PHST 683 - Survival Analysis
Final Project #2

Britt Kravets
University of Louisville
Department of Biostatistics

Introduction

Our area's Department of Child Protective Services (CPS) is testing a new program for victims of physical child abuse to determine if it is superior to the standard program, with the goal of reducing the rate of further instances of physical child abuse. The pilot program involves expanded in-person family counseling and access to a 24-hour telephone counseling hotline. Families who have had at least one call to CPS were included in the study. After a CPS contact event for suspected child abuse or neglect, the families were randomized into two groups, receiving either the pilot intervention or the standard intervention.

The families were then tracked for up to one year and the occurrence of secondary instances of physical child abuse were recorded. Additionally, data was recorded for whether or not each child was removed from the home and placed in a foster home and the time at which the removal occurred for those children. For children removed from the home at any time during the year of observation, investigators also measured whether or not the child was returned from home within the year and the time of return to home. Demographic information that may be related to occurrences of child abuse were also recorded for the families, including the age of the child at enrollment, biological sex of the child, whether the child was of a minority race, whether the family was living below the poverty threshold, whether there was a criminal history for anyone in the household, whether there was a history of substance abuse for anyone in the household, whether the initial event was the family's first contact with CPS, what metropolitan area region they resided in (north, south), and the agency of contact for the initial event (state, local). Children not experiencing physical child abuse within one year, or who moved away, were right censored.

Methods

Cox regression was performed with a time-dependent covariate (time in which the child was removed from their home) in order to determine whether the pilot program was better than the standard program at preventing secondary occurrences of child abuse. The analysis was done using R version 4.2.1 and RStudio version 2022.07.1+554. The code from the R analysis is included in Appendix I. The level of significance for all analyses was $\alpha = 0.05$.

First, I restructured the data set to recode the variables *removal*, *time.removal*, *return*, and *time.return* into a time dependent variable, *outofhome*. After restructuring, the data set had possible multiple lines of data for a single child. Each line for a single child indicated intervals of time where the child may have been out of home or in the home, or where a secondary occurrence of child abuse may have occurred or not.

Next, I analyzed the data from each treatment group (Pilot or Standard program) to ensure that there was a similar proportion of each variable in each group. For example, I checked whether there was a similar ratio of males to females in both the Pilot group and the Standard group. Then the Cox regression was performed (Appendix II). I started with a regression that included the time until a secondary instance of abuse as the dependent variable and the program (Pilot or Standard) as the only variable. Then I systematically added variables one by one until the model contained all variables that were statistically significant to the time until a secondary instance of abuse. Because our goal was to compare groups, I used a manual method of adding a covariate to the model and then testing it against the baseline model using ANOVA, rather than using the automated AIC testing in R. I then checked for interactions between the program group and the remaining variables. Finally, I checked that all the remaining covariates met the proportional hazards assumption (Appendix III), which is necessary in Cox regression, and determined if any stratification was necessary.

Results

I found that the intervention groups (Pilot and Standard) were balanced with respect to the other covariates. For the binary covariates (sex, minority, poverty, substance abuse, criminal history, first CPS contact, removal from home, region, agency) I used a two-sample test of proportions. As seen in Table 1, all p-values are greater than 0.05, indicating that there was not a statistically significant difference in proportions between the pilot group and the standard group for the binary covariates. For the one continuous covariate (age of the child), I first tested whether age was normally distributed in the Pilot program and the Standard program using the Shapiro-Wilk test (Table 2). I found that age was not normally distributed, and thus I should use a Wilcoxon Rank Sum test in order to determine whether the mean age was similar in both program groups. The Wilcoxon Rank Sum test had a p-value greater than 0.05 ($p = 0.6662$) indicating that there was not a statistically significant difference between the mean age between the pilot group and the standard group (Table 3). Thus, I considered the two programs to be balanced in regard to the other covariates.

