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

Benjamin Simsa

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

2023

Contents

Introduction 2

Compliance and missing data in ESM studies . . . 2

Methods 3

Multilevel AR(1) model . . . 3

Assumptions of the MLAR(1) model . . . 4

Estimation procedures for the MLAR(1) model . . . 4

Multilevel AR(1) model and statistical power . . . 4

Simulation study . . . 4

Simulation procedure . . . 4

Reproducibility and code/data availability . . . 9

Results 10

Simulation A . . . 10

Outcome: Estimation bias (MSE) . . . 10

Outcome: Standard error . . . 14

Outcome: Statistical power . . . 16

Simulation B . . . 18

Discussion 19

Limitations . . . 19

References 21

Introduction

Compliance and missing data in ESM studies

- Missing data happen quite often in intensive longitudinal designs. Average com

Average compliance in ESM studies.

Factors associated with compliance.

TODO

Methods

Multilevel AR(1) model

In this subchapter, I will describe the mathematical basis, assumptions and estimation procedures of the first-order multilevel autoregressive (MLAR(1)) model with random intercepts and random autoregressive effects, which is the focus of the simulation part of the thesis. While the notation for the model varies across different papers, the notation used by Lafit et al. (2020) will be adhered to throughout the thesis.

The MLAR 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 (inertia) 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). Following the recommendations by (Hamaker & Grasman, 2015), the predictor (lagged variable) was person-mean centered. The innovation ϵ_{it} (i.e., residuals, the part of the 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_e^2 (Lafit et al., 2020). The model used in the present thesis assumes that the innovation variance is 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 1i and the person-specific intercepts 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}. The random effects ν_{1i} themselves come from a normal distribution with mean of 0 and variance $\sigma_{\nu 1}^2$ (Lafit et al., 2020). The same holds for the person-specific intercepts γ_{0i}: they are a sum of a fixed effect β_{00} and a random effect ν_{0i} that comes from $N(0,\, \sigma_{\nu 0}^2)$.

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

Assumptions of the MLAR(1) model

In this part, the assumptions of the MLAR(1) model and the way they were taken into account in the present simulation study will be explained.

Stationarity. The MLAR(1) model is used to model stable processes in which no temporal trends are present. As such, it assumes weak stationarity: the (person-specific) process mean, innovation variance, and autoregressive parameter are assumed to not change through the time series (Rovine & Walls, 2006). For this reason, 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).

Exogeneity. TODO (see word file)

Equally spaced measurements. TODO (see word file again)

Estimation procedures for the MLAR(1) model

Initial conditions problem. TODO (see Word)

Multilevel AR(1) model and statistical power

Simulation study

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 data missing dependent on process value) on estimation performance/bias, standard error and statistical power for the estimation of the fixed autoregressive effect in the MLAR(1) model.

Simulation procedure

The study followed the general principles of a 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 in the simulation 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

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
Timepoints 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 runs per condition	1000	1000

effects variance set to either 0.05 or 0.1). Both random and fixed intercepts were estimated in Simulations A and B.

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). Each of the conditions was simulated in 1,000 simulation runs. As such, 544,000 datasets were generated (and the same number of models was estimated) in this simulation study. The manipulated variables (as well as the other parameters that remained fixed throughout all simulation conditions) are listed in Table 1 and 2.

Data generation. First, for each of the simulation conditions (i.e., combination of the parameters listed below), 1000 synthetic datasets were generated. Each dataset contained observations from N 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 with 1,000 observation was generated. 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 and random autoregressive effects $\nu_1 i$ were included in Simulation B. 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. Secondly, 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 were introduced to the data: a) data missing completely at random (MCAR); b) data missing in blocks of consecutive observations; c) lowest (100%-compliance) observations set as missing, and d) highest and lowest (100%-compliance)/2 observations set as missing.

