AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychological Assessment

© 2025 American Psychological Association
ISSN: 1040-3590

2025, Vol. 37, No. 10, 520–534
https://doi.org/10.1037/pas0001404

Evaluating the Feasibility of Episode-Contingent Experience Sampling Burst Designs

Yuanyuan Ji[1], Marieke J. Schreuder[1], Sigert Ariens[1], Egon Dejonckheere[1, 2], and Eva Ceulemans[1]
[1] Faculty of Psychology and Educational Sciences, KU Leuven
[2] Department of Medical and Clinical Psychology, Tilburg University

Experience sampling methodology has been widely used to study the links between emotion dynamics and mental health. Most studies rely on time-contingent sampling schemes, with momentary questionnaires being sent at prespecified times, usually multiple hours apart. The present study investigated the added value of episode-contingent (EC) burst designs, which may shed further light on emotion dynamics by triggering a series of closely timed beeps upon detecting emotional episodes. Using data from three EC studies ($N = 185$), we investigated the effectiveness and feasibility of two types of EC designs: signal-based (EC-Signal; bursts initiated when emotion ratings exceed thresholds) and event-based (EC-Event; bursts initiated by participants). Both EC designs are effective in capturing emotional episodes, but the quantity and intensity of the episodes varied depending on which of these two designs was used and the valence of the episodes. Regarding feasibility, compared to EC-Event, EC-Signal typically led to higher participant burden and lower compliance for both regular and follow-up beeps. Moreover, compliance tended to decrease over time and burden tended to increase for both EC-Signal and time-contingent, but not for EC-Event. In conclusion, both EC approaches showed feasibility but have distinct advantages and drawbacks. To select the best approach, researchers should carefully balance these trade-offs to maximize utility within specific research contexts.

Public Significance Statement
This study highlights the potential of episode-contingent burst experience sampling methodology designs for tracking fine-grained emotion dynamics in daily life. By evaluating the feasibility and effectiveness of two episode-contingent designs, we offer guidance on selecting the optimal approach to improve the study of emotion dynamics.

Keywords: episode-contingent design, measurement burst, ecological momentary assessment, emotion dynamics

Supplemental materials: https://doi.org/10.1037/pas0001404.supp

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Emotion research has extensively investigated the link between real-life emotion dynamics and mental health, with particular interest in how intensely people respond to emotional events and in how fast people recover from emotional episodes in daily life (e.g., Heininga et al., 2019; Houben et al., 2015; Koval et al., 2012; Kuppens et al., 2010; Sperry et al., 2020). Despite this long-standing interest, the findings on the relationship between emotion dynamics and mental health remain inconclusive (Dejonckheere et al., 2019; Schreuder et al., 2023), which may be partly due to limitations in study design (Revol et al., 2025).

A widely used approach for studying emotion dynamics in daily life is the experience sampling methodology (ESM; e.g., Myin-Germeys et al., 2009), where individuals receive notifications ("beeps") on mobile devices and self-report their momentary

Kristin Naragon-Gainey served as action editor.
Yuanyuan Ji https://orcid.org/0009-0003-2123-6311
Materials and code are available at https://osf.io/qxj7y/. The data are available from the corresponding author upon reasonable request.
The authors have no conflicts of interest to disclose. The research was supported by grants from the Research Council of KU Leuven (Grants C14/23/062 and iBOF/21/090) and by grant EOS40007528/G0I2422N from the research foundations Fonds Wetenschappelijk Onderzoek and Fonds De La Recherche Scientifique under the Excellence of Science program awarded to Eva Ceulemans.

Yuanyuan Ji played a lead role in formal analysis, visualization, and writing–original draft and an equal role in conceptualization, investigation, methodology, and writing–review and editing. Marieke J. Schreuder played an equal role in conceptualization, investigation, methodology, supervision, and writing–review and editing. Sigert Ariens played a supporting role in methodology and an equal role in conceptualization, data curation, investigation, and writing–review and editing. Egon Dejonckheere played a supporting role in methodology and an equal role in conceptualization, data curation, investigation, and writing–review and editing. Eva Ceulemans played a lead role in funding acquisition and project administration and an equal role in conceptualization, data curation, investigation, methodology, supervision, and writing–review and editing.

Correspondence concerning this article should be addressed to Yuanyuan Ji, Faculty of Psychology and Educational Sciences, KU Leuven, Tiensestraat 102, Box 3713, 3000 Leuven, Belgium. Email: yuanyuan.ji@kuleuven.be

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

emotions multiple times a day over several weeks. An important feature of ESM design that will determine how well specific emotion dynamics can be captured is the sampling schedule according to which the notifications are sent. Conventional ESM studies use time-contingent (TC) designs (Figure 1A), where questionnaires are sent at (random) predetermined moments that are hours apart (Vachon et al., 2019). While TC designs may allow assessing mean levels and variability of emotional experiences, they are ill-suited for capturing emotional episodes. First, TC designs may miss emotional episodes that occur in between the predetermined moments. To address this limitation, some studies have asked participants to retrospectively report the most significant event that occurred since the previous measurement (e.g., Bylsma et al., 2011; Heininga et al., 2019; Herres et al., 2018) or to create a drawing of how their emotions unfolded since the previous beep (Cloos et al., 2024). While such approaches help to capture emotional episodes, they are vulnerable to recall bias. Second, even when meaningful emotions are captured, TC designs do not allow to zoom in on the processes that follow emotional episodes (e.g., recovery, regulation), which may take place rapidly within only 1 to 2 hr (e.g., Kalokerinos et al., 2017; Verduyn & Lavrijsen, 2015)

An appealing alternative is the episode-contingent (EC) burst sampling scheme, which aims to collect data during intense emotional episodes rather than at random moments throughout the day (Dejonckheere & Mestdagh, 2021; Delecroix et al., 2024; Smyth et al., 2023). The measurement burst strategy, involving multiple closely scheduled (only minutes apart) measurements following the detection of an emotional episode (Revol et al., 2025; Schreuder et al., 2024), allows to better capture rapid emotion dynamics at the time scale they unfold (Revol et al., 2025). More broadly, EC burst designs can be adapted to capture episodic and fleeting incidents, making them particularly suitable in clinical contexts. For example, they have been used to investigate critical events such as self-harm behaviors (Kiekens et al., 2024), suicidal thoughts (Coppersmith et al., 2022), and binge eating (Biçaker et al., 2024), where detecting these events and understanding the regulatory processes that follow them are essential for both research and clinical intervention.

However, two crucial questions remain regarding the practical implementation of the EC burst design. The first concerns how to effectively detect emotional episodes (or, more generally, critical events). Two approaches have been used for this purpose: the signal-based approach (EC-Signal; Figure 1B), which initiates a measurement burst whenever a TC ("regular") beep yields an emotion score that exceeds a predefined threshold, and the event-based approach (EC-Event; Figure 1C), which relies on participants to trigger bursts themselves when experiencing emotional episodes. To date, no comparative study between both approaches has been conducted. Therefore, the effectiveness of EC-Signal and EC-Event designs in capturing relevant emotional episodes remains uncertain. The second question concerns the feasibility of EC burst designs, as frequent measurements may increase participant burden and lower compliance, potentially compromising data quantity and quality.

The Present Study

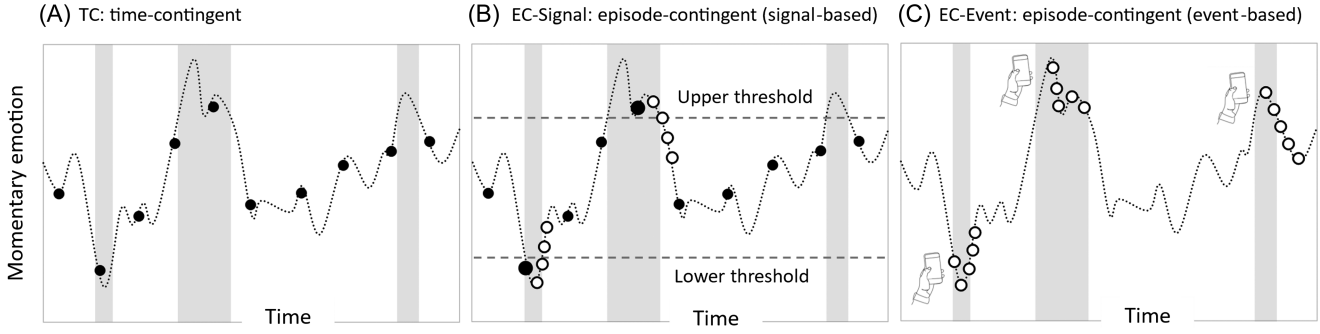
This article aimed to investigate (a) the effectiveness of EC burst designs in capturing emotional episodes and (b) the practical feasibility of EC burst designs. To achieve this, we report on three different EC burst studies and compare the resulting data. In the next section, we present our comparison criteria and expectations.

Effectiveness in Capturing Emotional Episodes

Quantity of Positive and Negative Episodes. The first criterion for evaluating the effectiveness of the two EC approaches is the number of detected emotional episodes. The EC-Event approach instructs participants to continuously monitor their emotions and start a burst whenever they feel that they are experiencing an emotional episode. Due to the continuous monitoring, EC-Event designs have the potential to capture many episodes. However, they also carry the risk that participants may forget or be unwilling to report their emotions consistently. In contrast, the EC-Signal approach relies on regular TC beeps to capture and follow up on moments of high emotional intensity that occur on scheduled assessments. Therefore, it does not solve the issue that episodes occurring in between scheduled assessments will be missed. The advantage is that EC-Signal designs

Figure 1

Schematic Illustration of Different Designs



Note. The dotted lines represent hypothetical emotion scores over time, with the gray shaded areas indicating emotional episodes. The black dots indicate time-contingent regular measurements, while the white dots indicate the bursts consisting of multiple follow-up measurements. The dotted lines represent the thresholds used to detect episodes in Episode-Contingent-Signal design.

reduce reliance on participants' engagement and ensure that participants are regularly reminded and prompted to self-report emotional experiences. Previous research that focused on reporting social interactions indicates that EC-Event strategies could lead to more frequent reporting of interactions compared to EC-Signal strategies (Himmelstein et al., 2019). We aimed to investigate whether this applies to the capturing of emotional episodes as well.

