Confounding and Effect Modification

John McGready
Johns Hopkins University

Lecture Topics

- Confounding

- Effect modification/statistical interaction

JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

Section A

Confounding: An Introduction

Confounding (Lurking Variable)

- Consider results from the following (fictitious) study:
  - This study was done to investigate the association between smoking and a certain disease in male and female adults
  - 210 smokers and 240 non-smokers were recruited for the study

Results for All Subjects			
	Smokers	**Non-Smokers**	**Totals**
Disease	52	64	116
No Disease	158	176	334
Totals	210	240	450

$$\hat{RR} = \frac{\hat{p}_{smokers}}{\hat{p}_{non-smokers}} = \frac{52/210}{64/240} \approx 0.93$$

$$\hat{OR} = \frac{\hat{p}_{smokers} / (1 - \hat{p}_{smokers})}{p_{non-smokers} / (1 - \hat{p}_{non-smokers})} = \frac{52 \times 176}{158 \times 64} \approx 0.91$$

What's Going On?

- Smoking is protective against disease?

- Most of the smokers are male and non-smokers are female

All Subjects			
	Smokers	**Non-Smokers**	**Totals**
Male	160	40	200
Female	50	200	250
Totals	210	240	450

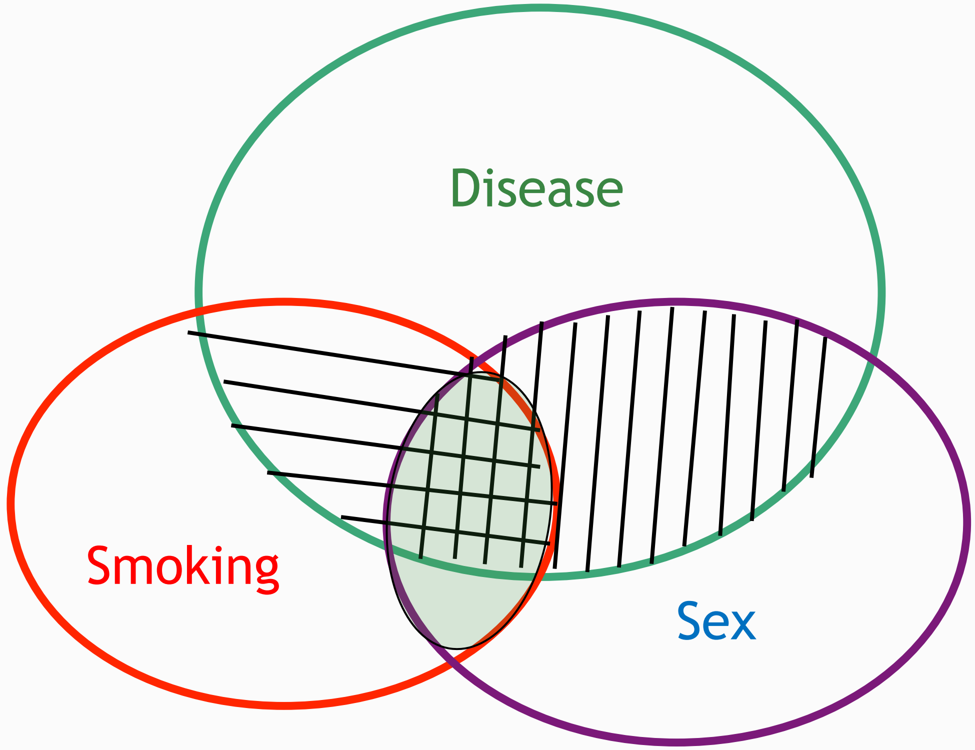
What's Going On?

- Smoking is protective against disease?
- Further, most of the persons with disease are female

All Subjects			
	Disease	**No Disease**	**Totals**
Male	33	167	200
Female	83	167	250
Totals	116	324	450

What’s Going On?

- A picture?



What’s Going On?

- The original outcome of interest is DISEASE

- The original exposure of interest is SMOKING

- In this sample, SEX is related to both the outcome and exposure
  - This relationship is possibly impacting overall relationship between DISEASE and SMOKING

- How can we look at the relationship between DISEASE and SMOKING removing any possible “interference” from SEX?
  - One approach—look at DISEASE and SMOKING relationship separately for males and females

Example

- Is smoking related to disease in males?

Results for Males			
	Smokers	**Non-Smokers**	**Totals**
Disease	29	4	33
No Disease	131	36	167
Totals	160	40	200

$$\hat{R}R_{males} = \frac{\hat{p}_{male\ smokers}}{\hat{p}_{male\ non-smokers}} = \frac{29/160}{4/40} \approx 1.8$$

$$\hat{O}R_{males} = \frac{\hat{p}_{male\ smokers} / (1 - \hat{p}_{male\ smokers})}{p_{male\ non-smokers} / (1 - \hat{p}_{male\ non-smokers})} = \frac{29 \times 36}{4 \times 131} \approx 2$$

Example

- Is smoking related to disease in females?

