**OBJECTIVE:** Our objective is to describe the smoking habits of the participants in the Framingham Heart study as they age and the impact of smoking on certain health outcomes.

For this analysis, we will be working as part of a research lab that studies smoking habits of adults. We are interested to describe the smoking habits of the participants in the Framingham Heart study as they age and the impact of smoking on certain health outcomes.

**STUDY DESIGN:** The Framingham Heart Study is a long-term prospective study of the etiology of cardiovascular disease among a population of free-living subjects in the community of Framingham, Massachusetts. Participant clinic data was collected during three examination periods, approximately 6 years apart, from roughly 1956 to 1968. Each participant has 1 to 3 observations depending on the number of exams the subject attended, and as a result there are 11,627 observations on the 4,434 participants. The Framingham heart study asks participants about their smoking habits at each visit. In particular, participants are asked if they are currently smoking at this visit (0 = Not a current smoker, 1 = Current smoker), which we will refer to as current smoking status. In addition, participants also report the number of cigarettes they are smoking per day. Clinic examination data has included cardiovascular disease risk factors and markers of disease such as blood pressure, blood chemistry, lung function, smoking history, health behaviors, ECG tracings, Echocardiography, and medication use.

**METHODS:** We first explored the data to evaluate missingness by using the “5% Rule” on all potential covariates. For each model, confounders were assessed using the “10% Rule.”

If the observed effect sizes in regression models differed by a relative magnitude of at least 10%, those covariates were retained in the model. The relationship between smoking status and age was explored graphically in addition to being fit using a generalized linear mixed-effects model.

The relationship between number of cigarettes smoked per day and age was explored graphically adjusted for sex. A linear mixed-effects model was fit to our four remaining questions of interest.

What we didn’t include:

* Why we chose not to do random slope
* Cross-sectional vs longitudinal
* Correlation

**RESULTS:** Upon applying the 5% rule to assess missing data, the following variables were removed: bpmeds (5.1%), glucose (12.4%), hdlc (74.0%), ldlc (74.0%). Because bpmeds are used to treat hypertension, thus correlated with sysbp and diasbp and prevhyp, this variable was removed. Because elevated glucose levels (hyperglycemia) are most often encountered clinically in the setting of diabetes mellitus, this variable was removed. Lastly, hdlc and ldlc are components of total cholesterol values, thus total serum cholesterol serves as an appropriate proxy. Observations for variables with less than 5% missing data were removed. No imputation was performed. Upon assessing the relationship between age and smoking status, we found that smoking status decreased with age. This trend was consistent among both males and females. We explored the data to evaluate the relationship between age and number of cigarettes smoked per day stratified by sex and found that the number of cigarettes smoked per day decreases with age. This trend is consistent with the downward trend for smoking status with age and was similar amongst both males and females.

The variability in number of cigarettes smoked per day also appears to decrease with age. (I saw this in a previous plot that was no longer in the rmd. Maybe just delete this point.) In addition, males appear to smoke more cigarettes per day on average compared to females of the same age group.

**With regard to systolic blood pressure:**

We found that

Among both non-smokers and smokers, there appears to be a generally linear relationship between systolic bp and age. No interaction was found.

Among both non-smokers and smokers, there appears to be a generally linear relationship between systolic bp and age.

In addition, there appears to be some interaction between bmi and current smoking status.

Because current smoking status is our primary covariate of interest, we included this in our model.

Among both males and females, there appears to be a generally linear relationship between systolic bp and age. While there seemed to be an interaction between age and sex, this interaction was not included in our model as our primary covariate of interest was current smoking status.

Among both males and females, there appears to be a generally linear relationship between systolic bp and bmi. Interaction between these bmi and sex were excluded from our model for the same reason stated previously.

Need interpretation of coefficients\*\*\*

Note -- systolic BP increases with age.

**With regard to diastolic blood pressure:**

Among both non-smokers and smokers, there appears to be a curvilinear relationship between Diastolic BP and age. No interaction was observed.

Among both non-smokers and smokers, there appears to be a generally linear relationship between Diastolic BP and bmi.

In addition, there appears to be some interaction between bmi and current smoking status. We included a product term for these two covariates to account for interaction.

Among both males and females, there appears to be a curvilinear relationship between Diastolic BP and age.

Among both males and females, there appears to be a curvilinear relationship between diastolic and bmi.

Need interpretation of coefficients\*\*\*

Note -- diastolic BP decreases with age (opposite systolic).

Curvilinear relationship should be reflected as a quadratic, no? (dbl-chk)

**With regard to serum total cholesterol:**

Among both non-smokers and smokers, there appears to be a curvilinear relationship between total cholesterol and age.

In addition, there appears to be some interaction between age and current smoking status becasue the two lines intersect. We included in model.

Among both non-smokers and smokers, there appears to be a curvilinear relationship between total cholesterol and bmi.

In addition, there appears to be some interaction between bmi and current smoking status. Included interaction term in model

Among both males and females, there appears to be a curvilinear relationship between total cholesterol and age.

Among both males and females, there appears to be a curvilinear relationship between total cholesterol and bmi.

Need interpretation of coefficients\*\*\*

Note -- diastolic BP decreases with age (opposite systolic).

Curvilinear relationship should be reflected as a quadratic, no? (dbl-chk)

CONCLUSION:

Our objectives were

(1) to describe the smoking habits of the participants in the Framingham Heart study as they age,

and

(2) to describe the impact of smoking on systolic blood pressure, diastolic blood pressure, and total serum cholesterol.

Significant differences were found \_\_\_

###

Appendix (2 tables/figures)

###

Appendix C: R code

## EDA for missing values  
smoke <- read.csv(file = 'frmgham2.csv') %>%   
 clean\_names()  
  
## checking missing values (5% rule)  
pMiss <- function(x){sum(is.na(x))/length(x)\*100}  
apply(smoke,2,pMiss) #2 indicates columns  
  
  
## Remove hdlc, ldlc, glucose, and bpmeds for having more than 5% of missing values.  
smoke\_vs1 = smoke %>%   
 filter(period==1) %>%   
 dplyr::select(c(randid,sex,age,cursmoke,totchol,bmi,heartrte,educ,diabp,sysbp,diabetes,prevap,prevchd,prevmi,prevstrk,prevhyp)) %>%   
 mutate(cursmoke = as.factor(cursmoke), sex=as.factor(sex), diabetes=as.factor(diabetes)) %>%   
 na.omit()  
## For those variables with less then 5% missing data, observations were removed. (i.e. no imputation performed on missing values)

# Question 1

## Identification of potential confounders

## Prevalent Hypertensive.   
glm1 <- glm(cursmoke ~ age+sex, data=smoke\_vs1,family = 'binomial')  
glm2 <- glm(cursmoke ~ age+sex+prevhyp, data=smoke\_vs1,family = 'binomial')  
  
#########################################  
## confounder check: 10% rule for age ##  
#########################################  
abs((exp(summary(glm1)$coefficients[2]) - exp(summary(glm2)$coefficients[2])) / exp(summary(glm2)$coefficients[2])) \* 100  
## [1] 0.390565 -- % diff between age coefficients b/t glm1&glm2; not a cf  
  
#########################################  
## confounder check: 10% rule for sex ##  
#########################################  
abs((exp(summary(glm1)$coefficients[3]) - exp(summary(glm2)$coefficients[3])) / exp(summary(glm2)$coefficients[3])) \* 100  
## [1] 0.4203757 -- % diff between sex coefficients b/t glm1&glm2; not a cf  
  
## prevhyp is not confounder for sex and age with cursmoke

## Prevalent Stroke  
glm1 <- glm(cursmoke ~ age+sex, data=smoke\_vs1,family = 'binomial')  
glm2 <- glm(cursmoke ~ age+sex+prevstrk, data=smoke\_vs1,family = 'binomial')  
  
#########################################  
## confounder check: 10% rule for age ##  
#########################################  
abs((exp(summary(glm1)$coefficients[2]) - exp(summary(glm2)$coefficients[2])) / exp(summary(glm2)$coefficients[2])) \* 100  
## [1] 0.01235394 -- % diff between age coefficients b/t glm1&glm2; not a cf  
  
#########################################  
## confounder check: 10% rule for sex ##  
#########################################  
abs((exp(summary(glm1)$coefficients[3]) - exp(summary(glm2)$coefficients[3])) / exp(summary(glm2)$coefficients[3])) \* 100  
## 1] 0.002629692 -- % diff between sex coefficients b/t glm1&glm2; not a cf  
  
## prevstrk is not confounder for sex and age with cursmoke

## Prevalent Myocardial Infarction  
glm1 <- glm(cursmoke ~ age+sex, data=smoke\_vs1,family = 'binomial')  
glm2 <- glm(cursmoke ~ age+sex+prevmi, data=smoke\_vs1,family = 'binomial')  
  
#########################################  
## confounder check: 10% rule for age ##  
#########################################  
abs((exp(summary(glm1)$coefficients[2]) - exp(summary(glm2)$coefficients[2])) / exp(summary(glm2)$coefficients[2])) \* 100  
## [1] 0.06734953 -- % diff between age coefficients b/t glm1&glm2; not a cf  
  
#########################################  
## confounder check: 10% rule for sex ##  
#########################################  
abs((exp(summary(glm1)$coefficients[3]) - exp(summary(glm2)$coefficients[3])) / exp(summary(glm2)$coefficients[3])) \* 100  
## [1] 1.067897 -- % diff between sex coefficients b/t glm1&glm2; not a cf  
  
## prevmi is not confounder for sex and age with cursmoke

## Prevalent Coronary Heart Disease   
glm1 <- glm(cursmoke ~ age+sex, data=smoke\_vs1,family = 'binomial')  
glm2 <- glm(cursmoke ~ age+sex+prevchd, data=smoke\_vs1,family = 'binomial')  
  
#########################################  
## confounder check: 10% rule for age ##  
#########################################  
abs((exp(summary(glm1)$coefficients[2]) - exp(summary(glm2)$coefficients[2])) / exp(summary(glm2)$coefficients[2])) \* 100  
## [1] 0.03469214 -- % diff between age coefficients b/t glm1&glm2; not a cf  
  
