IE through randomization: introduction and strategies

Harsha Thirumurthy

UNC-Chapel Hill

& The World Bank

Challenges in attributing impact

- Impact evaluation is important for resource allocation & policy-making
- But attributing impact can be challenging
  - For example, observation of drop in HIV incidence may be unrelated to a prevention program; could have occurred for a number of other reasons
- How a program is implemented has implications for what we can learn
  - Randomized design is arguably the best way to learn
- Goal here is to present various options

Outline

- Issues that arise in non-randomized evaluation
  - Challenge of finding counterfactual scenario; selection bias
- How can we use randomization to evaluate health programs?
  - Five strategies to consider
- Challenges and issues

NON-RANDOMIZED EVALUATION: KEY ISSUES

Overarching objective of impact evaluation (IE)

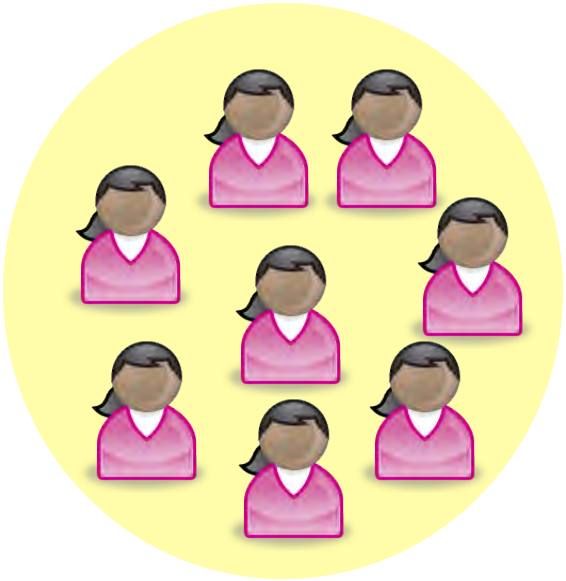
- To identify **causal effect** of intervention(s)
- **Counterfactual**: Need to find out what would have happened without intervention(s)
  - Cannot observe the same person or community with and without the program at the same point of time
- Simple comparisons of intervention and non-intervention groups can be problematic

Selection bias

- Selection bias can be a major issue in evaluations
  - BCC started at specific times/places for certain reasons
  - Participants may self-select into HIV testing
- Reason for self-selection may be correlated with outcome
  - Possible to statistically "control" for observed factors
  - But harder to control for unobserved factors
- Estimated impact of intervention may be biased

What we can observe…

Population



HIV-negative women aged 14-25 years

Matched for:

- socio-economic status
- education level
- ethnicity
- self-reported number of sex partners
- marital status
- other health status

What we may not observe…

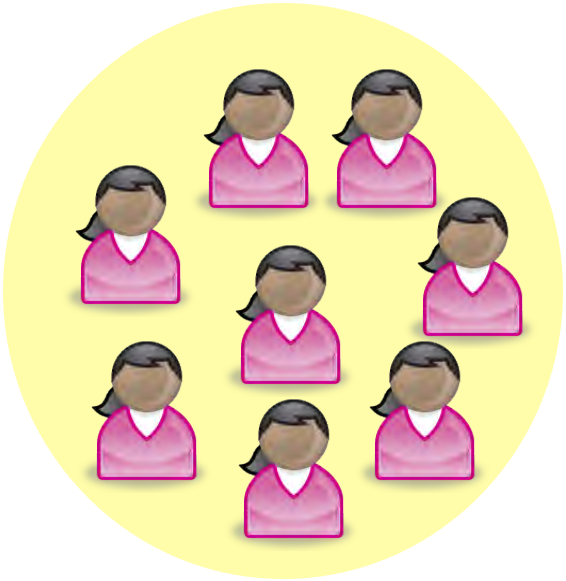
Population

Characteristics such as:

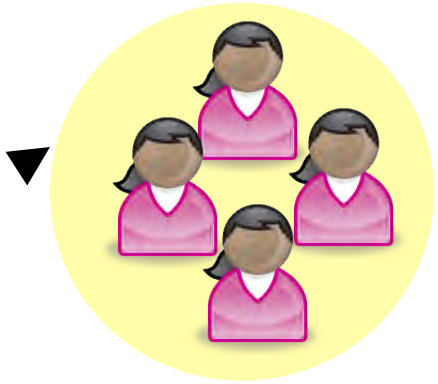
- Risk-related attitudes
- Decision-making power in household
- Other programs/interventions available to them

We can compare women with similar observables

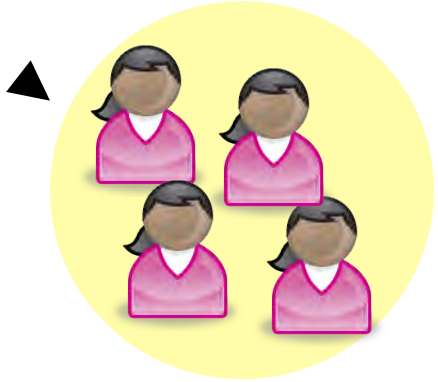
Population



Evaluation



Received intervention



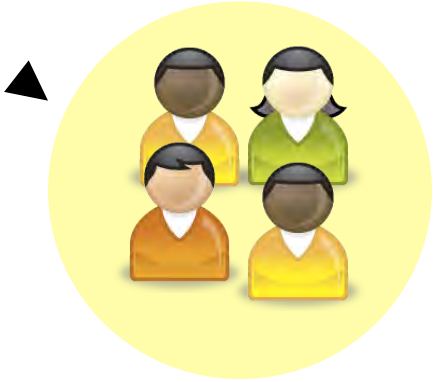
No intervention

But unobservables may vary between the two groups

Population

Evaluation

Received intervention



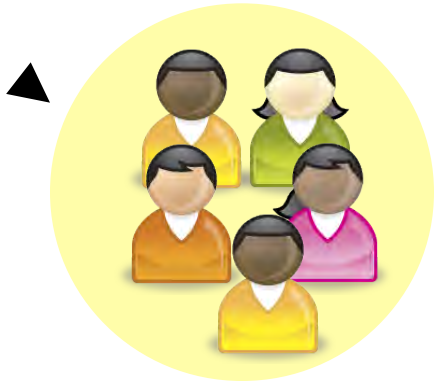
No intervention

With randomization, we can ensure that both groups are similar

Population

Evaluation

Received intervention



No intervention

Using randomization to develop proper counterfactual

- When done properly, randomization can provide comparison group that serves as valid counterfactual

Randomized evaluation designs

What unit and method of randomization is best?

