Optimal Design of Multisite-Randomized Trials Investigating Mediation Effects Under Unequal Costs

Zuchao "William" Shen
University of Kansas

Wei Li
University of Florida

Benjamin Kelcey
University of Cincinnati

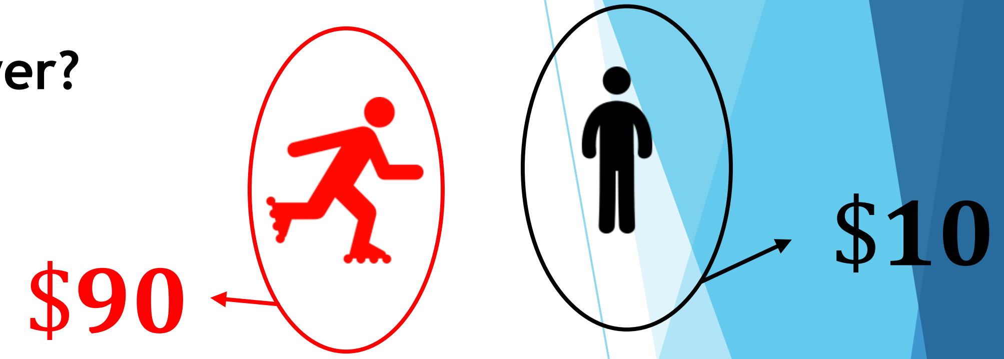
Walter Leite
University of Florida

Huibin Zhang
University of Tennessee, Knoxville

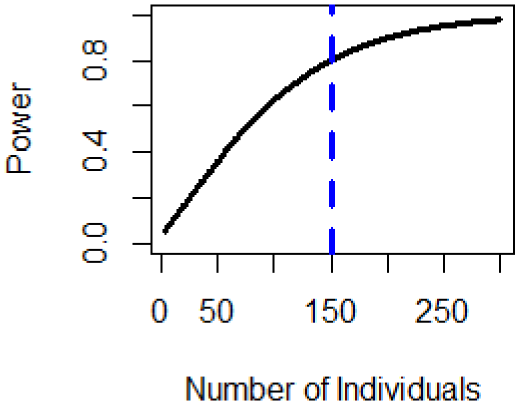
This presentation is based on research supported by the Spencer Foundation

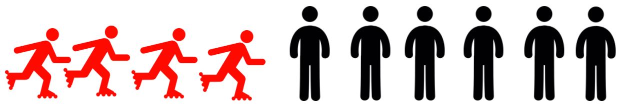
An Example About Sample Allocation, Statical Power & Budget

- How many individuals are needed to have .80% power?
  - Effect size: $d = 0.33$
  - Proportion of variance explained: $R^2 = 0.50$

