PSTAT 175 Final Project

Nicolette Phillips, Allison Yih, John Weisner

12/4/2020

Drug Treatment Survival Analysis

Introduction

What makes a successful and effective drug treatment program? This is an area of interest for many health care professionals and people looking to reduce risk of drug relapse. Therefore, finding the optimal conditions for treatment and being able to identify individuals that may be at a higher risk for relapse is an active area of research. The goal of this survival analysis project is to determine the treatment type that is the most effective at reducing drug relapse during and after treatment through a Cox proportional hazards model building process.

Data Overview

The dataset that we will use to analyze the UIS Drug Treatment study is UIS.dat, which is found from R's quantreg library. This was a 5-year(1989-1994) study comprised of two concurrent randomized trials of residential treatment for drug abuse. The purpose of the study was to compare treatment programs of different planned durations designed to reduce drug abuse and to prevent high-risk HIV behavior.[1] In the original dataset, there are 628 observations with 53 of the rows having missing values. Our cleaned dataset has a dimension of 575 rows x 18 columns, including 16 predictors. The variables `ID`, `HC`, `IV`, `RACE`, `TREAT`, `SITE`, `CENSOR`, and `IV3` are all coded as categorical variables. The time variable is defined as the number of days from admission into one of the treatment sites to self-reported return to drug use.

The columns and their values are:

- `ID`: Identification code (1 - 628).
- `AGE`: Age at enrollment to treatment site (Years).
- `BECK`: Beck Depression Score at admission to the treatment site(0.000 - 54.000). This number is based on a self-reporting questionnaire for evaluating the severity of depression in normal and psychiatric populations.
- `HC`: Heroin and/or cocaine use during the 3 months prior to admission to a treatment site(1 = Heroin & Cocaine, 2 = Heroin Only, 3 = Cocaine Only, 4 = Neither Heroin nor Cocaine).
- `IV`: History of IV drug use (1 = Never, 2 = Previous, 3 = Recent).
- `NDT`: Number of prior drug treatments (0 - 40).
- `RACE`: Subject's race (0 = White, 1 = Non-White).
- `TREAT`: Treatment randomization assigned to the subject(0 = Short, 1 = Long). Short versus long represents 3 months versus 6 months duration at Site A and 6 months versus 12 months duration at Site B.
- `SITE`: Treatment Site (0 = A, 1 = B). Treatment Site A was a comparison of 3 and 6 month modified therapeutic communities which incorporated elements of health education and relapse prevention. Treatment Site B was a 6 or 12 month therapeutic community program involving a highly-structured lifestyle in a communal living setting.
- `LEN.T`: Length of stay in treatment days (Admission Date to Exit Date).

[1] David W. Hosmer et al., *Applied Survival Analysis: Regression Modeling of Time to Event Data*, 10

- `TIME` - Time to drug relapse (Days Measured from Admission Date).
- `CENSOR` - Event for treating lost to follow-up as returned to drugs (1 = Returned to Drugs or Lost to Follow-Up, 0 = Otherwise).
- `Y` - log of TIME.

- `ND1` - Component of NDT. This variable and ND2 are the result of a transformation to make NDT a two-dimensional variable.

- `ND2` - Component of NDT. This variable and ND1 are the result of a transformation to make NDT a two-dimensional variable.
- `LNDTFRAC` - Compliance fraction (LEN.T/90 = short trt, LEN.T/180 = long trt).
- `IV3` - Recent IV use (1 = Yes, 0 = No). This is an altered version of IV variable, which transforms recent IV use into a binary variable. This removes the distinction between previous and recent in order to simplify the variable for modelling purposes.

Questions of Interest

The primary objective of this study is to determine the most effective treatment type and duration while controlling for other factors. The model building process we will be performing will seek to answer the questions raised by the study, such as the most effective treatment and whether treatment type/duration has an effect on the time until drug relapse. The aim of the treatment is to extend the time until drug relapse, so time until drug relapse will be the time variable we will be looking into.

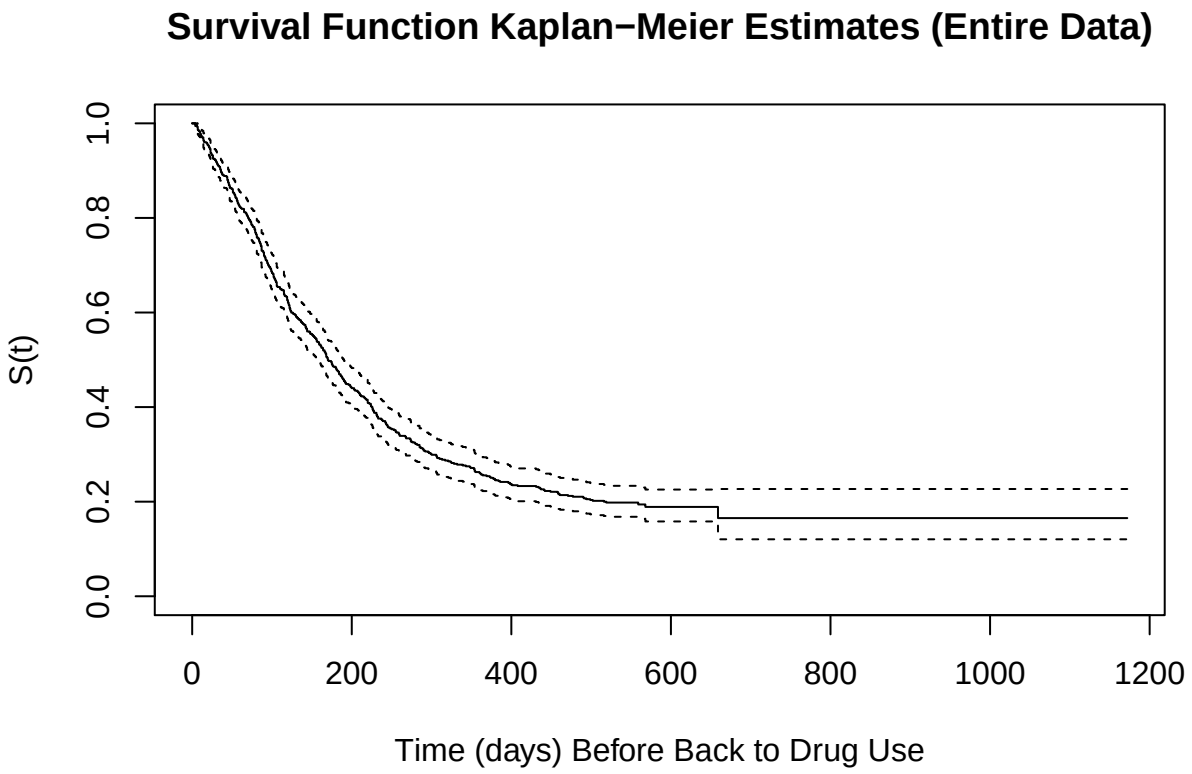
Since the time until drug relapse is measured starting at the admission into the treatment site, we are also interested in looking at the effect of treatment duration on the time to relapse. We will investigate the extent at which the time spent at the treatment site affects the time until relapse through a time varying covariate model. As an extension of this, we would also like to examine the effectiveness of different treatments for extending the time until relapse once the individuals have left the treatment site.

We will take all these factors into consideration while primarily investigating the effectiveness of different treatments on the time until drug relapse.