- **Simple randomization – individual or cluster-level**
- **Stepped wedge/randomized phase-in**
- **Selective promotion/encouragement**
- **Dose-response**
- **Multiple treatments**

1. SIMPLE RANDOMIZATION (LOTTERIES)

Simple randomization/lottery

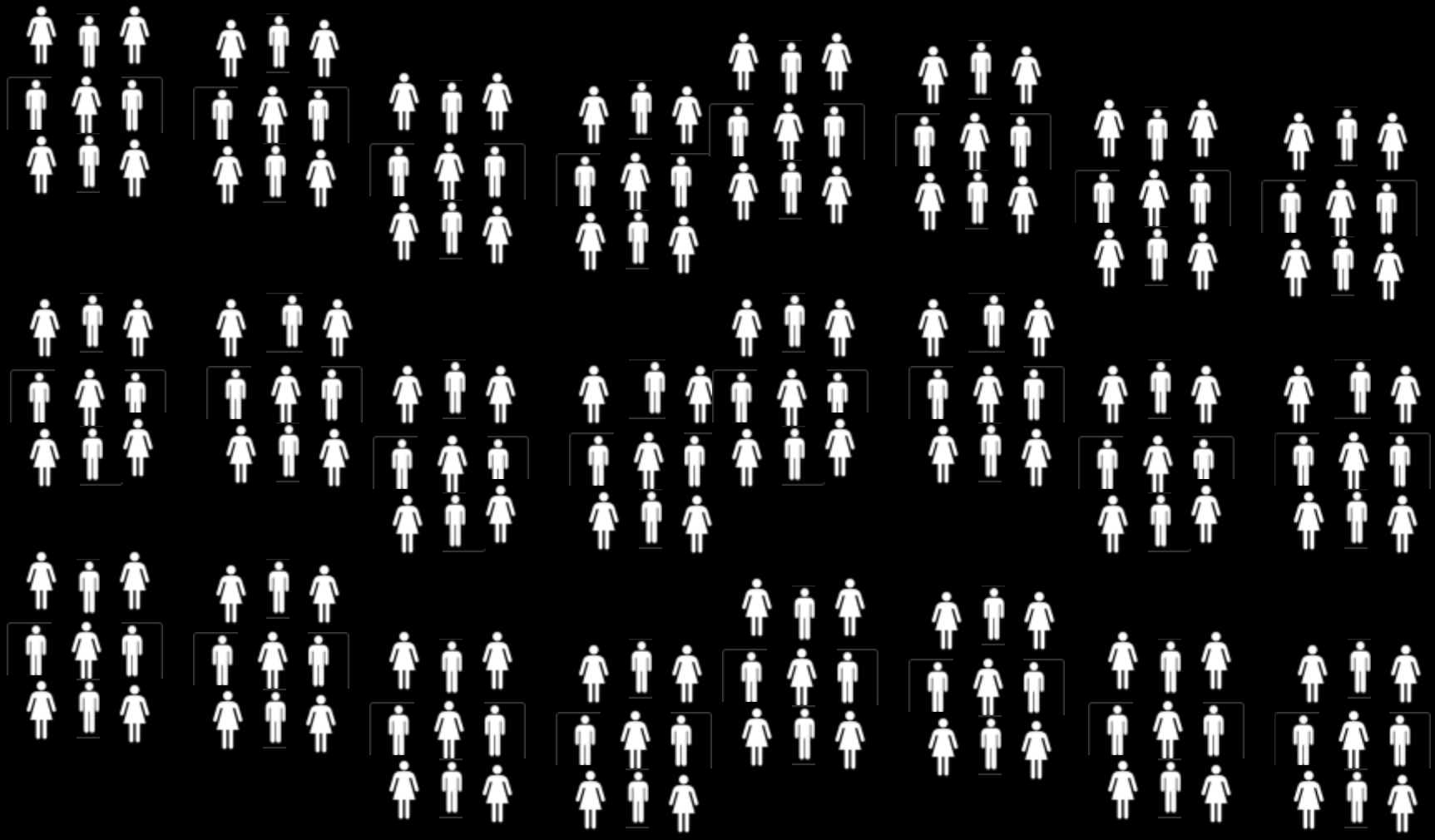
- Arguably, this is the most well known type of randomization design
- Individual-level randomization common in clinical trials to determine efficacy of new medication
- Advantages
  - Lotteries are simple, common and transparent
  - Randomly chosen from applicant pool or eligible list
  - Participants know the “winners” and “losers”
  - Simple lottery is useful when there is no a priori reason to discriminate
  - Often perceived as fair

Unit of randomization: options

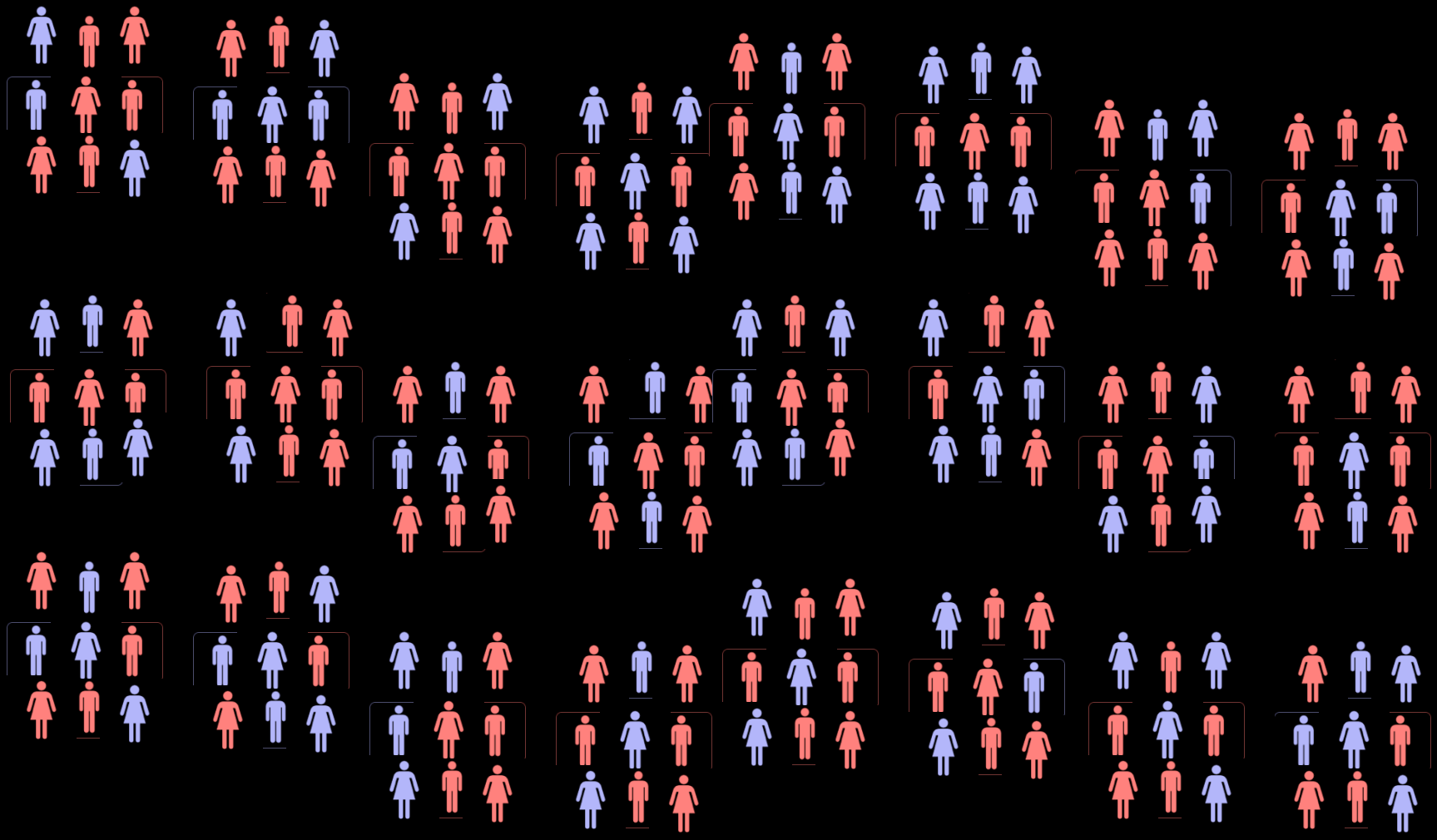
- One needs to decide the level at which randomization will take place
  - Randomizing at individual level
  - Randomizing at group level ("Cluster Randomized Trial")

- Which level to randomize?
  - What unit does the program target for treatment?
  - What is the unit of analysis?

