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

Amy Yang

Senior Statistical Analyst
Biostatistics Collaboration Center

Oct. 14 2016
BCC: Biostatistics Collaboration Center

Who We Are

Leah J. Welty, PhD
Assoc. Professor
BCC Director

Joan S. Chmiel, PhD
Professor

Jody D. Ciolino, PhD
Asst. Professor

Kwang-Youn A. Kim, PhD
Asst. Professor

Mary J. Kwasny, ScD
Assoc. Professor

Julia Lee, PhD, MPH
Assoc. Professor

Alfred W. Rademaker, PhD
Professor

Masha Kocherginsky, PhD
Assoc. Professor

Mary J. Kwasny, ScD
Assoc. Professor

Julia Lee, PhD, MPH
Assoc. Professor

Not Pictured:
1. David A. Aaby, MS
Senior Stat. Analyst

2. Tameka L. Brannon
Financial | Research Administrator

Hannah L. Palac, MS
Senior Stat. Analyst

Gerald W. Rouleau, MS
Stat. Analyst

Amy Yang, MS
Senior Stat. Analyst

Biostatistics Collaboration Center | 680 N. Lake Shore Drive, Suite 1400 | Chicago, IL 60611
Our mission is to support FSM investigators in the conduct of high-quality, innovative health-related research by providing expertise in biostatistics, statistical programming, and data management.
BCC: Biostatistics Collaboration Center

**How We Do It**

- **Are you writing a grant?**
  - **YES**
    - We provide:
      - Study Design
      - Analysis Plan
      - Power Sample Size
    - BCC faculty serve as Co-Investigators; analysts serve as Biostatisticians.
  - **NO**
    - **Short or long term collaboration?**
      - **Short**
        - Recharge Model (hourly rate)
      - **Long**
        - Subscription Model (salary support)

- The BCC recommends requesting grant support at least **6 - 8 weeks** before submission deadline.

- Statistical support for **Cancer-related projects** or **Lurie Children's** should be triaged through their available resources.

- Every investigator is provided a **FREE** initial consultation of up to 2 hours with BCC faculty of staff.

---

Northwestern Medicine
Feinberg School of Medicine
BCC: Biostatistics Collaboration Center

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

Amy Yang

Senior Statistical Analyst
Biostatistics Collaboration Center

Oct. 14 2016
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)
Before we can make a causal statement...
Rule out alternative explanations:
  Chance, Bias, **Confounding**

Smoking doubles your risk of getting heart disease
Confounding--Example

• Suppose that the smokers are much older than the 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
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
It turns out there are more blondes in the chemical X exposed group. 

- **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)

• 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

  ![](image)

• **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 -- \( OR = \frac{(261 \times 296)}{(645 \times 54)} = 2.22 \)

2. Stratify by confounder – SES

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
    
    \[ \text{OR}_{\text{LOW}} = 1.75 \quad \text{OR}_{\text{HIGH}} = 1.80 \]

  - 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:

<table>
<thead>
<tr>
<th>Low SES</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formula</td>
<td>219</td>
<td>447</td>
</tr>
<tr>
<td>Breastfeeding</td>
<td>33</td>
<td>118</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>High SES</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formula</td>
<td>42</td>
<td>198</td>
</tr>
<tr>
<td>Breastfeeding</td>
<td>21</td>
<td>178</td>
</tr>
</tbody>
</table>

\[
\text{OR}_{\text{MH}} = \frac{\sum w_i OR_i}{\sum w_i} = \frac{\sum (ad/Total)_i}{\sum (bc/Total)_i} = \frac{291*118 + 42*178}{817} + \frac{447*33 + 198*21}{439} = 1.77
\]

\[
\text{OR}_{\text{LOW}} = 1.75
\]

\[
\text{OR}_{\text{adjusted}} = 1.77
\]

\[
\text{OR}_{\text{HIGH}} = 1.8
\]
Modeling

- Stratification and MH estimation are equivalent to...

