

# Untitled

Jinglun Li; Fuyu Guo; Xinhao Li

12/10/2021

```
library(tidyverse)
library(survival)
library(ggpubr)
library(survminer)
library(glmnet)
library(car)
library(pROC)

setwd("D:\\BST_210_Heart_failure")
dta <- read.csv("heart_f.csv")
# select variables we will use
dta <- dplyr::select(dta,
                     "age", "sex", "anaemia",
                     "diabetes", "ejection_fraction", "smoking",
                     "platelets", "serum_creatinine", "serum_sodium",
                     "time", "DEATH_EVENT")
# rename the variables to make our work easier
names(dta) <- c("age", "sex", "anemia",
                "dbt", "ef", "smoking",
                "plat", "ser_crt", "ser_na",
                "time", "death")
dta$ser_crt_ab <- if_else(dta$ser_crt <= 1.5, 0, 1)
dta$ser_crt_group <- if_else(dta$ser_crt <= 1.5, "normal", "abnormal")
dta$age_65 <- if_else(dta$age >= 65, 1, 0)
dta$ef_group <- case_when(dta$ef <= 30 ~ "low",
                          dta$ef >30 & dta$ef < 45 ~ "normal",
                          dta$ef >=45 ~ "high")
```

## Descpritive analysis

```
library(table1)

## Warning: package 'table1' was built under R version 4.1.2
##
## Attaching package: 'table1'
## The following objects are masked from 'package:base':
##
##      units, units<-
table1(~.|ser_crt_group, data = dta)

## Get nicer `table1` LaTeX output by simply installing the `kableExtra` package
```

	abnormal	normal	Overall
	(N=67)	(N=232)	(N=299)
age			
Mean (SD)	65.0 (12.4)	59.6 (11.5)	60.8 (11.9)
Median [Min, Max]	60.0 [42.0, 95.0]	60.0 [40.0, 95.0]	60.0 [40.0, 95.0]
sex			
Mean (SD)	0.672 (0.473)	0.642 (0.480)	0.649 (0.478)
Median [Min, Max]	1.00 [0, 1.00]	1.00 [0, 1.00]	1.00 [0, 1.00]
anemia			
Mean (SD)	0.403 (0.494)	0.440 (0.497)	0.431 (0.496)
Median [Min, Max]	0 [0, 1.00]	0 [0, 1.00]	0 [0, 1.00]
dbt			
Mean (SD)	0.403 (0.494)	0.422 (0.495)	0.418 (0.494)
Median [Min, Max]	0 [0, 1.00]	0 [0, 1.00]	0 [0, 1.00]
ef			
Mean (SD)	33.8 (11.0)	39.3 (11.8)	38.1 (11.8)
Median [Min, Max]	35.0 [17.0, 70.0]	38.0 [14.0, 80.0]	38.0 [14.0, 80.0]
smoking			
Mean (SD)	0.269 (0.447)	0.336 (0.473)	0.321 (0.468)
Median [Min, Max]	0 [0, 1.00]	0 [0, 1.00]	0 [0, 1.00]
plat			
Mean (SD)	257000 (102000)	265000 (96800)	263000 (97800)
Median [Min, Max]	255000 [51000, 621000]	263000 [25100, 850000]	262000 [25100, 850000]
ser_crt			
Mean (SD)	2.67 (1.60)	1.03 (0.207)	1.39 (1.03)
Median [Min, Max]	2.10 [1.60, 9.40]	1.00 [0.500, 1.50]	1.10 [0.500, 9.40]
ser_na			
Mean (SD)	134 (5.89)	137 (3.61)	137 (4.41)
Median [Min, Max]	134 [113, 146]	137 [125, 148]	137 [113, 148]
time			
Mean (SD)	109 (80.7)	136 (75.8)	130 (77.6)
Median [Min, Max]	94.0 [4.00, 285]	121 [6.00, 280]	115 [4.00, 285]
death			
Mean (SD)	0.642 (0.483)	0.228 (0.421)	0.321 (0.468)
Median [Min, Max]	1.00 [0, 1.00]	0 [0, 1.00]	0 [0, 1.00]
ser_crt_ab			
Mean (SD)	1.00 (0)	0 (0)	0.224 (0.418)
Median [Min, Max]	1.00 [1.00, 1.00]	0 [0, 0]	0 [0, 1.00]
age_65			
Mean (SD)	0.478 (0.503)	0.358 (0.480)	0.385 (0.487)
Median [Min, Max]	0 [0, 1.00]	0 [0, 1.00]	0 [0, 1.00]
ef_group			
high	11 (16.4%)	69 (29.7%)	80 (26.8%)
low	32 (47.8%)	61 (26.3%)	93 (31.1%)
normal	24 (35.8%)	102 (44.0%)	126 (42.1%)