Each of these missingness patterns correspond to a hypothetical scenario in an ESM study. The MCAR pattern 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). When there is a block of missing data present, all missing observations follow each other, there is a block of non-missing values and missing values. The start and the endpoint of the missing block do not depend on the intensity of the emotion. This can correspond to a situation where a participant misses a series of beeps because they are attending a social event. For patterns c) and d), the missingness is dependent on the value of the process itself. Pattern c) represent a situation in which a participant does not respond to an ESM measure of a positive mood because they are not feeling well, while pattern d) represents a situation where a participant misses an ESM beep when they either do not feel well enough, or they feel too good to respond to their phone beeping.

It can be expected that the different missingness patterns will differ in their effects on the simulation outcomes (estimation bias, standard error, power etc.). This is because with identical proportion in missing data, datasets with different missingness patterns will have different proportions of effective observation-pairs (i.e., proportion of timepoints for which both the observation at t and the observation at t-1 are not missing) used to estimate the autoregressive effect. Figure 1 illustrates the four different missingness patterns on the same ESM time-series.

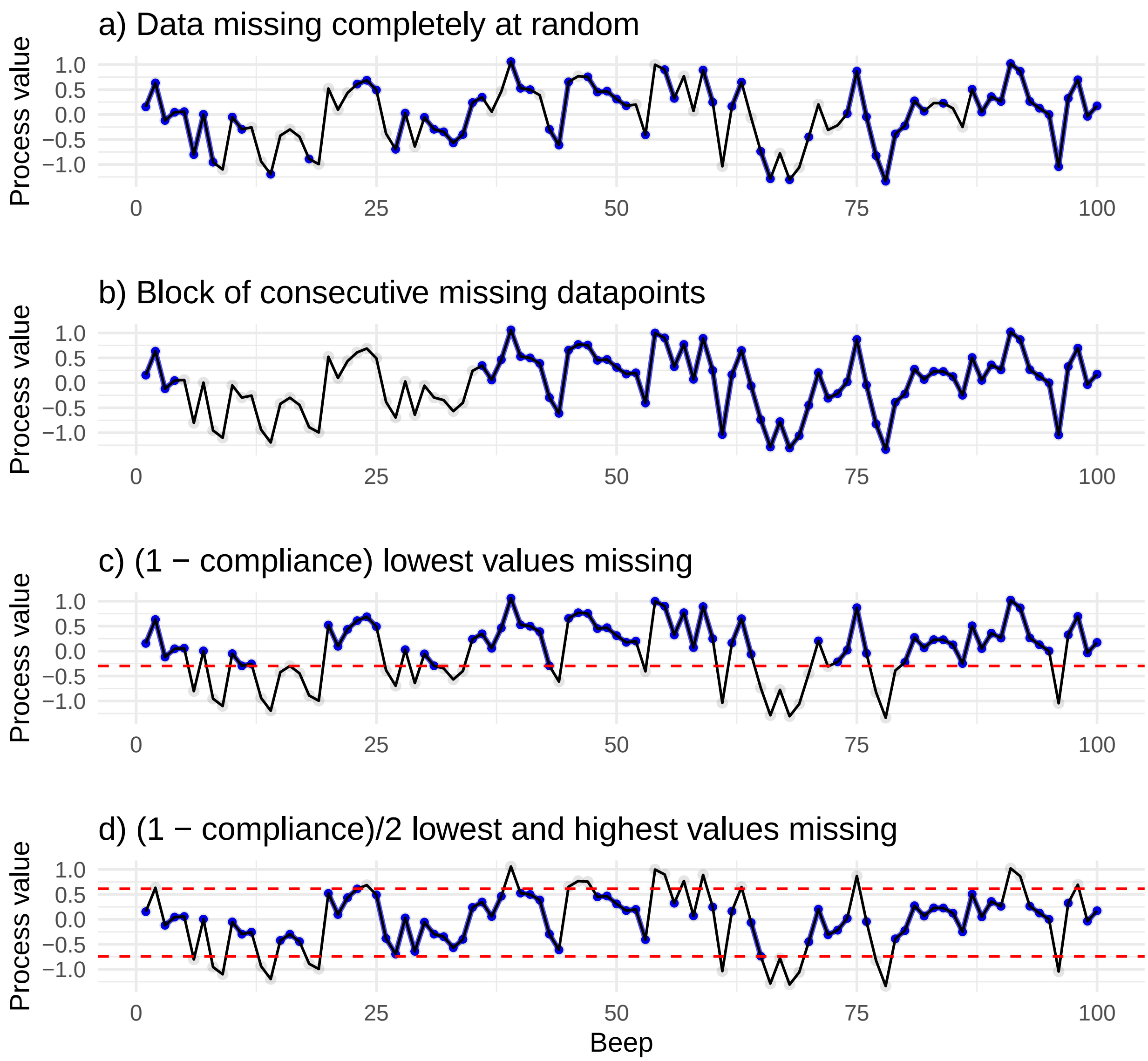

Figure 1: Illustration of the four different missingness pattern used in the simulation study. The blue dots represent observed datapoints, while the light gray dots represent the missing values.

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 was used to estimate the model.

The predictor (lagged) variable was person-mean centered. 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 Hamaker & Muthén (2020).

TODO: Describe person-mean centering more in detail

Simulation outcomes. Estimation bias (MSE), 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 real (simulated) fixed autoregressive effect β_{10} and the estimated fixed autoregressive effect $\hat{\beta_{10}}$ in each simulation run. As such, the dataset with estimation bias contained 1,000 rows per simulation condition.

