ACTG175_Nov2

Alvin Sheng

October 31, 2018

Set Up

Requisite Libraries

```
library(survival)
library(glmnet)
```

```
## Loading required package: Matrix
```

```
## Loading required package: foreach
```

```
## Loaded glmnet 2.0-13
```

```
library(polspline)
library(knitr)
```

Constructing the Design Matrix

I have omitted variables cd420 and cd820, on the basis of the SUGI paper's recommendations

The groups of people off-treatment by ~96 weeks and people not off-treatment by ~96 weeks are large enough for subgroup analysis.

```
actg175_mat <- model.matrix( ~ trt + age + wtkg + hemo + drugs +
                               karnof + oprior + preanti + race +
                               gender + symptom + offtrt + cd40 +
                               cd80, actg175)[,-1]

# Stratifying the data on offtrt for Subgroup Analysis
mat_not_off <- model.matrix( ~ trt + age + wtkg + hemo + drugs +
                               karnof + oprior + preanti + race +
                               gender + symptom + cd40 + cd80,
                               actg175[actg175$offtrt == 0,])[,-1]

mat_off <- model.matrix( ~ trt + age + wtkg + hemo + drugs +
                               karnof + oprior + preanti + race +
                               gender + symptom + cd40 + cd80,
                               actg175[actg175$offtrt == 1,])[,-1]
```

```
covariate_names <- colnames(actg175_mat)

kable(covariate_names, row.names = 1:length(covariate_names), col.names = "Covariates",
      caption = "These are the covariates of the data,
      in order, that will be used in glmnet and HARE")
```

Table 1: These are the covariates of the data, in order, that will be used in glmnet and HARE

	Covariates
1	trtZDV+ddi
2	trtZDV+ZAL
3	trtddi
4	age
5	wtkg
6	hemo1
7	drugs1
8	karnof
9	oprior1
10	preanti
11	race1
12	gender1
13	symptom1
14	offtrt1
15	cd40
16	cd80

```
covariate_names <- colnames(mat_not_off)

kable(covariate_names, row.names = 1:length(covariate_names), col.names = "Covariates",
      caption = "These are the covariates of the data stratified by offtrt, in order, that will be used
```

Table 2: These are the covariates of the data stratified by offtrt, in order, that will be used in glmnet and HARE

	Covariates
1	trtZDV+ddi
2	trtZDV+ZAL
3	trtddi
4	age
5	wtkg
6	hemo1
7	drugs1
8	karnof
9	oprior1
10	preanti
11	race1
12	gender1
13	symptom1
14	cd40
15	cd80

Survival Analysis

Proportional Hazards Modelling

Cox Proportional Hazards Model, unstratified dataset

```
(phm_full <- coxph(Surv(time, cid) ~ trt + age + wtkg + hemo + drugs +
                                karnof + oprior + preanti + race +
                                gender + symptom + offtrt + cd40 +
                                cd80, data = actg175))
```

```
## Call:
## coxph(formula = Surv(time, cid) ~ trt + age + wtkg + hemo + drugs +
##     karnof + oprior + preanti + race + gender + symptom + offtrt +
##     cd40 + cd80, data = actg175)
##
##                 coef exp(coef)  se(coef)     z       p
## trtZDV+ddi -7.57e-01  4.69e-01  1.25e-01 -6.06 1.3e-09
## trtZDV+ZAL -6.60e-01  5.17e-01  1.22e-01 -5.42 6.0e-08
## trtddi     -5.43e-01  5.81e-01  1.17e-01 -4.65 3.3e-06
## age         8.25e-03  1.01e+00  5.32e-03  1.55 0.12098
## wtkg        1.67e-03  1.00e+00  3.51e-03  0.47 0.63529
## hemo1      -4.92e-03  9.95e-01  1.71e-01 -0.03 0.97697
## drugs1     -3.98e-01  6.72e-01  1.49e-01 -2.67 0.00768
## karnof     -2.05e-02  9.80e-01  6.97e-03 -2.94 0.00330
## oprior1    -4.46e-02  9.56e-01  2.51e-01 -0.18 0.85865
## preanti     4.23e-04  1.00e+00  8.92e-05  4.74 2.1e-06
## race1      -5.64e-02  9.45e-01  1.09e-01 -0.52 0.60444
## gender1     2.44e-02  1.02e+00  1.37e-01  0.18 0.85880
## symptom1    3.82e-01  1.47e+00  1.04e-01  3.69 0.00023
## offtrt1     6.36e-01  1.89e+00  9.22e-02  6.89 5.4e-12
## cd40       -3.75e-03  9.96e-01  4.42e-04 -8.50 < 2e-16
## cd80        4.43e-04  1.00e+00  8.50e-05  5.22 1.8e-07
##
## Likelihood ratio test=282  on 16 df, p=0
## n= 2139, number of events= 521
```

```
print(cox.zph(phm_full))
```

```
##                rho   chisq        p
## trtZDV+ddi  0.1014  5.4925 0.019098
## trtZDV+ZAL  0.1220  7.7955 0.005238
## trtddi      0.0612  2.0058 0.156702
## age        -0.0271  0.4629 0.496276
## wtkg        0.0549  1.6044 0.205278
## hemo1       0.0540  1.5060 0.219747
## drugs1     -0.0186  0.1943 0.659357
## karnof     -0.0290  0.4460 0.504254
## oprior1    -0.0197  0.2133 0.644182
## preanti    -0.0267  0.3840 0.535477
## race1       0.0123  0.0808 0.776272
## gender1    -0.0140  0.1056 0.745252
## symptom1   -0.0104  0.0584 0.809112
## offtrt1    -0.1563 12.8731 0.000333
```

```
## cd40        0.1304 10.5127 0.001186
## cd80       -0.0969  4.7852 0.028705
## GLOBAL          NA 43.8346 0.000209
```