#########################################  
## confounder check: 10% rule for sex ##  
#########################################  
abs((exp(summary(glm1)$coefficients[3]) - exp(summary(glm2)$coefficients[3])) / exp(summary(glm2)$coefficients[3])) \* 100  
## [1] 0.2778121 -- % diff between sex coefficients b/t glm1&glm2; not a cf  
  
## prevchd is not confounder for sex and age with cursmoke

## Prevalent Angina Pectoris at exam  
glm1 <- glm(cursmoke ~ age+sex, data=smoke\_vs1,family = 'binomial')  
glm2 <- glm(cursmoke ~ age+sex+prevap, data=smoke\_vs1,family = 'binomial')  
  
#########################################  
## confounder check: 10% rule for age ##  
#########################################  
abs((exp(summary(glm1)$coefficients[2]) - exp(summary(glm2)$coefficients[2])) / exp(summary(glm2)$coefficients[2])) \* 100  
## [1] 0.03561141 -- % diff between age coefficients b/t glm1&glm2; not a cf  
  
#########################################  
## confounder check: 10% rule for sex ##  
#########################################  
abs((exp(summary(glm1)$coefficients[3]) - exp(summary(glm2)$coefficients[3])) / exp(summary(glm2)$coefficients[3])) \* 100  
## [1] 0.2606049 -- % diff between sex coefficients b/t glm1&glm2; not a cf  
  
## prevap is not confounder for sex and age with cursmoke

## Serum Total Cholesterol (mg/dL)  
glm1 <- glm(cursmoke ~ age+sex, data=smoke\_vs1,family = 'binomial')  
glm2 <- glm(cursmoke ~ age+sex+totchol, data=smoke\_vs1,family = 'binomial')  
  
#########################################  
## confounder check: 10% rule for age ##  
#########################################  
abs((exp(summary(glm1)$coefficients[2]) - exp(summary(glm2)$coefficients[2])) / exp(summary(glm2)$coefficients[2])) \* 100  
## [1] 0.1257789 -- % diff between age coefficients b/t glm1&glm2; not a cf  
  
#########################################  
## confounder check: 10% rule for sex ##  
#########################################  
abs((exp(summary(glm1)$coefficients[3]) - exp(summary(glm2)$coefficients[3])) / exp(summary(glm2)$coefficients[3])) \* 100  
## [1] 0.6115748 -- % diff between sex coefficients b/t glm1&glm2; not a cf  
  
## totchol is not confounder for sex and age with cursmoke

## Systolic Blood Pressure (mean of last two of three measurements) (mmHg)  
glm1 <- glm(cursmoke ~ age+sex, data=smoke\_vs1,family = 'binomial')  
glm2 <- glm(cursmoke ~ age+sex+sysbp, data=smoke\_vs1,family = 'binomial')  
  
#########################################  
## confounder check: 10% rule for age ##  
#########################################  
abs((exp(summary(glm1)$coefficients[2]) - exp(summary(glm2)$coefficients[2])) / exp(summary(glm2)$coefficients[2])) \* 100  
## [1] 0.5150948 -- % diff between age coefficients b/t glm1&glm2; not a cf  
  
#########################################  
## confounder check: 10% rule for sex ##  
#########################################  
abs((exp(summary(glm1)$coefficients[3]) - exp(summary(glm2)$coefficients[3])) / exp(summary(glm2)$coefficients[3])) \* 100  
## [1] 0.7407342 -- % diff between sex coefficients b/t glm1&glm2; not a cf  
  
## sysbp is not confounder for sex and age with cursmoke

## Diastolic Blood Pressure (mean of last two of three measurements) (mmHg)  
glm1 <- glm(cursmoke ~ age+sex, data=smoke\_vs1,family = 'binomial')  
glm2 <- glm(cursmoke ~ age+sex+diabp, data=smoke\_vs1,family = 'binomial')  
  
#########################################  
## confounder check: 10% rule for age ##  
#########################################  
abs((exp(summary(glm1)$coefficients[2]) - exp(summary(glm2)$coefficients[2])) / exp(summary(glm2)$coefficients[2])) \* 100  
## [1] 0.3804091 -- % diff between age coefficients b/t glm1&glm2; not a cf  
  
#########################################  
## confounder check: 10% rule for sex ##  
#########################################  
abs((exp(summary(glm1)$coefficients[3]) - exp(summary(glm2)$coefficients[3])) / exp(summary(glm2)$coefficients[3])) \* 100  
## [1] 2.271954 -- % diff between sex coefficients b/t glm1&glm2; not a cf  
  
## diabp is is not confounder for sex and age with cursmoke

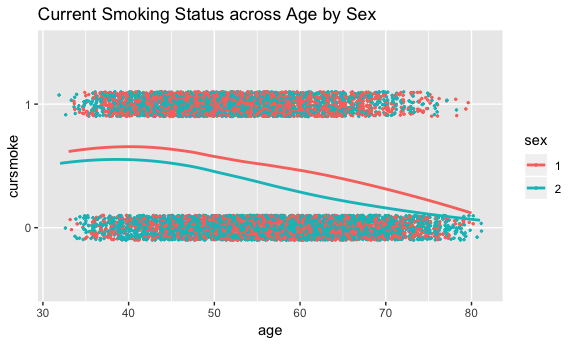
## Body Mass Index, weight in kilograms/height meters squared  
glm1 <- glm(cursmoke ~ age+sex, data=smoke\_vs1,family = 'binomial')  
glm2 <- glm(cursmoke ~ age+sex+bmi, data=smoke\_vs1,family = 'binomial')  
  
#########################################  
## confounder check: 10% rule for age ##  
#########################################  
abs((exp(summary(glm1)$coefficients[2]) - exp(summary(glm2)$coefficients[2])) / exp(summary(glm2)$coefficients[2])) \* 100  
## [1] 0.3996838 -- % diff between age coefficients b/t glm1&glm2; not a cf  
  
#########################################  
## confounder check: 10% rule for sex ##  
#########################################  
abs((exp(summary(glm1)$coefficients[3]) - exp(summary(glm2)$coefficients[3])) / exp(summary(glm2)$coefficients[3])) \* 100  
## [1] 8.611817 -- % diff between sex coefficients b/t glm1&glm2; not a cf  
  
## bmi is is not confounder for sex and age with cursmoke

## Heart rate (Ventricular rate) in beats/min  
glm1 <- glm(cursmoke ~ age+sex, data=smoke\_vs1,family = 'binomial')  
glm2 <- glm(cursmoke ~ age+sex+heartrte, data=smoke\_vs1,family = 'binomial')  
  
#########################################  
## confounder check: 10% rule for age ##  
#########################################  
abs((exp(summary(glm1)$coefficients[2]) - exp(summary(glm2)$coefficients[2])) / exp(summary(glm2)$coefficients[2])) \* 100  
## [1] 0.009174943 -- % diff between age coefficients b/t glm1&glm2; not a cf  
  
#########################################  
## confounder check: 10% rule for sex ##  
#########################################  
abs((exp(summary(glm1)$coefficients[3]) - exp(summary(glm2)$coefficients[3])) / exp(summary(glm2)$coefficients[3])) \* 100  
## [1] 4.304516 -- % diff between sex coefficients b/t glm1&glm2 is 4.305; not a cf  
  
## heartrte is is not confounder for sex and age with cursmoke

## Attained Education  
glm1 <- glm(cursmoke ~ age+sex, data=smoke\_vs1,family = 'binomial')  
glm2 <- glm(cursmoke ~ age+sex+educ, data=smoke\_vs1,family = 'binomial')  
  
#########################################  
## confounder check: 10% rule for age ##  
#########################################  
abs((exp(summary(glm1)$coefficients[2]) - exp(summary(glm2)$coefficients[2])) / exp(summary(glm2)$coefficients[2])) \* 100  
## [1] 0.09356206 -- % diff between age coefficients b/t glm1&glm2; not a cf  
  
#########################################  
## confounder check: 10% rule for sex ##  
#########################################  
abs((exp(summary(glm1)$coefficients[3]) - exp(summary(glm2)$coefficients[3])) / exp(summary(glm2)$coefficients[3])) \* 100  
## [1] 0.1610224 % diff between sex coefficients b/t glm1&glm2; not a cf  
  
## educ is not confounder for sex and age with cursmoke

## Diabetic according to criteria of first exam treated or first exam with casual glucose of 200 mg/dL or more  
glm1 <- glm(cursmoke ~ age+sex, data=smoke\_vs1,family = 'binomial')  
glm2 <- glm(cursmoke ~ age+sex+diabetes, data=smoke\_vs1,family = 'binomial')  
  
#########################################  
## confounder check: 10% rule for age ##  
#########################################  
abs((exp(summary(glm1)$coefficients[2]) - exp(summary(glm2)$coefficients[2])) / exp(summary(glm2)$coefficients[2])) \* 100  
## [1] 0.07727996 -- % diff between age coefficients b/t glm1&glm2; not a cf  
  
#########################################  
## confounder check: 10% rule for sex ##  
#########################################  
abs((exp(summary(glm1)$coefficients[3]) - exp(summary(glm2)$coefficients[3])) / exp(summary(glm2)$coefficients[3])) \* 100  
## [1] 0.3456463 -- % diff between sex coefficients b/t glm1&glm2; not a cf  
  
## diabetes is not confounder for sex and age with cursmoke

## no confounders identified; model with age:sex  
  
## limit data to covariates of interest  
smoke\_vs3 = smoke %>%   
 dplyr::select(c(randid,sex,age,cursmoke)) %>%   
 mutate(cursmoke = as.factor(cursmoke),   
 sex=as.factor(sex)) %>%   
 na.omit()  
  
## visualize data  
smoke\_vs3 %>%  
 ggplot(aes(age, cursmoke, group = sex, color = sex)) +  
 geom\_jitter(height = 0.1, size = 0.5) +  
 geom\_smooth(method = "loess", se = F) +  
 ggtitle("Current Smoking Status across Age by Sex")



## smoking generally declines with age  
  
  
## model data  
glmer\_1 <- glmer(cursmoke ~ age\*sex + (1 | randid), data = smoke\_vs3, family = binomial)  
summary(glmer\_1) ## all coefficient estimates were found to be significant

