New developments in Experience Sampling Methodology

Francis Tuerlinckx[1*], Peter Kuppens[1], Sigert Ariens[1], Leonie Cloos[1], Egon Dejonckheere[1,2,3], Ginette Lafit[1], Koen Niemeijer[1], Jordan Revol[1], Evelien Schat[1], Marieke Schreuder[1,2], Niels Vanhasbroeck[1,4], Eva Ceulemans[1]

1: Faculty of Psychology and Educational Sciences, KU Leuven - University of Leuven, Belgium

2: Tilburg School Social and Behavioral Sciences, Tilburg University, The Netherlands

3: m-Path Software

4: Psychological Methods, University of Amsterdam, The Netherlands

*: corresponding author (francis.tuerlinckx@kuleuven.be)

Author note

Correspondence concerning this article should be addressed to Francis Tuerlinckx, Faculty of Psychology and Educational Sciences, KU Leuven - University of Leuven, Tiensestraat 102, 3000 Leuven, Belgium. Email: francis.tuerlinckx@kuleuven.be.

The writing of this article was supported by a KU Leuven Research Council grant (C14/23/062) awarded to Francis Tuerlinckx, Eva Ceulemans and Peter Kuppens, IBOF grant (IBOF/21/090) awarded to Peter Kuppens and Eva Ceulemans, and the EOS excellence of science program (EOS 4000528/ G0I2422N) awarded to Eva Ceulemans.

Egon Dejonckheere, Francis Tuerlinckx and Peter Kuppens are shareholders and member of the board of directors of m-Path Software BV, the spin-off company commercializing m-Path, a platform for experience sampling research and clinical practice.

New developments in Experience Sampling Methodology

Abstract

Experience Sampling Methodology (ESM) has been widely used over the past decades to study feelings, behavior, and thoughts as they occur in daily life. Typically, participants complete several assessments per day via a smartphone, for multiple days. The growing adoption of ESM has spurred a number of methodological advancements. In this paper, we provide an overview of recent developments in ESM design, statistical analysis, and implementation. In terms of design, we discuss considerations around what to measure—including the reliability and validity of self-report measures as well as mobile sensing—as well as when to measure, where we focus on the pros and cons of how burst designs and advances in sample size planning methodology. Regarding statistical analysis, we highlight nonlinear models, survival analysis for understanding time-to-event data, and real-time monitoring of ESM time series. At the implementation level, we address open science practices and advances in data preprocessing. Although most of the topics discussed in this paper are generic, many of the examples are focused on the study of affect in daily life.

Introduction

A significant part of the psychological research community focuses nowadays on how complex psychological processes evolve dynamically across time within single individuals while going about their daily lives. Although the methodology to do this research has already been developed in the last century (Delespaul, 1995; DeVries, 1992; Larsen, 1987; Stone et al., 1999), the dynamic within-person paradigm (Molenaar, 2004) has been gaining momentum over the past decades in affective science (Kuppens et al., 2010), psychopathology research (Borsboom & Cramer, 2013), but in other fields as well, concerning for instance attachment (Verhees et al., 2023), suicide and self-harm (Coppersmith et al., 2023), personality (Costantini et al., 2015), attitudes (Dalege et al., 2016), and so on.

To study within-person psychological processes, researchers obviously need to repeatedly and frequently measure the process of interest in ecologically valid ways. Experience Sampling Methodology (ESM; also labeled Ecological Momentary Assessment (EMA) or Ambulatory Assessment) is the gold standard: Using a smartphone app, participants self-report on their momentary feelings, cognitions, behaviors and/or context by responding to a number of questions (also called items) at various moments throughout the day for the duration of several days (typically one or more weeks).

Several resources exist to introduce novice researchers into the common practices of conducting ESM research (e.g., Fritz et al., 2024; Myin-Germeys & Kuppens, 2025) or to point to open but pressing issues (Stone et al., 2023). Yet, there remain many challenges tied to using this method. The aim of this paper is to discuss new developments and upcoming trends in ESM methodology that deal with these challenges, with respect to three crucial aspects of ESM research in which methodology plays a significant role: design, statistical analysis, and implementation.

First, researchers have to design their study and decide which items they will include and when these will be sent to participants, as well as how many participants and measurements are needed. Second, the resulting data are complex, requiring to account for their multivariate, nested structure, and temporally dependent nature. As the questions that researchers want to answer using such complex data are very diverse, this requires an equally diverse statistical toolbox. Third, the future success of ESM research hinges to a large extent on the reproducibility and replicability of the results. To optimize this, researchers require easy-to-use and reliable tools to check and preprocess the data before analysis, and to transparently report on the different implementation steps taken.

Before starting our discussion of recent developments regarding design, statistical analysis and implementation, we mention a few side notes. First, we take the study of affective dynamics as a starting point, given its relative preponderance in contemporary ESM literature and our own familiarity with the topic and the resulting expertise. However, this is mostly without loss of generality as the issues and trends we describe are generic and therefore relevant for all domains relying on ESM. Second, the statistical analysis of ESM data pervades all three aspects we will discuss. Unless stated differently, we take multilevel (vector) autoregressive (VAR) modeling as the standard point of reference when discussing statistical developments. The core idea of VAR models is that current affect scores are predicted based on one's affect scores at the previous time point, and possibly additional predictor variables (see also below). Multilevel VAR modeling can be considered the field's reference point (Ariens et al., 2020; Hamaker et al., 2015; Hamaker & Wichers, 2017), as autoregressive modeling allows for controlling serial dependency while studying contextual effects onto psychological processes (Ariens et al., 2023; Bringmann et al., 2024). The multivariate (i.e., vector) nature of the model makes it the statistical workhorse behind dynamic network modeling (Bringmann et al., 2022).

Design

The most common practice in ESM research is to apply a time-contingent sampling scheme, which involves assessing a set of items at semi-random times throughout the day. Both the number and the content of the items (*What to measure*) and the timing of the measurements (*When to measure*), as well as their total number (*Sample size planning*) should be determined in order to achieve an optimal trade-off between participant burden and the amount of information gathered (Myin-Germeys & Kuppens, 2025). In the following sections, we will discuss the latest developments regarding what and when to measure, including alternative sampling schemes such as signal-based designs and burst designs, and regarding sample size planning.

What to measure?

Psychological processes, such as individuals' feelings and thoughts and the contexts in which these occur, are commonly assessed through self-report surveys. These questionnaires are essential for capturing fundamental aspects of subjective experiences, like emotions or thoughts (Gray & Watson, 2007). When implemented in ESM studies, self-report measures offer immediate insight on internal states, with a reduced risk of memory biases (Scollon et al., 2003).

The burdensome nature of repeated measures necessitates the use of short surveys. Researchers typically select a set of items from existing questionnaires to create an adapted scale, others use single items to reduce the length of the survey. Unfortunately, there is no widely agreed upon standard method to develop measures. For example, the multi-item scales widely vary across studies (e.g. Dejonckheere et al., 2019; Hall et al., 2021) and, many measures lack theoretical justification and psychometric evidence (Brose et al., 2020), undermining their validity. To facilitate and standardize the item selection process, an item repository (Kirtley et al., 2021; see below in the section on open science practices) has been developed with expert-based guidelines to select items (Eisele et al., 2024), encouraging researchers to carefully consider their measures and ensure both theoretical and empirical justification.

In the next subsections, we will discuss reliability and validity issues of self-report measures, followed by the use of mobile sensing in ESM as a potential alternative to self-report.

Reliability and validity of self-report

Regarding reliability, different methods to evaluate reliability exist, depending on whether the focus is on single items or on constructs measured by multiple items. For single items, Schuurman and Hamaker (2019) have developed a state space modeling approach while Dejonckheere et al. (2022) use a model-free approach, based on a specific design involving a test-retest component (i.e., one item is repeated at the end of an assessment). When applied to the same data, the method of Schuurman and Hamaker (2019) estimates that 30-50% of the total variability in participants' affective responses is due to error, while Dejonckheere et al. (2022) estimate this to be 27% of the total variability. On the other hand, for multi-item scales, usually factor modeling or an approach based on generalizability theory is used; see Castro-Alvarez et al. (2024) for a systematic overview of the various approaches. When applying some of these approaches, Haney et al. (2023) find in four studies that for a particular set of items often used in ESM studies (the PANAS-X or the Positive and Negative Affect Schedule – Expanded Form; Watson & Clark, 1999), the between-person reliability is very good and the reliability to detect within-person changes is also good, but may depend on additional factors (e.g., study design).

Regarding validity of measures, key construct validity criteria that have been investigated include their ability to capture momentary variability and contextual effects effectively (Cloos et al., 2023), as well as their ability to capture between-person differences in dynamics (Wenzel & Brose, 2023). A major finding from Cloos et al. (2023) is that, to capture variability in momentary affect, single-item measures (e.g., indicating your feeling on a scale ranging from negative to positive) are often an efficient, sufficient, and low-burden assessment tool for momentary affect (according to several criteria they even outperform multi-item scales). In Cloos et al. (2023), the approach taken was made to align with the perspective of validity described in the Standards (American Educational Research Association, 2014). Still other validity criteria could be implemented as well such as formal content validity (Spoto et al.,2025) or approaches focusing on systematic missingness patterns (Silvia et al. 2013; Rintala et al., 2019) or method reactivity (Barta, 2012).

In the preceding discussion on reliability and validity, we have assumed that item structure remained unaltered across persons, context, and time. However, the internal structure may show variation across these dimensions. To take this into account, there exists a scala of sophisticated methods for a more fine-grained analysis. For instance, multilevel factor analyses evaluate measurement invariance of within- and between-person measurement models (Rush & Hofer, 2014), and latent Markov factor analysis studies the time dynamics of the structure and the measurement models (Vogelsmeier et al., 2023).

