Experimental design for multi-level data: Improving our approach to power analysis using Monte Carlo simulation-based parameter recovery estimation

Chadwick, S.[1], & Davies, R.[1]

International Multilevel Conference 2019

[1]Department of Psychology, Lancaster University, United Kingdom

What's the point of research?

My research question:

"Are ratings of comprehension predictive of assessed comprehension?"

"'Will my study answer my research question?' is the most fundamental question a researcher can ask when designing a study"

(Johnson et al., 2015, p. 133)

Adequately designing a study

Question:

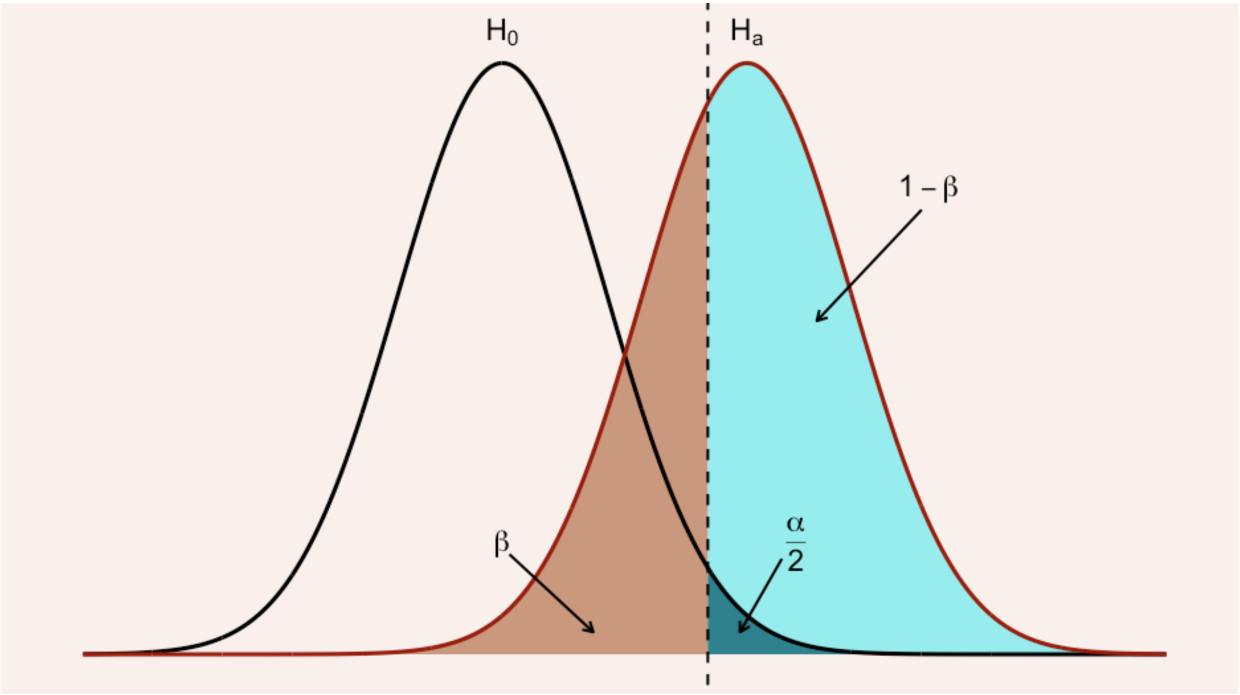
- Will my study be adequately powered to detect an effect of interest?

Answer:

- Do power analysis

What is power analysis?

Power = P(correctly reject H0)

