Policy Evaluation

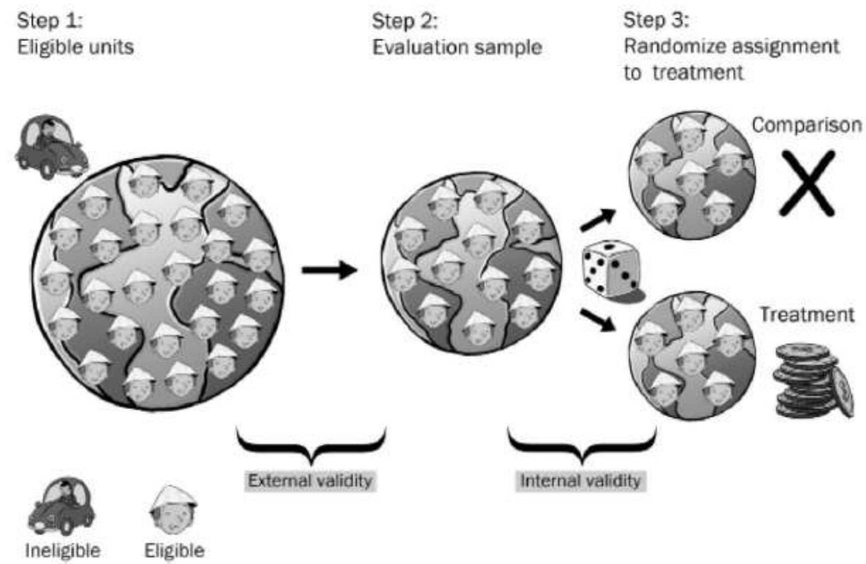
Lecture 3: Designing and Implementing an RCT

Edmund Malesky, Ph.D.

Duke University

Steps to Randomly Assign Treatment

1. **Define units eligible for program**
2. Determine sample size using power calculation
   - e.g. need larger N if minimum detectable effect small, Y rare or high variation, or if want to compare across subgroups
3. Select sample, ideally randomly
   - Use techniques from class
4. Assign T, C using transparent & ex ante rule for randomization
   - Coin, dice, lottery, random #
   - Record, or replicable w/ seed



Today's Class

- Real-world constraints
- Methods of randomization
- Variations on simple treatment-control

Constraints- Resources

- Most programs have limited resources
  - Vouchers, spaces in training programs, budget for community facilitators
- Results in more eligible recipients than resources will allow services for
- Limited resources can be an evaluation opportunity

Constraints-Fairness

- Lotteries are simple, common and transparent
- Useful when there is no a priori reason to discriminate
- Participants know the "winners" and "losers"
- Simple lottery is generally perceived as fair

Constraints: Contamination Spillovers/Crossovers

- Recall that the control group is meant to approximate the counterfactual
- If control group is different from the counterfactual, our results can be biased
- Can occur due to:
  - Spillovers
  - Crossovers

Spillovers

- In the presence of spillover effects, the simple treatment-control difference no longer gives the correct treatment effect.
  - Can be positive or negative.
- Spillover effects cause trouble for designs where the treatment saturation is blocked, but there are a couple of easy ways to use or create variation to measure them directly.

Experimental Estimation of Spillover Effects:

Miguel & Kremer, *'Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities'.*

- Deworming program randomized at the school level
- Controlling for the number of pupils within a given distance of an untreated unit, they look at how outcomes change as a function of the number of these pupils that were randomly assigned to treatment.
- Because treatment is randomized, localized intensity of treatment is incidentally randomized.

Baird, McIntosh, & Özler, *'Schooling, Income, & HIV Risk in Malawi'.*

- Conditional Cash Transfer Program run at the Village level
- Saturation of treatment in each village directly randomized to compare untreated girls in treatment villages to the control as a function of the share of girls in the corresponding village that were treated.

Constraints - Logistics

- Need to recognize logistical constraints in research designs.
  - E.g. individual de-worming treatment by health workers
  - Many responsibilities. Not just deworming.
- Serve members from both T/C groups
- Different procedures for different groups?

Constraints - Logistics

- Visibility of treatment status
- Randomizing at the child level within classes
- Randomizing at the class level within schools
- Randomizing at the community level