Table 1. Two-sample test of proportion for binary covariates

	Two-Sample Test of Proportion
Pilot Program vs. Standard Program	p-value
Sex	0.6297
Minority	0.6539
Poverty	0.3435
Substance Abuse	0.0624
Criminal History	0.3395
First CPS Contact	0.4360
Removed from Home	0.3324
Region	0.8020
Agency	0.5226

Table 2. Shapiro-Wilk Test of normality for the covariate age

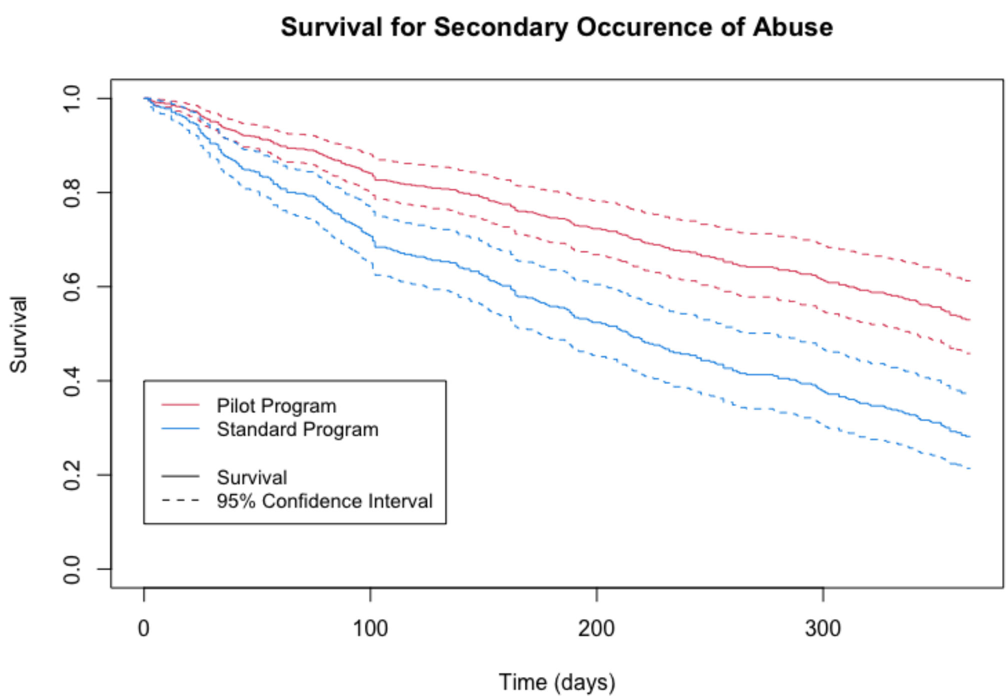
	Shapiro-Wilk Normality Test
	p-value
Age - Pilot Program	< 0.001
Age - Standard Program	< 0.001

Table 3. Wilcoxon Rank Sum test for comparison of mean age

	Wilcoxon Rank Sum Test
	p-value
Age - Pilot Program vs. Standard Program	0.6662

After controlling for age, criminal history in the household, first CPS contact, removal from the home, and substance about in the home, the Pilot Program was significantly better at preventing a secondary occurrence of abuse when compared to the Standard Program (hazard ratio = 1.9935, 95% CI = (1.5659, 2.5378), $p < 0.001$). Those in the Standard Program have almost double the hazard (99% increase) of experiencing a secondary occurrence of abuse compared to the pilot program.

Figure 1. Survival of secondary occurrence of abuse for pilot program vs. standard program



There were no significant interactions found between the program variable and the any of the other covariates. When testing for proportional hazards, all covariates were found to meet the proportional hazards assumption, so stratification was not necessary. However, plots of the model stratified by region (Figure 2) and agency (Figure 3) are included for additional information. Figure 1 shows survival for the pilot group in red and for the standard group in blue while holding all other covariates constant, with the dotted lines being the 95% confidence intervals. We can see that survival for the pilot group is consistently higher than that of the standard group, indicating that the pilot group experienced fewer secondary occurrences of child abuse over one year than the standard group. Figure 2 shows the survival for the pilot group vs. the standard group stratified by region, while holding all other covariates constant, and figure 3 shows the survival for the pilot group vs. the standard group stratified by agency, while holding all other covariates constant. As we can see, these also concur with the unstratified analysis, the pilot program clearly has better survival than the standard program in all cases, no matter whether the region is north or south, nor whether the agency is state or local.

