



# Ex post evaluation – Impact evaluation



## Why is ex post evaluation important?



## Lecture content

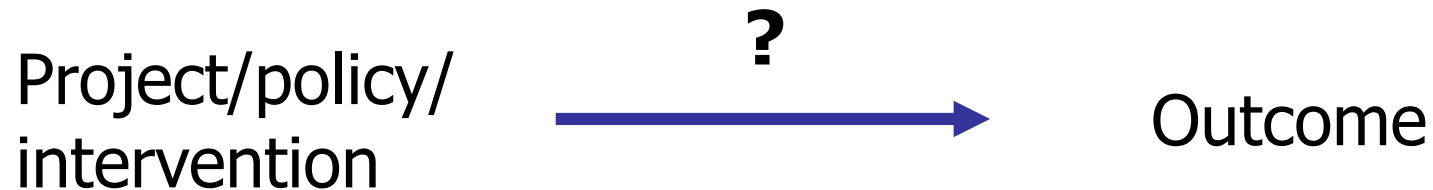
- Impact evaluations
- Quasi-experimental methods
- Difference-in-differences
- Propensity score matching
- Regression discontinuity design
- Examples

## Impact evaluation

*An impact evaluation assesses changes in the well-being of individuals, households, communities or firms that can be attributed to a particular project, program or policy. The central impact evaluation question is what would have happened to those receiving the intervention if they had not in fact received the program (World Bank, 2008).*



# Problem of impact evaluation: How do we estimate impact of the intervention?



## Impact evaluation

- Key feature - the establishment of a counterfactual
- Counterfactual – what the outcome would have been for programme participants if they had not participated in the programme
- To serve as a reliable counterfactual, members of the control group should be identical to those in the treatment group participating in the intervention programme



## Randomized control trials

- Subjects are randomly placed into a treatment and control group
- Outcomes are evaluated comparing the outcomes of both groups
- No selection bias when dividing subjects into the control and treatment groups
- Both observed and unobserved characteristics between the groups are balanced

**BUT:** What if we did not select the subjects randomly into the group before the intervention?

## Quasi experimental methods

- Mostly used when it is not possible to do randomized control trials
- Identification of a comparison group among the non treated units as similar as possible to the treatment group in terms of preintervention characteristics
- The comparison group captures what would have been the outcomes if the programme/policy had not been implemented



## Quasi experimental methods

- the outcome is approximated
- not assumption free (contrary to the randomized experiments)
- Rely on assumptions which cannot be tested
- REMARK: No intervention cost dimension in the design
  
- Difference-in-differences
- Propensity score matching
- Regression discontinuity design
- Instrumental variables (not covered in the lecture)

## Difference-in-differences (1)

- Simple outcome difference between treated and non-treated group does not reveal the true effect of the intervention
- Group difference already observed before the intervention  
->selection bias
- DiD controls for this selection bias by constructing the difference observed between the two groups before intervention from the difference observed after the intervention

