AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychological Assessment

© 2025 American Psychological Association
ISSN: 1040-3590

2025, Vol. 37, No. 10, 479–492
https://doi.org/10.1037/pas0001409

Cost-Effective Experience Sampling Method Studies: Integrating Budget Constraints Into Sample Size Decisions

Jordan Revol[1], Ginette Lafit[2], Olivia Kirtley[3], and Eva Ceulemans[1]
[1] Research Group of Quantitative Psychology and Individual Differences, KU Leuven
[2] Methodology of Educational Sciences Research Group, KU Leuven
[3] Center for Contextual Psychiatry, KU Leuven

The experience sampling method (ESM) plays a pivotal role in investigating the dynamics of psychopathological processes in daily life. A crucial question when designing ESM studies concerns the sample size needed, defined by the number of participants (N) and the number of measurement occasions per participant (T). Higher N and T increase power, but also increase researcher and participant burden, and study cost. Current approaches for sample size planning rarely account for these feasibility and financial constraints explicitly, despite significant variations in ESM studies' design, operational expenses, participant incentives, and compliance rates. This oversight can lead to suboptimal or unrealistic sample size planning. In this article, we extend the traditional power analysis framework to integrate budget constraints into sample size decisions. In particular, we demonstrate how to formalize budget considerations into cost functions for ESM studies and how to use these to optimally select N and T values. Through an illustrative example, we showcase how optimal sample size decisions strongly differ across ESM designs and associated cost functions, even when focusing on the same research questions.

Public Significance Statement
This study provided experience sampling method researchers with a method to account for different types of study costs when deciding on the number of participants and number of momentary questionnaires. This method allows to optimize resource management and may lead to more reliable studies.

Keywords: experience sampling method, optimal sample size planning, study cost, power analysis, compliance

Supplemental materials: https://doi.org/10.1037/pas0001409.supp

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

Clinical researchers have embraced the experience sampling method (ESM) as a key tool to shed light on within-person psychopathological processes in an ecologically valid way (Myin-Germeys et al., 2018; Myin-Germeys & Kuppens, 2021). In an ESM study, researchers send multiple (e.g., seven times a day for a week) brief questionnaires to each participant in order to track psychological constructs (e.g., mood, symptoms, behaviors) across time, while participants go about their daily lives. Given that disruptions of emotional experiences are a key aspect of psychopathology, researchers have used ESM to capture and quantify emotional experience features and to examine whether and how these features differ across persons (Abel & Minor, 2023; Boemo et al., 2022; Joseph et al., 2021; Kaurin et al., 2023; Trull & Ebner-Priemer, 2020; Zelenski & Larsen, 2000). Many studies have, for instance,

Kristin Naragon-Gainey served as action editor.
Jordan Revol https://orcid.org/0000-0001-5511-3617
Ginette Lafit https://orcid.org/0000-0002-8227-128X
Olivia Kirtley https://orcid.org/0000-0001-5879-4120
Eva Ceulemans https://orcid.org/0000-0002-7611-4683

For reproducibility purposes, the files required to conduct the simulations are available on GitLab (https://gitlab.kuleuven.be/ppw-okpiv/researchers/u0148925/cost-efficient-esm) and the Open Science Framework (https://osf.io/f9pqs/). We also provide Supplemental Material A, which applies the cost function approach to an alternative research question and statistical model. Supplemental Material B guides the readers through the detailed R code, step-by-step, to replicate the analysis of this article. Finally, the repository comes with the complete Git history of the article.

The authors declare that there were no conflicts of interest with respect to the authorship or the publication of this article. The research presented in this article

was supported by research grants from the Fonds voor Wetenschappelijk Onderzoek–Vlaanderen (Grant G0C9821N), the Research Council of KU Leuven (Grants C14/23/062 and iBOF/21/090), and the Excellence of Science program (Grant EOS 40007528/G0I2422N) awarded to Eva Ceulemans.

Jordan Revol played a lead role in conceptualization, formal analysis, methodology, software, visualization, and writing–original draft and an equal role in writing–review and editing. Ginette Lafit played a supporting role in software and an equal role in conceptualization, methodology, supervision, and writing–review and editing. Olivia Kirtley played an equal role in conceptualization, methodology, and writing–review and editing. Eva Ceulemans played a lead role in funding acquisition, supervision, and writing–review and editing and an equal role in conceptualization and methodology.

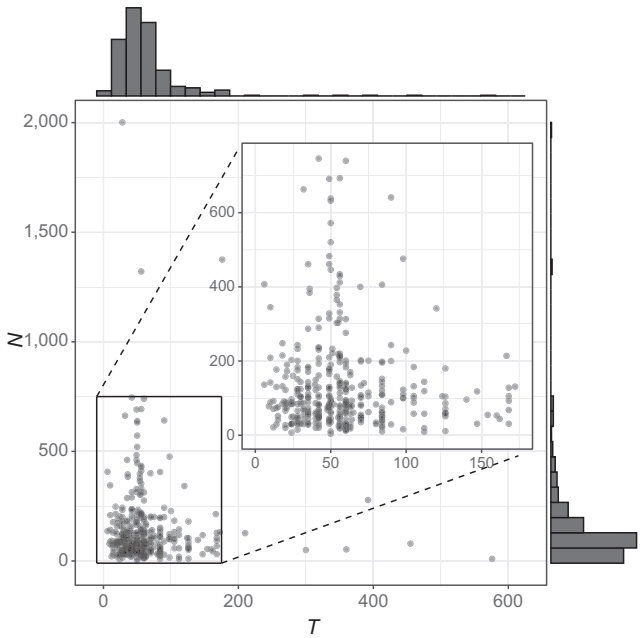
Correspondence concerning this article should be addressed to Jordan Revol, Research Group of Quantitative Psychology and Individual Differences, KU Leuven, Tiensestraat 102, Box 3713, 3000 Leuven, Belgium. Email: jordan.revol@kuleuven.be

focused on emotional inertia, which represents how long positive and negative emotions persist over time in daily life (Kuppens et al., 2010), and tested whether emotional inertia differs significantly across clinical and nonclinical populations (Brose et al., 2015; Thompson et al., 2012; Weermeijer et al., 2022).

Designing a (clinical) ESM study remains challenging because it involves many decisions. A crucial ESM design aspect that has recently attracted significant attention concerns the appropriate sample size (Kirtley et al., 2021; Lafit et al., 2021; Trull & Ebner-Priemer, 2020; Tuerlinckx et al., 2025; see also Nosek et al., 2022). In ESM studies, this involves making a decision regarding two dimensions: the number of participants (N) and the number of measurement occasions per participant (T). Figure 1, which is based on Wrzus and Neubauer's (2022) meta-analysis, demonstrates that selected T and N values considerably vary across ESM studies (see also Ammerman & Law, 2022; Bell et al., 2024; Fernández-Álvarez et al., 2024; Gee et al., 2020; Hall et al., 2021; J. J. Janssens et al., 2024; K. A. M. Janssens et al., 2018; Perski et al., 2022; Vachon et al., 2019). Since the selected values of N and T influence the quality and generalizability of study findings, researchers are urged to justify their sample size decisions (Kirtley et al., 2021; Lakens, 2022; Trull & Ebner-Priemer, 2020; see also Munafò et al., 2017). Traditionally, this justification is based on power analysis. The results of a power analysis can be visualized using a power contour plot (Baker et al., 2021; Westfall et al., 2014), representing power as a function of T and N (see Figure 2 for an example that we elaborate on in the remainder of this article). Such a contour plot typically reveals that multiple combinations of N and T can achieve a desired power level. For instance, as shown in Figure 2, to achieve 80 power, one could select $N = 90$ and $T = 127$, $N = 115$ and $T = 54$, or $N = 150$ and $T = 31$, all of which provide sufficient power for the research problem under investigation.

Figure 1

N and T of the Experience Sampling Method Studies Coded in a Meta-Analysis of Wrzus and Neubauer (2022)



Given that multiple N and T combinations yield sufficient power, how should one choose among them? We argue that researchers should keep in mind that the different combinations of N and T have different financial implications. A higher number of measurement occasions (T) for instance increase the total number of momentary questionnaires to complete and thus the time participants need to commit to the study. While incentives have received the most attention in the literature, primarily from a motivational and compliance perspective (Dejonckheere et al., 2024; Ludwigs et al., 2020; Wrzus & Neubauer, 2022), sources of financial costs in ESM studies are more diverse (Myin-Germeys & Kuppens, 2021), as they, among other things, also include hardware, software, and the workforce (e.g., the salary of student research assistant). Costs may rise further when running ESM studies in clinical populations (Deakin et al., 2022; Gabriel et al., 2019; Gross & Malzhacker, 2023; Piccirillo et al., 2024) because additional resources have to be allocated to recruit vulnerable populations or to improve participants' response rates and minimize dropout (e.g., flexible scheduling options, increased participant support, and trouble-shooting). It is thus important to account for cost considerations when designing one's (clinical) ESM study.

