

Untitled

KELVIN OFORI-MINTA

2022-06-22

```
lung$Chemo<-factor(lung$Chemo,
                  levels=c("No Chemo", "Chemo"),
                  labels=c("No Chemo", "Chemo"))

lung$AGE_cat<-factor(lung$AGE_cat,
                   levels = c("50-60", "60-70", "70-80", "Above 80", "Below 50"),
                   labels = c("50-60", "60-70", "70-80", "Above 80", "Below 50"))

lung$SEX<-factor(lung$SEX,
                levels=c("Female", "Male"),
                labels=c("Female", "Male"))

lung$CDCC_TOTAL_BEST<-factor(lung$CDCC_TOTAL_BEST,
                             levels = c("0", "1", "2", "3"),
                             labels = c("0", "1", "2", "3"))

lung$TUMOR_SIZE_cat<-factor(lung$TUMOR_SIZE_cat,
                            levels=c("<=1cm", "1cm-2cm", "2cm-3cm", "3cm-4cm", "4cm-5cm"),
                            labels=c("<=1cm", "1cm-2cm", "2cm-3cm", "3cm-4cm", "4cm-5cm"))

lung$GRADE<-factor(lung$GRADE,
                   levels=c("Moderately differentiated", "Poorly differentiated", "Undifferentiated", "Unknown"),
                   labels=c("Moderately differentiated", "Poorly differentiated", "Undifferentiated", "Unknown"))

lung$Pathology<-factor(lung$Pathology,
                      levels = c("Adenocarcinoma", "Other", "Squamous"),
                      labels = c("Adenocarcinoma", "Other", "Squamous"))

lung$Visceral_Pleural_Invasion<-factor(lung$Visceral_Pleural_Invasion,
                                       levels = c("Other", "Present"),
                                       labels = c("Other", "Present"))

lung$LYMPH_VASCULAR_INVASION2<-factor(lung$LYMPH_VASCULAR_INVASION2,
                                       levels=c("Absent", "Present", "Unknown"),
                                       labels=c("Absent", "Present", "Unknown"))

lung$Margins<-factor(lung$Margins,
                    levels = c("Other", "Positive", "Zero"),
                    labels = c("Other", "Positive", "Zero"))

lung$Lymph_Nodes_Sampled<-factor(lung$Lymph_Nodes_Sampled,
```

```

levels = c("<10", ">=10", "Unknown"),
labels = c("<10", ">=10", "Unknown"))

lung$Excision_less_than1<-factor(lung$Excision_less_than1,
levels = c("FALSE", "TRUE"),
labels = c("FALSE", "TRUE"))

```

Partition Data

```

require(caTools)

## Loading required package: caTools
## Warning: package 'caTools' was built under R version 4.0.5

set.seed(1)
split = sample.split(lung$DX_LASTCONTACT_DEATH_MONTHS, SplitRatio = 0.85)
train=subset(lung, split==T)
test=subset(lung, split==F)

