Case Study 7

Piotr Suder

2023-04-15

Basic dependencies:

- R>=4.1 # https://www.r-project.org/
- RStudio>=1.4.1717 # https://posit.co/download/rstudio-desktop/

R Package Dependencies

```r
options(warn=-1)

if (!require(pacman)) install.packages("pacman")
```

```
## Loading required package: pacman
```

```r
pacman::p_load(tidyverse, magrittr, ggpubr, latex2exp, xtable, tidybayes,
               survival, mice, ggsurvfit, gtsummary, survminer, dplyr, ggfortify)

dir.create("Figure", showWarnings = F)
theme_set(theme_classic(base_size = 12))
```

System Information

```r
sessionInfo()
```

```
## R version 4.1.3 (2022-03-10)
## Platform: x86_64-redhat-linux-gnu (64-bit)
## Running under: Fedora Linux 36 (MATE-Compiz)
##
## Matrix products: default
## BLAS/LAPACK: /usr/lib64/libflexiblas.so.3.3
##
## locale:
##  [1] LC_CTYPE=en_US.UTF-8       LC_NUMERIC=C
##  [3] LC_TIME=en_US.UTF-8        LC_COLLATE=en_US.UTF-8
##  [5] LC_MONETARY=en_US.UTF-8    LC_MESSAGES=en_US.UTF-8
##  [7] LC_PAPER=en_US.UTF-8       LC_NAME=C
##  [9] LC_ADDRESS=C               LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] ggfortify_0.4.16 survminer_0.4.9  gtsummary_1.7.0  ggsurvfit_0.3.0
```

```
##  [5] mice_3.15.0        survival_3.5-5    tidybayes_3.0.2   xtable_1.8-4
##  [9] latex2exp_0.9.6   ggpubr_0.6.0      magrittr_2.0.3    lubridate_1.9.2
## [13] forcats_1.0.0     stringr_1.5.0     dplyr_1.1.1       purrr_1.0.1
## [17] readr_2.1.4       tidyr_1.3.0       tibble_3.2.1      ggplot2_3.4.2
## [21] tidyverse_2.0.0   pacman_0.5.1
##
## loaded via a namespace (and not attached):
##  [1] splines_4.1.3         carData_3.0-5         RhpcBLASctl_0.23-42
##  [4] posterior_1.4.1       distributional_0.3.2  ggdist_3.2.1
##  [7] tensorA_0.36.2        yaml_2.3.7            pillar_1.9.0
## [10] backports_1.4.1       lattice_0.21-8        glue_1.6.2
## [13] arrayhelpers_1.1-0    digest_0.6.31        ggsignif_0.6.4
## [16] checkmate_2.1.0       colorspace_2.1-0      htmltools_0.5.5
## [19] Matrix_1.5-4          pkgconfig_2.0.3      broom_1.0.4
## [22] svUnit_1.0.6          scales_1.2.1         km.ci_0.5-6
## [25] KMsurv_0.1-5          tzdb_0.3.0            timechange_0.2.0
## [28] generics_0.1.3        farver_2.1.1          car_3.1-2
## [31] withr_2.5.0           cli_3.6.1             evaluate_0.20
## [34] fansi_1.0.4           broom.helpers_1.13.0 rstatix_0.7.2
## [37] xml2_1.3.3            tools_4.1.3          data.table_1.14.8
## [40] hms_1.1.3             lifecycle_1.0.3      munsell_0.5.0
## [43] compiler_4.1.3        rlang_1.1.0          grid_4.1.3
## [46] gt_0.9.0              rstudioapi_0.14      rmarkdown_2.21
## [49] gtable_0.3.3          abind_1.4-5          R6_2.5.1
## [52] gridExtra_2.3         zoo_1.8-11           knitr_1.42
## [55] fastmap_1.1.1         survMisc_0.5.6       utf8_1.2.3
## [58] stringi_1.7.12        parallel_4.1.3       Rcpp_1.0.10
## [61] vctrs_0.6.1           tidyselect_1.2.0     xfun_0.38
## [64] coda_0.19-4
```

