KU Leuven

Faculty of Psychology and Educational Sciences

THE EFFECT OF MISSING DATA ON THE ESTIMATION BIAS, VARIANCE, AND STATISTICAL POWER IN MULTILEVEL AUTOREGRESSIVE(1) MODELS

Master's thesis submitted for the
degree of Master of Science in
Master of Psychology: Theory and
Research by

Benjamín Šimsa

Supervisor: Prof. Dr. Eva Ceulemans
Co-supervisors: Dr. Ginette Lafit, Jordan Revol

2023

Summary

The multilevel autoregressive (MLAR) model represents one of the most popular methods for quantifying affective inertia, the resistance of affective processes to change. The present thesis aims to address a gap in knowledge about the effect of the presence of missing data on the statistical properties of the MLAR model. Although missed observations are common in intensive longitudinal datasets, none of the simulation studies that investigated the statistical properties of the MLAR model focused on this problem.

In four simulation studies, we have investigated how the presence of different missing data patterns (missing completely at random, a block of consecutive missing observations, and the most extreme observations missing), together with different values of compliance (i.e., the proportion of non-missing observations), influences the estimation bias, standard error, and statistical power to detect the fixed slope of the MLAR model.

The results suggest that when the lagged predictor is person-mean centered, the presence of missing data leads to an underestimation of the fixed autoregressive effect. The estimation bias is the worst for the conditions in which the most extreme observations are missing, and it is more severe when the compliance is low. However, the effect of missing data on estimation bias decreases as the number of timepoints (observations) per participant increases. The standard error is mostly influenced by the number of participants in the simulated dataset, and it is higher when the compliance is low. Similarly, the statistical power to detect the fixed autoregressive slope was found to be influenced by both compliance and the missingness pattern. These conclusions holds regardless of whether the random autoregressive effects are estimated included in the model or not. However, the supplementary simulations show that the bias is considerably smaller when the predictor is not person-mean centered.

While we only explored a small subset of plausible simulation parameters in our studies, the reproducible simulation code is available online. This allows researchers to appraise the effect of missing data on the estimation in their planned study, and to take missing data into account when planning the sample size for their study. The results of the present thesis are also relevant for the research design of intensive longitudinal studies. They suggest that if an accurate estimation of inertia is the main research goal, motivating participants to high compliance and collecting more observations per participant might be more effective than collecting data from a larger number of participants.

Acknowledgments

I want to thank my supervisors for their guidance, patience and thoughtful feedback at each step of the thesis-writing process. The thesis topics at KU Leuven are assigned through a lottery system – and I can truly say I have won the lottery.

Contribution and approach

The work on the present thesis, written in the form of a scientific article, entailed three main aspects: I learned the concepts of and reviewed the literature about autoregressive models and their use in psychological research, familiarized myself with the methodology of Monte Carlo simulation studies in R, and I learned about the different ways of ensuring maximum computational reproducibility and transparency of the results. As such, apart from gaining substantive knowledge about time-series data analysis in psychology, I also obtained a number of transferable skills, including dynamic reporting (the thesis was written using R Markdown), version control and reproducible simulation research workflows.

In the first year, I had several meetings with my supervisors, Eva Ceulemans, Ginette Lafit and Jordan Revol. In the initial phases of the process, I have read through the available literature about the statistical properties of multilevel autoregressive models. I adapted the code that my supervisors provided to explore how violating different assumptions of the model influences the estimation. At the end of the first year, we have converged on a plan to focus on the effect of missing observations on estimation bias, variance, and statistical power of the multilevel autoregressive model. We brainstormed several different missingness patterns that correspond to simplified situations arising in Experience sampling method research. The main outcome of the course *Master's Thesis Part 1: Project* was a thesis proposal, in which I outlined the plan for the simulation studies.

During the summer break between the first and second year, I adapted and extended the code for simulating the data and estimating the model, provided by Ginette and Jordan, to introduce the different missingness patterns in the simulated datasets. I conducted a set of preliminary simulation studies and met with Ginette and Jordan to discuss the results.

In the autumn of the second year, we held meetings to finalize the simulation study design, including the different simulation conditions and simulation parameters. I also drafted the introduction section of the thesis for which Ginette provided feedback. After the research design was finalized, I started conducting the final simulations while also writing the first draft of the thesis. The supervisors continuously provided advice and feedback about the R functions used to simulate the datasets, fit the model and extract relevant parameters. In February, I sent the supervisors the first full draft of the thesis, for which they provided a detailed feedback. In the following months, we iteratively improved the draft and decided to include two supplementary simulation studies to provide more insight into the mechanisms behind the results.

Contents

List of abbreviations . . . 2
List of symbols . . . 2

Introduction **3**

Multilevel AR(1) model **6**
Assumptions of the MLAR(1) model . . . 7
Estimation procedures . . . 7
Evidence from simulation studies . . . 8

Methods **9**
Simulation procedure . . . 9
Reproducibility and code/data availability . . . 13

Results **15**
Simulation A . . . 15
Outcome: Estimation bias . . . 15
Outcome: Standard error . . . 18
Outcome: Statistical power . . . 20
Simulation B . . . 22
Outcome: Estimation bias . . . 22
Outcome: Standard error . . . 22
Outcome: Statistical power . . . 24
Supplementary simulation C . . . 27
Supplementary simulation D . . . 31

Discussion **34**
Directions for future research . . . 36
Limitations . . . 37
Conclusion . . . 37

References **39**

Appendix 1: Full results from Simulation A **43**

Appendix 2: Full results from Simulation B **54**

List of abbreviations

ANOVA	Analysis of variance
AR	Autoregressive
ESM	Experience Sampling Method
MAR	Missing at random
MCAR	Missing completely at random
MCMC	Markov Chain Monte Carlo
MLE	Maximum likelihood estimation
MLAR(1)	Multilevel first-order autoregressive
MNAR	Missing not at random
N	Number of participants
NA	Negative affect
PA	Positive affect
SD	Standard deviation
SE	Standard error
T.obs	Number of observations (timepoints) per participant

List of symbols

Symbol	Definition
β_{00}	Fixed intercept
β_{10}	Fixed autoregressive parameter
ϵ_{it}	Innovation / residuals
esm_{it}	Process value of participant i at time t
$esm_{i,t-1}$	Lagged process value of participant i
γ_{0i}	Person-specific intercept
γ_{1i}	Person-specific autoregressive parameter
ν_{0i}	Random intercept
ν_{1i}	Random autoregressive parameter
σ^2_{ϵ}	Innovation variance
$\sigma^2_{\nu_0}$	Variance of the random intercepts
$\sigma^2_{\nu_1}$	Variance of the random autoregressive effects
$\sigma_{\nu_{01}}$	Correlation between the random intercepts and random autoregressive effects

Introduction

In recent years, the focus in diverse subfields of psychology has been shifting towards complexity, dynamics, and a within-person perspective (Hamaker, 2012). Among other things, this shift has been facilitated by the growing availability of smartphones and wearables. These devices allow researchers to use the Experience sampling method (ESM) to collect intensive longitudinal data with a high level of ecological validity (Myin-Germeys et al., 2018). Intensive longitudinal data consist of several repeated measurements per day, nested within individual participants (Larson & Csikszentmihalyi, 2014). The use of intensive longitudinal data considerably broadens the extent of research questions psychological researchers can investigate and statistical analyses they can conduct. Importantly, the multilevel structure of intensive longitudinal data allows scientists to investigate both within-person dynamic processes and the between-person differences therein (Wright & Zimmermann, 2019). One of the new research avenues that emerged with the growing popularity of intensive longitudinal data is the study of affect dynamics (for an overview, see Houben et al., 2015). The dynamic affect measures make use of the structure of intensive longitudinal data to take the fluctuating nature of affect/emotions into account.

There are two main approaches to capture affect dynamics: fitting models to the data (such as the first-order multilevel autoregressive (MLAR) model, Koval et al. 2021), and computing within-person descriptive statistics (for instance, aurocorrelation). Both the MLAR model and the within-person autocorrelation estimates quantify the degree to which a given person's current affect can be predicted by their affect at the preceding timepoint (Kuppens et al., 2010). As such, both approaches target emotional inertia: the degree to which affective states linger, the resistance to change (Kuppens & Verduyn, 2017).

Although some degree of inertia is to be expected in human emotional experiences, a high level of emotional inertia (i.e., a high temporal persistence of emotional states) has been linked to psychological maladjustment (Kuppens et al., 2010). A negative emotion process with a high inertia can get caught in a self-reinforced feedback loop (also called *critical slowing down,* Leemput 2014). This causes the process to be partially resistant to both external influences and internal processes, including emotional regulation (Koval et al., 2015).

The evidence about the association between emotional inertia and the well-being/psychopathology spectrum has grown steadily over the last two decades. A recent meta-analysis indicated an association between emotional inertia (of both positive and negative emotions) and psychological well-being/psychopathology (Houben et al., 2015). Specifically, higher emotional inertia has been linked to lower well-being and higher ocurrence of depressive symptoms (Brose et al., 2015), bipolar disorder (Mneimne et al., 2018), and lower response of depression and anxiety symptoms to cognitive-behavioral therapy (Bosley et al., 2019). However, recent evidence suggests that the association of inertia (of both positive and negative affect) and psychopathology/well-being is only limited when the mean affect intensities are taken into account (Bos et al., 2019; Bosley et al., 2019; Dejonckheere, 2019; Koval et al., 2013).

Despite the popularity of the MLAR(1) models in psychological research, there are several questions about their statistical properties that remain unanswered. One of them is the effect of missing observations on their estimation performance. This is a pressing issue, given that the presence of missing data in intensive longitudinal datasets is more of a rule than an exception: the average compliance (i.e., the ratio of answered ESM beeps to the beeps not answered by the participants) in ESM studies is around 79% (SD = 13.64%; Wrzus & Nebauer 2022). Furthermore, study compliance is associated with both participant and study design characteristics (Wrzus & Neubauer, 2022). Providing financial incentives for participation was found to be associated with increased compliance, while longer ESM questionnaires are associated with lower compliance (Eisele et al., 2020; Vachon et al., 2019). Participants with psychotic disorders tend to have lower compliance compared to general population, while the opposite is true for participants with depressive disorders (Rintala et al., 2019; Vachon et al., 2019). Higher positive affect, lower negative affect, lower stress, and less alcohol use are linked to higher compliance (Rintala et al., 2019, 2020).

While Ji et al. (2018) show that the presence of data missing completely at random (MCAR), missing at random (MAR) and not missing at random (MNAR) leads to a considerable bias in point estimates of cross-lagged and autoregressive parameters in person-specific vector autoregressive models when list-wise deletion is used, no similar evidence is available about the MLAR(1) model.

For an illustration on how the different missingness patterns can manifest themselves in the same ESM time-series, please refer to Figure 1. The MCAR missingness pattern (panel *a* in the figure) assumes that the participants miss responding to beeps randomly, and each beep has the same probability of being missed, regardless on any other factors (e.g., whether the previous beep was missed, or the intensity of the emotion measured by ESM). However, many different missing data scenarios can arise in ESM research. For instance, participants could be more likely to miss a series of beeps when they attend a social event. This will result in a block of consecutive missing datapoints, where neither the starting point nor the endpoint of the missing block depend on the intensity of the emotion (panel *b*). Alternatively, probability of an observation being missing can depend on the value of the emotion process itself. For example, a participant can miss responding to an ESM measure of a positive mood because they are not feeling well enough. In panel *c*, we observe that the participant misses all observations in which the true process mean is below the process mean and sufficiently far away from it. Additionally, they can be less likely to answer an ESM beep in situations that make them feel very good (e.g., they might skip responding an ESM beep when celebrating). These events can be represented as moments in which the ESM process is far away from its mean (panel *d*).

The goal of the present thesis is to investigate whether compliance (i.e., the inverse of the proportion of missing data for each participant of an ESM study) and the different patterns of missingness described in the previous paragraph (MCAR; data missing in a block; extreme observations missing) have an effect on estimation bias, variability, and statistical power of the multilevel AR(1) model. In the following part, I will describe the multilevel autoregressive model and its assumptions and summarize the already available evidence about its statistical properties from simulation studies.

Multilevel AR(1) model

In this subchapter, I will describe the mathematical basis and assumptions of the first-order multilevel autoregressive model (MLAR[1]) with random intercepts and random autoregressive effects, which will be the focus of the simulation part of the thesis. The notation used by Lafit et al. (2020) will be adhered to throughout the thesis.

The MLAR(1) model consists of two levels: the within-person Level 1 and the between-person Level 2. At Level 1, described by Equation (1) (Lafit et al., 2020), each participant's first-order autoregressive process is modelled: The person-specific autoregressive parameter γ_{1i} quantifies to what degree the process value esm_{it} of participant i at time t depends on the lagged process value $esm_{i,t-1}$. The person-specific intercept γ_{0i} represents the expected process value esm_{it} when the lagged variable $esm_{i,t-1}$ equals 0 (Jongerling et al., 2015). The innovation ϵ_{it} (i.e., residuals, the part of the process variance that is not explained by the lagged variable $esm_{i,t-1}$) is assumed to be independent and coming from a normal distribution with mean of 0 and variance σ^2_ϵ (Lafit et al., 2020). The model used in the present thesis assumes the innovation variance to be identical for all participants.

$$esm_{it} = \gamma_{0i} + \gamma_{1i} * esm_{i,t-1} + \epsilon_{it} \tag{1}$$

In the multilevel AR(1) model, the person-specific autoregressive effects γ_{1i} and the person-specific intercepts γ_{0i} are allowed to vary between participants. The Level 2 of the MLAR(1) model describes this between-person variability. The Level 2 is defined in Equation (2). Each person-specific autoregressive effect γ_{1i} is a sum of a fixed effect β_{10} and a person-specific random effect ν_{1i}. Similarly, the person-specific intercepts γ_{0i} are a sum of a fixed effect β_{00} and a random effect ν_{0i}. The random effects come from a bivariate normal distribution with variances denoted as $\sigma^2_{\nu_0}$ and $\sigma^2_{\nu_1}$ and correlation $\sigma_{\nu_{01}}$.

$$\begin{aligned} \gamma_{0i} &= \beta_{00} + \nu_{0i} \\ \gamma_{1i} &= \beta_{10} + \nu_{1i} \end{aligned} \tag{2}$$

To illustrate the process of estimating a multilevel autoregressive model, let us imagine a researcher who, similarly to Houben and Kuppens (2020), is interested in whether the inertia of negative affect is associated with borderline personality disorder features. To do so, she collects ESM data from 100 participants, obtaining self-reported values 10 times a day for 7 days via a mobile app.

The inertia of negative affect for each participant is operationalized as the person-specific autoregressive effect γ_{1i}, i.e., to what degree the intensity of negative affect at time t is influenced by the intensity at the previous timepoint, *t-1*. However, given the fluctuating nature of human affect, the magnitude of negative affect can almost never be perfectly predicted from its previous value. For example, the participant can feel worse as a consequence of an external event, or actively use emotion regulation techniques to manage their negative feelings. This variability in negative affect that cannot be predicted from the previous affect magnitude is included in the innovation parameter ϵ_{it} (Jongerling et al., 2015).

Each participant's person-specific autoregressive effect γ_{1i} consists of two components: the fixed effect β_{10}, which is identical across all participants, and the random effect ν_{1i}, which is estimated for each participant individually. As such, the person-specific effects are assumed to differ while also being assumed to come from the same distribution.

Assumptions of the MLAR(1) model

In this part, the assumptions of the MLAR(1) model will be explained.

Stationarity. The MLAR(1) model is used to model stable processes in which no temporal trends (i.e., changes in the process mean over time) are present. As such, it assumes weak stationarity: the (person-specific) process mean, innovation variance, and autoregressive parameter are assumed to not change throughout the time series (Rovine & Walls, 2006). Thus, the person-specific autoregressive effects γ_{1i} are assumed to be bounded by -1 and 1, as autoregressive effects larger than 1 (or lower than -1) cause a change in the process mean (Krone et al., 2016).

Equally spaced measurements. The time-periods that elapsed between each pair of consecutive measurement occasions are assumed to be equal in the following simulation study. In real-life ESM data, the lagged value of the last observation of each day is usually set as missing to account for the fact that the gap between the last night beep and the first morning beep is much larger than the time-gap between the other observations.

Estimation procedures

In the following subchapter, I will present two most-used approaches to estimating the MLAR (1) model in psychology: the maximum likelihood estimation (MLE) and Bayesian Markov Chain Monte Carlo (MCMC) estimation.

Maximum likelihood estimation (MLE). Thanks to its availability in standard software, such as the *nlme* R package (Pinheiro et al., 2022), ML is the most popular approach to estimating MLAR(1) models (Jongerling et al., 2015). The Full Information Maximum Likelihood method, which includes the regression coefficients and the variance components in the likelihood, is usually used for the estimation in combination with the Broyden-Fletcher-Goldfarb-Shanno algorithm, specifically designed to estimate stationary autocorrelation parameters (Krone et al., 2016).

Bayesian estimation. Bayesian MCMC is a flexible way of estimating the MLAR(1) model (Krone et al., 2016). When either the outcome or the lagged predictor is missing, it allows avoiding list-wise deletion of observation-pairs when either the outcome or the predictor are missing by using the estimated autoregressive parameter to estimate the value of the missing observations (Krone et al., 2016).

Evidence from simulation studies

Multiple simulation studies about the statistical properties of the MLAR(1) model have been published in recent years. In this subchapter, I will summarise the most important findings.

Jongerling et al. (2015) found that modelling innovation variance as fixed (i.e., identical for all participants) instead of random when it actually differs across participants leads to a considerable bias in the estimation of the fixed AR effect. There is an upward bias (overestimation) present when the correlation between the individual AR effects and individual innovation variances is positive, and vice versa. Additionally, Jongerling et al. point out that using the person-means to center the lagged predictor variable leads to a downward bias in the estimation of the fixed AR effect. The effect of person-mean centering the predictor on the estimation performance of the MLAR model was further studied by Hamaker & Grasman (2015). Their simulation study confirmed that person-mean centering leads to an underestimation of the fixed autoregressive effect, especially when the number of time points per participant *($T.obs$)* is low. Still, they recommend using person-mean centering when one is interested in the effect of a between-person predictor on inertia.

In their simulation study comparing the maximum likelihood and Bayesian approaches to estimating the MLAR model, Krone et al. (2016) show that the two estimation procedures have a comparable performance. Furthermore, a higher number of time points per participant leads to more precise estimates, while the effect of the number of participants on the estimation performance is small. They also show that a higher variance of the random AR effects leads to a worse estimation precision and that the estimation bias gets smaller when the real fixed AR effect increases.

Liu (2017) assessed how violating the normality of the random AR effect distribution influences the estimation performance of the MLAR model. The different distributions of the random AR effects were found to only have a small effect on the estimation performance.

While the simulation studies mentioned above provide an extensive body of evidence about the statistical properties of the MLAR model under different conditions, several questions remain unanswered. One of them is the effect of missing observations on estimation performance. The presence of missing values in an intensive longitudinal dataset decreases the number of observations per participant (or, more specifically, the number of observation-pairs that can be used for the estimation of the model). As such, it can be expected that lower compliance (i.e., lower proportion of ESM beeps that the given participant answered) will make estimation bias more severe. Additionally, different patterns of missingness might have different consequences on the estimation performance. Ji et al. (2018) show that the presence of data missing completely at random (MCAR), missing at random (MAR) and not missing at random (MNAR) leads to a considerable bias in point estimates of cross-lagged and autoregressive parameters in person-specific ($N = 1$) vector autoregressive models when list-wise deletion is used. However, no similar evidence is available about the MLAR(1) model.

Methods

The goal of the present exploratory simulation study is to assess the effects of four different patterns of missing data (data missing completely random, data missing in blocks, and two patterns of missingness dependent on the process value; see Figure 1) and study compliance on estimation performance and bias, standard error, and statistical power for the estimation of the fixed autoregressive effect in the MLAR(1) model. Apart from the missingness patterns and compliance, we manipulated the number of participants, the number of time points per participant, the simulated fixed autoregressive effect, and the variance of random AR effects. The values of the manipulated variables for both studies are reported in Table 1. The values of the manipulated variables were set considering realistic research questions in psychology. The study was exploratory; no apriori hypotheses were tested.

Simulation procedure

The study followed the general principles of the Monte Carlo simulation procedure described by Lane & Hennes (2018).

