



ABDUL LATIF JAMEEL
Poverty Action Lab

TRANSLATING RESEARCH INTO ACTION

YEF – ITCILO - JPAL

**Evaluating Youth Employment Programmes:
An Executive Course**

22 – 26 June 2015 | ITCILO Turin, Italy

Threats and Analysis

Bastien MICHEL

Aarhus University & TrygFonden's Centre

Course Overview



1. Introduction to Impact Evaluation
2. Measurements
3. How to Randomize
4. Sampling and Sample Size
- 5. Threats and Analysis**
6. Cost-Effectiveness Analysis and Scaling Up

Lecture Overview



- A. Attrition
- B. Spillovers
- C. Partial Compliance and Sample Selection Bias
- D. Intention to Treat & Treatment on Treated
- E. External validity
- F. Conclusion
- G. Some final recommendations

Lecture Overview



- A. Attrition**
- B. Spillovers
- C. Partial Compliance and Sample Selection Bias
- D. Intention to Treat & Treatment on Treated
- E. External validity
- F. Conclusion
- G. Some final recommendations

Attrition

- A. Is it a problem if some of the people in the experiment vanish before you collect your data?
 - A. It is a problem if the type of people who disappear is correlated with the treatment.
- B. Why is it a problem?
 - A. Loose the key property of RCT: two identical populations
- C. Why should we expect this to happen?
 - A. Treatment may change incentives to participate in the survey

Attrition bias: an example

- A. You want to evaluate the impact of a school feeding program
- B. The program was designed to:
 - Improve children's health
 - Increase children's school attendance (some children don't come to school because they are too weak)
- C. What impacts can we expect?
 - Increased attendance of weaker children
 - Improved nutrition
- D. As the main evaluator in charge of the study, you want to measure the impact on your main outcome: weight.
- E. Therefore, you go to all the schools in your sample (treatment and control) and measure everyone who is in school on a given day
- F. Will the treatment-control difference in weight be over-stated or understated?

Attrition bias: an example (no attrition)

Before Treatment		After Treatment	
T	C	T	C
20	20	22	20
25	25	27	25
30	30	32	30
Ave.			
Difference		Difference	

Attrition bias: an example (no attrition)

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

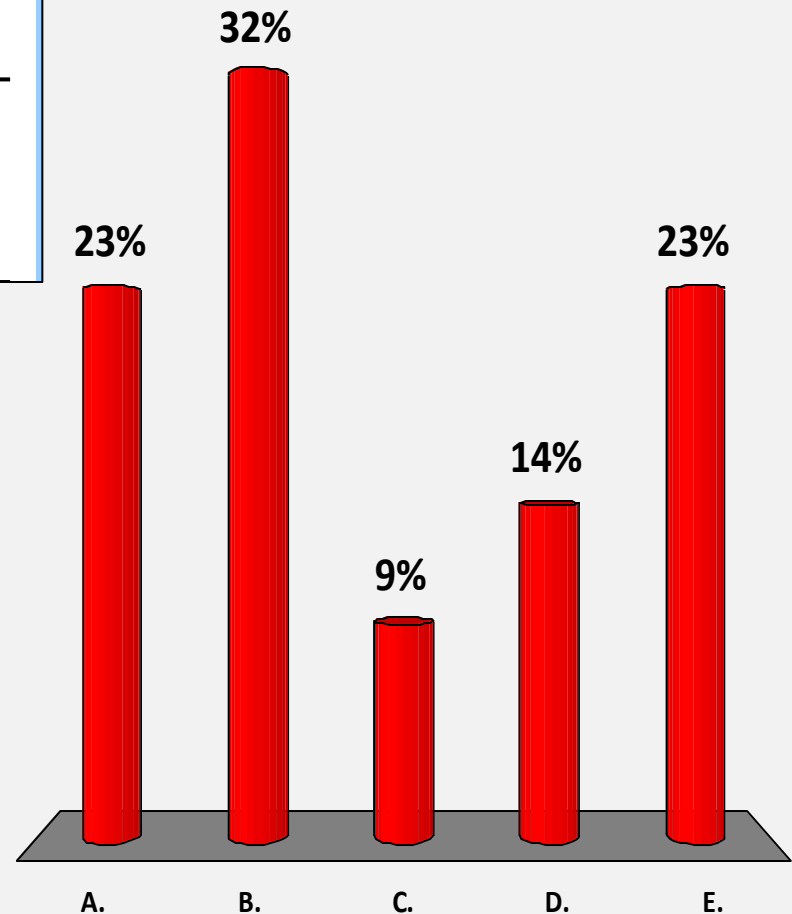
What if only children > 21 Kg come to school?



What if only children > 21 Kg come to school?

Before Treatment		After Treatment	
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
absent the program?

Before Treatment

T

C

[absent]

25

30

[absent]

25

30

Ave.

27,5

27,5

Difference

0

After Treatment

T

C

22

27

32

[absent]

25

30

27

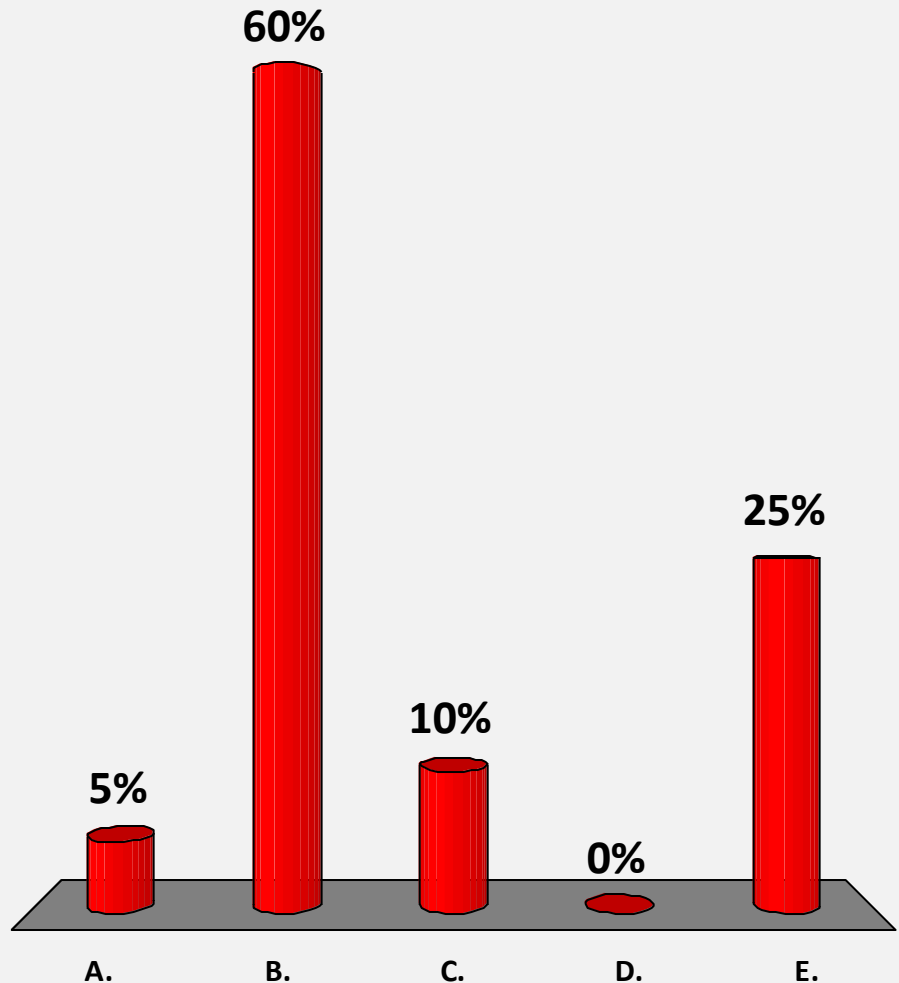
27,5

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



Attrition Bias



- A. Devote resources to tracking participants in the experiment
- B. If there is still attrition, check it does not induce any differences between your treatment and control groups.
- C. Good indication about validity of the first order property of the RCT:
 - Compare outcomes of two populations that only differ because one of them receive the program
- D. Internal validity

