

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. Randomized Evaluation from Start to Finish
7. Threats & Analysis
8. Ethical Considerations
9. Generalizing & Applying Evidence

Learning Objectives

- Develop a deeper understanding of how randomization works 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 Process & Design
- Balance & Stratification
- The Unit of Randomization
- Designing Randomized Evaluations for Different Program Specifications

Lecture Overview

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



Random Sampling and Random Assignment

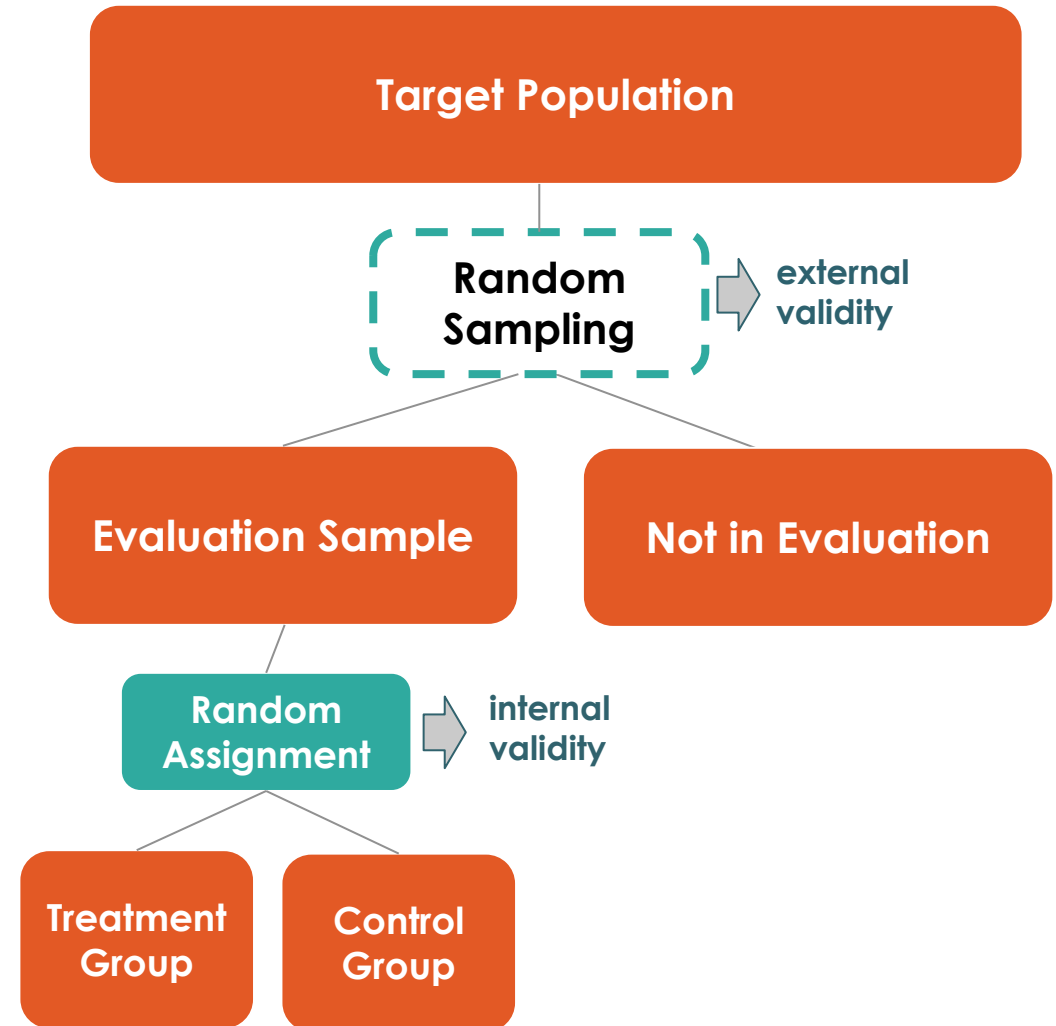
Defining terms

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.

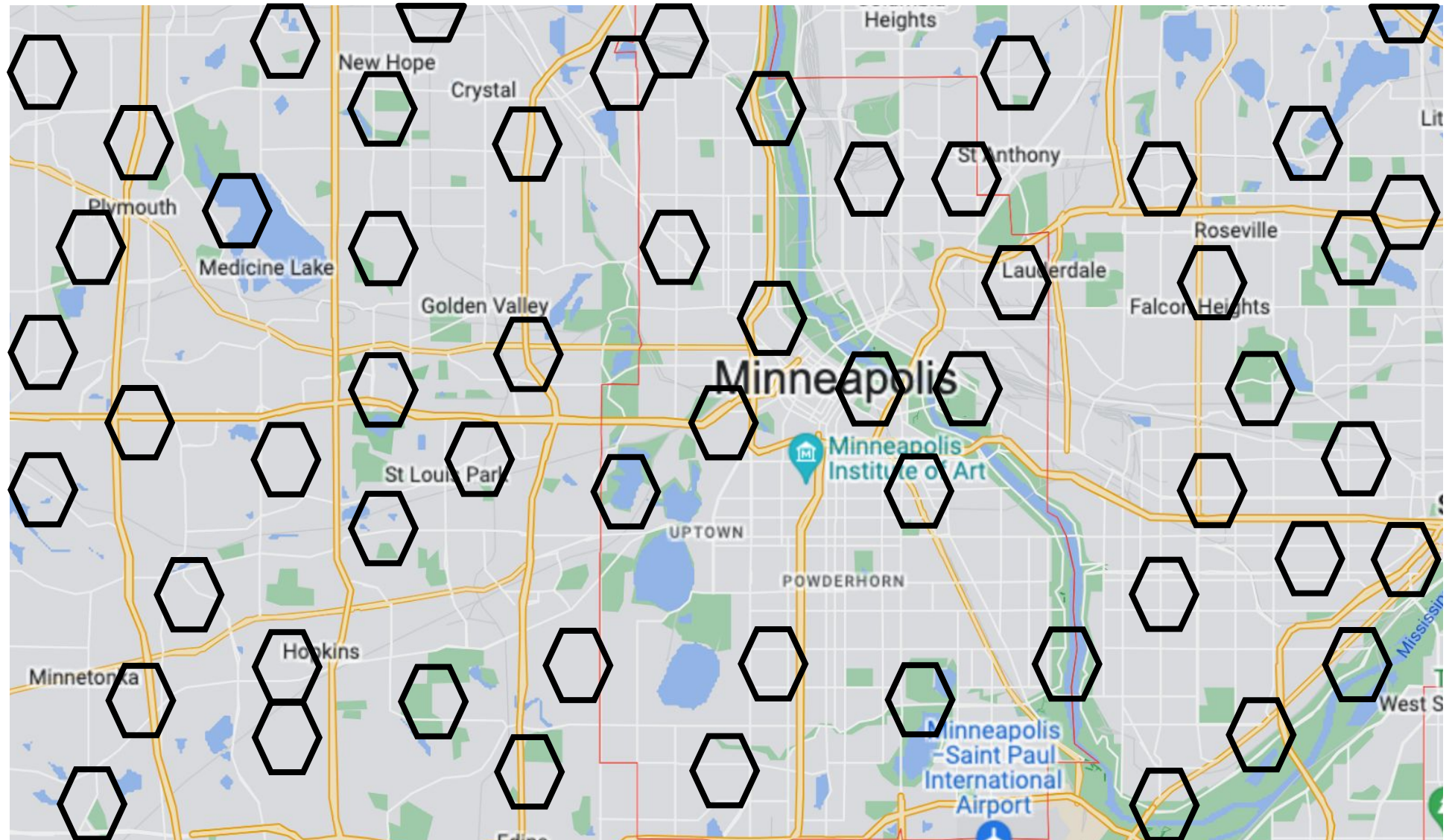
Random assignment: Taking a pool of eligible units and then allocating those units to treatment and control groups by means of a random process.

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



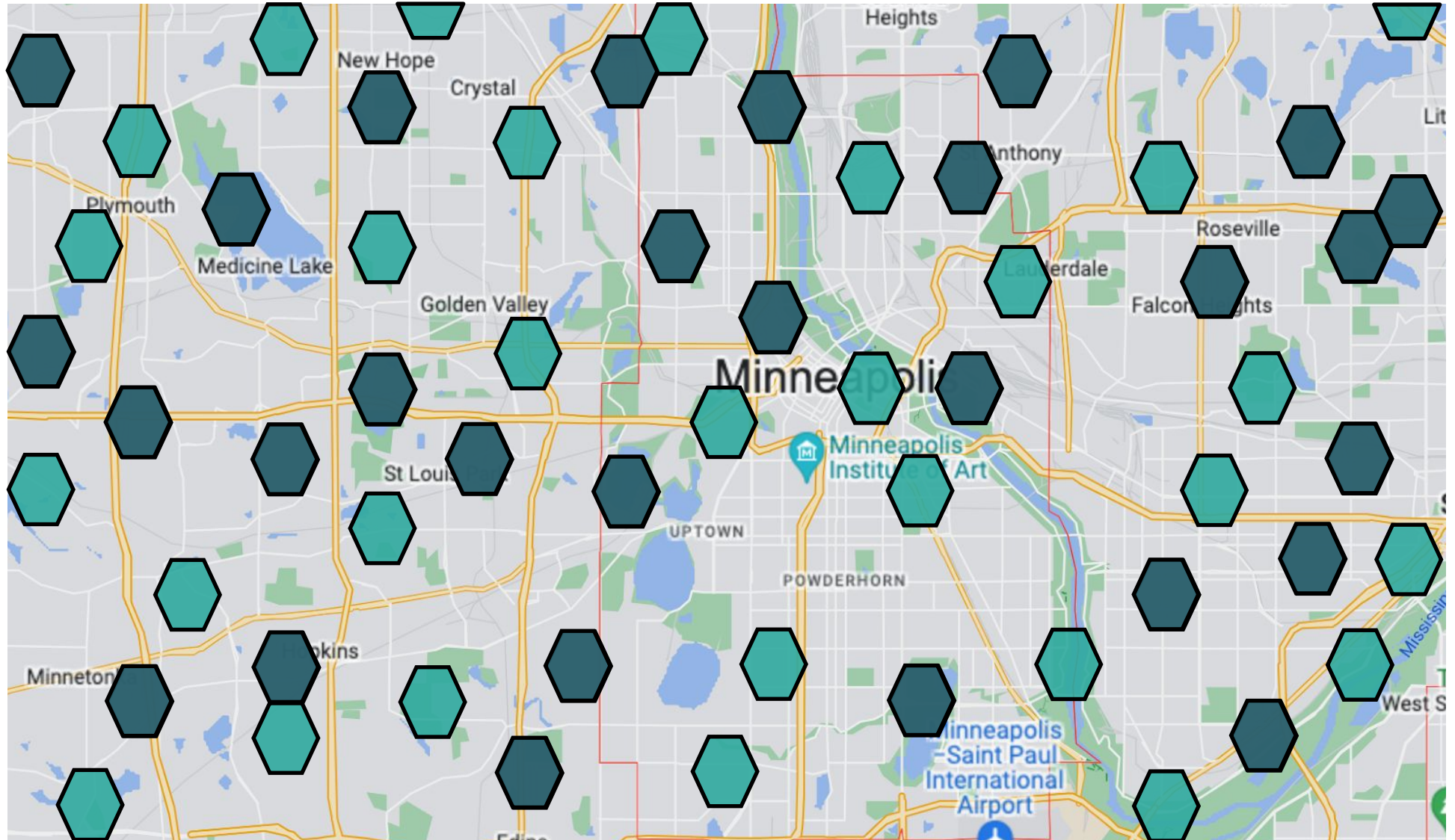
Random Sampling

Randomly
sample
from area of
interest

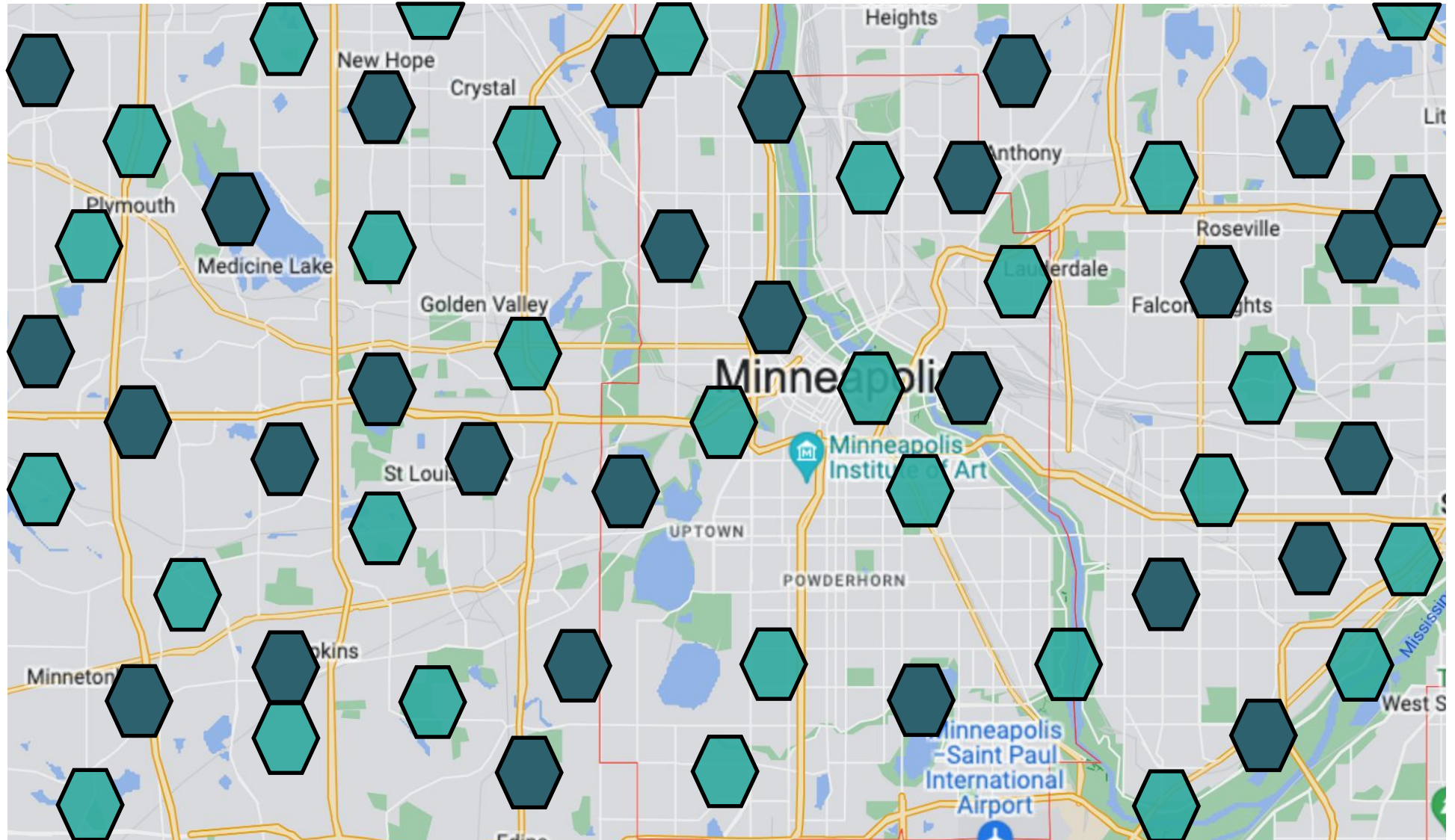
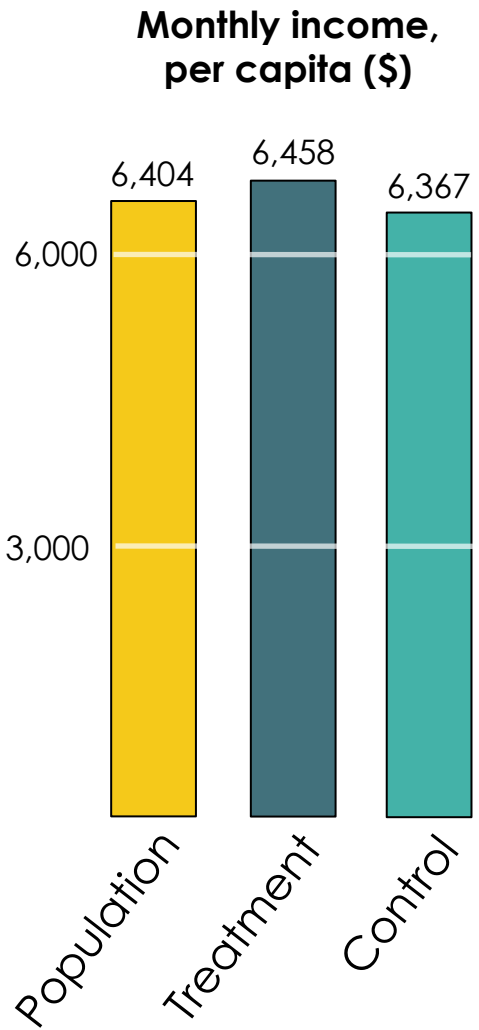