## KM-curves

```
# checked the overall person times
sum(dta$time)
```

```
## [1] 38948
```

```

sum(dta$death == 1)

## [1] 96
median(dta$time)

## [1] 115
fit.km_overall <- survfit(Surv(time, death == 1) ~ 1, data = dta)
quantile(fit.km_overall, 0.25)

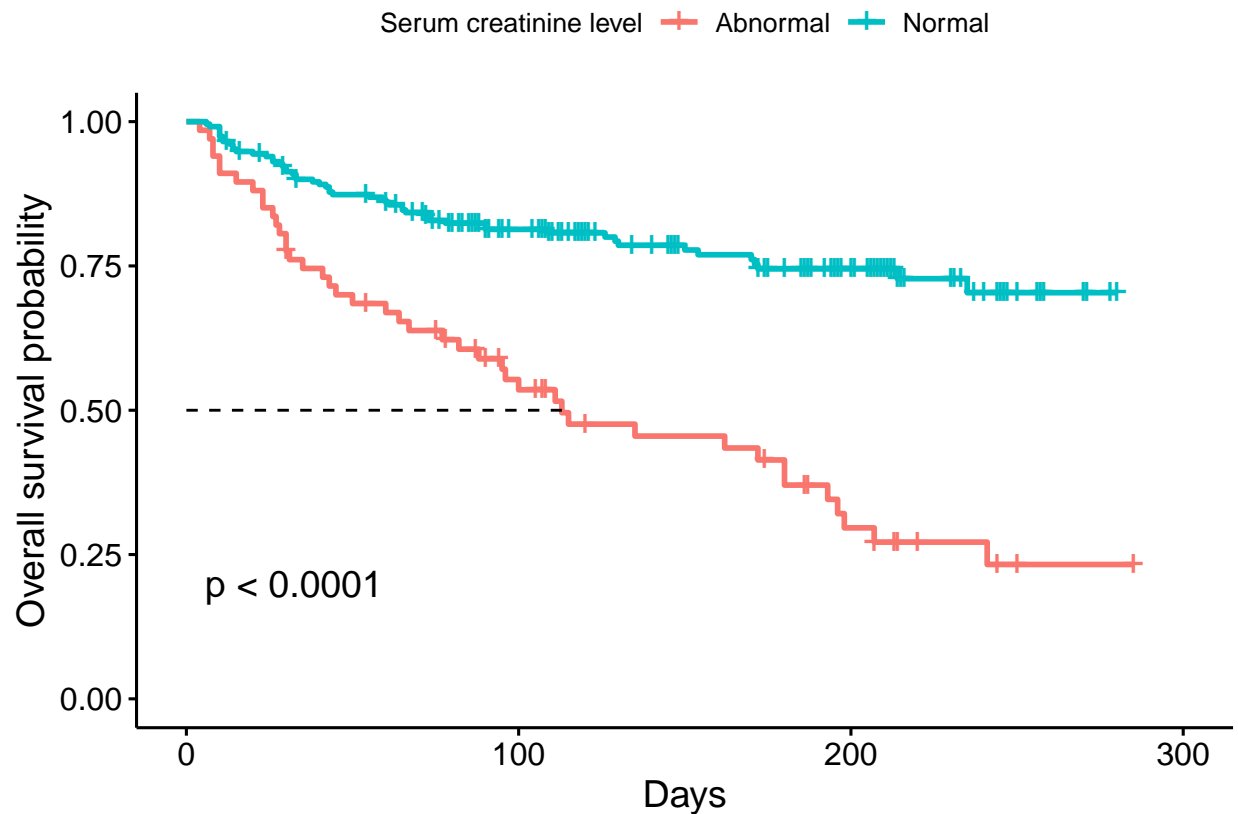
## $quantile
## 25
## 100
##
## $lower
## 25
## 72
##
## $upper
## 25
## 170

fit.km <- survfit(Surv(time, death == 1) ~ ser_crt_group, data = dta)
fit.km

## Call: survfit(formula = Surv(time, death == 1) ~ ser_crt_group, data = dta)
##
##               n events median 0.95LCL 0.95UCL
## ser_crt_group=abnormal  67     43    113     82    196
## ser_crt_group=normal  232     53     NA     NA     NA

ggsurvplot(
  fit = survfit(Surv(time, death) ~ ser_crt_group, data = dta,),
  xlab = "Days",
  ylab = "Overall survival probability",
  surv.median.line = "h",
  legend.title = "Serum creatinine level",
  legend.labs = c("Abnormal", "Normal"),
  pval = TRUE)

```

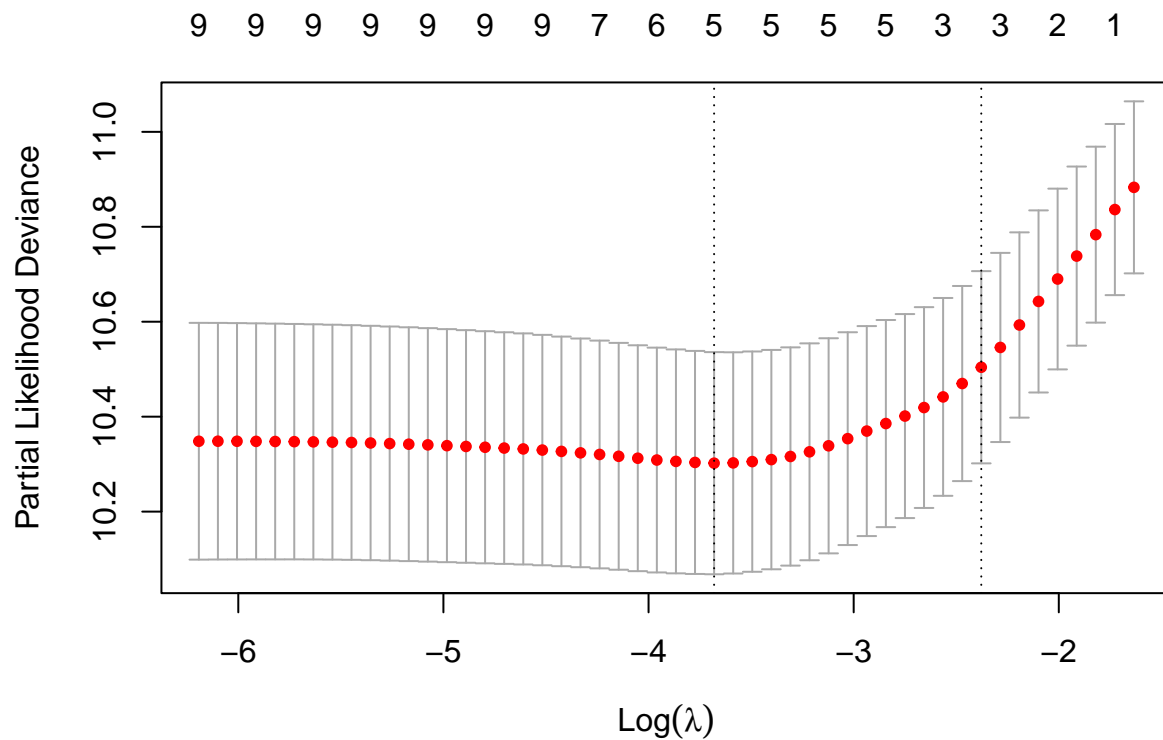


Kaplan-Meier analysis was done to estimate the probability of death events during the follow-up period. The upper survival curve for normal Serum creatinine level is above the lower curve for standard care across the entire. days of follow-up, visually indicating that populations with a normal Serum creatinine level showed a much better survival pattern.

## Lasso

### Binary serum creatinine

```
x <- dplyr::select(dta, -death, -time, -ser_crt, -age_65, -ef_group, -ser_crt_group) %>%
  as.matrix()
y <- dta %>% select( time, death) %>% mutate(status = death) %>% mutate(death = NULL) %>% as.matrix()
cvfit <- cv.glmnet(x, y, family = "cox", type.measure = "deviance")
plot(cvfit)
```



```
cvfit$lambda.min %>% log()
```

```
## [1] -3.680399
```

```
coef(cvfit, s = cvfit$lambda.min)
```

```
## 9 x 1 sparse Matrix of class "dgCMatrix"
```

```
##          1
## age      0.03376827
## sex      .
## anemia   0.27654566
## dbt      .
## ef       -0.02966855
## smoking  .
## plat     .
## ser_na   -0.02324259
## ser_crt_ab 0.76893937
```

```
cvfit$glmnet.fit
```

```
##
```

```
## Call: glmnet(x = x, y = y, family = "cox")
```

```
##
```

```
##      Df %Dev  Lambda
## 1    0 0.00 0.195200
## 2    1 0.69 0.177900
## 3    1 1.22 0.162100
## 4    2 1.65 0.147700
```

```

## 5  2 2.30 0.134600
## 6  2 2.83 0.122600
## 7  3 3.44 0.111700
## 8  3 4.00 0.101800
## 9  3 4.46 0.092740
## 10 3 4.85 0.084500
## 11 3 5.18 0.077000
## 12 4 5.45 0.070160
## 13 4 5.71 0.063920
## 14 5 5.99 0.058250
## 15 5 6.25 0.053070
## 16 5 6.46 0.048360
## 17 5 6.64 0.044060
## 18 5 6.79 0.040150
## 19 5 6.92 0.036580
## 20 5 7.02 0.033330
## 21 5 7.11 0.030370
## 22 5 7.18 0.027670
## 23 5 7.24 0.025210
## 24 6 7.31 0.022970
## 25 6 7.36 0.020930
## 26 6 7.41 0.019070
## 27 6 7.45 0.017380
## 28 7 7.49 0.015830
## 29 7 7.52 0.014430
## 30 9 7.55 0.013150
## 31 9 7.59 0.011980
## 32 9 7.63 0.010910
## 33 9 7.66 0.009944
## 34 9 7.68 0.009061
## 35 9 7.70 0.008256
## 36 9 7.72 0.007523
## 37 9 7.73 0.006854
## 38 9 7.74 0.006245
## 39 9 7.75 0.005691
## 40 9 7.76 0.005185
## 41 9 7.77 0.004724
## 42 9 7.77 0.004305
## 43 9 7.78 0.003922
## 44 9 7.78 0.003574
## 45 9 7.78 0.003256
## 46 9 7.79 0.002967
## 47 9 7.79 0.002703
## 48 9 7.79 0.002463
## 49 9 7.79 0.002244
## 50 9 7.79 0.002045