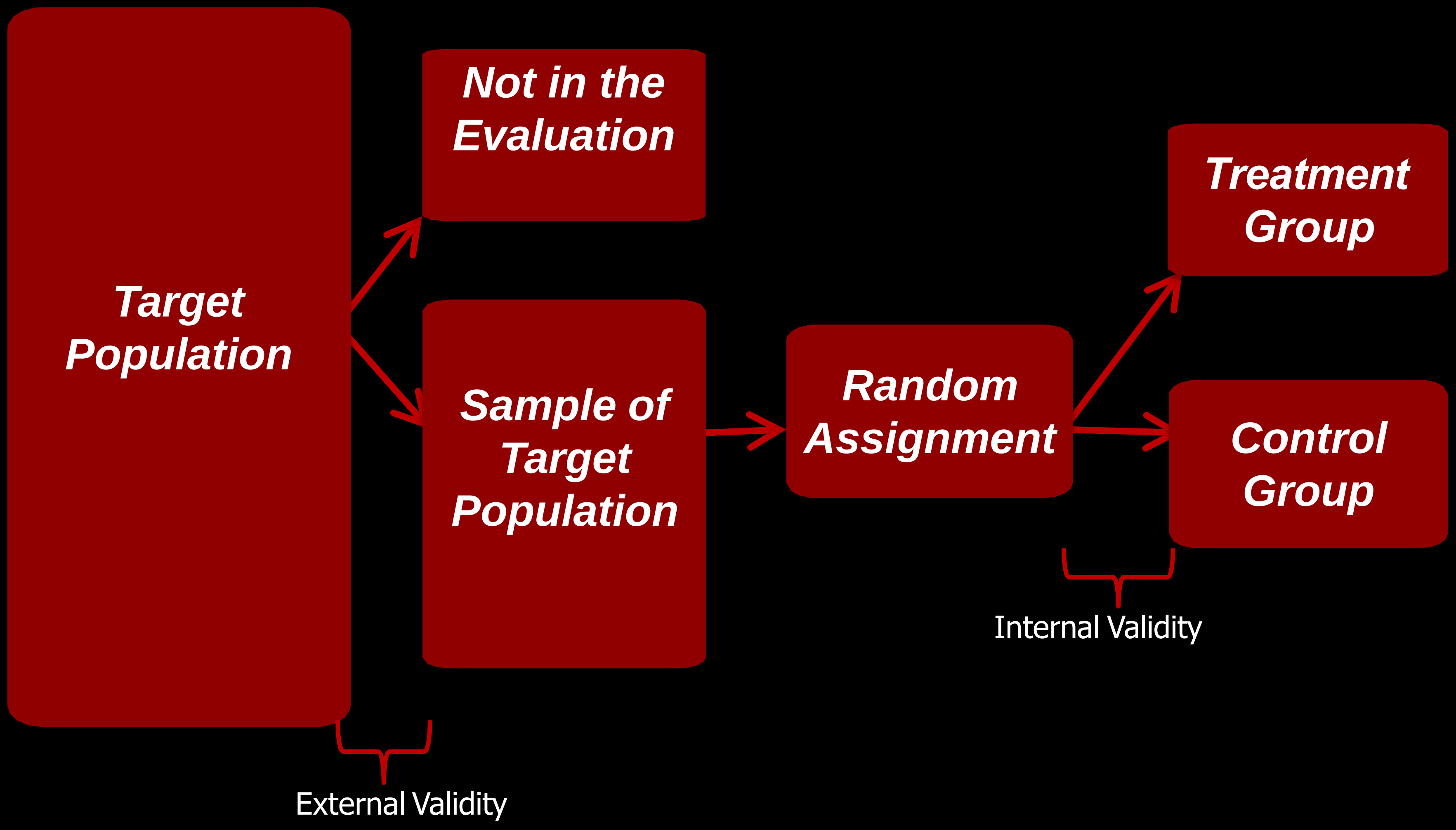
Constraints – Sample Size

- The program is only large enough to serve a handful of
- communities
- Primarily an issue of statistical power: too few observations means that we will be less likely to measure the impact with precision
- Desired sample size determined through a power calculation (not covered in this course)

Today's Class

- Real-world constraints
- Methods of randomization
- Variations on simple treatment-control

RCTs | Basic Structure



Basic RCT – Measuring Impact

- Data Required
  - Outcome data on treatment and control
  - Baseline data (if possible)
- Impact
  - Average Treatment Effect
    - Experiment - Conterfactual
    - Average Treatment minus Average Control