Random Assignment

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

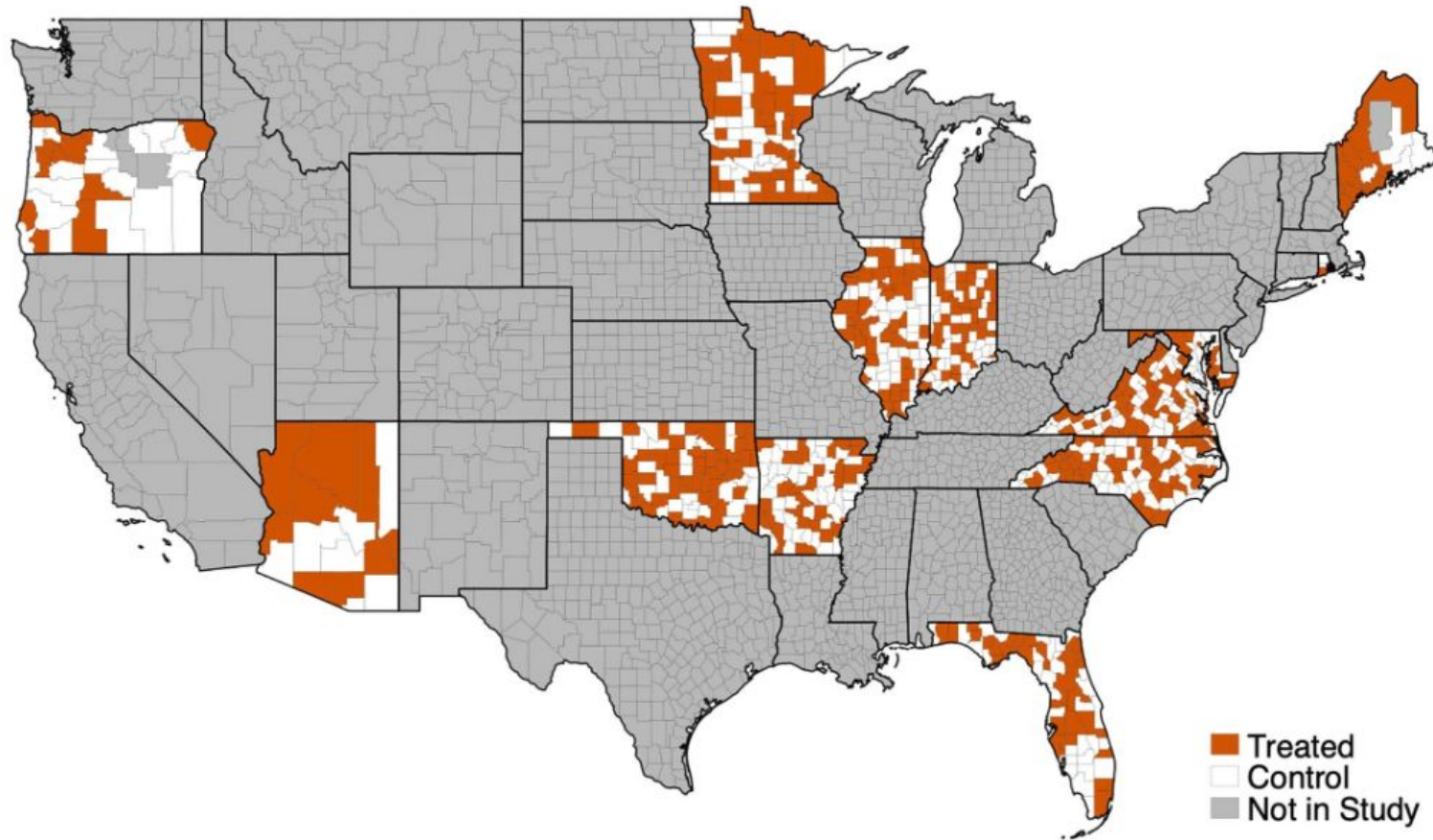


Random Assignment



A real-world example

Randomization of the program across zip codes in 13 states



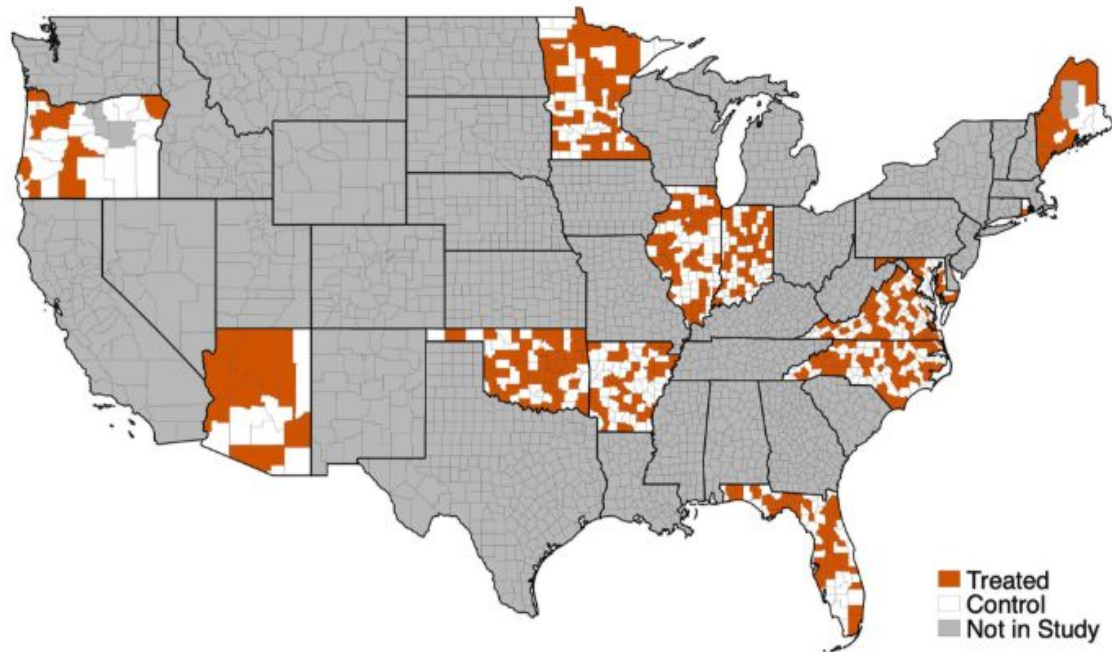
Randomized counties (Thanksgiving campaign).

Breza, Emily, Fatima Cody Stanford, Marcella Alsan, Burak Alsan, Abhijit Banerjee, Arun G. Chandrasekhar, Sarah Eichmeyer et al. "Effects of a large-scale social media advertising campaign on holiday travel and COVID-19 infections: a cluster randomized controlled trial." *Nature medicine* 27, no. 9 (2021): 1622-1628.

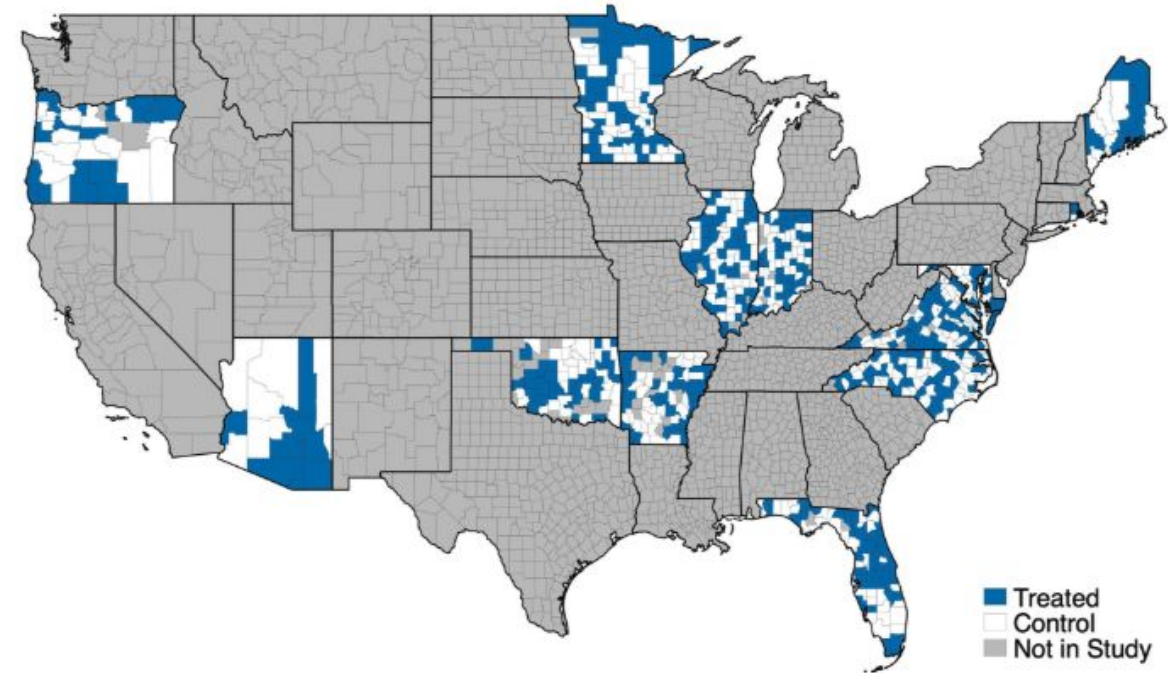
Effects of large-scale social media campaign on holiday travel and Covid-19

- **Context & intervention:** Before peak holiday travel in 2020, we shared brief videos over social media of medical providers encouraging people to stay home for the holidays, with the goal of reducing the spread of Covid-19
- **Sample selection**
 - Unit of randomization: US zip codes
 - Sample frame: All US states reporting weekly Covid-19 case count data at the zip code level

Randomization before Thanksgiving and before Christmas



Extended Data Fig. 1 | Randomized counties (Thanksgiving campaign).



Extended Data Fig. 2 | Randomized counties (Christmas campaign).

Breza, Emily, Fatima Cody Stanford, Marcella Alsan, Burak Alsan, Abhijit Banerjee, Arun G. Chandrasekhar, Sarah Eichmeyer et al. "Effects of a large-scale social media advertising campaign on holiday travel and COVID-19 infections: a cluster randomized controlled trial." *Nature medicine* 27, no. 9 (2021): 1622-1628.

Lecture Overview

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

Basic Randomization Process

- Randomization can be appropriate when a program is oversubscribed and resource constraints prevent everyone who is eligible from receiving the treatment
- A “**lottery**” can be easy to understand and implement
 - Randomly assign treatment and control from sample of those who are eligible (example: some housing authorities run lotteries for Housing Choice Vouchers)

“Rolling Randomization” Process

Randomize each participant upon entry into study/at the point of service

- Set the probability of assignment to treatment group to a fixed percentage (e.g., 50%, 75%)

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



Varying Intensity of Treatment

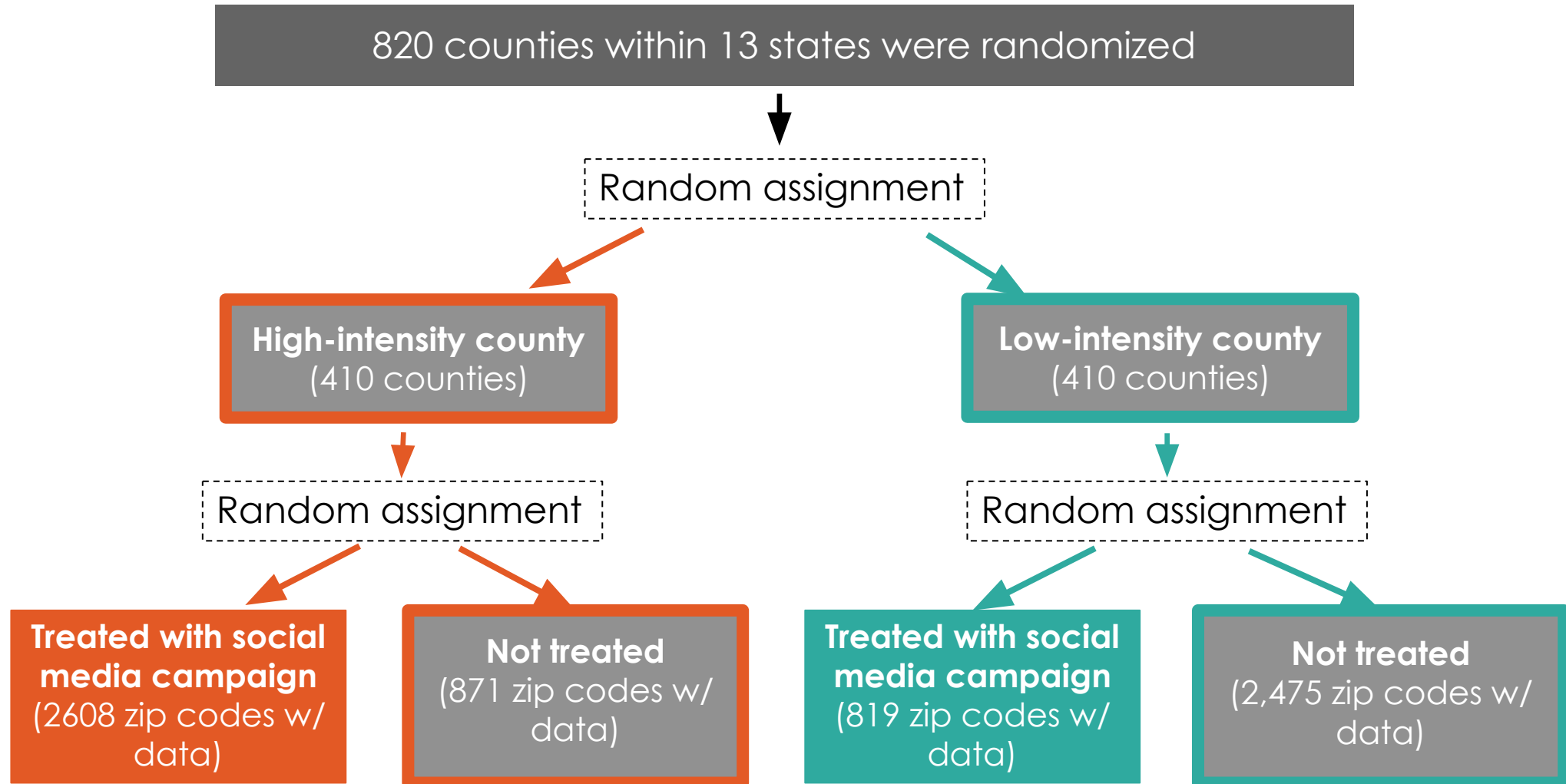
This type of research design can help answer the question: “What is the effect of varying the intensity of the treatment?”