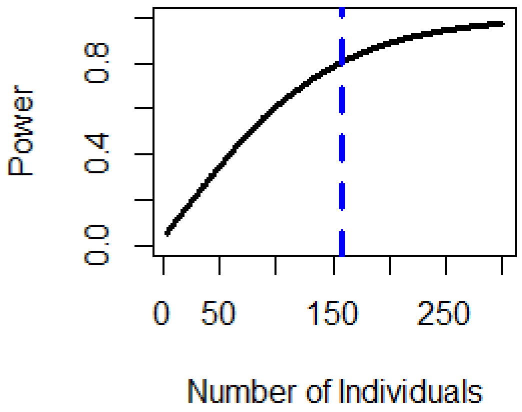


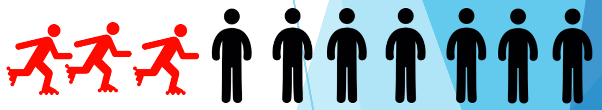
(1) $7,500



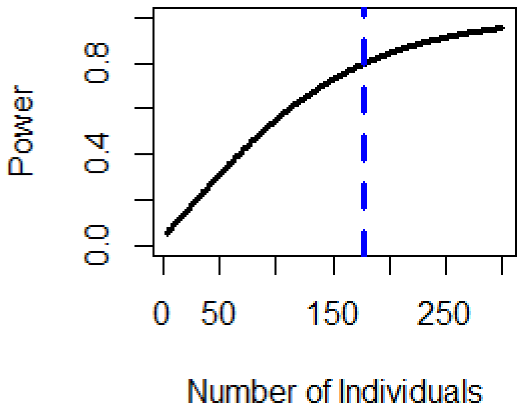


(2) $6,600

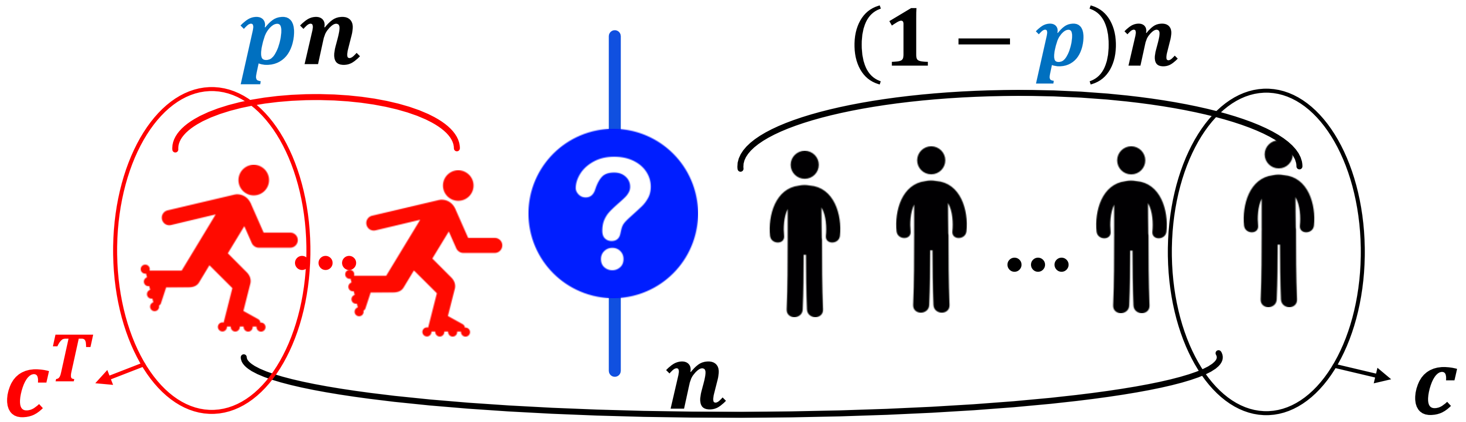




(3) $6,050



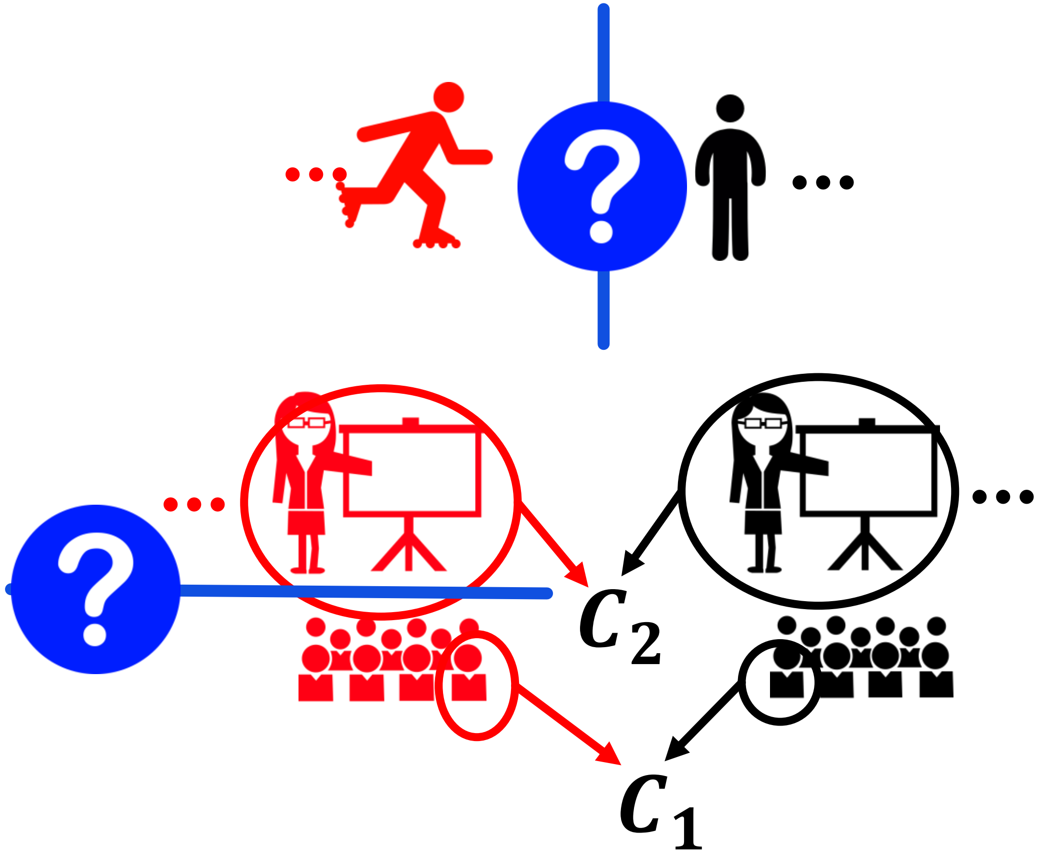
Optimal Design for Single-Level Experiments: Optimize the Sampling Ratio Between Conditions



