

# **AS.280.347**

# **CLASS 1.4**

- Look at interpretations of logistic regressions and propensity score results
- Effect modification



# Module 1: Smoking and risk of disease

- Question 1.1 (Q1.1): How does the risk of disease compare for smokers and otherwise similar non-smokers?
- Question 1.2 (Q1.2): Does the contribution of smoking to the risk of disease vary by sex or SES?
- To address each question, we want:
  - a data display
  - a statistical analysis
- We will answer these questions using data from the National Medical Expenditures Survey (NMES)

# Interpreting propensity score results

- Propensity scores based on **age** and **sex**:

```
propModel<-glm(eversmk ~ age + female, data=nmes.data,  
                family=binomial(link="logit"))  
  
predLogOdds<-predict(propModel)  
predProb<-exp(predLogOdds) / (1+exp(predLogOdds))  
probCutoffs<-quantile(predProb, probs=c(0,0.2, 0.4, 0.6, 0.8, 1))  
probQuintiles<-cut(predProb, probCutoffs, include.lowest = TRUE)
```

- Relationship between disease and smoking, adjusted for propensity score quintile:

```
glmOut<-glm(mscd ~ eversmk + probQuintiles, data=nmes.data,  
                family=binomial(link="logit"))  
summary(glmOut)
```

# Interpreting propensity score results

```
> summary(glmOut)
```

Call:

```
glm(formula = mscd ~ eversmk + probQuintiles, family = binomial(link = "logit"),  
     data = nmes.data)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-0.6362	-0.4860	-0.2074	-0.1107	3.3555

We estimate that the **log odds** of disease is **.53 higher** for smokers compared to non-smokers *of similar age and sex*.

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-5.6262	0.5077	-11.083	< 2e-16 ***
eversmkYes	<b>0.5336</b>	0.1413	3.776	<b>0.000159 ***</b>
probQuintiles(0.405, 0.419]	1.7977	0.5460	3.292	0.000994 ***
probQuintiles(0.419, 0.437]	3.5495	0.5121	6.931	4.18e-12 ***
probQuintiles(0.437, 0.606]	1.1060	0.5796	1.908	0.056352 .
probQuintiles(0.606, 0.638]	3.5979	0.5114	7.036	1.98e-12 ***

We would consider the *relationship statistically significant!*

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 2024.6 on 4077 degrees of freedom  
Residual deviance: 1710.1 on 4072 degrees of freedom  
AIC: 1722.1

Number of Fisher Scoring iterations: 8

We estimate that the **odds** of disease for smokers are **1.71 times** the odds for non-smokers *of similar age and sex*. The **odds** of disease are **71% higher** for smokers compared to non-smokers *of similar age and sex*.

# Interpreting propensity score results

```
> summary(glmOut)
```

Call:

```
glm(formula = mscd ~ eversmk + probQuintiles, family = binomial(link = "logit"),
  data = nmes.data)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-0.6362	-0.4860	-0.2074	-0.1107	3.3555

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-5.6262	0.5077	-11.083	< 2e-16 ***
eversmkYes	0.5336	0.1413	3.776	0.000159 ***
probQuintiles(0.405, 0.419]	1.7977	0.5460	3.292	0.000994 ***
probQuintiles(0.419, 0.437]	3.5495	0.5121	6.931	4.18e-12 ***
probQuintiles(0.437, 0.606]	1.1060	0.5796	1.908	0.056352 .
probQuintiles(0.606, 0.638]	3.5979	0.5114	7.036	1.98e-12 ***

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 2024.6 on 4077 degrees of freedom

Residual deviance: 1710.1 on 4072 degrees of freedom

AIC: 1722.1

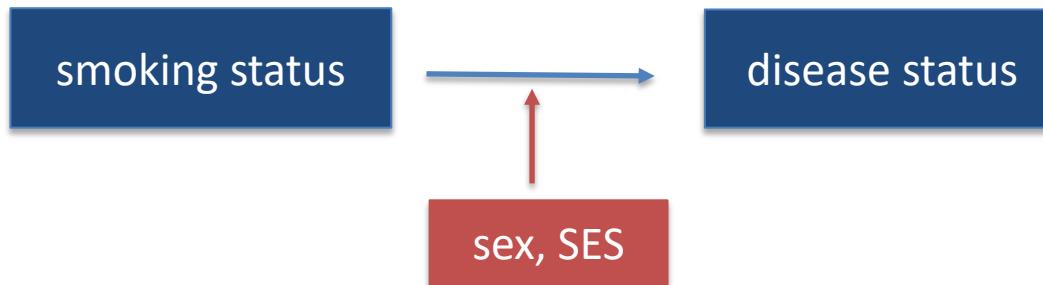
Number of Fisher Scoring iterations: 8

Don't need to interpret these since they are just for adjustment purposes and not the relationship of interest!

