Estimating Associations: Role of Adjustment Variables
Statistics 211 - Statistical Methods II

Presented January 30, 2018
Types of Adjustment Variables

Effect modifiers (interaction terms, moderators)

Suppose that we are interested in modeling the association between an outcome variable $Y$ and a predictor $X$.

I tend to classify adjustment covariates into five broad categories (this terminology is not universal).

Effect modifiers (interaction variables, moderators)

An effect modifier ($W$) is a covariate for which the association between the predictor of interest ($X$) and the outcome of interest ($Y$) differs with each level of $W$. 


Types of Adjustment Variables

Example: Effect modification

Example: The association between gender and the risk of chd differs by systolic blood pressure

<table>
<thead>
<tr>
<th>sbpgrp</th>
<th>Odds Ratio</th>
<th>chi2(1)</th>
<th>P&gt;chi2</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.394493</td>
<td>86.23</td>
<td>0.0000</td>
<td>0.32186 0.48351</td>
</tr>
<tr>
<td>2</td>
<td>0.429583</td>
<td>56.59</td>
<td>0.0000</td>
<td>0.34243 0.53892</td>
</tr>
<tr>
<td>3</td>
<td>0.597384</td>
<td>9.91</td>
<td>0.0016</td>
<td>0.43193 0.82621</td>
</tr>
<tr>
<td>4</td>
<td>0.741269</td>
<td>1.75</td>
<td>0.1858</td>
<td>0.47495 1.15693</td>
</tr>
</tbody>
</table>

How do we deal with effect modifiers?

Present stratified point estimates
Types of Adjustment Variables

Confounders

- **One definition**: A **confounder** is a variable that is causally related with the predictor of interest \((X)\) and the outcome of interest \((Y)\).
Types of Adjustment Variables

Example: Confounding

- Example: Weight may be a confounder in the relationship between diabetes and blood pressure:
  - Obesity is a leading cause of (type II) diabetes
  - Increased weight is associated with higher blood pressure

How do we deal with confounding?

- Adjust for the confounder
Types of Adjustment Variables

Mediator

- A mediator (or intervening, or intermittent variable) is a variable that lies in the causal pathway between the predictor of interest ($X$) and the outcome of interest ($Y$).

\[ \text{Predictor (}X\text{)} \quad \rightarrow \quad \text{Mediator (}W\text{)} \quad \rightarrow \quad \text{Outcome (}Y\text{)} \]
## Types of Adjustment Variables

### Example: Mediator

- **Example:** Diet can be viewed as a mediator in the relationship between socioeconomic status (SES) and obesity:
  - Lower SES generally leads to worse diet (cheaper fast foods)
  - Worse diet leads to obesity

- Adjustment for a mediator will attenuate or possibly remove the effect of the predictor of interest
  - Generally do not want to adjust for mediators unless attempting to dissect and explain the causal pathway of interest
Types of Adjustment Variables

**Precision variables**

- I define a **precision variable** as a covariate that is related to the outcome $Y$, but independent of the predictor of interest $X$.

Predictor ($X$)  ➔  Outcome ($Y$)

Precision Variable ($W$)
Types of Adjustment Variables

Example: Precision variable

- Example: In a controlled experiment, we randomize patients to an experimental cancer treatment or placebo and look at the proportion of patients who relapse on each arm:
  - Age may be associated with the probability of relapse
  - Because of randomization, age is independent of whether treatment was received

Why precision?

- Why do I refer to this as a precision variable? Coming soon...
- In many cases, adjustment for a precision variable is a good idea!
Types of Adjustment Variables

Nuisance variables

- I define a **nuisance variable** as a covariate that is independent of the outcome $Y$, but may or may not be related to the predictor of interest $X$. 

![Diagram showing the relationship between predictor (X), outcome (Y), and nuisance variable (W)]
Types of Adjustment Variables

Example: Nuisance variable

- Example In a controlled experiment, we randomize patients to an experimental cancer treatment or placebo and look at the proportion of patients who relapse on each arm:
  - Shoe color on the day of randomization is not likely to be associated with the probability of relapse

Adjustment for nuisance parameters is not a good thing

- We are trying to model the outcome $Y$
- We do not want to intentionally include covariates that (we believe) are not associated with $Y$
Result of Confounding

Adjusted vs. unadjusted covariate effects

1. Unadjusted model: \( E[Y_i] = \beta_0 + \beta_1 X_i \)
   - \( \beta_1 \) is the difference in the mean of \( Y \) for groups differing by 1-unit in \( X \)

2. Adjusted model: \( E[Y_i] = \gamma_0 + \gamma_1 X_i + \gamma_2 W_i \)
   - \( \gamma_1 \) is the difference in the mean of \( Y \) for groups differing by 1-unit in \( X \), but agreeing in their value of \( W \)
The Role of Adjustment Variables in Regression Modeling

Effect modifiers
Confounders
Mediators
Precision variables
Nuisance variables

Adjusted vs. unadjusted effects

Proposition 1: Let $\hat{\beta}_1$ denote the OLS estimate of $\beta_1$. Then under the adjusted model,

$$
E[\hat{\beta}_1] = \gamma_1 + \frac{\text{cov}(X, W)}{\text{var}(X)} \gamma_2
$$

$$
= \gamma_1 + r_{XW} \sqrt{\frac{\text{var}(W)}{\text{var}(X)}} \gamma_2
$$

where $r_{XW}$, $\text{var}(X)$, and $\text{var}(W)$ are the sample correlation between $X$ and $W$, sample variance of $X$, and sample variance of $W$, respectively.
Confounding and Bias

Proof:
Result of Confounding

The implication...

