Optimal Design of Experimental Studies Under Condition- and Unit-Specific Cost Structures

Zuchao "William" Shen
University of Kansas

Benjamin Kelcey
University of Cincinnati

This presentation is based on three papers published in JEBS, JREE, & JXE.

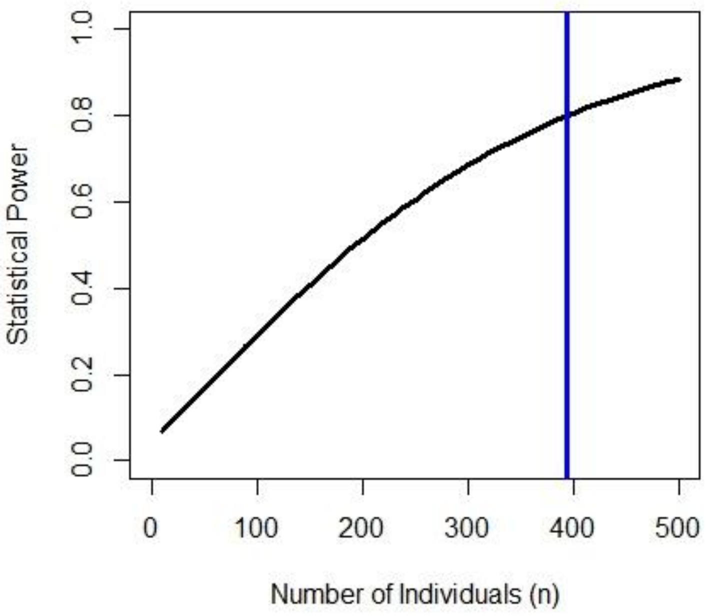
The research was supported by National Academy of Education and the Spencer Foundation.