Simulation conditions. Two simulation studies, Simulation A and Simulation B, were carried out to investigate the research questions. In Simulation A, no random autoregressive effects were simulated and estimated (i.e., each subject's time-series had the same simulated autoregressive effect, and only fixed autoregressive effects were estimated). In Simulation B, random autoregressive effects were simulated and estimated, with the random effects variance set to either 0.05 or 0.1 and the correlation between the random slopes and random intercepts set to 0. Both random and fixed intercepts were estimated in Simulations A and B. The multilevel autoregressive model estimated in Simulation A is defined in Equation (3), while Equation (4) describes the model estimated in Simulation B.

$$\begin{aligned} esm_{it} &= \gamma_{0i} + \gamma_{1i} * esm_{i,t-1} + \epsilon_{it} \\ \gamma_{0i} &= \beta_{00} + \nu_{0i} \\ \gamma_{1i} &= \beta_{10} \end{aligned} \tag{3}$$

$$\begin{aligned} esm_{it} &= \gamma_{0i} + \gamma_{1i} * esm_{i,t-1} + \epsilon_{it} \\ \gamma_{0i} &= \beta_{00} + \nu_{0i} \\ \gamma_{1i} &= \beta_{10} + \nu_{1i} \end{aligned} \tag{4}$$

Simulation A followed a 4 × 2 × 3 × 4 × 3 factorial design (yielding 288 simulation conditions in total), and Simulation B followed a 4 × 2 × 2 × 4 × 2 × 2 design (256 conditions in total). 1,000 replicates per cell (i.e., a combination of simulation conditions) were simulated. As such, 544,000 datasets were generated (and the same number of models was estimated) in this simulation study. The manipulated variables are listed in Table 1, and the parameters that remained fixed throughout all simulation conditions

Table 1: Values of the manipulated parameters used in the two simulation studies

Manipulated parameter	Simulation A	Simulation B
Missingness pattern	MCAR, block, extreme-onesided, extreme-twosided	MCAR, block, extreme-onesided, extreme-twosided
Simulated fixed AR effect	0.3, 0.5, 0.7	0.3, 0.7
Variance of random AR effects	-	0.05, 0.1
Compliance	0.4, 0.6, 0.8, 1	0.4, 0.6, 0.8, 1
Number of participants (N)	20, 50	20, 50
Time points per participant (T.obs)	20, 50, 100	50, 100

Table 2: Parameters used for the two simulation studies.

Simulation parameter	Simulation A	Simulation B
Fixed intercept	0	0
Variance of random intercepts	3	3
Innovation variance	3	3
Correlation between random intercepts and random slopes	0	0
Significance threshold	0.05	0.05
Simulation replicates per cell	1000	1000

are reported in Table 2.

Data generation. First, for each of the simulation conditions (i.e., combination of the parameters listed below), 1,000 synthetic datasets were generated. Each dataset contained observations from N simulated participants. A temporally dependent time-series of length *T.obs* was generated as nested within each simulated participant via a recursive equation. Additionally, for each time-series, a burn-in period containing 1,000 observations was generated and later discarded. The within-person error (innovation) vector ϵ_i was generated from a N(0, σ) distribution with σ set to 3 in all simulations. The fixed intercept β_{00} was set to 0 across all conditions. The random intercepts ν_{0i} for each simulated time-series were sampled from a $N(0, 3)$ distribution in both studies. In Simulation A, only fixed autoregressive effects β_{10} were simulated and manipulated, while both fixed autoregressive effects β_{10} and random autoregressive effects ν_{1i} were included in Simulation B. No night gaps were assumed in the simulations. For an overview of the values of all manipulated simulation parameters, please refer to Table 1.

Each time-series was then generated using Equation (1). The initial value was generated as a sum of the person-specific intercept γ_{0i} and the innovation ϵ_{ij}, and the following observations were calculated by multiplying the value of the time-series at *t-1* by the person-specific autoregressive effect γ_{1i} and adding the person-specific intercept γ_{0i} and the innovation ϵ_{ij}. Subsequently, after removing the burn-in datapoints, the first-order lagged version of the time-series was generated, setting the first lagged value as missing.

The non-manipulated simulation parameters $(\beta_{00}, \sigma_{\nu 0}, \sigma, \rho_{\nu})$ were set following a simulation design from Hamaker & Grasman (2015).

Introduction of missing values. After generating the time-series, missing data were introduced to each of the generated datasets according to the missing data pattern and compliance of the given simulation condition. Four different missingness patterns (corresponding to the hypothetical ESM study scenarios described in the Introduction) were introduced to the data. For an illustration of the different missing data patterns, see Figure 1.

a) Data missing completely at random (MCAR). In the simulation conditions with data MCAR, (1-compliance) observations were set at missing using the *delete_MCAR* function from the *missMethods R* package (Rockel, 2022). Each observation had an identical probability of being set as missing. For example, in the conditions with compliance of 0.6, each observation had probability of 0.4 of being set as missing.

b) Data missing in blocks of consecutive observations. This missingness pattern corresponds to a simplified situation in which the participants miss multiple consecutive observations, for example because they cannot respond to ESM beeps during work or social occassions. One block of missing observations was introduced in each individual time-series. For each simulated participant, the block started with the 5th observation of their time-series, and its length was directly proportional to compliance. For example, in a condition with compliance of 0.6 an 100 timepoints per participant, 40 (100×0.4) observations were set as missing, from beep number 5 to beep number 45.

c) Most extreme observations, one side. In missing patterns c) and d), missingness was directly dependent on the process value. In missingness pattern c), all observations below the (100%-compliance) percentile were set as missing. For example, with compliance set to 0.6, all observations below the 40th percentile were set as missing. The value of the threshold quantile was computed individually for each time-series.

d) Most extreme observations set as missing, two sides. In condition d), the highest and lowest (100%-compliance)/2 observations were set as missing. For example, with compliance set to 0.6, all observations below the 20th and above the 80th percentile were set as missing.

It can be expected that the different missingness patterns will differ in their effects on the simulation outcomes (estimation bias, standard error, power). Even with an identical proportion of missing data, datasets with different missingness patterns will have different proportions of effective observation-pairs (i.e., proportion of time points for which both the observation at t and the observation at t-1 are not missing) used to estimate the autoregressive effect. For example, when the data are missing in block, all missing observations follow each other, which means that most non-missing values can be used for estimation (i.e., both the observation at t and the lagged value will be non-missing). On the other hand, data MCAR will result in a smaller number of missing observation pairs, as there will be more timepoints for which either the process value or its lagged version will be missing. As an illustration, the average number of effective observation-pairs for each missingness pattern when the number of participants is 100, the number of timepoints is 100, and the compliance is 0.6 is included in Table 3.

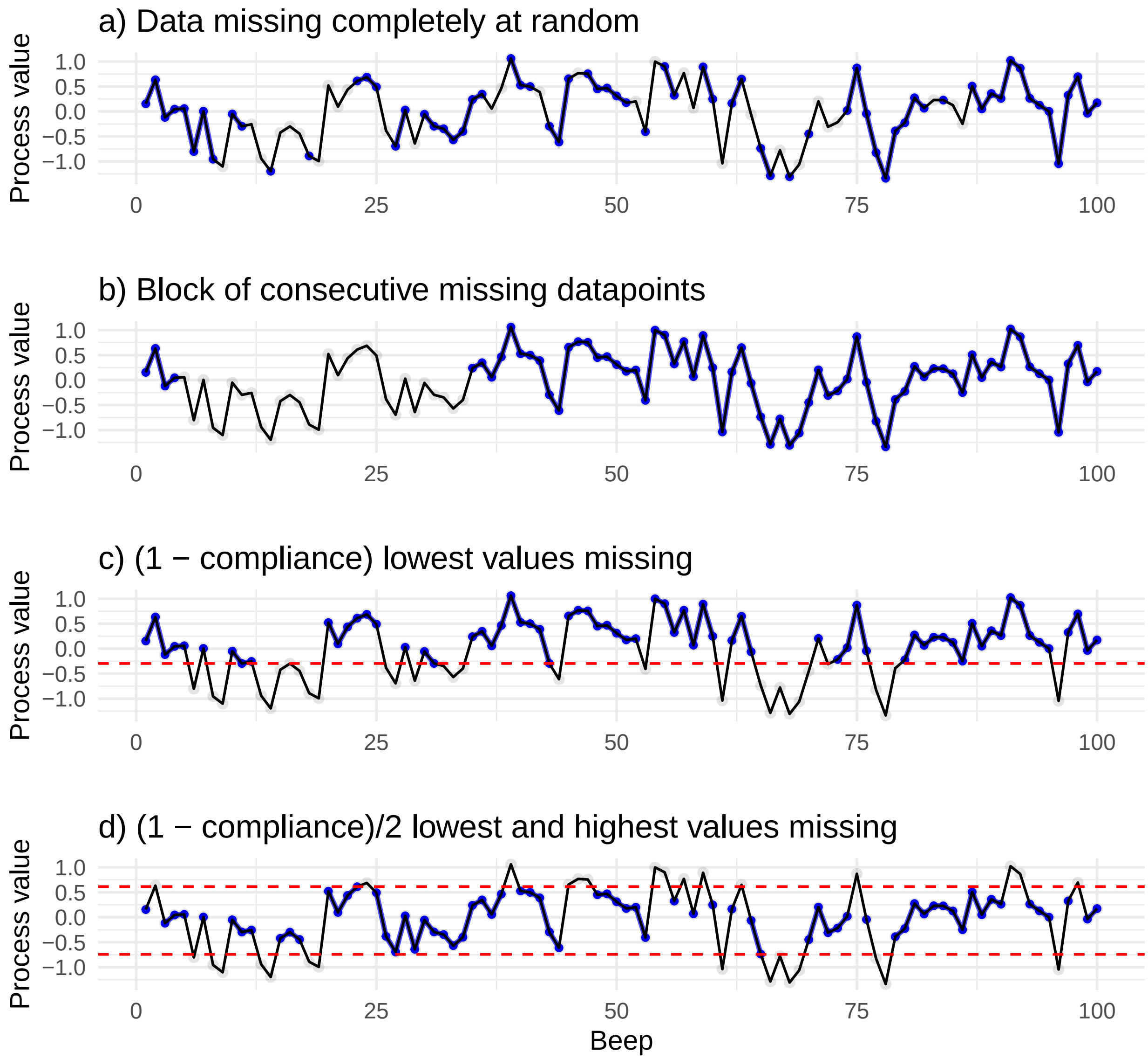

Figure 1: Illustration of the four different missingness pattern used in the simulation study. The blue dots represent observed datapoints, while the light gray dots represent missing values. Compliance is 0.7 in all four patterns. The total number of observations in each time-series is 100. There are 30 missing values in each time-series. The dashed red lines in plots c) and d) represent the threshold percentiles.

Table 3: Average number of effective observation-pairs per dataset for each missingness pattern. Number of participants = 100, timepoints per participant = 100, compliance = 0.6.

Missingness type	Effective observation-pairs
block	5700.00
extreme_oneside	4317.96
extreme_twosided	3866.45
mcar	3540.10

Fitting a multilevel autoregressive model. After missing values were introduced to the data, a MLAR(1) model was fitted to each of the simulated datasets using the *lme* function from the *nlme* R package (Pinheiro et al., 2022) with the value of the time-series at t as the outcome, the lagged *(t-1)* value of the time-series as the predictor, and the participant number as the grouping variable. We then extracted relevant parameters from the models that converged successfully. Missing values were treated by list-wise deletion. The restricted maximum log-likelihood method with the Broyden-Fletcher-Goldfarb-Shanno optimization algorithm was used to estimate the model.

Following the recommendations by Hamaker & Grasman (2015), the predictor (lagged variable) was person-mean centered (i.e., the observed mean of each respective participant's ESM process was subtracted from the value of the lagged variable at each time point) in the main simulation studies. Although person-mean centering results in an underestimation of the autoregressive effect (Hamaker & Grasman, 2015), it allows for a clearer interpretation of the within-person effects in multilevel models (Enders & Tofighi, 2007; Hamaker & Muthén, 2020). As a supplementary analysis, we also conducted the simulations without person-mean centering the predictor.

Simulation outcomes. Estimation bias, the standard error of the estimation, and the statistical power to estimate the fixed autoregressive effect β_{10} were the focal outcomes of the study. Additionally, we examined the effect of the manipulated variables on the proportion of models that successfully converged and the bias in the estimation of the person-mean used for centering of the predictor (lagged) variable.

Estimation bias was computed as the difference between the estimated fixed autoregressive effect $\hat{\beta}_{10}$ and the real (simulated) fixed autoregressive effect β_{10} in each simulation replicate. As such, the dataset with estimation bias contained 1,000 rows per simulation condition (the reported values are the average of the 1,000 results per conditions).

Standard error (SE) of the fixed autoregressive effect and statistical power were calculated for each simulation condition (i.e., 1 row per condition). Statistical power was computed as the proportion of simulation replicates (within the given simulation condition) in which the *p*-value for the estimated fixed autoregressive effect $\hat{\beta}_{10}$ was below the significance threshold ($\alpha = 0.05$) and the number of simulation replicates that converged successfully.

Reproducibility and code/data availability

The simulations were conducted in R version 4.2.1 (R Core Team, 2021). The study was conducted with emphasis on reproducibility of the results (Pawel et al., 2022). As such, we provide all data (simulation results) used for the reported analyses, as well as the full reproducible R code for the simulations (including the custom functions created for the purposes of the study), and the code used to generate the plots and result tables (available at https://github.com/benjsimsa/AR-missing-simulations) The repository also includes a *sessionInfo* document that lists the versions of the packages used for the study. The thesis was written using dynamic reporting in R Markdown (Allaire et al., 2022).

Additionally, the *renv* R package (Ushey, 2022) was used to set up a reproducible R environment and improve reproducibility by creating a project-local package library. For reproducible file referencing, the R package *here* (Müller, 2020) was used. For more information about the custom functions, simulation code, and the structure of the GitHub repository itself, please refer to the file README.md in the repository.

Results

Simulation A

The descriptive results for all 288 conditions included in Simulation A are reported in Table 20 (see Appendix).

Outcome: Estimation bias

ANOVA. We used a 4 × 2 × 3 × 4 × 3 factorial Type I ANOVA (with estimation bias for the fixed autoregressive effect as an outcome and number of participants, number of time points per participant, missingness type, compliance, and the simulated fixed autoregressive effect, as well as their two-ways interactions, as predictors) to assess which of the manipulated factors had a considerable influence on estimation bias. The results from every simulation run (i.e., 1,000 results per condition = 288,000 rows) were combined into a single dataset for the analysis. Given the very large sample size (which would make even negligible differences statistically significant) and the exploratory character of the analysis, p-values and significance thresholds were not used make inferences. Instead, we used a threshold of 0.14 for the partial ω^2, indicating a large effect size (Field et al., 2012). This cutoff will be used for all ANOVA results throughout the Results section. The partial ω^2 was chosen as the less biased alternative to partial η^2 (Okada, 2013). The results and effect sizes are reported in Table 4.

Four main effects above the effect size threshold of 0.14 were found: the main effect of missingness type ($\omega^2 = 0.73$), compliance ($\omega_p^2 = 0.63$), the number of time points per participant ($\omega_p^2 = 0.26$), and the simulated fixed slope ($\omega_p^2 = 0.14$). Furthermore, the interaction between the missingness type and compliance ($\omega_p^2 = 0.54$) had an effect size above the cut-off.

The main effects of missingness type and compliance are visualised in Figure 2 and Figure 3 (respectively), while the interaction between missingness type and compliance is depicted in Figure 4.

Figure 2 shows that while the underestimation of the fixed slopes is fairly low (although still considerable) when the observations are missing completely at random or in block, it becomes more severe when only the most extreme values (both at one side and at both sides) are missing. Additionally, the underestimation of the fixed slopes becomes more severe as the compliance gets lower.

Zooming in on the interaction between compliance and missingness type (Figure 4) suggests that the effect of compliance on estimation bias is dramatically more severe for the two conditions in which the most extreme values of the process were set as missing (as compared to the other two conditions, i.e., data MCAR and missing in blocks). In the worst-case scenario (low compliance of 0.4; the most extreme values on both sides missing), the average estimation bias was -0.48. The average estimation bias for all combinations of missingness type and compliance (averaged over the different values of the number of participants, time points per participant and simulated fixed slope) is reported in Table 5.

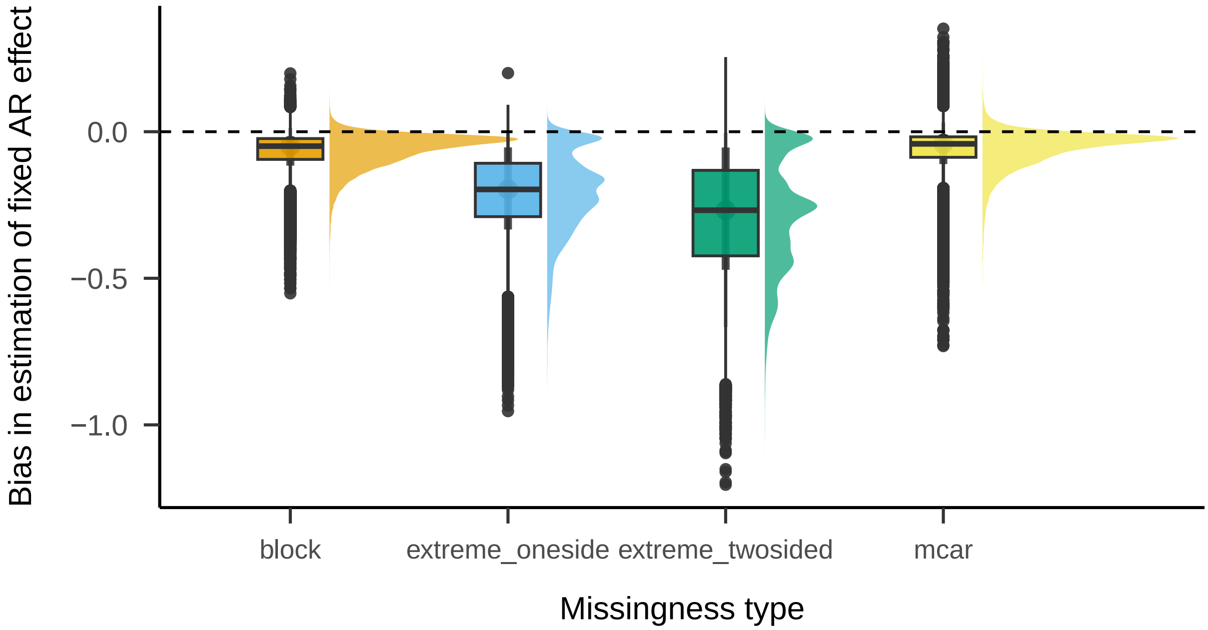

Figure 2: The effect of compliance on the bias in estimation of the fixed slopes.

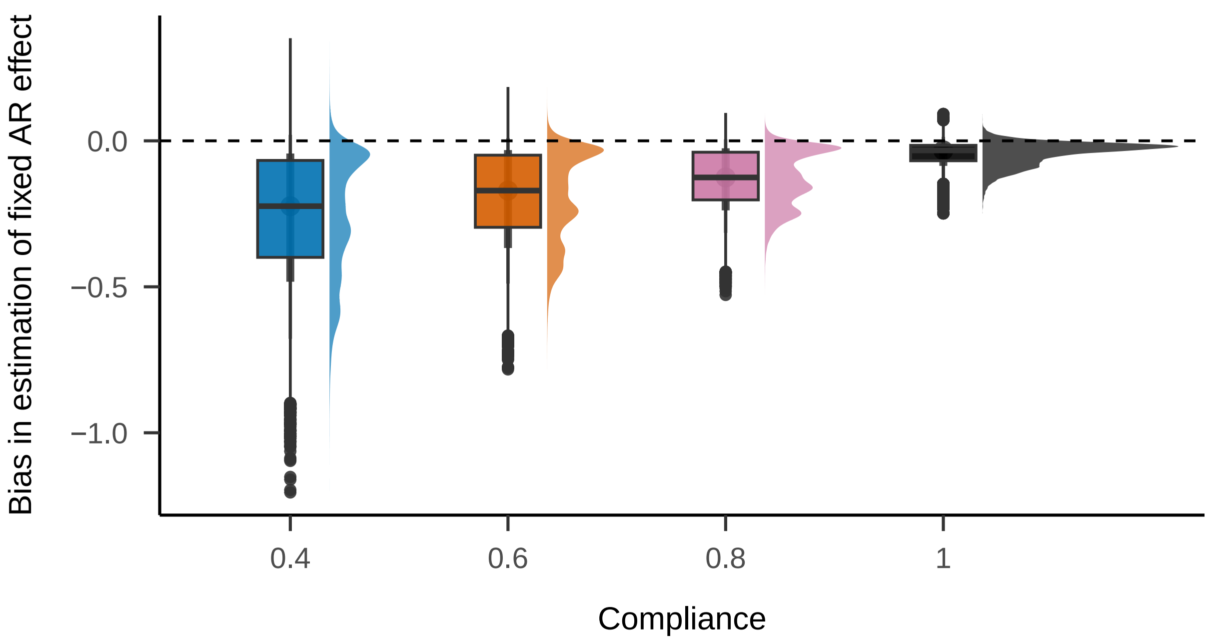

Figure 3: The effect of compliance on the bias in estimation of the fixed slopes.

