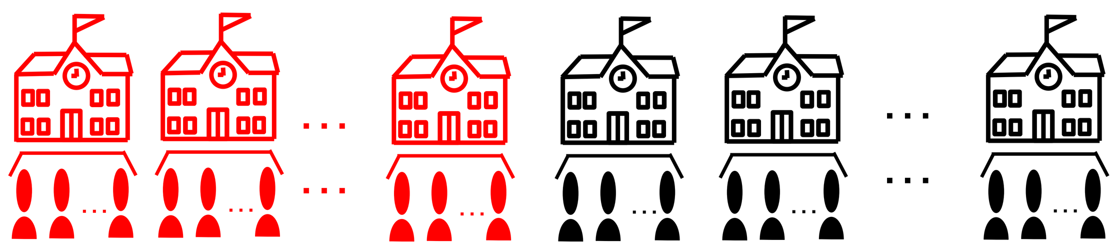
Optimal Allocation of Sample Sizes in Group-Randomized Trials

Zuchao Shen — University of Florida
Ben Kelcey — University of Cincinnati
& Gang Wang — Winthrop University

Introduction

Multilevel structures are extremely common in education and social sciences, such as students nested within schools and patients nested within physicians. Group-randomized trials are often used in the social sciences to account for the multilevel structures.



A key consideration in planning group-randomized trials is statistical power or the probability of detecting treatment effects if they exist. A related consideration is the optimal use of resources to achieve adequate power.

Optimal Sample Allocation

Sample allocation is the sample sizes (to be) allocated at different levels and treatment conditions in (group-randomized) experiments.

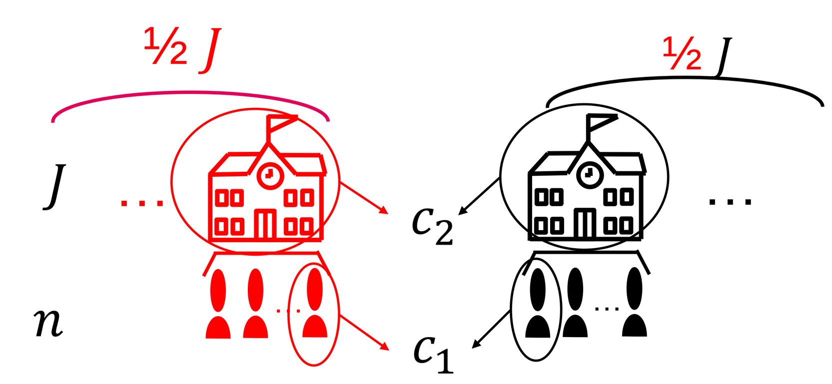
Optimal sample allocation is the sample allocation that achieves the maximum statistical power under a fixed budget, or uses the minimum resources to gain adequate statistical power.

Why optimal sample allocation? It can be used to improve design efficiency (e.g., less budget request) and/or statistical precision (more power, smaller effect sizes).

Methods

Framework & Cost Structure	Optimal Design Parameter(s)	Assumptions/Constraints & Visual Representation

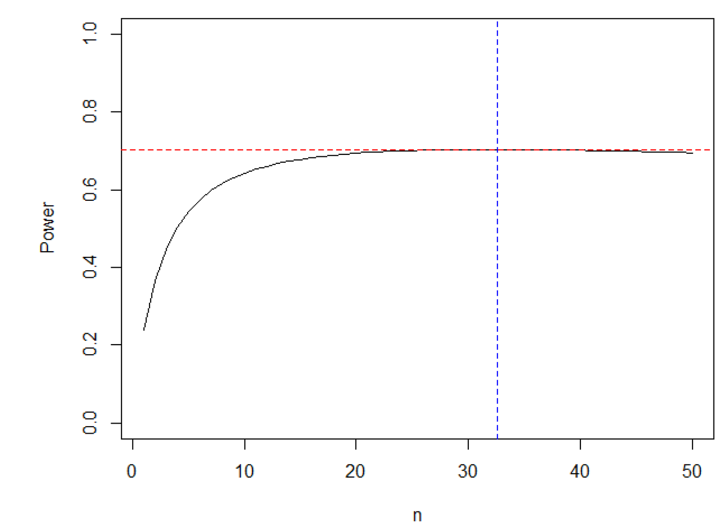
Raudenbush (1997)



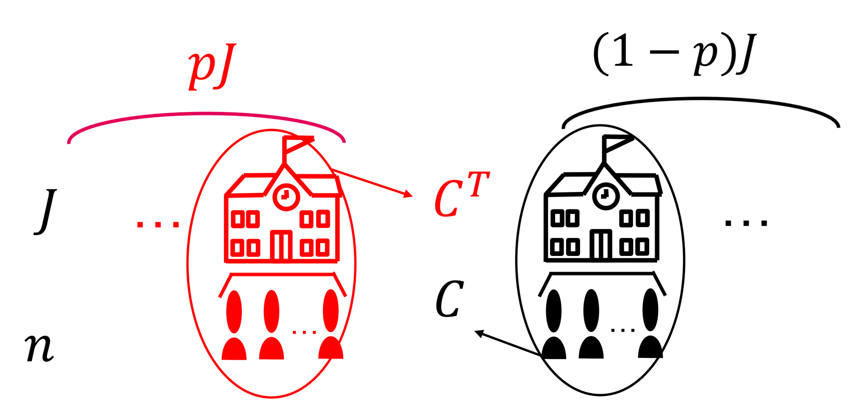
$$n = \sqrt{\frac{C_2}{C_1}}\sqrt{\frac{(1-\rho)(1-R_1^2)}{\rho(1-R_2^2)}}$$

- ❑ $R_\#^2$ represents the proportion of variance explained by covariate(s)
- ❑ ρ is the intraclass correlation coefficient or proportion of outcome variance at the group level.