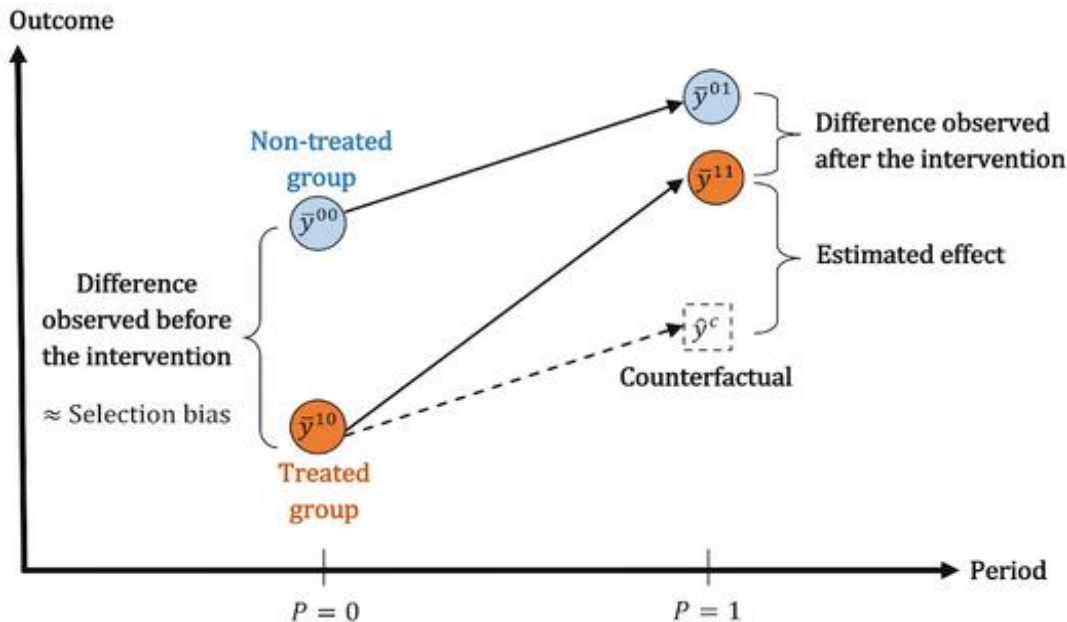
## Difference-in-differences (2)

- ASSUMPTION:
  - Without any intervention, the trend of the treated group would have been similar to that of the non-treated.
  - Counterfactual LEVELS for the treated and the non-treated can be different.

## Difference-in-differences (3)

- The first difference ( $\Delta$  after/before) controls for factors that are constant over time in that group, since we are comparing the same group to itself
  
- The second difference ( $\Delta$  treated/non-treated) measures the before-and-after change in outcomes for a group that did not enrol in the program but was exposed to the same set of environmental conditions – for capturing the time-varying factors
  - > elimination of the main source of bias that worried us in the simple before-and-after comparisons

	Before intervention P=0	After intervention P=1	$\Delta$ after/before
Non-treated group S=0	$\bar{y}^{00}$	$\bar{y}^{01}$	$\bar{y}^{01} - \bar{y}^{00}$
Treated group S=1	$\bar{y}^{10}$	$\bar{y}^{11}$	$\bar{y}^{11} - \bar{y}^{10}$
$\Delta$ treated/non-treated	$\bar{y}^{10} - \bar{y}^{00}$	$\bar{y}^{11} - \bar{y}^{01}$	$\bar{y}^{11} + \bar{y}^{00} - \bar{y}^{01} - \bar{y}^{10}$



## Example

- 20 municipalities
- Problem: Introduction of a new drug to decrease the mortality rate in municipalities
- 9 municipalities exposed to the treatment, 11 without the treatment
- Data of the mortality rate for each municipality before and after the treatment

Municipality	S	P	Mortality rate
1	0	0	15
1	0	1	14
2	0	0	16
2	0	1	15
⋮	⋮	⋮	⋮
10	0	0	8
10	0	1	7
11	0	0	10
11	0	1	9
12	1	0	19
12	1	1	15
13	1	0	22
13	1	1	18
⋮	⋮	⋮	⋮
19	1	0	21
19	1	1	18
20	1	0	20
20	1	1	17

## How to proceed (1):

1. Mean of the MR for the non-treated group before intervention

Municipality	S	P	Mortality rate
1	0	0	15
1	0	1	14
2	0	0	16
2	0	1	15
...	...	...	...
10	0	0	8
10	0	1	7
11	0	0	10
11	0	1	9
12	1	0	19
12	1	1	15
13	1	0	22
13	1	1	18
...	...	...	...
19	1	0	21
19	1	1	18
20	1	0	20
20	1	1	17

## How to proceed (1):

1. Mean of the MR for the non-treated group before intervention

Mortality	P=0	P=1
S=0	13.73	
S=1		



Municipality	S	P	Mortality rate
1	0	0	15
1	0	1	14
2	0	0	16
2	0	1	15
...	...	...	...
10	0	0	8
10	0	1	7
11	0	0	10
11	0	1	9
12	1	0	19
12	1	1	15
13	1	0	22
13	1	1	18
...	...	...	...
19	1	0	21
19	1	1	18
20	1	0	20
20	1	1	17

