Breast Cancer: Wisconsin

Jose Tamez

2023-04-26

Contents


```
library(survival)
library(FRESA.CAD)
```

```
## Loading required package: Rcpp
```

```
## Loading required package: stringr
```

```
## Loading required package: miscTools
```

```
## Loading required package: Hmisc
```

```
##
## Attaching package: 'Hmisc'
```

```
## The following objects are masked from 'package:base':
##
##     format.pval, units
```

```
## Loading required package: pROC
```

```
## Type 'citation("pROC")' for a citation.
```

```
##
## Attaching package: 'pROC'
```

```
## The following objects are masked from 'package:stats':
##
##     cov, smooth, var
```

```
source("~/GitHub/FRESA.CAD/R/RRPlot.R")
source("~/GitHub/FRESA.CAD/R/PoissonEventRiskCalibration.R")
op <- par(no.readonly = TRUE)
pander::panderOptions('digits', 3)
#pander::panderOptions('table.split.table', 400)
pander::panderOptions('keep.trailing.zeros',TRUE)
layout(matrix(1:1, nrow=1))
```

```
dataBreast <- read.csv("~/GitHub/RISKPLOTS/DATA/wpbc.data", header=FALSE)
table(dataBreast$V2)
```

```
##
##   N   R
## 151  47
```

```
rownames(dataBreast) <- dataBreast$V1
dataBreast$V1 <- NULL
dataBreast$status <- 1*(dataBreast$V2=="R")
dataBreast$V2 <- NULL
dataBreast$time <- dataBreast$V3
dataBreast$V3 <- NULL
dataBreast <- sapply(dataBreast,as.numeric)
```

```
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
```

```
dataBreast <- as.data.frame(dataBreast[complete.cases(dataBreast),])
table(dataBreast$status)
```

```
##
##   0   1
## 148  46
```

0.1 Modeling

```
ml <- BSWiMS.model(Surv(time,status)~1,data=dataBreast)
```

[+++++]

```
sm <- summary(ml)
pander::pander(sm$coefficients)
```

Table 1: Table continues below

	Estimate	lower	HR	upper	u.Accuracy	r.Accuracy
V26	8.07e-03	1	1.01	1.01	0.593	0.237
V27	4.13e-04	1	1.00	1.00	0.608	0.237
V24	7.71e-03	1	1.01	1.01	0.598	0.634
V7	1.03e-07	1	1.00	1.00	0.588	0.237
V35	8.65e-06	1	1.00	1.00	0.727	0.237
V34	9.13e-03	1	1.01	1.02	0.634	0.598

Table 2: Table continues below

	full.Accuracy	u.AUC	r.AUC	full.AUC	IDI	NRI	z.IDI
V26	0.593	0.598	0.500	0.598	0.0626	0.393	2.77
V27	0.608	0.608	0.500	0.608	0.0563	0.434	2.76
V24	0.603	0.609	0.618	0.613	0.0532	0.323	2.62
V7	0.588	0.595	0.500	0.595	0.0487	0.380	2.30
V35	0.727	0.641	0.500	0.641	0.0289	0.565	2.28
V34	0.603	0.618	0.609	0.613	0.0233	0.411	2.13

	z.NRI	Delta.AUC	Frequency
V26	2.38	0.09827	1
V27	2.63	0.10840	1
V24	1.94	-0.00529	1
V7	2.30	0.09489	1

	z.NRI	Delta.AUC	Frequency
V35	3.50	0.14116	1
V34	2.47	0.00338	1

0.2 Cox Model Performance

Here we evaluate the model using the RRPlot() function.