Cox Proportional Hazards Model

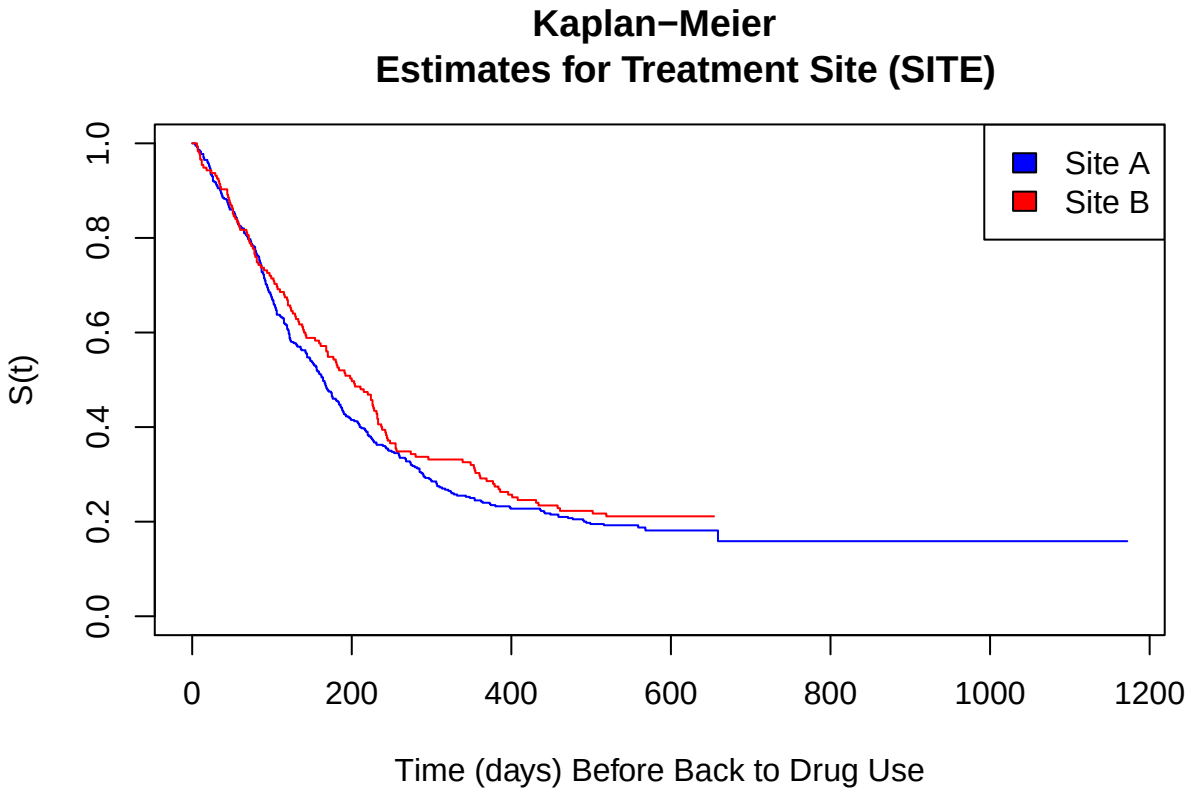
Before building and evaluating our Cox proportional hazards model, we will first plot the Kaplan-Meier estimates to get a preliminary view of our survival data. First we graphed the Kaplan-Meier estimates without splitting our data. This first graph gives a simple curve showing the survival rate for the entire data with a 95% confidence interval.

```
uis.censor <- uis$CENSOR
uis.time <- uis$TIME
uis.surv <- Surv(uis.time, uis.censor)
uis.km.fit <- survfit(uis.surv ~ 1)
plot(uis.km.fit, main = "Survival Function Kaplan-Meier Estimates (Entire Data)",
     xlab = "Time (days) Before Back to Drug Use",
     ylab = "S(t)", conf.int = T)
```



Next, we will look at the Kaplan-Meier graph with the data split between Site A and Site B.

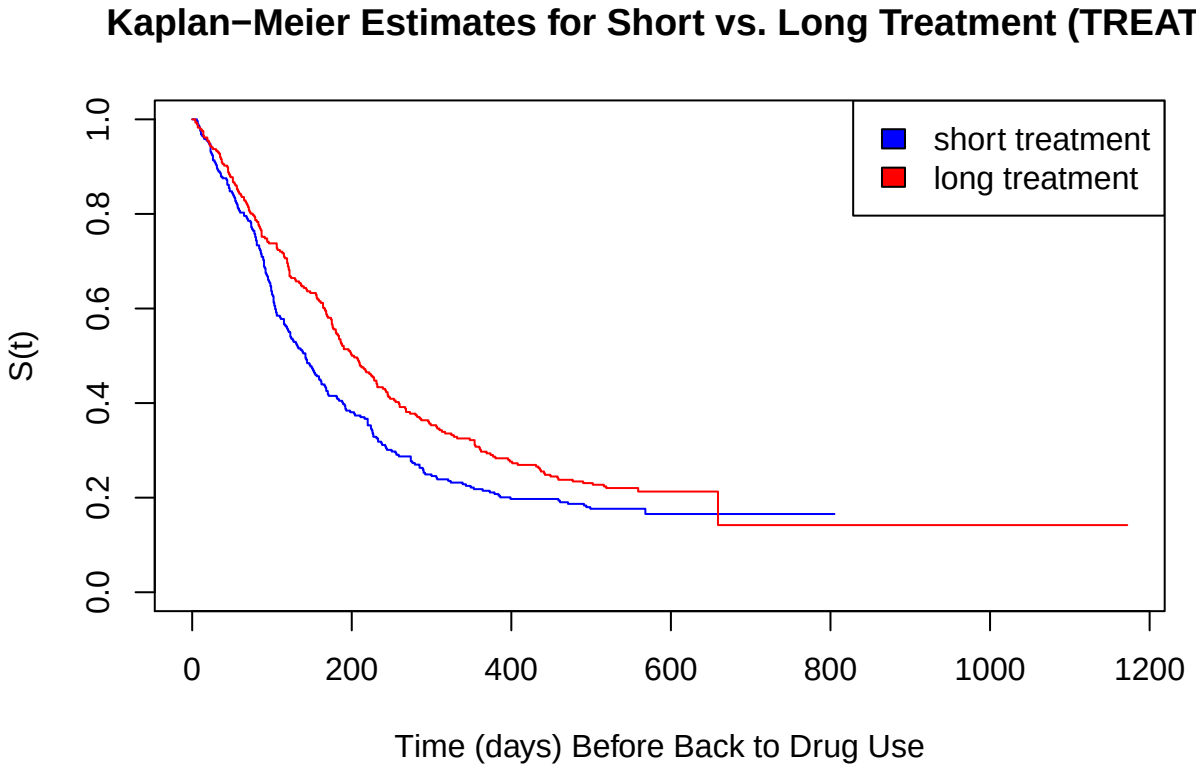
```
plot(survfit(uis.surv ~ uis$SITE), main = "Kaplan-Meier
     Estimates for Treatment Site (SITE)",
     xlab = "Time (days) Before Back to Drug Use",
     ylab = "S(t)",
     conf.int = F,
     col=c("blue", "red"))
legend(x="topright", c("Site A","Site B"), fill=c("blue","red"))
```



We can see that the two lines are relatively similar, with Site A showing lower survival probabilities overall. Intuitively this makes sense because Site A has shorter treatment durations, which we would expect to result in less time to drug relapse compared to Site B.

Next, we will look at the data split up by short vs long treatment.

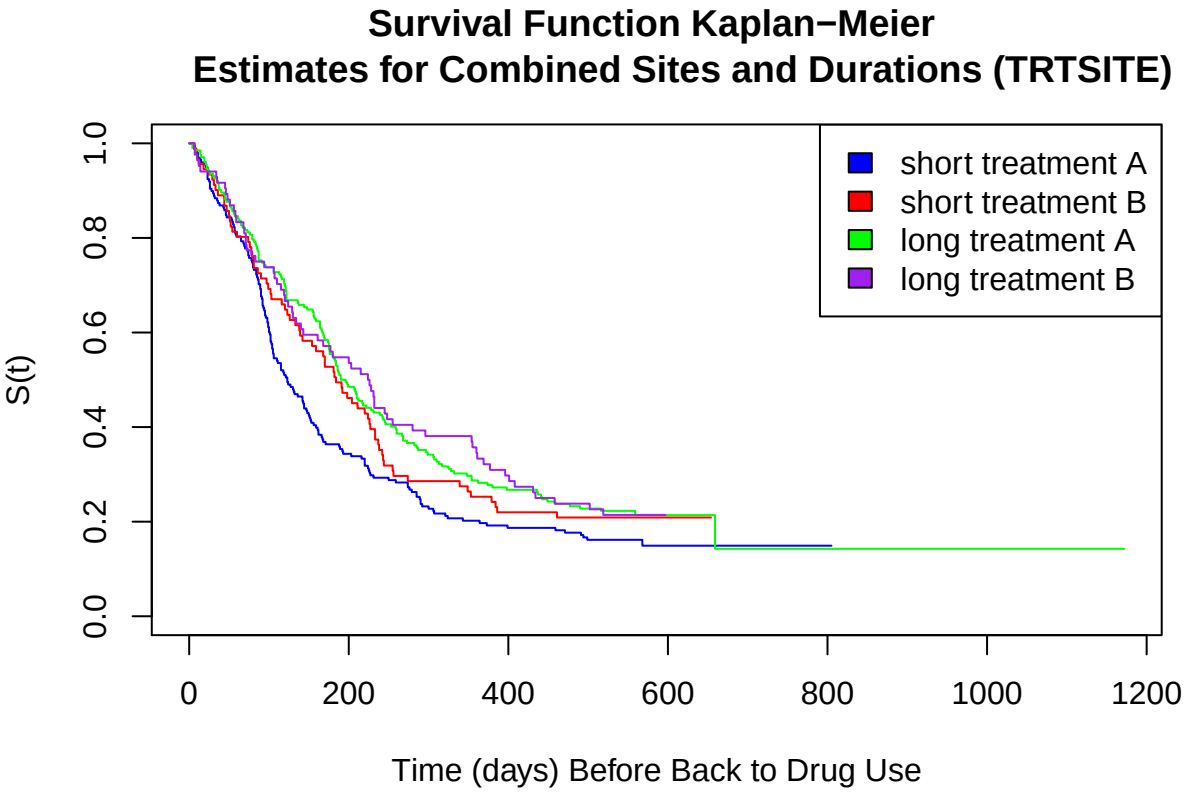
```
plot(survfit(uis.surv ~ uis$TREAT),
     main = "Kaplan-Meier Estimates for Short vs. Long Treatment (TREAT)",
     xlab = "Time (days) Before Back to Drug Use",
     ylab = "S(t)",
     conf.int = F,
     col=c("blue", "red"))
legend(x="topright", c("short treatment","long treatment"), fill=c("blue","red"))
```



