Lesson 10: Interactions

Nicky Wakim

2025-04-28

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

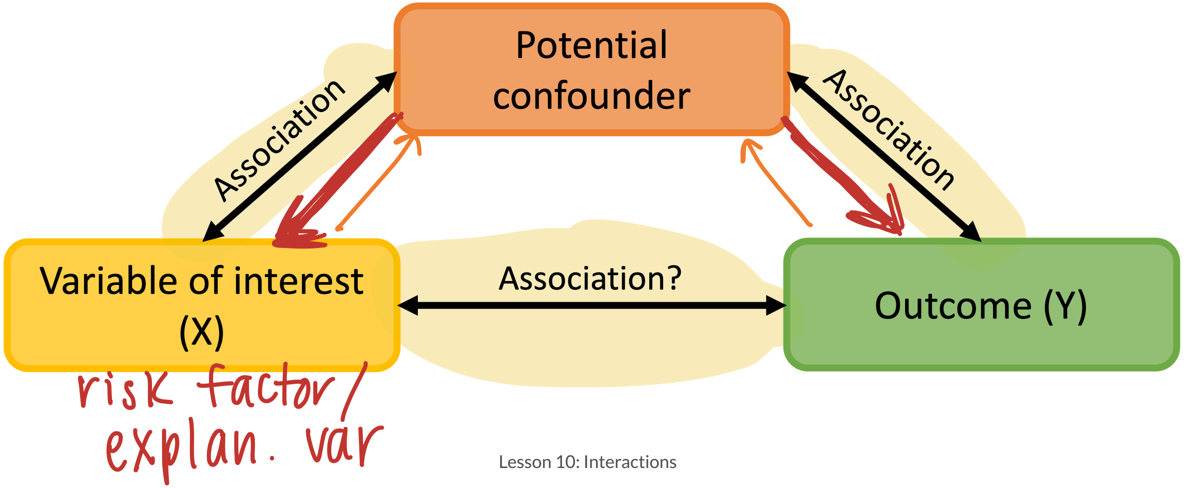
1. Connect understanding of confounding and interactions from linear regression to logistic regression.
2. Determine if an additional independent variable is a not a confounder nor effect modifier, is a confounder but not effect modifier, or is an effect modifier.
3. Calculate and interpret fitted interactions, 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…
  - 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



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

covariates

- 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

RF

X_1 vs. X_2

Y vs. X_2

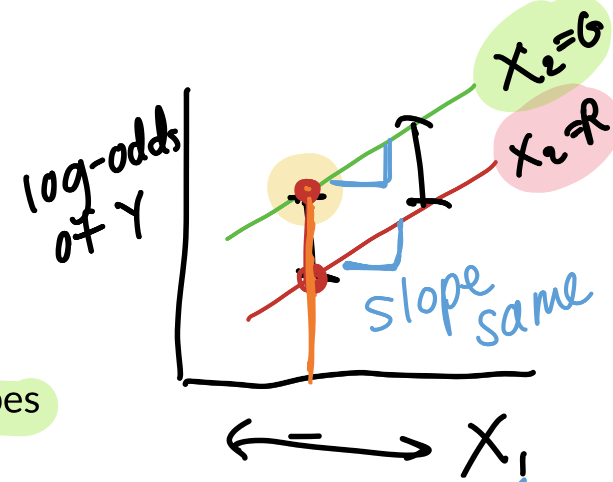
- *Hint hint:* This is why we did all those plots in Lab 2
  - To get a sense what might be important to include in the model

Including a confounder in the model

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

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

- And we assume that every level of the potential confounder, there is parallel slopes



every 1 unit inc of X_1, same slope regardless of X_2

- Note: to interpret β_1, we did not specify any value of X_2; only specified that it be held constant
  - Implicit assumption: effect of X_1 is equal across all values of X_2

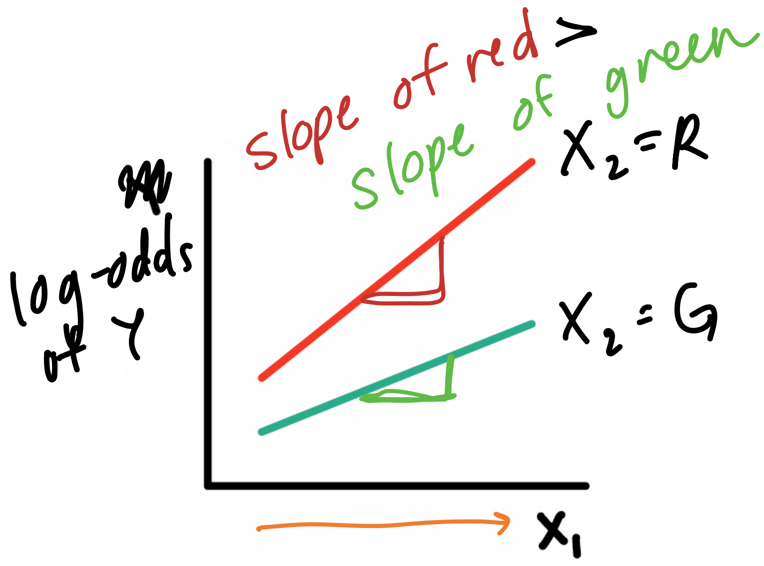
same X_1, diff estimated prob for $X_2=G$ vs. $X_2=R$

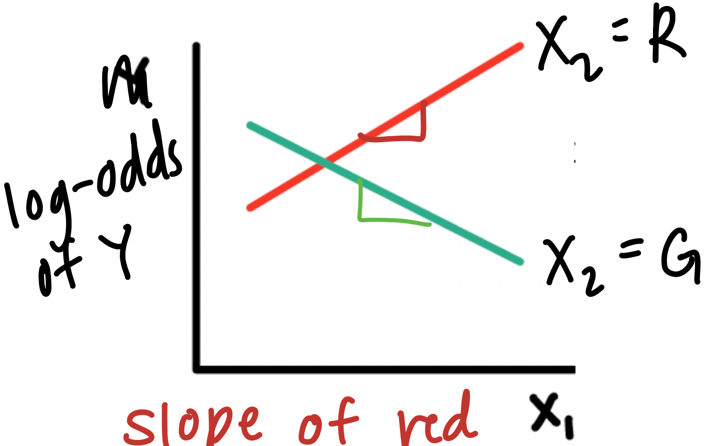
- The above model assumes that X_1 and X_2 do not *interact* (with respect to their effect on Y)
  - Epidemiology: no "effect modification"
  - Meaning the effect of X_1 is the same regardless of the values of X_2

What is an effect modifier?

