

Lesson 10: Interactions

Nicky Wakim

2026-04-29

Learning Objectives

1. Connect understanding of confounding and interactions from linear regression to logistic regression.
2. 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.
3. Calculate and interpret fitted interactions, including plotting the log-odds, predicted probability, and odds ratios.

Learning Objectives

1. Connect understanding of confounding and interactions from linear regression to logistic regression.
2. 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.
3. Calculate and interpret fitted interactions, including plotting the log-odds, predicted probability, and odds ratios.

Reference on slides

- I think these slides from Northwestern do a really good job differentiating between confounders, mediators, and effect modifiers
- <https://www.feinberg.northwestern.edu/sites/bcc/docs/2016lectures/impactofotherfactors.pdf>

Revisit from 512/612: What is a confounder?

- 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

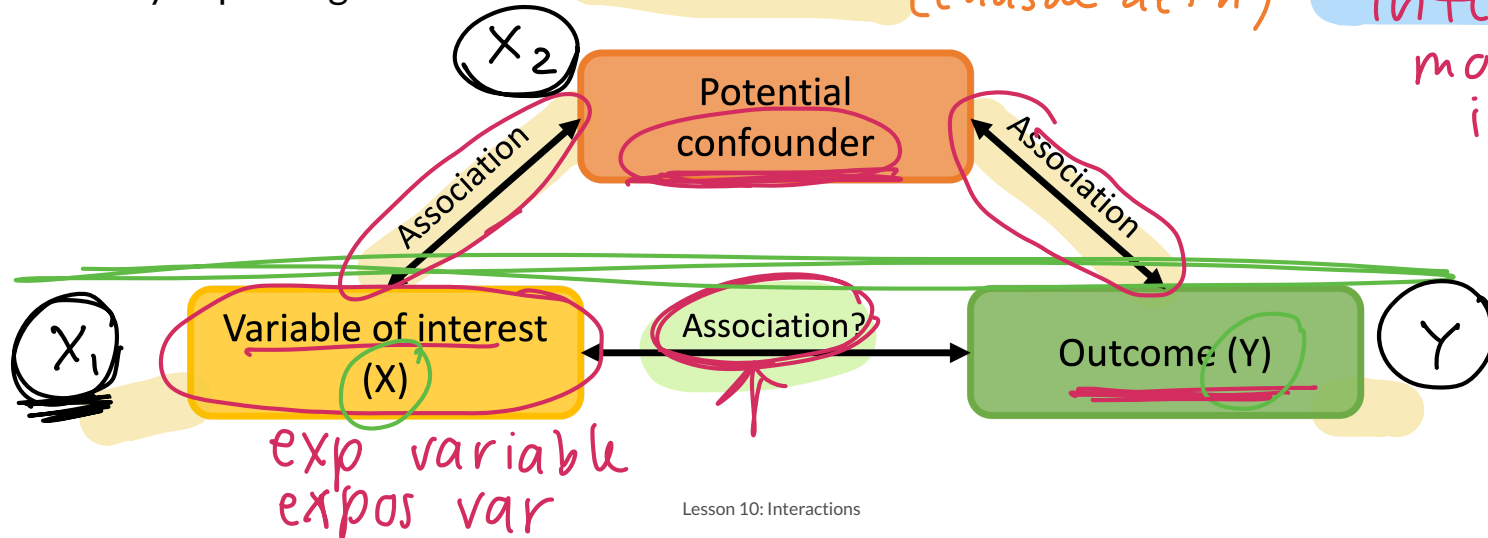
- A confounder must be... **EP/DEF'N W/IN CAUSAL INF**
 - Related to the outcome Y, but not a consequence of Y
 - Related to the explanatory variable X, but not a consequence of X
- In study that does not meet causal assumptions (retrospective, observational, case-control)...

- We need to adjust for **potential** confounders

- We have no way of proving if a variable is a true confounder (causal def'n)

effect
modifiers
need to
model
v/
interaction

interaction
model and
interpretation



Why do we include a covariate in the model?

- We want our model adjusted for other variables that may alter our estimated relationship between our risk factor/explanatory variable and the outcome
- Cases to include variables in the model:
 - Variable is a potential confounder
 - Variable is not balanced in our explanatory variable
 - Example: mean age is very different in treatment and control group
 - Variable is related to the outcome
- *Hint hint:* This is why we did all those plots in Lab 2
 - To get a sense what might be important to include in the model

Including a confounder in the model

- In the following model we have two variables, X_1 and X_2

$$\text{logit}(\pi(\mathbf{X})) = \beta_0 + \beta_1 \mathbf{X}_1 + \beta_2 \mathbf{X}_2$$

- And we assume that every level of the potential confounder, there is parallel slopes
 X_2 adjusting for X_2 of X_1
- Note: to interpret β_1 , we did not specify any value of X_2 ; only specified that it be held constant
adjust
- Implicit assumption: effect of X_1 is equal across all values of X_2

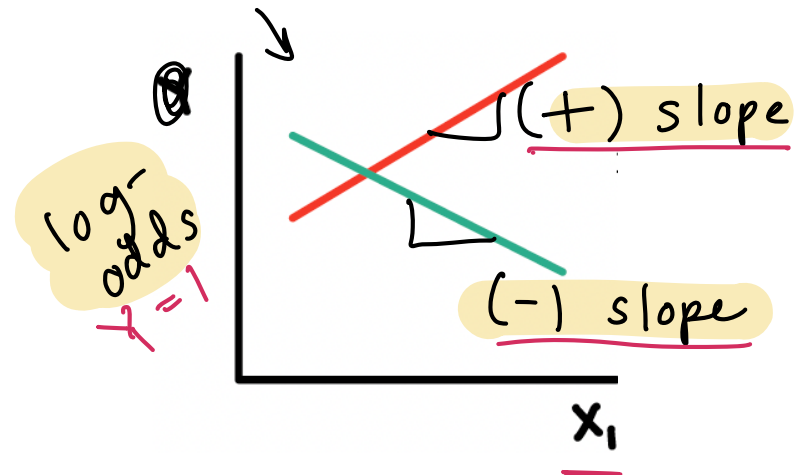
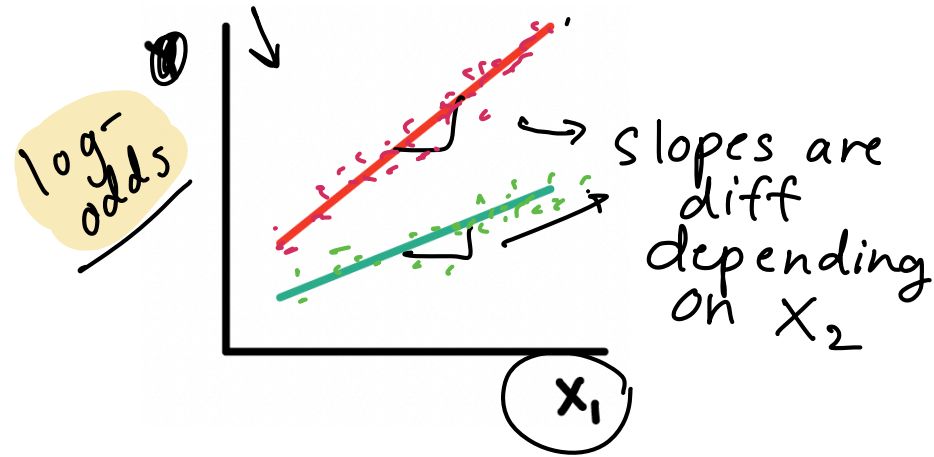
- The above model assumes that X_1 and X_2 do not interact (with respect to their effect on Y)
 - Epidemiology: no “effect modification” $\leftarrow$ if causal
 - Meaning the effect of X_1 is the same regardless of the values of X_2



What is an effect modifier?(interaction)

- An additional variable in the model
 - Outside of the main relationship between Y and X_1 that we are studying
- An effect modifier will change the effect of X_1 on Y depending on its value *depending on X_2*
 - Aka: as the effect modifier's values change, so does the association between Y and X_1
 - So the *the linear combo of coef* coefficient estimating the relationship between Y and X_1 changes with another variable
- We need to anchor our X_1 vs Y relationship in a value of X_2
 - Odds ratio for X_1 must be reported for a specific X_2

$$\log\text{-odds of } Y=1 = \text{logit}(\pi(X))$$



Confounding vs. Interaction

- **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*.
 - Hard to statistically distinguish between confounders and mediators
- **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

Note

Please refer to [Lesson 11 from BSTA 512/612](#) – lots of information about these concepts!

How do we include an effect modifier in the model?

