Principles Of Impact Evaluation And Randomized Trials
## Craig McIntosh
UCSD

Why are we here?

- What is the impact of the intervention?

  o What is the impact of NERICA on rice yields when it is used in practice?

  o What is the impact of improved information access on farmgate prices?

- Was this (observed) impact due to the program or something else?

  o Unbiased treatment or program effect

  o Attribution

Measuring Impact

"Treated" farmers yields

"Control" farmers yields
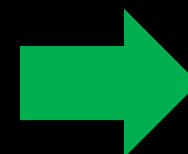
→ 10 kg

Farmers use seeds

Farmers don't use seeds
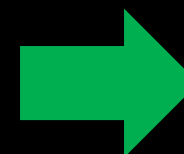
Offer farmers improved quality seeds

Is this unbiased? Too big or too small?

Measuring Impact
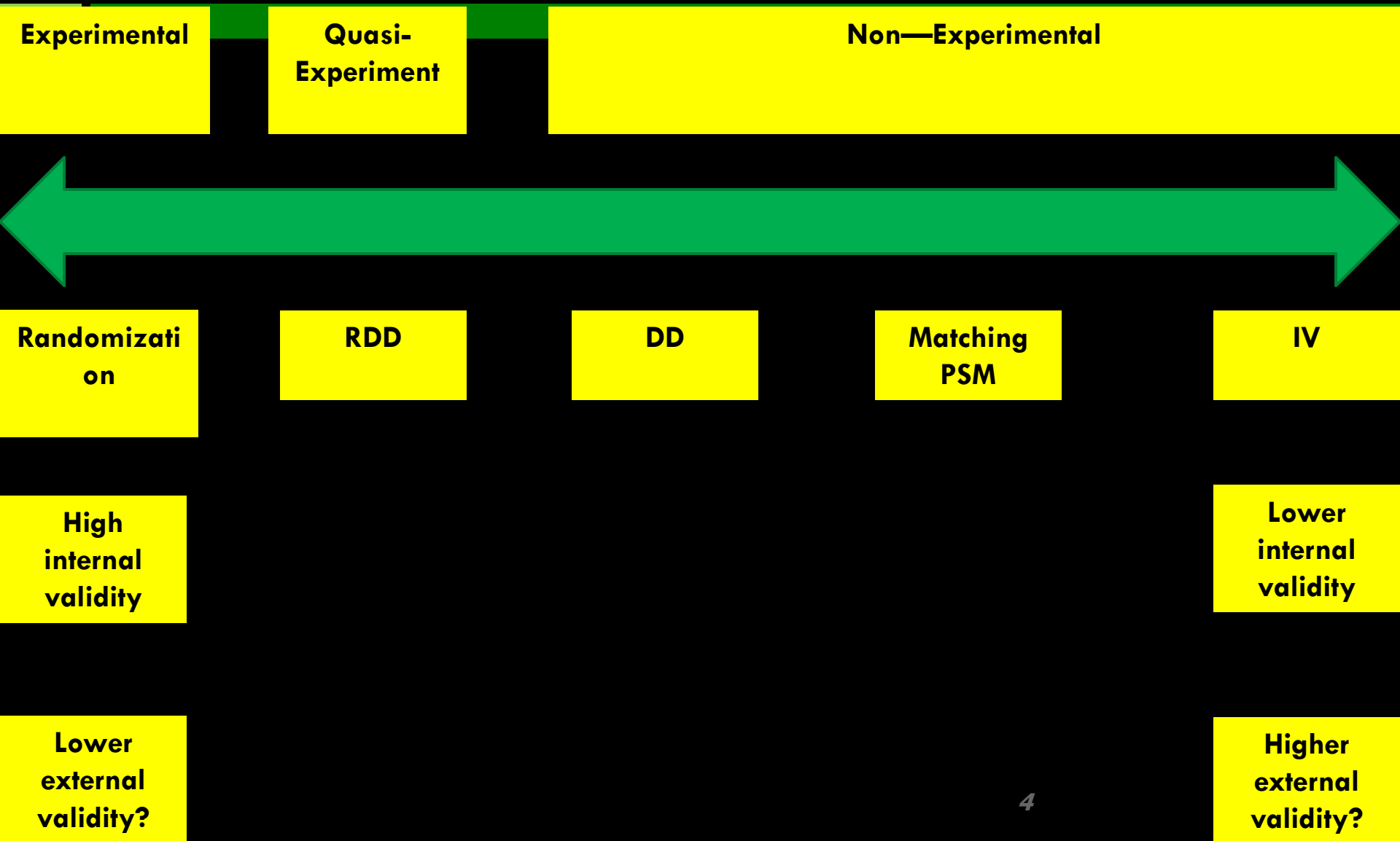
"Treated" farmers yields

"Control" farmers yields

→ 10 kg

Farmers in treated villages use seeds

Farmers in control villages don't use seeds

Choose villages away from a paved road to get seeds

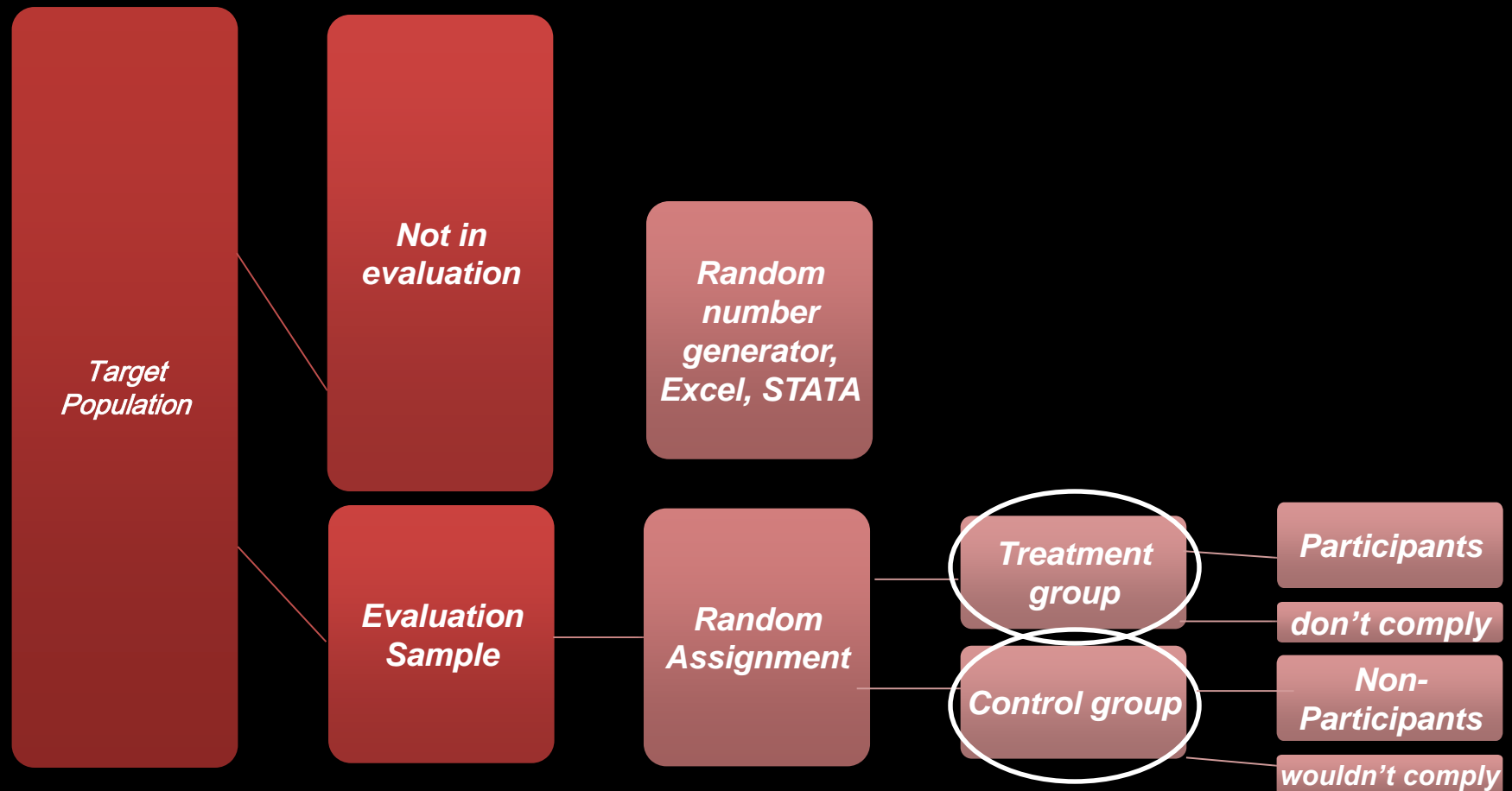Is this unbiased? Too big or too small?

What is randomization?

