Survival Analysis Project: HIV Clinical Trial

Juste Simanauskaite & Patricia Rivera

Contents

Introduction

HIV (Human Immunodeficiency Virus) is a disease known as an immune system disorder, which causes severe destruction of white blood cells that are responsible for fighting infection. The presence of this disorder is a lead-in for a human to be more prone to infections and cancer diseases. AIDS is the final stage of HIV, which is not always developed in HIV patients. Zidovudine (AZT) is known as antiretroviral medication for prevention of HIV/AIDS, whereas lamuvidine (3TC) is an inhibitor medication that works in decreasing HIV and hepatitis B. Previously, it has been founded that three-drug combinations, in particular, with a previous exposure to AZT, have shown the most significant resulted in reducing HIV-1 RNA concentrations. Therefore, this study used indinavir sulfate (a synthetic antiviral agent that inhibits HIV protease activity) in combination with AZT and 3TC as well as variation of placebo treatments to determine the potency of triple drug therapy in the cases of advanced HIV-1 patients. The study hypothesized that a three-drug combination, including a HIV-protease inhibitor and two nucleoside analogues (AZT and 3TC) would alter the progression of the HIV-1 disease. The study was successful in reaching significant data of the clinical superiority of a three-drug approach with inidavor over a treatment containing only a two-drug combination.

The current analysis of the data from a study conducted by Hammer et al. in 1997 considers the response variable to be *time*, which here describes the amount of time in days for the time of death, AIDS diagnosis, or the termination of the study. Another important variable used for the analysis is *censor*, which indicates the paarticipants of the study that survived till the termination of the study without dying or being diagnosed AIDS. The study explored the influence of the explanatory varibale *tx.* referring to the treatment group that was differentiated into: a control (placebo group) and a treatment group that included IDV (indinavir) # Methods

The study was a randomized, double-blind, and a placebo-controlled trial that compared a three-drug treatment of indinavir (Crixivan), zidovudine (AZT) and lamivudine (3TC) with a two-drug treatment. Patients were selected based on the factor that they had no more than 200 CD4 cells per cubic millimetear at least 3 months prioir to AZT therapy. The patients had to be more than 16 years old, with a diagnostic documentation of HIV-1 infection, having no more than 1 week of prior lamuvidine treatment, and a Karnofsky score of at least 70.

The approved patients received 200mg of open-label zidovudine three times daily and 150mg of lamuvidine two times daily and were randomly assigned to a placebo or a treatment of 800mg of indinavir every eight hours.

Some modifications were made to the protocol. In October of 1996 prioir exposure to AZT was reduced to at least 3 months and permitted patients with no tolerance for this drug to enter the study with stavudine as a substitute.

Patients diagnosed with AIDS-defining events were offored an open-label assignment of the indinavir treatment with nor reveal of their initial treatment assignments. All of these cases had to be reviewed via a blind procedure by the study chair.

Follow ups were made at weeks 4,8, and 16 and every eight weeks afterwards. CD4 cell counts and Plasma HIV-1 RNA concentrations were measured twice at baseline and at weeks 4,8,24, and 40.

The statistical analysis methods used to interpret results were Kaplan-Meier estimates, log-rank tests, and proportional hazards models. The p-values, estimates of treatment differences and 95% confidence intervals were not adjusted for repeated analysis.

The data and results have been reviewed again in 2019 and have been analyzed via statistic methhods such as Cox Proportional Hazards test, Kaplan-Meier estimates, Aalen model, Power analysis, and Schoenfeld residuals.

```r
knitr::opts_chunk$set(message=FALSE, warning=FALSE, fig.height=3, fig.width=5,
                      fig.align="center")
library(tidyverse)
library(broom)
library(plyr)
library(survival)
library(survminer)
library(coxed)


aids <- read.csv( "http://pages.pomona.edu/~jsh04747/courses/math150/AIDSdata.csv")
dim(aids)
```

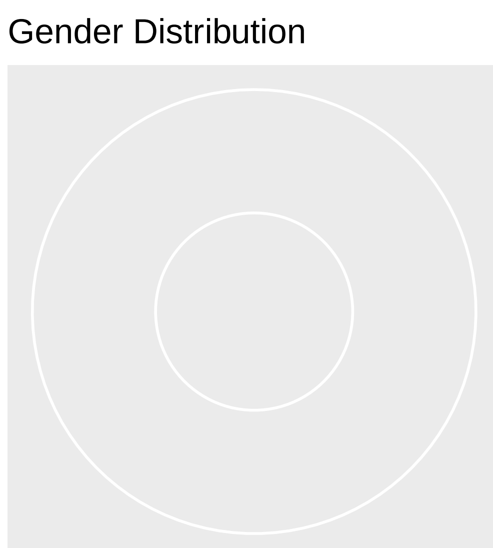
```
## [1] 851  16
```

```r
summary(aids)
```

```
##        id              time           censor            time_d     
##  Min.   :   1.0   Min.   :  1.0   Min.   :0.00000   Min.   :  1.0  
##  1st Qu.: 287.5   1st Qu.:179.5   1st Qu.:0.00000   1st Qu.:199.5  
##  Median : 581.0   Median :257.0   Median :0.00000   Median :266.0  
##  Mean   : 579.5   Mean   :231.8   Mean   :0.08108   Mean   :243.4  
##  3rd Qu.: 873.0   3rd Qu.:300.0   3rd Qu.:0.00000   3rd Qu.:306.0  
##  Max.   :1156.0   Max.   :362.0   Max.   :1.00000   Max.   :362.0  
##     censor_d            tx             txgrp           strat2      
##  Min.   :0.0000   Min.   :0.0000   Min.   :1.000   Min.   :0.0000  
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:1.000   1st Qu.:0.0000  
##  Median :0.0000   Median :1.0000   Median :2.000   Median :1.0000  
##  Mean   :0.0235   Mean   :0.5041   Mean   :1.504   Mean   :0.6157  
##  3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:2.000   3rd Qu.:1.0000  
##  Max.   :1.0000   Max.   :1.0000   Max.   :2.000   Max.   :1.0000  
##       sex            raceth          ivdrug         hemophil      
##  Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :0.00000  
##  1st Qu.:1.000   1st Qu.:1.000   1st Qu.:1.000   1st Qu.:0.00000  
##  Median :1.000   Median :1.000   Median :1.000   Median :0.00000  
##  Mean   :1.157   Mean   :1.706   Mean   :1.317   Mean   :0.03408  
##  3rd Qu.:1.000   3rd Qu.:2.000   3rd Qu.:1.000   3rd Qu.:0.00000  
##  Max.   :2.000   Max.   :5.000   Max.   :3.000   Max.   :1.00000  
##      karnof            cd4            priorzdv           age       
##  Min.   : 70.00   Min.   :  0.00   Min.   :  3.00   Min.   :15.00  
##  1st Qu.: 90.00   1st Qu.: 22.25   1st Qu.: 11.00   1st Qu.:33.00  
##  Median : 90.00   Median : 75.00   Median : 21.00   Median :38.00  
##  Mean   : 91.34   Mean   : 86.45   Mean   : 30.63   Mean   :38.81  
##  3rd Qu.:100.00   3rd Qu.:135.75   3rd Qu.: 44.00   3rd Qu.:44.00  
##  Max.   :100.00   Max.   :348.00   Max.   :288.00   Max.   :73.00  
```

The data set contains a sample size equal to 851 participants and 16 different variables. Out of these participants 782 were considered as uncensored data point, which indicates that these patients survived through the course of the study without diagnosis of AIDS and/or death. 69 were found to be censored meaning that either there was an occurence of death or AIDS diagnosis, out of which it is known that 20 patients died throughout the course of the study.

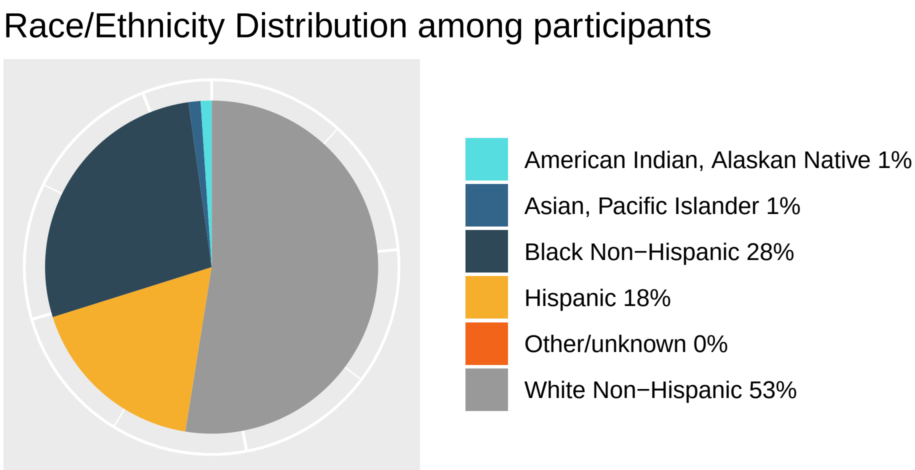
```
library(plotrix)
male<-sum(aids$sex=="male")
female<-sum(aids$sex=="female")
slices <- c(male, female)
lbls <- c("Male","Female")
pct <- round(slices/sum(slices)*100)
lbls <- paste(lbls, pct)
lbls <- paste(lbls,"%",sep="")
df = data.frame(slices = pct,labels =  lbls)
sexplot<- ggplot(df,aes(x = factor(1),y=slices,  fill = labels)) +
         geom_bar(stat="identity", width = 1)+
        coord_polar(theta = "y")+
        theme(axis.line = element_blank(),
          axis.text = element_blank(),
          axis.ticks = element_blank(),
          axis.title = element_blank())
print(sexplot + ggtitle("Gender Distribution")+ scale_fill_manual(values=c("#55DDE0", "#33658A","#2F4858")
```



The Pie Chart represents the gender distribution in the sample, with 84% male and 16% female. This shows the potential for the data to not be able to correctly represent the difference of the data variance by gender, if there were to be one. Therefore, gender is something to look into in future data analysis.

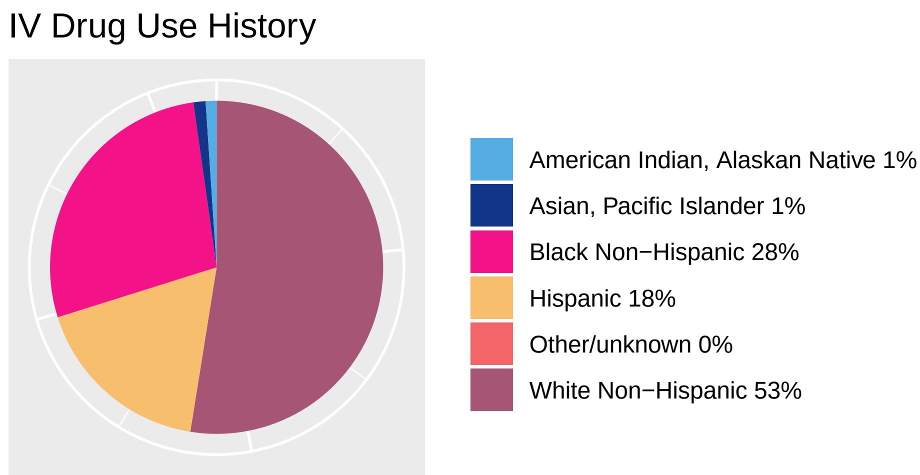
```
wnh<-sum(aids$raceth==1)
bnh<-sum(aids$raceth==2)
h<-sum(aids$raceth==3)
api<-sum(aids$raceth==4)
aian<-sum(aids$raceth==5)
oth<-sum(aids$raceth==6)
slices <- c(wnh,bnh,h,api,aian,oth)
lbls <- c("White Non-Hispanic", "Black Non-Hispanic", "Hispanic","Asian, Pacific Islander",
          "American Indian, Alaskan Native", "Other/unknown")
pct <- round(slices/sum(slices)*100)
lbls <- paste(lbls, pct)
```

```r
lbls <- paste(lbls,"%",sep="")
df = data.frame(slices = slices,labels =  lbls)

ethplot<- ggplot(df,aes(x = factor(1),y=slices,  fill = labels)) +
         geom_bar(stat="identity", width = 1)+
        coord_polar(theta = "y")+
        theme(axis.line = element_blank(),
          axis.text = element_blank(),
          axis.ticks = element_blank(),
          axis.title = element_blank())
print(ethplot + ggtitle("Race/Ethnicity Distribution among participants")+ scale_fill_manual(values=c("#55
```



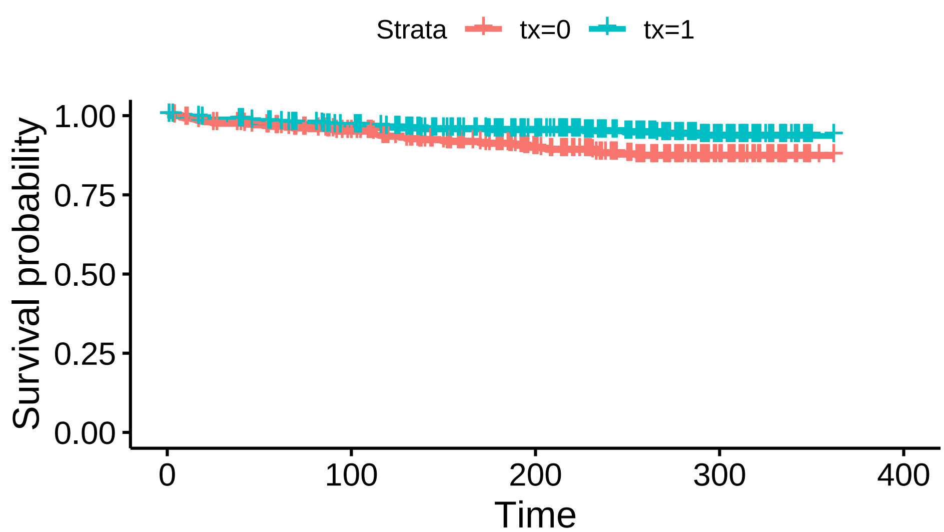
The distribution of race/ethnicity shows that the greatest number of participants consists of white non-Hispanic identifying individuals, with black non-Hispanic following and Hispanic as the 3rd largest represented group.

```r
never<-sum(aids$ivdrug==1)
cur<-sum(aids$ivdrug==2)
prev<-sum(aids$ivdrug==3)
slices <- c(never,cur,prev)
lbls <- c("Never", "Currently", "Previously")
pct <- round(slices/sum(slices)*100)
lbls <- paste(lbls, pct)
lbls <- paste(lbls,"%",sep="")
df = data.frame(slices = slices,labels =  lbls)
ivplot<- ggplot(df,aes(x = factor(1),y=slices,  fill = labels)) +
         geom_bar(stat="identity", width = 1)+
        coord_polar(theta = "y")+
        theme(axis.line = element_blank(),
          axis.text = element_blank(),
          axis.ticks = element_blank(),
          axis.title = element_blank())
print(ethplot + ggtitle("IV Drug Use History")+ scale_fill_manual(values=c("#56ADE3", "#12358A", "#F41288"
```

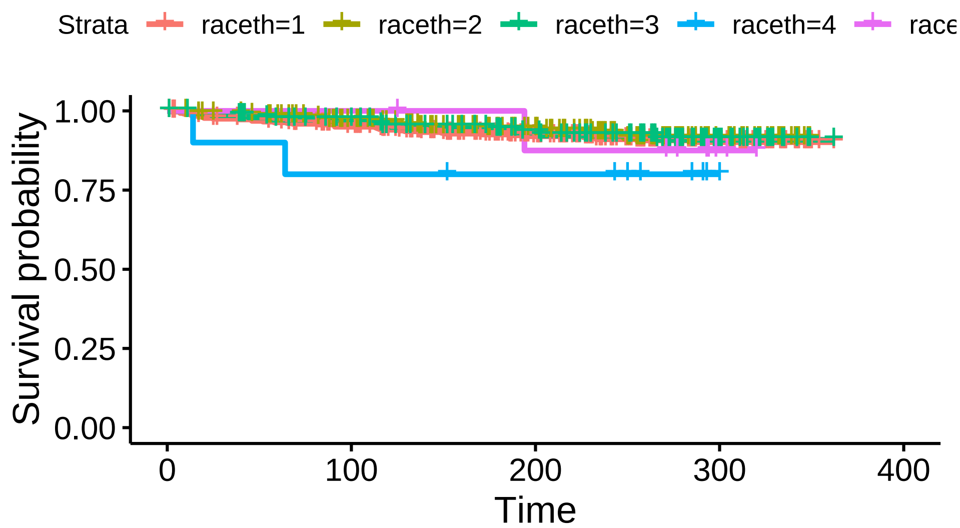


From this chart we see that most of the participants (84%) have never used IV drugs, whereas 16% of participants have some type of history of usage and none of the participants reported to be currently using the drugs.

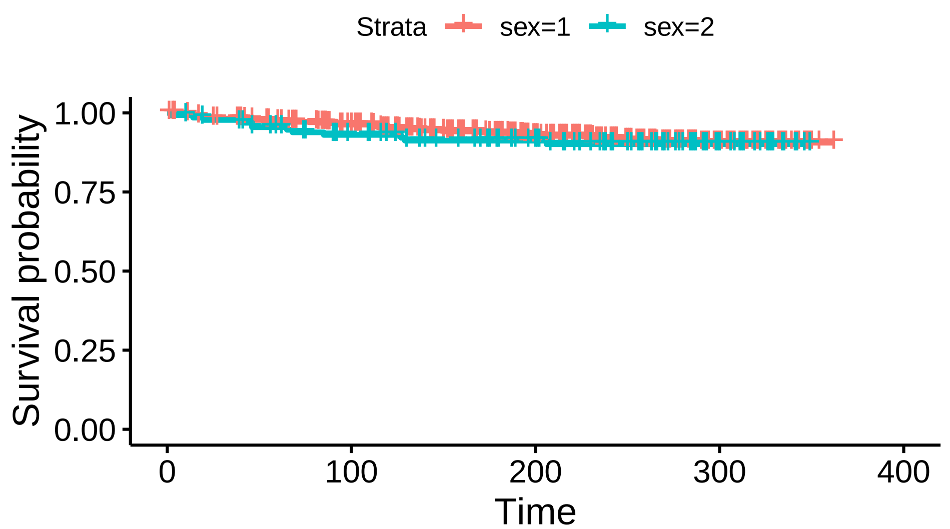
```
fit <- survfit(Surv(time,censor)~tx, data = aids)
ggsurvplot(fit,data = aids,conf.int = FALSE)
```



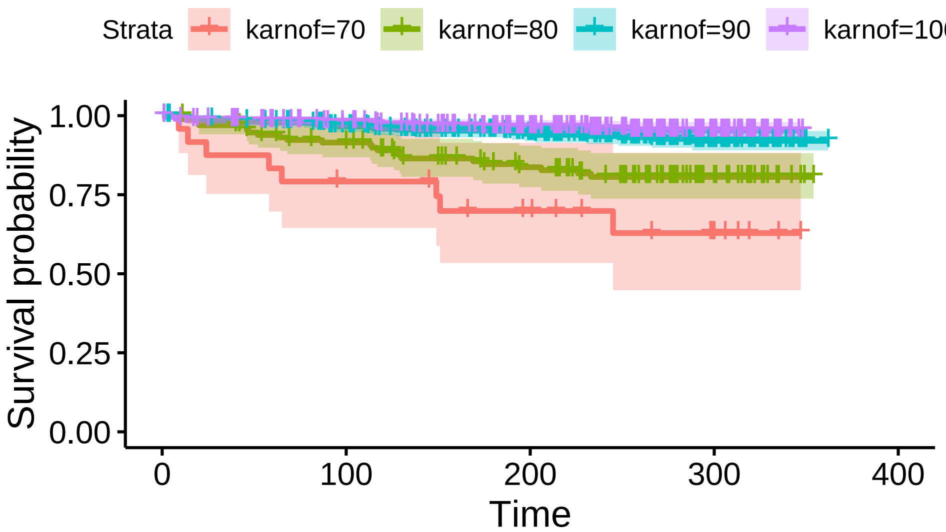

```
fit2 <- survfit(Surv(time,censor)~raceth, data = aids)
ggsurvplot(fit2,data = aids,conf.int = FALSE)
```




```
fit3 <- survfit(Surv(time,censor)~sex, data = aids)
ggsurvplot(fit3,data = aids,conf.int = FALSE)
```




```
aids_fit_time_k <- survfit(Surv(time, censor) ~karnof   , data=aids)
ggsurvplot(aids_fit_time_k, data=aids,  conf.int = TRUE)
```




```
#Survival Analysis
#mutation of age
aids <- read.csv( "http://pages.pomona.edu/~jsh04747/courses/math150/AIDSdata.csv")
aids <- aids %>%
  mutate(age = ifelse(age <= 20, "under20",
                             ifelse(age <=30, "20-30",
                                    ifelse(age <= 40, "30-40",
                                           ifelse(age <=50, "40-50",
                                                  ifelse(age <=60, "50-60",
                                                         ifelse(age <=70, "60-70",
                                                                "over70"))))))) %>%
  mutate(age = factor(age,
                      levels = c("under20", "20-30", "30-40","40-50", "50-60","60-70","over70")),
         sex = ifelse(sex == 2, "male","female"))

aids <- aids %>%
  mutate(cd4 = ifelse(cd4 <=50, "0-50",
                             ifelse(cd4 <=100, "50-100",
                                    ifelse(cd4 <= 150, "100-150",
                                           ifelse(cd4 <=200, "150-200",
                                                  ifelse(cd4 <=250, "200-250",
                                                         ifelse(cd4 <=300, "300-350", "350+")))))))
```

Since there are many values of the explanatory variable “age” in the original data, we’ve decided to mutate the variable into age categories from under 20 to over 70 in increments of 10 years.

```
library(survival)
library (survminer)
library(ggplot2)
library(broom)

##### backwards selection ######


#full model
cph_full<- coxph(Surv(time,censor)~.-time_d -censor_d, data = aids)
cph_full$loglik
```

```
## [1] -452.6325 -407.0079
```

```r
cph_full
```

```
## Call:
## coxph(formula = Surv(time, censor) ~ . - time_d - censor_d, data = aids)
##
##                 coef  exp(coef)   se(coef)      z        p
## id          5.197e-04  1.001e+00  3.636e-04  1.429  0.15295
## tx         -7.883e-01  4.546e-01  2.600e-01 -3.031  0.00243
## txgrp              NA         NA  0.000e+00     NA       NA
## strat2     -3.705e-01  6.904e-01  3.959e-01 -0.936  0.34940
## sexmale     3.388e-01  1.403e+00  3.320e-01  1.020  0.30756
## raceth     -5.390e-03  9.946e-01  1.440e-01 -0.037  0.97014
## ivdrug     -2.859e-01  7.513e-01  1.858e-01 -1.539  0.12377
## hemophil    3.339e-01  1.396e+00  6.874e-01  0.486  0.62711
## karnof     -6.481e-02  9.372e-01  1.426e-02 -4.546 5.47e-06
## cd4100-150 -2.623e+00  7.258e-02  1.069e+00 -2.453  0.01416
## cd4150-200 -1.987e+00  1.371e-01  1.077e+00 -1.845  0.06500
## cd4200-250 -1.365e+00  2.554e-01  1.068e+00 -1.278  0.20140
## cd4300-350 -1.670e+01  5.585e-08  3.685e+03 -0.005  0.99638
## cd4350+    -1.595e+01  1.183e-07  7.535e+03 -0.002  0.99831
## cd450-100  -9.339e-03  9.907e-01  4.222e-01 -0.022  0.98235
## priorzdv   -1.720e-03  9.983e-01  4.823e-03 -0.357  0.72134
## age20-30   -5.242e-01  5.920e-01  1.230e+00 -0.426  0.67005
## age30-40   -2.901e-01  7.482e-01  1.191e+00 -0.244  0.80758
## age40-50   -1.846e-01  8.315e-01  1.192e+00 -0.155  0.87689
## age50-60    6.235e-01  1.866e+00  1.211e+00  0.515  0.60671
## age60-70   -2.260e-01  7.977e-01  1.570e+00 -0.144  0.88550
## ageover70  -1.530e+01  2.262e-07  8.846e+03 -0.002  0.99862
##
## Likelihood ratio test=91.25  on 21 df, p=9.817e-11
## n= 851, number of events= 69
```

```r
#reduced model 1
cph_r1 <- coxph(Surv(time,censor)~.-time_d -censor_d -priorzdv, data = aids)
cph_r1$loglik
```

```
## [1] -452.6325 -407.0735
```

```r
cph_r1
```

```
## Call:
## coxph(formula = Surv(time, censor) ~ . - time_d - censor_d -
##     priorzdv, data = aids)
##
##                 coef  exp(coef)   se(coef)      z        p
## id          5.189e-04  1.001e+00  3.634e-04  1.428  0.15331
## tx         -7.933e-01  4.524e-01  2.598e-01 -3.053  0.00226
## txgrp              NA         NA  0.000e+00     NA       NA
## strat2     -3.761e-01  6.865e-01  3.979e-01 -0.945  0.34460
## sexmale     3.410e-01  1.406e+00  3.322e-01  1.027  0.30458
## raceth     -1.585e-03  9.984e-01  1.435e-01 -0.011  0.99119
## ivdrug     -2.882e-01  7.496e-01  1.859e-01 -1.550  0.12110
## hemophil    3.139e-01  1.369e+00  6.882e-01  0.456  0.64835
## karnof     -6.447e-02  9.376e-01  1.420e-02 -4.540 5.64e-06
## cd4100-150 -2.625e+00  7.244e-02  1.070e+00 -2.453  0.01417
## cd4150-200 -2.002e+00  1.351e-01  1.077e+00 -1.859  0.06298
## cd4200-250 -1.379e+00  2.519e-01  1.069e+00 -1.290  0.19702
## cd4300-350 -1.668e+01  5.700e-08  3.718e+03 -0.004  0.99642
```

```
## cd4350+    -1.596e+01  1.173e-07  7.581e+03 -0.002  0.99832
## cd450-100  -7.736e-03  9.923e-01  4.244e-01 -0.018  0.98546
## age20-30   -4.446e-01  6.411e-01  1.217e+00 -0.365  0.71491
## age30-40   -2.203e-01  8.023e-01  1.181e+00 -0.187  0.85203
## age40-50   -1.200e-01  8.870e-01  1.184e+00 -0.101  0.91928
## age50-60    6.912e-01  1.996e+00  1.203e+00  0.575  0.56563
## age60-70   -1.582e-01  8.537e-01  1.563e+00 -0.101  0.91936
## ageover70  -1.527e+01  2.324e-07  8.872e+03 -0.002  0.99863
##
## Likelihood ratio test=91.12  on 20 df, p=4.723e-11
## n= 851, number of events= 69
```

```
#likelihood ratio test
stat1<- 2*(cph_full$loglik[2]-cph_r1$loglik[2])
1-pchisq(stat1,1)
```

```
## [1] 0.7171496
```

```
#reduced model 2
cph_r2 <- coxph(Surv(time,censor)~. -time_d -censor_d -priorzdv -id, data = aids)
cph_r2$loglik
```

```
## [1] -452.6325 -408.1003
```

```
cph_r2
```

```
## Call:
## coxph(formula = Surv(time, censor) ~ . - time_d - censor_d -
##     priorzdv - id, data = aids)
##
##                  coef  exp(coef)   se(coef)      z        p
## tx         -7.689e-01  4.635e-01  2.591e-01 -2.968  0.00299
## txgrp              NA         NA  0.000e+00     NA       NA
## strat2     -3.535e-01  7.022e-01  4.031e-01 -0.877  0.38043
## sexmale     3.038e-01  1.355e+00  3.313e-01  0.917  0.35925
## raceth      2.684e-04  1.000e+00  1.438e-01  0.002  0.99851
## ivdrug     -2.634e-01  7.685e-01  1.848e-01 -1.425  0.15417
## hemophil    2.951e-01  1.343e+00  6.863e-01  0.430  0.66722
## karnof     -6.580e-02  9.363e-01  1.425e-02 -4.618 3.88e-06
## cd4100-150 -2.630e+00  7.211e-02  1.072e+00 -2.454  0.01413
## cd4150-200 -2.019e+00  1.327e-01  1.079e+00 -1.872  0.06118
## cd4200-250 -1.377e+00  2.523e-01  1.070e+00 -1.287  0.19810
## cd4300-350 -1.668e+01  5.675e-08  3.833e+03 -0.004  0.99653
## cd4350+    -1.591e+01  1.225e-07  7.620e+03 -0.002  0.99833
## cd450-100  -2.584e-02  9.745e-01  4.289e-01 -0.060  0.95196
## age20-30   -3.877e-01  6.786e-01  1.213e+00 -0.320  0.74931
## age30-40   -2.080e-01  8.122e-01  1.178e+00 -0.177  0.85989
## age40-50   -1.300e-01  8.781e-01  1.180e+00 -0.110  0.91223
## age50-60    6.972e-01  2.008e+00  1.200e+00  0.581  0.56112
## age60-70   -1.464e-01  8.638e-01  1.559e+00 -0.094  0.92521
## ageover70  -1.513e+01  2.698e-07  8.722e+03 -0.002  0.99862
##
## Likelihood ratio test=89.06  on 19 df, p=4.857e-11
## n= 851, number of events= 69
```

```
#likelihood ratio test
stat2 <- 2*(cph_r1$loglik[2]-cph_r2$loglik[2])
1-pchisq(stat2,1)
```

```
## [1] 0.1518505
```

```r
#reduced model 3
cph_r3 <- coxph(Surv(time,censor)~.-time_d -censor_d -priorzdv -id -hemophil, data = aids)
cph_r3$loglik
```

```
## [1] -452.6325 -408.1860
```

```r
cph_r3
```

```
## Call:
## coxph(formula = Surv(time, censor) ~ . - time_d - censor_d -
##     priorzdv - id - hemophil, data = aids)
##
##                  coef  exp(coef)   se(coef)      z        p
## tx         -7.753e-01  4.606e-01  2.584e-01 -3.001  0.00269
## txgrp              NA         NA  0.000e+00     NA       NA
## strat2     -3.426e-01  7.099e-01  4.018e-01 -0.853  0.39383
## sexmale     2.934e-01  1.341e+00  3.302e-01  0.888  0.37432
## raceth     -9.113e-04  9.991e-01  1.445e-01 -0.006  0.99497
## ivdrug     -2.634e-01  7.684e-01  1.846e-01 -1.427  0.15350
## karnof     -6.545e-02  9.366e-01  1.422e-02 -4.605 4.13e-06
## cd4100-150 -2.634e+00  7.177e-02  1.072e+00 -2.458  0.01396
## cd4150-200 -2.031e+00  1.312e-01  1.078e+00 -1.884  0.05956
## cd4200-250 -1.396e+00  2.477e-01  1.069e+00 -1.306  0.19161
## cd4300-350 -1.673e+01  5.441e-08  3.798e+03 -0.004  0.99649
## cd4350+    -1.583e+01  1.330e-07  7.501e+03 -0.002  0.99832
## cd450-100  -2.839e-02  9.720e-01  4.288e-01 -0.066  0.94721
## age20-30   -6.207e-01  5.375e-01  1.079e+00 -0.575  0.56520
## age30-40   -4.540e-01  6.351e-01  1.023e+00 -0.444  0.65729
## age40-50   -3.696e-01  6.910e-01  1.034e+00 -0.357  0.72086
## age50-60    4.517e-01  1.571e+00  1.048e+00  0.431  0.66648
## age60-70   -4.024e-01  6.687e-01  1.436e+00 -0.280  0.77935
## ageover70  -1.539e+01  2.080e-07  8.720e+03 -0.002  0.99859
##
## Likelihood ratio test=88.89  on 18 df, p=2.281e-11
## n= 851, number of events= 69
```

```r
#likelihood ratio test
stat3 <- 2*(cph_r3$loglik[2]-cph_r2$loglik[2])
1-pchisq(stat3,1)
```

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

```r
#reduced model 4
cph_r4 <- coxph(Surv(time,censor)~.-time_d -censor_d -priorzdv -id -hemophil -raceth, data = aids)
cph_r4$loglik
```

```
## [1] -452.6325 -408.1860
```

```r
cph_r4
```

```
## Call:
## coxph(formula = Surv(time, censor) ~ . - time_d - censor_d -
##     priorzdv - id - hemophil - raceth, data = aids)
##
##                  coef  exp(coef)   se(coef)      z        p
## tx         -7.753e-01  4.606e-01  2.583e-01 -3.001  0.00269
## txgrp              NA         NA  0.000e+00     NA       NA
## strat2     -3.426e-01  7.099e-01  4.017e-01 -0.853  0.39380
## sexmale     2.932e-01  1.341e+00  3.285e-01  0.892  0.37217
## ivdrug     -2.636e-01  7.683e-01  1.824e-01 -1.445  0.14843
```

```
## karnof     -6.546e-02  9.366e-01  1.420e-02 -4.611 4.01e-06
## cd4100-150 -2.634e+00  7.177e-02  1.072e+00 -2.458  0.01396
## cd4150-200 -2.031e+00  1.312e-01  1.078e+00 -1.884  0.05954
## cd4200-250 -1.395e+00  2.477e-01  1.068e+00 -1.306  0.19150
## cd4300-350 -1.673e+01  5.444e-08  3.798e+03 -0.004  0.99649
## cd4350+    -1.584e+01  1.325e-07  7.514e+03 -0.002  0.99832
## cd450-100  -2.835e-02  9.720e-01  4.287e-01 -0.066  0.94728
## age20-30   -6.207e-01  5.376e-01  1.079e+00 -0.575  0.56522
## age30-40   -4.537e-01  6.353e-01  1.022e+00 -0.444  0.65722
## age40-50   -3.693e-01  6.912e-01  1.034e+00 -0.357  0.72083
## age50-60    4.518e-01  1.571e+00  1.048e+00  0.431  0.66627
## age60-70   -4.018e-01  6.691e-01  1.434e+00 -0.280  0.77928
## ageover70  -1.537e+01  2.114e-07  8.652e+03 -0.002  0.99858
##
## Likelihood ratio test=88.89  on 17 df, p=9.7e-12
## n= 851, number of events= 69
```

```r
#likelihood ratio test
stat4 <- 2*(cph_r3$loglik[2]-cph_r4$loglik[2])
1-pchisq(stat4,1)
```

```
## [1] 0.9949662
```

```r
#reduced model 5
cph_r5 <- coxph(Surv(time,censor)~.-time_d -censor_d -priorzdv -id -hemophil -raceth, data = aids)
cph_r5$loglik
```

```
## [1] -452.6325 -408.1860
```

```r
cph_r5
```

```
## Call:
## coxph(formula = Surv(time, censor) ~ . - time_d - censor_d -
##     priorzdv - id - hemophil - raceth, data = aids)
##
##                  coef  exp(coef)   se(coef)      z        p
## tx         -7.753e-01  4.606e-01  2.583e-01 -3.001  0.00269
## txgrp              NA         NA  0.000e+00     NA       NA
## strat2     -3.426e-01  7.099e-01  4.017e-01 -0.853  0.39380
## sexmale     2.932e-01  1.341e+00  3.285e-01  0.892  0.37217
## ivdrug     -2.636e-01  7.683e-01  1.824e-01 -1.445  0.14843
## karnof     -6.546e-02  9.366e-01  1.420e-02 -4.611 4.01e-06
## cd4100-150 -2.634e+00  7.177e-02  1.072e+00 -2.458  0.01396
## cd4150-200 -2.031e+00  1.312e-01  1.078e+00 -1.884  0.05954
## cd4200-250 -1.395e+00  2.477e-01  1.068e+00 -1.306  0.19150
## cd4300-350 -1.673e+01  5.444e-08  3.798e+03 -0.004  0.99649
## cd4350+    -1.584e+01  1.325e-07  7.514e+03 -0.002  0.99832
## cd450-100  -2.835e-02  9.720e-01  4.287e-01 -0.066  0.94728
## age20-30   -6.207e-01  5.376e-01  1.079e+00 -0.575  0.56522
## age30-40   -4.537e-01  6.353e-01  1.022e+00 -0.444  0.65722
## age40-50   -3.693e-01  6.912e-01  1.034e+00 -0.357  0.72083
## age50-60    4.518e-01  1.571e+00  1.048e+00  0.431  0.66627
## age60-70   -4.018e-01  6.691e-01  1.434e+00 -0.280  0.77928
## ageover70  -1.537e+01  2.114e-07  8.652e+03 -0.002  0.99858
##
## Likelihood ratio test=88.89  on 17 df, p=9.7e-12
## n= 851, number of events= 69
```

```r
#likelihood ratio test
stat5 <- 2*(cph_r5$loglik[2]-cph_r4$loglik[2])
1-pchisq(stat5,1)
```

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

```r
#reduced model 6
cph_r6 <- coxph(Surv(time,censor)~.-time_d -censor_d -priorzdv -id -hemophil -raceth -strat2, data = aids)
cph_r6$loglik
```

```
## [1] -452.6325 -408.5655
```

```r
cph_r6
```

```
## Call:
## coxph(formula = Surv(time, censor) ~ . - time_d - censor_d -
##     priorzdv - id - hemophil - raceth - strat2, data = aids)
##
##                   coef  exp(coef)   se(coef)      z        p
## tx          -7.716e-01  4.623e-01  2.583e-01 -2.987  0.00282
## txgrp               NA         NA  0.000e+00     NA       NA
## sexmale      2.926e-01  1.340e+00  3.280e-01  0.892  0.37238
## ivdrug      -2.487e-01  7.798e-01  1.816e-01 -1.370  0.17066
## karnof      -6.579e-02  9.363e-01  1.415e-02 -4.649 3.33e-06
## cd4100-150 -2.929e+00  5.343e-02  1.012e+00 -2.894  0.00380
## cd4150-200 -2.333e+00  9.701e-02  1.016e+00 -2.296  0.02169
## cd4200-250 -1.669e+00  1.885e-01  1.018e+00 -1.640  0.10108
## cd4300-350 -1.703e+01  4.022e-08  3.788e+03 -0.004  0.99641
## cd4350+     -1.616e+01  9.565e-08  7.517e+03 -0.002  0.99828
## cd450-100   -2.890e-01  7.490e-01  3.013e-01 -0.959  0.33742
## age20-30    -6.658e-01  5.138e-01  1.078e+00 -0.617  0.53692
## age30-40    -4.846e-01  6.159e-01  1.022e+00 -0.474  0.63538
## age40-50    -4.076e-01  6.652e-01  1.033e+00 -0.395  0.69310
## age50-60     3.804e-01  1.463e+00  1.045e+00  0.364  0.71582
## age60-70    -4.572e-01  6.330e-01  1.433e+00 -0.319  0.74967
## ageover70   -1.544e+01  1.964e-07  8.671e+03 -0.002  0.99858
##
## Likelihood ratio test=88.13  on 16 df, p=5.513e-12
## n= 851, number of events= 69
```

```r
#likelihood ratio test
stat6 <- 2*(cph_r5$loglik[2]-cph_r6$loglik[2])
1-pchisq(stat6,1)
```

```
## [1] 0.3836735
```

```r
#reduced model 7
cph_r7 <- coxph(Surv(time,censor)~.-time_d -censor_d -priorzdv -id -hemophil -raceth -strat2 -sex, data = a
cph_r7$loglik
```

```
## [1] -452.6325 -408.9408
```

```r
cph_r7
```

```
## Call:
## coxph(formula = Surv(time, censor) ~ . - time_d - censor_d -
##     priorzdv - id - hemophil - raceth - strat2 - sex, data = aids)
##
##                   coef  exp(coef)   se(coef)      z        p
## tx          -7.732e-01  4.615e-01  2.585e-01 -2.991  0.00278
```

```
## txgrp                NA          NA  0.000e+00     NA        NA
## ivdrug     -2.491e-01  7.795e-01  1.815e-01 -1.372  0.16996
## karnof     -6.558e-02  9.365e-01  1.403e-02 -4.674 2.95e-06
## cd4100-150 -2.931e+00  5.334e-02  1.012e+00 -2.896  0.00378
## cd4150-200 -2.328e+00  9.754e-02  1.016e+00 -2.291  0.02198
## cd4200-250 -1.699e+00  1.828e-01  1.017e+00 -1.671  0.09470
## cd4300-350 -1.700e+01  4.139e-08  3.787e+03 -0.004  0.99642
## cd4350+    -1.621e+01  9.126e-08  7.495e+03 -0.002  0.99827
## cd450-100  -2.942e-01  7.451e-01  3.009e-01 -0.978  0.32818
## age20-30   -5.975e-01  5.502e-01  1.076e+00 -0.555  0.57855
## age30-40   -4.974e-01  6.081e-01  1.022e+00 -0.487  0.62645
## age40-50   -4.165e-01  6.594e-01  1.033e+00 -0.403  0.68686
## age50-60    3.915e-01  1.479e+00  1.045e+00  0.375  0.70796
## age60-70   -4.995e-01  6.068e-01  1.432e+00 -0.349  0.72719
## ageover70  -1.548e+01  1.893e-07  8.644e+03 -0.002  0.99857
##
## Likelihood ratio test=87.38  on 15 df, p=3.044e-12
## n= 851, number of events= 69
```

```r
#likelihood ratio test
stat7 <- 2*(cph_r6$loglik[2]-cph_r7$loglik[2])
1-pchisq(stat7,1)
```

```
## [1] 0.3862684
```

```r
#reduced model 8
cph_r8 <- coxph(Surv(time,censor)~.-time_d -censor_d -priorzdv -id -hemophil -raceth -strat2 -sex -txgrp -a
cph_r8$loglik
```

```
## [1] -452.6325 -412.2234
```

```r
cph_r8
```

```
## Call:
## coxph(formula = Surv(time, censor) ~ . - time_d - censor_d -
##     priorzdv - id - hemophil - raceth - strat2 - sex - txgrp -
##     age, data = aids)
##
##                  coef  exp(coef)   se(coef)      z       p
## tx         -7.170e-01  4.882e-01  2.560e-01 -2.800 0.00510
## ivdrug     -2.566e-01  7.737e-01  1.812e-01 -1.416 0.15683
## karnof     -6.572e-02  9.364e-01  1.403e-02 -4.685 2.8e-06
## cd4100-150 -2.937e+00  5.305e-02  1.012e+00 -2.902 0.00371
## cd4150-200 -2.395e+00  9.118e-02  1.013e+00 -2.365 0.01805
## cd4200-250 -1.581e+00  2.058e-01  1.015e+00 -1.558 0.11931
## cd4300-350 -1.594e+01  1.198e-07  2.411e+03 -0.007 0.99473
## cd4350+    -1.529e+01  2.291e-07  4.435e+03 -0.003 0.99725
## cd450-100  -3.081e-01  7.348e-01  2.983e-01 -1.033 0.30160
##
## Likelihood ratio test=80.82  on 9 df, p=1.111e-13
## n= 851, number of events= 69
```

```r
#likelihood ratio test
stat8 <- 2*(cph_r7$loglik[2]-cph_r8$loglik[2])
1-pchisq(stat8,1)
```

```
## [1] 0.01039962
```

```r
#reduced model 9
cph_r9 <- coxph(Surv(time,censor)~.-time_d -censor_d -priorzdv -id -hemophil -raceth -strat2 -sex -txgrp -a
```

```r
cph_r9$loglik
```

```
## [1] -452.6325 -416.3863
```

```r
cph_r9
```

```
## Call:
## coxph(formula = Surv(time, censor) ~ . - time_d - censor_d -
##     priorzdv - id - hemophil - raceth - strat2 - sex - txgrp -
##     age - tx, data = aids)
##
##                  coef  exp(coef)   se(coef)      z        p
## ivdrug     -2.252e-01  7.983e-01  1.804e-01 -1.249  0.21177
## karnof     -6.577e-02  9.363e-01  1.415e-02 -4.648 3.36e-06
## cd4100-150 -3.006e+00  4.949e-02  1.011e+00 -2.973  0.00295
## cd4150-200 -2.431e+00  8.798e-02  1.013e+00 -2.400  0.01640
## cd4200-250 -1.618e+00  1.983e-01  1.014e+00 -1.595  0.11070
## cd4300-350 -1.593e+01  1.205e-07  2.298e+03 -0.007  0.99447
## cd4350+    -1.549e+01  1.877e-07  4.449e+03 -0.003  0.99722
## cd450-100  -3.152e-01  7.296e-01  2.975e-01 -1.060  0.28935
##
## Likelihood ratio test=72.49  on 8 df, p=1.565e-12
## n= 851, number of events= 69
```

```r
###best model using backwards selection?
#likelihood ratio test
stat9 <- 2*(cph_r8$loglik[2]-cph_r9$loglik[2])
1-pchisq(stat9,1)
```

```
## [1] 0.003908424
```

```r
#reduced model 10
cph_r10 <- coxph(Surv(time,censor)~.-priorzdv -id -hemophil -raceth -time_d -strat2 -sex -txgrp -age -tx -
cph_r10$loglik
```

```
## [1] -452.6325 -416.3863
```

```r
cph_r10
```

```
## Call:
## coxph(formula = Surv(time, censor) ~ . - priorzdv - id - hemophil -
##     raceth - time_d - strat2 - sex - txgrp - age - tx - censor_d,
##     data = aids)
##
##                  coef  exp(coef)   se(coef)      z        p
## ivdrug     -2.252e-01  7.983e-01  1.804e-01 -1.249  0.21177
## karnof     -6.577e-02  9.363e-01  1.415e-02 -4.648 3.36e-06
## cd4100-150 -3.006e+00  4.949e-02  1.011e+00 -2.973  0.00295
## cd4150-200 -2.431e+00  8.798e-02  1.013e+00 -2.400  0.01640
## cd4200-250 -1.618e+00  1.983e-01  1.014e+00 -1.595  0.11070
## cd4300-350 -1.593e+01  1.205e-07  2.298e+03 -0.007  0.99447
## cd4350+    -1.549e+01  1.877e-07  4.449e+03 -0.003  0.99722
## cd450-100  -3.152e-01  7.296e-01  2.975e-01 -1.060  0.28935
##
## Likelihood ratio test=72.49  on 8 df, p=1.565e-12
## n= 851, number of events= 69
```

```r
#NOTE: should we take out censor_d anyways since its related to censor or keep it?
#likelihood ratio test
stat10 <- 2*(cph_r9$loglik[2]-cph_r10$loglik[2])
```

```r
1-pchisq(stat10,1)
```

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

```r
cph_r11 <- coxph(Surv(time,censor)~.-priorzdv -id -hemophil -raceth -time_d -strat2 -sex -txgrp -age -tx -
cph_r11$loglik
```

```
## [1] -452.6325 -417.2537
```

```r
cph_r11
```

```
## Call:
## coxph(formula = Surv(time, censor) ~ . - priorzdv - id - hemophil -
##     raceth - time_d - strat2 - sex - txgrp - age - tx - censor_d -
##     ivdrug, data = aids)
##
##                   coef  exp(coef)   se(coef)      z        p
## karnof     -6.358e-02  9.384e-01  1.401e-02 -4.539 5.65e-06
## cd4100-150 -2.997e+00  4.995e-02  1.011e+00 -2.964  0.00303
## cd4150-200 -2.441e+00  8.711e-02  1.013e+00 -2.409  0.01598
## cd4200-250 -1.634e+00  1.951e-01  1.014e+00 -1.611  0.10709
## cd4300-350 -1.689e+01  4.616e-08  3.748e+03 -0.005  0.99640
## cd4350+    -1.670e+01  5.596e-08  7.114e+03 -0.002  0.99813
## cd450-100  -3.191e-01  7.268e-01  2.975e-01 -1.073  0.28336
##
## Likelihood ratio test=70.76  on 7 df, p=1.038e-12
## n= 851, number of events= 69
```

```r
stat11 <- 2*(cph_r10$loglik[2]-cph_r11$loglik[2])
1-pchisq(stat11,1)
```

```
## [1] 0.1878145
```

```r
coxph(Surv(time_d,censor_d) ~ sex , data=aids)  %>% tidy()
```

```
## # A tibble: 1 x 7
##   term    estimate std.error statistic p.value conf.low conf.high
##   <chr>      <dbl>     <dbl>     <dbl>   <dbl>    <dbl>     <dbl>
## 1 sexmale    0.390     0.559     0.697   0.486   -0.706      1.49
```

```r
coxph(Surv(time,censor) ~ sex, data=aids) %>% tidy()
```

```
## # A tibble: 1 x 7
##   term    estimate std.error statistic p.value conf.low conf.high
##   <chr>      <dbl>     <dbl>     <dbl>   <dbl>    <dbl>     <dbl>
## 1 sexmale    0.199     0.318     0.625   0.532   -0.424     0.821
```

```r
coxph(Surv(time,censor) ~ age+ txgrp+ karnof, data=aids) %>% tidy()
```

```
## # A tibble: 8 x 7
##   term      estimate std.error statistic       p.value conf.low conf.high
##   <chr>        <dbl>     <dbl>     <dbl>         <dbl>    <dbl>     <dbl>
## 1 age20-30   -0.438     1.07    -0.409   0.682            -2.53     1.66
## 2 age30-40   -0.442     1.02    -0.434   0.665            -2.44     1.55
## 3 age40-50   -0.361     1.03    -0.352   0.725            -2.37     1.65
## 4 age50-60    0.460     1.04     0.442   0.659            -1.58     2.50
## 5 age60-70   -0.780     1.42    -0.551   0.582            -3.55     2.00
## 6 ageover70 -14.1    2688.      -0.00525 0.996          -Inf      Inf
## 7 txgrp      -0.844     0.257   -3.28    0.00103          -1.35    -0.340
## 8 karnof     -0.0814    0.0138  -5.89    0.00000000385   -0.109   -0.0543
```

```r
cox.zph(coxph(Surv(time,censor) ~ age + txgrp+karnof, data=aids))
```

```
##                rho    chisq     p
## age20-30   0.09054 5.70e-01 0.450
## age30-40   0.19294 2.53e+00 0.112
## age40-50   0.14871 1.50e+00 0.220
## age50-60   0.19861 2.69e+00 0.101
## age60-70   0.16251 1.81e+00 0.179
## ageover70  0.16355 2.57e-07 1.000
## txgrp     -0.10779 8.34e-01 0.361
## karnof     0.00121 1.03e-04 0.992
## GLOBAL          NA 7.98e+00 0.435
```

```r
coxph(Surv(time,censor) ~ age *txgrp*karnof, data=aids) %>% tidy()
```

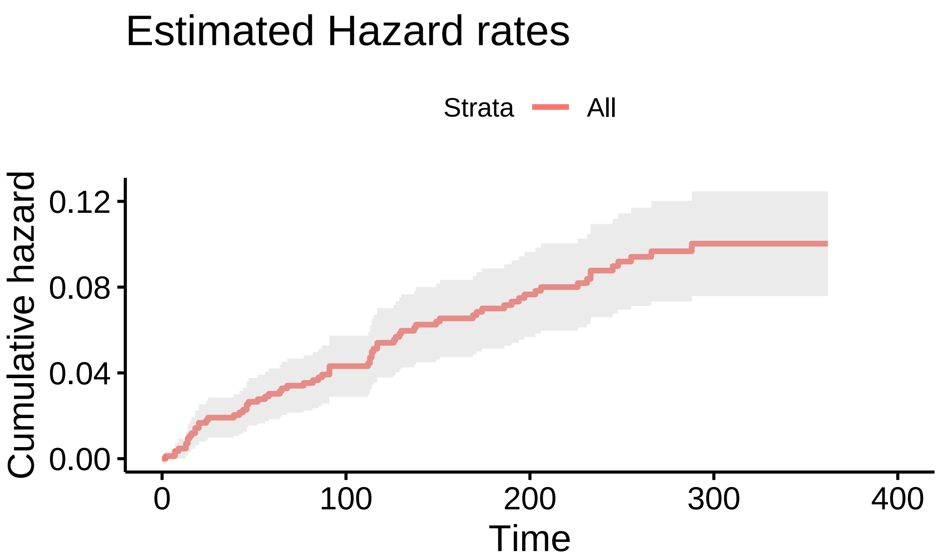
```
## # A tibble: 27 x 7
##    term           estimate std.error  statistic p.value conf.low conf.high
##    <chr>             <dbl>     <dbl>      <dbl>   <dbl>    <dbl>     <dbl>
##  1 age20-30         307.     138277.  0.00222     0.998     -Inf       Inf
##  2 age30-40         319.     138277.  0.00231     0.998     -Inf       Inf
##  3 age40-50         327.     138277.  0.00237     0.998     -Inf       Inf
##  4 age50-60         343.     138277.  0.00248     0.998     -Inf       Inf
##  5 age60-70         287.     176491.  0.00163     0.999     -Inf       Inf
##  6 ageover70         -1.66    29414. -0.0000565   1.000     -Inf       Inf
##  7 txgrp            150.      92392.  0.00163     0.999     -Inf       Inf
##  8 karnof             3.36     1424.  0.00236     0.998     -Inf       Inf
##  9 age20-30:txgrp  -144.      92392. -0.00156     0.999     -Inf       Inf
## 10 age30-40:txgrp  -146.      92392. -0.00158     0.999     -Inf       Inf
## # ... with 17 more rows
```

```r
cox.zph(coxph(Surv(time,censor) ~ age *txgrp*karnof, data=aids))
```

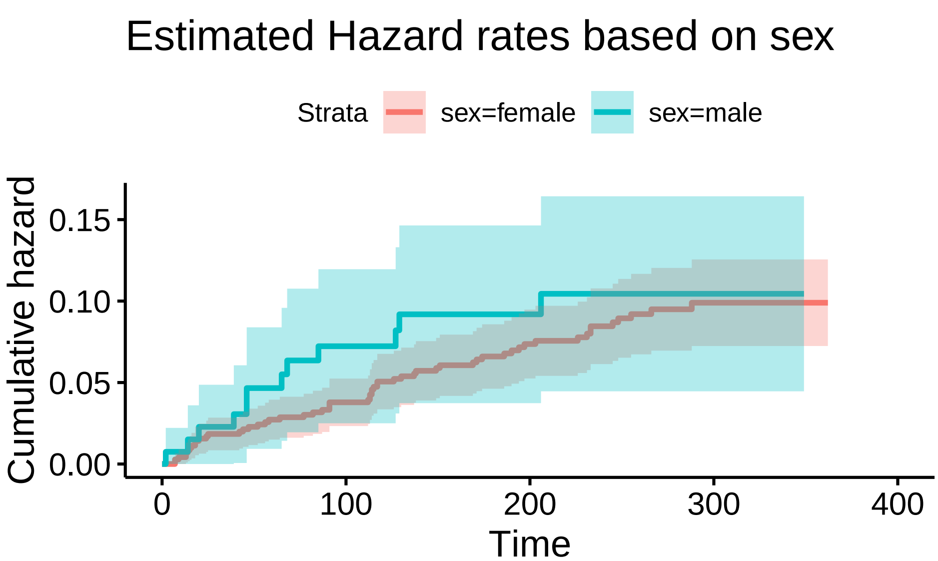
```
##                            rho    chisq     p
## age20-30               -0.1008 4.31e-08 1.000
## age30-40               -0.1583 3.15e-08 1.000
## age40-50               -0.0965 1.25e-08 1.000
## age50-60               -0.2071 6.53e-08 1.000
## age60-70               -0.2062 3.04e-08 1.000
## ageover70              -0.2493 7.81e-11 1.000
## txgrp                  -0.2032 2.68e-08 1.000
## karnof                 -0.1974 5.24e-08 1.000
## age20-30:txgrp          0.0921 2.14e-08 1.000
## age30-40:txgrp          0.1142 1.08e-08 1.000
## age40-50:txgrp          0.0826 5.64e-09 1.000
## age50-60:txgrp          0.1851 3.47e-08 1.000
## age60-70:txgrp          0.2102 2.15e-08 1.000
## ageover70:txgrp         0.1967 3.96e-11 1.000
## age20-30:karnof         0.0984 4.53e-08 1.000
## age30-40:karnof         0.1524 3.44e-08 1.000
## age40-50:karnof         0.0938 1.40e-08 1.000
## age50-60:karnof         0.2053 7.78e-08 1.000
## age60-70:karnof         0.1978 3.00e-08 1.000
## ageover70:karnof            NA      NaN   NaN
## txgrp:karnof            0.1996 2.81e-08 1.000
## age20-30:txgrp:karnof  -0.0910 2.15e-08 1.000
## age30-40:txgrp:karnof  -0.1020 9.71e-09 1.000
## age40-50:txgrp:karnof  -0.0823 6.23e-09 1.000
## age50-60:txgrp:karnof  -0.1796 3.72e-08 1.000
```

```
## age60-70:txgrp:karnof   -0.1981 1.98e-08 1.000
## ageover70:txgrp:karnof        NA      NaN   NaN
## GLOBAL                        NA 1.84e+01 0.891
```

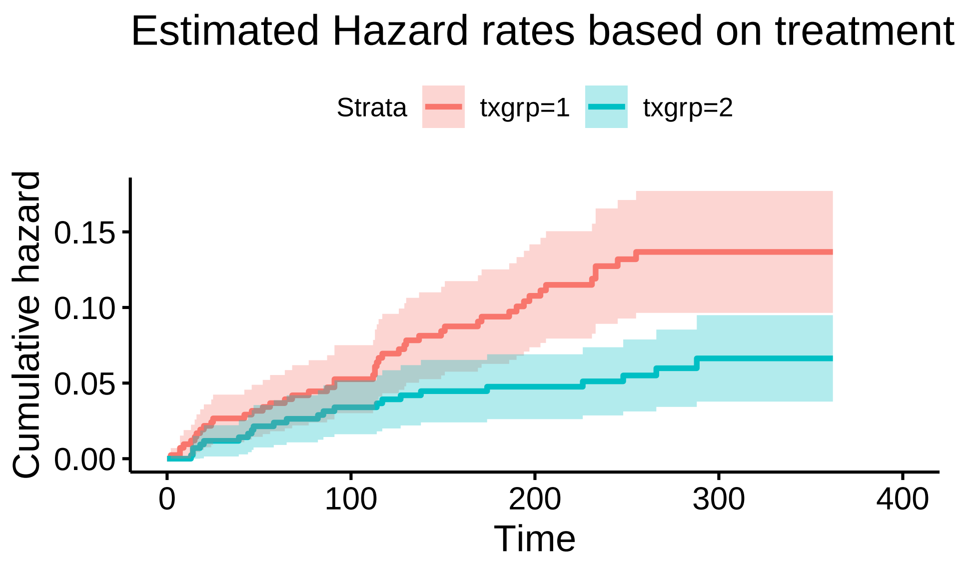
```r
ggsurvplot(survfit(Surv(time,censor) ~ 1, data=aids),
           censor=F, conf.int=T, fun="cumhaz") + ggtitle("Estimated Hazard rates")
```



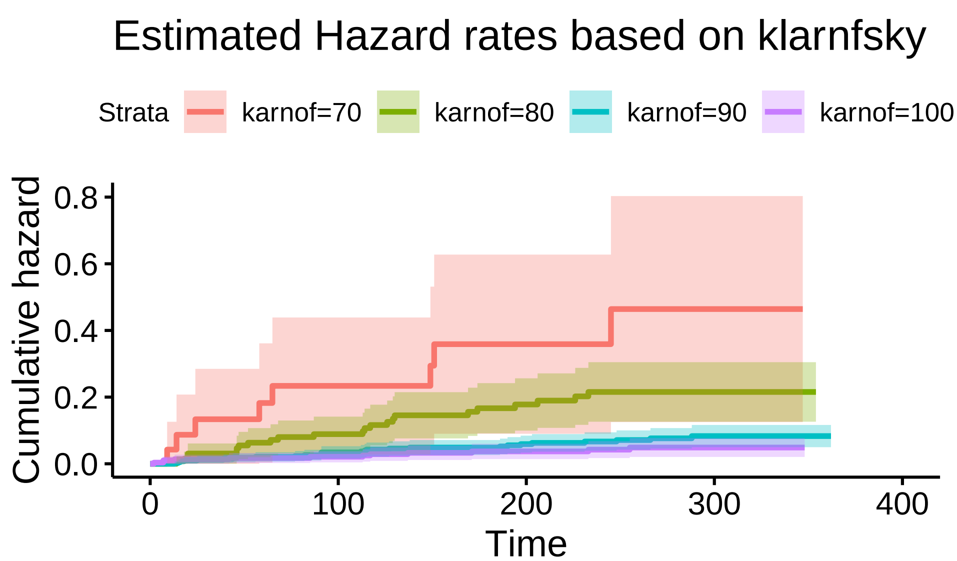

```r
ggsurvplot(survfit(Surv(time,censor) ~ sex, data=aids),
           censor=F, conf.int=T, fun="cumhaz") +
  ggtitle("Estimated Hazard rates based on sex")
```



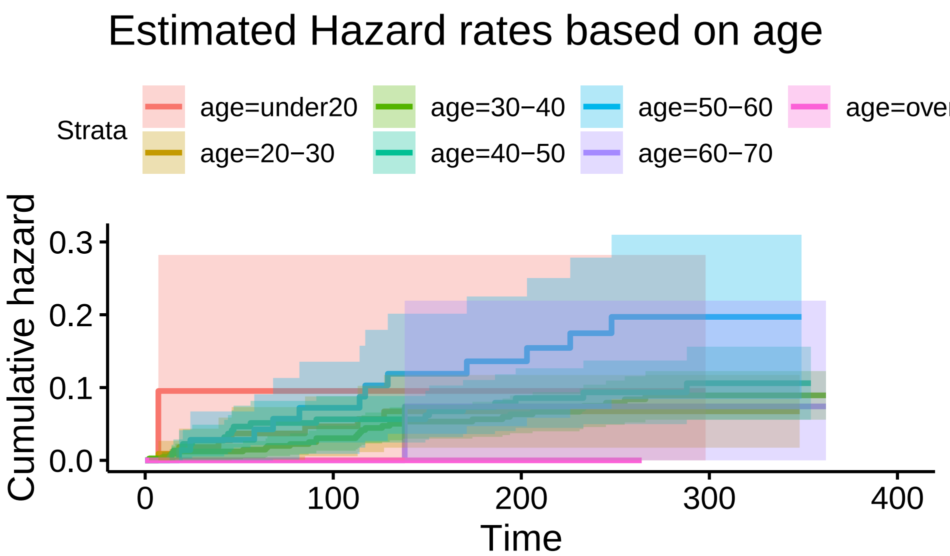

```r
ggsurvplot(survfit(Surv(time,censor) ~ txgrp, data=aids),
           censor=F, conf.int=T, fun="cumhaz") +
  ggtitle("Estimated Hazard rates based on treatment group")
```



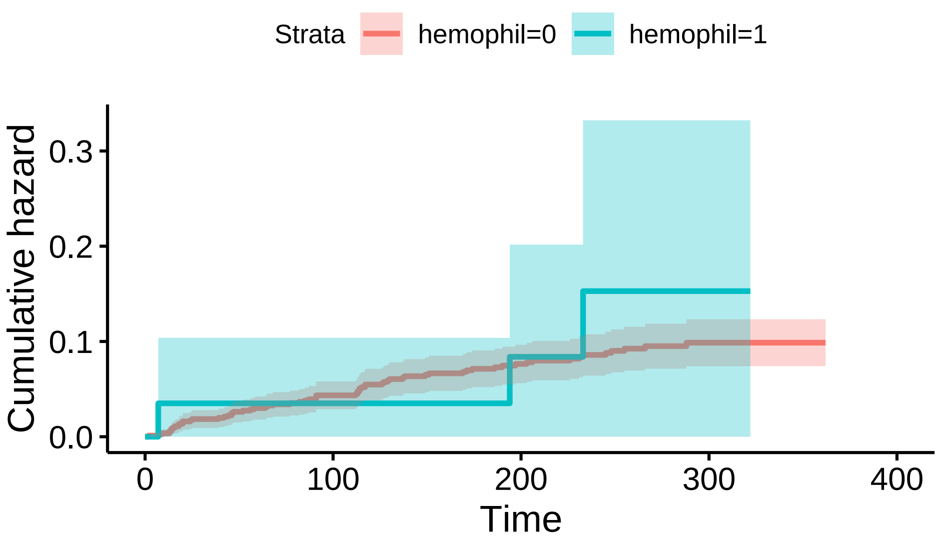

```r
ggsurvplot(survfit(Surv(time,censor) ~ karnof, data=aids),
           censor=F, conf.int=T, fun="cumhaz") +
  ggtitle("Estimated Hazard rates based on klarnfsky")
```



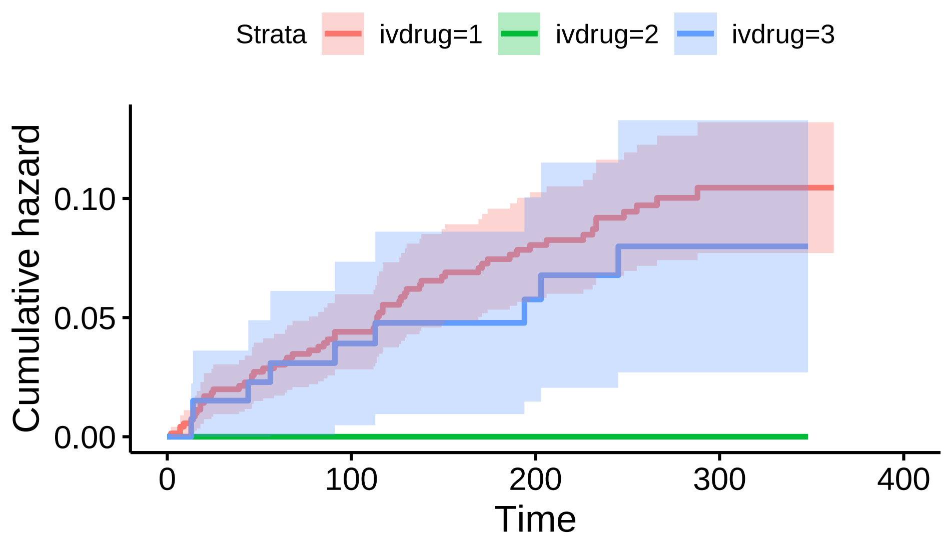

```r
ggsurvplot(survfit(Surv(time,censor) ~ age, data=aids),
           censor=F, conf.int=T, fun="cumhaz") +
  ggtitle("Estimated Hazard rates based on age")
```




```
ggsurvplot(survfit(Surv(time, censor)~hemophil, data = aids),
           censor=F, conf.int = T, fun = "cumhaz")
```




```
ggsurvplot(survfit(Surv(time, censor)~ivdrug, data = aids),
           censor=F, conf.int = T, fun = "cumhaz")
```




```
coxph(Surv(time,censor) ~ ivdrug, data=aids) %>% tidy()
```

```
## # A tibble: 1 x 7
##   term   estimate std.error statistic p.value conf.low conf.high
##   <chr>     <dbl>     <dbl>     <dbl>   <dbl>    <dbl>     <dbl>
## 1 ivdrug   -0.130     0.179    -0.723   0.470   -0.481     0.222
```

```
coxph(Surv(time,censor) ~ ivdrug*karnof, data=aids) %>% tidy()
```

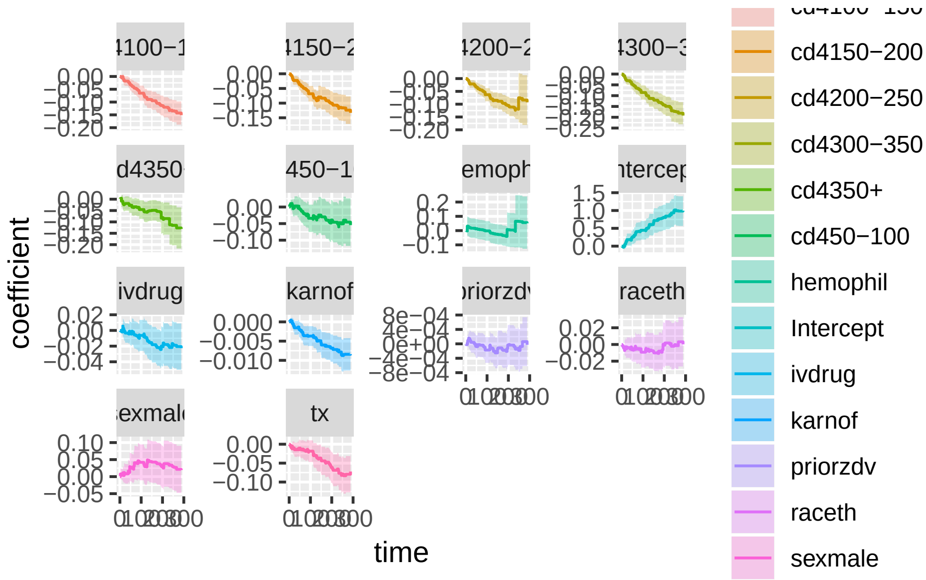
```
## # A tibble: 3 x 7
##   term          estimate std.error statistic p.value conf.low conf.high
##   <chr>            <dbl>     <dbl>     <dbl>   <dbl>    <dbl>     <dbl>
## 1 ivdrug        -0.711      1.71      -0.416 0.678    -4.07      2.64
## 2 karnof        -0.0903     0.0294    -3.07  0.00214  -0.148    -0.0326
## 3 ivdrug:karnof  0.00573    0.0201     0.285 0.775    -0.0336    0.0451
```

```
coxph(Surv(time,censor)~sex+tx+age+txgrp, data = aids) %>% tidy()
```

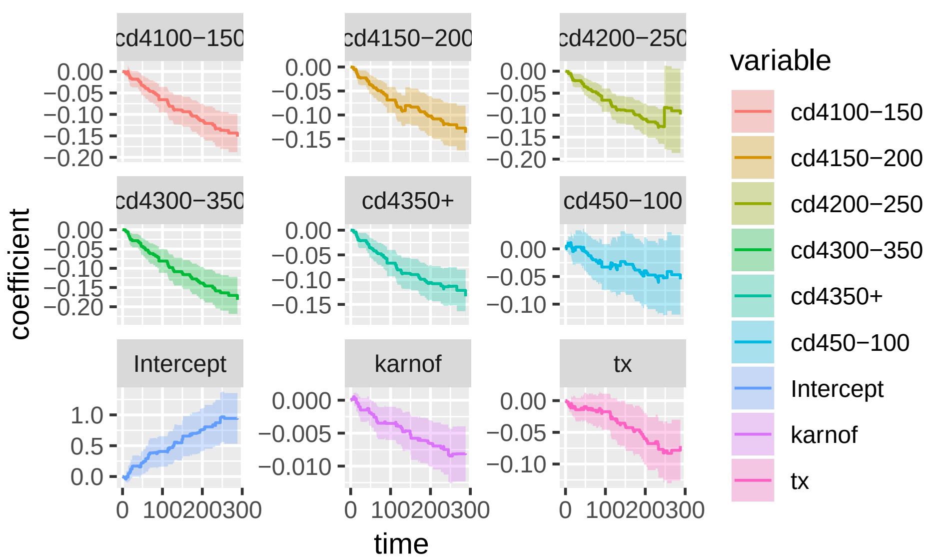
```
## # A tibble: 9 x 7
##   term      estimate std.error statistic  p.value conf.low conf.high
##   <chr>        <dbl>     <dbl>     <dbl>    <dbl>    <dbl>     <dbl>
## 1 sexmale     0.302      0.324   0.931    0.352     -0.333     0.937
## 2 tx         -0.790      0.256  -3.08     0.00205   -1.29     -0.288
## 3 age20-30   -0.424      1.07   -0.396    0.692     -2.52      1.67
## 4 age30-40   -0.214      1.02   -0.209    0.834     -2.21      1.79
## 5 age40-50   -0.0490     1.03   -0.0475   0.962     -2.07      1.97
## 6 age50-60    0.639      1.05    0.611    0.541     -1.41      2.69
## 7 age60-70   -0.328      1.42   -0.231    0.817     -3.11      2.46
## 8 ageover70 -14.1     2672.     -0.00528  0.996   -Inf       Inf
## 9 txgrp      NA          0       NA      NA         NA        NA
```

To better understand how nmuch the model fit changes with each different explanatory variable, we can use the graphical representation of the Aalen additive regression model. The Aalen model allows for time-varying covariate effects, while the Cox model allows only a common time-dependence through the baseline. In the Aalen model, we have the weighted comparisons of the crude estimate of the hazard rate of each group as compared to a baseline group, which here is defined as the estimate. As we can see, the selected explanatory variables in our model all have an inverse coefficient correlation with the baseline interceipt. The slope of an estimated cumulative regression function is positive when covariate increases and this fact correspond to an increasing hazard rate. On the other hand, if the slope is negative while the covariate increases, then this fact points to a decreasing hazard rate.

```r
library(ggfortify)

aa_fit <-aareg(Surv(time, censor) ~ cd4 + karnof+ priorzdv +hemophil +raceth +sex +tx +ivdrug, data = aids

autoplot(aa_fit,xlab="Coefficient", ylab="Time") + labs(x = "time", y = "coefficient")
```




```r
aa_fit2 <-aareg(Surv(time, censor) ~ cd4 + karnof+ tx , data = aids)

autoplot(aa_fit2) + labs(x = "time", y = "coefficient")
```



The Aalen model assumes that the cumulative hazard $H(t)$ for a subject can be expressed as $a(t)$ + X $B(t)$, where $a(t)$ is a time-dependent intercept term, X is the vector of covariates for the subject possibly time-dependent, and $B(t)$ is a time-dependent matrix of coefficients.

The plots show how the effects of the covariates change over time.

Patricia's "Something New"

I will be doing a power analysis by simulating survival analysis curves

1. What is the topic?

The topic is using sim.survdata in R to simulate survival data. Using that simulated data, we will make that the alternative and control for the coefficient beta by setting it equal to some value. Then using power analysis, we will see how many times we reject H_0.

2. How it is relevant? How it relates to survival analysis/analysis at hand?

Power analysis relates to survival analysis because if power is large after comparing our data to the simulated survival data, this tells us that there is a high chance that we would reject the null in favor of the alternative (control versus treatment?)

3. Resources to learn about the topic.

Below are some of the resources I have begun to use to learn about creating simulations of survival curves and performing power analysis:

a). https://cran.r-project.org/web/packages/coxed/vignettes/simulating_survival_data.html b). http://www.icssc.org/documents/advbiosgoa/tab%2026.00_survss.pdf

4. What will be challenging about learning something new?

Learning something new will be challenging because in this case, the concept of power analysis is something I just recently learned in Intro to Statistics. So learning to apply this concept in the context of survival analysis curves will be a challenge for me to learn. Learning how to simulate survival curves will also be challenging because I will have to learn how to use and interpret new functions in R.

Power Analysis code and simulation

```
simdata <- sim.survdata(N=1000, T=100, num.data.frames=1, beta = c(0.01,0.07,0.3))
head(simdata$data,10)
```

```
##             X1          X2          X3   y failed
## 1    1.8676723  1.24243351  0.06697356  70   TRUE
## 2   -0.4558316 -0.48072375  1.10672735   1   TRUE
## 3   -0.9415616 -0.05350077  0.69266665 100   TRUE
## 4   -1.2418018 -0.53717826 -1.84427787  73   TRUE
## 5    0.5513832 -1.90691481 -1.10560727  30   TRUE
## 6    0.6919124 -0.43256368  0.49735428  72   TRUE
## 7   -1.3167190 -0.46666253  1.73236153  68   TRUE
## 8    1.2012010 -1.22676536 -1.35483363  70   TRUE
## 9   -1.3159905 -0.20522213  0.05475034  69   TRUE
## 10 -0.4486984 -0.66101274 -0.48231292  30   TRUE
```

```
simdata$betas
```

```
##       [,1]
## [1,] 0.01
## [2,] 0.07
## [3,] 0.30
```

```r
head(simdata$baseline,10)
```

```
##    time  failure.PDF failure.CDF  survivor       hazard
## 1     1 0.0353157674  0.03531577 0.9646842 0.0353157674
## 2     2 0.0195015771  0.05481734 0.9451827 0.0202155031
## 3     3 0.0015243495  0.05634169 0.9436583 0.0016127565
## 4     4 0.0012480324  0.05758973 0.9424103 0.0013225469
## 5     5 0.0009993469  0.05858907 0.9414109 0.0010604160
## 6     6 0.0007782932  0.05936737 0.9406326 0.0008267306
## 7     7 0.0005848713  0.05995224 0.9400478 0.0006217850
## 8     8 0.0004190810  0.06037132 0.9396287 0.0004458082
## 9     9 0.0002809224  0.06065224 0.9393478 0.0002989717
## 10   10 0.0001703956  0.06082264 0.9391774 0.0001813977
```

```r
#ggsurvplot(survfit(Surv(y,failed) ~ X1 + X2 + X3, data = simdata$data))
model <- coxph(Surv(y, failed) ~ X1 + X2 + X3, data = simdata$data)

library(dplyr)
library(broom)
model %>% tidy()
```

```
## # A tibble: 3 x 7
##   term  estimate std.error statistic   p.value conf.low conf.high
##   <chr>    <dbl>     <dbl>     <dbl>     <dbl>    <dbl>     <dbl>
## 1 X1      0.0347    0.0334     1.04  0.299      -0.0307    0.100
## 2 X2      0.0146    0.0312     0.468 0.640      -0.0465    0.0758
## 3 X3      0.143     0.0340     4.20  0.0000265   0.0762    0.209
```

```r
set.seed(1234)
n.reps <- 100
simoutput <- c()
for(i in 1:n.reps){
  simdata <- sim.survdata(N=851, T=100, num.data.frames=1, xvars=2,beta = c(-0.058666,-0.015140))
  model <- coxph(Surv(y, failed) ~ X1 + X2, data = simdata$data)
  simoutput <- rbind(simoutput, cbind(rep = rep(i, 2), model %>% tidy()))

}

simoutput
```

```
##     rep term      estimate  std.error    statistic      p.value
## 1     1   X1 -3.080231e-02 0.03698786 -0.832767990 4.049756e-01
## 2     1   X2  1.066434e-03 0.03874869  0.027521797 9.780436e-01
## 3     2   X1 -6.231011e-02 0.03669308 -1.698143116 8.948075e-02
## 4     2   X2  3.813262e-02 0.03651322  1.044351001 2.963230e-01
## 5     3   X1 -2.936590e-02 0.03555698 -0.825882913 4.088705e-01
## 6     3   X2  5.716814e-03 0.03801301  0.150390973 8.804562e-01
## 7     4   X1 -6.646417e-02 0.03676089 -1.808013221 7.060444e-02
## 8     4   X2 -3.475956e-02 0.03589036 -0.968492756 3.327983e-01
## 9     5   X1 -6.056747e-02 0.03593891 -1.685289669 9.193270e-02
## 10    5   X2 -1.134861e-03 0.03556232 -0.031911889 9.745423e-01
## 11    6   X1 -4.093736e-02 0.03467662 -1.180546410 2.377830e-01
## 12    6   X2 -3.010143e-02 0.03567951 -0.843661508 3.988586e-01
## 13    7   X1 -5.380684e-02 0.03758346 -1.431662698 1.522404e-01
## 14    7   X2  7.248870e-03 0.03560391  0.203597617 8.386680e-01
## 15    8   X1 -1.573089e-02 0.03658278 -0.430008097 6.671898e-01
## 16    8   X2 -2.915203e-02 0.03611028 -0.807305405 4.194906e-01
## 17    9   X1 -5.047657e-03 0.03530936 -0.142955238 8.863255e-01
```

```
## 18    9   X2  4.814201e-02 0.03540877  1.359607250 1.739542e-01
## 19   10   X1 -3.256088e-02 0.03626670 -0.897817467 3.692829e-01
## 20   10   X2 -6.372910e-02 0.03541198 -1.799648208 7.191620e-02
## 21   11   X1  1.800873e-02 0.03692957  0.487650802 6.257972e-01
## 22   11   X2  6.744645e-03 0.03781808  0.178344478 8.584525e-01
## 23   12   X1 -1.000106e-01 0.03499724 -2.857669684 4.267644e-03
## 24   12   X2  2.381930e-02 0.03557772  0.669500498 5.031763e-01
## 25   13   X1 -3.765072e-02 0.03549326 -1.060785070 2.887876e-01
## 26   13   X2 -1.165382e-01 0.03662818 -3.181655293 1.464360e-03
## 27   14   X1 -4.046525e-02 0.03454971 -1.171218365 2.415110e-01
## 28   14   X2  2.899444e-02 0.03639968  0.796557468 4.257081e-01
## 29   15   X1 -5.911003e-02 0.03589099 -1.646932236 9.957197e-02
## 30   15   X2  1.961235e-02 0.03627130  0.540712534 5.887057e-01
## 31   16   X1 -5.069811e-02 0.03672001 -1.380667236 1.673813e-01
## 32   16   X2 -8.283286e-03 0.03785896 -0.218793292 8.268111e-01
## 33   17   X1 -4.563515e-02 0.03613469 -1.262918053 2.066186e-01
## 34   17   X2  1.231324e-02 0.03501290  0.351677257 7.250803e-01
## 35   18   X1 -1.057966e-01 0.03593094 -2.944443091 3.235364e-03
## 36   18   X2 -1.304522e-03 0.03942300 -0.033090374 9.736025e-01
## 37   19   X1 -7.138296e-02 0.03578789 -1.994612332 4.608518e-02
## 38   19   X2 -4.100998e-02 0.03570599 -1.148546257 2.507431e-01
## 39   20   X1 -8.139002e-02 0.03483776 -2.336258698 1.947776e-02
## 40   20   X2 -2.326160e-02 0.03545956 -0.656003767 5.118217e-01
## 41   21   X1 -3.792544e-02 0.03448879 -1.099645296 2.714867e-01
## 42   21   X2 -7.483697e-02 0.03543068 -2.112208403 3.466858e-02
## 43   22   X1 -1.799897e-02 0.03653650 -0.492629736 6.222742e-01
## 44   22   X2  9.424652e-03 0.03637984  0.259062461 7.955870e-01
## 45   23   X1 -4.471464e-02 0.03707767 -1.205972279 2.278282e-01
## 46   23   X2  2.672436e-03 0.03692515  0.072374413 9.423039e-01
## 47   24   X1 -1.142369e-01 0.03711164 -3.078196337 2.082577e-03
## 48   24   X2 -2.645912e-02 0.03656469 -0.723624913 4.692960e-01
## 49   25   X1 -8.838472e-03 0.03570041 -0.247573421 8.044645e-01
## 50   25   X2  4.896290e-02 0.03410790  1.435529818 1.511362e-01
## 51   26   X1 -1.057080e-01 0.03734521 -2.830564080 4.646600e-03
## 52   26   X2 -2.319656e-02 0.03750655 -0.618466930 5.362676e-01
## 53   27   X1 -5.685230e-02 0.03525417 -1.612640453 1.068226e-01
## 54   27   X2 -1.498365e-02 0.03433128 -0.436443159 6.625152e-01
## 55   28   X1  1.640367e-02 0.03735524  0.439126396 6.605700e-01
## 56   28   X2 -5.648880e-02 0.03549621 -1.591403967 1.115187e-01
## 57   29   X1  2.355667e-03 0.03527108  0.066787486 9.467509e-01
## 58   29   X2  5.744676e-02 0.03574311  1.607212087 1.080079e-01
## 59   30   X1 -6.185840e-02 0.03521468 -1.756608292 7.898460e-02
## 60   30   X2  4.665840e-02 0.03544273  1.316444820 1.880248e-01
## 61   31   X1 -1.056293e-01 0.03705663 -2.850482387 4.365297e-03
## 62   31   X2 -6.923403e-02 0.03928703 -1.762261572 7.802511e-02
## 63   32   X1 -1.877259e-02 0.03906026 -0.480605875 6.307966e-01
## 64   32   X2  7.244731e-03 0.03655606  0.198181376 8.429032e-01
## 65   33   X1 -6.576289e-02 0.03557666 -1.848484390 6.453230e-02
## 66   33   X2 -2.512402e-02 0.03576905 -0.702395624 4.824325e-01
## 67   34   X1 -8.914686e-02 0.03642262 -2.447568766 1.438237e-02
## 68   34   X2 -2.956608e-02 0.03682959 -0.802780657 4.221015e-01
## 69   35   X1 -5.276051e-02 0.03722538 -1.417326151 1.563876e-01
## 70   35   X2 -2.521038e-02 0.03717850 -0.678090319 4.977144e-01
## 71   36   X1 -8.694290e-02 0.03632656 -2.393370263 1.669438e-02
## 72   36   X2 -5.362627e-02 0.03639277 -1.473541958 1.406050e-01
## 73   37   X1 -2.183847e-02 0.03707865 -0.588976969 5.558767e-01
```

```
## 74   37   X2  9.244649e-03 0.03623776  0.255110961 7.986374e-01
## 75   38   X1 -7.786432e-02 0.03565127 -2.184054444 2.895825e-02
## 76   38   X2 -4.421072e-02 0.03629045 -1.218246832 2.231302e-01
## 77   39   X1 -4.632116e-02 0.03665276 -1.263783568 2.063077e-01
## 78   39   X2 -9.453473e-03 0.03830319 -0.246806441 8.050580e-01
## 79   40   X1 -3.638083e-02 0.03724658 -0.976756172 3.286899e-01
## 80   40   X2 -2.040805e-02 0.03549343 -0.574981031 5.653041e-01
## 81   41   X1 -6.334291e-02 0.03602343 -1.758381116 7.868268e-02
## 82   41   X2 -6.921611e-02 0.03509721 -1.972125762 4.859525e-02
## 83   42   X1  1.802987e-04 0.03671283  0.004911055 9.960816e-01
## 84   42   X2 -1.930876e-02 0.03757431 -0.513882081 6.073344e-01
## 85   43   X1 -5.052283e-02 0.03809801 -1.326127969 1.847973e-01
## 86   43   X2 -7.579427e-02 0.03558632 -2.129870891 3.318227e-02
## 87   44   X1 -7.548738e-02 0.03527618 -2.139896648 3.236312e-02
## 88   44   X2 -3.456986e-02 0.03612552 -0.956937308 3.385989e-01
## 89   45   X1  3.084174e-02 0.03423965  0.900760809 3.677155e-01
## 90   45   X2 -5.418818e-02 0.03560192 -1.522057545 1.279947e-01
## 91   46   X1 -1.018527e-01 0.03769951 -2.701697288 6.898654e-03
## 92   46   X2 -4.890103e-03 0.03662540 -0.133516716 8.937847e-01
## 93   47   X1 -8.520296e-02 0.03752713 -2.270436388 2.318112e-02
## 94   47   X2  3.610810e-02 0.03546209  1.018217028 3.085748e-01
## 95   48   X1 -2.906080e-02 0.03728146 -0.779497283 4.356868e-01
## 96   48   X2  2.283651e-02 0.03511308  0.650370318 5.154530e-01
## 97   49   X1 -8.842741e-02 0.03835545 -2.305471883 2.114016e-02
## 98   49   X2 -2.190517e-02 0.03719187 -0.588977466 5.558764e-01
## 99   50   X1 -5.974510e-02 0.03767678 -1.585727496 1.128011e-01
## 100  50   X2 -6.499511e-02 0.03637419 -1.786846827 7.396225e-02
## 101  51   X1 -4.020783e-02 0.03795873 -1.059251203 2.894854e-01
## 102  51   X2  7.409340e-02 0.03620525  2.046482040 4.070897e-02
## 103  52   X1 -6.318960e-02 0.03478679 -1.816482546 6.929638e-02
## 104  52   X2 -2.207501e-02 0.03651096 -0.604613333 5.454359e-01
## 105  53   X1 -5.227548e-02 0.03681688 -1.419878109 1.556432e-01
## 106  53   X2  4.902645e-03 0.03657769  0.134033771 8.933759e-01
## 107  54   X1  1.003317e-02 0.03735998  0.268553946 7.882730e-01
## 108  54   X2  2.330834e-02 0.03563178  0.654144760 5.130185e-01
## 109  55   X1 -7.003834e-02 0.03423092 -2.046054919 4.075097e-02
## 110  55   X2  2.885273e-02 0.03942551  0.731828946 4.642730e-01
## 111  56   X1 -5.751209e-02 0.03581104 -1.605987493 1.082767e-01
## 112  56   X2 -1.039759e-01 0.03834928 -2.711286846 6.702262e-03
## 113  57   X1 -6.475033e-02 0.03723026 -1.739185856 8.200208e-02
## 114  57   X2 -6.250097e-02 0.03702659 -1.688002160 9.141081e-02
## 115  58   X1 -8.592702e-03 0.03634645 -0.236411038 8.131137e-01
## 116  58   X2 -1.278233e-03 0.03649247 -0.035027296 9.720580e-01
## 117  59   X1 -3.642829e-02 0.03676996 -0.990707764 3.218283e-01
## 118  59   X2  5.205944e-04 0.03729703  0.013958065 9.888634e-01
## 119  60   X1  1.522112e-03 0.03832937  0.039711383 9.683232e-01
## 120  60   X2  1.119350e-02 0.03780201  0.296108656 7.671471e-01
## 121  61   X1 -3.376598e-02 0.03529062 -0.956797448 3.386695e-01
## 122  61   X2 -2.659074e-02 0.03478086 -0.764522071 4.445562e-01
## 123  62   X1 -1.003609e-01 0.03731212 -2.689765626 7.150222e-03
## 124  62   X2 -1.633653e-02 0.03732927 -0.437633165 6.616522e-01
## 125  63   X1 -6.952446e-02 0.03399719 -2.045005720 4.085429e-02
## 126  63   X2 -1.454377e-02 0.03441021 -0.422658677 6.725443e-01
## 127  64   X1 -3.275114e-04 0.03570896 -0.009171685 9.926822e-01
## 128  64   X2  7.293542e-03 0.03683797  0.197989808 8.430530e-01
## 129  65   X1 -2.863590e-02 0.03421151 -0.837025271 4.025784e-01
```

```
## 130  65   X2  5.567759e-02 0.03807521  1.462305546 1.436575e-01
## 131  66   X1  1.878338e-02 0.03565976  0.526738912 5.983749e-01
## 132  66   X2 -2.481082e-02 0.03566238 -0.695714272 4.866078e-01
## 133  67   X1 -9.077083e-02 0.03511369 -2.585055148 9.736342e-03
## 134  67   X2 -5.158984e-02 0.03643129 -1.416086066 1.567503e-01
## 135  68   X1  1.566573e-02 0.03782770  0.414133787 6.787761e-01
## 136  68   X2  3.735936e-03 0.03680882  0.101495698 9.191570e-01
## 137  69   X1 -2.028856e-02 0.03562242 -0.569544605 5.689866e-01
## 138  69   X2  2.453084e-02 0.03677380  0.667073826 5.047250e-01
## 139  70   X1 -2.406389e-02 0.03519404 -0.683748985 4.941337e-01
## 140  70   X2  4.020638e-02 0.03602180  1.116167894 2.643503e-01
## 141  71   X1 -1.155043e-01 0.03648636 -3.165684881 1.547183e-03
## 142  71   X2 -2.710149e-03 0.03488667 -0.077684367 9.380791e-01
## 143  72   X1  1.709765e-03 0.03564939  0.047960586 9.617477e-01
## 144  72   X2  1.911462e-03 0.03783233  0.050524577 9.597044e-01
## 145  73   X1 -9.809185e-02 0.03594931 -2.728615814 6.360075e-03
## 146  73   X2 -1.386936e-02 0.03411374 -0.406562160 6.843296e-01
## 147  74   X1 -7.739879e-02 0.03559845 -2.174217802 2.968878e-02
## 148  74   X2 -3.858319e-02 0.03704642 -1.041482213 2.976518e-01
## 149  75   X1 -1.762760e-03 0.03464773 -0.050876642 9.594238e-01
## 150  75   X2 -1.323995e-02 0.03688343 -0.358967422 7.196195e-01
## 151  76   X1 -3.772914e-02 0.03581286 -1.053507993 2.921082e-01
## 152  76   X2  3.226389e-02 0.03524507  0.915415649 3.599735e-01
## 153  77   X1 -8.293043e-02 0.03610254 -2.297079996 2.161421e-02
## 154  77   X2 -2.792196e-02 0.03452178 -0.808821586 4.186178e-01
## 155  78   X1 -6.975557e-02 0.03548690 -1.965670829 4.933665e-02
## 156  78   X2  1.086850e-02 0.03660476  0.296915025 7.665314e-01
## 157  79   X1 -1.307483e-02 0.03540944 -0.369247004 7.119436e-01
## 158  79   X2 -5.696438e-02 0.03602848 -1.581093139 1.138567e-01
## 159  80   X1 -6.309239e-02 0.03620861 -1.742469312 8.142636e-02
## 160  80   X2  5.617490e-03 0.03544641  0.158478366 8.740799e-01
## 161  81   X1 -3.665308e-02 0.03583397 -1.022858498 3.063748e-01
## 162  81   X2 -6.631743e-03 0.03567456 -0.185895556 8.525267e-01
## 163  82   X1 -5.644857e-02 0.03678069 -1.534733663 1.248493e-01
## 164  82   X2 -4.926207e-02 0.03581379 -1.375505752 1.689747e-01
## 165  83   X1 -5.895693e-02 0.03842937 -1.534163076 1.249895e-01
## 166  83   X2 -8.169088e-05 0.03853861 -0.002119715 9.983087e-01
## 167  84   X1 -7.020820e-02 0.03764587 -1.864964172 6.218643e-02
## 168  84   X2  8.784007e-03 0.03751347  0.234156087 8.148638e-01
## 169  85   X1 -1.466704e-02 0.03232981 -0.453669252 6.500669e-01
## 170  85   X2 -1.409855e-01 0.03570011 -3.949161930 7.842529e-05
## 171  86   X1 -2.958249e-02 0.03818998 -0.774613735 4.385679e-01
## 172  86   X2 -7.474306e-02 0.03562519 -2.098039388 3.590167e-02
## 173  87   X1 -5.465722e-02 0.03532603 -1.547222187 1.218097e-01
## 174  87   X2 -2.070036e-02 0.03648262 -0.567403427 5.704401e-01
## 175  88   X1 -6.315533e-02 0.03486020 -1.811674072 7.003657e-02
## 176  88   X2 -6.589827e-02 0.03558566 -1.851820744 6.405157e-02
## 177  89   X1  3.316564e-03 0.03516575  0.094312342 9.248610e-01
## 178  89   X2  4.932768e-02 0.03639212  1.355449347 1.752744e-01
## 179  90   X1  6.655681e-03 0.03790255  0.175599810 8.606084e-01
## 180  90   X2 -2.227758e-02 0.03588099 -0.620874267 5.346824e-01
## 181  91   X1 -4.433805e-02 0.03722575 -1.191058851 2.336305e-01
## 182  91   X2  2.470171e-02 0.03559615  0.693943134 4.877179e-01
## 183  92   X1 -2.450619e-02 0.03631060 -0.674904428 4.997365e-01
## 184  92   X2 -1.991391e-04 0.03687391 -0.005400543 9.956910e-01
## 185  93   X1 -7.462906e-02 0.03622628 -2.060080915 3.939081e-02
```

```
## 186  93   X2 -2.017764e-02 0.03668756 -0.549985732 5.823292e-01
## 187  94   X1 -9.114038e-02 0.03567458 -2.554771021 1.062577e-02
## 188  94   X2 -2.783179e-02 0.03590984 -0.775046330 4.383123e-01
## 189  95   X1 -7.411678e-02 0.03771190 -1.965342247 4.937464e-02
## 190  95   X2 -2.454443e-02 0.03768623 -0.651283695 5.148634e-01
## 191  96   X1 -6.704035e-02 0.03491517 -1.920092006 5.484628e-02
## 192  96   X2  5.291841e-02 0.03649199  1.450137865 1.470201e-01
## 193  97   X1 -1.346551e-01 0.03474430 -3.875602604 1.063611e-04
## 194  97   X2  3.959345e-02 0.03548847  1.115670815 2.645631e-01
## 195  98   X1 -1.813821e-02 0.03554822 -0.510242452 6.098816e-01
## 196  98   X2 -6.412166e-02 0.03636097 -1.763474863 7.782043e-02
## 197  99   X1 -4.857148e-02 0.03681186 -1.319451654 1.870182e-01
## 198  99   X2 -4.055132e-02 0.03589035 -1.129866733 2.585324e-01
## 199 100   X1 -1.627868e-01 0.03821047 -4.260267849 2.041821e-05
## 200 100   X2  2.910721e-02 0.03451161  0.843403398 3.990029e-01
##         conf.low     conf.high
## 1   -0.103297181  0.0416925682
## 2   -0.074879607  0.0770124742
## 3   -0.134227228  0.0096070149
## 4   -0.033431979  0.1096972155
## 5   -0.099056292  0.0403244936
## 6   -0.068787321  0.0802209491
## 7   -0.138514182  0.0055858448
## 8   -0.105103377  0.0355842630
## 9   -0.131006427  0.0098714943
## 10  -0.070835724  0.0685660026
## 11  -0.108902295  0.0270275695
## 12  -0.100031983  0.0398291251
## 13  -0.127469059  0.0198553893
## 14  -0.062533502  0.0770312428
## 15  -0.087431818  0.0559700366
## 16  -0.099926882  0.0416228281
## 17  -0.074252725  0.0641574099
## 18  -0.021257891  0.1175419182
## 19  -0.103642308  0.0385205511
## 20  -0.133135298  0.0056770985
## 21  -0.054371893  0.0903893616
## 22  -0.067377424  0.0808667147
## 23  -0.168603891 -0.0314172235
## 24  -0.045911743  0.0935503402
## 25  -0.107216231  0.0319147908
## 26  -0.188328136 -0.0447483243
## 27  -0.108181437  0.0272509311
## 28  -0.042347627  0.1003365037
## 29  -0.129455076  0.0112350193
## 30  -0.051478097  0.0907027907
## 31  -0.122668008  0.0212717817
## 32  -0.082485480  0.0659189077
## 33  -0.116457848  0.0251875398
## 34  -0.056310781  0.0809372621
## 35  -0.176219975 -0.0353732634
## 36  -0.078572184  0.0759631401
## 37  -0.141525926 -0.0012399911
## 38  -0.110992425  0.0289724705
## 39  -0.149670771 -0.0131092646
## 40  -0.092761053  0.0462378493
```

```
## 41  -0.105522226  0.0296713516
## 42  -0.144279818 -0.0053941226
## 43  -0.089609192  0.0536112585
## 44  -0.061878532  0.0807278353
## 45  -0.117385546  0.0279562570
## 46  -0.069699522  0.0750443933
## 47  -0.186974400 -0.0414994383
## 48  -0.098124607  0.0452063598
## 49  -0.078809983  0.0611330392
## 50  -0.017887346  0.1158131483
## 51  -0.178903257 -0.0325127399
## 52  -0.096708040  0.0503149222
## 53  -0.125949209  0.0122446033
## 54  -0.082271722  0.0523044183
## 55  -0.056811258  0.0896186052
## 56  -0.126060091  0.0130824826
## 57  -0.066774373  0.0714857058
## 58  -0.012608451  0.1275019794
## 59  -0.130877910  0.0071611060
## 60  -0.022808078  0.1161248815
## 61  -0.178258908 -0.0329996066
## 62  -0.146235200  0.0077671413
## 63  -0.095329289  0.0577841097
## 64  -0.064403838  0.0788933003
## 65  -0.135491854  0.0039660711
## 66  -0.095230064  0.0449820207
## 67  -0.160533879 -0.0177598424
## 68  -0.101750756  0.0426185890
## 69  -0.125720925  0.0201999022
## 70  -0.098078914  0.0476581460
## 71  -0.158141647 -0.0157441581
## 72  -0.124954782  0.0177022435
## 73  -0.094511296  0.0508343511
## 74  -0.061780052  0.0802693501
## 75  -0.147739530 -0.0079891100
## 76  -0.115338687  0.0269172460
## 77  -0.118159249  0.0255169344
## 78  -0.084526339  0.0656193929
## 79  -0.109382779  0.0366211272
## 80  -0.089973904  0.0491578015
## 81  -0.133947535  0.0072617059
## 82  -0.138005367 -0.0004268444
## 83  -0.071775530  0.0721361278
## 84  -0.092953054  0.0543355271
## 85  -0.125193552  0.0241478885
## 86  -0.145542171 -0.0060463615
## 87  -0.144627419 -0.0063473369
## 88  -0.105374575  0.0362348598
## 89  -0.036266748  0.0979502198
## 90  -0.123966668  0.0155903121
## 91  -0.175742348 -0.0279629824
## 92  -0.076674559  0.0668943542
## 93  -0.158754773 -0.0116511374
## 94  -0.033396316  0.1056125255
## 95  -0.102131123  0.0440095252
## 96  -0.045983869  0.0916568814
```

```
## 97  -0.163602708 -0.0132521106
## 98  -0.094799896  0.0509895505
## 99  -0.133590229  0.0141000249
## 100 -0.136287211  0.0062969965
## 101 -0.114605581  0.0341899153
## 102  0.003132408  0.1450543855
## 103 -0.131370453  0.0049912587
## 104 -0.093635169  0.0494851478
## 105 -0.124435224  0.0198842743
## 106 -0.066788303  0.0765935935
## 107 -0.063191045  0.0832573848
## 108 -0.046528667  0.0931453552
## 109 -0.137129706 -0.0029469718
## 110 -0.048419856  0.1061253209
## 111 -0.127700437  0.0126762669
## 112 -0.179139094 -0.0288126888
## 113 -0.137720294  0.0082196261
## 114 -0.135071764  0.0100698203
## 115 -0.079830439  0.0626450339
## 116 -0.072802159  0.0702456935
## 117 -0.108496095  0.0356395166
## 118 -0.072580248  0.0736214363
## 119 -0.073602071  0.0766462951
## 120 -0.062897075  0.0852840789
## 121 -0.102934319  0.0354023691
## 122 -0.094759968  0.0415784977
## 123 -0.173491258 -0.0272304447
## 124 -0.089500562  0.0568275049
## 125 -0.136157734 -0.0028911804
## 126 -0.081986543  0.0528989966
## 127 -0.070315793  0.0696607703
## 128 -0.064907550  0.0794946346
## 129 -0.095689232  0.0384174313
## 130 -0.018948450  0.1303036319
## 131 -0.051108465  0.0886752344
## 132 -0.094707800  0.0450861501
## 133 -0.159592405 -0.0219492590
## 134 -0.122993853  0.0198141732
## 135 -0.058475199  0.0898066548
## 136 -0.068408017  0.0758798899
## 137 -0.090107223  0.0495301064
## 138 -0.047544484  0.0966061627
## 139 -0.093042934  0.0449151588
## 140 -0.030395053  0.1108078030
## 141 -0.187016276 -0.0439923682
## 142 -0.071086756  0.0656664587
## 143 -0.068161748  0.0715812791
## 144 -0.072238542  0.0760614673
## 145 -0.168551199 -0.0276325013
## 146 -0.080731055  0.0529923440
## 147 -0.147170468 -0.0076271040
## 148 -0.111192851  0.0340264661
## 149 -0.069671060  0.0661455402
## 150 -0.085530136  0.0590502392
## 151 -0.107921053  0.0324627824
## 152 -0.036815182  0.1013429633
```

```
## 153 -0.153690109 -0.0121707449
## 154 -0.095583418  0.0397394896
## 155 -0.139308626 -0.0002025182
## 156 -0.060875509  0.0826125160
## 157 -0.082476056  0.0563263970
## 158 -0.127578908  0.0136501409
## 159 -0.134059959  0.0078751795
## 160 -0.063856202  0.0750911812
## 161 -0.106886374  0.0335802108
## 162 -0.076552600  0.0632891151
## 163 -0.128537407  0.0156402666
## 164 -0.119455797  0.0209316616
## 165 -0.134277112  0.0163632619
## 166 -0.075615979  0.0754525970
## 167 -0.143992753  0.0035763507
## 168 -0.064741044  0.0823090591
## 169 -0.078032293  0.0486982154
## 170 -0.210956425 -0.0710145788
## 171 -0.104433476  0.0452685055
## 172 -0.144567154 -0.0049189629
## 173 -0.123894979  0.0145805310
## 174 -0.092204985  0.0508042577
## 175 -0.131480072  0.0051694166
## 176 -0.135644884  0.0038483488
## 177 -0.065607036  0.0722401644
## 178 -0.021999571  0.1206549321
## 179 -0.067631952  0.0809433135
## 180 -0.092603036  0.0480478670
## 181 -0.117299174  0.0286230667
## 182 -0.045065473  0.0944688853
## 183 -0.095673657  0.0466612856
## 184 -0.072470671  0.0720723925
## 185 -0.145631257 -0.0036268636
## 186 -0.092083942  0.0517286680
## 187 -0.161061265 -0.0212194899
## 188 -0.098213777  0.0425502009
## 189 -0.148030738 -0.0002028245
## 190 -0.098408080  0.0493192259
## 191 -0.135472831  0.0013921371
## 192 -0.018604568  0.1244413926
## 193 -0.202752667 -0.0665575198
## 194 -0.029962669  0.1091495605
## 195 -0.087811439  0.0515350182
## 196 -0.135387845  0.0071445349
## 197 -0.120721403  0.0235784529
## 198 -0.110895114  0.0297924829
## 199 -0.237677995 -0.0878956966
## 200 -0.038534302  0.0967487184
```

```r
#sum(which(simoutput$p.value < 0.05))
sum(simoutput$p.value <0.05)
```

```
## [1] 36
```

```r
#simoutput%>%filter(term=="X1")%>%summarize(sum(p.value<0.05))

simoutput%>%dplyr::filter(term=="X1")%>%dplyr::summarize(sum(p.value<0.05))
```

```
##   sum(p.value < 0.05)
## 1                  28
```

```r
simoutput%>%dplyr::filter(term=="X2")%>%dplyr::summarize(sum(p.value<0.05))
```

```
##   sum(p.value < 0.05)
## 1                   8
```

```r
simoutput%>%dplyr::filter(term=="X3")%>%dplyr::summarize(sum(p.value<0.05))
```

```
##   sum(p.value < 0.05)
## 1                   0
```

```r
simoutput%>%dplyr::filter(term=="X4")%>%dplyr::summarize(sum(p.value<0.05))
```

```
##   sum(p.value < 0.05)
## 1                   0
```

Juste's "Something New": The Schoenfeld Residuals for the Cox PH model

Cox proportional hazards (PH) model is considered a great way to identify combined effects of several covariates on the relative risk (hazard). This model assumes that the hazards of the different strata formed by the levels of the covariates are proportional. This proportional hazards assumption is particularly important and can be tested via three different classes of tests. The first class is focused on the piece-wise estimation of models for subsets of data defined by stratification of time. The second one considers the interactions between covariates and some function of time. Final, third one is based on examinations of regression residuals. The Schoenfeld Residuals are a part of the third class of proportional hazard assumption testing and I will be exploring it in order to be able to eradicate a method for testing for the PH assumption in the current and future data set analyses. This topic is particularly important in relation to survival analysis since it provides an idea of whether the model is appropriate for the data set at hand and whether some covariates should be considered as variants of time in order to supply the best model for prediction of proportional hazards. Taking a completely new model of analyzing survival data is particularly difficult since the mathematical derivations and notations are also very varied from what we have seen in class. Although, I do remember some of the ideas behind parametric functions, their applications to statistical models are much more challenging than I have expected. Therefore, it will require me a lot of time and extensive research to be able to understand and learn how to apply this model to our data and other instances of survival analysis.

3. Resources to learn about the topic.

I have been researching articles and scientific journals that provide insights into the Schoenfeld residuals and their use in the Cox PH model. Sources include:

1. https://onlinelibrary.wiley.com/doi/full/10.1111/ajps.12176
2. https://rstudio-pubs-static.s3.amazonaws.com/39354_34153ff19e624116bd2fbdec7d2534aa.html

Explanation of the Theory Behind Schoenfeld Residuals

Let $z_{ij}(t)$ be the j^{th} covariate of the i^{th} unit, where $i = 1, 2, ..., n$ and $j = 1, 2..., p$

This notation indicates that z_{ij} is allowed to vary as a function of the time scale.

1) As we know from lecture, the Cox PH model assumes that $h(t)$ of the i^{th} individual satisfies:

- $h_i(t) = h_0(t)e^{z_i(t)\beta}$ where:
- h_0 -> baseline hazard
- $z_i(t)$ -> 1 x p vector of covariates for unit i each of which can be time fixed or time-varying.

2) However, another possibility has been presented by Therneau and Granbsh in 2000, where they proposed an idea that there could be an alternative to the current Cox model, where the coefficient of the estimate could also be varying as a function of time.

   The new hazard function would look like this: $h_i(t) = h_0(t)e^{z_i(t)\beta(t)}$

Therefore, in order to examine thee two models in a case when $\beta = \beta(t)$ requires a residual analysis that could indicate whether a model should consider a covariate as a variable with time.

Due to the fact that that some observations might be censored and in particular, regarding the Cox PH model, the baseline hazard is not estimated, in order to analyse the residuals a particular score process. The risk score for unit i at time t is thought to be $r_i(t) = e^{z_i(t)\beta}$, where $Y_i(t)$ is the indicator function and $Y_i(t) = 1$ indicates a point in which i is under risk and thus observation and it is equal to 0 in other occasions.

The Schoenfeld residuals are given by the equations:

1. $s_k = Z_{(k)} - \frac{\sum_i Y_i(t_k)r_i(t_k)Z_i(t_k)}{\sum_i Y_i(t_k)r_i(t_k)}$
2. $s_k = Z_{(k)} - \bar{z}(\hat{\beta}, t_k)$

In this case, the $Z(k)$ is the covariate vector of the particular unit that is experiencing the event at time k; $\hat{\beta}$ is the estimate of β and $\bar{z}(\hat{\beta}, t_k)$ is the weighted mean of covariate values.

Furthermore, the weighted variance can be represented by the derived equation at the k^{th} time as

$V(\beta, t_k) = \sum_i Y_i(t_k)r_i(t_k)Z_i(t_k) - \bar{z}(\hat{\beta}, t_k)'Z_i(t_k) - \frac{\bar{z}(\hat{\beta}, t_k)}{\sum_i Y_i(t_k)r_i(t_k)}$

From this, we can scale the Schoenfeld residuals by $V(\beta, t_k)$ of X at t_k via the equation:

$s_k^* = V^{-1}(\hat{\beta}, t_k)s_k$

The scaled Schoenfeld residuals can also be defined as follows:

$s_k^* = m\sum_{k=1}^{d} V(\hat{\beta}, t_k)s_k$

here, m is the total number of deaths in the data set.

Following the calculations, the residuals are plotted against time in order to test the proportional hazards assumption. If the assumption is correct, the residuals should be fitting around the line centered at zero (y=0). The further away this predicted line is form the horizontal of (y=0) the more likely one is to call the PH assumption to question and determine whether it is met through the model.

To go a little deeper into the analysis of the resiaul calculation, one can look at the calculations of the test statistic for this residual mdoel.

By producing a least squares slope of regression and assuming a relationship between s_{kj}^* and t_{kj} or some function $g(t_k)$ allows to derive a test statistic for the proportional hazards assumption in regards to the j^{th} covariate, which is given by:

$T_j = \frac{[\sum_{k=1}^{d}(g(t_k)-\hat{g})s_{kj}^{8}]^2}{dI^{jj}\sum_{k=1}^{d}(g(t_k)-\hat{g})^2}$

Here, the distribution is asymptotically as $X^2(1)$ stating the null hypothesis that the relationship between the covariate, in this case j and the event time follows the assumption of PH.

Interpretation of Schoenfeld Residuals from plots in R and the p-values presented.

The y-axis of the Schoenfeld residuals graph can be interpreted as the log of the hazard ratio for the explanatory variable– the coefficient in Cox's model if it were allow to vary over time. If the graph is flat, then the PH assumption is adequate. Furthermore, the Schoenfeld residuals are independent of time. A plot that shows a non-random pattern against time is evidence of violation of the PH assumption. The PH assumption is supported when there's a non-significant relationship between residuals and time. ### HIV Data Cox PH model analysis using Schoenfeld Residuals

Schoenfeld Residuals applied to our best Cox PH model for AIDS data where, we have an additive model of explanatory variables: baseline CD4 count, iv drug use history, and karnofsky performance scale score:

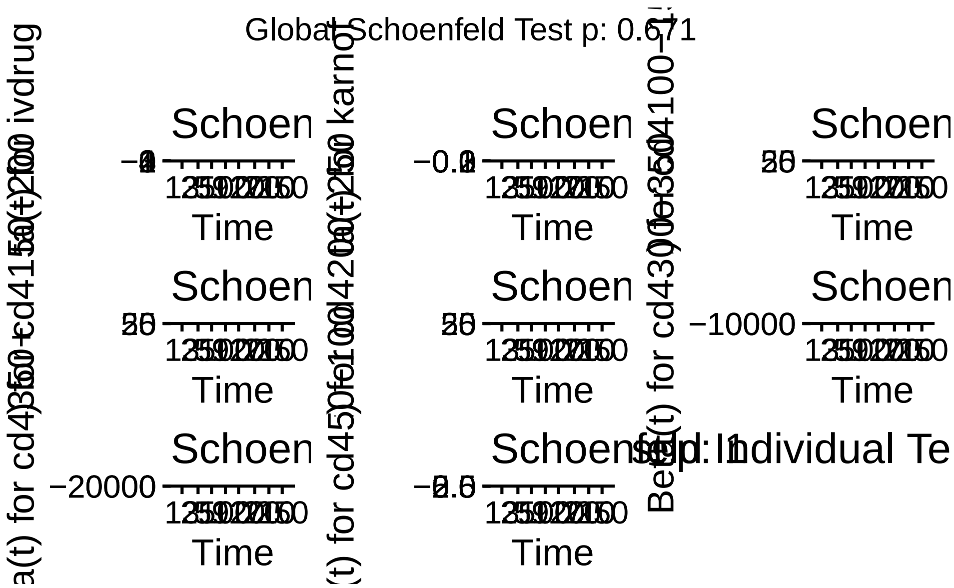
```
cph_r10 <- coxph(Surv(time,censor)~.-priorzdv -id -hemophil -raceth -time_d -strat2
                 -sex -txgrp -age -tx -censor_d, data = aids)
cph_r10
```

```
## Call:
## coxph(formula = Surv(time, censor) ~ . - priorzdv - id - hemophil -
##     raceth - time_d - strat2 - sex - txgrp - age - tx - censor_d,
##     data = aids)
##
##                   coef  exp(coef)   se(coef)      z        p
## ivdrug     -2.252e-01  7.983e-01  1.804e-01 -1.249  0.21177
## karnof     -6.577e-02  9.363e-01  1.415e-02 -4.648 3.36e-06
## cd4100-150 -3.006e+00  4.949e-02  1.011e+00 -2.973  0.00295
## cd4150-200 -2.431e+00  8.798e-02  1.013e+00 -2.400  0.01640
## cd4200-250 -1.618e+00  1.983e-01  1.014e+00 -1.595  0.11070
## cd4300-350 -1.593e+01  1.205e-07  2.298e+03 -0.007  0.99447
## cd4350+    -1.549e+01  1.877e-07  4.449e+03 -0.003  0.99722
## cd450-100  -3.152e-01  7.296e-01  2.975e-01 -1.060  0.28935
##
## Likelihood ratio test=72.49  on 8 df, p=1.565e-12
## n= 851, number of events= 69
```

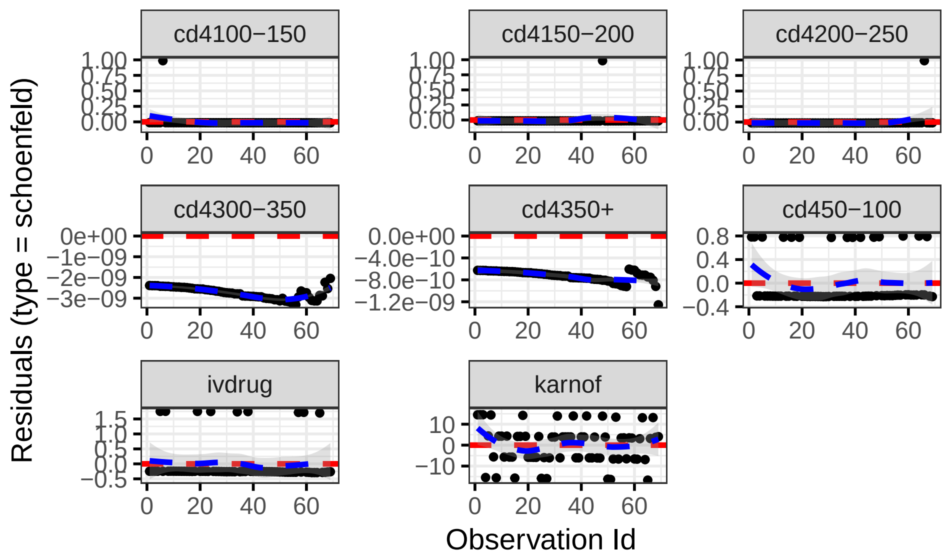
```
zph_r10 <- cox.zph(cph_r10)
zph_r10
```

```
##                 rho    chisq      p
## ivdrug     -0.03879 1.05e-01 0.7457
## karnof     -0.04324 1.29e-01 0.7192
## cd4100-150 -0.17538 2.09e+00 0.1485
## cd4150-200  0.07436 3.72e-01 0.5417
## cd4200-250  0.20640 2.98e+00 0.0846
## cd4300-350 -0.10566 7.61e-10 1.0000
## cd4350+     0.00334 2.36e-12 1.0000
## cd450-100  -0.02945 5.62e-02 0.8126
## GLOBAL           NA 5.79e+00 0.6710
```

```
ggcoxzph(zph_r10)
```



```r
ggcoxdiagnostics(cph_r10, type="schoenfeld")
```



Using the best determined Cox PH model for our data, we can look at the Schoenfeld residuals to determine if the PH assumption is met. Via the function "ggcoxzph()"", which produces, for each covariate, graphs of the scaled Schoenfeld residuals against the transformed time. Here, the solid line is a smoothing spline fit to the plot, with the dashed lines representing a +/- 2-standard-error. from these graphs, we don't see any patterns or significance of the residual fit regarding the graphs of the covariates with time. Therefore, the assumption of proportional hazards seems to be supported for the covariates: baseline CD4 count, iv drug use history, and karnofsky performance scale score.

Using the ggcoxdiagnostics() function we can provide another graphic representation of the residual distribution in regards to the covariates with time. Here, we also see that there's no particular pattern of the residuals around the line of fit, therefore again, we can state that the PH assumption has been met.

References: 1. http://www.ukm.my/jsm/pdf_files/SM-PDF-46-3-2017/15%20Aditif%20Aalen.pdf