- 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
  - Aka: as the effect modifier's values change, so does the association between Y and X_1
  - So the coefficient estimating the relationship between Y and X_1 changes with another variable X_2





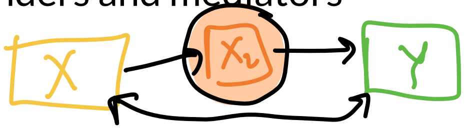
need to anchor X_1 v. Y relationship in value of X_2

slope of red pos, but slope of green negative

* interpretation of odds ratio (and value itself) must be reported for specific X_2

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*/ covariate
  - 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!! interaction term X_2 has an interaction w/ X_1
- We can incorporate interactions into our model through product terms:

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

- Terminology:
  - main effect parameters: β_1, β_2
    - The main effect models estimate the *average* X_1 and X_2 effects
  - interaction parameter: β_3

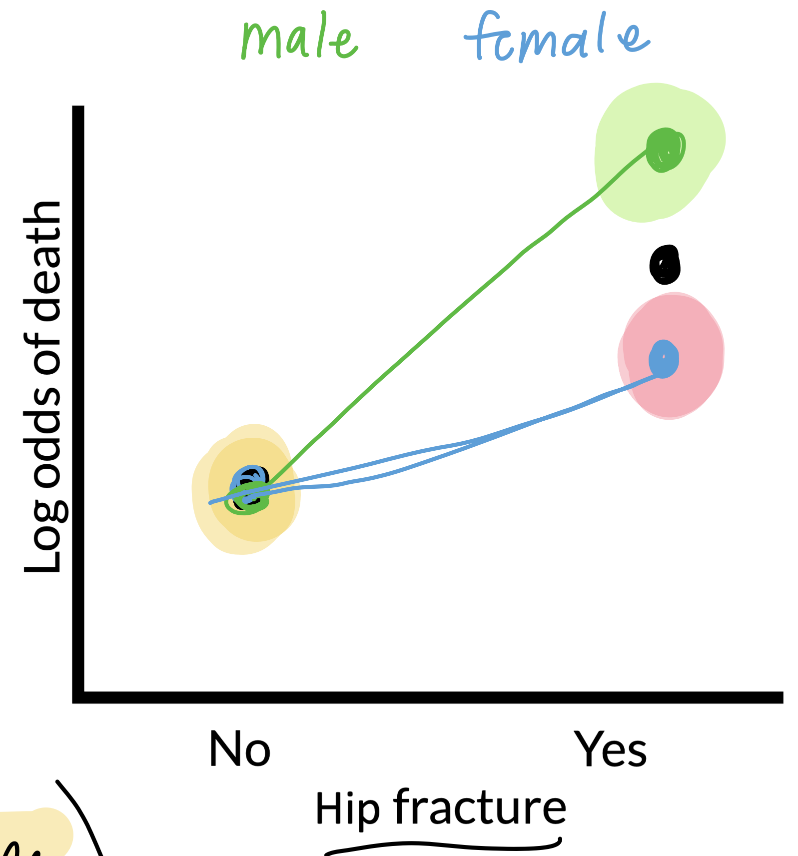
Example of interaction

- 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 of death comparing prev. hip fracture to no hip fracture



$$\frac{\text{odds death given HF}}{\text{odds death NO HF}}$$

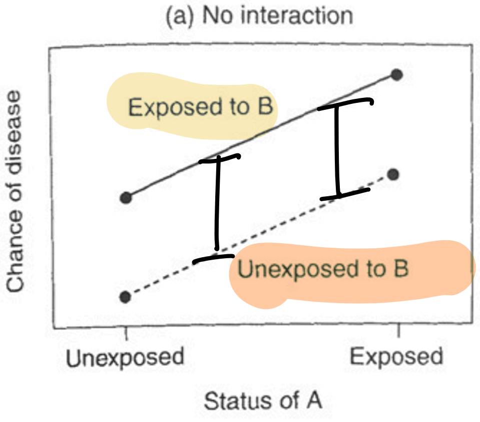
Types of interactions / non-interactions

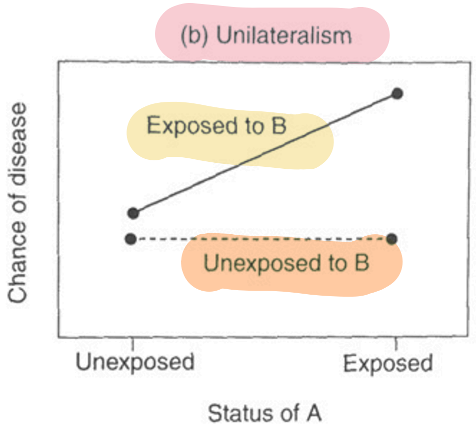
No interaction and three potential effects of interaction between two covariates A and B:

- **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

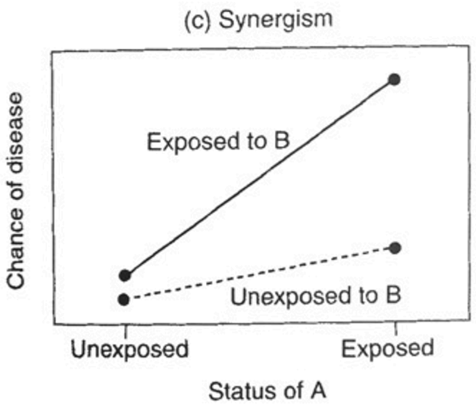
Confounder, but no interaction between A and B

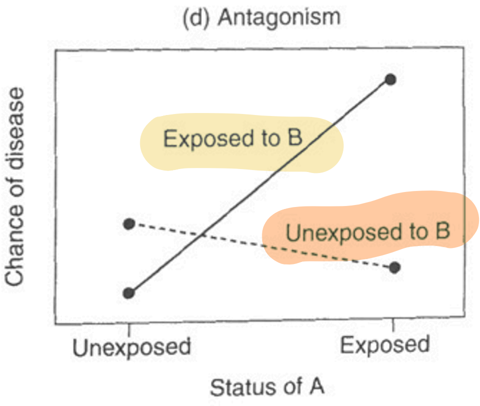




Exposure to A has no effect in the absence of exposure to B, but a considerable effect when B is present