## Generalized linear mixed model fit by maximum likelihood (Laplace  
## Approximation) [glmerMod]  
## Family: binomial ( logit )  
## Formula: cursmoke ~ age \* sex + (1 | randid)  
## Data: smoke\_vs3  
##   
## AIC BIC logLik deviance df.resid   
## 10840.6 10877.4 -5415.3 10830.6 11622   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -3.5920 -0.1412 -0.0524 0.1965 3.5958   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## randid (Intercept) 34.34 5.86   
## Number of obs: 11627, groups: randid, 4434  
##   
## Fixed effects:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 13.71738 0.84790 16.178 < 2e-16 \*\*\*  
## age -0.23847 0.01456 -16.383 < 2e-16 \*\*\*  
## sex2 -6.93043 1.09252 -6.344 2.25e-10 \*\*\*  
## age:sex2 0.05611 0.01671 3.358 0.000785 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr) age sex2   
## age -0.974   
## sex2 -0.833 0.810   
## age:sex2 0.761 -0.788 -0.950

######################  
## results table 1 ##  
######################  
coefs1 <- data.frame(round(coef(summary(glmer\_1)), 2)); coefs1

## Estimate Std..Error z.value Pr...z..  
## (Intercept) 13.72 0.85 16.18 0  
## age -0.24 0.01 -16.38 0  
## sex2 -6.93 1.09 -6.34 0  
## age:sex2 0.06 0.02 3.36 0

#LL CI  
LL\_CI\_glmer\_1 = rbind(  
round(coefs1[1, 1] - 1.96\*coefs1[1, 2], 5), ## Intercept  
round(coefs1[2, 1] - 1.96\*coefs1[2, 2], 5), ## CI age  
round(coefs1[3, 1] - 1.96\*coefs1[3, 2], 5), ## CI sex  
round(coefs1[4, 1] - 1.96\*coefs1[4, 2], 5)) ## CI age:sex  
  
## UL CI  
UL\_CI\_glmer\_1 = rbind(  
round(coefs1[1, 1] + 1.96\*coefs1[1, 2], 5), ## Intercept  
round(coefs1[2, 1] + 1.96\*coefs1[2, 2], 5), ## CI age  
round(coefs1[3, 1] + 1.96\*coefs1[3, 2], 5), ## CI sex  
round(coefs1[4, 1] + 1.96\*coefs1[4, 2], 5)) ## CI age:sex  
  
results\_1 = cbind(coefs1, LL\_CI\_glmer\_1, UL\_CI\_glmer\_1) %>%   
 dplyr::select(1, 2, 5, 6, 4)

# Question 2

## variable selection  
smoke\_vs2 = smoke %>%   
 filter(period==1) %>%   
 dplyr::select(c(randid,sex,age,cigpday,totchol,bmi,heartrte,educ,diabp,sysbp,diabetes)) %>%   
 mutate(sex=as.factor(sex), diabetes=as.factor(diabetes)) %>%   
 na.omit()

## Identification of potential confounders bet age and sex with # of cigarettes smoked per day

## Serum Total Cholesterol (mg/dL)  
lm1 <- lm(cigpday ~ age+sex, data=smoke\_vs2)  
lm2 <- lm(cigpday ~ age+sex+totchol, data=smoke\_vs2)  
  
#########################################  
## confounder check: 10% rule for age ##  
#########################################  
abs((summary(lm1)$coefficients[2] - summary(lm2)$coefficients[2]) / summary(lm2)$coefficients[2])\*100  
## [1] 5.98917-- % diff between age coefficients for lm1&lm2; not a cf  
  
#########################################  
## confounder check: 10% rule for sex ##  
#########################################  
abs((summary(lm1)$coefficients[3] - summary(lm2)$coefficients[3]) / summary(lm2)$coefficients[3])\*100  
## [1] 0.9520448 -- % diff between sex coefficients for lm1&lm2; not a cf  
  
## totchol is not confounder for age and sex with cigpday

## Body Mass Index, weight in kilograms/height meters squared  
lm1 <- lm(cigpday ~ age+sex, data=smoke\_vs2)  
lm2 <- lm(cigpday ~ age+sex+bmi, data=smoke\_vs2)  
  
##########################################  
## confounder check: 10% rule for age ##  
#########################################  
abs((summary(lm1)$coefficients[2] - summary(lm2)$coefficients[2]) / summary(lm2)$coefficients[2])\*100  
## [1] 7.09452 -- % diff between age coefficients for lm1&lm2; not a cf  
  
#########################################  
## confounder check: 10% rule for sex ##  
#########################################  
abs((summary(lm1)$coefficients[3] - summary(lm2)$coefficients[3]) / summary(lm2)$coefficients[3])\*100  
## [1] 2.097622 -- % diff between sex coefficients for lm1&lm2; not a cf  
  
## bmi is not confounder for age and sex with cigpday

## Heart rate (Ventricular rate) in beats/min  
lm1 <- lm(cigpday ~ age+sex, data=smoke\_vs2)  
lm2 <- lm(cigpday ~ age+sex+heartrte, data=smoke\_vs2)  
  
##########################################  
## confounder check: 10% rule for age ##  
#########################################  
abs((summary(lm1)$coefficients[2] - summary(lm2)$coefficients[2]) / summary(lm2)$coefficients[2])\*100  
## [1] 0.6952164 -- % diff between age coefficients for lm1&lm2; not a cf  
  
#########################################  
## confounder check: 10% rule for sex ##  
#########################################  
abs((summary(lm1)$coefficients[3] - summary(lm2)$coefficients[3]) / summary(lm2)$coefficients[3])\*100  
## [1] 3.411956 -- % diff between sex coefficients for lm1&lm2; not a cf  
  
## heartrte is not confounder for age and sex with cigpday

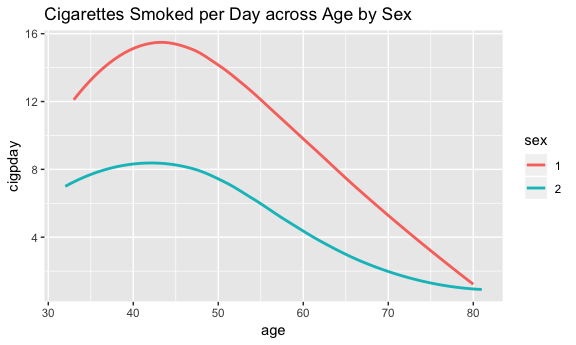
## Attained Education  
lm1 <- lm(cigpday ~ age+sex, data=smoke\_vs2)  
lm2 <- lm(cigpday ~ age+sex+educ, data=smoke\_vs2)  
  
##########################################  
## confounder check: 10% rule for age ##  
#########################################  
abs((summary(lm1)$coefficients[2] - summary(lm2)$coefficients[2]) / summary(lm2)$coefficients[2])\*100  
## [1] 2.764036 -- % diff between age coefficients for lm1&lm2; not a cf  
  
#########################################  
## confounder check: 10% rule for sex ##  
#########################################  
abs((summary(lm1)$coefficients[3] - summary(lm2)$coefficients[3]) / summary(lm2)$coefficients[3])\*100  
## [1] 0.1152682 -- % diff between sex coefficients for lm1&lm2; not a cf  
  
## educ is not confounder for age and sex with cigpday

## Diastolic Blood Pressure (mean of last two of three measurements) (mmHg)  
lm1 <- lm(cigpday ~ age+sex, data=smoke\_vs2)  
lm2 <- lm(cigpday ~ age+sex+diabp, data=smoke\_vs2)  
  
##########################################  
## confounder check: 10% rule for age ##  
#########################################  
abs((summary(lm1)$coefficients[2] - summary(lm2)$coefficients[2]) / summary(lm2)$coefficients[2])\*100  
## [1] 4.302741 -- % diff between age coefficients for lm1&lm2; not a cf  
  
#########################################  
## confounder check: 10% rule for sex ##  
#########################################  
abs((summary(lm1)$coefficients[3] - summary(lm2)$coefficients[3]) / summary(lm2)$coefficients[3])\*100  
## [1] 0.5923237 -- % diff between sex coefficients for lm1&lm2; not a cf  
  
## diabp is not confounder for age and sex with cigpday

## Systolic Blood Pressure (mean of last two of three measurements) (mmHg)  
lm1 <- lm(cigpday ~ age+sex, data=smoke\_vs2)  
lm2 <- lm(cigpday ~ age+sex+sysbp, data=smoke\_vs2)  
  
##########################################  
## confounder check: 10% rule for age ##  
#########################################  
abs((summary(lm1)$coefficients[2] - summary(lm2)$coefficients[2]) / summary(lm2)$coefficients[2])\*100  
## [1] 2.406528 -- % diff between age coefficients for lm1&lm2; not a cf  
  
#########################################  
## confounder check: 10% rule for sex ##  
#########################################  
abs((summary(lm1)$coefficients[3] - summary(lm2)$coefficients[3]) / summary(lm2)$coefficients[3])\*100  
## [1] 0.1367146 -- % diff between sex coefficients for lm1&lm2; not a cf  
  
## sysbp is not confounder for age and sex with cigpday

## Diabetic according to criteria of first exam treated or first exam with casual glucose of 200 mg/dL or more  
lm1 <- lm(cigpday ~ age+sex, data=smoke\_vs2)  
lm2 <- lm(cigpday ~ age+sex+diabetes, data=smoke\_vs2)  
  
##########################################  
## confounder check: 10% rule for age ##  
#########################################  
abs((summary(lm1)$coefficients[2] - summary(lm2)$coefficients[2]) / summary(lm2)$coefficients[2])\*100  
## [1] 1.550249 -- % diff between age coefficients for lm1&lm2; not a cf  
  
#########################################  
## confounder check: 10% rule for sex ##  
#########################################  
abs((summary(lm1)$coefficients[3] - summary(lm2)$coefficients[3]) / summary(lm2)$coefficients[3])\*100  
## [1] 0.1605627 -- % diff between sex coefficients for lm1&lm2; not a cf  
  
## diabetes is not confounder for age and sex with cigpday

## visualize data  
smoke %>%  
 mutate(sex = as.factor(sex)) %>%  
 ggplot(aes(age, cigpday, group = sex, color = sex)) +  
 geom\_smooth(method = "loess", se = F) +  
 ggtitle("Cigarettes Smoked per Day across Age by Sex")



## reduced dataset after assessing for confounders  
smoke\_vs4 = smoke %>%   
 dplyr::select(c(randid,cigpday,sex,age)) %>%   
 mutate(sex=as.factor(sex)) %>%   
 na.omit()  
  
