Statistical Power Analysis for Multi-level Models

Miao ("Michelle") Yang

Department of Psychology
Quantitative Study Group

Mar 19 2015

Outline of the Talk

1. Introduction to multilevel designs
2. Statistical power analysis for multilevel models
3. Software

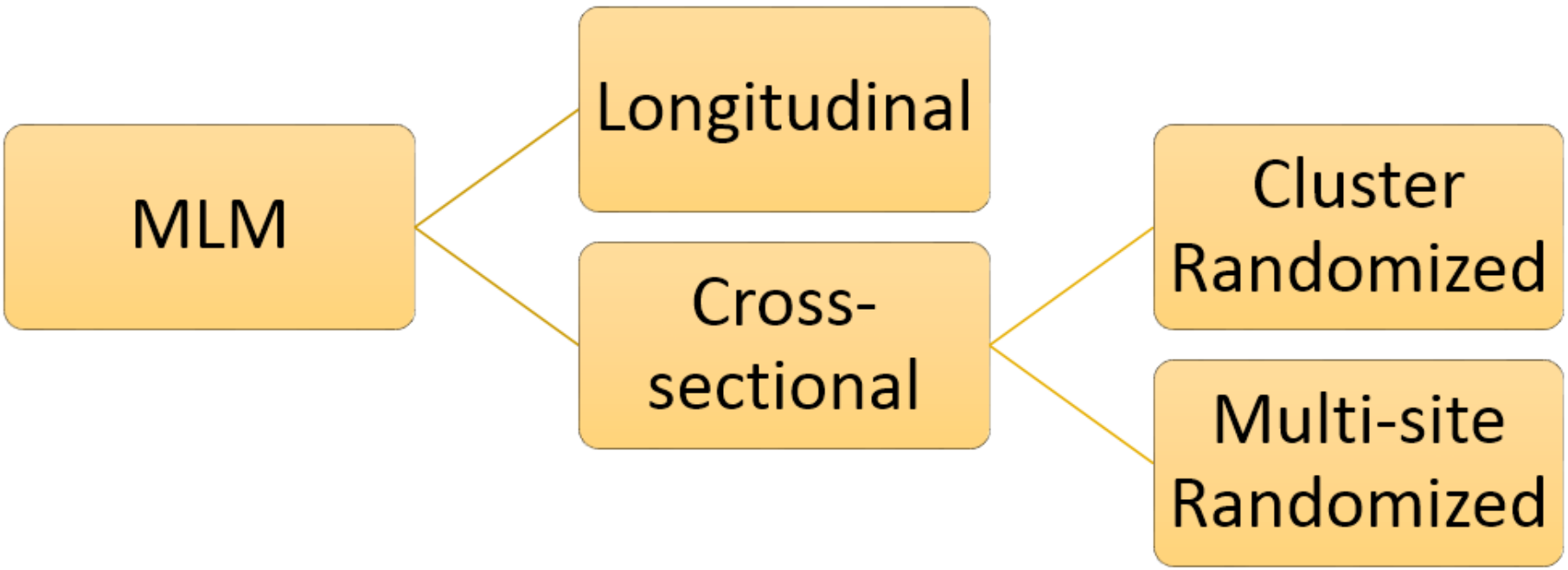
Outline of the Talk

1. Introduction to multilevel designs
2. Statistical power analysis for multilevel models
3. Software

Outline of the Talk

1. Introduction to multilevel designs
2. Statistical power analysis for multilevel models
3. Software

Multilevel Designs



Multilevel Modeling — When?

In educational studies, the total sample size is often a combination of students sampled from different classrooms or schools. When data exhibit such nested structure, multilevel modeling can be conducted.

Student (ID)	School (Name)	Verbal Score
1	Potato	88
2	Potato	85
3	Potato	92
4	Tomato	76
5	Tomato	78
6	Tomato	80
⋮	⋮	
60	Sheep	77

Table: An example of nested data

Multilevel Modeling — Why?

When data are nested, it is natural that the individuals within the same cluster (e.g., school) are correlated, which violates one of the assumptions of traditional models such as multiple regression and ANOVA. As a consequence, traditional models will produce biased estimates of parameter standard errors, and thus lead to significance tests with inflated type I error rates (e.g., Hox, 1998).

Advantages of using multilevel modeling:

- Handle nested data
- Allow us to know both individual and cluster differences
- More powerful

Multilevel Modeling (CRT vs MRT)

CRT:

- The entire site (school) is randomly assigned to treatment or control.
- Avoids a possible "spill over" effect within schools.

MRT:

- Students within schools are randomly assigned.
- More convenient and economical because we have a larger pool.
- Easy to manage because each cluster follows the same study design.

Power Analysis for CRT (1 treatment & 1 control)

$$Y_{ij} = \beta_{0j} + e_{ij},\ e_{ij} \sim N(0, \sigma_W^2)$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01} X_j + u_{0j},\ u_{0j} \sim N(0, \sigma_B^2)$$

$i = 1, 2, ..., n$ (individual); $j = 1, 2, ...J$ (cluster);

X_j: treatment indicator of cluster j, $X_j = \begin{cases} 0.5 & \text{treatment} \\ -0.5 & \text{control} \end{cases}$

γ_{00} : grand mean;

γ_{01} : treatment main effect (i.e., $\mu_D = \mu_T - \mu_C$)

β_{0j} : cluster mean

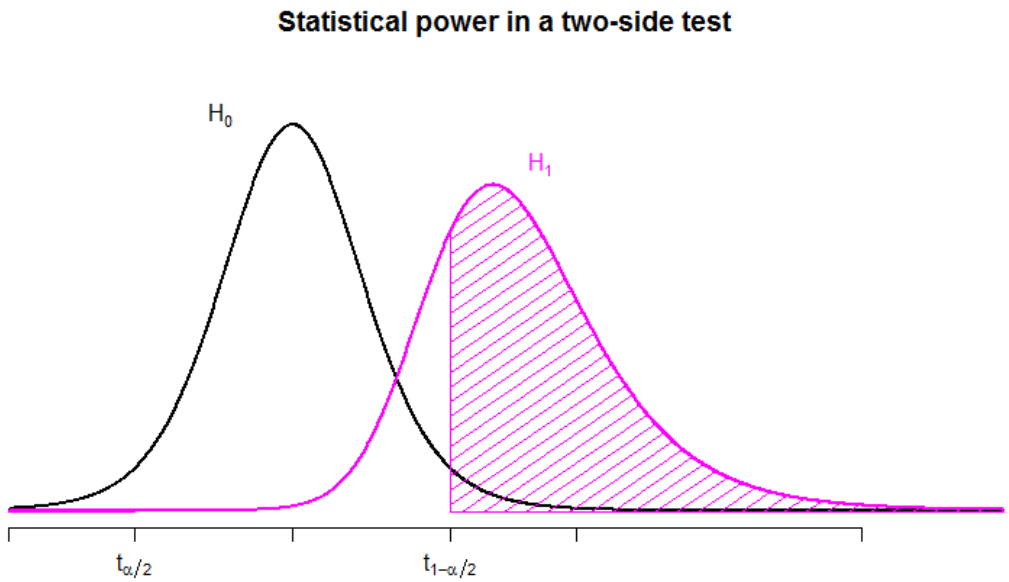
σ_W^2 : within-cluster variance; σ_B^2 : between-cluster variance

