Mapping methodological variation in experience sampling research from design to data analysis: A systematic review.

Lisa Peeters[1*], Wim Van Den Noortgate[1,3], M. Annelise Blanchard[1,4], Gudrun Eisele[2], Olivia Kirtley[2], Richard Artner[1], Ginette Lafit[1]

[1] Faculty of Psychology and Educational Sciences, KU Leuven, Belgium

[2] Center for Contextual Psychiatry, Department of Neurosciences, KU Leuven, Belgium

[3] ITEC – an imec research group at KU Leuven, Belgium

[4] Psychological Sciences Research Institute, Université catholique de Louvain, Belgium

* Corresponding author. *E-mail address*: lisa.peeters@kuleuven.be

Formatted: English (United States)
Formatted: Superscript
Formatted: Superscript
Formatted: French (France)
Formatted: French (France)
Formatted: French (France)

Stage 1 Abstract

Aim. The Experience Sampling Method (ESM) has become a widespread tool to study time-varying constructs ~~(e.g., emotions, substance craving)~~ across many subfields of psychological ~~(e.g., organizational psychology, clinical psychology)~~ and psychiatric research. This large variety in subfields of research and constructs of interest has contributed to considerable methodological variation. Despite the importance of the methodological choices made by ESM researchers for the quality and veracity of current ESM research and potential future innovation, few have attempted to systematically assess these choices, and to explore their justification. ~~Existing systematic reviews have focused on specific demographics (e.g., certain clinical populations) or subfields of psychology, while previous non-systematic attempts to map methodological variation are limited and focused exclusively on (pre-data-collection) design choices.~~

Therefore, the ~~first aim of the current systematic review is~~two aims of the current systematic review are to 1) describe the methodological variation ~~(from conception of the research question to data analysis)~~ in ESM study designs in the recent psychological literature~~. The second aim of the review is to~~ and 2) assess the transparency (i.e., reporting and open science practices) of ~~ESM~~ these studies~~., which encompasses reporting and open sciences practices.~~ These aims are a first step towards a broader goal to improve the methodological quality of ESM research in psychology, contributing to a more rigorous, credible science of daily life.

Methods. ~~To this end, we~~We developed an extensive list of data extraction items covering the entire workflow of an ESM study, from conception of the research question to reporting the results. This data was extracted from 150 recently published articles applying ESM in the field of psychology and psychiatry. ~~This systematic review – and its aim to describe the content and transparency of ESM research – is a first step towards a broader goal to improve~~

~~the methodological quality of ESM research in psychology, contributing to a more rigorous, credible science of daily life.~~

Results. [Updated at Stage 2 – we report a descriptive and narrative synthesis of our data]

Implications. [Updated at Stage 2]

Keywords: ESM, Experience sampling method, ecological momentary assessment, ambulatory assessment, study design, methodological variation, transparency, systematic review

Formatted: Highlight

Formatted: Highlight

Mapping methodological variation in ESM research from design to data analysis: A systematic review.

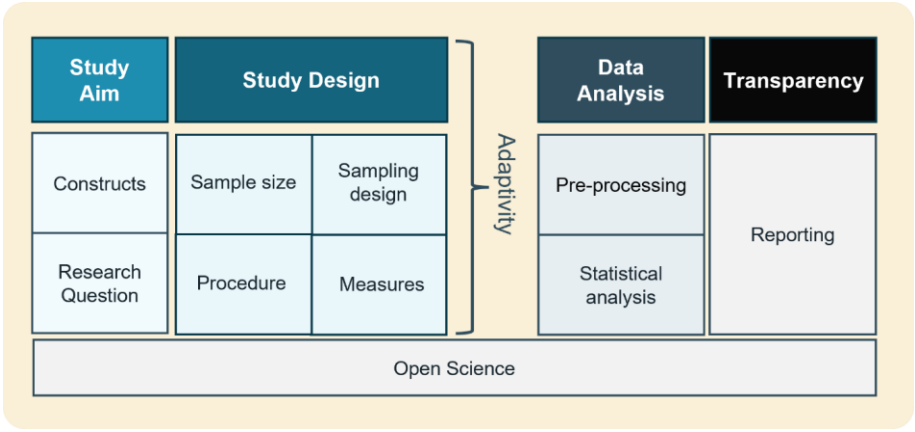
The Experience Sampling Method (ESM) has become the gold standard to study time-varying psychological processes, behaviors, and contexts in daily life. Examples include emotions (e.g., Blanke et al., 2020; Kalokerinos et al., 2019), substance use (e.g., Kurten et al., 2022; Weiss et al., 2022) and social experiences (e.g., Achterhof et al., 2022). The method is used across many subfields of psychological research, such as clinical psychology (e.g., Ader et al., 2022), personality psychology (e.g., Beck & Jackson, 2020; Kaurin et al., 2023) and organizational psychology (e.g., Shi et al., 2024). Also known by the closely related terms Ecological Momentary Assessment (EMA) and Ambulatory Assessment (AA), ESM is an intensive longitudinal data collection method aiming to capture constructs of interest in real-time and real-life contexts (Myin-Germeys & Kuppens, 2022). In their daily life, participants are prompted, usually multiple times per day, to complete self-report measurements. These measurements can inquire about participants' momentary thoughts, feelings, behavior and symptoms, as well as the context in which the measurement takes place (e.g., location, company; Myin-Germeys et al., 2009). Through the integration of data collection in participants' daily life, ESM allows the collection of ecologically valid data and a reduction of recall biases (i.e., biases that occur because participants' ability to retrospectively report past experiences is often challenged; Beal, 2015). In the past, participants were typically asked to fill in a pen-and-paper questionnaire each time they received a prompt from a beeper (Csikszentmihalyi & Larson, 1987). Technological advancement has facilitated the application of ESM by allowing participants to receive prompts and fill out questionnaires on a smartphone or on specialized research devices.

Despite its advantages and the technological advancement, ESM research does not come without its own challenges. Throughout the process of conducting an ESM study, the

researcher is presented with a myriad of – often ESM-specific – choices and considerations, mapped in Figure 1. In the first part of this work, we give a broad overview of the workflow of an ESM study, and present some of the literature that has aimed to map the methodological variation in (parts of) this process. This provides the backdrop for the second part of this work, a systematic review of the current ESM literature in psychology and psychiatry.

Figure 1

The Workflow of an ESM Study



The Workflow of an ESM Study

The first decision a researcher must take is whether an ESM study is appropriate for their research question. This is only the case if at least one of the constructs of interest varies over a relatively short time-frame (Myin-Germeys & Kuppens, 2022), such as state body image (Stieger et al., 2020) or affective experiences (Achterhof et al., 2022).