```

```
coef(cvfit, s = 0.070160)
```

```

## 9 x 1 sparse Matrix of class "dgCMatrix"
##              1
## age          2.122243e-02
## sex          .
## anemia       .
## dbt          .

```

```
## ef          -1.484026e-02
## smoking     .
## plat        .
## ser_na      -1.532881e-05
## ser_crt_ab  6.797683e-01
```

```
log(0.070160)
```

```
## [1] -2.656977
```

```
coef(cvfit, s = 0.019070)
```

```
## 9 x 1 sparse Matrix of class "dgCMatrix"
```

```
##          1
## age      0.03562731
## sex      -0.05220158
## anemia   0.32284543
## dbt      .
## ef       -0.03227083
## smoking  .
## plat     .
## ser_na   -0.02650398
## ser_crt_ab 0.78180514
```

```
log(0.019070)
```

```
## [1] -3.959639
```

### Continuous serum creatinine

```
x <- dplyr::select(dta, -death, -time, -ser_crt_ab, -age_65, -ef_group, -ser_crt_group) %>%
  as.matrix()
y <- dta %>% select( time, death) %>% mutate(status = death) %>% mutate(death = NULL) %>% as.matrix()
cvfit1 <- cv.glmnet(x, y, family = "cox", type.measure = "deviance")
```

```
png("lasso_ser_crt.png", width = 200, height = 200, units = "mm", res = 300)
par(mfrow = c(2,1))
plot(cvfit)
plot(cvfit1)
dev.off()
```

```
## pdf
## 2
```

```
cvfit1$glmnet.fit
```

```
##
## Call:  glmnet(x = x, y = y, family = "cox")
##
##      Df %Dev  Lambda
## 1    0 0.00 0.161500
## 2    2 0.67 0.147100
## 3    3 1.40 0.134100
## 4    3 2.38 0.122200
## 5    3 3.18 0.111300
## 6    3 3.85 0.101400
## 7    3 4.40 0.092410
```

```
## 8 3 4.86 0.084200
## 9 4 5.26 0.076720
## 10 4 5.61 0.069910
## 11 4 5.90 0.063700
## 12 4 6.13 0.058040
## 13 5 6.36 0.052880
## 14 5 6.58 0.048180
## 15 5 6.77 0.043900
## 16 5 6.92 0.040000
## 17 5 7.05 0.036450
## 18 5 7.15 0.033210
## 19 5 7.24 0.030260
## 20 5 7.32 0.027570
## 21 5 7.38 0.025120
## 22 5 7.43 0.022890
## 23 6 7.48 0.020860
## 24 6 7.52 0.019000
## 25 6 7.56 0.017320
## 26 7 7.60 0.015780
## 27 7 7.63 0.014380
## 28 7 7.66 0.013100
## 29 7 7.69 0.011940
## 30 7 7.70 0.010880
## 31 7 7.72 0.009909
## 32 8 7.74 0.009029
## 33 8 7.75 0.008227
## 34 8 7.77 0.007496
## 35 8 7.78 0.006830
## 36 9 7.79 0.006223
## 37 9 7.80 0.005670
## 38 9 7.80 0.005167
## 39 9 7.81 0.004708
## 40 9 7.82 0.004289
## 41 9 7.82 0.003908
## 42 9 7.82 0.003561
## 43 9 7.83 0.003245
## 44 9 7.83 0.002956
## 45 9 7.83 0.002694
## 46 9 7.83 0.002455
## 47 9 7.83 0.002236
## 48 9 7.84 0.002038
```

```
cvfit1$lambda.min
```

```
## [1] 0.0208574
```

```
coef(cvfit1, s = cvfit$lambda.min)
```

```
## 9 x 1 sparse Matrix of class "dgCMatrix"
```

```
## 1
## age 0.03628271
## sex .
## anemia 0.21602781
## dbt .
## ef -0.03714281
## smoking .
```



```

## plat      .
## ser_crt   0.28576411
## ser_na    -0.02677563
log(cvfit1$lambda.min)

## [1] -3.870047
coef(cvfit1, s = 0.020860) # choose the lambda to include 6 variables

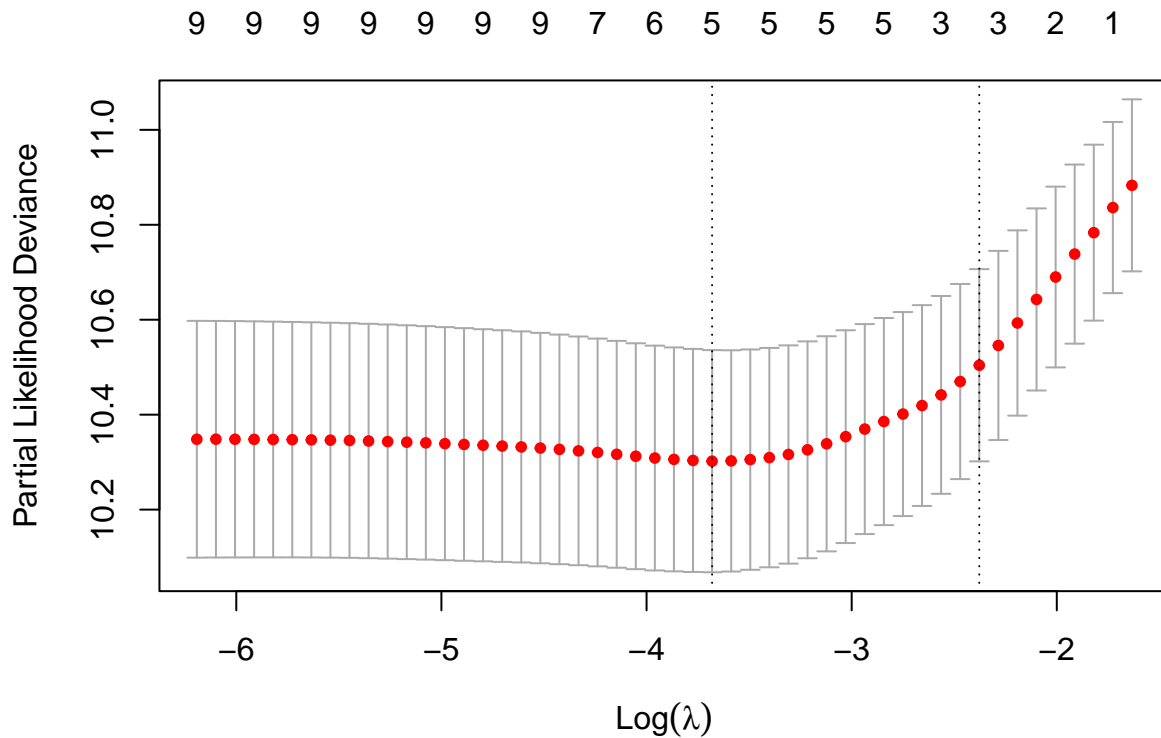
## 9 x 1 sparse Matrix of class "dgCMatrix"
##              1
## age          0.037551488
## sex          -0.008044765
## anemia       0.247750509
## dbt          .
## ef          -0.038790344
## smoking      .
## plat         .
## ser_crt      0.290099601
## ser_na      -0.029043395
log(0.020860)

## [1] -3.869922
coef(cvfit, s = 0.058040) # choose the lambda to include 4 variables

## 9 x 1 sparse Matrix of class "dgCMatrix"
##              1
## age          0.02468668
## sex          .
## anemia       0.02905881
## dbt          .
## ef          -0.01853485
## smoking      .
## plat         .
## ser_na      -0.00592348
## ser_crt_ab   0.70264285
log(0.058040)

## [1] -2.846623
### cvfit <- cv.glmnet(x, y, family = "cox", type.measure = "C")
plot(cvfit)