As we can see in the graph, the curve for the short treatment is lower than the curve for the long treatment. This is logical since we would expect shorter treatment durations to correspond to sooner drug relapse times.

The `TREAT` and `SITE` variables are interesting because they are interrelated and have a relationship with time. The treatment sites differ both in treatment type and treatment duration, and the `TREAT` variable is nested within the `SITE` variable. This is because it identifies whether the individual had a short or long treatment within their respective site. To better illustrate these two variables, we decided to create a new variable called `TRTSITE` that splits up the data in short treatment A, long treatment A, short treatment B, and long treatment B. Below is the Kaplan-Meier graph for combining treatment method and site into `TRTSITE`.

```
uis$TRTSITE<-paste(uis$TREAT,uis$SITE)
uis$TRTSITE<-as.factor(uis$TRTSITE)
plot(survfit(uis.surv ~ uis$TRTSITE),main = "Survival Function Kaplan-Meier
     Estimates for Combined Sites and Durations (TRTSITE)",
     xlab = "Time (days) Before Back to Drug Use",
     ylab = "S(t)",
     conf.int = F,
     col=c("blue", "red","green","purple"))
legend(x="topright",
       c("short treatment A","short treatment B","long treatment A","long treatment B"),
       fill=c("blue","red","green","purple"))
```



From the graph, it can be seen that there is some difference in the survival rates between the different treatments, with the short treatment A having visually a much lower survival rate. We can also see that long treatment B appears to have a higher survival rate, and short treatment B and long treatment A are generally in the middle. We will further examine the effectiveness of these different treatment types through our Cox proportional hazards model building process.

Next, we will use BIC to create the best Cox PH model for our data. We use forward selection by working from just one covariate up to as many covariates as are necessary. All covariates are tested and the one with the lowest BIC amount is chosen as the first covariate. Then, second covariates are added to the first covariate and tested, and the covariate with the lowest BIC is again added. This process continues until the BIC is at its minimum value, with the appropriate covariates being the ones that best fit our model.

To begin, we will choose our first covariate in our model by finding the lowest BIC of all the models with a single covariate. We will not be including the `TRTSITE` variable in the forward selection process, since this is our primary variable of interest and will be included in the model regardless. We will add `TRTSITE` to our model after determining the additional covariates for our model.

```
mod1<-coxph(uis.surv~uis$RACE)
mod2<-coxph(uis.surv ~ uis$NDT)
mod3<-coxph(uis.surv ~ uis$HC)
mod4<-coxph(uis.surv ~ uis$IV3)
mod5<-coxph(uis.surv ~ uis$AGE)
mod6<-coxph(uis.surv ~ uis$BECK)
BIC(mod1,mod2,mod3,mod4,mod5,mod6)
```

```
##      df      BIC
## mod1  1 5324.685
## mod2  1 5319.745
## mod3  1 5328.657
## mod4  1 5320.598
```

```
## mod5  1 5329.281
## mod6  1 5327.952
```

The first covariate that fits our model is number of prior drug treatments (`NDT`). This covariate had the lowest BIC at 5319.745. Next, we find the second covariate by testing the covariates again when paired with NDT.

```
mod2.1<-coxph(uis.surv ~ uis$NDT+uis$RACE)
mod2.2<-coxph(uis.surv ~ uis$NDT+uis$HC)
mod2.3<-coxph(uis.surv ~ uis$NDT+uis$IV3)
mod2.4<-coxph(uis.surv ~ uis$NDT+uis$AGE)
mod2.5<-coxph(uis.surv ~ uis$NDT+uis$BECK)
BIC(mod2.1,mod2.2,mod2.3,mod2.4,mod2.5,mod2)
```

```
##        df      BIC
## mod2.1  2 5319.744
## mod2.2  2 5324.175
## mod2.3  2 5319.236
## mod2.4  2 5318.419
## mod2.5  2 5321.471
## mod2    1 5319.745
```

The second covariate for our model is `AGE` because the `AGE` model had the lowest BIC at 5318.419. Because the BIC for `NDT+AGE` is lower than the BIC for only `NDT`, we add `AGE` to the model and continue to test for a third covariate.

```
mod2.4.1<-coxph(uis.surv ~ uis$NDT+uis$AGE+uis$RACE)
mod2.4.2<-coxph(uis.surv ~ uis$NDT+uis$AGE+uis$HC)
mod2.4.3<-coxph(uis.surv ~ uis$NDT+uis$AGE+uis$IV3)
mod2.4.4<-coxph(uis.surv ~ uis$NDT+uis$AGE+uis$BECK)
modint1<-coxph(uis.surv ~ uis$NDT*uis$IV3)
modint2<-coxph(uis.surv ~ uis$NDT*uis$AGE)
modint3<-coxph(uis.surv ~ uis$NDT*uis$HC)
BIC(mod2.4,mod2.4.1,mod2.4.2,mod2.4.3,mod2.4.4,modint1,modint2,modint3)
```

```
##          df      BIC
## mod2.4    2 5318.419
## mod2.4.1  3 5318.779
## mod2.4.2  3 5321.392
## mod2.4.3  3 5313.251
## mod2.4.4  3 5320.573
## modint1   3 5325.336
## modint2   3 5322.707
## modint3   3 5329.458
```

The third covariate that we consider adding to the model is recent IV drug use (`IV3`). The model with `IV3` had the lowest BIC value at 5313.251. The BIC for `NDT+AGE+IV3` is lower than the model with just `NDT+AGE`, so it is added into our model. Furthermore, we tested the BIC for interaction terms, but none of these test models resulted in a lower BIC than simply adding the factors. Therefore the interactions will not be added since they do not make a better fit. We will continue this process and test for a fourth covariate.

```
mod2.4.3.1<-coxph(uis.surv ~ uis$NDT+uis$AGE+uis$IV3+uis$RACE)
mod2.4.3.2<-coxph(uis.surv ~ uis$NDT+uis$AGE+uis$IV3+uis$HC)
mod2.4.3.3<-coxph(uis.surv ~ uis$NDT+uis$AGE+uis$IV3+uis$BECK)
modint4<-coxph(uis.surv ~ uis$NDT+uis$IV3*uis$AGE)
modint5<-coxph(uis.surv ~ uis$NDT*uis$IV3+uis$AGE)
BIC(mod2.4.3,mod2.4.3.1,mod2.4.3.2,mod2.4.3.3,modint4,modint5)
```

```
##            df      BIC
```

```
## mod2.4.3    3 5313.251
## mod2.4.3.1  4 5315.799
## mod2.4.3.2  4 5319.387
## mod2.4.3.3  4 5316.700
## modint4     4 5319.248
## modint5     4 5319.366
```

```
#IV3,AGE, and NDT give the best fit for the data
#interaction does not give better fit
```

