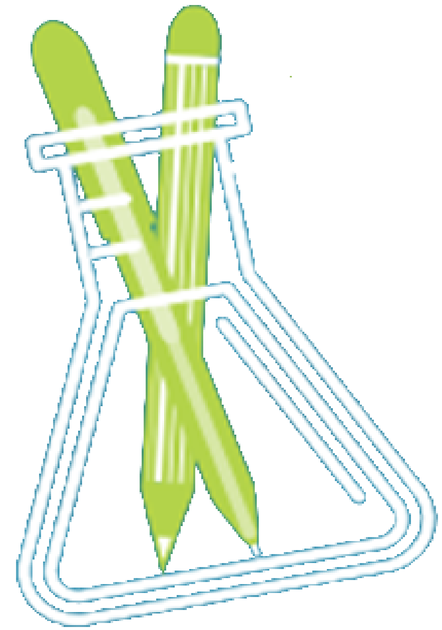
Measuring Impact

Causal Inference
Experimental
non-Experimental
Methods



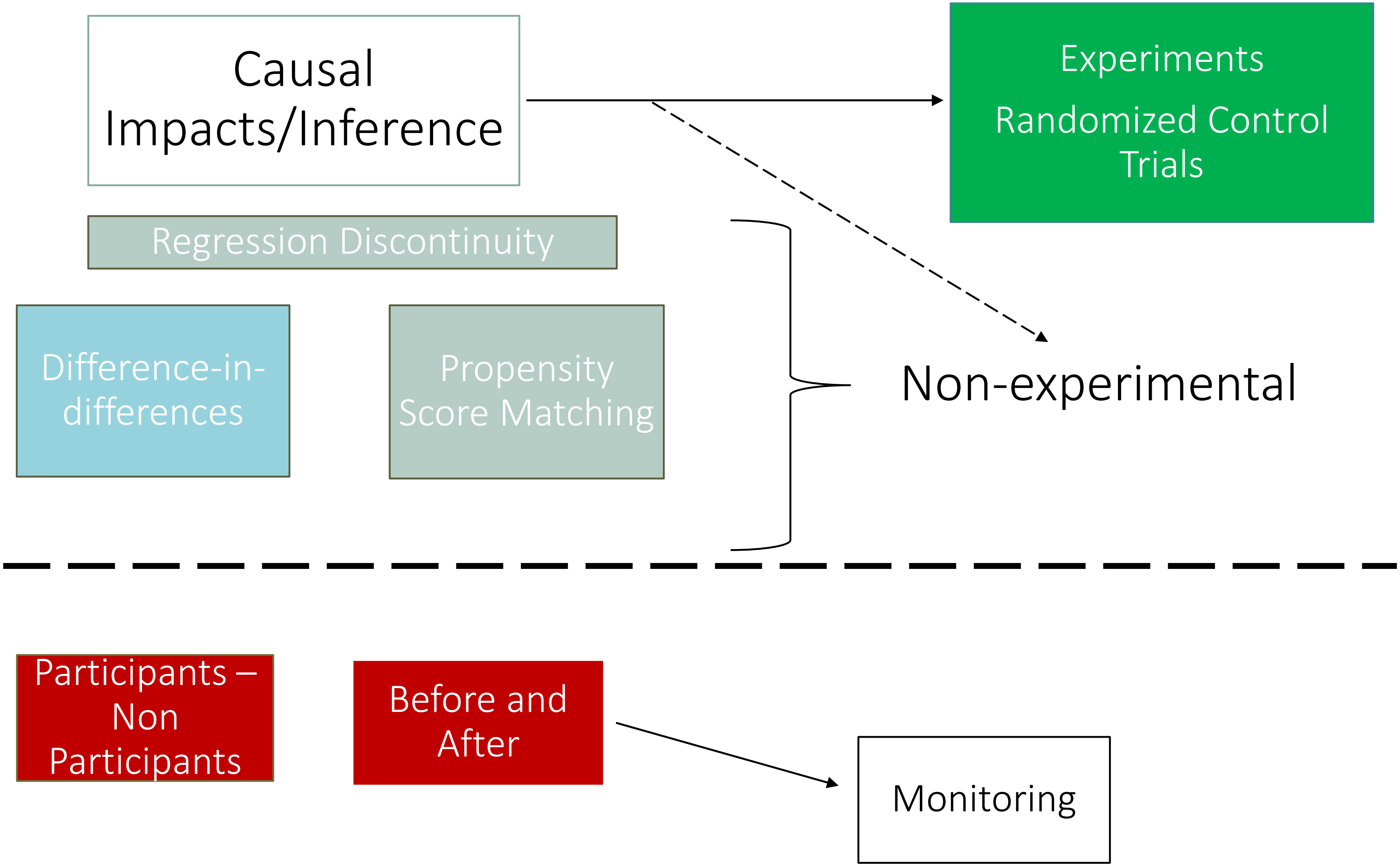
Vincenzo Di Maro, DIME

Bureaucracy Lab

Development Impact Evaluation | Global Governance Practice

October 22-25, 2019, Brussels, Belgium

IMPACT EVALUATION METHODS



Case study: *training program for civil servants - magistrate/judges*

Problem: Case disposition is slow

Intervention: Training. One-week training carried out by expert trainers on the principles and techniques to manage cases efficiently.