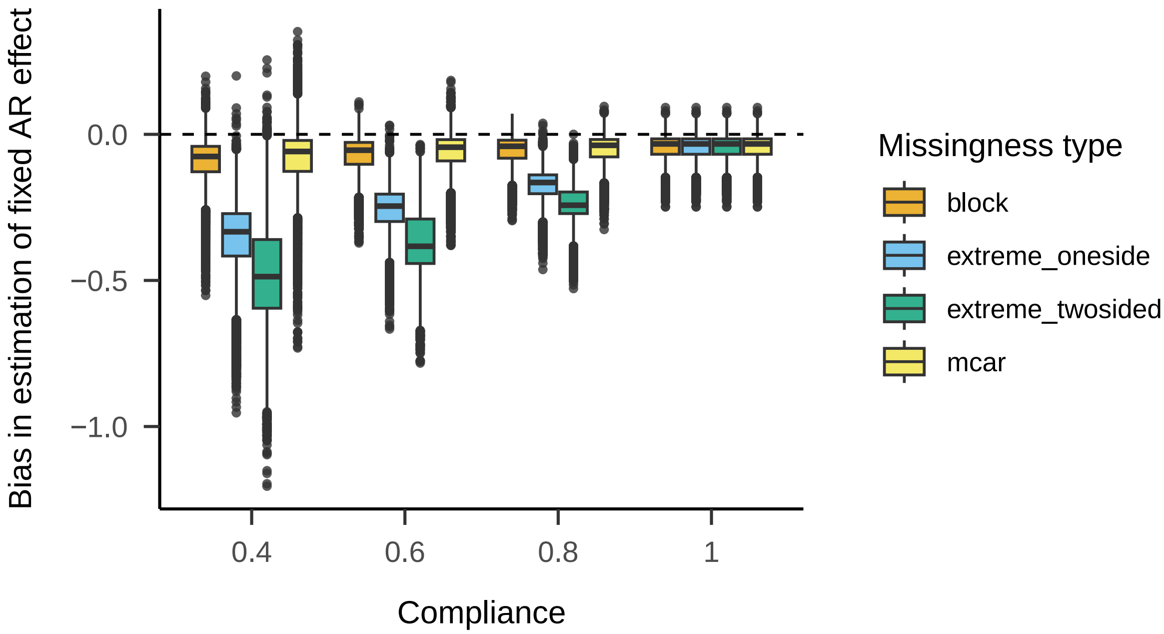

Figure 4: The effect of the interaction between missingness type and compliance on the bias in estimation of the fixed slopes.

Table 4: ANOVA results, simulation A. Outcome: Estimation bias

	Df	Sum Sq	Mean Sq	F value	p-value	Partial omega-squared
N	1	0.11	0.11	32.53	<0.001	0.00
T.obs	1	354.93	354.93	101753.95	<0.001	0.26
miss_type	3	2657.10	885.70	253921.29	<0.001	0.73
compliance	1	1706.99	1706.99	489377.69	<0.001	0.63
B1_sim	1	169.00	169.00	48449.29	<0.001	0.14
N:T.obs	1	0.02	0.02	4.40	0.0360	0.00
N:miss_type	3	0.00	0.00	0.31	0.8216	0.00
T.obs:miss_type	3	14.11	4.70	1348.32	<0.001	0.01
N:compliance	1	0.03	0.03	8.09	0.0044	0.00
T.obs:compliance	1	22.78	22.78	6529.83	<0.001	0.02
miss_type:compliance	3	1157.71	385.90	110634.77	<0.001	0.54
N:B1_sim	1	0.03	0.03	8.41	0.0037	0.00
T.obs:B1_sim	1	1.75	1.75	502.84	<0.001	0.00
miss_type:B1_sim	3	148.38	49.46	14179.70	<0.001	0.13
compliance:B1_sim	1	59.28	59.28	16994.38	<0.001	0.06
Residuals	287974	1004.48	0.00		NA	

Table 5: Simulation A. Average bias in estimation of the fixed slope for each combination of missingness type and compliance.

	Missingness type			
compliance	block	extreme_oneside	extreme_twosided	mcar
0.4	-0.09	-0.36	-0.48	-0.08
0.6	-0.07	-0.26	-0.37	-0.06
0.8	-0.05	-0.17	-0.24	-0.05
1.0	-0.04	-0.04	-0.04	-0.04

Table 6: Simulation A. Average standard error in the estimation of the fixed slope for each combination of number of participants, number of time points/participant, and compliance.

		Compliance			
N	T.obs	0.4	0.6	0.8	1
20	20	0.14	0.08	0.06	0.05
	50	0.07	0.05	0.04	0.03
	100	0.05	0.03	0.02	0.02
100	20	0.06	0.04	0.03	0.02
	50	0.03	0.02	0.02	0.01
	100	0.02	0.01	0.01	0.01

Outcome: Standard error

Descriptive statistics. The average standard errors for the different combinations of number of participants, time points per participant and compliance are reported in Table 6.

ANOVA. To examine the effect of the manipulated parameters on the standard error of the estimation of the fixed slopes, we combined the results for each condition (1,000 simulation runs) into a single row. As such, the dataset used for the following analyses had 288 rows in total. A 4 $\times$ 2 $\times$ 3 $\times$ 4 $\times$ 3 factorial Type I ANOVA was used to analyse the data. The full ANOVA results and effect sizes are reported in Table 7.

The main effects of the number of participants (ω_p^2 = 0.68), number of time points per participant (ω_p^2 = 0.68) and compliance (ω_p^2 = 0.66) crossed the cut-off for effect size.

Furthermore, the interactions between the number of time points per participant and compliance (ω_p^2 = 0.28), number of participants and time points per participant (ω_p^2 = 0.24), and between the number of participants and compliance (ω_p^2 = 0.22) were found.

Figure 5 depicts the interaction between the number of time points per participant and compliance, while Figure 6 shows the interaction between the number of participants and compliance.

Table 7: ANOVA results, simulation A. Outcome: Standard error

	Df	Sum Sq	Mean Sq	F value	p-value	Partial omega-squared
N	1	0.06	0.06	625.92	<0.001	0.68
T.obs	1	0.06	0.06	621.16	<0.001	0.68
miss_type	3	0.00	0.00	14.11	<0.001	0.12
compliance	1	0.05	0.05	556.59	<0.001	0.66
B1_sim	1	0.00	0.00	21.75	<0.001	0.07
N:T.obs	1	0.01	0.01	91.92	<0.001	0.24
N:miss_type	3	0.00	0.00	2.13	0.096	0.01
T.obs:miss_type	3	0.00	0.00	1.48	0.220	0.00
N:compliance	1	0.01	0.01	82.89	<0.001	0.22
T.obs:compliance	1	0.01	0.01	114.06	<0.001	0.28
miss_type:compliance	3	0.00	0.00	13.31	<0.001	0.11
N:B1_sim	1	0.00	0.00	3.14	0.078	0.01
T.obs:B1_sim	1	0.00	0.00	1.37	0.243	0.00
miss_type:B1_sim	3	0.00	0.00	0.20	0.895	0.00
compliance:B1_sim	1	0.00	0.00	1.55	0.214	0.00
Residuals	262	0.03	0.00		NA	

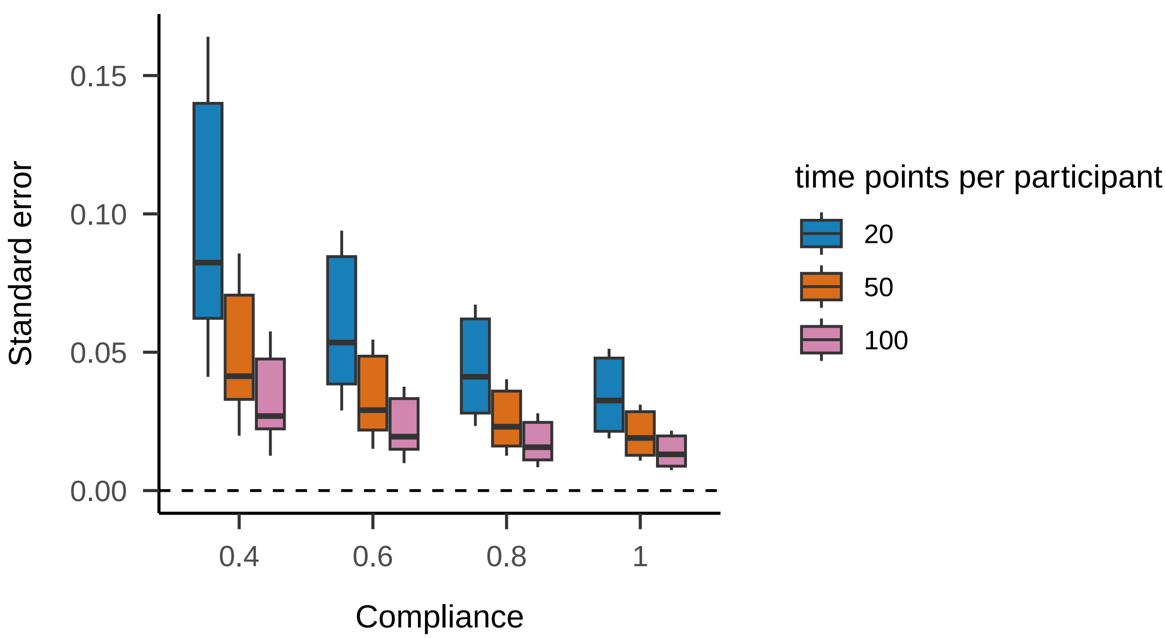

Figure 5: The effect of the interaction between number of time points and compliance on standard error of estimation of the fixed slopes. Simulation A.

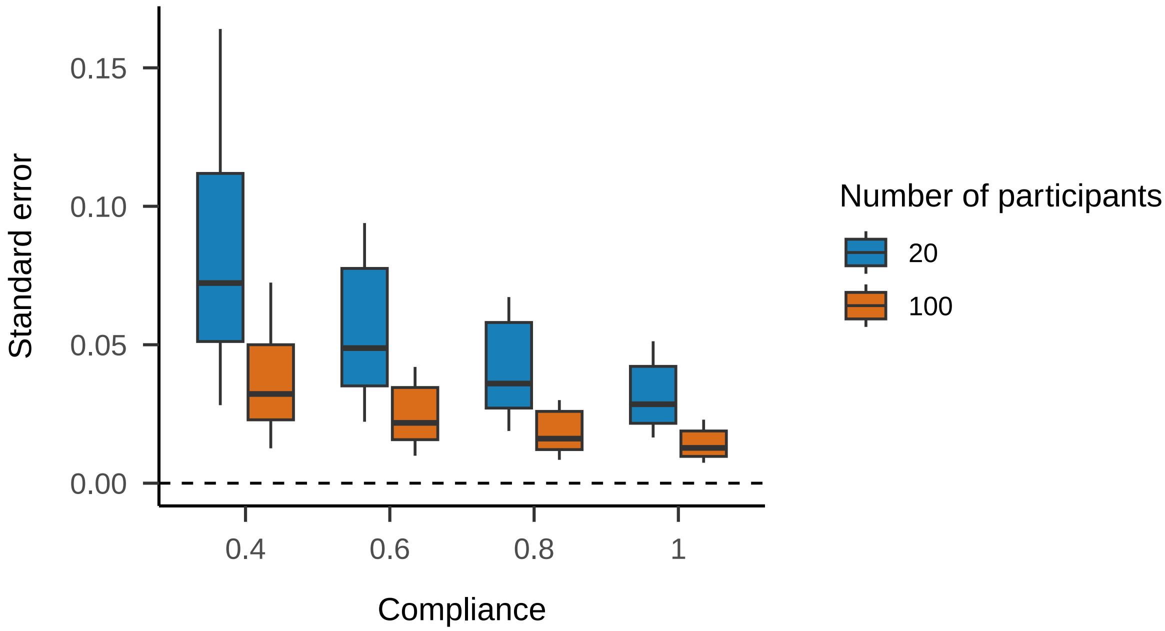

Figure 6: The effect of the interaction between number of participants and compliance on standard error of estimation of the fixed slopes. Simulation A.

Outcome: Statistical power

Descriptive statistics. The statistical power for each combination of the manipulated parameters is reported in Table 20. As an illustration, the effects of compliance, missingness type, the number of participants and the number of time points per participant when the simulated fixed slope is 0.3 are visualised in Figure 7. Consistent with the results about estimation bias, statistical power is the lowest in the two conditions with the most extreme datapoints missing. For the conditions with data missing completely at random and data missing in consecutive blocks, power is very high even when the compliance is low for most conditions (except for the two conditions with T = 20).

A peculiar pattern in the plot is worth pointing out: in the two conditions with T = 20 and the most extreme data missing at both sides (green dashed line), the statistical power is higher when compliance is 0.4 compared to when compliance is 0.6. This counterintuitive result is likely due to the fact that the underestimation is the most severe when the most extreme values at both sides are missing. As such, some of the estimates of the fixed slope will be negative, and their magnitude will be large enough for them to reach statistical significance.

ANOVA. A 4 × 2 × 3 × 4 × 3 factorial Type I ANOVA was used to analyse the effect of the manipulated parameters (288 conditions in total) on statistical power. The results are reported in Table 8.

Four main effect above the cut-off for the effect size were found: the effect of compliance ($\omega_p^2 = 0.43$), of missingness type ($\omega_p^2 = 0.36$), simulated fixed slope ($\omega_p^2 = 0.18$), and the effect of the number of time points per participant ($\omega_p^2 = 0.17$).

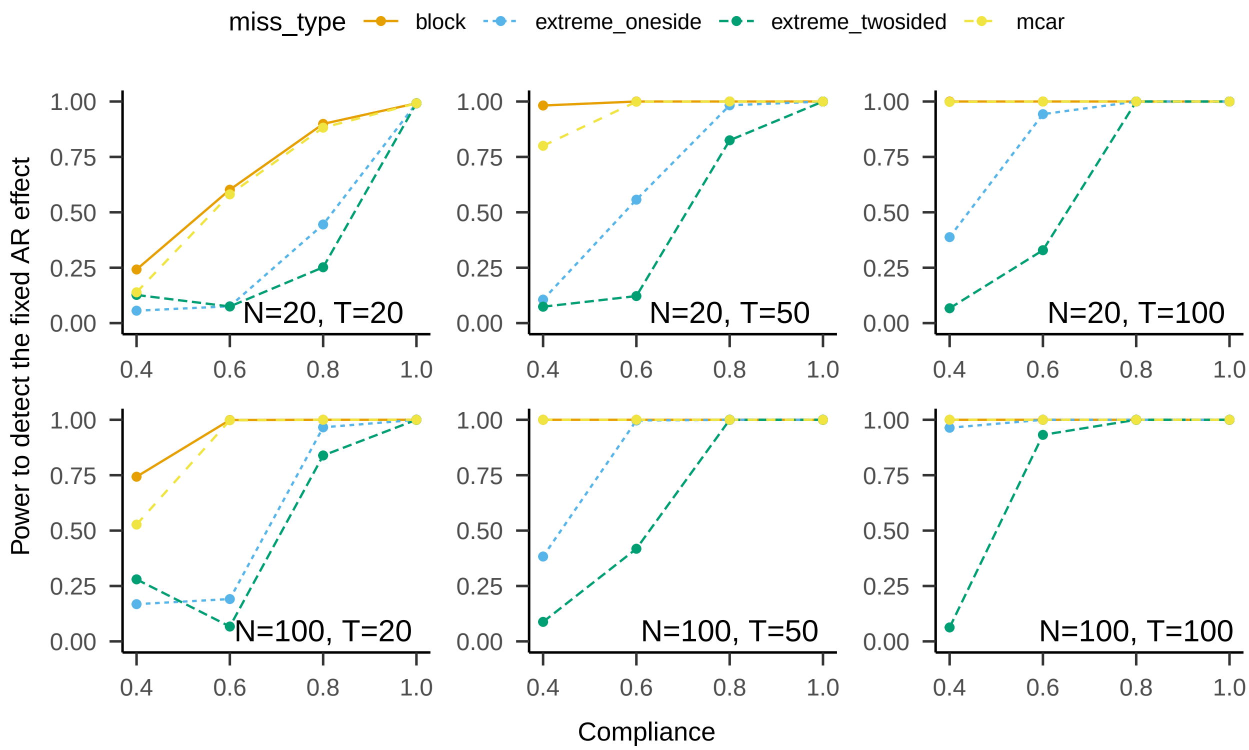

Figure 7: Simulation A. Statistical power to detect the fixed slope for all combinations of compliance, missingness type, number of participants and time points per participant when the simulated fixed slope is 0.3.

Table 8: ANOVA results, simulation A. Outcome: Power to detect the fixed slope

	Df	Sum Sq	Mean Sq	F value	p-value	Partial omega-squared
N	1	0.42	0.42	20.32	<0.001	0.06
T.obs	1	1.21	1.21	58.92	<0.001	0.17
miss_type	3	3.36	1.12	54.56	<0.001	0.36
compliance	1	4.47	4.47	217.19	<0.001	0.43
B1_sim	1	1.35	1.35	65.78	<0.001	0.18
N:T.obs	1	0.08	0.08	3.80	0.0524	0.01
N:miss_type	3	0.10	0.03	1.66	0.1753	0.01
T.obs:miss_type	3	0.34	0.11	5.49	0.0011	0.04
N:compliance	1	0.22	0.22	10.93	0.0011	0.03
T.obs:compliance	1	0.79	0.79	38.66	<0.001	0.12
miss_type:compliance	3	3.12	1.04	50.65	<0.001	0.34
N:B1_sim	1	0.09	0.09	4.25	0.0403	0.01
T.obs:B1_sim	1	0.24	0.24	11.50	<0.001	0.04
miss_type:B1_sim	3	0.44	0.15	7.19	<0.001	0.06
compliance:B1_sim	1	0.76	0.76	37.10	<0.001	0.11
Residuals	262	5.39	0.02		NA	

Simulation B

In Simulation B, random AR effects were included both in the data generating procedure and in the estimated models. The variance of random AR effects ($\sigma^2_{\nu 1}$) was manipulated as an additional simulation factor (2 values: 0.05 and 0.1). The correlation between the random slopes and intercepts was set to 0. For an overview of all manipulated and fixed simulation parameters, please refer to Table 1. The descriptive results for all 256 simulation conditions are reported in Table 21 in the Appendix.

Outcome: Estimation bias

To evaluate the effect of the number of participants, number of time points per participant, missingness type, compliance, the variance of random AR effects, and the simulated fixed autoregressive effect on the bias in the estimation of the fixed AR effect in Simulation B, a $4 \times 2 \times 2 \times 4 \times 2 \times 2$ factorial Type I ANOVA was used. The results from every simulation run (256 conditions * 1000 runs per condition) were combined into a single dataset. An identical inference criterion ($\omega^2_p \geq 0.14$) as in Simulation A was used. The ANOVA results are listed in Table 9.

Interestingly, compared to the results from Simulation A (see Table 4), the effect of the number of observations per participant on estimation bias ($\omega^2 = 0.04$) is much smaller and does not reach the effect size threshold. The three main effects that do reach the cut-off in Simulation B are the effect of missingness type ($\omega^2_p = 0.65$), compliance ($\omega^2_p = 0.47$), and the simulated fixed AR effect ($\omega^2_p = 0.46$). The only interaction that reached the cut-off was the interaction between missingness type and compliance ($\omega^2_p = 0.45$). The interaction is visualised in Figure 8. The pattern of the interaction is very similar to the pattern of the interaction between missingness type and compliance in Simulation A (see Figure 4). Interestingly, when compared to the results of Simulation A, the average estimation bias is slightly worse for the MCAR and block missingness types and slightly less severe for the two conditions with the extreme values missing (compare Table 10 and Table 4). However, the overall conclusion of Simulation A holds: there is a considerable downward estimation bias that becomes more severe the lower the compliance rate is, and it is most severe for the condition in which the most extreme data at both sides are missing.

Outcome: Standard error

ANOVA. The results of the $4 \times 2 \times 2 \times 4 \times 2 \times 2$ factorial ANOVA used to assess the influence of the manipulated factors on the standard error are reported in Table 12. Compared to Simulation A, more main effects of the manipulated factors crossed the effect-size cut-off: the effect of number of participants ($\omega^2_p = 0.93$), compliance ($\omega^2_p = 0.54$), number of time points per participant ($\omega^2_p = 0.32$), and the value of the simulated fixed AR effect ($\omega^2_p = 0.27$). Additionally, the interaction between compliance and the number of time points per participant ($\omega^2_p = 0.19$) crossed the effect size threshold.