Once it has been decided that ESM is appropriate for the research question and the constructs involved, the researcher must consider a number of complex design decisions. Choices regarding the sample size in an ESM study include the number of participants as well as the number of measurements per participant. The sampling design (e.g., time between

measurements, type of sampling scheme) is closely related to the sample size decisions in an ESM study. As for non-ESM research, a power analysis can be conducted for sample size planning (Lakens, 2022). Unfortunately, due to the multilevel nature and the temporal structure of the data, conducting a power analysis can be complex (Lafit et al., 2021). A number of other design-related decisions regarding the measures (e.g., use of an existing validated questionnaire vs. creation of new items; Eisele et al., 2024) and the procedure (e.g., incentives, software/devices, briefing) must also be considered carefully (Fritz et al., 2023).

On one hand, design decisions must be based on the research question. For example, the number of assessments per day should follow the (theoretical) temporal dynamics of the construct(s) of interest (Seizer et al., 2024; Wright & Zimmermann, 2019). On the other hand, researchers must be mindful of the burden that design decisions place on participants, and the quality (e.g., careless responding) and quantity of the resulting data. This is particularly relevant as methodological research has shown that there is a link between data quality and/or compliance, and certain aspects of the study design, such as the type of sampling scheme (Himmelstein et al., 2019) and questionnaire length (Eisele et al., 2022).

In recent years, there has been a growing interest in adapting certain aspects of the design, such as the questionnaire length or the sampling scheme, to the participants, context or incoming data (Schneider et al., 2024). ‘Adaptivity’ is quite a broad term, that as of yet is not used often in the ESM literature. Based on related literature in the field of educational sciences (Wauters et al., 2010), we formulated a definition of adaptivity in ESM designs. This definition is intentionally very broad, so that the current work can maximally inform a more precise conceptualization of adaptivity. We define adaptivity as the adjustment of one or more characteristics of the ESM study design to the individual participants’ characteristics and preferences, preceding measurements, and/or the context. This adaptivity can take place before the start of the data collection, in which case it is referred to as static adaptivity (e.g.,

tailoring of the questionnaire items to individual characteristics; Scholten et al., 2022). Dynamic adaptivity, on the other hand, refers to real-time updates to the study design during data collection (e.g., "episode-contingent" designs in which an event of interest triggers a follow-up period of more frequent measurements; Revol et al., 2023a; Schreuder et al., 2024). Based on our definition, conditional questionnaire branching depending on context (e.g., depending on the reported social context; Achterhof et al., 2022) is considered dynamic adaptivity. Adaptive designs can improve the data quality/quantity (Cai et al., 2022), reduce participant burden (Bos et al., 2022) and/or improve parameter estimates (Revol et al., 2023a), but researchers should be aware that they also bring forth additional complexities across the course of a study.

After the data collection, the raw data has to be processed (e.g., calculation of scale scores, missing data handling), which can be guided by several existing guidelines and recommendations for pre-processing (e.g., Trull & Ebner-Priemer, 2020; Viechtbauer, 2021a). Revol et al. (2023b) created an extensive framework to assist ESM researchers with pre-processing their data, including a website, templates, and an R package. The importance of careful consideration of pre-processing steps has been emphasized by recent findings that different pre-processing choices can lead to differences in the statistical results (Lafit et al., 2024; Weermeijer et al., 2022). After preprocessing the data, the researcher makes a number of decisions regarding data analysis (e.g., type of statistical analysis, software). For statistical analysis, many resources are available (e.g., Bringmann et al., 2013; Lafit, 2022; Viechtbauer, 2021b). Due to the hierarchical nature of the data (i.e., measurements nested within participants), resources almost always include an introduction to multi-level modeling (e.g., Bolger & Laurenceau, 2013). In addition, the temporal nature of the data (i.e., repeated measurements close together in time) often necessitates a statistical analysis that takes a possible autocorrelation between measurements into account. Common methods to capture

this temporal dependency are autoregressive (AR) and vector autoregressive (VAR) models (Ariens et al., 2020), which can be extended hierarchically (Bringmann et al., 2013).

Finally, researchers report their research to the scientific community. There are many ways in which meticulous reporting contributes to scientific progress. Thorough reporting of the entire research process allows reviewers and readers to make their own judgement about the quality of a study (Schiavone et al., 2023; Stone & Shiffman, 2002; Trull & Ebner-Priemer, 2020; van Roekel et al., 2019). It also allows other researchers to replicate or reproduce the results, which is essential to the credibility of scientific findings (Artner et al., 2021; Simons, 2014), especially considering the "replication crisis" in psychology (Wiggins & Christopherson, 2019). Furthermore, complete reporting of the research process allows the reader to interpret the results in a more informed manner, considering the fact that a different choice at any point in the "garden of forking paths" (Gelman & Loken, 2013) could have led to different results. In addition to high-quality reporting, open science practices can also increase the credibility of research findings in psychology. Examples of such open science practices are registration, open materials, data sharing, and code sharing. Pre-registration (i.e., recording the research plan before data collection) and post-registration (i.e., recording the analysis plan after data collection but before data access) aim to prevent exploitation of the "garden of forking paths" (Benning et al., 2019). Sharing study materials, for example, by contributing ESM items to the ESM Item Repository (Kirtley et al., 2024), increases measurement transparency and facilitates replication. Data and code sharing further increase transparency, and allow for reproduction of the data analysis by independent researchers. Fortunately, many guidelines are available for reporting (Flake & Fried, 2020; Schiavone et al., 2023; Stone & Shiffman, 2002; Trull & Ebner-Priemer, 2020) and open science in ESM research (pre-registration; Kirtley et al., 2021; data sharing; Alter & Gonzalez, 2018; sharing materials; Eisele et al., 2024; Kirtley et al., 2024).

All of the above can be summarized as follows: across all stages of an ESM study, researchers face a large number of complex methodological decisions, each of which requires careful consideration.

Mapping the Methodological Variation

The methodological variation has been acknowledged by many researchers, but few have attempted to systematically assess the methodological choices made by ESM researchers. Such a systematic assessment is a complex endeavor due to the breadth of research questions that prompt researchers to conduct ESM studies. Wrzus and Neubauer (2023) found that emotions were a topic of study in the majority of ESM studies in the psychological literature, but other popular topics included a wide range of psychological concepts such as personality states and health behaviors (Perski et al., 2022). There is also a wide variety in populations studied and sampling procedures in ESM research. Healthy, (young) adult convenience samples have been found to be the most common, but ESM studies are also conducted with clinical or mixed samples as well as with different age groups, and can be recruited in various ways (Wrzus & Neubauer, 2023). Existing systematic reviews on the use of ESM have often focused on specific populations (e.g., clinical populations), or specific topics of research. Examples are reviews of the use of ESM designs in adolescent samples (Van Roekel et al., 2019), in psychosis research (Bogudzińska et al., 2024; Deakin et al., 2022), in research on suicidal ideation (Ammerman & Law, 2022; Janssens et al., 2024), and in pediatric healthcare (van Dalen et al., 2023).

