

# Statistical Power Lab

Jessaca Spybrook

July 9, 2026

# Outline

- Software Options
- Designs
- Examples
- Exercises
- Writing up a Power Analysis

# Software Options

- Three widely used options
  - *CRT Power* (Borenstein & Hedges)
  - *Optimal Design Plus* (Raudenbush, Spybrook, Congdon, Liu, Martinez, Bloom, & Hill)
  - *PowerUp!* (Maynard & Dong)
- Plus a newer option 😊
  - The Generalizer plus Power (Tipton, Spybrook, & Miller)

# Software Options

- Key to keep in mind:
  - Much of the work is done before you get to the software 😊
  - All programs yield same results
  - Know the program you are using
    - Different programs require different parameters
    - Document the program and parameters

# Designs

- Be specific in naming a design
- Cluster randomized trial does NOT define the design
  - Consider three design elements
    - Number of levels
    - Level of random assignment
    - Level outcome data measured

|                             | Multisite Trial (MST) or Randomized Block Design (RBD) | 2-level Cluster Randomized Trial (2-level CRT) | 3-level Cluster Randomized Trial (3-level CRT) | 3-level Multisite Cluster Randomized Trial with Treatment at Level 2 (3-level MSCRT or 3-level RBD) | 4-level Multisite Cluster Randomized Trial with Treatment at Level 3 (4-level MSCRT or 4-level RBD) |
|-----------------------------|--------------------------------------------------------|------------------------------------------------|------------------------------------------------|-----------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------|
| Number of Levels            | 2                                                      | 2                                              | 3                                              | 3                                                                                                   | 4                                                                                                   |
| Level of Random Assignment  | 1                                                      | 2                                              | 3                                              | 2                                                                                                   | 3                                                                                                   |
| Level Outcome Data Measured | 1                                                      | 1                                              | 1                                              | 1                                                                                                   | 1                                                                                                   |
| Example of Level Structure  | L1: Students<br>L2: Schools                            | L1: Students<br>L2: Schools                    | L1: Students<br>L2: Teachers<br>L3: Schools    | L1: Students<br>L2: Teachers<br>L3: Schools                                                         | L1: Students<br>L2: Teachers<br>L3: Schools<br>L4: Districts                                        |

# Example 1

**Program:** A new math curriculum for 3<sup>rd</sup> graders. The curriculum is implemented at the school level. The researchers plan to randomly assign schools to the treatment or control condition.

**Research question:** Is the new math curriculum more effective than the traditional one?

Standardized math tests scores used as outcome measure. Based on previous research, 18% percent of the variation lies between schools. Researchers have access to last years test scores and expect the school level covariate explains 60 % of the variability in test scores.

# Example 1

- Scenario A
  - Researchers expect the new curriculum to boost test scores by 0.25 standard deviations. Assuming 90 students per school, how many schools do they need to detect an effect of 0.25 with power of 0.80? (Note: Assume teacher links and data are not available.)

# Example 1

- How many levels are in this study?
- What is the level of randomization?
- What is the level where outcome data is measured?
- What are good estimates of the intraclass correlation? Percent variance explained by covariate(s)?

<https://steppcenter.northwestern.edu/education-training/statistical-power-resources.html>

# Example 1

## Design Parameters

|                                                                             |            |
|-----------------------------------------------------------------------------|------------|
| Probability of Type I Error                                                 | $(\alpha)$ |
| One or Two-tailed Hypothesis Testing                                        |            |
| Effect Size                                                                 | $(\delta)$ |
| Number of Students (Level 1) per School (Level 2)                           | $(n)$      |
| Number of Schools (Level 2)                                                 | $(J)$      |
| Proportion of units randomly assigned to treatment condition                | $(P)$      |
| Proportion of variance in the outcome between schools (Level 2) ( $ICC_2$ ) | $(\rho_2)$ |
| Proportion of Level 1 variance explained by covariates                      | $(R_1^2)$  |
| Proportion of Level 2 variance explained by covariates                      | $(R_2^2)$  |
| Number of covariates at Level 2                                             | $(g_2)$    |

# Example 1

- Scenario B
  - Researchers have access to a total of 38 schools. What is the MDES? (Note: Assume teacher links and data are not available.)

# Example 2

**Program:** The following will be repeated across districts. 4 schools will be randomly assigned to either the new 3<sup>rd</sup> grade math curriculum or the regular curriculum. 100 students will be tested in each school.

**Research question:** Is the new math curriculum more effective than the traditional one?

Standardized math tests scores used as outcome measure. Based on previous research, within districts, 12% percent of the variation lies between schools within districts and 8% of the variance lies between districts. A school level covariate explains 48 % of the variability in test scores.

