J-PAL
ABDUL LATIF JAMEEL POVERTY ACTION LAB

How to Randomize

Course Overview

1. Why Evaluate
2. Theory of Change & Measurement
3. Why & When to Randomize
4. How to Randomize
5. Sample Size & Power
6. Ethical Considerations for Randomized Evaluations
7. Threats & Analysis
8. Randomized Evaluation from Start to Finish
9. Applying & Using Evidence
10. The Generalizability Framework

Learning Objectives

- Develop a deeper understanding of randomization and randomization design using real-world examples
- Determine how to select the appropriate level of randomization
- Review common program specifications and how these inform randomization design choices

Lecture Overview

- What is Randomization?
- Randomization Procedures
- The Unit of Randomization
- Balance & Stratification
- Designing Randomized Evaluations for Different Program Specifications

Lecture Overview

- What is Randomization?
- Randomization Procedures
- The Unit of Randomization
- Balance & Stratification
- Designing Randomized Evaluations for Different Program Specifications

Random Sampling and Random Assignment

ESP WORKBOOK

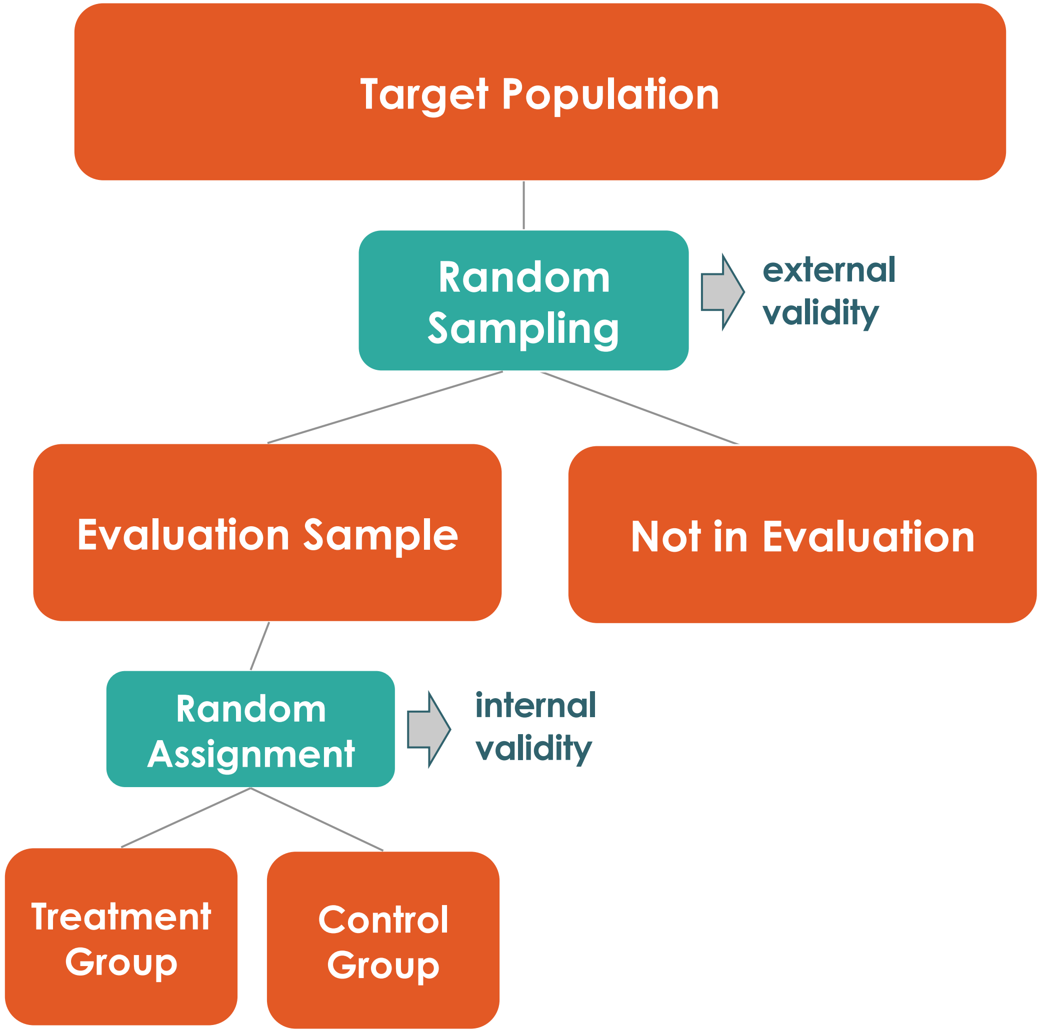
Defining terms

Random sampling: selecting units from a population of interest in a randomized manner to create a sample that is representative of the population.

External validity: the acceptability of results of an evaluation in contexts other than those in which the evaluation was conducted.

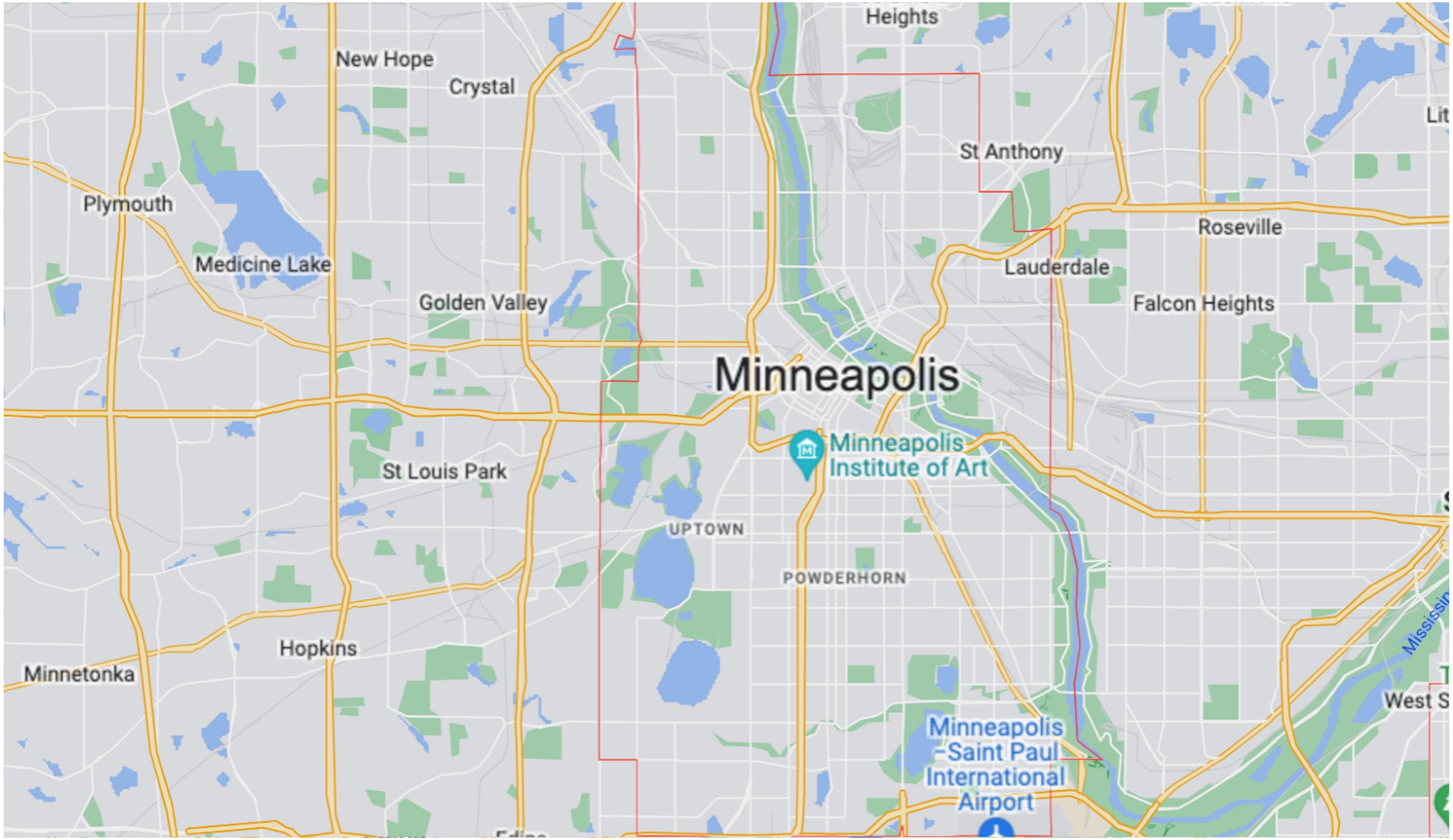
Random assignment: taking a pool of eligible units and then allocating those units to treatment and control groups by means of a random process

Internal validity: The acceptability of the results of an evaluation in terms of causal impact of the intervention.



Random Sampling

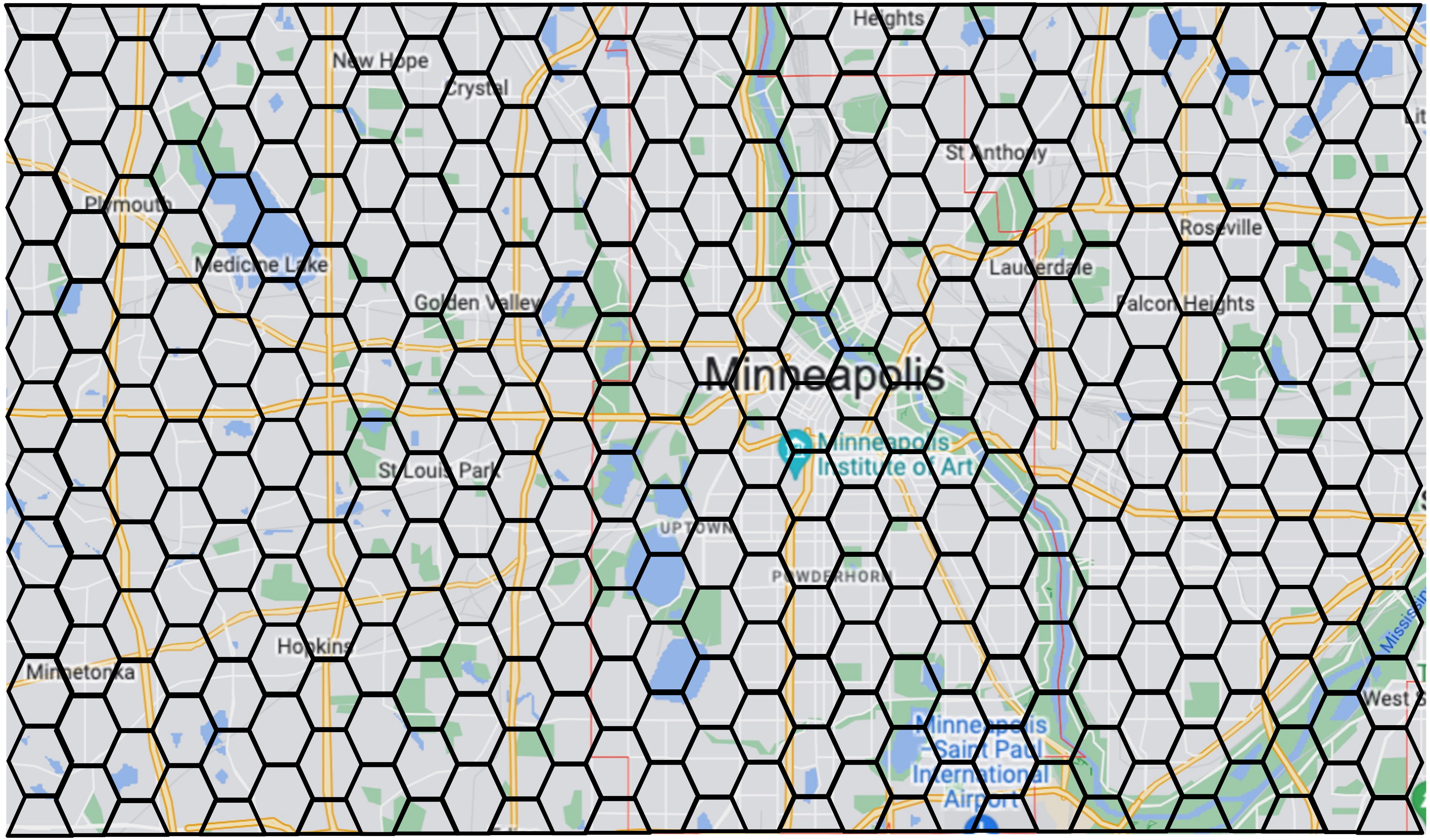
Area of interest



Google Maps

Random Sampling

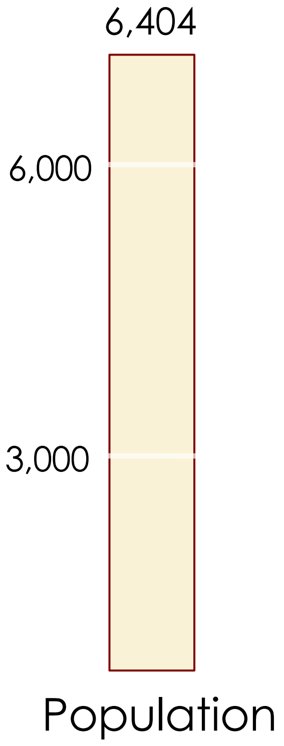
Area of interest

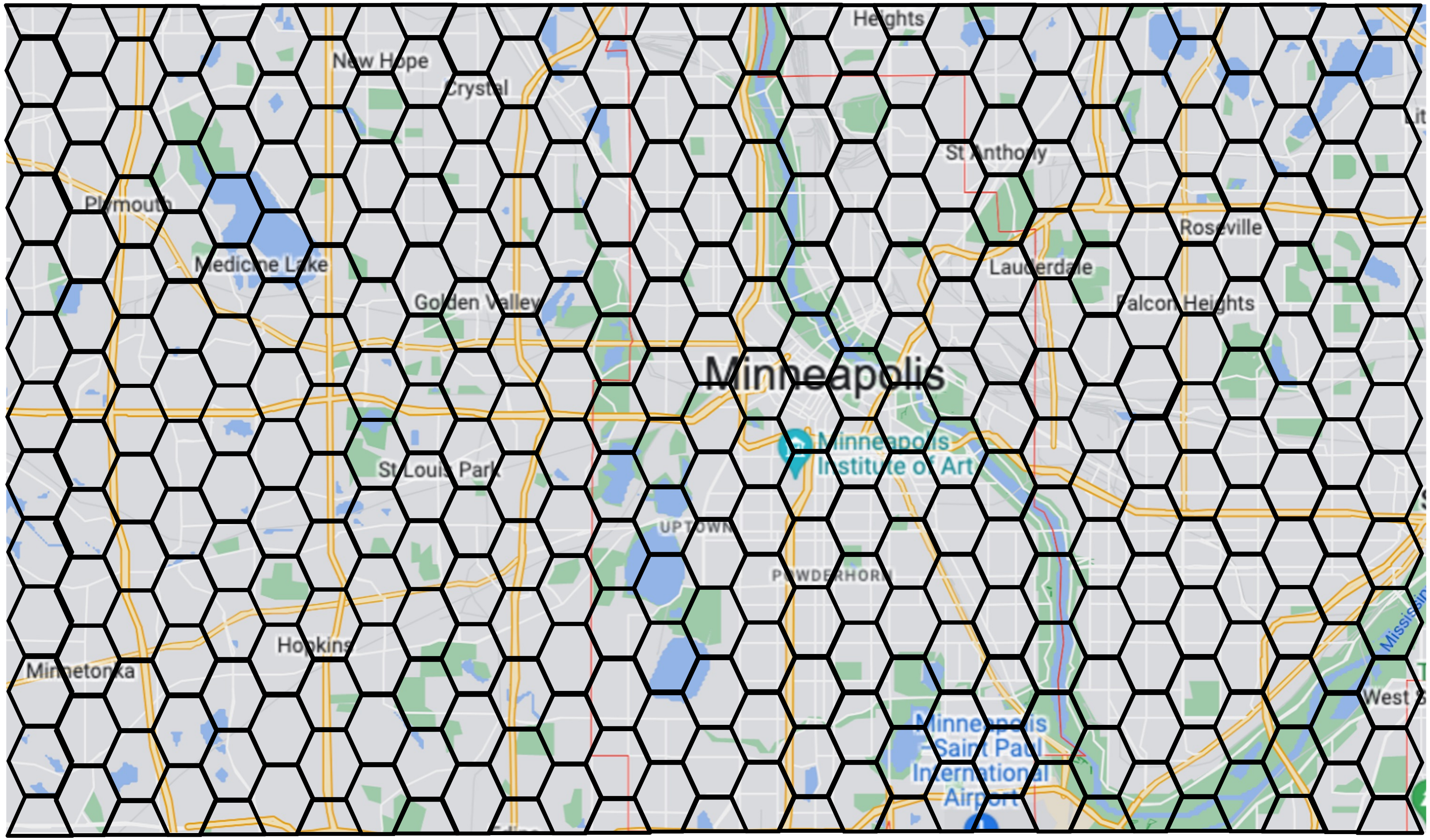


Google Maps

Random Sampling

Monthly income, per capita ($)