Invitation sent to 10000 magistrates

4030 decide to participate, 5970 did not

Main outcome: average case duration (in days)

How we can evaluate this?
Participants – Non Participants

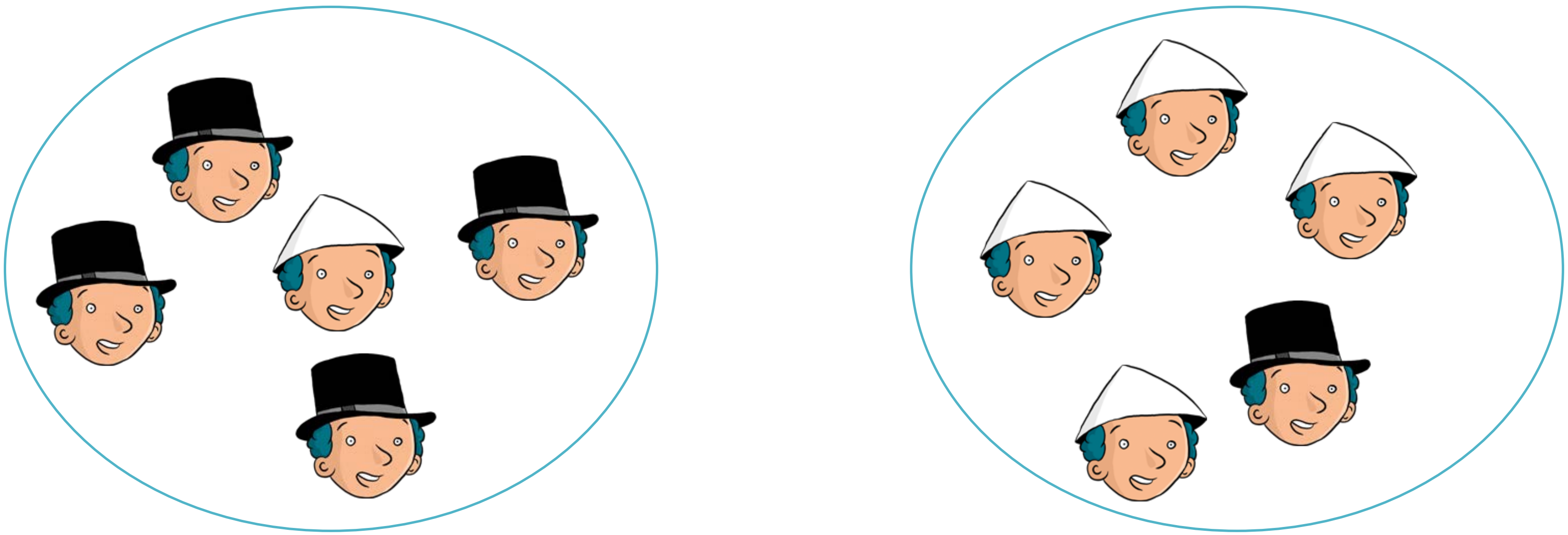
Participation to intervention is voluntary

The **training** was offered to all **10000 magistrates**.

Each magistrate could decide to receive the intervention **(opt in)** or just decline it and **(opt out)**.