```

Survival curve and logrank test for treatment groups

```

library("survival")
library("survminer")

fit1<-survfit(Surv(lung$DX_LASTCONTACT_DEATH_MONTHS,
lung$PUF_VITAL_STATUS) ~ lung$Chemo, data=lung)

#print(fit1)

summary(fit1)$table

##               records n.max n.start events    rmean se(rmean) median
## lung$Chemo=No Chemo   8607  8607   8607   7245 31.46107 0.1442387 31.05
## lung$Chemo=Chemo    1719  1719   1719   1514 31.91488 0.3064293 31.44
##               0.95LCL 0.95UCL
## lung$Chemo=No Chemo   30.72   31.34
## lung$Chemo=Chemo     30.72   32.13

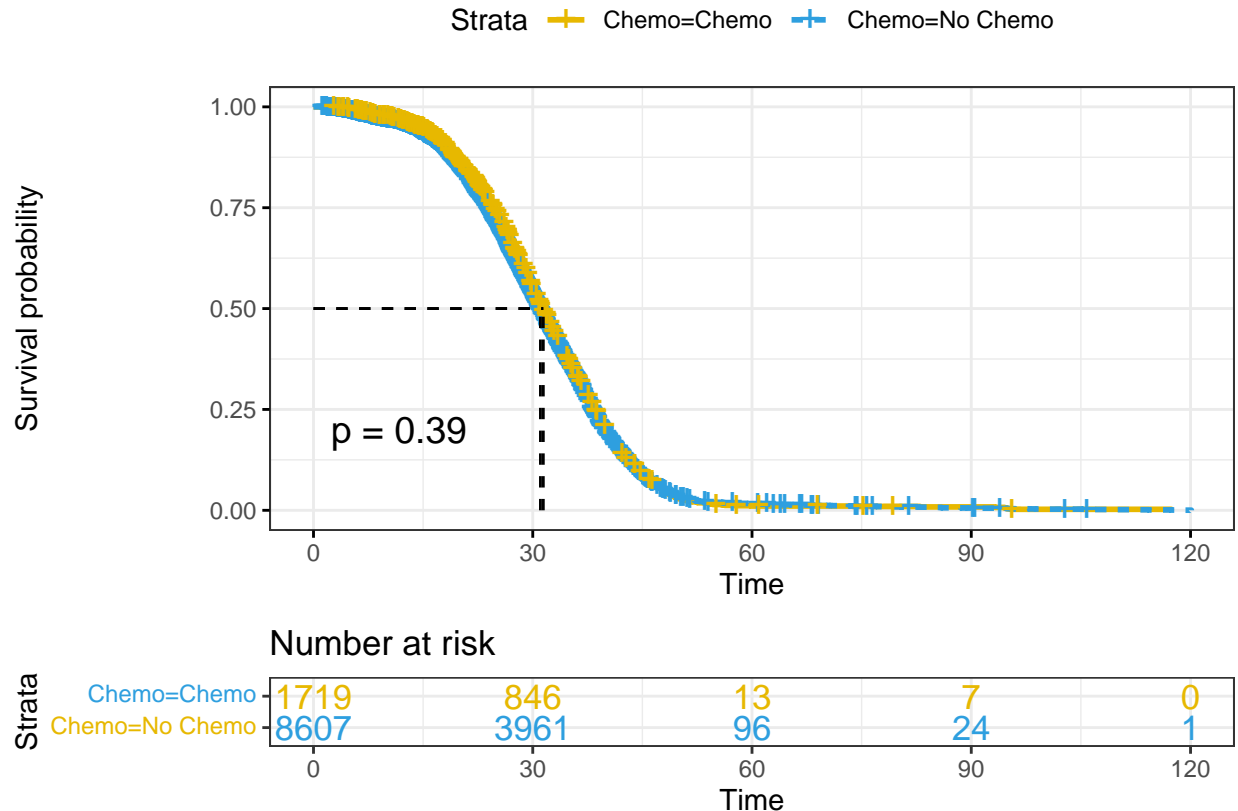
d=data.frame(time=fit1$time,
n.risk=fit1$n.risk,
n.event=fit1$n.event,
n.censor=fit1$n.censor,
surv=fit1$surv)

head(d)

##   time n.risk n.event n.censor    surv
## 1 1.05  8607      1        0 0.9998838
## 2 1.15  8606      1        1 0.9997676
## 3 1.31  8604      0        1 0.9997676
## 4 1.38  8603      0        1 0.9997676
## 5 1.41  8602      0        1 0.9997676
## 6 1.68  8601      0        3 0.9997676

