Interactions and Model Building Strategies

Announcements

- Midterm in 1 week (May 10 at 1-2:50pm)

- Still working on project instructions – please use similar division of work as last quarter

HW3 & 4 Redos

- **HW 4, Question 1, part h:** Make sure you use the fitted model with Age and CPR to calculate the 10 year increase in age.

↳ slack post estimable()

- Thank you all for your patience as I catch up on our project work, midterm, and class!

Class 9 Learning Objectives

Not coding in real time & not manual calc

1. Determine if an additional independent variable is a not a confounder nor effect modifier, is a confounder but not effect modifier, or is an effect modifier.

part 1, part 2
up to class 8

2. Calculate and interpret odds ratios for interactions

3. Understand the overall steps for purposeful selection as a model building strategy
   - Connect these steps back to the tests of coefficients in Class 8

Class 9 Learning Objectives

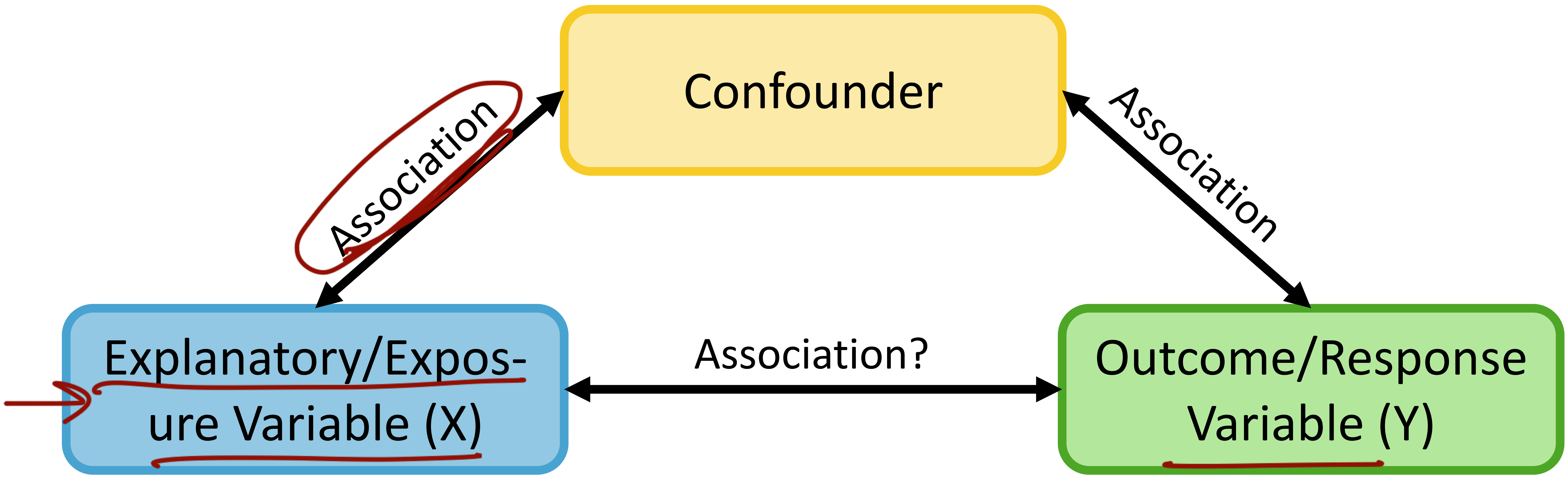
1. Determine if an additional independent variable is a not a confounder nor effect modifier, is a confounder but not effect modifier, or is an effect modifier.

2. Calculate and interpret odds ratios for interactions

3. Understand the overall steps for purposeful selection as a model building strategy
    - Connect these steps back to the tests of coefficients in Class 8

Re-visit Confounding

- A confounding variable, or **confounder**, is a factor/variable that wholly or partially accounts for the observed effect of the risk factor on the outcome
- A confounder must be
  - related to the outcome Y, but not a consequence of Y.
  - related to the explanatory variable X, but not a consequence of X.



Test for Confounding (I)

variable of interest

- We may check for confounding by comparing the estimated coefficient for the risk factor variable from models with and without the confounder
  - First fit a simple logistic regression model with single covariate x_1 (the risk factor), let θ_1 be the unadjusted estimate of the coefficient for X_1

$$\text{logit}(\pi(x)) = \beta_0 + \theta_1 x_1$$

  - Then fit a multiple logistic regression model with multiple covariate. For example, two covariate x_1 (the risk factor) and x_2 (the potential confounder), let β_1 be the estimate of the coefficient for x_1

$$\text{logit}(\pi(x)) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

Test for Confounding (I)

- Compute the percent change in the coefficient

w/out confounder

$$\Delta\hat{\beta}\% = 100\% \times \frac{(\hat{\theta}_1 - \hat{\beta}_1)}{\hat{\beta}_1}$$

w/ confounder

$\Delta\hat{\beta}\% > 20\% \Rightarrow$ consider x_2 to be a confounder

- What threshold should we use to include a confounder in a model?
  - **No real agreement in the statistical/epi/clinical literature**
  - Some say if change in coefficient is **greater than 20%,** a confounder should be included in the model
  - Others use changes as low as 10%, and others may use 25%
  - **For our class, let's follow the Hosmer and Lemeshow textbook, and use 20%**

We have the following two fitted models: $$\text{logit}(\pi(x)) = -3 + 0.4x_1$$ and $$\text{logit}(\pi(x)) = -3.05 + 0.49x_1 - 0.17x_2$$. Using the "greater than 20%...

https://www.polleverywhere.com/multiple_choice_polls/PApJp3fAtJfApDmFsAYjc

Poll Everywhere Question 1

Respond at **PollEv.com/nickywakim275**

We have the following two fitted models:

simple no x_2 → $\text{logit}(\pi(x)) = -3 + 0.4x_1$ **and**

fitted model including x_2 → $\text{logit}(\pi(x)) = -3.05 + 0.49x_1 - 0.17x_2$.

Using the "greater than 20%" rule, is x_2 a confounder?

x_2 is not a confounder — 73%

~~x_2~~ is a confounder — 27%

$$\Delta\beta\% = \frac{\hat\theta_1 - \hat\beta_1}{\hat\beta_1}$$

$\hat\theta_1 = 0.4$

$\hat\beta_1 = 0.49$

$$= \frac{0.4 - 0.49}{0.49}$$

$\Delta\beta\% = 18\%$

1 / 7 Instructions Responses More EXIT

Powered by Poll Everywhere

Test for Confounding (III)

- We may test the significance of any variable (including confounder) using hypothesis testing (say, likelihood ratio test)

  - It is generally suggested to **include a confounder** in the model if it is **statistically significantly associated with the outcome** itself

  - A **statistically insignificant confounder may stay in the model** if it is determined that the change in the estimated coefficient for the risk factor with and without the risk factor is **"clinically or biologically important"**

    - Whether it reaches a desired change threshold or not!

Confounding vs. Interaction

Please refer to your Day 10 notes from BSTA 512/612 – lots of information about these concepts!

