

Final_Project

Mengqi Zhu

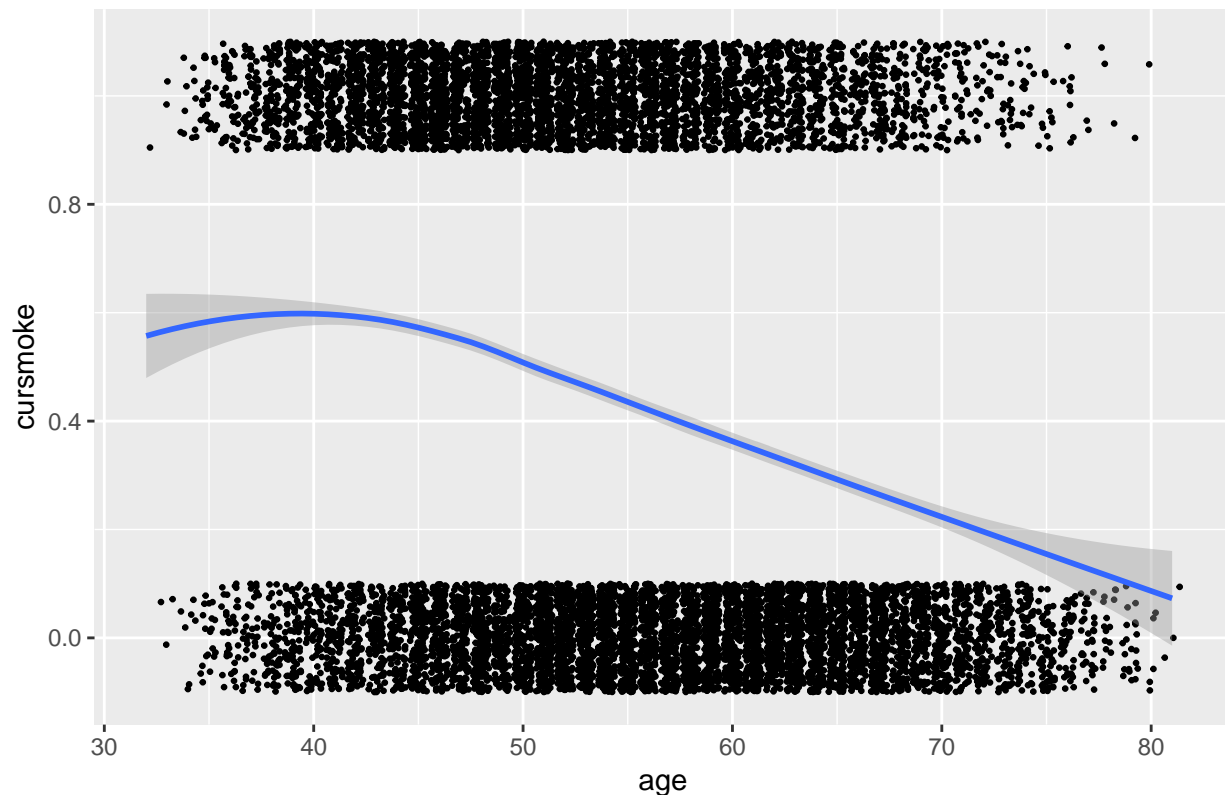
2018/11/29

```
smoke <- read.csv(file = 'frmgham2.csv') %>%  
  clean_names()
```

Figure 1 shows that as individuals age, the likelihood that they are smoking decreases. We can see that when we breaking individuals down by sex, it appears that the overall trend is the same between sexes with males having an overall higher likelihood of being smokers as age increases.

```
smoke %>%  
  ggplot(aes(age, cursmoke)) +  
  geom_jitter(height = 0.1, size = 0.5) +  
  geom_smooth(method = "loess") +  
  ggtitle("Figure 1: Current Smoking Status across Age")
```

Figure 1: Current Smoking Status across Age



```
#BY SEX  
smoke %>%  
  mutate(sex = as.factor(sex)) %>%  
  ggplot(aes(age, cursmoke, group = sex, color = sex)) +  
  geom_jitter(height = 0.1, size = 0.5) +  
  geom_smooth(method = "loess", se = F) +  
  ggtitle("Figure 2: Current Smoking Status across Age")
```

Figure 2: Current Smoking Status across Age

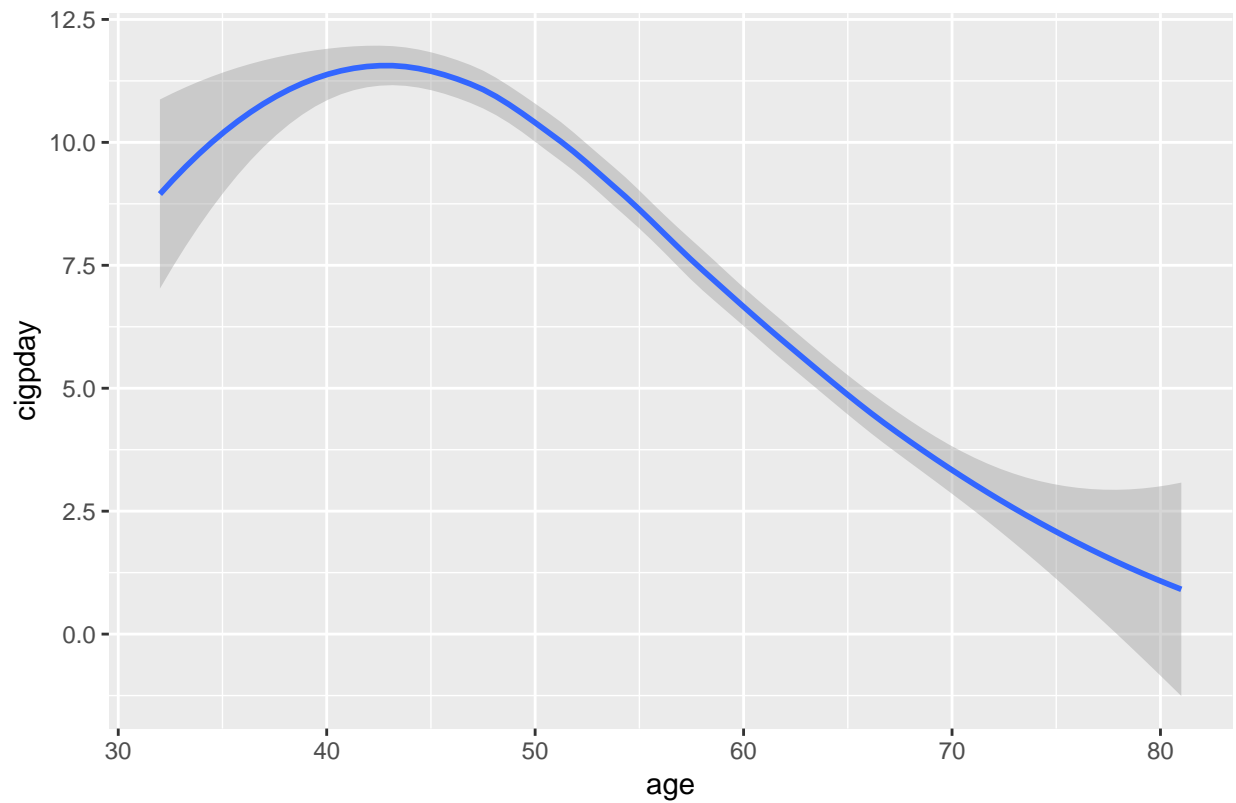


When looking at cigarette packs smoked per day, it appears that the number steadily decreases as individuals get older. The trend once again is the same in each sex however females are smoking less packs a day overall.

```
smoke %>%  
  ggplot(aes(age, cigpday)) +  
  geom_smooth(method = "loess") +  
  ggtitle("Figure 4: Cigarette per Day across Age")
```

```
## Warning: Removed 79 rows containing non-finite values (stat_smooth).
```

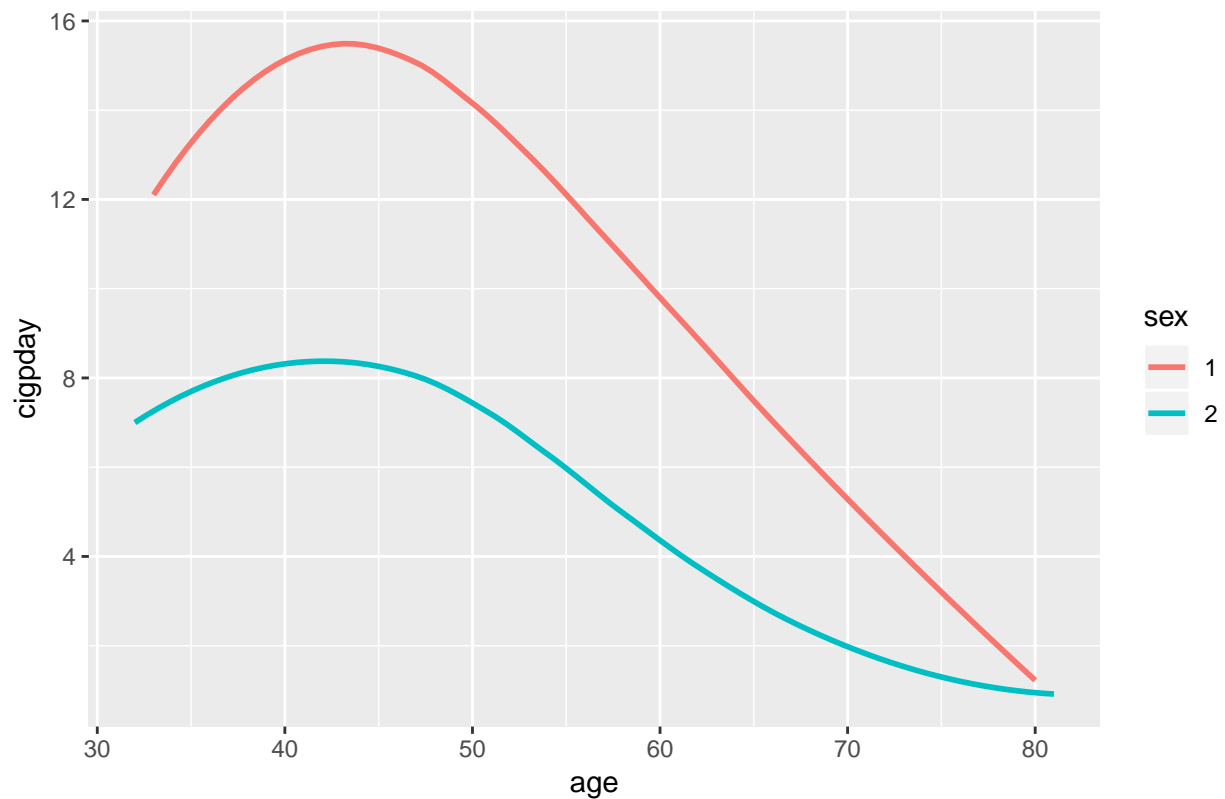
Figure 4: Cigarette per Day across Age



```
smoke %>%  
  mutate(sex = as.factor(sex)) %>%  
  ggplot(aes(age, cigpday, group = sex, color = sex)) +  
  geom_smooth(method = "loess", se = F) +  
  ggtitle("Figure 5: Cigarettes per Day across Age")
```

```
## Warning: Removed 79 rows containing non-finite values (stat_smooth).
```

