

Minimum Required Sample Size for Modelling Daily Cyclic Patterns in Ecological Momentary Assessment Data

Robin van de Maat¹ , Johan Lataster^{1,2} , Peter Verboon¹ 

[1] Faculty of Psychology, Open Universiteit, Heerlen, the Netherlands. [2] Department of Psychiatry and Psychology, School for Mental Health and Neuroscience, Maastricht University Medical Centre, Maastricht, the Netherlands.

Methodology, 2024, Vol. 20(4), 265–282, <https://doi.org/10.5964/meth.11399>

Received: 2023-02-17 • Accepted: 2024-11-07 • Published (VoR): 2024-12-23

Handling Editor: Isabel Benitez, University of Granada, Granada, Spain

Corresponding Author: Robin van de Maat, MA, P.O. Box 2960, 6401 DL, Heerlen, the Netherlands. E-mail: robin.vandemaat@ou.nl

Supplementary Materials: Code [see [Index of Supplementary Materials](#)]



Abstract

Cyclical patterns in ecological momentary assessment (EMA) data on emotions have remained relatively underresearched. Addressing such patterns can help to better understand emotion dynamics across time and contexts. However, no general rules of thumb are readily available for psychological researchers to determine the required sample size for measuring cyclical patterns in emotions. This study, therefore, estimates the minimum required sample sizes—in terms of the number of measurements per time period and subjects—to obtain a power of 80% given a certain underlying cyclical pattern based on input parameter values derived from an empirical EMA dataset. Estimated minimum required sample sizes varied between 50 subjects and 10 measurements per subject for accurately detecting cyclical patterns with a large magnitude, to 60 subjects and 30 measurements per subject for cyclical patterns of small magnitude. The resulting rules of thumb for sample sizes are discussed with a number of considerations in mind.

Keywords

ecological momentary assessment, experience sampling method, ambulatory assessment, intensive longitudinal data, cyclic models, sample sizes

Ecological Momentary Assessment (EMA) methods are increasingly being employed in psychological research to collect and assess data of everyday experiences over time



and (natural) contexts (Lafit et al., 2021; Liu & West, 2016). This is, to a large extent, driven by the expected advantages that EMA has over traditional retrospective methods (Ebner-Priemer & Trull, 2009; Kuppens & Verduyn, 2017; Reis et al., 2014; Shiffman et al., 2008; Wilhelm & Grossman, 2010). It is now well-recognized that EMA-based methods reduce the risk of recall bias, increase ecological validity, and have the ability to capture variations in emotions in daily life for more extended periods (Ebner-Priemer & Trull, 2009; Lazarus & Fisher, 2021; Lutz et al., 2018; Wilhelm & Grossman, 2010). This allows researchers to observe and address cyclical patterns in everyday experiences of emotions (van de Maat et al., 2020). However, not many studies have investigated cyclicity in EMA data (with a few exceptions, see for instance: Augustine & Larsen, 2012; Fok & Ramsay, 2006; Huh et al., 2015; Liu & West, 2016; Verboon & Leontjevas, 2018; West & Hepworth, 1991). Addressing these patterns can provide us with a clearer understanding of emotion dynamics across time and contexts (Beal & Ghandour, 2011; Verboon & Leontjevas, 2018), while ignoring these patterns may, alternatively, cause biased outcomes due to model misspecification resulting in spuriously high or low estimates of within- and between-subject relationships (Liu & West, 2016; Verboon & Leontjevas, 2018).

To overcome these issues, cyclic models using cyclic time predictors can be used to reduce error in within- and between-subject relations, isolate theoretically relevant effects, and let us better understand diurnal emotional patterns (van de Maat et al., 2020; Verboon & Leontjevas, 2018). However, the ability to reliably capture and account for cyclicity will partially depend on the amount of available EMA data, i.e., the total number of subjects and available measurements per subject (van de Maat et al., 2020), which, in turn, also depends on the overall response compliance of participants in the study.

Unfortunately, there are no general rules of thumb readily available for psychological researchers to determine the required sample size for measuring cyclical patterns in emotions data in daily life, or to determine how many measurements are allowed to be missing when it comes to cyclic modelling in EMA data. Therefore, this study aims to determine how many subjects and measurements are required to detect an underlying cyclical pattern of certain magnitude in EMA data, given that such a pattern exists, thereby allowing researchers to deal with these patterns. For this purpose, a power analysis will be performed, using Monte Carlo simulations based on an underlying cyclic model. We will focus exclusively on diurnal fluctuations, which are typically most relevant for EMA research (Chow et al., 2005; Stone et al., 1996), although our approach could be extended to encompass weekly or monthly cycles, if data would be suitable. After briefly discussing the rationale for considering cyclic models to analyse EMA data on emotions and how to test the statistical power in EMA data with various sample sizes and underlying cyclical patterns, we will perform and evaluate the results of a power analysis using different input parameters (e.g., sample sizes, cycle magnitudes, and random effects). Based on these results, possible rules of thumb are suggested for the number of subjects and measurements needed for detecting underlying cyclical patterns

of certain magnitude in EMA data. This facilitates researchers to better choose among different EMA sampling protocols in order to detect diurnal cyclical patterns and/or controlling for potential bias caused by such patterns in longitudinal emotion data.

