

proj3_survival

2024-03-07

- Goal is to check the analysis of the nejm paper
- And provide additional exploratory analyses that might “break” the results
- Are there other covariates that might lead to a different conclusion – PC for dim red on immunology variables
- Are the Kaplan-Meier curves actually different or not – Cox pH before 32 weeks and after 34 weeks
- How to deal with the multiple comparison issue – Numerous additional analyses on the same dataset
 - Look into different methods of p-value adjustment

Variables of Interests: Group Age: this is age at baseline Comorbidity indicator: diabetes + hypertension
Drug indicator: drugs addiction + alcohol use Previous pregnancy Education

Outcomes: GA.at.outcome (this is in days): this is gestational age at delivery; use this as event variable
Preg.ended...37.wk: This is the pre-term birth indicator; use this as censoring indicator - Survival as primary, logistic as secondary

Secondary Outcome: Birthweight → need to binarize to less than 2500 g - Logistic Regression

Check distributions: Birth.outcome

Subsetting the Dataframe

```
##  
## No Yes  
## 711 103
```

```
##  
## No Yes  
## 689 98
```

PCA

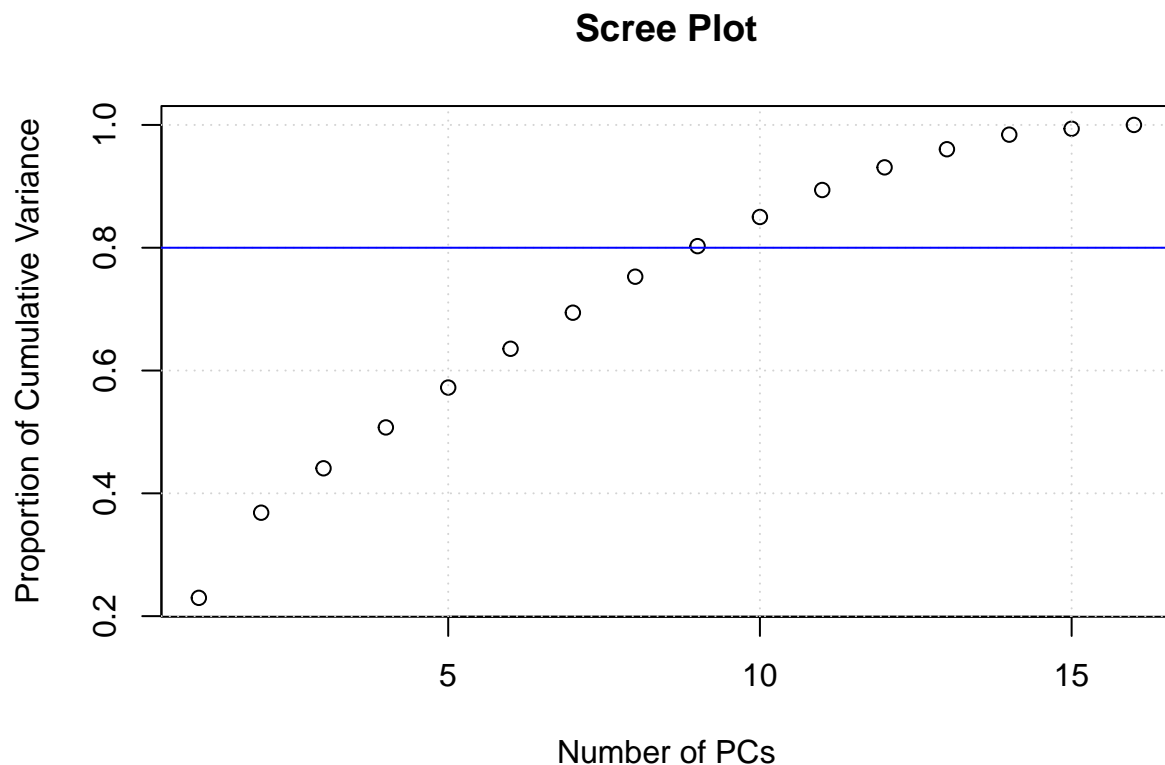
```
pca_complete_preterm <- prcomp(complete_preterm[,serum], scale = TRUE)  
summary(pca_complete_preterm)
```

```
## Importance of components:  
##          PC1      PC2      PC3      PC4      PC5      PC6      PC7  
## Standard deviation    1.9175 1.4895 1.07564 1.0323 1.01924 1.00503 0.96989  
## Proportion of Variance 0.2298 0.1387 0.07231 0.0666 0.06493 0.06313 0.05879  
## Cumulative Proportion 0.2298 0.3685 0.44078 0.5074 0.57231 0.63544 0.69423  
##          PC8      PC9      PC10     PC11     PC12     PC13     PC14  
## Standard deviation    0.96851 0.89214 0.87166 0.83851 0.76822 0.68729 0.61648  
## Proportion of Variance 0.05863 0.04974 0.04749 0.04394 0.03689 0.02952 0.02375
```

```
## Cumulative Proportion  0.75286 0.80260 0.85009 0.89403 0.93092 0.96044 0.98419
##                        PC15    PC16
## Standard deviation    0.39277 0.31410
## Proportion of Variance 0.00964 0.00617
## Cumulative Proportion  0.99383 1.00000
```

```
cumvars <- cumsum(pca_complete_preterm$sdev^2 / sum(pca_complete_preterm$sdev^2))

plot(cumvars, xlab = 'Number of PCs', ylab = 'Proportion of Cumulative Variance', main = 'Scree Plot')
grid()
abline(h = 0.8, col = 'blue')
```



```
rotated_dat <- as.data.frame(pca_complete_preterm$x)
# 9 PCs instead of 16

complete_preterm_extended <- cbind(complete_preterm[, -c(12:27)], rotated_dat[, 1:9])
complete_preterm_extended <- complete_preterm_extended[complete.cases(complete_preterm_extended), ]

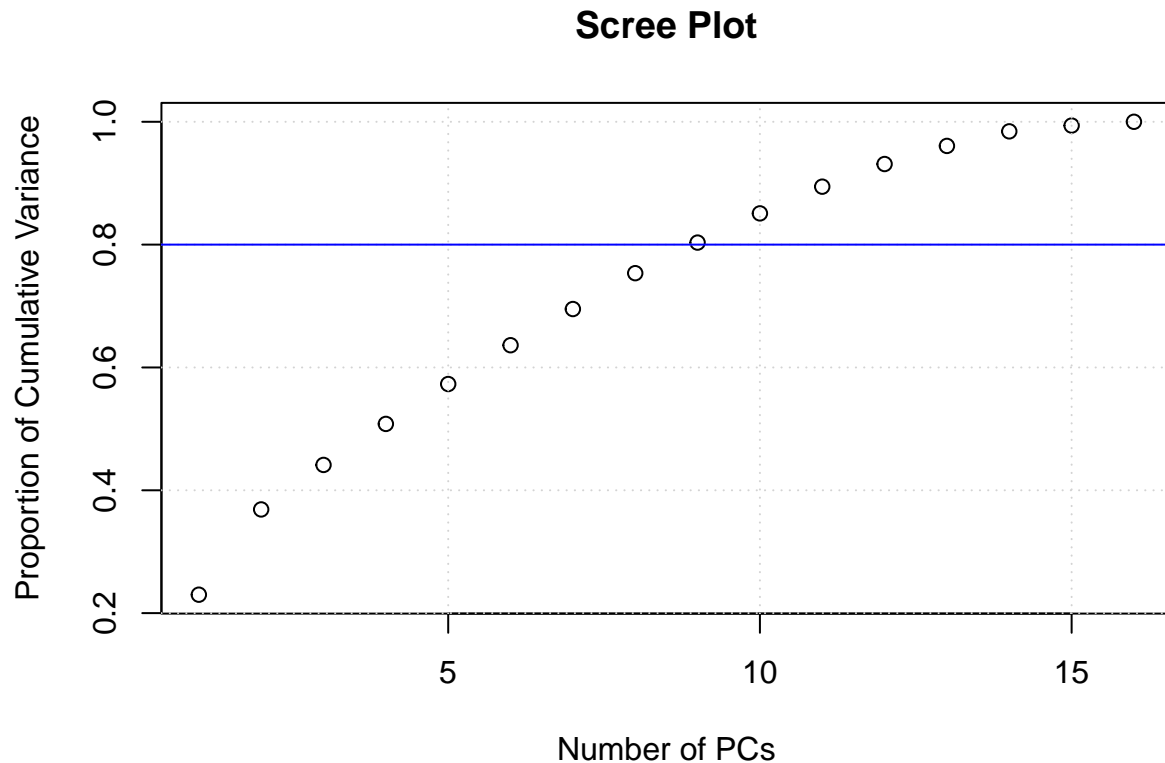
## For low birth weight
pca_complete_birthweight <- prcomp(complete_birthweight[, serum], scale = TRUE)
summary(pca_complete_birthweight)
```

```
## Importance of components:
##                PC1    PC2    PC3    PC4    PC5    PC6    PC7
## Standard deviation  1.918 1.4901 1.07696 1.0338 1.01945 1.00530 0.97107
```

```
## Proportion of Variance 0.230 0.1388 0.07249 0.0668 0.06496 0.06316 0.05894
## Cumulative Proportion 0.230 0.3688 0.44128 0.5081 0.57303 0.63620 0.69513
##          PC8      PC9      PC10      PC11      PC12      PC13      PC14
## Standard deviation  0.96640 0.8926 0.87219 0.83464 0.76770 0.68677 0.61624
## Proportion of Variance 0.05837 0.0498 0.04754 0.04354 0.03683 0.02948 0.02373
## Cumulative Proportion 0.75350 0.8033 0.85085 0.89439 0.93122 0.96070 0.98443
##          PC15      PC16
## Standard deviation  0.38865 0.31309
## Proportion of Variance 0.00944 0.00613
## Cumulative Proportion 0.99387 1.00000
```

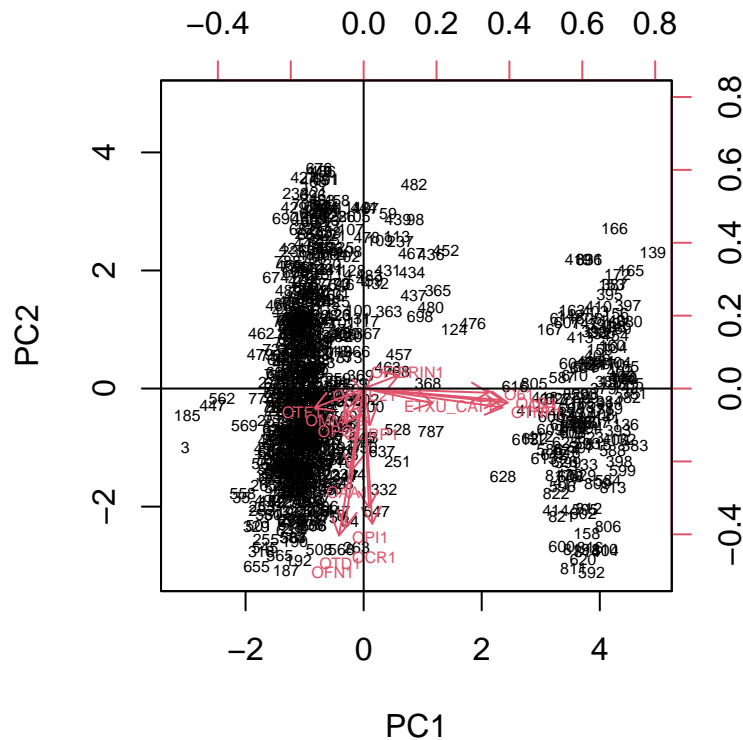
```
cumvars2 <- cumsum(pca_complete_birthweight$sdev^2 / sum(pca_complete_birthweight$sdev^2))

plot(cumvars2, xlab = 'Number of PCs', ylab = 'Proportion of Cumulative Variance', main = 'Scree Plot')
grid()
abline(h = 0.8, col = 'blue')
```



```
# 9 PCs instead of 16
rotated_dat <- as.data.frame(pca_complete_birthweight$x)

biplot(pca_complete_birthweight,scale=0,cex=0.5)
abline(h=0,v=0)
```



```
complete_birthweight_extended <- cbind(complete_birthweight[,-c(12:27)], rotated_dat[,1:9])
complete_birthweight_extended <- complete_birthweight_extended[complete.cases(complete_birthweight_exter
```

Exploratory Plots for initial data

```
library(survival)
library(survminer)
```

```
##      ggplot2
```

```
##      ggpubr
```

```
##
##      'survminer'
```

```
## The following object is masked from 'package:survival':
```

```
##
```

```
##      myeloma
```

```
complete_preterm$Preg.ended...37.wk <- ifelse(complete_preterm$Preg.ended...37.wk == 'Yes', 1, 0)

surv_obj <- with(complete_preterm, Surv(time = GA.at.outcome, event = Preg.ended...37.wk, type = 'right'))

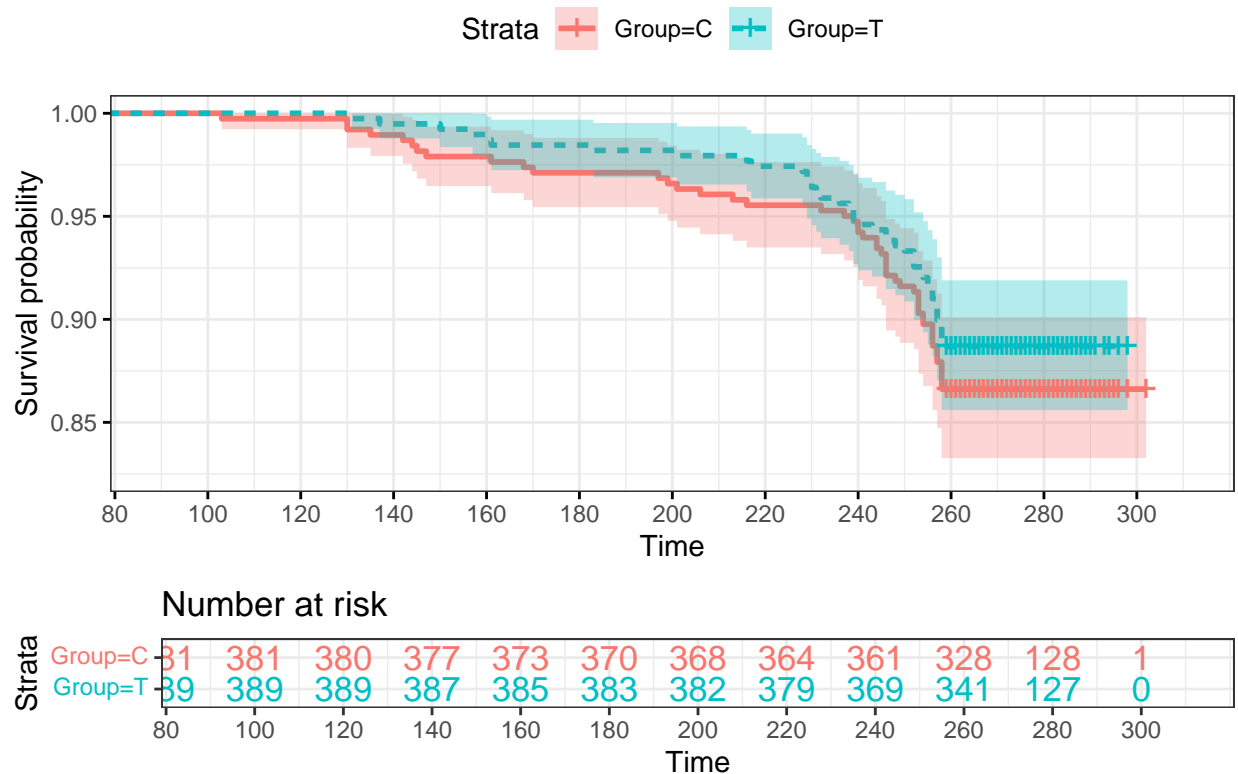
fit <- survfit(surv_obj ~ Group, data = complete_preterm)

ggs <- ggsvplot(fit,
  pval = TRUE, conf.int = TRUE,
  risk.table = TRUE, # Add risk table
  risk.table.col = "strata", # Change risk table color by groups
  linetype = "strata", # Change line type by groups
  ggtheme = theme_bw(), # Change ggplot2 theme
  palette = c("#F8766D", "#00BFC4"),
  ylim=c(0.825,1),
  xlim=c(90,310),
  title = 'Survival Curves for Control and Treatment Cohort',
  break.x.by = 20)

#ggs$plot <- ggs$plot + ylim(c(0.75,1)) + xlim(c(90, 310))
ggs
```

```
## Warning: Removed 1 row containing missing values or values outside the scale range
## (`geom_text()`).
## Removed 1 row containing missing values or values outside the scale range
## (`geom_text()`).
```

