Lecture 9 Regression Discontinuity

Filipa Sá

King's College London

Semester 2, 2019/20

Readings

- Angrist and Pischke chapter 4
- Carpenter and Dobkin (2011), "The Minimum Legal Drinking Age", *Journal of Economic Perspectives*, Vol. 25, No. 2, pp. 133-156

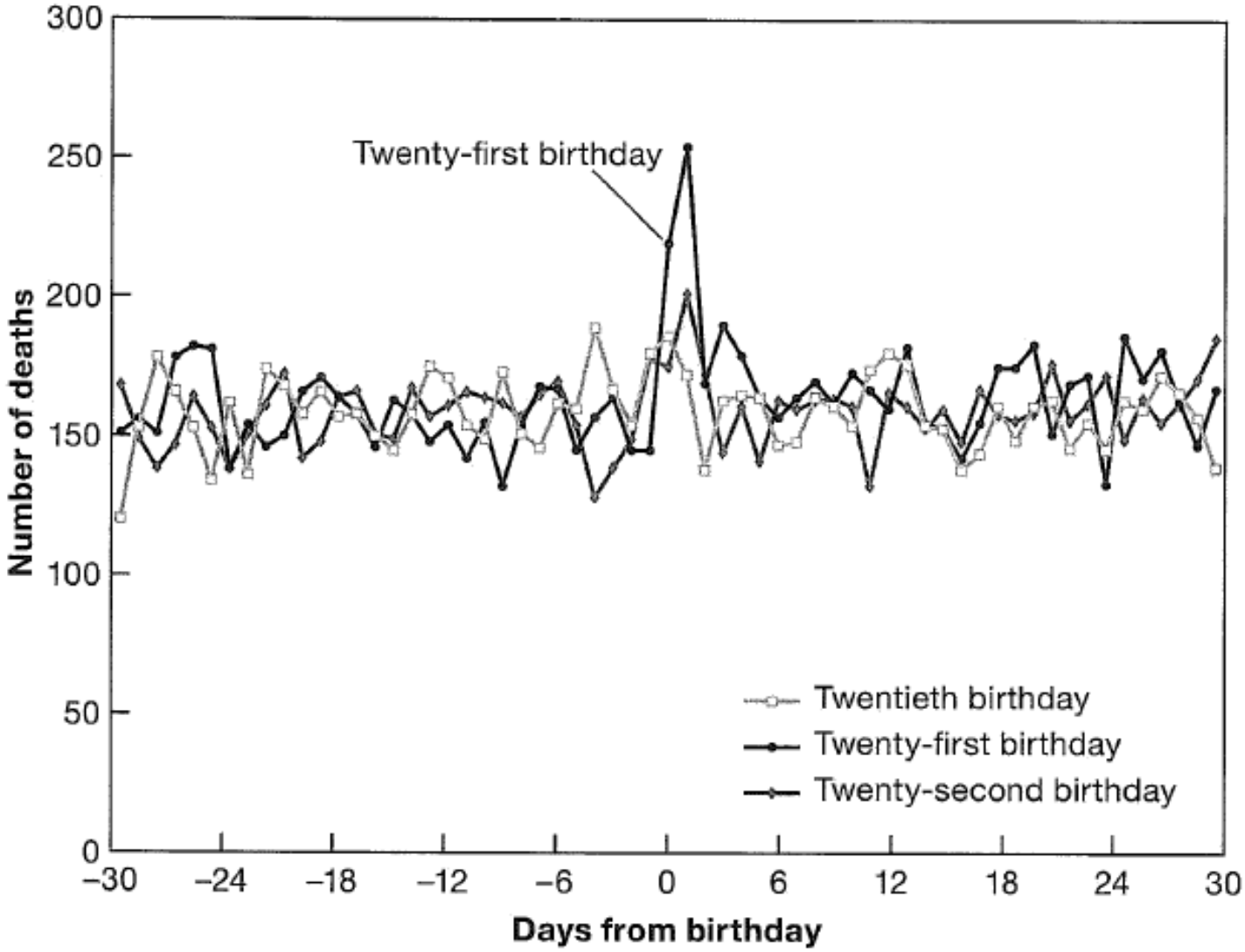
Regression discontinuity

- Our life has many rules which generate discontinuities:
  - We cannot retire until we reach a certain age
  - A party only wins an election if it has more than 50% of votes
  - ...
- Rigid rules create valuable experiments and regression discontinuity (RD) is a tool that allows us to study these experiments

MLDA experiment

- The Minimum Legal Drinking Age (MLDA) in the US is 21
- There is a spike in mortality risk after the 21st birthday
- The spike does not seem to reflect the effect of the birthday party itself. If that was the case, there should be a spike in mortality after the 22nd birthday as well.