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 1

13:42 Mon Apr 28 91%

Join by Web PollEv.com/nickywakim275

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

No interaction	0%
Unilateralism	11%
Synergism ✓	83%
Antogonism	6%

Powered by Poll Everywhere

log odds HF vs. NO HF does not dec

Example of interaction

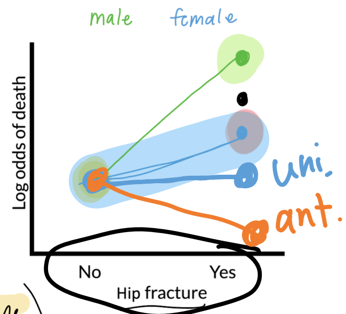
- In a cohort study of elderly people the chance of death (outcome) within 2 years was much higher for those who had previously suffered a hip fracture at the start of these 2 years, but the excess risk associated with a hip fracture was ...gher for males vs. females

...action between hip fracture status (yes/no) ...ar if assigned at birth or no)

...females < odds ratio for males

ratio of death ...mparing prev. hip fracture to no hip fracture

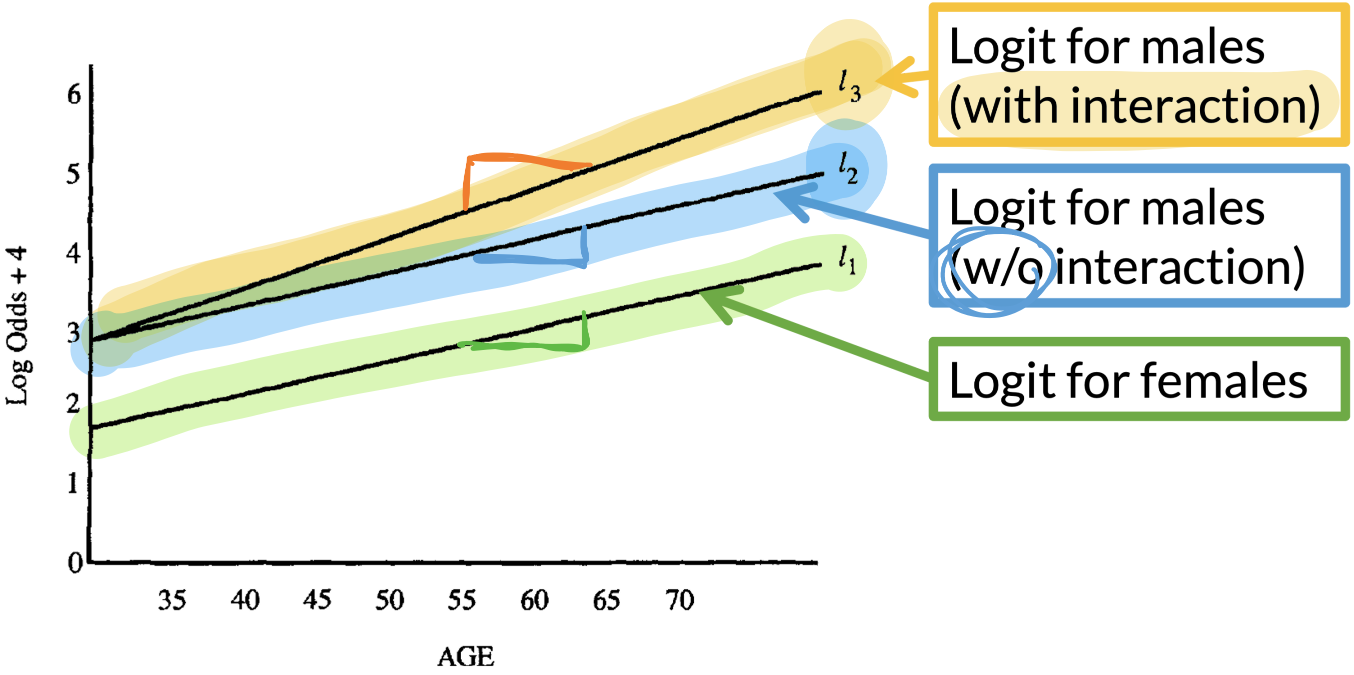
$\dfrac{\text{odds death given HF}}{\text{odds death NO HF}}$



Lesson 10: Interactions

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

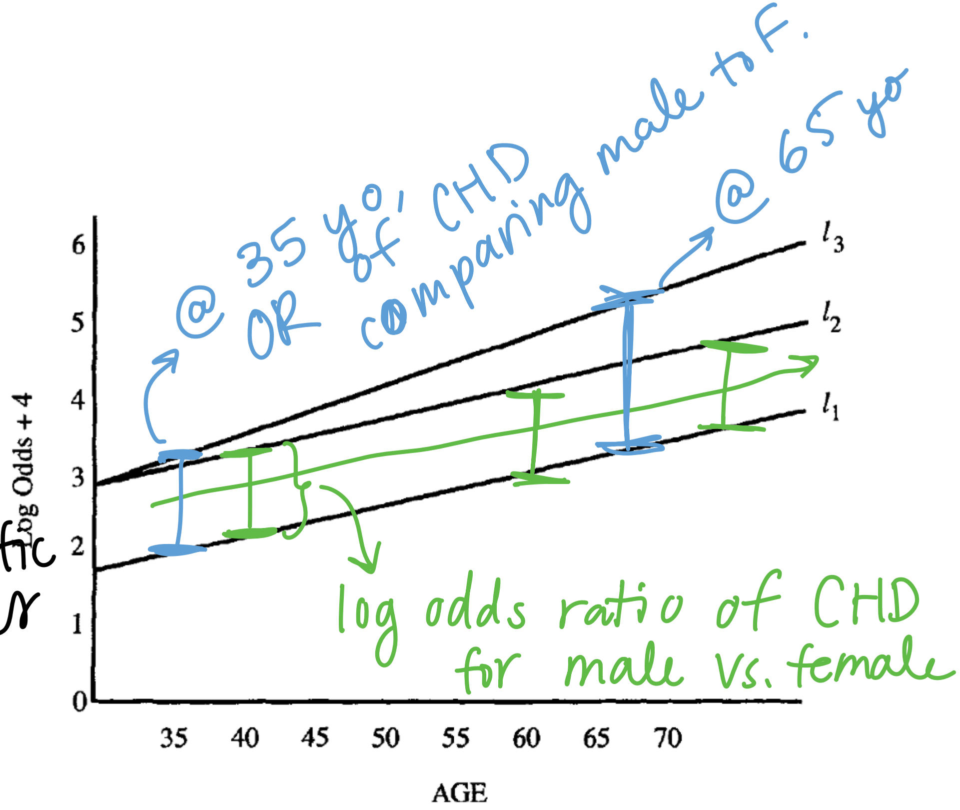


- Response variable: CHD
- Risk factor: sex
- Covariate to be controlled: age

Understand the interaction (2/3)

