Breast Cancer: Wisconsin

Jose Tamez

2023-05-18

Contents

1 Wisconsin Prognosis

1.0.1 Libraries

```
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
```

```
op <- par(no.readonly = TRUE)
pander::panderOptions('digits', 3)
#pander::panderOptions('table.split.table', 400)
pander::panderOptions('keep.trailing.zeros',TRUE)
```

1.1 The data

```
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
```

1.2 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
V24	4.69e-02	1.01	1.05	1.08	0.598	0.237
V26	4.72e-03	1.00	1.00	1.01	0.593	0.237
V27	2.42e-04	1.00	1.00	1.00	0.608	0.237
V34	1.19e-02	1.00	1.01	1.02	0.634	0.237
V7	6.05e-08	1.00	1.00	1.00	0.588	0.237
V35	5.06e-06	1.00	1.00	1.00	0.727	0.237

Table 2: Table continues below

	full.Accuracy	u.AUC	r.AUC	full.AUC	IDI	NRI	z.IDI
V24	0.598	0.609	0.5	0.609	0.0619	0.437	2.87
V26	0.593	0.598	0.5	0.598	0.0626	0.393	2.77
V27	0.608	0.608	0.5	0.608	0.0563	0.434	2.76
V34	0.634	0.618	0.5	0.618	0.0320	0.471	2.42
V7	0.588	0.595	0.5	0.595	0.0487	0.380	2.30
V35	0.727	0.641	0.5	0.641	0.0289	0.565	2.28

	z.NRI	Delta.AUC	Frequency
V24	2.67	0.1091	1
V26	2.38	0.0983	1
V27	2.63	0.1084	1
V34	2.85	0.1178	1
V7	2.30	0.0949	1
V35	3.50	0.1412	1