Pre-processing

The following code is used for data preprocessing and imputation. Due to long runtime this part is commented out and the result stored in the data_full.RDS object is loaded.

```
# data = read.csv('dukecathr.csv')
#
# ID.vec = unique(data$RSUBJID)
#
# death.adj = vector(length = nrow(data))
# time = vector(length = nrow(data))
# prev.cat = vector(length = nrow(data))
#
# count = 0
# cutoff = 2*365
# for (i in 1:length(ID.vec)){
#   if (i %% 300 == 0)
#   {
#     cat('progress: ', (i/length(ID.vec))*100, '% \n')
#   }
#   id = ID.vec[i]
#   cath.dat = data[data$RSUBJID == id,]
#   tot.cath = nrow(cath.dat)
#
```

```
#    #cat('id: ', id, ' tot.cath: ', tot.cath, '\n')
#    for (j in 1:tot.cath)
#    {
#      count = count + 1
#      if (j == 1)
#      {
#        prev.cat[count] = 0
#      }
#      if (j > 1)
#      {
#        t.diff = cath.dat$DAYS2LKA[j-1] - cath.dat$DAYS2LKA[j]
#        if (t.diff >= cutoff)
#        {
#          prev.cat[count] = 1
#        }
#        if (t.diff < cutoff)
#        {
#          prev.cat[count] = 2
#        }
#      }
#      if (j == tot.cath)
#      {
#        time[count] = cath.dat$DAYS2LKA[j]
#        death.adj[count] = cath.dat$DEATH[j]
#      }
#      if (j < tot.cath)
#      {
#        time[count] = cath.dat$DAYS2LKA[j] - cath.dat$DAYS2LKA[j+1]
#        death.adj[count] = 0
#      }
#    }
# }
#
# data$prev_cat = prev.cat
# data$death_adj = death.adj
# data$survtime = time
#
# # recode: DPCABG, DPMI, DPPCI,
# # collapse: CHFSEV
#
# # weight and height encode as BMI and maybe interact with age
# # recde age to the median of each category
#
#
# # potentially exclude: CATHAPPR (only take left heart)
#
# # TARGET: DAYS2LKA
#
# # Make binary and set as did not happen if there is a catherization happening before it: DSCABG, DSMI
#
#
#
# data$past_CABG = 1*(!is.na(data$DPCABG) & data$DPCABG > 2*365) + 2*(!is.na(data$DPCABG) & data$DPCABG
```

```r
# data$past_MI = 1*(!is.na(data$DPMI) & data$DPMI > 365) + 2*(!is.na(data$DPMI) & data$DPMI <= 365)
# data$past_PCI = 1*(!is.na(data$DPPCI) & data$DPPCI > 365) + 2*(!is.na(data$DPPCI) & data$DPPCI <= 365)
#
#
# data$CHF_severity = 0*(data$CHFSEV == 0) + 1*(data$CHFSEV == 1 | data$CHFSEV == 2) + 2*(data$CHFSEV ==
# data$CHF_severity[is.na(data$CHFSEV)] = NA
#
# data$BMI = data$WEIGHT_R / ((data$HEIGHT_R/100)^2)
#
#
# data$age = 21*(data$AGE_G == 1) + 27*(data$AGE_G == 2) + 32*(data$AGE_G == 3) + 37*(data$AGE_G == 4) +
#   42*(data$AGE_G == 5) + 47*(data$AGE_G == 6) + 52*(data$AGE_G == 7) + 57*(data$AGE_G == 8) +
#   62*(data$AGE_G == 9) + 67*(data$AGE_G == 10) + 72*(data$AGE_G == 11) + 77*(data$AGE_G == 12) +
#   85*(data$AGE_G == 13)
#
#
# # Change year into 3 categories: before coronary stents, before drug-eluting stents, and after both
# data = data|> mutate(year = case_when(
#   YRCATH_G %in% c(1,2) ~ 1, # coronary stents were approved by the FDA in 1994
#   YRCATH_G %in% c(3,4) ~ 2, # first drug-eluting stents were approved in 2003
#   TRUE ~ 3
# ))
#
# # DSCABG, DSMI, DSPCI,DSSTROKE
#
# data$subsequent_CABG = 1*(!is.na(data$DSCABG) & (data$DSCABG <= data$survtime))
# data$subsequent_MI = 1*(!is.na(data$DSMI) & (data$DSMI <= data$survtime))
# data$subsequent_PCI = 1*(!is.na(data$DSPCI) & (data$DSPCI <= data$survtime))
# data$subsequent_stroke = 1*(!is.na(data$DSSTROKE) & (data$DSSTROKE <= data$survtime))
#
# ####################################
#
# # Definitely exclude: FUPROTCL, RDAYSFROMINDEX, HXCHF (we already have CHF severity), NUMPRMI ?
