Evaluating the impact of social programs: Randomized Control Trials

Olivia D'Aoust

Development Economics and Management

April 4, 2014

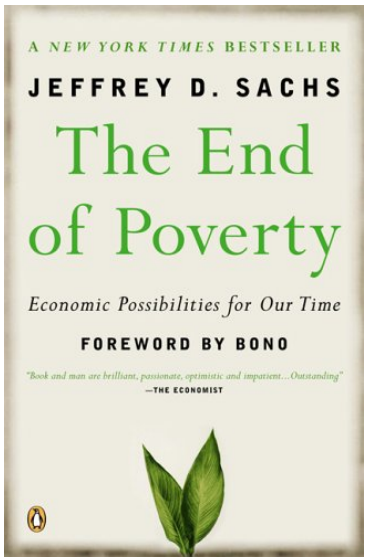
Introduction

What do we know about the impact of development programmes?

Introduction

What do we know about the impact of development programmes ?

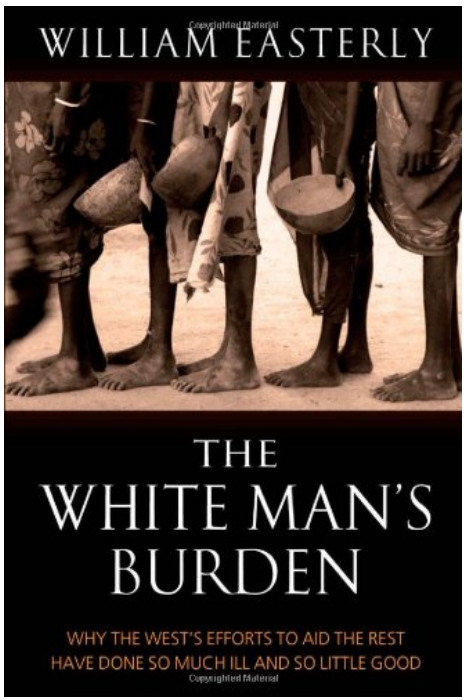
1. Aid optimists
   - Jeffrey Sachs - The end of poverty
   - *Do more* approach



Introduction

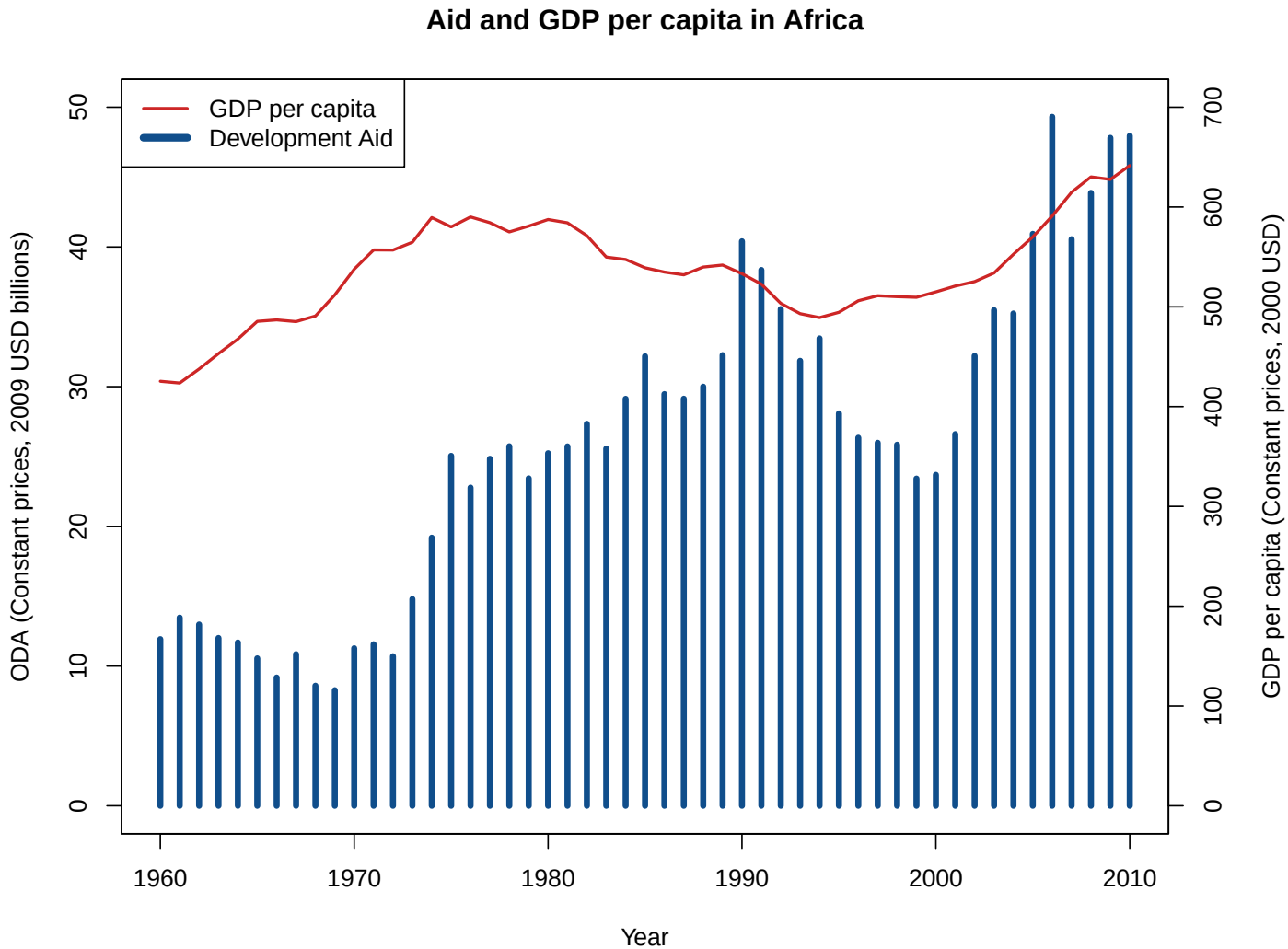
What do we know about the impact of development programmes ?

2. Aid pessimists
   - William Easterly - The White Man's Burden
   - *The poverty Puzzle* approach
   - what would have happen without the aid ? (counterfactual)



Introduction

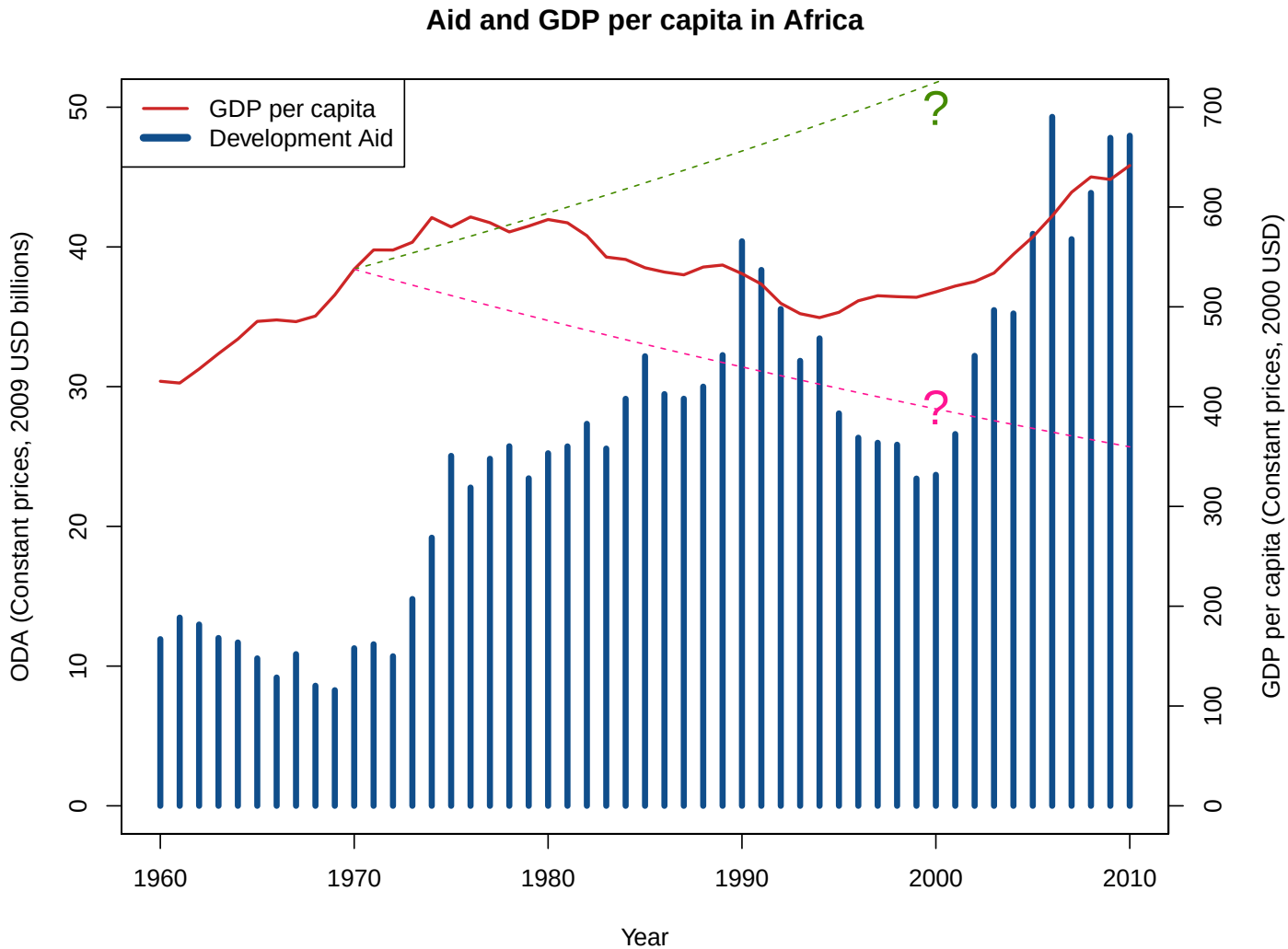
What do we know about the impact of development aid?



Sources: Easterly (2008) Can the West Save Africa ; WDI and QWIDS/DAC, 2011

Introduction

What would have happen if there had been no aid?



Sources: Easterly (2008) Can the West Save Africa ; WDI and QWIDS/DAC, 2011

Evaluating Social Programs I

When implementing a public policy, you want to know

- if there has been an impact?
- What is the nature and magnitude of the impact?

Examples of policies

- Organizing parent-school meetings on how to improve their involvement in their children's education
- Counselling the unemployed
- Distributing deworming drugs to children in Kenya
- Distributing bed nets
- Introducing microcredit in a region

Evaluating Social Programs II

Problem

How do you increase school attendance in developing countries?

Since low enrolment $\rightarrow$ High dropout $\rightarrow$ Low attendance $\rightarrow$ Low achievement

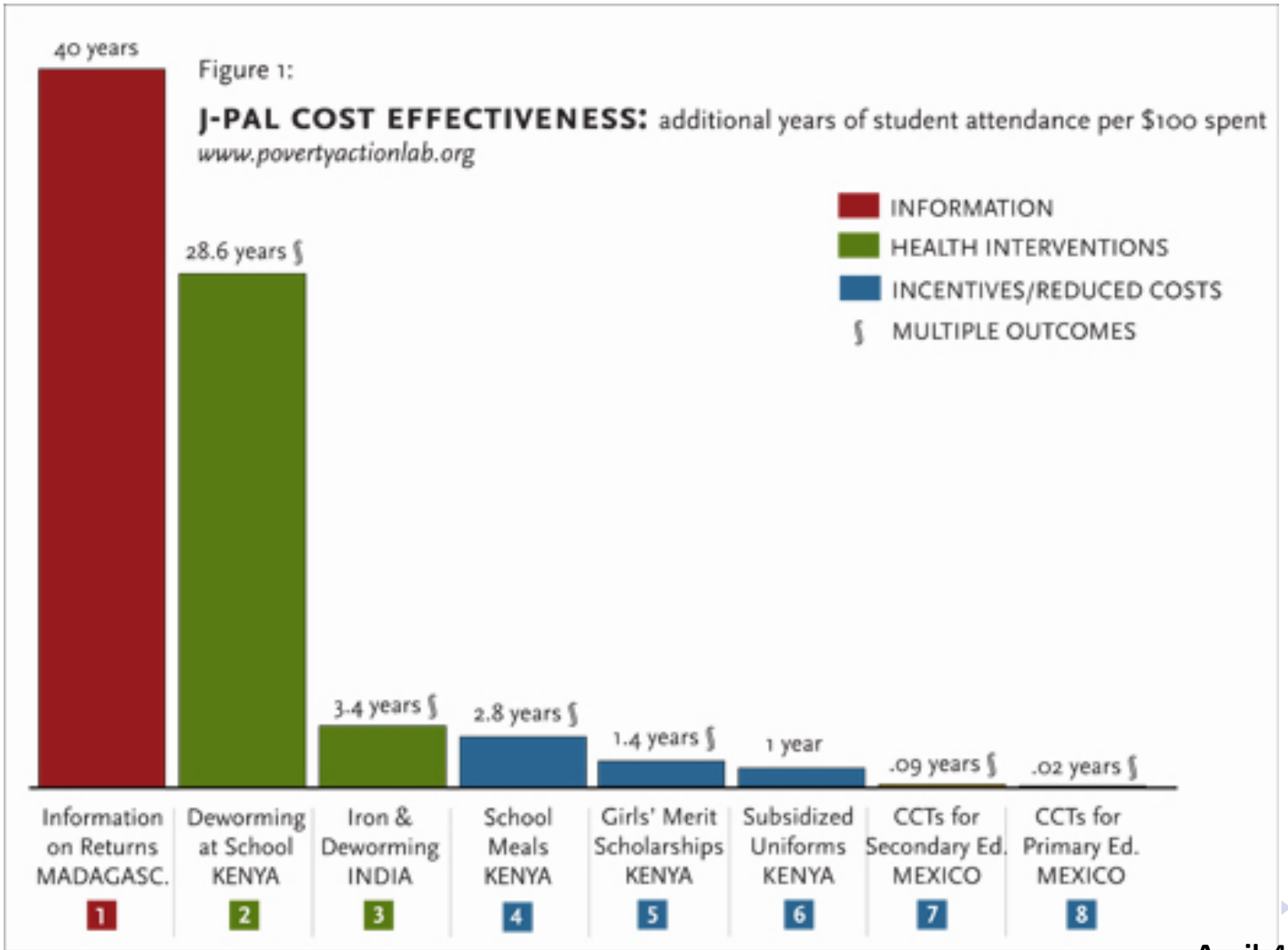
- Books for Free
- Building Improvements
- Cash Grants for Attendance
- Deworming
- Lunch for Free
- Remedial Education
- School Supplies (Blackboards, Chalk, etc.)
- Teacher Attendance (Monitoring)
- Uniforms for Free

??