Power Analysis for CRT (1 treatment & 1 control)

Test treatment main effect $H_0 : \gamma_{01} = 0$:

$$T = \frac{\hat{\gamma}_{01}}{\sqrt{Var(\hat{\gamma}_{01})}} = \frac{\bar{Y}_{..}^{T} - \bar{Y}_{..}^{C}}{\sqrt{4(\sigma_B^2 + \sigma_W^2/n)/J}}$$

Under H_0 : $T \sim t_{J-2}$.
Under H_1 : $T \sim t_{J-2,\lambda}$.



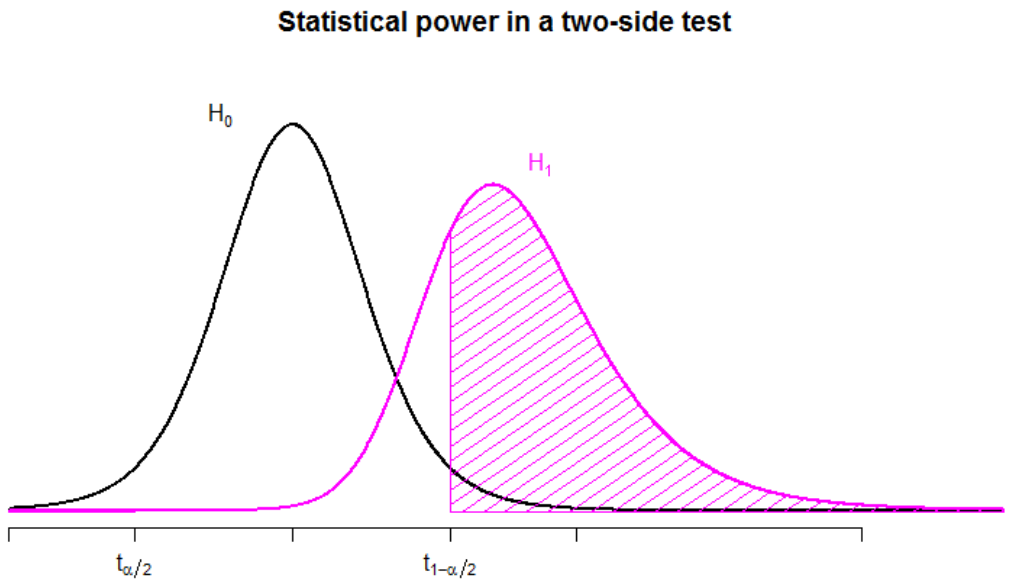
$$\begin{aligned}\text{Power} &= P(\text{reject } H_0 | H_1 \text{ true}) \\ &= \begin{cases} 1 - P[T_{J-2,\lambda} < t_0] + P[T_{J-2,\lambda} \leq -t_0] & \text{two} - \text{sided}; \\ 1 - P[T_{J-2,\lambda} < t_0] & \text{one} - \text{sided}, \end{cases}\end{aligned}$$

Power Analysis for CRT (1 treatment & 1 control)

Test treatment main effect $H_0 : \gamma_{01} = 0$:

$$T = \frac{\hat{\gamma}_{01}}{\sqrt{Var(\hat{\gamma}_{01})}} = \frac{\bar{Y}_{..}^{T} - \bar{Y}_{..}^{C}}{\sqrt{4(\sigma_B^2 + \sigma_W^2/n)/J}}$$

Under H_0 : $T \sim t_{J-2}$.
Under H_1 : $T \sim t_{J-2,\lambda}$.



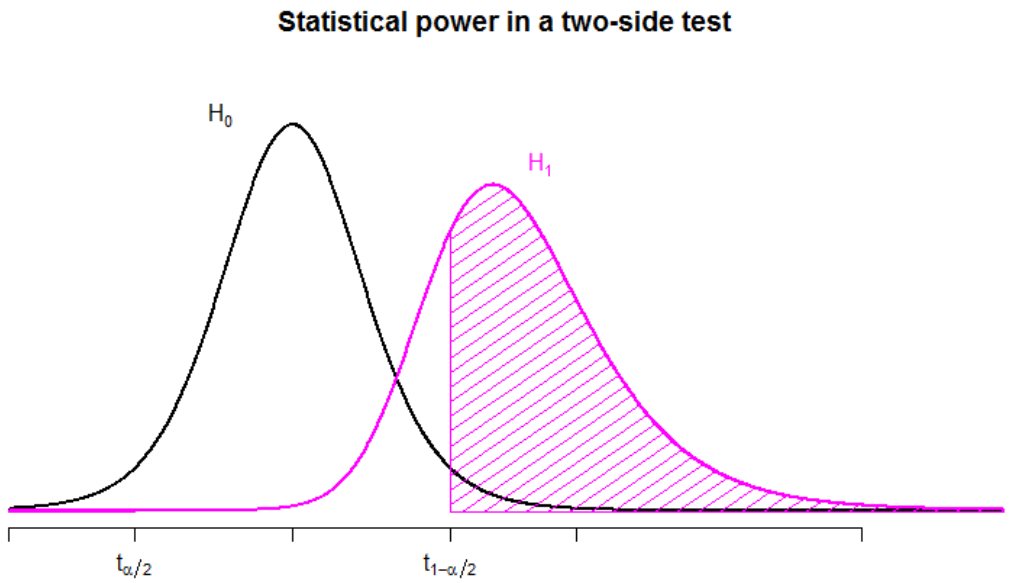
$$\begin{aligned}\text{Power} &= P(\text{reject } H_0 | H_1 \text{ true}) \\ &= \begin{cases} 1 - P[T_{J-2,\lambda} < t_0] + P[T_{J-2,\lambda} \leq -t_0] & \text{two} - \text{sided}; \\ 1 - P[T_{J-2,\lambda} < t_0] & \text{one} - \text{sided}, \end{cases}\end{aligned}$$

Power Analysis for CRT (1 treatment & 1 control)

Test treatment main effect $H_0 : \gamma_{01} = 0$:

$$T = \frac{\hat{\gamma}_{01}}{\sqrt{Var(\hat{\gamma}_{01})}} = \frac{\bar{Y}_{..}^{T} - \bar{Y}_{..}^{C}}{\sqrt{4(\sigma_B^2 + \sigma_W^2/n)/J}}$$

Under H_0 : $T \sim t_{J-2}$.
Under H_1 : $T \sim t_{J-2,\lambda}$.

