With Great Responsibility Comes Great Power:

Overcoming Methodological Challenges for Conducting Power Analysis in Intensive Longitudinal Research

Ginette Lafit

Research Group of Quantitative Psychology and Individual Differences

Center for Contextual Psychiatry

KU Leuven Open Science Study Day

May 3, 2021

KU LEUVEN

The most Frequently Asked Question to a Statistician

- How big does my sample size have to be?

- Sample size determines how much information is present to derive reliable conclusions

- Goal of a study is to test a hypothesis: a criterion to select the sample size is statistical power

- Lakens, *Sample size justification*
- Sormani, *The most frequently asked question to a statistician: The sample size*

Sample Size Planning

Design

Study is designed to yield sufficient statistical power to test specific hypotheses concerning parameters in the statistical model

Power

Probability of correctly rejecting the null hypothesis, when there is an effect of a certain size

- Cohen, *Statistical Power Analysis for the Behavioral Sciences*

Replicability of Scientific Findings

Low-powered studies

Reduce the probability of detecting a true effect (produce more false negatives)

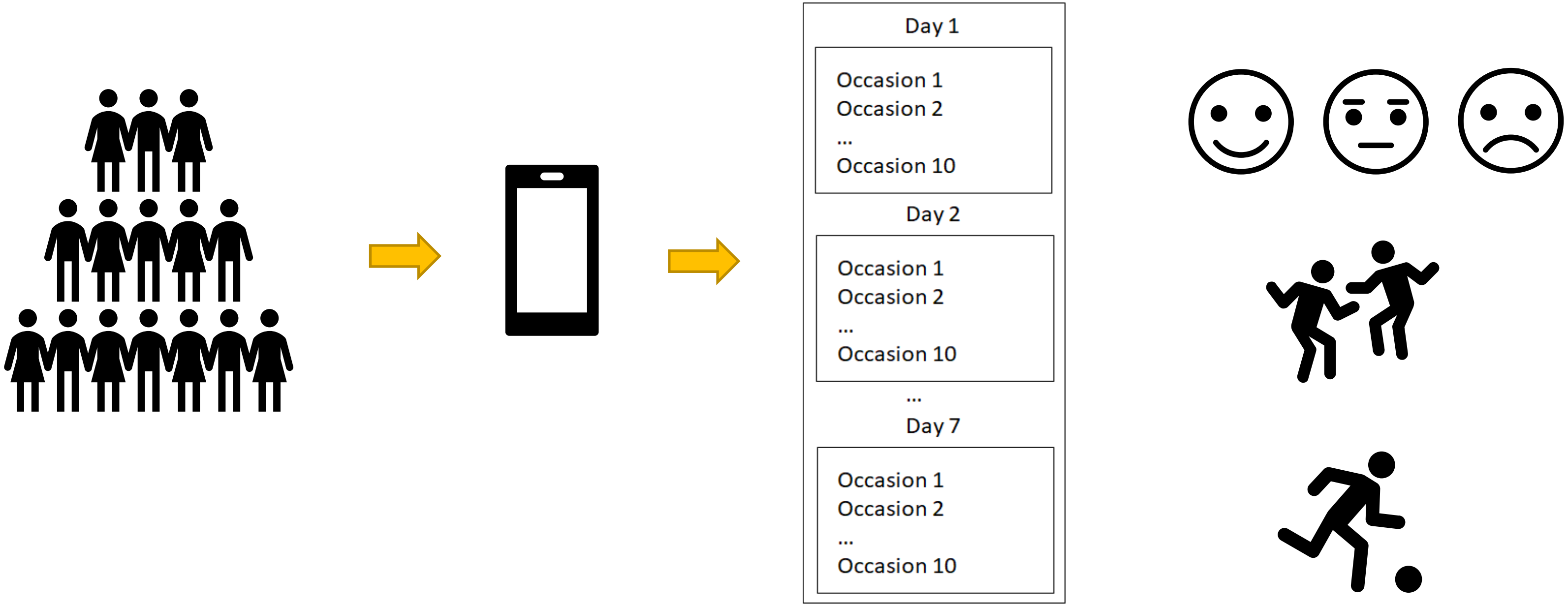
Produce gross overestimation of the effect-sizes