Furthermore, several methodological meta-analyses have assessed aspects of ESM study design in the context of their relation to compliance, retention, or data quality. Again, most methodological meta-analyses focused on a specific demographic or topic of research (e.g., substance use; Jones et al., 2019; severe mental disorders; Vachon et al., 2019). In their broad meta-analysis of compliance and dropout in ESM studies in psychology and related

disciplines, Wrzus and Neubauer (2023) provided an overview of some important sample and design characteristics. They selected a subsample (N = 477) out of the 1523 articles that fulfilled all inclusion criteria and coded ESM design characteristics, sample characteristics, and compliance and dropout.

Other, non-systematic attempts to map methodological variation in ESM designs further attest to the broad variation in design decisions. Janssens et al. (2018) conducted a qualitative study with an international sample of 47 researchers who had experience conducting ESM or diary studies. These researchers reported their design choices in a recent study and the reasoning behind their decisions. Even in a sample of only 47 studies, there was a large variety in certain aspects of the study design (e.g., sampling frequency, number of items). Rationales for these decisions were categorized into four main themes: statistical reasons, feasibility, reliability, and nature of the variables. The most common rationales differed for the different design choices that were assessed. Reliability of the measurements was an important rationale for the chosen study duration, nature of the items (retrospective vs. momentary), the maximum response delay and the sampling scheme (fixed vs. semi-random vs. random). Statistical reasons were also important for the choice of sampling scheme and study duration. The chosen measurement frequency was mostly motivated by participant burden (Janssens et al., 2018).

The common thread through this literature is the enormous breadth of the methodological variation in EMS study designs. Regarding the sample, most of these studies coded the sample size and found a wide range (e.g., M = 136.6, SD = 176.0, range = 4 - 2001 participants in Wrzus & Neubauer, 2023). The total number of measurements per participant was assessed in most cases by recording on one hand the study duration in days (e.g., Mdn = 17, range = 1 - 270 days in Janssens et al., 2018) and on the other hand the number of measurements per day (e.g., Mdn = 5, range = 1 - 50 in Janssens et al., 2018). However, none

of the aforementioned reviews or meta-analyses assessed whether a justification was provided for the reported sample size and/or number of measurements (e.g., a power analysis; Lakens, 2022).

Many methodological reviews and meta-analyses also assessed certain aspects of the sampling scheme. The findings regarding the most common sampling schemes are mixed: Janssens et al. (2018) found that 52% of the included studies used a fixed sampling scheme, while only 18% did so according to Vachon et al. (2019). Signal-contingent sampling was found to be most common, but other types of sampling (e.g., event-based) are often reported as well (Wrzus & Neubauer, 2023). Only the qualitative study by Janssens et al. (2018) assessed researchers' justifications regarding the chosen sampling schemes; existing reviews and meta-analyses did not.

The scarcity of systematic assessment of ESM measurements is striking, as there is a rising interest in assessing the quality of measurements in the psychological ESM literature (e.g., Brose et al., 2020; Cloos et al., 2023; Horstmann & Ziegler, 2020). Many previous reviews only recorded the number of items per questionnaire and found, again, a lot of variation. For example, in their meta-analysis, Vachon et al. (2019) found that the number of items in their set of primary ESM studies varied between 2 and 135. Other aspects of the measurements that have been assessed in a systematic manner, albeit more rarely, are response scales and questionnaire development. Findings indicated that Likert scales are the most common response scales (Vachon et al., 2019), and most studies use non-validated questionnaires (Deakin et al., 2022; Horstmann & Ziegler, 2020; Wright & Zimmermann, 2019).

Regarding the ESM procedure, it is evident that the use of electronic devices has become much more popular than pen-and-paper questionnaires (Ammerman & Law, 2022;

Vachon et al., 2019). Incentives are often of a financial nature (i.e., direct monetary payment/giftcards; Wrzus & Neubauer, 2023) and are often compliance-based (Jones et al., 2019; van Dalen et al., 2023). However, especially in healthcare settings it is increasingly common to receive personalized feedback as an incentive (van Dalen et al., 2023). Despite the importance of (de)briefing for compliance and data quality (Palmier-Claus et al., 2011), very few studies have systematically assessed (de)briefing in ESM studies. However, it has been found that many ESM studies include at least some training before the ESM period (Deakin et al., 2022; van Roekel et al., 2019).

The aforementioned literature undoubtedly provided valuable insights into the process of designing an ESM study. However, several arguments can be made for a need for further investigation of the research process in ESM studies. First, the field of ESM has, and continues to evolve quickly (Mestdagh & Dejonckheere, 2021). ESM designs continue to become more popular, more advanced, and often more complex (e.g., dyad studies; Griffith & Hankin, 2024; episode-contingent sampling designs; Revol et al., 2023a). Therefore, these previous findings might not be representative of the field of ESM research today. Second, as mentioned previously, most previous reviews and meta-analyses have been population-specific or subject-specific. While these works are of immense importance in their target subfield, we cannot assume that they are representative of the larger psychological ESM literature. Third, the methodological choices that were assessed in previous research were limited to a set of common design choices. Wrzus and Neubauer (2023) coded eight design features related to the sample size and sampling design (number of participants, total number of assessments, assessments per day, study duration) and the study procedure (incentives, reinforcement, participant care). Other reviews and meta-analyses coded similar sets of design choices. While these choices remain relevant today, they are only a subset of the myriad of methodological choices researchers are faced with over the course of an ESM study. Out of all

the steps of the workflow of a study (Figure 1), the study design is the only one that has been reviewed extensively in the ESM literature. However, a study does not end after its design, and the quality of its final product (communication of the newfound knowledge to the scientific community) depends on much more than the quality of the design alone. ESM researchers again face complex choices during the data collection, when they are pre-processing/analyzing the data and when they report their work (Figure 1).

Due to inconsistency in describing adaptivity, synthesis of the literature is difficult. As a consequence, to the best of our knowledge, no reviews or guidelines for adaptive methods in ESM research exist. As mentioned previously, guidelines and recommendations for preprocessing and analyzing data do exist. However, these resources only provide normative information. They do not describe the actual approaches taken by ESM researchers. In practice, pre-processing is often limited or not reported in detail (Blanchard et al., 2023; Revol et al., 2023b). In a crowdsourcing project, Bastiaansen et al. (2020) asked twelve teams of researchers to use the same raw ESM dataset to select target symptoms for treatment. The pre-processing included few steps for most teams, although all teams applied at least some form of pre-processing (e.g., clustering items, detrending, imputation of missing data; Bastiaansen et al., 2022). Researchers also rarely report checking statistical model assumptions before conducting the analyses (Blanchard et al., 2023).

