EMI, and the total usage time.

We calculated all adherence metrics for the entire study combined as well as separately for each week. This approach allowed us to track potential changes in adherence over time. Similar to the implementation metrics, time effects were tested statistically using separate LMMs per week-level adherence metric: $CM_{ij} = \beta_0 + \beta_{1weekj} + u_i + \epsilon_{ij}$, where

β_0 represents the intercept, β_{1weekj} the fixed effect of week, u_i the participant-specific random intercept, and ϵ_{ij} the residual error term.

fQ3: User experience

User experience (fQ3) was assessed with a shortened version of the uMARS questionnaire (Stoyanov et al., 2016). This scale is developed to assess the quality and usability of mobile applications. In the DynaM-INT study, we only assessed the functionality (Cronbach's $\alpha = 0.67$) and subjective quality subscales (Cronbach's $\alpha = 0.77$). The functionality subscale measured performance ("*How accurately/fast do the app features and components work?*"), ease of use ("*How easy is it to learn how to use the app; how clear are the menu labels, icons, and instructions?*"), navigation ("*Does moving between screens make sense; does the app have all necessary links between screens?*"), and gestural design ("*Do taps, swipes, pinches, and scrolls make sense, and are they consistent across all components/screens?*"). The subjective quality subscale assessed recommendation ("*Would you recommend this app to people who might benefit from it?*"), expected future use ("*How many times do you think you would use this app in the next 12 months if it was relevant to you?*"), willingness to pay ("*Would you pay for this app?*"), and overall rating ("*What is your overall (star) rating of the app?*"). Responses were given on a five-point Likert scale, and scores were calculated as the sum of the subscale items, divided by the maximum possible sum (range [0–100%]), following Marciniak et al. (2023c).

In addition, we added two specific questions related to our JITAI setup. First, we included one open item: "*What changes did you observe, for example, in your mood, behavior,* etc.*, while using the app?*" Here, we translated the reported answers to English using Google translate. Answers were manually categorized into common themes to describe the various reported user experiences. Second, we assessed "*Did the app help you use your skills during relatively stressful periods?*" to directly probe into the central concept of our JITAI setup. For this last question, we depict a distribution of answers, which were rated on a scale from 1 (not at all) to 7 (very much).

Results

Sample

The analyzed sample consisted of 176 students ($N = 131$ female, $N = 42$ male, $N = 3$ missing; mean age = 22.2 years, standard deviation [SD] = 2.20), with 85 participants in the ReApp group and 91 in the Imager group. They completed in total 176 calibration and 495 valid booster weeks, with 5,798 and 11,693 answered EMA beeps, respectively. In total, 3,512 EMIs (ReApp: $N = 1{,}519$; Imager: $N = 1{,}993$) were completed during the booster weeks. Despite wearing the Chill + wearable (on average 6 h and 5 min per day [SD = 2 h 31 min] during the calibration week, and 5 h 57 min [SD = 2 h 27 min], 6 h 5 m [SD = 2 h 38 min], and 6 h 11 min [SD = 2 h 38 min] in booster weeks 1 to 3), participants rarely engaged with its stress button. While the button could be pressed at any time, our current analysis focused only on presses that occurred during the 10-min EPA windows preceding each EMA beep, as these were the windows used for real-time triggering. Within those windows, out of 176 participants, 146 never used it, 20 pressed the button once, seven pressed it twice, and only three participants pressed it three times. Four participants were unable to use the Chill + wearable for the study due to skin irritations caused by the wristband; their data was hence omitted from the analysis assessing EPA performance.

In order to effectively illustrate the collected stress scores, we present descriptive figures for a subset of participants. To ensure clarity of the visual representation, we specifically selected participants with compliance rates of at least 50%. Figure 3 displays the individual distributions of stress scores collected during the calibration weeks of four randomly selected participants (compliance for participant A = 75% [45 beeps]; B = 68% [41 beeps]; C = 98% [59 beeps]; D = 80% [48 beeps]).

Figure 4 then depicts the stress scores collected during the booster weeks for one of these participants (see Supplemental Fig. 1 for the three other example participants). Crucially, this figure highlights each time the stress score exceeded the personalized stress threshold (represented by the horizontal line). At these moments, the JITAI algorithm triggered an intervention (indicated by the vertical lines). Although the JITAI algorithm generally functioned within the expected range of one to four interventions per day, the occurrence of more than four triggered interventions on two specific days (Booster 1 (B1)—day 2; B3—day 5) was not anticipated.

fQ1: Technical implementation

Decision pipeline

On average, we found that 68.36% (SD = 28.64%) of completed EMA beeps per week contained successfully extracted EPA features. Numbers were significantly higher during the calibration week (76.03%) than during the three booster weeks B1 (66.84%, $\beta = -9.19$, $P < .001$), B2 (65.00%, $\beta = -11.03$, $P < .001$), and B3 (64.34%, $\beta = -11.69$, $P < .001$). EPA uploads contained on average 9 min and 48 s (SD = 28 s) of data. This number was slightly lower during the first booster week (calibration = 9 min 51 s; B1: 9 min

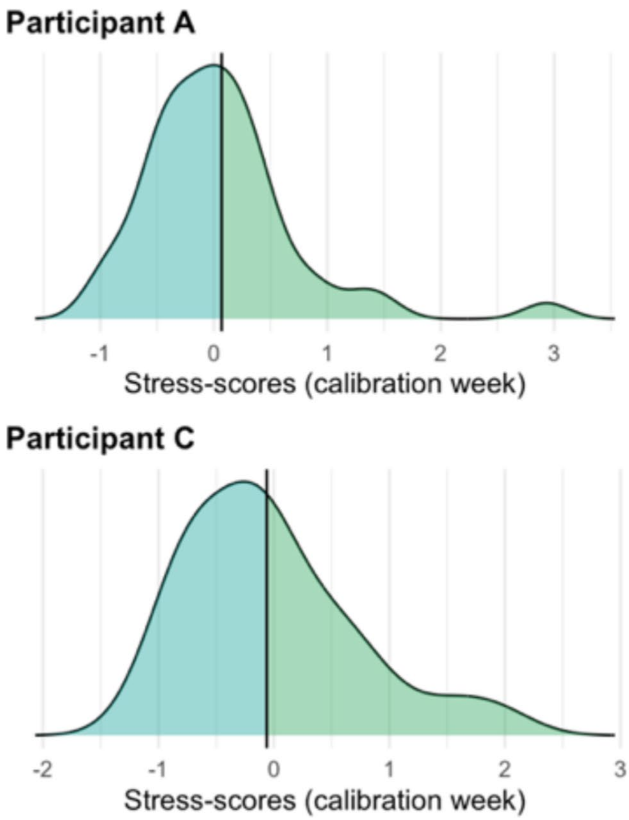

Fig. 3 Distribution of stress scores for selected participants. This figure shows stress scores collected during the calibration week for four randomly selected participants with compliance rates over 50%. Compliance rates were as follows: participant A = 75% (45 beeps); B = 68% (41 beeps); C = 98% (59 beeps); D = 80% (48 beeps). Stress scores exceeding the threshold (indicated by a black vertical line) are highlighted in green and labeled as stressful moments, whereas scores below this threshold are shown in blue

44 s, $\beta = -.11$, $P = .040$), but remained stable throughout the rest of the study (B2: 9 min 49 s, $\beta = -.03$, $P = .603$; B3: 9 min 46 s, $\beta = -.09$, $P = .100$). The average duration of EPA uploads remained close to the desired 10 min, demonstrating an effective data upload process. The uploaded EPA data contained high-quality signals, with quality indicators showing that 83.35% ($SD = 20.86\%$) of the SC-features and 87.08% ($SD = 9.58\%$) of the PPG-features were of high quality. Despite the overall success, it is important to consider the 32% of EMA beeps that did not result in successfully extracted EPA features, likely due to technical issues or an unstable Bluetooth connection.

On average, 1.61 ($SD = 1.26$) interventions were triggered per day. We found that the number of triggers decreased at the end of the study, from 1.74 triggered EMIs per day in B1 and 1.62 in B2 ($\beta = -.12$, $P = .175$) to 1.43 triggers per day in B3 ($\beta = -.31$, $P < .001$). We found that 42.57% ($SD = 27.43\%$) of completed EMA beeps per week triggered an intervention, on average. Here, we did not observe a difference between booster weeks (B1: 43.04%; B2: 42.57%, $\beta = -.46$, $P = .829$; B3: 42.73% $\beta = -.30$, $P = .890$). This suggests that the decrease in absolute number was not due to a decreasing performance of the JITAI pipeline but rather to a decreasing compliance of our participants (also see fQ2 Participant adherence).