1.3 Cox Model Performance

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

1.3.1 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")
```
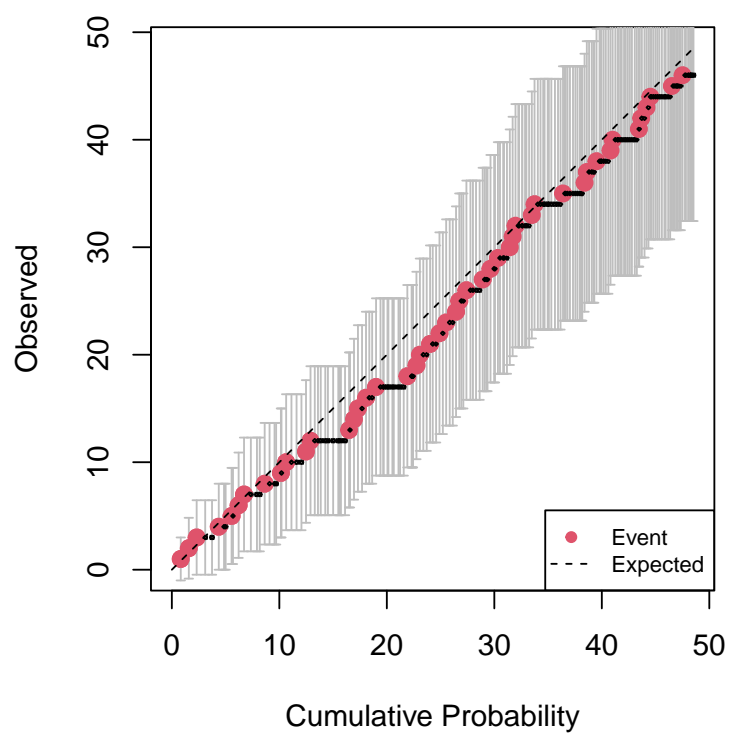
Table 4: Initial Parameters

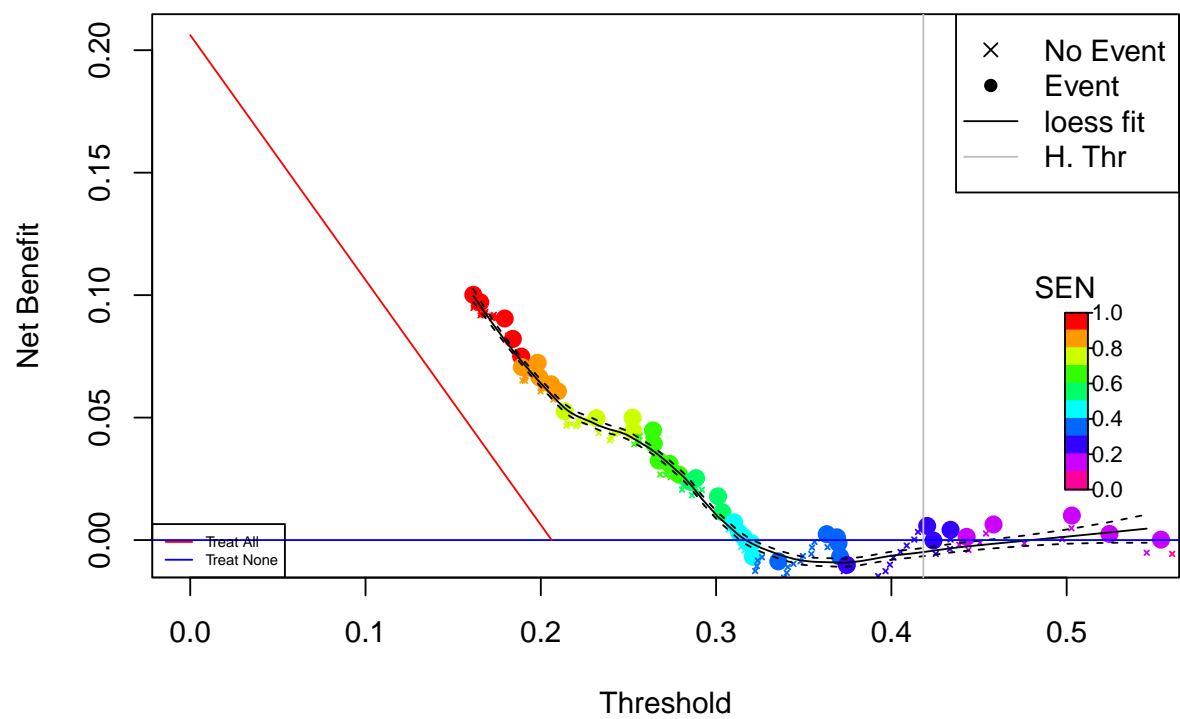| h0	timeinterval
0.323	51.1

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

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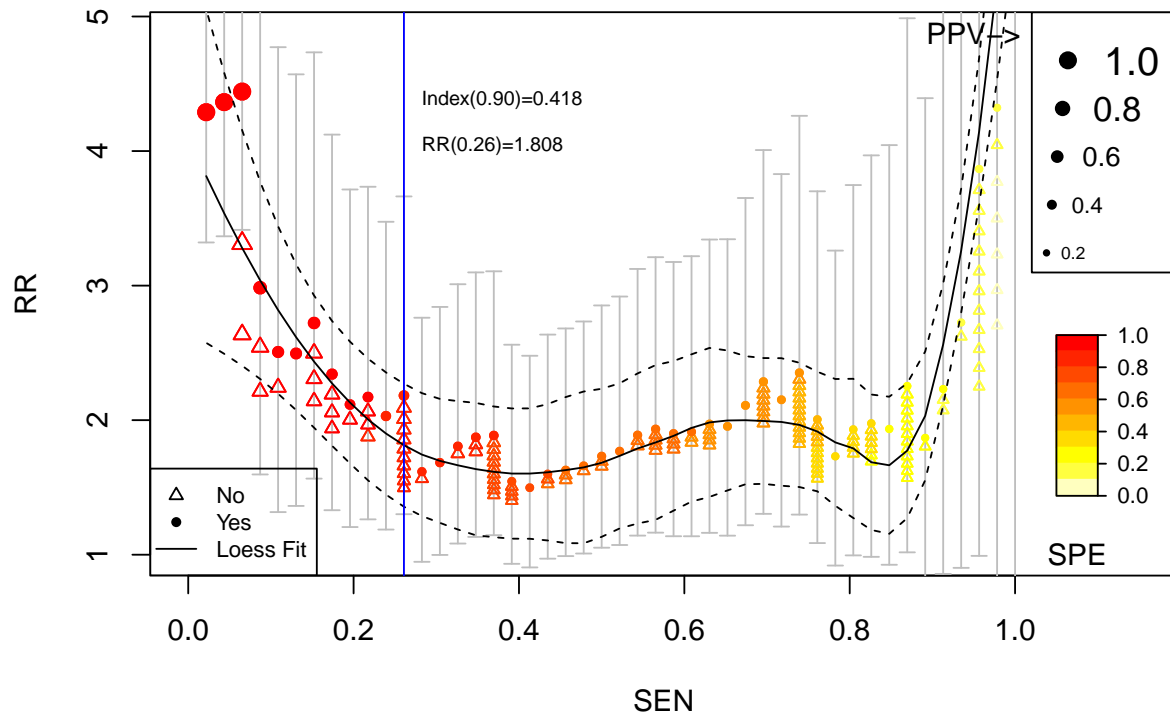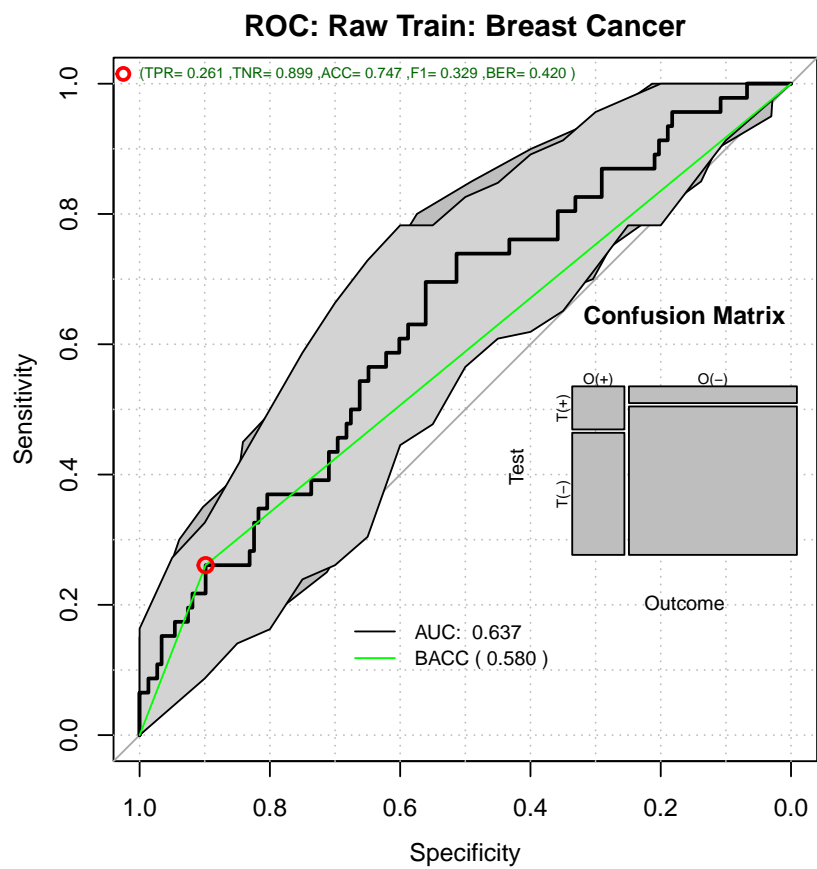
Cumulative vs. Observed: Raw Train: Breast Cancer

Decision Curve Analysis: Raw Train: Breast Cancer

Relative Risk: Raw Train: Breast Cancer

Index(0.90)=0.418

RR(0.26)=1.808

PPV→

SPE

△ No
● Yes
— Loess Fit

6

ROC: Raw Train: Breast Cancer

(TPR= 0.261 ,TNR= 0.899 ,ACC= 0.747 ,F1= 0.329 ,BER= 0.420)

Confusion Matrix

Test

O(+)   O(−)

T(+)

T(−)

Outcome

AUC: 0.637
BACC (0.580)

Sensitivity

Specificity

Time vs. Events: Raw Train: Breast Cancer
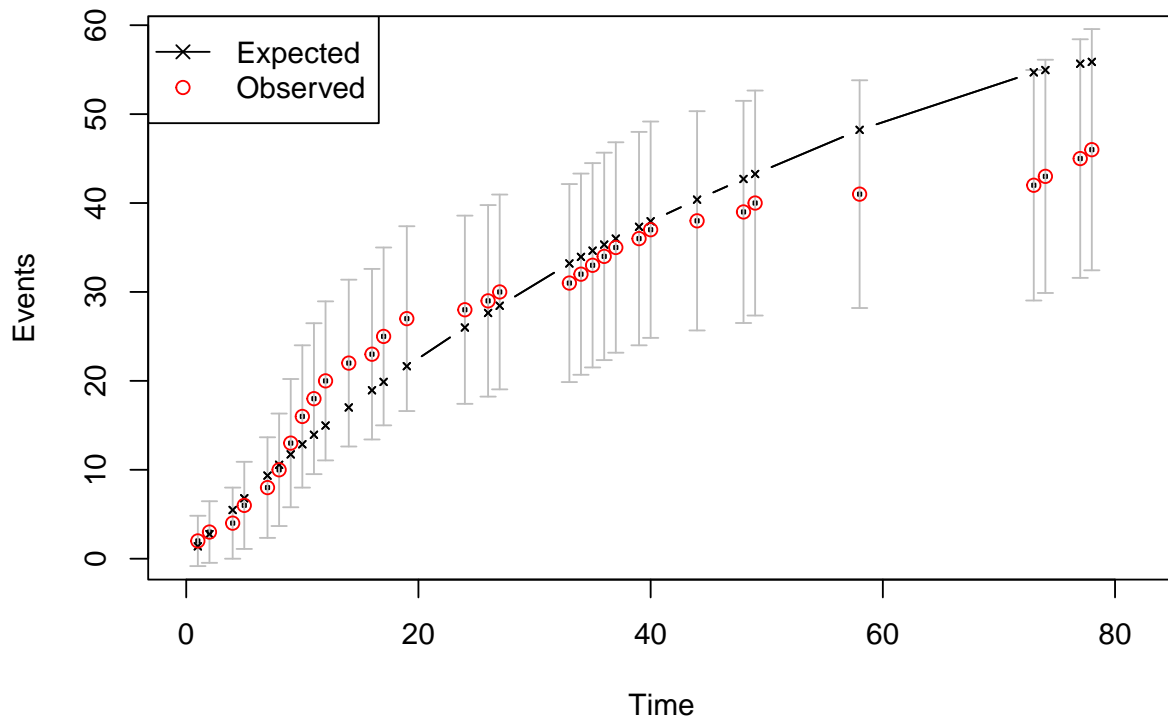
Kaplan–Meier: Raw Train: Breast Cancer
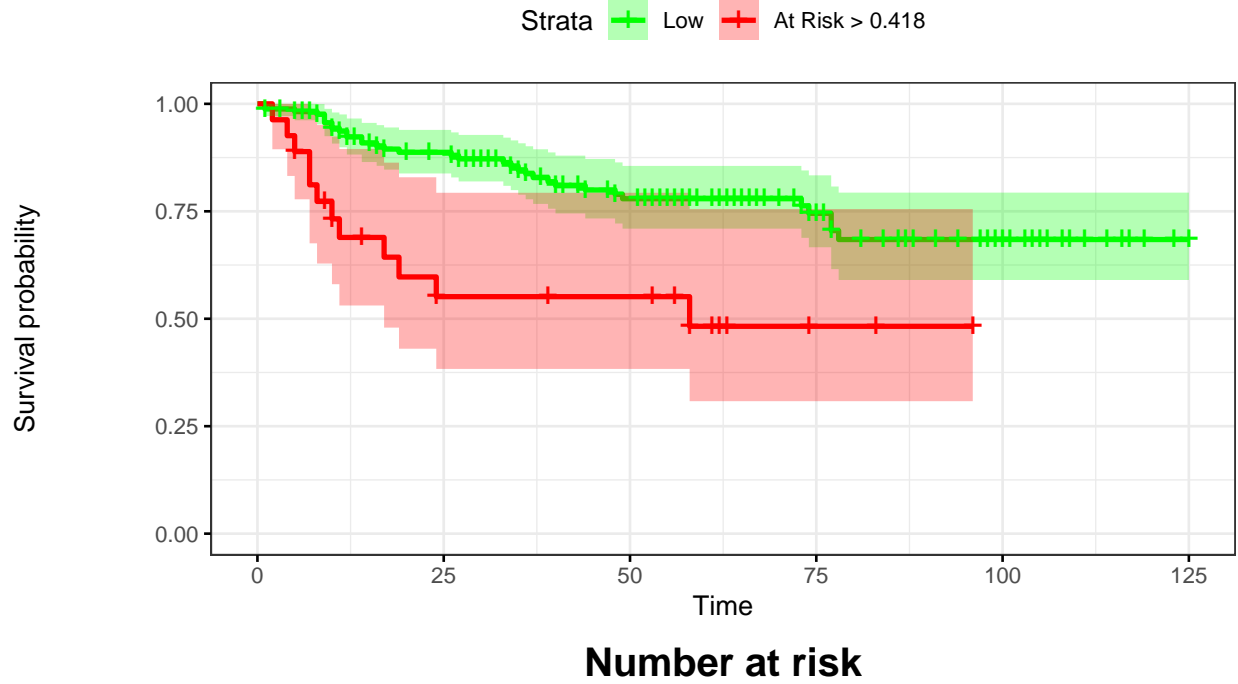
As we can see the Observed probability as well as the Time vs. Events are not calibrated.

1.3.2 Uncalibrated Performance Report

```
pander::pander(t(rrAnalysisTrain$keyPoints),caption="Threshold values")
```

Table 5: Threshold values

	@:0.9	@MAX_BACC	@MAX_RR	@SPE100	p(0.5)
Thr	0.42042	0.2640	0.1655	1.61e-01	0.497448
RR	2.18301	2.2857	4.3220	2.50e+01	2.307692
RR_LCI	1.30105	1.3037	0.6348	5.22e-02	1.275480
RR_UCI	3.66282	4.0074	29.4258	1.20e+04	4.175246
SEN	0.26087	0.6957	0.9783	1.00e+00	0.152174
SPE	0.89865	0.5608	0.1081	6.76e-02	0.952703
BACC	0.57976	0.6282	0.5432	5.34e-01	0.552438
NetBenefit	0.00578	0.0448	0.0971	1.00e-01	0.000374

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

9

Table 6: O/E Ratio

O/E	Low	Upper	p.value
0.823	0.603	1.1	0.203

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

Table 7: O/E Mean

mean	50%	2.5%	97.5%
1.01	1.01	0.95	1.08

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

Table 8: O/Acum Mean

mean	50%	2.5%	97.5%
0.92	0.92	0.912	0.928

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

mean.C Index	median	lower	upper
0.68	0.681	0.606	0.759

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

Table 10: ROC AUC

est	lower	upper
0.637	0.546	0.728

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

Table 11: Sensitivity

est	lower	upper
0.261	0.143	0.411

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

Table 12: Specificity

est	lower	upper
0.899	0.838	0.942

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

Table 13: Probability Thresholds

90%
0.418

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

Table 14: Logrank test Chisq = 11.608565 on 1 degrees of freedom,
p = 0.000656

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	167	34	41.1	1.23	11.6
class=1	27	12	4.9	10.27	11.6

1.3.3 Cox Calibration

