Assessing the Statistical Power/Precision Of Multisite Trials for Estimating Parameters Of Cross-site ITT Impact Distributions

Howard Bloom

Jessaca Spybrook

Presented to the Workshop on "Learning about and from Variation in Program Impacts?" at Stanford University on July 18, 2016. This presentation is based on research funded by the Spencer Foundation and the William T. Grant Foundation.

Overview

Questions

- How does the number and size of sites influence the precision of a multi-site trial?
- What other factors influence this precision?
- How should this information be used to design trials?

Designs

- Multi-site individually-randomized trials
- Multi-site cluster-randomized trials

Parameters

- The cross-site mean effect size
- The cross-site effect size standard deviation
- The difference in mean effect sizes for two categories of sites (i.e. the coefficient for a binary site-level moderator)

Reference

- Bloom, H.S. and J. Spybrook (under review) "Assessing the Precision of Multisite Trials for Estimating Parameters of Cross-site Distributions of ITT Program Effects."

Effect Sizes and Precision

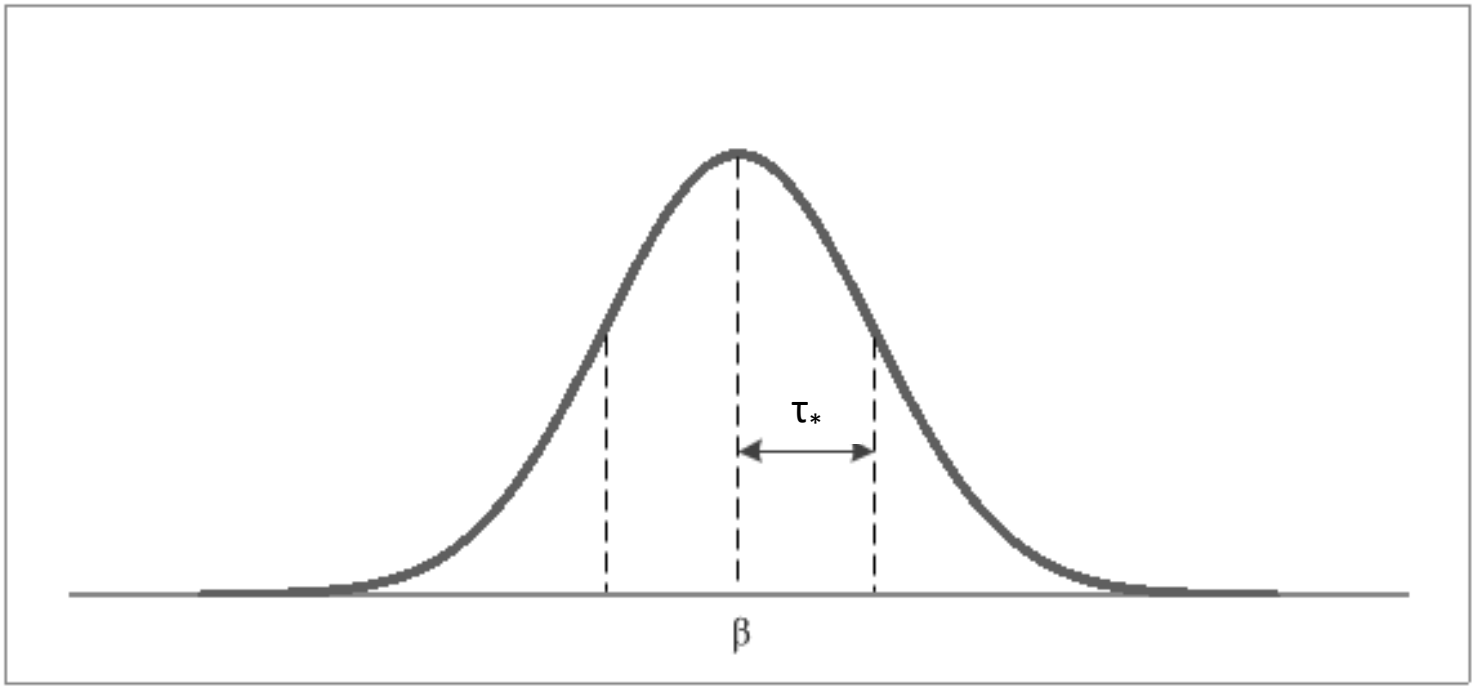
A Standardized Mean Difference Effect Size

$$Effect\ Size = \frac{Program\ Effect\ in\ Original\ Units}{Standard\ Deviation\ of\ Untreated\ Outcomes}$$

A Minimal Detectable Effect Size

The smallest program effect size that can be detected with 80 power at the 0.05 level of statistical significance.

