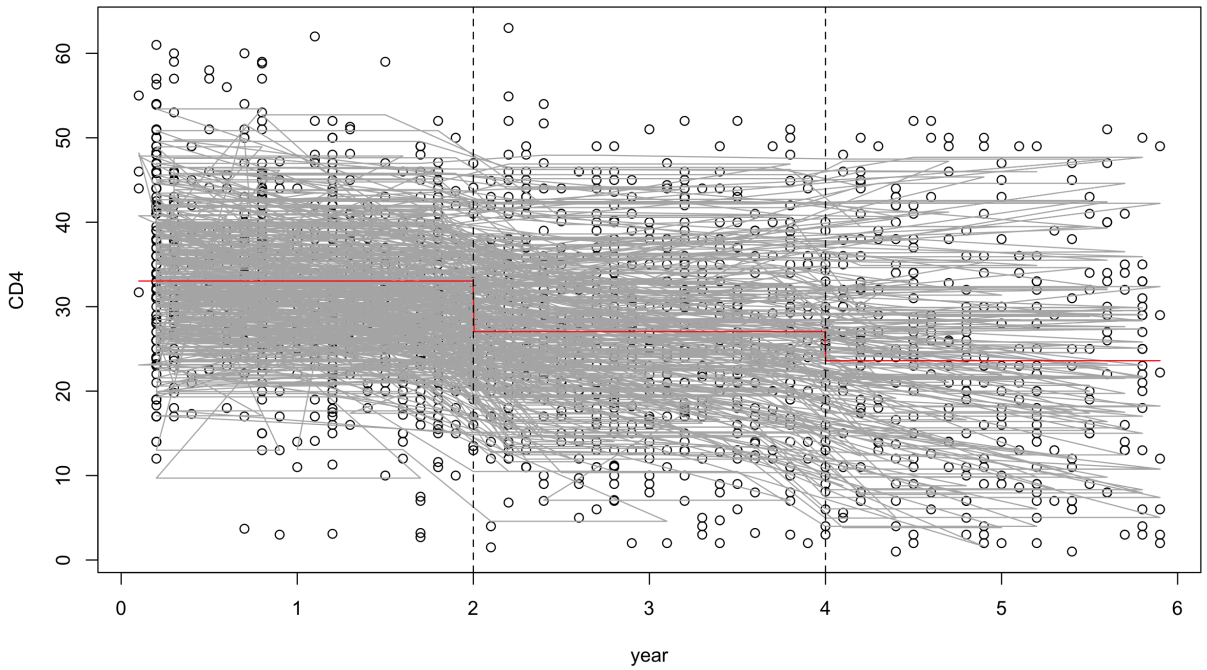
```
> library(lme4)
> library(npmlda)
> library(boot)
> set.seed(123)
> BMACS = read.csv("/Users/Aaron/Desktop/BMACS.csv")
> head(BMACS)
   ID Time Smoke   age preCD4 CD4
1 1022  0.2     0 26.25     38  17
2 1022  0.8     0 26.25     38  30
3 1022  1.2     0 26.25     38  23
4 1022  1.6     0 26.25     38  15
5 1022  2.5     0 26.25     38  21
6 1022  3.0     0 26.25     38  12
```

1.1

```
> basis1 = function(t){
+   b1 = ifelse(t<2,1,0)
+   b2 = ifelse(t>=2&t<4,1,0)
+   b3 = ifelse(t>=4,1,0)
+   return(matrix(c(b1,b2,b3),ncol=3))
+ }
> fit1 = lmer(CD4 ~ 0 + basis1(Time) + (1 + basis1(Time)|ID),data = BMACS)
Warning messages:
1: In checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv,  :
  unable to evaluate scaled gradient
2: In checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv,  :
  Model failed to converge: degenerate  Hessian with 1 negative eigenvalues
> summary(fit1)
Linear mixed model fit by REML ['lmerMod']
Formula: CD4 ~ 0 + basis1(Time) + (1 + basis1(Time) | ID)
   Data: BMACS

REML criterion at convergence: 12415.2

Scaled residuals:
    Min      1Q  Median      3Q     Max
-4.0327 -0.5234 -0.0134  0.5175  3.9661

Random effects:
 Groups   Name           Variance Std.Dev. Corr
 ID       (Intercept)    53.83    7.337
          basis1(Time)1  17.60    4.196    -0.04
          basis1(Time)2  23.20    4.817     0.30  0.26
          basis1(Time)3  61.60    7.848     0.26 -0.05  0.82
 Residual                31.08    5.575
Number of obs: 1817, groups:  ID, 283
```

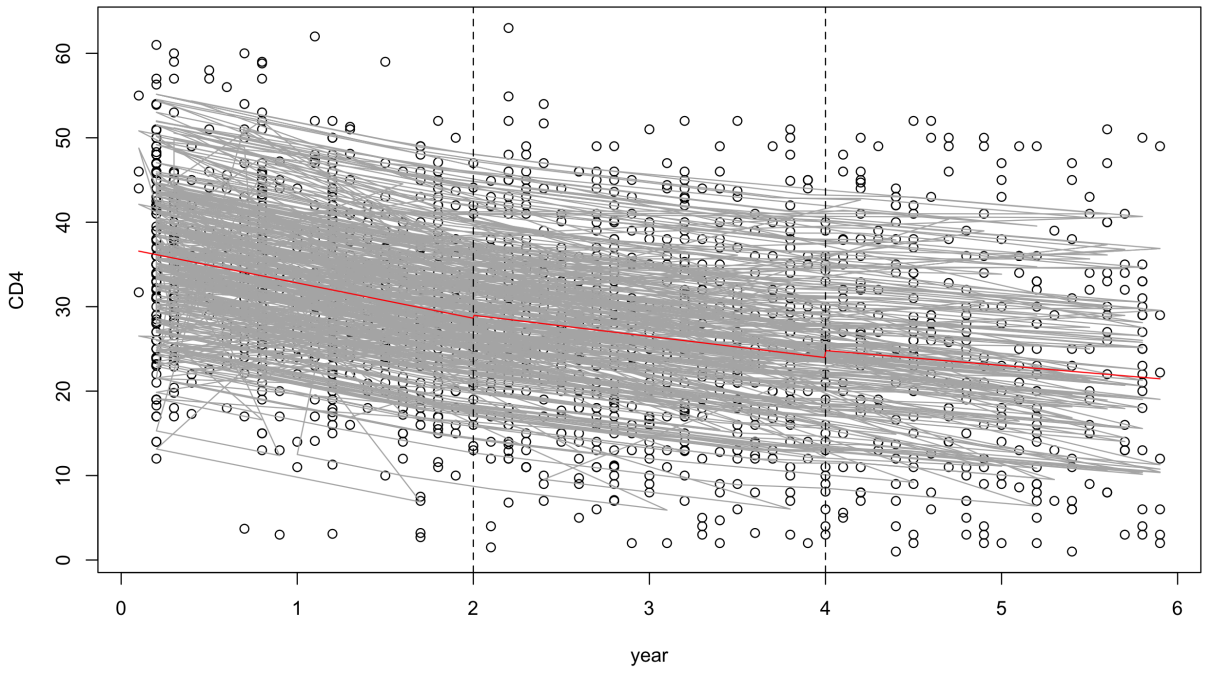
```
Fixed effects:
             Estimate Std. Error t value
basis1(Time)1  33.0369     0.5395   61.24
basis1(Time)2  27.0514     0.6766   39.98
basis1(Time)3  23.5988     0.9353   25.23

Correlation of Fixed Effects:
            b1(T)1 b1(T)2
basis1(Tm)2 0.678
basis1(Tm)3 0.469  0.751
convergence code: 0
unable to evaluate scaled gradient
Model failed to converge: degenerate  Hessian with 1 negative eigenvalues
> n = 1000
> Rtime = range(BMACS$Time)
> Tgrid = seq(from = Rtime[1], to = Rtime[2], length = n)
> grid = basis1(Tgrid)
> mean.hat1 = grid %*% fixef(fit1)
> plot(BMACS$Time, BMACS$CD4, xlab = "year", ylab = "CD4")
> lines(BMACS$Time, predict(fit1), col = "gray70")
> points(Tgrid, mean.hat1, col= "red", lty = 1, type = "l")
> abline(v = c(2,4), lty = 2)
```



1.2

```
> basis2 = function(t){
+   b1 = ifelse(t<2,1,0)
+   b2 = t*ifelse(t<2,1,0)
+   b3 = ifelse(t>=2&t<4,1,0)
+   b4 = t*ifelse(t>=2&t<4,1,0)
+   b5 = ifelse(t>=4,1,0)
+   b6 = t*ifelse(t>=4,1,0)
+   return(matrix(c(b1,b2,b3,b4,b5,b6),ncol=6))
+ }
> fit2 = lmer(CD4 ~ 0 + basis2(Time) + (1|ID),data = BMACS)
> summary(fit2)
Linear mixed model fit by REML ['lmerMod']
Formula: CD4 ~ 0 + basis2(Time) + (1 | ID)
   Data: BMACS

REML criterion at convergence: 12517.6

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.8798 -0.5707 -0.0159  0.5659  4.3474

Random effects:
 Groups   Name        Variance Std.Dev.
 ID       (Intercept) 77.49    8.803
 Residual             39.56    6.290
Number of obs: 1817, groups:  ID, 283

Fixed effects:
             Estimate Std. Error t value
basis2(Time)1  36.9955     0.6888  53.710
basis2(Time)2  -4.1829     0.4114 -10.167
basis2(Time)3  34.0024     1.4478  23.486
basis2(Time)4  -2.5080     0.4634  -5.412
basis2(Time)5  31.8591     3.0271  10.525
basis2(Time)6  -1.7632     0.6167  -2.859

Correlation of Fixed Effects:
            b2(T)1 b2(T)2 b2(T)3 b2(T)4 b2(T)5
basis2(Tm)2 -0.553
basis2(Tm)3  0.273  0.012
basis2(Tm)4 -0.001  0.007 -0.910
basis2(Tm)5  0.129  0.009  0.070 -0.002
basis2(Tm)6  0.002  0.000 -0.007  0.008 -0.976
> n = 1000
> Rtime = range(BMACS$Time)
```

```
> Tgrid = seq(from = Rtime[1],to = Rtime[2], length = n)
> grid = basis2(Tgrid)
> mean.hat2 = grid %*% fixef(fit2)
> plot(BMACS$Time, BMACS$CD4, xlab = "year", ylab = "CD4")
> lines(BMACS$Time, predict(fit2), col = "gray70")
> points(Tgrid, mean.hat2, col= "red", lty = 1, type = "l")
> abline(v = c(2,4), lty = 2)
```

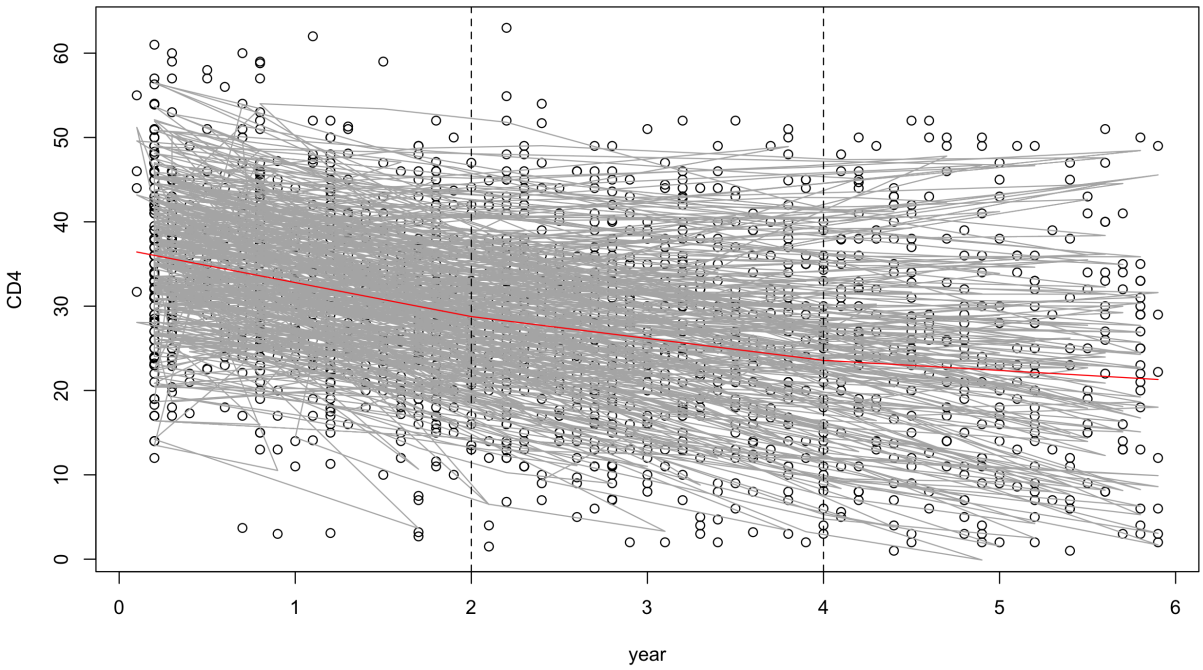


1.3

```
> basis3 = function(t){
+   b1 = ifelse(t>=0,1,0)
+   b2 = t
+   b3 = (t-2)*ifelse(t>=2,1,0)
+   b4 = (t-4)*ifelse(t>=4,1,0)
+   return(matrix(c(b1,b2,b3,b4),ncol=4))
+ }
> fit3 = lmer(CD4 ~ basis3(Time) + (1 + basis3(Time)|ID),data = BMACS)
fixed-effect model matrix is rank deficient so dropping 1 column / coefficient
Warning messages:
1: In checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv,  :
  unable to evaluate scaled gradient
2: In checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv,  :
  Model failed to converge: degenerate  Hessian with 1 negative eigenvalues
> summary(fit3)
```

```
Linear mixed model fit by REML ['lmerMod']
Formula: CD4 ~ basis3(Time) + (1 + basis3(Time) | ID)
   Data: BMACS

REML criterion at convergence: 12075.3

Scaled residuals:
    Min      1Q  Median      3Q     Max
-4.4906 -0.5297 -0.0253  0.4935  4.5624

Random effects:
 Groups   Name          Variance Std.Dev. Corr
 ID       (Intercept)   49.32    7.023
          basis3(Time)1 31.52    5.615     0.08
          basis3(Time)2 21.82    4.671    -0.29 -0.39
          basis3(Time)3 27.05    5.201     0.32  0.09 -0.77
          basis3(Time)4 15.15    3.893    -0.37  0.40  0.12 -0.53
 Residual               21.57    4.645
Number of obs: 1817, groups:  ID, 283

Fixed effects:
             Estimate Std. Error t value
(Intercept)    36.8266     0.6531  56.387
basis3(Time)2  -4.0310     0.3905 -10.322
basis3(Time)3   1.4369     0.5667   2.536
basis3(Time)4   1.3996     0.6738   2.077

Correlation of Fixed Effects:
            (Intr) b3(T)2 b3(T)3
basis3(Tm)2 -0.587
basis3(Tm)3  0.407 -0.782
basis3(Tm)4 -0.086  0.213 -0.556
fit warnings:
fixed-effect model matrix is rank deficient so dropping 1 column / coefficient
convergence code: 0
unable to evaluate scaled gradient
Model failed to converge: degenerate  Hessian with 1 negative eigenvalues

> n = 1000
> Rtime = range(BMACS$Time)
> Tgrid = seq(from = Rtime[1], to = Rtime[2], length = n)
> grid = basis3(Tgrid)
> mean.hat3 = grid %*% fixef(fit3)
> plot(BMACS$Time, BMACS$CD4, xlab = "year", ylab = "CD4")
> lines(BMACS$Time, predict(fit3), col = "gray70")
> points(Tgrid, mean.hat3, col= "red", lty = 1, type = "l")
```

> abline(v = c(2,4), lty = 2)



1.4

```
> anova(fit1,fit2)
refitting model(s) with ML (instead of REML)
Data: BMACS
Models:
fit2: CD4 ~ 0 + basis2(Time) + (1 | ID)
fit1: CD4 ~ 0 + basis1(Time) + (1 + basis1(Time) | ID)
     Df   AIC   BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
fit2  8 12535 12579 -6259.6    12519
fit1 14 12445 12522 -6208.5    12417 102.23      6  < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> anova(fit1,fit3)
refitting model(s) with ML (instead of REML)
Data: BMACS
Models:
fit1: CD4 ~ 0 + basis1(Time) + (1 + basis1(Time) | ID)
fit3: CD4 ~ basis3(Time) + (1 + basis3(Time) | ID)
     Df   AIC   BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
fit1 14 12445 12522 -6208.5    12417
fit3 20 12116 12226 -6038.0    12076   341      6  < 2.2e-16 ***
---
```

```
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> merBoot = bootMer(x = fit3, FUN = fixef, nsim = 100)
> CI.lower = apply(merBoot$t, 2, function(x) as.numeric(quantile(x, probs =.025,
na.rm=TRUE)))
> CI.upper = apply(merBoot$t, 2, function(x) as.numeric(quantile(x, probs =.975,
na.rm=TRUE)))
> (cbind(CI.lower,CI.upper))
                 CI.lower  CI.upper
(Intercept)   35.7114147 38.241289
basis3(Time)2 -4.8733722 -3.282402
basis3(Time)3  0.5172552  2.764431
basis3(Time)4 -0.2218978  2.382884
```

The result shows that the third model is best and we use bootstrap to find the confidence interval for each estimator in this model. The result shows that the influence of time is significant at first and then it decreases to a level.

2.1

```
> basis4 = function(t,s){
+   b1 = ifelse(t<2,1,0)*s
+   b2 = ifelse(t>=2&t<4,1,0)*s
+   b3 = ifelse(t>=4,1,0)*s
+   return(matrix(c(b1,b2,b3),ncol=3))
+ }
>
> fit4 = lmer(CD4 ~ 0 + basis1(Time) + basis4(Time,Smoke) + (0 + basis1(Time)|ID),data
= BMACS)
Warning message:
In checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv,  :
  Model failed to converge with max|grad| = 0.00343388 (tol = 0.002, component 1)
>
> summary(fit4)
Linear mixed model fit by REML ['lmerMod']
Formula: CD4 ~ 0 + basis1(Time) + basis4(Time, Smoke) + (0 + basis1(Time) |      ID)
   Data: BMACS

REML criterion at convergence: 12402.3

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.9959 -0.5247 -0.0138  0.5129  3.9590

Random effects:
 Groups   Name           Variance Std.Dev. Corr
 ID       basis1(Time)1  68.78    8.293
          basis1(Time)2  96.42    9.819   0.83
```

```
         basis1(Time)3 145.81   12.075   0.66 0.93
 Residual               31.09    5.576
Number of obs: 1817, groups:  ID, 283

Fixed effects:
                  Estimate Std. Error t value
basis1(Time)1      32.4066     0.6700  48.370
basis1(Time)2      25.9628     0.8298  31.288
basis1(Time)3      22.9001     1.1316  20.237
basis4(Time, Smoke)1  1.7244     1.1252   1.533
basis4(Time, Smoke)2  3.1522     1.4128   2.231
basis4(Time, Smoke)3  1.6618     2.0278   0.820

Correlation of Fixed Effects:
           b1(T)1 b1(T)2 b1(T)3 b4(T,S)1 b4(T,S)2
basis1(Tm)2  0.690
basis1(Tm)3  0.490  0.769
bss4(Tm,S)1 -0.595 -0.411 -0.292
bss4(Tm,S)2 -0.405 -0.587 -0.452  0.660
bss4(Tm,S)3 -0.273 -0.429 -0.558  0.445    0.730
convergence code: 0
Model failed to converge with max|grad| = 0.00343388 (tol = 0.002, component 1)

2.2
> basis5 = function(t,s){
+   b1 = ifelse(t<2,1,0)*s
+   b2 = t*ifelse(t<2,1,0)*s
+   b3 = ifelse(t>=2&t<4,1,0)*s
+   b4 = t*ifelse(t>=2&t<4,1,0)*s
+   b5 = ifelse(t>=4,1,0)*s
+   b6 = t*ifelse(t>=4,1,0)*s
+   return(matrix(c(b1,b2,b3,b4,b5,b6),ncol=6))
+ }
>
> fit5 = lmer(CD4 ~ 0 + basis2(Time) + basis5(Time,Smoke) + (1|ID),data = BMACS)
> summary(fit5)
Linear mixed model fit by REML ['lmerMod']
Formula: CD4 ~ 0 + basis2(Time) + basis5(Time, Smoke) + (1 | ID)
   Data: BMACS

REML criterion at convergence: 12494.6

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.8154 -0.5747 -0.0187  0.5501  4.3894
```

Random effects:

Groups	Name	Variance	Std.Dev.
ID	(Intercept)	76.79	8.763
Residual		39.47	6.283

Number of obs: 1817, groups: ID, 283

Fixed effects:

	Estimate	Std. Error	t value
basis2(Time)1	35.9639	0.8553	42.046
basis2(Time)2	-3.6777	0.4993	-7.365
basis2(Time)3	33.6282	1.7517	19.197
basis2(Time)4	-2.7295	0.5558	-4.911
basis2(Time)5	30.5919	3.5611	8.590
basis2(Time)6	-1.6118	0.7244	-2.225
basis5(Time, Smoke)1	2.9021	1.4353	2.022
basis5(Time, Smoke)2	-1.5746	0.8799	-1.789
basis5(Time, Smoke)3	0.6631	3.1070	0.213
basis5(Time, Smoke)4	0.8131	1.0043	0.810
basis5(Time, Smoke)5	4.1180	6.7449	0.611
basis5(Time, Smoke)6	-0.5629	1.3761	-0.409

Correlation of Fixed Effects:

	b2(T)1	b2(T)2	b2(T)3	b2(T)4	b2(T)5	b2(T)6	b5(T,S)1	b5(T,S)2	b5(T,S)3	b5(T,S)4	b5(T,S)5
basis2(Tm)2	-0.554										
basis2(Tm)3	0.286	0.010									
basis2(Tm)4	-0.003	0.007	-0.906								
basis2(Tm)5	0.137	0.008	0.075	-0.003							
basis2(Tm)6	0.002	0.001	-0.006	0.008	-0.974						
bss5(Tm,S)1	-0.596	0.330	-0.170	0.002	-0.082	-0.001					
bss5(Tm,S)2	0.314	-0.567	-0.006	-0.004	-0.004	-0.001	-0.555				
bss5(Tm,S)3	-0.161	-0.006	-0.564	0.511	-0.043	0.004	0.259	0.014			
bss5(Tm,S)4	0.002	-0.004	0.501	-0.553	0.002	-0.005	0.000	0.007	-0.914		
bss5(Tm,S)5	-0.072	-0.004	-0.040	0.002	-0.528	0.514	0.119	0.008	0.064	-0.002	
bss5(Tm,S)6	-0.001	-0.001	0.003	-0.004	0.513	-0.526	0.002	0.000	-0.007	0.007	-0.977

2.3

```
> basis6 = function(t,s){
+   b1 = ifelse(t>=0,1,0)*s
+   b2 = t*s
+   b3 = (t-2)*ifelse(t>=2,1,0)*s
+   b4 = (t-4)*ifelse(t>=4,1,0)*s
+   return(matrix(c(b1,b2,b3,b4),ncol=4))
+ }
> fit6 = lmer(CD4 ~ 0 + basis3(Time) + basis6(Time,Smoke) + (0 + basis3(Time)|ID),data
= BMACS)
Warning message:
In checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv,  :
  Model failed to converge with max|grad| = 0.00531314 (tol = 0.002, component 1)
>
> summary(fit6)
Linear mixed model fit by REML ['lmerMod']
Formula: CD4 ~ 0 + basis3(Time) + basis6(Time, Smoke) + (0 + basis3(Time) |      ID)
   Data: BMACS

REML criterion at convergence: 12063.5

Scaled residuals:
    Min      1Q  Median      3Q     Max
-4.4651 -0.5208 -0.0256  0.4926  4.5480

Random effects:
 Groups   Name          Variance Std.Dev. Corr
 ID       basis3(Time)1 87.13    9.334
          basis3(Time)2 22.05    4.696    -0.46
          basis3(Time)3 27.80    5.272     0.29 -0.77
          basis3(Time)4 15.71    3.963    -0.02  0.16 -0.56
 Residual               21.56    4.643
Number of obs: 1817, groups:  ID, 283

Fixed effects:
                    Estimate Std. Error t value
basis3(Time)1        36.0511     0.8120  44.395
basis3(Time)2        -3.9302     0.4822  -8.150
basis3(Time)3         1.1842     0.6896   1.717
basis3(Time)4         1.9510     0.7987   2.443
basis6(Time, Smoke)1  2.1685     1.3635   1.590
basis6(Time, Smoke)2 -0.2945     0.8270  -0.356
basis6(Time, Smoke)3  0.8097     1.2272   0.660
basis6(Time, Smoke)4 -2.0190     1.4980  -1.348

Correlation of Fixed Effects:
```

```
           b3(T)1 b3(T)2 b3(T)3 b3(T)4 b6(T,S)1 b6(T,S)2 b6(T,S)3
basis3(Tm)2 -0.587
basis3(Tm)3  0.412 -0.790
basis3(Tm)4 -0.084  0.227 -0.567
bss6(Tm,S)1 -0.596  0.349 -0.245  0.050
bss6(Tm,S)2  0.342 -0.583  0.461 -0.132 -0.590
bss6(Tm,S)3 -0.231  0.444 -0.562  0.318  0.402   -0.776
bss6(Tm,S)4  0.045 -0.121  0.302 -0.533 -0.079    0.220   -0.560
convergence code: 0
Model failed to converge with max|grad| = 0.00531314 (tol = 0.002, component 1)
```

2.4

```
> anova(fit4,fit5)
refitting model(s) with ML (instead of REML)
Data: BMACS
Models:
fit4: CD4 ~ 0 + basis1(Time) + basis4(Time, Smoke) + (0 + basis1(Time) |
fit4:     ID)
fit5: CD4 ~ 0 + basis2(Time) + basis5(Time, Smoke) + (1 | ID)
    Df   AIC   BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
fit4 13 12437 12508 -6205.4    12411
fit5 14 12535 12612 -6253.6    12507     0      1          1
> anova(fit4,fit6)
refitting model(s) with ML (instead of REML)
Data: BMACS
Models:
fit4: CD4 ~ 0 + basis1(Time) + basis4(Time, Smoke) + (0 + basis1(Time) |
fit4:     ID)
fit6: CD4 ~ 0 + basis3(Time) + basis6(Time, Smoke) + (0 + basis3(Time) |
fit6:     ID)
    Df   AIC   BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
fit4 13 12437 12508 -6205.4    12411
fit6 19 12109 12214 -6035.5    12071 339.65      6  < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> merBoot1 = bootMer(x = fit6, FUN = fixef, nsim = 100)
> CI.lower = apply(merBoot1$t, 2, function(x) as.numeric(quantile(x, probs=.025,
na.rm=TRUE)))
> CI.upper = apply(merBoot1$t, 2, function(x) as.numeric(quantile(x, probs=.975,
na.rm=TRUE)))
> (cbind(CI.lower,CI.upper))
                  CI.lower   CI.upper
basis3(Time)1       34.4492756 37.7571864
basis3(Time)2       -4.7985149 -3.0889339
basis3(Time)3        0.1145417  2.3330854
basis3(Time)4        0.6453930  3.4112040
```

basis6(Time, Smoke)1 -0.4295782 5.0035174
basis6(Time, Smoke)2 -2.0669538 1.4157211
basis6(Time, Smoke)3 -1.4992028 2.7555428
basis6(Time, Smoke)4 -4.3915045 0.4275039

The result shows that the third model is best and we use bootstrap to find the confidence interval for each estimator in this model. The result shows that the influence fluctuates in a relative small level. But we find a large increase in the confidence interval compared to question 1 and this may be because smoking status plays a role in HIV infection.