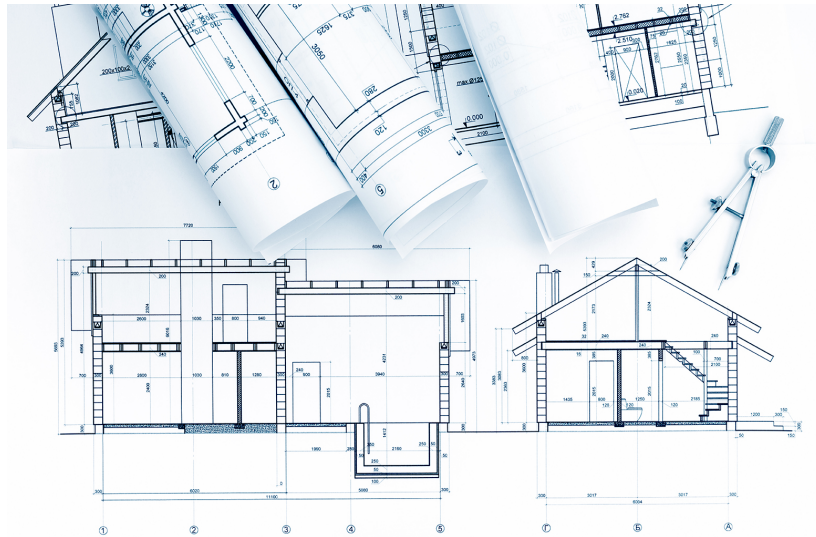# Threats and Analysis

J-PAL

# Course Overview

1. What is Evaluation?

2. Measurement: Outcomes, Impact, and Indicators

3. Why Randomize?

4. How to Randomize?

5. Sampling and Sample Size

6. Threats and Analysis

7. Start to Finish

8. Generalizability

# Introduction





The **conception phase** is important and allows us to design an evaluation enabling us to answer the research questions

But the **implementation phase** of the evaluation is also extremely important: many things can go wrong

# Objectives

- To be able to **identify** the main threats to validity during the implementation phase of the evaluation

- To define strategies to **mitigate** each of these threats

- To learn a few methods that can be used during **analysis** phase

# Lecture Overview

- Attrition

- Unexpected Spillovers

- Partial Compliance and Sample Selection Bias

    => Intention to Treat & Local Average Treatment Effect

- Behavioral Responses to Evaluations

- Research Transparency

# Lecture Overview

- Attrition

- Unexpected Spillovers

- Partial Compliance and Sample Selection Bias

    => Intention to Treat & Local Average Treatment Effect

- Behavioral Responses to Evaluations

- Research Transparency

# Attrition

- Is it a problem if some of the people in the experiment vanish before you collect your data?

  – It is a problem if the type of people who disappear is correlated with the treatment.

- Why is it a problem?

- Why should we expect this to happen?

# Attrition Bias: an Example

- The problem you want to address:
  - Some children don't come to school because they are too weak (undernourished)

- You start a school feeding program and want to do an evaluation
  - You have a treatment and a control group

- Weak, stunted children start going to school more if they are associated with a treatment school

- Measure effects on:

  - Enrollment/attendance

  - Child growth (e.g. weight of children)

- You go to all the schools (treatment and control) and <u>weigh</u> everyone who is in school on a given day

- Will the treatment-control difference in weight be over-stated or understated?

| | Before Treatment | | | After Treament | |
|---|---|---|---|---|---|
| | T | C | | T | C |
| | **20** | 20 | | **22** | 20 |
| | **25** | 25 | | **27** | 25 |
| | **30** | 30 | | **32** | 30 |
| Ave. | | | | | |
| Difference | | | Difference | | |

|  | **Before Treatment** | | | **After Treament** | |
| --- | --- | --- | --- | --- | --- |
|  | T | C |  | T | C |
|  | **20** | 20 |  | **22** | 20 |
|  | **25** | 25 |  | **27** | 25 |
|  | **30** | 30 |  | **32** | 30 |
| Ave. | **25** | **25** |  | **27** | **25** |
|  | Difference | 0 |  | Difference | 2 |

# What if Only Children > 21 Kg Come to School?

# What if Only Children > 21 Kg Come to School?

| | Before Treatment | | | After Treament | | |
|---|---|---|---|---|---|---|
| | T | C | | T | C | |
| | **20** | 20 | | **22** | 20 | |
| | **25** | 25 | | **27** | 25 | |
| | **30** | 30 | | **32** | 30 | |

A.  Will you underestimate the impact?

B.  Will you overestimate the impact?

C.  Neither

D.  Ambiguous

E.  Don't know

# What if Only Children > 21 Kg Come to School?

| | Before Treatment | | | After Treament | | |
|---|---|---|---|---|---|---|
| | **T** | **C** | | **T** | **C** | |
| | | | | | | |
| | [absent] | [absent] | | **22** | [absent] | |
| | **25** | 25 | | **27** | 25 | |
| | **30** | 30 | | **32** | 30 | |
| | | | | | | |
| Ave. | **27.5** | **27.5** | | **27** | **27.5** | |
| | | | | | | |
| Difference | | 0 | | Difference | -0.5 | |
| | | | | | | |

