Lecture 6: When Complete Randomization Fails

Attrition after random assignment

Xi Chen

Rotterdam School of Management
Erasmus University Rotterdam

June 3, 2023

Outline

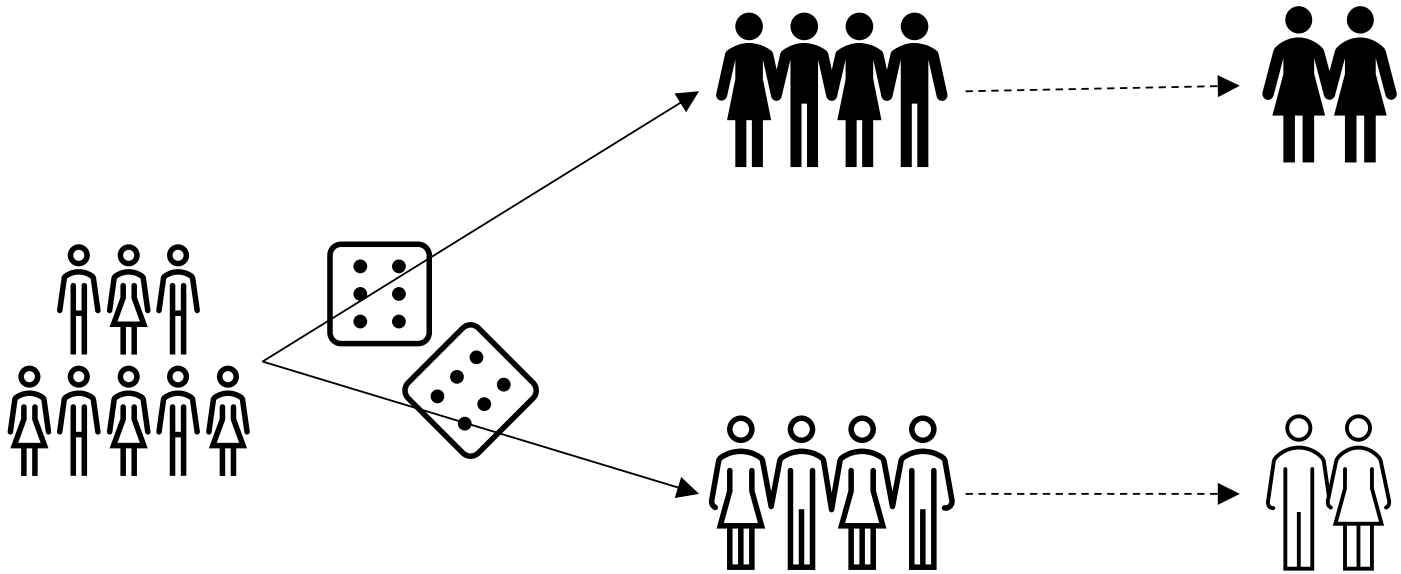
RSM Erasmus

Outline

RSM Erasmus

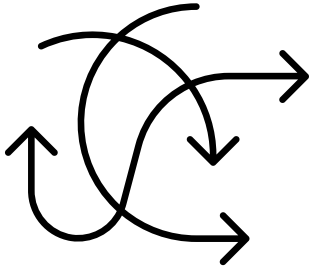
Failure to measure outcomes

The general idea of attrition: the failure to measure outcomes of some units.

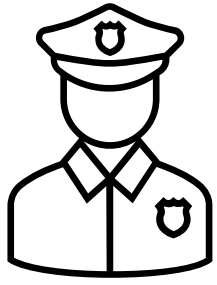


Various reasons of attrition

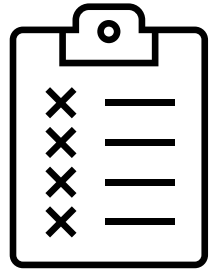
Refuse to cooperate
(two-stage studies)



Lose track of subjects
(long-term studies)



Blocked by authorities
(field experiments)



Unmeasurable
(e.g., decease)

RSM Erasmus

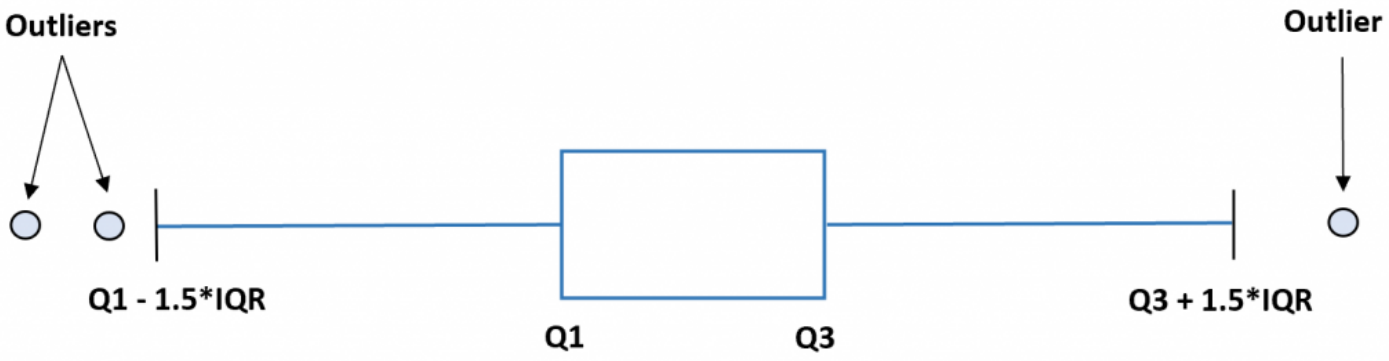
Attrition: an example

The introduction of e-commerce to countryside in China[1].