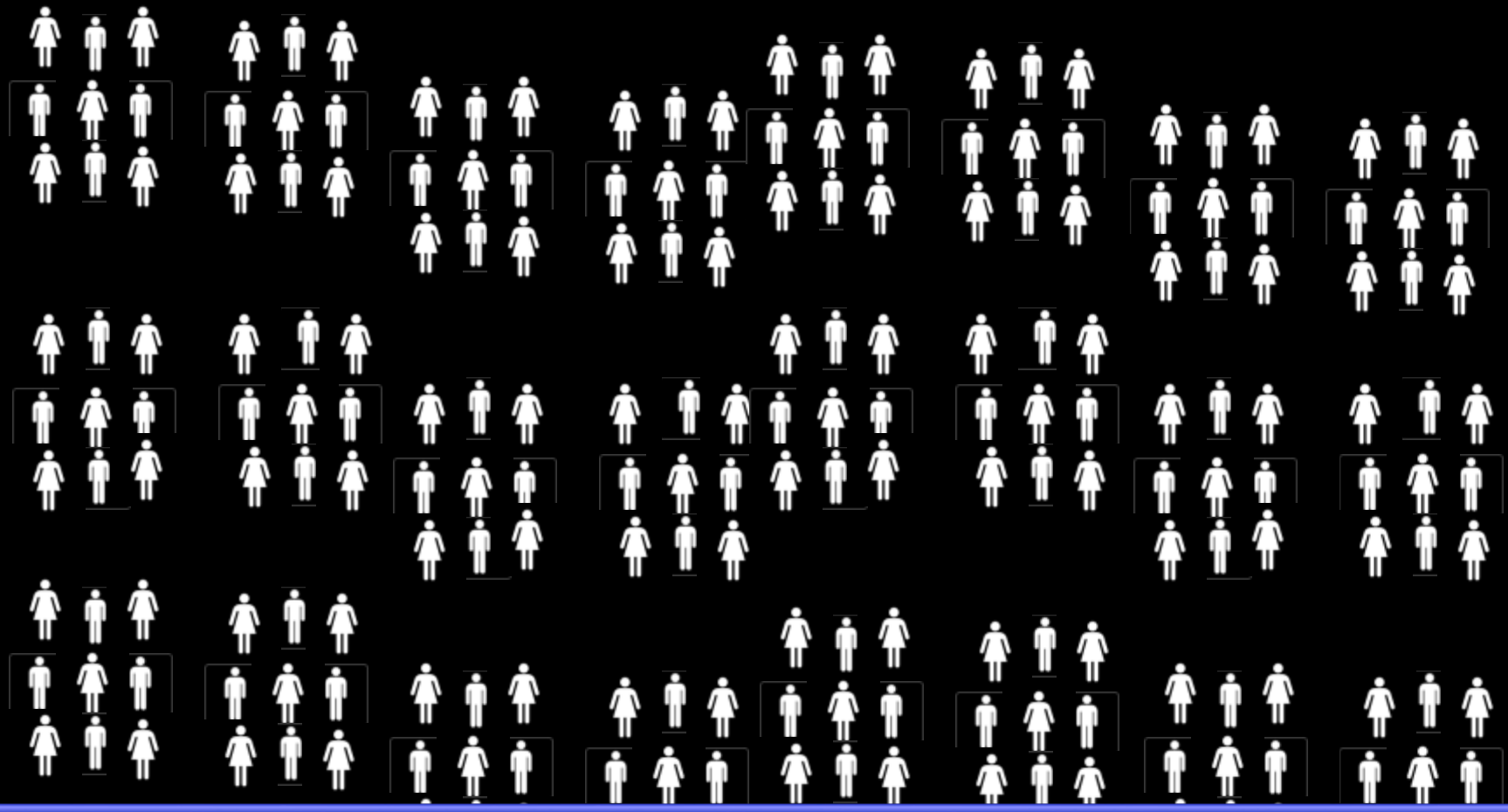
Unit of Randomization: Individual?



Unit of Randomization: Individual?