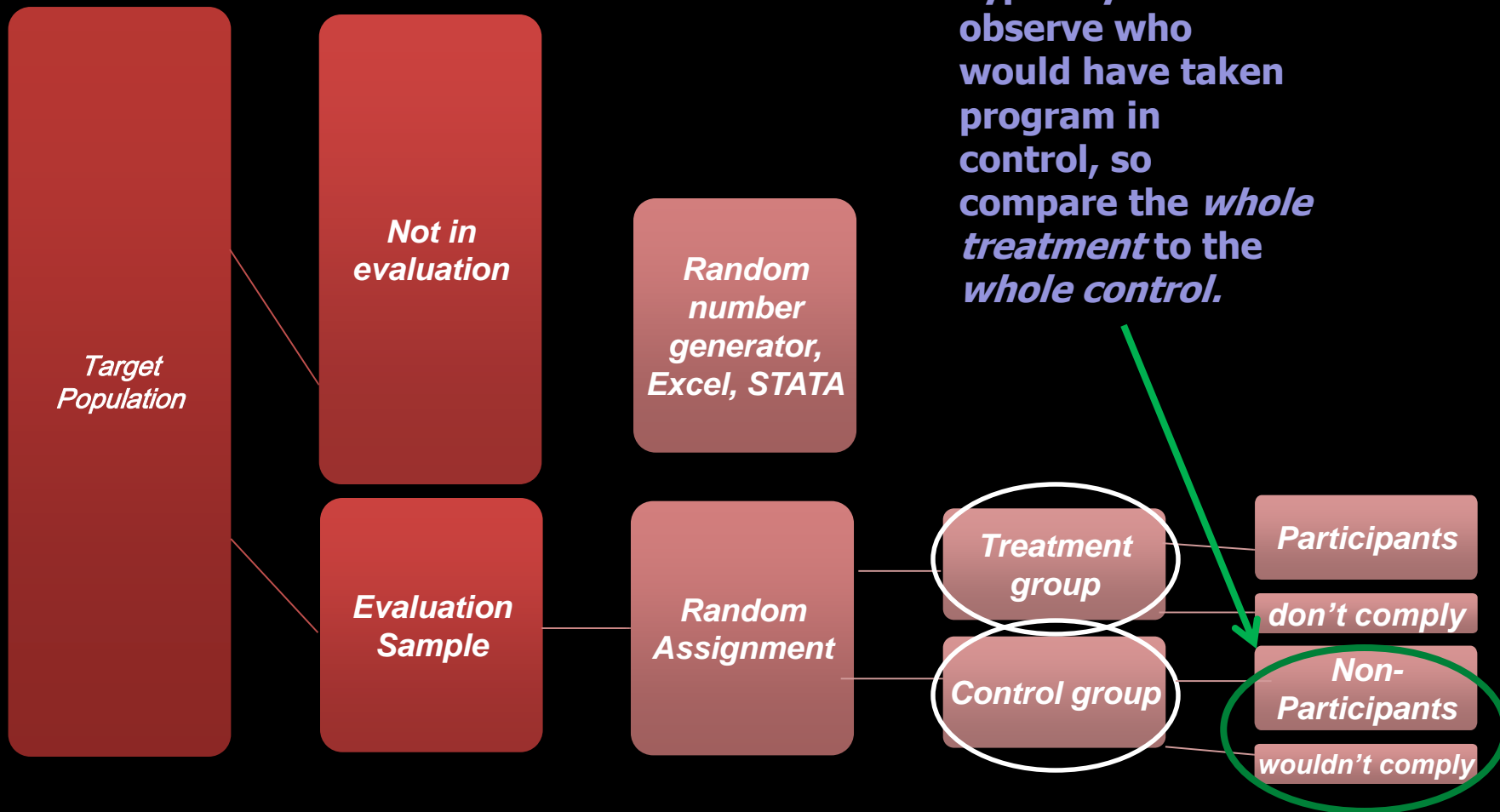- Randomization involves *randomly assigning* a potential participant (individual, household or village) to the treatment or control group

- It gives each potential participant a (usually equal) chance of being assigned to each group

- The objective is to ensure that the only systematic difference between the program participants (treatment) and non-participants (control) is the presence of the program

Basic setup

Basic setup

Target Population

Not in evaluation

Random number generator, Excel, STATA

Typically do not observe who would have taken program in control, so compare the *whole treatment* to the *whole control*.

Evaluation Sample

Random Assignment

Treatment group

Participants

don't comply

Control group

Non-Participants

wouldn't comply

How can randomization be useful to measure a program effect?

- On average (especially as sample size becomes large) *both un*observable *and* observable characteristics between program participants (treatment) and non-participants (control) are the same

- The only difference is the presence of the program

- Treatment effects very transparent for all involved in the study

- (But we need to check that it worked)

Lecture Overview

- Unit and method of randomization

- Real-world constraints

- Revisiting unit and method

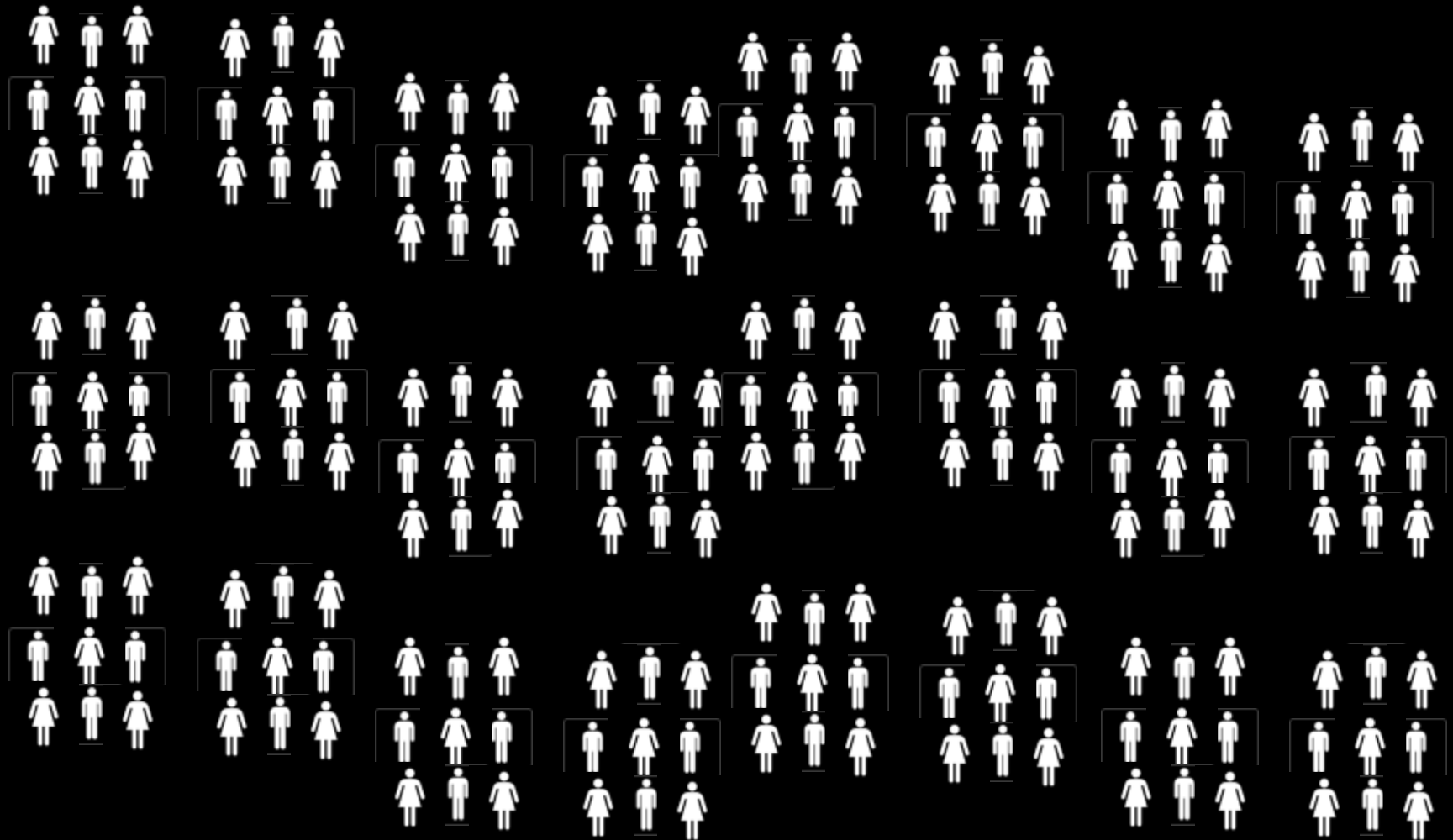- Variations on simple treatment-control

Lecture Overview

- **Unit and method of randomization**

- Real-world constraints

- Revisiting unit and method

- Variations on simple treatment-control
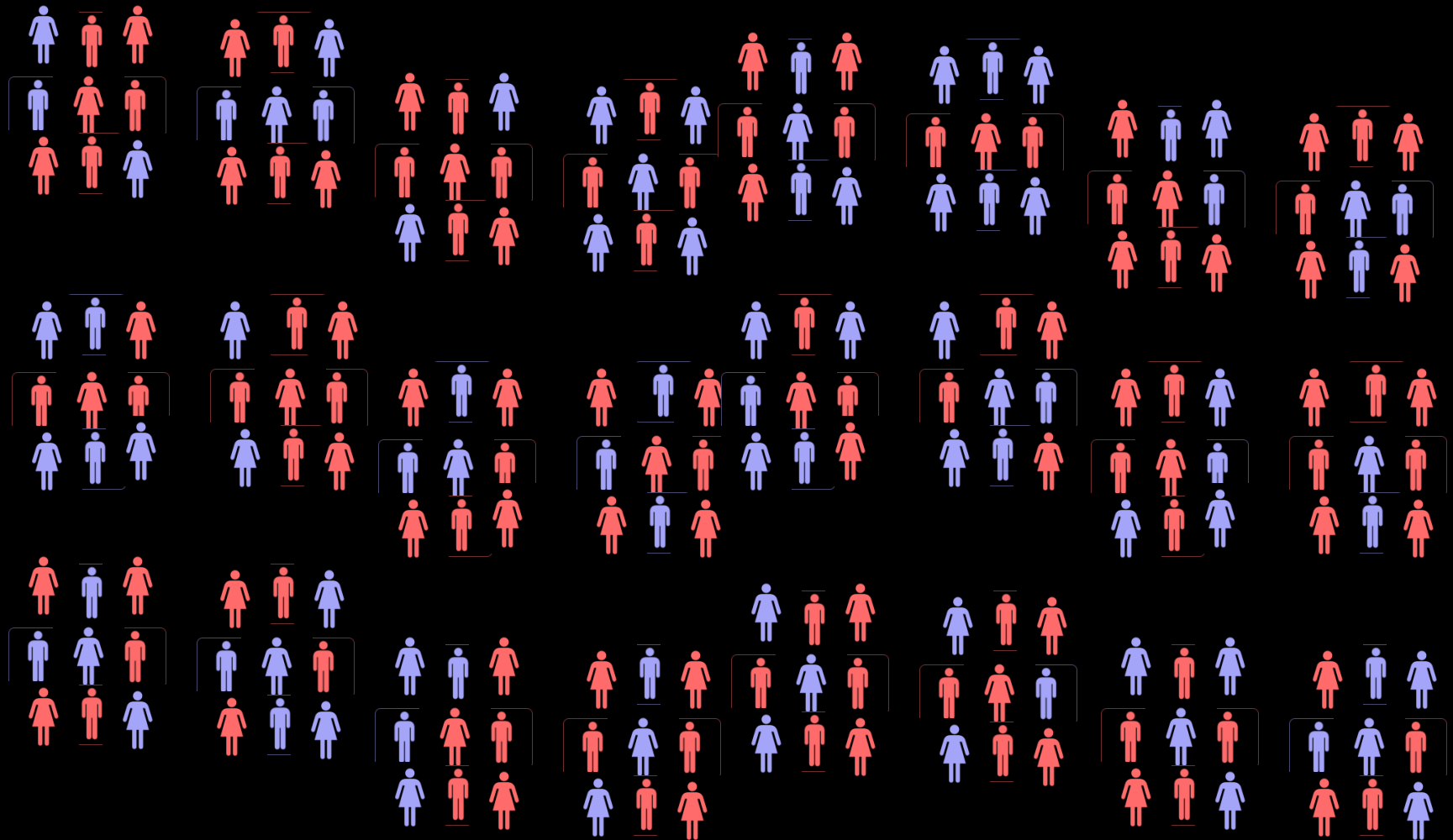
Unit of Randomization: Options

