J-PAL
ABDUL LATIF JAMEEL POVERTY ACTION LAB

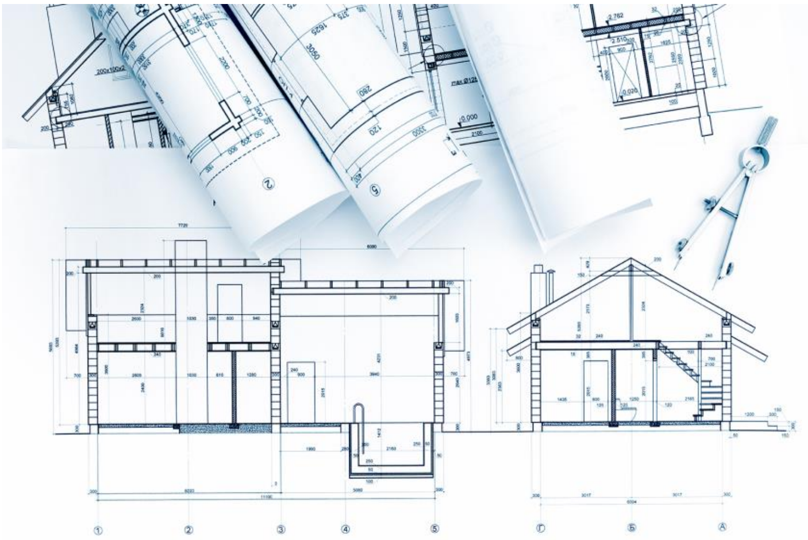
Post-Design Challenges

Professor Supreet Kaur
Department of Economics
UC Berkeley

Course Overview

1. What is Evaluation?
2. Outcomes, Impact, and Indicators
3. Why Randomize?
4. How to Randomize?
5. Sampling and Sample Size
6. Post-Design Challenges
7. From Evidence To Policy
8. Project from Start to Finish

Introduction



Conception phase is important and allows to design an evaluation enabling 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 **prevent** each of these threats
- To know some of the 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 live next to a treatment school
- First impact of your program: increased enrollment.
- In addition, you want to measure the impact on child’s growth
  - Second outcome of interest: Weight of children
- You go to all the schools (treatment and control) and measure 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

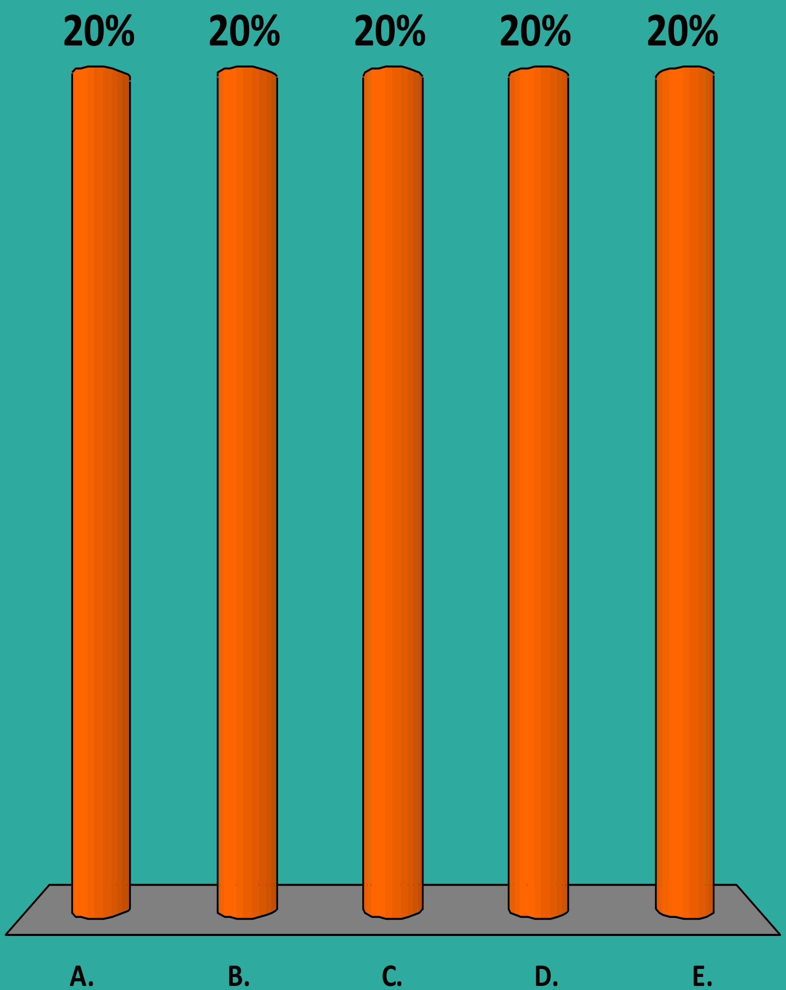
A. Will you underestimate the impact?

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 is attrition not a problem?

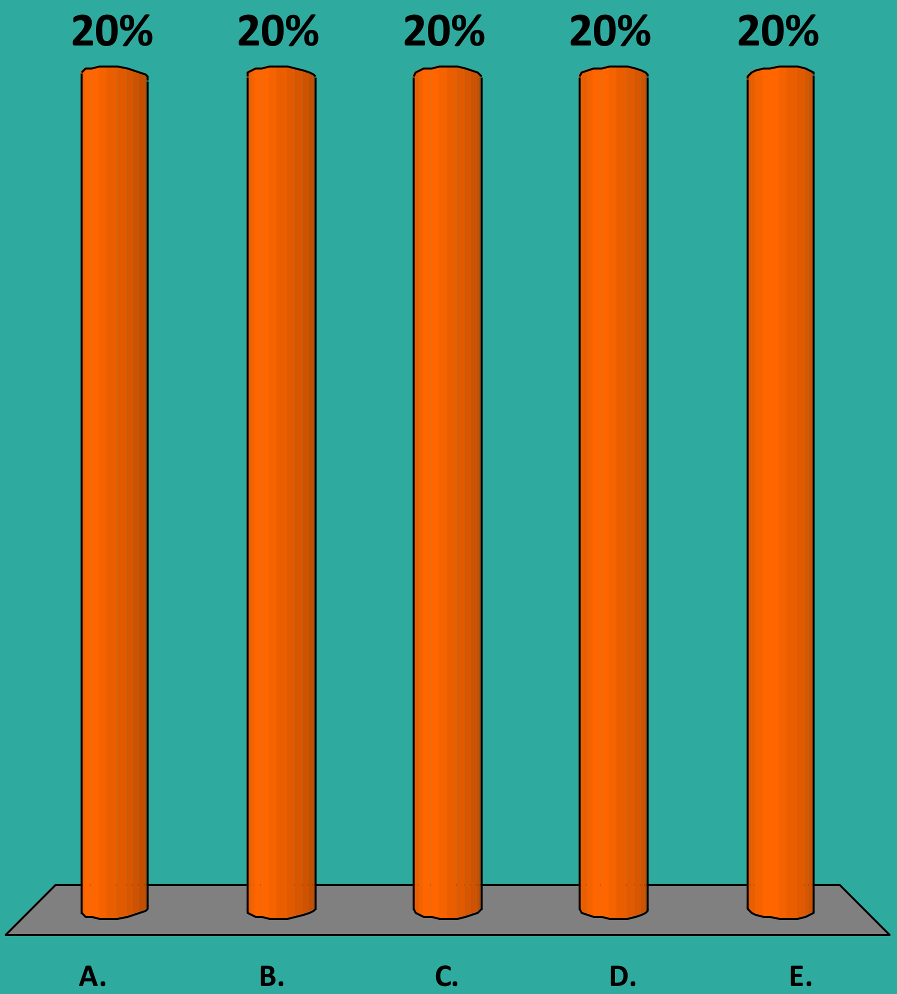
A. When it is less than 25% of the original sample