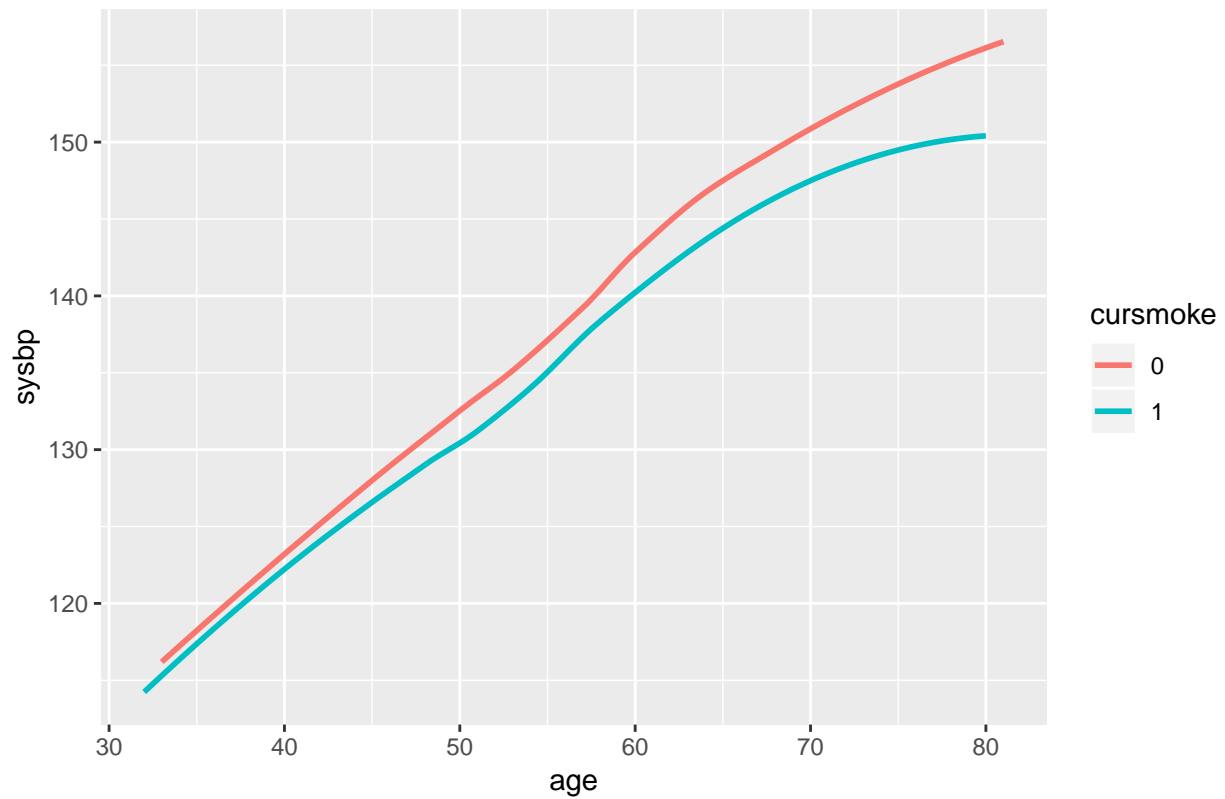
Figure 5: Cigarettes per Day across Age



In Figure 5, we see that as systolic blood pressure increases the likelihood of smoking decreases. The trend is not as profound in Figure 6 with diastolic BP or with serum total cholesterol in Figure 7.

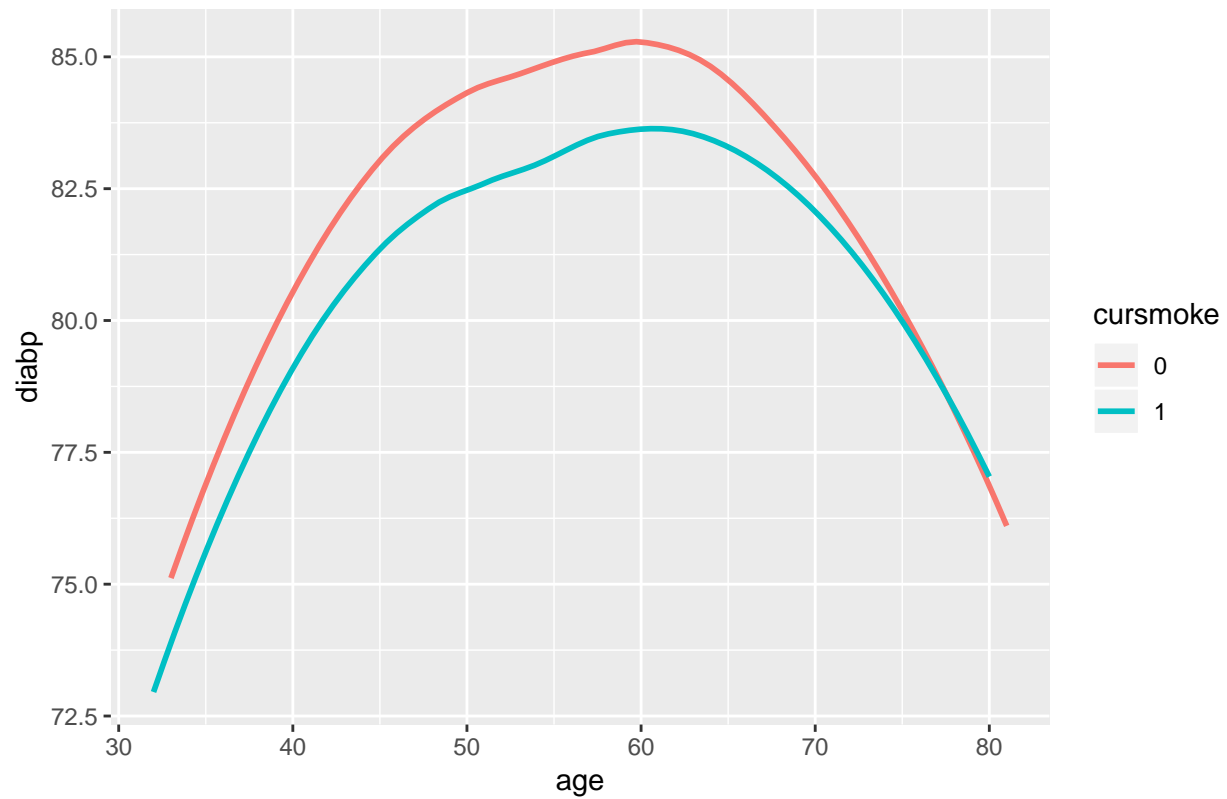
```
smoke %>%  
  mutate(cursmoke = as.factor(cursmoke)) %>%  
  ggplot(aes(age ,sysbp, group = cursmoke, color = cursmoke)) +  
  geom_smooth(method = "loess", se = F) +  
  ggtitle("Figure 5: Systolic Blood Pressure across Age")
```

Figure 5: Systolic Blood Pressure across Age



```
smoke %>%  
  mutate(cursmoke = as.factor(cursmoke)) %>%  
  ggplot(aes(age, diabp, group = cursmoke, color = cursmoke)) +  
  geom_smooth(method = "loess", se = F) +  
  ggtitle("Figure 6: Diastolic Blood Pressure across Age")
```

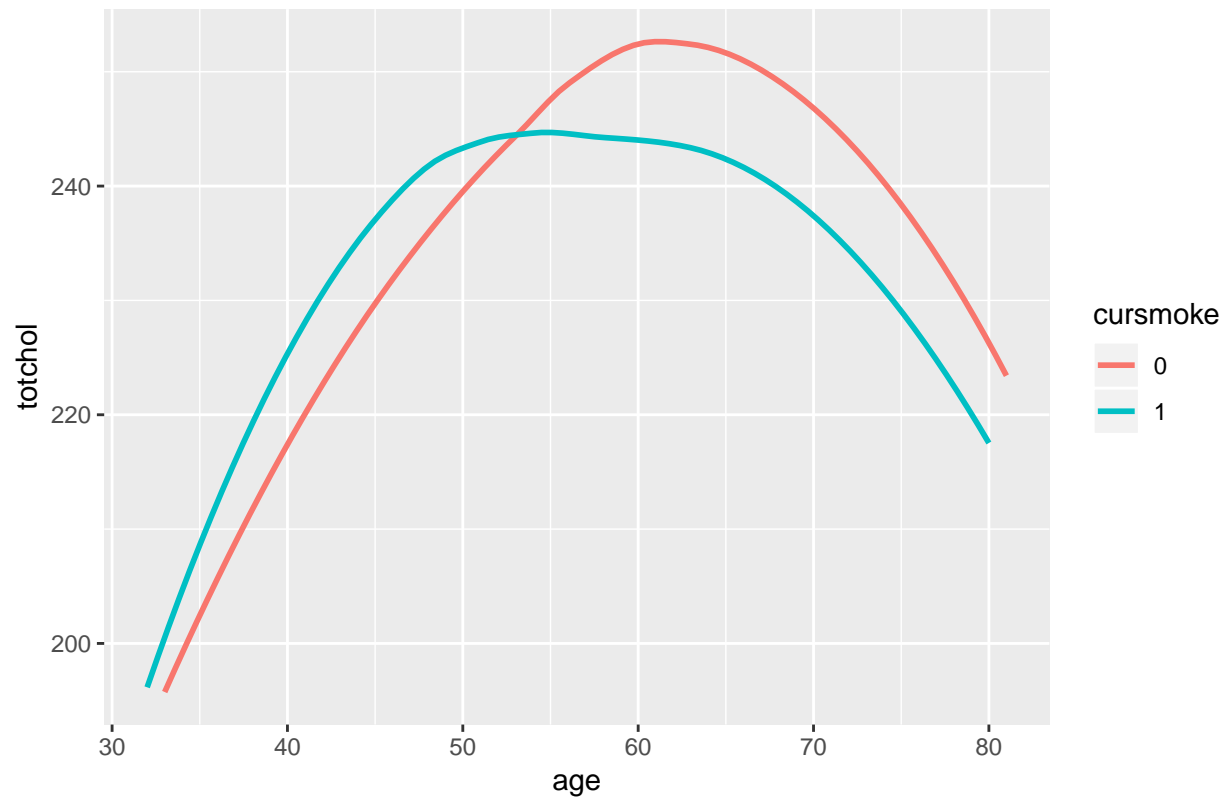
Figure 6: Diastolic Blood Pressure across Age



```
smoke %>%  
  mutate(cursmoke = as.factor(cursmoke)) %>%  
  ggplot(aes(age, totchol, cursmoke, group = cursmoke, color = cursmoke)) +  
  geom_smooth(method = "loess", se = F) +  
  ggtitle("Figure 7: Total Cholesterol across Age")
```

```
## Warning: Removed 409 rows containing non-finite values (stat_smooth).
```