A Cross-Site Distribution of Program Effect Sizes



β = The cross-site mean effect size
(*Associated design parameter*: Minimum detectable cross-site effect size, MDES)

τ_* = The cross-site standard deviation of effect sizes
(*Associated design parameter*: Minimum detectable effect-size standard deviation, MDESSD)

Prototypical Estimation Model for MDES and MDESSD: FIRC

Fixed Site-Specific Intercepts, Random Site-Specific Program Effects, a Single Level-One Covariate and a Single Level-One Residual Variance (for simplicity)

Level One: Individuals

$$Y_{ij} = \alpha_j + \beta_j \cdot T_{ij} + \theta \cdot X_{ij} + r_{ij}$$

$$r_{ij} \sim N(0, \sigma^2_{|x\alpha_j})$$

$$\sigma^2_{|x\alpha_j} = (1 - \rho_C)(1 - R^2_{C(within)})\sigma^2$$

Level Two: Sites

$$\alpha_j = \alpha_j$$

$$\beta_j = \beta + b_j$$

$$b_j \sim N(0, \tau^2_*)$$

MDES Computational Expression
(Individual Randomization)

$$MDES = M_{J-1}\sqrt{(\frac{1}{J})(\tau_*^2 + \frac{(1-\rho_C)(1-R^2_{C(within)})}{n\bar{T}(1-\bar{T})})}$$

where:

σ^2 is set equal to one.

M_{j-1} = a multiplier that rapidly approaches 2.8 as J increases (for a two-tail test at the 0.05 significance level with 80 percent power)

J = number of sites

n = number of individuals per site

$\bar{T}$ = proportion of individuals from each site randomized to treatment

τ_*= cross-site standard deviation of effect sizes

ρ_C = intra-class correlation for control group outcomes (i.e. the proportion of σ^2explained by site indicators)

$R^2_{C(within)}$ = the proportion of within-site outcome *variance* explained by our baseline covariate

Harmonic vs. Arithmetic Mean Site Sizes
(It Matters)

- You should always use the harmonic mean site sample size when assessing statistical precision.
- When site sample sizes vary, the harmonic mean will be smaller than the arithmetic mean – often by a lot.
- For example:
  - Welfare to work (site sample sizes range from 177 to 4,418)
    - Harmonic mean = 621
    - Arithmetic mean = 1,176
  - Head Start (site sample sizes range from 2 to 75)
    - Harmonic mean = 13
    - Arithmetic mean = 19

MDES for a Cross-Site Mean Effect Size

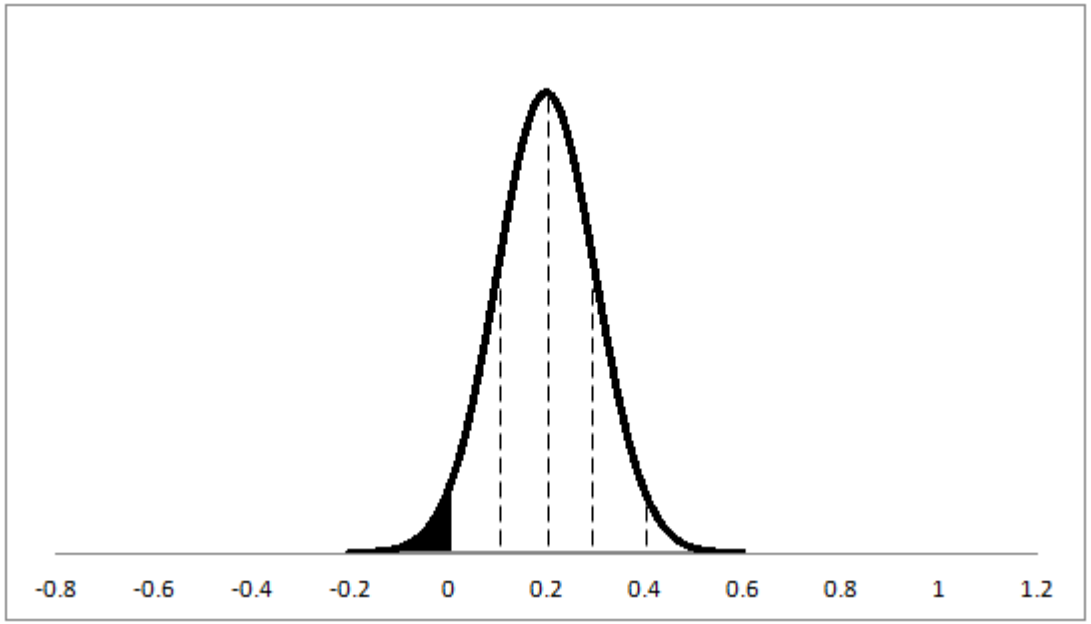
(Individual Randomization)