B. When it happens in the same proportion in both groups

C. When it is correlated with treatment assignment

D. All of the above

E. None of the above

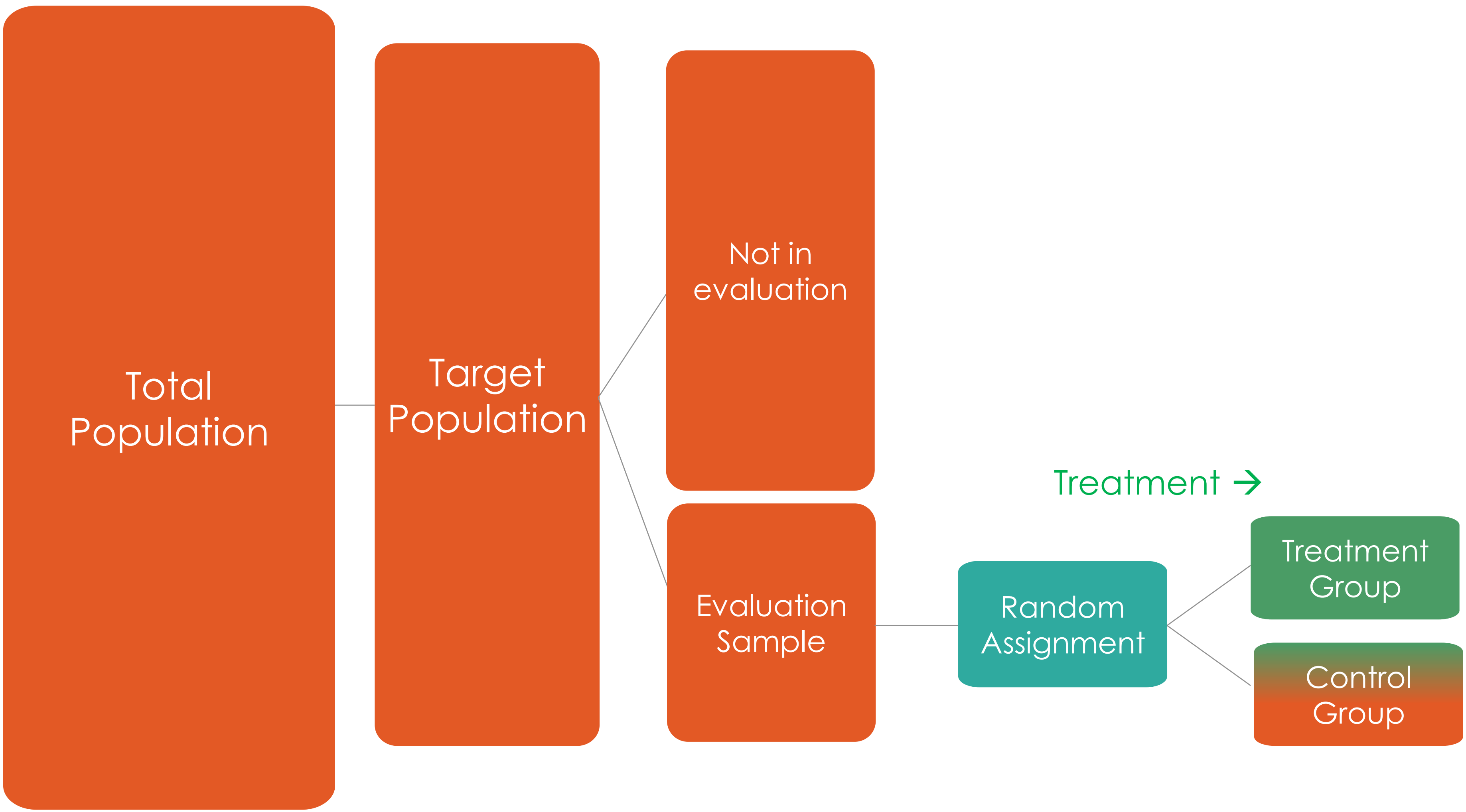


Lecture Overview

- 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

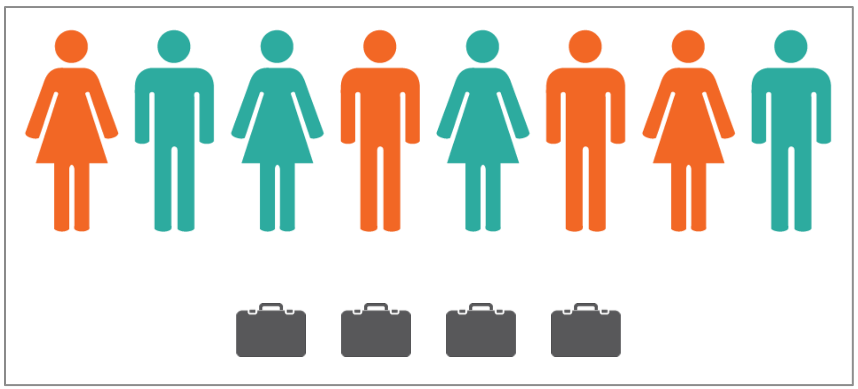


Reminder: Spillovers

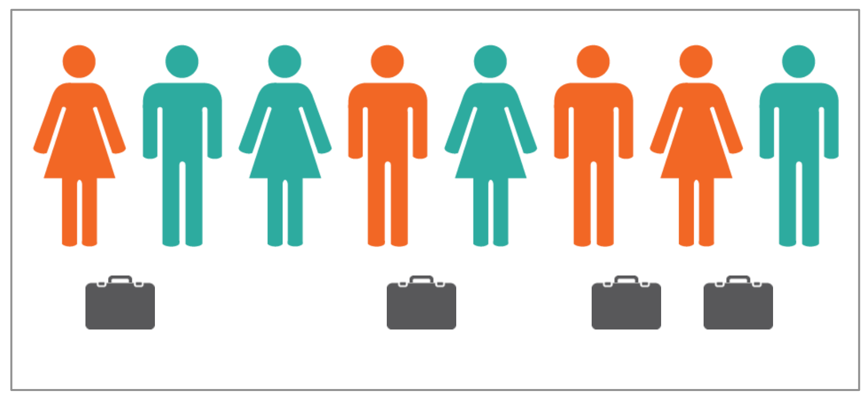
- **Different kinds** of spillovers (physical, informational, behavioral, general equilibrium)
- Can be **positive** or **negative**
- Make **hard or impossible** to measure the impact of the program
- Two strategies seen during design phase: **avoid** them or **measure** them

=> But what can we do if **unexpected spillovers** do happen?

