



How to Randomize

Santosh Harish

Energy Policy Institute at University of Chicago

4 July 2017



Course Overview

1. What is Evaluation?
2. Theory of Change
3. Outcome, Impact & Indicators
4. Why Randomize?
- 5. How to Randomize?**
6. Sampling and Sample Size
7. Threats and Analysis
8. Research to Policy
9. Project from Start to Finish

Lecture Overview

- What is randomization
- Simple Randomization
- Unit of Randomization
- Methods for different evaluation questions
- Real world challenges & design solutions

Lecture Overview

- What is Randomization
- Simple Randomization
- Unit of Randomization
- Methods for different evaluation questions
- Real world challenges & design solutions

Random Selection

Random Selection



Random Selection

Monthly income, per capita

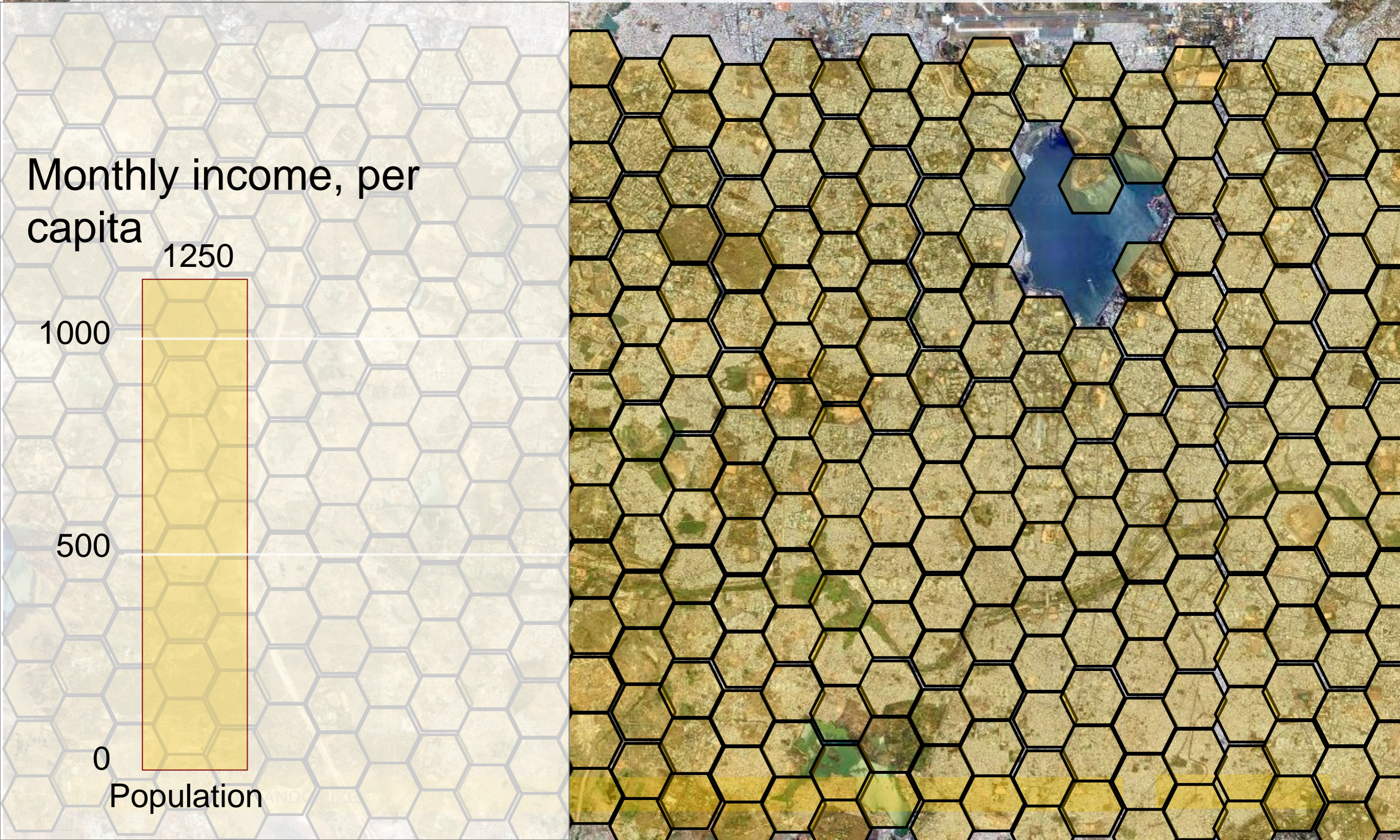
1250

1000

500

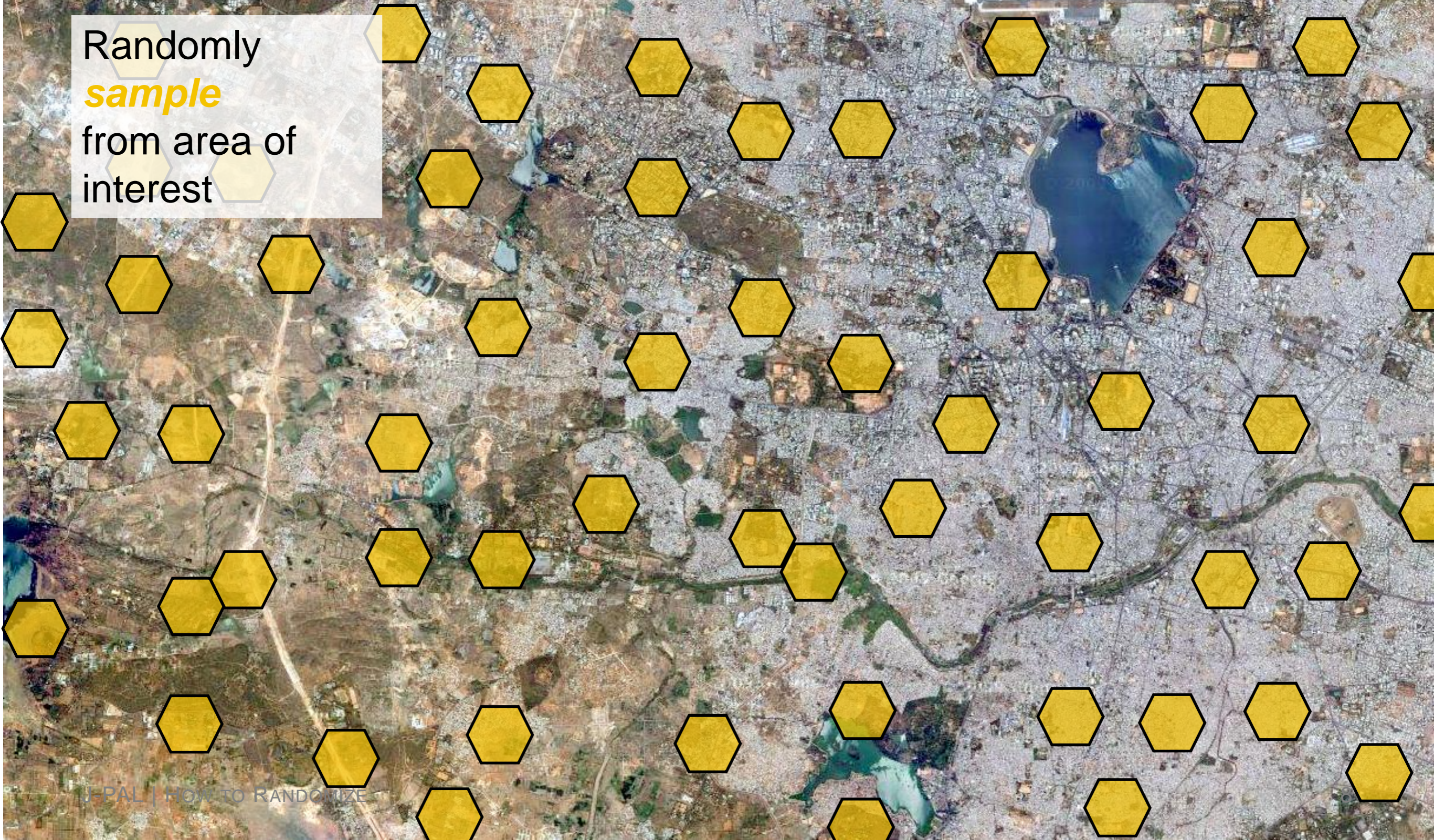
0

Population



Random Selection

Randomly
sample
from area of
interest



Random Selection

Monthly income, per capita

1250

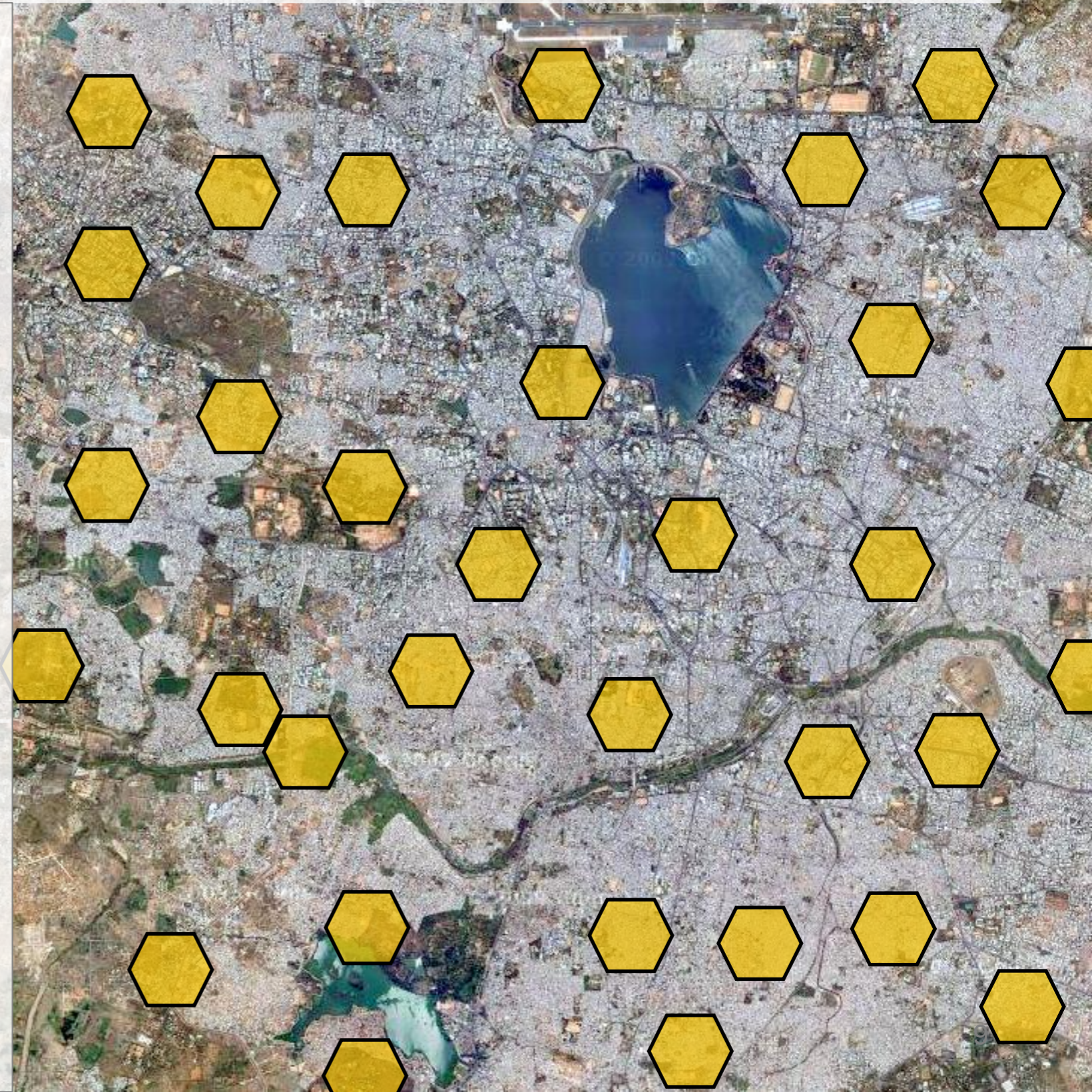
1252

1000

500

0

J-PRC
Population Sample



Random Assignment

Randomly
assign
to **treatment**

Random Assignment

Monthly income, per capita

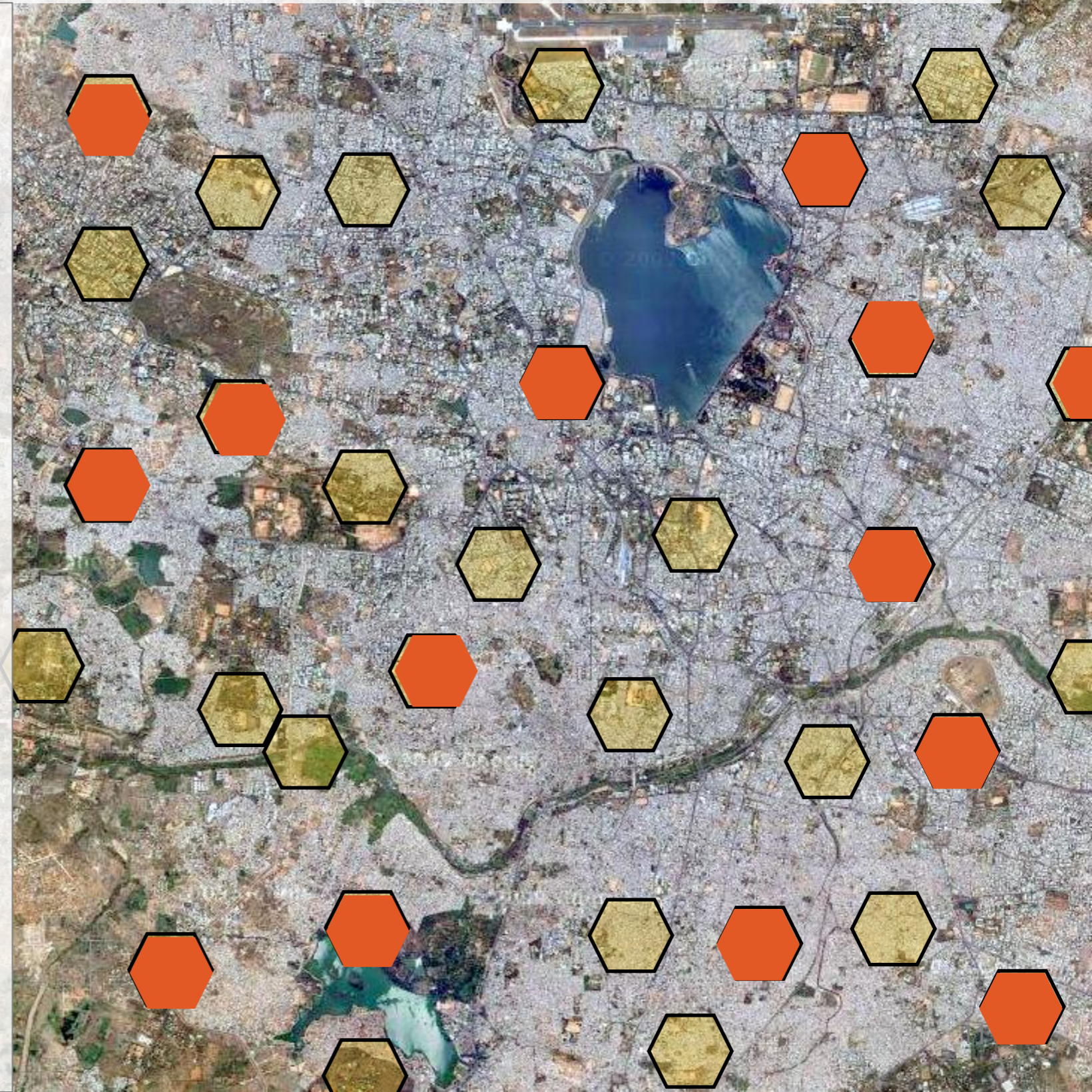
1250 1257

1000

500

0

J-PRG PHD RAND SITE
Population Treatment



Random Assignment

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

Random Assignment

Monthly income, per capita

1250

1257

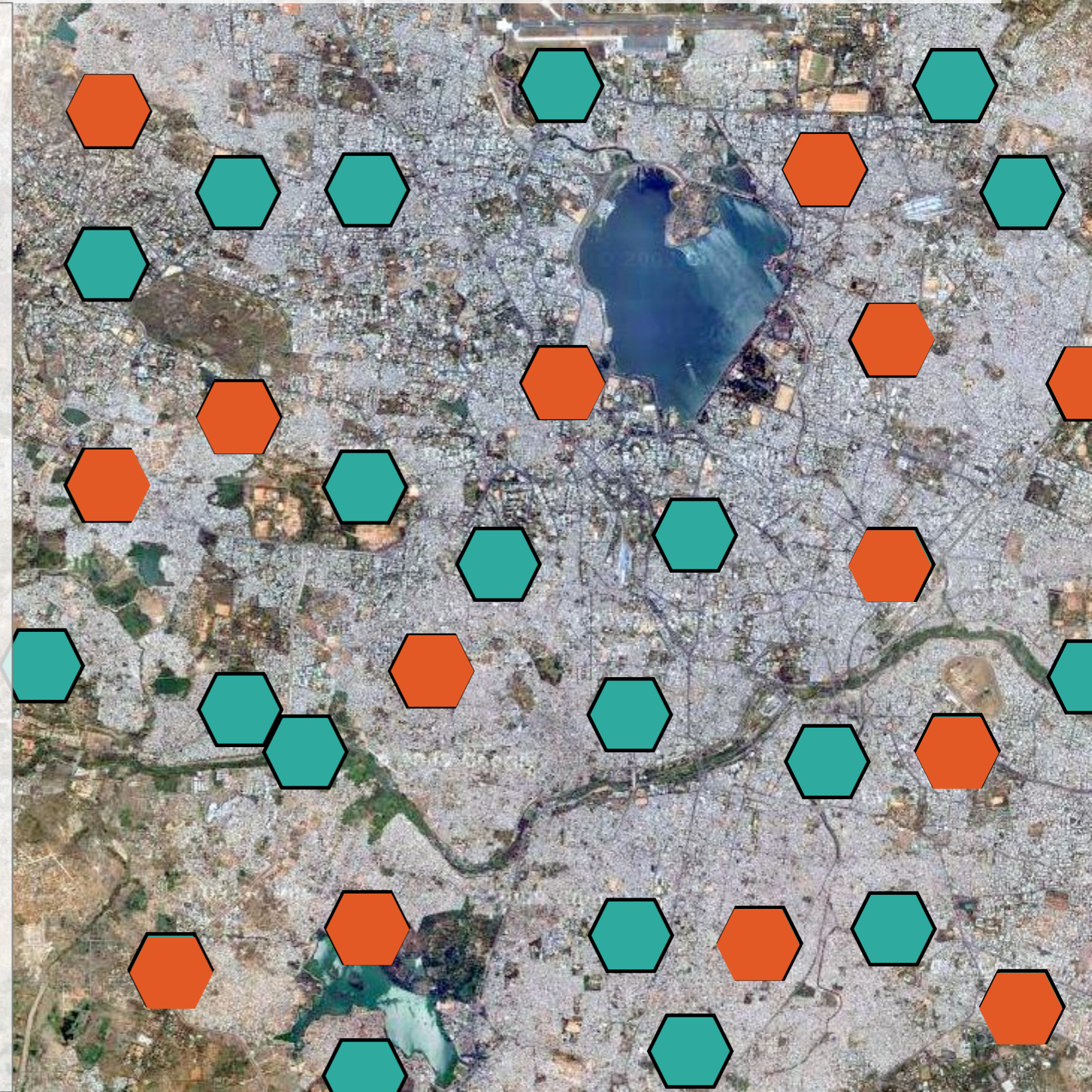
1244

1000

500

0

J-PRC PHD (RANDOM) 2012
Population Treatment Control

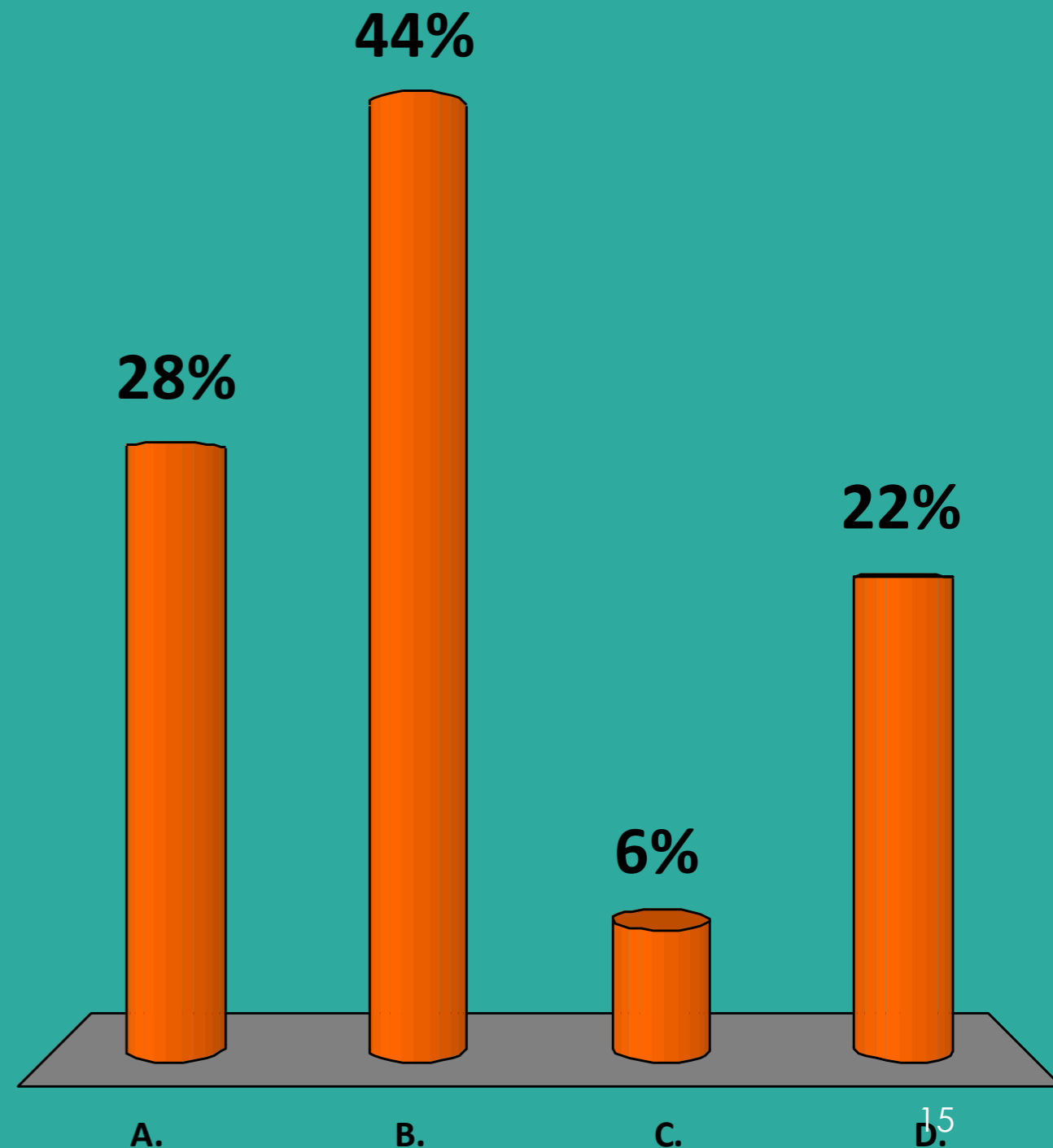


Alternate methods of Randomization?