# Now for Question 1.2

Question 1.2 (Q1.2): Does the contribution of smoking to the risk of disease vary by sex or SES?

- Data display?
- Statistical analysis?



- An **effect modification** (or interaction) is present when the relationship between a predictor of interest and the outcome varies by the level (subgroup) of another variable

# Ways to include effect modification

- Do separate analyses for each subgroup; report estimate with confidence interval by subgroup

OR

- Include interaction between smoking and indicator of subgroup in second step of propensity score analysis:

```
glm(mscd ~ eversmk*sex + ps.groups,  
     family=binomial(link="logit"))
```

- Be sure subgroup indicator is in propensity model

```
propModel <- glm(eversmk ~ ??? + sex,  
                  family=binomial(link="logit"))
```

# Interpreting effect modification results

- Analyses separately for males/females:

```
nmes.female <- nmes.data %>% filter(female=="Female")
nmes.male <- nmes.data %>% filter(female=="Male")
propModel<-glm(eversmk ~ age, data=nmes.female,
                family=binomial(link="logit"))
predLogOdds<-predict(propModel)
predProb<-exp(predLogOdds) / (1+exp(predLogOdds))
probCutoffs<-quantile(predProb, probs=c(0,0.2, 0.4, 0.6, 0.8, 1))
probQuintiles<-cut(predProb, probCutoffs, include.lowest = TRUE)
glmOut.female<-glm(mscd ~ eversmk + probQuintiles, data=nmes.female,
                     family=binomial(link="logit"))
summary(glmOut.female)
confint.default(glmOut.female)
```

← don't include sex here!

Then repeat the same analyses for males!

# Interpreting effect modification results

```
> summary(glmOut.female)
```

Call:

```
glm(formula = mscd ~ eversmk + probQuintiles, family = binomial(link = "logit"),  
     data = nmes.female)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-0.7251	-0.3578	-0.1647	-0.0933	3.4115

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	-1.8205	0.1390	-13.094	< 2e-16 ***	$\exp(0.6187) = 1.86$
eversmkYes	0.6187	0.1862	3.324	0.000888 ***	$\exp(0.2539) = 1.29$
probQuintiles(0.359, 0.402]	-0.8963	0.2035	-4.405	1.06e-05 ***	$\exp(0.9836) = 2.67$
probQuintiles(0.402, 0.437]	-2.4727	0.3608	-6.853	7.25e-12 ***	
probQuintiles(0.437, 0.462]	-3.6149	0.5936	-6.090	1.13e-09 ***	
probQuintiles(0.462, 0.491]	-3.9959	0.7198	-5.551	2.84e-08 ***	

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

```
> confint.default(glmOut.female)
```

	2.5 %	97.5 %
(Intercept)	-2.0929965	-1.5479856
eversmkYes	0.2538642	0.9835957
probQuintiles(0.359, 0.402]	-1.2950945	-0.4975151
probQuintiles(0.402, 0.437]	-3.1799374	-1.7654594
probQuintiles(0.437, 0.462]	-4.7783059	-2.4514382
probQuintiles(0.462, 0.491]	-5.4067054	-2.5849948

We estimate that the odds of disease for female smokers are **1.86 times** the odds for female non-smokers of similar ages (95% CI for OR: 1.29 to 2.67).

# Interpreting effect modification results

```
> summary(glmOut.male)
```

Call:  
glm(formula = mscd ~ eversmk + probQuintiles, family = binomial(link = "logit"),  
data = nmes.male)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-0.8627	-0.5177	-0.2145	-0.0958	3.4857

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	-6.0729	1.0098	-6.014	1.81e-09 ***	$\exp(0.6908) = 2.00$
eversmkYes	0.6908	0.2386	2.896	0.003785 **	$\exp(0.2232) = 1.25$
probQuintiles(0.488, 0.558]	1.2242	1.1219	1.091	0.275193	$\exp(1.158) = 3.18$
probQuintiles(0.558, 0.641]	2.3126	1.0495	2.204	0.027556 *	
probQuintiles(0.641, 0.731]	3.4397	1.0201	3.372	0.000747 ***	
probQuintiles(0.731, 0.857]	4.5853	1.0115	4.533	5.81e-06 ***	

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