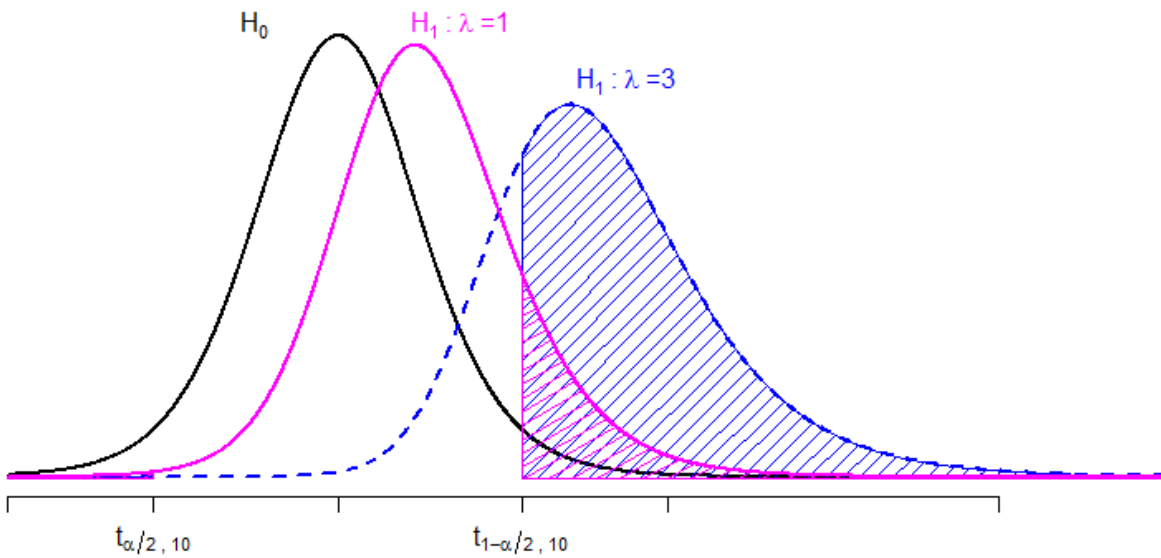


$$\begin{aligned}\text{Power} &= P(\text{reject } H_0 | H_1 \text{ true}) \\ &= \begin{cases} 1 - P[T_{J-2,\lambda} < t_0] + P[T_{J-2,\lambda} \leq -t_0] & \text{two} - \text{sided}; \\ 1 - P[T_{J-2,\lambda} < t_0] & \text{one} - \text{sided}, \end{cases}\end{aligned}$$

Power Analysis for CRT (1 treatment & 1 control)

$$\lambda = \frac{\mu_D}{\sqrt{4(\sigma_B^2 + \frac{\sigma_W^2}{n})/J}}.$$

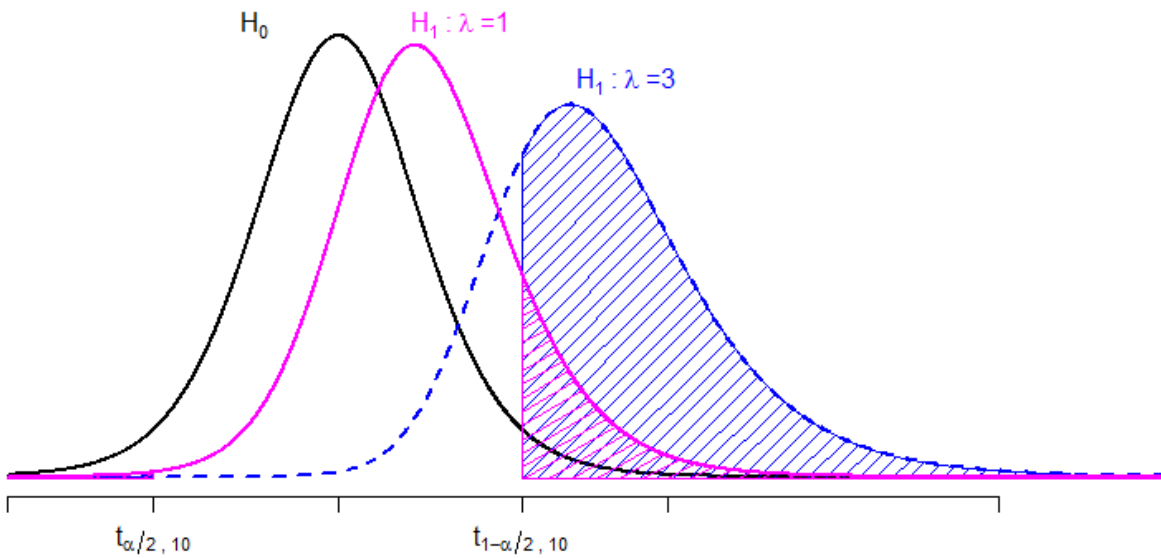
- As λ increases, power increases.
- λ is a function of μ_D, n, J, σ_B^2 and σ_W^2.
- To give more meaningful definition, we can reparameterize λ in terms of effect size and intra-class correlation (ICC).



Power Analysis for CRT (1 treatment & 1 control)

$$\lambda = \frac{\mu_D}{\sqrt{4(\sigma_B^2 + \frac{\sigma_W^2}{n})/J}}.$$

- As λ increases, power increases.
- λ is a function of μ_D, n, J, σ_B^2 and σ_W^2.
- To give more meaningful definition, we can reparameterize λ in terms of effect size and intra-class correlation (ICC).