Cyclic Patterns in Emotion Data and Why to Address Them

Everyday mental experiences in people's lives are seldom constant but continuously fluctuate over time and contexts due to internal or external events (Kuppens et al., 2010). The dynamics of these different experiences can be very distinct from one another and manifest themselves in various ways. Several studies have found, for instance, that (positive) emotions follow diurnal (Stone et al., 1996) or weekly cyclical patterns (Larsen & Kasimatis, 1990; Liu & West, 2016; Ram et al., 2005), whereas other emotions, such as negative mental states, are less likely to exhibit such cyclicity (van de Maat et al., 2020; Wood & Magnello, 1992). Simply put, a cycle is "a pattern of fluctuation (i.e., increase or decrease) that reoccurs across periods of time" (Jebb et al., 2015, p. 4). Furthermore, cyclic patterns in emotions can have different origins: they may be caused by exogenous factors (Beal & Ghandour, 2011) and thereby strongly associated with activities and locations, or depend on endogenous factors (Smyth & Stone, 2003), due to various cyclic physiological processes that influence psychological states (Stone et al., 1996). It has been suggested, for instance, that positive mood may have a biological component, whereas negative mood might be more environmentally determined (Stone et al., 1996). Other research has shown that positive moods tend to be higher when people are in company than when they are alone (Burgin et al., 2012; Silvia & Kwapil, 2011), suggesting it also has a large environmental component.

It has also become clear that people consistently differ in how much their emotions vary over time (Kuppens et al., 2010), meaning cyclic patterns in emotions are not uniform across individuals. Studies have shown systematic individual differences in both the presence and expression of these emotional cycles (Ram et al., 2005). Cycles can be more or less synchronized across individuals, e.g., due to similar work schedules, while other cycles are distinctly different across persons, such as specific daily morningness-eveningness rhythms (Liu & West, 2016). Furthermore, these cyclic patterns may also differ in magnitude between individuals: some may exhibit exaggerated cyclical patterns, whereas others may exhibit less pronounced fluctuations over time or contexts (Beal & Ghandour, 2011). Furthermore, the duration of cycles can differ, as emotional episodes may last for seconds, minutes, hours, or even for more extended time periods (Verduyn et al., 2015).

Researchers may have, in general, two reasons to be interested in cyclic patterns (van de Maat et al., 2020). First, they might be interested in exploring and capturing cyclic patterns (in emotions); their magnitude, duration, and how they differ across or within individuals. Another, second reason might be that the cyclic patterns found in the data are not of primary interest to the study but may cause bias in the effects the study tries

to uncover. Think, for example, of a study which attempts to measure affective reactivity to a particular environmental stimulus such as the workplace environment. It can be argued that (biologically driven) diurnal fluctuations in affect may confound the effect of interest and should be taken into account, similar to the correction for diurnal cortisol curves in EMA studies aiming to capture the “true” effects of environmental stressors on cortisol release (e.g., [Collip et al., 2011](#)). This seems to be particularly relevant if the occurrence of the environmental stimulus is (strongly) associated with the time of day, such as daily (work) schedules and related social interactions ([van de Maat et al., 2020](#)).

Various methods are available to capture cyclical patterns, all with their own advantages and disadvantages (for an overview, see [van de Maat et al., 2020](#)). In this paper, we will, however, rely on a cyclic model using cosine terms, which allows us to efficiently and accurately describe cyclical patterns more parsimoniously than some other traditional approaches ([van de Maat et al., 2020](#)). Furthermore, the cyclic model can easily be applied to hierarchical data, such as EMA datasets (for a tutorial, see: [van de Maat et al., 2020](#); [Verboon & Leontjevas, 2018](#)), and can therefore be very useful for the estimation of cyclical patterns that are common across participants, as well as patterns that may vary between participants ([Verboon & Leontjevas, 2018](#)). The following cyclic model can be used to capture cyclic patterns in the data ([Flury & Levri, 1999](#)):

$$\hat{Y}_{it} = b_{0i} + b_{1i} \cos\left(\frac{2\pi}{P}(t_i - b_{2i})\right) + e_i \quad (1)$$

Here b_{0i} represents the intercept or the mean value of the cyclic pattern (i.e., the vertical shift) for a specific participant, while b_{1i} is the amplitude or magnitude of the cyclic pattern, and b_{2i} the phase shift, i.e., the moment when the highest point of the cycle is established. Lastly, P is the number of measurements in a time period to complete a full diurnal, weekly or monthly cycle (i.e., the periodicity).

Sample Size in EMA Data

The required sample size for a study depends on many factors, such as the size and complexity of the model, distribution of the variables, nonresponse and amount of missing data, reliability of the variables, and strength of the relations among the variables ([Lafit, 2022](#); [Muthén & Muthén, 2002](#)). To capture specific cyclical patterns in EMA data, the length of the assessment period (e.g., day, week, or month), the total number of subjects and measurements, and the number of measurements per time period are essential ([van de Maat et al., 2020](#)). If the sample size is too small, the underlying cyclical patterns in the data may remain undetected or convergence difficulties may arise when a model is too complex for the data. EMA-based methods, and for that matter, any other intensive repeated measurement techniques, put high demands on participants and could lead to low response compliance, which, in turn, may affect the quality of the data ([Rintala et al., 2019](#)). Thus, one of the challenges for EMA researchers is to determine the optimal

sample size and assessment schedule that allows for accurately capturing patterns in the data while avoiding overburdening participants (Silvia et al., 2014).

Power Analyses in Hierarchical Structured Data

A power analysis, or in this case, *a priori power analysis*, aims to help researchers determine the smallest sample size that is required to detect the effect of a given test at the desired level of statistical power (Arend & Schäfer, 2019; Lafit, 2022). The power in a hypothesis test is the probability that the test will detect an effect that actually exists, for instance, the presence of a cyclical pattern in the data. However, estimating the power in hierarchical data is quite complex because sample sizes are allocated at multiple levels, and several types of effects (e.g., fixed and random effects) can be tested (Arend & Schäfer, 2019; Lafit et al., 2021). A recently developed method based on Monte Carlo Simulation in the statistical environment R (R Core Team, 2024) named *SIMR* (Green & MacLeod, 2016) allows us to perform such tests more easily and run power analyses for all types of models with different types of outcomes, input parameters, significance tests, and ranges of samples sizes (Arend & Schäfer, 2019). By generating random datasets using specific (true) population values from actual data, which increases the realism of the test, the “performance” of different research designs with different (cyclical) patterns can be evaluated. Therefore, the strength of this approach is that it can evaluate effects relative to the true population parameters (Silvia et al., 2014), using different sample sizes.

