BST 222 - Homework 4

Xiaowei Zeng

2023-10-31

The addicts data set is from a study by Caplehorn et al. ("Methadone Dosage and Retention of Patients in Maintenance Treatment," Med. J. Aust., 1991) in Sydney, Australia and contains a cohort of 238 heroin addicts who entered maintenance programs between February 1986 and August 1987. The time in the clinic was determined by dates of entry into and exit or transfer from the methadone maintenance program to which they had been assigned. There are two further covariates, namely, prison record and methadone dose, believed to affect the survival times.

The data set and R input code are on the website. The variables are as follows:

- id: Subject ID
- clinic: Clinic (1 or 2)
- status: Survival status (1 = left the clinic, 0 = censored, i.e. we do not know when they left)
- time: Survival time in days
- prison: Prison record (0 = none, 1 = any)
- methadone: Methadone dose (mg/day)

Load the R packages and the dataset.

```
library(survival)
library(KMsurv)
library(MASS)

vars <- c("id","clinic","status","time","prison","methadone")
addicts <- read.table("https://dmrocke.ucdavis.edu/Class/BST222.2023.Fall/addicts.txt",
                      header=F, col.names=vars)

#change variables to factors to be used in Cox PH
addicts$clinic <- factor(addicts$clinic,labels=c("Clinic1","Clinic2"))
addicts$prison <- factor(addicts$prison,labels=c("No","Yes"))

head(addicts)
```

```
##   id  clinic status time prison methadone
## 1  1 Clinic1      1  428     No        50
## 2  2 Clinic1      1  275    Yes        55
## 3  3 Clinic1      1  262     No        55
## 4  4 Clinic1      1  183     No        30
## 5  5 Clinic1      1  259    Yes        65
## 6  6 Clinic1      1  714     No        55
```

```
# Create a survival object
dfsurv <- Surv(addicts$time, addicts$status)
```

1. Test proportionality of hazards.

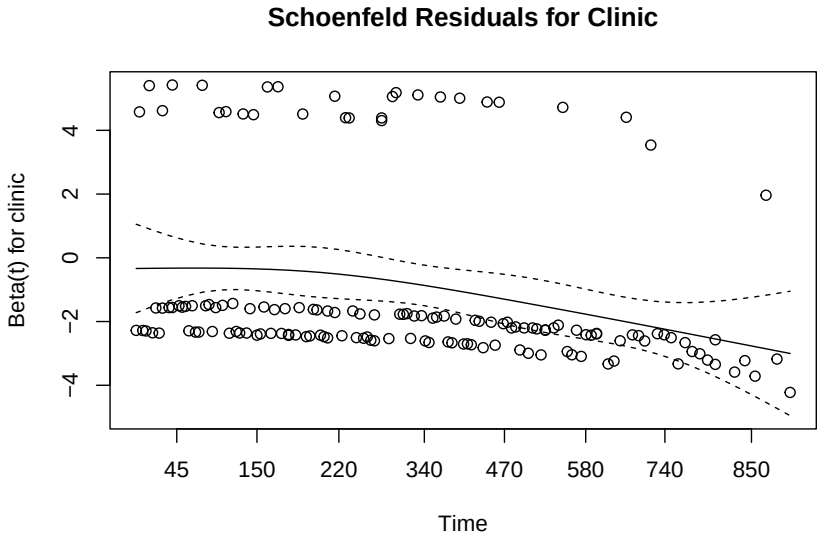
Fit the Cox model with clinic, methadone and prison and use `cox.zph` to test proportionality of hazards.

```
cox1 <- coxph(dfsurv ~ clinic + methadone + prison, data = addicts)
addicts.zph <- cox.zph(cox1)
print(addicts.zph)
```