Figure 7: Total Cholesterol across Age



```
#checking missing values (5% rule)
pMiss <- function(x){sum(is.na(x))/length(x)*100}
apply(smoke,2,pMiss) #2 indicates columns
```

```
##      randid      sex    totchol    age    sysbp    diabp
## 0.00000000 0.00000000 3.51767438 0.00000000 0.00000000 0.00000000
##      cursmoke    cigpday      bmi    diabetes    bpmeds    hearttrte
## 0.00000000 0.67945300 0.44723488 0.00000000 5.10019782 0.05160403
##      glucose      educ    prevchd    prevap    prevmi    prevstrk
## 12.38496603 2.53719790 0.00000000 0.00000000 0.00000000 0.00000000
##      prevhyp      time    period    hdlc    ldlc    death
## 0.00000000 0.00000000 0.00000000 73.96576933 73.97437000 0.00000000
##      angina    hospmi    mi_fchd    anychd    stroke    cvd
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##      hyperten    timeap    timemi    timemifc    timechd    timestrk
## 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
##      timecvd    timedth    timehyp
## 0.00000000 0.00000000 0.00000000
```

```
#Remove hdlc, ldlc, glucose, and bpmeds for having more than 5% of missing values.
```

It is okay as hdlc, ldlc are highly correlated with totchol, glucose is highly correlated with diabetes, bpmeds is highly correlated with sysbp and diabp. Therefore it won't lose much information to just drop these features.

Question 1

```
smoke_vs1 = smoke %>%
  filter(period==1) %>%
  dplyr::select(c(randid,sex,age,cursmoke,totchol,bmi,hearttrte,educ,diabp,sysbp,diabetes,prevap,prevchd)
  mutate(cursmoke = as.factor(cursmoke), sex=as.factor(sex), diabetes=as.factor(diabetes)) %>%
  na.omit()
```

```
glm1 <- glm(cursmoke ~ age+sex, data=smoke_vs1,family = 'binomial')
glm2 <- glm(cursmoke ~ age+sex+prevhyp, data=smoke_vs1,family = 'binomial')
a=(summary(glm1)$coefficients[2])-(1.96 * (summary(glm1)$coefficients[5]))
b= (summary(glm1)$coefficients[2])+(1.96 * (summary(glm1)$coefficients[5]))
c=summary(glm2)$coefficients[2]
exp(a)
```

```
## [1] 0.9422074
```

```
exp(b)
```

```
## [1] 0.9563489
```

```
exp(c)
```

```
## [1] 0.9529738
```

```
!(exp(c)>=exp(a) & exp(c)<=exp(b))
```

```
## [1] FALSE
```

```
d=(summary(glm1)$coefficients[3])-(1.96 * (summary(glm1)$coefficients[6]))
e=(summary(glm1)$coefficients[3])+(1.96 * (summary(glm1)$coefficients[6]))
f=summary(glm2)$coefficients[3]
exp(d)
```

```
## [1] 0.3737987
```

```
exp(e)
```

```
## [1] 0.4819127
```

```
exp(f)
```

```
## [1] 0.4226504
```

```
!(exp(f)>=exp(d) & exp(f)<=exp(e))
```

```
## [1] FALSE
```

prevhyp is not confounder for sex and age with cursmoke

```
glm1 <- glm(cursmoke ~ age+sex, data=smoke_vs1,family = 'binomial')
glm2 <- glm(cursmoke ~ age+sex+prevstrk, data=smoke_vs1,family = 'binomial')
a=(summary(glm1)$coefficients[2])-(1.96 * (summary(glm1)$coefficients[5]))
b= (summary(glm1)$coefficients[2])+(1.96 * (summary(glm1)$coefficients[5]))
c=summary(glm2)$coefficients[2]
exp(a)
```

```
## [1] 0.9422074
```

```
exp(b)
```



```
## [1] 0.9563489
```

```
exp(c)
```

```
## [1] 0.9493691
```

```
!(exp(c)>=exp(a) & exp(c)<=exp(b))
```

```
## [1] FALSE
```

```
d=(summary(glm1)$coefficients[3])-(1.96 * (summary(glm1)$coefficients[6]))  
e=(summary(glm1)$coefficients[3])+(1.96 * (summary(glm1)$coefficients[6]))  
f=summary(glm2)$coefficients[3]  
exp(d)
```

```
## [1] 0.3737987
```

```
exp(e)
```

```
## [1] 0.4819127
```

```
exp(f)
```

```
## [1] 0.4244159
```

```
!(exp(f)>=exp(d) & exp(f)<=exp(e))
```

```
## [1] FALSE
```

prevstrk is not confounder for sex and age with cursmoke

```
glm1 <- glm(cursmoke ~ age+sex, data=smoke_vs1,family = 'binomial')  
glm2 <- glm(cursmoke ~ age+sex+prevmi, data=smoke_vs1,family = 'binomial')  
a=(summary(glm1)$coefficients[2])-(1.96 * (summary(glm1)$coefficients[5]))  
b= (summary(glm1)$coefficients[2])+(1.96 * (summary(glm1)$coefficients[5]))  
c=summary(glm2)$coefficients[2]  
exp(a)
```

```
## [1] 0.9422074
```

```
exp(b)
```

```
## [1] 0.9563489
```

```
exp(c)
```

```
## [1] 0.948613
```

```
!(exp(c)>=exp(a) & exp(c)<=exp(b))
```

```
## [1] FALSE
```

```
d=(summary(glm1)$coefficients[3])-(1.96 * (summary(glm1)$coefficients[6]))  
e=(summary(glm1)$coefficients[3])+(1.96 * (summary(glm1)$coefficients[6]))  
f=summary(glm2)$coefficients[3]  
exp(d)
```

```
## [1] 0.3737987
```

```
exp(e)
```

```
## [1] 0.4819127
```

```

exp(f)

## [1] 0.4290085
!(exp(f)>=exp(d) & exp(f)<=exp(e))

## [1] FALSE
prevmi is not confounder for sex and age with cursmoke
glm1 <- glm(cursmoke ~ age+sex, data=smoke_vs1,family = 'binomial')
glm2 <- glm(cursmoke ~ age+sex+prevchd, data=smoke_vs1,family = 'binomial')
a=(summary(glm1)$coefficients[2])-(1.96 * (summary(glm1)$coefficients[5]))
b= (summary(glm1)$coefficients[2])+(1.96 * (summary(glm1)$coefficients[5]))
c=summary(glm2)$coefficients[2]
exp(a)

## [1] 0.9422074
exp(b)

## [1] 0.9563489
exp(c)

## [1] 0.9489226
!(exp(c)>=exp(a) & exp(c)<=exp(b))

## [1] FALSE
d=(summary(glm1)$coefficients[3])-(1.96 * (summary(glm1)$coefficients[6]))
e=(summary(glm1)$coefficients[3])+(1.96 * (summary(glm1)$coefficients[6]))
f=summary(glm2)$coefficients[3]
exp(d)

## [1] 0.3737987
exp(e)

## [1] 0.4819127
exp(f)

## [1] 0.4256095
!(exp(f)>=exp(d) & exp(f)<=exp(e))

## [1] FALSE
prevchd is not confounder for sex and age with cursmoke
glm1 <- glm(cursmoke ~ age+sex, data=smoke_vs1,family = 'binomial')
glm2 <- glm(cursmoke ~ age+sex+prevap, data=smoke_vs1,family = 'binomial')
a=(summary(glm1)$coefficients[2])-(1.96 * (summary(glm1)$coefficients[5]))
b= (summary(glm1)$coefficients[2])+(1.96 * (summary(glm1)$coefficients[5]))
c=summary(glm2)$coefficients[2]
exp(a)

## [1] 0.9422074

```

```

exp(b)

## [1] 0.9563489
exp(c)

## [1] 0.94959
!(exp(c)>=exp(a) & exp(c)<=exp(b))

## [1] FALSE
d=(summary(glm1)$coefficients[3])-(1.96 * (summary(glm1)$coefficients[6]))
e=(summary(glm1)$coefficients[3])+(1.96 * (summary(glm1)$coefficients[6]))
f=summary(glm2)$coefficients[3]
exp(d)

## [1] 0.3737987
exp(e)

## [1] 0.4819127
exp(f)

## [1] 0.4233239
!(exp(f)>=exp(d) & exp(f)<=exp(e))

## [1] FALSE
prevap is not confounder for sex and age with cursmoke
glm1 <- glm(cursmoke ~ age+sex, data=smoke_vs1,family = 'binomial')
glm2 <- glm(cursmoke ~ age+sex+totchol, data=smoke_vs1,family = 'binomial')
a=(summary(glm1)$coefficients[2])-(1.96 * (summary(glm1)$coefficients[5]))
b= (summary(glm1)$coefficients[2])+(1.96 * (summary(glm1)$coefficients[5]))
c=summary(glm2)$coefficients[2]
exp(a)

## [1] 0.9422074
exp(b)

## [1] 0.9563489
exp(c)

## [1] 0.9480594
!(exp(c)>=exp(a) & exp(c)<=exp(b))

## [1] FALSE
d=(summary(glm1)$coefficients[3])-(1.96 * (summary(glm1)$coefficients[6]))
e=(summary(glm1)$coefficients[3])+(1.96 * (summary(glm1)$coefficients[6]))
f=summary(glm2)$coefficients[3]
exp(d)

## [1] 0.3737987
exp(e)