#
# data_proc = data %>% select(c('survtime', 'death_adj', 'age','year', 'GENDER', 'RACE_G', 'ACS', 'CHF_
#                               'HXANGINA', 'HXCEREB', 'HXCOPD', 'HXDIAB', 'HXHTN', 'HXHYL', 'HXMI', 'H
#                               'NUMPRMI', 'DIASBP_R', 'PULSE_R', 'SYSBP_R', 'CBRUITS', 'BMI', 'S3',
#                               'CREATININE_R', 'HDL_R', 'LDL_R', 'TOTCHOL_R',
#                               'CATHAPPR', 'DIAGCATH', 'INTVCATH', 'CORDOM',
#                               'GRAFTST', 'LADST', 'LCXST', 'LMST', 'LVEF_R',
#                               'NUMDZV', 'PRXLADST', 'RCAST'))
#
# data_complete = data_proc %>% na.omit()
#
# head(data_complete)
#
# nrow(data_complete)
#
# data_mice = data_proc %>%
#   mutate(GENDER = factor(GENDER), RACE_G = factor(RACE_G), ACS = factor(ACS), year = factor(year),
#          CHF_severity = factor(CHF_severity), past_CABG = factor(past_CABG), past_MI = factor(past_MI
#          past_PCI = factor(past_PCI), HXANGINA = factor(HXANGINA), HXCEREB = factor(HXCEREB), HXCOPD=
#          HXDIAB = factor(HXDIAB), HXHTN = factor(HXHTN), HXHYL = factor(HXHYL), HXMI = factor(HXMI),
```
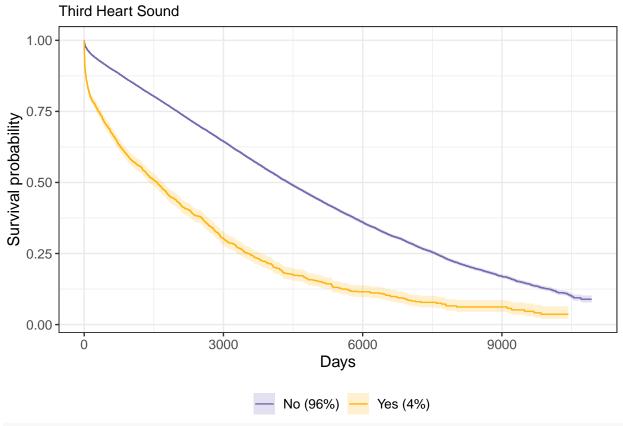
```r
#           HXSMOKE = factor(HXSMOKE), CBRUITS = factor(CBRUITS), S3 = factor(S3), CATHAPPR = factor(CATI
#           DIAGCATH = factor(DIAGCATH), INTVCATH = factor(INTVCATH), CORDO = factor(CORDOM))
#
#
# mice_obj <- mice(data = data_mice, m = 1)
# data_full <- complete(mice_obj)
#
# saveRDS(data_full, "data_full.RDS")

# Define colors
purple <- "#756bb1"
orange <- "#ffb50f"

data_full <- readRDS("data_full.RDS")
data.subject <- data_full %>%
  mutate(DEATH2 = ifelse(death_adj == 0, "Censored", "Dead"),
         DAYS2LKA = survtime, DEATH = death_adj)

PlotKMCurve <- function(group_var, var_name = NULL) {
  if (is.null(var_name)) var_name <- group_var
  survfit2(Surv(DAYS2LKA + 0.1, DEATH) ~ get(group_var), data = data.subject) %>%
    ggsurvfit() +
    add_confidence_interval() +
    labs(
      x = "Days",
      y = "Survival probability",
      fill = element_blank(), color = element_blank()
    )
}

fig1 <- PlotKMCurve("S3", "Third Heart Sound") +
  labs(subtitle = "Third Heart Sound") +
  scale_color_manual(values = c(purple, orange), labels = c("No (96%)", "Yes (4%)")) +
  scale_fill_manual(values = c(purple, orange), labels = c("No (96%)", "Yes (4%)"))
fig2 <- PlotKMCurve("past_CABG", "Past CABG") +
  labs(subtitle = "Closest Coronary Artery Bypass Surgery") +
  scale_color_manual(values = c("skyblue2", purple, orange), labels = c("Never (76%)", "More than 2 year
  scale_fill_manual(values = c("skyblue2", purple, orange), labels = c("Never (76%)", "More than 2 years
fig3 <- PlotKMCurve("CATHAPPR", "CATHAPPR") +
  labs(subtitle = "Type of Cardiac Catheterization") +
  scale_color_manual(values = c(orange, "skyblue2", purple, "darkgrey"),
                     labels = c("Unknown (2%)", "Right (2%)",
                                "Left (85%)",
                                "Right and Left (11%)")) +
  scale_fill_manual(values = c(orange, "skyblue2", purple, "black"),
                    labels = c("Unknown (2%)", "Right (2%)",
                               "Left (85%)",
                               "Right and Left (11%)"))

pdf("Figure/eda1.pdf", height = 3.5, width = 6)
fig1
dev.off()

## pdf
```
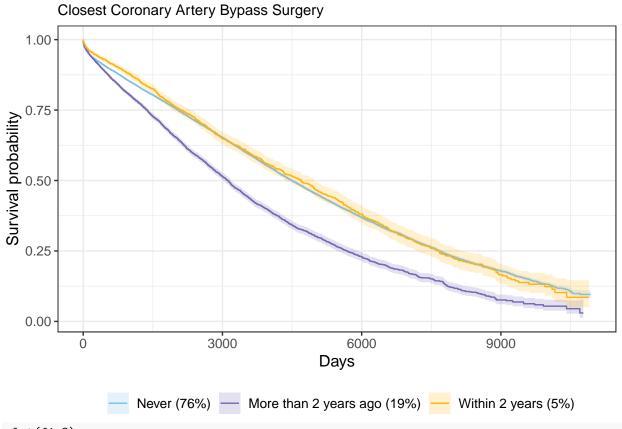
```
##     2
```
```
pdf("Figure/eda2.pdf", height = 3.5, width = 6)
fig2
dev.off()
```
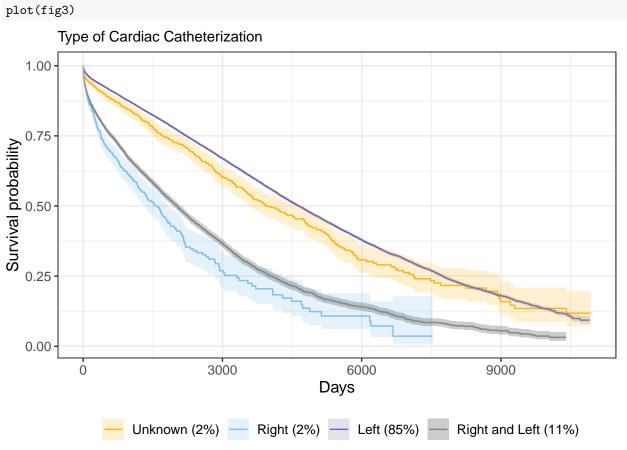```
pdf
2
```
```
pdf("Figure/eda3.pdf", height = 3.5, width = 6)
fig3
dev.off()
```
```
pdf
2
```
```
plot(fig1)
```