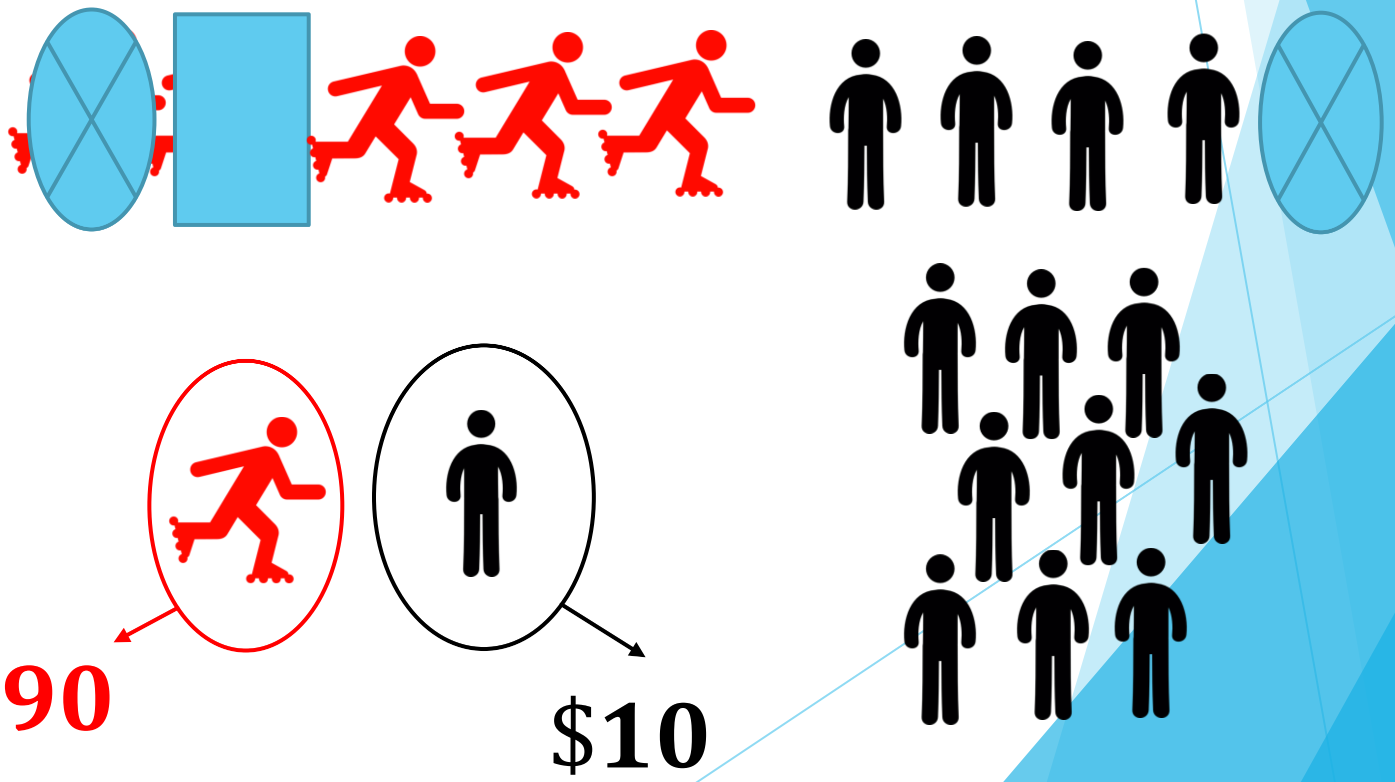
Email: zuchao.shen@gmail.com

An Example About Sample Allocation, Budget, & Statistical Power

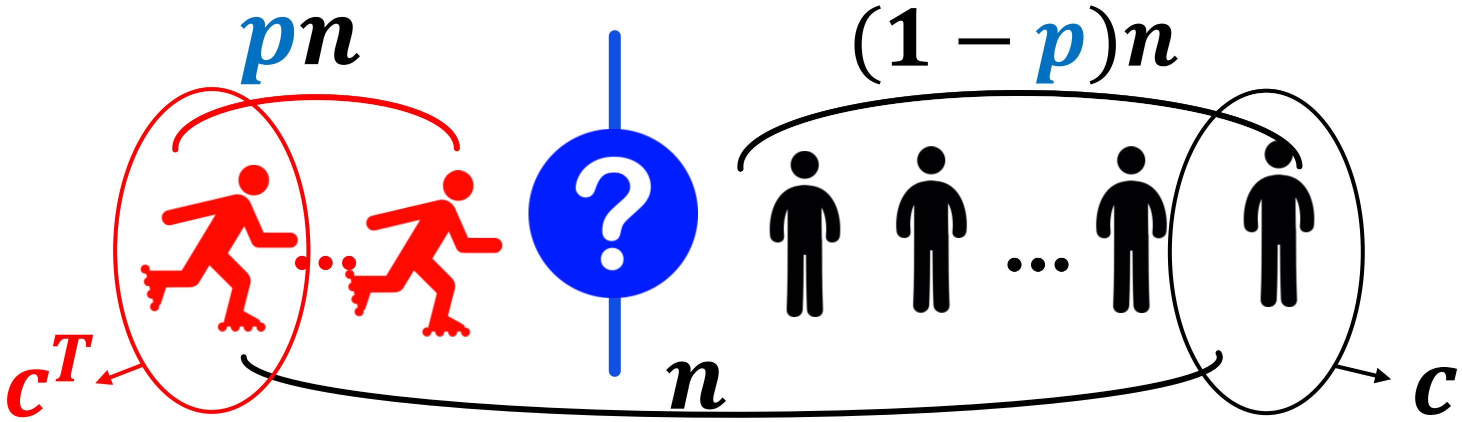
- How many individuals are needed to have .80% power?
  - Effect size: $d = 0.2$
  - Proportion of variance explained: $R^2 = 0.50$

$19,700 **$16,000**





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)

An Optimal Design Example

- $c = \$10,\ c^T = \$90$
- $d = 0.2,\ R^2 = 0.5$

$$\Rightarrow p = \frac{\sqrt{c/c^T}}{1+\sqrt{c/c^T}} = .25$$

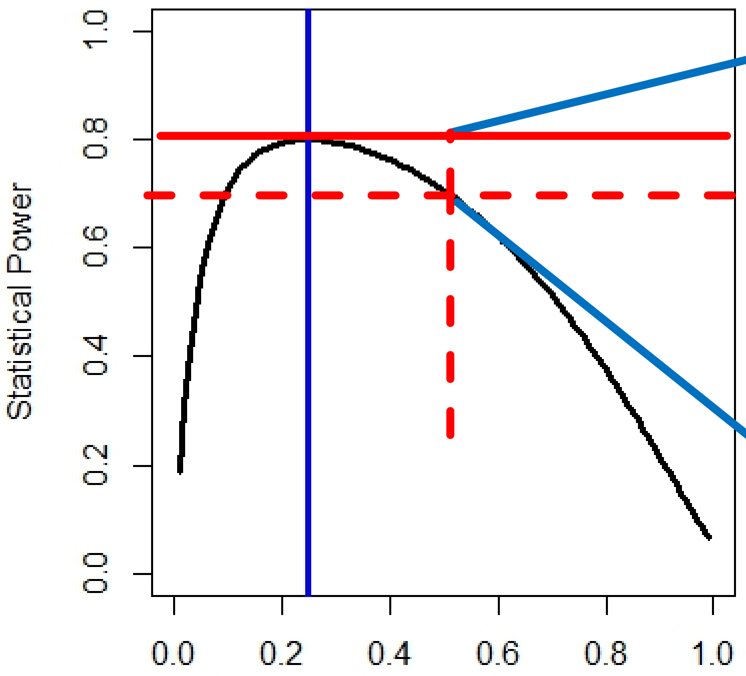
$15,780

(1)

$n^o = 526$

$(132, 395)$;

