

A Simple Algorithm Identifies Hypertensive Patients who Benefit from Intensive Blood Pressure Lowering

Shidan Wang

March 25, 2017

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

for SPRINT dataset

```
rm(list = ls())
library(ggplot2)
library(survival)
library(rpart)
library(rpart.plot)
```

Prepare Training & Testing data

Import data

```
baseline <- read.csv("../data/baseline.csv")
outcome <- read.csv("../data/outcomes.csv")
safety <- read.csv("../data/safety.csv")
ID <- baseline$MASKID
m <- length(baseline$MASKID)
```

Construct surv info

```
surv <- Surv(time = outcome[, 3], event = outcome[, 2])
surv.adverse.sae <- Surv(time = safety$SAE_DAYS, event = safety$SAE_EVNT)
```

Include all available variables

```
baseline.catagory <- data.frame(INTENSIVE = as.factor(baseline$INTENSIVE),
                                FRS = as.factor(baseline$INCLUSIONFRS),
```

```

        SBP = baseline$SBP,
        DBP = baseline$DBP,
        NOAGENTS = as.factor(baseline$NOAGENTS == 0),
        SMOKE = as.factor(c("Never", "Former",
"Current", NA)[baseline$SMOKE_3CAT]),
        ASPIRIN = as.factor(baseline$ASPIRIN),
        SUB.CKD = as.factor(baseline$SUB_CKD),
        SCREAT = baseline$SCREAT,
        RACE.BLACK = as.factor(baseline$RACE_BLACK),
        AGE = baseline$AGE,
        FEMALE = as.factor(baseline$FEMALE),
        SUB.CVD = as.factor(baseline$SUB_CVD),
        SUB.ClinicalCVD =
as.factor(baseline$SUB_CLINICALCVD),
        SUB.SubclinicalCVD =
as.factor(baseline$SUB_SUBCLINICALCVD),
        CHR = baseline$CHR,
        GLUR = baseline$GLUR,
        HDL = baseline$HDL,
        TRR = baseline$TRR,
        UMALCR = baseline$UMALCR,
        BMI = baseline$BMI,
        STATIN = as.factor(baseline$STATIN)
)

```

Building training & cross-validation datasets

```

m <- dim(baseline)[1]
all <- 1:m
all.cutoff <- 400
all.bad <- which(surv[all, 1] <= all.cutoff & surv[all, 2] == 1)
all.cutoff <- 1600
all.good <- which(surv[all, 1] > all.cutoff)
length(all.bad)

## [1] 207

length(all.good)

## [1] 466

all.true <- c(rep("high risk", length(all.bad)), rep("low risk",
length(all.good)))
n <- length(all.bad) + length(all.good)
all.select <- c(all.bad, all.good)
training <- 1:n
testing <- which(surv[, 1] > 400 & surv[, 1] <= 1600)

```

Setting weight for training

```

t <- table(all.true[training])
t

```

```
##
## high risk  low risk
##      207      466

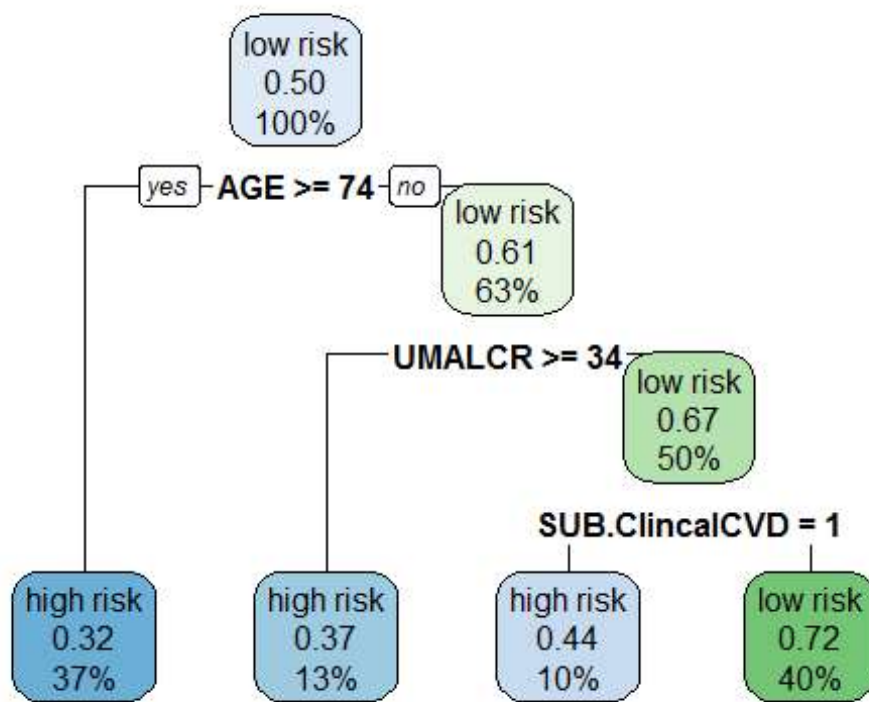
weights <- rep(1, length(training))
weights[which(all.true[training] == "high risk")] <- t[2]/t[1]
```

Training decision tree

```
fit <- rpart(all.true[training] ~ ., data =
baseline.catagory[all.select[training], ],
            control = rpart.control(cp = 0.025), weights = weights)
fit

## n= 673
##
## node), split, n, loss, yval, (yprob)
##      * denotes terminal node
##
## 1) root 673 466.0000 low risk (0.5000000 0.5000000)
##   2) AGE>=73.5 214 110.0000 high risk (0.6803493 0.3196507) *
##   3) AGE< 73.5 459 231.8744 low risk (0.3944285 0.6055715)
##     6) UMALCR>=34.145 78  44.0000 high risk (0.6349792 0.3650208) *
##     7) UMALCR< 34.145 381 155.3333 low risk (0.3323823 0.6676177)
##       14) SUB.ClinicalCVD=1 63  40.0000 high risk (0.5641646 0.4358354) *
##       15) SUB.ClinicalCVD=0 318 103.5556 low risk (0.2757396 0.7242604) *

rpart.plot(fit, tweak = 1.3)
```



Prognostic effect analysis

```
all.predict <- predict(fit, newdata = baseline.catagory[testing, ])
high.risk <- all.predict[, 1] > 0.5
sf <- survfit(surv[testing, ] ~ high.risk)
summary(sf, time = 365*4)
```

```
## Call: survfit(formula = surv[testing, ] ~ high.risk)
```

```
##
```

```
##           high.risk=FALSE
```

```
##      time      n.risk    n.event   survival   std.err
## 1.46e+03  6.26e+02   9.20e+01  9.71e-01  3.33e-03
```

```
## lower 95% CI upper 95% CI
```

```
## 9.65e-01  9.78e-01
```

```
##
```

```
##           high.risk=TRUE
```

```
##      time      n.risk    n.event   survival   std.err
## 1.46e+03  4.12e+02   2.60e+02  9.05e-01  7.23e-03
```

```
## lower 95% CI upper 95% CI
```

```
## 8.91e-01  9.19e-01
```

```
surv.test <- surv[testing, ]
```

```
par(mar = c(10, 10, 3, 1))
```

