

Risk-Evaluation: Breast Cancer Royston-Altman

Jose Tamez

2023-04-26

Contents

1	Evaluation of RISK survival models	1
1.1	The libraries	1
1.2	Breast Cancer Royston-Altman data	2
1.3	Cox Modeling	3
1.4	Cox Model Performance	4
1.5	Performance on the external data set	20
1.6	Logistic Model	36
1.7	Logistic Model Performance	37
1.8	Logistic Model Poisson Calibration	52
1.9	Comparing the COX and Logistic Models on the Independent Data	67

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)
layout(matrix(1:1, nrow=1))
```

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

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

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

```
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), "status"])

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)
```

	Estimate	lower HR	upper	u.Accuracy	Accuracy	full Accuracy	AUC	AUC	full AUC	AUCI	NRI	z.IDI	z.NRI	Delta AUC	Frequency
age_nodes	0.00716	0.001	1.001	1.001	0.626	0.600	0.632	0.630	0.601	0.634	0.0304	0.4594	12.81	14.37	0.0330561
size_grade	0.05649	0.005	1.006	1.006	0.598	0.623	0.632	0.599	0.626	0.634	0.01868	0.3914	9.82	11.29	0.0079471
nodes	0.08658	0.082	1.090	1.099	0.637	0.642	0.643	0.640	0.643	0.644	0.00740	0.0564	8.33	1.66	0.0001481
size	0.00688	0.005	1.007	1.009	0.595	0.641	0.643	0.595	0.642	0.644	0.01440	0.3587	8.05	9.97	0.0013221

	Estimate	lower HR	upper	Accuracy	Accuracy	Accuracy	AUC	AUC	Full.AUC	DI	NRI	z.IDI	z.NRI	Delta.AUC	Frequency
size_nodes	1.000	1.000	1.000	0.624	0.643	0.643	0.629	0.644	0.644	0.00340	0.343	07.25	9.57	-	1
	0.000378													0.000377	
age_size	-	1.000	1.000	0.567	0.627	0.632	0.568	0.630	0.634	0.00630	0.193	55.95	5.36	0.0040781	
	0.000149														
grade	0.204934	1.146	1.227	1.314	0.565	0.637	0.643	0.561	0.638	0.644	0.00920	0.206	95.88	6.31	0.0053441
age	-	0.996	0.997	0.998	0.513	0.628	0.643	0.513	0.628	0.644	0.00410	0.091	75.27	2.51	0.0154651
	0.003113														
grade_nodes	0.981	0.986	0.992	0.635	0.645	0.643	0.639	0.646	0.644	0.00207	-	5.03	-	-	1
	0.013784										0.0910		2.55	0.002609	

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")
```