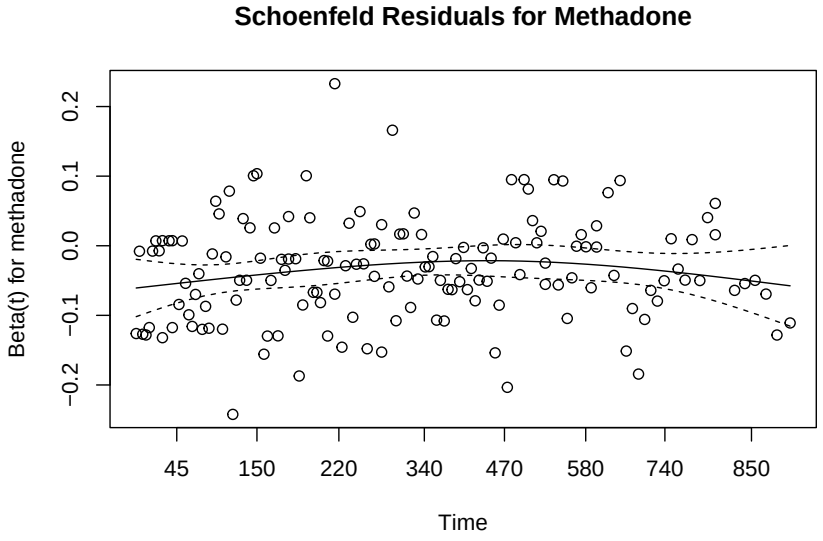
```
##            chisq df       p
## clinic    11.159  1 0.00084
## methadone  0.515  1 0.47310
## prison     0.839  1 0.35969
## GLOBAL    12.287  3 0.00646
```

There are three covariates in the model, so we will have three plots.

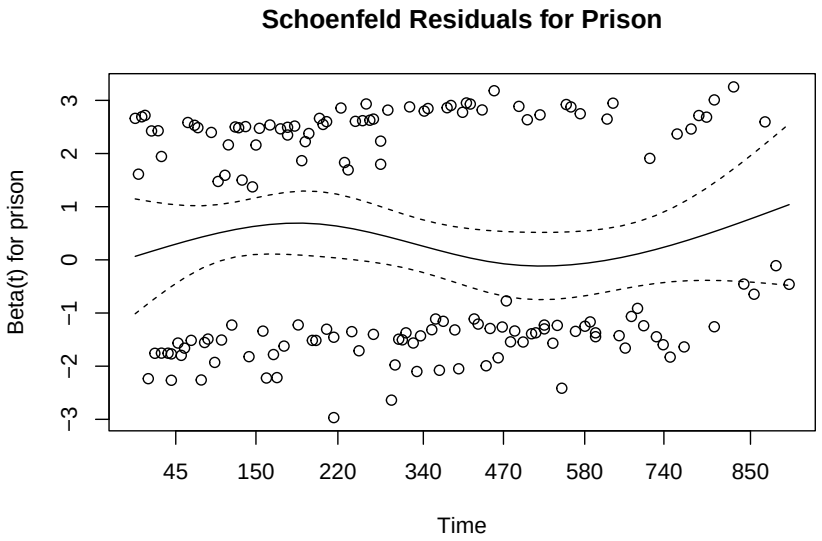
```
plot(addicts.zph[1], main = 'Schoenfeld Residuals for Clinic')
```




```
plot(addicts.zph[2], main = 'Schoenfeld Residuals for Methadone')
```



```
plot(addicts.zph[3], main = 'Schoenfeld Residuals for Prison')
```



We want to be sure of the PH assumption so we can use a higher threshold like 0.1 for the p-values. From the Schoenfeld Residuals for Clinic variable, we can find an evident pattern of residuals decreasing against time. The p-value of 0.00084 in the previous `zph` table also shows a problem in the proportionality assumption by clinic.

While in other plots, the residuals are roughly uncorrelated with time and the pattern of residuals vs time is nearly random; the p-values are both much larger than 0.1; so we can assume that these covariates meet the PH assumption.

Then have a quick look at the correlation with time.

```
cor.spear <- function(x) {cor.test(x, addicts.zph$x, method = "spearman")}
correlations <- apply(X = addicts.zph$y, MARGIN = 2, FUN = cor.spear)
correlations
```