```



## Cox models

*# Crude models*

```
crude_binary <- coxph(Surv(time, death) ~ ser_crt_ab, ties = "exact", data = dta)
summary(crude_binary)
```

## Call:

```
## coxph(formula = Surv(time, death) ~ ser_crt_ab, data = dta, ties = "exact")
```

##

```
## n= 299, number of events= 96
```

##

```
##          coef exp(coef) se(coef)      z Pr(>|z|)
```

```
## ser_crt_ab 1.2210    3.3906   0.2061 5.926 3.11e-09 ***
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

##

```
##          exp(coef) exp(-coef) lower .95 upper .95
```

```
## ser_crt_ab    3.391    0.2949    2.264    5.078
```

##

```
## Concordance= 0.615 (se = 0.025 )
```

```
## Likelihood ratio test= 31.94 on 1 df,  p=2e-08
```

```
## Wald test               = 35.11 on 1 df,  p=3e-09
```

```
## Score (logrank) test = 39.61 on 1 df,  p=3e-10
```

```
confint(crude_binary) %>% exp()
```

```
##           2.5 %   97.5 %
## ser_crt_ab 2.264055 5.077834

crude_cont <- coxph(Surv(time, death) ~ ser_crt, ties = "exact", data = dta)
summary(crude_cont)
```

```
## Call:
## coxph(formula = Surv(time, death) ~ ser_crt, data = dta, ties = "exact")
##
##      n= 299, number of events= 96
##
##           coef exp(coef) se(coef)      z Pr(>|z|)
## ser_crt 0.29071   1.33738   0.05539 5.249 1.53e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##           exp(coef) exp(-coef) lower .95 upper .95
## ser_crt      1.337      0.7477      1.2      1.491
##
## Concordance= 0.667 (se = 0.029 )
## Likelihood ratio test= 17.88 on 1 df,  p=2e-05
## Wald test              = 27.55 on 1 df,  p=2e-07
## Score (logrank) test = 31.67 on 1 df,  p=2e-08

confint(crude_cont) %>% exp()
```

```
##           2.5 %   97.5 %
## ser_crt 1.1998 1.49073
```

```
# Lasso selection models
```

```
cox1_binary<- coxph(Surv(time, death) ~ age + anemia + ef + ser_crt_ab + ser_na, ties = "exact", data =
summary(cox1_binary)
```

```
## Call:
## coxph(formula = Surv(time, death) ~ age + anemia + ef + ser_crt_ab +
##      ser_na, data = dta, ties = "exact")
##
##      n= 299, number of events= 96
##
##           coef exp(coef) se(coef)      z Pr(>|z|)
## age          0.041280  1.042144  0.008936  4.620 3.84e-06 ***
## anemia        0.469525  1.599234  0.210850  2.227 0.025960 *
## ef           -0.039356  0.961408  0.011081 -3.552 0.000383 ***
## ser_crt_ab    0.820491  2.271615  0.228127  3.597 0.000322 ***
## ser_na       -0.036569  0.964092  0.022376 -1.634 0.102196
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##           exp(coef) exp(-coef) lower .95 upper .95
## age          1.0421      0.9596      1.0241      1.0606
## anemia        1.5992      0.6253      1.0579      2.4176
## ef            0.9614      1.0401      0.9408      0.9825
## ser_crt_ab    2.2716      0.4402      1.4526      3.5524
## ser_na        0.9641      1.0372      0.9227      1.0073
##
```

```
## Concordance= 0.717 (se = 0.028 )
## Likelihood ratio test= 70.6 on 5 df, p=8e-14
## Wald test = 70.81 on 5 df, p=7e-14
## Score (logrank) test = 76.13 on 5 df, p=5e-15

cox1_cont<- coxph(Surv(time, death) ~ age + anemia + ef + ser_crt + ser_na, ties = "exact", data = dta)
summary(cox1_cont)

## Call:
## coxph(formula = Surv(time, death) ~ age + anemia + ef + ser_crt +
## ser_na, data = dta, ties = "exact")
##
## n= 299, number of events= 96
##
##          coef exp(coef) se(coef)      z Pr(>|z|)
## age      0.044071  1.045056  0.008944  4.928 8.33e-07 ***
## anemia   0.406114  1.500973  0.210068  1.933  0.0532 .
## ef      -0.047017  0.954071  0.010438 -4.504 6.66e-06 ***
## ser_crt  0.320567  1.377909  0.073201  4.379 1.19e-05 ***
## ser_na  -0.039726  0.961052  0.023483 -1.692  0.0907 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##          exp(coef) exp(-coef) lower .95 upper .95
## age      1.0451      0.9569      1.0269      1.0635
## anemia   1.5010      0.6662      0.9944      2.2656
## ef       0.9541      1.0481      0.9348      0.9738
## ser_crt  1.3779      0.7257      1.1937      1.5905
## ser_na   0.9611      1.0405      0.9178      1.0063
##
## Concordance= 0.725 (se = 0.028 )
## Likelihood ratio test= 72.15 on 5 df, p=4e-14
## Wald test = 75.43 on 5 df, p=8e-15
## Score (logrank) test = 78.09 on 5 df, p=2e-15