# When Might Attrition NOT be a Problem?

A. When the attrition rates are similar in both treatment and control groups

B. When the estimated treatment effect is zero (among those who remain in the study)

C. When the true treatment effect is zero

D. None of the above

# Lecture Overview

- Attrition

- Unexpected Spillovers

- Partial Compliance and Sample Selection Bias

     => Intention to Treat & Local Average Treatment Effect

- Behavioral Responses to Evaluations

- Research Transparency

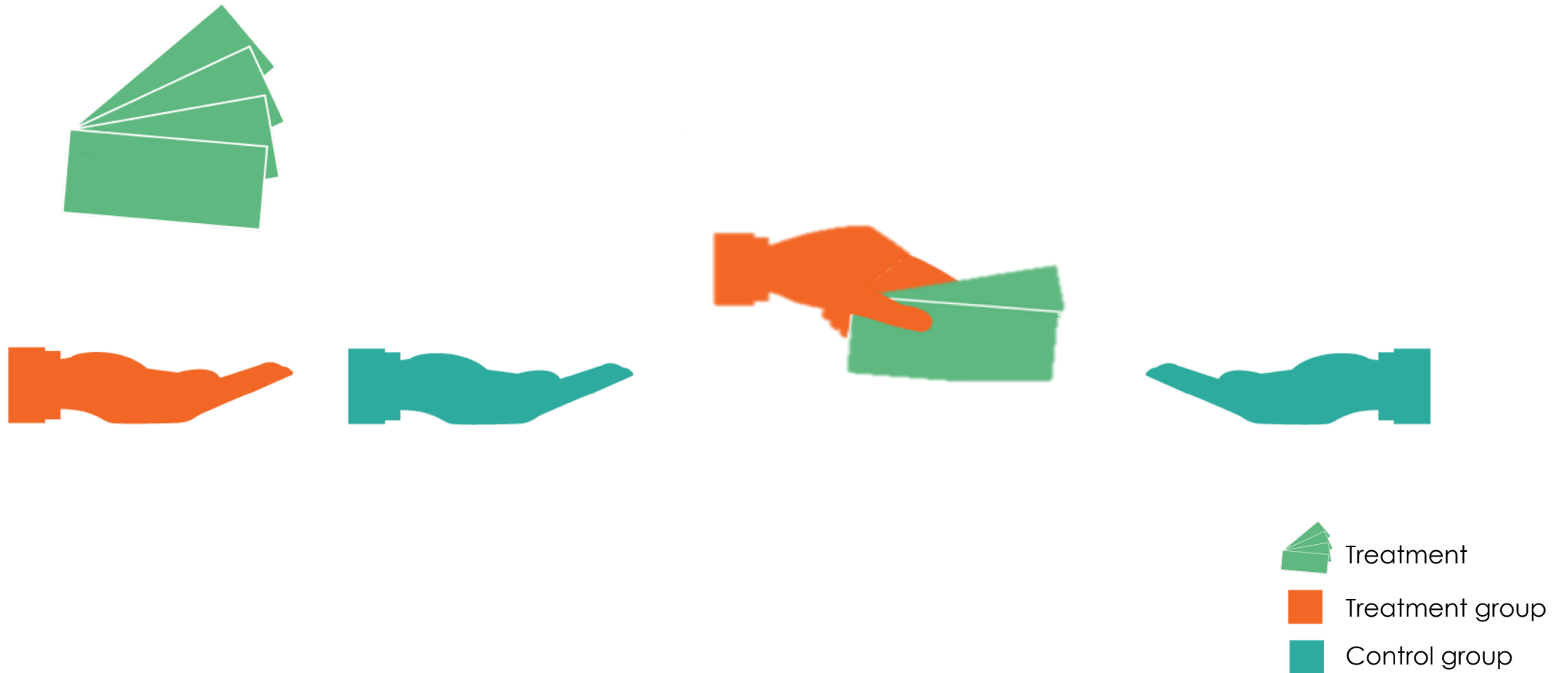# Reminder from Lecture 4: Spillovers
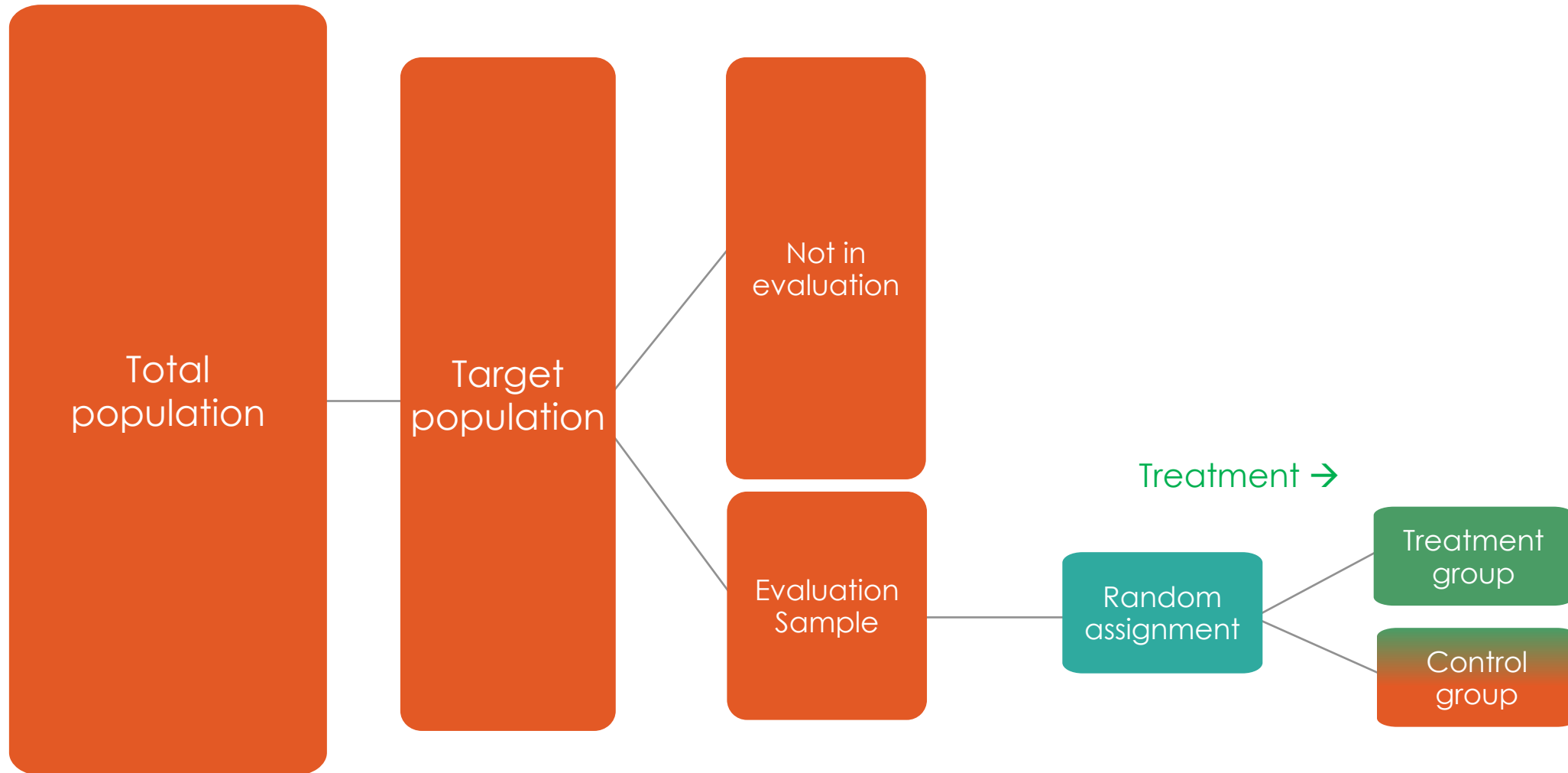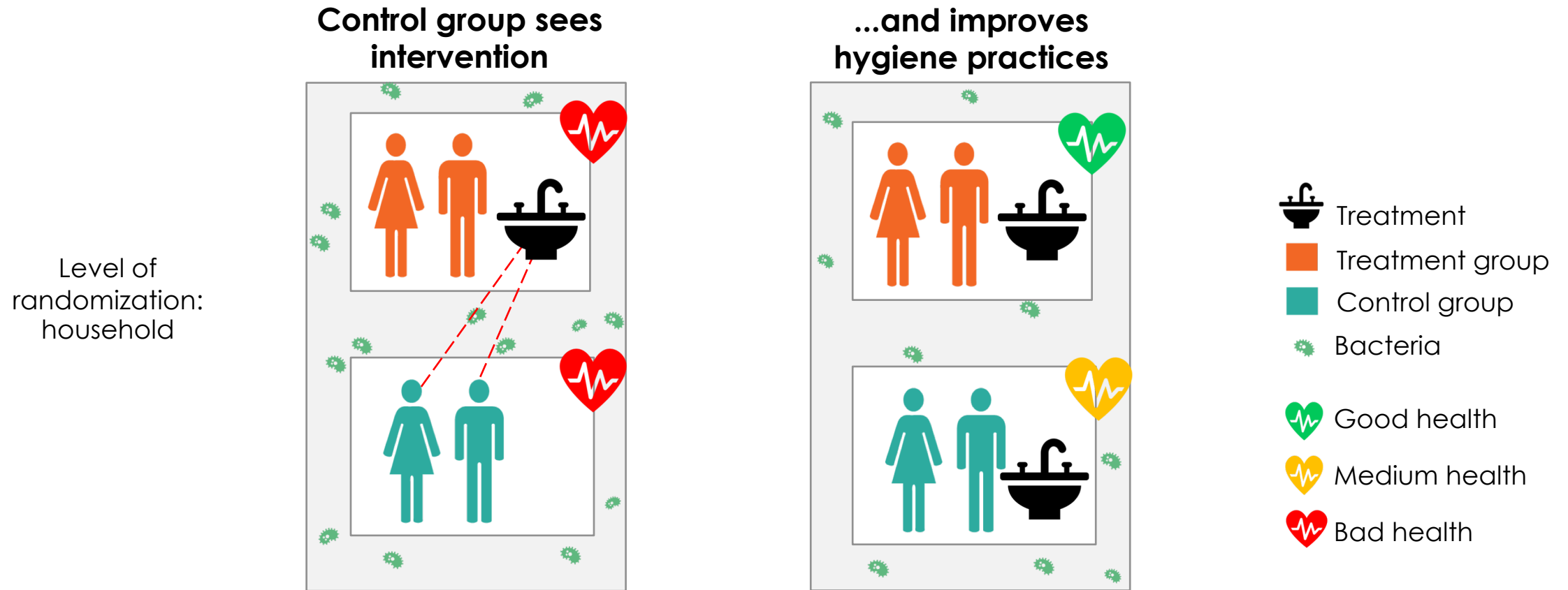
# Reminder: Spillovers