Idea: **compare case disposition speed** of magistrates that *opted in* with those that *opted out*

Participants – Non Participants (or Simple difference)



Problem: **Selection Bias.** *Why magistrates opted in?*

- *Better performers or higher capacity (observable)*
- *Stronger motivation (unobservable)*

Parts of this presentation build on material from Impact Evaluation in Practice www.worldbank.org/ieinpractice

How we can evaluate this?

Before - After

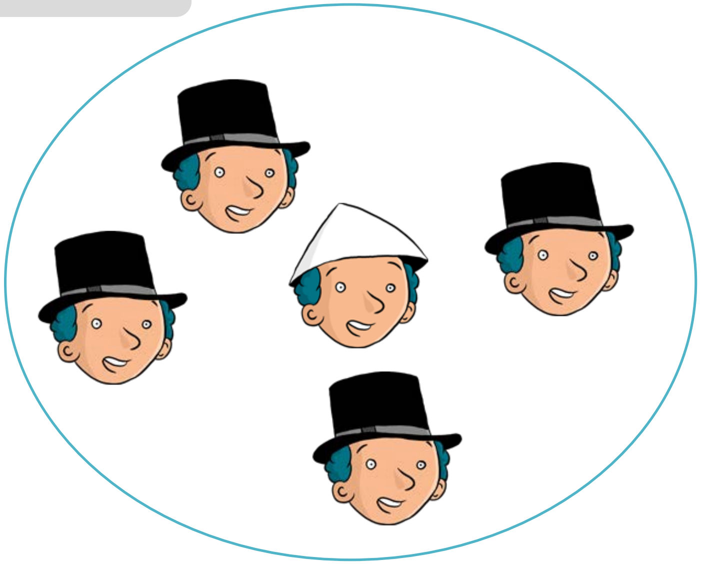
Idea: compare case disposition speed of magistrates that opted in **(treated) after** the intervention started with…

…same magistrates **(control) before** the interventions started

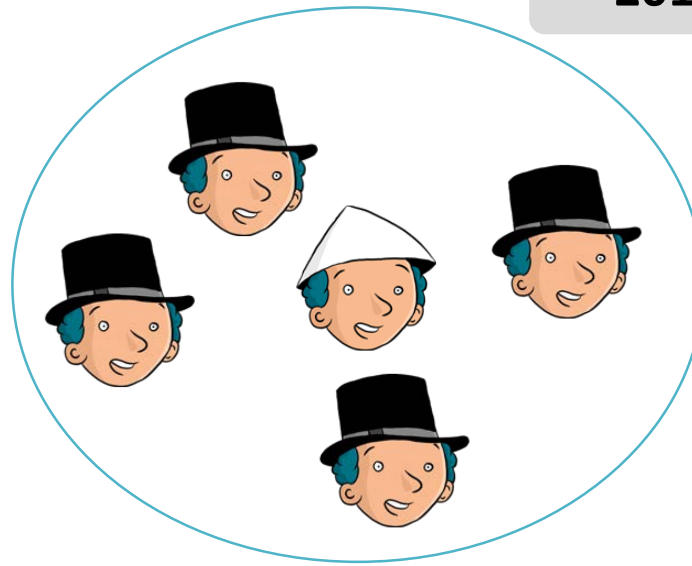
This is: case disposition speed for participants before and after the intervention

Before - After

2016



2018



Problem: **Time difference.** *Other things may have happened over time.*

- *Other programs for treated magistrates*
- *Overall economic conditions got better*

These 2 tools are *wrong* for IE

Before - After

Compare: Same subjects Before and After they receive an intervention.

Problem: Other things may have happened over time.

Participants – Non Participants

Compare: Group of subjects treated (participants) with group that **chooses** not to be treated (non participants)

Problem: Selection Bias. We do not know why they are not participating.

Both counterfactuals (comparison groups) may lead to **biased estimates** of the impact.

