Impact Evaluation Collaborative

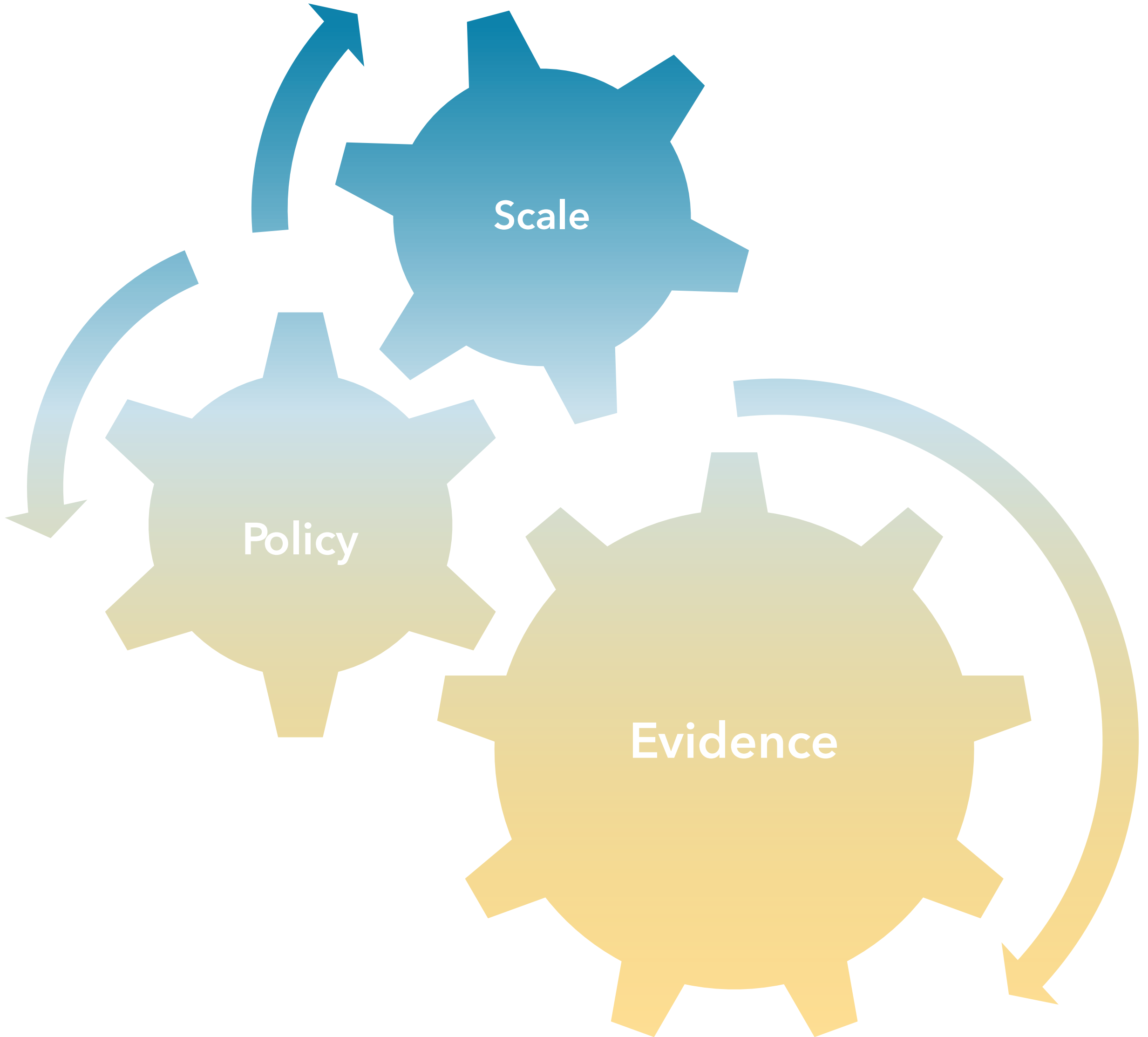
Moving Economic Inclusion to Scale

IE WORKSHOP

PARTNERSHIP FOR ECONOMIC INCLUSION

THE WORLD BANK
IBRD • IDA | WORLD BANK GROUP

i2i DIME
TRANSFORM DEVELOPMENT



Non-experimental impact evaluation methods

PARTNERSHIP FOR ECONOMIC INCLUSION
THE WORLD BANK
IBRD • IDA | WORLD BANK GROUP

i2i
DIME
TRANSFORM DEVELOPMENT

What we know so far

We want to isolate the causal effect ("impact") of our interventions on outcomes of interest

- Key problem is the search for a **counterfactual**: what would have happened to our participants in absence of the project?
- Challenges are:
  1. **Comparison over time:** Other things are happening at the same time, e.g., price and weather shocks
  2. **Comparison across households:** We don't know why certain people participate
- Objective is finding a suitable control group that acts as a **counterfactual**

What we know so far

- Randomizing the assignment to "treatment" is the "gold standard" methodology (simple, accurate, cheap)
    - Rely on few assumptions
    - Less data required
    - Easy to explain
- **What if we really cannot use randomization?**
    - e.g., large infrastructure projects that can't be randomized (roads, refugee camps, …)
    - There are other methods (difference-in-differences, matching, discontinuity)
    - Other methods rely on <u>key assumptions</u>
- Mixing of methods is possible!
    - RCTs and non-experimental methods are complementary, not substitutes!