- Interactions!!
- We can incorporate interactions into our model through product terms:

$$\text{logit}(\pi(\mathbf{X})) = \beta_0 + \beta_1 \mathbf{X}_1 + \beta_2 \mathbf{X}_2 + \beta_3 \mathbf{X}_1 \mathbf{X}_2$$

- Terminology:

- main effect parameters: β_1, β_2

main effects
of X_1 & X_2

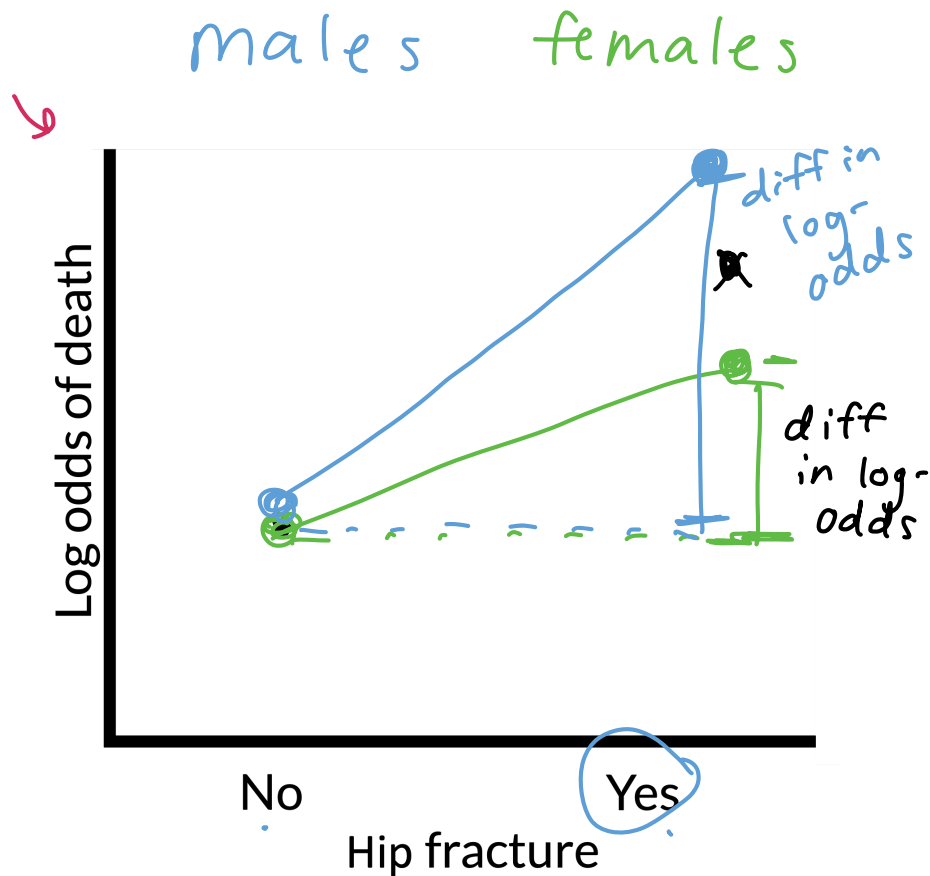
- The main effect models estimate the *average* X_1 and X_2 effects

- interaction parameter: β_3

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 males vs. females
- This is an interaction between hip fracture status (yes/no) and sex (unclear if assigned at birth or no)
- Odds ratio for females < odds ratio for males
 - Odds ratio = odds ratio of death, comparing previous hip fracture to no hip fracture

diff in log-odds male > female $\rightarrow$ log odds ratio male > female $\rightarrow$ odds ratio male > female



$$\log(a) - \log(b) = \log\left(\frac{a}{b}\right)$$

Types of interactions / non-interactions

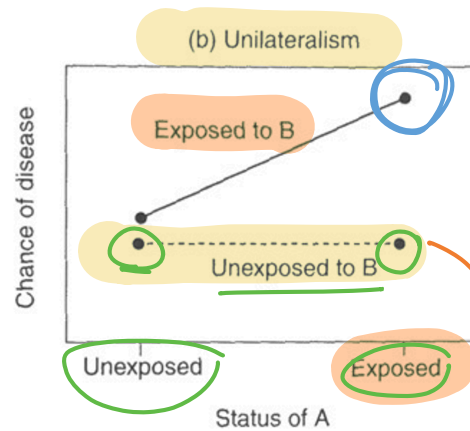
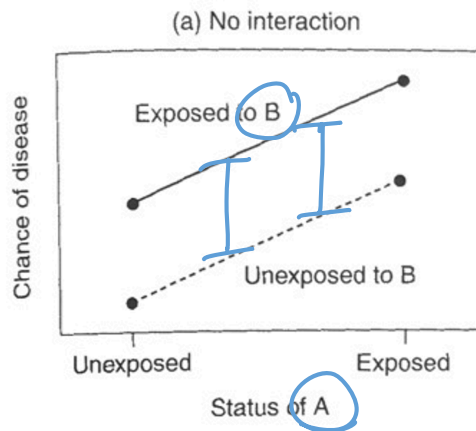
No interaction and three potential effects of interaction between two covariates A and B:

- **No interaction** between A and B (confounder with no interaction)
- **Unilateralism**: exposure to A has no effect in the absence of exposure to B, but a considerable effect when B is present.
- **Synergism**: the effect of A is in the same direction, but stronger in the presence of B.
- **Antagonism**: the effect of A works in the opposite direction in the presence of B.

Types of interactions / non-interactions

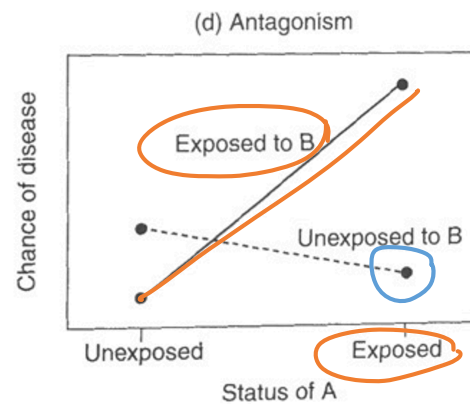
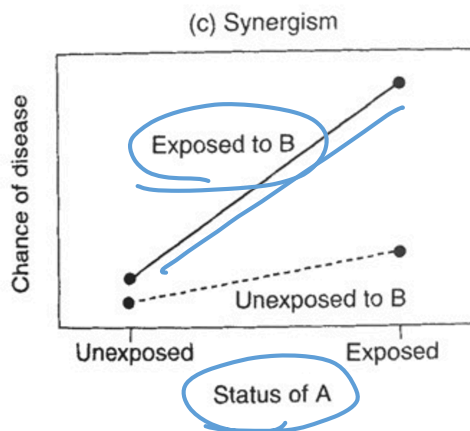
Confounder, but
no interaction
between A and B

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



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

flat when
not exposed to B
(A no association
w/ Y
if no
exp to
B)



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

Poll Everywhere Question 1

13:43 Wed Apr 29

Join by Web PollEv.com/nickywakim275

Going back to the example on hip fractures, what type of sex and hip fractures?

No interaction

Unilateralism

Synergism ☒

Antagonism

83%

6%

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 males vs. females

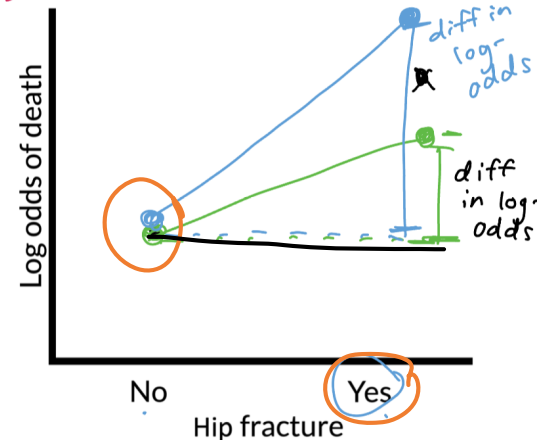
- This is an interaction between hip fracture status (yes/no) and sex (unclear if assigned at birth or no)

- Odds ratio for females ← odds ratio for males

- Odds ratio = odds ratio of death comparing previous hip fracture to no hip fracture

diff in log-odds male > female → log odds ratio male > female → odds ratio male > female

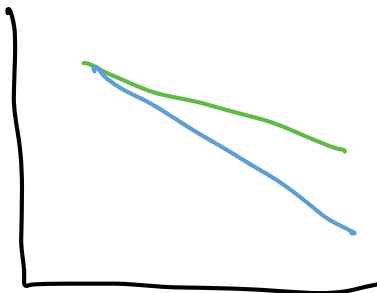
males females



$$\log(a) - \log(b) = \log\left(\frac{a}{b}\right)$$

Lesson 10: Interactions

11



Understand the interaction (1/3)