1. Randomizing at the individual level

2. Randomizing at the group level "Cluster Randomized Trial"

- At which level should we randomize?
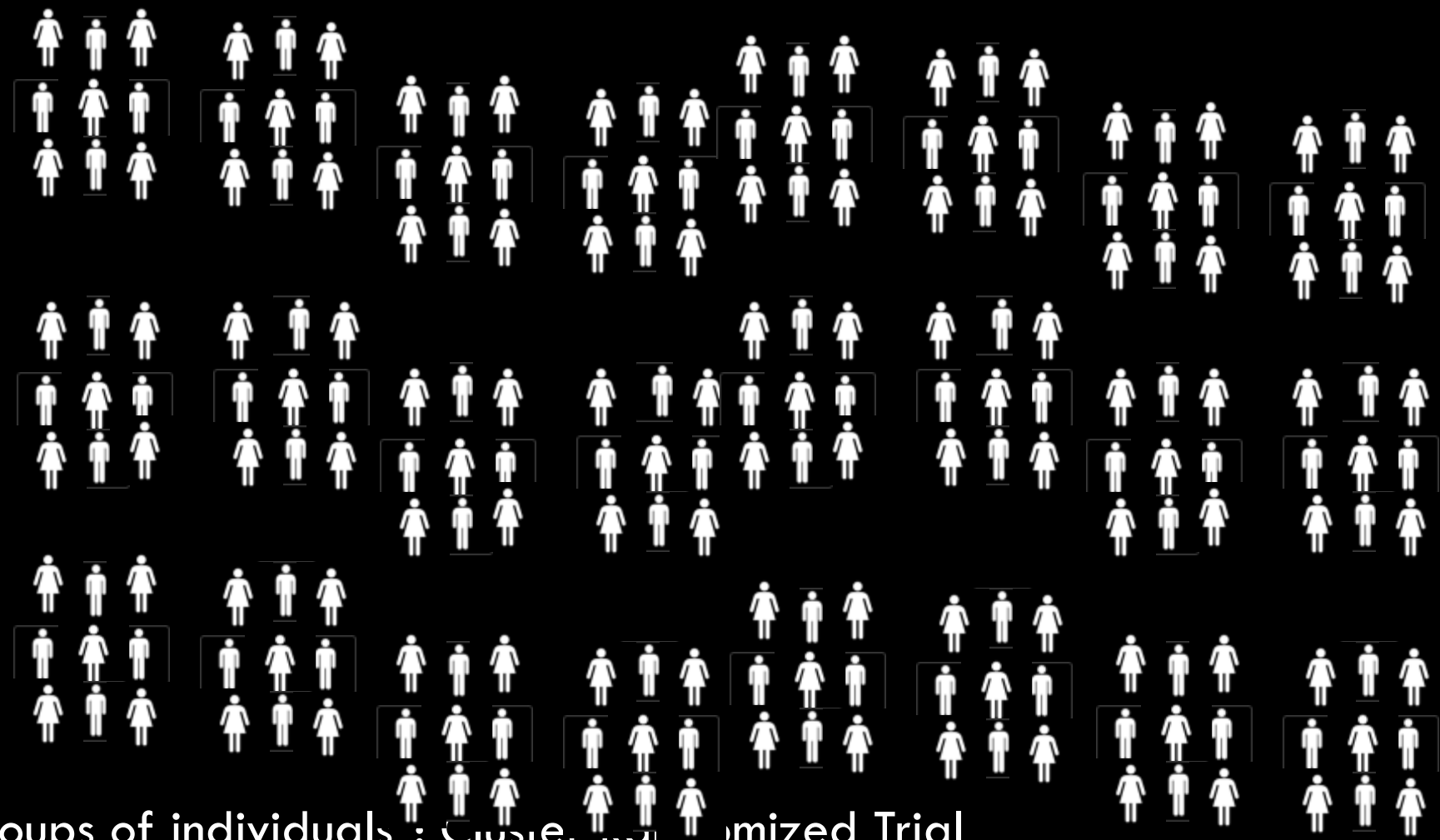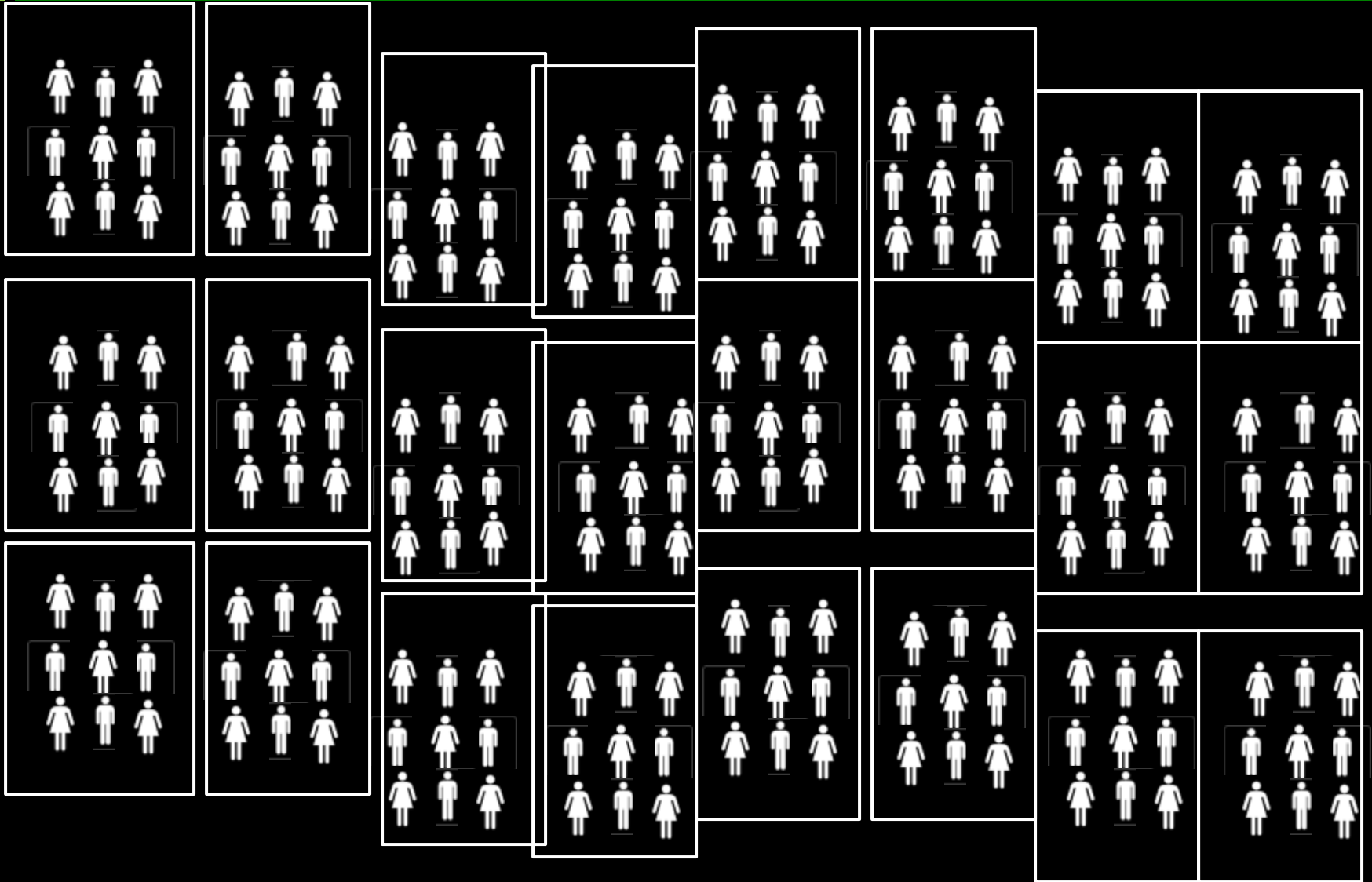
Unit of Randomization: Clusters?

