Breast Cancer: Wisconsin

Jose Tamez

2023-04-28

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
V26	1.24e-02	1	1.01	1.02	0.593	0.237
V27	2.04e-04	1	1.00	1.00	0.608	0.727
V24	1.18e-02	1	1.01	1.02	0.598	0.634
V7	1.59e-07	1	1.00	1.00	0.588	0.237
V35	1.40e-02	1	1.01	1.03	0.727	0.608
V34	1.40e-02	1	1.01	1.03	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	0.613	0.608	0.641	0.597	0.0562	0.447	2.72
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	0.613	0.641	0.608	0.597	0.0288	0.551	2.27
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	2.72	-0.04436	1
V24	1.94	-0.00529	1
V7	2.30	0.09489	1
V35	3.41	-0.01160	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()

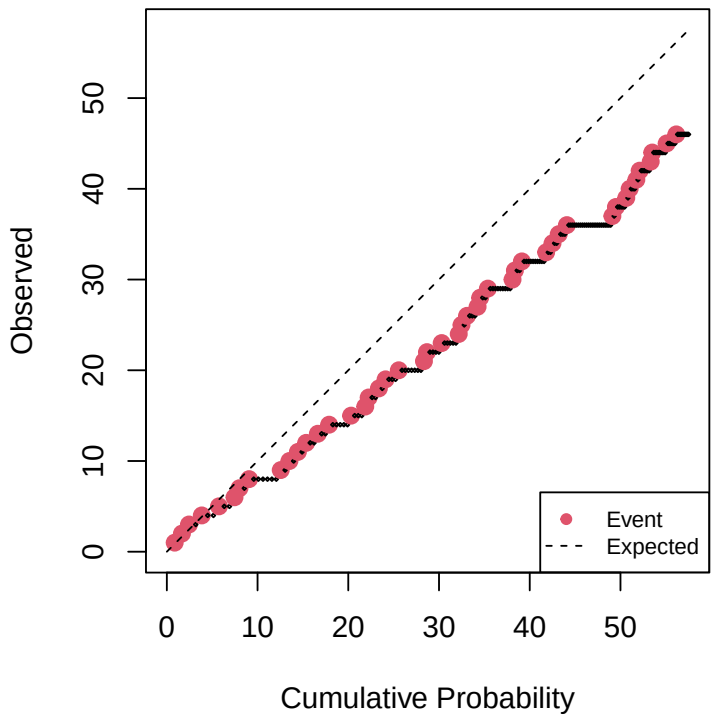
Here we will use the predicted event probability assuming a baseline hazard for events withing 5 years

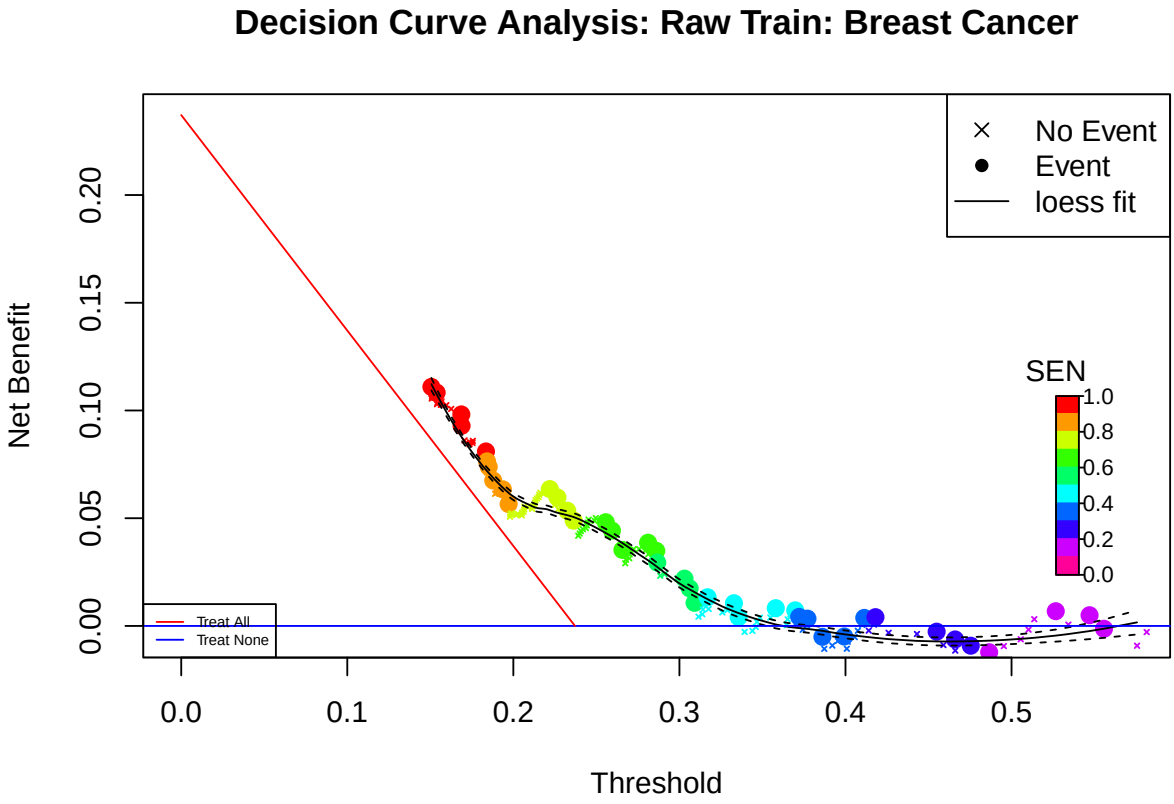
```
index <- predict(ml,dataBreast)
timeinterval <- 2*mean(subset(dataBreast,status==1)$time)

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")
```