Lesson learning and accountability

We should know

- which programs work
- which programs work **best**



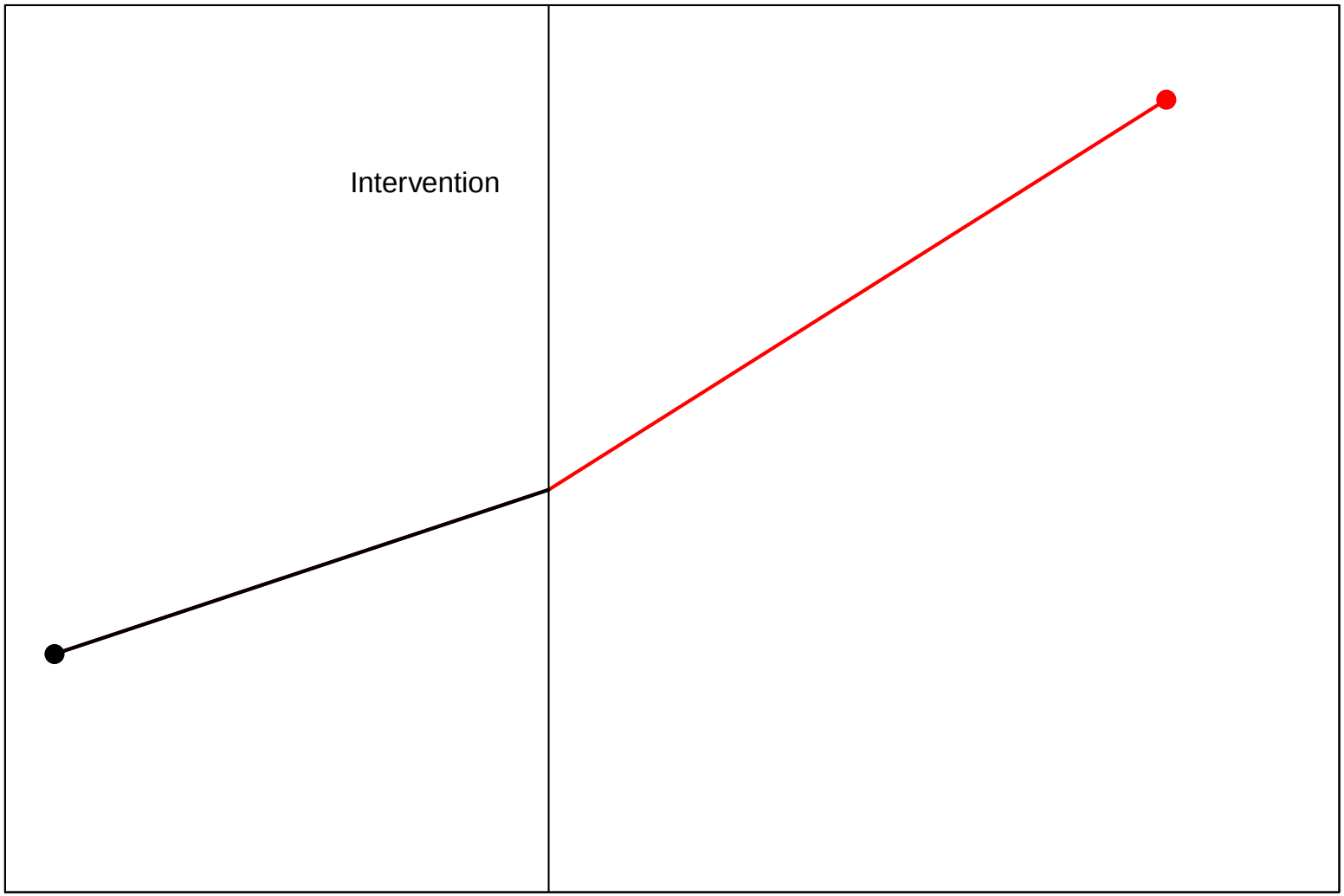
Placebo-controlled study

Randomized evaluations have been used in medicine for decades to test whether a new drug is beneficial for patients.

Before a drug is placed on the market, some patients are selected at random to receive it while others are given a placebo or no medical treatment.

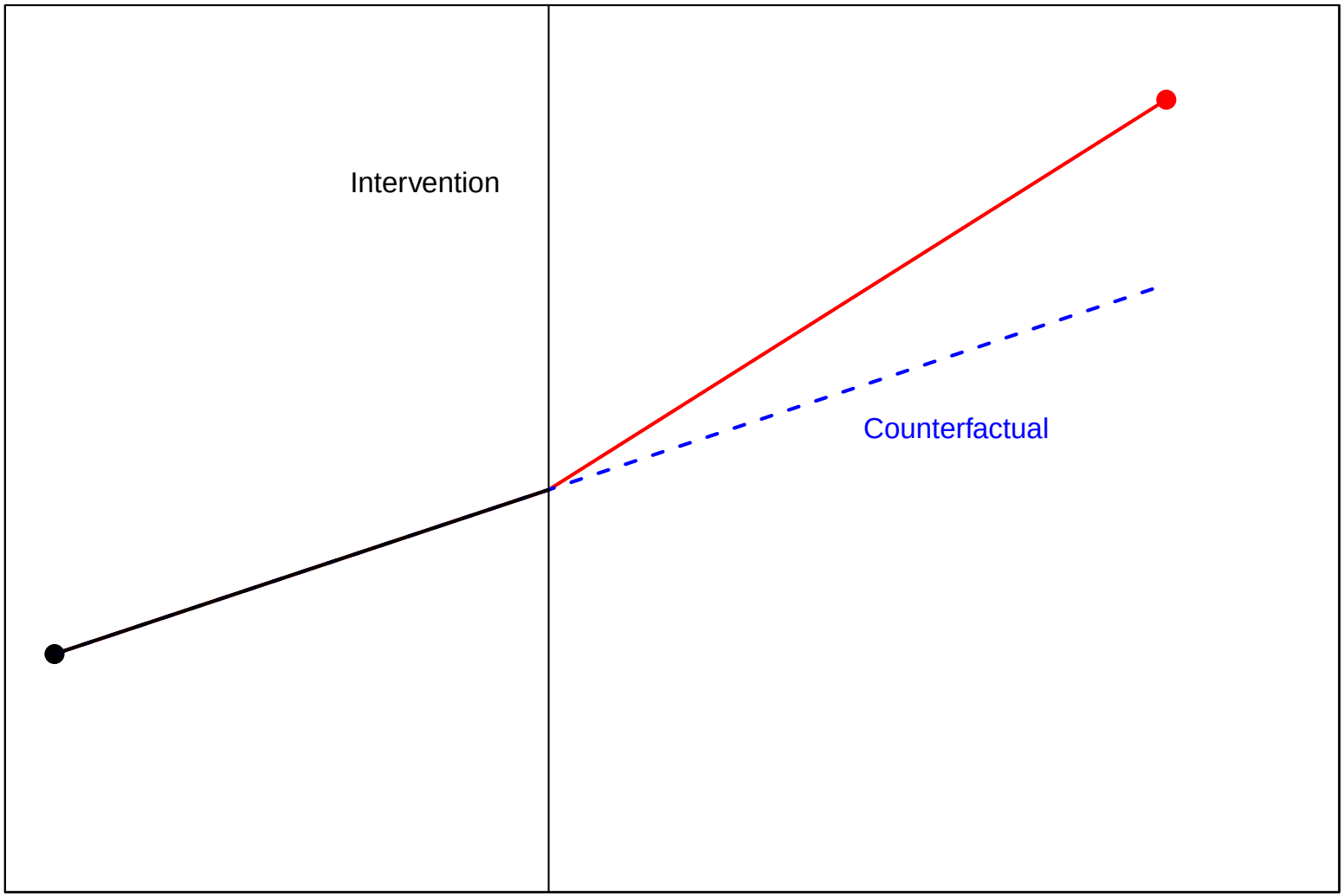
By comparing the two groups, scientists can prove objectively that any health benefits were caused by the drug itself.

Measuring impact of social programs



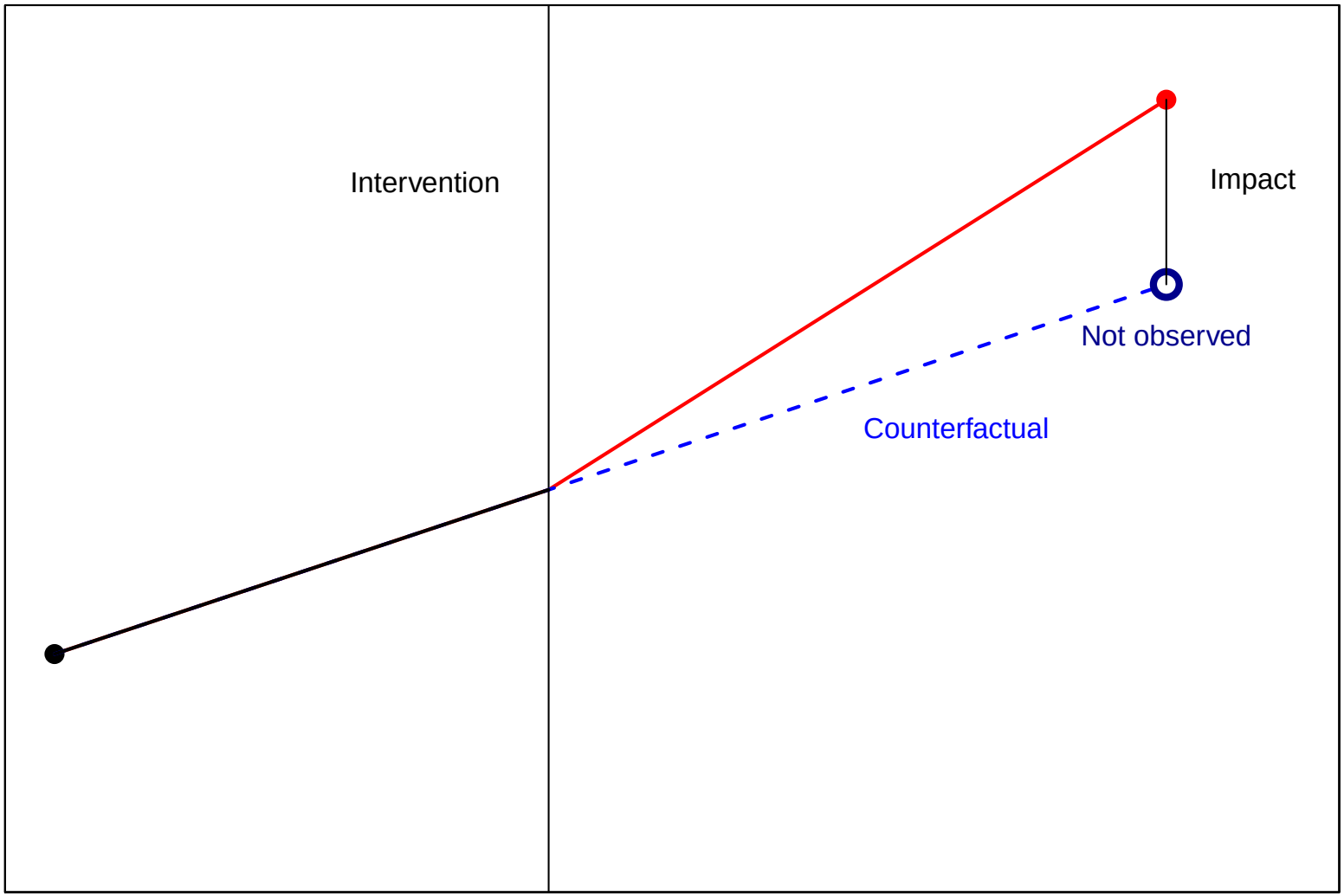
Measuring impact

What would have happen if there had been no program?



Measuring impact

What would have happen if there had been no program?



Measuring Impact

The effect of the treatment is not observable at an individual level.

- You observe treated individuals when treated
- You observe non-treated individuals when non-treated
- You DO NOT know what would have happened to the treated individual if he had not been treated
- You DO NOT know what would have happened to the non-treated individual if he had been treated

Individual cannot be cloned in reality.

Good example of this concept in the movie Sliding Doors (see trailer)

Measuring Impact

The impact of the program is defined as a comparison between:

- the outcome some time after the program
- the outcome at that same point in time had the program not been introduced (the **counterfactual**)

But how can you know what would have happen?

- You need to *construct or mimic* the counterfactual. The art of impact evaluation is to reconstruct the counterfactual correctly.
- Fundamental problem. Your impact evaluation is correct only if the estimation of the counterfactual is correct

Impact Evaluation Methods

1. Non- or Quasi-Experimental Methods
   - Pre-Post
   - Simple Difference
   - Difference-in-Difference
   - ...
2. Randomized Control Trials (RCTs)

An Example: Vote 2002 Campaign I

Problem: voter turnout has been declining since the 1960s in the U.S. In 2000 congressional and presidential elections, only 47% of eligible voters did vote.

Intervention : In the week preceding the 2002 congressional elections, Vote 2002 randomly placed phone calls to 60,000 voters and gave them the following message:

Hello, may I speak with Joe Iowa please? Hi. This is Marc Shotland calling from Vote 2002, a non-partisan effort working to encourage citizens to vote. We just wanted to remind you that elections are being held this Tuesday. The success of our democracy depends on whether we exercise our right to vote or not, so we hope you'll come out and vote this Tuesday. Can I count on you to vote next Tuesday?

An Example: Vote 2002 Campaign II

Impact? It should have changed certain outcomes (mainly voters turnout) for the ones who have participated.

1. Outcomes have changed
2. Observed change occurred among the participants of the program and did not occur among the non-participants
3. It is not something else that happened at the same time

Impact = Observed outcomes
- outcomes if there was no program

Impact Evaluation Methods

1. Non- or Quasi-Experimental Methods
   - Pre-Post
   - Simple Difference
   - Difference-in-Difference
   - ...
2. Randomized Control Trials (RCTs)

Impact Evaluation Methods

1. Non- or Quasi-Experimental Methods
   - Pre-Post
   - Simple Difference
   - Difference-in-Difference
   - ...
2. Randomized Control Trials (RCTs)

Method 1: Pre-Post (Before vs. After) I

Among the 60,000 people that were called, only 25,000 were reached.

