Risk-Evaluation: Breast Cancer Royston-Altman

Jose Tamez

2023-04-28

Contents

1 Evaluation of RISK survival models

This document highlights the use of

- RRPlot(),
- CoxRiskCalibration(), and
- CalibrationProbPoissonRisk(),

for the evaluation (RRPlot), and calibration of cox models (CoxRiskCalibration) or logistic models (CalibrationProbPoissonRisk) of survival data.

Furthermore, it can be used to evaluate any Risk index that reruns the probability of a future event on external data-set.

This document will use the survival::rotterdam, and survival::gbsg data-sets to train and predict the risk of cancer recurrence after surgery. Both Cox and Logistic models will be trained and evaluated.

Here are some sample plots returned by the evaluated functions:

1.1 The 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
```

```
#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)
```

1.2 Breast Cancer Royston-Altman data

1.2.1 data(gbsg, package="survival") and data(rotterdam, package="survival")

```
gbsgdata <- gbsg
rownames(gbsgdata) <- gbsgdata$pid
gbsgdata$pid <- NULL

odata <-rotterdam
rownames(odata) <- odata$pid
odata$pid <- NULL
odata$rfstime <- odata$rtime
odata$status <- odata$recur
odata$rtime <- NULL
odata$recur <- NULL

odata <- odata[,colnames(odata) %in% colnames(gbsgdata)]

odata$size <- 10*(odata$size=="<=20") +
  35*(odata$size=="20-50") +
  60*(odata$size==">50")

data <- as.data.frame(model.matrix(Surv(rfstime,status)~.*.,odata))

data$`(Intercept)` <- NULL

dataBrestCancerTrain <- cbind(time=odata[rownames(data),"rfstime"],status=odata[rownames(data),"status"

colnames(dataBrestCancerTrain) <-str_replace_all(colnames(dataBrestCancerTrain),":","_")
colnames(dataBrestCancerTrain) <-str_replace_all(colnames(dataBrestCancerTrain)," ","")
colnames(dataBrestCancerTrain) <-str_replace_all(colnames(dataBrestCancerTrain),"\\.","_")
colnames(dataBrestCancerTrain) <-str_replace_all(colnames(dataBrestCancerTrain),"-","_")
```

```
colnames(dataBrestCancerTrain) <-str_replace_all(colnames(dataBrestCancerTrain),">","_")
dataBrestCancerTrain$time <- dataBrestCancerTrain$time/365 ## To years


pander::pander(table(odata[rownames(data),"status"]),caption="rotterdam")
```

Table 1: rotterdam

0	1
1464	1518

1.2.2 data(gbsg, package="survival") data conditioning

```
gbsgdata <- gbsgdata[,colnames(odata)]
data <- as.data.frame(model.matrix(Surv(rfstime,status)~.*.,gbsgdata))

data$`(Intercept)` <- NULL

dataBrestCancerTest <- cbind(time=gbsgdata[rownames(data),"rfstime"],status=gbsgdata[rownames(data),"sta

colnames(dataBrestCancerTest) <-str_replace_all(colnames(dataBrestCancerTest),":","_")
colnames(dataBrestCancerTest) <-str_replace_all(colnames(dataBrestCancerTest)," ","")
colnames(dataBrestCancerTest) <-str_replace_all(colnames(dataBrestCancerTest),"\\.","_")
colnames(dataBrestCancerTest) <-str_replace_all(colnames(dataBrestCancerTest),"-","_")
colnames(dataBrestCancerTest) <-str_replace_all(colnames(dataBrestCancerTest),">","_")
dataBrestCancerTest$time <- dataBrestCancerTest$time/365

pander::pander(table(odata[rownames(data),"status"]), caption="gbsg")
```

Table 2: gbsg

0	1
499	183

1.3 Cox Modeling

```
ml <- BSWiMS.model(Surv(time,status)~.,data=dataBrestCancerTrain,loops=1,NumberofRepeats = 5)
```

—.

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

Table 3: Table continues below

	Estimate	lower	HR	upper	u.Accuracy	r.Accuracy
age_nodes	0.000716	1.001	1.001	1.001	0.626	0.600
size_grade	0.005649	1.005	1.006	1.006	0.598	0.623
nodes	0.086582	1.082	1.090	1.099	0.637	0.642
size	0.006888	1.005	1.007	1.009	0.595	0.641
size_nodes	-0.000378	1.000	1.000	1.000	0.624	0.643

	Estimate	lower	HR	upper	u.Accuracy	r.Accuracy
age__size	-0.000149	1.000	1.000	1.000	0.567	0.627
grade	0.204934	1.146	1.227	1.314	0.565	0.637
age	-0.003113	0.996	0.997	0.998	0.513	0.628
grade__nodes	-0.013784	0.981	0.986	0.992	0.635	0.645

Table 4: Table continues below

	full.Accuracy	u.AUC	r.AUC	full.AUC	IDI	NRI
age__nodes	0.632	0.630	0.601	0.634	0.03040	0.4594
size__grade	0.632	0.599	0.626	0.634	0.01868	0.3914
nodes	0.643	0.640	0.643	0.644	0.00745	0.0564
size	0.643	0.595	0.642	0.644	0.01447	0.3587
size__nodes	0.643	0.629	0.644	0.644	0.00346	0.3430
age__size	0.632	0.568	0.630	0.634	0.00635	0.1935
grade	0.643	0.561	0.638	0.644	0.00926	0.2069
age	0.643	0.513	0.628	0.644	0.00416	0.0917
grade__nodes	0.643	0.639	0.646	0.644	0.00207	-0.0910

	z.IDI	z.NRI	Delta.AUC	Frequency
age__nodes	12.81	14.37	0.033056	1
size__grade	9.82	11.29	0.007947	1
nodes	8.33	1.66	0.000148	1
size	8.05	9.97	0.001322	1
size__nodes	7.25	9.57	-0.000377	1
age__size	5.95	5.36	0.004078	1
grade	5.88	6.31	0.005344	1
age	5.27	2.51	0.015465	1
grade__nodes	5.03	-2.55	-0.002609	1

1.4 Cox Model Performance

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

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

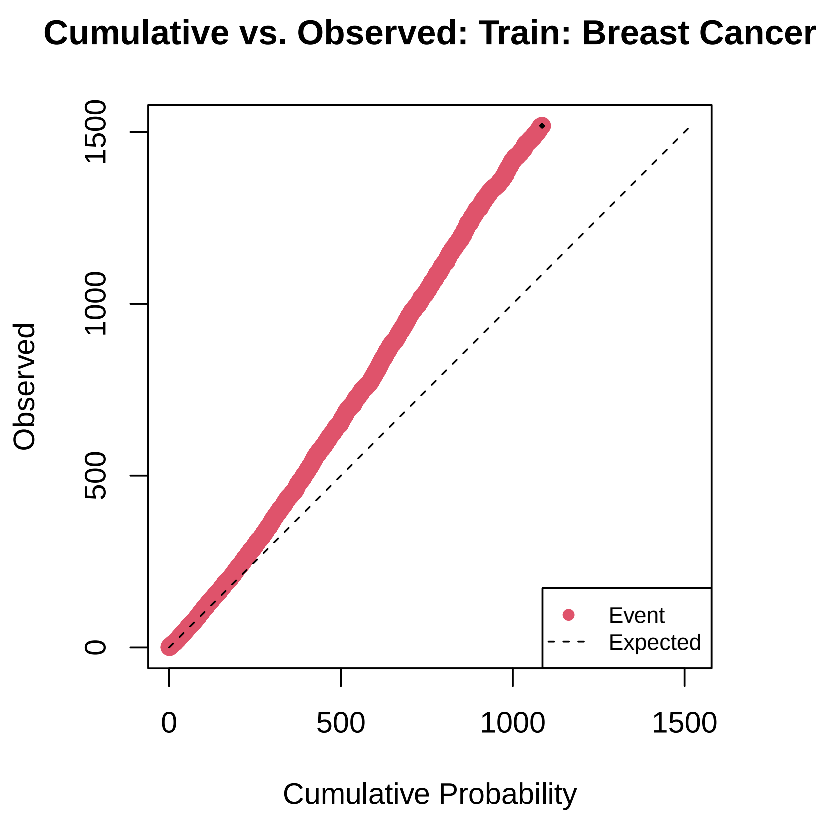
Here we will use the predicted event probability assuming a baseline hazard for events withing 5 years

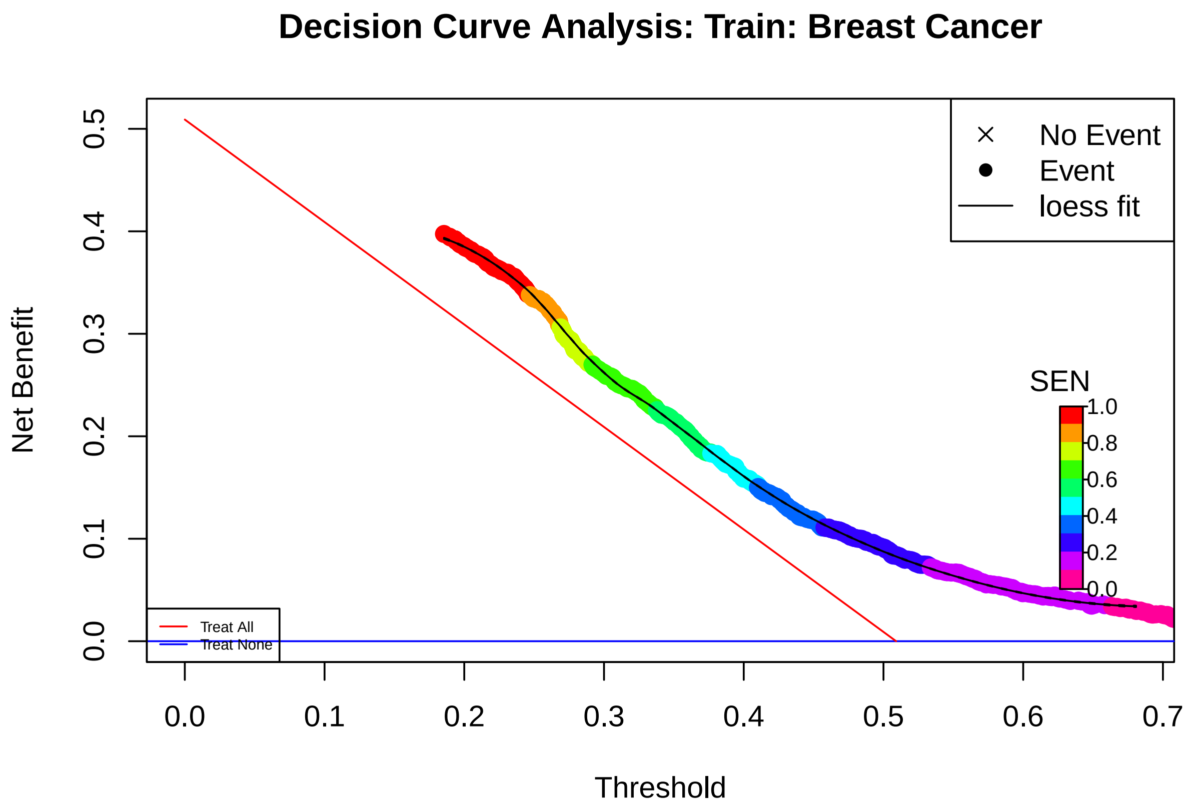
```
timeinterval <- 5 # Five years

h0 <- sum(dataBrestCancerTrain$status & dataBrestCancerTrain$time <= timeinterval)
h0 <- h0/sum((dataBrestCancerTrain$time > timeinterval) | (dataBrestCancerTrain$status==1))

pander::pander(t(c(h0=h0,timeinterval=timeinterval)),caption="Initial Parameters")
```

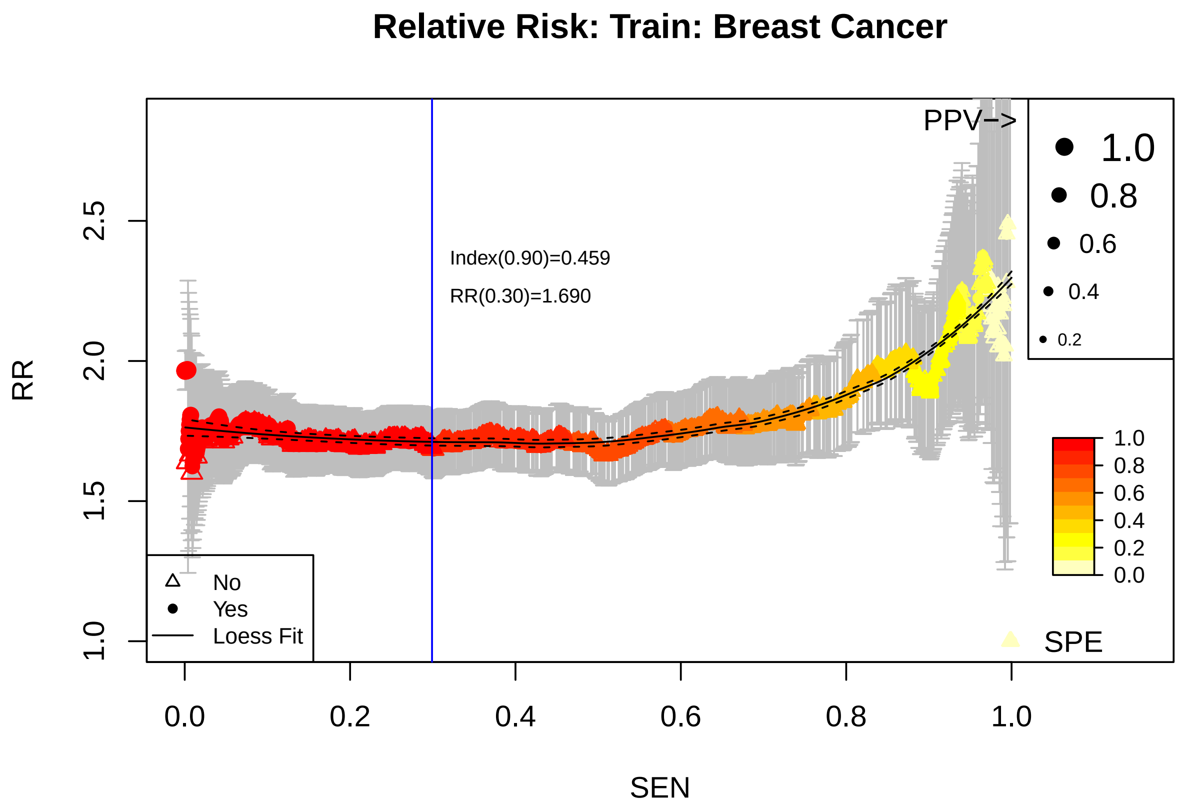
Table 6: Initial Parameters

h0	timeinterval
0.429	5

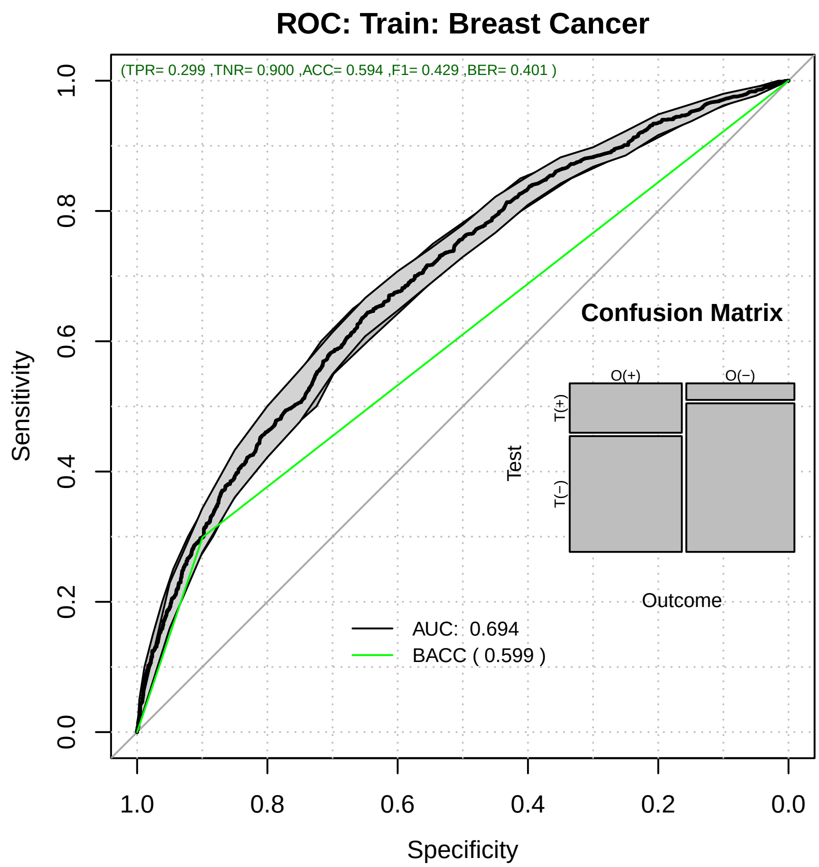
```
index <- predict(ml,dataBrestCancerTrain)
rdata <- cbind(dataBrestCancerTrain$status,ppoisGzero(index,h0))