Concerning data analysis, Bastiaansen et al. (2020) and Blanchard et al. (2023) found that analyses are most often conducted using R (R Core Team, 2024) and that VAR models are the most common statistical models fitted to the data. Almost all teams in Bastiaansen et al. (2022) used a VAR model to investigate autoregressive effects (i.e., whether a variable predicts itself at a later time point) and cross-lagged effects (i.e., whether a variable predicts another variable at a later time point). Several teams additionally investigated contemporaneous effects (i.e., whether a variable covaries with another variable at the same

time point). Some also used network approaches, in which multivariate data is analyzed by modeling statistical relationships between nodes in a network of variables (Borsboom et al., 2021). However, no two teams took exactly the same (or even very similar) approaches. Consequently, no two teams selected the same target symptoms for treatment. These results illustrate the critical role of the analysis strategy for the conclusions of a study. Moreover, they emphasize the importance of reporting clearly and completely how the results and conclusions of a study came to be.

However, current reporting practices do not necessarily reflect the abundance of resources available. In a review of ESM studies published in three major psychopathology journals between 2012 and 2018, Trull and Ebner-Priemer (2020) found a lot of room for improvement regarding reporting practices. For example, only 30% of the included studies reported psychometric properties of the items used in the ESM measurements, and only 32% reported the technical details of the sampling. Furthermore, only 17% provided a rationale for their sampling design, density, and scheduling. Regarding open science initiatives, as well, reviews find that their application in practice is far from perfect. For example, many preregistered studies deviate from their pre-registration without adequate disclosure of these deviations (Claesen et al., 2021). It is unclear how many researchers share their raw data and/or code.

The Current Study

We have illustrated the lack of a comprehensive overview of current practices in ESM research in psychology. The need for such an overview stems from two goals. On one hand, it allows for assessment and improvement of the quality of the use of ESM in psychological research. Knowledge of the current practices in ESM research can inform the development of guidelines and/or resources for all stages of the research workflow, which can in turn help researchers to apply ESM and report their work in an informed and methodologically valid

manner. On the other hand, a broad overview of the current state of the art in ESM research can facilitate innovation and improvement of the method itself. One example is the use of adaptivity in ESM designs: we hope to map how ESM researchers apply adaptivity today, and to use this information to identify how we can improve upon these methods in the future.

To address this need, the aim of the current study is twofold: the description of the variation in current methodological practices in ESM research, and the assessment of transparency in this field. Methodological practices here include the entire workflow of an ESM study, from conception of the research question to interpretation of the results (see Figure 1). Transparency refers to both reporting practices, and open science practices.

To this end, a systematic review of the ESM literature will be conducted. As discussed above, previous reviews were conducted in a specific subfield of psychology or a specific demographic and focused largely on a small set of design-related methodological choices. In contrast, the current review aims to describe a wide range of methodological choices in the entire field of psychological research. For each of these methodological choices, this review will assess 1) whether they were reported clearly and completely, 2) the decisions made (as far as they were reported), including the use of overarching open science practices (e.g., pre-registration), and 3) the rationale given for some of the decisions. For this review, we developed an extensive list of items covering methodological choices across the entire process of conducting an ESM study, from initial conception to reporting of the results.

Methods

Search Strategy

The systematic review will follow PRISMA guidelines (Page et al., 2021). The key component of this search is the experience sampling method (ESM) in the context of psychological research. The following databases will be searched: PsycARTICLES, ~~PsycInfo,~~

MEDLINE, Psychology Databases (via ProQuest), ~~and~~ Web of Science Core Collection (via Clarivate) and PubMed. To ensure the feasibility of the review, the search will be limited to works published at most one calendar year before the start of the search [Exact dates updated at Stage 2]. ~~in the year 2023 (1/1/2023–31/12/2023)~~. We argue that this focus on the recent literature reflects the aim of the review, namely, to assess the state of the art in psychological ESM research and to map the current research/reporting practices. As the search is limited to the year 2023, citation tracking would presumably identify very few additional records (if any) and will therefore not be carried out. To ensure the full understanding of the text by all reviewers, the search is limited to records written in English.

Formatted: Highlight

Search terms were compiled with the help of experts in both systematic reviews and experience sampling methodology. The search query includes only terms related to ESM. No search terms related to psychology were included; whether the research belongs to the field of psychology will be assessed in the screening phase. The full search query, tailored to each of the databases, can be found in the Supplementary Materials (https://osf.io/abvxp/).

Screening

Records are eligible for inclusion if they fulfill several criteria. They must describe an empirical study (i.e., no reviews or meta-analyses) in which ESM was applied with more than one measurement per day (i.e., no daily diary studies). Furthermore, ESM was used as a tool to answer a psychological research question (e.g., whether momentary affect is related to social interactions; Achterhof et al., 2022), not as a topic of research itself (e.g., whether a certain aspect of an ESM design is related to compliance; Eisele et al., 2022). This criterion thus excludes feasibility and methodological studies. Only records that described the entire research process (i.e., no protocols) will be included. Secondary data analyses are included if they report all stages of the research process. Eligible records must be published in a journal that is classified as “Psychology/Psychiatry” in the Web of Science database. The

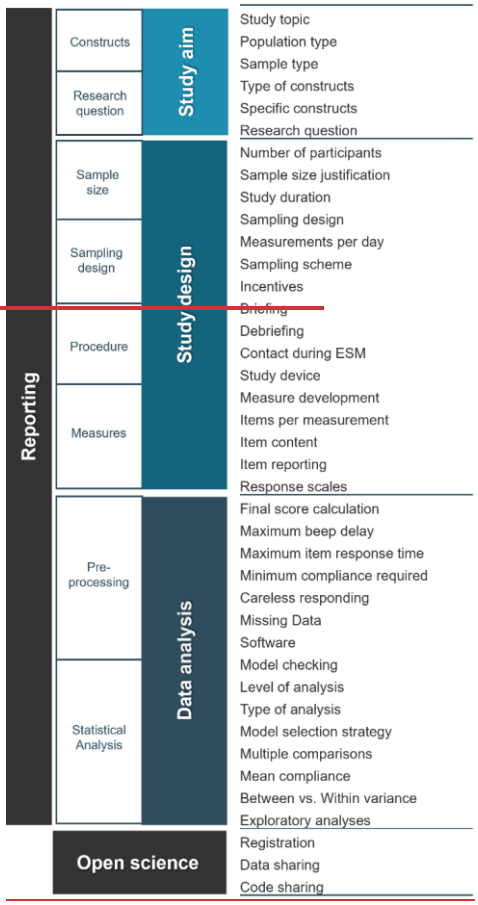
"Psychiatry"-category is included to ensure that the review did not exclude a large portion of the clinical ESM literature, which is often published in psychiatric journals. A full list of journals that fulfill this criterion can be found in the Supplementary Materials. Records are excluded if passive sensing data (e.g., accelerometry) were analyzed, or if the study was not exclusively observational (e.g., intervention studies, experimental studies). Longitudinal studies that analyzed data from multiple ESM study waves (i.e., periods of ESM measurement separated by non-measurement periods; Myin-Germeys & Kuppens, 2022) will also be excluded. While these design features are of increasing importance in the ESM literature, they lead to a different, additional set of design and analysis choices that fall beyond the scope of the current review.