$$E[Y_1 - Y_0 | D = 1]$$
$$= \bar{y}_T - \bar{y}_C$$

RANDOM ORDER OF PHASE-IN DESIGN

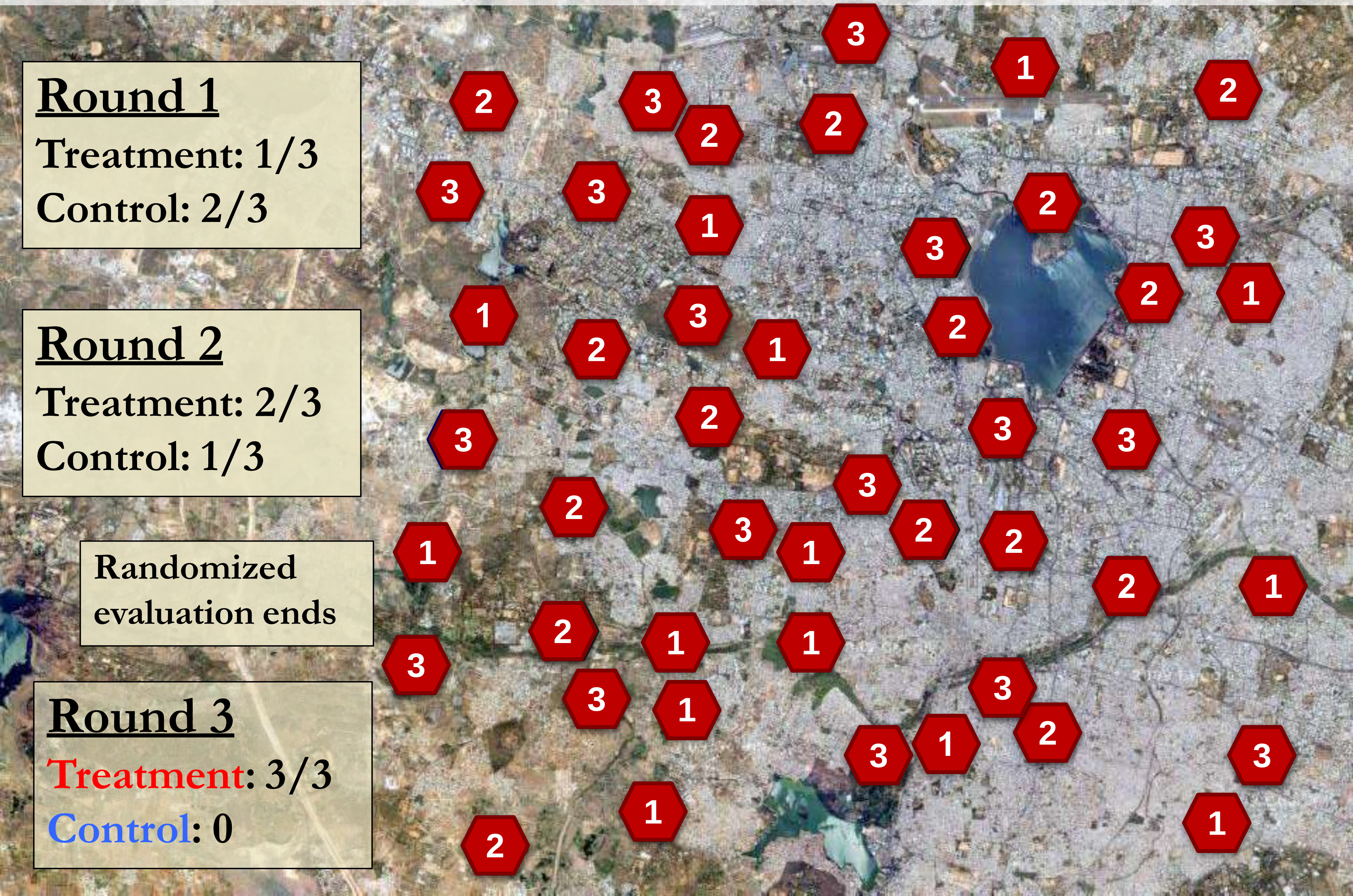
Phase-In: Takes Advantage of Expansion

- Ethical: Everyone gets program eventually
- Practical: Natural approach when expanding program faces resource constraints
- Randomization: What determines which schools, branches, etc. will be covered in which year?

Features of Phase-In Design

- **Counterfactual:**
  - After year 1, people/locations starting the intervention in Yr 2, 3… serve as the control group. After year 2, the participants starting the intervention in Yr. 3, 4… serve as the control group… and so on
- **Data required:**
  - Baseline (depending) and outcome data
- **Considerations:**
  - Over time, you loose the control group
  - Possible anticipatory effects by those to receive treatment in out-years