$15,780



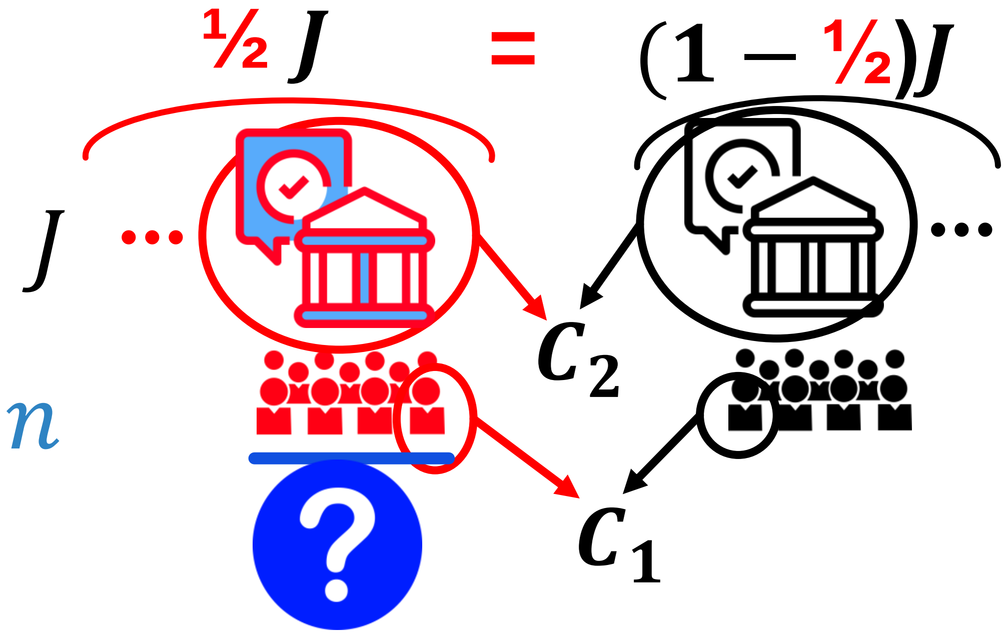
$n^C = 394\ (197, 197)$;

$197,00

25%

(2) $n^C = 316\ (158, 158)$; $\sim\$15,780$

Conventional Framework for 2-Level Cluster-Randomized Trials (CRTs)



Assumption & Constraint

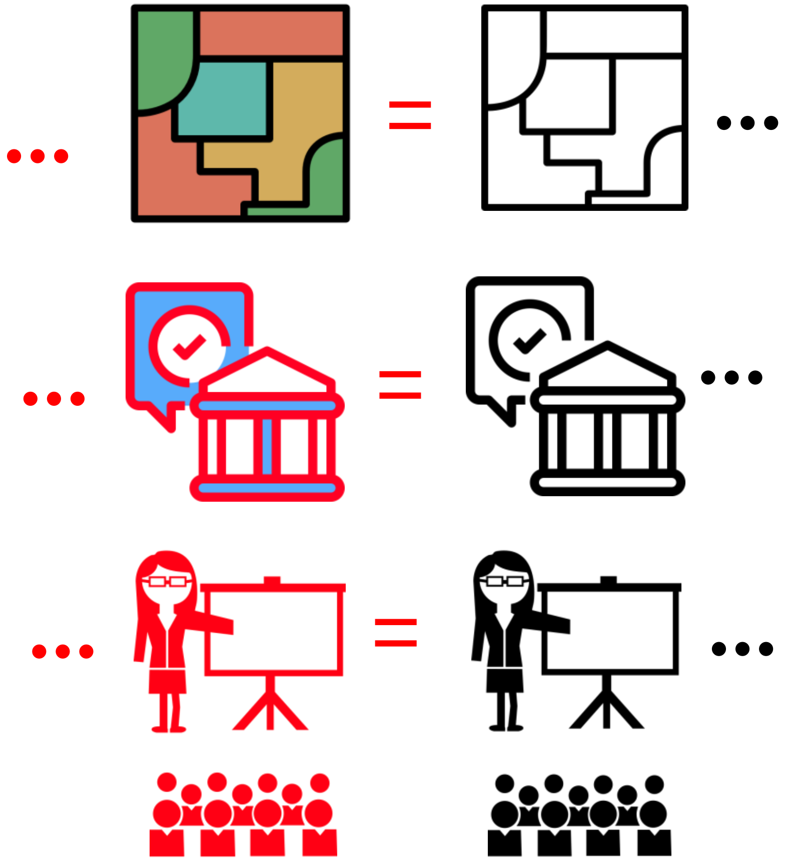
Optimal design parameter: $n = \sqrt{\frac{C_2}{C_1}}\sqrt{\frac{\sigma_{1A}^2}{\sigma_{2A}^2}} \Rightarrow n = \sqrt{\frac{C_2}{C_1}}\sqrt{\frac{(1-\rho)(1-R_1^2)}{\rho(1-R_2^2)}}$

(Raudenbush, 1997)

Conventional Framework: Extensions

Two-Level CRTs: Mediation (Kelcey & Shen, 2020; Cox & Kelcey, 2019)

Regression Discontinuity Designs (Rhoads & Dye, 2016)



Four-Level CRTs
(Hedges & Borenstein, 2014)

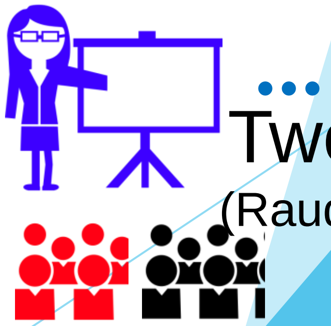
Four-Level MRTs
(Hedges & Borenstein, 2014)

Three-Level CRTs
(Konstantopoulos, 2009, 2011)

Three-Level MRTs
(Hedges & Borenstein, 2014)