Figure 2. Survival of secondary occurrence of abuse for pilot program vs. standard program stratified by region

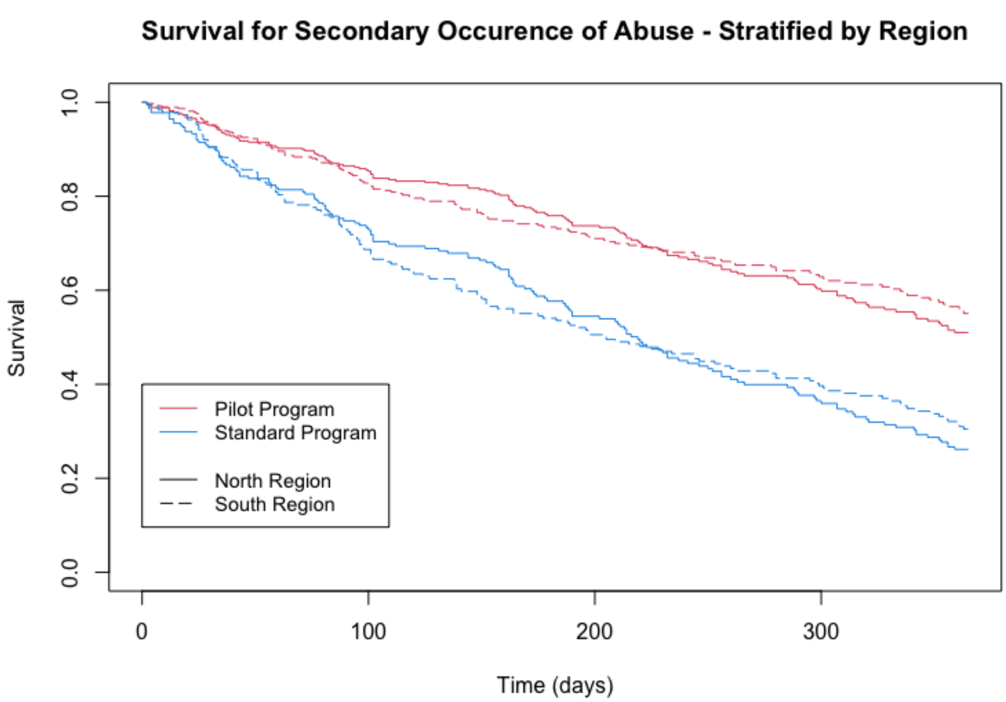
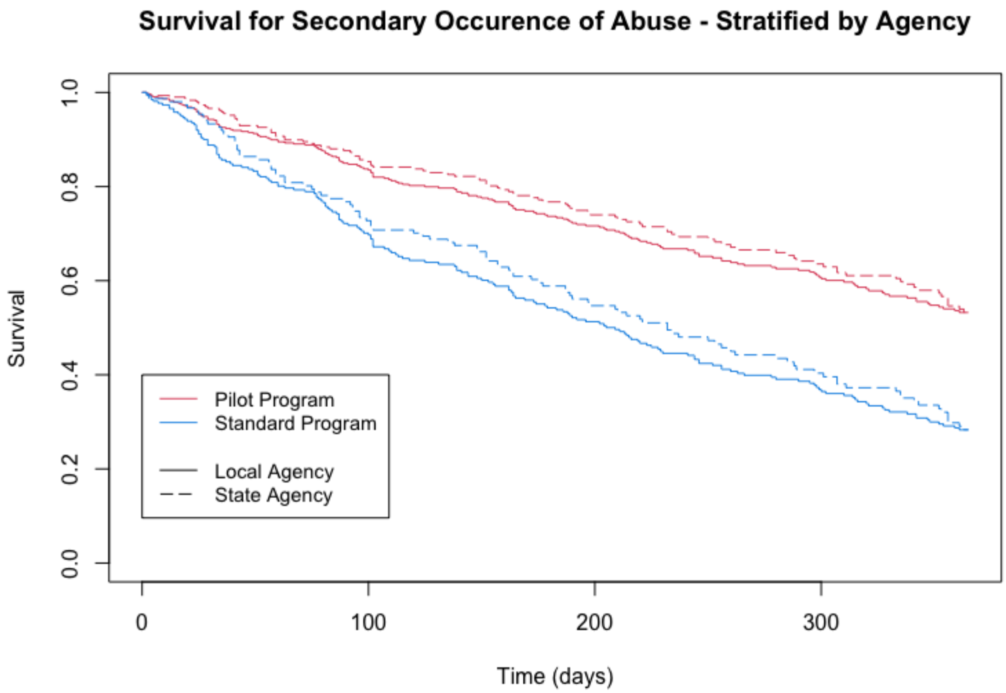