- **Different kinds** of spillovers (physical, informational, behavioral, general equilibrium)

- Can be **positive** or **negative**

- Make it **hard or impossible** to measure the impact of the program

- Two strategies seen during design phase: **avoid** them or **measure** them
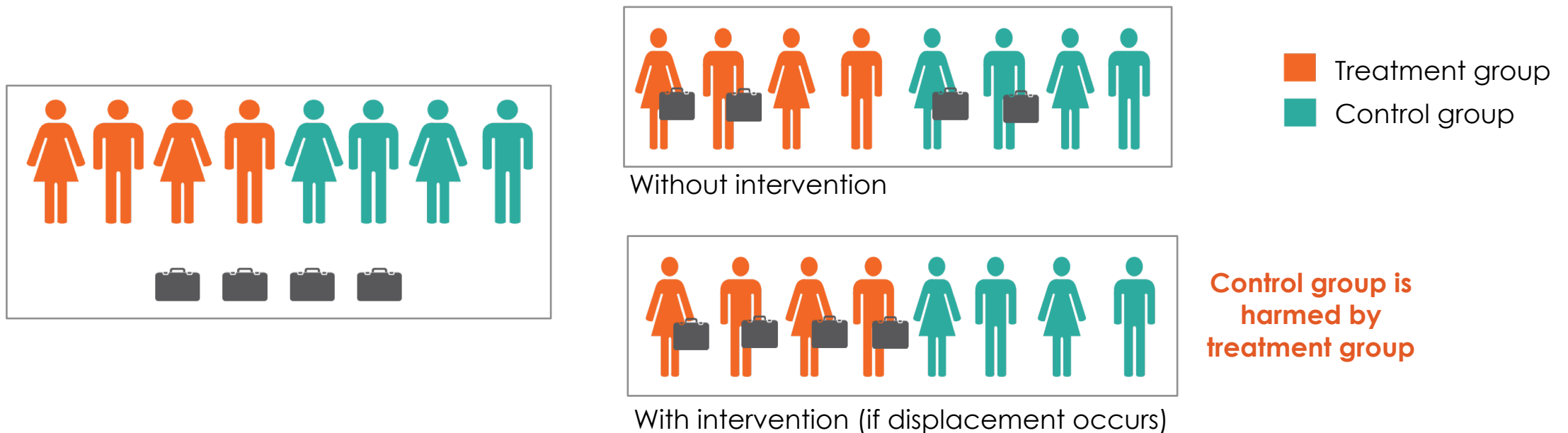
# Physical Spillover



Treatment

Treatment group

Control group

# Behavioral/Informational Spillover

Control group imitates neighbors' hygiene practices or learns about the health benefits



**Control group sees intervention**

**...and improves hygiene practices**

Level of randomization: household

Treatment

Treatment group

Control group

Bacteria

Good health

Medium health

Bad health

# Measuring Marketwide/General Equilibrium Effects
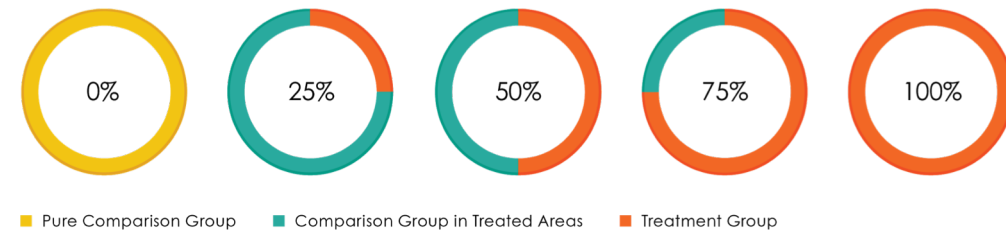
Example: displacement effects from job training programs

- Evaluations of job training programs traditionally compare employment outcomes between those who were trained (treatment) and those in the same area/population who were eligible but not trained (control)

- This does not take into account the possibility that the control group could be harmed if jobs are limited and treatment/control are in competition



Without intervention

With intervention (if displacement occurs)

Treatment group
Control group

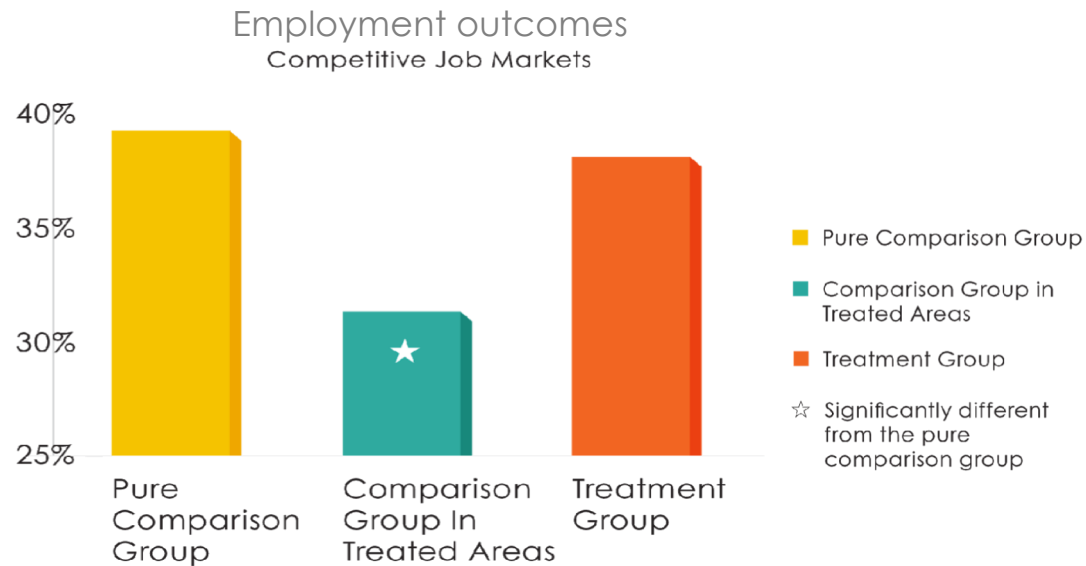**Control group is harmed by treatment group**