At the individual level (i.e., varying the dosage each individual gets)

At the group/cluster level (i.e., varying the proportion of treated people within an area)



Randomization (Thanksgiving campaign)



Breza, Emily, Fatima Cody Stanford, Marcella Alsan, Burak Alsan, Abhijit Banerjee, Arun G. Chandrasekhar, Sarah Eichmeyer et al. "Effects of a large-scale social media advertising campaign on holiday travel and COVID-19 infections: a cluster randomized controlled trial." *Nature medicine* 27, no. 9 (2021): 1622-1628.

Multiple treatments

How do different treatments compare?



How do different treatments interact?

		Text message	
		Yes	No
Video	Yes	Group 1 Text message + video	Group 2 Video
	No	Group 3 Text message	Group 4 Control

Impact of Tailored Public Health Messages for Black and Latinx Communities on Knowledge and Behavior

Context: Black and Latinx communities disproportionately affected by Covid-19 due to systemic racism and public health messaging that may not be framed to address these communities' concerns

Intervention: Three videos about generic health topics or Covid-19

- Delivered by a physician of the same or different race/ethnicity of the viewer
- With tailored acknowledgement of injustice or with standard statement

Effects of Tailored Messaging for Black and Latinx Communities on Covid-19 Knowledge and Behaviors

		Racial concordance	
		Yes	No
Tailored acknowledgement	Yes	Group 1 Racial concordance + tailored acknowledgement	Group 2 Tailored acknowledgement
	No	Group 3 Racial concordance	Group 4 Control

Lecture Overview

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

Balance & Stratification

Balance = the treatment and control groups are comparable on certain key characteristics

- Can check using **balance tests**. If you find that some variables are unbalanced, consider the number of imbalances and their magnitude.

Stratification = dividing the sample into **subgroups** (also known as **strata**) that share certain characteristics and then randomizing within each subsample

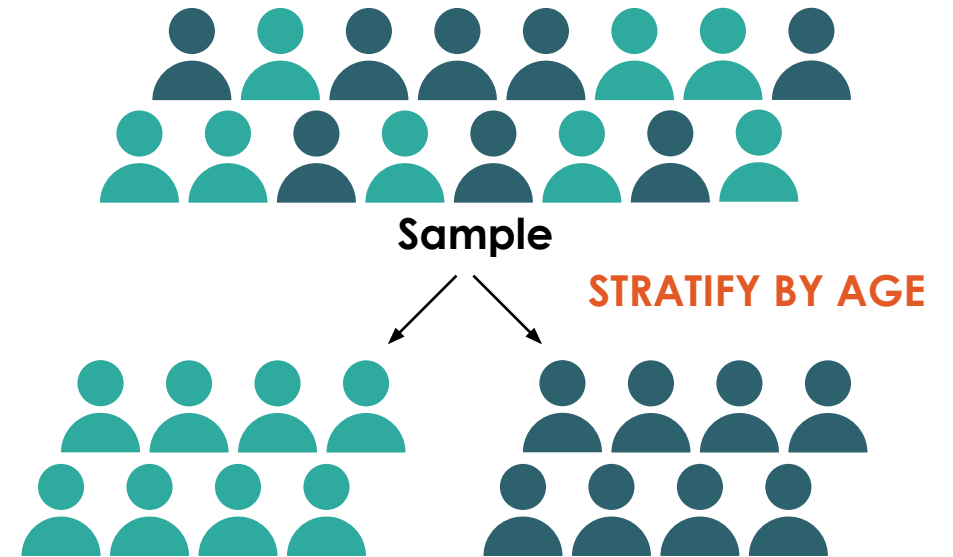
Stratification

Steps:

1. Form groups of similar units
2. Randomize within the group

Example:

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



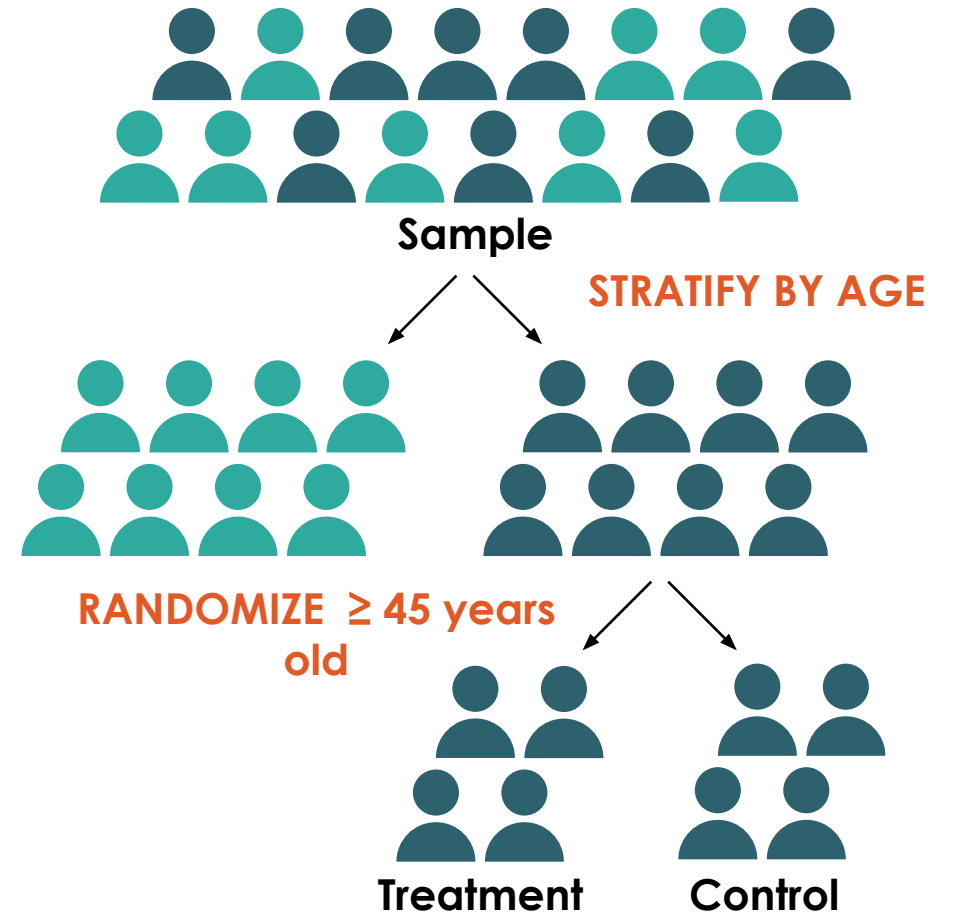
Stratification

Steps:

1. Form groups of similar units
2. Randomize *within* the group

Example:

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



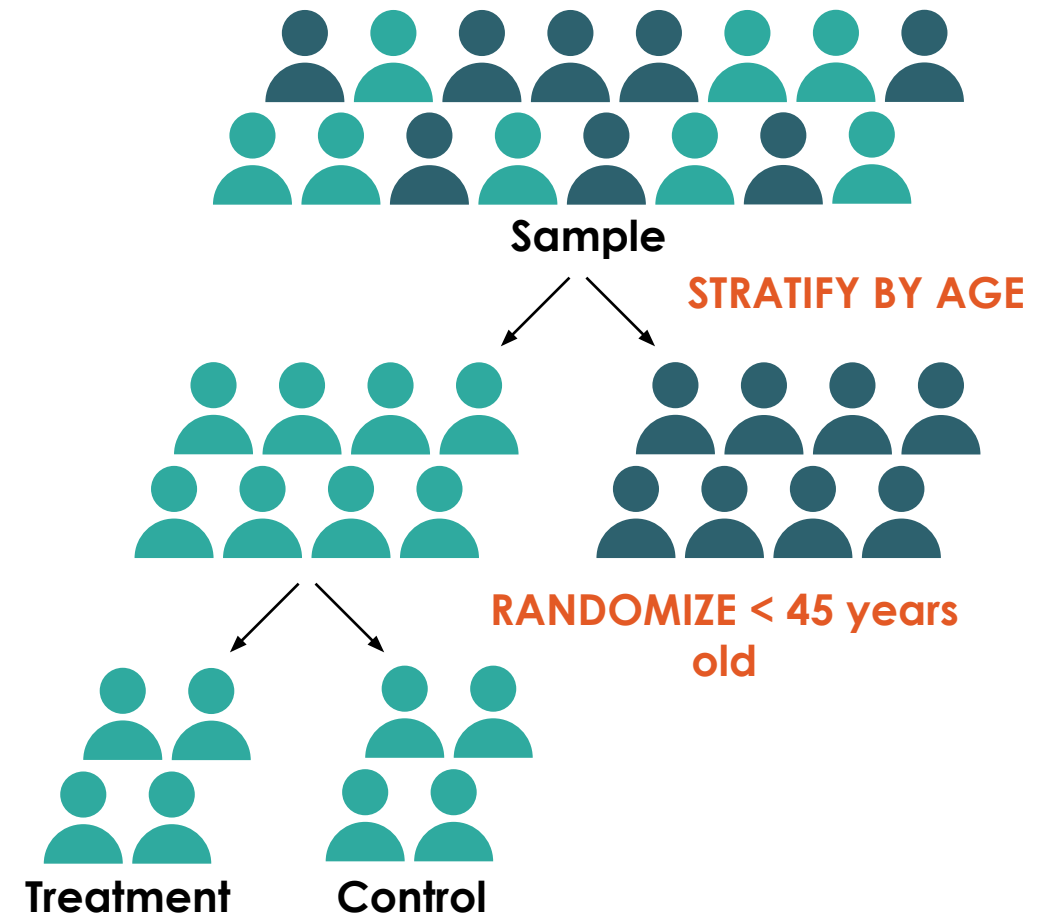
Stratification

Steps:

1. Form groups of similar units
2. Randomize *within* the group

Example:

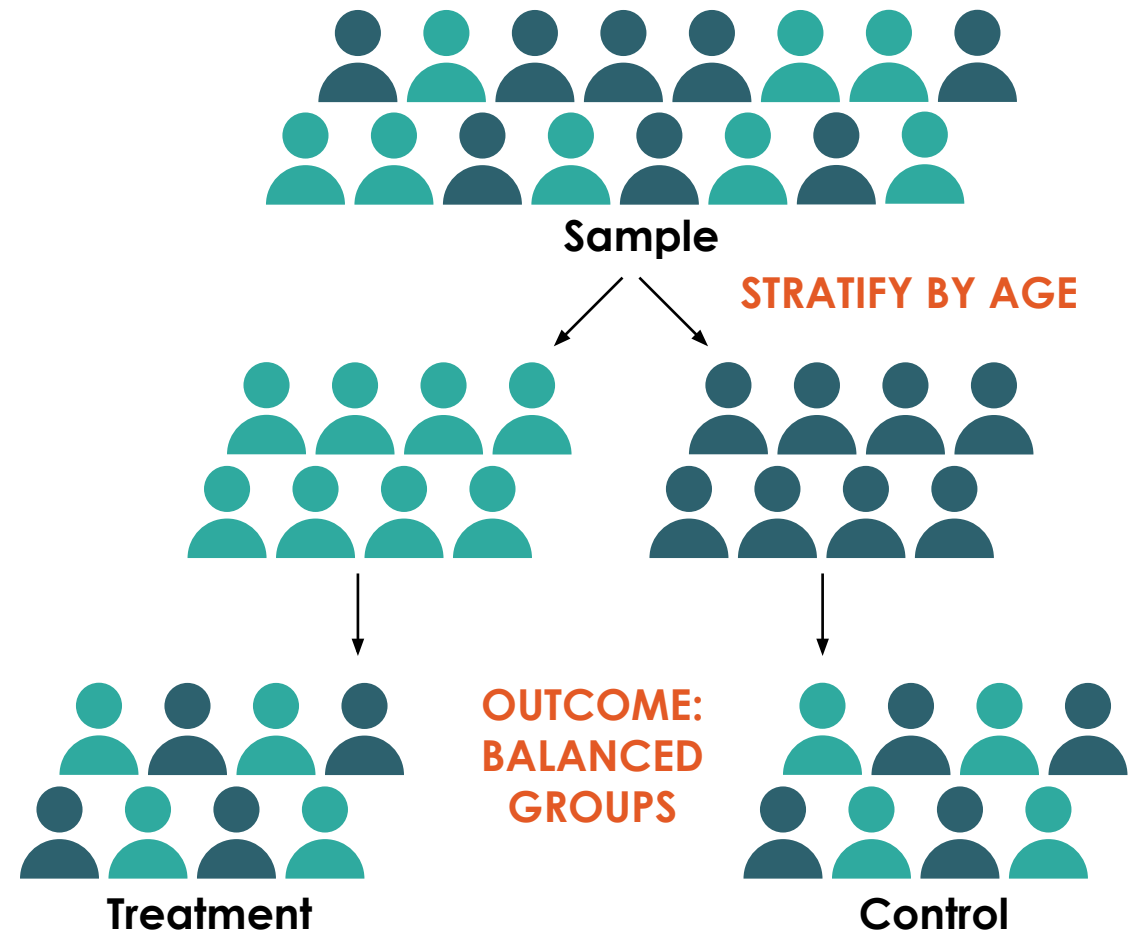
1. Divide sample into age groups
2. Randomize individuals separately within each education level



Stratification

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

But only for this one observable characteristic!



Why stratify?

1. To **achieve balance**
2. To **increase statistical power** (we will return to this tomorrow)
3. To analyze treatment effect by group (also known as **subgroup analysis** or heterogeneous treatment effects)
 - For example, differential effects by race or ethnicity, rural/urban settings, income level, etc.

During random sampling, in some cases, we may want to use stratification to ensure certain groups are overrepresented (**disproportionate stratification**)

Impact of Tailored Public Health Messages for Black and Latinx Communities on Knowledge and Behavior

Intervention: Three videos about generic health topics or Covid-19, varied by racial concordance and messaging

Sample and stratification:

- **Eligible sample:** 18 years or older, self-identifying as Black or Latinx, oversampled with a secondary education or less
- **Stratified** by sex, age (45 years and older), and location

Results: Physician messaging campaign increased Covid-19 knowledge and self-reported protective behaviors. Racial concordance or tailored messaging did not further increase effectiveness of messaging.