We will investigate not just the number of detected episodes but also the type of episodes (positive vs. negative) and hypothesize that the EC-Event design may capture an unequal number of positive versus negative episodes. Specifically, participants may experience episodes of different valences at varying frequencies. Additionally, response bias could occur, as participants may be less inclined to log their emotions immediately after experiencing stressful episodes, since these emotions are already burdensome (Dejonckheere & Erbaş, 2021; Eisele et al., 2021). On the other hand, it is also possible that during moments of positive emotions, participants may prefer to immerse themselves without interruption (i.e., emotional savoring; Bryant & Veroff, 2017), leading to underreporting of positive episodes. Such asymmetry may be less relevant for the EC-Signal design, as episodes should be detected based on predefined thresholds rather than participants' appraisal of events.

Intensity of Positive and Negative Episodes. Next, we consider the intensity of the emotional episodes detected by EC-Event and EC-Signal designs. In EC-Signal designs, intensity is directly linked to the threshold value used to trigger bursts, whereas in EC-Event designs, it may vary depending on participants' subjective interpretation of emotional episodes. For emotional episodes after social interactions, EC-Event might capture emotions with higher arousal, more positive valence, and lower variability compared to random TC prompts (Dawood et al., 2020). It is still unclear, however, how the intensity of positive and negative emotional episodes generally differs between designs, and we therefore aimed to investigate whether positive or negative emotional episodes are more extreme than the random emotion scores recorded using TC designs and how the two EC designs may differ.

Feasibility

Burden. Participant burden is a crucial criterion for evaluating the feasibility of the two EC designs, as it is closely associated with data quantity and quality (e.g., Dejonckheere et al., 2024; Fuller-Tyszkiewicz et al., 2013). In EC-Signal designs, follow-up beeps are added to regular beeps, increasing the sampling frequency, which may impose a greater burden on participants (Vachon et al., 2019). Moreover, the perceived burden may increase over time (Eisele et al., 2023). However, recent experimental studies suggest that increased sampling frequency does not significantly contribute to increased burden (Eisele et al., 2022; Hasselhorn et al., 2025), prompting us to explore how EC-Signal impacts perceived burden. In contrast, the EC-Event design relies on participants' self-initiation, with the sampling frequency and related burden depending on how often participants report their emotional episodes. Yet, the continuous self-monitoring required in EC-Event designs could lead to increased cognitive burden.

Compliance. Low compliance (i.e., failing to respond to notifications) is a common issue in ESM studies and worsens as studies progress (Rintala et al., 2019). While compliance relates to burden, in the sense that compliance lowers when burden increases, both are not synonymous: Participants may comply despite feeling burdened (Eisele et al., 2022), motivated by factors such as intrinsic interest in participation or external monetary compensation. Still, compliance could be lower in studies with higher sampling frequency (Dejonckheere et al., 2024). Therefore, EC-Signal designs—which combine regular beeps with bursts containing high-frequency follow-up beeps—may have a lower compliance compared to TC designs. In EC-Event designs, participants trigger assessments themselves, which might lead to higher compliance on follow-up beeps than in EC-Signal designs.

Method

Transparency and Openness

The code is available on the Open Science Framework at https://osf.io/qxj7y/ (Ji et al., 2025). The analysis plan was preregistered on the Open Science Framework at https://osf.io/fm2dy (Ji et al., 2024); however, some deviations were made to accommodate practical considerations and enhance coherence and clarity. All preregistered analyses and results, along with explanations for any deviations, are provided in the Supplemental Materials.

Data

We used data from three studies on emotion dynamics (see Table 1). Study 1 used the EC-Signal method for 2 weeks, after collecting TC data for 1 week (Revol et al., 2025; Schreuder et al., 2024); Study 2 used the EC-Event method; and Study 3 was a pilot study comparing both designs, where participants were included in a TC study beforehand (Cloos et al., 2023). All studies were approved by the Social and Societal Ethics Committee of KU Leuven.[1]

Participants

The total sample consisted of 185 participants across the three studies. Demographic information is provided in Table 1. Study 1 included 70 participants, with one participant excluded due to low compliance and the absence of burst data. The final sample ($N = 69$) included 52 females, and most of them were aged 21–25 years ($N = 57$), were from Belgium ($N = 65$), and had at least a bachelor's degree ($N = 55$). Study 2 originally had 97 participants, with two participants excluded for not reporting any emotional episode. The final sample ($N = 95$) included 84 females ($M_{age} = 22.26$, $SD_{age} = 7.28$). Study 3 had 20 participants, who were randomly assigned to one of the two groups (EC-Signal and EC-Event), with 10 participants in each group (EC-Signal: seven females, $M_{age} = 23.90$, $SD_{age} = 4.61$; EC-Event: five females, $M_{age} = 23.60$, $SD_{age} = 6.96$).

Protocols and Measures

Study 1. The study lasted 3 weeks and was divided into two phases: In the first week, the TC design was applied and in the second and third weeks the EC-Signal design. For the purpose of

[1] Study 1—the pros and cons of event-based sampling for intensive longitudinal studies (G-2021-2990-AMD); Study 2—assessing emotional recovery in daily life via a measurement burst design to predict individual differences in depressive symptoms (G-2022-4885-R2(MIN)); Study 3—the pros and cons of event-based sampling for intensive longitudinal studies (G-2021-2990).

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Table 1
Study Summary

Study	*N*	Data identifier	Age	Gender	Day	Design	Episode detection	Regular beep	Follow-up beep (time interval)
1 (Revol et al., 2025; Schreuder et al., 2024)	69	1-Time	18–20: 1 21–25: 57 26–30: 5 >31: 6	Female: 52 Male: 17	7 (Week 1)	TC		8 beeps/day [bipolar emotion; burden annoyance]	
		1-Signal			14 (Weeks 2, 3)	EC-Signal	Exceed 15th–85th personal quantile calculated from emotion scores in Data 1-Time	8 beeps/day [bipolar emotion; burden annoyance]	4 (every 10 min) [bipolar emotion; burden annoyance]
2	95	2-Event	Range: 18–60 *M*: 22.26 *SD*: 7.28	Female: 84 Male: 10 Other: 1	About a month	EC-Event	Self-initiate "uncommon" emotional event	2 reminders/day	7 (2, 5, 10, 15, 30, 60, 90) [bipolar emotion]
3	20	3-Time (Cloos et al., 2023)	See below	See below	14	TC		10 beeps/day [unipolar emotion]	
		3-Event (N = 10)	Range: 18–38 *M*: 23.6 *SD*: 6.96	Female: 5 Male: 4 Other: 1	14	EC-Event	Self-initiate emotional event	2 beeps/day [unipolar emotion; burden annoyance; burden effort; reason for missingness]	5 (every 10 min) [unipolar emotion; burden annoyance; reason for missingness]
		3-Signal (N = 10)	Range: 18–33 *M*: 23.9 *SD*: 4.61	Female: 7 Male: 3	14	EC-Signal	Exceed 75th overall quantile calculated from emotion scores in Data 3-Time	8 beeps/day [unipolar emotion; burden annoyance; burden effort; reason for missingness]	3 (every 10 min) [unipolar emotion; burden annoyance; reason for missingness]

Note. Measured variables for each study are provided in brackets. Two additional demographic variables were recorded in Study 1: nationality and education level. The nationality distribution was as follows: Belgian (65), Dutch (2), and other (2). For education level, participants reported the following: high school (13), associate's degree (1), academic bachelor (28), professional bachelor (19), and master's degree (8). TC = time-contingent; EC = episode-contingent.

clarity in the analysis and results, we labeled the associated data as 1-Time and 1-Signal, respectively. Participants received up to €60 depending on their compliance.

During the first week, participants received eight daily beeps on their smartphones from 10:00 a.m. to 8:30 p.m. at fixed 90-min intervals, prompting them to complete a questionnaire measuring (a) emotion using the bipolar item "How do you feel right now?" on a 0–100 visual analog scale (*very negative* to *very positive*) and (b) burden annoyance using "How disruptive was the beep?" on a visual slider scale from 0 (*not at all*) to 100 (*very intense*).

In the second and third weeks, in addition to the TC sampling scheme, four additional follow-up beeps were triggered at 10-min intervals when participants' emotion scores exceeded their personal thresholds. These personal thresholds were calculated based on the emotion distribution obtained during the first week. Specifically, if a participant's emotion score fell below the 15th percentile (i.e., negative episode) or above the 85th percentile (i.e., positive episode) of their reported emotion scores from the first week, four follow-up beeps were sent to prompt them to measure their emotion. At the final follow-up beep, participants were also asked to rate their burden annoyance.

Study 2. The study used the EC-Event design and was labeled as Data 2-Event in the analysis. Participants were instructed to log "uncommon emotional events" using the mobile app "m-Path" (https://www.m-path.io; Mestdagh et al., 2023) at any time. To ensure clarity, they were presented with multiple examples of suitable events. For instance, receiving important news, such as unexpectedly low or high exam grades, was considered an uncommon and emotional event that could be logged, while events like waking up to a rainy or sunny day, even if emotionally relevant, were deemed too ordinary to log. Participants were also informed about the criteria for earning a monetary bonus of €5 in addition to their fixed compensation of €20. Specifically, to receive the bonus, they needed to log at least 10 events and achieve a compliance rate of at least 71% by answering at least five out of the seven follow-up questions for each of these events.