- \( \hat{\beta}_1 \) is biased (and inconsistent) for \( \gamma_1 \) unless at least one of the following hold

  1. \( r_{XW} = 0 \): \( X \) and \( W \) are uncorrelated (in the sample), OR
  2. \( \gamma_2 = 0 \): \( W \) is not related to \( Y \)

- In either case, \( \hat{\beta}_1 \) is unbiased (and consistent) for \( \beta_1 \)

- Implication for confounders?
  
  - By definition, a confounder is related to the predictor of interest and the response
  
  - This implies that if \( W \) is a confounder, then both conditions above fail
  
  - Hence the parameter from the reduced model is biased for the adjusted estimate
The real question...

- We never know the ‘true’ model!

- Big Question: What do we want to hold constant when estimating the association between $Y$ and $X$?

- The answer to this defines the interpretation of our result...
Precision of Estimators

Relationship between the precision of unadjusted and adjusted estimates

- Consider the following linear regression models:

1. Unadjusted model: \( E[Y_i] = \beta_0 + \beta_1 X_i \)

2. Adjusted model: \( E[Y_i] = \gamma_0 + \gamma_1 X_i + \gamma_2 W_i \)
Precision of Estimators

Relationship between the precision of unadjusted and adjusted estimates

► Proposition 2:
  1. For the unadjusted model,

\[
\text{Var}[\hat{\beta}_1] = \frac{\sigma_{Y|X}^2}{n\text{var}(X)}
\]

  2. For the adjusted model,

\[
\text{Var}[\hat{\gamma}_1] = \frac{\sigma_{Y|X,W}^2}{n\text{var}(X)(1 - r_{XW}^2)}
\]

  where \( \sigma_{Y|X,W}^2 = \sigma_{Y|X}^2 - \gamma_2^2 \text{var}(W|X) \)

► Hence, if \( \gamma_2 \neq 0 \) then \( \sigma_{Y|X,W}^2 < \sigma_{Y|X}^2 \)

Proof: Homework 2!
## Implications of Propositions 1 & 2 (generalizeable to $p$ covariate case)

- **Case 1:** $r_{XW} = 0$ ($X$ and $W$ uncorrelated) and $\gamma_2 = 0$ ($W$ and $Y$ unrelated)
  - From Proposition 1, $\hat{\beta}_1$ unbiased for $\gamma_1$
  - From Proposition 2, $\text{Var}[\hat{\beta}_1] = \text{Var}[\hat{\gamma}_1]$
  - **Conclusion:** Lose 1 degree of freedom for hypothesis tests and CIs if adjusting for $W$
To adjust or not to adjust...

Implications of Propositions 1 & 2 (generalizeable to $p$ covariate case)

- **Case 2**: $r_{XW} \neq 0$ ($X$ and $W$ correlated) and $\gamma_2 = 0$ ($W$ and $Y$ unrelated)
  - From Proposition 1, $\hat{\beta}_1$ unbiased for $\gamma_1$
  - From Proposition 2, $\text{Var}[\hat{\beta}_1] < \text{Var}[\hat{\gamma}_1]$
  - **Conclusion**: Mathematically estimating the same quantity but lose precision when adjusting for $W$ (nuisance variable)
To adjust or not to adjust...

Implications of Propositions 1 & 2 (generalizeable to $p$ covariate case)

- Case 3: $r_{XW} = 0$ ($X$ and $W$ uncorrelated) and $\gamma_2 \neq 0$ ($W$ and $Y$ related)
  
  - From Proposition 1, $\hat{\beta}_1$ unbiased for $\gamma_1$
  
  - From Proposition 2, $\text{Var}[\hat{\beta}_1] > \text{Var}[\hat{\gamma}_1]$
  
  - **Conclusion**: Mathematically estimating the same quantity but *gain* precision when adjusting for $W$ (precision variable)
To adjust or not to adjust...

Implications of Propositions 1 & 2 (generalizeable to $p$ covariate case)

- Case 4: $r_{XW} \neq 0$ ($X$ and $W$ correlated) and $\gamma_2 \neq 0$ ($W$ and $Y$ related)
  - From Proposition 1, $\hat{\beta}_1$ biased for $\gamma_1$
  - From Proposition 2, no definitive statement about the variances
  - Conclusion: $W$ is a confounder and decision to adjust should be based on what you are trying to estimate.