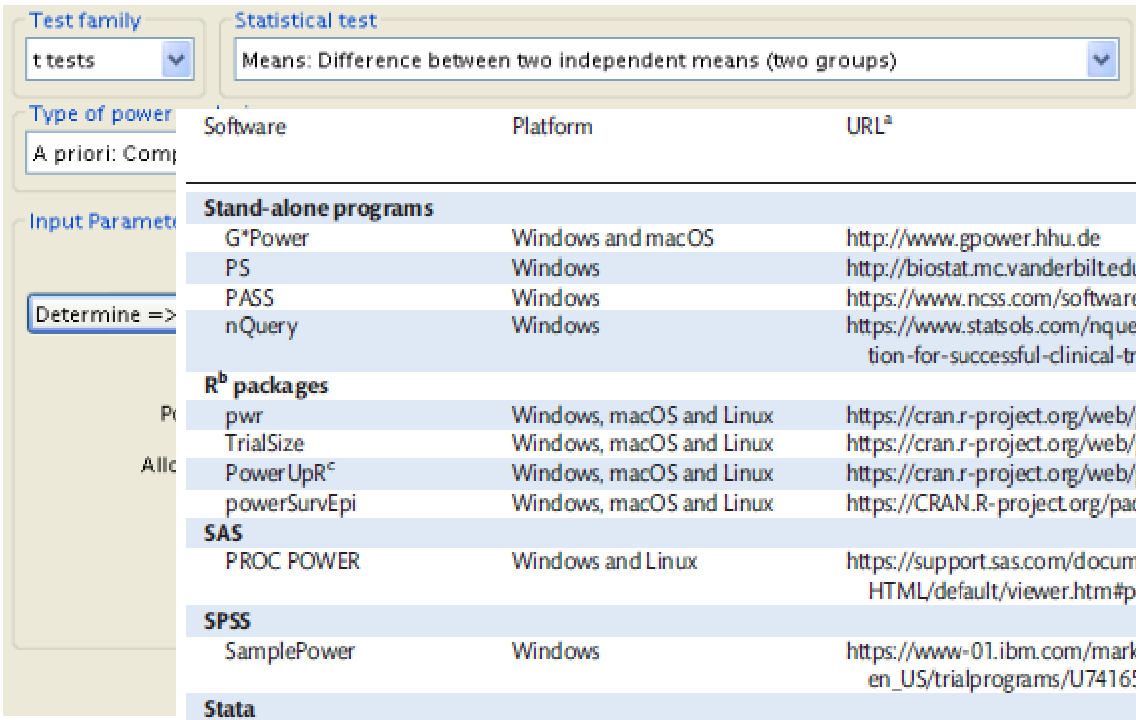


“Do power analysis”

1) Formulaic / analytic method

2) Simulation-based method

Formulaic / analytic approach



Software	Platform	URL[a]	Freely available?
Stand-alone programs			
G*Power	Windows and macOS	http://www.gpower.hhu.de	Yes
PS	Windows	http://biostat.mc.vanderbilt.edu/wiki/Main/PowerSampleSize	Yes
PASS	Windows	https://www.ncss.com/software/pass	No
nQuery	Windows	https://www.statsols.com/nquery-sample-size-and-power-calculation-for-successful-clinical-trials	No
R[b] packages			
pwr	Windows, macOS and Linux	https://cran.r-project.org/web/packages/pwr	Yes
TrialSize	Windows, macOS and Linux	https://cran.r-project.org/web/packages/TrialSize	Yes
PowerUpR[c]	Windows, macOS and Linux	https://cran.r-project.org/web/packages/PowerUpR	Yes
powerSurvEpi	Windows, macOS and Linux	https://CRAN.R-project.org/package=powerSurvEpi	Yes
SAS			
PROC POWER	Windows and Linux	https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#power_toc.htm	No
SPSS			
SamplePower	Windows	https://www-01.ibm.com/marketing/iwm/iwmdocs/tnd/data/web/en_US/trialprograms/U741655I36057W80.html	No
Stata			
power	Windows, macOS and Linux	https://www.stata.com/features/power-and-sample-size/	No
Microsoft Excel			
PowerUp[c]		http://www.causalevaluation.org/power-analysis.html	Yes[d]
Specialist simulation software			
IcebergSim	Windows	http://icebergsim.software.informer.com/versions/	Yes
FACTS	Windows	https://www.berryconsultants.com/software/	No
Clinical trial simulation	Windows and Linux	http://www.biopharmnet.com/innovation/trial_simulation/cts1.php	Yes[e]

(Hickey et al., 2018, Table 3)

Formulaic / analytic limitations

"advances [in specialist modelling techniques] have not been matched by the development of analytic formulae for sample size calculations under such models"

(Landau & Stahl, 2013, p. 325)

Off-the-shelf formula assumptions are rarely met

Bespoke closed-form equations can be designed, but can be difficult to define and inflexible

Simulation-based approach

Simulation-based power analyses can handle any design

Simulation-based power analyses can handle any data-generating mechanism

Separates the data-generating model from the analytic model (Landau & Stahl, 2013)

In ‘n’ steps or less

1. Define the data-generating mechanism
2. Simulate many datasets
3. Perform an analysis on each dataset
4. Calculate performance

(Arnold et al., 2011; Johnson et al., 2015; Kontopantelis et al., 2016; Landau & Stahl, 2013)

1. Define the data-generating mechanism

- Outcome distribution
- Sources of variance
- Covariate distributions
- Effect distributions

Making assumptions of the generative model

What's sensible is defensible:

A plausible range of parameter values should, with careful consideration and transparent justification, be assumed based on knowledge of the topic and study design.