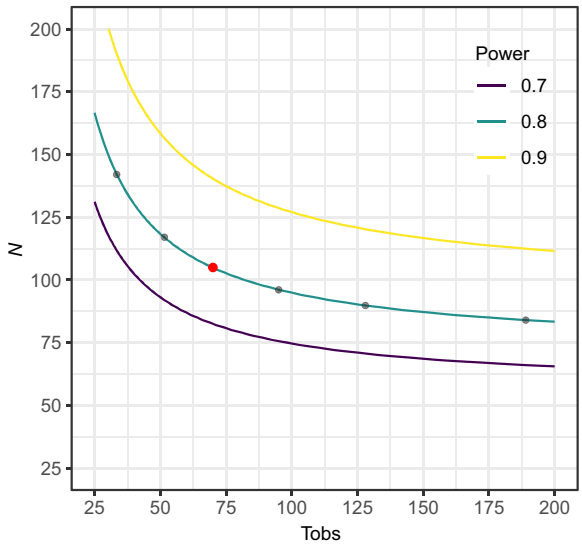
Many researchers have already argued decades ago that study costs should be accounted for when deciding on the sample size of psychological studies (Allison et al., 1997; Bacchetti et al., 2008; Brogden, 1949; Cascio & Silbey, 1979; Guo & Luh, 2013; Jóźwiak & Moerbeek, 2013; Liu & Colditz, 2017; Moerbeek, 2008, 2020, 2023; Schmidt et al., 1982; Tekle et al., 2011; Van Breukelen, 2013; Zimmer et al., 2023; Zitzmann et al., 2022). This can be achieved by implementing a cost function in one's power analysis, which allows researchers to identify the most cost-efficient study design that achieves the desired statistical power (e.g., Zitzmann et al., 2022). However, how to do this for ESM has not been investigated before. This is indeed challenging, given that ESM studies exhibit substantial heterogeneity in their design characteristics and sources of financial costs. This heterogeneity implies that each study requires its own tailored cost function. Yet, developing and implementing such tailored cost functions would allow ESM researchers to identify the most cost-efficient combination of N and T that achieves the desired statistical power, thereby optimizing resource allocation while minimizing potential resource waste.

In this article, we aim to assist ESM researchers by explaining how they can translate their ESM designs into study-specific cost functions. We will also demonstrate how to integrate these cost functions into their power analysis. To illustrate this translation process, we will use a specific research question that is studied by three different researchers, each employing a different design. This question focuses on the impact of a person-level variable on a person-specific model parameter: Does the emotional inertia of participants with major depressive disorder (MDD) differ significantly from that of healthy participants (i.e., participants without clinical diagnosis)? This example illustrates how optimal sample size decisions (i.e., decisions on T and N) depend on the underlying ESM study design and associated costs. To emphasize that the cost function approach is generic and can be used for a variety of research questions, we provide in Supplemental Material A the implementation steps and results for a second research question examining the association between two time-varying variables: Is there a contemporaneous relationship between anhedonia and negative affect (NA)

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

Figure 2

Example of a Power Contour Plot for a Specific Research Question



Note. See method section, step 3 power analysis for further clarification. The axes represent the number of participants (N) and the number of measurement occasions (T). The contour lines delineate *N* and *T* combinations for which power equals 70%, 80%, and 90%. The black dots represent various combinations of *N* and *T* that achieve 80% power. See the online article for the colour version of this figure.

in daily life? A discussion about the advantages, the limitations, and the future directions will close this article.

Method

In this section, we discuss the different steps an ESM researcher has to take to include cost considerations in a power analysis. First, researchers have to determine the research question for which they want to run the power analysis and the study design they want to implement. Second, researchers have to select the statistical model that will be used to answer their research question. Third, we recapitulate how to perform a power analysis and create a power contour plot, which provides clear insights into the relationship between sample size and power. Fourth, we introduce a novel step in the sample size determination procedure: integrating an ESM cost function that accounts for four costs sources, including participant incentives. We demonstrate how to incorporate this cost function into power analysis simulations and how power contour plots can be enriched with cost considerations. Finally, we show how to select *N* and *T* for one's study considering both power and cost constraints. To show the impact of various design choices and the associated differences in constraints, we discuss these five steps by introducing three different researchers—Peter, Laure, and Jean—who aim to investigate the same research question but envision quite different designs. All the materials used, including the scripts to run the power analysis and create the figures, are available on the Open Science Framework (Foster & Deardorff, 2017) at https://osf.io/f9pqs/, and GitLab at https://gitlab.kuleuven.be/ppw-okpiv/researchers/u0148925/cost-efficient-esm.[1] Additionally, the illustrations involve secondary analysis of data collected by Heininga et al. (2019) that received ethical approval.[2] Finally, this study was not preregistered.

Step 1: Research Question and Study Design

In this first step, researchers must specify their research question, and envisioned ESM design, as well as identify all potential sources of cost. Peter, Laure, and Jean are three researchers from three different research groups, who all want to study the effect of depression on emotional inertia in daily life.

Research Question

Previous research found that individuals diagnosed with MDD exhibit strong difficulties in regulating negative emotions throughout the day (Brose et al., 2015; Pe et al., 2015; Thompson et al., 2012, 2021). In other words, Peter, Laure, and Jean expect that persons diagnosed with MDD have stronger negative emotional inertia when compared with a nonclinical population.

ESM Design

To answer this research question, Peter, Laure, and Jean have each designed an ESM study, of which the specifics are outlined in the table of Table 1. The three studies have the following aspects in common. Participants diagnosed with MDD will be recruited from nearby hospitals, while the control group participants will be recruited in the general population through flyers. All participants will attend a 30-min briefing to receive instructions for the ESM study (e.g., how to install the ESM app on their smartphone). Subsequently, ESM questionnaires will be administered 10 times a day between 10:00 and 20:00 for 2 weeks at maximum, with an average between-questionnaire interval of 60 min. Upon completion of the ESM period, participants will receive feedback and compensation based on their compliance rate.

The three studies also differ in a number of cost-relevant ways. Peter needs to pay for an ESM data collection software (30 euros per participant). He has decided to provide a compensation bonus to participants who achieve 80% or higher compliance. His total study budget is capped at 7,000 euros.

Laure has planned to combine the ESM study with a qualitative approach (e.g., Eisele et al., 2023). Specifically, participants will attend 90-min interviews both before and after the ESM data collection period. Upon completion of the ESM period, participants will receive compensation for each completed ESM questionnaire. A student research assistant will be hired at 15 euros per hour to recruit participants, conduct interviews, facilitate briefings, and monitor and regularly check the data collection. The study involves

[1] Simulations were conducted using R (Version 4.4.1, R Core Team, 2024). We used the following R packages: for data manipulation and management: dplyr (Wickham et al., 2023), tidyr (Wickham et al., 2024), data.table (Barrett et al., 2024), and tibble (Müller & Wickham, 2023); for visualization: ggplot2 (Wickham, 2016), ggpubr (Kassambara, 2023), ggExtra (Attali & Baker, 2023), cowplot (Wilke, 2024), and viridis (Garnier et al., 2024); for statistical modeling: nlme (Pinheiro et al., 2023), GLMMadaptive (Rizopoulos, 2023), and MASS (Venables & Ripley, 2002); for utility functions: grid (R Core Team, 2024), latex2exp (Meschiari, 2022), flextable (Gohel & Skintzos, 2024), kableExtra (Zhu, 2024), officer (Gohel et al., 2024), gtable (Wickham & Pedersen, 2024), magick (Ooms, 2024), and parallel (R Core Team, 2024).

[2] KU Leuven Social and Societal Ethics Committee (S58526) and the KU Leuven Medical Ethics Committee (B322201627414).

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

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

Table 1
Study Design Characteristics

Category	Peter	Laure	Jean
Research question	Effect of depression on emotional inertia in daily life		
Hypothesis	Persons diagnosed with MDD have stronger negative emotional inertia when compared with persons without a clinical diagnosis.		
Model	First-order multilevel autoregressive model(1) with dichotomous Level 2 variable (MDD)		
Recruitment	• MDD: Nearby hospitals' health services • Control: Flyers in general population		
ESM	• 30 min online briefing • 10 beeps a day between 10 and 20 hr • Average time interval of 60 min • Simple feedback		
Additional tasks		• In-person 90-min interview before and after study • Task: Monitoring and checking data collection	Every morning: Complete 30-min cognitive tasks
Incentives	80% compliance or higher earning an additional bonus	Paid for each completed beep	80% compliance or higher earning an additional bonus
Others	ESM software license (30€/participant)	• Student research assistant (15€/hr) • ESM software license (15€/participant) • Laptop (1,500€)	• Student research assistant (20€/hr) • Programming ESM software • Laptop (1,500€)
Maximum budget	7,000€	16,000€	21,000€

Note. MDD = major depressive disorder; ESM = experience sampling method.

two equipment costs: an ESM software license (15 euros per participant) and a laptop for the research assistant (1,500 euros). Overall, Laure's maximal budget is 16,000 euros.