Mobile sensing

While ESM has many benefits, completing a number of self-report questions several times a day for multiple days becomes challenging when collecting fine-grained data over extended periods (Eisele et al., 2022). These demands often result in decreased data quality, lower compliance, and higher participant drop-out. Mobile sensing has emerged as a promising tool in mental health research, offering continuous, objective, and real-time data with no additional burden for participants (Insel, 2018; Onnela, 2021). For example, smartphone sensors can passively collect data on physical activity, GPS location, and device interactions, constructing a digital "fingerprint" that offers insights into emotional and physical health (Mehl et al., 2024).

Despite its potential, mobile sensing has yet to establish consistent and reliable links between sensing features and emotions or, more broadly, mental health (De Angel et al., 2022; Niemeijer & Kuppens, 2024). A significant barrier lies in methodological and analytical challenges, where researchers often face numerous arbitrary choices without clear guidelines. One solution is the adoption of multiverse analysis, a method that systematically explores the impact of different preprocessing and analytical decisions by considering multiple equally plausible pathways (Langener, Stulp, et al., 2024; Steegen et al., 2016; but see Del Giudice & Gangestad, 2021, for a critical perspective). Another development is the introduction of a preregistration template for mobile sensing studies (Langener, Siepe, et al., 2024) to document major methodological choices in advance, reducing biases and enhancing rigor and replicability in study design and analysis (see below in the section on "Open science practices in ESM"). Beyond methodological challenges, mobile sensing research must adapt to constant changes in operating system policies, which increasingly limit background data collection to protect user privacy. Addressing data protection and privacy concerns remains essential for responsible sensing research (Hong, 2024).

Aside from using mobile sensing to directly measure the phenomenon of interest, future applications of mobile sensing could involve replacing or corroborating specific ESM items. For instance, instead of asking participants about their location, researchers could measure it directly using GPS data. This approach is particularly beneficial when participants are unable or unwilling to accurately report on their feelings or behaviours (Wenz et al., 2024).

When to measure?

A key feature of most processes and states (such as affective experience) that one wants to measure with ESM is that they are not static but subject to continuous change and regulation (Waugh & Kuppens, 2021) and follow trends such as circadian rhythms (Seizer, et al. 2024). A crucial decision when designing an ESM study thus concerns the timing of the measurements: which sampling scheme allows to optimally capture such change and regulation, while allowing to adjust for trends. The advice that is given by experts remains relatively vague, however: "… assessments should occur at a timescale that is appropriate to the … processes of interest …" (Trull et al., 2015, p. 356). Given the lack of advice, many ESM studies implement a semirandom time-contingent sampling scheme, whereby measurement notifications are sent within predefined intervals that often last multiple hours.

Evidence is emerging that using such a time-contingent sampling scheme, might not be ideal when it comes to capturing the dynamics of affective experiences, which is crucial for the prediction, diagnosis and prevention of mental health problems (Kuppens & Verduyn, 2015; Lapate & Heller, 2020). Indeed, if measurements are hours apart, this comes with the risk of missing episodes of more intense affect and hence of missing information about how these episodes unfold and are regulated. To deal with this limitation, researchers have started to explore alternatives for scheduling measurements.

A first alternative is the episode-contingent 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). To this end, multiple closely scheduled (only minutes apart) measurements are sent following the detection of an emotional episode, allowing for more accurate estimation of the regulatory dynamics (Revol et al., 2025; Schreuder et al., 2023b). Such burst designs have been shown to be informative in clinical contexts, to investigate critical events such as self-harm behaviors (Kiekens et al., 2023), suicidal thoughts (Coppersmith et al., 2022), and binge eating (Biçaker et al., 2024).

An important question regarding such burst designs, is how to decide when to start a burst (e.g., Revol et al., 2025; Lathia et al., 2013; Pejovic et al., 2016). This can be achieved through participant initiation (Dejonckheere & Erbaş, 2021), by checking whether an ESM score exceeds a specific threshold (Revol et al., 2025), by using physiological data from wearables (Osotsi et al., 2020) or by using mobile sensing information, where underlying mobile sensing patterns, such as changes in physical activity, are used to trigger ESM prompts during moments of interest, a technique that has been labeled Geographical Ecological Momentary Assessment (Kestens & Kingsbury, 2024). Other important topics are the implications of burst designs for participant burden, compliance, and measurement validity (Doherty et al., 2020; Lathia et al., 2013; van Berkel et al., 2019). Indeed, there are first indications that the faster sampling scheme of burst designs leads to higher burden experience and lower compliance ([ANONYMIZED], 2025), which may lead to lower measurement validity.

A second alternative is to ask participants to provide drawings of their experienced affect trajectories between two standard ESM measurements, yielding continuous data (see Figure 1 for an example). This method has been shown to provide a substantial amount of informational gain (average 7%). On average, it reveals the occurrence of one positive and negative affect peaks or valley per day that would otherwise go unnoticed based on the typical momentary assessments (and the linear trajectories that can be interpolated from it) (Cloos, Mestdagh, et al., 2024). While this method is promising, further research is needed into how to extract meaningful parameters of affect dynamics from the drawings. Additionally, the retrospective nature of the task may introduce peak-end biases, which require careful investigation (Cloos, Vanpaemel, et al., 2024).

Figure 1. Continuous affect rating item displayed on a mobile device. The blue dot on the y-axis represents the momentary affect rating recorded at the previous assessment (15:14), while the blue dot at the opposite end of the graph represents the current momentary affect rating. The connecting line illustrates the fluctuation in negative affect between the two assessments. The y-coordinates represent the continuous affect intensity data for each timepoint on the x-axis, recorded per minute.

Sample size planning

As holds for all empirical research, a sufficiently large sample size (i.e., number of participants and number of measurement occasions) is needed in ESM research to reliably answer the research questions at hand (Arand & Schäfer, 2019; Schultzberg & Muthén, 2018). Moreover, careful sample size planning also prevents wasting valuable resources (e.g., the participants' time and public money). Yet, while ESM studies wildly vary in number of participants and number of measurement occasions (Wrzus & Neubauer, 2023), most of the recently published ESM studies unfortunately lack a rationale for the selected sample size (see Trull & Ebner-Priemer, 2020).

A good starting point for conducting ESM sample size planning in a more systematic way are the recently proposed frameworks for conducting simulation-based power analysis for various state-of-the-art statistical models popular in ESM research, as well as associated open-source software and step-by-step tutorials (e.g., person-specific VAR(1) models in Revol et al., 2024a; multilevel models in Lafit et al., 2021; Lane and Hennes, 2018; Lafit et al., 2022).

For example, to compute statistical power for a fixed effect of interest in a multilevel model, the simulation-based approach works as follows. First, the researcher selects the statistical model that will be used to investigate the effect and specifies plausible values of its parameters. Moreover, the researcher indicates which numbers of participants and measurements per participant are feasible. Using this input, a large number of data sets are simulated. The specified statistical model is then fitted to each of these data sets and the hypothesis of interest is tested. Finally, statistical power is estimated as the percentage of generated datasets in which the null hypothesis of interest was rejected at the selected alpha level. The user can then use this information to select the number of participants and measurements.

Regarding the optimal balance between number of participants and measurements per participant, first results obtained with such simulation based power analyses (Lafit et al., 2021) indicate that findings from multilevel models that do not account for temporal dependencies (e.g., Bell et al., 2014; Maas & Hox, 2005; Scherbaum & Ferreter, 2009) can still be informative for models commonly used in ESM research. Specifically, when testing the effect of a person-level (Level 2) predictor, the number of persons generally has a greater impact on statistical power than the number of time points. In contrast, when investigating within-person (Level 1) effects such as the auto-regressive or cross-regressive effects or the effect of time-varying predictors, the number of repeated measurements becomes critical. Nevertheless, optimizing the number of persons remains important, as it influences the estimation of random effect variance components. For cross-level interaction effects, both Level 1 and Level 2 sample sizes matter, though the number of persons typically has a greater effect on power.

Despite its flexibility and potential, simulation-based power analysis still faces challenges that must be addressed for practical use. First, this approach is computationally intensive because of the many simulations. Analytical power calculations – that is, power calculations based on formulas rather than simulations– can complement and speed up the simulation-based approach (see e.g., Lafit, Revol, et al., 2024). To this end, Lafit et al. (2024) derived analytical power calculations for multilevel models with autocorrelated within-person errors. Yet, such analytical methods are still missing for multilevel (V)AR models and come with multiple specific assumptions.

Second, statistical power depends on a hypothesized model, that is, an assumed data-generating process including the model's parameter values and effect sizes (see the next section on “Statistical analysis”). Ideally, researchers determine these values based on the smallest effect size of interest (SESOI; Lakens, 2022). However, SESOI is not commonly established in experience sampling method (ESM) research, as theories about time-varying relationships remain underdeveloped (Gabriel et al., 2019; Schreuder et al., 2020). Moreover, standardizing effect sizes in the complex statistical models in ESM research is challenging (Rights & Sterba, 2019). Therefore, researchers often rely on existing ESM data sets or pilot studies. To address the uncertainty associated when relying on previous studies, Lafit, Revol, et al. (2024) introduced a safeguard power analysis, which accounts for variations in study design and preprocessing decisions by using a lower-bound estimate of model parameters (e.g., the lower bound of a confidence interval) to calculate statistical power.