I. DIFFERENCE-IN-DIFFERENCES

II. REGRESSION DISCONTINUITY DESIGN

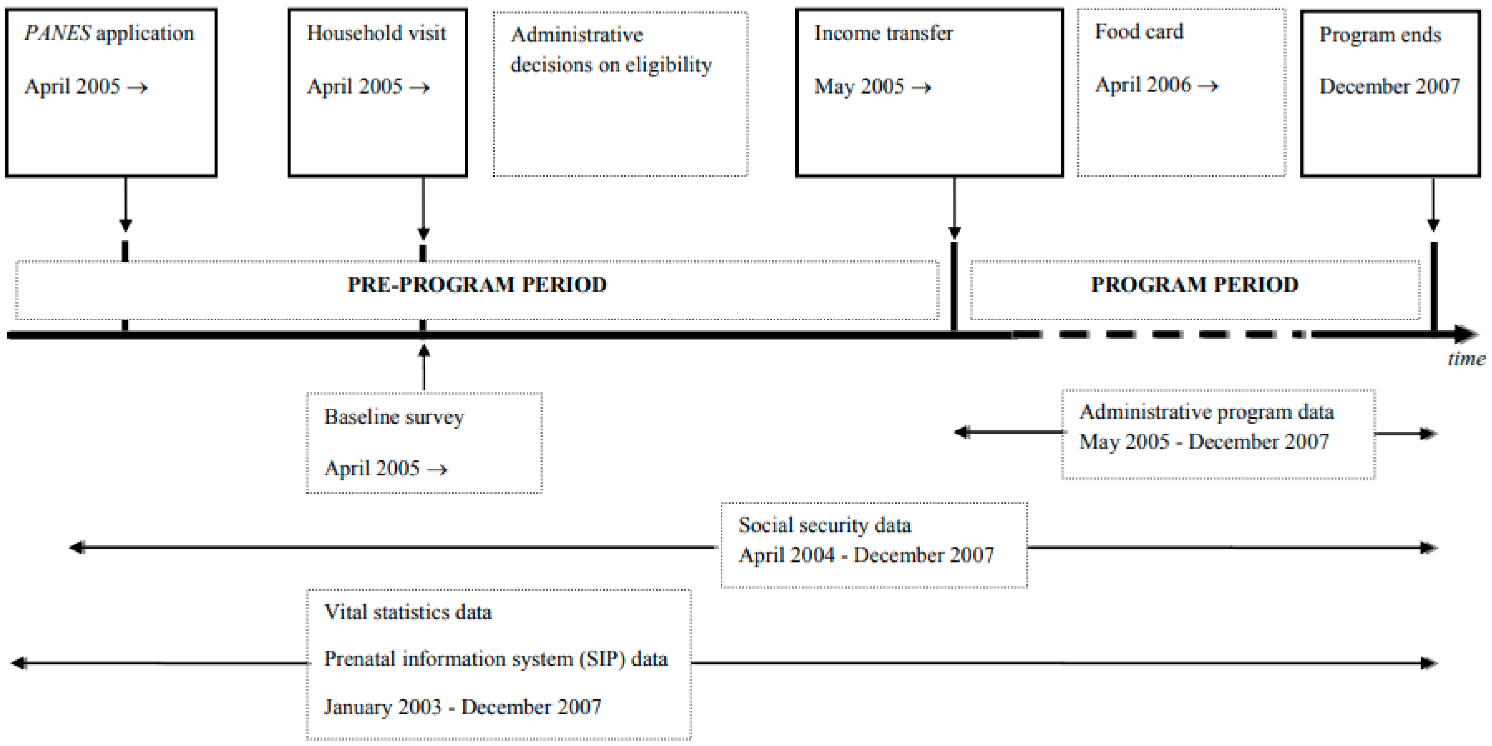
III. COMBINING METHODS

Case study 1: Uruguay PANES

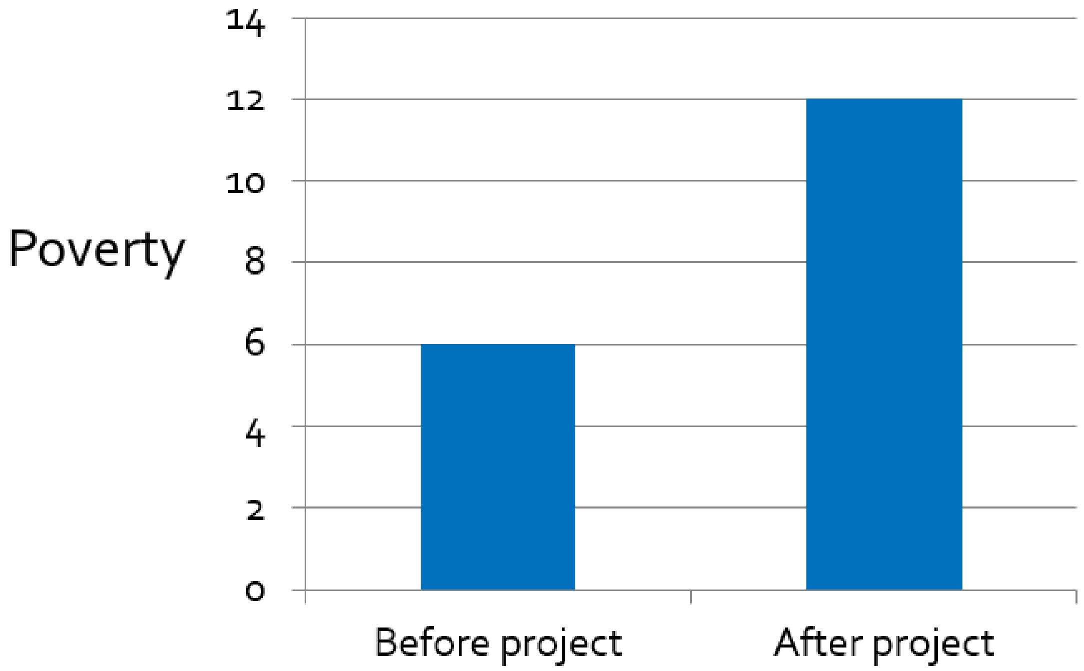
- Uruguay's 2005-2007 Plan de Atención Nacional a la Emergencia Social (PANES)
  - Temporary social protection program targeting poorest 20% of households below poverty line
  - Motivated by 2001-2002 crisis in neighboring countries
- PANES combined monthly cash transfer, food card for families, emergency employment, and trainings
- Amarante et al. (2011) use a mix of non-experimental methods to show the program reduced low birthweight by 15%

Case study 1: Uruguay PANES

Figure 1: Timing of *PANES* program activities and data collection

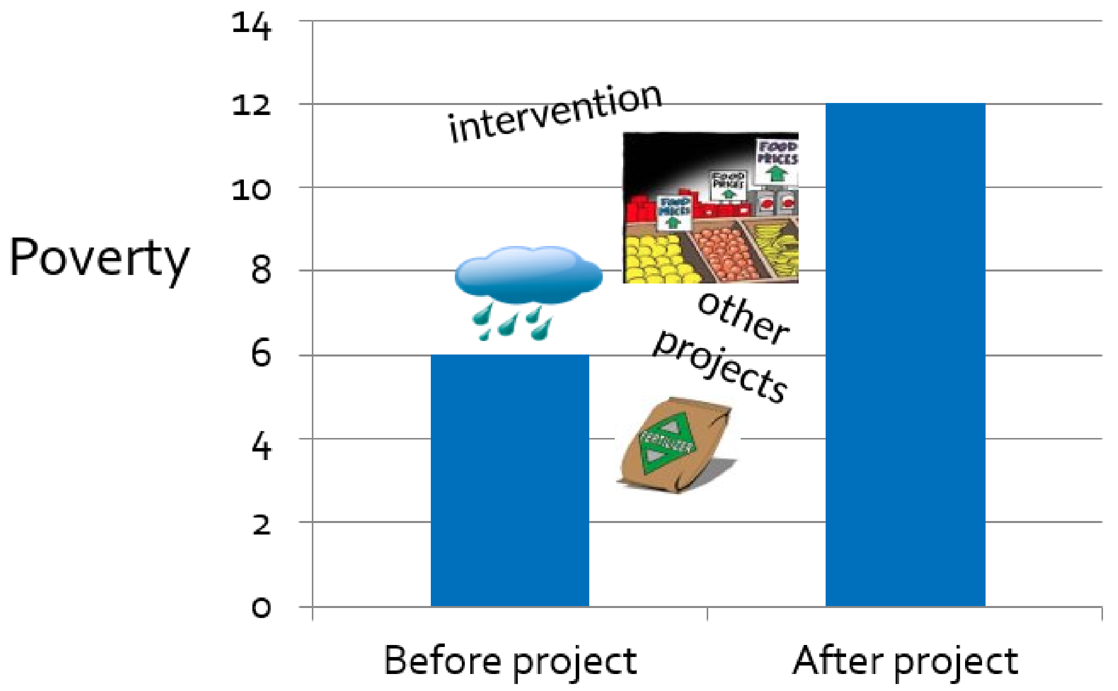


Compare beneficiaries before and after?