At the start of the study, a debriefing session was scheduled for each participant to occur 30 days later. There were no regular beeps each day, while two daily reminders to log emotional events were sent for 30 consecutive days. Some participants continued to log events beyond the 30-day period, resulting in a study length that varied across individuals (M = 35.72 days, SD = 10.41 days, with a maximum of 91 days). However, given that no reminders were sent and a debriefing session had taken place after the 30-day period, only data collected within the initial 30 days were included for further analysis.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

When participants logged an event, they were immediately prompted to complete a questionnaire to identify the event's valence (positive or negative), provide a description of the event, report the time delay between event occurrence and event logging, and rate their current emotion using a bipolar item, "How do you feel right now?," on a visual analog scale ranging from −20 (*very negative*) to 20 (*very positive*). Subsequently, seven follow-up beeps were sent at 2, 5, 10, 15, 30, 60, and 90 min after the initial report to measure emotion. In each follow-up beep, participants were shown their one-line description of the event and their most recent emotion rating.

Study 3. All participants were recruited from a prior 14-day TC study on emotions (Cloos et al., 2023), labeled as Data 3-Time. During this TC study, participants received 10 beeps each day within a 12-hr self-selected timeframe. The 12 hr were divided into 10 blocks of 72 min each, with one beep sent randomly within each block. Emotions were measured at each beep using multiple items. For this analysis, we only focus on the unipolar items for positive ("How positive do you feel?") and negative ("How negative do you feel?") emotions, both rated from 0 (*neutral*) to 100 (*very positive/negative*).

The EC study lasted for 2 weeks, with participants divided into two groups. The corresponding subsets of data were labeled as Data 3-Signal and Data 3-Event. Participants in the EC-Signal group (Data 3-Signal) received eight regular beeps per day within the 12-hr time window. These beeps were spaced at approximately 90-min intervals, with a minimum of 80 min between two consecutive TC beeps. At each beep, participants completed the following measures: (a) emotion using separate unipolar items for positive ("How positive are you feeling?") and negative ("How negative are you feeling?") emotions, both rated from 0 (*neutral*) to 100 (*very intense*); (b) burden using two items, "Completing the items was difficult" (burden effort) and "The beep bothered me" (burden annoyance), both rated on a scale from 0 (*not at all*) to 100 (*very intense*); and (c) reasons for missing data. At each beep, participants were asked, "Have you, since the last time you filled out the questions, not responded to a beep?" If yes, they were prompted to select a reason from multiple options, including "I couldn't (respond on time)," "I didn't want to respond," "I was too stressed to respond," "I felt disturbed by the notification," or "other." Additionally, three follow-up beeps were triggered at 10-min intervals if a positive emotion score is higher than 72, indicating a positive episode, or a negative emotion score is higher than 34, indicating a negative episode. These limits were determined as the 85th percentile of positive and negative emotion scores from Data 3-Time across all 20 participants. For each follow-up beep, emotion and burden annoyance were measured using the same items.

Participants in the EC-Event group (Data 3-Event) received two regular beeps per day at random moments within the 12-hr window, reminding them to continue reporting events and to measure unipolar positive and negative emotion, burden, and reasons for missing data. Additionally, participants could self-initiate the reporting of "emotional events" at any time by selecting "start questionnaire" in the "m-Path" app. They were instructed that most of their daily experiences could qualify as emotional events, and examples provided included receiving a message that evoked happiness or irritation, eating a delicious or unpleasant meal, or having a good or bad social interaction. Each self-reported event was immediately followed by five follow-up beeps at 10-min intervals to measure emotions and burden annoyance.

At the end of the study, semistructured interviews were conducted, during which participants were asked about their experience with the study, the accuracy and attentiveness of their responses, their perception of the study's burden, and reasons for not responding or reporting events (see the question list in the Supplemental Materials). The interviews lasted between 12 and 30 min. Participants received €50 for the 2-week study period and €8 for the interview.

Statistical Analyses

All significance tests were two-sided with an α of .05. Data analysis was performed using R (R Core Team, 2023).

Preprocessing

First, we took initial steps such as relabeling variables, merging data sets, and checking for technical issues and participant response behaviors. Second, we identified the episode valence in each data set. In Data 1-Signal and 3-Signal, the valence was identified by comparing the emotion score when triggering a burst with the predetermined threshold. Specifically, for Data 1-Signal, a bipolar emotion score above the personal positive threshold ($M = 80.69$, $SD = 8.31$, ranging from 62 to 100) indicated a positive episode, while a score below the personal negative threshold ($M = 52.24$, $SD = 13.75$, ranging from 15.75 to 80) indicated a negative episode. For Data 3-Signal, a positive emotion score above the overall threshold of 72 indicated a positive episode, and a negative emotion score above the overall threshold of 34 indicated a negative episode. Episodes with ambiguous classification (e.g., both high negative and high positive emotion scores) were labeled as "undefined." In Data 2-Event, participants reported event valence when logging events. Data 3-Event, however, lacked necessary information on event valence. We determined it by evaluating the positive and negative emotion scores upon reporting events. We applied two rules for classification: (1) If one of the emotion scores (positive or negative) was zero and the other was nonzero, the event was classified according to the nonzero score's valence; (2) if both emotion scores were nonzero, we assessed their relative extremity by positioning them into the participant's positive and negative emotion distributions from Data 3-Time and recording the associated quantiles. We then identified the event's valence by checking which of both quantiles was higher. Third, since emotion experience was measured using different bipolar scales in Data 1-Signal and 2-Event, we rescaled the scores of Data 2-Event to a 0–100 scale by adding 20 to the original score and then multiplying by 2.5.

Quantitative Analyses

Table 2 gives an overview of the operationalization, analyses, and main findings for the four criteria: (a) quantity and (b) intensity of positive and negative episodes, (c) burden, and (d) compliance, as well as the predictors that were included. Regarding the operationalization of the criteria, quantity was defined as the daily numbers of (positive and negative) episodes for each participant. Episode intensity was represented by the emotion score at the time of triggering the burst, with unipolar and bipolar emotions analyzed

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Table 2
Summary of Research Questions, Quantitative Analyses, and Results

Label	Outcome	Predictor	Excluded data	Reason for exclusion	Result
			Effectiveness in capturing emotional episodes		
Quantity of positive and negative episodes	Number of episodes each day	Data, day, episode valence	1-Time, 3-Time	Not measured	EC-Signal > EC-Event EC-Signal: PosE < NegE EC-Event: PosE > NegE Decrease over time
Intensity of positive and negative episodes	First bipolar emotion each episode + random emotion in TC	Data, day, episode valence	3-Signal, 3-Event, 3-Time	Not measured	PosE: EC-Signal = EC-Event > TC NegE: EC-Event > EC-Signal > TC No day effect
	First positive emotion each episode + random emotion in TC	Data, day, episode valence	1-Signal, 2-Event, 1-Time	Not measured	PosE: EC-Signal > EC-Event > TC No day effect
	First negative emotion each episode + random emotion in TC	Data, day, episode valence	1-Signal, 2-Event, 1-Time	Not measured	NegE: EC-Signal > EC-Event > TC No day effect
			Feasibility		
Burden	Burden annoyance	Data, day (regular beeps)	2-Event, 3-Time	Not measured (2-Event, 3-Time)	EC-Signal = TC > EC-Event Increase over time (1-Time, 1-Signal, 3-Signal)
		Data, day (follow-up beeps)	1-Time, 2-Event, 3-Time	Not measured (2-Event, 3-Time) No follow-up beeps (1-Time)	EC-Signal > EC-Event Increase over time (1-Signal, 3-Signal)
	Burden effort	Data, day (regular beeps)	1-Time, 1-Signal, 2-Event, 3-Time	Not measured	EC-Signal = EC-Event Increase over time (3-Signal)
Compliance	Proportion of nonmissing reports each day	Data, day (regular beeps)	2-Event	No regular beeps (2-Event)	TC > EC-Signal (no significant differences between EC-Event and other two designs) Decrease over time (1-Time, 1-Signal, 3-Time, 3-Signal)
		Data, day (follow-up beeps)	1-Time, 3-Time	No follow-up beeps (1-Time, 3-Time)	EC-Event > EC-Signal Decrease over time (3-Signal)

Note. "=" indicates the difference was not significant. "PosE" represents positive episodes, and "NegE" indicates negative episodes. EC = episode-contingent; TC = time-contingent.

separately. High intensity reflected the extremity of the emotion, indicating how far it deviated from the random emotion states generated by TC sampling. Burden was investigated by analyzing burden annoyance and burden effort separately. Compliance was calculated as the proportion of nonmissing regular and follow-up reports per participant per day.

Regarding analyses, the key predictors of interest remained consistent across analyses.[2] The first predictor was the data set identifier (e.g., 1-Time, 1-Signal), which was used to compare different data sets. To examine design differences, planned contrasts were employed to assess whether the EC-Signal, EC-Event, and/or TC designs differed in terms of the outcome criteria. The second key predictor was day number, which allowed us to investigate changes over time within data sets.[3] The interaction between data set and day number was also added to explore potential differences in time effects across data sets. Additionally, for quantity and intensity of episodes, episode valence was added as a predictor to investigate differences between positive and negative episodes. The analyses of burden and compliance, on the other hand, were conducted separately for regular and follow-up beeps due to their inherent differences in frequency, time interval, and questionnaire length.

As all outcome measures were nested within individuals, multilevel modeling with a random intercept for each individual was applied. To account for the fact that one participant could be included in multiple data sets,[4] the random intercept for each individual was calculated by pooling data from all data sets they participated in and using a consistent identifier across data sets.

Qualitative Analyses

The reasons for missingness were descriptively reported. The semistructured interview was analyzed to identify themes related to participants' experiences with EC designs, including but not limited to burden and compliance. We first reviewed the interviews and recorded all relevant information (see the Supplemental Materials for documentation of intermediate steps). Through an iterative process, we then identified and refined themes based on recurring topics and recorded the number of participants who reported each theme.

[2] We examined the potential effects of demographic variables (i.e., gender and age) in the Supplemental Materials. Almost no effects were significant except for one exception observed in Study 3.

