```
rm(list=ls())

library(foreign)
library(survival)
library(qvcalc)
library(boot)
library(ggplot2)
library(Hmisc)
library(ROCR)
library(calibrate)
library(splines)
library(readxl)
library(metafor)
library(dynpred)

fileDir <- "C:\\";

##############
#LOAD DATABASES
data<-read.dta(paste(fileDir,"SHHEC extract derivation.dta",sep=""))
data2<-read.dta(paste(fileDir,"SHHEC extract validate.dta",sep=""))

##############
#FIGURE 1: SBP
data$sbpcat<-as.factor(cut(data$sbp, c(0,120,140,160,180,260), right=FALSE,
labels=c(1:5)))

model<-coxph(Surv(dayscvd, cvdfu)~sbpcat ,data=data,x=T)
flv <- qvcalc(model, "sbpcat")
flvdata<-flv$qvframe[,c(1,3)]
medians<-round(tapply(data$sbp,data$sbpcat,median,na.rm=T),0)

par(xpd=NA, mfrow=c(1,2))

#FLOATING ABSOLUTE RISKS
plot(medians,flvdata[,1],ylim=c(-1,3),pch=15,cex=0.0001,xlab="Systolic blood
pressure, mm Hg",
                yaxt="n",ylab="Hazard ratio",bty="n",xaxt="n",xlim=c(100,200))
par(xpd=F)
axis(2,at=log(c(0,0.5,1,2,4,8,16)),labels=c(0,0.5,1,2,4,8,16),las=2)
axis(1,at=c(80,100, 120, 140,160,180,200),labels=c(80,100, 120,
140,160,180,200))
symbols(medians,flvdata[,1],circles=c(1.2,1.2,1.2,1.2,1.2),inches=F,add=T,bg='bl
ack',fg='black')
segments(medians,flvdata[,1]-1.96*flvdata[,2],medians,flvdata[,1]+1.96*flvdata[,
2])


#SPLINE

model <- coxph(Surv(dayscvd, cvdfu) ~ pspline(sbp, df=4), data=data)
ptemp <- termplot(model, se=TRUE, plot=FALSE)
```

```
sbpterm <- ptemp$sbp
center <- with(sbpterm, y[x==111])
ytemp <- sbpterm$y + outer(sbpterm$se, c(0, -1.96, 1.96), '*')

matplot(sbpterm$x, ytemp - center,
yaxt="n",ylim=c(-1,3),xlim=c(100,200),bty="n",
        type='l', lty=c(1,2,2), col=1,xlab="Systolic blood pressure, mm Hg",
        ylab="Hazard ratio",xaxt="n",lwd=1.2)

u95 <- (sbpterm$y + 1.96*sbpterm$se - center)
l95 <- (sbpterm$y - 1.96*sbpterm$se - center)

polyvecy <- c(u95, rev(l95))
polyvecx <- c(sbpterm$x, rev(sbpterm$x))
polycol <- rgb(.99, 0.2, 0.6, 0.25)
polygon(polyvecx, polyvecy, border=NA, col='grey80')

par(new=TRUE)
matplot(sbpterm$x, ytemp - center,
yaxt="n",ylim=c(-1,3),xlim=c(100,200),bty="n",
        type='l', lty=c(1,2,2), col=1,xlab="",
        ylab="",xaxt="n",lwd=1.2)
axis(2,at=log(c(0,0.5,1,2,4,8,16)),labels=c(0,0.5,1,2,4,8,16),las=2)
axis(1,at=c(80,100, 120, 140,160,180,200),labels=c(80,100, 120,
140,160,180,200))

##############
#FIGURE 2: BMI

data$bmicat<-as.factor(cut(data$bmi, c(15,20,25,30,35,40), right=FALSE,
labels=c(1:5)))
data<- within(data, bmicat <- relevel(bmicat, ref = 2))
model<-coxph(Surv(dayscvd, cvdfu)~bmicat ,data=data,x=T)
flv <- qvcalc(model, "bmicat")
flvdata<-flv$qvframe[,c(1,3)]

medians<-round(tapply(data$bmi,data$bmicat,median,na.rm=T),0)

results<-as.data.frame(matrix(NA,nrow=5,ncol=2))
results[1,]<-c(0,0)
results[2:5,1]<-log(summary(model)$conf.int[,1])
results[2:5,2]<-((log(summary(model)$conf.int[,4])-log(summary(model)$conf.int[,
1]))/1.96 +

(log(summary(model)$conf.int[,1])-log(summary(model)$conf.int[,3]))/1.96) / 2

par(xpd=NA, mfrow=c(1,2))

#FLOATING ABSOLUTE RISKS
plot(c(19,
25,30,35,40),flvdata[,1],ylim=c(-1,1.5),pch=15,cex=0.0001,xlab=expression(paste(