Finally, we simulated the JITAI pipeline without conducting the daily threshold adjustment. Keeping the initial threshold constant throughout the study would have resulted in 1.55 ($SD = 1.26$) daily interventions. Similar to the implemented design, the number of triggers would have decreased at the end of the study, from 1.68 triggered EMIs per day in B1 and 1.56 in B2 ($\beta = -.13$, $P = .161$) to 1.38 triggered EMI per day in B3 ($\beta = -.30$, $P < .001$). Without daily threshold adjustment, 41.49% ($SD = 27.86\%$) of completed EMA beeps per week would have triggered an intervention on average, and throughout the whole study (B1: 43.13%; B2: 41.46%, $\beta = -1.67$, $P = .426$; B3: 41.27%, $\beta = -1.86$, $P = .382$). Based on these numbers, the daily threshold adjustment did not add value to the developed JITAI pipeline, and it could be reconsidered that the initial threshold already achieved the goal of triggering intervention in 40% of the completed beeps. Supplemental Table S1 provides details of the implementation models.

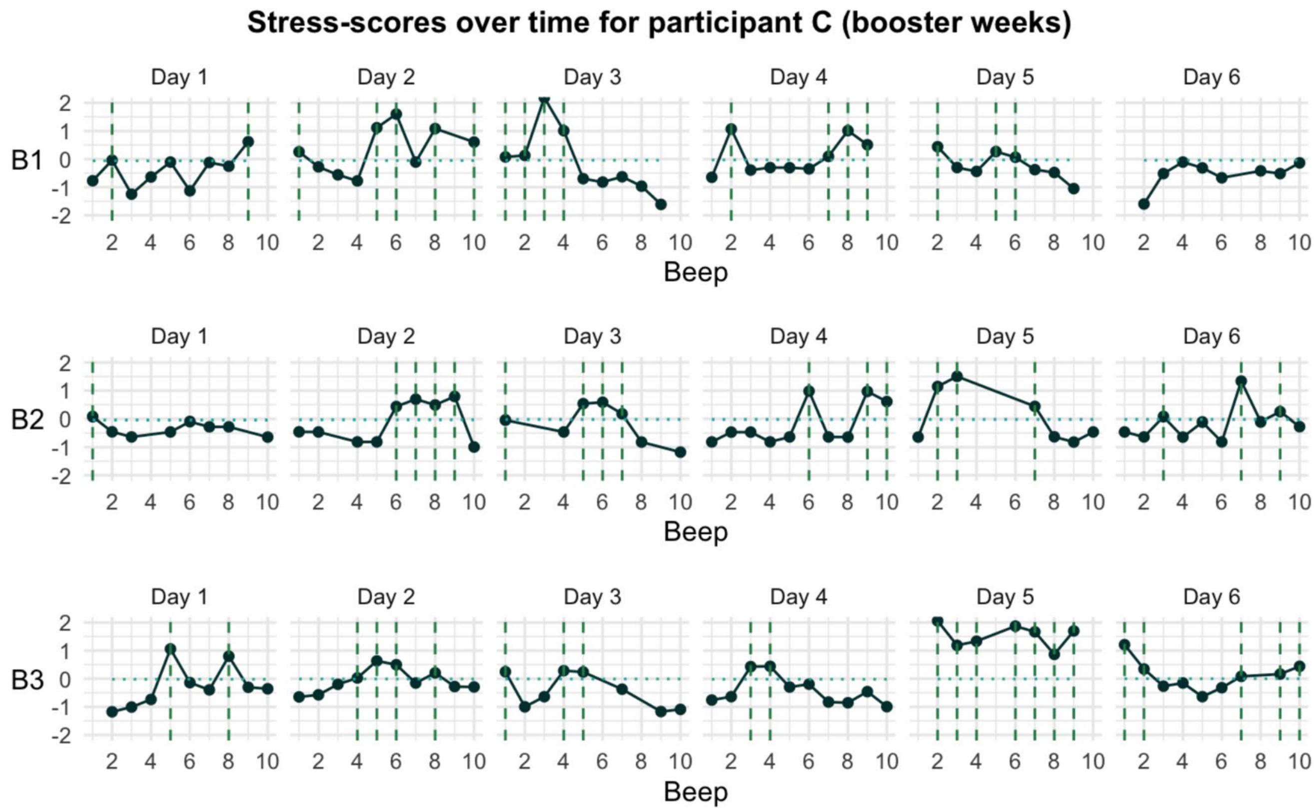

Fig. 4 Stress scores over time for one example participant with compliance over 50%. This figure shows stress scores across three booster weeks for one example participant with compliance rates over 50%. Compliance for this participant during B1 = 87% (52 beeps), B2 = 85% (51 beeps), and B3 = 88% (53 beeps). The figure depicts how stress levels fluctuate throughout the day and across weeks. Each dot represents a stress score, which was calculated in real time, upon completion of each EMA beep. Horizontal lines indicate the daily stress threshold, while vertical lines mark indicate when the JITAI algorithm triggered a JITAI. B = booster week

Stressful moments

All raw EMA and EPA features were significantly higher for beeps that triggered an intervention than for those that did not (NA: β = .75, rev-PA: β = .99, SC-number: β = 2.44, SC-magnitude: β = 1.29, HR-maximum: β = 11.54, HR-mean: β = 8.04; all $P < .001$; details in Supplemental Table S2). In addition, all raw features were significantly higher for triggered (NA: β = .37, rev-PA: β = .41, SC-number: β = .83, SC-magnitude: β = .47, HR-maximum: β = 6.63, HR-mean: β = 4.66; all $P < .001$) and lower for non-triggered beeps compared to baseline beeps that were collected during the calibration week (NA: β = −.39, rev-PA: β = −.59, SC-number: β = − 1.64, SC-magnitude: β = −.83, HR-maximum: β = − 4.77, HR-mean: β = − 3.30; all $P < .001$; details in Supplemental Table S3).

To validate the implementation of the triggering algorithm, the same comparisons were performed on the standardized features. These standardized features were computed using each participant's individual distribution from the calibration week and were used by the triggering algorithm in real time. These standardized features showed a similar pattern: all were significantly higher for triggered beeps compared to non-triggered beeps (NA: β = 1.03, rev-PA: β = 1.00, SC-number: β = 1.18, SC-magnitude: β = 2.82, HR-maximum: β = .86, HR-mean: β = .95; all $P < .001$; details in Supplemental Table S4) and compared to baseline (NA: β = .63, rev-PA: β = .46, SC-number: β = .95, SC-magnitude: β = 3.46, HR-maximum: β = .52*;* all $P < .001$; HR-mean: β = .72, $P = .017$; details in Supplemental Table S5). Compared to baseline, however, only half of the features of the non-triggered beeps were lower (NA: β = −.42, rev-PA: β = −.55, HR-maximum: β = −.33; all $P < .001$; SC-number: β = −.24, $P = .188$, SC-magnitude: β = .68, $P = .385$ and HR-mean: β = −.24, $P = .422$). In other words, SC-number, SC-magnitude, and HR-mean did not differ between baseline beeps during the calibration week and the moments that were classified as "not stressful" during the booster weeks. This is not problematic, as the calibration week was deliberately scheduled during a typical, non-stressful week, while booster weeks could coincide with both everyday situations and particularly stressful events, such as an important exam.

All composite scores were higher for beeps that triggered an intervention compared to non-triggered beeps (average EMA: β = 1.02, average EPA: β = 1.47, motion-corrected EPA: β = 1.59 and the final stress score: β = 1.38; all $P < .001$; details in Supplemental Table S6) and compared to baseline (average EMA: β = .55, average EPA: β = 1.43, motion-corrected EPA: β = 1.52 and the final stress score: β = 1.00; all $P < .001$; details in Supplemental Tables S6 and S7). Similar to the pattern described above, both EPA composite scores did not differ from baseline during "not stressful" moments in the booster weeks (average EPA: β = −.03, $P = .883$, motion-corrected EPA: β = −.05, $P = .820$). The average EMA and the stress score were lower during the non-triggered beeps compared to baseline (average EMA: β = −.49, final stress score: β = −.37; both $P < .001$).

Representativeness

ICCs were calculated for the hierarchical levels of participant and weeks, with lower ICCs (close to zero) implying greater variability and higher ICCs (close to one) implying greater similarity within units of the respective level (i.e., within participants or within weeks within participants; see Table 2). Hence, a representative calibration week would be characterized by features with a higher participant-level compared to week-level ICC, because this pattern would represent similarity between the different weeks of a given participant. For example, a participant ICC of.42 and week ICC of.13 for NA imply that 42% of the variance in NA is due to differences between individuals, 13% is due to week-to-week variations within individuals, and the remaining 45% is due to momentary fluctuations within weeks.