Jean's ESM study is part of a larger project investigating the associations between MDD, negative emotional inertia, and various cognitive capacities that are measured daily (see Schmiedek et al., 2010). Specifically, participants will visit the laboratory (or the day hospital for the clinical sample) to complete 30-min cognitive tasks every morning, before receiving their first ESM survey of the day. Due to the significant burden of the study, Jean wants to provide substantial incentives for the participants, irrespective of their compliance. Moreover, as in Peter's study, the participants who achieve 80% or higher compliance will earn a compliance bonus. To recruit participants, conduct briefings and interviews, and administer cognitive testing procedures, a student research assistant will be recruited at a higher rate of 20 euros per hour, reflecting the more specialized skills necessary for this study. Additionally, the student research assistant will dedicate three full-time weeks to programing the ESM study in the designated software. Regarding materials, Jean needs to buy a laptop (1,500 euros) for his student research assistant. Jean's total budget for the study is capped at 21,000 euros for his study.

Step 2: Select the Statistical Model

In this step, researchers must select an appropriate statistical model that aligns with their research questions. Peter, Laure, and Jean's research question involve estimating and testing if inertia of NA differs between a control group and a group diagnosed with MDD. Inertia, usually conceptualized as a person-specific effect, is estimated by an autoregressive parameter (i.e., regression of NA at time t on NA at time $t-1$). The impact of MDD, a binary group-level variable, on this person-specific effect can be assessed by examining its interaction with the autoregressive parameter. Hence, Peter, Laure, and Jean have both decided to use a multilevel regression model with between-group differences in the mean autoregressive effect (see Heininga et al., 2019). The equation of the model is

$$\begin{aligned} NA_{it} = {} & \beta_{00} + \beta_{10}NA_{it-1} + \beta_{01}MDD_i + \beta_{11}MDD_iNA_{it-1} \\ & + \nu_{0i} + \nu_{1i}NA_{it-1} + \varepsilon_{it}, \end{aligned} \tag{1}$$

where $i = (1, \ldots, N)$ denotes the person and $t = (1, \ldots, T)$ the measurement occasion. MDD is a dummy coded variable with 0 denoting the control group and 1 the group diagnosed with MDD. Hence, whereas β_{00} is the fixed intercept of the model,[3] $\beta_{00} + \beta_{01}$ represents the intercept for the MDD group. The random intercept $\nu_{0,i}$ captures the person-specific deviation from its group's intercept value. The mean autoregressive effect of the control group is captured by β_{10}, while $\beta_{10} + \beta_{11}$ represents the mean autoregressive effect of the group diagnosed with MDD. The person-specific deviation of the autoregressive effect from its group's value is quantified by ν_{1i}. Finally, ε_{it} are the within-person errors, theoretically representing various events occurring in daily life that influence NA and cannot be explained by the previous value of NA (NA_{it-1}). The within-person error is assumed to follow a Gaussian distribution centered on zero with variance σ^2_ε.

[3] It is crucial to note that the intercept in this model should not be confused with the mean of the process for a person i (see Lütkepohl, 2005).

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

This model also has a set of common assumptions, including the errors being identically and independently distributed, the time intervals between the measurement occasions being equal, and stationarity (for a comprehensive discussion, see Lafit et al., 2021). Note that the values of the model parameters (i.e., intercept, autoregressive effect, error terms) depend on the size of the interval between measurement occasions, implying that changing the time interval changes the parameter values. Finally, as commonly done in this type of analysis, Peter, Laure, and Jean decided to person-mean center the predictor NA_{it-1} (see Hamaker & Grasman, 2015).

Given Peter, Laure and Jean's research question, the parameter of interest is β_{11} as this parameter quantifies the extent to which being diagnosed with MDD is associated with increased inertia of NA. Hence, if β_{11} is significantly larger than zero, this would support Peter, Laure, and Jean's hypothesis.[4] Different tools and estimation methods can be used to fit this model (see Lafit et al., 2022). Peter, Jean, and Laure have decided to rely on the restricted maximum likelihood estimation implemented in the "lme" function from the nlme R package (Pinheiro et al., 2023).

Step 3: Power Analysis

Power is rooted in the hypothesis testing framework and represents the probability of detecting a true effect when it exists in the population under study (Cohen, 1977). To estimate power for a parameter of interest, researchers use power analysis. Whereas in other branches of psychology one can work with analytical formulas, in ESM research one usually works with simulations (see Lafit et al., 2023). This approach offers great flexibility making it possible to mimic aspects of the study (e.g., sampling scheme) and real-world conditions (e.g., missing data), and, more importantly, to implement any type of statistical model. Notably, software applications to estimate power for longitudinal multilevel models frequently use this approach (see Arend & Schäfer, 2019; Bolger et al., 2012; Lafit et al., 2021).

The simulation-based power analysis follows four substeps (Revol et al., 2024). The first substep involves specifying the necessary parameter values, including the sample size parameters, the simulation parameters, and the hypothesized model parameters. The sample size parameters are the number of measurement occasions T and the number of participants N. Given that we sample participants from two populations, N comprises N_0 control group participants and N_1 MDD-diagnosed participants. We will set $N_0 = N/2$ and $N_1 = N/2$.[5] The simulation parameters encompass the number of replicates (R, commonly 1,000), and the α value (α, commonly .05). R determines the number of artificial data sets generated in Substep 2, while α sets the statistical significance threshold of the hypothesis tests in Substep 3. The hypothesized model parameter values represent the researcher's best estimate of the true underlying values. Defining those values is often considered a bottleneck because they are difficult to determine (Lane & Hennes, 2018). A common practice to address this challenge is to estimate them using an existing data set (Lane & Hennes, 2018). Importantly, it is preferable that the data set closely matches the characteristics of the planned ESM study (e.g., items, time interval between beeps given its impact on the model parameters) and has a sufficiently large sample size to ensure reliable estimates (Albers & Lakens, 2018). In Jean and Laure's case, they were able to identify a suitable data set. This data set, collected by Heininga et al. (2019), contains the data of 40 control participants and 38 participants diagnosed with MDD. Participants took part in a 7-day ESM study, responding to 10 daily beeps with a fixed time interval of 66 min. At each of the 70 measurement occasions, participants rated 27 items including NA and anhedonia (variable used in Supplemental Material A) with a slider scale ranging from 0 (*not at all*) to 100 (*very much*). MDD severity was measured during a briefing interview with the Quick Inventory of Depressive Symptomatology (Rush et al., 2003). The data set is available in the EMOTE database (Kalokerinos et al., 2025).

In the second substep, the statistical model with the associated parameter values and the specified sample size is used to generate R artificial data sets. Importantly, this data generation process can incorporate various scenarios to better reflect real-world conditions. One such scenario, the occurrence of missing values, will be discussed in detail in a subsequent section. Another important assumption pertains to the time intervals between measurement occasions being equal. This assumption is violated by the night break, as no assessments are collected between 8 p.m. and 10 a.m. To address this issue, researchers typically employ within-day lagging of the outcome variable and avoid using the final observation of 1 day as a predictor for the first observation of the following day. This implies that if a protocol includes 10 measurement occasions per day, only the last nine observations per day are used as outcomes in the estimation procedure. Since this impacts power, we incorporated it into our simulation. Specifically, we generated a time series with 24 equidistant measurements per day and systematically excluded 14 consecutive ones after every sequence of 10 observations, thereby simulating the overnight measurement break.

In the third substep, the statistical model is fitted to each generated data set. Based on the chosen statistical significance value (α), it is determined whether the null hypothesis (H_0) associated with the parameter of interest is rejected for each replicate. The final substep involves estimating the empirical power. It is calculated by computing the proportion of replications for which H_0 that was rejected. While a power of 80% is often considered sufficient (for a discussion, see Giner-Sorolla et al., 2024), low power is expected to lead to biased results and compromises the replicability of studies, which has been one of the major factors of the recent replication crisis (Button et al., 2013; Fraley & Vazire, 2014; Koch et al., 2023; Munafò et al., 2017).

Power Contour Plot

To best visualize the power analysis results, one can build a power contour plot (Baker et al., 2021; Westfall et al., 2014). Creating a

[4] One may wonder whether the estimated autoregressive effect could be biased. Since this parameter is associated with a lagged dependent variable, bias can indeed arise when the available number of measurements is small (see dynamic panel bias, Falkenström, 2024; Zyphur et al., 2020), due to small sample sizes or low compliance rates. While bias is not assessed in the present illustrations for reasons of simplicity, it is expected to be negligible given the considered sample sizes (i.e., minimal sample size is $N = 50$ and $T = 50$) and compliance rates (88%). Note that bias can be computed from the simulation-based results (see Lafit et al., 2021).

[5] When N is an odd number, N_0 is rounded at the floor value and N_1 is rounded at the ceiling value.

power contour plot begins by establishing a large grid based on the two sample size parameters N and T. Then, we conduct a power analysis[6] for each combination of N and T within this grid. For instance, Peter, Laure, and Jean consider a range of values for both N and T from 50 to 200 with steps of five (see Figure 3). The results of the power analysis are then used to generate power contours using the R function "contourLines()" from the grDevices package (R Core Team, 2024). A power contour is a line that delineates all combinations of N and T that achieve a specific power value (e.g., 0.8). In Figure 3, the power contours are represented by continuous colored lines, respectively, purple for 70% power, blue for 80% power, and yellow for 90% power. The power contours suggest that increasing N has a greater impact on power than increasing T. This implies that prioritizing N over T would be more advantageous for power gain. Hence, if Peter, Laure, and Jean were to prioritize statistical power alone, they might both select, for instance, $N = 165$ and $T = 52$ achieving 80% power. However, other combinations are also viable, such as $N = 75$ and $T = 200$, $N = 100$ and $T = 163$, or $N = 130$ and $T = 83$.

