

```

t821.24sept09.r.txt
# Program for testicular cancer analyses: evaluate performance
# Ewout Steyerberg, version sept 2009
library(Hmisc)
library(Design)
library(foreign)

# Import testis data; available at www.clinicalpredictionmodels.org
t821 <- read.spss('f:/Book/SPSS/t821.sav',use.value.labels=F, to.data.frame=T)
t821$tum <- 1 - t821$nec
t821[1,]
#   patkey hosp study orchyr histr3 ter preafp prehcg   lnldhst   sqpost
# 1     94     1     1     79     2     0     0     1 -0.09553243 5.477226
#   reduc10 nec matter tum
# 1       7     0     1     1

n544 <- t821[t821$study==1,] # development set, n=544, JCO 1995
val <- t821[t821$study==3,] # Indiana validation set. n=273, J Urol 2001; no
LDH values
valother <- t821[t821$study==2 | t821$study==4,] # val 172 JCO 1997; n=105 BrJC
2003; total n=277

describe(n544)
describe(val)
describe(valother)

# Fit a full model in n544; 5 predictors
full <- lrm(tum ~ ter+preafp+prehcg+sqpost+reduc10,
            data=n544,x=T,y=T,linear.predictors=T)
full # c = 0.818

# Brier max
B <- mean((full$y) * (1-plogis(full$linear.predictors))^2 +
          (1-full$y) * plogis(full$linear.predictors)^2)
B
Bmax <- mean(full$y) * (1-mean(full$y))^2 + (1-mean(full$y)) * mean(full$y)^2
Bmax
Bscaled <- 1 - B/Bmax
Bscaled

#####
# Extend the model with LDH #
fullLDH <- update(full, ~ . + lnldhst)
fullLDH # c from 0.818 to 0.839

# Brier max
B <- mean((fullLDH$y) * (1-plogis(fullLDH$linear.predictors))^2 +
          (1-fullLDH$y) * plogis(fullLDH$linear.predictors)^2)
B
Bmax <- mean(fullLDH$y) * (1-mean(fullLDH$y))^2 + (1-mean(fullLDH$y)) *
mean(fullLDH$y)^2
Bmax
Bscaled <- 1 - B/Bmax
Bscaled

#####
# Validate the model without LDH #
lp.val <- predict(object=full, newdata = val)

# Brier max
B <- mean((val$TUM) * (1-plogis(lp.val))^2 + (1-val$TUM) * plogis(lp.val)^2)
B
Bmax <- mean(val$TUM) * (1-mean(val$TUM))^2 + (1-mean(val$TUM)) *
mean(val$TUM)^2
Bmax
Bscaled <- 1 - B/Bmax
Bscaled
#####