Produce inefficient and wasteful research

➢ Button et al., *Power failure: why small sample size undermines the reliability of neuroscience*

Opening the Black Box of Daily Life

ESM: study dynamic aspects of daily psychological functioning within individuals



- Myin-Germeys et al., *Experience sampling research in psychopathology: Opening the black box of daily life*

(Reliably) Opening the Black Box of Daily life

Goal Investigate how complex psychological processes evolve dynamically across time within single individuals

Design Intensive longitudinal designs: individuals are repeatedly measured

Planning Number of participants assuming a fixed number of (at least approximately) equidistant observations within individuals

Methods Multilevel regression models that account for temporal dependencies

➢ Lafit et al., *Selection of the Number of Participants in Intensive Longitudinal Studies*

Challenges for Conducting Power Analysis in Intensive Longitudinal Research

Design	Data have a multilevel structure: repeated observations are nested within individuals
Serial dependency	Observations are closer in time in comparison with traditional longitudinal designs
Method	Analytical formulations are not available for complex multilevel models

Power Analyses in Intensive Longitudinal Studies

- Analytical approach: formulas where the sample size is a function of the effect of interest, the standard deviation, and the test statistic

- Simulation-based approach

hypothesized population model

$$y_{it} = \beta_0 + \beta_1^* x_{it} + \nu_i x_{it} + \varepsilon_{it}$$

the effect of interest: β_1^* $(\neq 0)$

$$y_{it} = \beta_0 + \beta_1 x_{it} + \nu_i x_{it} + \varepsilon_{it} \Rightarrow H_0: \beta_1 = 0$$

$$y_{it} = \beta_0 + \beta_1 x_{it} + \nu_i x_{it} + \varepsilon_{it} \Rightarrow H_0: \beta_1 = 0$$

$$\vdots$$

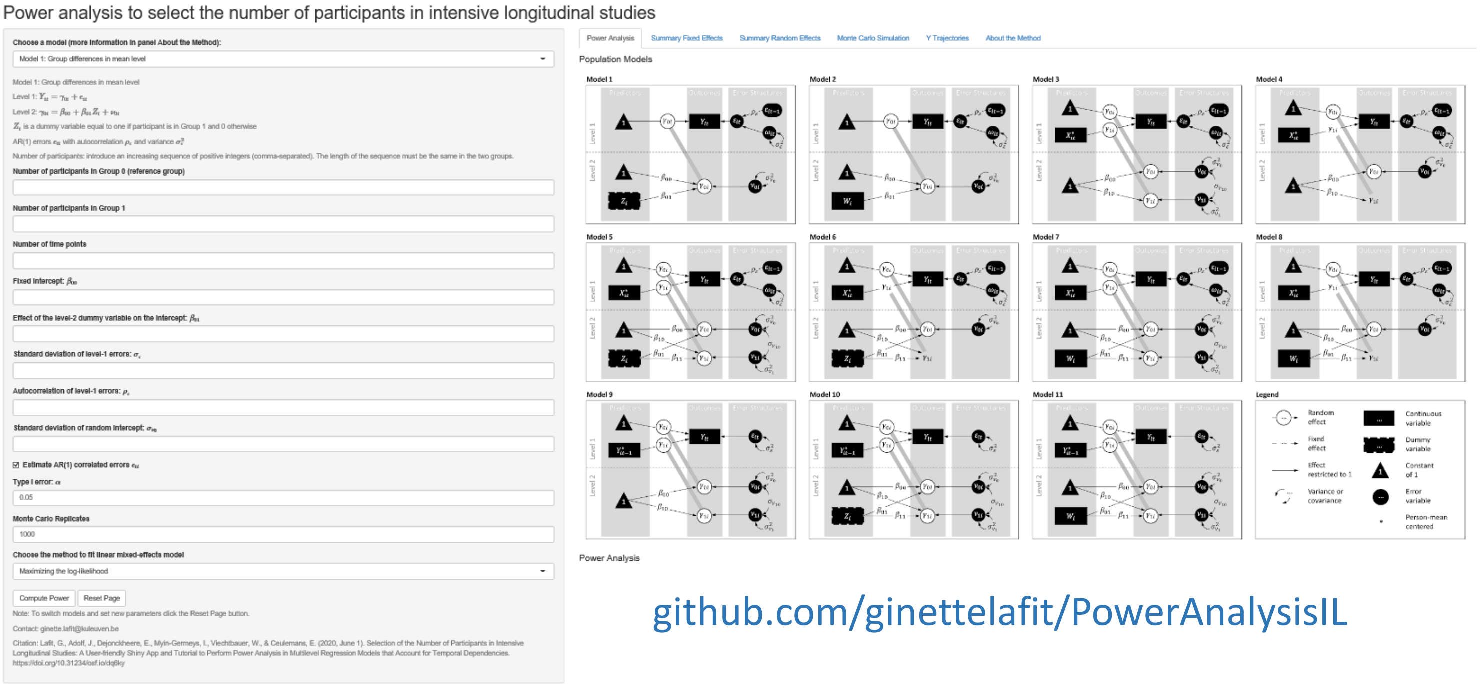
$$y_{it} = \beta_0 + \beta_1 x_{it} + \nu_i x_{it} + \varepsilon_{it} \Rightarrow H_0: \beta_1 = 0$$

