Experimental Methods: Lecture 3

Attrition

Raymond Duch

May 10, 2022

University of Oxford

Road Map to Lecture 3: Attrition

- Non-interference re-visited
- Attrition
- RAND case study
- Combating Attrition
- MIPO

Non-interference

$$\text{ATE} = \text{E}\left[Yi(1)_j D_i = 1\right] \text{E}\left[Yi(0)_j D_i = 0\right]$$

- Thus, average outcomes in the control and treatment groups in the sample are unbiased estimators of E [Yi (1)] and E [Yi (0)]
- But implicit in this description is the assumption that the researcher observes Yi for all subjects assigned
- Attrition occurs when outcome data are missing
- If attrition systematically related to potential outcomes, remaining subjects are no longer random samples of original group of subjects
- Thus, comparison of remaining group averages may no longer be an unbiased estimator of ATE

How Might Attrition Occur

- Subjects refuse to cooperate with researchers
  - Respondents refuse post-treatment questionnaire
- Researchers lose track of experimental subjects
  - Subjects change address or name
- Firms, organizations, or governments block researchers' access to outcomes
  - Common with experiments on sensitive topics, e.g. corruption
- Outcome variable unavailable for some subjects
  - For a job training program treatment that wants to measure wages six months later, how do you measure this outcome for subjects without jobs?
- Researchers deliberately discard observations
  - Subjects not understanding instructions get discarded

Rand Health Insurance Experiment: Setup

- Famous study illustrating the dangers of attrition for causal inference Cost $80 million (1974 USD), or about $300 million today
- Examines how copayment schemes in health insurance a ect health service consumption and outcomes
- Four groups: 5%, 50% 75%, and 100% cost coverage treatments
- Costs over $1,000 for last three groups fully insured
- Incentives so you "cannot lose financially by participating"
- As a related aside, Finkelstein et al's Oregon Health Insurance experiment

Rand Health Insurance Experiment: Setup

- Evaluation 3-5 years after treatment
- Those paying larger share made fewer physician visits and had lower rates of hospital admissions
- 100% group consumed 46% more in health services than 5% group
- 100% group no healthier on average than others, based on a wide array of health assessments
- Traditional interpretation of result: Requiring large co-pays does not have an adverse impact on health incomes for the average person
- But what about attrition? (Newhouse 1989)
- Any thoughts on how attrition might happen in a way that is systematically related to potential outcomes?

Rand Health Insurance Experiment: Critique

- 8% in free group refused to enroll
  - Military service, institutionalization, death, incomplete data collection
- 25% refusal rate in 5
  - 6.7% in free plan left voluntarily after enrolling
  - 0.4% in 5% treatment group left after enrolling
- One interpretation: Those who chose to remain in copay plans anticipated lower health costs
  - Those anticipating illness drop out to pick up cheaper private insurance
- Alternative: Those who dropped out have same unmeasured health outcomes and health service consumption than those remaining
- How would we know? How would we deal with it?

Four Strategies to Combat Attrition

- Assume missingness is independent of potential outcomes
  - Unbiased inferences but lose sample size
  - Cannot directly assess this
  - Newhouse points out average pre-treatment health outcomes and health service utilization are similar across experimental groups among subjects who do not drop out of the study
  - If one assumes prior health outcomes are indicative of subsequent unmeasured potential outcomes, one could interpret this evidence to mean missingness is independent of potential outcomes

Four Strategies to Combat Attrition

- Assume missingness independent of potential outcomes within subgroups defined by background attributes
  - Newhouse reports refusal to enroll is correlated with age/education
  - We could recover unbiased ATE by re-weighting the data to "fill in" age/education cells depleted by attrition
- Place bounds by guessing missing values (Manski 1995, 2007)
  - Assume those who disappear are extremely healthy or extremely ill
  - Alternatively, "trim" healthiest observations in free care group until attrition rate as high as cost-sharing groups
- Gather more data from missing subjects (Cochran 1977)
  - RAND eventually gathered health outcomes from 77

Special Forms of Attrition: MIPO

- If $R_i(z) \perp\!\!\!\perp Y_i(z)$, attrition is (relatively) innocuous
- Missing Independent of Potential Outcomes or MIPO
- Can only gather circumstantial evidence about plausibility
  - If brought about by random procedure, no systematic procedure between r_i and subject background attributes or experimental assignment should exist (regr F-test)
  - But do covariates at one's disposal include the systematic sources of missingness?
- MIPO assumption more convincing if subjects have little discretion over whether their outcomes will be reported
  - Tutoring: Students miss for midyear exam
  - Voter mobilization: Town clerk fails to record data in timely fashion
  - Arguments for/against MIPO in these cases?

Special Forms of Attrition: MIPO

- Alternative assumption: Attrition is unrelated to potential outcomes conditional on pre-treatment covariates X_i
- Not surprisingly, this is referred to as Missing Independent of Potential Outcomes given X, or MIPO | X
- Example: In student test example, we condition on attendance record prior to the treatment intervention as X_i
- Intuition: Assume $N_1 = 30$ Oxford, $N_2 = 10$ Cambridge students treated with tutoring vouchers
    - Let's say 15 Oxford students were missing at random
    - How might you recover an unbiased ATE in this example?