[3] Sensitivity analyses were conducted using the cumulative number of beeps sent by the end of each day as a proxy for both time (i.e., study duration) and sampling frequency. Our main findings remained robust.

[4] Participants from Study 1 provided data for both Data 1-Time and 1-Signal. Similarly, participants in the EC-Event group of Study 3 were included in both Data 3-Time and 3-Event, while those in the EC-Signal group were included in both Data 3-Time and 3-Signal.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Results

Preprocessing

To enhance transparency and reproducibility, we created preprocessing reports for all data sets, following the guidelines suggested by Revol et al. (2025). These reports, available on the Open Science Framework at https://osf.io/qxj7y (Ji et al., 2025), detail the initial steps we took, such as checking for technical issues and response behaviors. During preprocessing, we identified technical issues specific to follow-up beeps, such as a few duplicated beeps being sent to participants (<2.51% of beeps) and some scheduled beeps that were not sent (<0.01% of beeps). Additionally, some response behaviors warrant attention, such as response delays of more than 30 min (<1.12% of beeps). The summary of technical issues and response behaviors is provided as Supplemental Materials. Overall, we consider these issues to be minor and acceptable for all data sets.

Effectiveness in Detecting Episodes

Quantity of Positive and Negative Episodes

Multilevel modeling revealed significant main effects for data set identifier and episode valence, as well as their interaction (Figure 2 shows episode quantity for each data set; detailed statistical results for all four outcome criteria are provided in the Supplemental Materials). Planned contrasts were conducted to examine differences between study designs. Results indicated that the EC-Signal design ($M = 1.60$, $SD = 1.99$) yielded significantly more episodes than the EC-Event design, $M = 0.39$, $SD = 0.63$; $t(361) = 11.66$, $p < .001$. Further analyses indicated differences between positive and negative episodes. Specifically, in the EC-Signal design, negative episodes ($M = 0.91$, $SD = 1.59$) were triggered more frequently than positive episodes, $M = 0.66$, $SD = 1.41$; $t(6915) = 3.88$, $p = .001$. Conversely, in the EC-Event design, positive episodes ($M = 0.23$, $SD = 0.51$) were reported more frequently than negative episodes, $M = 0.16$, $SD = 0.40$; $t(6915) = -2.37$, $p = .018$.

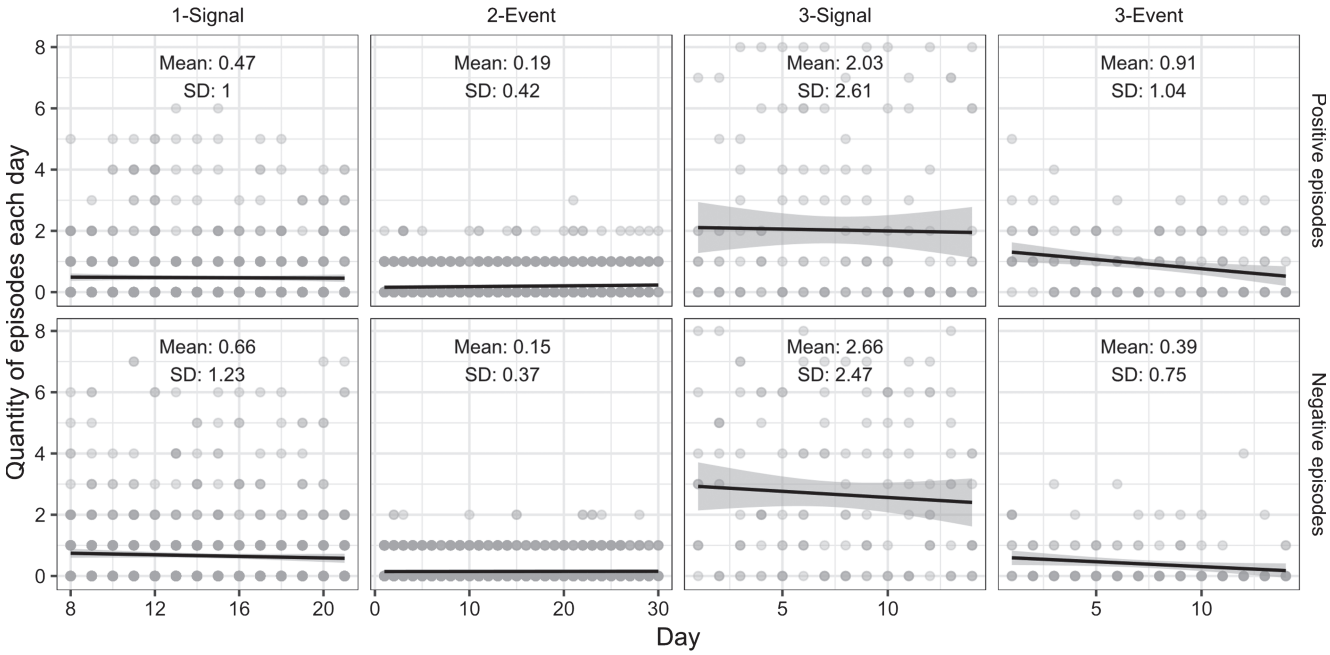
The effect of day was significant for the reference group (negative episode in Data 1-Signal), showing a slight decrease in quantity over time ($b = -0.01$, $p = .049$). All interaction effects with day were nonsignificant ($ps > .05$), indicating no differences in the day effect across data sets or between positive and negative episodes.

Intensity of Positive and Negative Episodes

For emotions measured on a bipolar scale (measured in Data 1-Time, 1-Signal, 2-Event), multilevel modeling revealed significant effects for both data set and episode valence, as well as their interaction (Figure 3). We found that both EC-Signal, $M = 81.72$, $SD = 8.18$; $t(4854) = 12.34$, $p < .001$, and EC-Event, $M = 85.26$, $SD = 12.93$; $t(368) = 13.56$, $p < .001$, captured positive episodes with significantly higher intensity (i.e., more positive) than the emotions recorded in TC designs ($M = 66.98$, $SD = 17.19$), while no significant difference was observed between EC-Signal and EC-Event designs, $t(961) = -1.37$, $p = .173$. In terms of negative episodes, both EC designs, EC-Signal: $M = 48.41$, $SD = 17.41$, $t(4844) = -19.07$,

Figure 2

Quantity of Positive and Negative Episodes Over Time



Note. The dots and their darkness represent how many participants report a specific number of episodes on a given day, with darker dots indicating a higher number of participants. The mean and standard deviation are provided. The black lines represent the fitted linear regression lines, and the shaded areas indicate the 95% confidence intervals. The fitted lines and confidence intervals were plotted using the "lm" method in the geom_smooth function from the ggplot2 package (Wickham, 2016), and the same function was used for Figures 4 and 5.

$p < .001$; EC-Event: $M = 19.87$, $SD = 14.67$; $t(399) = -28.20$, $p < .001$, captured negative episodes with significantly higher intensity (i.e., more negative) compared to random emotions in TC designs. Moreover, EC-Event designs captured negative episodes with significantly more intense negative emotions compared to EC-Signal designs, $t(711) = 11.70$, $p < .001$. The effect of day and its interaction with other predictors were nonsignificant, indicating that episode intensity did not significantly change over time ($ps > .05$). Notably, the average ratings of random emotional states in the TC design were above the midpoint of the scale (i.e., >50), suggesting a general tendency toward positivity. Similarly, emotions recorded during negative episodes—particularly in EC-Signal designs—were occasionally positive on the scale.

For emotions measured on a unipolar scale (Data 3-Time, 3-Signal, 3-Event), the results showed similar patterns. Specifically, positive episodes were associated with significantly more intense positive emotions in both EC-Signal, $M = 82.98$, $SD = 6.44$; $t(3313) = 17.80$, $p < .001$, and EC-Event designs, $M = 75.14$, $SD = 10.16$; $t(339) = 10.09$, $p < .001$, compared to TC designs ($M = 49.15$, $SD = 27.13$). And EC-Signal designs captured positive episodes with significantly more intense positive emotions compared to EC-Event, $t(3313) = 2.33$, $p = .020$. Negative episodes similarly were associated with significantly more intense negative emotion in both EC designs, EC-Signal: $M = 61.06$, $SD = 20.35$; $t(3317) = 27.73$, $p < .001$; EC-Event: $M = 43.52$, $SD = 24.32$; $t(3311) = 9.27$, $p < .001$, compared to TC design ($M = 22.24$, $SD = 22.08$), with EC-Signal capturing significantly more intense negative emotion than EC-Event, $t(3314) = 1.99$, $p = .047$. No significant day effects or interactions with day were observed.

Furthermore, we plotted the distribution of emotion ratings over time within bursts to gain insight into the recovery trends of emotional episodes, as shown in Figure 3. For emotions measured on a bipolar scale, we observed a clear recovery trend, with intensity decreasing over time during both negative and positive episodes. Similarly, for emotions measured on a unipolar scale, positive emotion scores decreased during positive episodes, and negative emotion scores decreased during negative episodes. It is noteworthy that the shape of the recovery differed across designs. For example, the 2-Event data set exhibited a stronger recovery trend, possibly reflecting the longer lasting impact of "uncommon emotional events," where emotions may not have fully returned to baseline even after the burst. In contrast, the 1-Signal data set showed a sharper change between the first two beeps, followed by a flatter trajectory, suggesting a rapid shift within the first 10 min and a slower recovery thereafter. Therefore, burst data may offer valuable high-resolution insights into emotional episodes, revealing details about episode duration and recovery trajectories that would often be overlooked by longer interval measurements.