```

```

t821.24sept09.r.txt
## H-L tests; available at www.clinicalpredictionmodels.org
source('f:/Book/R/HLtest.r')
hl.ext2(p=plogis(full$linear.predictor),y=full$y,g=10,df=8)
hl.ext2(p=plogis(fullLDH$linear.predictors),y=fullLDH$y,g=10,df=8)
hl.ext2(p=plogis(lp.val),y=val$TUM,g=10,df=9)

## CI around c stat
cstatNo <- rcorr.cens(full$linear.predictors, full$y)
cat(cstatNo[1], "[", cstatNo[1]-1.96/2*cstatNo[3], " - ",
cstatNo[1]+1.96/2*cstatNo[3],"]")
cstatYes <- rcorr.cens(fullLDH$linear.predictors, fullLDH$y)
cat(cstatYes[1], "[", cstatYes[1]-1.96/2*cstatYes[3], " - ",
cstatYes[1]+1.96/2*cstatYes[3],"]")
cstatYes - cstatNo
cstatVAL <- rcorr.cens(lp.val, val$TUM)
cat(cstatVAL[1], "[", cstatVAL[1]-1.96/2*cstatVAL[3], " - ",
cstatVAL[1]+1.96/2*cstatVAL[3],"]")

#####
# ROC plots
library(ROCR)
par(mfrow = c(2,1), pty='s', mar=c(4.2,4,4,1),cex=1, las=1, font =1, col=1)

# Development, without LDH
pred.full <- prediction(plogis(full$linear.predictor), full$y)
perf1 <- performance(pred.full,"tpr","fpr") # ROC curve

plot(perf1, xlab="", cex.lab=1.2)
lines(x=c(0,1), y=c(0,1), lwd=2)
title("A. Development data, n=544", adj=0)

# perf1@alpha.values are the cutoffs
# look for 20, 30, 40% points
cutoffs <- c(297,350, 393)
points(x=perf1@x.values[[1]][cutoffs], y=perf1@y.values[[1]][cutoffs], pch=1,
ps=3)
labROC <- perf1@alpha.values[[1]][cutoffs]
labROC <- paste(as.character(round(100*labROC)),"%", sep="")
text(x=perf1@x.values[[1]][cutoffs]+.01,
y=perf1@y.values[[1]][cutoffs]-0.02, labels=labROC, adj=0)
# text(x=1,y=1,labels="0%")

# Development, with LDH
pred.fullLDH <- prediction(plogis(fullLDH$linear.predictor), full$y)
perf2 <- performance(pred.fullLDH,"tpr","fpr") # ROC curve
plot(perf2, add=T, lty=2, lwd=2)
# perf1@alpha.values are the cutoffs
# look for 20, 30, 40% points
cutoffs <- c(331,387,454)
points(x=perf2@x.values[[1]][cutoffs], y=perf2@y.values[[1]][cutoffs], pch=2,
ps=3)

# Validation, without LDH
pred.val <- prediction(plogis(lp.val), val$tum)
perf3 <- performance(pred.val,"tpr","fpr") # ROC curve
plot(perf3, cex.lab=1.2)
lines(x=c(0,1), y=c(0,1), lwd=2)
title("B. Validation data, n=273", adj=0)

# look for 20, 30, 40% points; between 78-79; 84-85; 92-93
x3 <- c(mean(perf3@x.values[[1]][c(78,79)]),
mean(perf3@x.values[[1]][c(84,85)]), mean(perf3@x.values[[1]][c(92,93)]))
y3 <- c(mean(perf3@y.values[[1]][c(78,79)]),
mean(perf3@y.values[[1]][c(84,85)]), mean(perf3@y.values[[1]][c(92,93)]))
points(x=x3, y=y3, pch=1, ps=3)
text(x=x3+.0, y=y3-0.03, labels=labROC, adj=0)

# End ROC plots

```

#####

#####

```
# Validation plots; available at www.clinicalpredictionmodels.org
source('f:/Book/R/val.prob.ci.June09.r')
par(mfrow = c(1,1), pty='s', mar=c(4,3,4,1),cex=1, font =1, col=1 )
val.prob.ci(logit=full$linear.predictor,y=full$y, pl=T,smooth=T,logistic.cal=F,
g=10,
              xlab="", ylab="",riskdist='predicted',
cex.d01=.85,
              d1lab="Tumor", d0lab="Necrosis", dist.label=-0.95, cutoff=.2,
cex.lab=2)
mtext(side=2, text = "Observed masses with tumor", cex=1.1, line=2.5, las=0)
mtext(side=1, text = "Predicted risk without LDH", cex=1.1, line=2.2)
title("A. Development data, n=544", adj=0)
```

Graph B.

```
val.prob.ci(logit=fullLDH$linear.predictor,y=fullLDH$y,
pl=T,smooth=T,logistic.cal=F, g=10,
            xlab="", ylab="",riskdist='predicted',
cex.d01=.85,
            d1lab="Tumor", d0lab="Necrosis", dist.label=-0.95, cutoff=.2,
cex.lab=2)
mtext(side=2, text = "Observed masses with tumor", cex=1.1, line=2.5, las=0)
mtext(side=1, text = "Predicted risk with LDH", cex=1.1, line=2.2)
title("B. Development data, n=544", adj=0)
```

```
val.prob.ci(logit=lp.val,y=val$tum, pl=T,smooth=T,logistic.cal=F, g=10,
            xlab="", ylab="",riskdist='predicted',
cex.d01=.85,
            d1lab="Tumor", d0lab="Necrosis", dist.label=-0.95, cutoff=.2,
cex.lab=2)
mtext(side=2, text = "Observed masses with tumor", cex=1.1, line=2.5, las=0)
mtext(side=1, text = "Predicted risk without LDH", cex=1.1, line=2.2)
title("C. Validation data, n=273", adj=0)
# End validation plots
#####
```