Third, study costs are often overlooked in sample size planning. Conducting an ESM study involves overhead costs (e.g., data collection software, hiring research assistants) and participant renumeration. Ignoring these factors can lead to suboptimal or unrealistic sample sizes, wasting resources and comprised study reliability. While cost considerations have been incorporated in other behavioral science fields (e.g., Moerbeek, 2008), they remain underutilized in ESM research. Revol et al. (2024a) propose a framework for integrating cost into power analysis by specifying four key parameters: fixed costs, participant costs, measurement occasion costs, and incentive-related costs.

Finally, traditional power analysis focuses on estimating sample size for questions of interest about individual model parameters, which raises challenges when studying multiple effects or evaluating overall model quality. An alternative approach is predictive accuracy analysis, which evaluates a model's capacity to generalize by assessing how accurately it can predict future observations with a specified sample size (Yarkoni & Westfall, 2017). If predictive accuracy is too low, sample size must be increased. Building on this idea, Revol et al. (2024b) developed a predictive accuracy-based sample size planning method for the VAR(1) modeling framework.

Statistical analysis

As discussed by Ariens et al. (2020) and Hamaker et al. (2015), a wide variety of methods exist to analyze data from an ESM study (see also Bolger & Laurenceau, 2013; Walls & Schafer, 2006). Hamaker et al. (2015) classified the approaches based on eight dichotomies (e.g., single-subject vs. multiple-subject, univariate vs. multivariate, etc.). As stated in the introduction, a very common way of analyzing ESM data is through the multilevel (vector) autoregressive model with external predictors (e.g., coding for context information). Typically, the intercept and auto/crossregressive weight are assumed to vary randomly over persons (possibly together with other parameters).

Although we use in this paper the multilevel VAR as the reference model, we are well aware that many other approaches exist (often built off a VAR model) to analyze ESM data: dynamic structural equation models or DSEM (Hamaker et al., 2018), continuous time structural equation models (Driver et al., 2017), time-varying autoregressions (Bringmann et al., 2017), just to name a few. In the following subsections, we discuss a number of very recent developments that have not been part of the past overview papers. We acknowledge that this selection is idiosyncratic and to some extent arbitrary.

Nonlinear models

At the heart of the VAR-based models is a linear dynamical equation that is basically a multiple linear regression model. Such linear statistical models remain the standard for analyzing affect dynamical data, but the linearity assumption does not always conform with theoretical approaches on the nature of affect (e.g., Frijda, 2007; Scherer, 2009; Friston et al., 2006; see also Lewis, 2005) nor with empirical observations (e.g., Bonsall et al., 2013; Haslbeck et al., 2023; Loossens et al., 2020; Schimmack, 2001; van de Leemput, 2014). Vanhasbroeck et al. (2024) demonstrated that nonlinear phenomena occur in ESM data even after accounting for exogeneous influences from context, meaning that they cannot be explained away by such influences. In other words, by employing these linear models on affective data, we may be missing out on some important inherent characteristics of how affect changes nonlinearly over time.

To alleviate this difficulty, researchers have increasingly moved to models that can handle these nonlinear features. One way in which this is achieved is through the extension of linear statistical models to make them more flexible. An example of such a model is the regime-switching vector autoregressive model – an extension of the VAR that allows for changes in its parameters based on a probability of switching “regimes” (Hamaker et al., 2010; Hamilton, 1989). Another, related, example is the time-varying vector autoregressive model, which allows for more continuous changes in its parameters (Albers & Bringmann, 2020). Both extensions allow for capturing some often observed nonlinear patterns in affective data, more specifically for changes between emotional episodes.

Unfortunately, however, these types of extensions suffer from two limitations. First, the extensions’ abilities come at the cost of having to estimate a nontrivial number of parameters, making the models difficult to estimate. Additionally, these models are only able to deal with some, but not all observed nonlinearities, nor are they able to explain where these nonlinearities originate from, limiting their use in practice.

Because of these limitations, some researchers have introduced models whose primary aim it is to explain nonlinear dynamics from first principles. An example of such a model is the recently proposed Affective Ising Model (AIM; Loossens et al., 2020; Vanhasbroeck et al., 2024), which is specifically relevant for affect. The AIM is a continuous-time, nonlinear drift-diffusion model that was created with the purpose of being able to handle nonlinear features such as skew and multimodality in affective time-series. Unlike the extended linear models mentioned above, the AIM achieves this with a minimal number of parameters to be estimated. It is therefore able to capture a wide range of nonlinear phenomena while remaining parsimonious as a model.

Modeling emotional inertia and recovery with survival analysis

One of the reasons for the popularity of the (multilevel vector) autoregressive model is the close link between the autoregressive parameter(s) and the notions of emotional inertia and emotional recovery speed, which are important characteristics of affective dynamics. Emotional inertia indicates the extent to which intense emotional states tend to linger over time. Vice versa, emotional recovery is the process of returning to one’s emotional baseline after experiencing an emotional event. Higher autoregressive effects thus imply that intense emotions tend to linger and that emotional recovery takes more time. Despite its popularity, autoregressive modeling faces some limitations. First, it does not delineate emotional episodes but rather considers all obtained emotion scores over time. Therefore, autoregressive modeling does not directly shed light on the duration of specific emotional episodes and possible time-varying and person-level predictors thereof. Second, autoregressive modeling assumes that recovery presents itself as exponential decay, but this assumption is unsubstantiated.

These limitations can be addressed by applying time-to-event or survival analysis to model the duration of emotional episodes (i.e., the survival time), irrespective of the shape of the recovery trajectory. This technique is regularly used in medicine and epidemiology to model survival times of patients after a cancer treatment. To be able to apply this technique to ESM data, it was thus crucial to propose ways of identifying emotional episodes in ESM data. For instance, an episode has been defined to start after the occurrence of a stressful event or at a relatively extreme ESM score and to end when the ESM scores were close to the intensity before the event or close to the mean affect level. Applying survival analysis to these episode durations, has shown that emotionally recovering from stressful events takes around 63 minutes for non-depressed individuals, but 75 minutes for those with depression (median recovery times; De Calheiros Velozo et al., 2023). Additionally, multilevel survival analysis revealed that negative emotions seem to dissipate more quickly than positive emotions, especially when individuals report high state resilience (Schreuder et al., 2024). The potential usefulness of survival analysis is not limited to the field of affect dynamics (Lachowicz et al., 2023; Lougheed et al., 2019; Verduyn et al., 2009), but could also be relevant in uncovering processes that span longer timeframes (e.g., how long does it take before mental health problems emerge after observing the first indicators of risk).

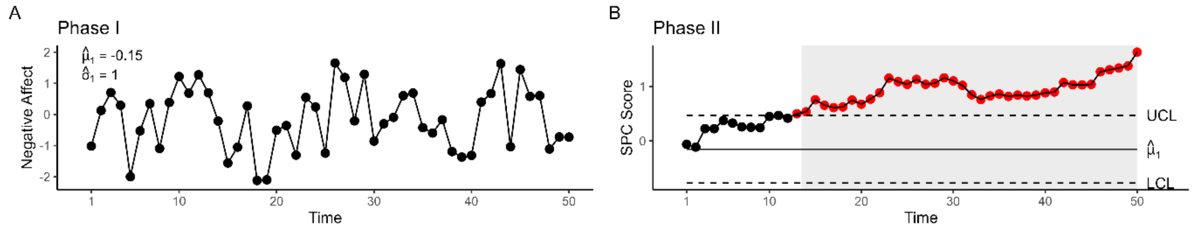
Real-time monitoring of ESM time series

The common way to use ESM data is retrospectively: After the data collection has finished, the data are analyzed at once. However, ESM can also be used prospectively. For example, do changing affect dynamics foreshadow shifts in mental health (e.g., Cramer et al., 2016; van de Leemput et al., 2014; Wichers & Groot, 2016)? Empirical work confirmed that this is indeed the case in some individuals. These findings opened up the possibility to monitor vulnerable people to detect in real time whether and when they are at risk of having a psychopathological episode (or risk for self-harm behavior or suicide attempt). Such risk signaling would allow for timely intervention. To fulfil this promise, methods are needed that can monitor and detect relevant changes in ESM data as they unfold.

Recent simulation and empirical work has shown that statistical process control (SPC; Roberts, 1959; Shewhart, 1931) procedures are particularly promising for real-time detection of changes in ESM data (Schat et al., 2023; Smit & Snippe, 2022). When applying SPC procedures in practice, one implements two distinct data collection and analysis phases. In phase I, baseline data are analyzed to capture the person's normal range for the monitored process (e.g., affective experiences), and used to compute control limits. Next, in phase II, continuously incoming ESM data are compared against these control limits. As long as the monitored scores fall within the control limits, the affective process is considered to be in control. As soon as a score falls beyond a control limit, the affective process is considered to be out of control, and this may a call to intervene. Figure 2 shows an example of an SPC procedure, where a first out-of-control score is detected at time point 13 and monitoring stops to allow for an intervention.

Research on applying SPC to ESM data has advanced in three directions. First, efforts have focused on adapting SPC to the unique characteristics of ESM data, which often violate the assumptions underlying SPC procedures (Schat et al., 2023; Schat, Tuerlinckx, De Ketelaere, et al., 2024). Specifically, EWMA, the most popular SPC procedure, imposes that the data are normally and independently distributed, whereas ESM data are often skewed and show serial dependence. Second, studies have explored alternative methods for determining phase I control limits, addressing the challenge of obtaining sufficient baseline data from individuals during periods of well-being (Schat, Tuerlinckx, Schreuder, et al., 2024; Schreuder, Kuppens, et al., 2024). Finally, SPC has been successfully applied to ESM data of 41 remitted individuals who gradually discontinued their antidepressant medication, 22 of whom experienced a recurrence of depression, demonstrating its potential to predict depression recurrence by monitoring and detecting changes in mental states (Schreuder, Schat, et al., 2024; Snippe et al., 2023).