Baseline characteristics/balance check

Appendix Table. Summary of Participant Characteristics*

Variable	Study Participants			Control			Any Intervention		
	All	African American	Latinx	All	African American	Latinx	All	African American	Latinx
Measured at baseline									
Observations, <i>n</i>	14 144	8767	5377	1617	1042	575	12 527	7725	4802
Gender, <i>n</i> (%)									
Female	8112 (57.4)	5277 (60.2)	2835 (52.7)	943 (58.3)	636 (61.0)	307 (53.4)	7169 (57.2)	4641 (60.1)	2528 (52.6)
Male	6032 (42.6)	3490 (39.8)	2542 (47.3)	674 (41.7)	406 (39.0)	268 (46.6)	5358 (42.8)	3084 (39.9)	2274 (47.4)
Mean age (SD), <i>y</i>	38.93 (15.3)	39.63 (15.7)	37.78 (14.4)	39.16 (15.6)	39.78 (16.1)	38.03 (14.7)	38.9 (15.2)	39.61 (15.7)	37.76 (14.4)
Stratum									
Female, age ≥45 y, coastal	1318 (9.3)	879 (10.0)	439 (8.2)	164 (10.1)	111 (10.7)	53 (9.2)	1154 (9.2)	768 (9.9)	386 (8.0)
Female, age ≥45 y, central	2013 (14.2)	380 (15.7)	633 (11.8)	229 (14.2)	162 (15.5)	67 (11.7)	1784 (14.2)	1218 (15.8)	566 (11.8)
Female, age ≤44 y, coastal	2010 (14.2)	1245 (14.2)	765 (14.2)	223 (13.8)	140 (13.4)	83 (14.4)	1787 (14.3)	1105 (14.3)	682 (14.2)
Female, age ≤44 y, central	2771 (19.6)	1773 (20.2)	998 (18.6)	327 (20.2)	223 (21.4)	104 (18.1)	2444 (19.5)	1550 (20.1)	894 (18.6)
Male, age ≥45 y, coastal	820 (5.8)	521 (5.9)	299 (5.6)	96 (5.9)	64 (6.1)	32 (5.6)	724 (5.8)	457 (5.9)	267 (5.6)
Male, age ≥45 y, central	996 (7.0)	641 (7.3)	355 (6.6)	111 (6.9)	75 (7.2)	36 (6.3)	885 (7.1)	566 (7.3)	319 (6.6)
Male, age ≤44 y, coastal	1939 (13.7)	1013 (11.6)	926 (17.2)	199 (12.3)	113 (10.8)	86 (15.0)	1740 (13.9)	900 (11.7)	840 (17.5)
Male, age ≤44 y, central	2277 (16.1)	1315 (15.0)	962 (17.9)	268 (16.6)	154 (14.8)	114 (19.8)	2009 (16.0)	1161 (15.0)	848 (17.7)
Mean household size (SD), <i>n</i>	3.22 (1.7)	3.04 (1.7)	3.5 (1.7)	3.23 (1.7)	3 (1.7)	3.64 (1.7)	3.21 (1.7)	3.05 (1.7)	3.48 (1.7)
Mean education index (SD)†	4.56 (1.5)	4.69 (1.5)	4.35 (1.4)	4.53 (1.5)	4.67 (1.5)	4.27 (1.4)	4.57 (1.5)	4.69 (1.5)	4.36 (1.4)

Alsan, Marcella, Fatima Cody Stanford, Abhijit Banerjee, Emily Breza, Arun G. Chandrasekhar, Sarah Eichmeyer, Paul Goldsmith-Pinkham et al. "Comparison of knowledge and information-seeking behavior after general COVID-19 public health messages and messages tailored for Black and Latinx communities: a randomized controlled trial." *Annals of internal medicine* 174, no. 4 (2021): 484-492.

What subgroup analyses are you interested in?
What variables might you want to stratify by?

Lecture Overview

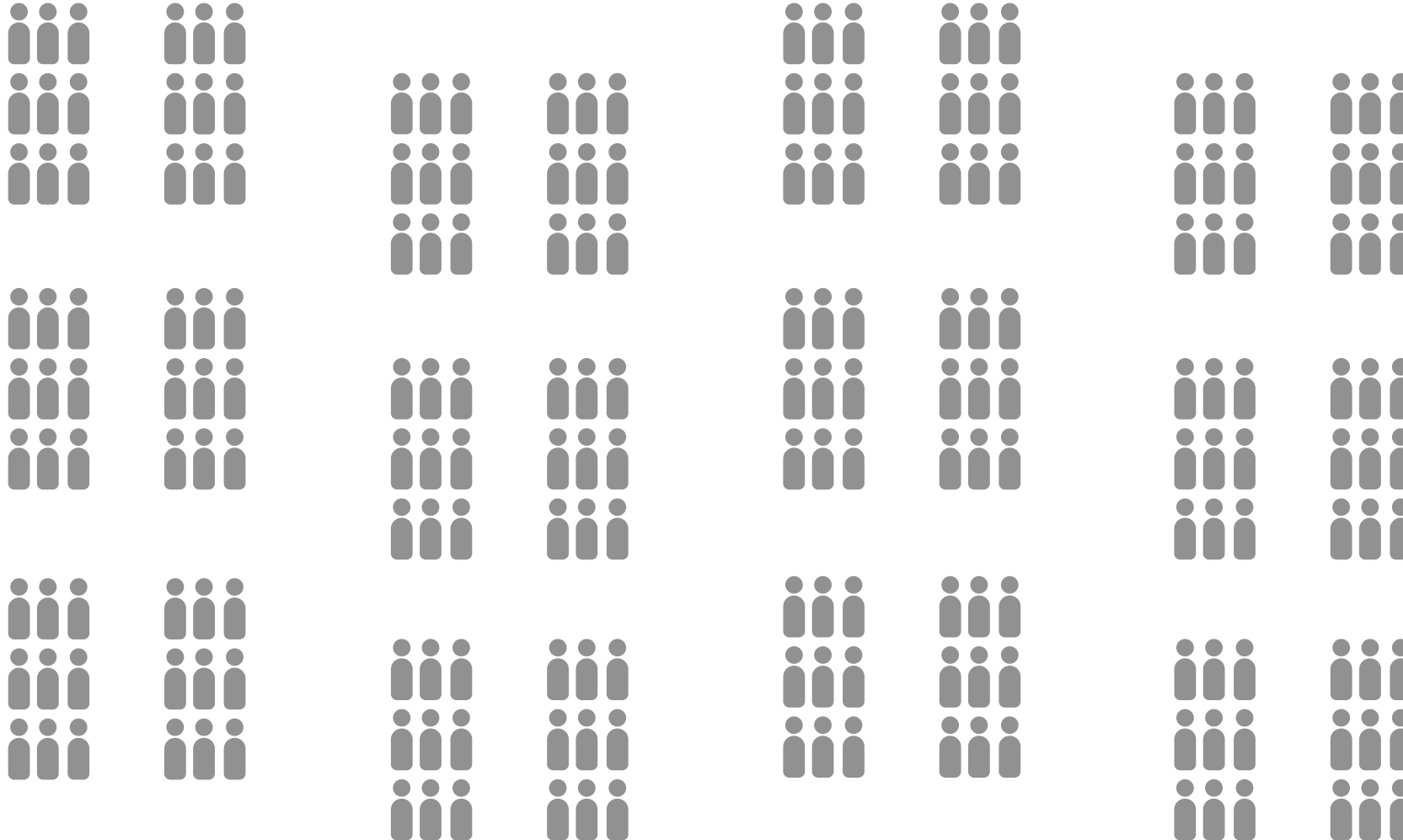
- What is Randomization?
- Randomization Process & Design
- Balance & Stratification
- The Unit of Randomization
- Designing Randomized Evaluations for Different Program Specifications

Units of Observation and the Level of Randomization

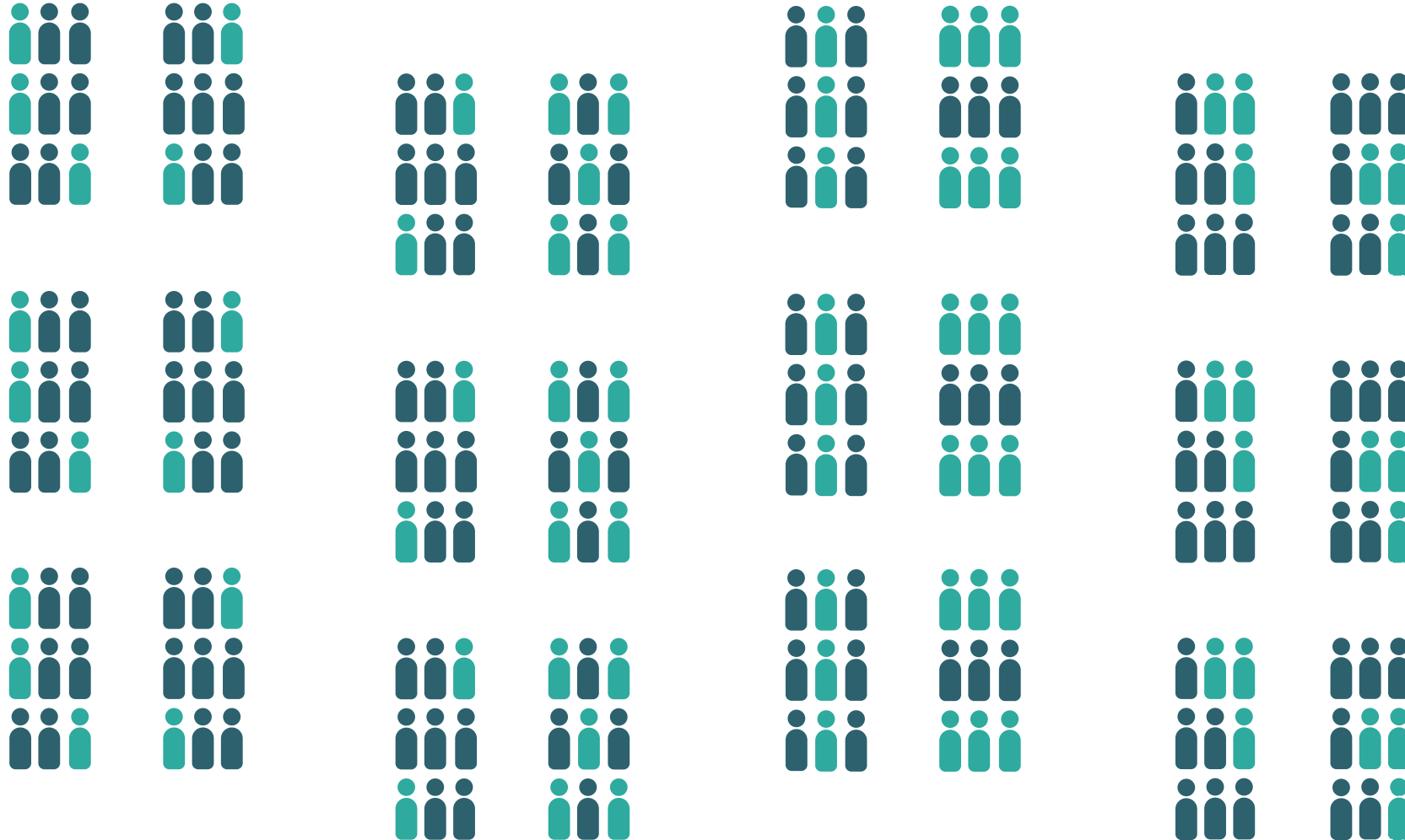
At which level should a study randomize?

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

Unit of Randomization: Student?



Unit of Randomization: Student?



Unit of Randomization: Classroom

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



Unit of Randomization: Classroom

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



Unit of Randomization: School

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



Unit of Randomization: School

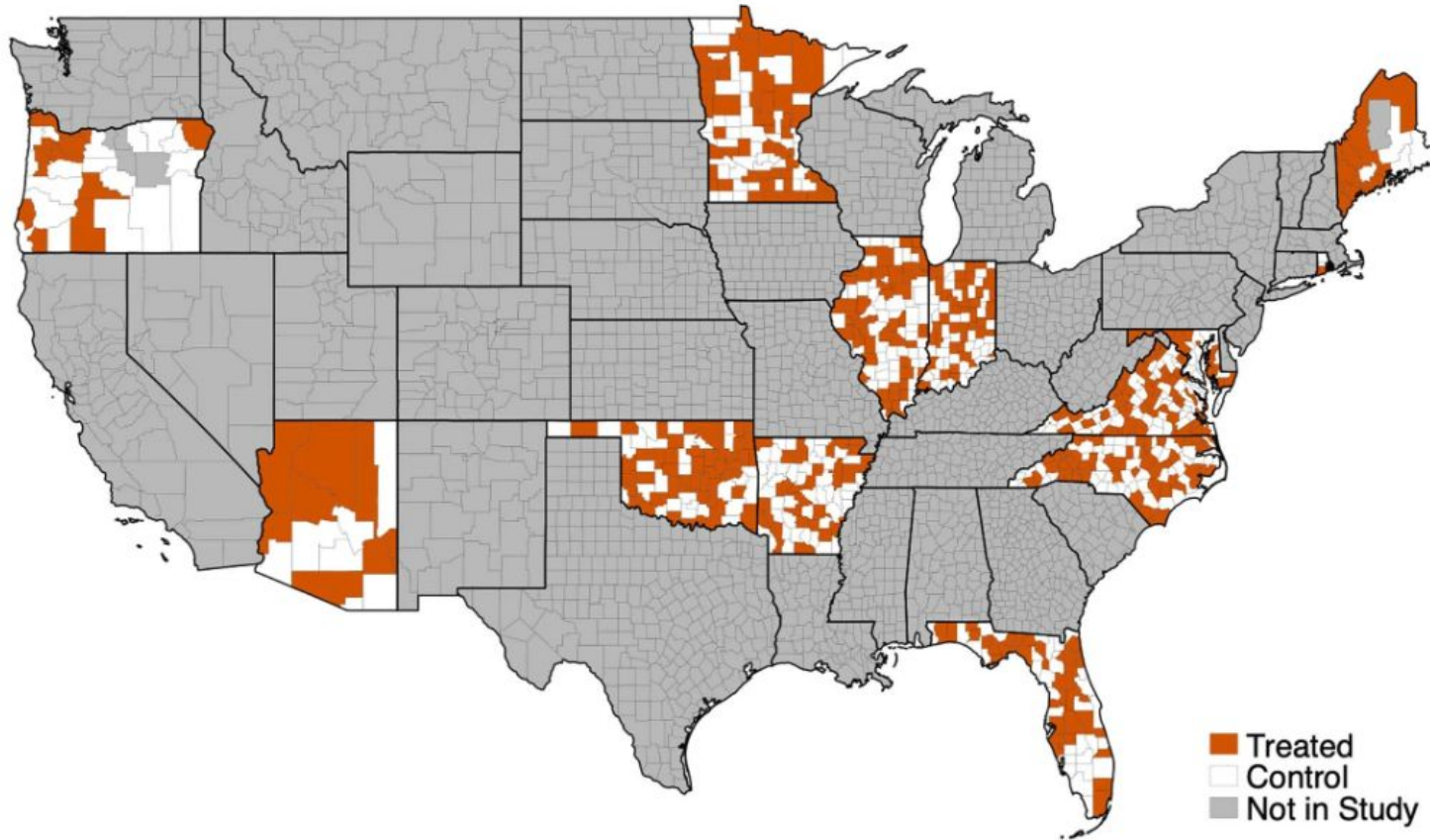
We call groups of units “clusters”
Randomization at the group level:
“cluster randomized trial”



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?