Figure 3. Survival of secondary occurrence of abuse for pilot program vs. standard program stratified by agency



The covariates of age, criminal history in the household, whether it was the first CPS contact, whether the child was removed from the home, and substance abuse in the household were all found to significantly influence secondary occurrences of child abuse. First CPS contact and higher age are protective against a secondary occurrence of abuse, while a criminal history in the home, removal from home to foster care, and substance abuse in the home are all associated with an increased hazard of secondary occurrence of abuse, after controlling for all included covariates.

Table 4. Hazard ratios, 95% confidence intervals, and p-values for Cox Regression model

	Cox Proportional Hazards Model		
	Hazard Ratio	**95% CI for HR**	**p-value**
Program - Standard	1.9935	1.5659 - 2.5378	< 0.001
Age (in years)	0.8063	0.7654 - 0.8495	< 0.001
Criminal History - Yes	2.3364	1.6687 - 3.2712	< 0.001
First CPS Contact - Yes	0.6412	0.4913 - 0.8368	0.0011
Removed from home (time-dependent) - Yes	1.5585	1.1568 - 2.0997	0.0035
Substance Abuse - Yes	1.6263	1.1286 - 2.3436	0.0091

The hazard of a secondary occurrence of child abuse decreases by about 19% for every one year increase in age (hazard ratio = 0.8063, 95% CI = (0.7654, 0.8495), $p < 0.001$), and it decreases by about 36% for those whom the initial incident was their first contact with CPS as compared to those who had had more than one contact with CPS (hazard ratio = 0.6412, 95% CI = (0.4913, 0.8368), $p = 0.0011$). Alternatively, the hazard of a secondary occurrence of child abuse increases by about 134% if someone in the household has a criminal history as compared to households without a criminal history (hazard ratio = 2.3364, 95% CI = (1.6687, 3.2712), $p < 0.001$), hazard increases by about 56% if the child has been removed from the home compared to if they were not removed (hazard ratio = 1.5585, 95% CI = (1.1568, 2.0997), $p = 0.0035$), and hazard increases by about 63% if someone in the household experiences substance abuse as compared to households with no substance abuse (hazard ratio = 1.6263, 95% CI = (1.1286, 2.3436), $p = 0.0091$). The covariates biological sex of the child, the child being of minority race, the family living below the poverty threshold, the region, and the agency were not found to have a significant influence on secondary occurrence of child abuse.

Based on these findings, there is a highly significant difference between the pilot program and the standard program in preventing secondary occurrences of physical child abuse. The Department of Child Protective Services should consider moving to phase out the standard program and to adopt the pilot program for all cases. Additionally, we have found three factors that increase the chances of secondary instances of physical abuse occurring in the home: criminal history in the home, removal from home to foster care, and substance abuse in the home. This information may also allow CPS to provide additional or individualized support to households that are experiencing these factors in order to further prevent future repeated occurrences of child abuse.