Individuals per site (n)	Sites					
	5	10	20	50	100	200
5	1.10	0.65	0.43	0.27	0.19	0.13
10	0.80	0.47	0.31	0.19	0.14	0.10
20	0.59	0.35	0.23	0.14	0.10	0.07
50	0.42	0.25	0.17	0.10	0.07	0.05
100	0.35	0.21	0.14	0.08	0.06	0.04
200	0.30	0.18	0.12	0.07	0.05	0.04
500	0.27	0.16	0.11	0.07	0.05	0.03

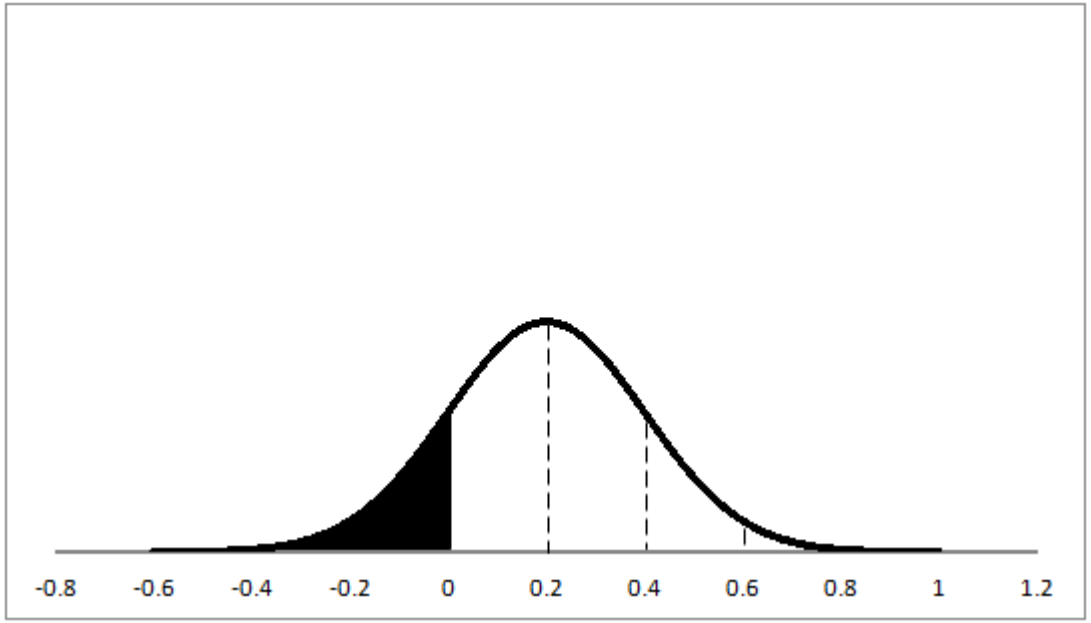
NOTES: Values in the table are for two-tail significance = 0.05, power = 80 percent, a single level-one baseline covariate, $\rho_C = 0.15$, $R^2_{C(within)} = 0.4$, $\bar{T} = 0.5$, constant n within sites, and $\tau_* = 0.15$.