```
op <- par(no.readonly = TRUE)


calprob <- CoxRiskCalibration(ml,dataBreast,"status","time")

pander::pander(c(h0=calprob$h0,
              Gain=calprob$hazardGain,
              DeltaTime=calprob$timeInterval),
           caption="Cox Calibration Parameters")
```

h0	Gain	DeltaTime
0.302	0.938	48.3

1.3.4 The RRplot() of the calibrated model

```
h0 <- calprob$h0
timeinterval <- calprob$timeInterval;

rdata <- cbind(dataBreast$status,calprob$prob)


rrAnalysisTrain <- RRPlot(rdata,atRate=c(0.90),
                  timetoEvent=dataBreast$time,
                  title="Calibrated Train: Breast",
                  ysurvlim=c(0.00,1.0),
                  riskTimeInterval=timeinterval)
```
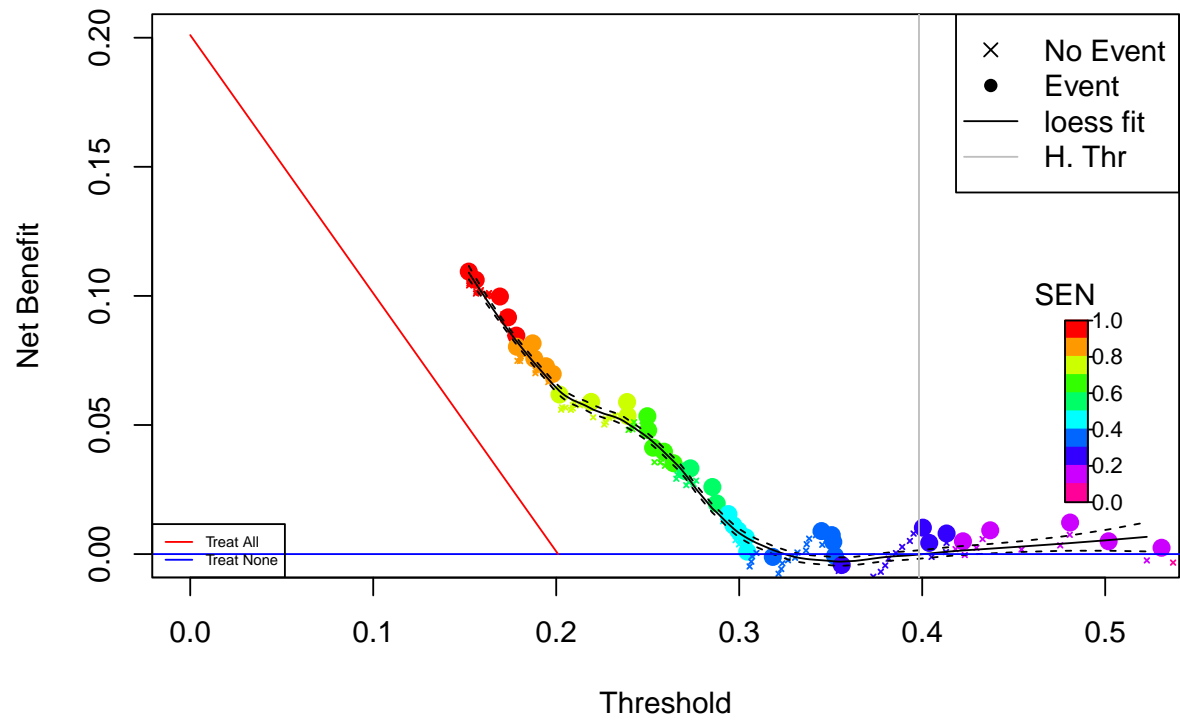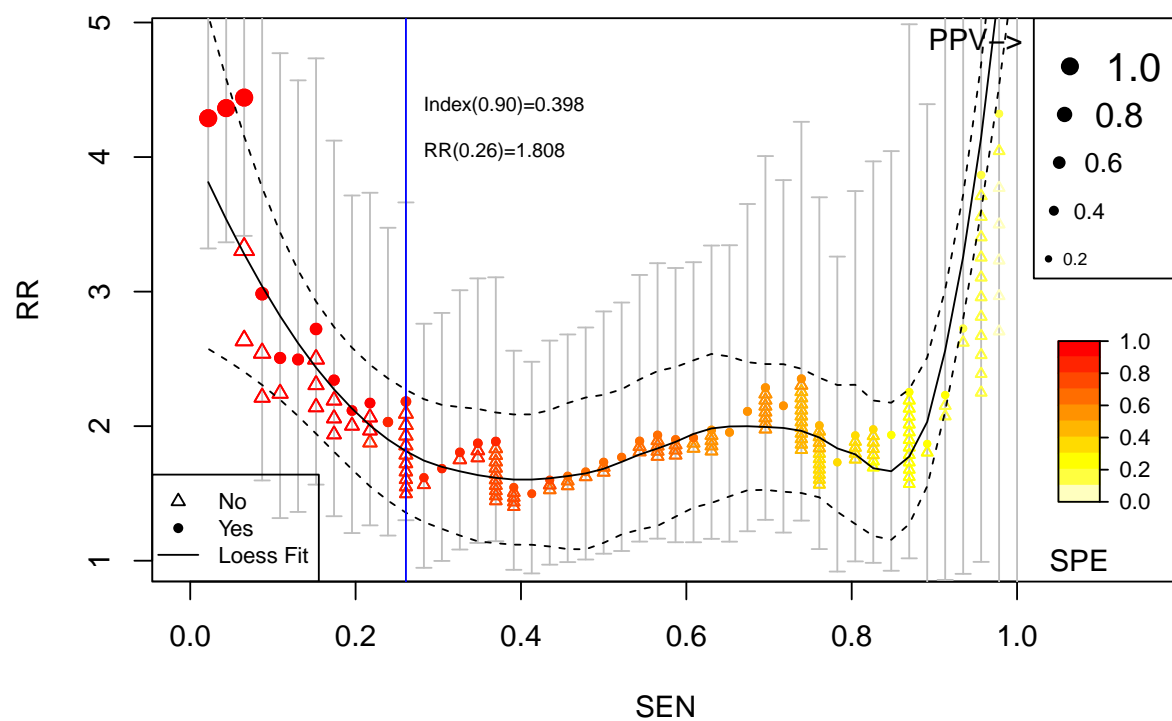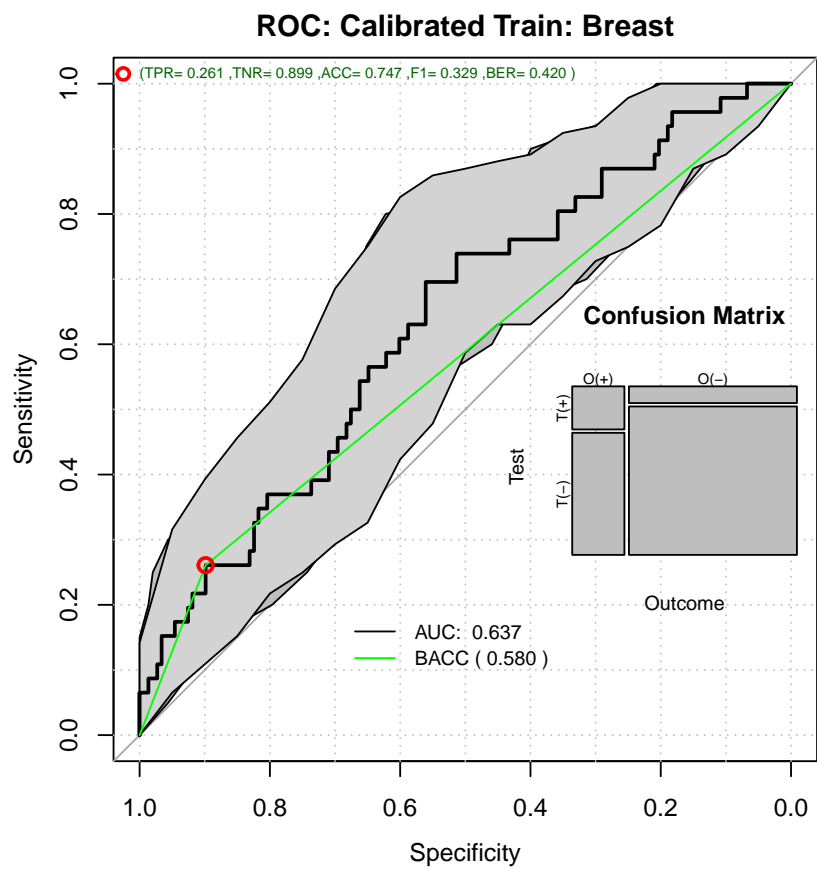
Cumulative vs. Observed: Calibrated Train: Breast
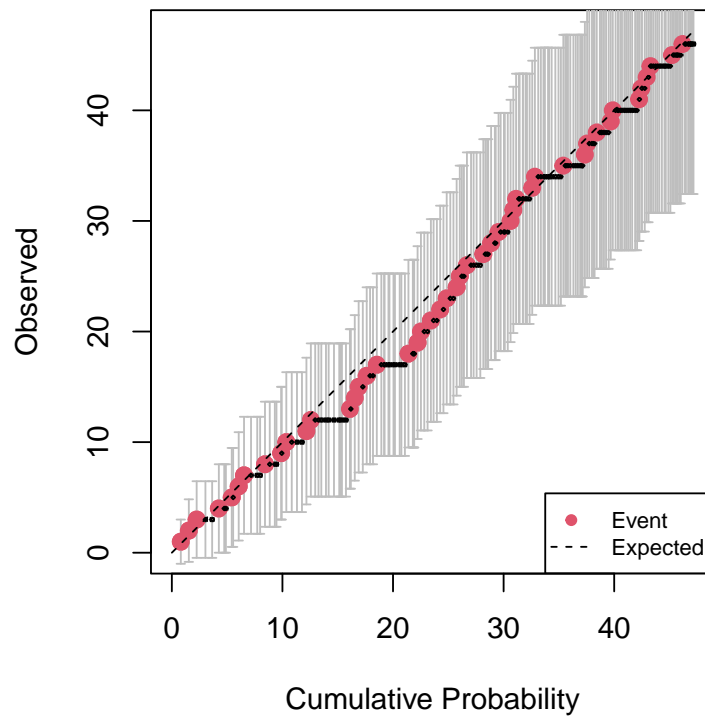
Decision Curve Analysis: Calibrated Train: Breast

Relative Risk: Calibrated Train: Breast

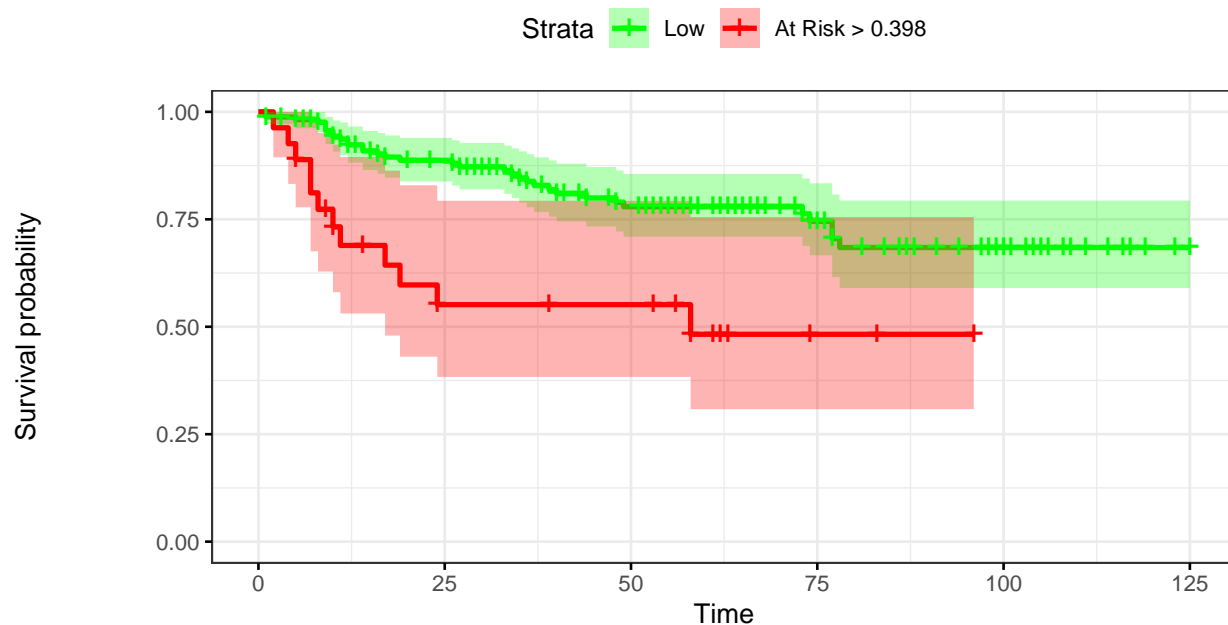# ROC: Calibrated Train: Breast

(TPR= 0.261 ,TNR= 0.899 ,ACC= 0.747 ,F1= 0.329 ,BER= 0.420)

Sensitivity

Specificity

Confusion Matrix

O(+)   O(−)

T(+)

Test

T(−)

Outcome

AUC: 0.637
BACC (0.580)

Time vs. Events: Calibrated Train: Breast
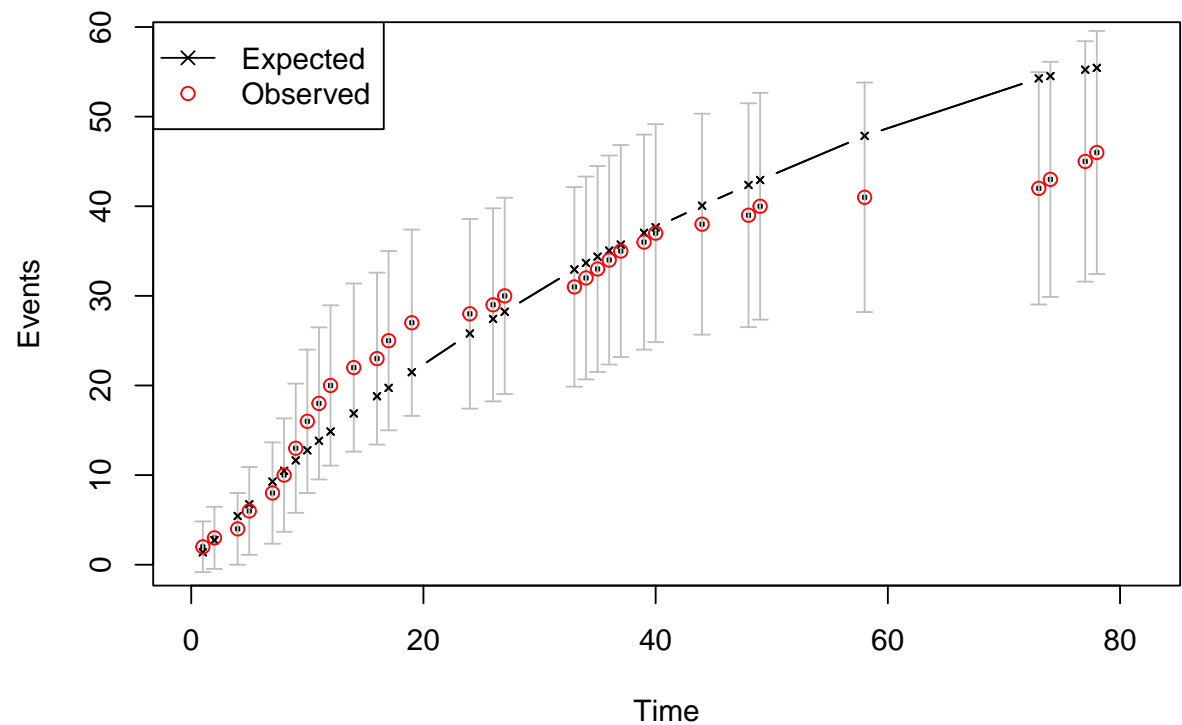
Kaplan–Meier: Calibrated Train: Breast

Number at risk

Low	167	116	76	42	20	1
At Risk > 0.398	27	11	10	2	0	0

1.3.5 Calibrated Train Performance