Results for Females			
	Smokers	**Non-Smokers**	**Totals**
Disease	23	60	83
No Disease	27	140	167
Totals	50	200	250

$$\hat{RR}_{females} = \frac{\hat{p}_{female\,smokers}}{\hat{p}_{female\,non-smokers}} = \frac{23/50}{60/200} \approx 1.5$$

$$\hat{OR}_{females} = \frac{\hat{p}_{female\,smokers} \big/ (1 - \hat{p}_{female\,smokers})}{p_{female\,non-smokers} \big/ (1 - \hat{p}_{female\,non-smokers})} = \frac{23 \times 140}{27 \times 60} \approx 2$$

Smoking, Disease, and Sex

- A recap
  - The overall (sometimes called crude, unadjusted) relationship (RR) between smoking and disease was nearly one (risk difference nearly 0)

$$\hat{R}R = 0.93; \hat{p}_{smokers} - \hat{p}_{non\text{-}smokers} = -0.02$$

  - The sex specific results showed similar positive associations between smoking and disease
  - Males : $\hat{R}R = 1.8; \hat{p}_{male\ smokers} - \hat{p}_{male\ non\text{-}smokers} \approx 0.08$
  - Females: $\hat{R}R = 1.5; \hat{p}_{female\ smokers} - \hat{p}_{female\ non\text{-}smokers} \approx 0.16$
  - (Note, for the moment we are not considering statistical significance, we are just using estimates to illustrate the point)

Simpson's Paradox

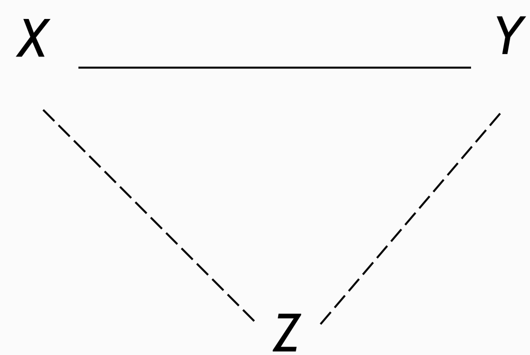
- The nature of an association can change (and even reverse direction) or disappear when data from several groups are combined to form a single group

- An association between an exposure X and a disease Y can be confounded by another lurking (hidden) variable Z

Confounding (Lurking Variable)

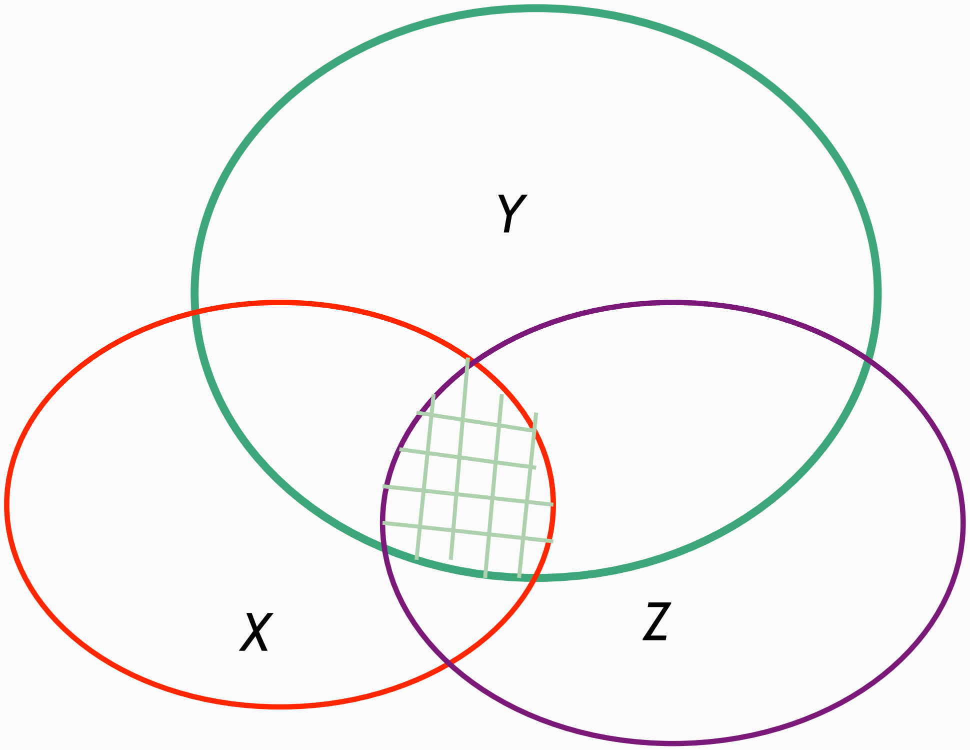
- A confounder Z distorts the true relation between X and Y

- This can happen if Z is related both to X and to Y



What’s Going On?

- A picture



What Is the Solution for Confounding?

- If you **DON'T KNOW** what the potential confounders are, there's not much you can do after the study is over
  - Randomization is the best protection
  - Randomization eliminates the potential links between the exposure of interest and potential confounders Z_1, Z_2, Z_3

- If you can't randomize but **KNOW** what the potential confounders are, or there are statistical methods to help control (adjust for confounders)
  - Potential confounders must be measured as part of study

How to Adjust for Confounding?

- Stratify
  - Look at tables separately
  - For example, male and females, clinic
  - Take weighted average of stratum specific estimates

- For example, in the disease/smoking situation
  - To get a sex adjusted relative risk for the smoking disease relationship we could weight the sex-specific relative risks by numbers of males and females

$$R\hat{R}_{sex\ adjusted} = \frac{n_{males} \times R\hat{R}_{males} + n_{females} \times R\hat{R}_{females}}{n_{males} + n_{females}}$$

$$R\hat{R}_{sex\ adjusted} = \frac{200 \times 1.8 + 250 \times 1.5}{200 + 250} \approx 1.6$$

How to Adjust for Confounding?

- There are better ways than this to take such a weighted average (weighting by standard error, for example), but this just illustrates the concept

- Confidence intervals can be computed for these adjusted measures of association

- One way to assess whether sex is a confounder: compare crude RR to sex adjusted RR, if it's "different" then sex is a confounder

How to Adjust for Confounding?

- Regression methods
  - Just around the corner!
  - More generalizable than weighted average approach, but the idea is similar