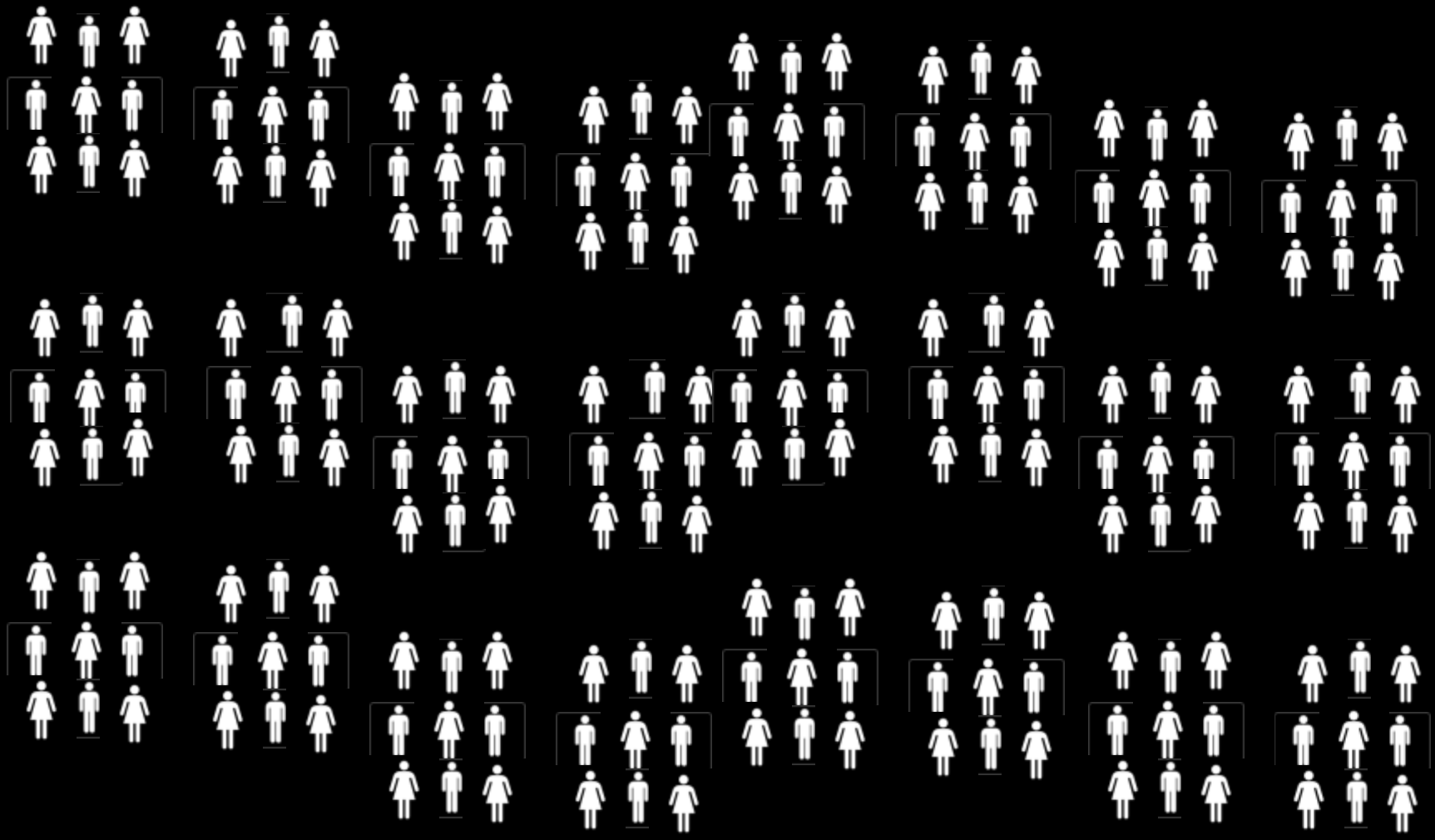


Unit of Randomization: Clusters?