rrAnalysisTrain <- RRPlot(rdata,atProb=c(0.90,0.80),
                     timetoEvent=dataBrestCancerTrain$time,
                     title="Train: Breast Cancer",
                     ysurvlim=c(0.00,1.0),
                     riskTimeInterval=timeinterval)
```



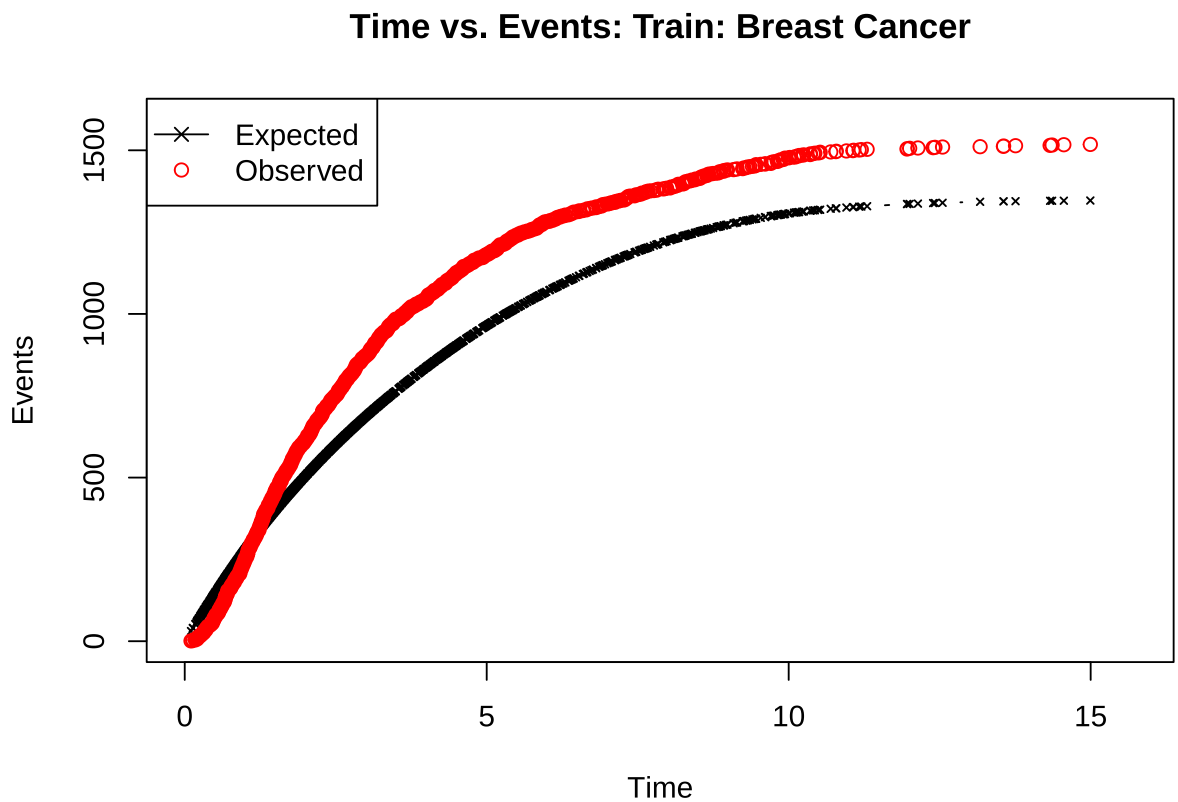

Decision Curve Analysis: Train: Breast Cancer
Net Benefit
Threshold
No Event
Event
loess fit
SEN
1.0
0.8
0.6
0.4
0.2
0.0
Treat All
Treat None
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7

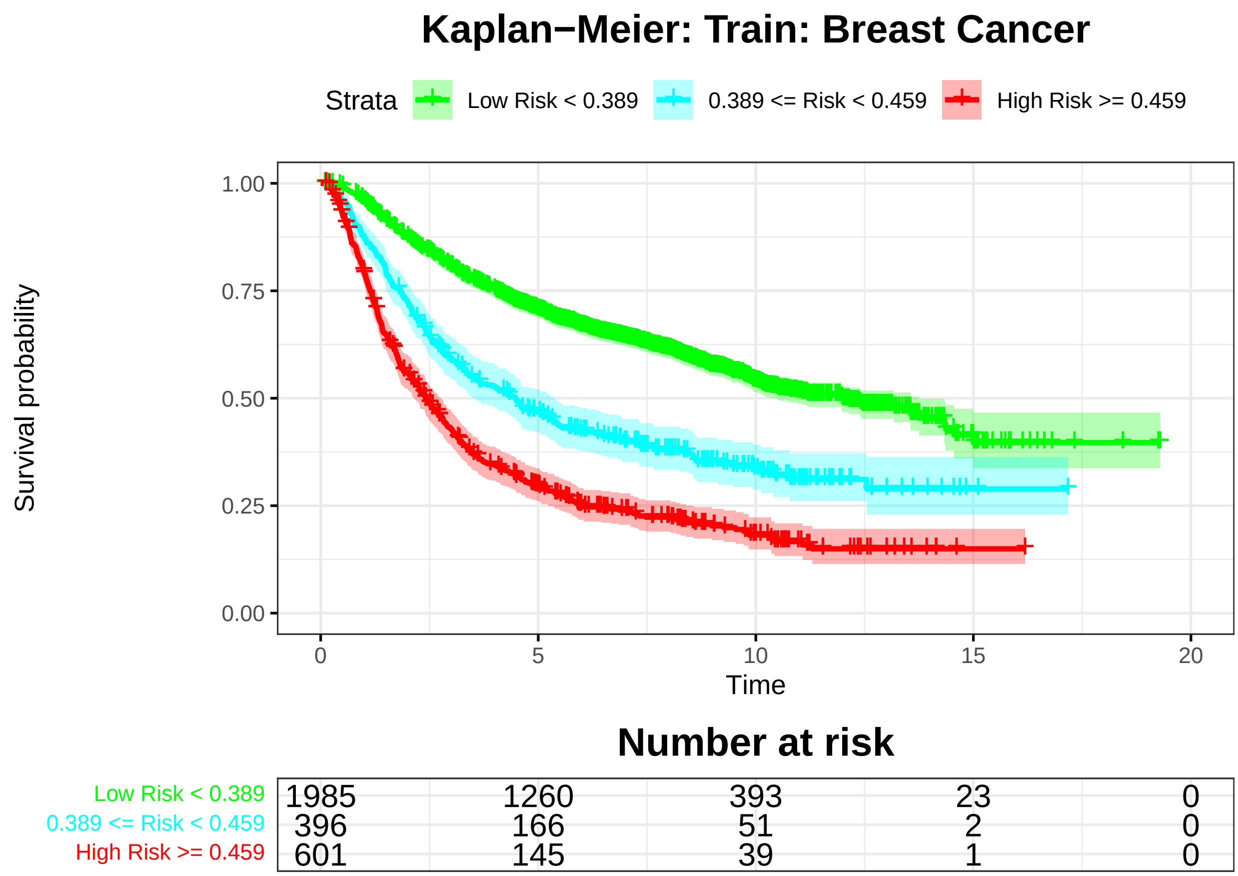


Relative Risk: Train: Breast Cancer
PPV–>
1.0
0.8
0.6
0.4
0.2
Index(0.90)=0.459
RR(0.30)=1.690
1.0
0.8
0.6
0.4
0.2
0.0
No
Yes
Loess Fit
SPE
RR
1.0
1.5
2.0
2.5
0.0
0.2
0.4
0.6
0.8
1.0
SEN



ROC: Train: Breast Cancer
(TPR= 0.299 ,TNR= 0.900 ,ACC= 0.594 ,F1= 0.429 ,BER= 0.401)
Sensitivity
Specificity
1.0
0.8
0.6
0.4
0.2
0.0
Confusion Matrix
O(+)
O(−)
T(+)
T(−)
Test
Outcome
AUC: 0.694
BACC (0.599)


Time vs. Events: Train: Breast Cancer
Expected
Observed
Events
0
500
1000
1500
0
5
10
15
Time



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

1.4.2 Uncalibrated Performance Report

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

Table 7: O/E Ratio

est	lower	upper
1.13	1.07	1.19

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

Table 8: O/E Ratio

mean	50%	2.5%	97.5%
1.13	1.13	1.12	1.14

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

Table 9: O/Acum Ratio

mean	50%	2.5%	97.5%
1.34	1.34	1.34	1.34

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

mean.C Index	median	lower	upper
0.676	0.677	0.663	0.69

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

Table 11: ROC AUC

est	lower	upper
0.694	0.675	0.713

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

Table 12: Sensitivity

est	lower	upper
0.299	0.276	0.323

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

Table 13: Specificity

est	lower	upper
0.9	0.883	0.915

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

Table 14: Probability Thresholds

90%	80%
0.459	0.389

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

Table 15: Risk Ratio

est	lower	upper
1.69	1.59	1.8

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

Table 16: Logrank test Chisq = 465.079317 on 2 degrees of freedom, p = 0.000000

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	1985	816	1144	93.9	385.7
class=1	396	248	177	28.0	31.8
class=2	601	454	197	336.3	391.3

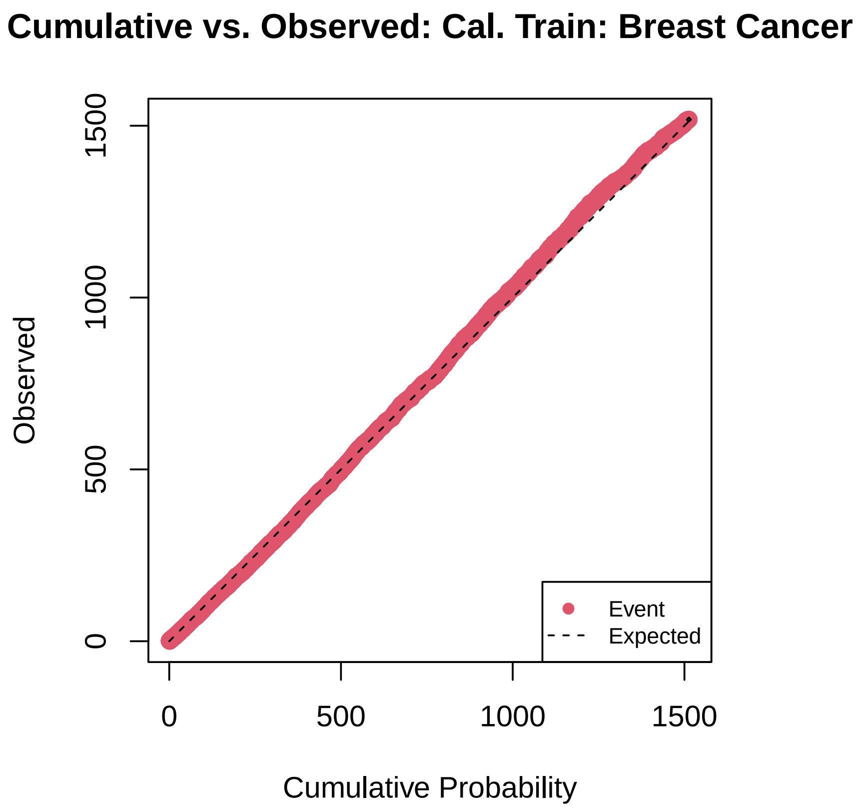
1.4.3 Cox Calibration

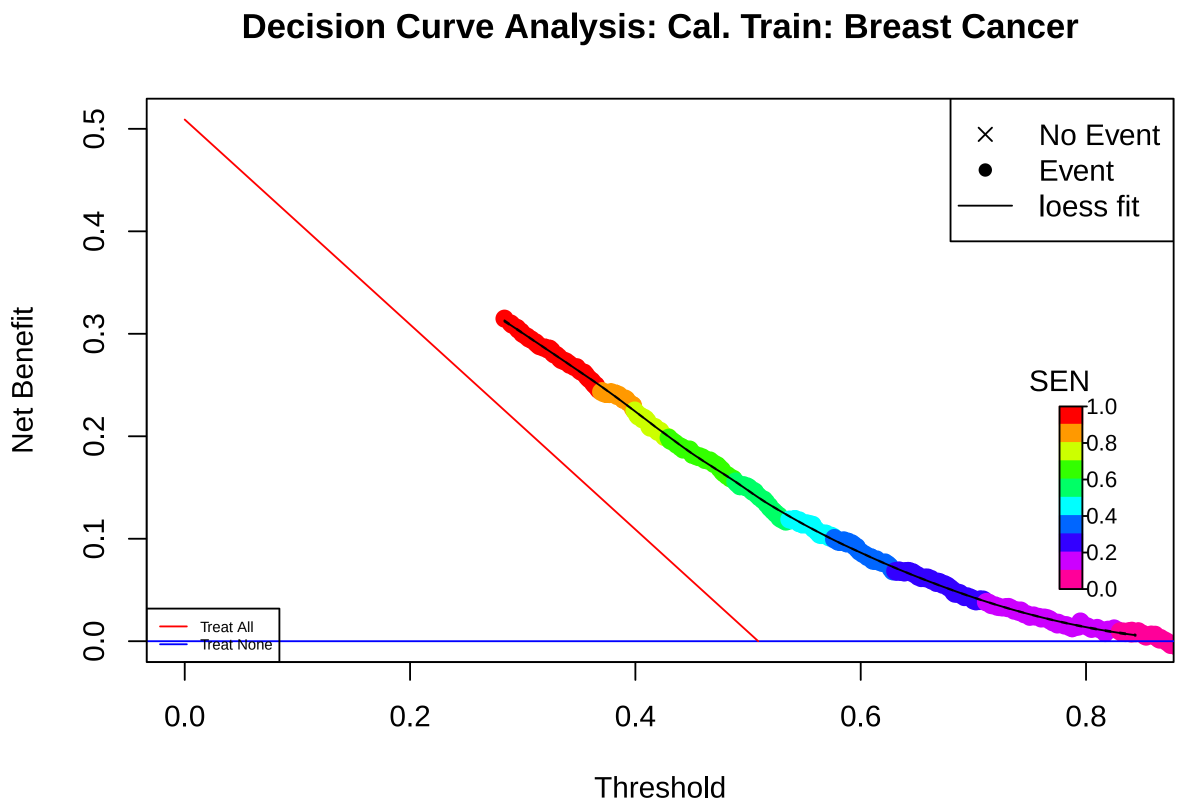
```
op <- par(no.readonly = TRUE)


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


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

h0	Gain	DeltaTime
0.698	1.35	6.97

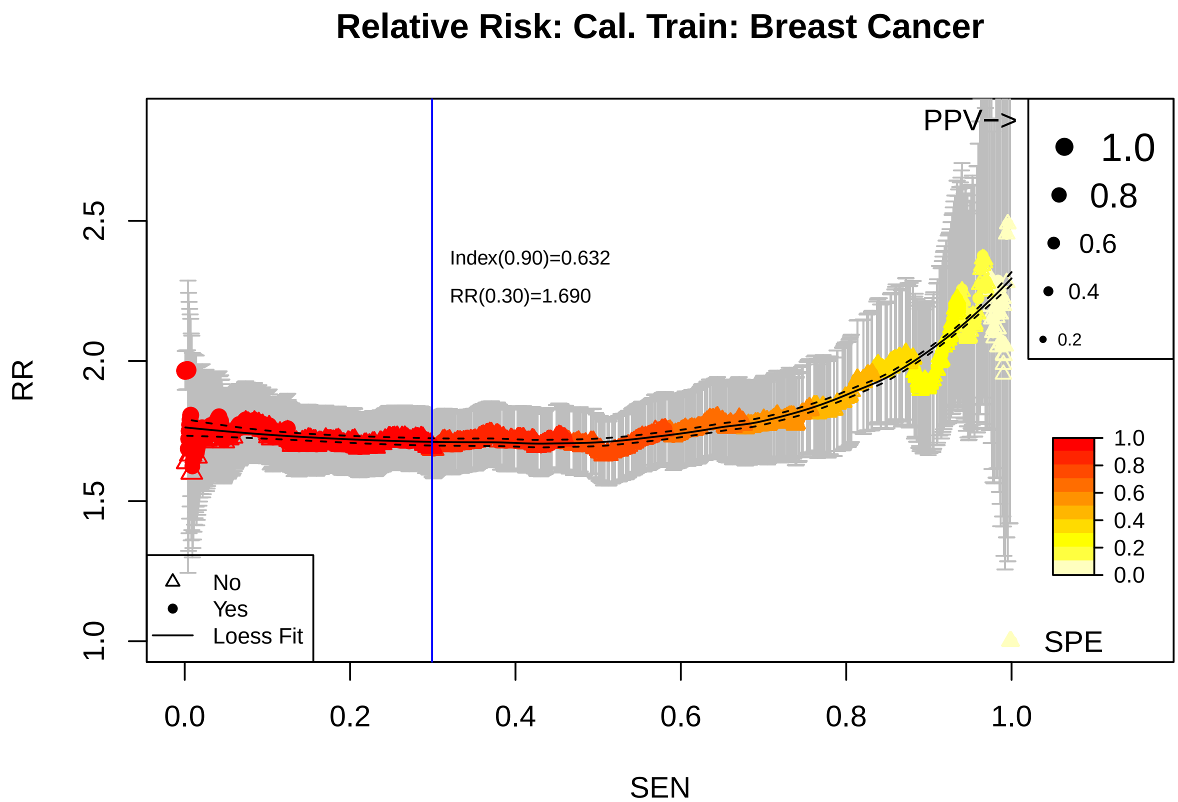
1.4.4 The RRplot() of the calibrated model

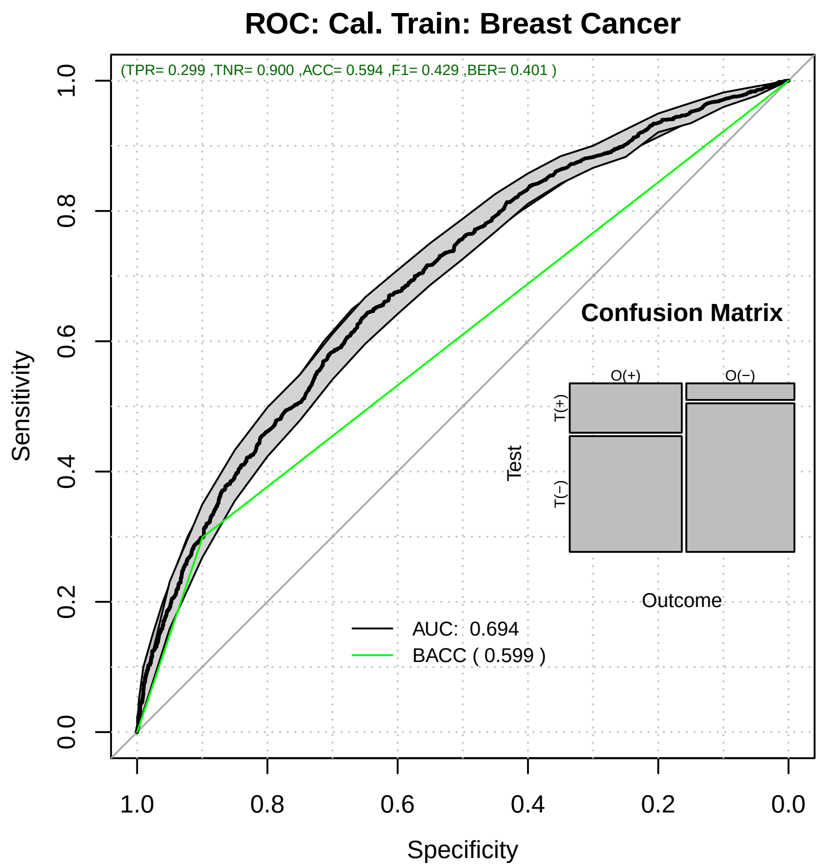
```
h0 <- calprob$h0
timeinterval <- calprob$timeInterval;

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


rrAnalysisTrain <- RRPlot(rdata,atProb=c(0.90,0.80),
                     timetoEvent=dataBrestCancerTrain$time,
                     title="Cal. Train: Breast Cancer",
                     ysurvlim=c(0.00,1.0),
                     riskTimeInterval=timeinterval)
```


Cumulative vs. Observed: Cal. Train: Breast Cancer
Observed
0
500
1000
1500
Event
Expected
Cumulative Probability

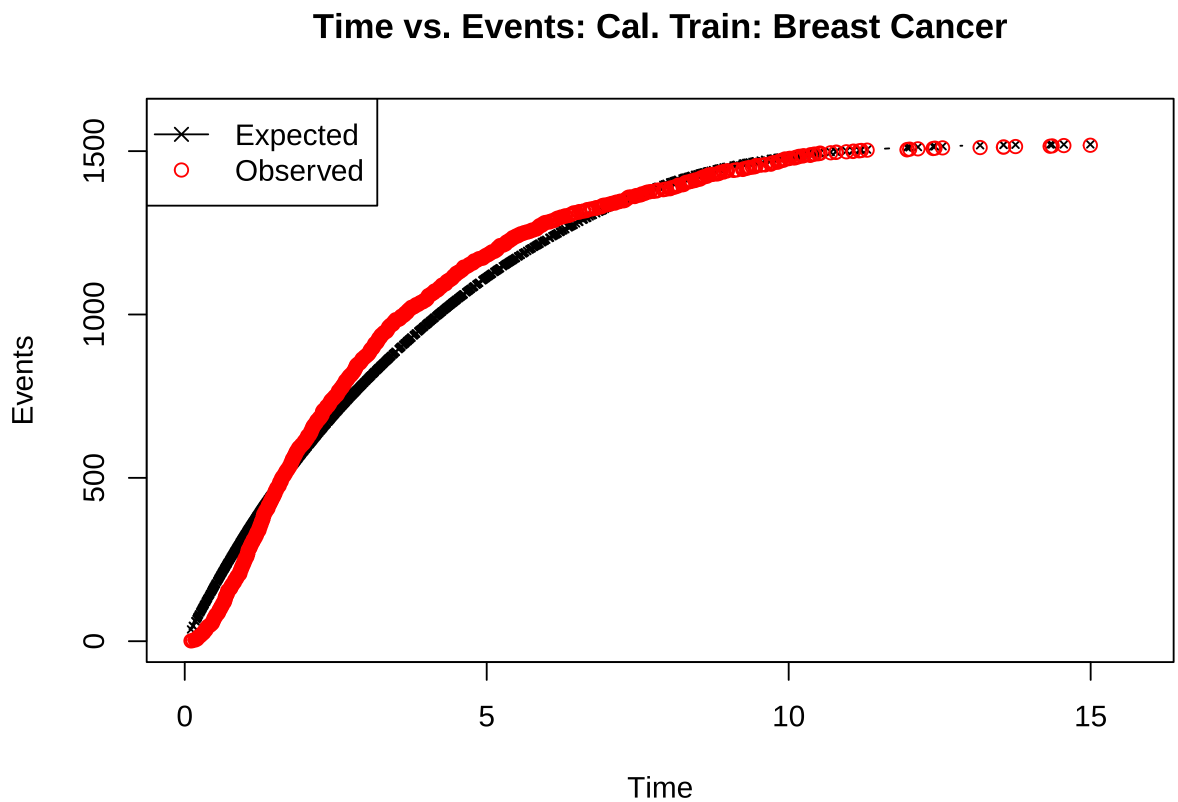


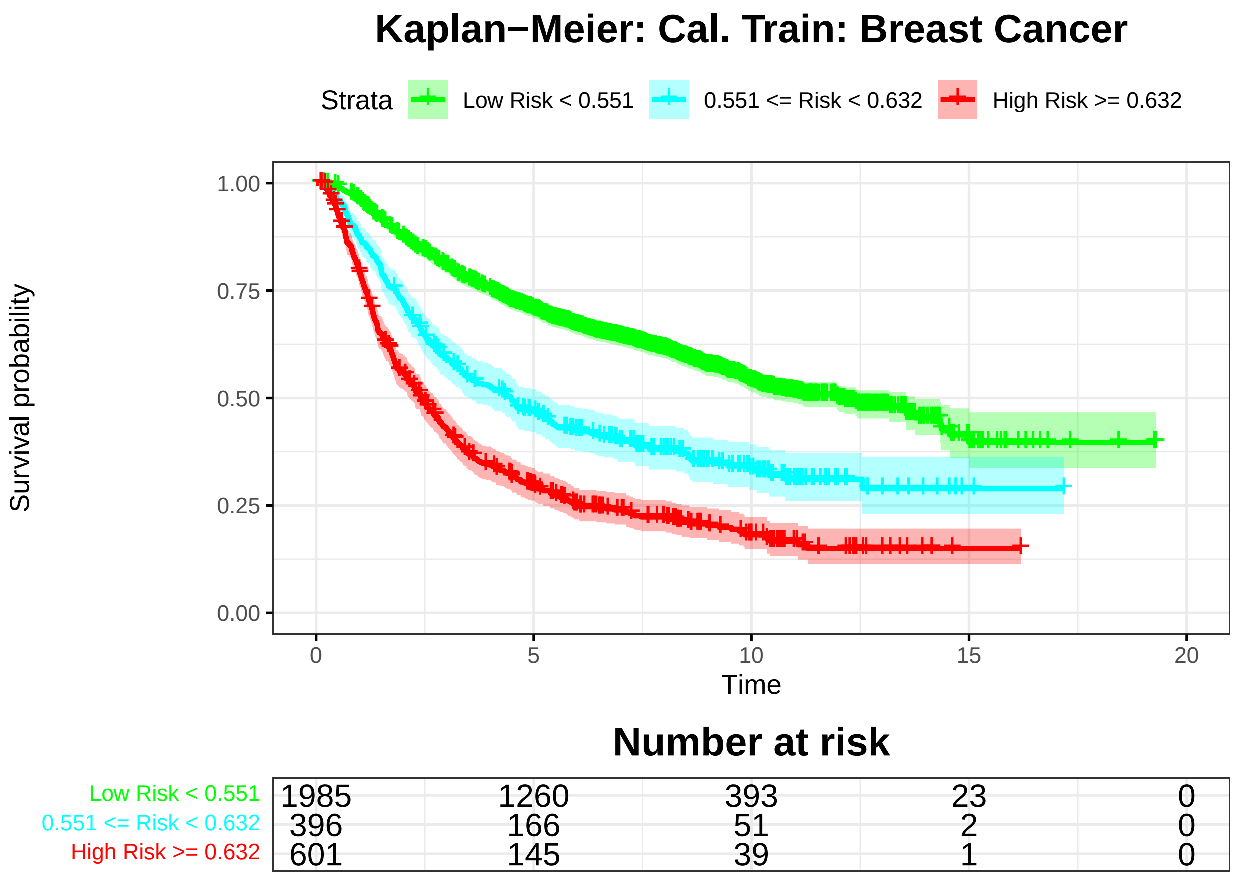
Decision Curve Analysis: Cal. Train: Breast Cancer
Net Benefit
Threshold
No Event
Event
loess fit
SEN
1.0
0.8
0.6
0.4
0.2
0.0
Treat All
Treat None
0.0
0.1
0.2
0.3
0.4
0.5
0.0
0.2
0.4
0.6
0.8


Relative Risk: Cal. Train: Breast Cancer
PPV–>
1.0
0.8
0.6
0.4
0.2
Index(0.90)=0.632
RR(0.30)=1.690
RR
2.5
2.0
1.5
1.0
No
Yes
Loess Fit
1.0
0.8
0.6
0.4
0.2
0.0
SPE
0.0
0.2
0.4
0.6
0.8
1.0
SEN



ROC: Cal. Train: Breast Cancer
(TPR= 0.299 ,TNR= 0.900 ,ACC= 0.594 ,F1= 0.429 ,BER= 0.401)
Sensitivity
Specificity
1.0
0.8
0.6
0.4
0.2
0.0
Confusion Matrix
O(+)
O(−)
T(+)
T(−)
Test
Outcome
AUC: 0.694
BACC (0.599)


Time vs. Events: Cal. Train: Breast Cancer
Expected
Observed
Events
0
500
1000
1500
0
5
10
15
Time



