International Initiative for Impact Evaluation

Causal Inference*

Sebastian Martinez, 3ie

June 27, 2023

*Source: Gertler et al. (2016) Impact Evaluation in Practice, 2nd Edition.

Objective

“

Estimate the causal effect (impact) of intervention (P) on outcome (Y).

(P) = Program or Treatment

(Y) = Indicator, Measure of Success

Example: What is the effect of an Aquaculture Productivity Enhancement Program **(P)** on Household Income **(Y)**?

Causal Inference

What is the impact of (P) on (Y)?

$$\delta = (Y \mid P=1) - (Y \mid P=0)$$

Problem of Missing Data

$$\delta = (Y \mid P=1) - (Y \mid P=0)$$

For a farmer in our program:

- we observe $(Y \mid P=1)$: Household Income (Y) with an Aquaculture program (P=1)
- but we do not observe $(Y \mid P=0)$: Household Income (Y) without an Aquaculture Program(P=0)

Solution

Estimate what ***would*** have happened to *Y* in the absence of *P*.

We call this the **Counterfactual**

The key to a good impact evaluation is a valid estimate of the **counterfactual!**

Estimating impact of *P* on *Y*

$$\delta = (Y \mid P=1) - (Y \mid P=0)$$

OBSERVE $(Y \mid P=1)$
Outcome with treatment

ESTIMATE $(Y \mid P=0)$
The Counterfactual

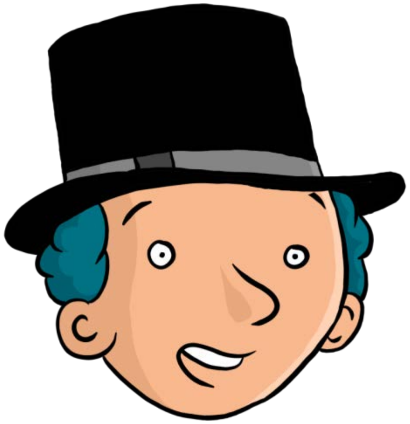
IMPACT = Outcome with treatment - counterfactual

- Intention to Treat (**ITT**) – *Those offered treatment*
- Treatment on the Treated (**TOT**) – *Those receiving treatment*

- Use **comparison** or **control** group

Example: What is the Impact of…

giving Ramesh



Aquaculture Enhancement

(P)

on Ramesh’s income

(Y)?