ICC in CRT

The intra-class correlation (ICC) quantifies the degree to which two randomly drawn observations within a cluster are correlated. In CRT, the ICC is defined as

$$\rho = corr(Y_{ij}, Y_{i'j}) = \frac{\sigma_B^2}{\sigma_B^2 + \sigma_W^2} = \frac{\sigma_B^2}{\sigma_T^2}.$$

- The proportion of total variance that is accounted for by clustering.
- $\rho = 0$, no between cluster variation.
- As ρ increases, more variation is due to between-cluster variability.
- For school-based data sets, ρ usually ranges between 0.10 to 0.30. (Bloom, Bos & Lee, 1999; Hedges & Hedberg, 2007)

Effect Size in CRT (1 treatment & 1 control)

The effect sizes used in educational and psychological research are typically standardized mean differences. Possible definitions for the effect size in CRT (Hedges, 2007):

- $f = \mu_D/\sigma_W$. This effect size might be of interest in a meta-analysis where the studies being compared are single-site studies.
- $f = \mu_D/\sigma_B$. This effect size might be of interest in a meta-analysis where the other studies are multisite studies that have been analyzed by using cluster means as the unit of analysis.
- $f = \mu_D/\sqrt{\sigma_B^2 + \sigma_W^2}$. This effect size might be of interest in a meta-analysis where the other studies are multisite studies or studies that sample from a broader population but do not include clusters. ♡♡♡

Effect Size in CRT (1 treatment & 1 control)

The effect sizes used in educational and psychological research are typically standardized mean differences. Possible definitions for the effect size in CRT (Hedges, 2007):

- $f = \mu_D/\sigma_W$. This effect size might be of interest in a meta-analysis where the studies being compared are single-site studies.
- $f = \mu_D/\sigma_B$. This effect size might be of interest in a meta-analysis where the other studies are multisite studies that have been analyzed by using cluster means as the unit of analysis.
- $f = \mu_D/\sqrt{\sigma_B^2 + \sigma_W^2}$. This effect size might be of interest in a meta-analysis where the other studies are multisite studies or studies that sample from a broader population but do not include clusters. ♡♡♡

Effect Size in CRT (1 treatment & 1 control)

The effect sizes used in educational and psychological research are typically standardized mean differences. Possible definitions for the effect size in CRT (Hedges, 2007):

- $f = \mu_D/\sigma_W$. This effect size might be of interest in a meta-analysis where the studies being compared are single-site studies.
- $f = \mu_D/\sigma_B$. This effect size might be of interest in a meta-analysis where the other studies are multisite studies that have been analyzed by using cluster means as the unit of analysis.
- $f = \mu_D/\sqrt{\sigma_B^2 + \sigma_W^2}$. This effect size might be of interest in a meta-analysis where the other studies are multisite studies or studies that sample from a broader population but do not include clusters. ♡♡♡

Power Analysis for CRT (1 treatment & 1 control)

Redefine λ in standardized notation:

$$\lambda = \frac{\mu_D}{\sqrt{4(\sigma_B^2 + \frac{\sigma_W^2}{n})/J}} = \frac{\sqrt{J}f}{\sqrt{4(\rho + \frac{1-\rho}{n})}}.$$

Now, λ is a function of n, J, f and ρ.

- As J or n increases, λ increases and thus power increases.
- As f increases, λ increases and thus power increases.
- As ρ increases, λ decreases and thus power decreases.

Power Analysis for CRT (2 treatments & 1 control)

$$Y_{ij} = \beta_{0j} + e_{ij},\ e_{ij} \sim N(0, \sigma_W^2)$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01}X_{1j} + \gamma_{02}X_{2j} + u_{0j},\ u_{0j} \sim N(0, \sigma_B^2)$$

$$X_{1j} = \begin{cases} 1/3 & \text{treatment1} \\ 1/3 & \text{treatment2}; \\ -1 & \text{control} \end{cases} \quad X_{2j} = \begin{cases} 1/2 & \text{treatment1} \\ -1/2 & \text{treatment2} \\ 0 & \text{control} \end{cases}$$

β_{0j} : cluster mean; γ_{00} : grand mean

γ_{01} : mean difference between the average of the two treatments and the control

γ_{02}: mean difference between the two treatments

$$Y_{ij} = \gamma_{00} + \gamma_{01}X_{1j} + \gamma_{02}X_{2j} + u_{0j} + e_{ij}$$

$$\begin{cases} \mu_{T1} = \gamma_{00} + \frac{1}{3}\gamma_{01} + \frac{1}{2}\gamma_{02} \\ \mu_{T2} = \gamma_{00} + \frac{1}{3}\gamma_{01} - \frac{1}{2}\gamma_{02} \\ \mu_C = \gamma_{00} - \gamma_{01} \end{cases} \implies \begin{cases} 0.5(\mu_{T1} + \mu_{T2}) - \mu_C = \gamma_{01} \\ \mu_{T1} - \mu_{T2} = \gamma_{02} \end{cases}$$

Power Analysis for CRT (2 treatments & 1 control)

We might be interested in three different types of test:

1. Test treatment main effect: $H_0: \gamma_{01} = 0 \Leftrightarrow \mu_D = 0.5(\mu_{T1} + \mu_{T2}) - \mu_C = 0$
   Under $H_0: T_1 \sim t_{J-3}$. Under $H_1: T_1 \sim t_{J-3,\lambda_1}$, where

$$\lambda_1 = \frac{\sqrt{J} f_1}{\sqrt{4.5(\rho + \frac{1-\rho}{n})}} \text{ and } f_1 = \frac{0.5(\mu_{T1} + \mu_{T2}) - \mu_C}{\sqrt{\sigma_B^2 + \sigma_W^2}}.$$

2. Comparing the two treatments: $H_0: \gamma_{02} = 0 \Leftrightarrow \mu_D = \mu_{T1} - \mu_{T2} = 0$
   Under $H_0: T_2 \sim t_{J-3}$. Under $H_1: T_2 \sim t_{J-3,\lambda_2}$, where

$$\lambda_2 = \frac{\sqrt{J} f_2}{\sqrt{6(\rho + \frac{1-\rho}{n})}} \text{ and } f_2 = \frac{\mu_{T1} - \mu_{T2}}{\sqrt{\sigma_B^2 + \sigma_W^2}}.$$

3. Ominibus test: $H_0: \gamma_{01} = \gamma_{02} = 0 \Leftrightarrow \mu_{T1} = \mu_{T2} = \mu_C$
   Under $H_0: F \sim F_{2,J-3}$. Under $H_1: F \sim F_{2,J-3,\lambda}$, where $\lambda = \lambda_1^2 + \lambda_2^2$.