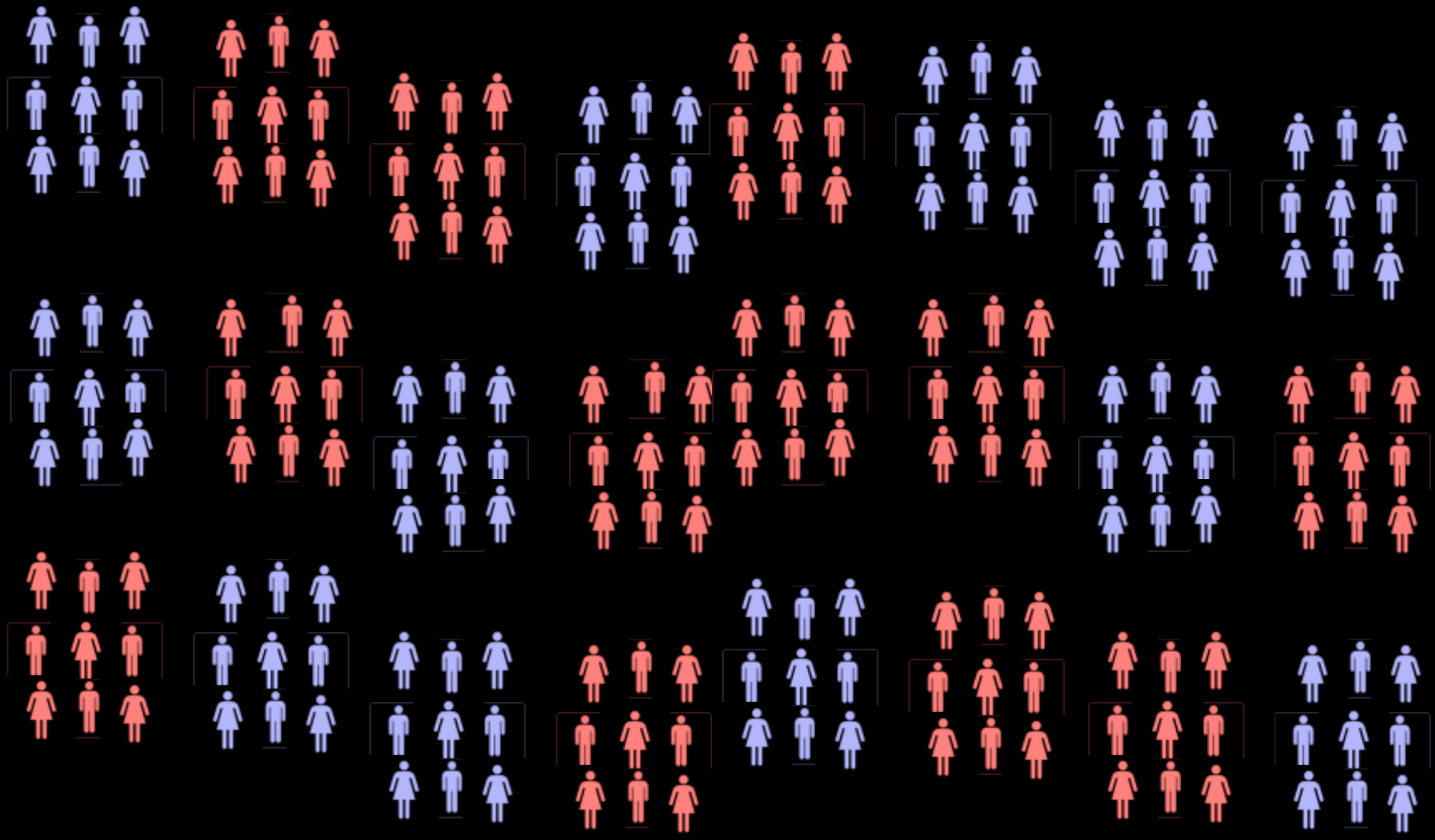


"Groups of individuals": Cluster Randomized Trial

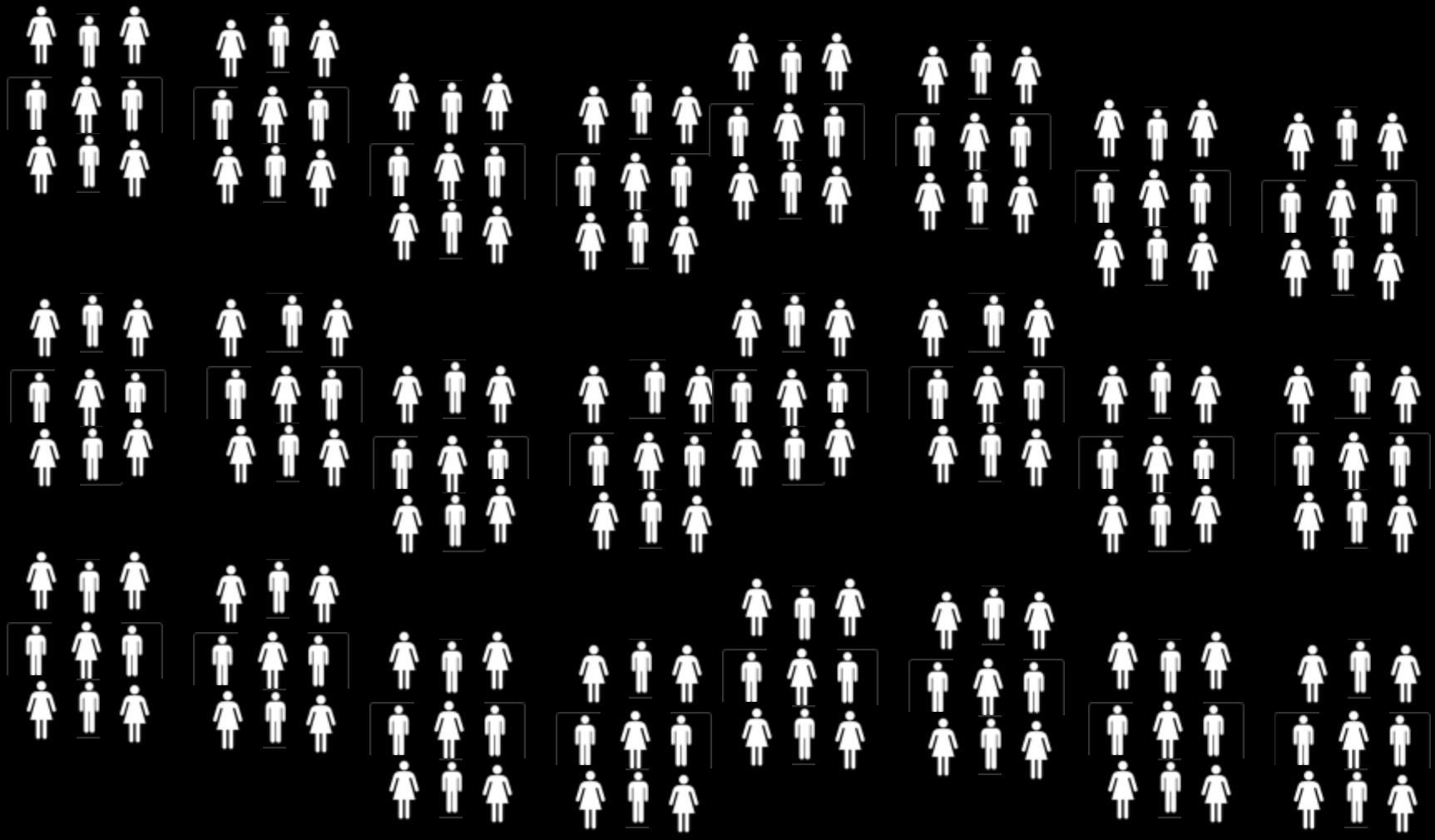
Unit of Randomization: Facility?



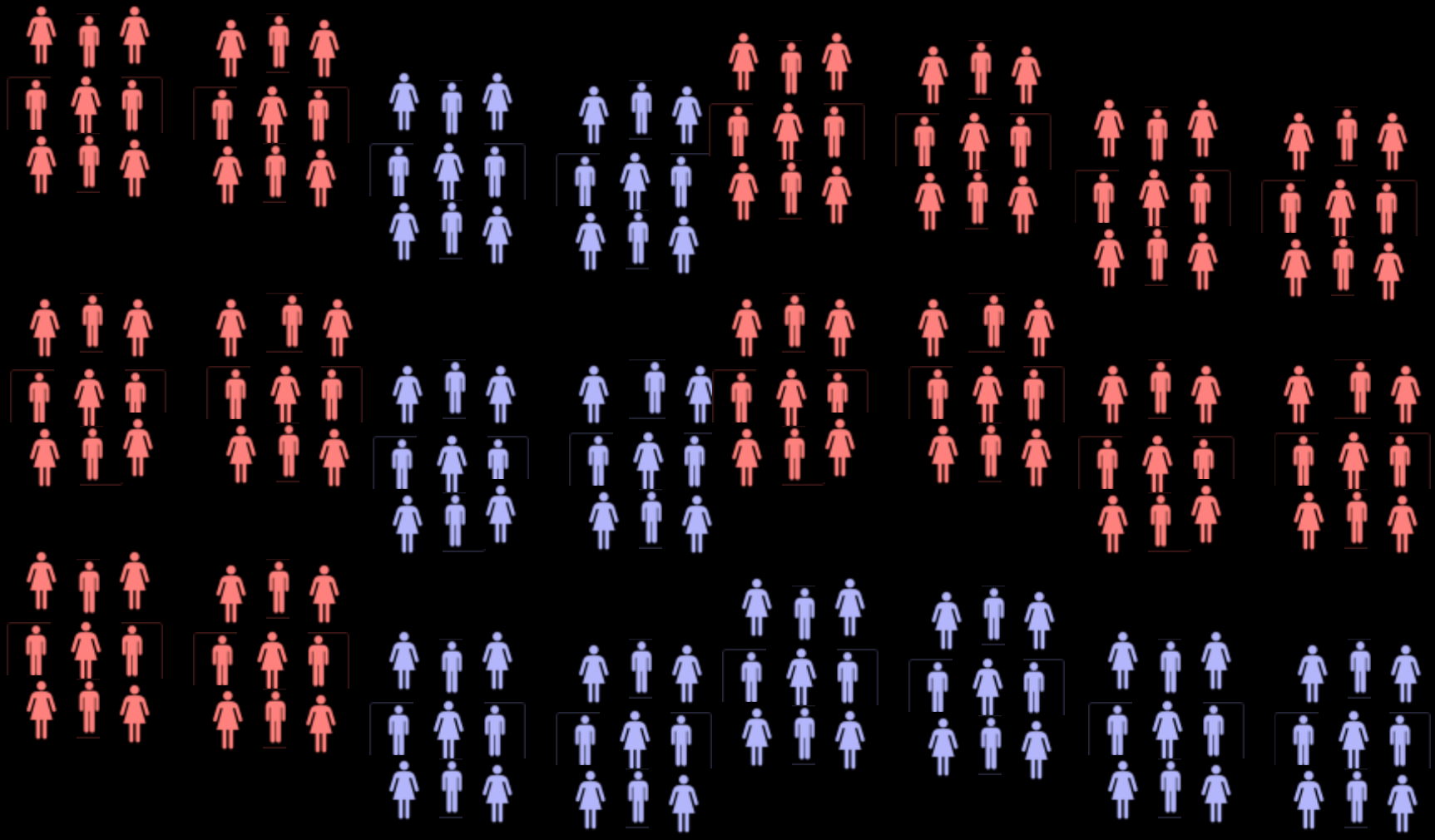
Unit of Randomization: Facility?



Unit of Randomization: District?



Unit of Randomization: District?