Figure 2. Example of an SPC procedure. (A) Phase I data used to compute the upper (UCL) and lower (LCL) control limits, based on the estimated mean and standard deviation (i.e., $\hat{\mu}_1$ and $\hat{\sigma}_1$). These limits reflect person's normal range for negative affect. (B) SPC procedure applied to the incoming phase II data, where a mean change (i.e., increase in negative affect) was introduced at the start of phase II. The dashed horizontal lines represent the UCL and LCL, the solid horizontal line denotes the phase I mean $\hat{\mu}_1$. The red dots indicate the out-of-control scores, with the first detected at time point 13, after which monitoring stops to allow for an intervention.



Implementation

Any successful implementation of ESM of course depends on data collection software, allowing to send notifications to participants' smartphones and register their responses. A variety of software platforms (e.g., movisensXS, Movisens GmbH; ExpiWell; Avicenna Research; m-Path; Mestdagh et al., 2023) are currently available for implementing ESM in both research and clinical settings (for a recent overview, see Weermeijer, et al., 2025). These tools all share that they support real-time data collection, allowing for the examination of behaviors, emotions, and contexts as they occur in daily life.

The practical implementation of ESM goes beyond data collection software. In fact, given that multiple things can go wrong during data collection, such as technical problems, participants missing or ignoring notifications, tools are needed to thoroughly preprocess the collected data. Moreover, open science tools to help researchers in making their decisions and in transparently reporting on them have been developed.

Open Science practices in ESM

ESM is characterized by numerous decisions related to design, data collection protocols, item selection, and data analysis strategies (Myin-Germeys & Kuppens 2025; Peeters et al., 2025; Trull & Ebner-Priemer, 2020). This dense 'garden of forking paths' (Gelman & Loken, 2013) has contributed to significant methodological variation across ESM studies, threatening the validity and credibility of scientific findings in daily life research. Consequently, three fundamental open science practices—(pre-)registration, registered reports, and open materials—have been specifically developed for ESM to enhance transparency and reproducibility.

The first open science practice, (pre-)registration of study design and analysis plan, has been proposed to tackle issues related to selective reporting and a lack of transparency in methodological decisions (Nosek et al., 2018; Munafò et al., 2017). While many (pre-)registration templates have been developed to assist researchers in delineating their study and analytical plans before conducting the research (Bowman et al., 2020), these templates often fail to capture the full range of complex methodological decisions associated with conducting an ESM study. For this reason, Kirtley et al. (2021) created a template specifically tailored to ESM studies to assist ESM researchers at various stages of (pre-)registration. This template integrates information about specific ESM methodological choices, such as the type of sampling scheme, the description of the number and type of ESM items, protocols for briefing and debriefing participants, and considerations regarding participant exclusion, missing data, and compliance. It is also accompanied by example templates illustrating registration before data collection and post-registration of secondary data analysis.

A second practice gaining traction in ESM research is registered reports. A registered report is a type of scientific article that undergoes a two-stage peer review process (Chambers & Tzavella, 2022). In the first stage, a protocol detailing the methods and proposed analysis is peer-reviewed prior to data collection or analysis. If accepted, the study is pre-registered and provisionally accepted for publication. In the second stage, the results and discussion are peer-reviewed to ensure no unjustified deviations from the pre-registration occurred. The implementation of registered reports in ESM research has faced several challenges related to project timelines (i.e., data collection cannot commence before Stage 1 acceptance) and issues with data quality and quantity, which can severely compromise pre-registered analysis (Kirtley et al., 2022). The issues related to data quality and quantity primarily concern compliance rates (Wrzus & Neubauer, 2023), participant burden, and careless responding (Eisele et al., 2022). Careless responding, in particular, can vary between individuals, over time, and across different sample types (Hasselhorn et al., 2023; Pihlajamäki et al., 2025). Furthermore, numerous studies have shown that ESM can induce changes in behavior or experience due to reactivity effects (Eisele et al., 2023; Kirtley et al., 2025). Despite these challenges, submitting Registered Reports allows researchers to ensure that data were not accessed prior to analysis. For example, Janssens et al. (2023) published the first Registered Report on youth self-harm using preexisting data from a large-scale ESM study.

Lastly, sharing open materials and data analysis code enhances study replicability and reproducibility, as key materials—such as ESM design protocols, items, and analysis code—can be accessed, evaluated, and reused (Munafò et al., 2017). An essential aspect of an ESM study revolves around the items included in the momentary questionnaires. Since there are no universal guidelines for developing and validating ESM items, many researchers construct their items or draw from previous studies (Eisele et al., 2024). Furthermore, the lack of transparency in measurement practices raises concerns about the validity of the measures and, ultimately, the validity of empirical findings (Flake and Fried, 2020). To address this issue, the Experience

Sampling Method Item Repository (Kirtley et al., 2025) was established as an open science initiative, allowing researchers to share their ESM questionnaires. Currently, the repository contains over 3,300 items contributed by researchers from 11 different countries, including descriptions of the ESM item, the measured construct, and the study population.

Data preprocessing and quality control

Preprocessing of ESM data has received little attention for a long time and has suffered from a lack of transparency, systematicity, and user-friendly tools; however, recommendations and software have recently been developed. Viechtbauer and Constantin (2023) developed esmpack, an R package that contains multiple functions that facilitate the preparation and management of ESM/EMA Data (see Viechtbauer, 2025). Revol et al. (2024c) have proposed an ESM preprocessing framework covering five key steps, ranging from validating the study protocol execution (e.g., adherence to planned questionnaire scheduling), assessing participants' response behaviors (e.g., response patterns, careless responses), and conducting exploratory analysis through statistical summaries and visualizations of the variables of interest. Additionally, Revol et al. (2024c) developed the esmtools R package along with the ESM preprocessing gallery (https://preprocess.esmtools.com/), which provides detailed tutorials for many of the preprocessing tasks within the framework. Finally, to facilitate transparent documentation of preprocessing pipelines and reporting of data characteristics, Revol et al. (2024c) provide RMarkdown templates to increase the reproducibility and transparency of ESM studies.

Another more specific development worth mentioning is that the past years have seen an increased attention to the detection and handling of careless responses. As participants are not seldomly renumerated for their participation, they are motivated to achieve high compliance rates. But this may come at the cost of data quality, even to the point where participants are quickly entering random responses to achieve good compliance. If such careless responding could be detected, its removal could mean a direct increase of data quality. While the abovementioned framework by Revol et al. (2024c) pays some attention to careless responding, other methods have been proposed with the sole purpose of directly identifying careless responses in ESM data (e.g., Jaso et al., 2022; Ulitzsch, et al., 2025; Vogelsmeier et al., 2025). We expect these methods to improve and become more widely available more available (e.g., by being implemented in the data collection software platform) in the years to come.

Conclusion and future directions

Substantive research using ESM has gained enormous popularity over the last few decades. For many researchers and even practitioners, ESM has become one of the standard tools to obtain insight into psychological phenomena and processes as they occur in daily life. This increased use of ESM is both the result of and the driving force behind much methodological research on how to conduct ESM. As we have shown in our synthesis above, this methodological ESM research covers a variety of aspects of the research cycle: design, statistical analysis, and implementation.

Because of the wide range of research questions and the surge of technological and methodological advancements the previous years, there is an enormous diversity in how ESM is applied across different labs and practitioners. We expect that the next phase in the methodological development of ESM will be one of further consolidation of good practices. At this point, the domain lacks clear and standardized procedures that enhance comparability and replicability. To achieve such a consolidation, it will be important that methodological ESM researchers provide well-studied and user-friendly tools and invest in workshops and online training. A prime example of an important domain in which such standardized procedures are needed is the reliability and validity of the measurement instruments. It would be extremely useful if information on the reliability and validity of single items and multiple item scales for a given target audience and research goal could be made available as part of the open online, item repository (Kirtley et al., 2025). In a next step, such information could also be integrated into the various software platforms for data collection. Similarly, integrating the preprocessing toolbox with the data collection software would signal a step forward for ESM users.

Consolidating ESM research will also require an increased global collaboration between research labs. Data collection is a costly enterprise and sharing the investment across labs may be helpful. A first good step in that direction is the EMOTE database (Kalokerinos et al., 2025) that curates and unlocks a large number of ESM data sets. Further global collaboration will also allow to use ESM to investigate cross-cultural variability in daily life dynamics.

Conducting larger studies is one way to advance ESM research. However, moving in the opposite direction, focusing squarely on the individual, may also be valuable. ESM can serve as a powerful tool for idiographic approaches by incorporating individualized questions, aligning with personalized models of psychopathology (Wright & Woods, 2020). Naturally, using such individualized questionnaires, some of the methods discussed in this paper may be less relevant (e.g., model-based reliability measures, survival analysis). However, other methods, such as online monitoring using statistical process control, remain fully applicable.

Because technology is constantly improving, we expect that new developments will have an impact on ESM as well. For example, many people nowadays wear a smartwatch or other wearables that constantly monitor bodily activity. Information read out by wearables can be very valuable in the context of ESM studies as it may provide insight in contextual information (Niemeijer & Kuppens, 2024). With further improvement , such information could enrich ESM data (e.g., informing in what type of contexts particular thoughts or feelings are (mal)adaptive) and ESM designs (e.g., informing when participants may be able to fill in a questionnaire).