For the EMA, the participant-level ICC was greater than the week-level ICC, indicating that the differences between participants were more pronounced than the differences between weeks within the same participant. While HR features followed a similar pattern, their absolute participant-level ICCs were relatively low (HR-mean = .19, HR-maximum = .08; compared to NA = .42; rev-PA = .30), indicating that heart rate measures exhibited more variability across weeks within the same participant. For SC features, the pattern was reversed, with week-level ICCs being greater than participant-level ICCs. This indicates that skin conductance responses varied more within the same participant across different weeks than between different participants. For all features (including SC), the greatest source of variability was present at the momentary or beep level. This high moment-to-moment variability suggests that the selected EMA and EPA features are highly dynamic, with significant fluctuations occurring within the week. This large variability at the momentary level indicates that despite the representativeness of the calibration week, most of the variance is due to immediate, situational changes rather than consistent patterns within a week.

fQ2: Participant adherence

From all participants enrolled in the RADAR-aRMT application ($N = 203$), 154 completed all booster weeks (76%), resulting in an attrition rate of 24%. On average, participants had an EMA compliance of 43.45% ($SD = 22.19\%$) and an EMI compliance of 42.77% ($SD = 30.43\%$). For both EMA and EMI, compliance dropped throughout the course of the study (Fig. 5). For EMA, numbers were significantly higher during the calibration week (54.91%), compared to B1 (41.50%, β = − 13.41, $P < .001$), B2 (38.70%, β = − 16.21, $P < .001$), and B3 (36.07%, β = − 18.84, $P < .001$). For EMI, compliance dropped from 47.12% in B1 to 39.20% in B2 (β = − 7.92, $P = .004$) and to 39.99% in B3 (β = − 7.13, $P = .010$).

Exploratively, we examined the start time of the EMI relative to the EMA that triggered the intervention. On average, participants started the EMI 13.73 min ($SD = 12.26$) after completing the EMA. This suggests that participants typically engaged with the EMI relatively quickly, despite it being available for 60 min, indicating that heightened stress levels did not delay engagement.

Participants completed 0.58 ($SD = 1.24$) self-triggered interventions per week and 2.48 ($SD = 2.35$) self-triggered evening interventions. Together with the number of triggered

Table 2 ICCs

Feature	Mean	Variance			ICC	
		Participant	Week	Beep	Participant	Week
NA	2.04	.50	.15	.54	.42	.13
rev-PA	3.36	.46	.21	.86	.30	.14
SC-number	3.85	1.42	3.07	12.30	.08	.18
SC-magnitude	1.20	.59	1.08	5.37	.08	.15
HR-mean	81.88	32.81	7.74	128.97	.19	.05
HR-maximum	94.55	26.92	10.10	289.24	.08	.03

Participant-ICC represents the proportion of total variance attributed to between-participant differences, and week-ICC indicates the proportion due to between-week (within-person) variance. The remaining proportion represents within-week fluctuations of the respective feature

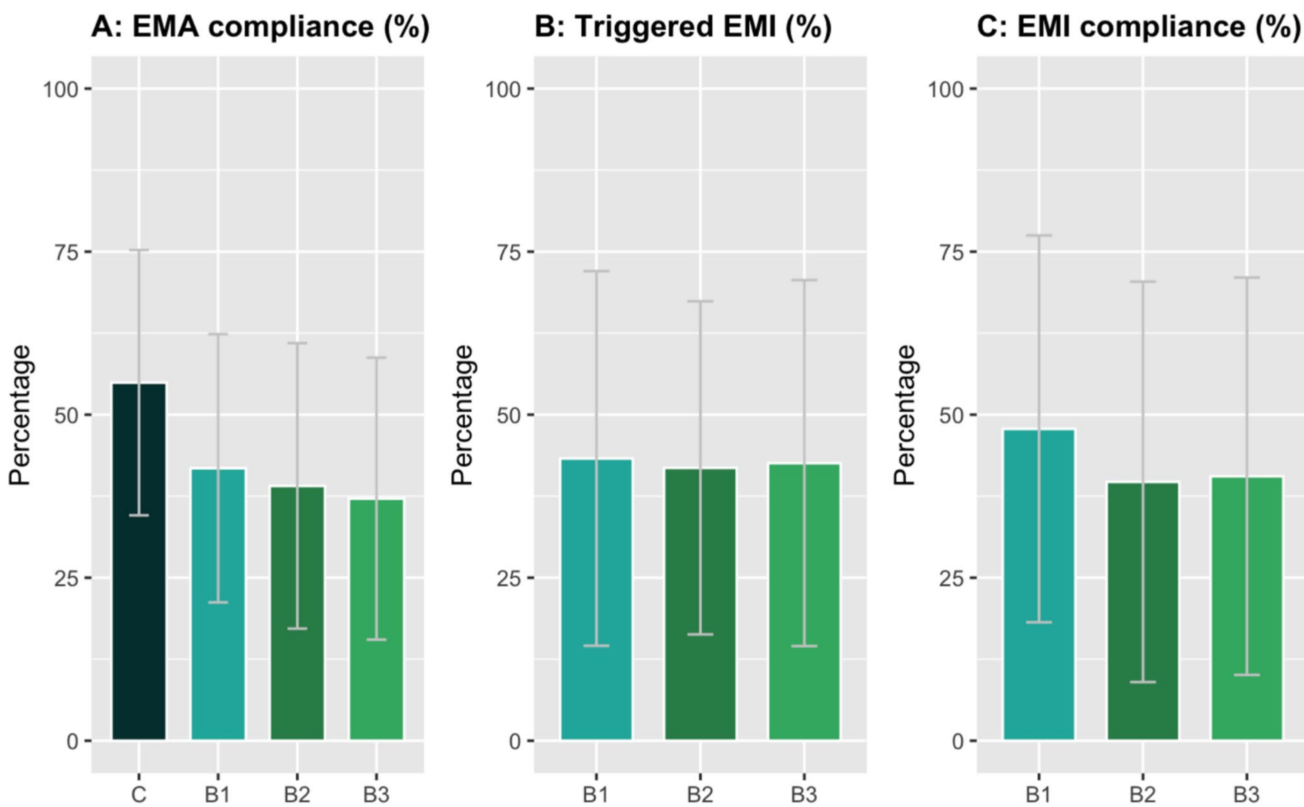

Fig. 5 Adherence and implementation in percentage per week. **A** EMA compliance in weeks (100% = 60 EMA beeps triggered per week); **B** percentage of beeps that triggered an EMI; **C** EMI compliance in weeks (100% = the number of triggered EMI beeps in that week). C = calibration week; B = booster week

interventions per week (4.04; $SD = 4.35$), this led to a total intervention adherence of 7.09 ($SD = 5.69$) EMIs per week. All intervention adherence dropped throughout the course of the study. For the total intervention adherence, numbers dropped from 8.49 completed EMIs in B1 to 6.65 in B2 ($\beta = -1.84$, $P < .001$) to 5.84 in B3 ($\beta = -2.65$, $P < .001$). Supplemental Table S8 provides details of the compliance models.

On average, participants spent 65.11 ($SD = 66.59$) min using the RADAR-aRMT application per week, divided over 49.09 ($SD = 49.19$) min answering EMA beeps and 16.02 ($SD = 29.71$) min completing EMIs. Similar to compliance, the time spent using the RADAR-aRMT application dropped throughout the study. For EMA, minutes remained relatively stable from the calibration week (67.26 min) to B1 (58.82 min), though there was a trend toward a decrease ($\beta = -8.44$, $P = .063$). This decline became more pronounced over time. Compared to the calibration week, it dropped to 50.67 in B2 ($\beta = -16.59$, $P < .001$) and to 48.47 in B3 ($\beta = -18.79$, $P < .001$). For EMI, minutes again remained similar at first, from 25.50 min in B1 to 21.25 min in B2 ($\beta = -4.25$, $P = .213$), but dropped to 15.70 min in B3 ($\beta = -9.80$, $P = .005$). For details, see Supplemental Table S9.

fQ3: User experience

In total, 126 participants completed the uMARS questionnaire (ReApp, $N = 61$; Imager, $N = 65$). For both interventions, high scores were assigned to the functionality subscale (ReApp, 79%; Imager, 80%, $t(124) = -.44$, $P = .662$). For the subjective subscale, which evaluates participant's intentions (including whether participants would recommend the app to others, use it in the next 12 months, consider paying for it, and what overall star rating they would assign), both interventions score lower yet similar ratings (ReApp, 54%; Imager, 51%, $t(124) = -.97$, $P = .335$).

The first additional question "*What changes did you observe, for example, in your mood, in your behavior etcetera, while using the app?"*, contained responses which described changes with respect to either the effectiveness of the specific interventions or the usability of the JITAI setup. All categorized responses are provided in the Supplemental Tables S10–S17. Here, we will focus on the usability of the JITAI setup. Twenty-four participants expressed usability concerns such as irritation over the frequent notifications and the need to watch the study phone (e.g., *"[…] the need to constantly check notifications/complete surveys forced us to use electronic devices more often and caused a feeling of* *stress."*) or technical issues like app crashes or slow performance. Also, annoyances generated by the phone (e.g., *"[…] Sometimes it annoyed me to always have to have my phone at hand, although nowadays people tend to consciously try to reduce cell phone use. […]"*) or Chill + wearables (e.g., *"[…] it looked like a bracelet from prison […]"*) were reported. For details, see Supplemental Tables S17.