Power Analysis for CRT (2 treatments & 1 control)

We might be interested in three different types of test:

1. Test treatment main effect: $H_0: \gamma_{01} = 0 \Leftrightarrow \mu_D = 0.5(\mu_{T1} + \mu_{T2}) - \mu_C = 0$
   Under $H_0: T_1 \sim t_{J-3}$. Under $H_1: T_1 \sim t_{J-3,\lambda_1}$, where

$$\lambda_1 = \frac{\sqrt{J} f_1}{\sqrt{4.5(\rho + \frac{1-\rho}{n})}} \text{ and } f_1 = \frac{0.5(\mu_{T1} + \mu_{T2}) - \mu_C}{\sqrt{\sigma_B^2 + \sigma_W^2}}.$$

2. Comparing the two treatments: $H_0: \gamma_{02} = 0 \Leftrightarrow \mu_D = \mu_{T1} - \mu_{T2} = 0$
   Under $H_0: T_2 \sim t_{J-3}$. Under $H_1: T_2 \sim t_{J-3,\lambda_2}$, where

$$\lambda_2 = \frac{\sqrt{J} f_2}{\sqrt{6(\rho + \frac{1-\rho}{n})}} \text{ and } f_2 = \frac{\mu_{T1} - \mu_{T2}}{\sqrt{\sigma_B^2 + \sigma_W^2}}.$$

3. Ominibus test: $H_0: \gamma_{01} = \gamma_{02} = 0 \Leftrightarrow \mu_{T1} = \mu_{T2} = \mu_C$
   Under $H_0: F \sim F_{2,J-3}$. Under $H_1: F \sim F_{2,J-3,\lambda}$, where $\lambda = \lambda_1^2 + \lambda_2^2$.

Power Analysis for CRT (2 treatments & 1 control)

We might be interested in three different types of test:

1. Test treatment main effect: $H_0: \gamma_{01} = 0 \Leftrightarrow \mu_D = 0.5(\mu_{T1} + \mu_{T2}) - \mu_C = 0$
   Under $H_0: T_1 \sim t_{J-3}$. Under $H_1: T_1 \sim t_{J-3,\lambda_1}$, where

$$\lambda_1 = \frac{\sqrt{J} f_1}{\sqrt{4.5(\rho + \frac{1-\rho}{n})}} \text{ and } f_1 = \frac{0.5(\mu_{T1} + \mu_{T2}) - \mu_C}{\sqrt{\sigma_B^2 + \sigma_W^2}}.$$

2. Comparing the two treatments: $H_0: \gamma_{02} = 0 \Leftrightarrow \mu_D = \mu_{T1} - \mu_{T2} = 0$
   Under $H_0: T_2 \sim t_{J-3}$. Under $H_1: T_2 \sim t_{J-3,\lambda_2}$, where

$$\lambda_2 = \frac{\sqrt{J} f_2}{\sqrt{6(\rho + \frac{1-\rho}{n})}} \text{ and } f_2 = \frac{\mu_{T1} - \mu_{T2}}{\sqrt{\sigma_B^2 + \sigma_W^2}}.$$

3. Ominibus test: $H_0: \gamma_{01} = \gamma_{02} = 0 \Leftrightarrow \mu_{T1} = \mu_{T2} = \mu_C$
   Under $H_0: F \sim F_{2,J-3}$. Under $H_1: F \sim F_{2,J-3,\lambda}$, where $\lambda = \lambda_1^2 + \lambda_2^2$.

Power Analysis for MRT (1 treatment & 1 control)

Let's move on to multisite randomized trials with 1 treatment and 1 control.

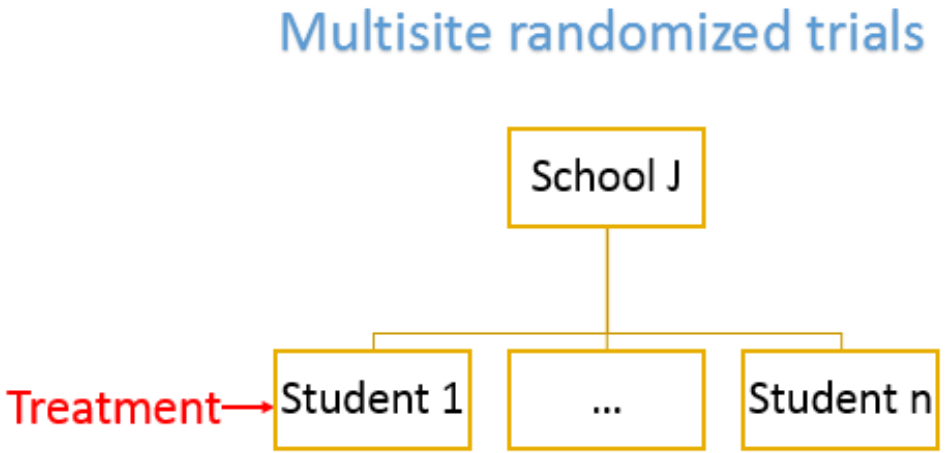
$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + e_{ij},\ e_{ij} \sim N(0, \sigma^2)$$

$$\beta_{0j} = \gamma_{00} + u_{0j},\ \beta_{1j} = \gamma_{10} + u_{1j}.\ \begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N(\mathbf{0}, \begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{bmatrix})$$

$i = 1, 2, ..., n$ (individual); $j = 1, 2, ...J$ (site);
X_{ij}: indicator of treatment assignment with

$$X_{ij} = \begin{cases} 0.5 & \text{treatment} \\ -0.5 & \text{control} \end{cases}$$

β_{0j} : mean at the jth site
β_{1j} : mean difference between treatment and control at the jth site
γ_{00} : grand mean;
γ_{10} : treatment main effect
σ^2 : between-person variation
τ_{00} : site variability
τ_{11} : variance of site-specfic treatment effects



Power Analysis for MRT (1 treatment & 1 control)

Let's move on to multisite randomized trials with 1 treatment and 1 control.