Cox Proportional Hazards Models, Stratified on Offtrt

Not off-treatment group

```
phm_not_off <- coxph(Surv(time, cid) ~ trt + age + wtkg + hemo + drugs +
                               karnof + oprior + preanti + race +
                               gender + symptom + cd40 + cd80,
                     data = actg175[actg175$offtrt == 0,])
phm_not_off
```

```
## Call:
## coxph(formula = Surv(time, cid) ~ trt + age + wtkg + hemo + drugs +
##     karnof + oprior + preanti + race + gender + symptom + cd40 +
##     cd80, data = actg175[actg175$offtrt == 0, ])
##
##                 coef exp(coef)  se(coef)     z       p
## trtZDV+ddi -0.918027  0.399306  0.165747 -5.54 3.0e-08
## trtZDV+ZAL -0.836859  0.433069  0.168584 -4.96 6.9e-07
## trtddi     -0.689483  0.501836  0.152634 -4.52 6.3e-06
## age        -0.004292  0.995717  0.007300 -0.59 0.55660
## wtkg        0.000939  1.000939  0.005035  0.19 0.85211
## hemo1       0.128185  1.136763  0.214582  0.60 0.55026
## drugs1     -0.243320  0.784021  0.222682 -1.09 0.27453
## karnof     -0.022326  0.977921  0.009906 -2.25 0.02421
## oprior1     0.308139  1.360890  0.316063  0.97 0.32960
## preanti     0.000371  1.000371  0.000124  2.98 0.00286
## race1       0.015824  1.015950  0.145333  0.11 0.91330
## gender1     0.003519  1.003525  0.191097  0.02 0.98531
## symptom1    0.336058  1.399420  0.148391  2.26 0.02353
## cd40       -0.004094  0.995914  0.000603 -6.79 1.1e-11
## cd80        0.000413  1.000413  0.000114  3.64 0.00028
##
## Likelihood ratio test=123  on 15 df, p=0
## n= 1363, number of events= 291
```

```
cox.zph(phm_not_off)
```

```
##                 rho   chisq       p
## trtZDV+ddi  0.08953  2.3616 0.12435
## trtZDV+ZAL  0.17330  8.7392 0.00311
## trtddi      0.05469  0.8832 0.34732
## age         0.02505  0.2177 0.64078
## wtkg        0.12262  4.7659 0.02903
## hemo1       0.03879  0.4435 0.50542
## drugs1     -0.06586  1.3520 0.24494
## karnof      0.01717  0.0896 0.76471
## oprior1     0.01372  0.0591 0.80797
## preanti     0.03587  0.3864 0.53421
## race1      -0.00913  0.0244 0.87583
## gender1    -0.04604  0.6344 0.42573
## symptom1    0.10698  3.3963 0.06534
```

```
## cd40        0.15831  9.6409 0.00190
## cd80       -0.18448  7.6441 0.00570
## GLOBAL           NA 30.8683 0.00915
```

Off-treatment group

```
phm_off <- coxph(Surv(time, cid) ~ trt + age + wtkg + hemo + drugs +
                              karnof + oprior + preanti + race +
                              gender + symptom + cd40 + cd80,
                    data = actg175[actg175$offtrt == 1,])
phm_off
```

```
## Call:
## coxph(formula = Surv(time, cid) ~ trt + age + wtkg + hemo + drugs +
##     karnof + oprior + preanti + race + gender + symptom + cd40 +
##     cd80, data = actg175[actg175$offtrt == 1, ])
##
##                 coef exp(coef)  se(coef)     z       p
## trtZDV+ddi -0.571702  0.564563  0.192543 -2.97 0.00299
## trtZDV+ZAL -0.482181  0.617435  0.178704 -2.70 0.00697
## trtddi     -0.367931  0.692165  0.183836 -2.00 0.04535
## age         0.025801  1.026137  0.007986  3.23 0.00123
## wtkg        0.004421  1.004430  0.004967  0.89 0.37351
## hemo1      -0.291446  0.747182  0.292683 -1.00 0.31936
## drugs1     -0.525043  0.591530  0.200929 -2.61 0.00897
## karnof     -0.018633  0.981539  0.009904 -1.88 0.05991
## oprior1    -0.462487  0.629716  0.420442 -1.10 0.27133
## preanti     0.000454  1.000454  0.000129  3.52 0.00043
## race1      -0.182109  0.833511  0.168925 -1.08 0.28101
## gender1     0.058549  1.060297  0.198503  0.29 0.76803
## symptom1    0.476667  1.610697  0.148362  3.21 0.00131
## cd40       -0.003219  0.996786  0.000663 -4.85 1.2e-06
## cd80        0.000398  1.000398  0.000130  3.06 0.00224
##
## Likelihood ratio test=111  on 15 df, p=1.11e-16
## n= 776, number of events= 230
```