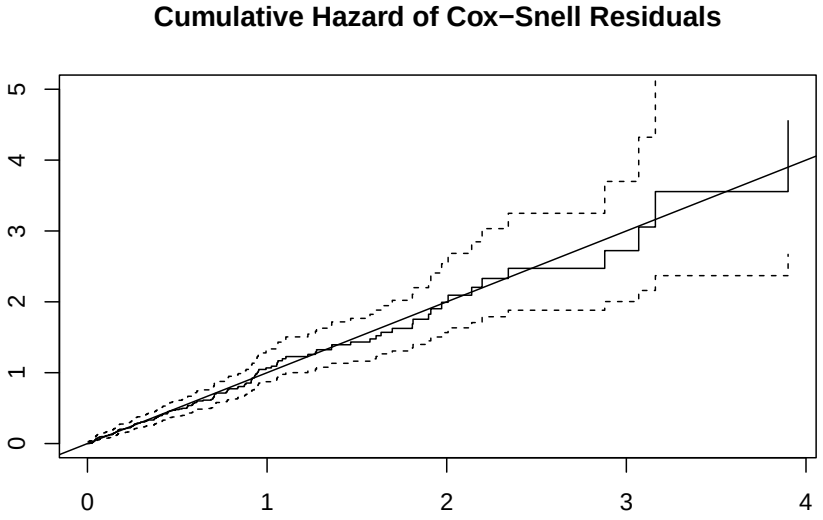
```
## $clinic
##
##  Spearman's rank correlation rho
##
## data:  x and addicts.zph$x
## S = 846709, p-value = 4.253e-11
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##        rho
## -0.5053265
##
##
## $methadone
##
##  Spearman's rank correlation rho
##
## data:  x and addicts.zph$x
## S = 513430, p-value = 0.2887
## alternative hypothesis: true rho is not equal to 0
```

```
## sample estimates:
##        rho
## 0.08719576
##
##
## $prison
##
##  Spearman's rank correlation rho
##
## data:  x and addicts.zph$x
## S = 492832, p-value = 0.1312
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##       rho
## 0.1238146
```

From the correlation test, the clinic variable induces statistically significant non-proportionality. This suggests that Cox model may not be an appropriate model for fitting clinic.

2. Plot the cumulative hazard of the Cox-Snell residuals.

```
addicts.mart <- residuals(cox1, type='martingale')
addicts.cs <- addicts$status - addicts.mart
surv.csr <- survfit(Surv(addicts.cs, addicts$status) ~ 1, type='fleming-harrington')
plot(surv.csr, fun='cumhaz', ylim=c(0, 5)); abline(0, 1)
title('Cumulative Hazard of Cox-Snell Residuals')
```

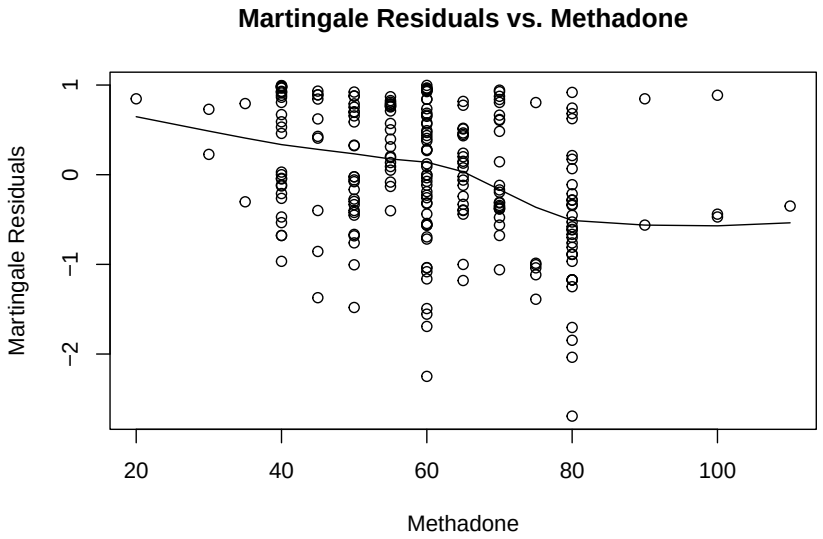


The line with slope 1 and intercept 0 fits the curve relatively well, so we don't see lack of fit based on these residuals.

3. Examine the functional form of methadone.

3.1. Plot the martingale residuals (omitting methadone) vs. methadone.

```
mres <- residuals(coxph(dfsurv~clinic+prison, data=addicts), type='martingale')
plot(addicts$methadone, mres, xlab='Methadone', ylab='Martingale Residuals')
lines(lowess(addicts$methadone, mres))
title('Martingale Residuals vs. Methadone')
```



The line is almost straight. It suggests that some modest transformation of Methadone may help improve the goodness of fit.