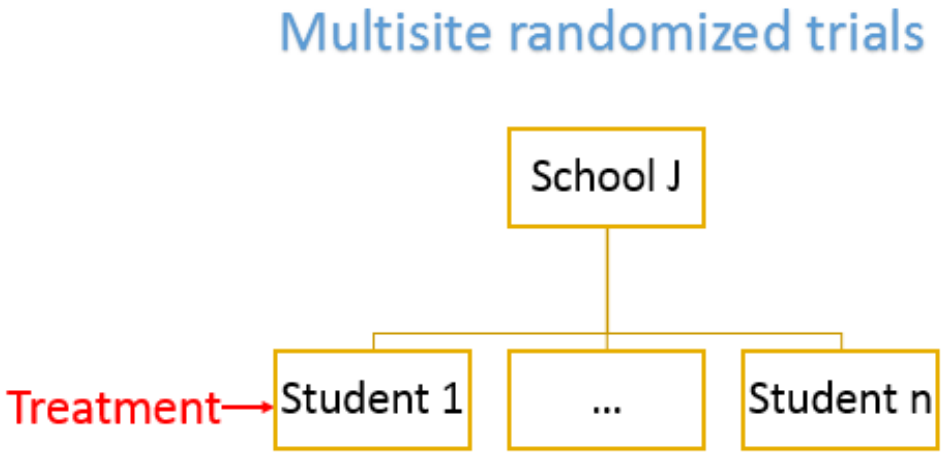
$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + e_{ij},\ e_{ij} \sim N(0, \sigma^2)$$

$$\beta_{0j} = \gamma_{00} + u_{0j},\ \beta_{1j} = \gamma_{10} + u_{1j}.\ \begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N(\mathbf{0}, \begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{bmatrix})$$

$i = 1, 2, ..., n$ (individual); $j = 1, 2, ...J$ (site);
X_{ij}: indicator of treatment assignment with

$$X_{ij} = \begin{cases} 0.5 & \text{treatment} \\ -0.5 & \text{control} \end{cases}$$

β_{0j} : mean at the jth site
β_{1j} : mean difference between treatment and control at the jth site
γ_{00} : grand mean;
γ_{10} : treatment main effect
σ^2 : between-person variation
τ_{00} : site variability
τ_{11} : variance of site-specfic treatment effects



Power Analysis for MRT (1 treatment & 1 control)

Test treatment main effect $H_0 : \gamma_{10} = 0 \Leftrightarrow \mu_D = \mu_T - \mu_C = 0$

Under $H_0 : T \sim t_{J-1}$. Under $H_1 : T \sim t_{J-1,\lambda}$, where

$$\lambda = \frac{\sqrt{J}\mu_D}{\sqrt{4\sigma^2/n + \tau_{11}}}.$$

$$\text{Power} = \begin{cases} 1 - P[T_{J-1,\lambda} < t_0] + P[T_{J-1,\lambda} \leq -t_0] & \text{two} - \text{sided} \\ 1 - P[T_{J-1,\lambda} < t_0] & \text{one} - \text{sided} \end{cases}$$

Following Raudenbush & Liu (2000), we define the effect size as $f = \frac{\mu_D}{\sqrt{\sigma^2}}$. Thus,

$$\lambda = \frac{\sqrt{J}f}{\sqrt{4/n + \tau_{11}/\sigma^2}}.$$

Power increases as

- the effect size (f) increases;
- the number of sites (J) or the number of individuals per site (n) increases;
- the variance of the treatment effect (τ_{11}) decreases;
- between-person variation (σ^2) increases.

Power Analysis for MRT (2 treatments & 1 control)

MRT (2 treatments & 1 control):

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + \beta_{2j}X_{2ij} + e_{ij}$$

$$\beta_{0j} = \gamma_{00} + u_{0j},\ \beta_{1j} = \gamma_{10} + u_{1j},\ \beta_{2j} = \gamma_{20} + u_{2j}$$

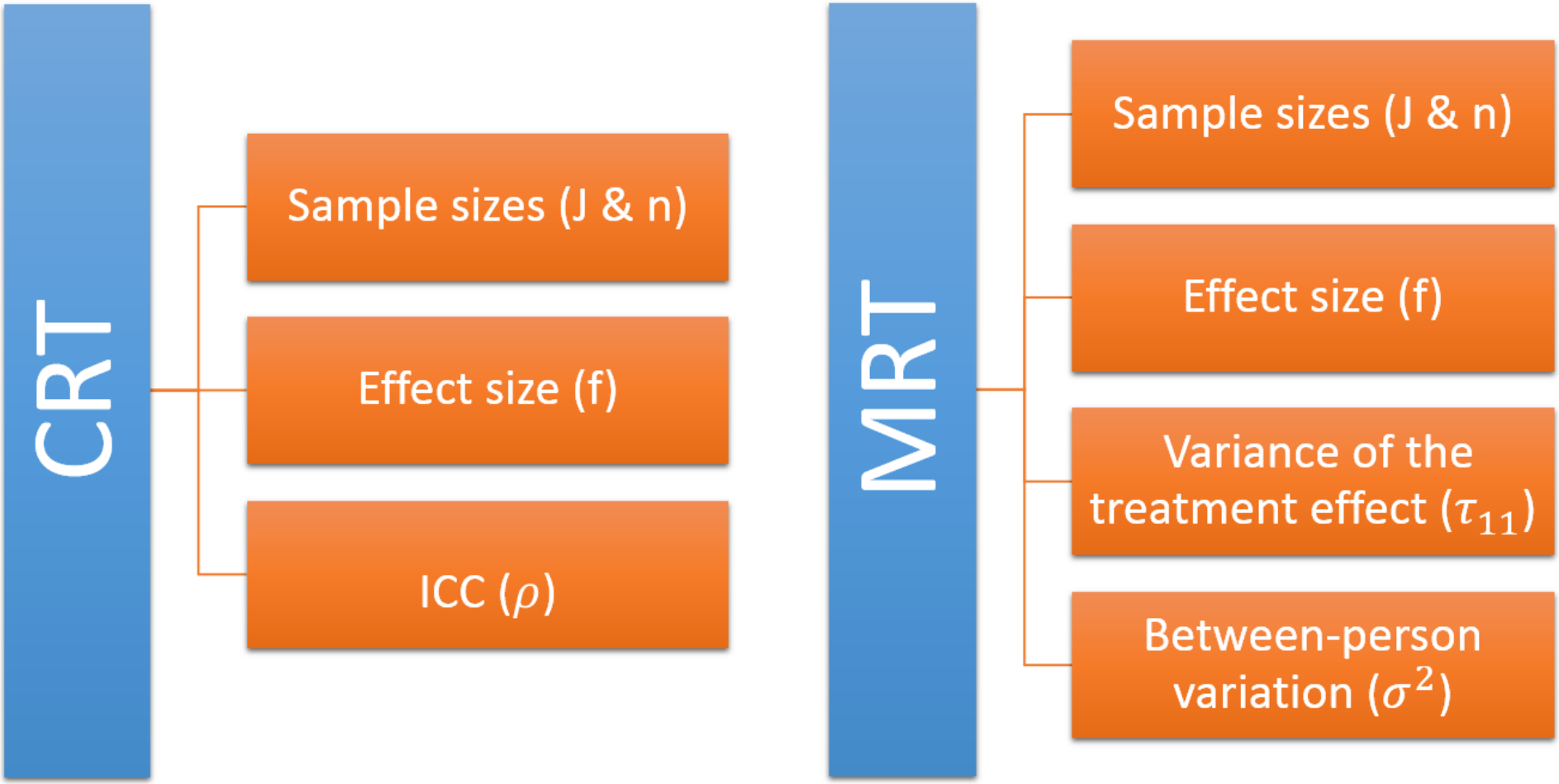
$$X_{1ij} = \begin{cases} 1/3 & \text{treatment1} \\ 1/3 & \text{treatment2}; \\ -1 & \text{control} \end{cases} \quad X_{2ij} = \begin{cases} 1/2 & \text{treatment1} \\ -1/2 & \text{treatment2} \\ 0 & \text{control} \end{cases}$$