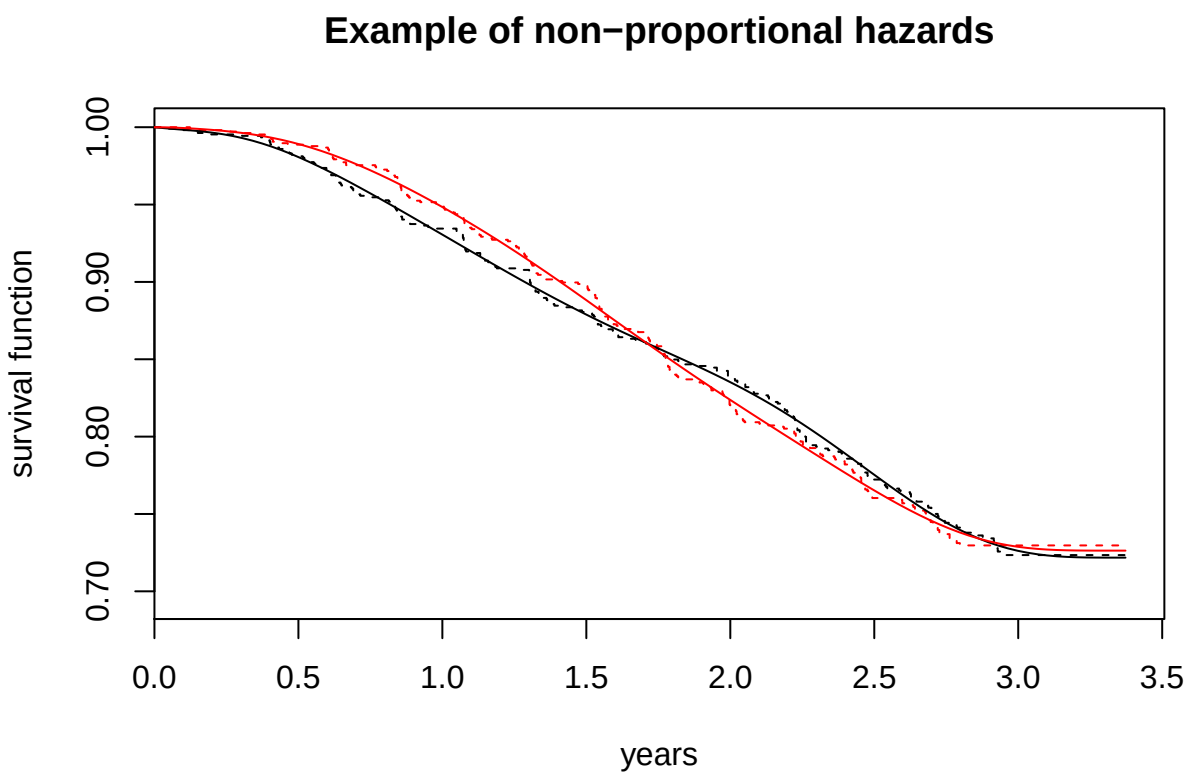
```
cox.zph(phm_off)
```

```
##                rho    chisq      p
## trtZDV+ddi  0.15779  6.03981 0.0140
## trtZDV+ZAL  0.14687  5.07131 0.0243
## trtddi      0.10890  2.93366 0.0868
## age        -0.02464  0.17777 0.6733
## wtkg       -0.03563  0.28871 0.5910
## hemo1       0.08734  1.74632 0.1863
## drugs1     -0.00575  0.00813 0.9282
## karnof     -0.08520  1.69962 0.1923
## oprior1    -0.08155  1.64443 0.1997
## preanti    -0.05853  0.83852 0.3598
## race1       0.01473  0.05340 0.8173
## gender1     0.00498  0.00619 0.9373
## symptom1   -0.11309  3.26190 0.0709
## cd40        0.11954  3.49329 0.0616
## cd80        0.00315  0.00292 0.9569
## GLOBAL           NA 22.19092 0.1029
```

Survival Curves using Bernstein Polynomials, Overlaid on the Kaplan-Meier Curves

Survival Curves for cd820

```
source("BPSurv.R")
pwr=0.4 #see Osman&Ghosh (2012) for details
yrs=seq(0,max(actg175$time),l=100)

above_med <- ifelse(actg175$cd820 > median(actg175$cd820), 1, 0)

pred=as.factor(above_med)
pred.levels=levels(pred); n.levels=length(pred.levels)
km.fit=survfit(Surv(time, cid) ~ pred, data=actg175)
lb=min(km.fit$lower)
plot(km.fit,col=1:n.levels,lty=2,ylab="survival function",xlab="years",ylim=c(lb,1))
for(j in 1:n.levels){
surv.time=actg175$time[pred==pred.levels[j]]; status=actg175$cid[pred==pred.levels[j]]
n=length(surv.time); m.est=ceiling(n^pwr)
bp.fit=BPsurv(y=surv.time,d=status,m=m.est)
S.bp=bp.fit$SFun(yrs)
lines(yrs, S.bp, col=j)
#log.haz=bp.fit$hFun(yrs)
#lines(yrs,log.haz)
              }
# legend("bottomleft",legend=c("Below Median cd820", "Above Median cd820"),col=1:n.levels,lty=1)
title("Example of non-proportional hazards",cex=0.75)
```



Lasso and Elastic-Net Regularized Generalized Linear Models

I'm using an alpha value of .95.

Finding the optimal lambda by running cross-validation 100 times and finding the median lambda.min

Unstratified Dataset

```
set.seed(123)

lambdas <- rep(0, 100)

for (i in 1:100) {
cv <- cv.glmnet(actg175_mat, Surv(actg175$time,
                                  actg175$cid),
                family = "cox", alpha = .95)
lambdas[i] <- cv$lambda.1se
}

median(lambdas)
```

```
## [1] 0.02131306
```

Using the median lambda to find the chosen coefficients

```
cv_phmnet <- cv.glmnet(actg175_mat,
                       Surv(actg175$time, actg175$cid),
                           family = "cox", alpha = .95)

med_lambda <- median(lambdas)

coef(cv_phmnet, s = med_lambda)
```

```
## 16 x 1 sparse Matrix of class "dgCMatrix"
##                          1
## trtZDV+ddi -0.2011000990
## trtZDV+ZAL -0.1339984190
## trtddi     -0.0295024973
## age         0.0005546806
## wtkg        .
## hemo1       .
## drugs1     -0.0932802322
## karnof     -0.0125510029
## oprior1     .
## preanti     0.0002700302
## race1       .
## gender1     .
## symptom1    0.3002258463
## offtrt1     0.4929877715
## cd40       -0.0027633650
## cd80        0.0002426843
```