# Add sex as fully adjusted model
cox2_binary <- coxph(Surv(time, death) ~ age + anemia + ef + ser_crt_ab + ser_na + sex, ties = "exact",
summary(cox2_binary)

## Call:
## coxph(formula = Surv(time, death) ~ age + anemia + ef + ser_crt_ab +
## ser_na + sex, data = dta, ties = "exact")
##
## n= 299, number of events= 96
##
##          coef exp(coef) se(coef)      z Pr(>|z|)
## age      0.042006  1.042900  0.008962  4.687 2.77e-06 ***
## anemia   0.473069  1.604913  0.211077  2.241 0.025012 *
## ef      -0.041090  0.959743  0.011222 -3.662 0.000251 ***
## ser_crt_ab 0.827171  2.286839  0.228770  3.616 0.000300 ***
## ser_na  -0.037021  0.963656  0.022410 -1.652 0.098533 .
## sex      -0.225638  0.798007  0.218249 -1.034 0.301202
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
##           exp(coef) exp(-coef) lower .95 upper .95
## age           1.0429    0.9589    1.0247    1.0614
## anemia        1.6049    0.6231    1.0612    2.4273
## ef            0.9597    1.0419    0.9389    0.9811
## ser_crt_ab    2.2868    0.4373    1.4605    3.5807
## ser_na        0.9637    1.0377    0.9222    1.0069
## sex           0.7980    1.2531    0.5203    1.2240
##
## Concordance= 0.718 (se = 0.028 )
## Likelihood ratio test= 71.65 on 6 df, p=2e-13
## Wald test           = 71 on 6 df, p=3e-13
## Score (logrank) test = 76.75 on 6 df, p=2e-14

cox2_cont <- coxph(Surv(time, death) ~ age + anemia + ef + ser_crt + ser_na + sex, ties = "exact", data = dta)
summary(cox2_cont)

## Call:
## coxph(formula = Surv(time, death) ~ age + anemia + ef + ser_crt +
##       ser_na + sex, data = dta, ties = "exact")
##
## n= 299, number of events= 96
##
##           coef exp(coef) se(coef)      z Pr(>|z|)
## age      0.044698  1.045712  0.008967  4.985 6.20e-07 ***
## anemia   0.413956  1.512791  0.210501  1.967  0.0492 *
## ef      -0.048485  0.952672  0.010562 -4.591 4.42e-06 ***
## ser_crt  0.317835  1.374149  0.072937  4.358 1.31e-05 ***
## ser_na  -0.040583  0.960230  0.023457 -1.730  0.0836 .
## sex     -0.194244  0.823457  0.218721 -0.888  0.3745
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##           exp(coef) exp(-coef) lower .95 upper .95
## age           1.0457    0.9563    1.0275    1.0643
## anemia        1.5128    0.6610    1.0014    2.2854
## ef            0.9527    1.0497    0.9332    0.9726
## ser_crt       1.3741    0.7277    1.1911    1.5853
## ser_na        0.9602    1.0414    0.9171    1.0054
## sex           0.8235    1.2144    0.5364    1.2642
##
## Concordance= 0.725 (se = 0.028 )
## Likelihood ratio test= 72.92 on 6 df, p=1e-13
## Wald test           = 75.81 on 6 df, p=3e-14
## Score (logrank) test = 78.54 on 6 df, p=7e-15
```

## Effect modification

```
cox_modi1 <- coxph(Surv(time, death) ~ age + sex + anemia + ef + ser_crt + ser_na + ser_crt*age, ties = "exact", data = dta)
summary(cox_modi1)

## Call:
## coxph(formula = Surv(time, death) ~ age + sex + anemia + ef +
##       ser_crt + ser_na + ser_crt * age, data = dta, ties = "exact")
##
## n= 299, number of events= 96
```

```
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## age          0.0463773 1.0474695 0.0142719 3.250 0.00116 **
## sex          -0.1918913 0.8253966 0.2192812 -0.875 0.38152
## anemia       0.4166415 1.5168586 0.2112173 1.973 0.04854 *
## ef          -0.0488370 0.9523363 0.0108121 -4.517 6.28e-06 ***
## ser_crt      0.3817630 1.4648649 0.4274677 0.893 0.37181
## ser_na      -0.0405270 0.9602833 0.0234366 -1.729 0.08377 .
## age:ser_crt -0.0009801 0.9990204 0.0064757 -0.151 0.87970
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## age          1.0475      0.9547      1.0186      1.0772
## sex          0.8254      1.2115      0.5370      1.2686
## anemia       1.5169      0.6593      1.0027      2.2947
## ef          0.9523      1.0500      0.9324      0.9727
## ser_crt      1.4649      0.6827      0.6338      3.3858
## ser_na       0.9603      1.0414      0.9172      1.0054
## age:ser_crt  0.9990      1.0010      0.9864      1.0118
##
## Concordance= 0.725 (se = 0.028 )
## Likelihood ratio test= 72.95 on 7 df, p=4e-13
## Wald test              = 74.97 on 7 df, p=1e-13
## Score (logrank) test = 93.01 on 7 df, p=<2e-16

cox_modi2 <- coxph(Surv(time, death) ~ age + sex + anemia + ef + ser_crt + ser_na + ser_crt*sex, ties = "exact")
summary(cox_modi2)

## Call:
## coxph(formula = Surv(time, death) ~ age + sex + anemia + ef +
##       ser_crt + ser_na + ser_crt * sex, data = dta, ties = "exact")
##
## n= 299, number of events= 96
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## age          0.045047 1.046077 0.009031 4.988 6.10e-07 ***
## sex          -0.108673 0.897024 0.334286 -0.325 0.745113
## anemia       0.421177 1.523755 0.211442 1.992 0.046378 *
## ef          -0.048999 0.952183 0.010621 -4.613 3.96e-06 ***
## ser_crt      0.342219 1.408069 0.100153 3.417 0.000633 ***
## ser_na      -0.040616 0.960198 0.023463 -1.731 0.083440 .
## sex:ser_crt -0.048295 0.952852 0.142306 -0.339 0.734326
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## age          1.0461      0.9560      1.0277      1.0648
## sex          0.8970      1.1148      0.4659      1.7272
## anemia       1.5238      0.6563      1.0068      2.3062
## ef          0.9522      1.0502      0.9326      0.9722
## ser_crt      1.4081      0.7102      1.1571      1.7135
## ser_na       0.9602      1.0415      0.9170      1.0054
## sex:ser_crt  0.9529      1.0495      0.7209      1.2594
##
```