```
plot(fig2)
```

Closest Coronary Artery Bypass Surgery

Never (76%) — More than 2 years ago (19%) — Within 2 years (5%)

```
plot(fig3)
```

Type of Cardiac Catheterization

Unknown (2%) — Right (2%) — Left (85%) — Right and Left (11%)

Standardize using the method by Andrew Gelman.

```
cont.names = c('age', 'DIASBP_R', 'PULSE_R', 'SYSBP_R', 'BMI', 'CREATININE_R', 'HDL_R', 'LDL_R', 'TOTCHO
                'GRAFTST', 'LADST', 'LCXST', 'LMST', 'LVEF_R', 'PRXLADST', 'RCAST')

data_full_std = data_full

for (i in 1:ncol(data_full_std))
{
  if (names(data_full_std)[i] %in% cont.names)
  {
    #print(names(data_full_std)[i] )
    data_full_std[,i] = (data_full_std[,i] - mean(data_full_std[,i]))/ (2*sd(data_full_std[,i]))
  }
}

head(data_full_std)
```

```
##   survtime death_adj       age year GENDER RACE_G ACS CHF_severity past_CABG
## 1      134         1 -0.4243467    3      0      1   0            2         0
## 2     5379         0 -0.4243467    2      0      1   0            0         1
## 3     3385         1  0.4082092    2      1      1   0            0         0
## 4        6         0 -0.8406246    1      0      2   1            0         0
## 5        3         0 -0.8406246    1      0      2   0            0         0
## 6        7         0 -0.8406246    1      0      2   1            0         0
##   past_MI past_PCI HXANGINA HXCEREB HXCOPD HXDIAB HXHTN HXHYL HXMI HXSMOKE
## 1       1        1        0       0      0      1     1     0    1       0
## 2       0        0        1       0      0      0     1     1    0       1
## 3       0        0        1       0      0      0     0     1    0       0
## 4       2        0        0       0      0      0     1     0    1       1
## 5       2        2        0       0      0      0     1     0    1       1
## 6       2        2        0       0      0      0     1     0    1       1
##   NUMPRMI    DIASBP_R     PULSE_R     SYSBP_R CBRUITS         BMI S3 CREATININE_R
## 1       1   0.7585088 -0.08448485   0.2680095       0  0.75738190  0   0.15819532
## 2       0   1.1849599  0.16466275   0.8601824       0  0.41391530  0  -0.09462577
## 3       0   0.3320577  0.60759183   0.7073636       0 -0.50210686  0  -0.09462577
## 4       1   0.4742080  0.52454263  -0.3241634       0 -0.02089169  0   0.66383751
## 5       1   0.2965201  0.19234582   0.1533954       0 -0.02089169  0   2.33877727
## 6       1  -0.2365438 -0.47204780  -0.5151869       0 -0.02089169  0  -0.12622841
##         HDL_R       LDL_R    TOTCHOL_R CATHAPPR DIAGCATH INTVCATH CORDOM
## 1 -0.25243793 -0.56403193 -0.76688913        3        1        0      2
## 2 -0.13330256 -0.01736401  0.10553713        2        1        0      2
## 3  0.19431972 -0.32486472 -0.07090863        2        1        1      2
## 4  0.32834702 -0.02875292 -0.17873660        3        1        1      2
## 5  0.03050858  0.27874778  0.36040323        2        1        0      2
## 6  0.10496819  0.06235840  0.11533967        2        1        1      2
##       GRAFTST       LADST       LCXST       LMST      LVEF_R NUMDZV    PRXLADST
## 1   0.2284010 -0.31793405 -0.4561525 -0.2814373 -0.88882519      1  -0.4129926
## 2   0.3044056  0.46027310 -0.1242779 -0.2814373  0.50339334      2   0.8544597
## 3   0.3044056  0.07116953 -0.1242779 -0.2814373  0.02457268      1  -0.4129926
## 4  -0.8356644 -1.09614119 -0.1242779 -0.2814373  0.09582829      1  -0.4129926
## 5  -1.2156877 -1.09614119 -0.7880272 -0.2814373 -0.08231073      1  -0.4129926
## 6  -0.8356644 -1.09614119  0.4730965 -0.2814373 -0.11793854      1  -0.4129926
##        RCAST CORDO
## 1  0.1374253     2
```

```
## 2  0.4763559      2
## 3 -0.5404357      2
## 4  0.4085698      2
## 5  0.4085698      2
## 6  0.4085698      2
```
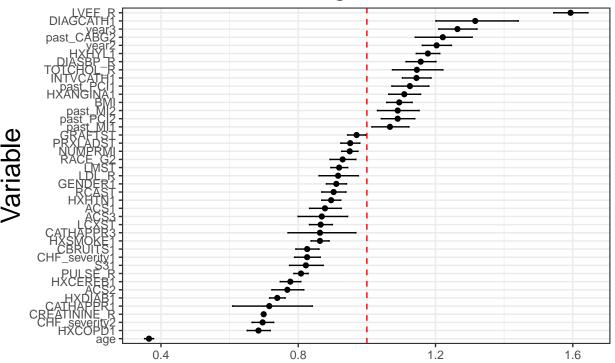
AFT model

```
survmodel = survreg(Surv(survtime + 0.1, death_adj)~1 + age + GENDER + RACE_G + year +  ACS +
                    CHF_severity + past_CABG + past_MI + past_PCI +
                    HXANGINA + HXCEREB + HXCOPD + HXDIAB + HXHTN +
                    HXHYL  + HXSMOKE + NUMPRMI + DIASBP_R + PULSE_R + SYSBP_R + #+ as.factor(HXMI)
                    + CBRUITS + BMI + S3 + CREATININE_R + HDL_R + LDL_R + TOTCHOL_R + CATHAPPR +
                    DIAGCATH + INTVCATH + CORDOM + GRAFTST + LADST + LCXST + LMST + LVEF_R +
                    NUMDZV + PRXLADST + RCAST, data = data_full_std, dist='weibull')
summary(survmodel)
```

```
##
## Call:
## survreg(formula = Surv(survtime + 0.1, death_adj) ~ 1 + age +
##     GENDER + RACE_G + year + ACS + CHF_severity + past_CABG +
##     past_MI + past_PCI + HXANGINA + HXCEREB + HXCOPD + HXDIAB +
##     HXHTN + HXHYL + HXSMOKE + NUMPRMI + DIASBP_R + PULSE_R +
##     SYSBP_R + +CBRUITS + BMI + S3 + CREATININE_R + HDL_R + LDL_R +
##     TOTCHOL_R + CATHAPPR + DIAGCATH + INTVCATH + CORDOM + GRAFTST +
##     LADST + LCXST + LMST + LVEF_R + NUMDZV + PRXLADST + RCAST,
##     data = data_full_std, dist = "weibull")
##                  Value Std. Error      z       p
## (Intercept)    8.48288    0.08709  97.41 < 2e-16
## age           -1.00787    0.01809 -55.72 < 2e-16
## GENDER1       -0.09331    0.01633  -5.71 1.1e-08
## RACE_G2       -0.07321    0.02059  -3.55 0.00038
## RACE_G3        0.03210    0.03681   0.87 0.38311
## year2          0.18503    0.01794  10.31 < 2e-16
## year3          0.23427    0.02242  10.45 < 2e-16
## ACS1          -0.13025    0.02672  -4.87 1.1e-06
## ACS2          -0.26369    0.03078  -8.57 < 2e-16
## ACS3          -0.14084    0.04230  -3.33 0.00087
## ACS4          -0.02727    0.01789  -1.52 0.12754
## CHF_severity1 -0.19137    0.02308  -8.29 < 2e-16
## CHF_severity2 -0.36211    0.02296 -15.77 < 2e-16
## past_CABG1    -0.03536    0.02113  -1.67 0.09421
## past_CABG2     0.19988    0.03454   5.79 7.2e-09
## past_MI1       0.06505    0.02572   2.53 0.01142
## past_MI2       0.08605    0.02834   3.04 0.00240
## past_PCI1      0.11813    0.02439   4.84 1.3e-06
## past_PCI2      0.08598    0.02278   3.77 0.00016
## HXANGINA1      0.10348    0.02131   4.86 1.2e-06
## HXCEREB1      -0.25241    0.01960 -12.88 < 2e-16
## HXCOPD1       -0.37974    0.02501 -15.18 < 2e-16
## HXDIAB1       -0.30268    0.01561 -19.39 < 2e-16
## HXHTN1        -0.11000    0.01562  -7.04 1.9e-12
## HXHYL1         0.16354    0.01480  11.05 < 2e-16
## HXSMOKE1      -0.14697    0.01547  -9.50 < 2e-16
```

9