The second additional question *"Did the App help you use skills during relatively stressful periods?"* further addressed the JITAI component of the intervention effectiveness. Figure 6 depicts the distribution of answers, with ReApp receiving an average rating of 3.98 (out of 7) and Imager receiving an average rating of 3.83 ($t(124) = .57$, $P = .568$), indicating no significant difference between the two interventions in their perceived effectiveness. On average, both apps were rated as moderately helpful in managing stress during the intervention period, with scores slightly below the scale midpoint.

Discussion

Principal results

This study evaluated the feasibility of the JITAI algorithm used within the DynaMORE interventional study (DynaMINT). The algorithm was designed to integrate psychological and physiological features that were collected through EMA and EPA in daily life. Features were integrated in real time several times a day, upon which they were compared to a personalized stress threshold. When moments of heightened stress were detected, the algorithm promptly triggered an EMI that was tailored to mitigate the adverse effects of stress. Findings from our study provide preliminary support for the technical feasibility of the JITAI algorithm. Participant adherence, while moderate, suggests that individuals could engage with the JITAI algorithm in their daily life. User feedback further supported the feasibility and acceptability of the developed JITAI algorithm, but also identified areas needing improvement.

The technical implementation of our JITAI algorithm was demonstrated to be successful, with the system performing real-time data processing operations as intended. Retrospective analyses revealed that both the individual EMA and EPA features, as well as the composite stress scores, were significantly higher during moments that triggered an EMI than during those that did not. Although these analyses were performed post hoc (i.e., after data collection), they revealed that the algorithm was able to successfully implement the predefined decision logic in real time. While our study demonstrated that both individual EMA and EPA features were significantly higher during moments that triggered an intervention, it raises the important question of whether either measure alone would be sufficient. Physiological measures can, however, be elevated in both highly negative and highly positive affective states (Tutunji et al., 2023), which could

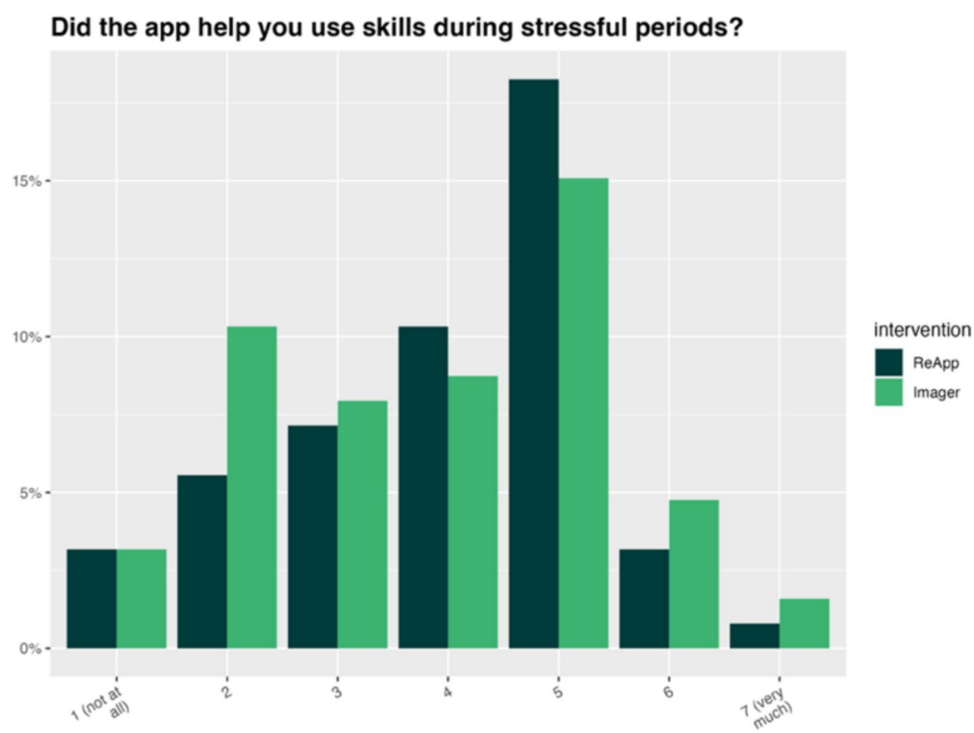

Fig. 6 Did the app help you use skills during relatively stressful periods? Depicted are subjective user ratings, categorized per intervention

lead to inappropriate triggers if used alone. Conversely, relying solely on affectivity might miss critical physiological indicators, leading to less precise triggering. This underscores the importance of integrating both modalities to ensure the validity and reliability of the triggers, as well as the need for continued refinement of the algorithm to better distinguish between different emotional states.

On average, 43% ($SD = 27\%$) of completed EMA beeps per week triggered an intervention, indicating that the stress scores frequently surpassed the personalized threshold. Thus, it appeared feasible to use the algorithm to compare momentary stress scores to a predetermined threshold. Moreover, we demonstrated intra-individual stability within the EMA features over time, supporting the use of a baseline week for personalized threshold calibration. However, this was less evident for physiological features. While heart rate followed a similar pattern, its relatively low participant-level ICCs suggest substantial variability across weeks, which may challenge the validity of a single baseline week. Additionally, for both heart rate and skin conductance, beep-level fluctuations accounted for the largest portion of variance. This suggests that while a baseline week is a suitable for EMA features, further research is needed to determine whether alternative calibration strategies would be more appropriate for physiological features.

We further showed that the majority of completed EMA beeps contained EPA features as well (68%). Instances without EPA features might suggest technical failure. However, the durations of these successful EPA uploads were close to the intended 10 min, suggesting that the Bluetooth connections were effective and stable. Additionally, the quality of the EPA uploads was high, with 83.35% ($SD = 20.86\%$) of the uploaded SC data and 87.08% ($SD = 9.58\%$) of the uploaded PPG signals marked as having good quality. Some missing data might have been collected but later treated as missing due to low quality indicators, suggesting that no uploaded segments met the quality threshold. It is more likely that the missing data resulted from participants not wearing the wearable. A thorough analysis of wearable use is beyond the scope of this paper, as it would require a detailed investigation of the complete Chill + wearable dataset, which includes not only the 10-min EPA segments but also the remaining EPA data recorded during the whole 16-h period. As the ratio of beeps without EPA data increased throughout the study, this could indeed indicate a decreasing compliance with the wearable device. Following this, subjective feedback also indicated challenges and discomfort related to the wearable device. This indicates the need for a more comfortable and aesthetically pleasing design. Using consumer-grade devices, like Garmin or Fitbit, could offer a solution for future studies (Henriksen et al., 2018). However, it is important to note that these devices currently provide no access to raw data, and most do not offer SC measures.

In addition to the primary real-time JITAI triggering algorithm, our pipeline further included a threshold adjustment algorithm. This second algorithm was designed to update individual thresholds each day, based on the cumulative number of interventions that were triggered throughout the day. The goal was to maintain a consistent triggering rate of three interventions per day. Retrospective simulations of the JITAI algorithm without these daily threshold adjustments showed that it did not influence the number of triggered interventions. This outcome might be due to low temporal drift in the current dataset, which is particularly notable given the study's long duration spanning several months. The stress score ICCs indicated moderate stability, with a participant-level ICC of 0.36 and a week-level ICC of 0.15, suggesting that stress levels were more consistent within individuals across weeks than between different weeks. Similar to the ICCs of the EMA features, this indicates a high degree of similarity between weeks for a given participant. Whether the same lack of temporal drift would be observed in other studies remains to be investigated. Despite the potentially unnecessary complication of our setup, we did manage to trigger interventions in approximately 40% of the completed beeps, as was intended. Unexpectedly, however, this resulted in only one to two triggered interventions per day. This outcome was lower than intended, likely due to the anticipated compliance rate of 70% not being met.

The average compliance rates for both EMA and EMI were around 43%. Recent meta-analyses have estimated higher EMA compliance rates: 78.7% on average across 65 studies (Vachon et al., 2019), 81.9% on average across 105 studies (Williams et al., 2021), and 79.2% on average across 477 studies (Wrzus & Neubauer, 2023). For EMI studies, however, adherence is reported far less consistently. A systematic review by Marciniak et al. (2020) reported that 19 out of 26 EMI studies did not report compliance. Among those that did, compliance varied from 33.8% to 93.3%. Notably, none of the EMI studies conducted in healthy populations reported compliance (Marciniak et al., 2020). Similarly, a more recent review by Dao et al. (2021) found that only four out of 17 EMI studies reported adherence. Among the included mental health-focused studies, only Hanssen et al. (2019) reported adherence of 49%, although in a schizophrenia sample and based on a different metric (responding to ≥ 1 prompt/day), limiting comparability to our results. Comparable adherence rates have also been reported in recent JITAI studies, with response rates ranging from 21 to 46% across diverse populations and intervention targets (e.g., Garland et al., 2023; Maria et al., 2021; Vinci et al., 2025). Although these studies all used different metrics, they do suggest that moderate adherence is not uncommon in JITAI implementations.