→ NOT causal inference

Before-After and Monitoring

Monitoring tracks indicators over time

- Among participants

It is descriptive before-after analysis

It tells us whether things are moving in the right direction

It does not tell us why things happen or how to make more happen

→ NOT causal inference

Impact Evaluation

Tracks mean outcomes over time in
the treatment group *relative to*
the control group

Compares
what DID happen with
what WOULD HAVE happened (counterfactual)

Identifies *cause-effect* link
controlling for ALL other time-varying factors

→ IS causal inference

Experiments

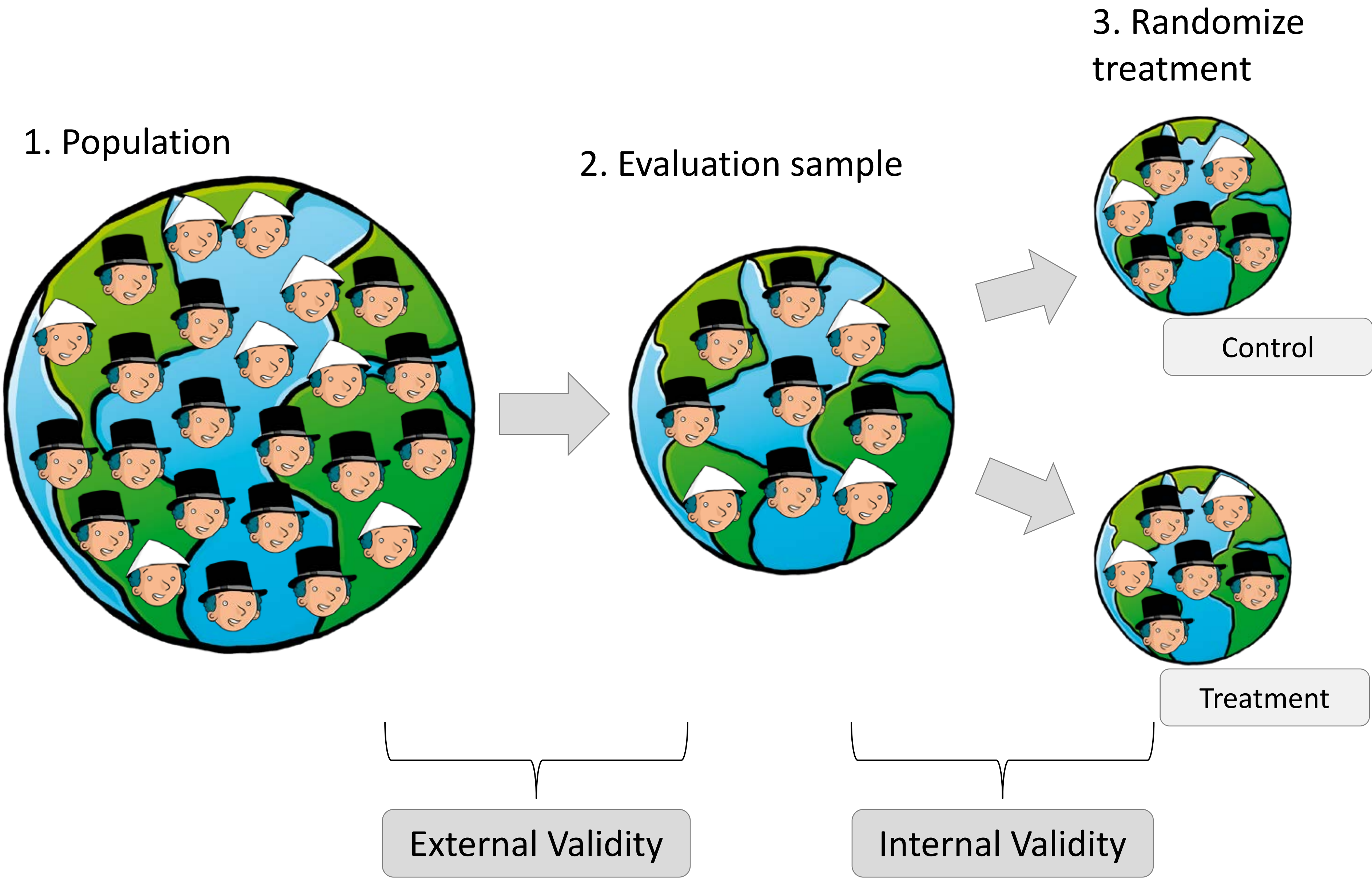
Other names: **Randomized Control Trials (RCTs)** or Randomization

Assignment to Treatment and Control is based on chance, it is random (like flipping a coin)

Treatment and Control groups will have exactly the same characteristics (balanced) at baseline.

