Confounding and Stratification

Epidemiological and Statistical Methods Day 1
Mutsa Bwakura Dangarembizi
2019
Objectives

• Understand the concept of confounding
• Explain the properties of a confounder
• Describe ways to control and assess for confounding
• Describe residual confounding and Interaction
What is confounding?
What is confounding?

• Distortion (inaccuracy) in the estimated measure of association between an exposure and an outcome
  – Occurs when the exposure of interest is mixed up with some other factor that is associated with the outcome
Consequences of confounding

• Can lead to overestimation or under-estimation of the true association between exposure and outcome
• Can change the direction of the observed effect
Example 1.

Does smoking influence the risk of coronary heart disease in men aged 18 to 64yr?
Effect of alcohol intake mixes up the estimate of the association between smoking and CHD and alcohol intake is also correlated with smoking.
Criteria for Confounding

Exposure → Confounding → Outcome

Confounding
Criteria for confounding

- Must be associated with the exposure of interest in the source population
- Must be a risk factor for the outcome of interest in those who are not exposed
- Must not be on the causal pathway between the exposure and outcome of interest
For example 1

1. Smokers generally have a higher alcohol intake
2. Alcohol intake is a risk factor for CHD in non-smokers
3. Alcohol intake is not on the causal pathway between smoking and CHD
An example of being in the causal pathway

Alcohol Consumption → ↑ HDL level → ↓ Heart disease
Controlling for confounding
Study design stage

• Randomization
  – Ensures that all known and unknown confounding factors are evenly distributed in the study groups
  – Need a large sample size
Restriction

• Limit participation to individuals who are similar in relation to the confounder
  – eg if smoking limit to non-smokers
  – Convenient and cheap
  – But
  – Reduces eligible participants
  – Limits generalisability
Matching

• Individuals are selected in such a way that the potential confounders are distributed in similarly among the study groups
  – Eg case control study of exercise and risk of MI
    • If age, sex and smoking are potential confounders, match the cases and controls
  – Difficult and expensive
  – Useful in case control studies
Controlling for confounding during analysis

• You must have measured the confounder

• Stratification
  – Stratify the data according to the potential confounder
  – Then measure the association between exposure and outcome in the different strata
  – If they are different from the “crude” measure of effect but similar to each other, this is evidence of confounding
Example: Case control study

- Coffee drinking and risk of pancreatic cancer
  - Cigarette smoking potential confounder

<table>
<thead>
<tr>
<th></th>
<th>Coffee drinker</th>
<th>Non coffee drinker</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ca pancreas</td>
<td>450</td>
<td>300</td>
</tr>
<tr>
<td>No Ca Pancreas</td>
<td>200</td>
<td>250</td>
</tr>
</tbody>
</table>

Estimated OR = 450/300 ÷ 200/250 = 1.9
Data analysis according to smoking habit

<table>
<thead>
<tr>
<th></th>
<th>Non smokers</th>
<th>Non smokers</th>
<th>Smokers</th>
<th>Smokers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coffee</td>
<td>Coffee</td>
<td>No coffee</td>
<td>coffee</td>
<td>No coffee</td>
</tr>
<tr>
<td>Ca pancreas</td>
<td>50</td>
<td>100</td>
<td>400</td>
<td>200</td>
</tr>
<tr>
<td>No Ca pancreas</td>
<td>100</td>
<td>200</td>
<td>100</td>
<td>50</td>
</tr>
</tbody>
</table>

Estimated OR = $\frac{50}{100} \div \frac{100}{200} = 1.0$

$\frac{400}{200} \div \frac{100}{50} = 1.0$

Smoking confounded the association between coffee consumption and Ca pancreas.

No effect of coffee on Ca pancreas for smokers or non smokers.
Example: cohort study

Is there an association between exposure to chemical X and the risk of developing lung cancer?

<table>
<thead>
<tr>
<th>Lung cancer</th>
<th>Exposure to chemical X</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>480</td>
<td>360</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>83,520</td>
<td>95,640</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>84,000</td>
<td>96,000</td>
</tr>
</tbody>
</table>

Occupational cohort study
Risk in the exposed = 480/84,000
Risk in unexposed = 360/96,000
Crude risk ratio = (480/84,000)/(360/96,000)
= 1.52 (95% CI 1.33-1.75)
p < 0.001

What is the interpretation?
Stratified analysis

Risk ratio among smokers = 
\[
\frac{80}{4,000} / \frac{160}{16,000} \quad = 2.00 \quad (95\% \text{ CI } 1.53-2.61)
\]

Risk ratio among non-smokers = 
\[
\frac{400}{80,000} / \frac{200}{80,000} \quad = 2.00 \quad (95\% \text{ CI } 1.69-2.37)
\]
Hypothetical examples of confounding

<table>
<thead>
<tr>
<th></th>
<th>Crude effect</th>
<th>Effect in stratum 1</th>
<th>Effect in stratum 1</th>
<th>Adjusted effect</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>No confounding</td>
</tr>
<tr>
<td>2</td>
<td>4.0</td>
<td>3.0</td>
<td>3.0</td>
<td>3.0</td>
<td>confounding</td>
</tr>
<tr>
<td>3</td>
<td>1.53</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>confounding</td>
</tr>
</tbody>
</table>
Summary of stratification analysis

• Confounding is present when there is a difference in the crude and adjusted estimates
  – Evidence of confounding if >10% difference
• A weighted average is used when strata have varying number of individuals
  – Mantel-Haensel method
• There is no statistical test for confounding
Controlling for confounding during analysis

• Standardization
  – Techniques used to remove the effects of differences in age or other confounding variables when comparing 2 or more populations

• Multivariate analysis
  – Allows adjustments for many confounders simultaneously
  – Involves construction of a mathematical model to describe the association between exposure and disease as well as confounders
Residual confounding

- Results from the incomplete adjustment for a confounder
  - Eg adjusting for social status, using broad categories (age)
  - Usually difficult to measure
Interaction (Effect Modification)

• Applied to situations in which the magnitude of the effect of an exposure of interest differs depending on the level of a third variable.
• A variable that differentially modifies the observed effect of a risk factor on a disease status
• Is associated with the outcome but not the exposure
• It must be detected and described
Example of interaction

• Reye’s syndrome is a rare but severe condition characterized by brain damage and liver dysfunction after a viral illness
  – Most common in 4-14 yr age-group treated with aspirin
  – Can also occur in adults but is rare

• Effect of aspirin treatment for a viral illness is clearly modified by age
Conclusion

- Confounding is concerned with “alternative” explanations for the effect seen between an exposure and the outcome.
- A confounder is associated with the exposure and a risk factor for the outcome but is not in the causal pathway.
- To prevent confounding, choose an appropriate study design and control for it using statistical methods.
- When confounding is present, the effect of the exposure of interest is the same in all strata of the confounding factor, but is different from the “crude” effect.
- Effect modification must be detected and described.
Sources

• Epidemiology in Medicine Hennekens and Buring
• ICEMS - LSHTM 2018
Thank you