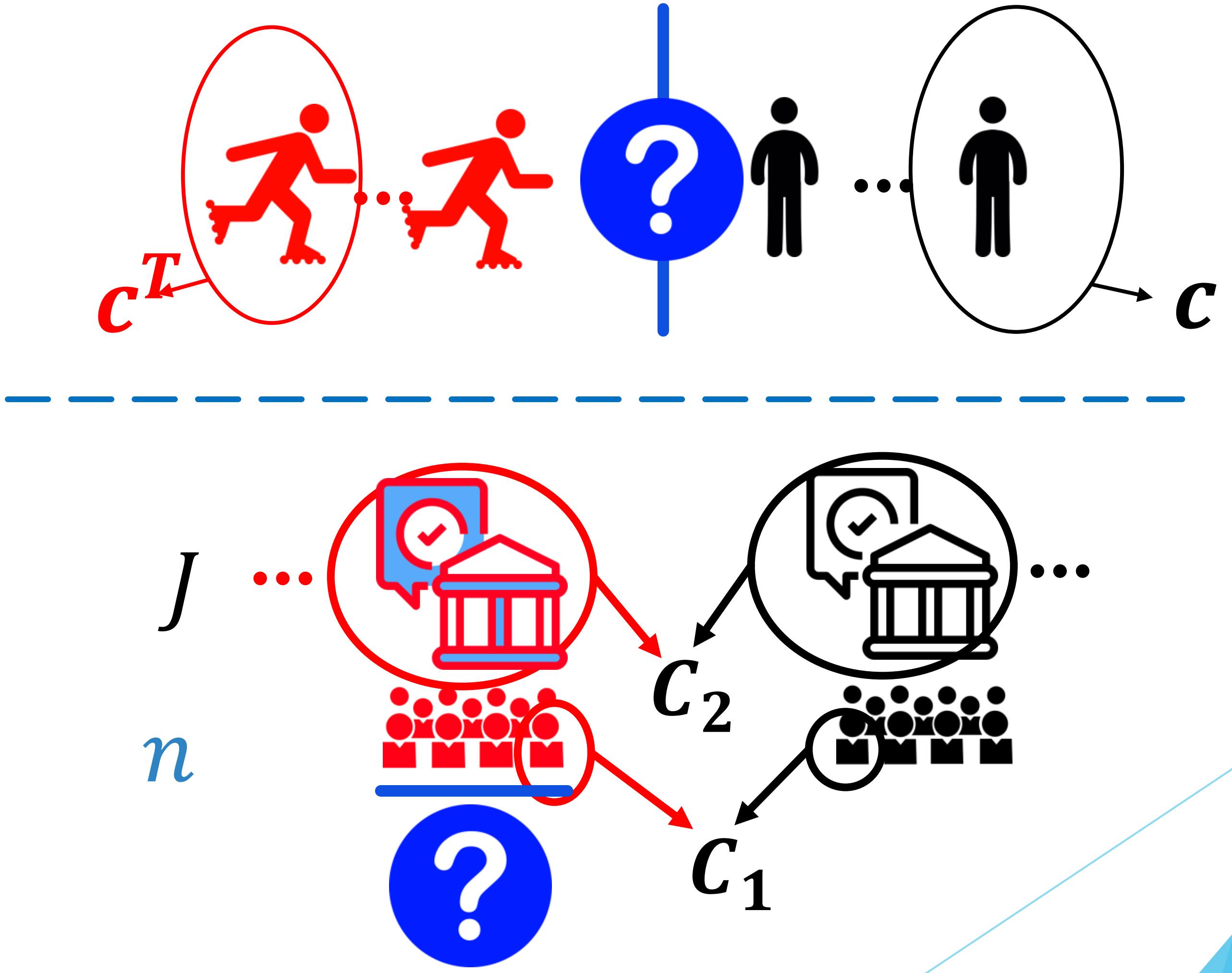
Assumption & Constraint

Two-Level CRTs
(Raudenbush, 1997)



Two-Level MRTs
(Raudenbush & Liu, 2000)

Inconsistency in Frameworks



Past & Current Research on Optimal Design

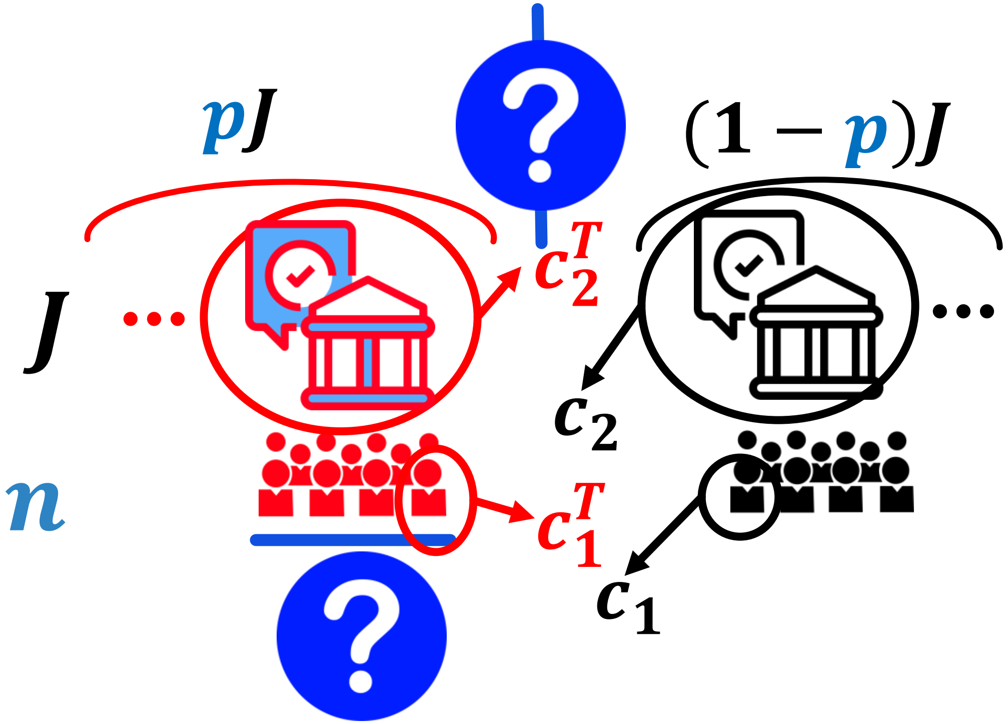
	Main Effect		Mediation/Moderation Effect	
	CRTs	MRTs	CRTs	MRTs
4-Level	(Shen, 2019)	(Shen, 2019)		
3-Level	(Shen & Kelcey, 2020)	(Shen & Kelcey, 2022b)		
2-Level	(Shen & Kelcey, 2020)	(Shen & Kelcey, 2022a)	(Shen et al., in prep)	(Shen et al., in prep)