Readers might also wonder about the impact of machine learning methods, since such methods have gained increased traction in psychological research. Although the ESM field is lagging behind, enthusiasm for machine learning applications is growing indeed (see e.g., Wang et al., 2025). The most prominent application involves employing supervised machine learning models to predict or forecast (clinically) relevant outcomes (e.g., relapse in depression), frequently incorporating mobile sensing and physiological sensors data (for example, see Niemeijer et al., 2022; Strakeljahn et al., 2024). Machine learning ideas are also employed for model selection, primarily through cross-validation. For example, Dejonckheere et al. (2019) use cross-validation to benchmark the usefulness of diverse affect dynamic measures derived from ESM in predicting psychological well-being. A few methodological studies have investigated how to optimally implement cross-validation in ESM research (Bulteel et al., 2018; Liu & Zhou, 2024). Moreover, Natural Language Processing, a branch of ML, has been used to analyze text and speech data (for example, see Weidman et al., 2020) and there are some first applications of Large Language Models (Zhao et al., 2024).

In conclusion, this paper highlighted the ongoing development of methodologies underpinning ESM. The growing adoption of ESM by substantive researchers as well as clinicians coupled with technological advancements has opened up an array of significant research questions waiting to be explored. These range from more practical and applied topics (e.g., personalization of questions and scheduling based on participants' profiles) to complex methodological challenges (e.g., advancing analytical methods, linking self-reported and passively collected data, distinguishing signal from noise). This range of scientific opportunities will make this field an active and exciting domain of research in the years ahead.

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

Albers, C. J., & Bringmann, L. F. (2020). Inspecting gradual and abrupt changes in emotion dynamics with the time-varying change point autoregressive model. *European Journal of Psychological Assessment, 36*(3), 492–499. https://doi.org/10.1027/1015-5759/a000589

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (Eds.). (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Arend, M. G., & Schäfer, T. (2019). Statistical power in two-level models: A tutorial based on Monte Carlo simulation. *Psychological Methods*, 24(1), 1-19. DOI: 10.1037/met0000195

Ariens, S., Ceulemans, E., & Adolf, J. K. (2020). Time series analysis of intensive longitudinal data in psychosomatic research: A methodological overview. *Journal of Psychosomatic Research, 137*, 110191, 1-15. https://doi.org/10.1016/j.jpsychores.2020.110191

Ariens, S., Adolf, J. K., & Ceulemans, E. (2023). One does not simply correct for serial dependence. *Structural Equation Modeling: A Multidisciplinary Journal, 30*, 807-821. https://doi.org/10.1080/10705511.2023.2173203

Bell, B. A., Morgan, G. B., Schoeneberger, J. A., Kromrey, J. D., & Ferron, J. M. (2014). How low can you go? An investigation of the influence of sample size and model complexity on point and interval estimates in two-level linear models. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 10(1)*, 1–11. https://doi.org/10.1027/1614-2241/a000062

Barta, W. D., Tennen, H., & Litt, M. D. (2012). Measurement reactivity in diary research. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 108–123). The Guilford Press.

Bonsall, M. B., Wallace-Hadrill, S. M. A., Geddes, J. R., Goodwin, G. M., & Holmes, E. A. (2012). Nonlinear time-series approaches in characterizing mood stability and mood instability in bipolar disorder. *Proceedings of the Royal Society B, 279*, 916–924. https://doi.org/10.1098/rspb.2011.1246

Boker, S. M., Molenaar, P. C. M., & Nesselroade, J. R. (2009). Issues in intraindividual variability: Individual differences in equilibria and dynamics over multiple time scales. *Psychology and Aging*, *24*(4), 858–862. http://dx.doi.org/10.1037/a0017912

Bolger, N., & Laurenceau, J. P. (2013). *Intensive longitudinal methods: An introduction to diary and experience sampling research*. Guilford press.

Borsboom, D., & Cramer, A. O. (2013). Network analysis: an integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology, 9(1),* 91-121. https://doi.org/10.1146/annurev-clinpsy-050212-185608

Bowman, S. D., DeHaven, A. C., Errington, T. M., Hardwicke, T. E., Mellor, D. T., Nosek, B. A., & Soderberg, C. K. (2020, January 22). OSF Prereg Template. https://doi.org/10.31222/osf.io/epgjd

Bringmann L. F., Hamaker E. L., Vigo D. E., Aubert A., Borsboom D., Tuerlinckx F. (2017). Changing dynamics: Time-varying autoregressive models using generalized additive modeling. *Psychological Methods, 22*, 409–425. https://doi.org/10.1037/met0000085

Bringmann, L. F., Ariens, S., Ernst, A., Snippe, E., & Ceulemans, E. (2024). Changing networks: Moderated idiographic psychological networks. *Advances in Psychology, 2,* e658296, 1-46. https://doi.org/10.56296/aip00014

Bringmann, L. F., Albers, C., Bockting, C., Borsboom, D., Ceulemans, E., Cramer, A., Epskamp, S., Eronen, M. I., Hamaker, E., Kuppens, P., Lutz, W., McNally, R. J., Molenaar, P., Tio, P., Voelkle, M. C., & Wichers, M. (2022). Psychopathological networks: Theory, methods and practice. *Behaviour Research and Therapy, 149*, 104011, 1-15. https://doi.org/10.1016/j.brat.2021.104011

Brose, A., Schmiedek, F., Gerstorf, D., & Voelkle, M. C. (2020). The measurement of within-person affect variation. *Emotion*, *20*(4), 677–699. https://doi.org/10.1037/emo0000583.supp

Bulteel, K., Mestdagh, M., Tuerlinckx, F., & Ceulemans, E. (2018). VAR(1) based models do not always outpredict AR(1) models in typical psychological applications. *Psychological Methods*, *23*, 740–756. https://doi.org/10.1037/met0000178

Cabrieto, J., Adolf, J. K., Tuerlinckx, F., Kuppens, P., & Ceulemans, E. (2019). An objective, comprehensive and flexible statistical framework for detecting early warning signs of mental health problems. *Psychotherapy and Psychosomatics*, *88*(3), 184–186. https://doi.org/10.1159/000494356

Cabrieto, J., Tuerlinckx, F., Kuppens, P., Wilhelm, F. H., Liedlgruber, M., & Ceulemans, E. (2018). Capturing correlation changes by applying kernel change point detection on the running correlations. *Information Sciences*, *447*, 117–139. https://doi.org/10.1016/j.ins.2018.03.010

Castro-Alvarez, S., Bringmann, L., Back, J., & Liu, S. (2024). The many reliabilities of psychological dynamics: An overview of statistical approaches to estimate the internal consistency reliability of intensive longitudinal data. OSF. https://doi.org/10.31234/osf.io/qyk2r

Chambers, C. D., & Tzavella, L. (2022). The past, present and future of Registered Reports. *Nature Human Behaviour*, *6*(1), 29-42.

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. 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*, *0*(0), 10731911241266286. https://doi.org/10.1177/10731911241266286

Cloos, L., Vanpaemel, W., Ceulemans, E., & Kuppens, P. (2024). *Remembering Affect Between Moments: Assessing Peak-End Effects In Continuous Affect Measures*. OSF. https://osf.io/nhmta

Coppersmith, D. D., Ryan, O., Fortgang, R. G., Millner, A. J., Kleiman, E. M., & Nock, M. K. (2023). Mapping the timescale of suicidal thinking. *Proceedings of the National Academy of Sciences, 120(17)*, e2215434120. https://doi.org/10.1073/pnas.2215434120

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

Dalege, J., Borsboom, D., Van Harreveld, F., Van den Berg, H., Conner, M., & Van der Maas, H. L. (2016). Toward a formalized account of attitudes: The Causal Attitude Network (CAN) model. *Psychological Review, 123(1),* 2-22. https://doi.org/10.1037/a0039802

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

De Angel, V., Lewis, S., White, K., Oetzmann, C., Leightley, D., Oprea, E., Lavelle, G., Matcham, F., Pace, A., Mohr, D. C., Dobson, R., & Hotopf, M. (2022). Digital health tools for the passive monitoring of depression: A systematic review of methods. *Npj Digital Medicine*, *5*(1), 1–14. https://doi.org/10.1038/s41746-021-00548-8

De Calheiros Velozo, J., Lafit, G., Viechtbauer, W., van Amelsvoort, T., Schruers, K., Marcelis, M., Goossens, L., Simons, C. J. P., Delespaul, P., Claes, S., Myin-Germeys, I., & Vaessen, T. (2023). Delayed affective recovery to daily-life stressors signals a risk for depression. *Journal of Affective Disorders*, *320*, 499–506. https://doi.org/10.1016/j.jad.2022.09.136

de Haan-Rietdijk, S., Voelkle, M. C., Keijsers, L., & Hamaker, E. L. (2017). Discrete- vs. Continuous-Time Modeling of Unequally Spaced Experience Sampling Method Data. *Frontiers in Psychology*, *8*(OCT), 1849. https://doi.org/10.3389/fpsyg.2017.01849

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*, Springer International Publishing. DOI: 10.1007/978-3-030-82965-0_7

Dejonckheere, E., Demeyer, F., Geusens, B., Piot, M., Tuerlinckx, F., Verdonck, S., & Mestdagh, M. (2022). Assessing the reliability of single-item momentary affective measurements in experience sampling. *Psychological Assessment*, *34*(12), 1138–1154. https://doi.org/10.1037/pas0001178

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., Penne, I., Briels, L. & Mestdagh, M. (2023). For better or for worse? Visualizing previous intensity levels improves emotion (dynamic) measurement in experience sampling. *Psychological Assessment, 36,* 215 – 234. DOI: 10.1037/pas0001296.