"Groups of individuals : Cluster Randomized Trial
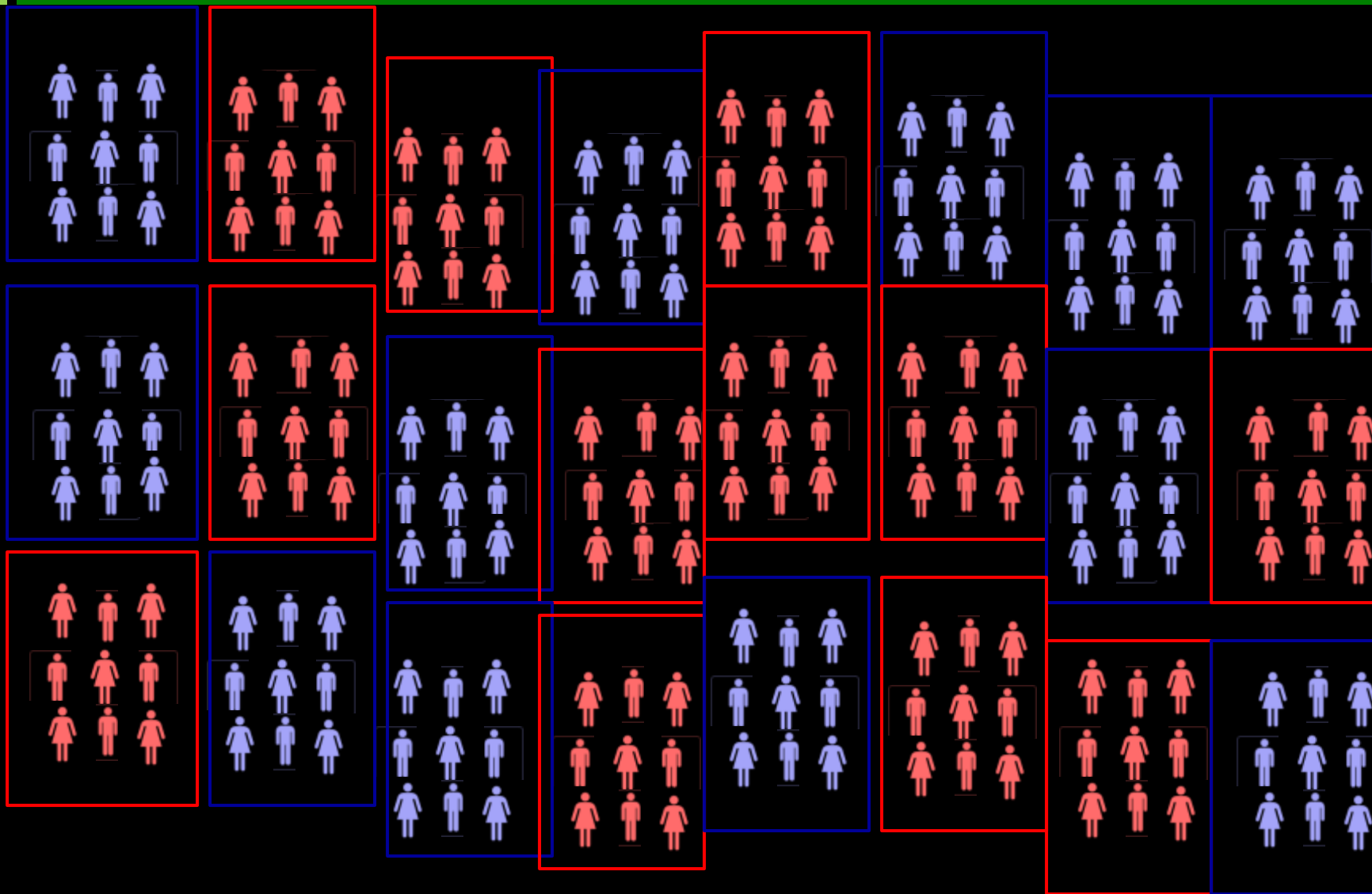
Unit of Randomization: Village?
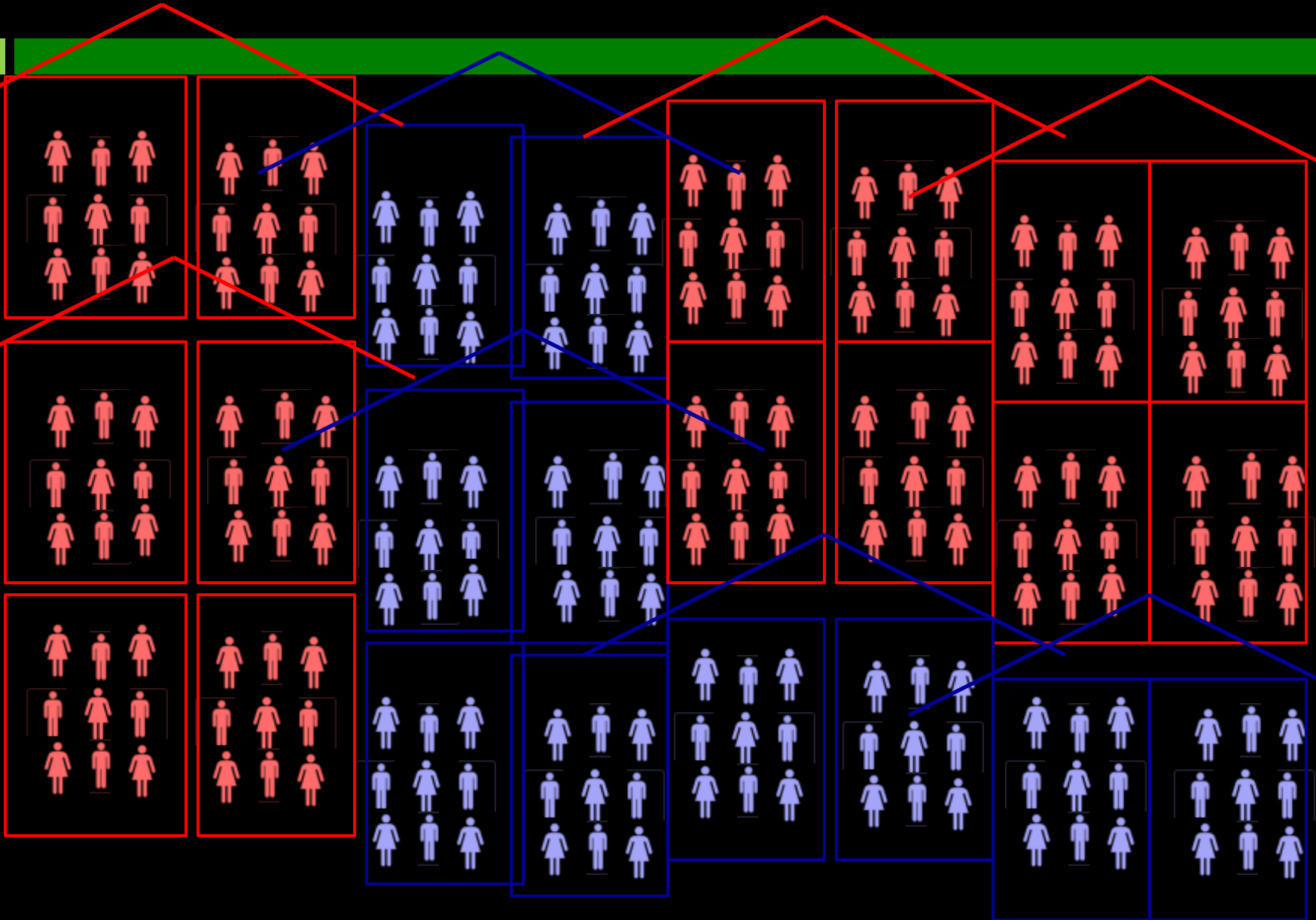
How do we choose the level?

- What unit does the program target for treatment?

- What is the unit of analysis?

How do we choose the level?

- Nature of the Treatment
  - How is the intervention administered?
  - What is the catchment area of each "unit of intervention"
  - How wide is the potential impact?
- Aggregation level of available data
- Power requirements: role of the 'design effect'.
  - power loss larger as those within cluster more similar
- Most natural to randomize at the level at which the treatment is administered.

Example: Individual design

- Intervention: A bank-linked mobile phone that permits account savings via airtime cards.

- Treatment Level: Individual.

- Randomization level: Individual.

- A self-employed, unbanked, and semi-urban sample drawn in, 5 towns in Sri Lanka.

- Offers of phones made directly at the individual level.

Example: Clustered design

- Intervention: Cash transfers for schooling

- Treatment level: Village

- Randomization level: Village
  - Sample of eligible households identified.
  - Households of eligible girls in treatment villages receive cash transfer if children remain in school.
  - Power lower than individual treatment, but school monitoring and transfers are both most natural at village level.

Example: Randomized Pricing

- Intervention: Rainfall-based index insurance for cooperativized farmers in Ethiopia.

- Treatment level: individual coop members

- Randomization level:
  - Treatment/Control: Village-level coops
  - Insurance price vouchers: Individual farmers
    - Twenty farmers selected in each village
    - Price vouchers for 100-700 birr are randomly distributed to individual members; gives information on demand curve for insurance.