Compared to these numbers, our EMA compliance rate was considerably low. Also, we observed a notable decrease in compliance over time, indicating decreased study engagement or motivation. Although previous research has demonstrated a decrease in compliance as the study progresses (Rintala et al., 2019), it is important to note that our EMA protocol spanned 24 days (4 weeks × 6 days) and was therefore quite extensive compared to studies included in the mentioned meta-analyses. These studies reported an average duration of 11.2 (*SD* = 19.0, range: 1–150) assessment days (Vachon et al., 2019), a median of 7 (range: 1–182) days (Williams et al., 2021), and an average duration of 12.4 (*SD* = 16.38, range: 2–180) days (Wrzus & Neubauer, 2023). Moreover, it is important to acknowledge that the current EMA protocol was embedded within the larger DynaM-INT study (Bögemann et al., 2023), which, in addition to the EMA-protocol, required participants to partake in other procedures, including biweekly repeated questionnaires, imposing another layer of burden.

Several other factors could have contributed to our lower compliance rates. First, the relationship between the number of days and the frequency of assessments per day could have influenced our compliance rates negatively (Wrzus & Neubauer, 2023). Studies often balance participant burden by adjusting the number of assessments and the duration of the study. For instance, studies with frequent assessments (e.g., six to ten times per day) are typically shorter, while longer studies usually contain fewer assessments per day (e.g., one or two). Second, our "booster" design, with intervention weeks scheduled once per month to lower the participant burden, did likely prevent the habit formation of daily application use. While this validated design was justified to reduce participant burden by interspersing EMI weeks with quiet periods, it might have contributed to the lower compliance rates instead (Marciniak et al., 2024). Participant feedback also suggests that technical errors in the arrival of notifications may have contributed to lower compliance. Third, the use of study phones might have impacted compliance, as participants may be less accustomed to carrying an additional device. Lastly, it is crucial to consider participant receptiveness to interventions, particularly in JITAI designs. A key question is whether moments of high stress are indeed the times when participants are most receptive to interventions. It is possible that individuals may perceive these JITAIs as intrusive or overwhelming rather than supportive, which could have further influenced compliance rates in our study. Taken together, these insights suggest that our EMA protocol may have been too intensive, contributing to lower compliance rates. Future JITAI studies should carefully balance the frequency and duration of assessments, ensure reliable technology, consider the practicalities of device usage, and optimize the timing of interventions to align with participant receptiveness to improve compliance rates.

Despite our demanding study protocol, the attrition rate, defined as the proportion of participants who discontinued the study during the EMI phase, was moderate, at 24%. This number is highly consistent with a recent meta-analysis investigating attrition rates in EMI-based randomized controlled trials for mental health problems (Linardon & Fuller-Tyszkiewicz, 2020). The estimated attrition rate was 24.1% (95% CI [19.3, 29.6]), on average across 83 studies. It is, however, essential to acknowledge that we did not consider the dropout rates before the start of the calibration week. Although these participants (*N* = 12) were not included in the current feasibility analyses, their early discontinuation provides critical insights into our study design. The predominant reasons for participant dropout revolved around the demanding time commitments of the study, scheduling conflicts, personal circumstances, and unforeseen life events such as moving to another city. These factors underscore the necessity of designing studies that achieve scientific objectives while fitting harmoniously into participants' daily lives.

Within this context, it is also crucial to consider the growing societal trend toward minimizing cell phone usage, as rightfully highlighted by one of our participants. As awareness of drawbacks of excessive screen time and smartphone addiction grows (Lanette et al., 2018), many individuals are actively seeking ways to decrease their dependence on digital devices. However, research on the mental health effects of smartphone use remains inconclusive, with both positive and negative effects being reported (Bayer et al., 2023; große Deters & Schoedel, 2024; Marciano et al., 2022; Roos & Wrzus, 2023). As recently suggested (Elmer et al., 2025), particularly vulnerable individuals, such as those in our sample, may be more susceptible to negative effects on mental health. This poses a challenge for the future of smartphone interventions, especially as the current implementation of our JITAI algorithm fundamentally relies on regular interactions with a mobile phone. To address this challenge, it is essential to develop minimally invasive triggering algorithms, preferably using passive data collection methods, such as smartphone-based digital phenotyping (Onnela & Rauch, 2016). Such digital phenotyping refers to data collection through smartphone sensors and logs without requiring active input from the user and may include global positioning system (GPS) data (to capture physical activity), call logs (to capture social interactions), keyboard inputs (to capture typing patterns), and app usage (to capture activities such as social media usage), among others (Choi et al., 2024). Future studies should further explore ways to integrate JITAI interventions in a way that minimizes participant burden while maintaining engagement, for example by leveraging more passive sensing and adaptive intervention strategies that align with individual usage patterns.

The functionality of both the ReApp and Imager interventions scored well, indicating that the technical aspects of both apps were well received. Subjective ratings were slightly lower, highlighting another point of improvement.

Qualitative feedback provided additional insights. Here, participants reported concerns such as irritation from frequent notifications and technical issues like app crashes, which sometimes resulted in increased rather than decreased stress levels. The JITAI received mixed responses, though most participants indicated that it helped them use the targeted skills during stressful periods. The efficacy of the JITAI will be assessed in a future study to determine whether our design enhances both target engagement (positive cognitive reappraisal for ReApp and reward sensitivity for Imager) and resilience, which is defined as "*the maintenance or quick recovery of mental health during and after times of adversity*" (Kalisch et al., 2017).

Limitations

This study has several limitations that highlight important directions for future research. First, physiological responses were only considered within the fixed 10-min windows preceding EMA, limiting stress detection throughout the entire day. Additionally, the developed pipeline relies on completion of the EMA for data uploads, further restricting continuous monitoring. Future studies could explore methods for passive, EMA-independent data collection to improve the real-time triggering pipeline. Such methods may include continuous monitoring of smartphone-derived features (e.g., activity patterns, keyboard interaction, or speech analysis), referred to as digital phenotyping (Insel, 2017). These features can be collected without active input from participants and may help detect stress unobtrusively throughout the day.

Second, while the pipeline successfully applied the predefined decision rules, its ability to detect real-world stress was not independently validated. Similarly, the stress detection threshold was pragmatically set to ensure a sufficient number of interventions, rather than identifying the most independently validated stressful moments. Moreover, the composite stress score was based on psychological (affective) and physiological indicators of the stress response. While we explored its association with subjective stress ratings (e.g., "*I feel stressed*") using a linear mixed-effects model with random intercepts for week nested within participant, this was not the primary focus of the current manuscript. Results showed a modest but significant within-subject association ($\beta = 0.19$, $p < .001$), suggesting partial overlap. Future studies should validate the composite stress score against established stress paradigms to assess its predictive validity and clarify its theoretical distinction from related constructs such as perceived distress.

Third, interventions were randomly assigned rather than tailored to individual baseline characteristics or momentary contextual factors. Developing adaptive intervention strategies, where content and timing are personalized based on individual stress responses and situational context, could enhance effectiveness. Together, these refinements will help advance the JITAI algorithm toward more personalized mental health support.

Conclusion

In conclusion, our study demonstrates preliminary feasibility of implementing a JITAI algorithm for providing support during moments of heightened stress. By validating the technical implementation of the developed JITAI, along with participant adherence and experience, our study significantly contributes to the field of personalized stress-related mental health interventions. Ultimately, our work lays a critical foundation for future advancements in mental health care aimed to deliver timely and personalized support to those in need.

Abbreviations DynaM-INT: DynaMORE interventional study; DynaMORE: Dynamic modelling of resilience; EMA: Ecological momentary assessment; EMI: Ecological momentary intervention; EPA: Ecological physiological assessment; HR: Heart rate; ICC: Intraclass correlation coefficient; JITAI: Just-in-time adaptive intervention; LMM: Linear mixed model; NA: Negative affect; PA: Positive affect; PPG: Photoplethysmography; RADAR: Remote assessment of disease and relapse; SC: Skin conductance; SQI: Signal quality indicator; uMARS: User version of the Mobile Application Rating Scale

Supplementary Information The online version contains supplementary material available at https://doi.org/10.3758/s13428-025-02870-5.