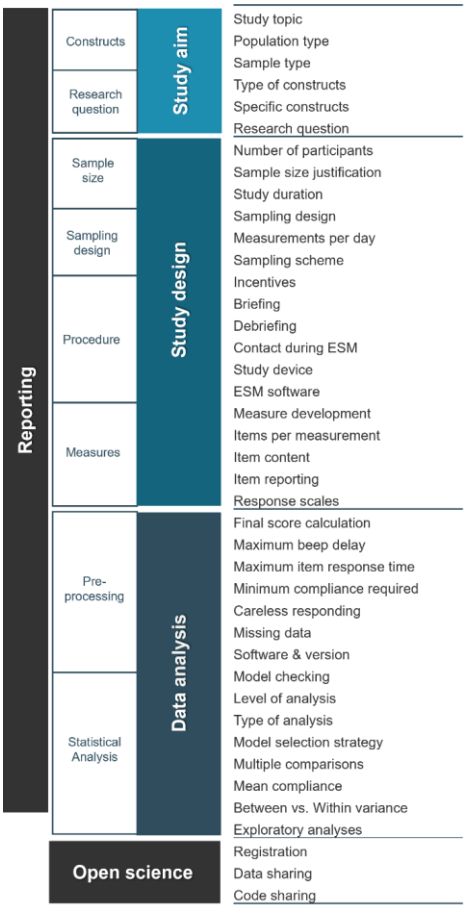
Study selection will take place in two stages. In Stage 1, the titles and abstracts are screened. In Stage 2, the full texts of records that survived the first stage are scrutinized. In both stages, one reviewer evaluates all records, while ~~another reviewer~~several other reviewers independently screen~~s~~ a subsample to allow for the evaluation of consistency in decision-making. Reviewers will use separate Zotero libraries and decision logs to ensure blinding of decisions. As a skewed distribution of decisions (few inclusions and many exclusions) could lead to a low Cohen's Kappa despite high values of percent agreement (Belur et al., 2021), both statistics will be calculated to assess inter-rater reliability. According to Landis and Koch (1977), Kappa values between .61 and .80 can be interpreted as "substantial", while values above .80 are "almost perfect". Discrepancy in decisions will be resolved through discussion. If the disagreement cannot be resolved through discussion, an ~~third~~ independent reviewer that was not involved in the screening stage will resolve the issue. Discrepancies, discussions, and resolutions will be recorded in a decision log; full-text screening decisions are also recorded along with reasons for exclusions.

~~To ensure the feasibility of the review (considering the small team), an upper limit was set for the final sample.~~ After all records are screened, a random sample of 150 records will be selected out of all eligible records. Random sampling of eligible records has been used in systematic reviews (e.g., Wrzus & Neubauer, 2023) to allow for sufficient coverage of the literature while safeguarding the review's feasibility (considering the small team). ~~This limit~~The sample size was decided based on a pilot ~~conducted on records published in 2022~~ (described in the Supplementary Materials), and a rough estimate of the number of relevant studies that are published in a single year based on previous systematic reviews (e.g., Wrzus & Neubauer, 2023). In case the number of eligible records is unexpectedly lower than this limit, the search will be expanded with an additional six months – to a total of 1.5 years before the start of the initial search. ~~, and allows for sufficient coverage of the current literature while safeguarding the review's feasibility.~~

405 **Figure 2**

406 *Methodological Choices Covered by the Data Extraction Items*


Reporting
Constructs
Research question
Study aim
Sample size
Sampling design
Procedure
Measures
Study design
Pre-processing
Statistical Analysis
Data analysis
Open science
Study topic
Population type
Sample type
Type of constructs
Specific constructs
Research question
Number of participants
Sample size justification
Study duration
Sampling design
Measurements per day
Sampling scheme
Incentives
Debriefing
Debriefing
Contact during ESM
Study device
Measure development
Items per measurement
Item content
Item reporting
Response scales
Final score calculation
Maximum beep delay
Maximum item response time
Minimum compliance required
Careless responding
Missing Data
Software
Model checking
Level of analysis
Type of analysis
Model selection strategy
Multiple comparisons
Mean compliance
Between vs. Within variance
Exploratory analyses
Registration
Data sharing
Code sharing



Data Extraction

The data that will be extracted consists of main items (Figure 2), with possible follow-up items in a branched structure (i.e., based on the response to the main item, different follow-up items are relevant). An initial set of items related to all stages of the research process was compiled based on a scoping search of older literature, available validity and reporting tools, and discussion with co-authors. The previously mentioned reporting recommendations (Flake & Fried, 2020; Schiavone et al., 2023; Stone & Shiffman, 2002; Trull & Ebner-Priemer, 2020) and meta-analyses (Wrzus & Neubauer, 2023) were particularly helpful in this endeavor. This set of items was subsequently refined in three rounds. In each

round, four to six co-authors independently evaluated the list of items through ~~MS~~Microsoft Ff~~f~~orms. At the end of each round, a group discussion was held with all reviewers to improve the list of items. Details of the review process, as well as the final list of data extraction items including follow-up items can be found in the Supplementary Materials. The last item, 'Is the word "Inertia" mentioned anywhere in the text?', was added in the interest of a separate study, and its results will not be discussed in this work.

For the data extraction, the reviewers will make use of a data extraction form created in Microsoft~~MS~~ Forms, which supports the branched structure of the items. ~~Two independent reviewers will extract the data.~~ Again, one reviewer extracts data from all records, while several ~~the~~ other reviewers only extract~~s~~ data from a subsample to allow for the evaluation of consistency in decision-making. Missing data will be coded as "not reported", and no attempt will be made to contact authors for more information. This is in line with the aim of this review, as part of this aim is to assess the quality of scientific reporting in the literature. Thus, the non-reporting of any data is also considered as valuable information.