Only difference is that treatment receives intervention, control does not

Random assignment

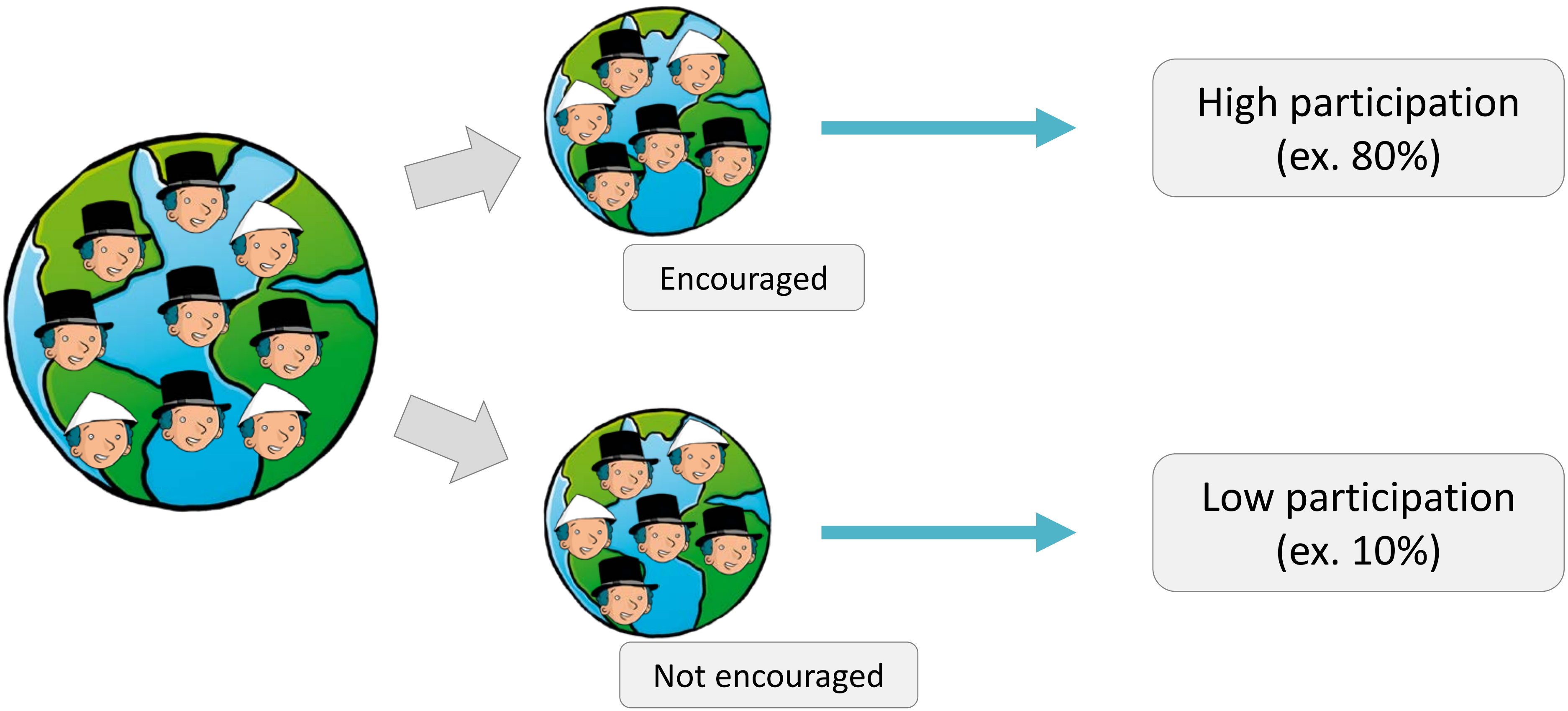


Encouragement design

- Not always possible to randomly assign to control group:
  - Political and ethical reasons
  - Participation is voluntary and all eligible
- Randomized promotion/encouragement
  - program available to everyone
  - But provide additional promotion, encouragement or incentives to a random sub-sample:
    - Additional Information.
    - Incentives (small gift or prize).
    - Transport (bus fare).

Encouragement design

Randomize Incentive to
participate. Ex. small gifts



NON-EXPERIMENTAL METHODS

Experiments are not always feasible:

- Implementation issues
- Political issues

Still possible to estimate impact rigorously

- Treatment and control/comparison assignment is **not random**
- Non-experimental approaches are valid but **only if we use the right method**
- **Some methods are not valid** for impact evaluation.

