

# Study design

Patrick Breheny

January 22, 2026

# Sampling in the ideal world

- In an ideal world,
  - We have a list of everyone in the population of interest
  - We randomly sample these people
  - It is equally costly to sample one person as it is another
  - No one ever refuses to be sampled

# Sampling in the real world

- In the real world, however, none of these things happen:
  - We have to get access to people somehow in order to sample them
  - It is often more cost effective to sample certain groups of people than other groups of people
  - Not all people are equally likely to participate in your study
- All of these factors may create bias

## Example: The 1936 Presidential Election

- In 1936, Franklin Delano Roosevelt was completing his first term of office as president of the United States, and was up for re-election
- The Republican candidate that year was Alfred Landon
- The country was still in the Great Depression, with 9 million unemployed and real income only two-thirds of what it was in 1929
- Roosevelt and Landon had different views about how active the government should be in terms of enacting policies to bring the country out of the depression

## Example: The 1936 Presidential Election (cont'd)

- The *Literary Digest* magazine had predicted the winner in every presidential election since 1916
- For the 1936 election, the *Digest* sampled 2.4 million people and predicted a landslide victory for Landon: 57% to 43%
- In the actual election, Roosevelt won 62% to 38%

# What went wrong?

- How could this poll have been so incredibly far off?
- The poll had an enormous sample size; variability of the estimate is not the issue
- Instead, the flaw was bias

# Where the bias came from, Part I

- The *Digest* mailed 10 million questionnaires to addresses gathered from telephone books and club membership lists
- This tended to screen out the poor, who were less likely to belong to clubs or to own telephones (at the time, only one out of every four households owned a telephone)
- This is called *selection bias*: instead of random sampling, certain subgroups of the population were more likely to be included than others

## Where the bias came from, Part II

- The *Digest*'s poll contained another flaw: only 2.4 million people replied, out of the 10 million who got the questionnaire
- Nonresponders can differ from responders in many important ways
- This type of bias is called *nonresponse bias*
- Thus, the 2.4 million respondents in the *Digest*'s poll do not even represent the 10 million people who were polled, let alone the entire population of voters

# Pharmaceutical trials and children

- Sample estimates only describe the population they are sampled from; attempts to generalize to other populations may be biased
- The issue of pharmaceutical trials and children is a biostatistical example of this last point
- For practical, ethical, and economic reasons, clinical trials usually involve only adults – children are excluded (only about 25% of drugs are subjected to pediatric studies)
- Physicians, however, are allowed to use any FDA-approved drug in any way that they think is beneficial, and aren't required to inform parents if the therapy hasn't been tested on children

# Propofol

- For example, propofol is a sedative that has consistently proved safe in adults
- In 1992, after several children who received propofol in the ICU died, the British government recommended against using it on patients under 16
- In the U.S., however, propofol continued to be widely used
- In 2001, the manufacturers of propofol conducted a randomized, controlled trial and found increased risk of death in children treated with propofol
- In response, the FDA added a warning to indicate this, although the administration of propofol to children in the ICU remained legal (and still is, with other studies finding no increase in risk . . . the issue remains controversial)

# Interpretation

- We'll discuss this more next week, but in this class I will often be giving you data from a study, then asking you to analyze it and finally, to provide a one-sentence summary of the main findings
- In doing so, it is always important to keep in mind the intended population of the study
- For example, this summary is a bit problematic: "There is significant evidence that Drug X reduces blood pressure"
- Providing context about the population is important: "There is significant evidence that Drug X reduces blood pressure *in adults with hypertension*"

# The method of comparison

- Identifying patients to sample is not the only important question in study design, however
- Suppose a new therapy is developed; how should we design an experiment to test whether or not it is effective?
- The only meaningful way to do this is to compare it with something else
- Thus, we are going to have to obtain two samples (or one sample, then split it into two groups):
  - The therapy is given to subjects in the *treatment group*
  - The other subjects are not treated and are used as *controls*

# Assigning treatment

- Subjects should be assigned to treatment or control *at random*
- Furthermore, the experiment should be *double blinded*:
  - The subjects should not know which group they are in
  - The doctors should not know which group the subjects are in