Figure 9 illustrates the main effect of the number of participants (as the most important factor) on standard error, while Figure 10 shows the interaction between compliance and the number of time points

Table 9: ANOVA results, simulation B. Outcome: Estimation bias

	Df	Sum Sq	Mean Sq	F value	p-value	Partial omega-squared
N	1	0.23	0.23	70.83	<0.001	0.00
T.obs	1	30.99	30.99	9725.42	<0.001	0.04
miss_type	3	1513.93	504.64	158363.00	<0.001	0.65
compliance	1	709.24	709.24	222568.23	<0.001	0.47
sigma_v1	1	33.87	33.87	10629.71	<0.001	0.04
B1_sim	1	701.88	701.88	220257.79	<0.001	0.46
N:T.obs	1	0.01	0.01	3.72	0.054	0.00
N:miss_type	3	0.03	0.01	3.19	0.023	0.00
T.obs:miss_type	3	1.22	0.41	128.00	<0.001	0.00
N:compliance	1	0.15	0.15	48.44	<0.001	0.00
T.obs:compliance	1	2.26	2.26	710.69	<0.001	0.00
miss_type:compliance	3	670.47	223.49	70133.70	<0.001	0.45
N:B1_sim	1	0.04	0.04	12.37	<0.001	0.00
T.obs:B1_sim	1	0.21	0.21	64.44	<0.001	0.00
miss_type:B1_sim	3	117.26	39.09	12265.54	<0.001	0.13
compliance:B1_sim	1	41.68	41.68	13080.86	<0.001	0.05
Residuals	255520	814.25	0.00		NA	

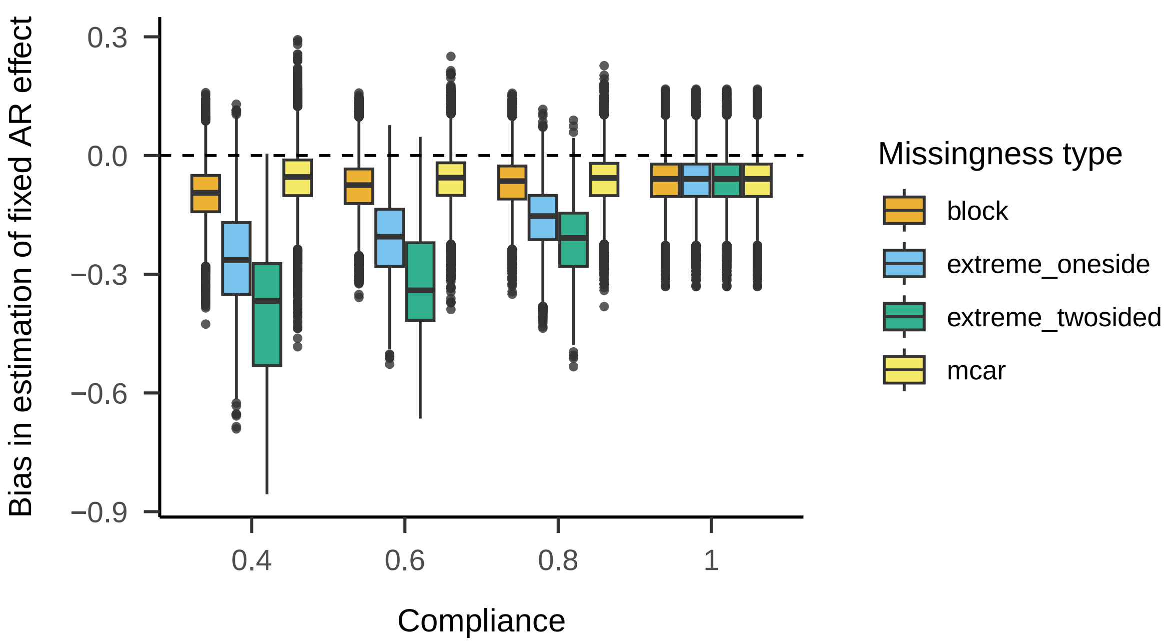

Figure 8: Simulation B: The effect of the interaction between missingness type and compliance on the bias in estimation of the fixed slopes.

Table 10: Simulation B. Average bias in estimation of the fixed slope for each combination of missingness type and compliance.

	Missingness type			
compliance	block	extreme_oneside	extreme_twosided	mcar
0.4	-0.10	-0.26	-0.40	-0.06
0.6	-0.08	-0.21	-0.32	-0.06
0.8	-0.07	-0.16	-0.21	-0.06
1.0	-0.06	-0.06	-0.06	-0.06

Table 11: ANOVA results, simulation B. Outcome: Standard error

	Df	Sum Sq	Mean Sq	F value	p-value	Partial omega-squared
N	1	0.07	0.07	3486.22	<0.001	0.93
T.obs	1	0.00	0.00	120.73	<0.001	0.32
miss_type	3	0.00	0.00	12.08	<0.001	0.11
sigma_v1	1	0.00	0.00	151.15	<0.001	0.37
compliance	1	0.01	0.01	295.94	<0.001	0.54
B1_sim	1	0.00	0.00	97.17	<0.001	0.27
N:T.obs	1	0.00	0.00	16.77	<0.001	0.06
N:miss_type	3	0.00	0.00	1.88	0.133	0.01
T.obs:miss_type	3	0.00	0.00	2.77	0.042	0.02
N:compliance	1	0.00	0.00	36.60	<0.001	0.12
T.obs:compliance	1	0.00	0.00	60.21	<0.001	0.19
miss_type:compliance	3	0.00	0.00	13.12	<0.001	0.12
N:B1_sim	1	0.00	0.00	15.03	<0.001	0.05
T.obs:B1_sim	1	0.00	0.00	3.03	0.083	0.01
miss_type:B1_sim	3	0.00	0.00	21.77	<0.001	0.20
compliance:B1_sim	1	0.00	0.00	2.34	0.127	0.01
Residuals	229	0.00	0.00		NA	

per participant. While the results are comparable to Simulation A, the SE is slightly higher for the same combinations of number of participants / time points per participant in Simulation B.

Outcome: Statistical power

Descriptive statistics. The statistical power for each combination of the manipulated parameters in Simulation B is reported in Table 21 (in the Appendix). The effects of compliance, missingness type, the number of participants and the number of time points per participant when the simulated fixed slope is 0.3 are shown in Figure 11. For the sake of clarity, only the results for simulation conditions in which the $\sigma^2_{\nu 1} = 0.1$ are visualised.

ANOVA. A $4 \times 2 \times 2 \times 4 \times 2 \times 2$ factorial Type I ANOVA was used to analyse the effect of the manipulated parameters on statistical power. The results are reported in Table 13.

Missingness type ($\omega^2_p = 0.4$) together with compliance ($\omega^2_p = 0.38$) were the two simulation factors

Table 12: Simulation B. Average standard error in the estimation of the fixed slope for each combination of number of participants, number of time points/participant, and compliance.

		Compliance			
N	T.obs	0.4	0.6	0.8	1
20	50	0.08	0.07	0.06	0.06
	100	0.06	0.06	0.05	0.05
100	50	0.04	0.03	0.03	0.03
	100	0.03	0.03	0.02	0.02

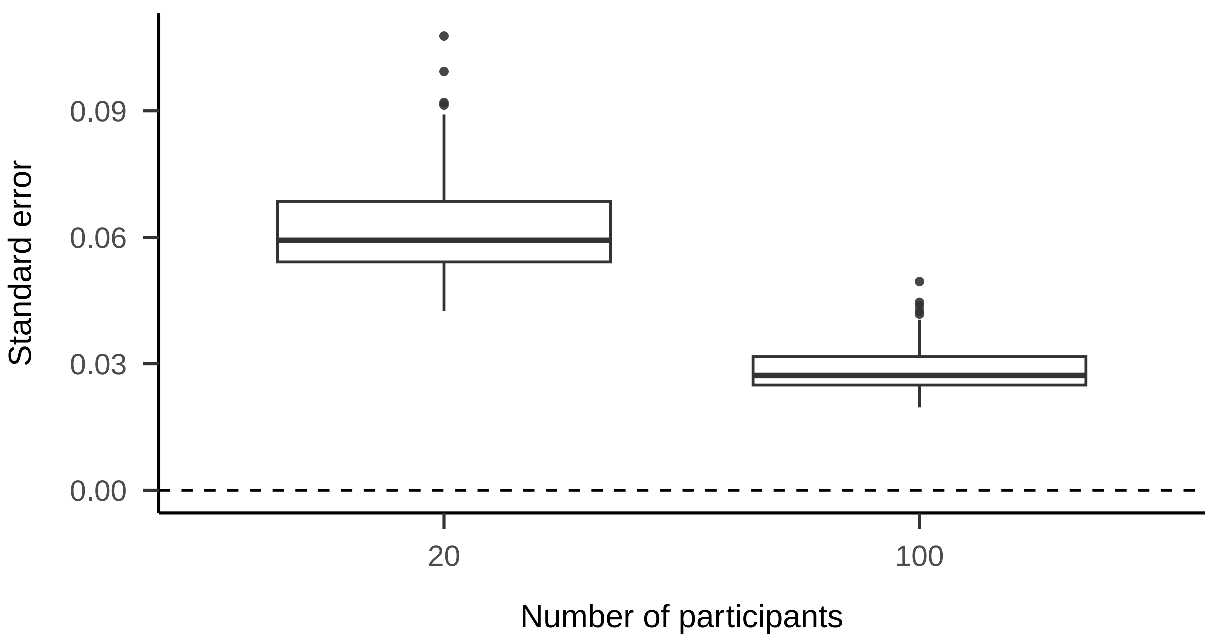

Figure 9: The effect of the number of participants on the standard error of estimation of the fixed slopes. Simulation B.

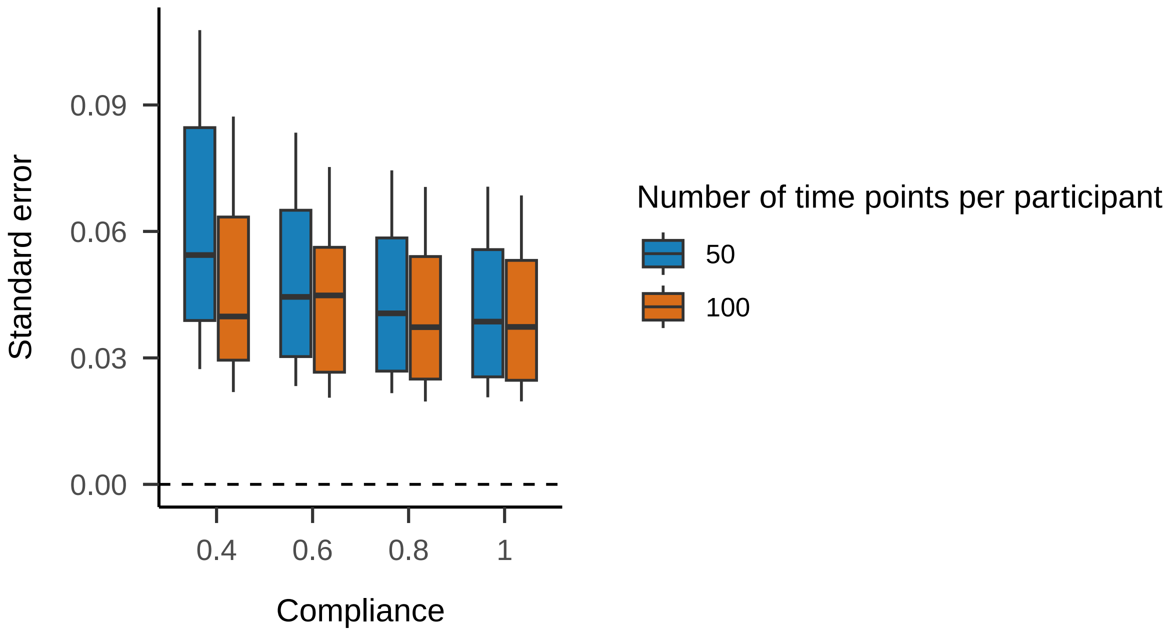

Figure 10: The effect of the interaction between number of time points per participant and compliance on standard error of estimation of the fixed slopes. Simulation B.

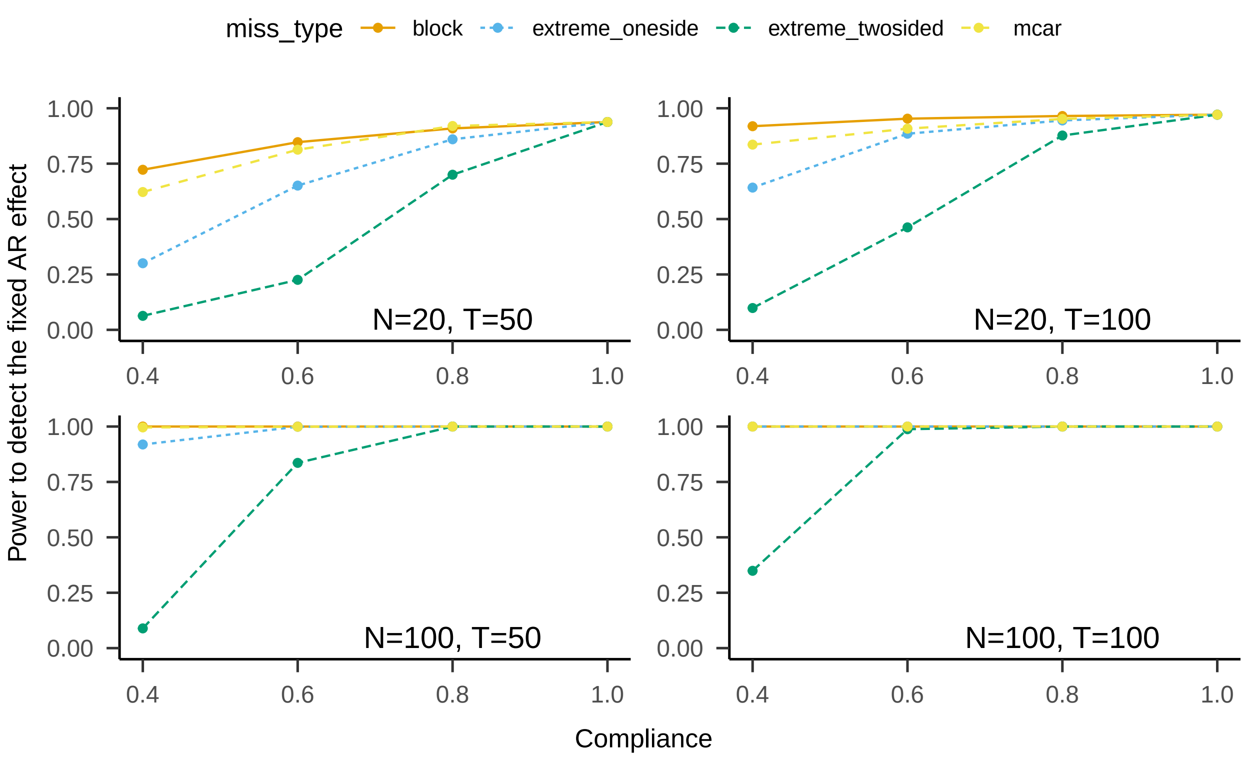

Figure 11: Simulation B. Statistical power to detect the fixed AR effect for the 4 combinations of compliance, missingness type, number of participants and time points per participant when the simulated fixed slope is 0.3 and the variance of random AR effects is 0.1.

Table 13: ANOVA results, simulation B. Outcome: Power to detect the fixed AR effect

	Df	Sum Sq	Mean Sq	F value	p-value	Partial omega-squared
N	1	0.41	0.41	42.23	<0.001	0.14
T.obs	1	0.09	0.09	9.86	0.0019	0.03
as.factor(sigma_v1)	1	0.00	0.00	0.00	0.9764	0.00
miss_type	3	1.69	0.56	58.45	<0.001	0.40
compliance	1	1.51	1.51	156.22	<0.001	0.38
B1_sim	1	1.03	1.03	106.64	<0.001	0.29
N:T.obs	1	0.06	0.06	6.58	0.0110	0.02
N:miss_type	3	0.12	0.04	4.18	0.0066	0.04
T.obs:miss_type	3	0.02	0.01	0.81	0.4912	0.00
N:compliance	1	0.20	0.20	21.09	<0.001	0.07
T.obs:compliance	1	0.05	0.05	5.47	0.0202	0.02
miss_type:compliance	3	1.78	0.59	61.49	<0.001	0.41
N:B1_sim	1	0.17	0.17	17.65	<0.001	0.06
T.obs:B1_sim	1	0.04	0.04	4.18	0.0421	0.01
miss_type:B1_sim	3	0.75	0.25	26.12	<0.001	0.23
compliance:B1_sim	1	0.75	0.75	77.43	<0.001	0.23
Residuals	229	2.21	0.01		NA	

with the strongest influence on the statistical power to detect the fixed AR effect. The value of the simulated AR effect has a large effect ($\omega_p^2 = 0.29$) as well. Three interactions crossed the effect size threshold: missingness type*compliance ($\omega_p^2 = 0.41$), missingness type*simulated fixed AR ($\omega_p^2 = 0.23$), and compliance*simulated fixed AR ($\omega_p^2 = 0.23$).

Supplementary simulation C

To investigate whether the results presented above hold when the lagged predictor is not person-mean centered, we conducted two supplementary simulation studies – Simulation C and D. All parameters were identical to Simulation A and B (respectively), except for the fact that the predictor was not person-mean centered in the supplementary simulations. We found that both the standard error and the bias in the estimation of the fixed autoregressive effect is considerably smaller when the predictor is not person-mean centered.

Estimation bias. In an ANOVA with the bias in the estimation of the fixed autoregressive effect as the outcome (Table 14), only the effect of the number of beeps per participant, missingness type, and the interaction between missingness type and compliance exceeded the effect size cut-off.

In contrast to the results of Simulation A, the effect of the manipulated factors on estimation bias is much smaller. Moreover, Figure 12 shows that while there is still slight underestimation when the most extreme values are missing, the magnitude of the underestimation is much smaller than in Simulation A. Additionally, when the observations are missing completely at random or in blocks, a slight overestimation of the fixed autoregressive effect occurs. The estimation becomes very precise as compliance gets higher

(Figure 13).

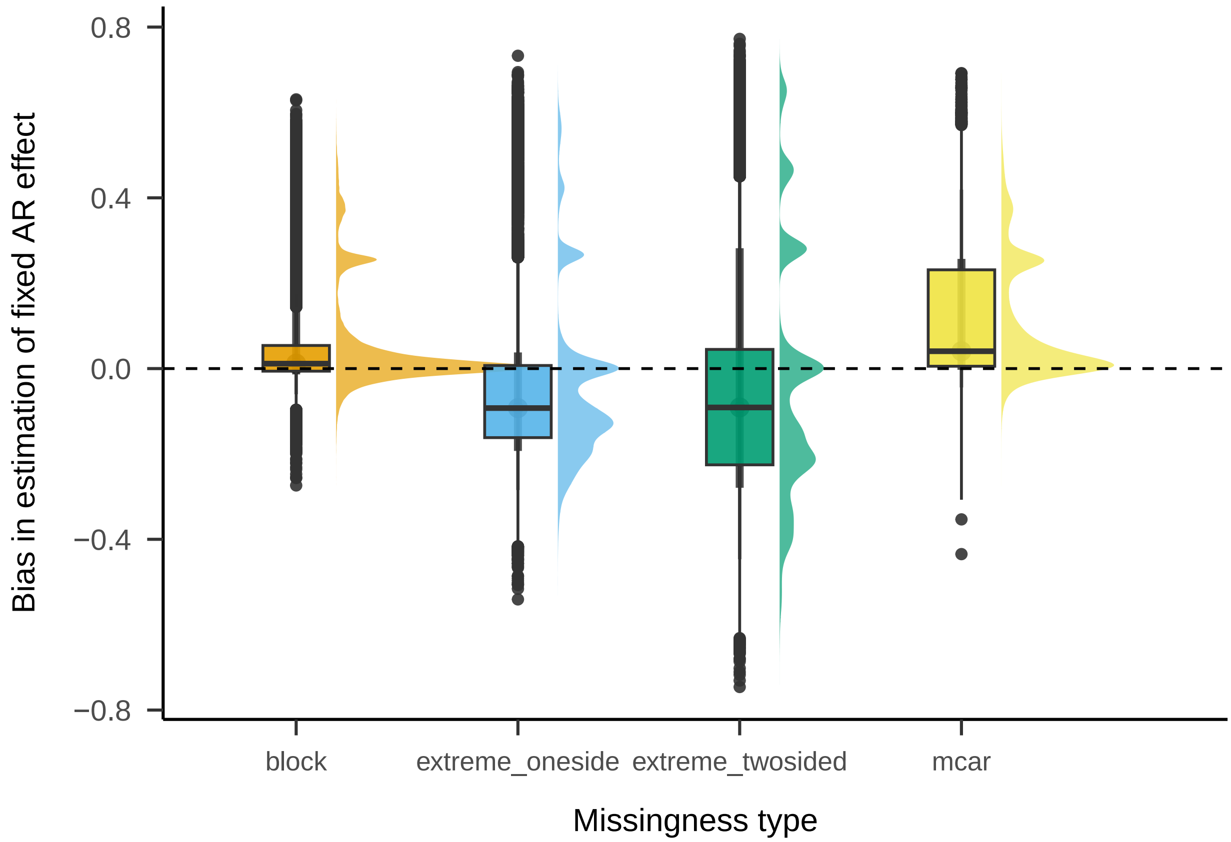

Figure 12: Supplementary Simulation C (no person-mean centering): The effect of missingnness type on estimation bias

Standard error. The results from ANOVA with Standard error of estimation as the outcome (when estimating the fixed autoregressive effect) are reported in Table 15. The main effects of the number of time points per participants and missingness type, as well as the interaction between compliance and the number of time points, crossed the effect size cut-off ($\omega_p^2 > 0.14$).

The standard errors in the estimation of fixed AR effect are considerably smaller when person-mean centering is not used (Table 16) compared to Simulation A (Table 5). Figure 14 illustrates the interaction between compliance and the number of time points.

Taken together, these results suggest that the results of MLAR(1) models obtained without person-mean centering the predictor are more robust to the presence of missing value with regards to estimation bias and standard error (i.e., the estimation bias and standard error are considerably lower when the predictor is not person-mean centered). Still, both the magnitude and the direction of the bias depend on the type of missingness.