## limited dataset to smokers only  
smoke\_vs4\_nonsmoker = smoke\_vs4 %>%   
 filter(cigpday == 0) %>%   
 group\_by(randid) %>%  
 summarize(cig\_count = sum(cigpday)) %>%   
 filter(cig\_count == 0)  
nonsmoker\_id = unique(smoke\_vs4\_nonsmoker$randid)  
smoke\_vs4\_smoker = smoke\_vs4 %>%   
 filter(!randid %in% nonsmoker\_id)  
  
####################################################################  
## LMER for smokers only ##  
## i.e. ignored those who had zero cigpday throughout entire study  
## i.e. those who did not smoke were excluded from this model)  
####################################################################  
lmer\_2 <- lmer(cigpday ~ age \* sex + (1 | randid), data = smoke\_vs4\_smoker)  
summary(lmer\_2)

## Linear mixed model fit by REML ['lmerMod']  
## Formula: cigpday ~ age \* sex + (1 | randid)  
## Data: smoke\_vs4\_smoker  
##   
## REML criterion at convergence: 28186.6  
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -3.5916 -0.4739 -0.0653 0.3832 5.9827   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## randid (Intercept) 75.99 8.717   
## Residual 41.15 6.415   
## Number of obs: 3895, groups: randid, 1584  
##   
## Fixed effects:  
## Estimate Std. Error t value  
## (Intercept) 26.99220 1.42418 18.953  
## age -0.06629 0.02622 -2.528  
## sex2 -15.59130 2.00555 -7.774  
## age:sex2 0.17357 0.03739 4.642  
##   
## Correlation of Fixed Effects:  
## (Intr) age sex2   
## age -0.971   
## sex2 -0.710 0.689   
## age:sex2 0.681 -0.701 -0.970

######################  
## results table 2 ##  
######################  
coefs2 <- data.frame(round(coef(summary(lmer\_2)), 2))  
  
## pval for lmer\_2  
coefs2$p\_value <- round(2\*(1 - pnorm(abs(coefs2$t.value))), 5)  
  
  
#LL CI  
LL\_CI\_lmer\_2 = rbind(  
round(coefs2[1, 1] - 1.96\*coefs2[1, 2], 2), ## Intercept  
round(coefs2[2, 1] - 1.96\*coefs2[2, 2], 2), ## CI age  
round(coefs2[3, 1] - 1.96\*coefs2[3, 2], 2), ## CI sex  
round(coefs2[4, 1] - 1.96\*coefs2[4, 2], 2)) ## CI age:sex  
  
## UL CI  
UL\_CI\_lmer\_2 = rbind(  
round(coefs2[1, 1] + 1.96\*coefs2[1, 2], 2), ## Intercept  
round(coefs2[2, 1] + 1.96\*coefs2[2, 2], 2), ## CI age  
round(coefs2[3, 1] + 1.96\*coefs2[3, 2], 2), ## CI sex  
round(coefs2[4, 1] + 1.96\*coefs2[4, 2], 2)) ## CI age:sex  
  
results\_2 = cbind(coefs2, LL\_CI\_lmer\_2, UL\_CI\_lmer\_2) %>%   
 dplyr::select(1, 2, 5, 6, 4)

# Question 3

## Identification of confounders b/t cursmoke and sysbp

## Participant sex  
lm1 <- lm(sysbp ~ cursmoke, data=smoke\_vs1)  
lm2 <- lm(sysbp ~ cursmoke + sex, data=smoke\_vs1)  
  
#########################################  
## confounder check: 10% rule for sex ##  
#########################################  
abs((summary(lm1)$coefficients[2] - summary(lm2)$coefficients[2]) / summary(lm2)$coefficients[2])\*100  
## [1] 3.147992 -- % diff between sex coefficients for lm1&lm2; not a cf  
  
## sex is not confounder

## Age at exam (years)  
lm1 <- lm(sysbp ~ cursmoke, data=smoke\_vs1)  
lm2 <- lm(sysbp ~ cursmoke + age, data=smoke\_vs1)  
  
#########################################  
## confounder check: 10% rule for age ##  
#########################################  
abs((summary(lm1)$coefficients[2] - summary(lm2)$coefficients[2]) / summary(lm2)$coefficients[2])\*100  
## [1] 149.7652 -- % diff between age coefficients for lm1&lm2; AGE is a CONFOUNDER  
  
## age is confounder

## Serum Total Cholesterol (mg/dL)  
lm1 <- lm(sysbp ~ cursmoke, data=smoke\_vs1)  
lm2 <- lm(sysbp ~ cursmoke + totchol, data=smoke\_vs1)  
  
#############################################  
## confounder check: 10% rule for totchol ##  
############################################  
abs((summary(lm1)$coefficients[2] - summary(lm2)$coefficients[2]) / summary(lm2)$coefficients[2])\*100  
## [1] 8.017793 -- % diff between totchol coefficients for lm1&lm2; not a cf  
  
## totchol is not confounder

## Body Mass Index, weight in kilograms/height meters squared  
lm1 <- lm(sysbp ~ cursmoke, data=smoke\_vs1)  
lm2 <- lm(sysbp ~ cursmoke + bmi, data=smoke\_vs1)  
  
#############################################  
## confounder check: 10% rule for bmi ##  
############################################  
abs((summary(lm1)$coefficients[2] - summary(lm2)$coefficients[2]) / summary(lm2)$coefficients[2])\*100  
## [1] 61.48888 -- % diff between bmi coefficients for lm1&lm2; BMI is a CONFOUNDER  
  
## bmi is confounder

## Heart rate (Ventricular rate) in beats/min  
lm1 <- lm(sysbp ~ cursmoke, data=smoke\_vs1)  
lm2 <- lm(sysbp ~ cursmoke + heartrte, data=smoke\_vs1)  
  
##############################################  
## confounder check: 10% rule for heartrte ##  
##############################################  
abs((summary(lm1)$coefficients[2] - summary(lm2)$coefficients[2]) / summary(lm2)$coefficients[2])\*100  
## [1] 7.409282 -- % diff between heartrte coefficients for lm1&lm2; not a cf  
  
## heartrte is not confounder

## Attained Education  
lm1 <- lm(sysbp ~ cursmoke, data=smoke\_vs1)  
lm2 <- lm(sysbp ~ cursmoke + educ, data=smoke\_vs1)  
  
##############################################  
## confounder check: 10% rule for educ ##  
##############################################  
abs((summary(lm1)$coefficients[2] - summary(lm2)$coefficients[2]) / summary(lm2)$coefficients[2])\*100  
## [1] 1.421427 -- % diff between educ coefficients for lm1&lm2; not a cf  
  
## educ is not confounder

## Diabetic according to criteria of first exam treated or first exam with casual glucose of 200 mg/dL or more  
lm1 <- lm(sysbp ~ cursmoke, data=smoke\_vs1)  
lm2 <- lm(sysbp ~ cursmoke + diabetes, data=smoke\_vs1)  
  
##############################################  
## confounder check: 10% rule for diabetes ##  
##############################################  
abs((summary(lm1)$coefficients[2] - summary(lm2)$coefficients[2]) / summary(lm2)$coefficients[2])\*100  
## [1] 3.723934 -- % diff between diabetes coefficients for lm1&lm2; not a cf  
  
## diabetes is not confounder

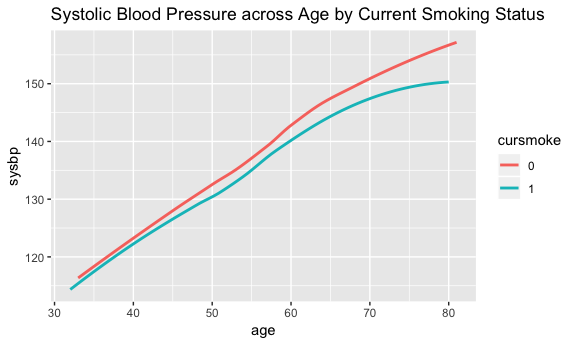
## Prevalent Angina Pectoris at exam  
lm1 <- lm(sysbp ~ cursmoke, data=smoke\_vs1)  
lm2 <- lm(sysbp ~ cursmoke + prevap, data=smoke\_vs1)  
  
##############################################  
## confounder check: 10% rule for angina ##  
##############################################  
abs((summary(lm1)$coefficients[2] - summary(lm2)$coefficients[2]) / summary(lm2)$coefficients[2])\*100  
## [1] 2.583982 -- % diff between angina coefficients for lm1&lm2; not a cf  
  
## prevap is not confounder

## Prevalent Coronary Heart Disease defined as pre-existing Angina Pectoris, Myocardial Infarction (hospitalized, silent or unrecognized), or Coronary Insufficiency (unstable angina)  
lm1 <- lm(sysbp ~ cursmoke, data=smoke\_vs1)  
lm2 <- lm(sysbp ~ cursmoke + prevchd, data=smoke\_vs1)  
  
##############################################  
## confounder check: 10% rule for chd ##  
##############################################  
abs((summary(lm1)$coefficients[2] - summary(lm2)$coefficients[2]) / summary(lm2)$coefficients[2])\*100  
## [1] 1.307933 -- % diff between chd coefficients for lm1&lm2; not a cf  
  
## chd not confounder

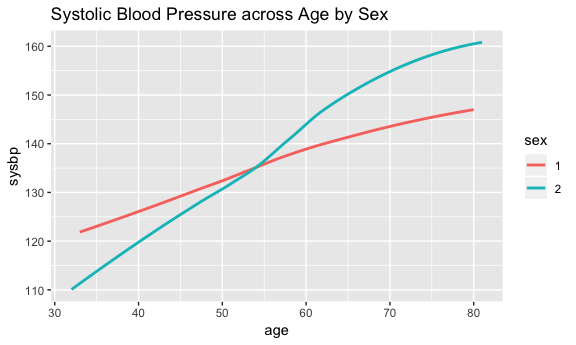
## Prevalent Myocardial Infarction  
lm1 <- lm(sysbp ~ cursmoke, data=smoke\_vs1)  
lm2 <- lm(sysbp ~ cursmoke + prevmi, data=smoke\_vs1)  
  
