Day 3 - Part 2: Public Policy Design and Evaluation

New Climate Economy Training Course

World Resources Institute

July 2021

Outline

1. Why evaluation?
2. Theory of Change (Toc)
3. Impact Evaluation: The basics
4. Selection bias

Program Evaluation

What is our goal?

Programs and policies are designed to achieve a goal (or a set of goals). For example:

Program Evaluation

What is our goal?

Programs and policies are designed to achieve a goal (or a set of goals). For example:

- The main objective of *Oportunidades* program in Mexico is to break intergenerational transmission of poverty by alleviating current poverty while investing in human capital of the next generation.
- In India, a local ONG designed a conditional cash transfer program that helps landless households participate in temporary migration during the hungry season, with the objective of decreasing rural poverty and inequality while also improving rural livelihoods

Program Evaluation

What is our goal?

Let's take the following example: **Immunization Incentives**

The Problem:

- Despite availability of free immunization, full coverage rates among children remains extremely low in many developing countries.

Intervention

- Reliable, monthly immunization camps set up in villages in Udaipur.
- Small incentives offered to mothers conditional on having child immunized; larger incentive when immunization course completed.

Program Evaluation

What do we measure?

What of these questions fits better with the description of an "impact evaluation"?

- What percentage of 3 year old children in Rajasthan were not fully immunized?
- Does holding regular immunization camps and providing incentives to parents improve immunization rates of children?

Program Evaluation

What do we measure?

What of these questions fits better with the description of an "impact evaluation"?

- What percentage of 3 year old children in Rajasthan were not fully immunized?
- Does holding regular immunization camps and providing incentives to parents improve immunization rates of children?

Program Evaluation

What do we measure?

What of these questions fits better with the description of an "impact evaluation"?

- What percentage of 3 year old children in Rajasthan were not fully immunized?
- Does holding regular immunization camps and providing incentives to parents improve immunization rates of children?

In simple words:

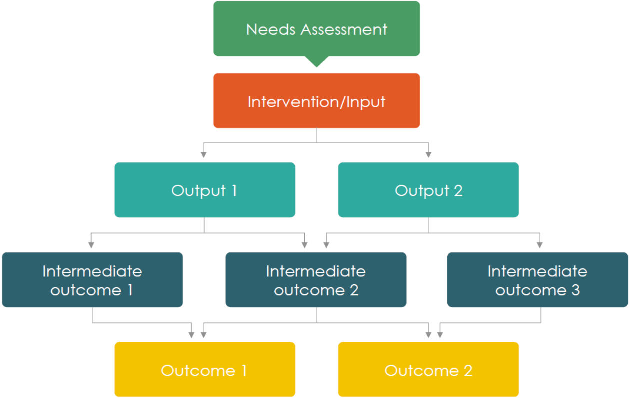
- An impact evaluation tries to determine if a program has an impact on some specific outcome(s) and, more importantly, how big is this impact.

Theory of Change (Toc)

Definition:

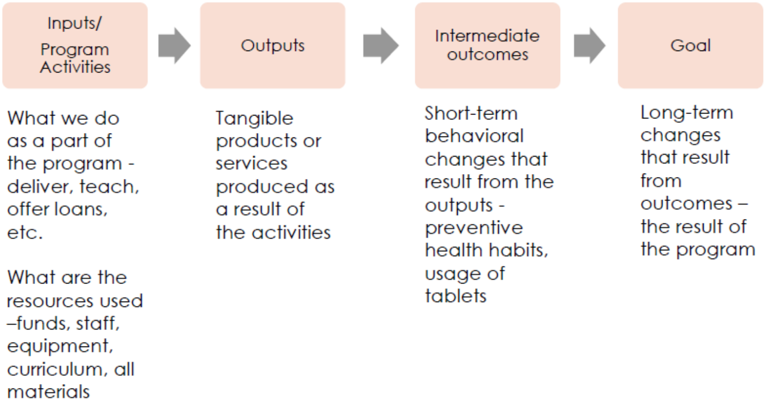
- ToC is and on-going **process** of reflection to explore change and how it happens...
  - And what that means in a particular context, sector, and/or group of people.
- Toc is an structured way of thinking about change and impact to be achieved.
  - It defines a set of connected building blocks, generally called inputs, outputs, and outcomes.

Theory of Change (Toc)



Source: JPAL.

Theory of Change (Toc)



Source: JPAL.

Theory of Change: Why do we need it?

- Helps design the intervention
  - Often done backwards.
  - Is each arrow really credible?

Theory of Change: Why do we need it?

- Helps design the intervention
  - Often done backwards.
  - Is each arrow really credible?
- Helps design the evaluation by:
  - Generating research questions.
  - Deciding which data to collect

Theory of Change: Why do we need it?

- Helps design the intervention
  - Often done backwards.
  - Is each arrow really credible?
- Helps design the evaluation by:
  - Generating research questions.
  - Deciding which data to collect
- By measuring the right intermediate variables, we can get "into the black box":
  - Allows to understand the "why", thus giving richer policy lessons
  - Gives more generalizable knowledge.