2-4. Simulation software options

SIMR (R; Green & MacLeod, 2016)

MLPowSim (MLwiN; Browne, Golalizadeh, & Parker, 2009)

POWERSIM (Stata; Luedicke, 2013)

Idpower (Stata; Kontopantelis, 2018)

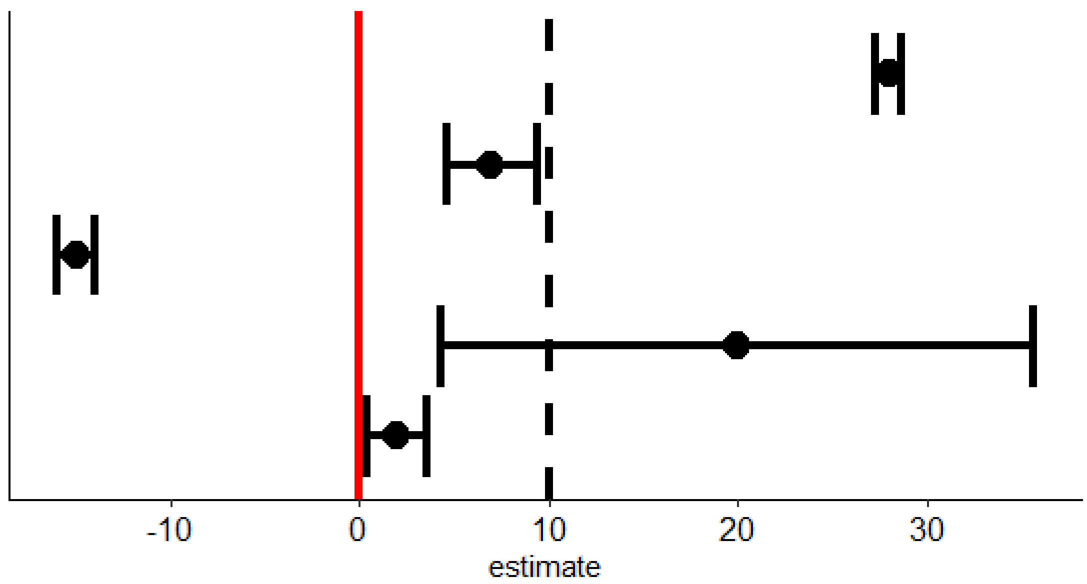
Power analysis is flawed

Lancaster University

"a narrow emphasis on statistical significance is placed as the primary focus of study design"

(Gelman & Carlin, 2014, p. 641)

Example: β = 10



Power analysis can be broader

Conventional power (NHST) is one form of power, and power analysis can be thought of more broadly, in terms of different goals.

(Gelman & Carlin, 2014; Hickey et al., 2018; Johnson et al., 2015; Kruschke, 2014; Landau & Stahl; 2013)

Reframe the question

Question:

~~- Will my study be adequately powered to detect an effect of interest?~~

- Will my study be adequately designed to accurately recover an effect of interest?

Answer:

~~- Do power analysis~~

- Do parameter recovery analysis

“Do parameter recovery analysis”

1. Define the data-generating mechanism
2. Simulate many datasets
3. Perform an analysis on each dataset
4. Calculate performance[1]

[1] in a more informative way

Defining parameter recovery

Two types of precision:

1. Estimate
2. Uncertainty

Parameter is recovered if:

The estimate is within a specified range <u>and</u> the associated uncertainty is within a specified range

Estimate precision

E.g. Estimate precision of β +/- 25%

Where β is the effect of interest

$\beta = 10$

$7.5 \geq \hat{\beta} \leq 12.5$

Frequentist error precision

E.g. Error precision of $SE_{\hat{\beta}} \leq 1.5$

Where $SE_{\hat{\beta}}$ is the estimated standard error associated with $\hat{\beta}$

$$95\%\ CI = \hat{\beta} +/- SE_{\hat{\beta}} * 1.96$$

$\hat{\beta} = 7.5$, 95% CI = [4.56, 10.44]

$\hat{\beta} = 10$, 95% CI = [7.06, 12.94]

$\hat{\beta} = 12.5$, 95% CI = [9.56, 15.44]