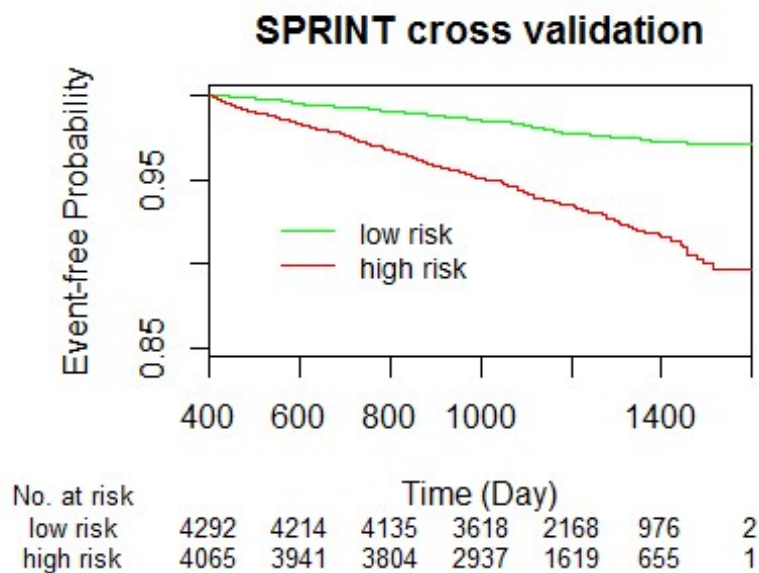
```
sf <- survfit(surv.test ~ high.risk)
```

```
plot(sf, col = c("green", "red"),
     ylim = c(0.85, 1.0), lwd = 1.5, cex = 0.9,
     xlim = c(400, 1600),
```

```

      xlab = "Time (Day)", ylab = "Event-free Probability", main = "SPRINT
cross validation")
legend(x = 500, y = 0.94,
      legend = c("low risk", "high risk"),
      col = c("green", "red"),
      bty = "n", lwd = 1.5, cex = 0.9)
#add patient num
to.write <- matrix(NA, nrow = 2, ncol = 7)
at <- c(400, 600, 800, 1000, 1200, 1400, 1600)
mylegend <- c("low risk", "high risk")
cex <- 0.8
for (i in 1:length(at)){
  temp <- try(summary(sf, times = at[i])$n.risk, silent = T)
  if(class(temp)=="try-error"){
    to.write[,i] <- NA
  }else{
    to.write[,i] <- summary(sf, times = at[i])$n.risk
  }
}
j <- c(1, 2)
for (i in 1:2){
  mtext(to.write[j[i],], side = 1, line = 3+i*0.8, at = at, cex = cex)
  mtext(mylegend[i], side = 1, line = 3+i*0.8, at = 100, cex = cex)
}
mtext("No. at risk", side = 1, line = 3, at = 100, cex = cex)

```



Predictive effect analysis

```
sf <- survfit(surv[testing, ] ~ high.risk + baseline.catagory[testing, 1])
surv.test <- surv[testing, ]
treatment.test <- baseline.catagory[testing, 1]
survdifff(surv.test[which(high.risk == T)] ~ treatment.test[which(high.risk ==
T)])

## Call:
## survdifff(formula = surv.test[which(high.risk == T)] ~
treatment.test[which(high.risk ==
##      T)])
##
##
##              N Observed Expected (O-E)^2/E
## treatment.test[which(high.risk == T)]=0 2052      158      131      5.47
## treatment.test[which(high.risk == T)]=1 2013      105      132      5.45
##              (O-E)^2/V
## treatment.test[which(high.risk == T)]=0      10.9
## treatment.test[which(high.risk == T)]=1      10.9
##
##  Chisq= 10.9  on 1 degrees of freedom, p= 0.000949

survdifff(surv.test[which(high.risk == F)] ~ treatment.test[which(high.risk ==
F)])

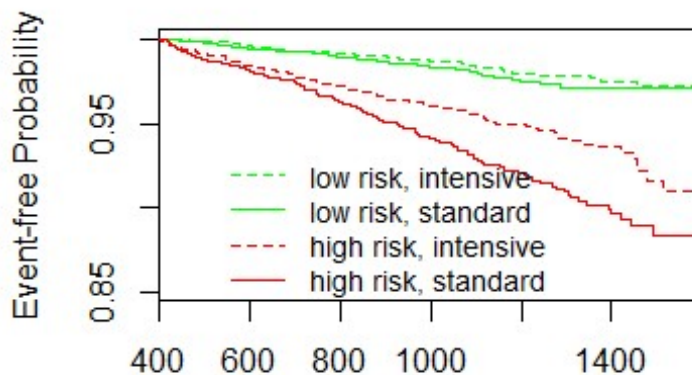
## Call:
## survdifff(formula = surv.test[which(high.risk == F)] ~
treatment.test[which(high.risk ==
##      F)])
##
##
##              N Observed Expected (O-E)^2/E
## treatment.test[which(high.risk == F)]=0 2132       50      45.7      0.414
## treatment.test[which(high.risk == F)]=1 2160       42      46.3      0.408
##              (O-E)^2/V
## treatment.test[which(high.risk == F)]=0      0.821
## treatment.test[which(high.risk == F)]=1      0.821
##
##  Chisq= 0.8  on 1 degrees of freedom, p= 0.365

par(mar = c(10, 10, 3, 1))
sf <- survfit(surv.test ~ high.risk + treatment.test)
plot(sf, col = c("green", "green", "red", "red"),
      ylim = c(0.85, 1.0), lty = c(1, 2, 1, 2), lwd = 1.5, cex = 0.9,
      xlim = c(400, 1600),
      xlab = "Time (Day)", ylab = "Event-free Probability")
legend(x = 500, y = 0.94,
      legend = c("low risk, intensive", "low risk, standard", "high risk,
intensive", "high risk, standard"),
      col = c("green", "green", "red", "red"),
      lty = c(2, 1, 2, 1), bty = "n", lwd = 1.5, cex = 0.9)
#add patient num
```

```

to.write <- matrix(NA, nrow = 4, ncol = 7)
at <- c(400, 600, 800, 1000, 1200, 1400, 1600)
mylegend <- c("low risk, intensive", "low risk, standard", "high risk,
intensive", "high risk, standard")
cex <- 0.8
for (i in 1:length(at)){
  temp <- try(summary(sf, times = at[i])$n.risk, silent = T)
  if(class(temp)=="try-error"){
    to.write[,i] <- NA
  }else{
    to.write[,i] <- summary(sf, times = at[i])$n.risk
  }
}
j <- c(2, 1, 4, 3)
for (i in 1:4){
  mtext(to.write[j[i],], side = 1, line = 3+i*0.8, at = at, cex = cex)
  mtext(mylegend[i], side = 1, line = 3+i*0.8, at = 100, cex = cex)
}
mtext("No. at risk", side = 1, line = 3, at = 100, cex = cex)

```



No. at risk		Time (Day)				
low risk, intensive	2160	2120	2083	1819	1084	500
low risk, standard	2132	2094	2052	1799	1084	476
high risk, intensive	2013	1955	1897	1482	833	348
high risk, standard	2052	1986	1907	1455	786	307

Serious adverse events (SAE) analysis

```

surv.adverse.sae.test <- surv.adverse.sae[testing, ]
treatment.test <- baseline.catagory[testing, 1]
all.predict.test <- predict(fit, newdata = baseline.catagory[testing, ])
high.risk.test <- all.predict.test[, 1] > 0.5