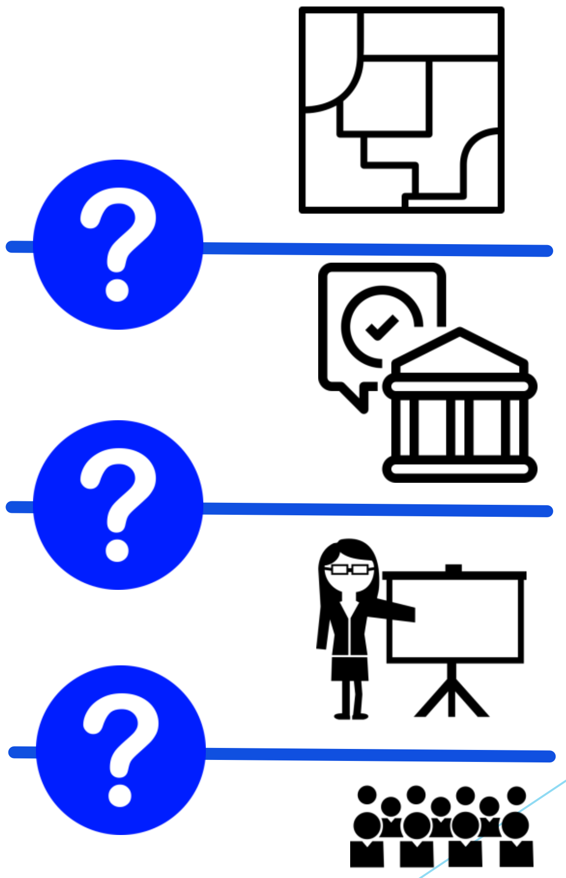
Optimal design parameter: $p = \frac{\sqrt{c/c^T}}{1+\sqrt{c/c^T}}$ **with** $0 < p < 1$

(Cochran, 1963; Nam, 1973)

Prior Optimal Design for Multilevel Experiments: Optimize Sampling Ratios Across Levels



(Raudenbush, 1997)

