Statistically Speaking Lecture Series

Sponsored by the Biostatistics Collaboration Center

The Impact of Other Factors:

Confounding, Mediation, and Effect Modification

Amy Yang, MSc

Senior Statistical Analyst

Department of Preventive Medicine

Northwestern Medicine®
Feinberg School of Medicine

BCC: Biostatistics Collaboration Center

Who We Are

Leah J. Welty, PhD
Assoc. Professor
BCC Director

Lauren Balmert, PhD
Asst. Professor

Jody D. Ciolino, PhD
Asst. Professor

Kwang-Youn A. Kim, PhD
Assoc. Professor

Masha Kocherginsky, PhD
Assoc. Professor

Mary J. Kwasny, ScD
Assoc. Professor

Julia Lee, PhD, MPH
Assoc. Professor

David Aaby, MS
Senior Stat. Analyst

Elizabeth Gray, MS
Stat. Analyst

Kimberly Koloms, MS
Stat. Analyst

Amy Yang, MS
Senior Stat. Analyst

Tameka L. Brannon
Financial | Research
Administrator

Biostatistics Collaboration Center (BCC)

Mission: to support investigators in the conduct of high-quality, innovative health-related research by providing expertise in biostatistics, statistical programming, and data management.

How do we accomplish this?

1. Every investigator is provided a **FREE** initial consultation of 1-2 hours, subsidized by **FSM Office for Research**. Thereafter:
   a) Grants
   b) Subscription
   c) Re-charge (Hourly) Rates

2. Grant writing (e.g. developing analysis plans, power/sample size calculations) is also supported by FSM at **no cost to the investigator**, with the goal of establishing successful collaborations.

BCC: Biostatistics Collaboration Center

What We Do

- Many areas of expertise, including:
  - Bayesian Methods
  - Big Data
  - Bioinformatics
  - Causal Inference
  - Clinical Trials
  - Database Design
  - Genomics
  - Longitudinal Data
  - Missing Data
  - Reproducibility
  - Survival Analysis

Many types of software, including:

BCC: Biostatistics Collaboration Center

Shared Statistical Resources

Northwestern Medicine | NUCATS Clinical and Translational Sciences Institute

Robert H. Lurie Comprehensive Cancer Center of Northwestern University

Stanley Manne Children's Research Institute | Ann & Robert H. Lurie Children's Hospital of Chicago

Biostatistics Collaboration Center (BCC)

- Supports non-cancer research at NU
- Provides investigators an initial 1-2 hour consultation subsidized by the FSM Office of Research
- Grant, Hourly, Subscription

Shirley Ryan AbilityLab

Quantitative Data Sciences Core (QDSC)

- Supports all cancer-related research at NU
- Provides free support to all Cancer Center members subsidized by RHLCCC
- Grant

Biostatistics Research Core (BRC)

- Supports Lurie Children's Hospital affiliates
- Provides investigators statistical support subsidized by the Stanley Manne Research Institute at Lurie Children's.
- Hourly

BCC: Biostatistics Collaboration Center

Shared Resources Contact Info

- Biostatistics Collaboration Center (BCC)
  - Website: http://www.feinberg.northwestern.edu/sites/bcc/index.html
  - Email: bcc@northwestern.edu
  - Phone: 312.503.2288

- Quantitative Data Sciences Core (QDSC)
  - Website: http://cancer.northwestern.edu/research/shared_resources/quantitative_data_sciences/index.cfm
  - Email: qdsc_rhlccc@northwestern.edu
  - Phone: 312.503.2288

- Biostatistics Research Core (BRC)
  - Website: https://www.luriechildrens.org/en-us/research/facilities/Pages/biostatistics.aspx
  - Email: mereed@luriechildrens.org
  - Phone: 773.755.6328

Northwestern Medicine®
Feinberg School of Medicine

The Impact of Other Factors: Confounding, Mediation, and Effect Modification

Amy Yang

Senior Statistical Analyst
Biostatistics Collaboration Center

Oct. 09 2017

Outline

- Confounding
  - Concept and definition
  - Identifying confounding
  - Quantifying confounding
  - Controlling confounding
- Mediation
- Effect Modification
  - Definition and examples
  - Confounding vs Effect Modification

Confounding--Example

- Cohort study -- Smoking and heart disease (HD)

- Suppose that the incidence of HD for smokers is twice that of non-smokers (Risk Ratio=2.0)