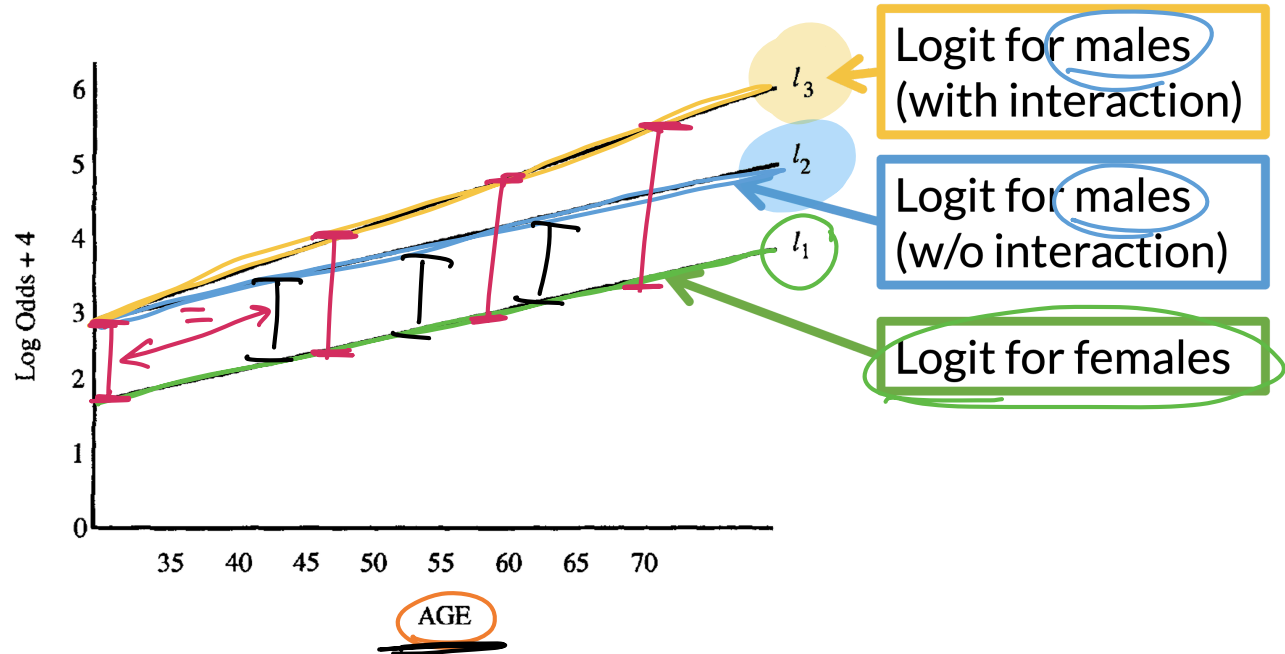
- Figure plots the logits (log-odds) under two different models:

- Model 1: No interaction between sex and age
- Model 2: Interaction between sex and age

- Response variable: CHD

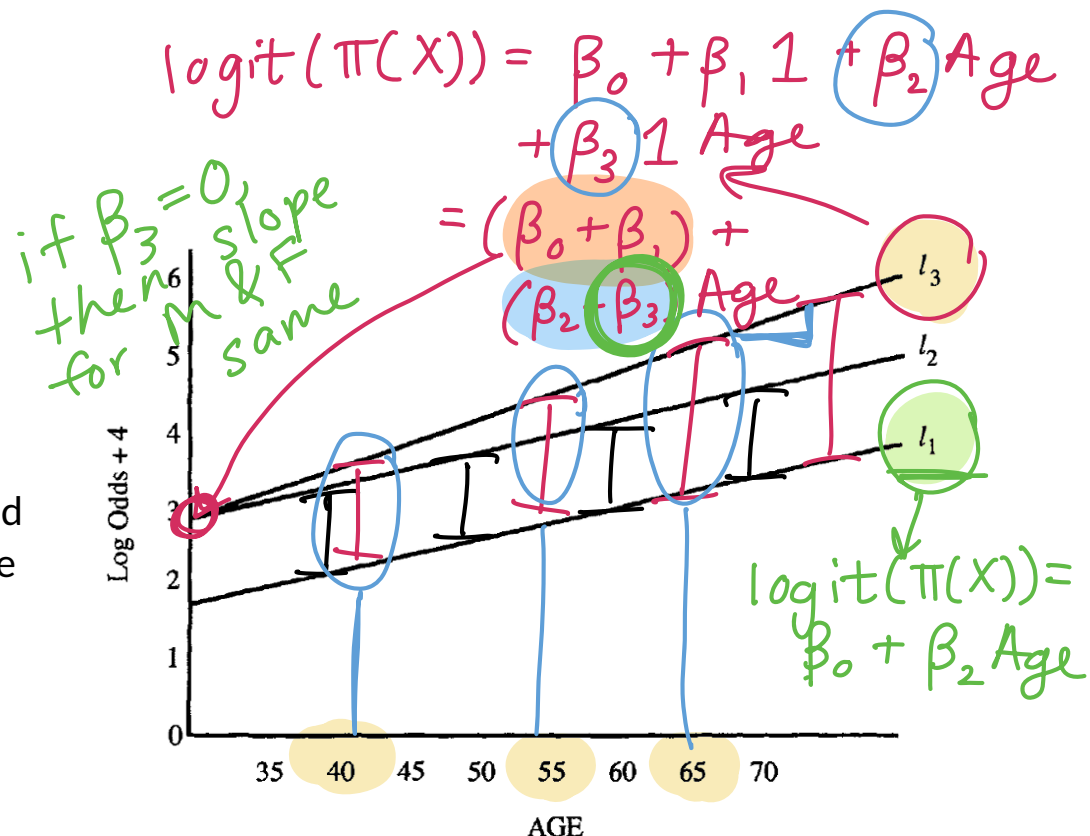
- Risk factor: sex

- Covariate to be controlled: age



Understand the interaction (2/3)

- 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 at specific values of 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



$$\text{logit}(\pi(X)) = \beta_0 + \beta_1 \cdot \mathbb{I}(\text{sex} = M) + \beta_2 \text{Age} + \beta_3 \cdot \mathbb{I}(\text{sex} = M) \cdot \text{Age}$$

Understand the interaction (3/3)

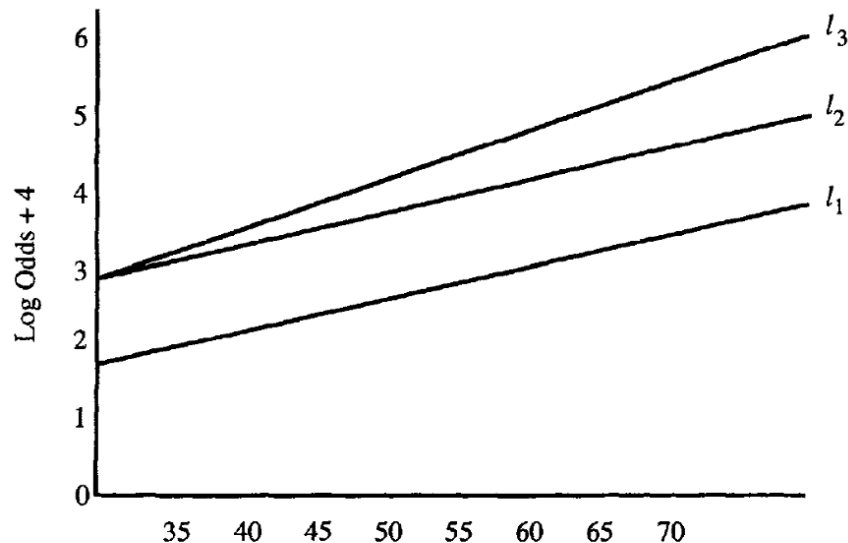
- In the real world, it is rare to see two completely parallel logit plots as we see l_2 and l_1
 - But we need to determine if the difference between l_2 and l_3 is important in the model

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

$$\beta_3 = 0? \text{ or } \beta_3 \neq 0?$$

- Likelihood ratio test (or Wald test sometimes) may be used to test the significance of an interaction term
 - Compare model **with** interaction to model **without** interaction

vs.



$$\begin{aligned} \text{logit}(\pi(x)) &= \beta_0 + \beta_1 I(\text{sex} = M) \\ &\quad + \beta_2 \text{Age} \\ \text{logit}(\pi(x)) &= \beta_0 + \beta_1 I(\text{sex} = M) \\ &\quad + \beta_2 \text{Age} + \beta_3 I(\text{sex} = M) \text{Age} \end{aligned}$$

Poll Everywhere Question 2

14:00 Wed Apr 29

85%



Join by Web [PollEv.com/nickywakim275](https://poll-ev.com/nickywakim275)

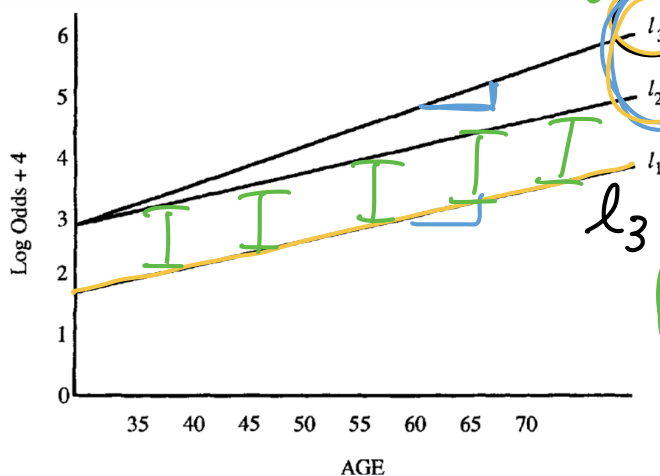


Let's say we fit the interaction model: $\text{logit}(\pi(\text{sex}_i, \text{age}_i)) = \beta_0 + \beta_1 \text{sex}_i + \beta_2 \text{age}_i + \beta_3 \text{sex}_i \text{age}_i$. From our Wald test of β_3 the p-value was

37

0.4. What does that mean for the log-odds in the plot?

$> 0.05 \rightarrow$ fail to reject null that $\beta_3 = 0$ (aka No interaction)