- ❑ Between-treatment equal cost assumption
- ❑ Balanced design constraint (i.e., $p = 0.5$).

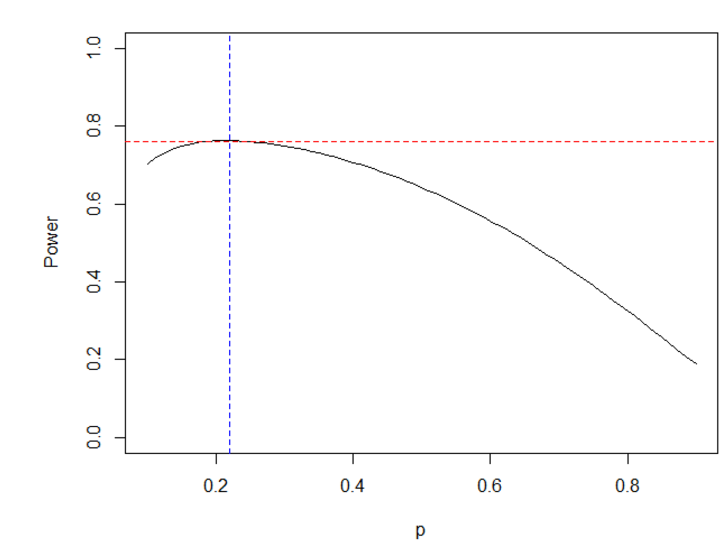


Liu (2003)

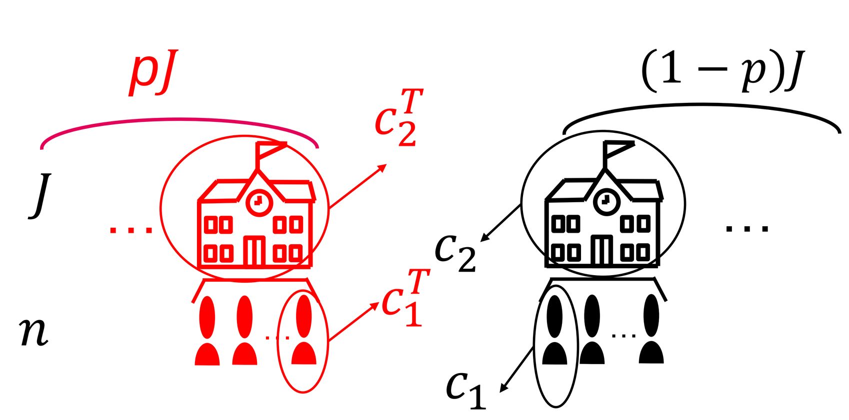


$$p = \frac{\sqrt{\frac{C}{C^T}}}{1+\sqrt{\frac{C}{C^T}}}$$

- ❑ Predetermined sample size constraint (i.e., $n = n_0$).



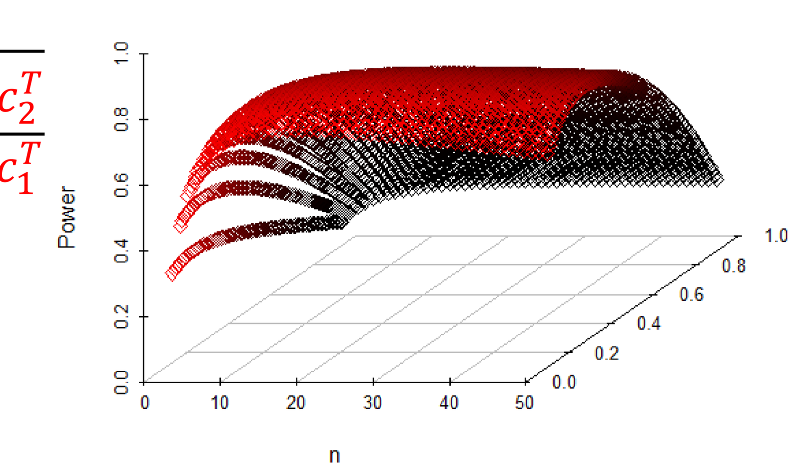
Proposed



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

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

- ❑ No constrained cost structure or sample size(s), but the framework can accommodate constraint(s)



An Example

A school-randomized trial

- ❑ intraclass correlation coefficient (ρ): 0.2
- ❑ How many schools and students do we need to have a power of 0.80 to detect an effect size of 0.20?
- ❑ What is the optimal sample allocation given the below cost structures of sampling?
  - one student in control: $10
  - one student in treatment: $10
  - one school in control: $300
  - one school in treatment: $5,000

Results

Framework	Sample Allocation	Total Costs
Raudenbush	$p = 0.50, n = 33, J = 178$	$530,440
Liu	$p = 0.22, n = 10, J = 322$	$462,500
Proposed	$p = 0.24, n = 24, J = 253$	$423,320

Conclusion

- ❑ The proposed framework
  - ✓ has more authentic assumptions
  - ✓ can identify more efficient designs
- ❑ It has been implemented in R
  - ✓ R package *odr* (Shen & Kelcey, 2018)