```

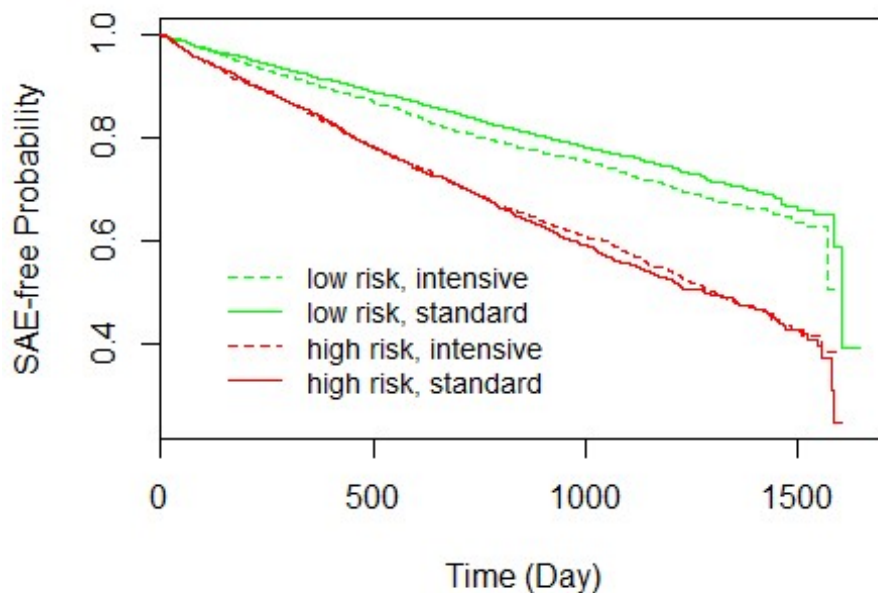
```

sf <- survfit(surv.adverse.sae.test ~ high.risk.test + treatment.test)
summary(sf, time = 365*4)

## Call: survfit(formula = surv.adverse.sae.test ~ high.risk.test +
treatment.test)
##
##
##           high.risk.test=FALSE, treatment.test=0
##      time      n.risk      n.event      survival      std.err
## 1.46e+03 1.81e+02 5.50e+02 6.76e-01 1.45e-02
## lower 95% CI upper 95% CI
## 6.48e-01 7.05e-01
##
##           high.risk.test=FALSE, treatment.test=1
##      time      n.risk      n.event      survival      std.err
## 1.46e+03 1.99e+02 6.28e+02 6.51e-01 1.33e-02
## lower 95% CI upper 95% CI
## 6.26e-01 6.78e-01
##
##           high.risk.test=TRUE, treatment.test=0
##      time      n.risk      n.event      survival      std.err
## 1460.000 100.000 935.000 0.442 0.017
## lower 95% CI upper 95% CI
## 0.410 0.476
##
##           high.risk.test=TRUE, treatment.test=1
##      time      n.risk      n.event      survival      std.err
## 1.46e+03 9.90e+01 9.06e+02 4.43e-01 1.71e-02
## lower 95% CI upper 95% CI
## 4.10e-01 4.77e-01

plot(sf, col = c("green", "green", "red", "red"), lty = c(1, 2, 1, 2), lwd =
1.5, cex = 0.9,
      xlab = "Time (Day)", ylab = "SAE-free Probability", ylim = c(0.25, 1))
legend(x = 100, y = 0.6, legend = c("low risk, intensive", "low risk,
standard", "high risk, intensive", "high risk, standard"),
      col = c("green", "green", "red", "red"), lty = c(2, 1, 2, 1), bty =
"n", lwd = 1.5, cex = 0.9)

```

```
survdif(surv.adverse.sae.test[which(high.risk.test == T)] ~
treatment.test[which(high.risk.test == T)])

## Call:
## survdif(formula = surv.adverse.sae.test[which(high.risk.test ==
##      T)] ~ treatment.test[which(high.risk.test == T)])
##
##
##              N Observed Expected
## treatment.test[which(high.risk.test == T)]=0 2052      943      928
## treatment.test[which(high.risk.test == T)]=1 2013      911      926
##
##              (O-E)^2/E (O-E)^2/V
## treatment.test[which(high.risk.test == T)]=0    0.245      0.49
## treatment.test[which(high.risk.test == T)]=1    0.245      0.49
##
## Chisq= 0.5  on 1 degrees of freedom, p= 0.484

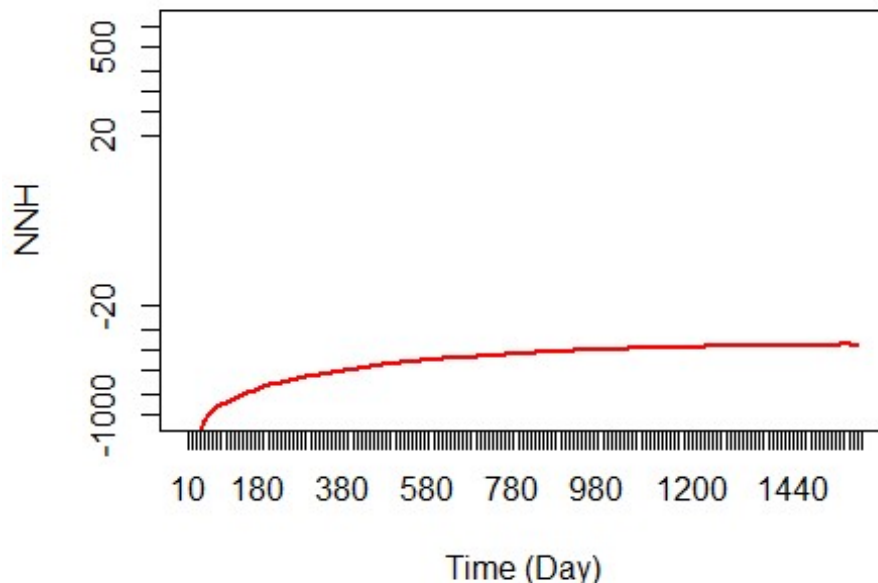
survdif(surv.adverse.sae.test[which(high.risk.test == F)] ~
treatment.test[which(high.risk.test == F)])

## Call:
## survdif(formula = surv.adverse.sae.test[which(high.risk.test ==
##      F)] ~ treatment.test[which(high.risk.test == F)])
##
##
##              N Observed Expected
## treatment.test[which(high.risk.test == F)]=0 2132      556      599
## treatment.test[which(high.risk.test == F)]=1 2160      635      592
##
##              (O-E)^2/E (O-E)^2/V
```

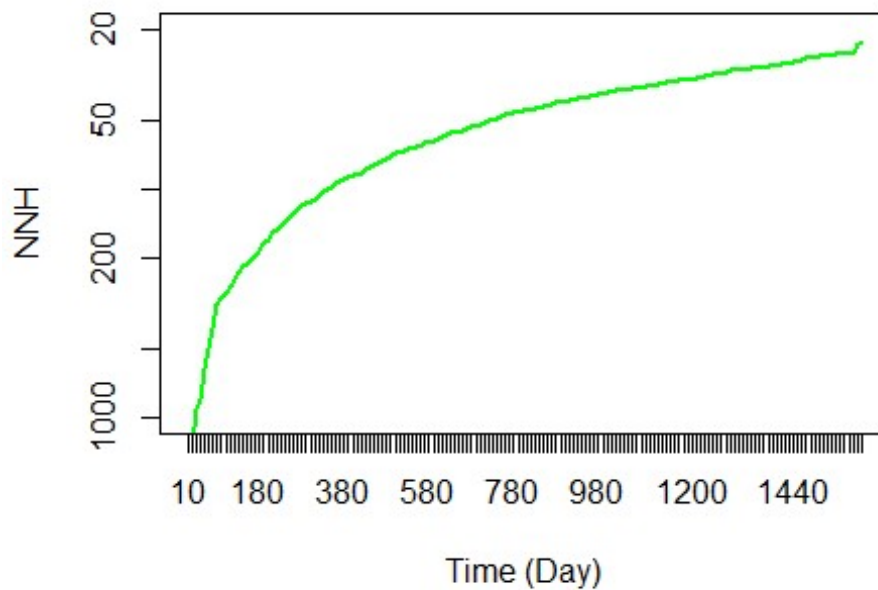
```
## treatment.test[which(high.risk.test == F)]=0      3.09      6.23
## treatment.test[which(high.risk.test == F)]=1      3.13      6.23
##
## Chisq= 6.2 on 1 degrees of freedom, p= 0.0125
```

Plot ARI curve overtime

```
time <- seq(10, 1600, 10)
source("NNH_NNT_plot.R")
NNH.high <- NNH_NNT_plot(surv.adverse.sae.test[which(high.risk == T), ],
group = treatment.test[which(high.risk == T)], NNT = F,
time = time, treatment = "1",
yaxis = T, ylabel = c(20, 50, 100, 200, 500, 1000), ylim =
c(1000, 0.1), lwd = 2,
new = F, col = "red", xlab = "Time (Day)", ylab = "NNH")
## trying to plot NNH with negative sign...
```