```

```
## [1] 0.4819127
```

```
exp(f)
```

```
## [1] 0.4218472
```

```
!(exp(f)>=exp(d) & exp(f)<=exp(e))
```

```
## [1] FALSE
```

totchol is not confounder for sex and age with cursmoke

```
glm1 <- glm(cursmoke ~ age+sex, data=smoke_vs1,family = 'binomial')
glm2 <- glm(cursmoke ~ age+sex+sysbp, data=smoke_vs1,family = 'binomial')
a=(summary(glm1)$coefficients[2])-(1.96 * (summary(glm1)$coefficients[5]))
b= (summary(glm1)$coefficients[2])+(1.96 * (summary(glm1)$coefficients[5]))
c=summary(glm2)$coefficients[2]
exp(a)
```

```
## [1] 0.9422074
```

```
exp(b)
```

```
## [1] 0.9563489
```

```
exp(c)
```

```
## [1] 0.9541667
```

```
!(exp(c)>=exp(a) & exp(c)<=exp(b))
```

```
## [1] FALSE
```

```
d=(summary(glm1)$coefficients[3])-(1.96 * (summary(glm1)$coefficients[6]))
e=(summary(glm1)$coefficients[3])+(1.96 * (summary(glm1)$coefficients[6]))
f=summary(glm2)$coefficients[3]
exp(d)
```

```
## [1] 0.3737987
```

```
exp(e)
```

```
## [1] 0.4819127
```

```
exp(f)
```

```
## [1] 0.4275944
```

```
!(exp(f)>=exp(d) & exp(f)<=exp(e))
```

```
## [1] FALSE
```

sysbp is not confounder for sex and age with cursmoke

```
glm1 <- glm(cursmoke ~ age+sex, data=smoke_vs1,family = 'binomial')
glm2 <- glm(cursmoke ~ age+sex+diabp, data=smoke_vs1,family = 'binomial')
a=(summary(glm1)$coefficients[2])-(1.96 * (summary(glm1)$coefficients[5]))
b= (summary(glm1)$coefficients[2])+(1.96 * (summary(glm1)$coefficients[5]))
c=summary(glm2)$coefficients[2]
exp(a)
```

```
## [1] 0.9422074
```

```

exp(b)

## [1] 0.9563489
exp(c)

## [1] 0.9528767
!(exp(c)>=exp(a) & exp(c)<=exp(b))

## [1] FALSE
d=(summary(glm1)$coefficients[3])-(1.96 * (summary(glm1)$coefficients[6]))
e=(summary(glm1)$coefficients[3])+(1.96 * (summary(glm1)$coefficients[6]))
f=summary(glm2)$coefficients[3]
exp(d)

## [1] 0.3737987
exp(e)

## [1] 0.4819127
exp(f)

## [1] 0.4149985
!(exp(f)>=exp(d) & exp(f)<=exp(e))

## [1] FALSE
diabp is is not confounder for sex and age with cursmoke
glm1 <- glm(cursmoke ~ age+sex, data=smoke_vs1,family = 'binomial')
glm2 <- glm(cursmoke ~ age+sex+bmi, data=smoke_vs1,family = 'binomial')
a=(summary(glm1)$coefficients[2])-(1.96 * (summary(glm1)$coefficients[5]))
b= (summary(glm1)$coefficients[2])+(1.96 * (summary(glm1)$coefficients[5]))
c=summary(glm2)$coefficients[2]
exp(a)

## [1] 0.9422074
exp(b)

## [1] 0.9563489
exp(c)

## [1] 0.9530611
!(exp(c)>=exp(a) & exp(c)<=exp(b))

## [1] FALSE
d=(summary(glm1)$coefficients[3])-(1.96 * (summary(glm1)$coefficients[6]))
e=(summary(glm1)$coefficients[3])+(1.96 * (summary(glm1)$coefficients[6]))
f=summary(glm2)$coefficients[3]
exp(d)

## [1] 0.3737987
exp(e)

```

```
## [1] 0.4819127
exp(f)

## [1] 0.3907743
!(exp(f)>=exp(d) & exp(f)<=exp(e))

## [1] FALSE
bmi is is not confounder for sex and age with cursmoke

glm1 <- glm(cursmoke ~ age+sex, data=smoke_vs1,family = 'binomial')
glm2 <- glm(cursmoke ~ age+sex+hearttrte, data=smoke_vs1,family = 'binomial')
a=(summary(glm1)$coefficients[2])-(1.96 * (summary(glm1)$coefficients[5]))
b= (summary(glm1)$coefficients[2])+(1.96 * (summary(glm1)$coefficients[5]))
c=summary(glm2)$coefficients[2]
exp(a)

## [1] 0.9422074
exp(b)

## [1] 0.9563489
exp(c)

## [1] 0.9491648
!(exp(c)>=exp(a) & exp(c)<=exp(b))

## [1] FALSE
d=(summary(glm1)$coefficients[3])-(1.96 * (summary(glm1)$coefficients[6]))
e=(summary(glm1)$coefficients[3])+(1.96 * (summary(glm1)$coefficients[6]))
f=summary(glm2)$coefficients[3]
exp(d)

## [1] 0.3737987
exp(e)

## [1] 0.4819127
exp(f)

## [1] 0.4069115
!(exp(f)>=exp(d) & exp(f)<=exp(e))

## [1] FALSE
hearttrte is is not confounder for sex and age with cursmoke

glm1 <- glm(cursmoke ~ age+sex, data=smoke_vs1,family = 'binomial')
glm2 <- glm(cursmoke ~ age+sex+educ, data=smoke_vs1,family = 'binomial')
a=(summary(glm1)$coefficients[2])-(1.96 * (summary(glm1)$coefficients[5]))
b= (summary(glm1)$coefficients[2])+(1.96 * (summary(glm1)$coefficients[5]))
c=summary(glm2)$coefficients[2]
exp(a)

## [1] 0.9422074
```

```

exp(b)

## [1] 0.9563489
exp(c)

## [1] 0.9483645
!(exp(c)>=exp(a) & exp(c)<=exp(b))

## [1] FALSE
d=(summary(glm1)$coefficients[3])-(1.96 * (summary(glm1)$coefficients[6]))
e=(summary(glm1)$coefficients[3])+(1.96 * (summary(glm1)$coefficients[6]))
f=summary(glm2)$coefficients[3]
exp(d)

## [1] 0.3737987
exp(e)

## [1] 0.4819127
exp(f)

## [1] 0.4237448
!(exp(f)>=exp(d) & exp(f)<=exp(e))

## [1] FALSE
educ is not confounder for sex and age with cursmoke
glm1 <- glm(cursmoke ~ age+sex, data=smoke_vs1,family = 'binomial')
glm2 <- glm(cursmoke ~ age+sex+diabetes, data=smoke_vs1,family = 'binomial')
a=(summary(glm1)$coefficients[2])-(1.96 * (summary(glm1)$coefficients[5]))
b= (summary(glm1)$coefficients[2])+(1.96 * (summary(glm1)$coefficients[5]))
c=summary(glm2)$coefficients[2]
exp(a)

## [1] 0.9422074
exp(b)

## [1] 0.9563489
exp(c)

## [1] 0.949986
!(exp(c)>=exp(a) & exp(c)<=exp(b))

## [1] FALSE
d=(summary(glm1)$coefficients[3])-(1.96 * (summary(glm1)$coefficients[6]))
e=(summary(glm1)$coefficients[3])+(1.96 * (summary(glm1)$coefficients[6]))
f=summary(glm2)$coefficients[3]
exp(d)

## [1] 0.3737987
exp(e)

```

```
## [1] 0.4819127
```

```
exp(f)
```

```
## [1] 0.4229651
```

```
!(exp(f)>=exp(d) & exp(f)<=exp(e))
```

```
## [1] FALSE
```

diabetes is not confounder for sex and age with cursmoke

None of these are confounders. It makes sense as nothing could affect age. Same for question 2. So we only put age and sex into the model.

```
smoke_vs3 = smoke %>%
  dplyr::select(c(randid,sex,age,cursmoke)) %>%
  mutate(cursmoke = as.factor(cursmoke), sex=as.factor(sex)) %>%
  na.omit()
glmer_1 <- glmer(cursmoke ~ age * sex + (1 | randid),
  data = smoke_vs3,
  family = binomial)
summary(glmer_1)
```

```
## 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
```

significant.


```
#CI age
-0.23847-1.96*0.01456
```

```
## [1] -0.2670076
```

```
-0.23847+1.96*0.01456
```

```
## [1] -0.2099324
```

```
#CI sex
-6.93043-1.96*1.09252
```

```
## [1] -9.071769
```

```
-6.93043+1.96*1.09252
```

```
## [1] -4.789091
```

```
#CIage:sex2
0.05611-1.96*0.01671
```

```
## [1] 0.0233584
```

```
0.05611+1.96*0.01671
```

```
## [1] 0.0888616
```

Question 2

Variable Selection and confounder identification:

```
smoke_vs2 = smoke %>%
  filter(period==1) %>%
  dplyr::select(c(randid,sex,age,cigpday,totchol,bmi,hearttrte,educ,diabp,sysbp,diabetes)) %>%
  mutate(sex=as.factor(sex), diabetes=as.factor(diabetes)) %>%
  na.omit()
```

```
lm1 <- lm(cigpday ~ age+sex, data=smoke_vs2)
lm2 <- lm(cigpday ~ age+sex+totchol, data=smoke_vs2)
a=(summary(lm1)$coefficients[2])-(1.96 * (summary(lm1)$coefficients[5]))
b= (summary(lm1)$coefficients[2])+(1.96 * (summary(lm1)$coefficients[5]))
c=summary(lm2)$coefficients[2]
a
```

```
## [1] -0.2971391
```

```
b
```

```
## [1] -0.2200154
```

```
c
```

```
## [1] -0.2750505
```

```
!(c=a & c<=b)
```

```
## [1] FALSE
```

```
d=(summary(lm1)$coefficients[3])-(1.96 * (summary(lm1)$coefficients[6]))
e=(summary(lm1)$coefficients[3])+(1.96 * (summary(lm1)$coefficients[6]))
f=summary(lm2)$coefficients[3]
d
```

```

## [1] -8.361345
e

## [1] -7.01624
f

## [1] -7.762696
!(f>=d & f<=e)

## [1] FALSE
totchol is not confounder for age and sex with cigpday

lm1 <- lm(cigpday ~ age+sex, data=smoke_vs2)
lm2 <- lm(cigpday ~ age+sex+bmi, data=smoke_vs2)
a=(summary(lm1)$coefficients[2])-(1.96 * (summary(lm1)$coefficients[5]))
b= (summary(lm1)$coefficients[2])+(1.96 * (summary(lm1)$coefficients[5]))
c=summary(lm2)$coefficients[2]
a

## [1] -0.2971391
b

## [1] -0.2200154
c

## [1] -0.2414477
!(c=a & c<=b)

## [1] FALSE
d=(summary(lm1)$coefficients[3])-(1.96 * (summary(lm1)$coefficients[6]))
e=(summary(lm1)$coefficients[3])+(1.96 * (summary(lm1)$coefficients[6]))
f=summary(lm2)$coefficients[3]
d

## [1] -8.361345
e

## [1] -7.01624
f

## [1] -7.853529
!(f>=d & f<=e)

## [1] FALSE
bmi is not confounder for age and sex with cigpday

lm1 <- lm(cigpday ~ age+sex, data=smoke_vs2)
lm2 <- lm(cigpday ~ age+sex+hearttrte, data=smoke_vs2)
a=(summary(lm1)$coefficients[2])-(1.96 * (summary(lm1)$coefficients[5]))
b= (summary(lm1)$coefficients[2])+(1.96 * (summary(lm1)$coefficients[5]))
c=summary(lm2)$coefficients[2]
a

## [1] -0.2971391