Step 4: Cost Function

While power contours help in pinpointing which sample size combinations can achieve a targeted power, they do not shed light on the financial feasibility or cost-effectiveness of these options. This limitation can be addressed by incorporating a cost function into the power contour.

A cost function is a mathematical formula that integrates various types of cost sources that can, for instance, be tied to including additional participants or adding more measurement occasions (Zitzmann et al., 2022). By combining these costs with specific N and T values, the cost function allows to compute the total study cost and thus to differentiate different sample size options in terms of cost-effectiveness. In the context of an ESM study, we argue that a cost function will usually comprise at least four terms (see Figure 4).

Fixed Costs (c_1)

The first term is a constant that is the sum of all fixed costs associated with an ESM study. This cost is thus independent of sample size, but ought to be acknowledged when computing the total cost of a study as it impacts financial feasibility. It includes materials (e.g., buying a new computer, buying equipment for briefing sessions), software licenses, training, and other one-time expenses necessary for study setup. In our example, Peter does not have any fixed costs ($c_1^P = 0$). Laure has to spend 1,500 euros for a new laptop ($c_1^L = 1{,}500$). Beyond buying a new laptop, Jean's study also requires the student research assistant to program the ESM study in the software, which is estimated to take 3 weeks of work (114 hr), hence 2,280 euros ($c_1^J = 3{,}780$).

Participant Cost (c_2)

The second term represents the cost per participant, implying that this cost directly scales with the number of participants (N) in the study. It may include material purchases for separate participants (e.g., smart watches for data collection), outsourced recruitment fees (e.g., payment to external recruitment agencies), travel reimbursements for participants visiting the laboratory, or personnel costs for student research assistants conducting participant-specific tasks (e.g., briefing, debriefing, interviews). In our example, Peter will recruit master's thesis students to handle the data collection for free; however, the ESM data collection software membership costs 30 euros per participant ($c_2^P = 30$). In Laure's case, programing the sampling schedule in the data collection application costs 15 euros per participant. Additionally, the 3 hr of interview and 1 hr of (de)briefing will be handled by a student research assistant paid 15 euros per hour, contributing 60 euros to the per-participant cost. Hence, the total cost per participant in Laure's study is 75 euros ($c_2^L = 75$). For Jean's study, the primary expense was the 30-min briefing by the student research assistant, paid at a rate of 20 euros per hour. This results in a per-participant cost of 10 euros ($c_2^J = 10$).

Measurement Occasion Cost (c_3)

The third term represents the cost per measurement occasion per participant. In other words, every additional measurement of a participant would increase the total cost. This c_3 term may include momentary expenses, such as the cost of sending a momentary questionnaire (which may vary depending on software subscription terms), or cumulative costs that increase with the number of measurement occasions (T), such as the time required for monitoring participants throughout the ESM data collection period. Importantly, all these costs must be expressed as a per-measurement cost. For example, daily costs (e.g., morning laboratory procedures) can be converted to a per-measurement cost by dividing them by the total number of beeps per day.

Peter, Laure, and Jean have to account for per measurement costs since they want to perform patient and data collection monitoring. Specifically, they want to consider two tasks: monitoring and managing the data collection process (labeled *Mon*; e.g., checking ESM collection progress on the online platform) and responding to participant questions (labeled *Que*; e.g., emails, requests, phone calls). Those tasks are performed by the student research assistant, hence, their cost is directly associated with the time spent by the student research assistant. This time can be estimated based on previous studies, considering the daily time required for monitoring data collection progress and the average time spent responding to each participant's inquiries across the study.

Peter does not have to take this cost into account as the monitoring of the data collection will be handled by master's thesis students ($c_3^P = 0$). In contrast, for Laure's previous study, checking the data collection progress took, on average, 5 min per day, and responding to participants' questions took on average 7 min per participant. This results in $Mon = 0.141$ euros, $Que = 0.025$ euros, and $c_3^L = 0.166$.

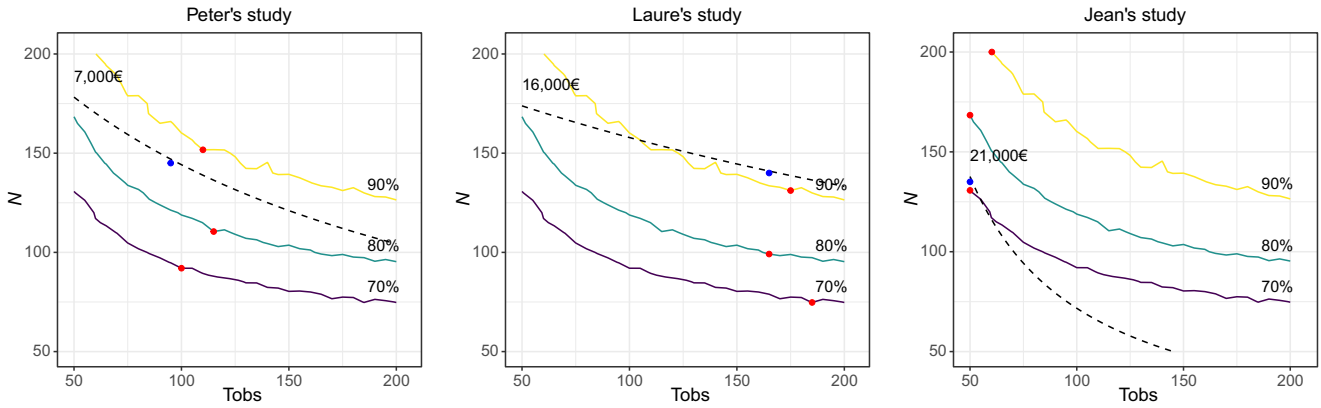
For Jean's previous study, checking the data collection progress took, on average, 10 min per day, and responding to participants' questions took on average 5 min per participant. This results in $Mon = 0.233$ euros and $Que = 0.017$ euros. In addition to these costs, daily laboratory session costs must also be considered, which amounted to a daily cost of 10 euros per participant. With 10

[6] To enhance the precision of the empirical power estimates, we decided to increase the number of R replicates from the usual 1,000 to 10,000.

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

Figure 3

Cost Enriched Power and Cost Contour Plots Obtained for Peter, Laure, and Jean's Studies



Note. The red dot indicates the most cost-effective combinations of N and T that achieve 70%, 80%, and 90% power. The blue dot indicates the highest achievable power within Peter, Laure, and Jean's budgets. The solid lines are the power contours, respectively, representing 70%, 80%, and 90% power; note that these power contours are the same in the three plots. The dashed cost contour line reflects the researcher's maximum budget. See the online article for the colour version of this figure.

measurement occasions per day, this translates to a cost of 1 euros per measurement occasion, consequently, $c_3^J = 1.25$.

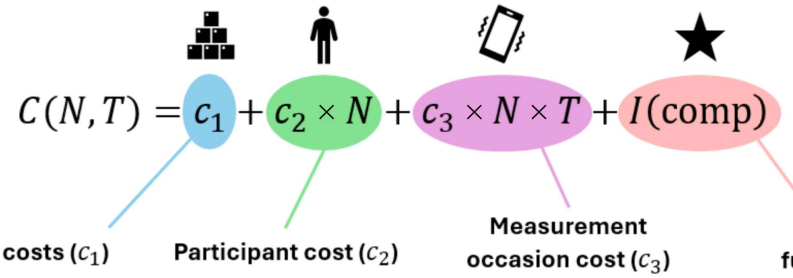
Incentive Function

The final term represents the expenses associated with compensating study participants (N), to increase participant motivation and compliance. This can take various forms (e.g., lottery, voucher) that can be formalized in a function, called the incentive function (I).

Peter, Laure, and Jean have adopted different approaches to participant compensation. In Peter and Jean's compensation scheme, each sent momentary questionnaire (whether answered or not) will yield a baseline reward of 0.15 euros for Peter and 0.8 euros for Jean. If at least 80% compliance is reached, participants respectively receive an additional 0.045 euros and 0.32 euros, per sent momentary questionnaire. If $T = 100$, Jean's participants will thus receive 80 euros (0.8×100) if they have a compliance lower than 80% and 112 euros if they reach 80% or more.

Figure 4

Summary of the Cost Sources and Associated Cost Parameter Values for Peter, Laure, and Jean



	Fixed costs (c_1)	Participant cost (c_2)	Measurement occasion cost (c_3)	Incentive function ($I()$)
Examples	• Computer • Advertisements (e.g., flyers) • Participant screening time • ESM platform/software • Server infrastructure • Project coordination	• Briefing session • Booking fee for research room • Interview • Professional recruitment agency (per pp) • Baseline and post-study measurements • Travel compensation	• Data collection monitoring • Technical support • Equipment usage/year • Daily tasks in the lab • Consumables	• Base participation payment • Per-measurement completion reward • Compliance bonus (e.g., >80%) • Lottery • Time-of-day bonus
Peter		• Data collection software: 30€ • Online briefing: 0€	• Monitoring and answering participants' questions: 0€	• Baseline reward per questionnaire: 0.15€ • If >80% compliance, 0.195€/measurement
Laure	• New laptop: 1,500€	• Data collection software: 15€ • Interview: 45€ • (De)briefing: 15€	• Monitoring: 0.141€ • Answering participants' questions: 0.025€	• Per completed questionnaire: 0.5€
Jean	• New laptop: 1,500€ • Program ESM study: 2,280€	• Briefing: 10€	• Cognitive tests: 1€ • Monitoring: 0.233€ • Answering participants' questions: 0.017€	• Baseline reward per questionnaire: 0.8€ • If >80% compliance, 1.12€/measurement

Note. ESM = experience sampling method; pp = participants. See the online article for the colour version of this figure.