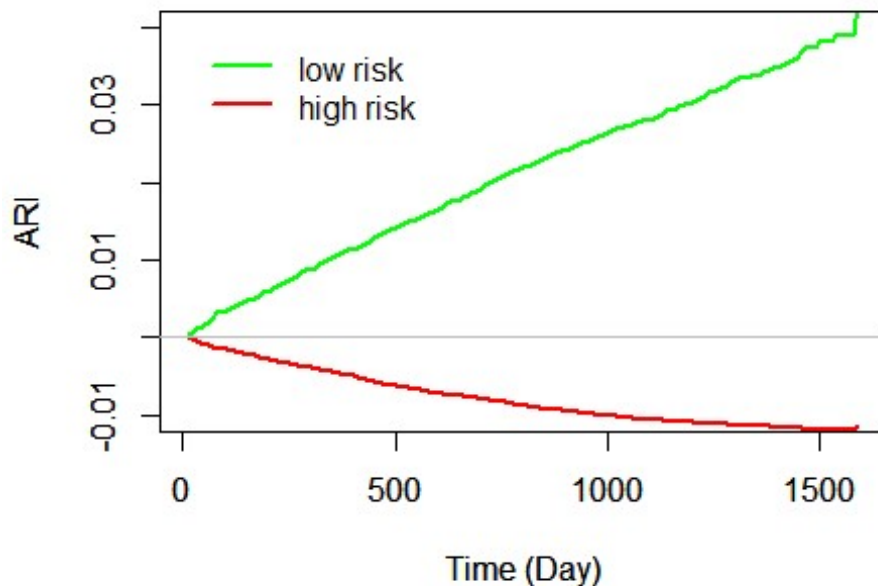


```
NNH.low <- NNH_NNT_plot(surv.adverse.sae.test[which(high.risk == F), ], group
= treatment.test[which(high.risk == F)], NNT = F,
time = time, treatment = "1",
yaxis = T, ylabel = c(20, 50, 100, 200, 500, 1000), ylim =
c(1000, 20), lwd = 2,
col = "green", xlab = "Time (Day)", ylab = "NNH")
## trying to plot NNH...
```



```
ARI_high <- 1/NNH.high[[1]]
ARI_low <- 1/NNH.low[[1]]
plot(time, ARI_high, col = "red", type = "l", xlab = "Time (Day)", ylab =
"ARI", ylim = c(-0.01, 0.04), lwd = 2, main = "ARI in SPRINT dataset")
par(new = T)
plot(time, ARI_low, col = "green", type = "l", xlab = "Time (Day)", ylab =
"ARI", ylim = c(-0.01, 0.04), lwd = 2)
abline(h = 0, col = "grey")
legend(x = 0, y = 0.04, legend = c("low risk", "high risk"), col = c("green",
"red"), lty = 1, bty = "n", lwd = 2)
```

ARI in SPRINT dataset



Plot systolic blood pressure against time, subgrouped by high/low risk

```

ID <- baseline$MASKID
bp <- read.csv("../data/bp.new.csv")
time <- sort(unique(bp$time))
bp$INTENSIVE <- baseline$INTENSIVE[match(bp$MASKID, baseline$MASKID)]
ID.high.risk <- baseline$MASKID[testing][which(high.risk == T)]
ID.low.risk <- baseline$MASKID[testing][which(high.risk == F)]
###among high risk patients
ave.intensive <- c()
ave.standard <- c()
sd.intensive <- c()
sd.standard <- c()
n.intensive <- c()
n.standard <- c()
for(i in 1:length(time)){
  to.select <- which(bp$time == time[i] & bp$MASKID %in% ID.high.risk)
  ave.intensive[i] <- ave(bp$SBP[to.select][which(bp$INTENSIVE[to.select] ==
1)))[1]
  sd.intensive[i] <- sd(bp$SBP[to.select][which(bp$INTENSIVE[to.select] ==
1)])
  ave.standard[i] <- ave(bp$SBP[to.select][which(bp$INTENSIVE[to.select] ==
0)))[1]
  sd.standard[i] <- sd(bp$SBP[to.select][which(bp$INTENSIVE[to.select] ==
0)])
  n.intensive[i] <- length(which(bp$INTENSIVE[to.select] == 1))

```

```

    n.standard[i] <- length(which(bp$INTENSIVE[to.select] == 0))
  }
  row.num <- 4
  mar = c(4+row.num, 8, 4, 2)
  par(mar = mar)
  plot(c(0, time), c(ave(baseline$SBP[which(baseline$INTENSIVE == 1)])[1],
    ave.intensive),
    xlim = range(time), ylim = c(110, 150), type = "l", col = "red", lwd =
  2, lty = 2,
    xlab = "Time (Month)", ylab = "Average SBP (mmHg)", main = "SPRINT cross
validation",
    xaxt = "n")
  axis(1, at = 0:10 * 6, labels = 0:10 * 6)
  par(new = T)
  plot(c(0, time), c(ave(baseline$SBP[which(baseline$INTENSIVE == 0)])[1],
    ave.standard),
    xlim = range(time), ylim = c(110, 150), type = "l", col = "red", lwd =
  2, lty = 1,
    xlab = "", ylab = "", main = "", xaxt = "n")
  arrows(time, ave.intensive + sd.intensive/sqrt(n.intensive), time,
    ave.intensive - sd.intensive/sqrt(n.intensive),
    length = 0.02, angle = 90, code = 3, col = "black", lwd = 2)
  arrows(time, ave.standard + sd.standard/sqrt(n.standard), time, ave.standard
    - sd.standard/sqrt(n.standard),
    length = 0.02, angle = 90, code = 3, col = "black", lwd = 2)
  #add number of medications
  med.intensive <- c()
  med.standard <- c()
  j <- 1
  for (i in (0:9) * 6){
    to.select <- which(bp$time == i & bp$MASKID %in% ID.high.risk)
    med.intensive[j] <-
    round(mean(bp$N_BPCLASSES[to.select][which(bp$INTENSIVE[to.select] == 1)],
    na.rm = T)[1], 1)
    mtext(med.intensive[j], side = 1, line = 3.8, at = i, cex = 1)
    med.standard[j] <-
    round(mean(bp$N_BPCLASSES[to.select][which(bp$INTENSIVE[to.select] == 0)],
    na.rm = T)[1], 1)
    mtext(med.standard[j], side = 1, line = 4.6, at = i, cex = 1)
    j <- j + 1
  }
  mtext("Mean number of medications", side = 1, line = 3.2, at = -6, cex = 1)
  mtext("high risk, intensive", side = 1, line = 3.8, at = -12, cex = 1)
  mtext("high risk, standard", side = 1, line = 4.6, at = -12, cex = 1)
  ###among low risk patients
  ave.intensive <- c()
  ave.standard <- c()
  sd.intensive <- c()
  sd.standard <- c()
  n.intensive <- c()