Phase-In Design



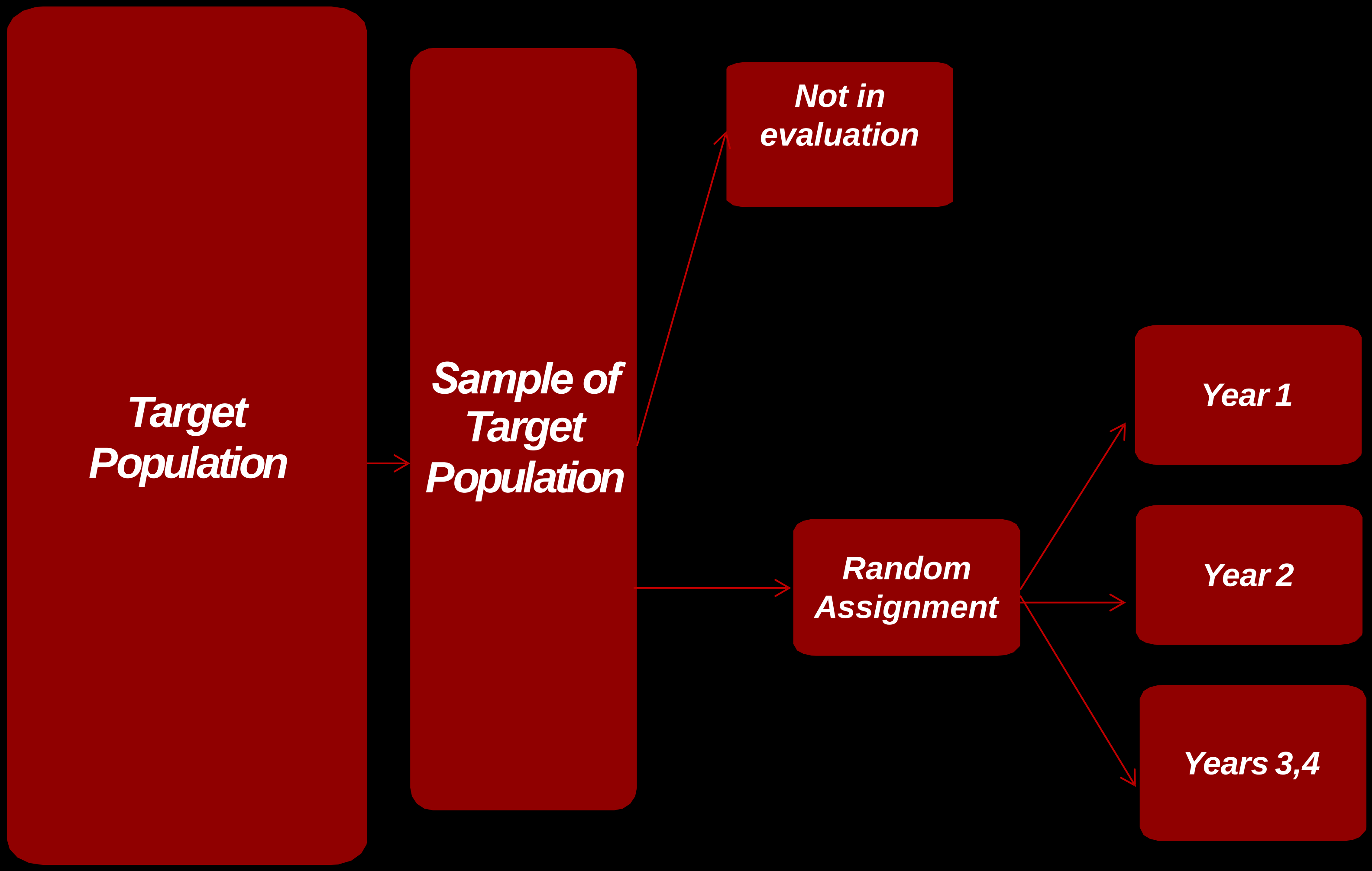
Round 1
Treatment: 1/3
Control: 2/3

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

Randomized evaluation ends

Round 3
Treatment: 3/3
Control: 0

RCTs | Phase In Design



Phase-In– Measuring Impact

- Impact
  - After Year 1: Average Treatment Group (those receiving in Year 1) minus those that will receive treatment in Year 2 &3.

$$E[Y_1 - Y_0|D = 1] = \bar{y}_{Y1} - \bar{y}_{Y2\&3}$$

  - After Year 2: Average Treatment Group (those receiving in Year 1 & 2) minus those that will receive treatment in Year 3 & 4.

$$E[Y_1 - Y_0|D = 1] = \bar{y}_{Y1\&2} - \bar{y}_{Y3\&4}$$

Phase-In: Pros and Cons

- Pros
  - Everyone gets something eventually Provides incentives to maintain contact
- Concerns
  - Can complicate estimating long-run effects
  - Over time, you loose the control group
  - Care required with phase-in windows
  - Do expectations of change actions today?
  - Possible anticipatory effects by those to receive treatment in out- years

ENCOURAGMENT DESIGN

Encouragment

- *What to do when you can't randomize access?*
  - Sometimes it's practically or ethically impossible to randomize program access
  - But many programs have less than 100% take-up
  - Randomize encouragement to receive treatment

What is Encouragment?

- Something that makes some folks more likely to use program than others
- Not itself a “treatment”
- For whom are we estimating the treatment effect?
- Think about who responds to encouragement

RCTs | Encouragement Design

- Data required:
  - Baseline (preferably) and outcome data for encouragement and non-encouragement groups
- Considerations:
  - The encouragement has to be calibrated to substantially increase enrollment,
  - The average treatment effect may be different between those enrolled because of encouragement (what we test) and the