```
pander::pander(t(rrAnalysisTrain$keyPoints),caption="Threshold values")
```

Table 16: Threshold values

	@:0.9	@MAX_BACC	@MAX_RR	@SPE100	p(0.5)
Thr	0.4004	0.2498	0.156	1.52e-01	0.50186
RR	2.1830	2.2857	4.322	2.50e+01	2.49545
RR_LCI	1.3011	1.3037	0.635	5.22e-02	1.36264
RR_UCI	3.6628	4.0074	29.426	1.20e+04	4.57001
SEN	0.2609	0.6957	0.978	1.00e+00	0.13043
SPE	0.8986	0.5608	0.108	6.76e-02	0.96622
BACC	0.5798	0.6282	0.543	5.34e-01	0.54833
NetBenefit	0.0102	0.0534	0.106	1.09e-01	0.00497

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

Table 17: O/E Ratio

O/E	Low	Upper	p.value
0.83	0.607	1.11	0.226

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

Table 18: O/E Mean

mean	50%	2.5%	97.5%
1.02	1.02	0.958	1.08

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

Table 19: O/Acum Mean

mean	50%	2.5%	97.5%
0.945	0.945	0.936	0.953

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

Table 20: C. Index

mean.C Index	median	lower	upper
0.68	0.681	0.596	0.755

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

Table 21: ROC AUC

est	lower	upper
0.637	0.546	0.728

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

Table 22: Sensitivity

est	lower	upper
0.261	0.143	0.411

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

Table 23: Specificity

est	lower	upper
0.899	0.838	0.942

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

Table 24: Probability Thresholds

90%
0.398

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

Table 25: Logrank test Chisq = 11.608565 on 1 degrees of freedom,
p = 0.000656

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	167	34	41.1	1.23	11.6
class=1	27	12	4.9	10.27	11.6

1.4 Cross-Validation

Here we use the estimated h0 and timeinterval from the full set