Standard error (SE) and statistical power were calculated for each simulation condition (i.e., 1 row per condition). TODO: Describe how SE was calculated. Statistical power was calculated as the proportion of simulation runs (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 runs that converged successfully.

The bias in the estimation of the person-mean of the time-series was computed as the average difference between the real process mean μ_i (3) and the observed person-mean $\hat{\mu_i}$ (computed after the missing data were introduced).

$$\mu_i = \beta_{00} + \nu_{0i}/1 - (\beta_{10} + \nu 1i) \tag{3}$$

Manipulated simulation parameters. Apart from the four different missingness patterns (described above), the following parameters were manipulated in bothSimulation A and Simulation B: The number of participants of each simulated study (N), the number of ESM observations within each participant (T.obs), the compliance rate (i.e., the proportion of timepoints that are not missing for each participant), and the simulated autoregressive effect.

Furthermore, in Simulation B, the variance of the random autoregressive effects was varied. The values of the manipulated variables were set considering realistic research questions in psychological research. The values of the manipulated variables for both studies are reported in Table 1.

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 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 present thesis was written using 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 the appendix (TODO).

Outcome: Estimation bias (MSE)

ANOVA. We used a $4 \times 2 \times 3 \times 4 \times 3$ factorial Type I ANOVA (with estimation bias as an outcome and number of participants, number of timepoints per participant, missingness type, compliance, and the simulated fixed autoregressive effect) 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 any difference 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 3.

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 timepoints 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 are 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 severe when only the most extreme values (both at one side and at both sides) are missing. SOMETHING ABOUT JANNE'S PAPER! Additionally, the underestimation of the fixed slopes becomes more severe as the compliance gets lower.

The average estimation bias when compliance is 0.8 (which is very close to the average compliance of ESM studies in psychology) is -0.13. As a consequence, many estimates of inertia in