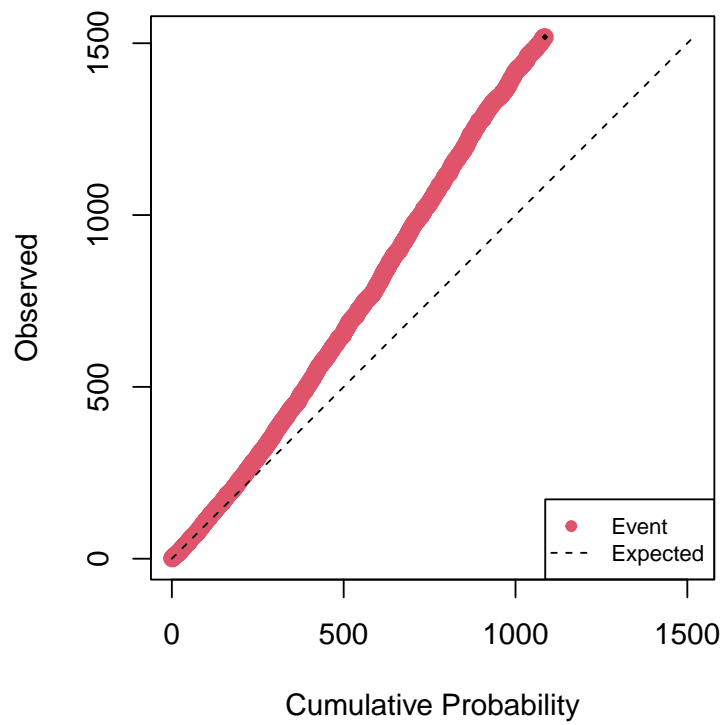
Table 4: Initial Parameters

h0	timeinterval
0.429	5

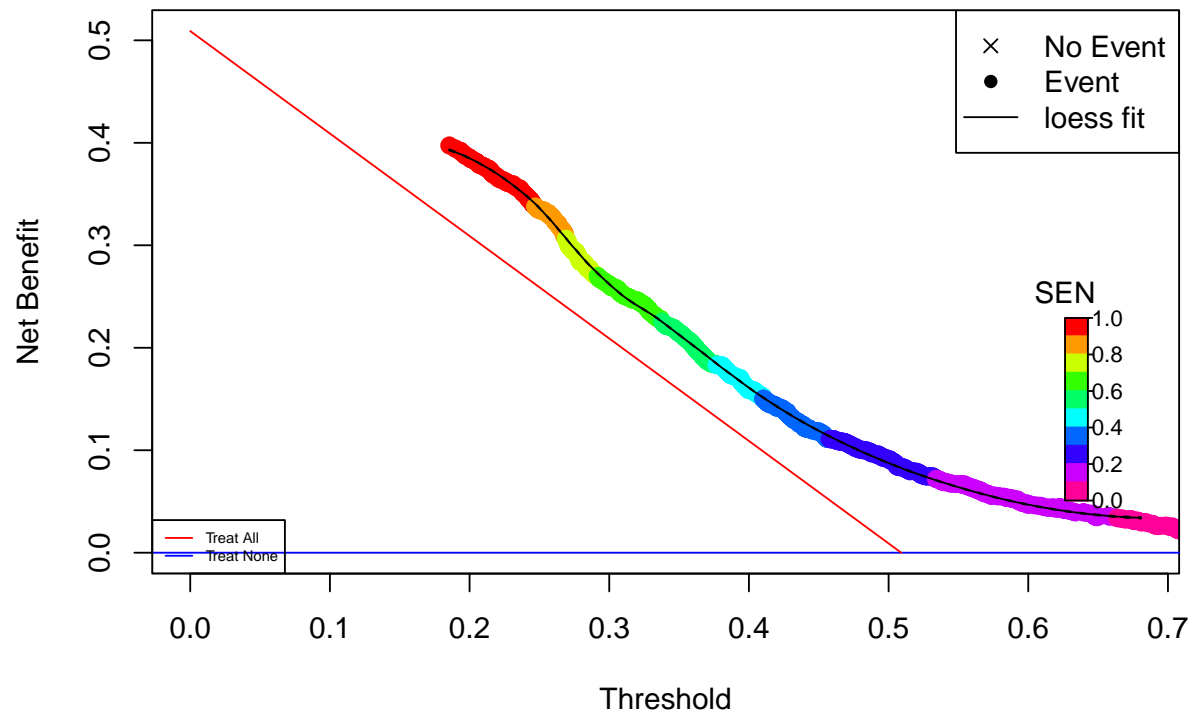
```
index <- predict(ml,dataBrestCancerTrain)
rdata <- cbind(dataBrestCancerTrain$status,ppoisGzero(index,h0))

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

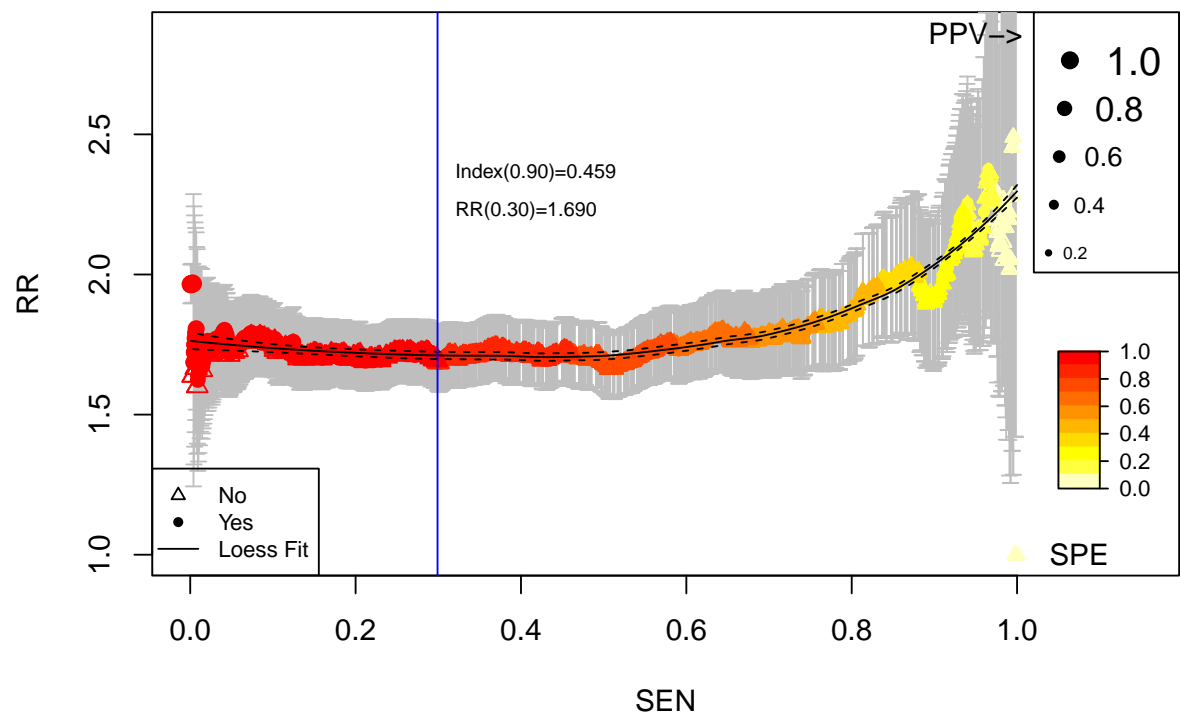
Cumulative vs. Observed: Train: Breast Cancer

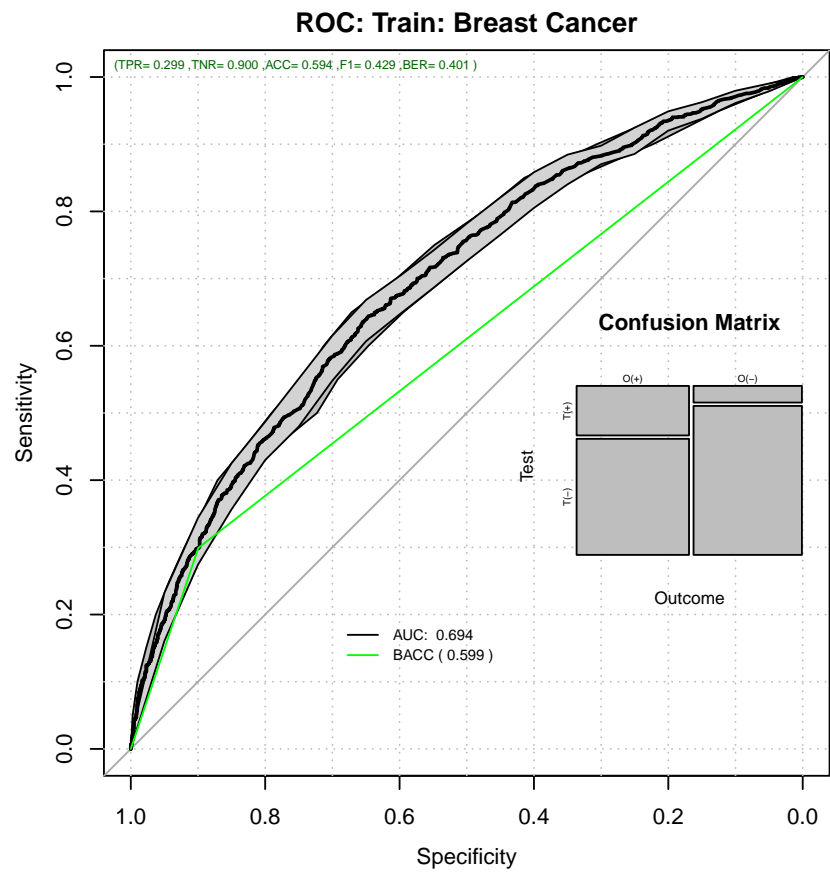


Decision Curve Analysis: Train: Breast Cancer

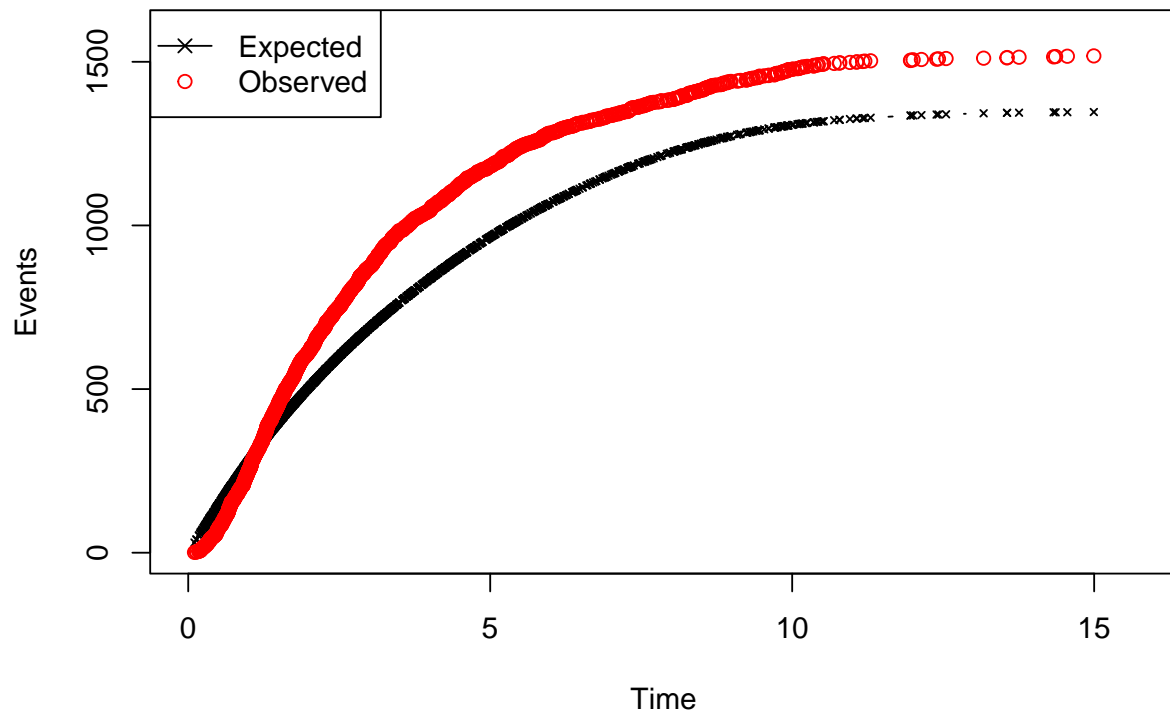


Relative Risk: Train: Breast Cancer

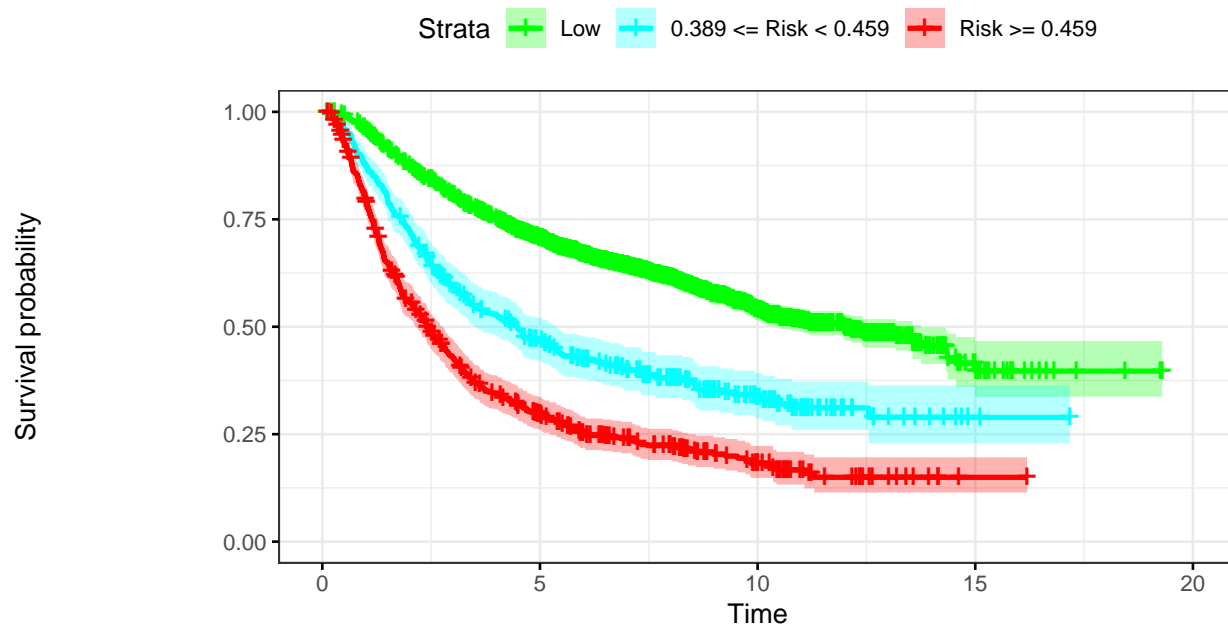




Time vs. Events: Train: Breast Cancer



Kaplan–Meier: Train: Breast Cancer



Number at risk

Low	1985	1260	393	23	0
$0.389 \leq \text{Risk} < 0.459$	396	166	51	2	0
$\text{Risk} \geq 0.459$	601	145	39	1	0

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 5: O/E Ratio

est	lower	upper
1.13	1.07	1.19

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

Table 6: 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 7: 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	0.676	0.662	0.69

