

# Biost 518 / Biost 515 Applied Biostatistics II / Biostatistics II



Timothy A. Thornton, Ph.D.  
Associate Professor of Biostatistics  
University of Washington

## Lecture 22 Adjusting for Confounding

# Outcomes often have multiple associated variables

- In this course we have covered a variety of methods to assess associations between two variables, i.e., by comparing the distribution of a response variable across groups defined by a single predictor variable of interest
  - Z test, t test,  $\chi^2$  test, linear regression, logistic regression, logrank test of Kaplan-Meier survival distributions
- Many outcomes of scientific interest, however, are expected to be influenced by many factors.

# Outcomes often have multiple associated variables

- Example: coronary heart disease (CHD)
  - CHD is a complex outcome that is associated with many risk factors.
  - People with diets that are high in fat diet tend to have higher low-density lipoprotein (LDL) cholesterol as compared to people with a low fat diet
  - High LDL is a risk factor for CHD
  - People with diets that are high in fat are also more likely to smoke and be overweight, factors which are also strongly associated with CHD risk.

# Definition of Confounding



- When assessing associations between two variables, we have to be aware of potential **confounding** due to other variables.
- A **confounding variable** is an extraneous variable that (1) is associated with both a predictor of interest (POI) variable and outcome (or response) variable, and (2) distorts the observed relationship between the POI and outcome.

# Definition of Confounding

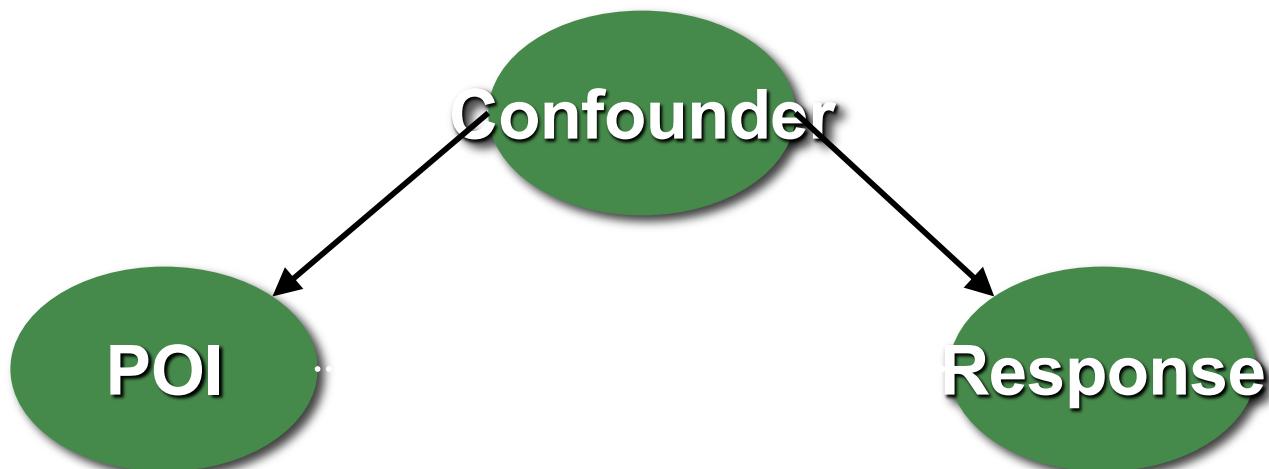


- The association between a predictor of interest and an outcome variable is **confounded** by a third variable if
  - The third variable is associated with the POI in the sample, e.g., the distribution of the confounder is different within subgroups defined by the POI
  - The third variable is associated with the outcome causally (in truth), e.g., is a risk factor (or protective factor) for the outcome of interest
  - The third variable is associated with the outcome in groups that are homogeneous with respect to the predictor of interest
  - The third variable is not an intermediate step in the causal pathway between the POI and outcome

# Classical Confounder

.....