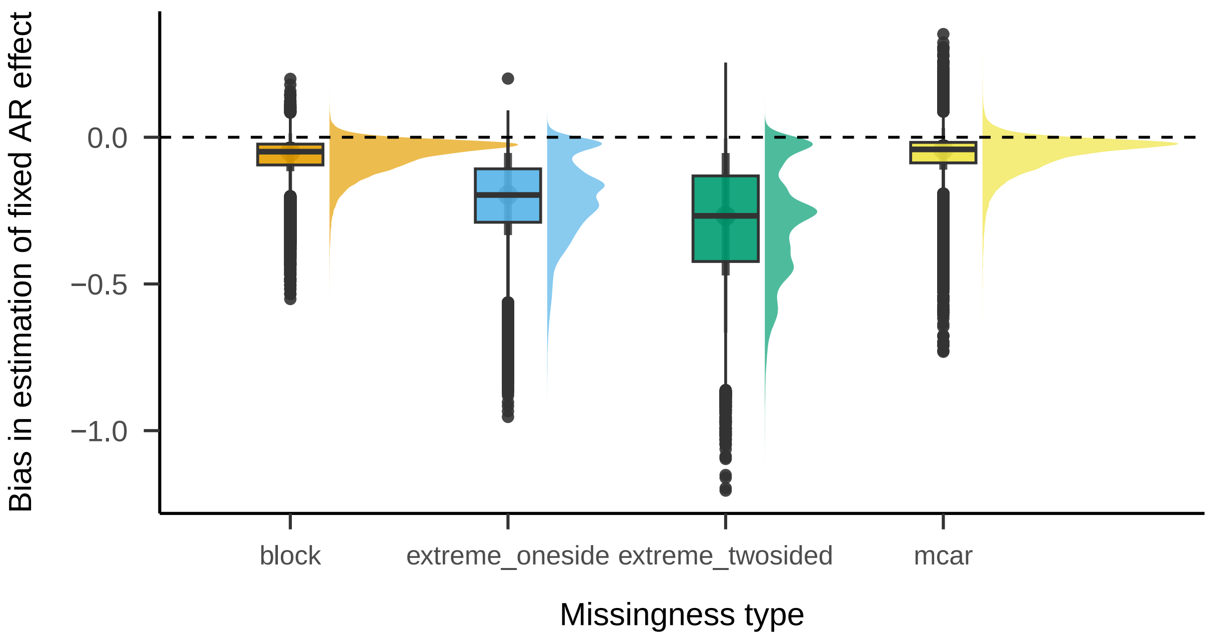

Figure 2: The effect of compliance on the bias in estimation of the fixed slopes.

psychological research could be seriously downward biased. Furthermore, the estimates are slightly biased even when compliance is 1 (i. e., there are no missing data; average bias: -0.04). This is in line with the findings about estimation biased caused by person-mean centering in multilevel autoregressive models (Hamaker & Grasman, 2015).

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 at both sides missing), the average estimation bias was -0.48). Given that the average simulated fixed slope was 0.5, these results imply that even rather large autocorrelations can be estimated as close to 0 in studies with the combination of low compliance and a non-random missingness pattern. At the same time, the results about data MCAR and missing in blocks are encouraging. Even in a low-compliance (0.4) condition, the average estimation bias was -0.08 for the former and -0.09 for the latter.

The average estimation bias for all combinations of missingness type and compliance (averaged over the different values of the number of participants, timepoints per participant and simulated fixed slope) is reported in Table 4.