Synthetic data sets

Power estimate: # of times that H_0 is rejected

➢ Bolger, *Power analysis for intensive longitudinal studies*

PowerAnalysisIL a Shiny App to Perform Power Analysis in Intensive Longitudinal Studies

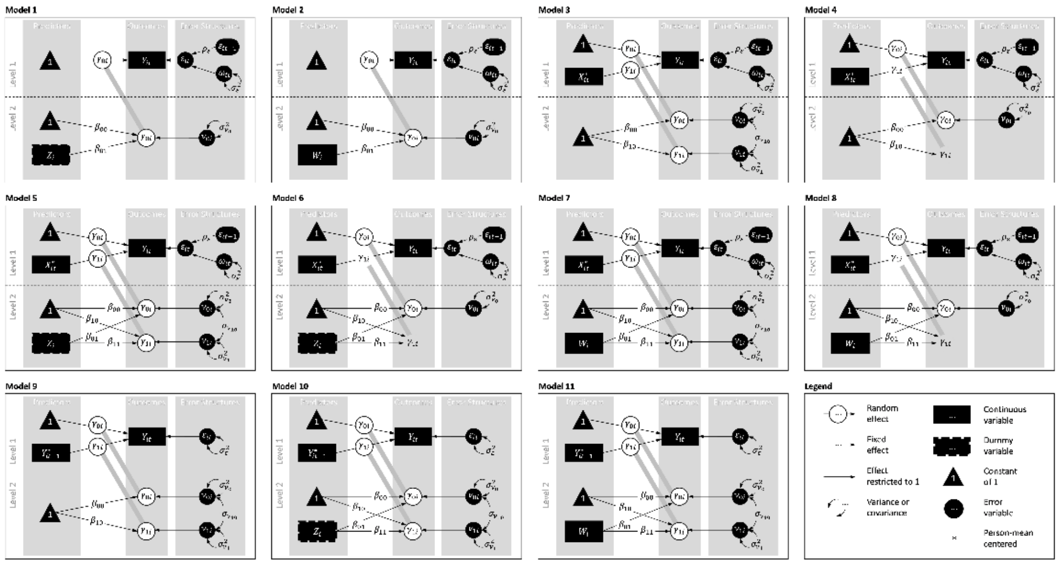


github.com/ginettelafit/PowerAnalysisIL

Lafit, G., Adolf, J., Dejonckheere, E., Germeys, I., Viechtbauer, W., Ceulemans, E. (2021). 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. *Advances in Methods and Practices in Psychological Science.*