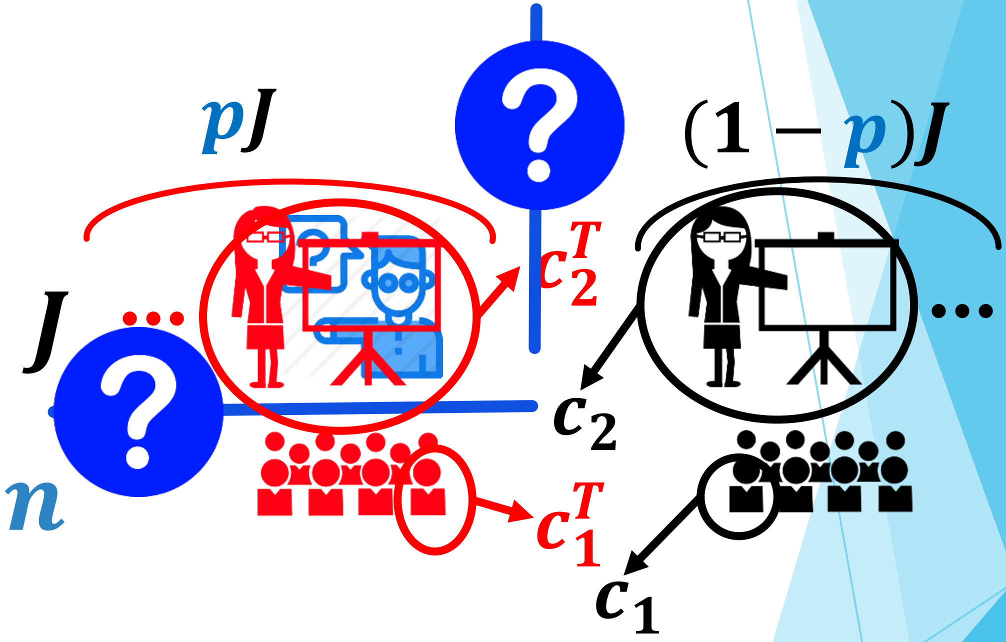


(Raudenbush & Liu, 2000)
(Konstantopoulos, 2009, 2011)
(Hedges & Borenstein, 2014)

A Flexible Framework for 2-Level Cluster-Randomized Trials (CRTs)

$$p = \frac{\sqrt{\frac{c_1 n + c_2}{c_1^T n + c_2^T}}}{1 + \sqrt{\frac{c_1 n + c_2}{c_1^T n + c_2^T}}}$$

$$n = \sqrt{\frac{(1-p)c_2 + pc_2^T}{(1-p)c_1 + pc_1^T}}\sqrt{\frac{(1-\rho)(1-R_1^2)}{\rho(1-R_2^2)}}$$



R package odr (Shen & Kelcey, 2023)

(Shen & Kelcey, 2020)