How to choose level of randomization

- Nature of the treatment/intervention
  - How is the intervention administered?
  - What is the catchment area of each "unit of intervention"
  - How wide is the potential impact?
- Power requirements
- Generally, best to randomize at the level at which the treatment is administered

When to choose individual/cluster randomization

- Most programs have limited resources
- More eligible recipients than resources will allow services for
- Learning rationale can also strengthen case for method that uses lotteries
  - Especially true it trying to decide whether to scale-up intervention

Cluster randomization

- Unit of randomization and unit of outcome measurement may differ
  - Evaluation study may measure outcomes for a sample of individuals within the group
- Effective sample size is number of clusters, not (just) number of individuals measured

2. STEPPED WEDGE DESIGN

Stepped wedge/randomized phase-in

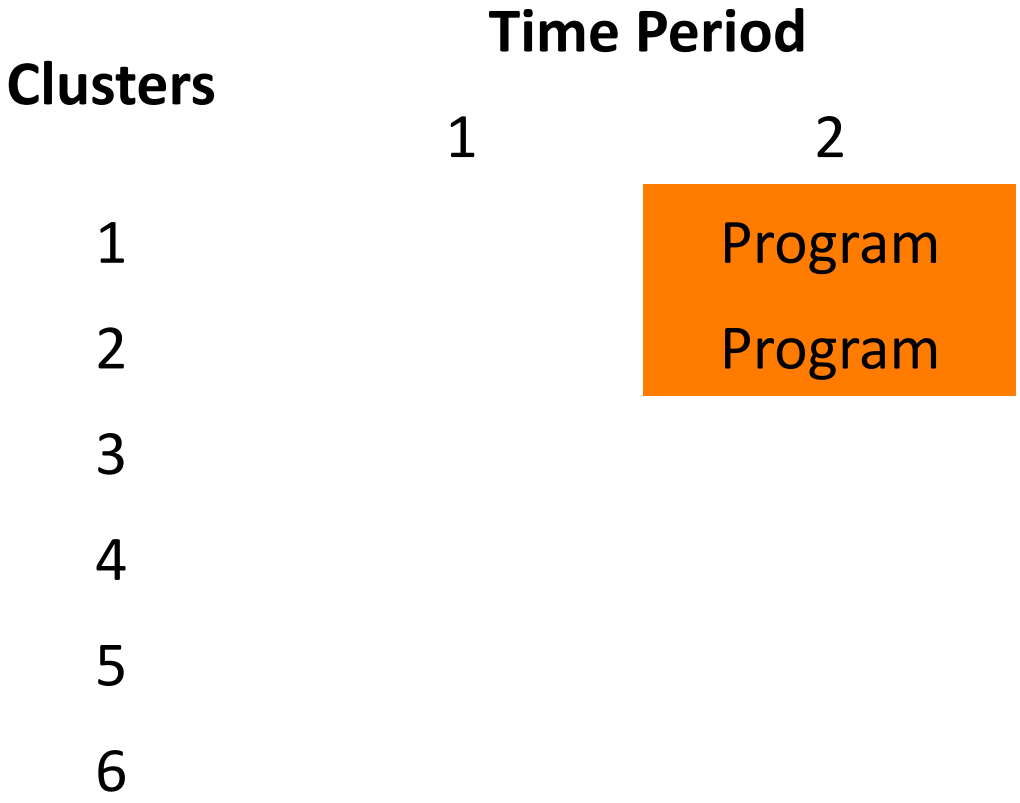
- What if a decision has been made to not withhold intervention from anyone?
- Often, a scale-up decision has been made but financial/logistical constraints limit ability to introduce program everywhere
- Randomized evaluation may still be possible
- Order of program roll-out is determined randomly
- Eventually, all communities receive program

Stepped wedge/randomized phase-in

Clusters	Time Period
	1
1	
2	
3	
4	
5	
6	

Brown CA, Lilford RJ. *BMC Medical Research Methodology,* 2006.

Stepped wedge/randomized phase-in



Brown CA, Lilford RJ. *BMC Medical Research Methodology,* 2006.

Stepped wedge/randomized phase-in

Clusters	Time Period 1	2	3
1		Program	Program
2		Program	Program
3			Program
4			Program
5			
6			

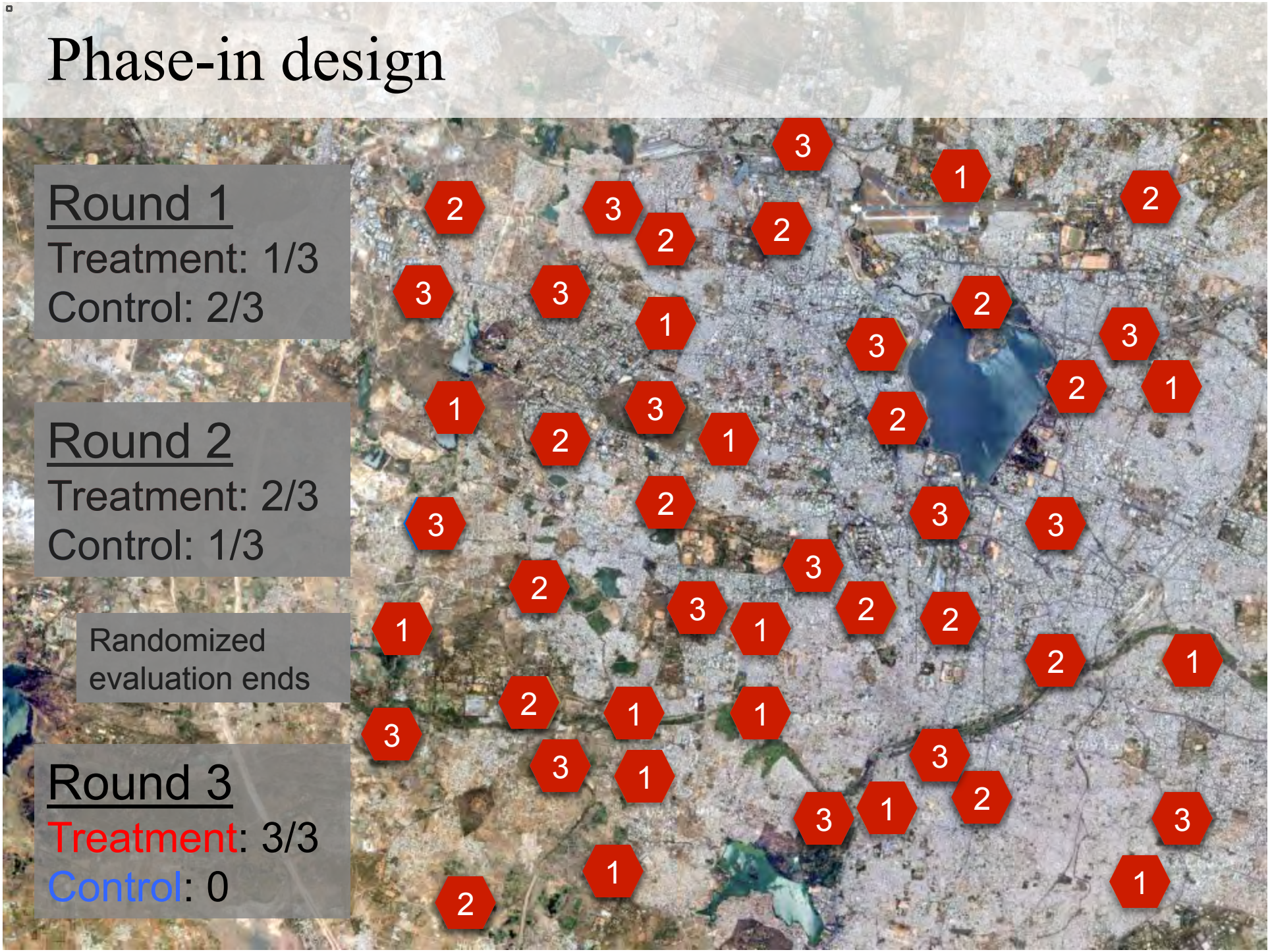
Brown CA, Lilford RJ. *BMC Medical Research Methodology,* 2006.

