Encouragement Designs, Missing Data, Open Discussion

Jake Bowers

06 June 2022

Core assumptions

Missing data in experiments

Non-compliance: Causal effects when we do not control the dose

Open Discussion

References

Key points for this lecture

▶ Attrition (missing data on outcomes) can lead to bias.

▶ Non-compliance can teach us about the effect of the "active ingredient" of the treatment.

Core assumptions

Review of core assumptions

1. Excludability

2. Non-interference

3. Random assignment of subjects treatment

Missing data in experiments

Attrition (missing data on outcomes)

- ▶ Some units may have missing data on outcomes (= units attrit) when:

    - ▶ some respondents can't be found or refuse to participate in endline data collection.

    - ▶ some records are lost.

- ▶ This is a problem when treatment affects missingness.

    - ▶ For example, units in control may be less willing to answer survey questions.
    - ▶ For example, treatment may have caused units to migrate and cannot be reached

- ▶ If we analyze the data by dropping units with missing outcomes, then we are no longer comparing similar treatment and control groups.

What can we do?

► Check whether attrition rates are similar in treatment and control groups.

► Check whether treatment and control groups have similar covariate profiles.

► Do not drop observations that are missing outcome data from your analysis.

► When outcome data are missing we can sometimes **bound** our estimates of treatment effects.

What can we do?

▶ But the best approach is to try to anticipate and prevent attrition.

   ▶ Blind people to their treatment status.

   ▶ Promise to deliver the treatment to the control group after the research is completed.

   ▶ Plan ex ante to reach all subjects at endline.

   ▶ Budget for intensive follow-up with a random sample of attriters.

Missing data on covariates is not as problematic

- Missing **background covariates** (i.e.,variables for which values do not change as a result of treatment) for some observations is less problematic.

  - We can still learn about the causal effect of an experiment without those covariates, as we saw in the Hypothesis Testing and the Estimation modules.

  - We can also use the background covariate as planned by imputing for the missing values.

  - We can also condition on that missingness directly.

Missing data on treatment assignment or blocks or clusters is a big problem

This would cause very similar problem as would missing data on outcomes.

Summary: Missing data in experiments

- ▶ Missing data on covariates is not a problem: it should be balanced by randomization.
- ▶ Missing data on outcomes is a problem: it can remove the benefits of the randomization.
- ▶ Missing data on treatment assignment is a problem: it can remove the benefits of the randomization.

Given missing data on outcomes:

- ▶ Check to see if treatment predicts missingness
- ▶ Check to see if missingness patterns relate strongly to covariates.
- ▶ Report results using bounds (set all outcomes to max and min, report both results)
- ▶ Report results using sensitivity analysis (like an observational study (see Rosenbaum (2017))).

Non-compliance: Causal effects when we do not control the dose

Non-compliance or Encouragement Designs I

▶ Sometimes units assigned to treatment don't take it. They don't comply with their assignment.

  ▶ If all units assigned to control do not take the treatment, but only some units assigned to treatment take the treatment, we have *one-sided non-compliance*.

▶ The effect of **treatment assignment** (often called the Intent to Treat Effect or ITT) is not the same as the effect of **receiving a dose of the treatment**.

  ▶ Not random who takes a dose of the treatment. Directly comparing people who take a dose with those who did not ("per protocol" or "as treated" comparisons) is not justified by randomization.

Non-compliance or Encouragement Designs I

▶ The effect of receiving the treatment is often called the "Local Average Treatment Effect" (LATE) or "Complier Average Causal Effect" (CACE) (same quantity).

  ▶ "Local" refers to the idea that the effect only occurs on the people who take the treatment when assigned to treatment (the kinds of people who 'comply' with the intentions of the researcher).

▶ Two main approaches to estimation:

  1. Instrumental variable (in which random assignment may be a good instrument and often we estimate the CACE using 2SLS) (See Gerber and Green (2012) Chapter 5) and
  2. Placebo-controlled experiments (Nickerson (2005), example Broockman and Kalla (2016)). In which we estimate the CACE by comparing outcomes of treated and placebo groups.

How to learn about the CACE/LATE using a Placebo

Imagine a door-to-door communication experiment where some houses are randomly assigned to receive a visit and some attitude is measured as an outcome later. Note that we now use Z (treatment assigned) and D (treatment taken) instead of just T.
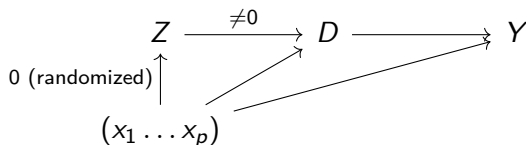
$$Z \xrightarrow{\ \neq 0\ } D \longrightarrow Y$$

0 (randomized) $\uparrow$

$(x_1 \ldots x_p)$

- ▶ Z_i is random assignment to a visit ($Z_i = 1$) or not ($Z_i = 0$).
- ▶ $D_i(Z_i = 1) = 1$ means that person i would open the door to have a conversation when assigned a visit (it is a potential dose)
- ▶ $D_i(Z_i = 1) = 0$ means that person i would not open the door to have a conversation when assigned a visit.
- ▶ $x_1 \ldots x_p$ are background covariates — confounders of the $D_i \to Y_i$ relationship.