```

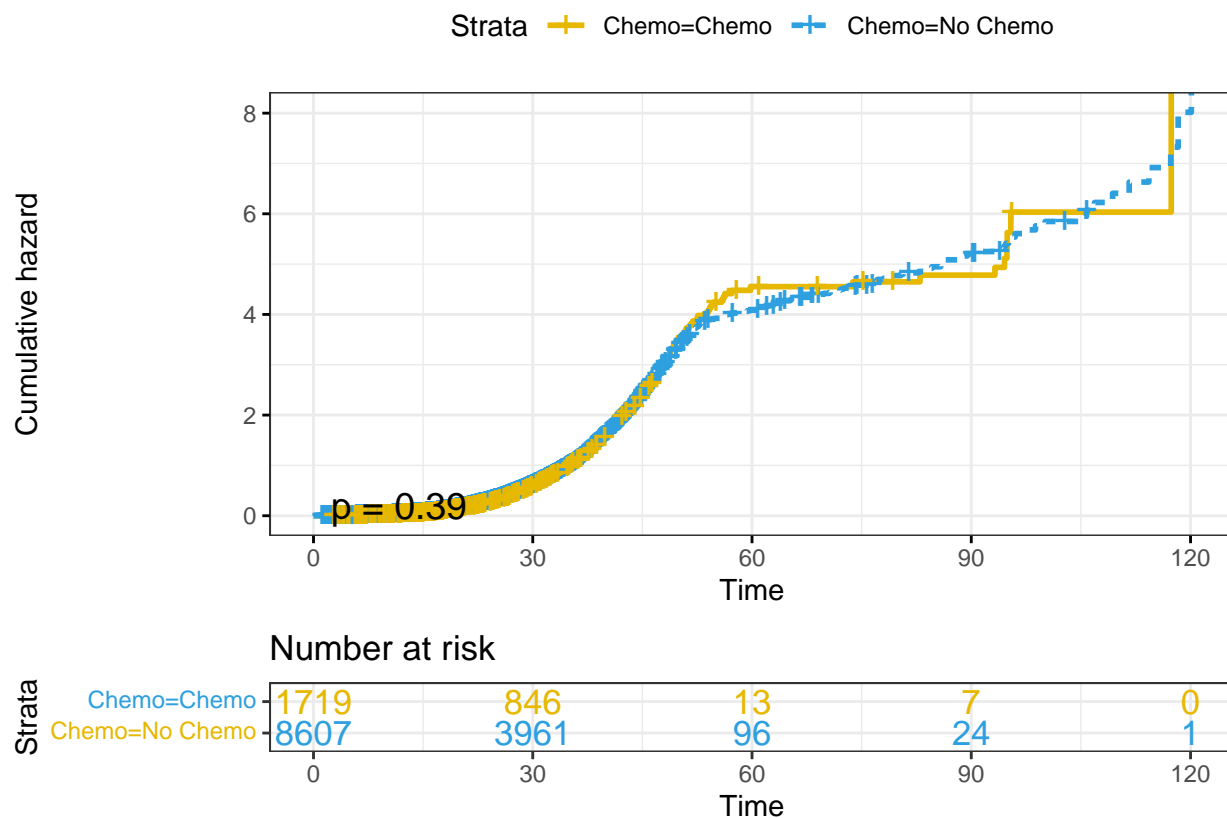
```
ggsurvplot(fit1,
  #legend.labs=c("tumor_free", "with_tumor"),
  pval = TRUE, conf.int = F,
  risk.table = TRUE, # Add risk table
  risk.table.col = "strata", # Change risk table color by groups
  linetype = "strata", # Change line type by groups
  surv.median.line = "hv", # Specify median survival
  ggtheme = theme_bw(), # Change ggplot2 theme
  palette = c("#E7B800", "#2E9FDF"))
```



Logrank test is insignificant = no significant difference between the two treatment groups, but the curve shows a slightly higher survival probability for chemo groups than no_chemo groups. This is violation of ph assumption.

#cumulative hazard plot of treatment groups

```
ggsurvplot(fit1,
  #legend.labs=c("tumor_free", "with_tumor"),
  pval = TRUE, conf.int = F,
  risk.table = TRUE, # Add risk table
  risk.table.col = "strata", # Change risk table color by groups
  linetype = "strata", # Change line type by groups
  #surv.median.line = "hv", # Specify median survival
  ggtheme = theme_bw(), # Change ggplot2 theme
  palette = c("#E7B800", "#2E9FDF"),
  fun = "cumhaz")
```



