



Stockholm
University

Pär Nyman

Methods in Political Science
27 April 2026

Evaluating experiments

Identification assumptions and potential outcomes

Introduction

1. The potential outcomes framework
2. Introducing the EITC example
3. Identification assumptions (90% of the lecture!)
 - Random assignment
 - Non-interference
 - Excludability
 - Compliance
4. Wrapping up

All examples include one treatment group and one control group, but same logic can be extended to multiple treatment groups.

Potential outcomes

- Potential outcome = what the outcome would be given different treatment conditions.
- You do not need to know more than what a potential outcome is, but the framework is quite common.
- Notation differs, but basic idea:
 - y_i^0 is the outcome for subject i if not treated.
 - y_i^1 is the outcome for subject i if treated.
 - D_i indicates whether subject i was treated (1) or not (0).

Potential outcomes

- For each subject we either observe y_i^0 or y_i^1 .
- The individual treatment effect $y_i^1 - y_i^0$ is not observable.
- Average treatment effect is estimated by comparing y_i^1 for subjects in the treatment group with y_i^0 in the control group.
- Often by just calculating the difference in means between the treated and the controls.

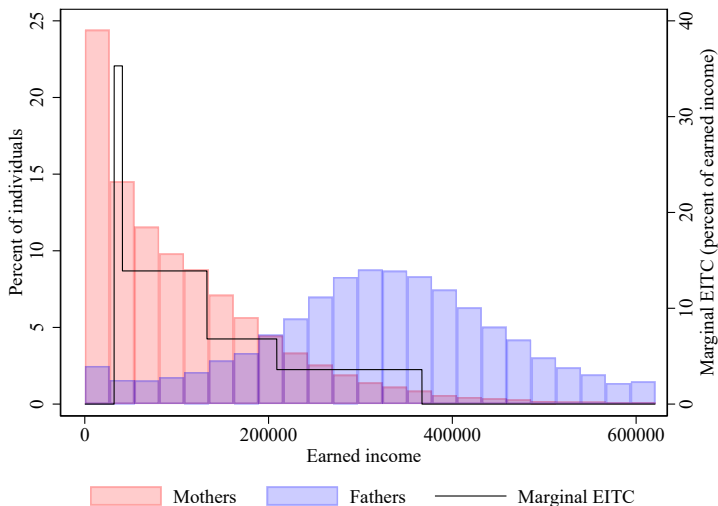
Example

Empirical example for illustration

- 'Jobbskatteavdraget' (EITC) is a tax cut on earned income.
- The annualized nature of taxation + the non-linear nature of the EITC makes it particularly beneficial to work part-time or a few months per year (earn 135,000 kr in 2018).
- Ideal setting for an information experiment on parents.
 - Large incentives for shorter or more equal leave.
 - Unknown incentives. I have never met someone who thought about this when planning the parental leave.
 - Testable. These individuals *can* decide to work more.

Example

Incentives – marginal EITC + income distribution, year after birth



Example

The EITC experiment summarized

- In 2018 we sent an information letter to 37,500 randomly selected couples who recently got a kid (the equally sized control group did not receive any letter).
- The letter informed about how the Swedish EITC reduces the marginal tax rate for mothers on parental leave.
- To estimate the effects of an experiment like this, it is enough to compare the mean of some outcome between the treatment and the control group.
- We found no effect on labour supply or the distribution of parental leave, and speculate that parents act as *satisficers*.

Assumptions

The difference-in-means is an unbiased estimator of the treatment effect under the following assumptions:

- Random assignment: same old exogeneity assumption.
- Non-interference: treatment of one unit does not affect the potential outcomes of other units.
- Excludability: the treatment assignment only affects the outcome through the treatment itself.
- Compliance: treatment is administered to the treatment group and not to the control group.

Random assignment

- Random assignment means that it is random whether a person is assigned to the treatment or the control group.
- It is how we meet the exogeneity assumption:
 - In expectation, potential outcomes do not differ between treatment and control group.
 - No omitted variables or reverse causation if x is random.
- Probability of treatment can differ as long as it is known (so we can control for it).
- Has nothing to do with random sampling.
- Nothing at all.

Random assignment

How is it done?