3.2. Make a categorized methadone variable.

Treat methadone as a categorical variable with categories $[20, 60)$, $[60, 80)$, and $[80, 110]$, and then label them respectively as 'low', 'median' or 'high'. Then fit a Cox model using categorized methadone as a factor.

```
meth2 <- cut(addicts$methadone, c(0, 59, 79, 110), labels=c('low', 'median', 'high'))
cox2 <- coxph(dfsurv~clinic+prison+meth2, data=addicts)
```

Use the LR test (`drop1`) to compare the statistical significance of the variables.

```
drop1(cox2, test='Chisq')
```

```
## Single term deletions
##
## Model:
## dfsurv ~ clinic + prison + meth2
##        Df    AIC    LRT  Pr(>Chi)
## <none>    1352.7
## clinic  1 1372.1 21.411 3.707e-06 ***
## prison  1 1353.6  2.881   0.08962 .
## meth2   2 1381.3 32.580 8.423e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The results show that at the significance level of 0.1, the coefficients of clinic, prison and categorized methadone are all significantly different from 0.

3.3. Compare the two full models by AIC.

Use `drop1` to obtain the AIC of the Cox model with numeric methadone variable.

```
drop1(cox1, test='Chisq')
```

```
## Single term deletions
##
## Model:
## dfsurv ~ clinic + methadone + prison
##           Df    AIC     LRT  Pr(>Chi)
## <none>       1352.5
## clinic     1 1376.9 26.3506 2.847e-07 ***
## methadone  1 1381.3 30.7820 2.887e-08 ***
## prison     1 1354.3  3.7727   0.05209 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

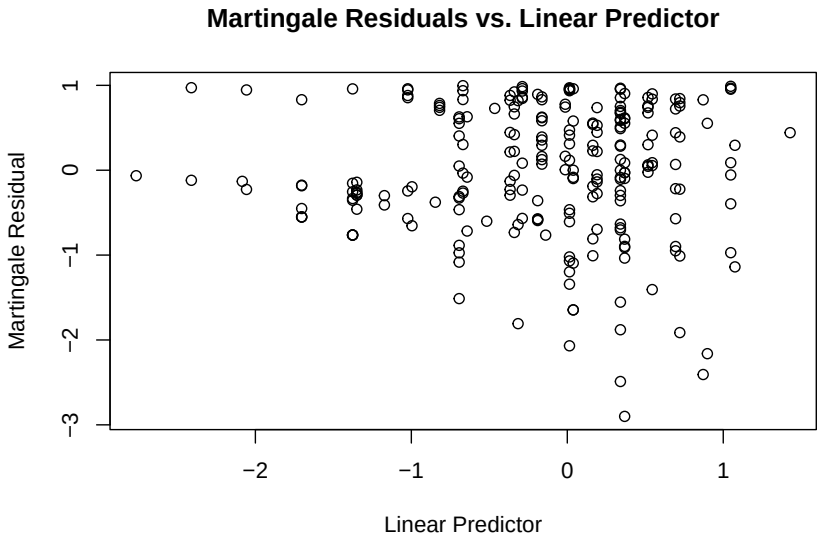
The model with numeric methadone variable has a lower AIC of 1352.5, while the AIC of the model with categorized methadone is 1352.7. So it actually make little difference to replace the quantitative variable by a 3-category factor, but this procedure can make it convenient and easy to explain the influence of methadone on survival.

4. Other model diagnostics.

Since the Cox model with numeric methadone variable has a lower AIC, in this section we continue diagnosing on this model.

Plot the martingale residuals vs. the linear predictor.

```
addicts.preds <- predict(cox1) # linear predictor
plot(addicts.preds, addicts.mart, xlab='Linear Predictor', ylab='Martingale Residual')
title('Martingale Residuals vs. Linear Predictor')
```