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

Table 9: ROC AUC

est	lower	upper
0.694	0.675	0.713

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

Table 10: Sensitivity

est	lower	upper
0.299	0.276	0.323

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

Table 11: Specificity

est	lower	upper
0.9	0.883	0.915

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

Table 12: Probability Thresholds

90%	80%
0.459	0.389

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

Table 13: Risk Ratio

est	lower	upper
1.69	1.59	1.8

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

Table 14: 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

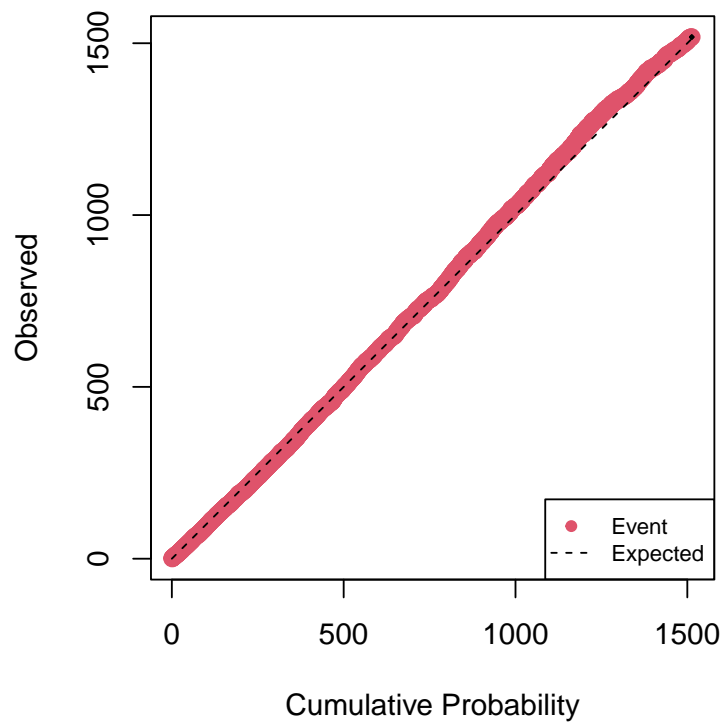
1.4.4 The RRplot() of the calibrated model

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

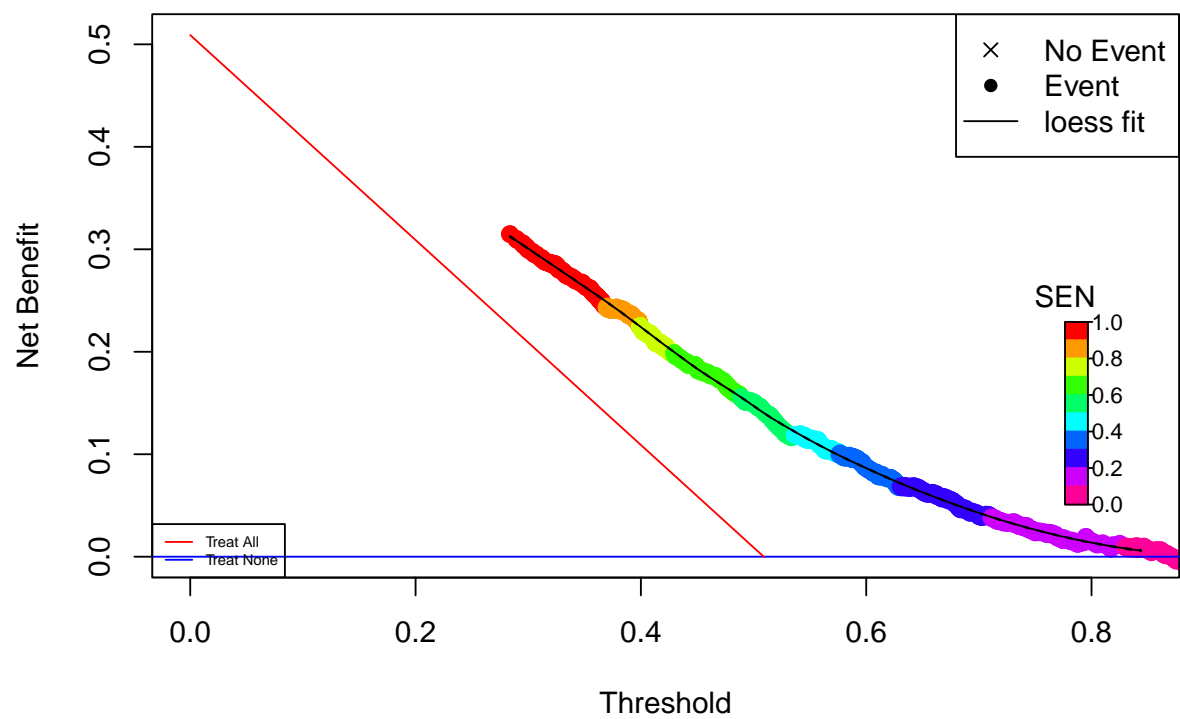
rdata <- cbind(dataBrestCancerTrain$status, calprob$prob)

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)
```