Survival Curves for Control and Treatment Cohort



```
surv_diff1 <- survdiff(Surv(time = GA.at.outcome, event = Preg.ended...37.wk, type = 'right') ~ Group, data = complete_preterm)
surv_diff1
```

```
## Call:
## survdiff(formula = Surv(time = GA.at.outcome, event = Preg.ended...37.wk,
##      type = "right") ~ Group, data = complete_preterm)
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## Group=C 381      51      46.7      0.405      0.799
## Group=T 389      44      48.3      0.391      0.799
##
## Chisq= 0.8  on 1 degrees of freedom, p= 0.4
```

Only checking the data pre-32.8 weeks aka 231 days or fewer

```
pre32 <- complete_preterm

pre32$Preg.ended...37.wk <- NULL
pre32$pre230 <- ifelse(pre32$GA.at.outcome <= 230, 1, 0)

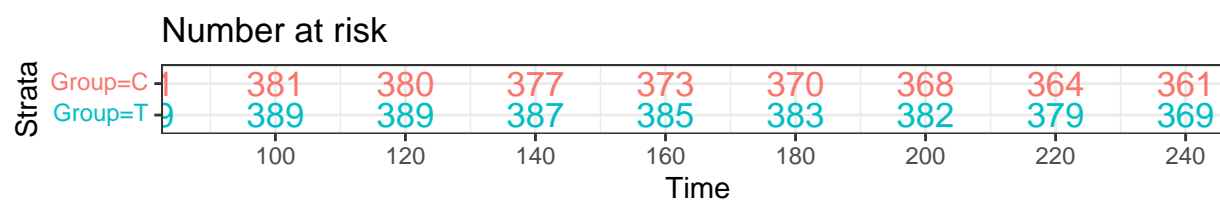
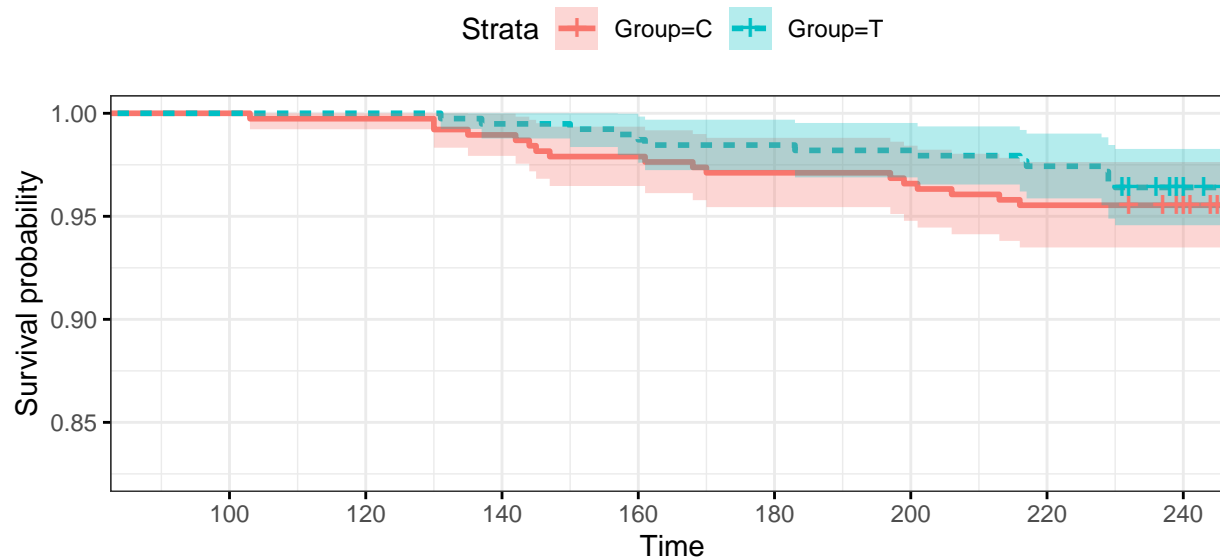
# new survival fit
surv_obj2 <- with(pre32, Surv(time = GA.at.outcome, event = pre230, type = 'right'))
fit2 <- survfit(surv_obj2 ~ Group, data = pre32)

ggs2 <- ggsurvplot(fit2,
  pval = TRUE, conf.int = TRUE,
  risk.table = TRUE, # Add risk table
  risk.table.col = "strata", # Change risk table color by groups
  linetype = "strata", # Change line type by groups
  ggtheme = theme_bw(), # Change ggplot2 theme
  palette = c("#F8766D", "#00BFC4"),
  ylim=c(0.825,1),
  xlim=c(90,240),
  title = 'Survival Curves for Control and Treatment Cohort',
  break.x.by = 20)

#ggs$plot <- ggs$plot + ylim(c(0.75,1)) + xlim(c(90, 310))
ggs2
```

```
## Warning: Removed 1 row containing missing values or values outside the scale range
## (`geom_text()`).
## Removed 1 row containing missing values or values outside the scale range
## (`geom_text()`).
```

Survival Curves for Control and Treatment Cohort



```
surv_diff2 <- survdiff(Surv(time = GA.at.outcome, event = pre230, type = 'right') ~ Group, data = pre32)
surv_diff2
```

```
## Call:
## survdiff(formula = Surv(time = GA.at.outcome, event = pre230,
##   type = "right") ~ Group, data = pre32)
##
##           N Observed Expected (O-E)^2/E (O-E)^2/V
## Group=C 381      17     15.2    0.202    0.398
## Group=T 389      14     15.8    0.196    0.398
##
## Chisq= 0.4  on 1 degrees of freedom, p= 0.5
```

Cox Proportional Hazards for preterm birth

```
fit.coxph <- coxph(Surv(time = GA.at.outcome, event = Preg.ended...37.wk) ~ ., data=complete_preterm)
summary(fit.coxph)
```

```
## Call:
## coxph(formula = Surv(time = GA.at.outcome, event = Preg.ended...37.wk) ~
##   ., data = complete_preterm)
##
## n= 770, number of events= 95
```

```

##
##          coef exp(coef) se(coef)      z Pr(>|z|)
## Age      3.029e-02 1.031e+00 2.008e-02 1.508 0.13151
## EducationLT 8 yrs -2.246e-01 7.988e-01 3.183e-01 -0.706 0.48042
## EducationMT 12 yrs 1.924e-01 1.212e+00 2.436e-01 0.790 0.42952
## HypertensionYes 1.215e+00 3.372e+00 3.737e-01 3.252 0.00115 **
## DiabetesYes 9.085e-01 2.481e+00 4.082e-01 2.225 0.02605 *
## Use.TobYes 3.281e-01 1.388e+00 3.138e-01 1.045 0.29589
## Use.AlcYes -4.693e-01 6.254e-01 8.081e-01 -0.581 0.56142
## Drug.AddYes 1.385e+00 3.995e+00 1.140e+00 1.216 0.22416
## Prev.pregYes 2.520e-01 1.287e+00 2.834e-01 0.889 0.37388
## GroupT -3.304e-01 7.186e-01 2.138e-01 -1.545 0.12230
## OAA1 -6.403e-04 9.994e-01 5.108e-04 -1.254 0.21001
## OCR1 1.312e-04 1.000e+00 5.735e-04 0.229 0.81902
## OFN1 -2.071e-04 9.998e-01 6.221e-04 -0.333 0.73920
## OPG1 3.436e-04 1.000e+00 4.708e-04 0.730 0.46558
## OPI1 -6.037e-04 9.994e-01 5.418e-04 -1.114 0.26520
## OTD1 1.767e-04 1.000e+00 5.469e-04 0.323 0.74657
## OTF1 -4.541e-04 9.995e-01 5.019e-04 -0.905 0.36561
## OCRP1 7.158e-05 1.000e+00 5.060e-04 0.141 0.88751
## O1B1 1.251e-02 1.013e+00 1.430e-02 0.875 0.38162
## O61 -1.101e-02 9.890e-01 2.156e-02 -0.511 0.60955
## O81 -1.865e-02 9.815e-01 3.767e-02 -0.495 0.62054
## OPGE21 -6.169e-04 9.994e-01 4.820e-04 -1.280 0.20053
## OTNF1 -1.075e-02 9.893e-01 1.220e-02 -0.881 0.37822
## OMMP91 -9.989e-04 9.990e-01 7.476e-04 -1.336 0.18149
## ETXU_CAT1 -3.415e-02 9.664e-01 1.112e-01 -0.307 0.75882
## OFIBRIN1 3.762e-05 1.000e+00 6.212e-04 0.061 0.95171
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##          exp(coef) exp(-coef) lower .95 upper .95
## Age      1.0308      0.9702      0.9910      1.072
## EducationLT 8 yrs 0.7988      1.2519      0.4280      1.491
## EducationMT 12 yrs 1.2122      0.8250      0.7520      1.954
## HypertensionYes 3.3717      0.2966      1.6208      7.014
## DiabetesYes 2.4805      0.4031      1.1145      5.521
## Use.TobYes 1.3883      0.7203      0.7505      2.568
## Use.AlcYes 0.6254      1.5989      0.1283      3.048
## Drug.AddYes 3.9955      0.2503      0.4281      37.286
## Prev.pregYes 1.2866      0.7773      0.7383      2.242
## GroupT 0.7186      1.3916      0.4726      1.093
## OAA1 0.9994      1.0006      0.9984      1.000
## OCR1 1.0001      0.9999      0.9990      1.001
## OFN1 0.9998      1.0002      0.9986      1.001
## OPG1 1.0003      0.9997      0.9994      1.001
## OPI1 0.9994      1.0006      0.9983      1.000
## OTD1 1.0002      0.9998      0.9991      1.001
## OTF1 0.9995      1.0005      0.9986      1.001
## OCRP1 1.0001      0.9999      0.9991      1.001
## O1B1 1.0126      0.9876      0.9846      1.041
## O61 0.9890      1.0111      0.9481      1.032
## O81 0.9815      1.0188      0.9117      1.057
## OPGE21 0.9994      1.0006      0.9984      1.000

```



```
## OTNF1          0.9893      1.0108      0.9659      1.013
## OMMP91         0.9990      1.0010      0.9975      1.000
## ETXU_CAT1      0.9664      1.0347      0.7772      1.202
## OFIBRIN1       1.0000      1.0000      0.9988      1.001
```

```
##
## Concordance= 0.677 (se = 0.028 )
## Likelihood ratio test= 49.54 on 26 df, p=0.004
## Wald test          = 61.58 on 26 df, p=1e-04
## Score (logrank) test = 71.19 on 26 df, p=4e-06
```

```
fit.coxph2 <- coxph(Surv(time = GA.at.outcome, event = pre230)~., data=pre32)
# ggsurvplot(survfit(fit.coxph2), data = pre32)
summary(fit.coxph2)
```

```
## Call:
## coxph(formula = Surv(time = GA.at.outcome, event = pre230) ~
##      ., data = pre32)
##
##      n= 770, number of events= 31
##
##              coef exp(coef)    se(coef)      z Pr(>|z|)
## Age          -3.491e-02  9.657e-01  3.717e-02 -0.939  0.3477
## EducationLT 8 yrs -5.625e-01  5.698e-01  6.418e-01 -0.876  0.3808
## EducationMT 12 yrs  6.207e-01  1.860e+00  4.271e-01  1.453  0.1461
## HypertensionYes  9.445e-01  2.572e+00  7.275e-01  1.298  0.1942
## DiabetesYes      1.340e+00  3.821e+00  6.812e-01  1.968  0.0491 *
## Use.TobYes       -2.978e-01  7.425e-01  6.968e-01 -0.427  0.6692
## Use.AlcYes       -8.719e-01  4.182e-01  1.639e+00 -0.532  0.5947
## Drug.AddYes      2.580e+00  1.320e+01  1.697e+00  1.521  0.1283
## Prev.pregYes     3.567e-01  1.429e+00  4.692e-01  0.760  0.4472
## GroupT          -4.483e-01  6.387e-01  3.857e-01 -1.162  0.2451
## OAA1             8.192e-05  1.000e+00  8.854e-04  0.093  0.9263
## OCR1             4.255e-04  1.000e+00  1.040e-03  0.409  0.6825
## OFN1             2.132e-03  1.002e+00  1.140e-03  1.870  0.0614 .
## OPG1            -7.681e-04  9.992e-01  8.226e-04 -0.934  0.3504
## OPI1            -3.487e-04  9.997e-01  9.700e-04 -0.359  0.7193
## OTD1            -4.924e-04  9.995e-01  9.664e-04 -0.510  0.6104
## OTF1             3.786e-04  1.000e+00  8.841e-04  0.428  0.6685
## OCRP1           -5.337e-04  9.995e-01  8.989e-04 -0.594  0.5527
## O1B1            -6.910e-02  9.332e-01  4.273e-02 -1.617  0.1059
## O61             4.353e-02  1.044e+00  4.077e-02  1.068  0.2857
## O81            -1.691e-03  9.983e-01  7.527e-02 -0.022  0.9821
## OPGE21          4.993e-05  1.000e+00  8.551e-04  0.058  0.9534
## OTNF1           9.931e-03  1.010e+00  1.607e-02  0.618  0.5365
## OMMP91          -5.851e-04  9.994e-01  1.287e-03 -0.454  0.6495
## ETXU_CAT1       1.568e-01  1.170e+00  1.784e-01  0.879  0.3795
## OFIBRIN1        5.916e-04  1.001e+00  1.031e-03  0.573  0.5663
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## Age              0.9657    1.03552    0.89784    1.039
## EducationLT 8 yrs  0.5698    1.75505    0.16197    2.004
## EducationMT 12 yrs  1.8603    0.53755    0.80541    4.297
```

```
## HypertensionYes      2.5716    0.38886    0.61795    10.702
## DiabetesYes          3.8208    0.26172    1.00533    14.521
## Use.TobYes           0.7425    1.34683    0.18947     2.910
## Use.AlcYes           0.4182    2.39138    0.01685    10.378
## Drug.AddYes         13.1976    0.07577    0.47464   366.962
## Prev.pregYes         1.4286    0.69998    0.56950     3.584
## GroupT               0.6387    1.56565    0.29990     1.360
## OAA1                 1.0001    0.99992    0.99835     1.002
## OCR1                 1.0004    0.99957    0.99839     1.002
## OFN1                 1.0021    0.99787    0.99990     1.004
## OPG1                 0.9992    1.00077    0.99762     1.001
## OPI1                 0.9997    1.00035    0.99775     1.002
## OTD1                 0.9995    1.00049    0.99762     1.001
## OTF1                 1.0004    0.99962    0.99865     1.002
## OCRP1                0.9995    1.00053    0.99771     1.001
## O1B1                 0.9332    1.07154    0.85826     1.015
## O61                  1.0445    0.95740    0.96427     1.131
## O81                  0.9983    1.00169    0.86139     1.157
## OPGE21               1.0000    0.99995    0.99838     1.002
## OTNF1               1.0100    0.99012    0.97867     1.042
## OMMP91              0.9994    1.00059    0.99690     1.002
## ETXU_CAT1            1.1697    0.85490    0.82457     1.659
## OFIBRIN1             1.0006    0.99941    0.99857     1.003
##
## Concordance= 0.749 (se = 0.047 )
## Likelihood ratio test= 28.85 on 26 df, p=0.3
## Wald test              = 29.17 on 26 df, p=0.3
## Score (logrank) test = 33.35 on 26 df, p=0.2
```