Attrition Bias

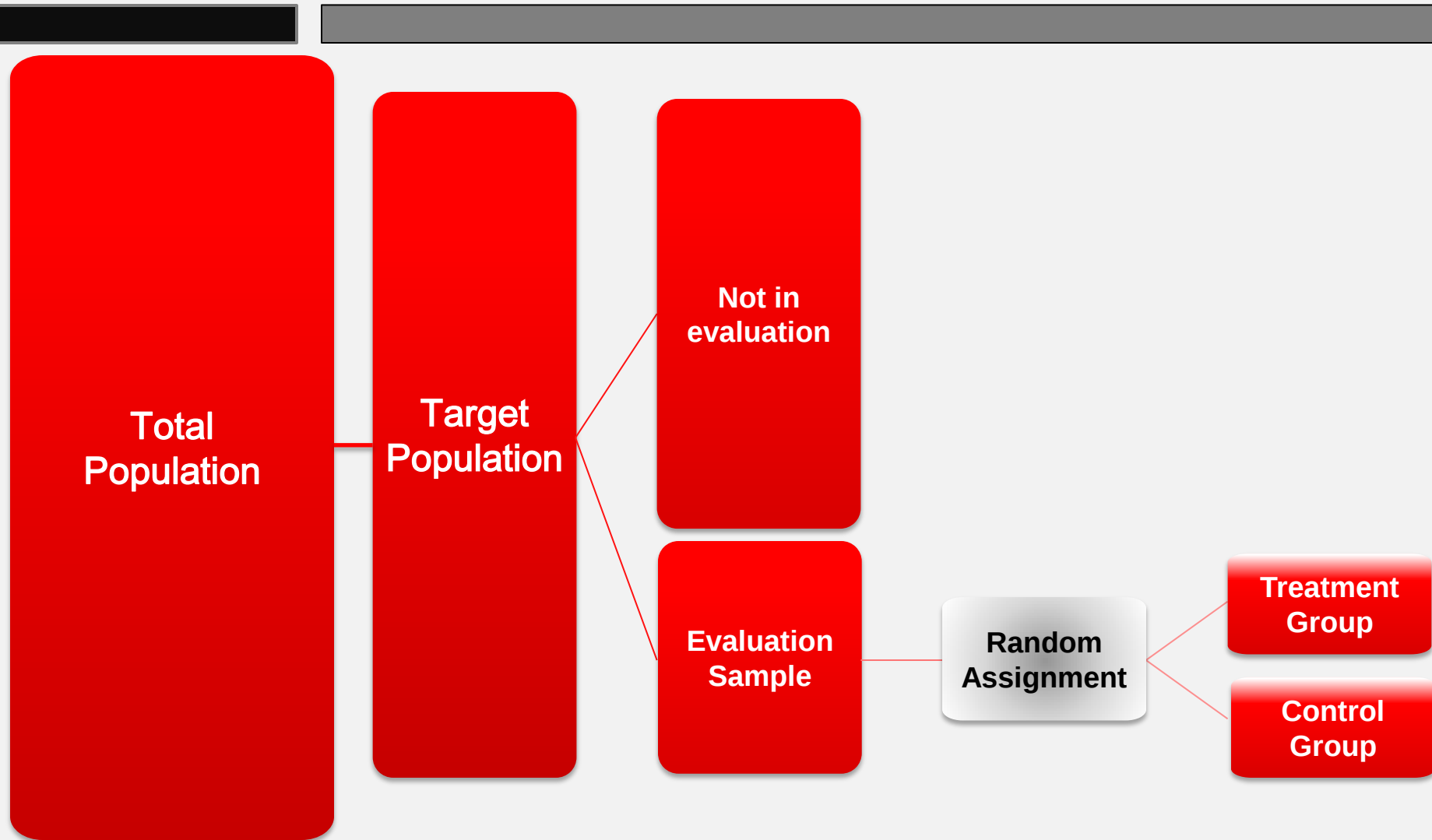
- A. If there is attrition but does not induce any imbalance between your treatment and control groups. Is this a problem?
- B. It can
- C. Assume only 50% of people in the test group and 50% in the control group answered the survey
- D. The comparison you are doing is a relevant parameter of the impact but... **on the population of respondent**
- E. But what about the population of non respondent
 - A. You know nothing!
 - B. Program impact can be very large on them,... or zero,... or negative!
- F. **External validity** might be at risk

Lecture Overview

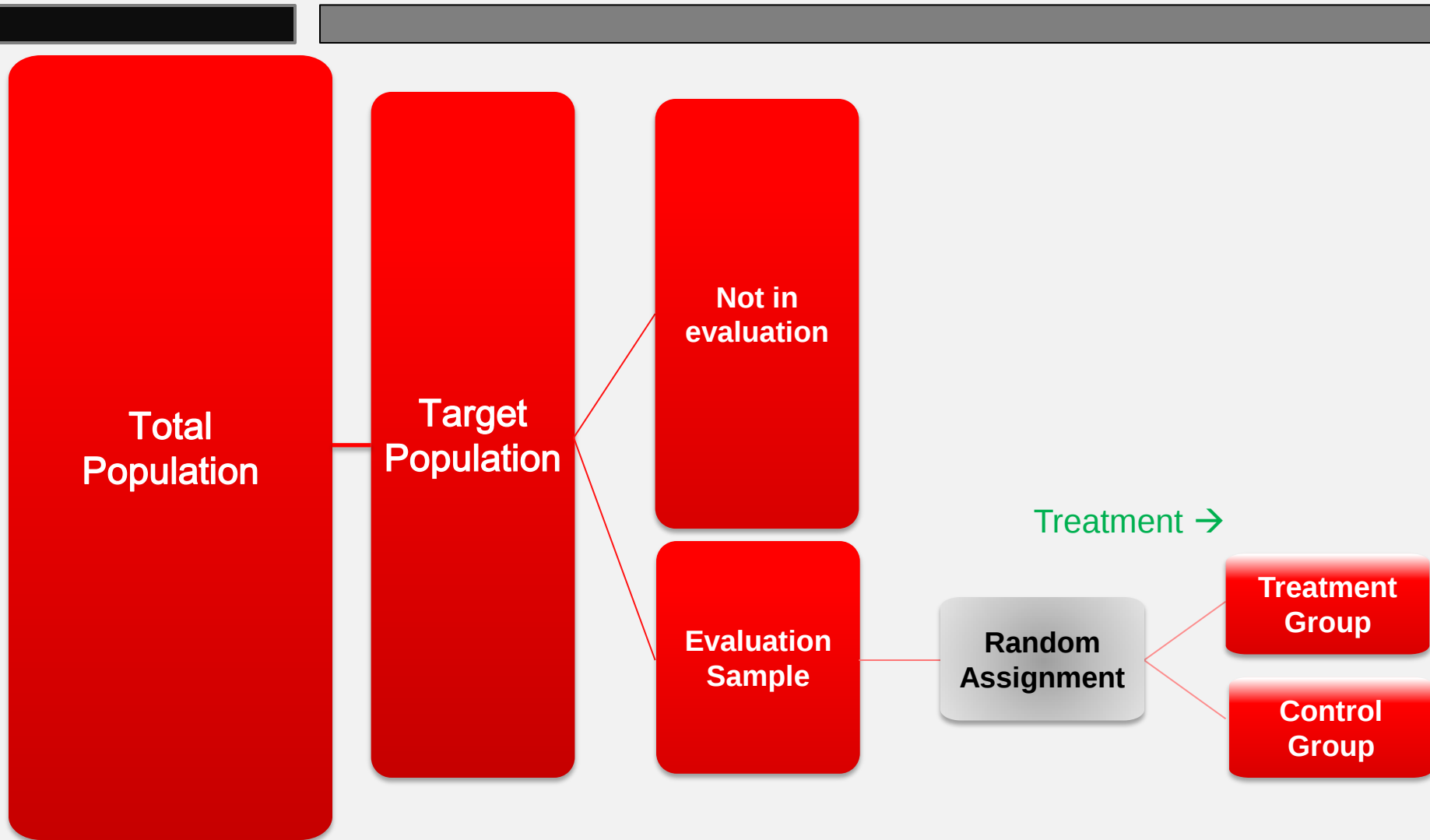


- A. Attrition
- B. Spillovers**
- C. Partial Compliance and Sample Selection Bias
- D. Intention to Treat & Treatment on Treated
- E. External validity
- F. Conclusion
- G. Some final recommendations

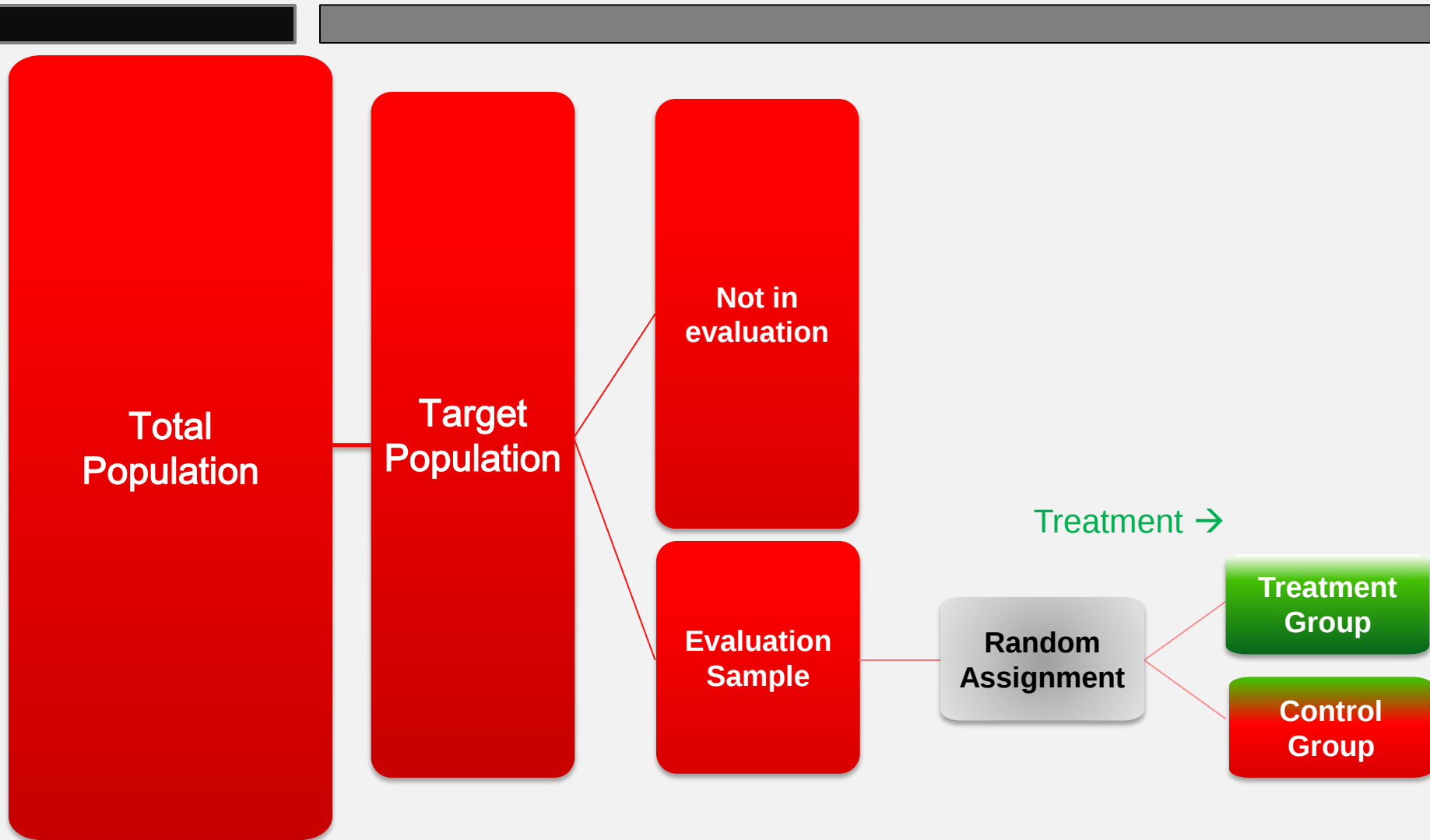
What else could go wrong?