Bayesian error precision

E.g. Error precision of $\hat{\beta}$ +/- 3…

Contained within the credible intervals or posterior HDI

$$95\%\ CI_{\hat{\beta}} = [\hat{\beta} - 3, \hat{\beta} + 3]$$

$$80\%\ HDI_{\hat{\beta}} = [\hat{\beta} - 3, \hat{\beta} + 3]$$

Example: My study

1. Define the data-generating mechanism

$Y_{ijkl} = \text{Bernoulli}(\Theta_{ijkl})$

$$\Theta_{ijkl} = \beta_{0ijkl} + \beta_{1i}x_{1i} + \beta_{2i}x_{2i} + \beta_{3i}x_{3i} + \beta_{4i}x_{4i}$$

x_1 = Comprehension ability
x_2 = Vocabulary
x_3 = Topic familiarity
x_4 = Rated comprehension

i = participant
j = text
k = question
l = observation

$\beta_{0ijkl} = \Upsilon_0 + u_{0il} + u_{0ij} + u_{0ik}$
$u_{0il} = N(u_{0i}, \sigma^2)$
$u_{0i} = N(0, \sigma^2)$
…

$\beta_{1i} = N(\mu, \sigma^2)$
$\beta_{2i} = N(\mu, \sigma^2)$
…

Example: My study

2. Simulate many datasets

Texts: 5, 10, 15

Participants: 50-500

Example: My study

3. Perform an analysis on the datasets

clmm(count ~ (1|participant) + (1|text) + comprehension.ability + vocabulary.score + topic.familiarity + rated.comprehension)

Example: My study

4. Calculate performance

Estimate precision: 50% of β

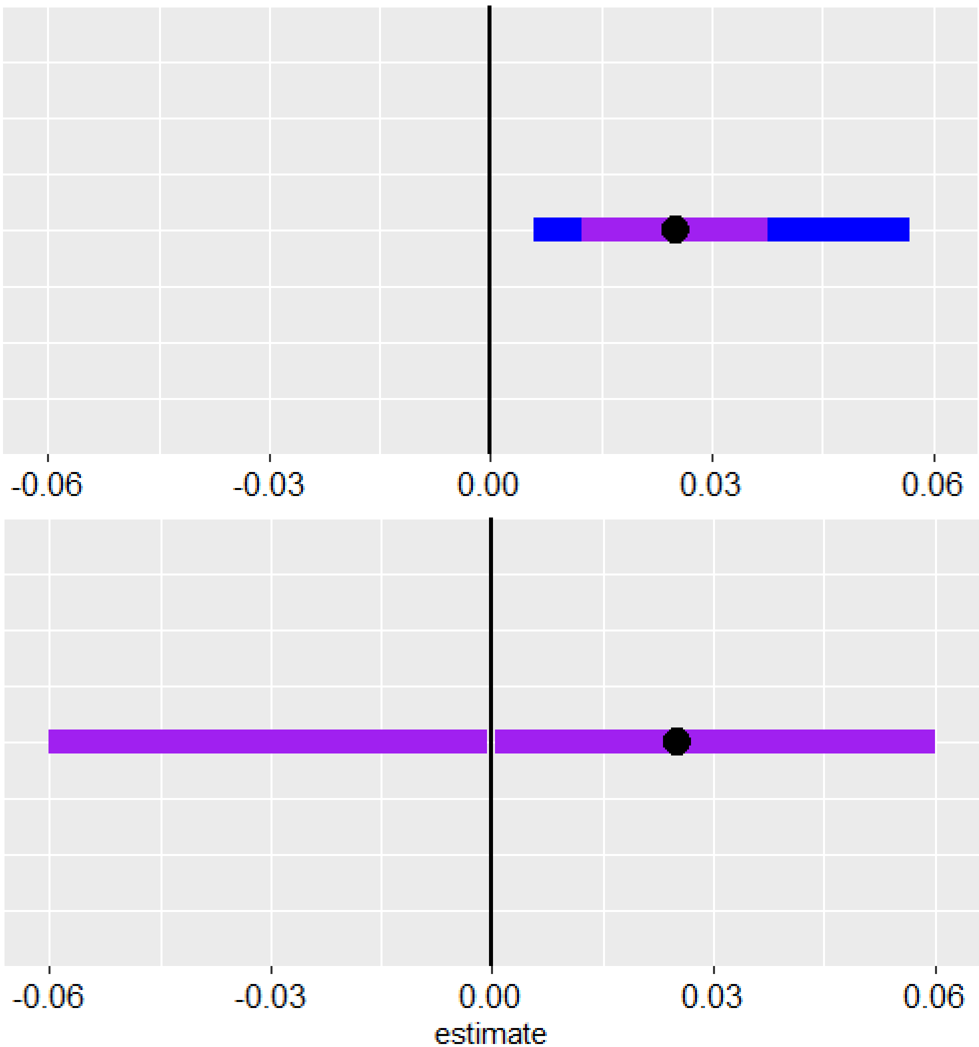
$$0.5\beta \geq \hat{\beta} \leq 1.5\beta$$