Table 14: ANOVA results, Supplementary simulation C (no person-mean centering of the predictor). Outcome: Estimation bias

	Df	Sum Sq	Mean Sq	F value	p-value	Partial omega-squared
N	1	18.79	18.79	924.46	<0.001	0.00
T.obs	1	2122.67	2122.67	104427.39	<0.001	0.27
miss_type	3	1146.68	382.23	18804.07	<0.001	0.16
compliance	1	610.86	610.86	30052.04	<0.001	0.09
B1_sim	1	8.00	8.00	393.36	<0.001	0.00
N:T.obs	1	7.14	7.14	351.43	<0.001	0.00
N:miss_type	3	0.76	0.25	12.51	<0.001	0.00
T.obs:miss_type	3	319.91	106.64	5246.11	<0.001	0.05
N:compliance	1	12.73	12.73	626.34	<0.001	0.00
T.obs:compliance	1	1264.44	1264.44	62205.42	<0.001	0.18
miss_type:compliance	3	259.36	86.45	4253.09	<0.001	0.04
N:B1_sim	1	0.62	0.62	30.46	<0.001	0.00
T.obs:B1_sim	1	40.76	40.76	2005.07	<0.001	0.01
miss_type:B1_sim	3	41.20	13.73	675.69	<0.001	0.01
compliance:B1_sim	1	12.11	12.11	595.87	<0.001	0.00
Residuals	286923	5832.23	0.02		NA	

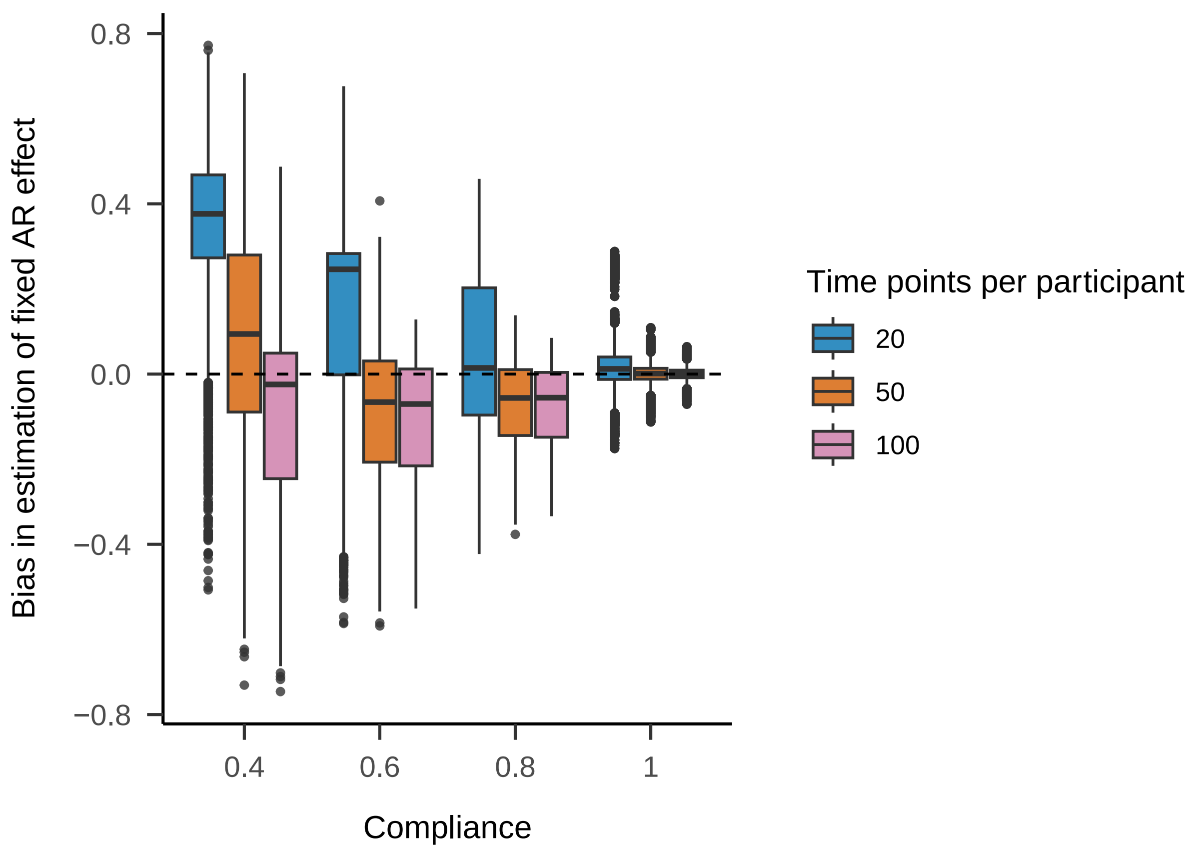

Figure 13: Supplementary Simulation C (no person-mean centering): The effect of the interaction between compliance and number of time points on estimation bias.

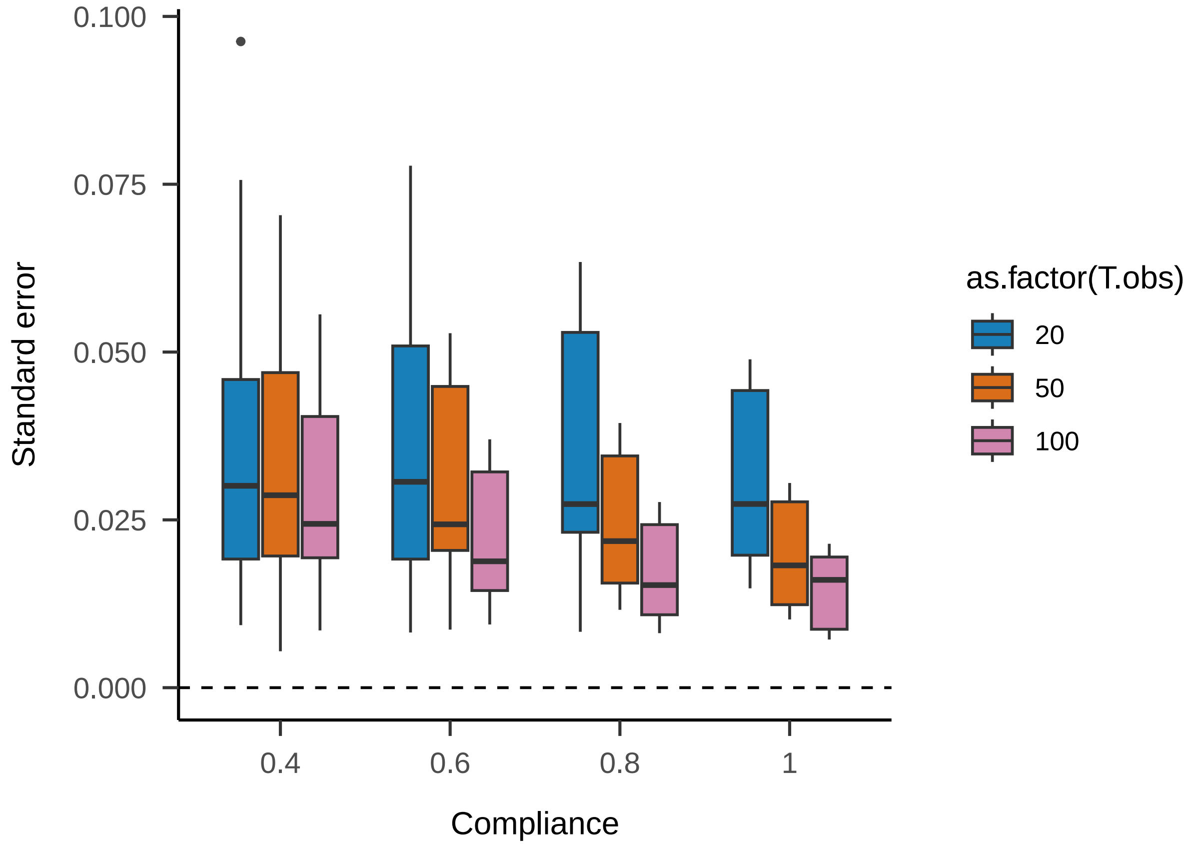

Figure 14:

Table 15: ANOVA results, supplementary Simulation C Outcome: Standard error

	Df	Sum Sq	Mean Sq	F value	p-value	Partial omega-squared
N	1	0.03	0.03	1032.24	<0.001	0.78
T.obs	1	0.01	0.01	267.88	<0.001	0.48
miss_type	3	0.00	0.00	6.19	<0.001	0.05
comp_mean	1	0.01	0.01	170.27	<0.001	0.37
B1_sim	1	0.01	0.01	311.30	<0.001	0.52
N:T.obs	1	0.00	0.00	46.53	<0.001	0.14
N:miss_type	3	0.00	0.00	0.88	0.452	0.00
T.obs:miss_type	3	0.00	0.00	3.75	0.012	0.03
N:comp_mean	1	0.00	0.00	33.31	<0.001	0.10
T.obs:comp_mean	1	0.00	0.00	18.65	<0.001	0.06
miss_type:comp_mean	3	0.00	0.00	6.14	<0.001	0.05
N:B1_sim	1	0.00	0.00	38.90	<0.001	0.12
T.obs:B1_sim	1	0.00	0.00	78.37	<0.001	0.21
miss_type:B1_sim	3	0.00	0.00	1.89	0.131	0.01
comp_mean:B1_sim	1	0.00	0.00	51.01	<0.001	0.15
Residuals	261	0.01	0.00		NA	

Table 16: Supplementary Simulation C, no person-mean centering. Average standard error in the estimation of the fixed slope for each combination of number of participants, number of time points/participant, and compliance.

		Compliance			
N	T.obs	0.4	0.6	0.8	1
20	20	0.05	0.05	0.05	0.04
	50	0.05	0.04	0.03	0.03
	100	0.04	0.03	0.02	0.02
100	20	0.02	0.02	0.02	0.02
	50	0.02	0.02	0.02	0.01
	100	0.02	0.01	0.01	0.01

Supplementary simulation D

In Simulation D, all simulation parameters were identical to Simulation B, except for the absence of person-mean centering of the lagged predictor.

Estimation bias. Interestingly, none of the main effects nor the two-way interactions reached the effect size threshold in supplementary Simulation D (see Table 17). For illustratory purposes, the average estimation bias for each combination of compliance and missingness type is included in Table 18 and visualised in Figure 15.

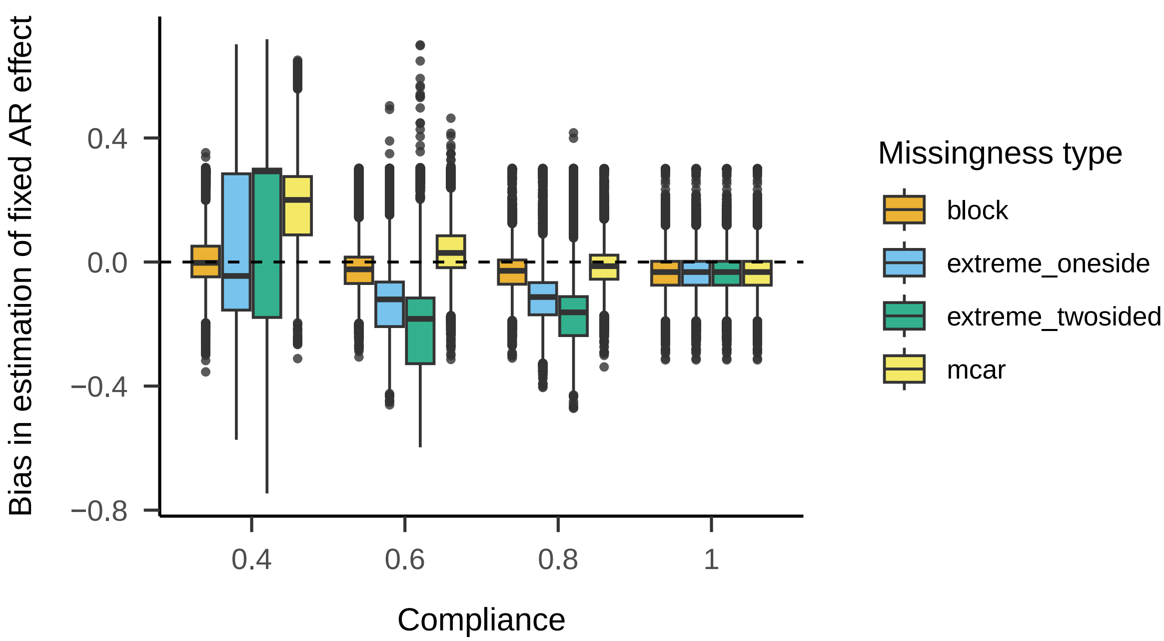

Figure 15: Supplementary Simulation D: The effect of the interaction between missingness type and compliance on the bias in estimation of the fixed slopes.

Standard error The results of the ANOVA with standard error as outcome are included in Table 19.

Table 17: ANOVA results, supplementary simulation D (no person-mean centering of the predictor). Outcome: Estimation bias

	Df	Sum Sq	Mean Sq	F value	p-value	Partial omega-squared
N	1	41.28	41.28	2179.86	<0.001	0.01
T.obs	1	293.13	293.13	15478.43	<0.001	0.06
miss_type	3	462.06	154.02	8132.86	<0.001	0.09
compliance	1	436.78	436.78	23063.43	<0.001	0.08
sigma_v1	1	19.46	19.46	1027.66	<0.001	0.00
B1_sim	1	4.74	4.74	250.22	<0.001	0.00
N:T.obs	1	4.12	4.12	217.68	<0.001	0.00
N:miss_type	3	20.00	6.67	351.99	<0.001	0.00
T.obs:miss_type	3	46.54	15.51	819.13	<0.001	0.01
N:compliance	1	38.37	38.37	2025.91	<0.001	0.01
T.obs:compliance	1	230.72	230.72	12182.88	<0.001	0.05
miss_type:compliance	3	171.48	57.16	3018.20	<0.001	0.03
N:B1_sim	1	12.63	12.63	667.07	<0.001	0.00
T.obs:B1_sim	1	36.82	36.82	1944.17	<0.001	0.01
miss_type:B1_sim	3	21.88	7.29	385.15	<0.001	0.00
compliance:B1_sim	1	214.38	214.38	11319.85	<0.001	0.04
Residuals	255025	4829.66	0.02		NA	

Table 18: Simulation D. Average bias in estimation of the fixed slope for each combination of missingness type and compliance.

	Missingness type			
compliance	block	extreme_oneside	extreme_twosided	mcar
0.4	0.02	0.00	0.12	0.18
0.6	-0.02	-0.11	-0.15	0.05
0.8	-0.03	-0.11	-0.16	-0.01
1.0	-0.04	-0.04	-0.04	-0.04

Table 19: ANOVA results, supplementary Simulation D. Outcome: Standard error

	Df	Sum Sq	Mean Sq	F value	p-value	omegasq_se_simD
N	1	0.06	0.06	1769.27	<0.001	0.87
T.obs	1	0.00	0.00	1.35	0.2472	0.00
miss_type	3	0.00	0.00	12.73	<0.001	0.12
sigma_v1	1	0.00	0.00	95.73	<0.001	0.27
compliance	1	0.00	0.00	8.11	0.0048	0.03
B1_sim	1	0.01	0.01	348.67	<0.001	0.58
N:T.obs	1	0.00	0.00	1.71	0.1923	0.00
N:miss_type	3	0.00	0.00	0.79	0.5000	0.00
T.obs:miss_type	3	0.00	0.00	0.05	0.9848	0.00
N:compliance	1	0.00	0.00	0.89	0.3452	0.00
T.obs:compliance	1	0.00	0.00	4.29	0.0394	0.01
miss_type:compliance	3	0.00	0.00	11.53	<0.001	0.11
N:B1_sim	1	0.00	0.00	26.84	<0.001	0.09
T.obs:B1_sim	1	0.00	0.00	40.26	<0.001	0.13
miss_type:B1_sim	3	0.00	0.00	7.93	<0.001	0.08
compliance:B1_sim	1	0.00	0.00	118.19	<0.001	0.31
Residuals	229	0.01	0.00		NA	

The main effects of the number of participants ($\omega_p^2 = 0.87$), simulated fixed autoregressive effect ($\omega_p^2 = 0.58$), and the variance of the random effects ($\omega_p^2 = 0.27$), as well as the interaction between compliance and simulated fixed autoregressive effect ($\omega_p^2 = 0.31$) crossed the effect size threshold.

Discussion

We conducted two Monte Carlo simulation studies (and two supplementary simulations) to address a gap in knowledge about the influence of missing data on the estimation performance of the first-order multilevel autoregressive model. In Simulation A, we only estimated and simulated fixed autoregressive effects (together with both fixed and random intercepts), while both fixed and random autoregressive effects were simulated and estimated in Simulation B. Three main outcomes were evaluated in both simulations: the estimation bias, standard error of the simulations, and statistical power for the estimates of fixed AR effects. Four values of compliance and four different missingness patterns (data MCAR, data missing in a block of consecutive observations, all values below a given percentile missing, and the most extreme/highest and lowest values missing) were varied across the simulations. The other manipulated factors included the number of participants, the number of time points per participant, the simulated value of the fixed AR effect, and the variance of the random AR effects.

The two simulation parameters related to missing data (compliance and missingness pattern) emerged as very important factors influencing all three outcomes. In both simulations, missingness type and compliance (and the interaction between the two) were the factors with the largest effect on the bias in the estimation of the fixed AR effect. Similarly, both missingness type and compliance had a strong influence on the statistical power to detect the fixed AR effect in both simulations. With regards to the standard error of the simulation results, compliance was found to have a very large effect (more so in Simulation A than in Simulation B), while the effect of missingness type was only moderate.

Our results corroborate the conclusions about the importance of the number of time points per participant for a precise estimation of the autoregressive effects (Hamaker & Grasman, 2015; Krone et al., 2016). In general, the estimation bias became considerably less severe as the ESM time-series length per participant increased. The number of observations per participant also had a large effect on statistical power. However, our simulations show that the context of missingness matters: when the compliance is low and the data are missing MCAR or in blocks, the underestimation of the fixed AR effect caused by the missing data (and the negative consequences for statistical power) becomes less severe very quickly as the number of observations per participant increases. On the other hand, when the missingness is dependent on the value of the process itself (i.e., the most extreme observations are missing), increasing compliance appears to be more important for estimation precision and statistical power than the length of the time-series. In other words, the presence of missing data exacerbates Nickell's bias (Nickell, 1981) – an estimation bias introduced by person-mean centering in multilevel models.

In Simulation A, the average estimation bias when compliance was 0.8, which roughly corresponds to the average compliance of ESM studies in psychology (Wrzus & Neubauer, 2022), was -0.13. This suggests that some of the estimates of emotional inertia in psychological research could be considerably downward biased. Furthermore, the estimates are slightly biased even when compliance is 1 (i.e., there are no missing data; average bias: 0.14). This is in line with the findings about the downward estimation bias caused by person-mean centering in multilevel autoregressive models (Hamaker & Grasman, 2015).

While Krone et al. (2016) found that estimation bias becomes smaller as the simulated fixed AR effect increases, we found an opposite pattern: overall, estimation bias increased with increasing simulated fixed AR effect. However, this only held true for the two missingness patterns in which the most extreme observations were set as missing. In conditions with data missing MCAR or in block, the estimation bias stayed almost identical regardless of the magnitude of the simulated fixed AR effect (in Simulation A) or increased slightly (Simulation B). The discrepancy between our results and those by Krone et al. (2016) can be explained by the fact that Krone et al. did not focus on missing data in their simulations.

Overall, there was always some degree of estimation bias present in the simulations, ranging from very severe (when the number of time points per participant and compliance were low, and the missingness of data was dependent on the process value) to mild (when compliance was high and the data were missing either MCAR or in block). The downward bias was especially pronounced in the simulations that used person-mean centering for the lagged predictor. The estimation bias resulting from the presence of missing data might be one of the driving forces behind the low value added by estimates of emotional inertia to the prediction of psychopathology and well–being, pointed out by Dejonckheere et al. (2019). Additionally, while our simulation studies did not explicitly assess the bias in the estimation of individual autoregressive effects, the results suggest that some individual differences in inertia estimates might not be caused by real differences in inertia, but due to the bias caused by missingness: for two individual participants with an identical real autoregressive parameter but different compliance and missingness patterns, the inertia estimates can vary considerably.

The supplementary simulations show that the estimation bias is considerably less severe (and the standard errors are smaller) when the lagged predictor is not person-mean centered. These results suggest that although using person-mean centering is usually recommended in the literature, it might not be the optimal choice when the number of observations and/or the compliance is low. However, as Hamaker & Grasman (2015) point out, the choice between centering and not centering the predictor should primarily be guided by the researchers' goals. If the researchers aim to obtain interpretable intercepts or to investigate how a Level 2 predictor influences the autoregressive effect, person-mean centering might still be preferable. On the other hand, the results of the present thesis support Hamaker and Grasman's (2015) claim that if the focal point of interest is the fixed autoregressive parameter, it is better to avoid person-mean centering the predictor.