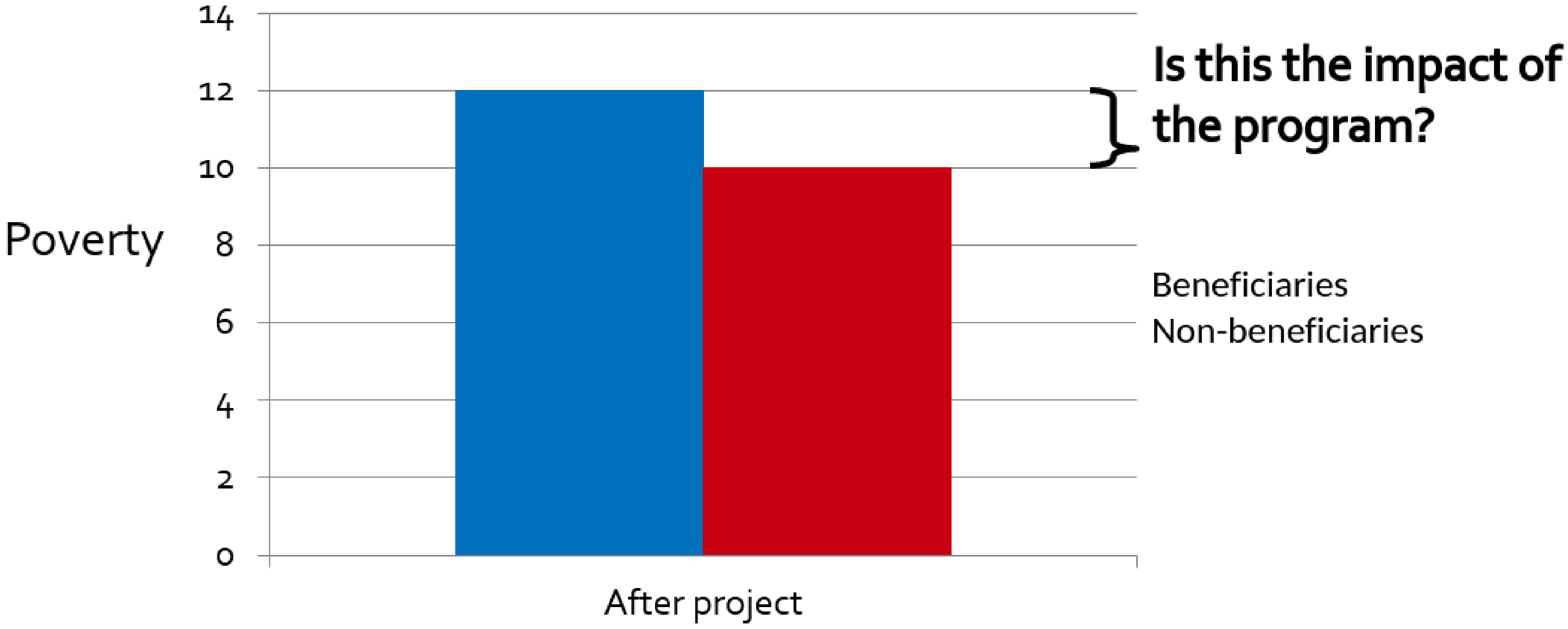
Is this the impact of the program?

Compare beneficiaries before and after?

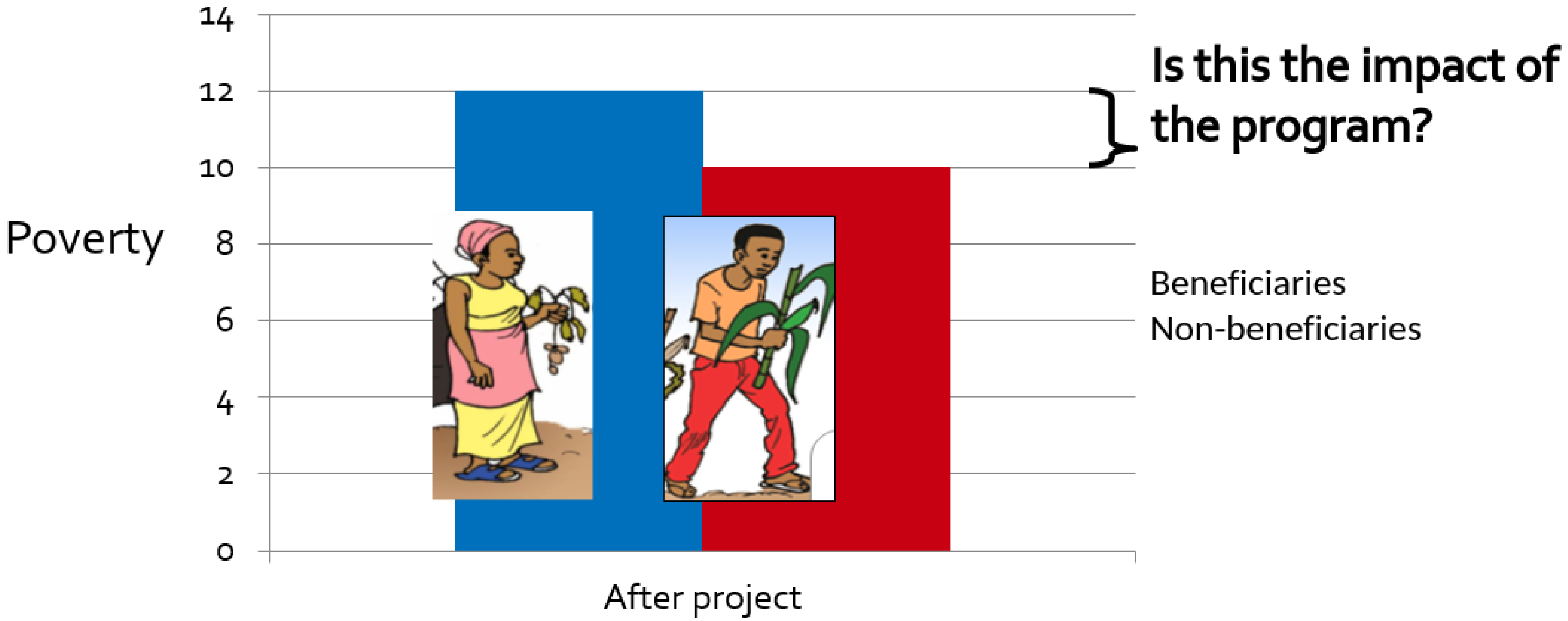


Is this the impact of the program?

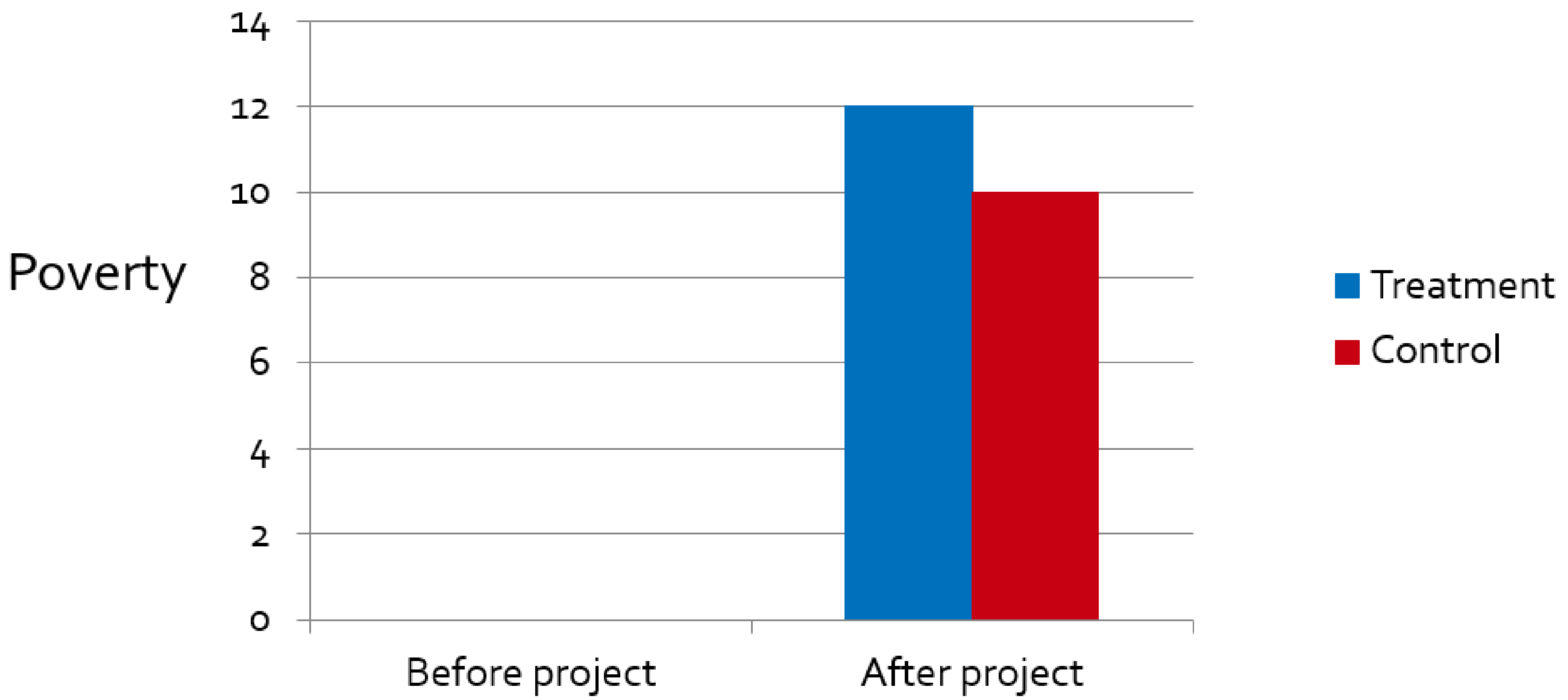
Compare beneficiaries and non-beneficiaries?