Population Model of Interest

- Group differences in mean level
- Effect of a level-2 continuous predictor on the mean level
- Effect of a level-1 continuous predictor
- Group differences in the effect of a level-1 continuous predictor
- Cross-level interaction between two continuous predictors
- Multilevel autoregressive models



Illustration

- Research questions: momentary anhedonia predicts momentary negative affect in individuals who have been diagnosed with MDD

- The data will be collected using an IL design including 70 measurement occasions per individual

- How many participants do we need to involve?

Step 1: select the model and set the sample size

Choose a model (more information in panel About the Method):

Model 3: Effect of a level-1 continuous predictor (random slope)

Model 3: Effect of a level-1 continuous predictor (random slope)

Level 1: $Y_{it} = \gamma_{0i} + \gamma_{1i} X_{it} + \epsilon_{it}$

Level 2: $\gamma_{0i} = \beta_{00} + \nu_{0i}$

Level 2: $\gamma_{1i} = \beta_{10} + \nu_{1i}$

AR(1) errors ϵ_{it} with autocorrelation ρ_ϵ and variance σ_ϵ^2

Number of participants: introduce an increasing sequence of positive integers (comma-separated).

Number of participants

15,20,30,40,60,100

Number of time points

70

Parameter values of the model of interest: we use the results from previous studies examining the same hypothesis or data from a pilot study

Step 2: set the value of the model parameters

Fixed intercept: β_{00}

42.90

Fixed slope: β_{10}

0.13

Standard deviation of level-1 errors: σ_ϵ

12.00

Autocorrelation of level-1 errors: ρ_ϵ

0.43

Standard deviation of random intercept: σ_{ν_0}

15.00

Standard deviation of random slope: σ_{ν_1}

0.12

Correlation between the random intercept and random slope: $\rho_{\nu_{01}}$

0.003

Mean of time-varying variable X:

51.70

Standard deviation of time-varying variable X:

23.70

☑ Person mean centering X_{it} using the individual mean

☑ Estimate AR(1) correlated errors ϵ_{it}

Type I error: α

0.05

Monte Carlo Replicates

1000

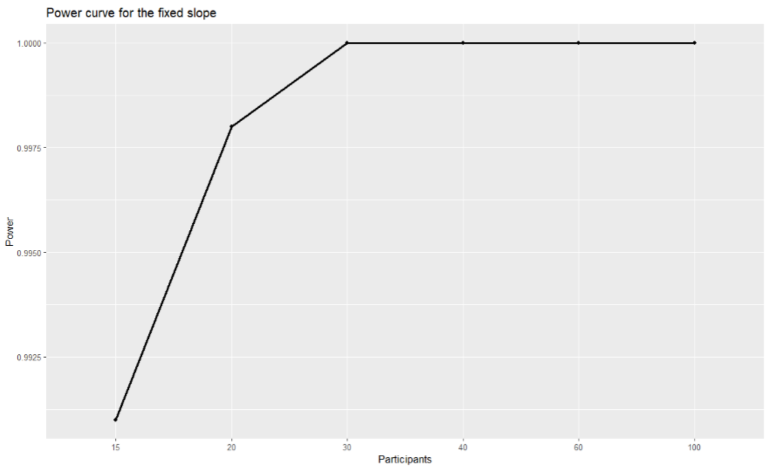
Choose the method to fit linear mixed-effects model