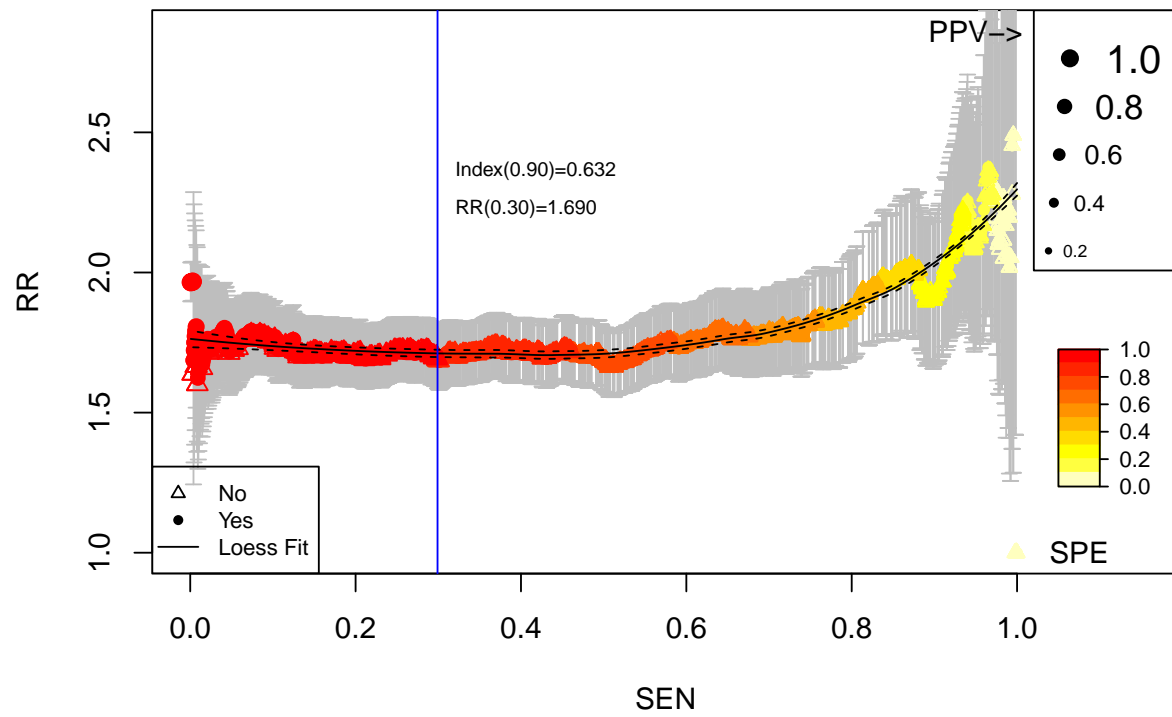
Cumulative vs. Observed: Cal. Train: Breast Cancer