##############################################  
## confounder check: 10% rule for prevmi ##  
##############################################  
abs((summary(lm1)$coefficients[2] - summary(lm2)$coefficients[2]) / summary(lm2)$coefficients[2])\*100  
## [1] 0.7243338 -- % diff between prevmi coefficients for lm1&lm2; not a cf  
  
## prevmi not confounder

## Prevalent Stroke  
lm1 <- lm(sysbp ~ cursmoke, data=smoke\_vs1)  
lm2 <- lm(sysbp ~ cursmoke + prevstrk, data=smoke\_vs1)  
  
##############################################  
## confounder check: 10% rule for prevstrk ##  
##############################################  
abs((summary(lm1)$coefficients[2] - summary(lm2)$coefficients[2]) / summary(lm2)$coefficients[2])\*100  
## [1] 1.192453 -- % diff between prevstrk coefficients for lm1&lm2; not a cf  
  
## prevstrk not confounder

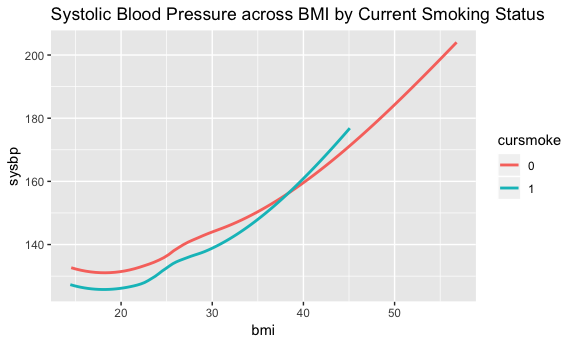
##########################  
## data prep for lmer\_3 ##  
##########################  
smoke\_vs5 = smoke %>%   
 dplyr::select(c(randid,cursmoke,sex,age,bmi,sysbp)) %>%   
 mutate(cursmoke=as.factor(cursmoke),   
 sex=as.factor(sex)) %>%   
 na.omit()  
  
####################  
## visualization ##  
####################  
smoke\_vs5 %>%  
 ggplot(aes(age, sysbp, group = cursmoke, color = cursmoke)) +  
 geom\_smooth(method = "loess", se = F) +  
 ggtitle("Systolic Blood Pressure across Age by Current Smoking Status")



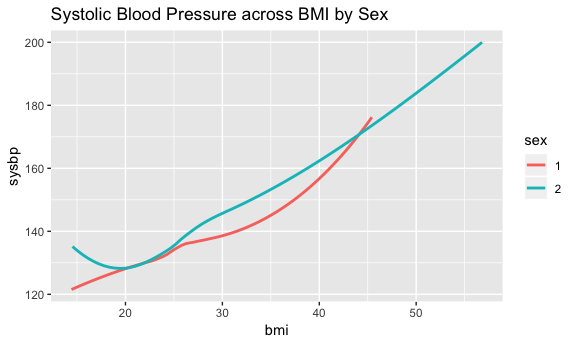
## Among both non-smokers and smokers, there appears to be a generally linear relationship between systolic bp and age.  
## No interaction  
  
smoke\_vs5 %>%   
 ggplot(aes(age, sysbp, group = sex, color = sex)) +  
 geom\_smooth(method = "loess", se = F) +  
 ggtitle("Systolic Blood Pressure across Age by Sex")



## Among both males and females, there appears to be a generally linear relationship between systolic bp and age.  
## In addition, there appears to be some interaction between age and sex becasue the two lines are not parallel (i.e. intersect).  
## age:sex -- not including bc not of interest  
  
smoke\_vs5 %>%   
 ggplot(aes(bmi, sysbp, group = cursmoke, color = cursmoke)) +  
 geom\_smooth(method = "loess", se = F) +  
 ggtitle("Systolic Blood Pressure across BMI by Current Smoking Status")



## Among both non-smokers and smokers, there appears to be a generally linear relationship between systolic bp and age.  
## In addition, there appears to be some interaction between bmi and cursmoke becasue the two lines are not parallel (i.e. intersect).  
## bmi:cursmoke  
  
smoke\_vs5 %>%   
 ggplot(aes(bmi, sysbp, group = sex, color = sex)) +  
 geom\_smooth(method = "loess", se = F) +  
 ggtitle("Systolic Blood Pressure across BMI by Sex")



## Among both males and females, there appears to be a generally linear relationship between systolic bp and bmi.  
## In addition, there appears to be some interaction between bmi and sex becasue the two lines are not parallel (i.e. intersect).  
## bmi:sex -- not including bc not of interest

## model lmer\_3  
lmer\_3 <- lmer(sysbp ~ cursmoke + age + sex + bmi\*cursmoke + (1|randid), data = smoke\_vs5)  
summary(lmer\_3)

## Linear mixed model fit by REML ['lmerMod']  
## Formula: sysbp ~ cursmoke + age + sex + bmi \* cursmoke + (1 | randid)  
## Data: smoke\_vs5  
##   
## REML criterion at convergence: 98640.2  
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -3.2526 -0.5330 -0.0505 0.4616 6.0695   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## randid (Intercept) 261.1 16.16   
## Residual 156.4 12.50   
## Number of obs: 11575, groups: randid, 4420  
##   
## Fixed effects:  
## Estimate Std. Error t value  
## (Intercept) 51.64809 2.22775 23.184  
## cursmoke1 -1.29947 2.55469 -0.509  
## age 0.85502 0.02058 41.537  
## sex2 2.61821 0.55285 4.736  
## bmi 1.42326 0.06937 20.516  
## cursmoke1:bmi 0.05004 0.09758 0.513  
##   
## Correlation of Fixed Effects:  
## (Intr) crsmk1 age sex2 bmi   
## cursmoke1 -0.511   
## age -0.519 0.049   
## sex2 -0.187 -0.009 0.010   
## bmi -0.823 0.570 -0.012 0.039   
## cursmok1:bm 0.464 -0.986 -0.008 0.031 -0.559

## added interaction term for bmi:cursmoke bc of plot and bc cursmoke is of interest  
  
  
######################  
## results table 3 ##  
######################  
coefs3 <- data.frame(round(coef(summary(lmer\_3)), 2))  
  
## pvalues for lmer\_3  
coefs3$p\_value <- round(2 \* (1 - pnorm(abs(coefs3$t.value))),5)  
  
#LL CI  
LL\_CI\_lmer\_3 = rbind(  
round(coefs3[1, 1] - 1.96\*coefs3[1, 2], 2), ## Intercept  
round(coefs3[2, 1] - 1.96\*coefs3[2, 2], 2), ## CI cursmoke  
round(coefs3[3, 1] - 1.96\*coefs3[3, 2], 2), ## CI age  
round(coefs3[4, 1] - 1.96\*coefs3[4, 2], 2), ## CI sex  
round(coefs3[5, 1] - 1.96\*coefs3[5, 2], 2), ## CI bmi  
round(coefs3[6, 1] - 1.96\*coefs3[6, 2], 2)) ## CI cursmoke:bmi  
  
## UL CI  
UL\_CI\_lmer\_3 = rbind(  
round(coefs3[1, 1] + 1.96\*coefs3[1, 2], 2), ## Intercept  
round(coefs3[2, 1] + 1.96\*coefs3[2, 2], 2), ## CI cursmoke  
round(coefs3[3, 1] + 1.96\*coefs3[3, 2], 2), ## CI age  
round(coefs3[4, 1] + 1.96\*coefs3[4, 2], 2), ## CI sex  
round(coefs3[5, 1] + 1.96\*coefs3[5, 2], 2), ## CI bmi  
round(coefs3[6, 1] + 1.96\*coefs3[6, 2], 2)) ## CI cursmoke:bmi  
  
results\_3 = cbind(coefs3, LL\_CI\_lmer\_3, UL\_CI\_lmer\_3) %>%   
 dplyr::select(1, 2, 5, 6, 4)

# Question 4

## Identification of confounders b/t cursmoke and diasbp

## Sex  
lm1 <- lm(diabp ~ cursmoke, data=smoke\_vs1)  
lm2 <- lm(diabp ~ cursmoke + sex, data=smoke\_vs1)  
  
##############################################  
## confounder check: 10% rule for sex ##  
##############################################  
abs((summary(lm1)$coefficients[2] - summary(lm2)$coefficients[2]) / summary(lm2)$coefficients[2])\*100  
## [1] 11.34421 -- % diff between sex coefficients for lm1&lm2; SEX is a CONFOUNDER

## Age  
lm1 <- lm(diabp ~ cursmoke, data=smoke\_vs1)  
lm2 <- lm(diabp ~ cursmoke + age, data=smoke\_vs1)  
  
##############################################  
## confounder check: 10% rule for age ##  
##############################################  
abs((summary(lm1)$coefficients[2] - summary(lm2)$coefficients[2]) / summary(lm2)$coefficients[2])\*100  
## [1] 59.05156 -- % diff between age coefficients for lm1&lm2; AGE is a CONFOUNDER

## Total Cholesterol  
lm1 <- lm(diabp ~ cursmoke, data=smoke\_vs1)  
lm2 <- lm(diabp ~ cursmoke + totchol, data=smoke\_vs1)  
  
##############################################  
## confounder check: 10% rule for totchol ##  
##############################################  
abs((summary(lm1)$coefficients[2] - summary(lm2)$coefficients[2]) / summary(lm2)$coefficients[2])\*100  
## [1] 8.297663 -- % diff between totchol coefficients for lm1&lm2; not a cf

## BMI  
lm1 <- lm(diabp ~ cursmoke, data=smoke\_vs1)  
lm2 <- lm(diabp ~ cursmoke + bmi, data=smoke\_vs1)  
  
##############################################  
## confounder check: 10% rule for bmi ##  
##############################################  
abs((summary(lm1)$coefficients[2] - summary(lm2)$coefficients[2]) / summary(lm2)$coefficients[2])\*100  
## [1] 126.3913 -- % diff between bmi coefficients for lm1&lm2; BMI is a CONFOUNDER

## Heartrate  
lm1 <- lm(diabp ~ cursmoke, data=smoke\_vs1)  
lm2 <- lm(diabp ~ cursmoke + heartrte, data=smoke\_vs1)  
  
#############################################  
## confounder check: 10% rule for heartrate ##  
##############################################  
abs((summary(lm1)$coefficients[2] - summary(lm2)$coefficients[2]) / summary(lm2)$coefficients[2])\*100  
## [1] 9.147251 -- % diff between heartrate coefficients for lm1&lm2; not a cf