Goals, Research Questions and Hypotheses

The goal of this study is to provide researchers with a rule of thumb for the minimum required number of data points in EMA studies, for detecting cyclical patterns with different magnitudes of fixed and random effects in emotion data. The number of data points is determined by the number of subjects and number of measurements per subject.

Method

Design, Sample and Procedure

A power analysis was performed based on the parameters of a previously collected EMA dataset (van de Maat et al., 2020). We followed steps and guiding principles from Paxton et al. (2001), together with the general recommendations for power analysis in two-level models from Arend and Schäfer (2019). Using the previously mentioned package *SIMR* in the statistical environment R version 4.4.1 (R Core Team, 2024), the performances of several cyclic model designs were tested across different within-person and between-person sample sizes, as well as different magnitudes of cyclical patterns and levels of random effects. The Monte Carlo study distinguishes four factors by which simulated random datasets were generated: the number of subjects in the sample size

($S = 10, \dots, 100$), the number of measurements per subject ($M = 10, 30, \text{ and } 50$), effect sizes (i.e., estimated using the cosine terms of the cyclic model) ($ES = .12, .20, \text{ and } .28$), and random effects ($RE = 0.5 \text{ and } 0, 0.4 \text{ and } 0.1, 0.1 \text{ and } 0.4$, for the intercept and slope variance respectively). It was assumed that the random intercept and slope effects were uncorrelated from each other in all conditions. In total, the current study design yields 270 conditions (see the specification of input parameters below), which were replicated 1,000 times (as recommended by [Green & MacLeod, 2016](#); [Schoemann et al., 2014](#)).

Measures

The simulated random EMA datasets generated for the power analysis are based on the parameters derived from a previously performed empirical EMA study. EMA data were collected from 109 Dutch-speaking participants from the general population ($M_{age} = 47.25, SD_{age} = 14.84$; 42% male), recruited through convenience sampling by undergraduate students of The Open University of the Netherlands. Participants with fewer than 17 valid reports (less than 33.3% of all signals, according to guidelines by [Delespaul, 1995](#)) were excluded from analyses, resulting in a final dataset of 2,783 measurements from 76 participants (see: [van de Maat et al., 2020](#) for a complete description of the study protocol and procedures). From the empirical study, the distribution (input) parameters from the measurement of *Positive affect* are derived. *Positive affect* was defined as the mean score on three self-report items from the Positive And Negative Affect Schedule (PANAS; [Engelen et al., 2006](#)), i.e., 'I feel cheerful,' 'I feel happy,' and 'I feel satisfied,' all scored on a 7-point Likert Scale from 1 = 'Not at all' to 7 = 'Very'. As mentioned earlier, negative mental states do not seem to exhibit cyclical patterns in healthy adult samples (see, for instance: [van de Maat et al., 2020](#); [Wood & Magnello, 1992](#)) or EMA measurements lack sensitivity to detect them, which is why the current study solely looked at positive affect.

Analysis

The earlier mentioned cyclic model from [Flury & Levri \(1999\)](#) was used to capture cyclic patterns in the simulated data. In our analysis, using a multilevel approach, we focus on both the fixed effects of diurnal cycles that all participants have in common, as well as random effects that capture the variability between individual participants. For more additional details on the cyclic model, we refer to previous literature on the subject (see, for instance: [van de Maat et al., 2020](#); [Verboon & Leontjevas, 2018](#)).

Specification of Input Parameters

To conduct a power analysis and determine the minimum required sample size for various possible magnitudes of cyclic patterns in EMA data, we must first specify the input parameters. For this, a 3 (effect sizes, defined as amplitudes of the cyclic pattern (b_1): small, medium, and large) \times 3 (types of random effects ($\sigma_{b_0}^2$ and $\sigma_{b_1}^2$): 0.5 and 0, 0.4 and 0.1,

0.1 and 0.4 for intercept and slope variance respectively) $\times 3$ (number of measurements in a 5 days-time period: from 10, representing two measurement per day, to 50, representing ten measurements per day, with increments of 20) $\times 10$ (number of subjects: 10 to 100 with increments of 10) design was used. The 3 different amplitudes (b_1): .15, .25, and .35, represent a small, medium and large cyclical pattern. The corresponding effect sizes (ES) for these amplitudes are respectively: .12, .20, and .28 for the small, medium, and large cyclical pattern. Here, the effect size is defined as: $ES = \frac{b_1}{\sqrt{(\tau_0^2 + \sigma^2)}}$, where the amplitude (b_1) is divided by the square root of the total variance, which is the sum of the within and between-subjects variance (see: Pustejovsky et al., 2014; Verboon et al., 2021 for a more comprehensive description of effect sizes). The chosen amplitude (b_1) for the very small cycle condition matches the smallest amplitude size from a previous study (van de Maat et al., 2020), which ranged from .05 to .09.

