# Chapter 2. Social Experiments (Randomized Control Trials and Natural Experiments)

Joan Llull

Quantitative Statistical Methods II
Barcelona GSE

# *Randomized Experiments*

In the treatment effect approach, a **randomized field trial** is regarded as the ideal design.

Long **history** of randomized field trials in social welfare in the U.S., beginning in the 1960s (see Moffitt (2003) for a review).

**Encouraged** by U.S. Federal Government, eventually almost mandatory. Legislation introduced in 1988.

**Resistance** from many states on ethical grounds (more so in other countries, where treatment groups are often areas for treatment instead of individuals).

Sometimes experiments are provided by nature: **natural experiments** (e.g. John Snow and the cholera case in SoHo).

# *Random Assignment and Treatment Effects*

In a controlled experiment, treatment status is **randomly assigned** by the researcher, which by construction, ensures **independence**:

$$Y_{1i}, Y_{0i} \perp\!\!\!\perp D_i.$$

This **eliminates the selection bias** (see Chapter 1), and implies $\alpha_{TT} = \beta$, as:

$$\mathbb{E}[Y_{0i}|D_i = 1] = \mathbb{E}[Y_{0i}|D_i = 0] = \mathbb{E}[Y_{0i}].$$

Also $\alpha_{ATE} = \alpha_{TT} = \beta$, as $\mathbb{E}[Y_{1i} - Y_{0i}|D_i = 1] = \mathbb{E}[Y_{1i} - Y_{0i}]$.

Thus, the average treatment effect can be estimated by a simple **linear regression** of the observed outcome $Y_i$ on the treatment dummy $D_i$ and a constant.

# *Standard Errors and Inference*

The standard default options for regression in statistical packages assume that residuals $U_i$ are **homoskedastic**. This implies:

$$\mathrm{Var}(Y_i|D_i = 1) = \mathrm{Var}(Y_i|D_i = 0) = \mathrm{Var}(U_i) = \mathrm{Var}(Y_{0i}) = \mathrm{Var}(Y_{1i}).$$

In the context of **heterogeneous** treatment effects, this is often violated.

$\mathrm{Var}(Y_{1i})$ can be expressed as:

$$\mathrm{Var}(Y_{1i}) = \mathrm{Var}(Y_{0i}) + \mathrm{Var}(Y_{1i} - Y_{0i}) + 2\,\mathrm{Cov}(Y_{1i} - Y_{0i}, Y_{0i}),$$

which is obtained by noting that $\mathrm{Var}(Y_{1i}) = \mathrm{Var}(Y_{0i} + (Y_{1i} - Y_{0i}))$.

**Homogeneous treatment effects** $\Rightarrow$ $Y_{1i} - Y_{0i}$ is a constant $\Rightarrow$ $\mathrm{Var}(Y_{1i} - Y_{0i}) = \mathrm{Cov}(Y_{1i} - Y_{0i}, Y_{0i}) = 0$.

**Heterogeneous treatment effects** $\Rightarrow$ not necessarily the case (e.g. effect of lottery ticket on wealth).

# *Standard Errors and Inference*

Standard error of the ATE is the **variance of the difference in means**:

$$\text{Var}(\beta^S) = \text{Var}(\bar{Y}_T - \bar{Y}_C) = \text{Var}(\bar{Y}_T) + \text{Var}(\bar{Y}_C) = \frac{\sigma_T^2}{N_1} + \frac{\sigma_C^2}{N_0},$$

where $\sigma_T^2$ and $\sigma_C^2$ are sample variances of the outcome computed on treated and control subsamples, and $N_1$ and $N_0$ are the sizes of each subsample.

Alternatively, from the regression we can compute **robust standard errors**:

$$\text{Var}\begin{pmatrix}\hat{\beta}_0\\ \hat{\beta}\end{pmatrix} = \frac{1}{N}\,\mathbb{E}[X_i X_i']^{-1}\,\mathbb{E}[X_i X_i' \hat{U}_i^2]\,\mathbb{E}[X_i X_i']^{-1},$$

where $X_i \equiv (1, D_i)'$. The **sample analog** is:

$$\begin{aligned}\widehat{\text{Var}}\begin{pmatrix}\hat{\beta}_0\\ \hat{\beta}\end{pmatrix} &= \frac{1}{N}\left(\frac{1}{N}\sum_{i=1}^{N} X_i X_i'\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} X_i X_i' U_i^2\right)\left(\frac{1}{N}\sum_{i=1}^{N} X_i X_i'\right)^{-1}\\ &= \begin{pmatrix}\frac{\sum_{i:D_i=0}\hat{U}_i^2}{N_0^2} & -\frac{\sum_{i:D_i=0}\hat{U}_i^2}{N_0^2}\\ -\frac{\sum_{i:D_i=0}\hat{U}_i^2}{N_0^2} & \frac{\sum_{i:D_i=0}\hat{U}_i^2}{N_0^2} + \frac{\sum_{i:D_i=1}\hat{U}_i^2}{N_1^2}\end{pmatrix}.\end{aligned}$$

Maybe observations are **not independent** $\Rightarrow$ **clustering** (e.g. Progresa).

## *Introduction of Additional Regressors*

Additional regressors $W_i$ are not needed for **consistency** as:

$$\gamma \frac{\mathrm{Cov}(W_i, D_i)}{\mathrm{Var}(D_i)} = 0.$$

Yet, it can be interesting to include them for several reasons:

- If they are relevant, they can increase **precision** (Frisch-Waugh Theorem).
- **Checking randomization**: are there statistical difference in these regressors between treated and controls?
- **Used in the randomization** (e.g. village-level randomization).

The last two lead to the context of **conditional independence**.

# ***Warning: Partial Complience***

So far we have assumed **perfect complience**: everyone elected takes the treatment and no control takes it.