Dataset Stratified on Offtrt

Not off-treatment group

```
set.seed(1019)

lambdas_not_off <- rep(0, 100)

for (i in 1:100) {
cv <- cv.glmnet(mat_not_off, Surv(actg175$time[actg175$offtrt == 0],
                                  actg175$cid[actg175$offtrt == 0]),
                family = "cox", alpha = .95)
lambdas_not_off[i] <- cv$lambda.1se
}

median(lambdas_not_off)
```

```
## [1] 0.02325531
```

Not off-treatment group

```
cv_phmnet_not_off <- cv.glmnet(mat_not_off,
                               Surv(actg175$time[actg175$offtrt == 0],
                                    actg175$cid[actg175$offtrt == 0]),
                               family = "cox", alpha = .95)

med_lambda_not_off <- median(lambdas_not_off)

coef(cv_phmnet_not_off, s = med_lambda_not_off)
```

```
## 15 x 1 sparse Matrix of class "dgCMatrix"
##                          1
## trtZDV+ddi -0.1838102599
## trtZDV+ZAL -0.1405427973
## trtddi     -0.0086159720
## age         .
## wtkg        .
## hemo1       .
## drugs1      .
## karnof     -0.0063399388
## oprior1     0.0426783109
## preanti     0.0001754859
## race1       .
## gender1     .
## symptom1    0.1470136559
## cd40       -0.0026907758
## cd80        0.0001632241
```

Exact same results as in Oct 19 report. Apparently, removing cd820 from mat_off had no effect.

Off-treatment group

```
set.seed(1102)

lambdas_off <- rep(0, 100)

for (i in 1:100) {
```

```
cv <- cv.glmnet(mat_off, Surv(actg175$time[actg175$offtrt == 1],
                              actg175$cid[actg175$offtrt == 1]),
                family = "cox", alpha = .95)
lambdas_off[i] <- cv$lambda.1se
}

median(lambdas_off)
```

```
## [1] 0.05485941
```

Off-treatment group

```
cv_phmnet_off <- cv.glmnet(mat_off, Surv(actg175$time[actg175$offtrt == 1],
                                         actg175$cid[actg175$offtrt == 1]),
                           family = "cox", alpha = .95)

med_lambda_off <- median(lambdas_off)


coef(cv_phmnet_off, s = med_lambda_off)
```

```
## 15 x 1 sparse Matrix of class "dgCMatrix"
##                          1
## trtZDV+ddi   .
## trtZDV+ZAL   .
## trtddi       .
## age          1.303510e-02
## wtkg         .
## hemo1        .
## drugs1       .
## karnof       .
## oprior1      .
## preanti      1.808237e-04
## race1        .
## gender1      .
## symptom1     2.751554e-01
## cd40        -1.166873e-03
## cd80         2.149170e-06
```

As compared with the Oct 19 report, this model is missing drugs and karnof.

HARE restricted to proportional hazards models, unstratified dataset

Restricting model selection to **proportional hazards models** through prophaz = TRUE

```
phm_hare <- hare(actg175$time, actg175$cid, actg175_mat, prophaz = TRUE)

summary(phm_hare)
```

```
## dim A/D   loglik       AIC        penalty
##                                   min    max
##   1 Add  -1714.79   3437.25   94.34     Inf
##   2 Add  -1670.85   3357.03      NA      NA
##   3 Add  -1620.45   3263.90   53.55   94.34
##   4 Add  -1593.67   3218.02   27.38   53.55
```

```
##   5 Add  -1579.98   3198.31   21.19   27.38
##   6 Add  -1569.39   3184.78   19.30   21.19
##   7 Add  -1560.07   3173.81      NA      NA
##   8 Add  -1551.71   3164.77      NA      NA
##   9 Del  -1540.44   3149.89   16.91   19.30
##  10 Del  -1531.99   3140.65   12.36   16.91
##  11 Add  -1525.80   3135.96   11.55   12.36
##  12 Add  -1521.97   3135.96      NA      NA
##  13 Add  -1518.26   3136.20      NA      NA
##  14 Add  -1510.68   3128.71      NA      NA
##  15 Add  -1506.80   3128.63      NA      NA
##  16 Del  -1496.92   3116.53    5.58   11.55
##  17 Del  -1494.13   3118.61    5.52    5.58
##  18 Del  -1491.36   3120.76    5.09    5.52
##  19 Del  -1488.82   3123.34    4.09    5.09
##  20 Del  -1486.92   3127.20      NA      NA
##  21 Del  -1484.73   3130.50    3.66    4.09
##  22 Del  -1482.90   3134.50    2.75    3.66
##  23 Del  -1481.53   3139.42    2.67    2.75
##  24 Del  -1480.19   3144.41    2.26    2.67
##  25 Del  -1479.06   3149.82    2.15    2.26
##  26 Del  -1477.98   3155.34    1.84    2.15
##  27 Add  -1477.07   3161.17    0.00    1.84
##
## the present optimal number of dimensions is 16.
## penalty(AIC) was the default: BIC=log(samplesize): log(2139)=7.67
##
##   dim1           dim2           beta        SE         Wald
## Constant                          -5.7        2.2   -2.64
## Co-15  linear                      0.04      0.016    2.54
## Time       1.3                    -0.78       0.21   -3.76
## Co-14  linear                     -0.61       0.36   -1.72
## Co-16  linear                    0.00047    8.4e-05    5.59
## Co-10  linear                    0.00047    8.6e-05    5.43
## Co-13  linear                       0.43        0.1    4.16
## Time      0.61                     -3.1       0.83   -3.78
## Co-1  linear                      -0.76       0.12   -6.09
## Co-2  linear                       -0.7       0.12   -5.78
## Co-3  linear                      -0.55       0.12   -4.68
## Co-4  linear   Co-14  linear        0.036     0.0094    3.78
## Co-4        43                    0.055      0.016    3.43
## Co-15   1.4e+02                   -0.056      0.017   -3.41
## Co-15   2.2e+02                    0.014     0.0029    4.80
## Co-4  linear                      -0.021     0.0083   -2.57
```

As compared with the corresponding HARE model with just cd420 deleted instead of both cd420 and cd820, the variables chosen are the same in this model (without cd820, of course), supporting the decision to omit cd820.