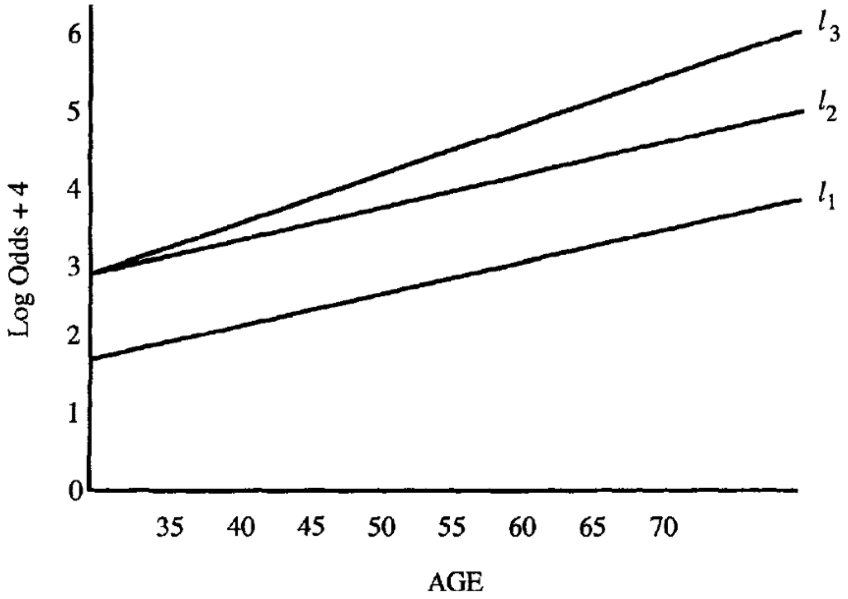
- **If age does not interact with sex**, the distance between l_2 and l_1 is the log odds ratio for sex, controlling for age ($l_2 - l_1$) stays the same for all values of age.

- **If age interacts with sex**, the distance between l_3 and l_1 is the log odds ratio for sex, ~~controlling for age~~. at specific ages
  - 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



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**

- Likelihood ratio test (or Wald test sometimes) may be used to test the significance of an interaction term

compare model w/ interaction
to model w/out interaction

Poll Everywhere Question 2

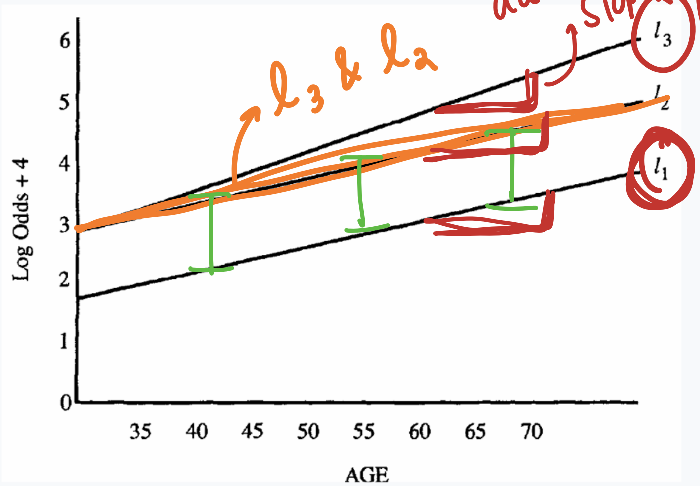
13:51 Mon Apr 28 ••• 88%

Join by Web **PollEv.com/nickywakim275**

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

$I(s=M)$

additional slope from β_3



l_3 is statistically different from l_2	7%
l_3 is statistically different from l_1	14%
l_3 is not statistically different from l_2 ✓	50%
l_3 is not statistically different from l_1	29%

pval 0.04

Powered by Poll Everywhere

$H_0: \beta_3 = 0$

$H_1: \beta_3 \neq 0$

p-val of 0.4

Fail to reject.

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

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

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.

Deciding between confounder and effect modifier

- This is more of a model selection question (in coming lectures)
- 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
   - 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!!**
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%

Step 1: Testing the interaction

$$\beta_0 + \beta_1 X_1 + \beta_2 I(X_2=1) + \beta_3 X_1 I(X_2=1)$$

- We test with $\alpha = 0.10$
- Follow the LRT procedure in ~~Lesson 6~~ Lesson 8, 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 cat

~~Binary~~ & continuous variables

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

$$\beta_0 + \beta_1 X_1 + \beta_2 I(X_2=1) + \beta_3 I(X_2=2) + \beta_4 X_1 I(X_2=1) + \beta_5 X_1 I(X_2=2)$$