Logistic regression for low birth weight

```
# Define indicator for low birth weight (birth weight less than 2500mg)
complete_birthweight$low.weight = ifelse(complete_birthweight$Birthweight <= 2500, 1, 0)

fit.logis <- glm(low.weight~., data=complete_birthweight[,-1])
summary(fit.logis)
```

```
##
## Call:
## glm(formula = low.weight ~ ., data = complete_birthweight[, -1])
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -5.060e-02  7.707e-02  -0.657   0.5117
## Age           2.669e-04  1.537e-03   0.174   0.8622
## Preg.ended...37.wkYes 6.602e-01  2.480e-02 26.623 <2e-16 ***
## EducationLT 8 yrs   2.110e-02  2.097e-02   1.006   0.3147
## EducationMT 12 yrs -1.254e-02  1.969e-02  -0.637   0.5242
## HypertensionYes     2.468e-02  4.715e-02   0.523   0.6008
## DiabetesYes        -9.521e-02  4.736e-02  -2.010   0.0448 *
## Use.TobYes         2.785e-02  2.581e-02   1.079   0.2809
## Use.AlcYes         1.140e-03  6.035e-02   0.019   0.9849
```

```
## Drug.AddYes          2.559e-02  1.176e-01   0.218   0.8278
## Prev.pregYes         -2.179e-02  1.927e-02  -1.131   0.2585
## GroupT               3.305e-03  1.562e-02   0.212   0.8325
## OAA1                 -4.821e-05  3.720e-05  -1.296   0.1954
## OCR1                 4.681e-05  4.098e-05   1.142   0.2537
## OFN1                 2.145e-05  4.641e-05   0.462   0.6441
## OPG1                -5.807e-05  3.475e-05  -1.671   0.0952 .
## OPI1                 4.189e-05  3.878e-05   1.080   0.2804
## OTD1                -1.950e-05  4.232e-05  -0.461   0.6451
## OTF1                 6.885e-05  3.770e-05   1.826   0.0682 .
## OCRP1                1.120e-05  3.743e-05   0.299   0.7649
## O1B1                -3.479e-04  1.104e-03  -0.315   0.7528
## O61                  1.017e-03  1.672e-03   0.609   0.5429
## O81                  8.639e-06  2.550e-03   0.003   0.9973
## OPGE21              -2.941e-05  3.472e-05  -0.847   0.3972
## OTNF1                2.252e-04  7.702e-04   0.292   0.7701
## OMMP91              -5.142e-06  5.341e-05  -0.096   0.9233
## ETXU_CAT1           2.627e-03  7.673e-03   0.342   0.7322
## OFIBRIN1            6.645e-05  4.520e-05   1.470   0.1419
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.04558677)
##
##      Null deviance: 69.270  on 766  degrees of freedom
## Residual deviance: 33.689  on 739  degrees of freedom
## AIC: -162.47
##
## Number of Fisher Scoring iterations: 2
```

```
fit.logis.step <- step(fit.logis, direction = "backward")
```

```
## Start:  AIC=-162.47
## low.weight ~ Age + Preg.ended...37.wk + Education + Hypertension +
##      Diabetes + Use.Tob + Use.Alc + Drug.Add + Prev.preg + Group +
##      OAA1 + OCR1 + OFN1 + OPG1 + OPI1 + OTD1 + OTF1 + OCRP1 +
##      O1B1 + O61 + O81 + OPGE21 + OTNF1 + OMMP91 + ETXU_CAT1 +
##      OFIBRIN1
##
##              Df Deviance      AIC
## - Education      2   33.774 -164.53
## - O81              1   33.689 -164.47
## - Use.Alc          1   33.689 -164.47
## - OMMP91           1   33.689 -164.46
## - Age              1   33.690 -164.44
## - Group            1   33.691 -164.43
## - Drug.Add         1   33.691 -164.42
## - OTNF1            1   33.693 -164.39
## - OCRP1            1   33.693 -164.38
## - O1B1             1   33.693 -164.37
## - ETXU_CAT1        1   33.694 -164.35
## - OTD1             1   33.698 -164.25
## - OFN1             1   33.698 -164.25
## - Hypertension     1   33.701 -164.19
```

```

## - O61                1    33.706 -164.09
## - OPGE21             1    33.721 -163.73
## - Use.Tob            1    33.742 -163.27
## - OPI1              1    33.742 -163.26
## - Prev.preg         1    33.747 -163.15
## - OCR1              1    33.748 -163.12
## - OAA1              1    33.765 -162.73
## <none>               33.689 -162.47
## - OFIBRIN1          1    33.787 -162.23
## - OPG1              1    33.816 -161.58
## - OTF1              1    33.841 -161.02
## - Diabetes           1    33.873 -160.29
## - Preg.ended...37.wk 1    66.000  351.33
##
## Step:  AIC=-164.53
## low.weight ~ Age + Preg.ended...37.wk + Hypertension + Diabetes +
##      Use.Tob + Use.Alc + Drug.Add + Prev.preg + Group + OAA1 +
##      OCR1 + OFN1 + OPG1 + OPI1 + OTD1 + OTF1 + OCRP1 + O1B1 +
##      O61 + O81 + OPGE21 + OTNF1 + OMMP91 + ETXU_CAT1 + OFIBRIN1
##
##              Df Deviance      AIC
## - O81                1    33.774 -166.53
## - Use.Alc            1    33.774 -166.52
## - Age                1    33.775 -166.51
## - OMMP91             1    33.776 -166.49
## - Group              1    33.776 -166.48
## - Drug.Add           1    33.777 -166.46
## - OTNF1              1    33.778 -166.44
## - O1B1               1    33.780 -166.41
## - OCRP1              1    33.781 -166.36
## - Hypertension       1    33.782 -166.34
## - ETXU_CAT1          1    33.783 -166.34
## - OFN1               1    33.785 -166.29
## - OTD1               1    33.785 -166.28
## - O61                1    33.791 -166.14
## - OPGE21             1    33.806 -165.80
## - Prev.preg          1    33.822 -165.44
## - Use.Tob            1    33.822 -165.44
## - OPI1              1    33.828 -165.31
## - OCR1              1    33.841 -165.01
## - OAA1              1    33.851 -164.78
## <none>               33.774 -164.53
## - OFIBRIN1          1    33.866 -164.44
## - OPG1              1    33.902 -163.64
## - OTF1              1    33.916 -163.30
## - Diabetes           1    33.965 -162.20
## - Preg.ended...37.wk 1    66.003  347.37
##
## Step:  AIC=-166.53
## low.weight ~ Age + Preg.ended...37.wk + Hypertension + Diabetes +
##      Use.Tob + Use.Alc + Drug.Add + Prev.preg + Group + OAA1 +
##      OCR1 + OFN1 + OPG1 + OPI1 + OTD1 + OTF1 + OCRP1 + O1B1 +
##      O61 + OPGE21 + OTNF1 + OMMP91 + ETXU_CAT1 + OFIBRIN1
##

```

```

##                               Df Deviance      AIC
## - Use.Alc                     1   33.774 -168.52
## - Age                         1   33.775 -168.51
## - OMMP91                      1   33.776 -168.49
## - Group                       1   33.776 -168.48
## - Drug.Add                    1   33.777 -168.46
## - OTNF1                       1   33.780 -168.41
## - O1B1                        1   33.781 -168.37
## - OCRP1                       1   33.782 -168.36
## - Hypertension                1   33.782 -168.34
## - ETXU_CAT1                   1   33.783 -168.33
## - OFN1                        1   33.785 -168.29
## - OTD1                        1   33.785 -168.28
## - O61                         1   33.794 -168.07
## - OPGE21                      1   33.806 -167.80
## - Use.Tob                     1   33.822 -167.43
## - Prev.preg                   1   33.823 -167.43
## - OPI1                        1   33.828 -167.31
## - OCR1                        1   33.842 -167.00
## - OAA1                        1   33.852 -166.77
## <none>                        33.774 -166.53
## - OFIBRIN1                    1   33.868 -166.41
## - OPG1                        1   33.902 -165.63
## - OTF1                        1   33.918 -165.26
## - Diabetes                    1   33.965 -164.20
## - Preg.ended...37.wk         1   66.010  345.44
##
## Step:  AIC=-168.52
## low.weight ~ Age + Preg.ended...37.wk + Hypertension + Diabetes +
##      Use.Tob + Drug.Add + Prev.preg + Group + OAA1 + OCR1 + OFN1 +
##      OPG1 + OPI1 + OTD1 + OTF1 + OCRP1 + O1B1 + O61 + OPGE21 +
##      OTNF1 + OMMP91 + ETXU_CAT1 + OFIBRIN1
##
##                               Df Deviance      AIC
## - Age                         1   33.775 -170.50
## - OMMP91                      1   33.776 -170.48
## - Group                       1   33.777 -170.47
## - Drug.Add                    1   33.777 -170.45
## - OTNF1                       1   33.780 -170.40
## - O1B1                        1   33.781 -170.37
## - OCRP1                       1   33.782 -170.36
## - Hypertension                1   33.783 -170.34
## - ETXU_CAT1                   1   33.783 -170.33
## - OFN1                        1   33.785 -170.28
## - OTD1                        1   33.786 -170.26
## - O61                         1   33.794 -170.07
## - OPGE21                      1   33.807 -169.79
## - Prev.preg                   1   33.823 -169.43
## - Use.Tob                     1   33.823 -169.42
## - OPI1                        1   33.828 -169.31
## - OCR1                        1   33.842 -169.00
## - OAA1                        1   33.852 -168.77
## <none>                        33.774 -168.52
## - OFIBRIN1                    1   33.869 -168.39

```

```

## - OPG1          1  33.902 -167.62
## - OTF1          1  33.919 -167.26
## - Diabetes      1  33.965 -166.20
## - Preg.ended...37.wk 1  66.032  343.70
##
## Step: AIC=-170.5
## low.weight ~ Preg.ended...37.wk + Hypertension + Diabetes + Use.Tob +
##   Drug.Add + Prev.preg + Group + OAA1 + OCR1 + OFN1 + OPG1 +
##   OPI1 + OTD1 + OTF1 + OCRP1 + O1B1 + O61 + OPGE21 + OTNF1 +
##   OMMP91 + ETXU_CAT1 + OFIBRIN1
##
##           Df Deviance      AIC
## - OMMP91      1  33.777 -172.46
## - Group        1  33.778 -172.45
## - Drug.Add     1  33.778 -172.43
## - OTNF1        1  33.781 -172.38
## - O1B1         1  33.782 -172.35
## - OCRP1        1  33.783 -172.33
## - Hypertension 1  33.784 -172.30
## - ETXU_CAT1    1  33.784 -172.30
## - OFN1         1  33.786 -172.27
## - OTD1         1  33.786 -172.25
## - O61          1  33.795 -172.06
## - OPGE21       1  33.807 -171.78
## - Use.Tob      1  33.823 -171.41
## - Prev.preg    1  33.825 -171.38
## - OPI1         1  33.829 -171.29
## - OCR1         1  33.843 -170.97
## - OAA1         1  33.852 -170.77
## <none>         33.775 -170.50
## - OFIBRIN1     1  33.870 -170.36
## - OPG1         1  33.904 -169.59
## - OTF1         1  33.920 -169.22
## - Diabetes     1  33.966 -168.19
## - Preg.ended...37.wk 1  66.235  344.05
##
## Step: AIC=-172.46
## low.weight ~ Preg.ended...37.wk + Hypertension + Diabetes + Use.Tob +
##   Drug.Add + Prev.preg + Group + OAA1 + OCR1 + OFN1 + OPG1 +
##   OPI1 + OTD1 + OTF1 + OCRP1 + O1B1 + O61 + OPGE21 + OTNF1 +
##   ETXU_CAT1 + OFIBRIN1
##
##           Df Deviance      AIC
## - Group        1  33.780 -174.40
## - Drug.Add     1  33.780 -174.39
## - OTNF1        1  33.783 -174.33
## - O1B1         1  33.784 -174.30
## - OCRP1        1  33.785 -174.28
## - ETXU_CAT1    1  33.786 -174.27
## - Hypertension 1  33.786 -174.26
## - OFN1         1  33.788 -174.22
## - OTD1         1  33.789 -174.19
## - O61          1  33.798 -174.00
## - OPGE21       1  33.810 -173.72

```

```

## - Use.Tob          1  33.823 -173.41
## - Prev.preg        1  33.827 -173.33
## - OPI1             1  33.830 -173.27
## - OCR1             1  33.845 -172.92
## - OAA1             1  33.854 -172.73
## <none>             33.777 -172.46
## - OFIBRIN1        1  33.873 -172.29
## - OPG1            1  33.906 -171.54
## - OTF1            1  33.922 -171.18
## - Diabetes         1  33.971 -170.06
## - Preg.ended...37.wk 1  66.344  343.31
##
## Step:  AIC=-174.4
## low.weight ~ Preg.ended...37.wk + Hypertension + Diabetes + Use.Tob +
##      Drug.Add + Prev.preg + OAA1 + OCR1 + OFN1 + OPG1 + OPI1 +
##      OTD1 + OTF1 + OCRP1 + O1B1 + O61 + OPGE21 + OTNF1 + ETXU_CAT1 +
##      OFIBRIN1
##
##              Df Deviance      AIC
## - Drug.Add      1  33.783 -176.33
## - OTNF1          1  33.786 -176.27
## - O1B1           1  33.786 -176.25
## - OCRP1          1  33.787 -176.24
## - ETXU_CAT1      1  33.788 -176.22
## - Hypertension   1  33.789 -176.20
## - OFN1           1  33.790 -176.16
## - OTD1           1  33.792 -176.13
## - O61            1  33.800 -175.95
## - OPGE21         1  33.811 -175.68
## - Use.Tob        1  33.826 -175.35
## - Prev.preg      1  33.829 -175.28
## - OPI1           1  33.833 -175.20
## - OCR1           1  33.847 -174.87
## - OAA1           1  33.857 -174.64
## <none>           33.780 -174.40
## - OFIBRIN1      1  33.875 -174.24
## - OPG1           1  33.908 -173.50
## - OTF1           1  33.924 -173.14
## - Diabetes       1  33.972 -172.05
## - Preg.ended...37.wk 1  66.376  341.69
##
## Step:  AIC=-176.33
## low.weight ~ Preg.ended...37.wk + Hypertension + Diabetes + Use.Tob +
##      Prev.preg + OAA1 + OCR1 + OFN1 + OPG1 + OPI1 + OTD1 + OTF1 +
##      OCRP1 + O1B1 + O61 + OPGE21 + OTNF1 + ETXU_CAT1 + OFIBRIN1
##
##              Df Deviance      AIC
## - OTNF1          1  33.789 -178.19
## - OCRP1          1  33.790 -178.18
## - O1B1           1  33.790 -178.17
## - ETXU_CAT1      1  33.791 -178.14
## - Hypertension   1  33.792 -178.13
## - OFN1           1  33.795 -178.06
## - OTD1           1  33.795 -178.05

```

```

## - O61                1    33.803 -177.87
## - OPGE21             1    33.814 -177.62
## - Prev.preg          1    33.833 -177.20
## - Use.Tob            1    33.834 -177.18
## - OPI1               1    33.836 -177.13
## - OCR1               1    33.850 -176.80
## - OAA1               1    33.862 -176.54
## <none>                33.783 -176.33
## - OFIBRIN1           1    33.879 -176.14
## - OPG1               1    33.913 -175.37
## - OTF1               1    33.927 -175.08
## - Diabetes           1    33.976 -173.96
## - Preg.ended...37.wk 1    66.426  340.27
##
## Step:  AIC=-178.19
## low.weight ~ Preg.ended...37.wk + Hypertension + Diabetes + Use.Tob +
##      Prev.preg + OAA1 + OCR1 + OFN1 + OPG1 + OPI1 + OTD1 + OTF1 +
##      OCRP1 + O1B1 + O61 + OPGE21 + ETXU_CAT1 + OFIBRIN1
##
##              Df Deviance      AIC
## - O1B1          1    33.791 -180.15
## - OCRP1          1    33.795 -180.05
## - Hypertension   1    33.798 -179.99
## - ETXU_CAT1      1    33.799 -179.98
## - OFN1           1    33.800 -179.94
## - OTD1           1    33.800 -179.93
## - O61            1    33.818 -179.53
## - OPGE21         1    33.820 -179.49
## - Prev.preg      1    33.838 -179.07
## - Use.Tob        1    33.839 -179.05
## - OPI1           1    33.843 -178.97
## - OCR1           1    33.861 -178.55
## - OAA1           1    33.872 -178.30
## <none>            33.789 -178.19
## - OFIBRIN1       1    33.883 -178.06
## - OPG1           1    33.920 -177.22
## - OTF1           1    33.929 -177.02
## - Diabetes       1    33.981 -175.85
## - Preg.ended...37.wk 1    66.510  339.24
##
## Step:  AIC=-180.15
## low.weight ~ Preg.ended...37.wk + Hypertension + Diabetes + Use.Tob +
##      Prev.preg + OAA1 + OCR1 + OFN1 + OPG1 + OPI1 + OTD1 + OTF1 +
##      OCRP1 + O61 + OPGE21 + ETXU_CAT1 + OFIBRIN1
##
##              Df Deviance      AIC
## - OCRP1          1    33.797 -182.01
## - ETXU_CAT1       1    33.799 -181.96
## - Hypertension    1    33.800 -181.95
## - OTD1            1    33.802 -181.90
## - OFN1            1    33.802 -181.90
## - OPGE21          1    33.822 -181.45
## - O61             1    33.824 -181.39
## - Prev.preg       1    33.839 -181.06