R package *odr* (Shen & Kelcey, 2023**); R Shiny App (**Shen & Kelcey, in progress**)**

NATIONAL ACADEMY of EDUCATION

spencer

A Flexible Framework for 2-Level CRTs

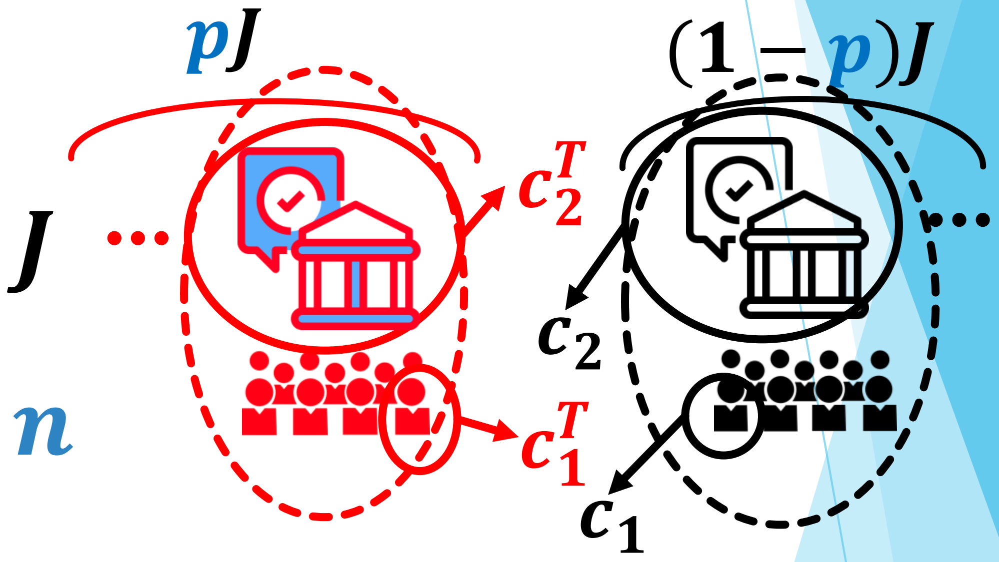


(Shen & Kelcey, 2020)

Optimal Design Parameters

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

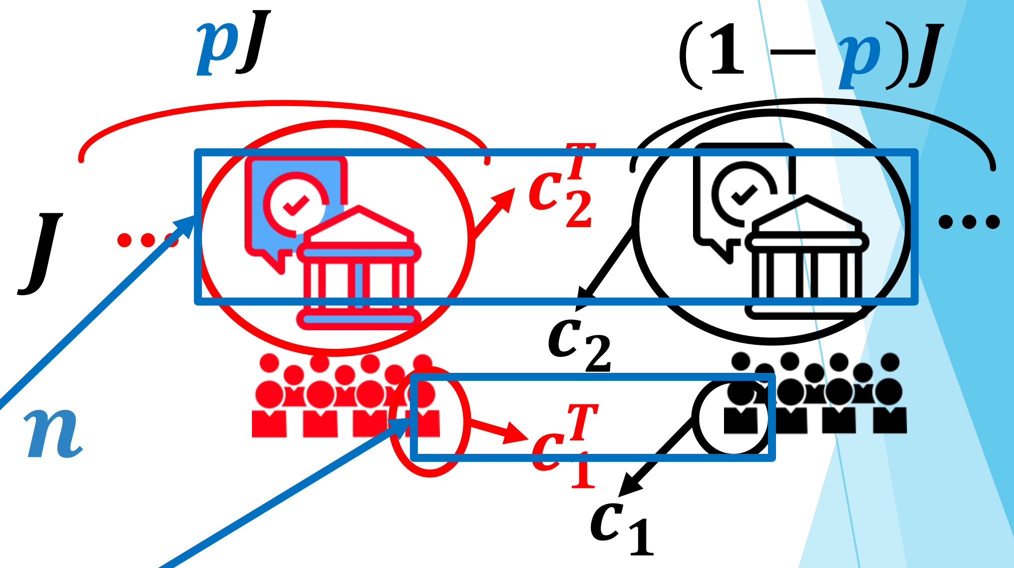
$$n = \sqrt{\frac{(1-p)c_2 + pc_2^T}{(1-p)c_1 + pc_1^T}} \sqrt{\frac{(1-\rho)(1-R_1^2)}{\rho(1-R_2^2)}}$$



(Shen & Kelcey, 2020, 2023)