Acknowledgements We thank S. Stöber and N. Donner for their help with project administration. For their help with study conduct, we thank the following people: At Charité: S. Blum, L. Do, E. Hodapp, A. Riepenhausen, C. Rohr, E. Rossi, C. Sachs, C. Wackerhagen; in Mainz: H. Fiehn, L. Knirsch, D. Laurila-Epe, A. Peschel, J. Piloth, C. Schultheis, C. Walter, M. Weber, K.S.L. Yuen; in Nijmegen: N. Aslan, E.J.C. Hermsen, J. Posthuma, M. Schepers; in Tel Aviv: S. Bar, S. Ben-Dor, S. Berman, I. David, D. Even-Or, O. Gafni, R. Horovich, L. Lagziel; in Warsaw: S. Matys, N. Robak, J. Szurnicka, A. Uściłko, M. Wasylkowska.

Authors' contributions S.A. Bögemann: conceptualization, software, formal analysis, investigation, data curation, writing –original draft, writing—review & editing, visualization.

F. Krause: conceptualization, methodology, writing—review & editing, supervision.

A. van Kraaij: conceptualization, data curation, writing—review & editing.

M. Marciniak: conceptualization, writing—review & editing.

J.M.C. van Leeuwen: conceptualization, software.

J. Weermeijer: conceptualization, software.

J. Mituniewicz: investigation, writing—review & editing.

L.M. Puhlmann: Investigation, methodology, writing—review & editing.

M. Zerban: investigation, writing—review & editing.

Z.C. Reppmann: investigation, writing—review & editing.

D. Kobylińska: writing—review & editing, funding acquisition.

K.S.L. Yuen: writing—review & editing, funding acquisition.

B. Kleim: writing—review & editing, funding acquisition.

H. Walter: writing—review & editing, funding acquisition.

I. Myin-Germeys: conceptualization, writing—review & editing, funding acquisition.

R. Kalisch: conceptualization, writing—review & editing, funding acquisition.

I.M. Veer: conceptualization, formal analysis, writing—review & editing, funding acquisition.

K. Roelofs: conceptualization, writing—review & editing, funding acquisition, Supervision.

E.J. Hermans: conceptualization, methodology, software, writing—review & editing, funding acquisition, supervision.

Funding This project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement numbers 777084 (DynaMORE project) and 101016127 (RESPOND project), Deutsche Forschungsgemeinschaft (DFG Grant CRC 1193, subprojects B01, C01, C04, Z03), the German Federal Ministry for Education and Research (BMBF) as part of the Network for University Medicine under Grant number 01KX2021 (CEOsys and EViPan projects), and the State of Rhineland-Palatinate, Germany (MARP program, DRZ program, Leibniz Institute for Resilience Research). OnePlanet Research Center acknowledges financial support from the Province of Gelderland. The work of LP was financially supported by the Technische Universität Dresden (Maria Reiche Postdoctoral Fellowship).

Declarations

Ethics approval The study was conducted in accordance with the Declaration of Helsinki and approved by the local ethics committees of all participating sites: the Ethics Committee of Charité – Universitätsmedizin Berlin, Germany; the Medical Ethical Committee of Radboud University Medical Center (METC Oost-Nederland), Nijmegen, the Netherlands; the Ethics Committee of the Tel Aviv University, Tel Aviv, Israel, and the Helsinki Committee of Tel Aviv Souraski Medical Center; the Ethics Committee of the State Medical Board of Rhineland-Palatinate, Mainz, Germany; the Ethics Committee for Scientific Research at the Faculty of Psychology, University of Warsaw (Komisja Etyki Badań Naukowych Wydziału Psychologii Uniwersytetu Warszawskiego), Warsaw, Poland.

Consent to participate Written informed consent was obtained from all individual participants.

Consent for publication Participants consented to the publication of their anonymized data for research purposes.

Conflict of interest The authors declare that they have no competing interests. RK has received advisory honoraria from JoyVentures, Herzlia, Israel.

Open Practices Statement The pseudonymized, preprocessed data and analysis scripts are available at OSF (https://osf.io/bnse8/), as well as additional supplemental information (https://osf.io/wvpg3). The study was not preregistered in an independent registry, although the analysis plan was prespecified and published in the protocol paper (Bögemann et al., 2023) prior to data analysis.

References

Balaskas, A., Schueller, S. M., Cox, A. L., & Dohertyid, G. (2021). Ecological momentary interventions for mental health: A scoping review.https://doi.org/10.1371/journal.pone.0248152

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1). https://doi.org/10.18637/jss.v067.i01

Bayer, J. B., Reinecke, L., & Abeele, M. M. P. V. (2023). Psychological perspectives on mobile media: A flyover review. *Mobile Media & Communication, 11*(1), 107–117. https://doi.org/10.1177/20501579221134369

Baxter, A. J., Vos, T., Scott, K. M., Ferrari, A. J., & Whiteford, H. A. (2014). The global burden of anxiety disorders in 2010. *Psychological Medicine, 44*(11), 2363–2374. https://doi.org/10.1017/S0033291713003243

Bögemann, S. A., Riepenhausen, A., Puhlmann, L. M. C., Bar, S., Hermsen, E. J. C., Mituniewicz, J., Reppmann, Z. C., Uściłko, A., van Leeuwen, J. M. C., Wackerhagen, C., Yuen, K. S. L., Zerban, M., Weermeijer, J., Marciniak, M. A., Mor, N., van Kraaij, A., Köber, G., Pooseh, S., Koval, P., … Walter, H. (2023). Investigating two mobile just-in-time adaptive interventions to foster psychological resilience: Research protocol of the DynaM-INT study. *BMC Psychology*. https://doi.org/10.1186/s40359-023-01249-5

Bonanno, G. A., Westphal, M., & Mancini, A. D. (2011). Resilience to loss and potential trauma. *Annual Review of Clinical Psychology, 7*, 511–535. https://doi.org/10.1146/annurev-clinpsy-032210-104526

Choi, A., Ooi, A., & Lottridge, D. (2024). Digital phenotyping for stress, anxiety, and mild depression: Systematic literature review. *JMIR Mhealth Uhealth, 12*, Article e40689. https://doi.org/10.2196/40689

Coyne, S. M., Stockdale, L., & Summers, K. (2019). Problematic cell phone use, depression, anxiety, and self-regulation: Evidence from a three year longitudinal study from adolescence to emerging adulthood. *Computers in Human Behavior*. https://doi.org/10.1016/j.chb.2019.02.014

Dao, K. P., De Cocker, K., Tong, H. L., Kocaballi, A. B., Chow, C., & Laranjo, L. (2021). Smartphone-delivered ecological momentary interventions based on ecological momentary assessments to promote health behaviors: Systematic review and adapted checklist for reporting ecological momentary assessment and intervention studies. *Journal of Medical Internet Research, mHealth and uHealth, 9*(11), Article e22890. https://doi.org/10.2196/22890

Dutcher, J. M. (2022). Brain reward circuits promote stress resilience and health: Implications for reward-based interventions. *Current Directions in Psychological Science, 32*(1), 65–72. https://doi.org/10.1177/09637214221121770

Dutcher, J. M., Creswell, J. D., Elsevier Ltd. (2018). The role of brain reward pathways in stress resilience and health. *Neuroscience and Biobehavioral Reviews, 95*, 559–567. https://doi.org/10.1016/j.neubiorev.2018.10.014

Elmer, T., Fernández, A., Stadel, M., Kas, M. J. H., & Langener, A. M. (2025). Bidirectional associations between smartphone usage and momentary well-being in young adults: Tackling methodological challenges by combining experience sampling methods with passive smartphone data. *Emotion*. https://doi.org/10.1037/emo0001485

Garland, E. L., Gullapalli, B. T., Prince, K. C., Hanley, A. W., Sanyer, M., Tuomenoksa, M., & Rahman, T. (2023). Zoom-based mindfulness-oriented recovery enhancement plus just-in-time mindfulness practice triggered by wearable sensors for opioid craving and chronic pain. *Mindfulness, 14*(6), 1329–1345. https://doi.org/10.1007/s12671-023-02137-0