```

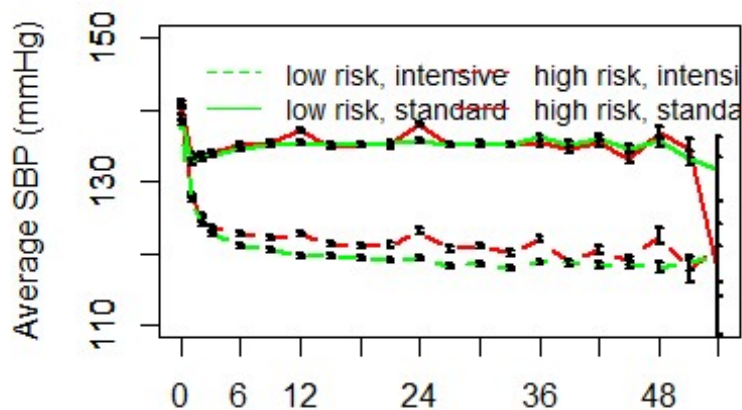
```

n.standard <- c()
for(i in 1:length(time)){
  to.select <- which(bp$time == time[i] & bp$MASKID %in% ID.low.risk)
  ave.intensive[i] <- ave(bp$SBP[to.select][which(bp$INTENSIVE[to.select] ==
1))][1]
  sd.intensive[i] <- sd(bp$SBP[to.select][which(bp$INTENSIVE[to.select] ==
1)])
  ave.standard[i] <- ave(bp$SBP[to.select][which(bp$INTENSIVE[to.select] ==
0))][1]
  sd.standard[i] <- sd(bp$SBP[to.select][which(bp$INTENSIVE[to.select] ==
0)])
  n.intensive[i] <- length(which(bp$INTENSIVE[to.select] == 1))
  n.standard[i] <- length(which(bp$INTENSIVE[to.select] == 0))
}
par(new = T)
plot(c(0, time), c(ave(baseline$SBP[which(baseline$INTENSIVE == 1))][1],
ave.intensive),
      xlim = range(time), ylim = c(110, 150), type = "l", col = "green", lwd =
2, lty = 2,
      xlab = "", ylab = "", main = "", xaxt = "n")
par(new = T)
plot(c(0, time), c(ave(baseline$SBP[which(baseline$INTENSIVE == 0))][1],
ave.standard),
      xlim = range(time), ylim = c(110, 150), type = "l", col = "green", lwd =
2, lty = 1,
      xlab = "", ylab = "", main = "", xaxt = "n")
arrows(time, ave.intensive + sd.intensive/sqrt(n.intensive), time,
ave.intensive - sd.intensive/sqrt(n.intensive),
      length = 0.02, angle = 90, code = 3, col = "black", lwd = 2)
arrows(time, ave.standard + sd.standard/sqrt(n.standard), time, ave.standard
- sd.standard/sqrt(n.standard),
      length = 0.02, angle = 90, code = 3, col = "black", lwd = 2)
#add number of medications
med.intensive <- c()
med.standard <- c()
j <- 1
for (i in (0:9) * 6){
  to.select <- which(bp$time == i & bp$MASKID %in% ID.low.risk)
  med.intensive[j] <-
round(mean(bp$N_BPCLASSES[to.select][which(bp$INTENSIVE[to.select] == 1)],
na.rm = T)[1], 1)
  mtext(med.intensive[j], side = 1, line = 5.4, at = i, cex = 1)
  med.standard[j] <-
round(mean(bp$N_BPCLASSES[to.select][which(bp$INTENSIVE[to.select] == 0)],
na.rm = T)[1], 1)
  mtext(med.standard[j], side = 1, line = 6.2, at = i, cex = 1)
  j <- j + 1
}
mtext("low risk, intensive", side = 1, line = 5.4, at = -12, cex = 1)
mtext("low risk, standard", side = 1, line = 6.2, at = -12, cex = 1)

```

```
#Legends
legend(x = 0, y = 150, legend = c("low risk, intensive", "low risk,
standard"),
      col = c("green", "green"), lty = c(2, 1), bty = "n", lwd = 1.5, cex =
0.9)
legend(x = 25, y = 150, legend = c("high risk, intensive", "high risk,
standard"),
      col = c("red", "red"), lty = c(2, 1), bty = "n", lwd = 1.5, cex = 0.9)
```

SPRINT cross validation



	Time (Month)									
Mean number of medications	0	6	12	18	24	30	36	42	48	54
high risk, intensive	2.8	2.9	2.9	2.9	2.9	2.9	2.9	3	3.7	
high risk, standard	2.1	1.9	1.9	1.9	1.9	1.9	2	2	2	2.5
low risk, intensive	2.1	2.6	2.7	2.7	2.8	2.8	2.7	2.7	2.8	4
low risk, standard	1.8	1.7	1.7	1.6	1.7	1.7	1.7	1.7	1.7	1.7

ACCORD External Validation

```
rm(list = ls())
baseline <- read.csv("../..//ACCORD/data/accord_key.csv", stringsAsFactors =
F) #[10251, 9] #arm: Intensive BP: 1, 3; Standard BP: 2, 4
outcome <- read.csv("../..//ACCORD/data/cvdoutcomes.csv", stringsAsFactors =
F) #[10251, 34] #2: primary event; #4: primary event time
labs <- read.csv("../..//ACCORD/data/otherlabs.csv", stringsAsFactors = F)
#[118662, 11] #uacr: UMACLR
sae <- read.csv("../..//ACCORD/data/sae.csv", stringsAsFactors = F)
```

```
#[253, 2] #1: t <= 18m; #2: t > 18m
meds <- read.csv("../..//ACCORD/data/concomitantmeds.csv")
bp <- read.csv("../..//ACCORD/data/bloodpressure.csv")
#[181991, 5]
```

Construct data

```
ID <- baseline$MaskID[which(baseline$arm %in% 1:2)]
ACCORDdat <- data.frame(ID = ID,
                        GENDER = baseline$female[which(baseline$MaskID %in%
ID)],
                        AGE = baseline$baseline_age[which(baseline$MaskID
%in% ID)],
                        BLACK = baseline$raceclass[which(baseline$MaskID %in%
ID)] == "Black",
                        subclinicalCVD =
baseline$cvd_hx_baseline[which(baseline$MaskID %in% ID)],
                        PrimaryTime = outcome$fuyrs_po[which(outcome$MaskID
%in% ID)],
                        PrimaryEvent = outcome$censor_po[which(outcome$MaskID
%in% ID)],
                        INTENSIVE = baseline$arm[which(baseline$MaskID %in%
ID)] == 1
)
baseline.labs <- labs[which(labs$Visit == "BLR"), ]
baseline.bp <- bp[which(bp$Visit == "BLR"), ]
ACCORDdat <- merge(ACCORDdat, baseline.labs, by.x = "ID", by.y = "MaskID")
ACCORDdat <- merge(ACCORDdat, baseline.bp, by.x = "ID", by.y = "MaskID",
all.x = T)
colnames(ACCORDdat)

## [1] "ID" "GENDER" "AGE" "BLACK"
## [5] "subclinicalCVD" "PrimaryTime" "PrimaryEvent" "INTENSIVE"
## [9] "Visit.x" "fpg" "alt" "cpk"
## [13] "potassium" "screat" "gfr" "ualb"
## [17] "ucreat" "uacr" "Visit.y" "sbp"
## [21] "dbp" "hr"

ACCORDdat <- ACCORDdat[-which(is.na(ACCORDdat$uacr)), ]
surv <- Surv(time = ACCORDdat$PrimaryTime, event = 1 -
ACCORDdat$PrimaryEvent)
```

Test decision tree performance in ACCORD

```
pred <- which(ACCORDdat$AGE >= 73.5 | ACCORDdat$uacr >= 34.145 |
ACCORDdat$subclinicalCVD == 1)
m <- dim(ACCORDdat)[1]
risk <- rep("low risk", m)
risk[pred] <- "high risk"
```