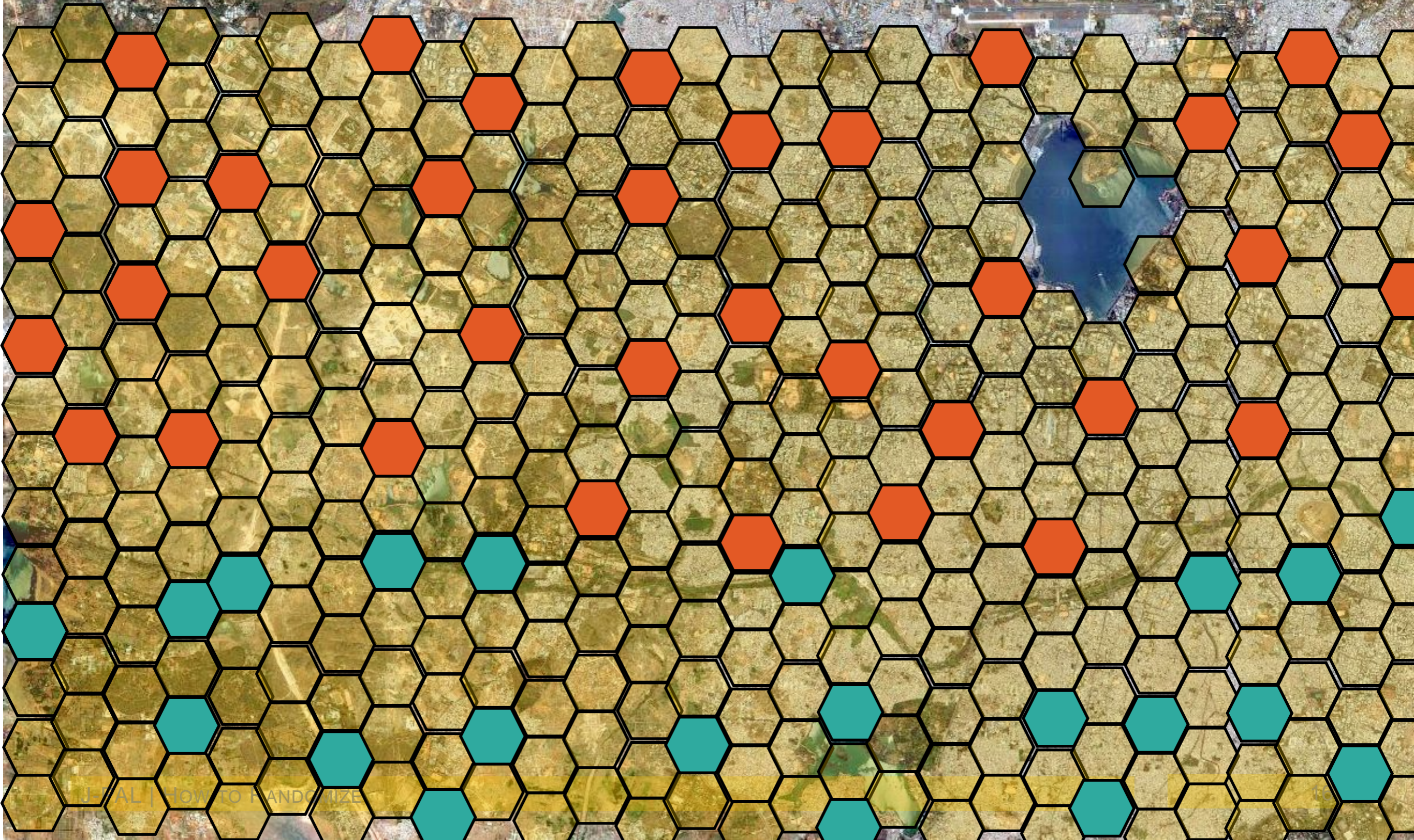


If we took a random sample from the South of the city (bottom of the screen) for C, and one from the North (top of the screen) as T, in expectation, the difference in income will be statistically insignificant.

- A. True
- B. False
- C. Don't know
- D. It depends



NOT Random Assignment



NOT Random Assignment

Monthly income, per capita

1453

1250

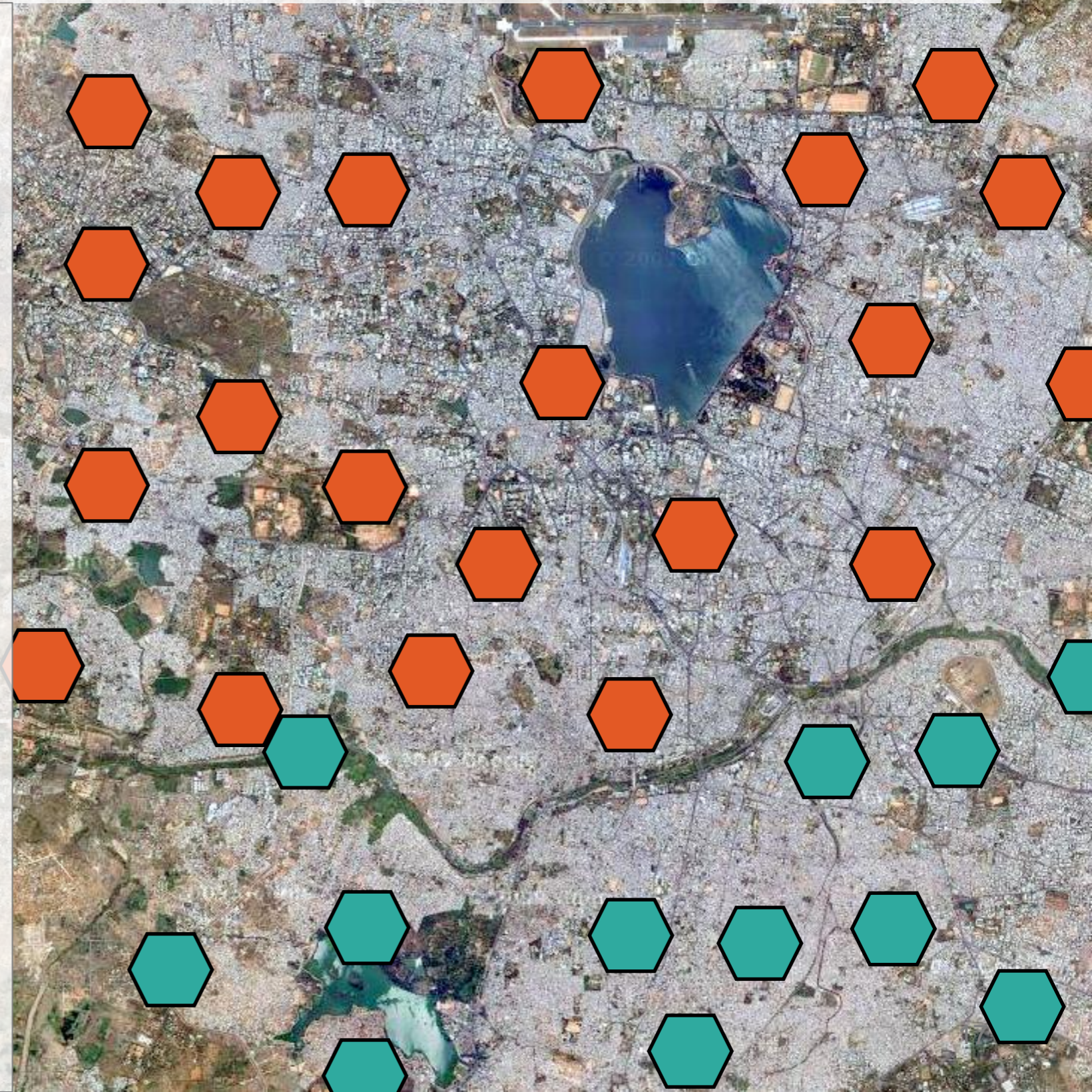
1000

942

500

0

J-PRG PHD (RANDOM) SITE
Population Treatment Control



Lecture Overview

- What is Randomization
- Simple Randomization
- Unit of Randomization
- Methods for different evaluation questions
- Real world challenges & design solutions

Simple randomization: Fixed probability

- For each member, set probability (e.g. 50%).
 - Spot randomization
 - Point-of-service randomization
- May end up with slightly more in one group and fewer in the other

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



Complete randomization: Fixed *proportion*

- Need sample frame
- Determine number in treatment (and in control)
- Pull out of a hat/bucket
-or-
- Use random number generator to order observations randomly
- What if no lists exist?

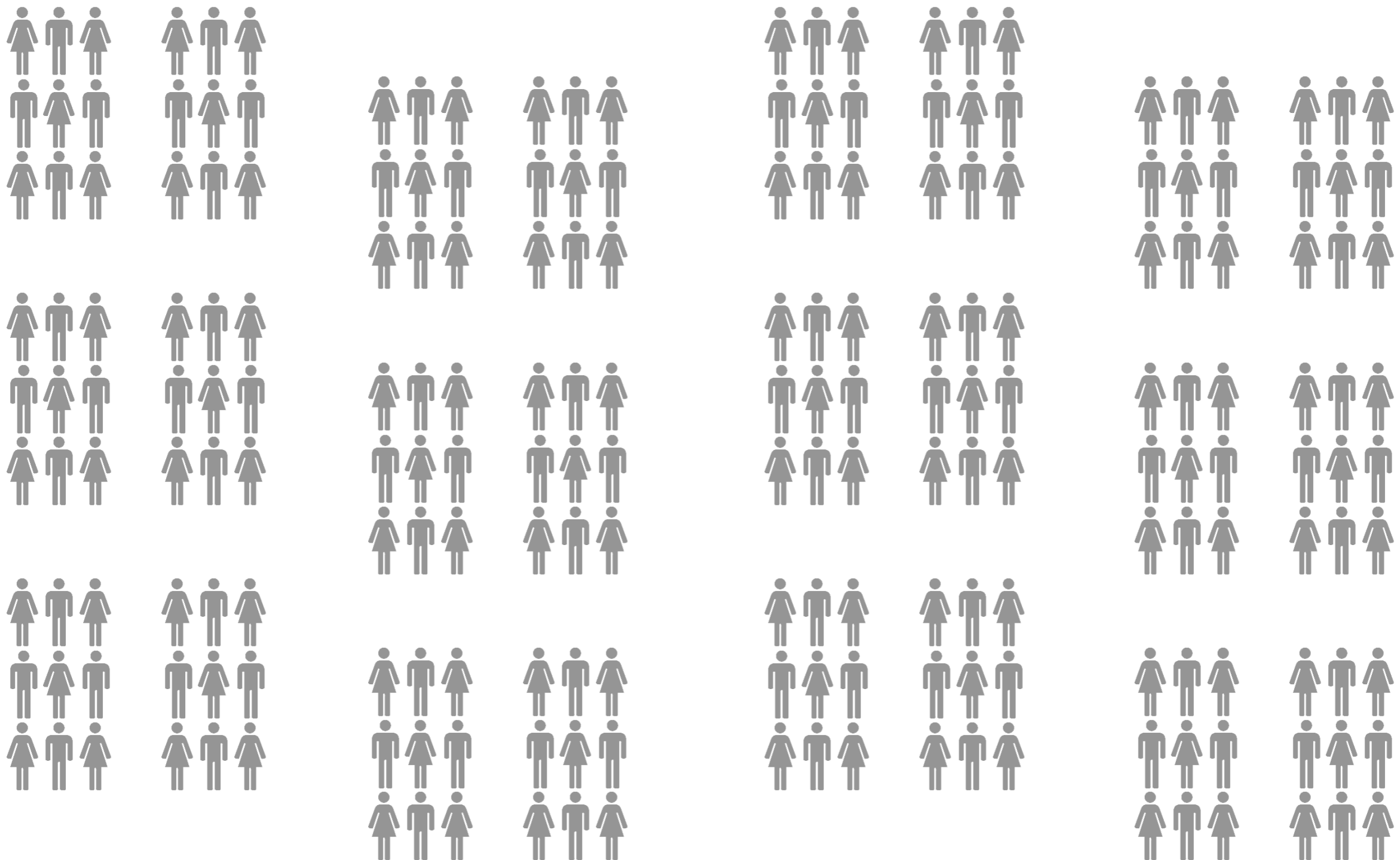


Source: Chris Blattman

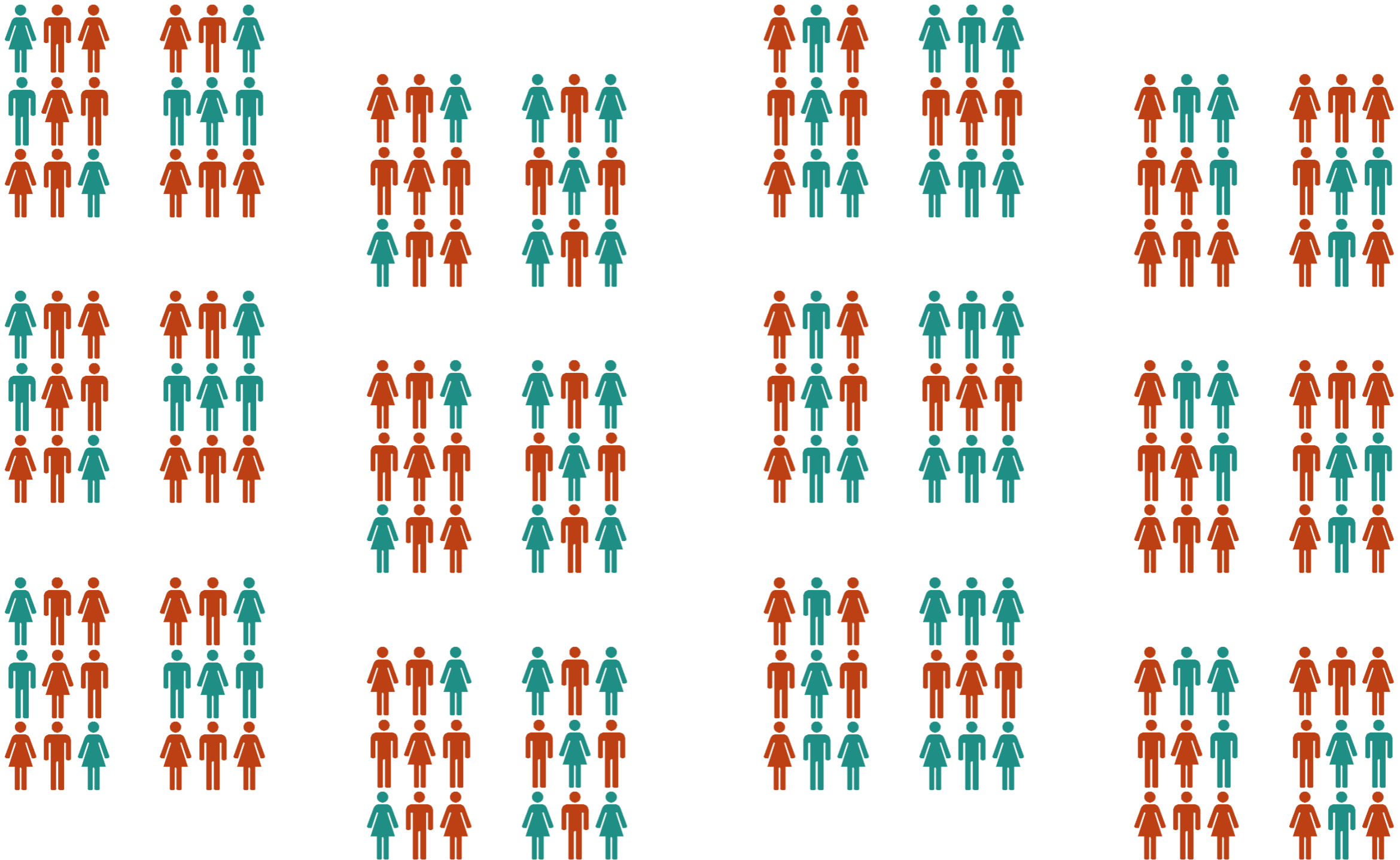
Lecture Overview

- What is Randomization
- Simple Randomization
- Unit of Randomization
- Methods for different evaluation questions
- Real world challenges & design solutions

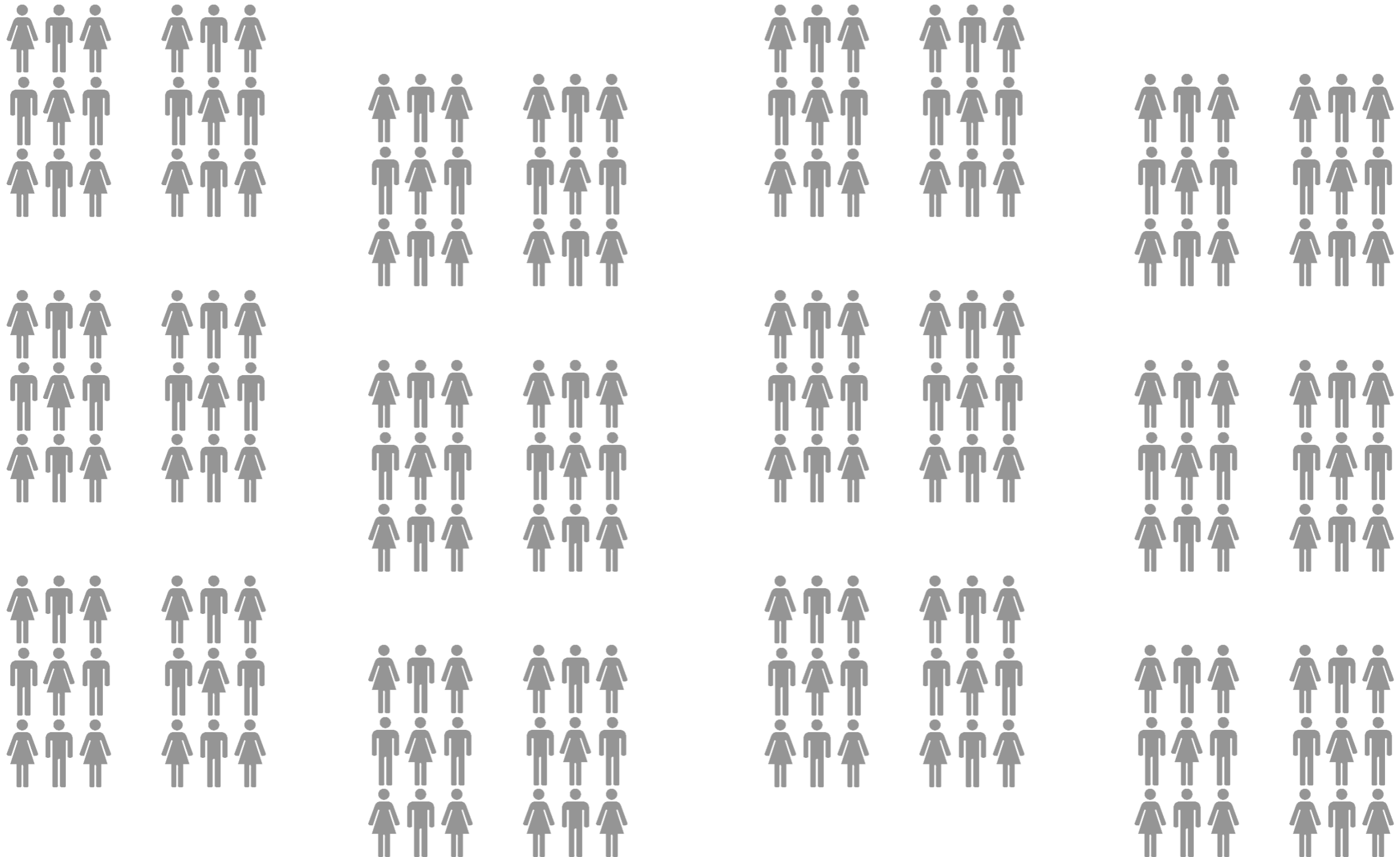
Unit of Randomization: Individual?



Unit of Randomization: Individual?



Unit of Randomization: Clusters?



Unit of Randomization: Class?



Unit of Randomization: Class?



Unit of Randomization: School?



