Breast Cancer: Wisconsin

Jose Tamez

2023-05-21

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
V27	5.36e-04	1	1.00	1.00	0.608	0.237
V26	4.93e-03	1	1.00	1.01	0.593	0.727
V24	1.00e-02	1	1.01	1.02	0.598	0.634
V7	1.34e-07	1	1.00	1.00	0.588	0.237
V35	1.46e-02	1	1.01	1.03	0.727	0.593
V34	1.18e-02	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
V27	0.608	0.608	0.500	0.608	0.0563	0.434	2.76
V26	0.619	0.598	0.641	0.615	0.0612	0.423	2.70
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	0.619	0.641	0.598	0.615	0.0275	0.551	2.24
V34	0.603	0.618	0.609	0.613	0.0233	0.411	2.13

	z.NRI	Delta.AUC	Frequency
V27	2.63	0.10840	1
V26	2.58	-0.02600	1
V24	1.94	-0.00529	1
V7	2.30	0.09489	1
V35	3.41	0.01689	1
V34	2.47	0.00338	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")
```
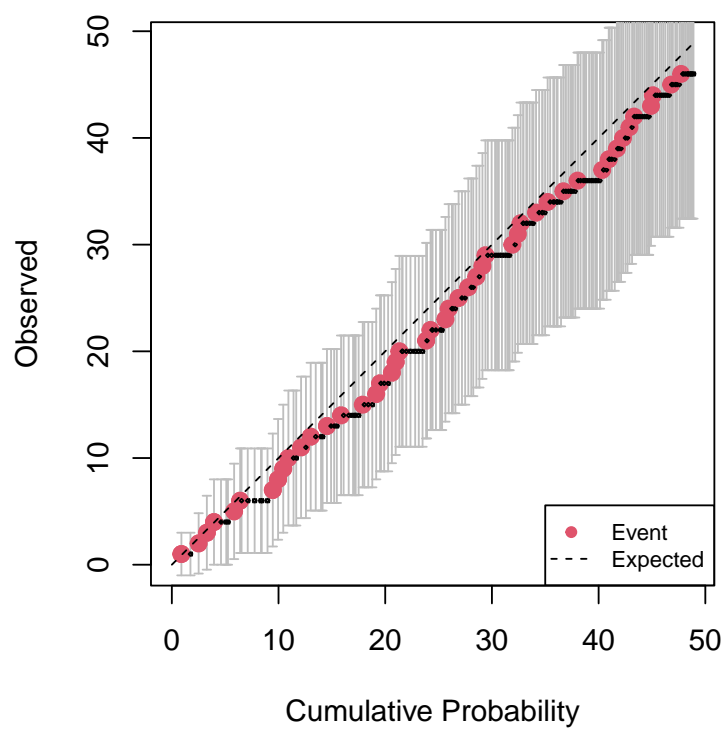
Table 4: Initial Parameters

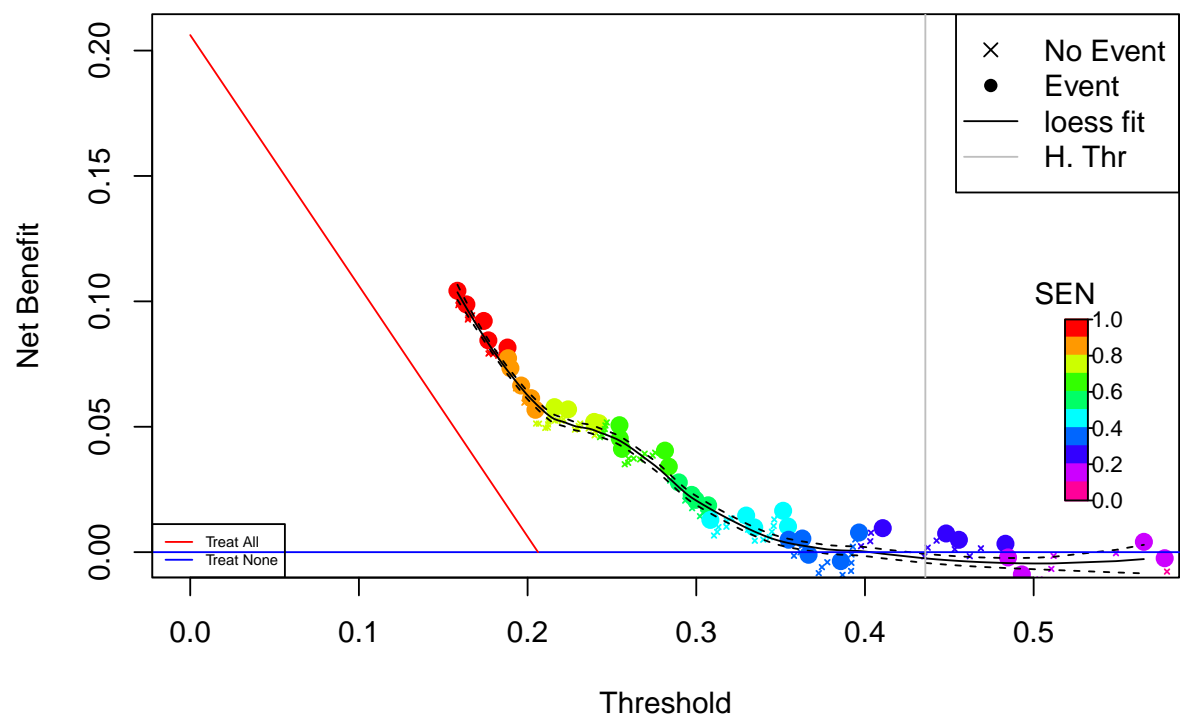| h0	timeinterval
0.323	51.1