Confounding--Example

Smoking doubles your risk of getting heart disease

Before we can make a causal statement…

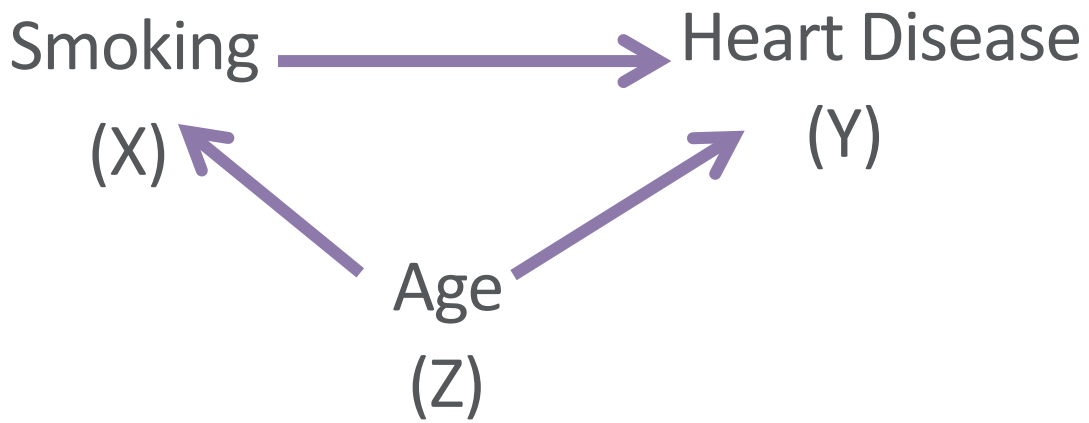
Rule out alternative explanations:

Chance, Bias, **Confounding**

Confounding--Example

- Suppose that the smokers are much older than the non-smokers
  - Twice the incidence of HD among old smokers vs. young non-smokers
- We know that age is a risk factor for heart disease
  - Implies the RR=2 is really reflecting the mixture of two effects (Older age and smoking)
- **Age** is a confounder in the study of association between smoking and HD

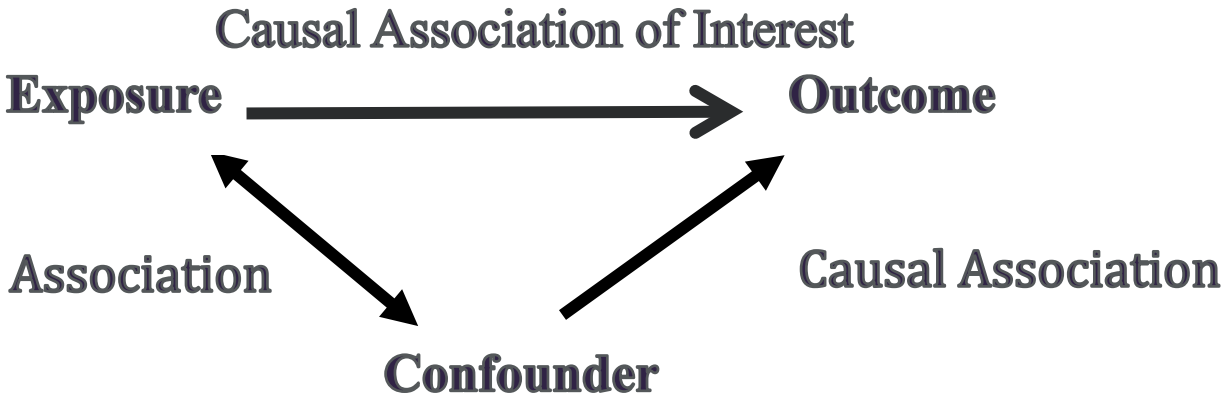
Confounding--Example



- Two pathways
  - Direct effect of smoking
  - Backdoor pathway through age → non-comparability

- **Confounding = Existence of backdoor pathway**

Confounding



Three properties of confounder:

- Should related to the exposure
- Should be an **independent** determinant of the outcome
- Should **not** be part of causal pathway from exposure to outcome
- Often taken as a **definition of a confounder**

Identifying Confounding

- **Not Recommended**
  - Approaches that are based *only* on statistical associations observed in study data

  e.g. Automated procedures (stepwise regression)

- **Recommended**
  - Three properties + knowledge/assumptions about causal relationships among variables
  - Study data are used to quantify confounding

What is not a Confounder--Example

Chemical X → Cognitive disability

- It turns out there are more blondes in the chemical X exposed group

Exposed

Non-Exposed

- **Question**: Is **hair color** a confounder?(Are blondes really…dumber?)
- Hair color is not a confounder, because hair color is not a risk factor for cognitive disability

