

Policy Evaluation
Lecture 4: Advanced RCTs

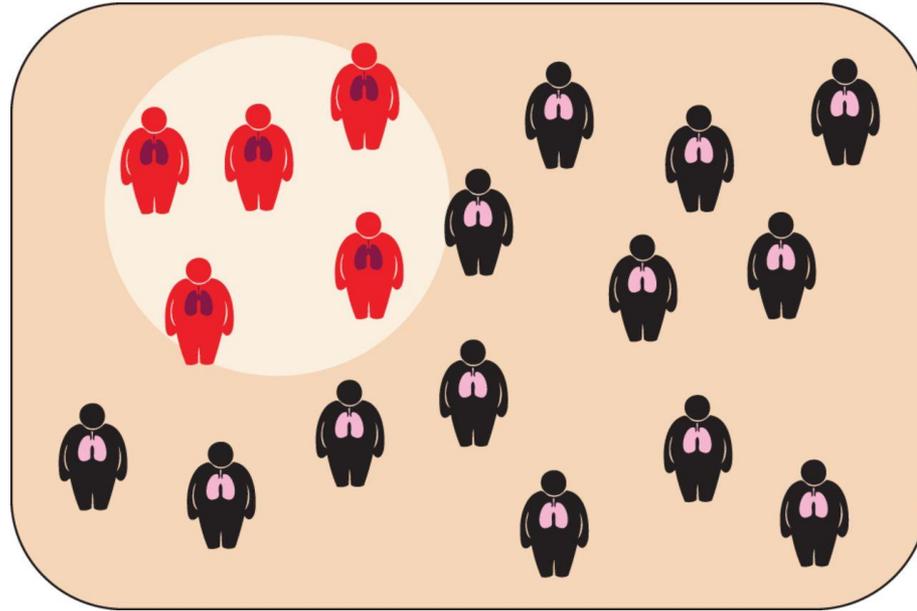
Edmund Malesky

July 6, 2020

Duke University & FUV

Special Issues

- Heterogenous Effects
- Multiple Treatments
- Spillovers
- Practical Issues
- Local Average Treatment Effects



SUB-GROUP ANALYSIS/ HETEROGENOUS TREATMENTS

What is a Heterogeneous Effect?

- Any given treatment might affect different experimental subjects in different ways.
 - For whom are there big effects?
 - For whom are there small effects?
 - For whom does treatment generate beneficial or adverse effects?
- Research on such questions can help inform theories about the conditions under which treatments are especially effective or ineffective.
- It can also help inform ways of designing and deploying policies so as to maximize their effectiveness

Conditional Average Treatment Effects (CATEs)

- A CATE is an average treatment effect specific to a subgroup of subjects, where the subgroup is defined by subjects' attributes (e.g., the ATE among female subjects) or attributes of the context in which the experiment occurs (e.g., the ATE among subjects at a specific site in a multi-site field experiment)

Using Interaction Effects

- In addition to CATEs, researchers are also interested in treatment-by-covariate interaction effects, or the difference between two CATEs when the covariate partitioning subjects into subgroups is not experimentally manipulated.
- The coefficient δ is the interaction effect and is interpreted as the difference between the ATE of treatment (X) among subjects in Z and the ATE of the job training program among subjects not in (Z)
- *If Z is not randomly assigned, not causal, but descriptive.*