Now: $D_i = \mathbb{1}\{\text{treatment taken}\}$ and $Z_i = \mathbb{1}\{\text{assigned to treatment}\}$.

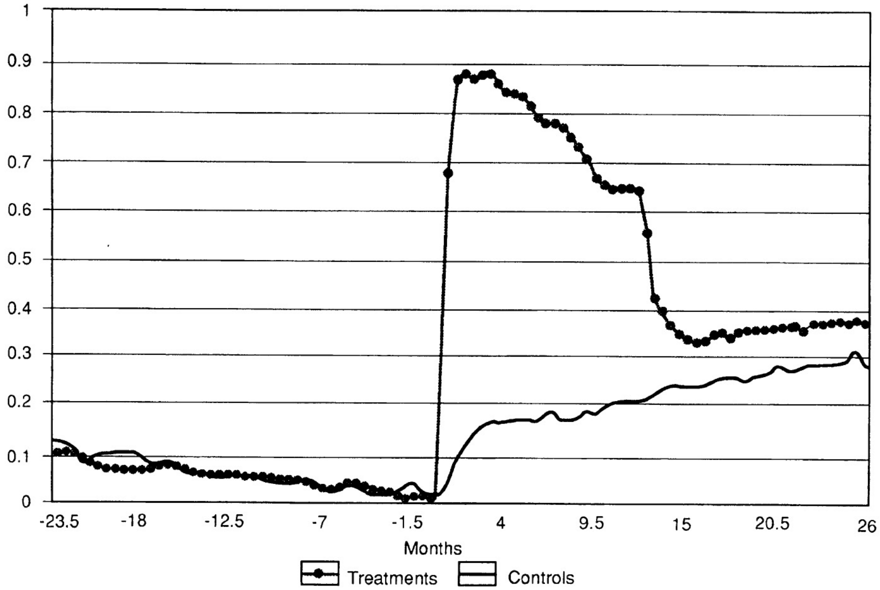
We may have $D_i = 0$ and $Z_i = 1$ (**no-shows**), and $D_i = 1$ and $Z_i = 0$ (**cross-overs**).

Now $Y_{1i}, Y_{0i} \not\perp\!\!\!\perp D_i$ but $Y_{1i}, Y_{0i} \perp\!\!\!\perp Z_i$. The latter can be used in an **IV fashion** to obtain $\alpha_{TT}$ (Chapter 4), or compute an **intention-to-treat** effect.

# *Warning: Longer Run Outcomes*

**National Supported Work program (NSW)**:

- designed in the U.S. in the mid 1970s
- training and job opportunities to disadvantaged workers
- NSW guaranteed to treated participants 12 months of subsidized employment (as trainees) in jobs with gradual increase in work standards.
- experimental design on women who volunteered for training
- Requirements: unemployed, a long-term AFDC recipient, and have no preschool children
- Participants were randomly assigned to treatment (275) and control groups (266) in 1976-1977
- Training in 1976, and then followed.
- Ham and LaLonde (1996) analyze the effects of the program.


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-23.5
-18
-12.5
-7
-1.5
4
9.5
15
20.5
26
Months
Treatments
Controls

# *Effects on Unemployment Rates*

Thanks to randomization, comparison of employment rates for treated and control gives an **unbiased estimate** of the effect of the program on employment at different horizons.

Initially, by construction there is a mechanical effect from the fact that treated women are offered a **subsidized job**.

Compliance with the treatment is decreasing over time, as women can decide to **drop from the subsidized job**.

The **employment growth for controls** is a reflection of the program's eligibility criteria.

Importantly, after the program ends, a **9 percentage points difference** in employment rates is sustained.

## *Ham and LaLonde's Additional Point*

But Ham and LaLonde (1996) make an important additional point: randomization **does not guarantee independence** for any possible outcomes.

**Two examples**: wages and unemployment durations (hazards).

Effect of training program on employment rates of the treated $\Rightarrow$ those who are working are a **selected sample**.

**Notation**: $W_i$ wages; $Y_i = 1$ if employed; $\eta_i = 1$ skilled type.

Suppose:

$$P(Y_i = 1|D_i = 1, \eta_i = j) > P(Y_i = 1|D_i = 0, \eta_i = j), \quad j = 0, 1$$

and:

$$\frac{P(Y_i = 1|D_i = 1, \eta_i = 0)}{P(Y_i = 1|D_i = 0, \eta_i = 0)} > \frac{P(Y_i = 1|D_i = 1, \eta_i = 1)}{P(Y_i = 1|D_i = 0, \eta_i = 1)}.$$

This implies that the **frequency of low skill** will be greater in the group of employed treatments than in the employed controls:

$$P(\eta_i = 0|Y_i = 1, D_i = 1) > P(\eta_i = 0|Y_i = 1, D_i = 0),$$

which is a way to say that $\eta_i$, which is unobserved, is **not independent** of $D_i$ given $Y_i = 1$, although, unconditionally, $\eta_i \perp\!\!\!\perp D_i$.

Consider the **conditional effects**:

$$\Delta_j \equiv \mathbb{E}[W|Y_i = 1, D_i = 1, \eta_i = j] - \mathbb{E}[W_i|Y_i = 1, D_i = 0, \eta_i = j], \quad j = 0, 1$$

Our effect of interest is:

$$\Delta_{ATE} = \Delta_0 P(\eta_i = 0) + \Delta_1 P(\eta_i = 1),$$

and comparison of average wages between treatments and controls is:

$$\Delta_W = \mathbb{E}[W_i|Y_i = 1, D_i = 1] - \mathbb{E}[W_i|Y_i = 1, D_i = 0] < \Delta_{ATE}.$$

$\Rightarrow$ may not be possible to correctly measure the effect on wages.