Risk only

```
cox <- coxph(surv ~ risk)
cox

## Call:
## coxph(formula = surv ~ risk)
##
##               coef exp(coef) se(coef)      z      p
## risklow risk -1.037      0.354    0.153 -6.78 1.2e-11
##
## Likelihood ratio test=53.5  on 1 df, p=2.55e-13
## n= 2258, number of events= 238

summary(cox)

## Call:
## coxph(formula = surv ~ risk)
##
##    n= 2258, number of events= 238
##
##               coef exp(coef) se(coef)      z Pr(>|z|)
## risklow risk -1.0371      0.3545    0.1529 -6.784 1.17e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##               exp(coef) exp(-coef) lower .95 upper .95
## risklow risk      0.3545      2.821    0.2627    0.4783
##
## Concordance= 0.614 (se = 0.017 )
## Rsquare= 0.023 (max possible= 0.788 )
## Likelihood ratio test= 53.52  on 1 df,  p=2.554e-13
## Wald test               = 46.02  on 1 df,  p=1.173e-11
## Score (logrank) test = 50.28  on 1 df,  p=1.335e-12

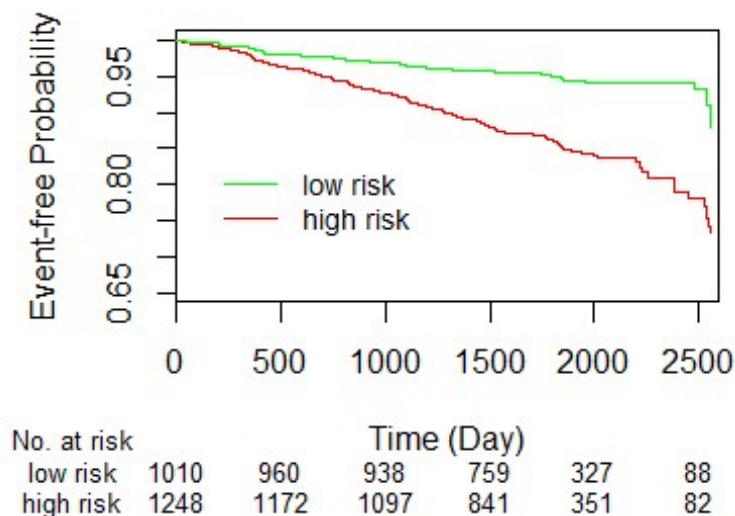
#plot with number at risk
par(mar = c(10, 10, 3, 1))
surv.temp <- surv
surv.temp[, 1] <- surv[, 1]*365
sf <- survfit(surv.temp ~ risk)
plot(sf, col = c("red", "green"),
     ylim = c(0.65, 1.0), lwd = 1.5, cex = 0.9,
     xlim = c(0, 2600),
     xlab = "Time (Day)", ylab = "Event-free Probability")
legend(x = 100, y = 0.85,
      legend = c("low risk", "high risk"),
      col = c("green", "red"),
      bty = "n", lwd = 1.5, cex = 0.9)

#add patient num
to.write <- matrix(NA, nrow = 2, ncol = 6)
at <- c(0, 500, 1000, 1500, 2000, 2500)
```

```

mylegend <- c("low risk", "high risk")
cex <- 0.8
for (i in 1:length(at)){
  temp <- try(summary(sf, times = at[i])$n.risk, silent = T)
  if(class(temp)=="try-error"){
    to.write[,i] <- NA
  }else{
    to.write[,i] <- summary(sf, times = at[i])$n.risk
  }
}
j <- c(2, 1)
for (i in 1:2){
  mtext(to.write[j[i],], side = 1, line = 3+i*0.8, at = at, cex = cex)
  mtext(mylegend[i], side = 1, line = 3+i*0.8, at = -at[2], cex = cex)
}
mtext("No. at risk", side = 1, line = 3, at = -at[2], cex = cex)

```



Risk + Treatment

```

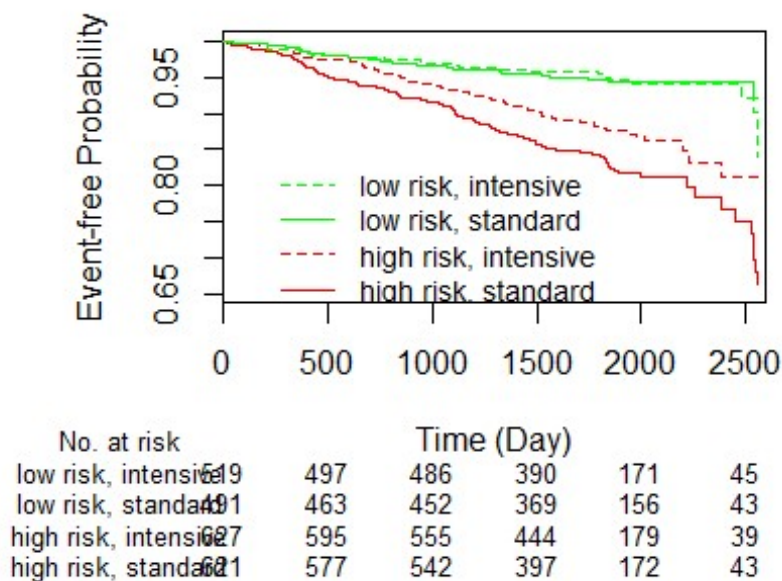
sf <- survfit(surv ~ risk + ACCORDdat$INTENSIVE)
par(mar = c(10, 10, 3, 1))
surv.temp <- surv
surv.temp[, 1] <- surv[, 1]*365
sf <- survfit(surv.temp ~ risk + ACCORDdat$INTENSIVE)
plot(sf, col = c("red", "red", "green", "green"),
      ylim = c(0.65, 1.0), lty = c(1, 2, 1, 2), lwd = 1.5, cex = 0.9,
      xlim = c(0, 2600),

```

```

xlab = "Time (Day)", ylab = "Event-free Probability")
legend(x = 150, y = 0.85,
      legend = c("low risk, intensive", "low risk, standard", "high risk,
intensive", "high risk, standard"),
      col = c("green", "green", "red", "red"),
      lty = c(2, 1, 2, 1), bty = "n", lwd = 1.5, cex = 0.9)
#add patient num
to.write <- matrix(NA, nrow = 4, ncol = 6)
at <- c(0, 500, 1000, 1500, 2000, 2500)
mylegend <- c("low risk, intensive", "low risk, standard", "high risk,
intensive", "high risk, standard")
cex <- 0.8
for (i in 1:length(at)){
  temp <- try(summary(sf, times = at[i])$n.risk, silent = T)
  if(class(temp)=="try-error"){
    to.write[,i] <- NA
  }else{
    to.write[,i] <- summary(sf, times = at[i])$n.risk
  }
}
j <- c(4, 3, 2, 1)
for (i in 1:4){
  mtext(to.write[j[i],], side = 1, line = 3+i*0.8, at = at, cex = cex)
  mtext(mylegend[i], side = 1, line = 3+i*0.8, at = -at[2], cex = cex)
}
mtext("No. at risk", side = 1, line = 3, at = -at[2], cex = cex)

```



```
survdifff(surv[which(risk == "high risk"), ] ~ ACCORDdat$INTENSIVE[which(risk == "high risk")])
```

```
## Call:
```

```
## survdifff(formula = surv[which(risk == "high risk"), ] ~  
ACCORDdat$INTENSIVE[which(risk ==  
## "high risk")])
```

```
##  
##  
## N Observed  
## ACCORDdat$INTENSIVE[which(risk == "high risk")]=FALSE 621 108  
## ACCORDdat$INTENSIVE[which(risk == "high risk")]=TRUE 627 74  
## Expected (O-E)^2/E  
## ACCORDdat$INTENSIVE[which(risk == "high risk")]=FALSE 89.8 3.68  
## ACCORDdat$INTENSIVE[which(risk == "high risk")]=TRUE 92.2 3.59  
## (O-E)^2/V  
## ACCORDdat$INTENSIVE[which(risk == "high risk")]=FALSE 7.28  
## ACCORDdat$INTENSIVE[which(risk == "high risk")]=TRUE 7.28  
##  
## Chisq= 7.3 on 1 degrees of freedom, p= 0.00699
```

```
survdifff(surv[which(risk == "low risk"), ] ~ ACCORDdat$INTENSIVE[which(risk == "low risk")])
```

```
## Call:
```

```
## survdifff(formula = surv[which(risk == "low risk"), ] ~  
ACCORDdat$INTENSIVE[which(risk ==  
## "low risk")])
```

```
##  
## N Observed Expected  
## ACCORDdat$INTENSIVE[which(risk == "low risk")]=FALSE 491 26 27.1  
## ACCORDdat$INTENSIVE[which(risk == "low risk")]=TRUE 519 30 28.9  
## (O-E)^2/E (O-E)^2/V  
## ACCORDdat$INTENSIVE[which(risk == "low risk")]=FALSE 0.0462 0.0898  
## ACCORDdat$INTENSIVE[which(risk == "low risk")]=TRUE 0.0434 0.0898  
##  
## Chisq= 0.1 on 1 degrees of freedom, p= 0.764
```