Although RACE was the next best covariate, it was not a better fit for the model compared to keeping just the three covariate model because the BIC when adding RACE was higher than when it was not added. So, using this process of forward selection, it is found that the covariates history of IV drug use (IV3), age of the patient (AGE), and number of previous drug treatments (NDT) give the best fit for the Cox PH model. Our final model will also include the TRTSITE variable, which is the key variable for addressing our questions of interest.

```
summary(coxph(uis.surv ~ uis$NDT+uis$AGE+uis$IV3+ uis$TRTSITE))
```

```
## Call:
## coxph(formula = uis.surv ~ uis$NDT + uis$AGE + uis$IV3 + uis$TRTSITE)
##
##   n= 575, number of events= 464
##
##                       coef exp(coef)  se(coef)      z Pr(>|z|)
## uis$NDT          0.031265  1.031758  0.008269  3.781 0.000156 ***
## uis$AGE         -0.029449  0.970980  0.008143 -3.616 0.000299 ***
## uis$IV3          0.312660  1.367057  0.103457  3.022 0.002510 **
## uis$TRTSITE0 1 -0.178351  0.836648  0.145588 -1.225 0.220561
## uis$TRTSITE1 0 -0.322899  0.724047  0.112237 -2.877 0.004015 **
## uis$TRTSITE1 1 -0.234024  0.791343  0.148662 -1.574 0.115442
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                exp(coef) exp(-coef) lower .95 upper .95
## uis$NDT           1.0318     0.9692    1.0152    1.0486
## uis$AGE           0.9710     1.0299    0.9556    0.9866
## uis$IV3           1.3671     0.7315    1.1162    1.6744
## uis$TRTSITE0 1    0.8366     1.1952    0.6290    1.1129
## uis$TRTSITE1 0    0.7240     1.3811    0.5811    0.9022
## uis$TRTSITE1 1    0.7913     1.2637    0.5913    1.0590
##
## Concordance= 0.586  (se = 0.014 )
## Likelihood ratio test= 39.93  on 6 df,   p=5e-07
## Wald test            = 41.33  on 6 df,   p=2e-07
## Score (logrank) test = 41.72  on 6 df,   p=2e-07
```

This summary output shows the p-values for each of the variables in our model. We can see that IV3, AGE, NDT, and TRTSITE long A all have p-values that are smaller than .05. Therefore we reject the null and determine that these variables are statistically significant. While some of the treatment sites have p values greater than .05 and are not shown to be statistically significant, since the treatment site is the primary variable we are examining we can allow this. The likelihood ratio test also indicates that this model is sufficient since the p-value is less than .05.

Setting a baseline ratio for the short A treatment, the following hazard ratios were computed in comparison: short B: 0.8366, long A: 0.724, long B: 0.7913. Because none of these ratios are greater than 1, the baseline

hazard rate is the highest. Highest to lowest, the hazard rate is short A, short B, long B, then long A.

For the hazard ratios, the sign of the coefficient tells if the hazard rate for the specific treatment is greater or less than the baseline rate for the short A treatment. Because the coefficients for all the hazard ratios for each of the treatment sites are negative, they all have a smaller hazard rate than the short treatment. Furthermore, the exp(coef) gives the ratio of each of the treatment sites compared to the baseline rate. These values of 0.837 for the short B treatment, 0.724 for the long A treatment, and 0.791 for the long B treatment show that the hazard rate of the short B treatment is 16.3% less than the baseline, the hazard rate of the long A treatment is 27.6% less than the baseline, and the hazard rate of the long B treatment is 20.39% than the baseline. Thus, the hazard ratios show that the long A treatment is most effective, followed by the long B treatment, then the short B treatment, and then the least effective being the short A treatment.

However, the 95% confidence interval contains 1 for the hazard ratio for the short B treatment (0.6290,1.1129) and the long B treatment (0.5913,1.0590), so we cannot conclude that the hazard rates for these two groups are significantly different than the hazard rate of the short A group. The confidence interval for the long A treatment (0.5811,0.9018) does not contain 1 so it can be concluded that the two hazard rates are significantly different. Furthermore, the likelihood ratio test gives a p-value of 5x10^-7, so there is a significant difference in the survival rates of at least one of the groups.

For the covariates, the hazard ratio is 1.3671 for `IV3`, 0.9710 for `AGE`, and 1.0318 for `NDT`. The confidence interval for `IV3` is (1.1162, 1.6744), for `AGE` is (0.9556, 0.9866), and for `NDT` is (1.0152, 1.0486). These confidence intervals represent the range of values that we are 95% confident the hazard proportion resides in. Since none of the confidence intervals include 1, we can determine that these covariates are statistically significant, which we expected from our previous p-test results. All the confidence intervals are quite narrow, so we can assume that the estimated hazard ratios for these covariates are reliable.

Checking Assumptions

Having built our model under the assumption that proportional hazards requirements hold, we must now verify that we can be operating under these assumptions. First, we will look at the log-log plots for each of the variables in our model to visually confirm that our assumption holds. Since both `AGE` and `NDT` are originally continuous variables, we will organize their values into bins in order to analyze their log-log graphs. We split `AGE` into younger than 32 and greater than or equal to 32, which is based on the median value for age. `NDT` was split into 0, 1-5, or greater than 5 previous drug treatments.

```
#alter age and ndt variables into groups for log-log analysis
summary(uis$AGE)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   20.00   27.00   32.00   32.38   37.00   56.00
```

```
uis$AGE.cat <- cut(uis$AGE, breaks=c(0, 33, 100), right = FALSE, labels= c("<32", ">=32"))
table(uis$AGE.cat)
```

```
##
##  <32 >=32
##  292  283
```

```
summary(uis$NDT)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   1.000   3.000   4.543   6.000  40.000
```

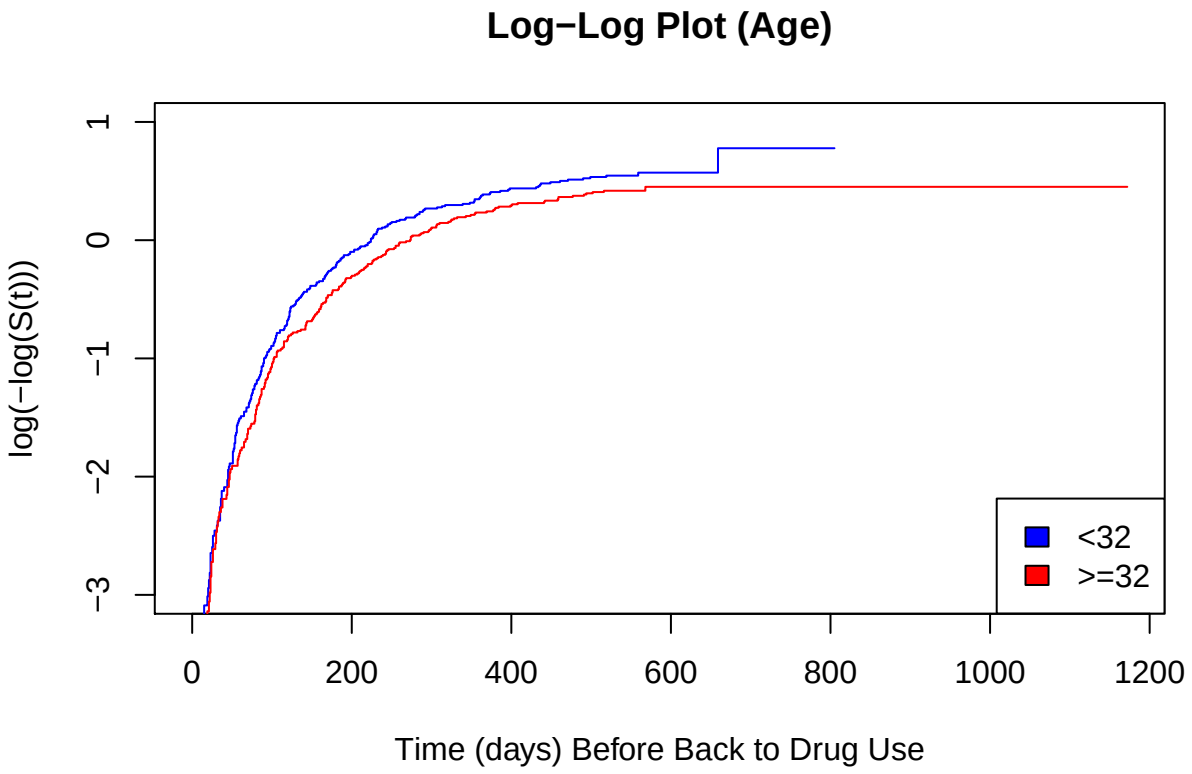
```
uis$NDT.cat <- cut(uis$NDT, breaks=c(0,1, 6, 50), right = FALSE,
                   labels= c("0", "1-5", ">5"))
table(uis$NDT.cat)
```

```
##
##   0 1-5  >5
```

```
##  79 337 159
```