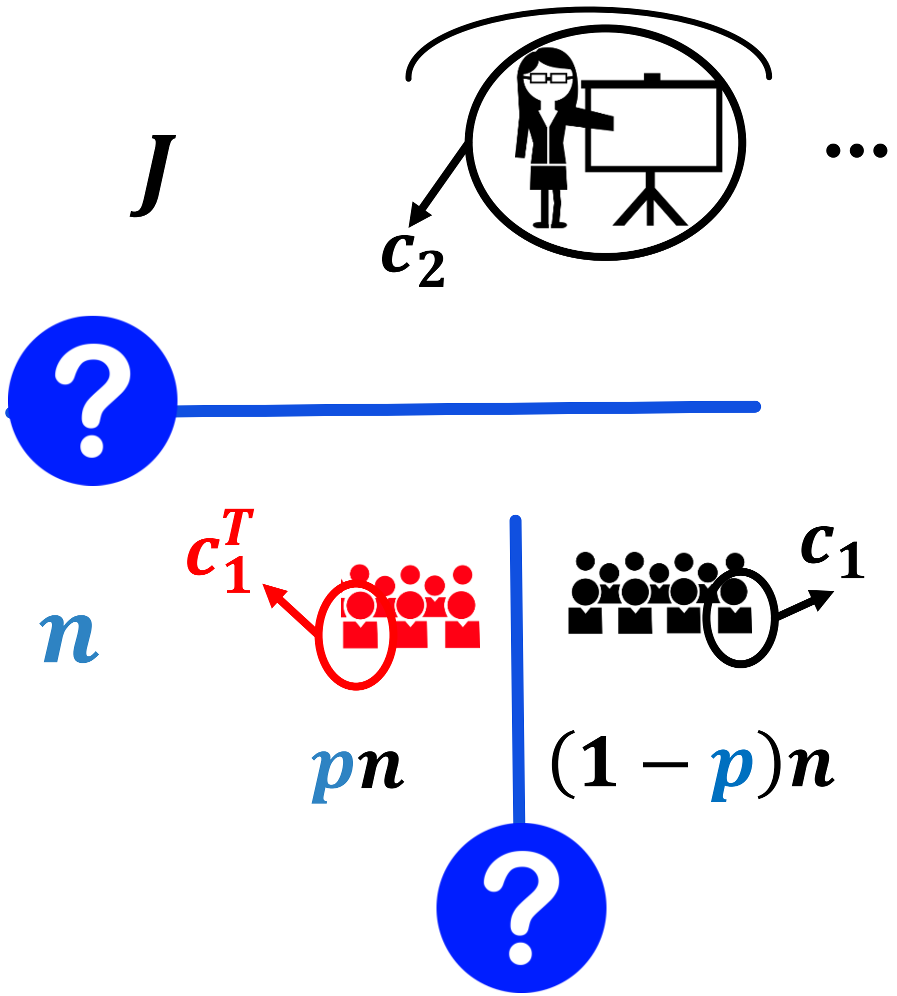
Optimal Design of Studies Investigating Main, Mediation, and Moderation Effects

	Main Effect		Mediation & Moderation	
	CRTs	MRTs	CRTs	MRTs
3-Level	(Shen & Kelcey, 2020)	(Shen & Kelcey, 2022a)		
2-Level	(Shen & Kelcey, 2020)	(Shen & Kelcey, 2022b)		

R package *odr* (Shen & Kelcey, 2023);

Shiny App (Shen & Kelcey, in progress)

Cost Structures of Sampling In Multisite Trials



Optimal Allocation in Multisite Trials Investigating Main Effects

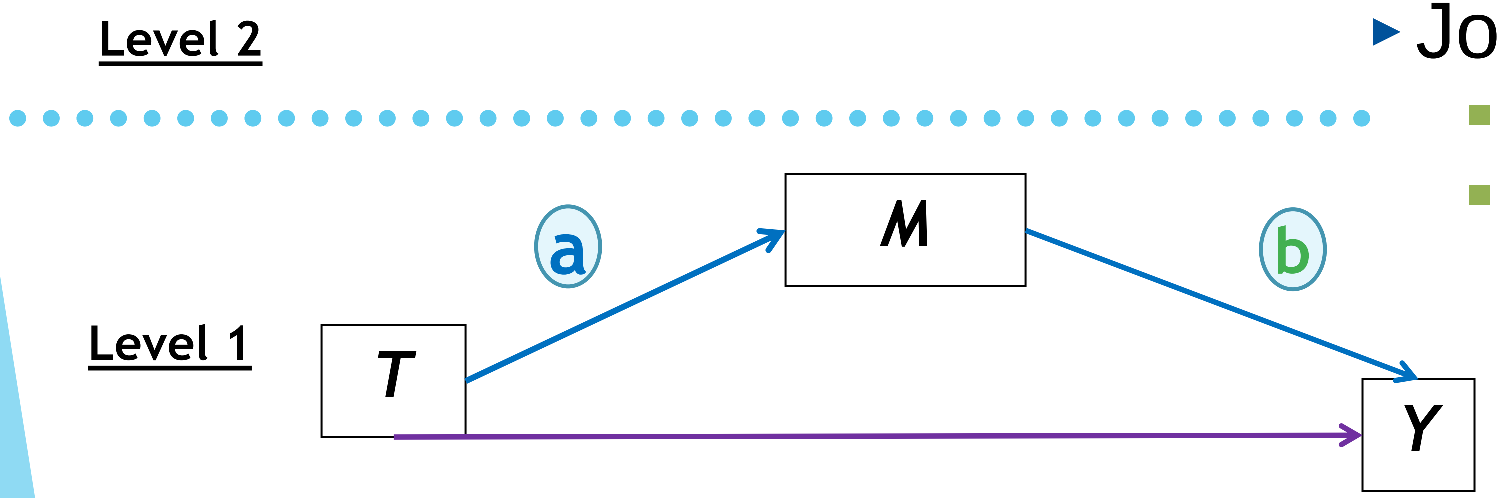
- p & n
  - Conditional variance of the outcome at the individual level
  - Conditional treatment-by-site variance
  - Cost information