$$Y_i = \alpha + \beta Z_i + \gamma X_i + \delta Z_i X_i + \varepsilon_i$$

Multiple Treatments

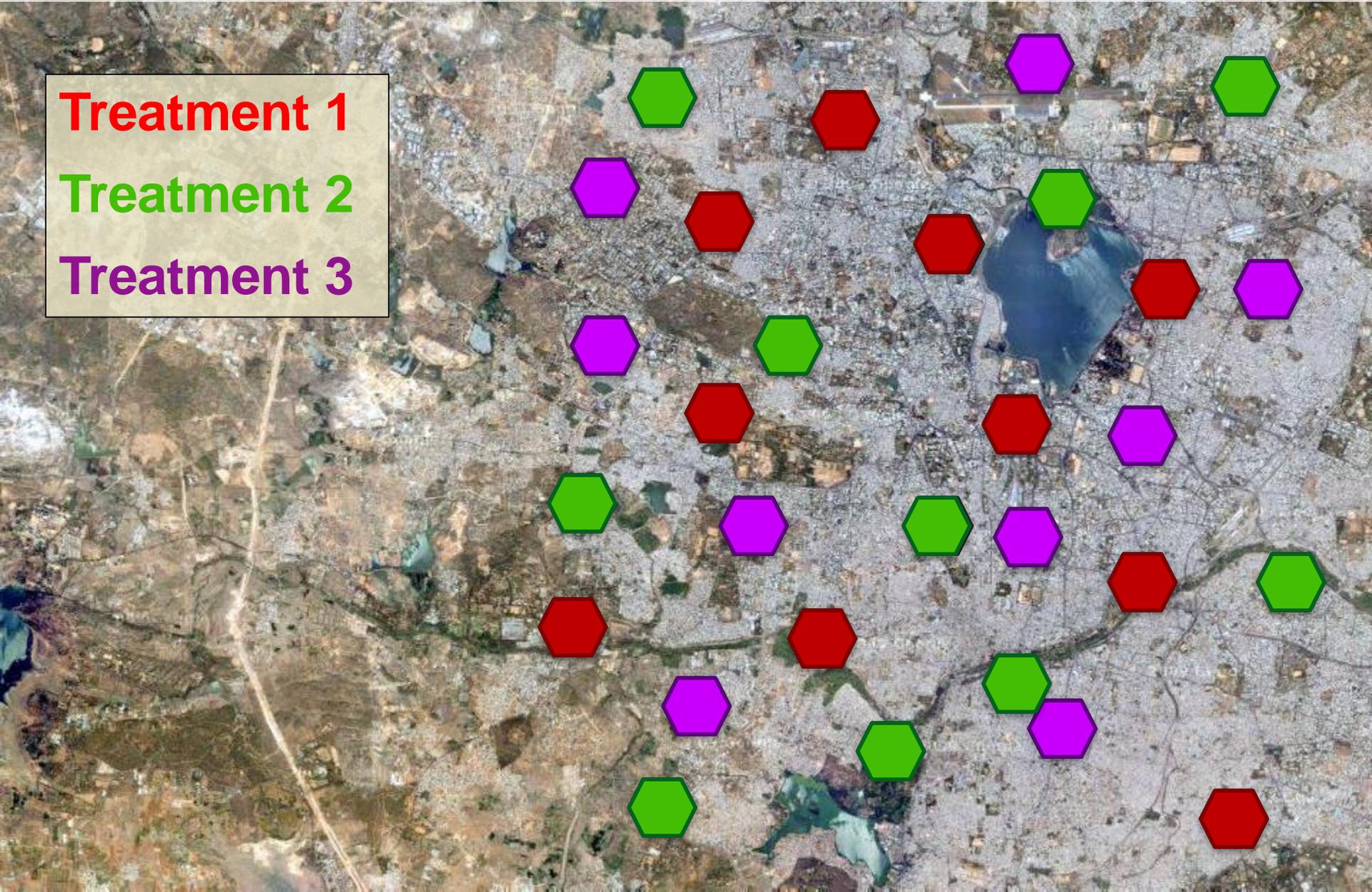
- Sometimes core question is deciding among different
- possible interventions
- You can randomize these programs
- Does this teach us about the benefit of any one intervention?
- Do you have a control group?