~~l_3 is statistically different from l_2~~

~~l_3 is statistically different from l_1~~

l_3 is not statistically different from l_2

l_3 is not statistically different from l_1

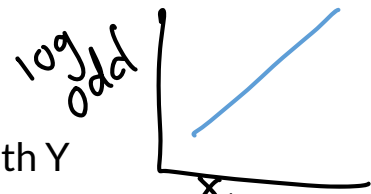
Correct version:
slope of l_3
no stat diff
than slope
of l_1

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 nor 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



$$\text{logit}(\pi(X)) = \beta_0 + \beta_1 X_1$$

2. It is a **confounder but not an effect modifier**. There is statistical adjustment but no statistical interaction

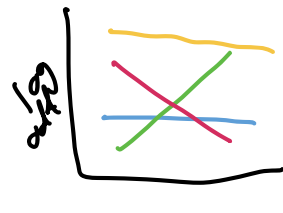
- Should include X_2 in the model as main effect

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

3. It is an **effect modifier**. There is statistical interaction.

- Should include X_2 in the model as main effect and interaction term

(cont X_1 & X_2)



$$\text{logit}(\pi(X)) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$$

Slopes vary & we capture that thru β_3

Learning Objectives

1. Connect understanding of confounding and interactions from linear regression to logistic regression.

2. 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. *interaction for effect mod*

3. Calculate and interpret fitted interactions, including plotting the log-odds, predicted probability, and odds ratios.

Deciding between confounder and effect modifier

- This is more of a model selection question (in coming lectures) $\rightarrow$ purp mod select (BSTA 512)
if checking ONLY 1 int $\rightarrow$ added conf then checked interactions
- But if we had a model with only TWO covariates, we could step through the following process:

1. Test the interaction (of potential effect modifier): use a partial F-test to test if interaction term(s) explain enough variation compared to model without interaction LRT

- Recall that for two continuous covariates, we will test a single coefficient
 - For a binary and continuous covariate, we will test a single coefficient
 - For two binary categorical covariates, we will test a single coefficient
 - For a multi-level categorical covariate (with any other type of covariate), we must test a group of coefficients!! $\rightarrow$ Need to Use LRT
- } Wald or LRT

2. Then look at the main effect (or potential confounder)

- If interaction already included, then automatically included as main effect (and thus not checked for confounding)
- For variables that are not included in any interactions:
 - Check to see if they are confounders by seeing whether exclusion of the variable changes any of the main effect of the primary explanatory variable by more than 10%

$\rightarrow$ include as confounder

Step 1: Testing the interaction

- We test with $\alpha = 0.10$
- Follow the LRT procedure in Lesson 6, slide 38
- Use the hypothesis tests for the specific covariate combo:

Binary & continuous variable

Testing a single coefficient for the interaction term using LRT comparing full model (with interaction) to reduced model (without interaction)

Multi-level categorical & continuous variables

Testing a group of coefficients for the interaction term using LRT comparing full model (with interaction) to reduced model (without interaction)

Binary & multi-level variable

Testing a group of coefficients for the interaction term using LRT comparing full model (with interaction) to reduced model (without interaction)

Two continuous variables

Testing a single coefficient for the interaction term using LRT comparing full model (with interaction) to reduced model (without interaction)

Poll Everywhere Question 2

13:25 Mon May 4

97%

Join by Web 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?

Wald test

8%

Likelihood ratio test ✓

88%

Chi-squared test

4%

Fisher Exact test

0%

REF $I(\text{age} = 65-74)$

$I(\text{age} = 75-90)$
 $I(PF=y)$

	55-64	65-74	75-90
Y			
N			

green + blue

is interact

$$\begin{matrix} \# \text{ cat} & b/c & 1 \\ & \text{ref} & \text{grp} \\ \downarrow & & \\ (3-1) & (2-1) \\ \uparrow & \uparrow \\ \# \text{ cat} & \text{ref} \\ & \text{grp} \end{matrix}$$

red model

$$\text{logit}(\pi(x)) =$$

$$\beta_0 + \beta_1 I(PF=y)$$

$$+ \beta_2 I(\text{age} = 65-74)$$

$$+ \beta_3 I(\text{age} = 75-90)$$

$$+ \beta_4 I(PF=y) \cdot$$

$$I(\text{age} = 65-74)$$

$$+ \beta_5 I(PF=y) I(\text{age} = 75-90)$$

Step 2: Testing a confounder

- If interaction already included: $\text{logit}(\pi(X)) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$
 - Meaning: LRT showed evidence for alternative/full model (*interaction model*)
 - Then the variable is an effect modifier and we don't need to consider it as a confounder
 - Then automatically included as main effect (and thus not checked for confounding)
- For variables that are not included in any interactions:
 - Check to see if they are confounders
 - One way to do this is by seeing whether exclusion of the variable changes any of the main effect of the primary explanatory variable by more than 10%
→ how much does X_1 's ^{main} effect change w/ & w/out X_2 ?
- If the main effect of the primary explanatory variable changes by less than 10%, then the additional variable is neither an effect modifier nor a confounder
 - We leave the variable out of the model

$$\text{logit}(\pi(X)) = \beta_0 + \beta_1 X_1$$

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

Testing for percent change ($\Delta\%$) in a coefficient

- Let's say we have X_1 and X_2 , and we specifically want to see if X_2 is a confounder for X_1 (the explanatory variable or variable of interest)
- If we are only considering X_1 and X_2 , then we need to run the following two models:

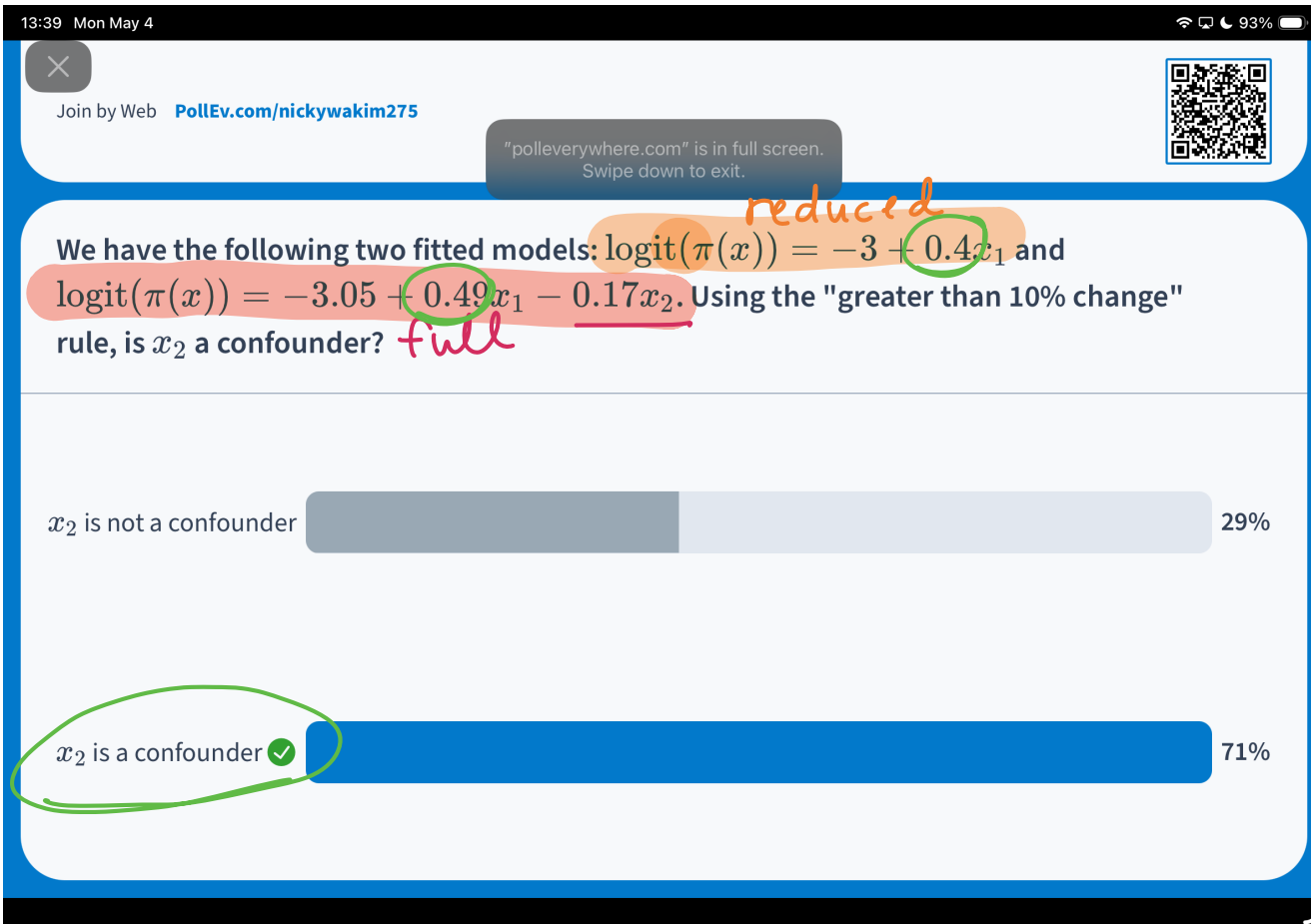
- ■ **Fitted model 1 / reduced model (mod1):** $\text{logit}(\hat{\pi}(\mathbf{X})) = \hat{\beta}_0 + \hat{\beta}_1 X_1$
- We call the above $\hat{\beta}_1$ the reduced model coefficient: $\hat{\beta}_{1,\text{mod1}}$ or $\hat{\beta}_{1,\text{red}}$
- **Fitted model 2 / full model (mod2):** $\text{logit}(\hat{\pi}(\mathbf{X})) = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2$
- We call this $\hat{\beta}_1$ the full model coefficient: $\hat{\beta}_{1,\text{mod2}}$ or $\hat{\beta}_{1,\text{full}}$

Calculation for % change in coefficient

$$\Delta\% = 100\% \cdot \frac{\hat{\beta}_{1,\text{mod1}} - \hat{\beta}_{1,\text{mod2}}}{\hat{\beta}_{1,\text{mod2}}} = 100\% \cdot \frac{\hat{\beta}_{1,\text{red}} - \hat{\beta}_{1,\text{full}}}{\hat{\beta}_{1,\text{full}}}$$

w/out X_2 , how much does β_1 change?