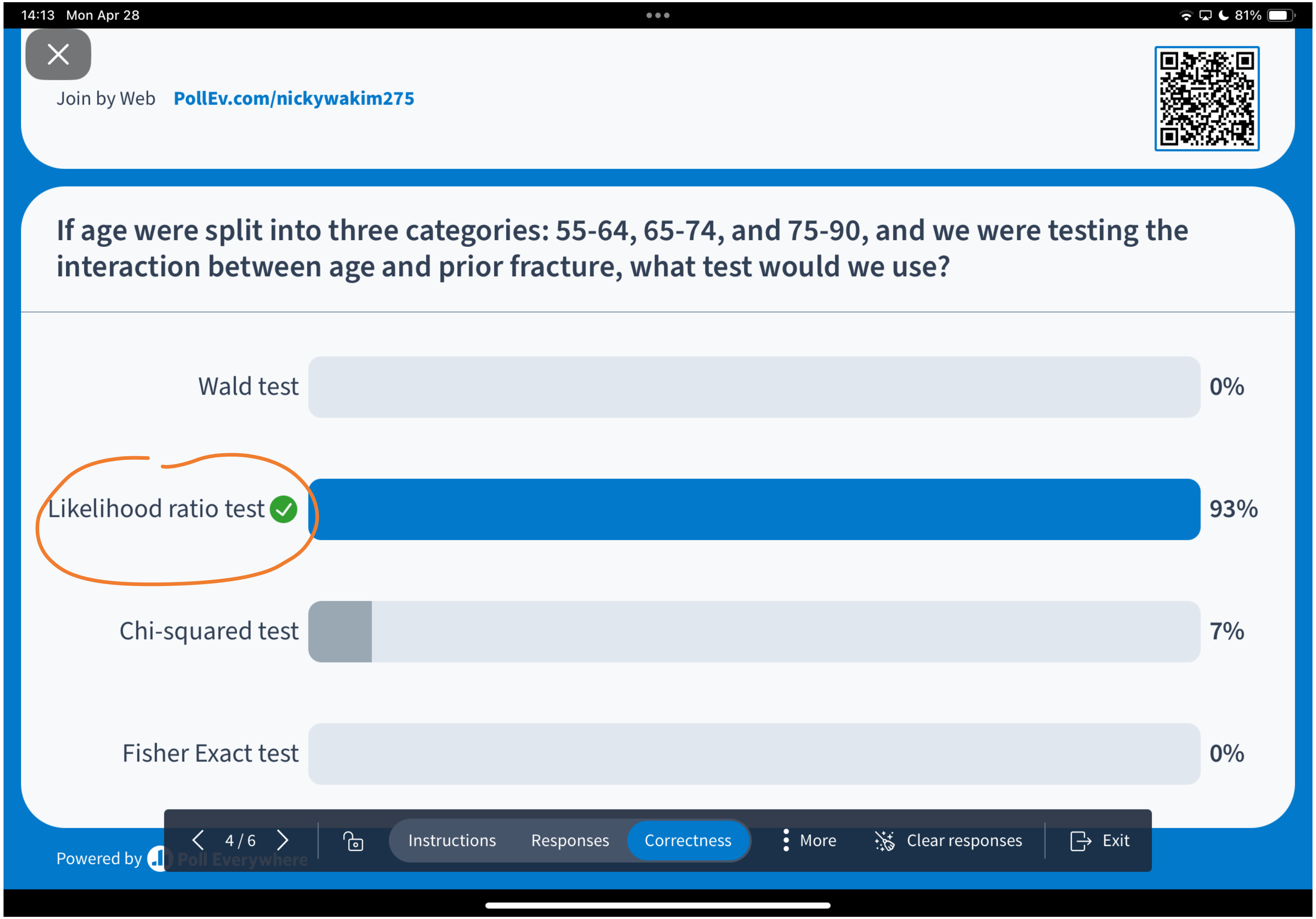
Binary & multi-level variable

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



Step 2: Testing a confounder

- If interaction already included:
  - Meaning: LRT showed evidence for alternative/full 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%
- 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

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

can be full − red or red − full w/ abs value

abs value

Calculation for % change in coefficient

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

full on bottom

Poll Everywhere Question 3

14:19 Mon Apr 28 79%

Join by Web PollEv.com/nickywakim275

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 10% change" rule, is x_2 a confounder?

$\hat{\beta}_{1,red}$

$\hat{\beta}_{1,full}$

x_2 is not a confounder	71%
x_2 is a confounder ✓	29%

3 / 6 Instructions Responses Correctness More Clear responses Exit

Powered by Poll Everywhere

$$\Delta\% = 100\% \times \left(\frac{\hat{\beta}_{1,full} - \hat{\beta}_{1,red}}{\hat{\beta}_{1,full}}\right)$$

$$= 100\% \times \left(\frac{0.49 - 0.4}{0.49}\right)$$

$$= 100\% \cdot 0.1837$$

$$= 18.37\%$$

$$> 10\%$$

Example: GLOW Study

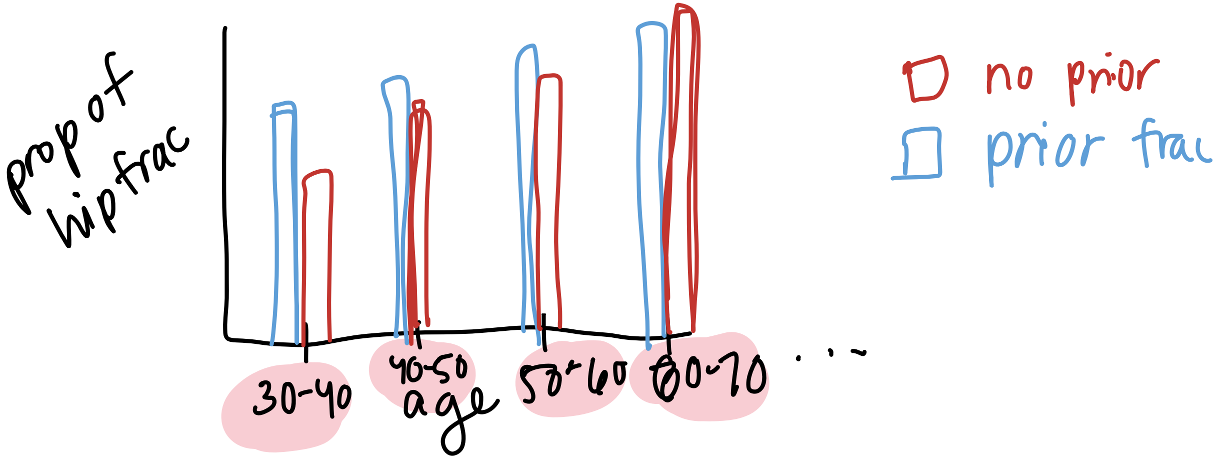
- From GLOW (Global Longitudinal Study of Osteoporosis in Women) study
- **Outcome variable:** any fracture in the first year of follow up (FRACTURE: 0 or 1)
- **Risk factor/variable of interest:** history of prior fracture (PRIORFRAC: 0 or 1)
- **Potential confounder or effect modifier:** age (AGE, a continuous variable)
  - Center age will be used! We will center around the rounded mean age of 69 years old

```r
library(aplore3)
mean_age = mean(glow500$age) %>% round()
glow = glow500 %>% mutate(age_c = age - 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

X2 X1 Y

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

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
```