```
## NUMPRMI       -0.05123     0.01264   -4.05 5.0e-05
## DIASBP_R       0.14570     0.01922    7.58 3.4e-14
## PULSE_R       -0.21321     0.01313  -16.24 < 2e-16
## SYSBP_R       -0.02078     0.01975   -1.05 0.29261
## CBRUITS1      -0.19117     0.02080   -9.19 < 2e-16
## BMI            0.09013     0.01732    5.20 1.9e-07
## S31           -0.19567     0.03037   -6.44 1.2e-10
## CREATININE_R  -0.35773     0.00367  -97.59 < 2e-16
## HDL_R         -0.01432     0.01869   -0.77 0.44339
## LDL_R         -0.08762     0.03185   -2.75 0.00594
## TOTCHOL_R      0.13599     0.03267    4.16 3.1e-05
## CATHAPPR1     -0.33463     0.08238   -4.06 4.9e-05
## CATHAPPR2      0.10283     0.05540    1.86 0.06344
## CATHAPPR3     -0.14693     0.05821   -2.52 0.01160
## DIAGCATH1      0.27445     0.04645    5.91 3.4e-09
## INTVCATH1      0.13496     0.01855    7.28 3.4e-13
## CORDOM         0.02786     0.02362    1.18 0.23818
## GRAFTST       -0.03038     0.01387   -2.19 0.02848
## LADST         -0.02399     0.01977   -1.21 0.22509
## LCXST         -0.14455     0.01963   -7.36 1.8e-13
## LMST          -0.08430     0.01359   -6.20 5.6e-10
## LVEF_R         0.46592     0.01593   29.25 < 2e-16
## NUMDZV         0.01378     0.01447    0.95 0.34075
## PRXLADST      -0.04984     0.01479   -3.37 0.00075
## RCAST         -0.10188     0.01986   -5.13 2.9e-07
## Log(scale)     0.06274     0.00519   12.10 < 2e-16
##
## Scale= 1.06
##
## Weibull distribution
## Loglik(model)= -229549.9   Loglik(intercept only)= -236038.4
##   Chisq= 12977.04 on 50 degrees of freedom, p= 0
## Number of Newton-Raphson Iterations: 23
## n= 83320
```

```r
xibeta = survmodel$linear.predictors


coeffs = survmodel$coefficients
lambda = exp(coeffs[1])
gamma = 1/survmodel$scale


low = confint(survmodel, level=.95)[-1,1]
high = confint(survmodel, level=.95)[-1,2]

df_results = data.frame(value = coeffs[-1], Q2.5= low, Q97.5 = high)
df_results$variable = rownames(df_results)

df_results$variable <- factor(df_results$variable,
                       levels = df_results$variable[order(df_results$value)])
df_signif = df_results[df_results$Q2.5 > 0 | df_results$Q97.5 < 0,]
p0 = ggplot(df_signif, aes(y =variable, x = exp(value))) +
  geom_errorbar(aes(xmax = exp(Q97.5), xmin =exp(Q2.5)),width=0.2) +
```

```
geom_point(position = position_dodge(0.9)) +
geom_vline(xintercept=1, color = "red", linetype = "dashed")+
xlab(expression("decreased survival"  %<->% "increased survival")) +
ylab("Variable") +
labs(title = "95% CI for change in survival time, AFT model")+
theme_bw()+
theme( legend.position = "none", axis.title = element_text(size = 20),
       plot.title = element_text(size = 22))


pdf("Figure/aft_fit.pdf", height = 7, width = 6)
p0
dev.off()
```

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

```
plot(p0)
```


Cox model

```
coxmodel <- coxph(Surv(survtime + 0.1, death_adj) ~ age + GENDER + RACE_G + year +  ACS +
                CHF_severity + past_CABG + past_MI + past_PCI +
                HXANGINA + HXCEREB + HXCOPD + HXDIAB + HXHTN +
                HXHYL  + HXSMOKE + NUMPRMI + DIASBP_R + PULSE_R + SYSBP_R + #+ as.factor(HXMI)
                + CBRUITS + BMI + S3 + CREATININE_R + HDL_R + LDL_R + TOTCHOL_R + CATHAPPR +
```

```
                    DIAGCATH + INTVCATH + CORDOM + GRAFTST + LADST + LCXST + LMST + LVEF_R +
                    NUMDZV + PRXLADST + RCAST, data = data_full_std)
summary(coxmodel)
```