```



```

## - Use.Tob          1  33.841 -181.00
## - OPI1             1  33.844 -180.94
## - OCR1             1  33.862 -180.54
## - OAA1             1  33.874 -180.27
## <none>             33.791 -180.15
## - OFIBRIN1        1  33.884 -180.03
## - OPG1            1  33.920 -179.21
## - OTF1            1  33.936 -178.86
## - Diabetes         1  33.983 -177.80
## - Preg.ended...37.wk 1  66.514  337.28
##
## Step:  AIC=-182.01
## low.weight ~ Preg.ended...37.wk + Hypertension + Diabetes + Use.Tob +
##      Prev.preg + OAA1 + OCR1 + OFN1 + OPG1 + OPI1 + OTD1 + OTF1 +
##      O61 + OPGE21 + ETXU_CAT1 + OFIBRIN1
##
##              Df Deviance      AIC
## - ETXU_CAT1      1  33.805 -183.83
## - Hypertension    1  33.806 -183.80
## - OTD1            1  33.806 -183.80
## - OFN1            1  33.808 -183.76
## - OPGE21          1  33.828 -183.32
## - O61             1  33.831 -183.25
## - Prev.preg       1  33.842 -182.99
## - Use.Tob         1  33.847 -182.88
## - OPI1            1  33.852 -182.75
## - OCR1            1  33.870 -182.36
## - OAA1            1  33.881 -182.12
## <none>            33.797 -182.01
## - OFIBRIN1        1  33.891 -181.88
## - OPG1            1  33.929 -181.03
## - OTF1            1  33.937 -180.83
## - Diabetes         1  33.992 -179.60
## - Preg.ended...37.wk 1  66.527  335.44
##
## Step:  AIC=-183.83
## low.weight ~ Preg.ended...37.wk + Hypertension + Diabetes + Use.Tob +
##      Prev.preg + OAA1 + OCR1 + OFN1 + OPG1 + OPI1 + OTD1 + OTF1 +
##      O61 + OPGE21 + OFIBRIN1
##
##              Df Deviance      AIC
## - Hypertension    1  33.815 -185.60
## - OTD1            1  33.815 -185.60
## - OFN1            1  33.816 -185.57
## - OPGE21          1  33.836 -185.12
## - O61             1  33.849 -184.83
## - Prev.preg       1  33.852 -184.77
## - Use.Tob         1  33.852 -184.76
## - OPI1            1  33.863 -184.52
## - OCR1            1  33.881 -184.11
## - OAA1            1  33.892 -183.86
## <none>            33.805 -183.83
## - OFIBRIN1        1  33.906 -183.54
## - OPG1            1  33.935 -182.88

```

```

## - OTF1          1  33.939 -182.80
## - Diabetes      1  34.004 -181.34
## - Preg.ended...37.wk 1  66.528  333.45
##
## Step: AIC=-185.6
## low.weight ~ Preg.ended...37.wk + Diabetes + Use.Tob + Prev.preg +
##      OAA1 + OCR1 + OFN1 + OPG1 + OPI1 + OTD1 + OTF1 + O61 + OPGE21 +
##      OFIBRIN1
##
##              Df Deviance      AIC
## - OTD1          1  33.824 -187.39
## - OFN1          1  33.826 -187.35
## - OPGE21        1  33.848 -186.85
## - O61           1  33.860 -186.59
## - Prev.preg     1  33.862 -186.54
## - Use.Tob       1  33.863 -186.50
## - OPI1          1  33.870 -186.35
## - OCR1          1  33.888 -185.94
## - OAA1          1  33.902 -185.63
## <none>          33.815 -185.60
## - OFIBRIN1     1  33.914 -185.35
## - OPG1          1  33.943 -184.70
## - OTF1          1  33.955 -184.44
## - Diabetes      1  34.007 -183.27
## - Preg.ended...37.wk 1  67.761  345.53
##
## Step: AIC=-187.39
## low.weight ~ Preg.ended...37.wk + Diabetes + Use.Tob + Prev.preg +
##      OAA1 + OCR1 + OFN1 + OPG1 + OPI1 + OTF1 + O61 + OPGE21 +
##      OFIBRIN1
##
##              Df Deviance      AIC
## - OFN1          1  33.831 -189.24
## - OPGE21        1  33.857 -188.64
## - O61           1  33.871 -188.34
## - OPI1          1  33.872 -188.31
## - Prev.preg     1  33.872 -188.31
## - Use.Tob       1  33.873 -188.28
## - OCR1          1  33.894 -187.81
## <none>          33.824 -187.39
## - OAA1          1  33.913 -187.37
## - OFIBRIN1     1  33.924 -187.14
## - OPG1          1  33.962 -186.28
## - OTF1          1  33.963 -186.26
## - Diabetes      1  34.016 -185.06
## - Preg.ended...37.wk 1  67.794  343.90
##
## Step: AIC=-189.24
## low.weight ~ Preg.ended...37.wk + Diabetes + Use.Tob + Prev.preg +
##      OAA1 + OCR1 + OPG1 + OPI1 + OTF1 + O61 + OPGE21 + OFIBRIN1
##
##              Df Deviance      AIC
## - OPGE21        1  33.866 -190.45
## - O61           1  33.873 -190.28

```

```

## - Use.Tob          1  33.877 -190.19
## - Prev.preg        1  33.879 -190.16
## - OPI1             1  33.887 -189.96
## - OAA1             1  33.917 -189.29
## <none>             33.831 -189.24
## - OFIBRIN1        1  33.928 -189.04
## - OCR1            1  33.936 -188.87
## - OPG1            1  33.965 -188.22
## - OTF1            1  33.975 -187.98
## - Diabetes         1  34.022 -186.92
## - Preg.ended...37.wk 1  67.795  341.91
##
## Step:  AIC=-190.45
## low.weight ~ Preg.ended...37.wk + Diabetes + Use.Tob + Prev.preg +
##      OAA1 + OCR1 + OPG1 + OPI1 + OTF1 + O61 + OFIBRIN1
##
##              Df Deviance      AIC
## - O61          1  33.909 -191.47
## - Use.Tob      1  33.912 -191.41
## - Prev.preg    1  33.917 -191.29
## - OPI1         1  33.919 -191.25
## <none>         33.866 -190.45
## - OFIBRIN1     1  33.955 -190.44
## - OAA1         1  33.958 -190.38
## - OCR1         1  33.973 -190.04
## - OPG1         1  34.007 -189.27
## - OTF1         1  34.013 -189.12
## - Diabetes     1  34.059 -188.08
## - Preg.ended...37.wk 1  68.160  344.03
##
## Step:  AIC=-191.47
## low.weight ~ Preg.ended...37.wk + Diabetes + Use.Tob + Prev.preg +
##      OAA1 + OCR1 + OPG1 + OPI1 + OTF1 + OFIBRIN1
##
##              Df Deviance      AIC
## - Use.Tob      1  33.951 -192.52
## - Prev.preg    1  33.959 -192.34
## - OPI1         1  33.972 -192.05
## <none>         33.909 -191.47
## - OAA1         1  34.009 -191.21
## - OCR1         1  34.017 -191.03
## - OFIBRIN1     1  34.020 -190.97
## - OTF1         1  34.035 -190.62
## - OPG1         1  34.066 -189.93
## - Diabetes     1  34.105 -189.06
## - Preg.ended...37.wk 1  68.209  342.58
##
## Step:  AIC=-192.52
## low.weight ~ Preg.ended...37.wk + Diabetes + Prev.preg + OAA1 +
##      OCR1 + OPG1 + OPI1 + OTF1 + OFIBRIN1
##
##              Df Deviance      AIC
## - Prev.preg    1  33.997 -193.49
## - OPI1         1  34.008 -193.23

```

```

## <none>                33.951 -192.52
## - OCR1                1  34.050 -192.29
## - OFIBRIN1            1  34.062 -192.01
## - OAA1                1  34.063 -191.99
## - OTF1                1  34.090 -191.39
## - OPG1                1  34.119 -190.74
## - Diabetes            1  34.149 -190.05
## - Preg.ended...37.wk  1  68.389  342.60
##
## Step:  AIC=-193.49
## low.weight ~ Preg.ended...37.wk + Diabetes + OAA1 + OCR1 + OPG1 +
##      OPI1 + OTF1 + OFIBRIN1
##
##              Df Deviance      AIC
## - OPI1        1  34.051 -194.28
## <none>                33.997 -193.49
## - OCR1        1  34.097 -193.23
## - OFIBRIN1    1  34.103 -193.11
## - OAA1        1  34.103 -193.10
## - OTF1        1  34.133 -192.43
## - OPG1        1  34.162 -191.76
## - Diabetes    1  34.199 -190.94
## - Preg.ended...37.wk  1  68.420  340.95
##
## Step:  AIC=-194.28
## low.weight ~ Preg.ended...37.wk + Diabetes + OAA1 + OCR1 + OPG1 +
##      OTF1 + OFIBRIN1
##
##              Df Deviance      AIC
## <none>                34.051 -194.28
## - OAA1        1  34.143 -194.19
## - OFIBRIN1    1  34.157 -193.89
## - OCR1        1  34.188 -193.19
## - OPG1        1  34.212 -192.65
## - OTF1        1  34.215 -192.58
## - Diabetes    1  34.248 -191.84
## - Preg.ended...37.wk  1  68.444  339.22

```

```
summary(fit.logis.step)
```

```

##
## Call:
## glm(formula = low.weight ~ Preg.ended...37.wk + Diabetes + OAA1 +
##      OCR1 + OPG1 + OTF1 + OFIBRIN1, data = complete_birthweight[,
##      -1])
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -3.240e-03  3.146e-02  -0.103   0.9180
## Preg.ended...37.wkYes  6.573e-01  2.374e-02  27.688 <2e-16 ***
## DiabetesYes      -9.693e-02  4.618e-02  -2.099   0.0362 *
## OAA1             -5.160e-05  3.591e-05  -1.437   0.1511
## OCR1              6.230e-05  3.563e-05   1.748   0.0808 .
## OPG1             -6.383e-05  3.367e-05  -1.895   0.0584 .

```

```
## OTF1                6.686e-05  3.493e-05   1.914   0.0560 .
## OFIBRIN1            6.689e-05  4.347e-05   1.539   0.1242
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.04486253)
##
##      Null deviance: 69.270  on 766  degrees of freedom
## Residual deviance: 34.051  on 759  degrees of freedom
## AIC: -194.28
##
## Number of Fisher Scoring iterations: 2
```

Exploratory Plots for extended data

```
library(survival)
library(survminer)

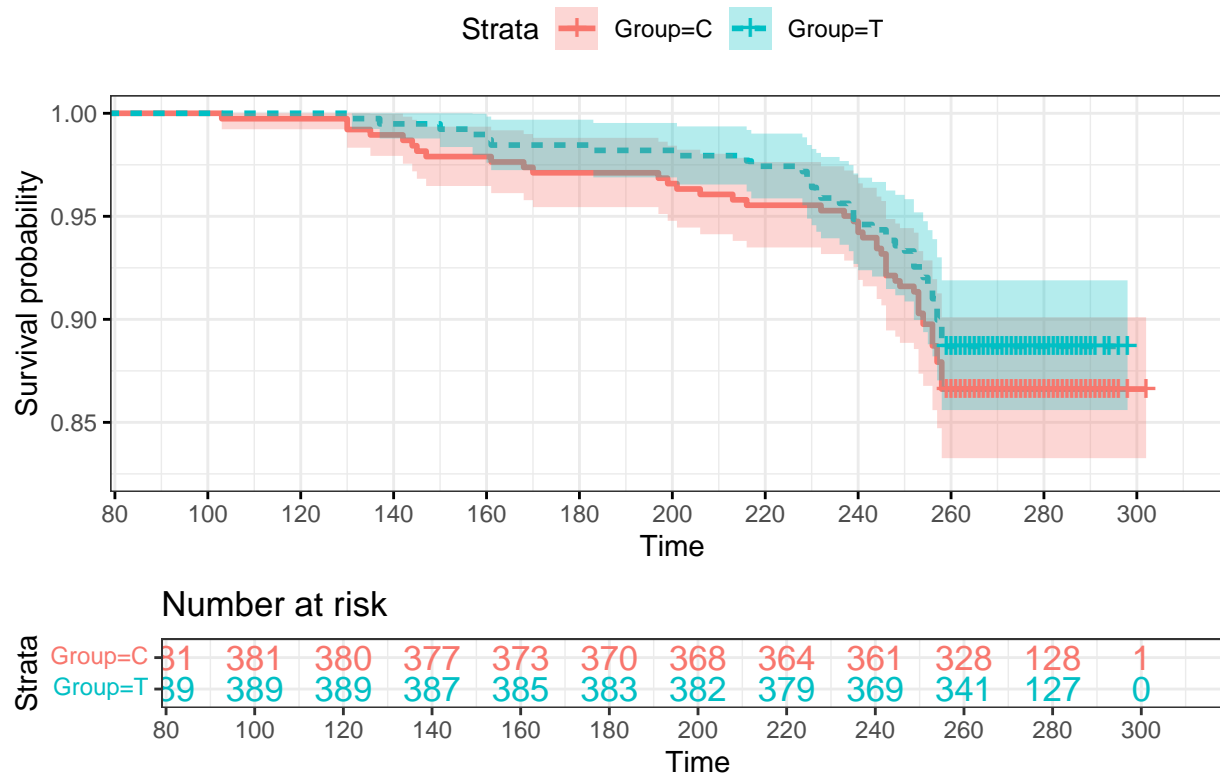
complete_preterm_extended$Preg.ended...37.wk <- ifelse(complete_preterm_extended$Preg.ended...37.wk ==
surv_obj <- with(complete_preterm_extended, Surv(time = GA.at.outcome, event = Preg.ended...37.wk, type
fit.ext <- survfit(surv_obj ~ Group, data = complete_preterm_extended)

ggs <- ggsurvplot(fit.ext,
  pval = TRUE, conf.int = TRUE,
  risk.table = TRUE, # Add risk table
  risk.table.col = "strata", # Change risk table color by groups
  linetype = "strata", # Change line type by groups
  ggtheme = theme_bw(), # Change ggplot2 theme
  palette = c("#F8766D", "#00BFC4"),
  ylim=c(0.825,1),
  xlim=c(90,310),
  title = 'Survival Curves for Control and Treatment Cohort',
  break.x.by = 20)

#ggs$plot <- ggs$plot + ylim(c(0.75,1)) + xlim(c(90, 310))
ggs

## Warning: Removed 1 row containing missing values or values outside the scale range
## (`geom_text()`).
## Removed 1 row containing missing values or values outside the scale range
## (`geom_text()`).
```

Survival Curves for Control and Treatment Cohort



```
surv_diff1_ext <- survdiff(Surv(time = GA.at.outcome, event = Preg.ended...37.wk, type = 'right') ~ Group, data = complete_preterm_extended)
surv_diff1_ext
```

```
## Call:
## survdiff(formula = Surv(time = GA.at.outcome, event = Preg.ended...37.wk,
##   type = "right") ~ Group, data = complete_preterm_extended)
##
##           N Observed Expected (O-E)^2/E (O-E)^2/V
## Group=C 381      51      46.7      0.405      0.799
## Group=T 389      44      48.3      0.391      0.799
##
## Chisq= 0.8  on 1 degrees of freedom, p= 0.4
```

Only checking the data pre-32.8 weeks aka 231 days or fewer

```
pre32_ext <- complete_preterm_extended

pre32_ext$Preg.ended...37.wk <- NULL
pre32_ext$pre230 <- ifelse(pre32_ext$GA.at.outcome <= 230, 1, 0)

# new survival fit
surv_obj2 <- with(pre32_ext, Surv(time = GA.at.outcome, event = pre230, type = 'right'))
fit2_ext <- survfit(surv_obj2 ~ Group, data = pre32_ext)
```

```

ggs2 <- ggsurvplot(fit2.ext,
  pval = TRUE, conf.int = TRUE,
  risk.table = TRUE, # Add risk table
  risk.table.col = "strata", # Change risk table color by groups
  linetype = "strata", # Change line type by groups
  ggtheme = theme_bw(), # Change ggplot2 theme
  palette = c("#F8766D", "#00BFC4"),
  ylim=c(0.825,1),
  xlim=c(90,240),
  title = 'Survival Curves for Control and Treatment Cohort',
  break.x.by = 20)

#ggs$plot <- ggs$plot + ylim(c(0.75,1)) + xlim(c(90, 310))
ggs2

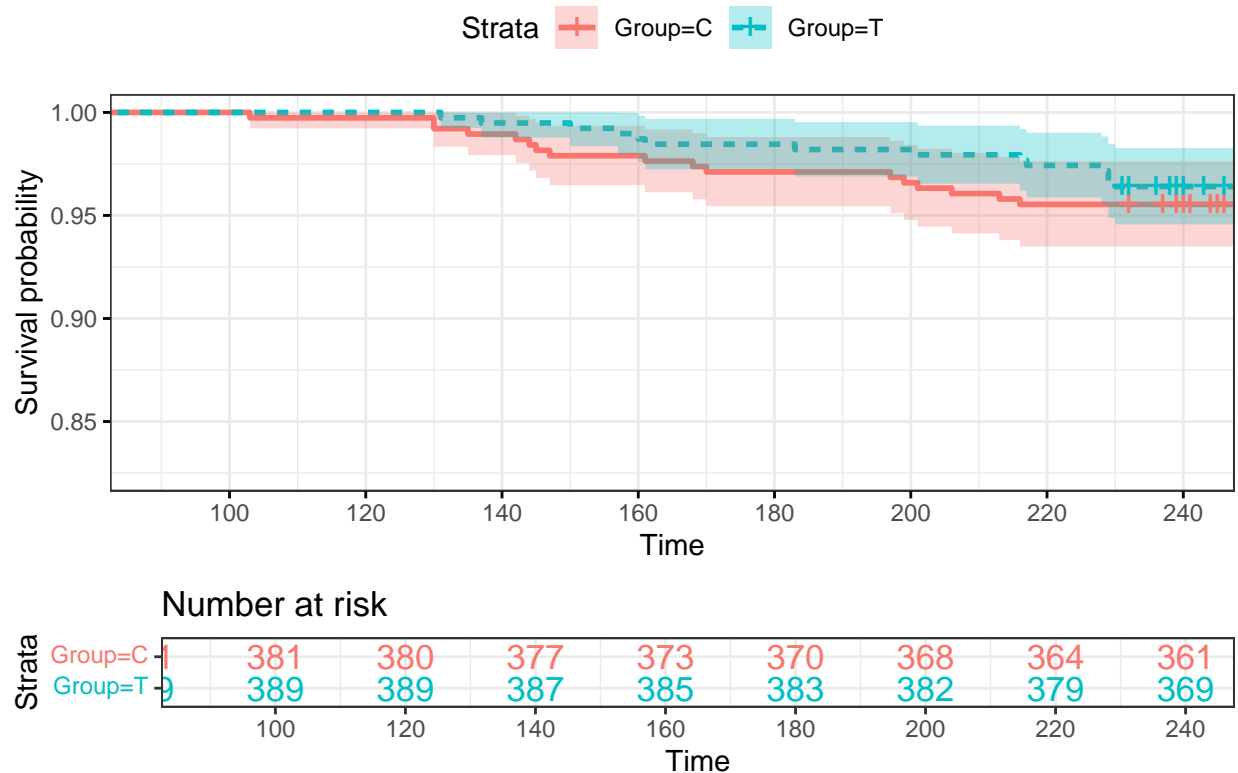
```

```

## Warning: Removed 1 row containing missing values or values outside the scale range
## (`geom_text()`).
## Removed 1 row containing missing values or values outside the scale range
## (`geom_text()`).

```

Survival Curves for Control and Treatment Cohort



```

surv_diff2_ext <- survdiff(Surv(time = GA.at.outcome, event = pre230, type = 'right') ~ Group, data = p
surv_diff2_ext

```

```
## Call:
```

```
## survdiff(formula = Surv(time = GA.at.outcome, event = pre230,
##      type = "right") ~ Group, data = pre32_ext)
##
##           N Observed Expected (O-E)^2/E (O-E)^2/V
## Group=C 381      17    15.2     0.202     0.398
## Group=T 389      14    15.8     0.196     0.398
##
##  Chisq= 0.4   on 1 degrees of freedom, p= 0.5
```

Cox Proportional Hazards for preterm birth

```
# Full-length of the observation period
fit.coxph.extend <- coxph(Surv(time = GA.at.outcome, event = Preg.ended...37.wk)~., data=complete_preterm)
summary(fit.coxph.extend)
```

```
## Call:
## coxph(formula = Surv(time = GA.at.outcome, event = Preg.ended...37.wk) ~
##      ., data = complete_preterm_extended)
##
##      n= 770, number of events= 95
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## Age              0.02917   1.02960  0.01982   1.472 0.141053
## EducationLT 8 yrs -0.23836   0.78792  0.31716 -0.752 0.452318
## EducationMT 12 yrs 0.19809   1.21908  0.24356  0.813 0.416034
## HypertensionYes    1.21544   3.37178  0.36408  3.338 0.000843 ***
## DiabetesYes        0.94257   2.56657  0.40052  2.353 0.018603 *
## Use.TobYes         0.34032   1.40539  0.31242  1.089 0.276019
## Use.AlcYes        -0.49980   0.60665  0.81352 -0.614 0.538973
## Drug.AddYes        1.31383   3.72039  1.13191  1.161 0.245757
## Prev.pregYes       0.22998   1.25857  0.27944  0.823 0.410517
## GroupT            -0.32660   0.72137  0.21318 -1.532 0.125521
## PC1               -0.11290   0.89324  0.06592 -1.713 0.086782 .
## PC2                0.11181   1.11830  0.07122  1.570 0.116441
## PC3               -0.04923   0.95196  0.09888 -0.498 0.618577
## PC4               -0.18203   0.83357  0.10517 -1.731 0.083484 .
## PC5               -0.10950   0.89629  0.10420 -1.051 0.293331
## PC6               -0.02876   0.97165  0.10516 -0.273 0.784494
## PC7               -0.02080   0.97941  0.10895 -0.191 0.848560
## PC8               -0.17238   0.84166  0.11088 -1.555 0.120019
## PC9               -0.07256   0.93001  0.12078 -0.601 0.547975
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## Age              1.0296    0.9712    0.9904    1.070
## EducationLT 8 yrs  0.7879    1.2692    0.4232    1.467
## EducationMT 12 yrs 1.2191    0.8203    0.7563    1.965
## HypertensionYes    3.3718    0.2966    1.6518    6.883
## DiabetesYes        2.5666    0.3896    1.1707    5.627
## Use.TobYes         1.4054    0.7115    0.7618    2.593
## Use.AlcYes         0.6067    1.6484    0.1232    2.988
```



```
## Drug.AddYes          3.7204      0.2688      0.4047      34.204
## Prev.pregYes         1.2586      0.7946      0.7278      2.176
## GroupT               0.7214      1.3862      0.4750      1.096
## PC1                  0.8932      1.1195      0.7850      1.016
## PC2                  1.1183      0.8942      0.9726      1.286
## PC3                  0.9520      1.0505      0.7843      1.156
## PC4                  0.8336      1.1997      0.6783      1.024
## PC5                  0.8963      1.1157      0.7307      1.099
## PC6                  0.9717      1.0292      0.7907      1.194
## PC7                  0.9794      1.0210      0.7911      1.213
## PC8                  0.8417      1.1881      0.6773      1.046
## PC9                  0.9300      1.0753      0.7340      1.178
##
## Concordance= 0.681 (se = 0.028 )
## Likelihood ratio test= 47.76 on 19 df, p=3e-04
## Wald test              = 59.81 on 19 df, p=4e-06
## Score (logrank) test = 69.15 on 19 df, p=1e-07
```

```
coef = round(cbind(summary(fit.coxph.extend)$coefficients[, c(2,5)], summary(fit.coxph.extend)$conf.int
colnames(coef) = c("Hazard", "P value", "2.5%", "97.5%")
knitr::kable(coef)
```

	Hazard	P value	2.5%	97.5%
Age	1.030	0.141	0.990	1.070
EducationLT 8 yrs	0.788	0.452	0.423	1.467
EducationMT 12 yrs	1.219	0.416	0.756	1.965
HypertensionYes	3.372	0.001	1.652	6.883
DiabetesYes	2.567	0.019	1.171	5.627
Use.TobYes	1.405	0.276	0.762	2.593
Use.AlcYes	0.607	0.539	0.123	2.988
Drug.AddYes	3.720	0.246	0.405	34.204
Prev.pregYes	1.259	0.411	0.728	2.176
GroupT	0.721	0.126	0.475	1.096
PC1	0.893	0.087	0.785	1.016
PC2	1.118	0.116	0.973	1.286
PC3	0.952	0.619	0.784	1.156
PC4	0.834	0.083	0.678	1.024
PC5	0.896	0.293	0.731	1.099
PC6	0.972	0.784	0.791	1.194
PC7	0.979	0.849	0.791	1.213
PC8	0.842	0.120	0.677	1.046
PC9	0.930	0.548	0.734	1.178

```
# 33 weeks of the observation period
fit.coxph2.extend <- coxph(Surv(time = GA.at.outcome, event = pre230)~., data=pre32_ext)
summary(fit.coxph2.extend)
```

```
## Call:
## coxph(formula = Surv(time = GA.at.outcome, event = pre230) ~
##       ., data = pre32_ext)
##
```

```
## n= 770, number of events= 31
##
##          coef exp(coef) se(coef)      z Pr(>|z|)
## Age      -0.03816   0.96256  0.03692 -1.034  0.3013
## EducationLT 8 yrs -0.53031   0.58842  0.63778 -0.832  0.4057
## EducationMT 12 yrs  0.55324   1.73888  0.41544  1.332  0.1830
## HypertensionYes    1.11255   3.04210  0.69435  1.602  0.1091
## DiabetesYes        1.16563   3.20796  0.66733  1.747  0.0807
## Use.TobYes         -0.41144   0.66270  0.68684 -0.599  0.5492
## Use.AlcYes         -0.73771   0.47821  1.54240 -0.478  0.6324
## Drug.AddYes        2.69194  14.76023  1.60254  1.680  0.0930
## Prev.pregYes       0.39203   1.47998  0.46815  0.837  0.4024
## GroupT            -0.46537   0.62790  0.38252 -1.217  0.2238
## PC1               -0.20253   0.81666  0.12960 -1.563  0.1181
## PC2               -0.13348   0.87504  0.12985 -1.028  0.3040
## PC3                0.06723   1.06954  0.16950  0.397  0.6916
## PC4                0.05609   1.05770  0.17416  0.322  0.7474
## PC5               -0.02139   0.97884  0.18413 -0.116  0.9075
## PC6                0.14587   1.15704  0.17867  0.816  0.4143
## PC7                0.27026   1.31030  0.19153  1.411  0.1582
## PC8               -0.14901   0.86156  0.19100 -0.780  0.4353
## PC9               -0.23522   0.79040  0.20908 -1.125  0.2606
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##          exp(coef) exp(-coef) lower .95 upper .95
## Age              0.9626    1.03890    0.89536    1.035
## EducationLT 8 yrs  0.5884    1.69947    0.16858    2.054
## EducationMT 12 yrs  1.7389    0.57508    0.77028    3.925
## HypertensionYes    3.0421    0.32872    0.78008   11.863
## DiabetesYes        3.2080    0.31172    0.86735   11.865
## Use.TobYes         0.6627    1.50899    0.17245    2.547
## Use.AlcYes         0.4782    2.09114    0.02327    9.829
## Drug.AddYes       14.7602    0.06775    0.63828  341.332
## Prev.pregYes       1.4800    0.67568    0.59124    3.705
## GroupT            0.6279    1.59260    0.29668    1.329
## PC1               0.8167    1.22449    0.63347    1.053
## PC2               0.8750    1.14280    0.67842    1.129
## PC3               1.0695    0.93498    0.76722    1.491
## PC4               1.0577    0.94545    0.75182    1.488
## PC5               0.9788    1.02162    0.68231    1.404
## PC6               1.1570    0.86427    0.81520    1.642
## PC7               1.3103    0.76318    0.90021    1.907
## PC8               0.8616    1.16069    0.59253    1.253
## PC9               0.7904    1.26518    0.52466    1.191
##
## Concordance= 0.717 (se = 0.051 )
## Likelihood ratio test= 22.47 on 19 df,  p=0.3
## Wald test              = 24.8 on 19 df,  p=0.2
## Score (logrank) test = 28.37 on 19 df,  p=0.08
```

```
coef2 = round(cbind(summary(fit.coxph2.extend)$coefficients[, c(2,5)], summary(fit.coxph2.extend)$conf.
colnames(coef2) = c("Hazard", "P value", "2.5%", "97.5%")
knitr::kable(coef2)
```

	Hazard	P value	2.5%	97.5%
Age	0.963	0.301	0.895	1.035
EducationLT 8 yrs	0.588	0.406	0.169	2.054
EducationMT 12 yrs	1.739	0.183	0.770	3.925
HypertensionYes	3.042	0.109	0.780	11.863
DiabetesYes	3.208	0.081	0.867	11.865
Use.TobYes	0.663	0.549	0.172	2.547
Use.AlcYes	0.478	0.632	0.023	9.829
Drug.AddYes	14.760	0.093	0.638	341.332
Prev.pregYes	1.480	0.402	0.591	3.705
GroupT	0.628	0.224	0.297	1.329
PC1	0.817	0.118	0.633	1.053
PC2	0.875	0.304	0.678	1.129
PC3	1.070	0.692	0.767	1.491
PC4	1.058	0.747	0.752	1.488
PC5	0.979	0.908	0.682	1.404
PC6	1.157	0.414	0.815	1.642
PC7	1.310	0.158	0.900	1.907
PC8	0.862	0.435	0.593	1.253
PC9	0.790	0.261	0.525	1.191

Logistic regression for low birth weight

```
# Define indicator for low birth weight (birth weight less than 2500mg)
complete_birthweight_extended$low.weight = ifelse(complete_birthweight_extended$Birthweight <= 2500, 1,

fit.logis.extend <- glm(low.weight~., data=complete_birthweight_extended[, -1])
summary(fit.logis.extend)
```

```
##
## Call:
## glm(formula = low.weight ~ ., data = complete_birthweight_extended[,
##      -1])
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.0257912  0.0380648   0.678  0.4983
## Age            0.0002958  0.0015252   0.194  0.8463
## Preg.ended...37.wkYes 0.6585377  0.0246351  26.732 <2e-16 ***
## EducationLT 8 yrs    0.0204414  0.0208579   0.980  0.3274
## EducationMT 12 yrs  -0.0152881  0.0195107  -0.784  0.4335
## HypertensionYes     0.0249145  0.0468241   0.532  0.5948
## DiabetesYes        -0.0948514  0.0469569  -2.020  0.0437 *
## Use.TobYes         0.0275443  0.0255394   1.079  0.2812
## Use.AlcYes         0.0012874  0.0600441   0.021  0.9829
## Drug.AddYes        0.0210102  0.1166075   0.180  0.8571
## Prev.pregYes      -0.0207677  0.0190639  -1.089  0.2763
## GroupT            0.0032362  0.0155414   0.208  0.8351
## PC1               0.0042191  0.0040491   1.042  0.2978
## PC2              -0.0051137  0.0052848  -0.968  0.3335
## PC3              -0.0167763  0.0072153  -2.325  0.0203 *
```

```
## PC4          0.0068448  0.0075772   0.903   0.3666
## PC5          0.0098890  0.0076209   1.298   0.1948
## PC6         -0.0065971  0.0077635  -0.850   0.3957
## PC7         -0.0109290  0.0079618  -1.373   0.1703
## PC8          0.0125149  0.0080042   1.564   0.1183
## PC9          0.0048154  0.0086789   0.555   0.5792
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.04527911)
##
## Null deviance: 69.270  on 766  degrees of freedom
## Residual deviance: 33.778  on 746  degrees of freedom
## AIC: -174.44
##
## Number of Fisher Scoring iterations: 2
```

```
fit.logis.step.extend <- step(fit.logis.extend, direction = "backward")
```

```
## Start:  AIC=-174.44
## low.weight ~ Age + Preg.ended...37.wk + Education + Hypertension +
## Diabetes + Use.Tob + Use.Alc + Drug.Add + Prev.preg + Group +
## PC1 + PC2 + PC3 + PC4 + PC5 + PC6 + PC7 + PC8 + PC9
##
##           Df Deviance    AIC
## - Use.Alc      1  33.778 -176.44
## - Drug.Add      1  33.780 -176.40
## - Age           1  33.780 -176.40
## - Group         1  33.780 -176.39
## - Education     2  33.874 -176.27
## - Hypertension  1  33.791 -176.15
## - PC9           1  33.792 -176.12
## - PC6           1  33.811 -175.69
## - PC4           1  33.815 -175.60
## - PC2           1  33.821 -175.47
## - PC1           1  33.827 -175.32
## - Use.Tob       1  33.831 -175.24
## - Prev.preg     1  33.832 -175.22
## - PC5           1  33.854 -174.71
## - PC7           1  33.864 -174.50
## <none>          33.778 -174.44
## - PC8           1  33.889 -173.93
## - Diabetes      1  33.963 -172.25
## - PC3           1  34.023 -170.90
## - Preg.ended...37.wk 1  66.134 338.88
##
## Step:  AIC=-176.44
## low.weight ~ Age + Preg.ended...37.wk + Education + Hypertension +
## Diabetes + Use.Tob + Drug.Add + Prev.preg + Group + PC1 +
## PC2 + PC3 + PC4 + PC5 + PC6 + PC7 + PC8 + PC9
##
##           Df Deviance    AIC
## - Age           1  33.780 -178.40
## - Drug.Add       1  33.780 -178.39
```

```

## - Group          1  33.780 -178.39
## - Education      2  33.874 -178.26
## - Hypertension   1  33.791 -178.14
## - PC9            1  33.792 -178.12
## - PC6            1  33.811 -177.69
## - PC4            1  33.815 -177.60
## - PC2            1  33.821 -177.46
## - PC1            1  33.827 -177.32
## - Prev.preg      1  33.832 -177.21
## - Use.Tob        1  33.833 -177.19
## - PC5            1  33.854 -176.71
## - PC7            1  33.864 -176.50
## <none>           33.778 -176.44
## - PC8            1  33.889 -175.93
## - Diabetes       1  33.963 -174.25
## - PC3            1  34.023 -172.90
## - Preg.ended...37.wk 1  66.152  337.10
##
## Step: AIC=-178.4
## low.weight ~ Preg.ended...37.wk + Education + Hypertension +
## Diabetes + Use.Tob + Drug.Add + Prev.preg + Group + PC1 +
## PC2 + PC3 + PC4 + PC5 + PC6 + PC7 + PC8 + PC9
##
##              Df Deviance      AIC
## - Drug.Add      1  33.782 -180.35
## - Group          1  33.782 -180.35
## - Education      2  33.875 -180.24
## - Hypertension   1  33.794 -180.08
## - PC9            1  33.794 -180.08
## - PC6            1  33.813 -179.65
## - PC4            1  33.817 -179.56
## - PC2            1  33.825 -179.38
## - PC1            1  33.830 -179.27
## - Prev.preg      1  33.834 -179.17
## - Use.Tob        1  33.834 -179.16
## - PC5            1  33.858 -178.64
## - PC7            1  33.867 -178.42
## <none>           33.780 -178.40
## - PC8            1  33.891 -177.88
## - Diabetes       1  33.963 -176.25
## - PC3            1  34.023 -174.89
## - Preg.ended...37.wk 1  66.343  337.30
##
## Step: AIC=-180.35
## low.weight ~ Preg.ended...37.wk + Education + Hypertension +
## Diabetes + Use.Tob + Prev.preg + Group + PC1 + PC2 + PC3 +
## PC4 + PC5 + PC6 + PC7 + PC8 + PC9
##
##              Df Deviance      AIC
## - Group          1  33.784 -182.30
## - Education      2  33.877 -182.19
## - Hypertension   1  33.795 -182.04
## - PC9            1  33.797 -182.02
## - PC6            1  33.814 -181.62

```

```

## - PC4          1  33.820 -181.49
## - PC2          1  33.827 -181.33
## - PC1          1  33.831 -181.23
## - Prev.preg    1  33.836 -181.13
## - Use.Tob      1  33.840 -181.04
## - PC5          1  33.859 -180.60
## <none>         33.782 -180.35
## - PC7          1  33.871 -180.34
## - PC8          1  33.893 -179.83
## - Diabetes     1  33.966 -178.19
## - PC3          1  34.027 -176.82
## - Preg.ended...37.wk 1  66.395  335.91
##
## Step:  AIC=-182.3
## low.weight ~ Preg.ended...37.wk + Education + Hypertension +
##   Diabetes + Use.Tob + Prev.preg + PC1 + PC2 + PC3 + PC4 +
##   PC5 + PC6 + PC7 + PC8 + PC9
##
##           Df Deviance      AIC
## - Education      2  33.880 -184.13
## - Hypertension    1  33.798 -183.98
## - PC9             1  33.800 -183.95
## - PC6             1  33.816 -183.59
## - PC4             1  33.823 -183.43
## - PC2             1  33.829 -183.29
## - PC1             1  33.834 -183.18
## - Prev.preg       1  33.838 -183.08
## - Use.Tob         1  33.843 -182.97
## - PC5             1  33.861 -182.56
## - PC7             1  33.872 -182.31
## <none>            33.784 -182.30
## - PC8             1  33.896 -181.76
## - Diabetes        1  33.966 -180.18
## - PC3             1  34.029 -178.77
## - Preg.ended...37.wk 1  66.429  334.29
##
## Step:  AIC=-184.13
## low.weight ~ Preg.ended...37.wk + Hypertension + Diabetes + Use.Tob +
##   Prev.preg + PC1 + PC2 + PC3 + PC4 + PC5 + PC6 + PC7 + PC8 +
##   PC9
##
##           Df Deviance      AIC
## - Hypertension    1  33.889 -185.93
## - PC9             1  33.900 -185.68
## - PC4             1  33.908 -185.49
## - PC6             1  33.914 -185.37
## - Prev.preg       1  33.924 -185.14
## - PC2             1  33.929 -185.03
## - Use.Tob         1  33.931 -184.97
## - PC1             1  33.937 -184.84
## - PC5             1  33.954 -184.45
## <none>            33.880 -184.13
## - PC7             1  33.972 -184.05
## - PC8             1  33.993 -183.58

```

```

## - Diabetes          1  34.072 -181.81
## - PC3               1  34.110 -180.94
## - Preg.ended...37.wk 1  66.433  330.35
##
## Step: AIC=-185.93
## low.weight ~ Preg.ended...37.wk + Diabetes + Use.Tob + Prev.preg +
##      PC1 + PC2 + PC3 + PC4 + PC5 + PC6 + PC7 + PC8 + PC9
##
##              Df Deviance      AIC
## - PC9          1  33.908 -187.50
## - PC4          1  33.918 -187.28
## - PC6          1  33.924 -187.14
## - Prev.preg    1  33.933 -186.94
## - PC2          1  33.937 -186.85
## - Use.Tob      1  33.942 -186.74
## - PC1          1  33.945 -186.66
## - PC5          1  33.961 -186.29
## <none>         33.889 -185.93
## - PC7          1  33.979 -185.90
## - PC8          1  34.003 -185.34
## - Diabetes     1  34.075 -183.74
## - PC3          1  34.124 -182.62
## - Preg.ended...37.wk 1  67.659  342.38
##
## Step: AIC=-187.5
## low.weight ~ Preg.ended...37.wk + Diabetes + Use.Tob + Prev.preg +
##      PC1 + PC2 + PC3 + PC4 + PC5 + PC6 + PC7 + PC8
##
##              Df Deviance      AIC
## - PC4          1  33.937 -188.85
## - PC6          1  33.943 -188.71
## - Prev.preg    1  33.952 -188.50
## - PC2          1  33.955 -188.42
## - Use.Tob      1  33.959 -188.35
## - PC1          1  33.964 -188.23
## - PC5          1  33.980 -187.86
## <none>         33.908 -187.50
## - PC7          1  33.998 -187.47
## - PC8          1  34.022 -186.92
## - Diabetes     1  34.093 -185.32
## - PC3          1  34.143 -184.19
## - Preg.ended...37.wk 1  67.725  341.12
##
## Step: AIC=-188.85
## low.weight ~ Preg.ended...37.wk + Diabetes + Use.Tob + Prev.preg +
##      PC1 + PC2 + PC3 + PC5 + PC6 + PC7 + PC8
##
##              Df Deviance      AIC
## - PC6          1  33.971 -190.06
## - Prev.preg    1  33.982 -189.82
## - PC2          1  33.985 -189.74
## - PC1          1  33.993 -189.57
## - Use.Tob      1  33.999 -189.44
## - PC5          1  34.009 -189.22

```

```

## <none>                33.937 -188.85
## - PC7                  1  34.027 -188.82
## - PC8                  1  34.052 -188.24
## - Diabetes             1  34.114 -186.84
## - PC3                  1  34.171 -185.58
## - Preg.ended...37.wk  1  67.767  339.59
##
## Step:  AIC=-190.06
## low.weight ~ Preg.ended...37.wk + Diabetes + Use.Tob + Prev.preg +
##      PC1 + PC2 + PC3 + PC5 + PC7 + PC8
##
##              Df Deviance      AIC
## - Prev.preg      1  34.010 -191.20
## - PC2            1  34.022 -190.93
## - PC1            1  34.029 -190.77
## - Use.Tob        1  34.042 -190.46
## - PC5            1  34.043 -190.46
## <none>            33.971 -190.06
## - PC7            1  34.061 -190.03
## - PC8            1  34.087 -189.45
## - Diabetes       1  34.147 -188.11
## - PC3            1  34.204 -186.82
## - Preg.ended...37.wk  1  67.976  339.96
##
## Step:  AIC=-191.2
## low.weight ~ Preg.ended...37.wk + Diabetes + Use.Tob + PC1 +
##      PC2 + PC3 + PC5 + PC7 + PC8
##
##              Df Deviance      AIC
## - PC2            1  34.059 -192.08
## - PC1            1  34.067 -191.90
## - PC5            1  34.071 -191.82
## - Use.Tob        1  34.075 -191.72
## <none>            34.010 -191.20
## - PC7            1  34.102 -191.11
## - PC8            1  34.126 -190.59
## - Diabetes       1  34.189 -189.16
## - PC3            1  34.247 -187.86
## - Preg.ended...37.wk  1  68.019  338.45
##
## Step:  AIC=-192.08
## low.weight ~ Preg.ended...37.wk + Diabetes + Use.Tob + PC1 +
##      PC3 + PC5 + PC7 + PC8
##
##              Df Deviance      AIC
## - Use.Tob        1  34.109 -192.95
## - PC1            1  34.115 -192.82
## - PC5            1  34.121 -192.69
## <none>            34.059 -192.08
## - PC7            1  34.152 -191.99
## - PC8            1  34.174 -191.49
## - Diabetes       1  34.237 -190.08
## - PC3            1  34.300 -188.68
## - Preg.ended...37.wk  1  68.039  336.67

```



```
##
## Step: AIC=-192.95
## low.weight ~ Preg.ended...37.wk + Diabetes + PC1 + PC3 + PC5 +
## PC7 + PC8
##
##           Df Deviance      AIC
## - PC1      1  34.158 -193.86
## - PC5      1  34.175 -193.49
## <none>      34.109 -192.95
## - PC7      1  34.202 -192.86
## - PC8      1  34.215 -192.58
## - Diabetes  1  34.291 -190.89
## - PC3      1  34.373 -189.06
## - Preg.ended...37.wk 1  68.321  337.84
##
## Step: AIC=-193.86
## low.weight ~ Preg.ended...37.wk + Diabetes + PC3 + PC5 + PC7 +
## PC8
##
##           Df Deviance      AIC
## - PC5      1  34.223 -194.41
## <none>      34.158 -193.86
## - PC7      1  34.251 -193.78
## - PC8      1  34.264 -193.49
## - Diabetes  1  34.335 -191.89
## - PC3      1  34.421 -189.98
## - Preg.ended...37.wk 1  68.365  336.34
##
## Step: AIC=-194.41
## low.weight ~ Preg.ended...37.wk + Diabetes + PC3 + PC7 + PC8
##
##           Df Deviance      AIC
## <none>      34.223 -194.41
## - PC7      1  34.316 -194.33
## - PC8      1  34.329 -194.03
## - Diabetes  1  34.404 -192.36
## - PC3      1  34.486 -190.54
## - Preg.ended...37.wk 1  68.365  334.34
```

```
summary(fit.logis.step.extend)
```

```
##
## Call:
## glm(formula = low.weight ~ Preg.ended...37.wk + Diabetes + PC3 +
## PC7 + PC8, data = complete_birthweight_extended[, -1])
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.024562   0.008228   2.985  0.00293 **
## Preg.ended...37.wkYes 0.654334   0.023747  27.554 < 2e-16 ***
## DiabetesYes     -0.092613   0.046154  -2.007  0.04514 *
## PC3             -0.017214   0.007118  -2.418  0.01582 *
## PC7             -0.011349   0.007890  -1.438  0.15076
## PC8              0.012229   0.007945   1.539  0.12416
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.04497061)
##
##      Null deviance: 69.270  on 766  degrees of freedom
## Residual deviance: 34.223  on 761  degrees of freedom
## AIC: -194.41
##
## Number of Fisher Scoring iterations: 2

coef.log = round(cbind(exp(summary(fit.logis.step.extend)$coefficients[,1]),drop1(fit.logis.step.extend

## Waiting for profiling to be done...

colnames(coef.log) = c("Odds ratio", "P value", "2.5%", "97.5%")
knitr::kable(coef.log)
```

	Odds ratio	P value	2.5%	97.5%
(Intercept)	1.025	NA	1.008	1.042
Preg.ended...37.wkYes	1.924	0.000	1.836	2.016
DiabetesYes	0.912	0.044	0.833	0.998
PC3	0.983	0.015	0.969	0.997
PC7	0.989	0.149	0.974	1.004
PC8	1.012	0.123	0.997	1.028

Sensitivity analysis

```
# Check missing value for the initial data
# View(data.frame(apply(opt_copy, 2, function(x){sum(is.na(x))})))

# Multiple imputation
library(mice)

opt_copy <- opt_copy %>% mutate_if(is.character, as.factor)

my_imp = mice(opt_copy, m=5, maxit = 10, seed = 2024)
```

```
##
## iter imp variable
## 1 1 Birthweight Preg.ended...37.wk Use.Tob Use.Alc Drug.Add OAA1 OCR1 OFN1 OPG1 OPI1
## 1 2 Birthweight Preg.ended...37.wk Use.Tob Use.Alc Drug.Add OAA1 OCR1 OFN1 OPG1 OPI1
## 1 3 Birthweight Preg.ended...37.wk Use.Tob Use.Alc Drug.Add OAA1 OCR1 OFN1 OPG1 OPI1
## 1 4 Birthweight Preg.ended...37.wk Use.Tob Use.Alc Drug.Add OAA1 OCR1 OFN1 OPG1 OPI1
## 1 5 Birthweight Preg.ended...37.wk Use.Tob Use.Alc Drug.Add OAA1 OCR1 OFN1 OPG1 OPI1
## 2 1 Birthweight Preg.ended...37.wk Use.Tob Use.Alc Drug.Add OAA1 OCR1 OFN1 OPG1 OPI1
## 2 2 Birthweight Preg.ended...37.wk Use.Tob Use.Alc Drug.Add OAA1 OCR1 OFN1 OPG1 OPI1
## 2 3 Birthweight Preg.ended...37.wk Use.Tob Use.Alc Drug.Add OAA1 OCR1 OFN1 OPG1 OPI1
```

##	2	4	Birthweight	Preg.ended...37.wk	Use.Tob	Use.Alc	Drug.Add	OAA1	OCR1	OFN1	OPG1	OPI1
##	2	5	Birthweight	Preg.ended...37.wk	Use.Tob	Use.Alc	Drug.Add	OAA1	OCR1	OFN1	OPG1	OPI1
##	3	1	Birthweight	Preg.ended...37.wk	Use.Tob	Use.Alc	Drug.Add	OAA1	OCR1	OFN1	OPG1	OPI1
##	3	2	Birthweight	Preg.ended...37.wk	Use.Tob	Use.Alc	Drug.Add	OAA1	OCR1	OFN1	OPG1	OPI1
##	3	3	Birthweight	Preg.ended...37.wk	Use.Tob	Use.Alc	Drug.Add	OAA1	OCR1	OFN1	OPG1	OPI1
##	3	4	Birthweight	Preg.ended...37.wk	Use.Tob	Use.Alc	Drug.Add	OAA1	OCR1	OFN1	OPG1	OPI1
##	3	5	Birthweight	Preg.ended...37.wk	Use.Tob	Use.Alc	Drug.Add	OAA1	OCR1	OFN1	OPG1	OPI1
##	4	1	Birthweight	Preg.ended...37.wk	Use.Tob	Use.Alc	Drug.Add	OAA1	OCR1	OFN1	OPG1	OPI1
##	4	2	Birthweight	Preg.ended...37.wk	Use.Tob	Use.Alc	Drug.Add	OAA1	OCR1	OFN1	OPG1	OPI1
##	4	3	Birthweight	Preg.ended...37.wk	Use.Tob	Use.Alc	Drug.Add	OAA1	OCR1	OFN1	OPG1	OPI1
##	4	4	Birthweight	Preg.ended...37.wk	Use.Tob	Use.Alc	Drug.Add	OAA1	OCR1	OFN1	OPG1	OPI1
##	4	5	Birthweight	Preg.ended...37.wk	Use.Tob	Use.Alc	Drug.Add	OAA1	OCR1	OFN1	OPG1	OPI1
##	5	1	Birthweight	Preg.ended...37.wk	Use.Tob	Use.Alc	Drug.Add	OAA1	OCR1	OFN1	OPG1	OPI1
##	5	2	Birthweight	Preg.ended...37.wk	Use.Tob	Use.Alc	Drug.Add	OAA1	OCR1	OFN1	OPG1	OPI1
##	5	3	Birthweight	Preg.ended...37.wk	Use.Tob	Use.Alc	Drug.Add	OAA1	OCR1	OFN1	OPG1	OPI1
##	5	4	Birthweight	Preg.ended...37.wk	Use.Tob	Use.Alc	Drug.Add	OAA1	OCR1	OFN1	OPG1	OPI1
##	5	5	Birthweight	Preg.ended...37.wk	Use.Tob	Use.Alc	Drug.Add	OAA1	OCR1	OFN1	OPG1	OPI1
##	6	1	Birthweight	Preg.ended...37.wk	Use.Tob	Use.Alc	Drug.Add	OAA1	OCR1	OFN1	OPG1	OPI1
##	6	2	Birthweight	Preg.ended...37.wk	Use.Tob	Use.Alc	Drug.Add	OAA1	OCR1	OFN1	OPG1	OPI1
##	6	3	Birthweight	Preg.ended...37.wk	Use.Tob	Use.Alc	Drug.Add	OAA1	OCR1	OFN1	OPG1	OPI1
##	6	4	Birthweight	Preg.ended...37.wk	Use.Tob	Use.Alc	Drug.Add	OAA1	OCR1	OFN1	OPG1	OPI1
##	6	5	Birthweight	Preg.ended...37.wk	Use.Tob	Use.Alc	Drug.Add	OAA1	OCR1	OFN1	OPG1	OPI1
##	7	1	Birthweight	Preg.ended...37.wk	Use.Tob	Use.Alc	Drug.Add	OAA1	OCR1	OFN1	OPG1	OPI1
##	7	2	Birthweight	Preg.ended...37.wk	Use.Tob	Use.Alc	Drug.Add	OAA1	OCR1	OFN1	OPG1	OPI1
##	7	3	Birthweight	Preg.ended...37.wk	Use.Tob	Use.Alc	Drug.Add	OAA1	OCR1	OFN1	OPG1	OPI1
##	7	4	Birthweight	Preg.ended...37.wk	Use.Tob	Use.Alc	Drug.Add	OAA1	OCR1	OFN1	OPG1	OPI1
##	7	5	Birthweight	Preg.ended...37.wk	Use.Tob	Use.Alc	Drug.Add	OAA1	OCR1	OFN1	OPG1	OPI1
##	8	1	Birthweight	Preg.ended...37.wk	Use.Tob	Use.Alc	Drug.Add	OAA1	OCR1	OFN1	OPG1	OPI1
##	8	2	Birthweight	Preg.ended...37.wk	Use.Tob	Use.Alc	Drug.Add	OAA1	OCR1	OFN1	OPG1	OPI1
##	8	3	Birthweight	Preg.ended...37.wk	Use.Tob	Use.Alc	Drug.Add	OAA1	OCR1	OFN1	OPG1	OPI1
##	8	4	Birthweight	Preg.ended...37.wk	Use.Tob	Use.Alc	Drug.Add	OAA1	OCR1	OFN1	OPG1	OPI1
##	8	5	Birthweight	Preg.ended...37.wk	Use.Tob	Use.Alc	Drug.Add	OAA1	OCR1	OFN1	OPG1	OPI1
##	9	1	Birthweight	Preg.ended...37.wk	Use.Tob	Use.Alc	Drug.Add	OAA1	OCR1	OFN1	OPG1	OPI1
##	9	2	Birthweight	Preg.ended...37.wk	Use.Tob	Use.Alc	Drug.Add	OAA1	OCR1	OFN1	OPG1	OPI1
##	9	3	Birthweight	Preg.ended...37.wk	Use.Tob	Use.Alc	Drug.Add	OAA1	OCR1	OFN1	OPG1	OPI1
##	9	4	Birthweight	Preg.ended...37.wk	Use.Tob	Use.Alc	Drug.Add	OAA1	OCR1	OFN1	OPG1	OPI1
##	9	5	Birthweight	Preg.ended...37.wk	Use.Tob	Use.Alc	Drug.Add	OAA1	OCR1	OFN1	OPG1	OPI1
##	10	1	Birthweight	Preg.ended...37.wk	Use.Tob	Use.Alc	Drug.Add	OAA1	OCR1	OFN1	OPG1	OPI1
##	10	2	Birthweight	Preg.ended...37.wk	Use.Tob	Use.Alc	Drug.Add	OAA1	OCR1	OFN1	OPG1	OPI1
##	10	3	Birthweight	Preg.ended...37.wk	Use.Tob	Use.Alc	Drug.Add	OAA1	OCR1	OFN1	OPG1	OPI1
##	10	4	Birthweight	Preg.ended...37.wk	Use.Tob	Use.Alc	Drug.Add	OAA1	OCR1	OFN1	OPG1	OPI1
##	10	5	Birthweight	Preg.ended...37.wk	Use.Tob	Use.Alc	Drug.Add	OAA1	OCR1	OFN1	OPG1	OPI1

```
opt_imp = complete(my_imp, 1)
```

```
# Check missing value for imputed data
```

```
# View(data.frame(apply(opt_imp, 2, function(x){sum(is.na(x))})))
```

```
itt_preterm <- opt_imp %>% select(-Birthweight)
```

```
itt_birthweight <- opt_imp %>% select(-GA.at.outcome)
```

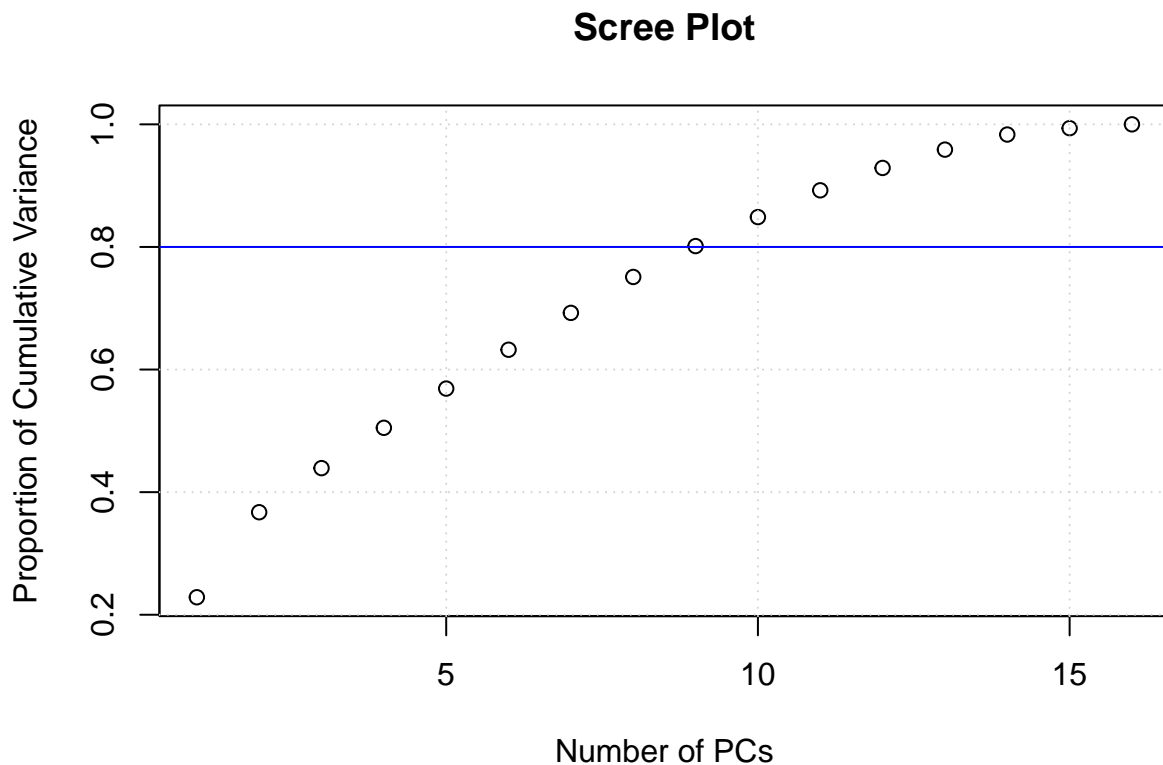
PCA

```
pca_itt_preterm <- prcomp(itt_preterm[,serum], scale = TRUE)
summary(pca_itt_preterm)
```

```
## Importance of components:
##              PC1      PC2      PC3      PC4      PC5      PC6      PC7
## Standard deviation  1.9124 1.4896 1.07306 1.02557 1.01182 1.00730 0.98054
## Proportion of Variance 0.2286 0.1387 0.07197 0.06574 0.06399 0.06342 0.06009
## Cumulative Proportion 0.2286 0.3673 0.43921 0.50495 0.56894 0.63235 0.69244
##              PC8      PC9      PC10     PC11     PC12     PC13     PC14
## Standard deviation  0.96848 0.89761 0.86969 0.83625 0.76396 0.69184 0.62756
## Proportion of Variance 0.05862 0.05036 0.04727 0.04371 0.03648 0.02991 0.02461
## Cumulative Proportion 0.75107 0.80142 0.84870 0.89240 0.92888 0.95879 0.98341
##              PC15     PC16
## Standard deviation  0.40650 0.31656
## Proportion of Variance 0.01033 0.00626
## Cumulative Proportion 0.99374 1.00000
```

```
cumvars <- cumsum(pca_itt_preterm$sdev^2 / sum(pca_itt_preterm$sdev^2))
```

```
plot(cumvars, xlab = 'Number of PCs', ylab = 'Proportion of Cumulative Variance', main = 'Scree Plot')
grid()
abline(h = 0.8, col = 'blue')
```



```

rotated_dat <- as.data.frame(pca_itt_preterm$x)
# 9 PCs instead of 16

itt_preterm_extended <- cbind(itt_preterm[, -c(12:27)], rotated_dat[, 1:9])
itt_preterm_extended <- itt_preterm_extended[complete.cases(itt_preterm_extended),]

## For low birth weight
pca_itt_birthweight <- prcomp(itt_birthweight[, serum], scale = TRUE)
summary(pca_itt_birthweight)

```

```

## Importance of components:
##
##          PC1      PC2      PC3      PC4      PC5      PC6      PC7
## Standard deviation  1.9124 1.4896 1.07306 1.02557 1.01182 1.00730 0.98054
## Proportion of Variance 0.2286 0.1387 0.07197 0.06574 0.06399 0.06342 0.06009
## Cumulative Proportion 0.2286 0.3673 0.43921 0.50495 0.56894 0.63235 0.69244
##
##          PC8      PC9      PC10     PC11     PC12     PC13     PC14
## Standard deviation  0.96848 0.89761 0.86969 0.83625 0.76396 0.69184 0.62756
## Proportion of Variance 0.05862 0.05036 0.04727 0.04371 0.03648 0.02991 0.02461
## Cumulative Proportion 0.75107 0.80142 0.84870 0.89240 0.92888 0.95879 0.98341
##
##          PC15     PC16
## Standard deviation  0.40650 0.31656
## Proportion of Variance 0.01033 0.00626
## Cumulative Proportion 0.99374 1.00000

```

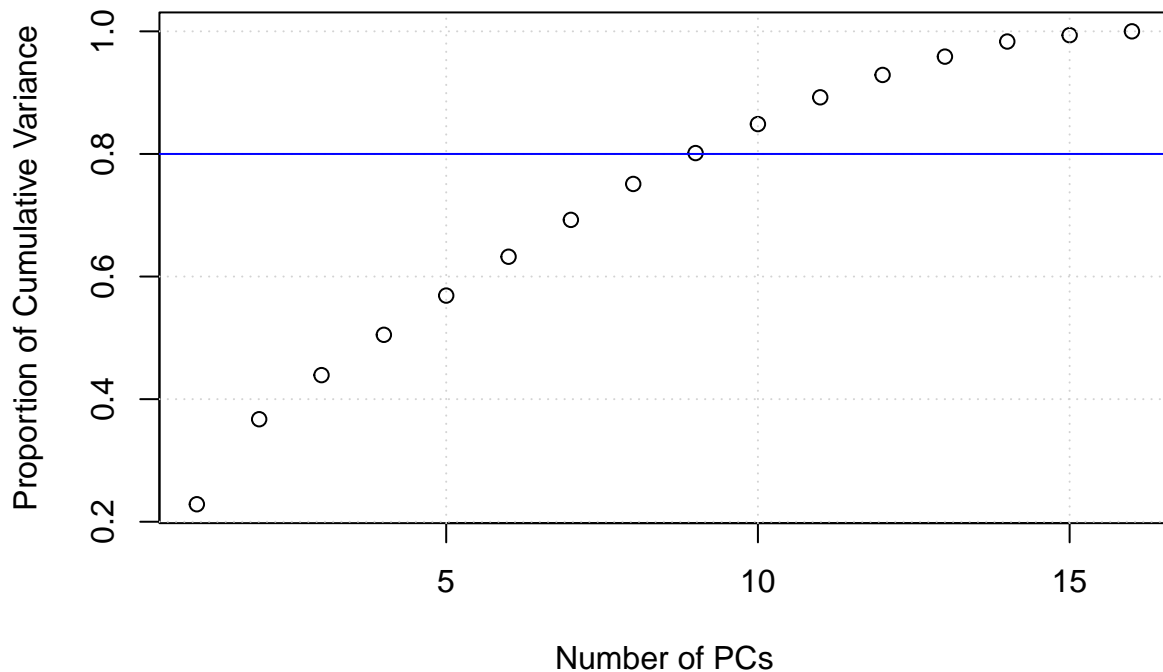
```

cumvars2 <- cumsum(pca_itt_birthweight$sdev^2 / sum(pca_itt_birthweight$sdev^2))

plot(cumvars2, xlab = 'Number of PCs', ylab = 'Proportion of Cumulative Variance', main = 'Scree Plot')
grid()
abline(h = 0.8, col = 'blue')

```

Scree Plot



```
# 9 PCs instead of 16
rotated_dat <- as.data.frame(pca_itt_birthweight$x)

itt_birthweight_extended <- cbind(itt_birthweight[, -c(12:27)], rotated_dat[, 1:9])
itt_birthweight_extended <- itt_birthweight_extended[complete.cases(itt_birthweight_extended),]
```

Logrank and plots

```
itt_preterm_extended$Preg.ended...37.wk <- ifelse(itt_preterm_extended$Preg.ended...37.wk == 'Yes', 1, 0)

surv_obj <- with(itt_preterm_extended, Surv(time = GA.at.outcome, event = Preg.ended...37.wk, type = 'r'))

fit.ext <- survfit(surv_obj ~ Group, data = itt_preterm_extended)

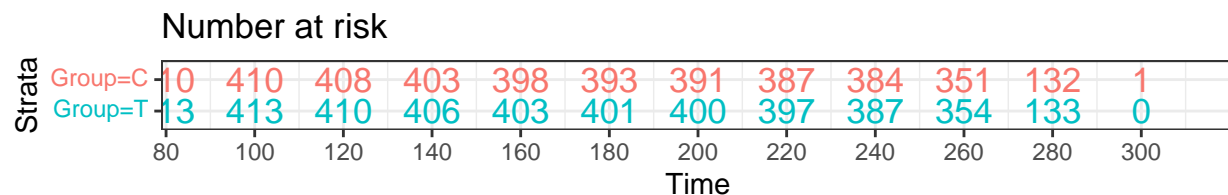
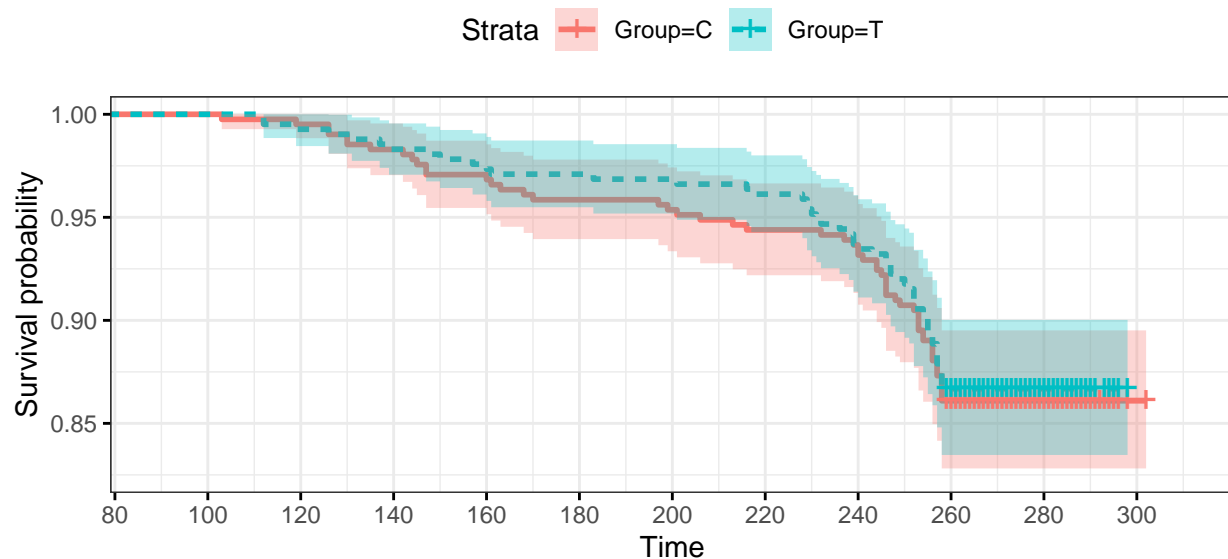
ggs <- ggsurvplot(fit.ext,
  pval = TRUE, conf.int = TRUE,
  risk.table = TRUE, # Add risk table
  risk.table.col = "strata", # Change risk table color by groups
  linetype = "strata", # Change line type by groups
  ggtheme = theme_bw(), # Change ggplot2 theme
  palette = c("#F8766D", "#00BFC4"),
  ylim=c(0.825,1),
  xlim=c(90,310),
  title = 'Survival Curves for Control and Treatment Cohort',
```

```
break.x.by = 20)

#ggs$plot <- ggs$plot + ylim(c(0.75,1)) + xlim(c(90, 310))
ggs
```

```
## Warning: Removed 1 row containing missing values or values outside the scale range
## (`geom_text()`).
## Removed 1 row containing missing values or values outside the scale range
## (`geom_text()`).
```

Survival Curves for Control and Treatment Cohort



```
surv_diff1_ext <- survdiff(Surv(time = GA.at.outcome, event = Preg.ended...37.wk, type = 'right') ~ Group)
surv_diff1_ext
```

```
## Call:
## survdiff(formula = Surv(time = GA.at.outcome, event = Preg.ended...37.wk,
##   type = "right") ~ Group, data = itt_preterm_extended)
##
##           N Observed Expected (O-E)^2/E (O-E)^2/V
## Group=C 410         57      55.6    0.0358    0.0714
## Group=T 413         55      56.4    0.0353    0.0714
##
## Chisq= 0.1  on 1 degrees of freedom, p= 0.8
```

Only checking the data pre-32.8 weeks aka 231 days or fewer

```
pre32_ext <- itt_preterm_extended

pre32_ext$Preg.ended...37.wk <- NULL
pre32_ext$pre230 <- ifelse(pre32_ext$GA.at.outcome <= 230, 1, 0)

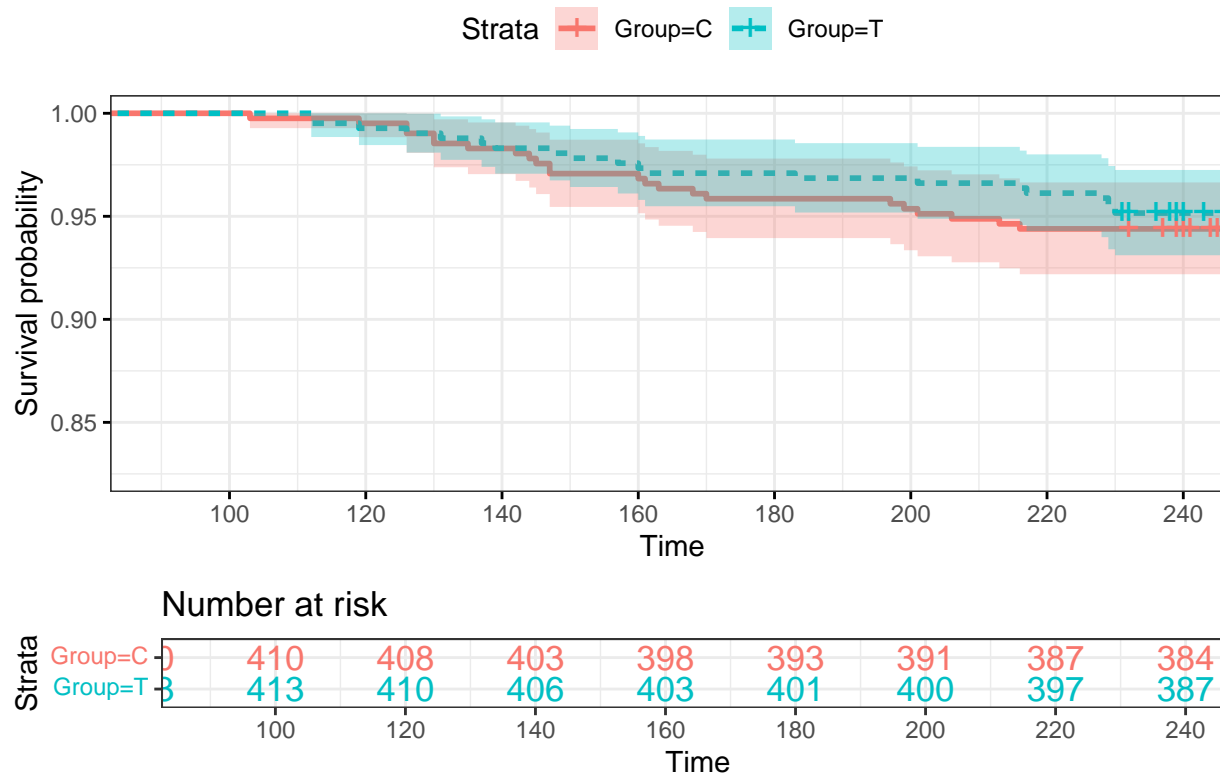
# new survival fit
surv_obj2 <- with(pre32_ext, Surv(time = GA.at.outcome, event = pre230, type = 'right'))
fit2_ext <- survfit(surv_obj2 ~ Group, data = pre32_ext)

ggs2 <- ggsurvplot(fit2_ext,
  pval = TRUE, conf.int = TRUE,
  risk.table = TRUE, # Add risk table
  risk.table.col = "strata", # Change risk table color by groups
  linetype = "strata", # Change line type by groups
  ggtheme = theme_bw(), # Change ggplot2 theme
  palette = c("#F8766D", "#00BFC4"),
  ylim=c(0.825,1),
  xlim=c(90,240),
  title = 'Survival Curves for Control and Treatment Cohort',
  break.x.by = 20)

#ggs$plot <- ggs$plot + ylim(c(0.75,1)) + xlim(c(90, 310))
ggs2
```

```
## Warning: Removed 1 row containing missing values or values outside the scale range
## (`geom_text()`).
## Removed 1 row containing missing values or values outside the scale range
## (`geom_text()`).
```


Survival Curves for Control and Treatment Cohort



```
surv_diff2_ext <- survdiff(Surv(time = GA.at.outcome, event = pre230, type = 'right') ~ Group, data = p
surv_diff2_ext
```

```
## Call:
## survdiff(formula = Surv(time = GA.at.outcome, event = pre230,
##   type = "right") ~ Group, data = pre32_ext)
##
##           N Observed Expected (O-E)^2/E (O-E)^2/V
## Group=C 410      23     21.3    0.129    0.256
## Group=T 413      20     21.7    0.127    0.256
##
## Chisq= 0.3  on 1 degrees of freedom, p= 0.6
```

Cox Proportional Hazards for preterm birth

```
fit.coxph.extend <- coxph(Surv(time = GA.at.outcome, event = Preg.ended...37.wk) ~ ., data=itt_preterm_ex
summary(fit.coxph.extend)
```

```
## Call:
## coxph(formula = Surv(time = GA.at.outcome, event = Preg.ended...37.wk) ~
##   ., data = itt_preterm_extended)
##
## n= 823, number of events= 112
```

```
##
##               coef exp(coef) se(coef)      z Pr(>|z|)
## Age           0.01841   1.01858  0.01871  0.984  0.32513
## EducationLT 8 yrs -0.36914   0.69133  0.30120 -1.226  0.22036
## EducationMT 12 yrs  0.07962   1.08288  0.22644  0.352  0.72512
## HypertensionYes    1.07842   2.94004  0.34672  3.110  0.00187 **
## DiabetesYes        1.03877   2.82574  0.37098  2.800  0.00511 **
## Use.TobYes         0.34344   1.40978  0.27600  1.244  0.21337
## Use.AlcYes         0.06295   1.06498  0.58704  0.107  0.91460
## Drug.AddYes        1.43148   4.18489  0.68710  2.083  0.03722 *
## Prev.pregYes       0.11772   1.12493  0.25005  0.471  0.63779
## GroupT            -0.22037   0.80223  0.19719 -1.118  0.26376
## PC1               -0.10924   0.89651  0.06080 -1.797  0.07238 .
## PC2                0.14519   1.15626  0.06566  2.211  0.02702 *
## PC3               -0.05390   0.94752  0.08988 -0.600  0.54869
## PC4               -0.16890   0.84459  0.09579 -1.763  0.07784 .
## PC5                0.04696   1.04808  0.09861  0.476  0.63394
## PC6                0.13172   1.14078  0.09691  1.359  0.17409
## PC7                0.08678   1.09066  0.09926  0.874  0.38197
## PC8               -0.01101   0.98905  0.10065 -0.109  0.91289
## PC9                0.07993   1.08321  0.11050  0.723  0.46948
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##               exp(coef) exp(-coef) lower .95 upper .95
## Age           1.0186    0.9818    0.9819    1.057
## EducationLT 8 yrs  0.6913    1.4465    0.3831    1.248
## EducationMT 12 yrs  1.0829    0.9235    0.6948    1.688
## HypertensionYes    2.9400    0.3401    1.4901    5.801
## DiabetesYes        2.8257    0.3539    1.3657    5.847
## Use.TobYes         1.4098    0.7093    0.8208    2.421
## Use.AlcYes         1.0650    0.9390    0.3370    3.365
## Drug.AddYes        4.1849    0.2390    1.0885   16.090
## Prev.pregYes       1.1249    0.8889    0.6891    1.836
## GroupT            0.8022    1.2465    0.5451    1.181
## PC1               0.8965    1.1154    0.7958    1.010
## PC2               1.1563    0.8649    1.0166    1.315
## PC3               0.9475    1.0554    0.7945    1.130
## PC4               0.8446    1.1840    0.7000    1.019
## PC5               1.0481    0.9541    0.8639    1.272
## PC6               1.1408    0.8766    0.9434    1.379
## PC7               1.0907    0.9169    0.8978    1.325
## PC8               0.9890    1.0111    0.8120    1.205
## PC9               1.0832    0.9232    0.8723    1.345
##
## Concordance= 0.659 (se = 0.026 )
## Likelihood ratio test= 47.62 on 19 df,  p=3e-04
## Wald test              = 57.01 on 19 df,  p=1e-05
## Score (logrank) test = 64.82 on 19 df,  p=7e-07
```

```
coef = round(cbind(summary(fit.coxph.extend)$coefficients[, c(2,5)], summary(fit.coxph.extend)$conf.int
colnames(coef) = c("Hazard", "P value", "2.5%", "97.5%")
knitr::kable(coef)
```

	Hazard	P value	2.5%	97.5%
Age	1.019	0.325	0.982	1.057
EducationLT 8 yrs	0.691	0.220	0.383	1.248
EducationMT 12 yrs	1.083	0.725	0.695	1.688
HypertensionYes	2.940	0.002	1.490	5.801
DiabetesYes	2.826	0.005	1.366	5.847
Use.TobYes	1.410	0.213	0.821	2.421
Use.AlcYes	1.065	0.915	0.337	3.365
Drug.AddYes	4.185	0.037	1.088	16.090
Prev.pregYes	1.125	0.638	0.689	1.836
GroupT	0.802	0.264	0.545	1.181
PC1	0.897	0.072	0.796	1.010
PC2	1.156	0.027	1.017	1.315
PC3	0.948	0.549	0.794	1.130
PC4	0.845	0.078	0.700	1.019
PC5	1.048	0.634	0.864	1.272
PC6	1.141	0.174	0.943	1.379
PC7	1.091	0.382	0.898	1.325
PC8	0.989	0.913	0.812	1.205
PC9	1.083	0.469	0.872	1.345

```
fit.coxph2.extend <- coxph(Surv(time = GA.at.outcome, event = pre230)~., data=pre32_ext)
summary(fit.coxph2.extend)
```

```
## Call:
## coxph(formula = Surv(time = GA.at.outcome, event = pre230) ~
##      ., data = pre32_ext)
##
##      n= 823, number of events= 43
##
##              coef exp(coef)  se(coef)      z Pr(>|z|)
## Age           -0.037285  0.963402  0.031960 -1.167  0.2434
## EducationLT 8 yrs -0.587947  0.555467  0.547835 -1.073  0.2832
## EducationMT 12 yrs  0.221135  1.247492  0.367136  0.602  0.5470
## HypertensionYes    1.002394  2.724797  0.641627  1.562  0.1182
## DiabetesYes        1.100047  3.004308  0.633945  1.735  0.0827
## Use.TobYes         0.237454  1.268017  0.456846  0.520  0.6032
## Use.AlcYes         0.135588  1.145209  0.794965  0.171  0.8646
## Drug.AddYes        2.058113  7.831180  0.848857  2.425  0.0153 *
## Prev.pregYes       0.266132  1.304907  0.391149  0.680  0.4963
## GroupT            -0.397387  0.672074  0.326489 -1.217  0.2235
## PC1               -0.190946  0.826177  0.112740 -1.694  0.0903
## PC2               -0.042026  0.958845  0.107677 -0.390  0.6963
## PC3                0.005892  1.005909  0.142849  0.041  0.9671
## PC4                0.059573  1.061383  0.150849  0.395  0.6929
## PC5                0.190863  1.210293  0.155943  1.224  0.2210
## PC6                0.034481  1.035082  0.158552  0.217  0.8278
## PC7               -0.083952  0.919475  0.160435 -0.523  0.6008
## PC8                0.162129  1.176012  0.164044  0.988  0.3230
## PC9                0.237334  1.267865  0.176273  1.346  0.1782
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
##               exp(coef) exp(-coef) lower .95 upper .95
## Age           0.9634    1.0380    0.9049    1.026
## EducationLT 8 yrs 0.5555    1.8003    0.1898    1.625
## EducationMT 12 yrs 1.2475    0.8016    0.6075    2.562
## HypertensionYes  2.7248    0.3670    0.7748    9.583
## DiabetesYes      3.0043    0.3329    0.8672   10.408
## Use.TobYes       1.2680    0.7886    0.5179    3.105
## Use.AlcYes       1.1452    0.8732    0.2411    5.440
## Drug.AddYes      7.8312    0.1277    1.4835   41.340
## Prev.pregYes     1.3049    0.7663    0.6062    2.809
## GroupT           0.6721    1.4879    0.3544    1.274
## PC1              0.8262    1.2104    0.6624    1.030
## PC2              0.9588    1.0429    0.7764    1.184
## PC3              1.0059    0.9941    0.7603    1.331
## PC4              1.0614    0.9422    0.7897    1.427
## PC5              1.2103    0.8262    0.8916    1.643
## PC6              1.0351    0.9661    0.7586    1.412
## PC7              0.9195    1.0876    0.6714    1.259
## PC8              1.1760    0.8503    0.8527    1.622
## PC9              1.2679    0.7887    0.8975    1.791
##
## Concordance= 0.7 (se = 0.041 )
## Likelihood ratio test= 25.73 on 19 df,  p=0.1
## Wald test              = 30.2 on 19 df,  p=0.05
## Score (logrank) test = 42.36 on 19 df,  p=0.002
```

```
coef2 = round(cbind(summary(fit.coxph2.extend)$coefficients[, c(2,5)], summary(fit.coxph2.extend)$conf.
colnames(coef2) = c("Hazard", "P value", "2.5%", "97.5%")
knitr::kable(coef2)
```

	Hazard	P value	2.5%	97.5%
Age	0.963	0.243	0.905	1.026
EducationLT 8 yrs	0.555	0.283	0.190	1.625
EducationMT 12 yrs	1.247	0.547	0.607	2.562
HypertensionYes	2.725	0.118	0.775	9.583
DiabetesYes	3.004	0.083	0.867	10.408
Use.TobYes	1.268	0.603	0.518	3.105
Use.AlcYes	1.145	0.865	0.241	5.440
Drug.AddYes	7.831	0.015	1.483	41.340
Prev.pregYes	1.305	0.496	0.606	2.809
GroupT	0.672	0.224	0.354	1.274
PC1	0.826	0.090	0.662	1.030
PC2	0.959	0.696	0.776	1.184
PC3	1.006	0.967	0.760	1.331
PC4	1.061	0.693	0.790	1.427
PC5	1.210	0.221	0.892	1.643
PC6	1.035	0.828	0.759	1.412
PC7	0.919	0.601	0.671	1.259
PC8	1.176	0.323	0.853	1.622
PC9	1.268	0.178	0.897	1.791

Logistic regression for low birth weight

```
# Define indicator for low birth weight (birth weight less than 2500mg)
itt_birthweight_extended$low.weight = ifelse(itt_birthweight_extended$Birthweight <= 2500, 1, 0)

# fit.logis.extend <- glm(low.weight~., data=itt_birthweight_extended[,-1])
# summary(fit.logis.extend)
# fit.logis.step.extend <- step(fit.logis.extend, direction = "backward")
# summary(fit.logis.step.extend)

fit.logis.itt <- glm(low.weight~Preg.ended...37.wk+Diabetes+PC3+PC7+PC8, data=itt_birthweight_extended[
summary(fit.logis.itt)
```

```
##
## Call:
## glm(formula = low.weight ~ Preg.ended...37.wk + Diabetes + PC3 +
##      PC7 + PC8, data = itt_birthweight_extended[, -1])
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.027748   0.008375   3.313 0.000962 ***
## Preg.ended...37.wkYes 0.686698   0.022733  30.207 < 2e-16 ***
## DiabetesYes       -0.114456   0.046312  -2.471 0.013660 *
## PC3                0.015575   0.007213   2.159 0.031124 *
## PC7                0.006260   0.007896   0.793 0.428095
## PC8                0.009031   0.007986   1.131 0.258474
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.04917254)
##
##      Null deviance: 85.567  on 822  degrees of freedom
## Residual deviance: 40.174  on 817  degrees of freedom
## AIC: -135.67
##
## Number of Fisher Scoring iterations: 2
```

```
coef.log2 = round(cbind(exp(summary(fit.logis.itt)$coefficients[,1]),drop1(fit.logis.itt, test = "Chisq
```

```
## Waiting for profiling to be done...
```

```
colnames(coef.log2) = c("Odds ratio", "P value", "2.5%", "97.5%")
knitr::kable(coef.log2)
```

	Odds ratio	P value	2.5%	97.5%
(Intercept)	1.028	NA	1.011	1.045
Preg.ended...37.wkYes	1.987	0.000	1.901	2.078
DiabetesYes	0.892	0.013	0.814	0.977

	Odds ratio	P value	2.5%	97.5%
PC3	1.016	0.030	1.001	1.030
PC7	1.006	0.426	0.991	1.022
PC8	1.009	0.257	0.993	1.025