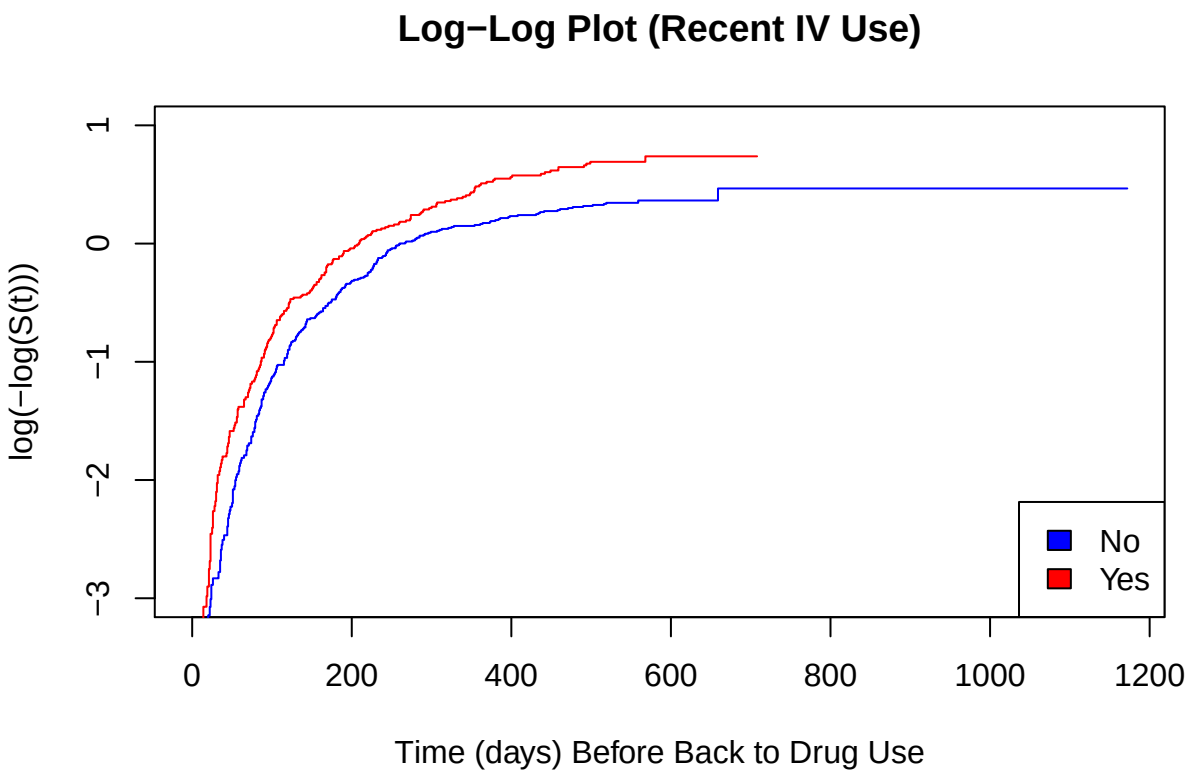
After sorting the continuous variables, we generated the log-log plot of each variable to check our assumptions.

```
#plot log-log graphs
#Age
cloglog <- function(x){log(-log(x))}
uis.age.fit<- survfit(uis.surv~uis$AGE.cat)
plot(uis.age.fit, fun=cloglog, xlab= "Time (days) Before Back to Drug Use",
     ylab= "log(-log(S(t)))",
     main= "Log-Log Plot (Age)",
     col=c("blue", "red"),
     ylim=c(-3, 1))
legend(x="bottomright", c("<32", ">=32"), fill=c("blue","red"))
```



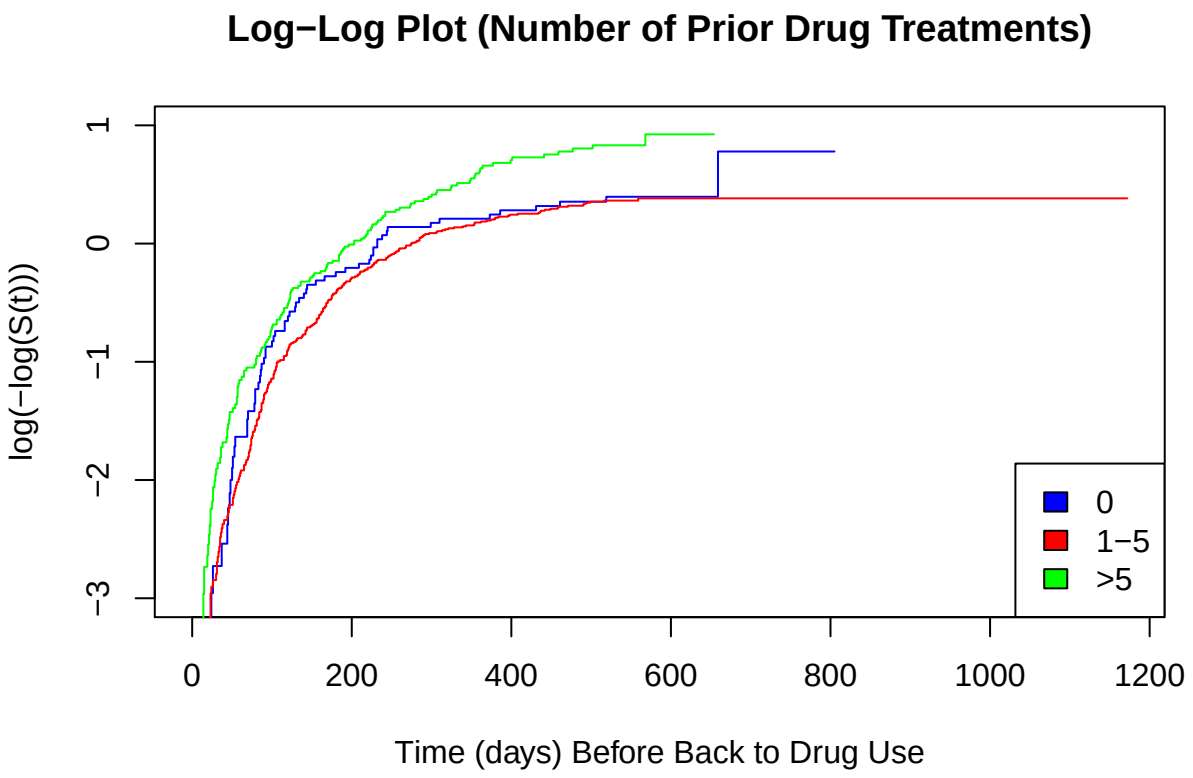
Looking at the log-log plot for age, we can see that the two age groups are generally parallel. There is some overlap at the far left side of the graph, but we can let this pass since there is a lot of inherent variance in this area. Therefore, we can assume that `AGE` fulfills the proportional hazards requirement.

```
#IV3
uis.iv3.fit<- survfit(uis.surv~uis$IV3)
plot(uis.iv3.fit, fun=cloglog, xlab= "Time (days) Before Back to Drug Use",
     ylab= "log(-log(S(t)))",
     main= "Log-Log Plot (Recent IV Use)",
     col=c("blue", "red"),
     ylim=c(-3, 1))
legend(x="bottomright", c("No", "Yes"), fill=c("blue","red"))
```