Delespaul P. (1995). *Assessing schizophrenia in daily life: The experience sampling method*. Maastricht, The Netherlands: University of Limburg.

Del Giudice M., & Gangestad, S.W. (2021). A traveler's guide to the multiverse: Promises, pitfalls, and a framework for the evaluation of analytic decisions. *Advances in Methods and Practices in Psychological Science, 4(1)*, 2515245920954925. doi:10.1177/2515245920954925

Doherty, K., Balaskas, A., & Doherty, G. (2020). The design of ecological momentary assessment technologies. *Interacting with Computers*, *32*(1), 257-278.

Driver, C. C., Oud, J. H. L., & Voelkle, M. C. (2017). Continuous time structural equation modeling with R package ctsem. *Journal of Statistical Software, 77(5)*, 1–35. https://doi.org/10.18637/jss.v077.i05

Eisele, G., Hiekkaranta, A. P., Kunkels, Y. K., Rot, M. aan het, Ballegooijen, W. van, Bartels, S. L., Bastiaansen, J. A., Beymer, P. N., Bylsma, L., Carpenter, R., Ellison, W. D., Fisher, A. J., Forkmann, P. D. T., Frumkin, M., Fulford, D., Naragon-Gainey, K., Greene, T., Heininga, V. E., Jones, A., ... Kirtley, O. (2024). *ESM-Q: A Consensus-Based Quality Assessment Tool for Experience Sampling Method Items*. OSF. https://doi.org/10.31234/osf.io/sjynv

Eisele, G., Vachon, H., Lafit, G., Tuyaerts, D., Houben, M., Kuppens, P., ... & 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. https://doi.org/10.1037/pas0001177

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

Elmer, T., Van Duijn, M. A. J., Ram, N., & Bringmann, L. (2023). Modeling categorical time-to-event data: The example of social interaction dynamics captured with event-contingent experience sampling methods. *Psychological Methods*. https://doi.org/doi.org/10.1037/met0000598

Flake, J. K., & Fried, E. I. (2020). Measurement schmeasurement: Questionable measurement practices and how to avoid them. *Advances in Methods and Practices in Psychological Science*, *3*(4), 456-465.

Frijda, N. H. (2007). *The laws of emotion*. Routledge.

Friston, K., Kilner, J., & Harrison, L. (2006). A free energy principle for the brain. *Journal of Physiology-Paris, 100*(1-3), 70-87. https://doi.org/10.1016/j.jphysparis.2006.10.001

Fritz, J., Piccirillo, M. L., Cohen, Z. D., Frumkin, M., Kirtley, O., Moeller, J., Neubauer, A.B., Noriss, L.A., Schuurman, N.K., Snippe, E., & Bringmann, L. F. (2024). So you want to do ESM? 10 essential topics for implementing the experience-sampling method. *Advances in Methods and Practices in Psychological Science, 7(3)*, doi:10.1177/25152459241267912

Gabriel, A. S., Podsakoff, N. P., Beal, D. J., Scott, B. A., Sonnentag, S., Trougakos, J. P., & Butts, M. M. (2019). Experience sampling methods: A discussion of critical trends and considerations for scholarly advancement. *Organizational Research Methods*, 22(4), 969-1006.

Gelman, A., & Loken, E. (2013). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time. Unpublished manuscript, Department of Statistics, Columbia University. http://stat.columbia.edu/~gelman/research/unpublished/forking.pdf

Gray, E. K., & Watson, D. (2007). Assessing positive and negative affect via self-report. In J. A. Coan & J. J. B. Allen (Eds.), *Handbook of emotion elicitation and assessment* (pp. 171–183). Oxford University Press.

Hall, M., Scherner, P. V., Kreidel, Y., & Rubel, J. A. (2021). A systematic review of momentary assessment designs for mood and anxiety symptoms. *Frontiers in Psychology*, *12*, 1–14. https://doi.org/10.3389/fpsyg.2021.642044

Hamaker, E. L., & Wichers, M. (2017). No time like the present: Discovering the hidden dynamics in intensive longitudinal data. *Current Directions in Psychological Science*, *26*(1), 10–15. https://doi.org/10.1177/0963721416666518

Hamaker, E. L., Grasman, R. P. P. P., & Kamphuis, J. H. (2010). Regime-switching models to study psychological processes. In P. C. M. Molenaar & K. M. Newell (Eds.), *Individual pathways of change: Statistical models for analyzing learning and development* (pp. 155–168). American Psychological Association. https://doi.org/10.1037/12140-009

Hamaker, E. L., Ceulemans, E., Grasman, R. P., & Tuerlinckx, F. (2015). Modeling affect dynamics: State of the art and future challenges. *Emotion Review, 7(4)*, 316-322.

Hamaker, E.L., Asparouhov, T., Brose, A., Schmiedek, F., Muthén, B. (2018). At the frontiers of modeling intensive longitudinal data: Dynamic Structural Equation Models for the affective measurements from the COGITO study. *Multivariate Behavioral Research, 53(6)*, 820-841. doi: 10.1080/00273171.2018.1446819.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica, 57*(2), 357-384.

Haslbeck, J., Ryan, O., & Dablander, F. (2023). Multimodality and skewness in emotion time series. *Emotion, 23*(8), 2117–2141. https://doi.org/10.1037/emo0001218

Hasselhorn, K., Ottenstein, C., & Lischetzke, T. (2021). The effects of assessment intensity on participant burden, compliance, within-person variance, and within-person relationships in ambulatory assessment. *Behavior Research Methods*. https://doi.org/10.3758/s13428-021-01683-6

Haney, A. M., Fleming, M. N., Wycoff, A. M., Griffin, S. A., & Trull, T. J. (2023). Measuring affect in daily life: A multilevel psychometric evaluation of the PANAS-X across four ecological momentary assessment samples. *Psychological assessment, 35(6),* 469–483. https://doi.org/10.1037/pas0001231

Hong, J. I. (2024). Designing for Privacy in Mobile Sensing Systems. In *Digital Phenotyping and Mobile Sensing: New Developments in Psychoinformatics*. Springer International Publishing.

Janssens, J. J., Myin-Germeys, I., Lafit, G., Achterhof, R., Hagemann, N., Hermans, K. S. F. M., Hiekkaranta, A. P., Lecei, A., & Kirtley, O. J. (2023). Lifetime and current self-harm thoughts and behaviors and their relationship to parent and peer attachment. *Crisis: The Journal of Crisis Intervention and Suicide Prevention, 44(5)*, 424–432. https://doi.org/10.1027/0227-5910/a000878

Jaso, B. A., Kraus, N. I., & Heller, A. S. (2022). Identification of careless responding in ecological momentary assessment research: From posthoc analyses to real-time data monitoring. *Psychological Methods, 27(6),* 958–981.

[ANONYMIZED] (2025). Evaluating the feasibility of episode-contingent experience-sampling designs. *Psychological Assessment*, in press.

Kestens, Y., & Kingsbury, C. (2024). In the right place at the right time: Leveraging geographic Ecological Momentary Assessment to analyze contexts and health. In Kuhn, S. (Ed.), *Environmental Neuroscience* (pp. 517-532). Cham: Springer Nature Switzerland.

Kirtley, O. J., Eisele, G., Kunkels, Y. K., Hiekkaranta, A. P., Van Heck, L., Pihlajamäki, M., ... Myin-Germeys, I. (2025, February 26). The Experience Sampling Method (ESM) Item Repository. https://doi.org/10.17605/OSF.IO/KG376

Kirtley, O. J., Janssens, J. J., & Kaurin, A. (2022). Open science in suicide research is open for business [Editorial]. *Crisis: The Journal of Crisis Intervention and Suicide Prevention, 43*(5), 355–360. https://doi.org/10.1027/0227-5910/a000859

Kirtley, O. J., Lafit, G., Achterhof, R., Hiekkaranta, A. P., & Myin-Germeys, I. (2021). Making the black box transparent: A template and tutorial for registration of studies using experience-sampling methods. *Advances in Methods and Practices in Psychological Science*, *4*(1). doi:10.1177/2515245920924686

Kirtley, O. J., Sohier, B., Simsa, B., Achterhof, R., Myin-Germeys, I., & Lafit, G. (2025). Reactivity to experience sampling among adolescents with and without a lifetime or current history of self-harm thoughts or behaviours. *Psychological Assessment*, in press. https://doi.org/10.31234/osf.io/tvw68

Kalokerinos, E. K., Russo-Batterham, D., Koval, P., Moeck, E. K., Grewal, K. K., Greenaway, K. H., Shrestha, K. M., Garrett, P., Michalewicz, A., Garber, J., & Kuppens, P. (2025). The EMOTE Database: An open, searchable database of experience sampling data mapping everyday life. *Manuscript in preparation*.

Kuppens, P., Allen, N. B., & Sheeber, L. B. (2010). Emotional inertia and psychological maladjustment. *Psychological science, 21(7)*, 984-991.

Lachowicz, A., Houben, M., Vaessen, T., Lafit, G., Achterhof, R., Akcaoglu, Z., Bamps, E., Hagemann, N., Hiekkaranta, A. P., Hermans, K., Janssens, J., Lecei, A., Kirtley, O., & Myin-Germeys, I. (2023). *Delayed Affective Recovery from Stress is Linked to Current but not Future Subclinical Anxiety Symptoms in Youth: An Experience Sampling Study*. OSF. https://doi.org/10.31234/osf.io/2fxyb

Lafit, G., Adolf, J. K., Dejonckheere, E., Myin-Germeys, I., Viechtbauer, W., & Ceulemans, E. (2021). Selection of the number of participants in intensive longitudinal studies: A user-friendly shiny app and tutorial for performing power analysis in multilevel regression models that account for temporal dependencies. *Advances in Methods and Practices in Psychological Science*, 4(1), 2515245920978738.