Spillovers, contamination



Spillovers, contamination



Example: Vaccination for chicken pox

- A. Suppose you randomize chicken pox vaccinations within schools
 - A. Suppose that prevents the transmission of disease, what problems does this create for evaluation?
 - B. Suppose externalities are local? How can we measure impact?

Externalities Within School

Without Externalities			
School A	Treated?	Outcome	
Pupil 1	Yes	no chicken pox	Total in Treatment with chicken pox
Pupil 2	No	chicken pox	Total in Control with chicken pox
Pupil 3	Yes	no chicken pox	Treatment Effect
Pupil 4	No	chicken pox	
Pupil 5	Yes	no chicken pox	
Pupil 6	No	chicken pox	

With Externalities			
Suppose, because prevalence is lower, some children are not re-infected with chicken pox			
School A	Treated?	Outcome	
Pupil 1	Yes	no chicken pox	Total in Treatment with chicken pox
Pupil 2	No	no chicken pox	Total in Control with chicken pox
Pupil 3	Yes	no chicken pox	Treatment Effect
Pupil 4	No	chicken pox	
Pupil 5	Yes	no chicken pox	
Pupil 6	No	chicken pox	

Externalities Within School

Without Externalities				
School A	Treated?	Outcome		
Pupil 1	Yes	no chicken pox	Total in Treatment with chicken pox	0%
Pupil 2	No	chicken pox	Total in Control with chicken pox	100%
Pupil 3	Yes	no chicken pox	Treatment Effect	-100%
Pupil 4	No	chicken pox		
Pupil 5	Yes	no chicken pox		
Pupil 6	No	chicken pox		

With Externalities				
Suppose, because prevalence is lower, some children are not re-infected with chicken pox				
School A	Treated?	Outcome		
Pupil 1	Yes	no chicken pox	Total in Treatment with chicken pox	0%
Pupil 2	No	no chicken pox	Total in Control with chicken pox	67%
Pupil 3	Yes	no chicken pox	Treatment Effect	-67%
Pupil 4	No	chicken pox		
Pupil 5	Yes	no chicken pox		
Pupil 6	No	chicken pox		

How to measure program impact in the presence of spillovers?



Design your experiment such that it allows you to:

- Avoid spillovers (higher unit of randomization)

or

- Measure spillovers

Example: Price Information

- A. Providing farmers with spot and futures price information by mobile phone
- B. Should we expect spillovers?
- C. Randomize: individual or village level?
- D. Village level randomization
 - A. Less statistical power
 - B. “Purer control groups”
- E. Individual level randomization
 - A. More statistical power (if spillovers small)
 - B. But spillovers might bias the measure of impact

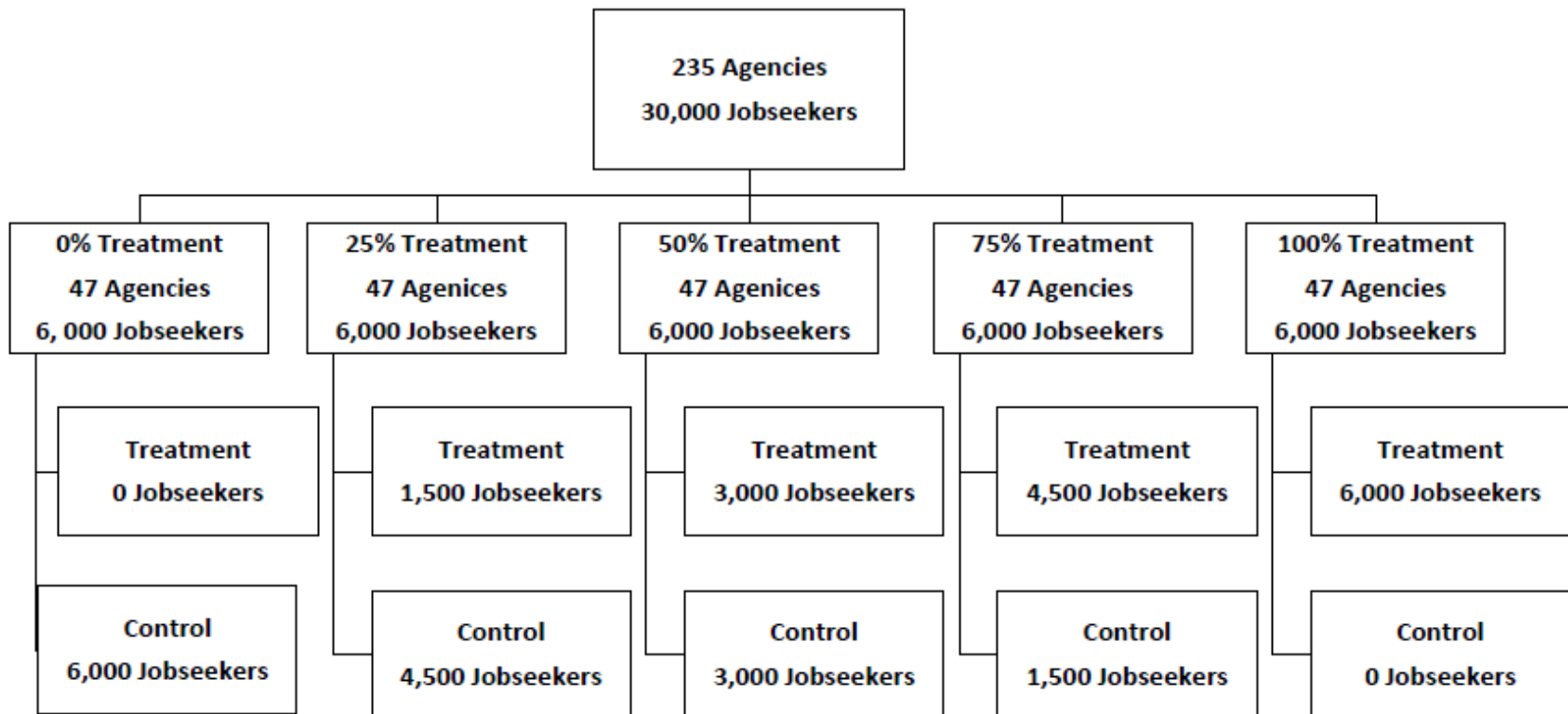
Example: Price Information

- A. Actually can do both together!
- B. Randomly assign villages into one of four groups, A, B and C
- C. Group A Villages
 - A. SMS price information to randomly selected 50% of individuals with phones
 - B. Two random groups: Test A and Control A
- D. Group B Villages
 - A. No SMS price information
- E. Allow to measure the true effect of the program: Test A/B
- F. Allow also to measure the spillover effect: Control A/B

How to measure program impact in the presence of spillovers?

Remember case study 2?