Encouragement Design

Encourage

Do not encourage

participated

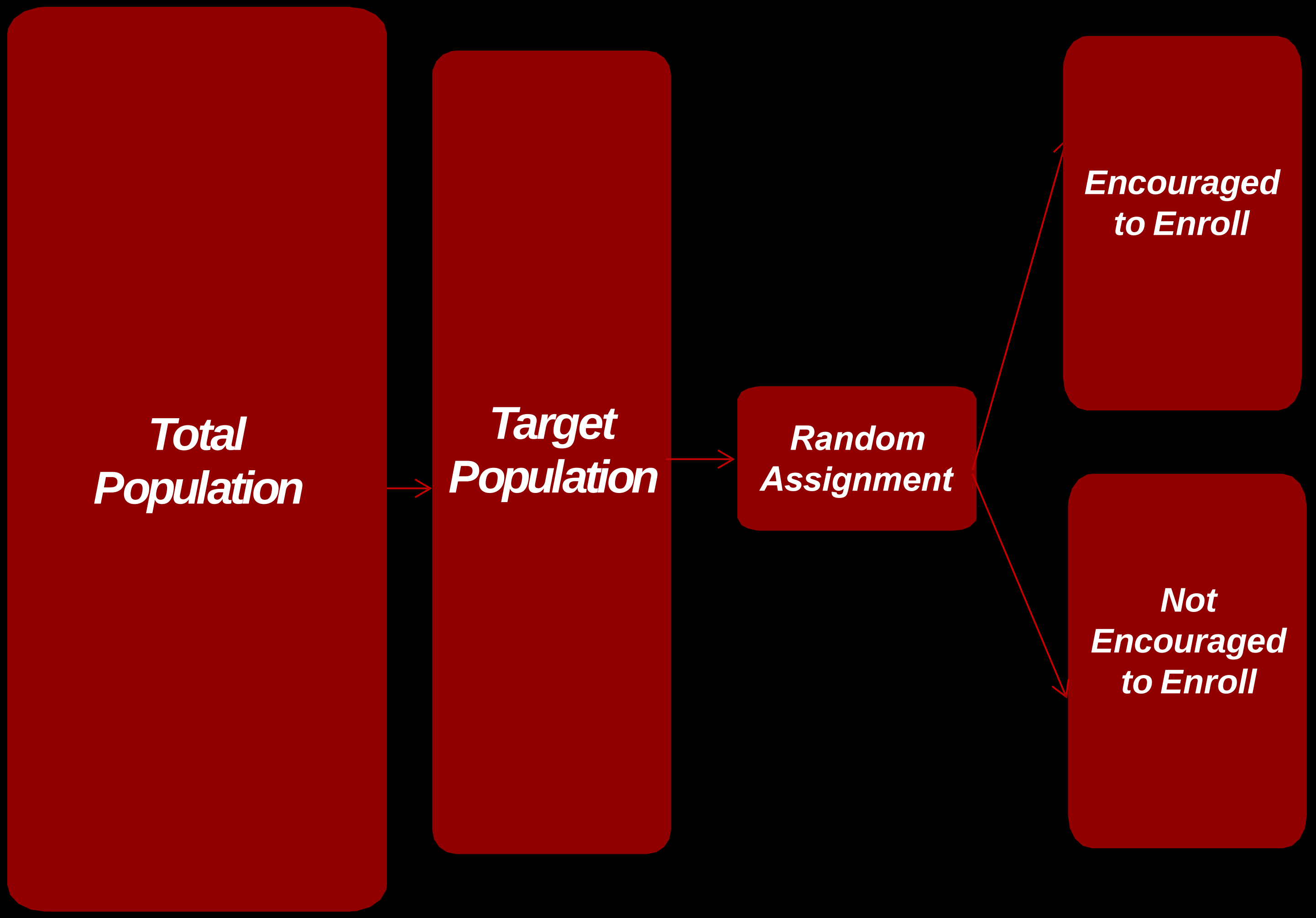
did not participate

compare
encouraged to not encouraged

These must be correlated

do not compare
participants to non-participants

RCTs | Encouragement Design



Encouragement– Measuring Impact

- Impact
  - Average Treatment Group (those with encouragement) minus those without encouragement.
  - Divide by percentage difference in enrollments (e).
  - Important: This is an intention to treat effect (ITE), not an ATE