1.4.5 Calibrated Train Performance

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

Table 18: O/E Ratio

est	lower	upper
0.998	0.949	1.05

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

Table 19: O/E Ratio

mean	50%	2.5%	97.5%
0.977	0.977	0.969	0.985

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

Table 20: O/Acum Ratio

mean	50%	2.5%	97.5%
1.01	1.01	1.01	1.01

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

mean.C Index	median	lower	upper
0.676	0.677	0.663	0.689

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

Table 22: ROC AUC

est	lower	upper
0.694	0.675	0.713

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

Table 23: Sensitivity

est	lower	upper
0.299	0.276	0.323

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

Table 24: Specificity

est	lower	upper
0.9	0.883	0.915

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

Table 25: Probability Thresholds

90%	80%
0.632	0.551

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

Table 26: Risk Ratio

est	lower	upper
1.69	1.59	1.8

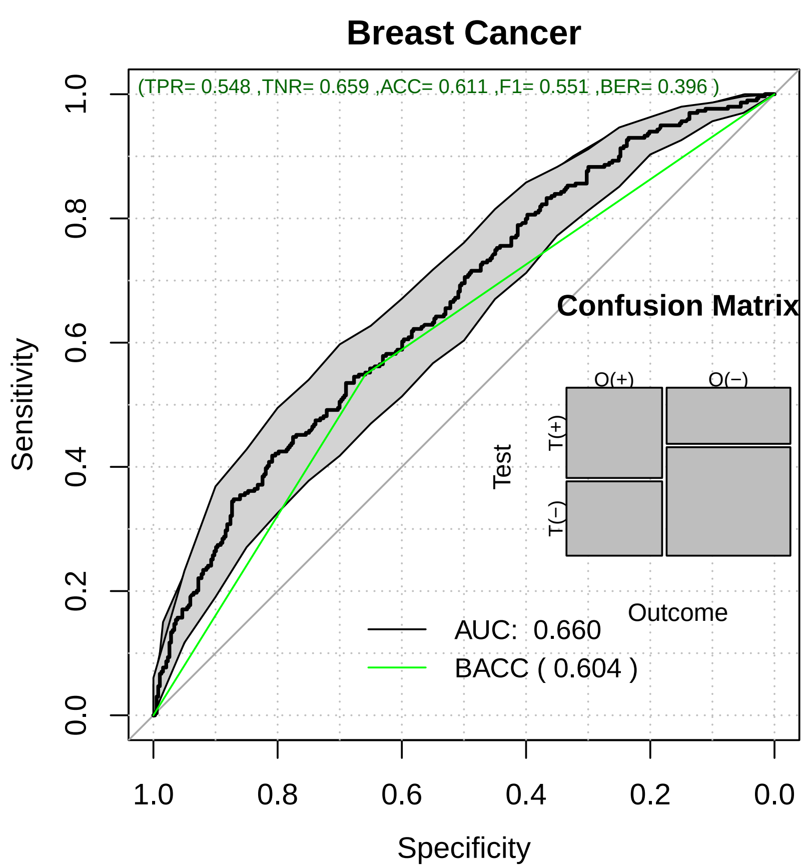
```
pander::pander(rrAnalysisTrain$surdif,caption="Logrank test")
```

Table 27: Logrank test Chisq = 465.079317 on 2 degrees of freedom, p = 0.000000

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	1985	816	1144	93.9	385.7
class=1	396	248	177	28.0	31.8
class=2	601	454	197	336.3	391.3

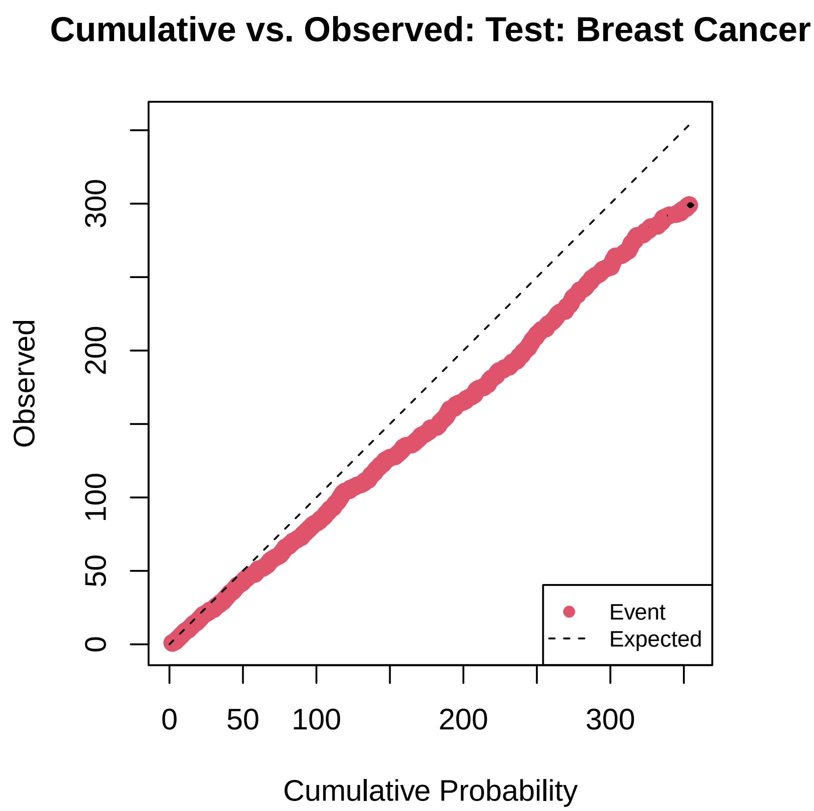
1.5 Performance on the external data set

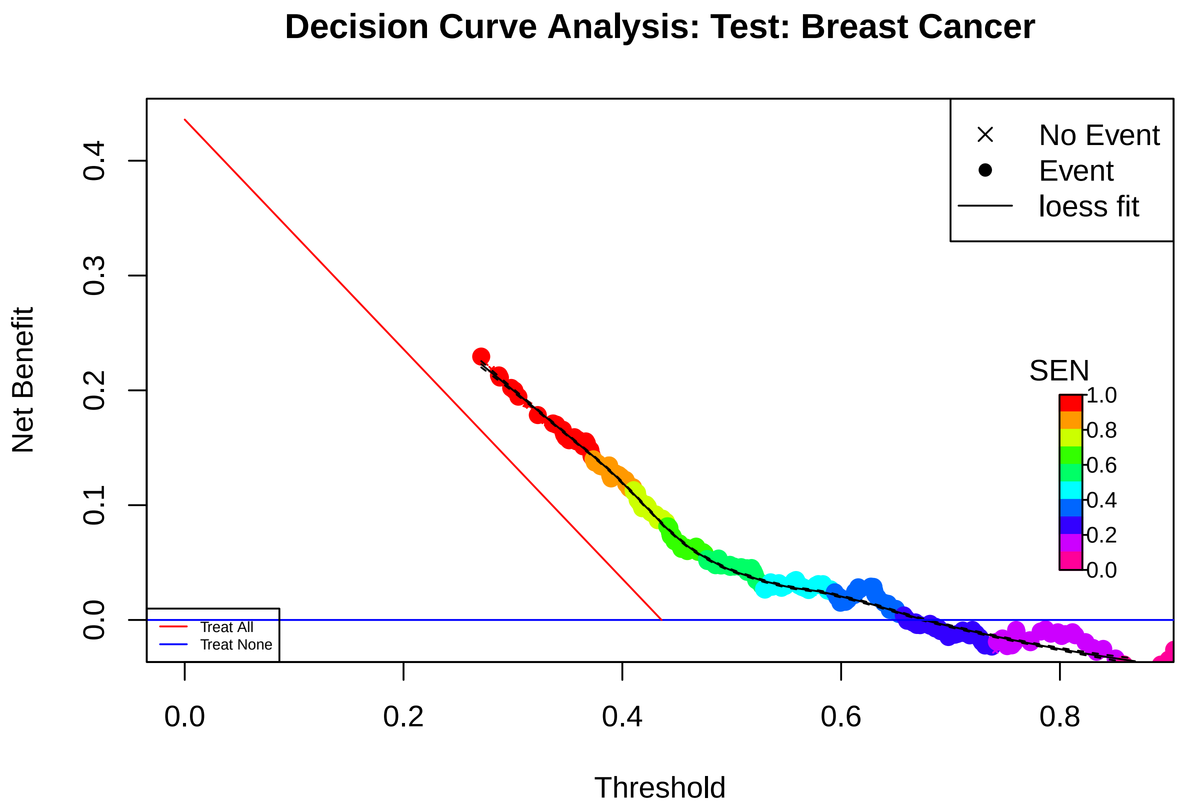
```
index <- predict(ml,dataBrestCancerTest)
pp <- predictionStats_binary(cbind(dataBrestCancerTest$status,index),plotname="Breast Cancer")
```

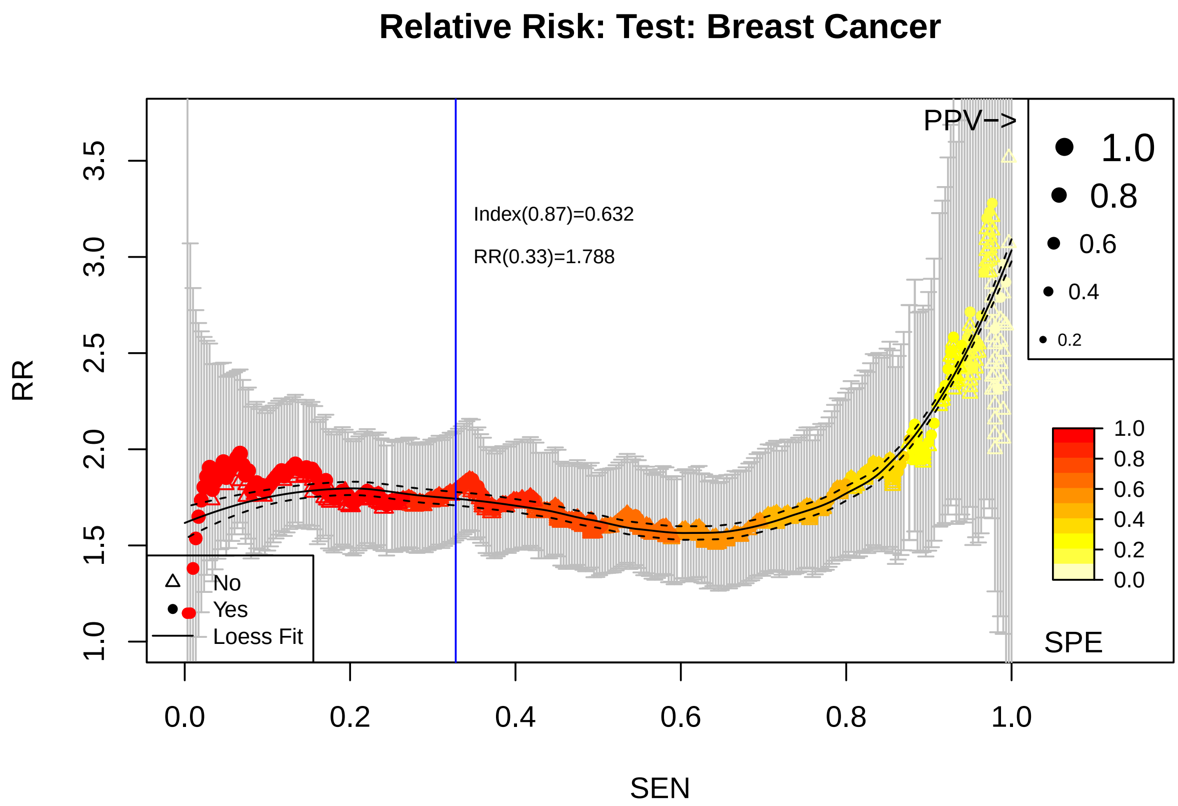



```
par(op)
```

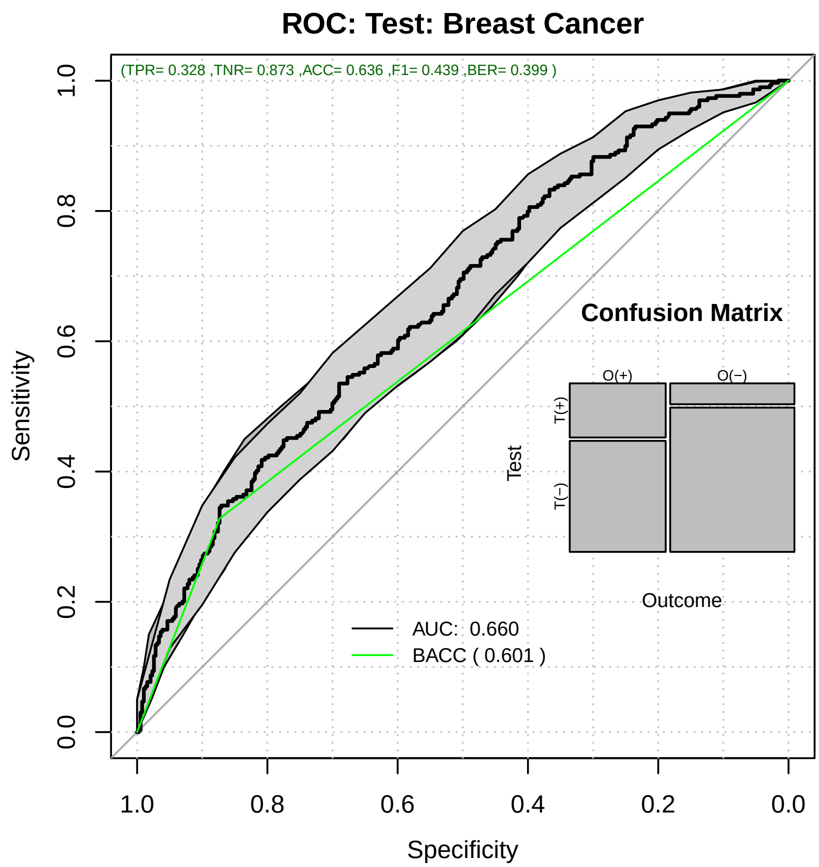
```
prob <- ppoisGzero(index,h0)
rdata <- cbind(dataBrestCancerTest$status,prob)
rrCoxTestAnalysis <- RRPlot(rdata,atThr=rrAnalysisTrain$thr_atP,
                     timetoEvent=dataBrestCancerTest$time,
                     title="Test: Breast Cancer",
                     ysurvlim=c(0.00,1.0),
                     riskTimeInterval=timeinterval)
```

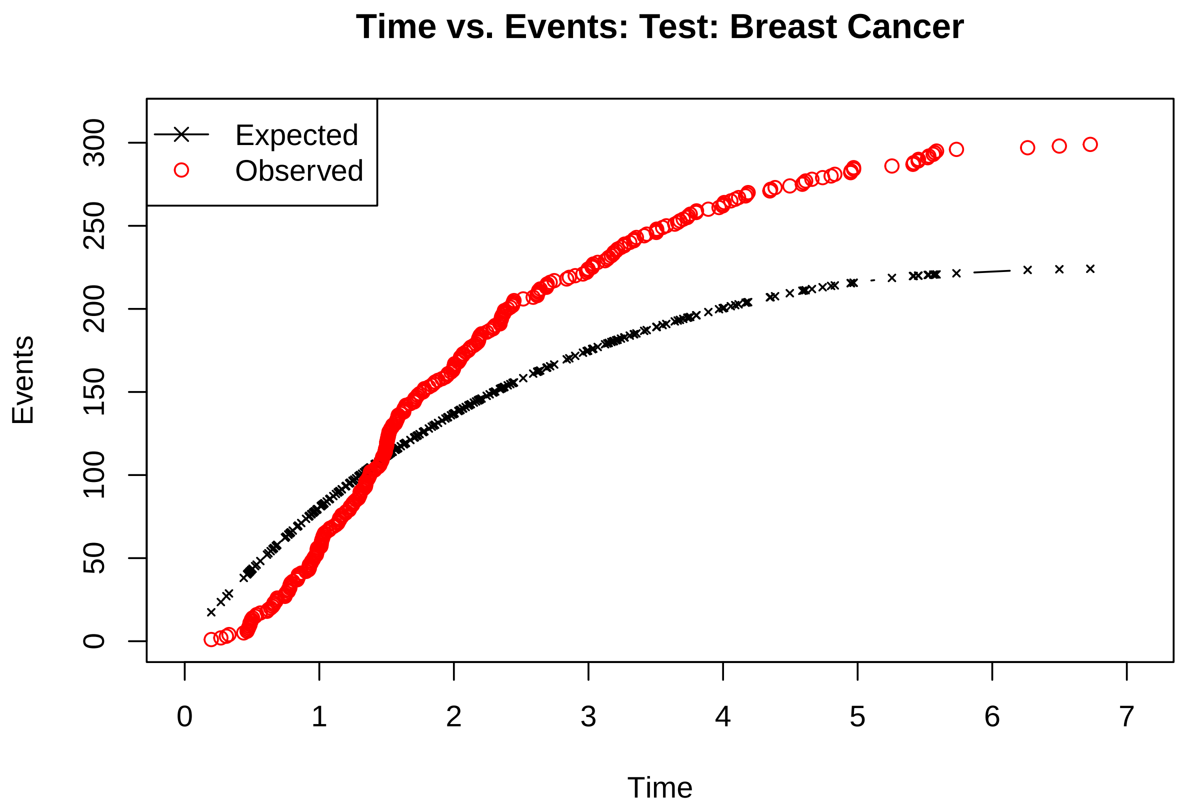


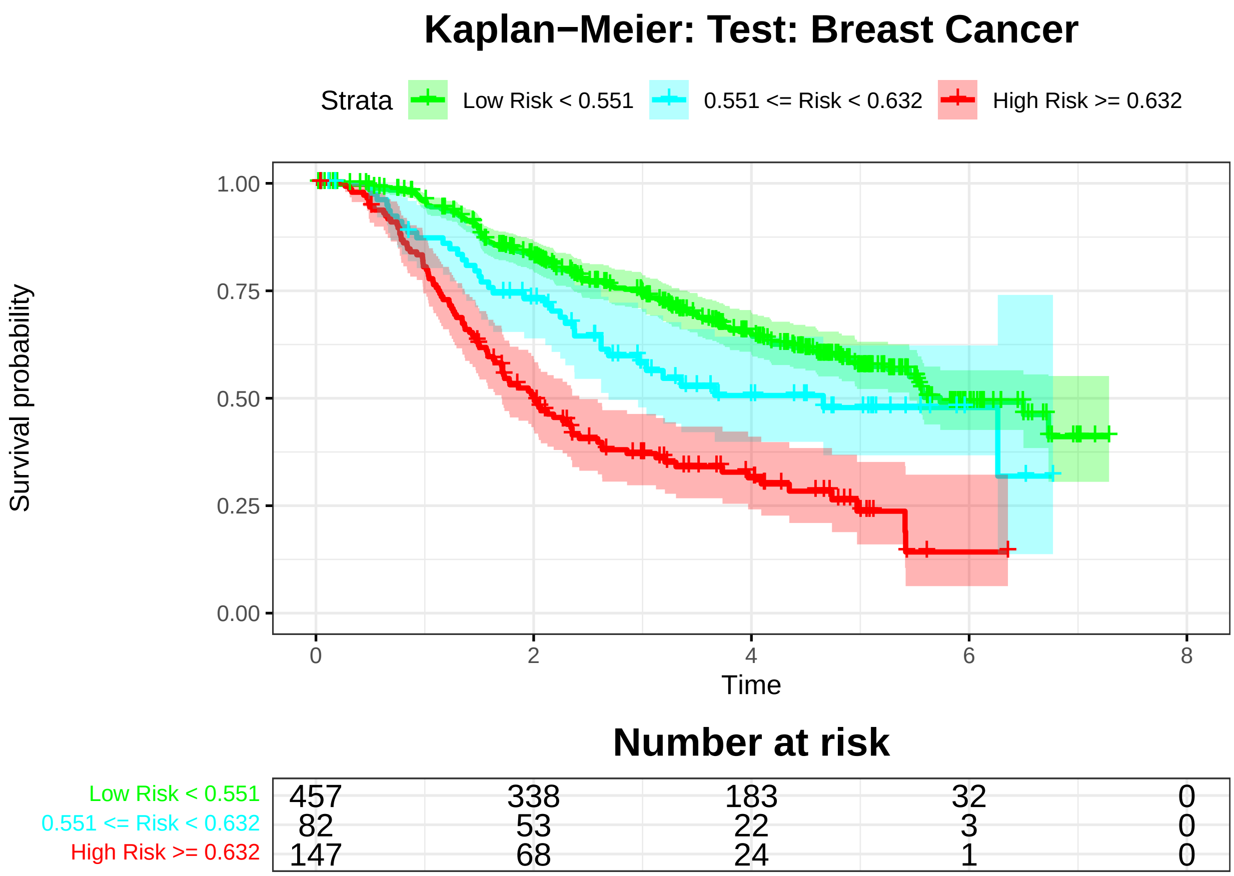

Decision Curve Analysis: Test: Breast Cancer
No Event
Event
loess fit
SEN
1.0
0.8
0.6
0.4
0.2
0.0
Treat All
Treat None
Net Benefit
0.0
0.1
0.2
0.3
0.4
0.0
0.2
0.4
0.6
0.8
Threshold



Relative Risk: Test: Breast Cancer
RR
SEN
PPV−>
1.0
0.8
0.6
0.4
0.2
SPE
Index(0.87)=0.632
RR(0.33)=1.788
No
Yes
Loess Fit


ROC: Test: Breast Cancer
(TPR= 0.328 ,TNR= 0.873 ,ACC= 0.636 ,F1= 0.439 ,BER= 0.399)
Sensitivity
Specificity
1.0
0.8
0.6
0.4
0.2
0.0
Confusion Matrix
O(+)
O(−)
T(+)
T(−)
Test
Outcome
AUC: 0.660
BACC (0.601)


Time vs. Events: Test: Breast Cancer
Expected
Observed
Events
0
50
100
150
200
250
300
0
1
2
3
4
5
6
7
Time




```
par(op)
```

1.5.1 External Data Report

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

Table 28: O/E Ratio

est	lower	upper
1.33	1.19	1.49

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

- **C Index**: *0.664*
- **Dxy**: *0.328*
- **S.D.**: *0.0311*
- **n**: *686*
- **missing**: *0*
- **uncensored**: *299*
- **Relevant Pairs**: *266144*