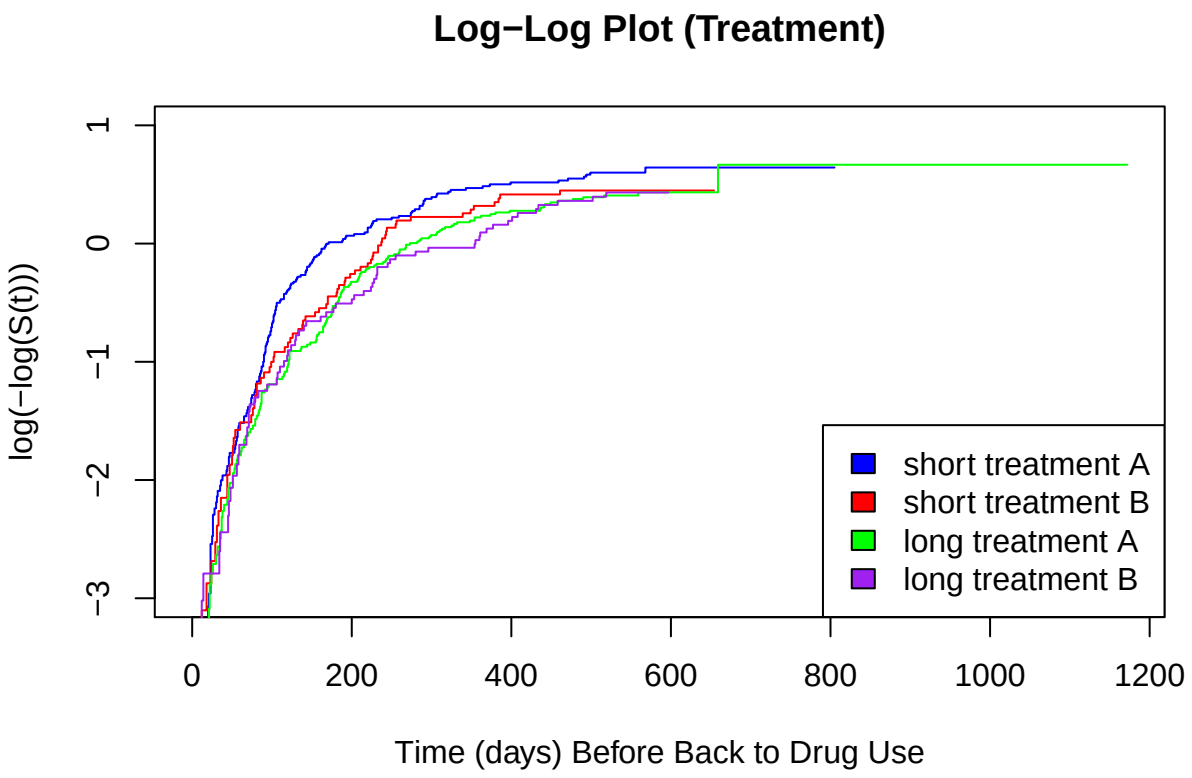
Now looking at the log-log plot for recent IV use, we can see that the two groups are again mostly parallel. The far left side of the graph does look a little concerning, but again we can let this pass since there is a lot of inherent variance in this area. Therefore, we can assume that `IV3` also fulfills the proportional hazards requirement.

```
#NDT
uis.ndt.fit<- survfit(uis.surv~uis$NDT.cat)
plot(uis.ndt.fit, fun=cloglog, xlab= "Time (days) Before Back to Drug Use",
     ylab= "log(-log(S(t)))",
     main= "Log-Log Plot (Number of Prior Drug Treatments)",
     col=c("blue", "red", "green"),
     ylim=c(-3, 1))
legend(x="bottomright", c("0","1-5", ">5"),
       fill=c("blue","red", "green"))
```



From the log-log graph for the number of previous drug treatments, we can see some overlap between the red and blue lines. This suggests that perhaps the effect of 0 drug treatments and 1-5 drug treatments is actually very similar. The blue line will also naturally have more variance since this group of observations is small. Therefore, if we instead look at the difference in slopes between the green line and and blue/red lines, we can see that they are mostly parallel. This is sufficient to declare that `NDT` complies with our proportional hazards assumption.

```
#TRTSITE
uis.trtsite.fit<- survfit(uis.surv~uis$TRTSITE)
plot(uis.trtsite.fit, fun=cloglog, xlab= "Time (days) Before Back to Drug Use",
     ylab= "log(-log(S(t)))",
     main= "Log-Log Plot (Treatment)",
     col=c("blue", "red", "green", "purple"),
     ylim=c(-3, 1))
legend(x="bottomright",
       c("short treatment A","short treatment B","long treatment A","long treatment B"),
       fill=c("blue","red","green","purple"))
```



Finally, looking at the log-log plot for treatment type, we can see a few concerning characteristics. There is a lot of overlap on the left side, which we will disregard for now. Aside from that, we can see overlap between the short treatment B, long treatment A, and long treatment B. We can see that short treatment A and long treatment B do look relatively seperate and parallel. Short treatment B and long treatment A look the most similar, likely due to the fact that they have the same treatment duration. If we consider this, we can look over some of the overlap and further evaluate the proprtional hazards assumption through testing. Since this is the primary variable we are interested in, we should include this in our model despite having some relationship with time.

To confirm our results from the log-log graphs, we will use the cox.zph function to test the proportional hazards assumption of our model. This test will allow us to verify that any inconsistencies in the log-log plots are indeed inconsequential, and also make sure the continuous versions of `AGE` and `NDT` still comply.

```
cox.zph(coxph(uis.surv ~ uis$NDT+uis$AGE+uis$IV3+ uis$TRTSITE))
```

```
##                chisq df    p
## uis$NDT      0.46876  1 0.49
## uis$AGE      0.61672  1 0.43
## uis$IV3      0.00192  1 0.97
## uis$TRTSITE 3.91223  3 0.27
## GLOBAL       4.94550  6 0.55
```

From the output above, each of the variables in our model have p-values greater than our α significance level of 0.05. We fail to reject the H0 and conclude that because there is no statistically significant difference with the proportional hazards model, our Cox proportional hazard model is appropriate. In particular, our `AGE`, `TRTSITE`, and `NDT` variables are not statistically different with the proportional hazards model, so we can move forward despite the log-log graph issues.

The first objective of this study was to find out if the treatment type affects the time until drug relapse for a patient and which treatment type is the most effective. To find if there was a significant difference, a Cox PH model was fitted with the covariates age, number of prior drug treatments, and a binary variable for recent history of IV drug use. From the likelihood ratio test included in the coxph calculation, it was concluded that

there is a significant difference between the hazard rates for at least one of the different treatments. This significant difference can be attributed to the difference between the hazard rates of the long A treatment and the short A treatment, with the long A treatment having a significantly lower hazard rate. This is echoed in the p-values of the coxph test, with the long A treatment being the only one having a statistically significant p-value. So treatment type does play a role in the time until drug relapse, with patients from the long treatment from site A being shown to have lower hazard rates and therefore longer times until drug relapse. Therefore, the long treatment from site A is the most effective treatment according to our model.

Cox Proportional Hazards Model with Time Varying Covariates

One of the potential issues for our Cox PH model goes back to the `TRTSITE` variable and its relationship with time. Since individuals are sorted into different lengths of treatment, it brings up the question of whether the instances of relapse are significantly lower while still at the treatment site. The effect of being in treatment on drug relapse time is not accounted for in our original model and can reduce the reliability of the model. Creating a model that separates the instances of relapse while still in the treatment site as well as after leaving the treatment site could help address the issues with our original Cox PH model.

By creating a time-dependent Cox proportional hazards model to separate the instances of relapse while being in and leaving the treatment site, this could help us better understand the extent at which treatment length affects the overall effectiveness of treatment.

To fit our data to a time-dependent model, we input start, stop, and event data into the Surv() function. Our event variable is set to a binary variable called `inTreatment`, which describes if the subject is in the treatment by checking that the drug relapse time equals the treatment length of stay. The `inTreatment` variable is coded to be 0 if the individual is in treatment and 1 otherwise. Our stop variable, `time2`, is equal to the time to drug relapse.

```
uis$inTreatment <- uis$CENSOR
uis$inTreatment[uis$TIME == uis$LEN.T] <- 0
uis$time2 <- uis$TIME  #time to relapse
uis$time2[uis$time2 == 0] <- 0.25
```

Because our subjects were not always in treatment for the specified time, we split our data based off the treatment length of stay for increased precision.