Poll Everywhere Question 3



$$\begin{aligned}\% \Delta &= \frac{|\beta_{\text{full}} - \beta_{\text{red}}|}{\beta_{\text{full}}} \times 100\% \\ &= \frac{|0.49 - 0.4|}{0.49} \\ &= \frac{0.09}{0.49} \times 100\% \\ &= 0.1837 \times 100\% \\ &= 18.37\% \\ &> 10\%\end{aligned}$$

Example: GLOW Study

- 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)

No Yes

logistic reg
for binary outcome

ASIDE ★ count of fractures: poisson reg (zero inf?)

- Risk factor/variable of interest: history of prior fracture (PRIORFRAC: 0 or 1)

No Yes

- Potential confounder or effect modifier: age (AGE, a continuous variable)

- Center age will be used! We will center around the rounded mean age of 69 years old

```
1 library(aplcore3)
2 mean_age = mean(glow500$age) %>% round()
3 glow = glow500 %>% mutate(age_c = age - mean_age)
```

$$\hookrightarrow \text{Age}_c = \text{Age} - \overline{\text{Age}}_{(\text{mean age})}$$

Example: GLOW Study: Try to visual the sample proportions

- Back in BSTA 512/612, we could visual the data to get a sense if there was an 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: Calculate the proportions

```
1 glow2 = glow %>%  
2   group_by(age, priorfrac, fracture) %>% # last one needs to be outcome  
3   summarise(n = n()) %>%  
4   mutate(freq = n / sum(n)) %>% # takes the proportion of yes/no  
5   filter(fracture == "Yes") # Filtering so only "success" shown  
6   #filter(freq != 1 | n != 1)  
7  
8 head(glow2)
```

A tibble: 6 × 5

Groups: age, priorfrac [6]

	age	priorfrac	fracture	n	freq
	<int>	<fct>	<fct>	<int>	<dbl>
1	56	No	Yes	3	0.15
2	56	Yes	Yes	1	1
3	57	No	Yes	3	0.158
4	57	Yes	Yes	1	0.333
5	58	No	Yes	3	0.15
6	59	Yes	Yes	2	0.667

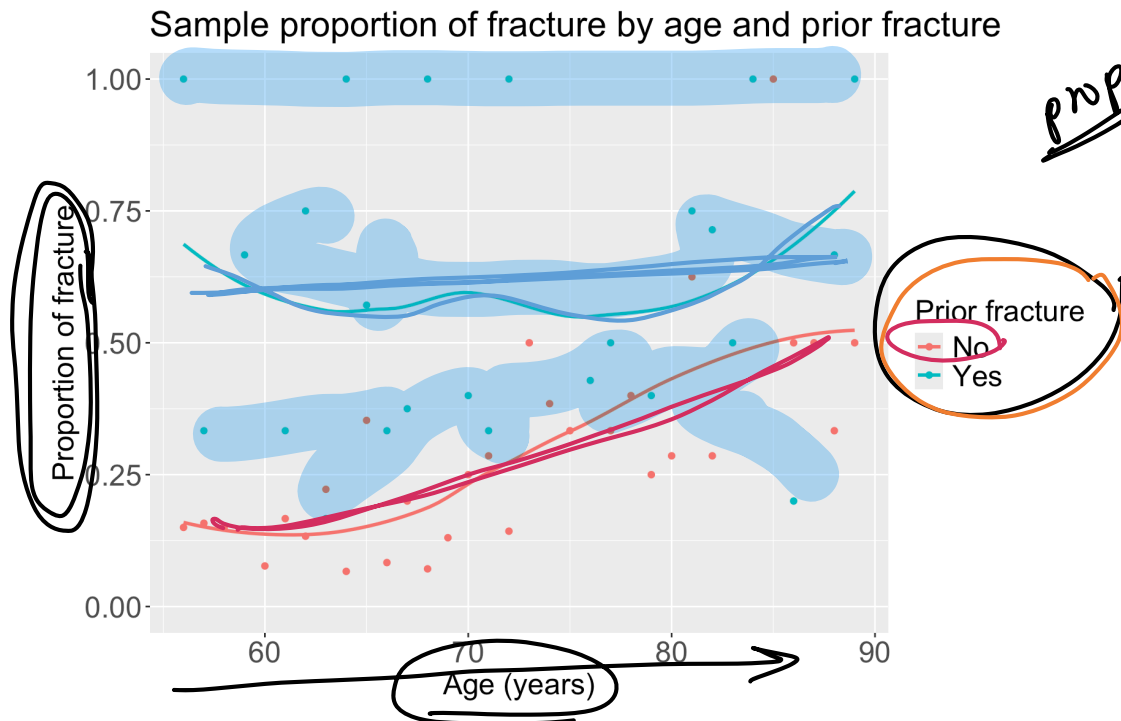
prop of fractures
for each age
& prior fract
combo

Example: GLOW Study: Plot the proportions

```
1 ggplot(data = glow2, aes(y = freq, x = age, color = priorfrac)) +  
2   geom_point() + ylim(0, 1) + geom_smooth(se = F) +  
3   labs(x = "Age (years)", y = "Proportion of fracture",  
4     color = "Prior fracture", title = "Sample proportion of fracture by age and prior fracture") +  
5   theme(axis.title = element_text(size = 18), axis.text = element_text(size = 18),  
6     title = element_text(size = 18), legend.text=element_text(size=18))
```

confounder

Not
log odds



- From sample proportions, looks like age and prior fracture may have an interaction!

Example: GLOW Study

We also could *jump right into model fitting* (connecting to the three possible roles of Age):

- **Model 1:** Age not included

$$\text{logit}(\pi(\mathbf{X})) = \beta_0 + \beta_1 \cdot I(\text{PF})$$

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

$$\text{logit}(\pi(\mathbf{X})) = \beta_0 + \beta_1 \cdot I(\text{PF}) + \beta_2 \cdot \text{Age}$$

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

$$\text{logit}(\pi(\mathbf{X})) = \beta_0 + \beta_1 \cdot I(\text{PF}) + \beta_2 \cdot \text{Age} + \beta_3 \cdot I(\text{PF}) \cdot \text{Age}$$

Example: GLOW Study

We also could *jump right into model fitting* (connecting to the three possible roles of Age):

- **Model 1:** Age not included

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

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

```
1 glow_m2 = glm(fracture ~ priorfrac + age_c,  
2               data = glow, family = binomial)
```

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

```
1 glow_m3 = glm(fracture ~ priorfrac + age_c + priorfrac*age_c,  
2               data = glow, family = binomial)
```

Example: GLOW Study: Age an effect modifier or confounder?

- **Model 1:** Age not included

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	-1.417	0.130	-10.859	0.000	-1.679	-1.167
priorfracYes	1.064	0.223	4.769	0.000	0.626	1.502

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

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	-1.372	0.132	-10.407	0.000	-1.637	-1.120
priorfracYes	0.839	0.234	3.582	0.000	0.378	1.297
age_c	0.041	0.012	3.382	0.001	0.017	0.065

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

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	-1.376	0.134	-10.270	0.000	-1.646	-1.120
priorfracYes	1.002	0.240	4.184	0.000	0.530	1.471
age_c	0.063	0.015	4.043	0.000	0.032	0.093
→ priorfracYes:age_c	-0.057	0.025	-2.294	0.022	-0.107	-0.008

Example: GLOW Study: Age an effect modifier or confounder?

- **Model 1:** Age not included

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	-1.417	0.130	-10.859	0.000	-1.679	-1.167
priorfracYes	1.064	0.223	4.769	0.000	0.626	1.502

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

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	-1.372	0.132	-10.407	0.000	-1.637	-1.120
priorfracYes	0.839	0.234	3.582	0.000	0.378	1.297
age_c	0.041	0.012	3.382	0.001	0.017	0.065

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

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	-1.376	0.134	-10.270	0.000	-1.646	-1.120
priorfracYes	1.002	0.240	4.184	0.000	0.530	1.471
age_c	0.063	0.015	4.043	0.000	0.032	0.093
priorfracYes:age_c	-0.057	0.025	-2.294	0.022	-0.107	-0.008

①

- Is age an effect modifier?

