Designing Adequately Powered Cluster-Randomized Trials using Optimal Design

Jessaca Spybrook, Western Michigan University
March 4, 2010

*Joint work with Stephen Raudenbush and Andres Martinez

Cluster-Randomized Trials

- CRT's
  - What?
  - When?
  - Why?

- Why do we care about power?

Different Cluster Randomized Trials

	Cluster Randomized Trial (2-level CRT)	Three-Level Model with Treatment at Level 3 (3-level CRT)	Multisite Cluster Randomized Trial (MSCRT)
Number of Levels	2	3	3
Level of Randomization	2	3	2
Blocking	No	No	Yes
Example of Level Structure	Students **Schools**	Students Classrooms **Schools**	Students **Schools** Districts

The Model (2-level CRT)

$$\text{L}-1: \quad y_{ij} = \beta_{0j} + e_{ij} \qquad\qquad e_{ij} \sim N\left(0, \sigma^2\right)$$

$$\text{L}-2: \quad \beta_{0j} = \gamma_{00} + \gamma_{01} W_j + u_{0j} \qquad u_{0j} \sim N(0, \tau)$$

Estimation

$$\hat{\gamma}_{01} = \bar{Y}_E - \bar{Y}_C$$

$$\operatorname{var}(\hat{\gamma}_{01}) = \frac{4(\tau + \sigma^2 / n)}{J}$$

Hypothesis Testing

$$H_0: \ \gamma_{01} = 0$$

$$H_a: \ \gamma_{01} \neq 0$$

$$Power = \Pr(F > f_{critical})$$

$$F = \frac{MS_{treatments}}{MS_{groups\ within\ treatments}}$$

Noncentrality Parameter

$$\frac{E(MS_{treatments})}{E(MS_{groups\ within\ treatments})} = 1 + \lambda$$

$$\lambda = \frac{\gamma_{01}^2}{\operatorname{var}\left(\hat{\gamma}_{01}\right)}$$

$$= \frac{J\gamma_{01}^2}{4(\tau + \sigma^2 / n)}$$

Standardize the Effect Size

$$\rho = \frac{\tau}{\tau + \sigma^2}$$

$$\delta = \frac{\gamma_{01}}{\sqrt{\tau + \sigma^2}}$$

$$\lambda = \frac{J\delta^2}{4(\rho + (1 - \rho)/n)}$$

2-level CRT: Key Parameters

- Intraclass correlation, ρ
- Effect size, δ
- Covariate-outcome correlation, R^2

- Where do they come from?
  - Pilot study
  - Review of the literature

Intraclass correlation

- Achievement
  - Bloom, Bos, & Lee (1999)
  - Schochet (2008)
  - Hedges & Hedberg (2007)
  - Bloom, Richburg-Hayes, and Black (2007)

- Health outcomes
  - Murray & Blitstein (2003)

Effect Size

- Bloom, Hill, Black, & Lipsey (2008)
- Interpret magnitude in terms of context, measures, samples
- Frame of reference
  - Normative changes, policy gaps, etc.

Covariate-outcome correlation

- Achievement
  - Bloom, Richburg-Hayes, & Black (2007)

Optimal Design

- Tool for planning studies
- Software and accompanying documentation
- Always being updated ☺

- Warning: Optimal Design is easy to use but is also easy to misuse

Optimal Design

- Organization
  - Person Randomized Trials
  - *Cluster Randomized Trials with Person-level Outcomes*
  - Cluster Randomized Trials with Cluster-level Outcomes

2-level CRT Example

Program: A whole school reform model is randomly assigned and implemented at the school level.

Research question: Is the whole school reform model more effective than the current one?

Standardized math tests scores used as outcome measure. Based on previous research:

- 18 percent of the variation lies between schools.
  (Students nested within schools).
- Researchers expect the treatment to boost test scores by 0.25 standard deviations.

With 90 students per school, how many schools are needed to detect the above effect size with 80% power?

Questions to Consider

- How many levels are in this study?
- What is the unit of randomization?
- Is there blocking?
- What is the minimum detectable effect size?
- What is a good estimate of the intraclass correlation? Percent of variance explained by the covariate?
- What information, if any, is missing?

2-level CRT Parameters for OD

- J : total number of clusters
- n : Total number of individuals per cluster
- ρ : Intraclass correlation
- R_{l2}^2 : percent of variation explained by cluster-level (level 2) covariate
- δ : effect size

2-level CRT exercise

Suppose a researcher is planning the evaluation of a whole school reform (WSR) model and is considering a cluster-randomized trial in which:

- Entire schools will be assigned at random to receive either the WSR or continue with their current practices
- The WSR model is considered effective if students in the treatment schools increase their scores in a standardized test significantly more than students in the control schools

The researcher:

- Will randomly select 300 students from each school
- Expect students in the treatment schools will outperform students in the control schools by 0.20 standard deviations
- Estimates that approximately 20 percent of the variation in student test scores lies between schools
- Has access to last year's test scores and expects those scores to be strongly correlated with the scores in the new test (say .8)

Some questions:

- How many schools are needed at 0.80 power? At 0.90 power?
- Does testing only 50 students per school affect power?
- What is the MDES for power = 0.80 if the researcher can only recruit 20 schools?