Stepped wedge/randomized phase-in

Clusters	Time Period 1	2	3	4
1		Program	Program	Program
2		Program	Program	Program
3			Program	Program
4			Program	Program
5				Program
6				Program

Brown CA, Lilford RJ. *BMC Medical Research Methodology,* 2006.

Phase-in design



Phase-in: takes advantage of expansion

- Everyone gets program eventually
- Natural approach when expanding program faces resource constraints
- What determines which slums, areas, villages, etc. will be covered in which year?

Concerns

- Can complicate estimating long-run effects
- Do expectations of the future change actions today?

Requirements and limitations

- Measure outcome on each unit at each time step
- Program must be rolled out in an area within the time step

- Key limitation: program/compoment must be effective within time step for impact to be detected

3. SELECTIVE PROMOTION

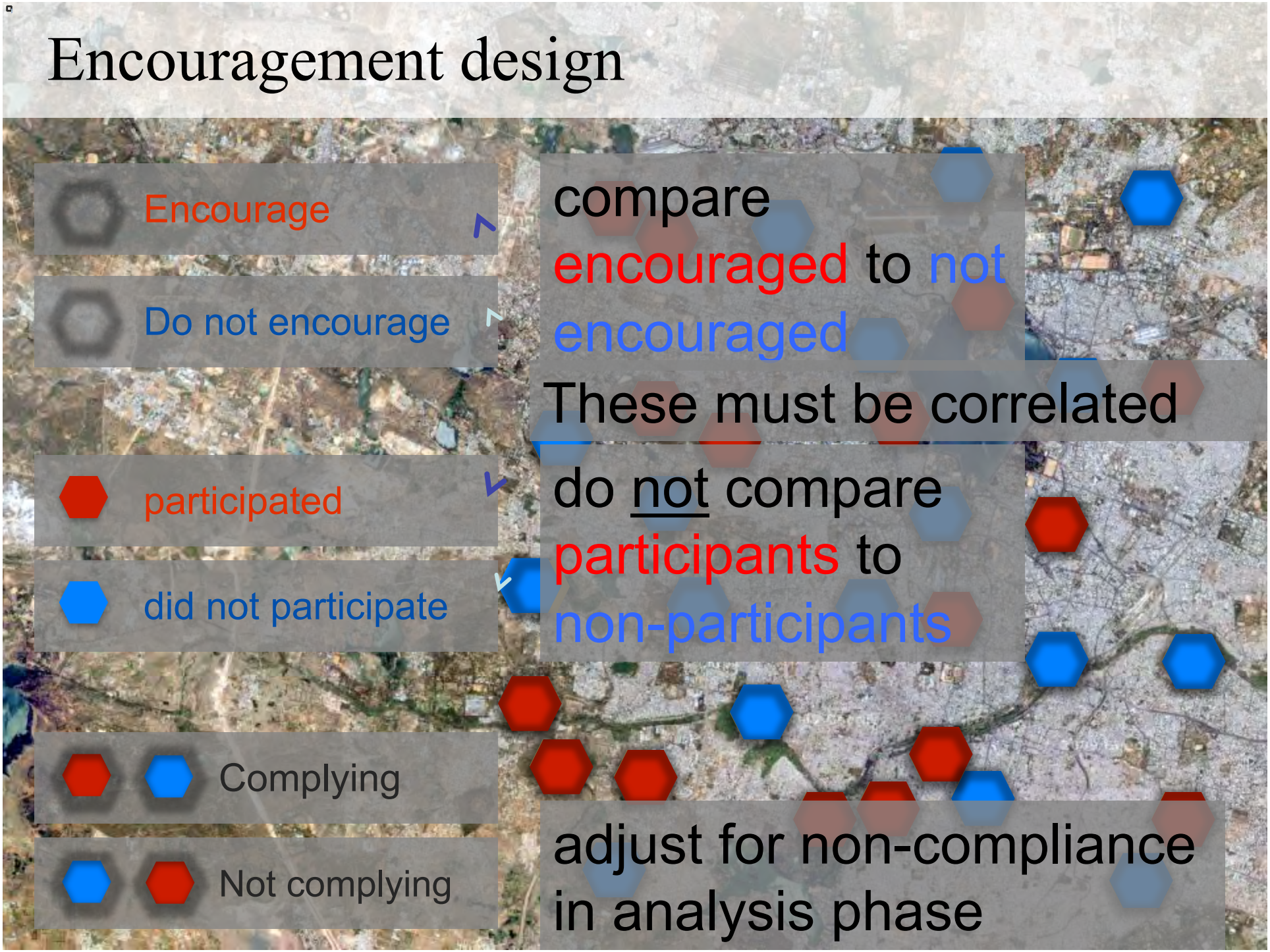
Encouragement design: What to do when you can't randomize access

- Sometimes it's practically or ethically impossible to randomize program access or even roll it out in a phased-in manner
- But most programs have less than 100% take-up, and this presents an opportunity to do a randomized evaluation
- Randomize encouragement to receive treatment
  - Information
  - Encouragement (small gift or prize)
  - Transport assistance

Selective promotion

- In this design, an intervention is made available everywhere and to everyone
- BUT, some communities/individuals (selected randomly) receive more information/incentives to uptake intervention
  - That is, intervention is "promoted" to some
- Promotion will result in a "treatment" group that gets more of intervention than "control" group

Encouragement design



Randomly promoting program

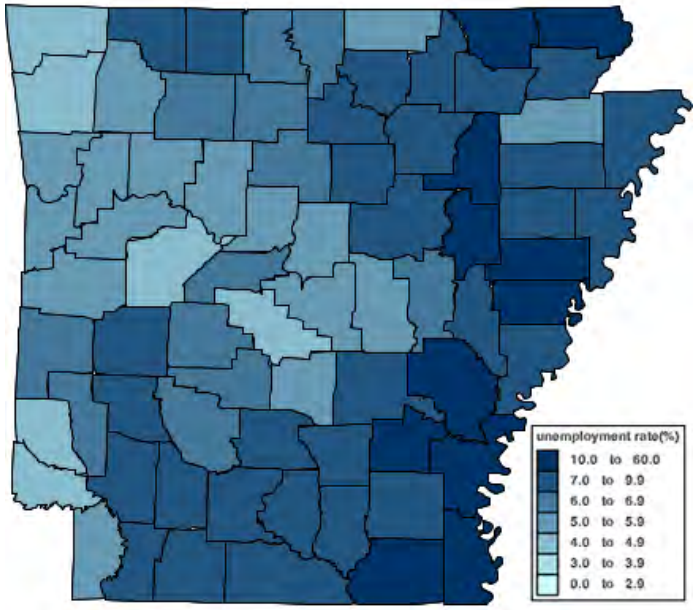
Necessary conditions

- Promoted and not-promoted groups are comparable
  - Whether or not you promote is not correlated with population characteristics
  - Guaranteed by randomization
- Promoted group has higher enrollment in the program
- Promotion of program does not affect outcomes directly

4. DOSE-RESPONSE

Dose-response evaluations

- Suitable when a program is already in place
- Examine differences in exposures (doses) or intensity across program areas
- Compare impact of program across varying levels of program intensity
- But this works best if dose is determined randomly