- Test the significance of the interaction term in Model 3
- We can use LRT (or Wald test bc only one coefficient for interaction)

b/c binary PF & cont age

- If not an effect modifier, check the change in coefficient for prior fracture between Model 1 and Model 2

Example: GLOW Study: Age an effect modifier or confounder?

- **Model 1:** Age not included

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	-1.417	0.130	-10.859	0.000	-1.679	-1.167
priorfracYes	1.064	0.223	4.769	0.000	0.626	1.502

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

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	-1.372	0.132	-10.407	0.000	-1.637	-1.120
priorfracYes	0.839	0.234	3.582	0.000	0.378	1.297
age_c	0.041	0.012	3.382	0.001	0.017	0.065

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

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	-1.376	0.134	-10.270	0.000	-1.646	-1.120
priorfracYes	1.002	0.240	4.184	0.000	0.530	1.471
age_c	0.063	0.015	4.043	0.000	0.032	0.093
priorfracYes:age_c	-0.057	0.025	-2.294	0.022	-0.107	-0.008

β_3

β_3

SE_{β_3}

W

p-val

Lesson 8

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

< 0.05

Poll Everywhere Question 4

Use the following model to answer: $\text{logit}(\pi(x)) = \beta_0 + \beta_1 PF + \beta_2 Age_c + \beta_3 PF * Age_c$. 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?

$$\text{logit}(\pi(x)) = \beta_0 + \beta_1 I(PF=Y) + \beta_2 Age_c + \beta_3 I(PF=Y) Age_c$$

$$\hookrightarrow I(PF=Y) = 0$$

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

β_2 ✓

8%

for every 1 yr inc in age, log odds inc by β_2 (when No PF).

$\beta_0 + \beta_2$

56%

log odds @ mean age (69yo)

$\beta_2 + \beta_3$

20%

$$\text{logit}(\pi(x)) = \beta_0 + \beta_1 + \beta_2 Age_c + \beta_3 Age_c$$

if PF: $I(PF=Y) = 1$

$\beta_0 + \beta_2 + \beta_3$

16%

Please please please reference your work from BSTA 512/612

- We had lessons and homeworks dedicated to this process!
- The process will be the same!
 - Only differences are t-test and F-test are replaced by Wald test and Likelihood ratio test, respectively!!

Learning Objectives

1. Connect understanding of confounding and interactions from linear regression to logistic regression.
2. 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.
3. Calculate and interpret fitted interactions, including plotting the log-odds, predicted probability, and odds ratios.

Example: GLOW Study

- Age is an **effect modifier** of prior fracture
- 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 on current fracture *by age*
 - Cannot summarize as a single (log) odds ratio

Example: GLOW – Interaction interpretation

- Model 3:

$$\begin{aligned}\text{logit}(\hat{\pi}(\mathbf{X})) &= \hat{\beta}_0 & + \hat{\beta}_1 \cdot I(\text{PF}) & + \hat{\beta}_2 \cdot \text{Age} & + \hat{\beta}_3 \cdot I(\text{PF}) \cdot \text{Age} \\ \text{logit}(\hat{\pi}(\mathbf{X})) &= -1.376 & + 1.002 \cdot I(\text{PF}) & + 0.063 \cdot \text{Age} & - 0.057 \cdot I(\text{PF}) \cdot \text{Age}\end{aligned}$$

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	-1.376	0.134	-10.270	0.000	-1.646	-1.120
priorfracYes	1.002	0.240	4.184	0.000	0.530	1.471
age_c	0.063	0.015	4.043	0.000	0.032	0.093
priorfracYes:age_c	-0.057	0.025	-2.294	0.022	-0.107	-0.008

Example: GLOW – Interaction interpretation

$$\text{logit}(\pi(X)) = \beta_0 + \beta_1 I(PF=1) + \beta_2 \text{Age}_c + \beta_3 I(PF) \text{Age}_c$$

• Model 3 estimated coefficients:

term	estimate	p.value	conf.low	conf
(Intercept)	-1.376	0.000	-1.646	-1
priorfracYes	1.002	0.000	0.530	1
age_c	0.063	0.000	0.032	0
priorfracYes:age_c	-0.057	0.022	-0.107	-0

• Model 3 estimated odds ratios:

term	estimate	p.value	conf.low	conf
(Intercept)	0.25	0.00	0.19	
priorfracYes	2.72	0.00	1.70	
age_c	1.06	0.00	1.03	
priorfracYes:age_c	0.94	0.02	0.90	

- $\hat{\beta}_3 = -0.057$

- The effect of having a prior fracture on the log odds of having a new fracture decreases by an estimated 0.057 for every one year increase in age (95% CI: 0.008, 0.107).

- Aka the log odds of a new fracture comparing prior fracture to no prior fracture gets closer to one another as age increases

- $\hat{\beta}_1 = 1.002$

$\rightarrow \text{Age}_c = 0$

- For individuals 69 years old, the estimated difference in log odds for a new fracture is 1.002 comparing individuals with a prior fracture to individuals with no prior fracture (95% CI: 0.530, 1.471).

- $\exp(\hat{\beta}_1) = 2.72$

- For individuals 69 years old, the estimated odds of a new fracture for individuals with prior fracture is 2.72 times the estimated odds of a new fracture for individuals with no prior fracture (95% CI: 1.70, 4.35).

Poll Everywhere Question 5

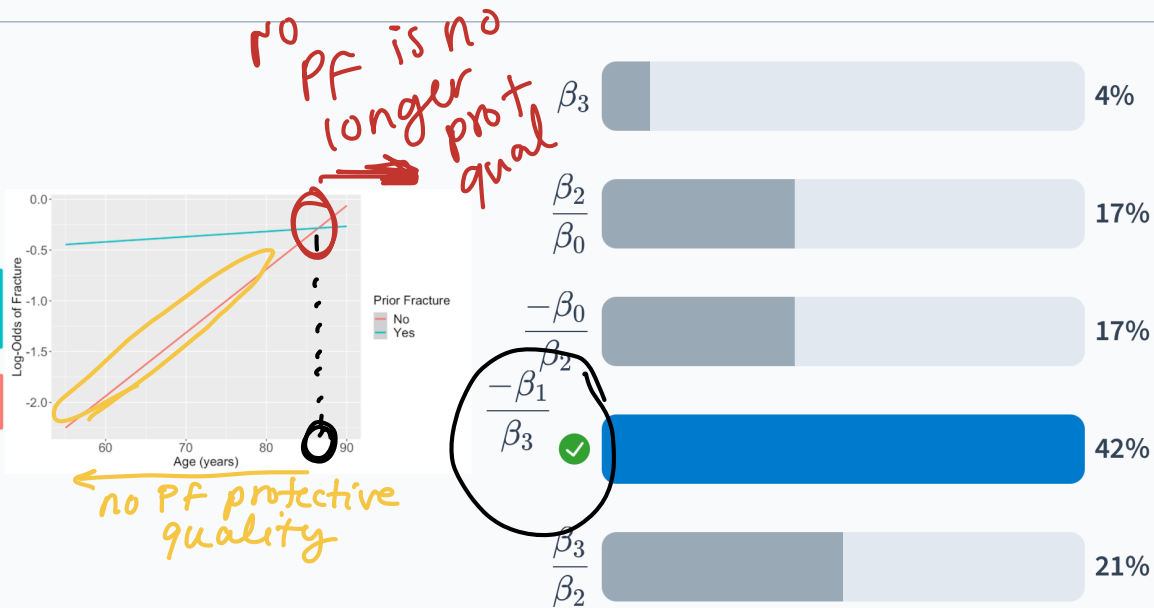
14:32 Mon May 4

76%

Join by Web [PollEv.com/nickywakim275](https://poller.com/nickywakim275)



Using the picture or previous slide, calculate the age at which the log-odds of a new fracture is equal regardless of prior fractures.



$$\beta_0 + \beta_1 + (\beta_2 + \beta_3) \text{ age} - \beta_0$$

=

$$\beta_0 + \beta_2 \text{ age} - \beta_0$$

$$\beta_1 + \beta_2 \text{ age} + \beta_3 \text{ age} = \beta_2 \text{ age}$$

$$\beta_1 + \beta_3 \text{ age} = 0$$

$$\frac{\beta_3 \text{ age}}{\beta_3} = \frac{-\beta_1}{\beta_3}$$

$$\text{age} = -\beta_1 / \beta_3$$

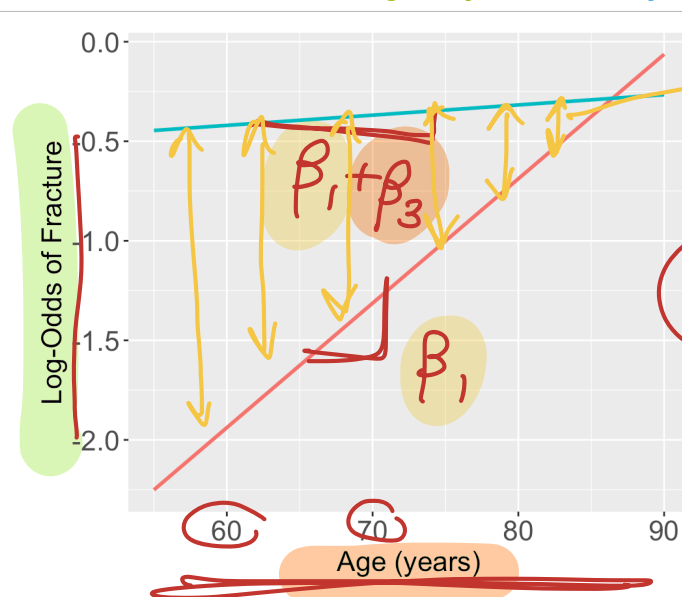
Plot of estimated log odds

```

1 prior_age = expand_grid(priorfrac = c("No", "Yes"), age_c = (55:90)-69)
2 frac_pred_log = predict(glow_m3, prior_age, se.fit = T, type="link")
3 pred_glow2 = prior_age %>% mutate(frac_pred_log = frac_pred_log$fit,
4                                   age = age_c + mean_age)
5
6 ggplot(pred_glow2) + #geom_point(aes(x = age, y = frac_pred, color = priorfrac)) +
7   geom_smooth(method = "loess", aes(x = age, y = frac_pred_log, color = priorfrac)) +
8   theme(text = element_text(size=20), title = element_text(size=16)) +
9   labs(color = "Prior Fracture", x = "Age (years)", y = "Log-Odds of Fracture")
  