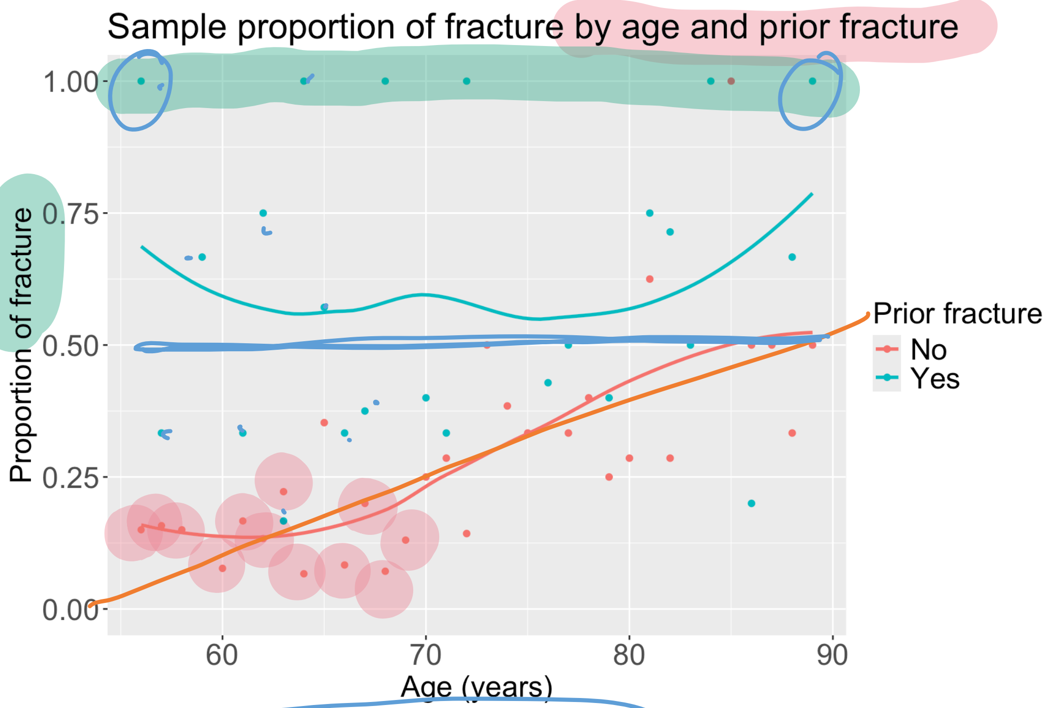
/No

1 or 0

age	prior	frac
56	No	No
56	No	Yes
56	Yes	No
56	Yes	Yes

Example: GLOW Study: Plot the proportions

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



- 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):

prob of future fracture

model 1

- Model 1: Age not included

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

prior fracture

if not model 2

2

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

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

2

if not model 3

1

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

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

1

see if interaction is sig.

Example: GLOW Study

We also *could 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_c,
              data = glow, family = binomial)
```

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

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

→ used on slide 52

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

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

- 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

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

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

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

Poll Everywhere Question 4

14:34 Mon Apr 28 74%

Join by Web PollEv.com/nickywakim275

log odds @ mean age & NO prior frac

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?

β_2 ✓	25%
$\beta_0 + \beta_2$	50%
$\beta_2 + \beta_3$ (w/ PF)	19%
$\beta_0 + \beta_2 + \beta_3$	6%

Powered by Poll Everywhere

if prior frac:

$\beta_0 + \beta_1(1) + \beta_2(age) + \beta_3(1)\,age$

age eff w/ PF:

$\beta_2 + \beta_3$

if NO PF:

$\beta_0 + \cancel{\beta_1(0)} + \beta_2(age) + \cancel{\beta_3(0)\,age}$

age eff w/ NPF:

β_2

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

a diff ages what are the diff ORs?

Example: GLOW – Interaction interpretation

- **Model 3:**

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

$$\text{logit}\left(\widehat{\pi}(\mathbf{X})\right) = \left[\underline{-1.376} \quad \underline{+1.002} \cdot I(\text{PF}) \quad \underline{+0.063} \cdot Age \right] \quad \underline{-0.057 \cdot I(\text{PF}) \cdot Age}$$

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

interaction: value of age changes the relationship b/w PF & Fract.

Example: GLOW – Interaction interpretation

$\text{logit}(\pi(X)) = \hat\beta_0 + \hat\beta_1 I(PF) + \hat\beta_2 \, age_c + \beta_3 I(PF) \cdot age_c$

prob of frac

$age_c = 0$

- Model 3 estimated **coefficients**:

term	estimate	p.value	conf.low	conf.high
(Intercept)	−1.376	0.000	−1.646	−1.120
priorfracYes	1.002	0.000	0.530	1.471
age_c	0.063	0.000	0.032	0.093
priorfracYes:age_c	−0.057	0.022	−0.107	−0.008

- Model 3 estimated **odds ratios**:

term	estimate	p.value	conf.low	conf.high
(Intercept)	0.25	0.00	0.19	0.33
priorfracYes	2.72	0.00	1.70	4.35
age_c	1.06	0.00	1.03	1.10
priorfracYes:age_c	0.94	0.02	0.90	0.99

when $age_c = 0$
(aka 69 yo)

- $\widehat{\beta}_3 = -0.057$
- 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
- $\widehat{\beta}_1 = 1.002$
- 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).
- $\exp(\widehat{\beta}_1) = 2.72$
- 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).

$\widehat{OR} = \exp(\hat\beta_1) = \dfrac{\widehat{odds}_{PF}}{odds_{NoPF}}$

estimated odds of new fract is 2.72 times for those w/ PF vs. noPF

Poll Everywhere Question 5

$\widehat{OR}$ of future fracture comparing PF to no PF

Age	PF	$\widehat{OR}$
50	no	–
	yes	
60	no	–
	yes	
70	no	–
	yes	

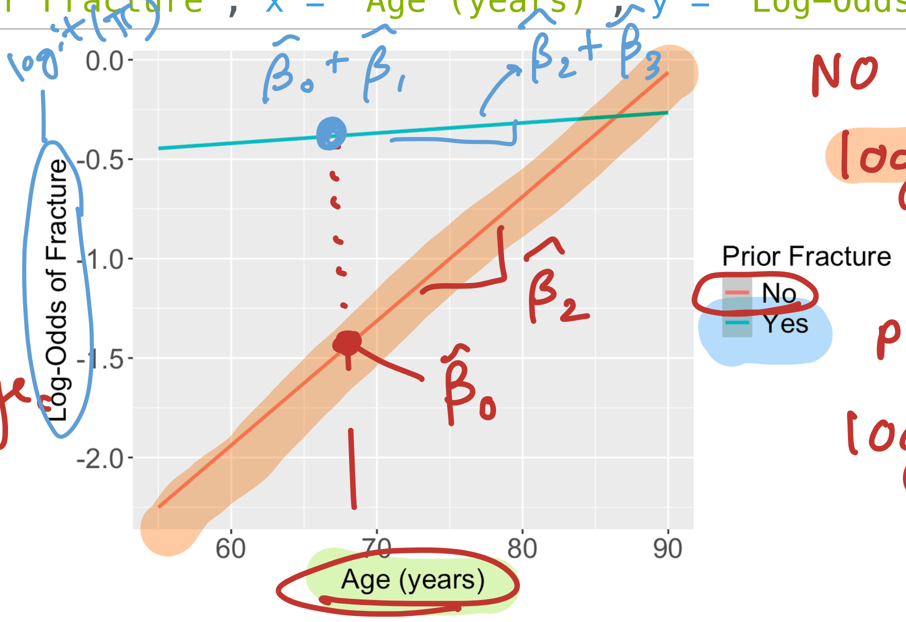
$$\frac{\widehat{\text{odds of fract}}\ (\text{for PF})}{\widehat{\text{odds of fract}}\ (\text{for NO PF})}$$

Plot of estimated log odds