APPENDIX I - R code

```
# Load Survival Library and data set ----------------------------------------
library(survival)
setwd("/Users/Brittany/Documents/My Schoolwork/MS Biostatistics Degree/PHST 683-Survival Anal
load("AbuseStudy.Rdata")
attach(dat)


# Restructure data set -------------------------------------------------------

project.frame <- data.frame(matrix(ncol = 15, nrow = 0))
colnames(project.frame) <- c('id','start','stop','event','outofhome','program',
                             'age','sex', 'minority','poverty','subst.abuse',
                             'crim.hist','first','region','agency')

for (i in 1:nrow(dat)) {
    start <- 0
    stop <- numeric(length=0)
    outofhome <- 0
    event <- numeric(length=0)
    if (dat$removal[i]==1) {
      start <- c(start, dat$time.removal[i])
      stop <- c(stop, dat$time.removal[i])
      outofhome <- c(outofhome,1)
      event <- c(event, 0)
    }
    if (dat$return[i]==1) {
      start <- c(start, dat$time.return[i])
      stop <- c(stop, dat$time.return[i])
      outofhome <- c(outofhome,0)
      event <- c(event, 0)
    }
    stop <- c(stop, dat$time[i])
    event <- c(event, dat$event[i])
    program <- dat$program[i]
    age <- dat$age[i]
    sex <- dat$sex[i]
    minority <- dat$minority[i]
    poverty <- dat$poverty[i]
    subst.abuse <- dat$subst.abuse[i]
    crim.hist <- dat$crim.hist[i]
    first <- dat$first[i]
    region <- dat$region[i]
    agency <- dat$agency[i]
    temp.frame <- data.frame(id=dat$id[i],start,stop,event,outofhome,program,
                             age,sex,minority,poverty,subst.abuse,crim.hist,first,
                             region,agency)
    project.frame <- rbind(project.frame, temp.frame)
}
save(project.frame, file = "projectframe.RData")

# Detach and clear environment -----------------------------------------------
detach(dat)
rm(list = ls())
load("projectframe.RData")
attach(project.frame)
```

```r
# Checking for balance ---------------------------------------------------------

# Sex
table.sex <- table(program, sex)
prop.test(table.sex)
# Minority
table.min <- table(program, minority)
prop.test(table.min)
# Poverty
table.pov <- table(program, poverty)
prop.test(table.pov)
# Substance Abuse
table.sub <- table(program, subst.abuse)
prop.test(table.sub)
# Criminal History
table.crim <- table(program, crim.hist)
prop.test(table.crim)
# First CPS contact
table.first <- table(program, first)
prop.test(table.first)
# Region
table.reg<- table(program, region)
prop.test(table.reg)
# Agency
table.agen <- table(program, agency)
prop.test(table.agen)
# Out of home
table.home <- table(program, outofhome)
prop.test(table.home)

#Age
# Shapiro-Wilk normality tests
with(project.frame, shapiro.test(age[program=="Pilot"]))
with(project.frame, shapiro.test(age[program=="Standard"]))

# Variance test
var.test(age ~ program, data = project.frame)

# Mann-Whitney test
x <- project.frame[ which(program=="Pilot"), ]
x <- x$age
y <- project.frame[ which(program=="Standard"), ]
y <- y$age
wilcox.test(x,y)


# Create surv object -----------------------------------------------------------
project.surv <- Surv(start,stop,event,type = "counting")
project.surv
```

```r
# Create model without interactions ----------------------------------------

# Model with only treatment factor
mod1 <- coxph(project.surv~program)
summary(mod1) # time-dependent variable is significant


# Test for covariates
mod.out <- coxph(project.surv~program+outofhome)
mod.age <- coxph(project.surv~program+age)
mod.sex <- coxph(project.surv~program+sex)
mod.min <- coxph(project.surv~program+minority)
mod.pov <- coxph(project.surv~program+poverty)
mod.sub <- coxph(project.surv~program+subst.abuse)
mod.crim <- coxph(project.surv~program+crim.hist)
mod.first <- coxph(project.surv~program+first)

anova(mod1, mod.out) #Out of home is significant
anova(mod1, mod.age) #Age is significant
anova(mod1, mod.sex) #Sex is NOT significant
anova(mod1, mod.min) #Minority is NOT significant
anova(mod1, mod.pov) #Poverty is significant
anova(mod1, mod.sub) #Substance abuse is NOT significant
anova(mod1, mod.crim) #Criminal history is significant
anova(mod1, mod.first) #First is significant

mod2 <- coxph(project.surv~program+age)

mod.out <- coxph(project.surv~program+age+outofhome)
mod.sex <- coxph(project.surv~program+age+sex)
mod.min <- coxph(project.surv~program+age+minority)
mod.pov <- coxph(project.surv~program+age+poverty)
mod.sub <- coxph(project.surv~program+age+subst.abuse)
mod.crim <- coxph(project.surv~program+age+crim.hist)
mod.first <- coxph(project.surv~program+age+first)

anova(mod2, mod.out) #Out of home is significant
anova(mod2, mod.sex) #Sex is NOT significant
anova(mod2, mod.min) #Minority is NOT significant
anova(mod2, mod.pov) #Poverty is significant
anova(mod2, mod.sub) #Substance abuse is NOT significant
anova(mod2, mod.crim) #Criminal history is significant