Quantifying and Controlling Confounding in the Analysis

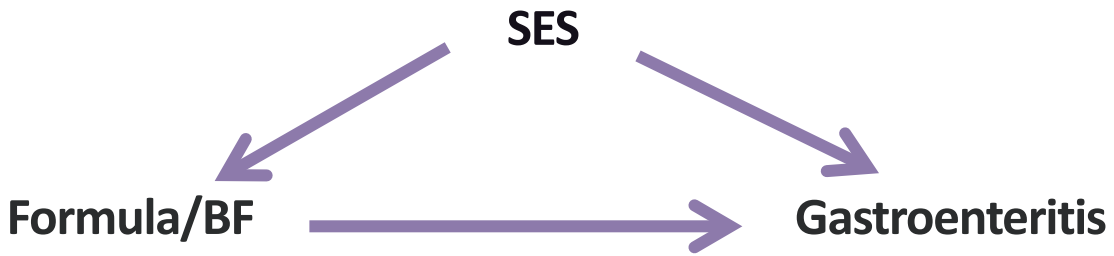
- Comparing the “crude” measure of association with the “adjusted” measures of association
- Stratification
  - Pooling (Weighted Averaging)
- Regression modeling

Example:

- Hypothetical case-control study examining the association between **formula vs. breastfeeding** and **gastroenteritis** among infants

Example:

- Concern about socioeconomic status (SES) as a confounder



- **Check the three properties:**
  1. SES affects whether people formula or breastfeed
  2. SES affects the outcome through the degree of crowding and hygiene issues
  3. SES is not in the pathway between feeding methods and Gastroenteritis

Quantifying and Controlling Confounding in the Analysis

- 1. Crude association -- OR=(261*296)/(645*54)=2.22

Gastroenteritis

	Yes	No
Formula	261	645
Breastfeeding	54	296

- 2. Stratify by confounder –SES

Low SES	Yes	No
Formula	219	447
Breastfeeding	33	118

$OR_{LOW} = 1.75$

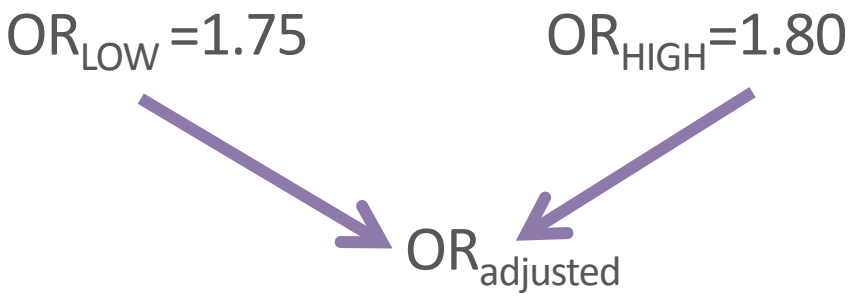
High SES	Yes	No
Formula	42	198
Breastfeeding	21	178

$OR_{HIGH} = 1.80$

- Positive confounder because crude OR 2.2 was larger than the stratified ORs 1.75 and 1.80

Quantifying and Controlling Confounding in the Analysis

- 3. Pooling (weighted averaging) –adjusted association
  - **If appropriate**, pool information over all strata by calculating (weighted) average of stratum specific measures
  - Assumption: constant effect across strata



- Mantel-Haenszel weights
  - Reflect amount of "information" within each stratum
  - Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease JNCI 22: 719-748, 1959

Mantel-Haenszel Estimation

- Case control data:

Low SES	Yes	No
Formula	219	447
Breastfeeding	33	118

High SES	Yes	No
Formula	42	198
Breastfeeding	21	178

$$OR_{MH} = \frac{\sum w_i OR_i}{\sum w_i} = \frac{\sum_{i=1}^{k} (ad/Total)_i}{\sum_{i=1}^{k} (bc/Total)_i} = \frac{\frac{291*118}{817} + \frac{42*178}{439}}{\frac{447*33}{817} + \frac{198*21}{439}} = 1.77$$

$OR_{LOW} = 1.75$ $OR_{HIGH} = 1.8$

$OR_{adjusted} = 1.77$

Modeling

- Stratification and MH estimation are equivalent to…
  - Calculating an unadjusted measure of association from a model

    Gastroenteritis ~ **b1***Formula/BF

  - Examining the measure of association after including the confounder in the model