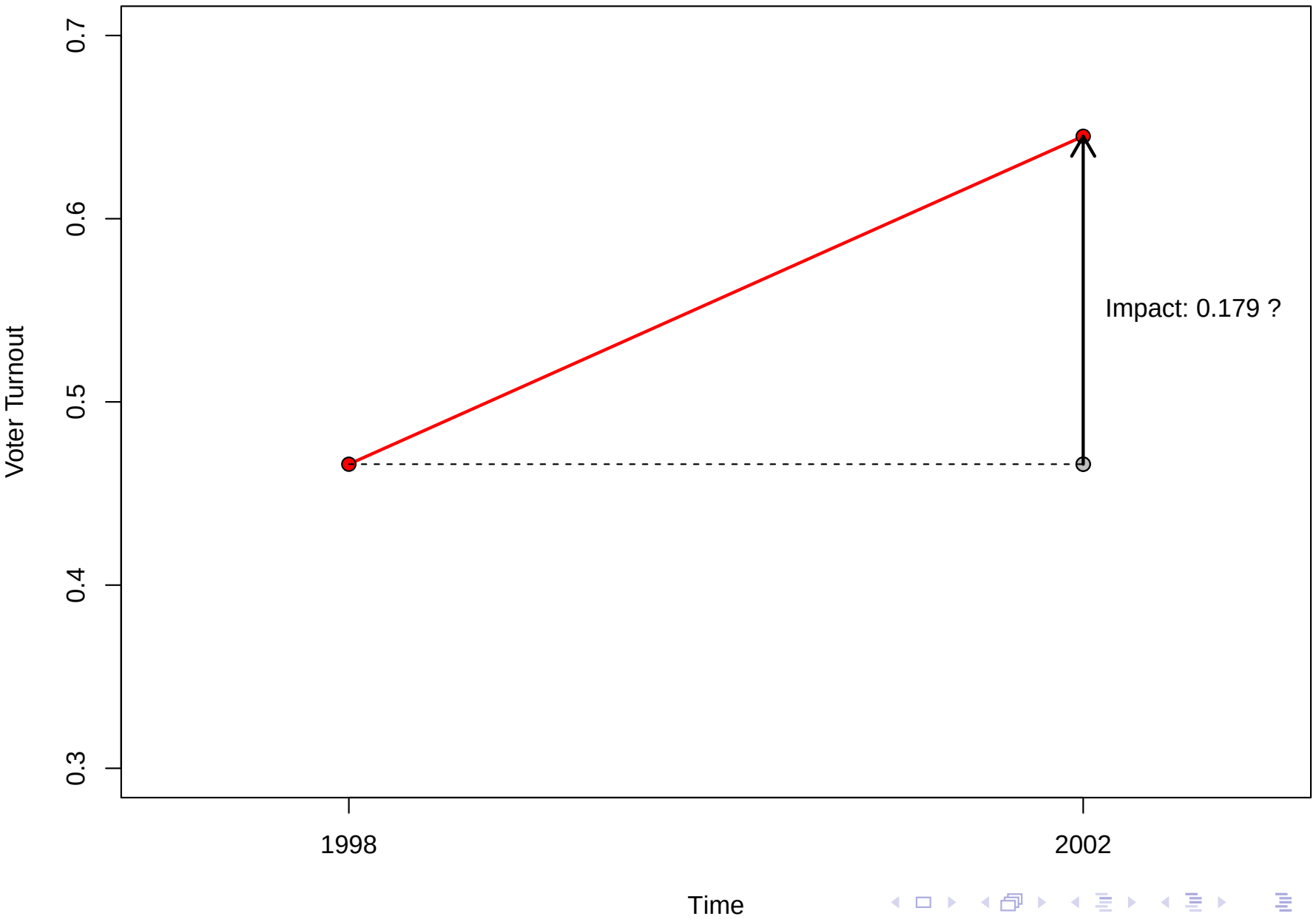
We also have data on voter turnout in 1998 elections. We can therefore look at the voter turnout before ("pre") and after ("post") the campaign for the people that were reached

	Voter turnout
Post-voter turnout (2002) if reached	64.5%
Pre-voter turnout (1998) if reached	46.6%
Difference	17.9pp*

* indicates statistically significant at the 5% level
pp means percentage points

Under what conditions can this difference be interpreted as the impact of the program?

Method 1: Pre-Post (Before vs. After) II



Impact Evaluation Methods

1. Non- or Quasi-Experimental Methods
   - Pre-Post
   - Simple Difference
   - Difference-in-Difference
   - ...
2. Randomized Control Trials (RCTs)

Impact Evaluation Methods

1. Non- or Quasi-Experimental Methods
   - Pre-Post
   - Simple Difference
   - Difference-in-Difference
   - ...
2. Randomized Control Trials (RCTs)

Method 2: Simple Difference I

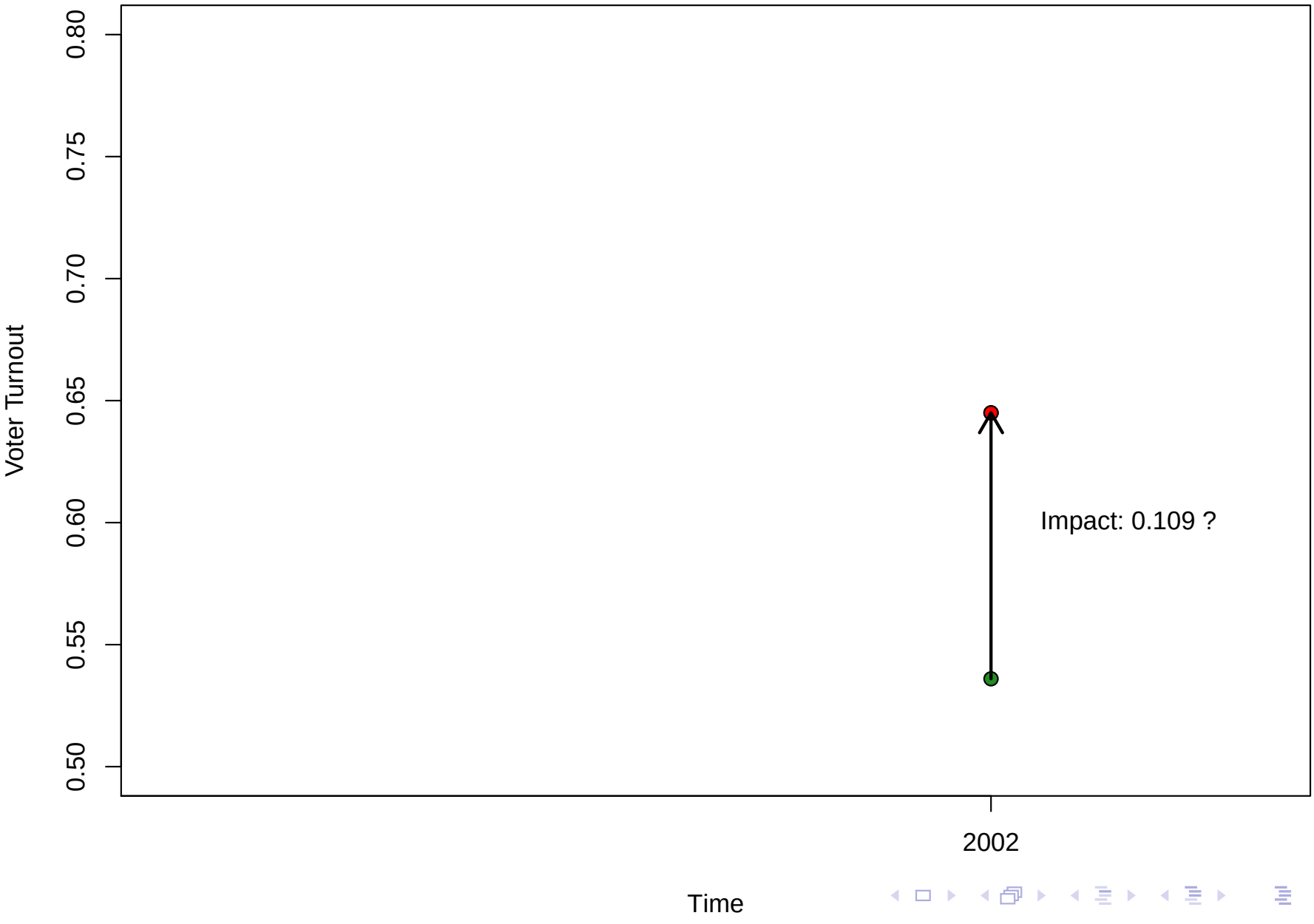
$\rightarrow$ The simple difference method will compare voter turnout between the reached (the participant group) and the not reached (the comparison group)

	Voter turnout
Voter turnout (2002) if reached	64.5%
Voter turnout (2002) if not reached	53.6%
Difference	10.9pp*

* indicates statistically significant at the 5% level

Under what conditions can this difference be interpreted as the impact of the program? Is the comparison group a good counterfactual?

Method 2: Simple Difference II



Method 2: Simple Difference III

Table : Characteristics of Reached and Not Reached groups

	Reached	Not reached	Difference
Voted in 2000	71.7%	63.3%	8.3pp*
Voted in 1998	46.6%	37.6%	9pp*
Household Size	1.56	1.50	0.06*
Average age	55.8	51.0	4.8*
Female	56.2%	53.8%	2.4pp*
Newly registered	7.3%	9.6%	-2.3pp*

* indicates statistically significant at the 5% level

Method 2: Simple Difference IV

Non-beneficiaries can be different than beneficiaries. Why?

- Programs often target beneficiaries according certain criteria (poverty level, demand for the service, etc.)
- People can choose to participate or not in the program (it may be that those who participated were more motivated?)

$\rightarrow$ Exposure to treatment may be correlated with unobserved characteristics related to the outcome studied

If beneficiaries are different than non-beneficiaries, they cannot represent a good counterfactual. The comparison will be biased because there is a *selection* of the beneficiaries. This is called **selection bias**

Vote 2002 Campaign

In the get-out-the-vote example, who do you think is more likely to be self-selected ?

Impact Evaluation Methods

1. Non- or Quasi-Experimental Methods
   - Pre-Post
   - Simple Difference
   - Difference-in-Difference
   - ...
2. Randomized Control Trials (RCTs)

Impact Evaluation Methods

1. Non- or Quasi-Experimental Methods
   - Pre-Post
   - Simple Difference
   - Difference-in-Difference
   - ...
2. Randomized Control Trials (RCTs)

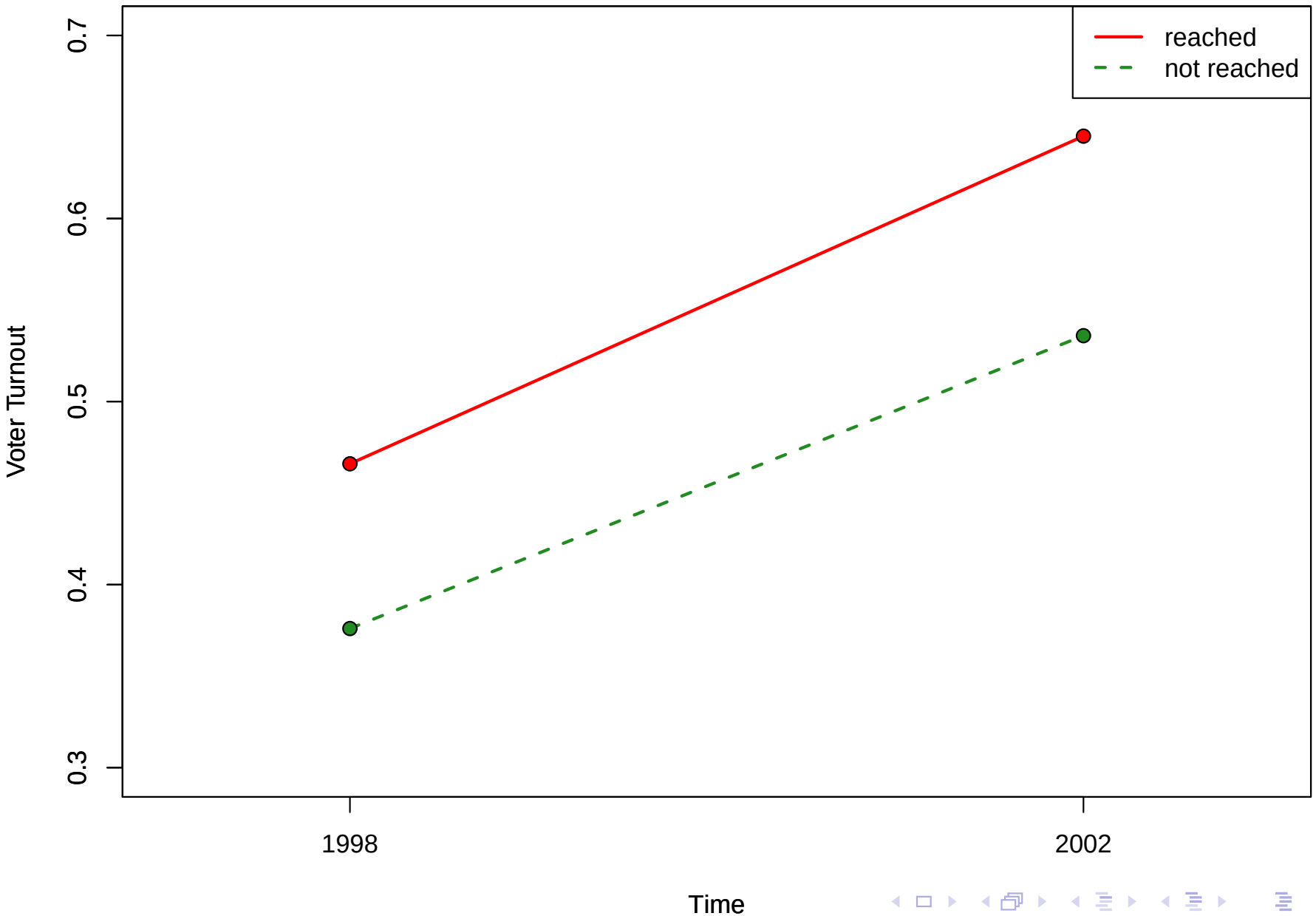
Method 3: Difference-in-Difference I

For the 60,000 people that were phoned, we know whether they voted in the 1998 elections (before the campaign). We can then exploit the panel dimension of the data to account for past voting behaviour in the two groups

	Voter turnout by group		
	Pre voter turnout	Post voter turnout	
Reached	46.6%	64.5%	17.9pp*
Not reached	37.6%	53.6%	16pp*
Diff in Diff	9pp*	10.9pp*	1.9pp*

* indicates statistically significant at the 5% level

Method 3: Difference-in-Difference II



Method 3: Difference-in-Difference III

Formally,

Let T denote the treated group and C denote the control group (the mimic of the counterfactual), and Y be the outcome. We consider two periods of time, before the intervention occurs: t = 0 and after it: t= 1

The expected impact is given by:

$$E(\text{Impact}) = \underbrace{\left[E(Y^T_{t=1}) - E(Y^T_{t=0})\right]}_{\text{difference in outcome among the treated before and after the program}} - \underbrace{\left[E(Y^C_{t=1}) - E(Y^C_{t=0})\right]}_{\text{difference in outcome among the controls before and after the program (accounts for time)}}$$