General Equilibrium



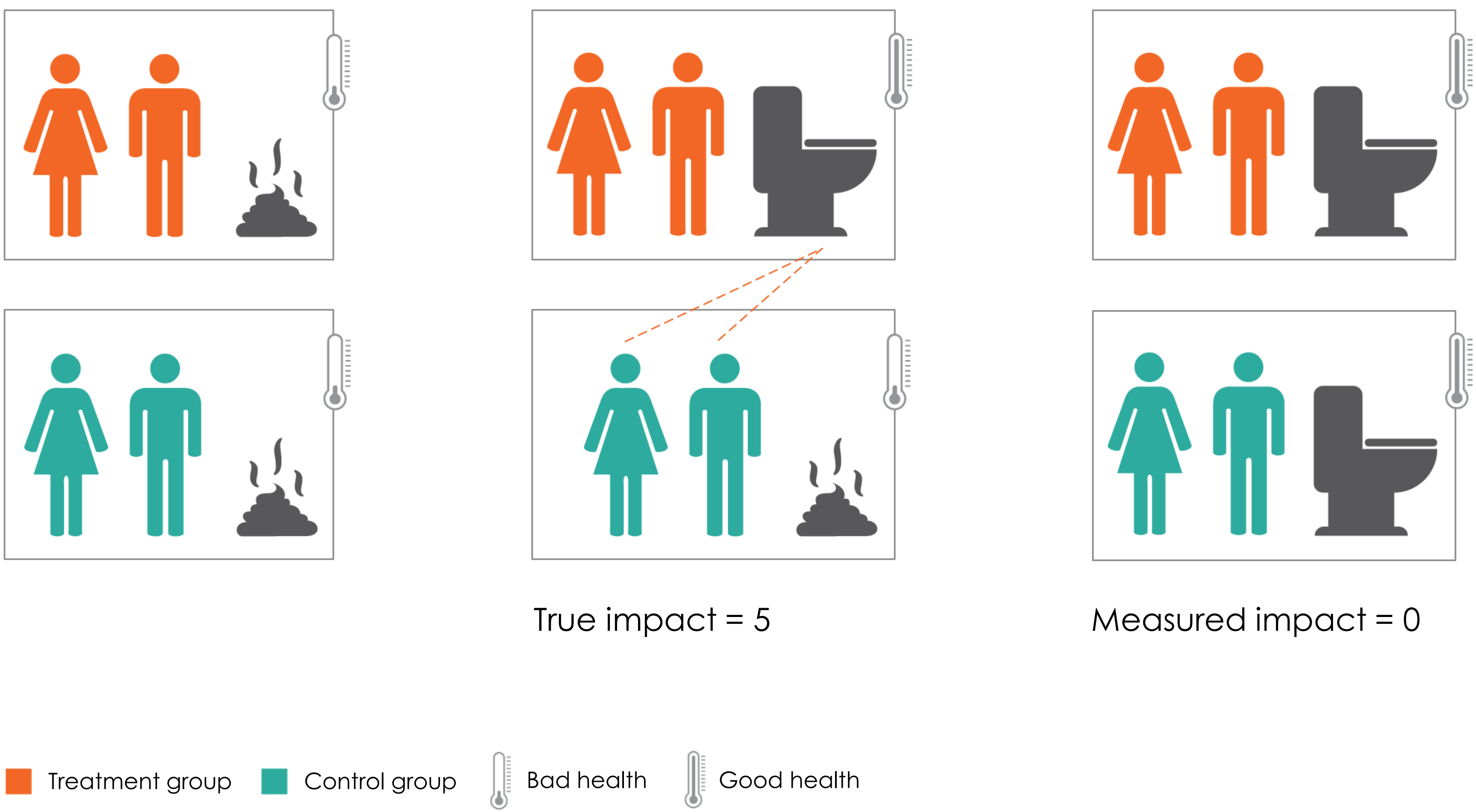
Without experiment



With experiment

Treatment group   Control group

Behavioral/Informational



Community Health

Physical

Treatment group Control group

Lecture Overview

- Attrition
- Unexpected Spillovers
- Partial Compliance and Sample Selection Bias