Attrition: **local governments denied access to some villages.**

[1] Please check Couture et al. (2021) for more details.

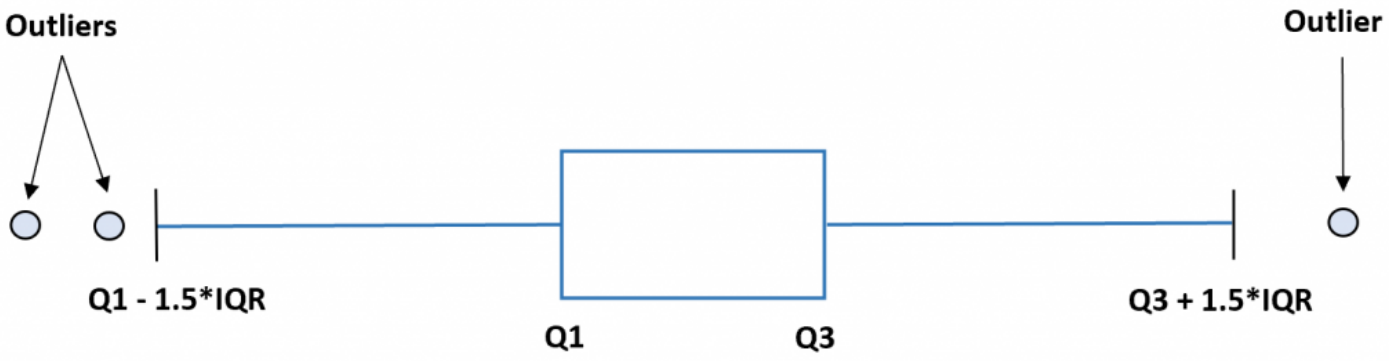
Attrition: the deletion of outliers in experimental data



Should we delete outliers in experimental data?

- Deletion leads to missing outcomes for some units.
- Leading to the failure of randomization, eventually non-identification of ATE.

Attrition: the deletion of outliers in experimental data

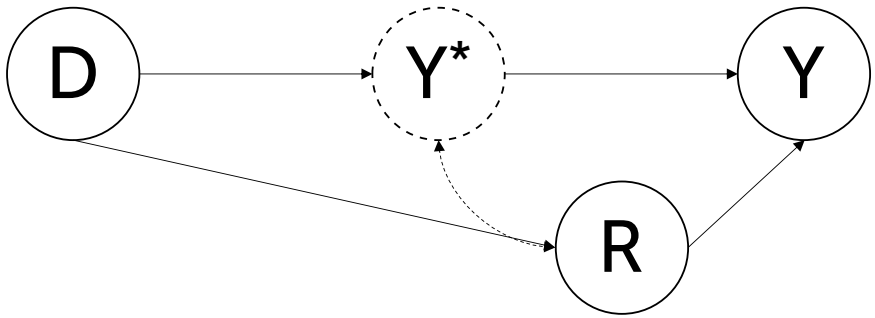


Should we delete outliers in experimental data?

- Deletion leads to missing outcomes for some units.
- Leading to the failure of randomization, eventually non-identification of ATE.

RSM Erasmus

A DAG representation of attrition



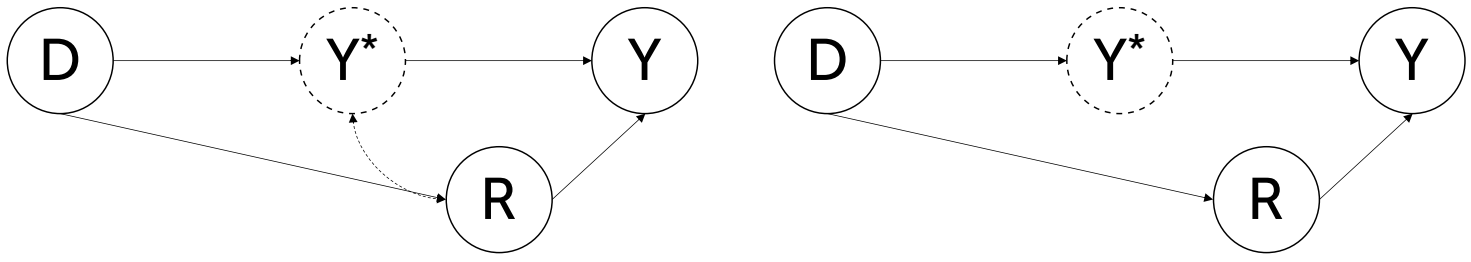
- D is the treatment, and **exogenous due to randomization**.
- Y^* is a latent outcome, partially observed by researchers.
- $R = \{0, 1\}$ is the binary state of attrition, caused by or causing Y^*.
- Y is the observed outcome with $Y = (1 - R) \times Y^*$

Fact (Intention-to-treat)

Without any assumption, we can identify the intention-to-treat effects with $P(Y \mid D)$.