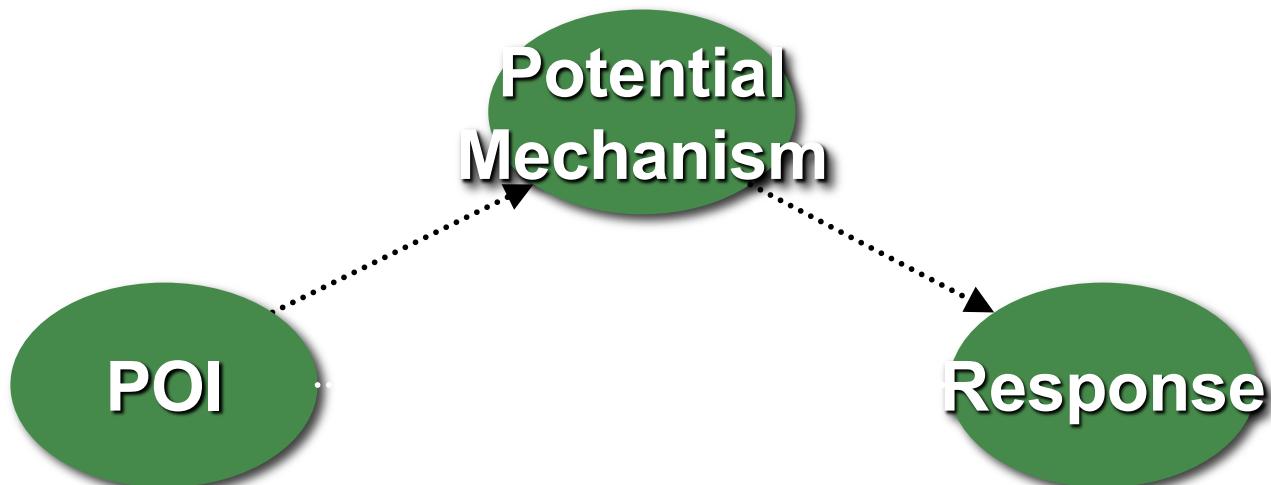
- A clear case of confounding is when some third variable is a “cause” of both the POI and response
- We generally adjust for such a confounder



# Causal Pathway

.....

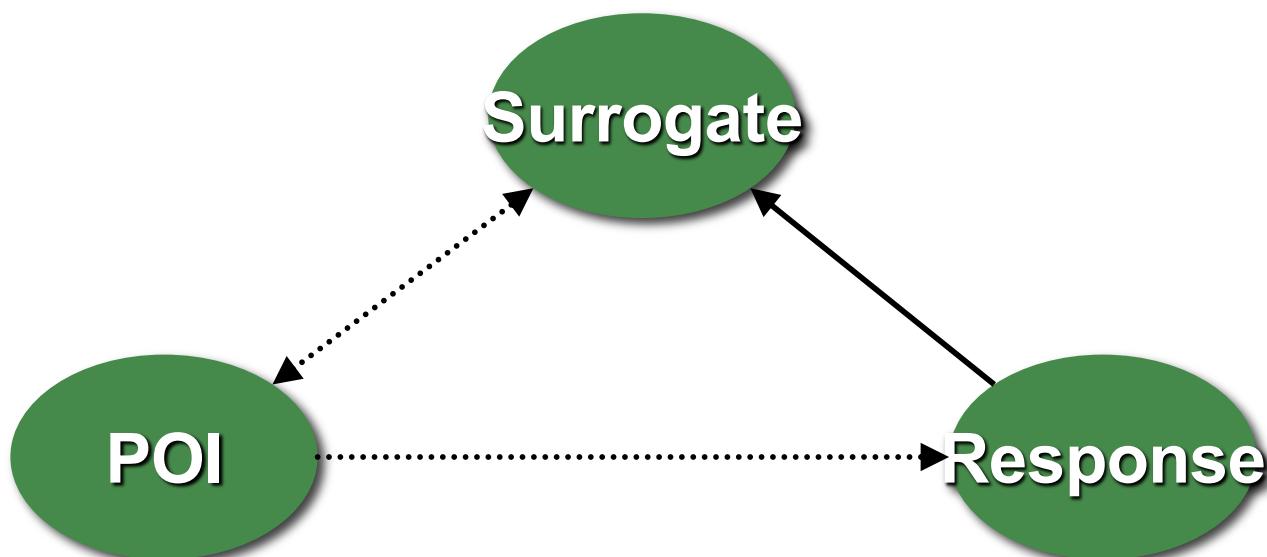
- A variable in the causal pathway of interest is not a confounder
- We **would not** adjust for such a variable (lest we lose the ability to detect the effect)



# Surrogate for Response

.....

- Adjustment for such a variable is a very BAD thing to do
- If interested in the association between pregnant mother's diet and birth weight of newborn child, we would not want to adjust for the head circumference of the newborn.