FIGURE 3



Lecture Overview



- A. Attrition
- B. Spillovers
- C. Partial Compliance and Sample Selection Bias**
- D. Intention to Treat & Treatment on Treated
- E. External validity
- F. Conclusion
- G. Some final recommendations

Sample selection bias



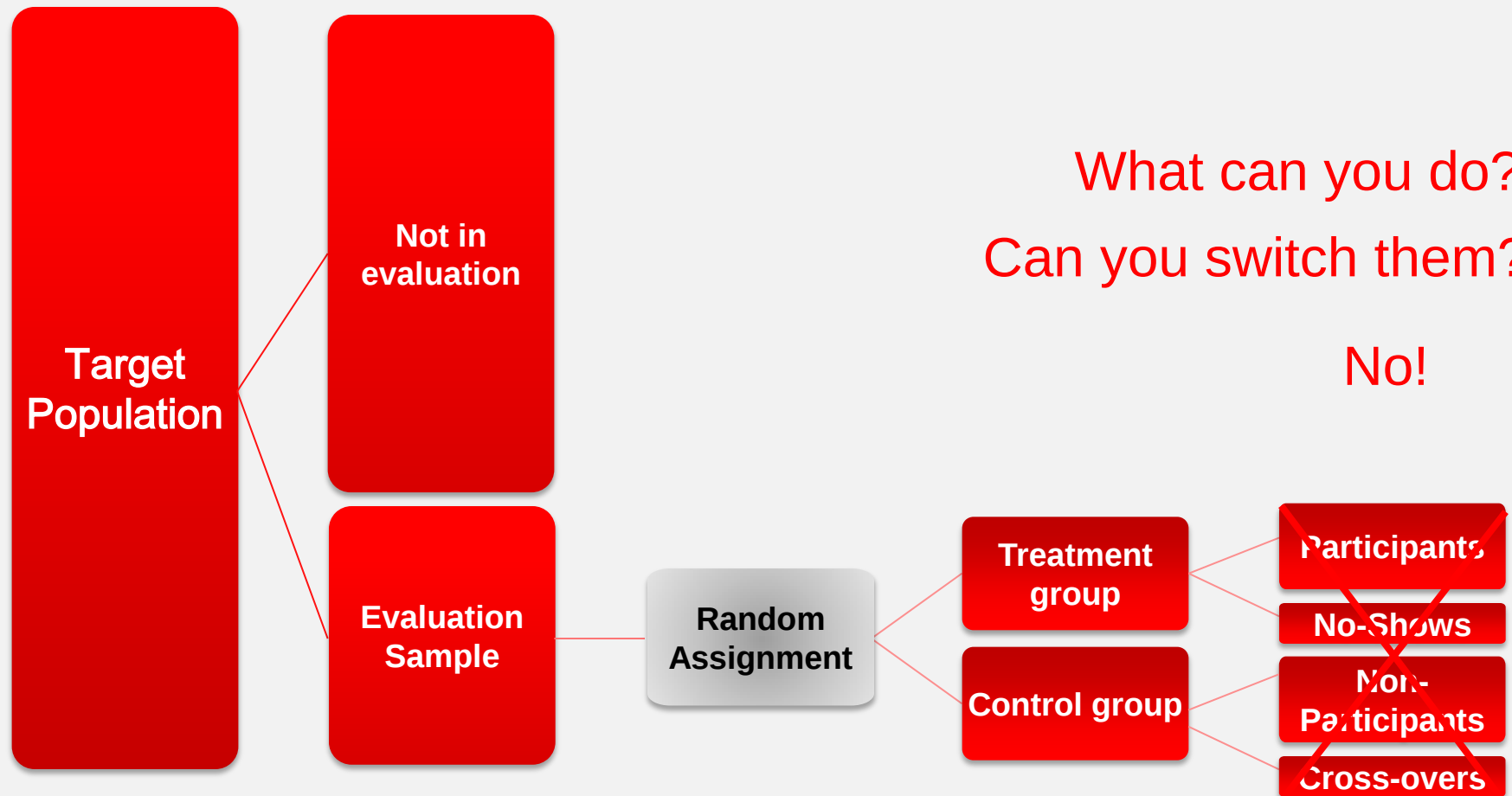
- A. Sample selection bias could arise if factors other than random assignment influence program allocation
- A. Even if intended allocation of program was random, the actual allocation may not be

Sample selection bias

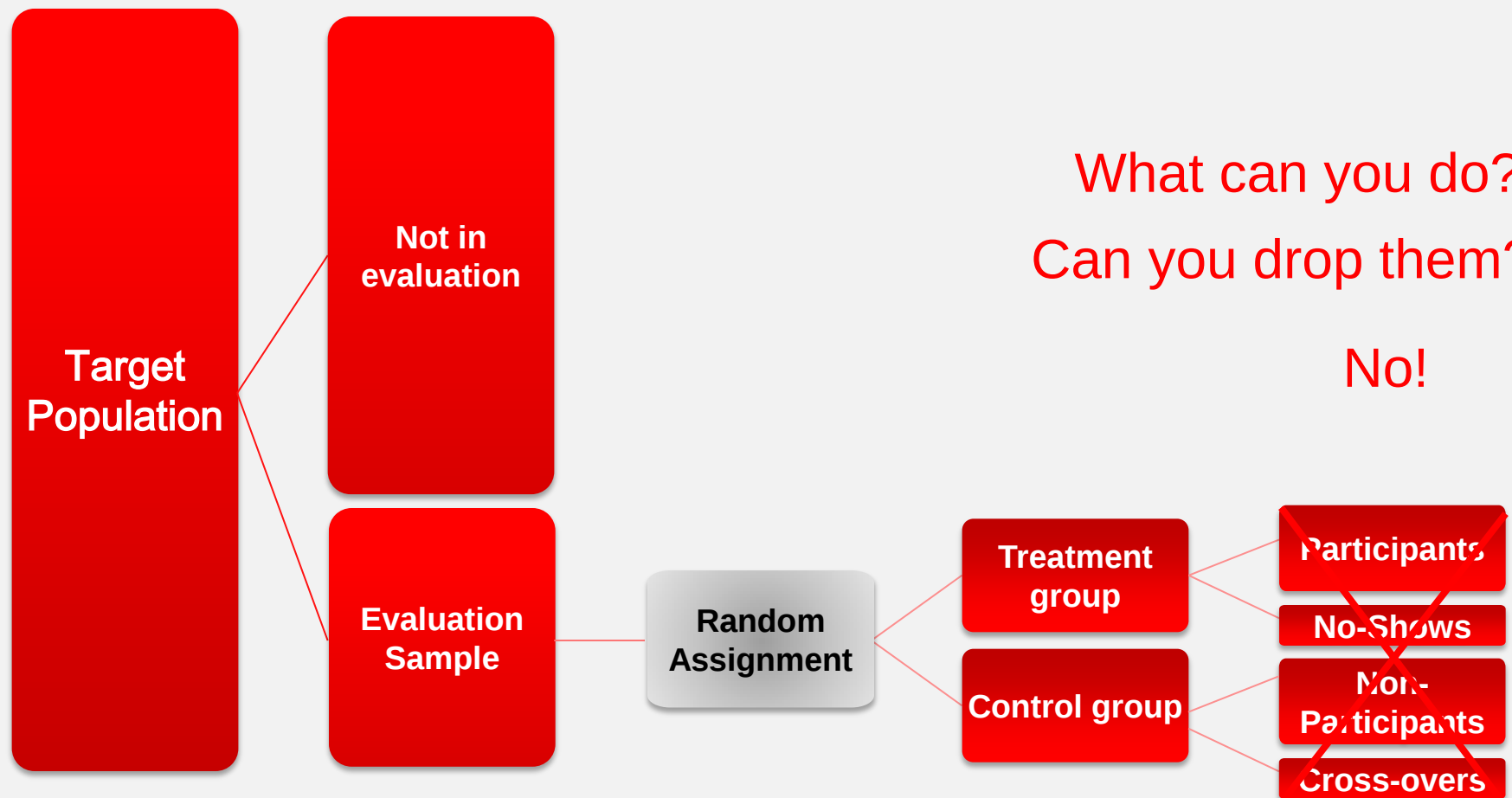


- A. Individuals assigned to comparison group could attempt to move into treatment group
 - A. School feeding program: parents could attempt to move their children from comparison school to treatment school
- B. Alternatively, individuals allocated to treatment group may not receive treatment
 - A. School feeding program: some students assigned to treatment schools bring and eat their own lunch anyway, or choose not to eat at all.