$$\omega(1-R_{2m}^2)(c_1^T-c_1)p^2(1-p)^2n^2+(1-\rho)(1-R_1^2)(c_1^T-c_1)p^2n$$
$$-(1-\rho)(1-R_1^2)(nc_1+c_2)(1-2p)$$
$$=0,$$

$$n=\sqrt{\frac{(1-\rho)(1-R_1^2)c_2}{p(1-p)\omega(1-R_{2m}^2)[pc_1^T+(1-p)c_1]}}.$$

(Shen & Kelcey, 2022b)

1-1-1 Mediation



- Sobel Test
  - $\sigma_{ab}^2 = a^2\sigma_b^2 + b^2\sigma_a^2$
- Joint significance test
  - Power(ab) = power(a)×power(b).
  - Statistical power for each path is calculated in a non-central *t*-distribution.

Optimal Design Parameters Under the Sobel Test

- a, b
- Conditional variance of the mediator at the individual level
- Conditional variance of the outcome at the individual level
- Conditional treatment-by-site variance
- Cost information

- $n = \sqrt{\frac{a^2(1-\rho)(1-R_1^2)c_2 p(1-p)}{b^2\{(1-\rho_M)(1-R_1^{2M})[c_1(1-p)+c_1^T p]+\omega(1-R_2^{2M})c_2\}(1-\rho_M)(1-R_1^{2M})}}$
- $a^2(1-\rho)(1-R_1^2)[-c_1 n + c_1^T n](p-p^2)n(1-\rho_M)(1-R_1^{2M})p(1-p) + b^2[\omega(1-R_2^{2M})n + (1-\rho_M)(1-R_1^{2M})][-c_1 n + c_1^T n]n(1-\rho_M)(1-R_1^{2M})n(1-\rho_M)(1-R_1^{2M})[p(1-p) - n(1-\rho_M)(1-R_1^{2M})(1-2p)] = 0$

Numerical solutions (Shen & Kelcey, 2020, 2022a)

Optimal Design of 1-1-1 Mediation Under the Joint Significance Test

- Ant colony optimization (ACO; Socha & Dorigo, 2008) algorithm
  - The ACO algorithm was inspired by the behavior of ant food foraging
  - The ACO algorithm creates artificial arts traveling through possible solution spaces to find an optimal solution that is linked to an objective function
- For the optimal design, we set the total cost as the objective function to be minimized

Steps of ACO

Initiation Stage

1. Initiate k (e.g., 50) sets of optimal design parameters of p and n (e.g., random sample 5 values for p and 10 values for n)

Iteration Stage

2. For each set of optimal design parameters, calculate the required number of sites to achieve a target power (80%), and calculate the required budget
3. Form/update a probability density function across optimal design parameters (p & n)
4. Sample additional sets of optima design parameters (p & n) according to the probability density function

Illustration

- $\rho = \rho_M = .20$
- $\omega = 0.01$
- $c_1 = \$10,\ c_1^T = \$480,\ c_2 = \$100$ (Gray et al., 2022)