## educ  
lm1 <- lm(diabp ~ cursmoke, data=smoke\_vs1)  
lm2 <- lm(diabp ~ cursmoke + educ, data=smoke\_vs1)  
  
#############################################  
## confounder check: 10% rule for educ ##  
##############################################  
abs((summary(lm1)$coefficients[2] - summary(lm2)$coefficients[2]) / summary(lm2)$coefficients[2])\*100  
## [1] 0.7884302 -- % diff between educ coefficients for lm1&lm2; not a cf

## diabetes  
lm1 <- lm(diabp ~ cursmoke, data=smoke\_vs1)  
lm2 <- lm(diabp ~ cursmoke + diabetes, data=smoke\_vs1)  
  
#############################################  
## confounder check: 10% rule for diabetes ##  
##############################################  
abs((summary(lm1)$coefficients[2] - summary(lm2)$coefficients[2]) / summary(lm2)$coefficients[2])\*100  
## [1] 1.952694 -- % diff between diabetes coefficients for lm1&lm2; not a cf

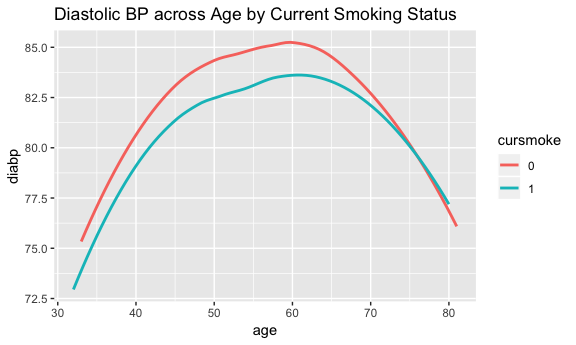
## Angina  
lm1 <- lm(diabp ~ cursmoke, data=smoke\_vs1)  
lm2 <- lm(diabp ~ cursmoke + prevap, data=smoke\_vs1)  
  
#############################################  
## confounder check: 10% rule for prevap ##  
##############################################  
abs((summary(lm1)$coefficients[2] - summary(lm2)$coefficients[2]) / summary(lm2)$coefficients[2])\*100  
## [1] 2.258716 -- % diff between prevap coefficients for lm1&lm2; not a cf

## CHD  
lm1 <- lm(diabp ~ cursmoke, data=smoke\_vs1)  
lm2 <- lm(diabp ~ cursmoke + prevchd, data=smoke\_vs1)  
  
#############################################  
## confounder check: 10% rule for chd ##  
##############################################  
abs((summary(lm1)$coefficients[2] - summary(lm2)$coefficients[2]) / summary(lm2)$coefficients[2])\*100  
## [1] 0.964231 -- % diff between chd coefficients for lm1&lm2; not a cf

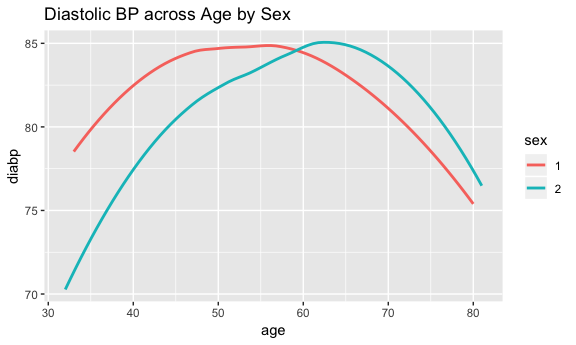
## Prevalent Myocardial Infarction  
lm1 <- lm(diabp ~ cursmoke, data=smoke\_vs1)  
lm2 <- lm(diabp ~ cursmoke + prevmi, data=smoke\_vs1)  
  
  
#############################################  
## confounder check: 10% rule for prevmi ##  
##############################################  
abs((summary(lm1)$coefficients[2] - summary(lm2)$coefficients[2]) / summary(lm2)$coefficients[2])\*100  
## [1] 0.3524298 -- % diff between prevmi coefficients for lm1&lm2; not a cf

## Prevalent Stroke  
lm1 <- lm(diabp ~ cursmoke, data=smoke\_vs1)  
lm2 <- lm(diabp ~ cursmoke + prevstrk, data=smoke\_vs1)  
  
#############################################  
## confounder check: 10% rule for stroke ##  
##############################################  
abs((summary(lm1)$coefficients[2] - summary(lm2)$coefficients[2]) / summary(lm2)$coefficients[2])\*100  
## [1] 0.9934274 -- % diff between stroke coefficients for lm1&lm2; not a cf

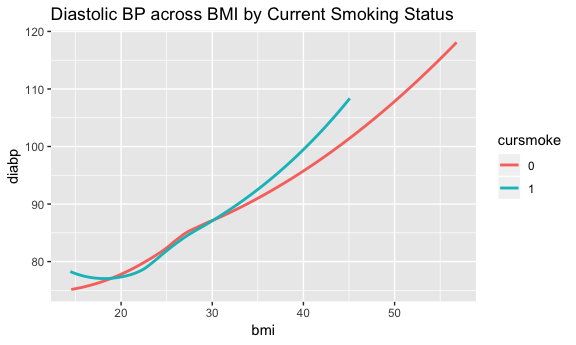
## data prep  
smoke\_vs6 = smoke %>%   
 dplyr::select(c(randid, cursmoke, sex, age, bmi, diabp)) %>%   
 mutate(sex=as.factor(sex),  
 cursmoke=as.factor(cursmoke)) %>%   
 na.omit()  
  
############################################  
## visualization & check for interaction ##  
############################################  
smoke\_vs6 %>%   
 ggplot(aes(age, diabp, group = cursmoke, color = cursmoke)) +  
 geom\_smooth(method = "loess", se = F) +  
 ggtitle("Diastolic BP across Age by Current Smoking Status")



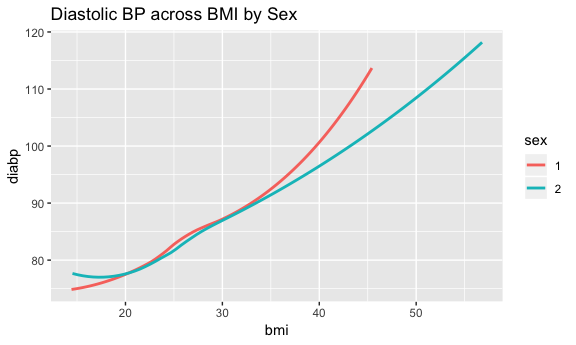
## Among both non-smokers and smokers, there appears to be a curvilinear relationship between Diastolic BP and age.  
## No interaction  
  
smoke\_vs6 %>%   
 ggplot(aes(age, diabp, group = sex, color = sex)) +  
 geom\_smooth(method = "loess", se = F) +  
 ggtitle("Diastolic BP across Age by Sex")



## Among both males (1 = male; red) and females (2 = females; blue-green), there appears to be a curvilinear relationship between Diastolic BP and age.  
## In addition, there appears to be some interaction between age and sex because the two plots intersect.  
## age:sex  
  
smoke\_vs6 %>%   
 ggplot(aes(bmi, diabp, group = cursmoke, color = cursmoke)) +  
 geom\_smooth(method = "loess", se = F) +  
 ggtitle("Diastolic BP across BMI by Current Smoking Status")



## Among both non-smokers and smokers, there appears to be a generally linear relationship between Diastolic BP and bmi.  
## In addition, there appears to be some interaction between bmi and current smoking status.  
## bmi:cursmoke  
  
smoke\_vs6 %>%   
 ggplot(aes(bmi, diabp, group = sex, color = sex)) +  
 geom\_smooth(method = "loess", se = F) +  
 ggtitle("Diastolic BP across BMI by Sex")



## Among both males and females, there appears to be a curvilinear relationship between diastolic and bmi.  
## In addition, there appears to be some interaction between bmi and sex.  
## bmi:sex

## lmer  
lmer\_4 <- lmer(diabp ~ cursmoke\*age + sex + bmi + (1|randid), data = smoke\_vs6)  
summary(lmer\_4)

## Linear mixed model fit by REML ['lmerMod']  
## Formula: diabp ~ cursmoke \* age + sex + bmi + (1 | randid)  
## Data: smoke\_vs6  
##   
## REML criterion at convergence: 85393.1  
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -5.6246 -0.5402 -0.0196 0.5121 4.5793   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## randid (Intercept) 69.18 8.318   
## Residual 53.76 7.332   
## Number of obs: 11575, groups: randid, 4420  
##   
## Fixed effects:  
## Estimate Std. Error t value  
## (Intercept) 60.14237 1.19720 50.236  
## cursmoke1 -8.81416 1.17670 -7.491  
## age -0.05110 0.01442 -3.544  
## sex2 -0.53287 0.29288 -1.819  
## bmi 1.02005 0.03139 32.496  
## cursmoke1:age 0.16118 0.02129 7.572  
##   
## Correlation of Fixed Effects:  
## (Intr) crsmk1 age sex2 bmi   
## cursmoke1 -0.475   
## age -0.682 0.623   
## sex2 -0.194 0.004 -0.005   
## bmi -0.697 0.044 -0.006 0.072   
## cursmoke1:g 0.417 -0.979 -0.596 0.023 -0.016

################  
## results\_4 ##  
################  
coefs4 <- data.frame(round(coef(summary(lmer\_4)), 2))  
  
## pvalues for lmer\_4  
coefs4$p\_value <- round(2 \* (1 - pnorm(abs(coefs4$t.value))),5); coefs4

## Estimate Std..Error t.value p\_value  
## (Intercept) 60.14 1.20 50.24 0.00000  
## cursmoke1 -8.81 1.18 -7.49 0.00000  
## age -0.05 0.01 -3.54 0.00040  
## sex2 -0.53 0.29 -1.82 0.06876  
## bmi 1.02 0.03 32.50 0.00000  
## cursmoke1:age 0.16 0.02 7.57 0.00000