```
## Concordance= 0.726 (se = 0.028 )
## Likelihood ratio test= 73.04 on 7 df, p=4e-13
## Wald test = 75.48 on 7 df, p=1e-13
## Score (logrank) test = 78.9 on 7 df, p=2e-14

cox_modi3 <- coxph(Surv(time, death) ~ age + sex + anemia + ef + ser_crt + ser_na + ser_crt*anemia, ties = "exact")
summary(cox_modi3)

## Call:
## coxph(formula = Surv(time, death) ~ age + sex + anemia + ef +
## ser_crt + ser_na + ser_crt * anemia, data = dta, ties = "exact")
##
## n= 299, number of events= 96
##
##          coef exp(coef) se(coef)      z Pr(>|z|)
## age          0.044702  1.045717  0.008959  4.990 6.05e-07 ***
## sex         -0.203784  0.815639  0.221974 -0.918  0.3586
## anemia        0.344707  1.411576  0.345485  0.998  0.3184
## ef          -0.049124  0.952063  0.010855 -4.525 6.03e-06 ***
## ser_crt       0.289072  1.335188  0.136462  2.118  0.0341 *
## ser_na      -0.041232  0.959606  0.023582 -1.748  0.0804 .
## anemia:ser_crt 0.040471  1.041301  0.160728  0.252  0.8012
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##          exp(coef) exp(-coef) lower .95 upper .95
## age          1.0457    0.9563    1.0275    1.0642
## sex           0.8156    1.2260    0.5279    1.2602
## anemia        1.4116    0.7084    0.7172    2.7783
## ef            0.9521    1.0504    0.9320    0.9725
## ser_crt       1.3352    0.7490    1.0218    1.7446
## ser_na        0.9596    1.0421    0.9163    1.0050
## anemia:ser_crt 1.0413    0.9603    0.7599    1.4269
##
## Concordance= 0.723 (se = 0.029 )
## Likelihood ratio test= 72.99 on 7 df, p=4e-13
## Wald test = 76.02 on 7 df, p=9e-14
## Score (logrank) test = 79.65 on 7 df, p=2e-14

cox_modi4 <- coxph(Surv(time, death) ~ age + sex + anemia + ef + ser_crt + ser_na + ser_crt*ef, ties = "exact")
summary(cox_modi4)

## Call:
## coxph(formula = Surv(time, death) ~ age + sex + anemia + ef +
## ser_crt + ser_na + ser_crt * ef, data = dta, ties = "exact")
##
## n= 299, number of events= 96
##
##          coef exp(coef) se(coef)      z Pr(>|z|)
## age          0.047525  1.048672  0.009187  5.173 2.30e-07 ***
## sex         -0.190397  0.826631  0.221130 -0.861  0.3892
## anemia        0.387438  1.473202  0.211567  1.831  0.0671 .
## ef          -0.063706  0.938281  0.015335 -4.154 3.27e-05 ***
## ser_crt       0.066794  1.069075  0.197854  0.338  0.7357
## ser_na      -0.045978  0.955063  0.023562 -1.951  0.0510 .
```

```

## ef:ser_crt 0.005615 1.005631 0.003920 1.432 0.1520
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##          exp(coef) exp(-coef) lower .95 upper .95
## age          1.0487    0.9536    1.0300    1.0677
## sex           0.8266    1.2097    0.5359    1.2751
## anemia        1.4732    0.6788    0.9731    2.2302
## ef            0.9383    1.0658    0.9105    0.9669
## ser_crt        1.0691    0.9354    0.7254    1.5755
## ser_na         0.9551    1.0471    0.9120    1.0002
## ef:ser_crt     1.0056    0.9944    0.9979    1.0134
##
## Concordance= 0.723 (se = 0.029 )
## Likelihood ratio test= 74.86 on 7 df, p=2e-13
## Wald test              = 71.18 on 7 df, p=9e-13
## Score (logrank) test = 79.29 on 7 df, p=2e-14

cox_modi5 <- coxph(Surv(time, death) ~ age + sex + anemia + ef + ser_crt + ser_na + ser_crt*ser_na, ties = "exact")
summary(cox_modi5)

## Call:
## coxph(formula = Surv(time, death) ~ age + sex + anemia + ef +
##       ser_crt + ser_na + ser_crt * ser_na, data = dta, ties = "exact")
##
## n= 299, number of events= 96
##
##          coef exp(coef) se(coef)      z Pr(>|z|)
## age          0.046489  1.047587  0.009097  5.110 3.22e-07 ***
## sex         -0.224778  0.798694  0.222559 -1.010 0.31251
## anemia        0.433041  1.541939  0.209525  2.067 0.03876 *
## ef          -0.054926  0.946556  0.011021 -4.984 6.23e-07 ***
## ser_crt     -4.824301  0.008032  2.163122 -2.230 0.02573 *
## ser_na      -0.109897  0.895927  0.038605 -2.847 0.00442 **
## ser_crt:ser_na 0.038140  1.038877  0.015934  2.394 0.01668 *
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##          exp(coef) exp(-coef) lower .95 upper .95
## age          1.047587    0.9546 1.0290732    1.0664
## sex           0.798694    1.2520 0.5163427    1.2354
## anemia        1.541939    0.6485 1.0226324    2.3250
## ef            0.946556    1.0565 0.9263294    0.9672
## ser_crt        0.008032 124.4995 0.0001158    0.5573
## ser_na         0.895927    1.1162 0.8306381    0.9663
## ser_crt:ser_na 1.038877    0.9626 1.0069333    1.0718
##
## Concordance= 0.728 (se = 0.028 )
## Likelihood ratio test= 77.71 on 7 df, p=4e-14
## Wald test              = 73.35 on 7 df, p=3e-13
## Score (logrank) test = 78.93 on 7 df, p=2e-14

#### modification of categorical
cox_modi1 <- coxph(Surv(time, death) ~ age + sex + anemia + ef + ser_crt_ab + ser_na + ser_crt_ab:age,
summary(cox_modi1)

```



```
## Call:
## coxph(formula = Surv(time, death) ~ age + sex + anemia + ef +
##       ser_crt_ab + ser_na + ser_crt_ab:age, data = dta, ties = "exact")
##
## n= 299, number of events= 96
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## age           0.04445   1.04546  0.01201  3.701 0.000214 ***
## sex          -0.22122   0.80154  0.21857 -1.012 0.311470
## anemia        0.48077   1.61732  0.21267  2.261 0.023784 *
## ef           -0.04100   0.95982  0.01122 -3.655 0.000257 ***
## ser_crt_ab     1.18046   3.25589  1.18026  1.000 0.317228
## ser_na       -0.03657   0.96409  0.02236 -1.636 0.101935
## age:ser_crt_ab -0.00537   0.99464  0.01761 -0.305 0.760398
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## age           1.0455      0.9565      1.0211      1.0704
## sex           0.8015      1.2476      0.5223      1.2302
## anemia        1.6173      0.6183      1.0660      2.4537
## ef            0.9598      1.0419      0.9389      0.9812
## ser_crt_ab     3.2559      0.3071      0.3221     32.9087
## ser_na         0.9641      1.0373      0.9227      1.0073
## age:ser_crt_ab 0.9946      1.0054      0.9609      1.0296
##
## Concordance= 0.718 (se = 0.028 )
## Likelihood ratio test= 71.74 on 7 df,  p=7e-13
## Wald test              = 69.67 on 7 df,  p=2e-12
## Score (logrank) test = 82.24 on 7 df,  p=5e-15

cox_modi2 <- coxph(Surv(time, death) ~ age + sex + anemia + ef + ser_crt_ab + ser_na + ser_crt_ab:sex,
summary(cox_modi2)