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

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

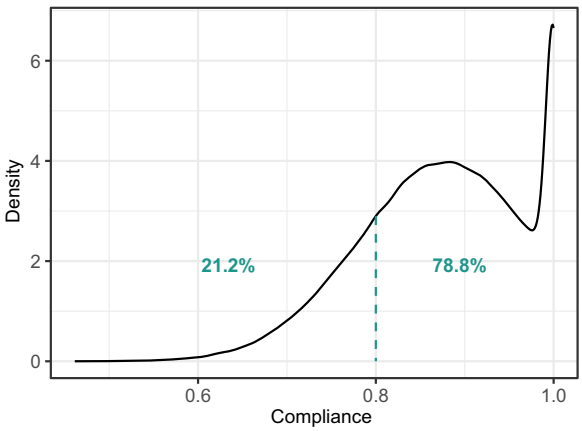
Laure's compensation scheme is directly tied to the number of momentary questionnaires answered by each participant at the end of the study (i.e., compliance rate). For every questionnaire answered, the participant receives 0.5 euros. For instance, if $T = 50$ and the participant responds to 35 of them, this results in a compliance rate of 70% and a total compensation of 17.5 euros. But how can one now compute the value of this fourth term? Before we explain that, we revisit the power analysis.

Implications of the Incentive Function: Revisiting the Power Analysis (Step 3)

As the previous paragraphs show that expected compliance may strongly impact the cost of a study design, it must be implemented in the power analysis in Substep 3, when simulating the replicates. To this end, we assume that participant compliance follows a truncated normal distribution. The truncated implies that the distribution is bounded within the interval [0, 1], as compliance rates cannot be negative or exceed 100%. The parameters of the distribution (i.e., mean and standard deviation) can either be set based on literature or estimated from data set(s) with similar characteristics to the study at hand. Following this second option, Laure and Jean decided to again use the clinical data set and estimated its compliance distribution. The analysis yielded a mean compliance of 0.88 with a standard deviation of 0.1 (see Figure 5). This distribution is used in the second substep of the power analysis to draw compliance scores for each simulated participant of each data set.

Based on the drawn compliance score, a proportion (1 − compliance) of measurement occasions from this participant is set as missing. This requires assumptions about the missingness pattern (Simsa et al., 2024; van Buuren, 2018). Laure and Jean adopt the simplest assumption, assuming a missing completely at random (van Buuren, 2018) pattern. Under missing completely at random, each observation has an equal probability of being missing, independently of any observed or unobserved variables.

Figure 5
Expected Compliance Distribution Derived From the Clinical Data Set



Note. The distribution is modeled as a bounded normal distribution within the interval [0, 1], with a mean of 0.88 and a standard deviation of 0.1. Following Jean's study, a compliance threshold is set at 0.8. Hence, 21.2% of participants are expected to fall below this threshold and 78.8% are expected to be above. See the online article for the colour version of this figure.

Computing the Cost Function Value for Specific T and N Values

Computing the first three terms of the cost function is easy given T, N, c_1, c_2, and c_3. The selected compliance distribution enables us to compute the fourth term, by calculating the expected compensation costs. Because Peter, Laure, and Jean follow different incentive strategies, their approaches to calculating these expected compensation costs differ. Laure has an incentive per questionnaire strategy in which the participant receives 0.5 euro ($\text{Inc}_{\text{questionnaire}}$) for each completed questionnaire. To calculate the expected cost per participant, one only needs to multiply the mean compliance (comp_{avg}), the number of measurement occasions (T), and the incentive per questionnaire ($\text{Inc}_{\text{questionnaire}}$). The total expected compensation cost is then derived by multiplying this result by the number of participants (N), as shown in Equation 2. For instance, for $N = 30$, $T = 50$, $\text{comp}_{\text{avg}} = 0.88$, and $\text{Inc}_{\text{questionnaire}} = 0.5$, the expected cost per participant is 22 and the expected compensation cost is 660 euros.

$$I(N,T) = N \times T \times \text{comp}_{\text{avg}} \times \text{Inc}_{\text{questionnaire}}. \tag{2}$$

In contrast, Peter and Jean use a compliance bonus strategy with an 80% threshold. Based on the compliance distribution (Figure 5), approximately 9.7% of participants are expected to fall below this threshold (Comp−), while 90.3% are expected to exceed it and get the bonus (Comp+). To calculate the expected cost per participant, one computes a weighted mean: one multiplies the probability of being below or above the compliance threshold by its associated incentive and sums these products. For Jean, if $T = 100$, Comp− participants receive 80 euros and Comp+ participants receive 112 euros. The expected cost per participant is thus $80 \times 0.21 + 112 \times 0.79 = 105.22$. The expected compensation cost of incentives is then obtained by multiplying this expected cost per participant by the number of participants (N). Hence, the expected compensation for 50 participants is $105.22 \times 50 = 5{,}261.03$ euros. The formula is given by

$$\begin{aligned} I(N,T) = N \times (&P(\text{Comp}-) \times \text{Pay}(T,\text{Comp}-) \\ &+ P(\text{Comp}+) \times \text{Pay}(T,\text{Comp}+)), \end{aligned} \tag{3}$$

with P(Comp−) and P(Comp+) denoting the probabilities of a participant falling into respectively the low (Comp−) or high (Comp+) compliance categories, and Pay(T, Comp−) and Pay(T, Comp+) representing the payment for a participant in the respective compliance category, given T measurement occasions. It is important to note that different study designs and pay schemes will lead to different incentive functions (see Supplemental Material B).

Adding the Cost Information to the Power Contour Plot

We can add the total cost information based on combining the different N and T combinations with the cost function to the power contour plot, which will yield different information for Peter, Laure, and Jean as they have different cost functions. Using this information,

we can determine two optimal sample sizes (Hecht et al., 2023; Manju et al., 2014; Zimmer et al., 2023). For a targeted power level (e.g., 80%), we can identify the combination of N and T that minimizes the total cost of the study (red dots in Figure 3), being $N = 111$ and $T = 115$ for Peter, $N = 100$ and $T = 165$ for Laure, and $N = 169$ and $T = 50$ for Jean, with a total cost of 5,697.48, 11,805.31, and 24,923.64 euros, respectively. On the other hand, for a fixed maximum budget, we can identify the combination of N and T that maximizes power (blue dots in Figure 3), being $N = 145$ and $T = 95$ achieving 0.869% power for Peter (6,904.8 euros), $N = 140$ and $T = 165$ achieving 0.913% power for Laure (15,898.44 euros), and $N = 135$ and $T = 50$ achieving 0.714% for Jean (20,669.89 euros). Additionally, one dashed maximum budget contour is drawn per researcher representing N and T combinations of which the costs equal their respective fixed maximum budgets.

With this enriched plot, a researcher has a comprehensive picture which allows to choose the optimal sample size for the study at hand taking into account financial considerations. While Peter, Laure, and Jean may opt for one of the two optimal points, they can also choose any N and T combination falling between the targeted power contour (e.g., 80%) and the maximal budget contour,[7] as all these combinations achieve sufficient power and remain within one's available budget.

We can further emphasize the importance of these differences in cost-efficiency balance, by noting that if Peter and Laure were to adopt Jean's optimal cost recommendation (red dot), their analysis would still achieve 80% power, but at a higher cost of 6,637.19 and 15,596.53 euros instead of 5,697.48 and 11,805.31 euros.

Step 5: Sample Size Decision

Remember that if Peter, Laure, and Jean were to prioritize statistical power alone, they would obtain the same power contours and might both select, for instance, $N = 165$ and $T = 52$ achieving 80% power. However, the cost-effectiveness analysis in Step 4 reveals a more nuanced picture. The optimal combination of N and T—whether minimizing costs to achieve 80% power or maximizing power within the budget—differs substantially across the three researchers. Moreover, while Jean's study achieves optimal cost-efficiency by prioritizing increases in N over T to maximize power, Laure's study achieves optimal cost-efficiency by prioritizing increases in T over N and Peter's study necessitates a more nuanced approach, balancing both N and T for maximum cost-effectiveness.

Having analyzed the outputs of the power and cost contour plot, Peter, Laure, and Jean must finalize their sample size decisions. Based on the power and cost contour plot, researchers choose any T and N combination between their 80% power contour (blue) and maximum budget contour (dashed line), as all these solutions yield a power of at least 80% while study costs remain within their maximum budgets. Peter aims to identify the optimal sample size that achieves 80% power with minimal cost. The power and contour plot analysis indicates $N = 111$ and $T = 115$ as being optimal, costing 5,697.48 euros.