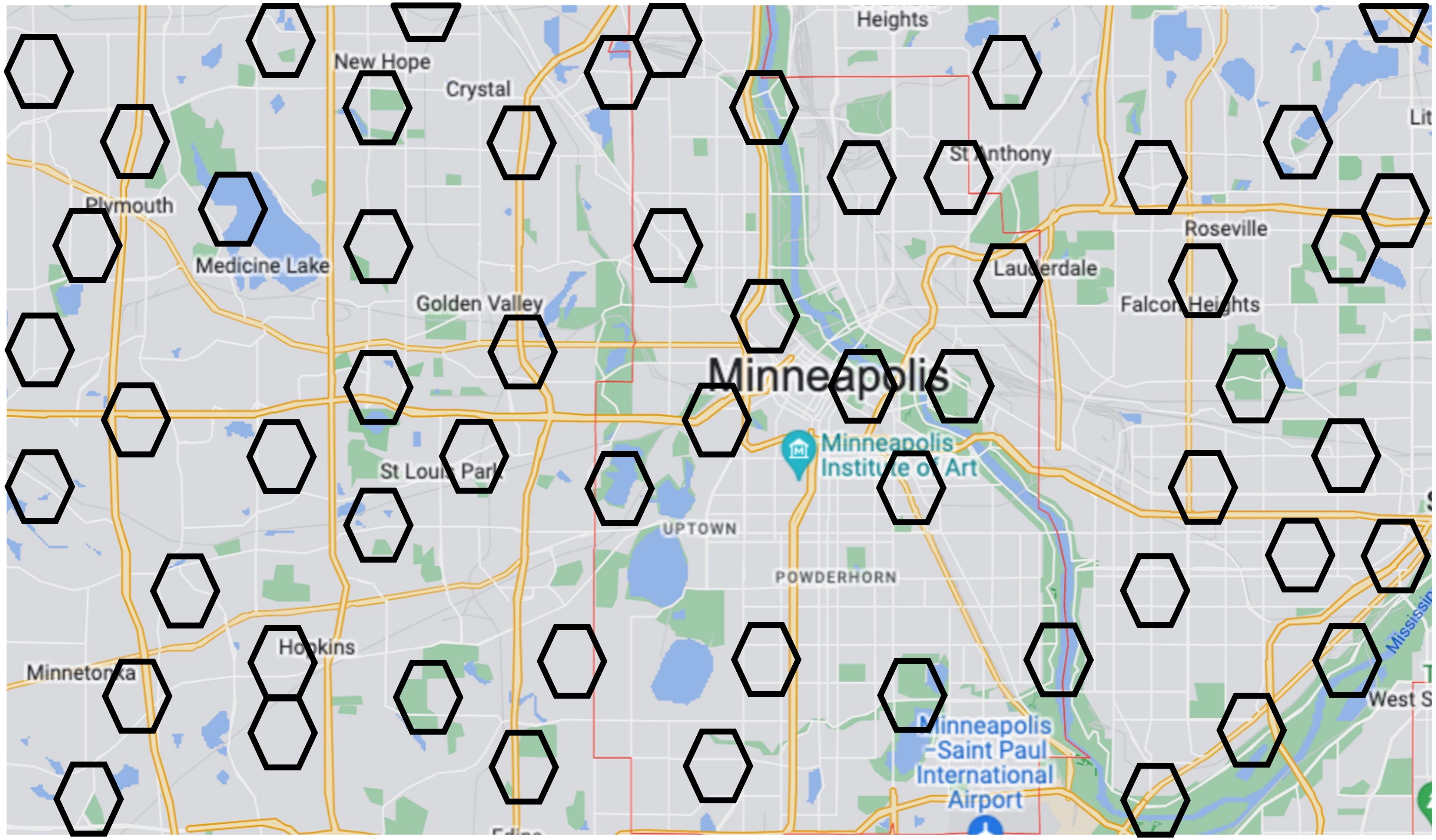




Google Maps

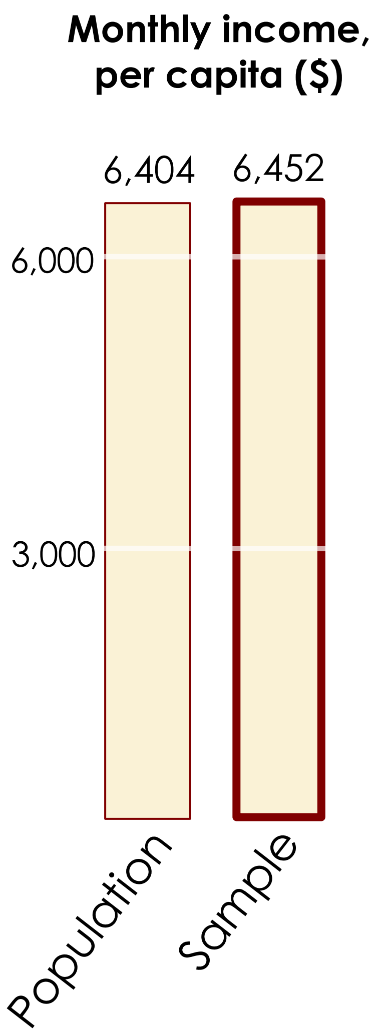
Random Sampling

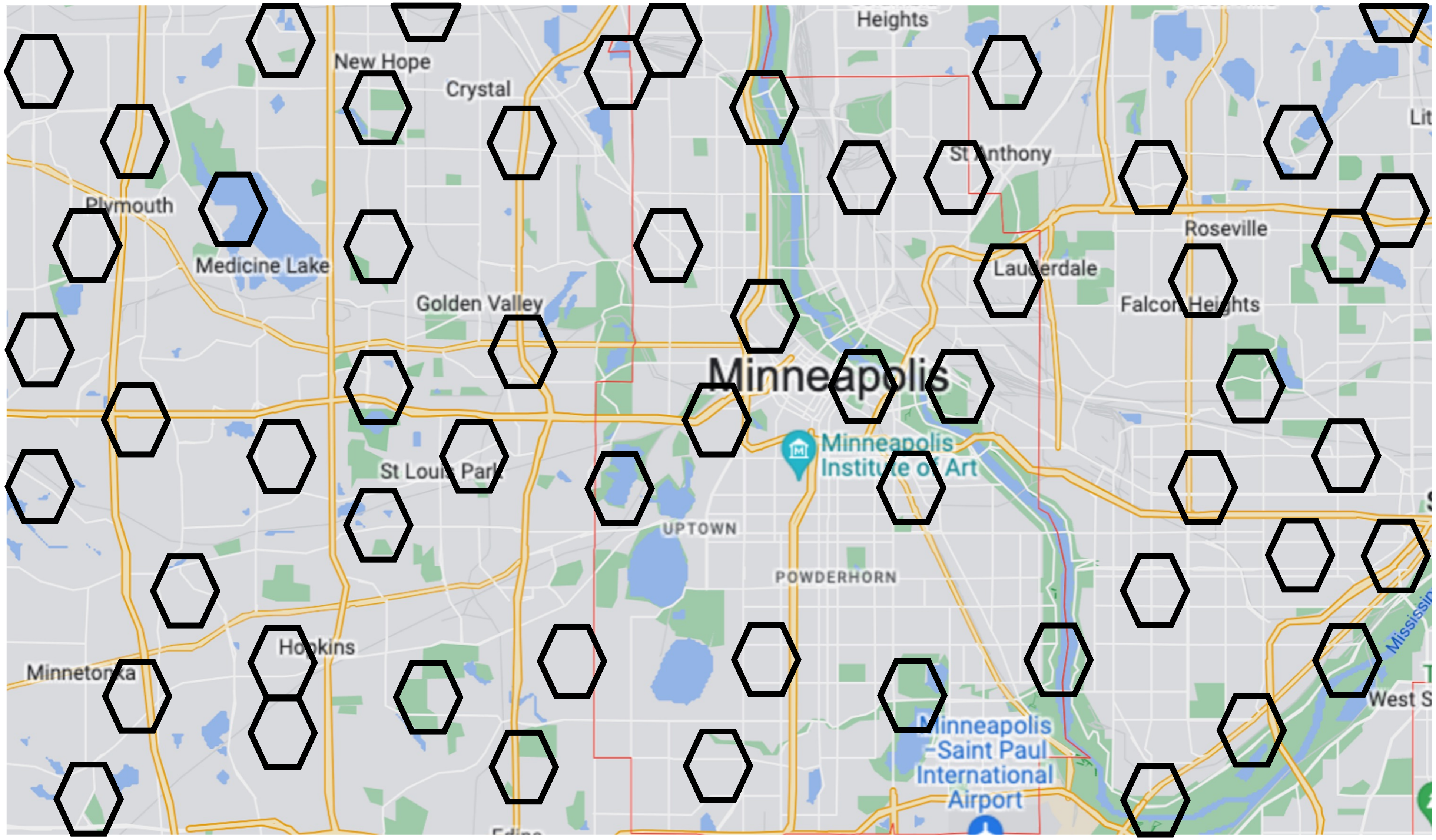
Randomly ***sample*** from area of interest



Google Maps

Random Sampling

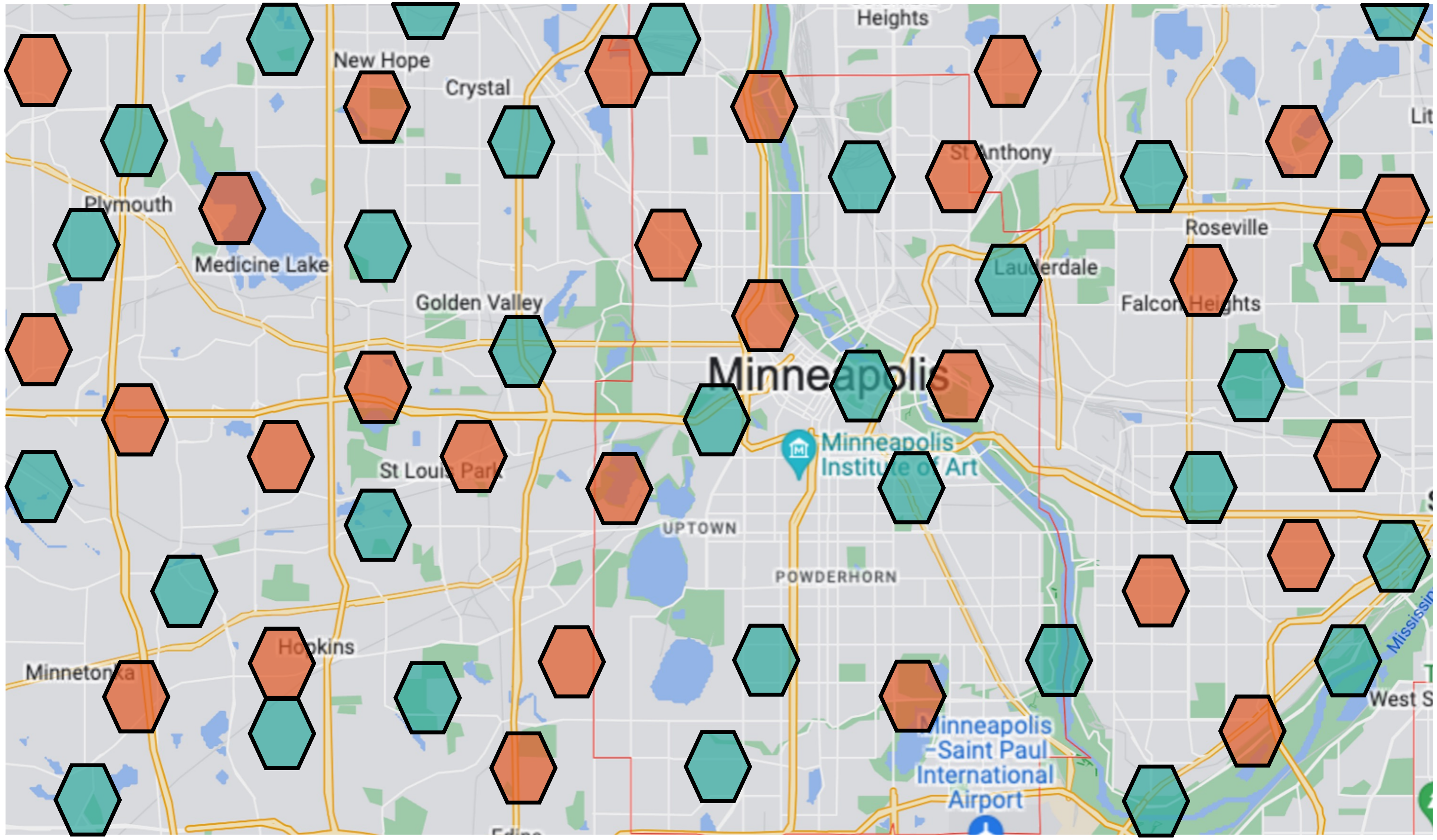




Google Maps

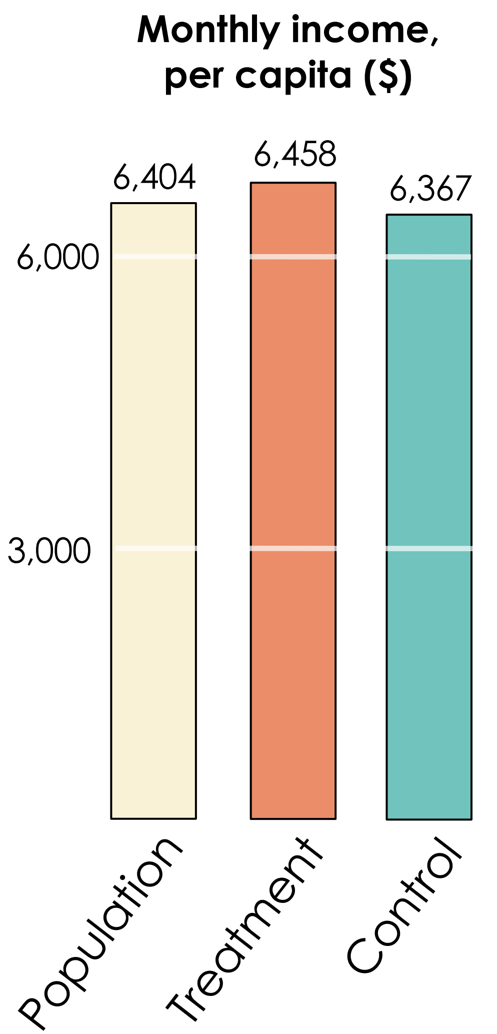
Random Assignment

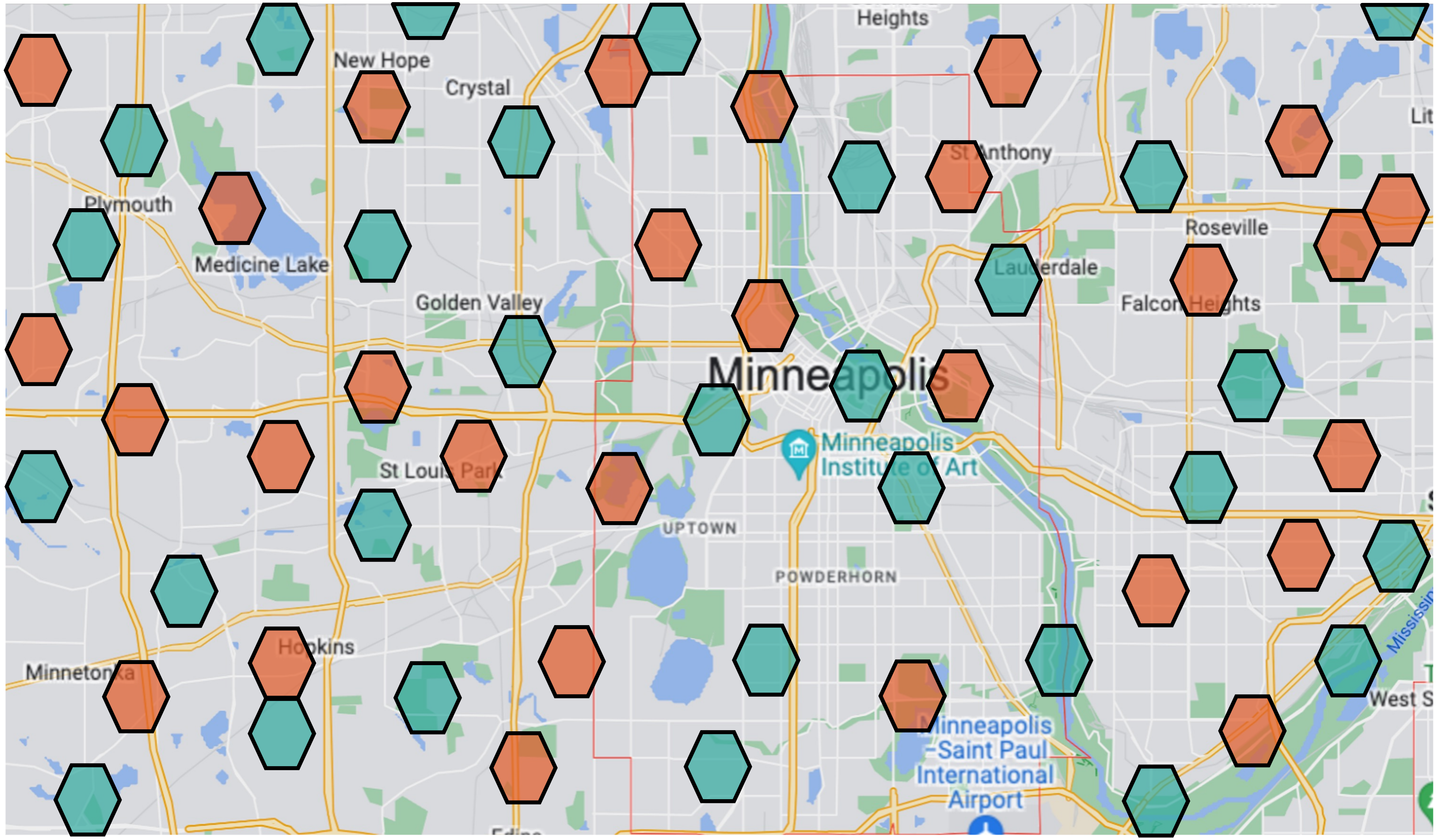
Randomly ***assign*** to **treatment** and **control**



Google Maps

Random Assignment





Google Maps

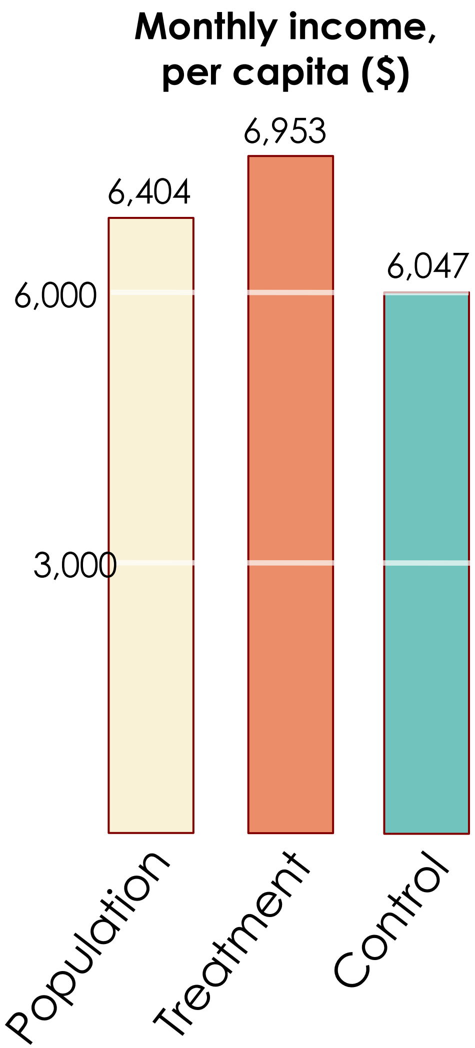
Suppose we take a random sample from our target population and assign the participants in the **northern half of the sample to the treatment**, and the participants in the **southern half of the sample to the control**. This is an example of:

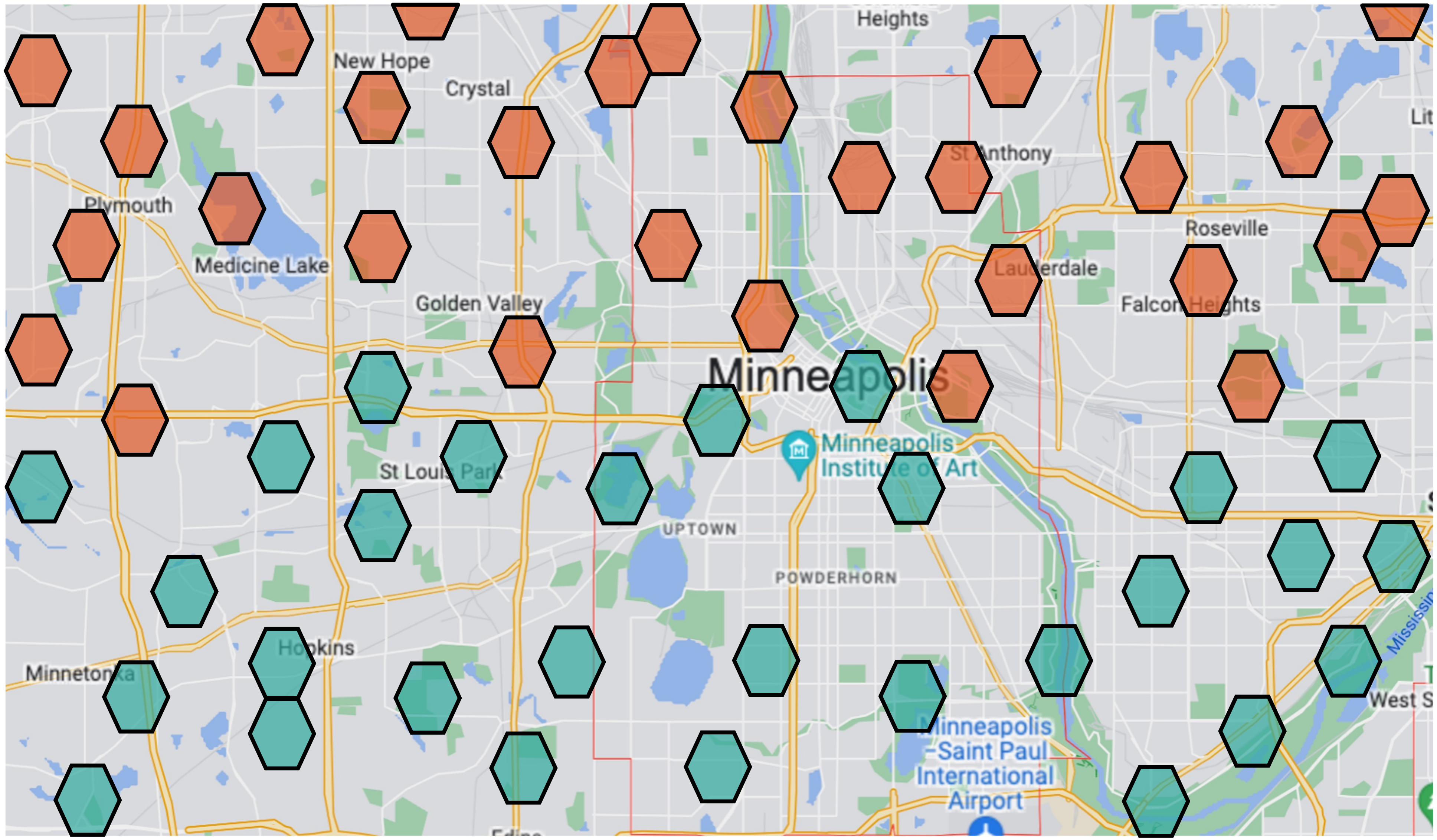
A. Random sampling but not random assignment
B. Random assignment but not random sampling
C. Random sampling and random assignment
D. Neither random sampling nor random assignment
E. Unsure

Suppose we take a random sample from our target population and assign the participants in the **northern half of the sample to the treatment**, and the participants in the **southern half of the sample to the control**. This is an example of:

A. **Random sampling but not random assignment**
B. Random assignment but not random sampling
C. Random sampling and random assignment
D. Neither random sampling nor random assignment
E. Unsure

Not Random Assignment



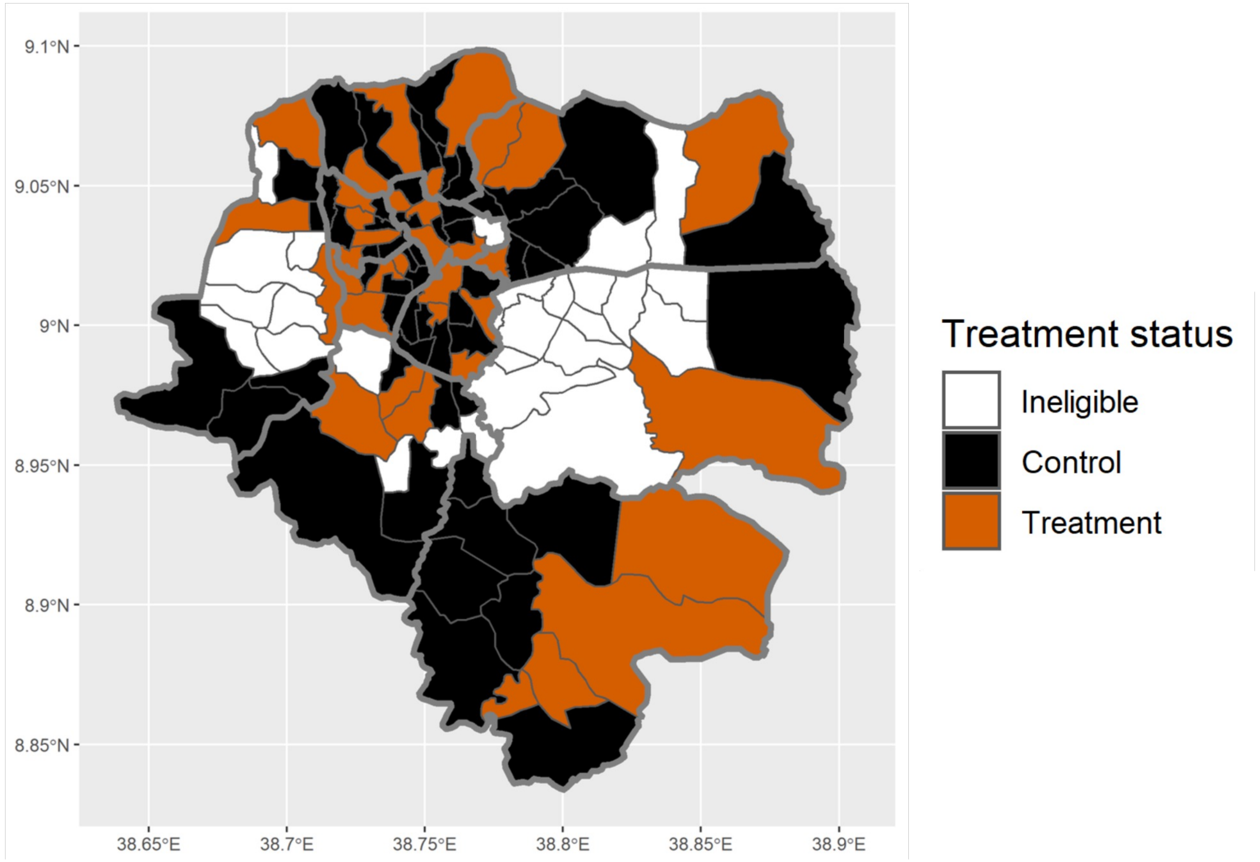


Google Maps

A real-world example

- In practice, your area of interest would be divided intentionally, such as by neighborhood or community boundaries
- A recent study in Addis Ababa, Ethiopia leveraged the gradual rollout of a large public works program across randomly selected neighborhoods (woredas) to measure the direct and indirect effects of the program

Randomization of the program across eligible woredas



Franklin, Simon, Clement Imbert, Girum Abebe, and Carolina Mejía Mantilla. *Urban Public Works in Spatial Equilibrium: Experimental Evidence from Ethiopia*. Centre for Economic Policy Research, 2021.

Lecture Overview

- What is Randomization?
- Randomization Procedures
- The Unit of Randomization
- Balance & Stratification
- Designing Randomized Evaluations for Different Program Specifications

Basic Randomization Design

- Randomization can be appropriate when a program is oversubscribed and resource constraints prevent everyone who is eligible from receiving the treatment
- A "lottery" design can be easy to understand and implement
  - Randomly assign treatment and control from sample of those who are eligible
  - Already used in the real world
    - Examples: charter school admissions; some housing authorities run lotteries for Housing Choice Vouchers

Conducting a "Lottery" Design:
Randomize with a Complete List of Study Participants

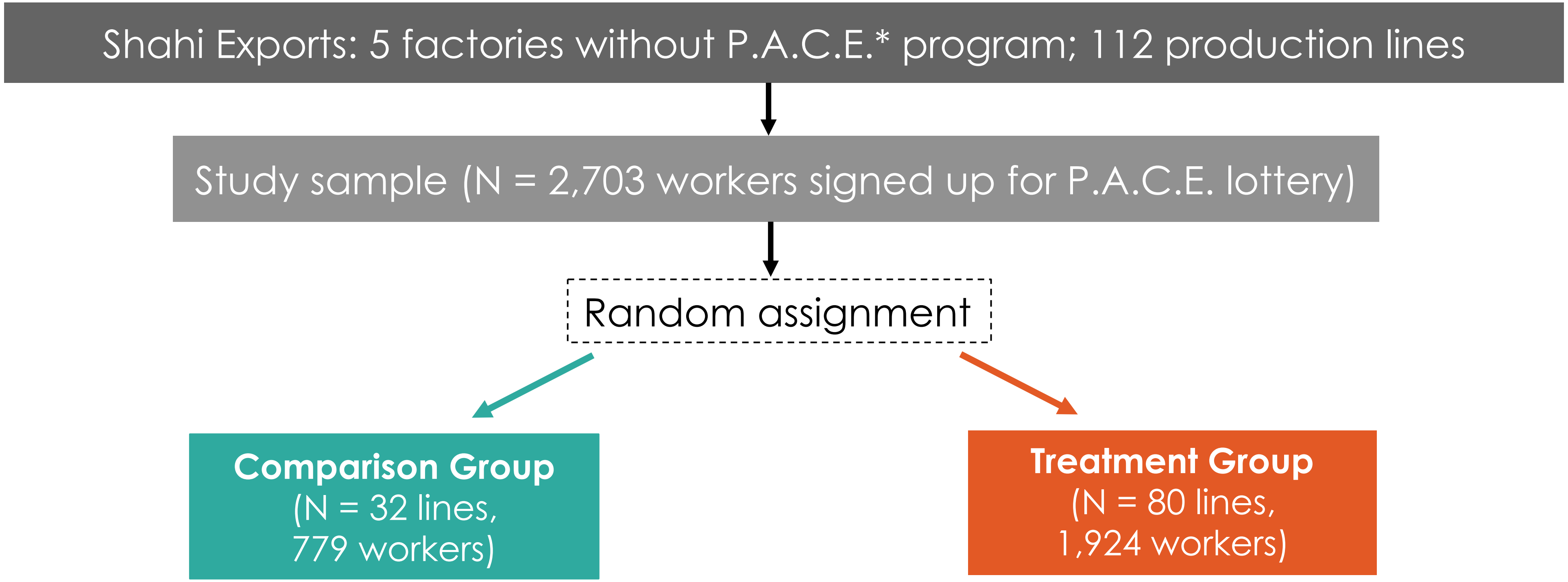
How does it work?

- Compile a list of all participants that will be in your study
- Determine the number of units that you want in treatment and in control
  - This number is based on logistical factors and on statistical power calculations
- Assign to treatment or control, typically using a random number generator

	schoolid	areaid	pretest_mean	random	treatment
50	503	5	22.1532	.3928247	1
51	130	1	18.2115	.3937532	1
52	617	6	35.6716	.4257155	1
53	215	2	23.2667	.4363684	1
54	603	6	14.7627	.4451973	1
55	408	4	25.0333	.4523951	1
56	529	5	14.7895	.4550135	1
57	602	6	37.8	.4553497	1
58	633	6	29.506	.4629052	1
59	514	5	25.2937	.4742467	1
60	639	6	20.4862	.4871492	1
61	509	5	23.2927	.5233575	0
62	346	3	36.3158	.5371823	0
63	625	6	28.8974	.5402876	0
64	411	4	20.6154	.5585051	0
65	132	1	34.7232	.5694133	0
66	314	3	22.9275	.5766308	0
67	320	3	36.5556	.5860199	0
68	502	5	16.8955	.595137	0
69	114	1	16.3226	.5979754	0
70	202	2	19.5256	.6041152	0

Example: For each unit (e.g. a school), a random number generator has assigned a random number in the "random" column. Treatment is assigned to those with a number of less than 0.5

Example: Personal Advancement and Career Enhancement (P.A.C.E) program



Adhvaryu, Achyuta, Namrata Kala, and Anant Nyshadham. "Returns to On-The-Job Soft Skills Training." *Journal of Political Economy*, February 1, 2023. https://doi.org/10.1086/724320.

*Personal Advancement and Career Enhancement program

"Rolling Randomization" Design: Randomize Each Participant Upon Entry into Study

How does it work?

- Set the probability of assignment to treatment group to a fixed percentage (e.g., 50%, 75%)
- Conduct a point-of-service randomization
  - Example: when a new participant comes to a study site, a research staff member runs randomization code through software to randomize the participant to treatment or control, after conducting a process of informed consent

ID	Coin	Treatment /Control
1	Heads	T
2	Heads	T
3	Tails	C
4	Heads	T
5	Tails	C
6	Heads	T
7	Tails	C
8	Tails	C
9	Heads	T
10	Heads	T
Count:		**T: 6** **C: 4**

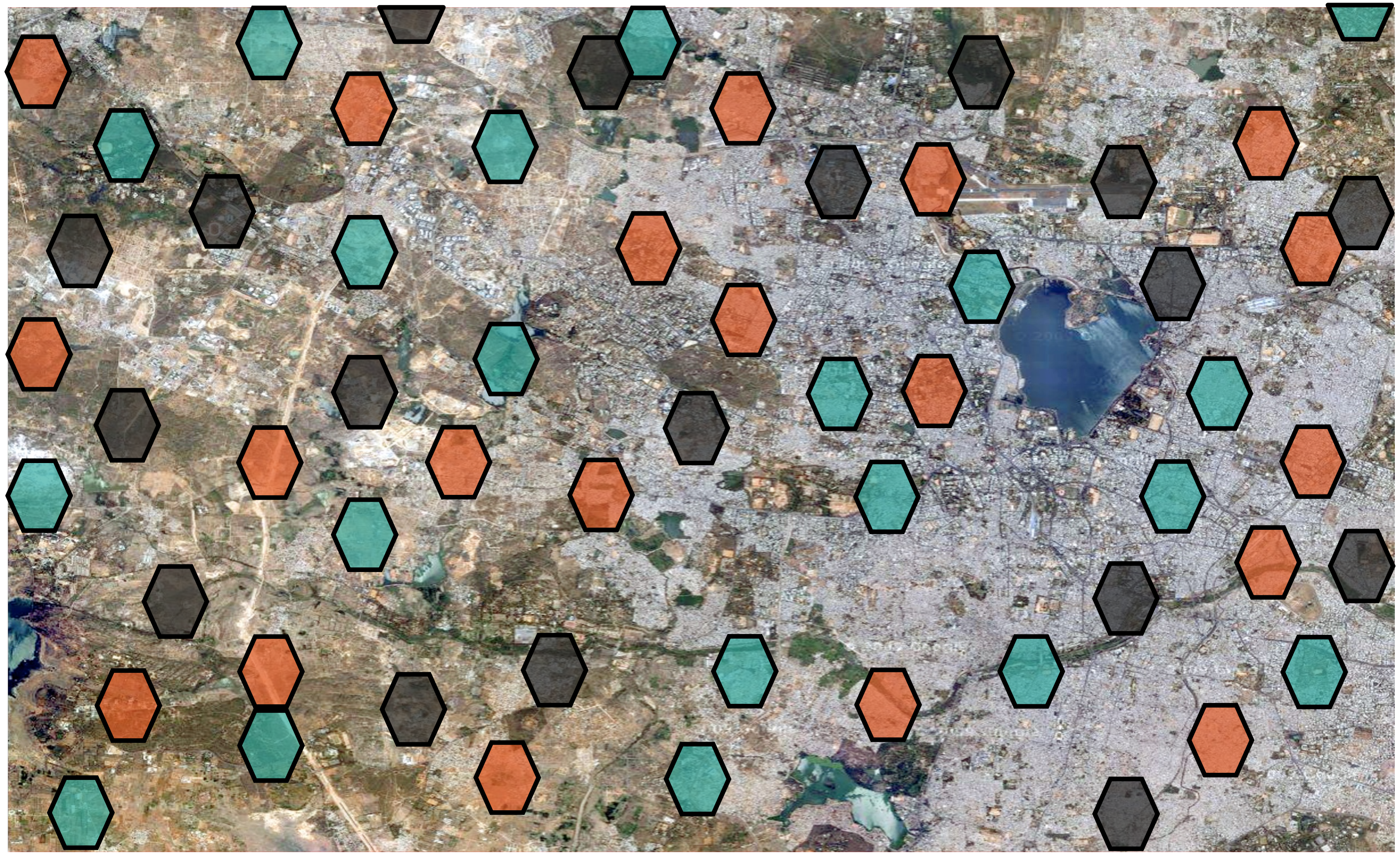
Multiple Treatments

Question: How do different treatments compare?

Treatment 1

Treatment 2

Control



Pictured: Spandana, Hyderabad, the site of a microcredit study (Banerjee et al. 2015). These are not the actual treatment and control neighborhoods.

Multiple Treatments

Question: How do different treatments compare?