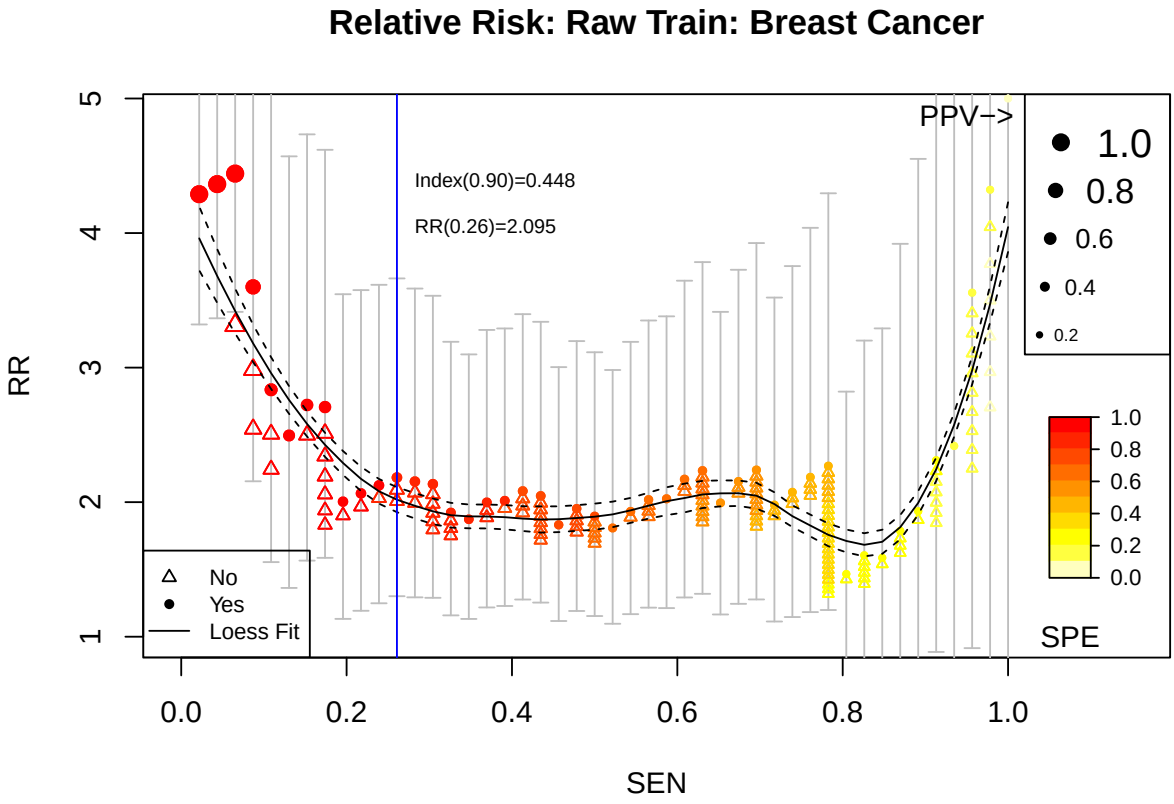
Table 4: Initial Parameters

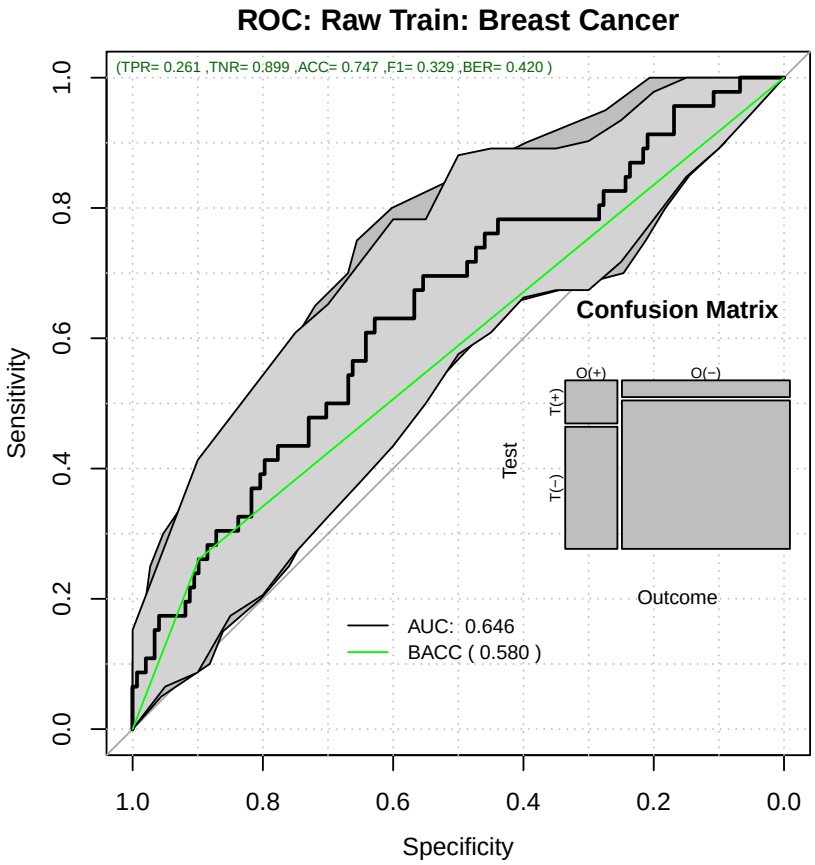
h0	timeinterval
0.323	51.1

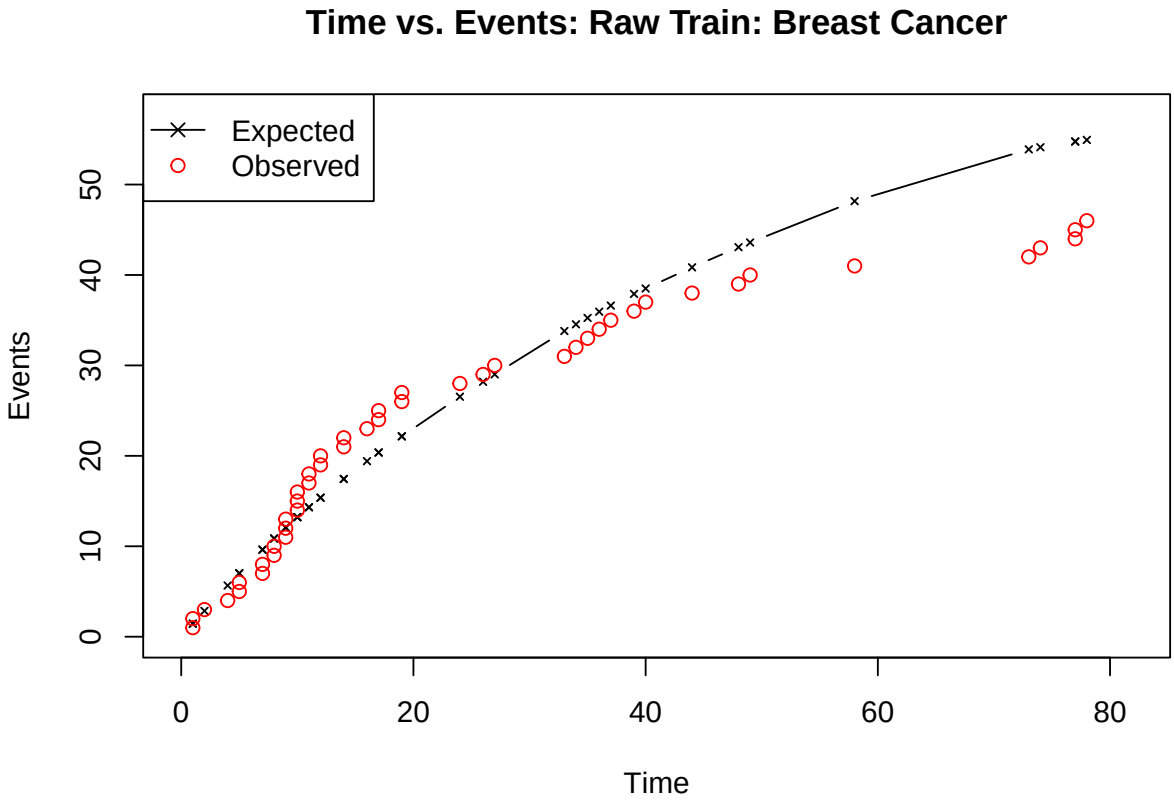
```
rdata <- cbind(dataBreast$status,ppoisGzero(index,h0))
rownames(rdata) <- rownames(dataBreast)

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

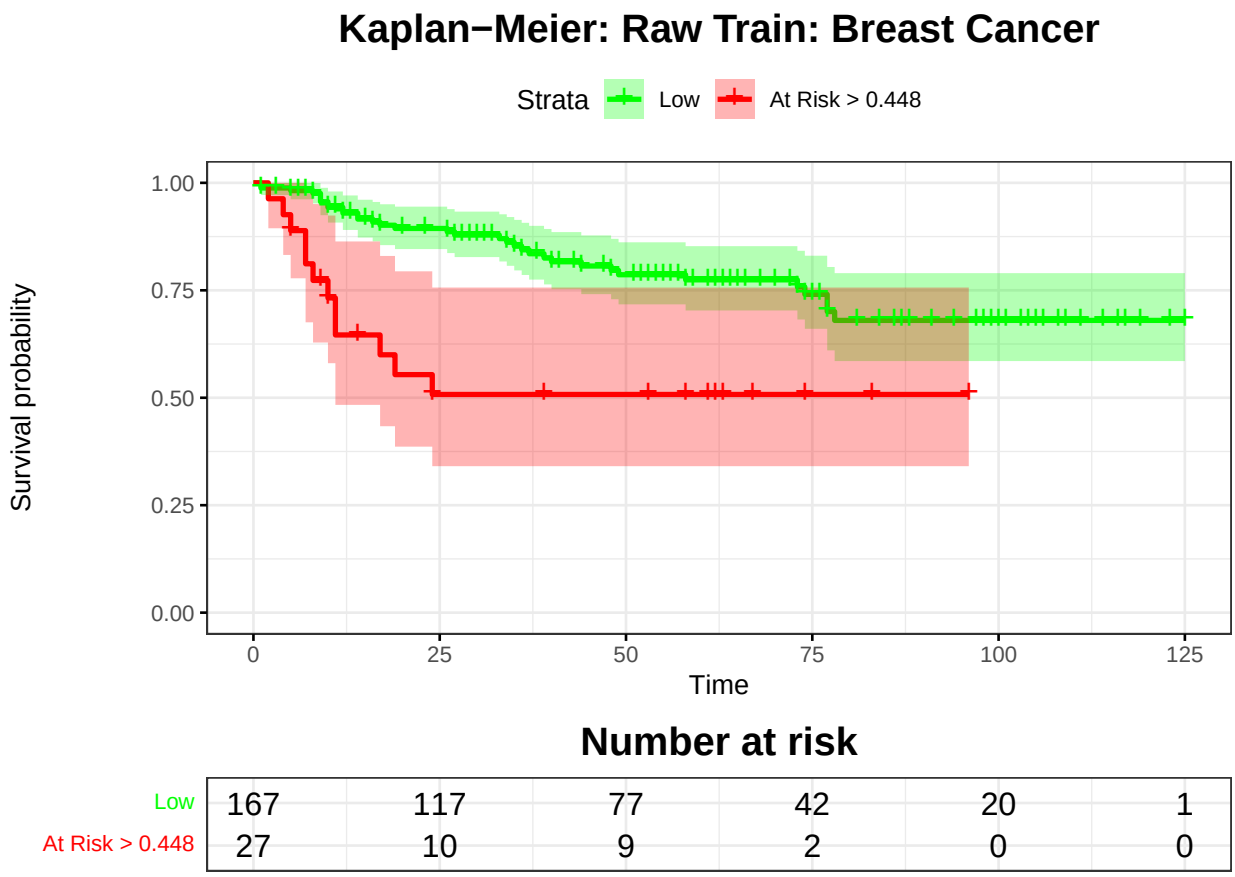

Cumulative vs. Observed: Raw Train: Breast Cancer
Observed
0
10
20
30
40
50
Cumulative Probability
0
10
20
30
40
50
Event
Expected


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–>
Index(0.90)=0.448
RR(0.26)=2.095
1.0
0.8
0.6
0.4
0.2
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: 0.646
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.

1.3.2 Uncalibrated Performance Report

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

Table 5: O/E Ratio

est	lower	upper
0.837	0.613	1.12

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

Table 6: O/E Ratio

mean	50%	2.5%	97.5%
0.995	0.996	0.947	1.04

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

Table 7: O/Acum Ratio

mean	50%	2.5%	97.5%
0.782	0.782	0.775	0.79

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

mean.C Index	median	lower	upper
0.686	0.687	0.603	0.762

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

Table 9: ROC AUC

est	lower	upper
0.646	0.553	0.739

```
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%
0.448

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

Table 13: Risk Ratio

est	lower	upper
2.09	1.24	3.53

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

Table 14: Logrank test Chisq = 12.659046 on 1 degrees of freedom, p = 0.000374

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	167	34	41.28	1.28	12.7
class=1	27	12	4.72	11.24	12.7