Preliminary results indicated that the required power of .80 for this very small effect size (ES) could not be achieved, even with 100 subjects and 50 measurements each (power = .71). Therefore, to reduce the size of the Monte Carlo study and avoid including uninformative conditions, we excluded the very small ES from further simulations. Other input parameters of the cyclic model in the simulations were held constant: the intercept (b_0) is set at 1, the periodicity (P) at 10, and the phase shift (b_2) has the value of 5, meaning that the highest point of the cycle is reached at the 5th measurement for a daily cycle that takes 10 measurements to complete (approximately equal to the empirical dataset where the phase shift was around noon). Lastly, the between-subject variance (i.e., the variance across the different subjects) was fixed at the value .50, while the intercept and slope variance varied across three types of random effect conditions: a large random intercept variance with no random slope variance ($\sigma_{b_0}^2 = .5$ and $\sigma_{b_1}^2 = .0$), a medium random intercept variance with a small random slope variance ($\sigma_{b_0}^2 = .4$ and $\sigma_{b_1}^2 = .1$), and a small random intercept variance with a large random slope variance ($\sigma_{b_0}^2 = .1$ and $\sigma_{b_1}^2 = .4$). The intra-class correlation coefficient (ICC) for all three random effect combinations is .33.

During the test-run simulations, it was established that different values for the parameters b_0 or b_2 did not majorly affect the results of our analysis. Adjusting the value for the periodicity (P), on the other hand, did affect the outcome of our analysis and therefore was set corresponding to the value of the EMA prompt design and dataset from the previous study (i.e., at 10). The number of measurements per subject for a profile condition in the simulation ranged from 10 to 50, giving a maximum of 5,000 measurements in the largest simulated sample size (i.e., 10 measurements per day for 5 consecutive days measured for 100 subjects). For the profile conditions in which the number of measurements per subject was below 50, the simulation randomly picked measurement moments within the total time period of 50 measurements up to the number of measurements set for that present profile condition.

Profile Conditions and Evaluating the Results

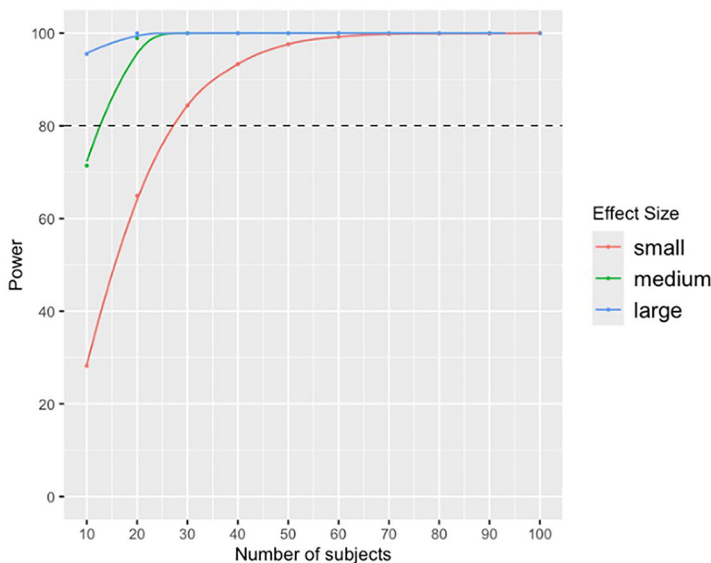
Each profile condition was evaluated using the randomly generated datasets (based on the population model derived from Equation 1 and the input parameters of the corresponding condition) to the extent to which it reached a sufficiently high power (i.e., a power of .80 or more; cf. Cohen, 1992). Stated differently, for all conditions the probabilities were estimated of correctly rejecting the H_0 (where H_0 means no cyclicity is present in the data) when cyclical patterns were present in the simulated data (with a significance level (α) of .05). The R scripts used for these analyses are available as supplementary materials (see van de Maat et al., 2024).

Results

The results of the power analyses performed for the small ($ES = .12$), medium ($ES = .20$), and large ($ES = .28$) cyclical pattern profile conditions are shown in Figure 2. To clarify the results, Figure 1 shows the power estimates for the large random intercept variance condition (with no random slope variance) across all three effect sizes, with 30 measurements per subject.

Figure 1

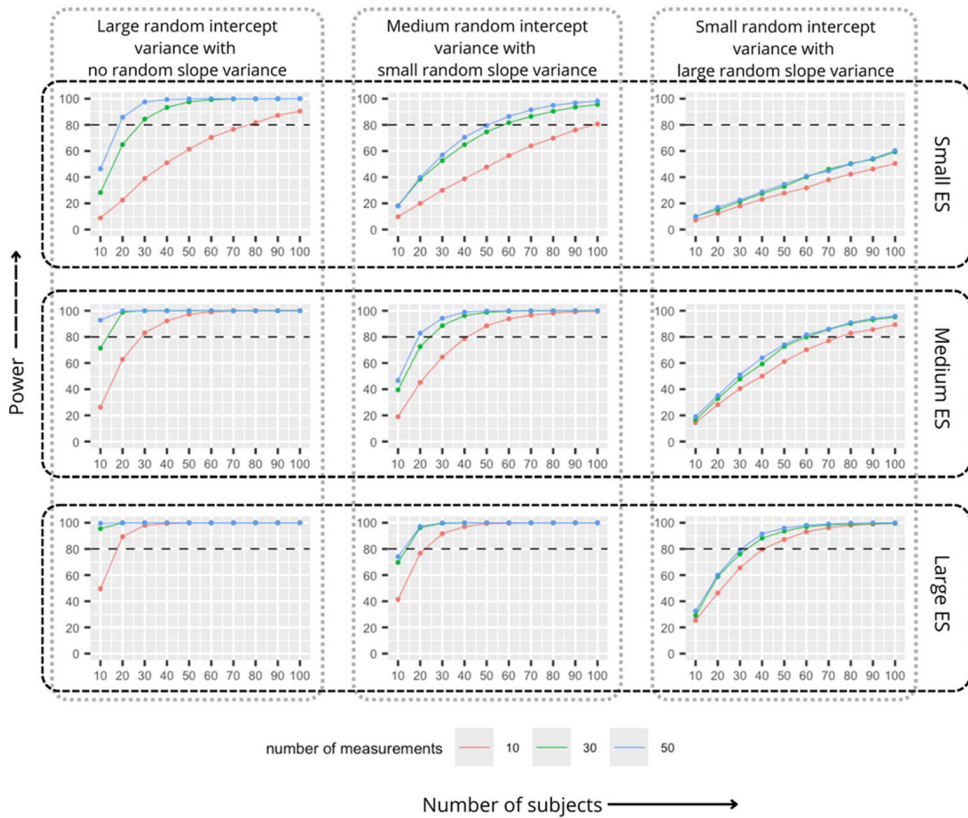
Power Estimates for the Large Random Intercept Variance Condition



Note. The three lines represent the power estimates for the small, medium, and large effect sizes in the large random intercept variance condition with no random slope variance, for sample sizes up to 100 subjects. The number of measurements per subject is held constant at 30.

