Attrition

- We saw that random assignment allows to have an unbiased estimator of the ATE
  - However, we are assuming that we observe all outcomes (for treatment and control)
- Attrition: data are missing ☹. Why may that happen? Example:
  - Participants may refuse to participate in post-treatment surveys
  - Participants moved away, and it is not possible to contact them
  - Outcome variable becomes unavailable (e.g. participants quitted their jobs or dropped school)
- Is this a problem? When?
  - It may be a serious problem

Attrition: Application

- RAND Health insurance study (Newhouse, 1989)
  - People (from low-income families) received a health insurance plan, which randomly covered 5%, 50%, 75% or 100% of health costs (for cost below US$1,000)
  - This experiment cost more than US$300 millions in 2010 dollars
  - Results: After a few years, patients with 100% coverage used much more hospital admissions and physician visits. But were not healthier
    - But, some people dropped out the study. Who?
    - What if those who had to pay dropped out because they anticipated serious health problems and preferred to pay for a different insurance plan? Maybe
    - What if those who dropped out were "the same" as the ones who remained? Maybe
  - Main issue: Those who drop out do not provide any health outcome → missing outcomes allows multiple interpretations

Attrition: What can we do?

- Key: Do subjects with missing outcomes have, on average, the same Y_i as subjects with available outcomes? (i.e. attrition is not correlated to potential outcomes)
  - Perhaps, yes - Would that be fine?
- For example, in the previous study, participants who dropped out and those who didn't have similar pretreatment health measures
- Another way to address attrition is filling in data with extreme values to get a conservative result