#####

Boxplots

```
par(mfrow = c(3,1), pty='s', mar=c(1.2,4,3,1),cex=0.95, font =1, col=1 )
boxplot(plogis(full$linear.predictors)~full$y,
        ylab="Predicted risk without LDH", xlab="Tumor",ylim=c(0,1),
names=Cs(Benign, Tumor), xaxt='n', cex.lab=1.1)
boxplot(c(mean(plogis(full$linear.predictors[full$y==0])),
mean(plogis(full$linear.predictors[full$y==1]))~ c(0,1),add=T,
        boxlty=0, staplelty=0, medlty=0, medlwd=0, medpch=15, xaxt='n')
mtext(side=3, text = paste("A. Development data, slope=",
round(mean(plogis(full$linear.predictors[full$y==1])) -
mean(plogis(full$linear.predictors[full$y==0])),2),sep=""), adj=0, cex=1.1,
line=1.2)

boxplot(plogis(fullLDH$linear.predictors)~fullLDH$y,
        ylab="Predicted risk with LDH", xlab="Tumor",ylim=c(0,1), xaxt='n',
cex.lab=1.1)
boxplot(c(mean(plogis(fullLDH$linear.predictors[fullLDH$y==0])),
mean(plogis(fullLDH$linear.predictors[fullLDH$y==1]))~ c(0,1),add=T,
        boxlty=0, staplelty=0, medlty=0, medlwd=0, medpch=15, xaxt='n')
mtext(side=3, text = paste("B. Development data, slope=",
round(mean(plogis(fullLDH$linear.predictors[fullLDH$y==1])) -
mean(plogis(fullLDH$linear.predictors[fullLDH$y==0])),2),sep=""), adj=0,
cex=1.1, line=1.2)
```

```
boxplot(plogis(lp.val)~val$tum,
```

```

t821.24sept09.r.txt
ylab="Predicted risk without LDH", xlab="Tumor at validation",ylim=c(0,1),
xaxt='n', cex.lab=1.1)
boxplot(c(mean(plogis(lp.val[val$tum==0])), mean(plogis(lp.val[val$tum==1]))~
c(0,1),add=T,
  boxlty=0, staplelty=0, medlty=0, medlwd=0, medpch=15, xaxt='n')
mtext(side=3, text = paste("C. Validation data, slope=",
round(mean(plogis(lp.val[val$tum==1])) -
      mean(plogis(lp.val[val$tum==0])),2),sep=""), adj=0, cex=1.1,
line=1.2)
mtext(side=1, text = "Benign          Tumor", cex=1.1, line=0)

# IDI is difference between slopes
slopeNo <- mean(plogis(full$linear.predictors[full$y==1])) -
mean(plogis(full$linear.predictors[full$y==0]))
slopeYes <- mean(plogis(fullLDH$linear.predictors[fullLDH$y==1])) -
mean(plogis(fullLDH$linear.predictors[fullLDH$y==0]))
slopeYes - slopeNo # IDI = 0.039

## End Boxplots
#####

#####
## scatterplot with model 1 against 2, markers by outcome
par(mfrow = c(1,1), pty='s', mar=c(4.2,3,4,1),cex=1)
plot(x= plogis(full$linear.predictors), y= plogis(fullLDH$linear.predictors),
  xlab="Predicted risk without LDH", ylab="Predicted risk with LDH",
cex.lab=1.2, cex=1,
  las=1,pty='s',xlim=c(0,1),ylim=c(0,1),
  pch=(full$y+.5)*2) # pch 1 and 3
abline(a=0,b=1, lwd=2)
title("Development data, n=544", adj=0)

#####
# Reclassification table
NoLDH <- cut(plogis(full$linear.predictors),breaks = c(0,.2,1), labels = c("No
LDH, <=20%", "No LDH, >20%"))
YesLDH <- cut(plogis(fullLDH$linear.predictors),breaks = c(0,.2,1), labels =
c("with LDH, <=20%", "with LDH, >20%"))
ValReclas <- cut(plogis(lp.val),breaks = c(0,.2,1), labels = c("val,
<=20%", "val, >20%"))
# Simple reclas
table(NoLDH)
table(YesLDH)
tabReclas <- table(NoLDH, YesLDH)
tabReclas
tabReclas[2]/(tabReclas[2]+tabReclas[4]) # reclassified among high risk
tabReclas[3]/(tabReclas[1]+tabReclas[3]) # reclassified among low risk

# By outcome
table(NoLDH, YesLDH, full$y)
# To get %
tab <- table(NoLDH, YesLDH, full$y )
tab[5:8]/(tab[1:4]+tab[5:8])

table(NoLDH, full$y )
tabNo <- table(NoLDH, full$y )
tabNo[3:4]+tabNo[1:2]
tabNo[3:4]/(tabNo[3:4]+tabNo[1:2])

table(YesLDH, full$y )
tabYes <- table(YesLDH, full$y )
tabYes[3:4]+tabYes[1:2]
tabYes[3:4]/(tabYes[3:4]+tabYes[1:2])

## NRI and IDI calculations with Harrell's functions
improveProb(x1=as.numeric(NoLDH)-1, x2=as.numeric(YesLDH)-1, y=full$y)

# Validation