If a record reported on more than one (sub)study that satisfied all inclusion criteria, data will be extracted for each study separately. When studies reported on multiple research questions, up to three main research questions per study will be considered. We include only research questions that are presented, including results, in the abstract of the record (following Artner et al., 2021). If more than three research questions satisfy this criterion, the most important research questions will be selected through discussion with the research team, with precedence for research questions with a priori hypotheses.

The data extraction form was piloted using records that were published in 2022 and would therefore not be included in the systematic review. A full description of the pilot can be found in the Supplementary Materials.

Data Synthesis

For numerical items (e.g., the study duration in days), summary statistics will be calculated. For binary or categorical items (e.g., whether or not a power analysis was conducted), proportions will be reported. The synthesis of open-ended items is dependent on the item's complexity: simpler items (e.g., the statistical software that was used) will be grouped into categories during data pre-processing and proportions will be reported. More complex items (e.g., a brief description of the research question) will be synthesized in a narrative way. As the synthesis of numerical and categorical data is largely predetermined, this synthesis will be carried out by one researcher and the reliability of the decisions will not be assessed. An RMarkdown document containing the preliminary analysis plan for this data can be found on https://osf.io/abvxp/. Open-ended items will be coded by two researchers independently, and reliability will be assessed. Decisions regarding presentation of the numerical results and the narrative synthesis (Popay et al., 2006) of the open-ended items will be discussed with multiple members of the research team. All statistical procedures will be conducted using the most recent version of R at the start of the data synthesis; the exact version will be reported in the Stage 2 report.

Formatted: Highlight

Study Design Template

Question	Hypo-thesis	Sampling plan	Analysis Plan	Rationale for deciding the sensitivity of the test for confirming or dis-confirming the hypothesis	Inter-pretation given different outcomes	Theory that could be shown wrong by the outcomes
What are the current methodological practices (throughout the entire process of conducting a study) in the	N/A	Systematic review of all experience sampling research published in 2023, with a limit of 150 publications to	For binary and categorical variables, percentages will be calculated for each category. For numerical variables, summary statistics (mean, standard deviation, range) will be calculated. Variables derived from short open	N/A	N/A	N/A

field of experience sampling research in psychology and psychiatry?		safeguard feasibility. For all publications, a large amount of data covering the entire research process will be extracted.	data extraction items will be made categorical after manual processing. Long open items will be summarized narratively.			
Are authors of papers describing experience sampling studies transparent about their methodological practices?	N/A		For a large number of variables, the percentage of publications that explicitly reported the relevant information will be calculated. For open science practices (categorical variables), percentages of each category will be calculated.	N/A	N/A	N/A

References

Achterhof, R., Kirtley, O. J., Schneider, M., Hagemann, N., Hermans, K. S. F. M., Hiekkaranta, A. P., Lecei, A., Lafit, G., & Myin-Germeys, I. (2022). Adolescents' real-time social and affective experiences of online and face-to-face interactions. *Computers in Human Behavior*, *129*, 107159. https://doi.org/10.1016/j.chb.2021.107159

Ader, L., Schick, A., Simons, C., Delespaul, P., Myin-Germeys, I., Vaessen, T., & Reininghaus, U. (2022). Positive affective recovery in daily life as a momentary mechanism across subclinical and clinical stages of mental disorder: Experience sampling study. *JMIR Mental Health*, *9*(11), e37394. https://doi.org/10.2196/37394

Alter, G., & Gonzalez, R. (2018). Responsible practices for data sharing. *The American Psychologist*, *73*(2), 146. https://doi.org/10.1037/amp0000258

Ammerman, B. A., & Law, K. C. (2022). Using intensive time sampling methods to capture daily suicidal ideation: A systematic review. *Journal of Affective Disorders*, *299*, 108–117. https://doi.org/10.1016/j.jad.2021.10.121

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. https://doi.org/10.1016/j.jpsychores.2020.110191

Field Code Changed

Artner, R., Verliefde, T., Steegen, S., Gomes, S., Traets, F., Tuerlinckx, F., & Vanpaemel, W. (2021). The reproducibility of statistical results in psychological research: An investigation using unpublished raw data. *Psychological Methods*, *26*(5), 527–546. https://doi.org/10.1037/met0000365

Bastiaansen, J. A., Kunkels, Y. K., Blaauw, F. J., Boker, S. M., Ceulemans, E., Chen, M., Chow, S.-M., de Jonge, P., Emerencia, A. C., Epskamp, S., Fisher, A. J., Hamaker, E. L., Kuppens, P., Lutz, W., Meyer, M. J., Moulder, R., Oravecz, Z., Riese, H., Rubel, J., … Bringmann, L. F. (2020). Time to get personal? The impact of researchers choices on the selection of treatment targets using the experience sampling methodology. *Journal of Psychosomatic Research*, *137*, 110211. https://doi.org/10.1016/j.jpsychores.2020.110211

Beal, D. J. (2015). ESM 2.0: State of the art and future potential of experience sampling methods in organizational research. *Annual Review of Organizational Psychology and Organizational Behavior*, *2*(2), 383–407. https://doi.org/10.1146/annurev-orgpsych-032414-111335

Beck, E. D., & Jackson, J. J. (2020). Consistency and change in idiographic personality: A longitudinal ESM network study. *Journal of Personality and Social Psychology*, *118*(5), 1080–1100. https://doi.org/10.1037/pspp0000249

Belur, J., Tompson, L., Thornton, A., & Simon, M. (2021). Interrater reliability in systematic review methodology: Exploring variation in coder decision-making. *Sociological Methods & Research*, *50*(2), 837–865. https://doi.org/10.1177/0049124118799372

Benning, S. D., Bachrach, R. L., Smith, E. A., Freeman, A. J., & Wright, A. G. C. (2019). The registration continuum in clinical science: A guide toward transparent practices. *Journal of Abnormal Psychology*, *128*(6), 528–540. https://doi.org/10.1037/abn0000451

Blanchard, M., Contreras, A., Kalkan, R., & Heeren, A. (2023). Auditing the research practices and statistical analyses of the group-level temporal network approach to psychological constructs:

Field Code Changed

A systematic scoping review. *Behavior Research Methods*, *55*(2), 767–787. https://doi.org/10.3758/s13428-022-01839-y

Blanke, E. S., Brose, A., Kalokerinos, E. K., Erbas, Y., Riediger, M., & Kuppens, P. (2020). Mix it to fix it: Emotion regulation variability in daily life. *Emotion*, *20*(3), 473–485. https://doi.org/10.1037/emo0000566