How we can evaluate this?

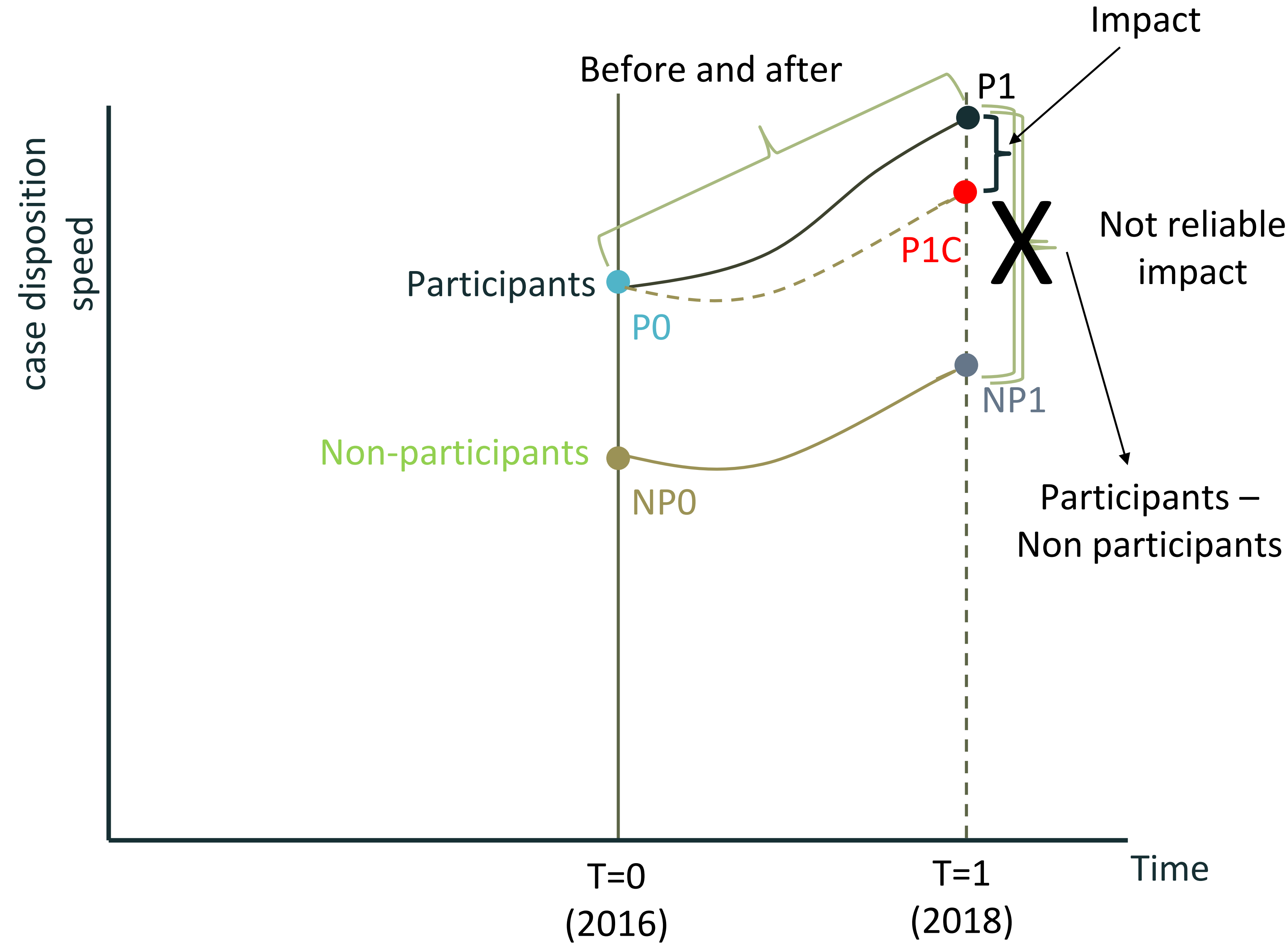
Difference-in-Differences

Idea: combine the **time dimension** (*before-after of control*) with the **participation choice** (*participants-non participants*)