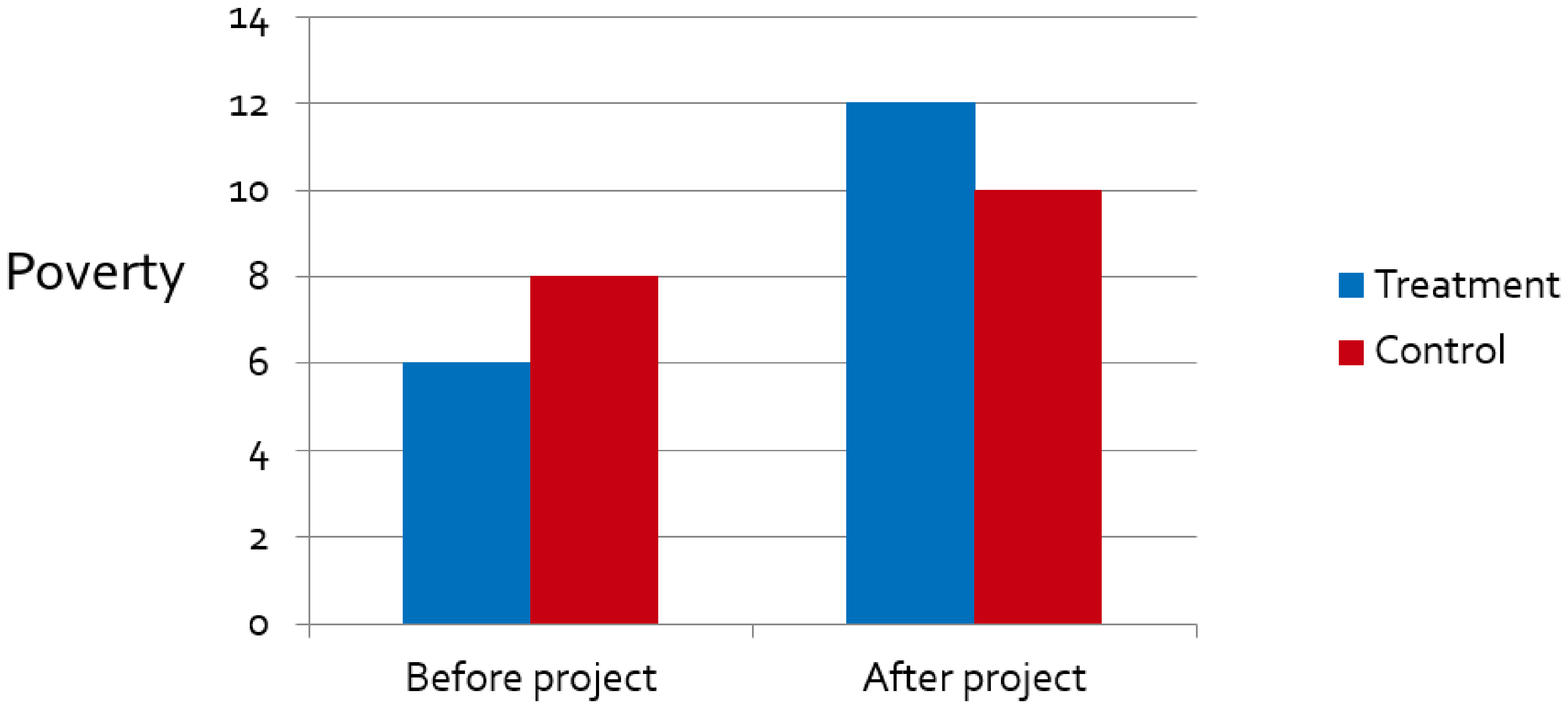
Compare beneficiaries and non-beneficiaries?



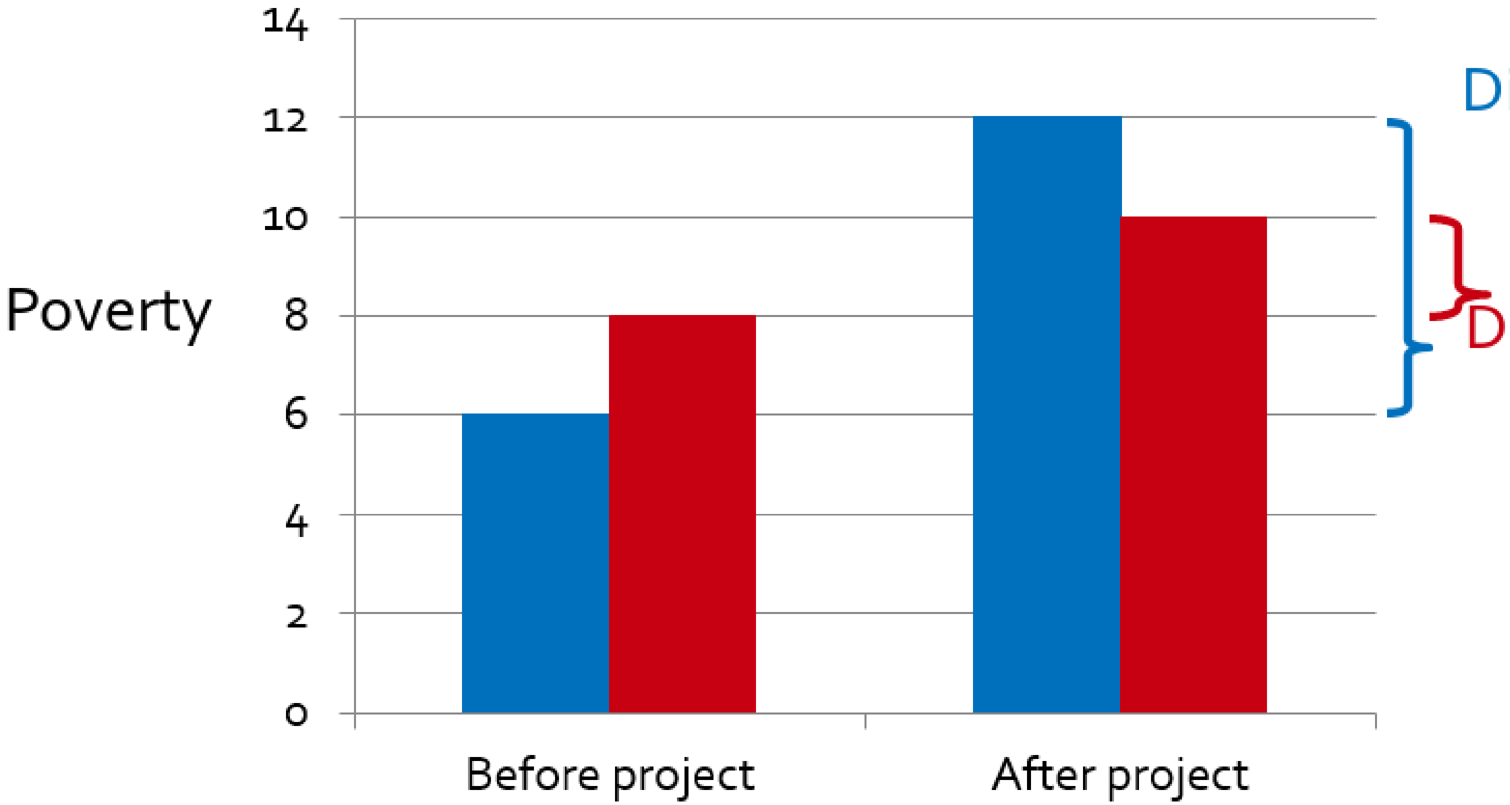
D-i-D: Combine the two differences?