# Example 2

- Scenario A
  - Assuming 16 districts are willing to participate, what is the MDES?

# Example 2

- How many levels are in this study?
- What is the level of randomization?
- What is the level where outcome data is measured?
- Given that there is blocking:
  - How are blocks being treated – as fixed or random effects?  
If random effects, what is the variance of treatment effect across blocks/sites?

# Effect Size Variability

- Weiss et al. (2017)
  - Multisite trials
  - Education outcomes early childhood thru postsecondary, other outcomes
  - Variability of average treatment effect, 0 to 0.35 SD units
- Thinking through Interpretation

Approximate Intervals around true treatment effect of 0.20 (assuming normality)

$$0.20 \pm 1.96\sqrt{SD}$$

| ESV  | SD   | Lower Bound | Upper Bound | Span |
|------|------|-------------|-------------|------|
| 0.01 | 0.10 | 0           | 0.4         | 0.4  |
| 0.03 | 0.17 | -0.14       | 0.54        | 0.68 |
| 0.05 | 0.22 | -0.24       | 0.64        | 0.88 |
| 0.10 | 0.32 | -0.44       | 0.84        | 1.28 |

# Example 2

## Design Parameters (Random effects case)

|                                                                                                                                         |            |
|-----------------------------------------------------------------------------------------------------------------------------------------|------------|
| Probability of Type I Error                                                                                                             | $(\alpha)$ |
| One or Two-tailed Hypothesis Testing                                                                                                    |            |
| Effect Size                                                                                                                             | $(\delta)$ |
| Number of Students (Level 1) per School (Level 2)                                                                                       | $(n)$      |
| Number of Schools (Level 2) per District (Level 3)                                                                                      | $(J)$      |
| Number of Districts (Level 3)                                                                                                           | $(K)$      |
| Proportion of units randomly assigned to treatment condition                                                                            | $(P)$      |
| Proportion of variance in the outcome between schools (Level 2) ( $ICC_2$ )                                                             | $(\rho_2)$ |
| Proportion of variance in the outcome between districts (Level 3) ( $ICC_3$ )                                                           | $(\rho_3)$ |
| Effect size variability as the ratio of the treatment effect variance between districts (Level 3) to the total variance in the outcome. | $(esv_3)$  |
| Proportion of Level 1 variance explained by covariates                                                                                  | $(R_1^2)$  |
| Proportion of Level 2 variance explained by covariates                                                                                  | $(R_2^2)$  |
| Proportion of treatment effect variance among districts (Level 3) explained by the districts (Level 3) covariates                       | $(R_3^2)$  |
| Number of covariates at Level 3                                                                                                         | $(g_3)$    |

# Directions for Exercises

For each of the exercises, please do the following:

1. Name the specific type of CRT.
2. Identify the sample size at each level (if known).
3. Estimate the design parameters using the specific resource noted. Document your assumptions, e.g. state, population, covariate set.
4. Conduct the power analysis using *The Generalizer plus Power*.
5. Document the results of the power analysis.

# Exercise 1

Suppose a group of researchers want to examine the effectiveness of a new reading curriculum. They plan to randomly assign 24 schools to the treatment and 24 schools to the control. In each school, they plan to test 80 5<sup>th</sup> graders. Teacher data is not available. The outcome of interest is student reading achievement. What is the MDES?

Use the [\*Intraclass Correlation Database\*](#) to estimate the unknown design parameters. Select the state and covariate set of your choice.

# Exercise 2

Suppose a group of researchers want to examine the effectiveness of a new math curriculum. Entire schools will be assigned at random to receive either the new curriculum or continue with current practice. All 5<sup>th</sup> grade teachers in the school will participate. Within each school, they expect to have 6 teachers and 25 students per teacher. The outcome of interest is student math achievement. They expect that students in the treatment schools will improve their math scores by 0.20 standard deviations. How many schools are necessary for 80 percent power?

Use findings from [here](#) (Tables 4-7) to estimate the unknown design parameters. Assume a pretest is available.  
Select a State.

# Exercise 3

Suppose a group of researchers want to examine the effects of a whole school reform model on 5<sup>th</sup> grade achievement (math or reading). Schools are the unit of random assignment. The researchers secure 22 districts, 4 schools per district, and 100 5<sup>th</sup> grade students per school. What is the MDES?

Use tables on next slides to estimate the unknown design parameters.

Assume 0.03 effect size variability.

Assume a school-level pretest is available.

Select Mi or Texas.