- **Confounders:** The adjusted odds ratio for one variable adjusting for confounders can be quite different from unadjusted odds ratio
  - Adjusting for them is called *controlling for confounding*.
- **Interactions:** When odds ratio for one variable is not constant over the levels of another variable, the two variables are said to have a *statistical interaction* (sometimes also called *effect modification)*
  - i.e.: the log odds of one variable is modified/changed with different values of the other variable
  - A variable is an ***effect modifier*** if it interacts with a risk factor

Interaction

- We learned that multiple logistic regression (MLR) model can be used for adjusting for confounders
  - Confounders can be either categorical or continuous independent variable
  - The adjusted OR using MLR is often very close to the adjusted OR using Mantel-Haenszel method (Class 4)

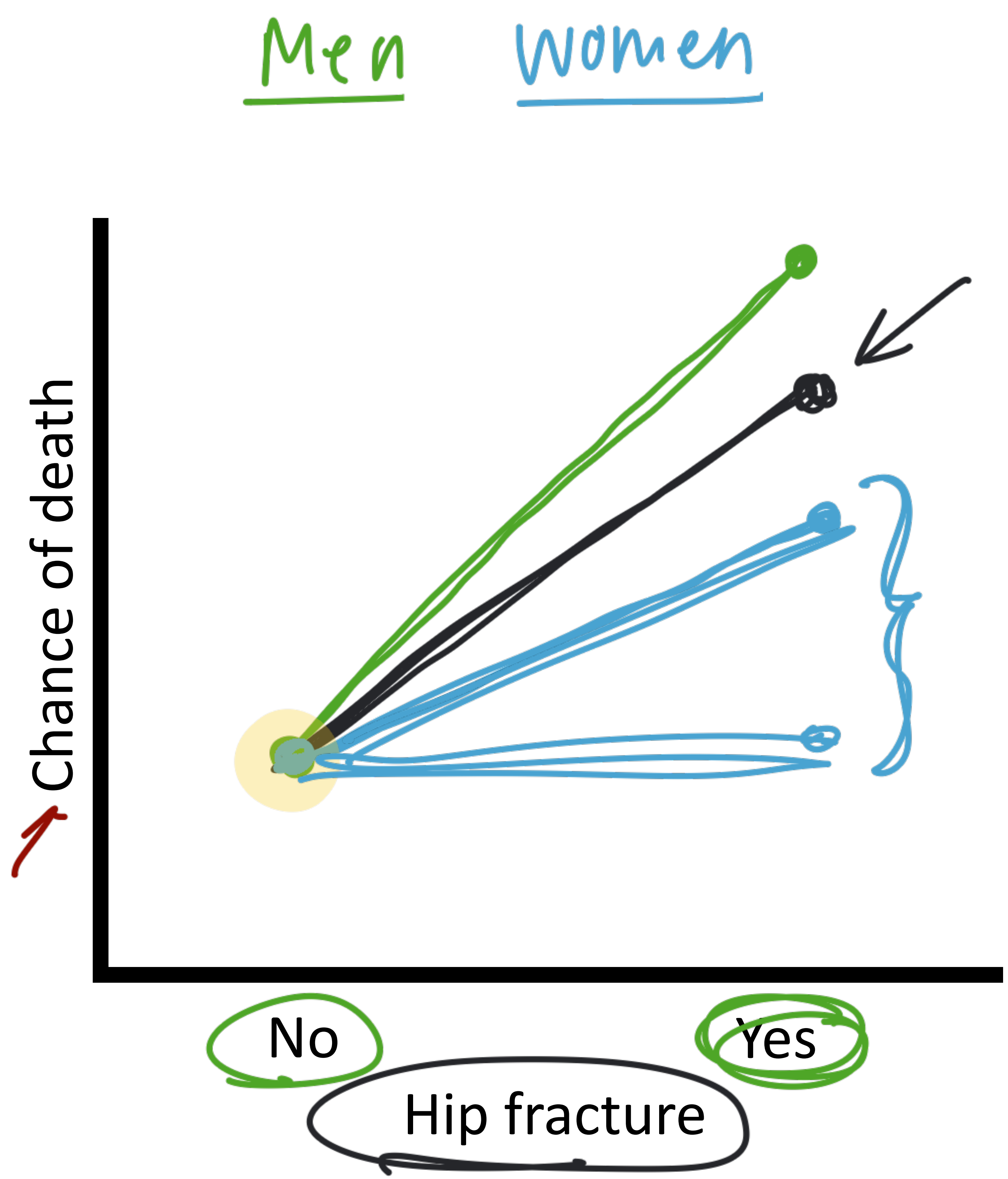
- Similarly, the single adjusted OR using MLR should ONLY be reported when there is no interaction
  - No interaction means the OR across strata is homogeneous

$$\text{logit}(\pi(x)) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

$$OR = e^{\hat{\beta}_1} \text{ controlling for } X_2$$

Example of interaction

- In a cohort study of elderly people the chance of death (outcome) within 2 years was much higher for those who had previously suffered a hip fracture at the start of these 2 years, but the excess risk associated with a hip fracture was significantly higher for men than women
- This is an interaction between hip fracture status (yes/no) and ~~sex~~ gender (unclear if assigned at birth or no)
- Odds ratio for ~~females~~ women ≠ odds ratio for ~~males~~ men



Types of Interaction (I)

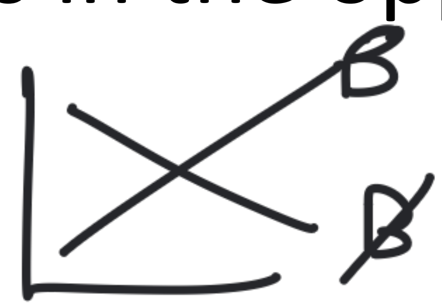
No interaction and three potential effects of interaction between two covariates A and B:

a) **No interaction** between A and B (confounder with no interaction)

b) **Unilateralism:** exposure to A has no effect in the absence of exposure to B, but a considerable effect when B is present.

c) **Synergism:** the effect of A is in the same direction, but stronger in the presence of B.

d) **Antagonism:** the effect of A works in the opposite direction in the presence of B.



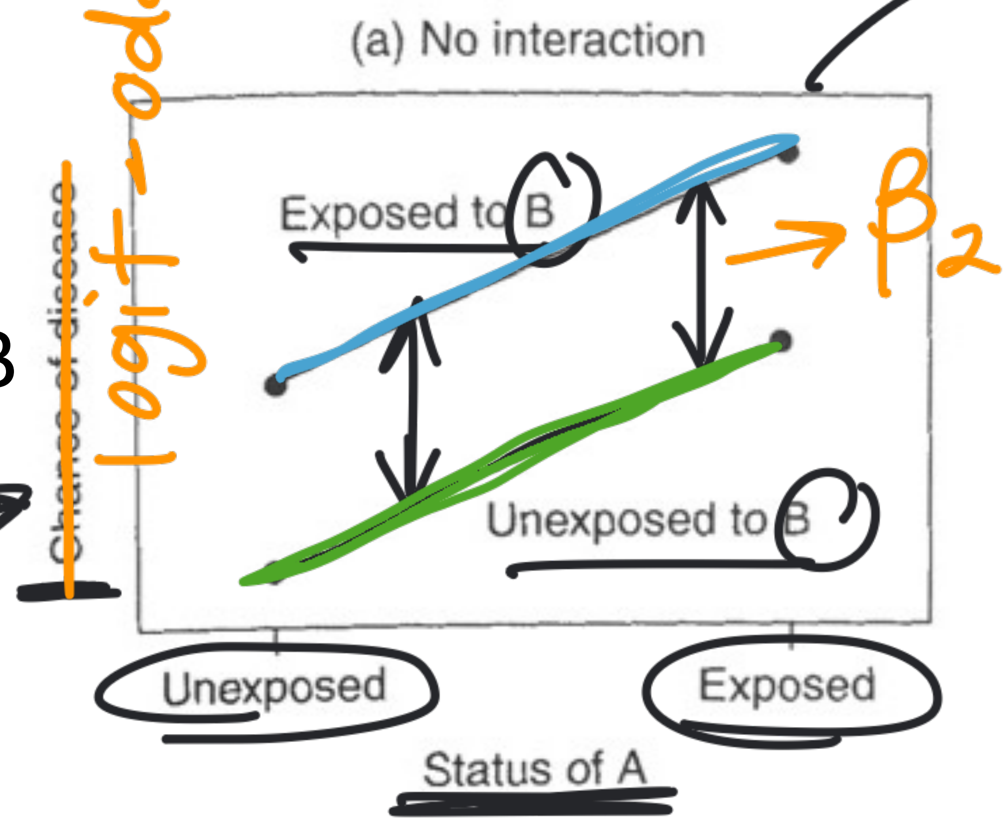
Types of Interaction (II)