- For each observation, we could just roll a die or use a random number function on the computer, and compare the outcome to a threshold (i.e., 1–3 control, 4–6 treatment).
- But to get an optimal number of subjects in each group, the standard procedure is as follows:
 1. Generate a random number for each subject.
 2. Sort the list according to this number.
 3. Assign the top to treatment and the bottom to control.

Random assignment

Group-based randomization

- Clustered randomization: Randomly assign entire groups of units to treatment or control.
 - Useful when treatment spillover is expected within clusters.
 - Important to cluster standard errors.
- Block randomization: Randomly assign treatment within blocks of units sharing similar characteristics.
 - Ensures balance between treatment and control groups for key characteristics.

Random assignment

Evaluating randomization in natural experiments

- Controlled experiments have random assignment unless someone made a mistake.
- In natural experiments it is something we must investigate:
 - Understanding the assignment process: interviewing decision-makers, reading documents, replicating the assignment, etc.
 - Discussing potential violations. Be creative!
- Treatment assignment does not need to be completely random as long as it is unrelated to potential outcomes.

Random assignment

Selection effects could violate random assignment

- Sometimes we observe a sub-sample of everyone who was once treated (or not).
- It is not a problem if selection into this sample is affected by treatment status or potential outcome alone.
- But potential outcomes would be related to treatment status if potential outcome affected selection in a way conditioned by treatment status (i.e., if the treatment made some people more likely to drop out than others).

Random assignment

Balance and placebo tests

- We can also evaluate the randomization by looking at observable differences between the treatment and control group on characteristics unaffected by the treatment.
- Balance tests examine whether treated and controls are similar in their baseline characteristics.
- Placebo outcomes are outcomes that should not be affected by the treatment.
- Placebo treatments are treatments that should not affect the outcome.

Random assignment

Statistical differences/effects do not necessarily mean that something is wrong

- We should expect some significant differences if we make a lot of tests, even if assignment was random.
- Sometimes we even get a very unbalanced sample.
- But it would probably make us doubt your as-if random natural experiment.

Random assignment

Control variables in experiments

- Sometimes control variables are used to adjust for unbalanced samples.
- I always include controls in my preferred specification.
 - Increase precision (because uncorrelated with x).
 - No need to add because of a failed balance test.
- If treatment probabilities differ you need to control for that!

Random assignment

Variable	Control	Treat	Dif	p
<i>Variables related to the child</i>				
Female	0.49	0.48	−0.00	0.44
<i>Variables related to the mother</i>				
Years of education	13.08	13.08	−0.00	0.97
Potential income (thousands)	280.93	280.65	−0.28	0.78
Year of birth	1986.23	1986.24	0.01	0.78
Employed 2017	0.81	0.82	0.00	0.79
Previous parental leave	128.97	128.73	−0.24	0.84
Missing prev. par. leave	0.49	0.49	0.00	0.65
<i>Variables related to the father</i>				
Years of education	12.68	12.65	−0.03	0.11
Potential income (thousands)	408.68	410.13	1.45	0.43
Year of birth	1983.22	1983.28	0.06	0.19
Employed 2017	0.92	0.92	0.00	0.92
Previous parental leave	29.08	29.76	0.69	0.12
Missing prev. par. leave	0.50	0.50	0.00	0.93

Non-interference

- Treatment of one unit does not affect the potential outcomes of other units.
- If it does, then we cannot separate the treatment effect from how the treatment affected others, unless the latter effect is identical for treated and controls.
- Also referred to as the stable unit treatment value assumption (SUTVA).

Non-interference

Violations of non-interference

- The treatment is transmitted from one subject to another.
- Subjects are aware of what treatments others receive and react on that.
- Resources to the treatment are taken from subjects that would otherwise have received them.

Non-interference

We discussed several sources of non-interference

- Did people discuss our letter, transmitting the information from treated to controls?
- The non-conformist newspaper Världen idag wrote an article about the experiment. Did that information reach our research subjects?
- Difficult to rule out, but arguably a minor problem?

Excludability

Excludability assumption

- The treatment assignment only affects the outcome through the treatment itself. No other pathways between treatment assignment and the outcome.
- If violated, we can still estimate the effect of being assigned to treatment, but we cannot separate the effect of our treatment from the effect of the other pathway.
- Also known as the exclusion restriction (especially for IV).

Excludability

Violations of excludability

- Subjects are aware of their treatment status and act on that information.
- Subjects in the treatment group are given some additional treatment.
- Different research assistants gather data about the treatment and the control group.