Bogudzińska, B., Jaworski, A., Zajdel, A., Skrzypek, K., & Misiak, B. (2024). The experience sampling methodology in psychosis risk states: A systematic review. *Journal of Psychiatric Research*, *175*, 34–41. https://doi.org/10.1016/j.jpsychires.2024.04.050

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

Borsboom, D., Deserno, M. K., Rhemtulla, M., Epskamp, S., Fried, E. I., McNally, R. J., Robinaugh, D. J., Perugini, M., Dalege, J., Costantini, G., Isvoranu, A.-M., Wysocki, A. C., van Borkulo, C. D., van Bork, R., & Waldorp, L. J. (2021). Network analysis of multivariate data in psychological science. *Nature Reviews Methods Primers*, *1*(1), 1–18. https://doi.org/10.1038/s43586-021-00055-w

Bos, F. M., Von Klipstein, L., Emerencia, A. C., Veermans, E., Verhage, T., Snippe, E., Doornbos, B., Hadders-Prins, G., Wichers, M., & Riese, H. (2022). A web-based application for personalized ecological momentary assessment in psychiatric care: User-centered development of the PETRA application. *JMIR Mental Health*, *9*(8), e36430. https://doi.org/10.2196/36430

Bringmann, L. F., Vissers, N., Wichers, M., Geschwind, N., Kuppens, P., Peeters, F., Borsboom, D., & Tuerlinckx, F. (2013). A network approach to psychopathology: New insights into clinical longitudinal data. *PLOS ONE*, *8*(4), e60188. https://doi.org/10.1371/journal.pone.0060188

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

Cai, L., Barnes, L. E., & Boukhechba, M. (2022). A framework for adaptive mobile ecological momentary assessments using reinforcement learning. In K. Arai (Ed.), *Intelligent Systems and Applications* (pp. 31–50). Springer International Publishing. https://doi.org/10.1007/978-3-030-82193-7_3

Claesen, A., Gomes, S., Tuerlinckx, F., & Vanpaemel, W. (2021). Comparing dream to reality: An assessment of adherence of the first generation of preregistered studies. *Royal Society Open Science*, *8*(10), 211037. https://doi.org/10.1098/rsos.211037

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

Csikszentmihalyi, M., & Larson, R. (1987). Validity and reliability of the Experience-Sampling Method. *The Journal of Nervous and Mental Disease*, *175*(9), 526–536. https://doi.org/10.1097/00005053-198709000-00004

Deakin, E., Ng, F., Young, E., Thorpe, N., Newby, C., Coupland, C., Craven, M., & Slade, M. (2022). Design decisions and data completeness for experience sampling methods used in psychosis: Systematic review. *BMC Psychiatry*, *22*(1), 669. https://doi.org/10.1186/s12888-022-04319-x

Eisele, G., Hiekkaranta, A. P., Kunkels, Y. K., Aan Het Rot, M., Van Ballegooijen, W., 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

Field Code Changed

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

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. https://doi.org/10.1177/2515245920952393

Fritz, J., Piccirillo, M., Cohen, Z. D., Frumkin, M., Kirtley, O., Moeller, J., Neubauer, A., Norris, L., Schuurman, N. K., Snippe, E., & Bringmann, L. (2023). *So you want to do ESM? Ten essential topics for implementing the Experience Sampling Method (ESM)*. OSF. https://doi.org/10.31219/osf.io/fverx

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.* Department of Statistics, Columbia University. Retrieved from http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf

Griffith, J. M., & Hankin, B. L. (2024). Longitudinal coupling of emotional wellbeing in parent-adolescent dyads: Evaluating the role of daily life positive affect socialization processes. *Development and Psychopathology*, 1–19. https://doi.org/10.1017/S0954579423001633

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

Horstmann, K. T., & Ziegler, M. (2020). Assessing personality states: What to consider when constructing personality state measures. *European Journal of Personality*, *34*(6), 1037–1059. https://doi.org/10.1002/per.2266

Janssens, J. J., Kiekens, G., Jaeken, M., & Kirtley, O. J. (2024). A systematic review of interpersonal processes and their measurement within experience sampling studies of self-injurious thoughts and behaviours. *Clinical Psychology Review*, *113*, 102467. https://doi.org/10.1016/j.cpr.2024.102467

Janssens, K. A. M., Bos, E. H., Rosmalen, J. G. M., Wichers, M. C., & Riese, H. (2018). A qualitative approach to guide choices for designing a diary study. *BMC Medical Research Methodology*, *18*(1), 140. https://doi.org/10.1186/s12874-018-0579-6

Jones, A., Remmerswaal, D., Verveer, I., Robinson, E., Franken, I. H. A., Wen, C. K. F., & Field, M. (2019). Compliance with ecological momentary assessment protocols in substance users: A meta-analysis. *Addiction*, *114*(4), 609–619. https://doi.org/10.1111/add.14503

Kalokerinos, E. K., Erbas, Y., Ceulemans, E., & Kuppens, P. (2019). Differentiate to regulate: Low negative emotion differentiation is associated with ineffective use but not selection of emotion-regulation strategies. *Psychological Science*, *30*(6), 863–879. https://doi.org/10.1177/0956797619838763

Kaurin, A., King, K. M., & Wright, A. G. C. (2023). Studying personality pathology with ecological momentary assessment: Harmonizing theory and method. *Personality Disorders: Theory, Research, and Treatment*, *14*(1), 62–72. https://doi.org/10.1037/per0000596

Kirtley, O. J., Eisele, G., Kunkels, Y. K., Hiekkaranta, A., Van Heck, L., Pihlajamäki, M. R., Kunc, B., Schoefs, S., Kemme, N., Biesemans, T., & Myin-Germeys, I. (2024). The Experience Sampling Method (ESM) Item Repository. https://doi.org/10.17605/OSF.IO/KG376

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), 2515245920924686. https://doi.org/10.1177/2515245920924686

Field Code Changed

Kurten, S., Vanherle, R., Beullens, K., Gebhardt, W. A., van den Putte, B., & Hendriks, H. (2022). Like to drink: Dynamics of liking alcohol posts and effects on alcohol use. *Computers in Human Behavior*, *129*, 107145. https://doi.org/10.1016/j.chb.2021.107145

Lafit, G. (2022). Non-normal, higher-level, and Var(1) models for the analysis of ESM data. In Myin-Germeys I, Kuppens P. (Eds.), *The open handbook of Experience Sampling Methodology: A step-by-step guide to designing, conducting, and analyzing ESM studies.* (pp. 185–214). The Center for Research on Experience Sampling and Ambulatory Methods Leuven (REAL).

