Sunthud Pornprasertmanit
W. Joel Schneider

Sample size estimation for Two-Condition Cluster Randomized Design

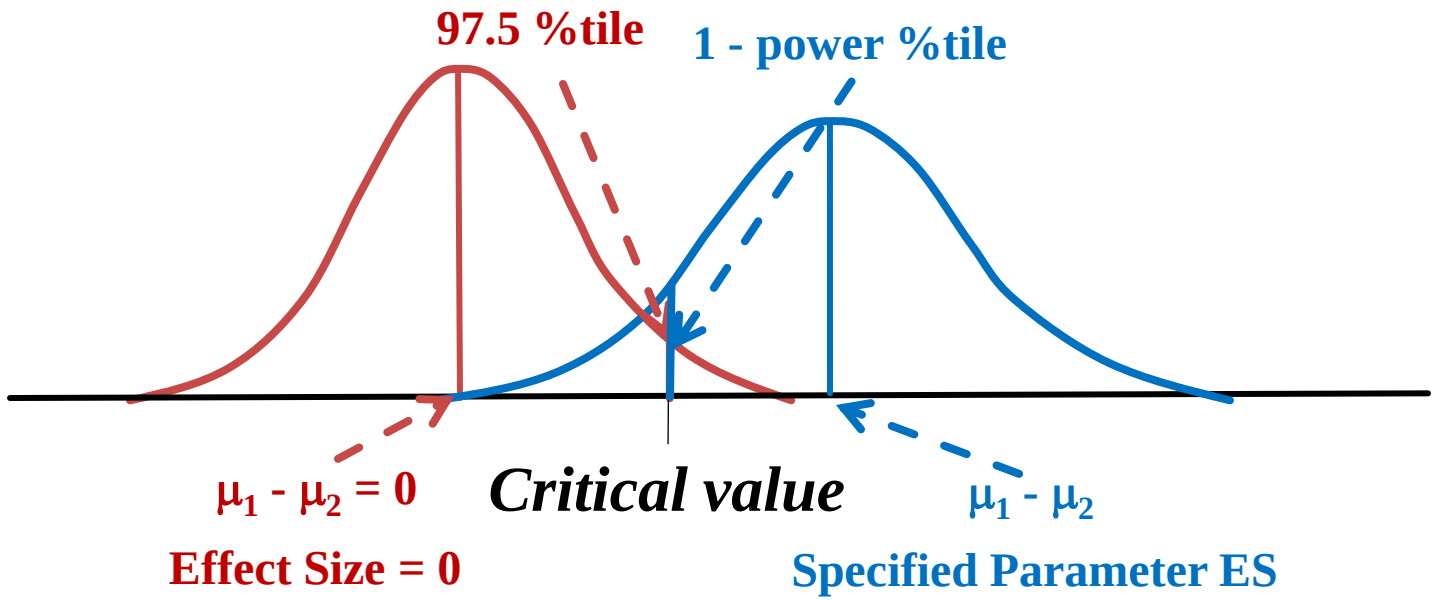
Outline

- Sample Size Estimation Approach
  - Power
  - Accuracy in Parameter Estimation
- Cluster Randomized Design (CRD)
- Sample Size Estimation in CRD
- Program Illustration

Two Approaches of Sample Size Estimation

- Power analysis
  - The probability of a significant result when there is a real effect in the population
- Width of Confidence Interval of Effect Size (*CI* of *ES*)
  - The accuracy of effect size estimation