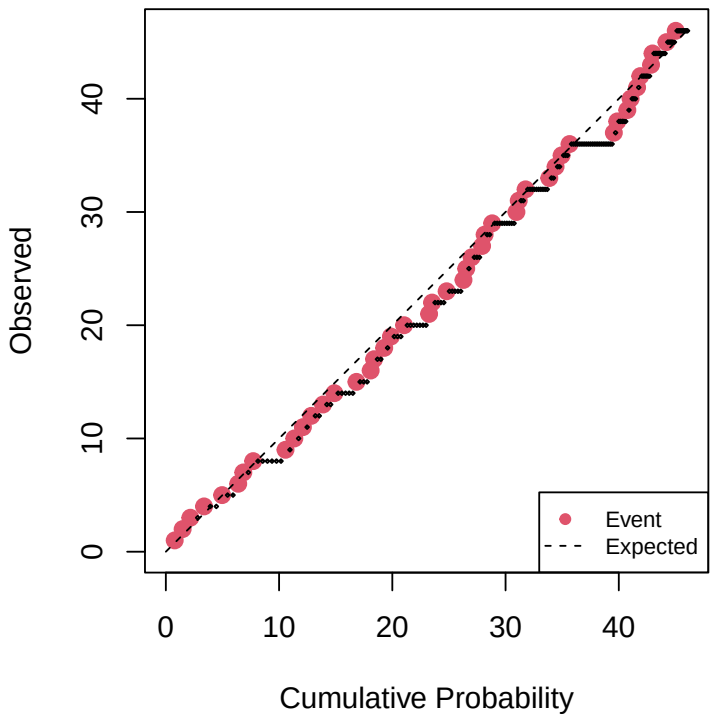
1.3.3 Cox Calibration

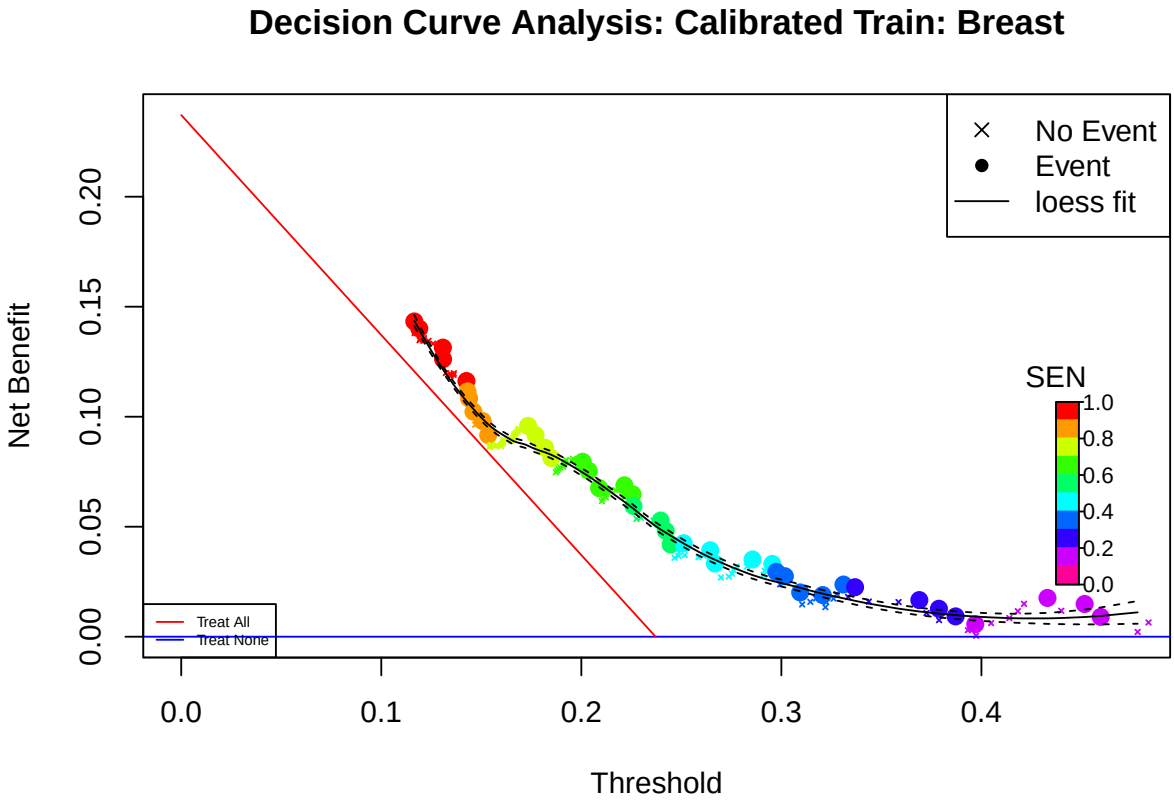
```
op <- par(no.readonly = TRUE)


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

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

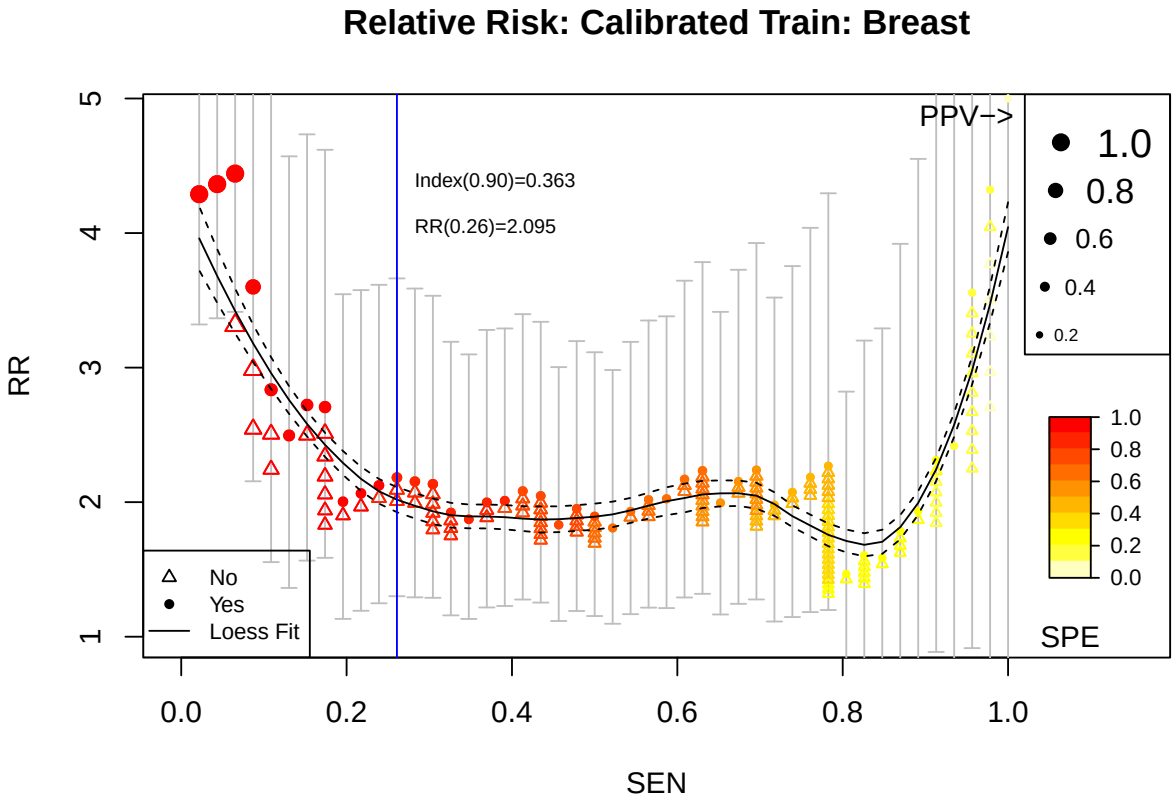
h0	Gain	DeltaTime
0.245	0.759	40.8

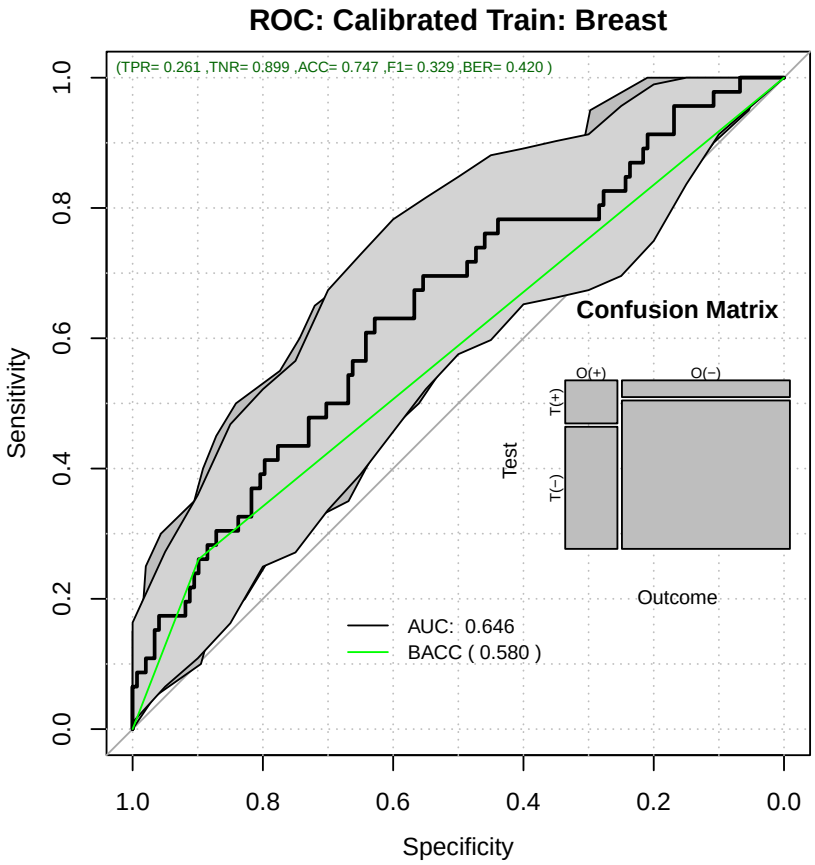
1.3.4 The RRplot() of the calibrated model

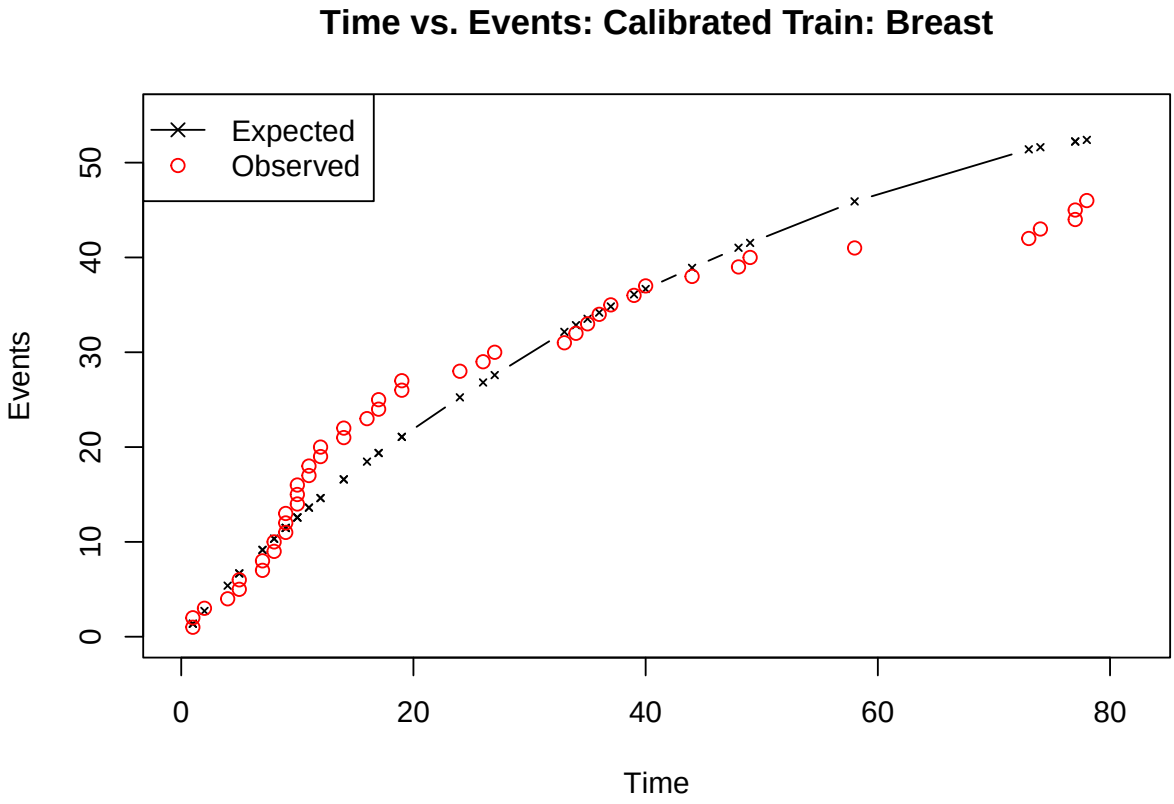
```
h0 <- calprob$h0
timeinterval <- calprob$timeInterval;