Thinking About the Magnitude of Cross-Site Effect-Size Variation

"Moderate" Variation ($\tau_* = 0.1$)



"Substantial" Variation ($\tau_* = 0.2$)



MDESSD Computational Expression
(Individual Randomization)

$$MDESSD = \sqrt{\left(\frac{(1-\rho_C)\left(1-R^2_{C(within)}\right)}{n\bar{T}(1-\bar{T})}\right)\left(\frac{F_{0.05}}{F_{0.80}} - 1\right)}$$

where:

σ^2 set equal to one

n = number of individuals per site

$\bar{T}$ = proportion of individuals at each site randomized to treatment

ρ_C = intra-class correlation for control group outcomes (i.e. the proportion of σ^2 explained by our site indicators)

$R^2_{C(within)}$ = proportion of within-site outcome variance explained by our baseline covariate

$F_{0.05}$ = 0.05 critical value for an F statistic with J-1 numerator degrees of freedom and J(n – 2) – K denominator degrees of freedom (for J sites, n individuals per site and K level-one covariates)

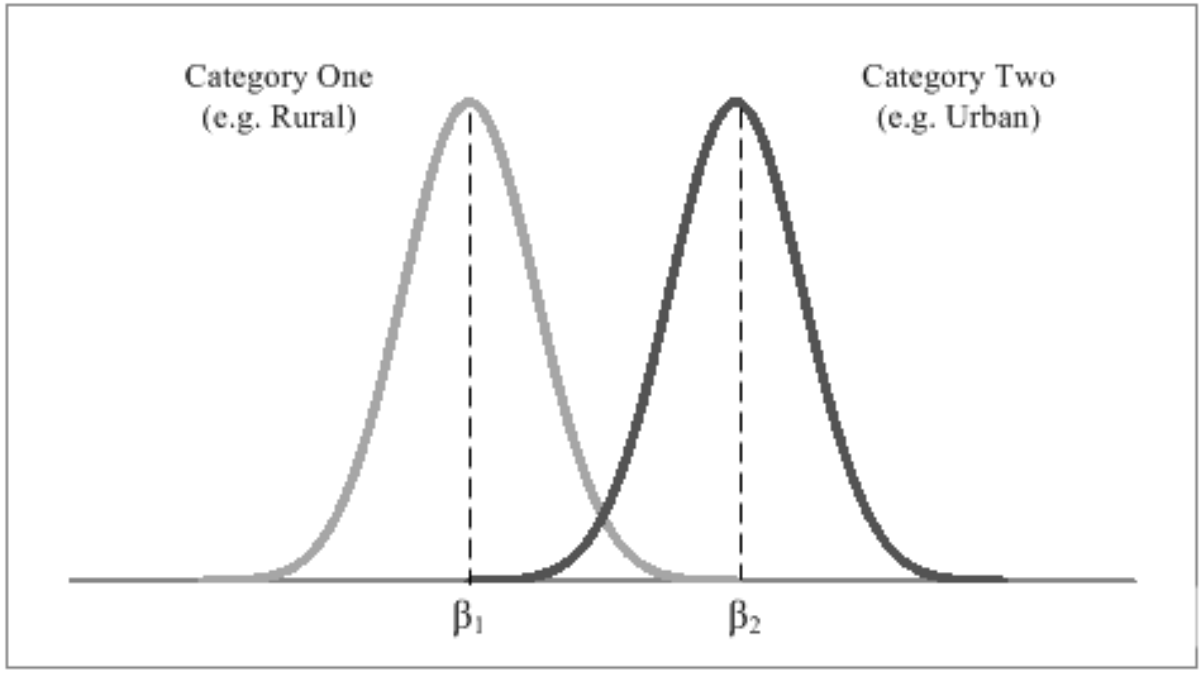
$F_{0.80}$ = the value that is below 80 percent of the distribution of possible values for an F statistic with J-1 numerator degrees of freedom and J(n – 2) – K denominator degrees of freedom

MDESSD for a Cross-Site Standard Deviation of Effect Sizes
(Individual Randomization)

	Number of Sites (J)					
Individuals/site (n)	**5**	**10**	**20**	**50**	**100**	**200**
5	1.65	1.07	0.78	0.57	0.45	0.37
10	1.05	0.70	0.52	0.38	0.30	0.35
20	0.72	0.48	0.36	0.26	0.21	0.17
50	0.45	0.30	0.22	0.16	0.13	0.11
100	0.31	0.21	0.16	0.11	0.09	0.08
200	0.22	0.15	0.11	0.08	0.07	0.05
500	0.14	0.09	0.07	0.05	0.03	0.03

NOTE: Values in the table are for significance = 0.05, power = 80 percent, a single level-one baseline covariate, $\rho_C = 0.15, R^2_{C(within)} = 0.4, \bar{T} = 0.5$ and constant *n* within sites.