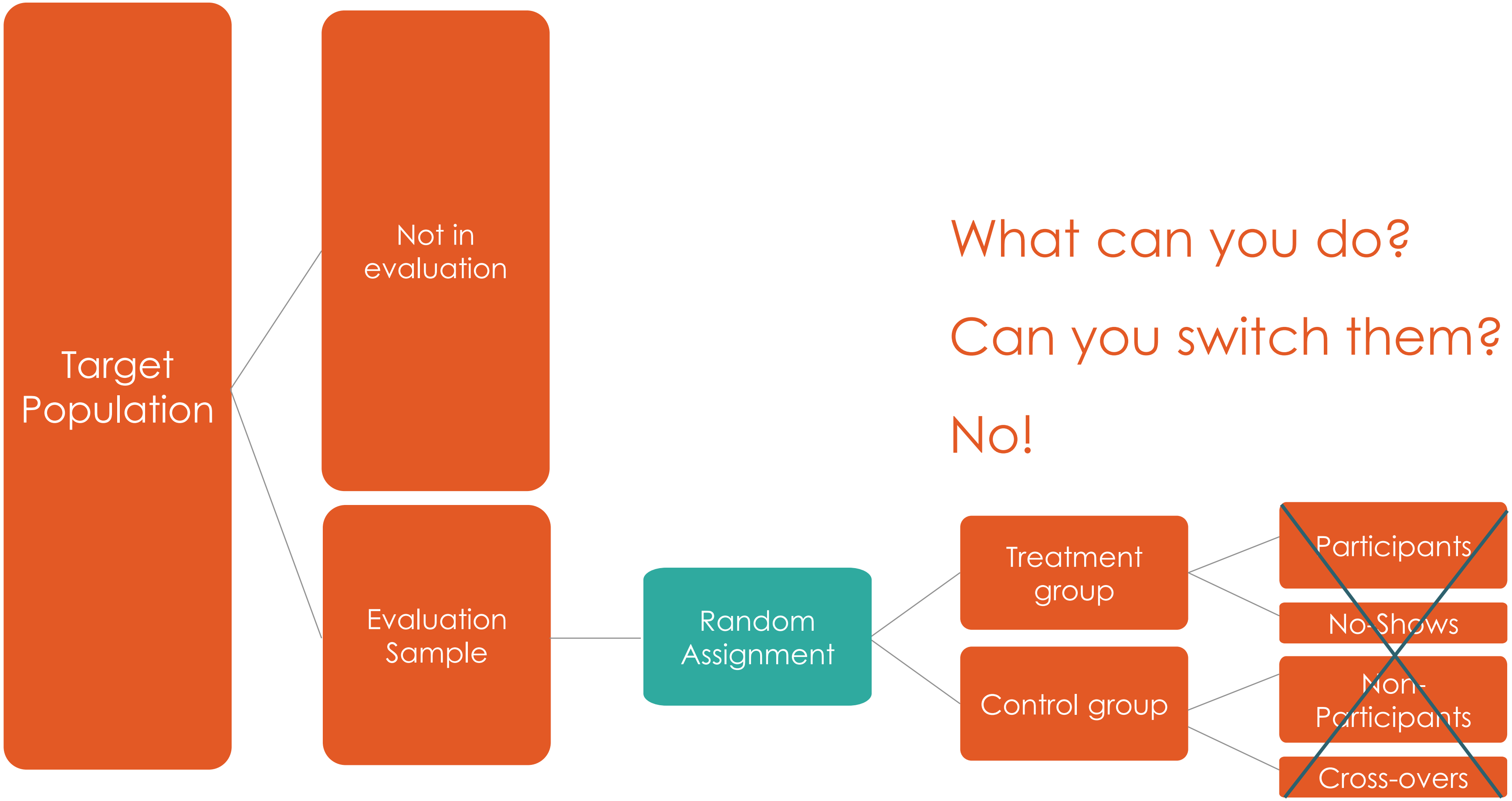
  => Intention to Treat & Local Average Treatment Effect
- Behavioral Responses to Evaluations
- Research Transparency

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

⇒ Can be due to project **implementers** or to **participants** themselves

Non compliers



What can you do?

Can you switch them?

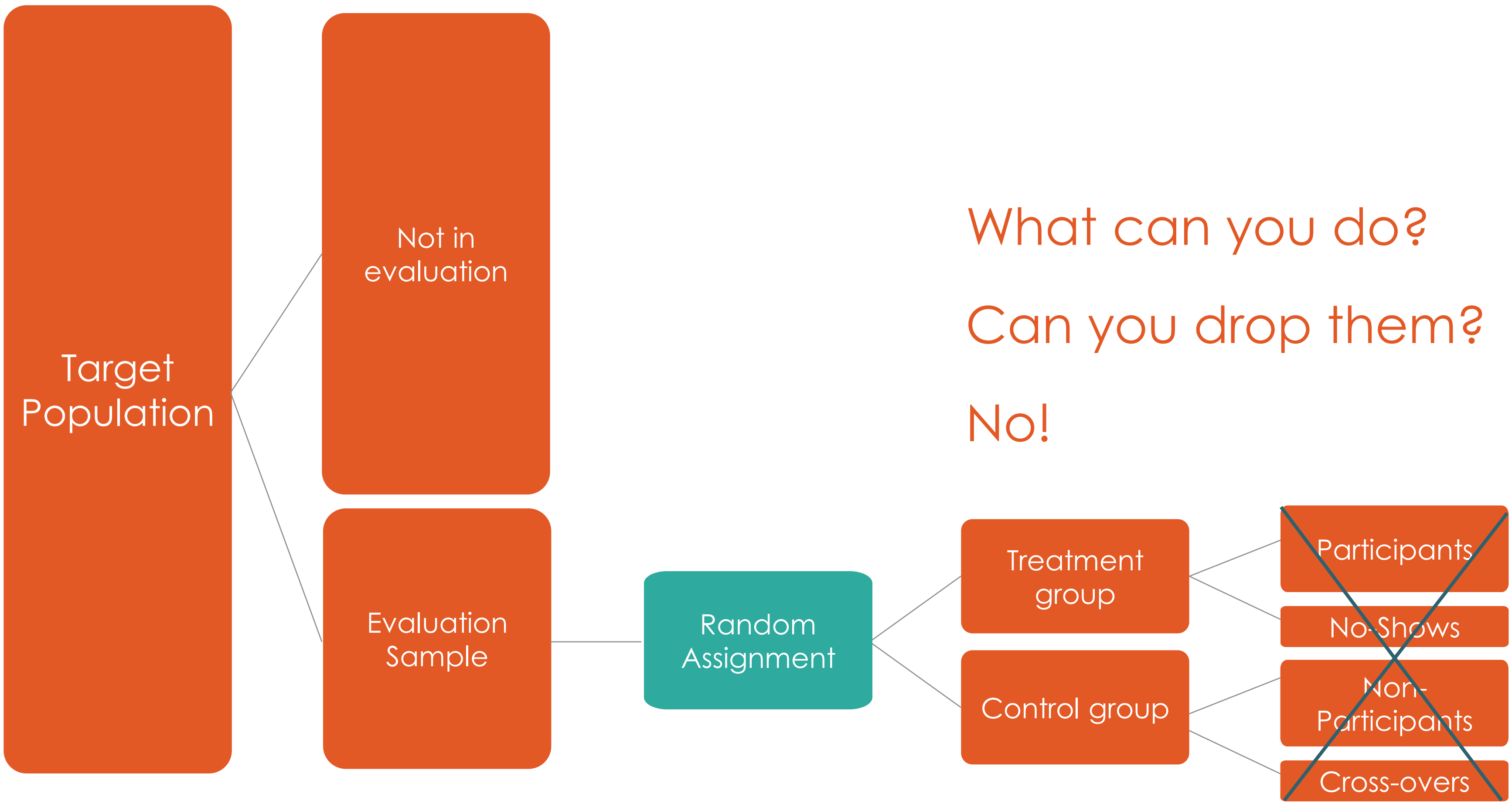
No!