Our results have several implications for the design choices in psychological studies that use the multilevel autoregressive model to estimate emotional inertia. However, it is important to note that these implications only hold for the simple MLAR model used in the simulations – in other designs, such as those that include a Level 2 predictor, the recommendations might differ. First, in line with previous simulation studies (Krone et al., 2016), we recommend for researchers to focus on increasing the number of time points per participant, rather than the number of participants, in order to increase the statistical power and the precision of the inertia estimates in situations where some amount of missing data can be expected. In other words, for optimal statistical performance, it is more effective to make the data collection period longer (or schedule more beeps per day) than to collect data from more participants. Secondly, while the time-series length is important, researchers should aim to design their ESM studies in a way that will make compliance as high as possible. According to recent evidence about compliance in ESM studies, these design choices include providing financial incentives to participants (Wrzus & Neubauer, 2022) and including less items in the ESM questionnaires (Eisele et al., 2020). Furthermore, the results suggest that the potential presence of missing data should be accounted for in power analyses for ESM studies. In an ideal case, a researcher should have an idea about the potential average compliance in their study and the missingness patterns that can be expected in the data. Of course, this is not entirely feasible, as it might be difficult to estimate the average compliance, and real-life ESM data will likely include a mixture of different missing data patterns, both at the within- and between-person level. Still, to avoid overestimating statistical power for planned studies, it is advisable to include several different missing data scenarios in the power simulations as a sensitivity check.

Directions for future research

The present thesis provides evidence about estimation bias being made more severe by lower compliance and by patterns of missingness that depend on the process value itself. Nonetheless, the insight into the mechanisms driving this bias remains limited. Several plausible explanations of the bias arise. First, the estimation bias was the most severe when the most extreme observations at both ends of the ESM process distribution were missing. This finding can be linked to the evidence about tails (i.e., the extreme ends) of distributions containing the most information about the scale of the distribution (Zheng & Gastwirth, 2002). However, further investigation is needed to provide a more detailed understanding about whether the tails also provide crucial information about the autoregressive effect.

Secondly, the differences in bias for different missingness patterns could be partly caused by the fact that the number of "effective" observation-pairs (i.e., pairs of current and lagged values where both values are not missing) used to estimate the autoregressive parameters differs between the missingnness patterns. For example, we have demonstrated that the number of effective observation-pairs is smaller when the data are MCAR compared to when they are missing in a block. However, the results of the present thesis seem to contradict this explanation. On the one hand, estimation bias was found to be very similar when data are MCAR and missing in blocks (although the two missingness patterns come with a very different number of observation-pairs). On the other hand, the estimation bias was much larger for the conditions with the most extreme values set as missing, compared to the condition with data MCAR (even though the number of effective observation-pairs is generally higher for the "extreme" conditions than when data are MCAR). As such, this explanation of estimation bias appears to be less plausible than the explanation described in the previous paragraph.

Limitations

First, while our simulation studies include a wide range of scenarios and parameter combinations, our results are far from comprehensive, and they depend on the specific simulation parameters we used. However, the reproducible code available from the GitHub repository (https://github.com/benjsimsa/AR-missing-simulations) provides a sufficient framework to rerun the analyses with different parameters and modify the code to better fit the peculiarities of their specific study sample and research questions.

Furthermore, the results depend on several assumptions, which were problematised as being too simplistic in previous research. We assumed that the innovation variance σ_ϵ was identical for all the participants, and that the random intercepts (and random slopes in Simulation B) came from a normal distribution. Additionally, we only focused on normally distributed affective processes. While a normal distribution can be assumed for ESM measures of positive emotions, negative affective processes are usually heavily right-skewed in the general population (Haslbeck et al., 2022). Additionally, we assumed that the analysed ESM time-series are measured without any error; however, recent evidence shows that this is very often not the case in real-world research (Dejonckheere et al., 2022; Schuurman & Hamaker, 2019), and unreliability can lead to further attenuation of the estimated parameters (Wenzel & Brose, 2022). While person-mean centering was carried out using observed means in the simulations we conducted, different ways of person-mean centering, such as using latent person-means, might be more appropriate (Gistelinck et al., 2021). Another assumption of the present simulations that is unlikely to hold in real-world data is the homogeneity of compliances and missing data patterns within each simulated dataset. In the real world, it can be expected for different missingness patterns to be present in the data at both the between- and within-person level. However, the violation of these assumptions can be simulated through modifications to the simulation code we provided.

Finally, although we took steps to make the simulations reproducible by making all code and results publicly available, using R packages *here* and *renv*, and reporting the sessionInfo for every simulation, a large number of packages with many dependencies were used, which might be detrimental to reproducibility in the long term.

Conclusion

To sum up, our results suggest that the presence of missing observations in ESM datasets can introduce bias in the results of multilevel autoregressive models and jeopardize the conclusions about inertia (and individual differences therein) drawn from them. The magnitude of estimation bias is higher when the compliance is low, when the missingness is dependent on process value, and when the number of observations per participant is low. Furthermore, the estimation bias is more pronounced when the researchers person-mean the predictor in the MLAR model. To minimize the estimation bias, we recommend the researchers to design their ESM data collection procedures in ways that promote high compliance, and, if possible, use designs with large numbers of observations per participant.

References

Allaire, J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., Wickham, H., Cheng, J., Chang, W., & Iannone, R. (2022). *Rmarkdown: Dynamic documents for r*. https://github.com/rstudio/rmarkdown

Bos, E. H., Jonge, P., & Cox, R. F. A. (2019). Affective variability in depression: Revisiting the inertia–instability paradox. *British Journal of Psychology*, *110*(4), 814–827. https://doi.org/10.1111/bjop.12372

Bosley, H. G., Soyster, P. D., & Fisher, A. J. (2019). Affect Dynamics as Predictors of Symptom Severity and Treatment Response in Mood and Anxiety Disorders: Evidence for Specificity. *Journal for Person-Oriented Research*, *5*(2), 101–113. https://doi.org/10.17505/jpor.2019.09

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

Dejonckheere, E. (2019). Complex affect dynamics add limited information to the prediction of psychological well-being. *Nature Human Behaviour*, *3*, 17.

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

Eisele, G., Vachon, H., Lafit, G., Kuppens, P., Houben, M., Myin-Germeys, I., & Viechtbauer, W. (2020). 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)*, 16.

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, *12*(2), 121–138. https://doi.org/10.1037/1082-989X.12.2.121

Field, A., Miles, J., & Field, Z. (2012). Discovering statistics using r (2012). *Great Britain: Sage Publications, Ltd*, *958*.

Gistelinck, F., Loeys, T., & Flamant, N. (2021). Multilevel Autoregressive Models when the Number of Time Points is Small. *Structural Equation Modeling: A Multidisciplinary Journal*, *28*(1), 15–27. https://doi.org/10.1080/10705511.2020.1753517

Hamaker, E. L. (2012). *Why researchers should think" within-person": A paradigmatic rationale.*

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*. https://doi.org/10.3389/fpsyg.2014.01492

Hamaker, E. L., & Muthén, B. (2020). The fixed versus random effects debate and how it relates to centering in multilevel modeling. *Psychological Methods*, *25*(3), 365–379. https://doi.org/10.1037/met0000239

Haslbeck, J. M. B., Ryan, O., & Dablander, F. (2022). *Multimodality and Skewness in Emotion Time Series*. https://doi.org/10.31234/osf.io/qudr6

Houben, M., & Kuppens, P. (2020). Emotion Dynamics and the Association With Depressive Features and Borderline Personality Disorder Traits: Unique, Specific, and Prospective Relationships. *Clinical Psychological Science*, *8*(2), 226–239. https://doi.org/10.1177/2167702619871962

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

Ji, L., Chow, S.-M., Schermerhorn, A. C., Jacobson, N. C., & Cummings, E. M. (2018). Handling Missing Data in the Modeling of Intensive Longitudinal Data. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(5), 715–736. https://doi.org/10.1080/10705511.2017.1417046

Jongerling, J., Laurenceau, J.-P., & Hamaker, E. L. (2015). A Multilevel AR(1) Model: Allowing for Inter-Individual Differences in Trait-Scores, Inertia, and Innovation Variance. *Multivariate Behavioral Research*, *50*(3), 334–349. https://doi.org/10.1080/00273171.2014.1003772

Koval, P., Burnett, P. T., & Zheng, Y. (2021). *Emotional Inertia: On the Conservation of Emotional Momentum* (C. E. Waugh & P. Kuppens, Eds.; pp. 63–94). Springer International Publishing. https://doi.org/10.1007/978-3-030-82965-0_4

Koval, P., Butler, E. A., Hollenstein, T., Lanteigne, D., & Kuppens, P. (2015). Emotion regulation and the temporal dynamics of emotions: Effects of cognitive reappraisal and expressive suppression on emotional inertia. *Cognition and Emotion*, *29*(5), 831–851. https://doi.org/10.1080/02699931.2014.948388

Koval, P., Pe, M. L., Meers, K., & Kuppens, P. (2013). Affect dynamics in relation to depressive symptoms: Variable, unstable or inert? *Emotion*, *13*(6), 1132–1141. https://doi.org/10.1037/a0033579

Krone, T., Albers, C. J., & Timmerman, M. E. (2016). Comparison of Estimation Procedures for Multilevel AR(1) Models. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00486

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

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., Adolf, J., Dejonckheere, E., Myin-Germeys, I., Viechtbauer, W., & Ceulemans, E. (2020). *Selection of the Number of Participants in Intensive Longitudinal Studies: A User-friendly Shiny App and Tutorial to Perform Power Analysis in Multilevel Regression Models that Account for Temporal Dependencies*. https://doi.org/10.31234/osf.io/dq6ky

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

Larson, R., & Csikszentmihalyi, M. (2014). *The experience sampling method* (p. 2134). Springer.

Leemput, I. A. van de, Wichers, M., Cramer, A. O. J., Borsboom, D., Tuerlinckx, F., Kuppens, P., Nes, E. H. van, Viechtbauer, W., Giltay, E. J., Aggen, S. H., Derom, C., Jacobs, N., Kendler, K. S., Maas, H. L. J. van der, Neale, M. C., Peeters, F., Thiery, E., Zachar, P., & Scheffer, M. (2014). Critical slowing down as early warning for the onset and termination of depression. *Proceedings of the National Academy of Sciences*, *111*(1), 87–92. https://doi.org/10.1073/pnas.1312114110

Liu, S. (2017). Person-specific versus multilevel autoregressive models: Accuracy in parameter estimates at the population and individual levels. *British Journal of Mathematical and Statistical Psychology*, *70*(3), 480–498. https://doi.org/10.1111/bmsp.12096

Mneimne, M., Fleeson, W., Arnold, E. M., & Furr, R. M. (2018). Differentiating the everyday emotion dynamics of borderline personality disorder from major depressive disorder and bipolar disorder. *Personality Disorders: Theory, Research, and Treatment*, *9*(2), 192–196. https://doi.org/10.1037/per0000255

Müller, K. (2020). *Here: A simpler way to find your files.* https://CRAN.R-project.org/package=here

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

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica*, *49*(6), 1417. https://doi.org/10.2307/1911408

Okada, K. (2013). Is omega squared less biased? A comparison of three major effect size indices in one-way ANOVA. *Behaviormetrika*, *40*(2), 129147.

Pawel, S., Kook, L., & Reeve, K. (2022). *Pitfalls and potentials in simulation studies.* http://arxiv.org/abs/2203.13076

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., Heisterkamp, S., Van Willigen, B., Ranke, J., Team, R. C., & Team, M. R. C. (2022). *Package 'nlme'.*

R Core Team. (2021). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. https://www.R-project.org/

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

Rintala, A., Wampers, M., Myin-Germeys, I., & Viechtbauer, W. (2020). Momentary predictors of compliance in studies using the experience sampling method. *Psychiatry Research*, *286*, 112896. https://doi.org/10.1016/j.psychres.2020.112896

Rockel, T. (2022). *missMethods: Methods for missing data.* https://CRAN.R-project.org/package=missMethods

Rovine, M. J., & Walls, T. A. (2006). Multilevel autoregressive modeling of interindividual differences in the stability of a process. *Models for Intensive Longitudinal Data*, 124147.

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

Ushey, K. (2022). *Renv: Project environments.* https://CRAN.R-project.org/package=renv

Vachon, H., Viechtbauer, W., Rintala, A., & Myin-Germeys, I. (2019). Compliance and Retention With the Experience Sampling Method Over the Continuum of Severe Mental Disorders: Meta-Analysis and Recommendations. *Journal of Medical Internet Research*, *21*(12), e14475. https://doi.org/10.2196/14475

Wenzel, M., & Brose, A. (2022). Addressing measurement issues in affect dynamic research: Modeling emotional inertia's reliability to improve its predictive validity of depressive symptoms. *Emotion*. https://doi.org/10.1037/emo0001108

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

Wrzus, C., & Neubauer, A. B. (2022). Ecological Momentary Assessment: A Meta-Analysis on Designs, Samples, and Compliance Across Research Fields. *Assessment*, 1–22. https://doi.org/10.1177/10731911211067538

Zheng, G., & Gastwirth, J. (2002). Do tails of symmetric distributions contain more fisher information about the scale parameter? *Sankhyā: The Indian Journal of Statistics, Series B*, 289300.

Appendix 1: Full results from Simulation A

Table 20: Simulation A. Full results.

Missingness pattern	Simulated fixed AR	Compliance	N participants	Beeps per participant	Estimated fixed AR	Power to detect fixed AR	Fixed AR SE	Fixed AR estimation bias
block	0.3	0.4	20	20	0.129	0.242	0.110	-0.171
block	0.3	0.4	20	50	0.226	0.982	0.054	-0.074
block	0.3	0.4	20	100	0.267	1.000	0.036	-0.033
block	0.3	0.4	100	20	0.130	0.743	0.049	-0.170
block	0.3	0.4	100	50	0.228	1.000	0.024	-0.072
block	0.3	0.4	100	100	0.266	1.000	0.016	-0.034
block	0.3	0.6	20	20	0.175	0.602	0.078	-0.125
block	0.3	0.6	20	50	0.253	1.000	0.042	-0.047
block	0.3	0.6	20	100	0.278	1.000	0.029	-0.022
block	0.3	0.6	100	20	0.179	0.999	0.035	-0.121
block	0.3	0.6	100	50	0.253	1.000	0.019	-0.047
block	0.3	0.6	100	100	0.278	1.000	0.013	-0.022
block	0.3	0.8	20	20	0.209	0.899	0.063	-0.091
block	0.3	0.8	20	50	0.264	1.000	0.036	-0.036
block	0.3	0.8	20	100	0.284	1.000	0.025	-0.016
block	0.3	0.8	100	20	0.210	1.000	0.028	-0.090
block	0.3	0.8	100	50	0.266	1.000	0.016	-0.034
block	0.3	0.8	100	100	0.283	1.000	0.011	-0.017
block	0.3	1.0	20	20	0.230	0.992	0.051	-0.070
block	0.3	1.0	20	50	0.273	1.000	0.031	-0.027
block	0.3	1.0	20	100	0.288	1.000	0.022	-0.012
block	0.3	1.0	100	20	0.230	1.000	0.023	-0.070
block	0.3	1.0	100	50	0.273	1.000	0.014	-0.027
block	0.3	1.0	100	100	0.287	1.000	0.010	-0.013
block	0.5	0.4	20	20	0.334	0.840	0.105	-0.166
block	0.5	0.4	20	50	0.415	1.000	0.051	-0.085
block	0.5	0.4	20	100	0.461	1.000	0.033	-0.039
block	0.5	0.4	100	20	0.339	1.000	0.047	-0.161
block	0.5	0.4	100	50	0.418	1.000	0.023	-0.082

Table 20: Simulation A. Full results. *(continued)*

Missingness pattern	Simulated fixed AR	Compliance	N participants	Beeps per participant	Estimated fixed AR	Power to detect fixed AR	Fixed AR SE	Fixed AR estimation bias
block	0.5	0.4	100	100	0.461	1.000	0.015	-0.039
block	0.5	0.6	20	20	0.365	0.997	0.074	-0.135
block	0.5	0.6	20	50	0.445	1.000	0.039	-0.055
block	0.5	0.6	20	100	0.474	1.000	0.026	-0.026
block	0.5	0.6	100	20	0.370	1.000	0.033	-0.130
block	0.5	0.6	100	50	0.446	1.000	0.018	-0.054
block	0.5	0.6	100	100	0.474	1.000	0.012	-0.026
block	0.5	0.8	20	20	0.396	1.000	0.059	-0.104
block	0.5	0.8	20	50	0.459	1.000	0.033	-0.041
block	0.5	0.8	20	100	0.481	1.000	0.022	-0.019
block	0.5	0.8	100	20	0.399	1.000	0.026	-0.101
block	0.5	0.8	100	50	0.460	1.000	0.015	-0.040
block	0.5	0.8	100	100	0.481	1.000	0.010	-0.019
block	0.5	1.0	20	20	0.416	1.000	0.048	-0.084
block	0.5	1.0	20	50	0.468	1.000	0.029	-0.032
block	0.5	1.0	20	100	0.485	1.000	0.020	-0.015
block	0.5	1.0	100	20	0.416	1.000	0.021	-0.084
block	0.5	1.0	100	50	0.469	1.000	0.013	-0.031
block	0.5	1.0	100	100	0.485	1.000	0.009	-0.015
block	0.7	0.4	20	20	0.565	0.997	0.092	-0.135
block	0.7	0.4	20	50	0.606	1.000	0.044	-0.094
block	0.7	0.4	20	100	0.654	1.000	0.028	-0.046
block	0.7	0.4	100	20	0.570	1.000	0.041	-0.130
block	0.7	0.4	100	50	0.610	1.000	0.020	-0.090
block	0.7	0.4	100	100	0.655	1.000	0.013	-0.045
block	0.7	0.6	20	20	0.569	1.000	0.065	-0.131
block	0.7	0.6	20	50	0.638	1.000	0.034	-0.062
block	0.7	0.6	20	100	0.670	1.000	0.022	-0.030
block	0.7	0.6	100	20	0.574	1.000	0.029	-0.126

Table 20: Simulation A. Full results. *(continued)*

Missingness pattern	Simulated fixed AR	Compliance	N participants	Beeps per participant	Estimated fixed AR	Power to detect fixed AR	Fixed AR SE	Fixed AR estimation bias
block	0.7	0.6	100	50	0.639	1.000	0.015	-0.061
block	0.7	0.6	100	100	0.670	1.000	0.010	-0.030
block	0.7	0.8	20	20	0.585	1.000	0.052	-0.115
block	0.7	0.8	20	50	0.653	1.000	0.028	-0.047
block	0.7	0.8	20	100	0.678	1.000	0.019	-0.022
block	0.7	0.8	100	20	0.588	1.000	0.023	-0.112
block	0.7	0.8	100	50	0.655	1.000	0.013	-0.045
block	0.7	0.8	100	100	0.678	1.000	0.008	-0.022
block	0.7	1.0	20	20	0.599	1.000	0.042	-0.101
block	0.7	1.0	20	50	0.663	1.000	0.024	-0.037
block	0.7	1.0	20	100	0.682	1.000	0.017	-0.018
block	0.7	1.0	100	20	0.600	1.000	0.019	-0.100
block	0.7	1.0	100	50	0.664	1.000	0.011	-0.036
block	0.7	1.0	100	100	0.682	1.000	0.007	-0.018
extreme_oneside	0.3	0.4	20	20	-0.053	0.056	0.139	-0.353
extreme_oneside	0.3	0.4	20	50	0.053	0.106	0.076	-0.247
extreme_oneside	0.3	0.4	20	100	0.086	0.388	0.051	-0.214
extreme_oneside	0.3	0.4	100	20	-0.061	0.168	0.062	-0.361
extreme_oneside	0.3	0.4	100	50	0.055	0.383	0.034	-0.245
extreme_oneside	0.3	0.4	100	100	0.087	0.964	0.023	-0.213
extreme_oneside	0.3	0.6	20	20	0.043	0.076	0.088	-0.257
extreme_oneside	0.3	0.6	20	50	0.111	0.557	0.052	-0.189
extreme_oneside	0.3	0.6	20	100	0.131	0.943	0.036	-0.169
extreme_oneside	0.3	0.6	100	20	0.042	0.191	0.040	-0.258
extreme_oneside	0.3	0.6	100	50	0.113	0.997	0.023	-0.187
extreme_oneside	0.3	0.6	100	100	0.133	1.000	0.016	-0.167
extreme_oneside	0.3	0.8	20	20	0.119	0.445	0.066	-0.181
extreme_oneside	0.3	0.8	20	50	0.166	0.983	0.039	-0.134
extreme_oneside	0.3	0.8	20	100	0.181	1.000	0.027	-0.119

