# Study Design

Ryan Miller

# Outline

1. Study design
2. Confounding variables
3. Randomized experiments

# Introduction

Consider three different hypothetical studies aimed at understanding the relationship between sun exposure and skin cancer:

1) Researchers identify people with and without skin cancer and ask them about their past frequencies of sun exposure
2) Researchers recruit willing participants, track their sun exposure over time, then see who develops cancer
3) Researchers recruit willing participants, randomly assign half of them to increase their sun exposure, then see who develops cancer

Assuming an association is found, which study design provides the *strongest evidence*? Are all of these study designs ethically viable?

# Study design

When interpreting associations found within a dataset, we need to consider the *study design*:

1) **Experimental study**, the explanatory variable is *manipulated* by the researcher, then the response is observed
2) **Observational study**, explanatory and response variables are both *observed* as they naturally occur

Each type of design has advantages and disadvantages.

# Two types of observational studies

- **Retrospective** studies identify cases then collect data on things that have already happened
  - For example, identifying cancer patients and determining their exposure to environmental toxins

# Two types of observational studies

- **Retrospective** studies identify cases then collect data on things that have already happened
  - For example, identifying cancer patients and determining their exposure to environmental toxins
- **Prospective** (or cohort) studies identify cases then follow them forward in time, collecting data as outcomes occur
  - For example, recruiting many young people, tracking their exposures, then waiting to see who develops cancer

# Two types of observational studies

- **Retrospective** studies identify cases then collect data on things that have already happened
  - For example, identifying cancer patients and determining their exposure to environmental toxins
- **Prospective** (or cohort) studies identify cases then follow them forward in time, collecting data as outcomes occur
  - For example, recruiting many young people, tracking their exposures, then waiting to see who develops cancer
  - Prospective studies are generally considered to be *stronger evidence* than retrospective studies

# Practice

In 1980, researchers at the University of Chicago collected data on all murders that took place during a felony in the state of Florida between 1972 and 1977. The researchers were interested in racial bias in the administration of the death penalty following the Civil Rights Act.

- Link to the data

1) How would you describe the type of study design used by these researchers?
2) Using StatKey, find the proportion of black offenders and the proportion of white offenders who were sentenced to death. Do you see evidence of racially biased sentencing?

# Practice (solution)

1) This is a retrospective observational study, as the verdicts were already determined when the data were collected and the explanatory variable (offender's race) was not manipulated by the researchers.
2) 21.1% of black offenders and 23.2% of white offenders were sentenced to death, which does not seem to indicate racially biased sentencing.

# Multivariate relationships

It's reasonable to presume the race of the victim could also impact the sentencing verdict. I've *stratified* these data into two subsets defined by the victim's race:

- Link to cases involving a white victim
- Link to cases involving a black victim

1) For the cases involving a *white victim*, how do death penalty rates compare across white and black offenders?
2) For cases involving a *black victim*, how do death penalty rates compare across white and black offenders?

# Confounding variables

- A **confounding variable** is one that is associated with *both* the explanatory and the response variable in an analysis

# Confounding variables

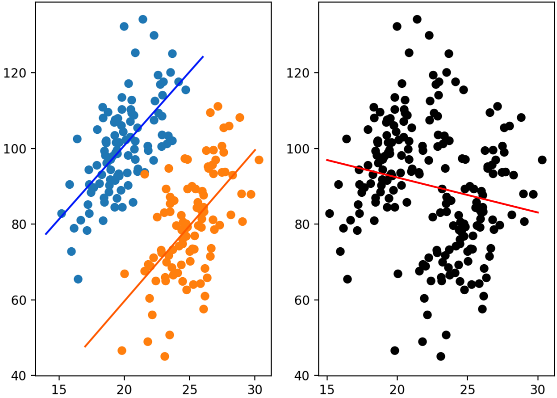
- A **confounding variable** is one that is associated with *both* the explanatory and the response variable in an analysis
  - A confounding variable will *obscure* the true relationship between explanatory and response variables
  - **Stratification** is one method for *controlling* for a confounding variable

# Confounding variables

- A **confounding variable** is one that is associated with *both* the explanatory and the response variable in an analysis
  - A confounding variable will *obscure* the true relationship between explanatory and response variables
  - **Stratification** is one method for *controlling* for a confounding variable
- *All* confounding variables must be controlled for in order to make a reliable claim of **causation**
  - The Bradford Hill criteria provide a framework for determining causation using observational data

# Simpson's paradox

**Simpson's paradox** occurs when the the impact of a confounding variable is so severe that it *reserves* a trend that was observed prior to stratification



# Practice

In the early 1970s, administrators at UC-Berkeley grew concerned that females were being discriminated against in admissions to the university's graduate programs. In 1973, 44% of the 8442 male applicants were admitted, while only 35% of the 4321 female applicants were admitted.