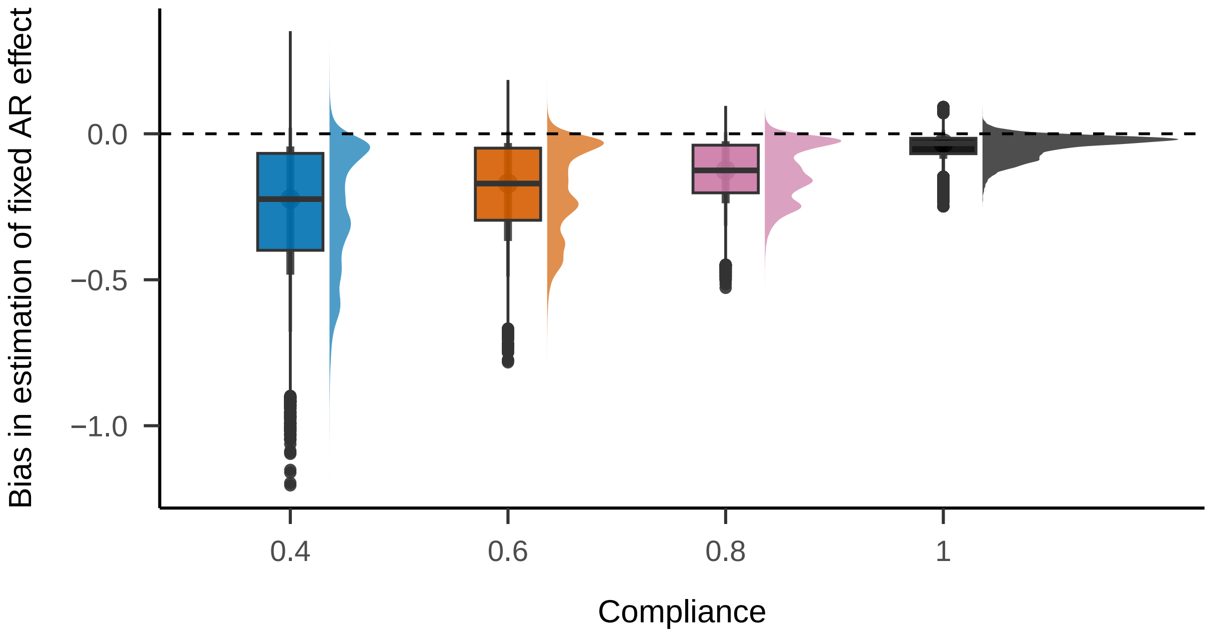

Figure 3: The effect of missingness type on the bias in estimation of the fixed slopes.

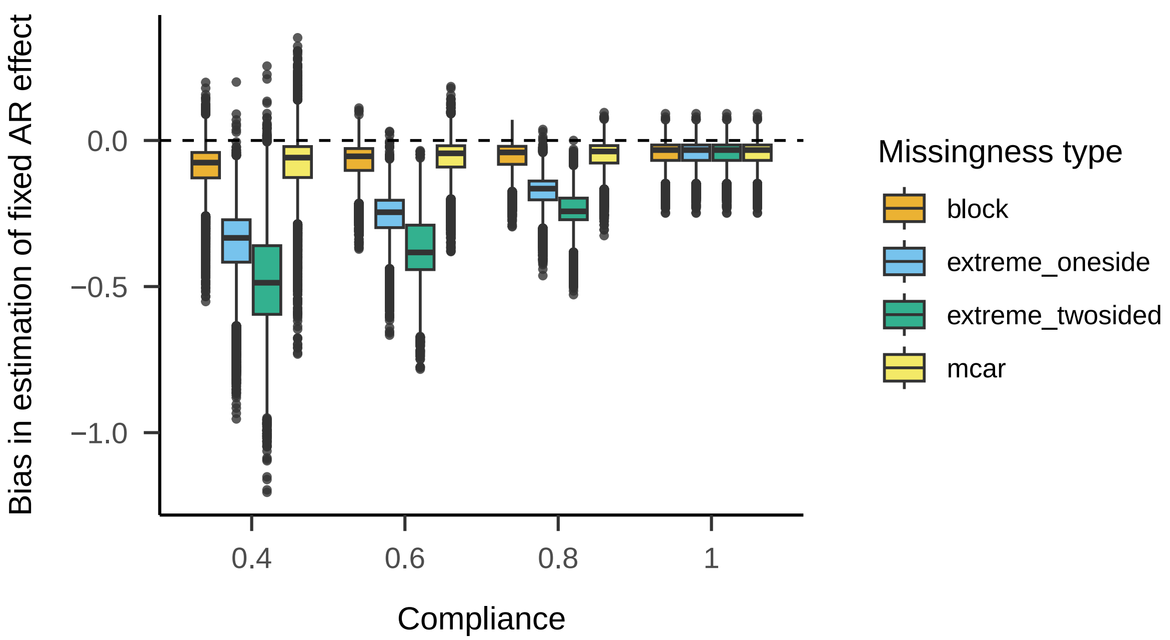

Figure 4: The effect of the interaction between missingness type and compliance on the bias in estimation of the fixed slopes.