# Measuring Marketwide/General Equilibrium Effects

Example: displacement effects from job training programs

- Crépon et al. (2012) evaluates the impact of a job placement program on unemployed populations across 235 labor markets in France

- Labor markets are randomly assigned to one of the following interventions:
  - None of the unemployed receive job training (pure control group)
  - 25% of unemployed are offered job training
  - 50% of unemployed are offered job training
  - 75% of unemployed are offered job training
  - 100% of unemployed are offered job training



0%  25%  50%  75%  100%

■ Pure Comparison Group   ■ Comparison Group in Treated Areas   ■ Treatment Group

- Study measures employment outcomes on treated groups AND control groups in treated areas

# General Equilibrium Effect: Untreated Job Seekers in Program Areas are Harmed by Treatment

## Employment outcomes
### Competitive Job Markets



- Pure Comparison Group
- Comparison Group in Treated Areas
- Treatment Group
- ☆ Significantly different from the pure comparison group

**Misleading comparison**: ■ versus ■

Ignoring the spillover effect, the study would have found that investing 100,000 euros into the job training program causes 9.7 people to find jobs within 8 months.

**Better comparison**: ■ versus ■

Comparing the treatment group to a pure control group provides a better sense of the treatment effect. However, this still fails to account for the spillover.
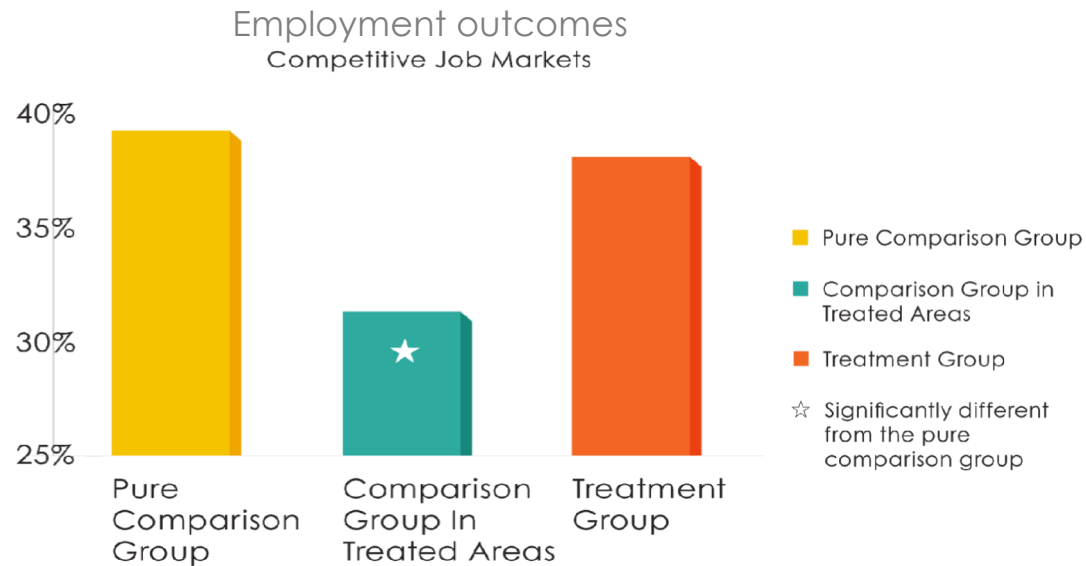
**Measuring the externality**: ■ versus ■

People living in areas with the job program that are not in the program have a harder time finding a job than people outside of those areas.

**Total treatment effect**: ■ versus ■ + ■ versus ■

When considering the spillover, the treatment is found to have no effect.

# General Equilibrium Effect: Untreated Job Seekers in Program Areas are Harmed by Treatment

## Employment outcomes
### Competitive Job Markets



- Pure Comparison Group
- Comparison Group in Treated Areas
- Treatment Group
- ☆ Significantly different from the pure comparison group

**Misleading comparison**: ▢ versus ▢

Ignoring the spillover effect, the study would have found that investing 100,000 euros into the job training program causes 9.7 people to find jobs within 8 months.

**Better comparison**: ▢ versus ▢

Comparing the treatment group to a pure control group provides a better sense of the treatment effect. However, this still fails to account for the spillover.
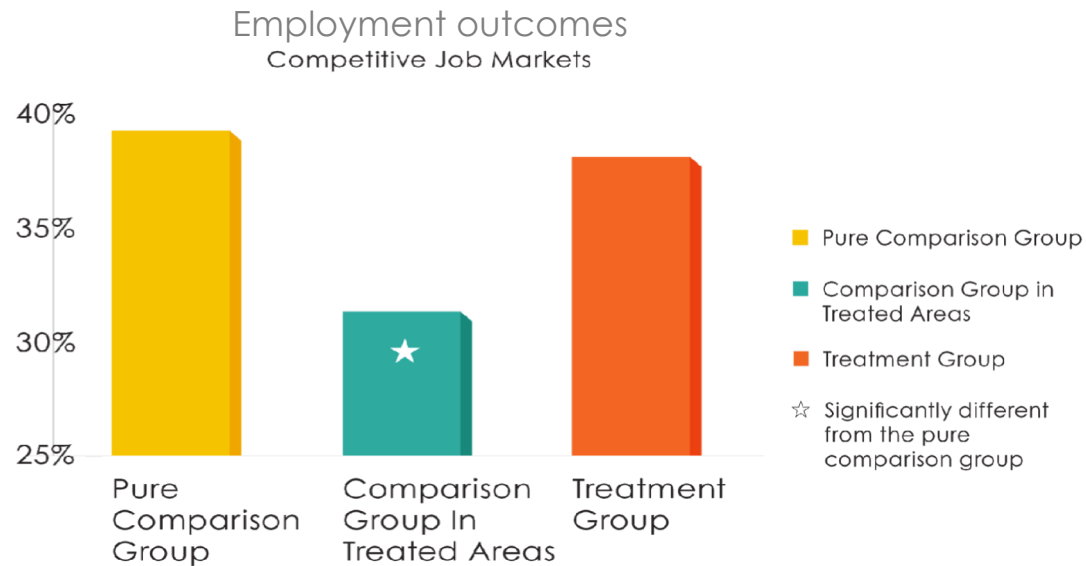
**Measuring the externality**: ▢ versus ▢

People living in areas with the job program that are not in the program have a harder time finding a job than people outside of those areas.

**Total treatment effect**: ▢ versus ▢ + ▢ versus ▢

When considering the spillover, the treatment is found to have no effect.

# General Equilibrium Effect: Untreated Job Seekers in Program Areas are Harmed by Treatment

## Employment outcomes
### Competitive Job Markets



Legend:
- Pure Comparison Group
- Comparison Group in Treated Areas
- Treatment Group
- ☆ Significantly different from the pure comparison group

**Misleading comparison:** ▢ versus ▢

Ignoring the spillover effect, the study would have found that investing 100,000 euros into the job training program causes 9.7 people to find jobs within 8 months.