```
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)
```
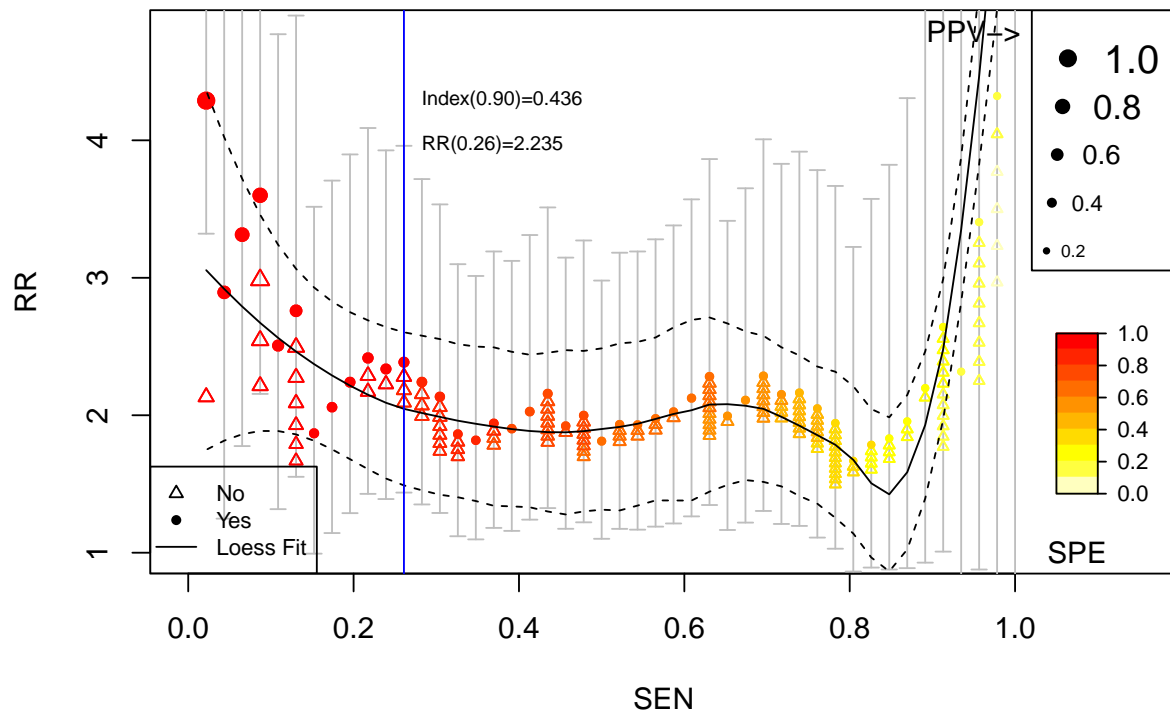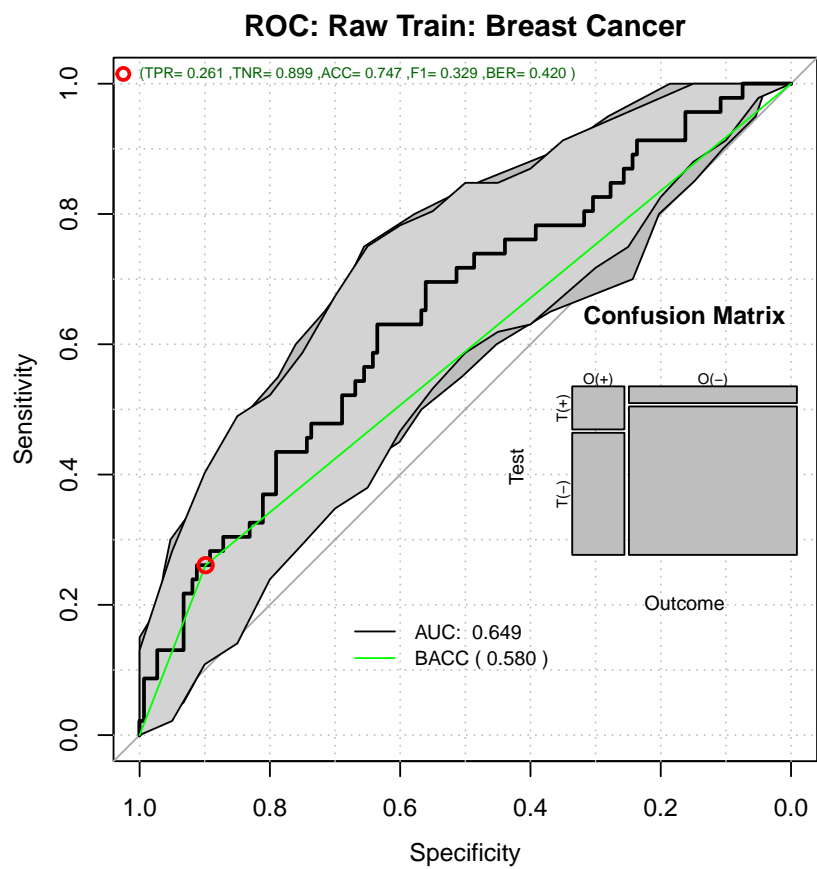
Cumulative vs. Observed: Raw Train: Breast Cancer
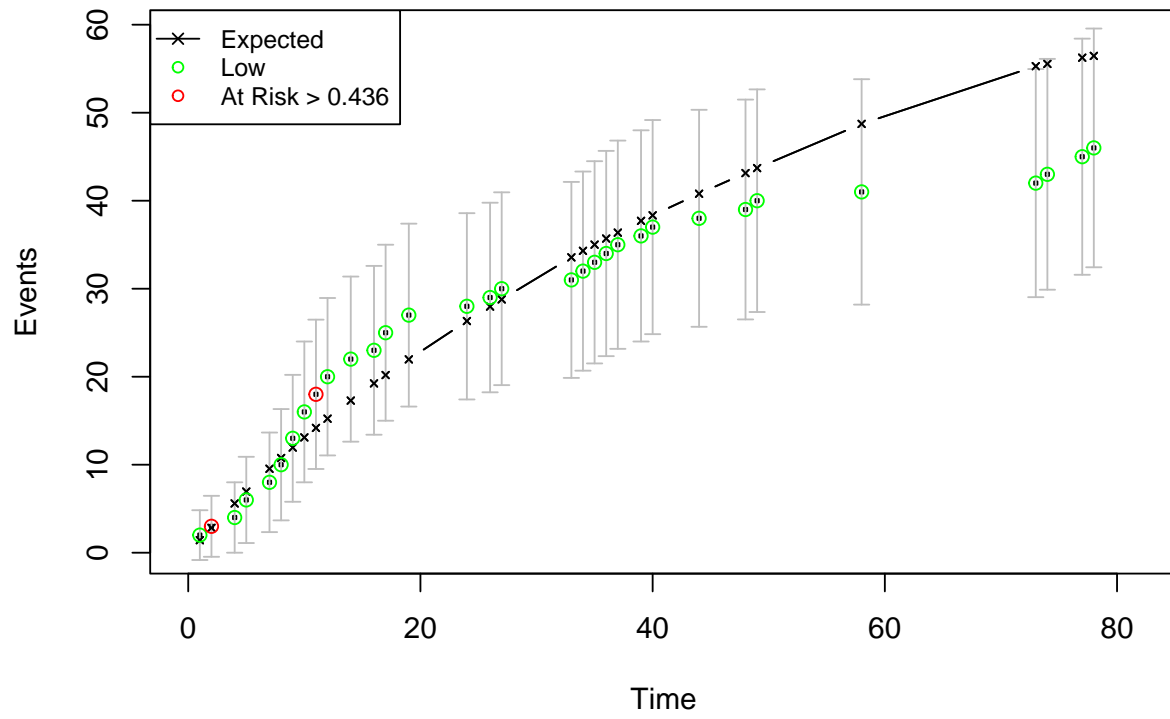
Decision Curve Analysis: Raw Train: Breast Cancer

Relative Risk: Raw Train: Breast Cancer

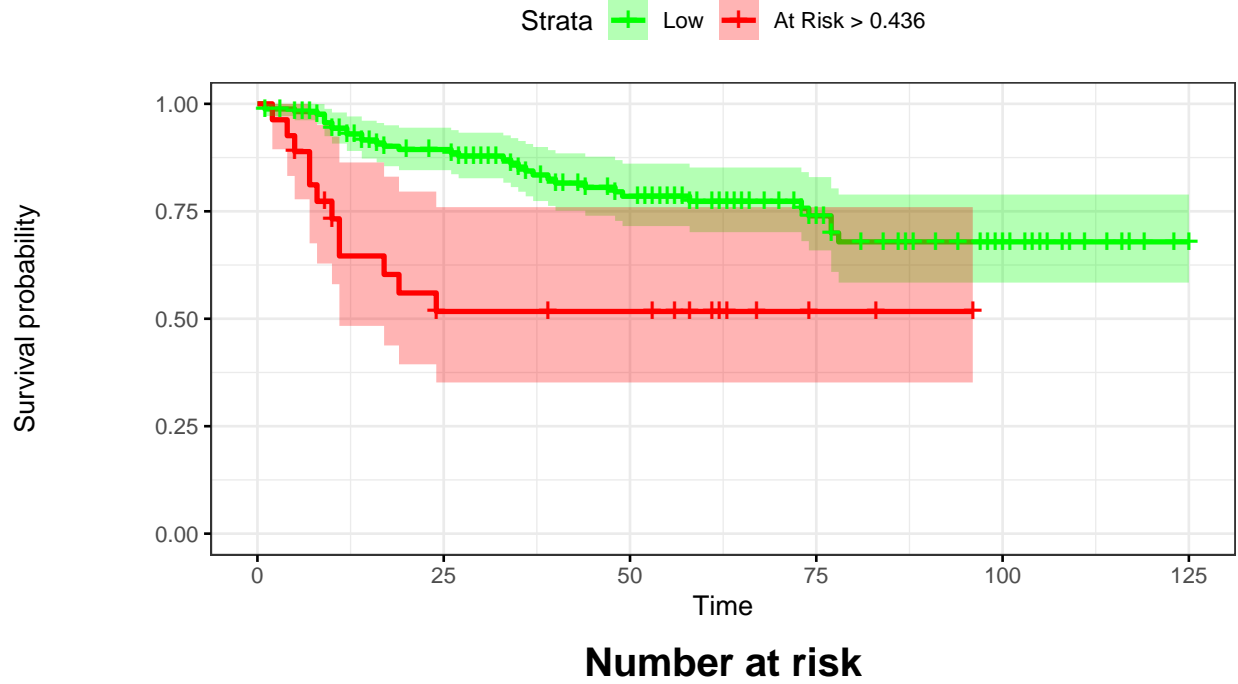ROC: Raw Train: Breast Cancer

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

Confusion Matrix

AUC: 0.649
BACC (0.580)

Time vs. Events: Raw Train: Breast Cancer

Legend:
- Expected
- Low
- At Risk > 0.436

X-axis: Time
Y-axis: Events

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	0.43519	0.2814	0.1637	1.58e-01	0.5003
RR	2.09244	2.2814	4.3220	2.77e+01	1.6687
RR_LCI	1.24116	1.3471	0.6348	5.75e-02	0.8379
RR_UCI	3.52758	3.8636	29.4258	1.33e+04	3.3235
SEN	0.26087	0.6304	0.9783	1.00e+00	0.1304
SPE	0.89189	0.6351	0.1081	7.43e-02	0.9324
BACC	0.57638	0.6328	0.5432	5.37e-01	0.5314
NetBenefit	-0.00168	0.0405	0.0988	1.04e-01	-0.0207

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

Table 6: O/E Ratio

O/E	Low	Upper	p.value
0.815	0.597	1.09	0.183

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

Table 7: O/E Mean

mean	50%	2.5%	97.5%
0.999	1	0.946	1.06

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

Table 8: O/Acum Mean

mean	50%	2.5%	97.5%
0.915	0.916	0.907	0.923

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

mean.C Index	median	lower	upper
0.687	0.687	0.602	0.762

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

Table 10: ROC AUC

est	lower	upper
0.649	0.557	0.74

```
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%
0.436

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

Table 14: Logrank test Chisq = 11.751276 on 1 degrees of freedom,
p = 0.000608

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	167	34	41.12	1.23	11.8
class=1	27	12	4.88	10.40	11.8