Serious adverse event (SAE) analysis

```
ACCORDdat$sae <- rep(0, m)  
for (i in 1:dim(sae)[1]){  
  j <- match(sae[i, 1], ACCORDdat$ID)  
  if (!is.na(j)){  
    ACCORDdat$sae[j] <- sae[i, 2]  
  }  
}  
  
t <- table(strata(risk, ACCORDdat$INTENSIVE), ACCORDdat$sae > 0)  
t
```

```
##
##                                     FALSE TRUE
## risk=high risk, ACCORDdat$INTENSIVE=FALSE    606  12
## risk=high risk, ACCORDdat$INTENSIVE=TRUE     601  25
## risk=low risk, ACCORDdat$INTENSIVE=FALSE     485   5
## risk=low risk, ACCORDdat$INTENSIVE=TRUE      508   9

ARI_high <- t[2, 2]/sum(t[2, ]) - t[1, 2]/sum(t[1, ])
ARI_low <- t[4, 2]/sum(t[4, ]) - t[3, 2]/sum(t[3, ])
SE <- function(a, n1, c, n2){
  return (sqrt((a/n1*(1-a/n1)/n1) + c/n2*(1-c/n2)/n2))
}
SE_high <- SE(t[2, 2], sum(t[2, ]), t[1, 2], sum(t[1, ]))
SE_low <- SE(t[4, 2], sum(t[4, ]), t[3, 2], sum(t[3, ]))
z <- (ARI_high-ARI_low)/sqrt(SE_high^2 + SE_low^2)
2 * (1 - pnorm(z))

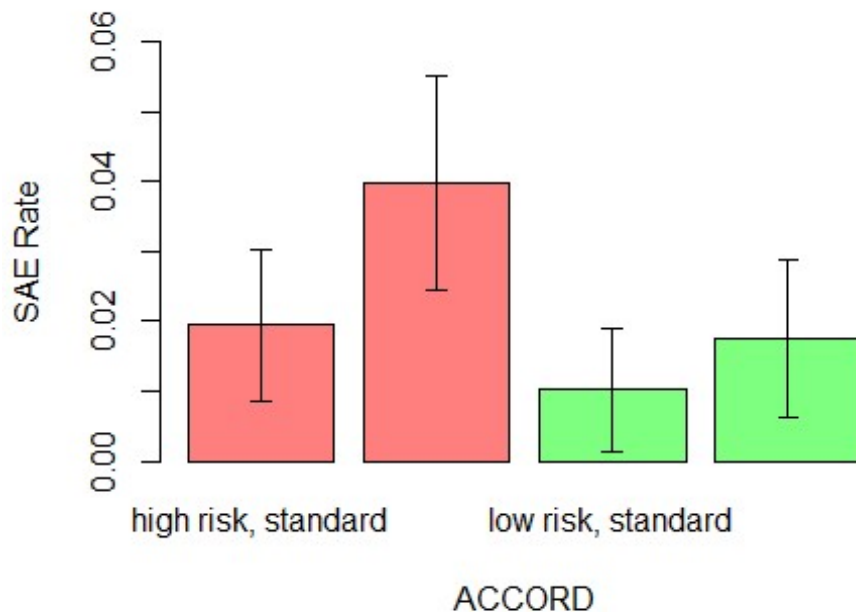
## [1] 0.2700931
```

Plot serious adverse events

```
t <- table(strata(risk, ACCORDdat$INTENSIVE), ACCORDdat$sae > 0)
t

##
##                                     FALSE TRUE
## risk=high risk, ACCORDdat$INTENSIVE=FALSE    606  12
## risk=high risk, ACCORDdat$INTENSIVE=TRUE     601  25
## risk=low risk, ACCORDdat$INTENSIVE=FALSE     485   5
## risk=low risk, ACCORDdat$INTENSIVE=TRUE      508   9

SAE.rate <- t[, 2]/(t[, 1] + t[, 2])
SE <- function(a, n){
  return (sqrt(a/n*(1-a/n)/n))
}
SE.SAE.rate <- c(SE(t[1, 1], t[1, 1] + t[1, 2]),
  SE(t[2, 1], t[2, 1] + t[2, 2]),
  SE(t[3, 1], t[3, 1] + t[3, 2]),
  SE(t[4, 1], t[4, 1] + t[4, 2]))
barcenters <- barplot(height = SAE.rate, ylim = c(0, 0.06),
  names.arg = c("high risk, standard", "high risk,
intensive", "low risk, standard", "low risk, intensive"),
  ylab = "SAE Rate", axes = T,
  col = c(rgb(1, 0, 0, 0.5), rgb(1, 0, 0, 0.5), rgb(0, 1,
0, 0.5), rgb(0, 1, 0, 0.5)), xlab = "ACCORD")
arrows(barccenters,
  SAE.rate - 1.96*SE.SAE.rate,
  barcenters,
  SAE.rate + 1.96*SE.SAE.rate,
  lwd = 1.5, angle = 90, code = 3, length = 0.05, col = "black")
```



Plot systolic blood pressure against time, subgrouped by high/low risk

```
par(mar = c(8, 8, 4, 2))
#Load bp
bp <- read.csv("../..//ACCORD/data/bloodpressure.csv", stringsAsFactors = F)
#[181991, 5]
bp$Visit[which(bp$Visit == "BLR")] <- 0
bp <- bp[-which(bp$Visit == "EXIT" | is.na(bp$Visit)), ]
bp$Visit <- as.numeric(gsub("F", "", bp$Visit))
bp$INTENSIVE <- ACCORDdat$INTENSIVE[match(bp$MaskID, ACCORDdat$ID)]
ID.high.risk <- ACCORDdat[pred, "ID"]
ID.low.risk <- ACCORDdat[-pred, "ID"]
time <- sort(unique(bp$Visit))
time <- c(0:3, (1:21) * 4)
#Load n_med
n_med <- read.csv("../..//ACCORD/data/concomitantmeds.csv", stringsAsFactors = F)
#[60560, 56]
names(n_med)

## [1] "MaskID"          "Visit"           "loop"
## [4] "thiazide"        "ksparing"        "potassium"
## [7] "a2rb"           "acei"            "dhp_ccb"
## [10] "nondhp_ccb"      "alpha_blocker"   "central_agent"
## [13] "beta_blocker"    "vasodilator"     "reserpine"
## [16] "other_bpmed"     "digitalis"        "antiarrhythmic"
## [19] "nitrate"         "other_cvmed"     "sulfonylurea"
```

```

## [22] "biguanide"          "meglitinide"      "ag_inhibitor"
## [25] "nphl_insulin"       "tzd"              "reg_insulin"
## [28] "la_insulin"         "othbol_insulin"   "premix_insulin"
## [31] "other_diabmed"      "bile_sequestrant" "statin"
## [34] "fibrate"           "other_lipidmed"   "cholest_abi"
## [37] "niacin"            "anti_coag"        "anti_inflam"
## [40] "platelet_agi"       "cox2"             "aspirin"
## [43] "thyroid"           "progestin"        "estrogen"
## [46] "oral_asthma"        "anti_depress"     "inhaled_asthma"
## [49] "oral_steroid"       "anti_psych"       "osteoporosis"
## [52] "fluid_retention"   "other_med"        "vitamin"
## [55] "otc"               "herbal"

n_med$Visit[which(n_med$Visit == "BLR")] <- 0
n_med$Sum <- rowSums(n_med[, c(3:5, 7:16)], na.rm = T) #select BP treatment only
n_med <- n_med[-which(n_med$Visit == "EXIT"), ]
n_med$Visit <- as.numeric(gsub("F", "", n_med$Visit))
n_med$INTENSIVE <- ACCORDdat$INTENSIVE[match(n_med$MaskID, ACCORDdat$ID)]
###among high risk patients
ave.intensive <- c()
ave.standard <- c()
sd.intensive <- c()
sd.standard <- c()
n.intensive <- c()
n.standard <- c()
for(i in 1:length(time)){
  to.select <- which(bp$Visit == time[i] & bp$MaskID %in% ID.high.risk)
  ave.intensive[i] <- ave(bp$sbp[to.select][which(bp$INTENSIVE[to.select] == T))][1]
  sd.intensive[i] <- sd(bp$sbp[to.select][which(bp$INTENSIVE[to.select] == T)])
  ave.standard[i] <- ave(bp$sbp[to.select][which(bp$INTENSIVE[to.select] == F))][1]
  sd.standard[i] <- sd(bp$sbp[to.select][which(bp$INTENSIVE[to.select] == F)])
  n.intensive[i] <- length(which(bp$INTENSIVE[to.select] == T))
  n.standard[i] <- length(which(bp$INTENSIVE[to.select] == F))
}
plot(time, ave.intensive,
      xlim = range(time), ylim = c(110, 150), type = "l", col = "red", lwd = 2, lty = 2,
      xlab = "Time (Month)", ylab = "Average SBP (mmHg)", main = "ACCORD external validation",
      xaxt = "n")
axis(1, at = 0:10 * 12, labels = 0:10 * 12)
par(new = T)
plot(time, ave.standard,
      xlim = range(time), ylim = c(110, 150), type = "l", col = "red", lwd = 2, lty = 1,