Laure wants to maximize her chances of detecting the effect of interest and therefore considers the most cost-effective combination achieving 90% power. The power and contour plot analysis indicates $N = 100$ and $T = 165$ as optimal for her, costing 11,805.31 euros.

Jean, however, faces a more complex decision-making process. His maximal budget only allows for 71.4%, falling short of the desired 80%. Consequently, he can follow two decisions. He can opt for the highest achievable power within his budget, choosing $N = 135$ and $T = 50$, to achieve 71.4% power for 12,764.97 euros. Alternatively, he could consider increasing his budget from 12,764.97 to 15,596.53 euros to reach the most cost-effective combination of $N = 169$ and $T = 50$, achieving 80% power.

Discussion

This article discussed how to integrate cost considerations in power-based sample size planning methods for ESM studies. Researchers can leverage this approach and the associated power and cost contour plots to evaluate different potential ESM designs and investigate their impact on optimal sample size and study feasibility. This enables a more comprehensive assessment of methodological options before data collection. Specifically, through translating Peter, Laure, and Jean's scenarios into cost functions, we demonstrated that different study designs lead to substantially different costs for identical sample sizes. This example underscores the critical importance of considering costs in ESM sample size determination to ensure that planned research remains within one's budgetary constraints. The generalizability of the cost function approach is further demonstrated in Supplemental Material A, in which we focus on a different research question involving a time-varying predictor. Additionally, we provide guided R code in Supplemental Material B to show how to implement this approach.

While this article establishes the importance of considering cost in ESM sample size planning, further research is needed. We identify five limitations.

First, implementing our method requires some programing expertise. The programmer must be capable of coding simulation-based power analysis for the chosen statistical model, incorporating compliance expectations, creating the power and cost plot, and identifying optimal sample sizes. To make this methodology more accessible to researchers, future research can provide user-friendly software tools, such as a dedicated R package or an interactive Shiny application (similar to Lafit et al., 2021).

Second, running simulation-based power analysis across a large grid of N and T values might be computationally intensive and exceed the computation power of standard computers.[8] While it is theoretically possible to optimize the algorithm to reduce the computation time, to our knowledge, no method has been proposed to overcome this issue. Existing approaches either rely on analytical power calculations (Baker et al., 2021; Hecht et al., 2023; Westfall et al., 2014; Zitzmann et al., 2022), are restricted to one-dimensional sample size problems (Constantin et al., 2023), or only provide a single optimal sample size (Zimmer et al., 2023).

However, when the computational load is too high researchers can still implement cost functions into a restricted grid search, implying that they evaluate only a few values for both N and T (e.g., $N = 50, 100, 150$ and $T = 50, 70, 140$) and select the N and T combination that yields sufficient power at the lowest cost. We explain how to implement this approach through guided R code in Supplemental Material B. Although researchers should be aware

[7] Assuming the target power can be reached with the defined maximal budget, which is not the case for Jean's study.

[8] The simulation required to generate 3 was executed on a high-performance computing cluster, requiring 2,880.72 hr of CPU time.

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

that one has a high chance of selecting sample size recommendations that would not be optimal if a larger grid would have been considered, this restricted grid approach is still better than not considering costs at all.

Third, although the proposed cost function will fit many ESM studies, some relevant aspects might be missing. Dropout (Wrzus & Neubauer, 2022) could be considered an additional cost source, as participants may require replacement and thus factoring in extra participants to reach the planned N. Participant burden, another important topic in ESM research (Eisele et al., 2022; Rintala et al., 2023; Schreuder et al., 2020; Tate et al., 2024; Van Genugten et al., 2020), can also be integrated into the cost function. For instance, one could try to optimize sample size selection further and find the least burdensome ESM study design while accounting for power and cost. Alternatively, participant burden can be addressed through using clear thresholds (e.g., $T < 70$). Finally, incentive options are not limited to the two proposed pay schemes and can include, for instance, vouchers, or lotteries,[9] necessitating a case-by-case reconsideration of the incentive function ($I()$). For instance, a lottery could be expressed as $\mathbb{D}(\mathbb{D},\mathbb{D}) = \mathbb{D}_4\mathbb{D}[\frac{N}{G}]$ where c_4 denotes the per measurement cost of providing a prize, G denotes the number of participants per prize, and represents $[\frac{N}{G}]$ the total number of prizes that have to be provided.

Fourth, a ubiquitous problem in power analysis is how to safeguard sample size decisions against imprecise power estimates (Lafit et al., 2024). Indeed, the accuracy of power estimation can be significantly impacted by uncertainty about the hypothesized parameter values, potentially leading to underpowered studies. To mitigate this risk, researchers (Lafit et al., 2024; Perugini et al., 2014) have proposed using a safeguard power analysis, which involves using the lower or upper bound of the 95% confidence interval of the parameter of interest estimate in the power analysis procedure.

Finally, while our approach allows to account for the expected compliance and possible interindividual differences therein, this comes with the challenge that one has to specify a compliance distribution. This is not trivial indeed. Although the effect of incentives on compliance is not fully understood, compliance is commonly accepted to be dependent on the incentive structure (Wrzus & Neubauer, 2022). Consequently, deriving compliance estimates from previous studies with different incentive structures is not recommended. Future research is needed to better understand the relationship between incentives and compliance rates, and to develop more sophisticated approaches for integrating this information in the cost function approach. In this article, to simulate compliance differences, we included missing observations following a missing completely at random pattern (Rubin, 1976) when generating data in the power analysis. However, this approach might misrepresent a complex and not fully understood reality. While the effect of missing data has been extensively explored in the literature, the missingness patterns in ESM studies remain largely unknown (Simsa et al., 2024), limiting the reliability of simulation results. Nonetheless, implementing compliance in this manner already provides more realistic sample size recommendations. For instance, we can observe that the simulation results behind Figure 3 differ from the one behind Figure 2: at $T = 100$, 80% power is achieved with $N = 95$ in Figure 2, but $N = 119$ is required in Figure 3, despite using the same underlying statistical model and associated parameter values. This difference in recommendations is due to accounting for the participants' compliance distribution ($M = 0.88$, $SD = 0.1$) and night breaks in Figure 3, which result in multiple missing measurement occasions.[10] These missing values lead to a smaller sample size than indicated by T when estimating the model and thus to lower power.

Conclusion

Cost considerations can be integrated into ESM sample size determination by means of the proposed cost function. It allows ESM researchers to assess financial feasibility and optimize resource allocation. Moreover, it forces researchers to consider compliance issues and how they can be counteracted by incentives, resulting in more realistic power estimates.

[9] Some ethical committees do not allow incentives in ESM studies, viewing them as potentially coercive.

[10] The figures also differ in power estimation methods. Figure 2 uses an analytical approach (Lafit et al., 2023), which reduces computation time but cannot currently integrate missing data. Importantly, both methods yield similar power estimates (Lafit et al., 2023).

References

Abel, D. B., & Minor, K. S. (2023). Social pleasure in daily life: A meta-analysis of experience sampling studies in schizophrenia. *Schizophrenia Research*, *260*, 56–64. https://doi.org/10.1016/j.schres.2023.08.009

Albers, C., & Lakens, D. (2018). When power analyses based on pilot data are biased: Inaccurate effect size estimators and follow-up bias. *Journal of Experimental Social Psychology*, *74*, 187–195. https://doi.org/10.1016/j.jesp.2017.09.004

Allison, D. B., Allison, R. L., Faith, M. S., Paultre, F., & Pi-Sunyer, F. X. (1997). Power and money: Designing statistically powerful studies while minimizing financial costs. *Psychological Methods*, *2*(1), 20–33. https://doi.org/10.1037/1082-989X.2.1.20

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

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. https://doi.org/10.1037/met0000195

Attali, D., & Baker, C. (2023). *ggExtra: Add marginal histograms to 'Ggplot2', and more 'ggplot2' enhancements* (Version 0.10.1) [Computer software]. https://cran.r-project.org/package=ggExtra

Bacchetti, P., McCulloch, C. E., & Segal, M. R. (2008). Simple, defensible sample sizes based on cost efficiency. *Biometrics*, *64*(2), 577–585. https://doi.org/10.1111/j.1541-0420.2008.01004_1.x

Baker, D. H., Vilidaite, G., Lygo, F. A., Smith, A. K., Flack, T. R., Gouws, A. D., & Andrews, T. J. (2021). Power contours: Optimising sample size and precision in experimental psychology and human neuroscience. *Psychological Methods*, *26*(3), 295–314. https://doi.org/10.1037/met0000337

Barrett, T., Dowle, M., Srinivasan, A., Gorecki, J., Chirico, M., Hocking, T., & Schwendinger, B. (2024). *Data.table: Extension of 'data.frame'* [Computer software]. https://cran.r-project.org/package=data.table

Bell, I. H., Eisner, E., Allan, S., Cartner, S., Torous, J., Bucci, S., & Thomas, N. (2024). Methodological characteristics and feasibility of ecological momentary assessment studies in psychosis: A systematic review and meta-analysis. *Schizophrenia Bulletin*, *50*(2), 238–265. https://doi.org/10.1093/schbul/sbad127

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

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

Boemo, T., Nieto, I., Vazquez, C., & Sanchez-Lopez, A. (2022). Relations between emotion regulation strategies and affect in daily life: A systematic review and meta-analysis of studies using ecological momentary assessments. *Neuroscience & Biobehavioral Reviews*, *139*, Article 104747. https://doi.org/10.1016/j.neubiorev.2022.104747

Bolger, N., Stadler, G., & Laurenceau, J.-P. (2012). Power analysis for intensive longitudinal studies. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 285–301). Guilford Press.