1.3.3 Cox Calibration

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

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

(48.41555 , 146.2793 , 0.9469027 , 46 , 50.46889)

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

h0	Gain	DeltaTime
0.3	0.929	48.4

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",
```
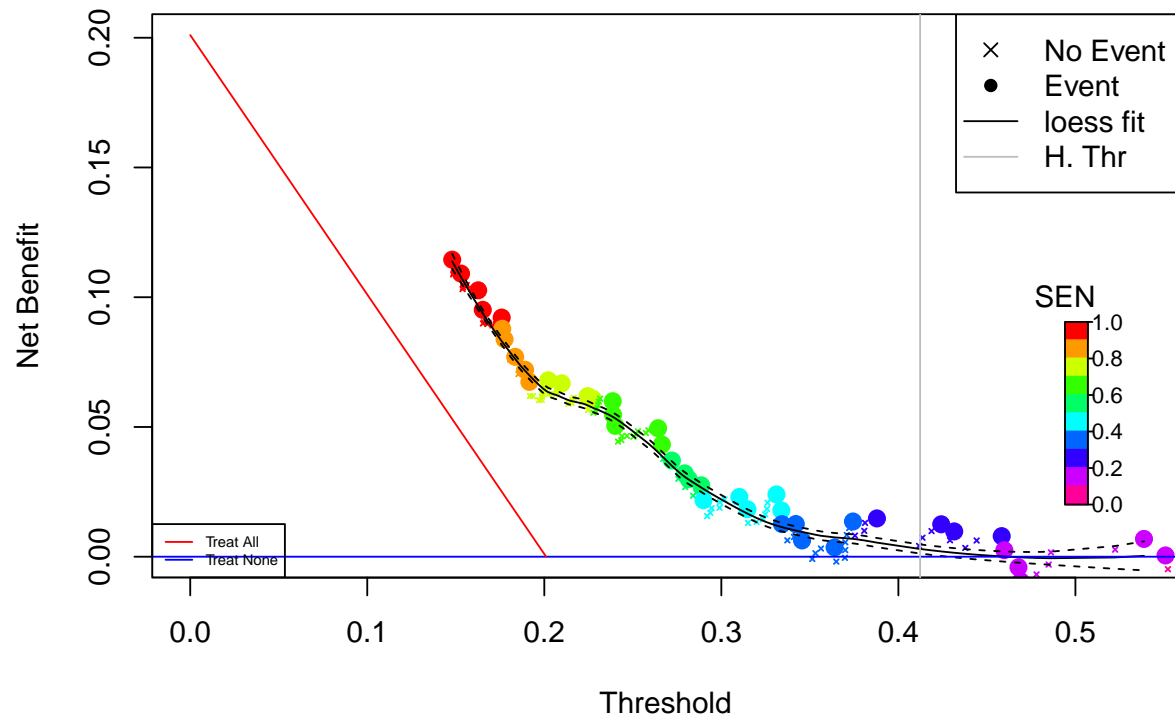
```
                ysurvlim=c(0.00,1.0),
                riskTimeInterval=timeinterval)
```
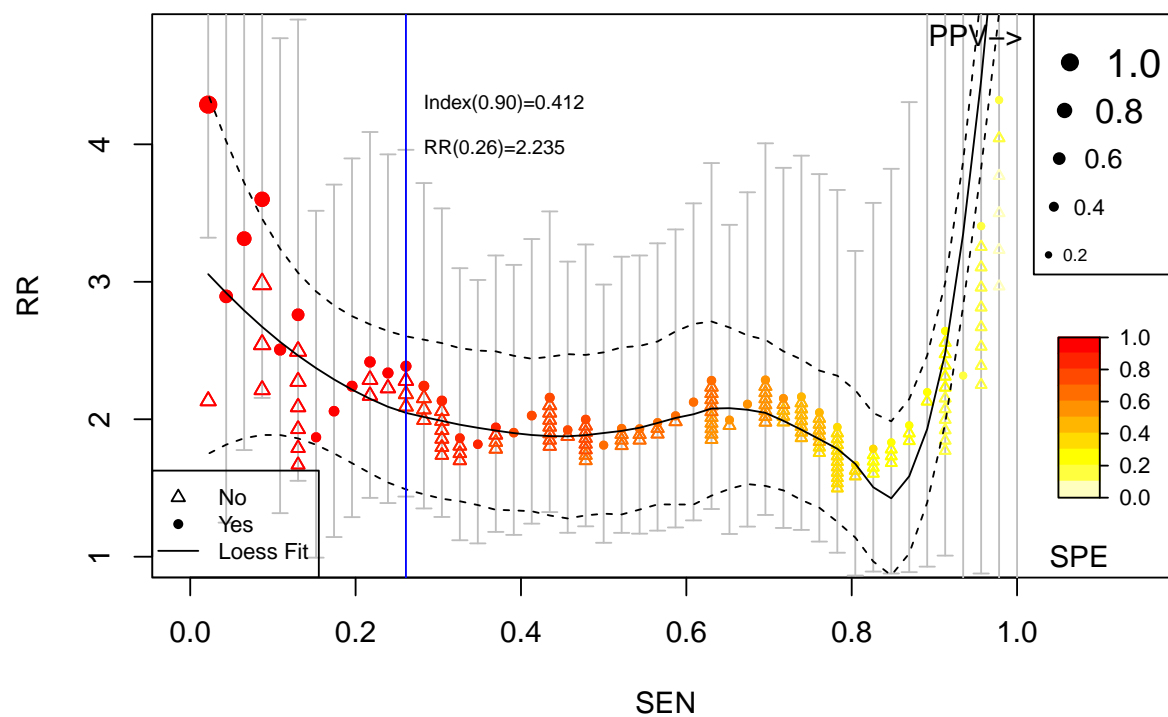
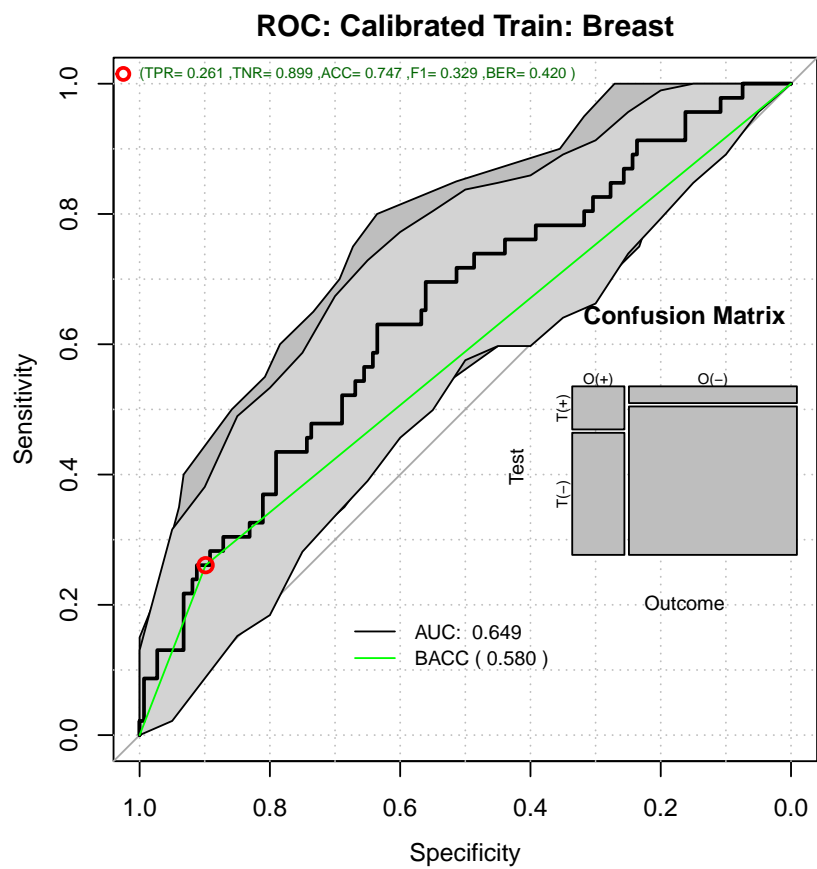## Cumulative vs. Observed: Calibrated Train: Breast
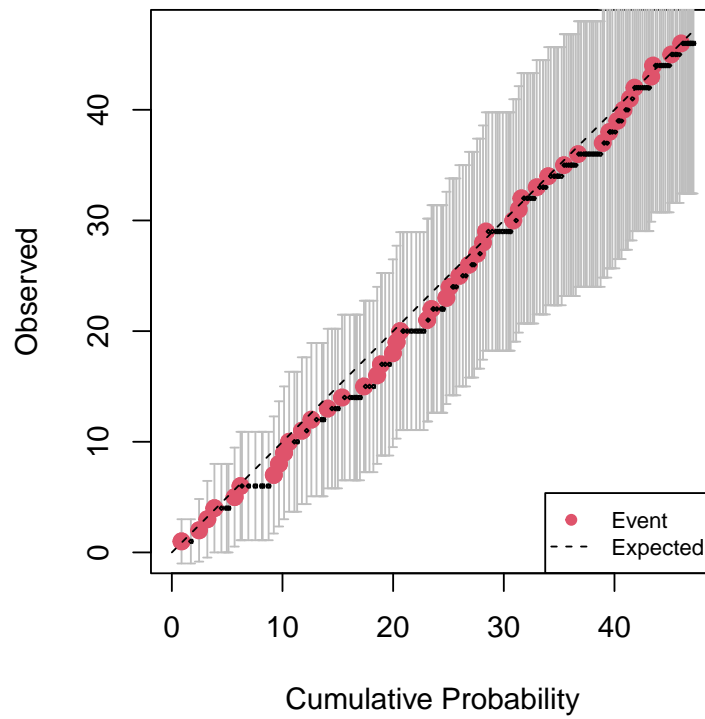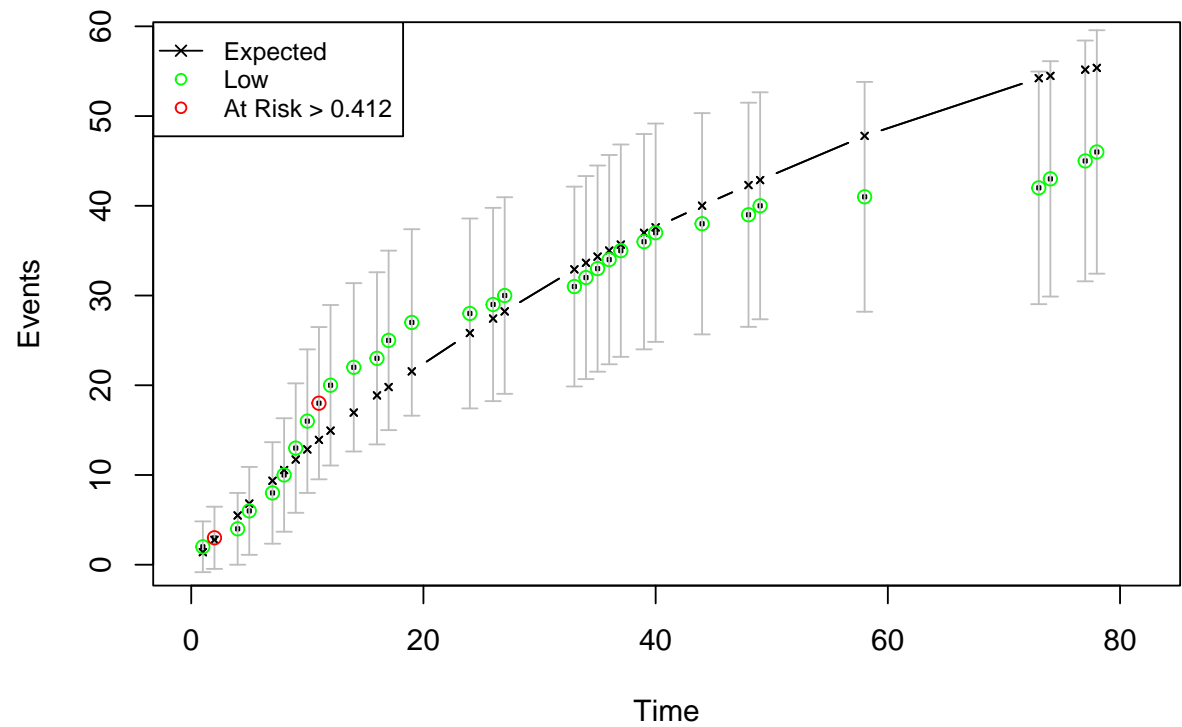
Decision Curve Analysis: Calibrated Train: Breast

Relative Risk: Calibrated Train: Breast

Index(0.90)=0.412

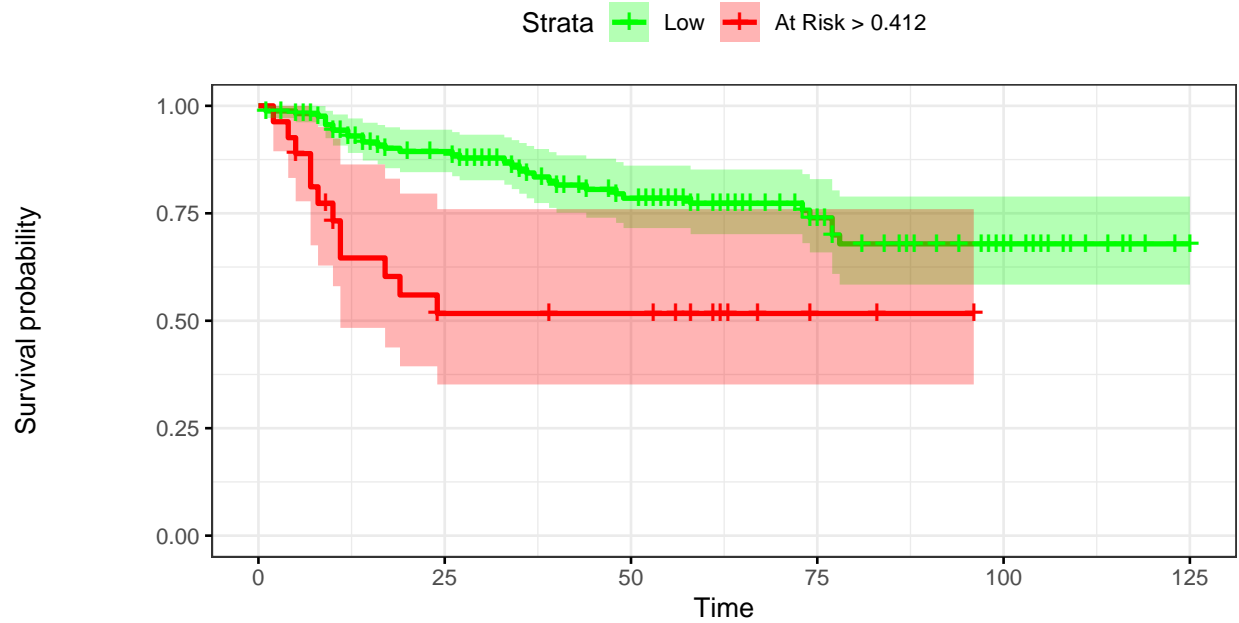RR(0.26)=2.235

PPV->

SPE

SEN

RR

△ No
● Yes
— Loess Fit

14

ROC: Calibrated Train: Breast

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

Confusion Matrix

AUC: 0.649
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 > 0.412	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	0.41170	0.2643	0.153	1.48e-01	0.48619
RR	2.09244	2.2814	4.322	2.77e+01	2.27500
RR_LCI	1.24116	1.3471	0.635	5.75e-02	1.21341
RR_UCI	3.52758	3.8636	29.426	1.33e+04	4.26534
SEN	0.26087	0.6304	0.978	1.00e+00	0.13043
SPE	0.89189	0.6351	0.108	7.43e-02	0.95946
BACC	0.57638	0.6328	0.543	5.37e-01	0.54495
NetBenefit	0.00415	0.0495	0.109	1.14e-01	0.00167

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

Table 17: O/E Ratio

O/E	Low	Upper	p.value
0.831	0.608	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	0.961	1.08

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

Table 19: O/Acum Mean

mean	50%	2.5%	97.5%
0.947	0.947	0.939	0.955

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

Table 20: C. Index

mean.C Index	median	lower	upper
0.687	0.688	0.602	0.765

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

Table 21: ROC AUC

est	lower	upper
0.649	0.557	0.74