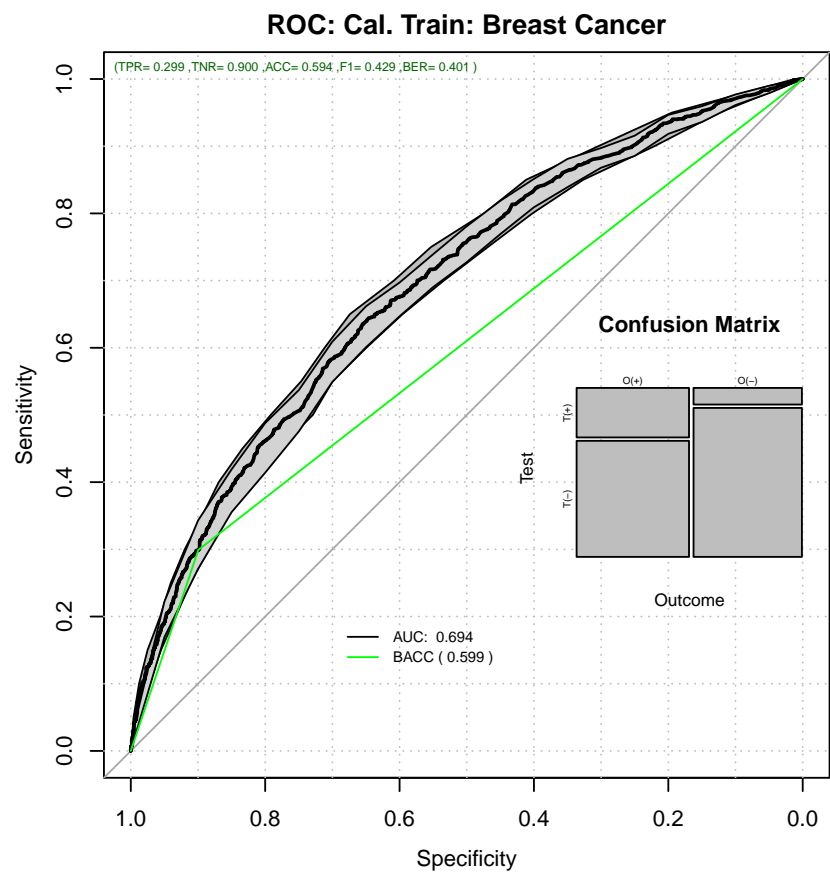


Decision Curve Analysis: Cal. Train: Breast Cancer

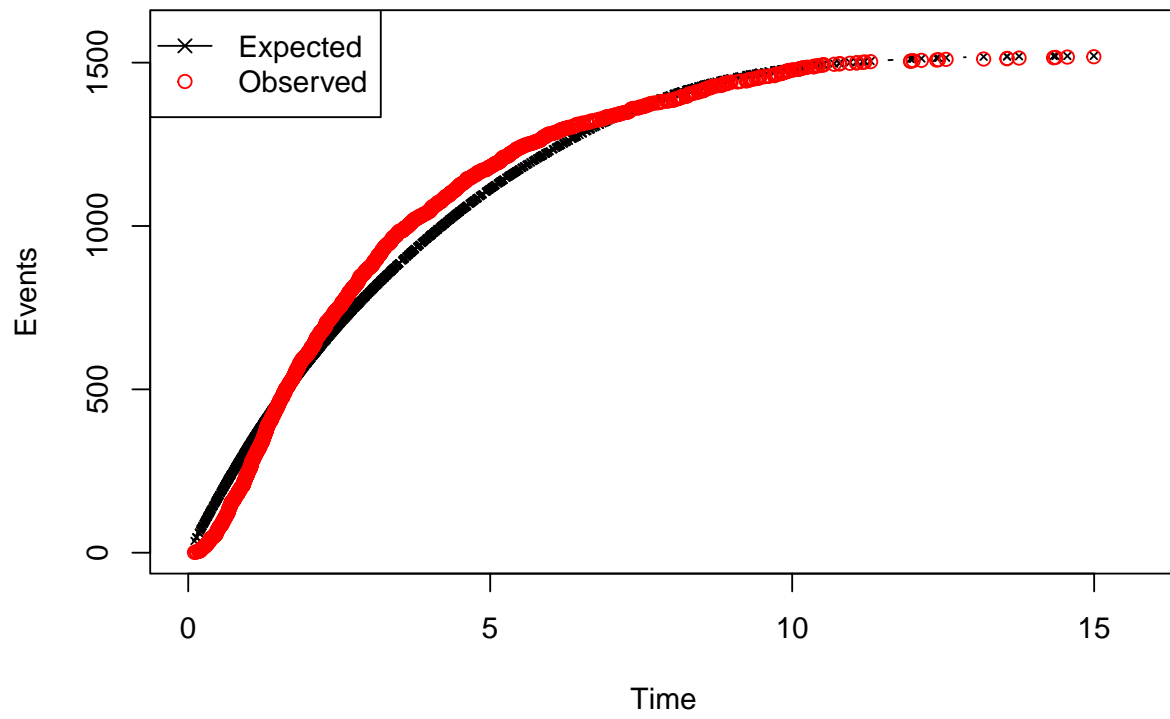


Relative Risk: Cal. Train: Breast Cancer

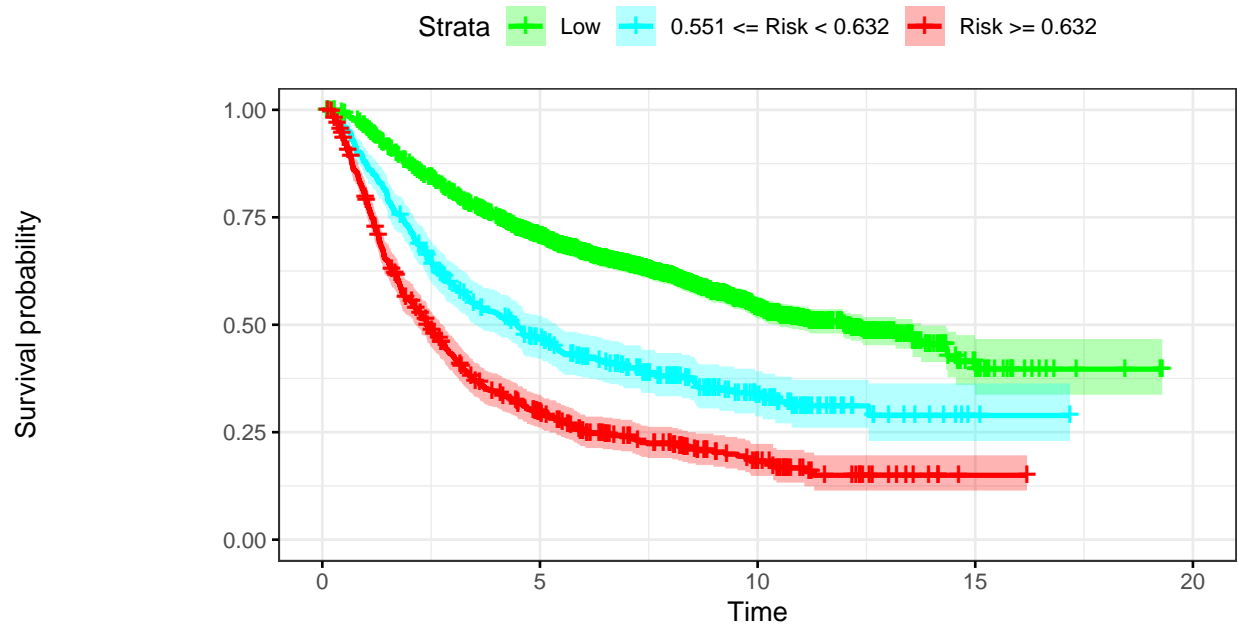




Time vs. Events: Cal. Train: Breast Cancer



Kaplan–Meier: Cal. Train: Breast Cancer



Number at risk

Low	1985	1260	393	23	0
$0.551 \leq \text{Risk} < 0.632$	396	166	51	2	0
$\text{Risk} \geq 0.632$	601	145	39	1	0