# Example of Confounding: Hospital Mortality and Rankings

The New York Times

U.S.

WORLD U.S. N.Y. / REGION BUSINESS TECHNOLOGY SCIENCE HEALTH SPORTS OPINION

POLITICS EDUCATION TEXAS

## Report on Hospital Death Rates Is Challenged

By MARTIN TOLCHIN, Special to The New York Times

Published: January 12, 1990

**WASHINGTON, Jan. 11**— The results of a Government study on death rates in nearly 6,000 hospitals were challenged today by researchers who said the Federal analysis failed to account for variations in the severity of patients' illness when they were hospitalized.

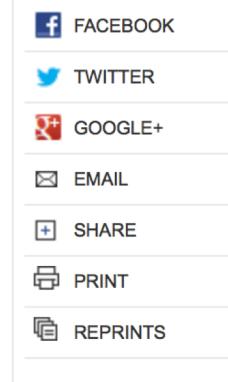
As a result, they said, some hospitals were unfairly treated in the findings, which named hospitals with higher-than-expected death rates.

The new study is being published Friday in The Journal of the American Medical Association.

The Government's report, issued in December by the Department of Health and Human Services, found that deaths exceeded Federal predictions in 196 hospitals. The Government study is intended to help patients and their families choose hospitals.

### Skepticism Is Recommended

Dr. Jesse Green, a health research analyst at the New York University School of Medicine who is the principal author of the new study, said, "The results suggest that consumers and health regulators should approach the Government's mortality rates with skepticism and be cautious about viewing high mortality ratings as an indication of quality of care."



# Example of Confounding: Hospital Mortality



ARTICLE | January 12, 1990

## The Importance of Severity of Illness in Assessing Hospital Mortality

Jesse Green, PhD; Neil Wintfeld, PhD; Phoebe Sharkey, PhD; Leigh J. Passman, MD, PhD

JAMA. 1990;263(2):241-246. doi:10.1001/jama.1990.03440020075036.

Text Size: A A A

Article

References

### ABSTRACT

ABSTRACT | REFERENCES



Each year, the Health Care Financing Administration (HCFA) releases a report comparing hospital mortality rates with predicted rates. Some argue that the HCFA's prediction model does not adequately account for patient severity. We tested this hypothesis by comparing the HCFA's model (replicated as closely as we could) to a second that added a severity measure (the Stage of Principal Diagnosis at Admission, a subscale of the Severity of Illness Index). In our simulation, the HCFA's model had very limited capacity to predict mortality (average  $R^2$ , 2.5%). Patients grouped according to admission severity had markedly different mortality rates, which the HCFA's model's predictions could not differentiate. The HCFA model also failed to predict large differences in mortality between hospitals with low- and high-severity admissions. Adding severity to the HCFA's model yielded more than an eightfold increase in the  $R^2$ , to 21.5%, and reduced instances of higher than expected hospital mortality to chance levels. These findings suggest that the HCFA's mortality release needs to be made much more sensitive to admission severity before it can be used to make valid inferences about the quality or effectiveness of hospital care.

(JAMA. 1990;263:241-246)

# Example of Confounding: Hospital Mortality

- Forgetting to account for severity of illness, a confounding variable, can affect our conclusions about the relationship between two variables, hospital quality of care and death rate
- Simplified Example of this: When looking only at hospital and death rate

	Hospital A	Hospital B
Died	63	16
Survived	2037	784
Total	2100	800

$$\widehat{OR} = \frac{(63)(784)}{(16)(2037)} = 1.515$$

	Hospital A	Hospital B
Died	3%	2%
Survived	97%	98%
Total	100%	100%

# Example of Confounding: Hospital Mortality

- Account for confounding variable: severity of illness

Patients not so severe		Patients severely ill			
	Hosp A	Hosp B			
Died	6	8	Died	57	8
Surv	594	592	Surv	1443	192
Total	600	600	Total	1500	200

$$\widehat{OR}_{NS} = \frac{(6)(592)}{(8)(594)} = 0.747$$

$$\widehat{OR}_S = \frac{(57)(192)}{(8)(1443)} = 0.948$$

Patients not so severe		Patients severely ill			
	Hosp A	Hosp B			
Died	1.00%	1.33%	Died	3.80%	4.00%
Surv	99.00%	98.67%	Surv	96.20%	96.00%
Total	100.00%	100.00%	Total	100.00%	100.00%