    Gastroenteritis ~ **b1'***Formula/BF + b2*SES

Preventing Confounding in Study Design

- Confounding is a **bias**

- We want to **prevent** in the conduct of the study and **remove** once we determine that it is present

- Study design strategies:
  - Randomization
  - Matching
  - Restriction

Preventing Confounding in Study Design

Randomization

- Subjects are allocated to exposure groups by a random method

- Gives subject **equal chance** of being in any exposure group

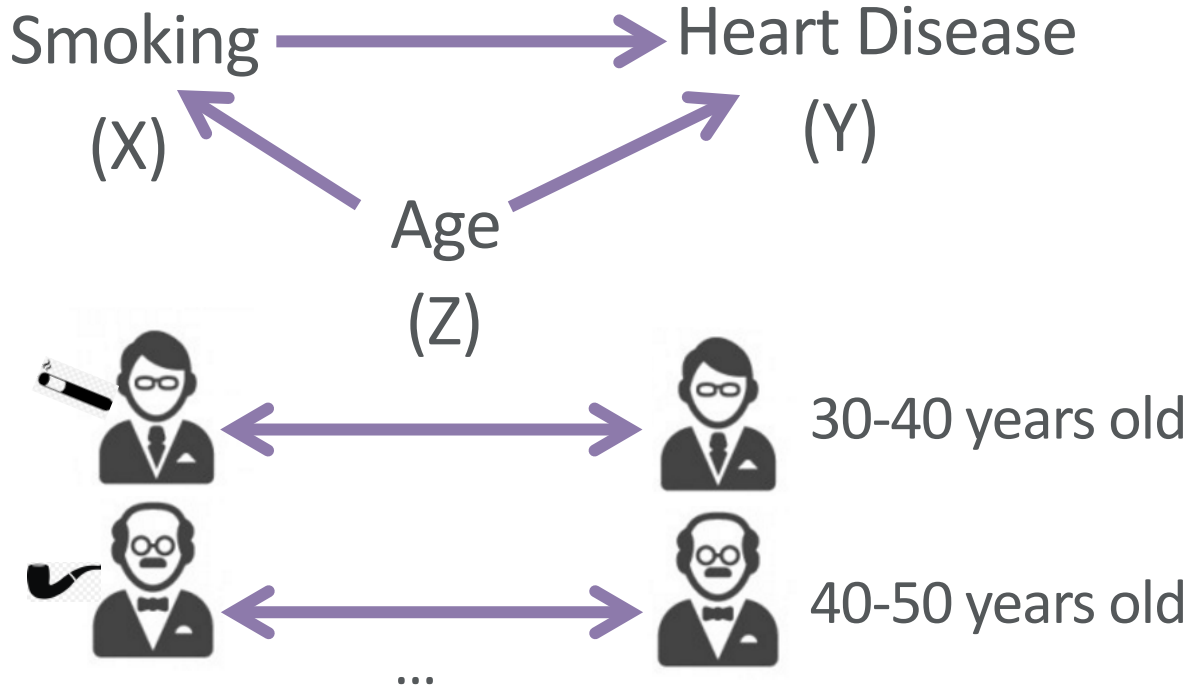
- Exposure groups will have similar distribution of
  - Age, gender, behavior …

- This includes both measured and unmeasured confounders
- Depending on the trial, confounders may still need to be considered in analysis (especially when n is small)

