Sample Size Planning for MLM

PSYC 575

Winnie Tse, Mark Lai

University of Southern California

Updated: 2021-11-13

Week Learning Objectives

- Describe the importance of having sufficient sample size for scientific research
- Describe conceptually the steps for sample size planning: precision analysis and power analysis
- Perform power analysis for MLM using the PowerUpR application and the `simr` package
- Understand the effect of uncertainty in parameter values and explore alternative approaches for sample size planning

Why Sample Size?

Small Sample Size is a Problem Because . . .

Low power

Misleading and noisy results[1]

- When coupled with publication bias (statistical significance filter)[2] [3]

Nonreproducible findings

[1] See Maxwell (2004)

[2] See the graph on this blog post

[3] See also Vasishth et al. (2018)

Review: Sampling distributions

Test yourself! -- Week 13 Quiz (ungraded)

What is the null distribution?

- Suppose we examine the effect of a therapy on eating disorder
- We test against the null hypothesis $H_0 : \gamma_{01} = 0$, where γ_{01} is the fixed effect of the therapy on eating disorder

What is the alternative distribution?

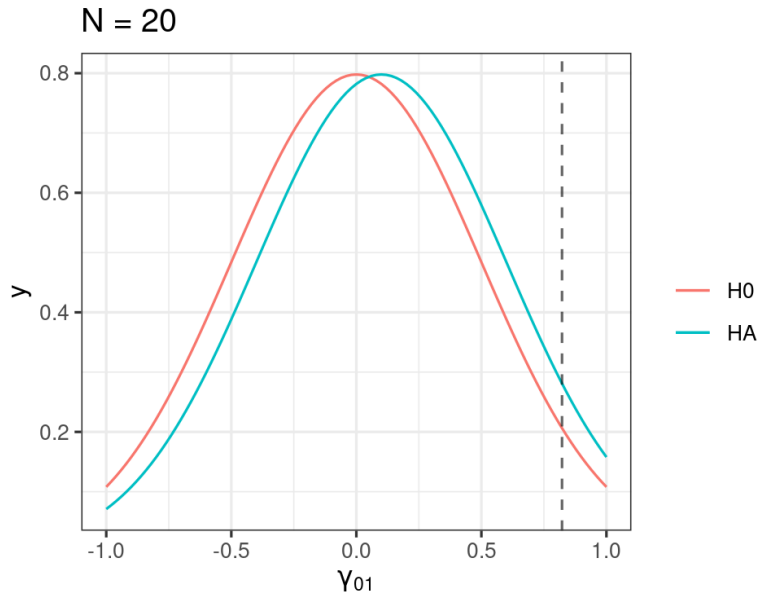
- Assume that the true effect of this therapy is $\gamma_{01} = .1$

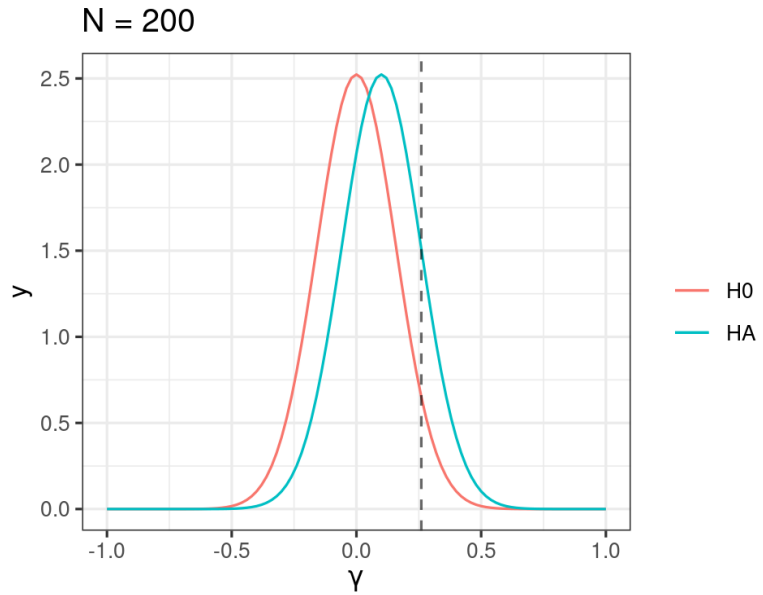
Sampling Distribution as a Function of Sample Size