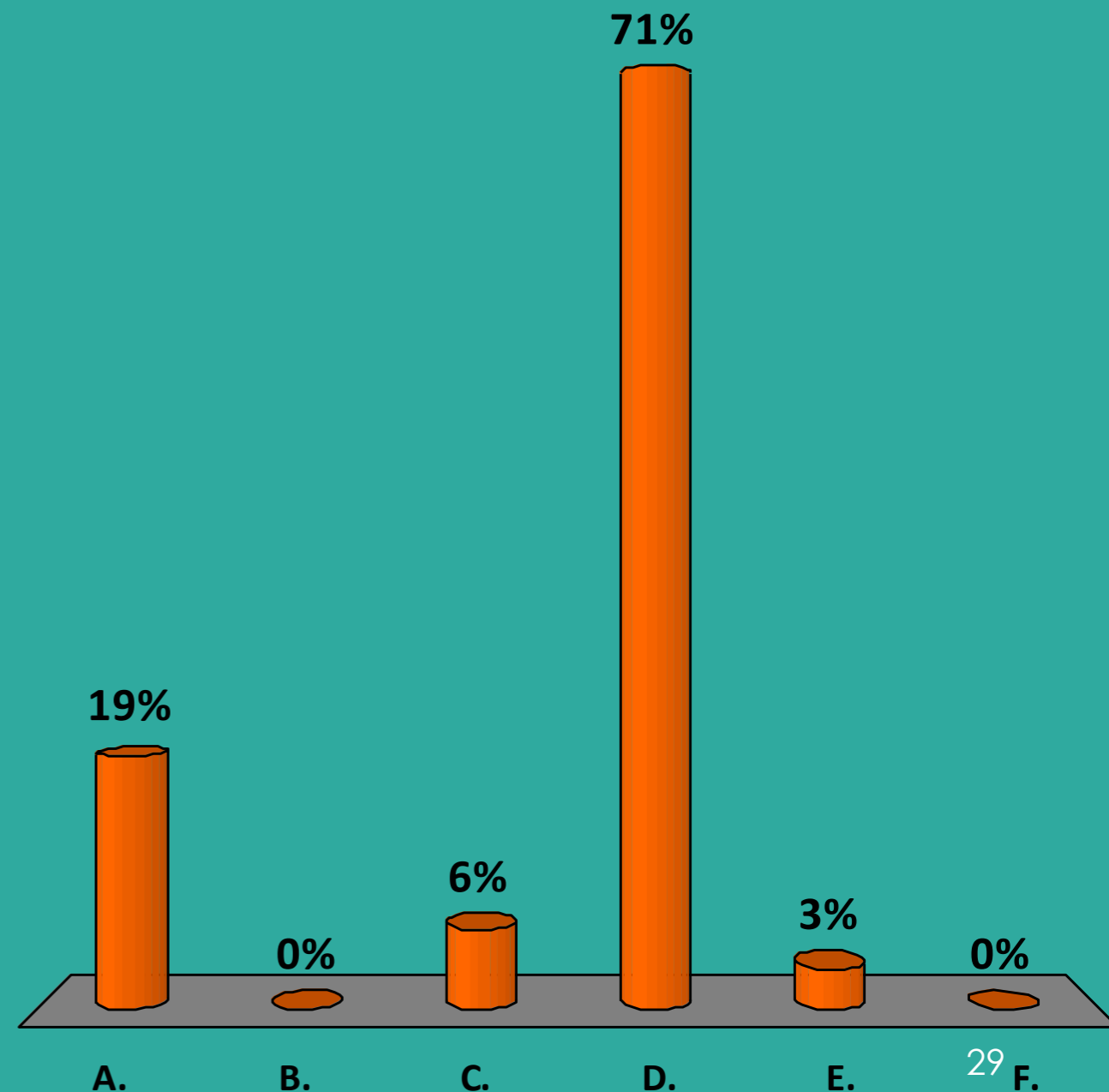
Unit of Randomization: School?



An education department wants to see if increasing the duration of recess can help reduce rates of obesity.

What is the appropriate unit of randomization?

- A. Child level
- B. Household level
- C. Classroom level
- D. School level
- E. Village level
- F. Don't know

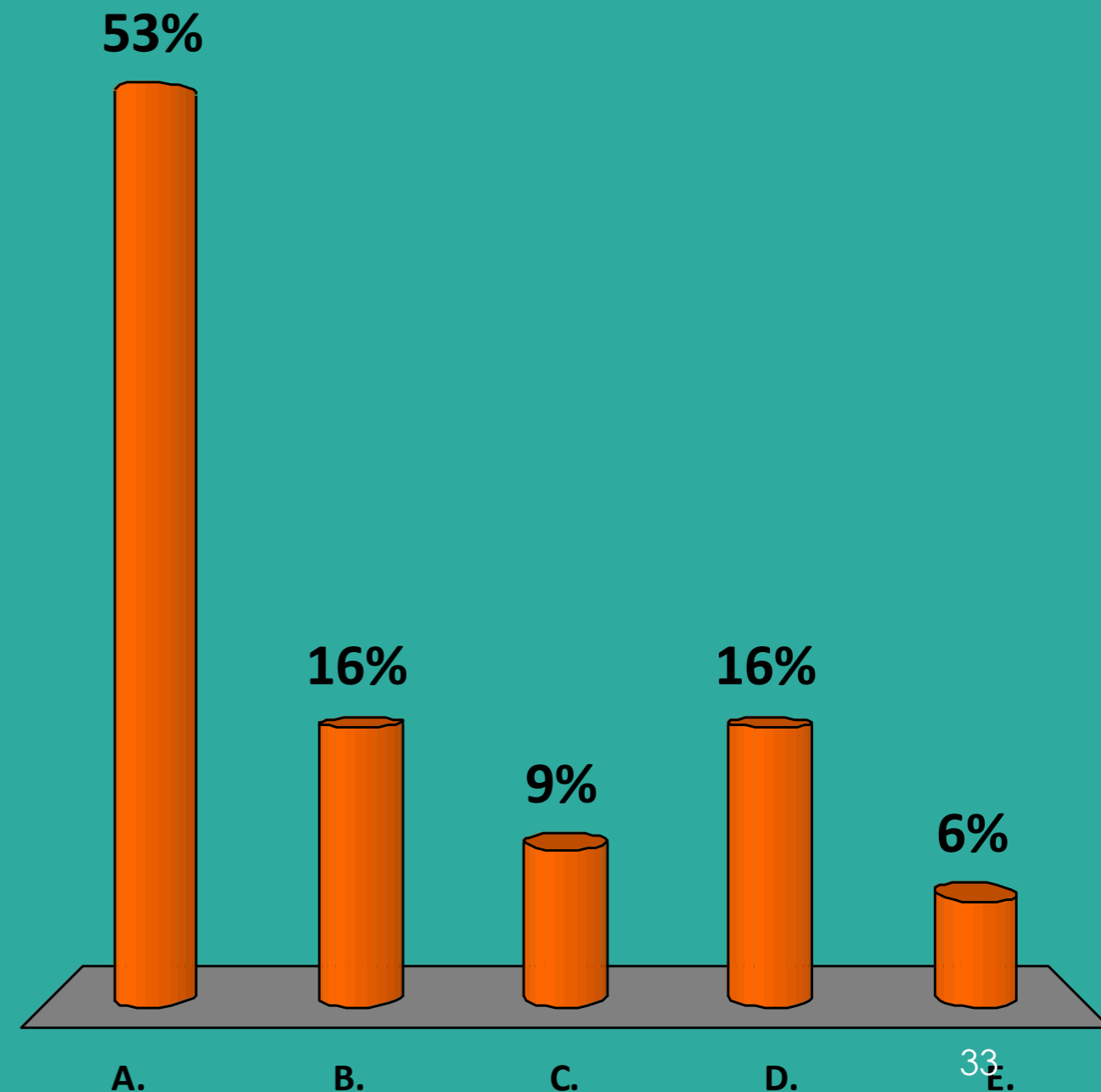


Lecture Overview

- What is Randomization
- Simple Randomization
- Unit of Randomization
- **Methods for different evaluation questions**
- Real world challenges & design solutions

The department of agriculture believes that if farmers used more fertilizer yields would improve. One advisor believes organic fertilizer will be more effective; a second believes inorganic fertilizer is better; a third believes neither will be effective.
Can we test all three beliefs within one single experiment?

- A. Yes, and we should
- B. No, they can only be answered with two separate experiments
- C. No they can only be answered with three separate experiments
- D. Yes, but best practice is to run separate experiments
- E. Don't know



1. Multiple treatments

Treatment 1

Treatment 2

Control



Reducing crime

The newly elected governor is looking for strategies to lower crime

- A. Advisor A suggests crime is an economic phenomenon. The best strategy to fight crime is employment. He proposes job training
- B. Advisor B says criminals already have a choice to work, and can always make more money committing crimes. We need to take the choice away with more law and order. He proposes hiring more police officers to patrol the streets.
- C. Advisor C says we need to simultaneously raise the cost of committing crime by increasing the chances of getting caught. So more officers are needed. But many criminals don't see formal employment as a choice. We need to reduce the cost of finding a job. Job training is also needed. He claims a combined strategy will be more cost effective than each individual strategy by itself.

Advisor A: Job Training

Advisor B: More police officers

Advisor C: Combined strategy

How many treatment arms should we use to test all three advisors' hypotheses?

A. 1

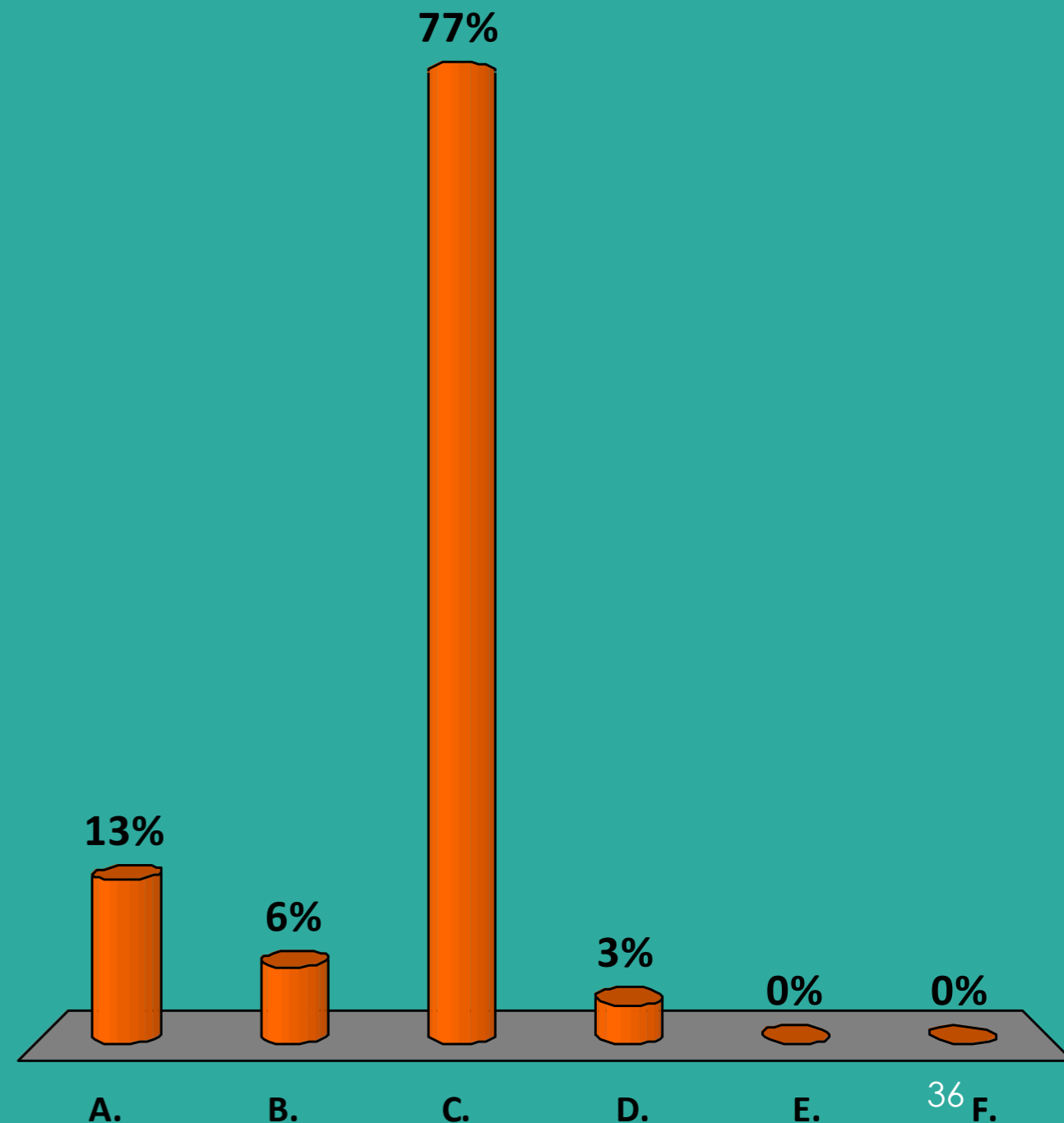
B. 2

C. 3







D. 4

E. 5



















































F. Don't know





































































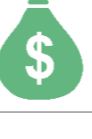

































2. Cross-cutting treatments: *Factorial Design*

		 Technical Knowledge	
		Y	N
Cash Grants 	Y	 Group 1 Training + Cash 	 Group 2 Cash
	N	Group 3 Training 	Group 4 Control

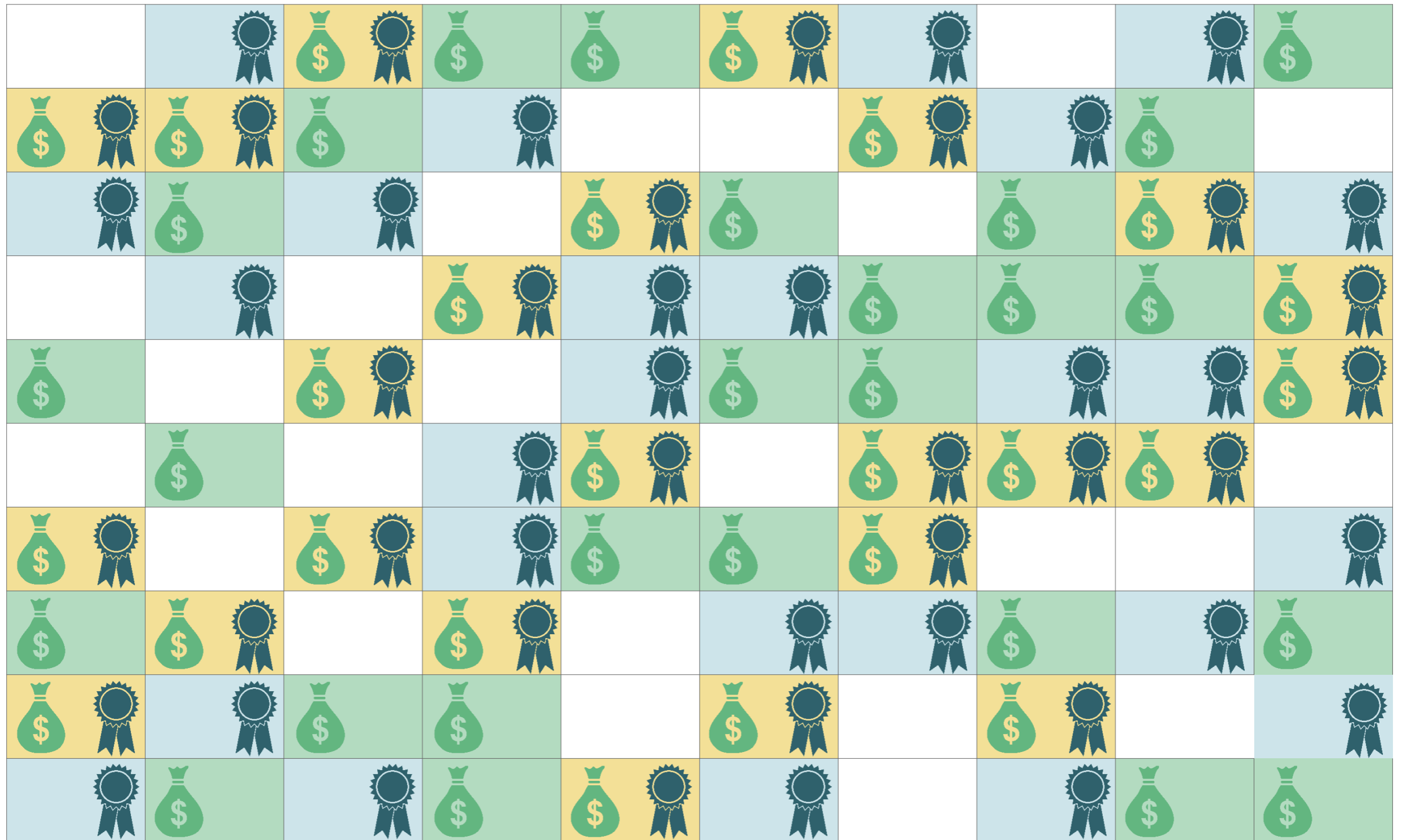
Cross-cutting treatments







Cross-cutting treatments

Cross-cutting treatments



Cross-cutting treatments

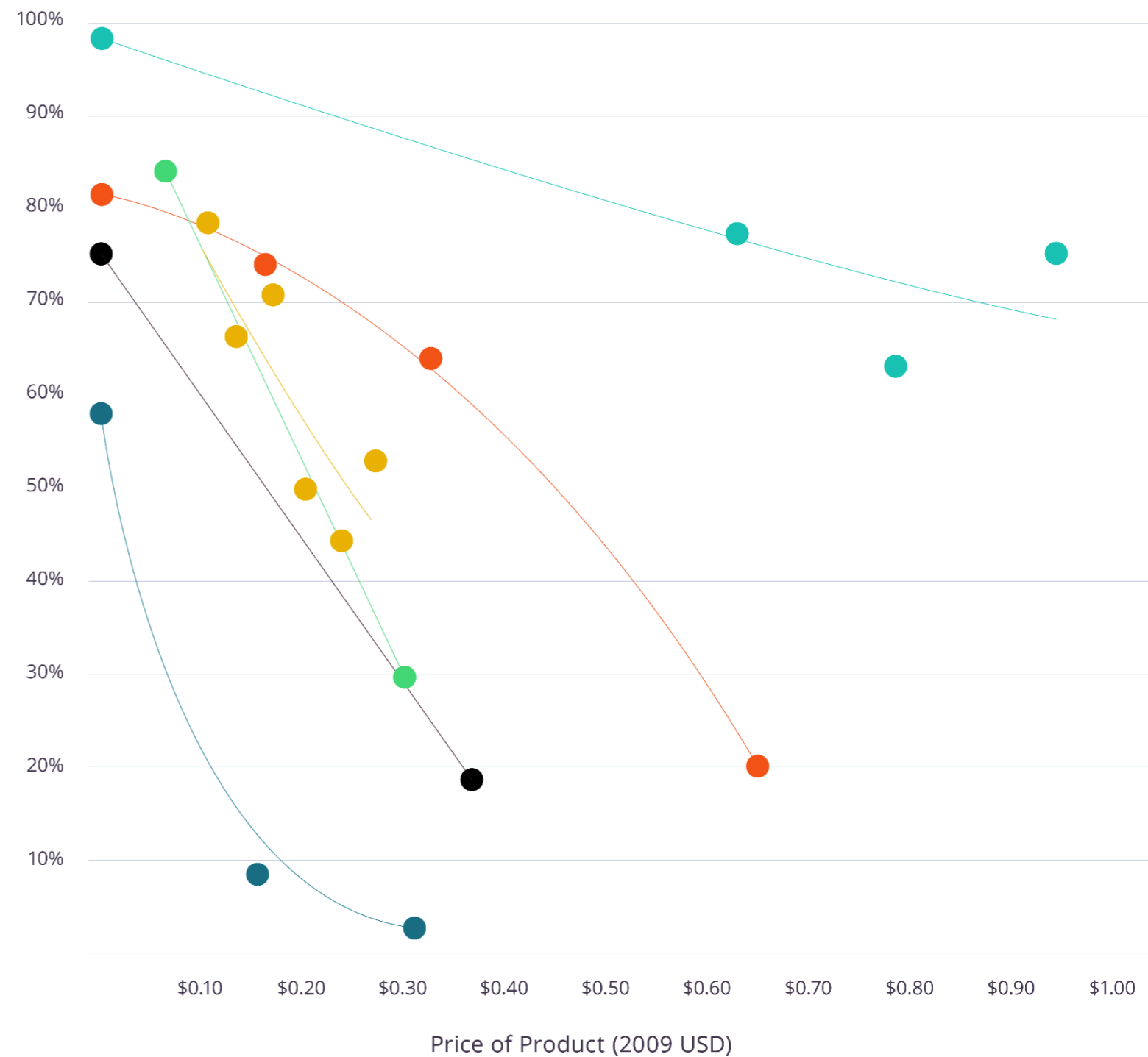
		 Technical Knowledge	
		Y	N
Cash Grants 	Y	 Group 1 Training + Cash 	 Group 2 Cash
	N	Group 3 Training 	Group 4 Control

3. Varying intensity of treatment

- To Measure:
 - Dosage
 - Sensitivity
 - Elasticity
 - Spillovers

Varying intensity of treatment (individual)

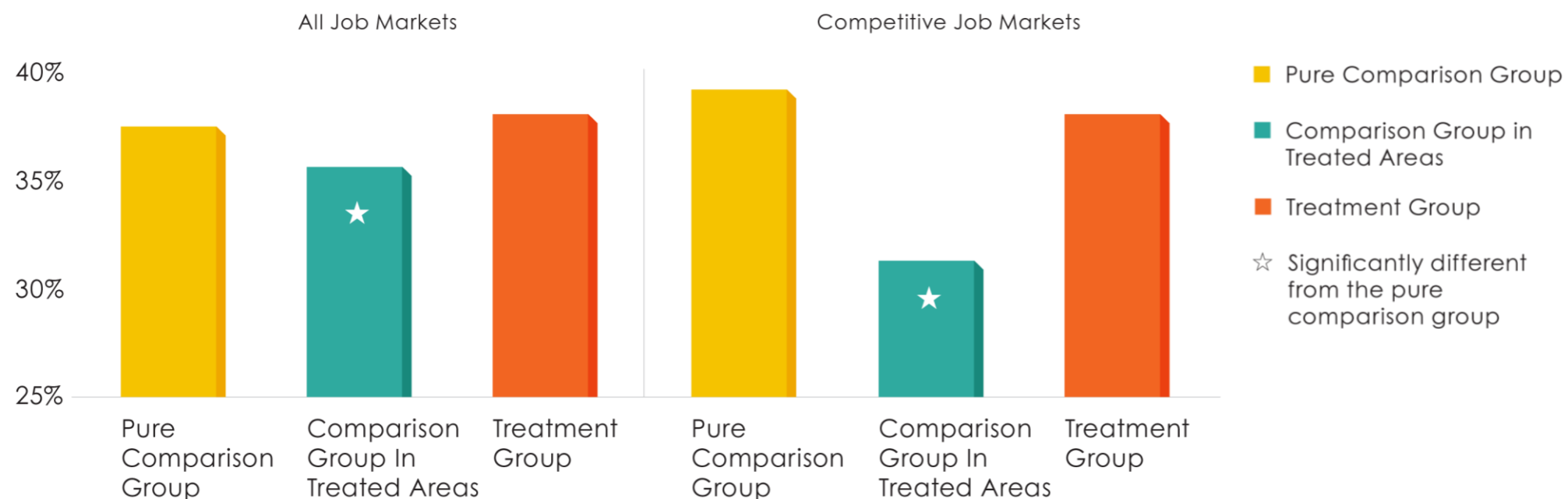
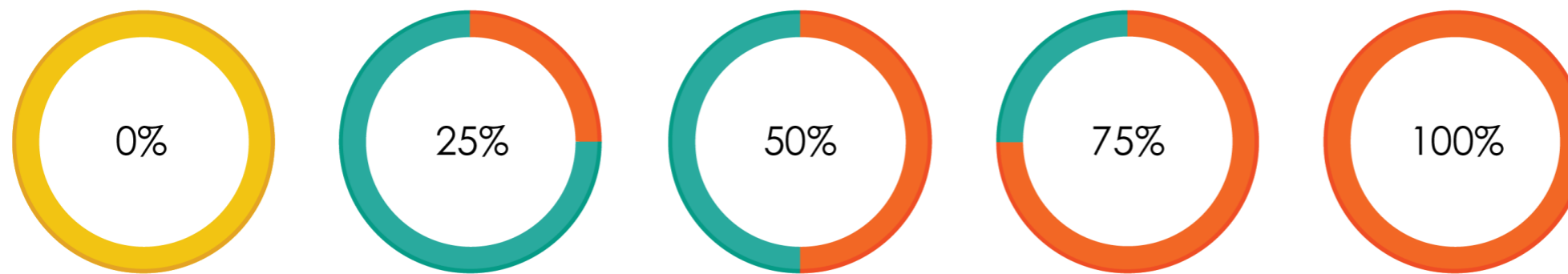
- Dosage
- Sensitivity
- Elasticity



- Deworming, Kenya 1
- Soap, India 4
- Water Disinfectant, Kenya 5
- Bednets in Clinics, Kenya 2
- Water Disinfectant, Zambia 3
- Bednet Vouchers, Kenya 6

Varying intensity of treatment (individual)

- Spillovers
- “General equilibrium”



Lecture Overview

- What is Randomization
- Simple Randomization
- Unit of Randomization
- Methods for different evaluation questions
- Real world challenges & design solutions

The New York Times

Expect the World®

To Test Housing Program, Some Are Denied Aid

By Cara Buckley

December 8, 2010

Half of the test subjects — people who are behind on rent and in danger of being evicted — are being denied assistance from the program for two years, with researchers tracking them to see if they end up homeless.