```
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%
0.412

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

Table 25: Logrank test Chisq = 11.751276 on 1 degrees of freedom, p = 0.000608

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	167	34	41.12	1.23	11.8
class=1	27	12	4.88	10.40	11.8

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: 131 Avg. Selected: 4.8 Min Tests: 1 Max Tests: 4 Mean Tests: 1.526718 . MAD: 0.4994976 .[+++++++].[+].[−].[−].[−].[−].[+++].[+++++++].[++++].[++++++]20 Tested: 163 Avg. Selected: 3.7 Min Tests: 1 Max Tests: 9 Mean Tests: 2.453988 . MAD: 0.4672188 .[+++].[+++++].[+++].[+++++].[+++++++].[+++++].[+++].[++++++++++].[++++++].[++]30 Tested: 181 Avg. Selected: 4.166667 Min Tests: 1 Max Tests: 10 Mean Tests: 3.314917 . MAD: 0.463722 .[++++++].[+++].[++++].[++++++].[-++].[+++].[+++].[++++].[+++++++].[++++++++]40 Tested: 193 Avg. Selected: 4.225 Min Tests: 1 Max Tests: 13 Mean Tests: 4.145078 . MAD: 0.4747876 .[+++++++].[+++++++].[+++++].[++++++++].[+++].[+].[+++++++].[++++++].[++++++++].[+++]50 Tested: 194 Avg. Selected: 4.46 Min Tests: 1 Max Tests: 14 Mean Tests: 5.154639 . MAD: 0.4797498 .[++++++++].[+++++].[+++++].[+++++++++].[++++++++].[+++++].[+++].[++].[++++++++].[++]60 Tested: 194 Avg. Selected: 4.683333 Min Tests: 1 Max Tests: 14 Mean Tests: 6.185567 . MAD: 0.4793757 .[++++].[+++++].[+++].[++++++].[+++++].[+++++++].[+++++++].[++++++].[++++++].[+++++++]70 Tested: 194 Avg. Selected: 4.757143 Min Tests: 2 Max Tests: 16 Mean Tests: 7.216495 . MAD: 0.4795574 .[+++].[++++++].[++++++++].[++++++++].[+++].[+++++].[++].[++++++].[++].[++++++++]80 Tested: 194 Avg. Selected: 4.8125 Min Tests: 2 Max Tests: 17 Mean Tests: 8.247423 . MAD: 0.4797133 .[−].[++++].[++++++++].[+++].[+++++++].[+++++].[+++++].[++++].[++++].[++++]90 Tested: 194 Avg. Selected: 4.788889 Min Tests: 2 Max Tests: 20 Mean Tests: 9.278351 . MAD: 0.4800506 .[+++++].[++++++++].[+++++].[++++].[+++++++].[++++++++].[+++++].[+++++++].[+++++].[+++++++]100 Tested: 194 Avg. Selected: 4.88 Min Tests: 3 Max Tests: 21 Mean Tests: 10.30928 . MAD: 0.4804926

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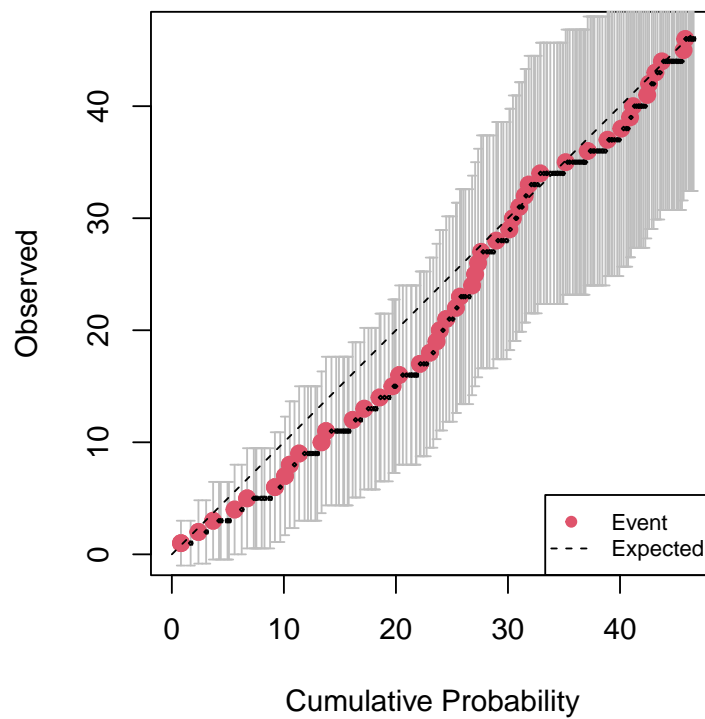
```
prob <- ppoisGzero(boxplot(unlist(stp[,4])~rownames(stp),plot=FALSE)$stats[3,],h0)

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)
```
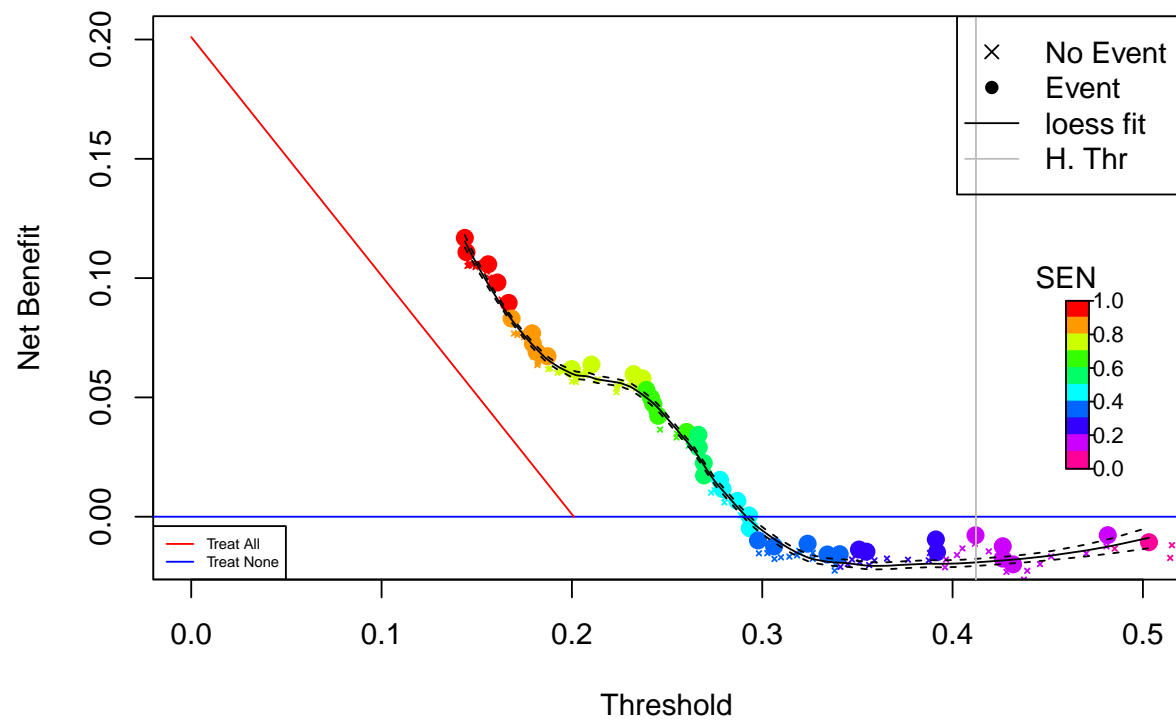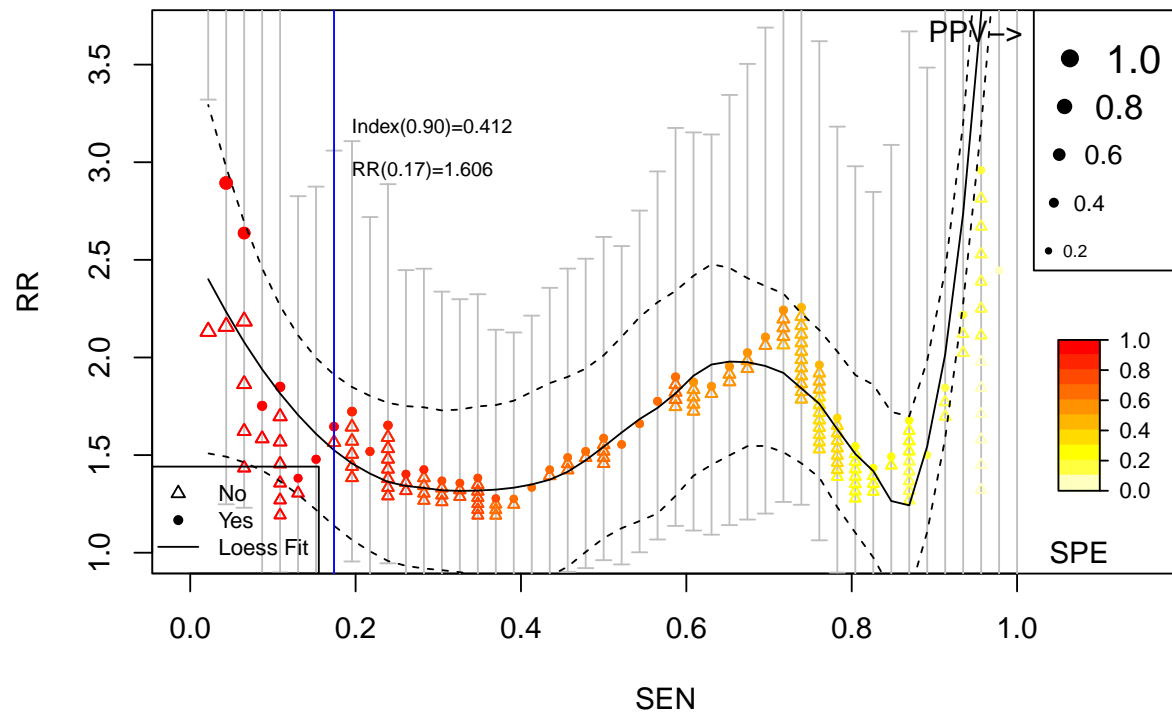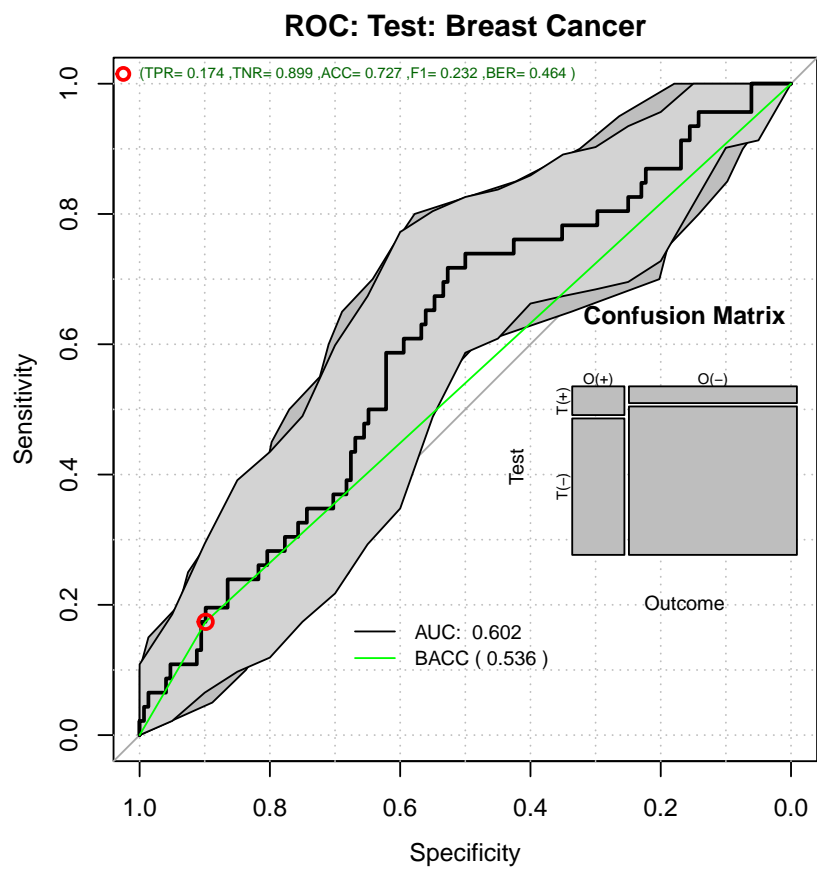
Cumulative vs. Observed: Test: Breast Cancer