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


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

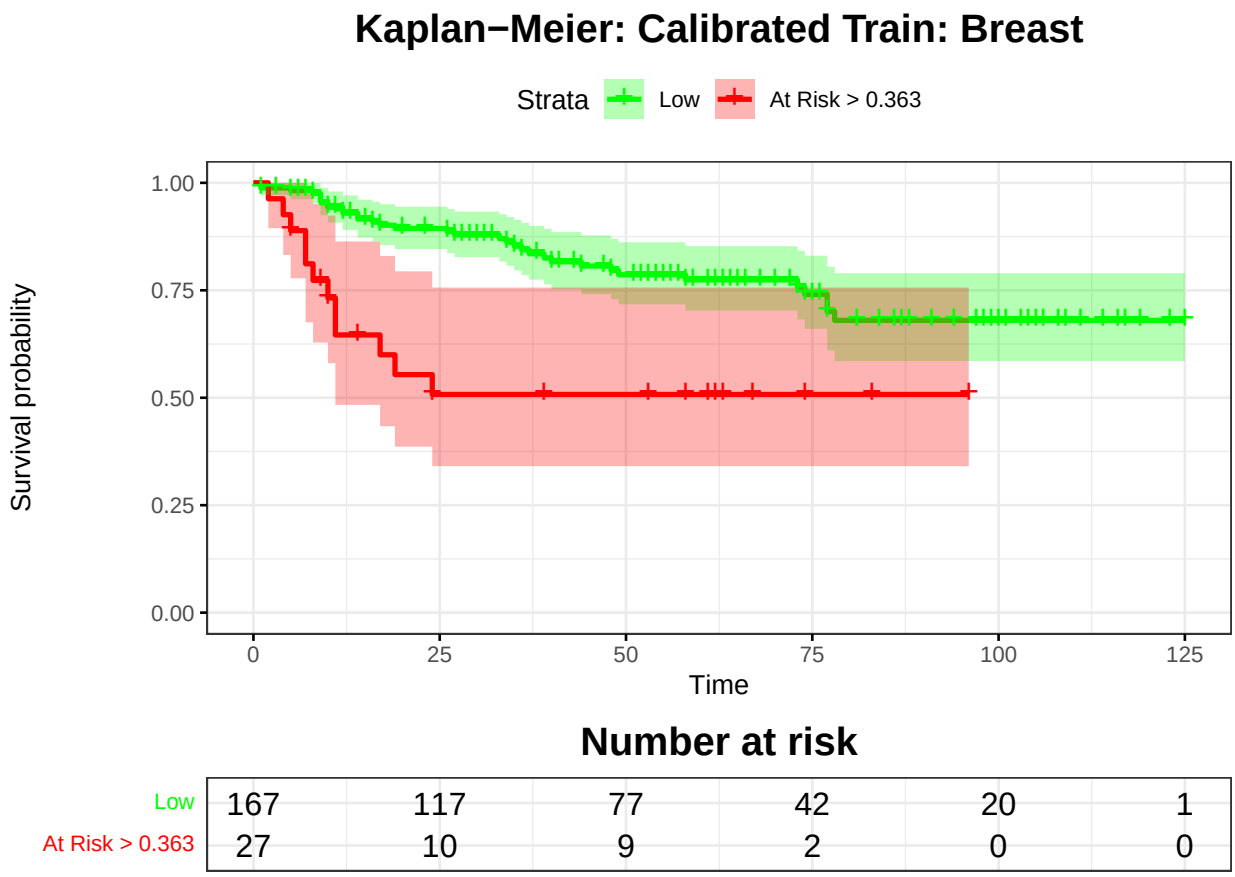

Cumulative vs. Observed: Calibrated Train: Breast
Observed
0
10
20
30
40
Cumulative Probability
0
10
20
30
40
Event
Expected


Decision Curve Analysis: Calibrated Train: Breast
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
Threshold


Relative Risk: Calibrated Train: Breast
PPV–>
Index(0.90)=0.363
RR(0.26)=2.095
1.0
0.8
0.6
0.4
0.2
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: Calibrated Train: Breast
(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: 0.646
BACC (0.580)


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



1.3.5 Calibrated Train Performance

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

Table 16: O/E Ratio

est	lower	upper
0.878	0.643	1.17

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

Table 17: O/E Ratio

mean	50%	2.5%	97.5%
1.05	1.05	0.995	1.1

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

Table 18: O/Acum Ratio

mean	50%	2.5%	97.5%
0.955	0.955	0.948	0.962

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

Table 19: C. Index

mean.C Index	median	lower	upper
0.686	0.685	0.603	0.765

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

Table 20: ROC AUC

est	lower	upper
0.646	0.553	0.739

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

Table 21: Sensitivity

est	lower	upper
0.261	0.143	0.411

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

Table 22: Specificity

est	lower	upper
0.899	0.838	0.942

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

Table 23: Probability Thresholds

90%
0.363

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

Table 24: Risk Ratio

est	lower	upper
2.09	1.24	3.53

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

Table 25: Logrank test Chisq = 12.659046 on 1 degrees of freedom, p = 0.000374

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	167	34	41.28	1.28	12.7
class=1	27	12	4.72	11.24	12.7

1.4 Cross-Validation

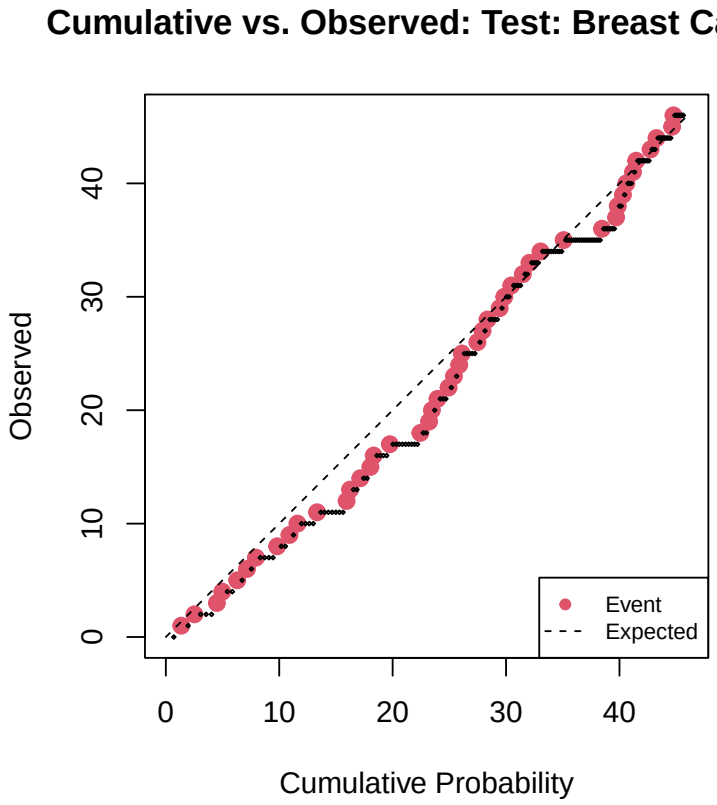
Here we use the estimated h0 and timeinterval from the full set