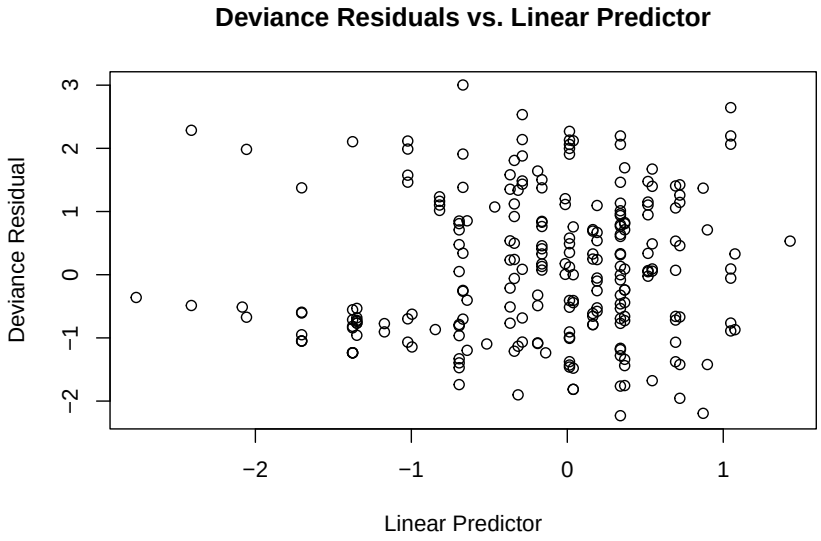
```
addicts.mart[addicts.mart < - 2.2]
```

```
##         8         9        26
## -2.489603 -2.900389 -2.407167
```

The smallest three martingale residuals in order are observations 8, 9, and 26.

Plot the deviance residuals vs. the linear predictor.

```
addicts.dev <- residuals(cox1, type='deviance')
plot(addicts.preds, addicts.dev, xlab='Linear Predictor', ylab='Deviance Residual')
title('Deviance Residuals vs. Linear Predictor')
```



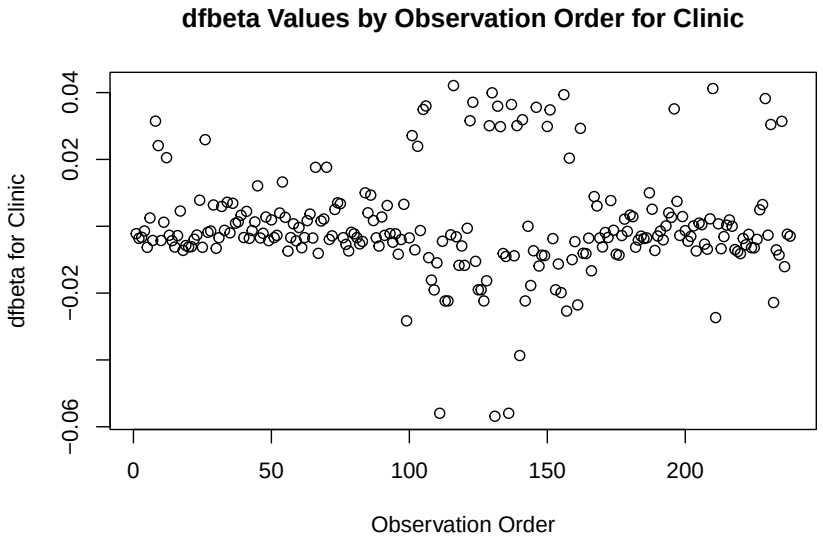

```
addicts.dev[addicts.dev > 3]
```

```
##      123
## 3.002669
```

The largest deviance residual is observation 123.

Plot the three dfbeta values by observation order.

```
addicts.dfb <- residuals(cox1, type='dfbeta')
plot(addicts.dfb[, 1], xlab='Observation Order', ylab='dfbeta for Clinic')
title('dfbeta Values by Observation Order for Clinic')
```

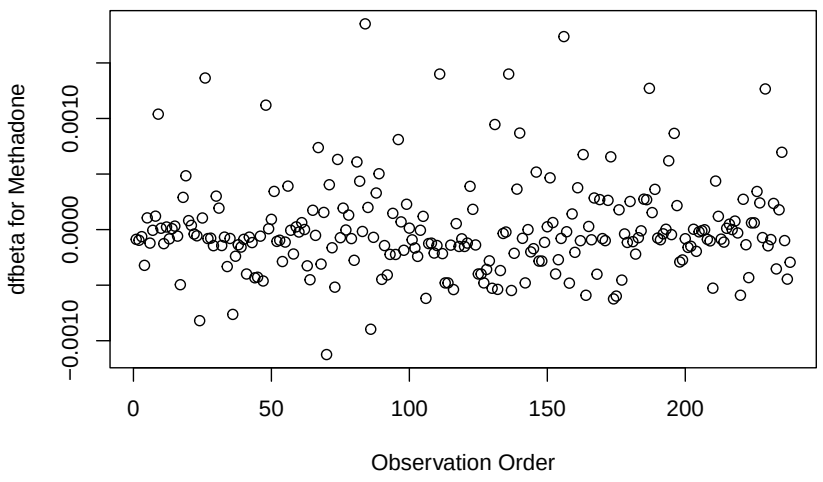



```
addicts.dfb[abs(addicts.dfb[, 1]) > 0.05, 1]
```

```
##         111         131         136
## -0.05596209 -0.05685669 -0.05596209
```