Traditional approaches: Random attrition

Assuming the attrition is random...



(a) DAG of Non-random Attrition (b) DAG of Random Attrition

Fact (Random Attrition)

Under random attrition, we can identify the average treatment effect with $P(Y^* \mid D) = P(Y^* \mid D, R)$.

Traditional approaches: Random attrition

Technical proof:

$$
\begin{aligned}
P\left(Y^* \mid do\left(D\right)\right) &= P\left(Y^* \mid D, R=0\right) P\left(R=0\right) + P\left(Y^* \mid D, R=1\right) P\left(R=1\right) \\
&\overset{Y^* \perp R \mid D}{=} P\left(Y^* \mid D\right) P\left(R=0\right) + P\left(Y^* \mid D\right) P\left(R=1\right) \\
&= P\left(Y^* \mid D\right)
\end{aligned}
$$

Because the attrition is at random, we can use observed outcome Y to calculate the conditional distribution:

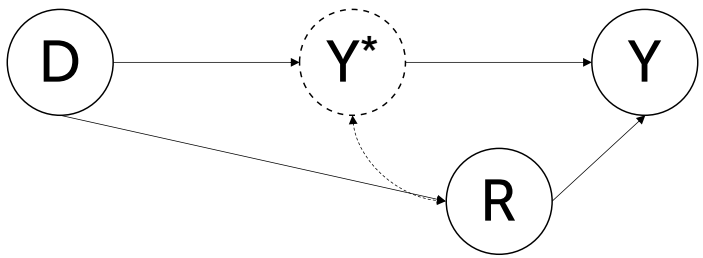
$$
P\left(Y^* \mid D\right) = P\left(Y^* \mid D, R=0\right) = P\left(Y \mid D\right)
$$

RSM Erasmus

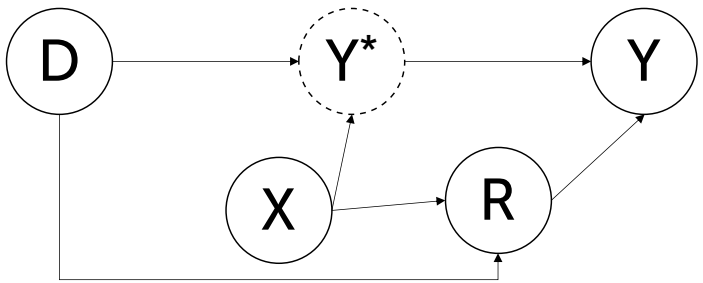
Traditional approaches: imputation

Assumption for Imputation

We observe **non-descendant** variables X that **separate** Y^* **and** R.



(c) DAG of Non-random Attrition



(d) DAG of Imputation Method

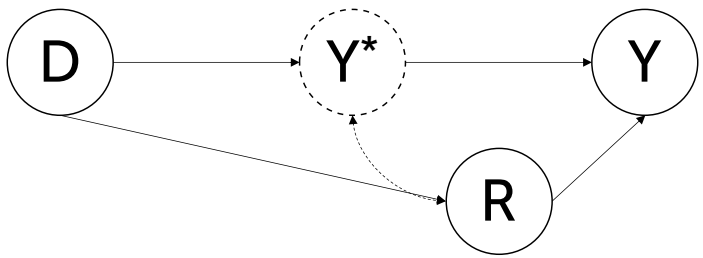
Fact (Imputation Method)

Under the assumption of the imputation method, we can identify the average treatment effect with
$P(Y^* \mid D, X) = P(Y^* \mid D, X, R)$.

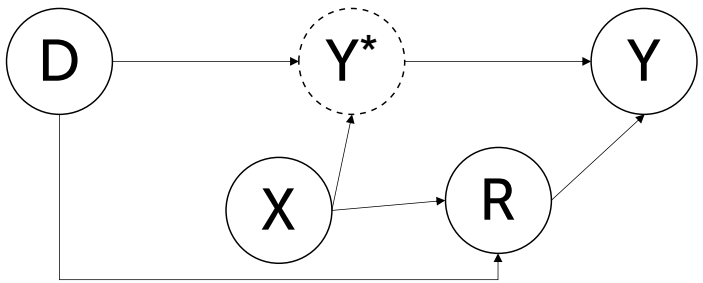
Traditional approaches: imputation

Assumption for Imputation

We observe **non-descendant** variables X that **separate** Y^* **and** R.



(e) DAG of Non-random Attrition



(f) DAG of Imputation Method

Fact (Imputation Method)

Under the assumption of the imputation method, we can identify the average treatment effect with
$P(Y^* \mid D, X) = P(Y^* \mid D, X, R)$.