**Better comparison:** ▢ versus ▢

Comparing the treatment group to a pure control group provides a better sense of the treatment effect. However, this still fails to account for the spillover.

**Measuring the externality:** ▢ versus ▢

People living in areas with the job program that are not in the program have a harder time finding a job than people outside of those areas.

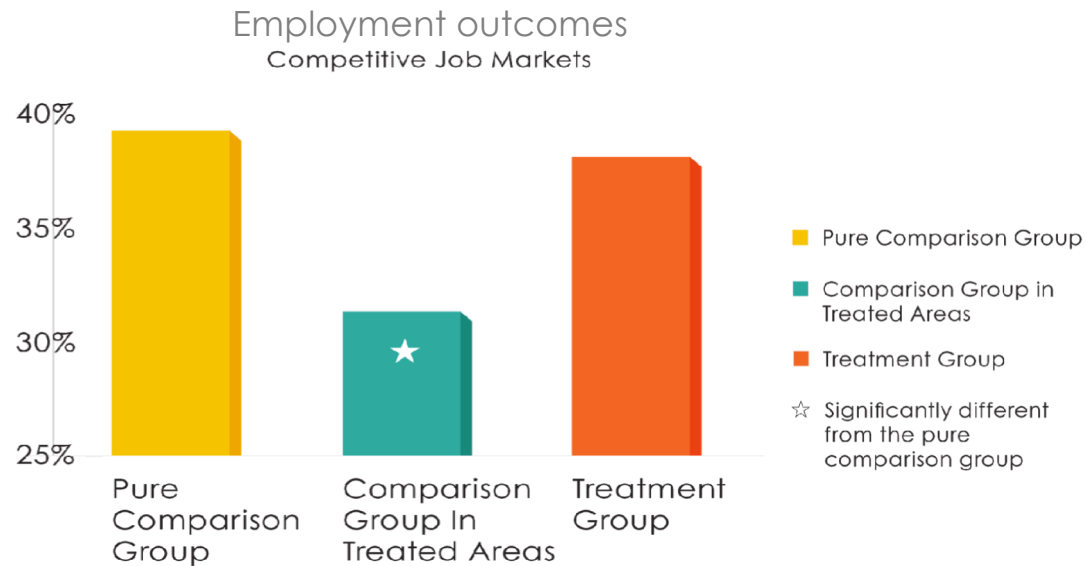**Total treatment effect:** ▢ versus ▢ + ▢ versus ▢

When considering the spillover, the treatment is found to have no effect.

# General Equilibrium Effect: Untreated Job Seekers in Program Areas are Harmed by Treatment

## Employment outcomes
### Competitive Job Markets



- ■ Pure Comparison Group
- ■ Comparison Group in Treated Areas
- ■ Treatment Group
- ☆ Significantly different from the pure comparison group

**Misleading comparison:** ▢ versus ▢

Ignoring the spillover effect, the study would have found that investing 100,000 euros into the job training program causes 9.7 people to find jobs within 8 months.

**Better comparison:** ▢ versus ▢

Comparing the treatment group to a pure control group provides a better sense of the treatment effect. However, this still fails to account for the spillover.

**Measuring the externality:** ▢ versus ▢

People living in areas with the job program that are not in the program have a harder time finding a job than people outside of those areas.

**Total treatment effect:** ▢ versus ▢ + ▢ versus ▢

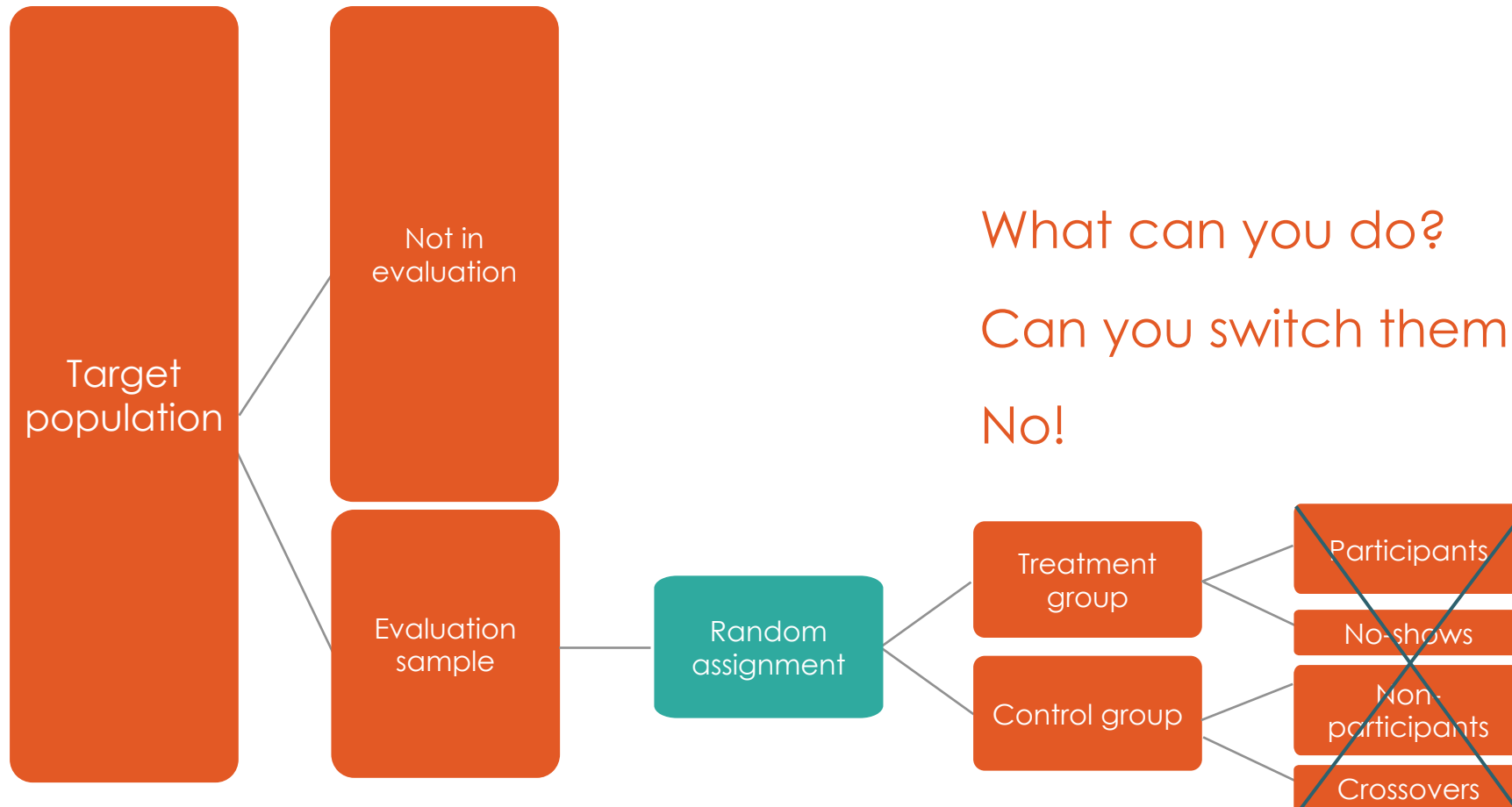When considering the spillover, the treatment is found to have no effect.