Cross-Site Effect-Size Distributions For Two Categories of Sites



β_1 = Grand mean effect size for category #1
β_2 = Grand mean effect size for category #2

$\beta_2 - \beta_1$ = The moderator coefficient
(Associated design parameter: Minimum detectable effect size difference, MDESD)

Estimation Model for MDESD: FIRC with a Binary Site-Level Impact Predictor

Fixed Site-Specific Intercepts, Random Site-Specific Program Effects, a Single Level-One Covariate, A Single Level-One Residual (for simplicity) and a Single Binary Site-Level Impact Predictor

Level One: Individuals

$$Y_{ij} = \alpha_j + \beta_j \cdot T_{ij} + \theta \cdot X_{ij} + r_{ij}$$

$$r_{ij} \sim N(0, \sigma^2_{|x\alpha_j})$$

$$\sigma^2_{|x\alpha_j} = (1 - \rho_C)(1 - R^2_{C(within)})\sigma^2$$

- Level Two: Sites

$$\alpha_j = \alpha_j$$

$$B_j = \beta' + \Delta W_j + b'_j$$

$$b_j \sim N(0, \tau^2_{*|W})$$

$$where\ \tau^2_{*|W} = \left(1 - R^2_{\Delta}\right)\tau^2_*$$

More on MDESD

A Few noteworthy distinctions:

Max Possible Effect Size difference is a function of τ_*^2

$$\Delta_{\max} = \sqrt{\frac{\tau_*^2}{\pi(1-\pi)}}$$

Relationship between R_Δ^2 **and** Δ

$$\Delta = \sqrt{\frac{\tau_*^2 R_\Delta^2}{\pi(1-\pi)}} \quad \text{or} \quad R_\Delta^2 = \frac{\Delta^2 \pi(1-\pi)}{\tau_*^2}$$

where π **is the proportion of sites in a particular category.**

MDESD Power Calculations
(Individual Randomization)

Steps

1. Determine $\Delta_{\max}$, $\Delta_{\max} = \sqrt{\frac{\tau_*^2}{\pi(1-\pi)}}$

Note: For $\Delta_{\max}$, $R_\Delta^2 = 1$

2. Calculate Power for $\Delta_{\max}$ where $\Delta = \Delta_{\max}$, and

$$POWER = 1 - \Phi\left(t_{critical} - \frac{\Delta}{SE\left(\hat{\Delta}\right)} \right)$$

where

$$se\left(\hat{\Delta}_{*|R_W^2}\right) = \sqrt{\left(\frac{1}{\pi(1-\pi)J}\right) + \left((1-R_W^2)\tau_*^2 + \frac{(1-\rho_C)(1-R_{C(within)}^2)}{n\bar{T}(1-\bar{T})}\right)}$$

MDESD Power Calculations
(Individual Randomization)

Steps (Continued)

3. If power less than 0.80, there is no MDESD for that particular sample size combination

4. If power greater than 0.80, select $\Delta < \Delta_{\max}$

5. Calculate $R_{\Delta}^{2} = \dfrac{\Delta^{2}\pi(1-\pi)}{\tau_{*}^{2}}$

6. Calculate Power (see step 2)

7. Repeat steps 4-7 until power = 0.80

Please email jessaca.spybrook@wmich.edu for R code for MDESD power calculations.

MDESD for the Coefficient (Δ) On a Binary Site-Level Moderator
(Individual Randomization)

Individuals per site (n)	Sites 5	10	20	50	100	200
5	--	--	--	--	--	0.26 (0.72)
10	--	--	--	--	0.26 (0.72)	0.19 (0.39)
20	--	--	--	0.27 (0.78)	0.20 (0.43)	0.14 (0.21)
50	--	--	0.28 (0.84)	0.19 (0.39)	0.14 (0.21)	0.10 (0.11)
100	--	0.30 (0.96)	0.23 (0.56)	0.16 (0.27)	0.12 (0.15)	0.08 (0.07)
200	--	0.26 (0.72)	0.20 (0.43)	0.14 (0.21)	0.10 (0.11)	0.07 (0.05)
500	0.30 (0.96)	0.24 (0.61)	0.18 (0.35)	0.13 (0.18)	0.09 (0.09)	0.07 (0.05)

NOTE: Values in the table are for two-tail significance = 0.05, power = 80 percent, $\rho_C = 0.15$, $R^2_{C(within)} = 0.4$, $\bar{T} = 0.5$, $\tau_* = 0.15$ and π = 0.6. Values in the table represent the MDESD. Values in parentheses represent R^2_W.