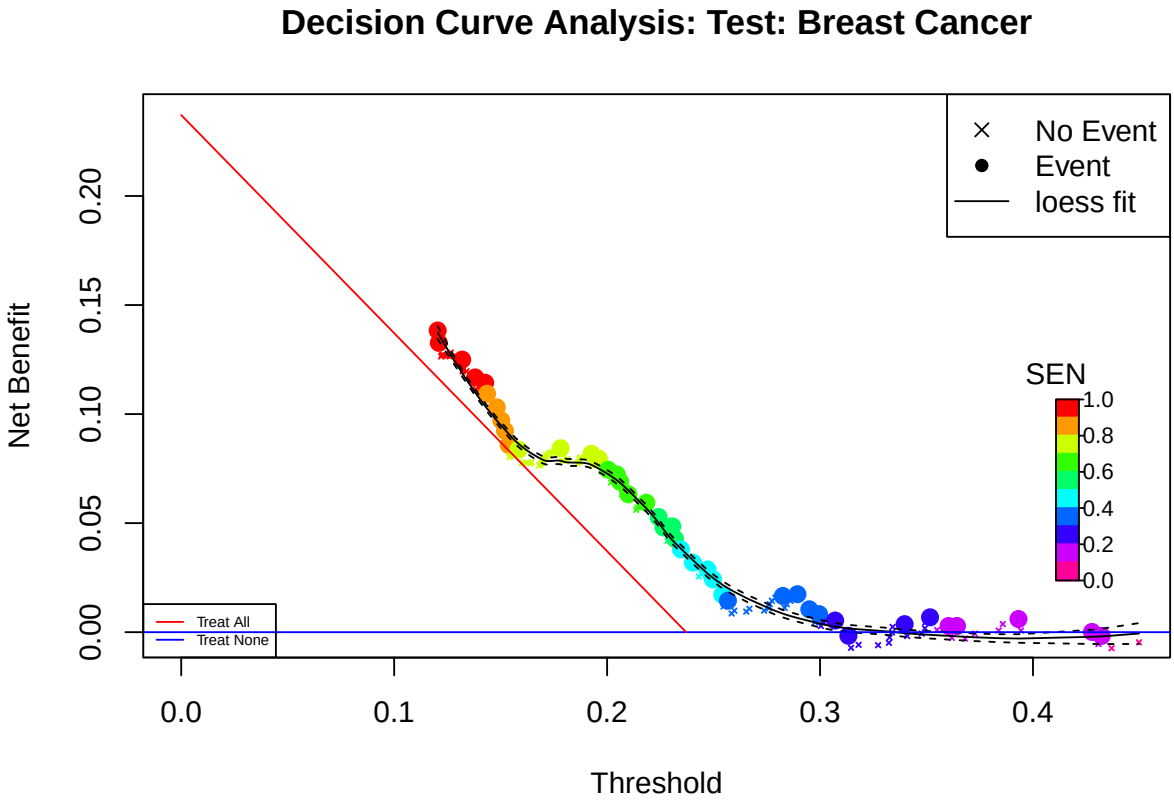
```
rcv <- randomCV(theData=dataBreast,
                theOutcome = Surv(time,status)~1,
                fittingFunction=BSWiMS.model,
                trainFraction = 0.9,
                repetitions=100,
                classSamplingType = "Pro"
        )
```

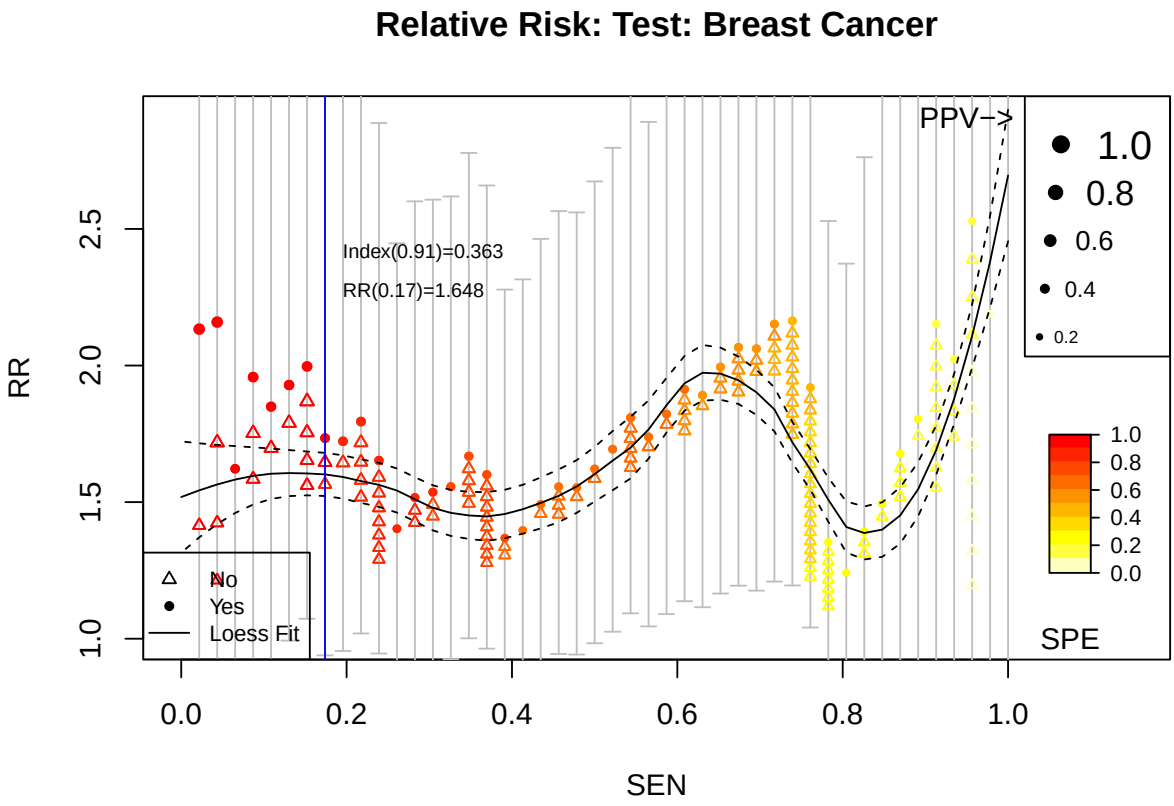
.[+++].[+++].[+++++].[+++].[+++].[+].[-++++].[+++++].[++++].[++++]10 Tested: 128 Avg. Selected: 4.1 Min Tests: 1 Max Tests: 4 Mean Tests: 1.5625 . MAD: 0.492001 .[+++].[+++].[++++].[+].[+].[++].[+++].[+++].[+
Tested: 173 Avg. Selected: 3.9 Min Tests: 1 Max Tests: 6 Mean Tests: 2.312139 . MAD: 0.4917073
.[++++].[+++++].[+].[++++].[++++++].[++++++].[+++++++].[++++].[++++].[+++++]30
Tested: 186 Avg. Selected: 4.133333 Min Tests: 1 Max Tests: 8 Mean Tests: 3.225806 . MAD: 0.4921404
.[+++++].[++++++].[++++++].[++++].[+++++].[+++].[++++++].[+++++].[+].[++++]40
Tested: 187 Avg. Selected: 4.325 Min Tests: 1 Max Tests: 10 Mean Tests: 4.278075 . MAD: 0.4883659
.[++++].[+++].[-].[+++++++++].[++++++++].[+++].[+++].[+++].[++++++].[++++++]50 Tested: 191 Avg. Selected: 4.4 Min Tests: 1 Max Tests: 13 Mean Tests: 5.235602 . MAD: 0.4852724 .[-].[++++++++].[-].[+++].[++++++].[+++].[++++++++].[+++++].[++++++].[-]60 Tested: 193 Avg. Selected: 4.316667 Min Tests: 1 Max Tests: 14 Mean Tests: 6.217617 . MAD: 0.4827781
.[++++].[++++++].[++].[++++].[++++].[+++++].[+++++++++].[+++].[++++].[++++++]70
Tested: 194 Avg. Selected: 4.371429 Min Tests: 1 Max Tests: 17 Mean Tests: 7.216495 . MAD: 0.4835372 .[+++].[+++++++++].[++++++].[+++++].[-].[++++].[+++++].[-].[++++].[+++]80
Tested: 194 Avg. Selected: 4.3875 Min Tests: 1 Max Tests: 18 Mean Tests: 8.247423 . MAD: 0.4819771
.[+++].[++++++].[+++++].[++++++++].[+++].[++++++++].[+++++++++].[++++++++].[+++++].[++]90
Tested: 194 Avg. Selected: 4.533333 Min Tests: 1 Max Tests: 18 Mean Tests: 9.278351 . MAD: 0.4803538
.[+++].[++++].[++++++++].[+++++].[+++++].[+++].[+].[++++++].[++++].[+++++++++]100
Tested: 194 Avg. Selected: 4.55 Min Tests: 1 Max Tests: 18 Mean Tests: 10.30928 . MAD: 0.4831203

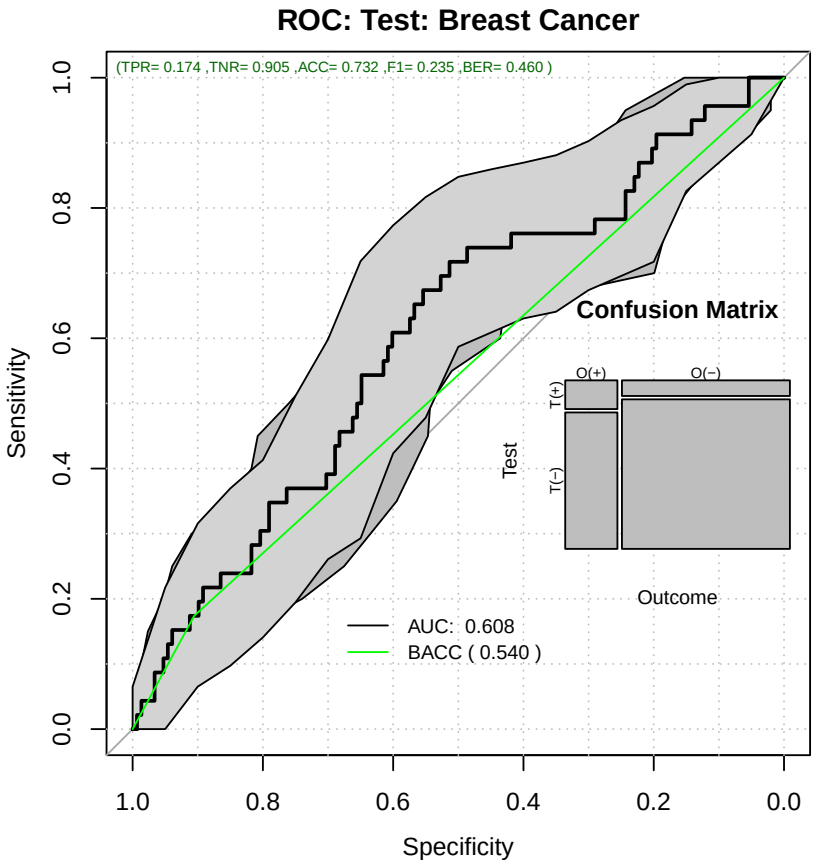
```
stp <- rcv$survTestPredictions
stp <- stp[!is.na(stp[,4]),]

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)
```