Method 3: Difference-in-Difference IV

You can also rewrite and have,

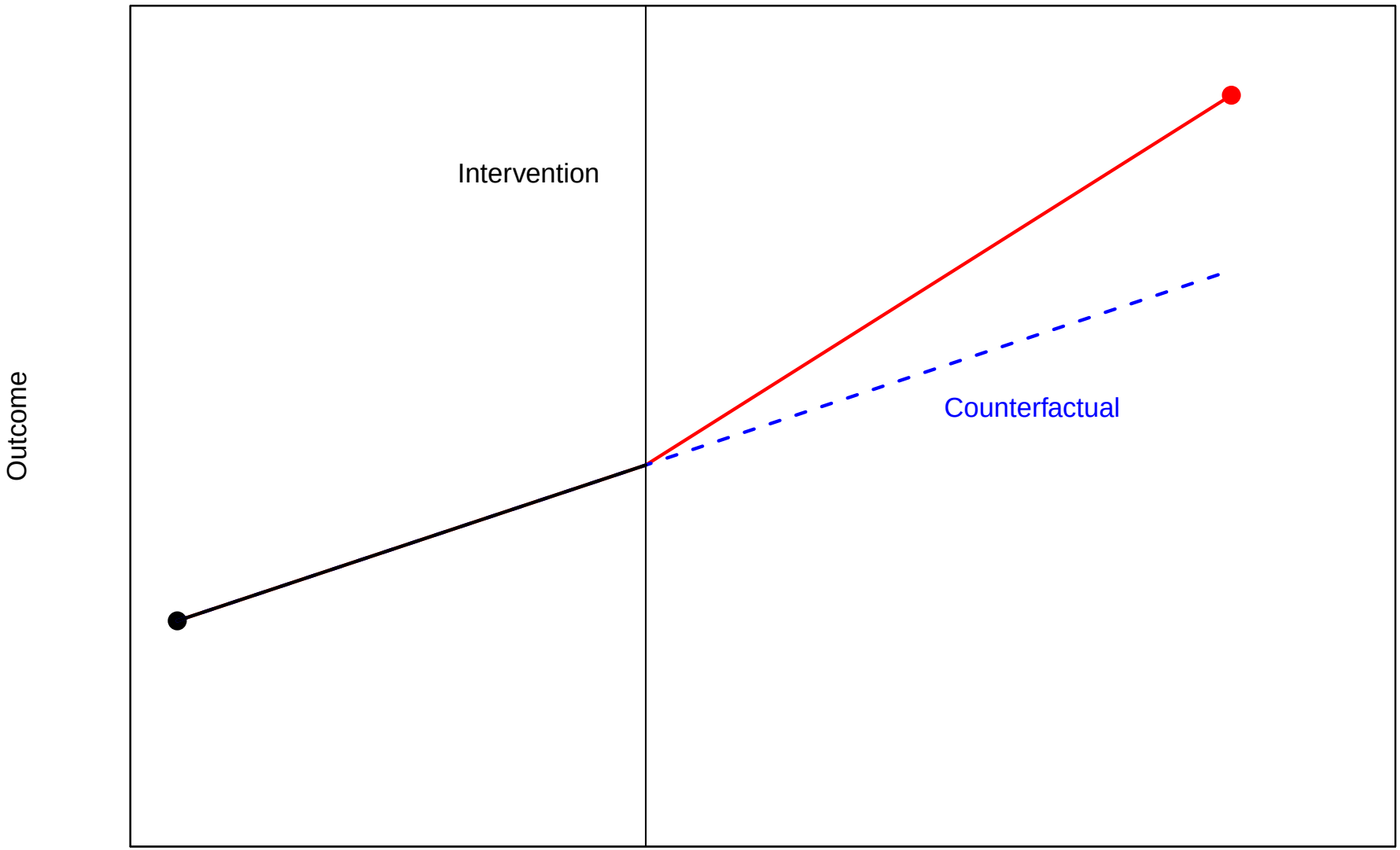
$$E(\text{Impact}) = \underbrace{\left[E(Y^T_{t=1}) - E(Y^C_{t=1})\right]}_{\text{difference in outcome among the treated and the control after the program}} - \underbrace{\left[E(Y^T_{t=0}) - E(Y^C_{t=0})\right]}_{\text{difference in outcome among the treated and the control before the treatment (accounts for intrinsec differences)}}$$

Method 3: Difference-in-Difference V

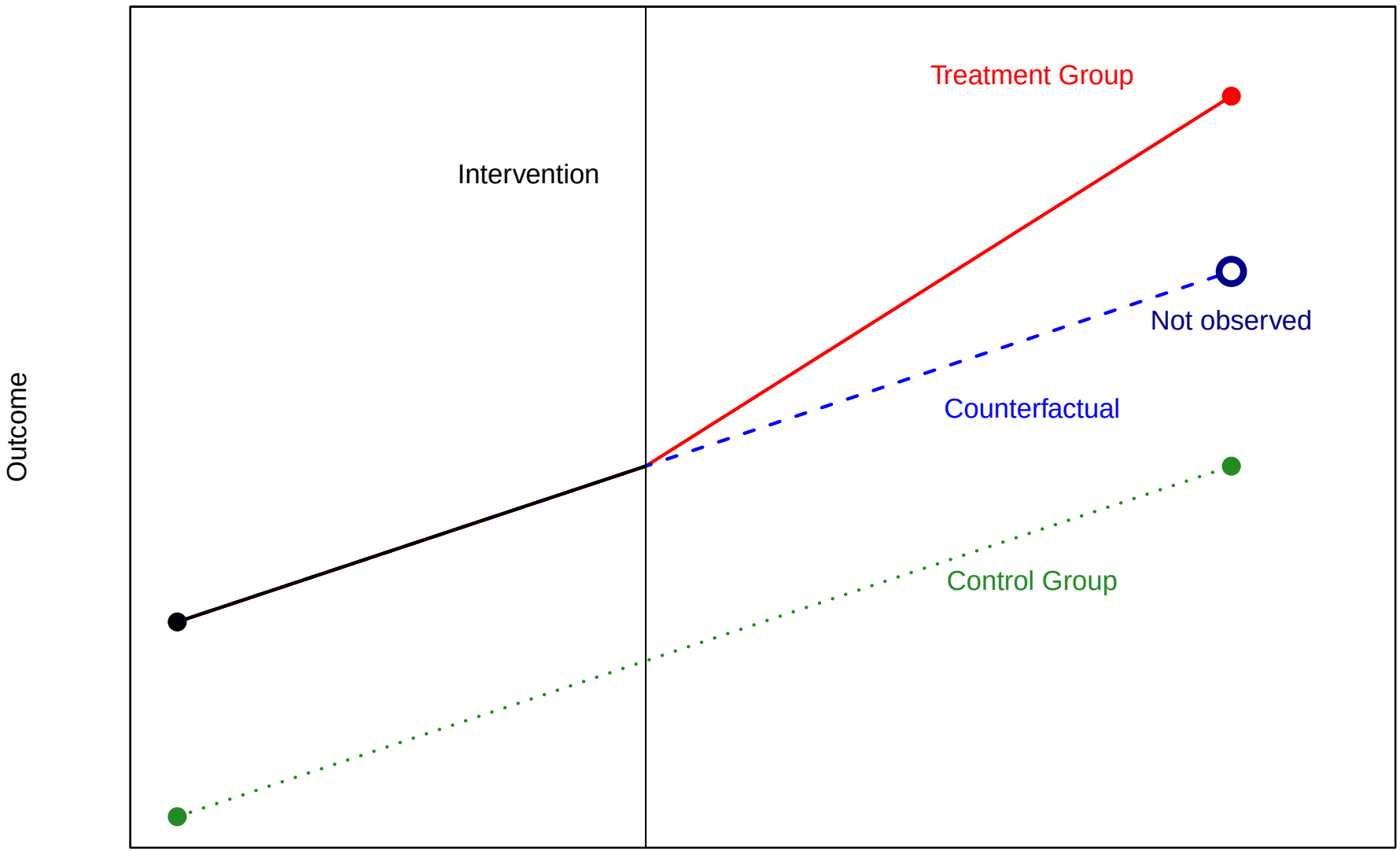
From the example,

$$E(\text{Impact}) = \underbrace{[64.5 - 53.6]}_{\substack{\text{difference in outcome among the treated and the control after the program}}} - \underbrace{[46.6 - 37.6]}_{\substack{\text{difference in outcome among the treated and the control before the treatment (accounts for intrinsic differences)}}}$$