## How to proceed (1):

1. Mean of the MR for the non-treated group before intervention
2. Mean of the MR for the non-treated group after intervention

Mortality	P=0	P=1
S=0	13.73	
S=1		

Municipality	S	P	Mortality rate
1	0	0	15
1	0	1	14
2	0	0	16
2	0	1	15
...	...	...	...
10	0	0	8
10	0	1	7
11	0	0	10
11	0	1	9
12	1	0	19
12	1	1	15
13	1	0	22
13	1	1	18
...	...	...	...
19	1	0	21
19	1	1	18
20	1	0	20
20	1	1	17

## How to proceed (1):

1. Mean of the MR for the non-treated group before intervention
2. Mean of the MR for the non-treated group after intervention

Mortality	P=0	P=1
S=0	13.73	
S=1		

Municipality	S	P	Mortality rate
1	0	0	15
1	0	1	14
2	0	0	16
2	0	1	15
...	...	...	...
10	0	0	8
10	0	1	7
11	0	0	10
11	0	1	9
12	1	0	19
12	1	1	15
13	1	0	22
13	1	1	18
...	...	...	...
19	1	0	21
19	1	1	18
20	1	0	20
20	1	1	17

## How to proceed (1):

1. Mean of the MR for the non-treated group before intervention
2. Mean of the MR for the non-treated group after intervention

Mortality	P=0	P=1
S=0	13.73	12.73
S=1		

Municipality	S	P	Mortality rate
1	0	0	15
1	0	1	14
2	0	0	16
2	0	1	15
...	...	...	...
10	0	0	8
10	0	1	7
11	0	0	10
11	0	1	9
12	1	0	19
12	1	1	15
13	1	0	22
13	1	1	18
...	...	...	...
19	1	0	21
19	1	1	18
20	1	0	20
20	1	1	17

## How to proceed (1):

1. Mean of the MR for the non-treated group before intervention
2. Mean of the MR for the non-treated group after intervention
3. Mean of the MR for the treated group before intervention

Mortality	P=0	P=1
S=0	13.73	12.73
S=1		

Municipality	S	P	Mortality rate
1	0	0	15
1	0	1	14
2	0	0	16
2	0	1	15
...	...	...	...
10	0	0	8
10	0	1	7
11	0	0	10
11	0	1	9
12	1	0	19
12	1	1	15
13	1	0	22
13	1	1	18
...	...	...	...
19	1	0	21
19	1	1	18
20	1	0	20
20	1	1	17

## How to proceed (1):

1. Mean of the MR for the non-treated group before intervention
2. Mean of the MR for the non-treated group after intervention
3. Mean of the MR for the treated group before intervention

Mortality	P=0	P=1
S=0	13.73	12.73
S=1		

Municipality	S	P	Mortality rate
1	0	0	15
1	0	1	14
2	0	0	16
2	0	1	15
...	...	...	...
10	0	0	8
10	0	1	7
11	0	0	10
11	0	1	9
12	1	0	19
12	1	1	15
13	1	0	22
13	1	1	18
...	...	...	...
19	1	0	21
19	1	1	18
20	1	0	20
20	1	1	17

## How to proceed (1):

1. Mean of the MR for the non-treated group before intervention
2. Mean of the MR for the non-treated group after intervention
3. Mean of the MR for the treated group before intervention

Mortality	P=0	P=1
S=0	13.73	12.73
S=1	20.44	

Municipality	S	P	Mortality rate
1	0	0	15
1	0	1	14
2	0	0	16
2	0	1	15
...	...	...	...
10	0	0	8
10	0	1	7
11	0	0	10
11	0	1	9
12	1	0	19
12	1	1	15
13	1	0	22
13	1	1	18
...	...	...	...
19	1	0	21
19	1	1	18
20	1	0	20
20	1	1	17