Preventing Confounding in Study Design

Matching

- On important potential confounders

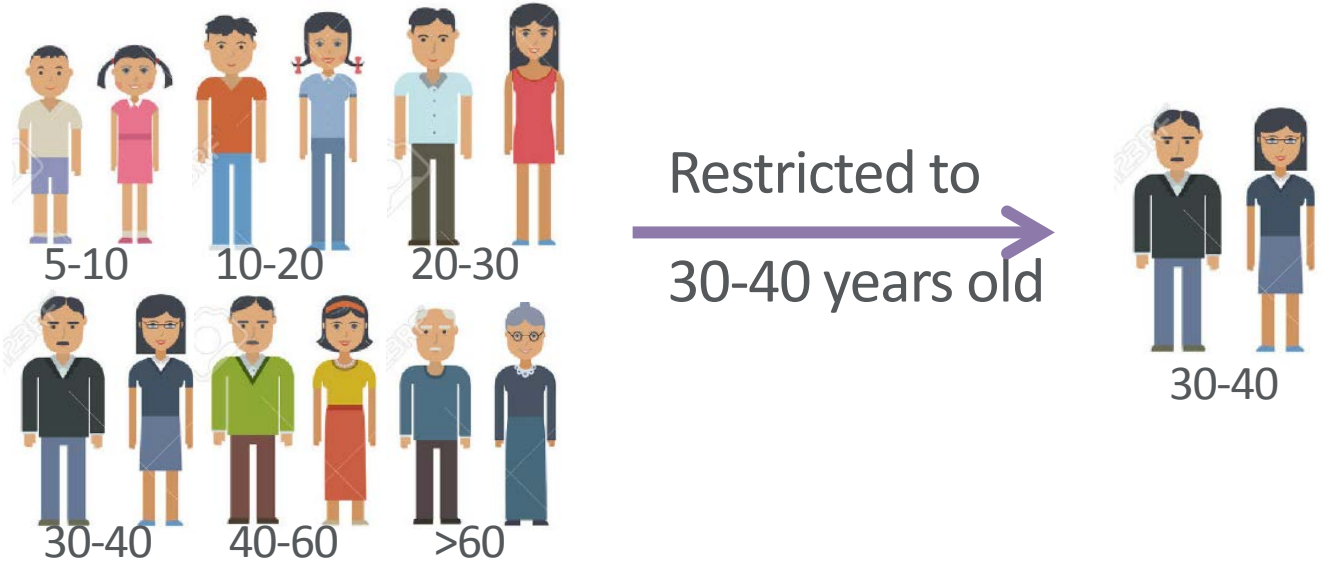


- Smoking and Non-Smoking groups are similar with respect to Age
- Analyses must account for matching

Preventing Confounding in Study Design

Restriction

- Restrict admission into the study to subjects who have the same level of the confounding factor
- E.g., Confounding by **Age** could be minimized by enroll subjects that are in the same age range



- **Be careful! Restriction limits generalizability**

Summary -- Confounding

- Three properties

- Control for confounding in the analysis
  - Stratification
  - MH estimation
  - Modeling

- Design strategies to prevent confounding
  - Randomization
  - Matching
  - Restriction

Mediation

- Confounder should *not* be in the pathway between the exposure and outcome

- If the other variable is in the pathway between the two, it is called a mediator

$$X \rightarrow Z \rightarrow Y$$

Mediation

Poverty

Burger
$0.99

Salad
$4.99

Limited access to healthy food

Diabetes

Mediation

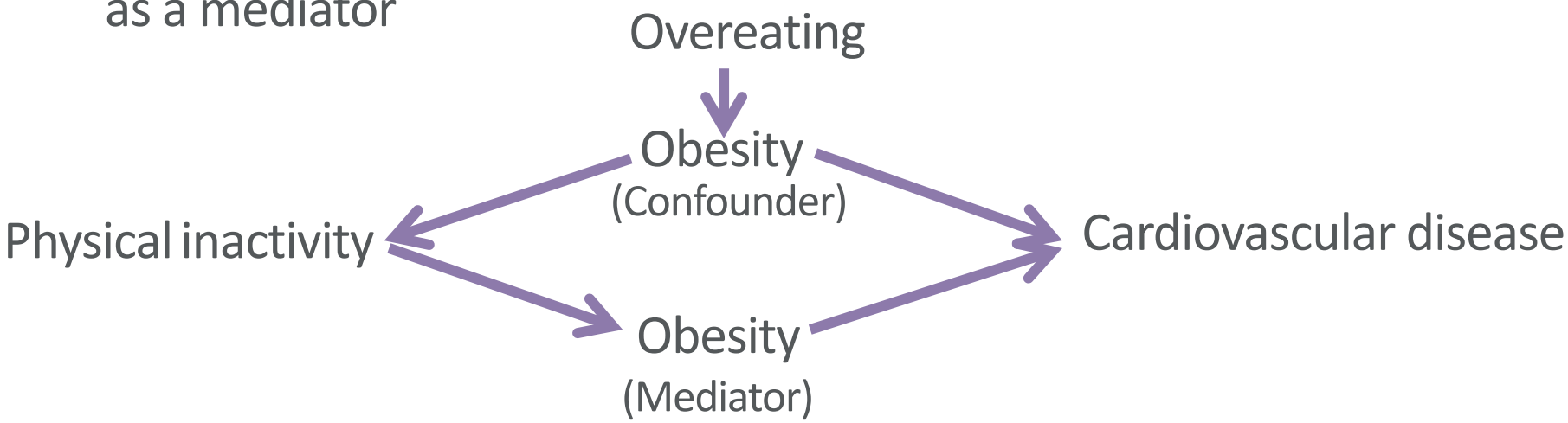
Multiple sexual partners

Increased risk of HPV infection

Cervical cancer

Mediation

- It is difficult to distinguish confounder and mediator **statistically**
- They should be separated from each other based on an **understanding of disease process**
- A variable can act partially as a confounder and partially as a mediator