Table 3: 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 4: 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 5: Simulation A. Average standard error in the estimation of the fixed slope for each combination of number of participants, number of timepoints/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, timepoints per participant and compliance are reported in Table \ref{tab:tab_aov_se}.

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 one row per simulation condition (288 rows in total). A a 4 × 2 × 3 × 4 × 3 factorial Type I ANOVA was used to analyse the data. The full ANOVA results and effect sizes are reported in Table DS.

The main effects of the number of participants ($\omega_p^2 = 0.68$), number of timepoints per participant ($\omega_p^2 = 0.68$) and compliance ($\omega_p^2 = 0.66$) crossed the cut-off for effect size.

Additionally, the interaction between the number of timepoints per participant and compliance ($\omega_p^2 = 0.28$), number of participants and timepoints per participants ($\omega_p^2 = 0.07$), and between the number of participants and compliance ($\omega_p^2 = 0.22$) was found.

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

Bias in the estimation of the fixed slope and bias in the estimation of the person-specific means. TODO.

Table 6: 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	

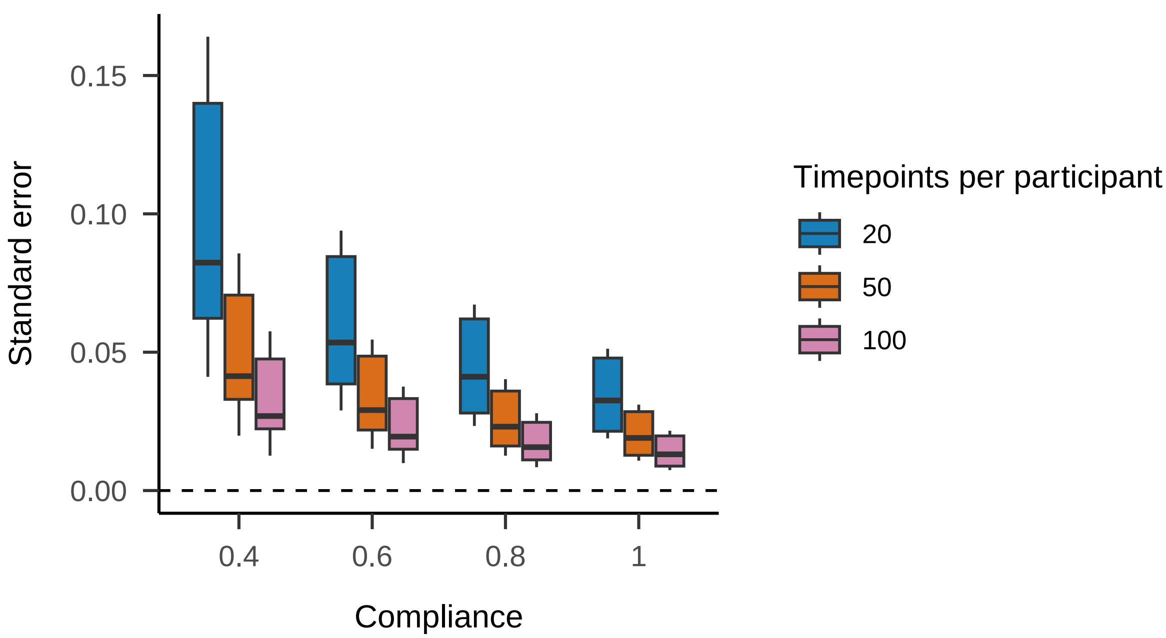

Figure 5: The effect of the interaction between number of timepoints and compliance on standard error of estimation of the fixed slopes. Simulation A.