## How to proceed (1):

1. Mean of the MR for the non-treated group before intervention
2. Mean of the MR for the non-treated group after intervention
3. Mean of the MR for the treated group before intervention
4. Mean of the MR for the treated group before after intervention

Mortality	P=0	P=1
S=0	13.73	12.73
S=1	20.44	

Municipality	S	P	Mortality rate
1	0	0	15
1	0	1	14
2	0	0	16
2	0	1	15
...	...	...	...
10	0	0	8
10	0	1	7
11	0	0	10
11	0	1	9
12	1	0	19
12	1	1	15
13	1	0	22
13	1	1	18
...	...	...	...
19	1	0	21
19	1	1	18
20	1	0	20
20	1	1	17

## How to proceed (1):

1. Mean of the MR for the non-treated group before intervention
2. Mean of the MR for the non-treated group after intervention
3. Mean of the MR for the treated group before intervention
4. Mean of the MR for the treated group before after intervention

Mortality	P=0	P=1
S=0	13.73	12.73
S=1	20.44	



Municipality	S	P	Mortality rate
1	0	0	15
1	0	1	14
2	0	0	16
2	0	1	15
...	...	...	...
10	0	0	8
10	0	1	7
11	0	0	10
11	0	1	9
12	1	0	19
12	1	1	15
13	1	0	22
13	1	1	18
...	...	...	...
19	1	0	21
19	1	1	18
20	1	0	20
20	1	1	17

## How to proceed (1):

1. Mean of the MR for the non-treated group before intervention
2. Mean of the MR for the non-treated group after intervention
3. Mean of the MR for the treated group before intervention
4. Mean of the MR for the treated group before after intervention

Mortality	P=0	P=1
S=0	13.73	12.73
S=1	20.44	17.22

## How to proceed (2):

5. The difference of the MR in the non-treated group before and after the treatment

Mortality	P=0	P=1	$\Delta$ after/before
S=0	13.73	12.73	
S=1	20.44	17.22	
$\Delta$ treated/non-treated			

## How to proceed (2):

5. The difference of the MR in the non-treated group before and after the treatment

Mortality	P=0	P=1	$\Delta$ after/before
S=0	13.73	12.73	-1.00
S=1	20.44	17.22	
$\Delta$ treated/non-treated			

## How to proceed (2):

5. The difference of the MR in the non-treated group before and after the treatment
6. The difference of the MR in the treated group before and after the treatment

Mortality	P=0	P=1	$\Delta$ after/before
S=0	13.73	12.73	-1.00
S=1	20.44	17.22	
$\Delta$ treated/non-treated			

## How to proceed (2):

5. The difference of the MR in the non-treated group before and after the treatment
6. The difference of the MR in the treated group before and after the treatment

Mortality	P=0	P=1	$\Delta$ after/before
S=0	13.73	12.73	-1.00
S=1	20.44	17.22	-3.22
$\Delta$ treated/non-treated			

## How to proceed (2):

5. The difference of the MR in the non-treated group before and after the treatment
6. The difference of the MR in the treated group before and after the treatment
7. The overall difference (treated and non-treated groups)

Mortality	P=0	P=1	$\Delta$ after/before
S=0	13.73	12.73	-1.00
S=1	20.44	17.22	-3.22
$\Delta$ treated/non-treated			

## How to proceed (2):

5. The difference of the MR in the non-treated group before and after the treatment
6. The difference of the MR in the treated group before and after the treatment
7. The overall difference (treated and non-treated groups)

Mortality	P=0	P=1	$\Delta$ after/before
S=0	13.73	12.73	-1.00
S=1	20.44	17.22	-3.22
$\Delta$ treated/non-treated			-2.22

## How to proceed (2):

5. The difference of the MR in the non-treated group before and after the treatment
6. The difference of the MR in the treated group before and after the treatment
7. The overall difference (treated and non-treated groups)

The same outcome is obtained by differentiating the differences between the groups

Mortality	P=0	P=1	$\Delta$ after/before
S=0	13.73	12.73	-1.00
S=1	20.44	17.22	-3.22
$\Delta$ treated/non-treated			-2.22