- **Concordant**: *176737*
- **Uncertain**: *203702*
- **cstatCI**:

mean.C Index	median	lower	upper
0.664	0.663	0.634	0.693

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

Table 30: ROC AUC

est	lower	upper
0.66	0.619	0.7

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

Table 31: Sensitivity

est	lower	upper
0.328	0.275	0.384

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

Table 32: Specificity

est	lower	upper
0.873	0.836	0.905

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

Table 33: Probability Thresholds

90%	80%
0.632	0.551

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

Table 34: Risk Ratio

est	lower	upper
1.79	1.53	2.09

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

Table 35: Logrank test Chisq = 81.471750 on 2 degrees of freedom, p = 0.000000

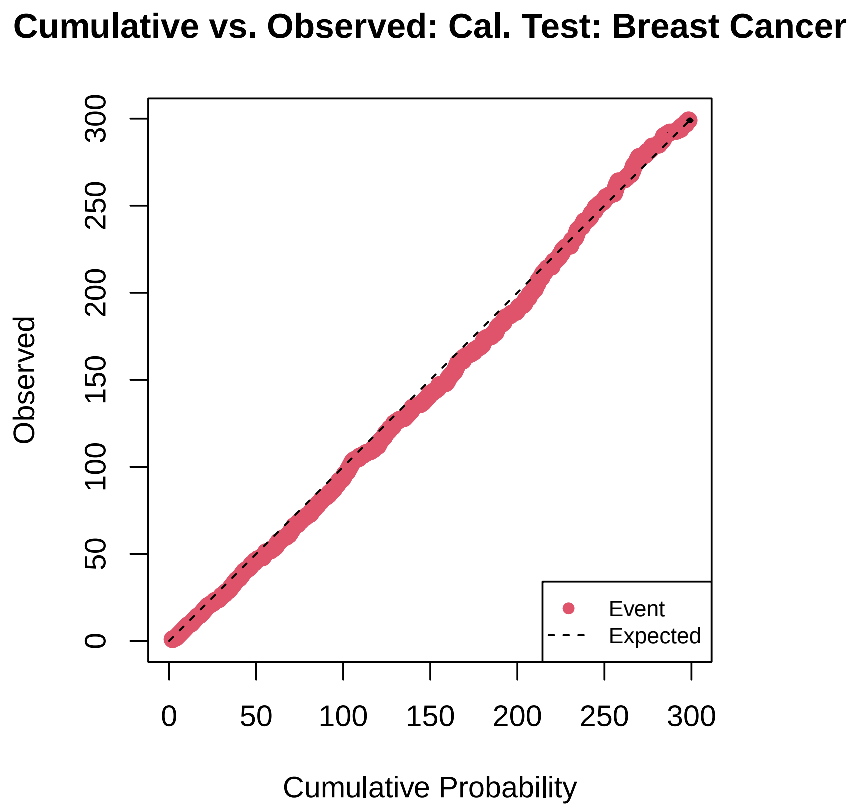
	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	457	164	221.4	14.888	58.181
class=1	82	37	33.2	0.438	0.494
class=2	147	98	44.4	64.710	77.254

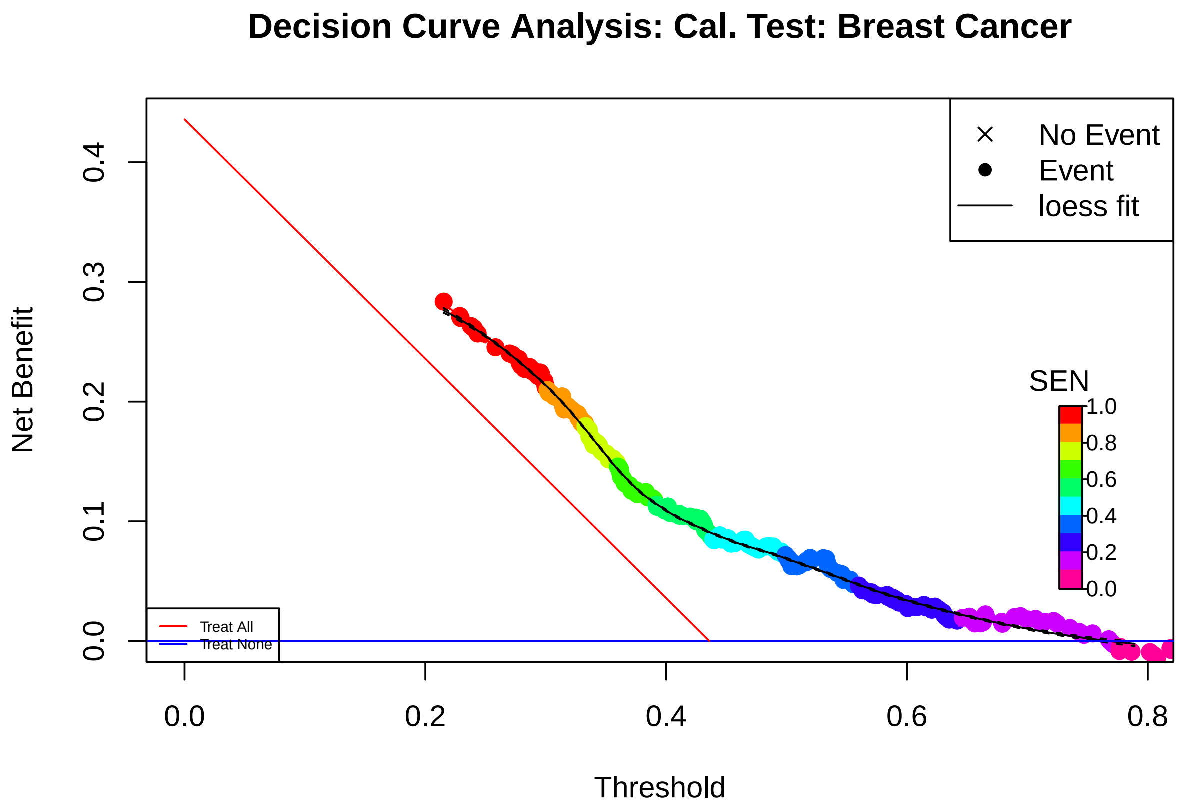
1.5.2 Calibrating the index on the test data

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

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

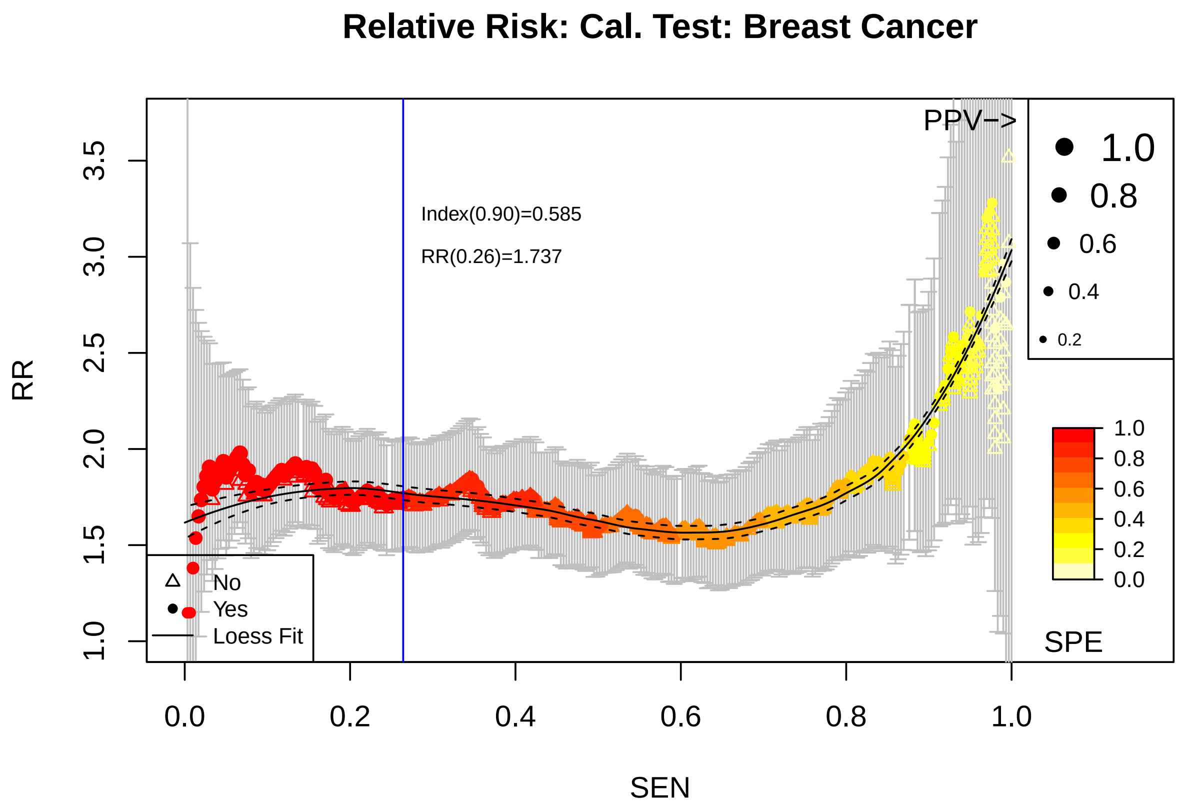
h0	Gain	DeltaTime
0.535	0.925	4.87

```
rdata <- cbind(dataBrestCancerTest$status,calprob$prob)

rrAnalysis <- RRPlot(rdata,atProb=c(0.90,0.80),
                     timetoEvent=dataBrestCancerTest$time,
                     title="Cal. Test: Breast Cancer",
                     ysurvlim=c(0.00,1.0),
                     riskTimeInterval=calprob$timeInterval)
```

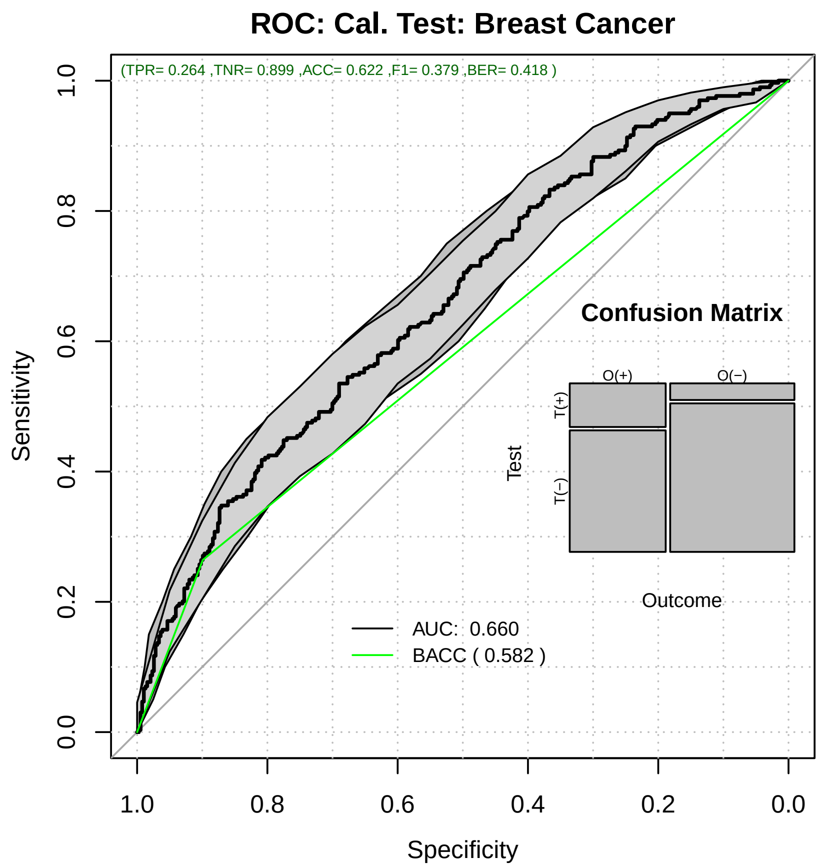

Cumulative vs. Observed: Cal. Test: Breast Cancer
Observed
0
50
100
150
200
250
300
Event
Expected
0
50
100
150
200
250
300
Cumulative Probability

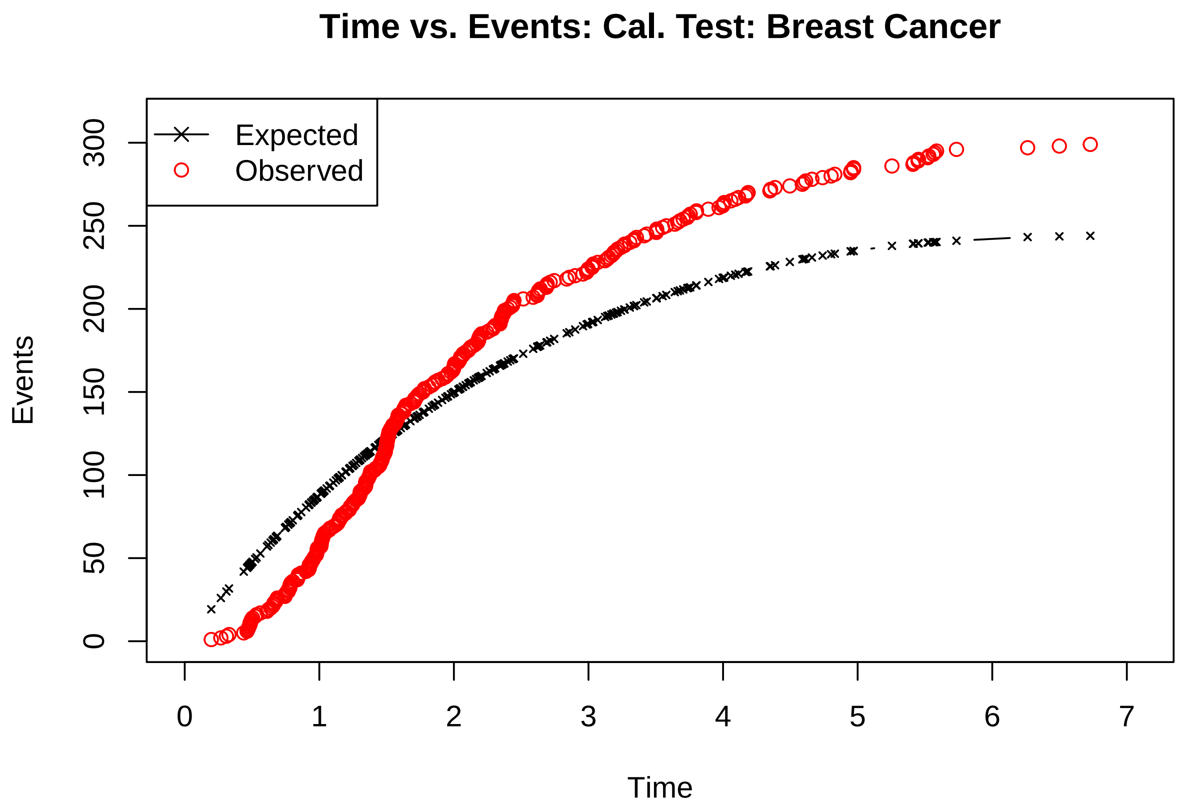


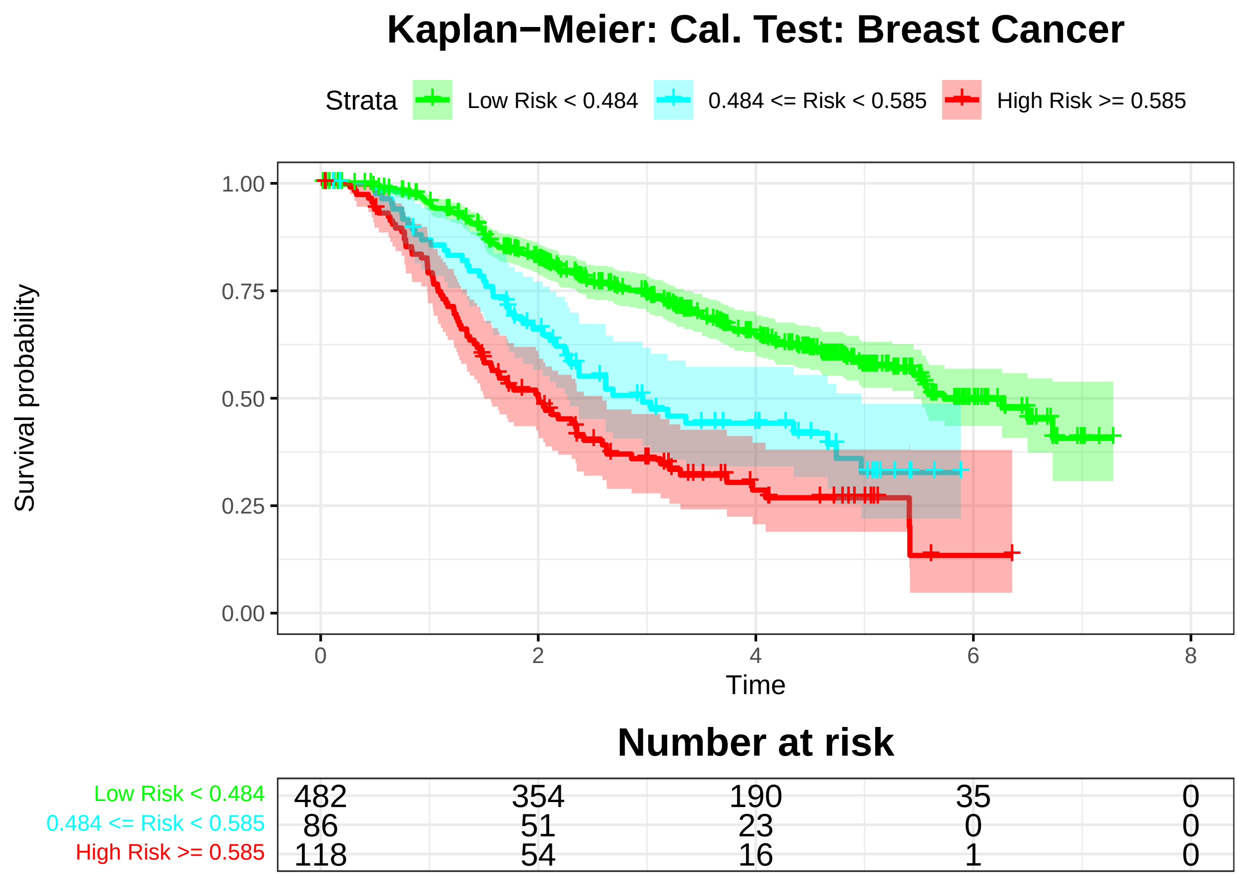
Decision Curve Analysis: Cal. Test: Breast Cancer
No Event
Event
loess fit
SEN
1.0
0.8
0.6
0.4
0.2
0.0
Net Benefit
0.0
0.1
0.2
0.3
0.4
Treat All
Treat None
0.0
0.2
0.4
0.6
0.8
Threshold



Relative Risk: Cal. Test: Breast Cancer
RR
SEN
PPV−>
1.0
0.8
0.6
0.4
0.2
Index(0.90)=0.585
RR(0.26)=1.737
No
Yes
Loess Fit
SPE
1.0
0.8
0.6
0.4
0.2
0.0
3.5
3.0
2.5
2.0
1.5
1.0
0.0
0.2
0.4
0.6
0.8
1.0


ROC: Cal. Test: Breast Cancer
(TPR= 0.264 ,TNR= 0.899 ,ACC= 0.622 ,F1= 0.379 ,BER= 0.418)
Sensitivity
Specificity
1.0
0.8
0.6
0.4
0.2
0.0
Confusion Matrix
O(+)
O(−)
T(+)
T(−)
Test
Outcome
AUC: 0.660
BACC (0.582)


Time vs. Events: Cal. Test: Breast Cancer
Expected
Observed
Events
300
250
200
150
100
50
0
0
1
2
3
4
5
6
7
Time



1.5.3 After Calibration Report

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

Table 37: O/E Ratio

est	lower	upper
1.23	1.09	1.37

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

- **C Index**: *0.664*
- **Dxy**: *0.328*
- **S.D.**: *0.0311*
- **n**: *686*
- **missing**: *0*
- **uncensored**: *299*
- **Relevant Pairs**: *266144*
- **Concordant**: *176737*
- **Uncertain**: *203702*

- **cstatCI**:

mean.C Index	median	lower	upper
0.664	0.664	0.633	0.691

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

Table 39: ROC AUC

est	lower	upper
0.66	0.619	0.7

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

Table 40: Sensitivity

est	lower	upper
0.264	0.215	0.318

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

Table 41: Specificity

est	lower	upper
0.899	0.865	0.927

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

Table 42: Probability Thresholds

90%	80%
0.585	0.484

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

Table 43: Risk Ratio

est	lower	upper
1.74	1.48	2.05

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

Table 44: Logrank test Chisq = 80.835092 on 2 degrees of freedom, p = 0.000000

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	482	173	232.4	15.20	69.5
class=1	86	47	32.0	7.02	7.9
class=2	118	79	34.6	57.14	65.4

1.6 Logistic Model

Here we train a logistic model on the same data set

```
## Only label subjects that present event withing five years

dataBrestCancerR <- subset(dataBrestCancerTrain, time>=5 | status==1)
dataBrestCancerR$status <- dataBrestCancerR$status * (dataBrestCancerR$time < 5)
dataBrestCancerR$time <- NULL

#ml <- BSWiMS.model(status~1,data=dataBrestCancerR,loops=20,NumberofRepeats = 5)
mlog <- BSWiMS.model(status~1,data=dataBrestCancerR,loops=1,NumberofRepeats = 5)
```

—..