D-i-D: Combine the two differences?



Difference-in-differences



Difference in treated group

—

Difference in control group

=

Treatment effect

Difference-in-differences

- Difference #1: compare over time, before and after the program
- Difference #2: compare treatment and control groups

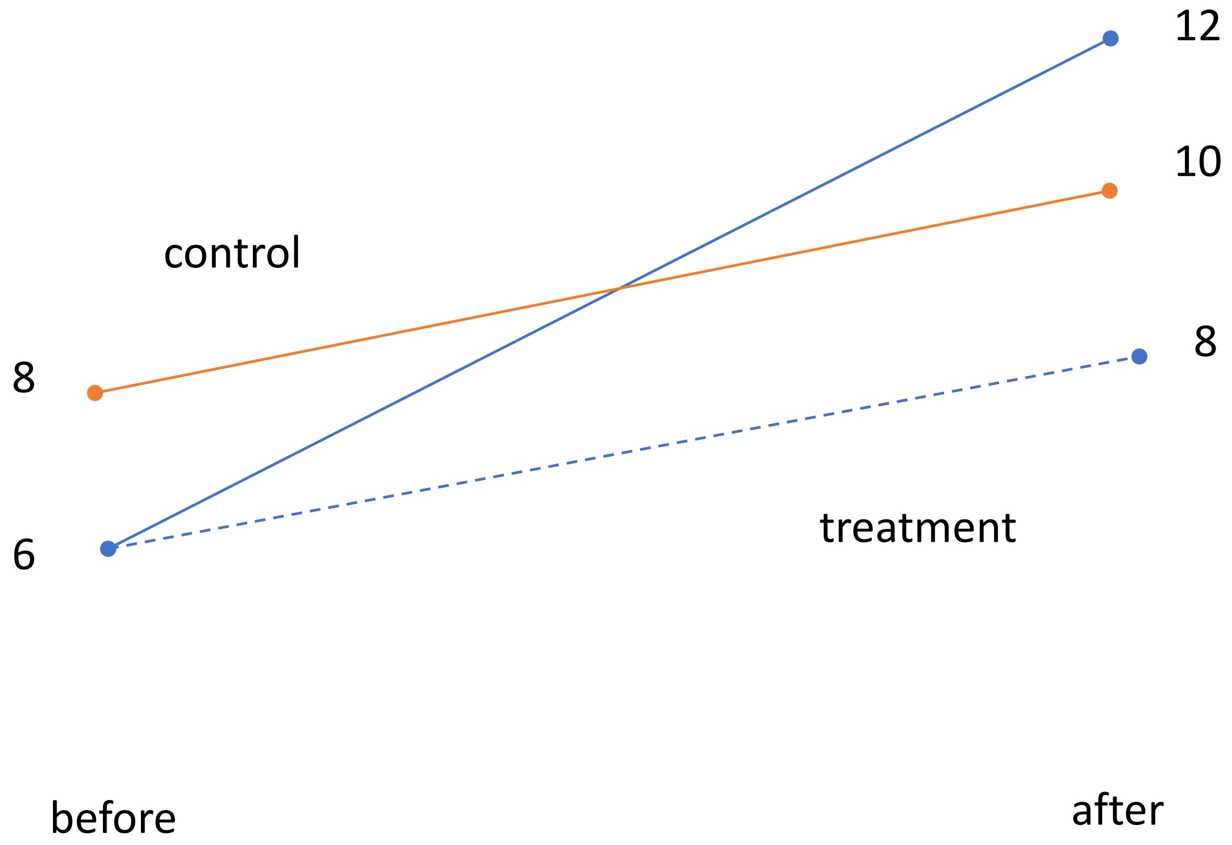
	Treatment	Control	Difference
Before	6	8	-2
After	12	10	2
Difference	6	2	4

