Observational studies

Patrick Breheny

September 3

Observational studies

- We have said that randomized controlled experiments are the gold standard for determining cause-and-effect relationships in human health
- However, such experiments are not always possible, ethical, or affordable
- A much simpler, more passive approach is to simply observe people's decisions and the consequences that seem to result from them, then attempt to link the two
- Such studies are called *observational studies*

Smoking

- For example, smoking studies are observational – no one is going to take up smoking for ten years just to please a researcher
- However, the idea of treatment/exposure (smokers) and control (nonsmokers) groups is still used, just as it was in controlled experiments
- The essential difference, however, is that the subject assigns themselves to the exposure/control group – the investigators just watch
- Because of this, confounding is possible: hundreds of studies have shown that smoking is *associated* with various diseases, but none can definitively prove *causation*

Controlling for confounders

- However, just because confounding is possible in such studies does not mean that investigators are powerless to address it
- Instead, well-conducted observational studies make strong efforts to identify confounders and *control for* their effect
- There are many techniques for doing so; the most direct approach is to make comparisons separately for smaller and more homogeneous groups

Controlling for confounders (cont'd)

- For example, studying the association between heart disease and smoking could be misleading, because men are more likely to have heart disease and also more likely to smoke
- A solution is to compare heart disease rates separately: compare male smokers to male nonsmokers, and the same for females
- Age is another common confounding factor that epidemiologists are often concerned with controlling for

The value of observational studies

- Hundreds of very carefully controlled and well-conducted studies of smoking have been conducted in the past several decades
- Most people would agree that these studies make a very strong case that smoking is dangerous, and that alerting the public to this danger has saved thousands of lives
- Observational studies are clearly a very powerful and necessary tool
- Furthermore, observational studies have tremendous value as initial studies to build up support for larger, more resource-intensive controlled experiments
- However, they can be very misleading – identifying confounders is not always easy, and is sometimes more art than science

Racial bias in Florida

- A study of racial bias in the administration of the death penalty was published in the *Florida Law Review*
- The sample consists of 674 defendants convicted of multiple homicides in Florida between 1976 and 1987, classified by the defendant's and the victims' races:

	White defendants		Black defendants	
Victims' race	Total	Death penalty	Total	Death penalty
White	467	53	48	11
Black	16	0	143	4

Evidence for racial bias against whites

- From the table, the overall percentage of white defendants who received the death penalty is

$$\frac{53+0}{467+16} = 11.0\%$$

- And for black defendants,

$$\frac{11+4}{48+143} = 7.9\%$$

- This would seem to be evidence of racial bias against white defendants

Controlling for victim's race

- However, let's control for the potentially confounding effect of victim's race by calculating the percent who received the death penalty separately for white victims and black victims:

	% sentenced to death	
Victims' race	White	Black
White	11.3	22.9
Black	0.0	2.8

- This table indicates racial bias against blacks

What's going on?

- This may seem paradoxical: if blacks are more likely to receive the death penalty for white victims, and also for black victims, how can whites be more likely to receive the death penalty overall?
- The answer is that both races are much more likely to be involved in murders in which the victim is the same race as the defendant (97% of white defendants were on trial for the murder of white victims; 75% of black defendants were on trial for the murder of black victims)
- Furthermore, Florida juries were much more likely to award the death penalty in cases involving white victims (12.5%) than black victims (2.5%)
- Thus, the apparent racial bias against whites could be due to the confounding factor of the victims' race

Weighted averages

- Due to the threat of confounding in observational studies, it is often useful to obtain an overall average that has been adjusted for the confounding factor
- One such method is to calculate a *weighted average*
- In a regular average, every observation gets an equal weight of $1/n$ – an equivalent way of writing the average is

$$\bar{x} = \sum_{i=1}^{n} \frac{1}{n} x_i$$

- In a weighted average, every observation gets its own weight w_i:

$$\bar{x}_w = \sum_{i=1}^{n} w_i x_i$$

where the weights must add up to 1

Death penalty rates as weighted averages

- We can express death penalty rates as weighted averages; this allows us to separate the confounder from the outcome
- I'll use the following notation: For a given defendant race (i.e., white or black):
  - Let w_w denote the proportion on trial for the murder of a white victim
  - Let w_b denote the proportion on trial for the murder of a black victim
  - Let $\bar{x}_w$ denote the percent sentenced to death for the murder of a white victim
  - Let $\bar{x}_b$ denote the percent sentenced to death for the murder of a black victim

Death penalty rates as weighted averages (cont'd)

- White defendants:

$$\begin{aligned}\bar{x} &= w_w \bar{x}_w + w_b \bar{x}_b \\ &= (.967)11.3 + (.033)0 \\ &= 11.0\end{aligned}$$

- Black defendants:

$$\begin{aligned}\bar{x} &= w_w \bar{x}_w + w_b \bar{x}_b \\ &= (.251)22.9 + (.749)2.8 \\ &= 7.9\end{aligned}$$

- This allows us to see directly the effect of confounding: the white-victim death penalty percentage gets 97% of the weight for white defendants, but only 25% of the weight for black defendants

Average controlled for victims' race

- What would happen if these weights were the same (*i.e.* if victims' race was not a confounding factor and both races were equally likely to be on trial for the murder of a white victim)?
- Overall, 76.4% (515/674) of the victims were white and 23.6% were black; using these as weights,

$$\begin{aligned} \text{Whites:} \quad & (.764)11.3 + (.236)0 = 8.6 \\ \text{Blacks:} \quad & (.764)22.9 + (.236)2.8 = 18.2 \end{aligned}$$

- By artificially forcing the distribution of victims' race to be the same for both groups, we obtain an average that is adjusted for the confounding factor of victim's race
- This allows us to isolate the effect of defendant's race upon his/her likelihood of receiving the death penalty, in the absence of the confounding effect of victim's race

Summary

- Randomized controlled trials are not always possible or practical; for these reasons observational studies also play an important role in science
- Observational studies are always limited by confounding, although known confounders can be accounted for, either through design or statistical calculations
- We did a simple example with a weighted average; more sophisticated approaches to adjusting for confounders are discussed in Biostatistical Methods II (BIOS 5720)