  - Calculating an unadjusted measure of association from a model
    \[ \text{Gastroenteritis} \sim b_1 \times \text{Formula/BF} \]

  - Examining the measure of association after including the confounder in the model
    \[ \text{Gastroenteritis} \sim b_1' \times \text{Formula/BF} + b_2 \times \text{SES} \]
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 (X) \rightarrow \text{Heart Disease (Y)}

Age (Z)

- 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 → 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

Physical inactivity \[\leftrightarrow\] Obesity (Confounder) \[\leftrightarrow\] Obesity (Mediator) \[\leftrightarrow\] Cardiovascular disease
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

Poverty → Limited access to healthy food → Diabetes

• 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\textsubscript{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 \textit{varies} across a third variable (Z)

- Gender modifies the effect of marital status on health outcomes
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

  Divorced $\rightarrow$ Suicide
  - Men RR=2.38
  - Women RR=1 no association

- 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

  \[ Y = X + Z + X \times Z \]

- Mathematically these two approaches are the same
Confounding vs Effect Modification

• 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: Study of relation between social support and depression
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
1. Calculate the crude measure of association between the exposure and outcome (e.g. RR, OR)

### Incident depression

<table>
<thead>
<tr>
<th>Low social support</th>
<th>Yes</th>
<th>No</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>191</td>
<td>7909</td>
<td>8100</td>
<td></td>
</tr>
<tr>
<td>High social support</td>
<td>50</td>
<td>7550</td>
<td>7600</td>
</tr>
</tbody>
</table>

| 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

<table>
<thead>
<tr>
<th></th>
<th>Men</th>
<th></th>
<th></th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low social support</td>
<td>26</td>
<td>257</td>
<td>4</td>
<td>2600</td>
</tr>
<tr>
<td>High social support</td>
<td>18</td>
<td>358</td>
<td>2</td>
<td>3600</td>
</tr>
<tr>
<td>Total</td>
<td>44</td>
<td>615</td>
<td>6</td>
<td>6200</td>
</tr>
</tbody>
</table>

Incident depression

<table>
<thead>
<tr>
<th></th>
<th>Women</th>
<th></th>
<th></th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low social support</td>
<td>16</td>
<td>533</td>
<td>5</td>
<td>5500</td>
</tr>
<tr>
<td>High social support</td>
<td>32</td>
<td>396</td>
<td>8</td>
<td>4000</td>
</tr>
<tr>
<td>Total</td>
<td>19</td>
<td>930</td>
<td>3</td>
<td>9500</td>
</tr>
</tbody>
</table>

$RR_{Men} = \frac{26/2600}{18/3600} = 2$

$RR_{Women} = \frac{165/5500}{32/4000} = 3.75$
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)
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**

\[
\text{RR}_{\text{MEN}} = 2 \quad \text{RR}_{\text{WOMEN}} = 3.75
\]
Exercise

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

<table>
<thead>
<tr>
<th>Z=0</th>
<th>Z=1</th>
<th>Crud.</th>
<th>Adjusted</th>
<th>Confounding?</th>
<th>EM?</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.25</td>
<td>1</td>
<td>1</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>8.4</td>
<td>1</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.25</td>
<td>1</td>
<td>2</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

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
    • $X \rightarrow Z \rightarrow Y$
<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
<th>Speaker</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Associate Professor, Division of Biostatistics, Department of Preventive Medicine</td>
<td></td>
</tr>
<tr>
<td>October 21</td>
<td>Clinical Trials: Highlights from Design to Conduct</td>
<td>Masha Kocherginsky, PhD</td>
<td>Clinical Trials: Highlights from Design to Conduct</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Associate Professor, Division of Biostatistics, Department of Preventive Medicine</td>
<td></td>
</tr>
<tr>
<td>October 25</td>
<td>Finding Signals in Big Data</td>
<td>Kwang-Youn A. Kim, PhD</td>
<td>Finding Signals in Big Data</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Assistant Professor, Division of Biostatistics, Department of Preventive Medicine</td>
<td></td>
</tr>
<tr>
<td>October 28</td>
<td>Enhancing Rigor and Transparency in Research: Adopting Tools that Support Reproducible Research</td>
<td>Leah J. Welty, PhD</td>
<td>Enhancing Rigor and Transparency in Research: Adopting Tools that Support Reproducible Research</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BCC Director, Associate Professor, Division of Biostatistics, Department of Preventive Medicine</td>
<td></td>
</tr>
</tbody>
</table>

All lectures will be held from noon to 1 pm in Hughes Auditorium, Robert H. Lurie Medical Research Center, 303 E. Superior St.