Assume true effect is $\gamma_{01} = 0.10$

Let's say

- when $N = 20, p < .05$ when $\hat{\gamma} \geq 0.82$
- when $N = 200, p < .05$ when $\hat{\gamma} \geq 0.26$





Steps for Sample Size Planning

Steps for Sample Size Planning

1. Write down your model equations
2. List out all parameters in the model
3. Determine if you want to achieve a desired level of

a. Power, or

b. Precision

Step 1: Write down model equations

Group-based therapy for eating disorder (cluster-randomized trial)

Step 1: Write down model equations

Group-based therapy for eating disorder (cluster-randomized trial)

Level-1

$$Y_{ij} = \beta_{0j} + \beta_{1j} X\_\mathrm{cmc}_{ij} + e_{ij}$$

$$e_{ij} \sim N(0, \sigma)$$

Level-2

$$\beta_{0j} = \gamma_{00} + \gamma_{01} W_j + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11} W_j + u_{1j}$$
$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_0^2 & \\ \tau_{01} & \tau_1^2 \end{bmatrix} \right)$$

- γ_{10}: X (purely level-1 with ICC = 0)
- γ_{01}: W (level-2)
- γ_{11}: $W \times X$ (cross-level interaction)

Step 2: List out all parameters

1. Fixed effects: $\gamma_{00}, \gamma_{01}, \gamma_{10}, \gamma_{11}$
2. Random effects: $\tau_0^2, \tau_1^2, \tau_{01}$
3. Number of clusters: J
4. Cluster size: n

Level-1

$$Y_{ij} = \beta_{0j} + \beta_{1j} X\_\text{cmc}_{ij} + e_{ij}$$

$$e_{ij} \sim N(0, \sigma)$$

Level-2

$$\beta_{0j} = \gamma_{00} + \gamma_{01} W_j + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11} W_j + u_{1j}$$
$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_0^2 & \\ \tau_{01} & \tau_1^2 \end{bmatrix} \right)$$

Standard Error and Precision Analysis

Sample Size and *SE*/Post. *SD*

In the previous graph, when $N = 20$, the sample estimate is likely to be anywhere between -0.4 and 0.6

$$SE \propto \frac{1}{\sqrt{N}}$$

One goal of sample size planning is to

> Have sufficient sample size to get precise (low *SE*) sample estimates of an effect

Analytic Formulas of *SE*

J = Number of clusters; n = Cluster size

- E.g., $J = 100$ schools; $n = 10$ students per school

Assuming $\tau_{01} = 0$

$$SE(\gamma_{01}) = \sqrt{\frac{1}{S_W^2}\left(\frac{\tau_0^2}{J} + \frac{\sigma^2}{Jn}\right)}$$

$$SE(\gamma_{10}) = \sqrt{\frac{\tau_1^2}{J} + \frac{\sigma^2}{JnS_X^2}}$$

$$SE(\gamma_{11}) = \sqrt{\frac{1}{S_W^2}\left(\frac{\tau_1^2}{J} + \frac{\sigma^2}{JnS_X^2}\right)}$$

Precision Analysis

Group-based therapy for eating disorder (cluster-randomized trial)

- Intervention at group level
- 10 participants per group
- Outcome standardized (i.e., *SD* = $\sqrt{\tau_0^2 + \sigma^2} = 1$)
  - γ = Cohen's d
- ICC = .3 (i.e., $\tau_0^2 = .3$)
- Goal: estimate J such that $SE(\gamma_{10}) \leq .1$
  - E.g., if we estimated the sample effect size to be $d = .25$, the 95% CI would be approximately [.05, .45].

Calculating J

When the predictor is binary (e.g., treatment-control), if half of the groups is in one condition, $S^2_W = 0.25$

- Otherwise, if 30% in one condition, $S^2_W = 0.3 \times 0.7$
- $\tau^2_0 = 0.3, \sigma^2 = 0.7, n = 10$

E.g., if $J = 30$

$$SE(\gamma_{01}) = \sqrt{\frac{1}{S^2_W}\left(\frac{\tau^2_0}{J} + \frac{\sigma^2}{Jn}\right)} = \sqrt{\frac{1}{0.25}\left(\frac{0.3}{30} + \frac{0.7}{(30)(10)}\right)} = 0.2221111$$

Keep trying, and you'll find ...