Table 20: Simulation A. Full results. *(continued)*

Missingness pattern	Simulated fixed AR	Compliance	N participants	Beeps per participant	Estimated fixed AR	Power to detect fixed AR	Fixed AR SE	Fixed AR estimation bias
extreme_oneside	0.3	0.8	100	20	0.119	0.966	0.029	-0.181
extreme_oneside	0.3	0.8	100	50	0.167	1.000	0.018	-0.133
extreme_oneside	0.3	0.8	100	100	0.182	1.000	0.012	-0.118
extreme_oneside	0.3	1.0	20	20	0.230	0.992	0.051	-0.070
extreme_oneside	0.3	1.0	20	50	0.273	1.000	0.031	-0.027
extreme_oneside	0.3	1.0	20	100	0.288	1.000	0.022	-0.012
extreme_oneside	0.3	1.0	100	20	0.230	1.000	0.023	-0.070
extreme_oneside	0.3	1.0	100	50	0.273	1.000	0.014	-0.027
extreme_oneside	0.3	1.0	100	100	0.287	1.000	0.010	-0.013
extreme_oneside	0.5	0.4	20	20	0.023	0.063	0.127	-0.477
extreme_oneside	0.5	0.4	20	50	0.160	0.622	0.069	-0.340
extreme_oneside	0.5	0.4	20	100	0.202	0.986	0.046	-0.298
extreme_oneside	0.5	0.4	100	20	0.021	0.074	0.057	-0.479
extreme_oneside	0.5	0.4	100	50	0.163	1.000	0.031	-0.337
extreme_oneside	0.5	0.4	100	100	0.205	1.000	0.021	-0.295
extreme_oneside	0.5	0.6	20	20	0.163	0.486	0.084	-0.337
extreme_oneside	0.5	0.6	20	50	0.244	0.998	0.048	-0.256
extreme_oneside	0.5	0.6	20	100	0.270	1.000	0.033	-0.230
extreme_oneside	0.5	0.6	100	20	0.162	0.982	0.038	-0.338
extreme_oneside	0.5	0.6	100	50	0.248	1.000	0.022	-0.252
extreme_oneside	0.5	0.6	100	100	0.272	1.000	0.015	-0.228
extreme_oneside	0.5	0.8	20	20	0.268	0.983	0.063	-0.232
extreme_oneside	0.5	0.8	20	50	0.325	1.000	0.037	-0.175
extreme_oneside	0.5	0.8	20	100	0.343	1.000	0.026	-0.157
extreme_oneside	0.5	0.8	100	20	0.269	1.000	0.028	-0.231
extreme_oneside	0.5	0.8	100	50	0.327	1.000	0.017	-0.173
extreme_oneside	0.5	0.8	100	100	0.345	1.000	0.011	-0.155
extreme_oneside	0.5	1.0	20	20	0.416	1.000	0.048	-0.084
extreme_oneside	0.5	1.0	20	50	0.468	1.000	0.029	-0.032

Table 20: Simulation A. Full results. *(continued)*

Missingness pattern	Simulated fixed AR	Compliance	N participants	Beeps per participant	Estimated fixed AR	Power to detect fixed AR	Fixed AR SE	Fixed AR estimation bias
extreme_oneside	0.5	1.0	20	100	0.485	1.000	0.020	-0.015
extreme_oneside	0.5	1.0	100	20	0.416	1.000	0.021	-0.084
extreme_oneside	0.5	1.0	100	50	0.469	1.000	0.013	-0.031
extreme_oneside	0.5	1.0	100	100	0.485	1.000	0.009	-0.015
extreme_oneside	0.7	0.4	20	20	0.136	0.228	0.116	-0.564
extreme_oneside	0.7	0.4	20	50	0.315	0.995	0.061	-0.385
extreme_oneside	0.7	0.4	20	100	0.375	1.000	0.041	-0.325
extreme_oneside	0.7	0.4	100	20	0.136	0.714	0.052	-0.564
extreme_oneside	0.7	0.4	100	50	0.325	1.000	0.027	-0.375
extreme_oneside	0.7	0.4	100	100	0.380	1.000	0.018	-0.320
extreme_oneside	0.7	0.6	20	20	0.317	0.967	0.078	-0.383
extreme_oneside	0.7	0.6	20	50	0.429	1.000	0.043	-0.271
extreme_oneside	0.7	0.6	20	100	0.462	1.000	0.029	-0.238
extreme_oneside	0.7	0.6	100	20	0.319	1.000	0.035	-0.381
extreme_oneside	0.7	0.6	100	50	0.433	1.000	0.019	-0.267
extreme_oneside	0.7	0.6	100	100	0.465	1.000	0.013	-0.235
extreme_oneside	0.7	0.8	20	20	0.447	1.000	0.058	-0.253
extreme_oneside	0.7	0.8	20	50	0.524	1.000	0.033	-0.176
extreme_oneside	0.7	0.8	20	100	0.546	1.000	0.023	-0.154
extreme_oneside	0.7	0.8	100	20	0.450	1.000	0.026	-0.250
extreme_oneside	0.7	0.8	100	50	0.526	1.000	0.015	-0.174
extreme_oneside	0.7	0.8	100	100	0.548	1.000	0.010	-0.152
extreme_oneside	0.7	1.0	20	20	0.599	1.000	0.042	-0.101
extreme_oneside	0.7	1.0	20	50	0.663	1.000	0.024	-0.037
extreme_oneside	0.7	1.0	20	100	0.682	1.000	0.017	-0.018
extreme_oneside	0.7	1.0	100	20	0.600	1.000	0.019	-0.100
extreme_oneside	0.7	1.0	100	50	0.664	1.000	0.011	-0.036
extreme_oneside	0.7	1.0	100	100	0.682	1.000	0.007	-0.018
extreme_twosided	0.3	0.4	20	20	-0.093	0.127	0.164	-0.393

Table 20: Simulation A. Full results. *(continued)*

Missingness pattern	Simulated fixed AR	Compliance	N participants	Beeps per participant	Estimated fixed AR	Power to detect fixed AR	Fixed AR SE	Fixed AR estimation bias
extreme_twosided	0.3	0.4	20	50	-0.018	0.074	0.086	-0.318
extreme_twosided	0.3	0.4	20	100	0.007	0.067	0.058	-0.293
extreme_twosided	0.3	0.4	100	20	-0.095	0.280	0.072	-0.395
extreme_twosided	0.3	0.4	100	50	-0.016	0.088	0.038	-0.316
extreme_twosided	0.3	0.4	100	100	0.009	0.063	0.026	-0.291
extreme_twosided	0.3	0.6	20	20	-0.007	0.075	0.094	-0.307
extreme_twosided	0.3	0.6	20	50	0.043	0.122	0.055	-0.257
extreme_twosided	0.3	0.6	20	100	0.056	0.329	0.038	-0.244
extreme_twosided	0.3	0.6	100	20	-0.003	0.067	0.042	-0.303
extreme_twosided	0.3	0.6	100	50	0.043	0.418	0.024	-0.257
extreme_twosided	0.3	0.6	100	100	0.057	0.932	0.017	-0.243
extreme_twosided	0.3	0.8	20	20	0.086	0.252	0.067	-0.214
extreme_twosided	0.3	0.8	20	50	0.120	0.825	0.040	-0.180
extreme_twosided	0.3	0.8	20	100	0.131	1.000	0.028	-0.169
extreme_twosided	0.3	0.8	100	20	0.090	0.839	0.030	-0.210
extreme_twosided	0.3	0.8	100	50	0.121	1.000	0.018	-0.179
extreme_twosided	0.3	0.8	100	100	0.131	1.000	0.012	-0.169
extreme_twosided	0.3	1.0	20	20	0.230	0.992	0.051	-0.070
extreme_twosided	0.3	1.0	20	50	0.273	1.000	0.031	-0.027
extreme_twosided	0.3	1.0	20	100	0.288	1.000	0.022	-0.012
extreme_twosided	0.3	1.0	100	20	0.230	1.000	0.023	-0.070
extreme_twosided	0.3	1.0	100	50	0.273	1.000	0.014	-0.027
extreme_twosided	0.3	1.0	100	100	0.287	1.000	0.010	-0.013
extreme_twosided	0.5	0.4	20	20	-0.045	0.091	0.158	-0.545
extreme_twosided	0.5	0.4	20	50	0.019	0.072	0.082	-0.481
extreme_twosided	0.5	0.4	20	100	0.037	0.110	0.055	-0.463
extreme_twosided	0.5	0.4	100	20	-0.050	0.143	0.070	-0.550
extreme_twosided	0.5	0.4	100	50	0.021	0.112	0.037	-0.479
extreme_twosided	0.5	0.4	100	100	0.040	0.371	0.025	-0.460

Table 20: Simulation A. Full results. *(continued)*

Missingness pattern	Simulated fixed AR	Compliance	N participants	Beeps per participant	Estimated fixed AR	Power to detect fixed AR	Fixed AR SE	Fixed AR estimation bias
extreme_twosided	0.5	0.6	20	20	0.075	0.144	0.092	-0.425
extreme_twosided	0.5	0.6	20	50	0.116	0.574	0.053	-0.384
extreme_twosided	0.5	0.6	20	100	0.128	0.940	0.036	-0.372
extreme_twosided	0.5	0.6	100	20	0.080	0.494	0.041	-0.420
extreme_twosided	0.5	0.6	100	50	0.117	0.998	0.024	-0.383
extreme_twosided	0.5	0.6	100	100	0.129	1.000	0.016	-0.371
extreme_twosided	0.5	0.8	20	20	0.217	0.880	0.065	-0.283
extreme_twosided	0.5	0.8	20	50	0.250	1.000	0.039	-0.250
extreme_twosided	0.5	0.8	20	100	0.258	1.000	0.027	-0.242
extreme_twosided	0.5	0.8	100	20	0.219	1.000	0.029	-0.281
extreme_twosided	0.5	0.8	100	50	0.250	1.000	0.017	-0.250
extreme_twosided	0.5	0.8	100	100	0.259	1.000	0.012	-0.241
extreme_twosided	0.5	1.0	20	20	0.416	1.000	0.048	-0.084
extreme_twosided	0.5	1.0	20	50	0.468	1.000	0.029	-0.032
extreme_twosided	0.5	1.0	20	100	0.485	1.000	0.020	-0.015
extreme_twosided	0.5	1.0	100	20	0.416	1.000	0.021	-0.084
extreme_twosided	0.5	1.0	100	50	0.469	1.000	0.013	-0.031
extreme_twosided	0.5	1.0	100	100	0.485	1.000	0.009	-0.015
extreme_twosided	0.7	0.4	20	20	0.040	0.089	0.149	-0.660
extreme_twosided	0.7	0.4	20	50	0.089	0.234	0.076	-0.611
extreme_twosided	0.7	0.4	20	100	0.108	0.552	0.051	-0.592
extreme_twosided	0.7	0.4	100	20	0.045	0.170	0.066	-0.655
extreme_twosided	0.7	0.4	100	50	0.093	0.752	0.034	-0.607
extreme_twosided	0.7	0.4	100	100	0.109	0.999	0.023	-0.591
extreme_twosided	0.7	0.6	20	20	0.216	0.656	0.087	-0.484
extreme_twosided	0.7	0.6	20	50	0.252	0.992	0.050	-0.448
extreme_twosided	0.7	0.6	20	100	0.258	1.000	0.034	-0.442
extreme_twosided	0.7	0.6	100	20	0.222	0.998	0.039	-0.478
extreme_twosided	0.7	0.6	100	50	0.255	1.000	0.022	-0.445

Table 20: Simulation A. Full results. *(continued)*

Missingness pattern	Simulated fixed AR	Compliance	N participants	Beeps per participant	Estimated fixed AR	Power to detect fixed AR	Fixed AR SE	Fixed AR estimation bias
extreme_twosided	0.7	0.6	100	100	0.260	1.000	0.015	-0.440
extreme_twosided	0.7	0.8	20	20	0.398	1.000	0.061	-0.302
extreme_twosided	0.7	0.8	20	50	0.437	1.000	0.035	-0.263
extreme_twosided	0.7	0.8	20	100	0.448	1.000	0.024	-0.252
extreme_twosided	0.7	0.8	100	20	0.400	1.000	0.027	-0.300
extreme_twosided	0.7	0.8	100	50	0.440	1.000	0.016	-0.260
extreme_twosided	0.7	0.8	100	100	0.448	1.000	0.011	-0.252
extreme_twosided	0.7	1.0	20	20	0.599	1.000	0.042	-0.101
extreme_twosided	0.7	1.0	20	50	0.663	1.000	0.024	-0.037
extreme_twosided	0.7	1.0	20	100	0.682	1.000	0.017	-0.018
extreme_twosided	0.7	1.0	100	20	0.600	1.000	0.019	-0.100
extreme_twosided	0.7	1.0	100	50	0.664	1.000	0.011	-0.036
extreme_twosided	0.7	1.0	100	100	0.682	1.000	0.007	-0.018
mcar	0.3	0.4	20	20	0.134	0.139	0.162	-0.166
mcar	0.3	0.4	20	50	0.244	0.800	0.085	-0.056
mcar	0.3	0.4	20	100	0.272	0.998	0.056	-0.028
mcar	0.3	0.4	100	20	0.143	0.527	0.072	-0.157
mcar	0.3	0.4	100	50	0.244	1.000	0.038	-0.056
mcar	0.3	0.4	100	100	0.271	1.000	0.025	-0.029
mcar	0.3	0.6	20	20	0.197	0.581	0.093	-0.103
mcar	0.3	0.6	20	50	0.260	0.999	0.053	-0.040
mcar	0.3	0.6	20	100	0.281	1.000	0.037	-0.019
mcar	0.3	0.6	100	20	0.193	0.998	0.041	-0.107
mcar	0.3	0.6	100	50	0.259	1.000	0.024	-0.041
mcar	0.3	0.6	100	100	0.281	1.000	0.016	-0.019
mcar	0.3	0.8	20	20	0.213	0.882	0.066	-0.087
mcar	0.3	0.8	20	50	0.268	1.000	0.039	-0.032
mcar	0.3	0.8	20	100	0.284	1.000	0.027	-0.016
mcar	0.3	0.8	100	20	0.216	1.000	0.029	-0.084

Table 20: Simulation A. Full results. *(continued)*

Missingness pattern	Simulated fixed AR	Compliance	N participants	Beeps per participant	Estimated fixed AR	Power to detect fixed AR	Fixed AR SE	Fixed AR estimation bias
mcar	0.3	0.8	100	50	0.268	1.000	0.018	-0.032
mcar	0.3	0.8	100	100	0.284	1.000	0.012	-0.016
mcar	0.3	1.0	20	20	0.230	0.992	0.051	-0.070
mcar	0.3	1.0	20	50	0.273	1.000	0.031	-0.027
mcar	0.3	1.0	20	100	0.288	1.000	0.022	-0.012
mcar	0.3	1.0	100	20	0.230	1.000	0.023	-0.070
mcar	0.3	1.0	100	50	0.273	1.000	0.014	-0.027
mcar	0.3	1.0	100	100	0.287	1.000	0.010	-0.013
mcar	0.5	0.4	20	20	0.330	0.545	0.157	-0.170
mcar	0.5	0.4	20	50	0.443	1.000	0.079	-0.057
mcar	0.5	0.4	20	100	0.472	1.000	0.052	-0.028
mcar	0.5	0.4	100	20	0.341	0.993	0.069	-0.159
mcar	0.5	0.4	100	50	0.443	1.000	0.035	-0.057
mcar	0.5	0.4	100	100	0.472	1.000	0.023	-0.028
mcar	0.5	0.6	20	20	0.387	0.986	0.087	-0.113
mcar	0.5	0.6	20	50	0.457	1.000	0.049	-0.043
mcar	0.5	0.6	20	100	0.480	1.000	0.033	-0.020
mcar	0.5	0.6	100	20	0.384	1.000	0.039	-0.116
mcar	0.5	0.6	100	50	0.456	1.000	0.022	-0.044
mcar	0.5	0.6	100	100	0.480	1.000	0.015	-0.020
mcar	0.5	0.8	20	20	0.402	1.000	0.062	-0.098
mcar	0.5	0.8	20	50	0.464	1.000	0.036	-0.036
mcar	0.5	0.8	20	100	0.482	1.000	0.025	-0.018
mcar	0.5	0.8	100	20	0.405	1.000	0.028	-0.095
mcar	0.5	0.8	100	50	0.464	1.000	0.016	-0.036
mcar	0.5	0.8	100	100	0.482	1.000	0.011	-0.018
mcar	0.5	1.0	20	20	0.416	1.000	0.048	-0.084
mcar	0.5	1.0	20	50	0.468	1.000	0.029	-0.032
mcar	0.5	1.0	20	100	0.485	1.000	0.020	-0.015

Table 20: Simulation A. Full results. *(continued)*

Missingness pattern	Simulated fixed AR	Compliance	N participants	Beeps per participant	Estimated fixed AR	Power to detect fixed AR	Fixed AR SE	Fixed AR estimation bias
mcar	0.5	1.0	100	20	0.416	1.000	0.021	-0.084
mcar	0.5	1.0	100	50	0.469	1.000	0.013	-0.031
mcar	0.5	1.0	100	100	0.485	1.000	0.009	-0.015
mcar	0.7	0.4	20	20	0.530	0.911	0.142	-0.170
mcar	0.7	0.4	20	50	0.644	1.000	0.067	-0.056
mcar	0.7	0.4	20	100	0.673	1.000	0.043	-0.027
mcar	0.7	0.4	100	20	0.541	1.000	0.062	-0.159
mcar	0.7	0.4	100	50	0.645	1.000	0.030	-0.055
mcar	0.7	0.4	100	100	0.673	1.000	0.019	-0.027
mcar	0.7	0.6	20	20	0.576	1.000	0.078	-0.124
mcar	0.7	0.6	20	50	0.655	1.000	0.042	-0.045
mcar	0.7	0.6	20	100	0.679	1.000	0.028	-0.021
mcar	0.7	0.6	100	20	0.575	1.000	0.035	-0.125
mcar	0.7	0.6	100	50	0.655	1.000	0.019	-0.045
mcar	0.7	0.6	100	100	0.679	1.000	0.013	-0.021
mcar	0.7	0.8	20	20	0.588	1.000	0.054	-0.112
mcar	0.7	0.8	20	50	0.659	1.000	0.031	-0.041
mcar	0.7	0.8	20	100	0.679	1.000	0.021	-0.021
mcar	0.7	0.8	100	20	0.590	1.000	0.024	-0.110
mcar	0.7	0.8	100	50	0.660	1.000	0.014	-0.040
mcar	0.7	0.8	100	100	0.681	1.000	0.009	-0.019
mcar	0.7	1.0	20	20	0.599	1.000	0.042	-0.101
mcar	0.7	1.0	20	50	0.663	1.000	0.024	-0.037
mcar	0.7	1.0	20	100	0.682	1.000	0.017	-0.018
mcar	0.7	1.0	100	20	0.600	1.000	0.019	-0.100
mcar	0.7	1.0	100	50	0.664	1.000	0.011	-0.036
mcar	0.7	1.0	100	100	0.682	1.000	0.007	-0.018

Appendix 2: Full results from Simulation B

Table 21: Simulation B. Full results.