(1) Test treatment main effect: $H_0: \gamma_{10} = 0 \Leftrightarrow \frac{1}{2}(\mu_{T1} + \mu_{T2}) = \mu_C$

(2) Comparing the two treatments: $H_0: \gamma_{20} = 0 \Leftrightarrow \mu_{T1} = \mu_{T2}$

(3) Ominibus test: $H_0: \gamma_{10} = \gamma_{20} = 0 \Leftrightarrow \mu_{T1} = \mu_{T2} = \mu_C$

Revisit


CRT
Sample sizes (J & n)
Effect size (f)
ICC (ρ)
MRT
Sample sizes (J & n)
Effect size (f)
Variance of the treatment effect (τ_{11})
Between-person variation (σ^2)

Software: WebPower

- CRT with 1 treatment and 1 control:

http://webpower.psychstat.org/models/mlm01/

- CRT with 2 treatments and 1 control:

http://webpower.psychstat.org/models/mlm02/

- MRT with 1 treatment and 1 control:

http://webpower.psychstat.org/models/mlm03/

- MRT with 2 treatments and 1 control:

http://webpower.psychstat.org/models/mlm04/

Software: WebPower

- CRT with 1 treatment and 1 control:

http://webpower.psychstat.org/models/mlm01/

- CRT with 2 treatments and 1 control:

http://webpower.psychstat.org/models/mlm02/

- MRT with 1 treatment and 1 control:

http://webpower.psychstat.org/models/mlm03/

- MRT with 2 treatments and 1 control:

http://webpower.psychstat.org/models/mlm04/

Application of WebPower

Example 1. A researcher plans to collect data from 20 clinics to examine the effect of certain behavioral therapies on recovering from anorexia. At each clinic, 30 anorexic girls will be randomly assigned to therapy 1, therapy 2, or the control group. Previous research suggests the therapy 1 might lead to an increase of 0.5 in BMI and therapy 2 might lead to an increase of 0.8 in BMI. Further, the between-person variation is 2.25 and the variance in treatment effects across sites is 0.4. What's the power for testing the treatment main effect ?

- Sample size $= 30$
- Effect size $= \frac{(0.5+0.8)/2}{\sqrt{2.25}} = 0.43$
- Number of clusters $= 20$
- Variance in treatment effects across sites $= 0.4$
- Between-person variation $= 2.25$

http://webpower.psychstat.org/models/mlm04/

Application of WebPower

Example 2. A group of educational researchers developed a new teaching method to help students improve their memory abilities. They decide to randomly assign 20 schools to either the new method or the standard method and test students on memory ability from these 20 schools. Suppose the new method might have a medium effect size and the intraclass correlation is 0.10. How many students in each school will be needed to obtain a power of 0.8?

- Effect size $= 0.5$
- Number of clusters $= 20$
- ICC $= 0.10$
- Power $= 0.8$

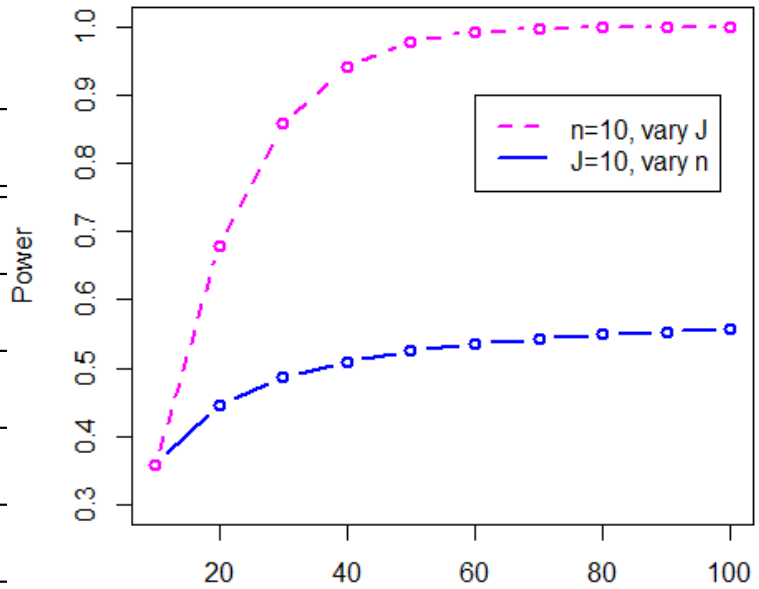
http://webpower.psychstat.org/models/mlm01/

Application of WebPower

Whether sample size (n) or cluster size (J) is more crucial in increasing the power for CRT?

Set effect size = 0.5, ICC = 0.1, signifcance level = 0.05.

	Varying n			Varying J		
Total n	n	J	Power	n	J	Power
100	10	10	0.359	10	10	0.359
200	20	10	0.447	10	20	0.680
300	30	10	0.487	10	30	0.858
400	40	10	0.510	10	40	0.942
500	50	10	0.525	10	50	0.978
600	60	10	0.535	10	60	0.992
700	70	10	0.543	10	70	0.997
800	80	10	0.549	10	80	0.999
900	90	10	0.553	10	90	1.000
1000	100	10	0.557	10	100	1.000