HARE restricted to proportional hazards models, stratified on Offtrt

Not Off-treatment group

```
phm_hare_not_off <- hare(actg175$time[actg175$offtrt == 0],
                                 actg175$cid[actg175$offtrt == 0], mat_not_off, prophaz = TRUE)

summary(phm_hare_not_off)
```

```
## dim A/D   loglik       AIC        penalty
##                                    min    max
##   1 Add  -1028.93   2065.08   88.22     Inf
##   2 Add   -984.82   1984.08   47.21   88.22
##   3 Add   -961.21   1944.08   14.94   47.21
##   4 Del   -954.42   1937.71      NA      NA
##   5 Del   -948.65   1933.39      NA      NA
##   6 Del   -938.80   1920.90   14.29   14.94
##   7 Del   -931.65   1913.83    9.04   14.29
##   8 Del   -927.14   1912.01    8.21    9.04
##   9 Del   -923.03   1911.02    6.16    8.21
##  10 Del   -919.95   1912.08    5.95    6.16
##  11 Del   -916.98   1913.34    4.92    5.95
##  12 Del   -914.67   1915.95      NA      NA
##  13 Del   -912.06   1917.94    3.85    4.92
##  14 Del   -910.15   1921.34      NA      NA
##  15 Del   -908.21   1924.68    3.74    3.85
##  16 Del   -906.34   1928.16    3.48    3.74
##  17 Del   -904.60   1931.89    2.27    3.48
##  18 Del   -903.96   1937.84      NA      NA
##  19 Del   -902.53   1942.19      NA      NA
##  20 Del   -901.20   1946.74    1.89    2.27
##  21 Del   -900.25   1952.07    1.64    1.89
##  22 Del   -899.43   1957.65    1.57    1.64
##  23 Del   -898.64   1963.29    1.07    1.57
##  24 Del   -898.11   1969.44    1.05    1.07
##  25 Add   -897.58   1975.60    0.00    1.05
##
## the present optimal number of dimensions is 9.
## penalty(AIC) was the default: BIC=log(samplesize): log(1363)=7.22
##
##   dim1           dim2           beta        SE         Wald
## Constant                          0.74       0.48    1.52
## Time       1.3                     -2       0.24   -8.20
## Co-14  linear                   -0.013     0.0022   -5.77
## Co-15  linear                   0.0011    0.00024    4.40
## Co-14   2.3e+02                 0.0093     0.0025    3.70
## Co-1  linear                    -0.99       0.17   -5.85
## Co-2  linear                    -0.88       0.17   -5.23
## Co-3  linear                    -0.69       0.15   -4.55
## Co-15   1.2e+03                -0.0011     0.0004   -2.73
```

Off-treatment group

```
phm_hare_off <- hare(actg175$time[actg175$offtrt == 1],
                                 actg175$cid[actg175$offtrt == 1], mat_off, prophaz = TRUE)

summary(phm_hare_off)
```

```
## dim A/D   loglik       AIC        penalty
```

```
##                                  min    max
##   1 Add   -657.77   1322.19   58.00     Inf
##   2 Add   -628.77   1270.84   28.52   58.00
##   3 Add   -614.51   1248.98   18.46   28.52
##   4 Add   -605.28   1237.17   17.00   18.46
##   5 Del   -596.78   1226.82   10.37   17.00
##   6 Del   -591.85   1223.62      NA      NA
##   7 Add   -586.97   1220.53      NA      NA
##   8 Add   -581.22   1215.66    9.69   10.37
##   9 Add   -576.98   1213.84      NA      NA
##  10 Del   -571.52   1209.58    5.18    9.69
##  11 Add   -569.13   1211.45      NA      NA
##  12 Add   -566.34   1212.53    4.36    5.18
##  13 Add   -564.16   1214.82    3.80    4.36
##  14 Del   -562.56   1218.28      NA      NA
##  15 Del   -560.42   1220.65      NA      NA
##  16 Del   -558.45   1223.37    3.46    3.80
##  17 Del   -557.14   1227.40      NA      NA
##  18 Del   -554.99   1229.76    2.62    3.46
##  19 Del   -553.80   1234.03      NA      NA
##  20 Del   -552.37   1237.82    2.60    2.62
##  21 Del   -551.07   1241.88    1.66    2.60
##  22 Add   -550.24   1246.88    0.00    1.66
##
## the present optimal number of dimensions is 10.
## penalty(AIC) was the default: BIC=log(samplesize): log(776)=6.65
##
##   dim1           dim2           beta        SE         Wald
## Constant                          -4.5       2.6   -1.72
## Time      0.61                      -5      0.82   -6.10
## Co-14  linear                     0.045     0.021    2.18
## Co-4  linear                    -0.0014     0.012   -0.11
## Co-13  linear                       0.6      0.15    4.12
## Co-15  linear                    -0.012    0.0034   -3.36
## Co-10  linear                   0.00043   0.00012    3.51
## Co-14   1.4e+02                   -0.05     0.021   -2.37
## Co-15   3.1e+02                   0.012    0.0034    3.49
## Co-4        40                    0.077     0.022    3.44
```