Hypothetical map

Dose–response evaluations

- This strategy can be helpful for identifying independent contribution of components
- Varying CD4 criteria in treatment for prevention
  - CD4<200 in some areas, test and treat in others
- Varying supply of MC
  - All fixed clinics in a region offer MC, but their capacity is limited and there are queues
  - Some areas are visited by mobile clinics that help rapidly increase MC coverage in those areas

Varying levels of treatment

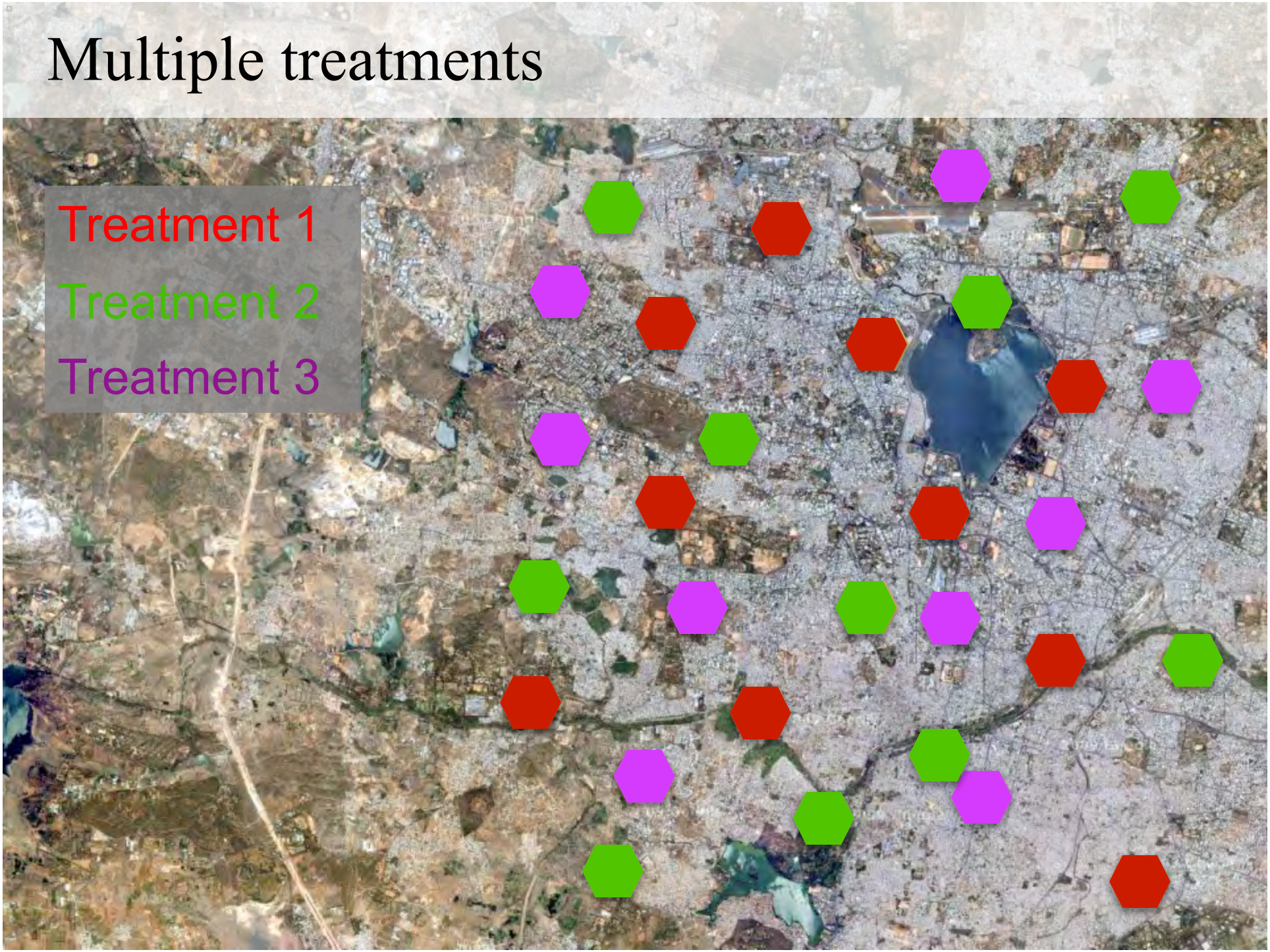
- Some schools are assigned full treatment
  - All kids get pills
- Some schools are assigned partial treatment
  - 50% are designated to get pills
- Testing subsidies and prices

5. MULTIPLE TREATMENTS

Multiple treatments

- Sometimes core question is deciding among different possible interventions
- You can randomize these programs
- Does this teach us about the benefit of any one intervention?
- Do you have a control group?

Multiple treatments



Cross-cutting treatments

- Test different components of treatment in different combinations
- Test whether components serve as substitutes or complements
- What is most cost-effective combination?
- Advantage: win-win for operations, can help answer questions, beyond simple “impact”

CAVEATS AND WAY FORWARD

Advantages of randomization

- Minimizes selection bias
  - Balances known and unknown confounders
- Simpler analyses and transparent results
- Decision makers understand (and believe) results
- Flexibility in design allows for multiple ways to estimate impact
- Impact of multiple interventions can be estimated

Some caveats of such designs

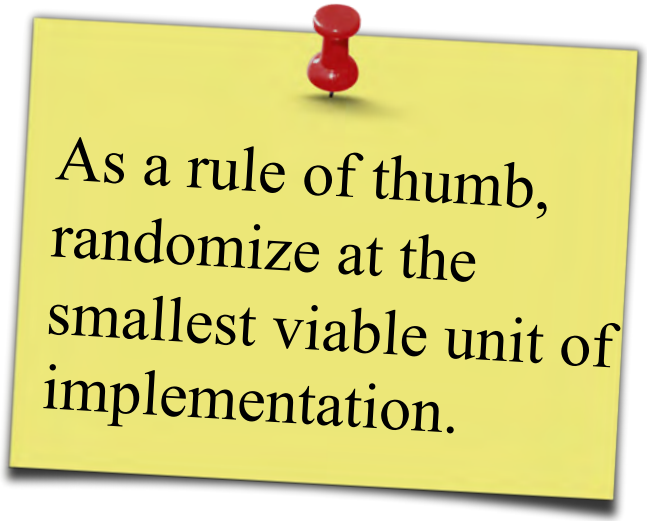
- Non-compliance
  - Not all treatment units will receive the treatment
  - Some control units may receive treatment
- Attrition: We may not be able to observe what happens to all units
- Hawthorne effect: just observing units makes them behave differently
- John Henry effect: the “controls” work harder to compensate

Choosing the most relevant design

- Context is important
- Design must take power considerations into account
- Generalizability of results is important
  - Internal validity: is result valid for “everyone” in **study population**?
  - External validity: is result valid at **entire population** level?

Unit of randomization

- Choose according to type of program
  - Individual/Household
  - School/Health Clinic/ catchment area
  - Block/Village/Community
  - Ward/District/Region



- Keep in mind
  - Need “sufficiently large” number of units to detect minimum desired impact: Power
  - Spillovers/contamination
  - Operational and survey costs

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
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