Mediation Effect

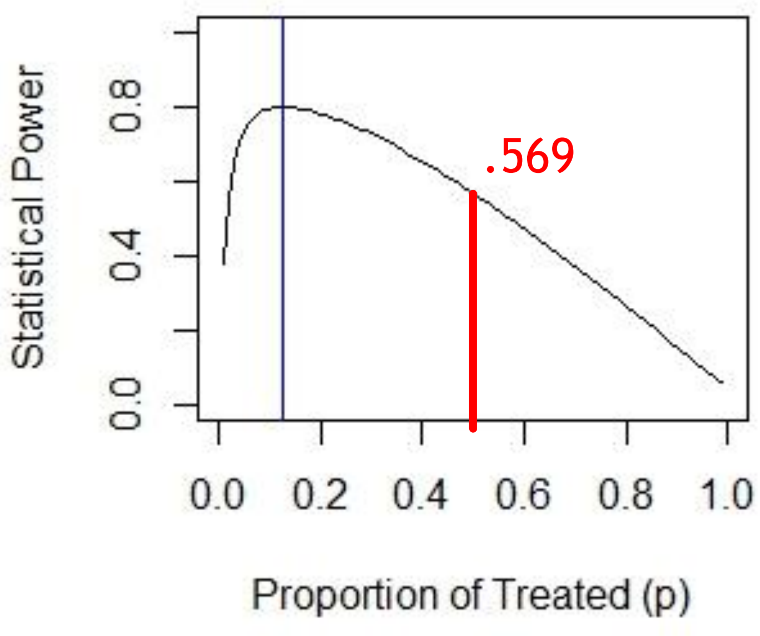
- $a = 0.2,\ b = .30$
- Joint Significance Test: $n = 32.25$ and $p = .126$ $J = 48$ ➔ ~$117,000
- Such a design can detect a main effect of .3 with 98.75% power
- If we use a conventional balanced design (p = .50 & n = 20), we will need 54% more budget to achieve 80% power for the mediation effect

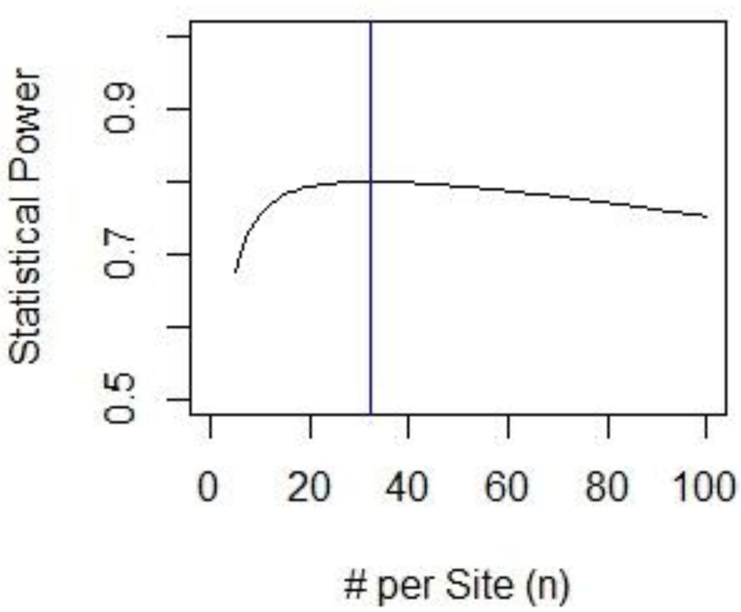
Main Effect

- d = .30
- Optimal allocation n = 20.3 and $p = .344$, $J = 18$ ➔ ~$66,000 (80% power)
- If we use a conventional balanced design (p = .50 & n = 20), we will need 30% more budget to achieve 80% power for the main effect
- Such an optimal design for main effect can detect the mediation effect ($a = 0.2,\ b = .30$) with 47.17% power

Optimization Under the Joint Significance Test

$117,000





Conclusion

- Optimal sample allocations are starting points
- Additional design parameters are needed in the design stage: Cost information
- A framework to simultaneously consider more than one effect in the same study design is needed

Contact Information:
Zuchao “William” Shen
zuchao.shen@gmail.com