```r
prior_age = expand_grid(priorfrac = c("No", "Yes"), age_c = (55:90)-69)
frac_pred_log = predict(glow_m3, prior_age, se.fit = T, type="link")
pred_glow2 = prior_age %>% mutate(frac_pred_log = frac_pred_log$fit,
                                  age = age_c + mean_age)

ggplot(pred_glow2) + #geom_point(aes(x = age, y = frac_pred, color = priorfrac)) +
  geom_smooth(method = "loess", aes(x = age, y = frac_pred_log, color = priorfrac)) +
  theme(text = element_text(size=20), title = element_text(size=16)) +
  labs(color = "Prior Fracture", x = "Age (years)", y = "Log-Odds of Fracture")
```

$\mathrm{logit}(\pi(x)) = \beta_0 + \beta_1 I(PF) + \beta_2 Age + \beta_3 I(PF) \cdot Age$



NO PF: $I(PF) = 0$

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

PF: $I(PF) = 1$

$\mathrm{logit}(\hat\pi(x)) = (\hat\beta_0 + \hat\beta_1) + (\hat\beta_2 + \hat\beta_3) Age$

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

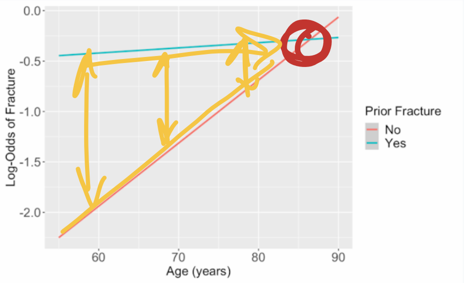
13:35 Wed Apr 30 46%

Join by Web **PollEv.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.

$$\text{logit}(\pi(\text{PF}=1, \text{Age}=a)) = \beta_0 + \beta_1 + \beta_2 a + \beta_3 a = (\beta_0 + \beta_1) + (\beta_2 + \beta_3)a$$

$$\text{logit}(\pi(\text{PF}=0, \text{Age}=a)) = \beta_0 + \beta_2 a$$



β_3	5%
$\frac{\beta_2}{\beta_0}$	26%
$\frac{-\beta_0}{\beta_2}$	5%
$\frac{-\beta_1}{\beta_3}$ ✓	37%
$\frac{\beta_3}{\beta_2}$	26%

Powered by **Poll Everywhere**

$$\cancel{\hat\beta_0} + \hat\beta_1 + (\cancel{\hat\beta_2} + \hat\beta_3)age = \cancel{\hat\beta_0} + \cancel{\hat\beta_2 age}$$

$$\hat\beta_1 + \hat\beta_3 age = 0$$

$$\hat\beta_3 age = -\hat\beta_1$$

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

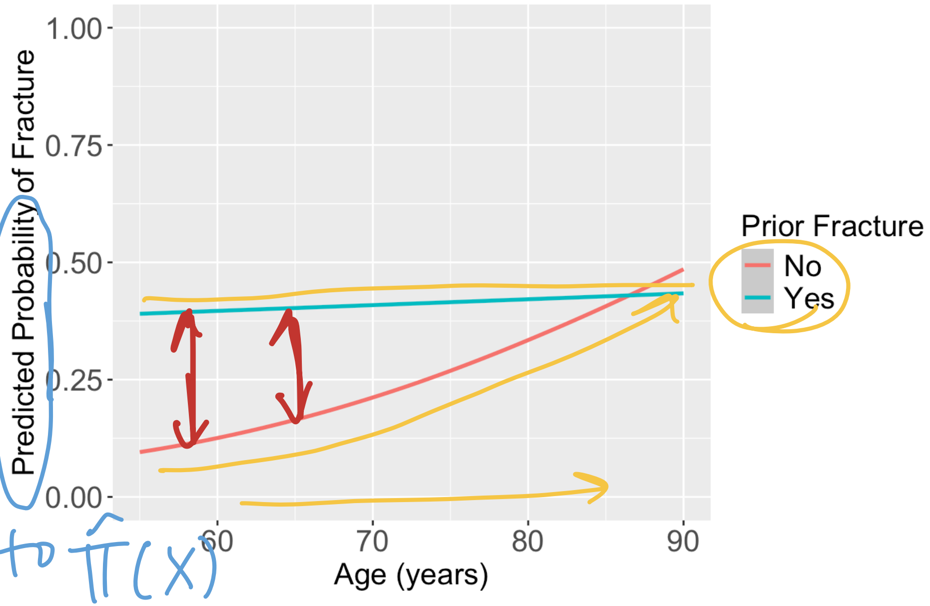
Plot the predicted probability of fracture

```
frac_pred = predict(glow_m3, prior_age, se.fit = T, type="response")
pred_glow = prior_age %>% mutate(frac_pred = frac_pred$fit,
                                 age = age_c + mean_age)