cox ph model for predictors of mortality

```
library(rms)

## Warning: package 'rms' was built under R version 4.0.5
## Loading required package: Hmisc
## Warning: package 'Hmisc' was built under R version 4.0.5
## Loading required package: lattice
## Loading required package: Formula
##
## Attaching package: 'Hmisc'
## The following objects are masked from 'package:base':
##
##   format.pval, units
## Loading required package: SparseM
##
## Attaching package: 'SparseM'
## The following object is masked from 'package:base':
##
##   backsolve
```

```
lung$Chemo=relevel(as.factor(lung$Chemo), ref="No Chemo")
cph_fit = cph(Surv(DX_LASTCONTACT_DEATH_MONTHS,PUF_VITAL_STATUS) ~ Chemo +
              AGE_cat+
              SEX +
              CDCC_TOTAL_BEST +
              TUMOR_SIZE_cat +
              GRADE +
              Visceral_Pleural_Invasion+
              LYMPH_VASCULAR_INVASION2+
              Margins +
              Lymph_Nodes_Sampled +
              Excision_less_than1,
              data = lung, surv = TRUE, x=TRUE, y=TRUE)
#summary(cph_fit)
cph_fit$coefficients
```

```
##              Chemo=Chemo              AGE_cat=60-70
##              -0.0328318140              -0.0427308354
##              AGE_cat=70-80              AGE_cat=Above 80
##              -0.0325111764              -0.0430066706
##              AGE_cat=Below 50              SEX=Male
##              -0.0909718147              -0.0510728192
##              CDCC_TOTAL_BEST=1              CDCC_TOTAL_BEST=2
##              -0.0021265612              -0.0085580328
##              CDCC_TOTAL_BEST=3              TUMOR_SIZE_cat=1cm-2cm
##              0.0285511916              -0.0839696086
##              TUMOR_SIZE_cat=2cm-3cm              TUMOR_SIZE_cat=3cm-4cm
##              -0.1181544824              -0.0802157062
##              TUMOR_SIZE_cat=4cm-5cm              GRADE=Poorly differentiated
##              -0.0791190101              -0.0002577231
##              GRADE=Undifferentiated              GRADE=Unknown
##              0.0907065052              0.0039066860
##              GRADE=Well differentiated Visceral_Pleural_Invasion=Present
##              -0.0500895804              0.0572584415
##              LYMPH_VASCULAR_INVASION2=Present LYMPH_VASCULAR_INVASION2=Unknown
##              -0.0234184225              -0.1408467926
##              Margins=Positive              Margins=Zero
##              0.0567441423              0.1806203997
##              Lymph_Nodes_Sampled=>=10              Lymph_Nodes_Sampled=Unknown
##              0.0522286042              -0.2225195325
##              Excision_less_than1=TRUE
##              -0.0199983933
```

```
exp(cph_fit$coefficients)
```

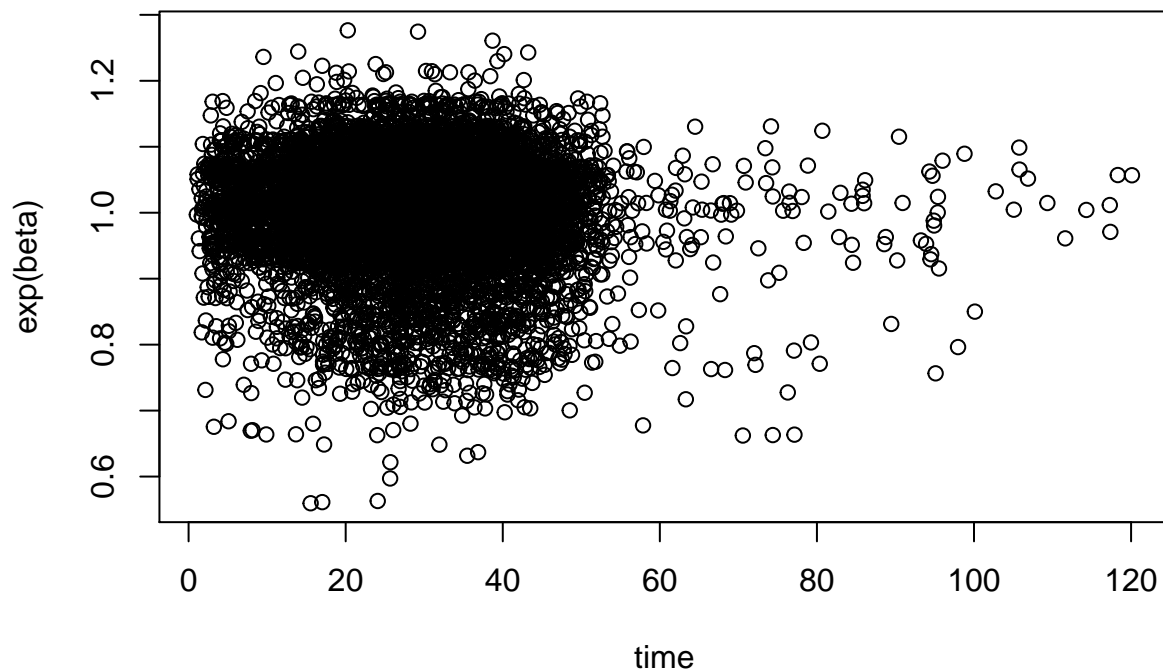
```
##              Chemo=Chemo              AGE_cat=60-70
##              0.96777013              0.9581693
##              AGE_cat=70-80              AGE_cat=Above 80
##              0.9680116              0.9579050
##              AGE_cat=Below 50              SEX=Male
##              0.9130434              0.9502095
##              CDCC_TOTAL_BEST=1              CDCC_TOTAL_BEST=2
##              0.9978757              0.9914785
##              CDCC_TOTAL_BEST=3              TUMOR_SIZE_cat=1cm-2cm
```

```
##          1.0289627          0.9194592
##      TUMOR_SIZE_cat=2cm-3cm      TUMOR_SIZE_cat=3cm-4cm
##          0.8885588          0.9229172
##      TUMOR_SIZE_cat=4cm-5cm      GRADE=Poorly differentiated
##          0.9239300          0.9997423
##      GRADE=Undifferentiated      GRADE=Unknown
##          1.0949476          1.0039143
##      GRADE=Well differentiated Visceral_Pleural_Invasion=Present
##          0.9511442          1.0589294
##      LYMPH_VASCULAR_INVASION2=Present LYMPH_VASCULAR_INVASION2=Unknown
##          0.9768537          0.8686224
##      Margins=Positive      Margins=Zero
##          1.0583850          1.1979603
##      Lymph_Nodes_Sampled=>=10      Lymph_Nodes_Sampled=Unknown
##          1.0536166          0.8004994
##      Excision_less_than1=TRUE
##          0.9802002
```

```
lp=exp(cph_fit$linear.predictors) #linear predictors for all patients
```

```
plot(lung$DX_LASTCONTACT_DEATH_MONTHS, exp(cph_fit$linear.predictors),
     type="p", main = "Hazard rates of all patients vs time",
     ylab = "exp(beta)", xlab = "time")
```