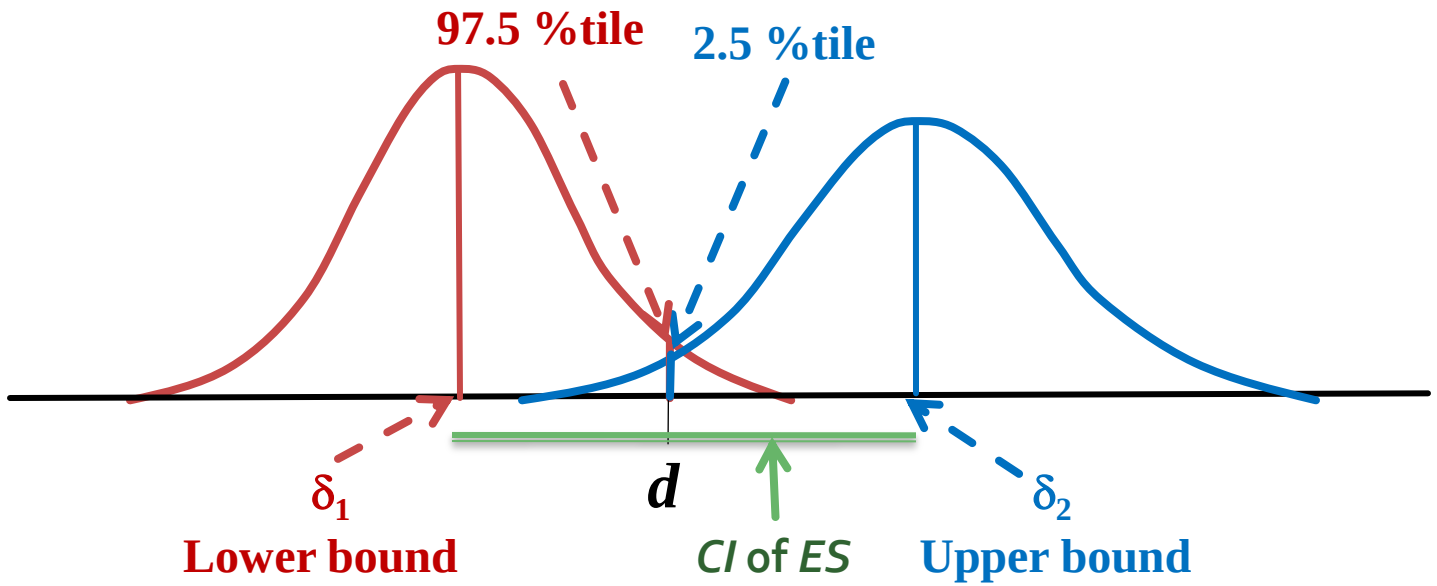
Power



- More *n* → Less *SE* → More power

Width of CI of ES

- 95 % CI of Cohen's *d*

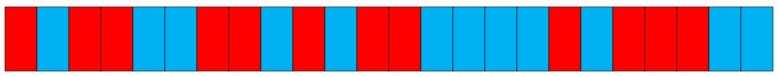


  - More *n* → Less *SE* → Narrower Width of *CI* of *ES*

Cluster Randomized Design

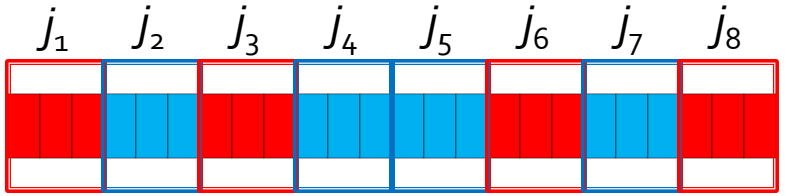
- CRD is the analysis of group differences when groups are randomly assigned to different conditions

Independent t-test



All sample size = 24

Two-condition CRD



All sample size = 24

$J = 8$

$n = 3$

Cluster Randomized Design

- Using Independent *t*-test
  - Independence of error terms assumption has been violated
    - Similar experience within clusters
  - Inflate type I error
- CRD accounts for interdependence