Figure 2

Power Estimates for Different Number of Subjects and Measurements per Subject



Note. Power estimates (y-axis) for different number of subjects (x-axis) under the following simulation condition: vertical shift (b_0) = 1, amplitude (b_1) = .05, phase shift (b_2) = 5, periodicity (P) = 10, number of simulations per test condition = 1,000. Number of measurements per subject: 10 (red line), 30 (green lines), and 50 (blue lines).

As expected, the power increases most strongly in the large *ES* condition as the number of subjects increases, followed by the medium *ES* condition. It increases least strongly, as opposed to the other conditions, for the small *ES*. With a large *ES* and 30 measurements per subject, sufficient power is reached if approximately 10 or more subjects are included. In the case of the medium *ES*, 20 or more subjects were needed, and for the small *ES*, at least 30 subjects.

Figure 2 shows the power estimates for different number of subjects (ranging from 10 to 100) and number of measurements per subject (10, 30 and 50) in the small (top row), medium (middle row), and large (bottom row) *ES* conditions, respectively. The

power estimates are shown for three types of variances across subjects: a large random intercept variance condition with no random slope variance (left column), a condition with a medium random intercept variance and a small random slope variance (middle column), and a condition with a small random intercept variance paired with a large random slope variance (right column). Compared to the other conditions, the power increases most strongly in the large random intercept variance conditions with a large *ES* when the number of measurements per subject or the number of subjects is increased. Power decreases as the random slope variance increases and/or the *ES* becomes smaller.

For the large *ES* with a few measurements per subject (10), around 50 subjects are needed to achieve sufficient power across all random effect conditions. With more measurements per subject (30 and 50), sufficient power is achieved with 40 subjects.

For the medium *ES* with 10 measurements per subject, 80 subjects are needed to achieve sufficient power in all random effect conditions. By increasing the measurements per subject to 30, 70 subjects are needed, and with 50 measurements per subject, 60 subjects are sufficient.

Lastly, power for detecting the small *ES* was insufficient with up to 50 measurements per subject and up to 100 subjects or less in the small random intercept variance condition with a large random slope variance. With a medium random intercept variance and a small random slope variance, 100 subjects and 10 measurements per subject provided sufficient power, and in the large random intercept variance condition with no random slope variance, at least 80 subjects are required. With 30 or 50 measurements per subject and a medium random intercept variance with a small random slope variance, sufficient power for detecting a small *ES* is achieved with 60 subjects. In the large random intercept variance condition (with no random slope variance), 30 subjects are sufficient with 30 measurements per subject, and 20 subjects were sufficient with 50 measurements per subject.

Overall, the power increases slightly stronger in all tested cycle conditions by increasing the number of measurements per subject rather than including additional subjects (as can be seen in [Figure 2](#)).

Discussion

With this study we have attempted to determine the minimum required sample sizes—in terms of the number of subjects and number of measurements per subject—to obtain a power of 80% for detecting different magnitudes of existing cyclical patterns in EMA data and, by using these results, provide (psychology) researchers with rules of thumb for sample sizes when using EMA-based methods for collecting data on emotions. For this aim, multiple power analyses, using Monte Carlo simulations, were performed for three different magnitudes of daily cyclical patterns: small, medium, and large cycles with three random variance conditions across subjects.

The power analyses performed in this study suggest that power in EMA studies increases most strongly in the large daily cycle conditions and least strongly for the small cycle conditions as the number of measurements per subject and the number of subjects increase. However, our findings suggest that power in EMA studies may generally benefit more from increasing the number of measurements per subject rather than including additional subjects in the study.

Depending on the magnitude of the daily cyclical pattern and the degree of between subject variance (whether small or large random intercept or slope effects), we suggest the following sample sizes as a rule of thumb (see also [Table 1](#)). Studies with expected large magnitudes of cyclical patterns are sufficiently powered with 40 subjects and 30 or more measurements per subject. Medium cycles should be detectable using 70 subjects in the study with at least 30 measurements per subject. For smaller cycle magnitudes, this study did not establish the required number of subjects and measurements per subject when the random slope variance exceeds the random intercept variance. However, if small cyclical patterns are expected to be found in the data with a relatively small random slope effect, a minimum of 60 subjects and 30 or more measurements per subject is recommended to find a cyclical pattern of this magnitude.

Additionally, the results seem to indicate that any sample size smaller than 100 subjects and 50 measurements per subject will not provide sufficient power to capture very small-scale cycle patterns. Alongside any practical difficulties for adding additional subjects into an EMA study, one could also argue that if such small cyclic patterns remain undetected, this may not be a major issue since it will not introduce much (additional) bias and confound “true” effects of interest. However, such very small fluctuations can nevertheless be meaningful for, for instance, certain physiological measurements, such as skin conductance responses, respiration, and heart rate variability (see for instance: [van Halem et al., 2020](#)). For those kinds of EMA studies, in which the researcher might be looking for very small fluctuations, one is likely to need a sample size with a large number of subjects as well as a large number of measurements per subject.

