

# Longitudinal Data HW1 Xinyue Lu

## 1.Data Summary

Read BMACS data

```
library(lme4)
```

```
## Loading required package: Matrix
```

```
BMACS <- read.csv("D:/luxinyve/00 Longitudinal Data/homework1/BMACS.csv", header=T)  
str(BMACS)
```

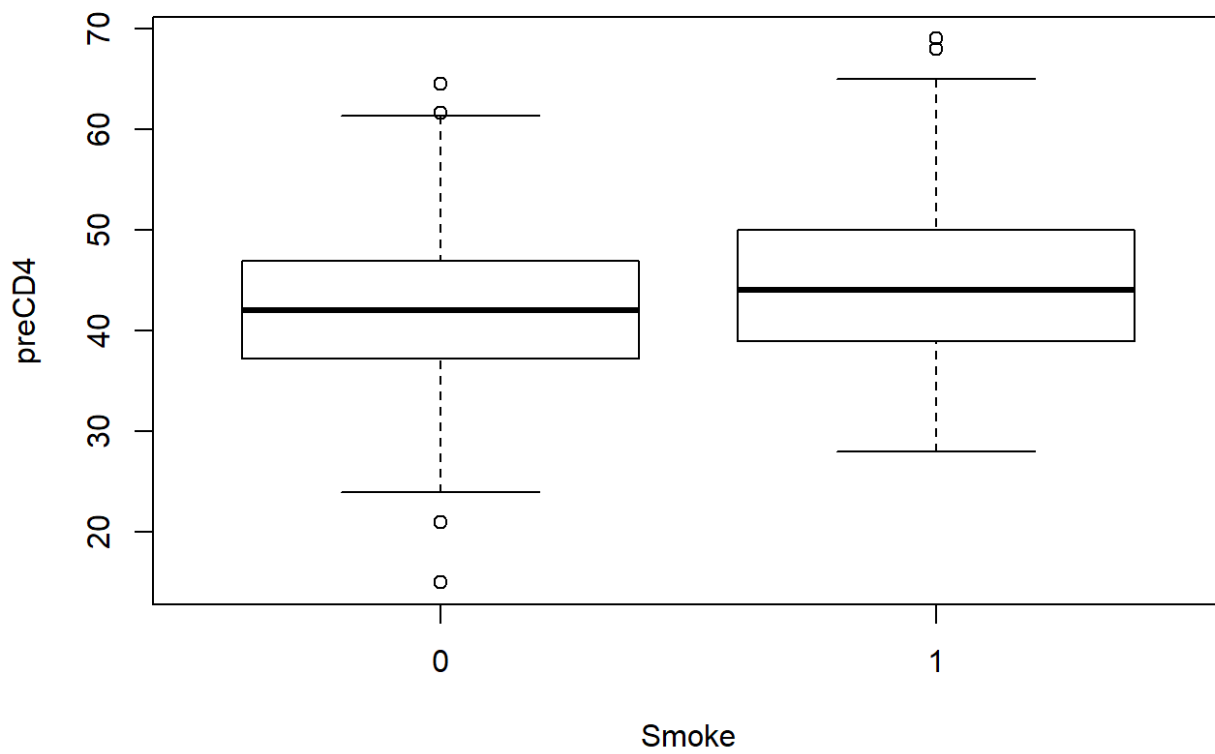
```
## 'data.frame': 1817 obs. of 6 variables:  
## $ ID : int 1022 1022 1022 1022 1022 1022 1022 1049 1049 1049 ...  
## $ Time : num 0.2 0.8 1.2 1.6 2.5 3 4.1 0.3 0.6 1 ...  
## $ Smoke : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ age : num 26.2 26.2 26.2 26.2 26.2 ...  
## $ preCD4: num 38 38 38 38 38 38 38 44.5 44.5 44.5 ...  
## $ CD4 : num 17 30 23 15 21 12 5 37 44 37 ...
```

```
summary(BMACS)
```

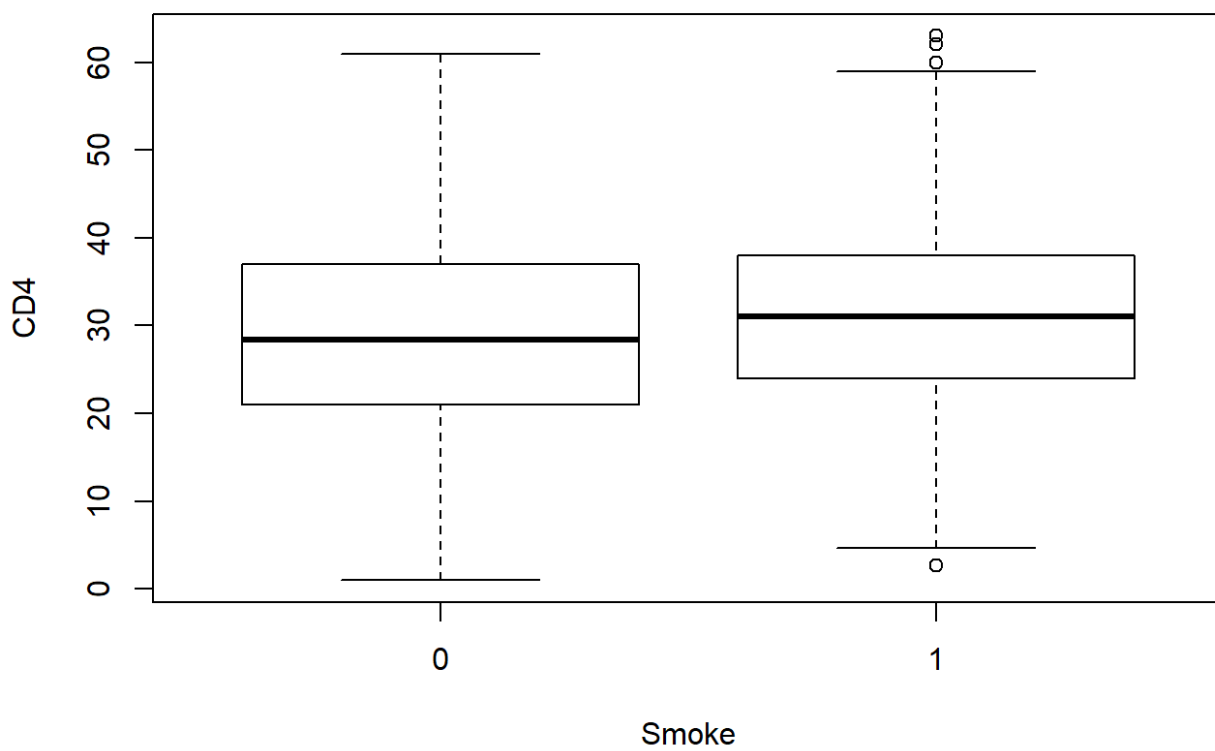
```
##           ID           Time           Smoke           age  
## Min.      :1022   Min.      :0.100   Min.      :0.000   Min.      :18.25  
## 1st Qu.:3334   1st Qu.:1.000   1st Qu.:0.000   1st Qu.:29.17  
## Median :5075   Median :2.200   Median :0.000   Median :32.58  
## Mean    :5401   Mean    :2.347   Mean    :0.311   Mean    :34.36  
## 3rd Qu.:7744   3rd Qu.:3.500   3rd Qu.:1.000   3rd Qu.:39.00  
## Max.    :9954   Max.    :5.900   Max.    :1.000   Max.    :59.92  
##      preCD4      CD4  
## Min.      :15.00   Min.      : 1.00  
## 1st Qu.:37.67   1st Qu.:21.20  
## Median :42.33   Median :29.00  
## Mean    :42.69   Mean    :29.26  
## 3rd Qu.:47.12   3rd Qu.:37.00  
## Max.     :69.00   Max.     :63.00
```

Use boxplot to show the relationship between CD4 and Smoke. It showed that people who always smoked had higher preCD4 and CD4.

```
boxplot(BMACS$preCD4~BMACS$Smoke, xlab="Smoke", ylab="preCD4")
```



```
boxplot(BMACS$CD4~BMACS$Smoke, xlab="Smoke", ylab="CD4")
```

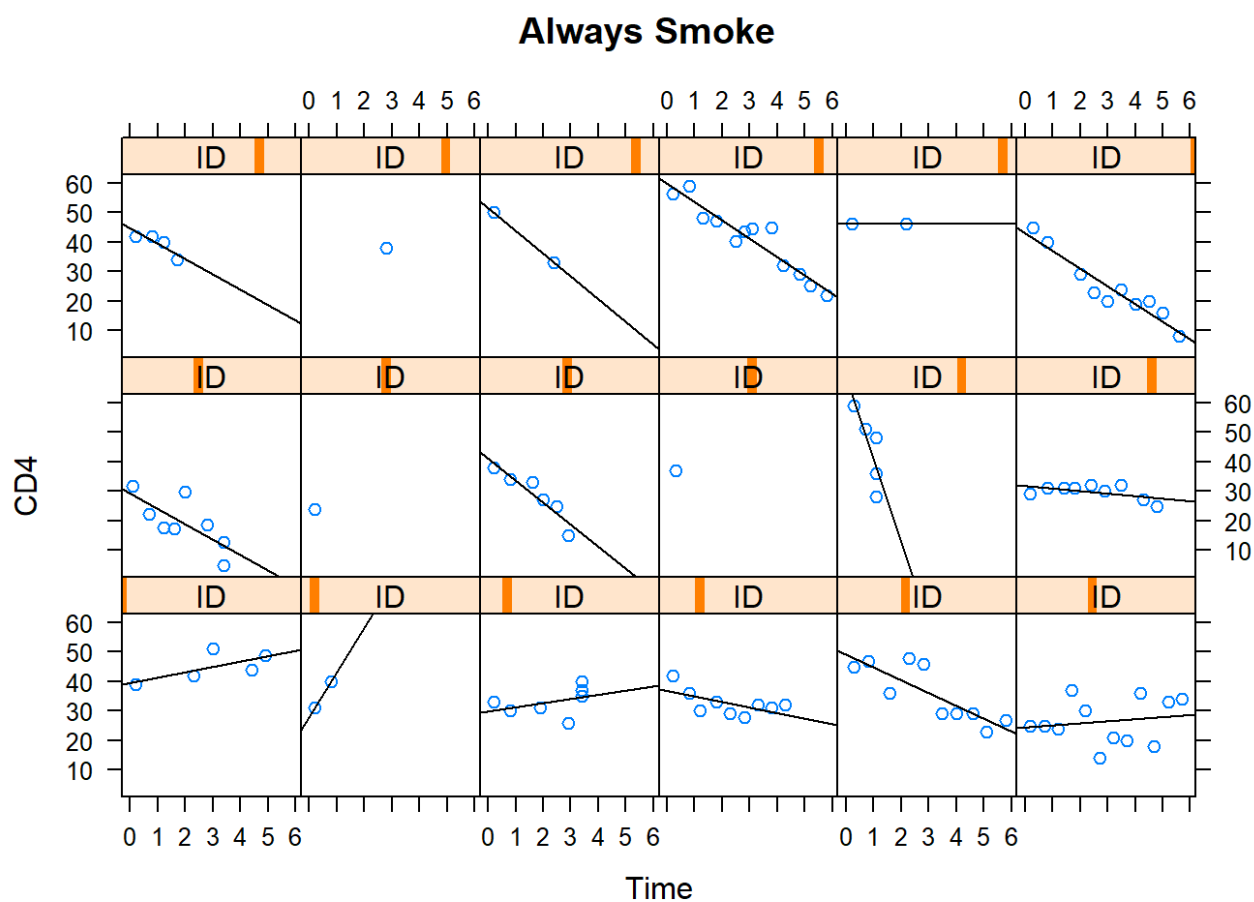


Plot the relationship of CD4 and Time. Choose 18 candidates from each smoking group.

For both groups of the people, no matter they smoke heavily or never, most people's CD4 went down or nearly unchanged as time passed by.

### Always Smoke

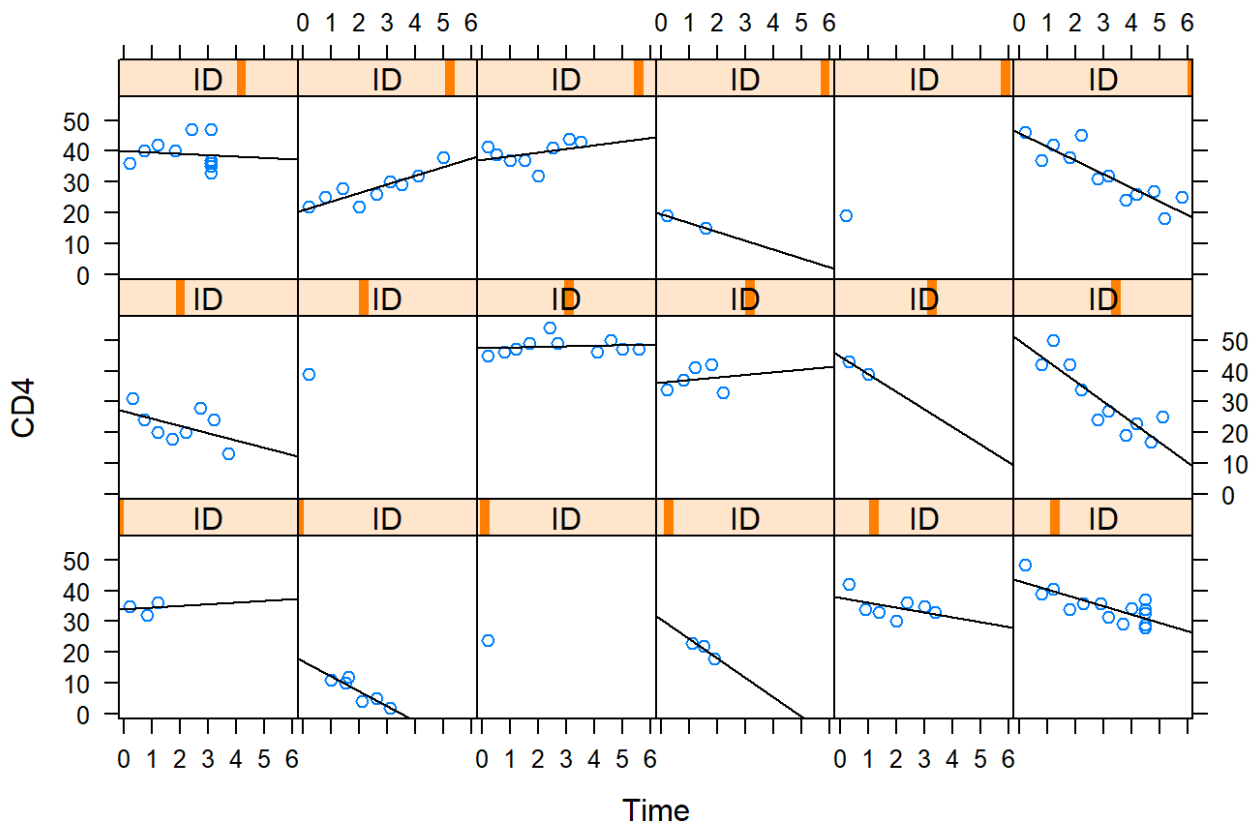
```
set.seed(1)
library(lattice)
BMACS=as.data.frame(BMACS)
Smoke1 <- sample(unique(BMACS$ID[BMACS$Smoke==1]), 18)
Smoke1.18=BMACS[is.element(BMACS$ID, Smoke1), ]
xyplot(CD4~Time|ID, data=Smoke1.18, main="Always Smoke", xlab="Time", ylab="CD4",
       layout=c(6, 3), panel=function(x, y) {
  panel.xyplot(x, y)
  panel.lmline(x, y)})
```



### Never Smoke

```
Smoke0 <- sample(unique(BMACS$ID[BMACS$Smoke==0]), 18)
Smoke0.18=BMACS[is.element(BMACS$ID, Smoke0), ]
xyplot(CD4~Time|ID, data=Smoke0.18, main="Never Smoke", xlab="Time", ylab="CD4",
       layout=c(6, 3), panel=function(x, y) {
  panel.xyplot(x, y)
  panel.lmline(x, y)})
```

## Never Smoke



If we fit CD4 and Time with simple linear regression models for each patient, most of the slopes of Time was negative, which proved the above conclusion.

```
lr1=lmList(CD4~Time|ID, subset=Smoke==1, data=BMACS)
head(coef(lr1))
```

```
##      (Intercept)      Time
## 1126    39.65189    1.806795
## 1265    24.82000   18.400000
## 1298    29.06217   -2.255773
## 1308    44.52326   -3.545325
## 1359    35.00000         NA
## 1505    50.24554   -6.032773
```

```
lr0=lmList(CD4~Time|ID, subset=Smoke==0, data=BMACS)
head(coef(lr0))
```

```
##      (Intercept)      Time
## 1022    25.76424   -4.2798248
## 1049    39.30046   -4.3893884
## 1120    43.61689   -0.2771174
## 1212    25.66667  -13.3333333
## 1219    41.45794   -6.3382353
## 1235    37.52175   -8.2307977
```

```
summary(coef(lr0))
```

```
##      (Intercept)          Time
## Min.      :-5.00    Min.      :-32.500
## 1st Qu. :27.32    1st Qu. : -4.766
## Median :34.18    Median : -2.550
## Mean      :34.73    Mean      : -3.210
## 3rd Qu. :41.22    3rd Qu. : -0.409
## Max.      :69.75    Max.      : 36.667
##              NA's    :14
```

```
summary(coef(lr1))
```

```
##      (Intercept)          Time
## Min.      : 5.20    Min.      :-33.0000
## 1st Qu. :30.37    1st Qu. : -5.8409
## Median :36.74    Median : -2.8596
## Mean      :37.41    Mean      : -3.8083
## 3rd Qu. :43.22    3rd Qu. : -0.3479
## Max.      :83.03    Max.      : 18.4000
##              NA's    :14
```

## 2. Fit Models

Define:

$y_{ij}$ : The post CD4 for the  $i$ th patient in the  $j$ th measurement.

$t_{ij}$ : Time of the  $j$ th measurement for the  $i$ th patient

$s_i = s_{ij}$ : Smoking status for the  $i$ th patient

$p_i = p_{ij}$ : PreCD4 for the  $i$ th patient

### 2.1 First Task: post-infection CD4 as a function of time

First, we consider a simple linear mixed-effects model of post-infection CD4 as a function of time since HIV-infection and use the two-stage modeling approach.

Model Expression:

First stage:

$$y_{ij} = \alpha_{1i} + \alpha_{2i}t_{ij} + \epsilon_{ij}, j = 1, 2, \dots, n_i$$

Second stage:

$$\alpha_{1i} = b_{1i} + \beta_1' Z_i$$

$$\alpha_{2i} = b_{2i} + \beta_2' Z_i$$

$\alpha_{1i}$  is the intercept,  $\alpha_{2i}$  is the slope.

$\beta_1$  and  $\beta_2$  are the fixed coefficient vectors.

$Z_i$  is the baseline covariates matrix for the  $i$ th patient

$\epsilon_{ij} \sim N(0, \sigma^2)$  is the measurement error ##### 2.1.1 Only random intercept ##### (a) Smoking status only Model

Expression:

First stage:

$$y_{ij} = \alpha_{1i} + \alpha_2 t_{ij} + \epsilon_{ij}, j = 1, 2, \dots, n_i$$

Second stage:

$$\begin{aligned}\alpha_{1i} &= \beta_0 + b_{1i} + \beta_1 s_i \\ \alpha_{2i} &= \beta_2\end{aligned}$$

Combination:

$$y_{ij} = \beta_0 + b_{1i} + \beta_1 s_i + \beta_2 t_{ij} + \epsilon_{ij}, j = 1, 2, \dots, n_i$$

$s_i$  is the smoking status for the  $i$ th patient.

$\beta_0, \beta_1, \beta_2$  are the fixed effect vector.

$\beta_0$  is the average post-CD4 level at the beginning time for non-smokers.

$\beta_0 + \beta_1$  is the average post-CD4 level at the beginning time for smokers.

$\beta_2$  is average slope of time.

$b_{1i} \sim N(0, \sigma_1^2)$  is the random intercept for the  $i$ th patient that deviated from the population average intercept.

```
model2.1.1.a <- lmer(CD4 ~ 1 + Time + Smoke + (1|ID), data = BMACS)
summary(model2.1.1.a)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: CD4 ~ 1 + Time + Smoke + (1 | ID)
## Data: BMACS
##
## REML criterion at convergence: 12548.8
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.7275 -0.5760 -0.0394  0.5680  4.5534
##
## Random effects:
## Groups   Name                Variance Std.Dev.
## ID       (Intercept)         77.28     8.791
## Residual                    40.26     6.345
## Number of obs: 1817, groups: ID, 283
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  34.6481    0.7271  47.649
## Time        -2.6709    0.1076 -24.813
## Smoke         2.0075    1.1595   1.731
##
## Correlation of Fixed Effects:
##      (Intr) Time
## Time  -0.305
## Smoke -0.575  0.021
```

(b) pre-CD4 level only

Combination:

$$y_{ij} = \beta_0 + b_{1i} + \beta_1 p_i + \beta_2 t_{ij} + \epsilon_{ij}, j = 1, 2, \dots, n_i$$

$\pi_i$  is the PreCD4 for the  $i$ th patient.

$\beta_0, \beta_1, \beta_2$  are the fixed effect vector.

$\beta_0$  is the post-CD4 level at the beginning time.

$\beta_1$  is the average slope of preCD4 level at the beginning time.

$\beta_2$  is the average slope of time.

```
model2.1.1.b <- lmer(CD4 ~ 1 + Time + preCD4 + (1|ID), data = BMACS)
summary(model2.1.1.b)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: CD4 ~ 1 + Time + preCD4 + (1 | ID)
## Data: BMACS
##
## REML criterion at convergence: 12521.5
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.8120 -0.5762 -0.0281  0.5802  4.4741
##
## Random effects:
## Groups   Name      Variance Std.Dev.
## ID       (Intercept) 67.90    8.240
## Residual                40.24    6.343
## Number of obs: 1817, groups: ID, 283
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) 18.53122    2.78267    6.66
## Time        -2.66454    0.10740   -24.81
## preCD4       0.39204    0.06344    6.18
##
## Correlation of Fixed Effects:
##      (Intr) Time
## Time   -0.089
## preCD4 -0.979  0.012
```

### (c) both smoking status and pre-CD4

Combination:

$$y_{ij} = \beta_0 + b_{0i} + \beta_1 s_i + \beta_2 p_i + \beta_3 t_{ij} + \epsilon_{ij}, \quad j = 1, 2, \dots, n_i$$

$\beta_0, \beta_1, \beta_2, \beta_3$  are the fixed effect vector.

```
model2.1.1.c <- lmer(CD4 ~ 1 + Time + Smoke + preCD4 + (1|ID), data = BMACS)
summary(model2.1.1.c)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: CD4 ~ 1 + Time + Smoke + preCD4 + (1 | ID)
## Data: BMACS
##
## REML criterion at convergence: 12519
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.8048 -0.5761 -0.0262  0.5793  4.4790
##
## Random effects:
## Groups   Name      Variance Std.Dev.
## ID       (Intercept) 68.03    8.248
## Residual                40.24    6.343
## Number of obs: 1817, groups: ID, 283
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) 18.60160    2.78678   6.675
## Time        -2.66302    0.10743  -24.790
## Smoke         0.81423    1.11434   0.731
## preCD4        0.38357    0.06454   5.943
##
## Correlation of Fixed Effects:
##              (Intr) Time   Smoke
## Time      -0.088
## Smoke      0.035  0.020
## preCD4    -0.969  0.008 -0.180
```

#### (d) Summary of only random intercept

model2.1.1.c was the best model because it had the lowest AIC and BIC. `summary(model2.1.1.c)[['call']]`

```
anova(model2.1.1.a, model2.1.1.b, model2.1.1.c, refit=F)
```

```
## Data: BMACS
## Models:
## model2.1.1.a: CD4 ~ 1 + Time + Smoke + (1 | ID)
## model2.1.1.b: CD4 ~ 1 + Time + preCD4 + (1 | ID)
## model2.1.1.c: CD4 ~ 1 + Time + Smoke + preCD4 + (1 | ID)
##
##              Df   AIC   BIC logLik deviance   Chisq Chi Df Pr(>Chisq)
## model2.1.1.a   5 12559 12586 -6274.4    12549
## model2.1.1.b   5 12532 12559 -6260.8    12522 27.2425      0    <2e-16 ***
## model2.1.1.c   6 12531 12564 -6259.5    12519  2.5879      1    0.1077
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

### 2.1.2 Only random slope

#### (a) Smoking status only

Model Expression:

First stage:

$$y_{ij} = \alpha_{1i} + \alpha_{2i}t_{ij} + \epsilon_{ij}, j = 1, 2, \dots, n_i$$

Second stage:



$$\alpha_{1i} = \beta_0$$

$$\alpha_{2i} = b_{1i} + \beta_1 + \beta_2 s_i$$

Combination:

$$y_{ij} = \beta_0 + (\beta_1 + b_{1i})t_{ij} + \beta_2 s_i t_{ij} + \epsilon_{ij}, j = 1, 2, \dots, n_i$$

$\beta_0, \beta_1, \beta_2$  are the fixed effect vector.

$\beta_0$  is the average post-CD4 level at the beginning time.

$\beta_1$  is the average slope of time for non-smokers.

$\beta_1 + \beta_2$  is the average slope of time for smokers.

```
model2.1.2.a <- lmer(CD4 ~ 1 + Time + Smoke + (Time-1|ID), data = BMACS)
summary(model2.1.2.a)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: CD4 ~ 1 + Time + Smoke + (Time - 1 | ID)
## Data: BMACS
##
## REML criterion at convergence: 12768.4
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.7788 -0.5584 -0.0119  0.5414  3.8732
##
## Random effects:
## Groups   Name  Variance Std.Dev.
## ID       Time  13.62    3.691
## Residual    47.58    6.898
## Number of obs: 1817, groups: ID, 283
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  35.3590    0.3740  94.538
## Time        -3.1370    0.2858 -10.976
## Smoke         1.6504    0.6026   2.739
##
## Correlation of Fixed Effects:
##      (Intr) Time
## Time  -0.368
## Smoke -0.524 -0.035
```

## (b) pre-CD4 level only

Combination:

$$y_{ij} = \beta_0 + (\beta_1 + b_{1i})t_{ij} + \beta_2 p_i t_{ij} + \epsilon_{ij}, j = 1, 2, \dots, n_i$$

$\beta_0, \beta_1, \beta_2$  are the fixed effect vector.

$\beta_0$  is the average post-CD4 level at the beginning time.

$\beta_1$  is the average slope of time.

$\beta_2$  is the average slope of the interaction term of time and preCD4 level.

```
model2.1.2.b <- lmer(CD4 ~ 1 + Time + preCD4 + (Time-1|ID), data = BMACS)
summary(model2.1.2.b)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: CD4 ~ 1 + Time + preCD4 + (Time - 1 | ID)
## Data: BMACS
##
## REML criterion at convergence: 12578.5
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.8322 -0.5637 -0.0120  0.5558  3.9629
##
## Random effects:
## Groups   Name Variance Std.Dev.
## ID       Time 13.64    3.693
## Residual    42.11    6.489
## Number of obs: 1817, groups: ID, 283
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) 15.7190    1.4062   11.18
## Time        -3.2288    0.2813  -11.48
## preCD4       0.4732    0.0322   14.70
##
## Correlation of Fixed Effects:
##      (Intr) Time
## Time   -0.068
## preCD4 -0.977 -0.027
```

### (c) both smoking status and pre-CD4

Combination:

$$y_{ij} = \beta_0 + (\beta_1 + b_{1i})t_{ij} + \beta_2 p_i t_{ij} + \epsilon_{ij}, j = 1, 2, \dots, n_i$$

$\beta_0, \beta_1, \beta_2$  are the fixed effect vector.

$\beta_0$  is the average post-CD4 level at the beginning time.

$\beta_1$  is the average slope of time.

$\beta_2$  is the average slope of the interaction term of time and preCD4 level.

```
model2.1.2.c <- lmer(CD4 ~ 1 + Time + Smoke + preCD4 + (Time-1|ID), data = BMACS)
summary(model2.1.2.c)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: CD4 ~ 1 + Time + Smoke + preCD4 + (Time - 1 | ID)
## Data: BMACS
##
## REML criterion at convergence: 12577.4
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.8255 -0.5694 -0.0135  0.5522  3.9447
##
## Random effects:
## Groups   Name Variance Std.Dev.
## ID       Time 13.66    3.696
## Residual    42.11    6.490
## Number of obs: 1817, groups: ID, 283
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) 15.73525    1.40687  11.185
## Time        -3.23447    0.28168 -11.483
## Smoke         0.34161    0.58042   0.589
## preCD4       0.47019    0.03261  14.417
##
## Correlation of Fixed Effects:
##      (Intr) Time  Smoke
## Time  -0.068
## Smoke   0.022 -0.033
## preCD4 -0.968 -0.021 -0.157
```

#### (d) Summary of only random slope

model2.1.2.b was the best model because it had the lowest AIC and BIC. `summary(model2.1.2.b)[['call']]`

```
anova(model2.1.2.a, model2.1.2.b, model2.1.2.c, refit=F)
```

```
## Data: BMACS
## Models:
## model2.1.2.a: CD4 ~ 1 + Time + Smoke + (Time - 1 | ID)
## model2.1.2.b: CD4 ~ 1 + Time + preCD4 + (Time - 1 | ID)
## model2.1.2.c: CD4 ~ 1 + Time + Smoke + preCD4 + (Time - 1 | ID)
##
##      Df   AIC   BIC logLik deviance   Chisq Chi Df Pr(>Chisq)
## model2.1.2.a  5 12778 12806 -6384.2    12768
## model2.1.2.b  5 12588 12616 -6289.3    12578 189.9289      0    <2e-16
## model2.1.2.c  6 12589 12622 -6288.7    12577   1.0959      1    0.2952
##
## model2.1.2.a
## model2.1.2.b ***
## model2.1.2.c
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

### 2.1.3 Both random intercept and random slope:

#### (a) Smoking status only

Model Expression:

First stage:

$$y_{ij} = \alpha_{1i} + \alpha_{2i}t_{ij} + \epsilon_{ij}, \quad j = 1, 2, \dots, n_i$$

Second stage:

$$\alpha_{1i} = \beta_0 + b_{1i} + \beta_1 s_i$$

$$\alpha_{2i} = \beta_2 + b_{2i} + \beta_3 s_i$$

Combination:

$$y_{ij} = \beta_0 + b_{1i} + \beta_1 s_i + (\beta_2 + b_{2i})t_{ij} + \beta_3 s_i t_{ij} + \epsilon_{ij}, j = 1, 2, \dots, n_i$$

$\beta_0, \beta_1, \beta_2, \beta_3$  are the fixed effect vector.

$\beta_0$  is the average post-CD4 level at the beginning time for non-smokers.

$\beta_0 + \beta_1$  is the average post-CD4 level at the beginning time for smokers.

$\beta_2$  is the average slope of time for non-smokers

$\beta_2 + \beta_3$  is the average slope of time for smokers.

$(b_{1i}, b_{2i})^T \sim N(0, (\sigma_{21}, \sigma_{12}, \sigma_{21}, \sigma_{22}))$  is the random deviations for each patient from the population's average intercept.

```
model2.1.3.a <- lmer(CD4 ~ 1 + Time + Smoke + (Time|ID), data = BMACS)
summary(model2.1.3.a)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: CD4 ~ 1 + Time + Smoke + (Time | ID)
## Data: BMACS
##
## REML criterion at convergence: 12149
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -4.1993 -0.5481 -0.0219  0.5205  4.2324
##
## Random effects:
## Groups   Name                Variance Std.Dev. Corr
## ID       (Intercept)  77.438     8.800
##          Time         9.329     3.054   -0.32
## Residual                    25.031     5.003
## Number of obs: 1817, groups: ID, 283
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  35.0357     0.7060  49.624
## Time        -3.0826     0.2355 -13.090
## Smoke         1.9986     1.1017   1.814
##
## Correlation of Fixed Effects:
##      (Intr) Time
## Time  -0.359
## Smoke -0.558 -0.002
```

(b) pre-CD4 level only

Combination:

$$y_{ij} = \beta_0 + b_{1i} + \beta_1 p_i + (\beta_2 + b_{2i})t_{ij} + \beta_3 p_i t_{ij} + \epsilon_{ij}, j = 1, 2, \dots, n_i$$

```
model2.1.3.b <- lmer(CD4 ~ 1 + Time + preCD4 + (Time|ID), data = BMACS)
summary(model2.1.3.b)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: CD4 ~ 1 + Time + preCD4 + (Time | ID)
## Data: BMACS
##
## REML criterion at convergence: 12102.7
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -4.2278 -0.5437 -0.0141  0.5250  4.3257
##
## Random effects:
## Groups   Name                Variance Std.Dev. Corr
## ID       (Intercept)  58.590     7.654
##          Time           9.428     3.070   -0.25
## Residual                24.989     4.999
## Number of obs: 1817, groups: ID, 283
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) 15.88791    2.54974    6.231
## Time        -3.10264    0.23641   -13.124
## preCD4       0.46352    0.05816    7.970
##
## Correlation of Fixed Effects:
##          (Intr) Time
## Time    -0.078
## preCD4  -0.979 -0.004
```

### (c) both smoking status and pre-CD4

Combination:

$$y_{ij} = \beta_0 + b_{1i} + \beta_1 s_i + \beta_2 p_i + (\beta_3 + b_{2i})t_{ij} + \beta_4 s_i t_{ij} + \beta_5 p_i t_{ij} + \epsilon_{ij}, j = 1, 2, \dots, n_i$$

```
model2.1.3.c <- lmer(CD4 ~ 1 + Time + Smoke + preCD4 + (Time|ID), data = BMACS)
summary(model2.1.3.c)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: CD4 ~ 1 + Time + Smoke + preCD4 + (Time | ID)
## Data: BMACS
##
## REML criterion at convergence: 12100.4
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -4.2258 -0.5449 -0.0140  0.5258  4.3168
##
## Random effects:
## Groups   Name                Variance Std.Dev. Corr
## ID       (Intercept)    58.743     7.664
##          Time           9.436     3.072   -0.25
## Residual                    24.985     4.998
## Number of obs: 1817, groups: ID, 283
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) 15.93566    2.55409   6.239
## Time        -3.10390    0.23650 -13.124
## Smoke         0.65610    1.02195   0.642
## preCD4       0.45699    0.05913   7.729
##
## Correlation of Fixed Effects:
##          (Intr) Time   Smoke
## Time    -0.078
## Smoke    0.031 -0.008
## preCD4  -0.969 -0.003 -0.174
```

## (d) Summary of random intercept and random slope

model2.1.3.b was the best model because it had the lowest BIC and high significance level.

`summary(model2.1.3.b)[['call']]`

```
anova(model2.1.3.a, model2.1.3.b, model2.1.3.c, refit=F)
```

```
## Data: BMACS
## Models:
## model2.1.3.a: CD4 ~ 1 + Time + Smoke + (Time | ID)
## model2.1.3.b: CD4 ~ 1 + Time + preCD4 + (Time | ID)
## model2.1.3.c: CD4 ~ 1 + Time + Smoke + preCD4 + (Time | ID)
##              Df    AIC    BIC logLik deviance   Chisq Chi Df Pr(>Chisq)
## model2.1.3.a   7 12163 12202 -6074.5    12149
## model2.1.3.b   7 12117 12155 -6051.4    12103 46.2838     0    <2e-16 ***
## model2.1.3.c   8 12116 12160 -6050.2    12100  2.2926     1      0.13
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## 2.2 Second Task

To allow some flexibility, we next consider the linear mixed-effect model of post-CD4 as a polynomial function of time since HIV-infection. Using the two-stage modeling approach

```
BMACS$Time2=BMACS$Time^2
```

Model Expression:

First stage:

$$y_{ij} = \alpha_{1i} + \alpha_{2i}t_{ij} + \alpha_{3i}t_{ij}^2 + \epsilon_{ij}, j = 1, 2, \dots, n_i$$

Second stage:

$$\alpha_{1i} = b_{1i} + \beta'_1 Z_i$$

$$\alpha_{2i} = b_{2i} + \beta'_2 Z_i$$

$$\alpha_{3i} = b_{3i} + \beta'_3 Z_i$$

#### 2.2.1 Only random intercept #### (a) Smoking status only

First stage:

$$y_{ij} = \alpha_{1i} + \alpha_{2i}t_{ij} + \alpha_{3i}t_{ij}^2 + \epsilon_{ij}, j = 1, 2, \dots, n_i$$

Second stage:

$$\alpha_{1i} = \beta_0 + b_{1i} + \beta_1 s_i$$

$$\alpha_{2i} = \beta_2$$

$$\alpha_{3i} = \beta_3$$

Combination:

$$y_{ij} = \beta_0 + b_{1i} + \beta_1 s_i + \beta_2 t_{ij} + \beta_3 t_{ij}^2 + \epsilon_{ij}, j = 1, 2, \dots, n_i$$

$\beta_3$  is average slope of time's square.

```
model2.2.1.a=lmer(CD4~1 + Time + Time2 + Smoke + (1|ID), data = BMACS)
summary(model2.2.1.a)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: CD4 ~ 1 + Time + Time2 + Smoke + (1 | ID)
## Data: BMACS
##
## REML criterion at convergence: 12521.2
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.8791 -0.5602 -0.0238  0.5537  4.3608
##
## Random effects:
## Groups   Name      Variance Std.Dev.
## ID       (Intercept) 76.96    8.773
## Residual                39.52    6.286
## Number of obs: 1817, groups: ID, 283
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) 36.31304    0.78314  46.369
## Time        -4.63168    0.36548 -12.673
## Time2         0.36336    0.06479   5.608
## Smoke         1.97176    1.15622   1.705
##
## Correlation of Fixed Effects:
##      (Intr) Time   Time2
## Time  -0.444
## Time2  0.379 -0.956
## Smoke -0.535  0.011 -0.006
```

## (b) pre-CD4 level only

Combination:

$$y_{ij} = \beta_0 + b_{1i} + \beta_1 p_i + \beta_2 t_{ij} + \beta_3 t_{ij}^2 + \epsilon_{ij}, j = 1, 2, \dots, n_i$$

```
model2.2.1.b <- lmer(CD4 ~ 1 + Time + Time2 + preCD4 + (1|ID), data = BMACS)
summary(model2.2.1.b)
```



```
## Linear mixed model fit by REML ['lmerMod']
## Formula: CD4 ~ 1 + Time + Time2 + preCD4 + (1 | ID)
## Data: BMACS
##
## REML criterion at convergence: 12493.2
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.9032 -0.5589 -0.0127  0.5666  4.2790
##
## Random effects:
## Groups   Name            Variance Std.Dev.
## ID       (Intercept)    67.42     8.211
## Residual                    39.49     6.284
## Number of obs: 1817, groups: ID, 283
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) 20.09881    2.78464   7.218
## Time        -4.64647    0.36478 -12.738
## Time2         0.36744    0.06470   5.679
## preCD4        0.39446    0.06317   6.244
##
## Correlation of Fixed Effects:
##      (Intr) Time   Time2
## Time   -0.121
## Time2   0.099 -0.956
## preCD4 -0.974 -0.003  0.007
```

### (c) both smoking status and pre-CD4

Combination:

$$y_{ij} = \beta_0 + b_{1i} + \beta_1 s_i + \beta_2 p_i + \beta_3 t_{ij} + \beta_4 t_{ij}^2 + \epsilon_{ij}, j = 1, 2, \dots, n_i$$

```
model2.2.1.c <- lmer(CD4 ~ 1 + Time + Time2 + Smoke + preCD4 + (1|ID), data = BMACS)
summary(model2.2.1.c)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: CD4 ~ 1 + Time + Time2 + Smoke + preCD4 + (1 | ID)
## Data: BMACS
##
## REML criterion at convergence: 12490.7
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.9070 -0.5569 -0.0135  0.5674  4.2839
##
## Random effects:
## Groups   Name                Variance Std.Dev.
## ID       (Intercept)         67.56    8.220
## Residual                    39.49    6.284
## Number of obs: 1817, groups: ID, 283
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) 20.16378    2.78888   7.230
## Time        -4.64325    0.36482  -12.727
## Time2         0.36711    0.06470   5.674
## Smoke         0.76977    1.10968   0.694
## preCD4        0.38646    0.06428   6.012
##
## Correlation of Fixed Effects:
##          (Intr) Time   Time2  Smoke
## Time    -0.120
## Time2    0.099 -0.956
## Smoke    0.034  0.013 -0.007
## preCD4  -0.963 -0.005  0.008 -0.180
```

#### (d) Summary of only random intercept

model2.2.1.b was the best model because it had the lowest AIC, BIC and high significance level.

`summary(model2.2.1.b)[['call']]`

```
anova(model2.2.1.a, model2.2.1.b, model2.2.1.c, refit=F)
```

```
## Data: BMACS
## Models:
## model2.2.1.a: CD4 ~ 1 + Time + Time2 + Smoke + (1 | ID)
## model2.2.1.b: CD4 ~ 1 + Time + Time2 + preCD4 + (1 | ID)
## model2.2.1.c: CD4 ~ 1 + Time + Time2 + Smoke + preCD4 + (1 | ID)
##              Df    AIC    BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
## model2.2.1.a  6 12533 12566 -6260.6    12521
## model2.2.1.b  6 12505 12538 -6246.6    12493 28.0355      0    <2e-16 ***
## model2.2.1.c  7 12505 12543 -6245.3    12491  2.5267      1    0.1119
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

### 2.2.2 Only random slope

#### (a) Smoking status only

First stage:

$$y_{ij} = \alpha_{1i} + \alpha_{2i}t_{ij} + \alpha_{3i}t_{ij}^2 + \epsilon_{ij}, j = 1, 2, \dots, n_i$$

Second stage:

$$\begin{aligned}\alpha_{1i} &= \beta_0 \\ \alpha_{2i} &= \beta_1 + b_{1i} + \beta_2 s_i \\ \alpha_{3i} &= \beta_3\end{aligned}$$

Combination:

$$y_{ij} = \beta_0 + (\beta_1 + b_{1i})t_{ij} + \beta_2 s_i t_{ij} + \beta_3 t_{ij}^2 + \epsilon_{ij}, \quad j = 1, 2, \dots, n_i$$

```
model2.2.2.a <- lmer(CD4 ~ 1 + Time + Time2 + Smoke:Time + (Time-1|ID), data = BMACS)
summary(model2.2.2.a)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: CD4 ~ 1 + Time + Time2 + Smoke:Time + (Time - 1 | ID)
## Data: BMACS
##
## REML criterion at convergence: 12761.3
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.9235 -0.5453 -0.0037  0.5383  3.6726
##
## Random effects:
## Groups   Name Variance Std.Dev.
## ID       Time 13.39    3.660
## Residual    47.40    6.885
## Number of obs: 1817, groups: ID, 283
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) 37.11488    0.43959  84.430
## Time        -4.85800    0.50040  -9.708
## Time2         0.30533    0.07516   4.062
## Time:Smoke   0.74348    0.53306   1.395
##
## Correlation of Fixed Effects:
##              (Intr) Time   Time2
## Time        -0.693
## Time2        0.690 -0.743
## Time:Smoke  -0.011 -0.365  0.012
```

(b) pre-CD4 level only

Combination:

$$y_{ij} = \beta_0 + (\beta_1 + b_{1i})t_{ij} + \beta_2 p_i t_{ij} + \beta_3 t_{ij}^2 + \epsilon_{ij}, \quad j = 1, 2, \dots, n_i$$

```
model2.2.2.b <- lmer(CD4 ~ 1 + Time + Time2 + preCD4:Time + (Time-1|ID), data = BMACS)
summary(model2.2.2.b)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: CD4 ~ 1 + Time + Time2 + preCD4:Time + (Time - 1 | ID)
## Data: BMACS
##
## REML criterion at convergence: 12749.4
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.9547 -0.5435  0.0054  0.5341  3.6798
##
## Random effects:
## Groups   Name Variance Std.Dev.
## ID       Time 12.57    3.545
## Residual    47.27    6.875
## Number of obs: 1817, groups: ID, 283
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  37.10536    0.43877  84.567
## Time        -10.26701    1.35066  -7.601
## Time2         0.31054    0.07501   4.140
## Time:preCD4   0.13210    0.02962   4.459
##
## Correlation of Fixed Effects:
##              (Intr) Time   Time2
## Time        -0.249
## Time2        0.691 -0.287
## Time:preCD4 -0.009 -0.940  0.015
```

### (c) both smoking status and pre-CD4

Combination:

$$y_{ij} = \beta_0 + (\beta_1 + b_{1i})t_{ij} + \beta_2 s_i t_{ij} + \beta_3 p_i t_{ij} + \beta_4 t_{ij}^2 + \epsilon_{ij}, \quad j = 1, 2, \dots, n_i$$

```
model2.2.2.c <- lmer(CD4 ~ 1 + Time + Time2 + Smoke:Time + preCD4:Time + (Time-1|ID), data = BMACS)
summary(model2.2.2.c)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: CD4 ~ 1 + Time + Time2 + Smoke:Time + preCD4:Time + (Time - 1 |
##      ID)
##      Data: BMACS
##
## REML criterion at convergence: 12748.6
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.9583 -0.5424  0.0043  0.5336  3.6763
##
## Random effects:
##      Groups   Name Variance Std.Dev.
##      ID       Time 12.62    3.553
##      Residual    47.26    6.875
## Number of obs: 1817, groups:  ID, 283
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  37.10292    0.43878  84.558
## Time        -10.23002    1.35483  -7.551
## Time2         0.31084    0.07502   4.144
## Time:Smoke    0.29614    0.53003   0.559
## Time:preCD4   0.12888    0.03027   4.258
##
## Correlation of Fixed Effects:
##              (Intr) Time   Time2  Tm:Smk
## Time         -0.249
## Time2         0.690 -0.286
## Time:Smoke   -0.010  0.054  0.010
## Time:preCD4 -0.007 -0.931  0.013 -0.196
```

#### (d) Summary of only random slope

model2.2.2.b was the best model because it had the lowest AIC, BIC and high significance level.

`summary(model2.2.2.b)[['call']]`

```
anova(model2.2.2.a, model2.2.2.b, model2.2.2.c, refit=F)
```

```
## Data: BMACS
## Models:
## model2.2.2.a: CD4 ~ 1 + Time + Time2 + Smoke:Time + (Time - 1 | ID)
## model2.2.2.b: CD4 ~ 1 + Time + Time2 + preCD4:Time + (Time - 1 | ID)
## model2.2.2.c: CD4 ~ 1 + Time + Time2 + Smoke:Time + preCD4:Time + (Time - 1 |
## model2.2.2.c:      ID)
##              Df   AIC   BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
## model2.2.2.a   6 12773 12806 -6380.6   12761
## model2.2.2.b   6 12761 12794 -6374.7   12749 11.8404     0    <2e-16 ***
## model2.2.2.c   7 12763 12801 -6374.3   12749  0.8793     1    0.3484
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

### 2.2.3 Both random intercept and random slope:

#### (a) Smoking status only

Model Expression:

First stage:

$$y_{ij} = \alpha_{1i} + \alpha_{2i}t_{ij} + \alpha_{3i}t_{ij}^2 + \epsilon_{ij}, j = 1, 2, \dots, n_i$$

Second stage:

$$\begin{aligned}\alpha_{1i} &= \beta_0 + b_{1i} + \beta_1 s_i \\ \alpha_{2i} &= \beta_2 + b_{2i} + \beta_3 s_i \\ \alpha_{3i} &= \beta_4\end{aligned}$$

Combination:

$$y_{ij} = \beta_0 + b_{1i} + \beta_1 s_i + (\beta_2 + b_{2i})t_{ij} + \beta_3 s_i t_{ij} + \beta_4 t_{ij}^2 + \epsilon_{ij}, j = 1, 2, \dots, n_i$$

```
model2.2.3.a <- lmer(CD4 ~ 1 + Time2 + Smoke*Time + (1 + Time|ID), data = BMACS)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl =
## control$checkConv, : Model failed to converge with max|grad| = 0.0022222
## (tol = 0.002, component 1)
```

```
summary(model2.2.3.a)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: CD4 ~ 1 + Time2 + Smoke * Time + (1 + Time | ID)
## Data: BMACS
##
## REML criterion at convergence: 12118.9
##
## Scaled residuals:
## Min      1Q  Median      3Q      Max
## -4.3295 -0.5443 -0.0195  0.5185  4.3035
##
## Random effects:
## Groups   Name                Variance Std.Dev. Corr
## ID       (Intercept)  77.955     8.829
##          Time          8.987     2.998   -0.32
## Residual                24.564     4.956
## Number of obs: 1817, groups: ID, 283
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  36.2739    0.7643  47.458
## Time2        0.3421    0.0591   5.789
## Smoke        2.0998    1.2303   1.707
## Time       -4.6312    0.4003 -11.570
## Smoke:Time  -0.1724    0.4957  -0.348
##
## Correlation of Fixed Effects:
##              (Intr) Time2  Smoke  Time
## Time2        0.304
## Smoke       -0.573 -0.030
## Time       -0.500 -0.711  0.198
## Smoke:Time  0.242  0.041 -0.443 -0.429
## convergence code: 0
## Model failed to converge with max|grad| = 0.0022222 (tol = 0.002, component 1)
```

## (b) pre-CD4 level only

Combination:

$$y_{ij} = \beta_0 + b_{1i} + \beta_1 p_i + (\beta_2 + b_{2i})t_{ij} + \beta_3 p_i t_{ij} + \beta_4 t_{ij}^2 + \epsilon_{ij}, j = 1, 2, \dots, n_i$$

```
model2.2.3.b <- lmer(CD4 ~ 1 + Time2 + preCD4*Time + (1 + Time|ID), data = BMACS)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl =  
## control$checkConv, : Model failed to converge with max|grad| = 0.0554679  
## (tol = 0.002, component 1)
```

```
summary(model2.2.3.b)
```

```
## Linear mixed model fit by REML ['lmerMod']  
## Formula: CD4 ~ 1 + Time2 + preCD4 * Time + (1 + Time | ID)  
## Data: BMACS  
##  
## REML criterion at convergence: 12070.7  
##  
## Scaled residuals:  
##      Min       1Q   Median       3Q      Max   
## -4.3204 -0.5418 -0.0060  0.5281  4.3904   
##  
## Random effects:  
## Groups   Name                Variance Std.Dev. Corr   
## ID       (Intercept)  59.756     7.730   
##          Time         8.804     2.967   -0.26   
## Residual                24.474     4.947   
## Number of obs: 1817, groups: ID, 283  
##  
## Fixed effects:  
##              Estimate Std. Error t value   
## (Intercept)  14.25482    2.78433   5.120   
## Time2         0.33253    0.05873   5.662   
## preCD4        0.53085    0.06350   8.360   
## Time        -1.22419    1.23687  -0.990   
## preCD4:Time -0.08064    0.02756  -2.926   
##  
## Correlation of Fixed Effects:  
##              (Intr) Time2  preCD4 Time   
## Time2         0.099   
## preCD4       -0.979 -0.021   
## Time        -0.405 -0.258  0.382   
## preCD4:Time  0.389  0.036 -0.398 -0.957   
## convergence code: 0  
## Model failed to converge with max|grad| = 0.0554679 (tol = 0.002, component 1)
```

## (c) both smoking status and pre-CD4

Combination:

$$y_{ij} = \beta_0 + b_{1i} + \beta_1 s_i + \beta_2 p_i + (\beta_3 + b_{2i})t_{ij} + \beta_4 s_i t_{ij} + \beta_5 p_i t_{ij} + \beta_5 t_{ij}^2 + \epsilon_{ij}, j = 1, 2, \dots, n_i$$

```
model2.2.3.c <- lmer(CD4 ~ 1 + Time2 + Smoke*Time + preCD4*Time + (1 + Time|ID), data = BMACS)
summary(model2.2.3.c)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: CD4 ~ 1 + Time2 + Smoke * Time + preCD4 * Time + (1 + Time |
##      ID)
##      Data: BMACS
##
## REML criterion at convergence: 12068
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -4.3191 -0.5410 -0.0096  0.5268  4.3825
##
## Random effects:
##      Groups   Name                Variance Std.Dev. Corr
##      ID       (Intercept)  59.803     7.733
##              Time          8.913     2.986   -0.26
##      Residual                24.464     4.946
## Number of obs: 1817, groups:  ID, 283
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) 14.27936    2.78603   5.125
## Time2        0.33173    0.05879   5.643
## Smoke        0.59713    1.12233   0.532
## Time       -1.19910    1.24307  -0.965
## preCD4       0.52529    0.06436   8.162
## Smoke:Time   0.01747    0.50123   0.035
## Time:preCD4 -0.08133    0.02811  -2.893
##
## Correlation of Fixed Effects:
##              (Intr) Time2  Smoke  Time  preCD4 Smk:Tm
## Time2         0.099
## Smoke         0.016 -0.031
## Time        -0.404 -0.256  0.009
## preCD4       -0.969 -0.016 -0.160  0.375
## Smoke:Time   0.006  0.036 -0.412  0.025  0.051
## Time:preCD4  0.382  0.029  0.052 -0.947 -0.393 -0.170
```

#### (d) Summary of random intercept and random slope

model2.2.3.b was the best model because it had the lowest AIC, BIC and high significance level.

```
summary(model2.2.3.b)[['call']]
```

```
anova(model2.2.3.a, model2.2.3.b, model2.2.3.c, refit=F)
```



```
## Data: BMACS
## Models:
## model2.2.3.a: CD4 ~ 1 + Time2 + Smoke * Time + (1 + Time | ID)
## model2.2.3.b: CD4 ~ 1 + Time2 + preCD4 * Time + (1 + Time | ID)
## model2.2.3.c: CD4 ~ 1 + Time2 + Smoke * Time + preCD4 * Time + (1 + Time |
## model2.2.3.c: ID)
##           Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## model2.2.3.a  9 12137 12186 -6059.4    12119
## model2.2.3.b  9 12089 12138 -6035.4    12071 48.133      0    <2e-16 ***
## model2.2.3.c 11 12090 12151 -6034.0    12068  2.698      2    0.2595
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## 2.2.4 Random intercept, random slope, and random coefficient for the second polynomial term

### (a) Smoking status only

Model Expression:

First stage:

$$y_{ij} = \alpha_{1i} + \alpha_{2i}t_{ij} + \alpha_{3i}t_{ij}^2 + \epsilon_{ij}, j = 1, 2, \dots, n_i$$

Second stage:

$$\alpha_{1i} = \beta_0 + b_{1i} + \beta_1 s_i$$

$$\alpha_{2i} = \beta_2 + b_{2i} + \beta_3 s_i$$

$$\alpha_{3i} = \beta_4 + b_{3i} + \beta_5 s_i$$

Combination:

$$y_{ij} = \beta_0 + b_{1i} + \beta_1 s_i + (\beta_2 + b_{2i})t_{ij} + \beta_3 s_i t_{ij} + (\beta_4 + b_{3i})t_{ij}^2 + \beta_5 s_i t_{ij}^2 + \epsilon_{ij}$$

```
model2.2.4.a <- lmer(CD4 ~ 1 + Smoke*Time2 + Smoke*Time + (1 + Time + Time2|ID), data = BMACS)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl =
## control$checkConv, : Model failed to converge with max|grad| = 0.0105201
## (tol = 0.002, component 1)
```

```
summary(model2.2.4.a)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: CD4 ~ 1 + Smoke * Time2 + Smoke * Time + (1 + Time + Time2 |
##      ID)
##      Data: BMACS
##
## REML criterion at convergence: 12087.9
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -4.3769 -0.5345 -0.0335  0.5142  4.4426
##
## Random effects:
##      Groups   Name                Variance Std.Dev. Corr
##      ID       (Intercept) 87.0701  9.3311
##              Time       27.8373  5.2761  -0.44
##              Time2      0.4494  0.6704   0.32 -0.87
## Residual                22.6486  4.7591
## Number of obs: 1817, groups: ID, 283
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) 36.40991    0.81402  44.729
## Smoke       1.89581    1.37245   1.381
## Time2       0.37759    0.09173   4.116
## Time      -4.79951    0.56722  -8.461
## Smoke:Time2 -0.02704    0.16677  -0.162
## Smoke:Time  0.04629    0.98965   0.047
##
## Correlation of Fixed Effects:
##              (Intr) Smoke  Time2  Time  Smk:T2
## Smoke       -0.593
## Time2       0.452 -0.268
## Time      -0.578  0.343 -0.882
## Smoke:Time2 -0.248  0.447 -0.550  0.485
## Smoke:Time  0.332 -0.585  0.505 -0.573 -0.875
## convergence code: 0
## Model failed to converge with max|grad| = 0.0105201 (tol = 0.002, component 1)
```

## (b) pre-CD4 level only

Combination:

$$y_{ij} = \beta_0 + b_{1i} + \beta_1 p_i + (\beta_2 + b_{2i})t_{ij} + \beta_3 p_i t_{ij} + (\beta_4 + b_{3i})t_{ij}^2 + \beta_5 p_i t_{ij}^2 + \epsilon_{ij}$$

```
model2.2.4.b <- lmer(CD4 ~ 1 + preCD4*Time2 + preCD4*Time + (1 + Time + Time2|ID), data = BMACS)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl =
## control$checkConv, : Model failed to converge with max|grad| = 0.0236645
## (tol = 0.002, component 1)
```

```
summary(model2.2.4.b)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: CD4 ~ 1 + preCD4 * Time2 + preCD4 * Time + (1 + Time + Time2 |
##      ID)
##      Data: BMACS
##
## REML criterion at convergence: 12042.4
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -4.3830 -0.5321 -0.0169  0.5219  4.5209
##
## Random effects:
##      Groups   Name                Variance Std.Dev. Corr
##      ID       (Intercept)  64.742    8.0462
##              Time         26.690    5.1663  -0.39
##              Time2        0.439    0.6626   0.28 -0.87
## Residual                22.590    4.7529
## Number of obs: 1817, groups: ID, 283
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) 12.063110   3.097507   3.894
## preCD4       0.584212   0.071004   8.228
## Time2       -0.219635   0.383612  -0.573
## Time         1.512948   2.371030   0.638
## preCD4:Time2 0.013845   0.008896   1.556
## preCD4:Time -0.147886   0.054580  -2.710
##
## Correlation of Fixed Effects:
##              (Intr) preCD4 Time2 Time  pCD4:T2
## preCD4       -0.982
## Time2        0.455 -0.449
## Time         -0.570  0.562 -0.878
## preCD4:Tim2 -0.444  0.456 -0.980  0.858
## preCD4:Time  0.559 -0.572  0.864 -0.981 -0.877
## convergence code: 0
## Model failed to converge with max|grad| = 0.0236645 (tol = 0.002, component 1)
```

### (c) both smoking status and pre-CD4

Combination:

$$y_{ij} = \beta_0 + b_{1i} + \beta_1 s_i + \beta_2 p_i + (\beta_3 + b_{2i})t_{ij} + \beta_4 s_i t_{ij} + \beta_5 p_i t_{ij} + (\beta_6 + b_{3i})t_{ij}^2 + \beta_7 s_i t_{ij}^2 + \beta_8 p_i t_{ij}^2 + \epsilon_{ij}$$

```
model2.2.4.c <- lmer(CD4 ~ 1 + preCD4*Time2 + preCD4*Time + Smoke*Time2 + Smoke*Time + (1 + Time + Time2|I
D), data = BMACS)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl =
## control$checkConv, : Model failed to converge with max|grad| = 0.00658647
## (tol = 0.002, component 1)
```

```
summary(model2.2.4.c)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula:
## CD4 ~ 1 + preCD4 * Time2 + preCD4 * Time + Smoke * Time2 + Smoke *
##   Time + (1 + Time + Time2 | ID)
##   Data: BMACS
##
## REML criterion at convergence: 12041.4
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -4.3778 -0.5342 -0.0192  0.5144  4.5095
##
## Random effects:
##   Groups   Name                Variance Std.Dev. Corr
##   ID       (Intercept) 65.1112  8.0691
##           Time      26.9065  5.1871  -0.40
##           Time2      0.4442  0.6665   0.29 -0.87
##   Residual                22.5861  4.7525
## Number of obs: 1817, groups: ID, 283
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) 12.078803   3.104175   3.891
## preCD4       0.581373   0.071982   8.077
## Time2       -0.226996   0.385035  -0.590
## Time        1.577355   2.378964   0.663
## Smoke       0.280275   1.251307   0.224
## preCD4:Time2 0.014435   0.009039   1.597
## preCD4:Time -0.152426   0.055552  -2.744
## Time2:Smoke -0.056490   0.168521  -0.335
## Time:Smoke  0.394455   0.994042   0.397
##
## Correlation of Fixed Effects:
##              (Intr) preCD4 Time2  Time   Smoke  pCD4:T2 prCD4:T Tm2:Sm
## preCD4       -0.972
## Time2        0.455 -0.444
## Time        -0.570  0.555 -0.878
## Smoke        0.009 -0.151  0.001 -0.001
## preCD4:Tim2 -0.439  0.454 -0.971  0.851 -0.065
## preCD4:Time  0.551 -0.569  0.855 -0.972  0.082 -0.878
## Time2:Smoke -0.001 -0.059  0.023 -0.025  0.452 -0.157  0.142
## Time:Smoke  0.000  0.078 -0.026  0.033 -0.575  0.147  -0.171  -0.875
## convergence code: 0
## Model failed to converge with max|grad| = 0.00658647 (tol = 0.002, component 1)
```

#### (d) Summary of only random intercept

model2.2.4.b was the best model because it had the lowest AIC, BIC and high significance level.

summary(model2.2.4.b)[['call']]

```
anova(model2.2.4.a, model2.2.4.b, model2.2.4.c, refit=F)
```

```
## Data: BMACS
## Models:
## model2.2.4.a: CD4 ~ 1 + Smoke * Time2 + Smoke * Time + (1 + Time + Time2 |
## model2.2.4.a:      ID)
## model2.2.4.b: CD4 ~ 1 + preCD4 * Time2 + preCD4 * Time + (1 + Time + Time2 |
## model2.2.4.b:      ID)
## model2.2.4.c: CD4 ~ 1 + preCD4 * Time2 + preCD4 * Time + Smoke * Time2 + Smoke *
## model2.2.4.c:      Time + (1 + Time + Time2 | ID)
##           Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## model2.2.4.a 13 12114 12186 -6044.0    12088
## model2.2.4.b 13 12068 12140 -6021.2    12042 45.522      0    <2e-16 ***
## model2.2.4.c 16 12073 12162 -6020.7    12041  0.986      3    0.8046
## ----
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## 2.3 Third Task: Final Model

Compare the fitness of the above models, and justify the final chosen linear mixedeffects model for the BMACS data.

```
anova(model2.1.1.a, model2.1.1.b, model2.1.1.c, model2.1.2.a, model2.1.2.b, model2.1.2.c, model2.1.3.a, model2.1.
3.b, model2.1.3.c, model2.2.1.a, model2.2.1.b, model2.2.1.c, model2.2.2.a, model2.2.2.b, model2.2.2.c, model2.2.3.
a, model2.2.3.b, model2.2.3.c, model2.2.4.a, model2.2.4.b, model2.2.4.c, refit=F)
```

```
## Data: BMACS
## Models:
## model2.1.1.a: CD4 ~ 1 + Time + Smoke + (1 | ID)
## model2.1.1.b: CD4 ~ 1 + Time + preCD4 + (1 | ID)
## model2.1.2.a: CD4 ~ 1 + Time + Smoke + (Time - 1 | ID)
## model2.1.2.b: CD4 ~ 1 + Time + preCD4 + (Time - 1 | ID)
## model2.1.1.c: CD4 ~ 1 + Time + Smoke + preCD4 + (1 | ID)
## model2.1.2.c: CD4 ~ 1 + Time + Smoke + preCD4 + (Time - 1 | ID)
## model2.2.1.a: CD4 ~ 1 + Time + Time2 + Smoke + (1 | ID)
## model2.2.1.b: CD4 ~ 1 + Time + Time2 + preCD4 + (1 | ID)
## model2.2.2.a: CD4 ~ 1 + Time + Time2 + Smoke:Time + (Time - 1 | ID)
## model2.2.2.b: CD4 ~ 1 + Time + Time2 + preCD4:Time + (Time - 1 | ID)
## model2.1.3.a: CD4 ~ 1 + Time + Smoke + (Time | ID)
## model2.1.3.b: CD4 ~ 1 + Time + preCD4 + (Time | ID)
## model2.2.1.c: CD4 ~ 1 + Time + Time2 + Smoke + preCD4 + (1 | ID)
## model2.2.2.c: CD4 ~ 1 + Time + Time2 + Smoke:Time + preCD4:Time + (Time - 1 |
## model2.2.2.c: ID)
## model2.1.3.c: CD4 ~ 1 + Time + Smoke + preCD4 + (Time | ID)
## model2.2.3.a: CD4 ~ 1 + Time2 + Smoke * Time + (1 + Time | ID)
## model2.2.3.b: CD4 ~ 1 + Time2 + preCD4 * Time + (1 + Time | ID)
## model2.2.3.c: CD4 ~ 1 + Time2 + Smoke * Time + preCD4 * Time + (1 + Time |
## model2.2.3.c: ID)
## model2.2.4.a: CD4 ~ 1 + Smoke * Time2 + Smoke * Time + (1 + Time + Time2 |
## model2.2.4.a: ID)
## model2.2.4.b: CD4 ~ 1 + preCD4 * Time2 + preCD4 * Time + (1 + Time + Time2 |
## model2.2.4.b: ID)
## model2.2.4.c: CD4 ~ 1 + preCD4 * Time2 + preCD4 * Time + Smoke * Time2 + Smoke *
## model2.2.4.c: Time + (1 + Time + Time2 | ID)
##
##           Df    AIC    BIC  logLik deviance   Chisq Chi Df Pr(>Chisq)
## model2.1.1.a  5 12559 12586 -6274.4    12549
## model2.1.1.b  5 12532 12559 -6260.8    12522  27.242    0 < 2.2e-16 ***
## model2.1.2.a  5 12778 12806 -6384.2    12768   0.000    0  1.0000
## model2.1.2.b  5 12588 12616 -6289.3    12578 189.929    0 < 2.2e-16 ***
## model2.1.1.c  6 12531 12564 -6259.5    12519  59.558    1 1.187e-14 ***
## model2.1.2.c  6 12589 12622 -6288.7    12577   0.000    0  1.0000
## model2.2.1.a  6 12533 12566 -6260.6    12521  56.178    0 < 2.2e-16 ***
## model2.2.1.b  6 12505 12538 -6246.6    12493  28.035    0 < 2.2e-16 ***
## model2.2.2.a  6 12773 12806 -6380.6    12761   0.000    0  1.0000
## model2.2.2.b  6 12761 12794 -6374.7    12749  11.840    0 < 2.2e-16 ***
## model2.1.3.a  7 12163 12202 -6074.5    12149 600.459    1 < 2.2e-16 ***
## model2.1.3.b  7 12117 12155 -6051.4    12103  46.284    0 < 2.2e-16 ***
## model2.2.1.c  7 12505 12543 -6245.3    12491   0.000    0  1.0000
## model2.2.2.c  7 12763 12801 -6374.3    12749   0.000    0  1.0000
## model2.1.3.c  8 12116 12160 -6050.2    12100 648.156    1 < 2.2e-16 ***
## model2.2.3.a  9 12137 12186 -6059.4    12119   0.000    1  1.0000
## model2.2.3.b  9 12089 12138 -6035.4    12071  48.133    0 < 2.2e-16 ***
## model2.2.3.c 11 12090 12151 -6034.0    12068   2.698    2   0.2595
## model2.2.4.a 13 12114 12186 -6044.0    12088   0.000    2  1.0000
## model2.2.4.b 13 12068 12140 -6021.2    12042 45.522    0 < 2.2e-16 ***
## model2.2.4.c 16 12073 12162 -6020.7    12041   0.986    3   0.8046
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

model2.2.4.b had the lowest AIC and the second lowest BIC, so it was the best model.  
model2.2.4.b:

$$y_{ij} = \beta_0 + b_{1i} + \beta_1 p_i + (\beta_2 + b_{2i}) t_{ij} + \beta_3 p_i t_{ij} + (\beta_4 + b_{3i}) t_{ij}^2 + \beta_5 p_i t_{ij}^2 + \epsilon_{ij}$$

### 3. Estimates and inferences of the best model

```
modelfinal=lmer(CD4~1+Smoke+Time+Smoke:Time+Time2+Smoke:Time2+(1+Time+Time2|ID), data=BMACS)
summary(modelfinal)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: CD4 ~ 1 + Smoke + Time + Smoke:Time + Time2 + Smoke:Time2 + (1 +
##      Time + Time2 | ID)
##      Data: BMACS
##
## REML criterion at convergence: 12087.9
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -4.3769 -0.5344 -0.0335  0.5142  4.4425
##
## Random effects:
##      Groups   Name                Variance Std.Dev. Corr
##      ID       (Intercept)  87.0574   9.3305
##              Time         27.8274   5.2752  -0.44
##              Time2        0.4491   0.6701   0.32 -0.87
## Residual                22.6498   4.7592
## Number of obs: 1817, groups:  ID, 283
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) 36.40991    0.81397  44.731
## Smoke        1.89572    1.37237   1.381
## Time        -4.79952    0.56716  -8.462
## Time2        0.37760    0.09172   4.117
## Smoke:Time   0.04643    0.98954   0.047
## Smoke:Time2 -0.02708    0.16675  -0.162
##
## Correlation of Fixed Effects:
##              (Intr) Smoke  Time   Time2  Smk:Tm
## Smoke        -0.593
## Time         -0.578  0.343
## Time2         0.452 -0.268 -0.882
## Smoke:Time    0.332 -0.585 -0.573  0.505
## Smoke:Time2  -0.248  0.447  0.485 -0.550 -0.875
```

```
attributes(summary(modelfinal))
```

```
## $names
## [1] "methTitle" "objClass" "devcomp" "isLmer"
## [5] "useScale" "logLik" "family" "link"
## [9] "ngrps" "coefficients" "sigma" "vcov"
## [13] "varcor" "AICtab" "call" "residuals"
## [17] "fitMsgs" "optinfo"
##
## $class
## [1] "summary.merMod"
```

The estimated coefficient of the fixed Effect ( $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ ), standard error, and t value.

```
summary(modelfinal)[['coefficients']]
```

```
##              Estimate Std. Error    t value
## (Intercept) 36.40991214 0.81397244 44.73113620
## Smoke       1.89571635 1.37236966  1.38134528
## Time        -4.79951678 0.56716260 -8.46232939
## Time2        0.37759643 0.09171841  4.11690993
## Smoke:Time   0.04643381 0.98954353  0.04692448
## Smoke:Time2 -0.02707749 0.16674574 -0.16238793
```

The measurement error  $\epsilon_{ij} \sim N(0, 22.649)$ .

```
summary(modelfinal)[['sigma']]
```

```
## [1] 4.75918
```

The standard variance and correlation matrix of the random effects

```
summary(modelfinal)[['varcor']]
```

```
## Groups   Name      Std.Dev. Corr
## ID       (Intercept) 9.33046
##          Time        5.27517 -0.445
##          Time2       0.67013  0.325 -0.869
## Residual                4.75918
```

```
summary(modelfinal)[['vcov']]
```

```
## 6 x 6 Matrix of class "dpoMatrix"
##              (Intercept)      Smoke      Time      Time2  Smoke:Time
## (Intercept)  0.66255114 -0.66255114 -0.26701107  0.033710155  0.26701107
## Smoke       -0.66255114  1.88339847  0.26701107 -0.033710155 -0.79448854
## Time        -0.26701107  0.26701107  0.32167342 -0.045856226 -0.32167342
## Time2        0.03371016 -0.03371016 -0.04585623  0.008412266  0.04585623
## Smoke:Time   0.26701107 -0.79448854 -0.32167342  0.045856226  0.97919639
## Smoke:Time2 -0.03371016  0.10233728  0.04585623 -0.008412266 -0.14445350
##              Smoke:Time2
## (Intercept) -0.033710155
## Smoke       0.102337282
## Time        0.045856226
## Time2       -0.008412266
## Smoke:Time  -0.144453500
## Smoke:Time2  0.027804142
```

## Conclusion

For the best model:

1. Time was negatively correlated with the post-infection CD4 level.
2. Smoking status was positively correlated with the post-infection CD4 level.
3. Time's square, Smoking status and Time's interaction and Smoking status and Time's square's interaction had



some relationship with post-infection CD4 level.

4. The variance of the random effect was relatively large. The deviation of each patient from the population is large.