## How to proceed (2):

5. The difference of the MR in the non-treated group before and after the treatment
6. The difference of the MR in the treated group before and after the treatment
7. The overall difference (treated and non-treated groups)

The same outcome is obtained by differentiating the differences between the groups

Mortality	P=0	P=1	$\Delta$ after/before
S=0	13.73	12.73	-1.00
S=1	20.44	17.22	-3.22
$\Delta$ treated/non-treated	6.71	4.49	-2.22

## Remarks

- Does not necessary require large set of data
- A good approach to calculate a quantitative impact estimate, but this method alone is not usually enough to address selection bias
- If the assumption about the parallel trend of both groups is violated, the DiD method alone would not provide an accurate assessment of the impact

## Propensity score matching

- Methods that match units from treated and non treated groups that have identical (similar) observable characteristics, except of the fact of receiving the intervention (mimicing the randomization)
- BUT: units may differ in more than one variable
- Instead, the **propensity score** is used - the likelihood that the individual will participate in the intervention (predicted likelihood of participation) given their observable characteristics
- Difference in mean outcome between treated and control group = the estimated impact of intervention
- PSM ensures that the average characteristics of the treatment and comparison groups are similar (sufficient for unbiased impact estimation)
- ASSUMPTION: all factors that affect the outcome are observed