Non-Proportional Hazards Modelling

Default HARE, unstratified dataset

HARE, with default settings. Please refer to Table 1 to interpret the basis functions in the HARE model.

```
nphm_hare <- hare(actg175$time, actg175$cid, actg175_mat)

summary(nphm_hare)
```

```
## dim A/D   loglik       AIC        penalty
##                                  min    max
##   1 Add  -1714.79   3437.25   94.34     Inf
##   2 Add  -1670.85   3357.03      NA      NA
##   3 Add  -1620.45   3263.90   53.55   94.34
```

```
##   4 Add  -1593.67   3218.02   27.38   53.55
##   5 Add  -1579.98   3198.31   21.19   27.38
##   6 Add  -1569.39   3184.78   19.30   21.19
##   7 Add  -1560.07   3173.81      NA      NA
##   8 Add  -1551.71   3164.77      NA      NA
##   9 Del  -1540.44   3149.89   16.91   19.30
##  10 Del  -1531.99   3140.65   14.03   16.91
##  11 Del  -1525.80   3135.96      NA      NA
##  12 Add  -1517.96   3127.93   11.47   14.03
##  13 Add  -1513.19   3126.07      NA      NA
##  14 Add  -1509.24   3125.82      NA      NA
##  15 Add  -1500.87   3116.77      NA      NA
##  16 Del  -1495.01   3112.71    9.57   11.47
##  17 Del  -1490.22   3110.81    9.50    9.57
##  18 Del  -1485.48   3108.98    5.30    9.50
##  19 Del  -1482.83   3111.35    5.29    5.30
##  20 Del  -1480.18   3113.73    4.57    5.29
##  21 Del  -1477.97   3116.96      NA      NA
##  22 Del  -1475.61   3119.93    3.81    4.57
##  23 Del  -1473.71   3123.78    3.81    3.81
##  24 Del  -1471.80   3127.64    3.29    3.81
##  25 Del  -1470.16   3132.02      NA      NA
##  26 Del  -1468.51   3136.40    2.70    3.29
##  27 Add  -1467.16   3141.37    0.00    2.70
##
## the present optimal number of dimensions is 18.
## penalty(AIC) was the default: BIC=log(samplesize): log(2139)=7.67
##
##   dim1           dim2           beta        SE         Wald
## Constant                          -5.6        2.2   -2.58
## Co-15  linear                     0.039      0.016    2.49
## Time       1.3                      -2       0.36   -5.49
## Co-14  linear                    -0.87       0.36   -2.41
## Co-16  linear                  0.00046    8.3e-05    5.55
## Co-10  linear                  0.00046    8.6e-05    5.38
## Co-13  linear                     0.43        0.1    4.19
## Time      0.61                   0.032        1.3    0.02
## Time       1.3 Co-14  linear       2.1       0.45    4.67
## Co-1  linear                     -0.75       0.12   -6.05
## Co-2  linear                      -0.7       0.12   -5.72
## Co-3  linear                     -0.54       0.12   -4.63
## Time      0.61 Co-14  linear      -5.4        1.7   -3.15
## Co-4        45                   0.055      0.017    3.22
## Co-15   1.4e+02                  -0.055      0.017   -3.36
## Co-15   2.2e+02                   0.014     0.0029    4.76
## Co-4  linear                    -0.018      0.008   -2.28
## Co-4  linear   Co-14  linear      0.035     0.0095    3.72
```

As compared with the corresponding HARE model with just cd420 deleted instead of both cd420 and cd820, the variables chosen are the same in this model (without cd820, of course), supporting the omission of cd820.

Compare with corresponding glmnet model

Default HARE, stratified on Offtrt

HARE, with default settings. Please refer to Table 2 to interpret the basis functions in the HARE model.

Not Off-treatment group

```
nphm_hare_not_off <- hare(actg175$time[actg175$offtrt == 0],
                          actg175$cid[actg175$offtrt == 0],
                          mat_not_off)

summary(nphm_hare_not_off)
```

```
## dim A/D   loglik       AIC        penalty
##                                   min    max
##   1 Add  -1028.93   2065.08   88.22     Inf
##   2 Add   -984.82   1984.08   47.21   88.22
##   3 Add   -961.21   1944.08   14.94   47.21
##   4 Del   -954.42   1937.71      NA      NA
##   5 Del   -948.65   1933.39      NA      NA
##   6 Del   -938.80   1920.90   14.29   14.94
##   7 Del   -931.65   1913.83   12.29   14.29
##   8 Del   -925.51   1908.76    8.25   12.29
##   9 Del   -921.39   1907.73    6.61    8.25
##  10 Del   -918.08   1908.34    4.69    6.61
##  11 Del   -915.74   1910.86    4.52    4.69
##  12 Add   -914.76   1916.13      NA      NA
##  13 Del   -911.71   1917.25      NA      NA
##  14 Del   -908.95   1918.95    3.94    4.52
##  15 Add   -906.98   1922.23    3.68    3.94
##  16 Add   -905.14   1925.76    3.65    3.68
##  17 Add   -903.31   1929.32    3.37    3.65
##  18 Del   -901.63   1933.17    3.27    3.37
##  19 Del   -899.99   1937.12    2.94    3.27
##  20 Del   -898.52   1941.39    2.85    2.94
##  21 Del   -897.10   1945.76    2.32    2.85
##  22 Del   -895.99   1950.77      NA      NA
##  23 Add   -894.78   1955.56    2.05    2.32
##  24 Add   -893.81   1960.84      NA      NA
##  25 Add   -892.73   1965.89    0.00    2.05
##
## the present optimal number of dimensions is 9.
## penalty(AIC) was the default: BIC=log(samplesize): log(1363)=7.22
##
##   dim1           dim2           beta        SE         Wald
## Constant                        -0.058       0.34   -0.17
## Time       1.3                      -2       0.24   -8.21
## Co-14  linear                   -0.0079      0.001   -7.74
## Co-15  linear                     0.001    0.00024    4.32
## Co-14   3.6e+02                   0.007     0.0018    3.95
## Co-1  linear                     -0.95       0.17   -5.77
## Co-2  linear                     -0.89       0.17   -5.28
## Co-3  linear                      -0.7       0.15   -4.59
## Co-15   1.2e+03                 -0.0011    0.00039   -2.74
```

No time interactions in the above model. It looks like offtrt bears the brunt of the time interaction.