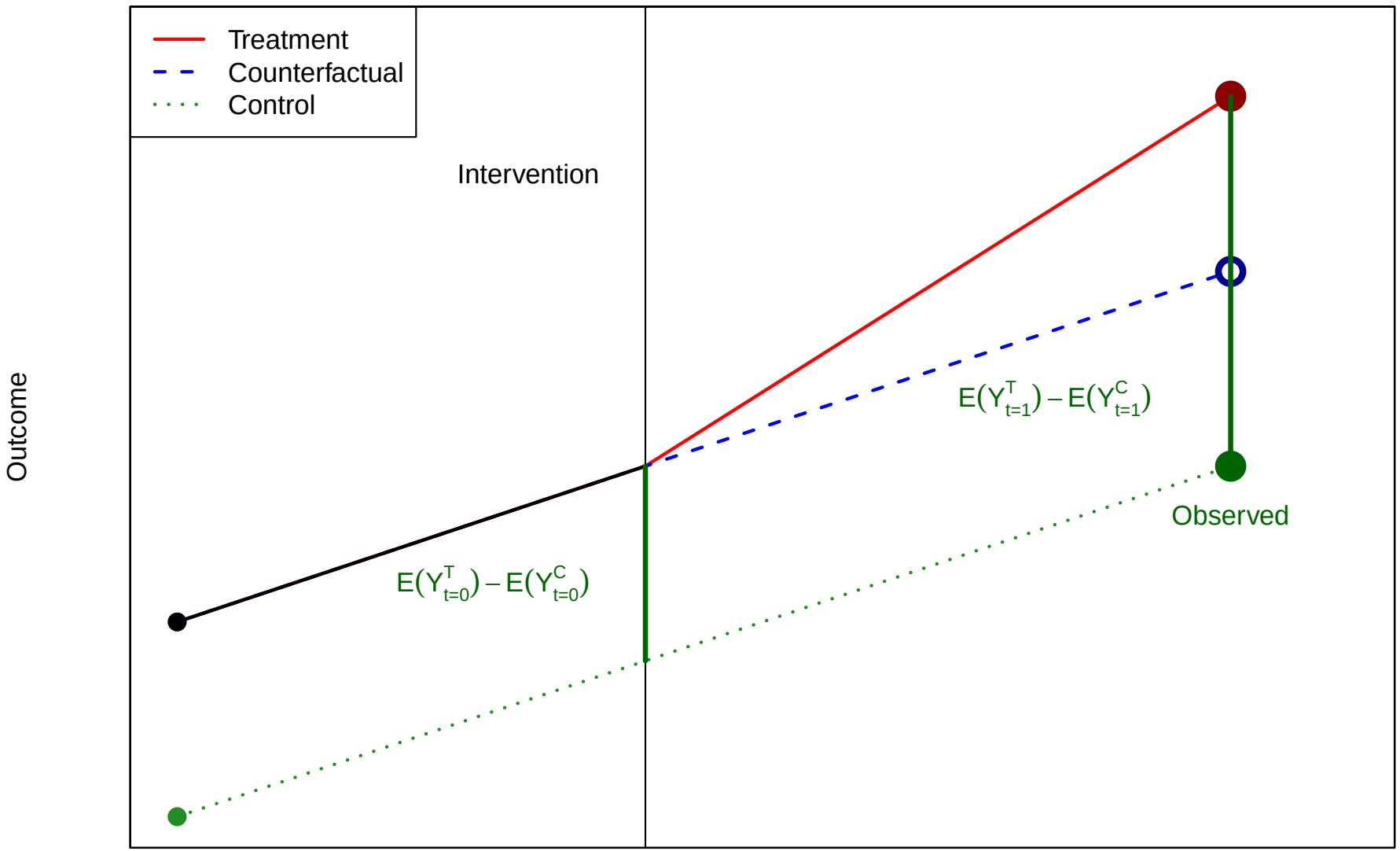
Method 3: Difference-in-Difference VI



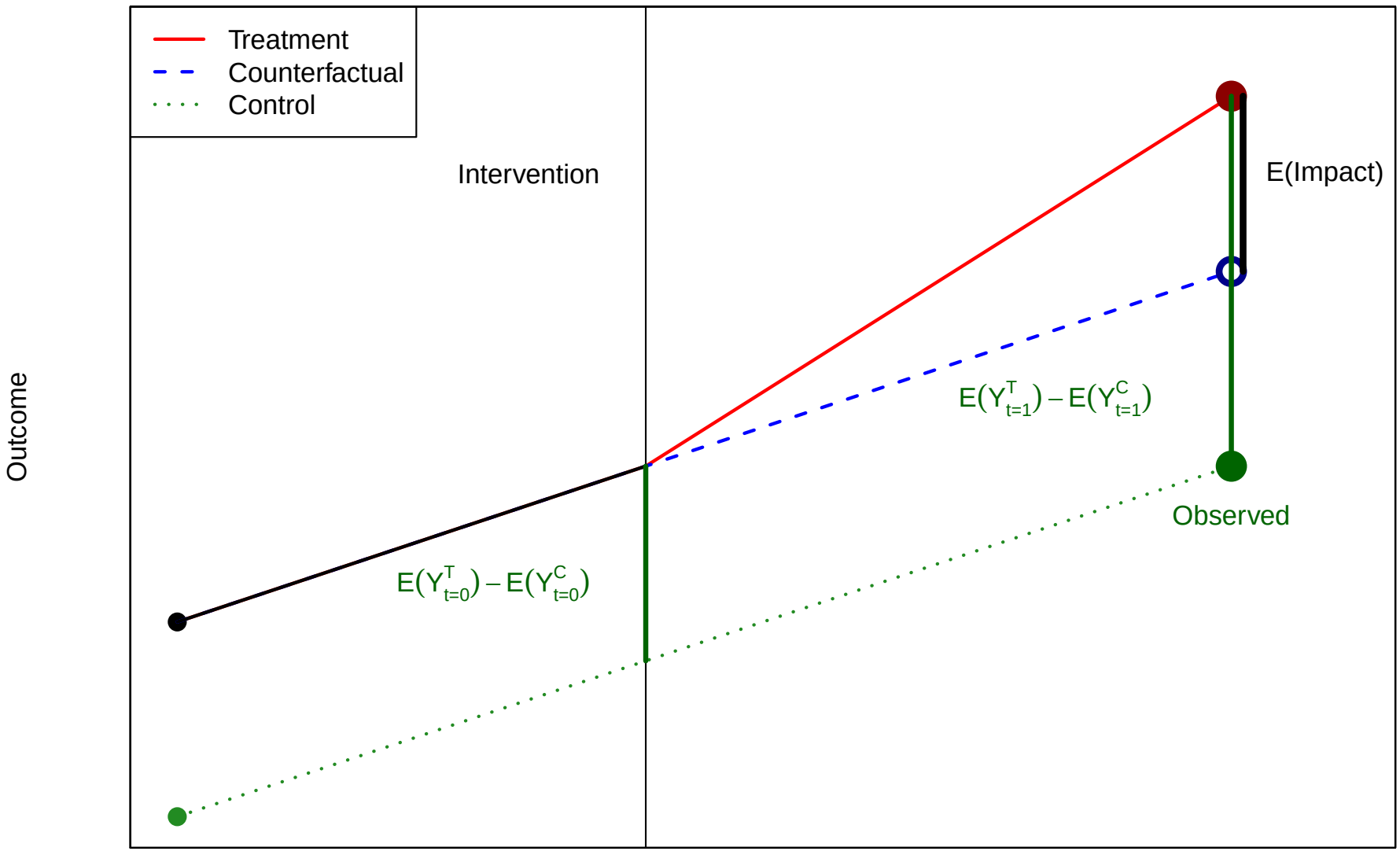
Method 3: Difference-in-Difference VII



Method 3: Difference-in-Difference VIII



Method 3: Difference-in-Difference IX



Method 3: Difference-in-Difference X

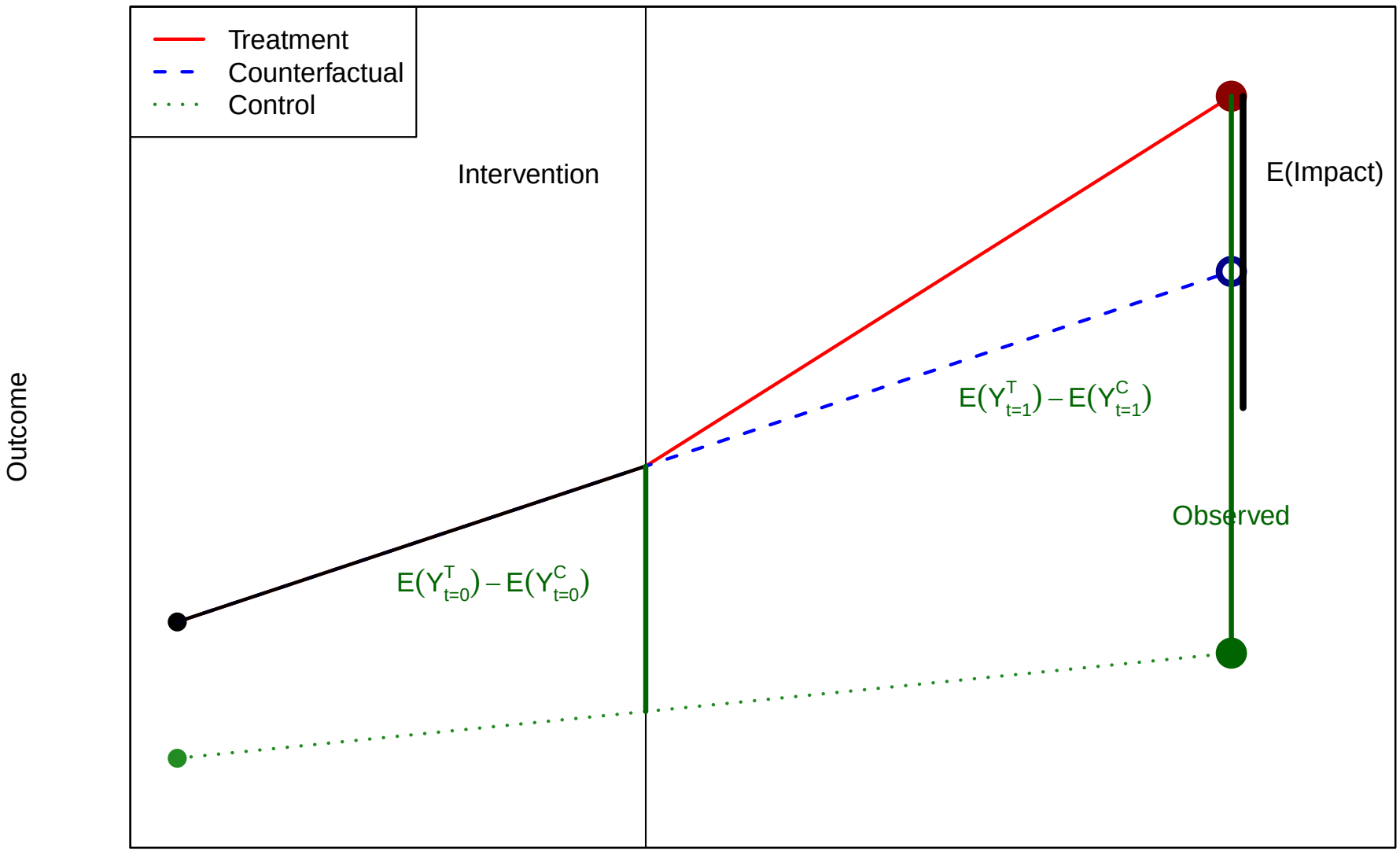
Important Remark: Under what conditions the diff-in-diff method provides a good impact estimate ?

Method 3: Difference-in-Difference XI

Important Remark: Under what conditions the diff-in-diff method provides a good impact estimate ?

- You assume that the treatment group would have behave the same without the program
- You assume that the control group and the counterfactual evolve along the same trend

Method 3: Difference-in-Difference XII



Impact Evaluation Methods

1. Non- or Quasi-Experimental Methods
   - Pre-Post
   - Simple Difference
   - Difference-in-Difference
   - ...
2. Randomized Control Trials (RCTs)

Impact Evaluation Methods

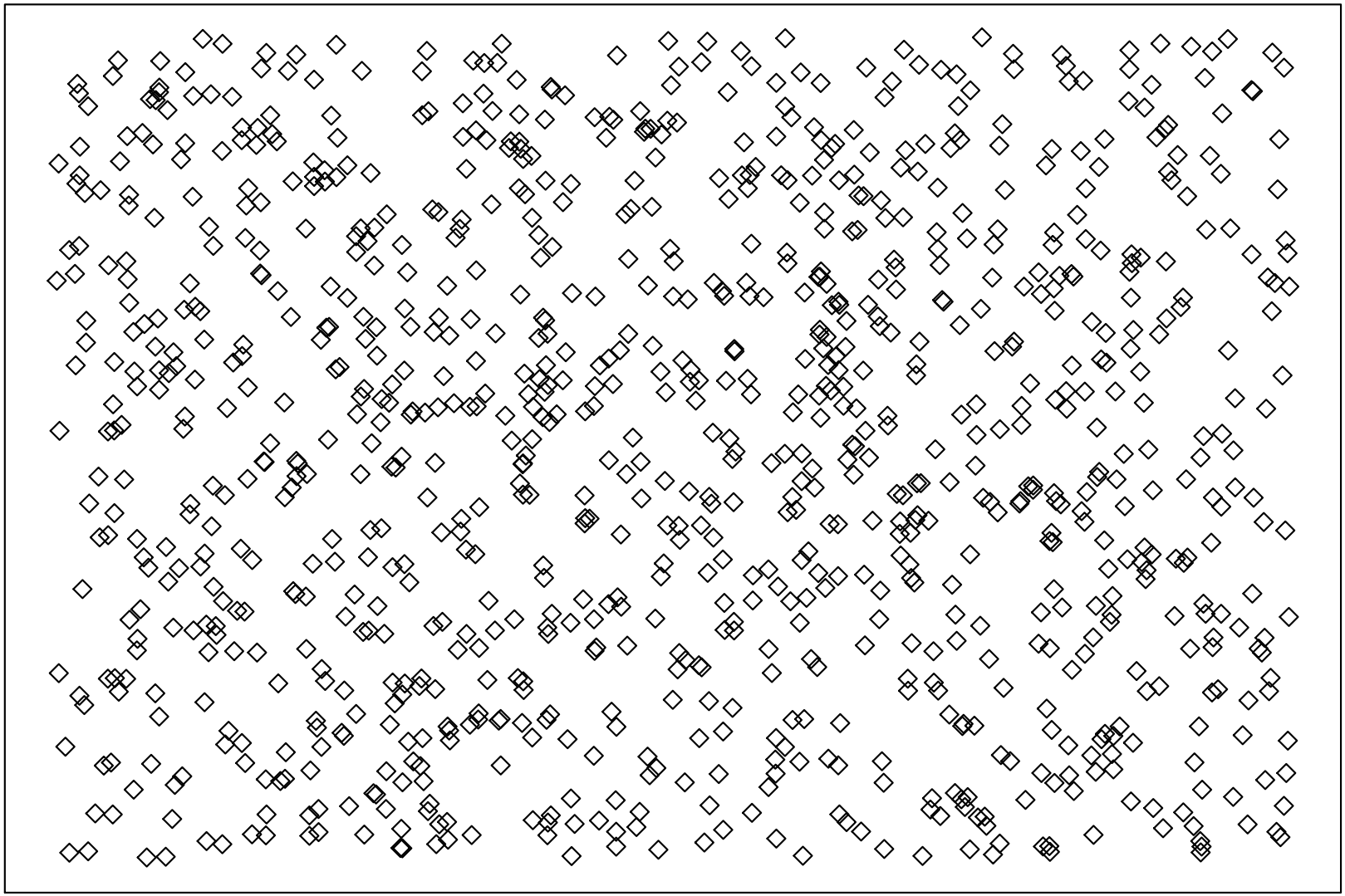
1. Non- or Quasi-Experimental Methods
   - Pre-Post
   - Simple Difference
   - Difference-in-Difference
   - ...
2. Randomized Control Trials (RCTs)

Method 4: Randomized Control Trials I

Back to our example,

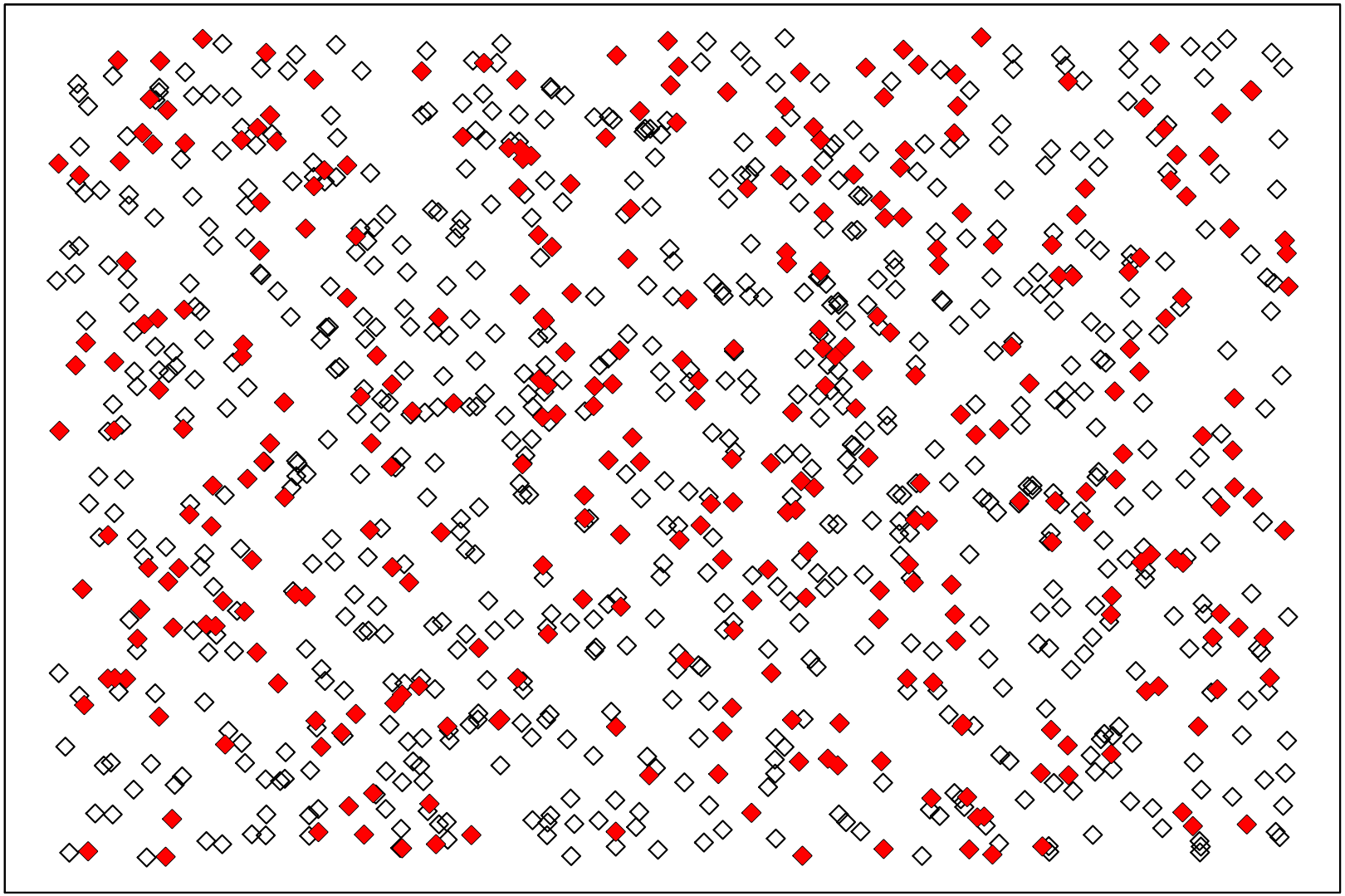
The 60,000 people that were called were in fact **randomly** selected from a larger population of 2 million potential voters.