# Simpson's Paradox



- Simpson's paradox: trends for an association that are observed within strata but disappear or reverse when the strata are combined.
- This can occur with continuous or binary outcomes
- Given binary variables  $Y$  (response),  $X$  (POI),  $Z$  (strata), as we saw in the previous examples, it is possible to have

$$\Pr( Y=1 | X=1, Z=1 ) > \Pr( Y=1 | X=0, Z=1 )$$

$$\Pr( Y=1 | X=1, Z=0 ) > \Pr( Y=1 | X=0, Z=0 )$$

but to have

$$\Pr( Y=1 | X=1 ) < \Pr( Y=1 | X=0 )$$

# Addressing Confounding



- How do we handle the confounding effects of alcohol use?
- We can address it in the *design* of our study.
  - Randomized control studies offers the best protection against confounding since randomization implies that, on average, treatment assignment is not associated with other factors
  - Can do a matched pair study design study where individuals who have similar characteristics (age, sex, health, etc.) at baseline that might be confounders are matched and compared
- We can also address confounding in the *statistical analysis* of the data.
- (Not an either/or choice!)

# Matched Pair Design



- We briefly discussed the matched pair design in lecture 7.
- A matched pair study design can be used to protect against confounding.
- In a matched pair design:
  - Subjects are matched in pairs according to some characteristics, e.g., potential confounders
  - Two measurements taken on the same subject, e.g. before treatment and after treatment
- To conduct statistical inference on such a sample, we analyze the ***difference*** using a one-sample *t* test procedure.
  - Under the null hypothesis of no effect, the difference should be 0.

# Diet Example: Matched Pair Design



-----  
I

Subject	Weight Before Diet	Weight After Diet	Difference
1	187	160	27
2	175	153	22
3	158	150	8
4	160	148	12
5	130	127	3
6	170	160	10
7	165	150	15
8	154	155	-1
9	177	167	10
10	190	184	6
11	139	130	9
12	145	147	-2
13	166	160	6
14	170	152	18
15	152	147	5
16	143	140	3
17	163	169	-6
18	154	139	15
19	182	160	22
20	145	140	3
mean	161.25	151.9	9.35
SD	16.36	13.46	8.56

# Diet Example: Matched Pair Design



To ascertain whether the diet reduces weight, we test

$$H_0 : \mu = 0 \quad H_a : \mu > 0$$

where  $\mu$  is the mean weight difference.

$$T\text{-statistic: } t = \frac{9.35 - 0}{8.56 / \sqrt{20}} = 4.88$$

$$p\text{-value: } p = P(t_{19} \geq 4.88) = 5.2 \times 10^{-5}$$

# Addressing Confounding by Analysis: Stratification



- **Stratification** is one useful approach for discovering and correcting confounding in an analysis, where a measure of association between an outcome and predictor is calculated within strata.
- In order to do this, we must first consider our beliefs about the causal relationships among the measured variables
- **Very Important:** We will not be able to assess causal relationships in our statistical analysis
  - Inference of causation comes only from study design
- However, consideration of hypothesized causal relationships helps us decide which statistical question to answer

# Stratification: Mantel-Haenszel test



- The Mantel–Haenszel test (MH) is often used in the analysis of stratified categorical data to protect against confounding.
- Using stratified  $2 \times 2$  tables, it allows for assessing association using between a binary outcome variable and a binary predictor by taking into account the stratification,

Notation for  $k$  stratified  $2 \times 2$  tables (i.e. a  $2 \times 2 \times k$  table)

	$Y=1$	$Y=0$	total
$X=1$	$A_j$	$B_j$	$A_j + B_j$
$X=0$	$C_j$	$D_j$	$C_j + D_j$
total	$A_j + C_j$	$B_j + D_j$	$n_j = A_j + B_j + C_j + D_j$

Within each strata  $j = 1 \dots k$ ,  $\widehat{OR}_j = \frac{A_j D_j}{B_j C_j}$

# Stratification: Mantel-Haenzsel test



- The Mantel-Haenzsel estimate of the odds ratio is a weighted average of the strata-specific estimates – it gives more weight to strata that provide the most information about the association

$$\widehat{OR}_{MH} = \frac{\sum_{j=1}^k A_j D_j / n_j}{\sum_{j=1}^k B_j C_j / n_j} = \sum_{j=1}^k \frac{B_j C_j / n_j}{\sum_{j'=1}^k B_{j'} C_{j'} / n_{j'}} \widehat{OR}_j$$