Example: Randomized Pricing

Demand Curve for Index Insurance
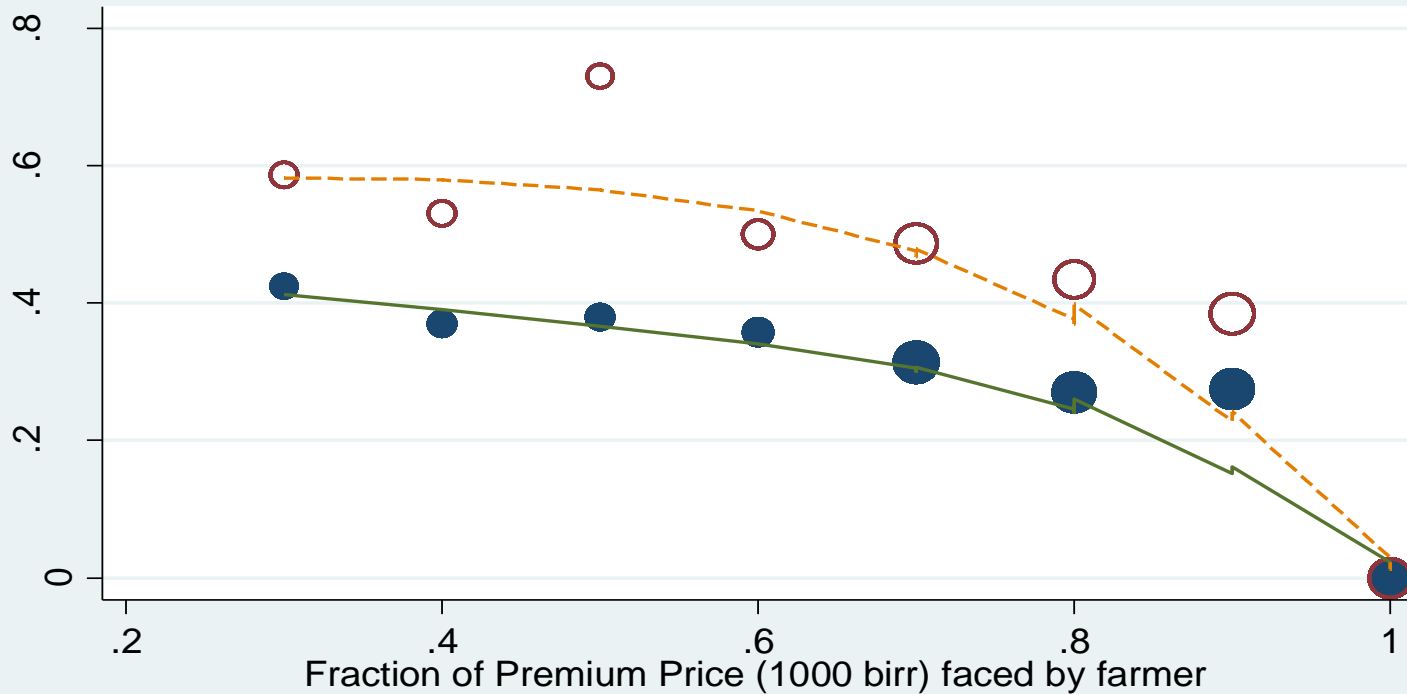
Circle size proportional to number of observations at each subsidy amount

Legend:
- All Treatments
- Kebeles with Any Uptake
- All, fitted
- Uptake, fitted

X-axis: Fraction of Premium Price (1000 birr) faced by farmer

Lecture Overview

- Unit and method of randomization

- **Real-world constraints**

- Revisiting unit and method

- Variations on simple treatment-control

Real-World Constraints

- Fairness and ethical issues

- Political Concerns

- Resources

- Crossovers/spillovers

- Logistics

- Sample size

Fairness

- Randomizing at the individual level within a farmers' association
    - Non-treated farmers might be unhappy
- Randomizing at the household-level within the village
    - Non-recipient households or the village chief might be unhappy
- Randomizing at the village or farmers' association level
    - Ministry of Agriculture might be unhappy

Political Concerns

- Lotteries are simple and common

- Randomly chosen from applicant pool

- Participants know the "winners" and "losers"

- Simple lottery is useful when there is no a priori reason to discriminate

- Can be perceived as fair

- Transparent

Resources

- Many programs have limited resources
  - Vouchers, Subsidies, Training
  - More eligible recipients than resources
- How will program recipients be chosen?
  - Clear-cut criteria
  - Arbitrary criteria
  - Random process
  - Some combination of the above

Spillovers/Crossovers

- Contamination of the control group can be due to:
  - Spillovers – positive or negative
  - Crossovers – movement to treatment (or control) group
- New designs make direct estimation of spillovers possible, but they require larger sample sizes.

Logistics

- Is it possible or feasible for staff to implement different programs in the same catchment area?
- Agricultural extension agent provides training in improved planting techniques
  - Training is one of many responsibilities of the agent
  - The agent might serve farmers from both treatment and control villages within his/her catchment areas
  - It might be difficult to train them to follow different procedures for different groups, and to keep track of what to give whom

Sample Size

- The program is only large enough to serve a handful of communities

- Might not be able to survey (or implement the program in) enough communities to detect a (statistical) effect

Lecture Overview

- Unit and method of randomization

- Real-world constraints

- **Revisiting unit and method**

- Variations on simple treatment-control

Possible Randomization Designs

- Simple lottery

- Randomization in the "bubble"

- Randomized phase-in

- Rotation

- Encouragement design

These are not mutually exclusive.
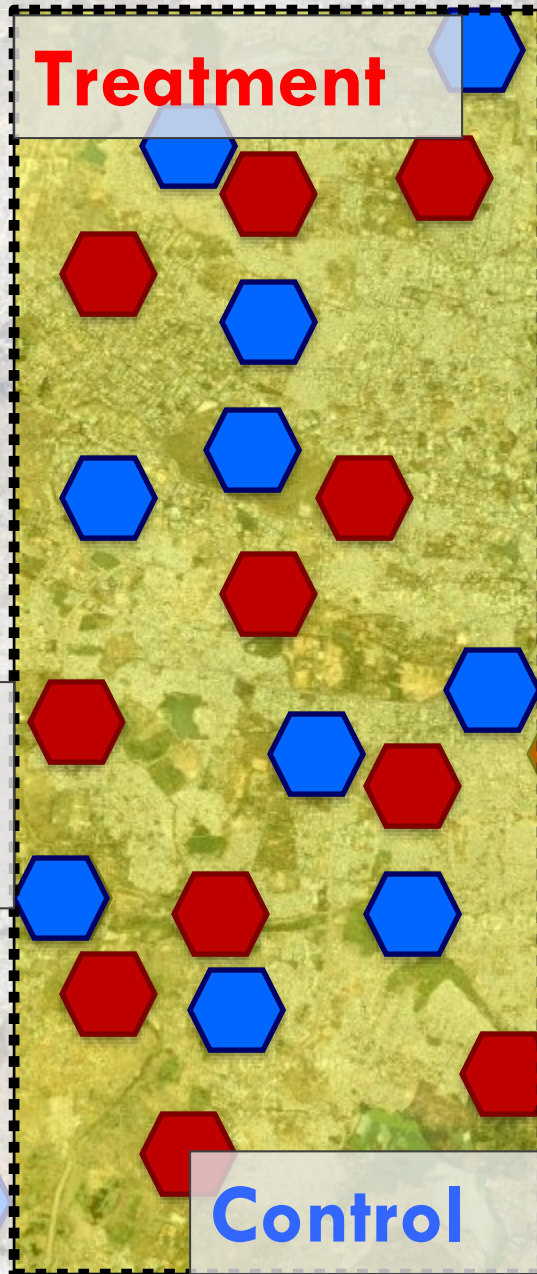
Randomization in "the bubble"

- A partner may not be willing to randomize among eligible people.

- However, a partner might be willing to randomize in "the bubble."

- People "in the bubble" are those who are borderline in terms of eligibility
  - Just above the threshold → not eligible, but almost

- What treatment effect do we measure? What does it mean for external validity?

Randomization in "the bubble"

Within the bubble, compare **treatment** to **control**

Treatment

Control

Non-participants > .25 ha

Participants <=.25 ha

Randomization in the Bubble

Must receive the program

Randomized assignment to the program

Ineligible
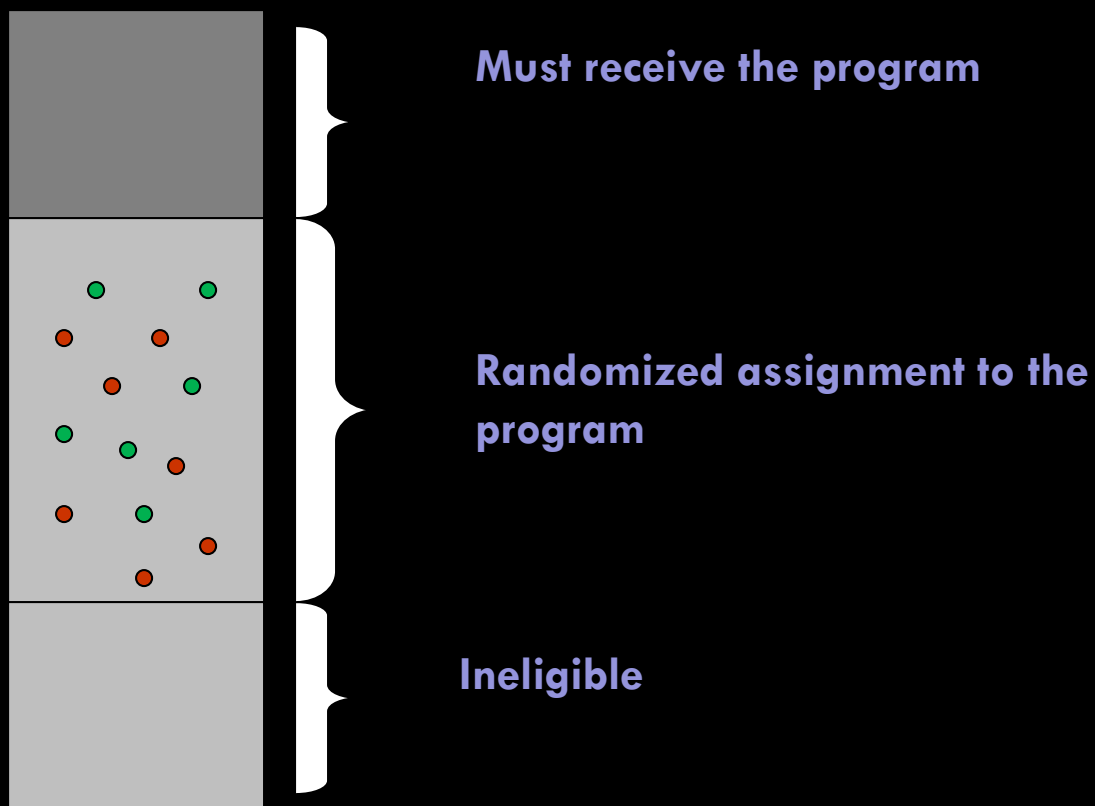
Randomization in "the bubble"

- Program still has discretion to treat "necessary" groups

- Example: Agricultural grant program in Niger (PRODEX)

- Example: Expansion of consumer credit in South Africa

Randomized Phase-In

- Takes advantage of the program expansion (ie, the NGO cannot implement in all villages the first year)

- Everyone gets program eventually

- If everyone is eligible for the program, what determines which villages, schools, branches, etc. will be covered in which year?