“They should immediately stop this experiment. The city shouldn’t be making guinea pigs out of its most vulnerable”

-Scott Stringer (Manhattan Borough President)

“I don’t think homeless people in our time, or any time, should be treated like lab rats”

-Councilwoman Annabel Palma

To Test Housing Program, Some Are Denied Aid

By Cara Buckley

December 8, 2010

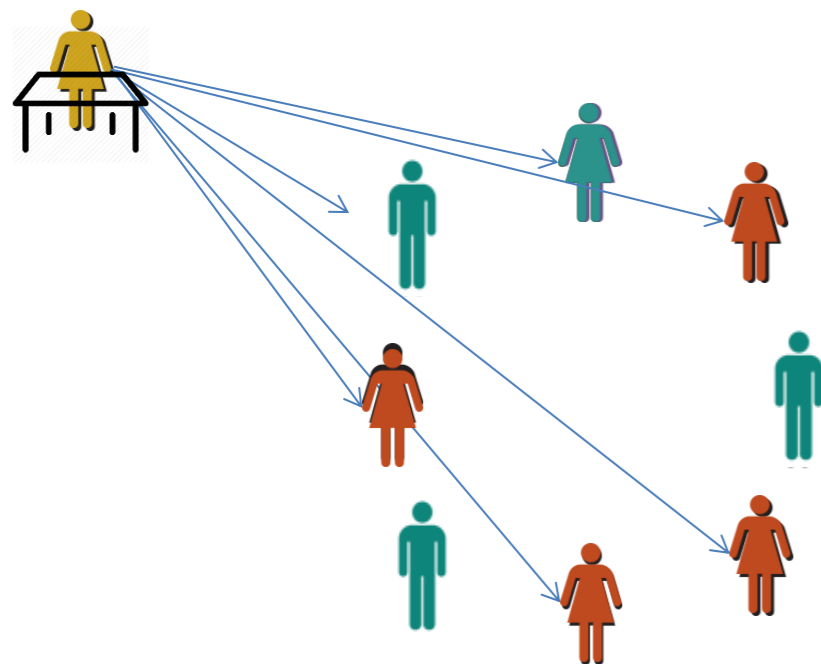
“It’s a very effective way to find out what works and what doesn’t, everybody, every country, has a limited budget and wants to find out what programs are effective.”

-Esther Duflo, Economist at M.I.T.

The firm’s institutional review board concluded that the study was ethical for several reasons, said Mary Maguire, a spokeswoman for Abt: because it was not an entitlement, meaning it was not available to everyone; because it could not serve all of the people who applied for it; and because the control group had access to other services.

C1: Identifying Control and Treatment

- Service providers have trouble distinguishing between treatment and comparison (or customizing service)



■ treatment

■ comparison



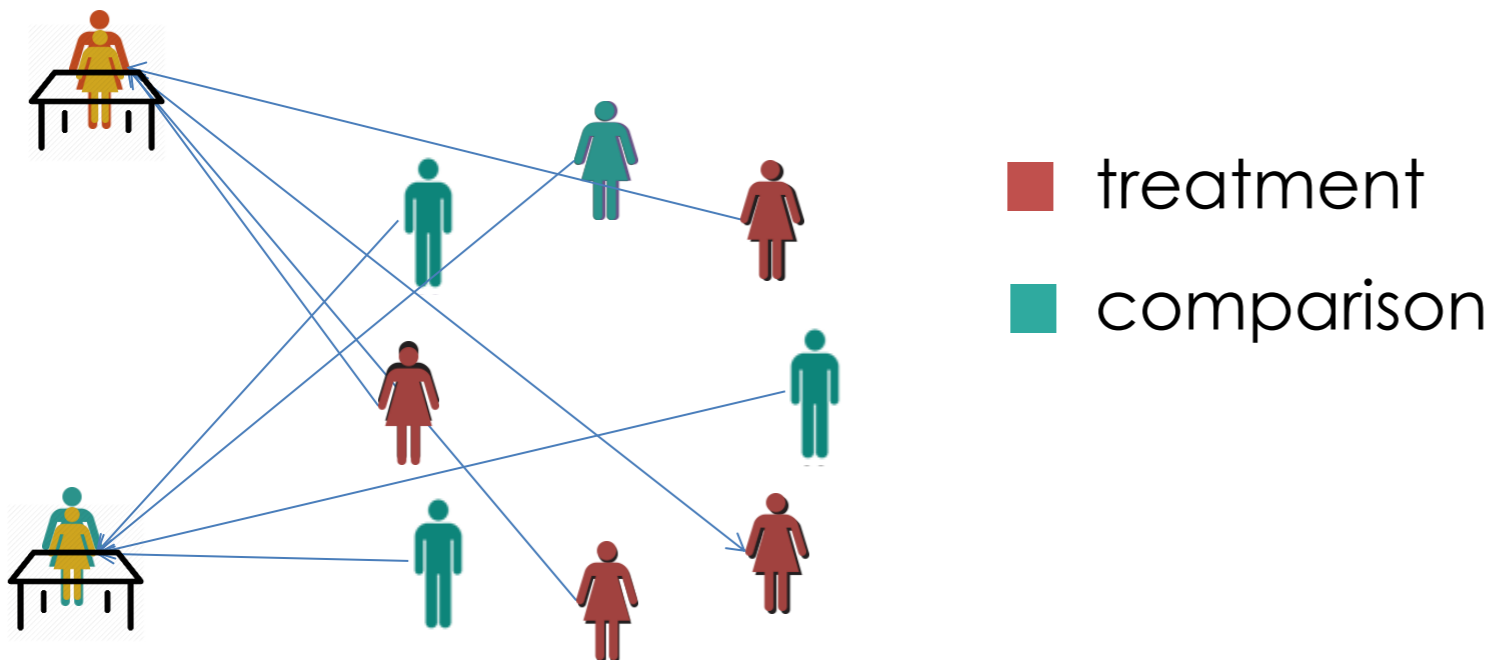
Services provided to both



Crossovers: Control receives intervention
(No longer represents pure counterfactual)

S1: Assign to Different Service Providers

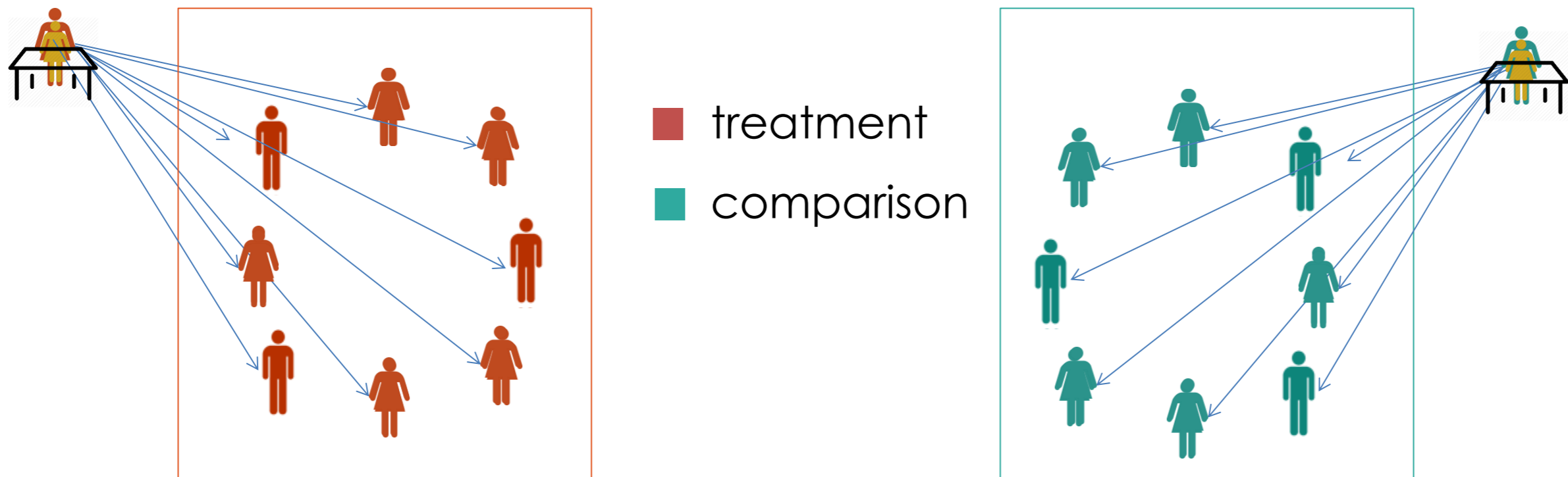
- Service providers have trouble distinguishing between treatment and comparison (or customizing service)



- Have different teams provide the different treatments
- Randomly assign to those teams

S1b: Randomize at a different unit

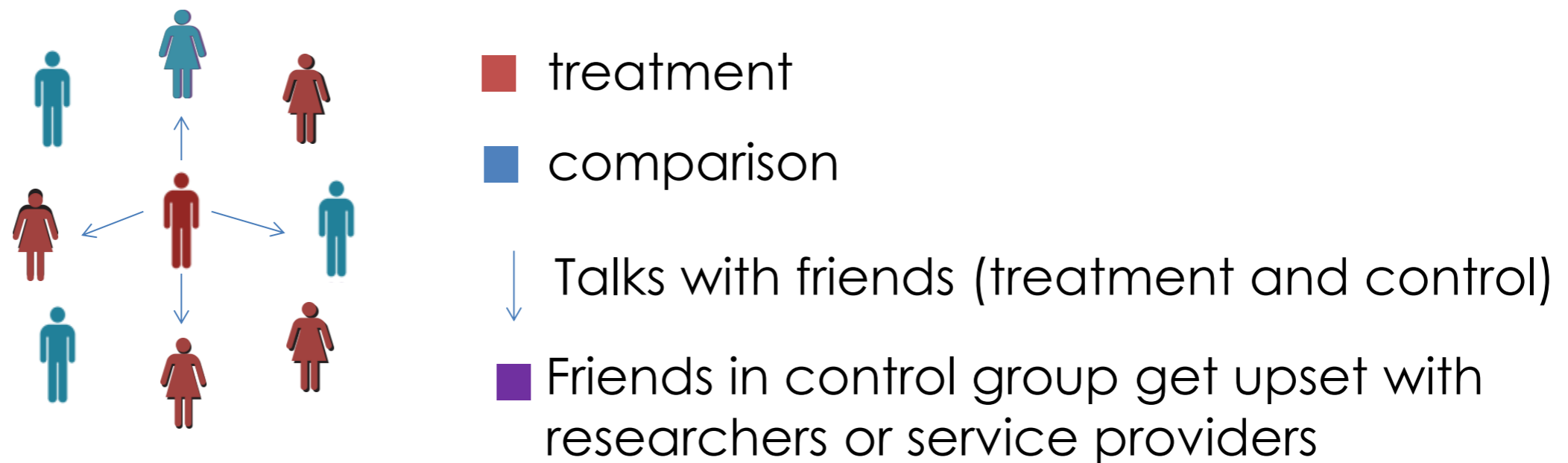
- Service providers have trouble distinguishing between treatment and comparison (or customizing service)



- Change the unit of random assignment
- Have providers treat entire clusters the same

C2a: Control group finds out about treatment

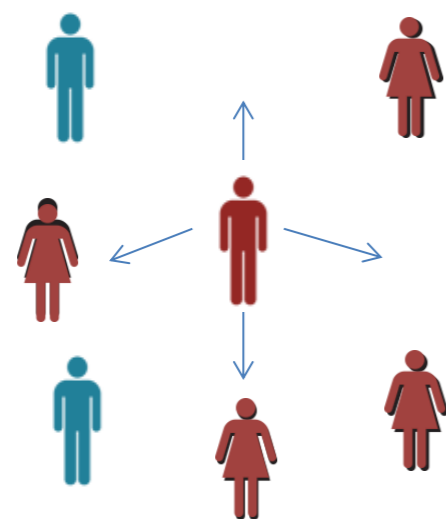
- If treatment and control individuals know each other, the control may get upset.



- Service providers may lose support of community

C2a: Control group finds out about treatment

- If treatment and control individuals know each other, the control may get upset.



■ treatment

■ comparison

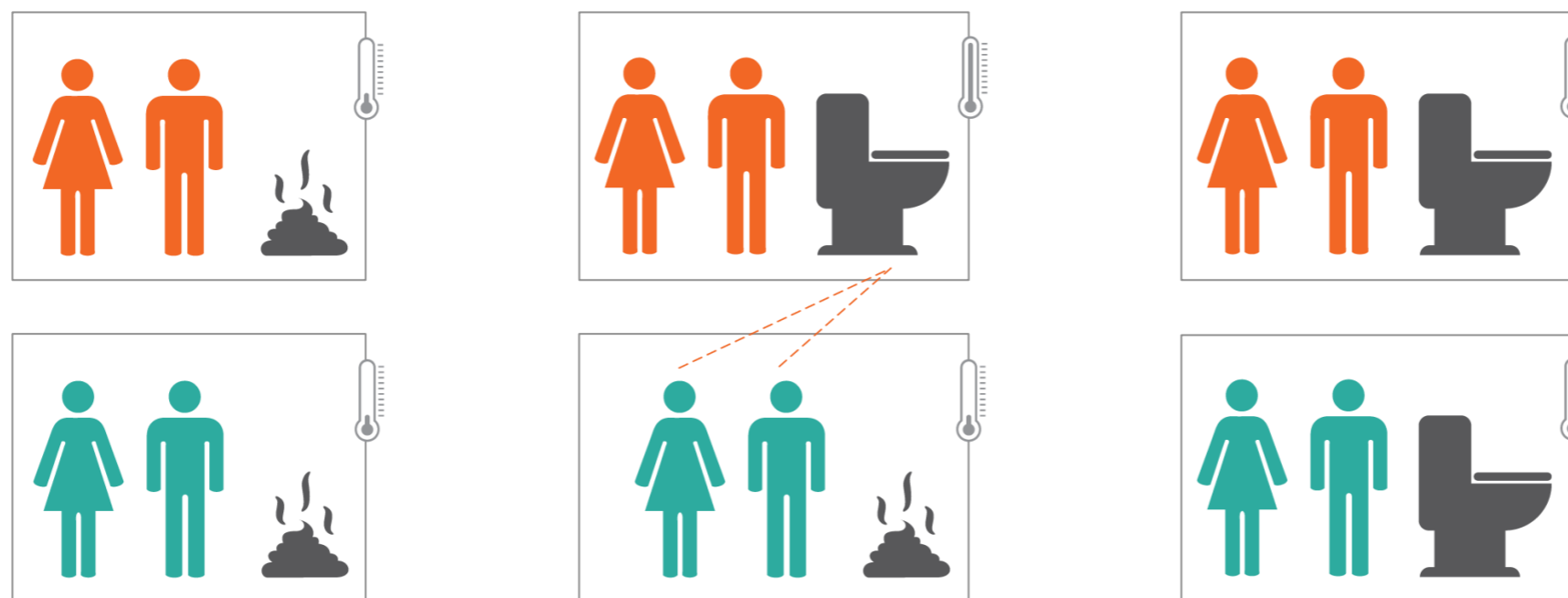
↓ Talks with friends (treatment and control)

■ Friends in control group get upset with researchers or service providers

- Service providers may lose support of community
- Attrition: Control withdraws participation from research

C2b: Control group benefits from treatment

- May change their behavior after seeing treatment

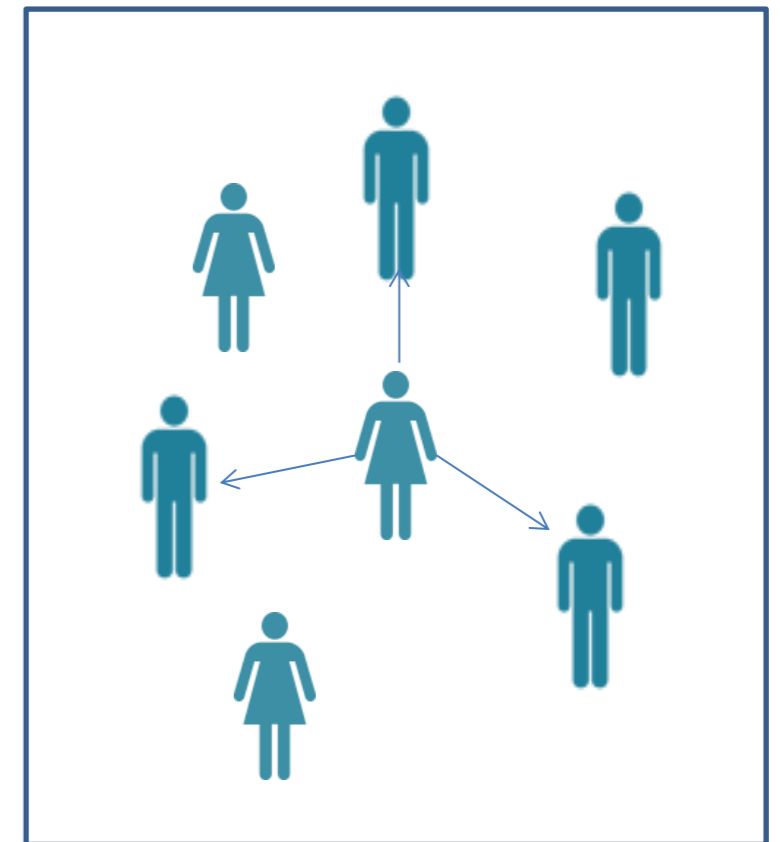
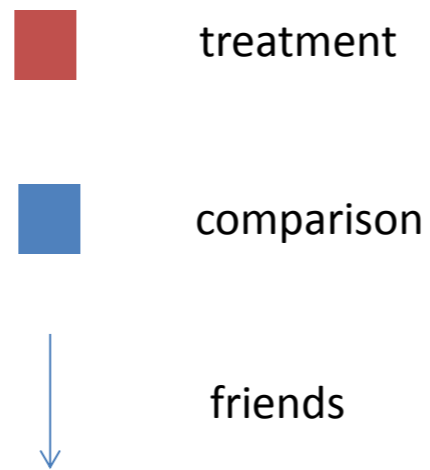
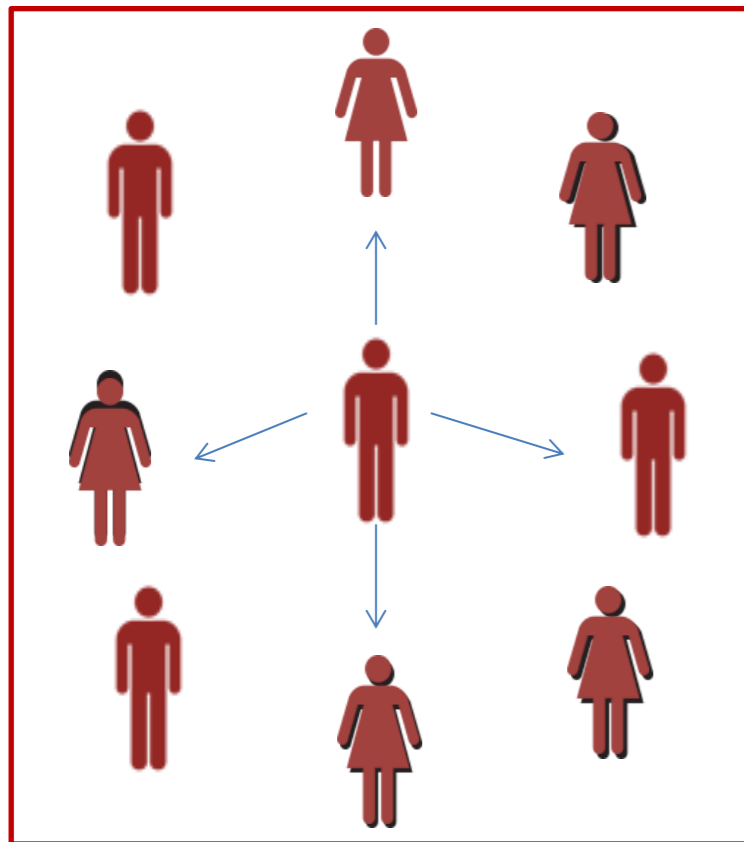