#LL CI  
LL\_CI\_lmer\_4 = rbind(  
round(coefs4[1, 1] - 1.96\*coefs4[1, 2], 2), ## Intercept  
round(coefs4[2, 1] - 1.96\*coefs4[2, 2], 2), ## CI cursmoke  
round(coefs4[3, 1] - 1.96\*coefs4[3, 2], 2), ## CI age  
round(coefs4[4, 1] - 1.96\*coefs4[4, 2], 2), ## CI sex  
round(coefs4[5, 1] - 1.96\*coefs4[5, 2], 2), ## CI bmi  
round(coefs4[6, 1] - 1.96\*coefs4[6, 2], 2)) ## CI cursmoke:age   
  
  
## UL CI  
UL\_CI\_lmer\_4 = rbind(  
round(coefs4[1, 1] + 1.96\*coefs4[1, 2], 2), ## Intercept  
round(coefs4[2, 1] + 1.96\*coefs4[2, 2], 2), ## CI cursmoke  
round(coefs4[3, 1] + 1.96\*coefs4[3, 2], 2), ## CI age  
round(coefs4[4, 1] + 1.96\*coefs4[4, 2], 2), ## CI sex  
round(coefs4[5, 1] + 1.96\*coefs4[5, 2], 2), ## CI bmi  
round(coefs4[6, 1] + 1.96\*coefs4[6, 2], 2)) ## CI cursmoke:age  
  
results\_4 = cbind(coefs4, LL\_CI\_lmer\_4, UL\_CI\_lmer\_4) %>%   
 dplyr::select(1, 2, 5, 6, 4)

# Question 5

## Identification of confounders

## Sex  
lm1 <- lm(totchol ~ cursmoke, data=smoke\_vs1)  
lm2 <- lm(totchol ~ cursmoke + sex, data=smoke\_vs1)  
  
#############################################  
## confounder check: 10% rule for sex ##  
##############################################  
abs((summary(lm1)$coefficients[2] - summary(lm2)$coefficients[2]) / summary(lm2)$coefficients[2])\*100  
## [1] 33.27236 -- % diff between sex coefficients for lm1&lm2; SEX is a CONFOUNDER

## Age  
lm1 <- lm(totchol ~ cursmoke, data=smoke\_vs1)  
lm2 <- lm(totchol ~ cursmoke + age, data=smoke\_vs1)  
  
#############################################  
## confounder check: 10% rule for age ##  
##############################################  
abs((summary(lm1)$coefficients[2] - summary(lm2)$coefficients[2]) / summary(lm2)$coefficients[2])\*100  
## [1] 1512.344 -- % diff between age coefficients for lm1&lm2; AGE is a CONFOUNDER

## diastolic bp  
lm1 <- lm(totchol ~ cursmoke, data=smoke\_vs1)  
lm2 <- lm(totchol ~ cursmoke + diabp, data=smoke\_vs1)  
  
#############################################  
## confounder check: 10% rule for diabp ##  
##############################################  
abs((summary(lm1)$coefficients[2] - summary(lm2)$coefficients[2]) / summary(lm2)$coefficients[2])\*100  
## [1] 54.94285 -- % diff between diabp coefficients for lm1&lm2; DIABP is a CONFOUNDER

## BMI  
lm1 <- lm(totchol ~ cursmoke, data=smoke\_vs1)  
lm2 <- lm(totchol ~ cursmoke + bmi, data=smoke\_vs1)  
  
#############################################  
## confounder check: 10% rule for bmi ##  
##############################################  
abs((summary(lm1)$coefficients[2] - summary(lm2)$coefficients[2]) / summary(lm2)$coefficients[2])\*100  
## [1] 61.62978 -- % diff between bmi coefficients for lm1&lm2; BMI is a CONFOUNDER

## Heartrate  
lm1 <- lm(totchol ~ cursmoke, data=smoke\_vs1)  
lm2 <- lm(totchol ~ cursmoke + heartrte, data=smoke\_vs1)  
  
#############################################  
## confounder check: 10% rule for heartrate ##  
##############################################  
abs((summary(lm1)$coefficients[2] - summary(lm2)$coefficients[2]) / summary(lm2)$coefficients[2])\*100  
## [1] 9.654717 -- % diff between heartrate coefficients for lm1&lm2; not a cf

## educ  
lm1 <- lm(totchol ~ cursmoke, data=smoke\_vs1)  
lm2 <- lm(totchol ~ cursmoke + educ, data=smoke\_vs1)  
  
#############################################  
## confounder check: 10% rule for educ ##  
##############################################  
abs((summary(lm1)$coefficients[2] - summary(lm2)$coefficients[2]) / summary(lm2)$coefficients[2])\*100  
## [1] 0.534969 -- % diff between educ coefficients for lm1&lm2; not a cf

## sysbp  
lm1 <- lm(totchol ~ cursmoke, data=smoke\_vs1)  
lm2 <- lm(totchol ~ cursmoke + sysbp, data=smoke\_vs1)  
  
#############################################  
## confounder check: 10% rule for sysbp ##  
##############################################  
abs((summary(lm1)$coefficients[2] - summary(lm2)$coefficients[2]) / summary(lm2)$coefficients[2])\*100  
## [1] 114.522 -- % diff between sysbp coefficients for lm1&lm2; SYSBP is a CONFOUNDER

## diabetes  
lm1 <- lm(totchol ~ cursmoke, data=smoke\_vs1)  
lm2 <- lm(totchol ~ cursmoke + diabetes, data=smoke\_vs1)  
  
#############################################  
## confounder check: 10% rule for diabetes ##  
##############################################  
abs((summary(lm1)$coefficients[2] - summary(lm2)$coefficients[2]) / summary(lm2)$coefficients[2])\*100  
## [1] 3.847935 -- % diff between diabetes coefficients for lm1&lm2; not a cf

## prevap  
lm1 <- lm(totchol ~ cursmoke, data=smoke\_vs1)  
lm2 <- lm(totchol ~ cursmoke + prevap, data=smoke\_vs1)  
  
#############################################  
## confounder check: 10% rule for prevap ##  
##############################################  
abs((summary(lm1)$coefficients[2] - summary(lm2)$coefficients[2]) / summary(lm2)$coefficients[2])\*100  
## [1] 1.778326 -- % diff between prevap coefficients for lm1&lm2; not a cf

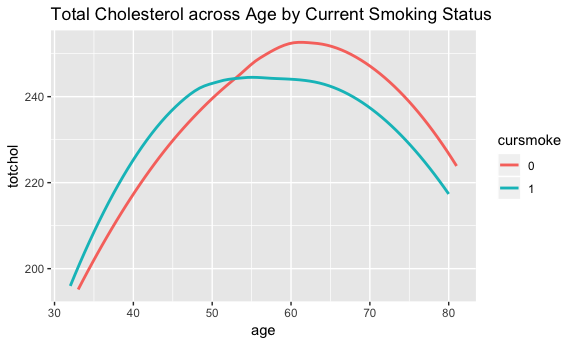
## prevchd  
lm1 <- lm(totchol ~ cursmoke, data=smoke\_vs1)  
lm2 <- lm(totchol ~ cursmoke + prevchd, data=smoke\_vs1)  
  
#############################################  
## confounder check: 10% rule for prevchd##  
##############################################  
abs((summary(lm1)$coefficients[2] - summary(lm2)$coefficients[2]) / summary(lm2)$coefficients[2])\*100  
## [1] 0.828466 -- % diff between prevchd coefficients for lm1&lm2; not a cf

## prevmi  
lm1 <- lm(totchol ~ cursmoke, data=smoke\_vs1)  
lm2 <- lm(totchol ~ cursmoke + prevmi, data=smoke\_vs1)  
  
#############################################  
## confounder check: 10% rule for prevmi ##  
##############################################  
abs((summary(lm1)$coefficients[2] - summary(lm2)$coefficients[2]) / summary(lm2)$coefficients[2])\*100  
## [1] 0.6781266 -- % diff between prevmi coefficients for lm1&lm2; not a cf

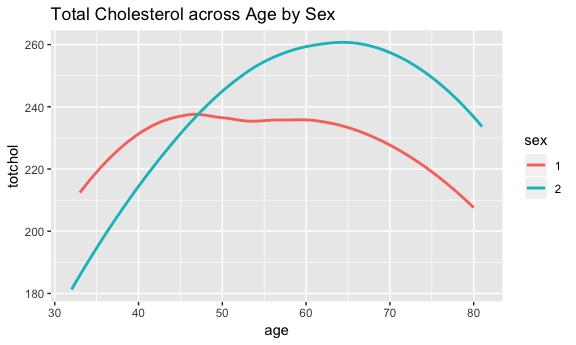
## stroke  
lm1 <- lm(totchol ~ cursmoke, data=smoke\_vs1)  
lm2 <- lm(totchol ~ cursmoke + prevstrk, data=smoke\_vs1)  
  
#############################################  
## confounder check: 10% rule for prevstroke ##  
##############################################  
abs((summary(lm1)$coefficients[2] - summary(lm2)$coefficients[2]) / summary(lm2)$coefficients[2])\*100  
## [1] 0.07024379 -- % diff between prevstroke coefficients for lm1&lm2; not a cf

## prevhyp  
lm1 <- lm(totchol ~ cursmoke, data=smoke\_vs1)  
lm2 <- lm(totchol ~ cursmoke + prevhyp, data=smoke\_vs1)  
  
  
#############################################  
## confounder check: 10% rule for prevhyp ##  
##############################################  
abs((summary(lm1)$coefficients[2] - summary(lm2)$coefficients[2]) / summary(lm2)$coefficients[2])\*100  
## [1] 53.84925 -- % diff between prevhyp coefficients for lm1&lm2; PREVHYP is a CONFOUNDER

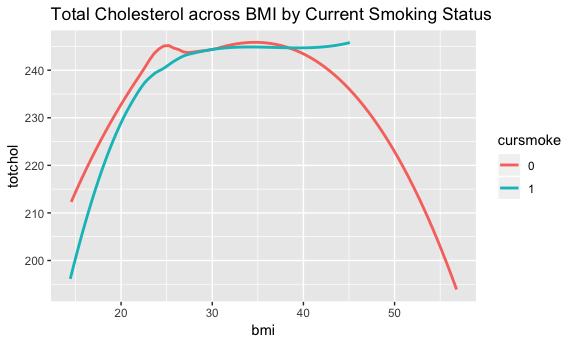
## data prep for modeling  
smoke\_vs7 = smoke %>%   
 dplyr::select(c(randid, totchol, cursmoke, sex, age, bmi)) %>%   
 mutate(cursmoke = as.factor(cursmoke),   
 sex = as.factor(sex)) %>%   
 na.omit()  
  