Method 4: Randomized Control Trials II



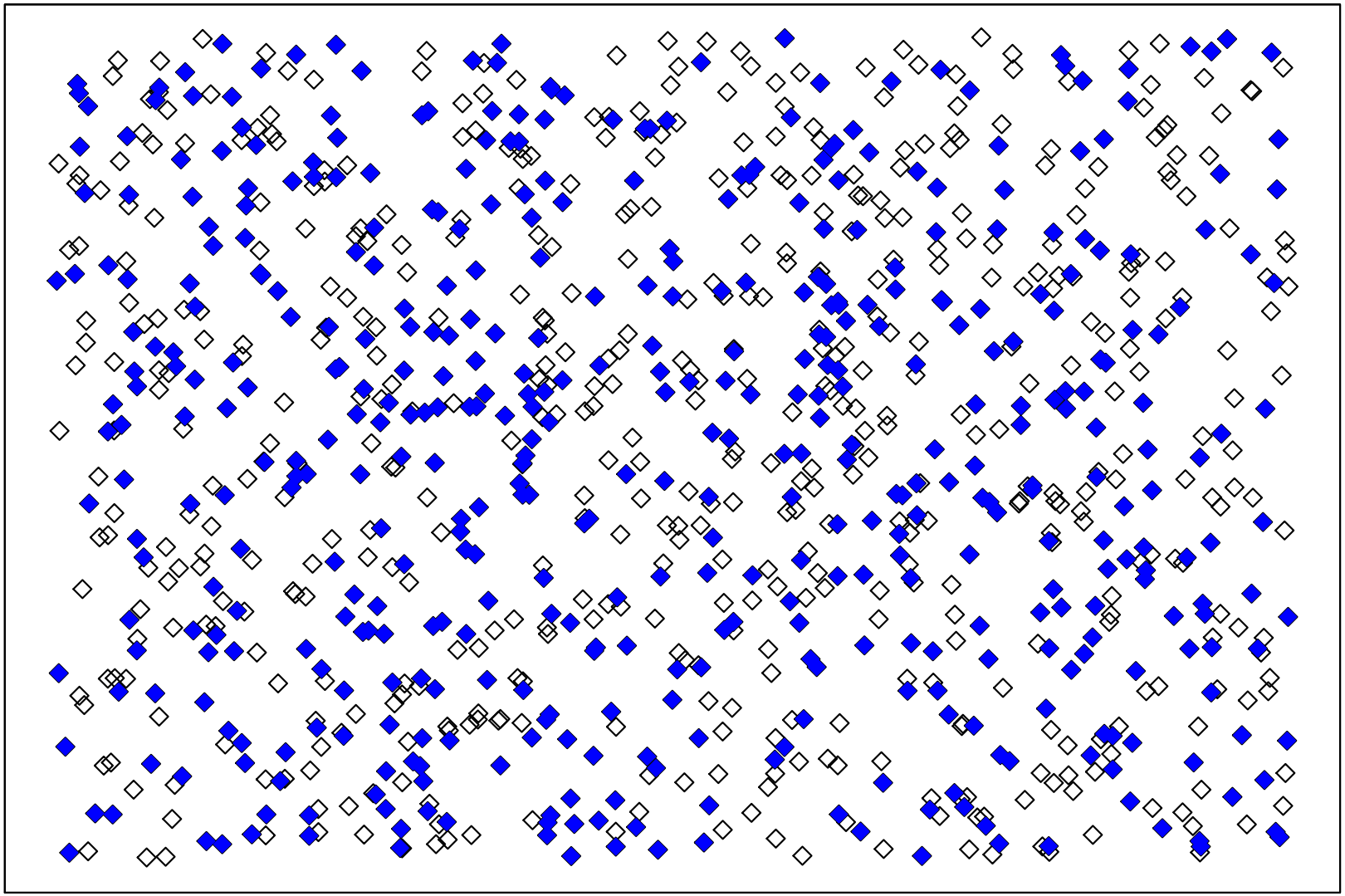
A population of 1000 individuals

Method 4: Randomized Control Trials III



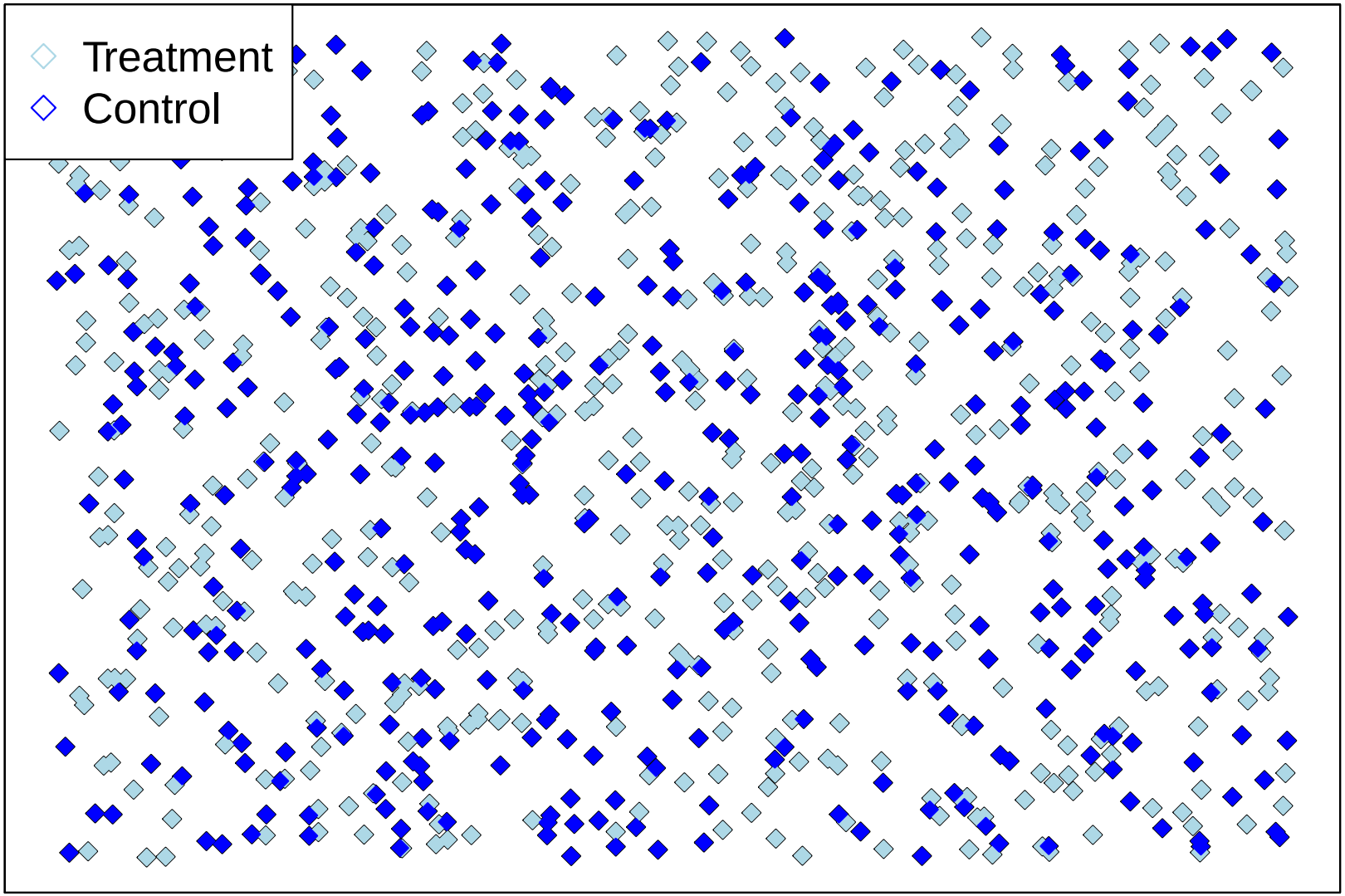
A population of 1000 individuals of which I randomly select a sample of 320

Method 4: Randomized Control Trials IV



A population of 1000 individuals of which I randomly select another sample of 473

Method 4: Randomized Control Trials V



I randomly assign 473 people to the treatment group, the 527 other constitues the control group

Method 4: Randomized Control Trials VI

Back to our example,

The 60,000 people that were called were in fact **randomly** selected from a larger population of 2 million potential voters.

- The treatment group is now the 60,000 people that were called (whether reached or not)
- and the control group is constituted by the rest.

From the randomization, we know that the 60,000 people is statistically identical to the 2 million individuals in both **observable and non-observable** characteristics, except for the fact that the 60,000 people were called.

Method 4: Randomized Control Trials VII

- Gold Standard
- $E(Y^T_{t=0}) - E(Y^C_{t=0}) = 0$, control and treatment group are similar at baseline (before the program takes place) in both observables and non-observables characteristics
- Why? Because of randomization

By construction,

$$E(\text{Impact}) = E(Y^T_{t=1}) - E(Y^C_{t=1})$$

Method 4: Randomized Control Trials VIII

Characteristics of called (T) and Not Called (C) groups before the program:

	Treatment	Control	Difference
Voted in 2000	56.7%	56.4%	0.4pp
Voted in 1998	22.7%	23.1%	-0.5pp
Household Size	1.50	1.50	0
Average age	52.0	52.2	-0.2
Female	54.6%	55.2%	-0.6pp
Newly registered	11.6%	11.7%	0pp

* indicates statistically significant at the 5% level

The two groups look very similar, as we expected.

Method 4: Randomized Control Trials IX

	Voter turnout by group		Impact estimate
	Treatment (60,000 called)	Control (2M not called)	
Simple Difference	58.2%	58%	0.2pp

BUT, only 25,000 were reached. We need to adjust for this (instrumental variable strategy) to obtain the **treatment effect on the treated**

0.4pp

Method 4: Randomized Control Trials X

Table : Summary of results for each method

Method	Impact estimate
Pre-Post	17.9pp*
Simple diff	10.8pp*
Diff-in-Diff	2.5pp*
Randomized experiment	0.4pp

In conclusion, the method is crucial in estimating the program impact. While the non-randomized method conclude to a positive significant impact of the program, the randomized experiment suggests that the campaign had no impact on the voter turnout.

Why? Because there is a **selection bias** (in this case, selection of those who picked up the phone) associated to the non-randomized methods.

Method 4: Randomized Control Trials XI

Why do Randomized Control Trials work?

- RCTs can measure the causal impact of a program.
- RCTs solve the problem of selection bias: members of the groups (treatment and control) are statistically equivalent.
- Changes over time do not invalidate results, because they affect both groups
- RCTs provide a valid counterfactual (What would have happened without the program?)

Impact Evaluation Methods

1. Non- or Quasi-Experimental Methods
   - Pre-Post
   - Simple Difference
   - Difference-in-Difference
   - ...
2. Randomized Control Trials (RCTs)
   - Unit of randomization
   - Multiple treatments
   - Spillovers

Impact Evaluation Methods

1. Non- or Quasi-Experimental Methods
   - Pre-Post
   - Simple Difference
   - Difference-in-Difference
   - ...
2. Randomized Control Trials (RCTs)
   - Unit of randomization
   - Multiple treatments
   - Spillovers

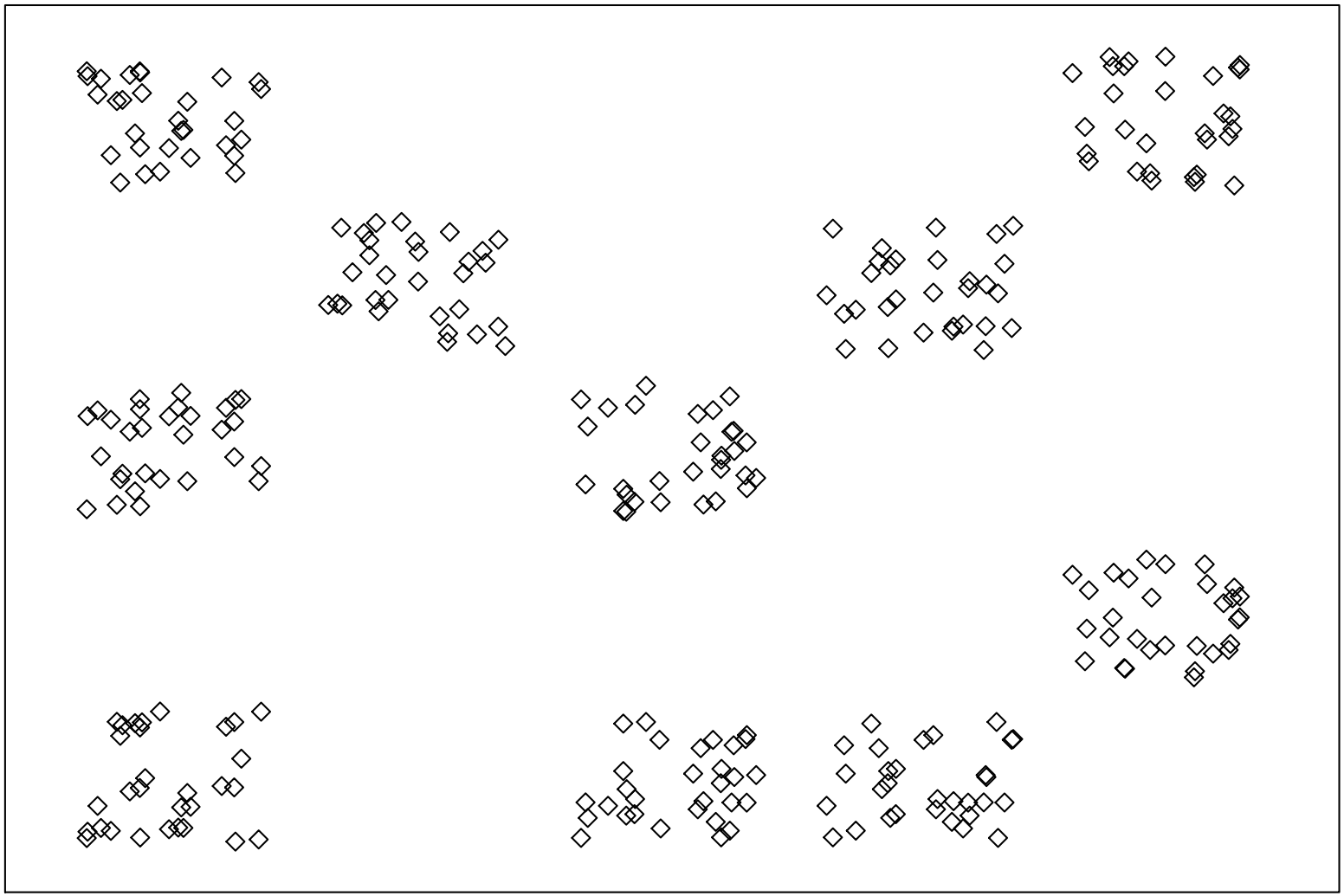
Unit of randomization I

At which level should we randomize?