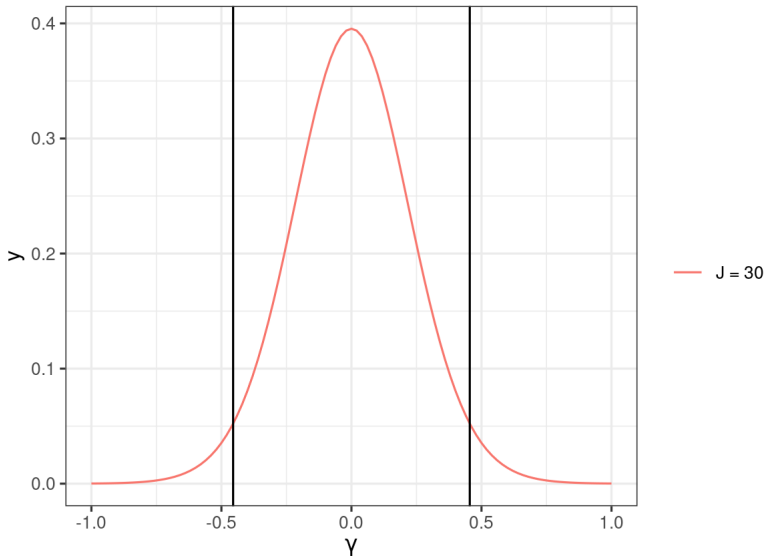
When J = 148, $SE(\gamma_{01}) = 0.1$

So you'll need 148 groups (74 treatment, 74 control)