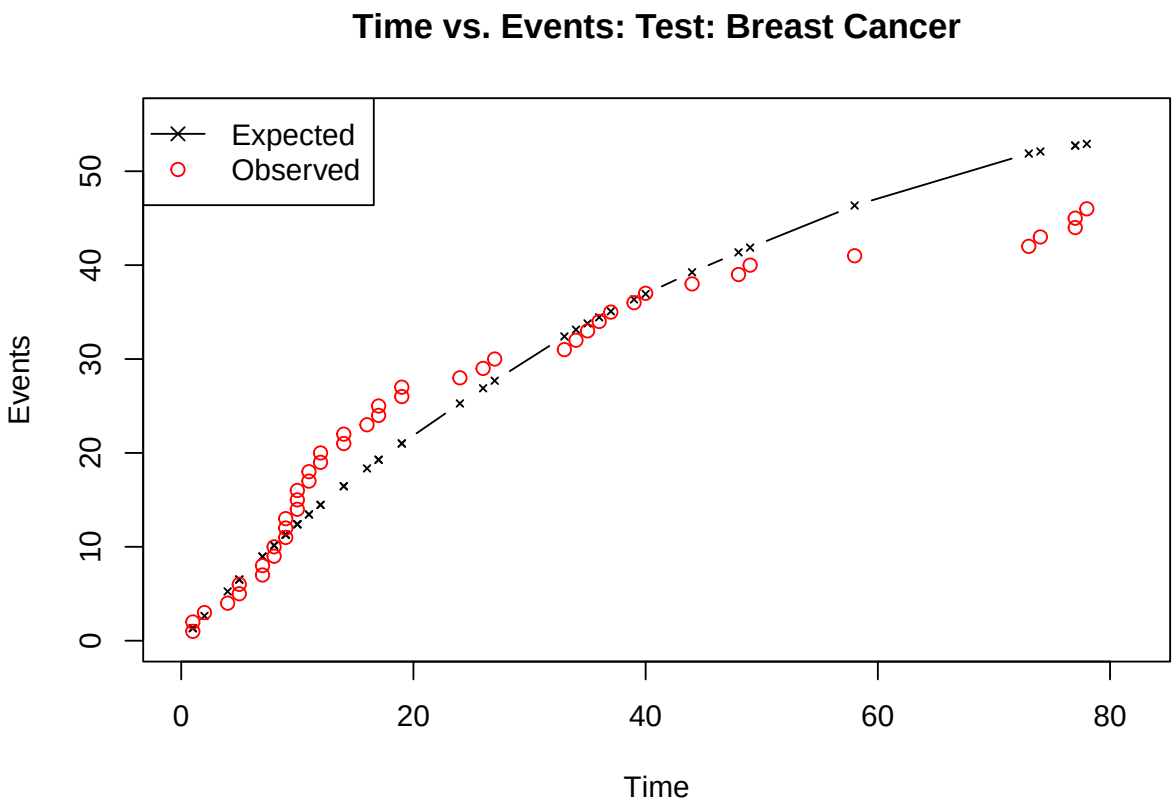
```
rdatacv <- cbind(status,prob)
rownames(rdatacv) <- bbx$names
names(times) <- bbx$names

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

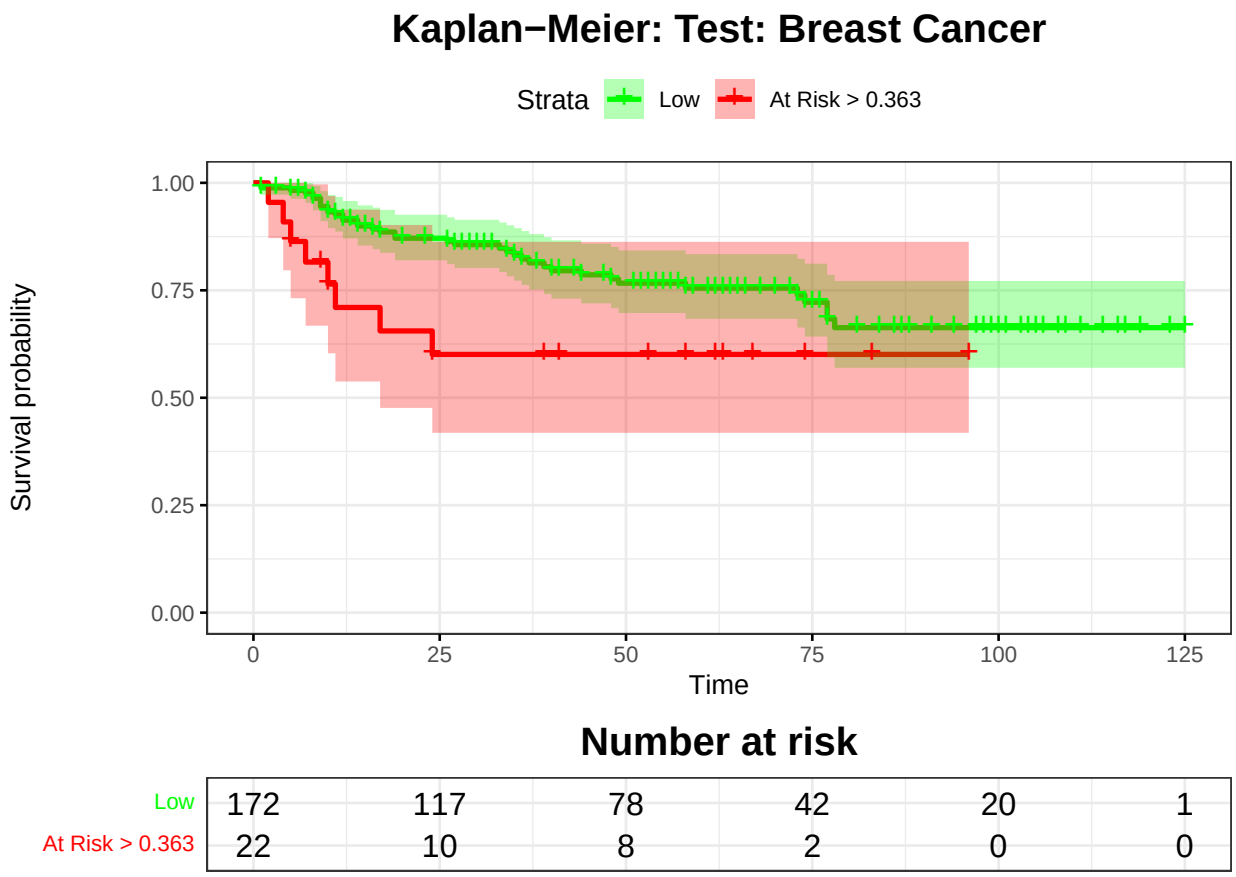



Decision Curve Analysis: Test: Breast Cancer
No Event
Event
loess fit
SEN
1.0
0.8
0.6
0.4
0.2
0.0
Treat All
Treat None
Net Benefit
0.00
0.05
0.10
0.15
0.20
0.0
0.1
0.2
0.3
0.4
Threshold


Relative Risk: Test: Breast Cancer
PPV–>
1.0
0.8
0.6
0.4
0.2
Index(0.91)=0.363
RR(0.17)=1.648
1.0
0.8
0.6
0.4
0.2
0.0
SPE
No
Yes
Loess Fit
RR
1.0
1.5
2.0
2.5
0.0
0.2
0.4
0.6
0.8
1.0
SEN


ROC: Test: Breast Cancer
(TPR= 0.174 ,TNR= 0.905 ,ACC= 0.732 ,F1= 0.235 ,BER= 0.460)
Sensitivity
Specificity
1.0
0.8
0.6
0.4
0.2
0.0
Confusion Matrix
O(+)
O(−)
T(+)
T(−)
Test
Outcome
AUC: 0.608
BACC (0.540)


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



1.4.1 Cross-Validation Test Performance

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

Table 26: O/E Ratio

est	lower	upper
0.869	0.636	1.16

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

Table 27: O/E Ratio

mean	50%	2.5%	97.5%
1.05	1.05	1	1.1

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