```
## Call:
## coxph(formula = Surv(survtime + 0.1, death_adj) ~ age + GENDER +
##     RACE_G + year + ACS + CHF_severity + past_CABG + past_MI +
##     past_PCI + HXANGINA + HXCEREB + HXCOPD + HXDIAB + HXHTN +
##     HXHYL + HXSMOKE + NUMPRMI + DIASBP_R + PULSE_R + SYSBP_R +
##     +CBRUITS + BMI + S3 + CREATININE_R + HDL_R + LDL_R + TOTCHOL_R +
##     CATHAPPR + DIAGCATH + INTVCATH + CORDOM + GRAFTST + LADST +
##     LCXST + LMST + LVEF_R + NUMDZV + PRXLADST + RCAST, data = data_full_std)
##
##   n= 83320, number of events= 24522
##
##                      coef exp(coef)  se(coef)       z Pr(>|z|)
## age              1.096892  2.994845  0.017169  63.886  < 2e-16 ***
## GENDER1          0.077373  1.080445  0.015305   5.055 4.30e-07 ***
## RACE_G2          0.139872  1.150126  0.019343   7.231 4.79e-13 ***
## RACE_G3         -0.007713  0.992317  0.034574  -0.223 0.823468
## year2           -0.071595  0.930908  0.017162  -4.172 3.02e-05 ***
## year3            0.052333  1.053727  0.021591   2.424 0.015359 *
## ACS1             0.098952  1.104013  0.025037   3.952 7.74e-05 ***
## ACS2             0.253526  1.288561  0.028931   8.763  < 2e-16 ***
## ACS3             0.125822  1.134080  0.039715   3.168 0.001534 **
## ACS4             0.035797  1.036445  0.016848   2.125 0.033614 *
## CHF_severity1    0.227922  1.255987  0.021812  10.449  < 2e-16 ***
## CHF_severity2    0.388590  1.474899  0.021769  17.851  < 2e-16 ***
## past_CABG1       0.060396  1.062257  0.019952   3.027 0.002469 **
## past_CABG2      -0.166530  0.846598  0.032419  -5.137 2.79e-07 ***
## past_MI1        -0.056268  0.945286  0.024161  -2.329 0.019863 *
## past_MI2        -0.078523  0.924481  0.026514  -2.962 0.003061 **
## past_PCI1       -0.115848  0.890611  0.023026  -5.031 4.87e-07 ***
## past_PCI2       -0.106098  0.899336  0.021425  -4.952 7.34e-07 ***
## HXANGINA1       -0.112527  0.893573  0.020026  -5.619 1.92e-08 ***
## HXCEREB1         0.281584  1.325228  0.018486  15.232  < 2e-16 ***
## HXCOPD1          0.408299  1.504257  0.023535  17.349  < 2e-16 ***
## HXDIAB1          0.361874  1.436018  0.014728  24.571  < 2e-16 ***
## HXHTN1           0.145249  1.156328  0.014717   9.870  < 2e-16 ***
## HXHYL1          -0.182457  0.833221  0.013889 -13.137  < 2e-16 ***
## HXSMOKE1         0.160805  1.174455  0.014526  11.070  < 2e-16 ***
## NUMPRMI          0.058834  1.060599  0.011803   4.985 6.21e-07 ***
## DIASBP_R        -0.164211  0.848563  0.018144  -9.050  < 2e-16 ***
## PULSE_R          0.227176  1.255051  0.012126  18.734  < 2e-16 ***
## SYSBP_R          0.061731  1.063677  0.018635   3.313 0.000924 ***
## CBRUITS1         0.213756  1.238320  0.019595  10.909  < 2e-16 ***
## BMI             -0.088891  0.914945  0.016399  -5.421 5.94e-08 ***
## S31              0.198361  1.219403  0.028585   6.939 3.94e-12 ***
## CREATININE_R     0.075954  1.078913  0.003822  19.875  < 2e-16 ***
## HDL_R            0.029299  1.029733  0.012360   2.370 0.017766 *
## LDL_R            0.039167  1.039944  0.029863   1.312 0.189663
## TOTCHOL_R       -0.070039  0.932357  0.030396  -2.304 0.021212 *
## CATHAPPR1        0.256863  1.292868  0.077480   3.315 0.000916 ***
## CATHAPPR2       -0.123424  0.883889  0.052198  -2.365 0.018054 *
```
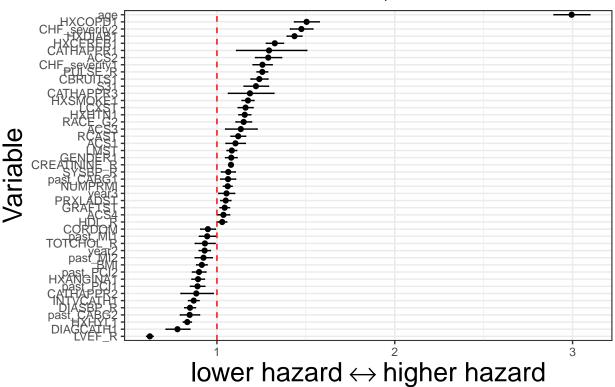
```
## CATHAPPR3      0.170244  1.185594  0.054838    3.104 0.001906 **
## DIAGCATH1     -0.250920  0.778085  0.043842   -5.723 1.04e-08 ***
## INTVCATH1     -0.139929  0.869420  0.017464   -8.012 1.12e-15 ***
## CORDOM        -0.052291  0.949052  0.022220   -2.353 0.018603 *
## GRAFTST        0.042665  1.043588  0.013039    3.272 0.001067 **
## LADST          0.019693  1.019888  0.018515    1.064 0.287510
## LCXST          0.149164  1.160863  0.018375    8.118 4.75e-16 ***
## LMST           0.080160  1.083461  0.012751    6.287 3.25e-10 ***
## LVEF_R        -0.473852  0.622599  0.015043  -31.500  < 2e-16 ***
## NUMDZV        -0.008111  0.991922  0.013625   -0.595 0.551655
## PRXLADST       0.048241  1.049423  0.013871    3.478 0.000505 ***
## RCAST          0.112958  1.119585  0.018688    6.045 1.50e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##               exp(coef) exp(-coef) lower .95 upper .95
## age              2.9948     0.3339    2.8957    3.0973
## GENDER1          1.0804     0.9255    1.0485    1.1133
## RACE_G2          1.1501     0.8695    1.1073    1.1946
## RACE_G3          0.9923     1.0077    0.9273    1.0619
## year2            0.9309     1.0742    0.9001    0.9628
## year3            1.0537     0.9490    1.0101    1.0993
## ACS1             1.1040     0.9058    1.0511    1.1595
## ACS2             1.2886     0.7761    1.2175    1.3637
## ACS3             1.1341     0.8818    1.0492    1.2259
## ACS4             1.0364     0.9648    1.0028    1.0712
## CHF_severity1    1.2560     0.7962    1.2034    1.3108
## CHF_severity2    1.4749     0.6780    1.4133    1.5392
## past_CABG1       1.0623     0.9414    1.0215    1.1046
## past_CABG2       0.8466     1.1812    0.7945    0.9021
## past_MI1         0.9453     1.0579    0.9016    0.9911
## past_MI2         0.9245     1.0817    0.8777    0.9738
## past_PCI1        0.8906     1.1228    0.8513    0.9317
## past_PCI2        0.8993     1.1119    0.8624    0.9379
## HXANGINA1        0.8936     1.1191    0.8592    0.9293
## HXCEREB1         1.3252     0.7546    1.2781    1.3741
## HXCOPD1          1.5043     0.6648    1.4364    1.5753
## HXDIAB1          1.4360     0.6964    1.3952    1.4781
## HXHTN1           1.1563     0.8648    1.1235    1.1902
## HXHYL1           0.8332     1.2002    0.8108    0.8562
## HXSMOKE1         1.1745     0.8515    1.1415    1.2084
## NUMPRMI          1.0606     0.9429    1.0363    1.0854
## DIASBP_R         0.8486     1.1785    0.8189    0.8793
## PULSE_R          1.2551     0.7968    1.2256    1.2852
## SYSBP_R          1.0637     0.9401    1.0255    1.1032
## CBRUITS1         1.2383     0.8075    1.1917    1.2868
## BMI              0.9149     1.0930    0.8860    0.9448
## S31              1.2194     0.8201    1.1530    1.2897
## CREATININE_R     1.0789     0.9269    1.0709    1.0870
## HDL_R            1.0297     0.9711    1.0051    1.0550
## LDL_R            1.0399     0.9616    0.9808    1.1026
## TOTCHOL_R        0.9324     1.0726    0.8784    0.9896
## CATHAPPR1        1.2929     0.7735    1.1107    1.5049
## CATHAPPR2        0.8839     1.1314    0.7979    0.9791
```