Optimal Design Parameters

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



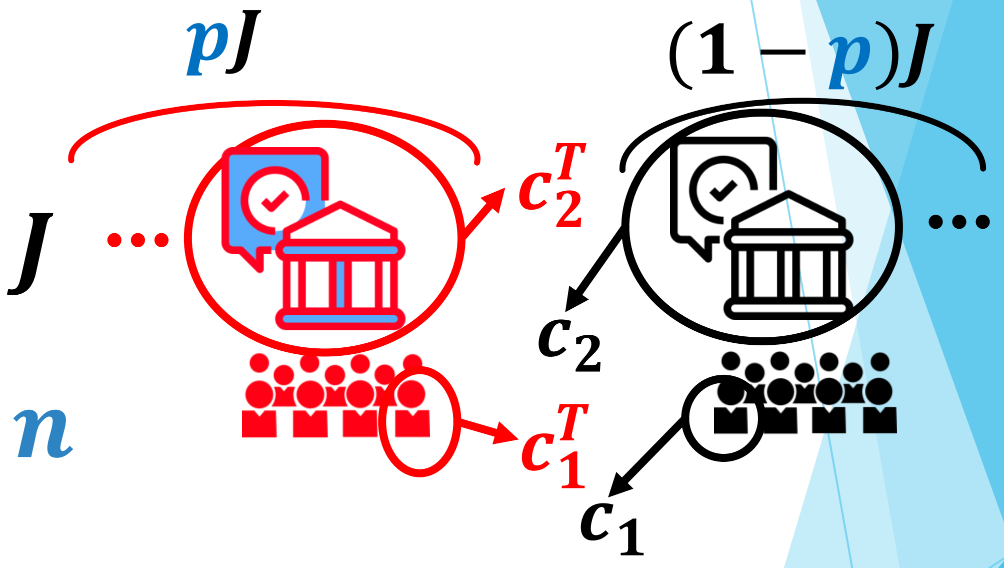
$$n = \sqrt{\frac{(1-p)c_2 + p c_2^T}{(1-p)c_1 + p c_1^T}} \sqrt{\frac{(1-\rho)(1-R_1^2)}{\rho(1-R_2^2)}}$$

(Shen & Kelcey, 2020, 2023)

No Closed-Form Solution

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

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



(Shen & Kelcey, 2020, 2023)

Iteration Method for Numerical Solutions

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

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

Sample $n = n_0$

Solve for $p = p_0$

Update $n = n_1$

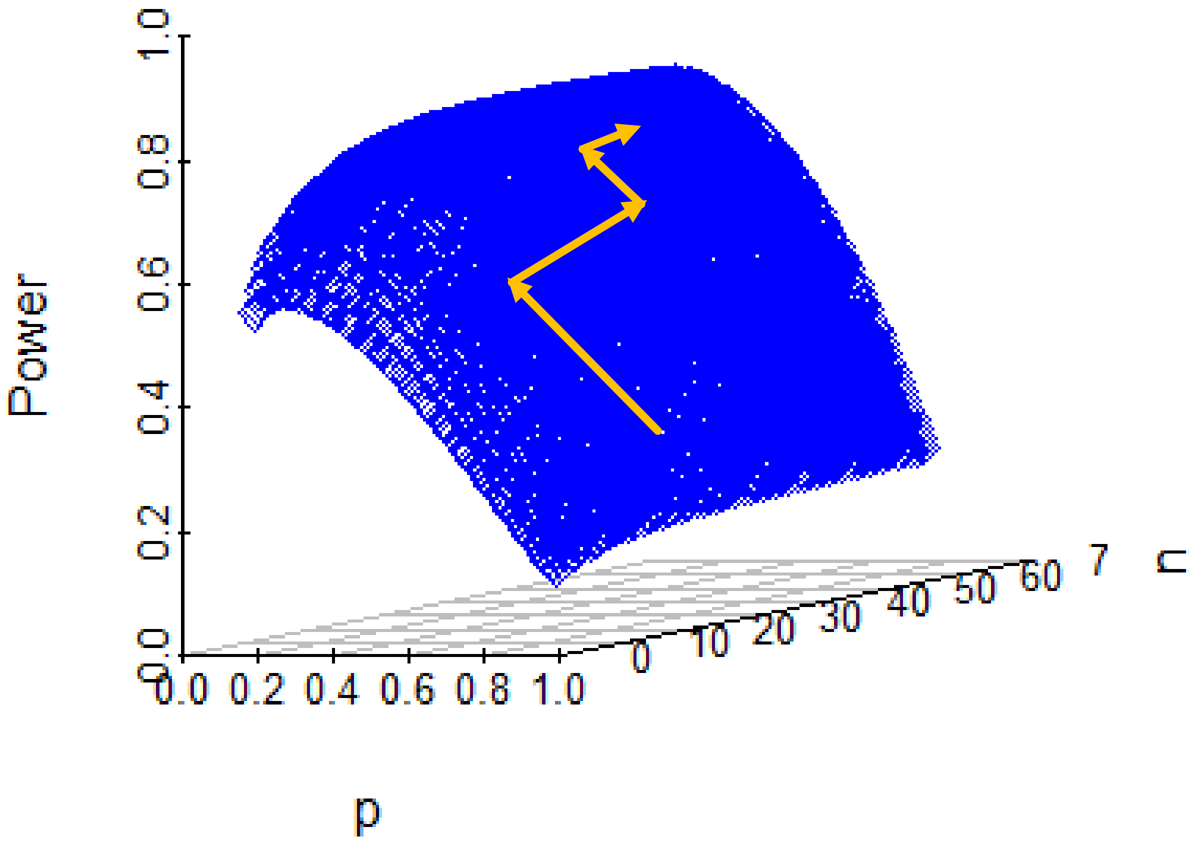
$p = p_1$

......