1.4.5 Calibrated Train Performance

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

Table 16: O/E Ratio

est	lower	upper
0.998	0.949	1.05

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

Table 17: 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 18: 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$ROCAalysis$aucs),caption="ROC AUC")
```

Table 20: ROC AUC

est	lower	upper
0.694	0.675	0.713

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

Table 21: Sensitivity

est	lower	upper
0.299	0.276	0.323

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

Table 22: Specificity

est	lower	upper
0.9	0.883	0.915

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

Table 23: Probability Thresholds

90%	80%
0.632	0.551

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

Table 24: Risk Ratio

est	lower	upper
1.69	1.59	1.8

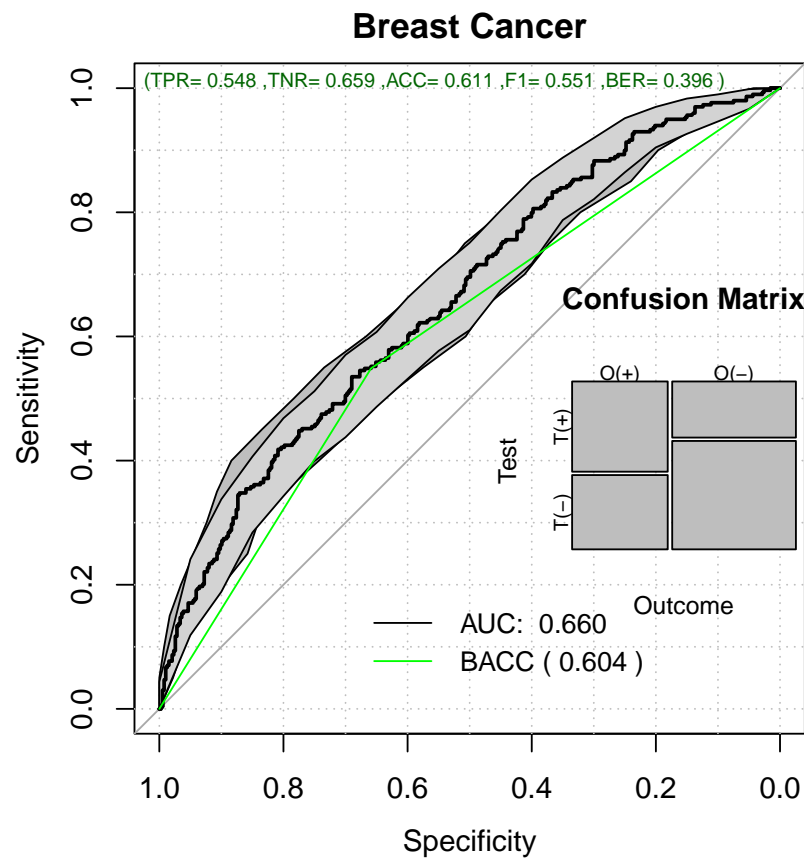
```
pander::pander(rrAnalysisTrain$sufdif,caption="Logrank test")
```

Table 25: 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")
```



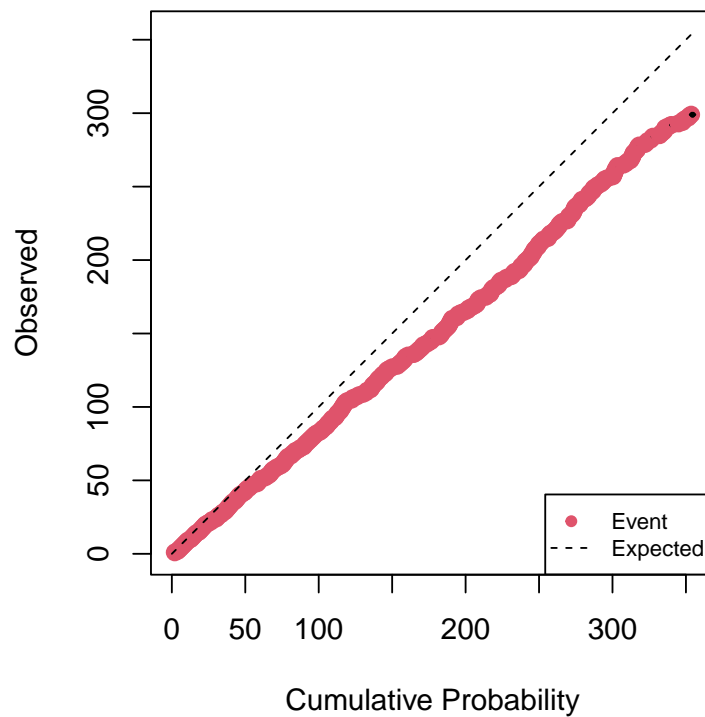
```
par(op)
```

```

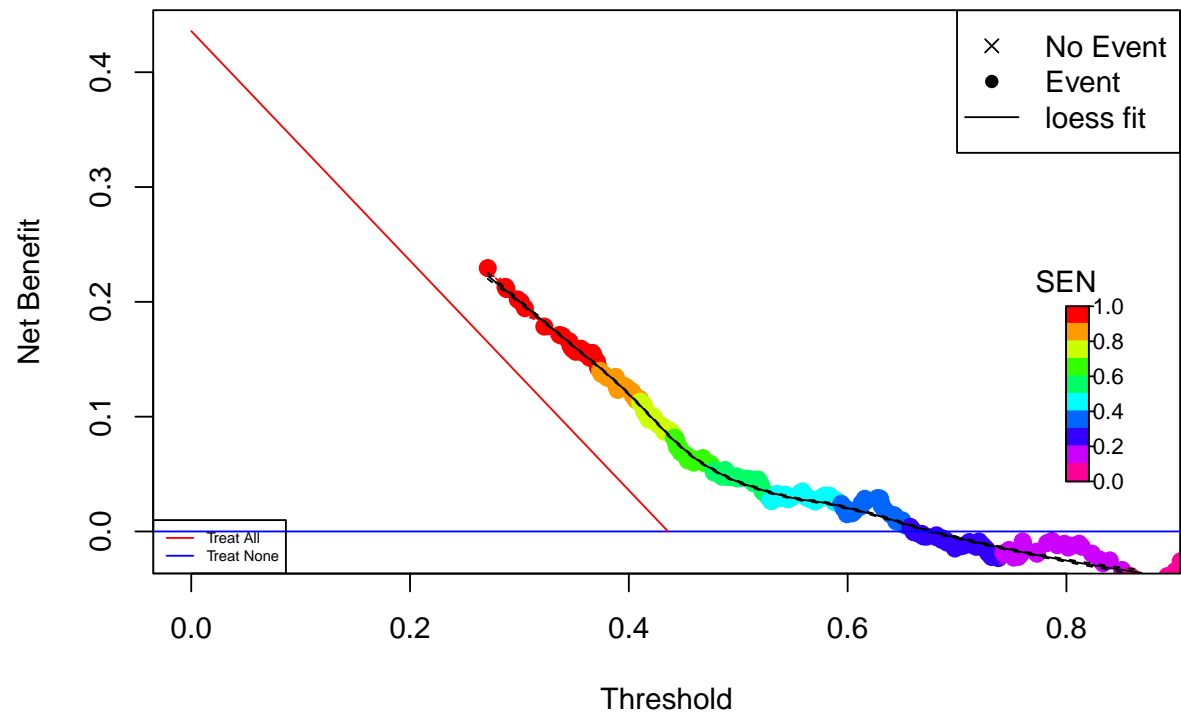
prob <- ppoisGzero(index,h0)
rdata <- cbind(dataBreastCancerTest$status,prob)
rrCoxTestAnalysis <- RRPlot(rdata,atThr=rrAnalysisTrain$thr_atP,
                             timetoEvent=dataBreastCancerTest$time,
                             title="Test: Breast Cancer",
                             ysurvlim=c(0.00,1.0),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