Three-level CRT

- Example:
  - Level 1: Students
  - Level 2: Teachers
  - Level 3: Schools
- Outcomes still at the student level
- Treatment is at the school level

The Model (3-level CRT)

- Model:

$$\text{L}-1: \quad y_{ijk} = \pi_{0jk} + e_{ijk} \qquad e_{ijk} \sim N\left(0, \sigma^2\right)$$

$$\text{L}-2: \quad \pi_{0jk} = \beta_{00k} + r_{0jk} \qquad r_{0jk} \sim N(0, \tau_{\pi})$$

$$\text{L}-3: \quad \beta_{00k} = \gamma_{000} + \gamma_{001} W_k + u_{00k} \qquad u_{00k} \sim N(0, \tau_{\beta_{00k}})$$

Estimation

$$\hat{\gamma}_{001} = \bar{Y}_E - \bar{Y}_C$$

$$Var(\hat{\gamma}_{001}) = \frac{4\left[\tau_{\beta_{00k}} + \left(\tau_{\pi} + \sigma^2 / n\right) / J\right]}{K}$$

Hypothesis Test

$$H_0: \ \gamma_{001} = 0$$

$$H_a: \ \gamma_{001} \neq 0$$

$$Power = \Pr(F > f_{critical})$$

$$F = \frac{MS_{treatments}}{MS_{groups\ within\ treatments}}$$

Noncentrality Parameter

$$\frac{E(MS_{treatments})}{E(MS_{groups\ within\ treatments})} = 1 + \lambda$$

$$\lambda = \frac{\gamma_{001}^2}{Var(\hat{\gamma}_{001})}$$

$$= \frac{K\gamma_{01}^2}{4(\tau_{\beta_{00k}} + (\tau_{\pi} + \sigma^2 / n) / J)}$$

Standardize the effect size

$$\rho_{l2} = \frac{\tau_{\pi}}{\tau_{\beta_{00k}} + \tau_{\pi} + \sigma^2}$$

$$\rho_{l3} = \frac{\tau_{\beta_{00k}}}{\tau_{\beta_{00k}} + \tau_{\pi} + \sigma^2}$$

$$\delta = \frac{\gamma_{001}}{\sqrt{\tau_{\beta_{00k}} + \tau_{\pi} + \sigma^2}}$$

$$\lambda = \frac{K\delta^2}{4\left\{\rho_{L3} + \left[\rho_{L2} + (1 - \rho_{L2} - \rho_{L3})/n\right]/J\right\}}$$

3-level CRT: Key Parameters

- K : total number of schools
- J : total number of classrooms per school
- n : total number of students per classroom
- ρ_π : classroom-level (level 2) intraclass correlation
- ρ_β : school-level (level 3) intraclass correlation
- R_{l3}^2 : percent of variation explained by school-level (level 3) covariate
- δ : effect size

3-level CRT Example

Program: A whole school reform model is randomly assigned and implemented at the school level.

Research question: Is the whole school reform model more effective than the current one?

Standardized math tests scores used as outcome measure. Researchers secure 8 teachers per school and 25 students per teacher. Based on previous research:

- 12% percent of the variation lies between schools. 8 percent of the variation lies between classes.

  (Students nested within schools).
- Expects pre-test to be strongly correlated with outcome (say 0.80)

What is the MDES with 30 total schools?

Questions to Consider

- How many levels are in this study?
- What is the unit of randomization?
- Is there blocking?
- What is the minimum detectable effect size?
- What is a good estimate of the intraclass correlations? Percent of variance explained by the covariate?
- What information, if any, is missing?

3-level CRT Parameters for OD

- K : total number of schools
- J : total number of classrooms per school
- n : total number of students per classroom
- ρ_π : classroom-level (level 2) intraclass correlation
- ρ_β : school-level (level 3) intraclass correlation
- R_{l3}^2 : percent of variation explained by school-level (level 3) covariate
- δ : effect size

3-level CRT exercise

Suppose a researcher is planning the evaluation of a 3rd grade math curriculum and is considering a cluster-randomized trial in which:

- Entire schools will be assigned at random to receive either the curriculum or continue with their current practices. All 3rd grade teachers will participate.
- The new math curriculum is considered effective if students in the treatment schools increase their scores in a standardized test significantly more than students in the control schools

The researcher:

- Will randomly select an average of 6 teachers per school and 25 students per teacher
- Expect students in the treatment schools will outperform students in the control schools by 0.20 standard deviations
- Estimates that approximately 12 percent of the variation in student test scores lies between schools and 6 percent lies between teachers
- Has access to last year's test scores and expects those scores to be strongly correlated with the scores in the new test (say .7)

Some questions:

- How many schools are needed at 0.80 power? At 0.90 power?
- What happens to the power if there are only 4 teachers available per school?
- What is the MDES for power = 0.80 if the researcher can only recruit 30 schools?