```
## CATHAPPR3       1.1856        0.8435        1.0648        1.3201
## DIAGCATH1       0.7781        1.2852        0.7140        0.8479
## INTVCATH1       0.8694        1.1502        0.8402        0.8997
## CORDOM          0.9491        1.0537        0.9086        0.9913
## GRAFTST         1.0436        0.9582        1.0173        1.0706
## LADST           1.0199        0.9805        0.9835        1.0576
## LCXST           1.1609        0.8614        1.1198        1.2034
## LMST            1.0835        0.9230        1.0567        1.1109
## LVEF_R          0.6226        1.6062        0.6045        0.6412
## NUMDZV          0.9919        1.0081        0.9658        1.0188
## PRXLADST        1.0494        0.9529        1.0213        1.0783
## RCAST           1.1196        0.8932        1.0793        1.1614
##
## Concordance= 0.752  (se = 0.002 )
## Likelihood ratio test= 17216  on 50 df,   p=<2e-16
## Wald test           = 17849  on 50 df,   p=<2e-16
## Score (logrank) test = 19318  on 50 df,   p=<2e-16
```

```r
coeffs2 = coxmodel$coefficients


low2 = confint(coxmodel, level=.95)[,1]
high2 = confint(coxmodel, level=.95)[,2]

df_results_cox = data.frame(value = coeffs2, Q2.5= low2, Q97.5 = high2)
df_results_cox$variable = rownames(df_results_cox)

df_results_cox$variable <- factor(df_results_cox$variable,
                        levels = df_results_cox$variable[order(df_results_cox$value)])
df_signif_cox = df_results_cox[df_results_cox$Q2.5 > 0 | df_results_cox$Q97.5 < 0,]

p1 = ggplot(df_signif_cox, aes(y =variable, x = exp(value))) +
  geom_errorbar(aes(xmax = exp(Q97.5), xmin = exp(Q2.5)),width=0.2) +
  geom_point(position = position_dodge(0.9)) +
  geom_vline(xintercept=1, color = "red", linetype = "dashed")+
  xlab(expression("lower hazard"  %<->% "higher hazard")) +
  ylab("Variable") +
  labs(title = "95% CI hazard ratio, Cox model")+
  theme_bw()+
  theme( legend.position = "none", axis.title = element_text(size = 20),
       plot.title = element_text(size = 22))


pdf("Figure/cox_fit.pdf", height = 7, width = 6)
p1
dev.off()
```

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

Table

```r
sigma = survmodel$scale
```

S3:

```r
exp(-coeffs["S31"]/sigma)
```

```
##      S31
## 1.201743
```

```r
exp(coeffs2["S31"])
```

```
##      S31
## 1.219403
```

Capp right

```r
exp(-coeffs["CATHAPPR1"]/sigma)
```

```
## CATHAPPR1
##  1.369274
```

```r
exp(coeffs2["CATHAPPR1"])
```

```
## CATHAPPR1
##  1.292868
```

Capp left

```r
exp(-coeffs["CATHAPPR2"]/sigma)
```

```
## CATHAPPR2
## 0.9079442
```

```r
exp(coeffs2["CATHAPPR2"])
```

```
## CATHAPPR2
## 0.8838892
```

Capp left and right

```r
exp(-coeffs["CATHAPPR3"]/sigma)
```

```
## CATHAPPR3
##  1.147972
```

```r
exp(coeffs2["CATHAPPR3"])
```

```
## CATHAPPR3
##  1.185594
```

Past CABG1

```r
exp(-coeffs["past_CABG1"]/sigma)
```

```
## past_CABG1
##   1.033769
```

```r
exp(coeffs2["past_CABG1"])
```

```
## past_CABG1
##   1.062257
```

Past CABG1

```r
exp(-coeffs["past_CABG2"]/sigma)
```

```
## past_CABG2
##  0.8288396
```

```r
exp(coeffs2["past_CABG2"])
```

```
## past_CABG2
##  0.8465978
```

```
plot(p1)
```


95% CI hazard ratio, Cox model

Diagnostics

```
source("http://myweb.uiowa.edu/pbreheny/7210/f18/notes/fun.R")


sfit <- survfit(coxmodel)
H0 <- -log(sfit$surv)
H <- approxfun(c(0, sfit$time), c(0, H0), method='constant')
e1 <- H(coxmodel$y[,1])*exp(coxmodel$linear.predictors)
e2 <- coxmodel$y[,2]-residuals(coxmodel)
head(e1)
```