```

```

table(ValReclass, val$TUM)
# End reclassification table calculations
#####

#####
## Start decision curves; available at www.decisioncurveanalysis.org
source('f:/Book/R/dcaNov08.r')
dcaNo <- dca(yvar=full$y, xmatrix=plogis(full$linear.predictor), prob="Y") #
tumor
dcaLDH <- dca(yvar=fullLDH$y, xmatrix=plogis(fullLDH$linear.predictor),
prob="Y") # tumor
dcaVal <- dca(yvar=val$tum, xmatrix=plogis(lp.val), prob="Y") # tumor

# Net benefit at 20%,30%, 40% thresholds
dcaNo[c(20,30,40),1]-dcaNo[c(20,30,40),2]
dcaLDH[c(20,30,40),1]-dcaLDH[c(20,30,40),2]
dcaVal[c(20,30,40),1]-dcaVal[c(20,30,40),2]

par(mfrow = c(2,1), pty='s', mar=c(4.2,4,4,1),cex=1)
plot(x=dcaNo$threshold, y=dcaNo[,1], type='l', lty=1, lwd=2, las=1,
ylim=c(-.08,.8),
ylab="Net Benefit", xlab="", cex.lab=1.2)
lines(x=dcaNo$threshold, y=dcaNo[,2], lty=3, lwd=2)
lines(x=dcaNo$threshold, y=dcaNo[,3], lty=1, lwd=2)
# LDH
lines(x=dcaLDH$threshold, y=dcaLDH[,1], lty=2, lwd=2)

arrows(20,max(dcaLDH[,2],na.rm=T)-.05,20,0)
text(x=20,y=max(dcaLDH[,2]+.02,na.rm=T), "Threshold")
text(x=80,y=-.05, "Resection in none")
text(x=60,y=.12, "Resection\nin all")
text(x=65,y=.4, "Resection\nif tumor risk\n> threshold")
title("A. Development data, n=544", adj=0)

# Validation
plot(x=dcaVal$threshold, y=dcaVal[,1], type='l', lty=1, lwd=2, las=1,
ylim=c(-.08,.8),
ylab="Net Benefit", xlab="Threshold risk for resection of tumor (%)",
cex.lab=1.2)
lines(x=dcaVal$threshold, y=dcaVal[,2], lty=3, lwd=2)
lines(x=dcaVal$threshold, y=dcaVal[,3], lty=1, lwd=2)

arrows(20,max(dcaVal[,2],na.rm=T)-.05,20,0)
text(x=20,y=max(dcaVal[,2]+.02,na.rm=T), "Threshold")
text(x=80,y=-.05, "Resection in none")
text(x=60,y=.2, "Resection\nin all")
text(x=75,y=.5, "Resection\nif tumor risk\n> threshold")
title("B. Validation data, n=273", adj=0)

## End decision curves ##
#####

```