Lafit, G., Artner, R., & Ceulemans, E. (2024). Enabling analytical power calculations for multilevel models with autocorrelated errors through deriving and approximating the precision matrix. *Behavior Research Methods*, 56(7), 8105-8131.

Lafit, G., Revol, J., Cloos, L., Kuppens, P., & Ceulemans, E. (2024). The effect of different construct operationalizations, study duration, and preprocessing choices on power-based sample size recommendations in intensive longitudinal research. *Assessment*, 10731911241286868.

Lafit, G., Revol, J., Cloos, L., Kuppens, P., & Ceulemans, E. (2024). The effect of different construct operationalizations, study duration, and preprocessing choices on power-based sample size recommendations in intensive longitudinal research. *Assessment, 32(2)*, 206-223. https://doi.org/10.1177/10731911241286868

Lafit, G., Sels, L., Adolf, J. K., Loeys, T., & Ceulemans, E. (2022). PowerLAPIM: An application to conduct power analysis for linear and quadratic longitudinal actor–partner interdependence models in intensive longitudinal dyadic designs. *Journal of Social and Personal Relationships*, 39(10), 3085-3115.

Lakens, D. (2022). Sample size justification. *Collabra: psychology, 8(1),* 33267.

Lane S. P., Hennes E. P. (2018). Power struggles: Estimating sample size for multilevel relationships research. *Journal of Social and Personal Relationships*, 35(1), 7–31.

Langener, A. M., Siepe, B. S., Elsherif, M., Niemeijer, K., Andresen, P. K., Akre, S., Bringmann, L. F., Cohen, Z. D., Choukas, N. R., Drexl, K., Fassi, L., Green, J., Hoffmann, T., Jagesar, R. R., Kas, M. J. H., Kurten, S., Schoedel, R., Stulp, G., Turner, G., & Jacobson, N. C. (2024). A template and tutorial for preregistering studies using passive smartphone measures. *Behavior Research Methods*. https://doi.org/10.3758/s13428-024-02474-5

Langener, A. M., Stulp, G., Jacobson, N. C., Costanzo, A., Jagesar, R. R., Kas, M. J., & Bringmann, L. F. (2024). It's All About Timing: Exploring Different Temporal Resolutions for Analyzing Digital-Phenotyping Data. *Advances in Methods and Practices in Psychological Science*, *7*(1), 25152459231202677. https://doi.org/10.1177/25152459231202677

Lathia, N., Rachuri, K. K., Mascolo, C., & Rentfrow, P. J. (2013). Contextual dissonance: Design bias in sensor-based experience sampling methods. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing* (pp. 183-192).

Lewis, M. D. (2005). Bridging emotion theory and neurobiology through dynamic systems modeling. *Behavioral and Brain Sciences, 28,* 169-245. https://doi.org/10.1017/S0140525X0500004X

Liu, S., & Zhou, D. J. (2024). Using cross-validation methods to select time series models: Promises and pitfalls. *British Journal of Mathematical and Statistical Psychology*, *77*(2), 337–355. https://doi.org/10.1111/bmsp.12330

Loossens, T., Mestdagh, M., Dejonckheere, E., Kuppens, P., Tuerlinckx, F., & Verdonck, S. (2020). The Affective Ising Model: A computational account of human affect dynamics. *PLoS Computational Biology, 16*, e1007860. https://doi.org/10.1371/journal.pcbi.1007860

Lougheed, J. P., Benson, L., Cole, P. M., & Ram, N. (2019). Multilevel survival analysis: Studying the timing of children's recurring behaviors. *Developmental Psychology*, *55*(1), 53–65. https://doi.org/10.1037/dev0000619

Maas, C. J., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology, 1(3)*, 86-92. https://doi.org/10.1027/1614-2241.1.3.86

Mehl, M. R., Eid, M., Wrzus, C., Harari, G. M., & Ebner-Priemer, U. (Eds.). (2024). *Mobile Sensing in Psychology: Methods and Applications*. Guilford Press.

Mestdagh, M., & Dejonckheere, E. (2021). Ambulatory assessment in psychopathology research: Current achievements and future ambitions. *Current Opinion in Psychology*, *41*, 1–8. https://doi.org/10.1016/j.copsyc.2021.01.004

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

Moerbeek, M. (2008). Powerful and cost-efficient designs for longitudinal intervention studies with two treatment groups. *Journal of Educational and Behavioral Statistics*, *33*(1), 41–61. https://doi.org/10.3102/1076998607302630

Movisens GmbH. (n.d.). movisensXS: Experience Sampling Platform. Retrieved from https://www.movisens.com/en/products/movisensxs/.

Munafò, M. R., Nosek, B. A., Bishop, D. V., Button, K. S., Chambers, C. D., Percie du Sert, N., ... & Ioannidis, J. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*(1), 1-9.

Myin-Germeys, I., & Kuppens, P. (Eds.). (2025) *The open handbook of experience sampling methodology: A step-by-step guide to designing, conducting, and analyzing ESM studies (3rd ed.)*. Leuven University Press.

Niemeijer, K., Mestdagh, M., & Kuppens, P. (2022). Tracking subjective sleep quality and mood with mobile sensing: multiverse study. *Journal of Medical Internet Research*, *24*(3), e25643. https://doi.org/10.2196/25643

Niemeijer, K., & Kuppens, P. (2024). Emotion detection with mobile sensing. In M. R. Mehl, M. Eid, C. Wrzus, G. M. Harari, & U. Ebner-Priemer (Eds.), *Mobile Sensing in Psychology: Methods and Applications* (pp. 561–580). Guilford Press.

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, *115*(11), 2600-2606.

Onnela, J.-P. (2021). Opportunities and challenges in the collection and analysis of digital phenotyping data. *Neuropsychopharmacology*, *46*(1), Article 1. https://doi.org/10.1038/s41386-020-0771-3

Osotsi, A., Oravecz, Z., Li, Q., Smyth, J., & Brick, T. R. (2020). Individualized modeling to distinguish between high and low arousal states using physiological data. *Journal of Healthcare Informatics Research, 4*, 91–109.

Peeters, L., Van Den Noortgate, W., Blanchard, M. A., Eisele, G., Kirtley, O., Artner, R., & Lafit, G. (2025). Mapping Methodological Variation in ESM Research from Design to Data Analysis: A Systematic Review [Stage 1 Registered Report]. In principle acceptance of Version 3 by Peer Community in Registered Reports. Registration link: https://doi.org/10.17605/OSF.IO/ZTVN3 Stage 1 recommendation and peer review history: https://rr.peercommunityin.org/articles/rec?id=894

Pejovic, V., Lathia, N., Mascolo, C., & Musolesi, M. (2016). Mobile-based experience sampling for behaviour research. In M. Tkalčič, B. De Carolis, M. de Gemmis, A. Odić, & A. Košir (Eds.), Emotions and personality in personalized services: Models, evaluation and applications (pp. 141–161). Springer International Publishing/Springer Nature. https://doi.org/10.1007/978-3-319-31413-6_8

Pihlajamäki, M., Eisele, G., Peeters, L., Kirtley, O. J., Myin-Germeys, I., & Lafit, G. (2025, February 21). An investigation of the temporal dynamics of careless responding across different populations in experience sampling data. Preprint. https://doi.org/10.31234/osf.io/6nrvp_v1

Ram, N., Brinberg, M., Pincus, A. L., & Conroy, D. E. (2017). The questionable ecological validity of ecological momentary assessment: Considerations for design and analysis. *Research in Human Development*, *14*(3), 253–270. https://doi.org/10.1080/15427609.2017.1340052

Revol, J., Lafit, G., Kirtley, O., & Ceulemans, E. (2024a). *Cost-effective ESM studies: Integrating budget constraints into sample size decisions*. OSF. https://osf.io/b4ev5

Revol, J., Lafit, G., & Ceulemans, E. (2024b). A new sample-size planning approach for person-specific VAR (1) studies: Predictive accuracy analysis. *Behavior Research Methods*, 1-16.

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

Revol, J., Carlier, C., Lafit, G., Verhees, M., Sels, L., & Ceulemans, E. (2024c). 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), 25152459241256609. 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

Roberts, S. W. (1959). Control chart tests based on geometric moving averages. *Technometrics*, *1*(3), 239–250. https://doi.org/10.2307/1266443

Rush, J., & Hofer, S. M. (2014). Differences in within- and between-person factor structure of positive and negative affect: Analysis of two intensive measurement studies using multilevel structural equation modeling. *Psychological Assessment*, *26*(2), 462–473. https://doi.org/10.1037/a0035666

Ryan, O., Kuiper, R. M., Hamaker, E. L., & Ryan, O. (2018). A continuous time approach to intensive longitudinal data: What, why, and how? In *Continuous time modeling in the behavioral and related sciences* (pp. 27–54). Springer.