In order to cut on different times for each treatment duration, we first divided our dataset by treatment length, and then we split our data into two observations per subject based on their time in treatment versus out of treatment. We needed to split our data in the four different durations (3 month A, 6 month A, 6 month B, 12 month B) to implement different cut times for each group. Once we had each group of data split, we merged the groups together to get our complete split data that we can work with in our Cox PH model.

```
uis.short.A<- subset(uis, TRTSITE=="0 0")
uis.long.A<- subset(uis, TRTSITE=="1 0")
uis.short.B<- subset(uis, TRTSITE=="0 1")
uis.long.B<- subset(uis, TRTSITE=="1 1")

uis.split.short.A <- survSplit(Surv(time2, inTreatment) ~ IV3 + AGE + NDT + TRTSITE,
                               data = uis.short.A, cut = 91,
                               id="Subject", episode="Episode")
uis.split.long.A <- survSplit(Surv(time2, inTreatment) ~ IV3 + AGE + NDT + TRTSITE,
                              data = uis.long.A, cut = 182,
                              id="Subject", episode="Episode")
uis.split.short.B <- survSplit(Surv(time2, inTreatment) ~ IV3 + AGE + NDT + TRTSITE,
                               data = uis.short.B, cut = 182,
                               id="Subject", episode="Episode")
uis.split.long.B <- survSplit(Surv(time2, inTreatment) ~ IV3 + AGE + NDT + TRTSITE,
                              data = uis.long.B, cut = 365,
```

```
                                id="Subject", episode="Episode")

uis.split <- rbind(uis.split.short.A, uis.split.long.A, uis.split.short.B,
                   uis.split.long.B)
```

Afterward, we use a Cox proportional hazards model to fit `IV3`, `AGE`, `NDT`, `TRTSITE`, which are the covariates that our original model identified as statistically significant. Additionally, we stratify the episode variable and place it in an interaction term with `TRTSITE` so we can get estimates of the baseline hazard function for each episode level. By implementing this time change effect, only the individuals in the group with the time that is between the start and stop times are included.

```
cox1 <- coxph(Surv(tstart, time2, inTreatment) ~ IV3 + AGE + NDT +
                      TRTSITE*strata(Episode), uis.split)
summary(cox1)
```

```
## Call:
## coxph(formula = Surv(tstart, time2, inTreatment) ~ IV3 + AGE +
##     NDT + TRTSITE * strata(Episode), data = uis.split)
##
##   n= 892, number of events= 393
##
##                                        coef exp(coef)  se(coef)      z
## IV3                                 0.227002  1.254833  0.112727  2.014
## AGE                                -0.027257  0.973111  0.008812 -3.093
## NDT                                 0.025106  1.025424  0.009529  2.635
## TRTSITE0 1                          0.183108  1.200944  0.254546  0.719
## TRTSITE1 0                          0.021554  1.021788  0.223262  0.097
## TRTSITE1 1                         -0.005966  0.994052  0.270730 -0.022
## TRTSITE0 1:strata(Episode)Episode=2 -0.095227  0.909166  0.358005 -0.266
## TRTSITE1 0:strata(Episode)Episode=2  0.099863  1.105020  0.298232  0.335
## TRTSITE1 1:strata(Episode)Episode=2  0.853410  2.347639  0.479115  1.781
##                                    Pr(>|z|)
## IV3                                 0.04404 *
## AGE                                 0.00198 **
## NDT                                 0.00842 **
## TRTSITE0 1                          0.47192
## TRTSITE1 0                          0.92309
## TRTSITE1 1                          0.98242
## TRTSITE0 1:strata(Episode)Episode=2 0.79024
## TRTSITE1 0:strata(Episode)Episode=2 0.73774
## TRTSITE1 1:strata(Episode)Episode=2 0.07488 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                                    exp(coef) exp(-coef) lower .95 upper .95
## IV3                                   1.2548     0.7969    1.0061    1.5651
## AGE                                   0.9731     1.0276    0.9564    0.9901
## NDT                                   1.0254     0.9752    1.0065    1.0448
## TRTSITE0 1                            1.2009     0.8327    0.7292    1.9778
## TRTSITE1 0                            1.0218     0.9787    0.6597    1.5827
## TRTSITE1 1                            0.9941     1.0060    0.5847    1.6899
## TRTSITE0 1:strata(Episode)Episode=2    0.9092     1.0999    0.4507    1.8339
## TRTSITE1 0:strata(Episode)Episode=2    1.1050     0.9050    0.6159    1.9826
## TRTSITE1 1:strata(Episode)Episode=2    2.3476     0.4260    0.9179    6.0042
```

```
##
## Concordance= 0.566  (se = 0.017 )
## Likelihood ratio test= 21.25  on 9 df,   p=0.01
## Wald test            = 22.16  on 9 df,   p=0.008
## Score (logrank) test = 22.37  on 9 df,   p=0.008
```

```
anova(cox1)
```

```
## Analysis of Deviance Table
##  Cox model: response is Surv(tstart, time2, inTreatment)
## Terms added sequentially (first to last)
##
##                           loglik  Chisq Df Pr(>|Chi|)
## NULL                     -2085.3
## IV3                      -2084.0 2.7513  1   0.097173 .
## AGE                      -2080.3 7.2794  1   0.006975 **
## NDT                      -2077.2 6.3078  1   0.012021 *
## TRTSITE                  -2076.6 1.0697  3   0.784397
## TRTSITE:strata(Episode) -2074.7 3.8415  3   0.279093
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the summary output, we see that `IV3`, `AGE`, and `NDT` are all significant predictors on the response since their p-values are less than .05 and their hazard ratio 95% confidence intervals do not contain 1. However we can see that the treatment type and the episode are not statistically significant since their p-values are greater than .05 and their confidence intervals include 1. The results from the ANOVA table also show that `TRTSITE` and the stratified Episode variable are not significant. From this, we can conclude that there is no significant difference in the time until relapse before and after leaving treatment.

To check if our Cox proportional hazard model is appropriate, we run a goodness of fit test on our fitted Cox proportional hazard model.

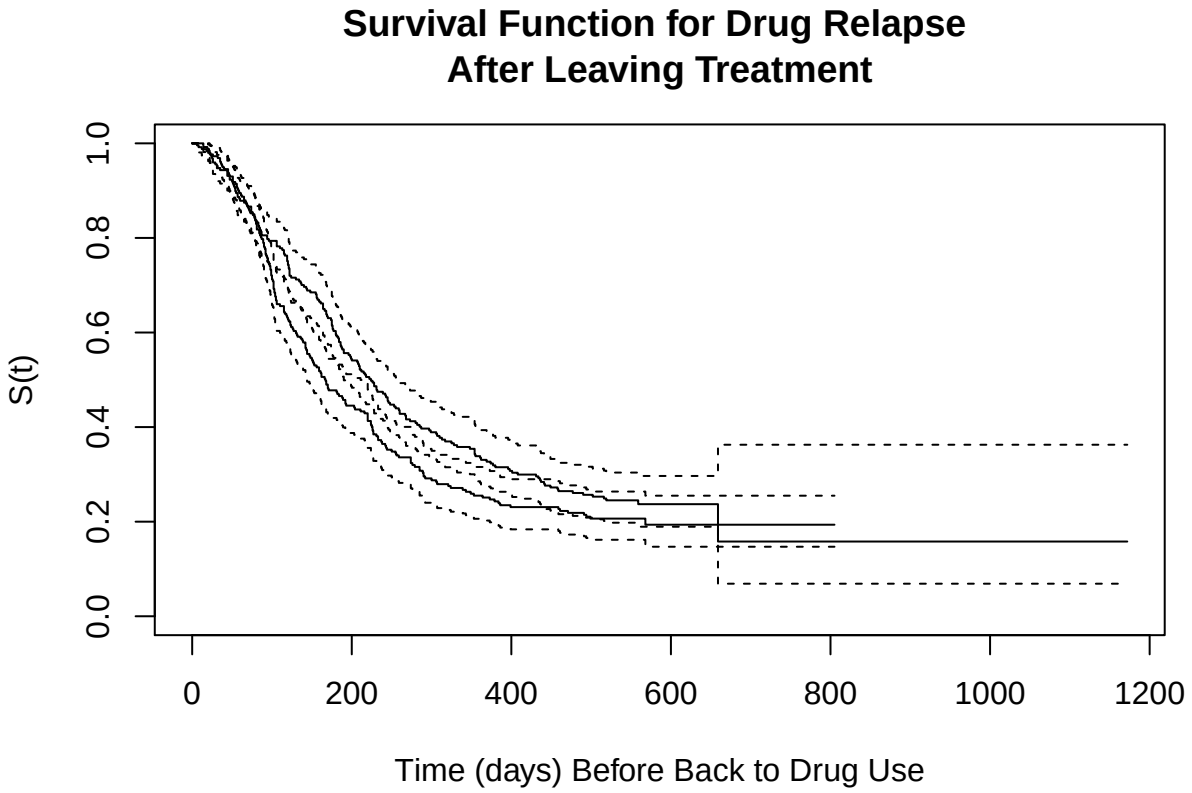
```
zp <- cox.zph(cox1, transform = "rank")
zp
```

```
##                          chisq df     p
## IV3                       4.04  1 0.045
## AGE                       2.18  1 0.140
## NDT                       3.31  1 0.069
## TRTSITE                   1.78  3 0.620
## TRTSITE:strata(Episode)   1.11  3 0.775
## GLOBAL                    8.95  9 0.442
```

From our goodness of fit test output, all of the covariates in our model, except `IV3`, have p-values greater than our α significance level of 0.05. We reject the H0 and conclude that because there is no statistically significant difference with the proportional hazards model for `IV3`, using our Cox proportional hazard model may not be appropriate.

Another question we explore using our time-varying covariates is whether the length of treatment affects the drug relapse time for subjects who have left the treatment, specifically those who did not leave during the study because they relapsed.