Multi-site Cluster Randomized Trial (MSCRT)

- A blocked or MSCRT:
  - Randomization of clusters into treatment conditions within *blocks* or *sites*, ie. districts
- Blocking used to:
  - Increase face validity
  - Increase precision of the estimates
- MSCRTs designs are very common
  - 3 out of every 4 designs funded by NCER and by NCEE between 2002 and 2006 (Spybrook, 2008)

Two Types of MSCRTs

- Multi-site CRT ("naturally blocked"). For example:
  - Randomization at classroom level, with schools as sites
  - Randomization at school level, with cities as sites

- Blocked CRT ("blocking embedded"). For example:
  - Blocks of schools are created based on percentage of free/reduced lunch
  - Blocks of communities are created based on SES

Blocks: Fixed or Random

	Random-effects	Fixed-effects
Conceptual:		
■ Generalization ■ Replication	To the "population of blocks"	Restricted to the particular blocks in the study
Operational:		
■ How to test for treatment effect and treatment effect heterogeneity ■ What and how to report	Report (a) the mean treatment effect and (b) the variance of the treatment effect	Report *K* simple effects

The Model: Random Block Effects

$$\text{L}-1: \quad y_{ijk} = \pi_{0jk} + e_{ijk} \qquad e_{ijk} \sim N\left(0, \sigma_e^2\right)$$

$$\text{L}-2: \quad \pi_{0jk} = \beta_{00k} + \beta_{01k} T_{jk} + r_{0jk} \qquad r_{jk} \sim N(0, \tau_\pi)$$

$$\text{L}-3: \quad \begin{matrix} \beta_{00k} = \gamma_{000} + u_{00k} \\ \beta_{01k} = \gamma_{010} + u_{01k} \end{matrix} \qquad \begin{pmatrix} u_{00k} \\ u_{01k} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{\beta_{00}} & \tau_{\beta_{01}} \\ \tau_{\beta_{10}} & \tau_{\beta_{11}} \end{pmatrix}\right]$$

Estimation: Random effects

$$\hat{\gamma}_{010} = \bar{Y}_E - \bar{Y}_C$$

$$Var(\hat{\gamma}_{010}) = \frac{\tau_{\beta_{11}} + 4(\tau_\pi + \sigma^2 / n) / J}{K}$$

$\tau_{\beta_{11}} = 0$

Hypothesis Testing: Random effects

$$H_0: \ \gamma_{010} = 0$$

$$H_a: \ \gamma_{010} \neq 0$$

$$Power = \Pr(F > f_{critical})$$

$$F = \frac{MS_{treatment}}{MS_{within-cell}}$$

$\tau_{\beta_{11}} = 0$

Noncentrality Parameter: Random effects

$$\frac{E(MS_{treatment})}{E(MS_{within-cell})} = 1 + \lambda$$

$$\lambda = \frac{\gamma_{010}^{2}}{Var(\hat{\gamma}_{010})}$$

$$= \frac{K\gamma_{010}^{2}}{\tau_{\beta_{11}} + 4(\tau_{\pi} + \sigma^{2}/n)/J}$$

$\tau_{\beta_{11}} = 0$

Standardize the Effect Size

$$\rho = \frac{\tau_\pi}{\tau_\pi + \sigma^2}$$

$$\delta = \frac{\gamma_{010}}{\sqrt{\tau_\pi + \sigma^2}}$$

$$\sigma_\delta^2 = \frac{\tau_{\beta_{11}}}{\tau_\pi + \sigma^2}$$

$$\lambda = \frac{K\delta^2}{\sigma_\delta^2 + 4\left[\rho + (1-\rho)/n\right]/J}$$

The Model: Fixed Block Effects

$$\text{L}-1: \quad y_{ijk} = \pi_{0jk} + e_{ijk} \qquad\qquad e_{ijk} \sim N\left(0, \sigma_e^2\right)$$

$$\text{L}-2: \quad \pi_{0jk} = \beta_{00k} + \beta_{01k} T_{jk} + r_{0jk} \qquad r_{jk} \sim N(0, \tau_\pi)$$

$$\text{L}-3: \quad \begin{aligned} \beta_{00k} &= \gamma_{000} + u_{00k} \\ \beta_{01k} &= \gamma_{010} + u_{01k} \end{aligned}$$

$\mathrm{u}_{00\mathrm{k}}$ are fixed effects associated with each site mean, constrained to have a mean of zero; and

$\mathrm{u}_{01\mathrm{k}}$ are fixed effects associated with each treatment-by-site interaction, constrained to have a mean of zero.