0.2.1 The evaluation of the raw Cox model with RRPlot()

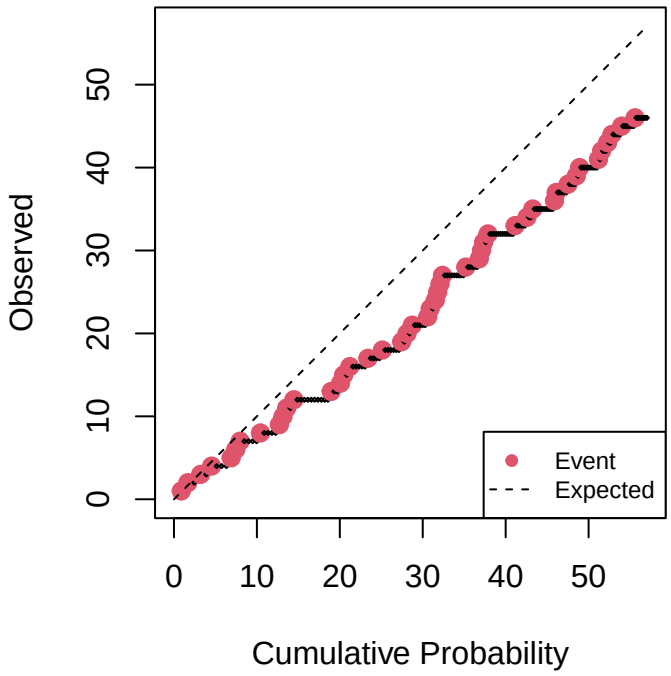
Here we will use the predicted event probability assuming a baseline hazard for events withing 5 years

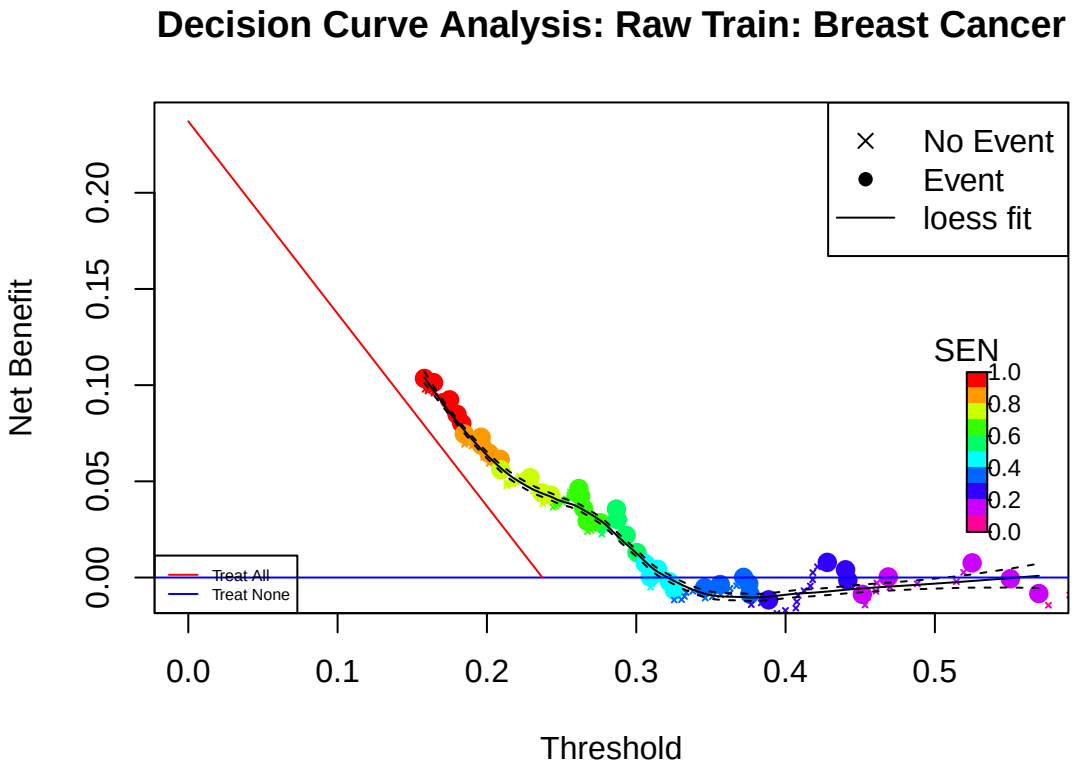
```
index <- predict(ml,dataBreast)
timeinterval <- 2*mean(subset(dataBreast,status==1)$time)

h0 <- sum(dataBreast$status & dataBreast$time <= timeinterval)
h0 <- h0/sum((dataBreast$time > timeinterval) | (dataBreast$status==1))
pander::pander(t(c(h0=h0,timeinterval=timeinterval)),caption="Initial Parameters")
```