```

```

b

## [1] -0.2200154
c

## [1] -0.256792
!(c=a & c<=b)

## [1] FALSE
d=(summary(lm1)$coefficients[3])-(1.96 * (summary(lm1)$coefficients[6]))
e=(summary(lm1)$coefficients[3])+(1.96 * (summary(lm1)$coefficients[6]))
f=summary(lm2)$coefficients[3]
d

## [1] -8.361345
e

## [1] -7.01624
f

## [1] -7.960397
!(f>=d & f<=e)

## [1] FALSE
hearttrte is not confounder for age and sex with cigpday
lm1 <- lm(cigpday ~ age+sex, data=smoke_vs2)
lm2 <- lm(cigpday ~ age+sex+educ, data=smoke_vs2)
a=(summary(lm1)$coefficients[2])-(1.96 * (summary(lm1)$coefficients[5]))
b= (summary(lm1)$coefficients[2])+(1.96 * (summary(lm1)$coefficients[5]))
c=summary(lm2)$coefficients[2]
a

## [1] -0.2971391
b

## [1] -0.2200154
c

## [1] -0.2659276
!(c=a & c<=b)

## [1] FALSE
d=(summary(lm1)$coefficients[3])-(1.96 * (summary(lm1)$coefficients[6]))
e=(summary(lm1)$coefficients[3])+(1.96 * (summary(lm1)$coefficients[6]))
f=summary(lm2)$coefficients[3]
d

## [1] -8.361345
e

## [1] -7.01624
f

```

```
## [1] -7.697665
```

```
!(f>=d & f<=e)
```

```
## [1] FALSE
```

educ is not confounder for age and sex with cigpday

```
lm1 <- lm(cigpday ~ age+sex, data=smoke_vs2)
```

```
lm2 <- lm(cigpday ~ age+sex+diabp, data=smoke_vs2)
```

```
a=(summary(lm1)$coefficients[2])-(1.96 * (summary(lm1)$coefficients[5]))
```

```
b= (summary(lm1)$coefficients[2])+(1.96 * (summary(lm1)$coefficients[5]))
```

```
c=summary(lm2)$coefficients[2]
```

```
a
```

```
## [1] -0.2971391
```

```
b
```

```
## [1] -0.2200154
```

```
c
```

```
## [1] -0.2479103
```

```
!(c=a & c<=b)
```

```
## [1] FALSE
```

```
d=(summary(lm1)$coefficients[3])-(1.96 * (summary(lm1)$coefficients[6]))
```

```
e=(summary(lm1)$coefficients[3])+(1.96 * (summary(lm1)$coefficients[6]))
```

```
f=summary(lm2)$coefficients[3]
```

```
d
```

```
## [1] -8.361345
```

```
e
```

```
## [1] -7.01624
```

```
f
```

```
## [1] -7.734606
```

```
!(f>=d & f<=e)
```

```
## [1] FALSE
```

diabp is not confounder for age and sex with cigpday

```
lm1 <- lm(cigpday ~ age+sex, data=smoke_vs2)
```

```
lm2 <- lm(cigpday ~ age+sex+sysbp, data=smoke_vs2)
```

```
a=(summary(lm1)$coefficients[2])-(1.96 * (summary(lm1)$coefficients[5]))
```

```
b= (summary(lm1)$coefficients[2])+(1.96 * (summary(lm1)$coefficients[5]))
```

```
c=summary(lm2)$coefficients[2]
```

```
a
```

```
## [1] -0.2971391
```

```
b
```

```
## [1] -0.2200154
```

```
c
```

```
## [1] -0.2525008
```

```

!(c=a & c<=b)

## [1] FALSE
d=(summary(lm1)$coefficients[3])-(1.96 * (summary(lm1)$coefficients[6]))
e=(summary(lm1)$coefficients[3])+(1.96 * (summary(lm1)$coefficients[6]))
f=summary(lm2)$coefficients[3]
d

## [1] -8.361345
e

## [1] -7.01624
f

## [1] -7.678295
!(f>=d & f<=e)

## [1] FALSE

sysbp is not confounder for age and sex with cigpday
lm1 <- lm(cigpday ~ age+sex, data=smoke_vs2)
lm2 <- lm(cigpday ~ age+sex+diabetes, data=smoke_vs2)
a=(summary(lm1)$coefficients[2])-(1.96 * (summary(lm1)$coefficients[5]))
b= (summary(lm1)$coefficients[2])+(1.96 * (summary(lm1)$coefficients[5]))
c=summary(lm2)$coefficients[2]
a

## [1] -0.2971391
b

## [1] -0.2200154
c

## [1] -0.2546299
!(c=a & c<=b)

## [1] FALSE
d=(summary(lm1)$coefficients[3])-(1.96 * (summary(lm1)$coefficients[6]))
e=(summary(lm1)$coefficients[3])+(1.96 * (summary(lm1)$coefficients[6]))
f=summary(lm2)$coefficients[3]
d

## [1] -8.361345
e

## [1] -7.01624
f

## [1] -7.701157
!(f>=d & f<=e)

## [1] FALSE

diabetes is not confounder for age and sex with cigpday

```

```

smoke_vs4 = smoke %>%
  dplyr::select(c(randid,cigpday,sex,age)) %>%
  mutate(sex=as.factor(sex)) %>%
  na.omit()

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_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

```

Ignore those who did not smoke through the whole study.

pvalue

```

#age
2*pnorm(-0.06629/0.02622)

```

```
## [1] 0.01146416
```

```

#CI age
-0.06629-1.96*0.02622

```

```
## [1] -0.1176812
```

```

-0.06629+1.96*0.02622

## [1] -0.0148988
#sex
2*pnorm(-15.59130/2.00555)

## [1] 7.599954e-15
#CI sex
-15.59130-1.96*2.00555

## [1] -19.52218
-15.59130+1.96*2.00555

## [1] -11.66042
#age:sex2
2*(1-pnorm(0.17357/0.03739))

## [1] 3.448019e-06
#CI age:sex2
0.17357-1.96*0.03739

## [1] 0.1002856
0.17357+1.96*0.03739

## [1] 0.2468544

```

Question 3

```

lm1 <- lm(sysbp ~ cursmoke, data=smoke_vs1)
lm2 <- lm(sysbp ~ cursmoke + sex, data=smoke_vs1)
a=(summary(lm1)$coefficients[2])-(1.96 * (summary(lm1)$coefficients[4]))
b=(summary(lm1)$coefficients[2])+(1.96 * (summary(lm1)$coefficients[4]))
c=summary(lm2)$coefficients[2]
a

## [1] -7.429258
b

## [1] -4.761322
c

## [1] -5.909267
!(c>=a & c<=b)

## [1] FALSE
sex is not confounder

lm1 <- lm(sysbp ~ cursmoke, data=smoke_vs1)
lm2 <- lm(sysbp ~ cursmoke + age, data=smoke_vs1)
a=(summary(lm1)$coefficients[2])-(1.96 * (summary(lm1)$coefficients[4]))
b=(summary(lm1)$coefficients[2])+(1.96 * (summary(lm1)$coefficients[4]))

```

```

c=summary(lm2)$coefficients[2]
a

## [1] -7.429258
b

## [1] -4.761322
c

## [1] -2.440408
!(c>=a & c<=b)

## [1] TRUE
age is not confounder

lm1 <- lm(sysbp ~ cursmoke, data=smoke_vs1)
lm2 <- lm(sysbp ~ cursmoke + totchol, data=smoke_vs1)
a=(summary(lm1)$coefficients[2])-(1.96 * (summary(lm1)$coefficients[4]))
b=(summary(lm1)$coefficients[2])+(1.96 * (summary(lm1)$coefficients[4]))
c=summary(lm2)$coefficients[2]
a

## [1] -7.429258
b

## [1] -4.761322
c

## [1] -5.642857
!(c>=a & c<=b)

## [1] FALSE
totchol is not confounder

lm1 <- lm(sysbp ~ cursmoke, data=smoke_vs1)
lm2 <- lm(sysbp ~ cursmoke + bmi, data=smoke_vs1)
a=(summary(lm1)$coefficients[2])-(1.96 * (summary(lm1)$coefficients[4]))
b=(summary(lm1)$coefficients[2])+(1.96 * (summary(lm1)$coefficients[4]))
c=summary(lm2)$coefficients[2]
a

## [1] -7.429258
b

## [1] -4.761322
c

## [1] -3.774433
!(c>=a & c<=b)

## [1] TRUE
bmi is confounder

```



```
lm1 <- lm(sysbp ~ cursmoke, data=smoke_vs1)
lm2 <- lm(sysbp ~ cursmoke + hearttrte, data=smoke_vs1)
a=(summary(lm1)$coefficients[2])-(1.96 * (summary(lm1)$coefficients[4]))
b=(summary(lm1)$coefficients[2])+(1.96 * (summary(lm1)$coefficients[4]))
c=summary(lm2)$coefficients[2]
a

## [1] -7.429258
b

## [1] -4.761322
c

## [1] -6.583046
!(c>=a & c<=b)
```

```
## [1] FALSE
hearttrte is not confounder
```

```
lm1 <- lm(sysbp ~ cursmoke, data=smoke_vs1)
lm2 <- lm(sysbp ~ cursmoke + educ, data=smoke_vs1)
a=(summary(lm1)$coefficients[2])-(1.96 * (summary(lm1)$coefficients[4]))
b=(summary(lm1)$coefficients[2])+(1.96 * (summary(lm1)$coefficients[4]))
c=summary(lm2)$coefficients[2]
a

## [1] -7.429258
b

## [1] -4.761322
c

## [1] -6.009864
!(c>=a & c<=b)
```

```
## [1] FALSE
educ is not confounder
```

```
lm1 <- lm(sysbp ~ cursmoke, data=smoke_vs1)
lm2 <- lm(sysbp ~ cursmoke + diabetes, data=smoke_vs1)
a=(summary(lm1)$coefficients[2])-(1.96 * (summary(lm1)$coefficients[4]))
b=(summary(lm1)$coefficients[2])+(1.96 * (summary(lm1)$coefficients[4]))
c=summary(lm2)$coefficients[2]
a

## [1] -7.429258
b

## [1] -4.761322
c

## [1] -5.876455
!(c>=a & c<=b)
```

```
## [1] FALSE
```

diabetes is not confounder

```
lm1 <- lm(sysbp ~ cursmoke, data=smoke_vs1)
lm2 <- lm(sysbp ~ cursmoke + prevap, data=smoke_vs1)
a=(summary(lm1)$coefficients[2])-(1.96 * (summary(lm1)$coefficients[4]))
b=(summary(lm1)$coefficients[2])+(1.96 * (summary(lm1)$coefficients[4]))
c=summary(lm2)$coefficients[2]
a

## [1] -7.429258
b

## [1] -4.761322
c

## [1] -5.941756
!(c>=a & c<=b)

## [1] FALSE
```

prevap is not confounder

```
lm1 <- lm(sysbp ~ cursmoke, data=smoke_vs1)
lm2 <- lm(sysbp ~ cursmoke + prevchd, data=smoke_vs1)
a=(summary(lm1)$coefficients[2])-(1.96 * (summary(lm1)$coefficients[4]))
b=(summary(lm1)$coefficients[2])+(1.96 * (summary(lm1)$coefficients[4]))
c=summary(lm2)$coefficients[2]
a

## [1] -7.429258
b

## [1] -4.761322
c

## [1] -6.016597
!(c>=a & c<=b)

## [1] FALSE
```

not confounder

```
lm1 <- lm(sysbp ~ cursmoke, data=smoke_vs1)
lm2 <- lm(sysbp ~ cursmoke + prevmi, data=smoke_vs1)
a=(summary(lm1)$coefficients[2])-(1.96 * (summary(lm1)$coefficients[4]))
b=(summary(lm1)$coefficients[2])+(1.96 * (summary(lm1)$coefficients[4]))
c=summary(lm2)$coefficients[2]
a

## [1] -7.429258
b

## [1] -4.761322
c

## [1] -6.139762
```

```
!(c>=a & c<=b)
```

```
## [1] FALSE
```

```
not confounder
```

```
lm1 <- lm(sysbp ~ cursmoke, data=smoke_vs1)
lm2 <- lm(sysbp ~ cursmoke + prevstrk, data=smoke_vs1)
a=(summary(lm1)$coefficients[2])-(1.96 * (summary(lm1)$coefficients[4]))
b=(summary(lm1)$coefficients[2])+(1.96 * (summary(lm1)$coefficients[4]))
c=summary(lm2)$coefficients[2]
a
```

```
## [1] -7.429258
```

```
b
```

```
## [1] -4.761322
```

```
c
```

```
## [1] -6.023463
```

```
!(c>=a & c<=b)
```

```
## [1] FALSE
```

```
not confounder
```