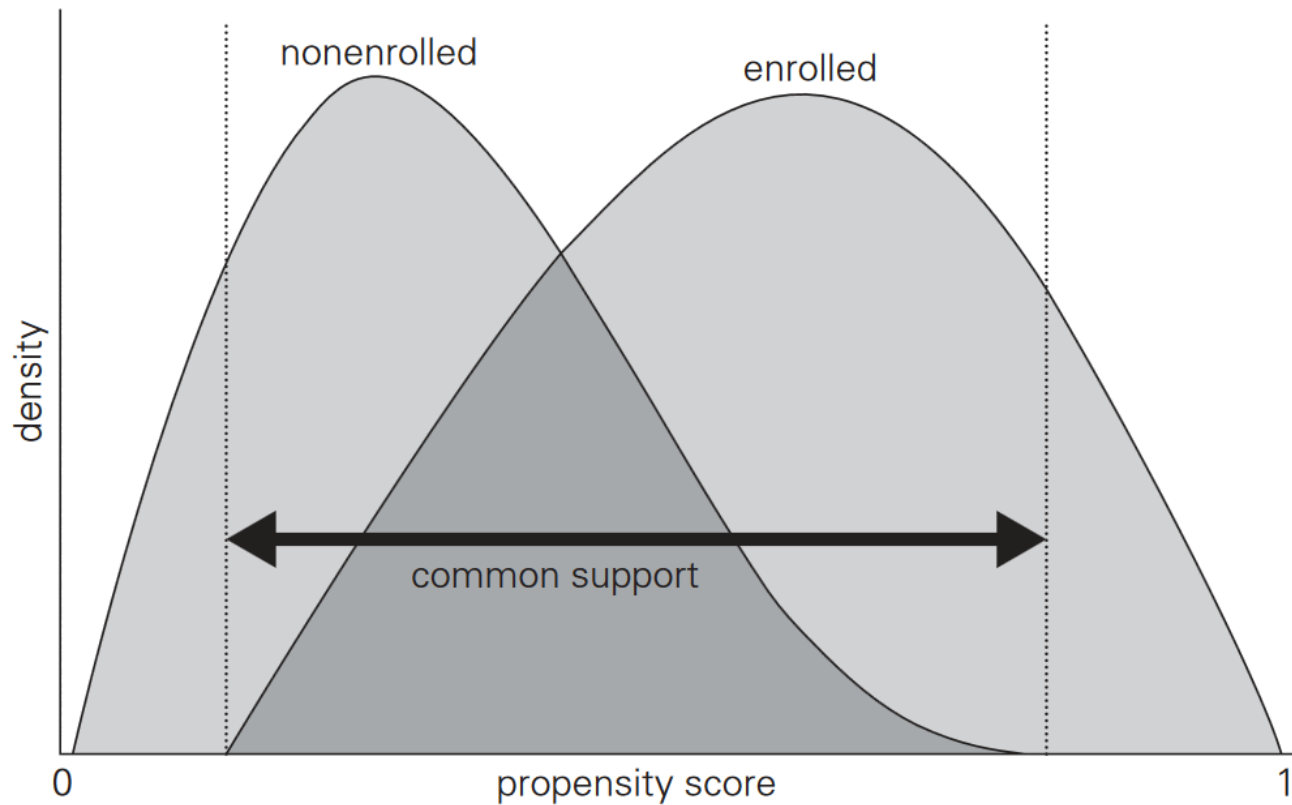
## PSM - technique

### 1. Computation of the propensity score:

- Calculation of unit's probability of being exposed to the intervention (by logistic regression)
- The probability is conditional on set of observable individual characteristics that may affect participation in the program

### 2. Sample restriction according to the propensity score distribution

# PSM – sample restriction



Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B., & Vermeersch, C. M. (2016). *Impact evaluation in practice*. World Bank Publications, pp. 110.

## PSM - technique

3. Matching of treated and controlled units based on their propensity score
4. By differentiating the average outcomes of matched treated and controlled units, the average treatment effect is estimated.

## Matching principle

Whether one unit can be matched once or multiple times:

- matching with replacement – each control unit can be matched to several treated obs.
- matching without replacement – each control unit is used no more than a one time

Methods of pairing the treated and control units:

- Nearest neighbour – treatment unit is matched with the closest controlled unit
- Calliper matching – standardized distance acceptable for any match

## Measurement of the intervention effect

- **Sample treatment effect for the treated group (ATT)**
  - Focus is on the treated units
  - Difference of the average outcome of the treated units and the average outcome of the matched controlled units
- **Sample average treatment effect for the controlled group (ATC)**
  - Units from controlled group are matched to their nearest neighbour treated units
  - Difference of the average outcome of the matched treated units and the average outcome of the controlled units
- **Average treatment effect (ATE)**
  - Combine the previous two approaches

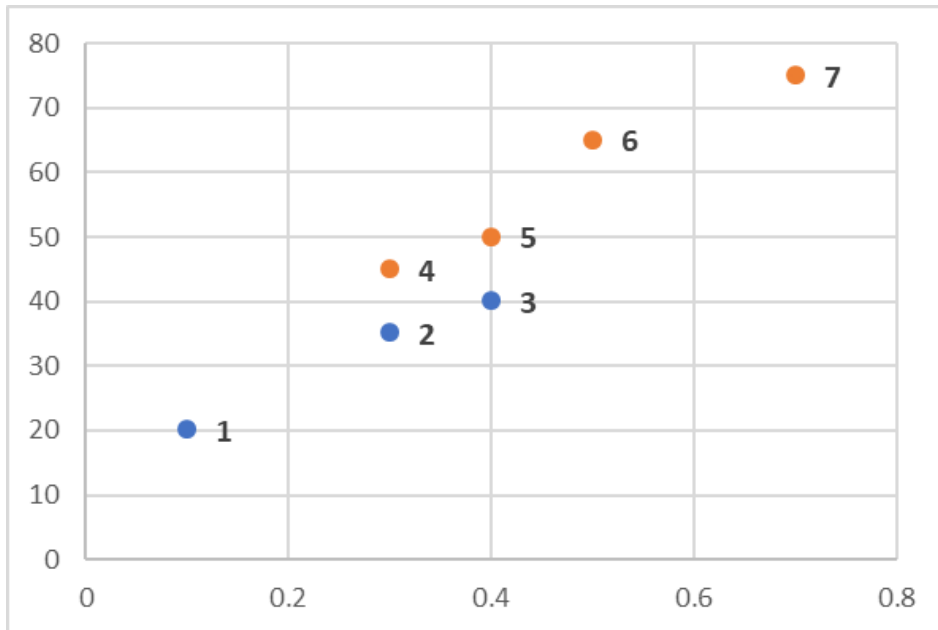


## Example

- 7 patients
- Intervention: Drug that releases from pain
- 4 patients participated, 3 did not
- Data: Number of hours that patient does not feel the pain