Until

$$n_m - n_{m-1} \leq \frac{1}{10^{10}}$$

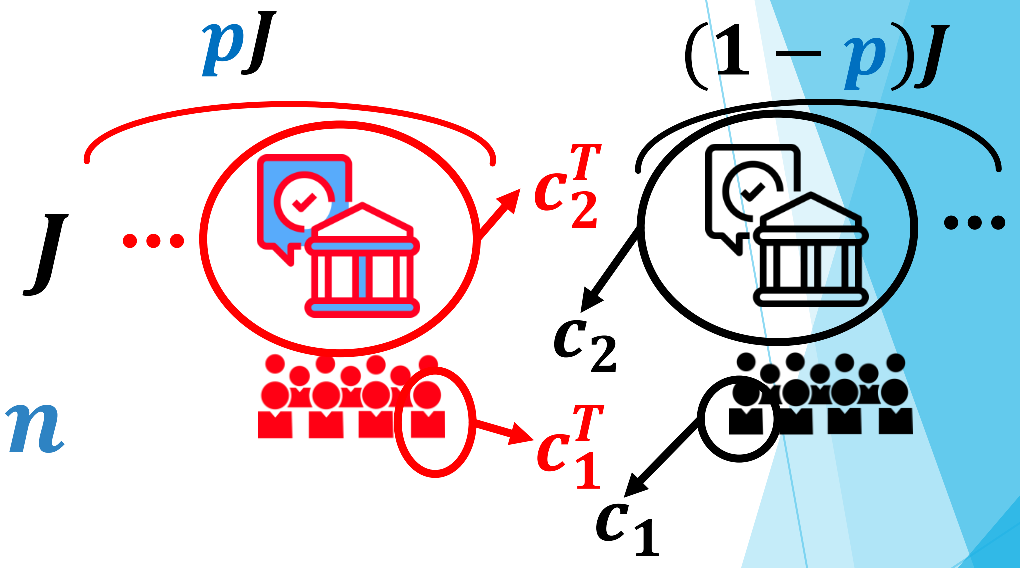
$$p_m - p_{m-1} \leq \frac{1}{10^{10}}$$



(Shen & Kelcey, 2020, 2023)

Optimal Design Parameters

Unit	Cost
A Student in Control c_1	$10
A Student in Treatment c_1^T	$10
A School in Control c_2	$500
A School in Treatment c_2^T	$10,000