# Lecture Overview

- Attrition

- Unexpected Spillovers

- Partial Compliance and Sample Selection Bias

   => Intention to Treat & Local Average Treatment Effect

- Behavioral Responses to Evaluations

- Research Transparency

# Sample Selection Bias

- Sample selection bias could arise if factors **other than random assignment** influence program allocation

- Individuals assigned to comparison group could move into treatment group

- Alternatively, individuals allocated to treatment group may not receive treatment

$\Rightarrow$ Can be due to project **implementers** or to **participants** themselves
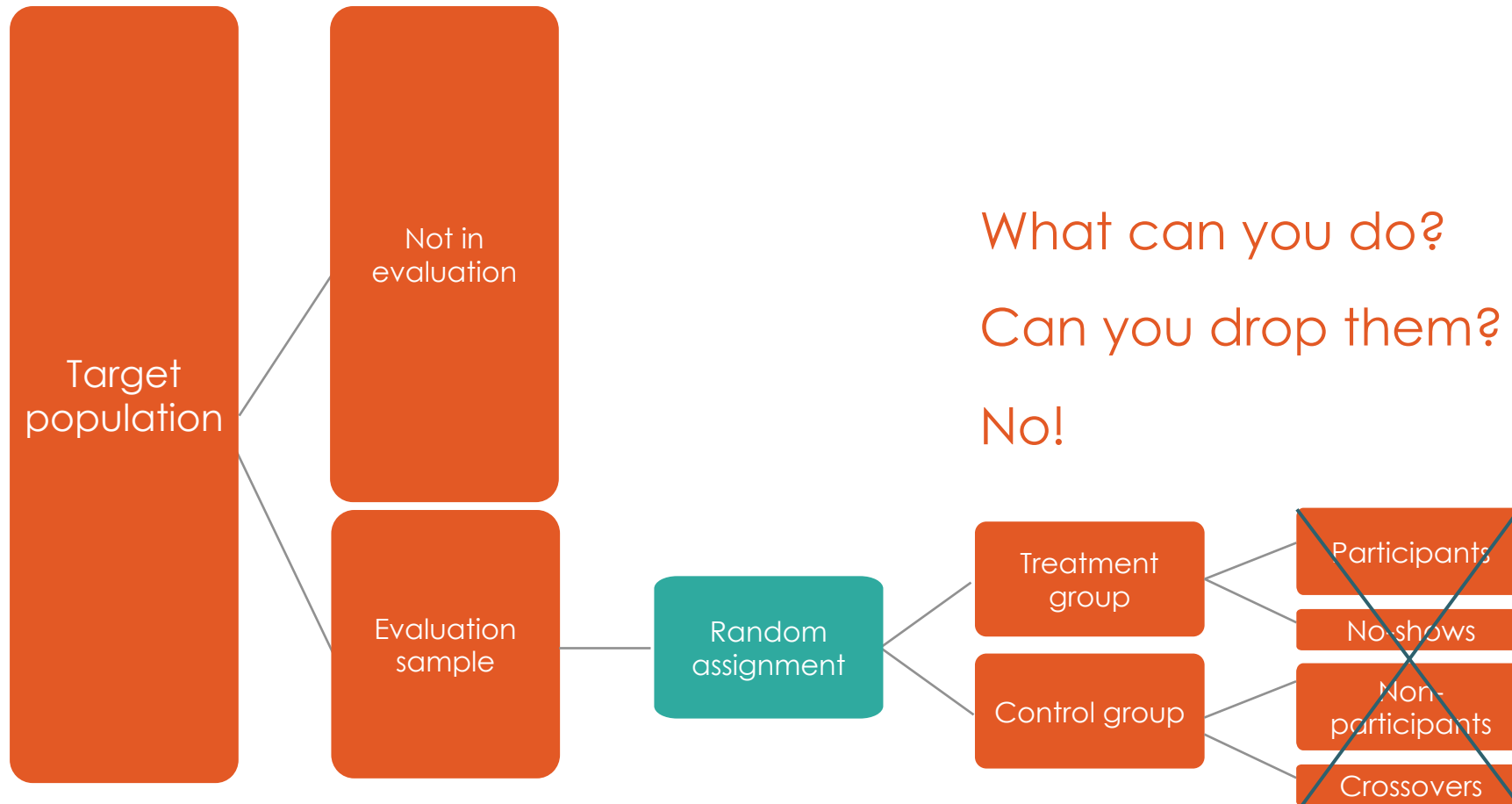
# Noncompliers

Target population

Not in evaluation

Evaluation sample

Random assignment

Treatment group

Control group

Participants

No-shows

Non-participants

Crossovers

What can you do?

Can you switch them?

No!

# Noncompliers

Target population

Not in evaluation

Evaluation sample

Random assignment

Treatment group

Control group

Participants

No-shows

Non-participants

Crossovers

What can you do?

Can you drop them?

No!

# Noncompliers



Target population

Not in evaluation

Evaluation sample

Random assignment

Treatment group

Control group

Participants

No-shows

Non-participants

Crossovers

You can compare the original groups
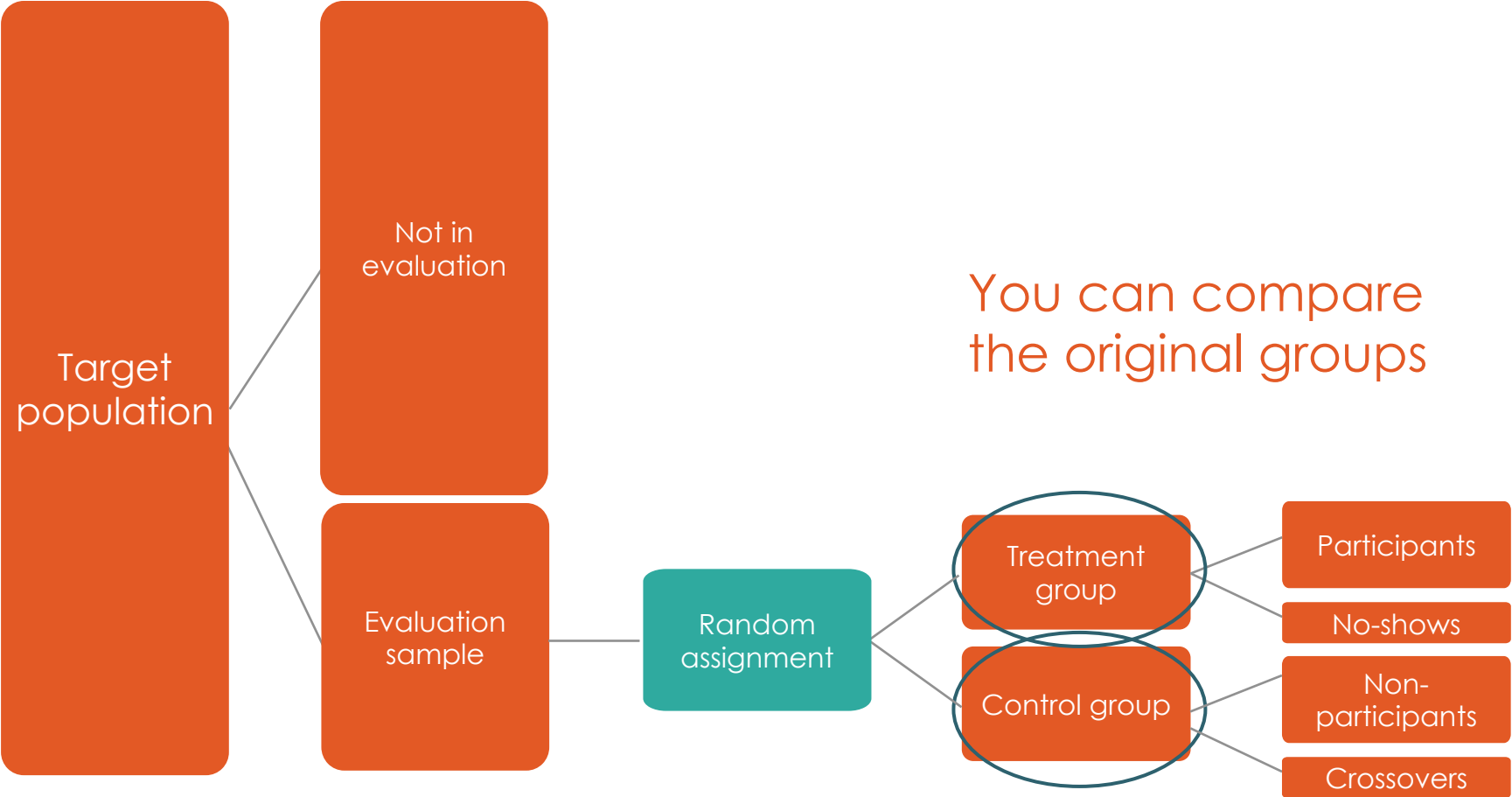
# Your Treatment Group for Analysis is…

