

w241: Experiments and Causality

Attrition, Mediation, and Generalizability

David Reiley, David Broockman, D. Alex Hughes

UC Berkeley, School of Information

Updated: 2021-07-19

Introduction:

Attrition, Mediation, and Generalizability

Attrition

Attrition: Outcomes for all subjects cannot be measured in an experiment

- The Y value may be missing for some subjects
- Why could attrition happen?

Example: Causes of Attrition

- Imagine studying a medicine that lowers blood pressure and therefore saves lives
- Groups
 - Treatment = 1
 - Control = 0
 - People present in both groups

Example: Causes of Attrition

Group	Status	Blood Pressure
0	Alive	70
0	Dead	?
0	Dead	?
1	Alive	100
1	Alive	90
1	Alive	80

Example: Causes of Attrition

Comparing mean of both groups may lead to an incorrect conclusion that this drug **raises** blood pressure

- Mean in treatment group: 90
- Mean in control group: 70
 - Apples-to-apples comparison lost due to attrition
 - Must consider how people came to have x values that may be used as comparisons
 - If subjects are lost from control group then the apples-to-apples comparison may be lost

Blood pressure comparisons between treatment and control groups cannot be made due to deaths in the control group

- Poor-quality trial for assessing the effects of this drug on blood pressure

Easy Example: Attrition by Movement

Example: Individual tax returns from cities as outcome data

- Experiment: What happens if we raise tax rate in cities?
- Concern: if taxes go up, people may have reduced incentive to work
- Due to higher taxes in some cities, people might move away
- Many cities in state, experiment only run in a portion of them

Easy Example: Attrition by Movement

People move out of the treatment cities where taxes are raised, cannot be observed

- No longer apples-to-apples comparison
- Cannot only look at the people who stayed
 - We don't know who in the treatment or control groups would have stayed and been observable
 - Without knowing who is observable and who is not, we can't be sure that we are going to conduct an apples-to-apples comparison

Another Example: Massachusetts and New Hampshire

- Many high income people have moved just into New Hampshire
- Along border, high income people tend to be in New Hampshire and low income people in Massachusetts
- We don't want to assess that living there are huge benefits to living in New Hampshire because people tend to be much richer

Harder Example: Attrition by Movement

Example: city-level data of total hours worked in one year for all jobs

- Some cities implement a policy to raise taxes
- Question: do higher taxes lead people to work less?
- May not see attrition in data received
 - Eg. City may give only total numbers of hours worked for all jobs
- Decrease in total hours worked in a treatment city may be driven by people moving away, not by people working fewer hours
- In this example, we only see aggregate hours worked, so we don't even know that there are people missing/attrition has occurred
- Tests for differential attrition may not illustrate that attrition is occurring

Important to consider conclusions when people opt out of being measured in an experiment

- This experiment *WILL* answer what determines how many hours are worked in a city in a given year
- This experiment *WILL NOT* answer what someone's individual behavior is because of differential attrition

Attrition by Survey Nonresponse

Most common cause of attrition

Example: Obama campaign asked "What is the effect of making phone calls to convince people to re-elect Barack Obama?"

Attrition by Survey Nonresponse

Scenario #1

Democrat	Monday Obama Call	Answers Tuesday Survey	Obama Vote
0	0	1	0
0	0	1	0
0	1	0*	0
0	1	0*	0

*Indicates that individual *did not* respond to Tuesday survey

-
- In this scenario:
 - People in the treatment group *did not* answer the survey
 - People in the control group *did* answer the survey

Attrition by Survey Nonresponse

Scenario #2

Democrat	Monday Obama Call	Answers Tuesday Survey	Obama Vote
1	0	0*	1
1	0	0*	1
1	1	1	1
1	1	1	1

*Indicates that individual *did not* respond to Tuesday survey

-
- In this scenario:
 - People in the treatment group *did* answer the survey
 - People in the control group *did not* answer the survey

Attrition by Survey Nonresponse

In both scenarios, we have no treatment effect by making the calls

- Everyone in the treatment group who answered the survey is a Democrat
 - Obama support in treatment group: 100%
- Everyone in the control group who answered the survey is a Republican
 - Obama support in control group: 0%

Calls *appear* to be extremely effective based on results

- Everyone in treatment group *who answered the survey* said they would vote for Barack Obama
- Everyone in the control group *who answered the survey* said they would vote for Mitt Romney

Attrition caused misleading results

- It appears that just calling people will convince all of them to vote for Barack Obama
- In reality, calls (treatment) didn't change minds; they only changed who responded to the survey

Detecting and Preventing Attrition

Average Attrition

"Average treatment effect" on ability to measure the outcome

- Possible test: Code 1 or 0 outcome variable to determine if the outcome was observable
- Is the measurement rate of the outcome the same in treatment and control?
- Possible test: simple regression of response to survey on treatment assignment
 - On average, are people more or less likely to be observable in treatment or control?
 - This test may detect some types of differential attrition

No (zero) average differential attrition can conceal big differential attrition

- Eg. Obama call campaign example from last section:

	Control	Treatment
Democrats	0%	100%
Republicans	100%	0%
Average	50%	50%

Balance on Ex Ante Covariates

Across Treatment Groups

Is there some indication that different groups were more or less likely to be observable differentially in treatment or control?