# Hospital Mortality Example: Mantel-Haenzsel test

- Account for confounding variable: severity of illness

Patients not so severe		Patients severely ill		
	Hosp A	Hosp B		
Died	6	8	Died	57
Surv	594	592	Surv	1443
Total	600	600	Total	1500
				200

Patients not so severe		Patients severely ill		
	Hosp A	Hosp B		
Died	1.00%	1.33%	Died	3.80%
Surv	99.00%	98.67%	Surv	96.20%
Total	100.00%	100.00%	Total	100.00%
				100.00%

$$\widehat{OR}_{MH} = \frac{(6)(592)/1200 + (57)(192)/1700}{(8)(594)/1200 + (8)(1443)/1700} = 0.874$$

- In R, can conduct a Mantel-Haenzel test using the **mantelhaen.test()** function

# Hospital Mortality Example: Mantel-Haenzsel test in R...

- Can perform a Mantel-Haenzsel test in R using the **mantelhaen.test()** function

```
> Hospital<-c(rep(1,2100),rep(0,800))
> Severity<-c(rep(0,600),rep(1,1500),rep(0,600),rep(1,200))
> Death<-c(rep(1,6),rep(0,594),rep(1,57),rep(0,1443),rep(1,8),rep(0,592),rep(1,8),rep(0,192))
> mantelhaen.test(Hospital,Death,Severity)
```

Mantel-Haenszel chi-squared test with continuity correction

```
data: Hospital and Death and Severity
Mantel-Haenszel X-squared = 0.073078, df = 1, p-value = 0.7869
alternative hypothesis: true common odds ratio is not equal to 1
95 percent confidence interval:
0.4753825 1.6074245
sample estimates:
common odds ratio
0.8741519
```

# Symptoms of Confounding: Unadjusted, ...Adjusted Analyses...

- Confounding typically produces a difference between unadjusted and adjusted analyses, but those symptoms are not proof of confounding
- Estimates of association from unadjusted analysis are markedly different from estimates of association from adjusted analysis are a symptom of confounding
  - Association within each stratum is similar to each other, but different from the association in the combined data
- Note that such a difference can occur at times when there is no confounding. Summary measures which are nonlinear functions of the mean sometimes show the above symptoms in the absence of confounding
  - Odds (and odds ratios)
  - Hazards (and hazard ratios)

# Adjusting for Confounding in Regression



- In linear regression, differences between adjusted and unadjusted analyses are a diagnostic of confounding
- Biost 515/518 will focus on multiple linear regression, which allows for adjustment of additional variables when assessing an association between a response and predictor of interest.

# Simple Linear Regression Model



- Simple linear regression of response  $Y$  on predictor of interest (POI)  $X$
- Mean for subpopulation with  $X = x$  from the simple linear regression model is:

Model

$$X_i = 0$$

$$X_i = x$$

$$X_i = x + 1$$

$$E[Y_i | X_i] = \beta_0 + \beta_1 \times X_i$$

$$E[Y_i | X_i = 0] = \beta_0$$

$$E[Y_i | X_i = x] = \beta_0 + \beta_1 \times x$$

$$E[Y_i | X_i = x + 1] = \beta_0 + \beta_1 \times x + \beta_1$$

# Multiple Regression Model



$Y_i$

Response for the  $i$ th subject

$X_i$

POI for the  $i$ th subject

$\vec{W}_i = (W_{1i}, W_{2i}, \dots, W_{pi})$  Additional  $p$  covariates for the  $i$ th subject

- Multiple linear regression model is:

$$E[Y_i | X_i, \vec{W}_i] = \beta_0 + \beta_1 X_i + \beta_2 W_{1i} + \cdots + \beta_{p+1} W_{pi}$$

- In the literature, multiple linear regression is often written as

$$E[Y_i | \vec{X}_i] = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_p X_{pi}$$

# FEV Dataset



- Association between lung function and self reported smoking in children
- Compare geometric means of FEV of children who smoke to comparable nonsmokers
- Restrict analysis to children 9 yo and older
  - No smokers less than 9
  - Still about 6 : 1 ratio of nonsmokers to smokers
    - Little precision gained by keeping younger children
    - Borrowing information from young kids problematic if not a linear relationship between  $\log(\text{FEV})$  and predictors
      - With confounding, want to get model correct