Off-treatment Group

```
nphm_hare_off <- hare(actg175$time[actg175$offtrt == 1],
                      actg175$cid[actg175$offtrt == 1],
                      mat_off)

summary(nphm_hare_off)
```

```
## dim A/D   loglik       AIC        penalty
##                                    min    max
##   1 Add   -657.77   1322.19   58.00     Inf
##   2 Add   -628.77   1270.84   28.52   58.00
##   3 Add   -614.51   1248.98   18.46   28.52
##   4 Add   -605.28   1237.17   17.00   18.46
##   5 Del   -596.78   1226.82   10.37   17.00
##   6 Del   -591.85   1223.62      NA      NA
##   7 Add   -586.97   1220.53      NA      NA
##   8 Add   -581.22   1215.66    9.69   10.37
##   9 Add   -576.98   1213.84      NA      NA
##  10 Del   -571.52   1209.58    5.18    9.69
##  11 Add   -569.13   1211.45      NA      NA
##  12 Add   -566.34   1212.53    4.36    5.18
##  13 Add   -564.16   1214.82    3.80    4.36
##  14 Del   -562.56   1218.28      NA      NA
##  15 Del   -560.42   1220.65      NA      NA
##  16 Del   -558.45   1223.37    3.46    3.80
##  17 Del   -557.14   1227.40      NA      NA
##  18 Del   -554.99   1229.76    2.62    3.46
##  19 Del   -553.80   1234.03      NA      NA
##  20 Del   -552.37   1237.82    2.60    2.62
##  21 Del   -551.07   1241.88    1.66    2.60
##  22 Add   -550.24   1246.88    0.00    1.66
##
## the present optimal number of dimensions is 10.
## penalty(AIC) was the default: BIC=log(samplesize): log(776)=6.65
##
##   dim1           dim2           beta        SE         Wald
## Constant                          -4.5        2.6   -1.72
## Time      0.61                      -5       0.82   -6.10
## Co-14  linear                     0.045      0.021    2.18
## Co-4  linear                    -0.0014      0.012   -0.11
## Co-13  linear                      0.6       0.15    4.12
## Co-15  linear                    -0.012     0.0034   -3.36
## Co-10  linear                   0.00043    0.00012    3.51
## Co-14   1.4e+02                   -0.05      0.021   -2.37
## Co-15   3.1e+02                   0.012     0.0034    3.49
## Co-4        40                    0.077      0.022    3.44
```

The above model has exactly the same results as its proportional hazards version. It's not the case for the not off-treatment group, however

HARE Survival Curves

Note: If we omit cd420, I think the plot that best represents non-proportional hazards is the plot below "Choosing patients based on offtrt". Based on the HARE summary, the only time interaction present was that for offtrt

Note: when choosing patients, if there are multiple patients that fit the desired characteristics, I choose the patient arbitrarily.

Choosing patients based on offtrt

Note: indexing by row number instead of row name

```
# Patient who did not go off treatment

# The below code chooses the patient with the desired characteristic with the MEDIAN time-to-event
ind <- which(actg175_mat[,14] == 0)
if (length(ind) %% 2 != 0) {
  mid <- ind[length(ind)/2 + .5]
} else {
  mid <- ind[length(ind)/2] # arbitrarily choose lesser one
}
# top number, 1228, is the row name (row number of original matrix). Bottom number, 1380, is the actual
mid
```

```
## 1228
## 1380
```

```
actg175_mat[mid,]
```

```
## trtZDV+ddi trtZDV+ZAL     trtddi        age       wtkg      hemo1
##     0.0000     1.0000     0.0000    37.0000    58.5144     0.0000
##     drugs1     karnof    oprior1    preanti      race1    gender1
##     0.0000   100.0000     0.0000     0.0000     1.0000     0.0000
##   symptom1    offtrt1       cd40       cd80
##     0.0000     0.0000   223.0000   404.0000
```

```
cat("\n")
```

```
# Patient who did go off treatment

ind <- which(actg175_mat[,14] == 1)
if (length(ind) %% 2 != 0) {
  mid <- ind[length(ind)/2 + .5]
} else {
  mid <- ind[length(ind)/2] # arbitrarily choose lesser one
}
mid
```

```
## 2392
##  575
```