$$\mathrm{logit}(\pi(x)) = \beta_0 + \beta_1 A + \beta_2 B$$

$= \beta_0 + \beta_1 A$

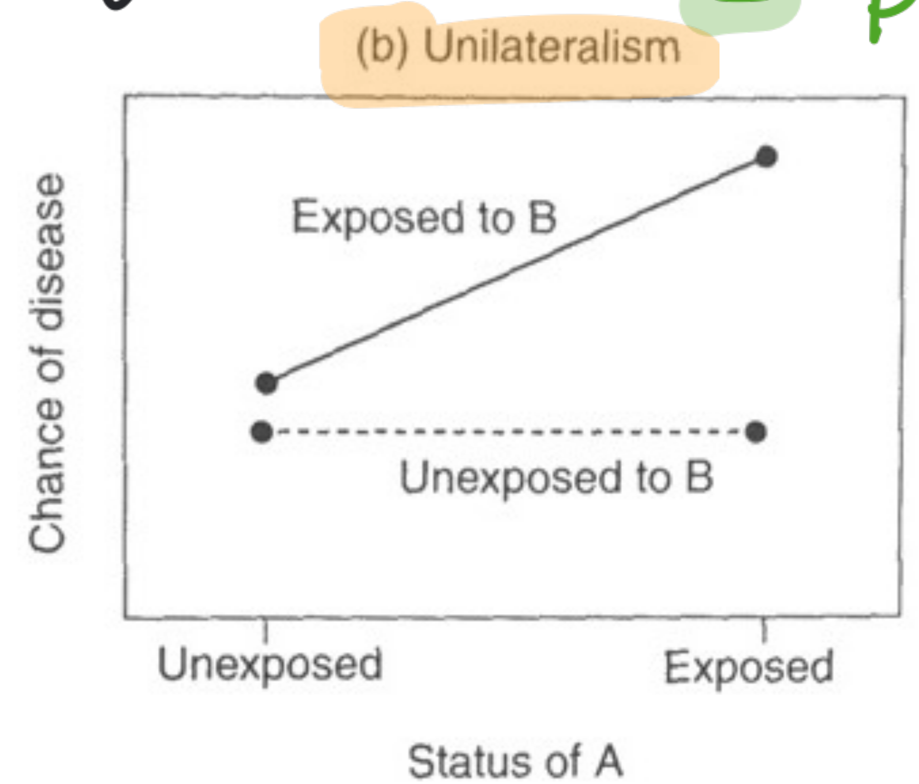
$= \beta_0 + \beta_1 A + \beta_2$

Confounder, but no interaction between A and B



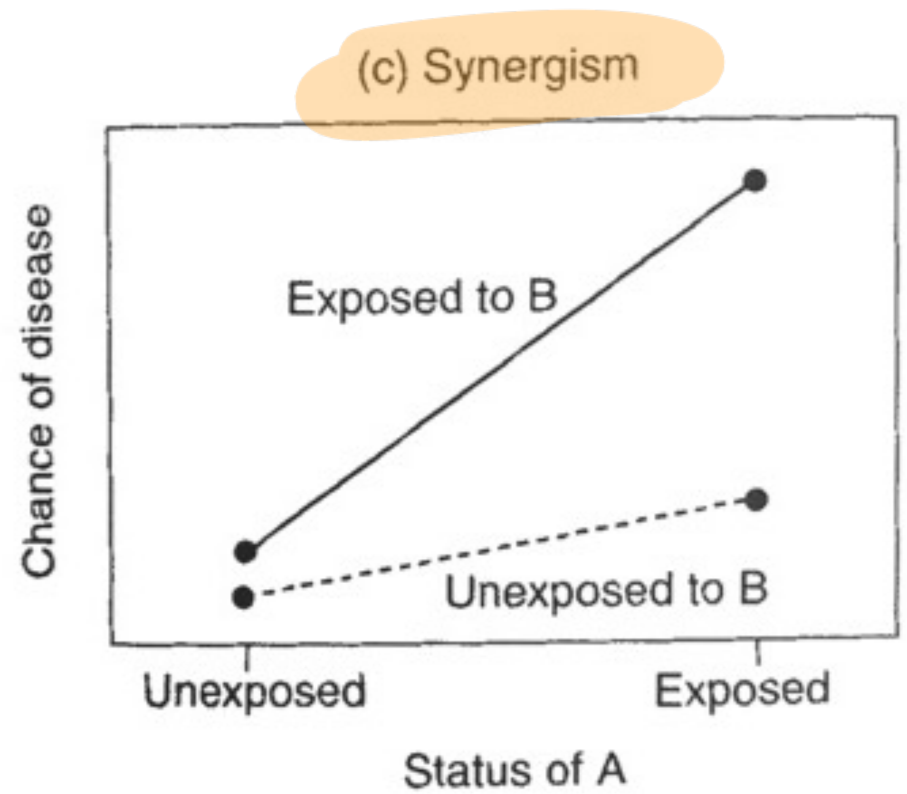
B=0

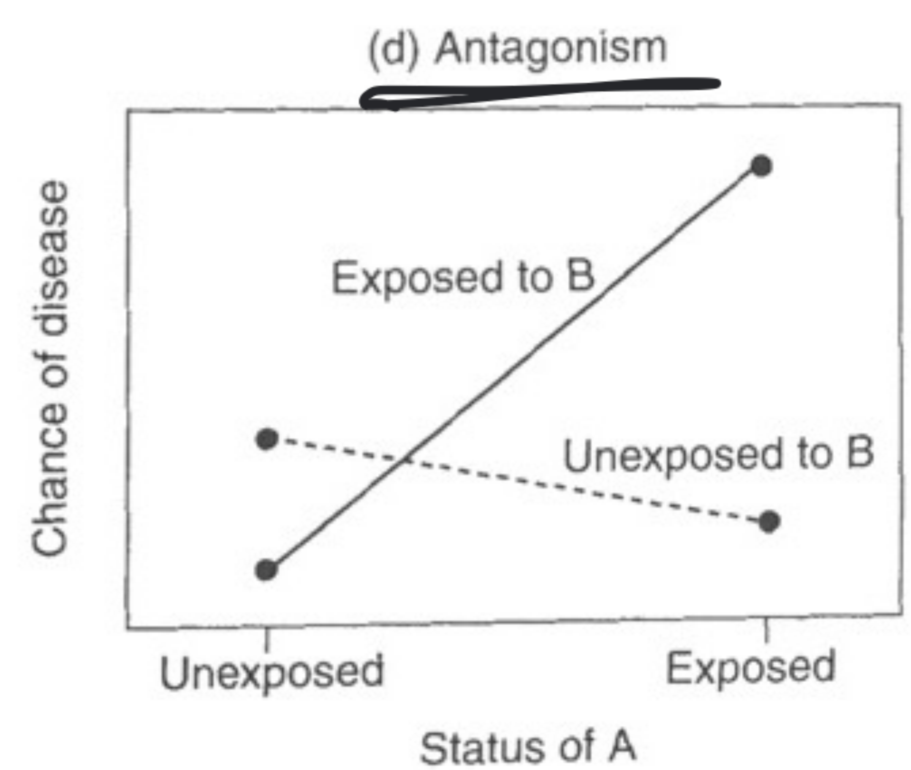
B=1



Exposure to A has no effect in the absence of exposure to B, but a considerable effect when B is present