Mediation

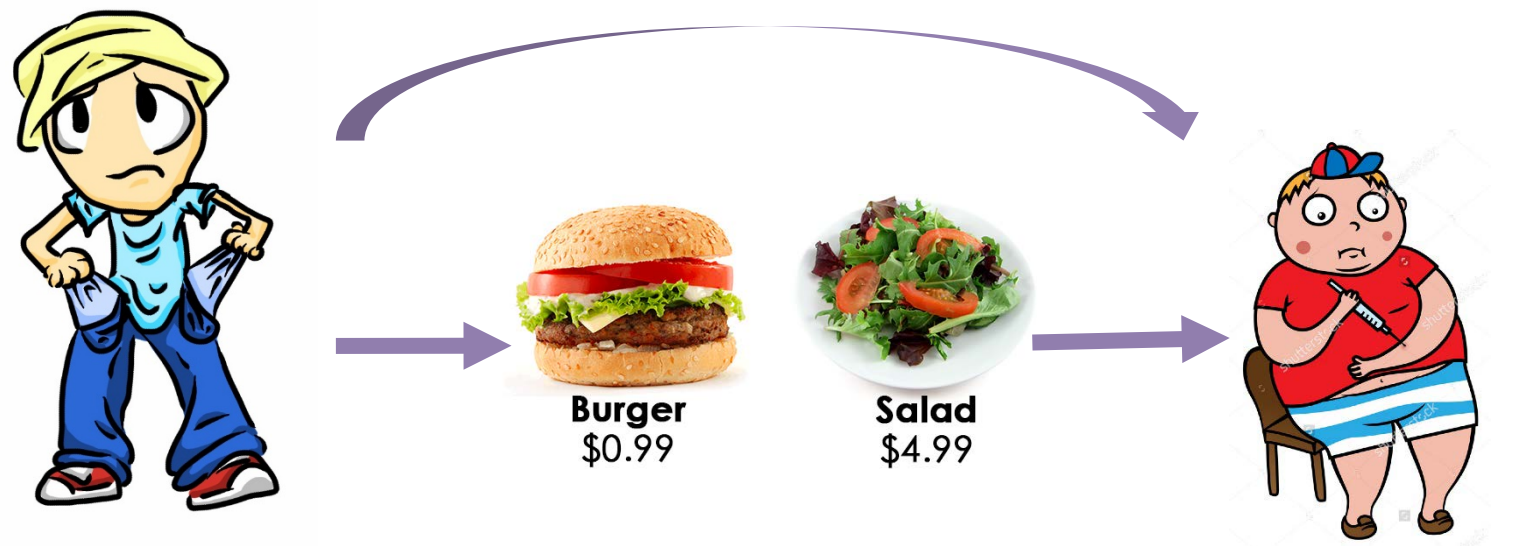
- **Question** : Should we adjust for mediators, as we do for confounders?
- We can, but the meaning of this adjustment is different

- Before adjustment, we have the **total effect** of the potential risk factor on the outcome

- After adjustment, we have the **remaining effect** of the risk factor after the partial effect of that mediator is considered

- **Remaining effect will be smaller than total effect**

Mediation



- If we do not adjust for the mediator
  - Crude OR = 2.4; **Total effect** of poverty on diabetes
- If we adjust for eating unhealthy food
  - OR_{adjust}=1.6; **Remaining** effect of poverty on diabetes

Effect Modification (Interaction)

- Effect modification is present when the measure of association between X and Y **varies** across a third variable (Z)

- **Gender** modifies the effect of **marital status** on health outcomes

J Epidemiol Community Health 2000;**54**:254-261 doi:10.1136/jech.54.4.254

Research report

Marital status and suicide in the National Longitudinal Mortality Study

Abstract

OBJECTIVES The purpose of the study was to examine the effect of marital status on the risk of suicide, using a large nationally representative sample. A related objective was to investigate the association between marital status and suicide by sex.

RESULTS For the entire sample, higher risks of suicide were found in divorced than in married persons. Divorced and separated persons were over twice as likely to commit suicide as married persons (RR=2.08, 95% confidence intervals (95% CI) 1.58, 2.72). Being single or widowed had no significant effect on suicide risk. When data were stratified by sex, it was observed that the risk of suicide among divorced men was over twice that of married men (RR=2.38, CI 1.77, 3.20). Among women, however, there were no statistically significant differentials in the risk of suicide by marital status categories.