Decision Curve Analysis: Test: Breast Cancer
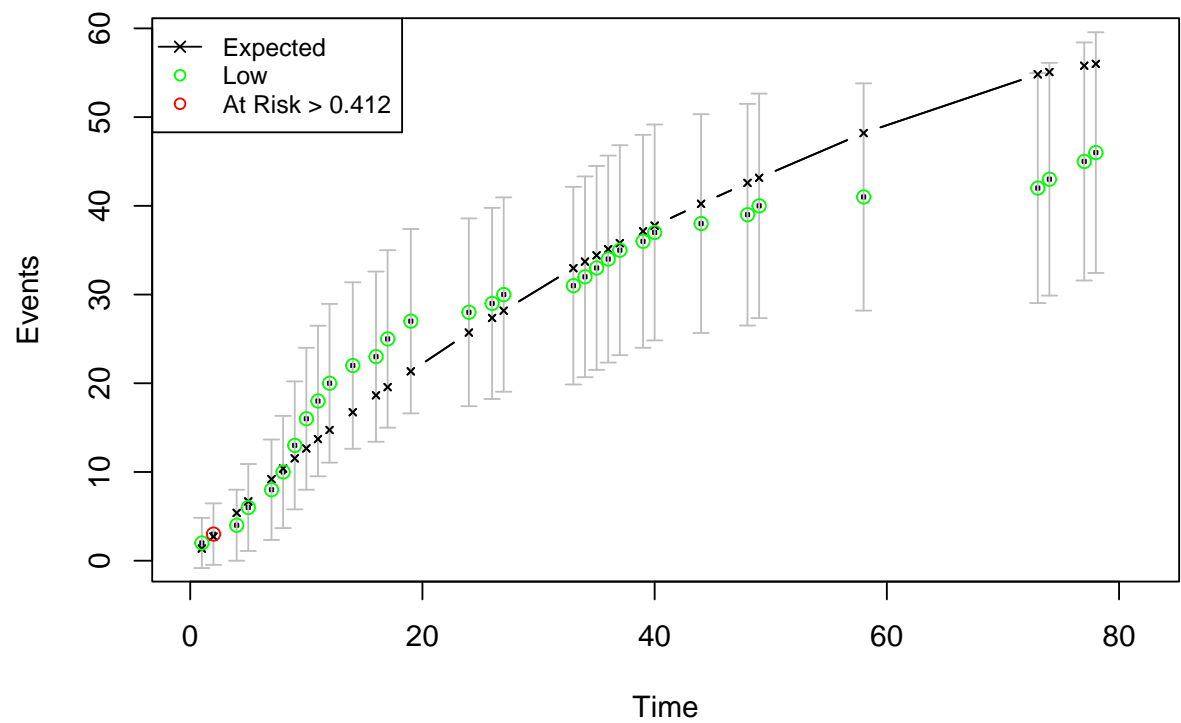
Relative Risk: Test: Breast Cancer

ROC: Test: Breast Cancer
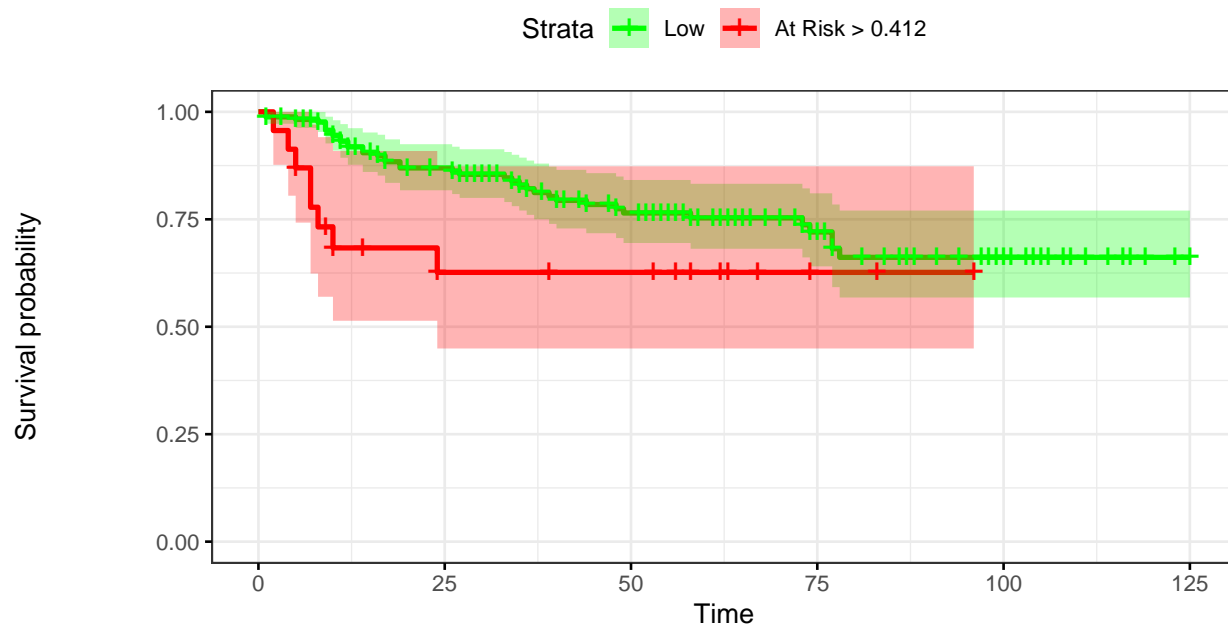

(TPR= 0.174 ,TNR= 0.899 ,ACC= 0.727 ,F1= 0.232 ,BER= 0.464)

AUC: 0.602
BACC (0.536)

Confusion Matrix

Sensitivity

Specificity

Test

Outcome

O(+) O(−)

T(+)

T(−)

Time vs. Events: Test: Breast Cancer

Kaplan–Meier: Test: Breast Cancer

Strata ✛ Low ✛ At Risk > 0.412

Number at risk

Low	171	117	77	42	20	1
At Risk > 0.412	23	10	9	2	0	0

1.4.1 Cross-Validation Test Performance

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

Table 26: Threshold values

	@:0.412	@MAX_BACC	@MAX_RR	@SPE100	p(0.5)
Thr	0.41205	0.237	0.156	1.44e-01	0.5032
RR	1.72297	2.243	2.959	2.24e+01	1.7524
RR_LCI	0.95517	1.260	0.768	4.69e-02	0.7841
RR_UCI	3.10798	3.992	11.398	1.07e+04	3.9166
SEN	0.19565	0.717	0.957	1.00e+00	0.0870
SPE	0.89865	0.527	0.142	6.08e-02	0.9595
BACC	0.54715	0.622	0.549	5.30e-01	0.5232
NetBenefit	-0.00779	0.058	0.106	1.17e-01	-0.0107