```
rcv <- randomCV(theData=dataBreast,
                theOutcome = Surv(time,status)~1,
                fittingFunction=BSWiMS.model,
                trainFraction = 0.9,
                repetitions=100,
                classSamplingType = "Pro"
        )
```

.[+++++].[+++++++++].[++].[+++++++++].[++].[++++].[++++].[+++++++++].[++++].[+++++]10
Tested: 127 Avg. Selected: 5.4 Min Tests: 1 Max Tests: 7 Mean Tests: 1.574803 . MAD: 0.4868298
.[+].[+++++++].[+++++++++].[+++].[++++].[+++++].[++++++++].[+++++++++].[+++++].[++]20
Tested: 173 Avg. Selected: 5.5 Min Tests: 1 Max Tests: 7 Mean Tests: 2.312139 . MAD: 0.4864678
.[+++++].[++++].[+++].[+++].[++].[+++++++++].[+++++++].[++++].[+].[++++++]30   Tested:
188 Avg. Selected: 5.133333 Min Tests: 1 Max Tests: 10 Mean Tests: 3.191489 . MAD: 0.4890094
.[++++].[+].[+++].[+++].[+++++++].[+++++++].[+++++++].[++++].[++++++++].[+++++]40
Tested: 193 Avg. Selected: 5.15 Min Tests: 1 Max Tests: 10 Mean Tests: 4.145078 . MAD:
0.4870537 .[+].[+++++++].[+].[+++++].[+++++].[+++].[++++++].[++++++].[+++++].[+++]50 Tested:
194 Avg. Selected: 4.92 Min Tests: 1 Max Tests: 13 Mean Tests: 5.154639 . MAD: 0.4864017
.[+++].[++++++].[++].[+++].[+++++++].[+++++].[+++++++].[+++].[+++++++].[++++]60 Tested:
194 Avg. Selected: 4.916667 Min Tests: 1 Max Tests: 14 Mean Tests: 6.185567 . MAD: 0.4855313
.[+++].[+++++].[+++++].[+++].[+++++].[+++++].[+++].[+++].[+++++].[++++++]70 Tested: 194
Avg. Selected: 4.842857 Min Tests: 2 Max Tests: 15 Mean Tests: 7.216495 . MAD: 0.4840267
.[+++++++].[++].[+++++].[++++++].[++++++].[+++++].[++].[++++++].[+++++++].[+++++]80 Tested:
194 Avg. Selected: 4.8625 Min Tests: 2 Max Tests: 16 Mean Tests: 8.247423 . MAD: 0.4834583
.[+++++++].[+++].[+++].[++++++].[++++++].[++++++++].[+++++].[++++].[+++].[++]90 Tested:
194 Avg. Selected: 4.844444 Min Tests: 2 Max Tests: 17 Mean Tests: 9.278351 . MAD: 0.4845374
.[++++++].[+++].[+++++].[++++++++].[++++++++].[-].[++++].[++++].[+++++++].[+++++++]100
Tested: 194 Avg. Selected: 4.87 Min Tests: 2 Max Tests: 19 Mean Tests: 10.30928 . MAD: 0.4841087

```
stp <- rcv$survTestPredictions
stp <- stp[!is.na(stp[,4]),]

bbx <- boxplot(unlist(stp[,1])~rownames(stp),plot=FALSE)
times <- bbx$stats[3,]
status <- boxplot(unlist(stp[,2])~rownames(stp),plot=FALSE)$stats[3,]
```
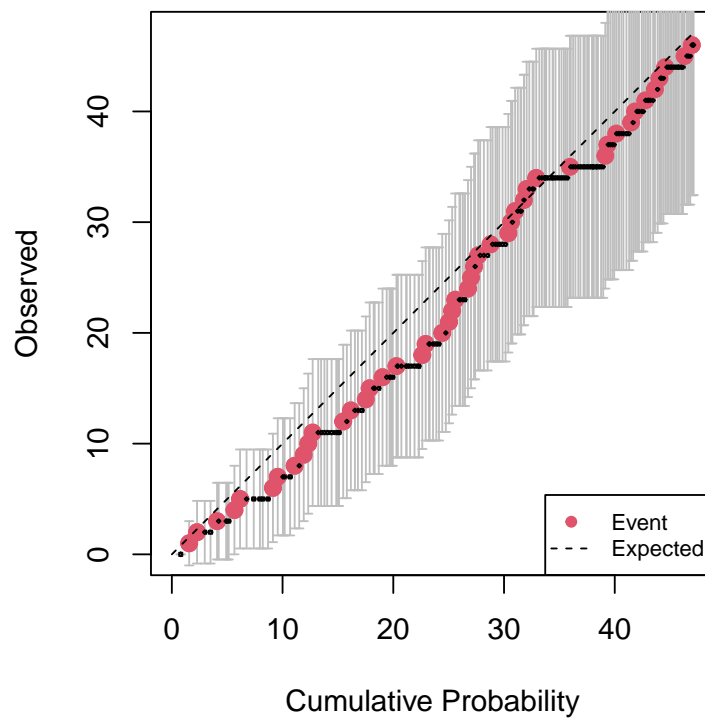
```
prob <- ppoisGzero(boxplot(unlist(stp[,4])~rownames(stp),plot=FALSE)$stats[3,],h0)

rdatacv <- cbind(status,prob)
rownames(rdatacv) <- bbx$names
names(times) <- bbx$names