Feasibility

Burden

We conducted multilevel modeling to examine differences in average burden levels for follow-up and regular beeps separately and to assess whether burden increased over time (Figure 4). For burden annoyance in regular beeps, planned pairwise contrasts showed that EC-Event participants ($M = 16.17$, $SD = 19.28$) reported lower burden annoyance than EC-Signal, $M = 44.76$, $SD = 32.23$; $t(94) = -3.96$, $p < .001$, and TC participants, $M = 30.16$, $SD = 28.76$; $t(99.8) = -3.20$, $p = .002$. No significant difference was observed between TC and EC-Signal designs, $t(108.5) = 1.83$, $p = .070$. A significant day effect was observed for the reference group, Data 1-Time ($B = 1.80$, $p < .001$), indicating an increase in burden over time. Significant interaction effects between day and data set were found, suggesting that the burden trajectories over time differed across data sets. By altering the reference group, we found that burden increased significantly over time for Data Sets 1-Time, 1-Signal ($B = 0.73$, $p < .001$), and 3-Signal ($B = 1.97$, $p < .001$), but not for Data 3-Event ($B = 0.31$, $p = .374$).

For burden annoyance elicited by follow-up beeps, similarly, EC-Event participants ($M = 18.60$, $SD = 20.96$) reported significantly lower burden annoyance than EC-Signal participants, $M = 55.04$, $SD = 33.80$; $t(71.1) = 4.22$, $p < .001$. Burden annoyance increased over time for Data 1-Signal ($B = 0.70$, $p = .001$) and 3-Signal ($B = 1.82$, $p < .001$), but not for Data 3-Event ($B = -0.23$, $p = .301$).

For burden effort during regular beeps, no significant difference was found between EC-Signal ($M = 29.60$, $SD = 25.17$) and EC-Event, $M = 15.17$, $SD = 16.98$; $t(25.5) = -0.849$, $p = .404$. Burden effort increased over time for Data 3-Signal ($B = 1.35$, $p < .001$), but not for Data 3-Event ($B = 0.07$, $p = .825$).

Compliance

Multilevel modeling was applied to investigate differences in average compliance levels for regular and follow-up beeps separately and how these different types of compliance change over time (Figure 5). For regular beeps, significant differences in compliance levels were observed across data sets. Planned pairwise contrasts showed that EC-Signal participants ($M = 0.76$, $SD = 0.26$) had lower compliance than TC design participants, $M = 0.85$, $SD = 0.20$; $t(1931) = -2.66$, $p = .008$. EC-Event ($M = 0.77$, $SD = 0.32$) did not differ significantly in compliance from TC, $t(337) = -1.41$, $p = .159$, or EC-Signal, $t(342) = 0.008$, $p = .993$. Compliance decreased significantly over time for the reference group, Data 1-Time ($B = -0.01$, $p = .002$). Significant interaction effects indicated differing compliance trajectories across data sets. Further analysis revealed a significant compliance decrease over time for Data 1-Time, 1-Signal ($B = -0.004$, $p = .002$), 3-Time ($B = -0.007$, $p = .005$), and 3-Signal ($B = -0.02$, $p < .001$), but not for Data 3-Event ($B = 0.002$, $p = .574$).

For follow-up beeps, EC-Event participants ($M = 0.88$, $SD = 0.18$) had significantly higher compliance rates than EC-Signal participants, $M = 0.63$, $SD = 0.30$; $t(264) = 5.71$, $p < .001$. Compliance decreased over time significantly for Data 3-Signal ($B = -0.02$, $p < .001$), but not for Data 1-Signal ($B = -0.002$, $p = .262$), 2-Event ($B = -0.0005$, $p = .577$), or 3-Event ($B = 0.0001$, $p = .985$).

Qualitative Analysis

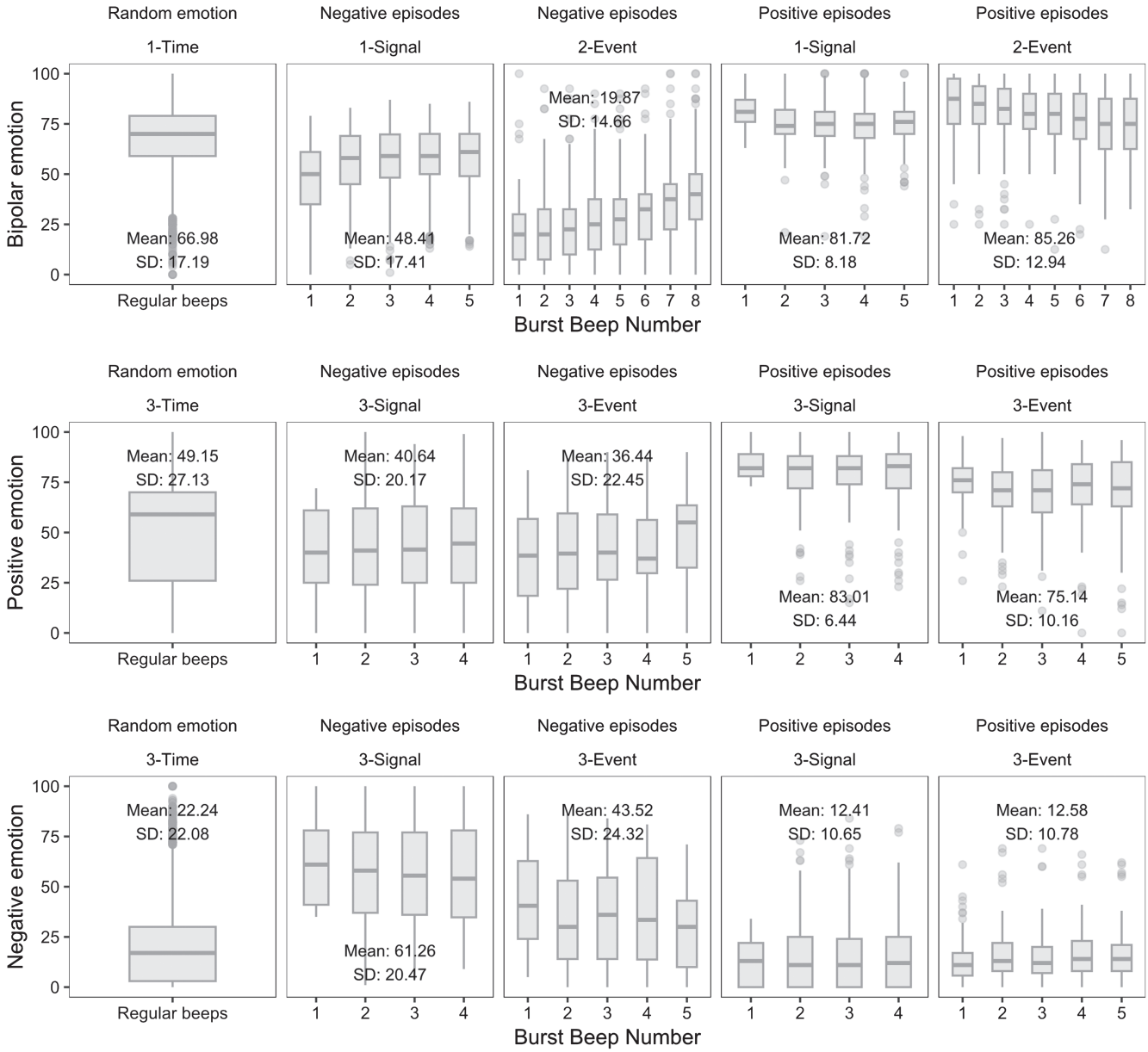
Reason for Missingness

Reasons for missingness were reported in 129 beeps in Data 3-Signal and 47 beeps in Data 3-Event, in which participants could select multiple reasons within a single beep. The primary reason reported by participants for missing the beeps was "I couldn't respond on time," with percentages of 86.05% for 3-Signal and 76.05% for 3-Event. This was followed by "I felt disturbed by the notification," reported in 11.63% of beeps for 3-Signal and 14.89%

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Figure 3

Intensity of Positive and Negative Episodes



Note. Since day effect was not significant, the figure aggregates data across all days for simplicity. The *x*-axis represents the sequence of burst beeps. Note that the time interval between burst beeps was different across data sets. For episode-contingent designs, the mean and standard deviation values presented in the figure were calculated based only on the first burst beep.

for 3-Event. Additionally, "I didn't want to respond" was indicated in 9.30% of beeps in 3-Signal and 14.89% in 3-Event.