Traditional approaches: imputation

Technical proof:

$$
\begin{aligned}
P\left(Y^{*} \mid do\left(D\right), X\right) &= P\left(Y^{*} \mid D, X, R=0\right) P\left(R=0\right) + P\left(Y^{*} \mid D, X, R=1\right) P\left(R=1\right. \\
&\overset{Y^{*} \perp R \mid X}{=} P\left(Y^{*} \mid D, X\right) P\left(R=0\right) + P\left(Y^{*} \mid D, X\right) P\left(R=1\right) \\
&= P\left(Y^{*} \mid D, X\right)
\end{aligned}
$$

Because the attrition is at random conditional on X, we can use observed outcome Y to calculate the conditional distribution:

$$
P\left(Y^{*} \mid D, X\right) = P\left(Y^{*} \mid D, X, R=0\right) = P\left(Y \mid D, X\right)
$$

We can then integrate over X to have the ATE:

$$
P\left(Y \mid D, X\right) = \sum_{x \in X} P\left(Y \mid D, X=x\right)
$$

RSM Erasmus

Issues of the traditional approaches

The law of credibility

The credibility of inference decreases with the strength of the assumptions maintained.

Assumptions are strong and cannot be justified.
Credibility is sacrificed for the "power" of conclusions.

- The random attrition is hardly a credible assumption.
- Difficult to find a set of X that d-seperate Y^* and R.

Could we adopt a better strategy?

RSM Erasmus

Outline

RSM Erasmus

Relaxing unrealistic assumptions

Many assumptions are non-refutable and unrealistic.

- Unconfoundedness
- Random attrition
- ...

Forego the point identification, make weaker assumptions and increase credibility of analyses.

Strong assumptions identify a point: •

Weak assumptions identify a set: $[- \; - \; - -]$

RSM Erasmus

Some types of bounds: no-data bounds

Assume the potential outcomes Y^0 and Y^1 are bounded $[0, 1]$.

$$-1 \leq Y_i^1 - Y_i^0 \leq 1 \Rightarrow E\left(Y_i^1 - Y_i^0\right) \in [-1, 1]$$

In more general terms, if Y^0 and $Y^1 \in [a, b]$

$$E\left(Y_i^1 - Y_i^0\right) \in [a - b, b - a]$$

The length of the bound is $2\,(b - a)$.

RSM Erasmus

Some types of bounds: no-assumption bounds

Assume we collect some data with $\{Y_i, D_i\}$, but random assignment is not ensured[2].
Can we do better with data?

$$E\left(Y_i^1 - Y_i^0\right) = \underbrace{\pi E\left(Y_i^1 \mid D_i = 1\right)}_{\text{Observed}} + (1-\pi)\underbrace{E\left(Y_i^1 \mid D_i = 0\right)}_{\text{Unobserved}}$$
$$- \underbrace{\pi E\left(Y_i^0 \mid D_i = 1\right)}_{\text{Unobserved}} - \underbrace{(1-\pi) E\left(Y_i^0 \mid D_i = 0\right)}_{\text{Observed}}$$

Under the bounds on potential outcomes $[a, b]$:

$$\tau^{obs} + (1-\pi)\, a - \pi b \leq E\left(Y_i^1 - Y_i^0\right) \leq \tau^{obs} + (1-\pi)\, b - \pi a$$

The length of the bound is $(b-a)$, half of the no-data bounds.

RSM Erasmus

[2] Let $\pi = P(D_i = 1)$ in the data collected.

Some types of bounds: monotone treatment response

Assume treatment always benefit people, or $Y_i^1 \geq Y_i^0, \forall i$
This means ITE is nonnegative, so the lower bound of ITE is 0.
Under this assumption, the lower bound of ATE is also 0:

$$E\left(Y_i^1 - Y_i^0\right) \geq 0$$

Using this assumption, we may further narrow the no-data and no-assumptions bounds.

Fact

Under monotone treatment response assumption, the no-data bound becomes $E\left(Y_i^1 - Y_i^0\right) \in [0, b-a]$.

RSM Erasmus

Some types of bounds: monotone treatment selection

Assume the treatment group potential outcomes are better than control groups, with

$$E\left(Y_i^1 \mid D_i = 1\right) \geq E\left(Y_i^1 \mid D_i = 0\right)$$
$$E\left(Y_i^0 \mid D_i = 1\right) \geq E\left(Y_i^0 \mid D_i = 0\right)$$

Under this assumption, the ATE is bounded from above,

$$E\left(Y_i^1 - Y_i^0\right) \leq E\left(Y_i^1 \mid D_i = 1\right) - E\left(Y_i^0 \mid D_i = 0\right)$$

We may use this assumption to further narrow the no-assumption bounds.

RSM Erasmus

Outline

RSM Erasmus

A running example

- We have a data from a randomized experiment.
- The outcome Y_i measured with a scale from 1 to 10 (i.e., $Y_i \in [1, 10]$).
- The experiment has an attrition problem, with NA denoting missing values.