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

Table 45: Table continues below

	Estimate	lower	OR	upper	u.Accuracy	r.Accuracy
size__nodes	1.05e-03	1.001	1.001	1.001	0.669	0.571
nodes	4.33e-02	1.040	1.044	1.048	0.676	0.634
grade__nodes	1.50e-02	1.014	1.015	1.016	0.682	0.637
age__nodes	1.06e-03	1.001	1.001	1.001	0.678	0.653
size__grade	1.75e-03	1.001	1.002	1.002	0.632	0.682
age__size	8.73e-05	1.000	1.000	1.000	0.608	0.682
grade	2.27e-01	1.168	1.254	1.347	0.571	0.683
age__meno	-6.04e-03	0.992	0.994	0.996	0.571	0.676
age__pgr	-5.42e-06	1.000	1.000	1.000	0.571	0.686
age__grade	-1.65e-03	0.997	0.998	0.999	0.574	0.690
meno__grade	1.02e-01	1.045	1.107	1.173	0.571	0.683
nodes__hormon	-1.38e-02	0.979	0.986	0.994	0.587	0.688
size	3.94e-03	1.002	1.004	1.006	0.611	0.693
meno__pgr	3.19e-04	1.000	1.000	1.001	0.571	0.687
pgr	-1.07e-04	1.000	1.000	1.000	0.571	0.689
meno__nodes	-2.60e-02	0.955	0.974	0.994	0.640	0.686
grade__pgr	-3.51e-05	1.000	1.000	1.000	0.571	0.669
meno__size	2.34e-03	1.000	1.002	1.004	0.604	0.691

Table 46: Table continues below

	full.Accuracy	u.AUC	r.AUC	full.AUC	IDI
size__nodes	0.668	0.627	0.500	0.628	0.11233

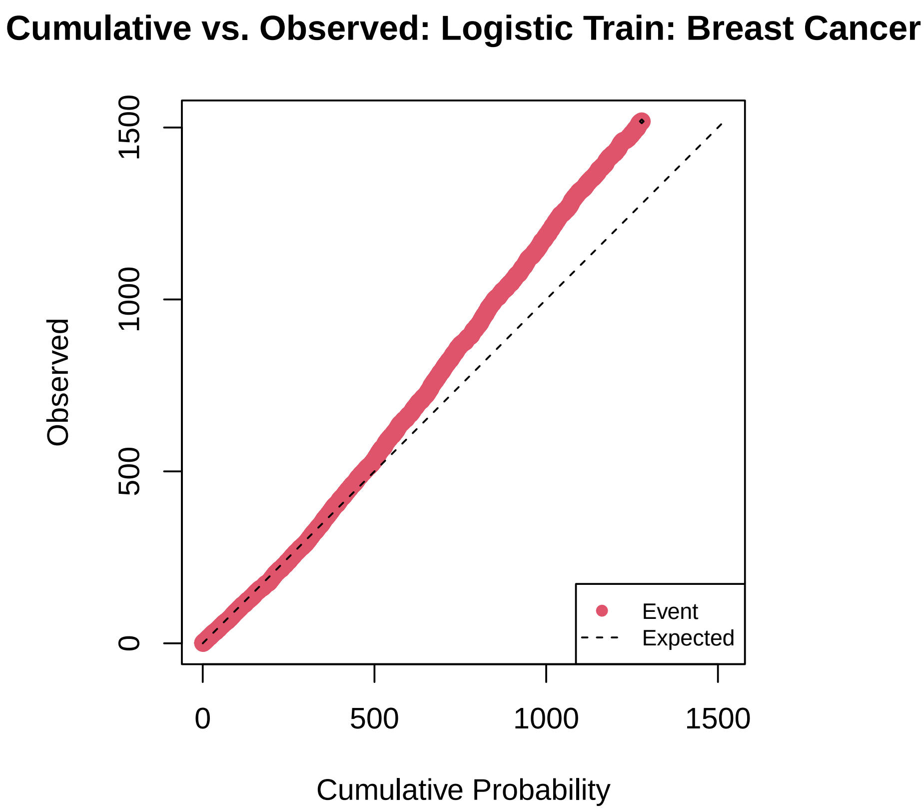
	full.Accuracy	u.AUC	r.AUC	full.AUC	IDI
nodes	0.690	0.639	0.621	0.662	0.07110
grade__nodes	0.686	0.649	0.624	0.655	0.06580
age__nodes	0.686	0.642	0.621	0.657	0.03346
size__grade	0.686	0.626	0.646	0.655	0.01787
age__size	0.686	0.577	0.649	0.657	0.01534
grade	0.690	0.500	0.653	0.662	0.01340
age__meno	0.686	0.500	0.645	0.657	0.00782
age__pgr	0.686	0.500	0.656	0.657	0.00512
age__grade	0.690	0.507	0.661	0.662	0.00454
meno__grade	0.686	0.500	0.652	0.657	0.00425
nodes__hormon	0.686	0.526	0.658	0.655	0.00280
size	0.690	0.618	0.663	0.662	0.00507
meno__pgr	0.686	0.500	0.657	0.657	0.00316
pgr	0.686	0.500	0.659	0.655	0.00257
meno__nodes	0.686	0.595	0.656	0.657	0.00264
grade__pgr	0.668	0.500	0.627	0.628	0.00241
meno__size	0.690	0.578	0.663	0.662	0.00185

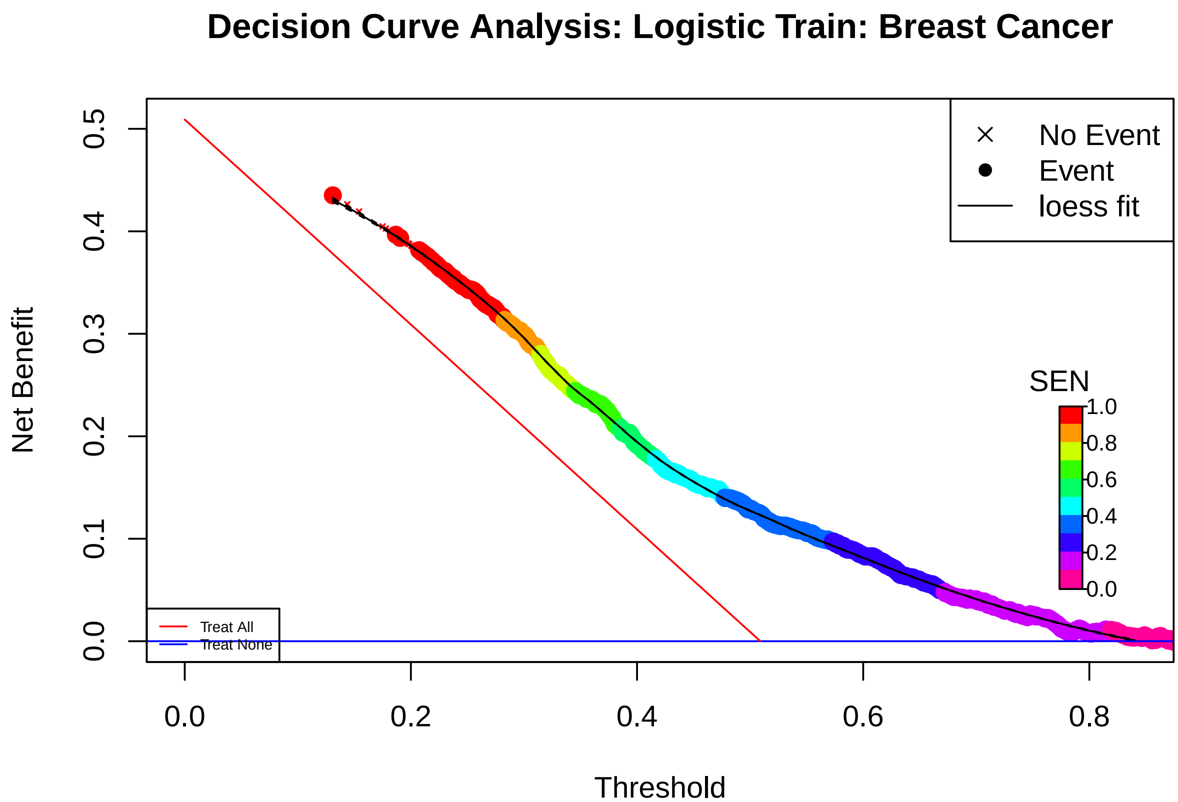
	NRI	z.IDI	z.NRI	Delta.AUC	Frequency
size__nodes	0.63654	17.86	18.870	0.128490	1
nodes	0.57106	14.13	16.179	0.040494	1
grade__nodes	0.54866	13.66	15.650	0.031087	1
age__nodes	0.21312	9.39	5.710	0.035896	1
size__grade	0.29411	6.74	7.728	0.008648	1
age__size	0.29152	6.41	7.652	0.007600	1
grade	0.19036	6.20	4.983	0.008461	1
age__meno	0.08057	4.76	2.337	0.012065	1
age__pgr	0.00745	4.11	0.194	0.000417	1
age__grade	0.11372	3.60	2.960	0.000315	1
meno__grade	0.20428	3.47	5.343	0.004441	1
nodes__hormon	0.45522	3.44	12.150	-0.002853	1
size	0.21050	3.42	5.600	-0.001075	1
meno__pgr	0.05977	3.35	1.558	-0.000429	1
pgr	0.19759	2.64	5.745	-0.004123	1
meno__nodes	-0.06329	2.59	-1.645	0.000631	1
grade__pgr	0.17471	2.55	5.058	0.001252	1
meno__size	0.10227	2.43	2.662	-0.001378	1

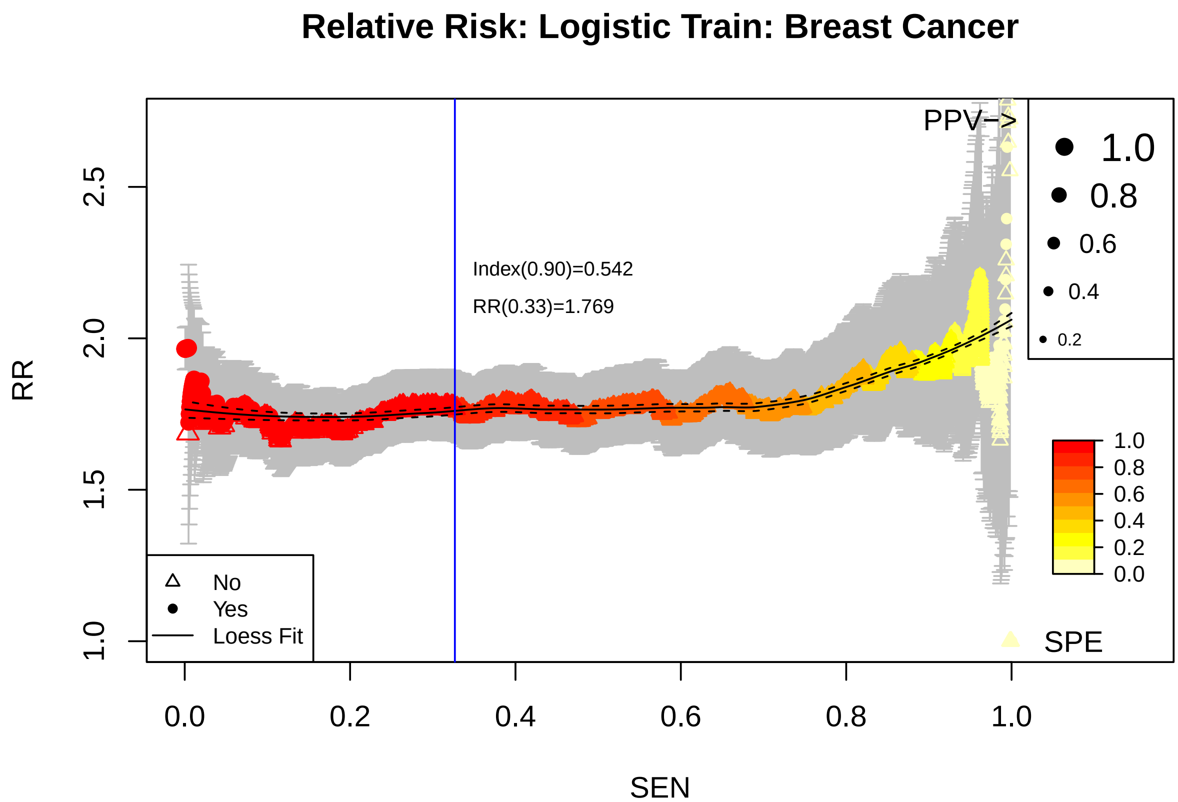
1.7 Logistic Model Performance

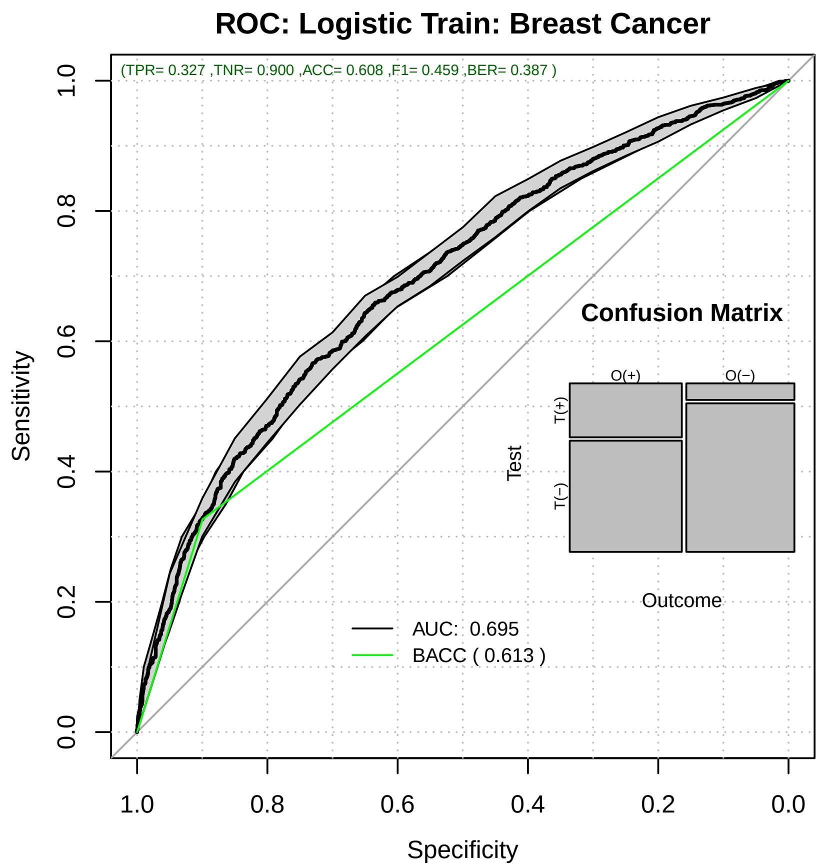
```
op <- par(no.readonly = TRUE)


cprob <- predict(mlog,dataBrestCancerTrain)

rdata <- cbind(dataBrestCancerTrain$status,cprob)
rrAnalysisTrain <- RRPlot(rdata,atProb=c(0.90,0.80),
                    timetoEvent=dataBrestCancerTrain$time,
                    title="Logistic Train: Breast Cancer",
```

```
            ysurvlim=c(0.00,1.0),
            riskTimeInterval=5.0)
```

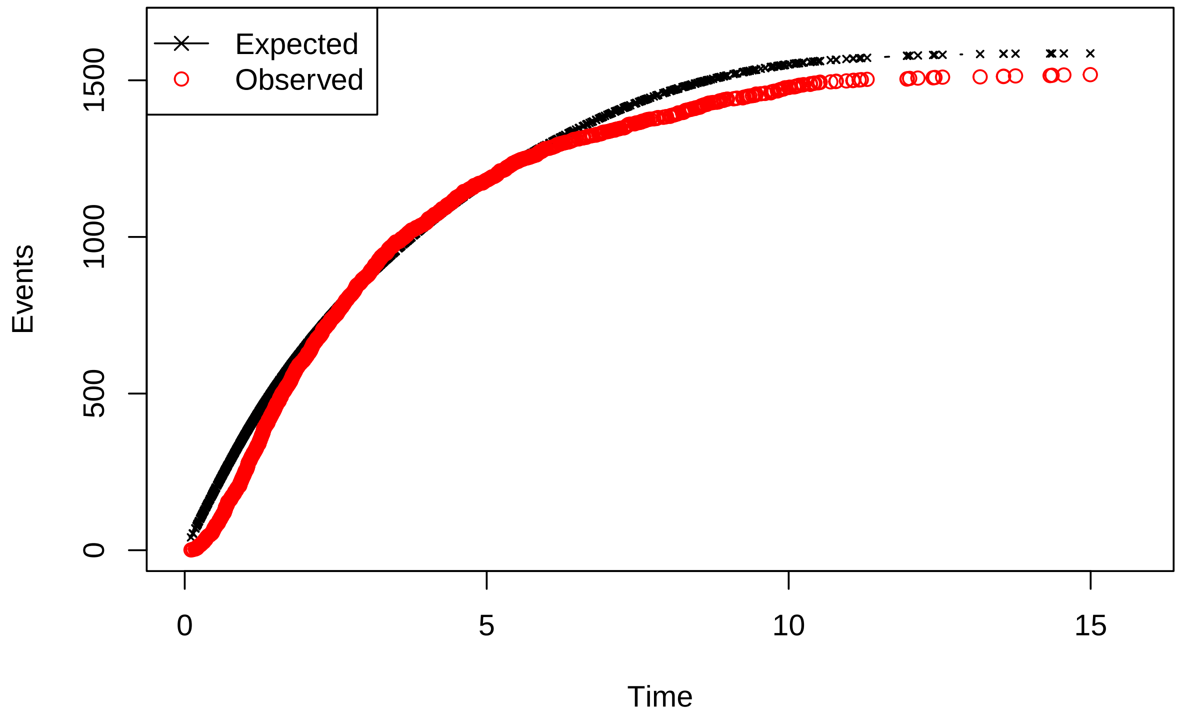


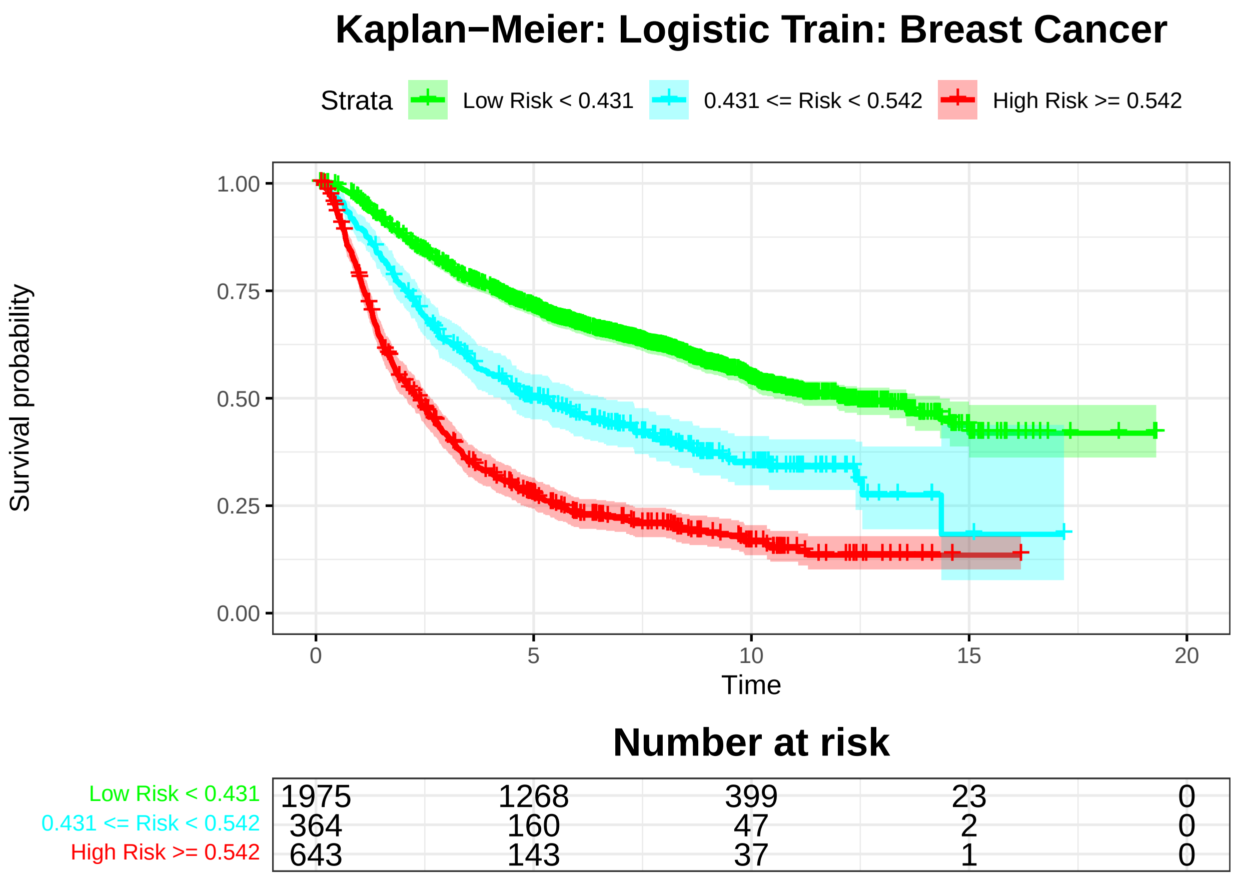

Decision Curve Analysis: Logistic Train: Breast Cancer
Net Benefit
Threshold
0.0
0.1
0.2
0.3
0.4
0.5
0.0
0.2
0.4
0.6
0.8
No Event
Event
loess fit
SEN
1.0
0.8
0.6
0.4
0.2
0.0
Treat All
Treat None


Relative Risk: Logistic Train: Breast Cancer
PPV->
1.0
0.8
0.6
0.4
0.2
Index(0.90)=0.542
RR(0.33)=1.769
RR
2.5
2.0
1.5
1.0
No
Yes
Loess Fit
1.0
0.8
0.6
0.4
0.2
0.0
SPE
0.0
0.2
0.4
0.6
0.8
1.0
SEN


ROC: Logistic Train: Breast Cancer
(TPR= 0.327 ,TNR= 0.900 ,ACC= 0.608 ,F1= 0.459 ,BER= 0.387)
Sensitivity
Specificity
1.0
0.8
0.6
0.4
0.2
0.0
Confusion Matrix
O(+)
O(−)
T(+)
T(−)
Test
Outcome
AUC: 0.695
BACC (0.613)

Time vs. Events: Logistic Train: Breast Cancer






```
par(op)
```

1.7.1 Training Report

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

Table 48: O/E Ratio

est	lower	upper
0.957	0.91	1.01

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

- **C Index**: *0.68*
- **Dxy**: *0.36*
- **S.D.**: *0.014*
- **n**: *2982*
- **missing**: *0*
- **uncensored**: *1518*
- **Relevant Pairs**: *6184528*