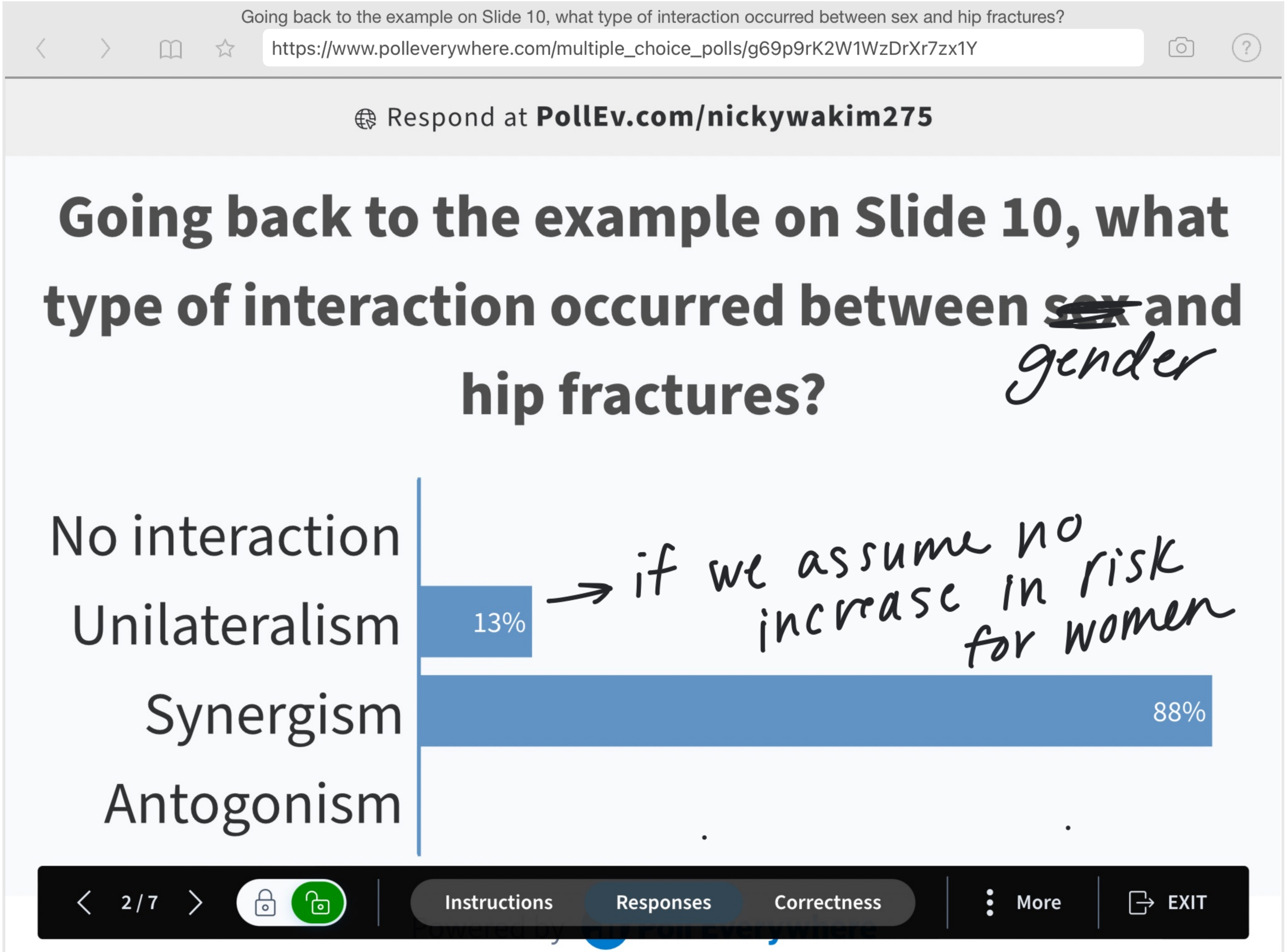
Effect of A is in the same direction, but stronger in the presence of B





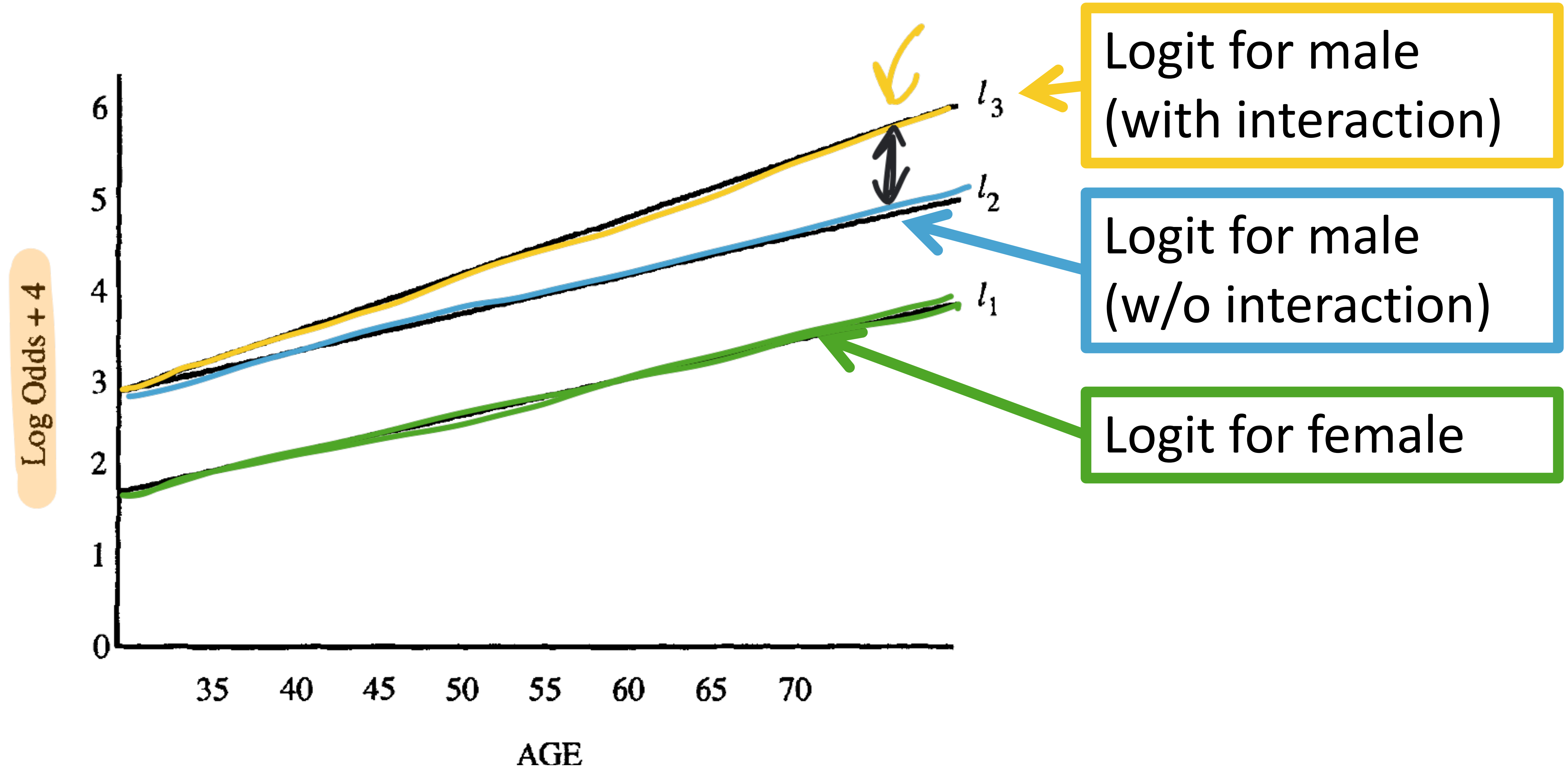
Effect of A works in the opposite direction in the presence of B