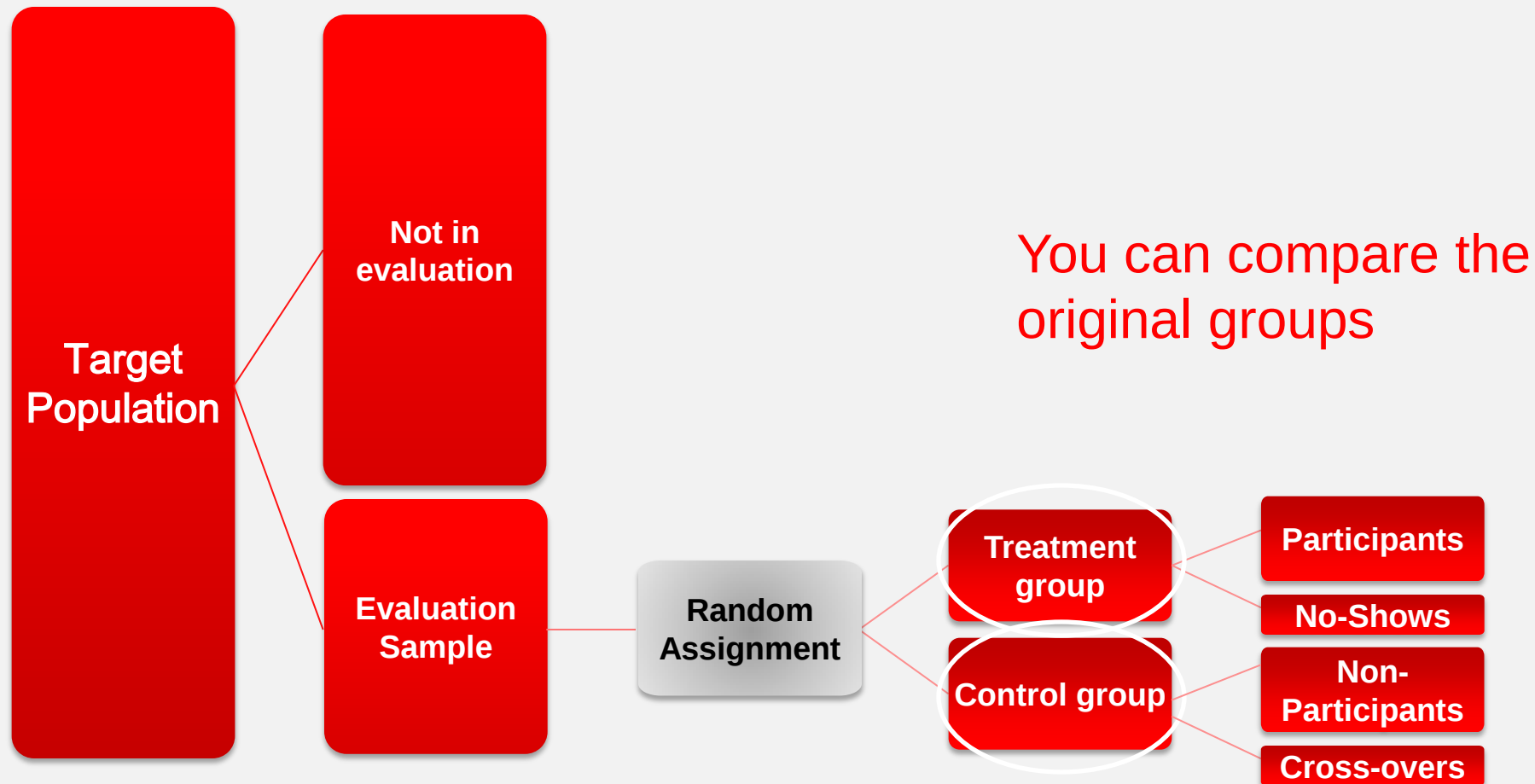
Non compliers



Non compliers



Non compliers



Lecture Overview



- A. Attrition
- B. Spillovers
- C. Partial Compliance and Sample Selection Bias
- D. Intention to Treat & Treatment on Treated**
- E. External validity
- F. Conclusion
- G. Some final recommendations

ITT and ToT



- A. Deworming campaign in villages
- B. Some people in treatment villages not treated
 - A. 78% of people assigned to receive treatment received some treatment
- C. How should we estimate the impact of the intervention? What can we do?

Intention to Treat (ITT)

A. What does “*intention to treat*” measure?

“What happened to the average child who is in a treated school in this population?”

A. Is this difference a causal effect? Yes because we compare two identical populations

B. But a causal effect of what?

A. Clearly not a measure of the deworming

B. Actually a measure of the global impact of the intervention

When is ITT useful?

- A. May relate more to actual programs
- B. For example, we may not be interested in the medical effect of deworming, but what would happen under an actual deworming program.
- C. If students often miss school and therefore don't get the treatment, the intention to treat estimate may actually be most relevant.

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	Intention to Treat ?	Treated?	Observed 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	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	Intention to Treat ?	Treated?	Observed 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	0
Pupil 10	no	no	0
Avg. Change among Not-Treated B=			0.9

School 1:

Avg. Change among Treated (A)

School 2:

Avg. Change among not-treated (B)

A-B

From ITT to effect of Treatment On the Treated

A. What about the impact on those who received the treatment?

Treatment On the Treated (TOT)

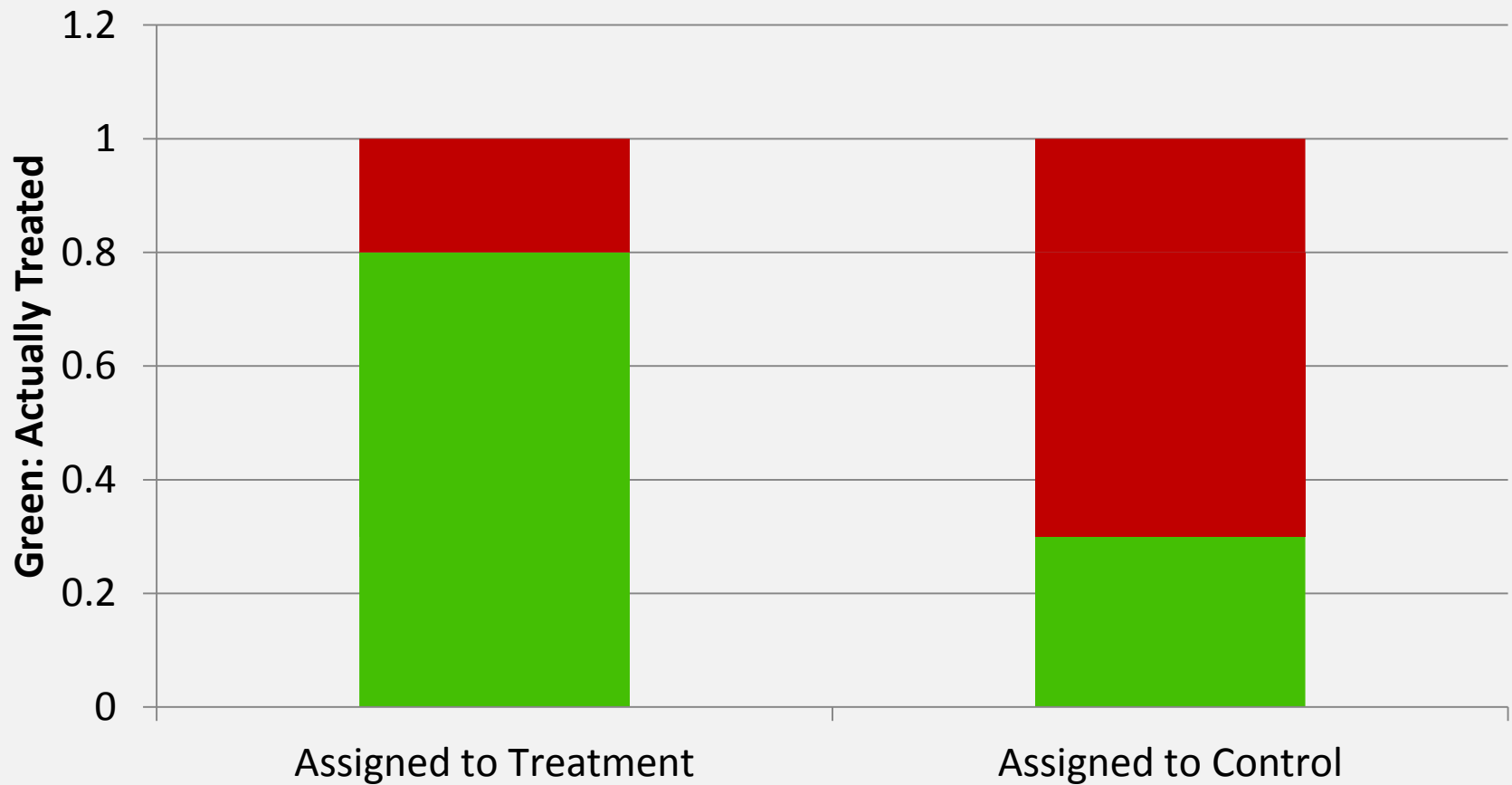
B. Is it possible to measure this parameter?

The answer is yes

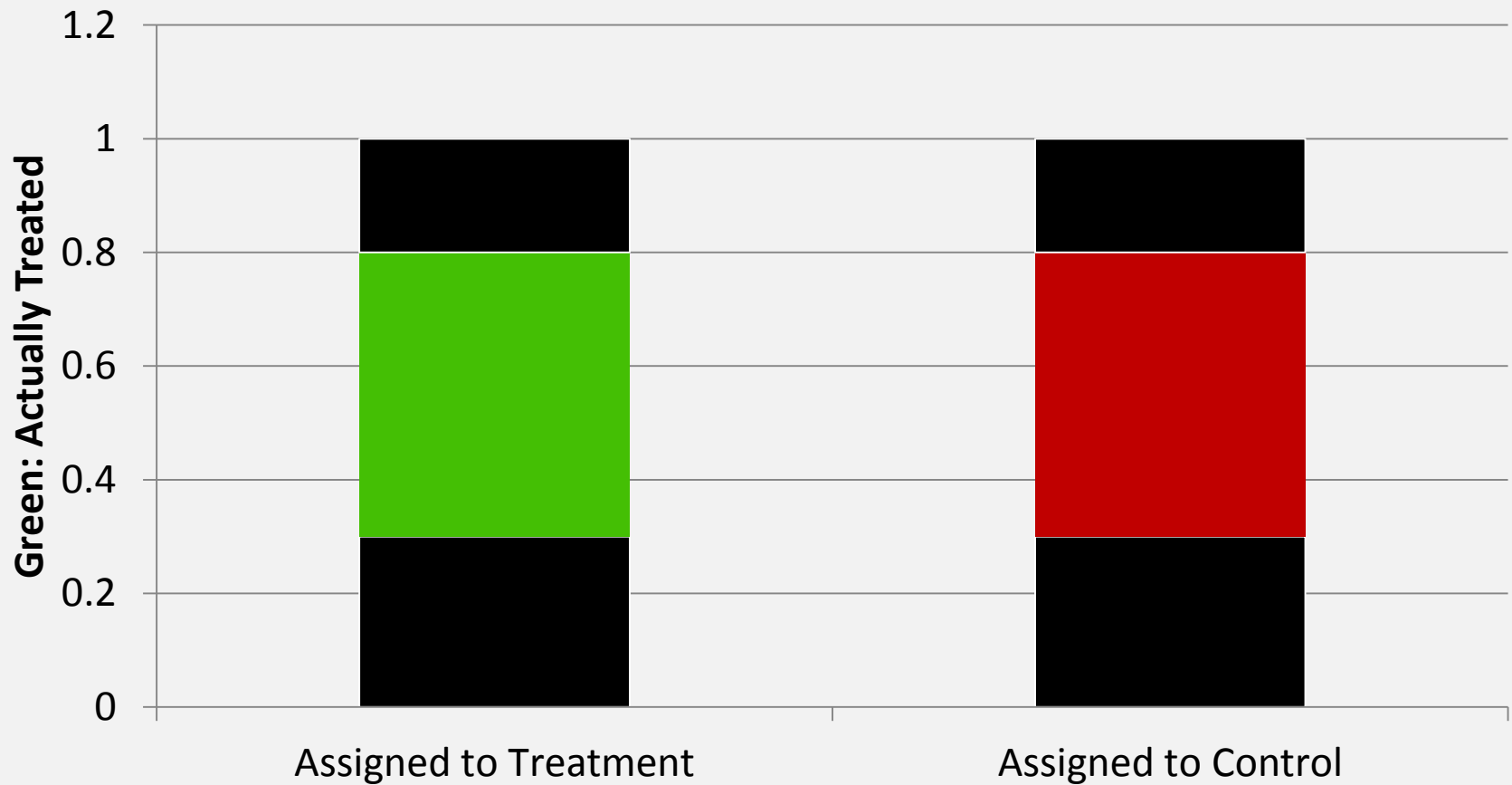
From ITT to effect of Treatment On the Treated (TOT)