Randomized Phase-In

Round 1
Treatment: 1/3
Control: 2/3

Round 2
Treatment: 2/3
Control: 1/3

Randomized evaluation ends

Round 3
Treatment: 3/3
Control: 0

Randomized Phase-In
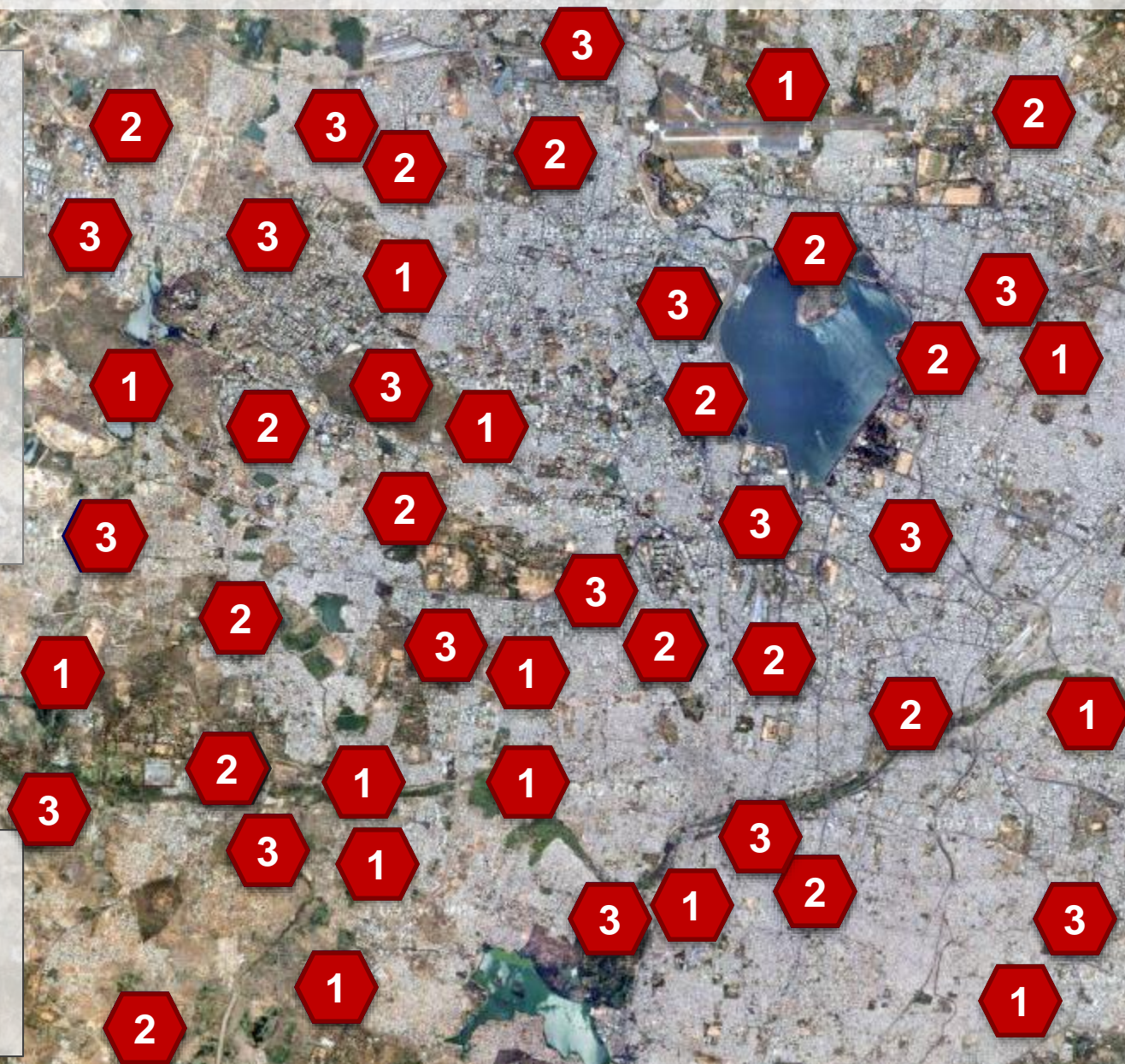
Advantages

Everyone gets something eventually

Provides incentives to maintain contact

Concerns

Can complicate estimating long-run effects

Be careful with phase-in windows

Do expectations of change actions today?

Rotation

- **Groups get treatment in turns**
  - Group A gets treatment in the first period
  - Group B gets treatment in the second period

Rotation design

Round 1
Treatment: 1/2
Control: 1/2

Round 2
Treatment
from Round 1
→ **Control**

Control from
Round 1 →
Treatment

Rotation
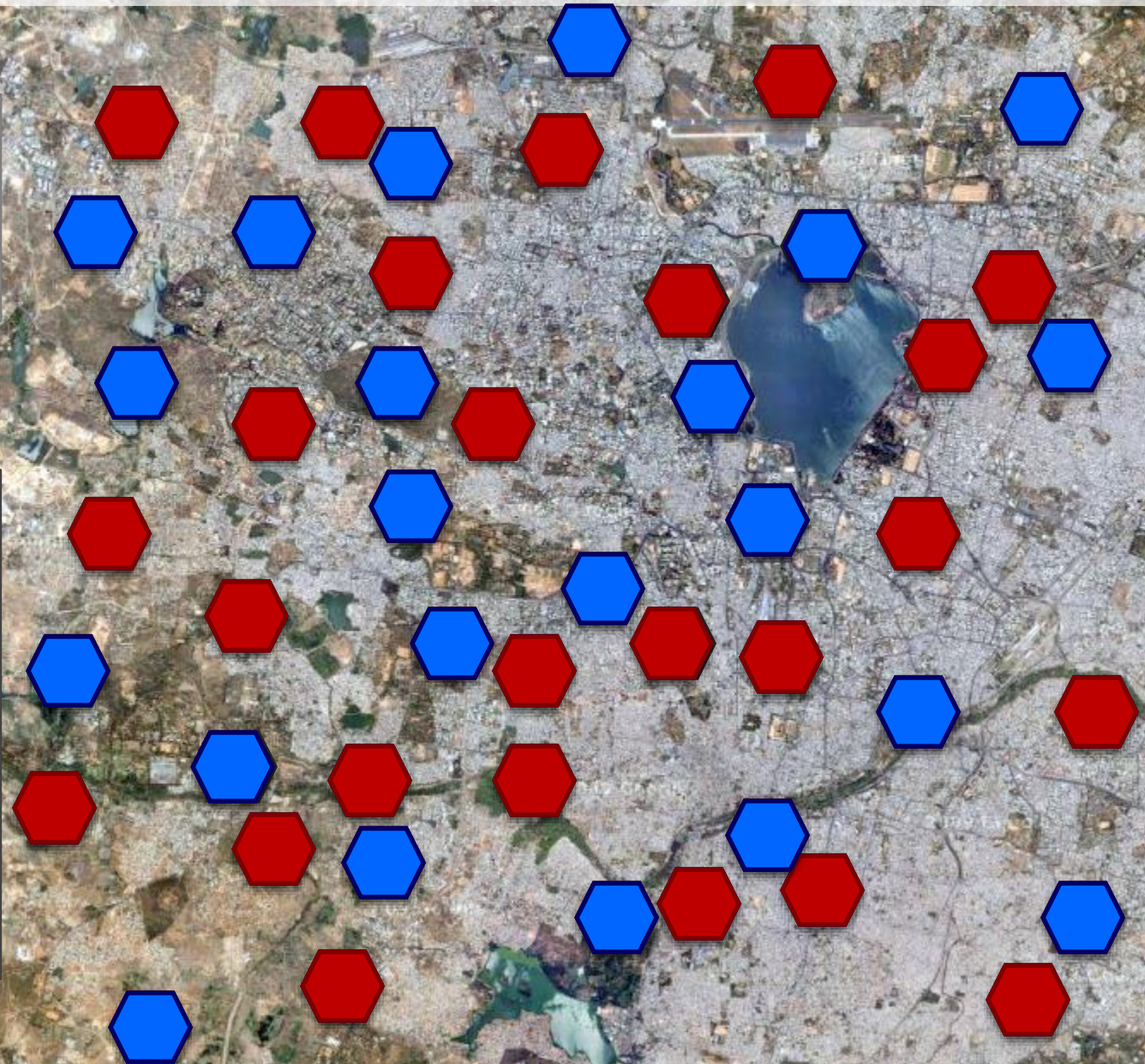
- **Advantages:**
    - Might be perceived as fairer, therefore easier to get accepted

- **Disadvantages:**
    - If those in Group B anticipate treatment, they might change their behavior
    - Cannot measure long-term impact because no pure control group

Randomized Encouragement

- Sometimes it's not possible to randomize program access (vaccines, savings program, etc)

- But many programs have less than 100% take-up

- Randomize encouragement to receive treatment

Encouragement design

Encourage

Do not encourage

participated

did not participate

Complying

Not complying

compare **encouraged** to **not encouraged**

These must be correlated

do not compare **participants** to **non-participants**

adjust for non-compliance in analysis phase

What is "encouragement"?