rrAnalysisTest <- RRPlot(rdatacv,atThr = rrAnalysisTrain$thr_atP,
                    timetoEvent=times,
                    title="Test: Breast Cancer",
                    ysurvlim=c(0.00,1.0),
                    riskTimeInterval=timeinterval)
```
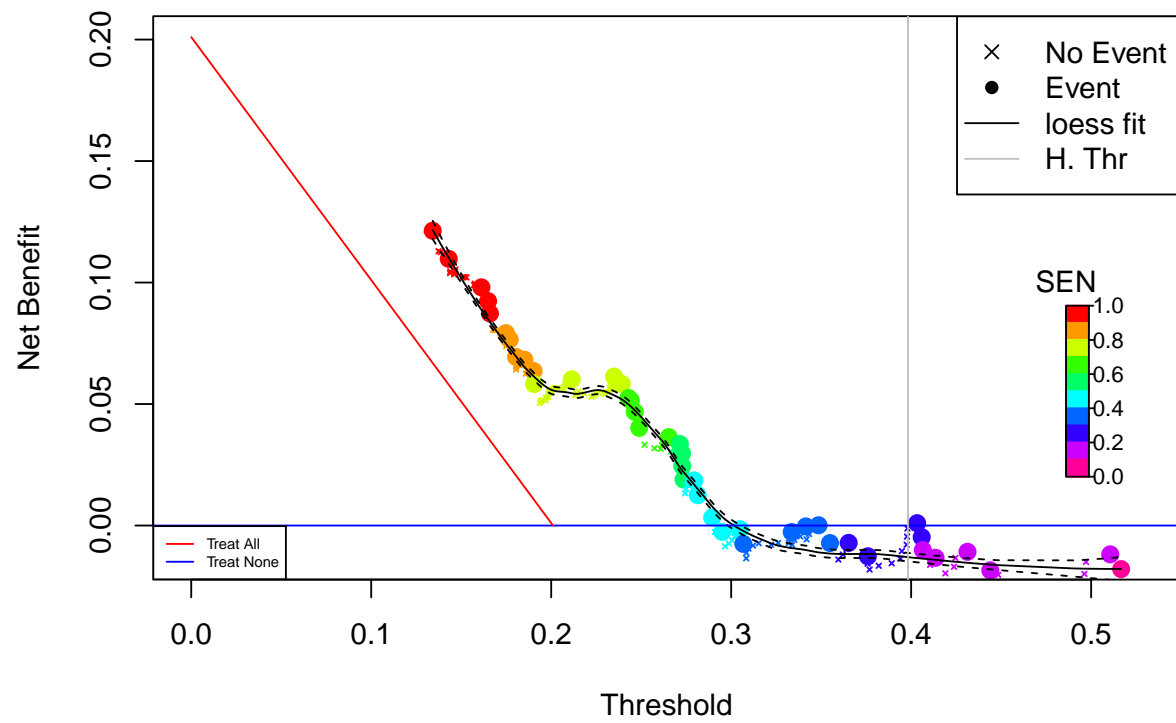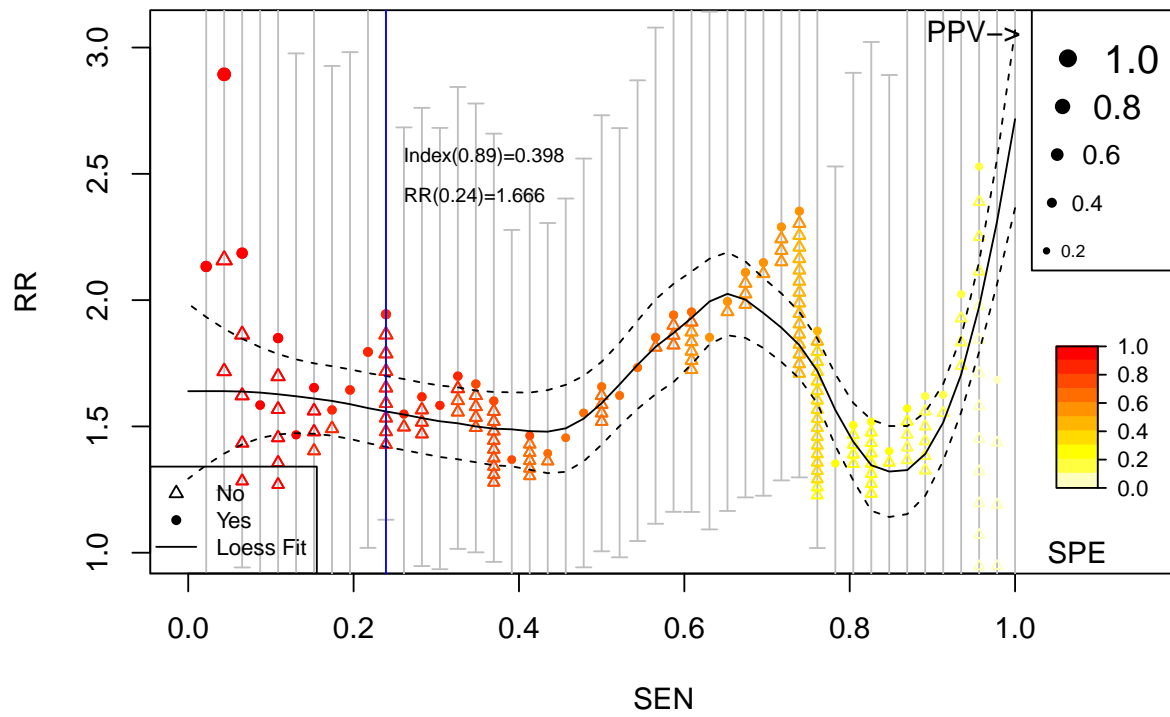
Cumulative vs. Observed: Test: Breast Cancer

Decision Curve Analysis: Test: Breast Cancer

Relative Risk: Test: Breast Cancer

Index(0.89)=0.398

RR(0.24)=1.666

PPV–>

SPE

- △ No
- ● Yes
- — Loess Fit

RR

SEN

ROC: Test: Breast Cancer

(TPR= 0.239 ,TNR= 0.892 ,ACC= 0.737 ,F1= 0.301 ,BER= 0.434)

Confusion Matrix

AUC: 0.608
BACC (0.566)
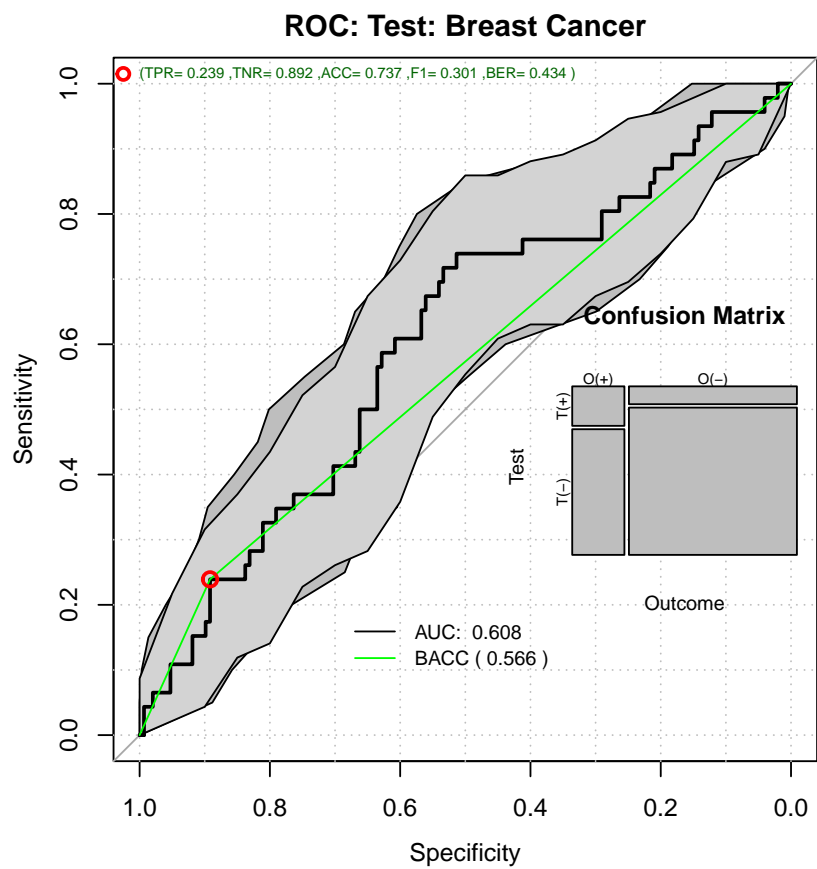
Time vs. Events: Test: Breast Cancer
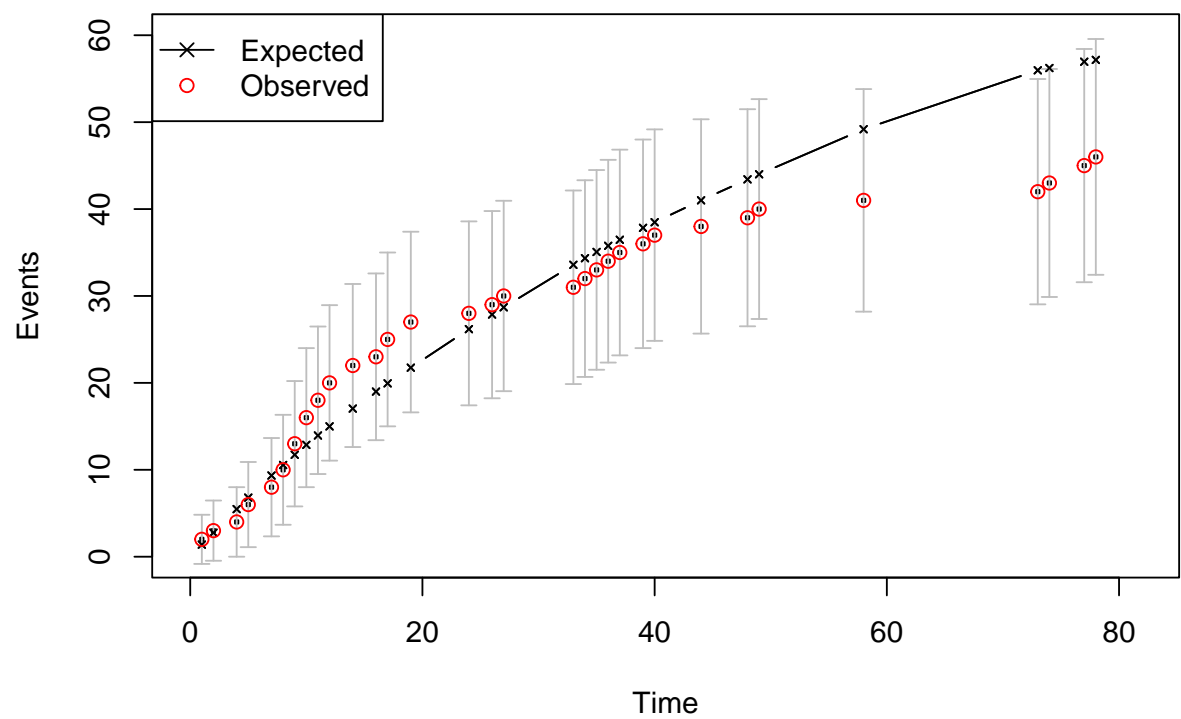
Kaplan–Meier: Test: Breast Cancer

Strata ┼ Low ┼ At Risk > 0.398

Number at risk

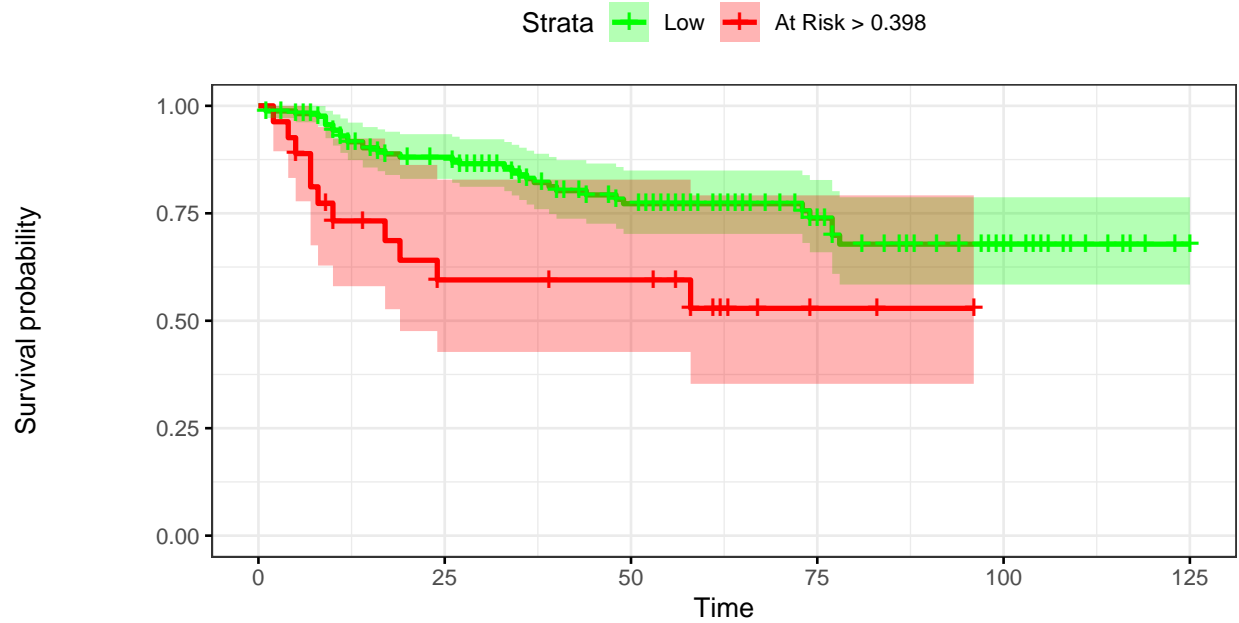|					
Low	167	115	75	42	20	1
At Risk > 0.398	27	12	11	2	0	0

1.4.1 Cross-Validation Test Performance

```
pander::pander(t(rrAnalysisTest$keyPoints),caption="Threshold values")
```

Table 26: Threshold values

	@:0.398	@MAX_BACC	@MAX_RR	@SPE100	p(0.5)
Thr	0.39796	0.2349	0.161	1.34e-01	0.497
RR	1.86327	2.3522	2.529	7.23e+00	1.698
RR_LCI	1.07879	1.2983	0.663	1.62e-02	0.811
RR_UCI	3.21820	4.2617	9.651	3.22e+03	3.553
SEN	0.23913	0.7391	0.957	1.00e+00	0.109
SPE	0.88514	0.5135	0.122	2.03e-02	0.946
BACC	0.56213	0.6263	0.539	5.10e-01	0.527
NetBenefit	-0.00121	0.0613	0.098	1.21e-01	-0.015

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

Table 27: O/E Ratio

O/E	Low	Upper	p.value
0.805	0.589	1.07	0.146