Non compliers

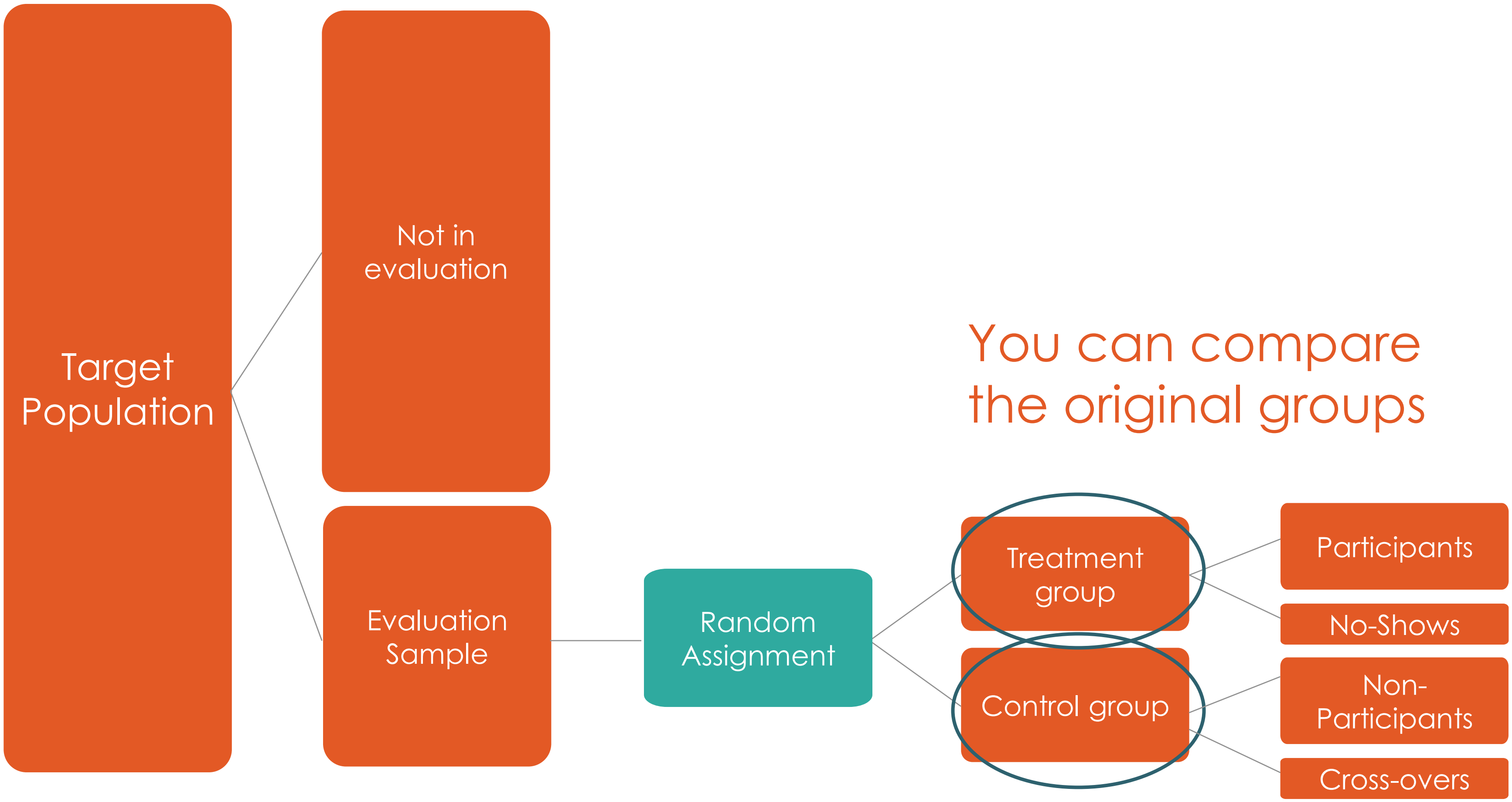
What can you do?

Can you drop them?

No!



Non compliers



You can compare the original groups

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

Compliance is imperfect

School 1	Intention to treat?	Treated?
Pupil 1	Yes	Yes
Pupil 2	Yes	Yes
Pupil 3	Yes	Yes
Pupil 4	Yes	**No**
Pupil 5	Yes	Yes
Pupil 6	Yes	**No**
Pupil 7	Yes	**No**
Pupil 8	Yes	Yes
Pupil 9	Yes	Yes
Pupil 10	Yes	**No**

School 2	Intention to Treat?	Treated?
Pupil 1	No	No
Pupil 2	No	No
Pupil 3	No	Yes
Pupil 4	No	No
Pupil 5	No	No
Pupil 6	No	Yes
Pupil 7	No	No
Pupil 8	No	No
Pupil 9	No	No
Pupil 10	No	No

ITT / LATE

Intention To Treat	Local Average Treatment Effect
What happened to the average child who is in a treated school in this population?	What happened to a child that actually received the treatment?
Measuring the impact of launching the program	Measuring the impact of the program itself

- ITT and LATE are two different ways to analyze the data
- ITT may relate more to actual programs, especially if imperfect compliance is likely to happen

=> Let's now see how we do it

Intention To Treat

School 1	Intention to treat?	Treated?	Observed 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
		Avg. Change among Treated A =	

School 2			
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**	**0**
Pupil 10	No	**No**	**0**
		Avg. Change among Not-Treated B =	

School 1: Avg. Change among Treated	(A)
School 2: Avg. Change among Not-Treated	(B)
A-B	

School 1	Intention to treat?	Treated?	Observed 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
		Avg. Change among Treated A =	**3**

School 2			
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**	**0**
Pupil 10	No	**No**	**0**
		Avg. Change among Not-Treated B =	**0.9**

School 1: Avg. Change among Treated	(A) **3**
School 2: Avg. Change among Not-Treated	(B) **0.9**
A-B	**2.1**

