Lecture 1 Randomised Trials

Filipa Sá

King's College London

Semester 1, 2019/20

Contact details

- Email: filipa.sa@kcl.ac.uk
- Office Hours: Mondays 3:00-4:00 pm and Tuesdays 2:00-3:00 pm, Bush House (N) 4.17

Structure of the course

- Methods that economists (and other scientists) use to study causal questions
  - Randomised trials
  - Regression
  - Instrumental variables
  - Difference-in-Differences
  - Regression Discontinuity Designs
- Focus on real-world questions and data applications using Stata

Structure of the course

- Assessment
  - 80% coursework — take-home extended problem set, posted on December 6 and due on December 20
  - 20% tutorial assessment — two short problem sets, to be submitted on KEATS, due on October 14 and November 11

Readings

- Main textbook:
  - Angrist and Pischke (2015), *Mastering 'Metrics*, Princeton University Press
- This will be complemented with:
  - Wooldridge (2016), *Introductory Econometrics: a Modern Approach*, Cengage Learning, 6th edition
  - Some academic papers

Readings — randomised trials

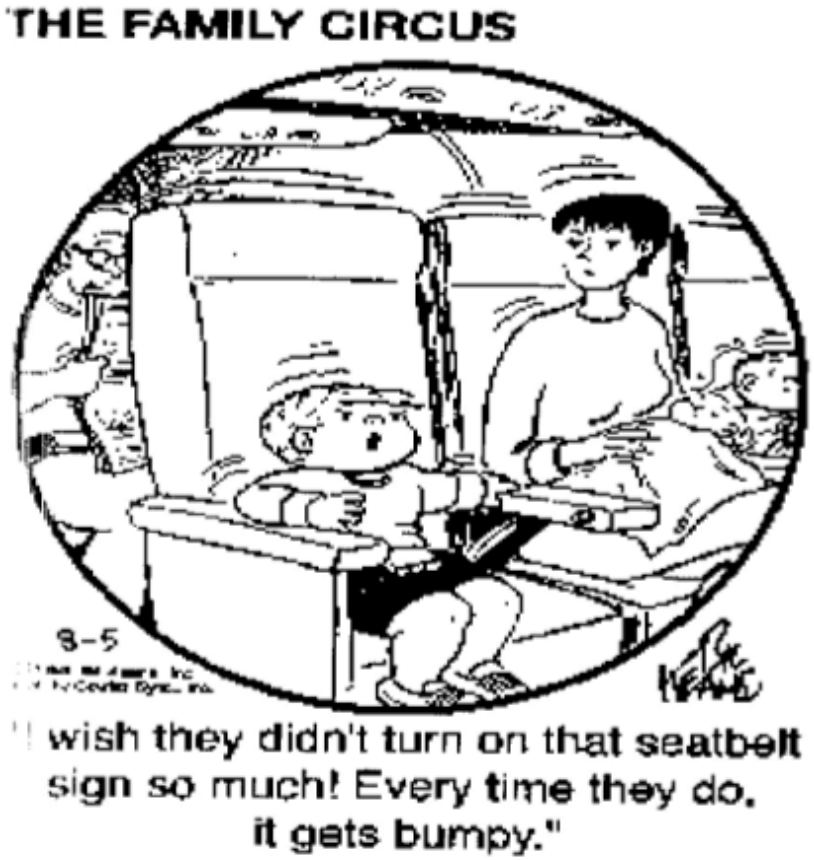
- Angrist and Pischke chapter 1
- Wooldridge chapter 1

What do economists do?

- Apply models to understand the world
- Use data to answer cause-and-effect questions:
  - Does free health insurance make people healthier?
  - Is there racial discrimination in the labour market?
  - Does going to a high-rank university increase earnings?
  - Does immigration increase house prices?
  - Do children in smaller families have better opportunities?
  - Does a reduction in the minimum legal drinking age increase youth death?
  - Does central bank lending to troubled banks help prevent bank failures?
  - Do higher tuition fees decrease university applications?

Thinking like an economist

- Correlation is not causality



Thinking like an economist

- Correlation is not causality
  - People who smoke have higher levels of stress
  - Is smoking a cause of stress?