How to learn about the CACE/LATE using a Placebo
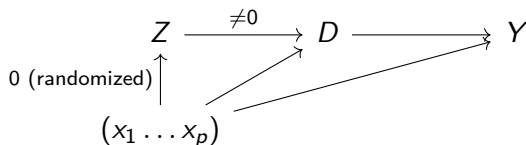
$$Z \xrightarrow{\neq 0} D \longrightarrow Y$$

0 (randomized) $\uparrow$

$(x_1 \ldots x_p)$

- Z_i is random assignment to a visit ($Z_i = 1$) or not ($Z_i = 0$).
- $D_i(Z_i = 1) = 1$ means that person i would open the door to have a conversation when assigned a visit (it is a potential dose)
- $D_i(Z_i = 1) = 0$ means that person i would not open the door to have a conversation when assigned a visit.
- $x_1 \ldots x_p$ are background covariates — confounders of the $D_i \to Y_i$ relationship.
- So, D_i is **an intermediate outcome:** $D_i = Z_i * D_i(1) + (1 - Z_i)D_i(0)$.
- $Y_i(Z_i = 1, D_i(Z_i = 1) = 1)$ is primary outcome when i is assigned to treatment **and** would open the door.

Defining causal effects

So now we have four potential outcomes per person.

- $Y_i(Z_i = 1, D_i = 1)$ is the potential outcome for people who were assigned a visit and who opened the door. ("Compliers" or "Always-takers")

- $Y_i(1, D_i = 0)$ is the potential outcome for people who were assigned a visit and who did not open the door. ("Never-takers" or "Defiers")

- $Y_i(0, D_i = 1)$ is the potential outcome for people who were not assigned a visit and who opened the door. ("Defiers" or "Always-takers")

- $Y_i(0, D_i = 0)$ is the potential outcome for people who were not assigned a visit and who did not open the door. ("Compliers" or "Never-takers")

Defining causal effects

The average causal effect of the "intention to visit people" is the ITT (averaging over whether or not people openned the door):

$$ITT = ITT_Y = \delta = \bar{Y}(Z = 1) - \bar{Y}(Z = 0).$$

The average causal effect of talking with someone at the door is the average treatment effect on compliers (the type of person who would open the door if visited) or the CACE:

$$CACE = \bar{Y}(Z_i = 1, D_i = 1 | D_i(1) = 1) - \bar{Y}(Z_i = 0, D_i = 0 | D_i(1) = 1)$$

How to estimate the CACE using a Placebo Design?

We can estimate the average outcome among those assigned to treatment who opened their doors ($\bar{Y}(Z_i = 1, D_i = 1|D_i(1) = 1)$) because there are no "always-takers" (i.e. people can only talk with a canvasser if a canvesser shows up).

$$\hat{\bar{Y}}_i(Z_i = 1, D_i = 1|D_i(1) = 1) = \sum_i \frac{(Z_i D_i Y_i)}{\sum_i D_i}$$

Key here is that we use the design to claim that only compliers open the door when visited.

How to estimate the CACE using a Placebo Design?

What about $\bar{Y}(Z_i = 0, D_i = 0 | D_i(1) = 1)$? How to learn about the outcome of people not assigned treatment but who would have opened their door if visited? (Hmm... We don't know whether they would have openned the door because they were not assigned a visit!)

What about a placebo? The Nickerson placebo-controlled design adds an arm to the experiment. For example, $Z_i = 2$, where **the compliance should be the same as treatment** $(D_i(Z_i = 2) = D_i(Z_i = 1))$ and **outcome should be the same as control** $(Y_i(Z_i = 2) = Y_i(Z_i = 0))$. And where the $Z_i = 2$ is independent of all $x_1 \ldots x_p$ because of randomization. So people in $Z_i = 2$ are not systematically different from people in $Z_i = 1$.

How to estimate the CACE using a Placebo Design?

This means that for:

$$CACE = \bar{Y}(Z_i = 1, D_i = 1 | D_i(1) = 1) - \bar{Y}(Z_i = 0, D_i = 0 | D_i(1) = 1)$$

We can estimate it with

$$\widehat{CACE} = \left(\sum_i \frac{(I(Z_i = 1)D_i Y_i)}{\sum_i I(Z_i = 1)D_i} \right) - \left(\sum_i \frac{(I(Z_i = 2)D_i Y_i)}{\sum_i I(Z_i = 2)D_i} \right)$$

Notice: we can check the assumptions:

▶ same compliance pattern (and same complier types between treatment and placebo): $D_i(Z_i = 2) = D_i(Z_i = 1)$
▶ same outcome pattern between placebo and control: $Y_i(Z_i = 2) = Y_i(Z_i = 0)$

Example from Broockman and Kalla

For example see that paper.

Summary

- ▶ Analyze as you randomized, even when you don't control the dose, to get the ITT (very common in policy contexts).
- ▶ The danger of per-protocol analysis (comparing based on D_i is an observational study, not a randomized experiment)
- ▶ If a placebo-controlled study doesn't work for you, you can use an instrumental variable approach (see Gerber and Green (2012) chapter 5, or end of the slides from our Estimation day).

Open Discussion

Your questions?

References

References I

Broockman, David, and Joshua Kalla. 2016. "Durably reducing transphobia: A field experiment on door-to-door canvassing." *Science* 352 (6282): 220–24.

Gerber, Alan S., and Donald P. Green. 2012. *Field Experiments: Design, Analysis, and Interpretation*. New York, NY: W. W. Norton & Company.

Nickerson, D. W. 2005. "Scalable Protocols Offer Efficient Design for Field Experiments." *Political Analysis* 13 (3): 233.

Rosenbaum, Paul R. 2017. *Observation and Experiment: An Introduction to Causal Inference*. Harvard University Press.