```

```

      xlab = "", ylab = "", main = "", xaxt = "n")
arrows(time, ave.intensive + sd.intensive/sqrt(n.intensive), time,
ave.intensive - sd.intensive/sqrt(n.intensive),
      length = 0.02, angle = 90, code = 3, col = "black", lwd = 2)
arrows(time, ave.standard + sd.standard/sqrt(n.standard), time, ave.standard
- sd.standard/sqrt(n.standard),
      length = 0.02, angle = 90, code = 3, col = "black", lwd = 2)
#add number of medications
med.intensive <- c()
med.standard <- c()
j <- 1
for (i in (0:7) * 12){
  to.select <- which(n_med$Visit == i & n_med$MaskID %in% ID.high.risk)
  med.intensive[j] <-
round(mean(n_med$Sum[to.select][which(n_med$INTENSIVE[to.select] == T)],
na.rm = T)[1], 1)
  mtext(med.intensive[j], side = 1, line = 3.8, at = i, cex = 1)
  med.standard[j] <-
round(mean(n_med$Sum[to.select][which(n_med$INTENSIVE[to.select] == F)],
na.rm = T)[1], 1)
  mtext(med.standard[j], side = 1, line = 4.6, at = i, cex = 1)
  j <- j + 1
}
mtext("Mean number of medications", side = 1, line = 3.2, at = -12, cex = 1)
mtext("high risk, intensive", side = 1, line = 3.8, at = -24, cex = 1)
mtext("high risk, standard", side = 1, line = 4.6, at = -24, cex = 1)
###among low risk patients
ave.intensive <- c()
ave.standard <- c()
sd.intensive <- c()
sd.standard <- c()
n.intensive <- c()
n.standard <- c()
for(i in 1:length(time)){
  to.select <- which(bp$Visit == time[i] & bp$MaskID %in% ID.low.risk)
  ave.intensive[i] <- ave(bp$sbp[to.select][which(bp$INTENSIVE[to.select] ==
T))][1]
  sd.intensive[i] <- sd(bp$sbp[to.select][which(bp$INTENSIVE[to.select] ==
T))])
  ave.standard[i] <- ave(bp$sbp[to.select][which(bp$INTENSIVE[to.select] ==
F))][1]
  sd.standard[i] <- sd(bp$sbp[to.select][which(bp$INTENSIVE[to.select] ==
F))])
  n.intensive[i] <- length(which(bp$INTENSIVE[to.select] == T))
  n.standard[i] <- length(which(bp$INTENSIVE[to.select] == F))
}
par(new = T)
plot(time, ave.intensive,
      xlim = range(time), ylim = c(110, 150), type = "l", col = "green", lwd =
2, lty = 2,

```

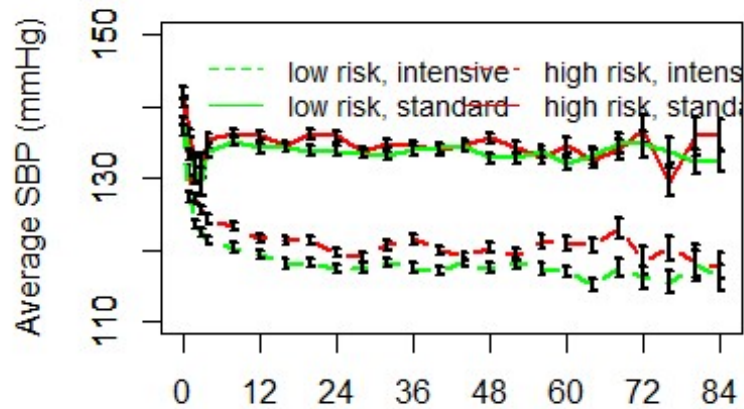


```

      xlab = "", ylab = "", main = "", xaxt = "n")
par(new = T)
plot(time, ave.standard,
      xlim = range(time), ylim = c(110, 150), type = "l", col = "green", lwd =
2, lty = 1,
      xlab = "", ylab = "", main = "", xaxt = "n")
arrows(time, ave.intensive + sd.intensive/sqrt(n.intensive), time,
ave.intensive - sd.intensive/sqrt(n.intensive),
      length = 0.02, angle = 90, code = 3, col = "black", lwd = 2)
arrows(time, ave.standard + sd.standard/sqrt(n.standard), time, ave.standard
- sd.standard/sqrt(n.standard),
      length = 0.02, angle = 90, code = 3, col = "black", lwd = 2)
#add number of medications
med.intensive <- c()
med.standard <- c()
j <- 1
for (i in (0:7) * 12){
  to.select <- which(n_med$Visit == i & n_med$MaskID %in% ID.low.risk)
  med.intensive[j] <-
round(mean(n_med$Sum[to.select][which(n_med$INTENSIVE[to.select] == T)],
na.rm = T)[1], 1)
  mtext(med.intensive[j], side = 1, line = 5.4, at = i, cex = 1)
  med.standard[j] <-
round(mean(n_med$Sum[to.select][which(n_med$INTENSIVE[to.select] == F)],
na.rm = T)[1], 1)
  mtext(med.standard[j], side = 1, line = 6.2, at = i, cex = 1)
  j <- j + 1
}
mtext("low risk, intensive", side = 1, line = 5.4, at = -24, cex = 1)
mtext("low risk, standard", side = 1, line = 6.2, at = -24, cex = 1)
#Legends
legend(x = 0, y = 150, legend = c("low risk, intensive", "low risk,
standard"),
      col = c("green", "green"), lty = c(2, 1), bty = "n", lwd = 1.5, cex =
0.9)
legend(x = 40, y = 150, legend = c("high risk, intensive", "high risk,
standard"),
      col = c("red", "red"), lty = c(2, 1), bty = "n", lwd = 1.5, cex = 0.9)

```

ACCORD external validation



	Time (Month)							
Mean number of medications	0	12	24	36	48	60	72	84
high risk, intensive	1.9	3.1	3.3	3.3	3.5	3.5	3.3	3.1
high risk, standard	2	2.1	2.2	2.3	2.3	2.3	2.4	2.4
low risk, intensive	1.5	2.7	2.8	2.9	2.9	3	3	2.9
low risk, standard	1.4	1.6	1.7	1.7	1.9	1.8	1.8	1.8

```
rm(list = ls())
```

Session info

```
sessionInfo()
```

```
## R version 3.3.2 (2016-10-31)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 14393)
##
## locale:
## [1] LC_COLLATE=English_United States.1252
## [2] LC_CTYPE=English_United States.1252
## [3] LC_MONETARY=English_United States.1252
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United States.1252
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods   base
##
## other attached packages:
## [1] rpart.plot_2.1.0 rpart_4.1-10      survival_2.40-1    ggplot2_2.1.0
##
## loaded via a namespace (and not attached):
## [1] Rcpp_0.12.6      lattice_0.20-34    digest_0.6.10      rprojroot_1.2
## [5] plyr_1.8.4       grid_3.3.2         gtable_0.2.0       backports_1.0.5
```

```
## [9] magrittr_1.5      evaluate_0.10      scales_0.4.0      stringi_1.1.1
## [13] Matrix_1.2-7.1    rmarkdown_1.4      splines_3.3.2      tools_3.3.2
## [17] stringr_1.0.0     munsell_0.4.3      yaml_2.1.14        colorspace_1.2-6
## [21] htmltools_0.3.5   knitr_1.15.1
```