Table 28: O/Acum Ratio

mean	50%	2.5%	97.5%
0.917	0.917	0.903	0.929

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

mean.C Index	median	lower	upper
0.654	0.653	0.562	0.733

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

Table 30: ROC AUC

est	lower	upper
0.608	0.514	0.701

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

Table 31: Sensitivity

est	lower	upper
0.174	0.0782	0.314

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

Table 32: Specificity

est	lower	upper
0.905	0.846	0.947

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

Table 33: Probability Thresholds

90%
0.363

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

Table 34: Risk Ratio

est	lower	upper
1.65	0.887	3.06

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

Table 35: Logrank test Chisq = 3.877078 on 1 degrees of freedom, p = 0.048950

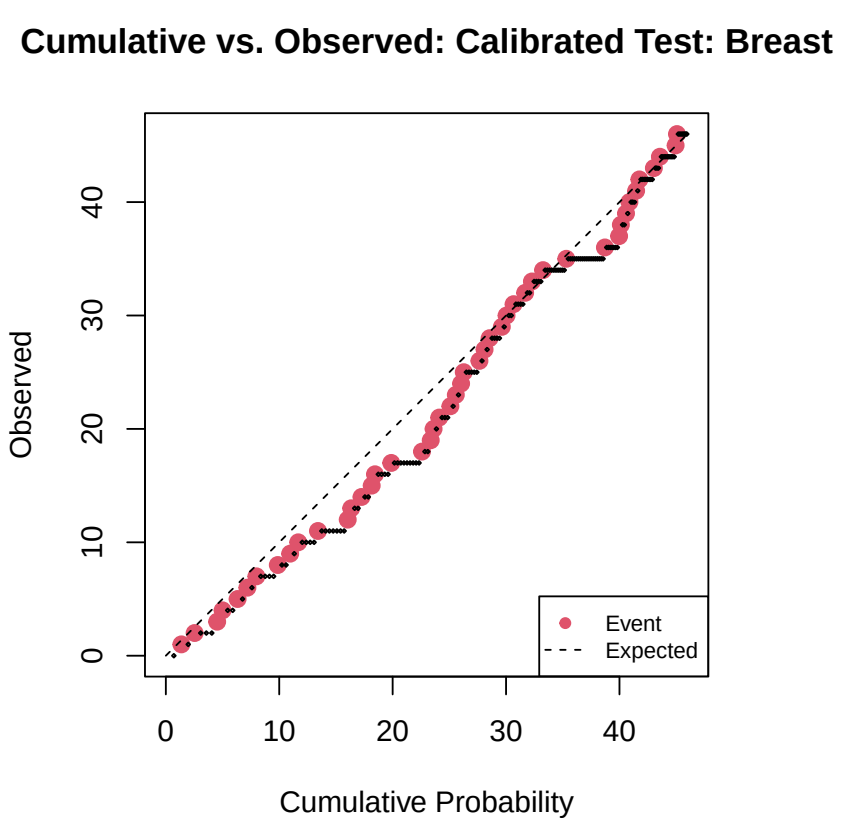
	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	172	38	41.82	0.349	3.88
class=1	22	8	4.18	3.499	3.88

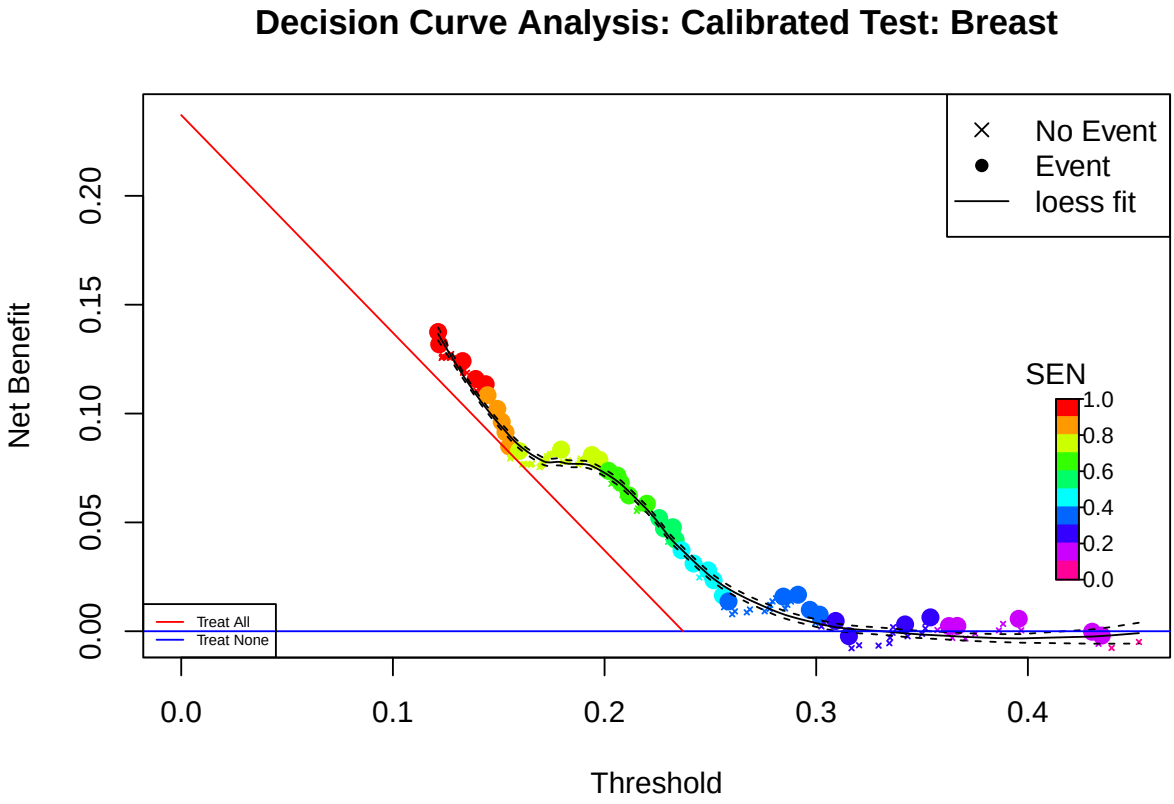
1.4.2 Calibrating the test results

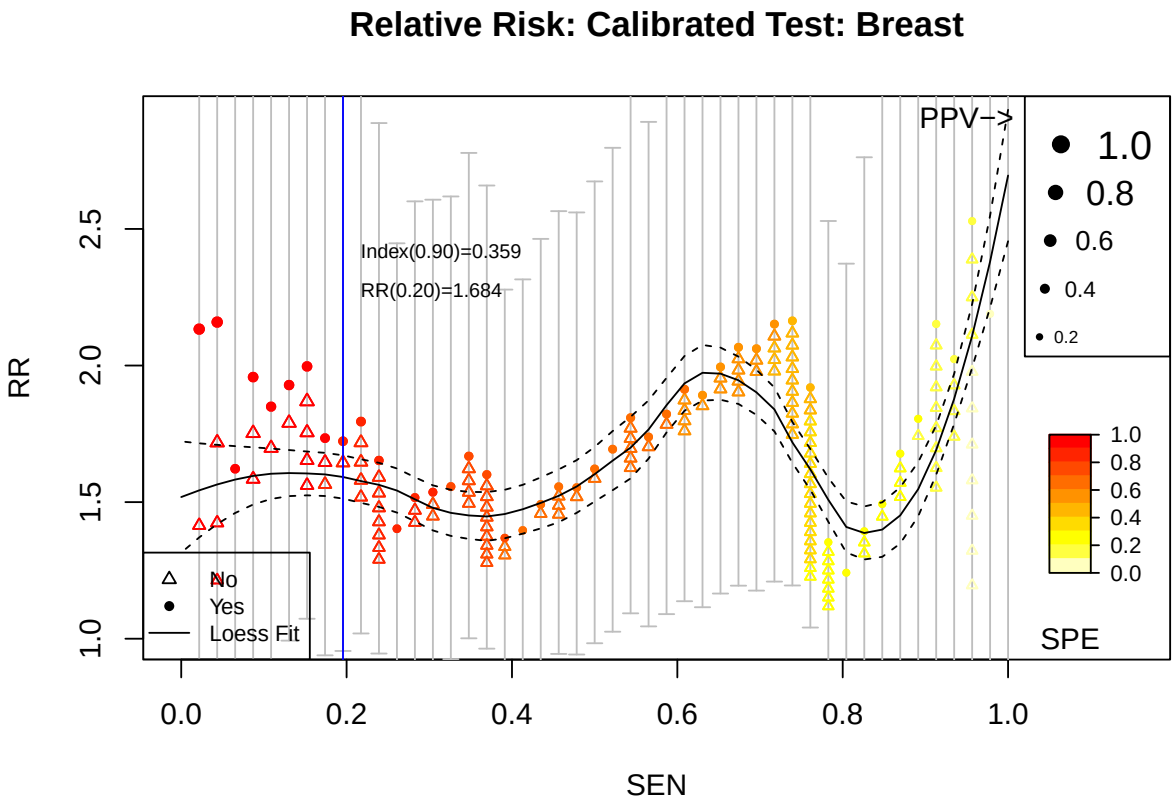
```
rdatacv <- cbind(status,prob,times)
calprob <- CalibrationProbPoissonRisk(rdatacv)