A. Individuals assigned to treatment who were *actually* treated

B. All individuals who were *actually* treated

C. Individuals assigned to treatment, regardless of whether or not they were treated

D. Don't know

# What Can be Done?

- Ideally: **prevent** it during design or implementation phase

  => cannot always be done

- **Monitor** it during implementation phase

  => important to be aware that it happens

- **Interpret** it during analysis phase

  => see next section

# Lecture Overview

- Attrition

- Unexpected Spillovers

- Partial Compliance and Sample Selection Bias

  => Intention to Treat & Local Average Treatment Effect

- Behavioral Responses to Evaluations

- Research Transparency

# A School Feeding Program



- Let's take the example of a school feeding program

- Some schools receive the program, some don't (random allocation)

- But allocation is imperfectly respected

# Intention to Treat (ITT)

- Easiest way to deal with partial compliance: Calculate the Intent to Treat (ITT):

  – The difference in between the average outcome of the group that was randomly assigned to treatment and the group that was randomly assigned to control, *regardless of whether they actually received the treatment.*

- What does "intention to treat" measure?

  *"What happened to the average child who is in a treated school in this population?"*

- Is this difference the causal effect of the intervention?

## School 1: Treatment

| | Intention to treat? | Treated | Change in weight |
|---|---|---|---|
| Pupil 1 | Yes | Yes | 4 |
| Pupil 2 | Yes | Yes | 4 |
| Pupil 3 | Yes | Yes | 4 |
| Pupil 4 | Yes | No | 0 |
| Pupil 5 | Yes | Yes | 4 |
| Pupil 6 | Yes | No | 2 |
| Pupil 7 | Yes | No | 0 |
| Pupil 8 | Yes | Yes | 6 |
| Pupil 9 | Yes | Yes | 6 |
| Pupil 10 | Yes | No | 0 |

## School 2: Control

| | Intention to treat? | Treated | Change in weight |
|---|---|---|---|
| Pupil 1 | No | No | 2 |
| Pupil 2 | No | No | 1 |
| Pupil 3 | No | Yes | 3 |
| Pupil 4 | No | No | 0 |
| Pupil 5 | No | No | 0 |
| Pupil 6 | No | Yes | 3 |
| Pupil 7 | No | No | 0 |
| Pupil 8 | No | No | 0 |
| Pupil 9 | No | No | 1 |
| Pupil 10 | No | No | 0 |

**School 1: Treatment**

| | Intention to treat? | Treated | Change in weight |
|---|---|---|---|
| Pupil 1 | Yes | Yes | 4 |
| Pupil 2 | Yes | Yes | 4 |
| Pupil 3 | Yes | Yes | 4 |
| Pupil 4 | Yes | No | 0 |
| Pupil 5 | Yes | Yes | 4 |
| Pupil 6 | Yes | No | 2 |
| Pupil 7 | Yes | No | 0 |
| Pupil 8 | Yes | Yes | 6 |
| Pupil 9 | Yes | Yes | 6 |
| Pupil 10 | Yes | No | 0 |

**School 2: Control**

| | Intention to treat? | Treated | Change in weight |
|---|---|---|---|
| Pupil 1 | No | No | 2 |
| Pupil 2 | No | No | 1 |
| Pupil 3 | No | Yes | 3 |
| Pupil 4 | No | No | 0 |
| Pupil 5 | No | No | 0 |
| Pupil 6 | No | Yes | 3 |
| Pupil 7 | No | No | 0 |
| Pupil 8 | No | No | 0 |
| Pupil 9 | No | No | 1 |
| Pupil 10 | No | No | 0 |

NOT correct

| Mean treated in school 1 | 4.6 |
| Mean not treated in school 2 | 0.5 |
| Difference... | |

**School 1: Treatment**

| | Intention to treat? | Treated | Change in weight |
|---|---|---|---|
| Pupil 1 | Yes | Yes | 4 |
| Pupil 2 | Yes | Yes | 4 |
| Pupil 3 | Yes | Yes | 4 |
| Pupil 4 | Yes | No | 0 |
| Pupil 5 | Yes | Yes | 4 |
| Pupil 6 | Yes | No | 2 |
| Pupil 7 | Yes | No | 0 |
| Pupil 8 | Yes | Yes | 6 |
| Pupil 9 | Yes | Yes | 6 |
| Pupil 10 | Yes | No | 0 |

**School 2: Control**

| | Intention to treat? | Treated | Change in weight |
|---|---|---|---|
| Pupil 1 | No | No | 2 |
| Pupil 2 | No | No | 1 |
| Pupil 3 | No | Yes | 3 |
| Pupil 4 | No | No | 0 |
| Pupil 5 | No | No | 0 |
| Pupil 6 | No | Yes | 3 |
| Pupil 7 | No | No | 0 |
| Pupil 8 | No | No | 0 |
| Pupil 9 | No | No | 1 |
| Pupil 10 | No | No | 0 |

The *Intent to Treat:*

| | |
|---|---|
| Mean in school 1 : | 3.0 |
| Mean in school 2 : | 1.0 |
| **Difference:** | **2.0** |

**School 1: Treatment**

| | Intention to treat? | Treated | Change in weight |
|---|---|---|---|
| Pupil 1 | Yes | Yes | 4 |
| Pupil 2 | Yes | Yes | 4 |
| Pupil 3 | Yes | Yes | 4 |
| Pupil 4 | Yes | No | 0 |
| Pupil 5 | Yes | Yes | 4 |
| Pupil 6 | Yes | No | 2 |
| Pupil 7 | Yes | No | 0 |
| Pupil 8 | Yes | Yes | 6 |
| Pupil 9 | Yes | Yes | 6 |
| Pupil 10 | Yes | No | 0 |