## Call:
## coxph(formula = Surv(time, death) ~ age + sex + anemia + ef +
##       ser_crt_ab + ser_na + ser_crt_ab:sex, data = dta, ties = "exact")
##
## n= 299, number of events= 96
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## age           0.043070   1.044010  0.009066  4.750 2.03e-06 ***
## sex           0.094411   1.099011  0.297058  0.318 0.750623
## anemia        0.484534   1.623419  0.210160  2.306 0.021136 *
## ef           -0.041799   0.959063  0.011205 -3.730 0.000191 ***
## ser_crt_ab     1.305057   3.687900  0.362646  3.599 0.000320 ***
## ser_na       -0.034168   0.966409  0.022768 -1.501 0.133421
## sex:ser_crt_ab -0.729885   0.481964  0.435215 -1.677 0.093530 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## age           1.0440      0.9578      1.0256      1.0627
## sex           1.0990      0.9099      0.6140      1.9672
## anemia        1.6234      0.6160      1.0753      2.4509
```

```
## ef            0.9591      1.0427      0.9382      0.9804
## ser_crt_ab    3.6879      0.2712      1.8117      7.5070
## ser_na        0.9664      1.0348      0.9242      1.0105
## sex:ser_crt_ab 0.4820      2.0748      0.2054      1.1310
##
## Concordance= 0.723 (se = 0.028 )
## Likelihood ratio test= 74.46 on 7 df, p=2e-13
## Wald test          = 73.52 on 7 df, p=3e-13
## Score (logrank) test = 81.2 on 7 df, p=8e-15

cox_modi3 <- coxph(Surv(time, death) ~ age + sex + anemia + ef + ser_crt_ab + ser_na + ser_crt_ab:anemia, data = dta, ties = "exact")
summary(cox_modi3)
```

```
## Call:
## coxph(formula = Surv(time, death) ~ age + sex + anemia + ef +
##       ser_crt_ab + ser_na + ser_crt_ab:anemia, data = dta, ties = "exact")
##
## n= 299, number of events= 96
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## age            0.041516  1.042390  0.009021  4.602 4.18e-06 ***
## sex            -0.229061  0.795280  0.218452 -1.049 0.294378
## anemia          0.393345  1.481930  0.275876  1.426 0.153924
## ef             -0.041245  0.959594  0.011220 -3.676 0.000237 ***
## ser_crt_ab      0.738893  2.093616  0.302152  2.445 0.014468 *
## ser_na         -0.037728  0.962975  0.022508 -1.676 0.093706 .
## anemia:ser_crt_ab 0.189347  1.208460  0.422791  0.448 0.654261
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##              exp(coef) exp(-coef) lower .95 upper .95
## age            1.0424      0.9593      1.0241      1.0610
## sex            0.7953      1.2574      0.5183      1.2203
## anemia          1.4819      0.6748      0.8630      2.5448
## ef             0.9596      1.0421      0.9387      0.9809
## ser_crt_ab      2.0936      0.4776      1.1580      3.7852
## ser_na          0.9630      1.0384      0.9214      1.0064
## anemia:ser_crt_ab 1.2085      0.8275      0.5277      2.7677
##
## Concordance= 0.718 (se = 0.028 )
## Likelihood ratio test= 71.85 on 7 df, p=6e-13
## Wald test          = 72.51 on 7 df, p=5e-13
## Score (logrank) test = 81.29 on 7 df, p=8e-15

cox_modi4 <- coxph(Surv(time, death) ~ age + sex + anemia + ef + ser_crt_ab + ser_na + ser_crt_ab:ef, data = dta, ties = "exact")
summary(cox_modi4)
```

```
## Call:
## coxph(formula = Surv(time, death) ~ age + sex + anemia + ef +
##       ser_crt_ab + ser_na + ser_crt_ab:ef, data = dta, ties = "exact")
##
## n= 299, number of events= 96
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## age            0.041612  1.042490  0.008971  4.639 3.51e-06 ***
```

```

## sex          -0.218871  0.803425  0.218772 -1.000 0.317092
## anemia       0.452170  1.571720  0.210898  2.144 0.032031 *
## ef          -0.058515  0.943164  0.015054 -3.887 0.000101 ***
## ser_crt_ab   -0.494692  0.609759  0.752107 -0.658 0.510704
## ser_na       -0.040167  0.960629  0.022161 -1.813 0.069908 .
## ef:ser_crt_ab 0.039612  1.040407  0.021352  1.855 0.063565 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##               exp(coef) exp(-coef) lower .95 upper .95
## age              1.0425      0.9592   1.0243   1.0610
## sex              0.8034      1.2447   0.5233   1.2336
## anemia           1.5717      0.6362   1.0396   2.3762
## ef              0.9432      1.0603   0.9157   0.9714
## ser_crt_ab       0.6098      1.6400   0.1396   2.6629
## ser_na           0.9606      1.0410   0.9198   1.0033
## ef:ser_crt_ab    1.0404      0.9612   0.9978   1.0849
##
## Concordance= 0.725 (se = 0.028 )
## Likelihood ratio test= 75.02 on 7 df,  p=1e-13
## Wald test              = 67.99 on 7 df,  p=4e-12
## Score (logrank) test = 76.84 on 7 df,  p=6e-14

cox_modi5 <- coxph(Surv(time, death) ~ age + sex + anemia + ef + ser_crt_ab + ser_na + ser_crt_ab:ser_na)
summary(cox_modi5)