Poll Everywhere Question 2



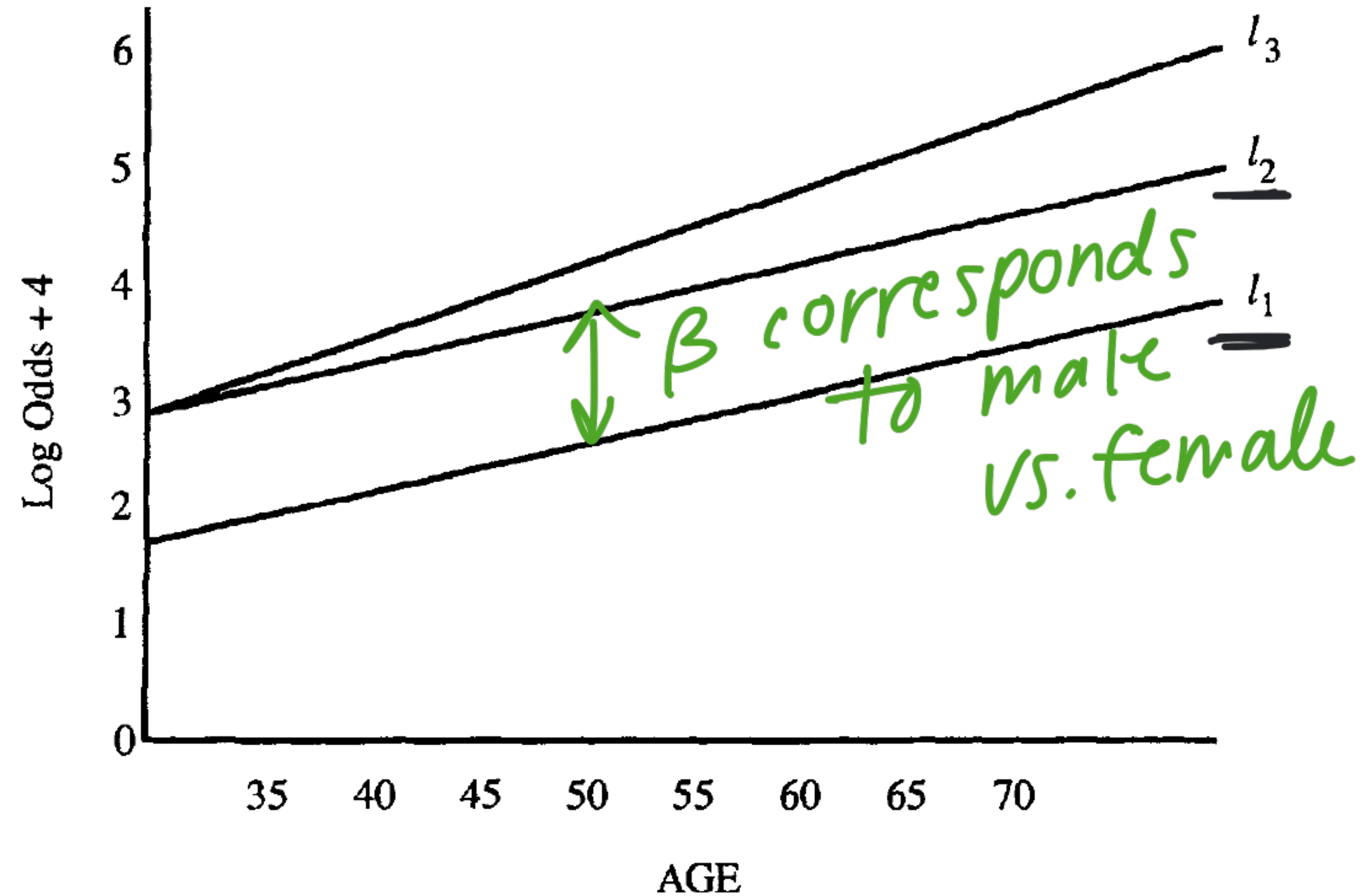
Understand the Interaction (I)

- Figure plots the logits under three different models showing the presence and absence of interaction.
- Response variable: CHD
- Risk factor: sex
- Covariate to be controlled: age



Understand the Interaction (II)

- **If age does not interact with sex**, the distance between l_2 and l_1 is the log odds ratio for sex, controlling for age
  - $(l_2 - l_1)$ stays the same for all values of age.
- **If age interacts with sex,** the distance between l_3 and l_1 is the log odds ratio for sex, controlling for age.
  - Age values need to be specified because $(l_3 - l_1)$ differs for different values of age.
  - Must specify age when reporting odds ratio comparing sex



Let's say we fit the interaction model: $$logit(\pi(sex_i, age_i))=\beta_0+\beta_1sex_i +\beta_2age_i +\beta_3sex_iage_i$$. From our Wald test of $$\beta_3$$ the...

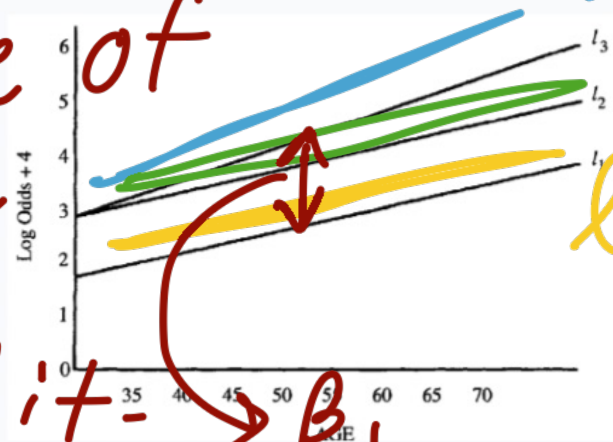
https://www.polleverywhere.com/multiple_choice_polls/Oo9MDkw9gClCaFmBG07Bo

Respond at **PollEv.com/nickywakim275**

Pc

Let's say we fit the interaction model: $logit(\pi(sex_i, age_i)) = \beta_0 + \beta_1 sex_i + \beta_2 age_i + \beta_3 sex_i age_i$. From our Wald test of β_3 the p-value was 0.4. What does that mean for the logits in the plot?