- A. If there is such imperfect compliance, the difference observed between those **assigned to treatment** and those **assigned to control** is smaller than under compliance (the effect is somewhat diluted)
- B. But at the same time, the share of treated individuals in the treatment group is less than 1 and the share of treated individuals in the control group is superior to 0. We need to account for the share of individuals/probability of being treated.
- C. The **TOT** parameter “corrects” the ITT, scaling it up by this “take-up” difference

Estimating ToT from ITT: Wald



Interpreting ToT from ITT: Wald



Estimating TOT

- A. What values do we need?
 - B. $Y(AT)$ the average value over the Assigned to Treatment group (AT)
 - C. $Y(AC)$ the average value over the Assigned to Control group (AC)
-
- A. $\text{Prob}[T | AT] = \text{Proportion of treated in AT group}$
 - B. $\text{Prob}[T | AC] = \text{Proportion of treated in AC group}$
 - C. These proportion are called **take-up** of the program

Treatment on the treated (TOT)

A. Starting from a regression model

$$Y_i = a + B \cdot T_i + e_i$$

A. Angrist and Pischke show

$$B = [E(Y_i | Z_i=1) - E(Y_i | Z_i=0)] / [P(T_i=1 | Z_i=1) - P(T_i=1 | Z_i=0)]$$

A. With $Z=1$ is assignment to treatment group

Treatment on the treated (TOT)

$$B = [E(Y_i | Z_i=1) - E(Y_i | Z_i=0)] / [P(T_i=1 | Z_i=1) - E(T_i=1 | Z_i=0)]$$

A. Estimates will be

$$[Y(\mathbf{AT}) - Y(\mathbf{AC})] / [\text{Prob}[T | \mathbf{AT}] - \text{Prob}[T | \mathbf{AC}]]$$

A. The **ratio** of the **ITT** estimates on the **difference in take-up**

TOT estimate

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)=			

A = Gain if Treated
B = Gain if not Treated

TOT Estimator: A-B

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

School 2	Intention to Treat ?	Treated?	Observed 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	0
Pupil 10	no	no	0
Avg. Change Y(C) =			

Y(T)

Y(C)

Prob(Treated|T)

Prob(Treated|C)

Y(T)-Y(C)

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

A-B

TOT 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

A = Gain if Treated
B = Gain if not Treated

TOT Estimator: A-B

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

School 2	Intention to Treat ?	Treated?	Observed 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	0
Pupil 10	no	no	0
Avg. Change Y(C) =			0.9

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
-----	------

Generalizing the ToT Approach: Instrumental Variables

1. First stage regression

$$T = a_0 + a_1 Z + Xc + u$$

(a_1 is the difference in take-up)

2. Get predicted value of treatment:

$$\text{Pred}(T | Z, X) = a_0 + a_1 Z + Xc$$

3. Perform the regression of Y on predicted treatment instead on treatment

$$Y = b_0 + b_1 \text{Pred}(T | Z, X) + Xd + v$$

Requirements for Instrumental Variables

A. First stage

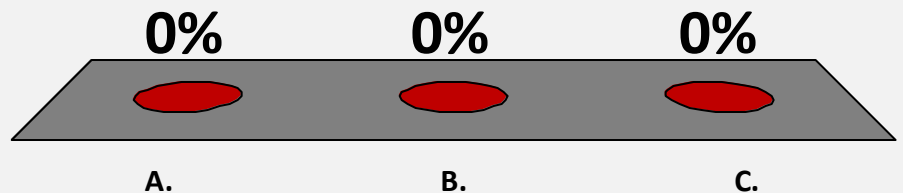
- A. Your experiment (or instrument) meaningfully affects probability of treatment
- B. Actually the experiment is “good” if there is a large effect of assignment to treatment on treatment participation (the difference in take-up)

B. Exclusion restriction

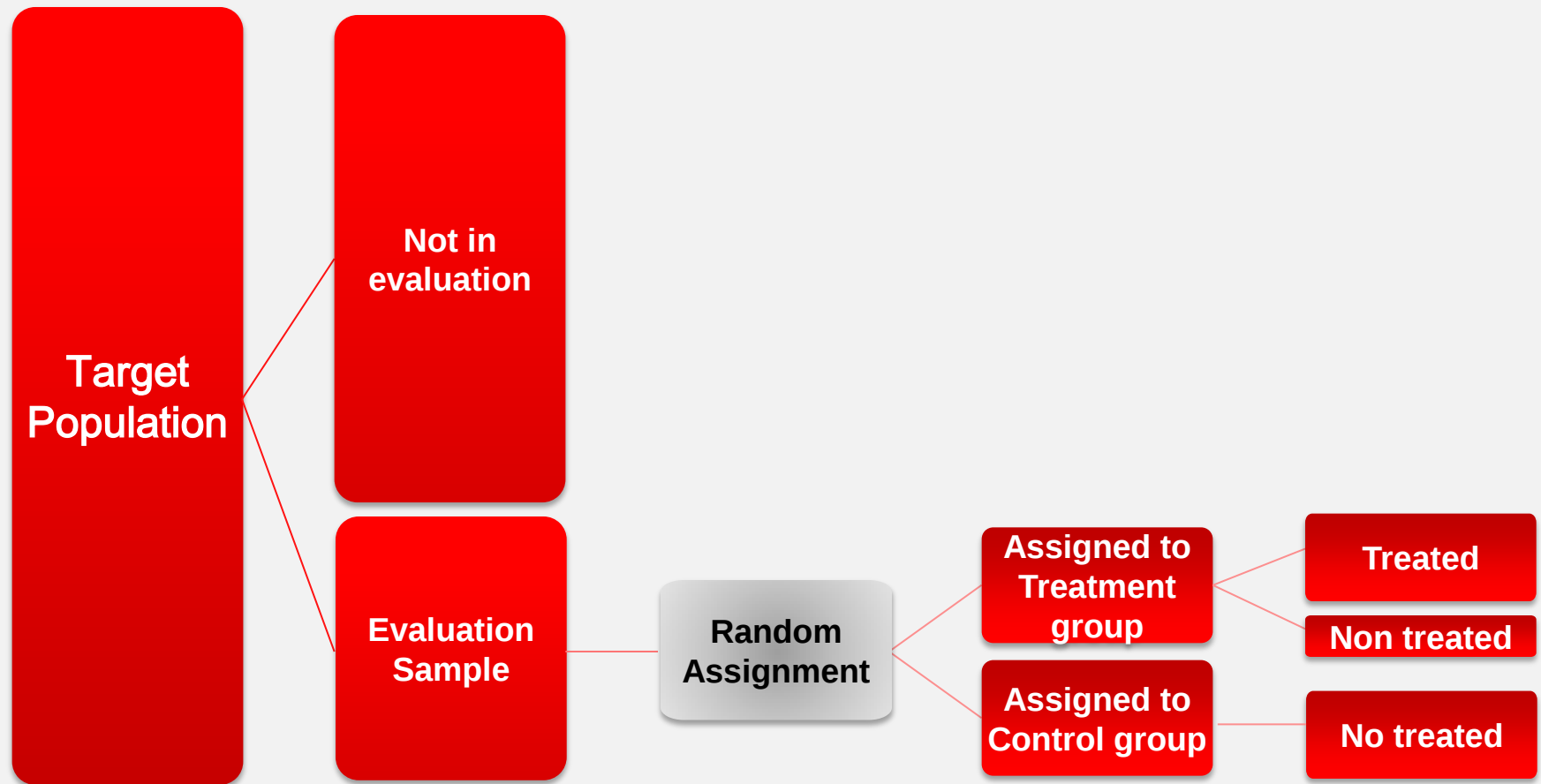
- A. Your experiment (or instrument) does not affect outcomes through another channel