Theory of Change (Toc)

In sum, the design of a “good” ToC can be build in 6 steps

- Situation analysis – Specifying the context
- Clarify the program goal
- Design the program /product
- Map the causal pathway
- Explicate assumptions
- Design SMART indicators

Impact Evaluation: The basics

Counterfactual: What is it?

Counterfactual $\rightarrow$ The state of the world that program participants would have experienced in the absence of the program (i.e. had they not participated in the program)

Impact Evaluation: The basics

Counterfactual: What is it?

Counterfactual $\rightarrow$ The state of the world that program participants would have experienced in the absence of the program (i.e. had they not participated in the program)

But wait, is that even possible?...

Impact Evaluation: The basics

Counterfactual: What is it?

Counterfactual $\rightarrow$ The state of the world that program participants would have experienced in the absence of the program (i.e. had they not participated in the program)

But wait, is that even possible?...

Problem: Counterfactual cannot be observed in practice. You can only observe one "state of the world".

Impact Evaluation: The basics

Comparison group

Fortunately, there is a solution to this huge problem $\rightarrow$ We can "construct" the counterfactual!

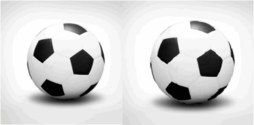
- This is done (usually) selecting a group of individuals that **did not** participate in the program.
- This group is known as the **control group** or **comparison group**.
- The selection of this group is one the most important decisions in the design of an impact evaluation.

Impact Evaluation: The basics

Comparison group

There is no recipe on how to select a control group. However, the principle behind this idea is to select a group that:

- Is exactly like the group of participants in all dimensions but one: the exposure to the program being evaluated.



- In this way, we will be able to attribute differences in outcomes between the group of participants and the comparison group to the program (and not to other factors).

Impact Evaluation: The basics

Causal inference

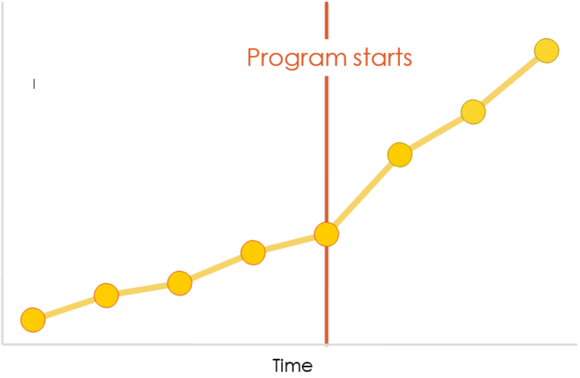
After generating your comparison group, you will be able to compute the **impact** of the program[1]

- The comparison between the outcome after the program has been introduced and the outcome at the same point in time had the program not been introduced (the **counterfactual**!)
- The impact of a program is also known in the literature as **causal effect**.
- Our goal is to estimate the size of this effect accurately and with confidence

[1] In practice, there are many more design considerations that you should take into account, but those topics goes far beyond the scope of the course.

Impact Evaluation: The basics

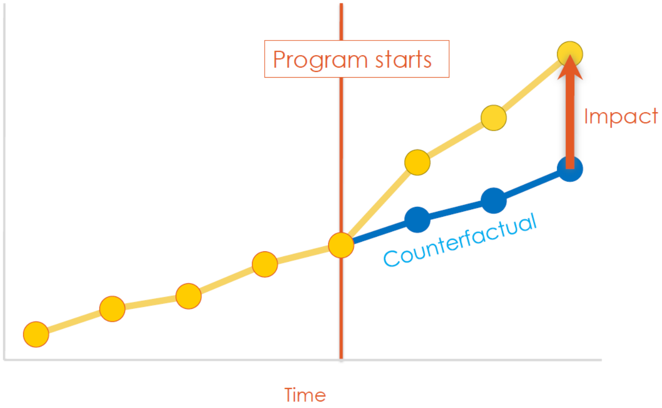
Causal Inference: What is the impact of the program?



Source: JPAL.

Impact Evaluation: The basics

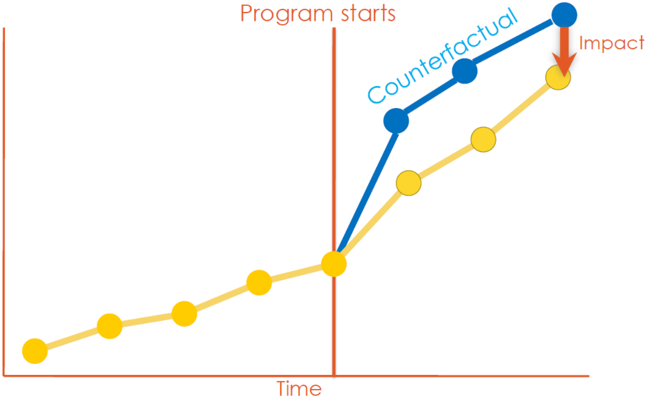
Causal Inference: What is the impact of the program?



Source: JPAL.