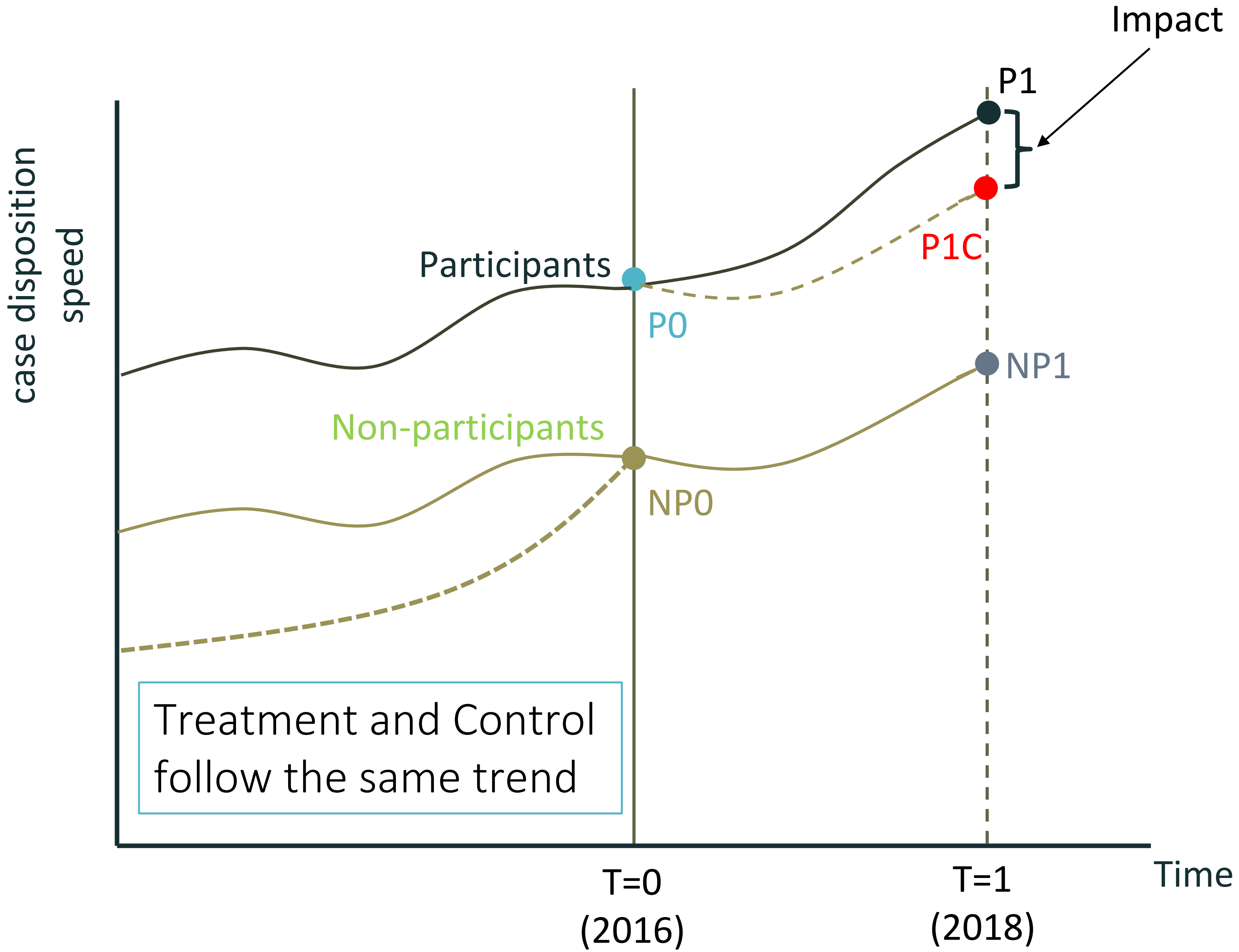
(under some assumptions) this deals with the problems above:

- **Time differences.** Other things may have happened over time.
- **Selection Bias.** We do not know why they are not participating.

Difference-in-Differences



Assumptions for Diff-in-Diff



Difference-in-Differences

Difference-in-differences combines *Participation decision* with *Time dimension.*

It deals with problems of previous methods under the…

…fundamental assumption
Trends –slopes- are the same in treatments and controls

Possible to test if you have data pre-treatment

Deals with unobservables only if constant over time

Improve diff-in-diff if you match groups based on observable characteristics (propensity score matching)

IMPACT EVALUATION METHODS