- **Concordant**: *4206592*
- **Uncertain**: *2703838*
- **cstatCI**:

mean.C Index	median	lower	upper
0.68	0.68	0.666	0.693

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

Table 50: ROC AUC

est	lower	upper
0.695	0.677	0.714

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

Table 51: Sensitivity

est	lower	upper
0.327	0.303	0.351

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

Table 52: Specificity

est	lower	upper
0.9	0.883	0.915

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

Table 53: Probability Thresholds

90%	80%
0.542	0.431

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

Table 54: Risk Ratio

est	lower	upper
1.77	1.66	1.88

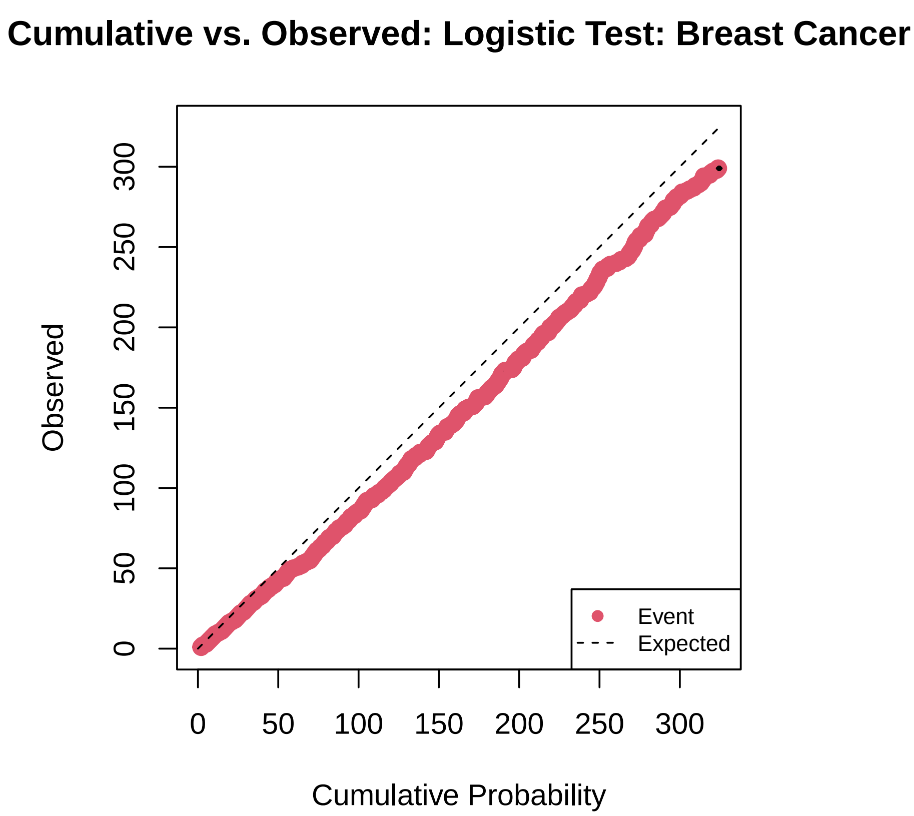
```
pander::pander(rrAnalysisTrain$surdif,caption="Logrank test")
```

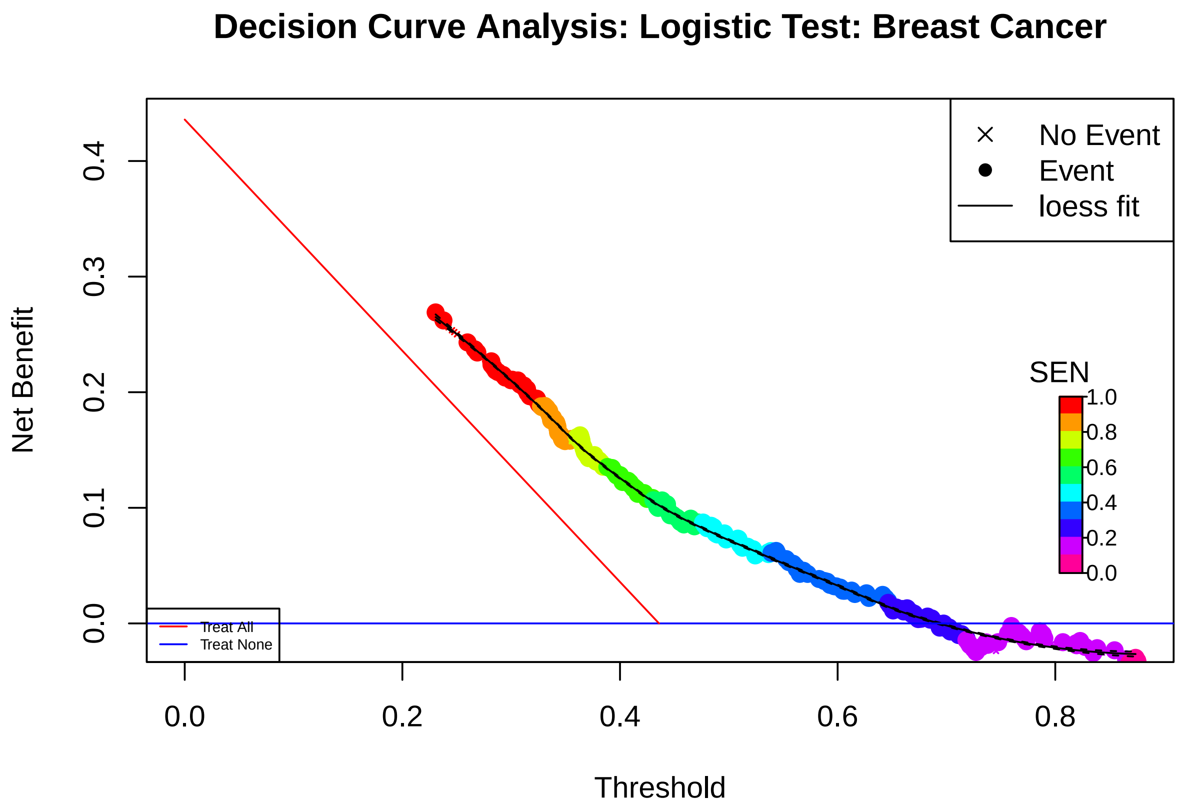
Table 55: Logrank test Chisq = 543.347175 on 2 degrees of freedom, p = 0.000000

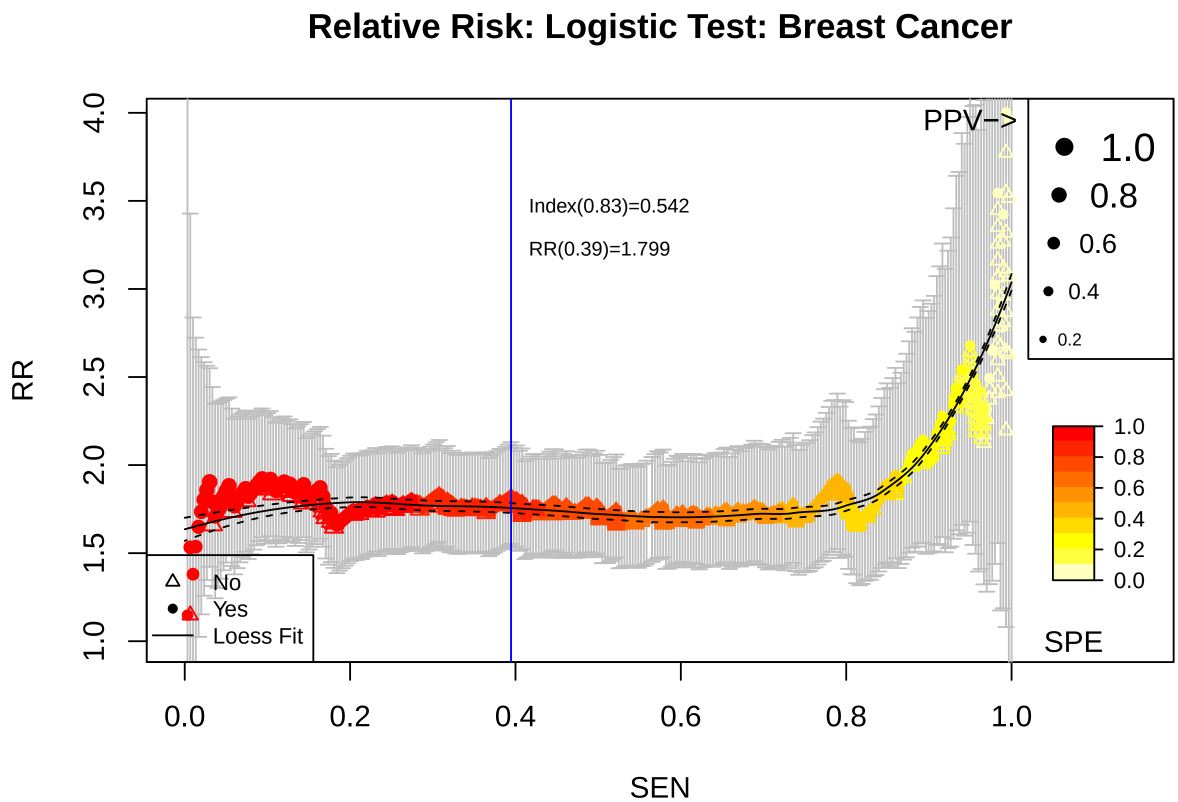
	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	1975	804	1145	101.5	418.9
class=1	364	218	169	14.1	15.9
class=2	643	496	204	418.2	490.7

1.7.2 Results on the validation set using Logistic model

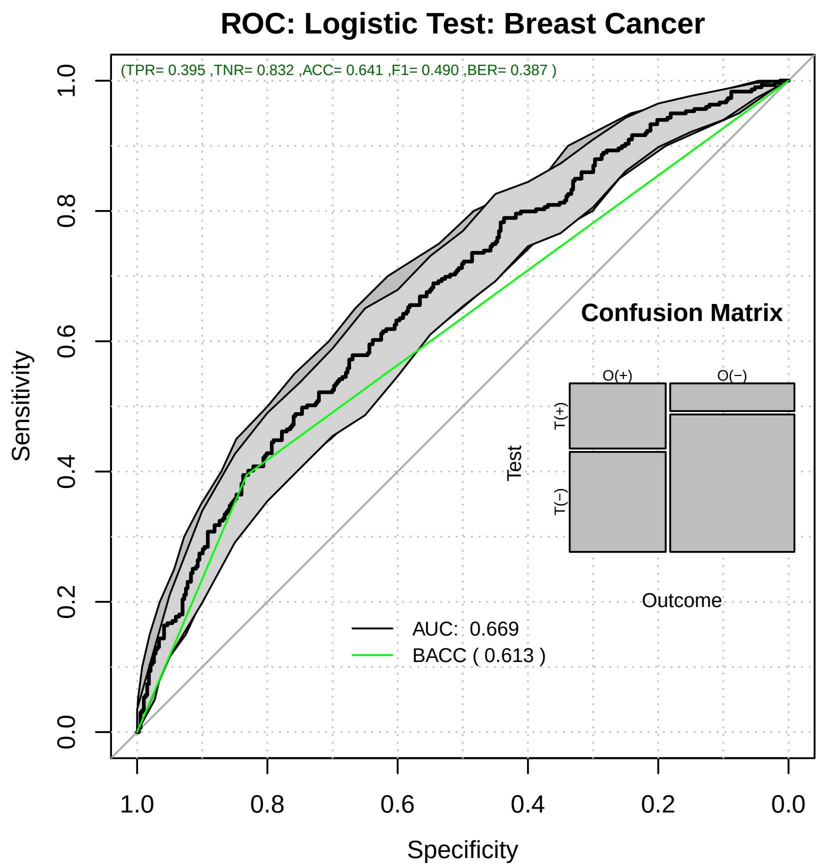
```
pre <- predict(mlog,dataBrestCancerTest)
rdata <- cbind(dataBrestCancerTest$status,pre)

rrAnalysis <- RRPlot(rdata,atThr=rrAnalysisTrain$thr_atP,
                     timetoEvent=dataBrestCancerTest$time,
                     title="Logistic Test: Breast Cancer",
                     ysurvlim=c(0.00,1.0),
                     riskTimeInterval=5)
```

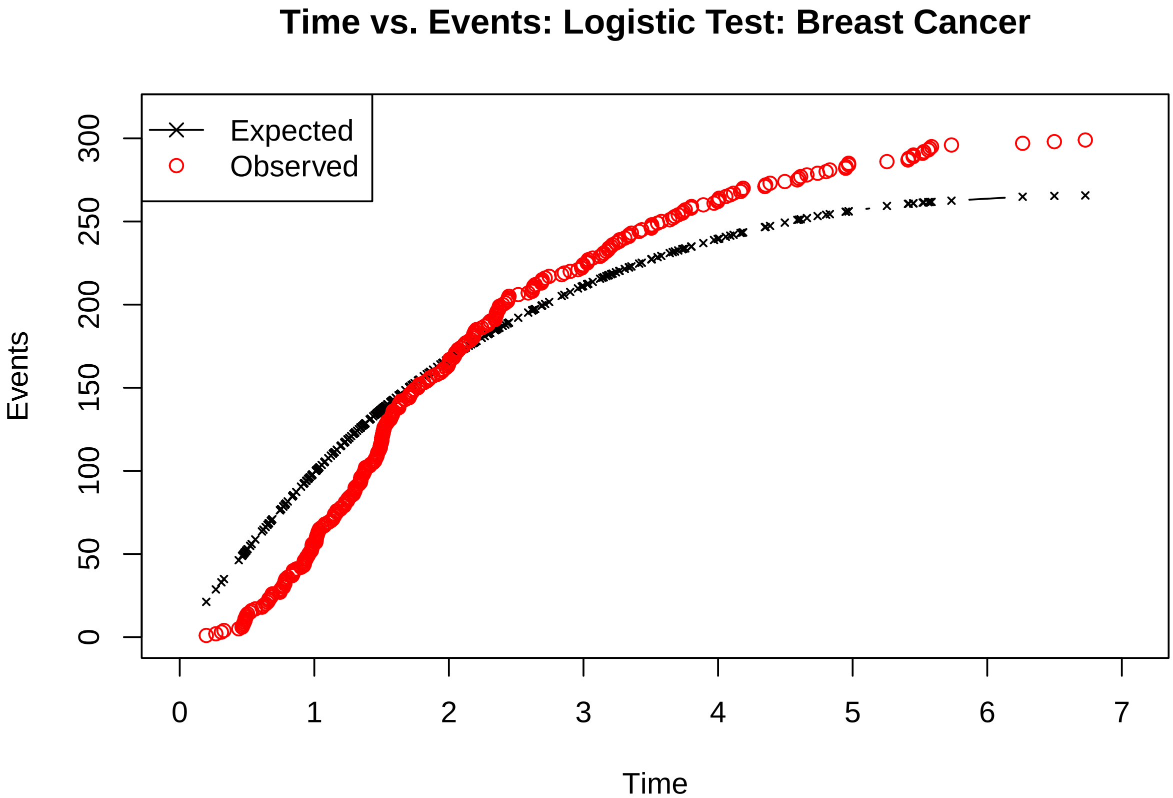



Decision Curve Analysis: Logistic Test: Breast Cancer
Net Benefit
Threshold
No Event
Event
loess fit
SEN
Treat All
Treat None
0.0
0.1
0.2
0.3
0.4
0.0
0.2
0.4
0.6
0.8
1.0
0.8
0.6
0.4
0.2
0.0

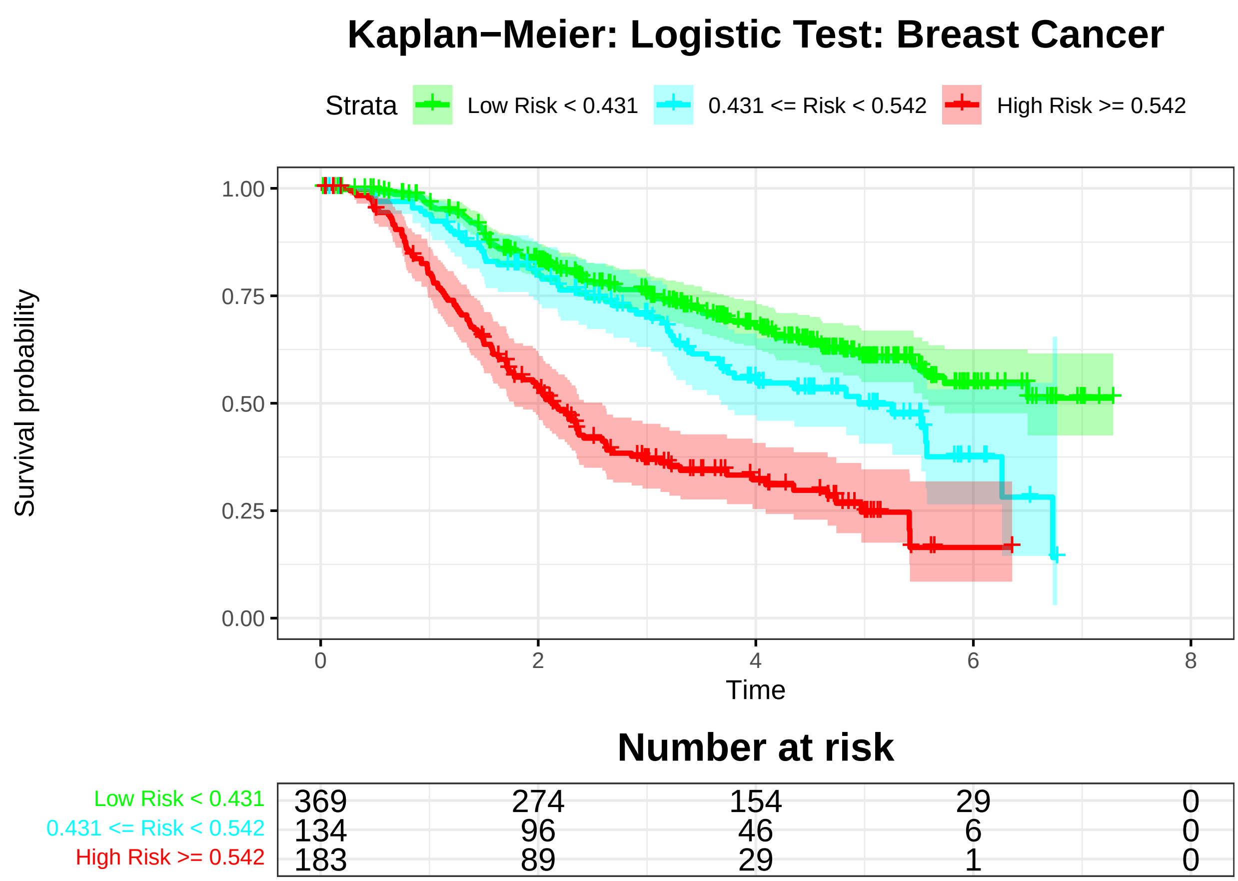


Relative Risk: Logistic Test: Breast Cancer
Index(0.83)=0.542
RR(0.39)=1.799
PPV->
1.0
0.8
0.6
0.4
0.2
SPE
No
Yes
Loess Fit
RR
SEN


ROC: Logistic Test: Breast Cancer
(TPR= 0.395 ,TNR= 0.832 ,ACC= 0.641 ,F1= 0.490 ,BER= 0.387)
Sensitivity
Specificity
0.0
0.2
0.4
0.6
0.8
1.0
1.0
0.8
0.6
0.4
0.2
0.0
Confusion Matrix
O(+)
O(−)
T(+)
T(−)
Test
Outcome
AUC: 0.669
BACC (0.613)



Time vs. Events: Logistic Test: Breast Cancer
Expected
Observed
Events
Time
0
50
100
150
200
250
300
0
1
2
3
4
5
6
7




```
par(op)
```

1.7.3 Validation Report

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

Table 56: O/E Ratio

est	lower	upper
1.13	1	1.26

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

- **C Index**: *0.669*
- **Dxy**: *0.338*
- **S.D.**: *0.0309*
- **n**: *686*
- **missing**: *0*
- **uncensored**: *299*
- **Relevant Pairs**: *266144*

- **Concordant**: *178115*
- **Uncertain**: *203702*
- **cstatCI**:

mean.C Index	median	lower	upper
0.669	0.669	0.64	0.698

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

Table 58: ROC AUC

est	lower	upper
0.669	0.628	0.709

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

Table 59: Sensitivity

est	lower	upper
0.395	0.339	0.453

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

Table 60: Specificity

est	lower	upper
0.832	0.791	0.868

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

Table 61: Probability Thresholds

90%	80%
0.542	0.431

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

Table 62: Risk Ratio

est	lower	upper
1.8	1.54	2.11

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

Table 63: Logrank test Chisq = 92.507991 on 2 degrees of freedom, p = 0.000000

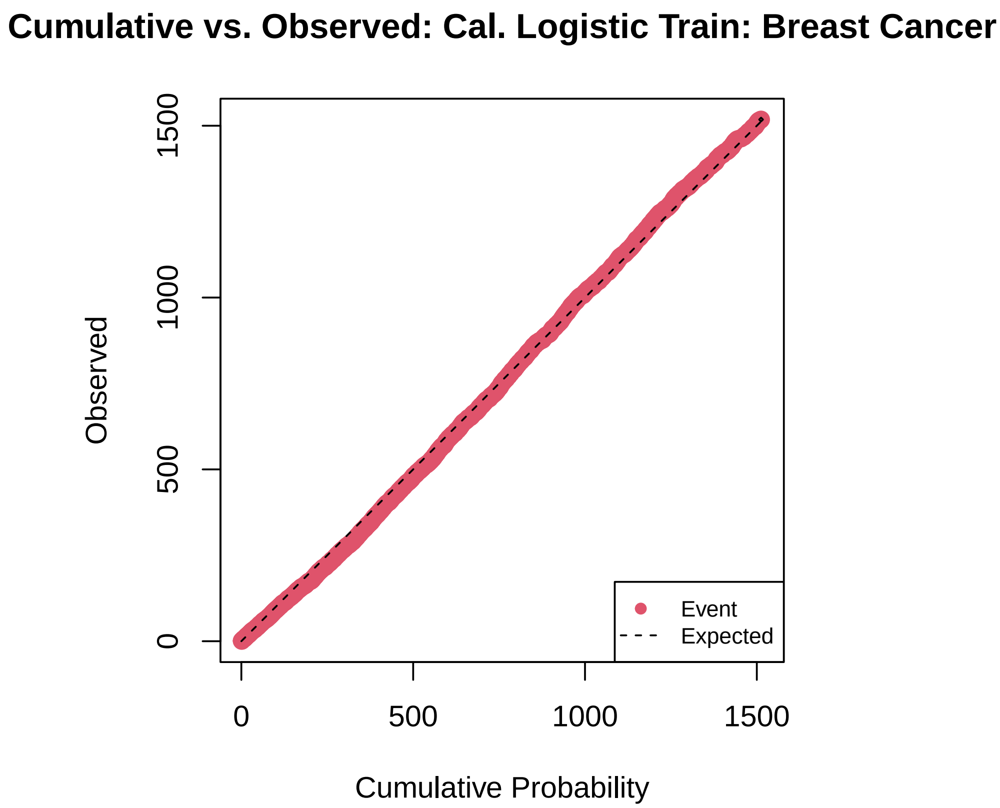
	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	369	121	181.7	20.2997	52.3868
class=1	134	60	61.7	0.0479	0.0604
class=2	183	118	55.5	70.2342	88.0195