"Body mass index, kg/", m^2, sep = "")),
```

```
              yaxt="n",xaxt="n",ylab="Hazard ratio",bty="n")
par(xpd=F)
axis(2,at=log(c(0,0.5,1,2,4)),labels=c(0,0.5,1,2,4),las=2)
axis(1,at=c(0, 20, 25,30,35,40),labels=c(0, 20, 25,30,35,40))
symbols(medians,flvdata[,1],circles=c(.25,.25,.25,.25,.25),inches=F,add=T,bg='bl
ack',fg='black')
segments(medians,flvdata[,1]-1.96*flvdata[,2],medians,flvdata[,1]+1.96*flvdata[,
2])

#SPLINE
par(xpd=F)

model <- coxph(Surv(dayscvd, cvdfu) ~ pspline(bmi, df=4), data=data)
ptemp <- termplot(model, se=TRUE, plot=FALSE)
bmiterm <- ptemp$bmi # this will be a data frame
center <- with(bmiterm, y[x==22.5])

ytemp <- bmiterm$y + outer(bmiterm$se, c(0, -1.96, 1.96), '*')
matplot(bmiterm$x, ytemp - center,
yaxt="n",xaxt="n",ylim=c(-1,1.5),xlim=c(19,40),bty="n",
        type='l', lty=c(1,2,2), col=1,xlab=expression(paste("Body mass index,
kg/", m^2, sep = "")),
        ylab="Hazard ratio")

u95 <- (bmiterm$y + 1.96*bmiterm$se - center)
l95 <- (bmiterm$y - 1.96*bmiterm$se - center)

polyvecy <- c(u95, rev(l95))
polyvecx <- c(bmiterm$x, rev(bmiterm$x))
polygon(polyvecx, polyvecy, border=NA, col='grey80')

par(new=TRUE)
matplot(bmiterm$x, ytemp - center,
yaxt="n",ylim=c(-1,1.5),xlim=c(19,40),bty="n",
        type='l', lty=c(1,2,2), col=1,xaxt="n",lwd=1.2,ylab='',xlab='')

axis(2,at=log(c(0,0.5,1,2,4)),labels=c(0,0.5,1,2,4),las=2)
axis(1,at=c(0, 20, 25,30,35,40),labels=c(0, 20, 25,30,35,40))


###############
#FIGURE 3 - AIC

data$bmi4<-cut(data$bmi, c(0,20,25,30,100), right=FALSE, labels=c(1:4))

results<-as.data.frame(matrix(NA,nrow=128,ncol=9))

#EXTRACT AIC AND NUMBER OF VARIABLES FOR ALL MODELS (ONLY SHOWN FOR MODEL 1, 2,
3, 7, AND 128)
results[1,1:2]<-extractAIC(coxph(Surv(dayscvd,cvdfu)~ 1, data=data))
results[2,1:2]<-extractAIC(coxph(Surv(dayscvd,cvdfu)~ age, data=data))
results[3,1:2]<-extractAIC(coxph(Surv(dayscvd,cvdfu)~ sbp, data=data))
```

```
results[7,1:2]<-extractAIC(coxph(Surv(dayscvd,cvdfu)~ bmi4, data=data))
results[128,1:2]<-extractAIC(coxph(Surv(dayscvd,cvdfu)~
age+sbp+tc+hdlc+diabetes+smoker+bmi4, data=data))

#EXTRACT VARIABLE NAMES (ONLY SHOWN FOR MODEL 2, 3, 7, AND 128)
results[2,3]<-names(coef(coxph(Surv(dayscvd,cvdfu)~ age, data=data)))
results[3,3]<-names(coef(coxph(Surv(dayscvd,cvdfu)~ sbp, data=data)))
results[3,3]<-names(coef(coxph(Surv(dayscvd,cvdfu)~ as.numeric(bmi4),
data=data)))
results[128,3:9]<-names(coef(coxph(Surv(dayscvd,cvdfu)~
age+sbp+tc+hdlc+diabetes+smoker+as.numeric(bmi4), data=data)))

#TAG MODELS WITH AGE AND/OR BMI
results$colour<-0                               #       NEITHER AGE NOR BMI
results$colour[c(2,.....)]<-1           #       WITH AGE, NOT BMI
results$colour[c(7,.....)]<-2           #       WITH BMI, NOT AGE
results$colour[c(128,...)]<-3           #       WITH AGE AND BMI

par(mar=c(5,5,2,2),cex.lab=1.2,cex.axis=1.1)
par(xpd=NA)
par(xaxs='i',yaxs='i')

plot(results[,1],results[,2],ylab="",
        xlab="Number of prognostic variables",pch=16,

col=c("yellow2","black","red","green3")[unclass(as.factor(results$colour))],yaxt
="n",
        bty="n",ylim=c(4750,5050))