anova(mod2, mod.first) #First is significant

mod3 <- coxph(project.surv~program+age+crim.hist)

mod.out <- coxph(project.surv~program+age+crim.hist+outofhome)
mod.sex <- coxph(project.surv~program+age+crim.hist+sex)
mod.min <- coxph(project.surv~program+age+crim.hist+minority)
mod.pov <- coxph(project.surv~program+age+crim.hist+poverty)
mod.sub <- coxph(project.surv~program+age+crim.hist+subst.abuse)
mod.first <- coxph(project.surv~program+age+crim.hist+first)
```

```r
anova(mod3, mod.out) #Out of home is significant
anova(mod3, mod.sex) #Sex is NOT significant
anova(mod3, mod.min) #Minority is NOT significant
anova(mod3, mod.pov) #Poverty is NOT significant
anova(mod3, mod.sub) #Substance abuse is significant
anova(mod3, mod.first) #First is significant

mod4 <- coxph(project.surv~program+age+crim.hist+first)

mod.out <- coxph(project.surv~program+age+crim.hist+first+outofhome)
mod.sex <- coxph(project.surv~program+age+crim.hist+first+sex)
mod.min <- coxph(project.surv~program+age+crim.hist+first+minority)
mod.pov <- coxph(project.surv~program+age+crim.hist+first+poverty)
mod.sub <- coxph(project.surv~program+age+crim.hist+first+subst.abuse)

anova(mod4, mod.out) #Out of home is significant
anova(mod4, mod.sex) #Sex is NOT significant
anova(mod4, mod.min) #Minority is NOT significant
anova(mod4, mod.pov) #Poverty is NOT significant
anova(mod4, mod.sub) #Substance abuse is significant

mod5 <- coxph(project.surv~program+age+crim.hist+first+outofhome)

mod.sex <- coxph(project.surv~program+age+crim.hist+first+outofhome+sex)
mod.min <- coxph(project.surv~program+age+crim.hist+first+outofhome+minority)
mod.pov <- coxph(project.surv~program+age+crim.hist+first+outofhome+poverty)
mod.sub <- coxph(project.surv~program+age+crim.hist+first+outofhome+subst.abuse)

anova(mod5, mod.sex) #Sex is NOT significant
anova(mod5, mod.min) #Minority is NOT significant
anova(mod5, mod.pov) #Poverty is NOT significant
anova(mod5, mod.sub) #Substance abuse is significant

mod6 <- coxph(project.surv~program+age+crim.hist+first+outofhome+subst.abuse)

mod.sex <- coxph(project.surv~program+age+crim.hist+first+outofhome+subst.abuse+sex)
mod.min <- coxph(project.surv~program+age+crim.hist+first+outofhome+subst.abuse+minority)
mod.pov <- coxph(project.surv~program+age+crim.hist+first+outofhome+subst.abuse+poverty)

anova(mod6, mod.sex) #Sex is NOT significant
anova(mod6, mod.min) #Minority is NOT significant
anova(mod6, mod.pov) #Poverty is NOT significant

# Final model without interactions
mod.noint <- mod6
summary(mod.noint)
drop1(mod.noint, test="Chisq")

# Testing for interactions -----------------------------------------------

mod.age <- update(mod.noint, .~.+program*age)
mod.crim <- update(mod.noint, .~.+program*crim.hist)
mod.first <- update(mod.noint, .~.+program*first)
mod.out <- update(mod.noint, .~.+program*outofhome)
mod.sub <- update(mod.noint, .~.+program*subst.abuse)

anova(mod.noint, mod.age)   #NOT significant
anova(mod.noint, mod.crim)  #NOT significant
anova(mod.noint, mod.first) #NOT significant
anova(mod.noint, mod.out)   #NOT significant
anova(mod.noint, mod.sub)   #NOT significant

mod.final <- mod.noint
```

```r
# Check Proportional Hazards Assumption ------------------------------------

cox.zph(mod.final)
drop1(mod.final, test="Chisq")    #Stratification is not required


# Stratification just for additional information ----------------------------

mod.regstrat <- update(mod.final, .~.+strata(region))
mod.agstrat <- update(mod.final, .~.+strata(agency))

summary(mod.regstrat)
summary(mod.agstrat)

cox.zph(mod.regstrat)
cox.zph(mod.agstrat)

drop1(mod.regstrat, test = "Chisq")
drop1(mod.agstrat, test = "Chisq")

# Plots ------------------------------------

mean(age)
new.frame <- data.frame(age=7.078519, program=c("Pilot","Standard"), crim.hist=0,
                        first=0, outofhome=0, subst.abuse=0)
fit1 <- survfit(mod.final, new.frame)
plot(fit1, col=c(2,4), conf.int = T, xlab="Time (days)", ylab="Survival",
     main="Survival for Secondary Occurence of Abuse")
legend(0, 0.4, legend=c("Pilot Program", "Standard Program", "", "Survival",
      "95% Confidence Interval"), lty=c(1,1,0,1,2), col=c(2,4,0,1,1), cex = 0.9)

fit2 <- survfit(mod.regstrat, new.frame)
plot(fit2, col=c(2,2,4,4), lty=c(1,5,1,5), conf.int = F, xlab="Time (days)", ylab="Survival",
     main="Survival for Secondary Occurence of Abuse - Stratified by Region")
legend(0, 0.4, legend=c("Pilot Program", "Standard Program", "", "North Region", "South Region"),
       lty=c(1,1,0,1,5), col=c(2,4,0,1,1), cex = 0.9)

fit3 <- survfit(mod.agstrat, new.frame)
plot(fit3, col=c(2,2,4,4), lty=c(1,5,1,5), conf.int = F, xlab="Time (days)", ylab="Survival",
     main="Survival for Secondary Occurence of Abuse - Stratified by Agency")
legend(0, 0.4, legend=c("Pilot Program", "Standard Program", "", "Local Agency", "State Agency"),
       lty=c(1,1,0,1,5), col=c(2,4,0,1,1), cex = 0.9)
```