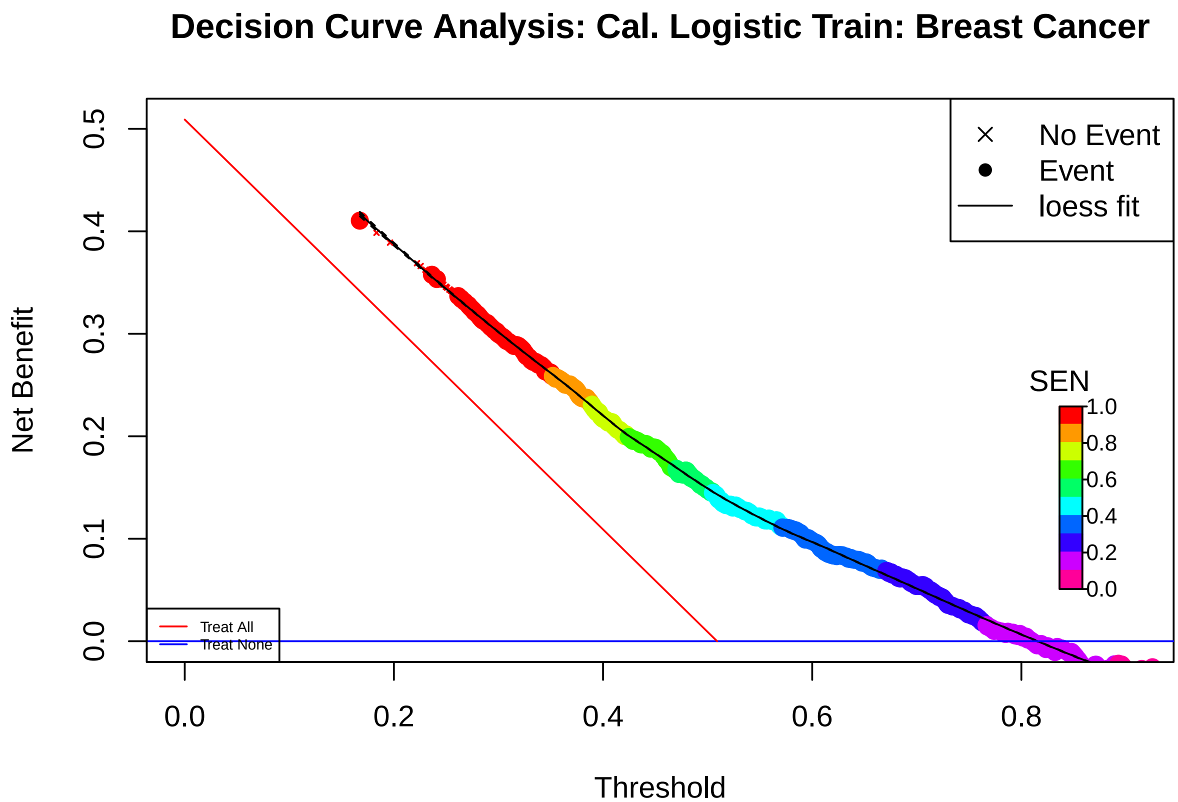
1.8 Logistic Model Poisson Calibration

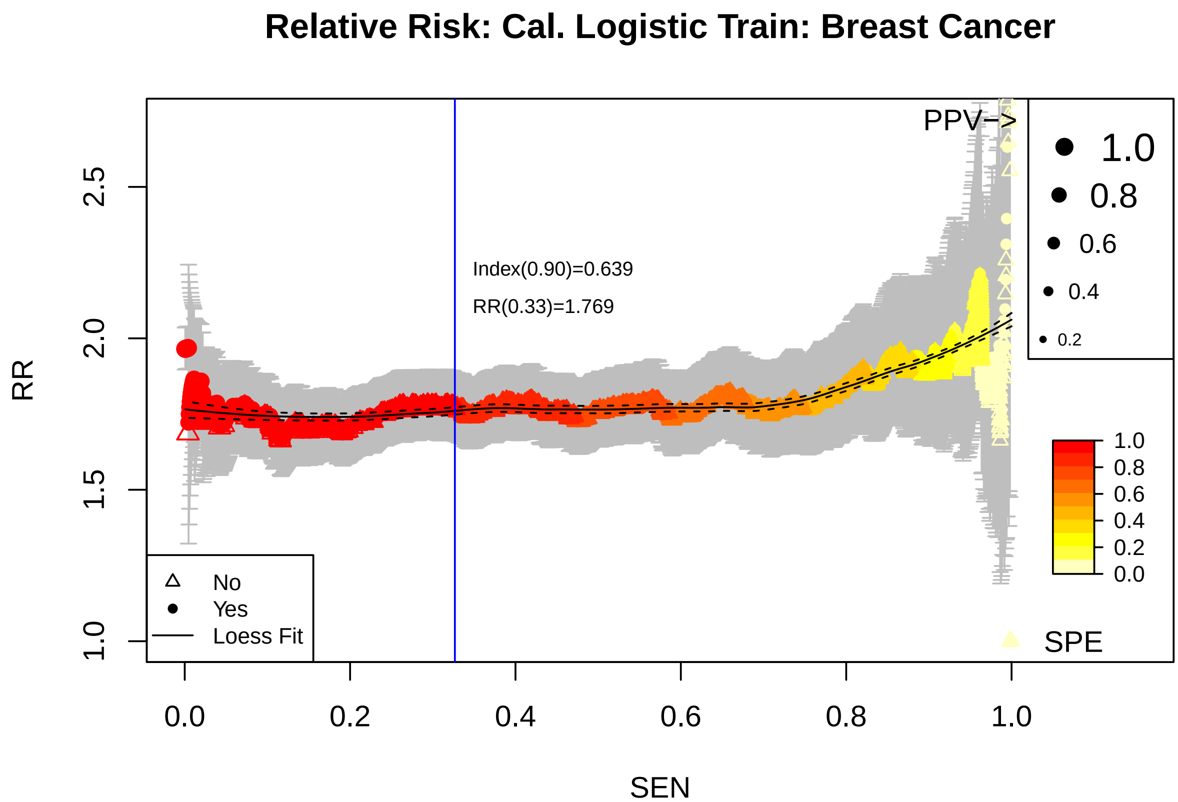
```
riskdata <- cbind(dataBrestCancerTrain$status,predict(mlog,dataBrestCancerTrain,type="prob"),dataBrestCa
calprob <- CalibrationProbPoissonRisk(riskdata)

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

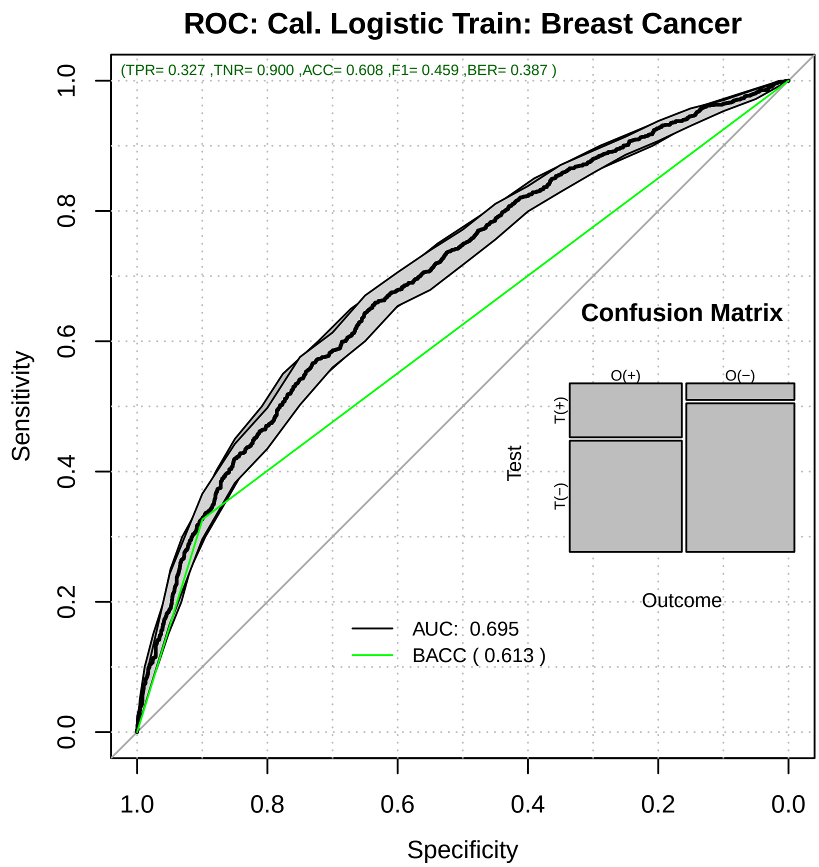
h0	Gain	DeltaTime
0.676	1.31	7.14

```
timeinterval <- calprob$timeInterval;
gain <- calprob$hazardGain

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


rrAnalysisTrain <- RRPlot(rdata,atProb=c(0.90,0.80),
                     timetoEvent=dataBrestCancerTrain$time,
                     title="Cal. Logistic Train: Breast Cancer",
                     ysurvlim=c(0.00,1.0),
                     riskTimeInterval=timeinterval)
```

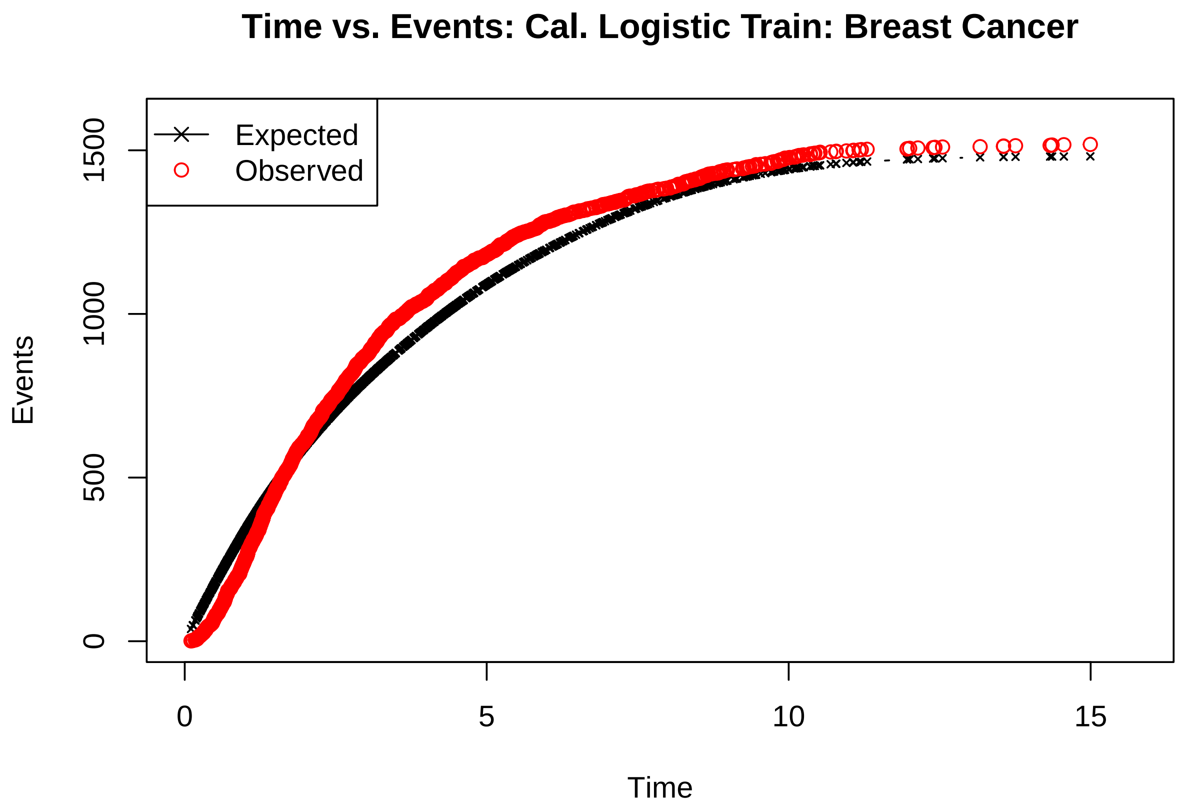

Cumulative vs. Observed: Cal. Logistic Train: Breast Cancer
Observed
0
500
1000
1500
Cumulative Probability
0
500
1000
1500
Event
Expected

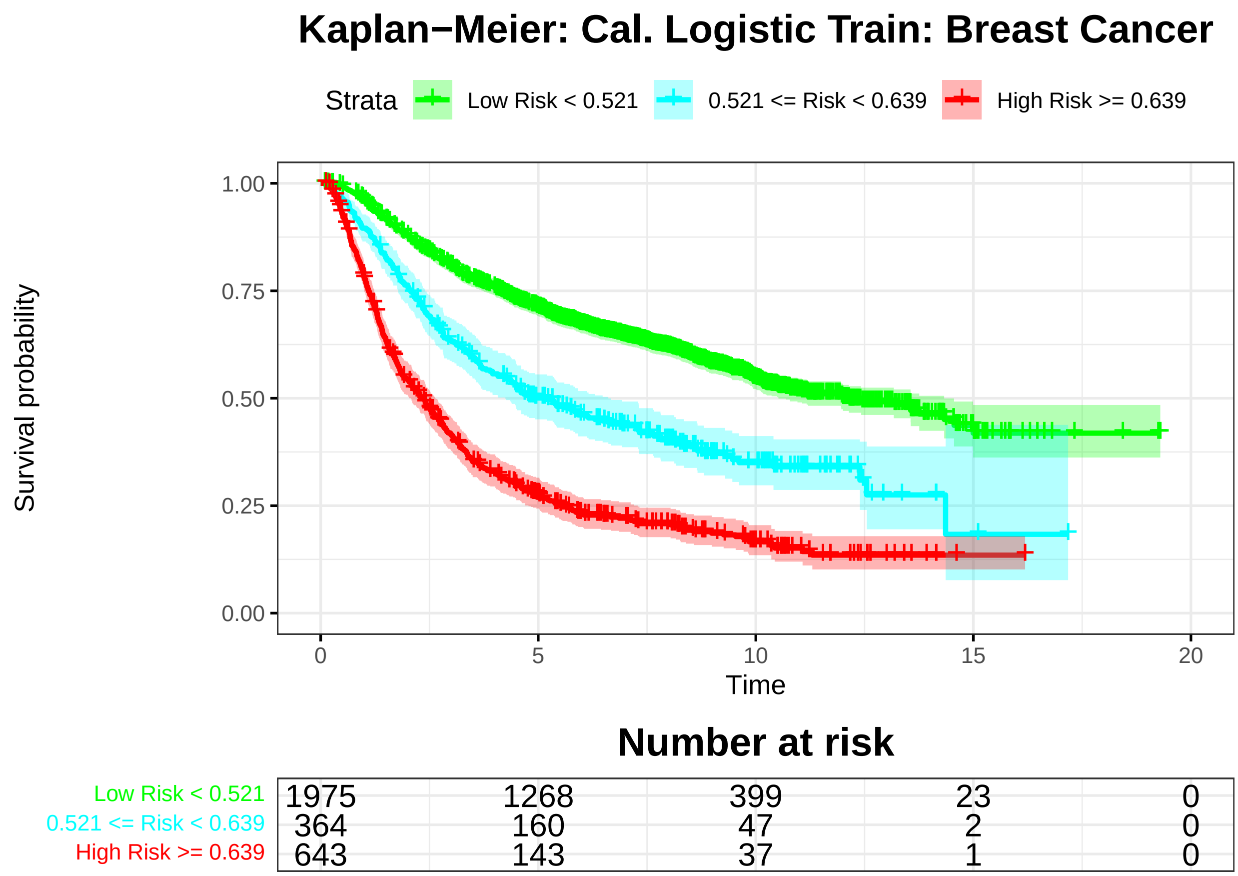

Decision Curve Analysis: Cal. Logistic Train: Breast Cancer
Net Benefit
Threshold
No Event
Event
loess fit
SEN
Treat All
Treat None


Relative Risk: Cal. Logistic Train: Breast Cancer
RR
SEN
PPV->
1.0
0.8
0.6
0.4
0.2
Index(0.90)=0.639
RR(0.33)=1.769
No
Yes
Loess Fit
SPE
1.0
0.8
0.6
0.4
0.2
0.0



ROC: Cal. Logistic Train: Breast Cancer
(TPR= 0.327 ,TNR= 0.900 ,ACC= 0.608 ,F1= 0.459 ,BER= 0.387)
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: 0.695
BACC (0.613)


Time vs. Events: Cal. Logistic Train: Breast Cancer
Expected
Observed
Events
Time
0
500
1000
1500
0
5
10
15




```
par(op)
```

1.8.1 Report of the calibrated logistic: training

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

Table 65: O/E Ratio

est	lower	upper
1.02	0.974	1.08

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

- **C Index**: *0.68*
- **Dxy**: *0.36*
- **S.D.**: *0.014*
- **n**: *2982*
- **missing**: *0*
- **uncensored**: *1518*
- **Relevant Pairs**: *6184528*

- **Concordant**: *4206586*
- **Uncertain**: *2703838*
- **cstatCI**:

mean.C Index	median	lower	upper
0.68	0.68	0.666	0.694

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

Table 67: ROC AUC

est	lower	upper
0.695	0.677	0.714

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

Table 68: Sensitivity

est	lower	upper
0.327	0.303	0.351

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

Table 69: Specificity

est	lower	upper
0.9	0.883	0.915

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

Table 70: Probability Thresholds

90%	80%
0.639	0.521

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

Table 71: Risk Ratio

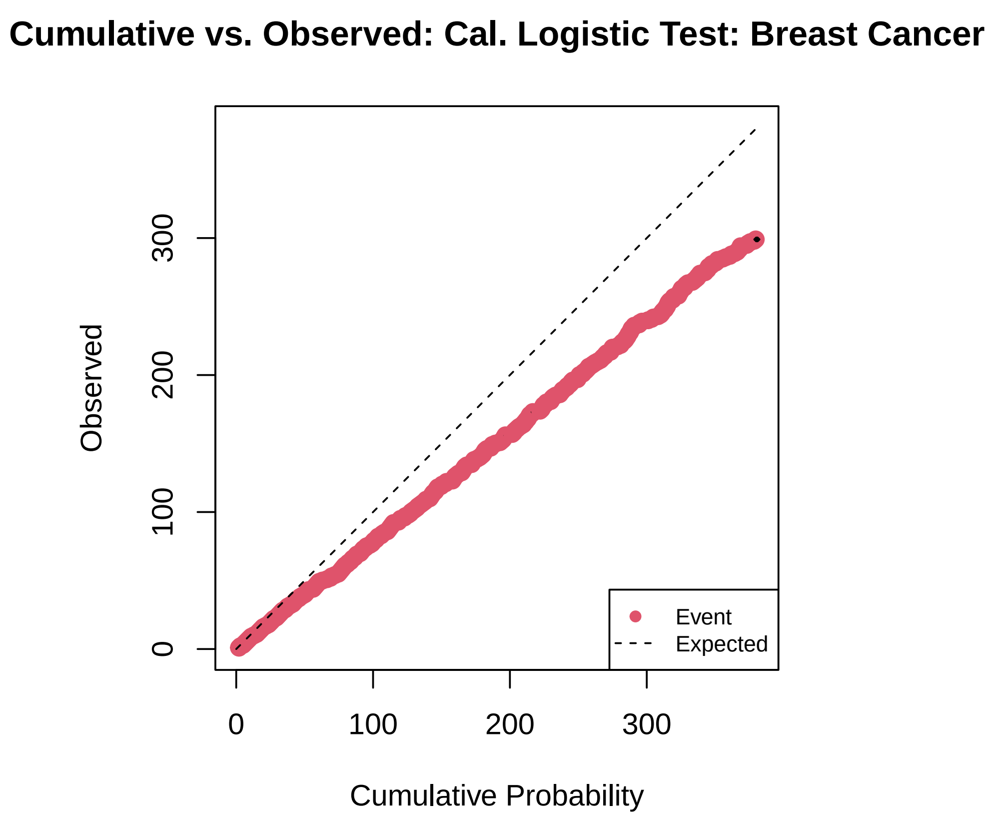
est	lower	upper
1.77	1.66	1.88

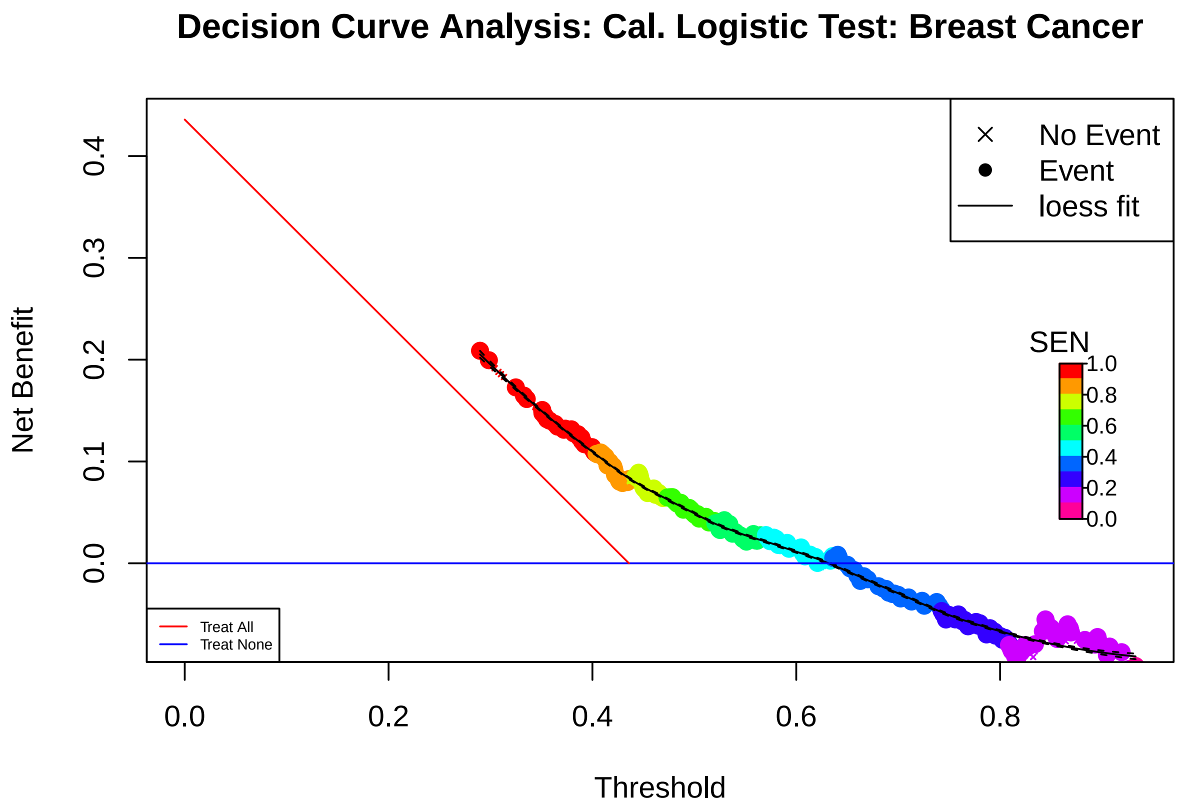
```
pander::pander(rrAnalysisTrain$surdif,caption="Logrank test")
```

Table 72: Logrank test Chisq = 543.347175 on 2 degrees of freedom, p = 0.000000

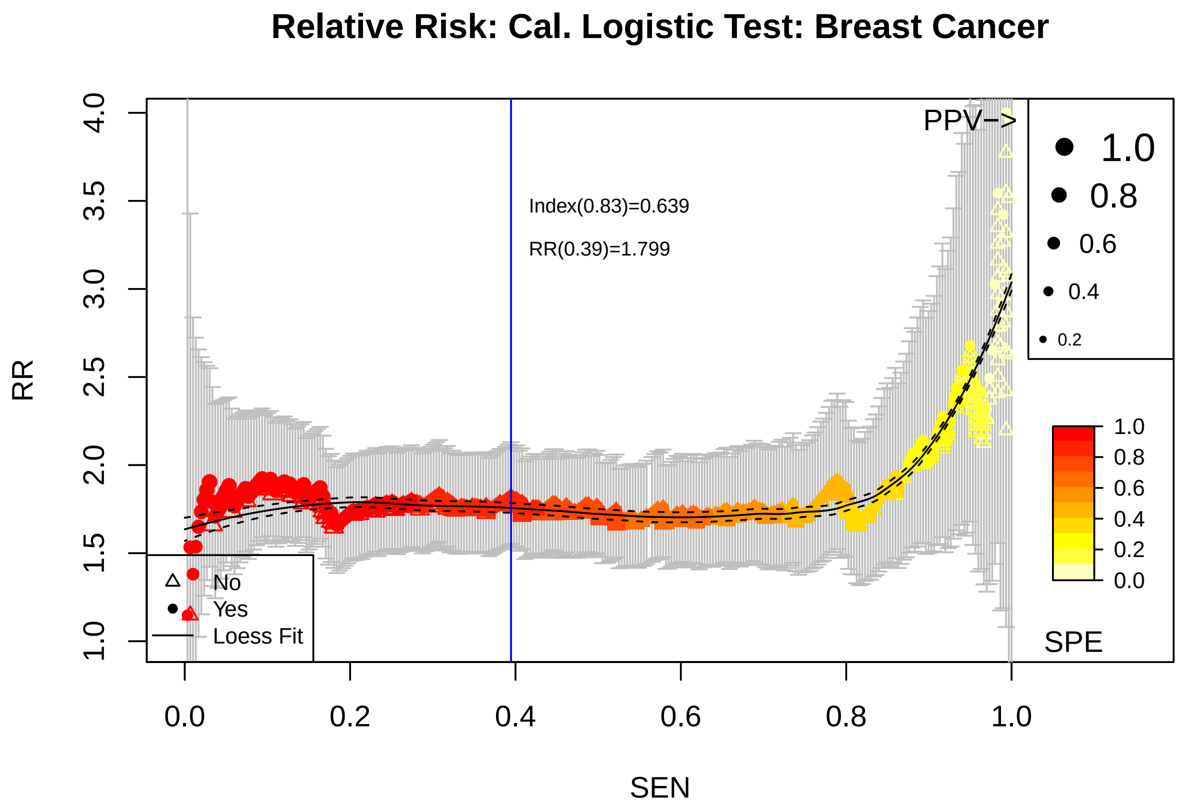
	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	1975	804	1145	101.5	418.9
class=1	364	218	169	14.1	15.9
class=2	643	496	204	418.2	490.7

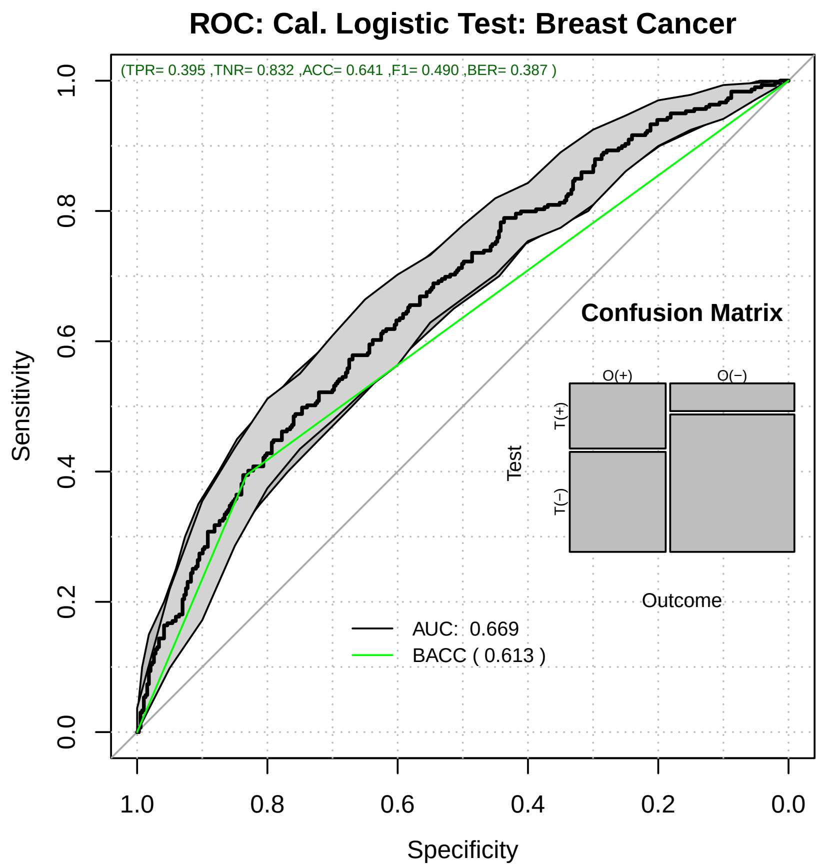
```
probLog <- predict(mlog,dataBrestCancerTest)
aprob <- adjustProb(probLog,gain)