True impact = 5

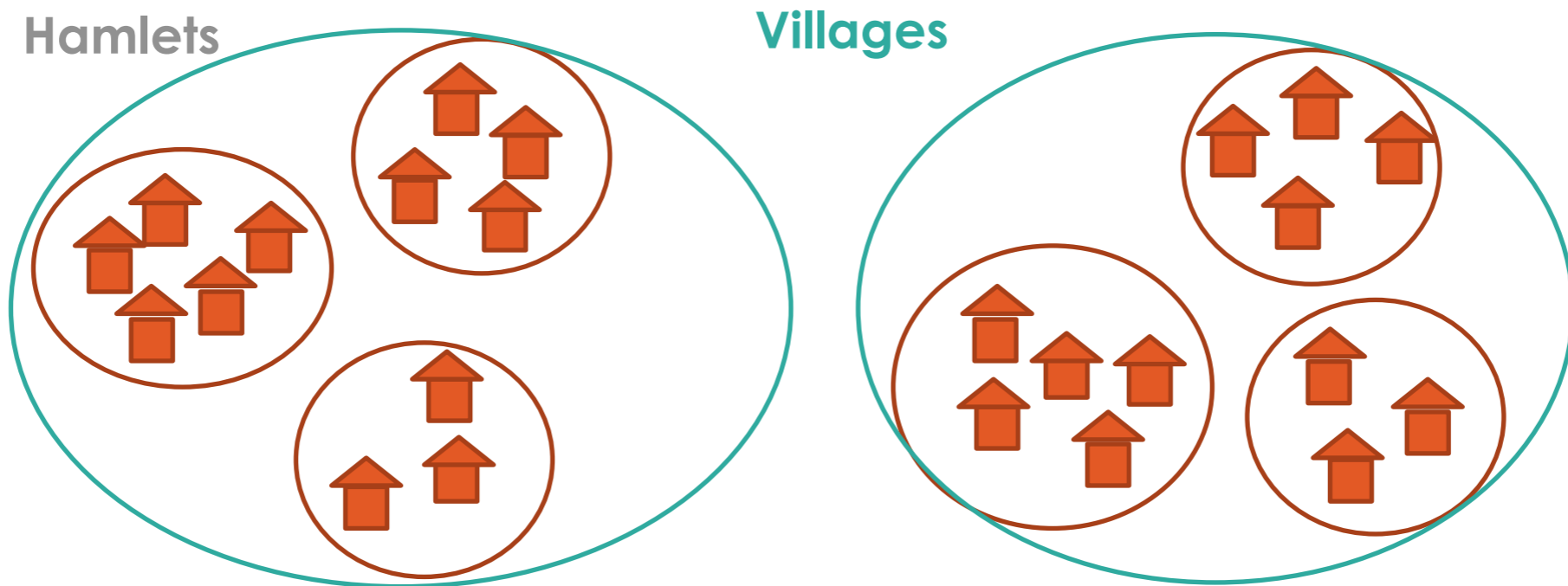
Measured impact = 0

■ Treatment group ■ Control group 🌡️ Bad health 🌡️ Good health

S2a: Varying the unit to contain spillovers



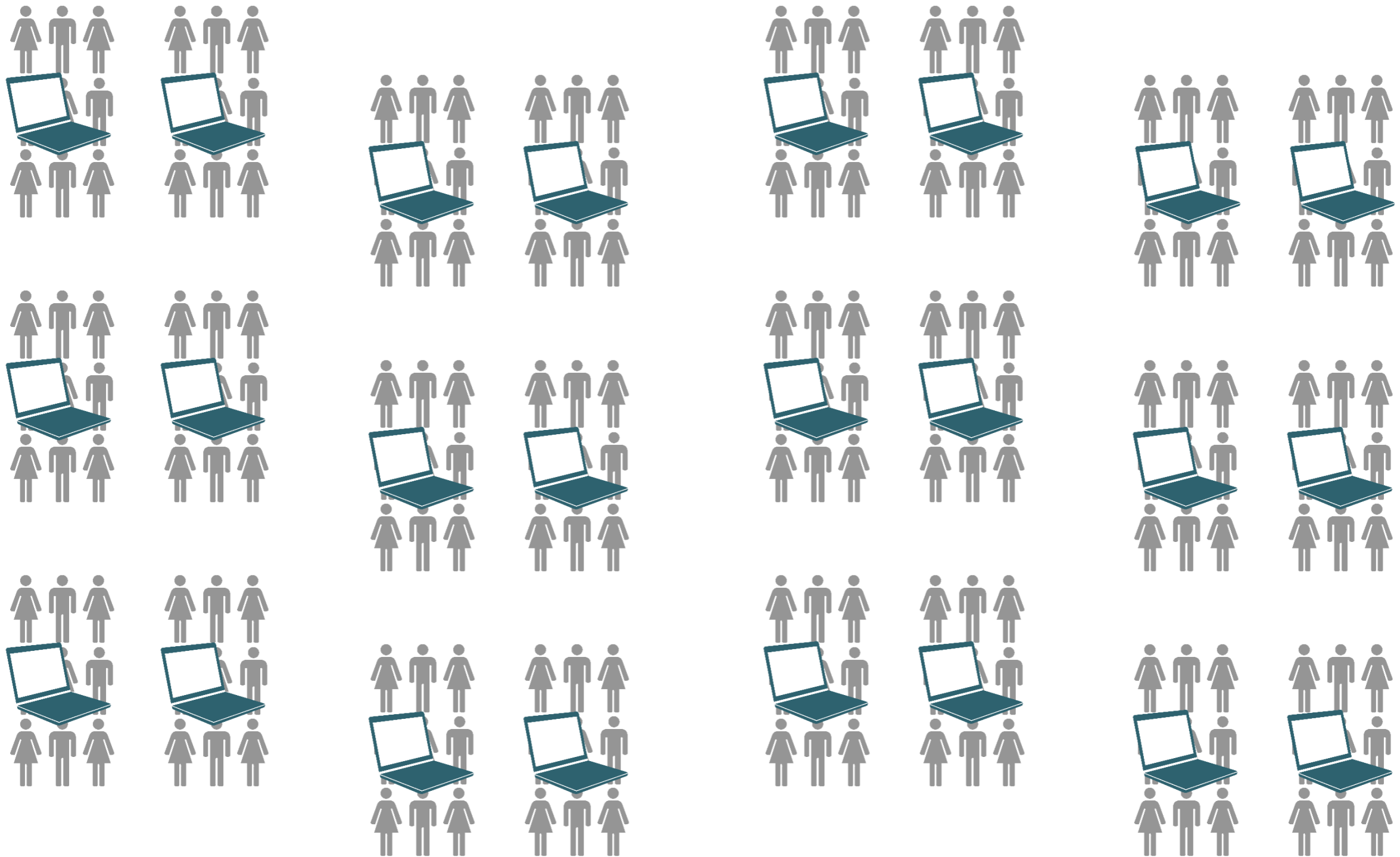
Unit of Randomization



S2b: Creating a Buffer

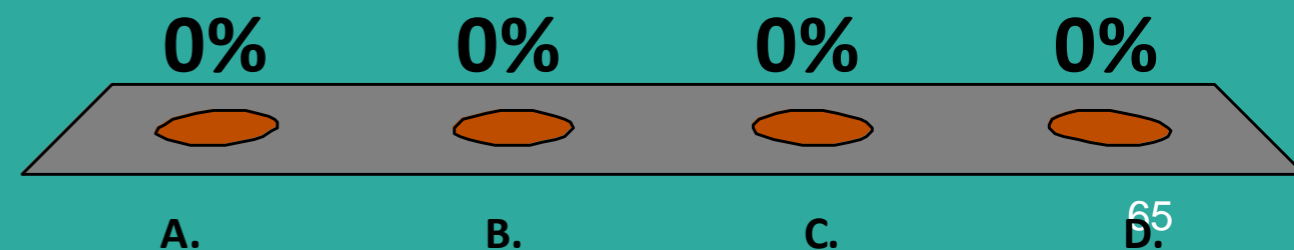


C3: Have resources to treat everyone. (Where's the control group?)



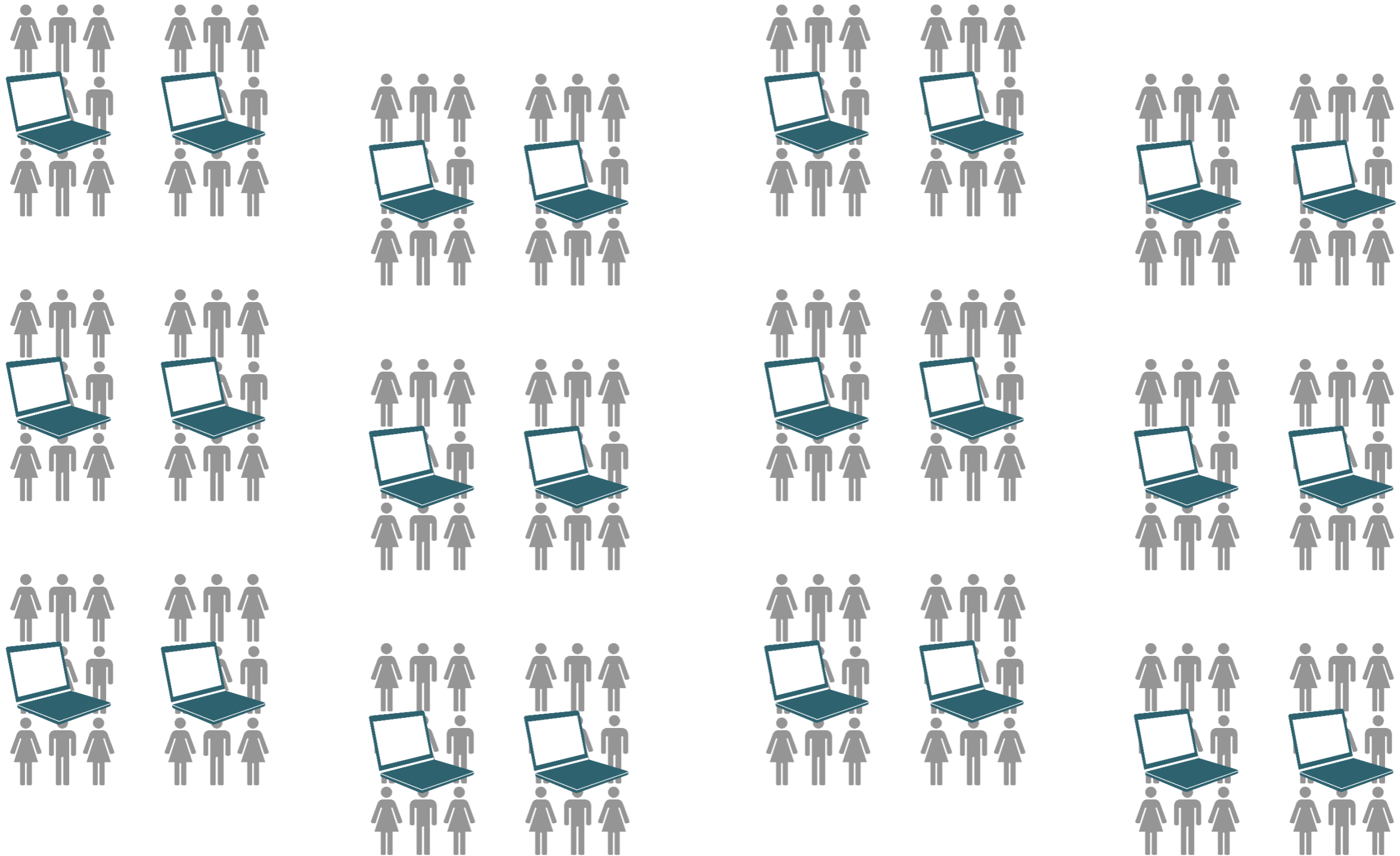
If you had enough resources to provide everyone what would you do?

- A. It's an important question, let's run the experiment anyway
- B. Give the control group an alternate treatment
- C. Scrap the experiment and look elsewhere
- D. Something else



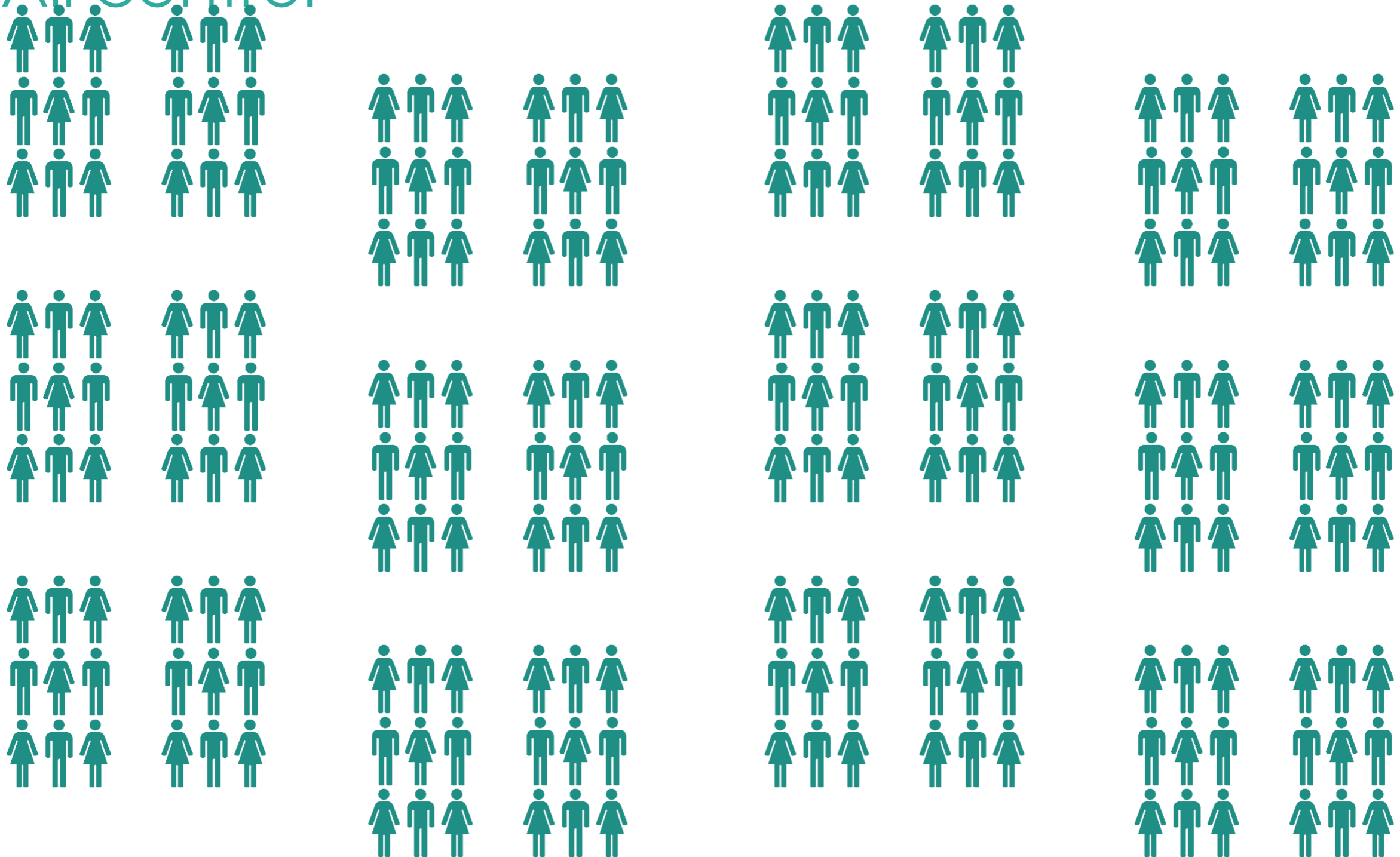


S3: Phase In



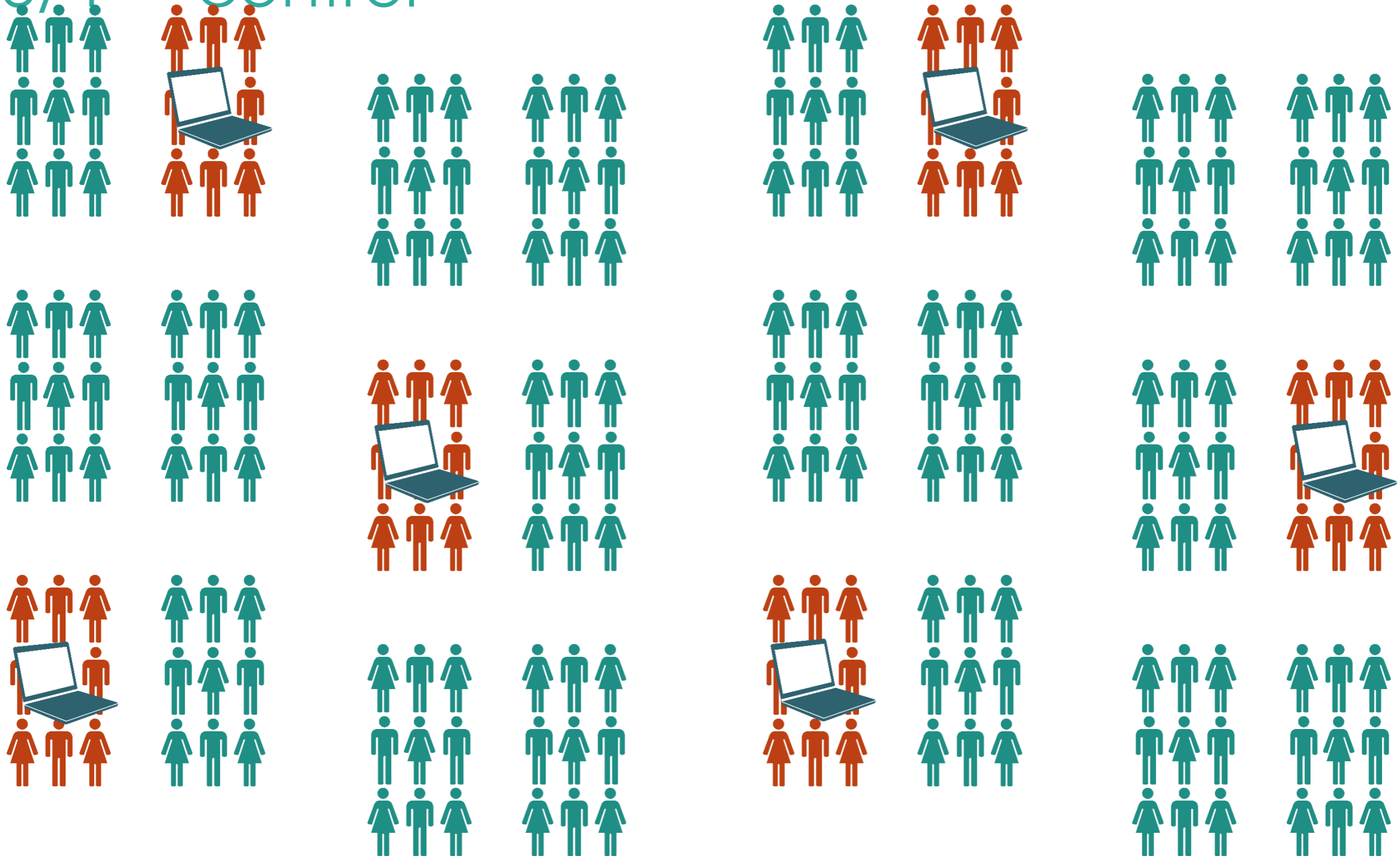
Phase 0: No one treated yet

All control



Phase 1:

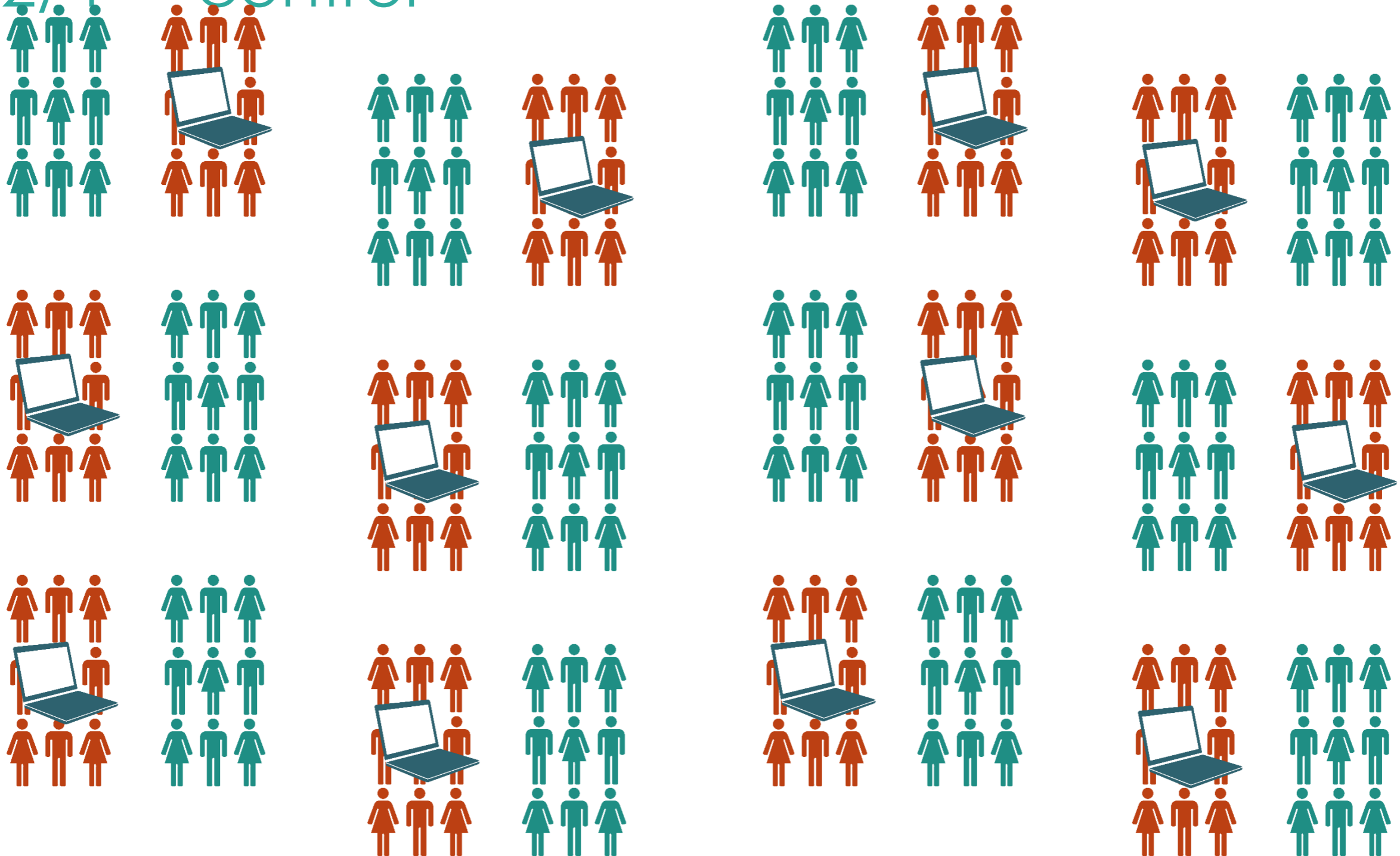
$1/4^{\text{th}}$ treated
 $3/4^{\text{th}}$ control



Phase 2:

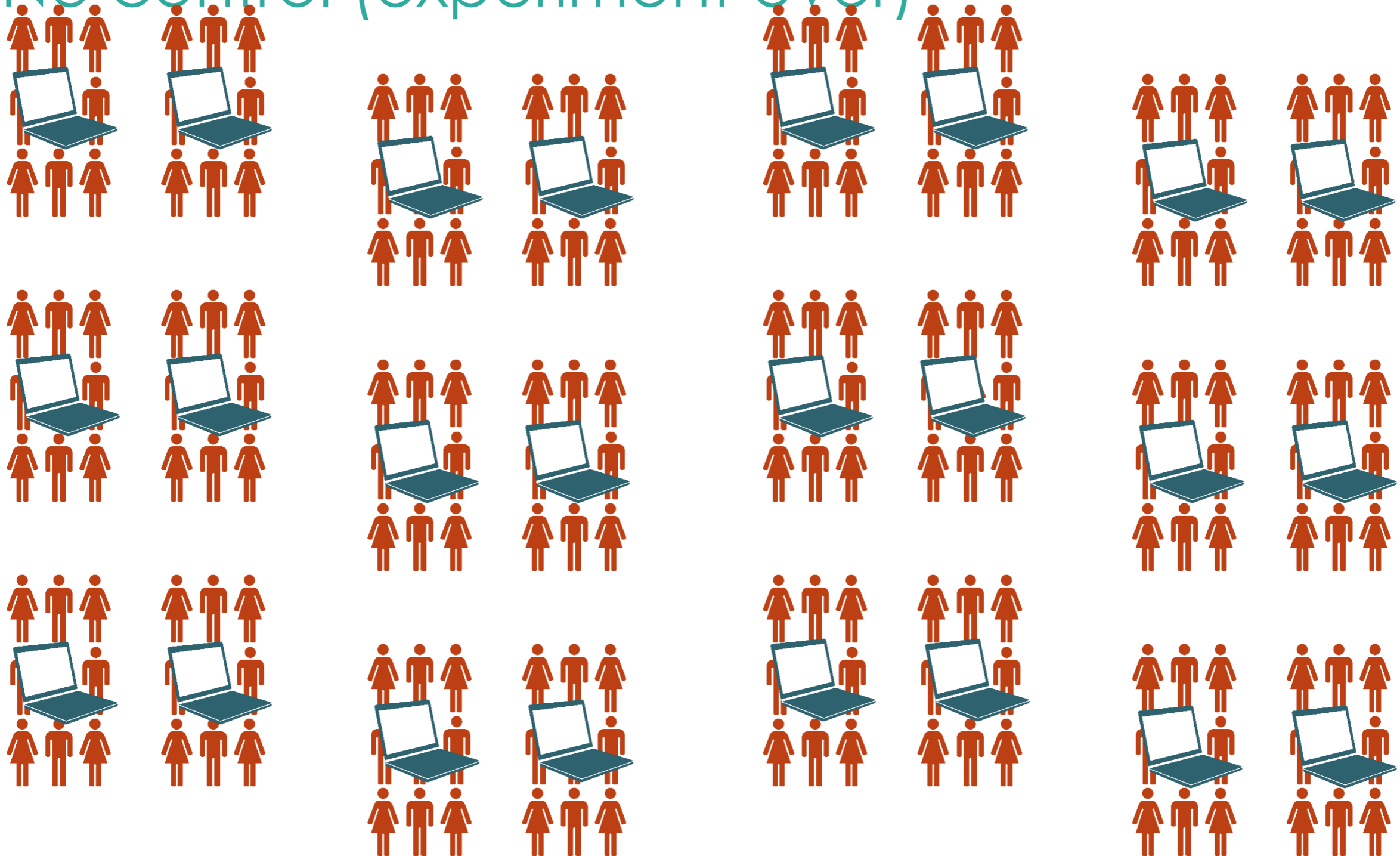
2/4^{ths} treated

2/4^{ths} control



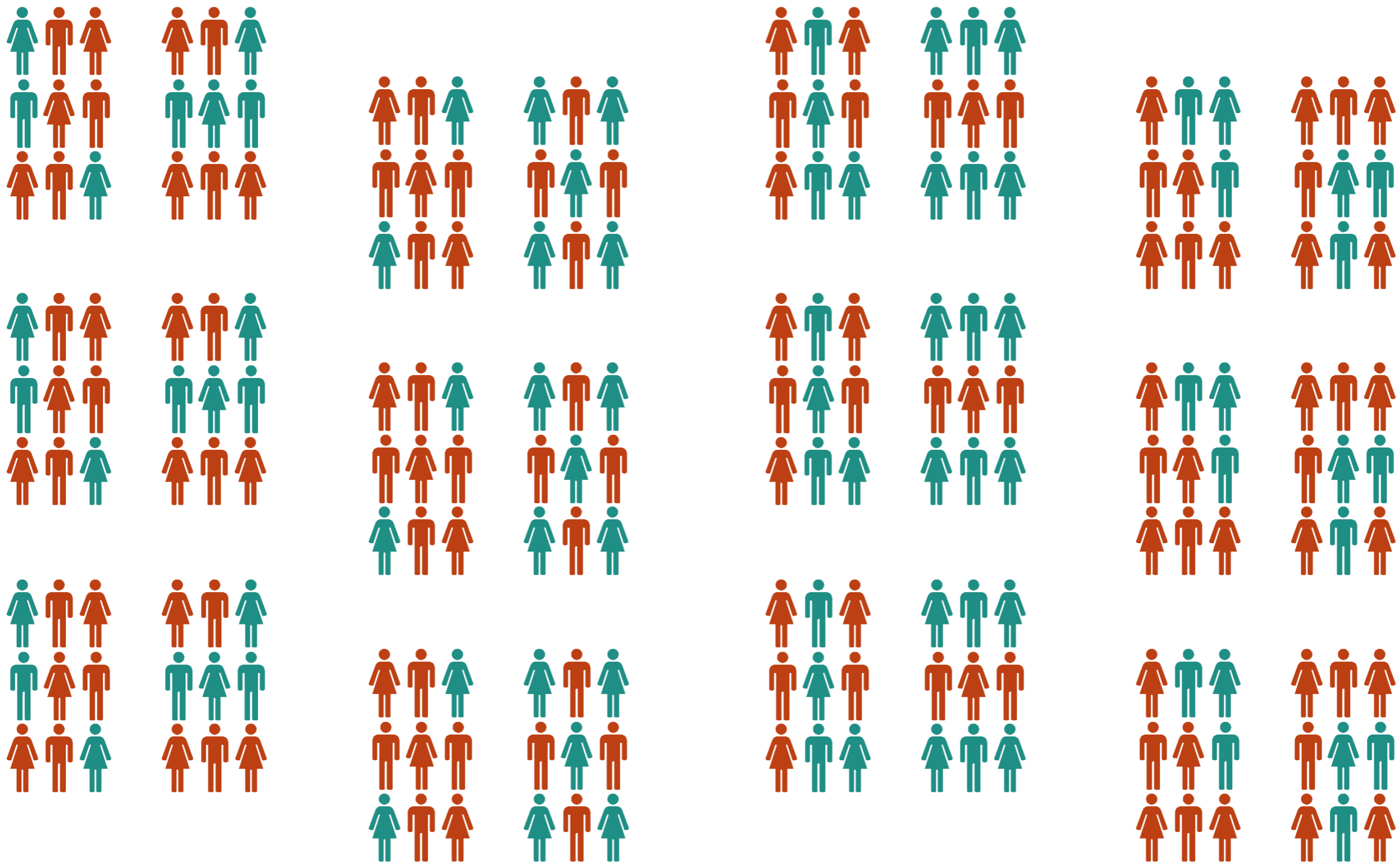
Phase 4: All treated

No control (experiment over)



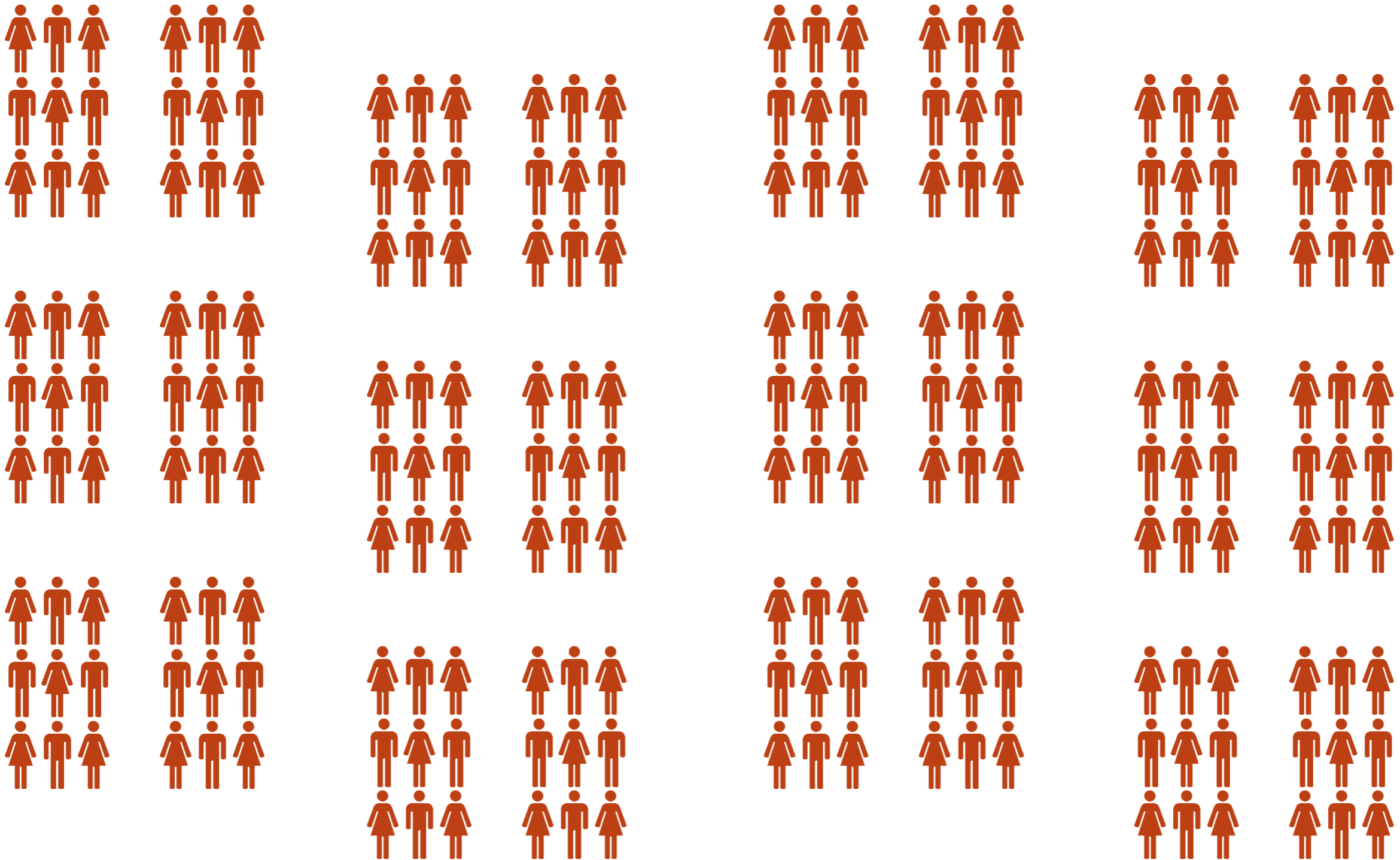
Phase-in: Can be any level, any # of phases