```
> confint.default(glmOut.male)
```

	2.5 %	97.5 %
(Intercept)	-8.0520901	-4.093645
eversmkYes	0.2232090	1.158432
probQuintiles(0.488, 0.558]	-0.9747158	3.423182
probQuintiles(0.558, 0.641]	0.2556370	4.369488
probQuintiles(0.641, 0.731]	1.4403205	5.439146
probQuintiles(0.731, 0.857]	2.6027590	6.567759

We estimate that the odds of disease for male smokers are 2.0 times the odds for male non-smokers of similar ages (95% CI for OR: 1.25 to 3.18).

# Interpreting effect modification results

- Including interaction term along with propensity adjustment:

```
propModel<-glm(eversmk ~ age + female, data=nmes.data,  
family=binomial(link="logit"))  
  
predLogOdds<-predict(propModel)  
predProb<-exp(predLogOdds) / (1+exp(predLogOdds))  
probCutoffs<-quantile(predProb, probs=c(0,0.2, 0.4, 0.6, 0.8, 1))  
probQuintiles<-cut(predProb, probCutoffs, include.lowest = TRUE)  
  
glmOut<-glm(mscd ~ eversmk*female + probQuintiles, data=nmes.data,  
family=binomial(link="logit"))  
summary(glmOut)  
confint.default(glmOut)
```

be sure to include sex here!

include interaction with sex here

# Interpreting effect modification results

```
> summary(glmOut)
```

Call:

```
glm(formula = mscd ~ eversmk * female + probQuintiles, family = binomial(link = "logit"),  
     data = nmes.data)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-0.6441	-0.4871	-0.2080	-0.1100	3.3538

No evidence of effect modification  
(interaction) between sex and smoking  
status because the interaction term is not  
statistically significant!

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-8.5671	0.8065	-10.622	< 2e-16 ***
eversmkYes	0.6661	0.2323	2.867	0.004139 **
femaleFemale	2.9468	0.6270	4.700	2.60e-06 ***
probQuintiles(0.405, 0.419]	1.7979	0.5460	3.293	0.000992 ***
probQuintiles(0.419, 0.437]	3.5482	0.5123	6.926	4.34e-12 ***
probQuintiles(0.437, 0.606]	3.4615	0.6949	4.981	6.32e-07 ***
probQuintiles(0.606, 0.638]	6.4333	0.7985	8.057	7.85e-16 ***
eversmkYes:femaleFemale	-0.1430	0.2961	-0.483	0.629056

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

# Interpreting effect modification results

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-8.5671	0.8065	-10.622	< 2e-16 ***
eversmkYes	0.6661	0.2323	2.867	0.004139 **
femaleFemale	2.9468	0.6270	4.700	2.60e-06 ***
probQuintiles(0.405, 0.419]	1.7979	0.5460	3.293	0.000992 ***
probQuintiles(0.419, 0.437]	3.5482	0.5123	6.926	4.34e-12 ***
probQuintiles(0.437, 0.606]	3.4615	0.6949	4.981	6.32e-07 ***
probQuintiles(0.606, 0.638]	6.4333	0.7985	8.057	7.85e-16 ***
eversmkYes:femaleFemale	-0.1430	0.2961	-0.483	0.629056

If the interaction term HAD been statistically significant, we would have interpreted the interaction term the following way:

$$\begin{aligned}\text{Females: } \log(\text{odds of disease}) &= \beta_0 + \beta_1 eversmk + \beta_2(1) + \cdots + \beta_7 eversmk * (1) \\ &= (\beta_0 + \beta_2) + (\beta_1 + \beta_7) eversmk + \cdots\end{aligned}$$

$$\begin{aligned}\text{Males: } \log(\text{odds of disease}) &= \beta_0 + \beta_1 eversmk + \beta_2(0) + \cdots + \beta_7 eversmk * (0) \\ &= \beta_0 + \beta_1 eversmk + \cdots\end{aligned}$$

The coefficient on the interaction term tells the additional affect of smoking on disease for females compared to males. So the difference in the log odds of disease for females is **.14 lower** than the difference in the log odds of disease for males of a similar age. So the OR of disease for females is **0.87 times** the OR of disease for males of a similar age. So the OR of disease for females is **13% lower** than the odds ratio of disease for males of a similar age.

# Assignment 1.4

Finalize your report for Module 1 to answer Q1.1 and Q1.2.

- ***For each question, you should have a data display and a statistical analysis to address the question.***
- ***Provide a caption for your data display(s).***  
[\(http://abacus.bates.edu/~ganderso/biology/resources/writing/HTWtablefigs.html\)](http://abacus.bates.edu/~ganderso/biology/resources/writing/HTWtablefigs.html)
- ***Write up your results in a few paragraphs to answer both questions. In your write-up, you should refer to your data display(s) and your analysis results. Be numerate!***
- Work together in groups!
- Submit your assignment in R markdown through Blackboard by Sunday @ midnight.