Effect Modification

- Conceptualization of effect modification
  - Approach one

  The "effect" of variable X on Y is not the same across levels of variable Z

  - Approach two

  The "effect" of variables X and Z on Y combined is larger or smaller than you would expect given the "effect" of each on Y individually

$$Y=X+Z+X*Z$$

- Mathematically these two approaches are the same

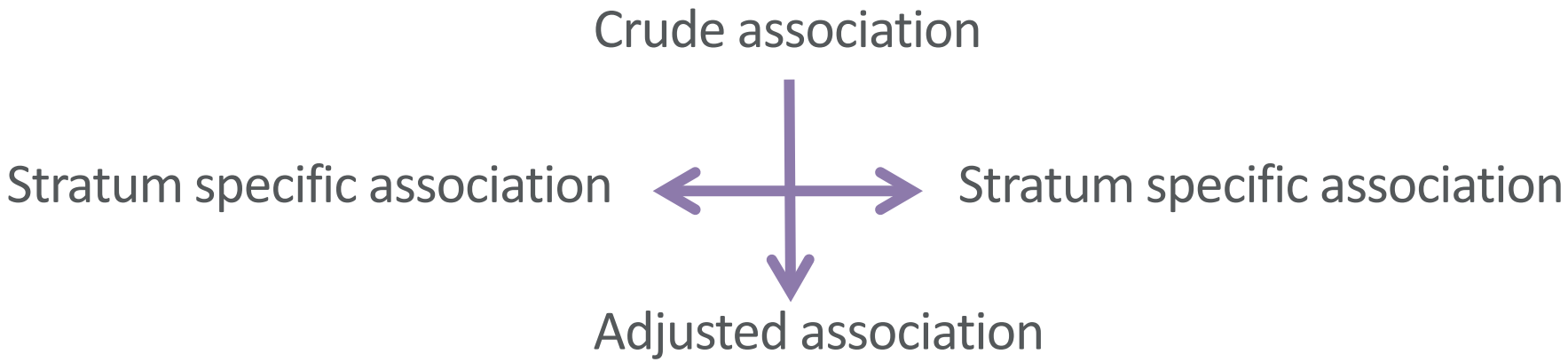
Confounding vs Effect Modification

- People confuse them because the same technique – stratification is used

- Stratification is a step in the process of **adjusting for confounding**
  - Bias we want to remove

- Stratification is a step in the process of **describing effect modification**
  - We want to describe effect modification

Confounding vs Effect Modification

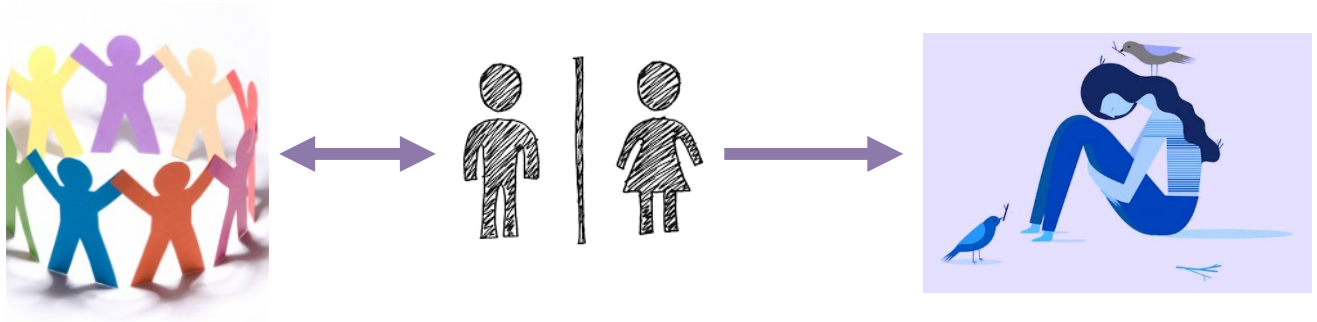


- Confounding
  - Association is **similar** in different strata of Z
  - Compare the adjusted association with the crude association

- Effect modification
  - Association is **different** in different strata of Z
  - Compare associations across strata

Confounding vs Effect Modification

- A factor could be confounder and/ or modifier
- Example: Gender effect on the study of relation between social support and depression



- Confounder -- more women in the high social support group and women tend to have higher depression level

- Effect modifier -- social support is more effect in lowering depression in women than in men

Road Map

- 1. Calculate the crude measure of association
- 2. Stratify the data by the potential confounder/ effect modifier
- 3. Calculate the stratified measure of association
- 4. Compare 3 using the **Test for Homogeneity** (**Breslow-Day Test**)
- 5. Are the associations homogeneous?