Unit	S	Outcome (Y)	PSM
1	0	20	0.1
2	0	35	0.3
3	0	40	0.4
4	1	45	0.3
5	1	50	0.4
6	1	65	0.5
7	1	75	0.7

## The matching process



1. Matching the treated units with the controlled units

$4 - 2, 5 - 3, 6 - 3, 7 - 3$

2. Matching the controlled units with the treated units

$1 - 4, 2 - 4, 3 - 5$

Unit	S	Outcome (Y)	PSM
1	0	20	0.1
2	0	35	0.3
3	0	40	0.4
4	1	45	0.3
5	1	50	0.4
6	1	65	0.5
7	1	75	0.7

Treated units paired with the controlled units

4 – 2, 5 – 3, 6 – 3, 7 – 3

Controlled units paired with the treated units

1 – 4, 2 – 4, 3 – 5

$$\mathbf{ATT} = (45+50+65+75)/4 - (35+40+40+40)/4 = 20$$

$$\mathbf{ATC} = (45+45+50)/3 - (20+35+40)/3 = 15$$

$$\mathbf{ATE} = ((45+50+65+75) + (45+45+50))/7 - ((35+40+40+40) + (20+30+40))/7 = 18,57$$

## Remarks

- Always feasible if data are available
- The assumption that no selection bias has occurred stemming from unobserved characteristics is very strong, and most problematic, it cannot be tested
- Any variable that is thought to influence the exposure of the intervention and the outcome should be included – identification problem

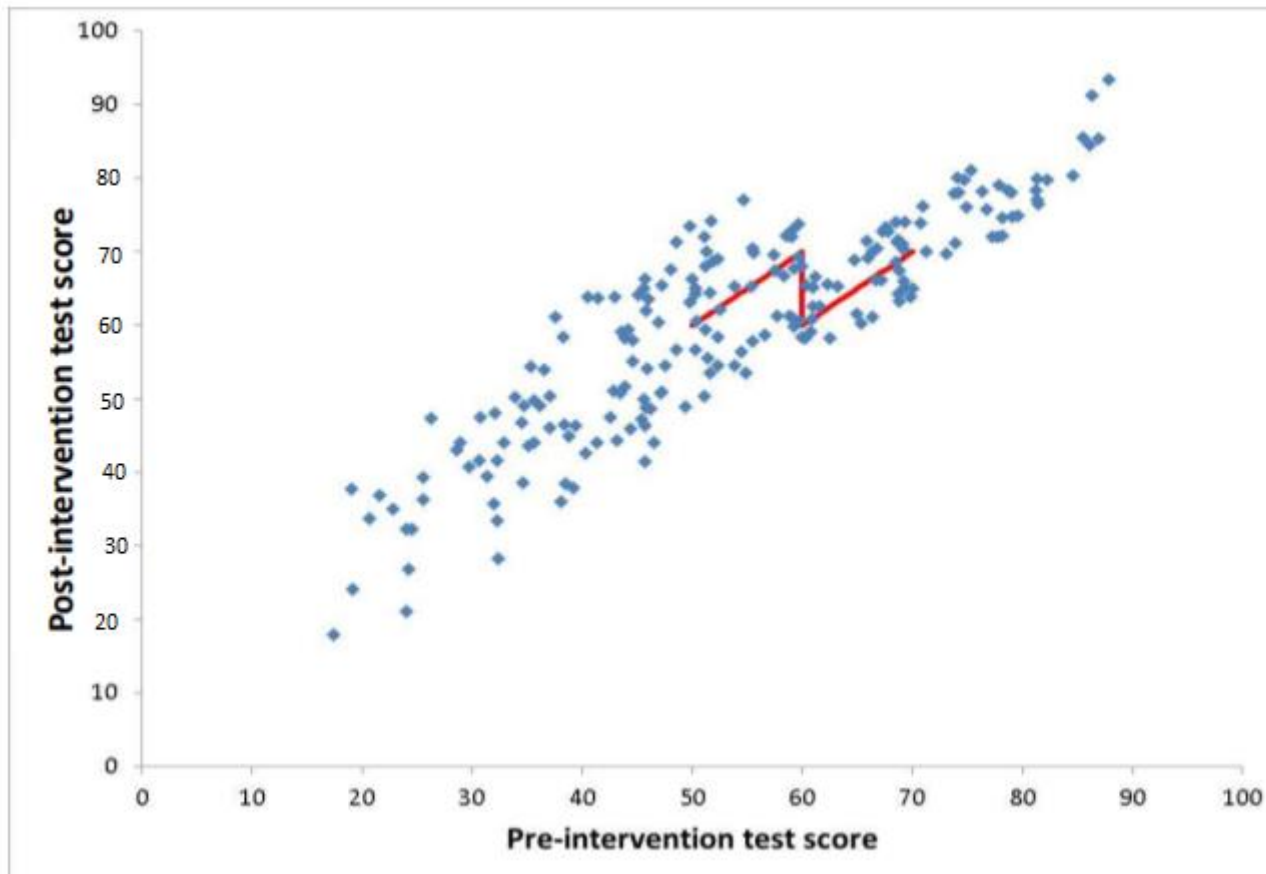
## Regression discontinuity design (RDD)