Multiple treatments

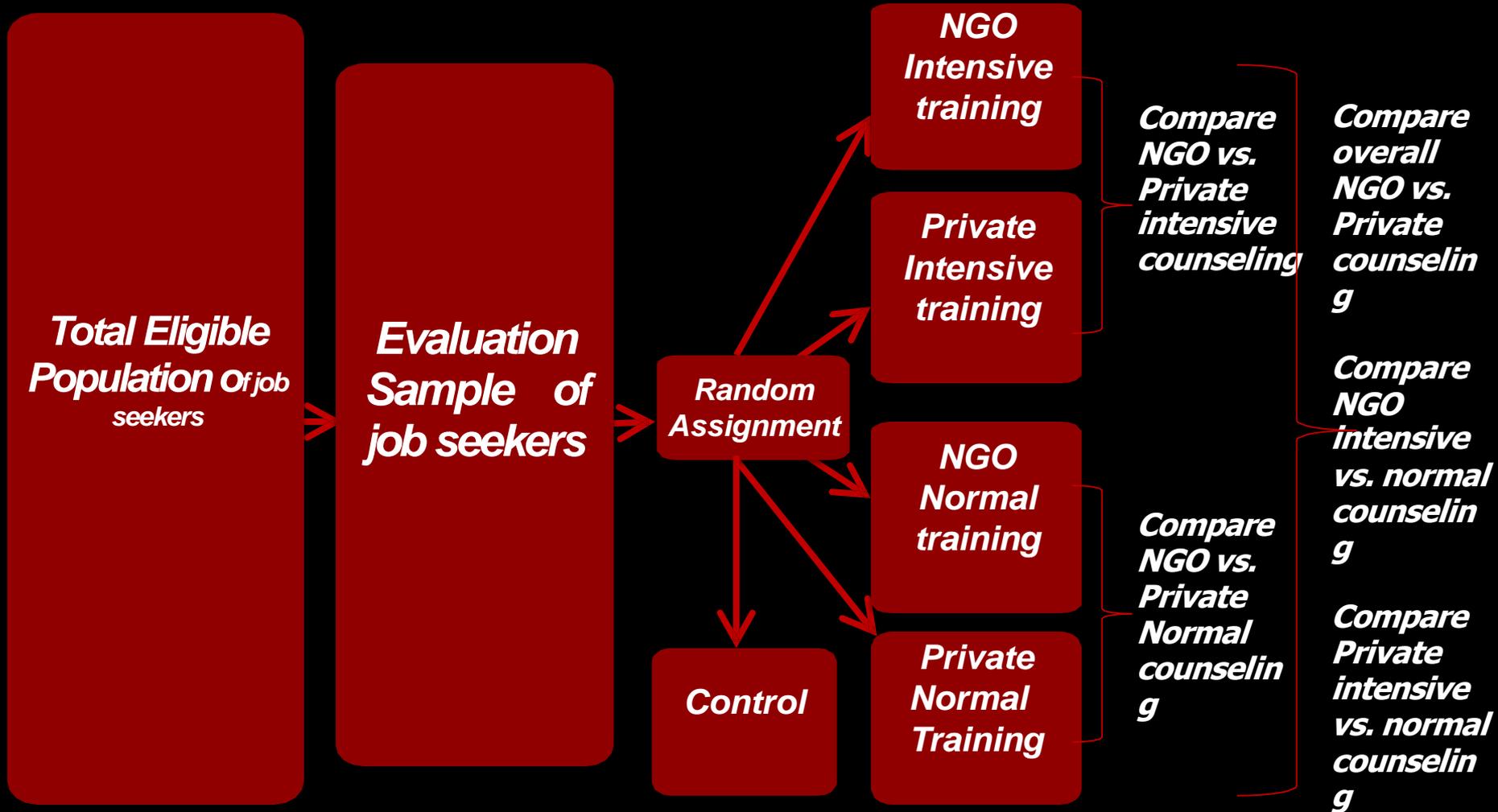
Treatment 1

Treatment 2

Treatment 3



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

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.



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

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.



FINAL PRACTICAL ISSUES

What is easily randomized?

1. Information:
 - Trainings
 - Political Message dissemination
 - Dissemination of information about politician quality, corruption
 - Mailers offering product variation
 - Promotion of the treatment.
 - Problem with all of these is that they may be peripheral to the key variation that you really care about.
 - This has led to a great deal of research which studies that which can be randomized, rather than that which we are interested in.
2. Decentralized, Individual-level treatments:
 - Makes evaluation of many of the central questions in policy difficult.
 - Voting systems, national policies, representative-level effects, international agreements not easily tractable.
 - Voter outreach, message framing, redistricting, audits much more straightforward.

Practical Issues in the Design of Field Trials:

1. Do you control the implementation directly?
 - If so, you can be more ambitious in research design.
 - If not, you need to be brutally realistic about the strategic interests of the agency which will be doing the actual implementation. Keep it simple.
 - Has the implementing agency placed any field staff on the ground whose primary responsibility is to guard the sanctity of the research design? If not, you MUST do this.
2. Does the program have a complex selection process?
 - If so, you must build the evaluation around this process.
 - Either pre-select and estimate TET, or go for ITE.
 - If uptake is low, you need to pre-select a sample with high uptake rates in order to detect the ITE
3. Is there a natural constraint to the implementation of the program?
 - If so, use it to identify:
 - ‘Oversubscription’ method
 - If rollout will be staggered anyway, then you can often motivate the rationale behind a randomized order to the implementer.

Imperfect Randomization

- Local Average Treatment Effect (LATE)
- Partial Compliance
 - Try to choose the design with the highest level of compliance
- Externalities
 - Spillover effects both within and outside groups.
 - If spillover is likely, design the program to address them (Miguel and Kremer).
- Attrition
 - Random attrition will impact calculation of standard errors.
 - Systematic attrition will actually bias results.
 - Make sure to record and track attrition.