```

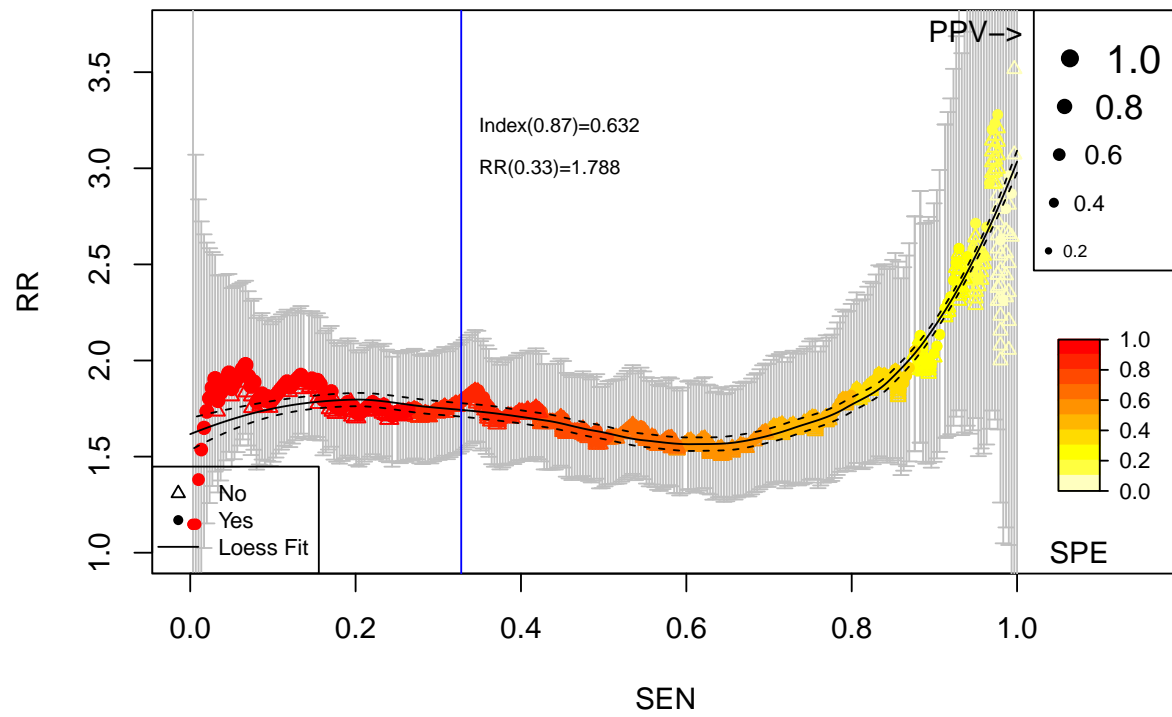
Cumulative vs. Observed: Test: Breast Cancer