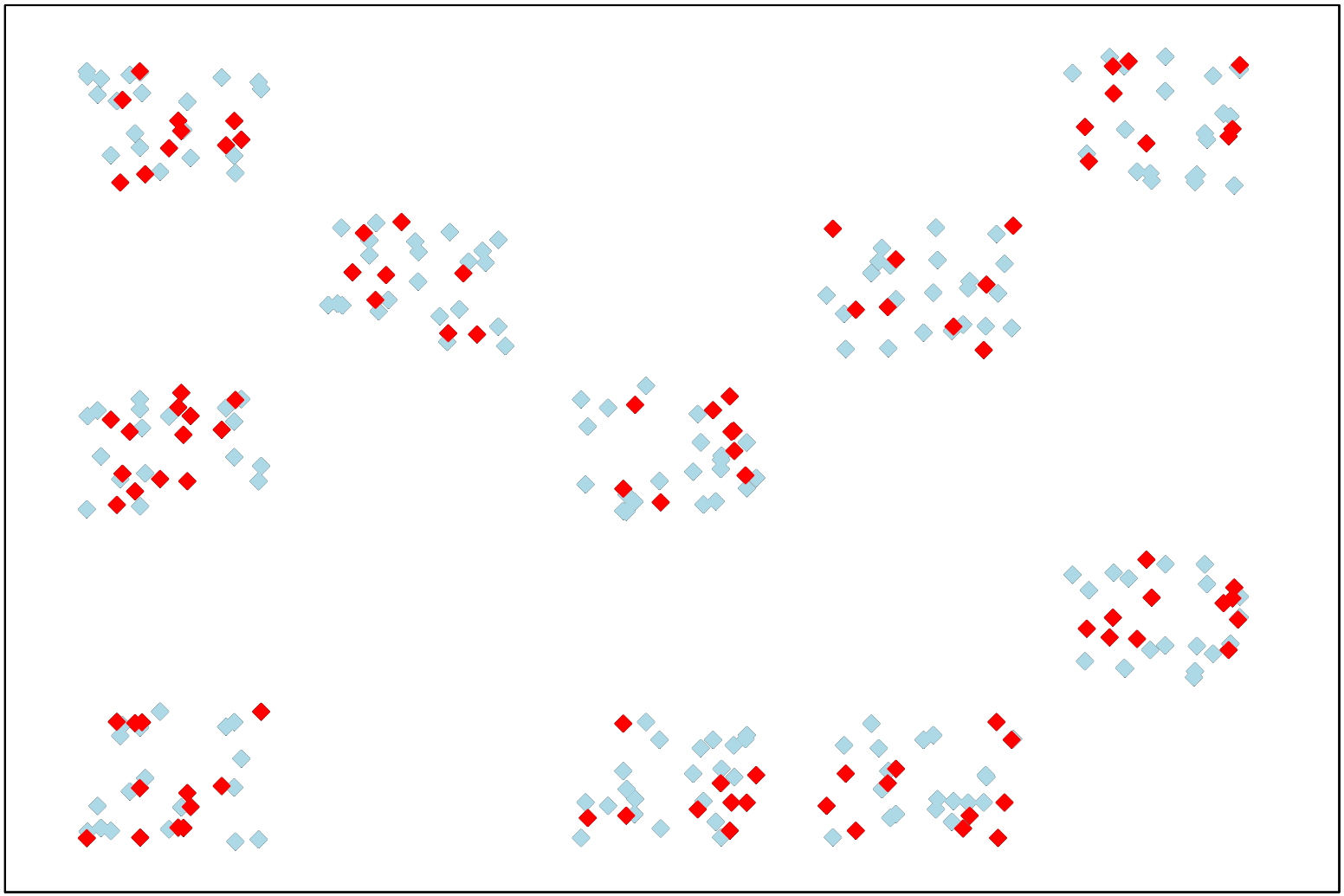
- **Individuals:** a child receives a scholarship
- **Households:** a family receives discounted water disinfectant
- **Schools:** all children in a school receive deworming treatment
- **Clinics:** a community clinic gives pregnant women free bed nets
- **Villages:** the community well in a village is improved

Unit of randomization II

Imagine a district of 10 villages...

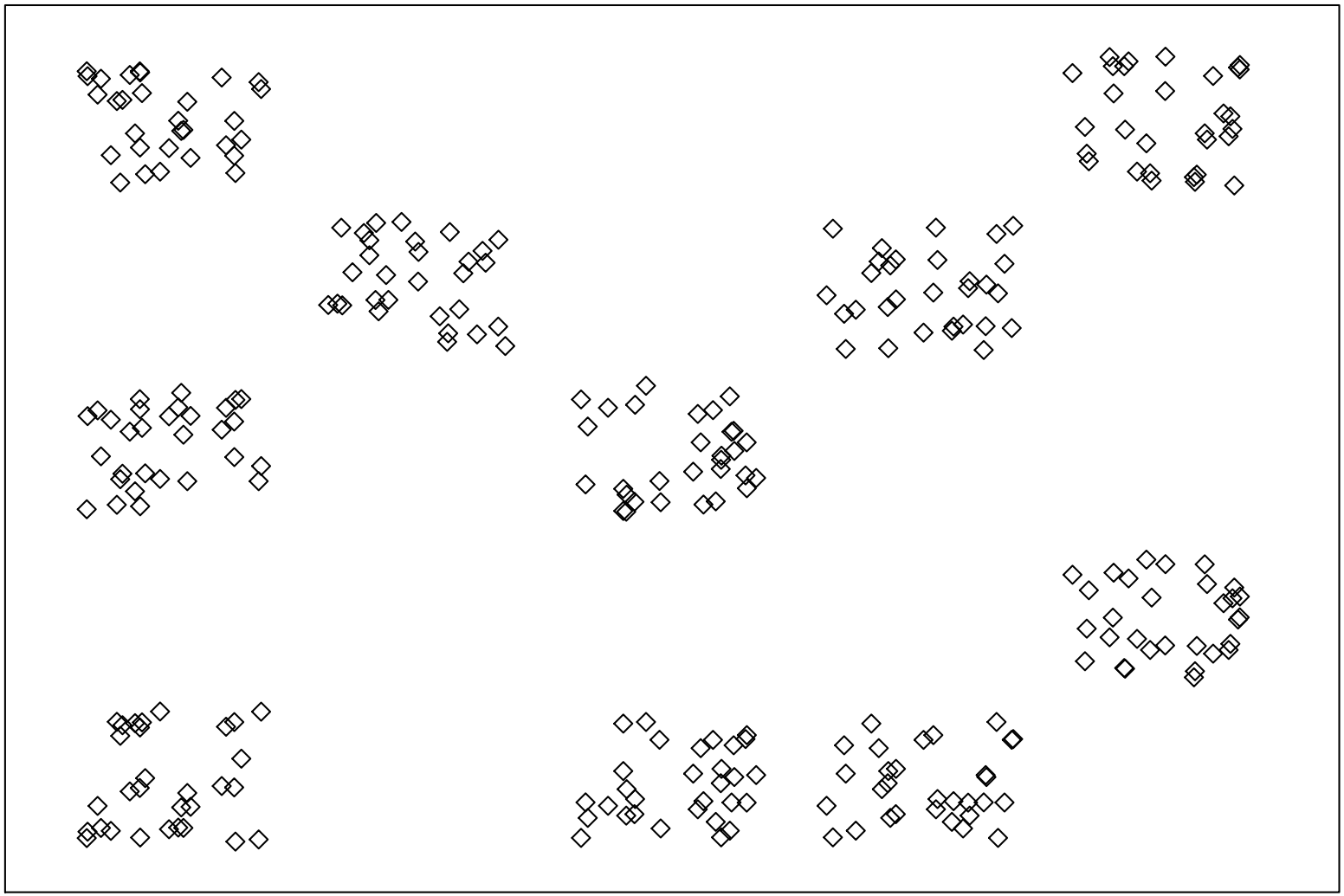


Unit of randomization III

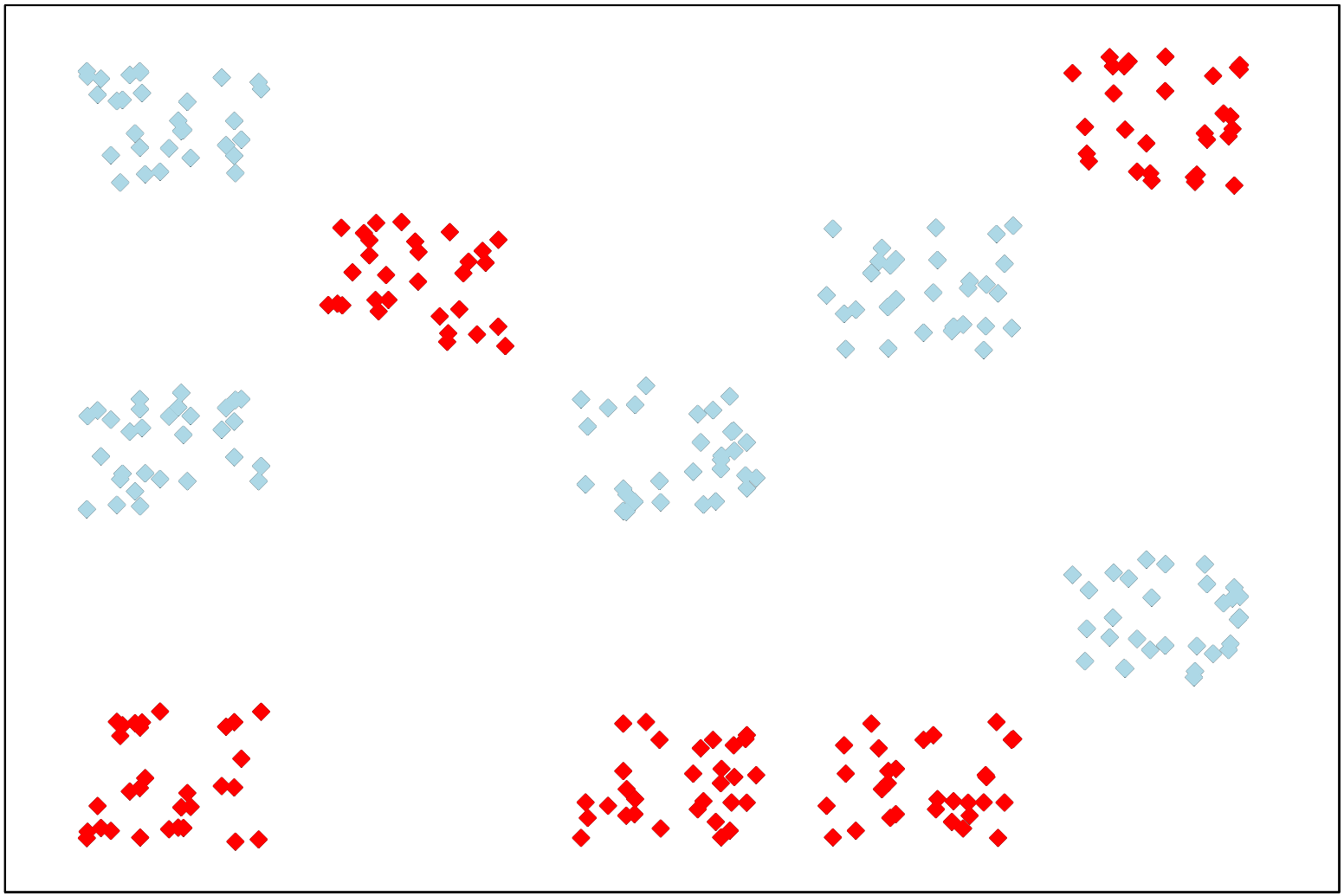


Unit of randomization IV

Imagine a district of 10 villages...



Unit of randomization V



Impact Evaluation Methods

1. Non- or Quasi-Experimental Methods
   - Pre-Post
   - Simple Difference
   - Difference-in-Difference
   - ...
2. Randomized Control Trials (RCTs)
   - Unit of randomization
   - Multiple treatments
   - Spillovers

Impact Evaluation Methods

1. Non- or Quasi-Experimental Methods
   - Pre-Post
   - Simple Difference
   - Difference-in-Difference
   - ...
2. Randomized Control Trials (RCTs)
   - Unit of randomization
   - Multiple treatments
   - Spillovers

Multiple treatments I

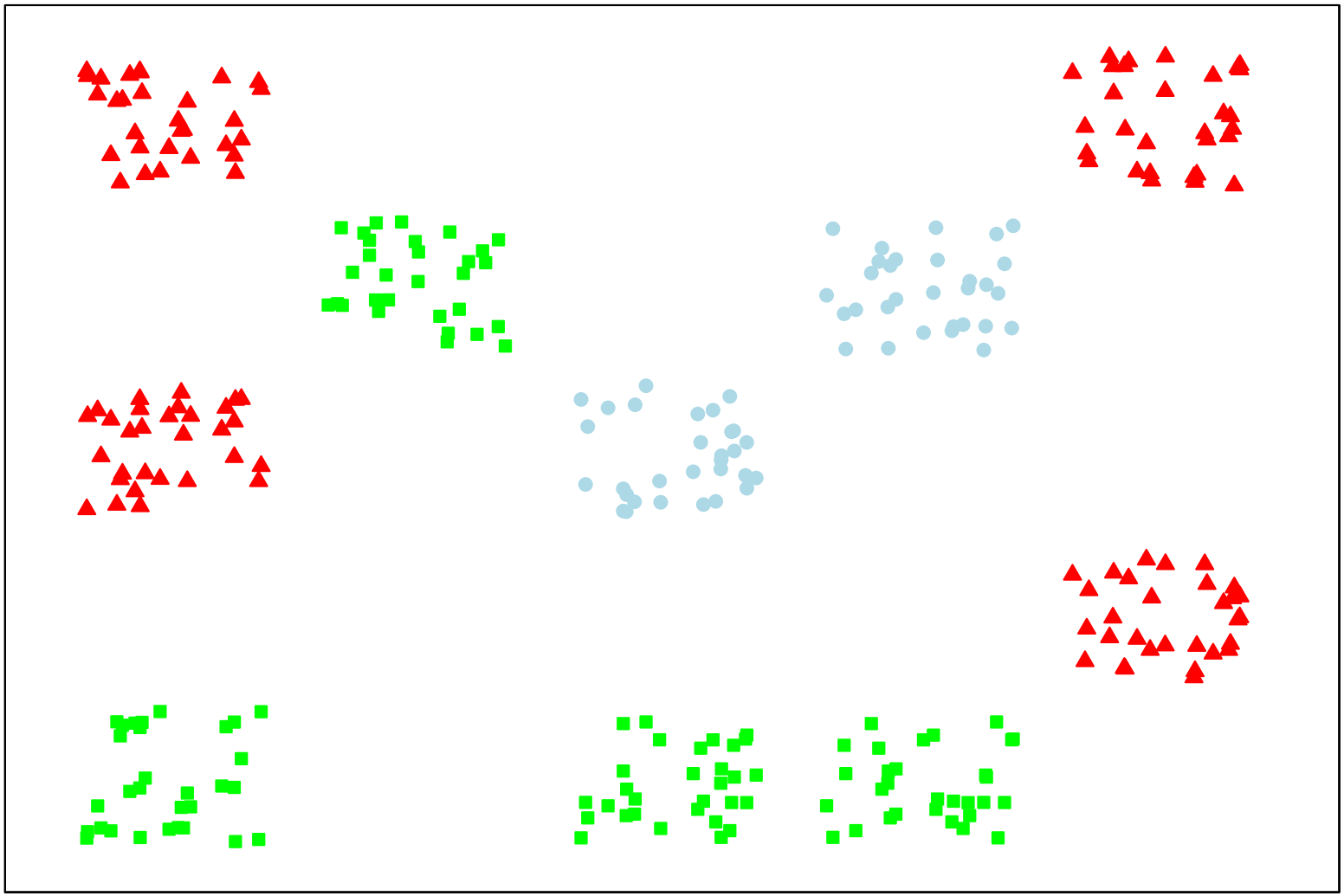
- Sometimes core question is deciding among different possible interventions, or different degree of treatment
- You can randomize these programs

Randomized Evaluation in Micro-insurance

From 300 villages,

- random selection of 120 villages will receive a "soft" insurance package, with low premium and coverage,
- random selection 120 villages will receive a "complete" insurance package, with high premium and benefits,
- The remaining 60 villages will serve as a control group, with no available formal insurance.

Multiple treatments II



Impact Evaluation Methods

1. Non- or Quasi-Experimental Methods
   - Pre-Post
   - Simple Difference
   - Difference-in-Difference
   - ...
2. Randomized Control Trials (RCTs)
   - Unit of randomization
   - Multiple treatments
   - Spillovers

Impact Evaluation Methods

1. Non- or Quasi-Experimental Methods
   - Pre-Post
   - Simple Difference
   - Difference-in-Difference
   - ...
2. Randomized Control Trials (RCTs)
   - Unit of randomization
   - Multiple treatments
   - Spillovers

Spillovers I

What if there are treatment externalities affecting the untreated?