Treatment - Control

Before - After

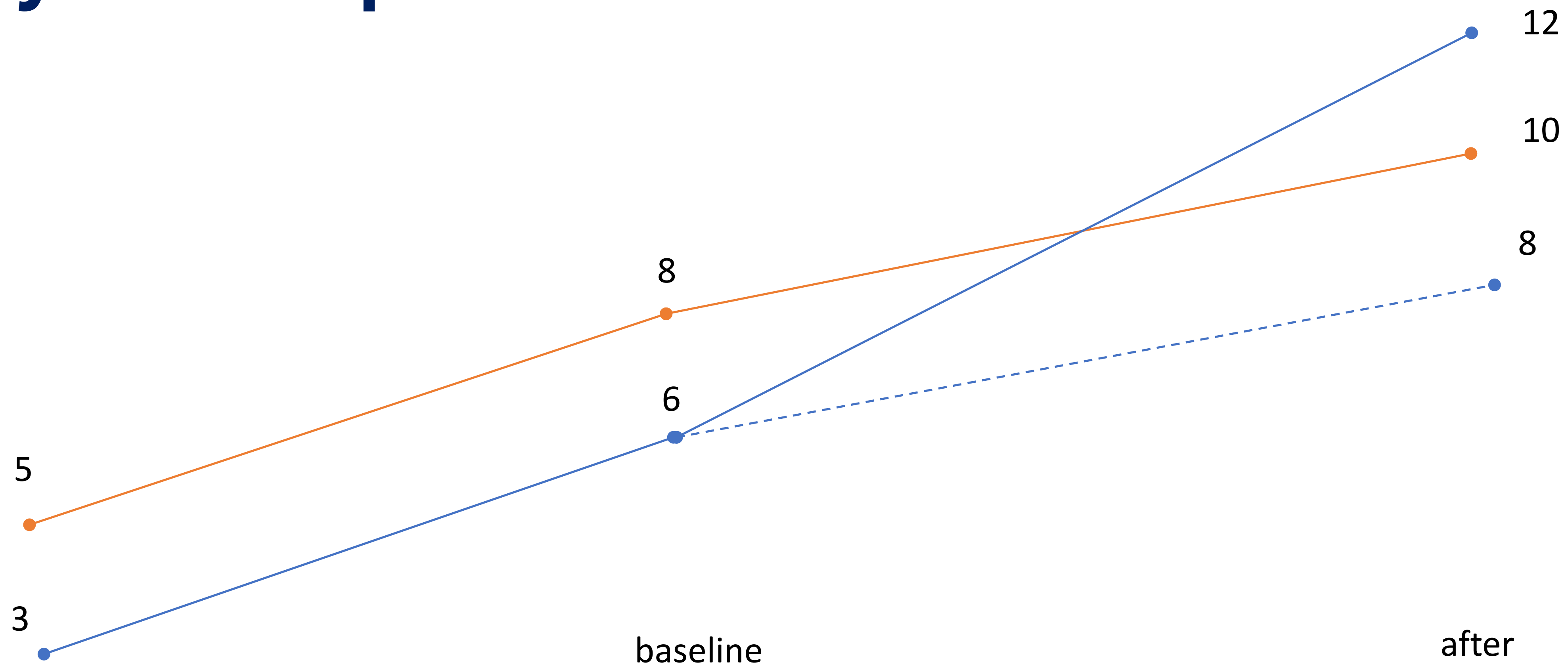
Difference-in-differences

Key assumption: Parallel trends

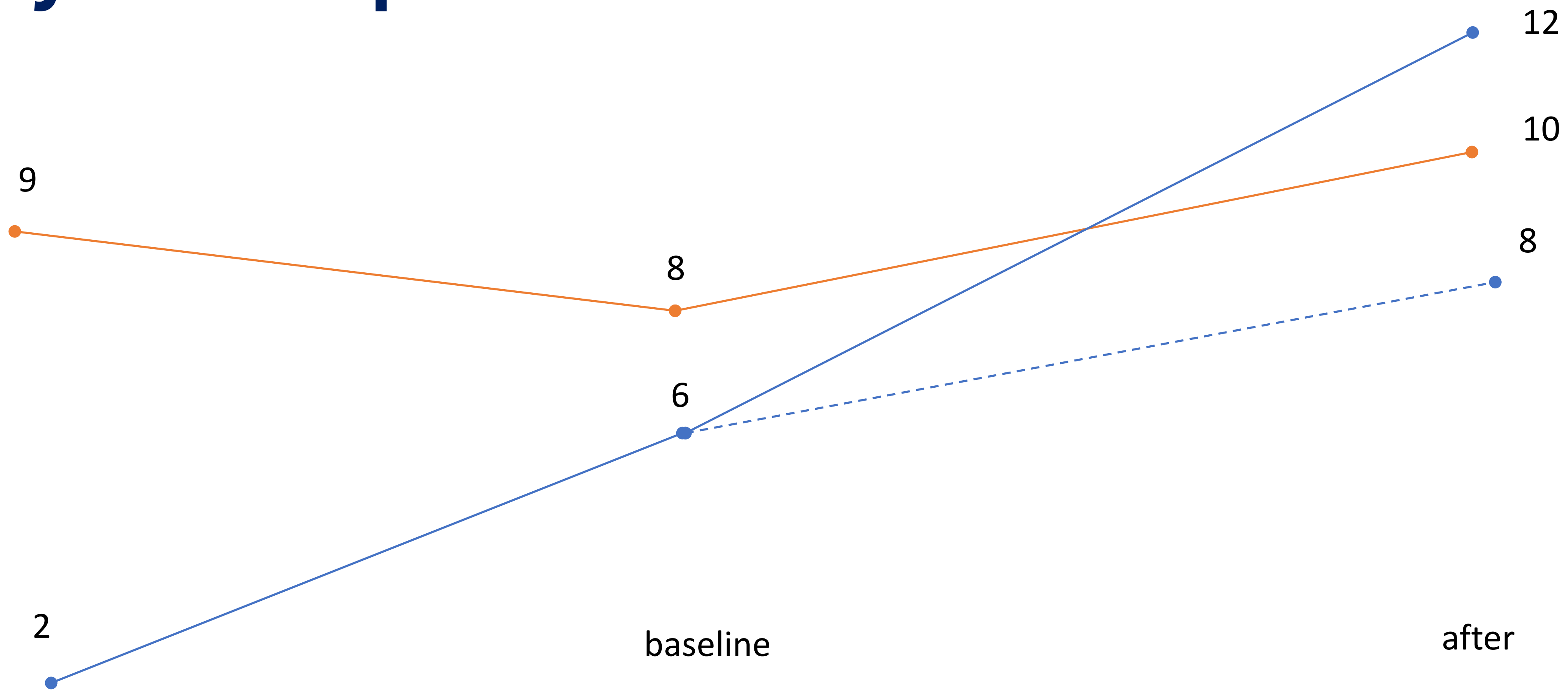


Non-experimental impact evaluation methods

Key assumption: Parallel trends



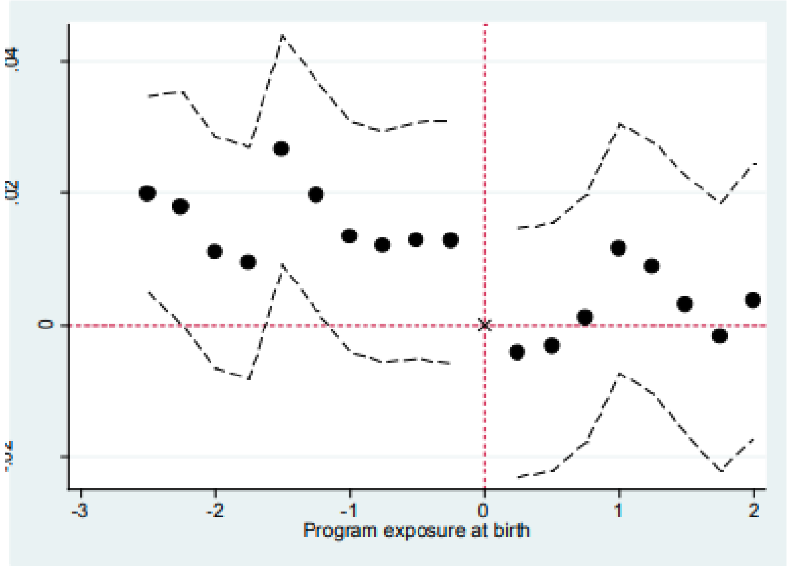
Key assumption: Parallel trends



Case study 1: Difference-in-differences

- Eligible PANES applicants
  - 2003 – 2005 (pre-PANES) low birthweight: 0.102
  - 2005 – 2007 (post-PANES) low birthweight: 0.091
- Ineligible PANES applicants
  - 2003 – 2005 (pre-PANES) low birthweight: 0.093
  - 2005 – 2007 (post-PANES) low birthweight: 0.091
- Difference-in-differences estimate
  - (Treated after – Treated before) - (Control after – Control before) =
  - (0.091 - 0.102) - (0.091 - 0.093) = -0.009
  - 10% decrease in low birthweight

A: Y-axis: Low birthweight



How do we know?

- Compare history of control and treatment groups before baseline
    - Sometimes administrative data is available, but often limited
- More likely to hold when groups are similar at baseline, and treatment selection is based on criteria other than the outcome indicator of interest
    - Often not the case: Targeting is frequently determined by outcomes we care the most about, e.g., poverty

Identifying a control group

Which of these two groups could serve as a counterfactual? Choose!

1. Households in the same country but without PANES?
2. Households outside the country and without PANES?
3. Both?
4. Neither?

Why didn't those communities get the intervention?

- Selection criteria are sources of differences!