# Compare Alternative Models



- Real life:
  - We should choose a single model in advance of looking at the data
- Academic exercise for this lecture
  - Observe what happens to parameter estimates and SE across models
    - Smoking
    - Smoking adjusted for age

# Ex: FEV and Smoking



```
> library(uwIntroStats)
> fevdata2<-subset(fevdata,age>=9)
> fevdata2$logfev<-log(fevdata2$fev)
> fevdata2$smoker<-(2-fevdata2$smoke)
> logregmod1<- regress ("mean",logfev~smoker,data=fevdata2)
> logregmod1
```

Call:

```
regress(fnctl = "mean", formula = logfev ~ smoker, data = fevdata2)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.68110	-0.15903	-0.00525	0.15579	0.69848

Coefficients:

	Estimate	Naive SE	Robust SE	95%L	95%H	F stat	df	Pr(>F)
[1] Intercept	1.058	0.01281	0.01293	1.033	1.084	6693.87	1	< 0.00005
[2] smoker	0.1023	0.03328	0.03165	0.04009	0.1645	10.45	1	0.0013

Residual standard error: 0.2477 on 437 degrees of freedom

Multiple R-squared: 0.02117, Adjusted R-squared: 0.01893

F-statistic: 10.45 on 1 and 437 DF, p-value: 0.001322

# Unadjusted Interpretation: Intercept



- Geometric mean of FEV in nonsmokers is 2.88 l/sec
  - The scientific relevance is questionable here, because we do not really know the population our sample represents
    - Comparing smokers to nonsmokers is more useful than looking at either group by itself
  - (Calculations:  $e^{1.058} = 2.881$ )
  - (The P value is of no importance whatsoever, it is testing that the log geometric mean is 0 or that the geometric mean is 1. Why would we care?)
- (Because smoker is a binary variable, the estimate corresponds to the sample geometric mean)

# Unadjusted Interpretation: Smoker Slope



- Geometric mean of FEV is 10.8% higher in smokers than in nonsmokers (95% CI: 4.1% to 17.9% higher)
  - These results are atypical of what we might expect with no true difference between groups:  $P = 0.001$
  - (Calculations:  $e^{0.102} = 1.108$ ;  $e^{0.040} = 1.041$ ;  $e^{0.165} = 1.179$ )
    - (Note that  $\exp(x)$  is approx  $1+x$  for  $x$  close to 0)
- (Because smoker is a binary (0-1) variable, this analysis is nearly identical to a two sample t test allowing for unequal variances)

# Ex: Adjusted for Age



```
> logregmod2<- regress ("mean",logfev~smoker+age,data=fevdata2)
> logregmod2
```

Call:

```
regress(fnctl = "mean", formula = logfev ~ smoker + age, data = fevdata2)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.6107	-0.1506	0.0028	0.1462	0.5383

Coefficients:

	Estimate	Naive SE	Robust SE	95%L	95%H	F stat	df	Pr(>F)
[1] Intercept	0.3518	0.05452	0.05750	0.2388	0.4648	37.44	1	< 0.00005
[2] smoker	-0.05135	0.03046	0.03438	-0.1189	0.01623	2.23	1	0.136
[3] age	0.06360	4.811e-03	5.14e-03	0.05349	0.07370	153.08	1	< 0.00005

Residual standard error: 0.2095 on 436 degrees of freedom

Multiple R-squared: 0.3012, Adjusted R-squared: 0.298

F-statistic: 82.28 on 2 and 436 DF, p-value: < 2.2e-16

# Age Adjusted Interpretation: Intercept



- Geometric mean of FEV in newborn nonsmokers is 1.42 l/sec
  - Intercept corresponds to the log geometric mean in a group having all predictors equal to 0
  - There is no scientific relevance is here, because we are extrapolating outside our data
  - (Calculations:  $e^{0.352} = 1.422$ )

# Age Adjusted Interpretation: Age Slope



- Geometric mean of FEV is 6.6% higher for each year difference in age between two groups with similar smoking status(95% CI: 5.5% to 7.6% higher for each year difference in age)
- These results are highly atypical of what we might expect with no true difference in the geometric mean FEV between age groups having similar smoking status:  $P < 0.0005$