$$E[Y\_1 - Y\_0 \ |D = 1 = \frac{y_E - \bar{y}_C}{e_E - e_C}$$

$$E[Y\_1 - Y\_0 \ |D = 1 = \frac{y_E - \bar{y}_C}{e_E - e_C}$$

Encouragement Design Impact Example

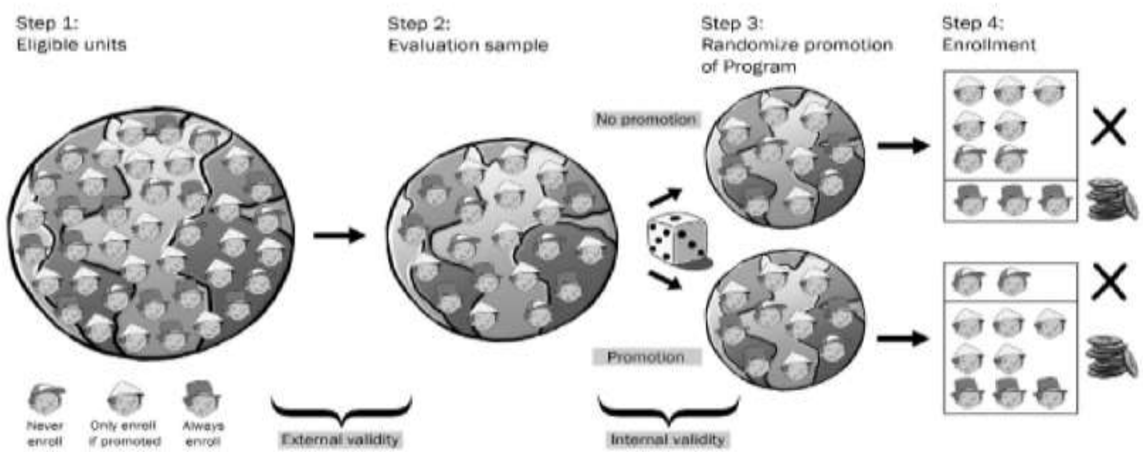
- You launch a universally available job training program and randomly assign certain areas in which individuals receive encouragement to enroll.

- You find that the overall percentage of the population that enrolls is 25% higher in encouragement areas. The average income in encouragement areas after one year is \$100; it is \$80 in non-encouragement areas.

- The impact (ITE) of the program is therefore: ($100-$80)/.25 = \$25

Encouragement Design/Randomized Promotion

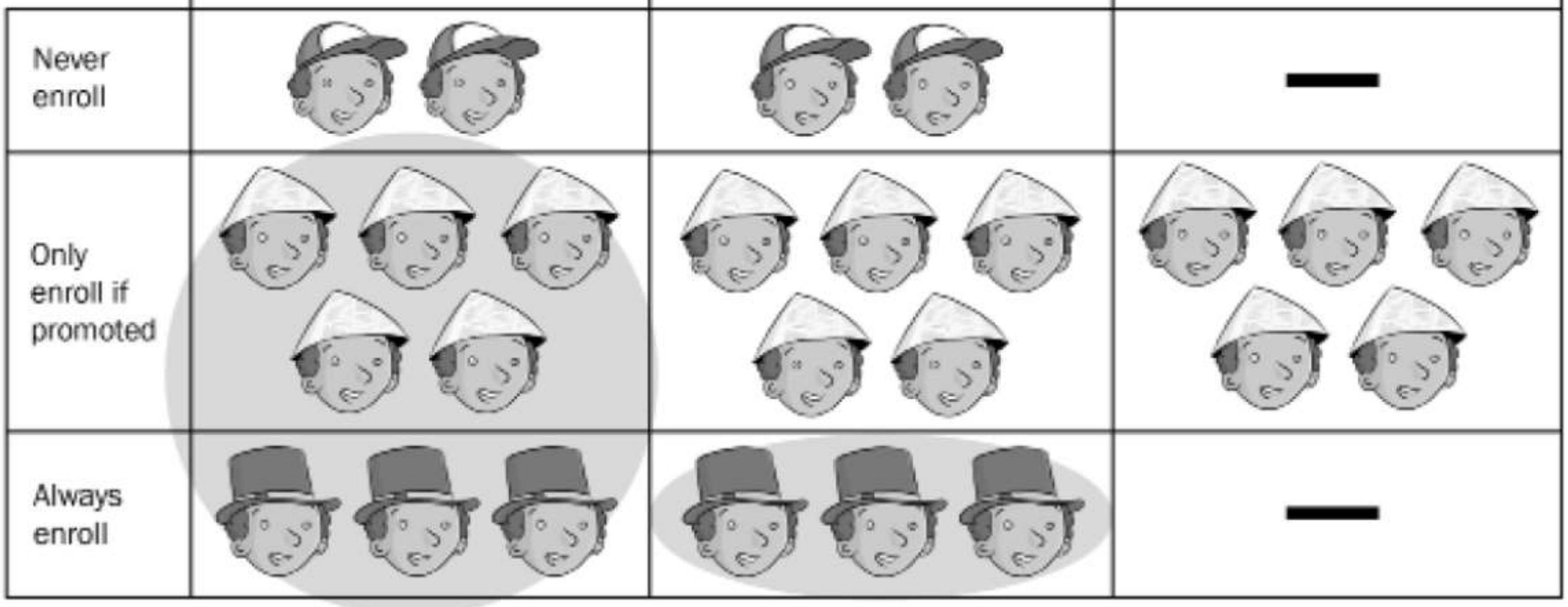
- Randomized promotion estimates impact if can't control participation
  - Budget exists, not politically or ethically OK to exclude
  - Randomly select units to offer *promo (*not treatment, as all can get)
- Info campaign or incentive increases take-up in random sample of population
  - Needs to increase enrollment, but not outcomes directly (not lots of $)
  - Promo affects *compliers*, though always-takers & never-takers exist
  - Creates equivalent of comparison group (promotion is an IV for T)



Source: Gertler, 2011.