Yes
(i.e. did not reject H0)

- 6. Calculate the adjusted measure of association – **Mantel-Haenszel estimation**
- 7. Compare 6 and 1 to describe direction and magnitude of the confounding

No
(i.e. rejected H0)

6. Present measures of association stratified by effect modifier

Road Map Step 1

- 1. Calculate the crude measure of association between the exposure and outcome (e.g. RR, OR)

Incident depression

	Yes	No	Total
Low social support	191	7909	8100
High social support	50	7550	7600
Total	241	15459	15700

Risk ratio = (191/8100)/(50/7600)=3.6

Road Map Step 2 & 3

- 2. Stratify the data by the potential confounder/ effect modifier

Incident depression

Men	Yes	No	Total
Low social support	26	2574	2600
High social support	18	3582	3600
Total	44	6156	6200

Incident depression

Women	Yes	No	Total
Low social support	165	5335	5500
High social support	32	3968	4000
Total	197	9303	9500

- 3. Calculate the stratified measure of association

RR_{Men}= (26/2600)/(18/3600)=2 RR_{Women}= (165/5500)/(32/4000)=3.75

Road Map Step 4

- 4. Compare the RRs using the **Test for Homogeneity** (**Breslow-Day Test**)
  - Equivalent to test statistics for interaction term in regression model
  - Null hypothesis: the measure of association is homogeneous across strata

- If the test of homogeneity is "significant"
  - Reject homogeneity
  - Evidence for heterogeneity (i.e. effect modification)

- The choice of significant level (e.g. $p<0.05$) is open to interpretation
  - One "conservative" approach is using significant level of larger than 0.05 (maybe 0.10 or 0.20)

Road Map Step 5 & 6

- In our example χ^2=3.08, DF=1, P=0.08
- 5. **Question**: Does it appear we have homogeneous association (H0: Association the same across strata)?
- Assume we used conservative 10% level of significance…
- No (p=0.08<0.10)
- Reject H0; we have evidence of effect modification

- 6. Present measures of association stratified by gender

RR_{MEN}= 2 RR_{WOMEN}=3.75

Exercise

- X-Y association stratified by potential confounder/EM Z

Z=0	Z=1	Crude	Adjusted	Confounding?	EM?
4	0.25	1	1		
1	1	8.4	1		
4	0.25	1	2		

Adjusted estimate not relevant – present stratified associations when there is effect modification

Properties of Stratification

- Pro:
  - Simple and intuitive

- Con:
  - Not practical when there are multiple factors
  - With continuous variables (e.g. age) have to create categories
  - In these situations, regression models have many strengths

Summary

- Other variables in a study can be
  - Confounders
    - Bias
    - Prevent in study design
    - Adjust for in analysis
  - Effect modifiers
    - Personalized medicine; effects in a subgroup
    - Stratify and report
  - Mediators
    - X→Z→Y

Statistically Speaking

What's next

- **Monday, October 16**
  Using REDCap for Data Capture in Clinical Studies: Database Management on a Budget
  Jody D. Ciolino, PhD, Assistant Professor of Preventive Medicine

- **Monday, October 30**
  Using R for Statistical Graphics: The Do's and Don'ts of Data Visualization
  David Aaby, MS, Senior Statistical Analyst of Biostatistics Collaboration Center

- **Wednesday, November 1**
  Time-to-Event Analysis: A 'Survival' Guide
  Lauren C. Balmert, PhD, Assistant Professor of Preventive Medicine

BCC: Biostatistics Collaboration Center

Contact Us

- Request an Appointment
  - http://www.feinberg.northwestern.edu/sites/bcc/contact-us/request-form.html
- General Inquiries
  - bcc@northwestern.edu
  - 312.503.2288
- Visit Our Website
  - http://www.feinberg.northwestern.edu/sites/bcc/index.html

Biostatistics Collaboration Center |680 N. Lake Shore Drive, Suite 1400 |Chicago, IL 60611

Northwestern Medicine®
Feinberg School of Medicine

Your feedback is important to us! (And helps us plan future lectures).

Complete the evaluation survey to be entered in to a drawing to win 2 free hours of biostatistics consultation.