Spybrook, J., Westine, C., & Taylor, J. (2016). Design parameters for impact research in science education: A multistate analysis. *AERA Open*, 2(1), 1-15.

TABLE 5

*Unconditional School-Level and District-Level ICC Averages for Science Achievement Outcomes by Grade and State: Three-Level Model*

| Grade | Michigan <sup>a</sup> |      |                   |      | Texas <sup>b</sup> |      |                   |      | Wisconsin <sup>c</sup> |      |                   |      |
|-------|-----------------------|------|-------------------|------|--------------------|------|-------------------|------|------------------------|------|-------------------|------|
|       | ICC <sub>L2</sub>     | SE   | ICC <sub>L3</sub> | SE   | ICC <sub>L2</sub>  | SE   | ICC <sub>L3</sub> | SE   | ICC <sub>L2</sub>      | SE   | ICC <sub>L3</sub> | SE   |
| 4     |                       |      |                   |      |                    |      |                   |      | .095                   | .005 | .037              | .005 |
| 5     | .076                  | .004 | .146              | .010 | .118               | .003 | .079              | .007 |                        |      |                   |      |
| 8     | .102                  | .009 | .117              | .011 | .104               | .005 | .060              | .007 | .107                   | .008 | .036              | .006 |
| 10    |                       |      |                   |      | .136               | .008 | .055              | .008 | .079                   | .008 | .042              | .006 |
| 11    | .270                  | .013 | .031              | .008 | .127               | .008 | .059              | .008 |                        |      |                   |      |

*Note.* ICC = intraclass correlation.

<sup>a</sup>Unconditional ICCs for Michigan are averages across 4 years of data (2007-2008 through 2010-2011).

<sup>b</sup>Unconditional ICCs for Texas are averages across 5 years of data (2006-2007 through 2010-2011).

<sup>c</sup>Unconditional ICCs for Wisconsin are averages across 6 years of data (2005-2006 through 2010-2011).

TABLE 7

*Average  $R^2$  Values for Most Recent School-Level Pretest Covariates in Two-Level and Three-Level Models by Subject, Grade, and State*

| Grade       | Three-level HLM <sup>c</sup> |             |                    |             |                        |             |
|-------------|------------------------------|-------------|--------------------|-------------|------------------------|-------------|
|             | Michigan <sup>a</sup>        |             | Texas <sup>b</sup> |             | Wisconsin <sup>c</sup> |             |
|             | $R^2_{1,2}$                  | $R^2_{1,3}$ | $R^2_{1,2}$        | $R^2_{1,3}$ | $R^2_{1,2}$            | $R^2_{1,3}$ |
| Science     |                              |             |                    |             |                        |             |
| 4           |                              |             |                    |             | .682                   | .880        |
| 5           | .541                         | .971        | .546               | .917        |                        |             |
| 8           | .642                         | .946        | .739               | .856        | .817                   | .844        |
| 10          |                              |             | .858               | .859        | .818                   | .798        |
| 11          | .942                         | .844        | .865               | .836        |                        |             |
| Reading     |                              |             |                    |             |                        |             |
| 4           |                              |             |                    |             | .619                   | .818        |
| 5           | .416                         | .918        | .472               | .713        |                        |             |
| 8           | .467                         | .857        | .619               | .623        | .749                   | .688        |
| 10          |                              |             | .634               | .525        | .753                   | .639        |
| 11          | .893                         | .743        | .583               | .586        |                        |             |
| Mathematics |                              |             |                    |             |                        |             |
| 4           |                              |             |                    |             | .575                   | .609        |
| 5           | .396                         | .657        | .440               | .755        |                        |             |
| 8           | .509                         | .780        | .630               | .617        | .682                   | .567        |
| 10          |                              |             | .797               | .581        | .789                   | .632        |
| 11          | .928                         | .692        | .803               | .679        |                        |             |

*Note.* HLM = hierarchical linear model

<sup>a</sup> $R^2$  for 1-year lagged school-level pretest covariates for Michigan are averages across 3 years of data (2008-2009 through 2010-2011).

<sup>b</sup> $R^2$  for 1-year lagged school-level pretest covariates for Texas are averages across 4 years of data (2007-2008 through 2010-2011).

<sup>c</sup> $R^2$  for 1-year lagged school-level pretest covariates for Wisconsin are averages across 5 years of data (2006-2007 through 2010-2011).

<sup>e</sup>The three-level model refers to a conditional HLM with students nested in schools nested in districts. School mean pretest covariates are included at Level 2 and aggregated at Level 3.

# Writing up a Power Analysis

- Differs depending on the design
- Differs depending on software
- Critical to identify your design and software
- TG Plus Power provides report template