- Something that makes some individuals more likely to use program than others
    - Not in itself a treatment
    - E.g., vouchers, training, visit from agent, etc
- For whom are we estimating the treatment effect?
    - Think about who responds to encouragement (compliers)

Summary: Experimental Designs

- Simple lottery

- Randomization in the "bubble"

- Randomized phase-in

- Rotation

- Encouragement design

These are not mutually exclusive.

Methods of randomization - recap

Design	Most useful when…	Advantages	Disadvantages
Basic Lottery	•Program oversubscribed	•Familiar •Easy to understand •Easy to implement •Can be implemented in public	•Control group may not cooperate •Differential attrition

Methods of randomization - recap

Design	Most useful when…	Advantages	Disadvantages
Phase-In	•Expanding over time •Everyone must receive treatment eventually	•Easy to understand •Constraint is easy to explain •Control group complies because they expect to benefit later	•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	•Everyone must receive something at some point •Not enough resources per given time period for all	•More data points than phase-in	•Difficult to measure long-term impact

Methods of randomization - recap

Design	Most useful when…	Advantages	Disadvantages
Encouragement	•Program has to be open to all comers •When take-up is low, but can be easily improved with an incentive	•Can randomize at individual level even when the program is not administered at that level	•Measures impact of those who respond to the incentive •Need large enough inducement to improve take-up •Encouragement itself may have direct effect

Lecture Overview

- Unit and method of randomization

- Real-world constraints

- Revisiting unit and method

- **Variations on simple treatment-control**

Variations on Simple Treatment and Control

- Multiple treatments

- Crossing or interacting treatments

- Randomizing incentives to comply

- Stratified randomization

- Multiple-stage randomization

- Discontinuity in eligibility

Multiple treatments

- Sometimes the core question is deciding among different possible interventions
  - Example: in-person extension agent visits versus a call-in hotline
- You can randomize these interventions
- Does this teach us about the benefit of any one intervention?
- Do you have a control group?

Multiple treatments
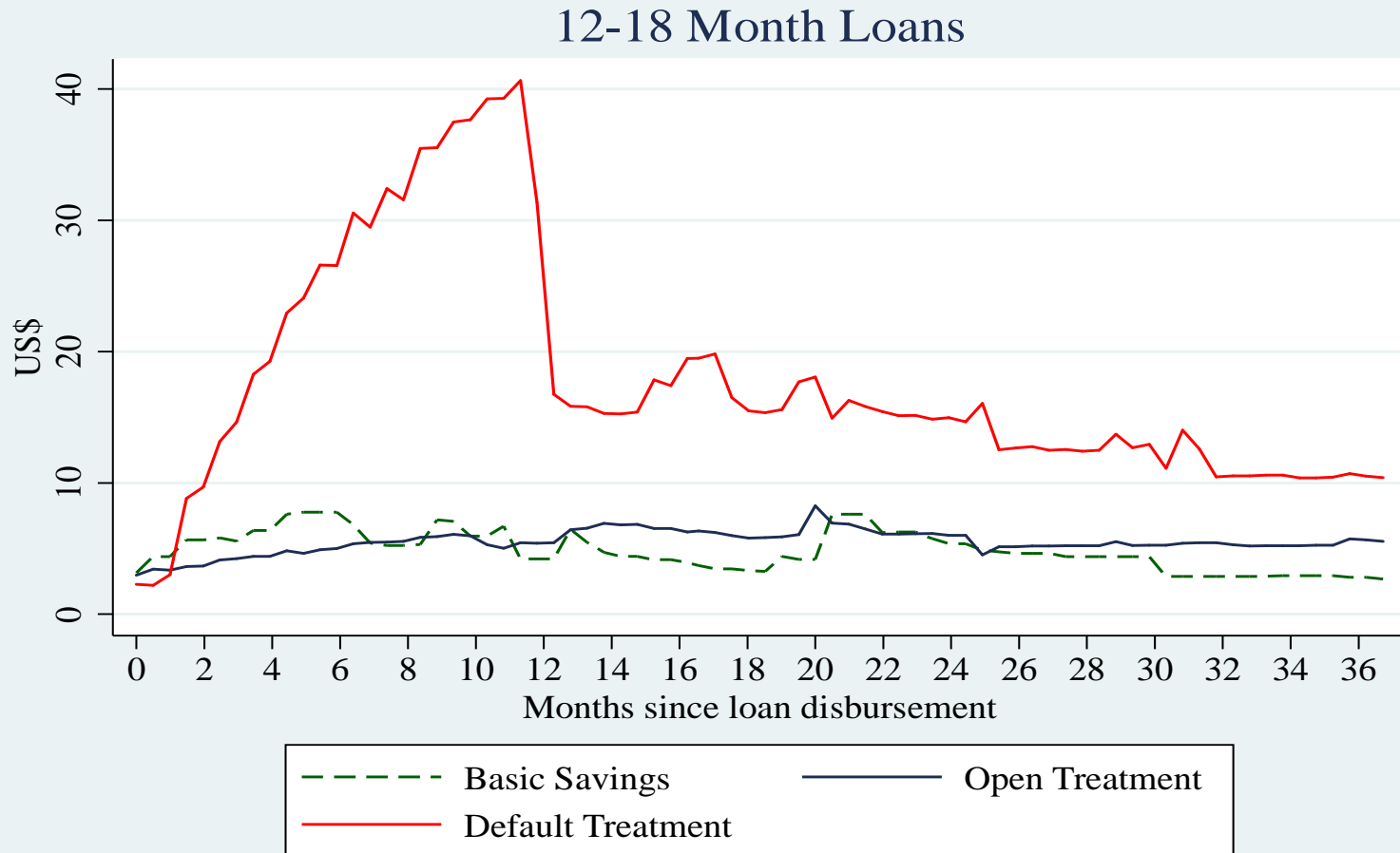
Treatment 1
Treatment 2
Treatment 3
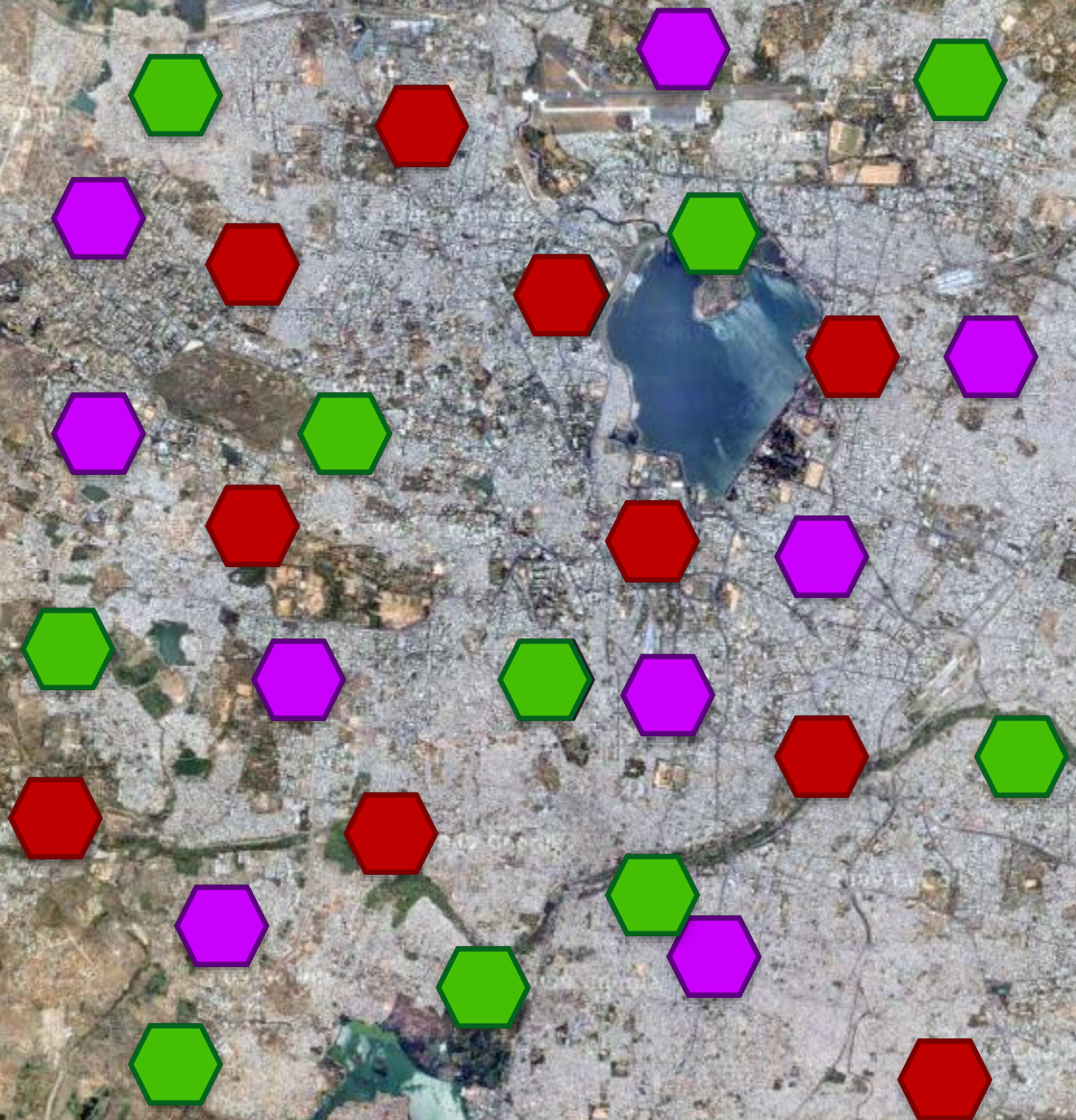
New products vs. 'standard'.

- Many institutions capture data only on clients/beneficiaries, makes controls expensive.

- In a product innovation, the standard product is a natural control group.

- Makes it relatively easy to experiment, capture outcomes of most interest to implementer.

- However, these designs do not measure the impact of the standard product at all.

New products vs. 'standard'.