# The U.S. polio epidemic

- A polio epidemic hit the United States in 1916
- During the next forty years, hundreds of thousands of people, especially children, fell victim to the disease
- By the 1950s, several vaccines had been developed
- One in particular, developed by Jonas Salk, seemed very promising based on laboratory studies

# Field trials

- In 1954, the Public Health Service organized an experiment to see whether the Salk vaccine would protect children from polio outside the laboratory
- The subjects were children in the most vulnerable age groups: grades 1, 2, and 3
- Two million children were involved: some were vaccinated, some refused treatment, and some were deliberately left unvaccinated

# Ethical issues

- This raises issues of medical ethics, which are always a consideration in medical studies
- Is it ethical to leave some children deliberately unvaccinated?
- Maybe a more ethical design would be to offer the vaccine to all children, and the children whose parents refused vaccination would serve as the controls

# Confounding

- There is a big problem with the “ethical” design
- Higher-income parents are more likely to consent to treatment, and their children are more likely to suffer from polio
- The reason for this is that children from poorer backgrounds are more likely to contract mild cases of polio early in childhood, while still protected by antibodies from their mothers
- Thus, differences between treatment and control groups could be due to parental income, not the treatment
- Family background here is said to be a *confounding factor*; confounding is a major source of bias

# Making treatment and control groups similar

- So, the “ethical” design isn’t really all that ethical, in the sense that it won’t correctly determine whether the vaccine works or not
- To avoid bias and confounding, it is important that the treatment and control groups be as similar as possible — except for the treatment
- It is necessary, then, that both treatment and control groups be chosen from the same population: children whose parents consented to treatment

## Ethical issues, Part II

- This means that in order to truly learn whether the vaccine works or not, it is necessary to withhold the vaccine from consenting patients
- This certainly raises ethical considerations, but it is important to remember that when a new therapy first emerges, no one really knows whether or not its benefits will outweigh its risks
- This uncertainty is what justifies withholding treatment from the control group

## How to choose who receives treatment?

- As we have said, it is important that the treatment and control groups be as similar as possible except for the treatment; but how should we decide which children go in which group?
- One approach would be to use human judgment to try to make the treatment and control group as similar as possible with respect to all the relevant variables
- Experience shows, however, that this is still flawed
- Not only does human judgment often result in unseen bias, but we are also limited to balancing on confounders that we know about

# Randomized controlled trials

- It is much better to use a carefully designed random procedure
- For the Salk vaccine trial, this is equivalent to flipping a coin: heads, the child gets the vaccine; tails, the child does not
- Experiments in which an impartial chance procedure determines whether a subject is placed into the treatment or control group are called *randomized controlled experiments*

# Placebos

- Another basic precaution is the use of a *placebo*
- In the Salk vaccine trial, children in the control group were given an injection of saline solution
- Therefore, the children did not know whether they had received the treatment; this ensures that their response is due to the vaccine itself, not the idea of treatment
- This may not seem important, but the placebo effect can be surprisingly strong

# Double-blinding

- To see whether or not polio was being prevented, physicians would have to determine whether or not the children had contracted polio
- Many forms of polio are hard to diagnose, and borderline cases could be influenced by a physician's knowledge of whether the child was vaccinated
- So, another precaution taken in the Salk vaccine trial is that the doctors were not told which group the child belonged to
- Thus, neither the subjects nor the doctors knew who was in the treatment group and who was in the control group
- The Salk vaccine trial was therefore a randomized controlled double-blind experiment
- This is pretty much the best design there is

# The results of the trial

|            | Size of group<br>(n) | Polio cases per<br>100,000 children |
|------------|----------------------|-------------------------------------|
| Treatment  | 200,000              | 28                                  |
| Control    | 200,000              | 71                                  |
| No consent | 350,000              | 46                                  |

$$\text{Rate} = \frac{\#\text{Cases}}{n} \cdot 100,000$$

## Another design

- Randomized controlled double-blind experiments are now recognized to be the gold standard for experiments, but this was not the case in the 1950s
- There was a lot of disagreement over the best way to design this study
- In addition to the design we just talked about, a second design proposed by the National Foundation for Infantile Paralysis (NFIP) was carried out
- In the NFIP design, all second graders would be offered the vaccine, and children in grades 1 and 3 would serve as controls