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

- A. True
- B. False
- C. Don't Know



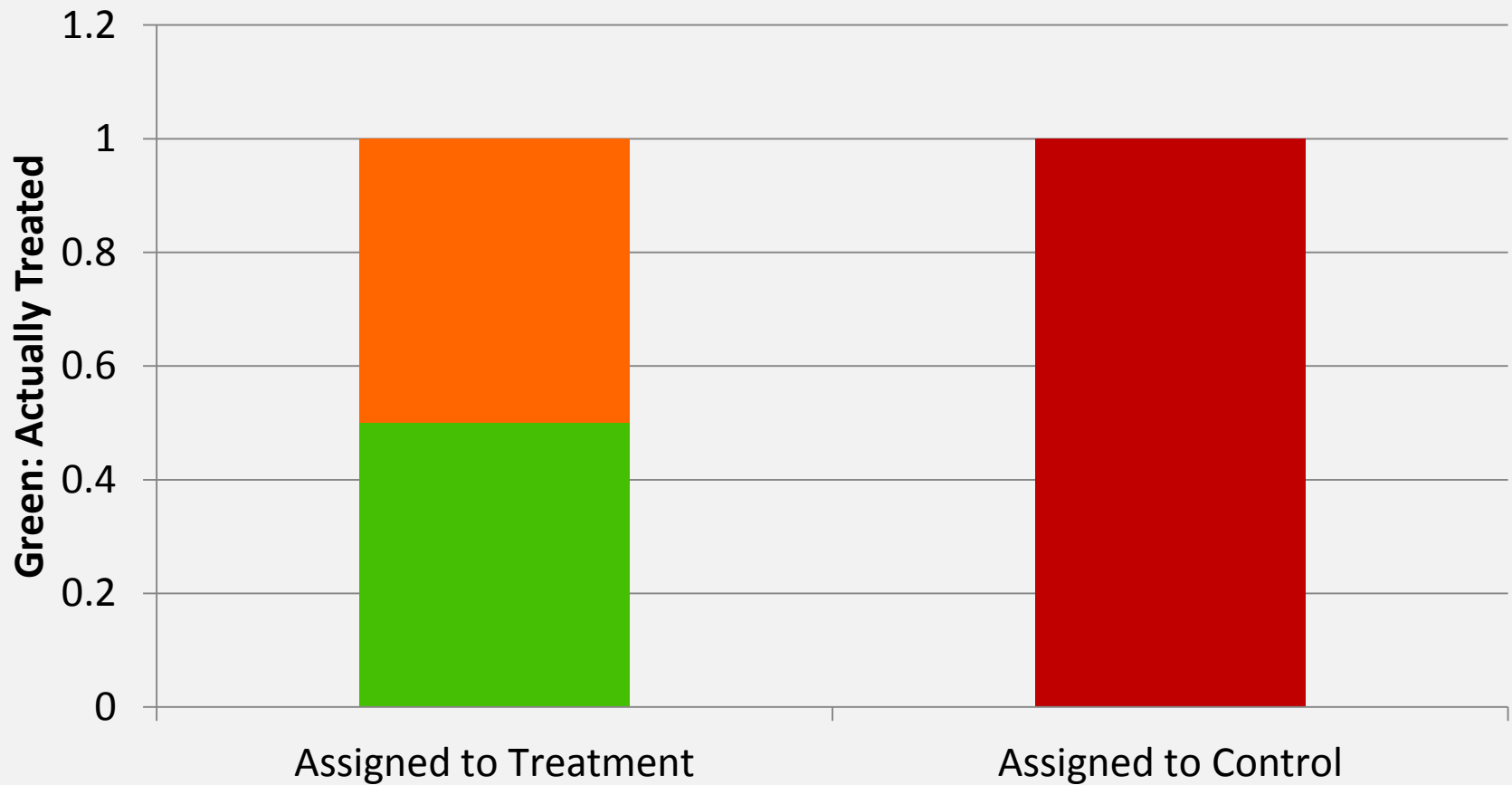
TOT not always appropriate...



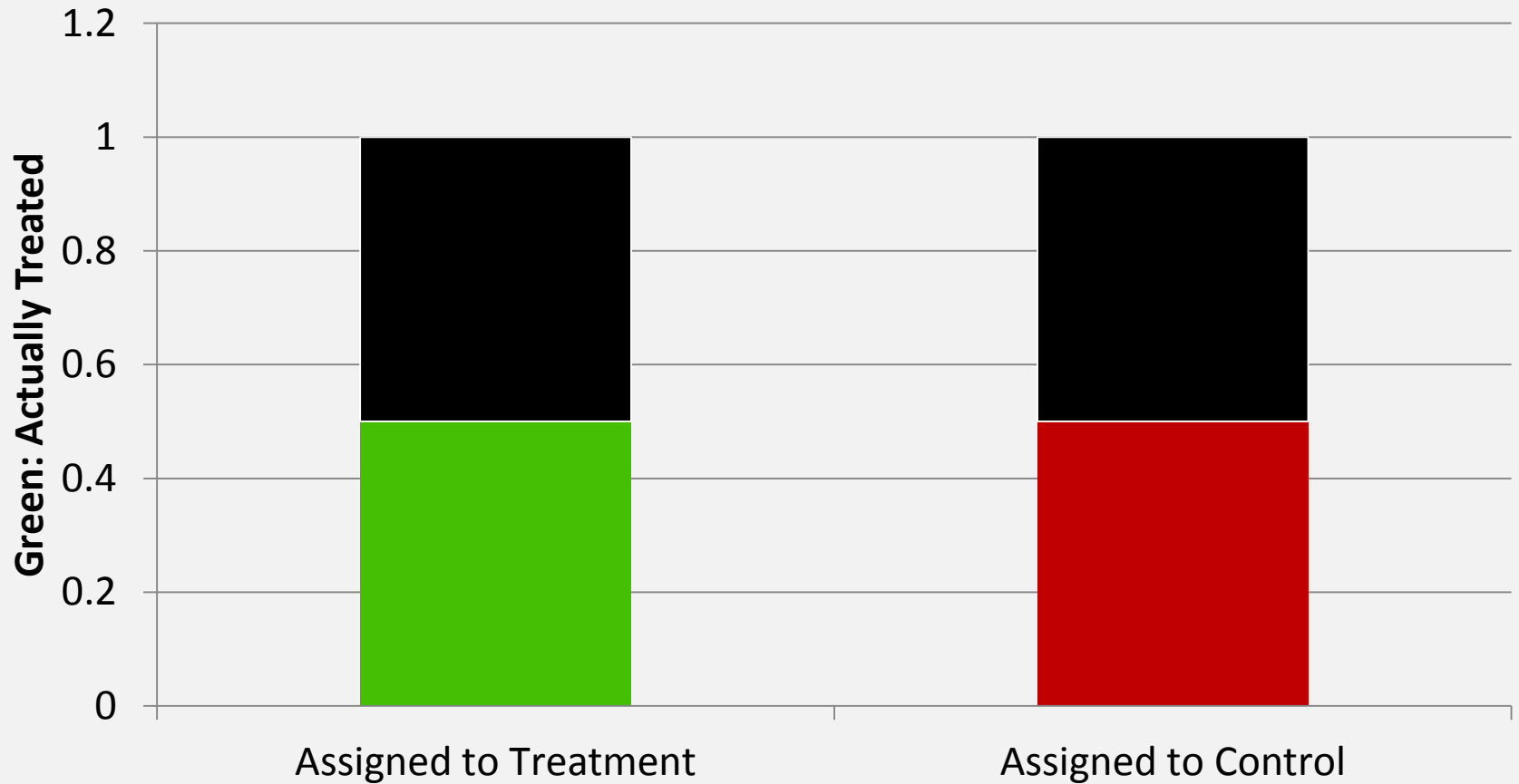
TOT not always appropriate...

- A. Example: send 50% of retired people in Paris a letter warning of flu season, encourage them to get vaccines
- B. Suppose 50% in treatment, 0% in control get vaccines
- C. Suppose incidence of flu in treated group drops 35% relative to control group
- D. Is $(.35) / (.5 - 0) = 70\%$ the correct estimate?
- E. What effect might letter alone have?
- F. Some retired people in the assignment to treatment group might consider it is better not to get a vaccine but... to stay home
- G. They didn't get the treatment but they have been influenced by the letter

Non treated in the AT group impacted



Non treated in AT group do not cancel out



Lecture Overview

- A. Spillovers
- B. Partial Compliance and Sample Selection Bias
- C. Intention to Treat & Treatment on Treated
- D. External validity**
- E. Conclusion
- F. Some final recommendations

Threat to external validity:



- A. Behavioral responses to evaluations
- B. Generalizability of results

Threat to external validity:

Behavioral responses to evaluations

- One limitation of evaluations is that the evaluation itself may cause the treatment or comparison group to change its behavior
 - Treatment group behavior changes: Hawthorne effect
 - Comparison group behavior changes: John Henry effect
- Minimize impact of the evaluation itself on respondents' life, should be as discrete as possible
- Consider including controls who are measured at endline only
- Use good admin data if available

Generalizability of results



A. Depend on three factors:

- A. Program Implementation: can it be replicated at a large (national) scale?
- B. Sensitivity of results: would a similar, but slightly different program, have same impact?
- C. Study Sample: is it representative?