# Practice

In the early 1970s, administrators at UC-Berkeley grew concerned that females were being discriminated against in admissions to the university's graduate programs. In 1973, 44% of the 8442 male applicants were admitted, while only 35% of the 4321 female applicants were admitted.

| Department | All | | Men | | Women | |
|---|---|---|---|---|---|---|
| | Applicants | Admitted | Applicants | Admitted | Applicants | Admitted |
| **A** | 933 | 64% | ***825*** | 62% | 108 | **82%** |
| **B** | 585 | 63% | ***560*** | 63% | 25 | **68%** |
| **C** | 918 | 35% | 325 | **37%** | ***593*** | 34% |
| **D** | 792 | 34% | 417 | 33% | 375 | **35%** |
| **E** | 584 | 25% | 191 | **28%** | ***393*** | 24% |
| **F** | 714 | 6% | 373 | 6% | 341 | **7%** |

1) Identify the explanatory, response, and confounding variable in the stratified table presented above
2) Explain *why* the overall admission rates are so different than those within each department

# Practice (solution)

1) Sex is the explanatory variable, admission is the response variable, and program/department is the confounding variable.
2) Men more frequently applied to less competitive programs (such as A and B) that admitted both men and women at high rates, thereby "inflating" their overall admissions rate relative to that of women.

# Balance

The underlying issue created by confounding variables can be described as **imbalanced groups**

- In the death penalty example, offenders were more likely to victimize their own race (with crimes against whites being punished more harshly)
  - This led to the groups of white offenders and black offenders being *systematically different* in an *important way* (victims race)

# Balance

The underlying issue created by confounding variables can be described as **imbalanced groups**

- In the death penalty example, offenders were more likely to victimize their own race (with crimes against whites being punished more harshly)
  - This led to the groups of white offenders and black offenders being *systematically different* in an *important way* (victims race)
- In the UC-Berkeley example, men were more likely to apply to less competitive programs (which had higher overall admissions rates)
  - This led to the groups of male and female applicants being *systematically different* in an *important way* (department)

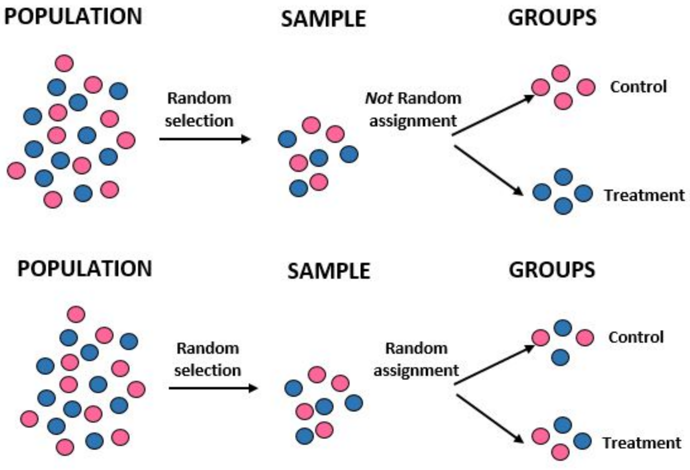
# Obtaining “balanced” groups

- **Stratification** is a way of *forcing* balance
  - For example, when we compare male and female admissions within department A, both groups cannot be imbalanced in terms of department (since they’ve all applied to department A)

# Obtaining "balanced" groups

- **Stratification** is a way of *forcing* balance
  - For example, when we compare male and female admissions within department A, both groups cannot be imbalanced in terms of department (since they've all applied to department A)
- **Random assignment** of the explanatory variable is another way to achieve balanced groups
  - On average, both groups (defined by the explanatory variable) will have the *same distribution* for *any* potential confounding variables
  - Random assignment is *only* possible in an *experimental design*

# Random assignment



# Example #1 - Feasibility of Randomized Experiments

Suppose we want to know: "Is arthroscopic surgery is effective in treating arthritis of the knee?" Describe both an *observational study* and a *randomized experiment* that you could conduct to answer this question. Be sure to address the following during your discussion:

1. How costly will it be for the researchers to collect data with each design?
2. Are there any feasibility problems or ethical issues with each design?