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

Table 27: O/E Ratio

O/E	Low	Upper	p.value
0.821	0.601	1.1	0.204

25

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

Table 28: O/E Mean

mean	50%	2.5%	97.5%
1.02	1.02	0.961	1.09

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

Table 29: O/Acum Mean

mean	50%	2.5%	97.5%
0.892	0.892	0.874	0.907

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

mean.C Index	median	lower	upper
0.654	0.655	0.573	0.74

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

Table 31: ROC AUC

est	lower	upper
0.602	0.51	0.695

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

Table 32: Sensitivity

est	lower	upper
0.174	0.0782	0.314

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

Table 33: Specificity

est	lower	upper
0.899	0.838	0.942

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

Table 34: Probability Thresholds

90%
0.412

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

Table 35: Logrank test Chisq = 3.859070 on 1 degrees of freedom,
p = 0.049478

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	171	38	41.82	0.348	3.86
class=1	23	8	4.18	3.480	3.86

1.4.2 Calibrating the test results

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

(48.90536 , 146.2793 , 0.9564824 , 46 , 51.51018)

```
pander::pander(c(h0=calprob$h0,
                Gain=calprob$hazardGain,
                DeltaTime=calprob$timeInterval),
            caption="Cox Calibration Parameters")
```
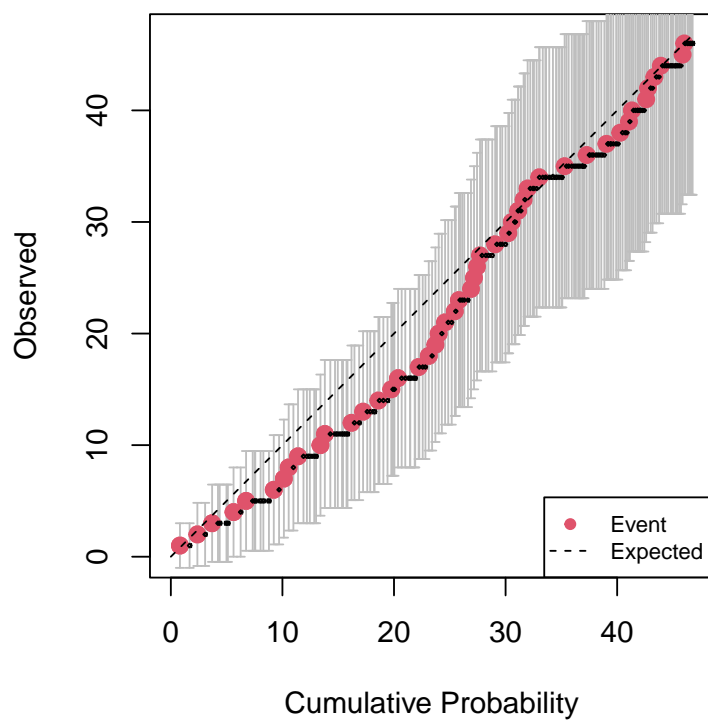
h0	Gain	DeltaTime
0.326	1.01	48.9

```
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)
```
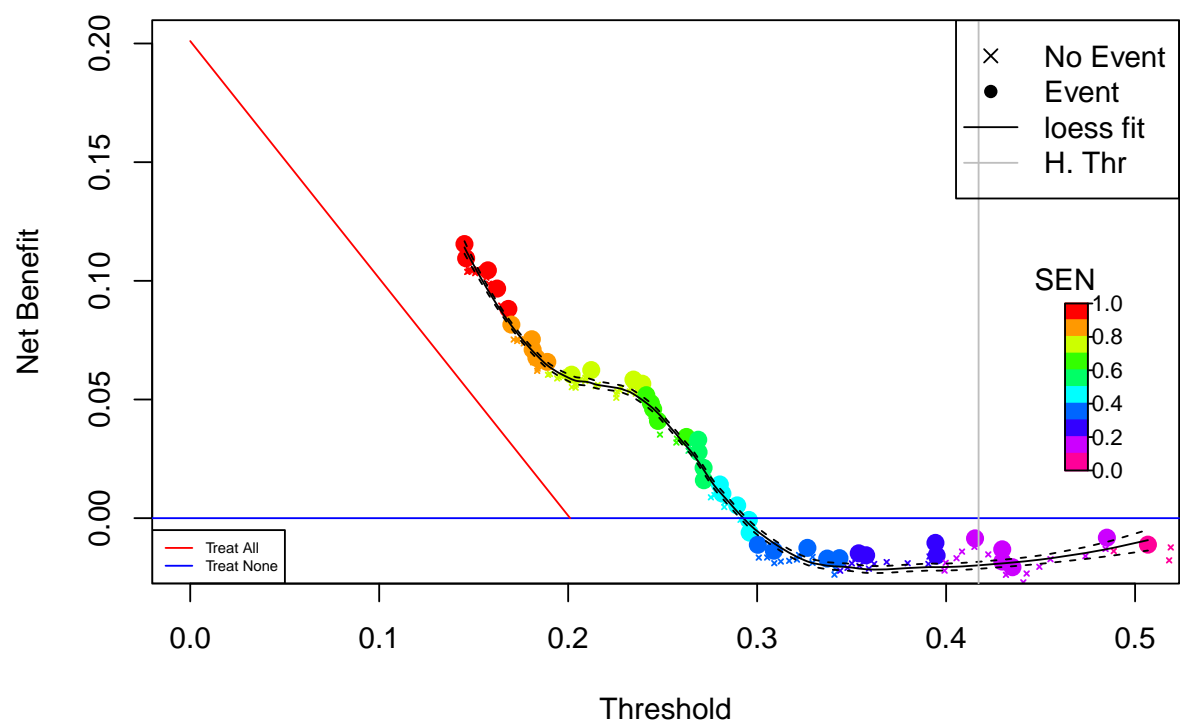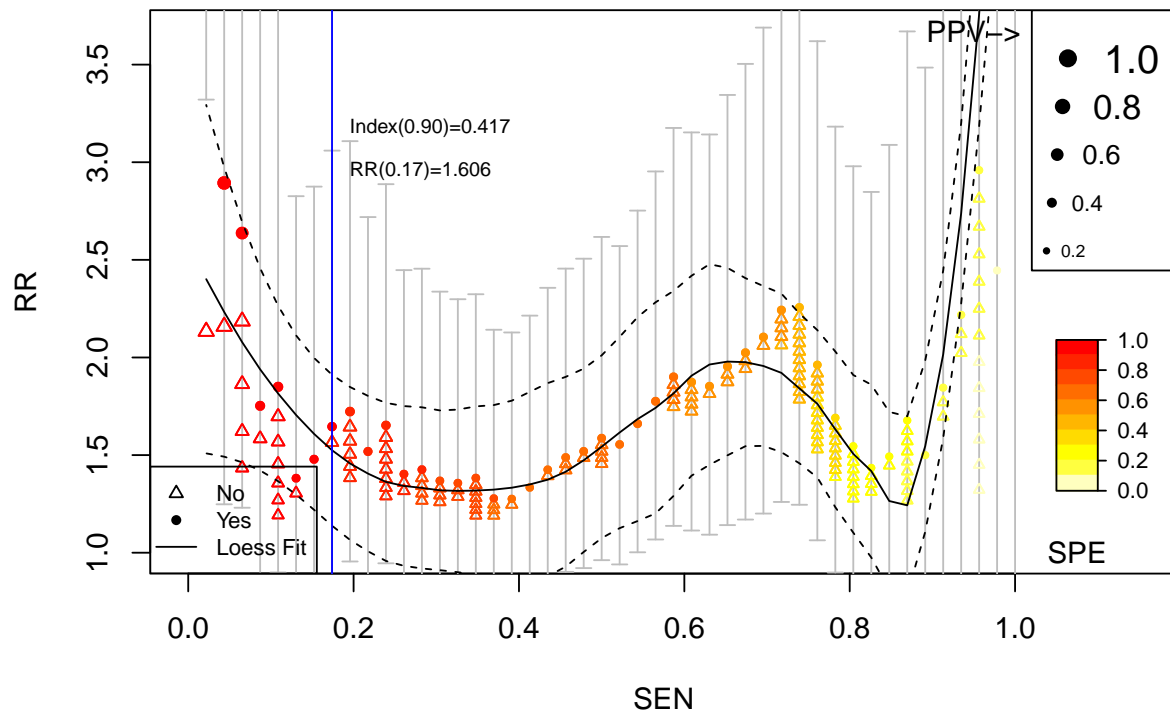
Cumulative vs. Observed: Calibrated Test: Breast