Semistructured Interview

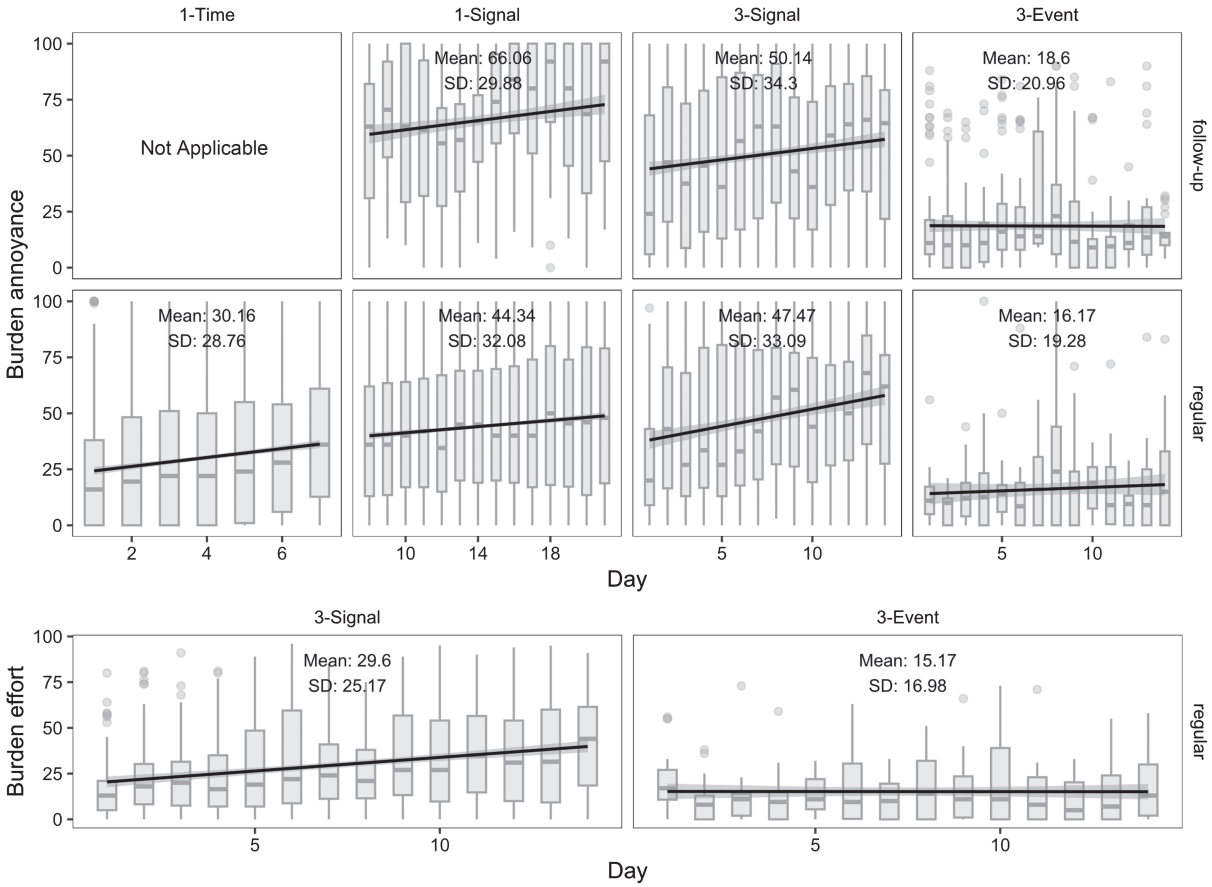
A total of 18 participants from Study 3 were interviewed, of whom nine belonged to the EC-Signal (3-Signal) and nine to the EC-Event (3-Event) group. In terms of effectiveness in detecting episodes, most participants in the EC-Event group reported that they sometimes forgot to press the event button (7/9).[5] Some participants found it difficult to use the button (3/9), were uncertain about how many events they should report (1/9), or were uncertain about what constituted an emotional event (2/9). Additionally, participants sometimes intentionally chose not to report an event, either because they knew they did not have time to respond to follow-ups (2/9) or

[5] Numbers in the parentheses indicate the number of participants who expressed this idea/number of total interviewees in the group. Participants not included did not necessarily hold an opposing view; mostly, they simply did not mention it during their interviews.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Figure 4

Burden Annoyance and Burden Effort Over Time



Note. The upper eight panels and lower two panels present the average level of burden annoyance and burden effort over time, respectively. The boxplots show the distribution of burden scores on each day. The black lines represent the fitted linear regression lines, and the shaded areas indicate the 95% confidence intervals.

because they were unwilling to report personal or negative events that felt too confronting (3/9).

Regarding burden, five out of nine participants in the EC-Signal group reported experiencing higher levels of burden compared to TC designs. This was primarily due to the high frequency of measurements (4/9), the lack of novelty after the earlier TC study (3/9), and personal contextual stressors such as deadlines and exams (2/9). However, the remaining participants found the EC-Signal study less burdensome due to the use of short, repetitive questionnaires that were easy to complete (4/9). Burden was reported to increase over time for both TC and EC-Signal designs (5/9). In contrast, no participants in the EC-Event group experienced more burden compared to the TC designs; some of them even stated that the EC-Event design was not burdensome at all (2/9), primarily due to lower measurement frequency (4/9), shorter questionnaires (1/9), and more freedom and flexibility (2/9). However, some participants noted that it required a lot of effort to remember to log events throughout the study (4/9). EC-Event participants mentioned that the burden of EC-Event did not increase over time (3/9), but rather even decreased a bit after the initial phase (1/9). In both groups, participants frequently mentioned that the beeps were most burdensome when interfering with activities or busy moments, such as meetings or while driving (11/18).

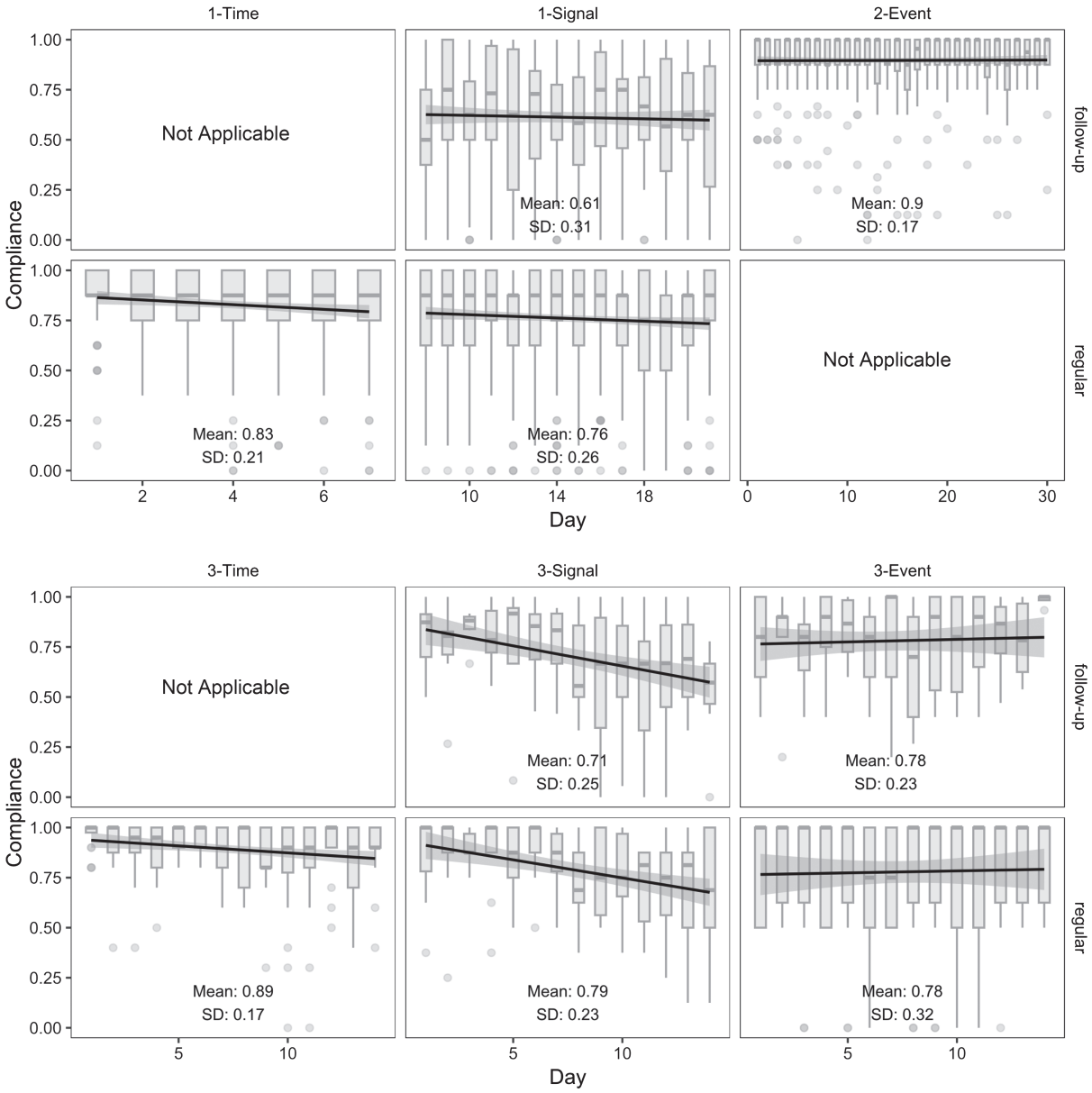
Importantly, participants indicated that burden did not significantly impact the affective responses (2/18). Participants from both groups stated that they were able to report accurately and attentively (6/18), with no noticeable difference between EC and TC designs in attention levels (5/18) and compliance (2/18). Missed beeps were typically due to situations where it was physically impossible to respond, such as job interviews, oversleeping, or simply forgetting (11/18). However, one participant did mention occasionally choosing not to respond when feeling very negative (1/18).

Some additional relevant information emerged. First, while the shorter questionnaire was preferred by some participants, some others felt that the division of positive and negative emotions might not fully reflect their emotional experience and was too simplistic (9/18). Second, some participants reported that the follow-up measurements were so frequent that their emotion experience had not changed across these beeps (4/18). Third, a habituation effect was mentioned in both groups, where participants gradually became familiar with the consistent order of questions, allowing them to respond more swiftly (11/18). However, the repetition could also

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Figure 5

Compliance Rate Over Time



Note. The boxplots show the distribution of compliance for regular and follow-up beeps on each day. The black lines represent the fitted linear regression lines, and the shaded areas indicate the 95% confidence intervals.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

lead participants to experience a lack of novelty and a decline in attention (5/18). Fourth, two participants mentioned that they found it a shame the study did not ask for more detailed information about the events or about what happened between beeps (2/18). Overall, most participants reported their experiences with EC designs to be positive (13/18).

Discussion

In the present study, we investigated an innovative sampling method for measuring emotion dynamics—the EC sampling method, which triggers subsequent measurement bursts upon detecting emotional episodes—regarding its effectiveness and feasibility. This study is a first exploration and comparison of two types of EC designs: signal-based (EC-Signal) and event-based (EC-Event).

Quantitative Results

Generally, both EC-Signal and EC-Event showed promise in capturing emotions of high intensity as compared to the random emotion reports generated by TC designs. The result of the comparison

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

between EC-Signal and TC was expected given that EC-Signal follows up on intense EC-Signal scores and thus by definition yields scores of higher intensity. The comparison of EC-Event and TC designs, on the other hand, provided clear evidence that participants reported emotions of higher intensity relative to the baseline levels reflected by TC. However, both methods might miss emotional episodes for distinct reasons: EC-Signal builds on TC designs for episode detection, implying that episodes that occur in between regular beeps will be missed. EC-Event may miss episodes because participants could forget or choose not to report emotional episodes. Once episodes are detected, both approaches make it possible to monitor how emotions evolve over a more fine-grained timescale, which may provide much-needed insight into the timescale at which emotion dynamics unfold (Adolf et al., 2021; Voelkle et al., 2018).