Randomized Promotion

Promoted group	Non-promoted group	Impact
% enrolled = 80% Average Y for promoted group = 110	% enrolled = 30% Average Y for non-promoted group = 70	Δ% enrolled = 50% ΔY = 40 Impact = 40/50% = 80



Randomized Promotion: **Health Insurance Program**

Table 4.4 Case 4—HISP Impact Using Randomized Promotion (Comparison of Means)

	Promoted villages	Nonpromoted villages	Difference	*t*-stat
Household health expenditures baseline	17.1	17.2	−0.1	−0.47
Household health expenditures follow-up	14.9	18.8	−3.9	−18.3
Enrollment in HISP	49.2%	8.4%	40.4%	

Table 4.5 Case 4—HISP Impact Using Randomized Promotion (Regression Analysis)

	Linear regression	Multivariate linear regression
Estimated impact on household health expenditures	−9.4** (0.51)	−9.7** (0.45)

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

Today's Class

- Real-world constraints
- Methods of randomization
- Variations on simple treatment-control

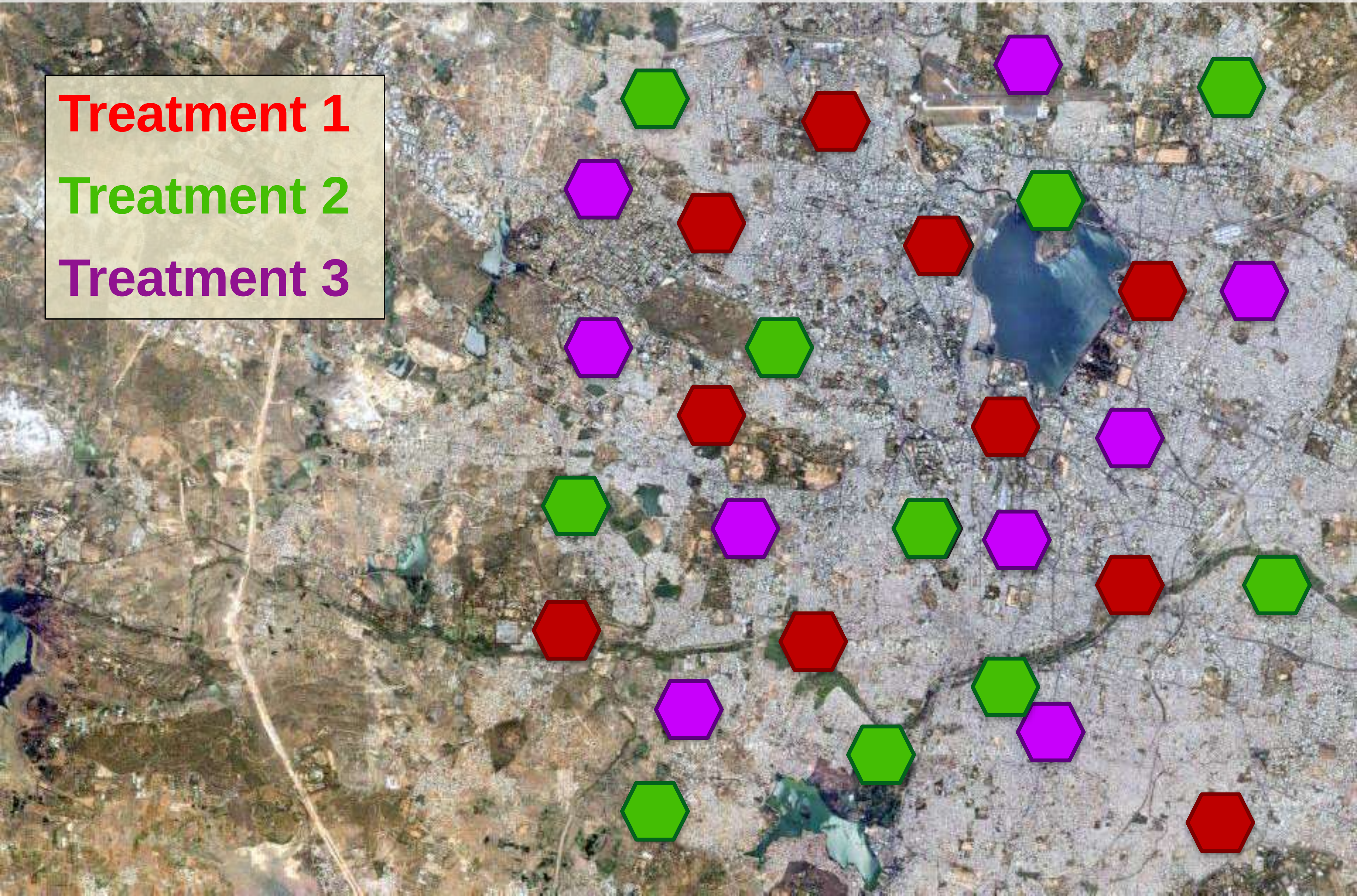
Multiple Treatments

- 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?