```
pander::pander(t(rrAnalysisTest$OE95ci),caption="O/E Mean")
```

Table 28: O/E Mean

mean	50%	2.5%	97.5%
1	1	0.944	1.07

```
pander::pander(t(rrAnalysisTest$OAcum95ci),caption="O/Acum Mean")
```

Table 29: O/Acum Mean

mean	50%	2.5%	97.5%
0.897	0.897	0.883	0.909

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

mean.C Index	median	lower	upper
0.654	0.654	0.568	0.733

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

Table 31: ROC AUC

est	lower	upper
0.608	0.513	0.702

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

Table 32: Sensitivity

est	lower	upper
0.239	0.126	0.388

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

Table 33: Specificity

est	lower	upper
0.892	0.83	0.937

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

Table 34: Probability Thresholds

90%
0.398

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

Table 35: Logrank test Chisq = 7.737689 on 1 degrees of freedom,
p = 0.005408

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	167	35	40.9	0.85	7.74
class=1	27	11	5.1	6.81	7.74

1.4.2 Calibrating the test results

```
rdatacv <- cbind(status,prob,times)
calprob <- CalibrationProbPoissonRisk(rdatacv)

pander::pander(c(h0=calprob$h0,
                Gain=calprob$hazardGain,
                DeltaTime=calprob$timeInterval),
             caption="Cox Calibration Parameters")
```
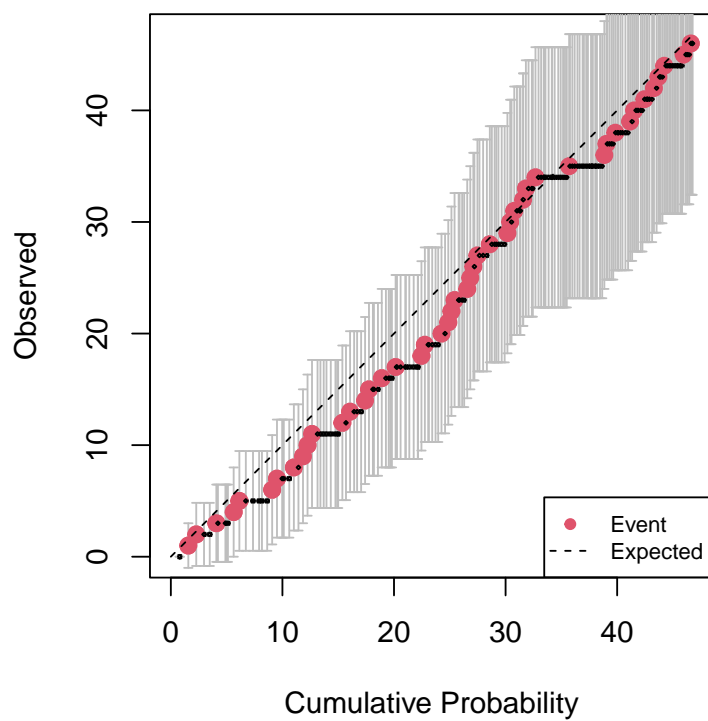
Table: Cox Calibration Parameters

h0	Gain	DeltaTime
0.323	1	49.2

```
timeinterval <- calprob$timeInterval;

rdata <- cbind(status,calprob$prob)