From ITT to LATE

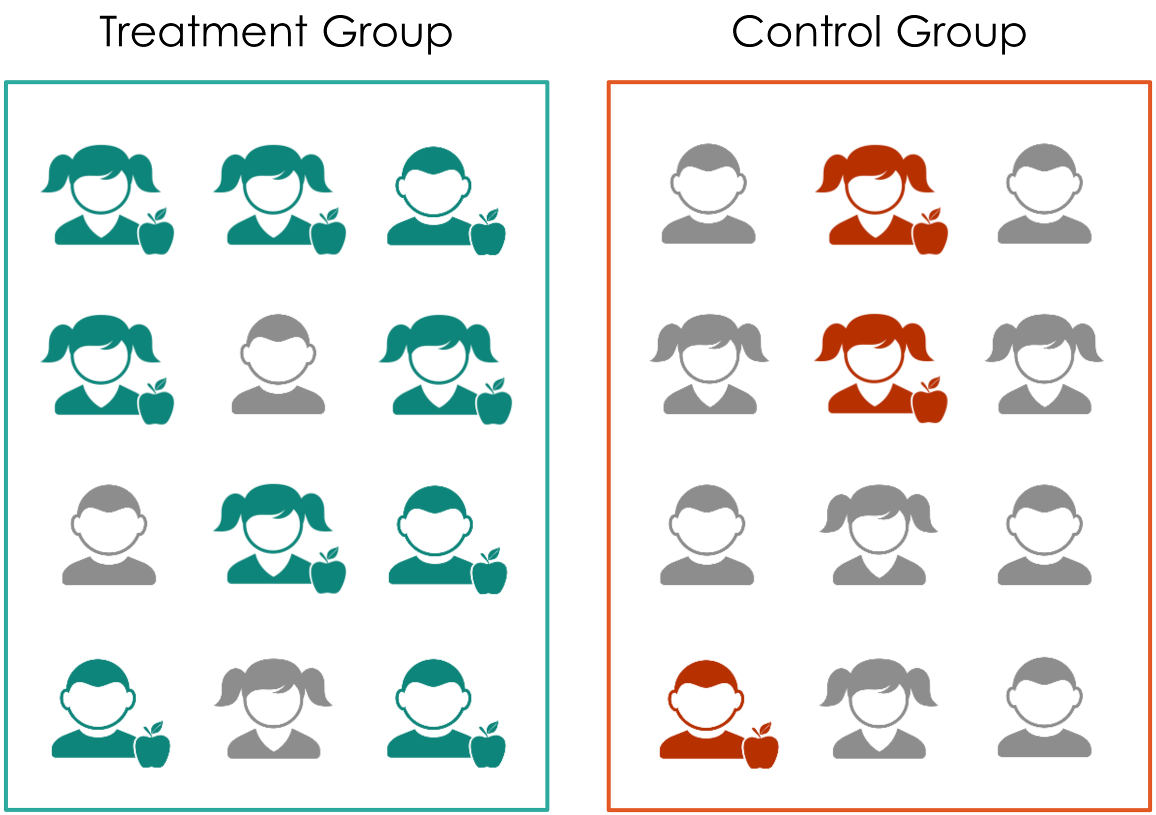
We conceptually divide our treatment and control groups into three categories:

1/ The "**always takers**", who will get the meals no matter if they are in the treatment or the control group

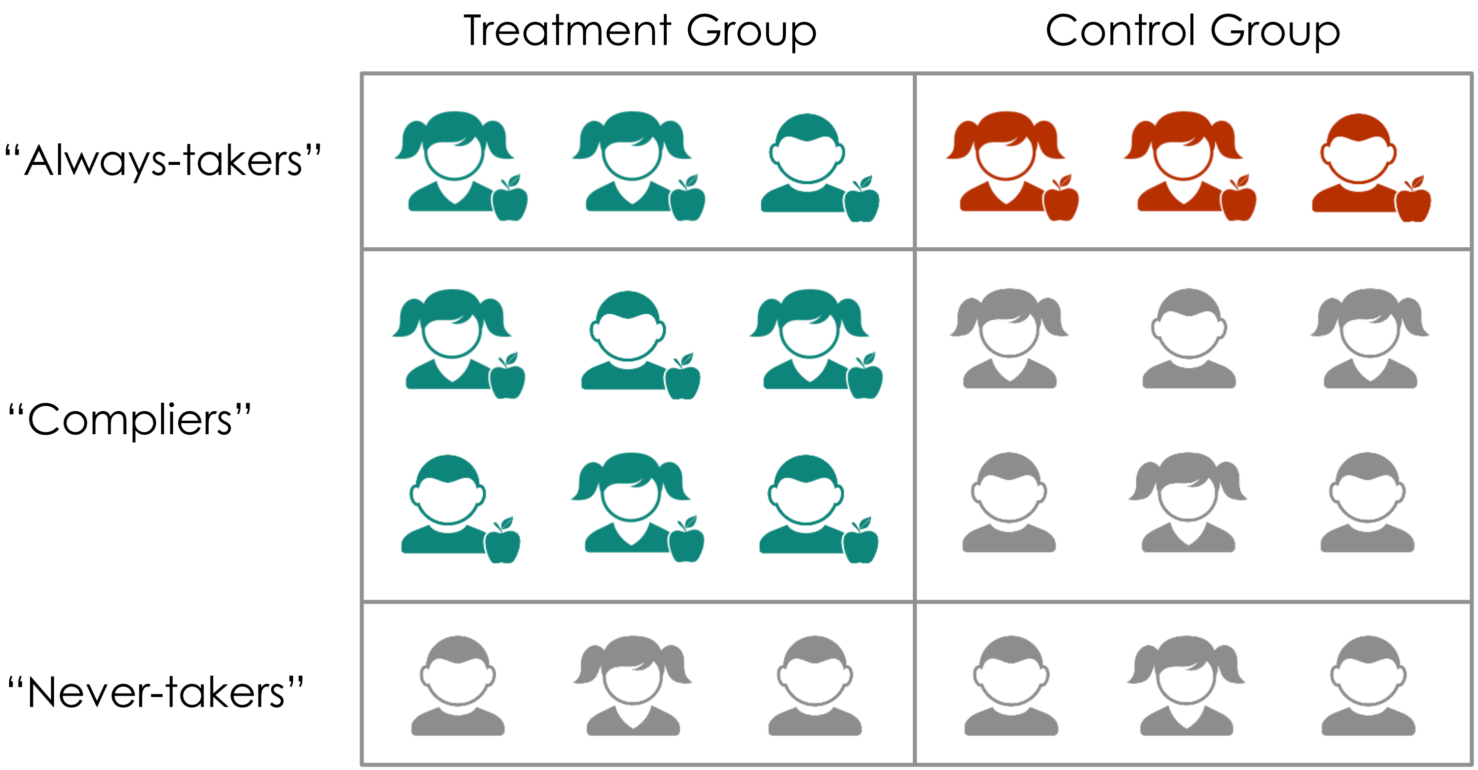
2/ The "**never takers**", who won't get the meals no matter if they are in the treatment or the control group

3/ The "**compliers**", who will behave according to the group they have been assigned to