# Age Adjusted Interpretation: Smoker Slope

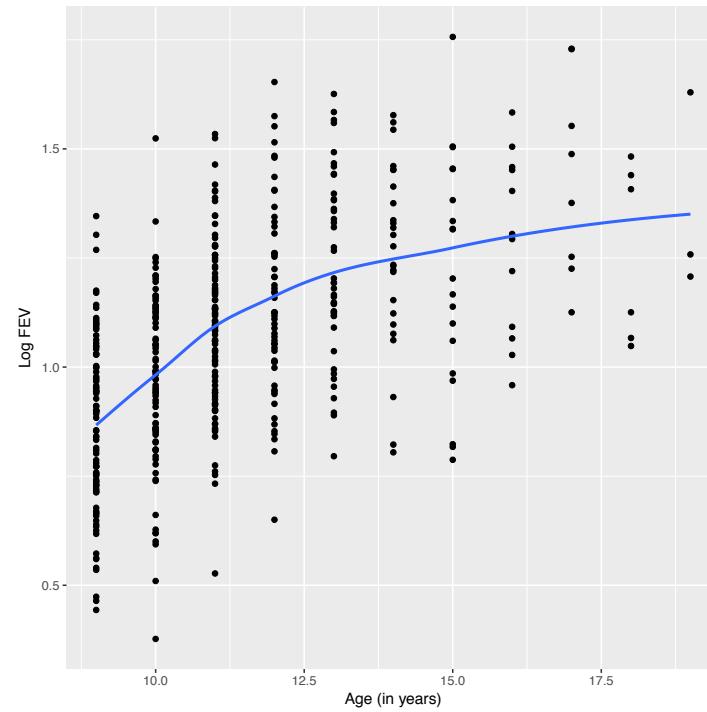
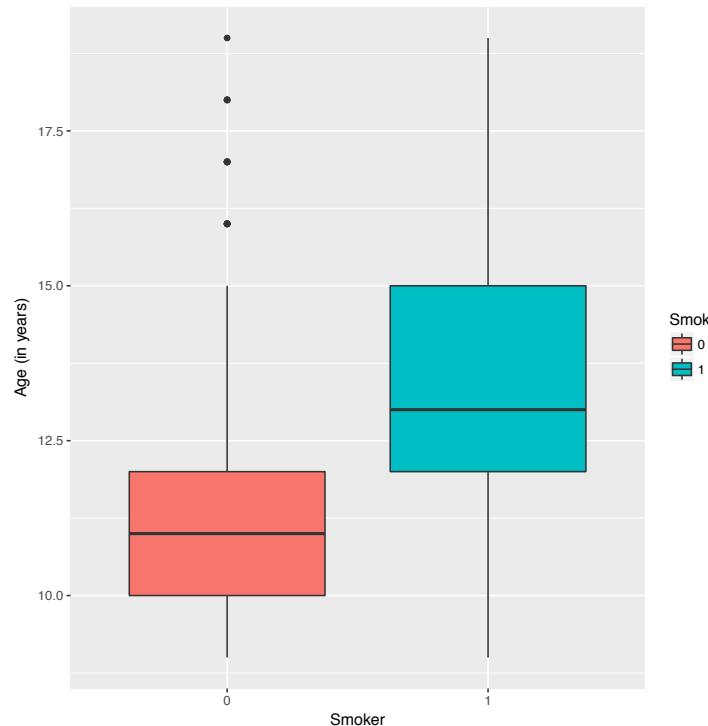


- Geometric mean of FEV is 5.0% lower in smokers than in nonsmokers of the same age (95% CI: 11.2% lower to 1.6% higher)
- These results are not atypical of what we might expect with no true difference between groups of the same age:  $P = 0.136$ 
  - Lack of statistical significance is also evident because the confidence interval contains 1 (as a ratio) or 0 (as a percent difference)
- (Calculations:  $e^{-0.051} = 0.950$ ;  $e^{-0.119} = 0.888$ ;  $e^{0.016} = 1.016$ )
  - (Note that  $\exp(x)$  is approx  $1+x$  for  $x$  close to 0)

# Age Adjusted Comments



- Comparing unadjusted and age adjusted smoking-FEV association analyses
- Marked difference in effect of smoking suggests that there was indeed confounding
  - Age is a relatively strong predictor of FEV
  - Age is associated with smoking in the sample
    - Mean (SD) of age in analyzed nonsmokers: 11.1 (2.04)
    - Mean (SD) of age in analyzed smokers: 13.5 (2.34)



# Age Adjusted Comments