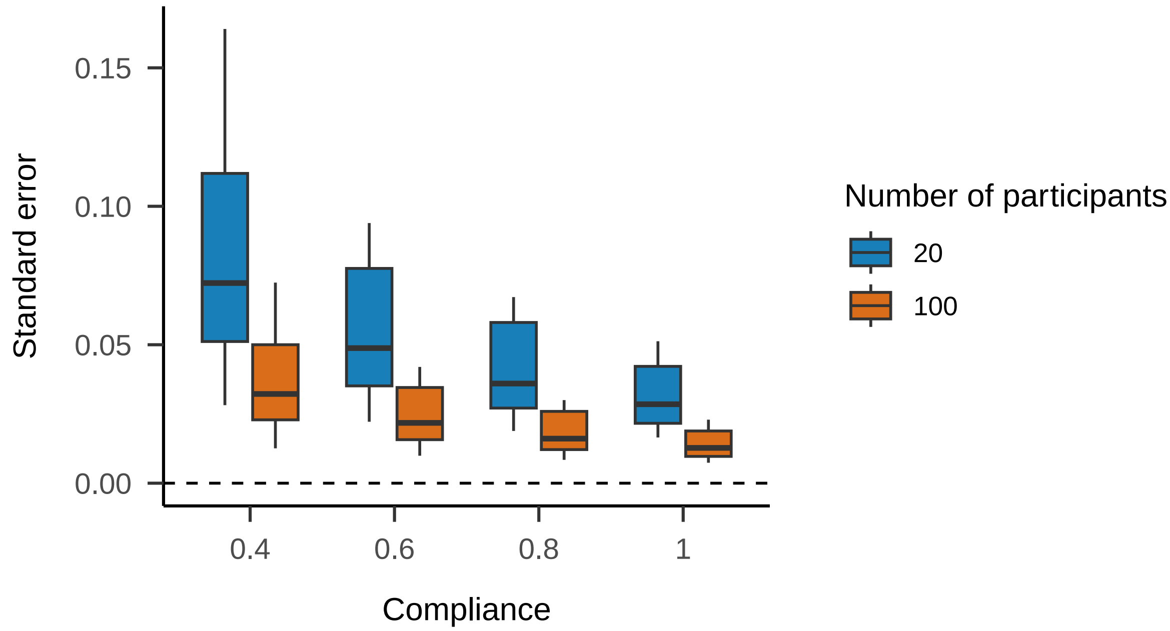

Figure 6: The effect of the interaction between number of participants and compliance on standard error of estimation of the fixed slopes. Simulation A.

Outcome: Statistical power

Descriptive statistics. The statistical power for each combination of the manupulated parameters is reported in Table (BIG TABLE TODO). As an illustration, the effect of compliance, missingness type, the number of participants and the number of timepoints 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 is worth pointing out in the plot: 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. 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. Aa $4 \times 2 \times 3 \times 4 \times 3$ factorial Type I ANOVA was used to analyse the effect of the manipulated parameters (288 conditions in total) on statistical power. The results are