Hazard rates of all patients vs time



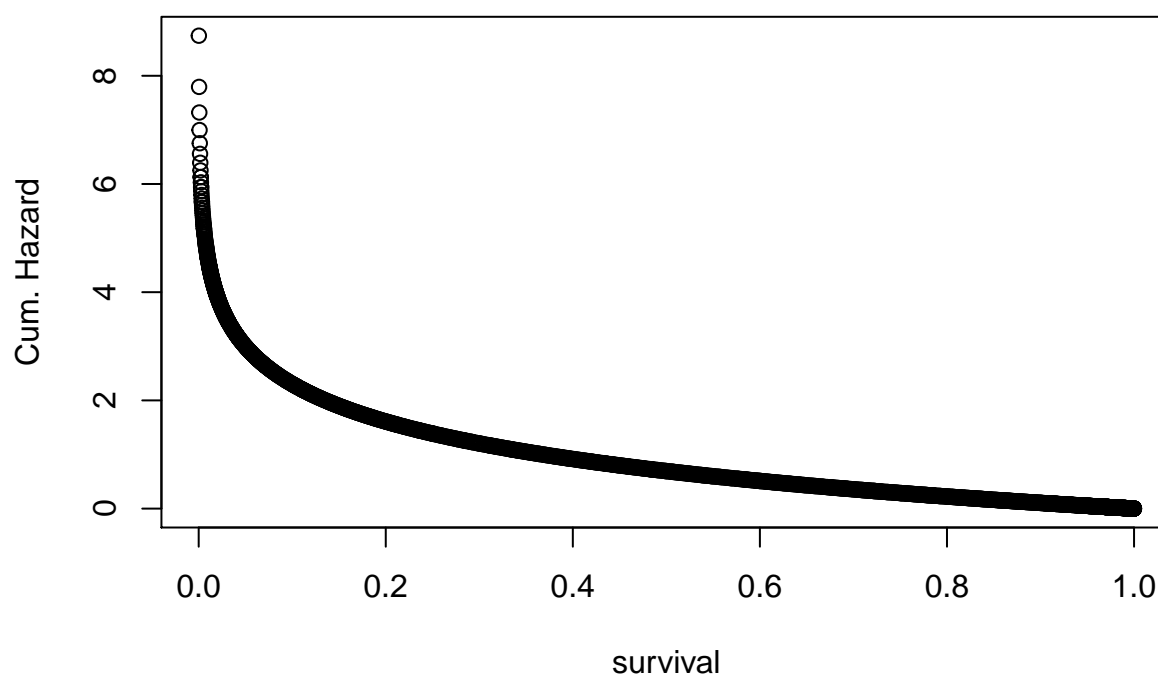
Output Suppressed : Same as Cph model

cumulative Hazard vs Survival Probability

```
surv100=survest(cph_fit)

plot(surv100$surv,surv100$cumhaz,
     main = "Plot of Cum. Hazard vs Survival Probability",
     xlab="survival", ylab = "Cum. Hazard")
```