Error precision: 50% of $\hat{\beta}$

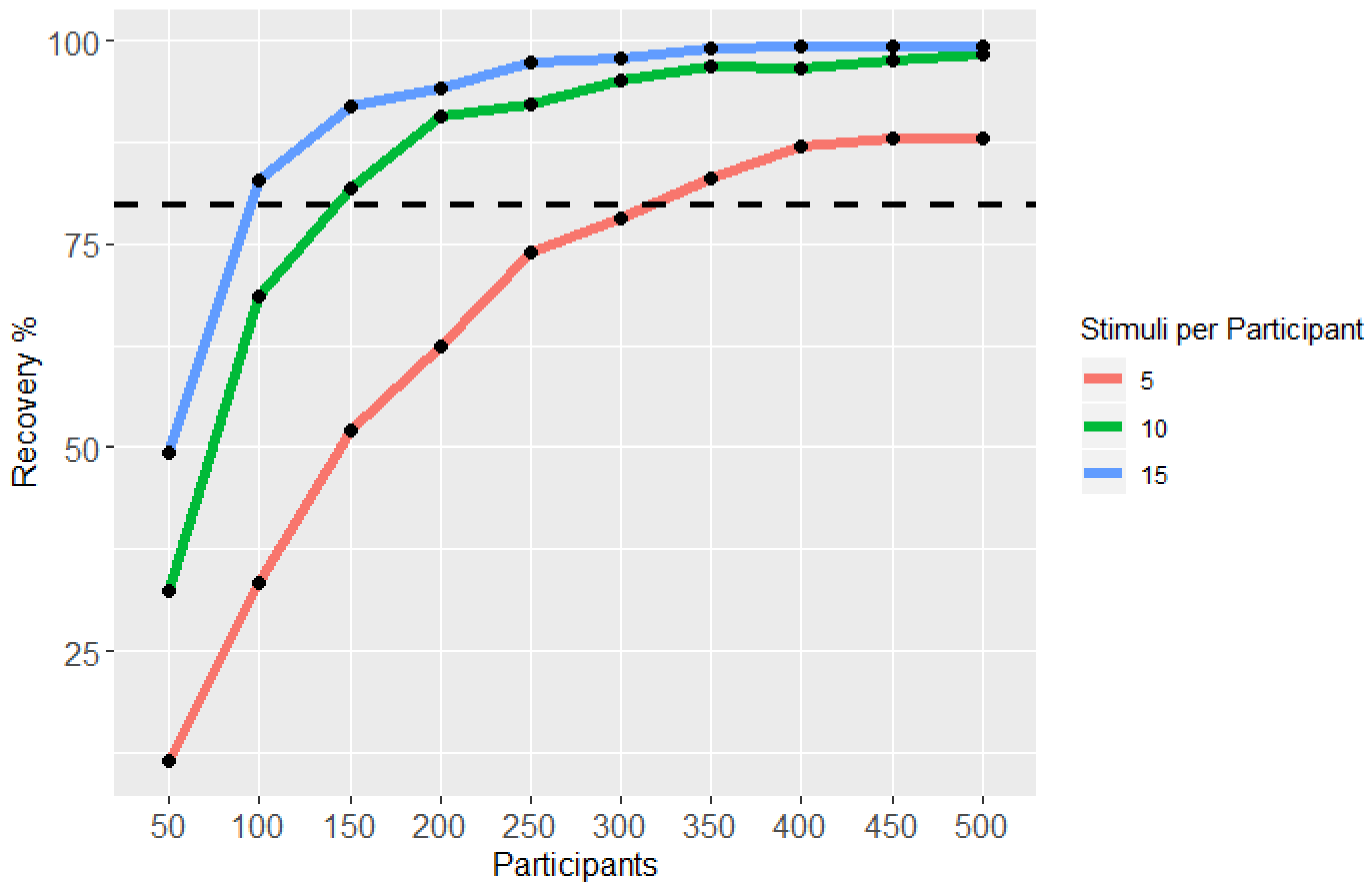
$$95\%\ UCI_{\hat{\beta}} \geq 0.5*0.5\beta \text{ and } 95\%\ LCI_{\hat{\beta}} \leq 0.5*1.5\beta$$