Impact Evaluation: The basics

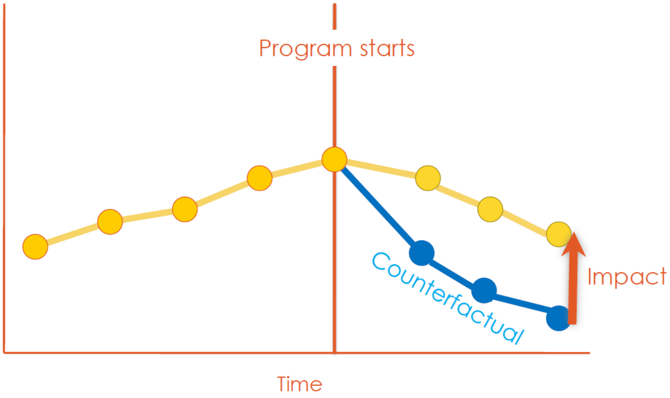
Causal Inference: What is the impact of the program?



Source: JPAL.

Impact Evaluation: The basics

Causal Inference: What is the impact of the program?



Source: JPAL.

The selection problem

Now we are going to introduce some notation to this problem, following Angrist and Pischke (2008). For this, think about a hospital treatment as described by a binary variable $D_i = \{0, 1\}$.

The outcome of interest (a measure of health status in our example) is denoted by Y_i. An the question we would like to study is whether Y_i is affected by hospital care.

The selection problem

To address this question, we **asume** we can imagine what might have happened to someone who went to the hospital if they had not gone and vice versa. Hence, for any individual there are two potential health variables:

$$potential\,outcome \quad = \quad \begin{cases} Y_{1i} & if \quad D_i = 1 \\ Y_{0i} & if \quad D_i = 0 \end{cases}$$

Y_{0i} is the health status of an individual had not gone to the hospital, irrespective of whether he actually went, while Y_{1i} is the individual's health status if the goes.

The selection problem

In this context, we would like to know the difference between Y_{1i} and Y_{0i} $\Rightarrow$ **Causal effect** or the **impact**(in our example, the causal effect of going to the hospital by individual i).

The selection problem

In this context, we would like to know the difference between Y_{1i} and Y_{0i} $\Rightarrow$ **Causal effect** or the **impact**(in our example, the causal effect of going to the hospital by individual i).

In other words, this is what we would measure if we could go back in time and change a person's treatment status. Now, in terms of potential outcomes:

The selection problem

In this context, we would like to know the difference between Y_{1i} and Y_{0i} $\Rightarrow$ **Causal effect** or the **impact**(in our example, the causal effect of going to the hospital by individual i).

In other words, this is what we would measure if we could go back in time and change a person's treatment status. Now, in terms of potential outcomes:

$$\begin{aligned} Y_i &= \begin{cases} Y_{1i} & if \quad D_i = 1 \\ Y_{0i} & if \quad D_i = 0 \end{cases} \\ &= Y_{0i} + (Y_{1i} - Y_{0i})D_i \end{aligned}$$

$*(Y_{1i} - Y_{0i})$ corresponds to the causal effect of hospitalization for an individual.

The selection problem

However...

The selection problem

However...

We never see both potential outcomes for any person $\Rightarrow$ Compare the average value of those who were treated and not treated.

The selection problem

However...

We never see both potential outcomes for any person $\Rightarrow$ Compare the average value of those who were treated and not treated.

But, does a naive comparison of averages tells us everything we would like to know?....

The selection problem

However...

We never see both potential outcomes for any person $\Rightarrow$ Compare the average value of those who were treated and not treated.

But, does a naive comparison of averages tells us everything we would like to know?....Not necessarily!

$$\underbrace{E[Y_i|D_i=1]-E[Y_i|D_i=0]}_{Observed\,difference\,in\,average\,health} = \underbrace{E[Y_{1i}|D_i=1]-E[Y_{0i}|D_i=1]}_{Average\,treatment\,effect\,on\,the\,treated} + \underbrace{E[Y_{0i}|D_i=1]-E[Y_{0i}|D_i=0]}_{Selection\,bias}$$

The selection problem

The *selection bias* term may be so large (in absolute value) $\Rightarrow$ Completely masks a positive treatment effect.

- In our example, the term reflects the difference in average Y_{0i} between those who were and were not hospitalized.
- Because the sick are more likely than healthy to seek treatment, those who were hospitalized have worse Y_{0i}, making selection bias negative.

The selection problem

The *selection bias* term may be so large (in absolute value) $\Rightarrow$ Completely masks a positive treatment effect.

- In our example, the term reflects the difference in average Y_{0i} between those who were and were not hospitalized.
- Because the sick are more likely than healthy to seek treatment, those who were hospitalized have worse Y_{0i}, making selection bias negative.

The selection problem

The *selection bias* term may be so large (in absolute value) $\Rightarrow$ Completely masks a positive treatment effect.

- In our example, the term reflects the difference in average Y_{0i} between those who were and were not hospitalized.
- Because the sick are more likely than healthy to seek treatment, those who were hospitalized have worse Y_{0i}, making selection bias negative.

How can we overcome this issue?....Let's randomize!

References

Angrist, Joshua D and Jorn-Steffen Pischke (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.