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-griendly 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. https://doi.org/10.1177/2515245920978738

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*, *0*(0). https://doi.org/10.1177/10731911241286868

Lakens, D. (2022). Sample size justification. *Collabra: Psychology*, *8*(1), 33267. https://doi.org/10.1525/collabra.33267

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174. https://doi.org/10.2307/2529310

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

Formatted: French (France)

Formatted: English (United States)

Myin-Germeys, I., & Kuppens, P. (Eds.). (2022) *The open handbook of experience sampling methodology: A step-by-step guide to designing, conducting, and analyzing ESM studies* (2nd ed.). Leuven: Center for Research on Experience Sampling and Ambulatory Methods Leuven.

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

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, n71. https://doi.org/10.1136/bmj.n71

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

Perski, O., Keller, J., Kale, D., Asare, B. Y.-A., Schneider, V., Powell, D., Naughton, F., ten Hoor, G., Verboon, P., & Kwasnicka, D. (2022). Understanding health behaviours in context: A systematic review and meta-analysis of ecological momentary assessment studies of five key health behaviours. *Health Psychology Review*, *16*(4), 576–601. https://doi.org/10.1080/17437199.2022.2112258

Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., & Britten, N. (2006). *Narrative synthesis in systematic reviews: A product from the ESRC methods programme.* Lancaster University. https://doi.org/10.13140/2.1.1018.4643

Formatted: Font: Italic

Formatted: Font: Not Italic

R Core Team. (2024). *R: A Language and Environment for Statistical Computing*. Vienna, Austria. Retrieved from https://www.R-project.org/

Revol, J., Ariens, S., Lafit, G., Adolf, J., & Ceulemans, E. (2023a). *Episode-contingent experience-sampling designs for accurate estimates of autoregressive dynamics*. OSF. https://doi.org/10.31234/osf.io/wbu3x

Revol, J., Carlier, C., Lafit, G., Verhees, M., Sels, L., & Ceulemans, E. (2023b). *Preprocessing ESM data: A step-by-step framework, tutorial website, R package, and reporting templates*. OSF. https://doi.org/10.31234/osf.io/hnu2t

Schiavone, S. R., Quinn, K. A., & Vazire, S. (2023). *A consensus-based tool for evaluating threats to the validity of empirical research*. PsyArXiv. https://doi.org/10.31234/osf.io/fc8v3

Schneider, S., Junghaenel, D. U., Smyth, J. M., Fred Wen, C. K., & Stone, A. A. (2024). Just-in-time adaptive ecological momentary assessment (JITA-EMA). *Behavior Research Methods*, *56*(2), 765–783. https://doi.org/10.3758/s13428-023-02083-8

Scholten, S., Lischetzke, T., & Glombiewski, J. A. (2022). Integrating theory-based and data-driven methods to case conceptualization: A functional analysis approach with ecological momentary assessment. *Psychotherapy Research*. https://doi.org/10.1080/10503307.2021.1916639

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

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

Field Code Changed

Field Code Changed

Formatted: French (France)

Formatted: French (France)

Formatted: German (Germany)

Formatted: German (Germany)

Field Code Changed

Shi, G., Wu, Y., Pang, H., & Liu, Z. (2024). The double-edged sword effect of leaders' fear of missing out on leaders' creativity: An experience sampling method study. *Psychology Research and Behavior Management*, *17*, 259–271. https://doi.org/10.2147/PRBM.S449490

Simons, D. J. (2014). The value of direct replication. *Perspectives on Psychological Science*, *9*(1), 76–80. https://doi.org/10.1177/1745691613514755

Stieger, S., Aichinger, I., & Swami, V. (2022). The impact of nature exposure on body image and happiness: An experience sampling study. *International Journal of Environmental Health Research*. https://www-tandfonline-com.kuleuven.e-bronnen.be/doi/abs/10.1080/09603123.2020.1803805

Stone, A. A., & Shiffman, S. (2002). Capturing momentary, self-report data: A proposal for reporting guidelines. *Annals of Behavioral Medicine*, *24*(3), 236–243. https://doi.org/10.1207/S15324796ABM2403_09

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), 56–63. https://doi.org/10.1037/abn0000473

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

Van Dalen, M., Snijders, A., Dietvorst, E., Bracké, K., Nijhof, S. L., Keijsers, L., Hillegers, M. H. J., & Legerstee, J. S. (2023). Applications of the experience sampling method (ESM) in paediatric healthcare: A systematic review. *Pediatric Research*, 1–14. https://doi.org/10.1038/s41390-023-02918-2

Field Code Changed

Van Roekel, E., Keijsers, L., & Chung, J. M. (2019). A review of current ambulatory assessment studies in adolescent samples and practical recommendations. *Journal of Research on Adolescence*, *29*(3), 560–577. https://doi.org/10.1111/jora.12471

Wiechtbauer, W. (2021a). Structuring, checking, and preparing the data. In I. Myin-Germeys & P. Kuppens (Eds.), *The open handbook of Experience Sampling Methodology* (pp. 137–152). Center for Research on Experience Sampling and Ambulatory Methods Leuven (REAL).

Wiechtbauer, W. (2021b). Statistical methods for ESM data. In I. Myin-Germeys & P. Kuppens (Eds.), *The open handbook of Experience Sampling Methodology* (pp. 153–183). Center for Research on Experience Sampling and Ambulatory Methods Leuven (REAL).

Weermeijer, J., Lafit, G., Kiekens, G., Wampers, M., Eisele, G., Kasanova, Z., Vaessen, T., Kuppens, P., & Myin-Germeys, I. (2022). Applying multiverse analysis to experience sampling data: Investigating whether preprocessing choices affect robustness of conclusions. *Behavior Research Methods*, *54*(6), 2981–2992. https://doi.org/10.3758/s13428-021-01777-1

Weiss, D. M., Bernier, E., Robbins, D. R., Elacqua, K. M., Johnson, K. A., Powers, K., Mesholam-Gately, R. I., & Woodberry, K. A. (2022). Using experience sampling methodology data to characterize the substance use of youth with or at-risk of psychosis. *Frontiers in Psychiatry*, *13*, 874246. https://doi.org/10.3389/fpsyt.2022.874246

Wiggins, B. J., & Christopherson, C. D. (2019). The replication crisis in psychology: An overview for theoretical and philosophical psychology. *Journal of Theoretical and Philosophical Psychology*, *39*(4), 202–217. https://doi.org/10.1037/teo0000137

Wright, A. G. C., & Zimmermann, J. (2019). Applied ambulatory assessment: Integrating idiographic and nomothetic principles of measurement. *Psychological Assessment*, *31*(12), 1467–1480. https://doi.org/10.1037/pas0000685

Field Code Changed

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