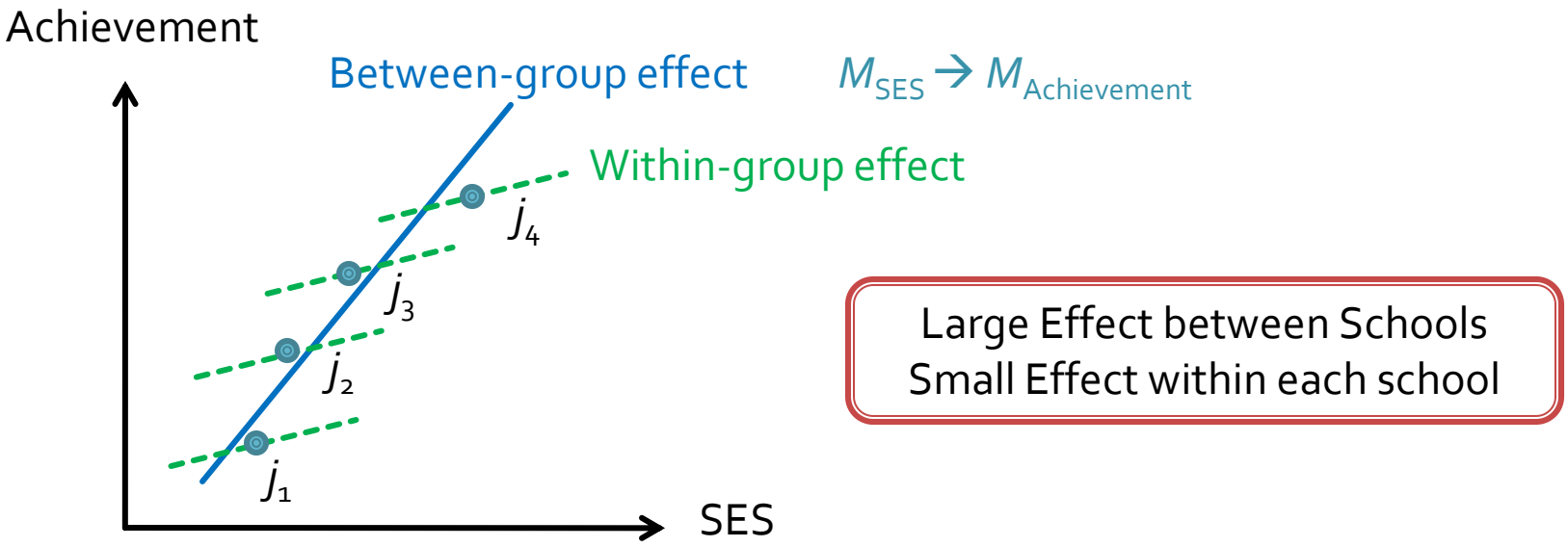
Basic Concepts in CRD

- Two types of errors in CRD
  - Group-level error variance
  - Individual-level error variance
- Intraclass correlation (ICC)

$$\text{ICC} = \frac{\text{Group error variance}}{\text{Group error variance} + \text{Individual error variance}}$$

Basic Concepts in CRD

- Covariate Effect in CRD



Effect Size

- Effect Size Definition

$$\delta = \frac{\mu_1 - \mu_2}{\sigma}$$

- In single level design, σ is pooled *SD* or $\sqrt{MS_{error}}$
- In CRD, three types of pooled *SD*
  - Group or $\sqrt{\tau}$
  - Individual or $\sqrt{\sigma^2}$
  - Total or $\sqrt{\tau + \sigma^2}$

Effect Size

- Hedges (2007) guideline
- In this study, use only individual pooled *SD*
- Assume $\sigma = 1$ → Effect Size = Condition Difference

CI of *ES*

- Formula by Hedges (2007)
- Phantom Variable Method by SEM packages

$$\text{Effect Size} = \frac{\gamma_{Y \text{ on } X}}{\sqrt{\sigma^2_{Within}}}$$

  - Find *CI* of *ES* based on Wald Statistic

Finding Sample Size

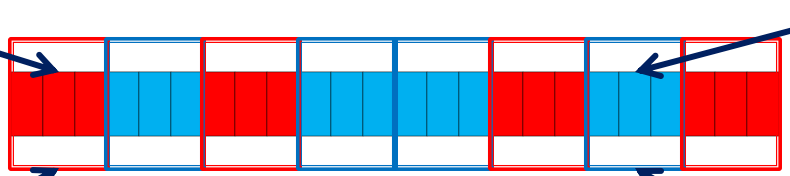
- Different Combination of three factors can yield the same power or width of *CI*
  - Number of Clusters (*J*)
  - Cluster size (*n*)
  - Proportion of treatment clusters (*p*)
- Different Combination also yield same costs

Finding Sample Size

- Four costs

Treatment Individual Cost (TIC)

Control Individual Cost (CIC)



Treatment Group Cost (TGC)

Control Group Cost (CGC)

Each Treatment Group Cost = **TGC** + (*n* x **TIC**)

Each Control Group Cost = **CGC** + (*n* x **CIC**)

Number of Treatment Groups = pJ

Number of Control Groups = $(1-p)J$

$$\text{Total Cost} = pJ(\mathbf{TGC} + (n \times \mathbf{TIC})) + (1-p)J(\mathbf{CGC} + (n \times \mathbf{CIC}))$$

Finding Sample Size

- Three criteria
  - Minimize number of overall individuals by specified power/width
    - Find various *n*, *J*, *p* for given power/width → Find lowest *nJ*
  - Minimize cost by specified power/width
    - Find various *n*, *J*, *p* for given power/width → Find lowest cost
  - Maximize power/ Minimize width by specified cost
    - Find various *n*, *J*, *p* for given cost → Find highest power/width

Program Logic

- Find starting values by Wald Statistic formula using normal approximation
  - Given individual error variance = 1
- Find more accurate result by a priori Monte Carlo Simulation by Mplus

Program Illustration

1. What happens when a covariate is added? (Post Hoc)
2. How many classrooms are required to detect a small effect? (A priori)

Example 1

- Effectiveness of training to administer cognitive behavioral therapy (King et al., 2002)
- 84 therapists assigned to two conditions
- 4 patients each
- DV = Beck Depression Inventory (BDI) Score
- ES with individual-level SD = 0.09
- Intraclass correlation = 0.013

Example 1

- Result = *ns*
- Post Hoc power = 0.124
- If the researchers collected BDI scores of therapists,
  - Cluster-level variable
  - Cluster-level Error Variance Explained = 10%
- Can the covariate help to achieve high power?

Example 2

- A new teaching method
- DV = Academic Achievement
- Intraclass correlation = 0.25
- Classroom size = 25
- Power = 0.8
- Meaningful ES = 0.2

Example 2

- Cost

	Treatment	Control
Cluster Cost	600	300
Individual Cost	2	2

- How many classrooms should be used?

Any Question???