Plot of Cum. Hazard vs Survival Probability



Smoothing of the Hazard Function - Non Parametric Survival

```
library(bshazard)
```

```
## Warning: package 'bshazard' was built under R version 4.0.5
```

```
## Loading required package: splines
```

```
## Loading required package: Epi
```

```
## Warning: package 'Epi' was built under R version 4.0.5
```

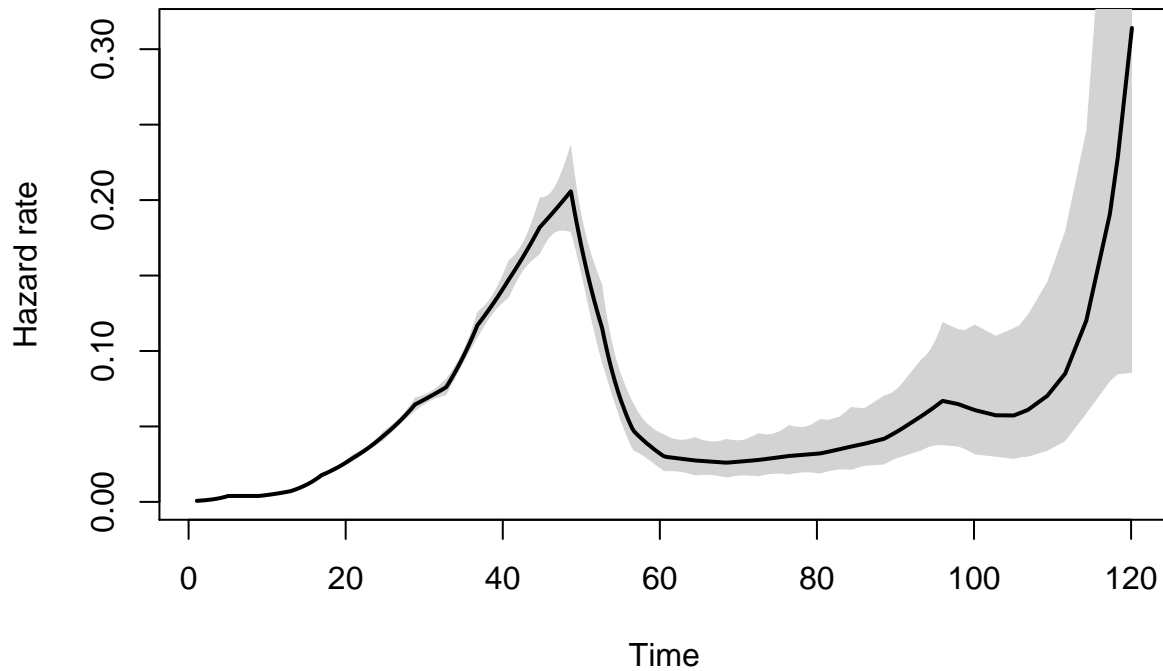
```
mod2<-bshazard(Surv(lung$DX_LASTCONTACT_DEATH_MONTHS,
                    lung$PUF_VITAL_STATUS) ~ 1, data=lung)
```

```
## Iterations: relative error in phi-hat = 1e-04
```

```
## phi= 1.371273   sv2= 0.1795889   df= 16.74352   lambda= 7.635624
```

```
## phi= 1.366469 sv2= 0.2158626 df= 17.46658 lambda= 6.330271
## phi= 1.363351 sv2= 0.2439175 df= 17.96468 lambda= 5.589395
## phi= 1.361396 sv2= 0.2637204 df= 18.2931 lambda= 5.162272
## phi= 1.360199 sv2= 0.2768153 df= 18.50186 lambda= 4.913741
## phi= 1.359476 sv2= 0.2851058 df= 18.63103 lambda= 4.76832
## phi= 1.359043 sv2= 0.2902113 df= 18.70953 lambda= 4.682944
## phi= 1.358786 sv2= 0.2933021 df= 18.75669 lambda= 4.632718
## phi= 1.358633 sv2= 0.2951539 df= 18.78481 lambda= 4.603136
## phi= 1.358543 sv2= 0.2962565 df= 18.80152 lambda= 4.585699
## phi= 1.35849 sv2= 0.2969105 df= 18.81141 lambda= 4.575418
## phi= 1.358458 sv2= 0.2972977 df= 18.81726 lambda= 4.569353
## phi= 1.35844 sv2= 0.2975266 df= 18.82071 lambda= 4.565776
## phi= 1.358429 sv2= 0.2976617 df= 18.82275 lambda= 4.563666
## phi= 1.358422 sv2= 0.2977416 df= 18.82396 lambda= 4.562421
## phi= 1.358418 sv2= 0.2977887 df= 18.82467 lambda= 4.561686
## phi= 1.358416 sv2= 0.2978165 df= 18.82509 lambda= 4.561252
```

```
plot(mod2)
```



```
#predicting risks
```

```
# train$Chemo=relevel(as.factor(train$Chemo), ref="No Chemo")
cphfit_train = cph(
  Surv(DX_LASTCONTACT_DEATH_MONTHS, PUF_VITAL_STATUS) ~ Chemo +
    AGE+
    SEX +
    CDCC_TOTAL_BEST +
```