- Check for balance on ex ante covariates across treatment groups
- Example: Obama campaign has 100,000 phone numbers they intend to call
 - There is an equal distribution of Republicans and Democrats among the people they are calling
 - Campaign conducts treatment and surveys

Is there, on average, a similar likelihood of being a Democrat among those who answered the survey in both treatment and control?

- No, this test would fail. There are more Democrats in the treatment group than the control group
- May wrongly estimate the effect of treatment on the outcome

Balance on Ex Ante Covariates

Across Treatment Groups

- In a political example, look for differences in income or political party
- If treatment causes Republicans to attrit (leave the sample), party affiliation will no longer be balanced in the final sample

$$P \supset Q \rightarrow \neg Q \supset \neg P$$

$$P \supset Q \nrightarrow \neg P \supset \neg Q$$

- \supset : superset
- \rightarrow : implies
- \nrightarrow : does not imply
- \neg : not

Differential Attrition Concealed

Although something could be diagnostic of differential attrition, differential attrition may still exist

- There could see there is no average differential attrition and observable covariates are still balanced, *BUT*,

There could be differential attrition in something you can't measure

- There might be differential attrition due to the treatment
- Especially applies in political and commercial examples when:
 - People voluntarily agree to be measured
 - Treatment may change some peoples' behavior
- We may not be able to observe all the reasons why people chose to take or not to take our survey

Differential Attrition Concealed

Generally:

- Attrition can make us worried
- If attrition happens randomly across treatment and control and treatment groups is fine
- The population in the control is the same as the population in the treatment, even though attrition has occurred
- However, we don't know if it is actually random
 - In the Obama survey case it was not random
 - There was differential attrition by party

A simple way to look at this: Is the average attrition different across groups?

- If we have 80% response from the treatment group and only 50% from the control group, we're going to get different populations
- Even though there is 50% attrition in both treatment and control, you can still have a completely different composition of Republicans and Democrats in the groups
- Initial conclusion: Treatment was effective (calls convinced people to vote for Obama)
- Looking at the covariate balance revealed this was a false conclusion

Differential attrition can be totally undetectable, even when it biases your conclusions

Avoiding Attrition with Surveys

Conceal the purpose of the survey

- Convince people to take the survey without knowing it is related to a treatment they have already received
- In the political calls example, you could ask other questions about other things going on in their state first

Increase response rate over all

- More likely to get more representative sample when response rate is higher
- For example, people are more likely to answer face-to-face surveys because they don't want to be rude
- **Example:** YouTube ads replaced surveys so that people must take the survey
 - Also eliminated differential attrition
 - If forced to take the survey, less likely to take due to treatment

Secure commitments from subjects

- Only do randomized treatment and control groups within those who have committed to take survey

Other Methods to Avoid Attrition

Locate a subject pool unlikely to attrit

- In the income tax study you could choose subjects less likely to move, such as:
 - Subjects over 65
 - Subjects with children

Consider other measurement strategies

- **Example:** study using website survey
- One option is to call everyone who saw the ad and ask if they purchased the product
- A better option is to measure conversions on a website
 - With this, someone would have to clear the cookies on their computer to attrit
 - After seeing the ad, people are less likely to clear their cookies than if asked to answer a survey
- We are more worried about differential attrition in the case of surveys, where people are actively deciding whether to answer, than in the case of website conversions, where the data is being collected passively

Yoga:

Demo of Attrition

Yoga:

Extreme Values Bounds

Review:

Extreme Values Bounds

Mediation

Mediation

Mediation: The study of why a treatment has a certain effect

- Treatment may not have a direct effect on Y
- The mediator is an intervening variable the treatment changes that causes Y to change

$$T \rightarrow M \rightarrow Y$$

Mediation

Example 1: Chewing bark → less pain

- *Why?* Because there is aspirin in the bark (mediator)
- *Why does aspirin work?* Because there is a biological mechanism that reduces pain (new mediator)
- *Why?* ...