**School 2: Control**

| | Intention to treat? | Treated | Change in weight |
|---|---|---|---|
| Pupil 1 | No | No | 2 |
| Pupil 2 | No | No | 1 |
| Pupil 3 | No | Yes | 3 |
| Pupil 4 | No | No | 0 |
| Pupil 5 | No | No | 0 |
| Pupil 6 | No | Yes | 3 |
| Pupil 7 | No | No | 0 |
| Pupil 8 | No | No | 0 |
| Pupil 9 | No | No | 1 |
| Pupil 10 | No | No | 0 |

The *Intent to Treat:*

| Mean in school 1 : | 3.0 |
|---|---|
| Mean in school 2 : | 1.0 |
| Difference: | 2.0 |

Treatment Probability:

| Fraction treated in school 1: | 0.6 |
|---|---|
| Fraction treated in school 2: | 0.2 |
| Difference: | 0.4 |

# Local Average Treatment Effect (LATE)

- Sometimes we want to know the actual causal effect of the program, not just the ITT:

  – This is called the estimate of the "Local Average Treatment Effect": LATE

- The intuitive idea:

  – Let's say the ITT effect of afterschool classes is a 3 point test score difference between treatment and control schools.

  – But only 50% of the children in the treatment schools actually went to the classes (for simplicity let's assume no children in control schools got the classes).

- If the effect of 50% take-up is to increase scores by 3 points, then we can say that if everyone were to take the classes, the effect would be

$$3 \times \frac{1}{.5} = 3 \times 2 = 6 \text{ points}$$

# Local Average Treatment Effect (LATE)

- In general, the Local Average Treatment Effect (LATE) is:

$$LATE = \frac{ITT}{\text{(fraction of take-up in treatment)-(fraction take-up in control)}}$$

- What does the LATE measure?

    *The effect of the program on those people who choose to*

    *take it up due to the intervention.*

- Note: Effects on those people who didn't take it up might have been quite different.

- Very similar: "Treatment on the Treated" (TOT)

## School 1: Treatment

| | Intention to treat? | Treated | Change in weight |
|---|---|---|---|
| Pupil 1 | Yes | Yes | 4 |
| Pupil 2 | Yes | Yes | 4 |
| Pupil 3 | Yes | Yes | 4 |
| Pupil 4 | Yes | No | 0 |
| Pupil 5 | Yes | Yes | 4 |
| Pupil 6 | Yes | No | 2 |
| Pupil 7 | Yes | No | 0 |
| Pupil 8 | Yes | Yes | 6 |
| Pupil 9 | Yes | Yes | 6 |
| Pupil 10 | Yes | No | 0 |

## School 2: Control

| | Intention to treat? | Treated | Change in weight |
|---|---|---|---|
| Pupil 1 | No | No | 2 |
| Pupil 2 | No | No | 1 |
| Pupil 3 | No | Yes | 3 |
| Pupil 4 | No | No | 0 |
| Pupil 5 | No | No | 0 |
| Pupil 6 | No | Yes | 3 |
| Pupil 7 | No | No | 0 |
| Pupil 8 | No | No | 0 |
| Pupil 9 | No | No | 1 |
| Pupil 10 | No | No | 0 |

### The *Intent to Treat:*

| Mean in school 1 : | 3.0 |
|---|---|
| Mean in school 2 : | 1.0 |
| Difference: | 2.0 |

### Treatment Probability:

| Fraction treated in school 1: | 0.6 |
|---|---|
| Fraction treated in school 2: | 0.2 |
| Difference: | 0.4 |

### *Local Average Treatment Effect:*

$$2/.4 = 5$$

# ITT vs LATE

- If obtaining estimate is easy, why not always use LATE?

- ITT may be policy relevant parameter of interest

  - For example, we may not be interested in the medical effect of deworming treatment, but what would happen under an actual deworming program.

  - If students often miss school and therefore don't get the deworming medicine, the intention to treat estimate may actually be most relevant.

# ITT / LATE: Conclusions

- Both ITT and LATE can provide valuable information to decision-makers

- LATE gives the effect of the intervention on the ones that comply with the programme

- ITT gives the overall effect of the intervention, admitting that partial compliance can happen (which is inherent to any policy)

# Lecture Overview

- Attrition

- Unexpected Spillovers

- Partial Compliance and Sample Selection Bias

    => Intention to Treat & Local Average Treatment Effect

- Behavioral Responses to Evaluations

- Research Transparency

# Behavioral Responses to Evaluations

One limitation of evaluations is that they may cause changes in behavior.  How?

- **Treatment group** changes its behavior:
  - Hawthorne effect
  - Demand effect

- **Comparison group** changes its behavior:
  - John Henry effect
  - Resentment and demoralization effects
  - Anticipation effects

- **Both groups** can be affected: survey effects

# Hawthorne Effect

- Experiments from 1924-32 at Hawthorne Works, a Western Electric Factory

- Different experiments to increase workers productivity, including lighting studies

- Productivity gains as a result of the attention paid to workers

- When the experiment stops, gains disappear

Productivity increases
Productivity decreases

# John Henry Effect

- A legendary American railway worker in the 1870s

- Heard that his output was compared to the output of a machine

- Worked harder to outperform the machine (and died)

# How to Limit Evaluation-Driven Effects?

- Use a **different level** of randomization


- **Minimize salience** of evaluation as much as possible:

  - Do not announce phase-in
    - Downside is that this can be useful to reduce attrition!
  - Make sure staff is impartial and treats both groups similarly (ex: blind data collection staff to treatment arm)


- Measure the evaluation-driven effects in a **subset** of the sample

# Lecture Overview

- Attrition

- Unexpected Spillovers

- Partial Compliance and Sample Selection Bias

    => Intention to Treat & Local Average Treatment Effect

- Behavioral Responses to Evaluations

- Research Transparency

# Multiple Outcomes

- Can we look at various outcomes?

- The more outcomes you look at, the higher the chance you find at least one significantly affected by the program

    – Pre-specify outcomes of interest

    – Report results on all measured outcomes, even null results

    – Correct statistical tests (Bonferroni)

    – Group outcomes together and form indices

# Covariates

- Why include covariates?
    - May explain variation, improve statistical power

- Why not include covariates?
    - Appearances of "specification searching"

- What to control for?
    - If stratified randomization: add strata fixed effects
    - Other covariates

General Guideline: Report both "raw" differences and regression-adjusted results

# The AEA RCT Registry

# To Do or Not to Do a Pre-Analysis Plan?

- Particularly useful when:

- Many ways to measure the outcome

- Many different subgroups

- But some drawbacks:

- What about unexpected outcomes?

- How to adapt to the main findings?

$\Rightarrow$ We can do conditional PAPs… but costly and time-consuming

$\Rightarrow$ Up to each J-PAL affiliate to do or not to do a PAP

# Conclusions

- Internal validity is the great strength of Randomized Evaluations…

- …so everything undermining it must be carefully considered

- Design phase and power calculation are important…

- …but so is the ability to face challenges during implementation phase

- Distinguish well between attrition, spillovers and partial compliance

- Be aware of experimental effects

# Further Resources

- Using Randomization in Development Economics Research: A Toolkit (Duflo, Glennerster, Kremer)

- Mostly Harmless Econometrics (Angrist and Pischke)

- Identification and Estimation of Local Average Treatment Effects (Imbens and Angrist, Econometrica, 1994).

# References, Reuse, and Citation