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

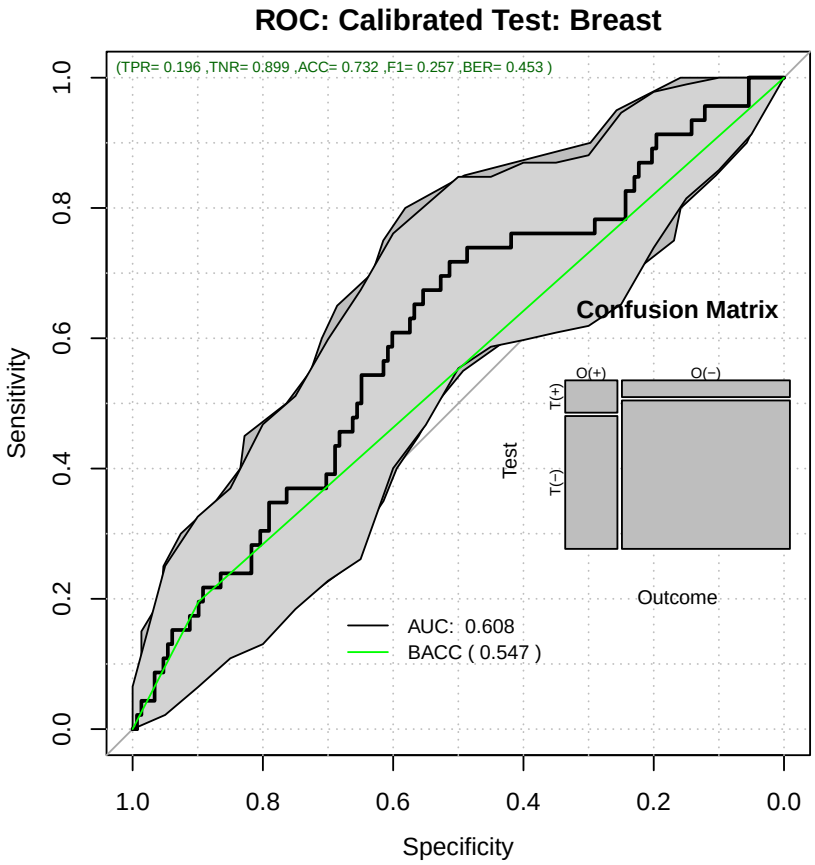
h0	Gain	DeltaTime
0.325	1.01	41.2

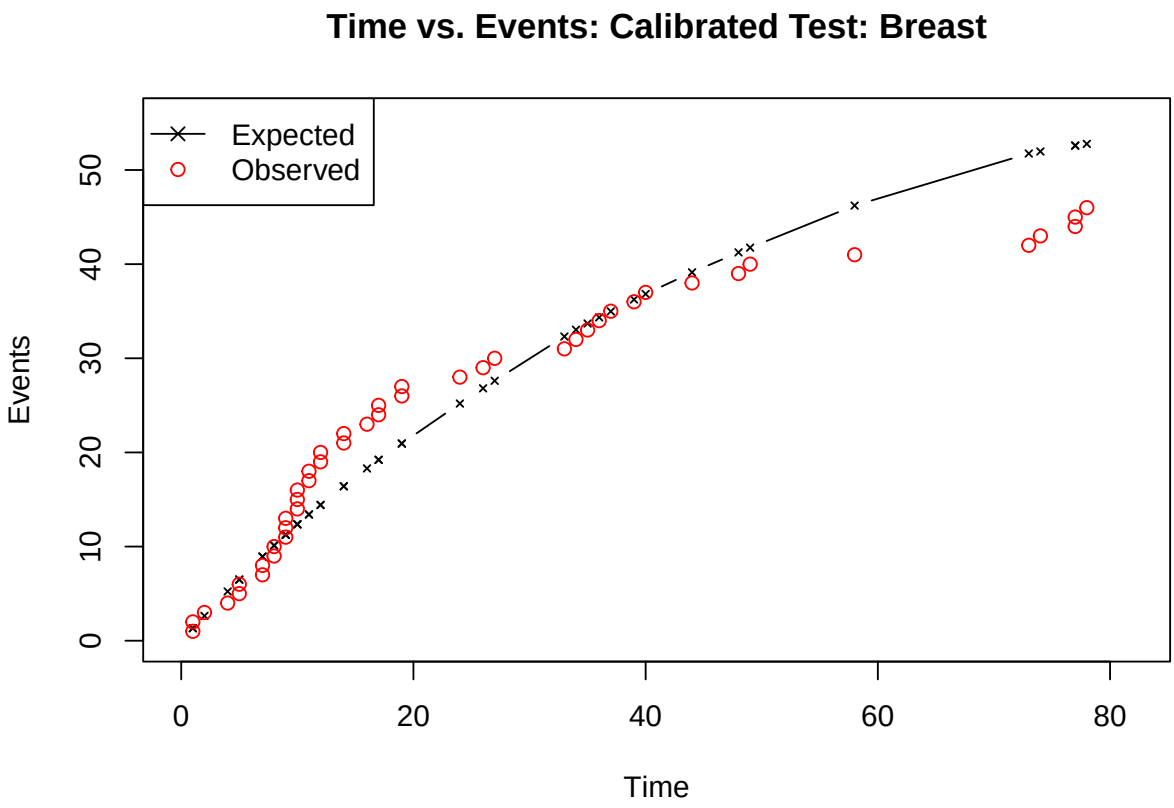
```
timeinterval <- calprob$timeInterval;

rdata <- cbind(status,calprob$prob)


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

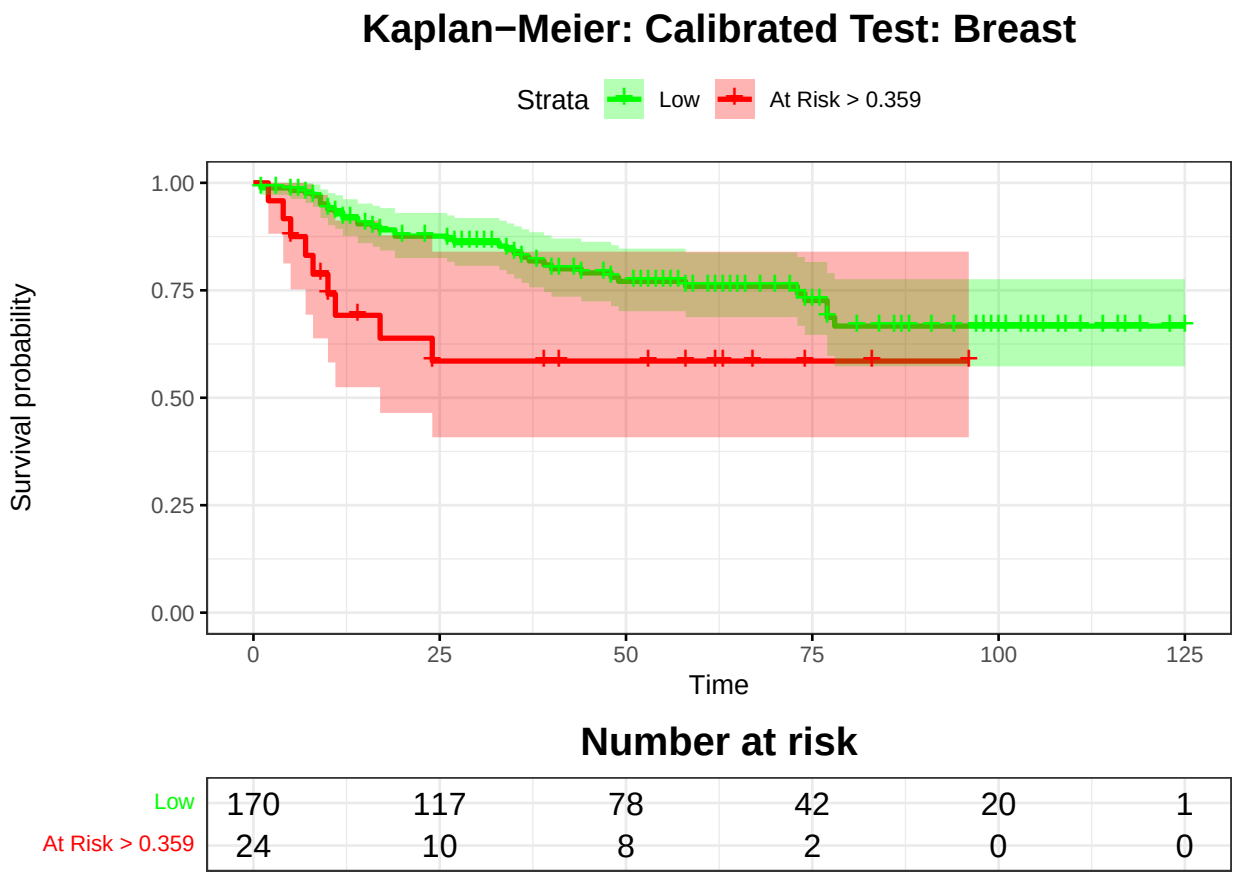

Cumulative vs. Observed: Calibrated Test: Breast
Observed
0
10
20
30
40
Cumulative Probability
0
10
20
30
40
Event
Expected


Decision Curve Analysis: Calibrated Test: Breast
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
Threshold


Relative Risk: Calibrated Test: Breast
PPV–>
1.0
0.8
0.6
0.4
0.2
Index(0.90)=0.359
RR(0.20)=1.684
No
Yes
Loess Fit
1.0
0.8
0.6
0.4
0.2
0.0
SPE
RR
2.5
2.0
1.5
1.0
0.0
0.2
0.4
0.6
0.8
1.0
SEN


ROC: Calibrated Test: Breast
(TPR= 0.196 ,TNR= 0.899 ,ACC= 0.732 ,F1= 0.257 ,BER= 0.453)
Sensitivity
Specificity
1.0
0.8
0.6
0.4
0.2
0.0
Confusion Matrix
O(+)
O(−)
T(+)
T(−)
Test
Outcome
AUC: 0.608
BACC (0.547)


Time vs. Events: Calibrated Test: Breast
Expected
Observed
Events
0
10
20
30
40
50
0
20
40
60
80
Time



Calibrated Test Performance

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

Table 37: O/E Ratio

est	lower	upper
0.872	0.638	1.16

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

Table 38: O/E Ratio

mean	50%	2.5%	97.5%
1.05	1.05	1	1.11

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

Table 39: O/Acum Ratio

mean	50%	2.5%	97.5%
0.911	0.911	0.898	0.924

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

mean.C Index	median	lower	upper
0.654	0.653	0.571	0.736

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

Table 41: ROC AUC

est	lower	upper
0.608	0.514	0.701

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

Table 42: Sensitivity

est	lower	upper
0.196	0.0936	0.339

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

Table 43: Specificity

est	lower	upper
0.899	0.838	0.942

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

Table 44: Probability Thresholds

90%
0.359

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

Table 45: Risk Ratio

est	lower	upper
1.68	0.931	3.04

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

Table 46: Logrank test Chisq = 5.501458 on 1 degrees of freedom, p = 0.019001

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	170	37	41.64	0.517	5.5
class=1	24	9	4.36	4.937	5.5