A situation of imperfect compliance

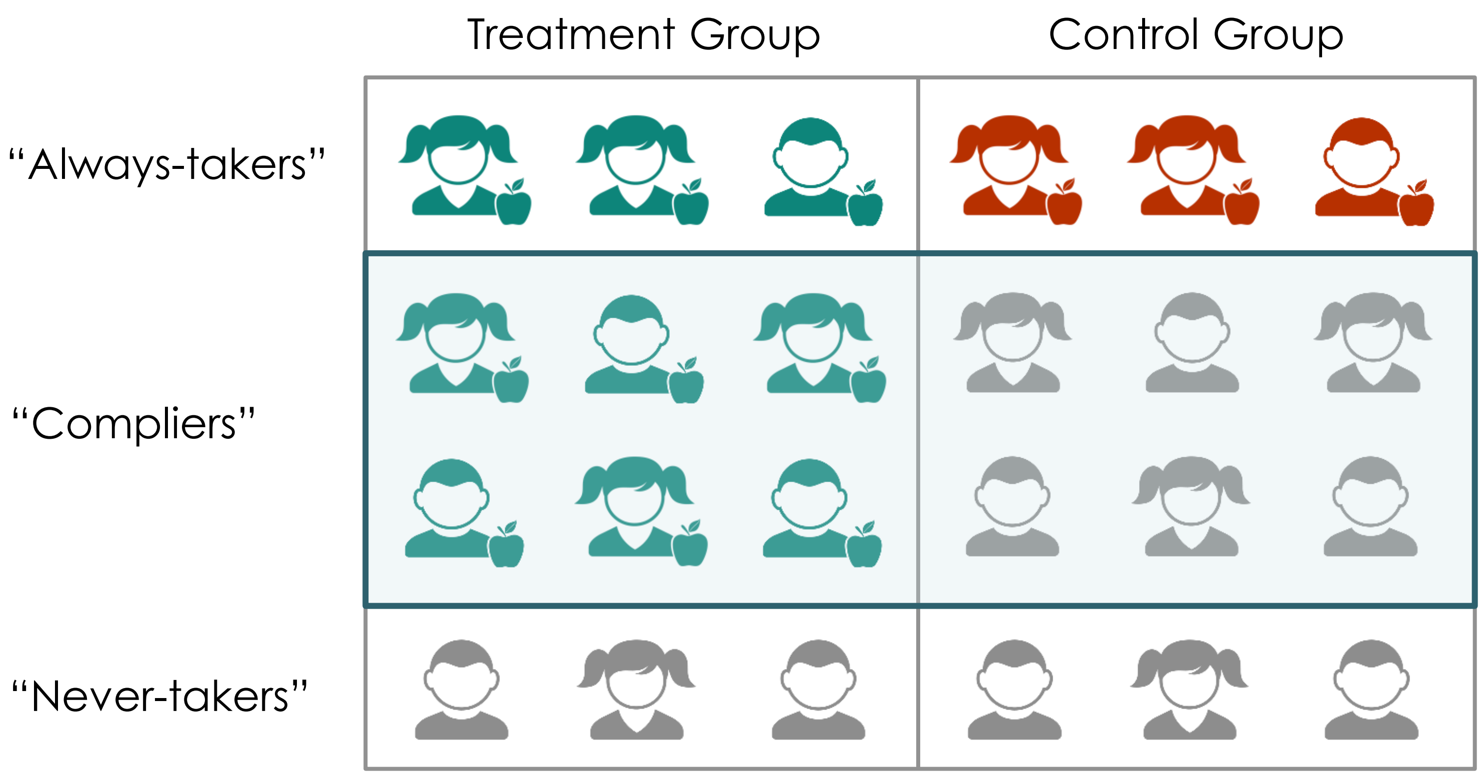


Division into the three categories



As the assignation was done randomly, the proportion of each category should be **similar in Treatment and Control**

Comparing the compliers



- To measure the impact of receiving the treatment, we compare ***compliers*** from Treatment and Control
- This measure of the impact is **"local"**: it is only valid for compliers. It can have a different impact for *always-takers* or *never-takers*.

LATE Estimator

What values do we need?

- Y(T)
- Y(C)
- Prob[treated | T]
- Prob[treated | C]

$$\frac{Y(T) - Y(C)}{Prob[treated|T] - Prob[treated|C]}$$

LATE estimator

School 1	Intention to treat?	Treated?	Observed 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
		Avg. Change Y(T) =	

School 2			
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	0
Pupil 10	No	No	0
		Avg. Change Y(C) =	

A = Gain if Treated
B = Gain if not Treated

ToT Estimator: A-B

$$\textbf{A-B} = \frac{Y(T)-Y(C)}{Prob(Treated \mid T)-Prob(Treated \mid C)}$$

Y(T)	
Y(C)	
Prob(Treated | T)	
Prob(Treated | C)	

Y(T)-Y(C)	
Prob(Treated | T)-Prob(Treated | C)	

A-B	

LATE estimator

School 1	Intention to treat?	Treated?	Observed 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
		Avg. Change Y(T) =	**3**
School 2			
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	0
Pupil 10	No	No	0
		Avg. Change Y(C) =	**0.9**

A = Gain if Treated
B = Gain if not Treated

ToT Estimator: A-B

$$\textbf{A-B} = \frac{Y(T)-Y(C)}{Prob(Treated \mid T)-Prob(Treated \mid C)}$$

Y(T)	3
Y(C)	0.9
Prob(Treated | T)	60%
Prob(Treated | C)	20%

Y(T)-Y(C)	2.1
Prob(Treated | T)-Prob(Treated | C)	40%

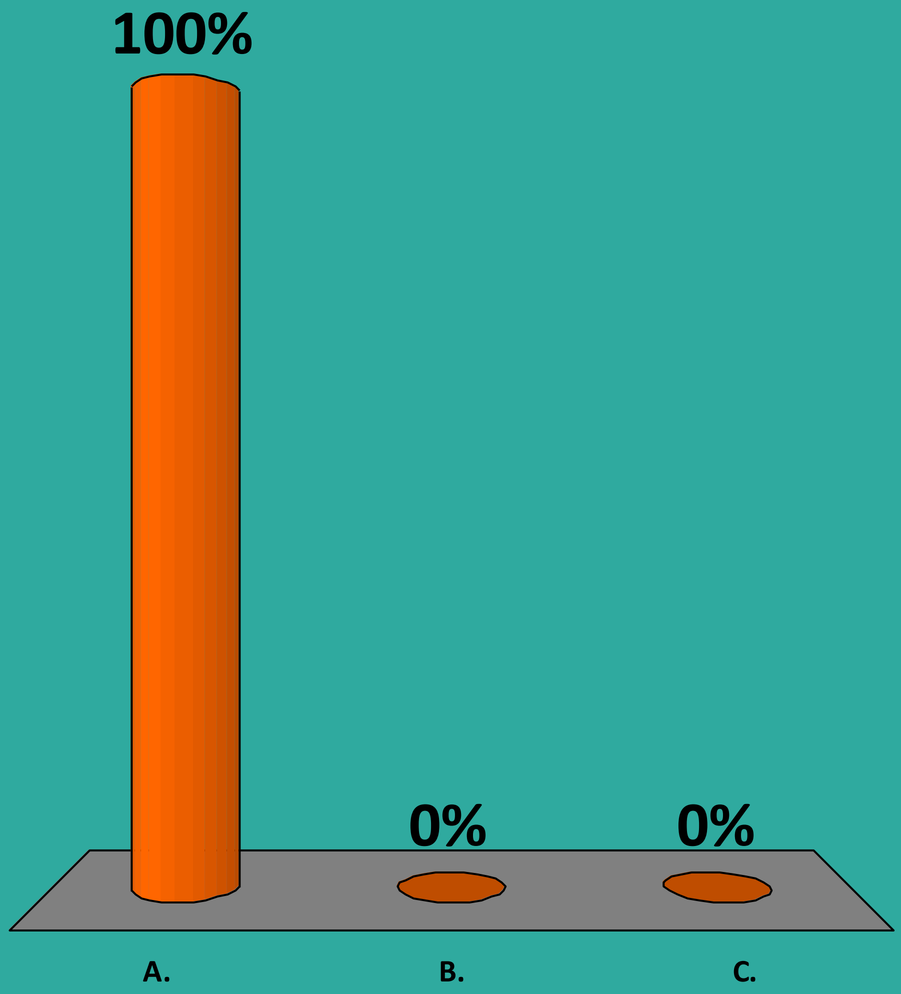
A-B	5.25

The ITT estimate will always be smaller (e.g., closer to zero) than the LATE estimate