Returning to Covid-19 social media campaign example



Randomized counties (Thanksgiving campaign).

Breza, Emily, Fatima Cody Stanford, Marcella Alsan, Burak Alsan, Abhijit Banerjee, Arun G. Chandrasekhar, Sarah Eichmeyer et al. "Effects of a large-scale social media advertising campaign on holiday travel and COVID-19 infections: a cluster randomized controlled trial." *Nature medicine* 27, no. 9 (2021): 1622-1628.

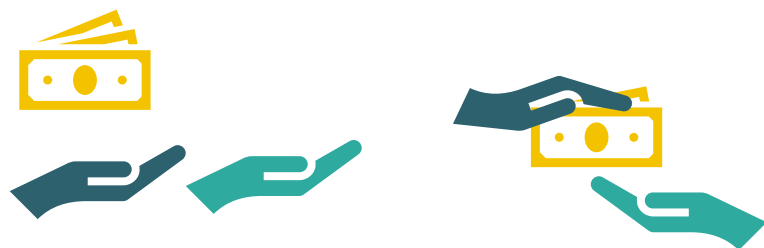
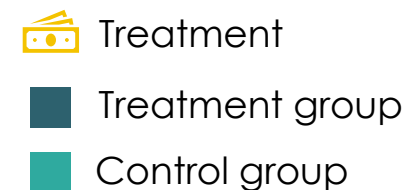
What are some considerations and constraints for the level of randomization in your contexts?

Level of Randomization: Considerations

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



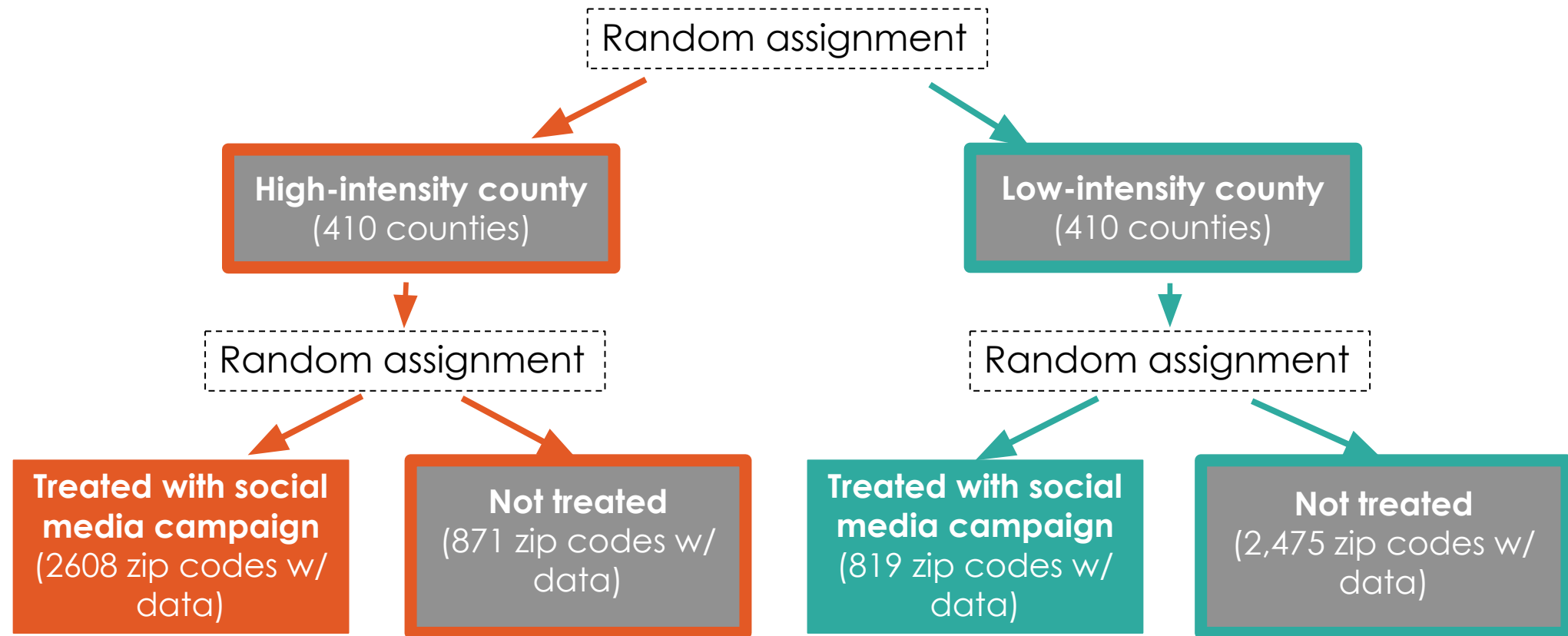
Noncompliance: When participants do not follow (“comply with”) their treatment assignment. For example, if participants assigned to the control group join the treatment group, or vice versa.



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

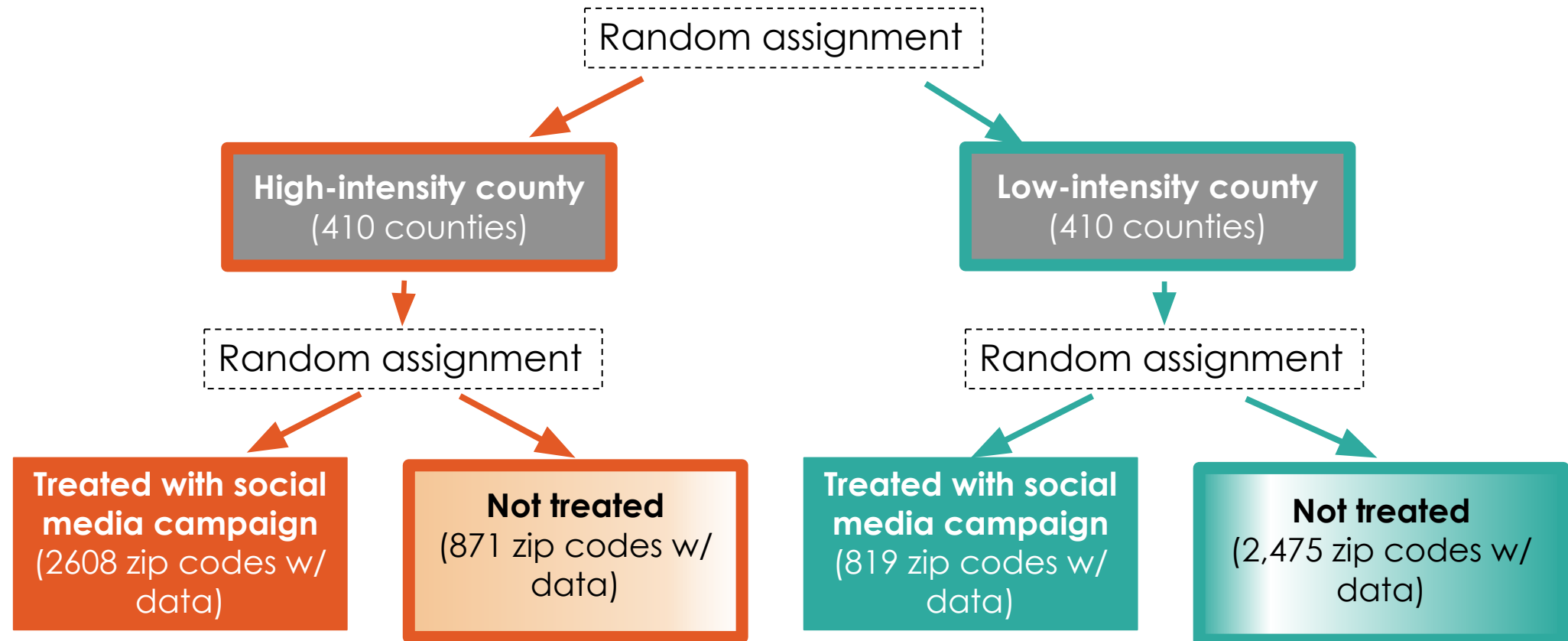
Measuring Spillovers in Covid-19 Social Media Campaign

- Investigating spillover in high-intensity zip codes to untreated zip codes vs. spillover in low-intensity zip codes to untreated zip codes



Measuring Spillovers in Covid-19 Social Media Campaign

- Investigating spillover in high-intensity zip codes to untreated zip codes vs. spillover in low-intensity zip codes to untreated zip codes



Lecture Overview

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

Program Specifications

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

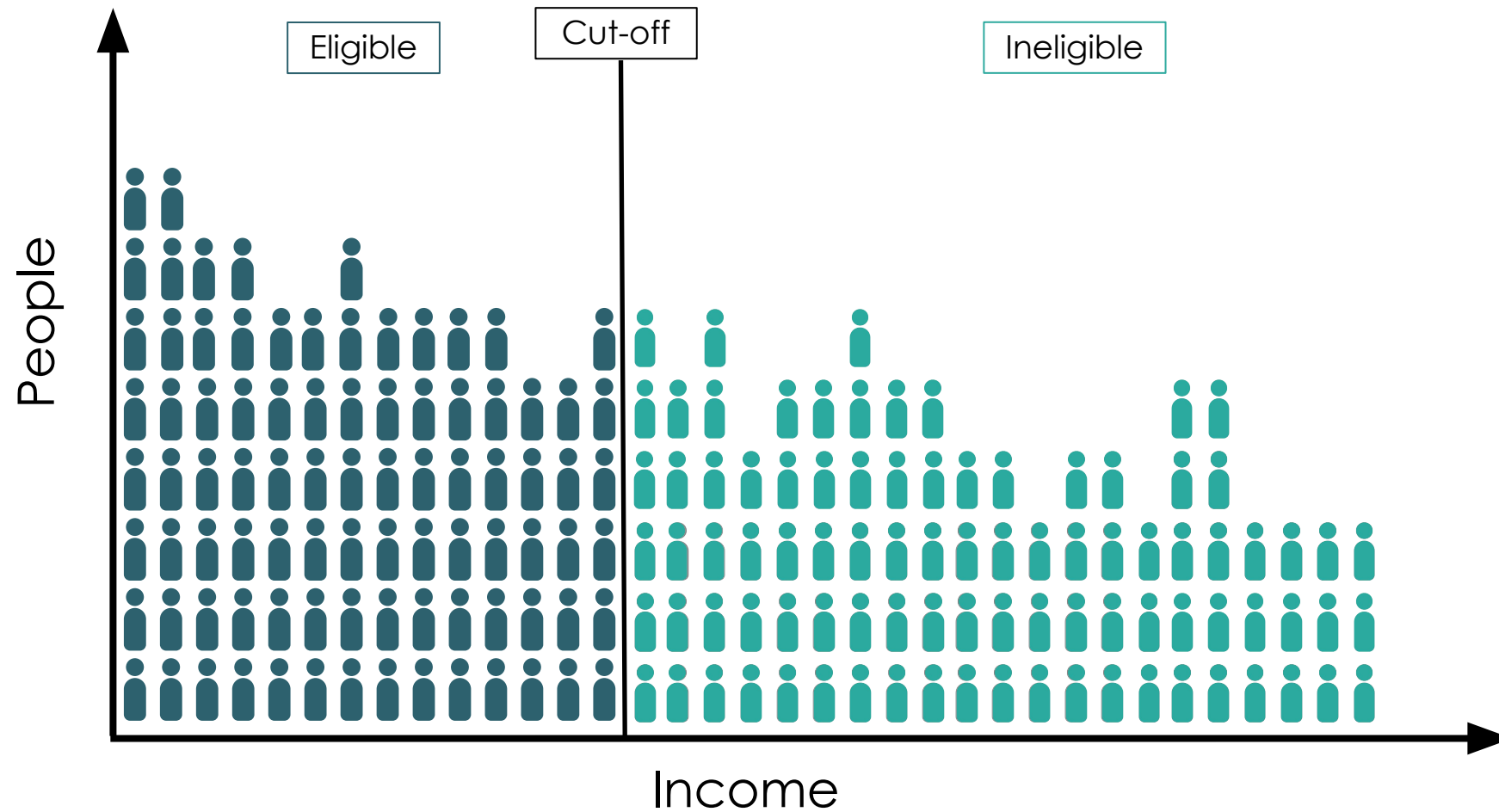
- Eligibility is determined by a cutoff (e.g., income, credit score)
- There are enough resources to extend the program to everyone in the study area
- The intervention is a current entitlement program, and participation cannot and/or should not be withheld from anyone who is eligible

Program Specification:

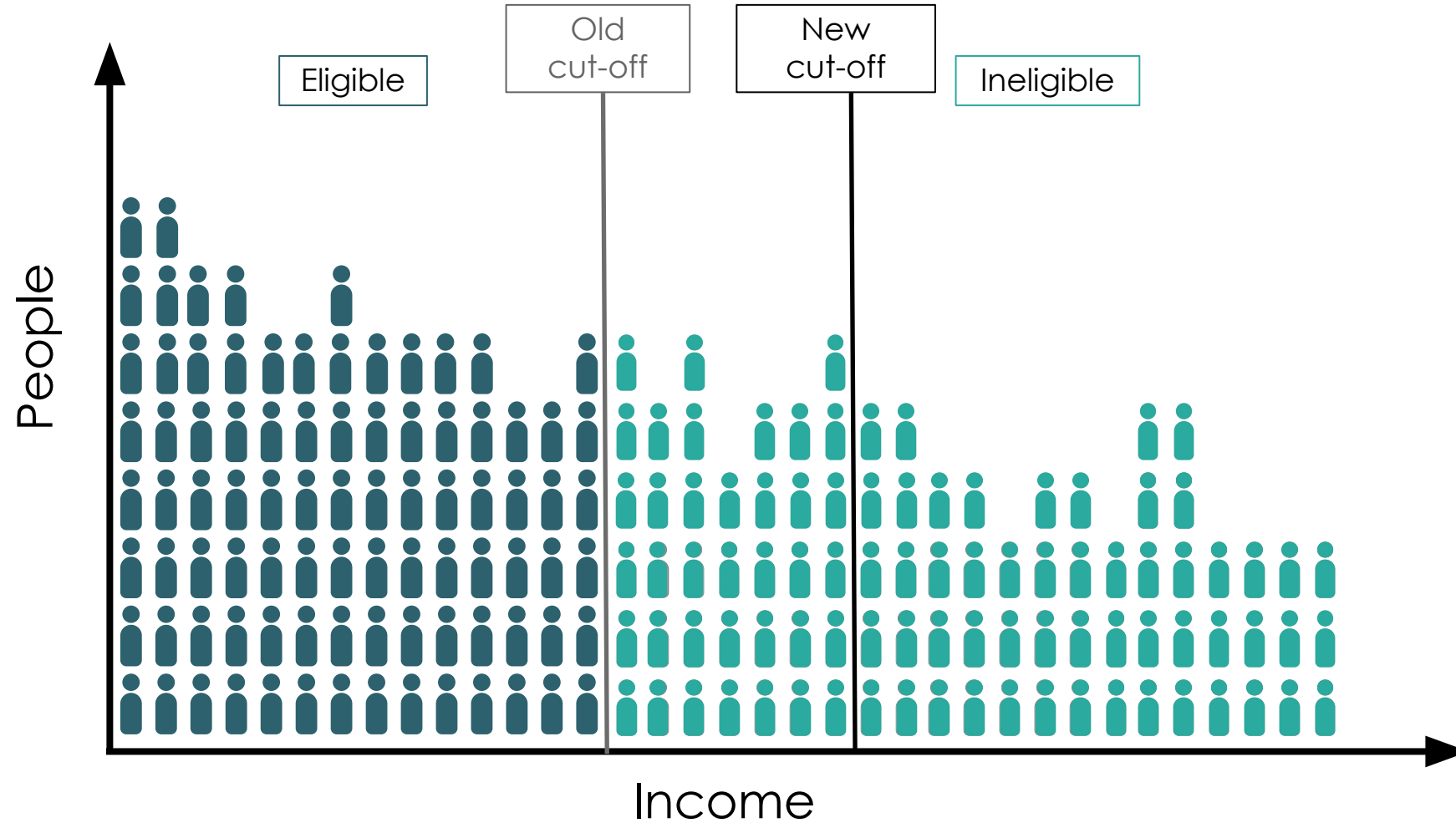
Program has Strict Eligibility Criteria



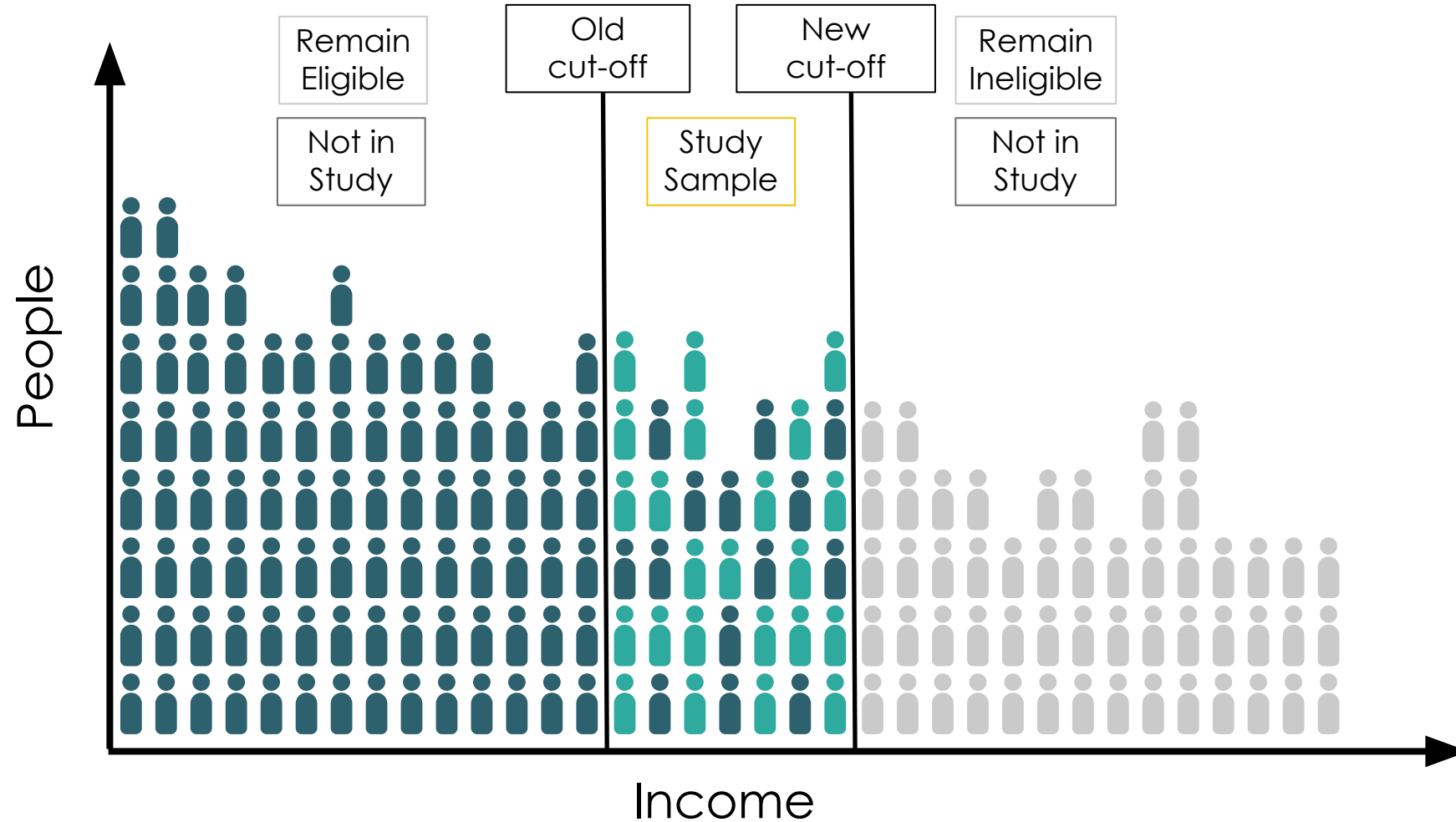
Program with an Eligibility Cut-off



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



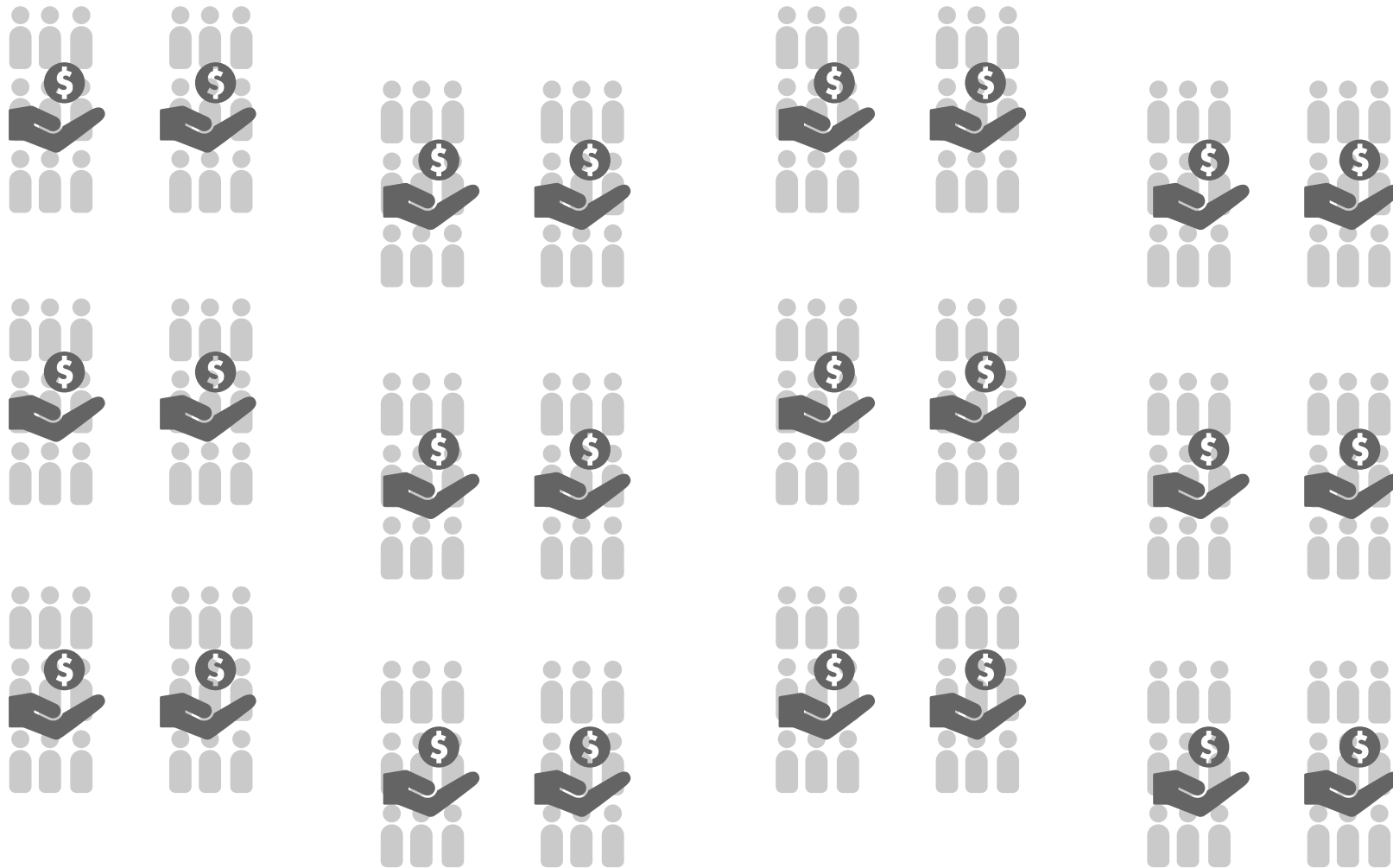
If it is not possible to randomize access to a program because you **cannot withhold treatment from the control group**, you might be able to conduct a randomized controlled trial by:

- 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:

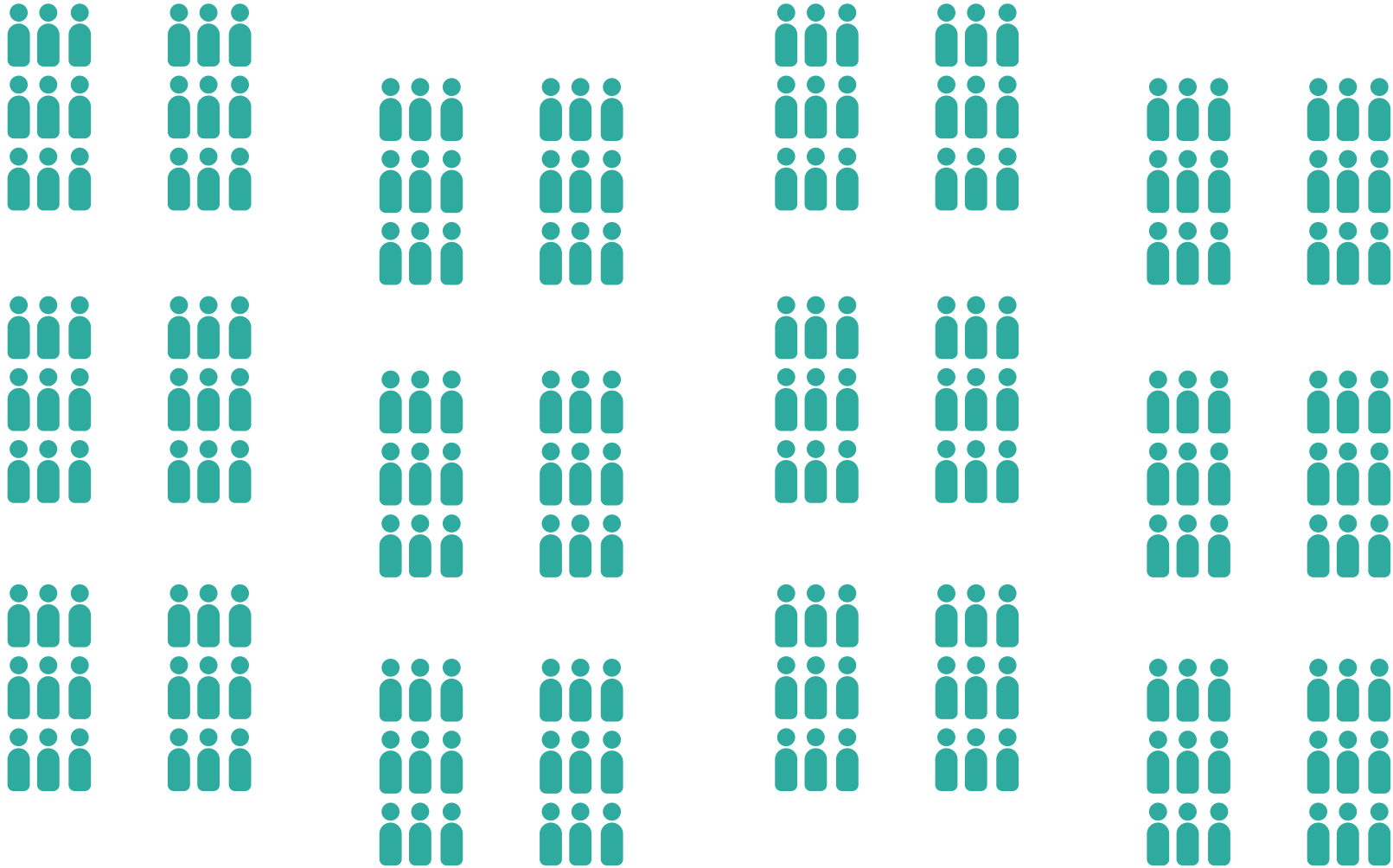
- A. Randomizing timing of access to program
- B. Randomizing encouragement to take up the program
- C. Either A or B – depending on the context, either of these options might be appropriate.**
- D. Not sure