Example: Study A

Testing one treatment, compared to a control group that continues with "business as usual"

Treatment group: New intervention

Control group

Example: Study B

Testing multiple treatments, compared to a control group that continues with "business as usual"

Treatment group 1: New intervention #1

Treatment group 2: New intervention #2

Control group

Example: Study C

Comparing multiple treatments to each other

Treatment group 1: New intervention #1

Treatment group 2: New intervention #2

Multiple Treatments

Example: A randomized evaluation testing the impact of different delivery methods of Covid-19 information on mental health and knowledge of Indian migrant workers

Intervention Design: Randomly assign migrant workers to one of three groups

1:
Covid-19 information via **text message**

2:
Covid-19 information via **pre-recorded audio message**

3:
Covid-19 information via **live phone call**

Results: Phone calls as effective as other methods at increasing knowledge, but had better mental health outcomes

Sadish, D., Achyuta Adhvaryu, and Anant Nyshadham. "(Mis) information and anxiety: Evidence from a randomized Covid-19 information campaign." *Journal of Development Economics* 152 (2021): 102699.

Cross-cutting Treatments: Factorial Design

Question: How do different treatments interact?

		Reduced search cost: Yes	Reduced search cost: No
Improved information	Yes	Group 1 Reduced search cost + Improved information	Group 2 Improved information
	No	Group 3 Reduced search cost	Group 4 Control

Adhvaryu, Achyuta et al. 2021. "Using AI to Expand the Job Search of Displaced Workers in the Aftermath of the Covid-19 Crisis." AEA RCT Registry. January 29. https://doi.org/10.1257/rct.7099-1.0

When is it necessary/not necessary to have a “pure control” group?

… meaning, a group that is not assigned to any treatment in an evaluation and continues with “business as usual”

Do we always need a "pure control" group?

When do we need it?

If you need to estimate the impact of a treatment compared to a counterfactual of no treatment (a "business as usual" approach), then you would need a "pure control" group.

When do we not?

If you are more interested in testing the effectiveness of different treatments relative to each other, then you could assign participants to different treatment groups, with no "pure control" group.

Varying Intensity of Treatment

This type of research design can help answer the question: “What is the effect of varying the intensity of the treatment?”

At the individual level:

- Dosage/level
    - Sensitivity
- Price reduction
    - Demand elasticity

At the group/cluster level:

- Externalities
    - Ex. Saturating labor markets

Varying Intensity of Treatment

This type of research design can help answer the question: “What is the effect of varying the intensity of the treatment?”

Example: What is the impact of managerial training for garment production line supervisors on productivity?

- Low floor saturation: 0-33% of supervisors trained
- Medium floor saturation: 35-63% of supervisors trained
- High floor saturation: 66-100% of supervisors trained

Outcomes

- What kinds of spillover effects do we find?
- At what level of saturation do we find spillover effects?
- At what point in time do we find spillover effects?

Adhvaryu, Achyuta, Emir Murathanoglu, and Anant Nyshadham. "On the allocation and impacts of managerial training." (2022).

Lecture Overview

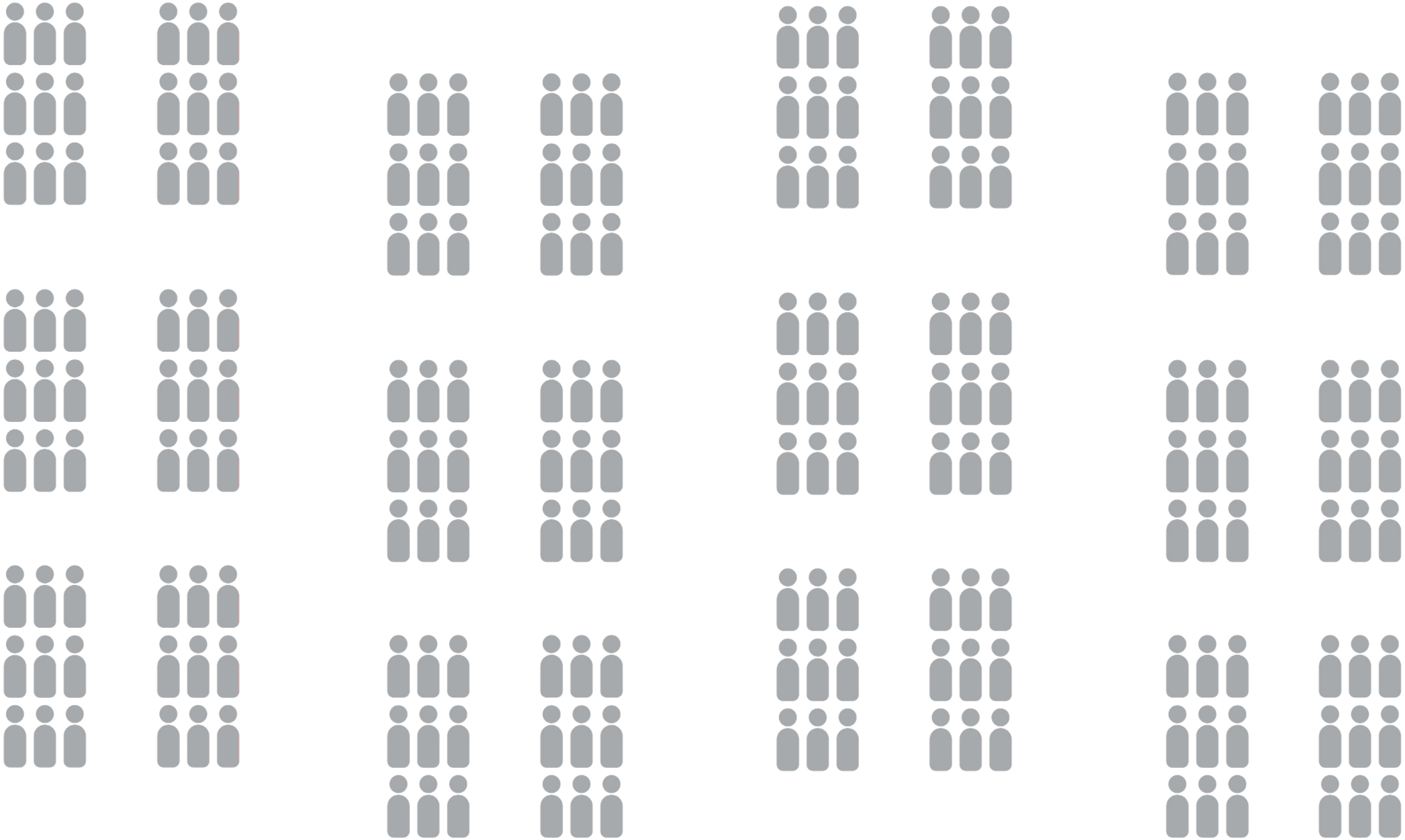
- What is Randomization?
- Randomization Procedures
- The Unit of Randomization
- Balance & Stratification
- Designing Randomized Evaluations for Different Program Specifications

Units of Observation and the Level of Randomization

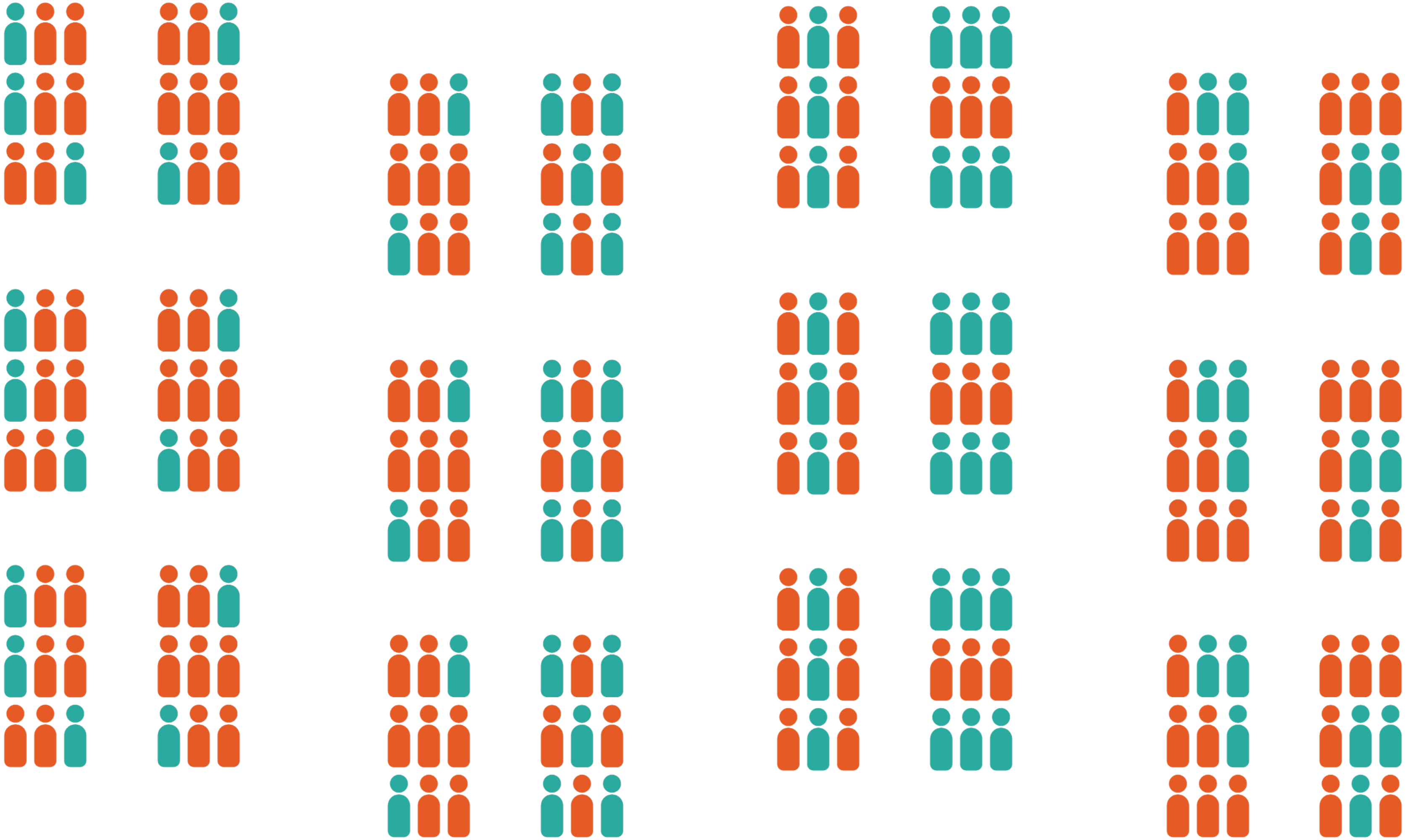
At which level should a study randomize?

- Randomizing at the individual level
  - e.g., people, patients, or students
- Randomizing at the group level
  - e.g., villages, clinics, or schools
  - *Outcomes can still be measured at the individual level*

Unit of Randomization: Individual?



Unit of Randomization: Individual worker?

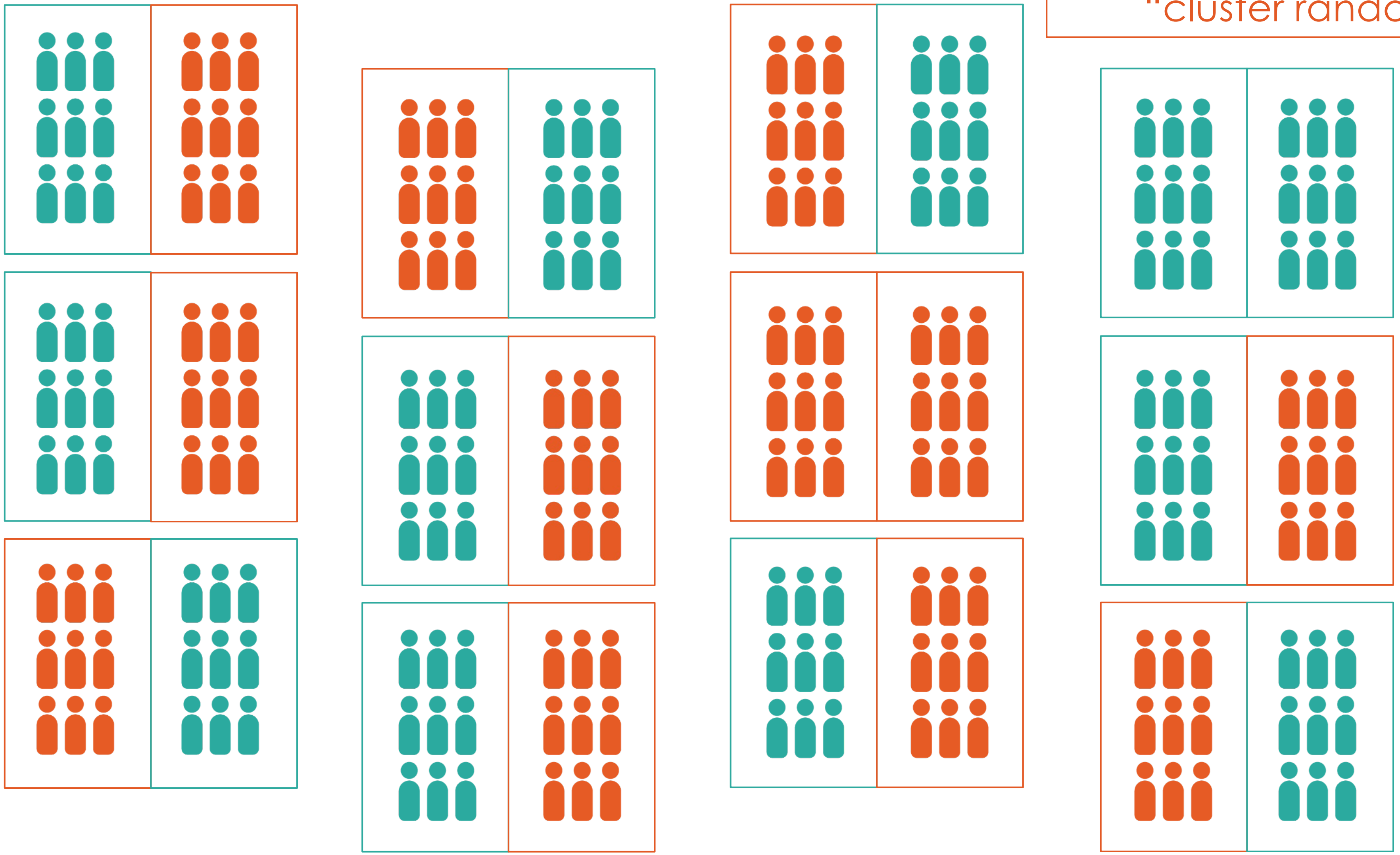


Unit of Randomization: Factory line?

We call groups of units "clusters"
Randomization at the group level:
"cluster randomized trial"

Unit of Randomization: Factory line?

We call groups of units "clusters"
Randomization at the group level:
"cluster randomized trial"



Unit of Randomization: Factory?

We call groups of units "clusters"
Randomization at the group level:
"cluster randomized trial"

Unit of Randomization: Factory?

We call groups of units "clusters"
Randomization at the group level:
"cluster randomized trial"

Level of Randomization: Considerations

- What is the unit of measurement?
  - What are the outcomes we care about, and at what level are we able to measure them?

- What unit does the program target for treatment?
  - At what level is the intervention administered? (e.g., classrooms for educational programs, factory lines for workforce projects)

- Which level gives us the greatest probability of detecting an effect?
  (This involves statistical power, which is the focus of tomorrow's first session)

- Reality check: Which level is feasible ethically, financially, politically, and logistically?

A state education department wants to see if increasing the duration of recess (a break from class for playtime) can help improve children's respiratory health. **What is an appropriate unit of randomization, and why?**

A. Student level
B. Classroom level
C. School level
D. District level
E. Unsure

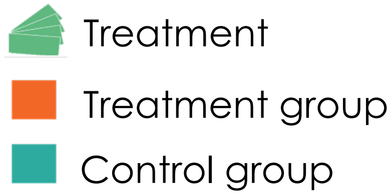
A state education department wants to see if increasing the duration of recess can help improve children's respiratory health. **What is an appropriate unit of randomization, and why?**

A. Student level
B. Classroom level
C. **School level –** *There is no single correct answer, but in this scenario, randomizing at the school level might be the most feasible and appropriate option. If students in the same classroom or classes within the same school had different lengths of recess, this might cause conflict or confusion. Randomizing at the district level might not provide a large enough sample size.*
D. District level
E. Unsure

Level of Randomization: Considerations

Choose a level of randomization to minimize **noncompliance** and to measure or contain **spillovers** (these are two different concepts)

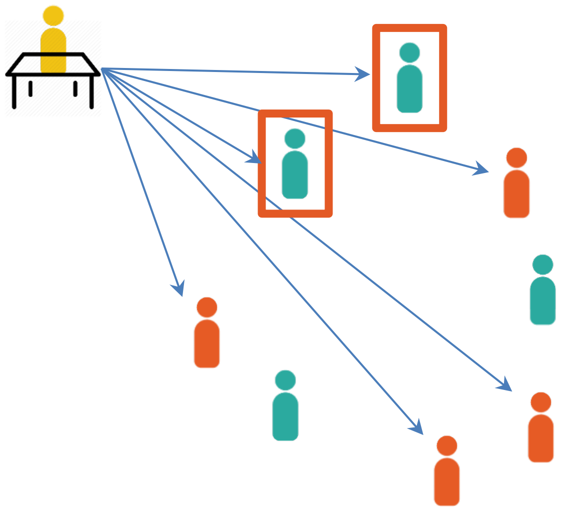
Noncompliance: When participants do not follow (“comply with”) their treatment assignment. For example, if participants assigned to the control group join the treatment group, or vice versa.



Spillover: When the treatment indirectly affects those who have not been treated. Spillover effects can be positive or negative.

Potential Sources of Noncompliance

- Logistical or political challenges: For example, service providers may find it difficult to administer treatment alongside their other responsibilities.
- Service providers might have trouble distinguishing between treatment and control (or customizing service), or may be unwilling to provide differential treatment.