The aforementioned rules of thumb should be used with some considerations and precautions in mind. First, the minimum required sample sizes do not account for non-response or other (quality) issues related to gathering EMA data. Missing data due to nonresponse increases standard errors and reduces statistical power, in addition to other potentially harmful consequences in psychological research ([Enders, 2010](#); [Schoemann et al., 2017](#); [Silvia et al., 2014](#)). [Rintala et al. \(2019\)](#) found, for instance, that data collection using EMA based-methods showed an average response rate of 78% and that compliance depended on the time within a day (highest compliance was found in the afternoon and lowest in the early morning) and declined on the fifth assessment day. Factors that (additionally) may influence compliance are participants' health (psychosis patients were less compliant) and possibly the number of items per assessment ([Rintala et al., 2019](#)). These aspects influence the final sample size, the distribution, and the density

Table 1
Rules of Thumb for the Required Number of Subjects to Obtain Sufficient Power With 10 or 30 or More Measurements per Subject

Expected Effect Size (ES)	Expected Random Effect		
	Large random intercept variance with no random slope variance	Medium random intercept variance with small random slope variance	Small random intercept variance with large random slope variance
Small ES	> 90 subjects with 10 measurements > 30 subjects with > 30 measurements	> 100 subjects with 10 measurements > 60 subjects with > 30 measurements	unknown ^a
Medium ES	> 30 subjects with 10 measurements > 20 subjects with > 30 measurements	> 50 subjects with 10 measurements > 30 subjects with > 30 measurements	> 80 subjects with 10 measurements > 70 subjects with > 30 measurements
Large ES	> 20 subjects with 10 measurements > 10 subjects with > 30 measurements	> 40 subjects with 10 measurements > 20 subjects with > 30 measurements	> 50 subjects with 10 measurements > 40 subjects with > 30 measurements

^a The required number of subjects to obtain sufficient power in the condition with a small ES and a large random slope variance was not established.

of responses on the time within a day or days of the week and require researchers to increase the number of participants or measurement prompts in the EMA study-setting. Therefore, recruiting additional participants or increasing the number of measurement moments in test settings is recommended to compensate for nonresponse but not so much to overburden participants resulting in lower response compliance and power. Researchers should find the right balance between the number of measurement prompts and overall response compliance. Additionally, researchers should invest in ways to improve compliance within existing sampling prompts.

It also requires further research on how skewness in the distributions of measurement moments—caused by the concentration of responses within a day or days of the week—affects the power of the study and the minimum sample size to capture daily cycle patterns. Recent developments for analysing missing data brought new attention to methods that mitigate the problem of missing data (Silvia et al., 2014). For instance, a planned missing-data design allows researchers to increase the quality of collected data in a study, especially when researchers intend to collect several items to measure a dependent variable (Schoemann et al., 2014).

Another and second important consideration is the issue of (intra)individual variations in cyclical patterns in EMA data on emotions. While fixed effects are generally the types of effects most frequently analysed in two-level models (Arend & Schäfer, 2019), cycles may vary from day to day (e.g., workdays may show different trajectories in the data than weekend days) and from person to person. The amount of within or between-subject variance has an impact on power (Hox et al., 2017). In our power analysis, we included several random effect conditions as an initial step toward addressing this complexity. While this only scratched the surface of the issue, it provides some insights into how these components may influence power. Notably, the power to detect cyclical patterns appears to decrease more sharply with increases in random slope variance compared to increases in random intercept variance. This suggests that a larger minimum sample size is required when a large random slope variance is anticipated. Since we kept the between-subject variance constant, future studies should more thoroughly examine how varying levels of total between-subject variance affect the power to detect cyclical patterns. This may also help explain our somewhat contradictory finding—compared to other studies—that an increase of the number of measurements per subject benefits EMA-studies more than including additional subjects (rather than vice versa as, for instance, Snijders (2005) concludes). In order to perform such analyses, different values for input parameters, such as intraclass correlation coefficient (ICC) and cross-level interaction (CLI) effects, need to be specified (Arend & Schäfer, 2019, provide a helpful and comprehensive guide for performing power analyses with multilevel EMA data).

Third and lastly, (psychology) researchers may, of course, be inclined to examine other theoretical relevant relationships in the data on emotions rather than solely cyclical patterns. However, adding additional predictors to the model to find these relevant

effects of interest increases the complexity of the model. This, in addition to smaller sample sizes, puts researchers at risk for more convergence failures and higher standard errors (Silvia et al., 2014). Therefore, the rules of thumb proposed in this study do not apply to all situations regarding sample size requirements. Future research might add additional (contextual) predictors to the cyclic model to further increase complexity, allowing evaluation of possible convergence issues when too small sample sizes are involved and any confounding effects associated with these variables.

Our study provides general rules of thumb for the minimum required sample size to capture underlying cyclical patterns in EMA data. This enables researchers to implement the appropriate EMA study design that adequately considers sample sizes to explore and control for these cyclical patterns.

Funding: The authors have no funding to report.

Acknowledgments: The authors have no additional (i.e., non-financial) support to report.

Competing Interests: The authors have declared that no competing interests exist.

Supplementary Materials

For this article, the following Supplementary Materials are available:

- R scripts for analyses of study plots and simulations. (van de Maat et al., 2024).