β_2 = is when sex = 0 = female
↳ slope of age



$l_3 \rightarrow$

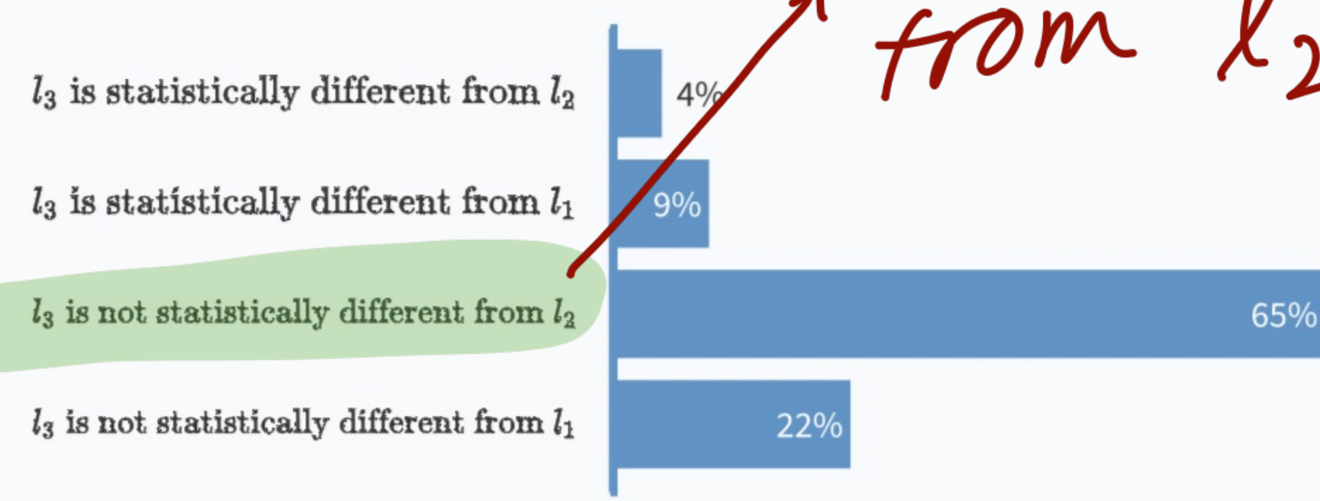
$l_2 \rightarrow \beta_2$

$l_1 \rightarrow \beta_2$

β_3 = additional slope of age for males

β_1

no additional slope from l_2

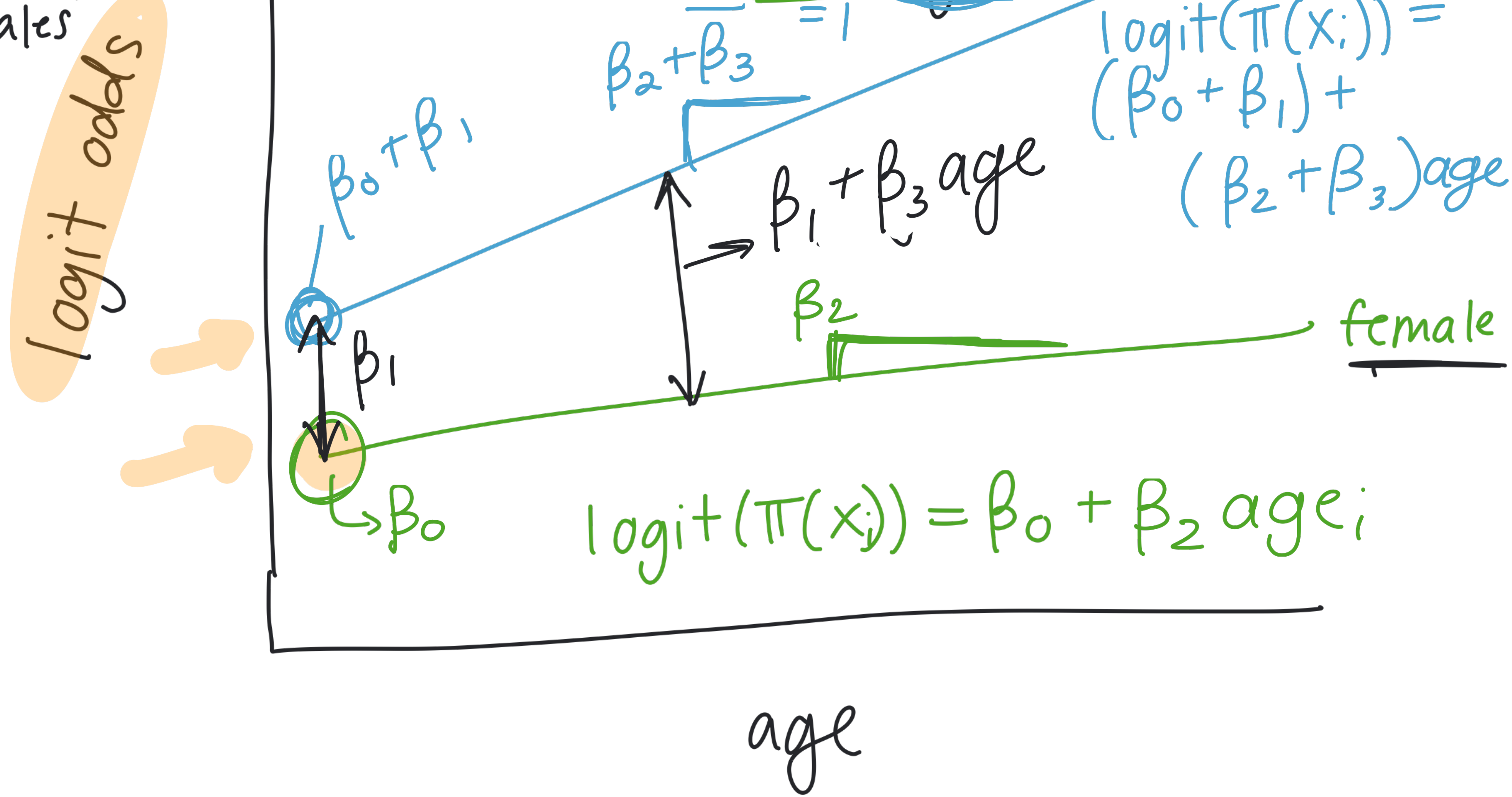


Powered by Poll Everywhere

sex = 0 for females

sex = 1 for males

$$\text{logit}(\pi(x)) = \beta_0 + \beta_1 \underbrace{\text{sex}_i}_{=1} + \beta_2 \text{age}_i + \beta_3 \underbrace{\text{sex}_i}_{=1} \text{age}_i$$



Understand the Interaction (III)

- In the real world, it is rare to see two completely parallel logit plots as we see l_2 and l_1
  - But we need to determine if the difference between l_2 and l_3 is important in the model

- We may not want to include the interaction term unless it is statistically significant and/or clinically meaningful

→ we can use set of coefficient

- Likelihood ratio test (or Wald test) may be used to test the significance of coefficients for variables of the interaction term

Summary

- In a logistic model with two covariates : x_1 (the risk factor, a binary variable) and x_2 (potential confounder/effect modifier)
- The role of x_2 can be one of the three possibilities:
  1. **Not a confounder or effect modifier**, and not significantly associated with Y
     - *No need to include x_2 in the model* (for your dataset)
     - May still be nice to include if other literature in the field includes it
  2. It is a **confounder but not an effect modifier**. There is statistical adjustment but no statistical interaction
     - Should *include x_2 in the model as main effect*
  3. It is an **effect modifier**. There is statistical interaction.
     - Should *include x_2 in the model as main effect and interaction term*

Example: GLOW Study (I)

- From GLOW (Global Longitudinal Study of Osteoporosis in Women) study
  - **Outcome variable:** any fracture in the first year of follow up (FRACTURE: 0 or 1)
  - **Risk factor/variable of interest:** history of prior fracture (PRIORFRAC: 0 or 1)
  - **Potential confounder or effect modifier**: age (AGE, a continuous variable)

Example: GLOW Study (II)

- Back in BSTA 512/612, we could visual the data to get a sense of the interaction before fitting a model

- With a binary outcome, this is a little harder
  - We could use a contingency table or plot proportions of the outcome
  - Hard to do this when our potential confounder or effect modifier is continuous

Example: GLOW Study (III)

Instead, we jump right into model fitting (connecting to the three possible roles of Age):

- **Model 1:** Age not included

$$\longrightarrow \operatorname{logit}\big(\pi(\boldsymbol{x})\big) = \beta_0 + \underline{\theta_1 \mathrm{PF}}$$

- **Model 2:** Age as main effect (age as potential confounder)

$$\operatorname{logit}\big(\pi(\boldsymbol{x})\big) = \beta_0 + \eta_1 \mathrm{PF} + \underline{\beta_2 \mathrm{Age}}$$

- **Model 3:** Age and Prior Fracture interaction (age as potential effect modifier)

$$\operatorname{logit}\big(\pi(\boldsymbol{x})\big) = \beta_0 + \beta_1 \mathrm{PF} + \underline{\beta_2 \mathrm{Age}} + \underline{\beta_3 \mathrm{PF} \times \mathrm{Age}}$$

Example: GLOW Study (III)

Instead, we jump right into model fitting (connecting to the three possible roles of Age):

- **Model 1:** Age not included

```
glow_m1 = glm(fracture ~ priorfrac, data = glow, family = binomial)
```

- **Model 2:** Age as main effect (age as potential confounder)

```
glow_m2 = glm(fracture ~ priorfrac + age, data = glow, family = binomial)
```

- **Model 3:** Age and Prior Fracture interaction (age as potential effect modifier)

```
glow_m3 = glm(fracture ~ priorfrac + age + priorfrac*age, data = glow, family = bin
omial)
```