Service provider

Participant assigned to treatment group

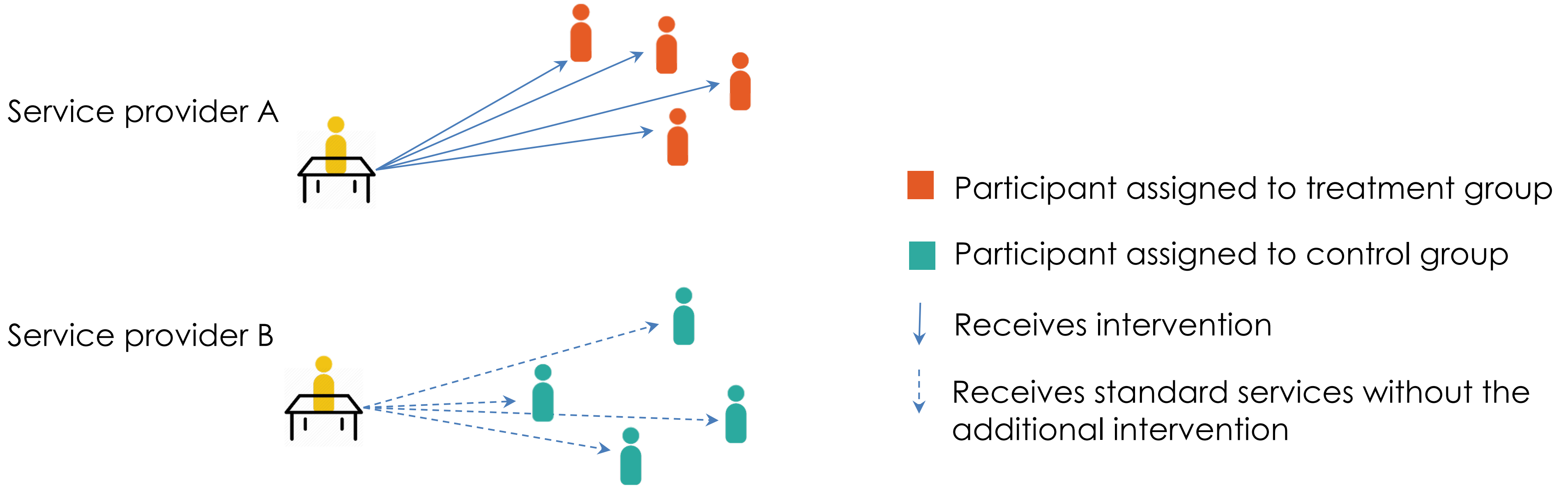
Participant assigned to control group

Service provider treats this participant

Members of the control group who "cross over" into the treatment group and receive intervention
(No longer represents the counterfactual)

Noncompliance Solution: Change Unit of Randomization to a Different Level–Randomize Service Providers

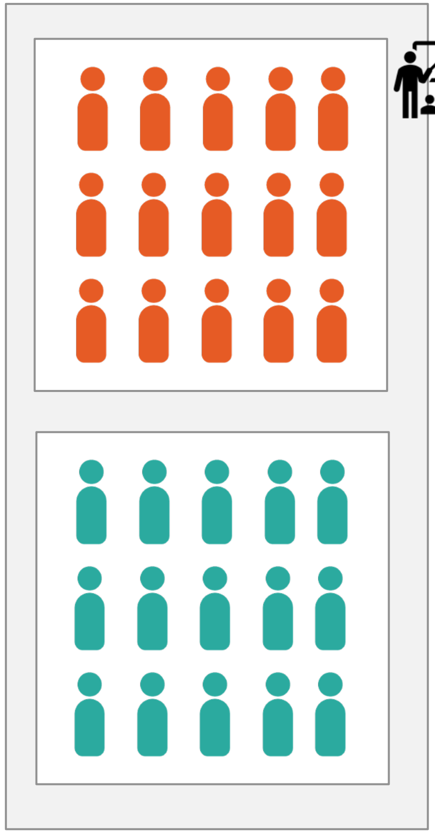
- Have different service providers administer the different treatments
- Randomly assign treatment/control to those service providers
- This enables providers to treat entire clusters the same



Spillovers Solution 1: Randomize at a Different Level

Randomizing at the production line level contains the positive spillovers within the treated production lines

Level of randomization: **production line**



Intervention: **training**

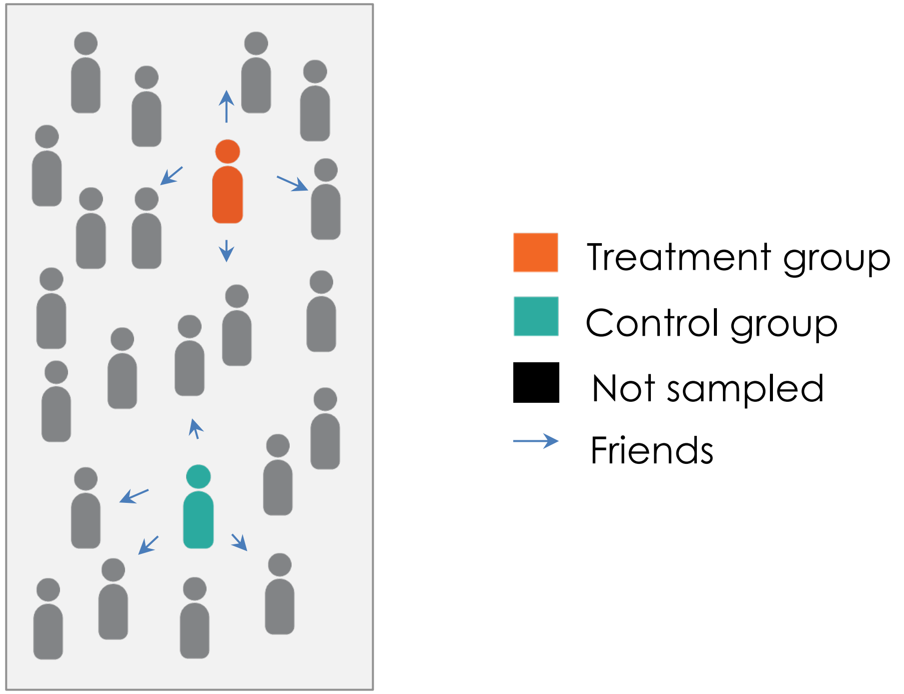
Treatment group

Control group

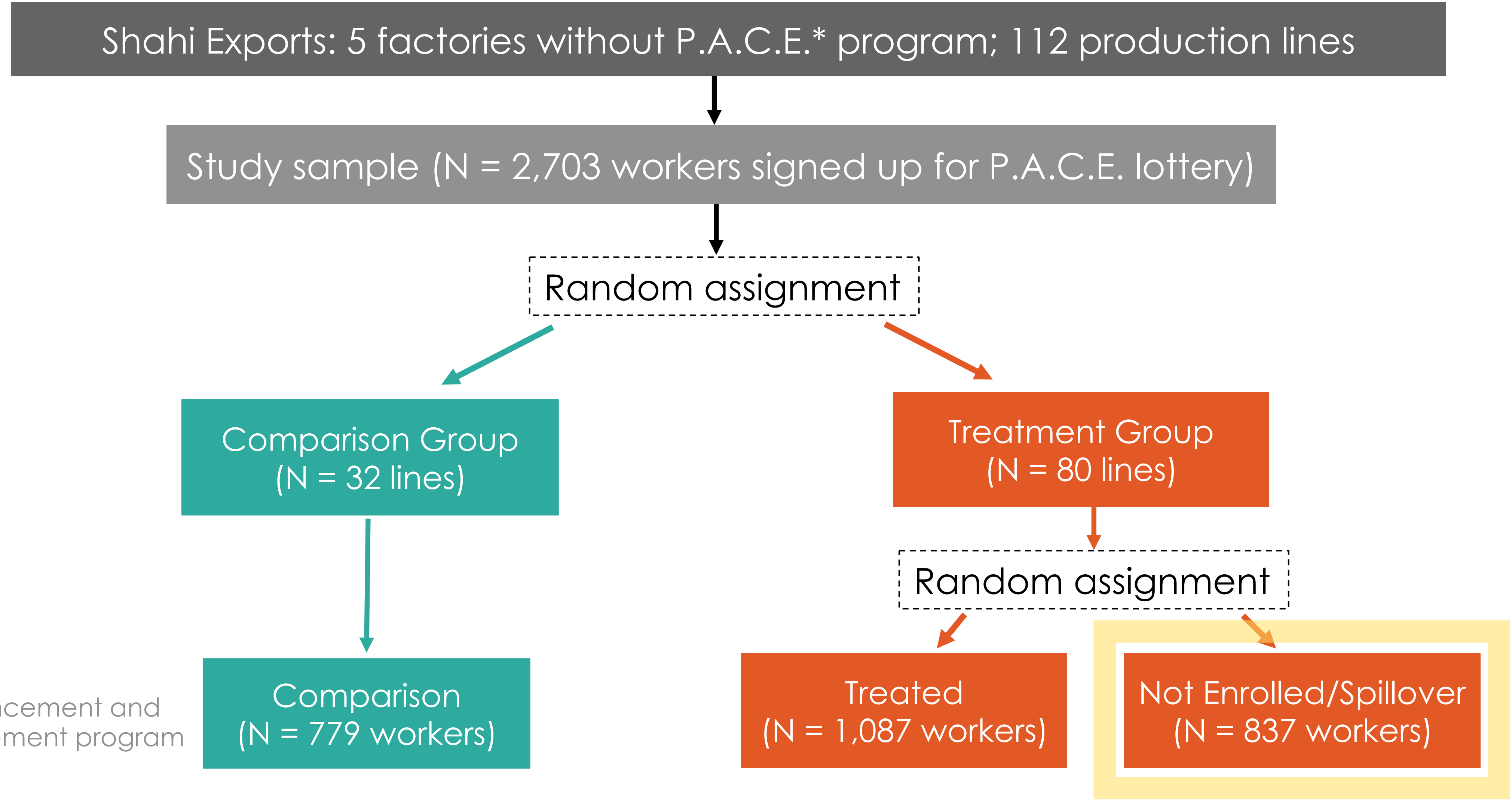
Spillovers Solution 2: Build in Buffers

Providing a buffer between the treatment and control subjects prevents treatment from spilling into control

Level of randomization: **individual**



Spillovers Solution 3: Measure Spillovers



*Personal Advancement and Career Enhancement program

Adhvaryu, Achyuta, Namrata Kala, and Anant Nyshadham. "Returns to On-The-Job Soft Skills Training." *Journal of Political Economy*, February 1, 2023.

Lecture Overview

- What is Randomization?
- Randomization Procedures
- The Unit of Randomization
- Balance & Stratification
- Designing Randomized Evaluations for Different Program Specifications

Balance

- **Balance** = the distribution of baseline characteristics in the treatment group is the same or similar to that in the control group

- Due to random assignment, the treatment and comparison groups will be balanced on average across baseline characteristics

- Even when randomization is done correctly, some of these values will be different by chance

Treatment

Comparison

Source: freepik

Covid-19 information study balance checks

Table V: Randomization Balance

Variable	TM	VR	PC	TM - VR	VR - PC	TM - PC
Female	0.58	0.56	0.56	0.02	0.00	0.02
				(0.04)	(0.04)	(0.04)
Age	23.86	24.23	24.00	-0.37	0.23	-0.14
				(0.49)	(0.45)	(0.50)
Educated above grade 10	0.36	0.33	0.32	0.03	0.01	0.04
				(0.04)	(0.04)	(0.04)
Left job	0.54	0.50	0.50	0.04	0.00	0.04
				(0.04)	(0.04)	(0.04)
Smartphone	0.75	0.75	0.75	-0.00	0.01	0.01
				(0.04)	(0.03)	(0.04)
Lives in hostel	0.23	0.30	0.30	-0.06	0.00	-0.06
				(0.04)	(0.03)	(0.04)
Numeracy Score	2.21	2.18	2.20	0.03	-0.02	0.01
				(0.09)	(0.08)	(0.09)
Short-Term Memory	5.24	5.34	5.15	-0.10	0.19	0.09
				(0.17)	(0.14)	(0.16)
Knowledge	-0.02	-0.00	0.02	-0.02	-0.02	-0.04
				(0.09)	(0.07)	(0.09)
PHQ-4	3.07	2.92	3.04	0.14	-0.11	0.03
				(0.24)	(0.21)	(0.24)

The table reports statistics from t-tests comparing differences in means between the three intervention groups: Text Message (TM), Voice Recordings (VM), and Phone Calls (PC). Standard errors are in parentheses.

Sadish, D., Achyuta Adhvaryu, and Anant Nyshadham. "(Mis) information and anxiety: Evidence from a randomized Covid-19 information campaign." *Journal of Development Economics* 152 (2021): 102699.

Stratification

- Ensures groups are balanced on select characteristics
- Divide the sample into **subgroups** (also known as **strata**) that share certain characteristics and then select some from each subsample
  - Can be used for both sampling and treatment assignment
  - Example: men and women
- Usually, the same proportion is selected for treatment from each stratum, though this is not always the case
  - Especially in sampling: may want some groups overrepresented in sample to ensure they appear with high enough frequency (also called **disproportionate stratification**)
    - Example: widowed household heads

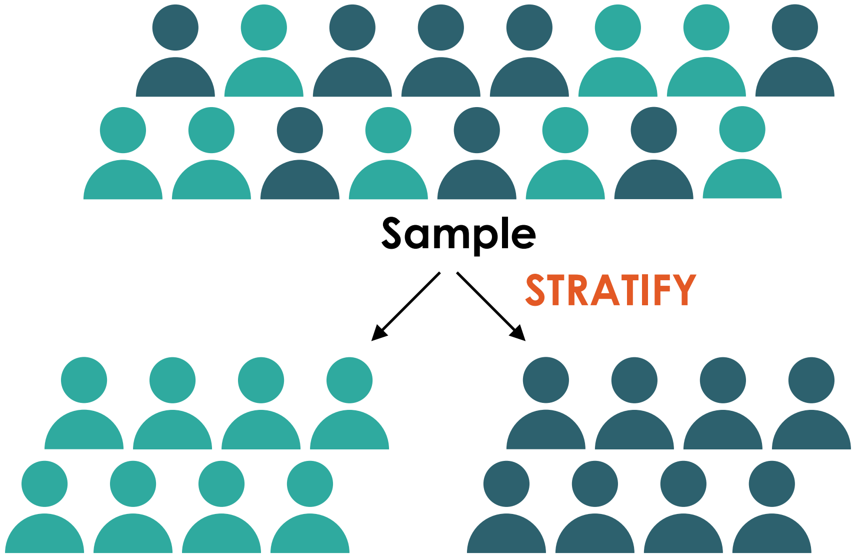
RCT Design: Stratification

Steps:

1. Form groups of similar units (individuals, households, schools)
2. Randomize within the group

Example:

1. Divide sample into education level
2. Randomize individuals separately within each education level



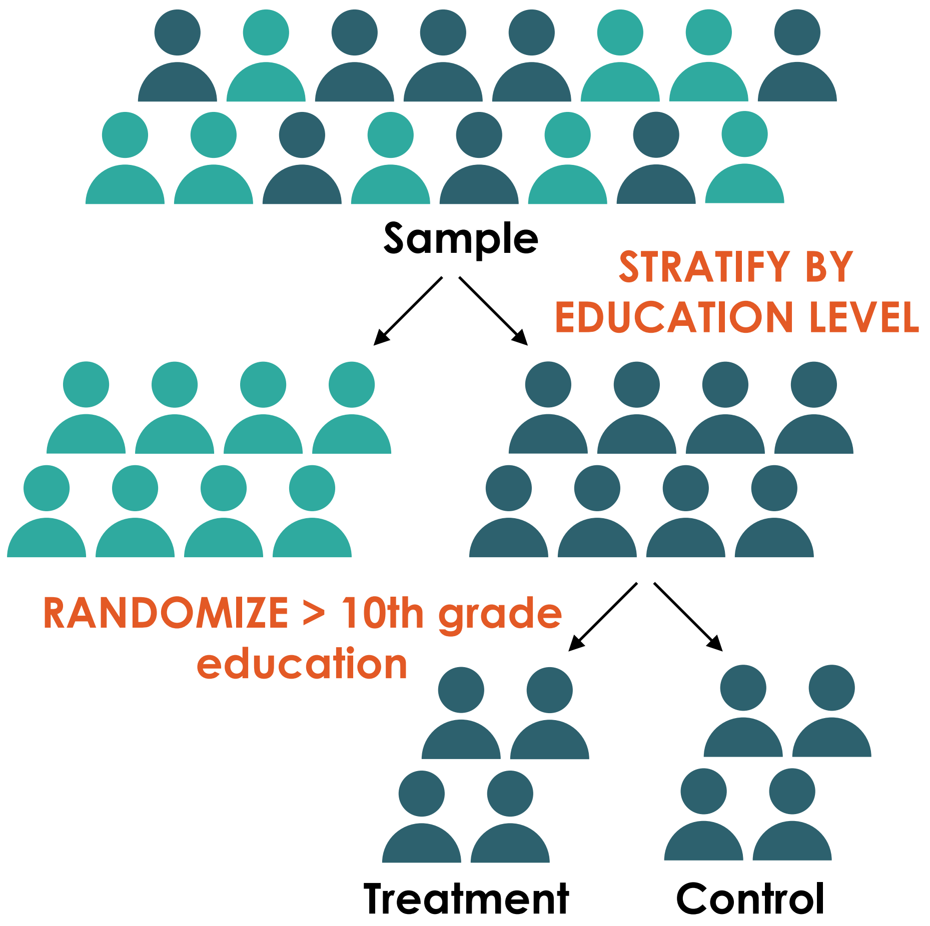
RCT Design: Stratification

Steps:

1. Form groups of similar units (individuals, households, schools)
2. Randomize *within* the group

Example:

1. Divide sample into education level
2. Randomize individuals separately within each education level



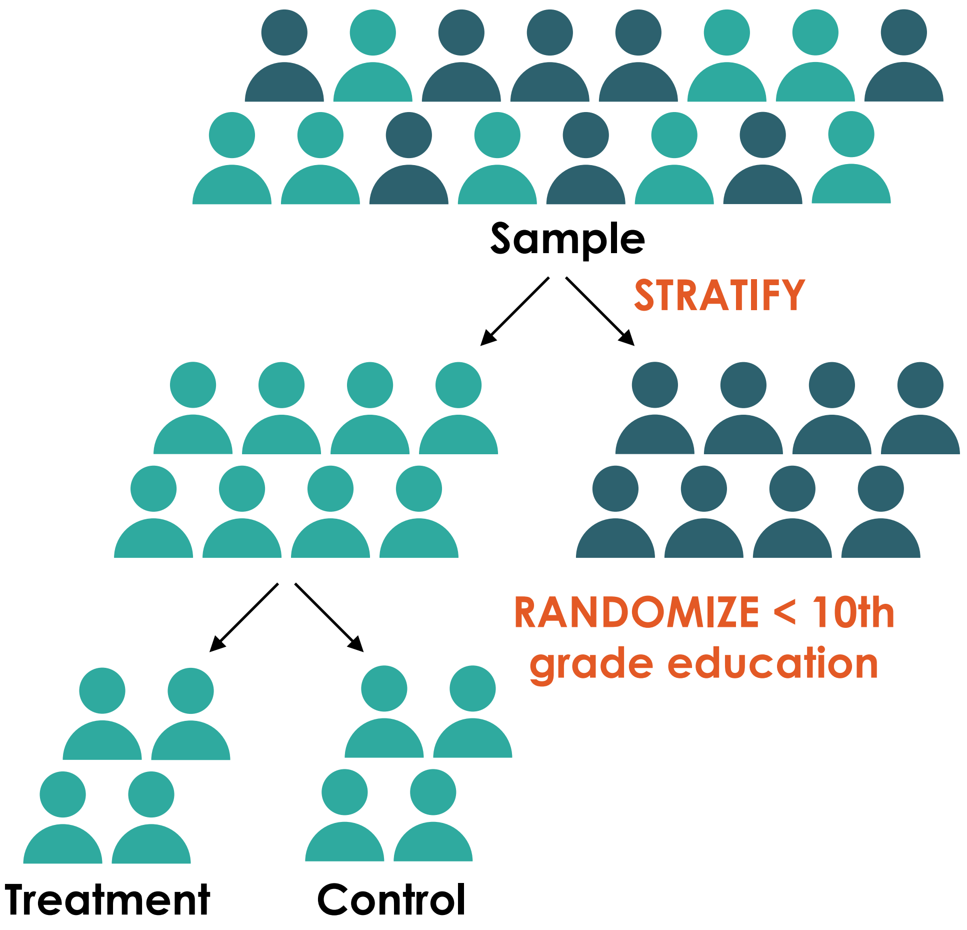
RCT Design: Stratification

Steps:

1. Form groups of similar units (individuals, households, schools)
2. Randomize *within* the group

Example:

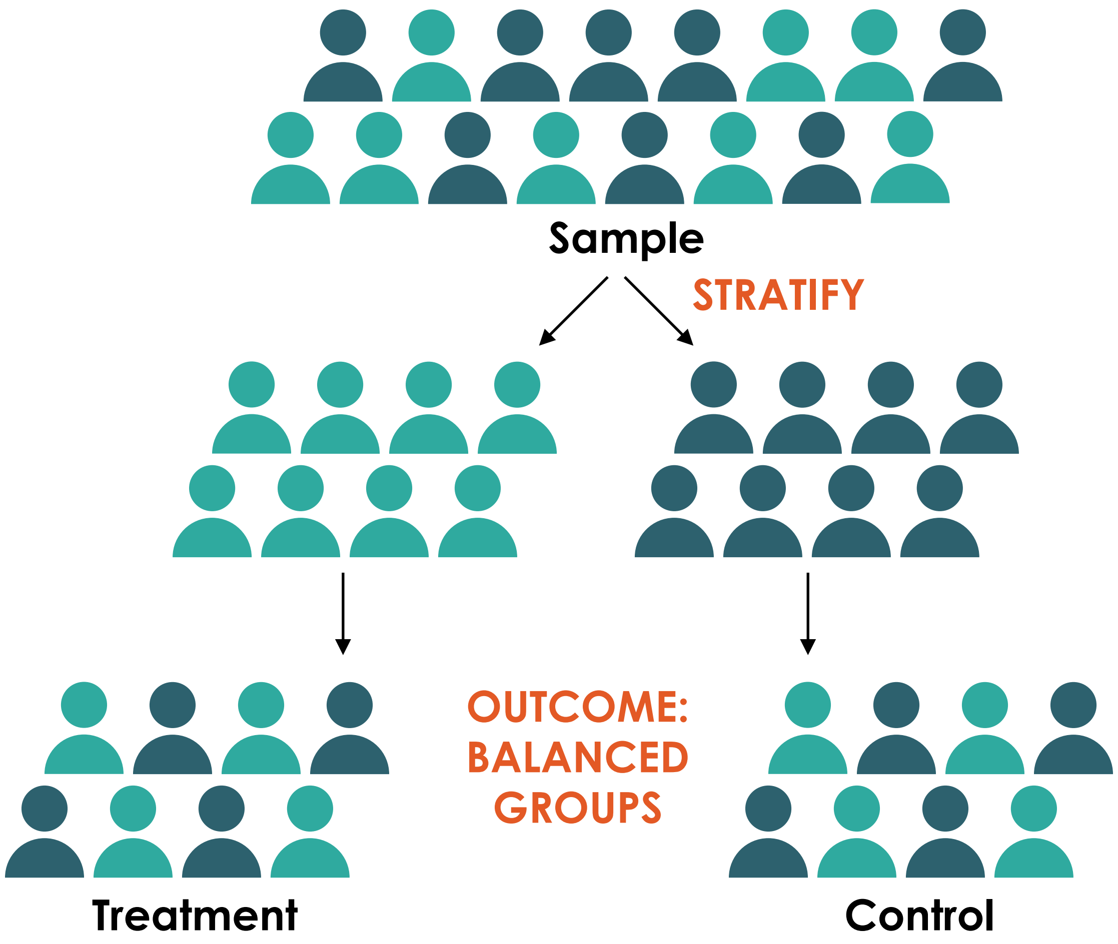
1. Divide sample into education level
2. Randomize individuals separately within each education level



RCT Design: Stratification

Randomization sets up an expectation of balance. Stratification ensures it.

But only for this one observable characteristic!



When to stratify

1. To **achieve balance**, i.e., ensure the treatment and control groups are comparable on key characteristics

1. To **increase statistical power** (we will return to this tomorrow)

1. To analyze treatment effect by group (also known as **subgroup analysis** or heterogeneous treatment effects)
    1. For example, younger versus older workers
    2. We may also care about differential effects by rural/urban, education level, race/ethnicity, etc.

Lecture Overview

- What is Randomization?
- Randomization Procedures
- The Unit of Randomization
- Balance & Stratification
- Designing Randomized Evaluations for Different Program Specifications
  - Program has Strict Eligibility Criteria
  - Resources Exist to Extend the Program to Everyone in the Study Area
  - Program is an Entitlement

Program Specifications

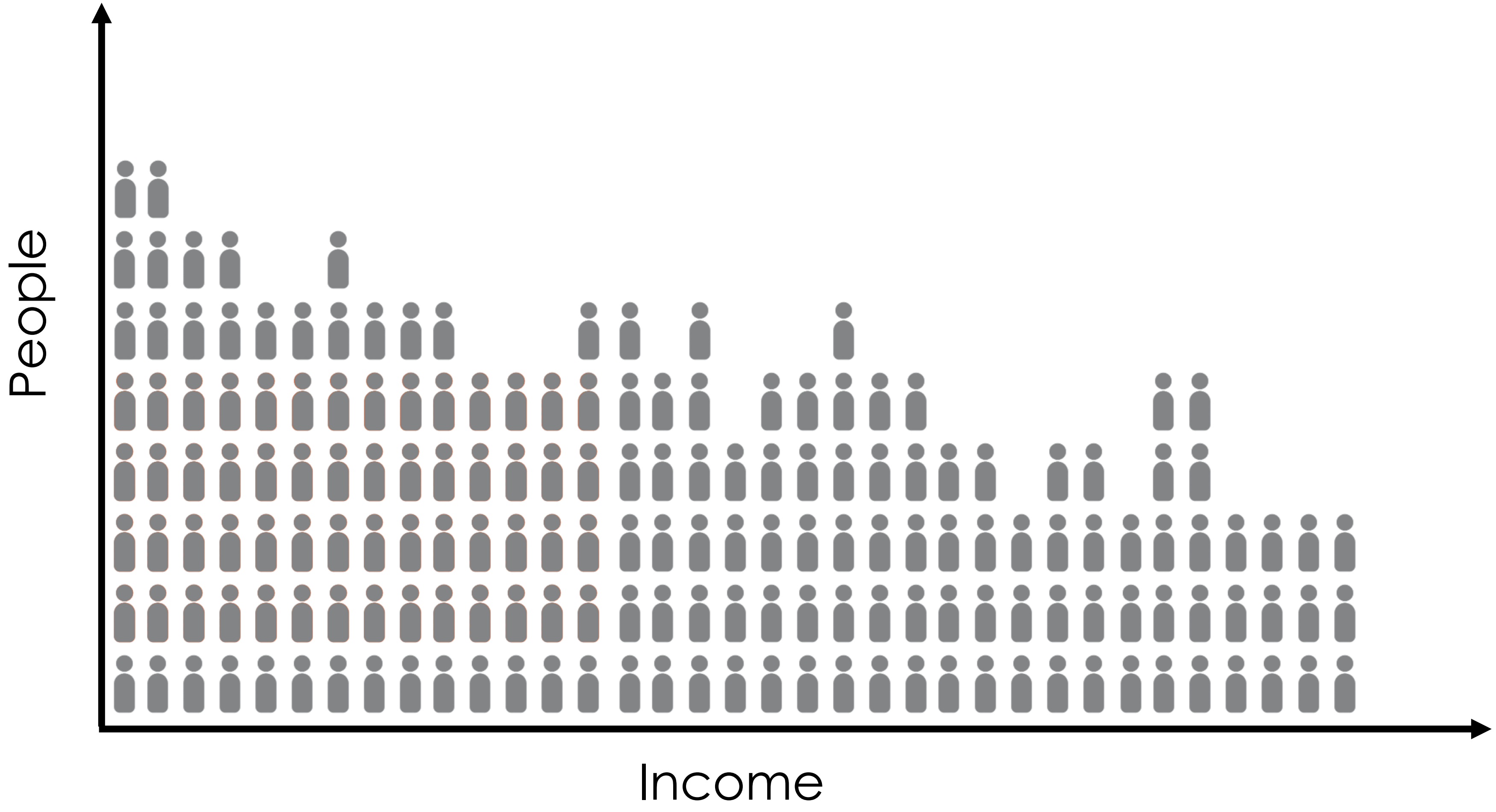
Evaluating programs with certain specifications requires adjustments. For example, adjustments might be required if:

- Eligibility is determined by a cutoff (e.g., income, credit score)
- There are enough resources to extend the program to everyone in the study area
- The intervention is a current entitlement, and participation cannot and/or should not be withheld from anyone who is eligible

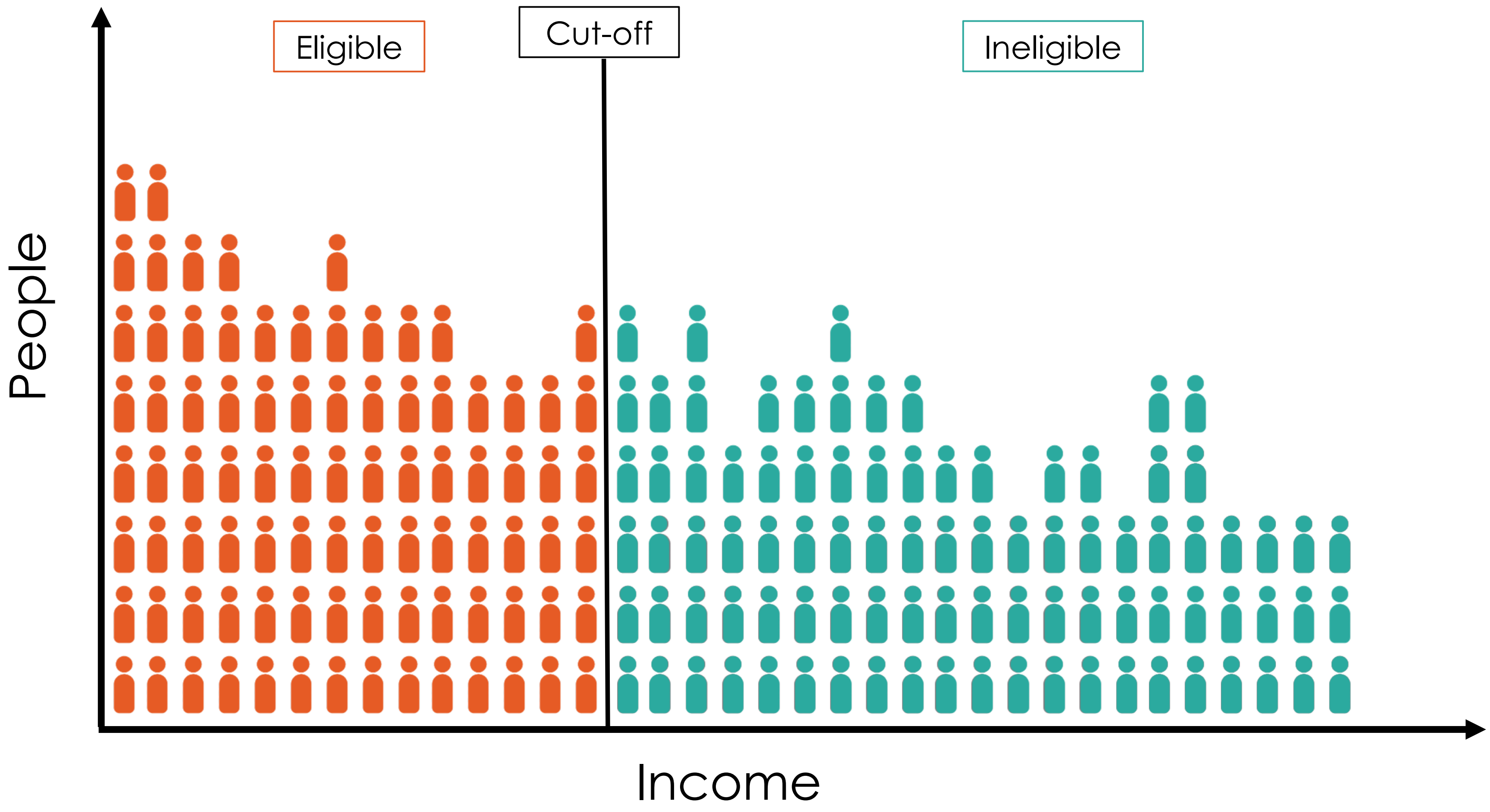
Program Specification:

Program has Strict Eligibility Criteria

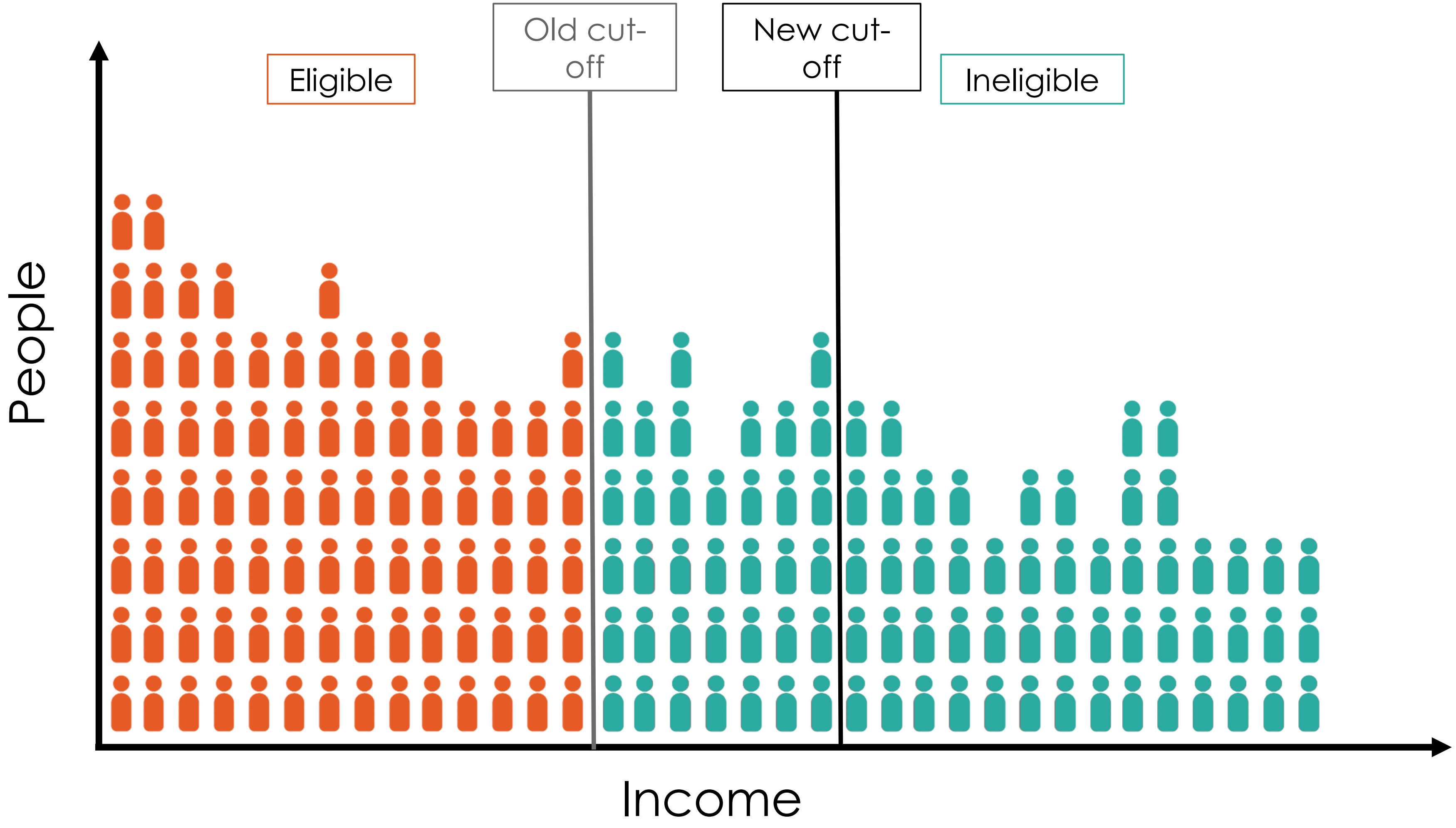
Program with an Eligibility Cut-off



Program with an Eligibility Cut-off



RCT Design: Expand Eligibility and Randomize among the Newly Eligible



RCT Design: Expand Eligibility and Randomize among the Newly Eligible

Program Specification:

Resources Exist to Extend the Program to Everyone in the Study Area

If it is not possible to randomize access to a program because you **cannot withhold treatment from the control group**, you might be able to conduct a randomized controlled trial by:

A. Randomizing timing of access to program

B. Randomizing encouragement to take up the program

C. Either A or B

D. Not sure

If it is not possible to randomize access to a program because you cannot withhold treatment from the control group, you might be able to conduct a randomized controlled trial by:

A. Randomizing timing of access to program

B. Randomizing encouragement to take up the program

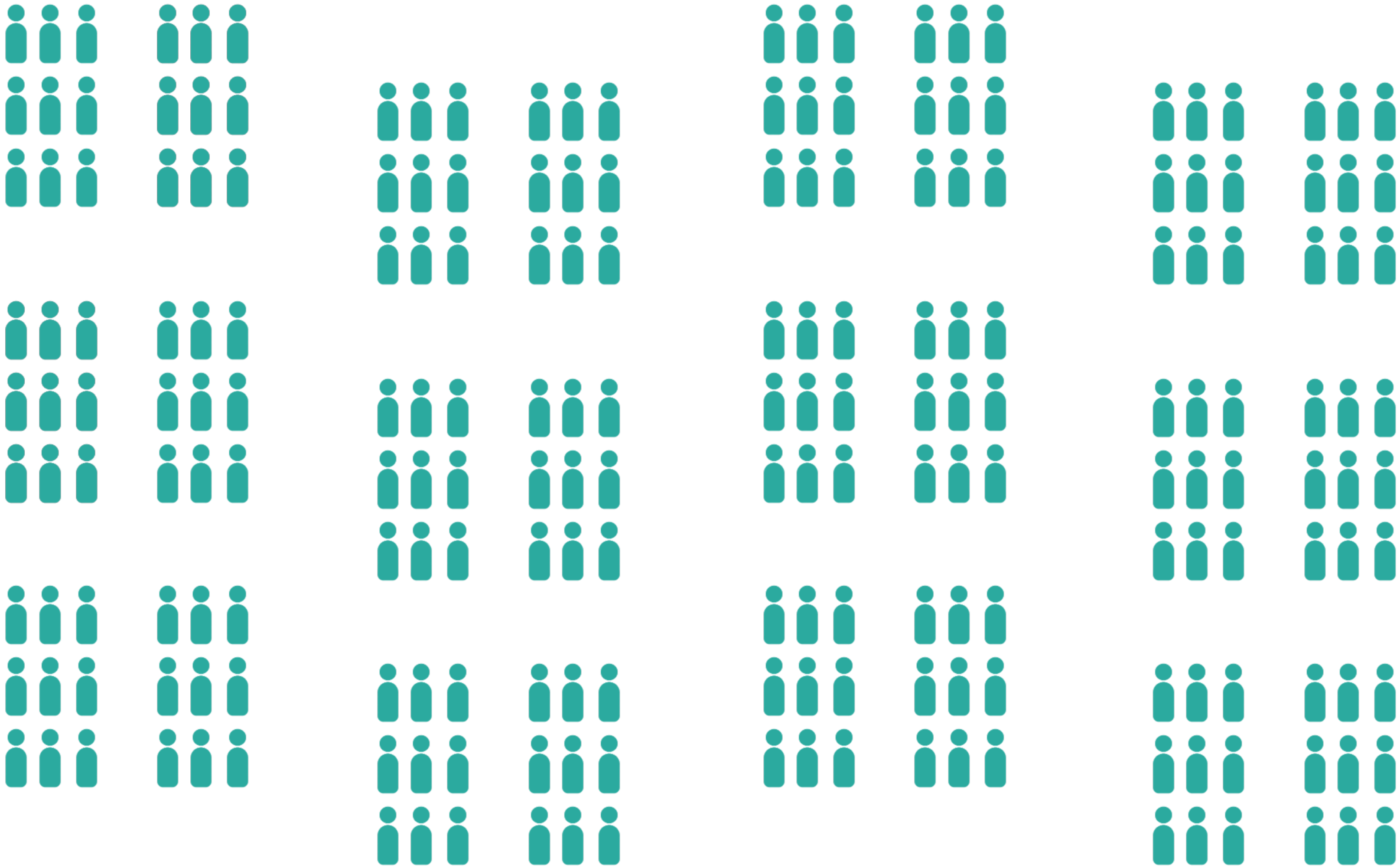
C. Either A or B – depending on the context, either of these options might be appropriate.

D. Not sure

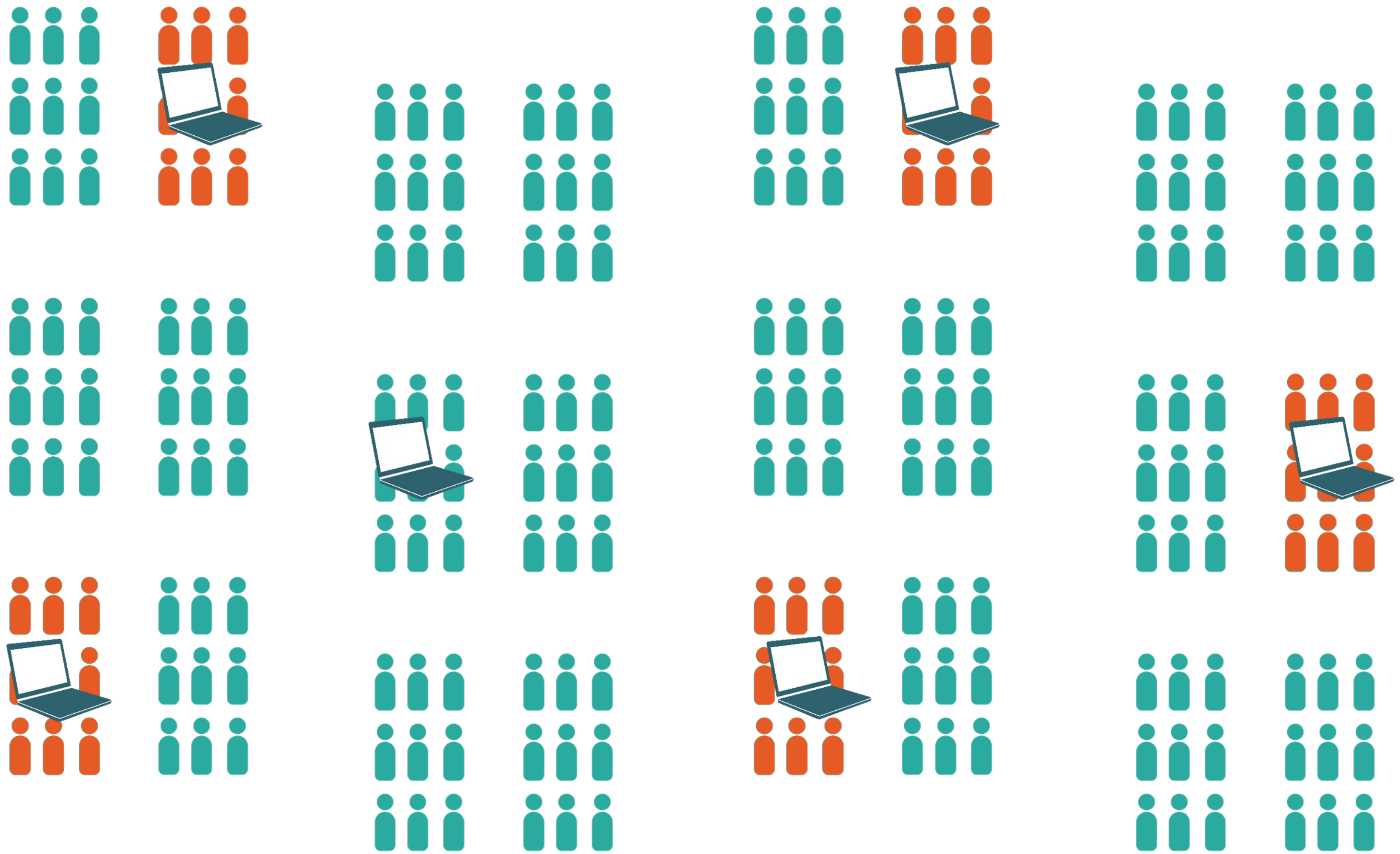
RCT Design: Phase-in

Phase 0: No One Treated Yet

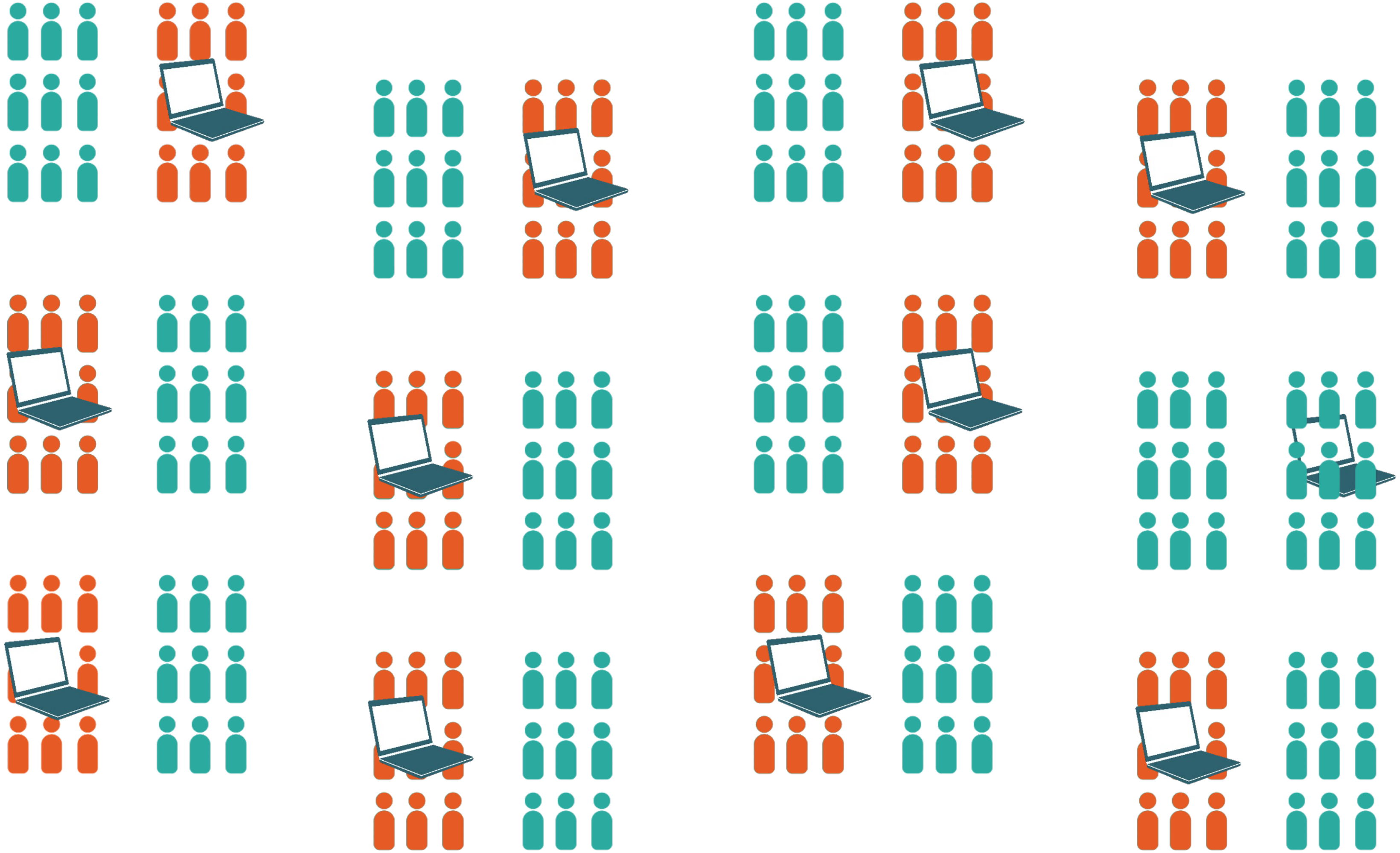
All Control



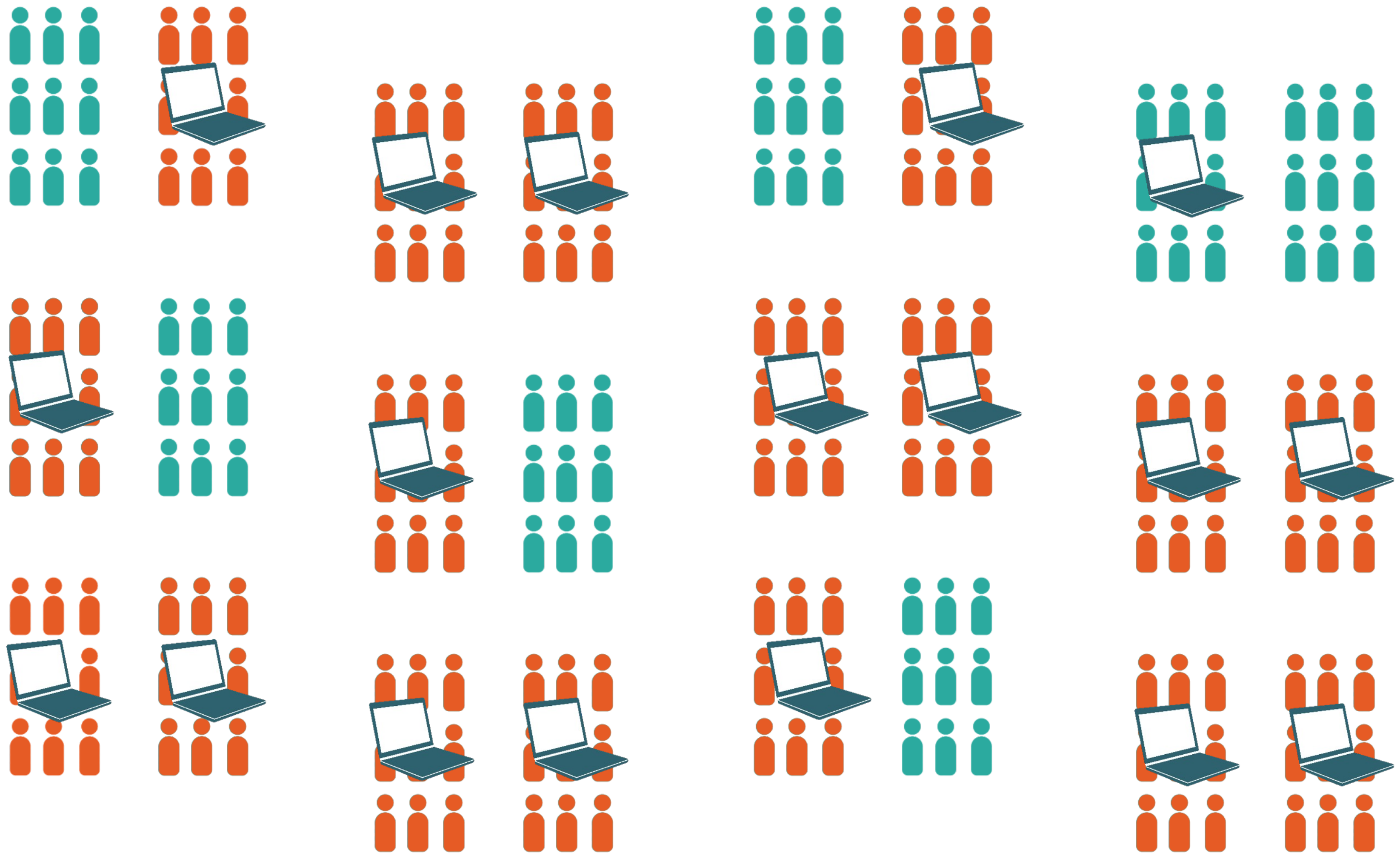
Phase 1: 25% Treated
75% Control



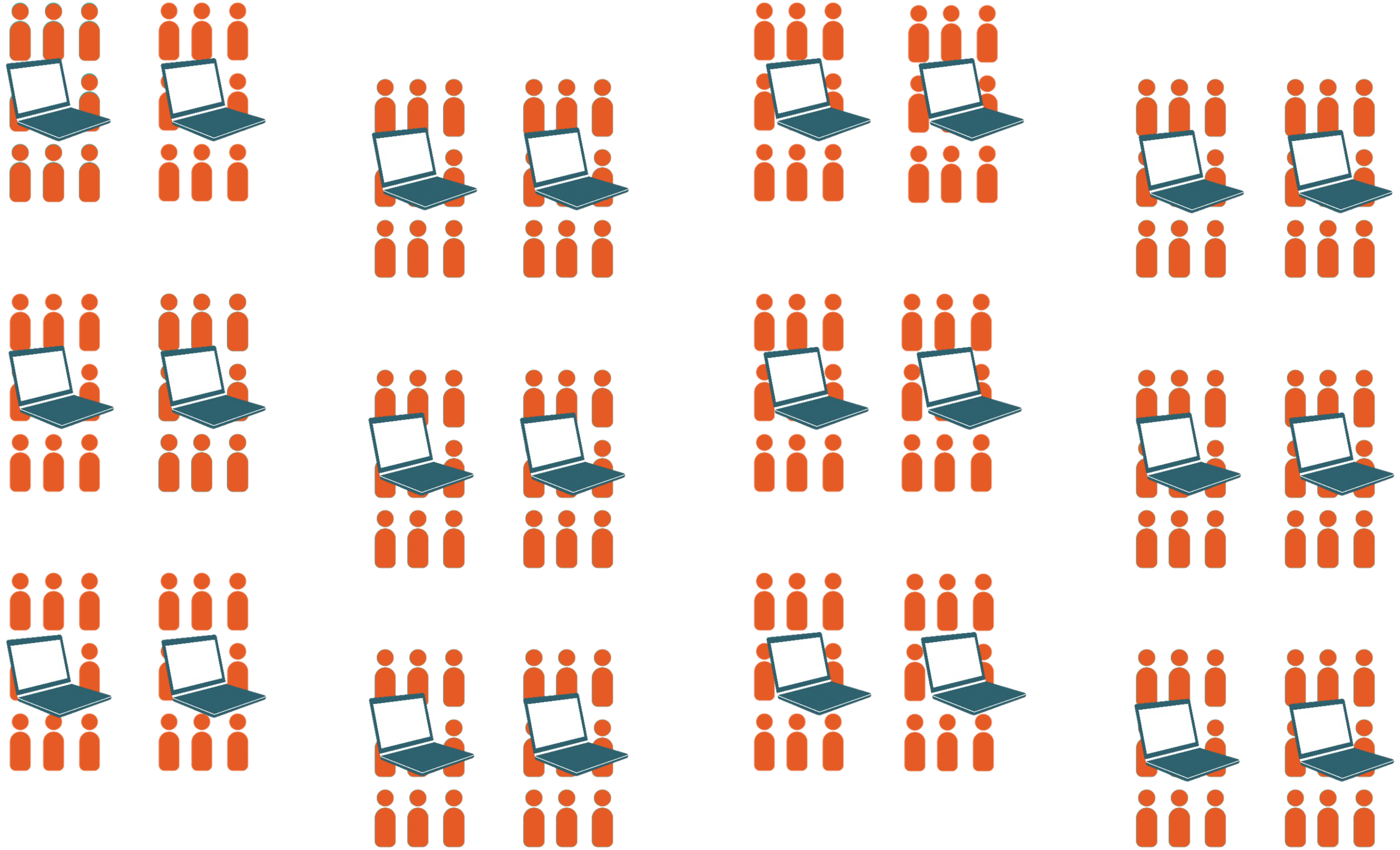
Phase 2: Half Treated
Half Control



Phase 3: 75% Treated
25% Control



Phase 4: All Treated
No Control (Experiment Over)