By rule of thumb, bmi is confounder. Based on the literature view, age and sex can be potential confounders as they can affect smoke status and sysbp at the same time. So we still put them into model.

```
smoke_vs5 = smoke %>%
  dplyr::select(c(randid,cursmoke,sex,age,bmi,sysbp)) %>%
  mutate(sex=as.factor(sex),cursmoke=as.factor(cursmoke)) %>%
  na.omit()
```

no interaction term because not significant

```
lmer_3 <- lmer(sysbp ~ cursmoke + bmi + sex + age + (1|randid), data = smoke_vs5)
summary(lmer_3)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: sysbp ~ cursmoke + bmi + sex + age + (1 | randid)
## Data: smoke_vs5
##
## REML criterion at convergence: 98637.7
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.2514 -0.5336 -0.0511  0.4610  6.0665
##
## Random effects:
##  Groups   Name                Variance Std.Dev.
##  randid   (Intercept)        261.1     16.16
##  Residual                    156.3     12.50
## Number of obs: 11575, groups:  randid, 4420
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  51.118109   1.973388  25.904
```

```
## cursmoke1    -0.008418    0.432847   -0.019
## bmi          1.443149    0.057512   25.093
## sex2         2.609485    0.552576    4.722
## age          0.855104    0.020583   41.544
##
## Correlation of Fixed Effects:
##           (Intr) crsmk1 bmi      sex2
## cursmoke1 -0.355
## bmi        -0.767  0.132
## sex2        -0.227  0.124  0.068
## age         -0.582  0.244 -0.020  0.010
```

Calculating p value using normal approximation:

```
#cursmoke
2*(pnorm(-0.008418/0.432847))
```

```
## [1] 0.9844837
```

Not significant, but include because this is our interest

```
#cursmoke 95% CI
-0.008418-1.96*0.432847
```

```
## [1] -0.8567981
```

```
-0.008418+1.96*0.432847
```

```
## [1] 0.8399621
```

```
#bmi
2*(1-pnorm(1.443149/0.057512))
```

```
## [1] 0
```

```
#CI bmi
1.443149-1.96*0.057512
```

```
## [1] 1.330425
```

```
1.443149+1.96*0.057512
```

```
## [1] 1.555873
```

```
#sex
2*(1-pnorm(2.609485/0.552576))
```

```
## [1] 2.330775e-06
```

```
#95%CI sex
2.609485-1.96*0.552576
```

```
## [1] 1.526436
```

```
2.609485+1.96*0.552576
```

```
## [1] 3.692534
```

```
#age
2*(1-pnorm(0.855104/0.020583))
```

```
## [1] 0
```

```
#age
0.855104-1.96*0.020583
```

```
## [1] 0.8147613
```

```
0.855104+1.96*0.020583
```

```
## [1] 0.8954467
```

Question 4

```
lm1 <- lm(diabp ~ cursmoke, data=smoke_vs1)
lm2 <- lm(diabp ~ cursmoke + sex, data=smoke_vs1)
a=(summary(lm1)$coefficients[2])-(1.96 * (summary(lm1)$coefficients[4]))
b=(summary(lm1)$coefficients[2])+(1.96 * (summary(lm1)$coefficients[4]))
c=summary(lm2)$coefficients[2]
a
```

```
## [1] -3.365883
```

```
b
```

```
## [1] -1.922971
```

```
c
```

```
## [1] -2.982803
```

```
!(c>=a & c<=b)
```

```
## [1] FALSE
```

sex is not confounder

```
lm1 <- lm(diabp ~ cursmoke, data=smoke_vs1)
lm2 <- lm(diabp ~ cursmoke + age, data=smoke_vs1)
a=(summary(lm1)$coefficients[2])-(1.96 * (summary(lm1)$coefficients[4]))
b=(summary(lm1)$coefficients[2])+(1.96 * (summary(lm1)$coefficients[4]))
c=summary(lm2)$coefficients[2]
a
```

```
## [1] -3.365883
```

```
b
```

```
## [1] -1.922971
```

```
c
```

```
## [1] -1.662623
```

```
!(c>=a & c<=b)
```

```
## [1] TRUE
```

age is confounder

```
lm1 <- lm(diabp ~ cursmoke, data=smoke_vs1)
lm2 <- lm(diabp ~ cursmoke + totchol, data=smoke_vs1)
a=(summary(lm1)$coefficients[2])-(1.96 * (summary(lm1)$coefficients[4]))
b=(summary(lm1)$coefficients[2])+(1.96 * (summary(lm1)$coefficients[4]))
```

```

c=summary(lm2)$coefficients[2]
a

## [1] -3.365883
b

## [1] -1.922971
c

## [1] -2.441814
!(c>=a & c<=b)

## [1] FALSE
totchol is not confounder

lm1 <- lm(diabp ~ cursmoke, data=smoke_vs1)
lm2 <- lm(diabp ~ cursmoke + bmi, data=smoke_vs1)
a=(summary(lm1)$coefficients[2])-(1.96 * (summary(lm1)$coefficients[4]))
b=(summary(lm1)$coefficients[2])+(1.96 * (summary(lm1)$coefficients[4]))
c=summary(lm2)$coefficients[2]
a

## [1] -3.365883
b

## [1] -1.922971
c

## [1] -1.168078
!(c>=a & c<=b)

## [1] TRUE
bmi is confounder

lm1 <- lm(diabp ~ cursmoke, data=smoke_vs1)
lm2 <- lm(diabp ~ cursmoke + hearttrte, data=smoke_vs1)
a=(summary(lm1)$coefficients[2])-(1.96 * (summary(lm1)$coefficients[4]))
b=(summary(lm1)$coefficients[2])+(1.96 * (summary(lm1)$coefficients[4]))
c=summary(lm2)$coefficients[2]
a

## [1] -3.365883
b

## [1] -1.922971
c

## [1] -2.910674
!(c>=a & c<=b)

## [1] FALSE
hearttrte is not confounder

```

```
lm1 <- lm(diabp ~ cursmoke, data=smoke_vs1)
lm2 <- lm(diabp ~ cursmoke + educ, data=smoke_vs1)
a=(summary(lm1)$coefficients[2])-(1.96 * (summary(lm1)$coefficients[4]))
b=(summary(lm1)$coefficients[2])+(1.96 * (summary(lm1)$coefficients[4]))
c=summary(lm2)$coefficients[2]
a

## [1] -3.365883
b

## [1] -1.922971
c

## [1] -2.623741
!(c>=a & c<=b)
```

```
## [1] FALSE
educ is not confounder
```

```
lm1 <- lm(diabp ~ cursmoke, data=smoke_vs1)
lm2 <- lm(diabp ~ cursmoke + diabetes, data=smoke_vs1)
a=(summary(lm1)$coefficients[2])-(1.96 * (summary(lm1)$coefficients[4]))
b=(summary(lm1)$coefficients[2])+(1.96 * (summary(lm1)$coefficients[4]))
c=summary(lm2)$coefficients[2]
a

## [1] -3.365883
b

## [1] -1.922971
c

## [1] -2.593779
!(c>=a & c<=b)
```

```
## [1] FALSE
diabetes is not confounder
```

```
lm1 <- lm(diabp ~ cursmoke, data=smoke_vs1)
lm2 <- lm(diabp ~ cursmoke + prevap, data=smoke_vs1)
a=(summary(lm1)$coefficients[2])-(1.96 * (summary(lm1)$coefficients[4]))
b=(summary(lm1)$coefficients[2])+(1.96 * (summary(lm1)$coefficients[4]))
c=summary(lm2)$coefficients[2]
a

## [1] -3.365883
b

## [1] -1.922971
c

## [1] -2.586016
!(c>=a & c<=b)
```

```
## [1] FALSE
```

not confounder

```
lm1 <- lm(diabp ~ cursmoke, data=smoke_vs1)
lm2 <- lm(diabp ~ cursmoke + prevchd, data=smoke_vs1)
a=(summary(lm1)$coefficients[2])-(1.96 * (summary(lm1)$coefficients[4]))
b=(summary(lm1)$coefficients[2])+(1.96 * (summary(lm1)$coefficients[4]))
c=summary(lm2)$coefficients[2]
```

a

```
## [1] -3.365883
```

b

```
## [1] -1.922971
```

c

```
## [1] -2.619172
```

```
!(c>=a & c<=b)
```

```
## [1] FALSE
```

not confounder

```
lm1 <- lm(diabp ~ cursmoke, data=smoke_vs1)
lm2 <- lm(diabp ~ cursmoke + prevmi, data=smoke_vs1)
a=(summary(lm1)$coefficients[2])-(1.96 * (summary(lm1)$coefficients[4]))
b=(summary(lm1)$coefficients[2])+(1.96 * (summary(lm1)$coefficients[4]))
c=summary(lm2)$coefficients[2]
```

a

```
## [1] -3.365883
```

b

```
## [1] -1.922971
```

c

```
## [1] -2.65378
```

```
!(c>=a & c<=b)
```

```
## [1] FALSE
```

not confounder

```
lm1 <- lm(diabp ~ cursmoke, data=smoke_vs1)
lm2 <- lm(diabp ~ cursmoke + prevstrk, data=smoke_vs1)
a=(summary(lm1)$coefficients[2])-(1.96 * (summary(lm1)$coefficients[4]))
b=(summary(lm1)$coefficients[2])+(1.96 * (summary(lm1)$coefficients[4]))
c=summary(lm2)$coefficients[2]
```