Estimation: Fixed effects

$$\hat{\gamma}_{010} = \bar{Y}_E - \bar{Y}_C$$

$$Var(\hat{\gamma}_{010}) = \frac{4(\tau_\pi + \sigma^2 / n) / J}{K}$$

$\tau_{\beta_{11}} = 0$

Hypothesis Testing: Fixed effects

$$H_0 : \ \gamma_{010} = 0$$

$$H_a : \ \gamma_{010} \neq 0$$

$$Power = \Pr(F > f_{critical})$$

$$F = \frac{MS_{treatment}}{MS_{treatment-by-site}}$$

$\tau_{\beta_{11}} = 0$

Noncentrality Parameter: Fixed effects

$$\frac{E(MS_{treatment})}{E(MS_{treatment-by-site})} = 1 + \lambda$$

$$\lambda = \frac{\gamma_{010}^{2}}{Var(\hat{\gamma}_{010})}$$

$$= \frac{K\gamma_{010}^{2}}{4(\tau_{\pi} + \sigma^{2}/n)/J)}$$

$\tau_{\beta_{11}} = 0$

Standardize the Effect Size

$$\rho = \frac{\tau_\pi}{\tau_\pi + \sigma^2}$$

$$\delta = \frac{\gamma_{010}}{\sqrt{\tau_\pi + \sigma^2}}$$

$$\lambda = \frac{K\delta^2}{4\left[\rho + (1-\rho)/n\right]/J}$$

MSCRT: Key Parameters

- K : total number of sites or blocks
- J : total number of clusters per site
- n : total number of students per cluster
- ρ : intraclass correlation
- R_{l2}^2 : percent of variation explained by cluster-level (level 2) covariate
- δ : effect size
- σ_δ^2: effect size variability, for the random site effects > 0, for the fixed site effects set to 0

MSCRT Exercise 1

Program: A whole school reform model is randomly assigned and implemented at the school level.

Research question: Is the whole school reform model more effective than the current one?

Standardized math tests scores used as outcome measure.
Researchers secure 3 large districts with 100 students per school.
Based on previous research:

- 12% percent lies between districts within schools
- School-level covariate will explain 50 percent of outcome variation
- Interested in an MDES of 0.25

How should we treat the district effects? Fixed or random.
How many schools per district at 0.80 power?
What if we select 12 districts instead?

Questions to Consider

- How many levels are in this study?
- What is the unit of randomization?
- Is there blocking?
- What is the minimum detectable effect size?
- What is a good estimate of the intraclass correlation? The percent of variance explained by the covariate?
- What information, if any, is missing?

MSCRT parameters for OD

- K : total number of blocks/sites
- J : total number of clusters per site
- n : total number of individuals per cluster
- ρ : intraclass correlation
- δ *:* effect size
- σ_{δ}^{2} : effect size variability (specifying greater than 0 treats blocks as random effects)
- B : percent of variation explained by blocking
- R_{l2}^{2} : the percent of variation explained by site-level (level-2) covariate

MSCRT exercise

Suppose a researcher is planning the evaluation of a school-wide literacy program and is considering a cluster-randomized trial in which:

- Entire schools will be assigned at random to receive either the literacy program or continue with their current practices.
- The new math curriculum is considered effective if students in the treatment schools increase their scores in a standardized test significantly more than students in the comparison schools

The researcher:

- Secured 10 districts to participate in the study
- Within each district, 6 schools will be randomly assigned to either the new literacy program or the comparison group
- 100 students per school
- Estimates that approximately .16 percent of the variation in student test scores lies between schools within districts
- Has access to last year's test scores and expects those scores to be strongly correlated with the scores in the new test (say .7)

Some questions:

- What is the MDES if the researcher treats the blocks as fixed effects?
- What is the MDES if the researcher treats the blocks as random effects (ESV = 0.01?

Writing up a Power Analysis

	Two-level Cluster Randomized Trial (2-level HLM)	Three-level Cluster Randomized Trial (3-level HLM)	Three-level Multi-site Cluster Randomized Trial (3-level RBD)
Number of level-one units	X	X	X
Number of level-two units	X	X	X
Number of level-three units		X	X
Level 2 intra-class correlation	X	X	X
Level 3 intra-class correlation		X	
Percent of variance explained by covariate	X	X	X
Percent of variance explained by blocking			X
Treatment of blocks – fixed versus random			X
Effect size variability			X