Example: GLOW Model Coefficient Estimates

Model		Variable	Coeff.	Std. Err.	z	p	95% CI
1	$\hat{\theta}_1$	PRIORFRAC	1.064	0.2231	4.77	<0.001	0.627, 1.501
	$\hat{\beta}_0$	Constant	−1.417	0.1305	−10.86	<0.001	−1.672, −1.161
2	$\hat{\eta}_1$	PRIORFRAC	0.839	0.2342	3.58	<0.001	0.380, 1.298
	$\hat{\beta}_2$	AGE	0.041	0.0122	3.38	0.001	0.017, 0.065
	$\hat{\beta}_0$	Constant	−4.214	0.8478	−4.97	<0.001	−5.876, −2.553
3	$\hat{\beta}_1$	PRIORFRAC	4.961	1.8102	2.74	0.006	1.413, 8.509
	$\hat{\beta}_2$	AGE	0.063	0.0155	4.04	<0.001	0.032, 0.093
	$\hat{\beta}_3$	PRIORFRAC × AGE	−0.057	0.0250	−2.29	0.022	−0.106, −0.008
	$\hat{\beta}_0$	Constant	−5.689	1.0841	−5.25	<0.001	−7.814, −3.565

Table from Hosmer & Lemeshow, page 72

Example: GLOW – Age **confounder**? (I)

- Is age a confounder?
  - Test change in prior fracture coefficient estimate between Model 1 and Model 2

Model		Variable	Coeff.	Std. Err.	z	p	95% CI	
1	$\hat{\theta}_1$	PRIORFRAC	1.064	0.2231	4.77	<0.001	0.627,	1.501
	$\hat{\beta}_0$	Constant	−1.417	0.1305	−10.86	<0.001	−1.672,	−1.161
2	$\hat{\eta}_1$	PRIORFRAC	0.839	0.2342	3.58	<0.001	0.380,	1.298
	$\hat{\beta}_2$	AGE	0.041	0.0122	3.38	0.001	0.017,	0.065
	$\hat{\beta}_0$	Constant	−4.214	0.8478	−4.97	<0.001	−5.876,	−2.553
3	$\hat{\beta}_1$	PRIORFRAC	4.961	1.8102	2.74	0.006	1.413,	8.509
	$\hat{\beta}_2$	AGE	0.063	0.0155	4.04	<0.001	0.032,	0.093
	$\hat{\beta}_3$	PRIORFRAC × AGE	−0.057	0.0250	−2.29	0.022	−0.106,	−0.008
	$\hat{\beta}_0$	Constant	−5.689	1.0841	−5.25	<0.001	−7.814,	−3.565

$$\underline{\Delta\hat{\beta}\%} = 100\% \times \frac{\hat{\theta}_1 - \hat{\eta}_1}{\hat{\eta}_1}$$

$$= 100\% \times \frac{1.064 - 0.839}{0.839}$$

$$= 26.8\% > 20\%$$

Table from Hosmer & Lemeshow, page 72

Example: GLOW – Age **confounder**? (II)

- Is age a confounder?
  - Test change in prior fracture coefficient estimate between Model 1 and Model 2

Model		Variable	Coeff.	Std. Err.	z	p	95% CI	
1	$\hat{\theta}_1$	PRIORFRAC	1.064	0.2231	4.77	<0.001	0.627,	1.501
	$\hat{\beta}_0$	Constant	−1.417	0.1305	−10.86	<0.001	−1.672,	−1.161
2	$\hat{\eta}_1$	PRIORFRAC	0.839	0.2342	3.58	<0.001	0.380,	1.298
	$\hat{\beta}_2$	AGE	0.041	0.0122	3.38	0.001	0.017,	0.065
	$\hat{\beta}_0$	Constant	−4.214	0.8478	−4.97	<0.001	−5.876,	−2.553
3	$\hat{\beta}_1$	PRIORFRAC	4.961	1.8102	2.74	0.006	1.413,	8.509
	$\hat{\beta}_2$	AGE	0.063	0.0155	4.04	<0.001	0.032,	0.093
	$\hat{\beta}_3$	PRIORFRAC × AGE	−0.057	0.0250	−2.29	0.022	−0.106,	−0.008
	$\hat{\beta}_0$	Constant	−5.689	1.0841	−5.25	<0.001	−7.814,	−3.565

Coefficient for prior fracture from the univariable model **overestimates the effect by 26.8%**. Hence, at this point, we could **conclude that age is a confounder** and including age in the model provides an important statistical adjustment to the effect of prior fracture.

Table from Hosmer & Lemeshow, page 72

Example: GLOW – Age **effect modifier**? (I)

- Is age an effect modifier?
  - Test the significance of the interaction term in Model 3
  - We use the Wald test

Model		Variable	Coeff.	Std. Err.	z	p	95% CI
1	$\hat{\theta}_1$	PRIORFRAC	1.064	0.2231	4.77	<0.001	0.627, 1.501
	$\hat{\beta}_0$	Constant	−1.417	0.1305	−10.86	<0.001	−1.672, −1.161
2	$\hat{\eta}_1$	PRIORFRAC	0.839	0.2342	3.58	<0.001	0.380, 1.298
	$\hat{\beta}_2$	AGE	0.041	0.0122	3.38	0.001	0.017, 0.065
	$\hat{\beta}_0$	Constant	−4.214	0.8478	−4.97	<0.001	−5.876, −2.553
3	$\hat{\beta}_1$	PRIORFRAC	4.961	1.8102	2.74	0.006	1.413, 8.509
	$\hat{\beta}_2$	AGE	0.063	0.0155	4.04	<0.001	0.032, 0.093
	$\hat{\beta}_3$	PRIORFRAC × AGE	−0.057	0.0250	−2.29	0.022	−0.106, −0.008
	$\hat{\beta}_0$	Constant	−5.689	1.0841	−5.25	<0.001	−7.814, −3.565

Wald statistic for the interaction coefficient, $\hat{\beta}_3$, is **statistically significant** with $p = 0.022$. Thus, there is **considerable evidence of a statistical interaction** between these age and prior fracture.

Table from Hosmer & Lemeshow, page 72

Poll Everywhere Question 4

Age 65-74 × PF

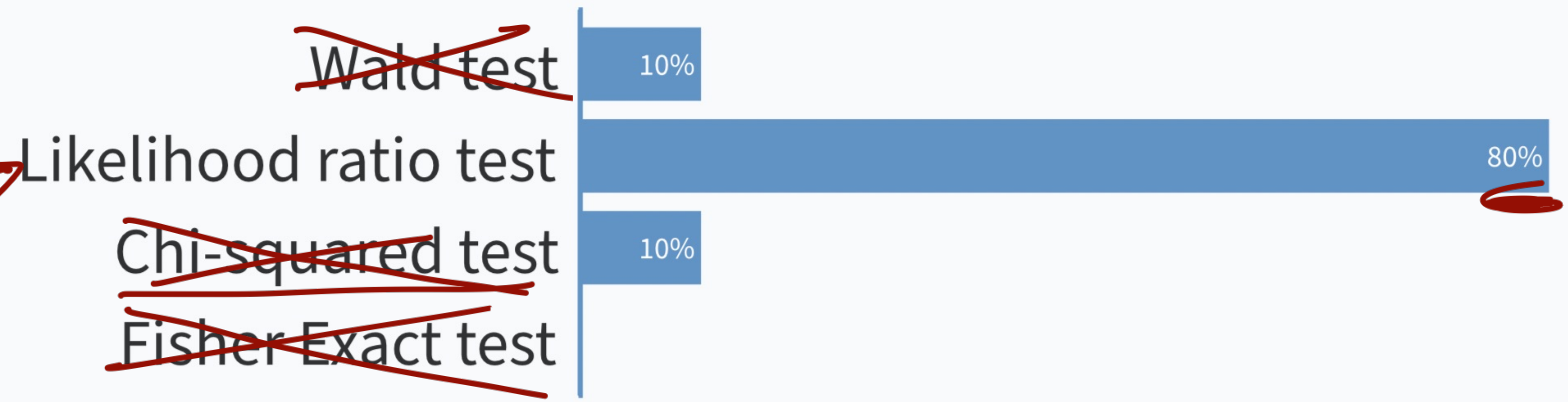
Age 75-90 × PF

If age were split into three categories: 55-64, 65-74, and 75-90, and we were testing the interaction between age and prior fracture, what test would we use?

https://www.polleverywhere.com/multiple_choice_polls/ID0DAGbV7afGKwrhbRhsJ

Respond at **PollEv.com/nickywakim275**

If age were split into three categories: 55-64, 65-74, and 75-90, and we were testing the interaction between age and prior fracture, what test would we use?



