



BOLD
THINKERS
DRIVING
REAL-WORLD
IMPACT

50 YEARS



Session 2: On the Frontier of the “Where” of Social Experiments: Enhancing External Validity in the Relevant Policy Space

*Social Experiments in Practice:
The Why, When, Where, and How of Experimental Design & Analysis*
MEMBER FORUM | WASHINGTON, DC



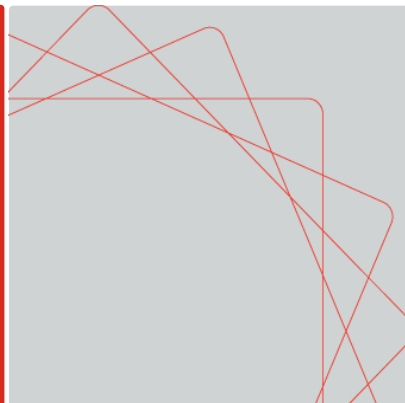
BOLD
THINKERS
DRIVING
REAL-WORLD
IMPACT

50 YEARS

**Design Innovations
for Improved External
Validity: Selecting
Sites for Social
Experiments**

Presented by:
Rob Olsen

APPAM-Abt Institutional Member Forum
Washington, DC



Background



- Almost all social experiments (evaluations based on random assignment) are conducted in a nonrandomly selected sample of sites.
- Researchers use various approaches to obtain samples for these social experiments.
- The resulting samples are often called “convenience samples” or “purposive samples.”

Abt Associates | pg 3

Nonrandom Site Selection – Two Examples



- Example 1: One-at-a-time recruiting of favored sites.
 - Favor sites with certain characteristics.
 - Stratify to meet distributional goals.
- Example 2: Mass invitation asking sites to volunteer.
 - Competition led by research sponsor or evaluator.
 - Could be competitive process or first-come first-serve.

Abt Associates | pg 4

Problem



- When sites are selected nonrandomly, ***it is unclear if we can learn anything with confidence that extends beyond the sites in the study sample.***

Why don't evaluations select sites randomly?

- Some do (e.g., the Head Start Impact Study and an ongoing evaluation of the Workforce Investment Act).
- But most don't because they don't think they can successfully (or cheaply) recruit a random sample.

Abt Associates | pg 5

Question



How can we select more representative samples of sites in social experiments?

Abt Associates | pg 6

Overview of Recommendations



1. Define and identify the population of interest.
2. Define site strata based within the population.
3. Select sites randomly within strata.
4. Select larger samples of sites.
5. Conduct a quasi-experiment in a broader sample of sites as a supplement to the experiment.

Abt Associates | pg 7

Overview of Recommendations



1. Define the population of interest.

2. Define site strata based on observed factors.
3. Select sites randomly within strata.
4. Select larger samples of sites.
5. Conduct a quasi-experiment in a broader sample of sites as a supplement to the experiment.

Abt Associates | pg 8

Define the Population of Interest



- Identify the:
 1. Important policy decisions that the study could inform.
 2. Individuals whom these decisions could affect.

- The population(s) of interest should be identified by the research sponsor—but they may need help from their evaluator.

Abt Associates | pg 9

Define the Population of Interest



Define the *primary* population of interest
as:

**the intended beneficiaries of
the most consequential policy decision
that the study could inform.**

Abt Associates | pg 10

Examples of the Population



- If the policy decision is whether to keep or eliminate an existing program:
 - The population of interest could be current participants in places where the program currently operates.
 - Examples: Most evaluations of federal programs.

- If the policy decision is whether to expand the program:
 - The population of interest could be potential beneficiaries if program were scaled up.
 - Examples: Most evaluations of demonstration programs.

Abt Associates | pg 11

Overview of Recommendations



1. Define the population of interest.

- 2. Define site strata within the population.**

3. Select sites randomly within strata.

4. Select larger samples of sites.

5. Conduct a quasi-experiment in a broader sample of sites as a supplement to the experiment.

Abt Associates | pg 12

Define Site Strata Within the Population



- Identify site-level factors likely to be related to the impacts of the intervention.
- This will help us select a representative sample—or reweight an unrepresentative sample.

How do we choose the factors on which to stratify?

1. Based on “theory” or evidence on impact variation.
2. Based on the data (e.g., cluster analysis).

Abt Associates | pg 13

Overview of Recommendations



1. Define the population of interest.
2. Define site strata within the population.
- 3. Select sites randomly within strata.**
4. Select larger samples of sites.
5. Conduct a quasi-experiment in a broader sample of sites as a supplement to the experiment.