```
## [1] 0.059423141 0.298388708 0.339458867 0.004207226 0.002054084 0.003097070
```

```
head(e2)
```

```
##           1           2           3           4           5           6
## 0.059264304 0.298388708 0.339433247 0.004207226 0.002054084 0.003097070
```

```
# Slide 5: Diagnostic plot
efit <- survfit(Surv(e1, coxmodel$y[,2])~1)
lim <- c(0,15)

pdf("Figure/cox_snail_resid.pdf", height = 3.5, width = 6)
plot(efit, fun='cumhaz', mark.time=FALSE, bty='n', conf.int=FALSE, lwd=1, las=1,
```

```
      xlab='Residual', ylab='Cumulative hazard', xlim=lim, ylim=lim)
ciband(efit, fun=function(x) -log(x))
lines(lim, lim, col='red', lwd=1)
dev.off()
```

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

```
plot(efit, fun='cumhaz', mark.time=FALSE, bty='n', conf.int=FALSE, lwd=1, las=1,
      xlab='Residual', ylab='Cumulative hazard', xlim=lim, ylim=lim)
ciband(efit, fun=function(x) -log(x))
lines(lim, lim, col='red', lwd=1)
```


Oout-of sample-performance - Concordance

```
get.concordance = function(pred_test, truth_test, death)
{
  nvalid = length(pred_test)
  agree.count = 0
  pair.count = 0
  for (i in 2:nvalid)
  {
    for (j in 1:(i-1))
    {
      pair.count = pair.count + death[j]
      agree.count = agree.count + death[j]*((pred_test[i] >= pred_test[j]) == (truth_test[i] >= truth_te
    }
  }

  concord = agree.count / pair.count
  return(concord)
}
```

```r
ndata = nrow(data_full_std)
death_ind = which(data_full$death_adj == 1)
nvalid = floor(0.1*nrow(data_full_std))

set.seed(4)
index = sample(1:nrow(data_full_std), size = nvalid, replace = FALSE)

test_data = data_full_std[index,]
train_data = data_full_std[-index,]

# AFT

survtrain_aft = survreg(Surv(survtime + 0.1, death_adj)~1 + age + GENDER + RACE_G + year +  ACS +
                            CHF_severity + past_CABG + past_MI + past_PCI +
                            HXANGINA + HXCEREB + HXCOPD + HXDIAB + HXHTN +
                            HXHYL  + HXSMOKE + NUMPRMI + DIASBP_R + PULSE_R + SYSBP_R + #+ as.factor(HXMI
                            + CBRUITS + BMI + S3 + CREATININE_R + HDL_R + LDL_R + TOTCHOL_R + CATHAPPR +
                            DIAGCATH + INTVCATH + CORDOM + GRAFTST + LADST + LCXST + LMST + LVEF_R +
                            NUMDZV + PRXLADST + RCAST, data = train_data, dist='weibull')

pred_test_aft <- predict(survtrain_aft, newdata=test_data, type='response', se=FALSE)
truth_test = test_data$survtime

coxmodel_train <- coxph(Surv(survtime + 0.1, death_adj) ~ age + GENDER + RACE_G + year +  ACS +
                            CHF_severity + past_CABG + past_MI + past_PCI +
                            HXANGINA + HXCEREB + HXCOPD + HXDIAB + HXHTN +
                            HXHYL  + HXSMOKE + NUMPRMI + DIASBP_R + PULSE_R + SYSBP_R + #+ as.factor(HXMI
                            + CBRUITS + BMI + S3 + CREATININE_R + HDL_R + LDL_R + TOTCHOL_R + CATHAPPR +
                            DIAGCATH + INTVCATH + CORDOM + GRAFTST + LADST + LCXST + LMST + LVEF_R +
                            NUMDZV + PRXLADST + RCAST, data = train_data)

pred_test_cox <- predict(coxmodel_train, newdata=test_data, type='risk', se=FALSE)


survtrain_aft_loglog = survreg(Surv(survtime + 0.1, death_adj)~1 + age + GENDER + RACE_G + year +  ACS
                                CHF_severity + past_CABG + past_MI + past_PCI +
                                HXANGINA + HXCEREB + HXCOPD + HXDIAB + HXHTN +
                                HXHYL  + HXSMOKE + NUMPRMI + DIASBP_R + PULSE_R + SYSBP_R + #+ as.fact
                                + CBRUITS + BMI + S3 + CREATININE_R + HDL_R + LDL_R + TOTCHOL_R + CATH
                                DIAGCATH + INTVCATH + CORDOM + GRAFTST + LADST + LCXST + LMST + LVEF_R
                                NUMDZV + PRXLADST + RCAST, data = train_data, dist='lognormal')


pred_test_aft_loglog <- predict(survtrain_aft_loglog, newdata=test_data, type='response', se=FALSE)


# AFT Concordance
concord.aft = get.concordance(pred_test = pred_test_aft, truth_test = truth_test, death = data_full_std
print(concord.aft)
```

```
##      18179
## 0.5394875
```

```
# Cox Concordance - the prediction is for risk, so we take negative
concord.cox = get.concordance(pred_test = -pred_test_cox, truth_test = truth_test, death = data_full_st
print(concord.cox)
```

```
##      18179
## 0.5383177
```