Powered by Poll Everywhere

Class 9 Learning Objectives

1. Determine if an additional independent variable is a not a confounder nor effect modifier, is a confounder but not effect modifier, or is an effect modifier.

2. Calculate and interpret odds ratios for interactions

3. Understand the overall steps for purposeful selection as a model building strategy
    - Connect these steps back to the tests of coefficients in Class 8

Example: GLOW study (IV)

- Age is an effect modifier (and possibly also a confounder)

- When a covariate is an effect modifier, its status as a confounder is of secondary importance since the estimate of the effect of the risk factor depends on the specific value of the covariate
  - Must summarize the effect of prior fracture by age
  - Cannot summarize as a single (log) odds ratio

Example: GLOW – Interaction interpretation

Model	Variable	Coeff.	Std. Err.	z	p	95% CI
3	$\hat{\beta}_1$ PRIORFRAC	4.961	1.8102	2.74	0.006	1.413, 8.509
	$\hat{\beta}_2$ AGE	0.063	0.0155	4.04	<0.001	0.032, 0.093
	$\hat{\beta}_3$ PRIORFRAC × AGE	−0.057	0.0250	−2.29	0.022	−0.106, −0.008
	$\hat{\beta}_0$ Constant	−5.689	1.0841	−5.25	<0.001	−7.814, −3.565

- $\hat{\beta}_3 = -0.057$
- The **effect** of prior fracture on the log of odds of having a new fracture is additionally decreases by 0.057 for every one year increase in age
  - Aka the log odds of a new fracture comparing prior fracture to no prior fracture gets closer to one another as age increases

- $\hat{\beta}_1 = 4.961$
- The log odds ratio for a new fracture comparing prior fracture to no prior fracture **when the individual is 0 years old**.

Use the following model to answer: $$\text{logit}(\pi(\bold{x}))=\beta_0+\beta_1 PF+\beta_2 Age+\beta_3 PF*Age$$. What value corresponds to the effect of age...

https://www.polleverywhere.com/multiple_choice_polls/Jn1yKYfHvJ6HPA8NxMYk9

Poll Eve
Que

Respond at **PollEv.com/nickywakim275**

Use the following model to answer: $\text{logit}(\pi(\mathbf{x})) = \beta_0 + \beta_1 PF + \beta_2 Age + \beta_3 PF * Age$**. What value corresponds to the effect of age on the log odds of a new fracture when an individual did not have a prior fracture?**

$\text{logit}(\pi(x)) = \beta_0 + \beta_2 Age$

$PF = 0$ or 1?

β_2	21%
$\beta_0 + \beta_2$	50%
$\beta_2 + \cancel{\beta_3}$	21%
$\beta_0 + \beta_2 + \cancel{\beta_3}$	7%

$PF = 0$ & $age = 0$

Powered by **Poll Everywhere**

Graphically Showing Interaction

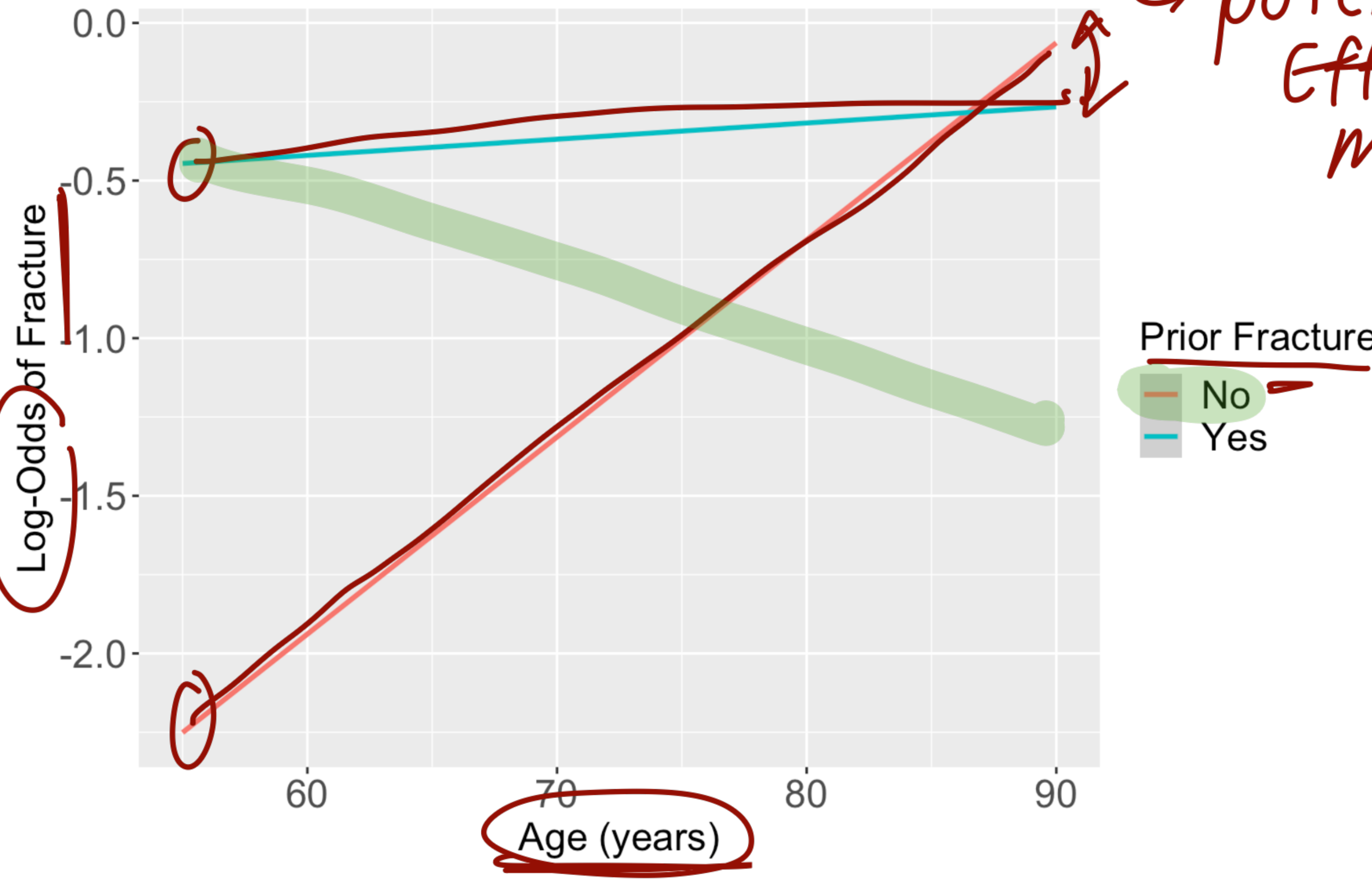
Risk factor

- One easy way to see the nature of the interaction between F and X is to plot the two logit functions

↳ potential Effect mod.

- Using the GLOW fracture example:

synergistic interaction



Wrap-up

- 4-minute exit ticket

- **Next class**
  - Midterm review on Monday
    - Important concepts
    - Practice problems
  - Midterm next Wednesday!
  - We will go through a model building example

Class 9 Exit Ticket

https://forms.office.com/r/YJ6wcNFGBK