Decision Curve Analysis: Calibrated Test: Breast
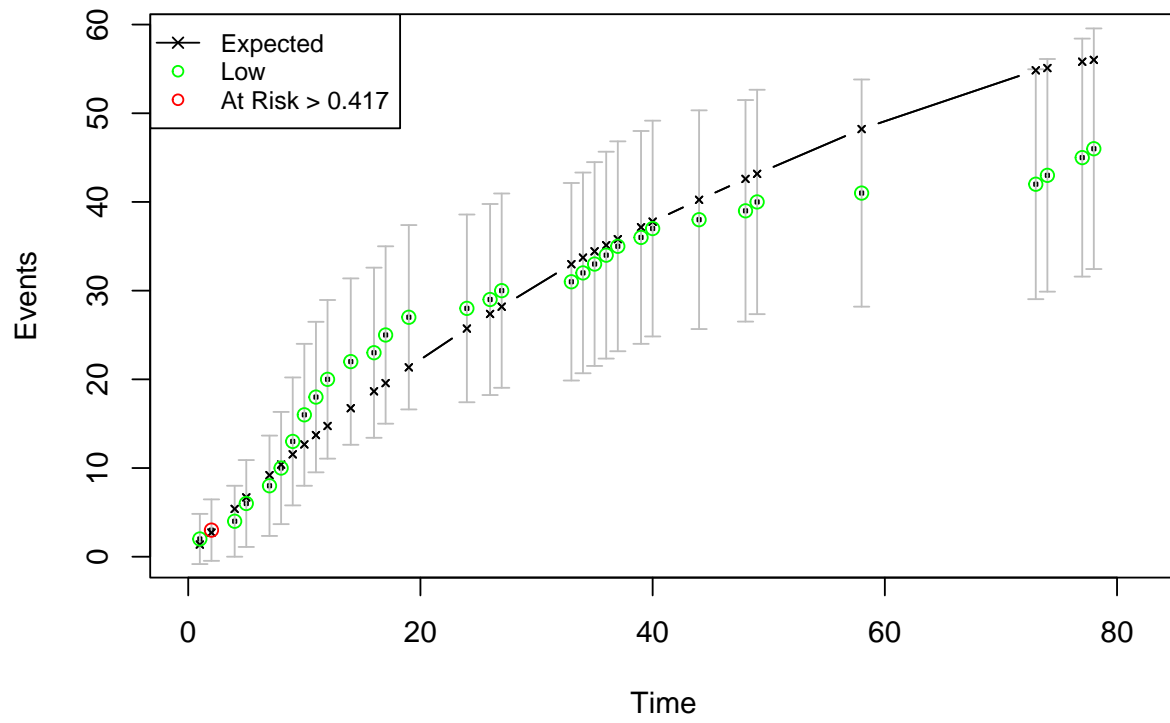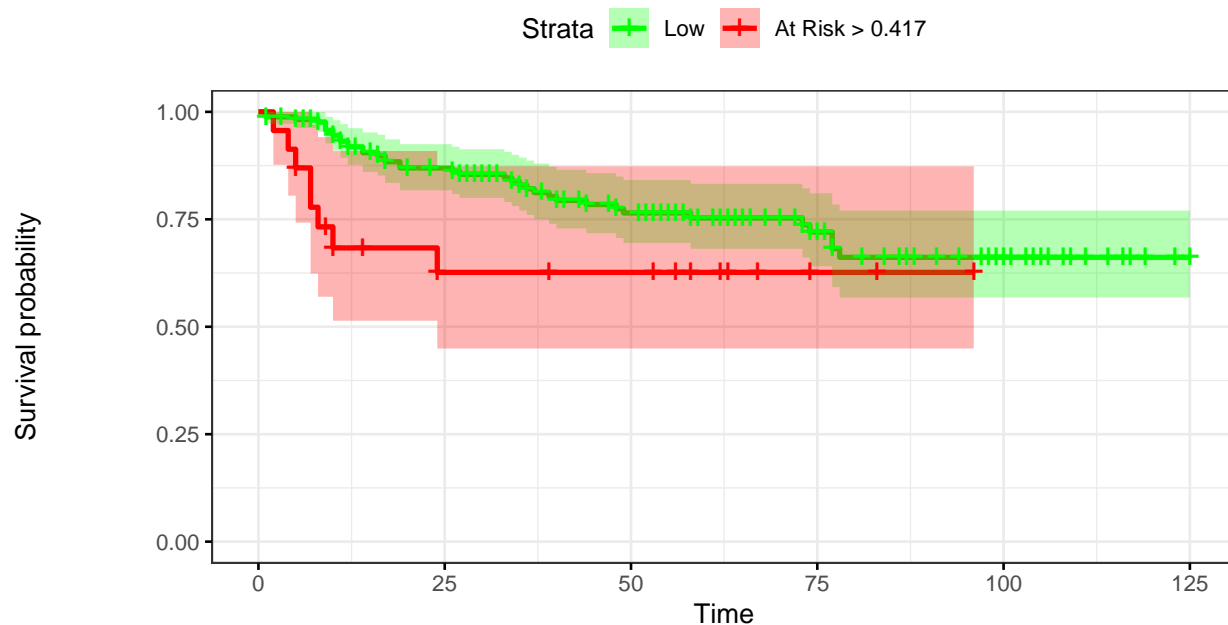
Relative Risk: Calibrated Test: Breast

ROC: Calibrated Test: Breast

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

Confusion Matrix

AUC: 0.602

BACC (0.536)
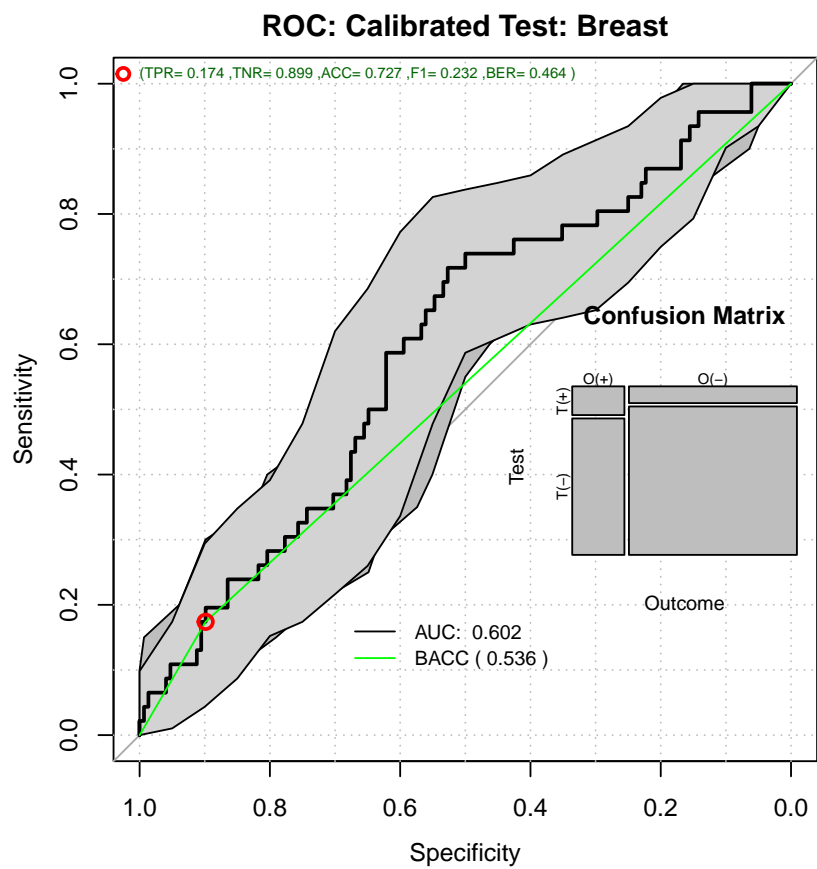
Time vs. Events: Calibrated Test: Breast

Kaplan–Meier: Calibrated Test: Breast

Strata ✚ Low ✚ At Risk > 0.417

Number at risk

Low	171	117	77	42	20	1
At Risk > 0.417	23	10	9	2	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	0.41530	0.2392	0.158	1.45e-01	0.5068
RR	1.72297	2.2427	2.959	2.24e+01	1.7524
RR_LCI	0.95517	1.2600	0.768	4.69e-02	0.7841
RR_UCI	3.10798	3.9918	11.398	1.07e+04	3.9166
SEN	0.19565	0.7174	0.957	1.00e+00	0.0870
SPE	0.89865	0.5270	0.142	6.08e-02	0.9595
BACC	0.54715	0.6222	0.549	5.30e-01	0.5232
NetBenefit	-0.00852	0.0567	0.104	1.15e-01	-0.0112

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

Table 38: O/E Ratio

O/E	Low	Upper	p.value
0.821	0.601	1.1	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	0.961	1.09

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

Table 40: O/Acum Mean

mean	50%	2.5%	97.5%
0.888	0.888	0.871	0.905

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

mean.C Index	median	lower	upper
0.654	0.655	0.572	0.741

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

Table 42: ROC AUC

est	lower	upper
0.602	0.51	0.695

```
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%
0.417

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

Table 46: Logrank test Chisq = 3.859070 on 1 degrees of freedom,
p = 0.049478

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	171	38	41.82	0.348	3.86
class=1	23	8	4.18	3.480	3.86