Moreover, each design showed its own pros and cons. In the present study, EC-Signal captured more emotional episodes compared to EC-Event.[6] Positive episodes were more likely to be reported in EC-Event, while negative episodes were more likely to be detected by the EC-Signal design. Additionally, comparisons of emotional intensity between the two designs yielded mixed results: In Study 3, EC-Signal captured higher intensity episodes than EC-Event, while in the comparison between Data 1-Signal and 2-Event, EC-Event captured more intense negative episodes and similarly intense positive episodes compared to EC-Signal. However, we urge caution when interpreting this finding, as it is influenced by factors such as the instructions provided to participants for the EC-Event design (e.g., minor vs. uncommon emotional events) and thresholds used in the EC-Signal designs. These factors can influence—or be influenced by—how participants or researchers identify emotional episodes, potentially affecting the number and type of episodes captured. For example, Data 2-Event instructed participants to report uncommon emotional events, yielding high-intensity, low-frequency episodes, while Data 3-Event allowed for minor emotional events, resulting in lower intensity episodes that occur more frequently.

Regarding feasibility, we found that the burden of regular beeps in EC-Signal designs was not higher than in TC designs, but the burden of EC-Signal was consistently higher than that of EC-Event for both regular and follow-up beeps. Compliance followed a similar pattern: EC-Signal designs had the lowest compliance (averaging around 63% for follow-up beeps and 76% for regular beeps), while EC-Event designs maintained relatively high compliance for follow-up beeps (averaging 88%). Participants in EC-Signal and TC typically reported increasing burden and showed decreasing compliance over time, whereas EC-Event participants exhibited relatively stable levels of burden and compliance throughout the study. Overall, despite the potential differences between the two designs, both tentatively demonstrated feasibility, with participants generally reporting a positive experience and compliance remaining acceptable.

Qualitative Results

The qualitative analyses of the interviews provided some additional information. First, shorter questionnaires were often preferred by participants, which was consistent with Eisele et al. (2022), who found that longer questionnaires were associated with increased burden and decreased data quantity and quality. However, it is noteworthy that the shorter questionnaire may come at the cost of overly simple questions, and participants might feel they could not fully reflect their emotional experiences. Second, the consistency of question order was advantageous as it allowed participants to "internalize" the questions, thereby reducing their cognitive burden when answering. However, there is a risk that participants may become bored with the repetition. Third, participants reported that they sometimes forgot to press the event button, which can be considered a form of noncompliance (i.e., noncompliance with event logging) that cannot be directly captured in our quantitative data. This form of noncompliance might also influence the results for follow-up beeps, as participants who chose to report an event might have been more motivated or available to respond to subsequent follow-up prompts. Fourth, although participants often mentioned being disturbed or burdened by frequent beeps, they usually had positive experiences with EC designs and could attentively follow the study. Altogether, these findings tentatively support the feasibility of EC designs.

Practical Implications

Overall, our findings offered some initial insights into the applicability of EC designs in capturing emotional and symptom-related episodes (and other events of interest). First, EC-Signal designs are well-fitted for close (emotion) monitoring over shorter time periods, such as exam weeks, the start of a new job, or acute phases of mental health treatment. However, when implementing EC-Signal designs, researchers must weigh the benefit of capturing more episodes against the potential increase in participant burden, which could impact compliance and data quality. EC-Signal designs also require careful predetermination of thresholds, which should be closely aligned with the specific research questions. EC-Event designs, on the other hand, may be preferable for studies focusing on long-term monitoring of (less frequent) emotions/events or for populations sensitive to survey burden. Examples of such events include experiencing specific social interactions (Elmer et al., 2023; Stadel et al., 2024) or, in clinical settings, restrictive eating episodes among patients with eating disorders (Biçaker et al., 2024) and suicidal thoughts among high-risk populations (Coppersmith et al., 2022).

Constraints on Generality and Limitations

Despite the insights provided by our findings, the present study comes with several limitations. First, due to the limited number of existing EC studies in emotion research, our analyses included only three data sets with relatively small sample sizes. Additionally, participants in Study 3 were recruited from a previous ESM study and were willing to continue. This may introduce potential selection bias, as these participants might be more likely to engage in the study and have positive attitudes toward participation.

Moreover, we included only healthy samples, which may limit the generalizability of the findings to (sub)clinical populations. This is particularly important, as vulnerable populations may be more

[6] The pattern of "EC-Signal > EC-Event" for episode quantity was also observed in an additional analysis comparing the two groups within Study 3, where most design variables (e.g., sampling intervals within bursts, questionnaire length) remained constant across groups. It is also noteworthy that minor events were allowed in the EC-Event group in Study 3, meaning that it should be able to capture even more events. This provides preliminary support for the generalizability of this pattern.

sensitive to the burden of repeatedly reporting their momentary experiences, especially when the events are confronting or stressful. Indeed, a meta-analysis showed that psychosis (but not other mental disorders) relates to lower compliance in TC designs (Vachon et al., 2019). Despite this, EC designs may still be feasible in clinical contexts. For instance, Kiekens et al. (2024) used an EC-Signal design with three follow-up measurements at 5- to 10-min intervals triggered when nonsuicidal self-injury urges exceeded a threshold of 5 on a scale from 1 to 7 in regular ESM surveys and reported a median compliance rate of 78.6%. Similarly, Biçaker et al. (2024) implemented EC-Event designs for eating disorders, where participants received one follow-up measurement after reporting an eating episode. The study achieved a relatively high average compliance rate of 82%, with participants reporting an average of nine restrictive eating episodes and 15 skipped meals per individual over a 2-week period. Hence, EC designs need not be restricted to healthy samples and also have the potential for capturing critical clinical incidents.

Third, variations in study designs across data sets may impact comparability. For instance, EC-Signal designs focus on detecting episodes based on high intensity, whereas EC-Event designs emphasize capturing specific "emotional events." This distinction may highlight a divergence in defining what constitutes an emotional episode: EC-Signal tends to adhere to more "objective" criteria, whereas EC-Event leans toward more "subjective" criteria. More detailed variations also existed, such as Data 1-Signal detected emotions using personal quantiles (15th to 85th), while Data 3-Signal used an overall 75th quantile. Moreover, Data 2-Event instructed participants to report "uncommon emotional events," excluding common or minor events, while Data 3-Event allowed for minor emotional events. Such design variations, together with other factors such as sampling frequency, study duration, and question phrasing, could all influence the results. Given these variations, caution is warranted when interpreting our findings.

Future Directions

Considering the design variations and limited sample size in this study, it is important to replicate our findings in future research to establish their robustness. Additionally, future studies should aim to include a broader range of designs or systematically manipulate design variables to enable a more comprehensive and rigorous comparison of EC designs.

An interesting finding of the present study is that some negative episodes had emotion scores on the positive end of the scale (i.e., >50 on a bipolar scale), particularly in EC-Signal designs. In EC-Signal designs, the intensity of an episode is closely tied to the (personal) threshold value, which is calculated based on a specific quantile range of the (personal) emotion distribution. This indicates that the detected episode intensity was distribution-wise negative, while it may not necessarily be scale-wise negative. Similarly, in EC-Event designs, people identified positive and negative episodes based on their subjective appraisal of the event, which could also be "relative" to their own emotional baseline. Although capturing relative emotion dynamics is common in ESM research, when the research focus is on absolute negative emotions, the threshold should reflect scale-wise negativity. Therefore, it is crucial for future research to investigate how to appropriately select thresholds in EC-Signal and formulate clear EC-Event instructions to ensure that the nature of detected episodes aligns with the research goals.

Additionally, future research is needed to further improve these EC designs. For example, adaptive sampling schemes for follow-up beeps could be implemented, depending on how rapidly the emotions are changing during each emotional episode. For instance, to optimally capture emotional recovery, it would perhaps be useful to continue the burst as long as emotions have not yet returned to baseline. If recovery is only taking place slowly, one could increase the time interval between burst beeps in order to minimize participant burden.

Conclusion

EC designs may offer promising avenues for improving our understanding of emotional dynamics and could potentially be applied to capture the unfolding of other critical, clinically relevant incidences. Therefore, validating their practicality is crucial. Our findings provide initial insights and tentative support for the basic feasibility of EC designs, showing their ability to capture emotional episodes with high intensity and provide a fine-grained understanding of their temporal unfolding while maintaining manageable participant burden and generally positive participant experiences. This study also provides the first comparison of two EC approaches for capturing emotion episodes, where EC-Signal captures more episodes than EC-Event but with higher participant burden and lower compliance. These findings underscore the importance of carefully considering the trade-offs between these designs to maximize their effectiveness and feasibility.