# Sham knee surgery

*In the 1990s a study was conducted in 10 men with arthritic knees that were scheduled for surgery. They were all treated identically expect for one key distinction: only half of them actually got surgery! Once each subject was in the operating room and anesthetized, the surgeon looked at a randomly generated code indicating whether he should do the full surgery or just make three small incisions in the knee and stitch up the patient to leave a scar. All patients received the same post-operative care, rehabilitation, and were later evaluated by staff who didn't know whether they had actually received the surgery or not. The result? Both the sham knee surgery and the real knee surgery showed indistinguishable levels of improvement*

Source: https://www.nytimes.com/2000/01/09/magazine/the-placebo-prescription.html

# Vocabulary for randomized experiments

The Sham Knee Surgery example illustrates several important aspects of a well-designed experiment that we've yet to discuss:

- **Control Group** - Some patients were randomly assigned not to receive the knee surgery, providing a comparison group that is, on average, balanced with surgery group in all baseline characteristics
- **Placebo** - Patients in the control group received a fake surgery
- **Blinding** - Using a placebo is not helpful if patients know which group they're in. Similarly, the staff interacting with the patients might treat them differently if they knew the patient's group
    - **Single-blind** - the participants don't know the treatment assignments
    - **Double-blind** - the participants *and* everyone interacting with the participants don't know the treatment assignments

# Example #2 - Can Randomization Fail?

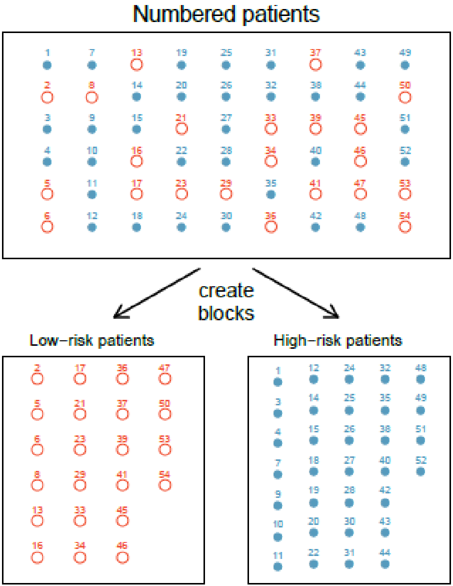
- A University of Iowa researcher was conducting an experiment on lab monkeys
- Lab monkeys are expensive, so his experiment had $n = 8$
- Having taken a statistics course, he randomly assigned treatment/control groups
- After conducting the experiment and seeing surprising results, the researcher recognizes that the 4 monkeys in the control group were also the oldest 4 monkeys
- The researcher knew that the age of the monkey had an important on the outcome variable, but he expected randomization to handle that

Should he report his results? What could he have done differently?

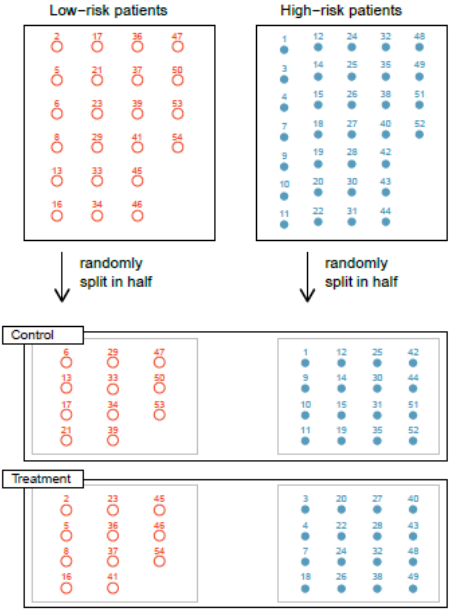
# Blocking

- Randomization is not guaranteed to properly balance the treatment and control groups unless the sample size is relatively large
- At smaller sample sizes, strategies such as **blocking** can be used
  - In this design, cases are first split using a **blocking variable**, then random assignment is done within each block
  - This ensures the blocking variable is balanced in each group

# Blocking


Numbered patients
create blocks
Low−risk patients
High−risk patients

# Blocking


Low–risk patients
High–risk patients
randomly
split in half
randomly
split in half
Control
Treatment