Comparison With Other Software

Optimal Design (Raudenbush et al., 2011)
— Graphic-based power analysis

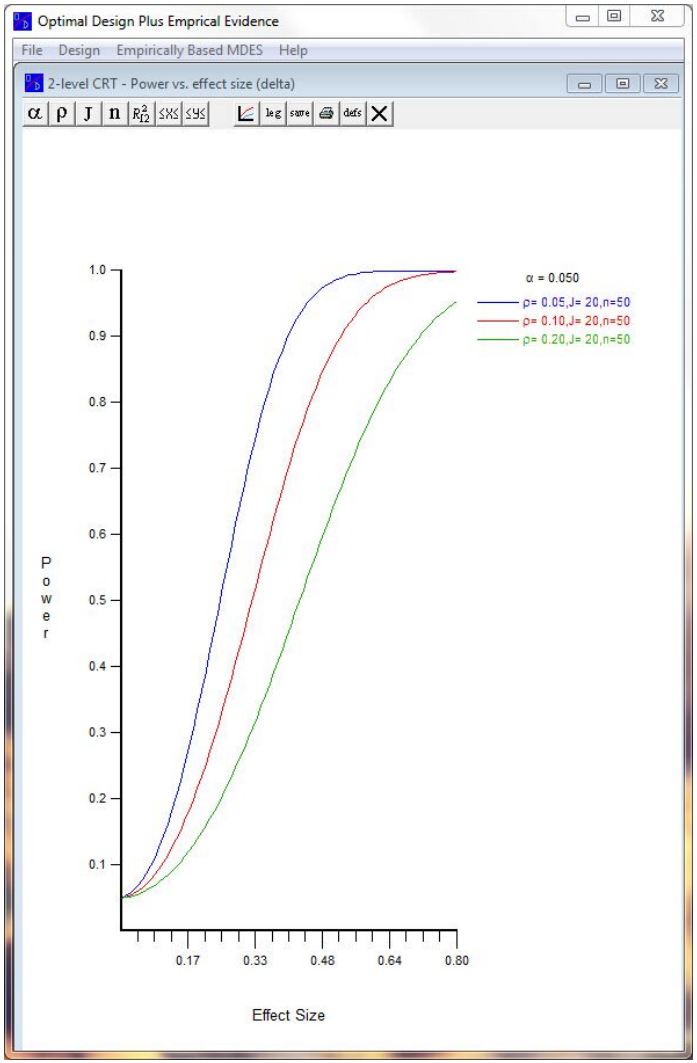
- Pros: Available for three-level designs.
- Cons: No exact value for power. Only considers 1 treatment and 1 control.

R function MRTpower() (Usami,2014):

- Pros: Extended to three-level designs and unbalanced designs.
- Cons: Only estimates sample size. To do power or effect size calculation, readers should understand the technical details of the paper and write their own syntax.

What can be improved in WebPower?

- Generalize to three-level and unbalanced designs.



Comparison With Other Software

Optimal Design (Raudenbush et al., 2011)
— Graphic-based power analysis

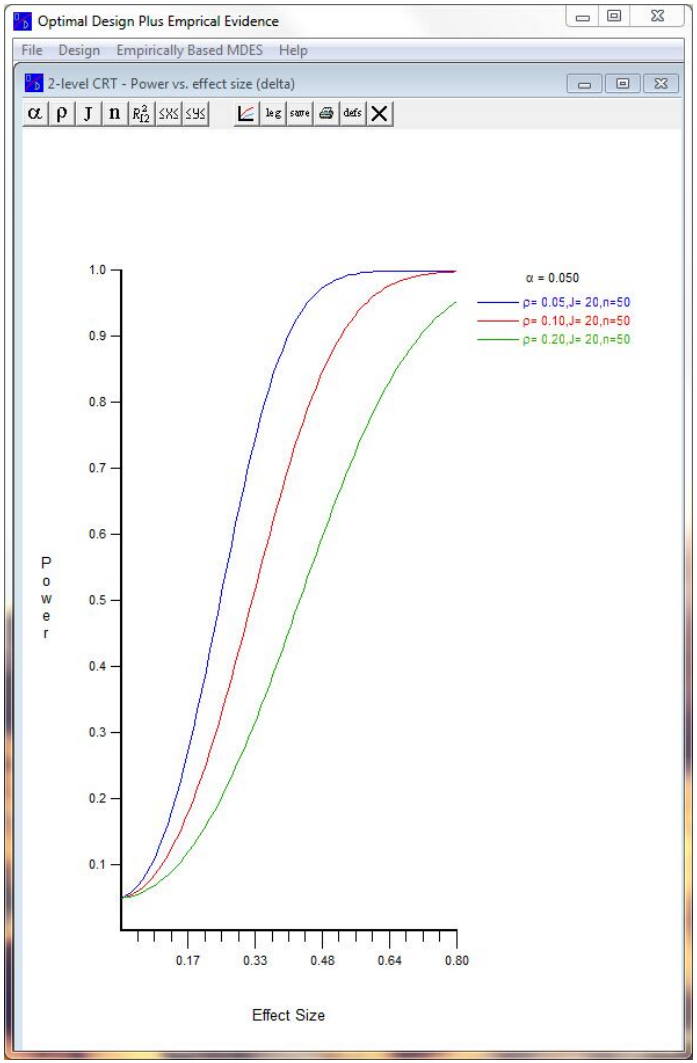
- Pros: Available for three-level designs.
- Cons: No exact value for power. Only considers 1 treatment and 1 control.

R function MRTpower() (Usami,2014):

- Pros: Extended to three-level designs and unbalanced designs.
- Cons: Only estimates sample size. To do power or effect size calculation, readers should understand the technical details of the paper and write their own syntax.

What can be improved in WebPower?

- Generalize to three-level and unbalanced designs.



References

Bloom, H. S., Johannes, M. B. & Lee, S-W. (1999). "Using Cluster Random Assignment to Measure Program Impacts: Statistical Implications for the Evaluation of Education Programs." Evaluation Review 23(4): 445-69.

Hedges, L. V. (2007). Effect sizes in cluster-randomized designs. *Journal of Educational and Behavioral Statistics, 32,* 341-370.

Hedges, L. V. & Hedberg, E. C. (2007). Intraclass correlations for planning group randomized experiments in rural education. *Journal of Research in Rural Education, 22(10).*

Hox, J. (1998). Multilevel modeling: When and why. *Classification, data analysis, and data highways,* 147-154.

Raudenbush, S. W. & Liu, X. (2000). Statistical power and optimal design for multisite randomized trials. *Psychological Methods,* Vol 5(2), 199-213.

Raudenbush, S. W., et al. (2011). Optimal Design Software for Multi-level and Longitudinal Research (Version 3.01) [Software]. Available from www.wtgrantfoundation.org.

Usami, S. (2014). Generalized sample size determination formulas for experimental research with hierarchical data. *Behavior Research Methods,* Vol 46(2), 346-356.

Thanks

Johnny & Ke-Hai

Supported by the Department of Education (R305D140037)

Gabrielle, Agung & Haiyan

All of you