```
surv_afterTrt <- Surv(uis$TIME[uis$TIME - uis$LEN.T > 0],
                      uis$CENSOR[uis$TIME - uis$LEN.T > 0])
km_afterTrt <- survfit(surv_afterTrt ~ uis$TREAT[uis$TIME - uis$LEN.T > 0])
plot(km_afterTrt, conf.int = TRUE, main = "Survival Function for Drug Relapse
After Leaving Treatment", xlab = "Time (days) Before Back to Drug Use",
ylab = "S(t)")
```



From our Kaplan-Meier curves, there seems to be a slight difference between the drug relapse time after leaving the treatment and treatment lengths, as the survival curves overlap with the 95% confidence interval estimates for the treatment length groups. For the long treatment group, the KM curves seem to depict that their survival time was longer.

To check whether the survival curves for the short and long treatments are actually different, we use the survdiff() function to calculate the log-rank test result.

```
survdiff(surv_afterTrt ~ uis$TREAT[uis$TIME - uis$LEN.T >0])
```

```
## Call:
## survdiff(formula = surv_afterTrt ~ uis$TREAT[uis$TIME - uis$LEN.T >
##     0])
##
##                                         N Observed Expected (O-E)^2/E (O-E)^2/V
## uis$TREAT[uis$TIME - uis$LEN.T > 0]=0 247      197      177      2.18         4
## uis$TREAT[uis$TIME - uis$LEN.T > 0]=1 257      196      216      1.79         4
##
##  Chisq= 4  on 1 degrees of freedom, p= 0.05
```

From our log-rank test output, we get a p-value that is equal to our alpha significance level 0.05, so we reject H0 and conclude that there is a statistically significant difference between the survival estimate curves.

Then, to estimate the hazard ratio for the different treatment lengths, we fit our survival object to a Cox PH model.

```
cox_afterTrt <- coxph(surv_afterTrt ~ uis$TREAT[uis$TIME - uis$LEN.T >0] )
summary(cox_afterTrt)
```

```
## Call:
## coxph(formula = surv_afterTrt ~ uis$TREAT[uis$TIME - uis$LEN.T >
##     0])
```

```
##
##   n= 504, number of events= 393
##
##                                       coef exp(coef) se(coef)      z Pr(>|z|)
## uis$TREAT[uis$TIME - uis$LEN.T > 0] -0.2020    0.8171   0.1010 -1.999   0.0456
##
## uis$TREAT[uis$TIME - uis$LEN.T > 0] *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                                     exp(coef) exp(-coef) lower .95 upper .95
## uis$TREAT[uis$TIME - uis$LEN.T > 0]    0.8171      1.224    0.6703    0.9961
##
## Concordance= 0.533  (se = 0.013 )
## Likelihood ratio test= 3.99  on 1 df,   p=0.05
## Wald test            = 3.99  on 1 df,   p=0.05
## Score (logrank) test = 4.01  on 1 df,   p=0.05
```

From the summary output, we see that the 95% hazard ratio confidence interval for the long treatment group is between 0.6703 and 0.9961, and the hazard ratio is 0.8171. This indicates that individuals having a longer treatment length would be 18% less likely to go to relapse than subjects in the shorter treatment length. Additionally, because the interval does not contain 1, this would mean that the long treatment group's hazard rate is lower than than the group with a short treatment length. However, because the upper bound of the confidence interval is close to 1, this could potentially mean that there is no difference between the time to relapse after leaving treatment between the two types of lengths.

Finally, to check if our Cox proportional hazard model is appropriate, we run a goodness of fit test on our fitted Cox proportional hazard model.

```
cox.zph(cox_afterTrt)
```

```
##                                     chisq df     p
## uis$TREAT[uis$TIME - uis$LEN.T > 0]  2.86  1 0.091
## GLOBAL                               2.86  1 0.091
```

From our goodness of fit test output, the model's predictor has a p-value greater than our α significance level of 0.05, so we reject the H0 and conclude that using our Cox proportional hazard model is appropriate.

One of the reasons for using a time-varying model was to determine the treatment's effectiveness of extending time until relapse after an individual left the treatment site. From our 95% confidence intervals for each of the treatment sites, they contain 1. Therefore, we could not conclude that there is a distinct difference between the treatment sites and time until relapse when using this time split model.

Another purpose of this study was to look at how much each treatment site impacted the time until relapse. By looking at the hazard ratios, we conclude that the time spent at the long B treatment site would increase the time until relapse, given that its hazard ratio is smaller than 1. However, the other treatment sites' hazard ratios were greater than 1, so we could not conclude that relapse time would lengthen significantly. However the estimated hazard ratios are also not reliable due to the confidence intervals. We also could not conclude that the time spent in treatment has a measurable effect on time to relapse since our model showed no significant difference in the time to relapse from before and after leaving treatment.

Additionally, because our Cox PH assumption was not met, this would mean that our time-varying model is not reliable. By violating the assumption, our parameters that our Cox model is estimating could possibly not be a meaningful way to quantify the difference between the treatment site groups, as the relationships between the covariates may be more complicated than the Cox PH model's linear combination assumption.

Conclusion

After modeling the UIS data with a simple Cox PH model and also an extended time varying Cox PH model, we have determined several characteristics about the effectiveness of the treatment sites and how they affect the time until relapse. We first built a Cox proportional hazards model through forward selection, evaluated this model, and checked our proportional hazards assumptions. In regards to our questions of interest, this model showed that the treatment type does affect the time until relapse and we concluded that the long treatment at site A was the most effective at increasing the time until relapse. This model also complied with our proportional hazards assumptions. We hypothesized that this model was somewhat lacking due to the time-varying aspects of our `TRTSITE` variable, so we also evaluated a time varying Cox PH model.

The time varying model was used to to separate the instances of relapse while being in and leaving the treatment site, in order to help us understand the extent at which treatment length affects the overall effectiveness of treatment. After splitting the UIS data on before and after treatment and fitting the new model, we evaluated the results and checked our assumptions. The time varying model p-values, hazard ratio confidence intervals, and ANOVA tables all concluded that there was not a significant difference in the time to relapse from before and after treatment. We also found that our proportional hazards assumptions fell through when looking at the goodness of fit test. In regards to our questions of interest, we could not conclude that there is a distinct difference between the treatment sites and time until relapse when using this time split model. When evaluating the hazard ratios, we could see that the long treatment at site B was the most effective, but this is not a reliable observation when we look at the confidence intervals for the hazard ratios.

When comparing the two different models in terms of reliability and the conclusions that can be drawn from them, the simple Cox PH model is superior. Our original model complies with the proportional hazards assumptions and contains more statistically significant covariates. The time varying model raises concerns about our assumptions and also shows statistical insignificance for many of our important covariates. The biggest takeaway from our time varying model is the conclusion that the time to relapse is not affected by the individual being in treatment or not. We wanted to explore the time varying model because we thought our original model was leaving out a significant interaction between treatment status and time to relapse. When this was shown to not be the case, our original model gained legitimacy. Therefore our final conclusions from this analysis is that treatment type does affect the time until relapse, and under our model we can see that the long treatment as Site A is the most effective form of treatment for increasing time to drug relapse.

References

[1] Hosmer, D.W. and Lemeshow, S. (1998) Applied Survival Analysis: Regression Modeling of Time to Event Data, Table 1.3, John Wiley and Sons Inc., New York, NY

[2] Therneau, Terry, et al. “Using Time Dependent Covariates and Time Dependent Coefficients in the Cox Model.” Cran R-Project, 25 Sept. 2020.

[3] Therneau, T (2020). A Package for Survival Analysis in R. R package version 3.2-7, https://CRAN.R-project.org/package=survival.