lines(0:7,tapply(results[,2],results[,1],min,na.rm=T))
axis(2,at=c(4750,4800,4850,4900,4950,5000,5050),labels=c(4750,4800,4850,4900,495
0,5000,5050),las=2)
legend(5,5050,legend=c("Neither age nor BMI","BMI, not age","Age, not BMI",
                                 "Age and BMI"),

cex=1,col=c("yellow2","red","black","green3"),pch=16,bty = "n")

text(-1.42,4900,"Akaike information criterion",srt=90,xpd=T, cex=1.2)


##############
#FIGURE 4: ROC
data$cvdfu10<-ifelse(data$dayscvd>=3652.5,0,data$cvdfu)
data2$cvdfu10<-ifelse(data2$dayscvd>=3652.5,0,data2$cvdfu)

model1<-glm(cvdfu10~age , data=data, family=binomial)
model2<-glm(cvdfu10~age+sbp+tc+hdlc+diabetes+smoker , data=data,
family=binomial)

p1<-predict(model1,newdata=data2,type="response")
pred1 <- prediction(p1, data2$cvdfu)
stats1 <- performance(pred1, 'tpr', 'fpr')
```

```
auc1 <- performance(pred1, measure = "auc")
auc1 <- auc1@y.values[[1]]
auc1

p2<-predict(model2,newdata=data2,type="response")
pred2 <- prediction(p2, data2$cvdfu)
stats2 <- performance(pred2, 'tpr', 'fpr')

auc2 <- performance(pred2, measure = "auc")
auc2 <- auc2@y.values[[1]]
auc2

par(xpd=FALSE)
par(xaxs='i',yaxs='i')

plot(stats1@x.values[[1]], stats1@y.values[[1]], type='s', ylab="Sensitivity", 
xlab="1-specificity",
         col='black', lwd=1,xlim=c(0,1),ylim=c(0,1),bty="n",yaxt="n",xaxt="n")
lines(stats2@x.values[[1]], stats2@y.values[[1]], type='s', col=2, lty=1)
axis(2,at=c(0,0.2,0.4,0.6,0.8,1.0),labels=c(0.0,0.2,0.4,0.6,0.8,1.0),las=2,cex=1
.2)
axis(1,at=c(0,0.2,0.4,0.6,0.8,1.0),labels=c(0.0,0.2,0.4,0.6,0.8,1.0),cex=1.2)
segments(0,0,1,1,lwd=1,lty=2,col="gray40")
legend (0,1,legend=c("Best model       0.72 (0.69; 0.76)",
                           "Age only            0.66 (0.62; 
0.70)"),cex=1.2,pt.bg=c("red","black"),pch=c(22,22),bty = "n")

par(xpd=NA)

text(0.35,1.01,"AUC (95% CI)", cex=1.2,font=2)

#######################
#FIGURE 5: C-STATISTIC

data2$bmi4<-cut(data2$bmi, c(0,20,25,30,100), right=FALSE, labels=c(1:4))

results <-as.data.frame(matrix(NA,nrow=7,ncol=4))

results[1,1]<-"Age"
results[2,1]<-"+ Smoking status"
results[3,1]<-"+ SBP"
results[4,1]<-"+ Diabetes"
results[5,1]<-"+ TC"
results[6,1]<-"+ HDL-C"
results[7,1]<-"+ BMI"

results[1,2]<-cindex(coxph(Surv(dayscvd,cvdfu)~ age, data=data),data2)$cindex
results[2,2]<-cindex(coxph(Surv(dayscvd,cvdfu)~ age+smoker, 
data=data),data2)$cindex
results[3,2]<-cindex(coxph(Surv(dayscvd,cvdfu)~ age+smoker+sbp, 
data=data),data2)$cindex
results[4,2]<-cindex(coxph(Surv(dayscvd,cvdfu)~ age+smoker+sbp+diabetes, 
```

```
data=data),data2)$cindex
results[5,2]<-cindex(coxph(Surv(dayscvd,cvdfu)~ age+smoker+sbp+diabetes+tc,
data=data),data2)$cindex
results[6,2]<-cindex(coxph(Surv(dayscvd,cvdfu)~ age+smoker+sbp+diabetes+tc+hdlc,
data=data),data2)$cindex
results[7,2]<-cindex(coxph(Surv(dayscvd,cvdfu)~
age+smoker+sbp+diabetes+tc+hdlc+bmi4, data=data),data2)$cindex

#STANDARD ERRORS ARE OBTAINED THROUGH BOOTSTRAPPING (N=500)
results[1,3]<-0.0155102
results[2,3]<-0.01295918
results[3,3]<-0.01290816
results[4,3]<-0.01316327
results[5,3]<-0.01336735
results[6,3]<-0.01280612
results[7,3]<-0.01331633