References

- 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>
- Augustine, A. A., & Larsen, R. J. (2012). Emotion research. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (Vol. 27, pp. 497–510). Guilford Press.
- Beal, D. J., & Ghandour, L. (2011). Stability, change, and the stability of change in daily workplace affect. *Journal of Organizational Behavior*, 32(4), 526–546. <https://doi.org/10.1002/job.713>
- Burgin, C. J., Brown, L. H., Royal, A., Silvia, P. J., Barrantes-Vidal, N., & Kwapil, T. R. (2012). Being with others and feeling happy: Emotional expressivity in everyday life. *Personality and Individual Differences*, 53(3), 185–190. <https://doi.org/10.1016/j.paid.2012.03.006>
- Chow, S.-M., Ram, N., Boker, S. M., Fujita, F., & Clore, G. (2005). Emotion as a thermostat: Representing emotion regulation using a damped oscillator model. *Emotion (Washington, D.C.)*, 5(2), 208–225. <https://doi.org/10.1037/1528-3542.5.2.208>
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159. <https://doi.org/10.1037/0033-2909.112.1.155>

- Collip, D., Nicolson, N. A., Lardinois, M., Lataster, T., van Os, J., & Myin-Germeys, I. (2011). Daily cortisol, stress reactivity and psychotic experiences in individuals at above average genetic risk for psychosis. *Psychological Medicine*, *41*(11), 2305–2315.
<https://doi.org/10.1017/S0033291711000602>
- Delespaul, P. (1995). *Assessing schizophrenia in daily life: the experience sampling method*. Datawyse/Universitaire Pers Maastricht.
- Ebner-Priemer, U. W., & Trull, T. J. (2009). Ecological momentary assessment of mood disorders and mood dysregulation. *Psychological Assessment*, *21*(4), 463–475. <https://doi.org/10.1037/a0017075>
- Enders, C. K. (2010). *Applied missing data analysis*. Guilford Press.
- Engelen, U., De Peuter, S., Victoir, A., Van Diest, I., & Van den Bergh, O. (2006). Verdere validering van de “Positive and Negative Affect Schedule” (PANAS) en vergelijking van twee Nederlandstalige versie [Further validation of the Positive and Negative Affect Schedule (PANAS) and comparison of two Dutch versions]. *Gedrag en Gezondheid*, *34*, 61–70.
<https://doi.org/10.1007/BF03087979>
- Flury, B. D., & Levri, E. P. (1999). Periodic logistic regression. *Ecology*, *80*(7), 2254–2260.
[https://doi.org/10.1890/0012-9658\(1999\)080\[2254:PLR\]2.0.CO;2](https://doi.org/10.1890/0012-9658(1999)080[2254:PLR]2.0.CO;2)
- Fok, C. C. T., & Ramsay, J. O. (2006). Fitting curves with periodic and nonperiodic trends and their interactions with intensive longitudinal data. In T. A. Walls & J. L. Schafer (Eds.), *Models for intensive longitudinal data* (pp. 109–123). Oxford University Press.
<https://doi.org/10.1093/acprof:oso/9780195173444.003.0005>
- Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498.
<https://doi.org/10.1111/2041-210X.12504>
- Hox, J. J., Moerbeek, M., & van de Schoot, R. (2017). *Multilevel analysis: Techniques and applications* (3rd ed.). Routledge. <https://doi.org/10.4324/9781315650982>
- Huh, D., Kaysen, D. L., & Atkins, D. C. (2015). Modeling cyclical patterns in daily college drinking data with many zeroes. *Multivariate Behavioral Research*, *50*(2), 184–196.
<https://doi.org/10.1080/00273171.2014.977433>
- Jebb, A. T., Tay, L., Wang, W., & Huang, Q. (2015). Time series analysis for psychological research: Examining and forecasting change. *Frontiers in Psychology*, *6*, Article 727.
<https://doi.org/10.3389/fpsyg.2015.00727>
- Kuppens, P., Oravecz, Z., & Tuerlinckx, F. (2010). Feelings change: Accounting for individual differences in the temporal dynamics of affect. *Journal of Personality and Social Psychology*, *99*(6), 1042–1060. <https://doi.org/10.1037/a0020962>
- Kuppens, P., & Verduyn, P. (2017). Emotion dynamics. *Current Opinion in Psychology*, *17*, 22–26.
<https://doi.org/10.1016/j.copsyc.2017.06.004>
- Lafit, G. (2022). Sample size selection in ESM studies. In I. Myin-Germeys & P. Kuppens (Eds.), *The open handbook of experience sampling methodology: A step-by-step guide to designing, conducting, and analyzing ESM studies* (2nd ed., pp. 217–236). Center for Research on Experience Sampling and Ambulatory Methods.

- 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>
- Larsen, R. J., & Kasimatis, M. (1990). Individual differences in entrainment of mood to the weekly calendar. *Journal of Personality and Social Psychology*, 58(1), 164–171. <https://doi.org/10.1037/0022-3514.58.1.164>
- Lazarus, G., & Fisher, A. J. (2021). Negative emotion differentiation predicts psychotherapy outcome: Preliminary findings. *Frontiers in Psychology*, 12, Article 689407. <https://doi.org/10.3389/fpsyg.2021.689407>
- Liu, Y., & West, S. G. (2016). Weekly cycles in daily report data: An overlooked issue. *Journal of Personality*, 84(5), 560–579. <https://doi.org/10.1111/jopy.12182>
- Lutz, W., Schwartz, B., Hofmann, S. G., Fisher, A. J., Husen, K., & Rubel, J. A. (2018). Using network analysis for the prediction of treatment dropout in patients with mood and anxiety disorders: A methodological proof-of-concept study. *Scientific Reports*, 8(1), Article 7819. <https://doi.org/10.1038/s41598-018-25953-0>
- Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*, 9(4), 599–620. https://doi.org/10.1207/S15328007SEM0904_8
- Paxton, P., Curran, P. J., Bollen, K. A., Kirby, J., & Chen, F. (2001). Monte Carlo experiments: Design and implementation. *Structural Equation Modeling*, 8(2), 287–312. https://doi.org/10.1207/S15328007SEM0802_7
- Pustejovsky, J. E., Hedges, L. V., & Shadish, W. R. (2014). Design-comparable effect sizes in multiple baseline designs: A general modeling framework. *Journal of Educational and Behavioral Statistics*, 39(5), 368–393. <https://doi.org/10.3102/1076998614547577>
- R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Ram, N., Chow, S.-M., Bowles, R. P., Wang, L., Grimm, K., Fujita, F., & Nesselroade, J. R. (2005). Examining interindividual differences in cyclicality of pleasant and unpleasant affects using spectral analysis and item response modeling. *Psychometrika*, 70(4), 773–790. <https://doi.org/10.1007/s11336-001-1270-5>
- Reis, H. T., Gable, S. L., & Maniaci, M. R. (2014). Methods for studying everyday experience in its natural context. In *Handbook of research methods in social and personality psychology* (2nd ed., pp. 373–403). Cambridge University Press.
- Rintala, A., Wampers, M., Myin-Germeys, I., & Viechtbauer, W. (2019). Response compliance and predictors thereof in studies using the experience sampling method. *Psychological Assessment*, 31(2), 226–235. <https://doi.org/10.1037/pas0000662>