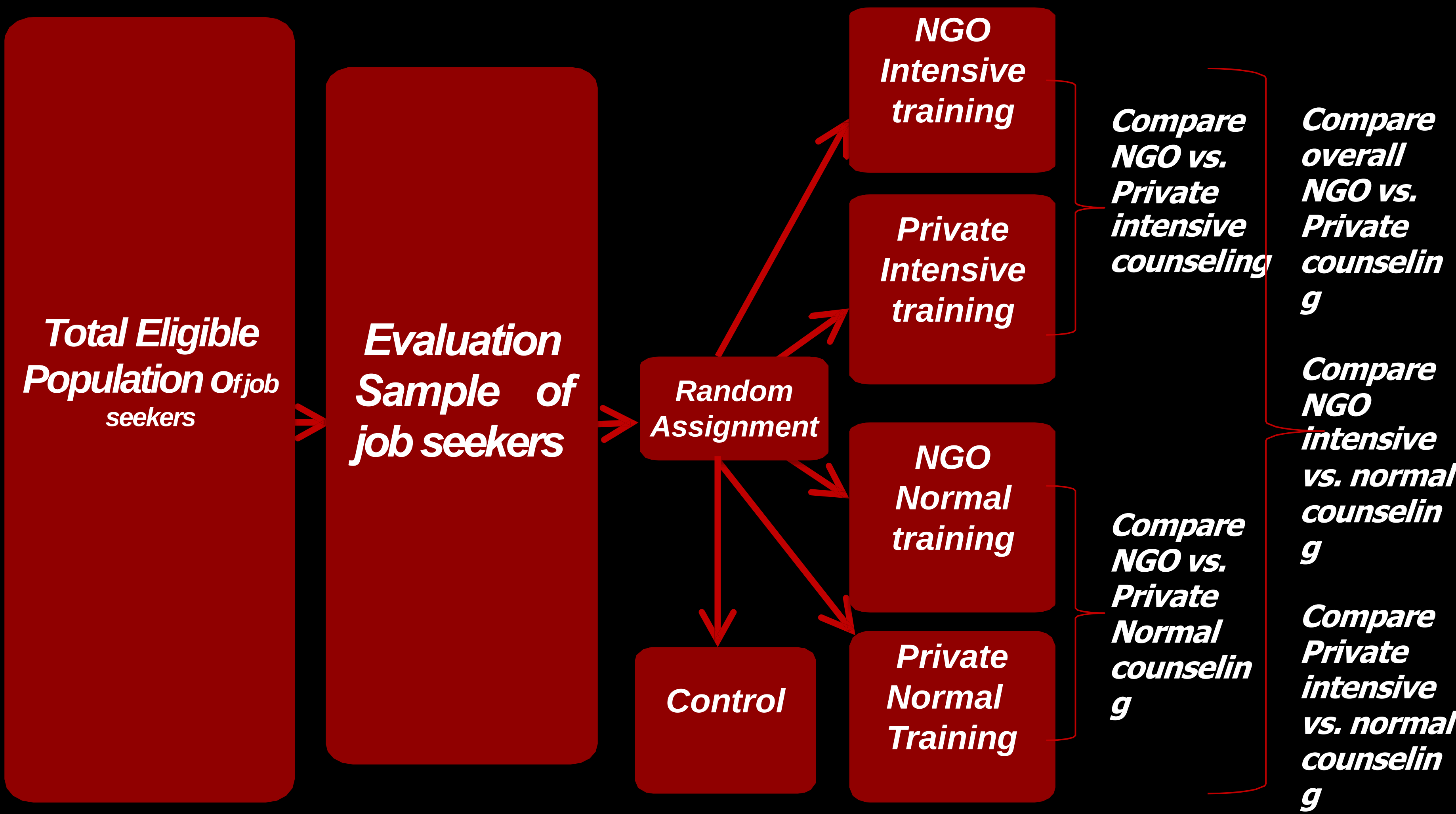
Multiple treatments

- 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



RCTs | Multi-Arm RCTs



Cross-cutting treatments

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

Factorial Design

2 Treatments

1. Training program for entrepreneurs
2. Micro-credit loan program

	Loans	No Loans
Training	Loans + Training	Training Only
No Training	Loans Only	No Training or Loans

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

Summing Up

- Implementation issues are critical
- Many different approaches to get around in-field problems.
- Design and analysis differ slightly depending on which experimental type you are using.
- Pay attention to ATE, ITE, TET, and LATE in write-up.