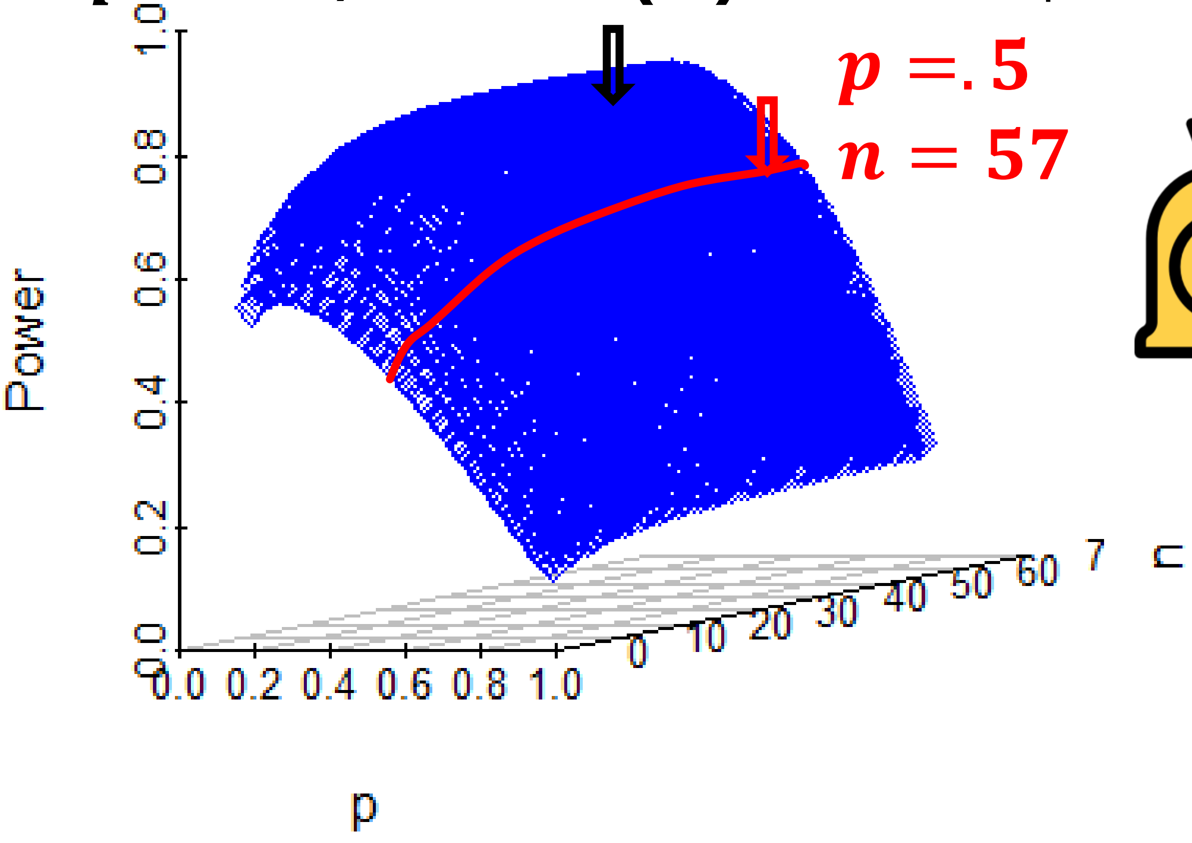


$$\rho = .2,\ R_1^2 = .45,\ R_2^2 = .65$$

$$p = .23$$
$$n = 41$$
$$J = 91$$
$$m = \$283K$$

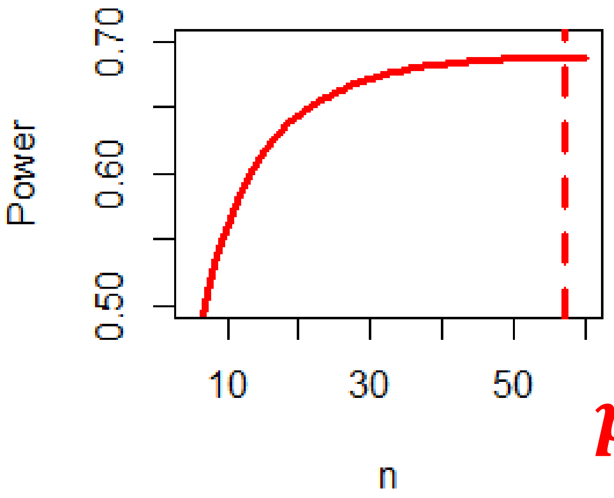
Graphic Illustration

$p = .23, n = 41$ (.8) $m^o = \$283K$



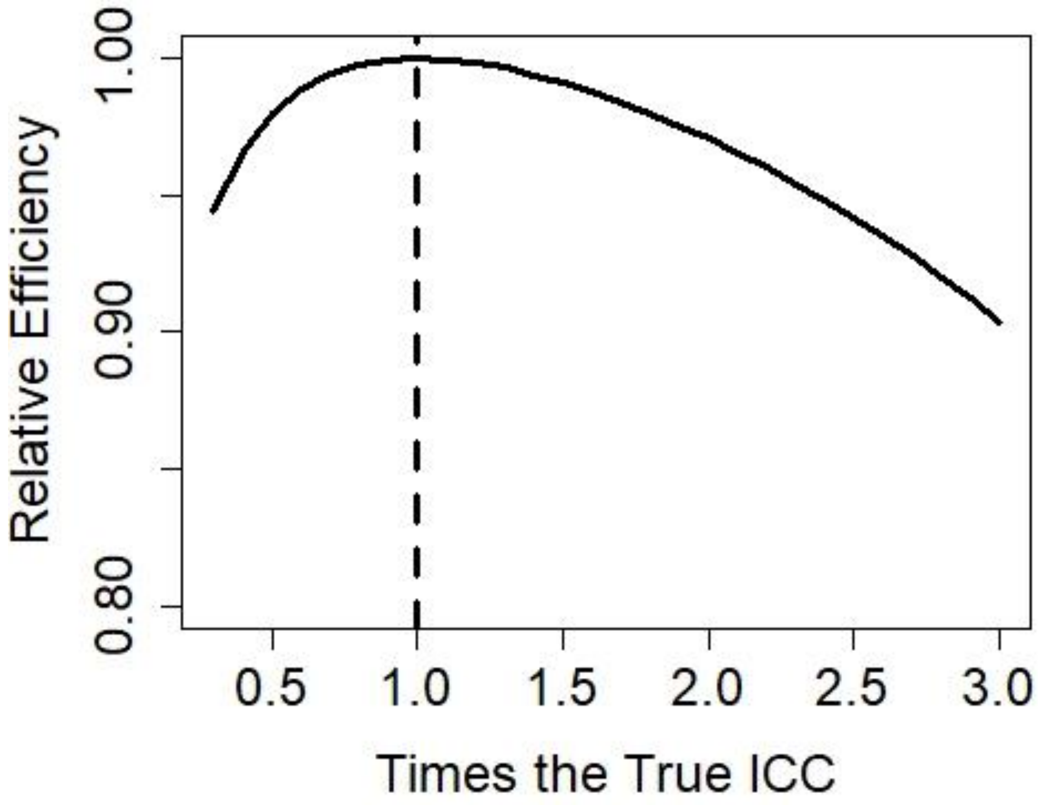
$p = .5$
$n = 57$

$m = \$367K$ (30%) ⬆



$p = .5$

Robustness to Misspecification



True value = 1 time (i.e., $\rho = .20$)

- Relative Efficiency

$$= \frac{m^o}{m}$$

$$= \frac{\sigma_d^{2^o}}{\sigma_d^2}$$

$$= f(cost, variance, OD\ parameters)$$

(e.g., Shen & Kelcey, 2020, 2022a, 2022b)

Contributions

Efficient and effective designs

Accelerating innovation

Theory

Framework

Tools

Email: zuchao.shen@gmail.com