The Perfect *Clone*

Ramesh

Ramesh’s Clone

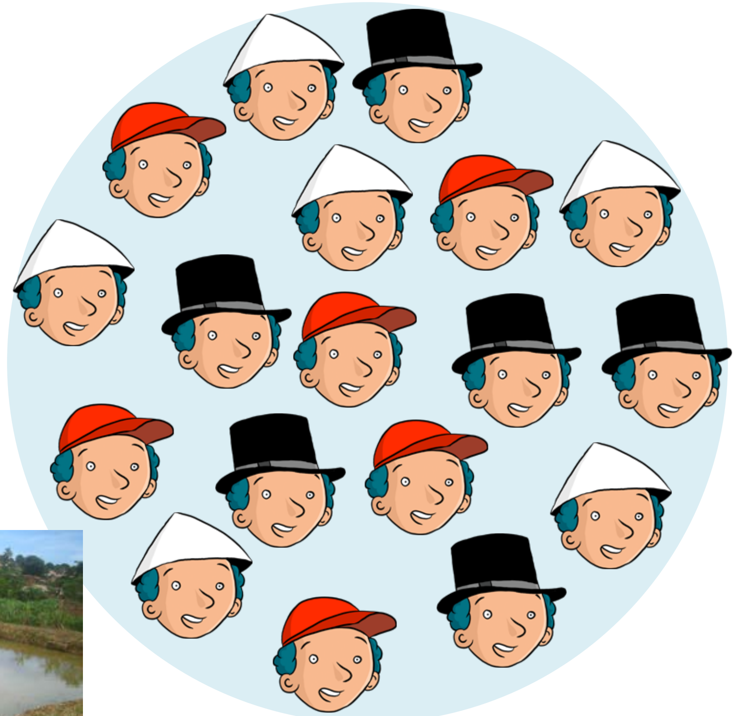
600

400

IMPACT=600-400=200

In reality, use statistics

Treatment



Comparison

Average Y=600

Average Y=400

IMPACT=600-400=200

Two common but problematic estimates of the counterfactual

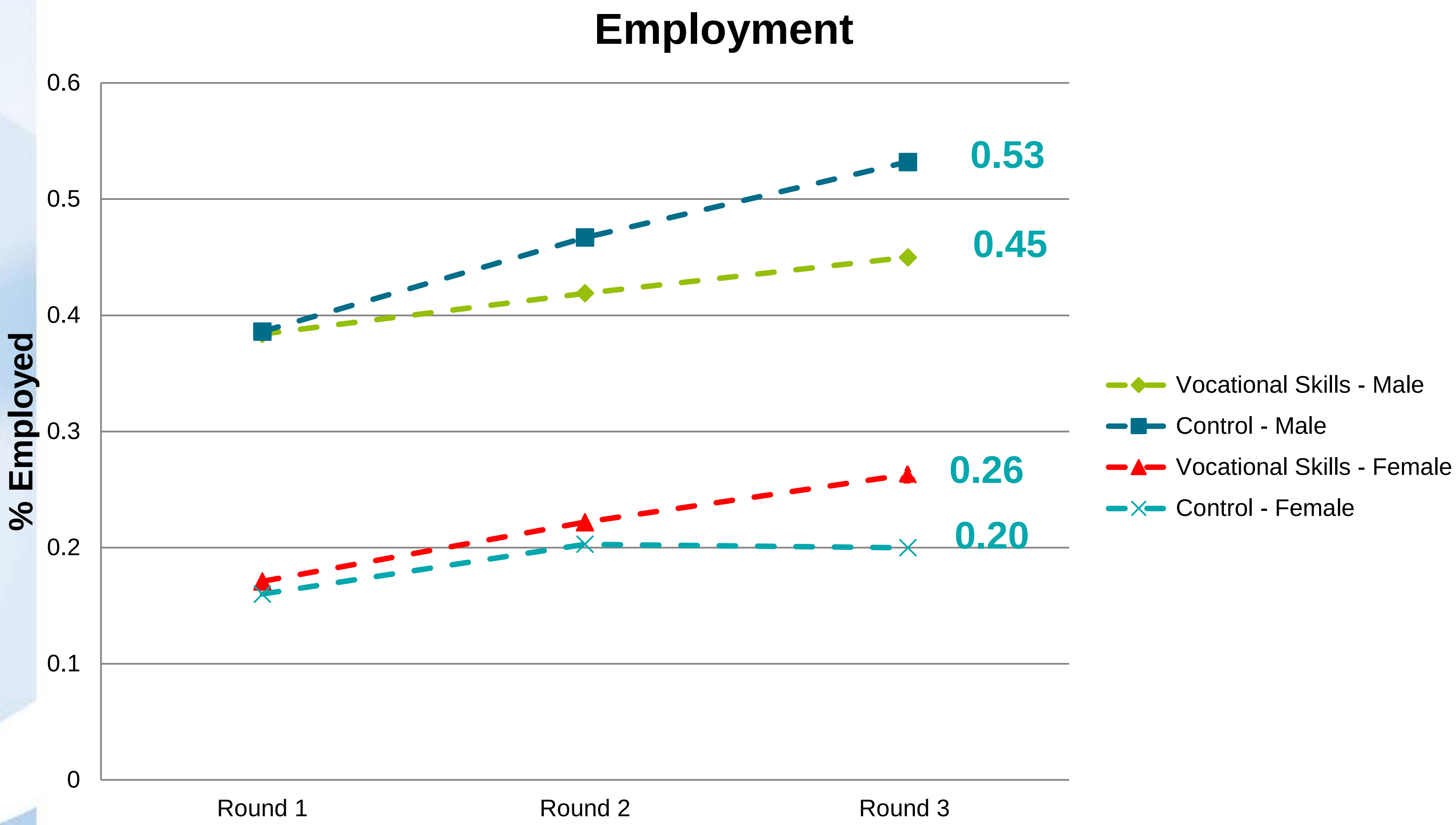
Impact of an Employment Program

Baseline

6 month follow-up

12 month follow-up

Impact of an Employment Program



Baseline

6 month follow-up

12 month follow-up

Baseline condition (Pre-Post)