Phase 1: Half get treatment, half control

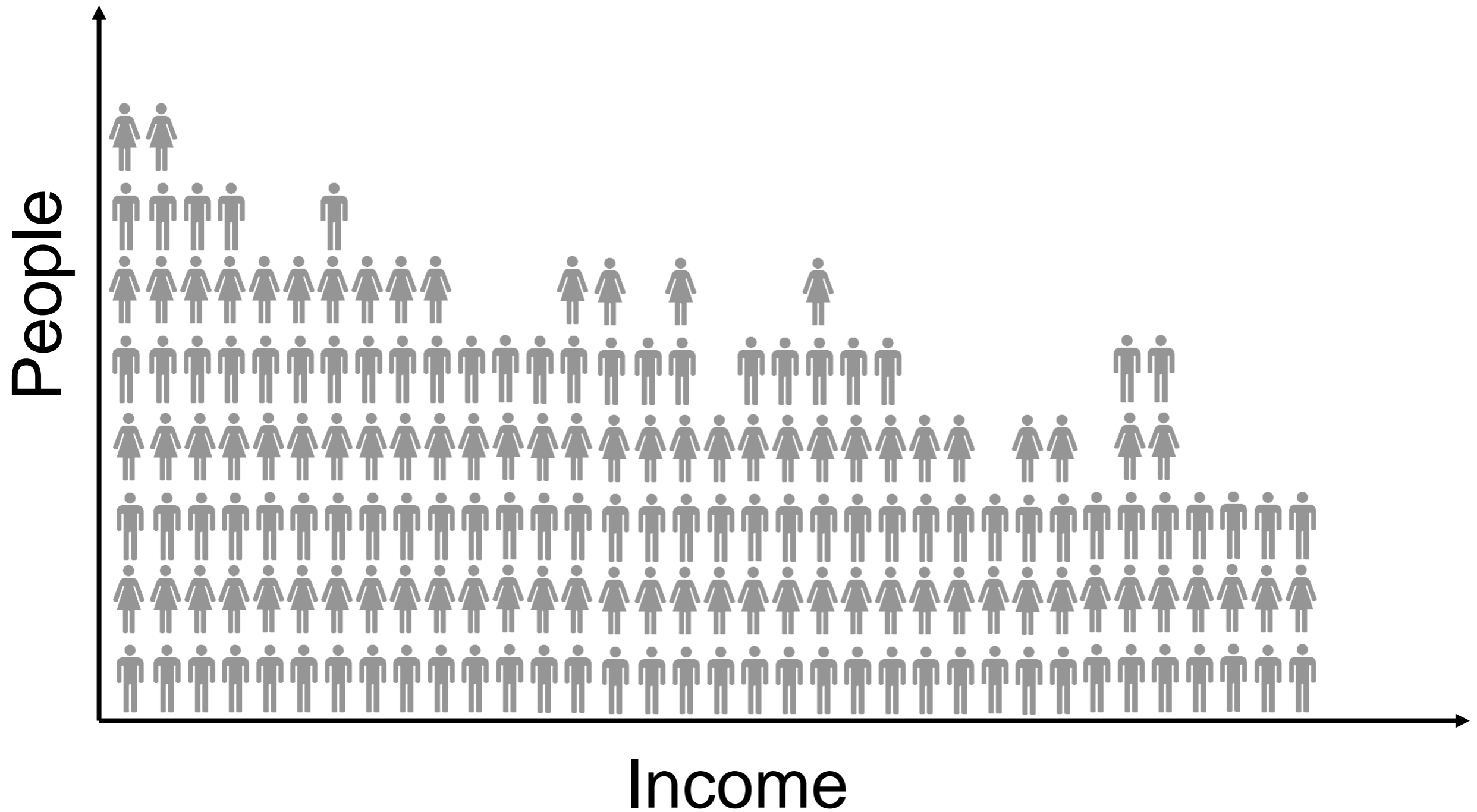


Phase-in: Can be any level, any # of phases

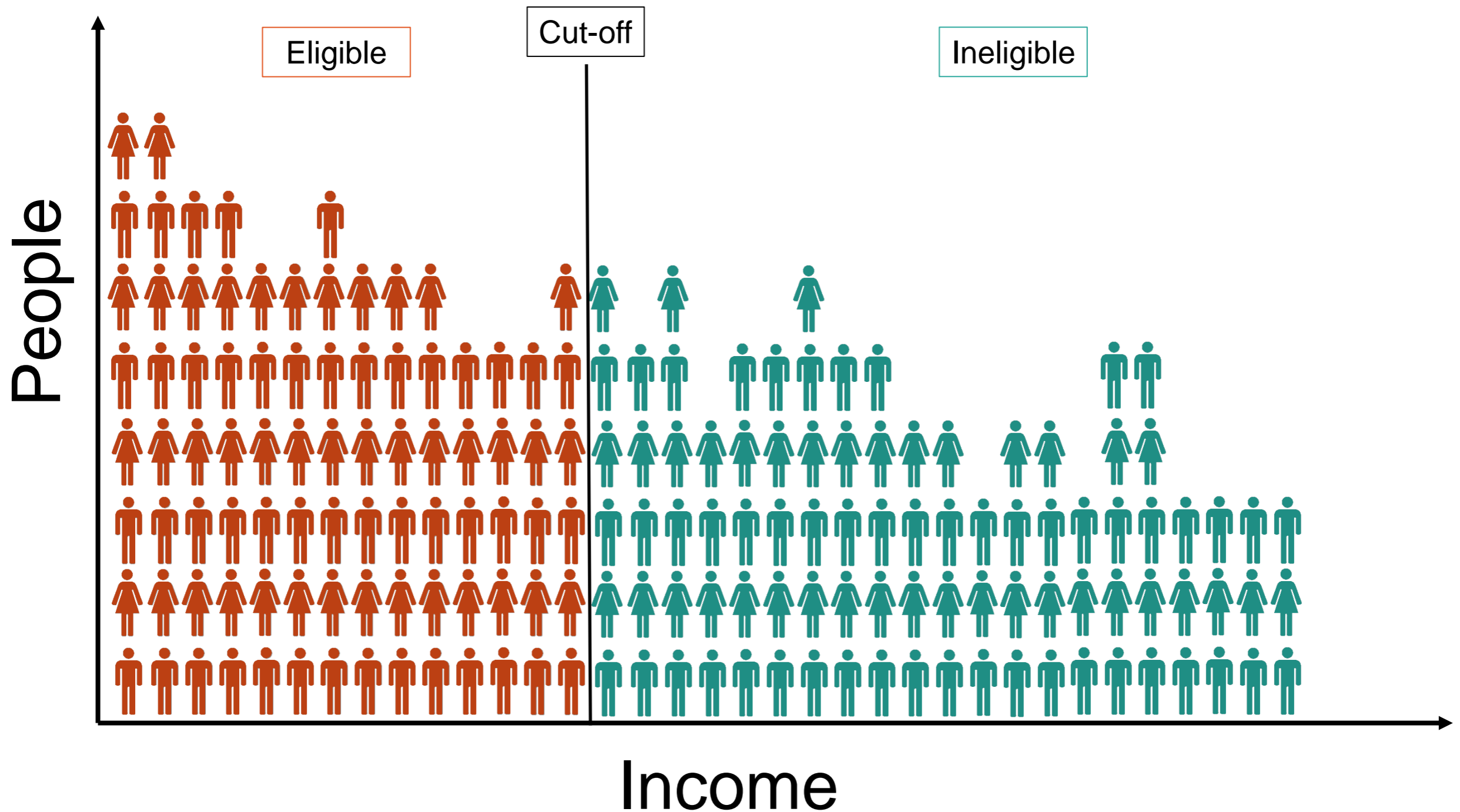
Phase 2: Everyone is treated



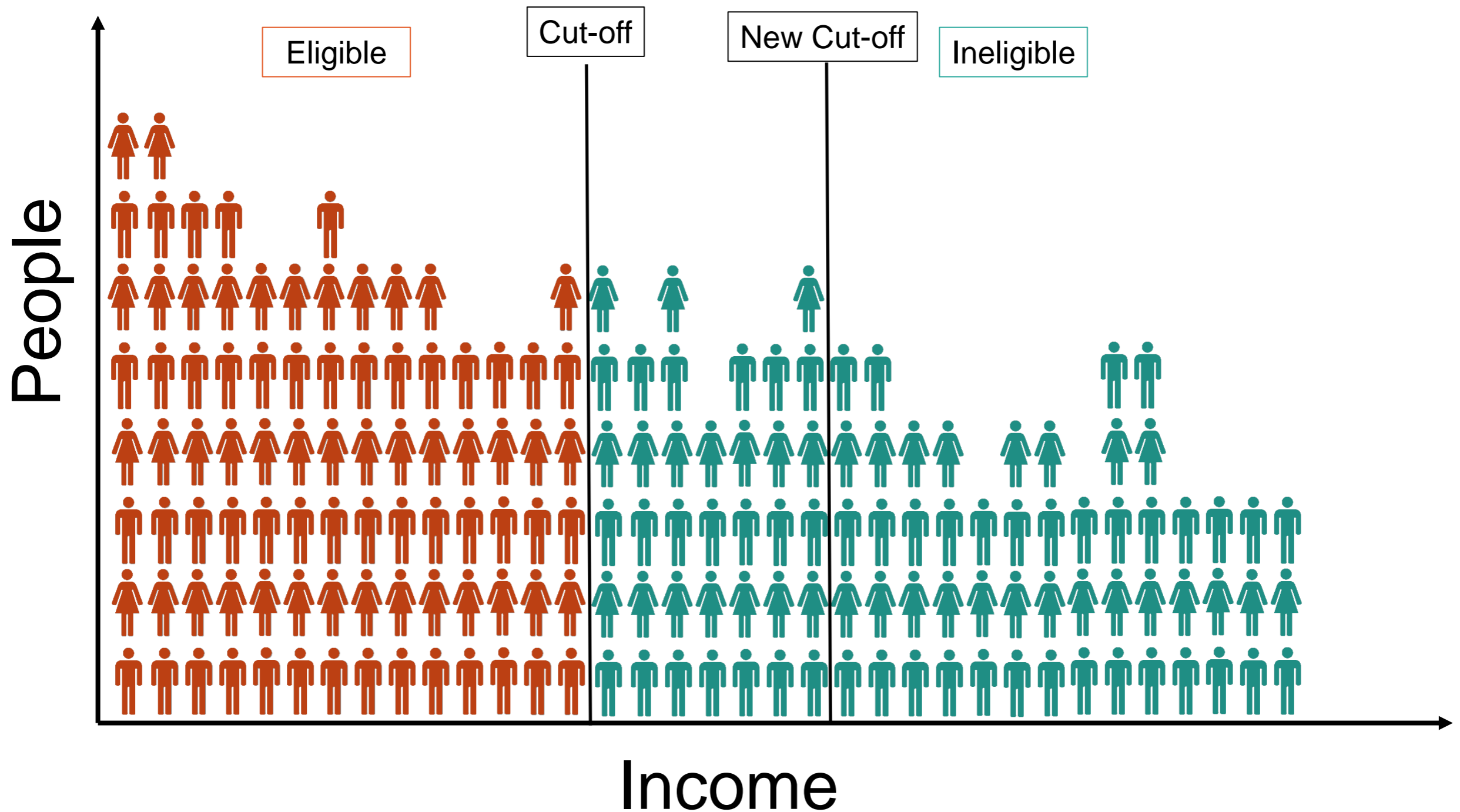
C4: There's an eligibility criteria



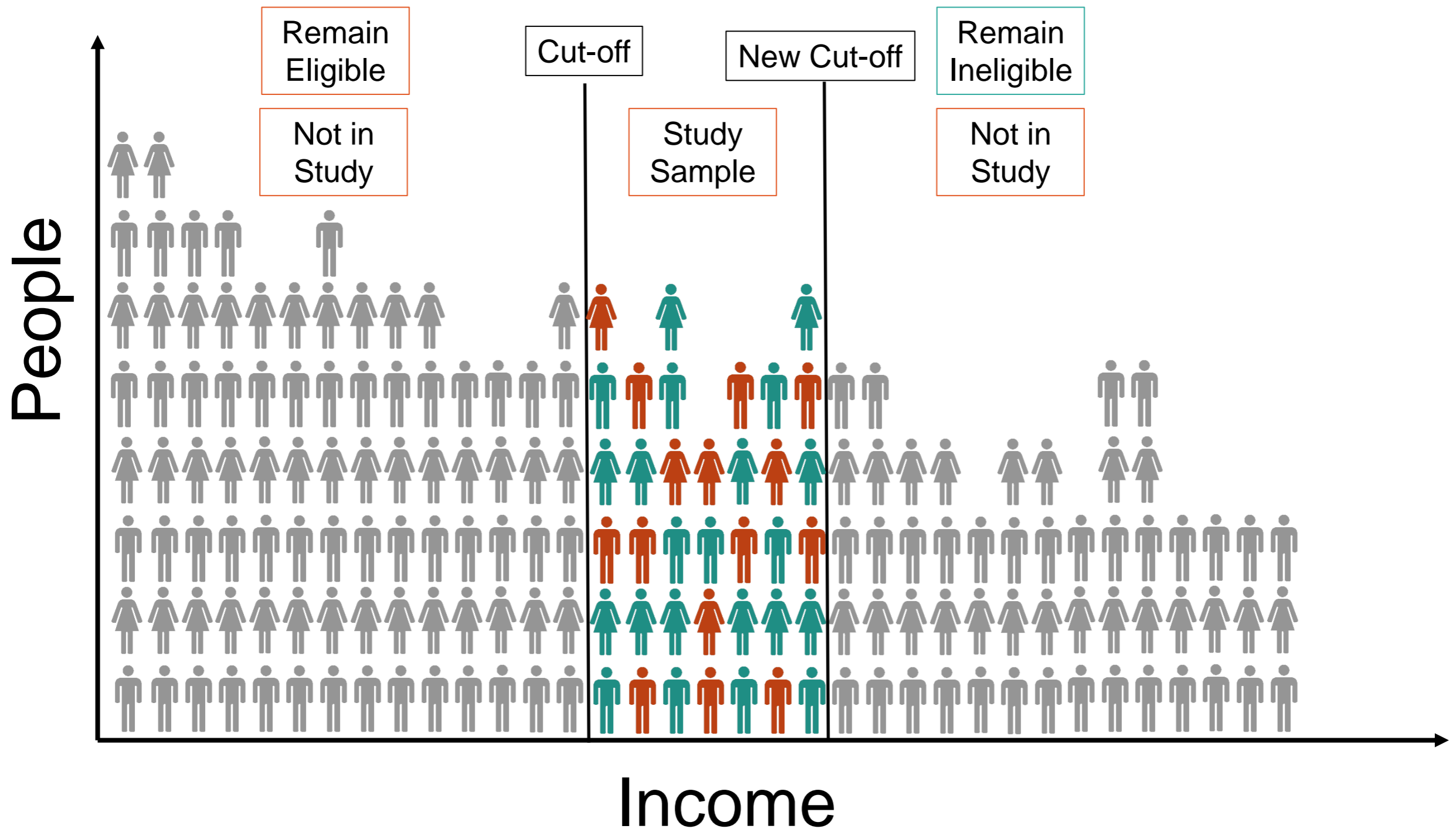
C4: There's an eligibility criteria



S4: Relax the eligibility criteria



S4: Randomize “on the bubble”

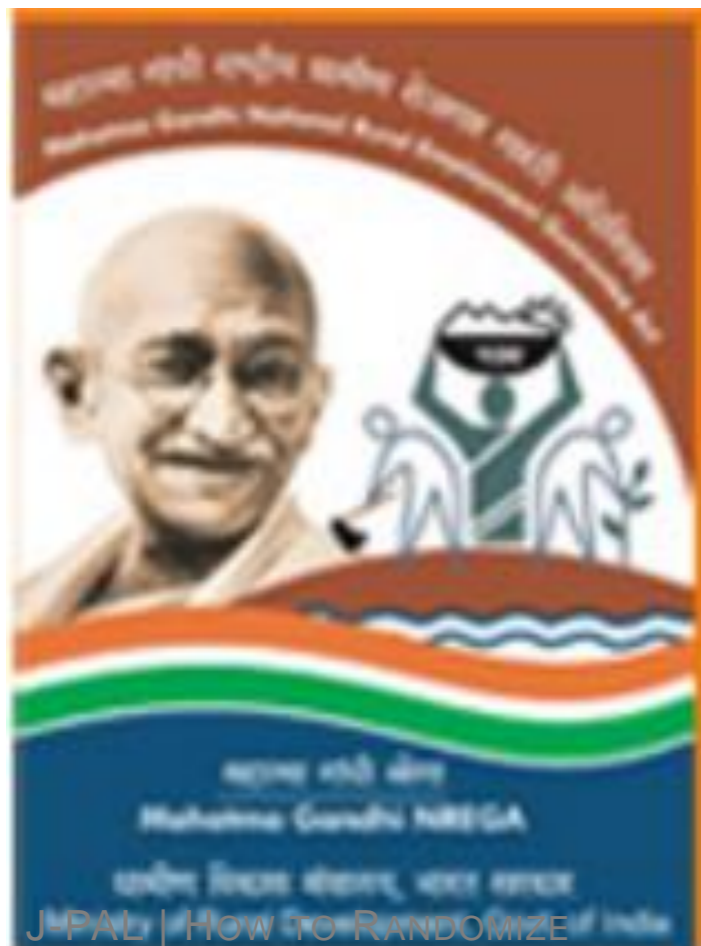


C5: Program is an entitlement Cannot force nor deny intervention

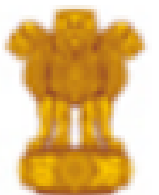


The Mahatma Gandhi National Rural Employment Guarantee Act 2005

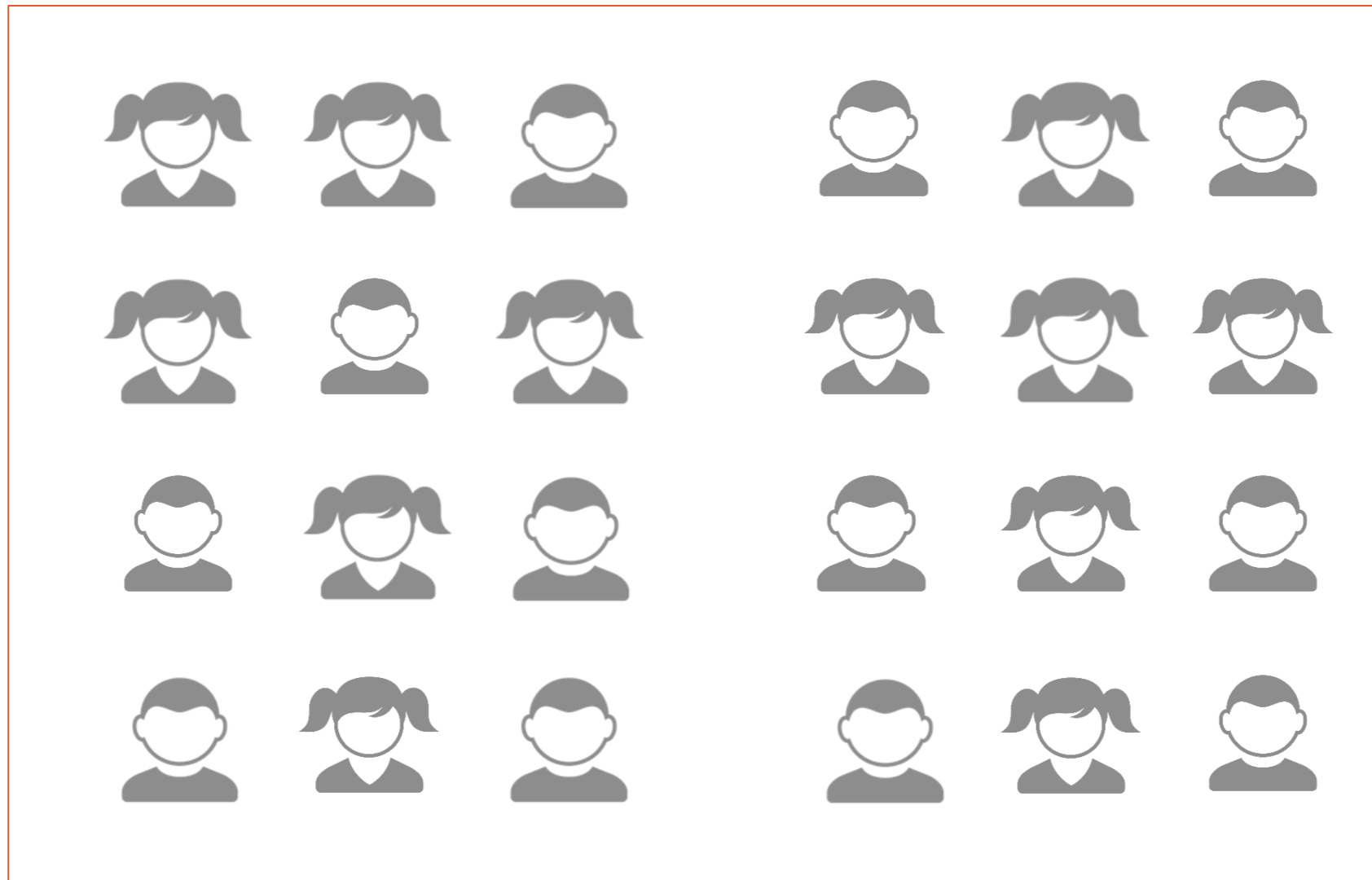
Ministry of Rural Development
Government of India



Ministry of Rural Development
Government of India



C5: Program is an entitlement

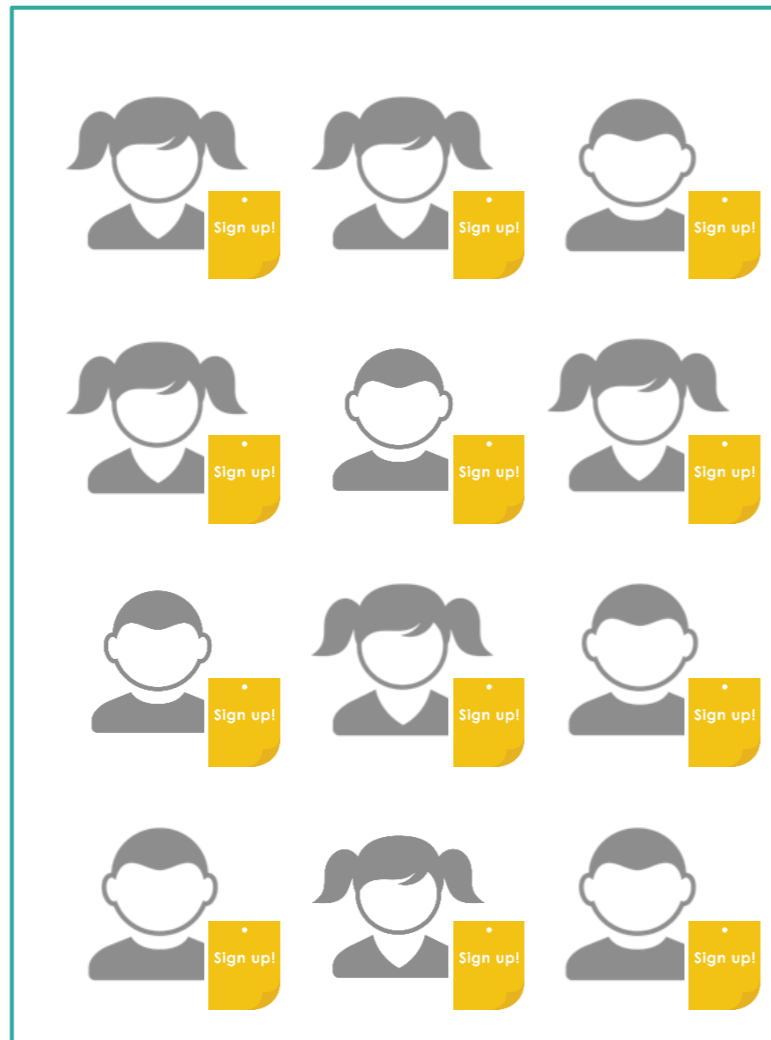


S5: Encouragement



Treatment Group

Control Group



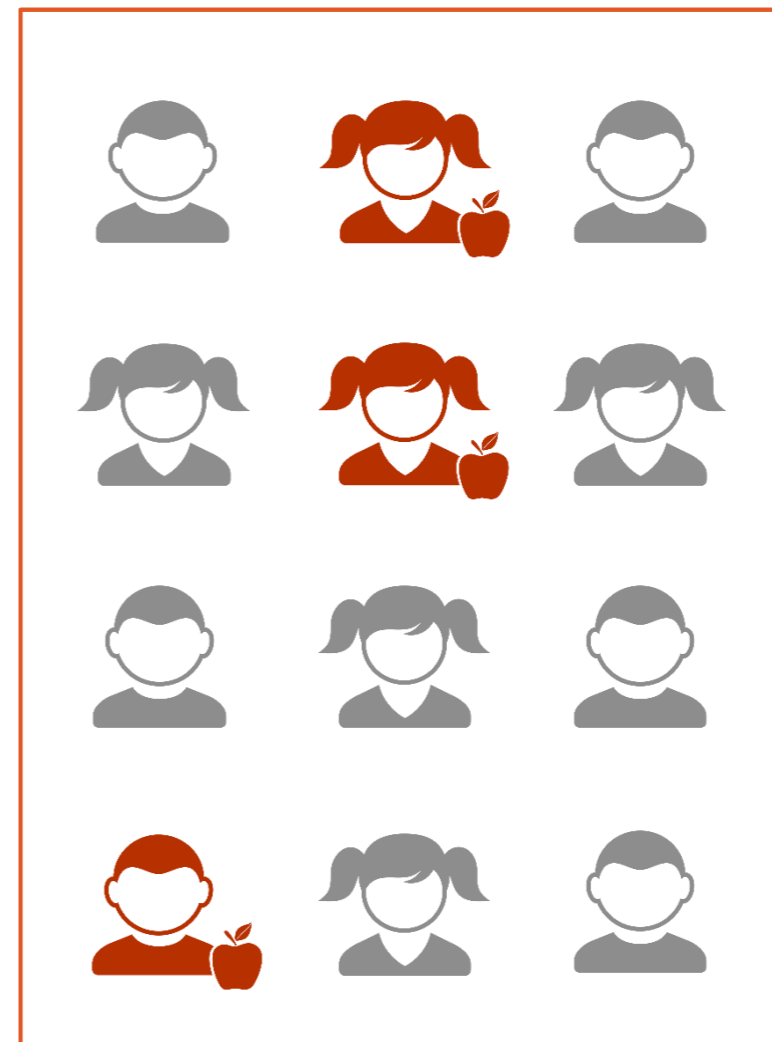
S5: Encouragement

Treatment Group



$3/4^{\text{th}}$ take-up

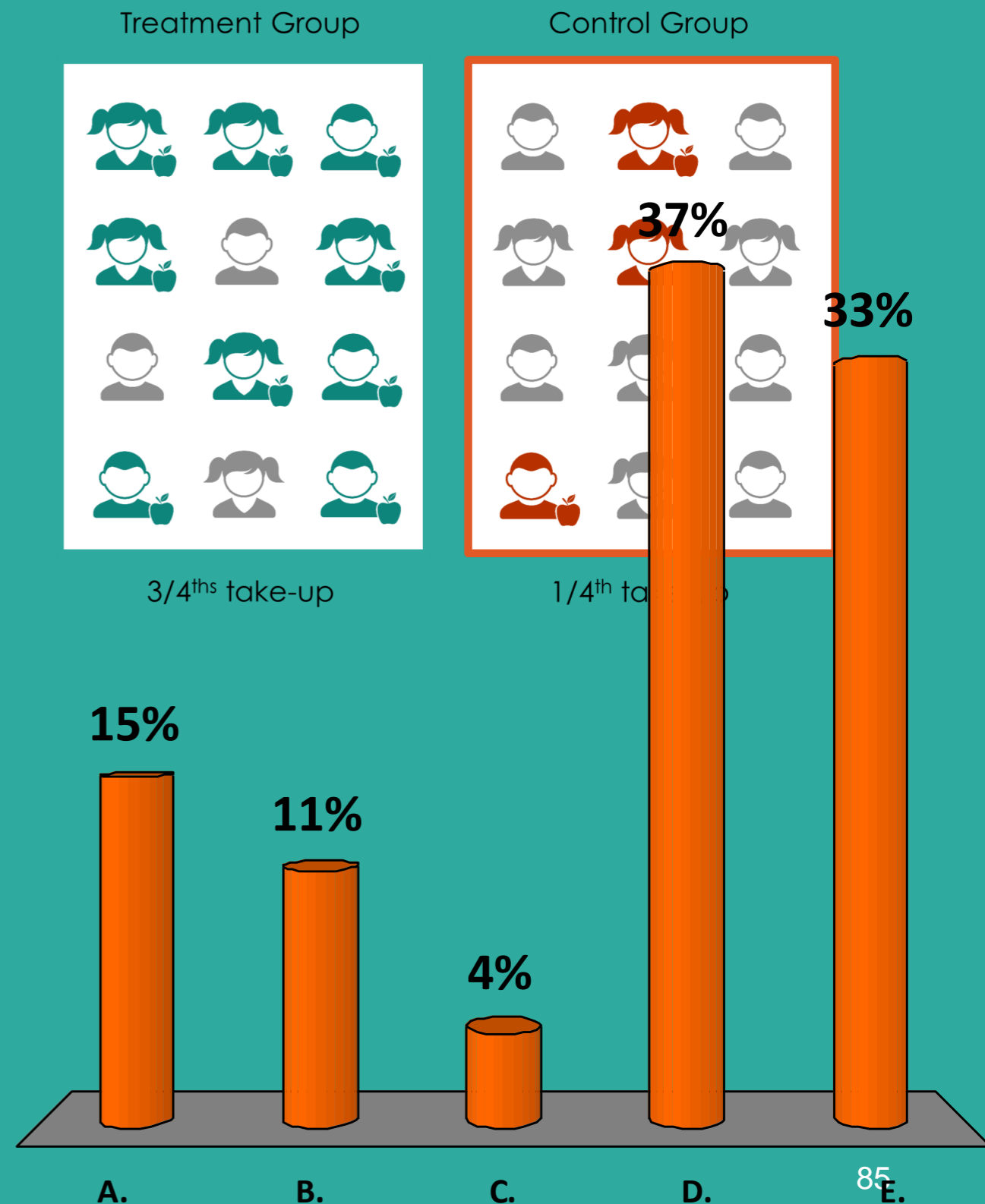
Control Group



$1/4^{\text{th}}$ take-up

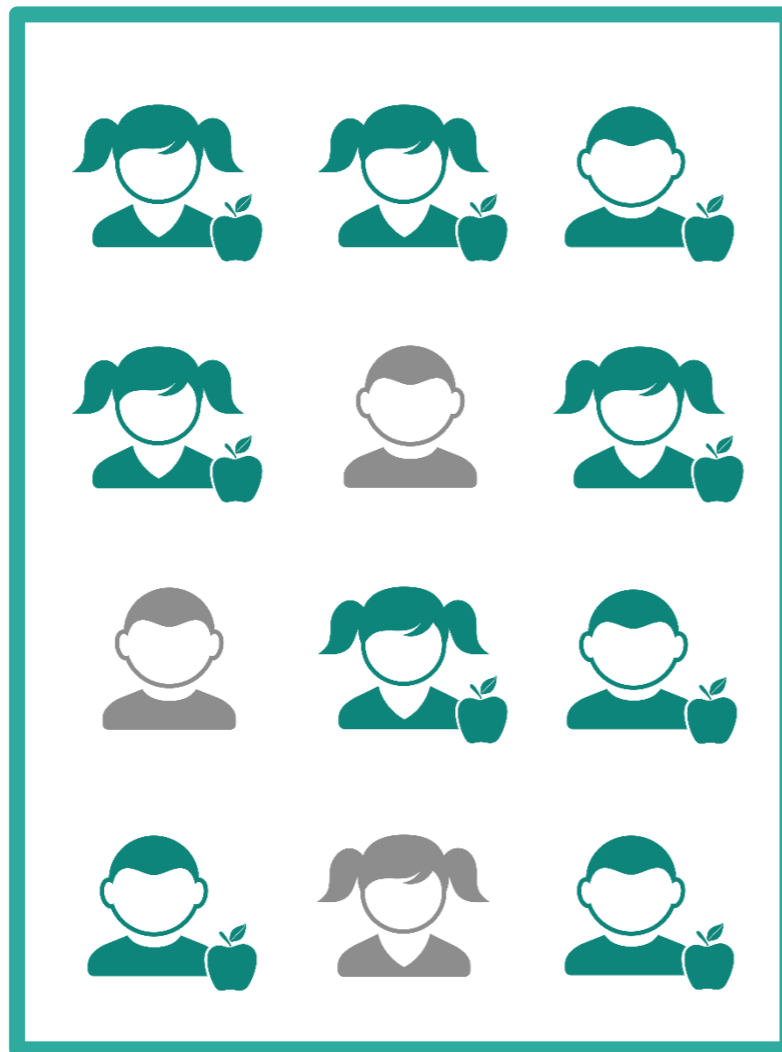
To evaluate the effect of this program, you would first:

- A. Compare those who enrolled to those who didn't
- B. Drop those who didn't enroll from the treatment group
- C. Drop those who did enroll from the control group
- D. Both B&C
- E. Compare treatment group to entire control group