Lecture Overview

- A. Spillovers
- B. Partial Compliance and Sample Selection Bias
- C. Intention to Treat & Treatment on Treated
- D. External validity
- E. Conclusion**
- F. Some final recommendations

Conclusion



- A. There are many threats to the internal and external validity of randomized evaluations...
- B. ...as there are for every other type of study
- C. Randomized trials:
 - A. Facilitate simple and transparent analysis
 - B. Allow clear tests of validity of experiment

Lecture Overview



- A. Spillovers
- B. Partial Compliance and Sample Selection Bias
- C. Intention to Treat & Treatment on Treated
- D. External validity
- E. Conclusion
- F. Some final recommendations**

Recommendations



Start early...

- A. Design
- B. IRB
- C. Funding
- D. Authorizations
- E. Sampling
- F. Questionnaires
- G. ...

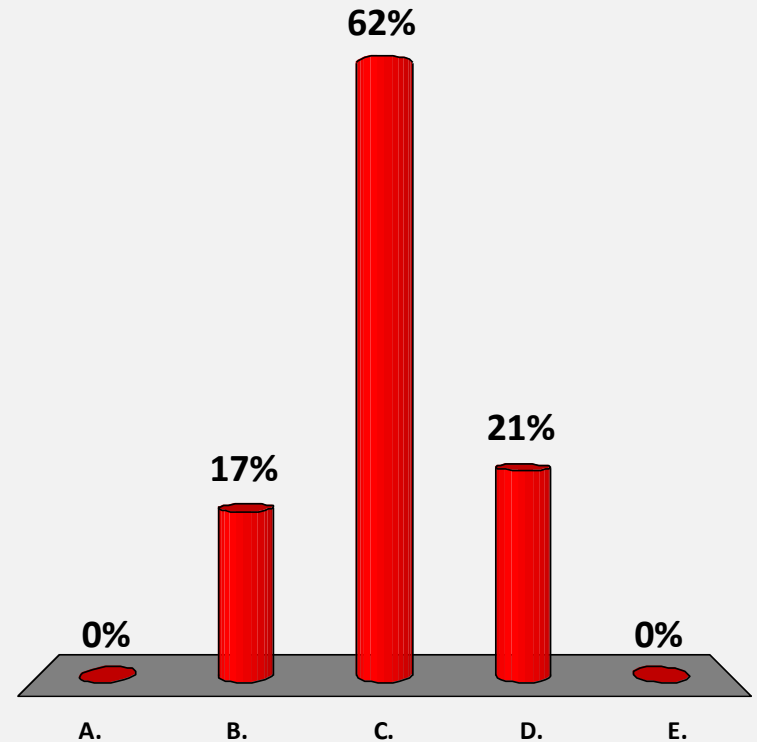
Further resources



- A. Using Randomization in Development Economics Research: A Toolkit (Duflo, Glennerster, Kremer)
- B. Mostly Harmless Econometrics (Angrist and Pischke)
- C. Identification and Estimation of Local Average Treatment Effects (Imbens and Angrist, Econometrica, 1994).

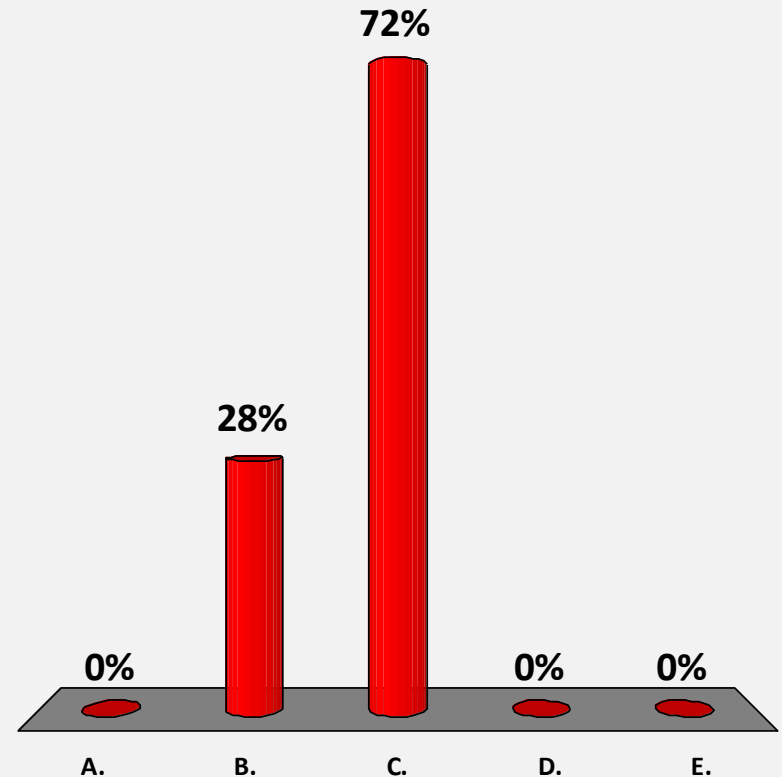
How was the length of this presentation?

- A. Unbearably long
- B. Too long
- C. Just right
- D. Not long enough
- E. Too short – more time, please!



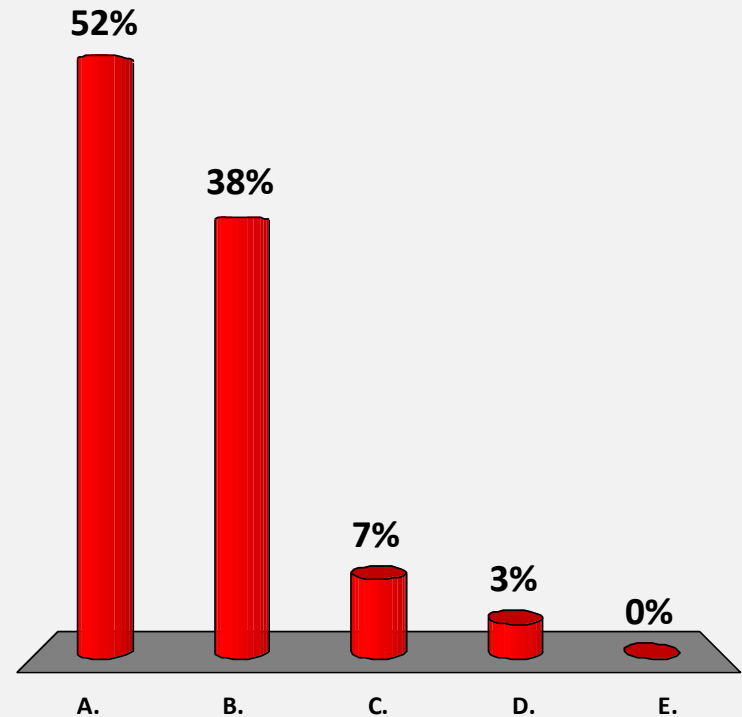
How was the pace of this presentation?

- A. Too fast! I couldn't keep up.
- B. Rushed
- C. Just right
- D. Slow
- E. Too slow, I fell asleep.



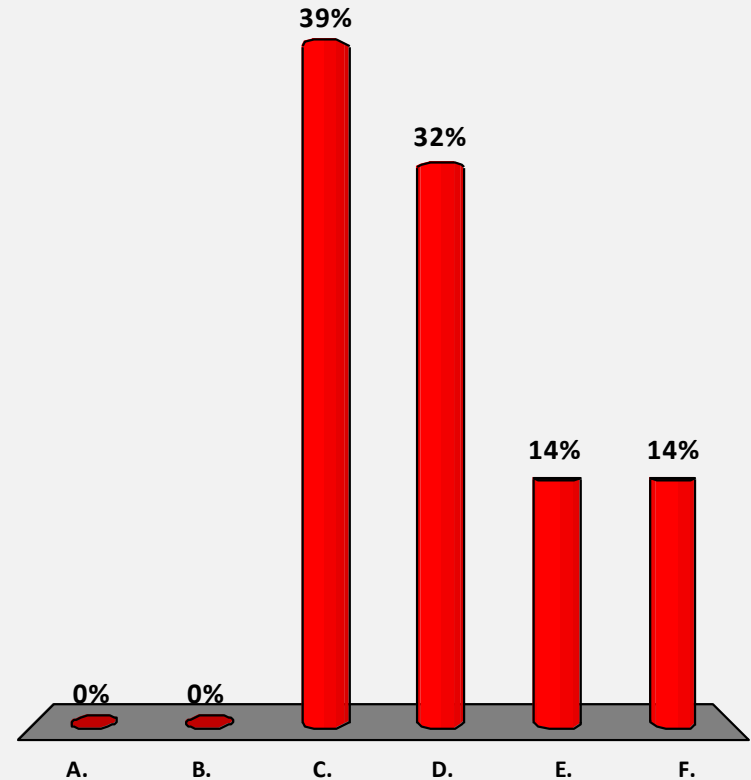
Was the content relevant to your work?

- A. Very relevant
- B. Quite useful
- C. Perhaps
- D. Not really
- E. No – not useful at all.



Before today, how much of this material did you already feel comfortable/ proficient in?

- A. 100%
- B. 80%
- C. 60%
- D. 40%
- E. 20%
- F. < 20%



After this presentation, how much of this material do you feel proficient in?

- A. 100%
- B. 80%
- C. 60%
- D. 40%
- E. 20%
- F. < 20%