The three largest absolute dfbetas for clinic are observations 111, 131 and 136.

```
plot(addicts.dfb[, 2], xlab='Observation Order', ylab='dfbeta for Methadone')
title('dfbeta Values by Observation Order for Methadone')
```

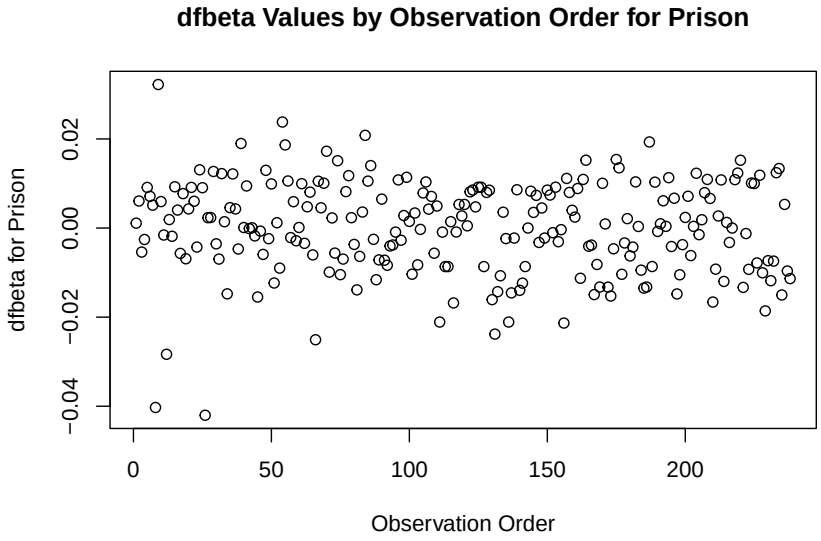



```
addicts.dfb[abs(addicts.dfb[, 2]) > 0.0013, 2]
```

```
##          26          84         111         136         156
## 0.001364216 0.001851482 0.001399712 0.001399712 0.001736957
```

The five largest absolute dfbetas for methadone are observations 26, 84, 111, 136 and 156.

```
plot(addicts.dfb[, 3], xlab='Observation Order', ylab='dfbeta for Prison')
title('dfbeta Values by Observation Order for Prison')
```




```
addicts.dfb[abs(addicts.dfb[, 3]) > 0.04, 3]
```

```
##           8          26
## -0.04028373 -0.04201940
```

The two largest absolute dfbetas for prison are observations 8 and 26.

Table: Observations to Examine by Residuals and Influence

Model Diagnostics	Observations
Martingale Residuals	8, 9, 26
Deviance Residuals	123
Clinic Influence	111, 131, 136
Methadone Influence	26, 84, 111, 136, 156
Prison Influence	8, 26

So the most important observations to examine seem to be 8, 26, 111, and 136.

```
with(addicts, summary(time[status==1]))
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     7.0   157.8   283.0   345.5   521.0   899.0
```

```
with(addicts, summary(methadone))
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    20.0    50.0    60.0    60.4    70.0   110.0
```

```
with(addicts, summary(prison))
```

```
##  No Yes
## 127 111
```

```
addicts[c(8, 26, 111, 136), -1]
```

```
##      clinic status time prison methadone
## 8   Clinic1      0  796    Yes        60
## 26  Clinic1      0  566    Yes        45
## 111 Clinic2      0  769    Yes        40
## 136 Clinic2      0  769    Yes        40
```

- 8: Clinic1, keeping staying in clinic, with prison record, median methadone doze.
- 26: Clinic1, keeping staying in clinic, with prison record, low methadone doze.
- 111: Clinic2, keeping staying in clinic, with prison record, low methadone doze.
- 136: Clinic2, keeping staying in clinic, with prison record, low methadone doze.

It turns out that nearly all the unusual observations have features of not leaving the clinic, having at least one prison record and low methadone doze. But this can actually happen in real case.

Since the clinic hazards are clearly not proportional, this is not the final answer, but it will do for the moment. Then we may need to use strata, between which the base hazards are permitted to be different.