Case study 2: Turkey ESSN

- In some cases, matching methods can help improve balance between control and treatment groups in a difference-in-differences design
  - Özler et al. (2021) find an emergency cash transfer program targeting refugees in Turkey increased household consumption, but induced children to shift from treated to control households
  - To improve credibility of difference-in-difference estimates, they compare changes in outcomes among matched treated and control households with similar characteristics

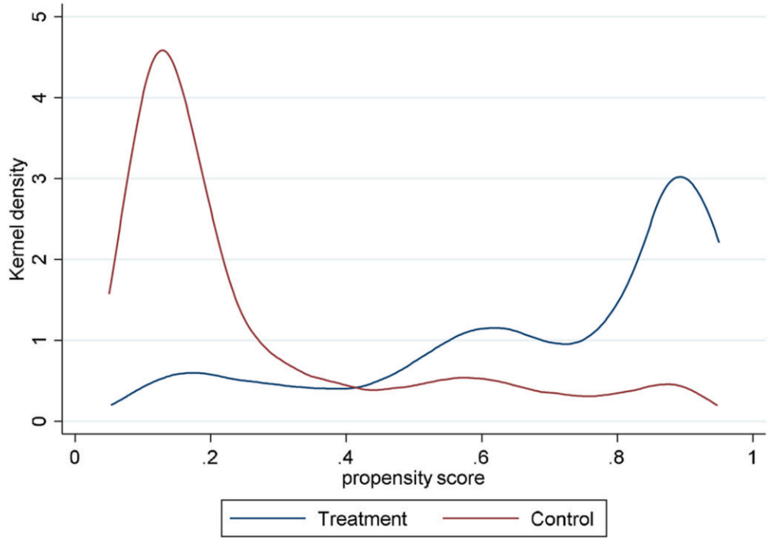
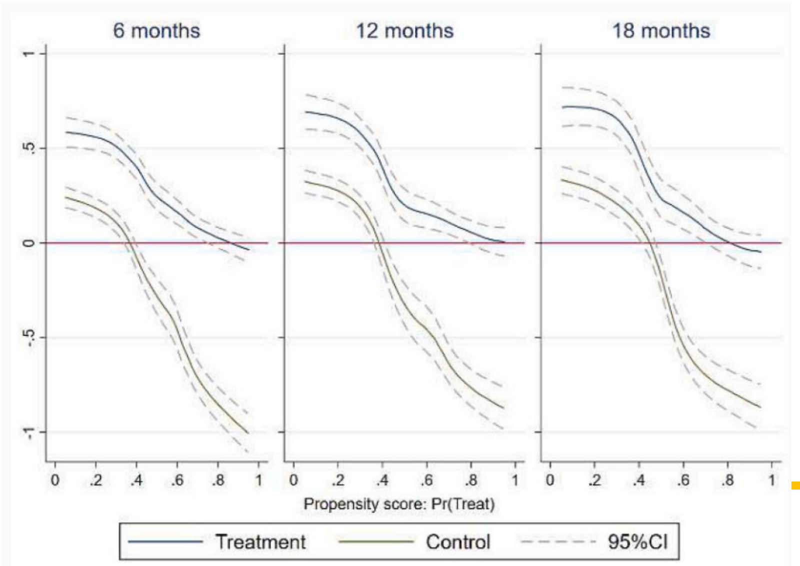


Fig. 1. Kernel density smoothing of propensity score across treatment and control samples.

Panel B: Changes in the number of children aged 0-17 years since baseline by propensity score



Summary: Difference-in-differences

- Compare treatment and control groups before AND after the project interventions
- Pick control that is as similar as possible to the treatment group at baseline
    - Create a long list of sites that could receive the project
    - Use historical data available to pick comparison sites
- If selection already happened, we need historical data to make the parallel trends assumption credible

Regression discontinuity designs

- Regression discontinuity designs (RDD) are more similar to randomization
  - Identifying "almost random" assignment from selection process
- Need a clear and enforced eligibility rule
  - A simple, quantifiable score ("threshold")
- Assignment to treatment must be based on this rule
  - e.g., target households with poverty score above a threshold
  - e.g., target households with children below a certain age
- Basic idea: Compare individuals just above and just below threshold

RDD logic

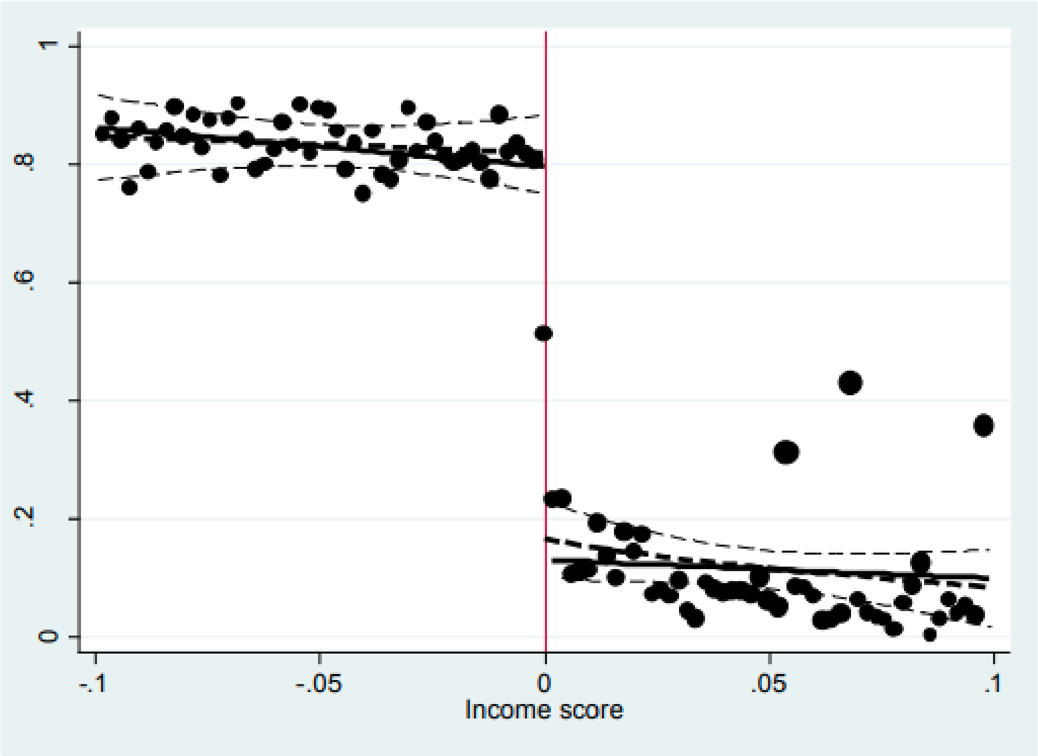
- Assignment to treatment depends on continuous "score" or ranking (e.g., child's age)
    - Potential beneficiaries are ordered by score
    - There is a cut-off point for "eligibility" -- clearly defined pre-determined criterion
    - Cut-off determines assignment to treatment
- This usually results from administrative decisions
    - e.g., resource contraints limit coverage
    - e.g., very targeted intervention expected to be more suitable for some people
    - Transparent rules rather than discretion used

Case study 3: Uruguay PANES

- Two approaches to estimating the impacts of Uruguay PANES
  - **Difference-in-differences**: Compare eligible and ineligible beneficiaries, before and after the program
  - **Regression discontinuity design**: Compare just barely eligible and just barely ineligible beneficiaries