RCT Design: Phase-in

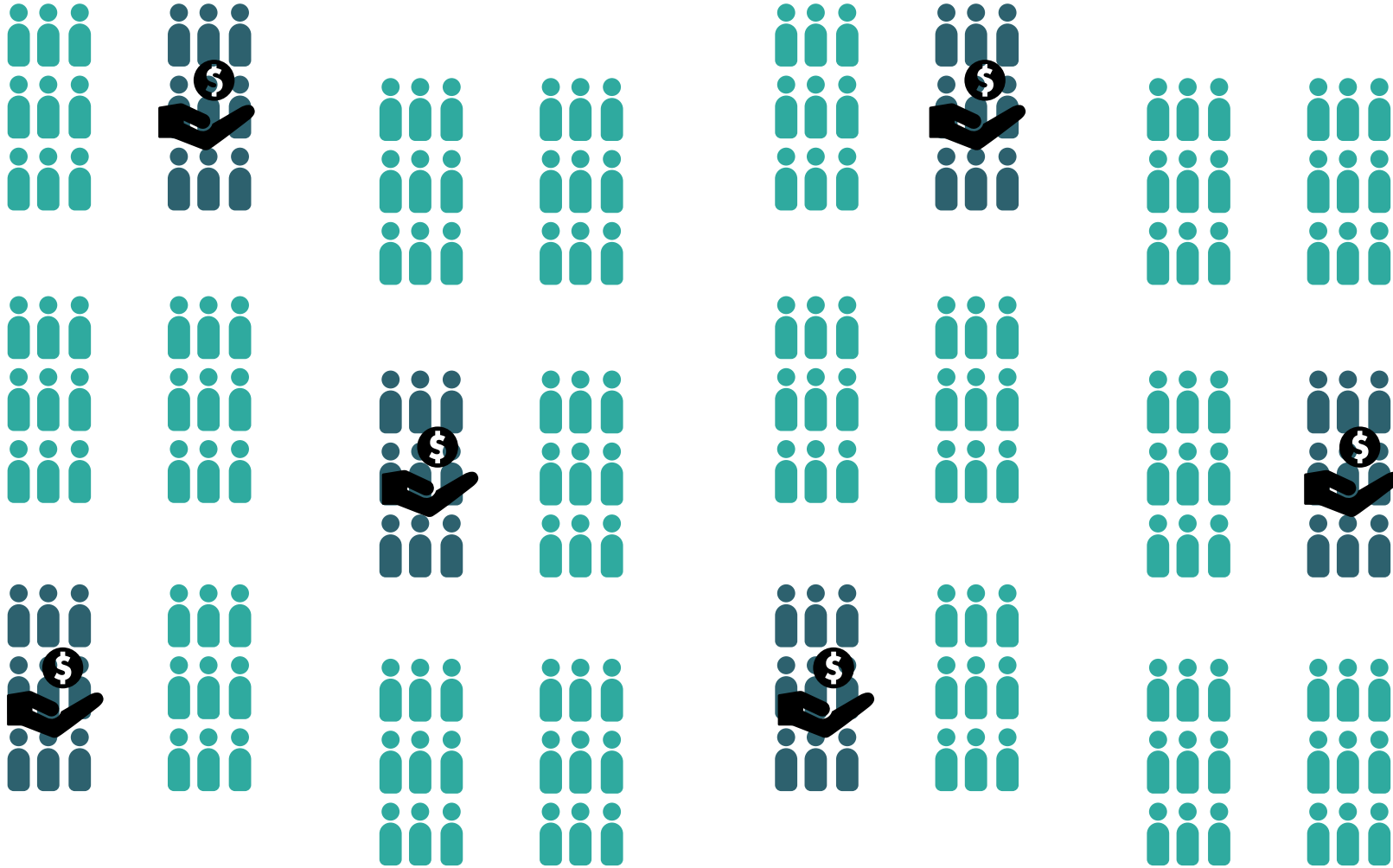


Phase 0: No One Treated Yet

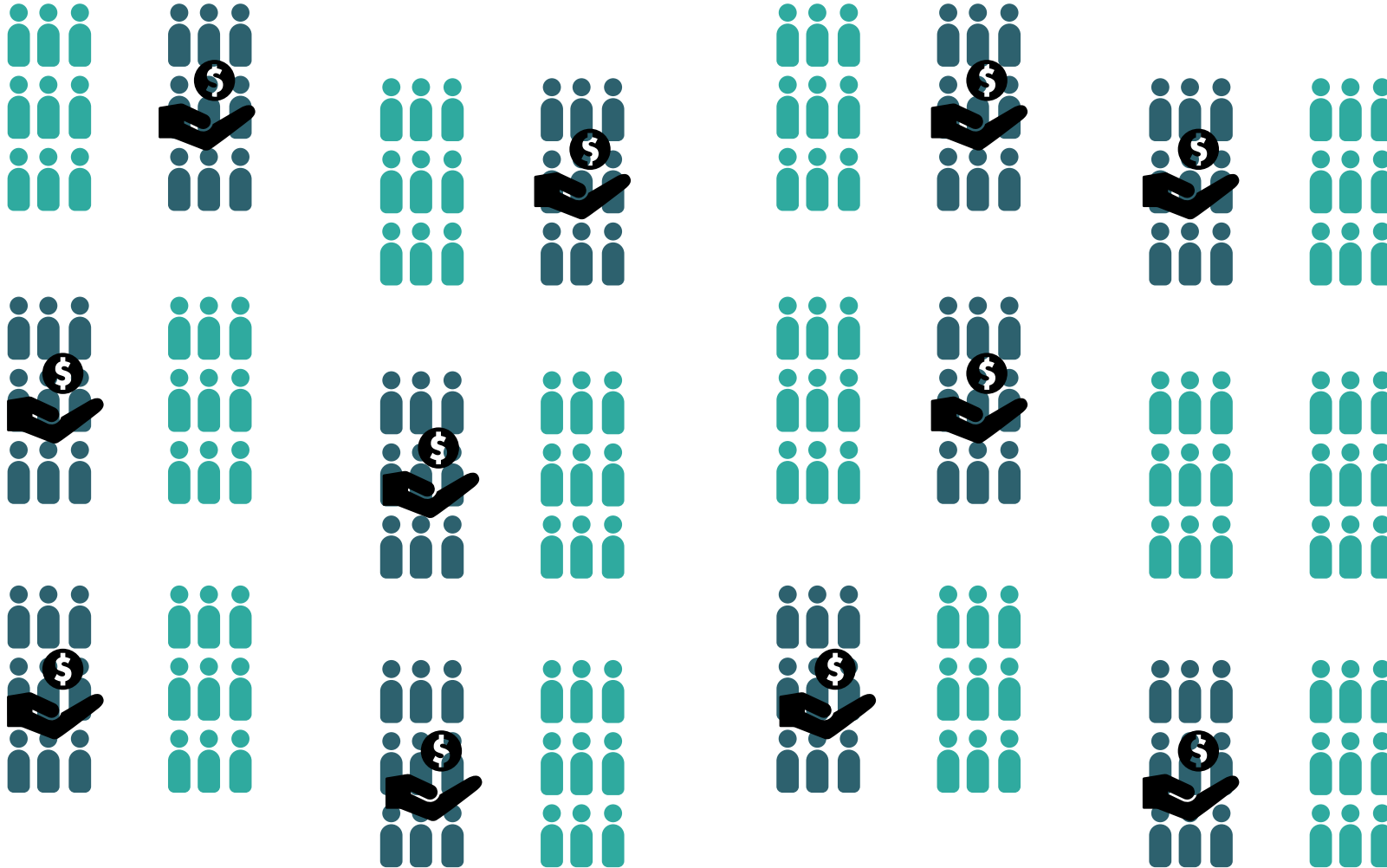
All Control



Phase 1: 25% Treated 75% Control

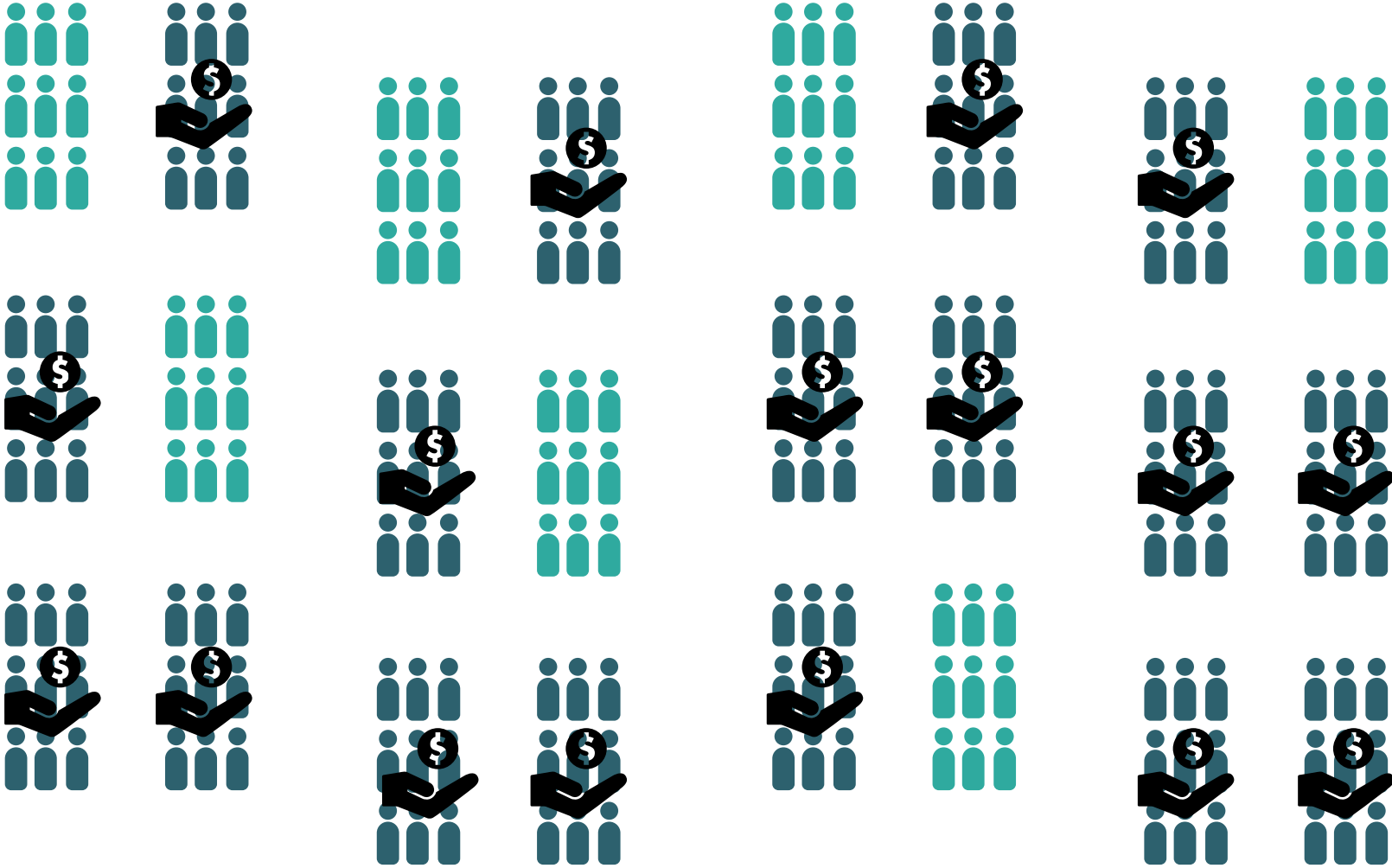


Phase 2: Half Treated Half Control



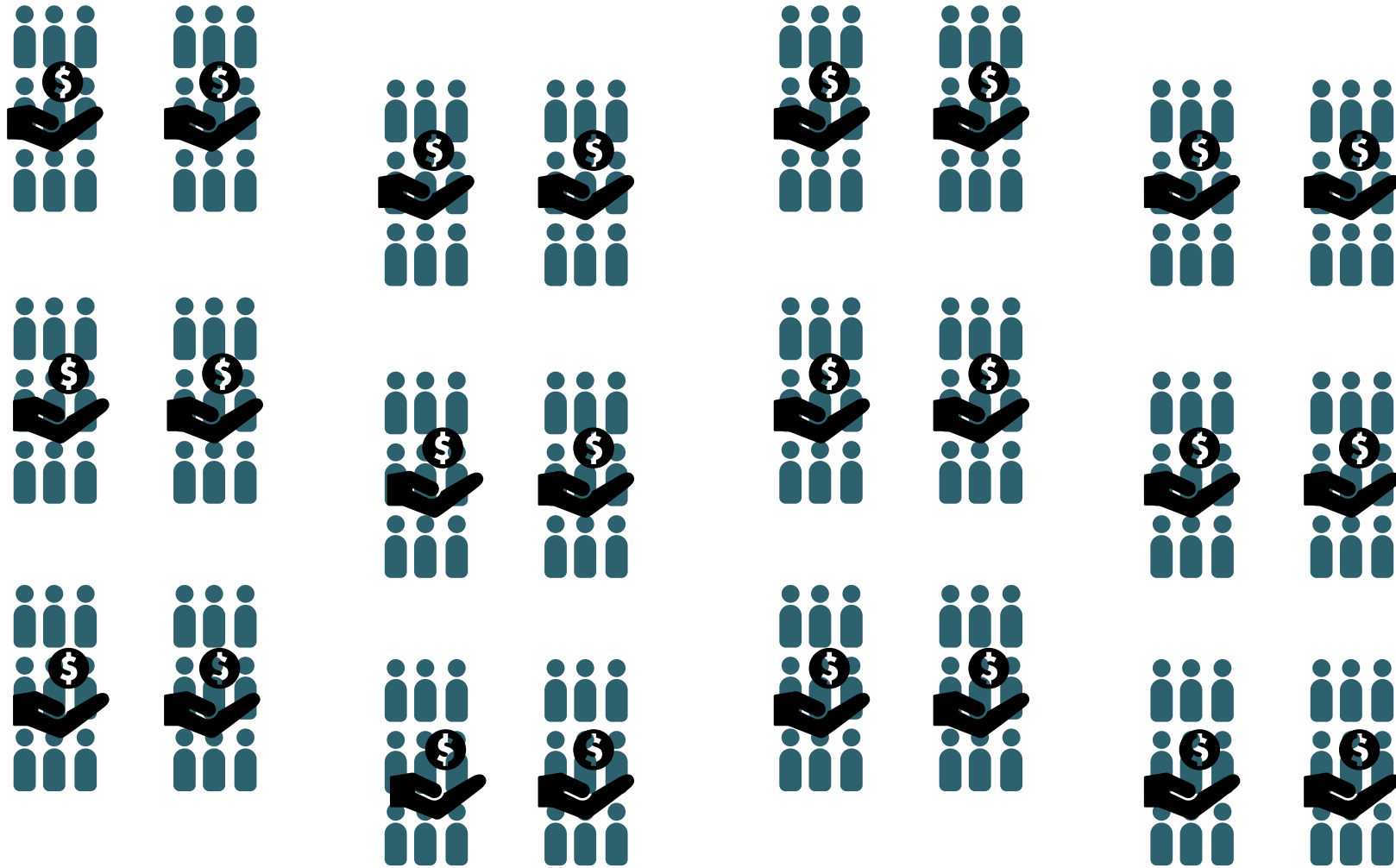
Phase 3: 75% Treated

25% Control



Phase 4: All Treated

No Control (Experiment Over)