Brogden, H. E. (1949). When testing pays off. *Personnel Psychology*, *2*(2), 171–183. https://doi.org/10.1111/j.1744-6570.1949.tb01397.x

Brose, A., Schmiedek, F., Koval, P., & Kuppens, P. (2015). Emotional inertia contributes to depressive symptoms beyond perseverative thinking. *Cognition and Emotion*, *29*(3), 527–538. https://doi.org/10.1080/02699931.2014.916252

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376. https://doi.org/10.1038/nrn3475

Cascio, W. F., & Silbey, V. (1979). Utility of the assessment center as a selection device. *Journal of Applied Psychology*, *64*(2), 107–118. https://doi.org/10.1037/0021-9010.64.2.107

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (Rev. ed.). Academic Press.

Constantin, M. A., Schuurman, N. K., & Vermunt, J. K. (2023). A general Monte Carlo method for sample size analysis in the context of network models. *Psychological Methods*. Advance online publication. https://doi.org/10.1037/met0000555

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), Article 669. https://doi.org/10.1186/s12888-022-04319-x

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

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

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

Falkenström, F. (2024). Time-lagged panel models in psychotherapy process and mechanisms of change research: Methodological challenges and advances. *Clinical Psychology Review*, *110*, Article 102435. https://doi.org/10.1016/j.cpr.2024.102435

Fernández-Álvarez, J., Colombo, D., Gómez Penedo, J. M., Pierantonelli, M., Baños, R. M., & Botella, C. (2024). Studies of social anxiety using ambulatory assessment: Systematic review. *JMIR Mental Health*, *11*, Article e46593. https://doi.org/10.2196/46593

Foster, E. D., & Deardorff, A. (2017). Open Science Framework (OSF). *Journal of the Medical Library Association*, *105*(2), 203–206. https://doi.org/10.5195/jmla.2017.88

Fraley, R. C., & Vazire, S. (2014). The N-pact factor: Evaluating the quality of empirical journals with respect to sample size and statistical power. *PLOS ONE*, *9*(10), Article e109019. https://doi.org/10.1371/journal.pone.0109019

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. https://doi.org/10.1177/1094428118802626

Garnier, S., Ross, N., Rudis, R., Camargo, P. A., Sciaini, M., & Scherer, C. (2024). *Viridis(Lite)—Colorblind-friendly color maps for R* (Version 0.6.5) [Computer software]. https://doi.org/10.5281/zenodo.4679423

Gee, B. L., Han, J., Benassi, H., & Batterham, P. J. (2020). Suicidal thoughts, suicidal behaviours and self-harm in daily life: A systematic review of ecological momentary assessment studies. *Digital Health*, *6*. https://doi.org/10.1177/2055207620963958

Giner-Sorolla, R., Montoya, A. K., Reifman, A., Carpenter, T., Lewis, N. A., Aberson, C. L., Bostyn, D. H., Conrique, B. G., Ng, B. W., Schoemann, A. M., & Soderberg, C. (2024). Power to detect what? Considerations for planning and evaluating sample size. *Personality and Social Psychology Review*, *28*(3), 276–301. https://doi.org/10.1177/10888683241228328

Gohel, D., Moog, S., & Heckmann, M. (2024). *Officer: Manipulation of Microsoft Word and PowerPoint documents* (Version 0.6.7) [Computer software]. https://cran.r-project.org/package=officer

Gohel, D., & Skintzos, P. (2024). *Flextable: Functions for tabular reporting* (Version 0.9.7) [Computer software]. https://cran.r-project.org/package=flextable

Gross, T., & Malzhacker, T. (2023). The experience sampling method and its tools: A review for developers, study administrators, and participants. *Proceedings of the ACM on Human–Computer Interaction* (Vol. 7, pp. 1–29). Association for Computing Machinery. https://doi.org/10.1145/3593234

Guo, J.-H., & Luh, W.-M. (2013). Efficient sample size allocation with cost constraints for heterogeneous-variance group comparison. *Journal of Applied Statistics*, *40*(12), 2549–2563. https://doi.org/10.1080/02664763.2013.819417

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*, Article 642044. https://doi.org/10.3389/fpsyg.2021.642044

Hamaker, E. L., & Grasman, R. P. P. P. (2015). To center or not to center? Investigating inertia with a multilevel autoregressive model. *Frontiers in Psychology*, *5*, Article 1492. https://doi.org/10.3389/fpsyg.2014.01492

Hecht, M., Walther, J.-K., Arnold, M., & Zitzmann, S. (2023). Finding the optimal number of persons (*N*) and time points (*T*) for maximal power in dynamic longitudinal models given a fixed budget. *Structural Equation Modeling: A Multidisciplinary Journal*, *31*(3), 535–551. https://doi.org/10.1080/10705511.2023.2230520

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

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*, Article 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), Article 140. https://doi.org/10.1186/s12874-018-0579-6

Joseph, N. T., Jiang, Y., & Zilioli, S. (2021). Momentary emotions and salivary cortisol: A systematic review and meta-analysis of ecological momentary assessment studies. *Neuroscience & Biobehavioral Reviews*, *125*, 365–379. https://doi.org/10.1016/j.neubiorev.2021.02.042

Jóźwiak, K., & Moerbeek, M. (2013). Optimal treatment allocation and study duration for trials with discrete-time survival endpoints. *Journal of Statistical Planning and Inference*, *143*(5), 971–982. https://doi.org/10.1016/j.jspi.2012.11.006

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].

Kassambara, A. (2023). *Ggpubr: 'ggplot2' based publication ready plots* (Version 0.6.0) [Computer software]. https://cran.r-project.org/package=ggpubr

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

Koch, T. K., Pargent, F., Kleine, A.-K., Lermer, E., & Gaube, S. (2023). *A tutorial on tailored simulation-based power analysis for experimental designs with generalized linear mixed models*. PsyArXiv. https://doi.org/10.31234/osf.io/rpjem

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

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). https://doi.org/10.1177/2515245920978738

Lafit, G., Artner, R., & Ceulemans, E. (2023). *Enabling analytical power calculations for multilevel models with autocorrelated errors through deriving and approximating the information matrix*. https://doi.org/10.31234/osf.io/tj5bv

Lafit, G., Meers, K., & Ceulemans, E. (2022). A systematic study into the factors that affect the predictive accuracy of multilevel VAR(1) models. *Psychometrika*, *87*(2), 432–476. https://doi.org/10.1007/s11336-021-09803-z

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

Lakens, D. (2022). Sample size justification. *Collabra: Psychology*, *8*(1), Article 33267. https://doi.org/10.1525/collabra.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. https://doi.org/10.1177/0265407517710342

Liu, J., & Colditz, G. A. (2017). Optimal design of longitudinal data analysis using generalized estimating equation models. *Biometrical Journal*, *59*(2), 315–330. https://doi.org/10.1002/bimj.201600107

Ludwigs, K., Lucas, R., Veenhoven, R., Richter, D., & Arends, L. (2020). Can happiness apps generate nationally representative datasets?—A case study collecting data on people's happiness using the German socio-economic panel. *Applied Research in Quality of Life*, *15*(4), 1135–1149. https://doi.org/10.1007/s11482-019-09723-2

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer.

Manju, M. A., Candel, M. J. J. M., & Berger, M. P. F. (2014). Sample size calculation in cost-effectiveness cluster randomized trials: Optimal and maximin approaches. *Statistics in Medicine*, *33*(15), 2538–2553. https://doi.org/10.1002/sim.6112

Meschiari, S. (2022). *Latex2exp: Use LaTeX expressions in plots* (Version 0.9.6) [Computer software]. https://cran.r-project.org/package=latex2exp

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

Moerbeek, M. (2020). Optimal designs for group randomized trials and group administered treatments with outcomes at the subject and group level. *Statistical Methods in Medical Research*, *29*(3), 797–810. https://doi.org/10.1177/0962280219846149

Moerbeek, M. (2023). Optimal placebo-treatment comparisons in trials with an incomplete within-subject design and heterogeneous costs and variances. *PLOS ONE*, *18*(4), Article e0283382. https://doi.org/10.1371/journal.pone.0283382

Müller, K., & Wickham, H. (2023). *Tibble: Simple data frames* (Version 3.2.1) [Computer software]. https://cran.r-project.org/package=tibble

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., Simonsohn, U., Wagenmakers, E.-J., Ware, J. J., & Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*(1), Article 0021. https://doi.org/10.1038/s41562-016-0021

Myin-Germeys, I., Kasanova, Z., Vaessen, T., Vachon, H., Kirtley, O., Viechtbauer, W., & Reininghaus, U. (2018). Experience sampling methodology in mental health research: New insights and technical developments. *World Psychiatry*, *17*(2), 123–132. https://doi.org/10.1002/wps.20513

Myin-Germeys, I., & Kuppens, P. (Eds.). (2021). *The open handbook of experience sampling methodology: A step-by-step guide to designing, conducting, and analyzing ESM studies*. Center for Research on Experience Sampling and Ambulatory Methods.

Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., Fidler, F., Hilgard, J., Kline Struhl, M., Nuijten, M. B., Rohrer, J. M., Romero, F., Scheel, A. M., Scherer, L. D., Schönbrodt, F. D., & Vazire, S. (2022). Replicability, robustness, and reproducibility in psychological science. *Annual Review of Psychology*, *73*(1), 719–748. https://doi.org/10.1146/annurev-psych-020821-114157

Ooms, J. (2024). *Magick: Advanced graphics and image-processing in R* (Version 2.8.5) [Computer software]. https://cran.r-project.org/package=magick

Pe, M. L., Kircanski, K., Thompson, R. J., Bringmann, L. F., Tuerlinckx, F., Mestdagh, M., Mata, J., Jaeggi, S. M., Buschkuehl, M., Jonides, J., Kuppens, P., & Gotlib, I. H. (2015). Emotion-network density in major depressive disorder. *Clinical Psychological Science*, *3*(2), 292–300. https://doi.org/10.1177/2167702614540645

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

Perugini, M., Gallucci, M., & Costantini, G. (2014). Safeguard power as a protection against imprecise power estimates. *Perspectives on Psychological Science*, *9*(3), 319–332. https://doi.org/10.1177/1745691614528519

Piccirillo, M., Fritz, J., Cohen, Z. D., Frumkin, M., Kirtley, O., Moeller, J., Neubauer, A., Norris, L., Schuurman, N. K., Snippe, E., & Bringmann, L. F. (2024). *A momentary assessment of the future of experience sampling researc*. https://doi.org/10.31234/osf.io/82bnf

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

Pinheiro, J., Bates, D., & R Core Team. (2023). *Nlme: Linear and nonlinear mixed effects models* (Version 3.1-167) [Computer software]. https://cran.r-project.org/package=nlme

R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Revol, J., Lafit, G., & Ceulemans, E. (2024). A new sample-size planning approach for person-specific VAR(1) studies: Predictive accuracy analysis. *Behavior Research Methods*, *56*(7), 7152–7167. https://doi.org/10.3758/s13428-024-02413-4

Rintala, A., Wampers, M., Lafit, G., Myin-Germeys, I., & Viechtbauer, W. (2023). Perceived disturbance and predictors thereof in studies using the experience sampling method. *Current Psychology*, *42*(8), 6287–6301. https://doi.org/10.1007/s12144-021-01974-3

Rizopoulos, D. (2023). *GLMMadaptive: Generalized linear mixed models using adaptive gaussian quadrature* (Version 0.9.1) [Computer software]. https://cran.r-project.org/package=GLMMadaptive

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581–592. https://doi.org/10.1093/biomet/63.3.581

Rush, A. J., Trivedi, M. H., Ibrahim, H. M., Carmody, T. J., Arnow, B., Klein, D. N., Markowitz, J. C., Ninan, P. T., Kornstein, S., Manber, R., Thase, M. E., Kocsis, J. H., & Keller, M. B. (2003). The 16-item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): A psychometric evaluation in patients with chronic major depression. *Biological Psychiatry*, *54*(5), 573–583. https://doi.org/10.1016/S0006-3223(02)01866-8

Schmidt, F. L., Hunter, J. E., & Pearlman, K. (1982). Assessing the economic impact of personnel programs on workforce productivity. *Personnel Psychology*, *35*(2), 333–347. https://doi.org/10.1111/j.1744-6570.1982.tb02199.x

Schmiedek, F., Lövdén, M., & Lindenberger, U. (2010). Hundred days of cognitive training enhance broad cognitive abilities in adulthood: Findings from the COGITO study. *Frontiers in Aging Neuroscience*, *2*, Article 27. https://doi.org/10.3389/fnagi.2010.00027

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

Simsa, B., Revol, J., Lafit, G., Cloos, L., Adolf, J., & Ceulemans, E. (2024). *Navigating missing data in multilevel autoregressive modeling: Consequences of missingness mechanisms and temporal patterns of missings on estimation performance*. PsyArXiv. https://doi.org/10.31234/osf.io/57vua

Tate, A. D., Fertig, A. R., De Brito, J. N., Ellis, É. M., Carr, C. P., Trofholz, A., & Berge, J. M. (2024). Momentary factors and study characteristics associated with participant burden and protocol adherence: Ecological momentary assessment. *JMIR Formative Research*, *8*, Article e49512. https://doi.org/10.2196/49512

Tekle, F. B., Tan, F. E. S., & Berger, M. P. F. (2011). Too many cohorts and repeated measurements are a waste of resources. *Journal of Clinical Epidemiology*, *64*(12), 1383–1390. https://doi.org/10.1016/j.jclinepi.2010.11.023

Thompson, R. J., Bailen, N. H., & English, T. (2021). Everyday emotional experiences in current and remitted major depressive disorder: An experience-sampling study. *Clinical Psychological Science*, *9*(5), 866–878. https://doi.org/10.1177/2167702621992082

Thompson, R. J., Mata, J., Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Gotlib, I. H. (2012). The everyday emotional experience of adults with major depressive disorder: Examining emotional instability, inertia, and reactivity. *Journal of Abnormal Psychology*, *121*(4), 819–829. https://doi.org/10.1037/a0027978

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

Tuerlinckx, F., Kuppens, P., Ariens, S., Cloos, L., Dejonckheere, E., Lafit, G., Niemeijer, K., Revol, J., Schat, E., Schreuder, M., Vanhasbroeck, N., & Ceulemans, E. (2025). New developments in experience sampling methodology. *British Journal of Mathematical and Statistical Psychology*. Advance online publication. https://doi.org/10.1111/bmsp.12398

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

Van Breukelen, G. J. P. (2013). Optimal experimental design with nesting of persons in organizations. *Zeitschrift Für Psychologie*, *221*(3), 145–159. https://doi.org/10.1027/2151-2604/a000143

van Buuren, S. (2018). *Flexible imputation of missing data* (2nd ed.). CRC Press.

Van Genugten, C. R., Schuurmans, J., Lamers, F., Riese, H., Penninx, B. W. J. H., Schoevers, R. A., Riper, H. M., & Smit, J. H. (2020). Experienced burden of and adherence to smartphone-based ecological momentary assessment in persons with affective disorders. *Journal of Clinical Medicine*, *9*(2), Article 322. https://doi.org/10.3390/jcm9020322

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). Springer.

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

Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General*, *143*(5), 2020–2045. https://doi.org/10.1037/xge0000014

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. Springer.

Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *Dplyr: A grammar of data manipulation* (Version 1.1.4) [Computer software]. https://cran.r-project.org/package=dplyr

Wickham, H., & Pedersen, T. L. (2024). *Gtable: Arrange 'grobs' in tables* (Version 0.3.6) [Computer software]. https://CRAN.R-project.org/package=gtable

Wickham, H., Vaughan, D., & Girlich, M. (2024). *Tidyr: Tidy messy data* (Version 1.3.1) [Computer software]. https://CRAN.R-project.org/package=tidyr

Wilke, C. O. (2024). *Cowplot: Streamlined plot theme and plot annotations for 'Ggplot2'* (Version 1.1.3) [Computer software]. https://CRAN.R-project.org/package=cowplot

Wrzus, C., & Neubauer, A. B. (2022). Ecological momentary assessment: A meta-analysis on designs, samples, and compliance across research fields. *Assessment*, *30*(3), 825–846. https://doi.org/10.1177/1073191121106753 8

Zelenski, J. M., & Larsen, R. J. (2000). The distribution of basic emotions in everyday life: A state and trait perspective from experience sampling data. *Journal of Research in Personality*, *34*(2), 178–197. https://doi.org/10.1006/jrpe.1999.2275

Zhu, H. (2024). *kableExtra: Construct complex table with 'kable' and pipe syntax* (Version 1.4.0) [Computer software]. https://CRAN.R-project.org/package=kableExtra

Zimmer, F., Henninger, M., & Debelak, R. (2023). Sample size planning for complex study designs: A tutorial for the mlpwr package. *Behavior Research Methods*, *56*(5), 5246–5263. https://doi.org/10.3758/s13428-023-02269-0

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

Zitzmann, S., Wagner, W., Hecht, M., Helm, C., Fischer, C., Bardach, L., & Göllner, R. (2022). How many classes and students should ideally be sampled when assessing the role of classroom climate via student ratings on a limited budget? An optimal design perspective. *Educational Psychology Review*, *34*(2), 511–536. https://doi.org/10.1007/s10648-021-09635-4

Zyphur, M. J., Voelkle, M. C., Tay, L., Allison, P. D., Preacher, K. J., Zhang, Z., Hamaker, E. L., Shamsollahi, A., Pierides, D. C., Koval, P., & Diener, E. (2020). From data to causes II: Comparing approaches to panel data analysis. *Organizational Research Methods*, *23*(4), 688–716. https://doi.org/10.1177/1094428119847280

Received October 15, 2024
Revision received May 22, 2025
Accepted May 28, 2025 ■

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