Obs	D_i	Y_i
1	1	10
2	1	8
3	1	NA
4	0	3
5	0	NA
6	0	6

$Y_i \in [0, 10]$

RSM Erasmus

A running example

Obs	D_i	Y_i
1	1	10
2	1	8
3	1	NA
4	0	3
5	0	NA
6	0	6

$Y_i \in [0, 10]$

- Without additional assumptions, the intention-to-treat is: $\text{ITT}_Y = \frac{10+8}{3} - \frac{3+6}{3} = 3.$
- Under random attrition, the ATE is: $\text{ATE}_{\text{random}} = \frac{10+8}{2} - \frac{3+6}{2} = 4.5.$

RSM Erasmus

Bounding approach: Manski bounds

Manski bounds have no additional assumptions. See Horowitz and Manski (2000).

- Upper bound: substituting NA's in **treatment with max values and control with min values**.
- Lower bound: substituting NA's in **treatment with min values and treatment with max values**.

Obs	D_i	Y_i
1	1	10
2	1	8
3	1	NA
4	0	3
5	0	NA
6	0	6

$Y_i \in [0, 10]$

RSM Erasmus

Bounding approach: Manski bound

Obs	D_i	Y_i
1	1	10
2	1	8
3	1	NA
4	0	3
5	0	NA
6	0	6

- Manski bound: $\left[\frac{10+8+1}{3} - \frac{3+10+6}{3}, \frac{10+8+10}{3} - \frac{3+1+6}{3}\right] = [0, 6]$
- $\text{ITT}_Y = 3$ and $\text{ATE}_{\text{random}} = 4.5$

In practice, use *max* and *min* values of observed outcomes.
Uninformative but wide, as a benchmark for further analysis.
Can we narrow the bound?

RSM Erasmus

Lee bounds Lee (2009)

Assumptions: **monotonicity of treatment selection**.

- Treatment status either increases / decreases attrition for all people.
- The effect of treatment on attrition has the same sign ($+$ or $-$).

How to use the condition?

- Trimming samples such that share of observed individuals is equal in the treatment and control.
- Trimming from above or below to obtain lower or upper bound.

RSM Erasmus

Lee bounds: key idea

- Suppose we have a treatment group with 40% of attrition and a control group with 50%.
- The treatment decreases the attrition likelihood for 10%.
- To get the same level of "missing," we need to trim $x\%$ of people from the treatment.
- To calculate $x\%$, we do:

$$\frac{40\%}{1-x\%} = 50\% \Rightarrow x\% = \frac{50\% - 40\%}{50\%} = 20\%$$

RSM Erasmus

Lee bounds: key idea

- Suppose we have a treatment group with 40% of attrition and a control group with 50%.
- The treatment decreases the attrition likelihood for 10%.
- To get the same level of "missing," we need to trim $x\%$ of people from the treatment.
- To calculate $x\%$, we do:

$$\frac{40\%}{1-x\%} = 50\% \Rightarrow x\% = \frac{50\% - 40\%}{50\%} = 20\%$$

RSM Erasmus

Lee bounds: key idea

- Suppose we have a treatment group with 40% of attrition and a control group with 50%.
- The treatment decreases the attrition likelihood for 10%.
- To get the same level of "missing," we need to trim $x\%$ of people from the treatment.
- To calculate $x\%$, we do:

$$\frac{40\%}{1-x\%} = 50\% \Rightarrow x\% = \frac{50\% - 40\%}{50\%} = 20\%$$

RSM Erasmus

Lee bounds: key idea

- Suppose we have a treatment group with 40% of attrition and a control group with 50%.
- The treatment decreases the attrition likelihood for 10%.
- To get the same level of "missing," we need to trim $x\%$ of people from the treatment.
- To calculate $x\%$, we do:

$$\frac{40\%}{1-x\%} = 50\% \Rightarrow x\% = \frac{50\% - 40\%}{50\%} = 20\%$$

RSM Erasmus

Lee bounds: key idea

Question: **which 20% of the treated units to trim?**

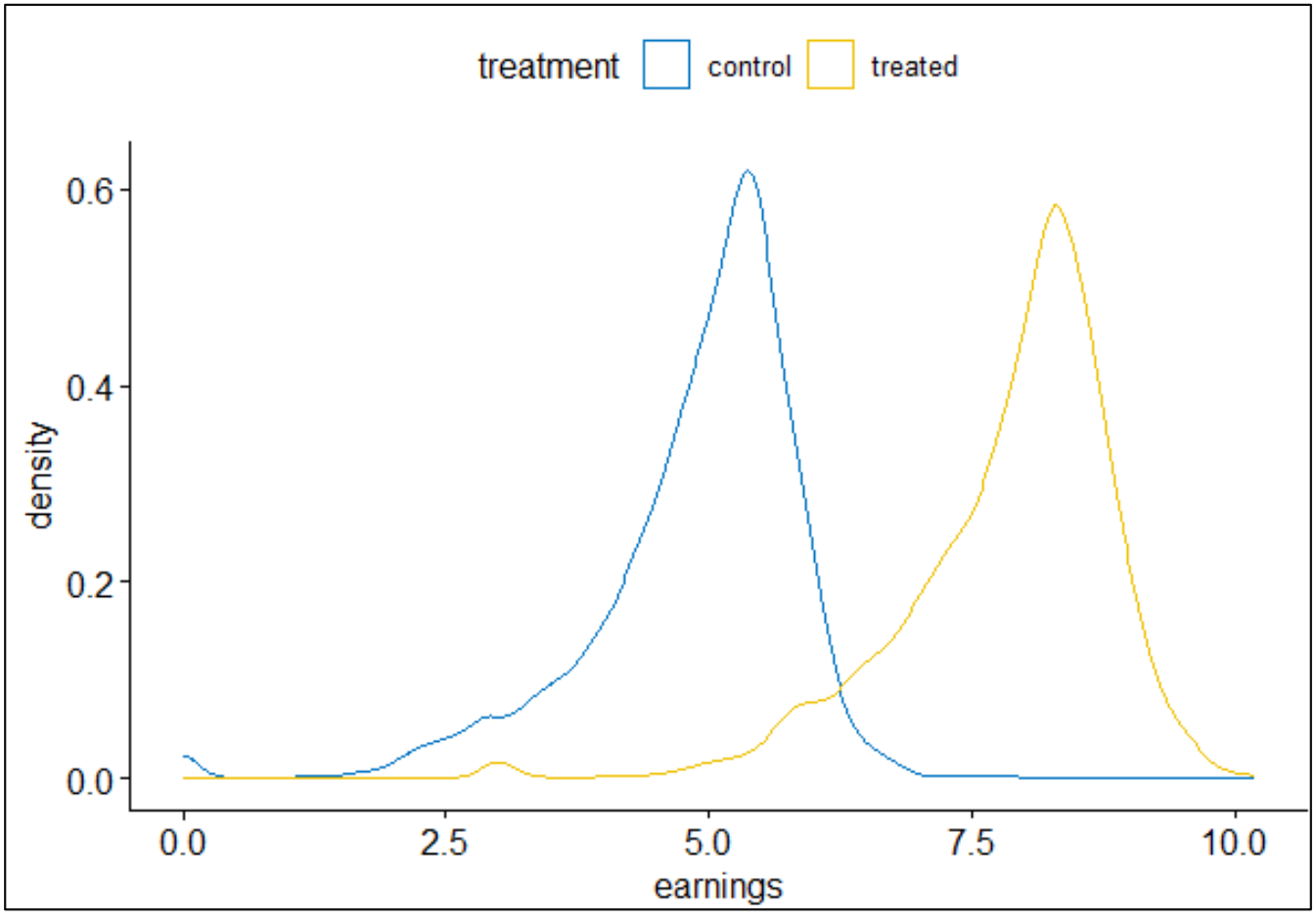
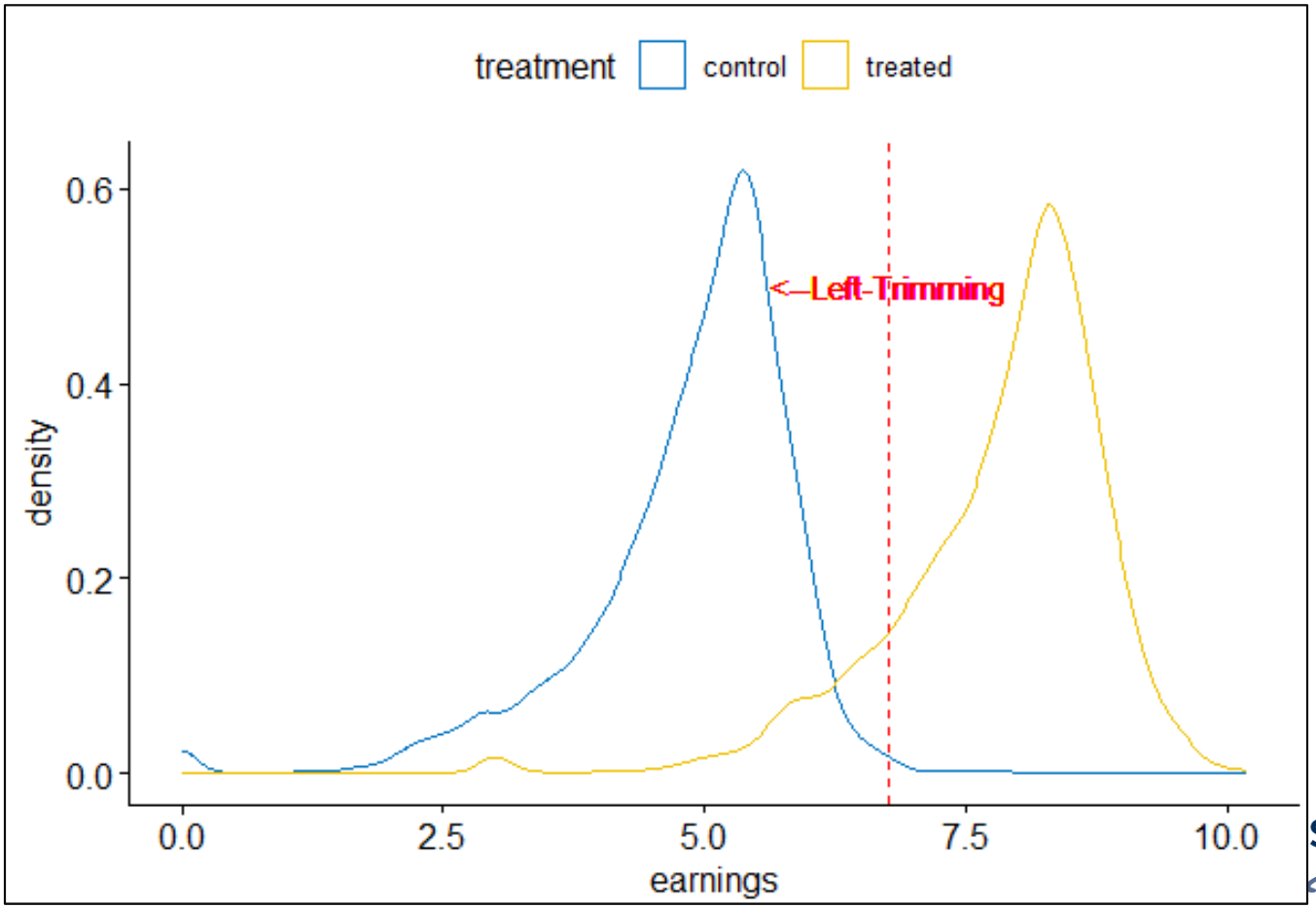