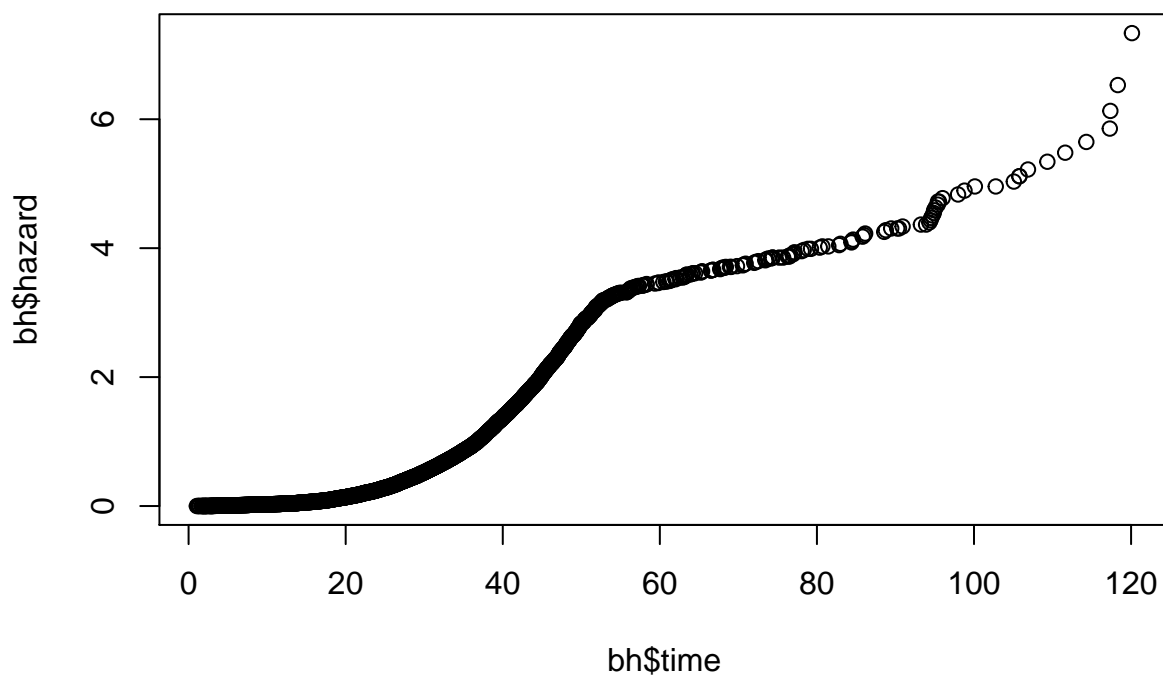
```

TUMOR_SIZE_cat +
GRADE +
Pathology +
Visceral_Pleural_Invasion+
LYMPH_VASCULAR_INVASION2+
Margins +
Lymph_Nodes_Sampled +
Excision_less_than1 ,
data = train, surv = TRUE, x=TRUE, y=TRUE, time.inc=24)

#baseline cumulative hazard function for all covariates set to zero(centered=False)
bh=basehaz(cphfit_train, centered = F)

#plot baseline hazard vs time
plot(bh$time, bh$hazard)

```



```

#length(bh$hazard) #945

```

predict linear predictors for test data