Abt Associates | pg 14

Select Sites Randomly within Strata



- Recent research has focused on systematic site selection within strata (Tipton et al., 2014).
- But it's hard to beat random selection because it ensures that sample selected matches the population on both observed and unobserved characteristics.
- Within each stratum, select a random sample of:
 1. Sites to recruit.
 2. Replacement sites to recruit if needed.
 3. Sites that initially refused for additional recruiting.

Abt Associates | pg 15

Select Sites Randomly within Strata



- Random site selection would allow us to:
 - Weight the sample by the selection probabilities—as we do in surveys.
 - Adjust the weights to account for non-participation of selected sites—as we account for non-response in surveys.
- Dr. Elizabeth Stuart will discuss these approaches to weighting and reweighting the sample later in this conference.

Abt Associates | pg 16

Overview of Recommendations



1. Define the population of interest.
2. Define site strata within the population.
3. Select sites randomly within strata.

4. Select larger samples of sites.

5. Conduct a quasi-experiment in a broader sample of sites as a supplement to the experiment.

Abt Associates | pg 17

Select Larger Samples



- Generalization to the population requires a larger sample than simply detecting effects in the sample of sites that participate in the study.
- If we select sites randomly, we have to account for an additional source of sampling error.
 - Impacts may vary across sites in the population.
 - We may select sites with large or small impacts by chance.
- This sampling error increases the variance of the estimate unless we compensate w/larger samples.

Abt Associates | pg 18

Overview of Recommendations



1. Define the population of interest.
2. Define site strata within the population.
3. Select sites randomly within strata.
4. Select larger samples of sites.

5. Conduct a quasi-experiment in a broader sample of sites as a supplement to the experiment.

Abt Associates | pg 19

Conduct a Quasi-Experiment as a Supplement to the Experiment



- Some sites that won't agree to random assignment may be willing to participate in a study based on a quasi-experimental design (QED).
- We could conduct the experiment—but also conduct a QED in a broader set of sites. This would allow us to:
 - Estimate effects for a broader sample of sites.
 - Compare the effects of the program in the experimental sample to the effects of the program in the broader sample.
- Example: Evaluation of KIPP charter schools.

Abt Associates | pg 20

Costs of Recommendations



| Less Expensive | More Expensive |
|------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ul style="list-style-type: none"> Define the population of interest Define site strata Select sites randomly | <ul style="list-style-type: none"> Try harder to recruit a random sample of initial refusers Select larger samples of sites Conduct a quasi-experiment in a broader sample of sites |

Abt Associates | pg 21

Conclusion



- The relevance of social experiments for informing policy decisions has been limited by the way we choose sites.
- The approaches that I have described are some examples of how we can try to do better.
- They won't eliminate the problem, but...
- Let's not make the perfect the enemy of the good.

Abt Associates | pg 22

Acknowledgments



- My collaborators in this research area:
 - Stephen Bell (Abt Associates)
 - Larry Orr (Johns Hopkins University)
 - Elizabeth Stuart (Johns Hopkins University)

- Research funders:
 - Institute of Education Sciences (Grant R305D100041)
 - National Science Foundation (Grant 1335843)
 - National Institute of Mental Health (Grant K25MH083846)

Abt Associates | pg 23

For more information, please contact:

Rob Olsen

Principal Scientist

Social & Economic Policy Division

T: 301.634.1716

E: Rob_Olsen@abtassoc.com



BOLD
THINKERS
DRIVING
REAL-WORLD
IMPACT

50 YEARS

EXTRA SLIDES



Examples of the Population



- Some evaluations select sites that can implement the intervention with fidelity, or in which the “contrast” between the experimental conditions is large.
- The implicit population is participants in locations with conditions “favorable” for positive impacts.
- Could inform the policy decision of whether to sponsor a large “scale-up” or effectiveness study:
 - Logic: If the intervention isn’t effective under favorable conditions, it won’t be effective under typical conditions



Select Larger Samples



- Generalization to large populations requires larger samples.

- Suppose that:
 - We select sites randomly and randomize 100 individuals per site.
 - We want to determine if the population average effect is positive.
 - The program is effective for half of sites (impact = 0.20 standard deviations) and ineffective for the other half (zero impact).

- To detect impacts of ≥ 0.10 standard deviations, we need:
 - 26 sites to detect this impact *in participating sites*.*
 - 7 additional sites to detect this impact *in the population*.

*Technical note: Assumes $R^2 = .20$, 80 percent power, two-tailed test at 5-percent level.