Excludability

Excludability in the EITC experiment

- Not a problem in our study if we define the treatment as receiving the letter (nothing else happened).
- More problematic if we define it more narrowly as an increase in information about the tax system.
 - The letter may signal pro-work or pro-equality norms.
 - The letter may have initiated a discussion.
- The substantive problem cannot be solved by the treatment definition; only how we label it.

Compliance

- Compliers sometimes refer to subjects in the treatment group that receive the treatment, and those in the control group that do not (the others are non-compliers).
- Sometimes we distinguish between four types of behaviour:
 - Compliers: Follow their assigned treatment
 - Always-takers: Take the treatment regardless of assignment
 - Never-takers: Refuse the treatment regardless of assignment
 - Defiers: Do the opposite of their assigned treatment
- If there is non-compliance we must adjust for that when estimating the treatment effect.

Compliance

Compliance in the EITC experiment

- If we think about the treatment as receiving the letter, then we have close to perfect compliance.
 - Statistics Sweden have up-to-date addresses and PostNord are more reliable than their reputation says.
 - People in the control group are unlikely to receive the letter.
- If we think about the treatment as reading the letter, and understanding the information, then it is more complicated.
- We had no way of quantifying the latter kind of compliance, but argue for small effects also under low compliance.

Compliance

- Under non-compliance we must distinguish between different treatment effects.
- The average treatment effect (ATE) if everyone was treated.
- Intent-to-treat effect (ITT) is the effect of being assigned to treatment.
- Local average treatment effect (LATE) or complier average causal effect (CACE): the treatment effect for compliers.
- Not obvious which one that is most interesting.

Compliance

Estimate the LATE/CACE

- The proportion that receives the treatment is p larger in the treatment group than in the control group.
- Then the expected difference between treated and controls is $p \times LATE$.
- If we divide the observed difference by p we get our estimate of LATE.

Compliance

Illustration with one-sided non-compliance

- Assume that you ran a GOTV experiment with 50 percent compliance in the treatment group and a treatment effect of 4 percentage points.
- In expectation, that would increase voter turnout with 2 percentage points (0.5×4) in the treatment group.
- We could retrieve the treatment effect by dividing the difference in turnout with the compliance rate ($2/0.5 = 4$)
- Only true in expectation. Actual difference between treatment and control can be smaller or larger than 2.

Compliance

Illustration with two-sided non-compliance

- In addition, if 25 percent were non-compliers in the control group, their expected turnout would increase by 1 p.p.
- We could divide the difference in turnout ($2 - 1 = 1$) with the difference in treatment rates ($0.5 - 0.25 = 0.25$) to retrieve the treatment effect ($1/0.25 = 4$).
- Only true in expectation.
- By dividing the difference with a small number, we are inflating the sampling errors (still unbiased, but bad for precision).

Compliance

Estimate the LATE/CACE

- If we want to use a regression framework this is equivalent to an IV regression of the following form:

$$\begin{aligned} Outcome &= a_1 + b_1 \widehat{Treatment} + e \\ \widehat{Treatment} &= a_2 + b_2 Assignment \end{aligned}$$

- You do not need to see that, but we will try it out at the workshop.

Compliance

We should never remove the non-compliers from the analysis.

- Compliance is not randomly assigned and the compliers probably have other potential outcomes than the non-compliers.
- People who spend a lot of time at home, who open the door to strangers or who are good friends with the canvassers were more(?) likely to vote even before the treatment.
- Unless we can do it before randomization, or in extreme cases based on some observable characteristic that is balanced between treated and controls.

Wrapping up

Conducting the experiment

- Focus has been on evaluating the experiments, but these are also the things to keep in mind when conducting them.
- Can you identify likely non-compliers and remove them before the random assignment?
- Can you increase compliance through incentives, reminders, or monitoring?
- How can you define/assign the treatment to avoid spillovers and excludability violations?
- My experience: dialogue is key in field experiments.

Wrapping up

Pre-registration

- If you are serious about conducting an experiment, it could make sense to pre-register your hypothesis and analysis.
- The idea is to avoid data mining, i.e., to search for statistically significant relationships with a large probability of finding false positives.
- Most good journals expect experiments to be pre-registered.

Wrapping up

Thank you for today!