- if treatment prevents contagious epidemics, we expect transmission to be prevented (e.g. deworming)
- if treatment is a prevention/information campaign, there can be imitation and/or communication with peers (e.g. bed nets)
- if treatment is a cash transfer, we may observe local economy effects (e.g. PROGRESA)

Spillovers II

Deworming

Suppose we randomize pupils within schools. We then give deworming drug to the pupils that are assigned to the treatment group.

At the end of the program, if there are **no externalities**, we observe that all children that received treatment don't have worms anymore

	Treated	Outcome
pupil 1	Yes	No worms
pupil 2	Yes	No worms
pupil 3	No	Worms
pupil 4	Yes	No worms
pupil 5	No	Worms
pupil 6	No	Worms

$$\text{Impact} = E(Worms^{T}_{t=1}) - E(Worms^{C}_{t=1})$$

Impact = 0% - 100% = -100 %

Spillovers III

At the end of the program, if there are **externalities**, prevalence is lower and we observe that some children are not reinfected with worms, even though they did not receive treatment

	Treated	Outcome
pupil 1	Yes	No worms
pupil 2	Yes	No worms
pupil 3	No	Worms
pupil 4	Yes	No worms
pupil 5	No	Worms
pupil 6	No	No Worms

$$\text{Impact} = E(Worms^{T}_{t=1}) - E(Worms^{C}_{t=1})$$

$$\text{Impact} = 0\% - 67\% = -67\ \%$$

More children that are dewormed, but the estimated impact is lower.

$\rightarrow$ Spillovers lead to an underestimation of the program

For more information, see Miguel, E., & Kremer, M. (2004). Worms: identifying impacts on education and health in the presence of treatment externalities. Econometrica, 72(1), 159-217

Spillovers IV

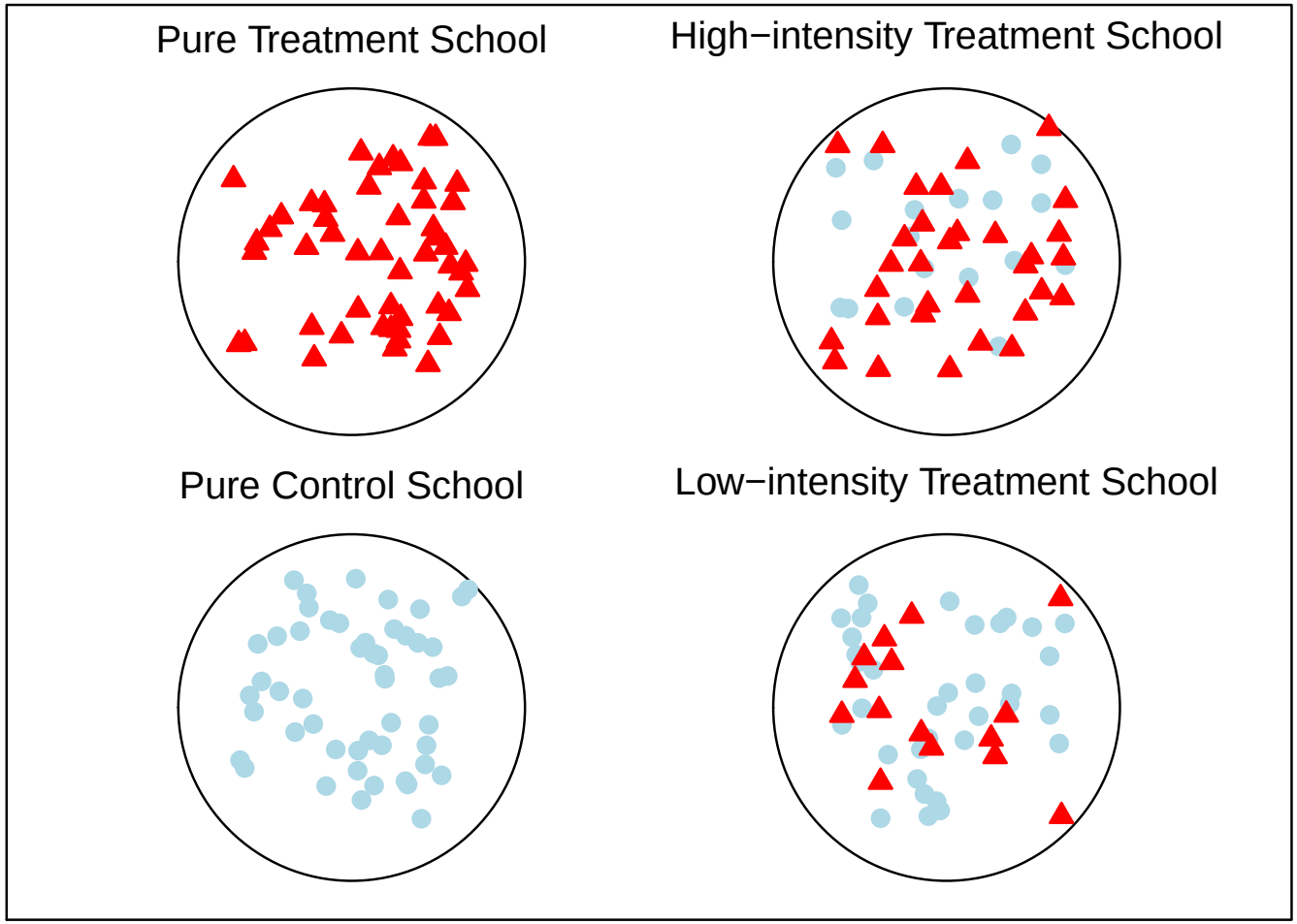
Solution?

- Design the unit of randomization to encompass the spillovers
- In the deworming case, if we expect spillovers to be present within schools, we should randomize at the school level
- And then assign the school to different treatment-intensity groups
  - Pure control: no pupils get treatment
  - Pure treatment: all pupils get treatment
  - Low-intensity treatment: 30% of pupils get treatment
  - High-intensity treatment: 75% of pupils get treatment

You can then measure the extent of the spillovers by comparing pure controls schools to the ones that received partial to full treatment.

Spillovers V

Measuring impact in the presence of spillovers



Threats

Potential threats randomized evaluations should be thought about:

- the sampled population may be specific if knowing there is randomization being conducted
- behaviour in control or treatment group may change (knowing they are part of an experiment)

On the ground,

- RCTs pose ethical problems
- RCTs are politically constraint
- Results only apply to specifics contexts (external validity?)

see Cilliers et al. (2013), Reddy (2011) and Ravallion (2009) for thorough discussions

Low-Hanging fruits for better health

What are the challenges faced?

- Irregular service delivery and non-qualified medical staff (supply-side)
- Tendency to postpone and beliefs matter (demand-side)

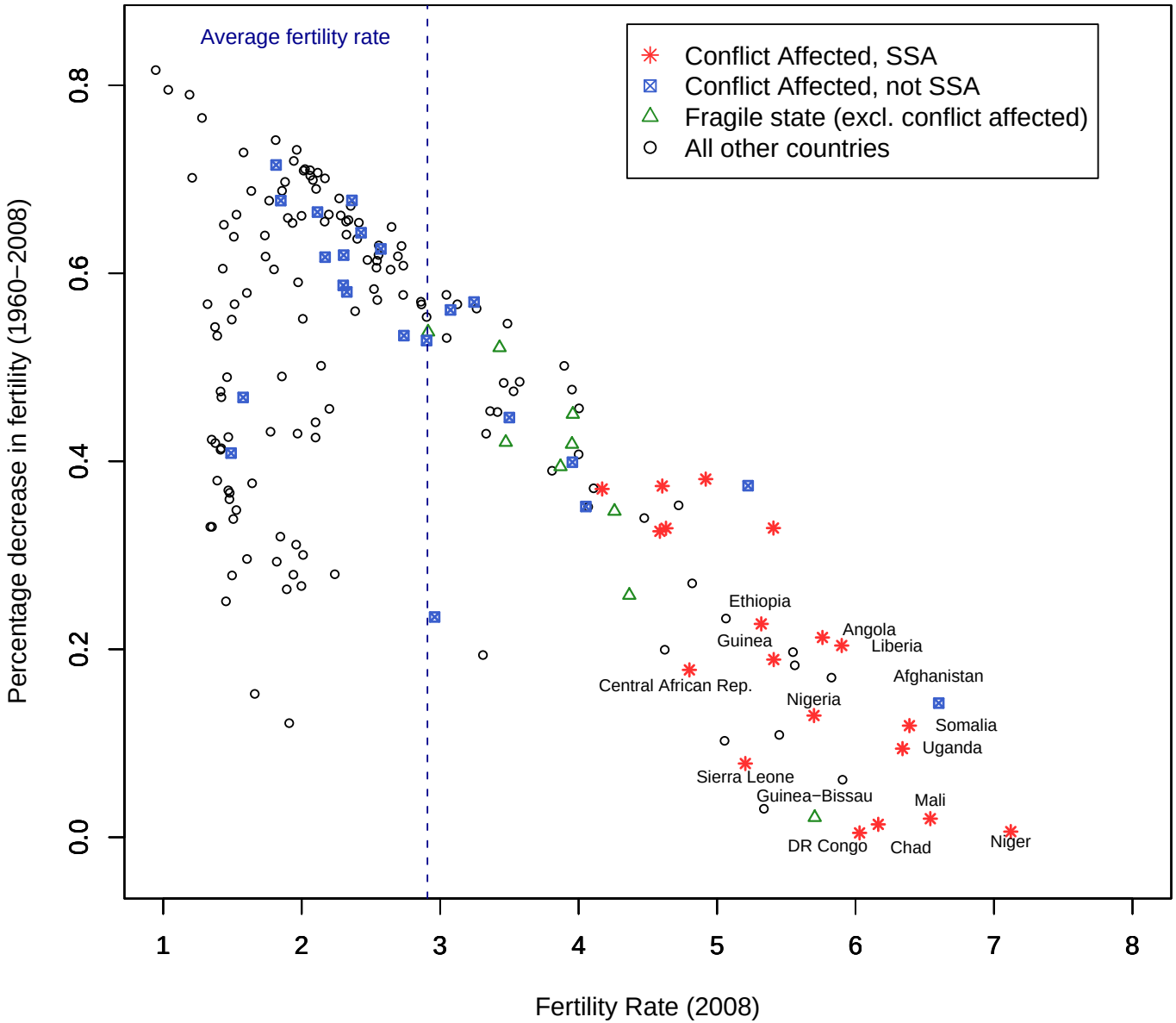
Prevention, chlorine, immunization, etc. are cheap but could have a high impact on improving health of the poor.

- Should people receive free bed-nets or pay for them?
- Should we "bribe" them to convince them to immunize their children?

We are surrounded by invisible nudges: clean water from the tap, compulsory immunization, health insurance, no worries about the next meal... people in the developing world don't have that luxury at their doorstep.

Family size I

Family size II



What causes what ?

- Is it poverty that pushes families to have more kids (insurance, replacement effects?)
- Or can it be that when you have to feed more persons, the pieces of the cake are smaller ? (unless technology (infinitely) increase the size of the cake)

Hard to say, and even harder to say what will happen as population continues to increase... *Citing The Economist, "To celebrate falling fertility is like congratulating the captain of the Titanic on heading towards the iceberg more slowly..."*

What can we do ?

- Force reduction (e.g. one child policy in China)?
- Provide access to contraception methods (supply-side)?
- or increase their usage (demand-side) ?

What should we do?

Do Teenagers Respond to HIV Risk Information? Evidence from a Field Experiment in Kenya - Pascaline Dupas (2011)

We know that

- better access to contraceptive can help teenagers to postpone pregnancies
- there is little done about unwanted teenage pregnancies... nor about the related issue of the spread of MSTs (including HIV/AIDS)
- standard message to teenage girls: "Abstain, Be faithful, use a Condom... or you Die", also referred as the ABCD strategy

Three programs were implemented :

1. ABCD alone ($\sim$ standard message) $\rightarrow$ statut-quo
2. Informing about "suggar daddies" $\rightarrow$ 1.2pp reduction in pregnancies
3. Paying school uniforms $\rightarrow$ 3pp reduction in pregnancies
4. ABCD + uniforms $\rightarrow$ undoes the effect of the uniforms

In the end...

- Fertility is a choice, and it depends on preferences
- Men and women have different preferences regarding family size
- Thinking about programs to be implemented, keep in mind that one household is not one person
- Final outcome depends on various factors (such as the legal, social, political and economic environment)
- It is NECESSARY for poor parents to have large families, as it constitutes their safety-nets in the present, and their insurance and support when they become too old to work

Reluctant entrepreneurs

Reluctant entrepreneurs

- Who are (micro-)entrepreneurs ?
- Why is micro-credit not enough to help tiny business grow bigger?
  - Risk and job stability
  - Time-inconsistency
  - Aspirations

In Sri Lanka, grants were randomly assigned to business owners:

- No grant
- US$ 250
- US$ 500

The men that invested large shares of both grants and got high returns. In contrast, women invest only the US$ 500 grant, and do not get return on their investment.

Conclusion

RCTs are a good method to evaluate the impact social programs. They can be THE credible way to evaluate impact IF

- Designed well
- Conducted well
- Interpreted well

and most importantly, the fundamental ethical requirements are met.

Still, very hard to meet all the "theoretical" requirements that would make them the perfect method - if only such method existed. Keep critical eye !

References

- Banerjee, A. and E. Duflo (2011), *Poor Economics: A Radical Rethinking of the way to fight global Poverty*, Public Affairs, New-York
- Sachs, J. (2006). The end of poverty: economic possibilities for our time. Penguin.
- Easterly, W. (2006). The white man's burden: why the West's efforts to aid the rest have done so much ill and so little good. Oxford University Press, UK
- Arceneaux, K., A. Gerber, D. Green (2006) Comparing Experimental and Matching Methods Using a Large Scale Field Experiment on Voter Mobilization, *Political Analysis* 14:1-36 (you can find the data and stata codes here)
- Miguel, E., & Kremer, M. (2004). Worms: identifying impacts on education and health in the presence of treatment externalities. Econometrica, 72(1), 159-217.
- Cilliers, J., Dube, O., & Siddiqi, B. (2013). "White Man's Burden"? A Field Experiment on Generosity and Foreigner Presence
- Reddy, S. G. (2011). Randomise This ! On Poor Economics.
- Ravallion, M. (2009). Evaluation in the Practice of Development. The World Bank Research Observer, 24(1), 29-53.
- Dupas, P. (2011). Do Teenagers Respond to HIV Risk Information? Evidence from a field experiment in Kenya. American Economic Journal: Applied Economics, 3(1), 1-34.
- De Mel, S., Mckenzie, D., & Woodruff, C. (2009). Are Women More Credit Constrained? Experimental Evidence on Gender and Microenterprise Returns. American Economic Journal: Applied Economics, 1(3), 1-32.

Questions ?

E-mail : odaoust@ulb.ac.be

Office : H.4.242 (send me an e-mail beforehand)