```

expand grid w/ create df of every combo of & PF

log-odds (logit)



diff of log odd b/w PF & no PF @ diff ages

$\beta_3 = 0$: parallel lines (no interact)

$\beta_3 \neq 0$: Slopes are diff $\Rightarrow$ int

~~Poll Everywhere Question 6 (Bonus q if we're feeling it)~~

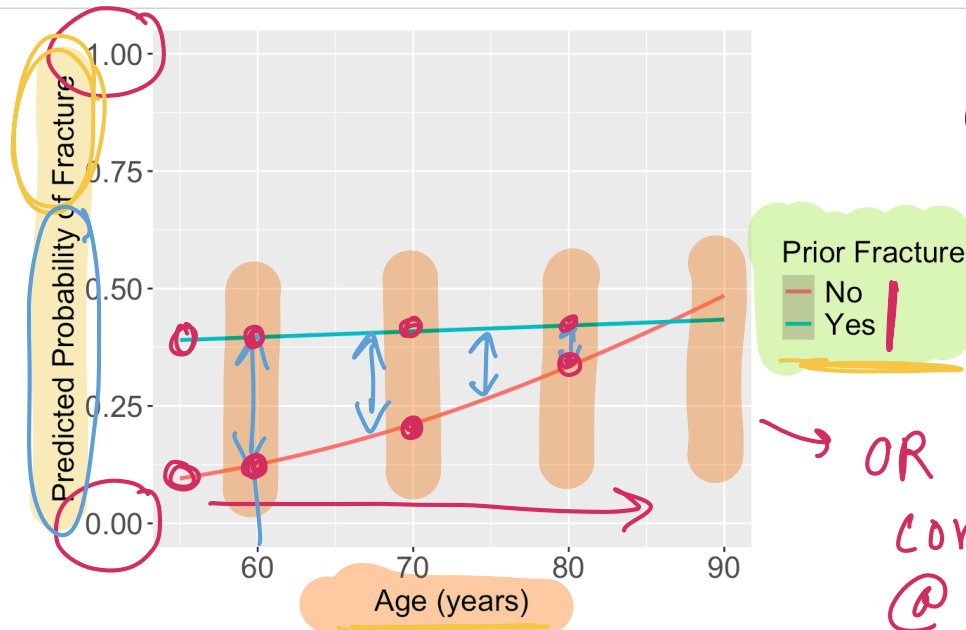
none? huh?

Age	PF	OR	95%CI
60	no PF		
	PF	4.55	
70	no PF		
	PF	2.57	
80	no PF		
	PF	1.1 ??	

Calc @ home

Plot the predicted probability of fracture

```
1 frac_pred = predict(glow_m3, prior_age, se.fit = T, type="response")
2 pred_glow = prior_age %>% mutate(frac_pred = frac_pred$fit,
3                                 age = age_c + mean_age)
4
5 ggplot(pred_glow) + #geom_point(aes(x = age, y = frac_pred, color = priorfrac)) +
6   geom_smooth(method = "loess", aes(x = age, y = frac_pred, color = priorfrac)) +
7   theme(text = element_text(size=20), title = element_text(size=16)) + ylim(0,1) +
8   labs(color = "Prior Fracture", x = "Age (years)", y = "Predicted Probability of Fracture")
```



PF becomes
less protective
as age
increases

→ OR of future fracture
comparing PF to no PF
@ a specific age

Odds Ratio in the Presence of Interaction (1/2)

- When interaction exists between a risk factor (F) and another variable (X), the estimate of the odds ratio for F depends on the value of X

Age

PF

Age

PF

- When an interaction term (F*X) exists in the model

- $\widehat{OR}_F \neq \exp(\widehat{\beta}_F)$ in general

$$\exp(\widehat{\beta}_{INT}) \neq OR$$

- Assume we want to compute the odds ratio for $(F = f_1 \text{ and } F = f_0)$, the correct model-based estimate is

$$\widehat{OR}_F = \exp(\widehat{g}(F = f_1, X = x) - \widehat{g}(F = f_0, X = x))$$

- Let's work this out on the next slide!

need to set
specific x val

(in ex: specific
age)

placeholder
for x val

Odds Ratio in the Presence of Interaction (2/2)

- We may write the two logits (log-odds) for given x as below:

$$\begin{aligned}\hat{g}(F = f_1, X = x) &= \hat{\beta}_0 + \hat{\beta}_1 f_1 + \hat{\beta}_2 x + \hat{\beta}_3 f_1 \cdot x \\ \hat{g}(F = f_0, X = x) &= \hat{\beta}_0 + \hat{\beta}_1 f_0 + \hat{\beta}_2 x + \hat{\beta}_3 f_0 \cdot x\end{aligned}$$

PF
no PF

- The difference in two logits (log-odds) is:

$$\hat{g}(f_1, x) - \hat{g}(f_0, x) = [\hat{\beta}_0 + \hat{\beta}_1 f_1 + \hat{\beta}_2 x + \hat{\beta}_3 f_1 \cdot x] - [\hat{\beta}_0 + \hat{\beta}_1 f_0 + \hat{\beta}_2 x + \hat{\beta}_3 f_0 \cdot x]$$

$$\hat{g}(f_1, x) - \hat{g}(f_0, x) = \hat{\beta}_1 f_1 - \hat{\beta}_1 f_0 + \hat{\beta}_3 x f_1 - \hat{\beta}_3 x f_0$$

$$\hat{g}(f_1, x) - \hat{g}(f_0, x) = \hat{\beta}_1 (f_1 - f_0) + \hat{\beta}_3 x (f_1 - f_0)$$

interaction

- Therefore,

$$\widehat{OR}(F = \overset{f_1}{\cancel{f_1}}, F = f_0, X = x) = \widehat{OR}_F = \exp [\hat{\beta}_1 (f_1 - f_0) + \hat{\beta}_3 x (f_1 - f_0)]$$

Our ex

Steps to compute OR under interaction

- Note: You don't need to know the math itself, but I think it's helpful to think of it this way

1. Identify two sets of values that you want to compare with only one variable changed

- In previous slides, one set was $(F = f_1, X = x)$ and the other was $(F = f_0, X = x)$

2. Substitute values in the fitted log-odds model

- You should have two equations, one for each set of values

3. Take the difference of the two log-odds

4. Exponentiate the resulting difference $\rightarrow$ to get OR

Example: GLOW Study

1. Identify two sets of values that you want to compare with only one variable changed

$$\left[\begin{array}{l} \bullet \text{ Set 1: } PF = 1, Age = a \\ \bullet \text{ Set 2: } PF = 0, Age = a \end{array} \right. \rightarrow I(PF = \text{yes}) = 1$$
$$\hookrightarrow I(PF = \text{yes}) = 0$$

2. Substitute values in the fitted log-odds model

$$\text{logit}(\hat{\pi}(\mathbf{X})) = \hat{\beta}_0 + \hat{\beta}_1 \cdot I(PF) + \hat{\beta}_2 \cdot Age + \hat{\beta}_3 \cdot I(PF) \cdot Age$$

$$\begin{aligned} \text{logit}(\hat{\pi}(PF = 1, Age = a)) &= \hat{\beta}_0 + \hat{\beta}_1 \cdot 1 + \hat{\beta}_2 \cdot a + \hat{\beta}_3 \cdot 1 \cdot a \\ &= \hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 \cdot a + \hat{\beta}_3 \cdot a \end{aligned}$$

$$\begin{aligned} \text{logit}(\hat{\pi}(PF = 0, Age = a)) &= \hat{\beta}_0 + \hat{\beta}_1 \cdot 0 + \hat{\beta}_2 \cdot a + \hat{\beta}_3 \cdot 0 \cdot a \\ &= \hat{\beta}_0 + \hat{\beta}_2 \cdot a \end{aligned}$$