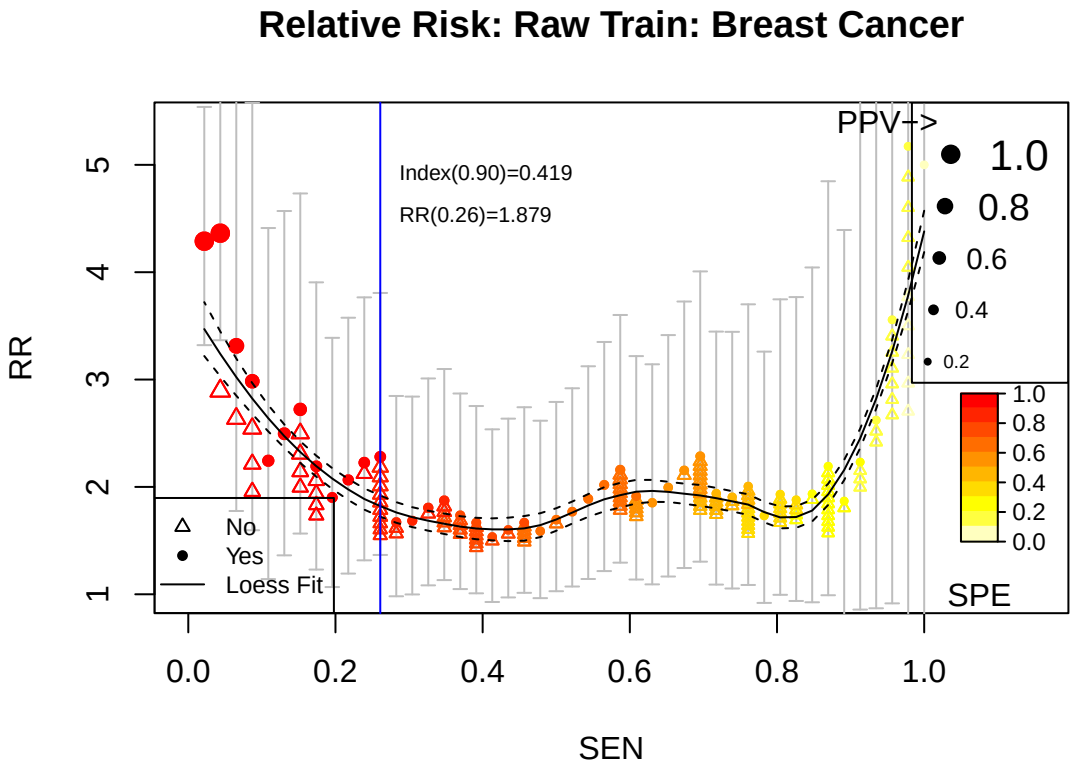
Table 4: Initial Parameters

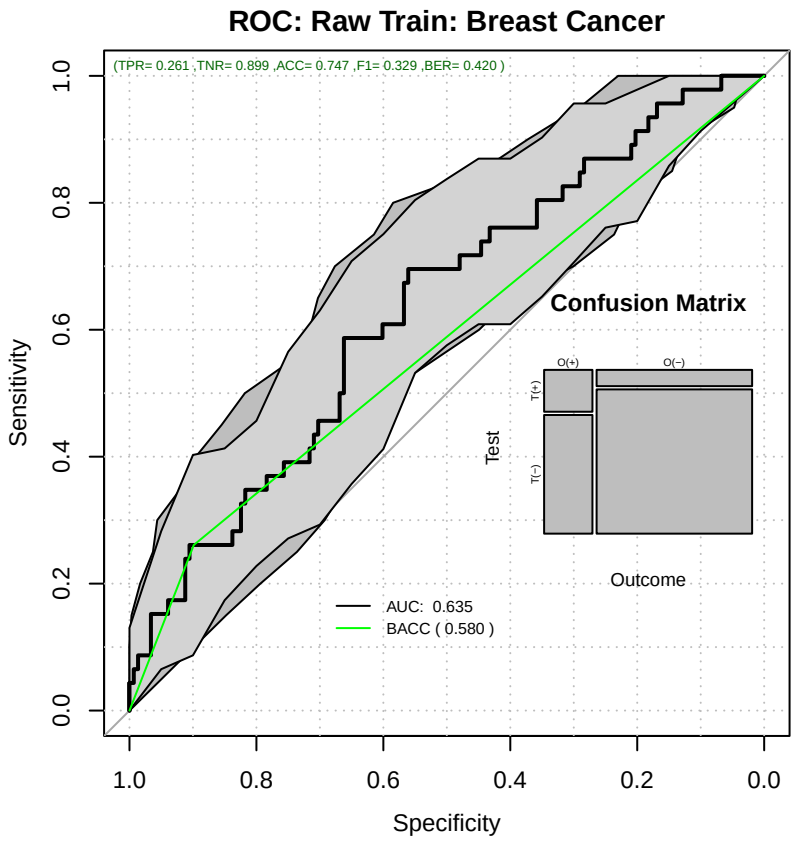
h0	timeinterval
0.323	51.1

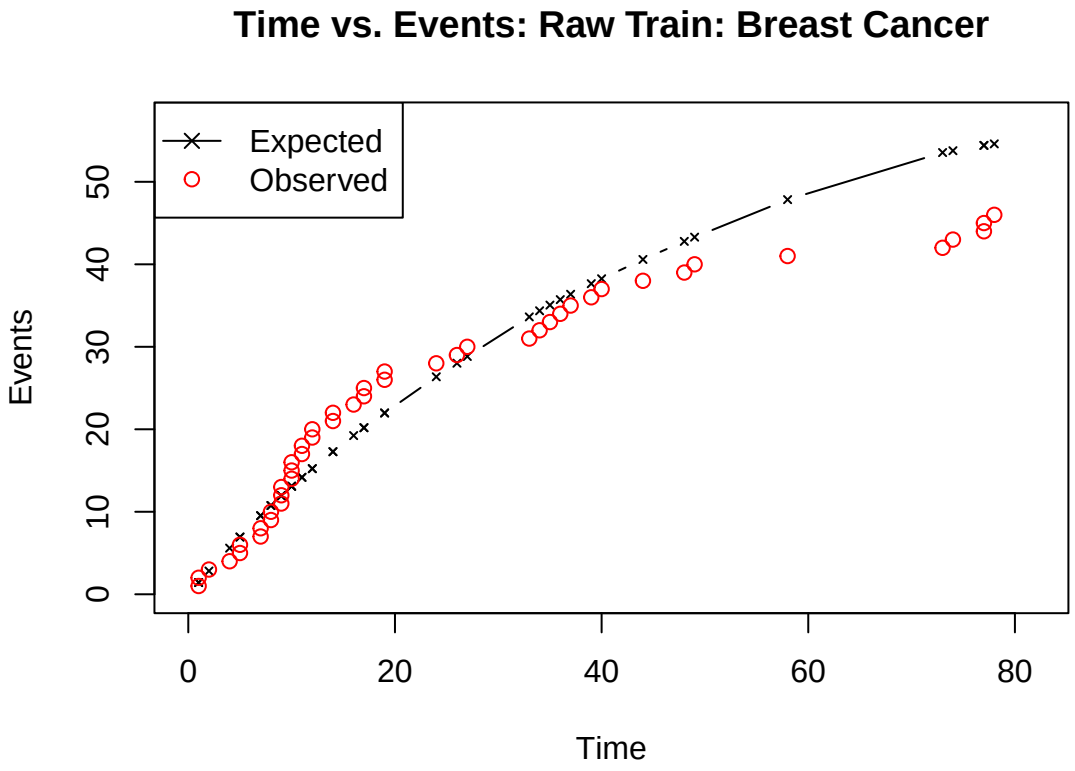
```
rdata <- cbind(dataBreast$status,ppoisGzero(index,h0))
rownames(rdata) <- rownames(dataBreast)

rrAnalysisTrain <- RRPlot(rdata,atProb=c(0.90),
                      timetoEvent=dataBreast$time,
                      title="Raw Train: Breast Cancer",
                      ysurvlim=c(0.00,1.0),
                      riskTimeInterval=timeinterval)
```