a

```
## [1] -3.365883
```

b

```
## [1] -1.922971
```

c

```
## [1] -2.618415
```



```
!(c>=a & c<=b)
```

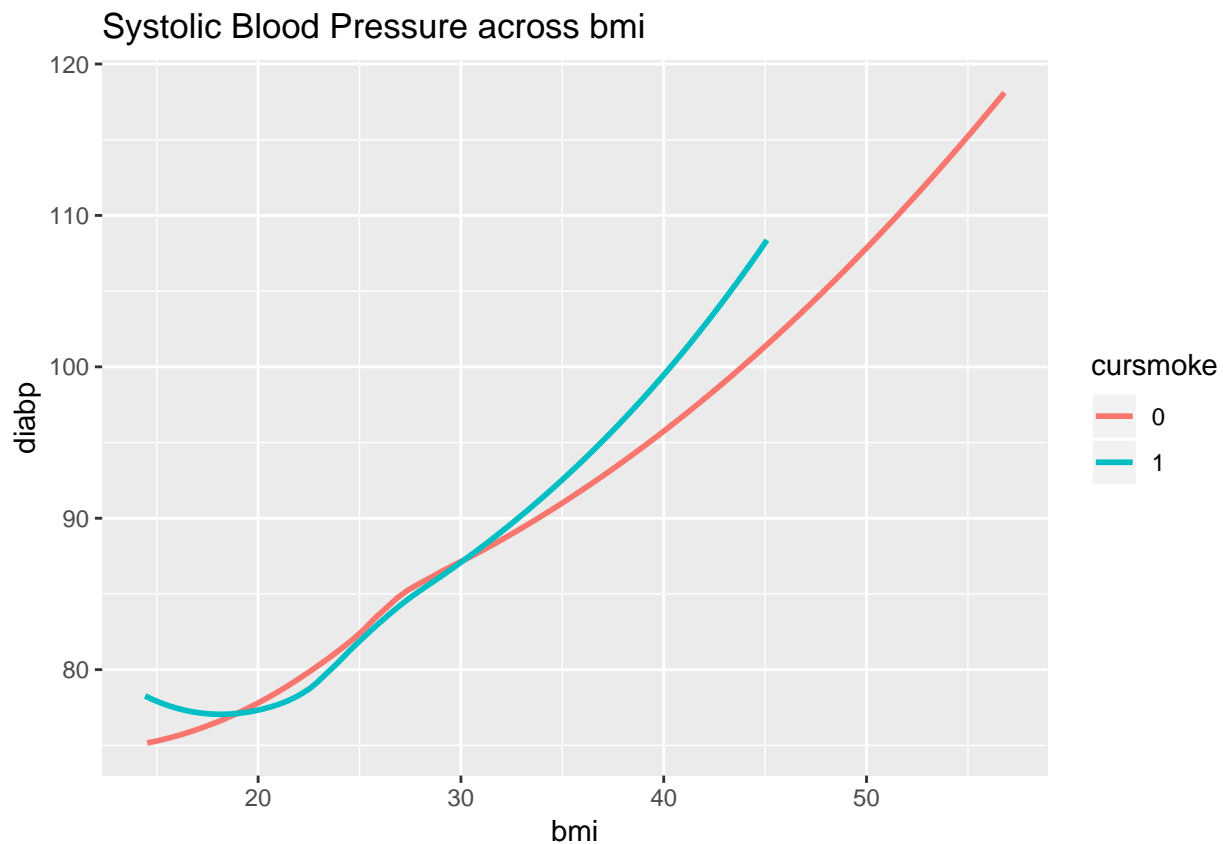
```
## [1] FALSE
```

not confounder

By rule of thumb, age, bmi is confounder. Based on the literature view, sex can be potential confounders as it can affect smoke status and diabp at the same time. So we still put them into model.

checking interactions

```
smoke_vs6 = smoke %>%  
  dplyr::select(c(randid, cursmoke, sex, age, bmi, diabp)) %>%  
  mutate(sex=as.factor(sex), cursmoke=as.factor(cursmoke)) %>%  
  na.omit()  
smoke_vs6 %>%  
  mutate(cursmoke = as.factor(cursmoke)) %>%  
  ggplot(aes(bmi, diabp, group = cursmoke, color = cursmoke)) +  
  geom_smooth(method = "loess", se = F) +  
  ggtitle("Systolic Blood Pressure across bmi")
```



This plot argues for some interaction of the two predictors, as the lines are not parallel.

```
lmer_4 <- lmer(diabp ~ cursmoke*bmi + sex + cursmoke*age + (1|randid), data = smoke_vs6)  
summary(lmer_4)
```

```
## Linear mixed model fit by REML ['lmerMod']  
## Formula: diabp ~ cursmoke * bmi + sex + cursmoke * age + (1 | randid)  
## Data: smoke_vs6
```

```
##
## REML criterion at convergence: 85394.2
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -5.6356 -0.5401 -0.0204  0.5116  4.5821
##
## Random effects:
##   Groups   Name      Variance Std.Dev.
##   randid   (Intercept) 69.18    8.318
##   Residual                53.74    7.331
## Number of obs: 11575, groups:  randid, 4420
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)   61.12030    1.32853  46.006
## cursmoke1    -11.20762    1.83644  -6.103
## bmi           0.98322    0.03816  25.763
## sex2         -0.51623    0.29304  -1.762
## age          -0.05122    0.01442  -3.553
## cursmoke1:bmi  0.09296    0.05477   1.697
## cursmoke1:age  0.16113    0.02129   7.570
##
## Correlation of Fixed Effects:
##              (Intr) crsmk1 bmi    sex2    age    crsmk1:b
## cursmoke1    -0.607
## bmi          -0.763  0.460
## sex2         -0.160 -0.023  0.040
## age          -0.616  0.403 -0.003 -0.005
## cursmok1:bm  0.434 -0.768 -0.569  0.033 -0.005
## cursmoke1:g  0.375 -0.626 -0.013  0.023 -0.596 -0.001
```

pvalue

```
coefs4<- data.frame(coef(summary(lmer_4)))
pvalue<-round(2*(1-pnorm(abs(coefs4$t.value))),3)
pvalue
```

```
## [1] 0.000 0.000 0.000 0.078 0.000 0.090 0.000
```

cursmoke1:bmi not significant, remove this interaction

```
lmer_41 <- lmer(diabp ~ bmi + sex + cursmoke*age + (1|randid), data = smoke_vs6)
summary(lmer_41)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: diabp ~ bmi + sex + cursmoke * age + (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
## bmi 1.02005 0.03139 32.496
## sex2 -0.53287 0.29288 -1.819
## cursmoke1 -8.81416 1.17670 -7.491
## age -0.05110 0.01442 -3.544
## cursmoke1:age 0.16118 0.02129 7.572
##
## Correlation of Fixed Effects:
## (Intr) bmi sex2 crsmk1 age
## bmi -0.697
## sex2 -0.194 0.072
## cursmoke1 -0.475 0.044 0.004
## age -0.682 -0.006 -0.005 0.623
## cursmoke1:g 0.417 -0.016 0.023 -0.979 -0.596
```

```
#cursmoke
2*(pnorm(-8.81416/1.17670))
```

```
## [1] 6.857235e-14
```

```
#CI cursmoke
-8.81416-1.96*1.17670
```

```
## [1] -11.12049
```

```
-8.81416+1.96*1.17670
```

```
## [1] -6.507828
```

```
#bmi
2*(1-pnorm(1.02005/0.03139))
```

```
## [1] 0
```

```
#CIbmi
1.02005-1.96*0.03139
```

```
## [1] 0.9585256
```

```
1.02005+1.96*0.03139
```

```
## [1] 1.081574
```

```
#sex
2*(pnorm(-0.53287/0.29288))
```

```
## [1] 0.06884828
```

although not significant, it is confounder, so we still put in into the model

```
#CI sex
-0.53287-1.96*0.29288
```

```
## [1] -1.106915
```

```

-0.53287+1.96*0.29288

## [1] 0.0411748
#age
2*pnorm(-0.05110/0.01442)

## [1] 0.0003945699
#CI age
-0.05110-1.96*0.01442

## [1] -0.0793632
-0.05110+1.96*0.01442

## [1] -0.0228368
#cursmoke1:age
2*(1-pnorm(0.16118/0.02129 ))

## [1] 3.708145e-14
#cursmoke1:age
0.16118-1.96*0.02129

## [1] 0.1194516
0.16118+1.96*0.02129

## [1] 0.2029084

```

Question 5

```

lm1 <- lm(totchol ~ cursmoke, data=smoke_vs1)
lm2 <- lm(totchol ~ cursmoke + sex, data=smoke_vs1)
a=(summary(lm1)$coefficients[2])-(1.96 * (summary(lm1)$coefficients[4]))
b=(summary(lm1)$coefficients[2])+(1.96 * (summary(lm1)$coefficients[4]))
c=summary(lm2)$coefficients[2]
a

## [1] -7.188361
b

## [1] -1.871145
c

## [1] -3.398869
!(c>=a & c<=b)

## [1] FALSE
sex is not confounder

lm1 <- lm(totchol ~ cursmoke, data=smoke_vs1)
lm2 <- lm(totchol ~ cursmoke + age, data=smoke_vs1)
a=(summary(lm1)$coefficients[2])-(1.96 * (summary(lm1)$coefficients[4]))
b=(summary(lm1)$coefficients[2])+(1.96 * (summary(lm1)$coefficients[4]))

```

```

c=summary(lm2)$coefficients[2]
a

## [1] -7.188361
b

## [1] -1.871145
c

## [1] 0.3207259
!(c>=a & c<=b)

## [1] TRUE
age is confounder
lm1 <- lm(totchol ~ cursmoke, data=smoke_vs1)
lm2 <- lm(totchol ~ cursmoke + diabp, data=smoke_vs1)
a=(summary(lm1)$coefficients[2])-(1.96 * (summary(lm1)$coefficients[4]))
b=(summary(lm1)$coefficients[2])+(1.96 * (summary(lm1)$coefficients[4]))
c=summary(lm2)$coefficients[2]
a

## [1] -7.188361
b

## [1] -1.871145
c

## [1] -2.923499
!(c>=a & c<=b)

## [1] FALSE
diabp is not confounder
lm1 <- lm(totchol ~ cursmoke, data=smoke_vs1)
lm2 <- lm(totchol ~ cursmoke + bmi, data=smoke_vs1)
a=(summary(lm1)$coefficients[2])-(1.96 * (summary(lm1)$coefficients[4]))
b=(summary(lm1)$coefficients[2])+(1.96 * (summary(lm1)$coefficients[4]))
c=summary(lm2)$coefficients[2]
a

## [1] -7.188361
b

## [1] -1.871145
c

## [1] -2.802548
!(c>=a & c<=b)

## [1] FALSE
bmi is confounder

```

```
lm1 <- lm(totchol ~ cursmoke, data=smoke_vs1)
lm2 <- lm(totchol ~ cursmoke + hearttrte, data=smoke_vs1)
a=(summary(lm1)$coefficients[2])-(1.96 * (summary(lm1)$coefficients[4]))
b=(summary(lm1)$coefficients[2])+(1.96 * (summary(lm1)$coefficients[4]))
c=summary(lm2)$coefficients[2]
a
```

```
## [1] -7.188361
```

```
b
```

```
## [1] -1.871145
```

```
c
```

```
## [1] -5.013823
```

```
!(c>=a & c<=b)
```

```
## [1] FALSE
```

hearttrte is not confounder

```
lm1 <- lm(totchol ~ cursmoke, data=smoke_vs1)
lm2 <- lm(totchol ~ cursmoke + educ, data=smoke_vs1)
a=(summary(lm1)$coefficients[2])-(1.96 * (summary(lm1)$coefficients[4]))
b=(summary(lm1)$coefficients[2])+(1.96 * (summary(lm1)$coefficients[4]))
c=summary(lm2)$coefficients[2]
a
```

```
## [1] -7.188361
```

```
b
```

```
## [1] -1.871145
```

```
c
```

```
## [1] -4.505649
```

```
!(c>=a & c<=b)
```

```
## [1] FALSE
```

educ is not confounder

```
lm1 <- lm(totchol ~ cursmoke, data=smoke_vs1)
lm2 <- lm(totchol ~ cursmoke + sysbp, data=smoke_vs1)
a=(summary(lm1)$coefficients[2])-(1.96 * (summary(lm1)$coefficients[4]))
b=(summary(lm1)$coefficients[2])+(1.96 * (summary(lm1)$coefficients[4]))
c=summary(lm2)$coefficients[2]
a
```

```
## [1] -7.188361
```

```
b
```

```
## [1] -1.871145
```

```
c
```

```
## [1] -2.111556
```

```
!(c>=a & c<=b)
```

```
## [1] FALSE
```

sysbp is not confounder

```
lm1 <- lm(totchol ~ cursmoke, data=smoke_vs1)
lm2 <- lm(totchol ~ cursmoke + diabetes, data=smoke_vs1)
a=(summary(lm1)$coefficients[2])-(1.96 * (summary(lm1)$coefficients[4]))
b=(summary(lm1)$coefficients[2])+(1.96 * (summary(lm1)$coefficients[4]))
c=summary(lm2)$coefficients[2]
a

## [1] -7.188361
b

## [1] -1.871145
c

## [1] -4.361909
!(c>=a & c<=b)

## [1] FALSE
```

diabetes is not confounder

```
lm1 <- lm(totchol ~ cursmoke, data=smoke_vs1)
lm2 <- lm(totchol ~ cursmoke + prevap, data=smoke_vs1)
a=(summary(lm1)$coefficients[2])-(1.96 * (summary(lm1)$coefficients[4]))
b=(summary(lm1)$coefficients[2])+(1.96 * (summary(lm1)$coefficients[4]))
c=summary(lm2)$coefficients[2]
a

## [1] -7.188361
b

## [1] -1.871145
c

## [1] -4.450606
!(c>=a & c<=b)

## [1] FALSE
```

not confounder

```
lm1 <- lm(totchol ~ cursmoke, data=smoke_vs1)
lm2 <- lm(totchol ~ cursmoke + prevchd, data=smoke_vs1)
a=(summary(lm1)$coefficients[2])-(1.96 * (summary(lm1)$coefficients[4]))
b=(summary(lm1)$coefficients[2])+(1.96 * (summary(lm1)$coefficients[4]))
c=summary(lm2)$coefficients[2]
a

## [1] -7.188361
b

## [1] -1.871145
c

## [1] -4.492534
```

```
!(c>=a & c<=b)
```

```
## [1] FALSE
```

```
not confounder
```

```
lm1 <- lm(totchol ~ cursmoke, data=smoke_vs1)
lm2 <- lm(totchol ~ cursmoke + prevmi, data=smoke_vs1)
a=(summary(lm1)$coefficients[2])-(1.96 * (summary(lm1)$coefficients[4]))
b=(summary(lm1)$coefficients[2])+(1.96 * (summary(lm1)$coefficients[4]))
c=summary(lm2)$coefficients[2]
a
```

```
## [1] -7.188361
```

```
b
```

```
## [1] -1.871145
```

```
c
```

```
## [1] -4.56068
```

```
!(c>=a & c<=b)
```

```
## [1] FALSE
```

```
not confounder
```

```
lm1 <- lm(totchol ~ cursmoke, data=smoke_vs1)
lm2 <- lm(totchol ~ cursmoke + prevstrk, data=smoke_vs1)
a=(summary(lm1)$coefficients[2])-(1.96 * (summary(lm1)$coefficients[4]))
b=(summary(lm1)$coefficients[2])+(1.96 * (summary(lm1)$coefficients[4]))
c=summary(lm2)$coefficients[2]
a
```

```
## [1] -7.188361
```

```
b
```

```
## [1] -1.871145
```

```
c
```

```
## [1] -4.526573
```

```
!(c>=a & c<=b)
```

```
## [1] FALSE
```

```
not confounder
```

```
lm1 <- lm(totchol ~ cursmoke, data=smoke_vs1)
lm2 <- lm(totchol ~ cursmoke + prevhyp, data=smoke_vs1)
a=(summary(lm1)$coefficients[2])-(1.96 * (summary(lm1)$coefficients[4]))
b=(summary(lm1)$coefficients[2])+(1.96 * (summary(lm1)$coefficients[4]))
c=summary(lm2)$coefficients[2]
a
```

```
## [1] -7.188361
```

```
b
```

```
## [1] -1.871145
```



```
c
```

```
## [1] -2.94428
```

```
!(c>=a & c<=b)
```

```
## [1] FALSE
```

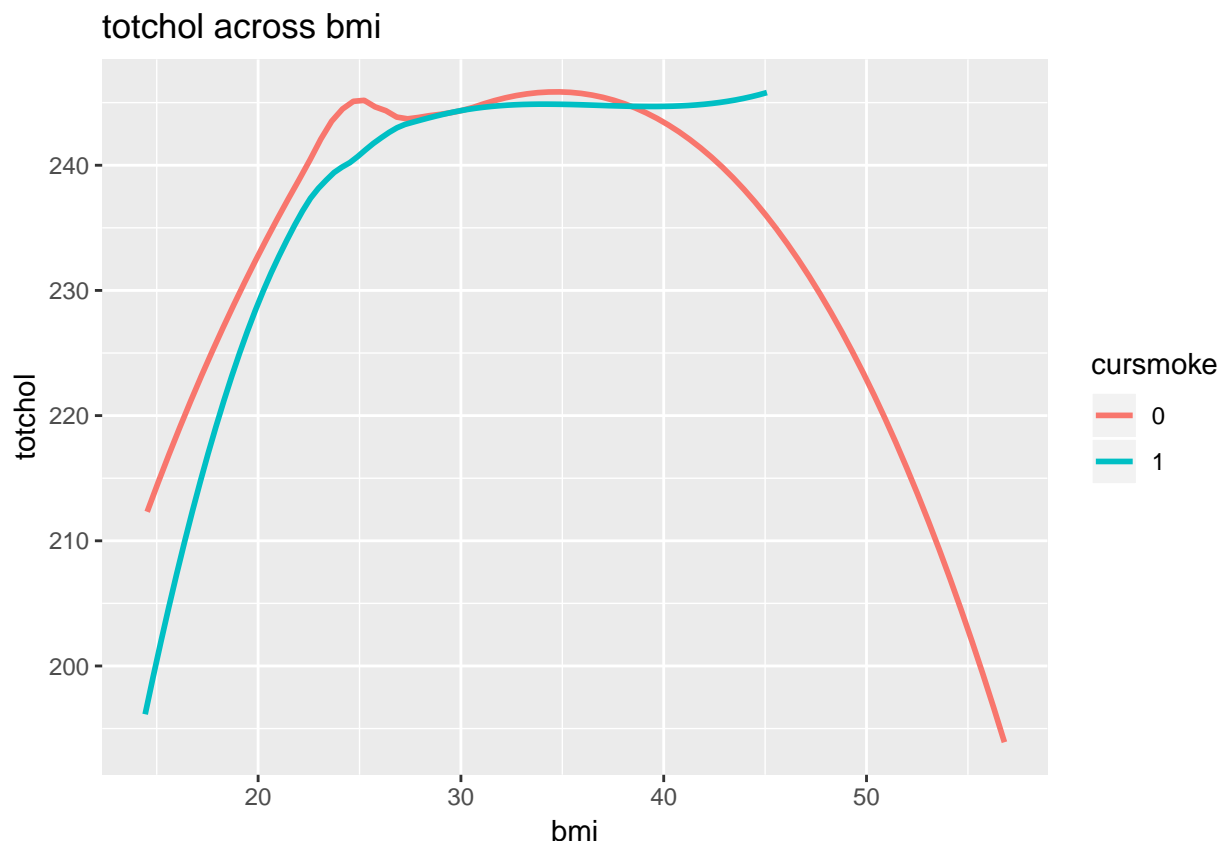
not confounder

By rule of thumb, age, bmi is confounder. Based on the literature view, sex can be potential confounders as it can affect smoke status and totchol at the same time. So we still put them into model.

```
smoke_vs7 = smoke %>%  
  dplyr::select(c(randid, cursmoke, sex, age, bmi, totchol)) %>%  
  mutate(sex=as.factor(sex), cursmoke=as.factor(cursmoke)) %>%  
  na.omit()
```

checking interactions

```
smoke_vs7 %>%  
  mutate(cursmoke = as.factor(cursmoke)) %>%  
  ggplot(aes(bmi, totchol, group = cursmoke, color = cursmoke)) +  
  geom_smooth(method = "loess", se = F) +  
  ggtitle("totchol across bmi")
```



This plot argues for some interaction of the two predictors, as the lines are not parallel.

```
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

#cursmoke
2*pnorm(-29.38692/7.05633)

## [1] 3.118741e-05

#CI cursmoke
-29.38692-1.96*7.05633

## [1] -43.21733

-29.38692+1.96*7.05633

## [1] -15.55651

#bmi
2*(1-pnorm(1.45643/0.15257))

## [1] 0

#CI bmi
1.45643-1.96*0.15257

## [1] 1.157393

```

```

1.45643+1.96*0.15257
## [1] 1.755467
#sex
2*(1-pnorm(13.94202/1.22545))
## [1] 0
#CI sex
13.94202-1.96*1.22545
## [1] 11.54014
13.94202+1.96*1.22545
## [1] 16.3439
#age
2*(1-pnorm(0.32861/0.05514))
## [1] 2.529219e-09
#CI age
0.32861-1.96*0.05514
## [1] 0.2205356
0.32861+1.96*0.05514
## [1] 0.4366844
#curSmoke1:bmi
2*(1-pnorm(0.74739/0.21240))
## [1] 0.0004335274
#CI curSmoke1:bmi
0.74739-1.96*0.21240
## [1] 0.331086
0.74739+1.96*0.21240
## [1] 1.163694
#curSmoke1:age
2*(1-pnorm(0.26381/0.08053))
## [1] 0.001053176
#CI curSmoke1:age
0.26381-1.96*0.08053
## [1] 0.1059712
0.26381+1.96*0.08053
## [1] 0.4216488

```

Please include a table which shows point estimate, 95 CI and p value for each term in the model(calculated above)