MLDA experiment



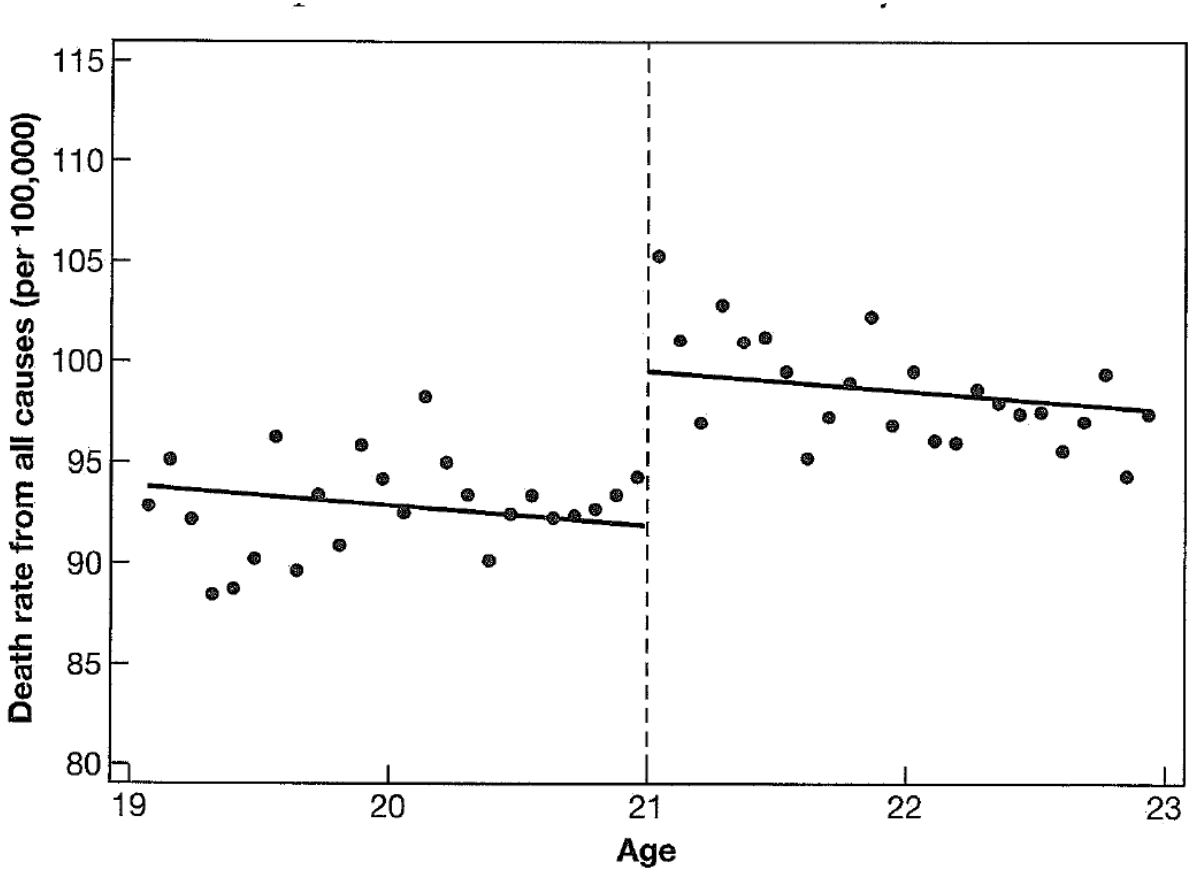
MLDA experiment

- Causal question: what is the effect of legal access to alcohol on death rates?
- Treatment is a discontinuous function of age:

$$D_a = \begin{cases} 1 & \text{if } a \geq 21 \\ 0 & \text{if } a < 21. \end{cases}$$

- The variable that determines treatment — in this case, age — is called the *running variable*

MLDA experiment



MLDA experiment

- The RD analysis of the MLDA is based on a regression:

$$M_a = \alpha + \rho D_a + \gamma a + \varepsilon_a$$

- M_a is the death rate in month a (month is defined as a 30-day interval counting from the 21st birthday)
- The regression controls for treatment D_a and age. It is important to control for age because it affects mortality in ways other than the MLDA:
  - older people are more likely to die from internal causes, i.e, illness
  - older people are less likely to die from accidents and homicides

MLDA experiment

- This regression produces the lines in the figure
- γ captures the smooth decline in death rates among young people as they mature
- ρ captures the jump in deaths at age 21
- The regression generates an estimate of ρ of 7.7, which is quite a large effect when compared to an average death rate of around 95.

MLDA experiment

- Should we not control for other things?
- Recall the OVB formula:
  OVB = Relation between omitted variable and $D_a \times$ Effect of omitted variable in long
- But we know that D_a is determined only by age, so there are no other variables related with D_a. There is no OVB in this short regression.
- The lack of OVB is a payoff that comes from the fact that we know where treatment comes from. Although treatment is not randomly assigned, it is determined only by age.

How could RD go wrong?

- If the relation between the outcome variable (Y) and the running variable (X) is linear with a clear jump in $E(Y|X)$ at the cutoff, then RD should capture the causal effect of treatment — first panel in the figure
- If the relation is nonlinear, we may still see a jump at the cutoff point — second panel — or we may not see a jump — third panel
- In the third panel, if the regression is linear on the running variable, we may conclude that there is a discontinuity, when in fact there is nonlinearity.
- We will see how to deal with this next week.

How could RD go wrong?