Power Analysis

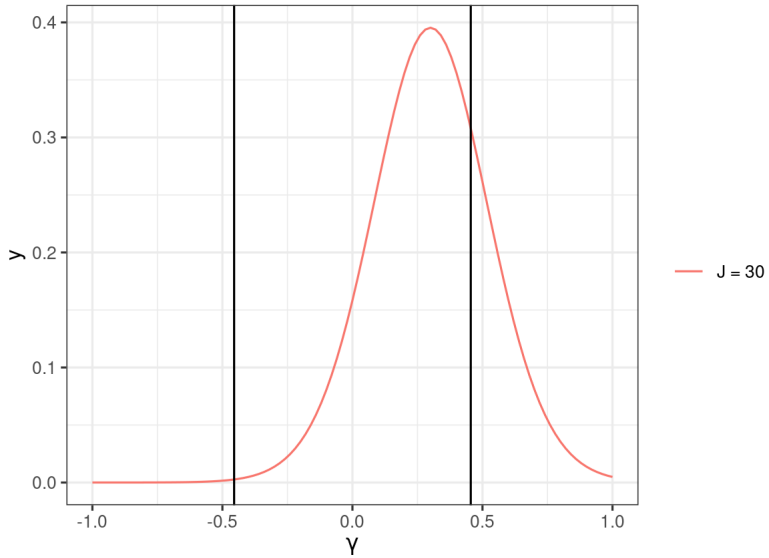
Two-tailed test, $\alpha = .05$

$H_0 : \gamma_{01} = 0$



Critical region: $\hat{\gamma}_{01} \leq -0.45$ or $\hat{\gamma}_{01} \geq 0.45$

$H_1 : \gamma_{01} = 0.3$

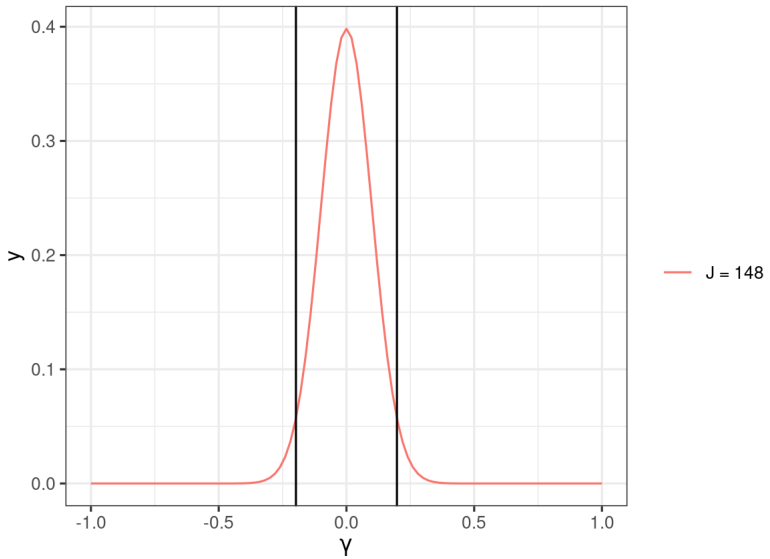


Power[1]
$\approx P(\hat{\gamma}_{01} \leq -0.45) + P(\hat{\gamma}_{01} \geq 0.45) = 0.2465731$

[1] In practice, we need to incorporate the sampling variability of the standard error as well, so this power calculation is only a rough approximation.

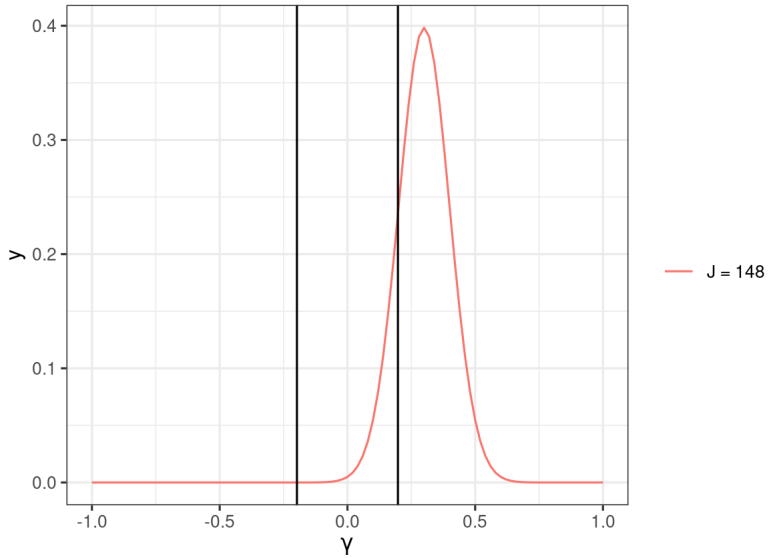
Two-tailed test, $\alpha = .05$

$H_0 : \gamma_{01} = 0$



Critical region: $\hat{\gamma}_{01} \leq -0.2$ or $\hat{\gamma}_{01} \geq 0.2$

$H_1 : \gamma_{01} = 0.3$



Power
$\approx P(\hat{\gamma}_{01} \leq -0.2) + P(\hat{\gamma}_{01} \geq 0.2) = 0.8461551$

Tools for Power Analysis

1. Stand-alone programs
    - Optimal Design
    - PinT
2. R packages
    - `simr`
3. Spreadsheet/Webapp
    - PowerUp!

See more discussion in Arend & Schäfer (2019)

PowerUpR Shiny App

https://powerupr.shinyapps.io/index/

Monte Carlo Simulation for Power Analysis

- Simulate a large number (e.g., R = 1,000) of data sets based on given effect size, ICC, etc
- Fit an MLM to each simulated data
- Power $\approx$ Proportion of times $p < \alpha$

See sample R code for using `simr`

Uncertainty in Parameter Values

Uncertainty in Parameter Values

In the PowerUpR demo, to calculate the number of clusters J need to achieve 80% power, we determined

1. Type I error rate = .05
2. Two tailed test = TRUE
3. `g2`, `r21`, `r22` = 0, as we did not include any covariates
4. `p` = .5, for a balanced design (half treatment, half control)