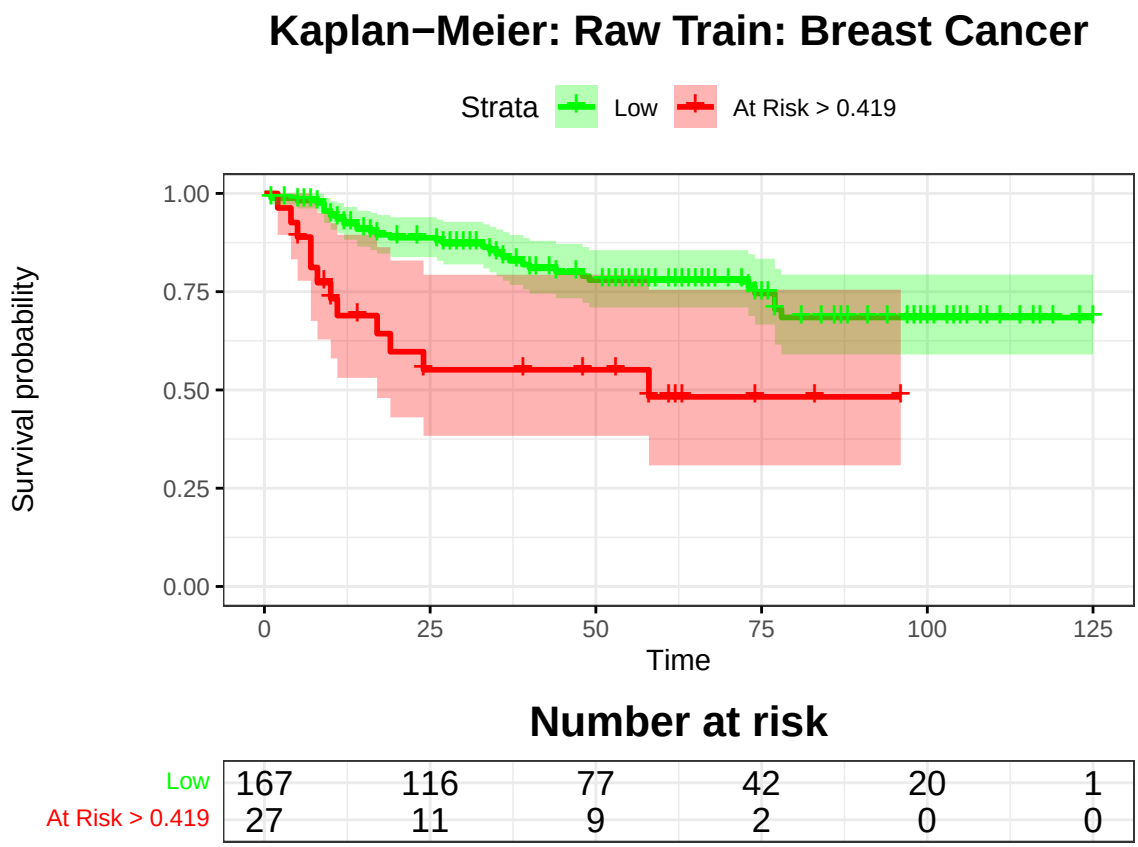

Cumulative vs. Observed: Raw Train: Breast Cancer
Observed
0
10
20
30
40
50
Cumulative Probability
Event
Expected