Example 2: Fruit juice → cures scurvy

- *Why?* Because there is vitamin C in the bark (new mediator)
- *Why does vitamin C cure scurvy?* Because there is a biological mechanism
- *Why?* ...

Asking questions is important!

- But there are limits to what we know, or even *could know* in many contexts
- Maintain an eye for deeper understanding
- But acknowledge where a usefully shallow understanding can suffice

Mediation

Mediation: The study of why a treatment has a certain effect

- Treatment may not have a direct effect on Y
- The mediator is an intervening variable the treatment changes that causes Y to change

$$T \rightarrow M \rightarrow Y$$

- Never-ending chain of "whys"
- Common criticism of experiments: statistics doesn't tell us why something has an effect
- Always more questions to ask

Typical Approach to Mediation Analysis

DO NOT use this approach because the analysis is typically biased

Example: Lime treats scurvy by increasing vitamin C levels

- Mediator = $\alpha + \beta \times \text{treatment}$
 - Does lime juice lead to higher levels of vitamin C?
- Outcome = $\alpha + \beta \times \text{treatment}$
 - Is there an average treatment effect on the outcome?
 - Does lime juice lead to less scurvy?
- Outcome = $\alpha + \beta_1 \times \text{treatment} + \beta_2 \times \text{mediator}$
 - What is the effect of lime juice (treatment) on scurvy (outcome) once we control for vitamin C levels (mediator)?

Typical Approach to Mediation Analysis

DO NOT use this approach!

Analysis is typically biased

Result: The treatment is no longer effective when we control for the mediator

- However, doesn't definitively establish that the mediator caused the treatment to lose effect
- Mediator was not randomly assigned
- Treatment may have affected other mediators
- Other potential mediators correlated with the mediator that was measured
- May see the same outcome with different mediators

This type of mediation analysis cannot determine if a particular mediator is responsible for why a treatment has an effect

Why Mediation is Difficult to Study

- Mediation analyses are always suggestive, never definitive
- Experimental estimates of indirect effects must affect **only** the mediator in question
 - What is in fruit juice that cures scurvy?
 - Randomly assign vitamin C pills
- In the social sciences, treatment may have several potential causal pathways
- Mediators often aren't directly measurable, such as thoughts or interests
 - Cannot be sure which mediators are being changed with the treatment
 - Cannot be sure that all possible mediators have been measured
- Causal heterogeneity
 - Subjects are differentially affected by changes in X and M

These "why" questions are really difficult to study in general

Program of Research

Mediation usually isn't critical but suggests more research

- **Example:** Fruit juice cures scurvy
- The reason fruit juice cures scurvy may not matter
- However, mediation analysis may help us to refine our knowledge
- For example it may help us be able to send vitamin-C pills on ships rather than crates of limes

Mediation analyses can guide thinking and suggest hypotheses for future research

- But will not be certain a particular mediator is the cause of why a treatment had an effect

Implicit Mediation Analysis

Theory testing by looking for subgroup effects

- **Example:** Politics
 - *Theory:* people are more likely to adopt the views of politicians of their own parties
 - *Treatment group:* receives letter from Democratic politician stating his position on an issue
 - *Control group:* does not receive letter
- Theory leads you to believe Democrats will accept the position in the letter more than Republicans
- Theory is tested but is not dispositive ; there may be other reasons Democrats adopted the position in the letter
- Cannot be certain why a treatment had an effect
- Tests theories to see how consistent they are

Bolger and Amarel Study

Social psychologists hypothesized: Social support reduces stress due to an increase in self-confidence

- Common and ineffective way to study this: Measure self-confidence after someone has received social support
- Find that the effect of social support goes away when we control for post-treatment self-confidence

Bolger and Amarel study

- Randomly assign the mediator
- Conducted an implicit mediation analysis with an experiment that randomly assigns whether an increase in self-confidence is present
- Randomly assigned a confederate (research assistant who posed as a peer) to either speak in a self-confidence-enhancing way or not after giving subjects a stressful task

Bolger and Amarel Study

Result

- Confidence-enhancing support worked
- Non-confidence enhancing support did not work

What we don't necessarily learn

- Whether confidence itself is responsible for reduced stress
- For example, maybe the words used to enhance confidence also affected happiness which reduced stress

Takeaway from this study

- It may be a good strategy to use words from the confidence-enhancing support to reduce stress in other settings
- Eg. hospital patients

Mediation analysis

- can be useful for suggesting other treatments
- can never be dispositive about what causes a treatment to work in the real world

Generalizability

Generalizability: The limitations of any one experiment to apply to other subjects in other contexts

Generalizability of Subjects

In an experiment, results apply to the estimated causal effects only among the given sample

