DISCOVERY **#3: Observational Studies & Confounders**

January 18, 2019

Karle Flanagan and Wade Fagen-Ulmschneider

CONTROLLED EXPERIMENTS	**OBSERVATIONAL STUDIES**
The researcher decides how to divide the subjects into treatment and control groups. Can be randomized and non-randomized.	The researcher has NO power over assignment into treatment and control groups. Subjects themselves or simply fate determines who gets the treatment and who doesn't, the researcher just observes what happens.

Observational studies are done out of necessity. Whenever possible, it's better to do a randomized controlled experiment. Why?

Main Problem with Observational Studies: They can show an association, but it's difficult to make conclusions about causality. Since the treatment and control groups just "happened" they are often very different from each other. These differences confound (mix up) the results when you try to reach a conclusion.

An underlying difference between the two groups (other than the treatment) is called a confounding factor. *Confounding factors (or confounders) are common in observational studies. Good studies take great care to reduce confounding*.

Possible **confounders** make it difficult to prove **causation** by **association**. Did the treatment **cause** the response or is the treatment simply **associated** with the response? Maybe both treatment and response were caused by a third **confounding factor**.

*If the treatment truly does cause the response, there will be a causal link explaining how or why the treatment itself is causing the response.

*If the treatment does not cause the response, then there could be a confounder that's making it look like the treatment is causing the response.

PRACTICE PROBLEMS

Example 1: Coffee and lung cancer. Studies in the 1960s showed that coffee drinkers had higher rates of lung cancer than those who did not drink coffee. Because of this, some people identified coffee as a cause of lung cancer. Do you think this is true?

Example 2: A recent study was done and found that people who wear sunscreen regularly were **more likely** to get skin cancer than those who do not. This goes against what doctors had originally thought. Answer the following questions below.

a) What type of study is this?

b) Does the study show that sunscreen causes skin cancer?

c) Based only on the info below, classify the following as confounders, causal links, or neither.
   i) Chemicals in Sunscreen- When applying sunscreen to your skin, you're exposing yourself to certain harmful chemicals that may cause skin cancer.

   ii) Genetics- Cancer is known to run in some families, making it more likely for you to get skin cancer whether you wear sunscreen or not.

   iii) Health problems- Studies have shown that too much sun exposure is bad for your health.

   iv) Fair Skin- People with fair or lighter skin easily get sunburned and therefore wear more sunscreen to prevent this. Also, people with fair skin have a higher risk of getting skin cancer.

Example 3: One day, while scrolling through Facebook this summer, I came across an article entitled: "The Secret to a Long Marriage Is Drinking Together." As a newlywed, I quickly became interested and clicked on the article. It stated that couples who drink together, stay together. Over 2,000 couples were involved, and they found that couples who reported drinking alcohol together even just a few times each year were less likely to get on each other's' nerves and more likely to have a positive outlook on their marriage. The study highlighted that what's most important isn't how much the couples drink, but whether they BOTH drink. If both partners drank, they were more likely to have a happier marriage than if just one of them drank.

a) Based only on the information above, this study is an example of

b) What can we conclude from this study?
   i) We see that there is an association between drinking alcohol together and happy marriages, however, we aren't sure if there is a causal relationship- there could be other variables confounding the data.
   ii) Drinking alcohol together causes couples to stay together! Steve and I should drink together as much as possible if we want to remain happy in our marriage.
   iii) Drinking alcohol and happy marriages are not related to each other at all.
   iv) When one partner drinks and the other doesn't, it causes marital problems.

c) Based only on the info below, state whether the following are confounders, causal links, or neither:
   i) Shared Interests- Couples with shared interests are both more likely to drink together and to engage in other activities together that lead to happier marriages.

   ii) Endorphins- Alcohol releases endorphins making you feel happier which leads to happier marriages.

   iii) Commitment - Marital happiness depends on a strong commitment from both partners to continually work on resolving their problems rather than running away from them.

   iv) Communication—Clear communication about expectations going into the marriage and the long-term goals contribute to longer lasting and happier marriages.

How to deal with confounders:
STRATIFICATION- Statisticians adjust for these confounding variables by breaking the control and treatment groups into **smaller more homogeneous sub-groups, where the confounding factor is the same.** This is called **stratification**. Stratification plays a similar role in observational studies as blocking does in randomized experiments.

Back to the blocking example from before: *This time as an observational study.*
Does computer-graded homework improve final exam scores compared to hand-graded?
Suppose the 20 students in data science DISCOVERY were allowed to *choose* themselves (instead of by random assignment) to be in the computer or hand graded homework group. Let's say 8 "A" students and 2 "B" students chose computer grading, while 2 "A" students and 8 "B" students choose hand grading.

To eliminate the effect of the confounder we would **stratify on the confounding variable and compare the A and B students separately.**

With observational data you need a much bigger sample size than you do with experimental data because each time you stratify for a possible confounder the comparison groups get smaller and smaller, leaving more room for chance error.

*How would you stratify to account for a second confounding variable?

SUMMARY

In this course, you will learn how to conduct and analyze your own randomized experiments. That will involve more detail than has been presented in this section. For now, just focus on the main idea: to try to establish causality, run a randomized controlled experiment if possible. If you are conducting an observational study, you might be able to establish an association but not causation. Be extremely careful about confounding factors before making conclusions about causality based on an observational study.

Stratification helps removes confounding in observational studies by comparing sub-groups for which the confounding variable is the same.

References

1. *Stat 100 Incomplete Lecture Notes Workbook by Ellen Fireman, Karle Flanagan, & John Marden*
2. *Stat 200 Incomplete Lecture Notes Workbook by Ellen Fireman, Karle Flanagan, & John Marden*
3. *Data 8 Online Textbook: The Foundations of Data Science by Ani Adhikari & John DeNero*