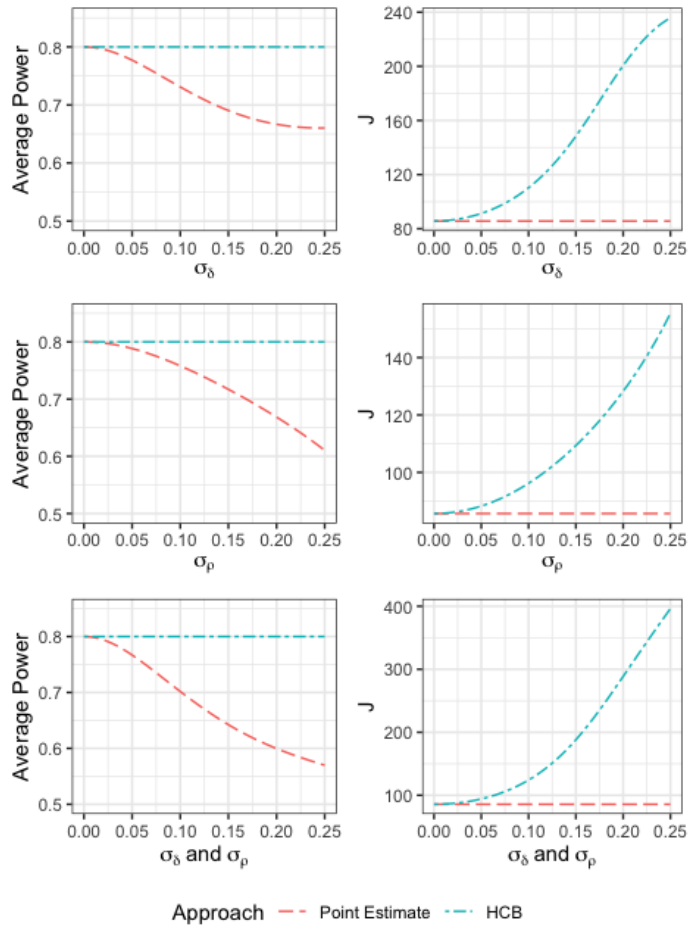
However, we need to guess the values of

1. Effect size = .3?
2. ICC = .3?

The Effect of Uncertainty in Power

Ignoring uncertainty

- The more uncertainty we have but ignore about a parameter value, the more power loss we will have in our study (red curve)
- Uncertainty in both effect size and ICC can further reduce our power
- The more uncertainty we have, the more samples we need to achieve 80% power

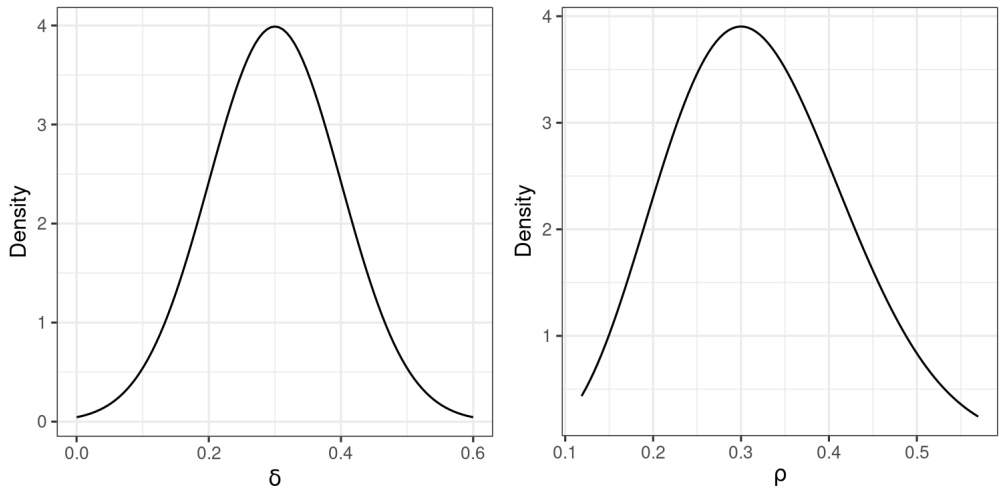


Hybrid Classical-Bayesian approach

- Incorporates uncertainty for sample size planning
- Instead of plugging in a point value of a guess, we can specify how much uncertainty we have (e.g., standard error of γ_{01} from a previous study)

$$\delta \sim N(.3, .1) \quad \rho \sim \text{Beta}(a, b)$$

- where a, b can be calculated by $\hat{\rho} = .3$ and $\sigma_\rho = .1$ (estimate and uncertainty about ρ)



hcbr Shiny App

http://winnie-wy-tse.shinyapps.io/hcb_shiny

Additional Notes on Power

- Increasing J usually leads to higher power than increasing n
- Balanced designs generally have higher power than unbalanced designs
- Larger sample size required for testing level-2 predictors
- Testing an interaction requires a much larger sample size
    - E.g., 16 times larger than for a main effect