A. True

B. False

C. Don't Know



LATE / ToT

- In academic papers, you will often see **"Treatment on the Treated"** (ToT)
- It is a way of analyzing the data that constitutes a **subset** of Local Average Treatment Effect (LATE)
- We talk of ToT when there are **non-compliers** in the Treatment group but **not in the Control group**

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 take-up 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:

- **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

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 limit evaluation-driven effects?

- Use a **different level** of randomization
- **Minimize salience** of evaluation as much as possible:
  - Do not announce phase-in (but useful to reduce attrition!)
  - Make sure staff is impartial and treats both groups similarly
- Consider including controls who are measured **at end-line** only
- Measure the evaluation-driven effects on 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

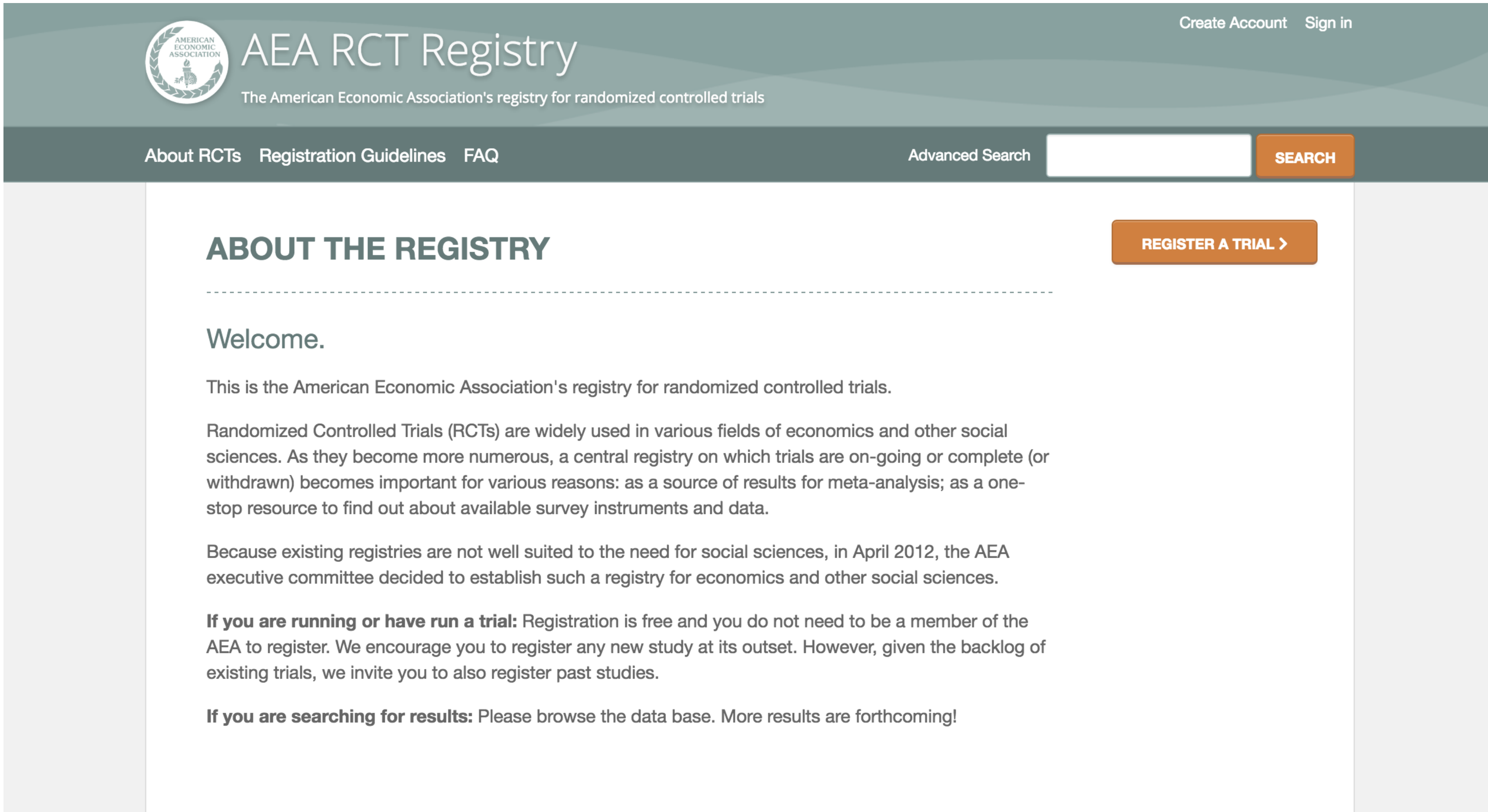
- Can we look at various outcomes?
- The more outcomes you look at, the higher the chance you find at least one significantly affected by the program
  - Pre-specify outcomes of interest
  - Report results on all measured outcomes, even null results
  - Correct statistical tests (Bonferroni)

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

Rule: Report both "raw" differences and regression-adjusted results

The AEA RCT Registry



Create Account Sign in

AEA RCT Registry

The American Economic Association's registry for randomized controlled trials

About RCTs Registration Guidelines FAQ Advanced Search SEARCH

REGISTER A TRIAL

ABOUT THE REGISTRY

Welcome.

This is the American Economic Association's registry for randomized controlled trials.

Randomized Controlled Trials (RCTs) are widely used in various fields of economics and other social sciences. As they become more numerous, a central registry on which trials are on-going or complete (or withdrawn) becomes important for various reasons: as a source of results for meta-analysis; as a one-stop resource to find out about available survey instruments and data.

Because existing registries are not well suited to the need for social sciences, in April 2012, the AEA executive committee decided to establish such a registry for economics and other social sciences.

If you are running or have run a trial: Registration is free and you do not need to be a member of the AEA to register. We encourage you to register any new study at its outset. However, given the backlog of existing trials, we invite you to also register past studies.

If you are searching for results: Please browse the data base. More results are forthcoming!

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).