Thinking like an economist

- Interesting correlations. . .
  http://www.tylervigen.com/spurious-correlations

Other things equal

- Suppose we are interested in studying the causal effect of health insurance on health — are people healthier when the government provides free health care (e.g. the NHS)?
- In the US, there is no NHS. Elderly people are covered by Medicare and poor people are covered by Medicaid. Those who are not covered, can choose to participate in an employer-provided insurance plan or buy private insurance. But many Americans are uninsured. When they are unwell, they rely on A&E.
- How can we test whether health insurance improves health?
  - We can compare the health of someone covered by insurance with the health of someone who is uninsured
  - *Ceteris paribus (other things equal)* — we assume that these two people have the same characteristics, except for insurance coverage
  - But... perhaps people with insurance have different characteristics

Other things equal

- Those with health insurance are more educated, have higher income and are more likely to be working than the uninsured —Table 1.1
- Differences in health could be partly due to these different characteristics — more educated people tend to smoke less, eat better and be healthier
- Comparing the two groups is like comparing apples with oranges

Other things equal

TABLE 1.1
Health and demographic characteristics of insured and uninsured couples in the NHIS

	Husbands			Wives		
	Some HI (1)	No HI (2)	Difference (3)	Some HI (4)	No HI (5)	Difference (6)
A. Health						
Health index	4.01 [.93]	3.70 [1.01]	.31 (.03)	4.02 [.92]	3.62 [1.01]	.39 (.04)
B. Characteristics						
Nonwhite	.16	.17	−.01 (.01)	.15	.17	−.02 (.01)
Age	43.98	41.26	2.71 (.29)	42.24	39.62	2.62 (.30)
Education	14.31	11.56	2.74 (.10)	14.44	11.80	2.64 (.11)
Family size	3.50	3.98	−.47 (.05)	3.49	3.93	−.43 (.05)
Employed	.92	.85	.07 (.01)	.77	.56	.21 (.02)
Family income	106,467	45,656	60,810 (1,355)	106,212	46,385	59,828 (1,406)
Sample size	8,114	1,281		8,264	1,131	

Notes: This table reports average characteristics for insured and uninsured married couples in the 2009 National Health Interview Survey (NHIS). Columns (1), (2), (4), and (5) show average characteristics of the group of individuals specified by the column heading. Columns (3) and (6) report the difference between the average characteristic for individuals with and without health insurance (HI). Standard deviations are in brackets; standard errors are reported in parentheses.

The language of experiments

- **Outcome** — a measure of health (for example, an index for "Would you say your health in general is excellent, very good, good, fair or poor?")
- **Treatment** — a variable that indicates coverage by private health insurance
- **Treatment group** — people with insurance
- **Control group** — people without insurance

The language of experiments

- Outcome of person i without health insurance: Y_{i0}
- Outcome of person i with health insurance: Y_{i1}
- Causal effect: $Y_{i1} - Y_{i0}$
- But for each person, we either observe Y_{i1} or Y_{i0}
- We don't observe the **counterfactual** — what would have been the health of someone with insurance if they had not taken insurance
- Treatment status is captured by a dummy variable:
  - $D_i = 1$ if i is insured
  - $D_i = 0$ otherwise

Selection bias

- Suppose that health insurance makes everyone healthier by a constant amount κ :

$$Y_{i1} = Y_{i0} + \kappa \tag{1}$$

- We compare the average health of people with and without insurance:

$$Avg[Y_{i1}|D_i = 1] - Avg[Y_{i0}|D_i = 0] \tag{2}$$

- How does this compare with the causal effect κ?
- From equation (1):

$$Avg[Y_{i1}|D_i = 1] = Avg[Y_{i0}|D_i = 1] + \kappa$$

Selection bias

- Substituting in equation (2):

$$\begin{aligned} Avg[Y_{i1}|D_i &= 1] - Avg[Y_{i0}|D_i = 0] = \\ \kappa + [Avg[Y_{i0}|D_i &= 1] - Avg[Y_{i0}|D_i = 0]] \end{aligned}$$

  This equation tells us that:
  Difference in group means = Average causal effect + Selection bias
- The **selection bias** is the difference in average Y_{i0} between the groups being compared
- If people who have health insurance are healthier for all sorts of reasons (because they have higher education, higher income, are more likely to be employed, etc), the selection bias is positive
- We would conclude that health insurance improves health, but we are not comparing like for like

Selection bias

- If selection bias is only due to observable characteristics, then it is relatively easy to deal with:
  - We can use regression and control for those characteristics — more on this in lectures 3 and 4
  - This is equivalent to focusing on samples of people with the same education, income and employment status
  - But if there are unobservable differences between the insured and the uninsured, it is more challenging — we will study a solution to this next week!