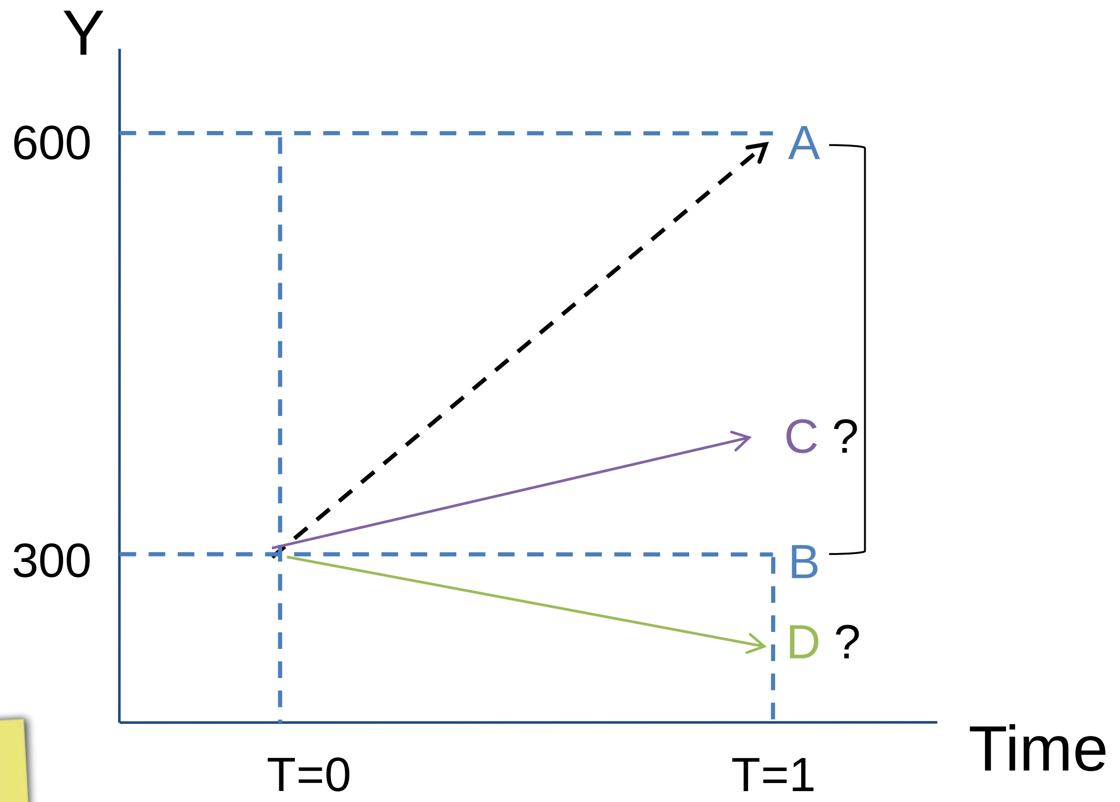
Good Weather:

- o Real Impact=A-C
- o A-B is an *overestimate*

Bad Weather:

- o Real Impact=A-D
- o A-B is an *underestimate*

Before & After doesn't control for other time-varying factors!



Self Selection

- If we have post-treatment data on
  - Enrolled: *choose to participate*
  - Not-enrolled: *choose NOT to participate*
    - *High potential for biased estimate of counterfactual*

- Selection Bias
  - Reason for not enrolling may be correlated with outcome (Y)
    - *Control for observables.*
    - *But not un-observables!*
  - Estimated impact is confounded with other factors that determine treatment status

Solution:

Implement an **identification strategy** to control for time varying factors and minimize selection bias.

Compare:

To

Random Assignment (RCT)

Regression Discontinuity Design

Instrumental Variables

Difference-in-Differences

Interrupted Time Series

Synthetic Control

Matching

IE Methods Toolbox

Questions?

Thank you

www.3ieimpact.org

@3ieNews

/3ieimpact

/3ievideos

international-initiative-for-impact-evaluation

New Delhi London Washington, DC