# Results of both trials

## Randomized controlled double-blind experiment

|            | Size    | Rate |
|------------|---------|------|
| Treatment  | 200,000 | 28   |
| Control    | 200,000 | 71   |
| No consent | 350,000 | 46   |

## NFIP study

|                        | Size    | Rate |
|------------------------|---------|------|
| Grade 2 (vaccine)      | 225,000 | 25   |
| Grades 1 & 3 (control) | 725,000 | 54   |
| Grade 2 (no consent)   | 125,000 | 44   |

# Conclusions

- We observed lower incidence of Polio in the vaccine group than the placebo group; what could be causing it?
  - Confounding?
  - Perception bias?
  - Diagnostic bias?
  - Chance?
- The incidence of Polio could be lower in the vaccine group simply by random chance
- Next week, we will investigate whether or not this is a plausible explanation

# Clofibrate

The Coronary Drug Project Research Group published an article in the *New England Journal of Medicine* (1980) describing a randomized controlled double-blind experiment involving the drug clofibrate, which reduces the level of cholesterol in the blood

|             | Clofibrate |        |
|-------------|------------|--------|
|             | Number     | Deaths |
| Adherers    | 708        | 15%    |
| Nonadherers | 357        | 25%    |
| Total       | 1,065      | 18%    |

Subjects who took more than 80% of their prescribed medicine were called “adherers”

# Interpreting the clofibrate results

- This looks like strong evidence that clofibrate is effective, but caution is in order
- Subjects were randomized with respect to whether they received the drug; they were **not** randomized with respect to their adherence
- Thus, confounding is possible

# Clofibrate and placebo results

|             | Clofibrate |        | Placebo |        |
|-------------|------------|--------|---------|--------|
|             | Number     | Deaths | Number  | Deaths |
| Adherers    | 708        | 15%    | 1,813   | 15%    |
| Nonadherers | 357        | 25%    | 882     | 28%    |
| Total       | 1,065      | 18%    | 2,695   | 19%    |

- Taking into account the placebo results as well, clofibrate no longer looks effective
- One possibility is that adherers are more concerned with their health, and take better care of themselves in general
- Take-home message: comparing subjects *as they were randomized* is the only completely valid way of carrying out a controlled experiment; all other comparisons are subject to confounding and bias

# Portacaval shunts

- Patients with cirrhosis of the liver may start to hemorrhage and bleed to death
- One treatment involves surgery to redirect the flow of blood through what is called a *portacaval shunt*, a long and hazardous operation
- A bunch of studies were done in the 1950s and 1960s trying to determine whether the benefits of this surgery outweighed its risks

# Portacaval shunt studies

---

| Design                       | Degree of enthusiasm |          |      |
|------------------------------|----------------------|----------|------|
|                              | Marked               | Moderate | None |
| No controls                  | 24                   | 7        | 1    |
| Controls, but not randomized | 10                   | 3        | 2    |
| Randomized controlled        | 0                    | 1        | 3    |

---

# Conclusions

- The poorly designed studies greatly exaggerated the value of the surgery
- One possible explanation is that in an experiment without randomized controls, many physicians have a natural tendency to treat only the patients who are in relatively good shape
- This biases the study in favor of the treatment

# Diethylstibesterol

- DES (Diethylstibesterol) is an artificial hormone used to prevent miscarriage in pregnant women
- Five studies of DES were carried out using “historical controls” (outcome rates of patients from the past); all had favorable conclusions regarding the value of the therapy
- Three randomized controlled designs were carried out, and all were negative about the value of DES
- Doctors paid attention to the positive studies and ignored the randomized controlled studies, giving the drug to 50,000 women each year throughout the 1960s
- This turned out to be a medical tragedy – DES has the disastrous side effect of causing cancer in female offspring; DES was banned in 1971

# Summary

- Poor study design can bias results
- Samples are subject to selection bias and nonresponse bias
- Bias arises whenever the sample is not representative of the population
- Another source of bias is the perception of benefit from a treatment (placebo effect)
- Randomization is the only way to guarantee the similarity of the treatment and control groups
- Randomized controlled double-blind experiments reduce bias to a minimum, and that is why they are the most convincing study design
  - Sometimes, controlled randomized experiments are not possible/ethical and observational studies are necessary; we will discuss such studies in our next lecture