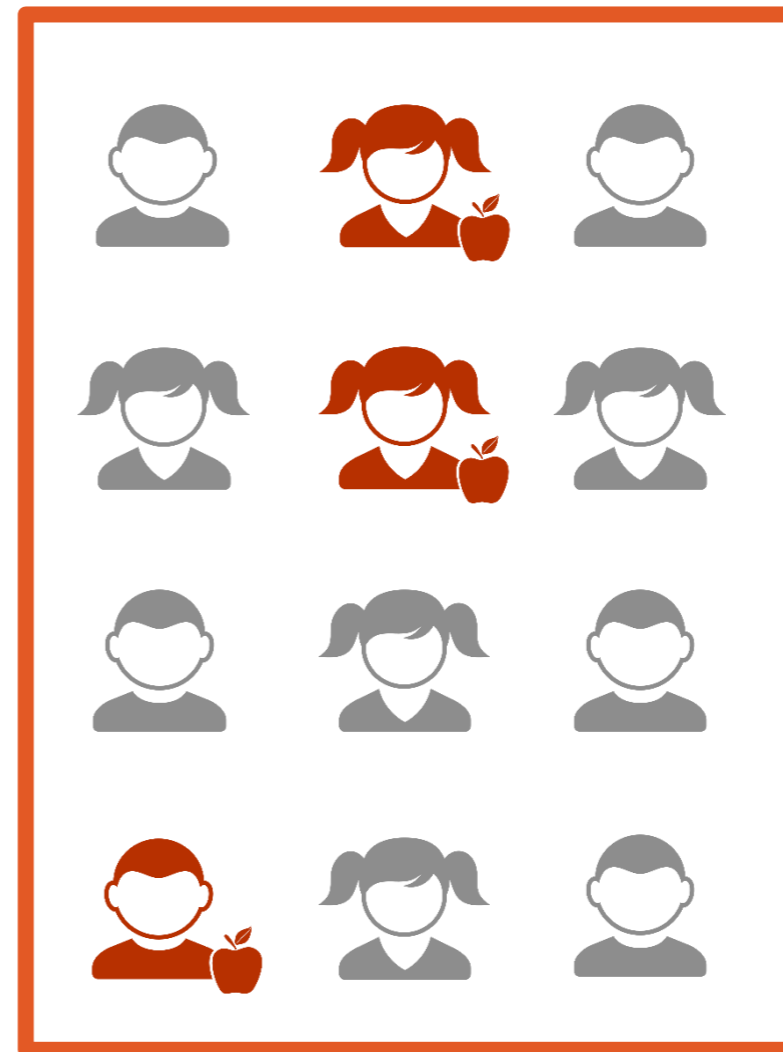
S5: Encouragement

Treatment Group



3/4^{ths} take-up

Control Group



1/4th take-up

Compare

Entire Treatment Group

to

Entire Control Group

C6: Sample size is small



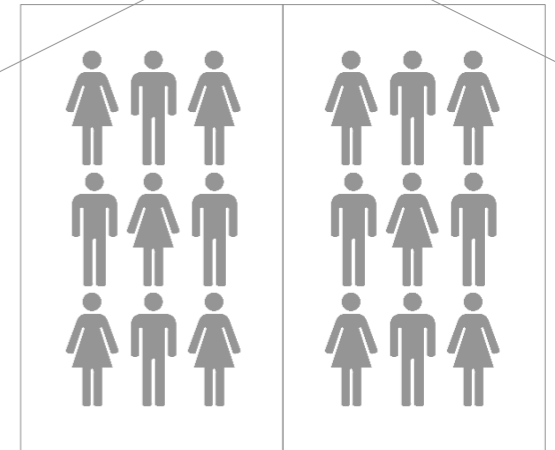
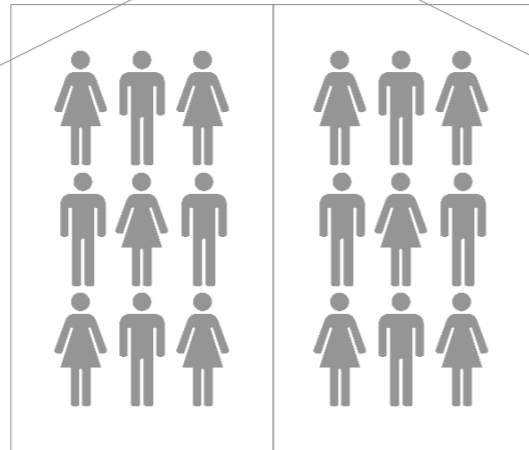
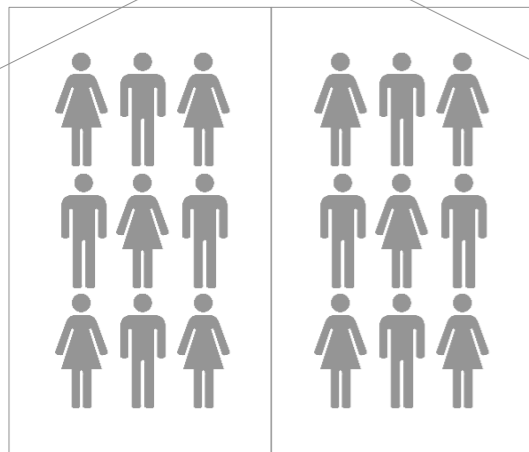
S6a: Change the unit of randomization

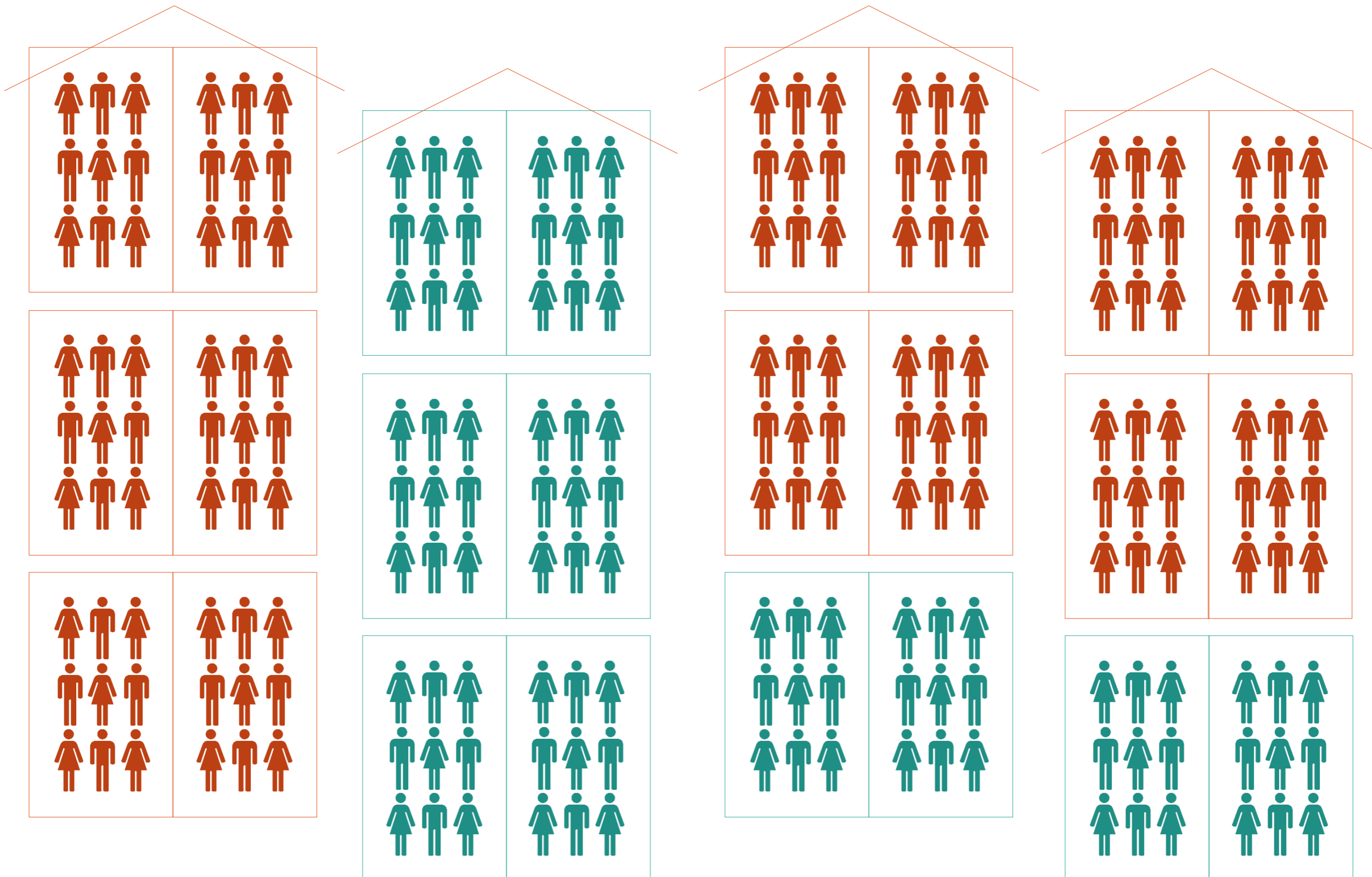


S6a: Change the unit of randomization



Stratification: Why Stratify?

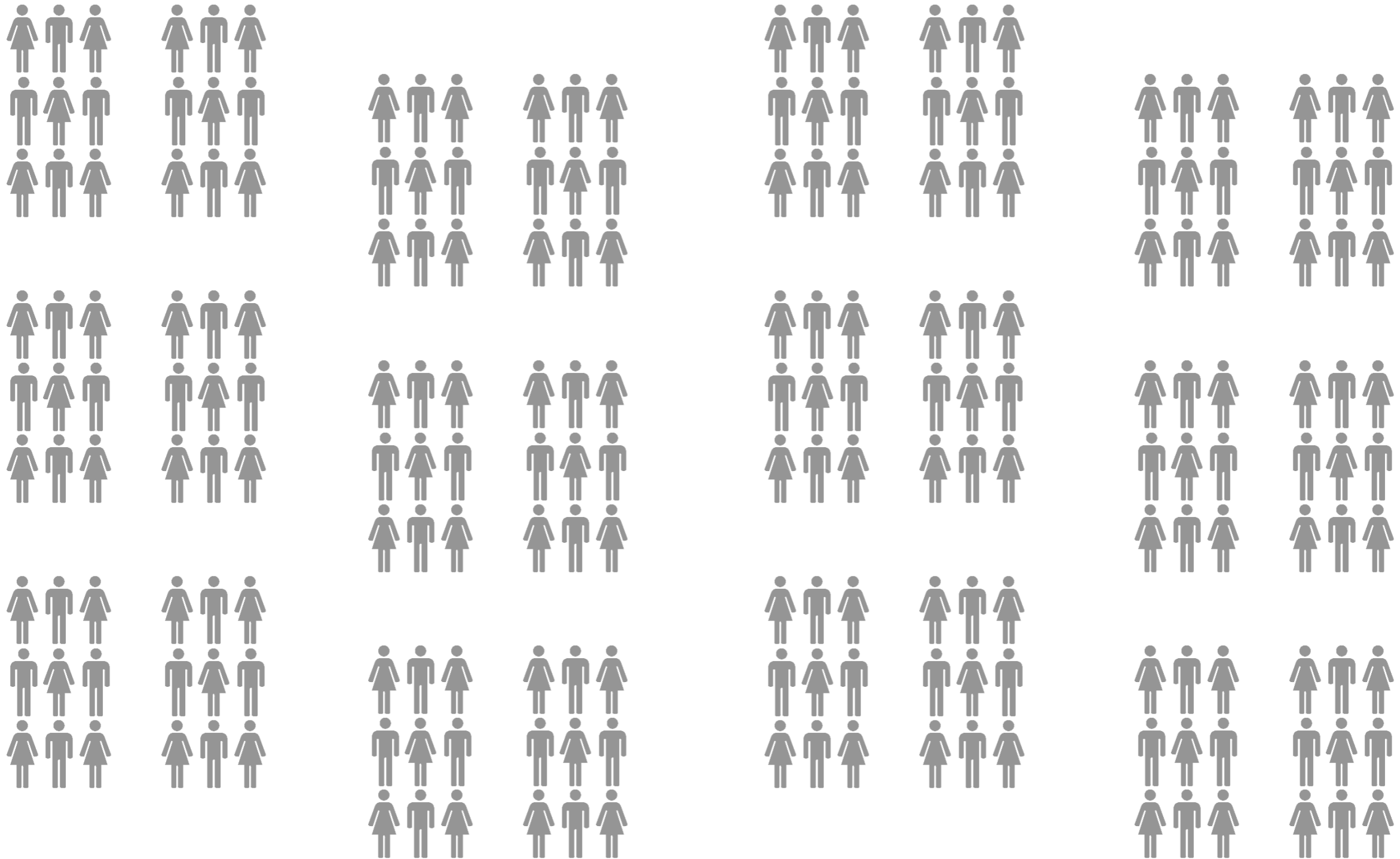




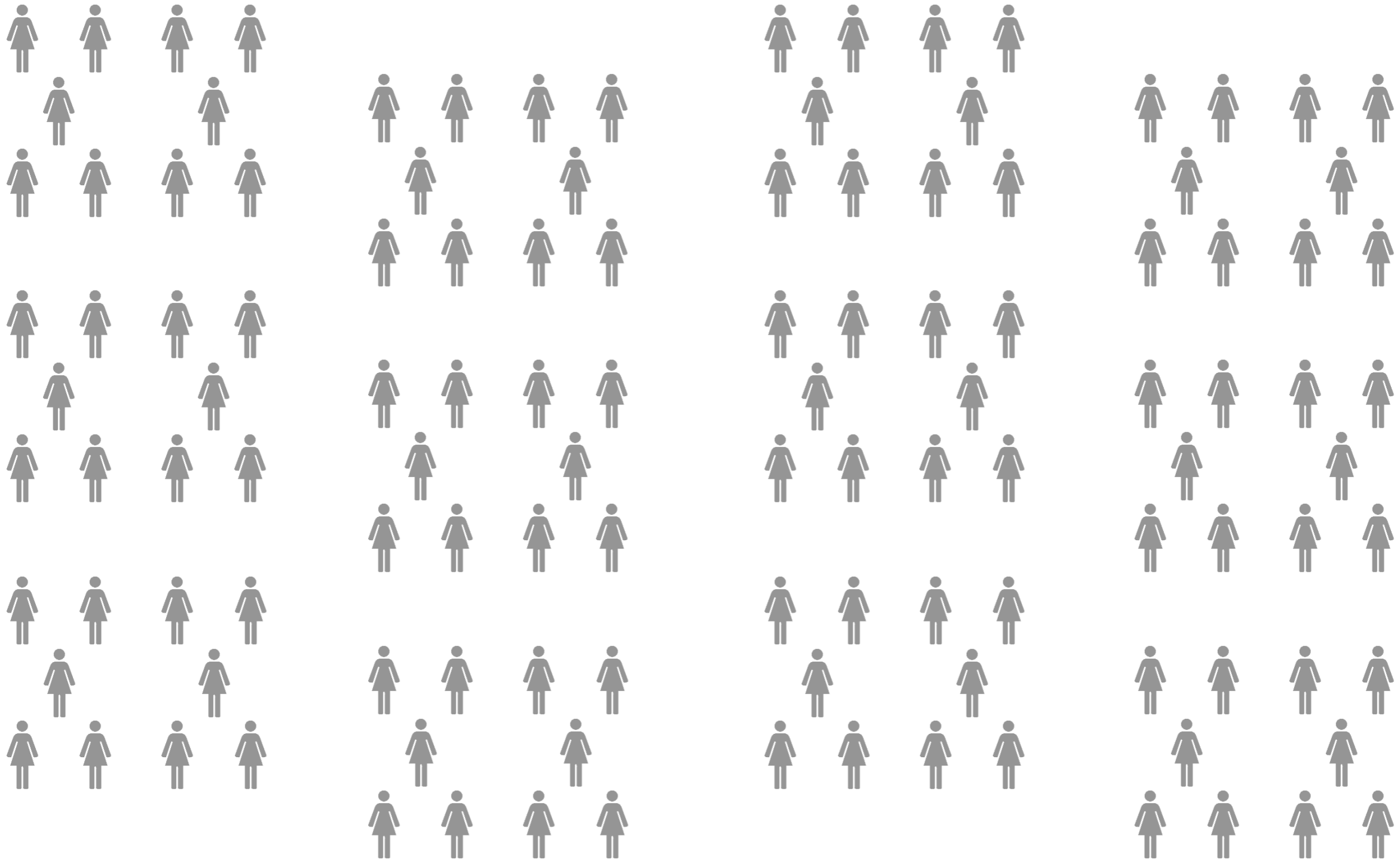
Stratification



Stratification II: By Gender



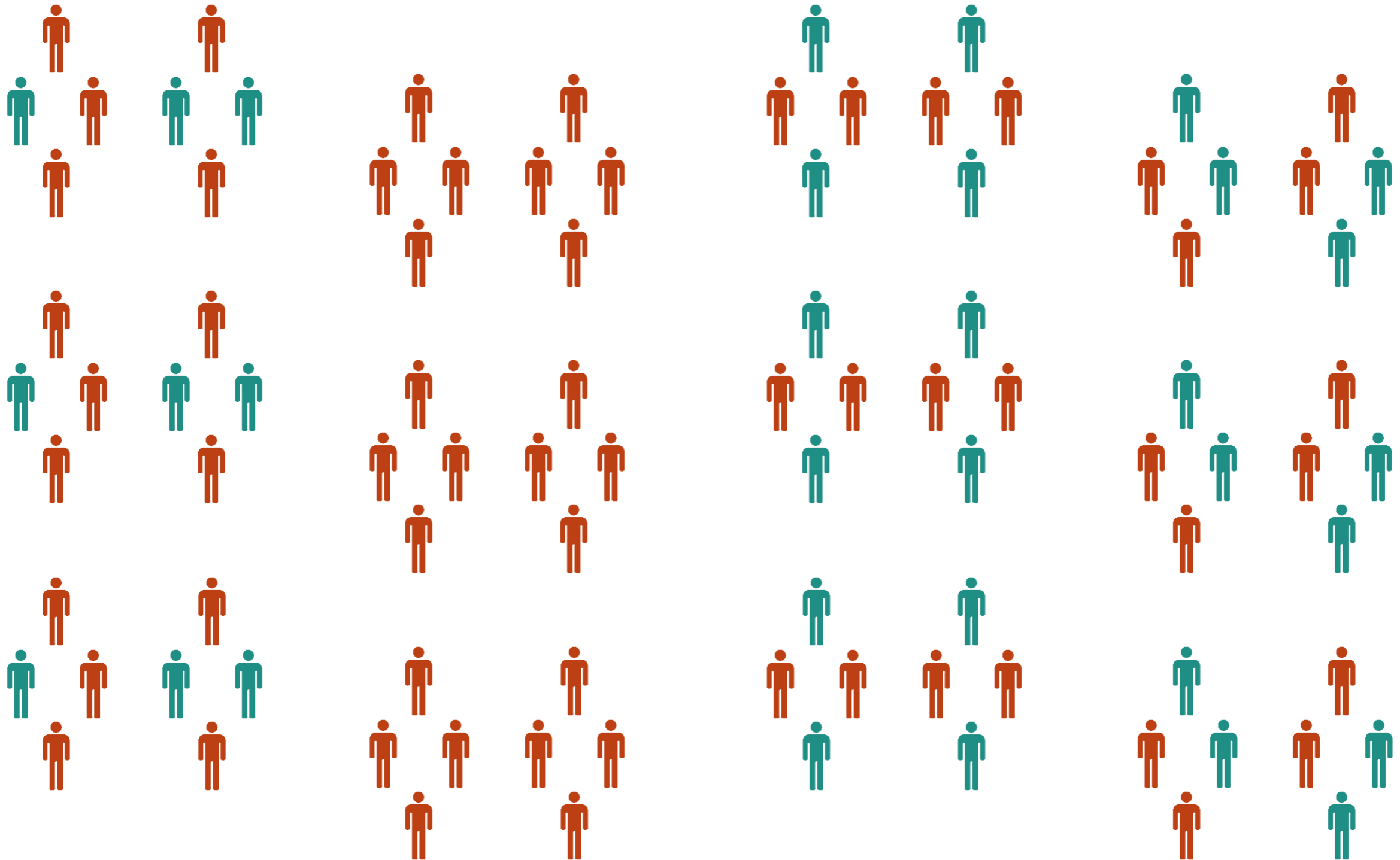
Take the women



Randomize Women separately



Randomize the men separately



Stratification by Gender



Recap on Challenges

Challenge	Implication	Solution
<ul style="list-style-type: none"> Service provider can't distinguish between T & C 	<ul style="list-style-type: none"> Crossovers 	<ul style="list-style-type: none"> Change Unit of Randomization Create a buffer
<ul style="list-style-type: none"> Control group finds out, benefits or is harmed 	<ul style="list-style-type: none"> Spillovers Crossovers Attrition 	<ul style="list-style-type: none"> Change Unit of Randomization Create a buffer
<ul style="list-style-type: none"> Resources to treat all 	<ul style="list-style-type: none"> No control group 	<ul style="list-style-type: none"> Phase in
<ul style="list-style-type: none"> Strict eligibility criteria 	<ul style="list-style-type: none"> Can't be randomized 	<ul style="list-style-type: none"> Randomization on the bubble
<ul style="list-style-type: none"> Program is an entitlement 	<ul style="list-style-type: none"> Can't force/deny program 	<ul style="list-style-type: none"> Encouragement Design
<ul style="list-style-type: none"> Sample size is small 	<ul style="list-style-type: none"> Insufficient power 	<ul style="list-style-type: none"> Change unit of randomization Stratification



Questions?



Follow us on Twitter @JPAL_SA

Visit our website www.povertyactionlab.org for more resources



Appendix:

Methods of randomization - recap

Design	Most useful when...	Advantages	Disadvantages
Basic Lottery	<ul style="list-style-type: none">•Program oversubscribed	<ul style="list-style-type: none">•Familiar•Easy to understand•Easy to implement•Can be implemented in public	<ul style="list-style-type: none">•Control group may not cooperate•Differential attrition

Methods of randomization - recap

Design	Most useful when...	Advantages	Disadvantages
Phase-In	<ul style="list-style-type: none">•Expanding over time•Everyone must receive treatment eventually	<ul style="list-style-type: none">•Easy to understand•Constraint is easy to explain•Control group complies because they expect to benefit later	<ul style="list-style-type: none">•Anticipation of treatment may impact short-run behavior•Difficult to measure long-term impact

Methods of randomization - recap

Design	Most useful when...	Advantages	Disadvantages
Rotation	<ul style="list-style-type: none">•Everyone must receive something at some point•Not enough resources per given time period for all	<ul style="list-style-type: none">•More data points than phase-in	<ul style="list-style-type: none">•Difficult to measure long-term impact

Methods of randomization - recap

Design	Most useful when...	Advantages	Disadvantages
Encouragement	<ul style="list-style-type: none">• Program has to be open to all comers• When take-up is low, but can be easily improved with an incentive	<ul style="list-style-type: none">• Can randomize at individual level even when the program is not administered at that level	<ul style="list-style-type: none">• Measures impact of those who respond to the incentive• Need large enough inducement to improve take-up• Encouragement itself may have direct effect