12-18 Month Loans

Top 1% excluded.

Cross-cutting treatments

- Test different components of treatment in different combinations
    - Improved seeds only, improved seeds plus training, training only, no treatment
- Test whether components serve as substitutes or complements
- What is most cost-effective combination?
- Advantage: win-win for operations, can help answer questions for them, beyond simple "impact"!

Varying incentives to comply

- Testing subsidies and prices
    - Vary the price of seeds, inputs or access to market
    - Vary information, via mobile phone
    - Provide temporary subsidies and see whether this 'incentive to adopt' can have lasting consequences for adoption
- Testing social networks
    - Who are the pivotal actors whose behavior is influential for the decisions of others?

Stratified Randomization

- Randomization should, in principle, ensure balance in the treatment and control groups if the sample size is large enough

- What happens when it is small?

- Stratified randomization can help to ensure balance across groups when there is a small(er) sample
  - Divide the sample into different subgroups
  - Select treatment and control from each subgroup

- What happens if you don't stratify?

Stratified Randomization

- Stratify on variables that could have important impact on outcome variable (bit of a guess)

- Stratify on subgroups that you are particularly interested in (where may think impact of program may be different)

- Stratification more important when small data set

- Can get complex to stratify on too many variables

- Makes the draw less transparent the more you stratify

- You can also stratify on index variables you create
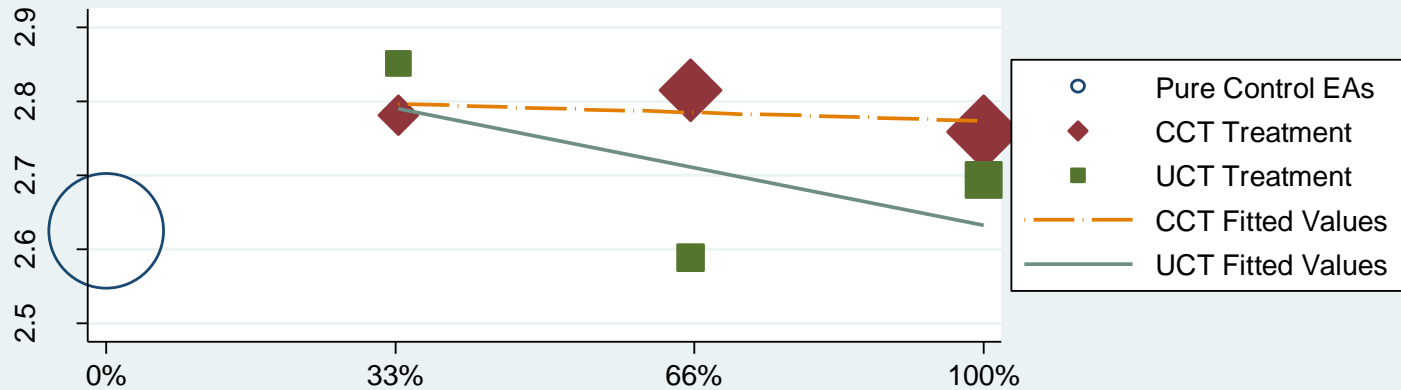
Multi-Stage Randomization

- Can use these designs to measure spillover effects.
- Two stages:
1. Randomize the fraction of a 'cluster' to be treated
2. Randomly pick the individual units to be treated based on the cluster-level saturation.

- Compare treated to untreated (normal impact)
- Compare within-cluster controls to pure controls (spillover impact)
- Compare impact for different intensities of treatment (saturation and threshold effects)
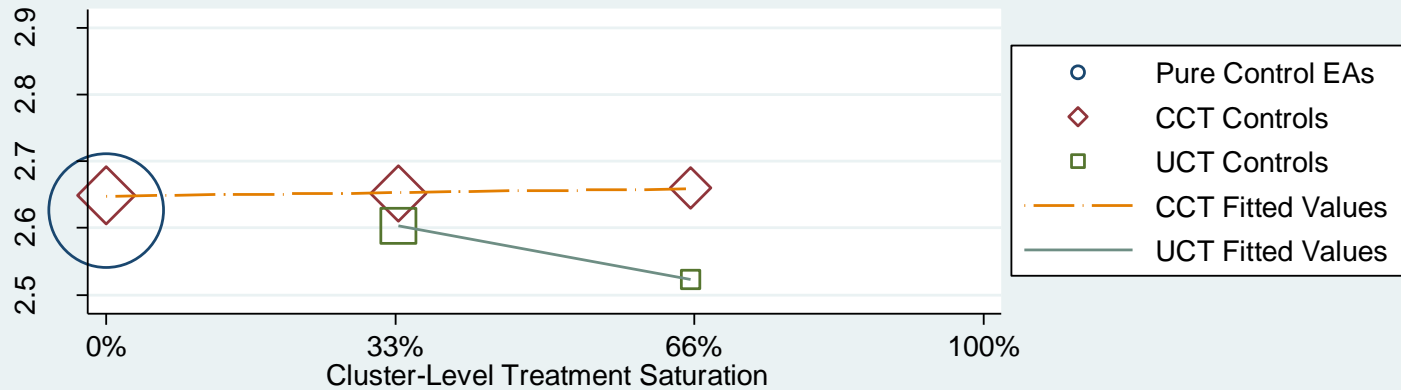
Multi-Stage Randomization

Enrollment by EA-Level Treatment Saturation

Discontinuity design

- If program has a sharp eligibility threshold, those just eligible and just ineligible are 'as if randomized'.
- Allows a clean estimation of impact.
  - Only provides impact *at that eligibility threshold;* not for any other type of person.
  - However, care most about this impact because this is the margin of expansion?

- Can be straightforward way of getting impact, but requires strict adherence to a rule of eligibility.

Regression Discontinuity (RD):

Figure 4: Program Participation, Refrigerators
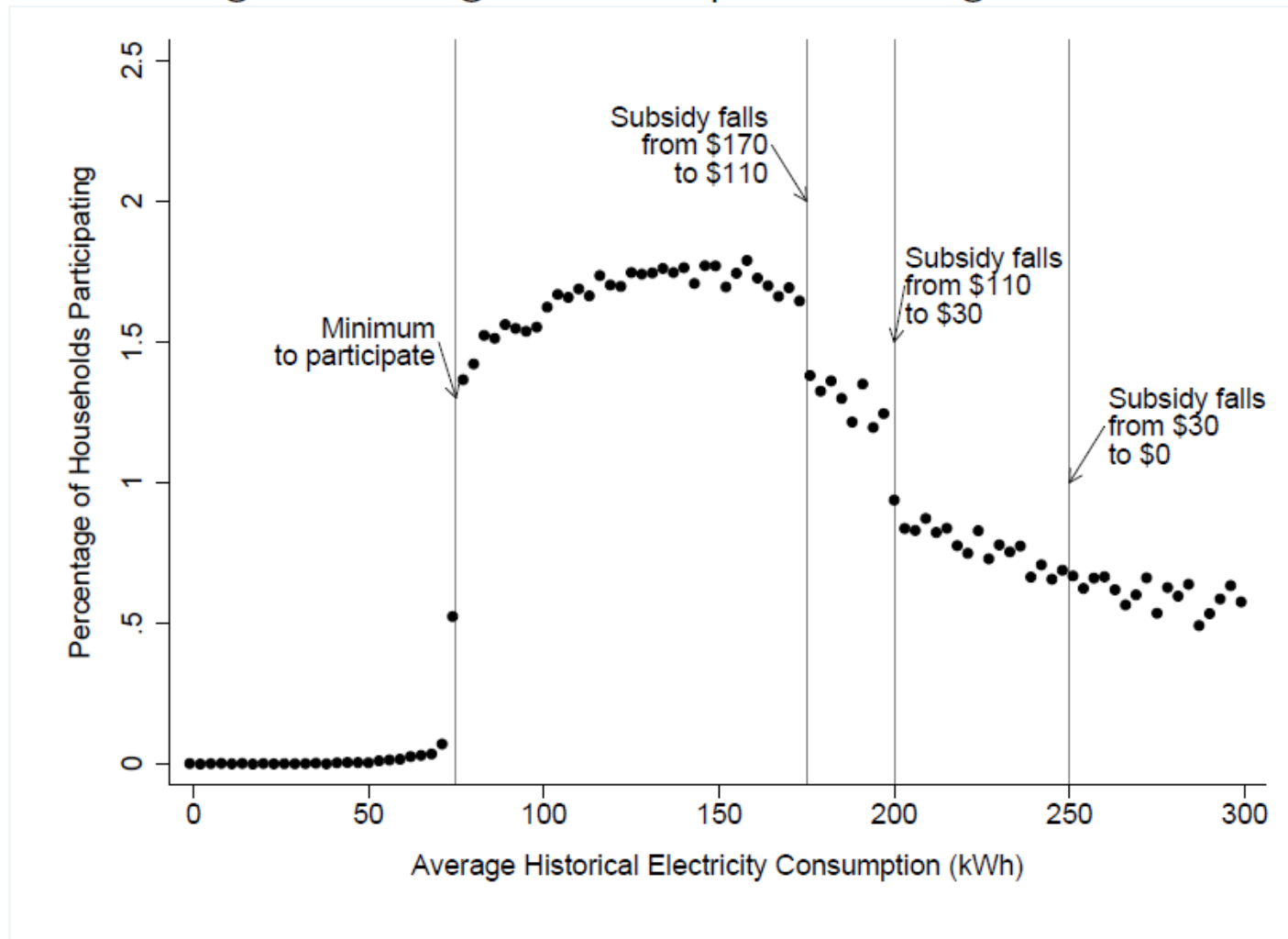
Source: Boomhower and Davis (2013).

Mechanics of Randomization

- Need sample frame

- Pull out of a hat?

- Use random number generator in spreadsheet program to order observations randomly?

- Stata program code

- What if no existing list:
  - listing exercise
  - random sampling rules

Source: Jenny Aker

Thank you!