ggplot(pred_glow) + #geom_point(aes(x = age, y = frac_pred, color = priorfrac)) +
  geom_smooth(method = "loess", aes(x = age, y = frac_pred, color = priorfrac)) +
  theme(text = element_text(size=20), title = element_text(size=16)) + ylim(0,1) +
  labs(color = "Prior Fracture", x = "Age (years)", y = "Predicted Probability of Fracture"
```

w/ inv logit

to convert from $\text{logit}(\hat{\pi}(x))$ to $\hat{\pi}(x)$



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

of Y

- 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

PF   age

- When an interaction term (F*X) exists in the model
  - $\widehat{OR}_F \neq \exp(\widehat{\beta}_F)$ in general

PF   no PF

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

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

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

$$\text{logit}(\hat{\pi}(\text{age}=a, I(PF)=1)$$
$$- \text{logit}(\hat{\pi}(\text{age}=a, I(PF)=0))$$

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

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

$$\hat{g}\left(F=f_1, X=x\right)=\hat{\beta}_0+\hat{\beta}_1 f_1+\hat{\beta}_2 x+\hat{\beta}_3 f_1 \cdot\ x$$
$$\hat{g}\left(F=f_0, X=x\right)=\hat{\beta}_0+\hat{\beta}_1 f_0+\hat{\beta}_2 x+\hat{\beta}_3 f_0 \cdot\ x$$

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

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

- Therefore,

$$\widehat{OR}(F=f-1, F=f_0, X=x)=\widehat{OR}_F=\exp\left[\hat{\beta}_1\left(f_1-f_0\right)+\hat{\beta}_3 x\left(f_1-f_0\right)\right]$$

Steps to compute OR under interation

- 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

Example: GLOW Study

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

- Set 1: $PF = 1, Age = a$
- Set 2: $PF = 0, Age = a$

2. Substitute values in the fitted log-odds model

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

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

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

Example: GLOW Study

3. Take the difference of the two log-odds

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

4. Exponentiate the resulting difference

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

odds ratio of fract comparing PF to no PF @ age a

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

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

At age 60, estimated odds of fracture comparing PF to no PF is 4.55 times

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

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

age centered
(centered @ mean 69 yo)

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

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

```r
library(gmodels)
glow_m3 %>% estimable(
                    cbind("(Intercept)"                          = 0,    # beta0
                          "priorfracYes"                         = 1,    # beta1
                          "age_c"                                = 0,    # beta2
                          "priorfracYes:age_c"                   = (55:90)-69),
                    conf.int = 0.95) %>%
  exp(.) %>%
  head(5)
```

$\hat\beta_1 + \hat\beta_3 a$

given age (centered)

to get ORs

```
                 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
```

55

56

57

58

59

55 - 69

gives you CI

Calculate odds ratios across values (predict())

```r
prior_age = expand_grid(priorfrac = c("No", "Yes"), age_c = (55:90)-69)
frac_pred_logit = predict(glow_m3, prior_age, se.fit = T, type="link")
pred_glow2 = prior_age %>% mutate(frac_pred = frac_pred_logit$fit,
                                  age = age_c + mean_age) %>%

                pivot_wider(names_from = priorfrac, values_from = frac_pred) %>%

                mutate(OR_YN = exp(Yes - No))
head(pred_glow2)
```

log-odds for no PF

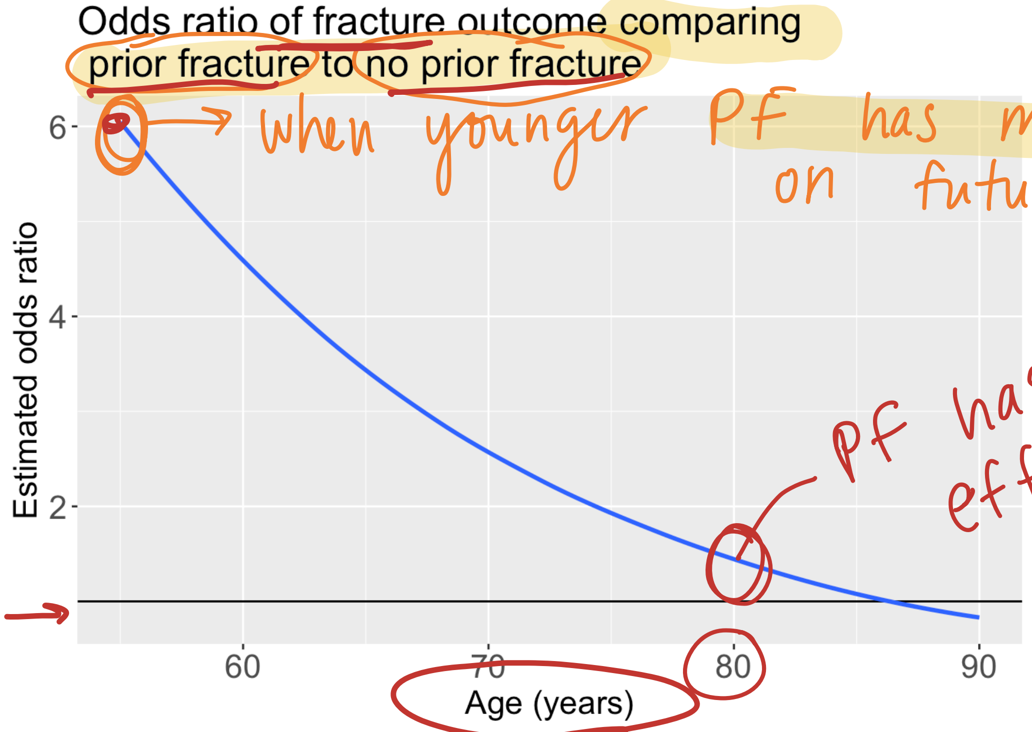
log-odds for PF

```
# A tibble: 6 × 5
  age_c   age    No    Yes OR_YN
  <dbl> <dbl> <dbl>  <dbl> <dbl>
1   -14    55 -2.25 -0.446  6.08
2   -13    56 -2.19 -0.441  5.74
3   -12    57 -2.13 -0.436  5.42
4   -11    58 -2.06 -0.430  5.12
5   -10    59 -2.00 -0.425  4.83
6    -9    60 -1.94 -0.420  4.56
```

exp(Yes - No)

Plotting the odds ratio for an interaction

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



when younger PF has more effect on future fracture

PF has less effect on future fract.

@1

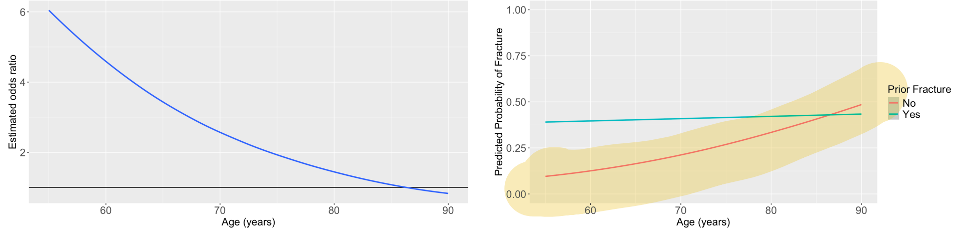
How would I report these results?

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

$\exp(\hat{\beta}_1)$

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). As seen in Figure 1 (a), the estimated odds ratio of a new fracture when comparing prior fracture status to no PF 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].

$-\hat{\beta}_1/\hat{\beta}_3$



(a) Odds ratio of fracture outcome comparing prior fracture to no prior fracture

(b) Predicted probability of fracture

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