####################  
## visualization ##  
####################  
smoke\_vs7 %>%   
 ggplot(aes(age, totchol, group = cursmoke, color = cursmoke)) +  
 geom\_smooth(method = "loess", se = F) +  
 ggtitle("Total Cholesterol across Age by Current Smoking Status")



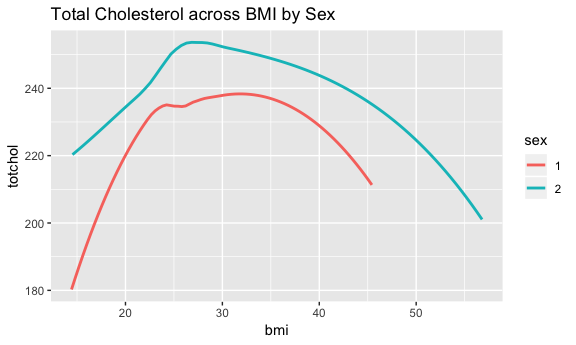
## Among both non-smokers and smokers, there appears to be a curvilinear relationship between total cholesterol and age.  
## In addition, there appears to be some interaction between age and current smoking status becasue the two lines intersect(?).  
## cursmoke:age  
  
smoke\_vs7 %>%   
 ggplot(aes(age, totchol, group = sex, color = sex)) +  
 geom\_smooth(method = "loess", se = F) +  
 ggtitle("Total Cholesterol across Age by Sex")



## Among both males and females, there appears to be a curvilinear relationship between total cholesterol and age.  
## In addition, there appears to be some interaction between age and sex because the two plots intersect.  
## age:sex --- will not include  
  
smoke\_vs7 %>%   
 ggplot(aes(bmi, totchol, group = cursmoke, color = cursmoke)) +  
 geom\_smooth(method = "loess", se = F) +  
 ggtitle("Total Cholesterol across BMI by Current Smoking Status")



## Among both non-smokers and smokers, there appears to be a curvilinear relationship between total cholesterol and bmi.  
## In addition, there appears to be some interaction between bmi and current smoking status.  
## bmi:cursmoke  
  
smoke\_vs7 %>%   
 ggplot(aes(bmi, totchol, group = sex, color = sex)) +  
 geom\_smooth(method = "loess", se = F) +  
 ggtitle("Total Cholesterol across BMI by Sex")



## Among both males and females, there appears to be a curvilinear relationship between total cholesterol and bmi.  
## In addition, there appears to be no interaction between bmi and current smoking status.

lmer\_5 <- lmer(totchol ~ cursmoke\*bmi + sex + cursmoke\*age + (1|randid), data = smoke\_vs7)  
summary(lmer\_5)

## Linear mixed model fit by REML ['lmerMod']  
## Formula: totchol ~ cursmoke \* bmi + sex + cursmoke \* age + (1 | randid)  
## Data: smoke\_vs7  
##   
## REML criterion at convergence: 112160.6  
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -8.8193 -0.5216 -0.0140 0.4864 9.1134   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## randid (Intercept) 1305.5 36.13   
## Residual 674.1 25.96   
## Number of obs: 11173, groups: randid, 4405  
##   
## Fixed effects:  
## Estimate Std. Error t value  
## (Intercept) 176.51441 5.22462 33.785  
## cursmoke1 -29.38692 7.05633 -4.165  
## bmi 1.45643 0.15257 9.546  
## sex2 13.94202 1.22545 11.377  
## age 0.32861 0.05514 5.960  
## cursmoke1:bmi 0.74739 0.21240 3.519  
## cursmoke1:age 0.26381 0.08053 3.276  
##   
## Correlation of Fixed Effects:  
## (Intr) crsmk1 bmi sex2 age crsmk1:b  
## cursmoke1 -0.599   
## bmi -0.775 0.458   
## sex2 -0.169 -0.019 0.040   
## age -0.599 0.393 -0.002 -0.003   
## cursmok1:bm 0.434 -0.776 -0.559 0.029 -0.005   
## cursmoke1:g 0.363 -0.617 -0.016 0.022 -0.589 0.001

## MOP Q: why not include quadratic terms since the relationships with tochol are all curvilinear?  
  
#############  
## results\_5  
#############  
coefs5 <- data.frame(round(coef(summary(lmer\_5)), 2))  
coefs5$p\_value <- round(2 \* (1 - pnorm(abs(coefs5$t.value))),5)  
  
#LL CI  
LL\_CI\_lmer\_5 = rbind(  
round(coefs5[1, 1] - 1.96\*coefs5[1, 2], 2), ## Intercept  
round(coefs5[2, 1] - 1.96\*coefs5[2, 2], 2), ## CI cursmoke  
round(coefs5[3, 1] - 1.96\*coefs5[3, 2], 2), ## CI bmi  
round(coefs5[4, 1] - 1.96\*coefs5[4, 2], 2), ## CI sex  
round(coefs5[5, 1] - 1.96\*coefs5[5, 2], 2), ## CI age  
round(coefs5[6, 1] - 1.96\*coefs5[6, 2], 2), ## CI cursmoke:bmi   
round(coefs5[7, 1] - 1.96\*coefs5[7, 2], 2)) ## CI cursmoke:age   
  
## UL CI  
UL\_CI\_lmer\_5 = rbind(  
round(coefs5[1, 1] + 1.96\*coefs5[1, 2], 2), ## Intercept  
round(coefs5[2, 1] + 1.96\*coefs5[2, 2], 2), ## CI cursmoke  
round(coefs5[3, 1] + 1.96\*coefs5[3, 2], 2), ## CI bmi  
round(coefs5[4, 1] + 1.96\*coefs5[4, 2], 2), ## CI sex  
round(coefs5[5, 1] + 1.96\*coefs5[5, 2], 2), ## CI age  
round(coefs5[6, 1] + 1.96\*coefs5[6, 2], 2), ## CI cursmoke:bmi   
round(coefs5[7, 1] + 1.96\*coefs5[7, 2], 2)) ## CI cursmoke:age   
  
results\_5 = cbind(coefs5, LL\_CI\_lmer\_5, UL\_CI\_lmer\_5) %>%   
 dplyr::select(1, 2, 5, 6, 4); results\_5

## Estimate Std..Error LL\_CI\_lmer\_5 UL\_CI\_lmer\_5 p\_value  
## (Intercept) 176.51 5.22 166.28 186.74 0.00000  
## cursmoke1 -29.39 7.06 -43.23 -15.55 0.00003  
## bmi 1.46 0.15 1.17 1.75 0.00000  
## sex2 13.94 1.23 11.53 16.35 0.00000  
## age 0.33 0.06 0.21 0.45 0.00000  
## cursmoke1:bmi 0.75 0.21 0.34 1.16 0.00043  
## cursmoke1:age 0.26 0.08 0.10 0.42 0.00104

kable(results\_1)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Estimate | Std..Error | LL\_CI\_glmer\_1 | UL\_CI\_glmer\_1 | Pr…z.. |
| (Intercept) | 13.72 | 0.85 | 12.0540 | 15.3860 | 0 |
| age | -0.24 | 0.01 | -0.2596 | -0.2204 | 0 |
| sex2 | -6.93 | 1.09 | -9.0664 | -4.7936 | 0 |
| age:sex2 | 0.06 | 0.02 | 0.0208 | 0.0992 | 0 |

kable(results\_2)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Estimate | Std..Error | LL\_CI\_lmer\_2 | UL\_CI\_lmer\_2 | p\_value |
| (Intercept) | 26.99 | 1.42 | 24.21 | 29.77 | 0.00000 |
| age | -0.07 | 0.03 | -0.13 | -0.01 | 0.01141 |
| sex2 | -15.59 | 2.01 | -19.53 | -11.65 | 0.00000 |
| age:sex2 | 0.17 | 0.04 | 0.09 | 0.25 | 0.00000 |

kable(results\_3)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Estimate | Std..Error | LL\_CI\_lmer\_3 | UL\_CI\_lmer\_3 | p\_value |
| (Intercept) | 51.65 | 2.23 | 47.28 | 56.02 | 0.00000 |
| cursmoke1 | -1.30 | 2.55 | -6.30 | 3.70 | 0.61005 |
| age | 0.86 | 0.02 | 0.82 | 0.90 | 0.00000 |
| sex2 | 2.62 | 0.55 | 1.54 | 3.70 | 0.00000 |
| bmi | 1.42 | 0.07 | 1.28 | 1.56 | 0.00000 |
| cursmoke1:bmi | 0.05 | 0.10 | -0.15 | 0.25 | 0.61005 |

kable(results\_4)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Estimate | Std..Error | LL\_CI\_lmer\_4 | UL\_CI\_lmer\_4 | p\_value |
| (Intercept) | 60.14 | 1.20 | 57.79 | 62.49 | 0.00000 |
| cursmoke1 | -8.81 | 1.18 | -11.12 | -6.50 | 0.00000 |
| age | -0.05 | 0.01 | -0.07 | -0.03 | 0.00040 |
| sex2 | -0.53 | 0.29 | -1.10 | 0.04 | 0.06876 |
| bmi | 1.02 | 0.03 | 0.96 | 1.08 | 0.00000 |
| cursmoke1:age | 0.16 | 0.02 | 0.12 | 0.20 | 0.00000 |

kable(results\_5)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Estimate | Std..Error | LL\_CI\_lmer\_5 | UL\_CI\_lmer\_5 | p\_value |
| (Intercept) | 176.51 | 5.22 | 166.28 | 186.74 | 0.00000 |
| cursmoke1 | -29.39 | 7.06 | -43.23 | -15.55 | 0.00003 |
| bmi | 1.46 | 0.15 | 1.17 | 1.75 | 0.00000 |
| sex2 | 13.94 | 1.23 | 11.53 | 16.35 | 0.00000 |
| age | 0.33 | 0.06 | 0.21 | 0.45 | 0.00000 |
| cursmoke1:bmi | 0.75 | 0.21 | 0.34 | 1.16 | 0.00043 |
| cursmoke1:age | 0.26 | 0.08 | 0.10 | 0.42 | 0.00104 |