*a is
placeholder
for a
value of
age*

Example: GLOW Study

3. Take the difference of the two log-odds

$$\begin{aligned} & [\text{logit}(\pi(PF = 1, \text{Age} = a))] - [\text{logit}(\pi(PF = 0, \text{Age} = a))] \\ &= [\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 a + \hat{\beta}_3 a] - [\hat{\beta}_0 + \hat{\beta}_2 a] \\ &= \hat{\beta}_1 + \hat{\beta}_3 a \end{aligned}$$

4. Exponentiate the resulting difference

comparing

$$\widehat{OR}[(PF = 1, \text{Age} = a), (PF = 0, \text{Age} = a)] = \exp(\hat{\beta}_1 + \hat{\beta}_3 a)$$

OF FRACTURE

set 1

set 2

We can put in values for age to see how the OR changes

- If we let $a = 60$, i.e., compute OR for age = 60, then

$$\widehat{OR}_{a=60} = \exp(1.002 - 0.057 \cdot (60 - 69)) = 4.55$$

- If we let $a = 70$, i.e., compute OR for age = 70, then

$$\widehat{OR}_{a=70} = \exp(1.002 - 0.057 \cdot (70 - 69)) = 2.57$$

Calculate odds ratios across values (`estimable()`)

- We could use `estimable()` to calculate this linear combination of coefficients (see [BSTA 512 Lesson](#))

```
1 library(gmodels)
2 glow_m3 %>% estimable(
3   cbind("(Intercept)"
4         "priorfracYes"
5         "age_c"
6         "priorfracYes:age_c"
7         conf.int = 0.95) %>%
8   exp(.) %>%
9   head(5)
```

	Estimate	Std. Error	X^2 value	DF	Pr(> X^2)	Lower.CI	Upper.CI
(0 1 0 -14)	6.081961	1.613392	1535383	2.718282	1.000161	2.354473	15.71063
(0 1 0 -13)	5.742790	1.578600	2324162	2.718282	1.000129	2.321435	14.20657
(0 1 0 -12)	5.422533	1.545094	3605103	2.718282	1.000102	2.287290	12.85533
(0 1 0 -11)	5.120136	1.512903	5721950	2.718282	1.000080	2.251861	11.64184
(0 1 0 -10)	4.834603	1.482070	9262667	2.718282	1.000062	2.214941	10.55260

match output
of coef table
summary(glow_m3)

```
0, # beta0
1, # beta1
0, # beta2
(55:90)-69, # beta3
```

$$OR = e^{\hat{\beta}_1 + \hat{\beta}_3 a}$$

$$0\hat{\beta}_0 + 1\hat{\beta}_1 + 0\hat{\beta}_2 + (55-69)\hat{\beta}_3$$

55 56 57 58 ... 90
-14 -13 -12 -11 ... -2

Calculate odds ratios across values (`predict()`)

```

1 prior_age = expand_grid(priorfrac = c("No", "Yes"), age_c = (55:90)-69)
2 frac_pred_logit = predict(glow_m3, prior_age, se.fit = T, type="link")
3 pred_glow2 = prior_age %>% mutate(frac_pred = frac_pred_logit$fit,
4                                   age = age_c + mean_age) %>%
5
6     pivot_wider(names_from = priorfrac, values_from = frac_pred) %>%
7
8     mutate(OR_YN = exp(Yes - No))
9 head(pred_glow2)

```

every combo
priorfrac &
age_c

log odds

```

# A tibble: 6 × 5
  age_c   age priorfrac Yes OR_YN
<dbl> <dbl> <dbl> <dbl> <dbl>
1   -14    55      No    Yes  6.08
2   -13    56      No    Yes  5.74
3   -12    57      No    Yes  5.42
4   -11    58      No    Yes  5.12
5   -10    59      No    Yes  4.83
6    -9    60      No    Yes  4.56

```

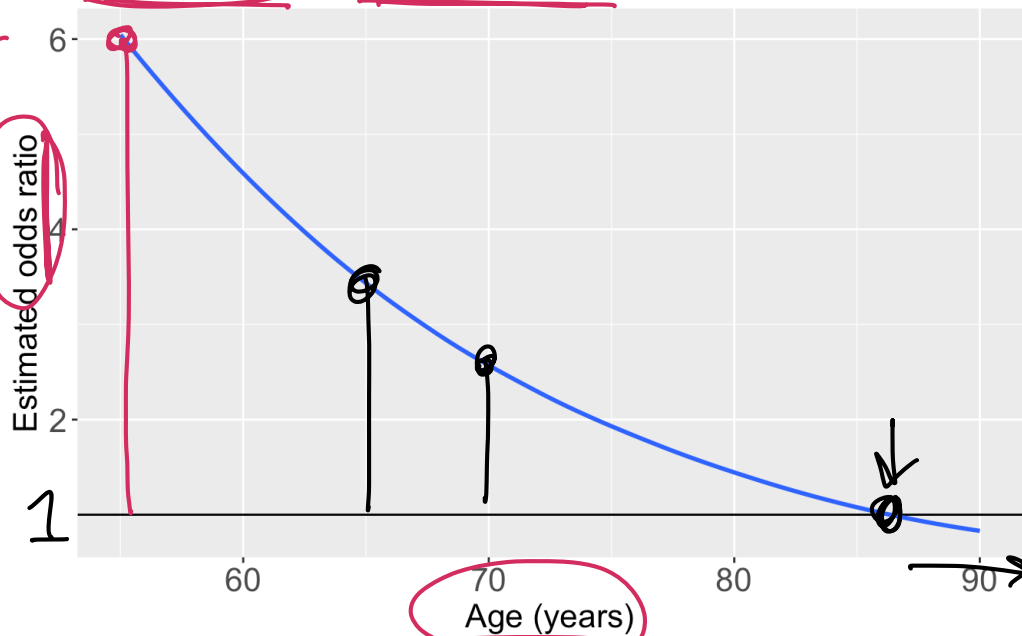
	PF	log odds
-14 55	No	-2.25
-14 55	Yes	-0.446
-13 56	No	-2.19
-13 56	Yes	-0.441

Plotting the odds ratio for an interaction

```
1 ggplot(pred_glow2) +  
2   geom_hline(yintercept = 1) +  
3   geom_smooth(method = "loess", aes(x = age, y = OR_YN)) +  
4   theme(text = element_text(size=20), title = element_text(size=16)) +  
5   labs(x = "Age (years)", y = "Estimated odds ratio", title = "Odds ratio of fracture outcome
```

Odds ratio of fracture outcome comparing
prior fracture to no prior fracture

@ 55 yo
est. odds of
fracture is
6x when
comparing
PF to PF



How would I report these results?

- Remember our main covariate is prior fracture, so we want to focus on how age changes the relationship between prior fracture and a new fracture!

For individuals $age_i = 0$ 69 years old, the estimated odds of a new fracture for individuals with prior fracture is 2.72 times the estimated odds of a new fracture for individuals with no prior fracture (95% CI: 1.70, 4.35). As seen in Figure 1 (a), the estimated odds ratio of a new fracture when comparing prior fracture status decreases with age, indicating that the effect of prior fractures on new fractures decreases as individuals get older. In Figure 1 (b), it is evident that for both prior fracture statuses, the predicted probability of a new fracture increases as age increases. However, the predicted probability of new fracture for those without a prior fracture increases at a higher rate than that of individuals with a prior fracture. Thus, the predicted probabilities of a new fracture converge at age [insert age here].

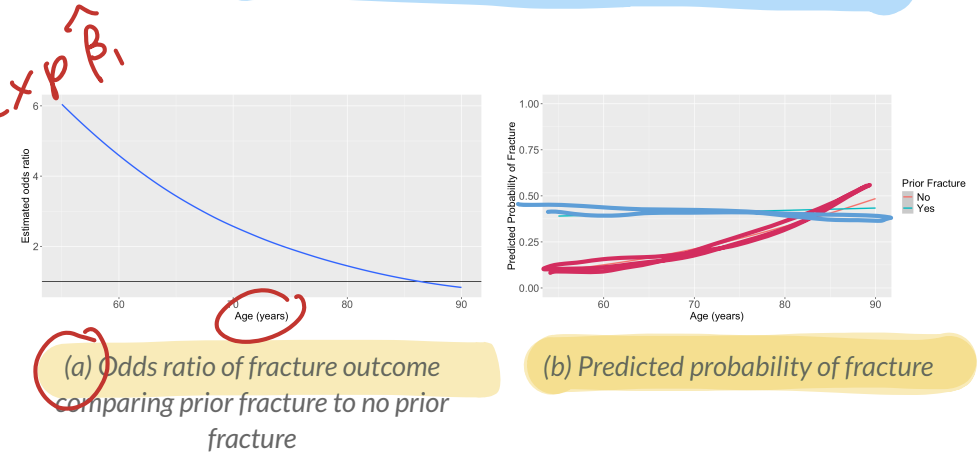


Figure 1: Plots of odds ratio and predicted probability from fitted interaction model

Age	PF	OR of new frac (95% CI)
60	no PF	
	PF	
70	no PF	
	PF	
80	no PF	
	PF	

$$\begin{aligned} \text{logit}(\pi(x)) = & \beta_0 + \beta_1 I(\text{age} > 90) \\ & + \beta_2 \text{Age} \\ & + \beta_3 I(\text{age} > 90) \text{Age} \end{aligned}$$