Program Specification:

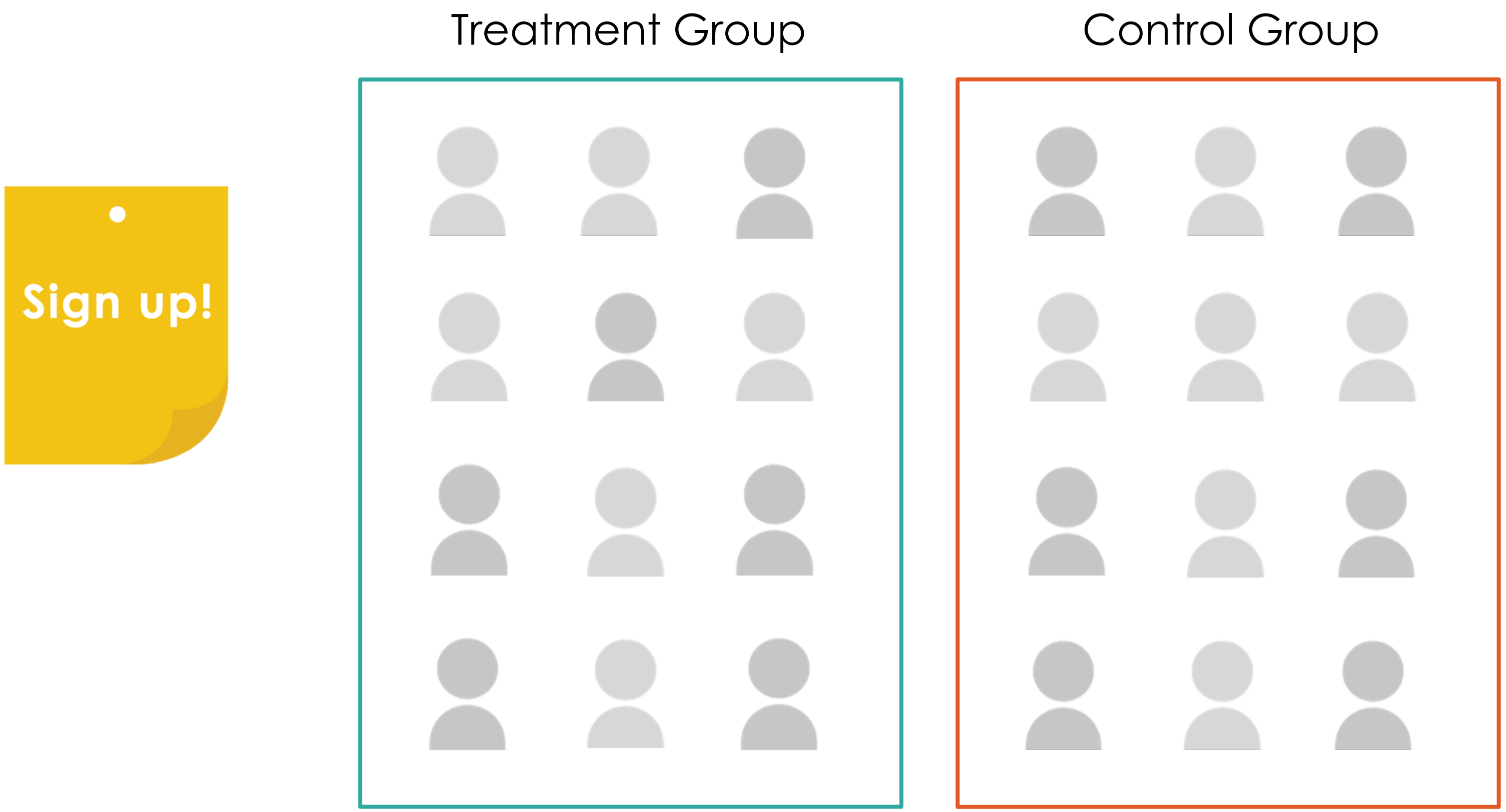
Program is an Entitlement

Program is an Entitlement: Cannot Mandate nor Deny Intervention

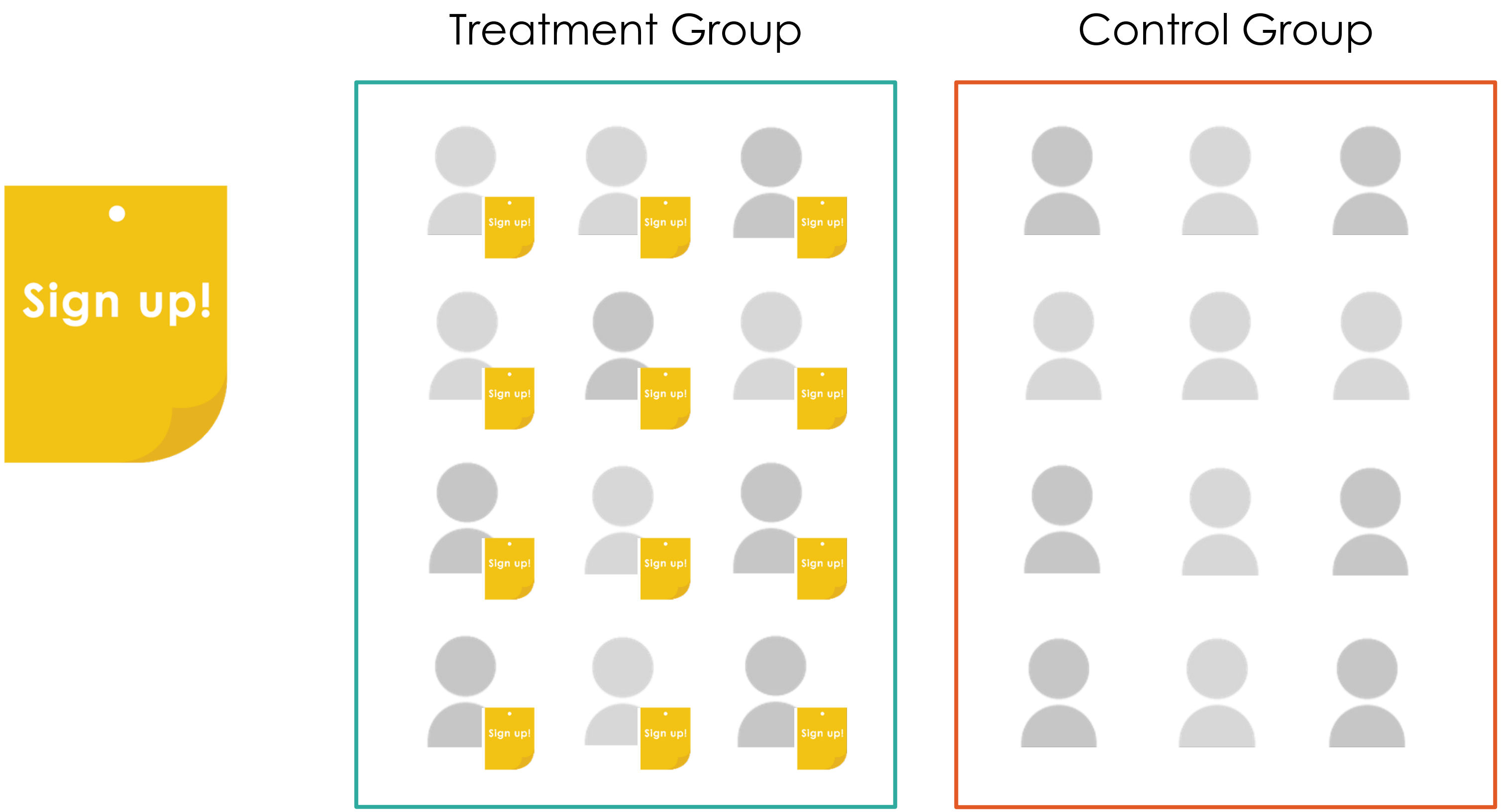
SNAP

Supplemental
Nutrition
Assistance
Program

RCT Design: Encouragement

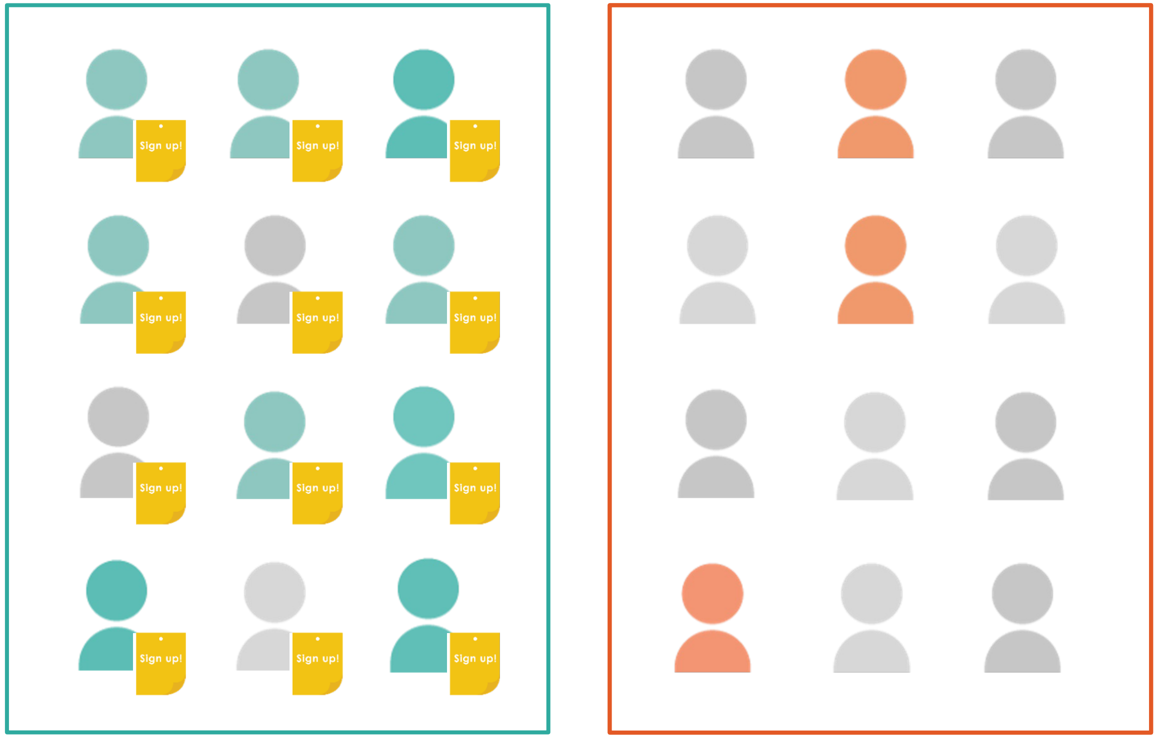


RCT Design: Encouragement



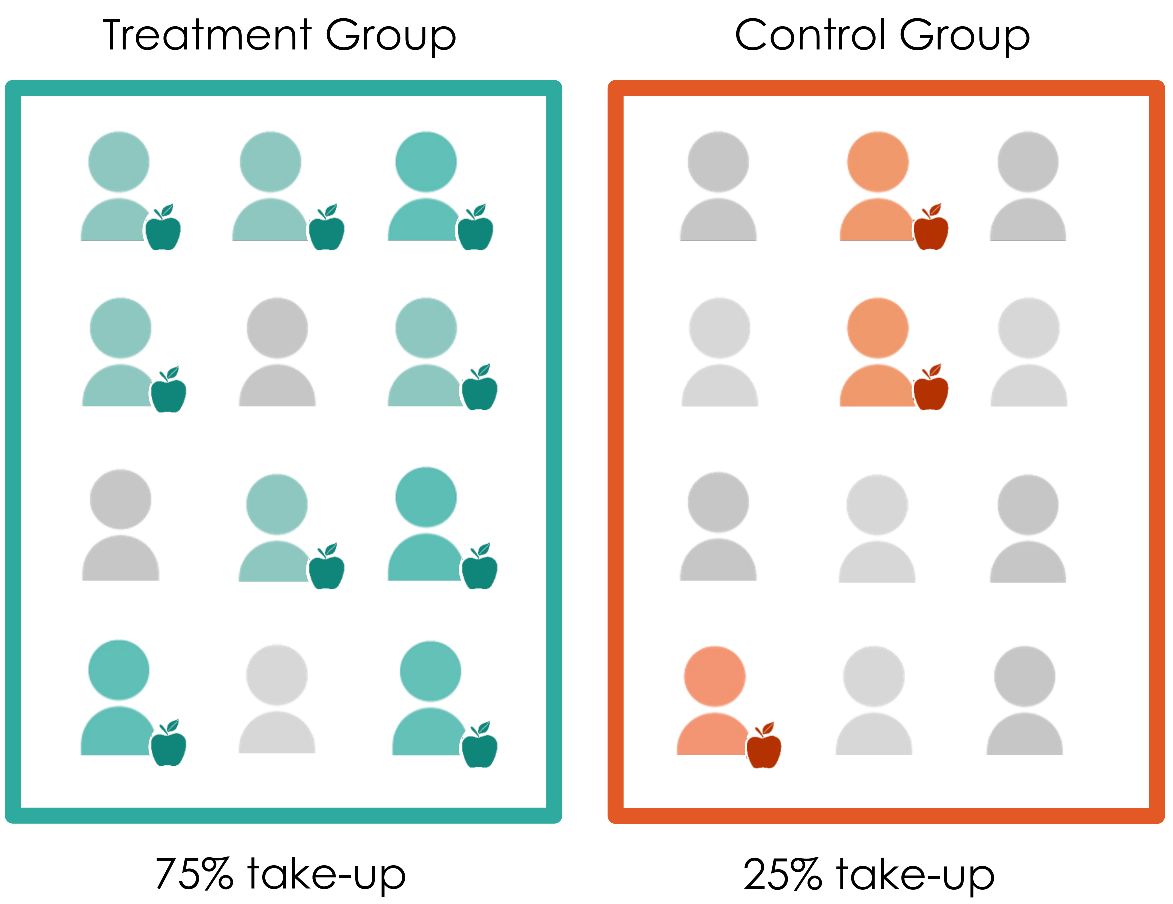
RCT Design: Encouragement

Treatment Group | Control Group



75% take-up | 25% take-up

RCT Design: Encouragement



Compared

Entire Treatment Group to Entire Control Group

Evaluation: https://www.povertyactionlab.org/evaluation/snap-take-evaluation

Appendix

How to randomize

ESP WORKBOOK

Population of interest:

|Type answers here.

Method of sampling and/or enrolling participants (or groups of participants):

|Type answers here.

Number of units that will be included in the evaluation (if known):

|Type answers here.

Randomization process

(e.g. randomize at the point of service, or randomize from a complete list of participants)

|Type answers here.

Treatment group 1

Intervention that treatment group 1 receives:

|Type answers here.

Number of units in treatment group 1 (if known):

|Type answers here.

Treatment group 2
(if more than one treatment arm)

Intervention that treatment group 2 receives:

|Type answers here.

Number of units in treatment group 2 (if known):

|Type answers here.

Control group

Services ("business as usual") that the control group receives, if any:

|Type answers here.

Number of units in the control group (if known):

|Type answers here.

Which variables to use for stratification

- Variables that are **discrete**
  - Practical constraint (can't create "buckets" out of continuous variables)
- Variables you believe might be **correlated with the outcome of interest** or possible shocks
  - For example, female headed households have different access to inputs
- Variables for which you are interested in **heterogeneous (differential) treatment effects**

Unbalanced groups

What do you do if you conduct a balance test and find that some variables are unbalanced?

- Consider the number of variables that are unbalanced (remember that if you are testing at a 5% significance level, then you would expect to see differences between the groups roughly 5% of the time)
- What is the magnitude of difference? (a large magnitude of difference is more concerning than a small one)
- Which variables are unbalanced? (Imbalances can be especially problematic if the variables are correlated with take-up of treatment, attrition, or the outcome variables)

Recap: Randomization designs

Design	Most useful when:	Advantages	Disadvantages
Basic lottery	Program is oversubscribed/not enough resources to deliver to all eligible Resources are constant for the evaluation period	Familiar and well understood Easy to implement Often seen as fair Transparent process	Control group may be less interested in answering questions/being involved, which could create problems for the analysis (differential attrition)

Recap: Randomization designs

Design	Most useful when:	Advantages	Disadvantages
Randomized phase-in/roll out	Resources to administer program expands over time Everyone must receive treatment eventually Logistically difficult to roll out to everyone at one time	Easy to understand Constraint is easy to explain Control group likely to comply because they expect to benefit later	Anticipation of receiving the program may impact short-run behavior for the control group Only possible for outcomes that can be measured before everyone is treated Difficult to measure long-term impact

Recap: Randomization designs

Design	Most useful when:	Advantages	Disadvantages
Encouragement designs	Program must be open to everyone in target population Take-up is low in the relevant population	Allows evaluation of programs that cannot exclude anyone Can randomize at individual level even if the program is administered at a group level	Measures impact of those who respond to the encouragement (not the general population) Need large enough inducement to improve take-up Encouragement itself may have direct effect on outcomes

Recap: Randomization designs

Design	Most useful when:	Advantages	Disadvantages
Expand eligibility and randomize near an eligibility cutoff	There are enough funds to expand the number of people served by an intervention The intervention is essential for certain people who are already being served	Everyone previously eligible for a program remains eligible	Does not measure the intervention's effect on people who were already eligible. Only can measure effects for people who are near the eligibility cutoff

Adaptive experiments

- A more flexible form of assignment
- Use information on treatment outcomes from earlier waves of the study to place more participants over time in the treatment conditions that are working the best
  - Example: algorithm placed participants in job assistance interventions that were showing to be most effective
- Enables quicker response to research findings

Resource:

- Adaptive experiments for policy research (VoxDev)

J-PAL studies:

- Adaptive Trial to Identify Maximally Effective COVID Information Campaign
- An Adaptive Targeted Field Experiment: Job Search Assistance for Refugees in Jordan

Additional J-PAL resources

- Randomization (research resource)
- Real-world challenges to randomization and their solutions
- Sample size and calculating statistical power (research resource)
- Rules of thumb for statistical power

Reuse and citation

To reference this lecture, please cite as:

J-PAL. “Lecture: How to Randomize.” Abdul Latif Jameel Poverty Action Lab. 2023. Cambridge, MA