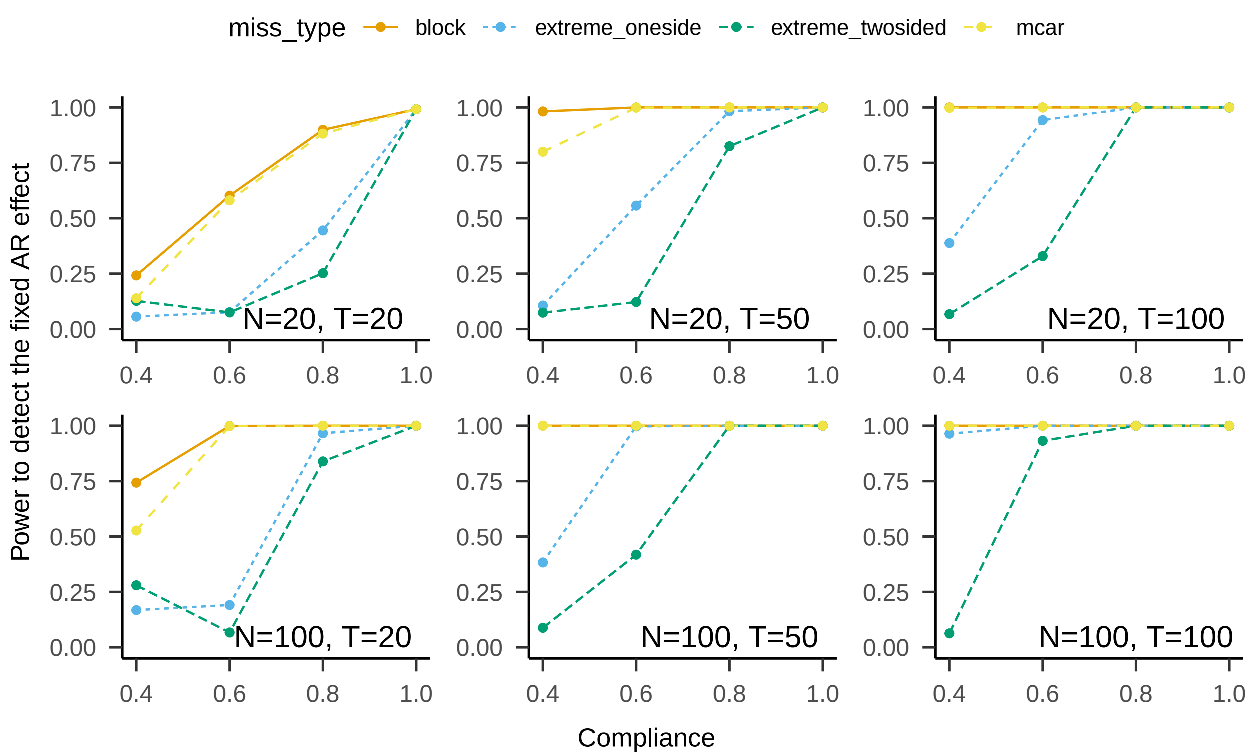

Figure 7: Simulation A. Statistical power to detect the fixed slope for all combinations of compliance, missingness type, number of participants and timepoints per participant when the simulated fixed slope is 0.3.

Table 7: 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	

reported in Table 7.

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

Simulation B

Discussion

- The effect of compliance on all 3 outcomes is quite high, and it largely depends on the missingness pattern.

- while the previous research suggested that increasing T is very effective (and this is true), it is very important to keep compliance as high as possible
- consequences of misestimation: could this be one of the reasons why the inertia of NA/PA does not predict psychopathology/well-being (Dejonckheere, 2019)?
- in studies estimating inertia, there might be some individual differences in inertia estimates that are due to different missingness patterns, not due to changes in real inertia (i. e., two participants with the same real inertia but different missingness patterns could have different inertia estimates)
- also, this makes computing power a bit less straightforward (ideally, to get realistic estimates of power, one should have an idea about what the average compliance could be and what missingness patterns might be present in the data - of course, in real life, we can expect there to be a mix of missing data patterns in every dataset).
- of course, not entirely realistic - real life data will likely have a mixture of missing data patterns at both the within- and between-person level.

Limitations

- focused on normally distributed processes. a normal dist can be assumed for positive affective proceses, however, negative affective processes are usually heavily right-skewed in the general population (Haslbeck et al., 2022).

- o We assumed that the analysed time-series is measured without any error; this is not the case in real-world research Schuurman & Hamaker (2019) . Also, this unreliability can lead to further attenuation of the estimated parameters (Wenzel & Brose, 2022).

o The person-mean centering was carried out using observed means. However, using latent means might be more appropriate; they describe this in (Gistelinck et al., 2021).

- o Only used very specific combination of parameters, BUT I provide a framework for people to reproduce the study

- Limitation-code: Although I took steps to maximise reproducibility of the simulations (shared all code and results, made a package containing the custom functions, used packages here() and renv(), and reported the sessionInfo for every simulation), I also used quite a lot of packages with many dependencies, which might be detrimental to reproducibility

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

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

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., & 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

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

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

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

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

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/

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

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