rdata <- cbind(dataBrestCancerTest$status,aprob)
rrAnalysisTestLogistic <- RRPlot(rdata,atThr=rrAnalysisTrain$thr_atP,
                     timetoEvent=dataBrestCancerTest$time,
                     title="Cal. Logistic Test: Breast Cancer",
                     ysurvlim=c(0.00,1.0),
                     riskTimeInterval=timeinterval)
```

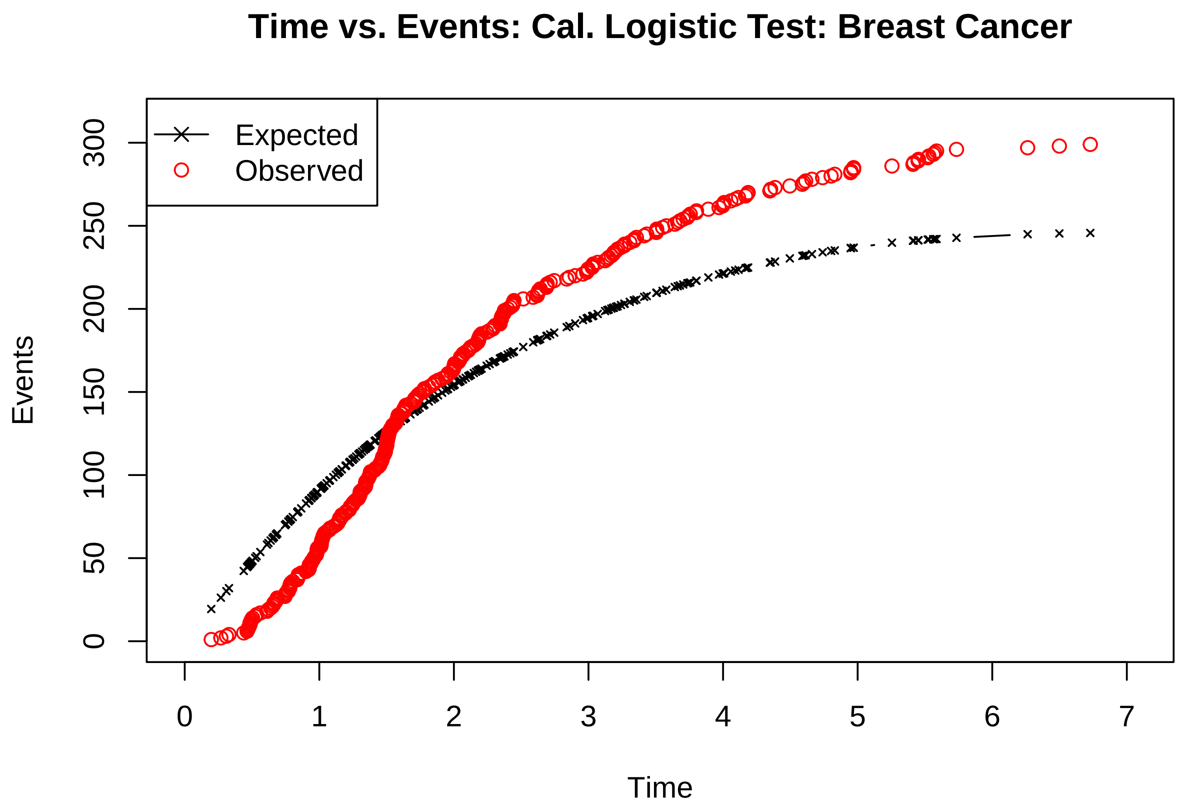




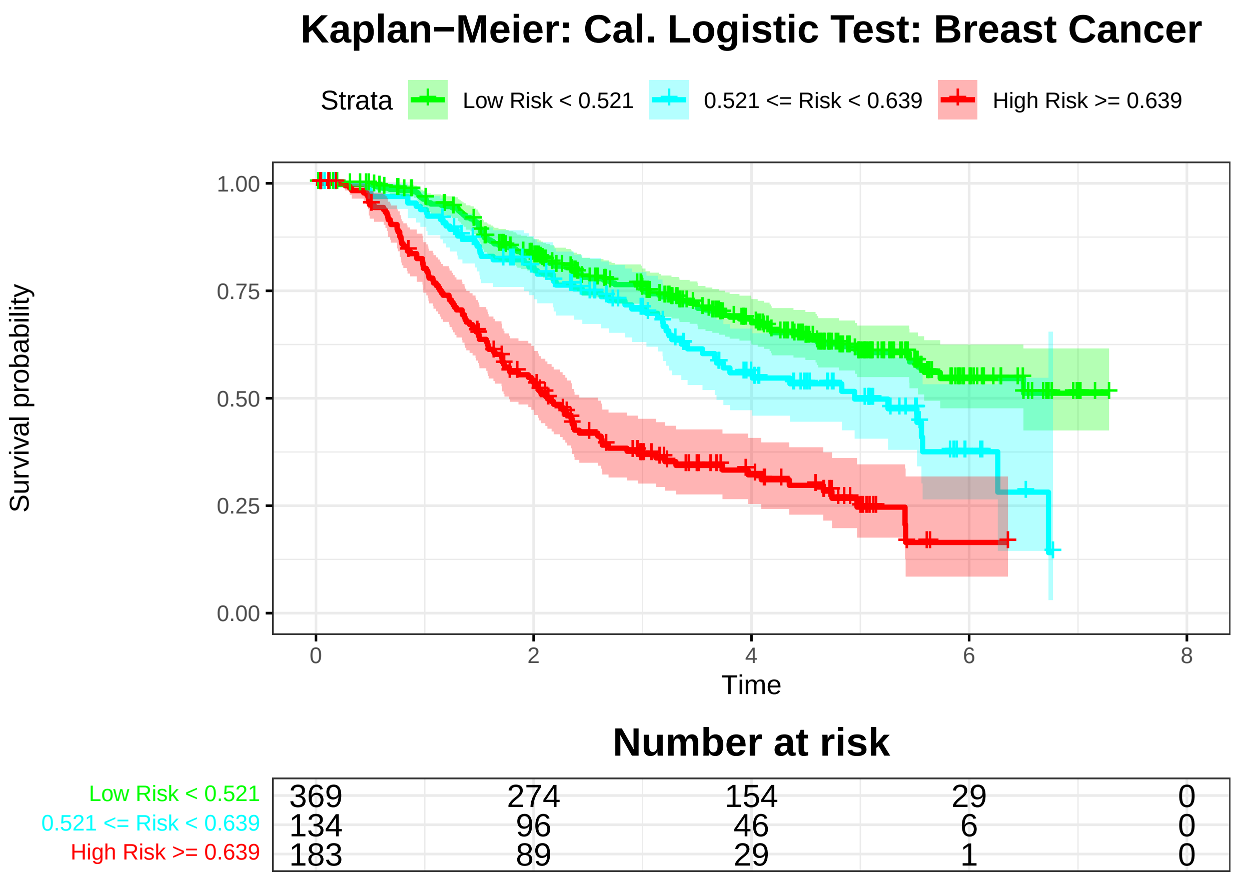
Decision Curve Analysis: Cal. Logistic Test: Breast Cancer
Net Benefit
0.4
0.3
0.2
0.1
0.0
Threshold
0.0
0.2
0.4
0.6
0.8
No Event
Event
loess fit
SEN
1.0
0.8
0.6
0.4
0.2
0.0
Treat All
Treat None


Relative Risk: Cal. Logistic Test: Breast Cancer
Index(0.83)=0.639
RR(0.39)=1.799
PPV−>
1.0
0.8
0.6
0.4
0.2
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
3.0
3.5
4.0
0.0
0.2
0.4
0.6
0.8
1.0
SEN


ROC: Cal. Logistic Test: Breast Cancer
(TPR= 0.395 ,TNR= 0.832 ,ACC= 0.641 ,F1= 0.490 ,BER= 0.387)
Sensitivity
Specificity
1.0
0.8
0.6
0.4
0.2
0.0
Confusion Matrix
O(+)
O(−)
T(+)
T(−)
Test
Outcome
AUC: 0.669
BACC (0.613)



Time vs. Events: Cal. Logistic Test: Breast Cancer
Expected
Observed
Events
Time
0
50
100
150
200
250
300
0
1
2
3
4
5
6
7




```
par(op)
```

1.8.2 Report of the calibrated validation

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

Table 73: O/E Ratio

est	lower	upper
1.22	1.08	1.36

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

- **C Index**: *0.669*
- **Dxy**: *0.338*
- **S.D.**: *0.0309*
- **n**: *686*
- **missing**: *0*
- **uncensored**: *299*
- **Relevant Pairs**: *266144*

- **Concordant**: *178115*
- **Uncertain**: *203702*
- **cstatCI**:

mean.C Index	median	lower	upper
0.669	0.669	0.639	0.701

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

Table 75: ROC AUC

est	lower	upper
0.669	0.628	0.709

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

Table 76: Sensitivity

est	lower	upper
0.395	0.339	0.453

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

Table 77: Specificity

est	lower	upper
0.832	0.791	0.868

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

Table 78: Probability Thresholds

90%	80%
0.639	0.521

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

Table 79: Risk Ratio

est	lower	upper
1.8	1.54	2.11

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

Table 80: Logrank test Chisq = 92.507991 on 2 degrees of freedom, p = 0.000000

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	369	121	181.7	20.2997	52.3868
class=1	134	60	61.7	0.0479	0.0604
class=2	183	118	55.5	70.2342	88.0195

1.9 Comparing the COX and Logistic Models on the Independent Data

```
pander::pander(t(rrCoxTestAnalysis$OAcum95ci))
```

mean	50%	2.5%	97.5%
0.841	0.841	0.839	0.842

```
pander::pander(t(rrAnalysisTestLogistic$OAcum95ci))
```

mean	50%	2.5%	97.5%
0.791	0.791	0.791	0.792

```
pander::pander(t(rrCoxTestAnalysis$OE95ci))
```

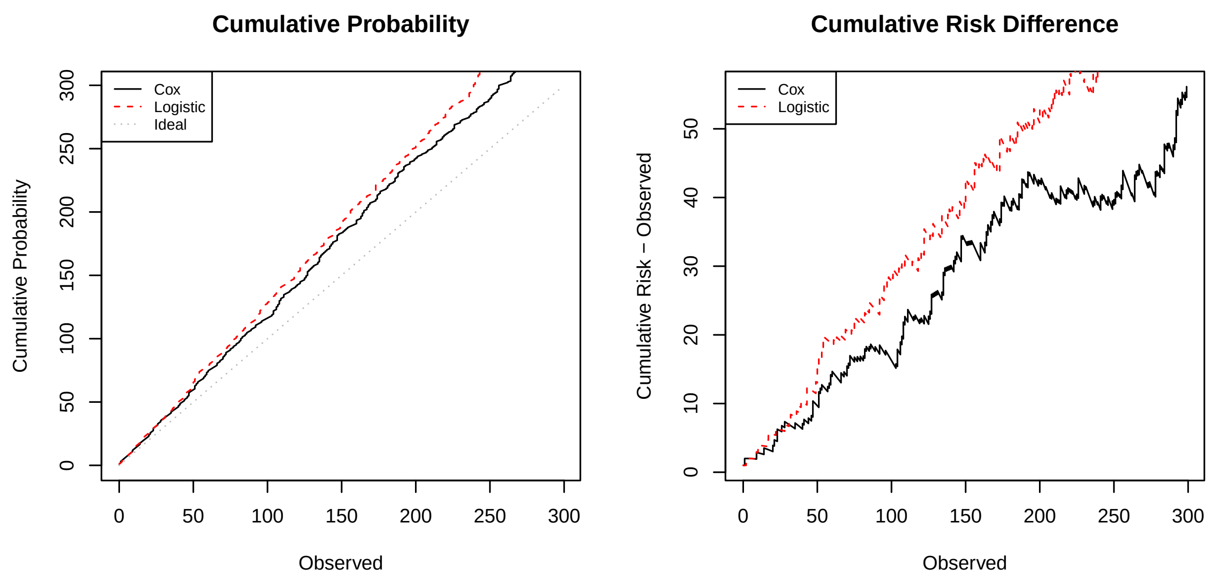
mean	50%	2.5%	97.5%
1.07	1.07	1.04	1.1

```
pander::pander(t(rrAnalysisTestLogistic$OE95ci))
```

mean	50%	2.5%	97.5%
0.955	0.955	0.925	0.986

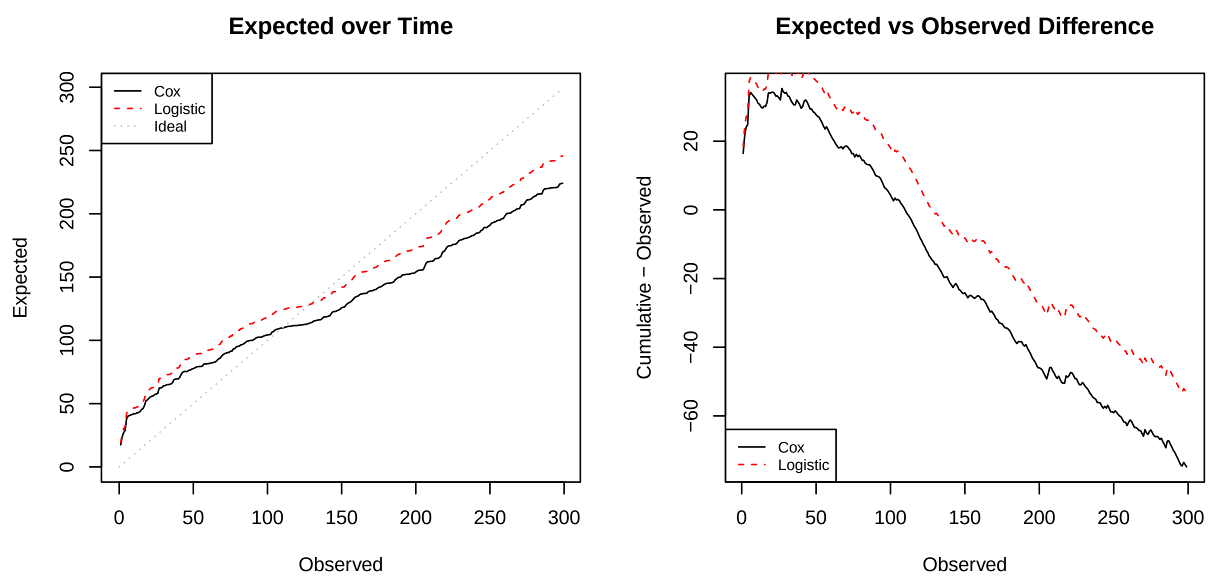
```
maxobs <- sum(dataBrestCancerTest$status)

par(mfrow=c(1,2),cex=0.75)

plot(rrCoxTestAnalysis$CumulativeOvs,type="l",lty=1,
     main="Cumulative Probability",
     xlab="Observed",
     ylab="Cumulative Probability",
     ylim=c(0,maxobs),
     xlim=c(0,maxobs))
lines(rrAnalysisTestLogistic$CumulativeOvs,lty=2,col="red")
lines(x=c(0,maxobs),y=c(0,maxobs),lty=3,col="gray")
legend("topleft",legend = c("Cox","Logistic","Ideal"),
       col=c("black","red","gray"),
       lty=c(1,2,3),
       cex=0.75
```

```
)


plot(rrCoxTestAnalysis$CumulativeOvs$Observed,
     rrCoxTestAnalysis$CumulativeOvs$Cumulative-
       rrCoxTestAnalysis$CumulativeOvs$Observed,
     main="Cumulative Risk Difference",
     xlab="Observed",
     ylab="Cumulative Risk - Observed",
     type="l",
     lty=1)
lines(rrAnalysisTestLogistic$CumulativeOvs$Observed,
     rrAnalysisTestLogistic$CumulativeOvs$Cumulative-
       rrAnalysisTestLogistic$CumulativeOvs$Observed,
     lty=2,
     col="red")
legend("topleft",legend = c("Cox","Logistic"),
       col=c("black","red"),
       lty=c(1,2),
       cex=0.75
)
```




```
plot(rrCoxTestAnalysis$OEData[,2:3],type="l",lty=1,
     main="Expected over Time",
     xlab="Observed",
     ylab="Expected",
     ylim=c(0,maxobs),
     xlim=c(0,maxobs))
lines(rrAnalysisTestLogistic$OEData[,2:3],lty=2,col="red")
lines(x=c(0,maxobs),y=c(0,maxobs),lty=3,col="gray")
legend("topleft",legend = c("Cox","Logistic","Ideal"),
       col=c("black","red","gray"),
       lty=c(1,2,3),
```

```
        cex=0.75
)

plot(rrCoxTestAnalysis$OEData$Observed,
     rrCoxTestAnalysis$OEData$Expected-
       rrCoxTestAnalysis$OEData$Observed,
     main="Expected vs Observed Difference",
     xlab="Observed",
     ylab="Cumulative - Observed",
     type="l",
     lty=1)
lines(rrAnalysisTestLogistic$OEData$Observed,
     rrAnalysisTestLogistic$OEData$Expected-
       rrAnalysisTestLogistic$OEData$Observed,
     lty=2,col="red")

legend("bottomleft",legend = c("Cox","Logistic"),
        col=c("black","red"),
        lty=c(1,2),
        cex=0.75
)
```




```
par(op)
```