Missingness pattern	Simulated fixed AR	Compliance	N participants	Beeps per participant	Estimated fixed AR	Power to detect fixed AR	Fixed AR SE	Fixed AR estimation bias
block	0.3	0.4	20	50	0.221	0.852	0.071	-0.079
block	0.3	0.4	20	50	0.213	0.722	0.081	-0.087
block	0.3	0.4	20	100	0.264	0.985	0.059	-0.036
block	0.3	0.4	20	100	0.252	0.919	0.072	-0.048
block	0.3	0.4	100	50	0.226	1.000	0.032	-0.074
block	0.3	0.4	100	50	0.218	1.000	0.037	-0.082
block	0.3	0.4	100	100	0.263	1.000	0.027	-0.037
block	0.3	0.4	100	100	0.253	1.000	0.033	-0.047
block	0.3	0.6	20	50	0.248	0.964	0.062	-0.052
block	0.3	0.6	20	50	0.237	0.847	0.075	-0.063
block	0.3	0.6	20	100	0.278	0.997	0.055	-0.022
block	0.3	0.6	20	100	0.263	0.953	0.070	-0.037
block	0.3	0.6	100	50	0.251	1.000	0.028	-0.049
block	0.3	0.6	100	50	0.242	1.000	0.034	-0.058
block	0.3	0.6	100	100	0.275	1.000	0.025	-0.025
block	0.3	0.6	100	100	0.265	1.000	0.032	-0.035
block	0.3	0.8	20	50	0.260	0.989	0.059	-0.040
block	0.3	0.8	20	50	0.250	0.909	0.072	-0.050
block	0.3	0.8	20	100	0.283	0.997	0.054	-0.017
block	0.3	0.8	20	100	0.270	0.965	0.069	-0.030
block	0.3	0.8	100	50	0.263	1.000	0.027	-0.037
block	0.3	0.8	100	50	0.254	1.000	0.033	-0.046
block	0.3	0.8	100	100	0.281	1.000	0.024	-0.019
block	0.3	0.8	100	100	0.271	1.000	0.031	-0.029
block	0.3	1.0	20	50	0.268	0.994	0.056	-0.032
block	0.3	1.0	20	50	0.258	0.938	0.071	-0.042
block	0.3	1.0	20	100	0.287	0.998	0.053	-0.013
block	0.3	1.0	20	100	0.273	0.971	0.069	-0.027
block	0.3	1.0	100	50	0.270	1.000	0.026	-0.030

Table 21: Simulation B. Full results. *(continued)*

Missingness pattern	Simulated fixed AR	Compliance	N participants	Beeps per participant	Estimated fixed AR	Power to detect fixed AR	Fixed AR SE	Fixed AR estimation bias
block	0.3	1.0	100	50	0.261	1.000	0.032	-0.039
block	0.3	1.0	100	100	0.285	1.000	0.024	-0.015
block	0.3	1.0	100	100	0.274	1.000	0.031	-0.026
block	0.7	0.4	20	50	0.574	1.000	0.059	-0.126
block	0.7	0.4	20	50	0.519	1.000	0.067	-0.181
block	0.7	0.4	20	100	0.614	1.000	0.048	-0.086
block	0.7	0.4	20	100	0.556	1.000	0.058	-0.144
block	0.7	0.4	100	50	0.576	1.000	0.027	-0.124
block	0.7	0.4	100	50	0.522	1.000	0.031	-0.178
block	0.7	0.4	100	100	0.615	1.000	0.022	-0.085
block	0.7	0.4	100	100	0.560	1.000	0.026	-0.140
block	0.7	0.6	20	50	0.600	1.000	0.051	-0.100
block	0.7	0.6	20	50	0.544	1.000	0.060	-0.156
block	0.7	0.6	20	100	0.629	1.000	0.045	-0.071
block	0.7	0.6	20	100	0.570	1.000	0.056	-0.130
block	0.7	0.6	100	50	0.600	1.000	0.023	-0.100
block	0.7	0.6	100	50	0.545	1.000	0.027	-0.155
block	0.7	0.6	100	100	0.629	1.000	0.021	-0.071
block	0.7	0.6	100	100	0.574	1.000	0.025	-0.126
block	0.7	0.8	20	50	0.615	1.000	0.047	-0.085
block	0.7	0.8	20	50	0.558	1.000	0.057	-0.142
block	0.7	0.8	20	100	0.636	1.000	0.044	-0.064
block	0.7	0.8	20	100	0.577	1.000	0.055	-0.123
block	0.7	0.8	100	50	0.613	1.000	0.022	-0.087
block	0.7	0.8	100	50	0.557	1.000	0.026	-0.143
block	0.7	0.8	100	100	0.637	1.000	0.020	-0.063
block	0.7	0.8	100	100	0.581	1.000	0.025	-0.119
block	0.7	1.0	20	50	0.621	1.000	0.045	-0.079
block	0.7	1.0	20	50	0.564	1.000	0.056	-0.136

Table 21: Simulation B. Full results. *(continued)*

Missingness pattern	Simulated fixed AR	Compliance	N participants	Beeps per participant	Estimated fixed AR	Power to detect fixed AR	Fixed AR SE	Fixed AR estimation bias
block	0.7	1.0	20	100	0.640	1.000	0.044	-0.060
block	0.7	1.0	20	100	0.581	1.000	0.055	-0.119
block	0.7	1.0	100	50	0.620	1.000	0.021	-0.080
block	0.7	1.0	100	50	0.563	1.000	0.025	-0.137
block	0.7	1.0	100	100	0.641	1.000	0.020	-0.059
block	0.7	1.0	100	100	0.585	1.000	0.025	-0.115
extreme_oneside	0.3	0.4	20	50	0.086	0.169	0.084	-0.214
extreme_oneside	0.3	0.4	20	50	0.123	0.301	0.086	-0.177
extreme_oneside	0.3	0.4	20	100	0.129	0.565	0.061	-0.171
extreme_oneside	0.3	0.4	20	100	0.157	0.642	0.066	-0.143
extreme_oneside	0.3	0.4	100	50	0.098	0.731	0.038	-0.202
extreme_oneside	0.3	0.4	100	50	0.131	0.919	0.040	-0.169
extreme_oneside	0.3	0.4	100	100	0.134	0.997	0.028	-0.166
extreme_oneside	0.3	0.4	100	100	0.164	1.000	0.031	-0.136
extreme_oneside	0.3	0.6	20	50	0.139	0.620	0.062	-0.161
extreme_oneside	0.3	0.6	20	50	0.161	0.651	0.068	-0.139
extreme_oneside	0.3	0.6	20	100	0.165	0.911	0.049	-0.135
extreme_oneside	0.3	0.6	20	100	0.180	0.885	0.058	-0.120
extreme_oneside	0.3	0.6	100	50	0.144	1.000	0.028	-0.156
extreme_oneside	0.3	0.6	100	50	0.166	0.999	0.031	-0.134
extreme_oneside	0.3	0.6	100	100	0.166	1.000	0.023	-0.134
extreme_oneside	0.3	0.6	100	100	0.182	1.000	0.027	-0.118
extreme_oneside	0.3	0.8	20	50	0.184	0.915	0.054	-0.116
extreme_oneside	0.3	0.8	20	50	0.192	0.860	0.064	-0.108
extreme_oneside	0.3	0.8	20	100	0.202	0.990	0.047	-0.098
extreme_oneside	0.3	0.8	20	100	0.207	0.945	0.058	-0.093
extreme_oneside	0.3	0.8	100	50	0.187	1.000	0.025	-0.113
extreme_oneside	0.3	0.8	100	50	0.196	1.000	0.029	-0.104
extreme_oneside	0.3	0.8	100	100	0.202	1.000	0.022	-0.098

Table 21: Simulation B. Full results. *(continued)*

Missingness pattern	Simulated fixed AR	Compliance	N participants	Beeps per participant	Estimated fixed AR	Power to detect fixed AR	Fixed AR SE	Fixed AR estimation bias
extreme_oneside	0.3	0.8	100	100	0.209	1.000	0.027	-0.091
extreme_oneside	0.3	1.0	20	50	0.268	0.994	0.056	-0.032
extreme_oneside	0.3	1.0	20	50	0.258	0.938	0.071	-0.042
extreme_oneside	0.3	1.0	20	100	0.287	0.998	0.053	-0.013
extreme_oneside	0.3	1.0	20	100	0.273	0.971	0.069	-0.027
extreme_oneside	0.3	1.0	100	50	0.270	1.000	0.026	-0.030
extreme_oneside	0.3	1.0	100	50	0.261	1.000	0.032	-0.039
extreme_oneside	0.3	1.0	100	100	0.285	1.000	0.024	-0.015
extreme_oneside	0.3	1.0	100	100	0.274	1.000	0.031	-0.026
extreme_oneside	0.7	0.4	20	50	0.325	0.976	0.075	-0.375
extreme_oneside	0.7	0.4	20	50	0.306	0.933	0.078	-0.394
extreme_oneside	0.7	0.4	20	100	0.387	1.000	0.060	-0.313
extreme_oneside	0.7	0.4	20	100	0.360	1.000	0.064	-0.340
extreme_oneside	0.7	0.4	100	50	0.333	1.000	0.035	-0.367
extreme_oneside	0.7	0.4	100	50	0.313	1.000	0.036	-0.387
extreme_oneside	0.7	0.4	100	100	0.390	1.000	0.028	-0.310
extreme_oneside	0.7	0.4	100	100	0.365	1.000	0.030	-0.335
extreme_oneside	0.7	0.6	20	50	0.418	1.000	0.061	-0.282
extreme_oneside	0.7	0.6	20	50	0.385	0.999	0.065	-0.315
extreme_oneside	0.7	0.6	20	100	0.454	1.000	0.054	-0.246
extreme_oneside	0.7	0.6	20	100	0.416	1.000	0.059	-0.284
extreme_oneside	0.7	0.6	100	50	0.421	1.000	0.028	-0.279
extreme_oneside	0.7	0.6	100	50	0.387	1.000	0.030	-0.313
extreme_oneside	0.7	0.6	100	100	0.456	1.000	0.025	-0.244
extreme_oneside	0.7	0.6	100	100	0.419	1.000	0.027	-0.281
extreme_oneside	0.7	0.8	20	50	0.498	1.000	0.054	-0.202
extreme_oneside	0.7	0.8	20	50	0.453	1.000	0.060	-0.247
extreme_oneside	0.7	0.8	20	100	0.522	1.000	0.050	-0.178
extreme_oneside	0.7	0.8	20	100	0.473	1.000	0.057	-0.227

Table 21: Simulation B. Full results. *(continued)*

Missingness pattern	Simulated fixed AR	Compliance	N participants	Beeps per participant	Estimated fixed AR	Power to detect fixed AR	Fixed AR SE	Fixed AR estimation bias
extreme_oneside	0.7	0.8	100	50	0.498	1.000	0.025	-0.202
extreme_oneside	0.7	0.8	100	50	0.454	1.000	0.028	-0.246
extreme_oneside	0.7	0.8	100	100	0.524	1.000	0.023	-0.176
extreme_oneside	0.7	0.8	100	100	0.478	1.000	0.026	-0.222
extreme_oneside	0.7	1.0	20	50	0.621	1.000	0.045	-0.079
extreme_oneside	0.7	1.0	20	50	0.564	1.000	0.056	-0.136
extreme_oneside	0.7	1.0	20	100	0.640	1.000	0.044	-0.060
extreme_oneside	0.7	1.0	20	100	0.581	1.000	0.055	-0.119
extreme_oneside	0.7	1.0	100	50	0.620	1.000	0.021	-0.080
extreme_oneside	0.7	1.0	100	50	0.563	1.000	0.025	-0.137
extreme_oneside	0.7	1.0	100	100	0.641	1.000	0.020	-0.059
extreme_oneside	0.7	1.0	100	100	0.585	1.000	0.025	-0.115
extreme_twosided	0.3	0.4	20	50	-0.008	0.053	0.092	-0.308
extreme_twosided	0.3	0.4	20	50	0.009	0.063	0.091	-0.291
extreme_twosided	0.3	0.4	20	100	0.021	0.065	0.062	-0.279
extreme_twosided	0.3	0.4	20	100	0.037	0.098	0.063	-0.263
extreme_twosided	0.3	0.4	100	50	0.002	0.069	0.040	-0.298
extreme_twosided	0.3	0.4	100	50	0.023	0.089	0.042	-0.277
extreme_twosided	0.3	0.4	100	100	0.026	0.155	0.027	-0.274
extreme_twosided	0.3	0.4	100	100	0.026	0.155	0.027	-0.274
extreme_twosided	0.3	0.4	100	100	0.046	0.349	0.029	-0.254
extreme_twosided	0.3	0.6	20	50	0.063	0.182	0.061	-0.237
extreme_twosided	0.3	0.6	20	50	0.079	0.226	0.065	-0.221
extreme_twosided	0.3	0.6	20	100	0.080	0.428	0.045	-0.220
extreme_twosided	0.3	0.6	20	100	0.096	0.462	0.050	-0.204
extreme_twosided	0.3	0.6	100	50	0.069	0.693	0.028	-0.231
extreme_twosided	0.3	0.6	100	50	0.088	0.836	0.030	-0.212
extreme_twosided	0.3	0.6	100	100	0.099	0.988	0.024	-0.201
extreme_twosided	0.3	0.8	20	50	0.141	0.760	0.052	-0.159

Table 21: Simulation B. Full results. *(continued)*

Missingness pattern	Simulated fixed AR	Compliance	N participants	Beeps per participant	Estimated fixed AR	Power to detect fixed AR	Fixed AR SE	Fixed AR estimation bias
extreme_twosided	0.3	0.8	20	50	0.151	0.700	0.061	-0.149
extreme_twosided	0.3	0.8	20	100	0.156	0.959	0.043	-0.144
extreme_twosided	0.3	0.8	20	100	0.162	0.877	0.053	-0.138
extreme_twosided	0.3	0.8	100	50	0.145	1.000	0.024	-0.155
extreme_twosided	0.3	0.8	100	50	0.156	1.000	0.028	-0.144
extreme_twosided	0.3	0.8	100	100	0.155	1.000	0.020	-0.145
extreme_twosided	0.3	0.8	100	100	0.164	1.000	0.025	-0.136
extreme_twosided	0.3	1.0	20	50	0.268	0.994	0.056	-0.032
extreme_twosided	0.3	1.0	20	50	0.258	0.938	0.071	-0.042
extreme_twosided	0.3	1.0	20	100	0.287	0.998	0.053	-0.013
extreme_twosided	0.3	1.0	20	100	0.273	0.971	0.069	-0.027
extreme_twosided	0.3	1.0	100	50	0.270	1.000	0.026	-0.030
extreme_twosided	0.3	1.0	100	50	0.261	1.000	0.032	-0.039
extreme_twosided	0.3	1.0	100	100	0.285	1.000	0.024	-0.015
extreme_twosided	0.3	1.0	100	100	0.274	1.000	0.031	-0.026
extreme_twosided	0.7	0.4	20	50	0.150	0.384	0.089	-0.550
extreme_twosided	0.7	0.4	20	50	0.144	0.363	0.089	-0.556
extreme_twosided	0.7	0.4	20	100	0.178	0.755	0.065	-0.522
extreme_twosided	0.7	0.4	20	100	0.169	0.674	0.068	-0.531
extreme_twosided	0.7	0.4	100	50	0.158	0.959	0.042	-0.542
extreme_twosided	0.7	0.4	100	50	0.154	0.937	0.044	-0.546
extreme_twosided	0.7	0.4	100	100	0.181	1.000	0.032	-0.519
extreme_twosided	0.7	0.4	100	100	0.177	1.000	0.033	-0.523
extreme_twosided	0.7	0.6	20	50	0.290	0.981	0.067	-0.410
extreme_twosided	0.7	0.6	20	50	0.268	0.956	0.069	-0.432
extreme_twosided	0.7	0.6	20	100	0.303	1.000	0.056	-0.397
extreme_twosided	0.7	0.6	20	100	0.278	0.996	0.060	-0.422
extreme_twosided	0.7	0.6	100	50	0.289	1.000	0.032	-0.411
extreme_twosided	0.7	0.6	100	50	0.270	1.000	0.033	-0.430

Table 21: Simulation B. Full results. *(continued)*

Missingness pattern	Simulated fixed AR	Compliance	N participants	Beeps per participant	Estimated fixed AR	Power to detect fixed AR	Fixed AR SE	Fixed AR estimation bias
extreme_twosided	0.7	0.6	100	100	0.305	1.000	0.026	-0.395
extreme_twosided	0.7	0.6	100	100	0.284	1.000	0.028	-0.416
extreme_twosided	0.7	0.8	20	50	0.434	1.000	0.057	-0.266
extreme_twosided	0.7	0.8	20	50	0.393	1.000	0.062	-0.307
extreme_twosided	0.7	0.8	20	100	0.446	1.000	0.053	-0.254
extreme_twosided	0.7	0.8	20	100	0.403	1.000	0.058	-0.297
extreme_twosided	0.7	0.8	100	50	0.431	1.000	0.026	-0.269
extreme_twosided	0.7	0.8	100	50	0.394	1.000	0.029	-0.306
extreme_twosided	0.7	0.8	100	100	0.448	1.000	0.024	-0.252
extreme_twosided	0.7	0.8	100	100	0.409	1.000	0.027	-0.291
extreme_twosided	0.7	1.0	20	50	0.621	1.000	0.045	-0.079
extreme_twosided	0.7	1.0	20	50	0.564	1.000	0.056	-0.136
extreme_twosided	0.7	1.0	20	100	0.640	1.000	0.044	-0.060
extreme_twosided	0.7	1.0	20	100	0.581	1.000	0.055	-0.119
extreme_twosided	0.7	1.0	100	50	0.620	1.000	0.021	-0.080
extreme_twosided	0.7	1.0	100	50	0.563	1.000	0.025	-0.137
extreme_twosided	0.7	1.0	100	100	0.641	1.000	0.020	-0.059
extreme_twosided	0.7	1.0	100	100	0.585	1.000	0.025	-0.115
mcar	0.3	0.4	20	50	0.255	0.702	0.099	-0.045
mcar	0.3	0.4	20	50	0.251	0.622	0.108	-0.049
mcar	0.3	0.4	20	100	0.281	0.956	0.075	-0.019
mcar	0.3	0.4	20	100	0.268	0.836	0.087	-0.032
mcar	0.3	0.4	100	50	0.256	1.000	0.045	-0.044
mcar	0.3	0.4	100	50	0.248	0.996	0.049	-0.052
mcar	0.3	0.4	100	100	0.283	1.000	0.034	-0.017
mcar	0.3	0.4	100	100	0.276	1.000	0.040	-0.024
mcar	0.3	0.6	20	50	0.265	0.932	0.071	-0.035
mcar	0.3	0.6	20	50	0.248	0.813	0.083	-0.052
mcar	0.3	0.6	20	100	0.284	0.995	0.061	-0.016

Table 21: Simulation B. Full results. *(continued)*

Missingness pattern	Simulated fixed AR	Compliance	N participants	Beeps per participant	Estimated fixed AR	Power to detect fixed AR	Fixed AR SE	Fixed AR estimation bias
mcar	0.3	0.6	20	100	0.270	0.908	0.075	-0.030
mcar	0.3	0.6	100	50	0.264	1.000	0.033	-0.036
mcar	0.3	0.6	100	50	0.256	0.999	0.038	-0.044
mcar	0.3	0.6	100	100	0.284	1.000	0.028	-0.016
mcar	0.3	0.6	100	100	0.273	1.000	0.034	-0.027
mcar	0.3	0.8	20	50	0.269	0.985	0.061	-0.031
mcar	0.3	0.8	20	50	0.263	0.920	0.074	-0.037
mcar	0.3	0.8	20	100	0.283	0.999	0.055	-0.017
mcar	0.3	0.8	20	100	0.273	0.952	0.071	-0.027
mcar	0.3	0.8	100	50	0.270	1.000	0.028	-0.030
mcar	0.3	0.8	100	50	0.260	1.000	0.034	-0.040
mcar	0.3	0.8	100	100	0.286	1.000	0.025	-0.014
mcar	0.3	0.8	100	100	0.275	1.000	0.032	-0.025
mcar	0.3	1.0	20	50	0.268	0.994	0.056	-0.032
mcar	0.3	1.0	20	50	0.258	0.938	0.071	-0.042
mcar	0.3	1.0	20	100	0.287	0.998	0.053	-0.013
mcar	0.3	1.0	20	100	0.273	0.971	0.069	-0.027
mcar	0.3	1.0	100	50	0.270	1.000	0.026	-0.030
mcar	0.3	1.0	100	50	0.261	1.000	0.032	-0.039
mcar	0.3	1.0	100	100	0.285	1.000	0.024	-0.015
mcar	0.3	1.0	100	100	0.274	1.000	0.031	-0.026
mcar	0.7	0.4	20	50	0.626	1.000	0.078	-0.074
mcar	0.7	0.4	20	50	0.578	1.000	0.086	-0.122
mcar	0.7	0.4	20	100	0.659	1.000	0.058	-0.041
mcar	0.7	0.4	20	100	0.604	1.000	0.068	-0.096
mcar	0.7	0.4	100	50	0.638	1.000	0.035	-0.062
mcar	0.7	0.4	100	50	0.584	1.000	0.039	-0.116
mcar	0.7	0.4	100	100	0.664	1.000	0.027	-0.036
mcar	0.7	0.4	100	100	0.610	1.000	0.031	-0.090

Table 21: Simulation B. Full results. *(continued)*

Missingness pattern	Simulated fixed AR	Compliance	N participants	Beeps per participant	Estimated fixed AR	Power to detect fixed AR	Fixed AR SE	Fixed AR estimation bias
mcar	0.7	0.6	20	50	0.628	1.000	0.057	-0.072
mcar	0.7	0.6	20	50	0.571	1.000	0.066	-0.129
mcar	0.7	0.6	20	100	0.652	1.000	0.048	-0.048
mcar	0.7	0.6	20	100	0.593	1.000	0.060	-0.107
mcar	0.7	0.6	100	50	0.631	1.000	0.026	-0.069
mcar	0.7	0.6	100	50	0.576	1.000	0.030	-0.124
mcar	0.7	0.6	100	100	0.652	1.000	0.022	-0.048
mcar	0.7	0.6	100	100	0.592	1.000	0.027	-0.108
mcar	0.7	0.8	20	50	0.626	1.000	0.049	-0.074
mcar	0.7	0.8	20	50	0.571	1.000	0.058	-0.129
mcar	0.7	0.8	20	100	0.645	1.000	0.045	-0.055
mcar	0.7	0.8	20	100	0.585	1.000	0.057	-0.115
mcar	0.7	0.8	100	50	0.625	1.000	0.022	-0.075
mcar	0.7	0.8	100	50	0.568	1.000	0.027	-0.132
mcar	0.7	0.8	100	100	0.646	1.000	0.020	-0.054
mcar	0.7	0.8	100	100	0.587	1.000	0.026	-0.113
mcar	0.7	1.0	20	50	0.621	1.000	0.045	-0.079
mcar	0.7	1.0	20	50	0.564	1.000	0.056	-0.136
mcar	0.7	1.0	20	100	0.640	1.000	0.044	-0.060
mcar	0.7	1.0	20	100	0.581	1.000	0.055	-0.119
mcar	0.7	1.0	100	50	0.620	1.000	0.021	-0.080
mcar	0.7	1.0	100	50	0.563	1.000	0.025	-0.137
mcar	0.7	1.0	100	100	0.641	1.000	0.020	-0.059
mcar	0.7	1.0	100	100	0.585	1.000	0.025	-0.115