APPENDIX II - Summary output for final Cox Regression Model

```
coxph(formula = project.surv ~ program + age + crim.hist + first +
    outofhome + subst.abuse)

  n= 675, number of events= 277

                   coef exp(coef) se(coef)      z Pr(>|z|)
programStandard  0.6899    1.9935   0.1232  5.601 2.13e-08 ***
age             -0.2153    0.8063   0.0266 -8.093 5.82e-16 ***
crim.hist        0.8486    2.3364   0.1717  4.942 7.73e-07 ***
first           -0.4444    0.6412   0.1358 -3.271  0.00107 **
outofhome        0.4437    1.5585   0.1521  2.918  0.00353 **
subst.abuse      0.4863    1.6263   0.1864  2.609  0.00909 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

                exp(coef) exp(-coef) lower .95 upper .95
programStandard    1.9935     0.5016    1.5659    2.5378
age                0.8063     1.2402    0.7654    0.8495
crim.hist          2.3364     0.4280    1.6687    3.2712
first              0.6412     1.5596    0.4913    0.8368
outofhome          1.5585     0.6417    1.1568    2.0997
subst.abuse        1.6263     0.6149    1.1286    2.3436

Concordance= 0.694  (se = 0.016 )
Likelihood ratio test= 132.7  on 6 df,   p=<2e-16
Wald test            = 129.5  on 6 df,   p=<2e-16
Score (logrank) test = 134.1  on 6 df,   p=<2e-16
```

APPENDIX III - Test of proportional hazards for final Cox Regression Model

```
> cox.zph(mod.final)
              chisq df    p
program      0.4326  1 0.51
age          0.0834  1 0.77
crim.hist    0.0190  1 0.89
first        0.1393  1 0.71
outofhome    1.6635  1 0.20
subst.abuse  1.5969  1 0.21
GLOBAL       4.3406  6 0.63
```