rrAnalysisTest <- RRPlot(rdata,atRate=c(0.90),
                  timetoEvent=times,
                  title="Calibrated Test: Breast",
                  ysurvlim=c(0.00,1.0),
                  riskTimeInterval=timeinterval)
```
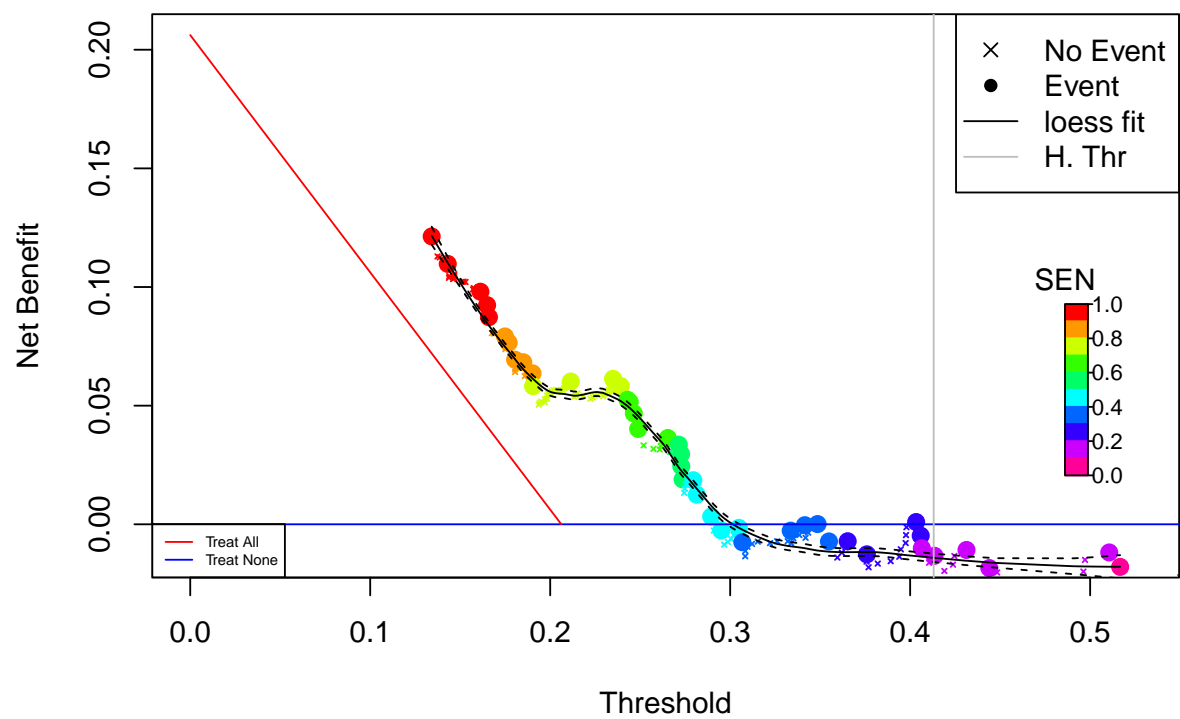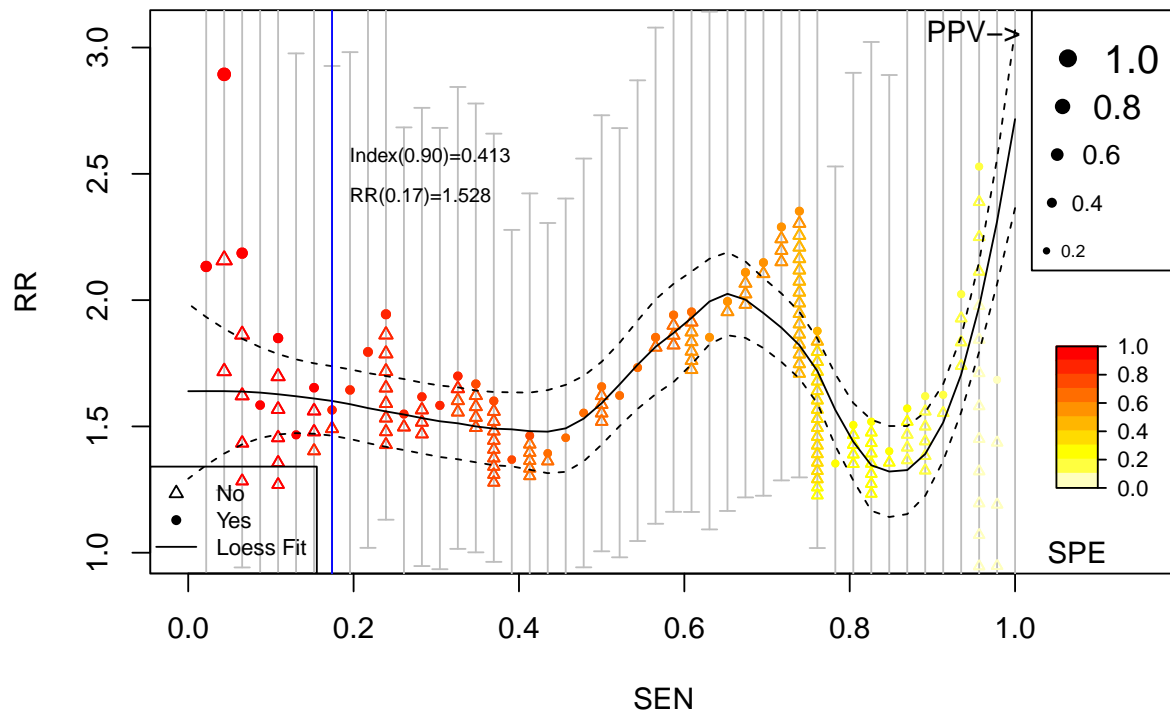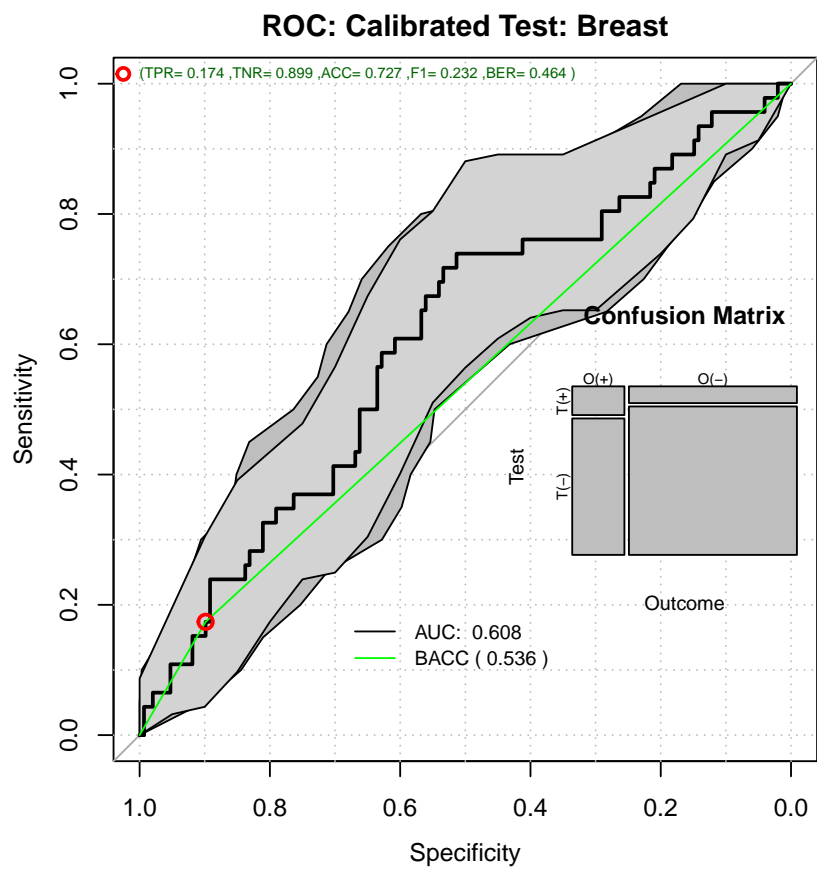
Cumulative vs. Observed: Calibrated Test: Breast

Observed

Cumulative Probability

- Event
- Expected

Decision Curve Analysis: Calibrated Test: Breast

Relative Risk: Calibrated Test: Breast

Index(0.90)=0.413

RR(0.17)=1.528

ROC: Calibrated Test: Breast

(TPR= 0.174 ,TNR= 0.899 ,ACC= 0.727 ,F1= 0.232 ,BER= 0.464)

Sensitivity

Specificity

AUC: 0.608
BACC (0.536)

Confusion Matrix

Test

O(+)   O(−)

T(+)

T(−)

Outcome

Time vs. Events: Calibrated Test: Breast
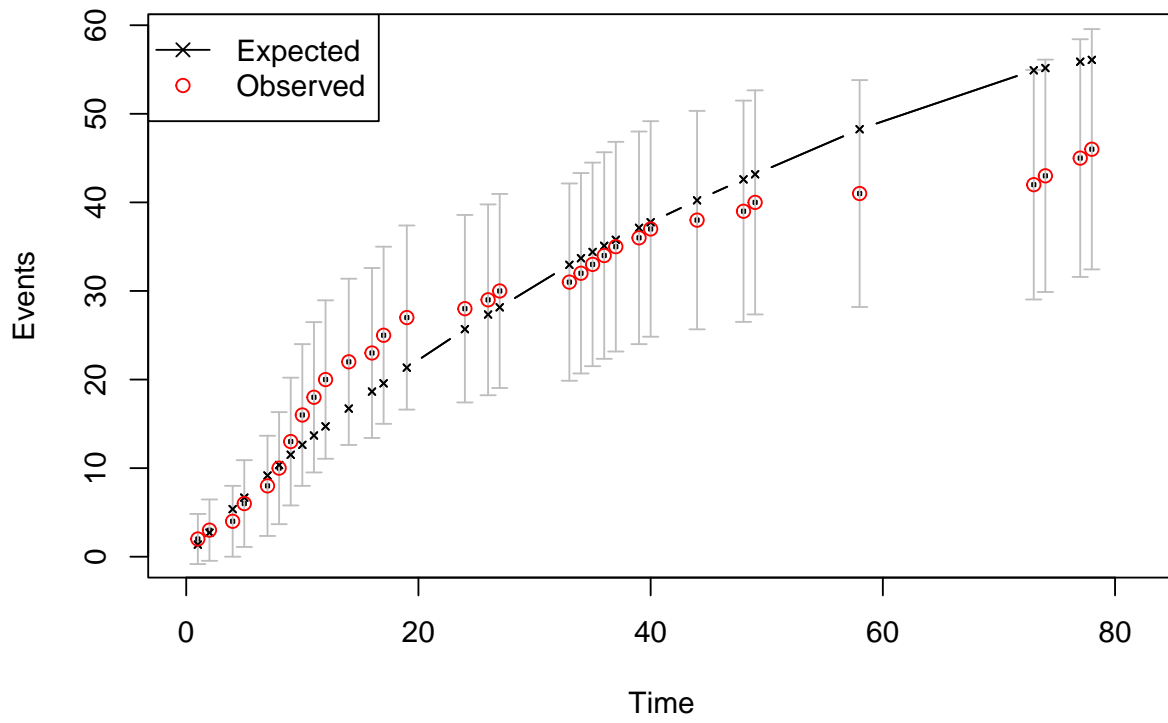
Kaplan–Meier: Calibrated Test: Breast

Strata + Low + At Risk > 0.413
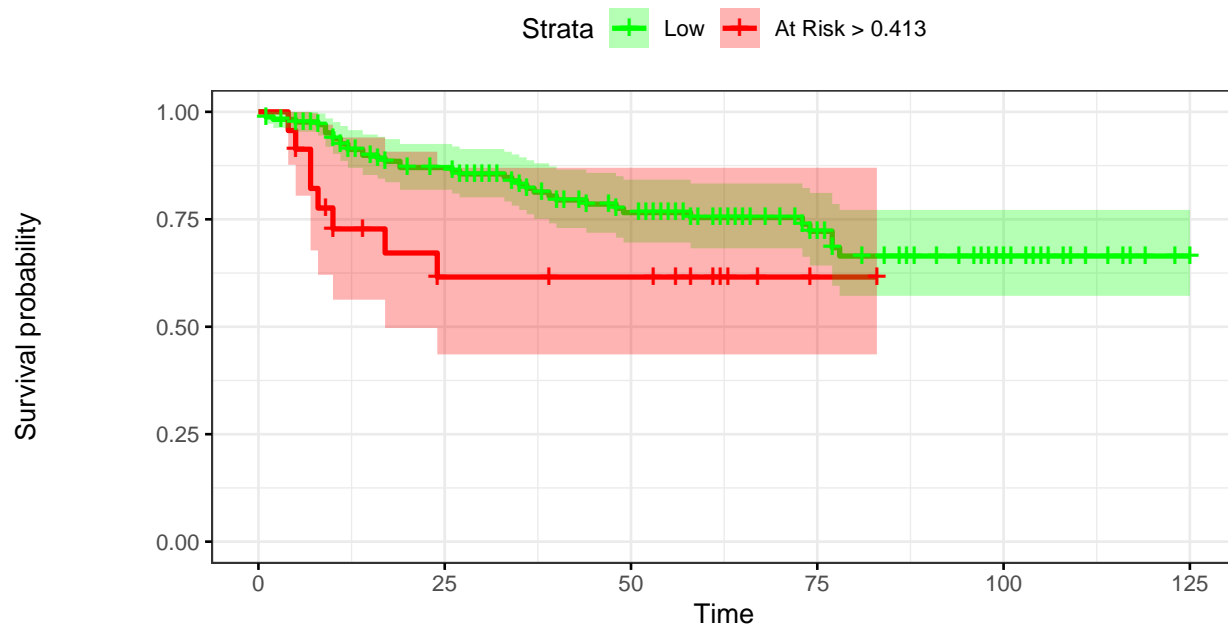
Number at risk

Low	171	117	77	43	20	1
At Risk > 0.413	23	10	9	1	0	0

Calibrated Test Performance

```
pander::pander(t(rrAnalysisTest$keyPoints),caption="Threshold values")
```

Table 37: Threshold values

	@:0.9	@MAX_BACC	@MAX_RR	@SPE100	p(0.5)
Thr	0.4135	0.2349	0.161	1.34e-01	0.497
RR	1.5652	2.3522	2.529	7.23e+00	1.698
RR_LCI	0.8370	1.2983	0.663	1.62e-02	0.811
RR_UCI	2.9270	4.2617	9.651	3.22e+03	3.553
SEN	0.1739	0.7391	0.957	1.00e+00	0.109
SPE	0.8986	0.5135	0.122	2.03e-02	0.946
BACC	0.5363	0.6263	0.539	5.10e-01	0.527
NetBenefit	-0.0133	0.0613	0.098	1.21e-01	-0.015

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

Table 38: O/E Ratio

O/E	Low	Upper	p.value
0.82	0.6	1.09	0.204

```
pander::pander(t(rrAnalysisTest$OE95ci),caption="O/E Mean")
```

Table 39: O/E Mean

mean	50%	2.5%	97.5%
1.02	1.02	0.958	1.09

```
pander::pander(t(rrAnalysisTest$OAcum95ci),caption="O/Acum Mean")
```

Table 40: O/Acum Mean

mean	50%	2.5%	97.5%
0.903	0.903	0.889	0.917

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

mean.C Index	median	lower	upper
0.654	0.654	0.57	0.733

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

Table 42: ROC AUC

est	lower	upper
0.608	0.513	0.702

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

Table 43: Sensitivity

est	lower	upper
0.174	0.0782	0.314

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

Table 44: Specificity

est	lower	upper
0.899	0.838	0.942

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

Table 45: Probability Thresholds

90%
0.413

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

Table 46: Logrank test Chisq = 3.831057 on 1 degrees of freedom,
p = 0.050311

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	171	38	41.8	0.346	3.83
class=1	23	8	4.2	3.443	3.83