Figure: Density Plots of Outcomes for the Treated and Control Group

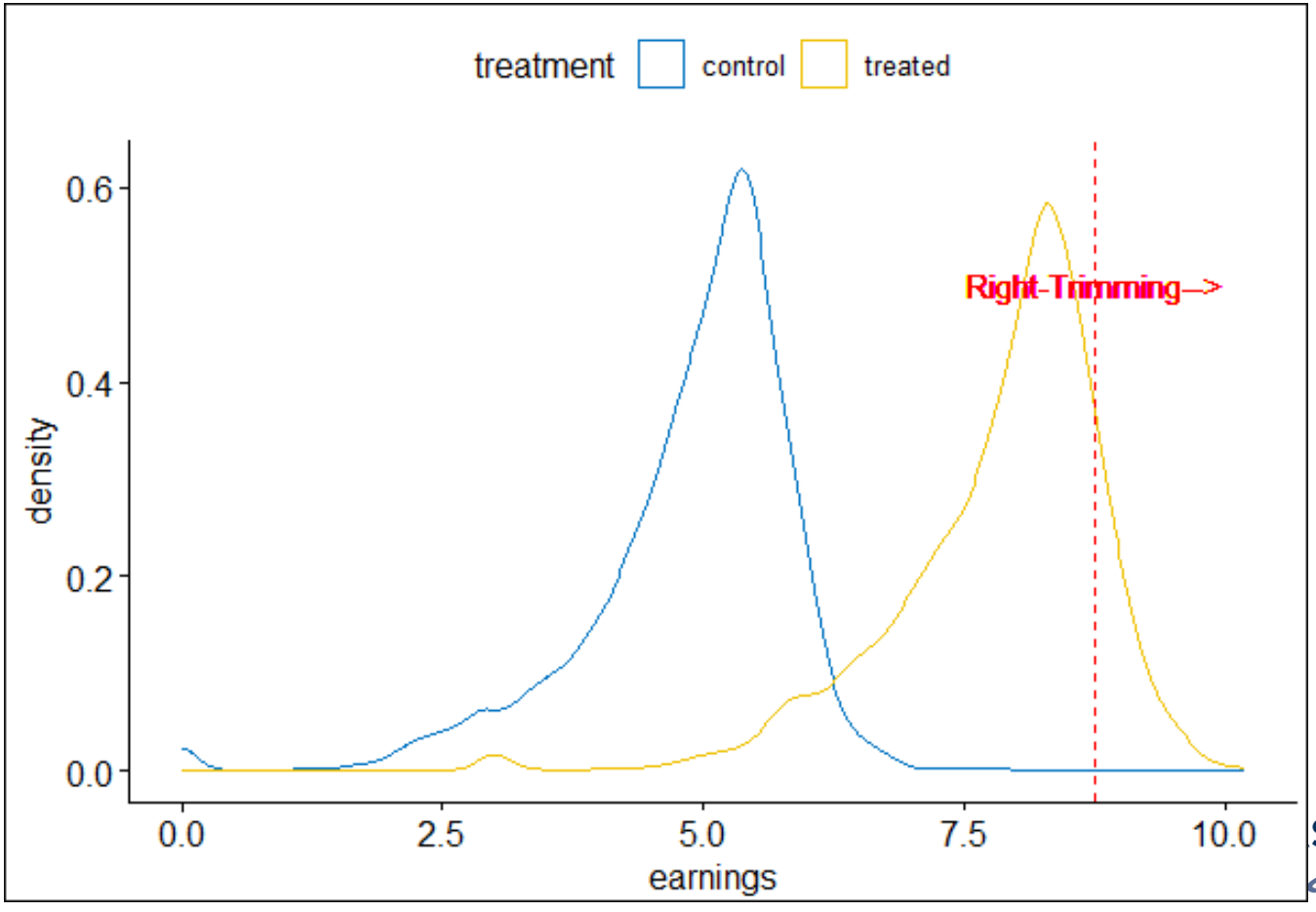
Lee bounds: key idea

To trim the people in the 20% lower quantile to get the upper bound:



Lee bounds: key idea

To trim the people in the 20% upper quantile to get the lower bound:



Lee bounds: procedure

Y the outcomes, D the treatment, and W the non-attrition.