- When there is some kind of criterion that must be met before people can participate in the intervention (a threshold)
  - F.e: students below a certain test score are enrolled in a remedial programme; women above a certain age are eligible for participation in a health programme; central government makes funds available for municipalities with less than five thousands inhabitants..
- A comparison of treatment and controlled units around a threshold (above and below), which intervention is dispensed
- ASSUMPTION: considering observations lying close to either side of the threshold, the selection bias should be eliminated

## RDD application

1. Threshold definition
2. To determine the margin around the threshold
  - A small margin can be set up, and the resulting treatment and comparison groups can be tested for their balance or similarity
3. Once the sample is established, a regression line is fitted

# Regression discontinuity design



White, H., & S. Sabarwal (2014). Quasi-experimental Design and Methods, *Methodological Briefs: Impact Evaluation 8*, UNICEF Office of Research, Florence.

## Remarks

- Deals with non-observable characteristics more convincingly than other quasi-experimental methods
- ASSUMPTION: Unit cannot manipulate their treatment status (cannot influence the treatment status)
- The status manipulation may produce biased estimates
- The impact estimate is valid for those close to the threshold, but the impact on those further from the threshold may be different
- However, from previous researches, the difference is not great -> RDD acceptable method for estimating the effects of a programme or policy.



## Activity

- In groups, discuss a current problem (in Czech, in your country, anywhere) that you think should be solved (can be a simple problem)
- Formulate an intervention that you would use in order to solve/reduce the problem
- After the intervention is done, you would like to estimate the impact of your intervention/programme/project
- Choose what is the indicator that could be used to assess the impact
- Think about the factors that may influence your outcome and therefore should be considered in the estimation

## Sources:

Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B., & Vermeersch, C. M. (2016). *Impact evaluation in practice*. World Bank Publications.

Grant, T., & K. Crowland (2017). *A Practical Guide to Getting Started with Propensity Scores*. Data & Information Management Enhancement (DIME) Kaiser Permanente.

Josselin, J. M., & Le Maux, B. (2017). *Statistical Tools for Program Evaluation: Methods and Applications to Economic Policy, Public Health, and Education*. Springer.

White, H., & S. Sabarwal (2014). Quasi-experimental Design and Methods, *Methodological Briefs: Impact Evaluation 8*, UNICEF Office of Research, Florence.

White, H., Raitzer, D.A. (2017). *Impact Evaluation of Development Interventions A Practical Guide*. Asian Development Bank.