Goldberg, D. P., Gater, R., Sartorius, N., Ustun, T. B., Piccinelli, M., Gureje, O., & Rutter, C. (1997). The validity of two versions of the GHQ in the WHO study of mental illness in general health care. *Psychological Medicine, 27*(1), 191–197. https://doi.org/10.1017/S0033291796004242

große Deters, F., & Schoedel, R. (2024). Keep on scrolling? Using intensive longitudinal smartphone sensing data to assess how

everyday smartphone usage behaviors are related to well-being. *Computers in Human Behavior, 150*, Article 107977. https://doi.org/10.1016/J.CHB.2023.107977
Hanssen, E., Balvert, S., Oorschot, M., Borkelmans, K., Van Os, J., Delespaul, P., & Fett, A. (2019). An ecological momentary intervention incorporating personalised feedback to improve symptoms and social functioning in schizophrenia spectrum disorders. *Psychiatry Research, 284*, 112695. https://doi.org/10.1016/j.psychres.2019.112695
Healey, J., & Picard, R. (2000). SmartCar: detecting driver stress. Proceedings 15th International Conference on Pattern Recognition. ICPR-2000, 218–221. https://doi.org/10.1109/ICPR.2000.902898
Hedges, L. V., Hedberg, E. C. C., & Kuyper, A. M. (2012). The variance of intraclass correlations in three- and four-level models. *Educational and Psychological Measurement, 72*(6), 893–909. https://doi.org/10.1177/0013164412445193
Henriksen, A., Haugen Mikalsen, M., Woldaregay, A. Z., Muzny, M., Hartvigsen, G., Hopstock, L. A., & Grimsgaard, S. (2018). Using fitness trackers and smartwatches to measure physical activity in research: Analysis of consumer wrist-worn wearables. *Journal of Medical Internet Research, 20*(3), Article e110. https://doi.org/10.2196/jmir.9157
Heron, K. E., & Smyth, J. M. (2010). Ecological momentary interventions: Incorporating mobile technology into psychosocial and health behaviour treatments. *British Journal of Health Psychology, 15*(1), 1–39. https://doi.org/10.1348/135910709X466063
Insel, T. R., American Medical Association. (2017). Digital phenotyping: Technology for a new science of behavior. *JAMA - Journal of the American Medical Association, 318*(13), 1215–1216. https://doi.org/10.1001/jama.2017.11295
Ivanova, A., Gorbaniuk, O., Błachnio, A., Przepiórka, A., Mraka, N., Polishchuk, V., & Gorbaniuk, J. (2020). Mobile phone addiction, phubbing, and depression among men and women: A moderated mediation analysis. *Psychiatric Quarterly*. https://doi.org/10.1007/s11126-020-09723-8
James, S. L., Abate, D., Abate, K. H., Abay, S. M., Abbafati, C., Abbasi, N., Abbastabar, H., Abd-Allah, F., Abdela, J., Abdelalim, A., Abdollahpour, I., Abdulkader, R. S., Abebe, Z., Abera, S. F., Abil, O. Z., Abraha, H. N., Abu-Raddad, L. J., Abu-Rmeileh, N. M. E., Accrombessi, M. M. K., … Murray, C. J. L. (2018). Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: A systematic analysis for the Global Burden of Disease Study 2017. *The Lancet, 392*(10159), 1789. https://doi.org/10.1016/S0140-6736(18)32279-7
Jorm, A. F., Patten, S. B., Brugha, T. S., & Mojtabai, R. (2017). Has increased provision of treatment reduced the prevalence of common mental disorders? Review of the evidence from four countries. *World Psychiatry, 16*(1), 90–99. https://doi.org/10.1002/wps.20388
Kalisch, R., Baker, D., Basten, U., Boks, M., Bonanno, G., Brummelman, E., Chmitorz, A., Fernàndez, G., Fiebach, C., Galatzer-Levy, I., Geuze, E., Groppa, S., Helmreich, I., Hendler, T., Hermans, E., Jovanovic, T., Kubiak, T., Lieb, K., Lutz, B., … Kleim, B. (2017). The resilience framework as a strategy to combat stress-related disorders. *Nature Human Behaviour, 1*(11), 784–790. https://doi.org/10.1038/s41562-017-0200-8
Kalisch, R., Müller, M. B., & Tüscher, O. (2015). A conceptual framework for the neurobiological study of resilience. *Behavioral and Brain Sciences, 38*, Article e92. https://doi.org/10.1017/S0140525X1400082X
Kalisch, R., Russo, S. J., & Müller, M. B. (2024). Neurobiology and systems biology of stress resilience. *Physiological Reviews, 104*(3), 1205–1263. https://doi.org/10.1152/physrev.00042.2023
Kocielnik, R., Sidorova, N., Maggi, F. M., Ouwerkerk, M., & Westerink, J. H. D. M. (2013). Smart technologies for long-term stress monitoring at work. Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems, 53–58. https://doi.org/10.1109/CBMS.2013.6627764
Lanette, S., Chua, P. K., Hayes, G., & Mazmanian, M. (2018). How Much is "Too Much"? *Proceedings of the ACM on Human-Computer Interaction, 2*(CSCW), 1–22. https://doi.org/10.1145/3274370
Lazarus, R. S., & Folkman, S. (1987). Transactional theory and research on emotions and coping. *European Journal of Personality, 1*(3), 141–169. https://doi.org/10.1002/PER.2410010304
Linardon, J., & Fuller-Tyszkiewicz, M. (2020). Attrition and adherence in smartphone-delivered interventions for mental health problems: A systematic and meta-analytic review. *Journal of Consulting and Clinical Psychology, 88*(1), 1–13. https://doi.org/10.1037/ccp0000459
Liu, Q., He, H., Yang, J., Feng, X., Zhao, F., & Lyu, J. (2020). Changes in the global burden of depression from 1990 to 2017: Findings from the Global Burden of Disease study. *Journal of Psychiatric Research, 126*, 134–140. https://doi.org/10.1016/j.jpsychires.2019.08.002
Lupien, S. J., Leclaire, S., Majeur, D., Raymond, C., Jean Baptiste, F., & Giguère, C.-E. (2022). Doctor, I am so stressed out!' A descriptive study of biological, psychological, and socioemotional markers of stress in individuals who self-identify as being 'very stressed out' or 'zen. *Neurobiology of Stress, 18*, 100454. https://doi.org/10.1016/j.ynstr.2022.100454
Marciano, L., Driver, C. C., Schulz, P. J., & Camerini, A. L. (2022). Dynamics of adolescents' smartphone use and well-being are positive but ephemeral. *Scientific Reports, 12*(1), 1–15. https://doi.org/10.1038/s41598-022-05291-y
Marciniak, M. A., Shanahan, L., Binder, H., Kalisch, R., & Kleim, B. (2023a). Positive prospective mental imagery characteristics in young adults and their associations with depressive symptoms. *Cognitive Therapy and Research, 47*(4), 695–706. https://doi.org/10.1007/s10608-023-10378-5
Marciniak, M. A., Shanahan, L., Myin-Germeys, I., Milos Veer, I., Yuen, K. S. L., Binder, H., Walter, H., Hermans, E. J., Kobylinska, D., Tuescher, O., Kalisch, R., & Kleim, B. (2023b). *ReApp-an mHealth app increasing reappraisal: results from two randomized controlled trials*. https://doi.org/10.31234/osf.io/u4f5e
Marciniak, M. A., Shanahan, L., Rohde, J., Schulz, A., Wackerhagen, C., Kobylińska, D., Tuescher, O., Binder, H., Walter, H., Kalisch, R., & Kleim, B. (2020). Standalone smartphone cognitive behavioral therapy–based ecological momentary interventions to increase mental health: Narrative review. *JMIR Mhealth Uhealth, 8*(11), Article e19836. https://doi.org/10.2196/19836
Marciniak, M. A., Shanahan, L., Yuen, K. S. L., Milos Veer, I., Walter, H., Tuescher, O., Kobylińska, D., Kalisch, R., Hermans, E., Binder, H., & Kleim, B. (2024). *Burst versus continuous delivery design in digital mental health interventions: evidence from a randomized clinical trial*. https://doi.org/10.1177/20552076241249267
Marciniak, M., Shanahan, L., Myin-Germeys, I., Veer, I., Yuen, K. S., Binder, H., Walter, H., Hermans, E., Kalisch, R., & Kleim, B. (2023c). Imager—A mobile health mental imagery-based ecological momentary intervention targeting reward sensitivity: A randomized controlled trial. *Applied Psychology: Health and Well-Being*. https://doi.org/10.1111/APHW.12505
Maria, D. S., Padhye, N., Businelle, M., Yang, Y., Jones, J., Sims, A., & Lightfoot, M. (2021). Efficacy of a just-in-time adaptive intervention to promote HIV risk reduction behaviors among young adults experiencing homelessness: Pilot randomized controlled trial. *Journal of Medical Internet Research, 23*(7), Article e26704. https://doi.org/10.2196/26704
Myin-Germeys, I., & Kuppens, P. (Eds.). (2022). The open handbook of experience sampling methodology: A step-by-step guide to designing, conducting, and analyzing ESM studies (2nd ed.). Center for Research on Experience Sampling and Ambulatory Methods Leuven.