Example: My study

$\beta_{4\mu}$:
0.025



Example: Result



Caution

Assumptions on parameters

Choosing parameter estimates is difficult

Time

Convergence

Available code

Lancaster University

R package on GitHub – chaddlewick/spr (under development)

```
observedvariables = as.list(c(participant = "rep(1:20, each = 40)",
                qriscore = "rnorm(participant, 10, 2)",
                hlvascore = "rnorm(participant, 8, 0.5)",
                texts = "rep(1:10, times = 20, each = 4)",
                question = "rep(1:800)"))
effectvariables = as.list(c(intercept = "0.15",
                bparticipant = "rnorm(participant, mean=0, sd=0.4)",
                bqriscore = "rnorm(participant, 0.025, 0.001)",
                bhlvascore = "rnorm(participant, 0.02, 0.001)",
                btexts = "rnorm(texts, 0, 0.02)",
                bquestion = "rnorm(question, 0, 0.015)"))
outcomegeneration = as.list(c(outcome= "rbinom(observation, 1, dataset$py)",
                py = "dataset$intercept + dataset$bparticipant + dataset$bqriscore*dataset$qriscore +
                dataset$bhlvascore*dataset$hlvascore + dataset$btexts + dataset$bquestion"))
analyticmodel = "brm(outcome ~ (1|participant) + (1|texts) + qriscore + hlvascore, data=dataset, family =
                bernoulli(), cores = 2)"
```

References & Resources

Anderson, S.F., Kelley, K., & Maxwell, S.E. (2017). Sample-size planning for more accurate statistical power: A method adjusting sample effect sizes for publication bias and uncertainty. Association for Psychological Science, 28, 1547-1562. DOI: 10.1177/0956797617723724

Arnold, B.F., Hogan, D.R., Colford, J.M., & Hubbard, A.E. (2011). Simulation methods to estimate design power: An overview for applied research. BMC Medical Research Methodology, 11, 1-10. DOI: 10.1186/1471-2288-11-94

Browne, W.J., Golalizadeh, M., & Parker, R.M.A. (2009) - A Guide to Sample Size Calculations for Random Effect Models via Simulation and the MLPowSim Software Package. Retrieved March 2019, from http://www.bristol.ac.uk/cmm/software/mlpowsim/

Gelman, A., & Carlin, J. (2014). Beyond power calculations: Assessing type S (size) and type M (magnitude) errors. Association for Psychological Science, 9, 641-651. DOI: 10.1177/1745691614551642

Green, P., & MacLeod, C.J. (2016). SIMR: an R package for power analysis of generalized linear mixed models by simulation. Methods in Ecology and Evolution, 7, 493-498. DOI: 10.1111/2041-210X.12504

Hickey, L., Grant, S.W., Dunning, J., & Siepe, M. (2018). Statistical primer: Sample size and power calculations – why, when and how?

Johnson, P.C.D., Barry, S.J.E., Ferguson, H.M., & Müller, P. (2015). Power analysis for generalized linear mixed models in ecology and evolution. Methods in Ecology and Evolution, 6, 133-142. DOI: 10.1111/2041-210X.12306

Kontopantelis, E., Springate, D.A., Parisi, R., & Reeves, D. (2016). Simulation-based power calculations for mixed effects modelling: ipdpower in Stata. Journal of statistical software, 74, 1-25. DOI: 10.18637/jss.v074.i12

Kruschke, J.K. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. New York, NY: Academic Press

Kruschke, J.K. (2018). Rejecting or accepting parameter values in Bayesian estimation. Advances in Methods and Practices in Psychological Science, 1, 270-280. DOI: 10.1177/2515245918771304

Landau, S., & Stahl, D. (2013). Sample size and power calculations for medical studies by simulation when closed form expressions are not available. *Statistical Methods in Medical Research, 22*, 324-345. DOI: 10.1177/0962280212439578

Luedicke, J. (2013). Powersim: Simulation-based power analysis for linear and generalised linear models. 2013 Stata Conference, Stata Users Group.

Thank you

Do you have any questions or feedback?

s.chadwick4@Lancaster.ac.uk | @chaddlewick | github.com/chaddlewick