# Example #3 - Limitations of Randomization

Even with a large sample, randomization might not produce the intended results:

- Police departments have long been uncertain about how to best respond to cases of domestic abuse
- Minneapolis Police conducted a study comparing three different response strategies:
  - Arrest
  - Advice
  - Seperate

# Example #3 - Limitations of Randomization

Even with a large sample, randomization might not produce the intended results:

- Police departments have long been uncertain about how to best respond to cases of domestic abuse
- Minneapolis Police conducted a study comparing three different response strategies:
  - Arrest
  - Advice
  - Seperate

Officers were randomly assigned a strategy to use on each case, but they were given discretion to change the strategy if necessary. Precautions were taken to ensure the officers were as faithful as possible to each assigned strategy. The outcome of interest was whether or not violence reoccurred.

# Minneapolis Police Case Study

The columns of the table indicate the strategy assigned, the rows indicate the strategy actually used:

| | Arrest | Advice | Seperate |
|---|---|---|---|
| Arrest | 91 | 18 | 26 |
| Advice | 0 | 84 | 5 |
| Seperate | 1 | 6 | 82 |

Overall there was 82% adherence to the randomly assigned strategy, but do you see any problems?

Minneapolis Police Article Link

# Minneapolis Police Case Study

- A common pattern was to “upgrade” to the arrest strategy
- The advice and separate groups likely lost their highest risk members to the arrest group
- This seemingly well-designed experiment ended up needing to be bailed out by complex statistical approaches used to jointly model reoccurance of violence (the outcome variable) and adherence to the randomized strategy

# The Intention-to-Treat Principle

- The Food and Drug Administration (FDA) mandates an **intent-to-treat principle** (ITT) as the primary design and analysis strategy for clinical trials
- This means that all subjects who are randomized be included in the final analysis, even if they, cross-over, do not adhere to any protocol, or drop out of the study
- It also means that clinical trials estimate the effect of the *treatment assignment* rather than the treatment itself

Intent-to-treat Article Link

# Intention-to-Treat Example

Below are the results of the MN police case study *as randomized*:

| | Recurrence | No Recurrence |
|---|---|---|
| Arrest | 10 | 82 |
| Advice | 24 | 84 |
| Seperate | 26 | 87 |

Below are the results *by treatment used*:

| | Recurrence | No Recurrence |
|---|---|---|
| Arrest | 18 | 117 |
| Advice | 16 | 73 |
| Seperate | 26 | 63 |

With your group, compare the difference in recurrence proportions of the Advice and Separate treatments using an ITT analysis. Compare this with the same analysis done using the treatments used.

# Intention-to-Treat Example - Solution

Using ITT, $p_{re|advice} = 24/106 = 0.22$ and $p_{re|separate} = 26/113 = 0.23$

- We can conclude that *telling* an officer to use the “Advice” strategy leads to essentially the same outcomes as telling the officer to use the “Separate” strategy

Using as treated, $p_{re|advice} = 16/89 = 0.18$ and $p_{re|separate} = 26/89 = 0.29$

- We have trouble concluding anything; while there is a clear difference, we don’t know if it is due to the strategies themselves, or a disproportionate switching of high/low risk cases into different strategies

# Study Design and Statistical Inference

- The overarching goal of a statistician is to *rule out* possible explanations for an observed association

# Study Design and Statistical Inference

- The overarching goal of a statistician is to *rule out* possible explanations for an observed association
- The following explanations can be addressed by study design:
  - **Sampling bias** - Simple random sampling
  - **Confounding variables** - Random assignment, stratification
  - **Other sources of bias** - Placebo, double-blinding, etc.

# Study Design and Statistical Inference

- The overarching goal of a statistician is to *rule out* possible explanations for an observed association
- The following explanations can be addressed by study design:
  - **Sampling bias** - Simple random sampling
  - **Confounding variables** - Random assignment, stratification
  - **Other sources of bias** - Placebo, double-blinding, etc.
- Ideally, we can use careful study design to reduce the viable explanations to either **random chance** or a **real relationship**

# Summary

1. Observational designs
    - Study design is an additional cause for concern when trying to infer causation
2. Confounding variables
    - In observational designs, confounding variables can obscure the underlying relationship between explanatory and response variables
3. Randomized experiments
    - Random assignment of the explanatory variable eliminates the possibility confounding variables
    - Other measures should be taken to prevent bias (ie: placebo, blinding, etc.)