Decision Curve Analysis: Raw Train: Breast Cancer
No Event
Event
loess fit
SEN
1.0
0.8
0.6
0.4
0.2
0.0
Treat All
Treat None
Net Benefit
0.00
0.05
0.10
0.15
0.20
0.0
0.1
0.2
0.3
0.4
0.5
Threshold


Relative Risk: Raw Train: Breast Cancer
PPV–>
1.0
0.8
0.6
0.4
0.2
Index(0.90)=0.419
RR(0.26)=1.879
No
Yes
Loess Fit
SPE
RR
SEN
0.0
0.2
0.4
0.6
0.8
1.0
1
2
3
4
5


ROC: Raw Train: Breast Cancer
(TPR= 0.261 ,TNR= 0.899 ,ACC= 0.747 ,F1= 0.329 ,BER= 0.420)
Sensitivity
1.0
0.8
0.6
0.4
0.2
0.0
Specificity
1.0
0.8
0.6
0.4
0.2
0.0
Confusion Matrix
O(+)
O(−)
T(+)
T(−)
Test
Outcome
AUC: 0.635
BACC (0.580)


Time vs. Events: Raw Train: Breast Cancer
Expected
Observed
Events
0
10
20
30
40
50
0
20
40
60
80
Time



As we can see the Observed probability as well as the Time vs. Events are not calibrated.

0.2.2 Uncalibrated Performance Report

```
pander::pander(t(rrAnalysisTrain$OERatio),caption="O/E Ratio")
```

Table 5: O/E Ratio

est	lower	upper
0.843	0.617	1.12

```
pander::pander(t(rrAnalysisTrain$OE95ci),caption="O/E Ratio")
```

Table 6: O/E Ratio

mean	50%	2.5%	97.5%
1	1	0.956	1.06

```
pander::pander(t(rrAnalysisTrain$OAcum95ci),caption="O/Acum Ratio")
```

Table 7: O/Acum Ratio

mean	50%	2.5%	97.5%
0.776	0.776	0.769	0.783

```
pander::pander(rrAnalysisTrain$c.index$cstatCI,caption="C. Index")
```

mean.C Index	median	lower	upper
0.677	0.679	0.595	0.751

```
pander::pander(t(rrAnalysisTrain$ROCAnalysis$aucs),caption="ROC AUC")
```

Table 9: ROC AUC

est	lower	upper
0.635	0.543	0.727

```
pander::pander((rrAnalysisTrain$ROCAnalysis$sensitivity),caption="Sensitivity")
```

Table 10: Sensitivity

est	lower	upper
0.261	0.143	0.411

```
pander::pander((rrAnalysisTrain$ROCAnalysis$specificity),caption="Specificity")
```

Table 11: Specificity

est	lower	upper
0.899	0.838	0.942

```
pander::pander(t(rrAnalysisTrain$thr_atP),caption="Probability Thresholds")
```

Table 12: Probability Thresholds

90%
0.419

```
pander::pander(t(rrAnalysisTrain$RR_atP),caption="Risk Ratio")
```

Table 13: Risk Ratio

est	lower	upper
1.88	1.1	3.2

```
pander::pander(rrAnalysisTrain$surdif,caption="Logrank test")
```

Table 14: Logrank test Chisq = 11.670372 on 1 degrees of freedom, p = 0.000635

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	167	34	41.11	1.23	11.7
class=1	27	12	4.89	10.33	11.7