Maximizing the restricted log-likelihood

Estimate Computational Time | Compute Power | Reset Page

Step 3: inspect simulation results

Power curve



Summary fixed effects

	True value	Mean	Std.error	Bias	(1-alpha)% Coverage	Power
Fixed intercept N 15	42.90	42.9567	0.1179	0.0567	0.939	1.000
Fixed intercept N 20	42.90	42.9296	0.1059	0.0296	0.934	1.000
Fixed intercept N 30	42.90	42.8422	0.0900	-0.0578	0.941	1.000
Fixed intercept N 40	42.90	42.9985	0.0754	0.0985	0.946	1.000
Fixed intercept N 60	42.90	42.9181	0.0596	0.0181	0.953	1.000
Fixed intercept N 100	42.90	42.9109	0.0471	0.0109	0.942	1.000
Fixed slope N 15	0.14	0.1413	0.0011	0.0013	0.925	0.991
Fixed slope N 20	0.14	0.1385	0.0009	-0.0015	0.940	0.998
Fixed slope N 30	0.14	0.1407	0.0008	0.0007	0.932	1.000
Fixed slope N 40	0.14	0.1392	0.0006	-0.0008	0.941	1.000
Fixed slope N 60	0.14	0.1407	0.0005	0.0007	0.944	1.000
Fixed slope N 100	0.14	0.1402	0.0004	0.0002	0.946	1.000

Limitations and Future Research

Prior information	Accommodating uncertainty about the hypothesized model parameters: sensitivity analysis
Optimal design	Selection of the numbers of measurement occasions and persons
Models Precision Prediction	Extend the approach to include models/hypothesis Sample size planning for statistical accuracy Sample size planning & prediction

- Albers & Lakens, *When power analyses based on pilot data are biased: Inaccurate effect size estimators and follow-up bias*
- Moerbeek, *The effects of the number of cohorts, degree of overlap among cohorts, and frequency of observation on power in accelerated longitudinal designs*
- Maxwell & Rausch, *Sample size planning for statistical power and accuracy in parameter estimation*

Thanks to my amazing collaborators!

Janne Adolf

Egon Dejonckheere

Inez Myin-Germeys

Wolfgang Viechtbauer

Eva Ceulemans

Thanks for your attention!

Open access materials:
Shiny app: https://github.com/ginettelafit/PowerAnalysisIL
OSF page of the project: https://osf.io/vguey/
Tutorial: https://doi.org/10.1177/2515245920978738

Ginette Lafit
ginette.lafit@kuleuven.be
@GinetteLafit
github.com/ginettelafit

References

- Albers, C., & Lakens, D. (2018). When power analyses based on pilot data are biased: Inaccurate effect size estimators and follow-up bias. *Journal of Experimental Social Psychology*, 74, 187–195.
- Bolger, N. (2011). *Power analysis for intensive longitudinal studies*. In N. Bolger, G. Stadler, & J.-P. Laurenceau (Eds.), Handbook of research methods for studying daily life (pp. 285–301). Guilford Press.
- Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power Failure: Why Small Sample Size Undermines the Reliability of Neuroscience. *Nature Reviews Neuroscience* 14 (5): 365.
- Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Academic press.
- Lafit, G., Adolf, J., Dejonckheere, E., Germeys, I., Viechtbauer, W., Ceulemans, E. (2021). 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. *Advances in Methods and Practices in Psychological Science*.
- Lakens, D. (2021, January 4). Sample Size Justification. https://doi.org/10.31234/osf.io/9d3yf
- Maxwell, S. E., Kelley, K., & Rausch, J. R. (2008). Sample size planning for statistical power and accuracy in parameter estimation. *Annual Review of Psychology*, 59, 537–563.
- Moerbeek, M. (2011). The effects of the number of cohorts, degree of overlap among cohorts, and frequency of observation on power in accelerated longitudinal designs. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 7(1), 11–24.
- Myin-Germeys, I., Oorschot, M., Collip, D., Lataster, J., Delespaul, P., & Van Os, J. (2009). Experience sampling research in psychopathology: opening the black box of daily life. *Psychological Medicine*, 39(9), 1533.
- Sormani, M. P. (2017). The most frequently asked question to a statistician: The sample size. *Multiple Sclerosis Journal*, 23(5), 644-646.