Appendix

Extensions to Cluster Randomized Multisite Trials

Extension to Multi-site Cluster Randomized Trials

Trial Design:

- Students in schools (clusters), schools in districts (sites)
- Treatment assigned at the school (cluster) level

Additional Challenges:

- Three sample sizes
  - Number of individuals per cluster
  - Number of clusters per site
  - Number of sites

Parameters Examined:

- MDES, MDESSD, MDESD

MDES for a Cross-Site Mean Effect Size

(Cluster Randomization)

Clusters per site	Sites					
	4	6	8	10	12	20
4	0.43	0.29	0.23	0.20	0.18	0.13
6	0.37	0.25	0.20	0.17	0.15	0.11
8	0.34	0.23	0.18	0.16	0.14	0.10
10	0.32	0.21	0.17	0.15	0.13	0.10
12	0.30	0.20	0.16	0.14	0.13	0.09
20	0.27	0.18	0.15	0.13	0.11	0.08

NOTES. Values in the table are for two-tail significance = 0.05, power = 80 percent , no level-one covariates, a single level two covariate, $R^2_{CC} = 0.74, \bar{T} = 0.5, \rho_{CS} = 0.07, \rho_{CC} = 0.10$, constant n = 200 within each cluster, constant J clusters per site and $\tau_* = 0.10$.

MDESSD for a Cross-Site Standard Deviation of Effect Sizes
(Cluster Randomization)

Clusters per site	Sites					
	4	6	8	10	12	20
4	0.60	0.43	0.35	0.31	0.28	0.23
6	0.41	0.31	0.26	0.23	0.21	0.17
8	0.35	0.26	0.22	0.19	0.18	0.14
10	0.31	0.23	0.19	0.17	0.16	0.13
12	0.27	0.20	0.17	0.15	0.14	0.11
20	0.21	0.15	0.13	0.12	0.11	0.09

NOTES. Values in the table are for two-tail significance = 0.05, power = 80 percent , no level-one covariates, a single level two covariate, $R^2_{CC} = 0.74, \bar{T} = 0.5, \rho_{CS} = 0.07, \rho_{CC} = 0.10$, constant n = 200 within each cluster and constant J clusters per site and $\pi = 0.60$.

MDESD for the Coefficient On a Binary Site-Level Moderator
(Cluster Randomization)

Clusters per site	Sites					
	4	6	8	10	12	20
4	--	--	--	--	--	--
6	--	--	--	--	--	0.20 (0.92)
8	--	--	--	--	--	0.18 (0.77)
10	--	--	--	--	--	0.17 (0.68)
12	--	--	--	--	0.19 (0.90)	0.16 (0.61)
20	--	--	0.20 (0.94)	0.18 (0.82)	0.17 (0.72)	0.14 (0.49)

NOTES. Values in the table are for two-tail significance = 0.05, power = 80 percent, no level-one covariates, a single level two covariate, $R^2_{CC} = 0.74, \bar{T} = 0.5, \rho_{CS} = 0.07, \rho_{CC} = 0.10$, constant n = 200 within each cluster, constant J clusters per site and $\tau_* = 0.10$. Values in the table represent the MDESD. Values in parentheses represent R^2_W.

Conclusions

Challenges (Individual Randomization)

-MDES, total **number of sites** is key sample size

-MDESSD, total **number of individuals per site** is key sample size

-MDESD, total **number of sites** is key sample size (and there is a maximum possible value for the difference)

Challenges (Cluster Randomization)

-MDES, total **number of sites** is key sample size

-MDESSD, total **number of clusters per site** is key sample size

-MDESD, total **number of sites** is key sample size (and there is a maximum possible value for the difference)

- **Bottom line** = Power is a substantial limitation for this design