- Step 1: Get the share of individuals with observed Y in the treatment and control, as q_1 and q_0.
- Step 2: Obtain the attrition difference:

$$q = \begin{cases} \frac{q_1 - q_0}{q_1} & \text{if } q_1 > q_0 \\ \frac{q_0 - q_1}{q_0} & \text{if } q_0 > q_1 \end{cases}$$

- Step 3: Find the qth and $(1-q)$th quantile of observed outcomes in the treatment group as Y_q^1 and Y_{1-q}^1.

RSM Erasmus

Lee bounds: procedure

Y the outcomes, D the treatment, and W the non-attrition.

- Step 4: Upper and lower bound:

$$\overline{\tau} = \underbrace{\frac{\sum 1\left(Y_i \geq Y_q^1\right) D_i W_i Y_i}{\sum 1\left(Y_i \geq Y_q^1\right) D_i W_i}}_{\text{Non-attritors in treatment with higher outcome}} - \underbrace{\frac{\sum (1 - D_i) W_i Y_i}{\sum (1 - D_i) W_i}}_{\text{Non-attriters in control}}$$

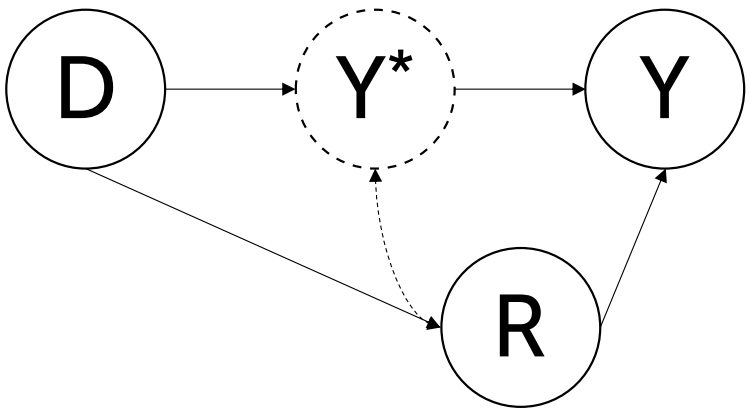
$$\underline{\tau} = \underbrace{\frac{\sum 1\left(Y_i \leq Y_{1-q}^1\right) D_i W_i Y_i}{\sum 1\left(Y_i \leq Y_{1-q}^1\right) D_i W_i}}_{\text{Non-attritors in treatment with lower outcome}} - \underbrace{\frac{\sum (1 - D_i) W_i Y_i}{\sum (1 - D_i) W_i}}_{\text{Non-attriters in control}}$$

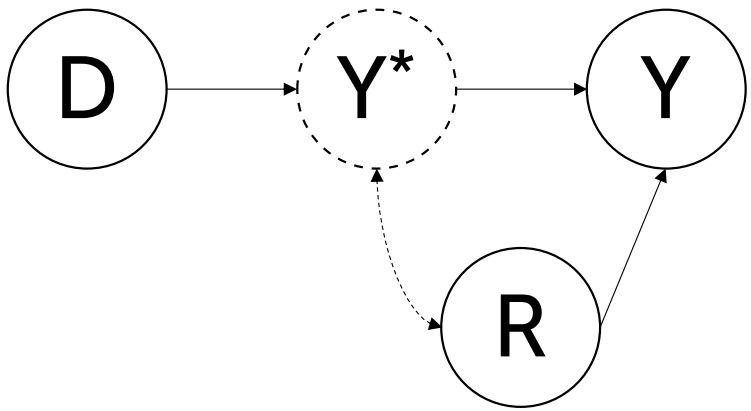
RSM Erasmus

Lee bounds: summary

Treatment influences attrition in one direction:

- By comparing the treatment and control, we may "partial out" the effect of treatment on attrition.
- ATE on the non-attritors!





References I

- Couture, Victor, Benjamin Faber, Yizhen Gu, and Lizhi Liu. 2021. "Connecting the Countryside via E-Commerce: Evidence from China." American Economic Review: Insights, 3 (1): 35-50.
- Horowitz, J. L., & Manski, C. F. (2000). Nonparametric analysis of randomized experiments with missing covariate and outcome data. Journal of the American statistical Association, 95(449), 77-84.
- Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. The Review of Economic Studies, 76(3), 1071-1102.

RSM Erasmus