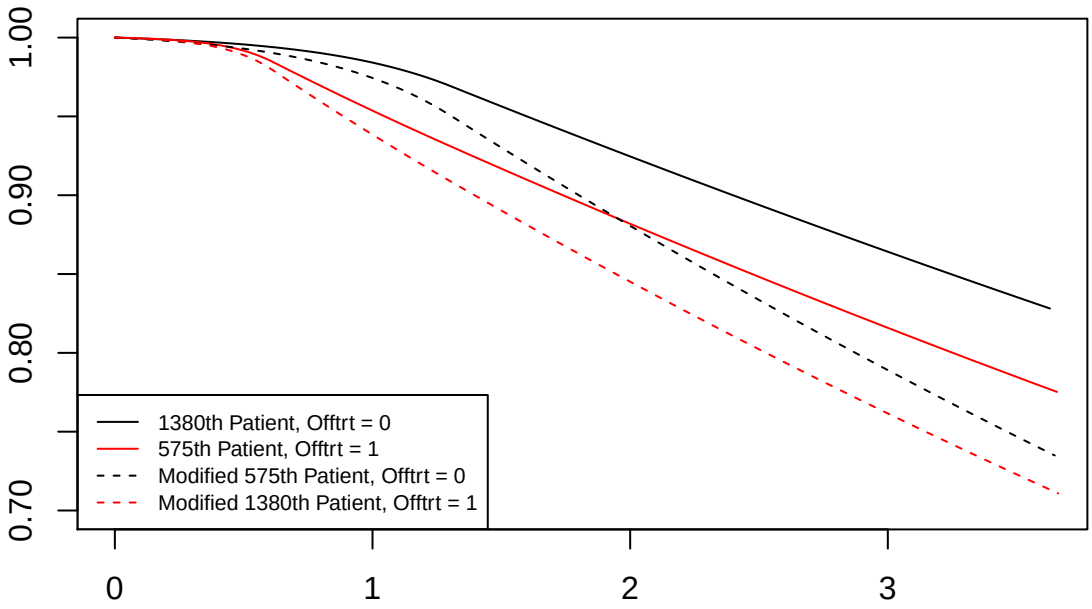
```
actg175_mat[mid,]
```

```
## trtZDV+ddi trtZDV+ZAL     trtddi        age       wtkg      hemo1
##        0.0        1.0        0.0       15.0       46.8        1.0
##     drugs1     karnof    oprior1    preanti      race1    gender1
##        0.0       90.0        0.0      481.0        0.0        1.0
```

```
##   symptom1    offtrt1       cd40       cd80
##        0.0        1.0      311.0      699.0
```

```
cat("\n")
```

```
plot(nphm_hare, actg175_mat[1380,], what = "s", col = 1, ylim = c(.70,1))

plot(nphm_hare, actg175_mat[575,], what = "s", col = 2, add = TRUE)



# patient with same offtrt value as row number 1380 but with other covariates being row 575's
x1 <- actg175_mat[575,]
x1[14] <- 0
plot(nphm_hare, x1, what = "s", col = 1, lty = 2, add = TRUE)

# patient with same offtrt value as row number 575 but with other covariates being row 1380's
x2 <- actg175_mat[1380,]
x2[14] <- 1
plot(nphm_hare, x2, what = "s", col = 2, lty = 2, add = TRUE)

legend("bottomleft",legend=c("1380th Patient, Offtrt = 0", "575th Patient, Offtrt = 1", "Modified 575th
```



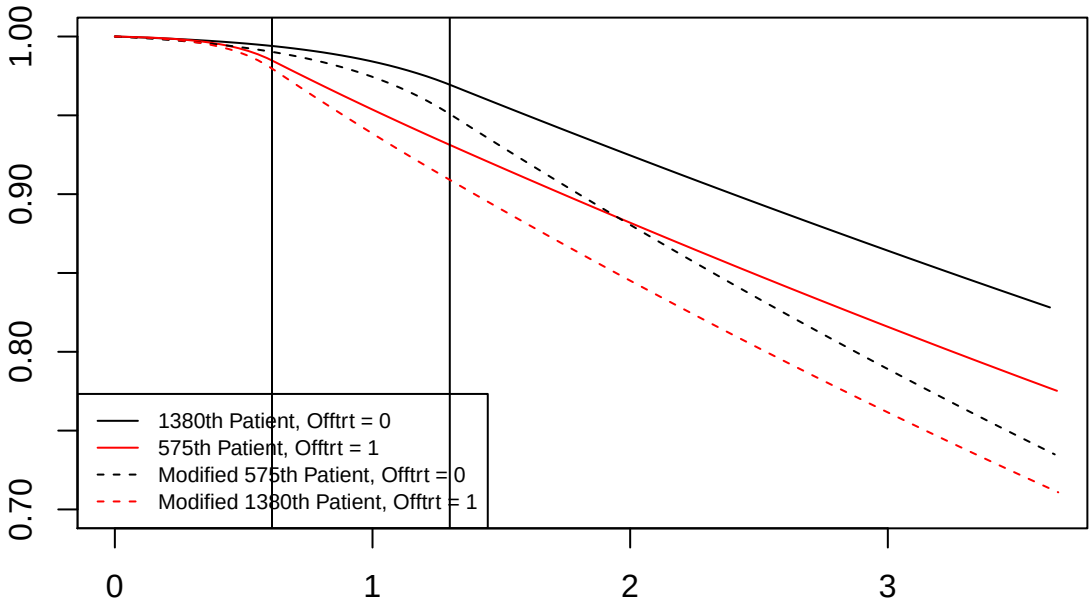
Caption for the above graph: The black and red solid lines refer to the survival curves conditioned on the covariates for the 1380th and 575th patients in the ACTG-175 data set, respectively. Each dashed line has the same Offtrt value as the solid line of the same color, but all other covariates take on the values of the other patient.

Note that when the solid red line is compared with the dashed black line, where only the Offtrt value is different between them, the hazards seem to cross.

Showcases quite a bit of interaction

```
plot(nphm_hare, actg175_mat[1380,], what = "s", col = 1, ylim = c(.70,1))

plot(nphm_hare, actg175_mat[575,], what = "s", col = 2, add = TRUE)
```

```r
abline(v = .61)

abline(v = 1.3)

# patient with same offtrt value as row number 1380 but with other covariates being row 575's
x1 <- actg175_mat[575,]
x1[14] <- 0
plot(nphm_hare, x1, what = "s", col = 1, lty = 2, add = TRUE)

# patient with same offtrt value as row number 575 but with other covariates being row 1380's
x2 <- actg175_mat[1380,]
x2[14] <- 1
plot(nphm_hare, x2, what = "s", col = 2, lty = 2, add = TRUE)

legend("bottomleft",legend=c("1380th Patient, Offtrt = 0", "575th Patient, Offtrt = 1", "Modified 575th
```



The vertical lines indicate the knots of time when interacting with offtrt. See the nphm_hare output above.