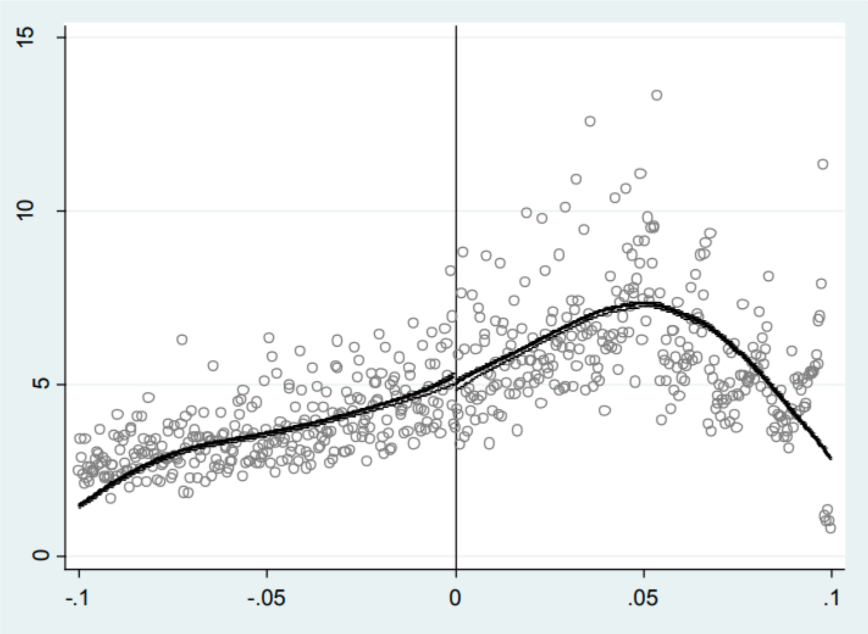
Below threshold, most receive PANES

C: Treated



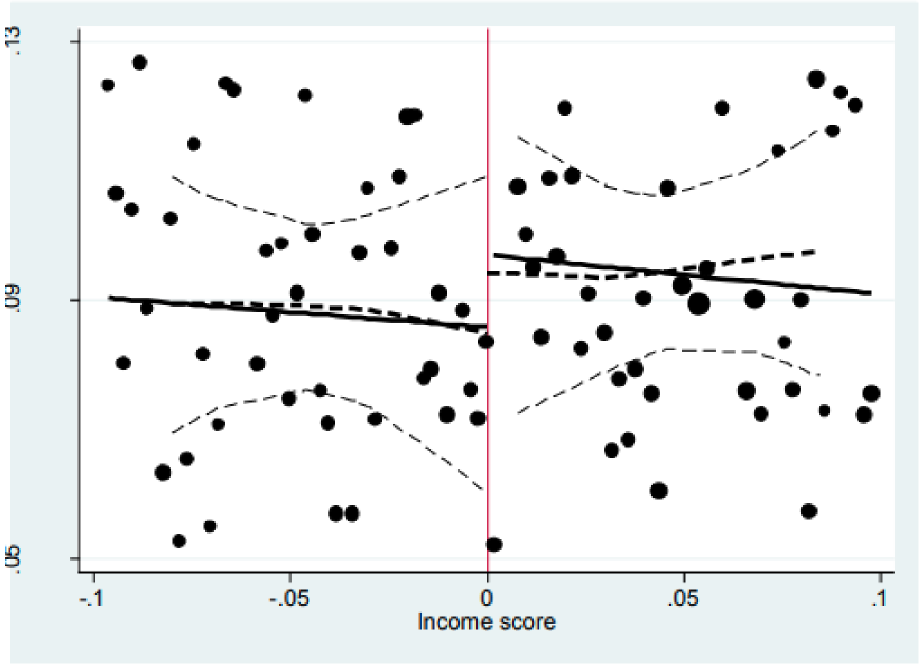
Many observations near threshold, no evidence of manipulation

Panel A: Distribution of the standardized *PANES* predicted income score, McCrary (2008) test

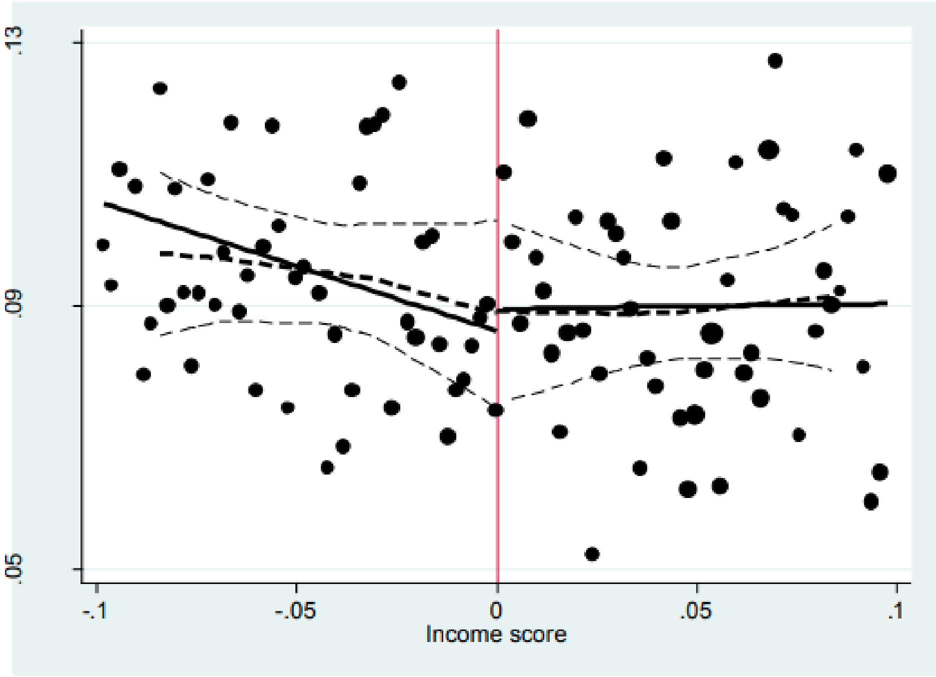


Differences in birthweight only emerge after program

A: Low birthweight, program period



B: Low birthweight, pre-program period



RDD drawbacks

- How generalizable are the results?
  - They only tell us about the impact of economic inclusion programs on birthweight for households at the threshold!
  - Economic inclusion may have different impacts on birthweight for the poorest households, or richer households!
- Hard to know in advance how many households will be close to the cut-off

Using RDDs

- Major advantages
  - Transparency
  - Graphical, intuitive presentation
- Major shortcomings
  - Requires many observations around cut-off
  - Not guaranteed ex-ante
- Why?
  - Can only estimate impacts using sample close to cut-off
  - Results therefore most applicable only to households close to cut-off

Summary: RDD

- Randomized control trials require minimal assumptions and provide intuitive estimates
    - Not always feasible
- Non-experimental methods require assumptions that must be carefully tested
    - More data-intensive
    - Not always testable
    - Challenging to use when you want to "unbundle" impacts

How to fit things together

- **Get creative**
  - Mix-and-match types of methods!
    - In example of PANES: Finding similar results using regression discontinuity and difference-in-differences improves credibility of results
    - Sometimes, may be possible to implement multiple methods on one data set (e.g., administrative data), but for other outcomes may need to use just one method
      - **Difference-in-differences requires many observations of the same individuals over time**
      - **Regression discontinuity requires many observations of individuals near the threshold**
  - Address relevant questions with relevant techniques

Thank you!

Presenter's name
Contact

PEI FUNDING PARTNERS

PARTNERSHIP FOR ECONOMIC INCLUSION

THE WORLD BANK
IBRD • IDA | WORLD BANK GROUP

brac

Co— Impact

german cooperation
DEUTSCHE ZUSAMMENARBEIT

Implemented by

giz Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH

Irish Aid
Rialtas na hÉireann
Government of Ireland