Program Specification:
Entitlement Programs

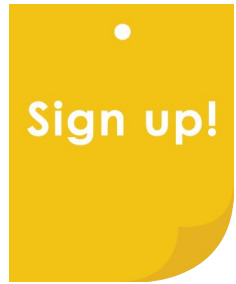


Entitlement Programs: Cannot Mandate nor Deny Intervention

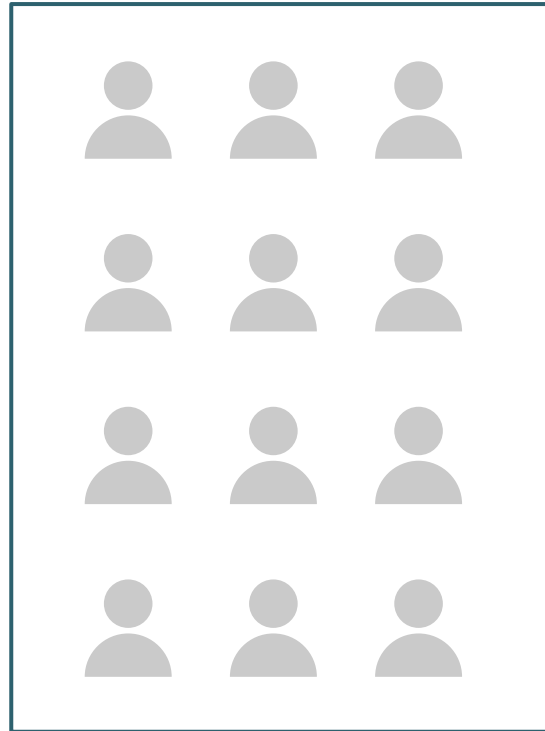


Supplemental
Nutrition
Assistance
Program

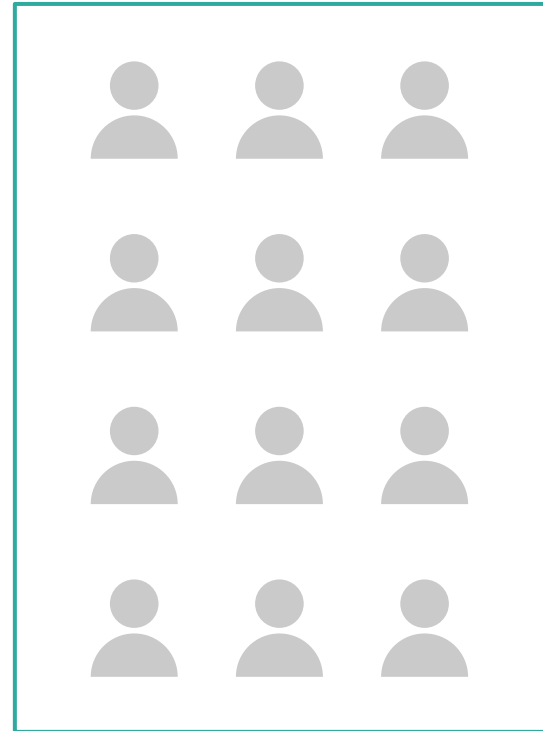
RCT Design: Encouragement



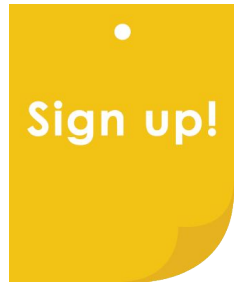
Treatment Group



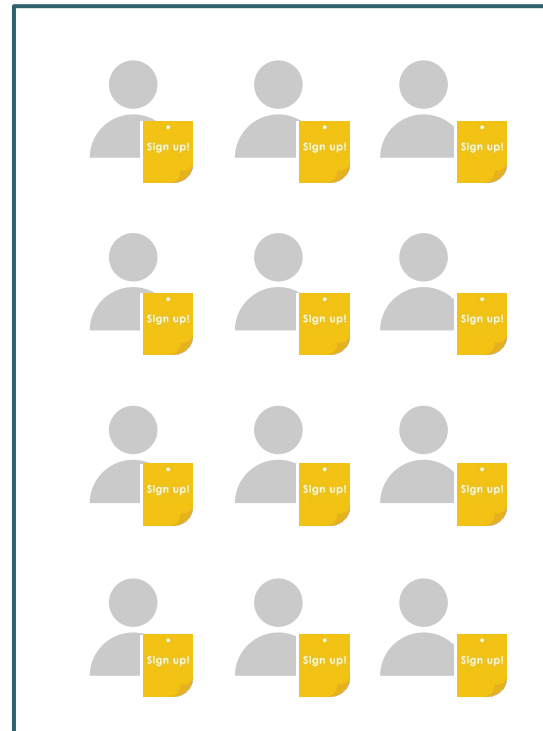
Control Group



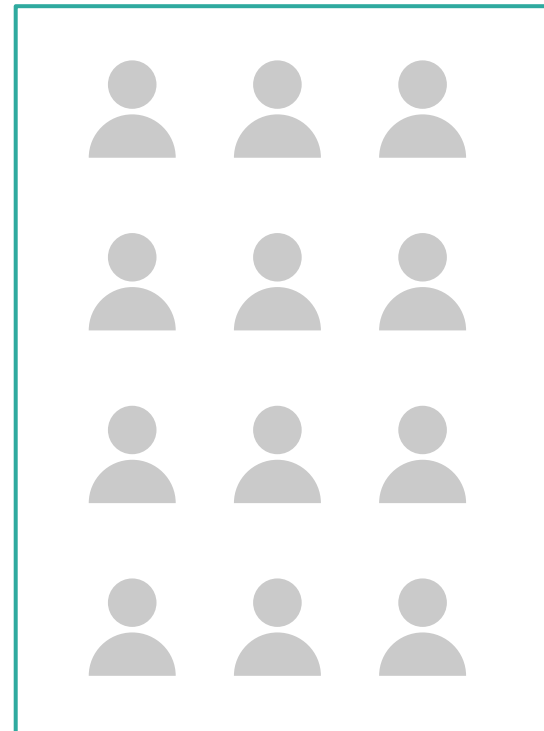
RCT Design: Encouragement



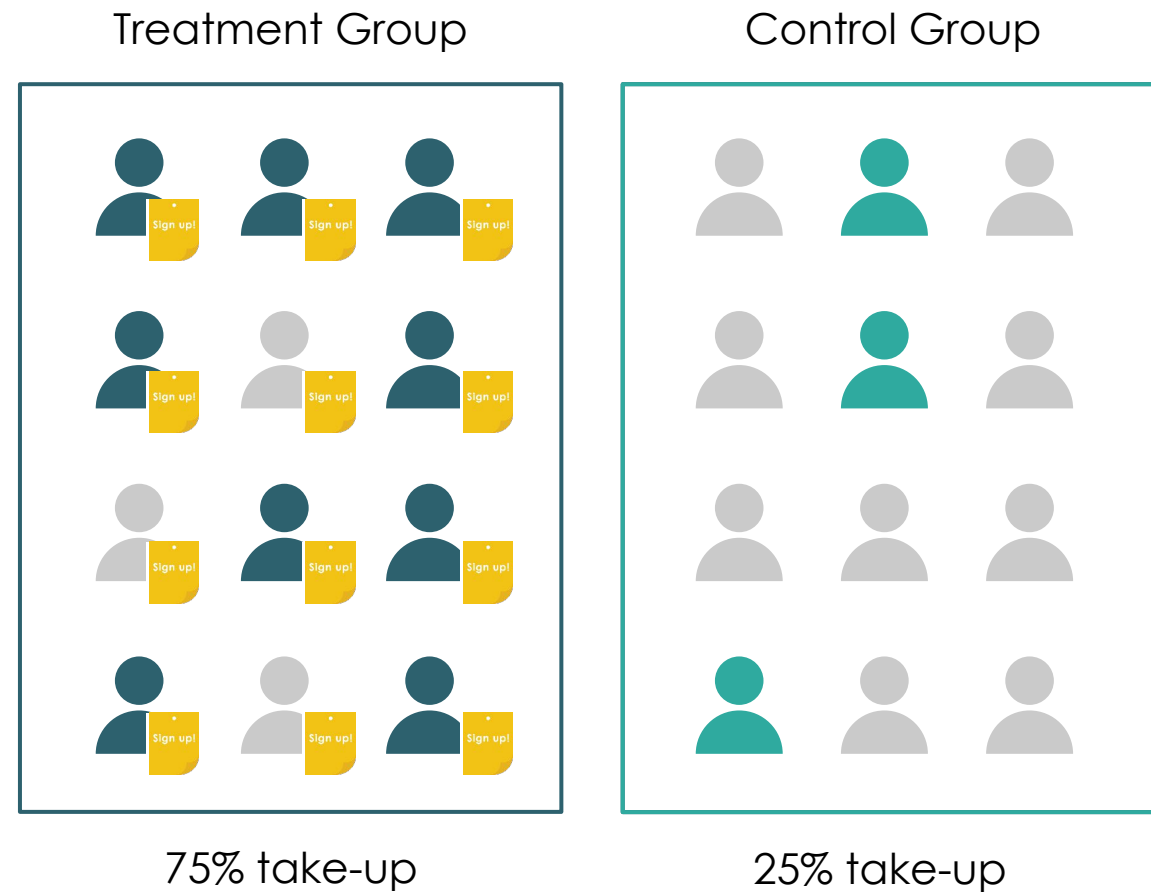
Treatment Group



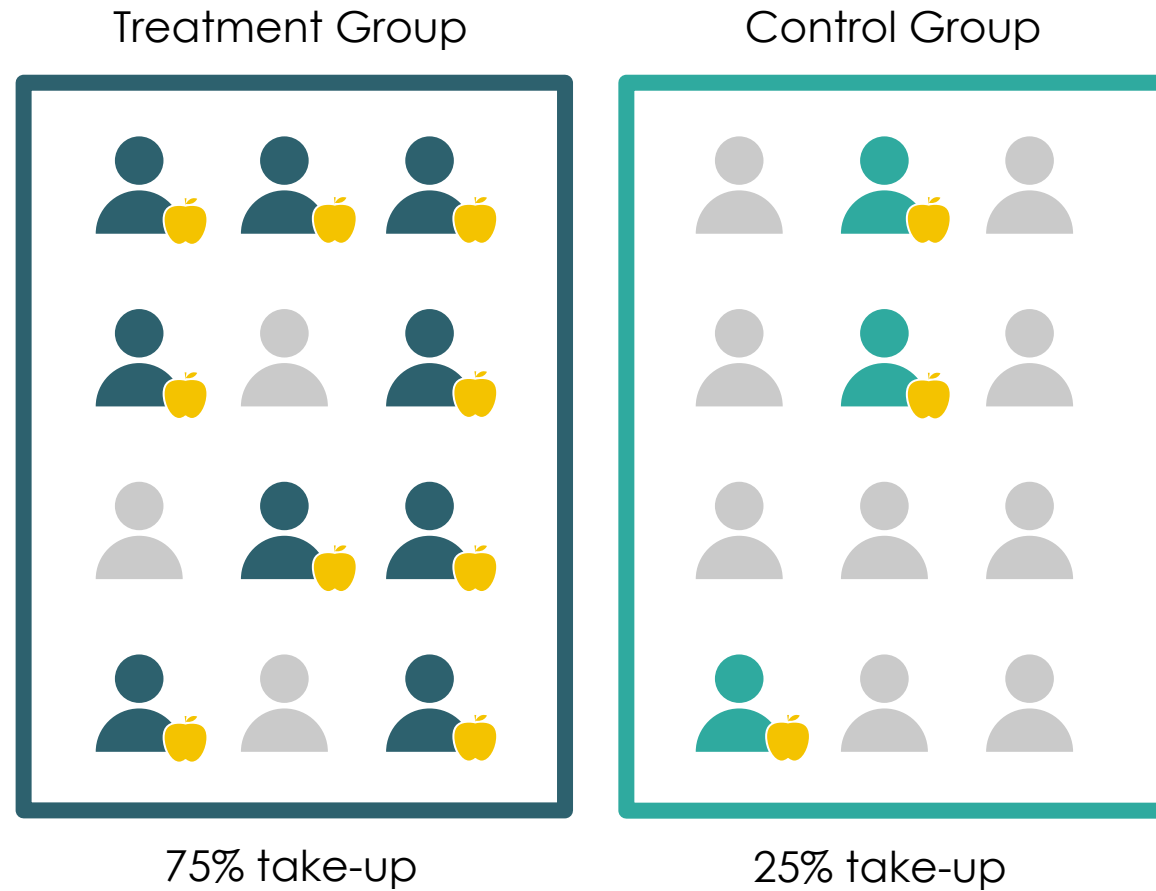
Control Group



RCT Design: Encouragement



RCT Design: Encouragement



Compare
Entire Treatment Group to **Entire Control Group**

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

How might expanding eligibility, phase-in design, and encouragement design be of use in your contexts?

A Very Quick Recap

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

References

Breza, Emily, Fatima Cody Stanford, Marcella Alsan, Burak Alsan, Abhijit Banerjee, Arun G. Chandrasekhar, Sarah Eichmeyer et al. "Effects of a large-scale social media advertising campaign on holiday travel and COVID-19 infections: a cluster randomized controlled trial." *Nature medicine* 27, no. 9 (2021): 1622-1628.

<https://www.nature.com/articles/s41591-021-01487-3>

Alsan, Marcella, Fatima Cody Stanford, Abhijit Banerjee, Emily Breza, Arun G. Chandrasekhar, Sarah Eichmeyer, Paul Goldsmith-Pinkham et al. "Comparison of knowledge and information-seeking behavior after general COVID-19 public health messages and messages tailored for Black and Latinx communities: a randomized controlled trial." *Annals of internal medicine* 174, no. 4 (2021): 484-492.

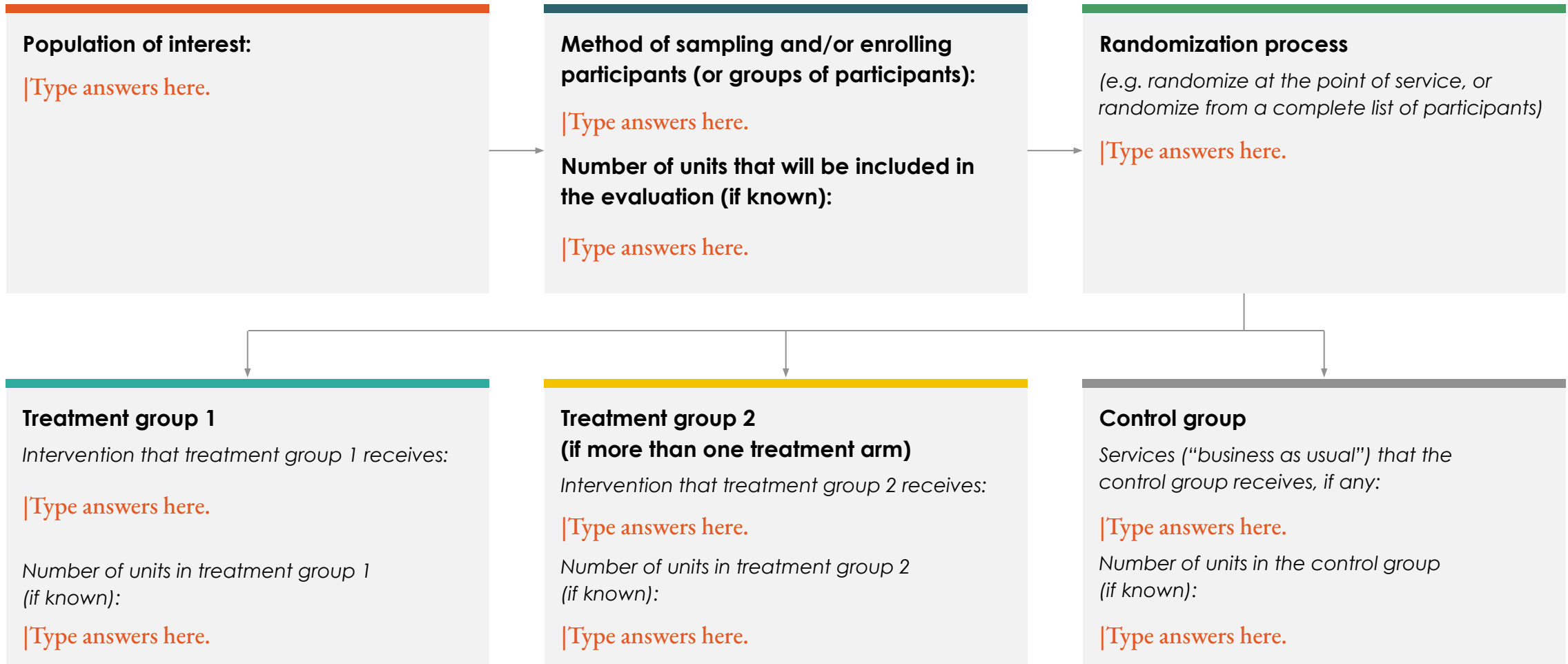
<https://www.acpjournals.org/doi/full/10.7326/M20-6141>

Appendix





How to randomize



Which variables to use for stratification

- Variables that are **discrete**
 - Practical constraint (it is challenging to 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	<p>Program is oversubscribed/not enough resources to deliver to all eligible</p> <p>Resources are constant for the evaluation period</p>	<p>Familiar and well understood</p> <p>Easy to implement</p> <p>Often seen as fair</p> <p>Transparent process</p>	<p>Control group may be less interested in answering questions/being involved, which could create problems for the analysis (differential attrition)</p>

Recap: Randomization designs

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

Recap: Randomization designs

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

Recap: Randomization designs

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

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](#)

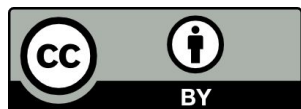
Additional J-PAL resources

- [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. 2024. “Lecture: How to Randomize.” Abdul Latif Jameel Poverty Action Lab. Cambridge, MA



J-PAL, 2024

This lecture is made available under a Creative Commons Attribution 4.0 License (international): <https://creativecommons.org/licenses/by/4.0/>