- For example if I walk down the street and do an experiment, the kind of people I meet on the street may show a different effect than if I called people in India on the phone or walk down the street in Australia
- Many differences across people and areas that might lead us to different conclusions
- Eg. Heterogeneous treatment effects

Example: Cancer Drug Trials

- New cancer drug that has not been tried in humans
- People volunteer for medical trial and are assigned to either treatment or control
- Volunteers are already in various serious conditions
- Results show a large positive treatment effect among dire medical cases
- Results may not apply to patients with a more mild stage of cancer
- Medical trials need a wide variety of populations in different stages of the disease

Consider how volunteers may differ from the broader population when patients self-select into an experiment

- Results for dire patients may not generalize to patients with milder cancer

Generalizability of Treatments

- Treatments first tested in research hospitals, with highly trained and educated staff, tend to be more effective than when tested in other hospitals
 - Research hospital doctors may be better at following the protocol and managing patients with side effects
 - Therefore, treatment is different in research hospitals compared to others

Generalizability of Treatments

Example: Job Training

Will the a job-training program that shows effectiveness in an experiment work if used company-wide?

Experimental Study

- Study to evaluate effect of a job training program
- Send letters to job training programs asking them to randomly choose 500 people to be evaluated
- 20 programs agree to participate
- Results show all programs highly effective in placing people in jobs
- However, selection of programs was not random
 - 1,000 sites were asked to participate
 - Only 20 sites agreed to be evaluated
 - Maybe only those programs who knew they were effective agreed to participate

Example: Opower Program

Allcott (2014)

Program sent people letters that compared their electric usage with that of their neighbors

- People who were using more electricity than their neighbors reduced energy consumption over the next few months
- Environmentally conscious areas were the first to sign up for the initial pilot program
- The program in these areas showed large effects
- Subsequent cities signed up and showed smaller effects
- People and partners most willing to try a new intervention may be most likely to benefit
- Reverse can be true too; those least willing to give up something may benefit the most

Consider how results might change when you try an experiment with another population

Example: Auction Study

Lucking-Reiley (1999)

Reported results showing that Dutch auctions raised more revenue than first-price sealed-bid auctions

- *Dutch auction*: price continues to lower until a bidder accepts that price
- *First-price sealed-bid auction*: bids simultaneously submitted, highest bid wins
- Does the same result hold for goods other than *Magic: The Gathering* cards?
- Does the result hold for in-person bidding, or bidding on websites, instead of just bids submitted via e-mail?
- Does the results hold when the Dutch auction is fast rather than slow?
 - Reiley's auctions had one price change per day; flower auctions in Holland change price in fractions of a second

There are many ways that a particular experiment's results might differ across contexts

Example: Online Advertising

Lewis and Reiley (2014)

Reported that online advertising increased in-store purchases

- Do the results hold for other retailers?
- Do the results hold when the subjects are not existing customers of the retailer?
- Do the results hold for advertising done on sites other than *Yahoo!* ?
- Do the results hold for offline advertising?
- Were the results specific to the actual ad creatives used?
- Do the results hold in time periods other than fall 2007?

In general, we don't know if the results generalize until we replicate

Example: Get Out the Vote Campaigns

Green, Gerber, and Larimer (2008)

- Professors tried a strategy in an off-year Michigan primary election where few people typically vote or know about the election
 - **Results:** Get Out the Vote strategy is very effective
 - But, when treatment was tried in presidential election, the effects were much smaller
- How treatment effects differ across people and across context may not be obvious until we conduct further experiments
 - After many experiments, Get Out the Vote can now make empirical generalizations
 - Hundreds of experiments allowed them to find regularities so we can accurately forecast causal effects

Key Takeaway

No experiment applies to all circumstances

Summary:

Attrition, Mediation, Generalizability

Attrition

Treatment has an effect on which units are being observed

- Ruins comparability

Strategies for detecting attrition:

- Is there an average effect on attrition on whether people's outcomes are measurable?
- Are covariates still balanced among those for whom the outcomes are measurable?

Try to reduce attrition in design

- Design a measurement procedure that is less sensitive to attrition
- Eg. measure behavior rather than using surveys

Mediation

The study of *WHY* a treatment has an effect

Fundamentally unanswerable question

- Cannot know for sure why a treatment worked

"Implicit mediation" analysis can guide hypothesis generation

- Eg. Limes vs vitamin C example
- A way to generate hypotheses more than anything that can ultimately be answered for sure

Generalizability

Your results pertain only to the sample and context of that particular study

- Cannot ever know for sure how your experiment will differ in other contexts
- Perform other experiments to validate that your results would replicate in other circumstances
- With many experiments you can start to build more general theories