Schat, E., Tuerlinckx, F., De Ketelaere, B., & Ceulemans, E. (2024). Real-time detection of mean and variance changes in experience sampling data: A comparison of existing and novel statistical process control approaches. *Behavior Research Methods*, *56*, 1459–1475. https://doi.org/10.3758/s13428-023-02103-7

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

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

Scherbaum, C. A., & Ferreter, J. M. (2009). Estimating statistical power and required sample sizes for organizational research using multilevel modeling. *Organizational Research Methods, 12(2)*, 347–367. https://doi.org/10.1177/1094428107308906

Scherer, K. R. (2000). Emotions as episodes of subsystem synchronization driven by nonlinear appraisal processes. In M. Lewis & I. Granic (Eds.), *Emotion, Development, and Self-Organization: Dynamic Systems Approaches to Emotional Development (Cambridge Studies in Social and Emotional Development)*. Cambridge University Press. https://doi.org/10.1017/CBO9780511527883.005

Schimmack, U. (2001). Pleasure, displeasure, and mixed feelings: Are semantic opposites mutually exclusive? *Cognition and Emotion, 15*, 81–97. https://doi.org/10.1080/02699930126097

Schreuder, M. J., Groen, R. N., Wigman, J. T., Hartman, C. A., & Wichers, M. (2020). Measuring psychopathology as it unfolds in daily life: Addressing key assumptions of intensive longitudinal methods in the TRAILS TRANS-ID study. *BMC Psychiatry*, 20, 1-14.

Schreuder, M. J., Kuppens, P., Schat, E., de Jonge, P., Hartman, C. A., & Ceulemans, E. (2024). Warning signals for mental health problems in at-risk young adults may be informed by momentary emotions reported by the general population: A novel application of the principles of statistical process control. *Mental Health Science*, *2*(3), e27. https://doi.org/10.1002/mhs2.72

Schreuder, M. J., Schat, E., Smit, A. C., Snippe, E., & Ceulemans, E. (2024). Monitoring emotional intensity and variability in real-time to forecast depression recurrence in remitted adults. *Journal of Consulting and Clinical Psychology*, *92*(8), 505–516. https://doi.org/10.1037/ccp0000871

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*, *0*(0), 1–10. https://doi.org/10.1080/02699931.2024.2423706

Schultzberg, M., & Muthén, B. (2018). Number of subjects and time points needed for multilevel time-series analysis: A simulation study of dynamic structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(4), 495-515.

Schuurman, N. K. (2023). *A "Within/Between Problem" Primer: About (Not) Separating Within-Person Variance and Between-Person Variance in Psychology.* PsyArXiv. https://doi.org/10.31234/osf.io/7zgkx

Schuurman, N. K., Houtveen, J. H., & Hamaker, E. L. (2015). Incorporating measurement error in n = 1 psychological autoregressive modeling. *Frontiers in Psychology*, *6*, 1038. https://doi.org/10.3389/fpsyg.2015.01038

Schuurman, N. K., & Hamaker, E. L. (2019). Measurement error and person-specific reliability in multilevel autoregressive modeling. *Psychological Methods, 24(1),* 70–91. https://doi.org/10.1037/met0000188

Scollon, C. N., Kim-Prieto, C., & Diener, E. (2003). Experience Sampling: Promises and pitfalls, strengths and weaknesses. *Journal of Happiness Studies: An Interdisciplinary Forum on Subjective Well-Being*, *4*(1), 5–34. https://doi.org/10.1023/A:1023605205115

Seizer, L., Schiepek, G., Cornelissen, G., & Löchner, J. (2024). A primer on sampling rates of ambulatory assessments.Psychological Methods. Advance online publication. https://doi.org/10.1037/met0000656

Shewhart, W. A. (1931). *Economic control of quality of manufactured product*. Macmillan And Co Ltd.

Silvia, P. J., Kwapil, T. R., Eddington, K. M., & Brown, L. H. (2013). Missed beeps and missing data: Dispositional and situational predictors of nonresponse in experience sampling research. *Social Science Computer Review, 31(4),* 471-481. https://doi.org/10.1177/0894439313479902

Smit, A. C., & Snippe, E. (2022). Real-time monitoring of increases in restlessness to assess idiographic risk of recurrence of depressive symptoms. *Psychological Medicine*, 1–10. https://doi.org/10.1017/S0033291722002069

Snippe, E., Smit, A. C., Kuppens, P., Burger, H., & Ceulemans, E. (2023). Recurrence of depression can be foreseen by monitoring mental states with statistical process control. *Journal of Psychopathology and Clinical Science*, *132*(2), 145–155. https://doi.org/10.1037/abn0000812

Sonnemans, J., & Frijda, N. H. (1994). The structure of subjective emotional intensity. *Cognition and Emotion*, *8*(4), 329–350. https://doi.org/10.1080/02699939408408945

Spoto, A., Nucci, M., Prunetti, E., & Vicovaro, M. (2025). Improving content validity evaluation of assessment instruments through formal content validity analysis. *Psychological Methods, 30(2),* 203–222. https://doi.org/10.1037/met0000545

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, *11*(5), 702–712. https://doi.org/10.1177/1745691616658637

Stone, A. A., Schneider, S., & Smyth, J. M. (2023). Evaluation of pressing issues in ecological momentary assessment. *Annual Review of Clinical Psychology, 19(1)*, 107-131. https://doi.org/10.1146/annurev-clinpsy-080921-083128

Strakeljahn, F., Lincoln, T., Krkovic, K., & Schlier, B. (2024). Predicting the onset of psychotic experiences in daily life with the use of ambulatory sensor data – A proof-of-concept study. *Schizophrenia Research*, *267*, 349–355. https://doi.org/10.1016/j.schres.2024.03.049

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

Trull, T. J., & Ebner-Priemer, U. W. (2020). Ambulatory assessment in psychopathology research: A review of recommended reporting guidelines and current practices. *Journal of Abnormal Psychology*, *129(1)*, 56.

Ulitzsch, E., Nestler, S., Lüdtke, O., & Nagy, G. (2024). A screen-time-based mixture model for identifying and monitoring careless and insufficient effort responding in ecological momentary assessment data. *Psychological Methods.* Advance online publication. https://doi.org/10.1037/met0000636

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

van Berkel, N., Goncalves, J., Lovén, L., Ferreira, D., Hosio, S., & Kostakos, V. (2019). Effect of experience sampling schedules on response rate and recall accuracy of objective self-reports. *International Journal of Human-Computer Studies, 125*, 118-128.

Vanhasbroeck, N., Niemeijer, K., & Tuerlinckx, F. (2024). Nonlinearity in affect dynamics persists after accounting for the valence of daily-life events. *Emotion, 24*(5), 1206–1223. https://doi.org/10.1037/emo0001336

Verduyn, P., Delvaux, E., Van Coillie, H., Tuerlinckx, F., & Mechelen, I. V. (2009). Predicting the Duration of Emotional Experience: Two Experience Sampling Studies. *Emotion*, *9*(1), 83–91. https://doi.org/10.1037/a0014610

Viechtbauer, W. (2025). Structuring, checking, and preparing the data. In Myin-Germeys & Kuppens (Eds.), *The Open Handbook of Experience Sampling Methodology: A step-by-step guide to designing, conducting, and analyzing ESM studies* (pp. 137–152). Leuven University Press.

Viechtbauer, W., & Constantin, M. (2023). *esmpack: Functions that facilitate preparation and management of ESM/EMA data*. R package version 0.1-20.

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

Vogelsmeier, L. V. D. E., Cloos, L., Kuppens, P., & Ceulemans, E. (2023). Evaluating dynamics in affect structure with latent Markov factor analysis. *Emotion, 24(3),* 782. https://doi.org/10.1037/emo0001307

Vogelsmeier, L. V., Uglanova, I., Rein, M. T., & Ulitzsch, E. (2025). Investigating dynamics in attentive and inattentive responding together with their contextual correlates using a novel mixture IRT model for intensive longitudinal data. *British Journal of Mathematical and Statistical Psychology*, in press. https://doi.org/10.1111/bmsp.12373

Walls, T. A., & Schafer, J. L. (Eds.). (2006). *Models for intensive longitudinal data*. Oxford University Press.

Wang, P., Liu, A., & Sun, X. (2025). Integrating emotion dynamics in mental health: A trimodal framework combining ecological momentary assessment, physiological measurements, and speech emotion recognition. *Interdisciplinary Medicine*, e20240095. https://doi.org/10.1002/INMD.20240095

Watson, D., & Clark, L.A. (1999). *The PANAS-X: Manual for the Positive and Negative Affect Schedule—Expanded Form*. University of Iowa.

Weermeijer, J., Kiekens, G., & Wampers, M. (2025). Experience Sampling Platforms. In *The Open Handbook of Experience Sampling Methods. A step-by-step guide to designing, conducting, and analyzing ESM studies* (3rd ed., pp. 103–118). Leuven University Press.

Weidman, A. C., Sun, J., Vazire, S., Quoidbach, J., Ungar, L. H., & Dunn, E. W. (2020). (Not) hearing happiness: Predicting fluctuations in happy mood from acoustic cues using machine learning. *Emotion*, *20*(4), 642–658. https://doi.org/10.1037/emo0000571

Wenz, A., Keusch, F., & Bach, R. L. (2024). Measuring Smartphone Use: Survey Versus Digital Behavioral Data. *Social Science Computer Review*, 08944393231224540. https://doi.org/10.1177/08944393231224540

Wenzel, M., & Brose, A. (2023). Addressing measurement issues in affect dynamic research: Modeling emotional inertia's reliability to improve its predictive validity of depressive symptoms. *Emotion*, *23*(2), 412–424. https://doi.org/10.1037/emo0001108

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

Wright, A. G. C., & Woods, W. C. (2020). Personalized Models of Psychopathology. *Annual Review of Clinical Psychology, 16*, 49–74. https://doi.org/10.1146/annurev-clinpsy-102419-125032

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

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, 12(6), 1100–1122. https://doi.org/10.1177/1745691617693393

Zhao, F., Shang, Y., & Trull, T. J. (2024). FENAP: Foundation Models for EMA-Derived Narrative Analysis and Prediction. *Journal of Healthcare Informatics Research*. https://doi.org/10.1007/s41666-024-00170-9