- Schoemann, A. M., Boulton, A. J., & Short, S. D. (2017). Determining power and sample size for simple and complex mediation models. *Social Psychological & Personality Science*, 8(4), 379–386. <https://doi.org/10.1177/1948550617715068>
- Schoemann, A. M., Miller, P., Pornprasertmanit, S., & Wu, W. (2014). Using Monte Carlo simulations to determine power and sample size for planned missing designs. *International Journal of Behavioral Development*, 38(5), 471–479. <https://doi.org/10.1177/0165025413515169>
- Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology*, 4, 1–32. <https://doi.org/10.1146/annurev.clinpsy.3.022806.091415>
- Silvia, P. J., & Kwapil, T. R. (2011). Aberrant asociality: How individual differences in social anhedonia illuminate the need to belong. *Journal of Personality*, 79(6), 1315–1332. <https://doi.org/10.1111/j.1467-6494.2010.00702.x>
- Silvia, P. J., Kwapil, T. R., Walsh, M. A., & Myin-Germeys, I. (2014). Planned missing-data designs in experience-sampling research: Monte Carlo simulations of efficient designs for assessing within-person constructs. *Behavior Research Methods*, 46(1), 41–54. <https://doi.org/10.3758/s13428-013-0353-y>
- Smyth, J. M., & Stone, A. A. (2003). Ecological momentary assessment research in behavioral medicine. *Journal of Happiness Studies*, 4(1), 35–52. <https://doi.org/10.1023/A:1023657221954>
- Snijders, T. A. B. (2005). Power and sample size in multilevel linear models. In *Encyclopedia of statistics in behavioral science*. <https://doi.org/https://doi.org/10.1002/0470013192.bsa492>
- Stone, A. A., Smyth, J. M., Pickering, T., & Schwartz, J. (1996). Daily mood variability: Form of diurnal patterns and determinants of diurnal patterns. *Journal of Applied Social Psychology*, 26(14), 1286–1305. <https://doi.org/10.1111/j.1559-1816.1996.tb01781.x>
- van de Maat, R., Lataster, J., & Verboon, P. (2020). Why and how to deal with diurnal cyclic patterns in ambulatory assessment of emotions. *European Journal of Psychological Assessment*, 36(3), 471–481. <https://doi.org/10.1027/1015-5759/a000579>
- van de Maat, R., Lataster, J., & Verboon, P. (2024). Supplementary materials to “Minimum required sample size for modelling daily cyclic patterns in ecological momentary assessment data” [R scripts for analyses of study plots and simulations]. PsychOpen GOLD. <https://doi.org/10.23668/psycharchives.15817>
- van Halem, S., van Roekel, E., Kroencke, L., Kuper, N., & Denissen, J. (2020). Moments that matter? On the complexity of using triggers based on skin conductance to sample arousing events within an experience sampling framework. *European Journal of Personality*, 34(5), 794–807. <https://doi.org/10.1002/per.2252>
- Verboon, P., Duif, M., & van Tuijl, P. (2021). *Single case design analyses*. Open University Press. <https://ou-books.gitlab.io/scda---single-case-design-analyses/>
- Verboon, P., & Leontjevas, R. (2018). Analyzing cyclic patterns in psychological data. *Tutorials in quantitative methods for psychology*, 14(4), 218–234. <https://doi.org/10.20982/tqmp.14.4.p218>
- Verduyn, P., Delaveau, P., Rotgé, J.-Y., Fossati, P., & Van Mechelen, I. (2015). Determinants of emotion duration and underlying psychological and neural mechanisms. *Emotion Review*, 7(4), 330–335. <https://doi.org/10.1177/1754073915590618>

- West, S. G., & Hepworth, J. T. (1991). Statistical issues in the study of temporal data: Daily experiences. *Journal of Personality*, *59*(3), 609–662. <https://doi.org/10.1111/j.1467-6494.1991.tb00261.x>
- Wilhelm, F. H., & Grossman, P. (2010). Emotions beyond the laboratory: Theoretical fundamentals, study design, and analytic strategies for advanced ambulatory assessment. *Biological Psychology*, *84*(3), 552–569. <https://doi.org/10.1016/j.biopsycho.2010.01.017>
- Wood, C., & Magnello, M. E. (1992). Diurnal changes in perceptions of energy and mood. *Journal of the Royal Society of Medicine*, *85*(4), 191–194. <https://doi.org/10.1177/014107689208500404>



Methodology is the official journal of the European Association of Methodology (EAM).



leibniz-psychology.org

PsychOpen GOLD is a publishing service by Leibniz Institute for Psychology (ZPID), Germany.