results[,4]<-c("0.6488 (0.6184, 0.6792)",
                        "0.6768 (0.6514, 0.7022)",
                        "0.6959 (0.6706, 0.7212)",
                        "0.7017 (0.6759, 0.7275)",
                        "0.7056 (0.6794, 0.7318)",
                        "0.7201 (0.6950, 0.7452)",
                        "0.7206 (0.6945, 0.7467)")

par(font=1)
par(xpd=NA)

forest(x=results[,2], vi=results[,3]^2,annotate=F,slab=results[,1],xlab="",
digits=2, psize=1.5,
        xlim=c(0.49,0.95),at=c(0.6,0.65,0.70,0.75,.8),cex=.8,ylim=c(1,10))
segments(results[6,2],8,results[6,2],.65,lty=2)

text(0.92,7:1,results[,4], adj=1, cex=.8)

par(font=2)
text(0.57, 9, "Variables included", adj=1, cex=.8)
text(0.72, -1, "c-statistic", adj=1, cex=.8)
text(0.91, 9, "c-statistic (95% CI)", adj=1, cex=.8)


######################
#FIGURE 6: CALIBRATION

fit<-coxph(Surv(dayscvd,cvdfu)~ age+sbp+tc+hdlc+smoker+diabetes, data=data)

sfit <- survfit(fit)
index <- max((1:length(sfit$time))[sfit$time < 10*365.25])
surv10 <- sfit$surv[index]

data2$w = fit$coef["age"]*(data2$age - fit$means["age"]) +
                 fit$coef["tc"] *(data2$tc - fit$means["tc"]) +
```

```
                fit$coef["hdlc"]*(data2$hdlc - fit$means["hdlc"]) +
                fit$coef["sbp"] *(data2$sbp - fit$means["sbp"]) +
                fit$coef["diabetes"]*(data2$diabetes -fit$means["diabetes"]) +
                fit$coef["smoker"]*(data2$smoker -fit$means["smoker"])
data2$risk10<- 1-surv10^exp(data2$w)

data2$deciles<-cut_number(data2$risk10, 10, labels=c(1:10))

erisk10<-as.numeric(tapply(data2$risk10,data2$deciles,mean))
orisk10<-as.numeric(table(data2$cvdfu10,data2$deciles)[2,]/table(data2$deciles))

par(xaxs='i',yaxs='i')

plot(orisk10,erisk10,yaxt="n",xaxt="n"
        ,ylab="Expected risk", xlab="Observed
risk",bty="n",ylim=c(0,0.2),xlim=c(0,0.2),pch=16,cex=1.25)
axis(2,at=c(0,0.05,0.1,0.15,0.2),labels=c(0,0.05,0.1,0.15,0.2),las=2)
axis(1,at=c(0,0.05,0.1,0.15,0.2),labels=c(0,0.05,0.1,0.15,0.2))
segments(0,0,0.2,0.2,lwd=1,lty=2,col="gray40")


#########################
#FIGURE 7: INVERSE O-GIVE

Risks<-read_excel("C:\\Risks all.xlsx",1)

risks<-as.numeric(quantile(Risks$Risk, prob = seq(0, 1, length = 101)))
probs<-c(0:100)

plot(risks,probs,xlim=c(0,50))

par(mar=c(5,5,5,7),cex.lab=1.2,cex.axis=1.1)
par(xaxs='i',yaxs='i')

Ecdf(Risks$Risk,what=c('1-F'),yaxt="n",bty="n",ylab="Percentage with this risk
or higher",
        xlab="10-year risk",subtitles=F,xlim=c(0,50),lwd=2,col="black")
axis(2,at=c(0,0.2,0.4,0.6,0.8,1.0),labels=c(0,20,40,60,80,100),las=2)
axis(4,at=c(0,200/1560,400/1560,600/1560,800/1560,1000/1560,1200/1560,1400/1560,
1),labels=c(0,200,400,600,800,1000,1200,1400,1560),las=2)
text(57,0.5,"Number with this risk or higher (thousands)",srt=-90,xpd=T,
cex=1.2)
segments(10,0,10,0.3,lwd=2.5,lty=2,col="blue")
segments(0,0.3,10,0.3,lwd=2.5,lty=2,col="blue")
segments(50,0.3,10,0.3,lwd=2.5,lty=2,col="blue")

segments(20,0,20,0.11,lwd=2.5,lty=3,col="red")
segments(0,0.11,20,0.11,lwd=2.5,lty=3,col="red")
segments(50,0.11,20,0.11,lwd=2.5,lty=3,col="red")

quantile(data$risk*100, prob = seq(0, 1, length = 100))
```