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*. https://doi.org/10.1007/s12160-016-9830-8

Nielson, D. M., Keren, H., O'Callaghan, G., Jackson, S. M., Douka, I., Vidal-Ribas, P., Pornpattananangkul, N., Camp, C. C., Gorham, L. S., Wei, C., Kirwan, S., Zheng, C. Y., & Stringaris, A. (2021). Great expectations: A critical review of and suggestions for the study of reward processing as a cause and predictor of depression. *Biological Psychiatry, 89*(2), 134–143. https://doi.org/10.1016/j.biopsych.2020.06.012

Onnela, J.-P., & Rauch, S. L. (2016). Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. *Neuropsychopharmacology, 41*, 1691–1696. https://doi.org/10.1038/npp.2016.7

R Core Team. (2021). R: A language and environment for statistical computing. In 2021. https://www.r-project.org/

Ranjan, Y., Rashid, Z., Stewart, C., Conde, P., Begale, M., Verbeeck, D., Boettcher, S., Dobson, R., & Folarin, A. (2019). RADAR-base: Open source mobile health platform for collecting, monitoring, and analyzing data using sensors, wearables, and mobile devices. *JMIR Mhealth Uhealth, 7*(8), Article e11734. https://doi.org/10.2196/11734

Reavley, N., & Jorm, A. F. (2010). Prevention and early intervention to improve mental health in higher education students: A review. *Early Intervention in Psychiatry, 4*(2), 132–142. https://doi.org/10.1111/j.1751-7893.2010.00167.x

Riley, W. T., Rivera, D. E., Atienza, A. A., Nilsen, W., Allison, S. M., & Mermelstein, R. (2011). Health behavior models in the age of mobile interventions: Are our theories up to the task? *Translational Behavioral Medicine, 1*, 53–71. https://doi.org/10.1007/s13142-011-0021-7

Rintala, A., Wampers, M., Myin-Germeys, I., & Viechtbauer, W. (2019). Response compliance and predictors thereof in studies using the experience sampling method. *Psychological Assessment, 31*(2), 226–235. https://doi.org/10.1037/pas0000662

Roos, Y., & Wrzus, C. (2023). Is the smartphone friend and foe? Benefits and costs of self-reported smartphone use for important life domains in a representative German sample. *Current Psychology, 42*(28), 24717–24731. https://doi.org/10.1007/S12144-022-03593-y

Schlotz, W. (2018). Investigating associations between momentary stress and cortisol in daily life: What have we learned so far? https://doi.org/10.1016/j.psyneuen.2018.11.038

Selye, H., & Fortier, C. (1950). Adaptive reaction to stress. *Psychosomatic Medicine, 12*(3), 149–157. https://doi.org/10.1097/00006842-195005000-00003

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., & Dunbar, G. C. (1998). The Mini-International Neuropsychiatric Interview (M.I.N.I.): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *The Journal of Clinical Psychiatry, 59*(suppl 20), 11980.

Smets, E., De Raedt, W., & Van Hoof, C. (2019). Into the wild: The challenges of physiological stress detection in laboratory and ambulatory settings. *IEEE Journal of Biomedical and Health Informatics, 23*(2), 463–473. https://doi.org/10.1109/JBHI.2018.2883751

Smyth, J., & Heron, K. (2016). Is providing mobile interventions •just-in-time• helpful? An experimental proof of concept study of just-in-time intervention for stress management. 10.0/Linux-x86_64

Smyth, J. M., Zawadzki, M. J., Marcusson-Clavertz, D., Scott, S. B., Johnson, J. A., Kim, J., Toledo, M. J., Stawski, R. S., Sliwinski, M. J., & Almeida, D. M. (2023). Computing components of everyday stress responses: Exploring conceptual challenges and new opportunities. *Perspectives on Psychological Science, 18*(1), 110–124. https://doi.org/10.1177/17456916221082108

Stoyanov, S. R., Hides, L., Kavanagh, D. J., & Wilson, H. (2016). Development and validation of the user version of the Mobile Application Rating Scale (uMARS). *JMIR Mhealth Uhealth, 4*(2), Article e72. https://doi.org/10.2196/mhealth.5849

The MathWorks Inc. (2022). The MathWorks Inc. (https://www.mathworks.com).

Tutunji, R., Kogias, N., Kapteijns, B., Krentz, M., Krause, F., Vassena, E., & Hermans, E. J. (2023). Detecting prolonged stress in real life using wearable biosensors and ecological momentary assessments: Naturalistic experimental study. *Journal of Medical Internet Research*. https://doi.org/10.2196/39995

Vachon, H., Viechtbauer, W., Rintala, A., & Myin-Germeys, I. (2019). Compliance and retention with the experience sampling method over the continuum of severe mental disorders: Meta-analysis and recommendations. *Journal of Medical Internet Research, 21*(12), e14475. https://doi.org/10.2196/14475

Vinci, C., Sutton, S. K., Yang, M., Jones, S. R., Kumar, S., & Wetter, D. W. (2025). Proximal effects of a just-in-time adaptive intervention for smoking cessation with wearable sensors: Microrandomized trial. *PubMed, 13*, e55379. https://doi.org/10.2196/55379

Vos, T., Flaxman, A. D., Naghavi, M., Lozano, R., Michaud, C., Ezzati, M., Shibuya, K., Salomon, J. A., Abdalla, S., Aboyans, V., Abraham, J., Ackerman, I., Aggarwal, R., Ahn, S. Y., Ali, M. K., Almazroa, M. A., Alvarado, M., Anderson, H. R., Anderson, L. M., … Murray, C. J. L. (2012). Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990–2010: A systematic analysis for the Global Burden of Disease Study 2010. *The Lancet, 380*(9859), 2163–2196. https://doi.org/10.1016/S0140-6736(12)61729-2

Wang, L., & Miller, L. C. (2020). Just-in-the-moment adaptive interventions (JITAI): A meta-analytical review. *Health Communication, 35*, 1531–1544. https://doi.org/10.1080/10410236.2019.1652388

Wijsman, J. L. P. (2014). Sensing Stress: Stress detection from physiological variables in controlled and uncontrolled conditions. https://doi.org/10.3990/1.9789036537858

Williams, M. T., Lewthwaite, H., Fraysse, F., Gajewska, A., Ignatavicius, J., & Ferrar, K. (2021). Compliance with mobile ecological momentary assessment of self-reported health-related behaviors and psychological constructs in adults: Systematic review and meta-analysis. *Journal of Medical Internet Research, 23*(3), Article e17023. https://doi.org/10.2196/17023

Wrzus, C., & Neubauer, A. B. (2023). Ecological momentary assessment: A meta-analysis on designs, samples, and compliance across research fields. *Assessment, 30*(3), 825–846. https://doi.org/10.1177/10731911211067538

Xu, Z., & Smit, E. (2023). Using a complexity science approach to evaluate the effectiveness of just-in-time adaptive interventions: A meta-analysis. *Digital Health*. https://doi.org/10.1177/20552076231183543

Zhang, G., Yang, X., Tu, X., Ding, N., & Lau, J. T. F. (2019). Prospective relationships between mobile phone dependence and mental health status among Chinese undergraduate students with college adjustment as a mediator. https://doi.org/10.1016/j.jad.2019.09.047

Authors and Affiliations

S. A. Bögemann[1] · **F. Krause**[1] · **A. van Kraaij**[2] · **M. A. Marciniak**[3,4,5] · **J. M. van Leeuwen**[1] · **J. Weermeijer**[6] · **J. Mituniewicz**[7] · **L. M. C. Puhlmann**[8,9] · **M. Zerban**[10] · **Z. C. Reppmann**[11] · **D. Kobylińska**[7] · **K. S. L. Yuen**[8,10] · **B. Kleim**[3,4] · **H. Walter**[11,12] · **I. Myin-Germeys**[6] · **R. Kalisch**[8,10] · **I. M. Veer**[13] · **K. Roelofs**[14,15] · **E. J. Hermans**[1]

✉ S. A. Bögemann
sophie.bogemann@donders.ru.nl

1 Donders Institute for Brain, Cognition, and Behaviour, Radboud University Medical Center, Nijmegen, The Netherlands

2 OnePlanet Research Center, OnePlanet Research Center, Wageningen, The Netherlands

3 Division of Experimental Psychopathology and Psychotherapy, Department of Psychology, University of Zurich, Zurich, Switzerland

4 Department of Psychiatry, Psychotherapy and Psychosomatics, Psychiatric University Hospital (PUK), University of Zurich, Zurich, Switzerland

5 Department of Psychology, Education and Child Studies, Erasmus University Rotterdam, Rotterdam, The Netherlands

6 Center for Contextual Psychiatry, Department of Neurosciences, KU Leuven, Leuven, Belgium

7 Faculty of Psychology, University of Warsaw, Warsaw, Poland

8 Leibniz Institute for Resilience Research, Mainz, Germany

9 Clinical Psychology and Behavioral Neuroscience, Faculty of Psychology, Technische Universität Dresden, Dresden, Germany

10 Neuroimaging Center, Focus Program Translational Neuroscience (FTN), Johannes Gutenberg University Medical Center, Mainz, Germany

11 Department of Psychiatry and Psychotherapy, Charité Universitätsmedizin Berlin, Corporate Member of FU Berlin and Humboldt Universität Zu Berlin, Berlin, Germany

12 German Center of Mental Health (DZPG), Partner Site Berlin/Potsdam, 10177 Berlin, Germany

13 Department of Developmental Psychology, University of Amsterdam, Amsterdam, The Netherlands

14 Behavioural Science Institute, Radboud University, Nijmegen, The Netherlands

15 Center for Cognitive Neuroimaging, Donders Institute for Brain Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands