Efficacy of Additional IDV with Two Nucleoside Analogues in AIDS Treatment

Hailey Lee

due 5/3/2019

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

Introduction

Despite today's cutting edge technology, HIV still remains as a detrimental infection with no cure, killing millions of people annually[1]. However, according to World Health Organization, a good consistdent treatment can decelerate the progression of HIV in the patient's body almost to a halt[4]. In fact, with numerous ongoing researches and development for new treaments for HIV, many HIV patients are able to remain well and productive for a considerable amount of time. One of the prospective treatments is the combination of three drugs, two of which are nucleoside analogues with an additional potent HIV-protease inhibitor[1]. Specifically, a recent research discovered that when a patient who has been under zidovudine (ZDV) treatment is exposed to indinavir (IDV), zidovudine (ZDV), and lamivudine (3TC) the progress of HIV is hindered[1]. This discovery has raised a crucial question of the efficacy and safety of including indinavir in the three drug treatment for HIV patients. If the efficacy and safety of the new regimen are to be successful, its impact would include prolonging millions of HIV patients' lives as well as breaking a new progress in the goal of HIV cure. This report will attempt to address the effect of IDV in the three drug regimen through statistical analysis.

Methods

The data used in the statistical analysis is obtained from Hammer et al's AIDS Clinical Trial, which is a double-blind, placebo-controlled trial.The data includes variables such as time to AIDS defining diagnosis or death, race/ethinicity, sex, medical history and others from 851 patients. It should be noted that the trial was stopped early as the efficacy results met a pre-specified level of significance at an interim analysis. Moreover, the variable txgrp will not be considered as it overlaps with the variable tx. The dataset for the txgrp contains only the values of "1= ZDV + 3TC", which is equivalent to the control group of tx, and "2= ZDV + 3TC + IDV", which is equivalent to the treatment group of tx. We do not have any data for the values "3= d4T + 3TC" and "4= d4t + 3TC + IDV". Therefore our question of interest is the effect of including IDV in the ZDV and 3TC regimen.

The analysis starts with some explanatory data analysis, looking at the distributions of the sample variables and correlations. The next step is survival analysis. Several Kaplan-Meier curves were created to examine the difference in survival probabiity and cumulative hazard between the treatment and the control group. Then, a Cox PH model is created based on the variables chosen through model building techniques to best represent the survival information. For model building, Akaike's Information Criteria through backward selection and the Drop-in-deviance test were used to select the best variables. The next step is to examine the assumptions of the Cox PH model to validate the inferences that will be made. One of the assumptions is

Proportional Hazard assumption and it is examined through Schoenfeld Residual Test. After checking that all the assumptions are met, inferences and inpretations based on the results are made.

Explanatory Data Analysis

```r
AIDSdata <- read.csv("~/Desktop/MATH150/AIDSdata.csv")

time <- AIDSdata$time
censor <- AIDSdata$censor
tx <- AIDSdata$tx
strat2 <- AIDSdata$strat2
sex <- AIDSdata$sex
raceth <- AIDSdata$raceth
ivdrug <- AIDSdata$ivdrug
hemophil <- AIDSdata$hemophil
karnof <-AIDSdata$karnof
cd4 <- AIDSdata$cd4
priorzdv <- AIDSdata$priorzdv
age <- AIDSdata$age

tx <- as.factor(tx)
strat2 <- as.factor(strat2)
sex <- as.factor(sex)
ivdrug <- as.factor(ivdrug)
raceth <-as.factor(raceth)
hemophil <- as.factor(hemophil)
karnof <- as.factor(karnof)

cor(AIDSdata)
```

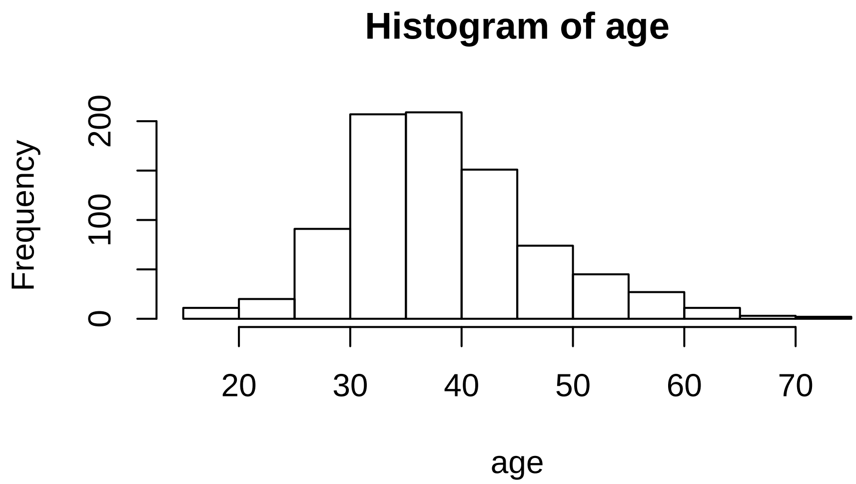
```
##                     id        time       censor       time_d    censor_d
## id        1.000000000  0.04138381  0.041107525  0.063846722  0.05601720
## time      0.041383808  1.00000000 -0.416135794  0.834109840 -0.20775886
## censor    0.041107525 -0.41613579  1.000000000  0.021591192  0.52226788
## time_d    0.063846722  0.83410984  0.021591192  1.000000000 -0.18968665
## censor_d  0.056017198 -0.20775886  0.522267880 -0.189686654  1.00000000
## tx        0.050049647  0.08763450 -0.101461559  0.048541143 -0.06333291
## txgrp     0.050049647  0.08763450 -0.101461559  0.048541143 -0.06333291
## strat2    0.017105707  0.01272685 -0.190163658 -0.066541275 -0.11664989
## sex      -0.020573391 -0.09122431  0.013416438 -0.097845557  0.01811803
## raceth   -0.007873309 -0.05137833 -0.003609676 -0.061490288  0.03454138
## ivdrug    0.057392910  0.03697266 -0.022980899  0.017491697  0.01760321
## hemophil  0.007664485 -0.03017628  0.015391333 -0.024502469  0.01361504
## karnof   -0.005175123  0.05457945 -0.201799857 -0.034821371 -0.14726808
## cd4       0.038704422  0.01168518 -0.220298640 -0.091783660 -0.11901881
## priorzdv  0.030567460  0.01794619 -0.022675657  0.007837262 -0.02213377
## age       0.005587959  0.07501085  0.054215271  0.076584803  0.10874854
##                     tx       txgrp       strat2          sex       raceth
## id        0.050049647  0.050049647  0.017105707 -0.0205733908 -0.007873309
## time      0.087634501  0.087634501  0.012726848 -0.0912243146 -0.051378325
## censor   -0.101461559 -0.101461559 -0.190163658  0.0134164380 -0.003609676
## time_d    0.048541143  0.048541143 -0.066541275 -0.0978455566 -0.061490288
```

```
## censor_d -0.063332908 -0.063332908 -0.116649888  0.0181180344  0.034541381
## tx        1.000000000  1.000000000  0.008914054  0.0351593022 -0.002624371
## txgrp     1.000000000  1.000000000  0.008914054  0.0351593022 -0.002624371
## strat2    0.008914054  0.008914054  1.000000000 -0.0100150156 -0.041813756
## sex       0.035159302  0.035159302 -0.010015016  1.0000000000  0.134806594
## raceth   -0.002624371 -0.002624371 -0.041813756  0.1348065940  1.000000000
## ivdrug   -0.016474382 -0.016474382 -0.014087491  0.0419733646  0.137637052
## hemophil -0.020976327 -0.020976327 -0.011406518 -0.0634178933 -0.003576786
## karnof   -0.034862386 -0.034862386  0.221230873  0.0335869646 -0.016607576
## cd4       0.050973592  0.050973592  0.726044130  0.0095299068 -0.031270266
## priorzdv  0.040931105  0.040931105  0.069456743  0.0003210823 -0.019273829
## age       0.022304493  0.022304493  0.036272351 -0.1358319895 -0.062772335
##                ivdrug     hemophil       karnof          cd4      priorzdv
## id        0.057392910  0.007664485 -0.005175123  0.038704422  0.0305674596
## time      0.036972665 -0.030176276  0.054579454  0.011685180  0.0179461923
## censor   -0.022980899  0.015391333 -0.201799857 -0.220298640 -0.0226756572
## time_d    0.017491697 -0.024502469 -0.034821371 -0.091783660  0.0078372618
## censor_d  0.017603205  0.013615038 -0.147268081 -0.119018806 -0.0221337729
## tx       -0.016474382 -0.020976327 -0.034862386  0.050973592  0.0409311048
## txgrp    -0.016474382 -0.020976327 -0.034862386  0.050973592  0.0409311048
## strat2   -0.014087491 -0.011406518  0.221230873  0.726044130  0.0694567426
## sex       0.041973365 -0.063417893  0.033586965  0.009529907  0.0003210823
## raceth    0.137637052 -0.003576786 -0.016607576 -0.031270266 -0.0192738294
## ivdrug    1.000000000 -0.063971783 -0.079568200  0.022764282 -0.0020785943
## hemophil -0.063971783  1.000000000  0.042849516 -0.002285840  0.0906000417
## karnof   -0.079568200  0.042849516  1.000000000  0.207462189 -0.0355326431
## cd4       0.022764282 -0.002285840  0.207462189  1.000000000  0.1069861581
## priorzdv -0.002078594  0.090600042 -0.035532643  0.106986158  1.0000000000
## age       0.072886854 -0.111125222 -0.100729467  0.007134258  0.1255918449
##                   age
## id        0.005587959
## time      0.075010850
## censor    0.054215271
## time_d    0.076584803
## censor_d  0.108748539
## tx        0.022304493
## txgrp     0.022304493
## strat2    0.036272351
## sex      -0.135831990
## raceth   -0.062772335
## ivdrug    0.072886854
## hemophil -0.111125222
## karnof   -0.100729467
## cd4       0.007134258
## priorzdv  0.125591845
## age       1.000000000
```

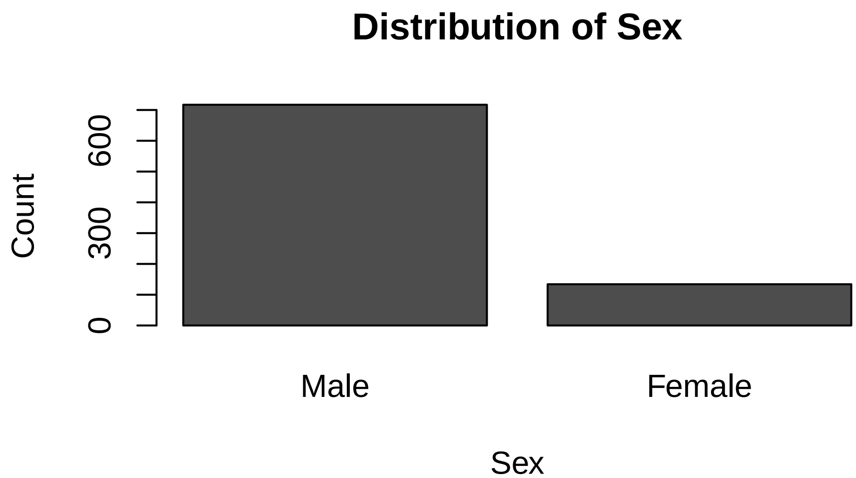
Observing the correlation matrix, we find that time and time_d (0.83), cd4 and strat2(0.72) have noticeably high corrleation. Looking into the context of the variables, it makes sense that time to AIDS diagnosis or death would be correlated to timet to death. Similarly, it also makes sense that the baseline CD4 count would correlate to CD4 stratum at screening. This is important to note as multicollinearity may cause problems in later analysis and inference.

Three charts for distribution of age, sex, and race/ethinicty are created to see how represenatative the sample is.

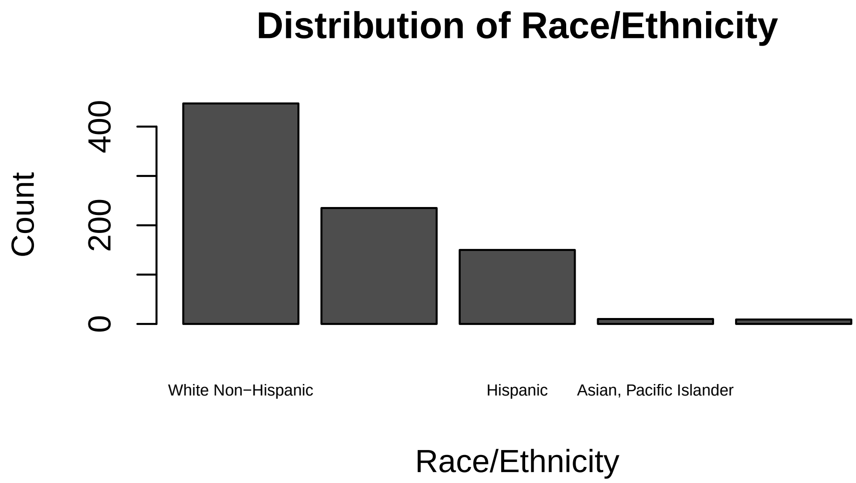
```r
hist(age)
```




```r
matrix_sex <- matrix(c(717, 134),ncol=2)
colnames(matrix_sex) <- c("Male","Female")
rownames(matrix_sex) <- c("Count")
matrix_sex <- as.table(matrix_sex)
barplot(matrix_sex, main="Distribution of Sex",xlab = "Sex",ylab = "Count")
```




```r
matrix_raceth <- matrix(c(447,235,150,10,9),ncol=5)
colnames(matrix_raceth) <- c("White Non-Hispanic","Black Non-Hispanic","Hispanic","Asian, Pacific Islan
rownames(matrix_raceth) <- c("Count")
matrix_raceth <- as.table(matrix_raceth)
barplot(matrix_raceth, main="Distribution of Race/Ethnicity",xlab = "Race/Ethnicity",cex.names=.5,ylab =
```



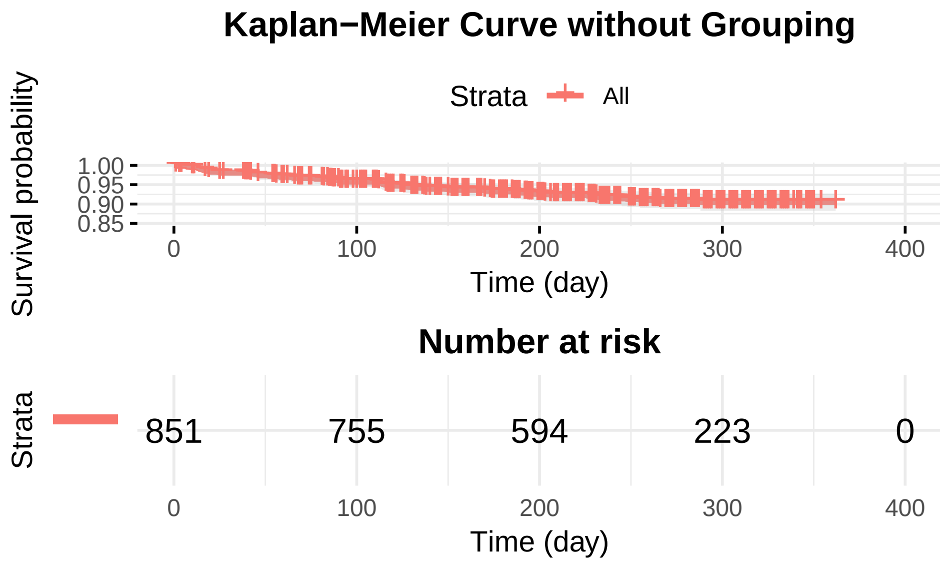
Looking at the age distribution, we can see that the samples are fairly normally distributed. The distribution of sex is highly skewed towards male. Similarly, the distribution of race/ethnicity shows that the sample is mainly from white non-hispanic patients. However, according to CDC, most of those living with HIV are male and white non-hispanic ranks the highest of the population with black non-hispanic following after[4].

Survival Curves

First, the figure below is the Kaplan-Meier curve without any grouping. This is the null model that shows the overall survival probability of event, namely AIDS diagnosis or death, following the time (in day) axis. The second figure is the side-by-side view of the Kaplan-Meirer curve for treatment and control group, and the corresponding cumulative hazard function for the same groups. From the Kaplan-Meier curve we can observe that there is higher probability of survival in the treatment group (with IDV) compared to the control group (without IDV). Similarly, there exists more cumulative hazard for the treatment group compared to the control group. Note that in both Kaplan-Meier curves, the survival probability axis is scaled to range from 0.85 to 1.00, not 0 to 1.

```
surv.obj <- Surv(time,censor)
surv.everyone <- survfit(surv.obj ~ 1, data = AIDSdata)

ggsurvplot(surv.everyone, ylim=c(0.85,1), title="Kaplan-Meier Curve without Grouping", xlab="Time (day)
, risk.table.y.text.col = TRUE, risk.table.y.text = FALSE,risk.table.height = 0.45)
```

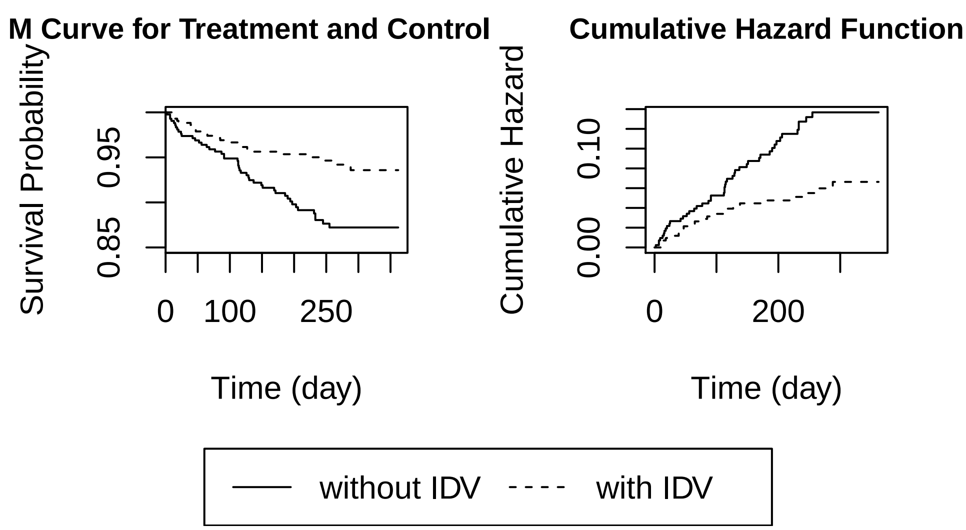



```
KM<- survfit(surv.obj~tx,type="kaplan-meier", data=AIDSdata)

par(oma = c(2,0,0,0))

par(mfrow=c(1,2))
plot(KM,lty=1:2, ylim=c(0.85,1),main="KM Curve for Treatment and Control Group", xlab="Time (day)",ylab=
plot(KM, fun="cumhaz",lty=1:2, main="Cumulative Hazard Function", cex.main=0.9, xlab="Time (day)", ylab=

par(fig = c(0, 1, 0, 1), oma = c(0, 0, 0, 0), mar = c(0, 0, 0, 0), new = TRUE)
plot(0, 0, type = "n", bty = "n", xaxt = "n", yaxt = "n")
legend("bottom",c("without IDV","with IDV"),lty=1:2,xpd = TRUE, horiz = TRUE, inset = c(0,0))
```



Model Building

I selected Akaike's Information Criteria for model building technique. This process conducts the backward model selection by starting from a full model with every variable included and eliminates insignificant variables downward.The objective is to find the model that provides the least AIC and produces the best survival information.

The variables that will be tested through the step function are: tx, sex, karnof, cd4, log(priorzdv), age, and all interaction terms among these variables. Notice that the list of variables does not contain strat2, hemophil, and raceth. As discussed in the Explanatory Data Analysis, cd4 and strat2 has a high corrleation, which leads to a problem of multicolinearilty. To avoid this problem, I selected one of the variables, cd4, to include in the AIC model. I choose cd4 because it is a continuous variable, compared to strat2 which is a discrete/binary variable. cd4 will provide more power and sensitivity to the model. Next omitted variables are hemophil and raceth. These two variables are omitted from the AIC model due to failure of convergence in the step function. However, following the results of AIC model, I will be conducnting the drop in deviance test to examine the model including these variables.

```r
model.AIC <- coxph(surv.obj ~ (tx + sex + cd4 + log(priorzdv) + age + ivdrug + karnof)^2, data = AIDSda

step(model.AIC, data = AIDSdata, direction = "backward", k = 2)
```

```
## Start:  AIC=848.18
## surv.obj ~ (tx + sex + cd4 + log(priorzdv) + age + ivdrug + karnof)^2
##
##                        Df    AIC
## - cd4:log(priorzdv)     1 846.20
## - log(priorzdv):karnof  1 846.22
## - tx:karnof             1 846.22
## - tx:ivdrug             1 846.23
## - age:ivdrug            1 846.29
## - sex:log(priorzdv)     1 846.37
## - sex:ivdrug            1 846.41
## - log(priorzdv):age     1 846.60
## - tx:age                1 846.69
## - sex:cd4               1 846.76
## - tx:cd4                1 846.90
## - sex:age               1 847.35
## - cd4:ivdrug            1 847.52
## <none>                    848.18
## - ivdrug:karnof         1 848.23
## - cd4:age               1 848.24
## - cd4:karnof            1 849.08
## - age:karnof            1 849.71
## - sex:karnof            1 849.94
## - tx:sex                1 851.03
## - log(priorzdv):ivdrug  1 852.82
## - tx:log(priorzdv)      1 853.75
##
## Step:  AIC=846.2
## surv.obj ~ tx + sex + cd4 + log(priorzdv) + age + ivdrug + karnof +
##     tx:sex + tx:cd4 + tx:log(priorzdv) + tx:age + tx:ivdrug +
##     tx:karnof + sex:cd4 + sex:log(priorzdv) + sex:age + sex:ivdrug +
##     sex:karnof + cd4:age + cd4:ivdrug + cd4:karnof + log(priorzdv):age +
##     log(priorzdv):ivdrug + log(priorzdv):karnof + age:ivdrug +
##     age:karnof + ivdrug:karnof
##
##                        Df    AIC
## - log(priorzdv):karnof  1 844.22
## - tx:karnof             1 844.24
## - tx:ivdrug             1 844.24
## - age:ivdrug            1 844.30
## - sex:log(priorzdv)     1 844.38
```

```
## - sex:ivdrug            1 844.43
## - tx:age                1 844.70
## - log(priorzdv):age     1 844.71
## - sex:cd4               1 844.78
## - tx:cd4                1 845.01
## - sex:age               1 845.35
## - cd4:ivdrug            1 845.78
## <none>                    846.20
## - cd4:age               1 846.24
## - ivdrug:karnof         1 846.28
## - cd4:karnof            1 847.10
## - age:karnof            1 847.72
## - sex:karnof            1 847.94
## - tx:sex                1 849.04
## - log(priorzdv):ivdrug  1 850.89
## - tx:log(priorzdv)      1 851.90
##
## Step:  AIC=844.22
## surv.obj ~ tx + sex + cd4 + log(priorzdv) + age + ivdrug + karnof +
##     tx:sex + tx:cd4 + tx:log(priorzdv) + tx:age + tx:ivdrug +
##     tx:karnof + sex:cd4 + sex:log(priorzdv) + sex:age + sex:ivdrug +
##     sex:karnof + cd4:age + cd4:ivdrug + cd4:karnof + log(priorzdv):age +
##     log(priorzdv):ivdrug + age:ivdrug + age:karnof + ivdrug:karnof
##
##                        Df    AIC
## - tx:ivdrug             1 842.27
## - tx:karnof             1 842.28
## - age:ivdrug            1 842.33
## - sex:log(priorzdv)     1 842.41
## - sex:ivdrug            1 842.46
## - tx:age                1 842.71
## - log(priorzdv):age     1 842.77
## - sex:cd4               1 842.81
## - tx:cd4                1 843.03
## - sex:age               1 843.37
## - cd4:ivdrug            1 843.79
## <none>                    844.22
## - cd4:age               1 844.26
## - ivdrug:karnof         1 844.42
## - cd4:karnof            1 845.10
## - age:karnof            1 845.77
## - sex:karnof            1 845.94
## - tx:sex                1 847.04
## - log(priorzdv):ivdrug  1 848.91
## - tx:log(priorzdv)      1 849.90
##
## Step:  AIC=842.27
## surv.obj ~ tx + sex + cd4 + log(priorzdv) + age + ivdrug + karnof +
##     tx:sex + tx:cd4 + tx:log(priorzdv) + tx:age + tx:karnof +
##     sex:cd4 + sex:log(priorzdv) + sex:age + sex:ivdrug + sex:karnof +
##     cd4:age + cd4:ivdrug + cd4:karnof + log(priorzdv):age + log(priorzdv):ivdrug +
##     age:ivdrug + age:karnof + ivdrug:karnof
##
##                        Df    AIC
```

```
## - tx:karnof             1 840.32
## - age:ivdrug            1 840.42
## - sex:log(priorzdv)     1 840.47
## - sex:ivdrug            1 840.47
## - tx:age                1 840.82
## - log(priorzdv):age     1 840.83
## - sex:cd4               1 840.85
## - tx:cd4                1 841.04
## - sex:age               1 841.44
## - cd4:ivdrug            1 841.83
## <none>                    842.27
## - cd4:age               1 842.32
## - ivdrug:karnof         1 842.56
## - cd4:karnof            1 843.15
## - age:karnof            1 843.84
## - sex:karnof            1 844.02
## - tx:sex                1 845.07
## - log(priorzdv):ivdrug  1 846.93
## - tx:log(priorzdv)      1 848.33
##
## Step:  AIC=840.32
## surv.obj ~ tx + sex + cd4 + log(priorzdv) + age + ivdrug + karnof +
##     tx:sex + tx:cd4 + tx:log(priorzdv) + tx:age + sex:cd4 + sex:log(priorzdv) +
##     sex:age + sex:ivdrug + sex:karnof + cd4:age + cd4:ivdrug +
##     cd4:karnof + log(priorzdv):age + log(priorzdv):ivdrug + age:ivdrug +
##     age:karnof + ivdrug:karnof
##
##                        Df    AIC
## - age:ivdrug            1 838.48
## - sex:log(priorzdv)     1 838.51
## - sex:ivdrug            1 838.52
## - sex:cd4               1 838.89
## - log(priorzdv):age     1 838.89
## - tx:age                1 839.01
## - tx:cd4                1 839.04
## - sex:age               1 839.56
## - cd4:ivdrug            1 839.89
## <none>                    840.32
## - cd4:age               1 840.38
## - ivdrug:karnof         1 840.70
## - cd4:karnof            1 841.17
## - age:karnof            1 842.04
## - sex:karnof            1 842.69
## - tx:sex                1 843.37
## - log(priorzdv):ivdrug  1 845.00
## - tx:log(priorzdv)      1 846.34
##
## Step:  AIC=838.48
## surv.obj ~ tx + sex + cd4 + log(priorzdv) + age + ivdrug + karnof +
##     tx:sex + tx:cd4 + tx:log(priorzdv) + tx:age + sex:cd4 + sex:log(priorzdv) +
##     sex:age + sex:ivdrug + sex:karnof + cd4:age + cd4:ivdrug +
##     cd4:karnof + log(priorzdv):age + log(priorzdv):ivdrug + age:karnof +
##     ivdrug:karnof
##
```

```
##                       Df    AIC
## - sex:log(priorzdv)     1 836.66
## - sex:ivdrug            1 836.73
## - log(priorzdv):age     1 836.97
## - sex:cd4               1 837.04
## - tx:age                1 837.12
## - tx:cd4                1 837.20
## - sex:age               1 837.80
## - cd4:ivdrug            1 838.00
## - cd4:age               1 838.44
## <none>                    838.48
## - ivdrug:karnof         1 838.99
## - cd4:karnof            1 839.30
## - age:karnof            1 840.19
## - sex:karnof            1 840.96
## - tx:sex                1 841.46
## - log(priorzdv):ivdrug  1 843.18
## - tx:log(priorzdv)      1 844.42
##
## Step:  AIC=836.66
## surv.obj ~ tx + sex + cd4 + log(priorzdv) + age + ivdrug + karnof +
##     tx:sex + tx:cd4 + tx:log(priorzdv) + tx:age + sex:cd4 + sex:age +
##     sex:ivdrug + sex:karnof + cd4:age + cd4:ivdrug + cd4:karnof +
##     log(priorzdv):age + log(priorzdv):ivdrug + age:karnof + ivdrug:karnof
##
##                       Df    AIC
## - log(priorzdv):age     1 835.08
## - sex:ivdrug            1 835.18
## - tx:age                1 835.29
## - sex:cd4               1 835.30
## - tx:cd4                1 835.39
## - sex:age               1 835.88
## - cd4:ivdrug            1 836.20
## - cd4:age               1 836.64
## <none>                    836.66
## - ivdrug:karnof         1 837.22
## - cd4:karnof            1 837.53
## - age:karnof            1 838.41
## - sex:karnof            1 839.14
## - tx:sex                1 839.59
## - log(priorzdv):ivdrug  1 841.23
## - tx:log(priorzdv)      1 843.61
##
## Step:  AIC=835.08
## surv.obj ~ tx + sex + cd4 + log(priorzdv) + age + ivdrug + karnof +
##     tx:sex + tx:cd4 + tx:log(priorzdv) + tx:age + sex:cd4 + sex:age +
##     sex:ivdrug + sex:karnof + cd4:age + cd4:ivdrug + cd4:karnof +
##     log(priorzdv):ivdrug + age:karnof + ivdrug:karnof
##
##                       Df    AIC
## - sex:ivdrug            1 833.56
## - sex:cd4               1 833.70
## - tx:age                1 833.93
## - tx:cd4                1 833.96
```

```
## - sex:age               1 834.36
## - cd4:ivdrug            1 834.57
## - cd4:age               1 834.87
## <none>                    835.08
## - ivdrug:karnof         1 835.64
## - cd4:karnof            1 836.01
## - age:karnof            1 836.68
## - sex:karnof            1 837.47
## - tx:sex                1 838.06
## - log(priorzdv):ivdrug  1 839.72
## - tx:log(priorzdv)      1 842.56
##
## Step:  AIC=833.56
## surv.obj ~ tx + sex + cd4 + log(priorzdv) + age + ivdrug + karnof +
##     tx:sex + tx:cd4 + tx:log(priorzdv) + tx:age + sex:cd4 + sex:age +
##     sex:karnof + cd4:age + cd4:ivdrug + cd4:karnof + log(priorzdv):ivdrug +
##     age:karnof + ivdrug:karnof
##
##                        Df    AIC
## - sex:cd4               1 832.12
## - tx:age                1 832.44
## - tx:cd4                1 832.44
## - cd4:ivdrug            1 832.92
## - sex:age               1 832.92
## - cd4:age               1 833.31
## <none>                    833.56
## - ivdrug:karnof         1 833.98
## - cd4:karnof            1 834.50
## - age:karnof            1 835.06
## - sex:karnof            1 836.03
## - tx:sex                1 836.75
## - log(priorzdv):ivdrug  1 838.28
## - tx:log(priorzdv)      1 841.53
##
## Step:  AIC=832.12
## surv.obj ~ tx + sex + cd4 + log(priorzdv) + age + ivdrug + karnof +
##     tx:sex + tx:cd4 + tx:log(priorzdv) + tx:age + sex:age + sex:karnof +
##     cd4:age + cd4:ivdrug + cd4:karnof + log(priorzdv):ivdrug +
##     age:karnof + ivdrug:karnof
##
##                        Df    AIC
## - tx:cd4                1 830.81
## - tx:age                1 830.98
## - cd4:ivdrug            1 831.53
## - sex:age               1 831.83
## <none>                    832.12
## - cd4:age               1 832.33
## - ivdrug:karnof         1 832.56
## - cd4:karnof            1 833.29
## - age:karnof            1 833.96
## - tx:sex                1 834.79
## - sex:karnof            1 836.25
## - log(priorzdv):ivdrug  1 836.86
## - tx:log(priorzdv)      1 840.11
```

```
##
## Step:  AIC=830.81
## surv.obj ~ tx + sex + cd4 + log(priorzdv) + age + ivdrug + karnof +
##     tx:sex + tx:log(priorzdv) + tx:age + sex:age + sex:karnof +
##     cd4:age + cd4:ivdrug + cd4:karnof + log(priorzdv):ivdrug +
##     age:karnof + ivdrug:karnof
##
##                        Df    AIC
## - tx:age                1 829.96
## - sex:age               1 830.38
## - cd4:ivdrug            1 830.47
## <none>                    830.81
## - cd4:age               1 830.99
## - ivdrug:karnof         1 831.20
## - cd4:karnof            1 831.80
## - age:karnof            1 832.52
## - tx:sex                1 833.35
## - sex:karnof            1 834.59
## - log(priorzdv):ivdrug  1 835.44
## - tx:log(priorzdv)      1 838.47
##
## Step:  AIC=829.96
## surv.obj ~ tx + sex + cd4 + log(priorzdv) + age + ivdrug + karnof +
##     tx:sex + tx:log(priorzdv) + sex:age + sex:karnof + cd4:age +
##     cd4:ivdrug + cd4:karnof + log(priorzdv):ivdrug + age:karnof +
##     ivdrug:karnof
##
##                        Df    AIC
## - sex:age               1 829.23
## - cd4:ivdrug            1 829.61
## - cd4:age               1 829.82
## <none>                    829.96
## - ivdrug:karnof         1 830.33
## - cd4:karnof            1 830.87
## - age:karnof            1 831.52
## - tx:sex                1 831.94
## - sex:karnof            1 833.50
## - log(priorzdv):ivdrug  1 834.44
## - tx:log(priorzdv)      1 837.19
##
## Step:  AIC=829.23
## surv.obj ~ tx + sex + cd4 + log(priorzdv) + age + ivdrug + karnof +
##     tx:sex + tx:log(priorzdv) + sex:karnof + cd4:age + cd4:ivdrug +
##     cd4:karnof + log(priorzdv):ivdrug + age:karnof + ivdrug:karnof
##
##                        Df    AIC
## - cd4:ivdrug            1 828.95
## <none>                    829.23
## - cd4:age               1 829.43
## - ivdrug:karnof         1 829.77
## - cd4:karnof            1 829.93
## - age:karnof            1 830.37
## - tx:sex                1 830.87
## - sex:karnof            1 831.63
```

```
## - log(priorzdv):ivdrug  1 833.57
## - tx:log(priorzdv)      1 836.42
##
## Step:  AIC=828.95
## surv.obj ~ tx + sex + cd4 + log(priorzdv) + age + ivdrug + karnof +
##     tx:sex + tx:log(priorzdv) + sex:karnof + cd4:age + cd4:karnof +
##     log(priorzdv):ivdrug + age:karnof + ivdrug:karnof
##
##                        Df    AIC
## - ivdrug:karnof         1 828.16
## - cd4:age               1 828.94
## <none>                    828.95
## - cd4:karnof            1 829.39
## - age:karnof            1 829.88
## - tx:sex                1 830.42
## - sex:karnof            1 831.19
## - log(priorzdv):ivdrug  1 832.34
## - tx:log(priorzdv)      1 835.58
##
## Step:  AIC=828.16
## surv.obj ~ tx + sex + cd4 + log(priorzdv) + age + ivdrug + karnof +
##     tx:sex + tx:log(priorzdv) + sex:karnof + cd4:age + cd4:karnof +
##     log(priorzdv):ivdrug + age:karnof
##
##                        Df    AIC
## - cd4:age               1 828.09
## <none>                    828.16
## - cd4:karnof            1 828.52
## - age:karnof            1 828.94
## - tx:sex                1 829.50
## - sex:karnof            1 830.57
## - log(priorzdv):ivdrug  1 830.72
## - tx:log(priorzdv)      1 834.63
##
## Step:  AIC=828.09
## surv.obj ~ tx + sex + cd4 + log(priorzdv) + age + ivdrug + karnof +
##     tx:sex + tx:log(priorzdv) + sex:karnof + cd4:karnof + log(priorzdv):ivdrug +
##     age:karnof
##
##                        Df    AIC
## - cd4:karnof            1 827.65
## - age:karnof            1 827.93
## <none>                    828.09
## - tx:sex                1 829.47
## - sex:karnof            1 830.14
## - log(priorzdv):ivdrug  1 830.70
## - tx:log(priorzdv)      1 835.30
##
## Step:  AIC=827.65
## surv.obj ~ tx + sex + cd4 + log(priorzdv) + age + ivdrug + karnof +
##     tx:sex + tx:log(priorzdv) + sex:karnof + log(priorzdv):ivdrug +
##     age:karnof
##
##                        Df    AIC
```

```
## - age:karnof            1 826.91
## <none>                    827.65
## - tx:sex                1 829.04
## - sex:karnof            1 829.36
## - log(priorzdv):ivdrug  1 830.57
## - tx:log(priorzdv)      1 835.00
## - cd4                   1 861.07
##
## Step:  AIC=826.91
## surv.obj ~ tx + sex + cd4 + log(priorzdv) + age + ivdrug + karnof +
##     tx:sex + tx:log(priorzdv) + sex:karnof + log(priorzdv):ivdrug
##
##                        Df    AIC
## <none>                    826.91
## - age                   1 827.24
## - tx:sex                1 828.15
## - sex:karnof            1 828.17
## - log(priorzdv):ivdrug  1 830.01
## - tx:log(priorzdv)      1 834.14
## - cd4                   1 859.86
```

```
## Call:
## coxph(formula = surv.obj ~ tx + sex + cd4 + log(priorzdv) + age +
##     ivdrug + karnof + tx:sex + tx:log(priorzdv) + sex:karnof +
##     log(priorzdv):ivdrug, data = AIDSdata)
##
##                           coef exp(coef)  se(coef)     z      p
## tx                    7.50e-01  2.12e+00  1.19e+00  0.63 0.5267
## sex                   6.91e+00  1.01e+03  3.78e+00  1.83 0.0674
## cd4                  -1.55e-02  9.85e-01  3.18e-03 -4.89  1e-06
## log(priorzdv)        -3.45e-01  7.08e-01  3.78e-01 -0.91 0.3621
## age                   2.17e-02  1.02e+00  1.40e-02  1.55 0.1202
## ivdrug               -2.41e+00  8.96e-02  1.09e+00 -2.22 0.0264
## karnof                3.35e-02  1.03e+00  5.17e-02  0.65 0.5178
## tx:sex                1.22e+00  3.38e+00  6.71e-01  1.82 0.0690
## tx:log(priorzdv)     -9.73e-01  3.78e-01  3.30e-01 -2.95 0.0031
## sex:karnof           -7.98e-02  9.23e-01  4.46e-02 -1.79 0.0740
## log(priorzdv):ivdrug  6.21e-01  1.86e+00  2.93e-01  2.12 0.0339
##
## Likelihood ratio test=100  on 11 df, p=1.11e-16
## n= 851, number of events= 69
```

From Akaike's Information Criteria, we obtain the model coxph(formula = surv.obj ~ tx + sex + cd4 + log(priorzdv) + age + ivdrug + karnof + tx:sex + tx:log(priorzdv) + sex:karnof + log(priorzdv):ivdrug, data = AIDSdata)" to best represent the cox PH model. However, we can see in the p-values that some of the terms are not significant. Specifically, there are 7 terms with high p-values: tx with p-value = 0.5267 > 0.05, sex with p-value = 0.0674 > 0.05, log(priorzdv) with p-value = 0.3621 > 0.05, age with p-value = 0.1202 > 0.05, karnof with p-value = 0.5178, tx:sex with p-value = 0.0690 > 0.05, and sex:karnof with p-value = 0.0740 > 0.05. We will be testing if these terms are necessary in the model. We will also incorporate the variables that were omitted from the first AIC process, namely hemophil, raceth and the interaction between the two.

We first eliminate hemophil:raceth, hemophil, raceth. We obtain G-statistic of 0.143704 with 3 degrees of freedom, which gives us p-value of 0.9861204 > 0.05. Thus, we can reject the null hypothesis and conclude that we do not need those terms in the model.

```r
model_2 <- coxph(surv.obj ~ tx + sex + cd4 + log(priorzdv) + age + ivdrug + karnof + tx:sex + tx:log(pr

model_3 <- coxph(surv.obj ~ tx + sex + cd4 + log(priorzdv) + age + ivdrug + karnof + tx:sex + tx:log(pr

G3 = 2*(model_2$loglik[2] - model_3$loglik[2])
G3
```

```
## [1] 0.143704
```

```r
1-pchisq(G3, 3)
```

```
## [1] 0.9861204
```

Similarly, we proceed to eliminate the terms with highest p-values, namely tx, karnof, log(priorzdv), and age. We obtain G-statistic of 4.15885 with 4 degrees of freedom, which corresponds to the p-value of $0.3526382 > 0.05$. Thus, we can reject the null hypothesis and conclude that we do not need those terms in the model.

```r
model_4 <- coxph(surv.obj ~ sex + cd4 + ivdrug + tx:sex + tx:log(priorzdv) + sex:karnof + log(priorzdv)

G4 = 2*(model_3$loglik[2] - model_4$loglik[2])
G4
```

```
## [1] 4.415885
```

```r
1-pchisq(G4, 4)
```

```
## [1] 0.3526382
```

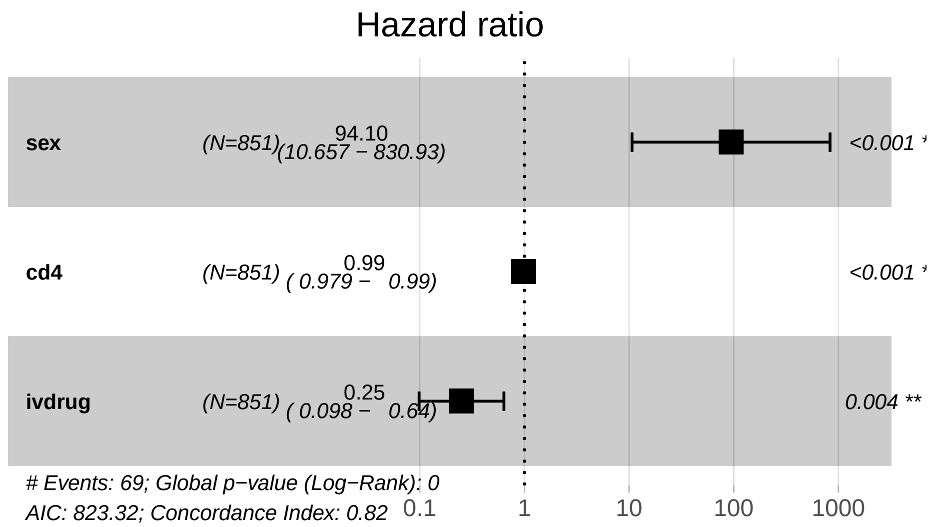
The final model and its summary is shown below.

```r
model_final <- model_4
summary(model_final)
```

```
## Call:
## coxph(formula = surv.obj ~ sex + cd4 + ivdrug + tx:sex + tx:log(priorzdv) +
##     sex:karnof + log(priorzdv):ivdrug, data = AIDSdata)
##
##   n= 851, number of events= 69
##
##                           coef exp(coef)  se(coef)      z Pr(>|z|)
## sex                   4.544371 94.101217  1.111333  4.089 4.33e-05 ***
## cd4                  -0.014842  0.985267  0.003092 -4.800 1.59e-06 ***
## ivdrug               -1.384768  0.250382  0.476076 -2.909 0.003629 **
## sex:tx                1.531421  4.624742  0.544041  2.815 0.004879 **
## tx:log(priorzdv)     -0.847839  0.428340  0.231357 -3.665 0.000248 ***
## sex:karnof           -0.054329  0.947121  0.012229 -4.442 8.89e-06 ***
## ivdrug:log(priorzdv)  0.327097  1.386936  0.118390  2.763 0.005729 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                      exp(coef) exp(-coef) lower .95 upper .95
## sex                    94.1012    0.01063  10.65680  830.9282
## cd4                     0.9853    1.01495   0.97931    0.9913
## ivdrug                  0.2504    3.99390   0.09848    0.6366
## sex:tx                  4.6247    0.21623   1.59220   13.4331
## tx:log(priorzdv)        0.4283    2.33460   0.27218    0.6741
## sex:karnof              0.9471    1.05583   0.92469    0.9701
## ivdrug:log(priorzdv)    1.3869    0.72101   1.09972    1.7492
```

```
##
## Concordance= 0.815  (se = 0.035 )
## Rsquare= 0.107   (max possible= 0.655 )
## Likelihood ratio test= 95.94  on 7 df,   p=0
## Wald test            = 79.45  on 7 df,   p=1.776e-14
## Score (logrank) test = 91.55  on 7 df,   p=1.11e-16
```

```
ggforest(model_final, data = AIDSdata)
```



The summary of Cox PH model delivers a complete report of the model. We can see that the model includes "sex","cd4","ivdrug" as well as some interaction terms such as "sex:tx","tx:log(priorzdv)","sex:karnof","ivdrug:log(priorzdv)". All other variables have been eliminiated from the model. We can also see that the p-values for the Wald Test and Score (logrank) test are statistically significant, implying the model is fairly significant.

In the summary of the model, the hazard ratio (HR) is shown through the colomn of exp(coef). A positive sign implies that the hazard (or risk of AIS diagnosis/death) is higher for patients with higher values of the corresponding variable. The given HR is the ratio of the second group relative to the first group. For example, in sex, the number 94.101217 means that female (coded 2) have less hazard than male (coded 1) by a factor of 94.10, or 5.9%. Below the summary, there exists a full HR table that shows the list of HR with corresponding confidence intervals for each factor.

Overall, while sex, sex:tx, and ivdrug:log(priorzdv) have a positive coefficient, all the other variables such as cd4, ivdrug, tx:log(priorzdv), and sex:karnof have negative coefficient. Therefore, sex (specifically, being female), and the interactions terms, sex:tx and ivdrug:log(priorzdv), are associated with better probability of survival.

As discussed in the method section, the data is obtained from a double-blind, placebo-controlled trial. As it is an experiment, not an observational study, we can assume causation. In other words, the IDV treatment does increase the probability of survival, prolonging the time until AIDS diagnosis or death.

In addition, this conclusion may be extended to the general HIV patient population whose CD4 cells per cubic millimeter are no more than 200 and those who've had at least three months of prior zidovudine therapy.

Schoenfeld Residuals

There are three technical conditions (or assumptions) for Cox PH model. First assumption is that all the observations must be independent. Similarly, second assumption is that all the censorings are independent.

Last assumption is the Proportional Hazard Assumption. In other words, the hazard ratios are not dependent on time.

The Schoenfeld Residuals can test the last assumption, the proportional Hazard assumption. It tests the independence between residuals and time.

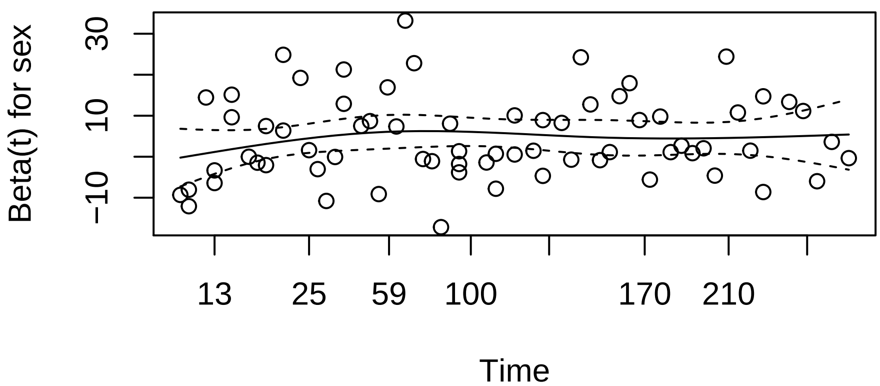
The Schoenfeld Residual for x_l and any subject i who is still alive at time t_j is difference between the covariate x_{i1} for that subject and the weighted average of the covariates in the risk set[2]. Mathematically, Schoenfeld residual$_i = x_{i1} - \overline{x}_1(\beta, t_j)$ where x_{i1} is the covariate value for the individual that failed and $\overline{m}_1(\beta, t_j)$ is the average value for the lth covariate.[2] Therefore, we obtain a separate residual graph for each covariate. The graph will plot the residuals against time axis. The Schoenfeld residuals to check the fit of Cox PH model. Here, if the plot of Schoenfeld residuals against time shows a non-random pattern,we can say that the PH assumption has been violated.

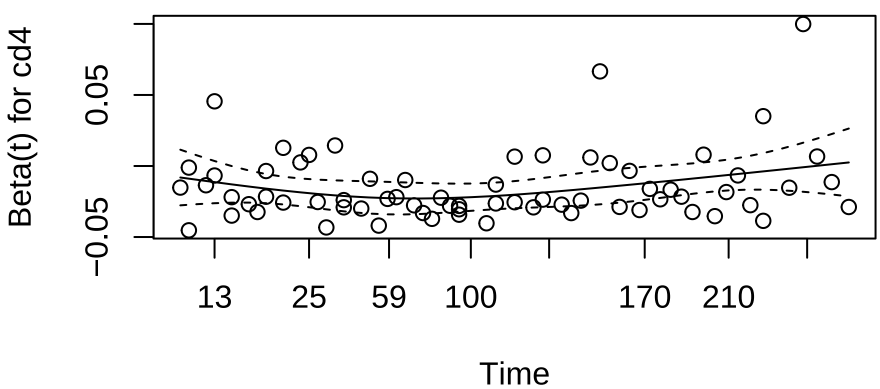
```
fit <- cox.zph(model_final)
fit
```

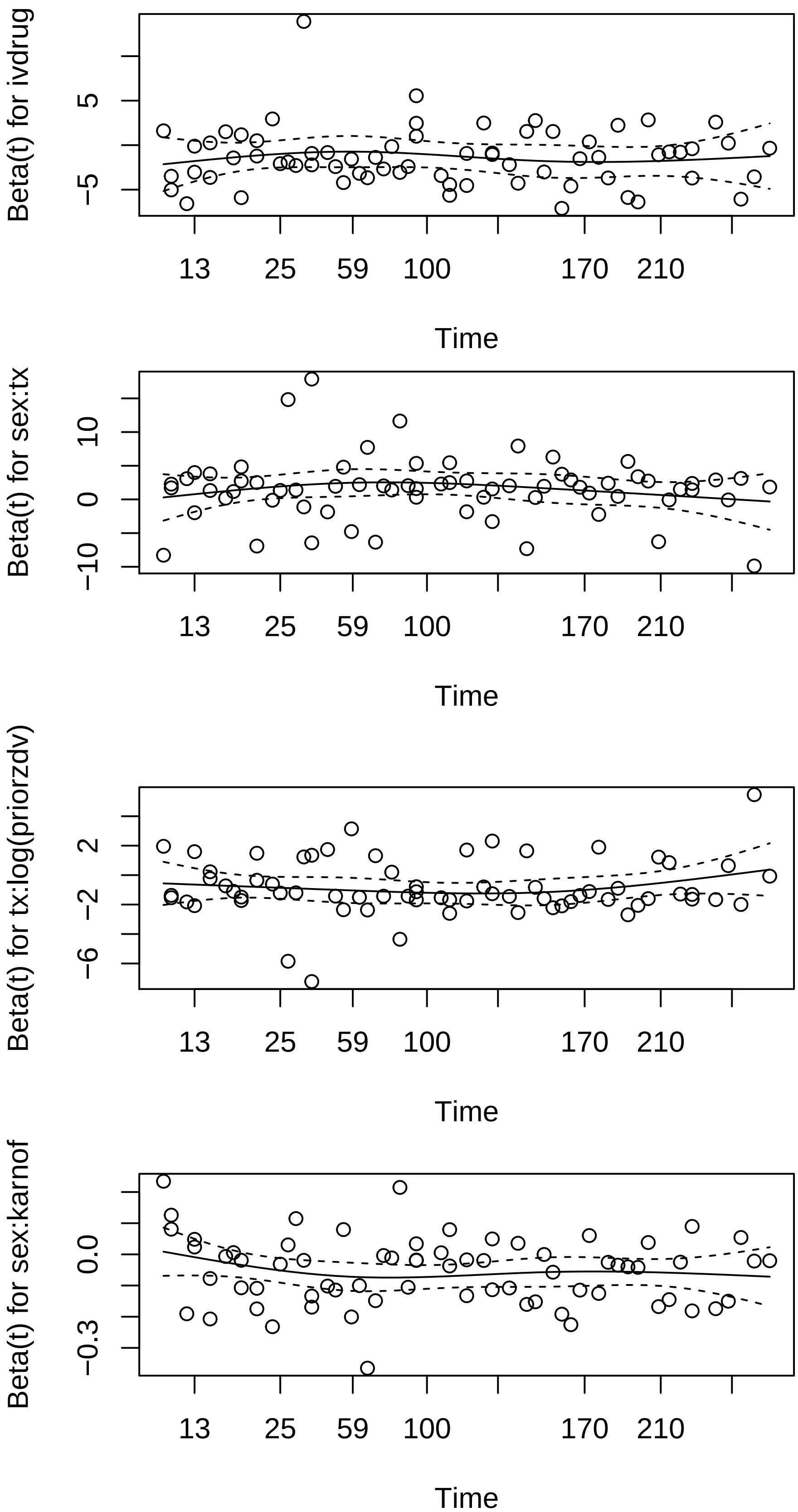
```
##                         rho  chisq     p
## sex                  0.0745 0.4609 0.497
## cd4                  0.1539 1.5078 0.219
## ivdrug              -0.0379 0.0677 0.795
## sex:tx              -0.0655 0.3052 0.581
## tx:log(priorzdv)     0.0625 0.2653 0.606
## sex:karnof          -0.1101 0.9488 0.330
## ivdrug:log(priorzdv)  0.0183 0.0189 0.891
## GLOBAL                   NA 3.9482 0.786
```

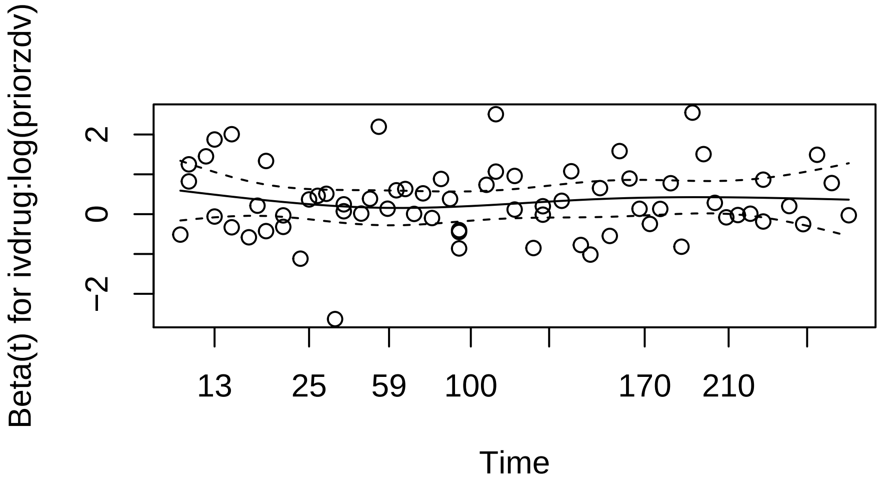
We can see from the output above that the test is not significant for all the covariates, and the global test is also not statistically significant. Therefore, we can assume the proportional hazard. We expect to see random residual plots for all covariates.

```
plot(fit)
```






Beta(t) for ivdrug
5
−5
13
25
59
100
170
210
Time
Beta(t) for sex:tx
10
0
−10
13
25
59
100
170
210
Time
Beta(t) for tx:log(priorzdv)
2
−2
−6
13
25
59
100
170
210
Time
Beta(t) for sex:karnof
0.0
−0.3
13
25
59
100
170
210
Time



The plots contain 1 solid line that smoothes out the fit of the plot with 2 dotted lines above and below for +/-2 standard error. Looking at the graphs we can verify that there is no pattern along the time axis. For example, the Schoenfeld Residual plot for sex covariate does not appear to have any noticeable patterns, which means that men are just as likely to have AIDS diagnosis or death as women, and vice versa. It shows that the hazard ratio between two groups is constant. Therefore, we can conclude that the proportional hazard assumption has been met.

References

[1] Hammer, Scott M., et al. "A Controlled Trial of Two Nucleoside Analogues plus Indinavir in Persons with Human Immunodeficiency Virus Infection and CD4 Cell Counts of 200 per Cubic Millimeter or Less | NEJM." The New England Journal of Medicine, 11 Sept. 1997.

[2] Hardin, Jo. "Methods in Biostatistics Online Notes." ST47S, 1 May 2019, st47s.com/Math150/Notes/.

[3] "HIV/AIDS." Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, 9 Mar. 2017, www.cdc.gov/hiv/group/gender/men/index.html.

[4] "HIV/AIDS." World Health Organization, World Health Organization, 19 Dec. 2017, www.who.int/features/qa/71/en/.