References

Adolf, J. K., Loossens, T., Tuerlinckx, F., & Ceulemans, E. (2021). Optimal sampling rates for reliable continuous-time first-order autoregressive and vector autoregressive modeling. *Psychological Methods*, *26*(6), 701–718. https://doi.org/10.1037/met0000398

Biçaker, E., Trolio, V., Miller, A. E., Zhu, L. Y., White, C., & Racine, S. E. (2024). Affective processes underlying restrictive eating in atypical anorexia nervosa: Comparisons to anorexia nervosa and bulimia nervosa using ecological momentary assessment. *International Journal of Eating Disorders*, *57*(4), 879–891. https://doi.org/10.1002/eat.24134

Bryant, F. B., & Veroff, J. (2017). *Savoring: A new model of positive experience* (1st ed.). Psychology Press. https://doi.org/10.4324/9781315088426

Bylsma, L. M., Taylor-Clift, A., & Rottenberg, J. (2011). Emotional reactivity to daily events in major and minor depression. *Journal of Abnormal Psychology*, *120*(1), 155–167. https://doi.org/10.1037/a0021662

Cloos, L., Ceulemans, E., & Kuppens, P. (2023). Development, validation, and comparison of self-report measures for positive and negative affect in intensive longitudinal research. *Psychological Assessment*, *35*(3), 189–204. https://doi.org/10.1037/pas0001200

Cloos, L., Mestdagh, M., Vanpaemel, W., Ceulemans, E., & Kuppens, P. (2024). Measuring continuous affect in daily life with intensity profile drawings. *Assessment*. Advance online publication. https://doi.org/10.1177/10731911241266286

Coppersmith, D. D. L., Fortgang, R. G., Kleiman, E. M., Millner, A. J., Yeager, A. L., Mair, P., & Nock, M. K. (2022). Effect of frequent assessment of suicidal thinking on its incidence and severity: High-resolution real-time monitoring study. *The British Journal of Psychiatry*, *220*(1), 41–43. https://doi.org/10.1192/bjp.2021.97

Dawood, S., Hallquist, M. N., Pincus, A. L., Ram, N., Newman, M. G., Wilson, S. J., & Levy, K. N. (2020). Comparing signal-contingent and

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

event-contingent experience sampling ratings of affect in a sample of psychotherapy outpatients. *Journal of Psychopathology and Behavioral Assessment*, *42*(1), 13–24. https://doi.org/10.1007/s10862-019-09766-7

Dejonckheere, E., & Erbaş, Y. (2021). Designing an experience sampling study. In I. Myin-Germeys & P. Kuppens (Eds.), *The open handbook of experience sampling methodology* (pp. 33–69). Amazon Kindle Direct Publishing.

Dejonckheere, E., & Mestdagh, M. (2021). On the signal-to-noise ratio in real-life emotional time series. In C. E. Waugh & P. Kuppens (Eds.), *Affect dynamics* (pp. 131–152). Springer International Publishing. https://doi.org/10.1007/978-3-030-82965-0_7

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

Dejonckheere, E., Verdonck, S., Andries, J., Röhrig, N., Piot, M., Kilani, G., & Mestdagh, M. (2024). Real-time incentivizing survey completion with game-based rewards in experience sampling research may increase data quantity, but reduces data quality. *Computers in Human Behavior*, *160*, Article 108360. https://doi.org/10.1016/j.chb.2024.108360

Delecroix, C., van Nes, E. H., Scheffer, M., & van de Leemput, I. A. (2024). Monitoring resilience in bursts. *Proceedings of the National Academy of Sciences of the United States of America*, *121*(31), Article e2407148121. https://doi.org/10.1073/pnas.2407148121

Eisele, G., Vachon, H., Lafit, G., Kuppens, P., Houben, M., Myin-Germeys, I., & Viechtbauer, W. (2022). The effects of sampling frequency and questionnaire length on perceived burden, compliance, and careless responding in experience sampling data in a student population. *Assessment*, *29*(2), 136–151. https://doi.org/10.1177/1073191120957102

Eisele, G., Vachon, H., Lafit, G., Tuyaerts, D., Houben, M., Kuppens, P., Myin-Germeys, I., & Viechtbauer, W. (2023). A mixed-method investigation into measurement reactivity to the experience sampling method: The role of sampling protocol and individual characteristics. *Psychological Assessment*, *35*(1), 68–81. https://doi.org/10.1037/pas0001177

Eisele, G., Vachon, H., Myin-Germeys, I., & Viechtbauer, W. (2021). Reported affect changes as a function of response delay: Findings from a pooled dataset of nine experience sampling studies. *Frontiers in Psychology*, *12*, Article 580684. https://doi.org/10.3389/fpsyg.2021.580684

Elmer, T., Ram, N., Gloster, A. T., & Bringmann, L. F. (2023). Studying daily social interaction quantity and quality in relation to depression change: A multi-phase experience sampling study. *Personality and Social Psychology Bulletin*. Advance online publication. https://doi.org/10.1177/01461672231211469

Fuller-Tyszkiewicz, M., Skouteris, H., Richardson, B., Blore, J., Holmes, M., & Mills, J. (2013). Does the burden of the experience sampling method undermine data quality in state body image research? *Body Image*, *10*(4), 607–613. https://doi.org/10.1016/j.bodyim.2013.06.003

Hasselhorn, K., Ottenstein, C., & Lischetzke, T. (2025). Modeling careless responding in ambulatory assessment studies using multilevel latent class analysis: Factors influencing careless responding. *Psychological Methods*, *30*(2), 374–392. https://doi.org/10.1037/met0000580

Heininga, V. E., Dejonckheere, E., Houben, M., Obbels, J., Sienaert, P., Leroy, B., van Roy, J., & Kuppens, P. (2019). The dynamical signature of anhedonia in major depressive disorder: Positive emotion dynamics, reactivity, and recovery. *BMC Psychiatry*, *19*(1), Article 59. https://doi.org/10.1186/s12888-018-1983-5

Herres, J., Caporino, N. E., Cummings, C. M., & Kendall, P. C. (2018). Emotional reactivity to daily events in youth with anxiety disorders. *Anxiety, Stress, and Coping*, *31*(4), 387–401. https://doi.org/10.1080/10615806.2018.1472492

Himmelstein, P. H., Woods, W. C., & Wright, A. G. C. (2019). A comparison of signal- and event-contingent ambulatory assessment of interpersonal behavior and affect in social situations. *Psychological Assessment*, *31*(7), 952–960. https://doi.org/10.1037/pas0000718

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

Ji, Y., Schreuder, M. J., Ariens, S., Dejonckheere, E., & Ceulemans, E. (2024). *An evaluation of the feasibility and outcomes of burst ESM designs*. https://doi.org/10.17605/OSF.IO/FM2DY

Ji, Y., Schreuder, M. J., Ariens, S., Dejonckheere, E., & Ceulemans, E. (2025). *Evaluating the feasibility of episode-contingent experience-sampling burst designs*. https://doi.org/10.17605/OSF.IO/QXJ7Y

Kalokerinos, E. K., Résibois, M., Verduyn, P., & Kuppens, P. (2017). The temporal deployment of emotion regulation strategies during negative emotional episodes. *Emotion*, *17*(3), 450–458. https://doi.org/10.1037/emo0000248

Kiekens, G., Claes, L., Kleiman, E. M., Luyckx, K., Coppersmith, D. D. L., Fortgang, R. G., Myin-Germeys, I., & Nock, M. K. (2024). The short-term course of nonsuicidal self-injury among individuals seeking psychiatric treatment. *JAMA Network Open*, *7*(10), Article e2440510. https://doi.org/10.1001/jamanetworkopen.2024.40510

Koval, P., Kuppens, P., Allen, N. B., & Sheeber, L. (2012). Getting stuck in depression: The roles of rumination and emotional inertia. *Cognition and Emotion*, *26*(8), 1412–1427. https://doi.org/10.1080/02699931.2012.667392

Kuppens, P., Allen, N. B., & Sheeber, L. B. (2010). Emotional inertia and psychological maladjustment. *Psychological Science*, *21*(7), 984–991. https://doi.org/10.1177/0956797610372634

Mestdagh, M., Verdonck, S., Piot, M., Niemeijer, K., Kilani, G., Tuerlinckx, F., Kuppens, P., & Dejonckheere, E. (2023). m-Path: An easy-to-use and highly tailorable platform for ecological momentary assessment and intervention in behavioral research and clinical practice. *Frontiers in Digital Health*, *5*, Article 1182175. https://doi.org/10.3389/fdgth.2023.1182175

Myin-Germeys, I., Oorschot, M., Collip, D., Lataster, J., Delespaul, P., & van Os, J. (2009). Experience sampling research in psychopathology: Opening the black box of daily life. *Psychological Medicine*, *39*(9), 1533–1547. https://doi.org/10.1017/S0033291708004947

R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Revol, J., Ariens, S., Lafit, G., Adolf, J., & Ceulemans, E. (2025). Episode-contingent experience-sampling designs for accurate estimates of autoregressive dynamics. *Psychological Methods*. Advance online publication. https://doi.org/10.1037/met0000758

Revol, J., Carlier, C., Lafit, G., Verhees, M., Sels, L., & Ceulemans, E. (2024). Preprocessing experience-sampling-method data: A step-by-step framework, tutorial website, r package, and reporting templates. *Advances in Methods and Practices in Psychological Science*, *7*(4), 1–18. https://doi.org/10.1177/25152459241256609

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

Schreuder, M. J., Ariens, S., Lafit, G., & Ceulemans, E. (2024). Bouncing back from emotional ups and downs: Insights in emotional recovery using survival analyses of burst ESM data. *Cognition and Emotion*. Advance online publication. https://doi.org/10.1080/02699931.2024.2423706

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

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.

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

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

Stadel, M., van Duijn, M. A. J., Wright, A. G. C., Bringmann, L. F., & Elmer, T. (2024). Considering the 'With Whom': Differences between event- and signal-contingent ESM Data of person-specific social interactions. *Multivariate Behavioral Research*, *59*(4), 841–858. https://doi.org/10.1080/00273171.2024.2335405

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), Article e14475. https://doi.org/10.2196/14475

Verduyn, P., & Lavrijsen, S. (2015). Which emotions last longest and why: The role of event importance and rumination. *Motivation and Emotion*, *39*(1), 119–127. https://doi.org/10.1007/s11031-014-9445-y

Voelkle, M. C., Gische, C., Driver, C. C., & Lindenberger, U. (2018). The role of time in the quest for understanding psychological mechanisms. *Multivariate Behavioral Research*, *53*(6), 782–805. https://doi.org/10.1080/00273171.2018.1496813

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag. https://ggplot2.tidyverse.org

Received September 21, 2024
Revision received April 30, 2025
Accepted May 7, 2025 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
All rights, including for text and data mining, AI training, and similar technologies, are reserved.