## Call:
## coxph(formula = Surv(time, death) ~ age + sex + anemia + ef +
##       ser_crt_ab + ser_na + ser_crt_ab:ser_na, data = dta, ties = "exact")
##
## n= 299, number of events= 96
##
##               coef exp(coef) se(coef)      z Pr(>|z|)
## age              0.04187  1.04276  0.00901  4.647 3.36e-06 ***
## sex             -0.23694  0.78904  0.21959 -1.079 0.280597
## anemia           0.47024  1.60038  0.21078  2.231 0.025683 *
## ef             -0.04134  0.95950  0.01123 -3.682 0.000231 ***
## ser_crt_ab      -2.39868  0.09084  6.28589 -0.382 0.702760
## ser_na          -0.05191  0.94941  0.03658 -1.419 0.155850
## ser_crt_ab:ser_na 0.02379  1.02407  0.04633  0.513 0.607671
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##               exp(coef) exp(-coef) lower .95 upper .95
## age              1.04276      0.9590 1.025e+00 1.061e+00
## sex              0.78904      1.2674 5.131e-01 1.213e+00
## anemia           1.60037      0.6249 1.059e+00 2.419e+00
## ef              0.95950      1.0422 9.386e-01 9.809e-01
## ser_crt_ab       0.09084     11.0086 4.052e-07 2.036e+04
## ser_na           0.94941      1.0533 8.837e-01 1.020e+00
## ser_crt_ab:ser_na 1.02407      0.9765 9.352e-01 1.121e+00
##
## Concordance= 0.717 (se = 0.028 )
## Likelihood ratio test= 71.91 on 7 df,  p=6e-13
## Wald test              = 70.22 on 7 df,  p=1e-12

```

```
## Score (logrank) test = 76.99 on 7 df, p=6e-14
library(splines2)
library(splines)
nonlinear <- coxph(Surv(time, death) ~ age + sex + anemia + ef + ns(ser_crt, df = 3) + ser_na, ties = "none", data = dta)
range(dta$ser_crt)

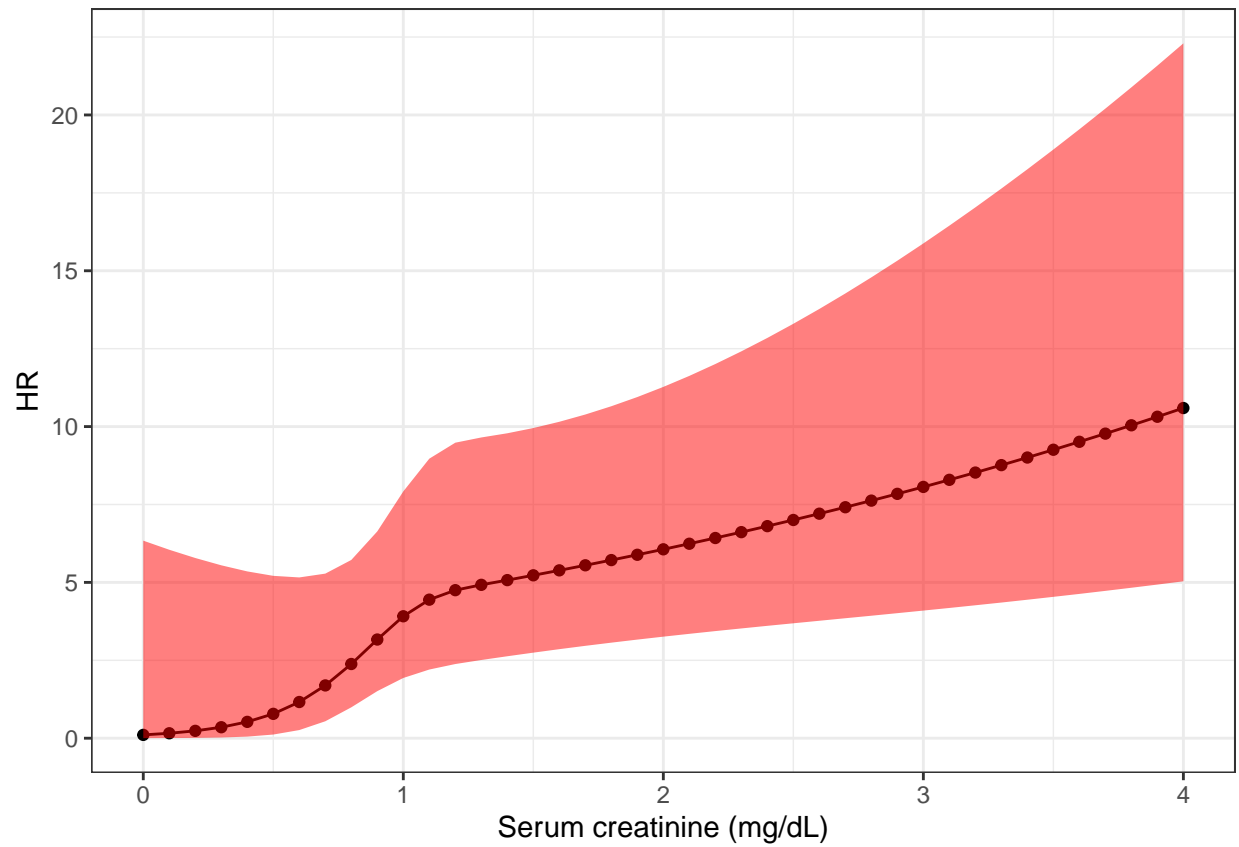
## [1] 0.5 9.4

df <- dta[rep(1, 41),]
df$ser_crt <- seq(0, 4, by = 0.1)
predict <- predict(nonlinear, df, type = "lp", se = T, reference = "strata")
predict$se.fit

##          1          1.1          1.2          1.3          1.4          1.5          1.6          1.7
## 2.0850953 1.8576100 1.6313004 1.4067340 1.1849026 0.9676885 0.7607629 0.5802599
##          1.8          1.9          1.10          1.11          1.12          1.13          1.14          1.15
## 0.4472391 0.3775470 0.3597444 0.3581293 0.3526797 0.3435218 0.3351840 0.3285421
##          1.16          1.17          1.18          1.19          1.20          1.21          1.22          1.23
## 0.3234911 0.3199088 0.3176594 0.3165989 0.3165794 0.3174544 0.3190825 0.3213306
##          1.24          1.25          1.26          1.27          1.28          1.29          1.30          1.31
## 0.3240760 0.3272080 0.3306283 0.3342514 0.3380036 0.3418229 0.3456580 0.3494675
##          1.32          1.33          1.34          1.35          1.36          1.37          1.38          1.39
## 0.3532191 0.3568885 0.3604588 0.3639199 0.3672676 0.3705031 0.3736324 0.3766662
##          1.40
## 0.3796189

linear_differece <- predict$fit - predict$fit[1]
linear_se <- sqrt(predict$se.fit^2 + (predict$se.fit[1])^2)

gg_df <- tibble(crt = seq(0, 4, by = 0.1),
               est = exp(predict$fit),
               low = exp(predict$fit - 1.96*predict$se.fit),
               up = exp(predict$fit + 1.96*predict$se.fit))
ggplot(gg_df) +
  geom_point(aes(x = crt, y = est), color = "black") +
  geom_line(aes(x = crt, y = est), color = "black") +
  geom_ribbon(aes(x = crt, y = est, ymin = low, ymax = up), alpha = 0.5, fill = "red") +
  theme_bw() +
  labs(y = "HR", x = "Serum creatinine (mg/dL)")
```



```
ggsave("nonlinear.png", width = 125, height = 75, units = "mm", dpi = 300)
```