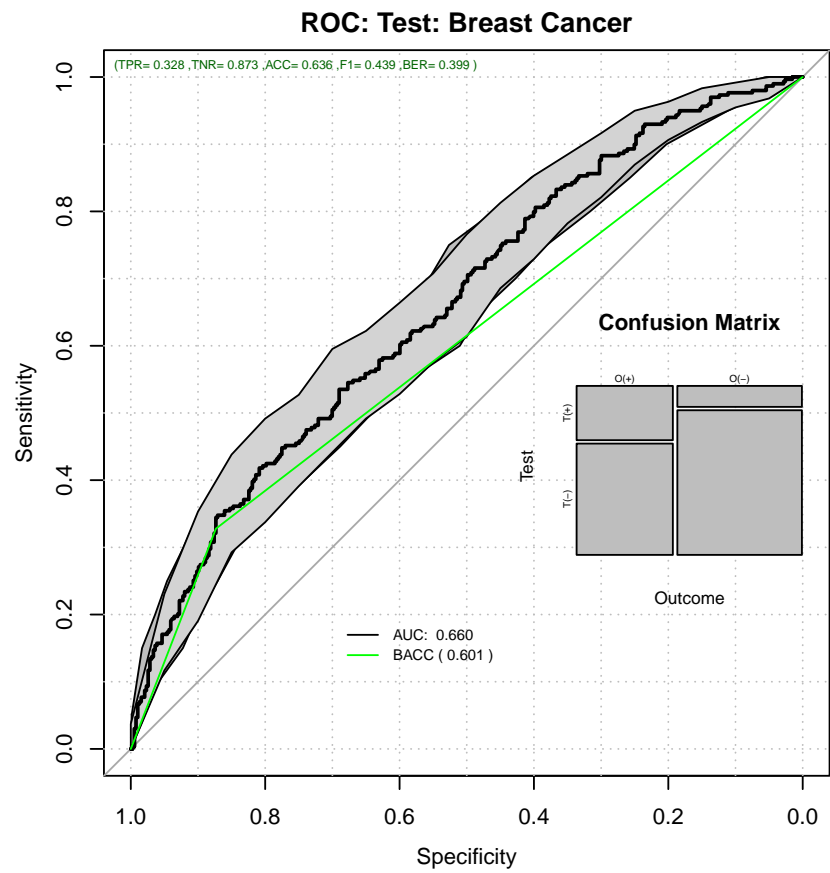


Decision Curve Analysis: Test: Breast Cancer

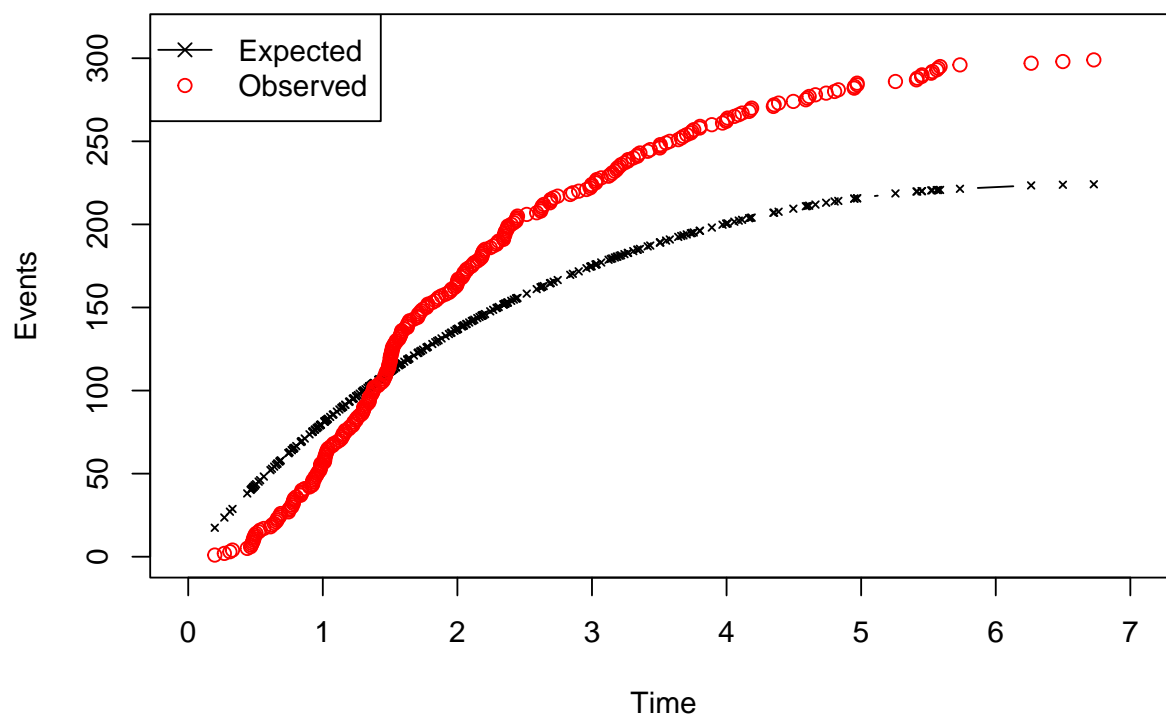


Relative Risk: Test: Breast Cancer

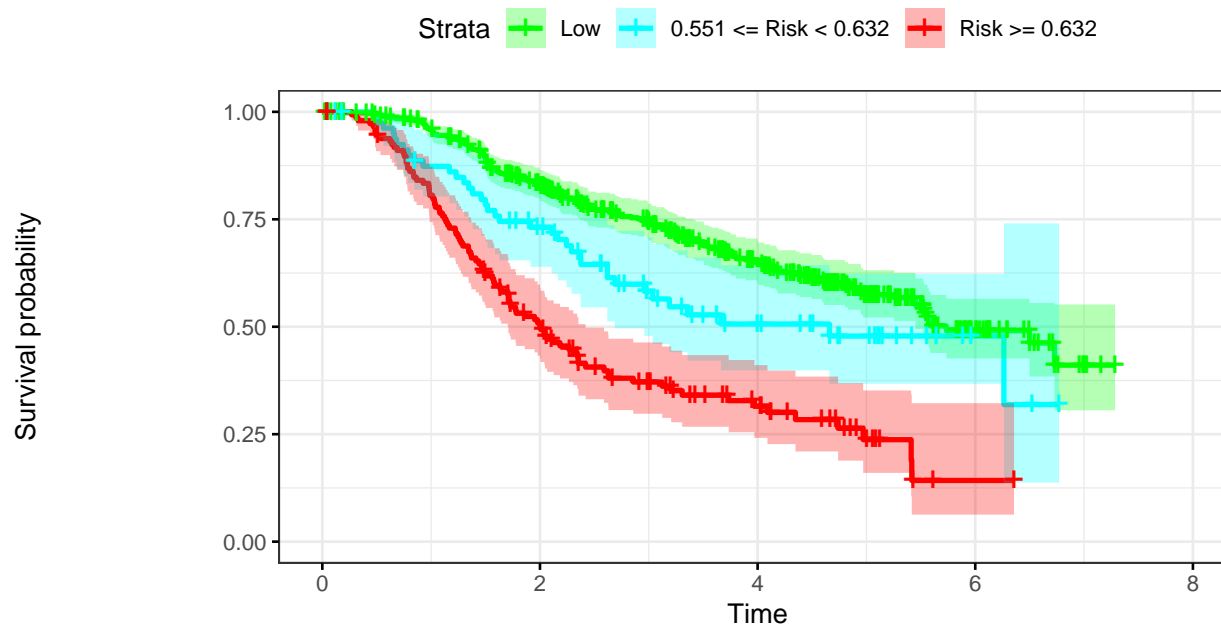




Time vs. Events: Test: Breast Cancer



Kaplan–Meier: Test: Breast Cancer



Number at risk

Low	457	338	183	32	0
0.551 <= Risk < 0.632	82	53	22	3	0
Risk >= 0.632	147	68	24	1	0

```
par(op)
```

1.5.1 External Data Report

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

Table 26: 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	0.664	0.632	0.693

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

Table 28: ROC AUC

est	lower	upper
0.66	0.619	0.7

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

Table 29: Sensitivity

est	lower	upper
0.328	0.275	0.384

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

Table 30: Specificity

est	lower	upper
0.873	0.836	0.905

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

Table 31: Probability Thresholds

90%	80%
0.632	0.551

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

Table 32: Risk Ratio

est	lower	upper
1.79	1.53	2.09

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

Table 33: 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")

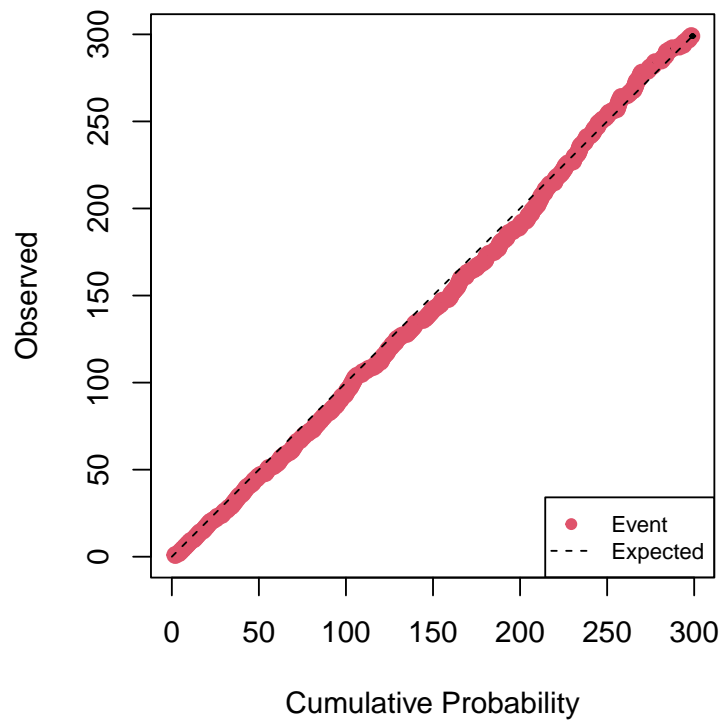
pander::pander(c(h0=calprob$h0,
                 Gain=calprob$hazardGain,
                 DeltaTime=calprob$timeInterval),
               caption="Cox Calibration Parameters")
```

h0	Gain	DeltaTime
0.535	0.925	4.87

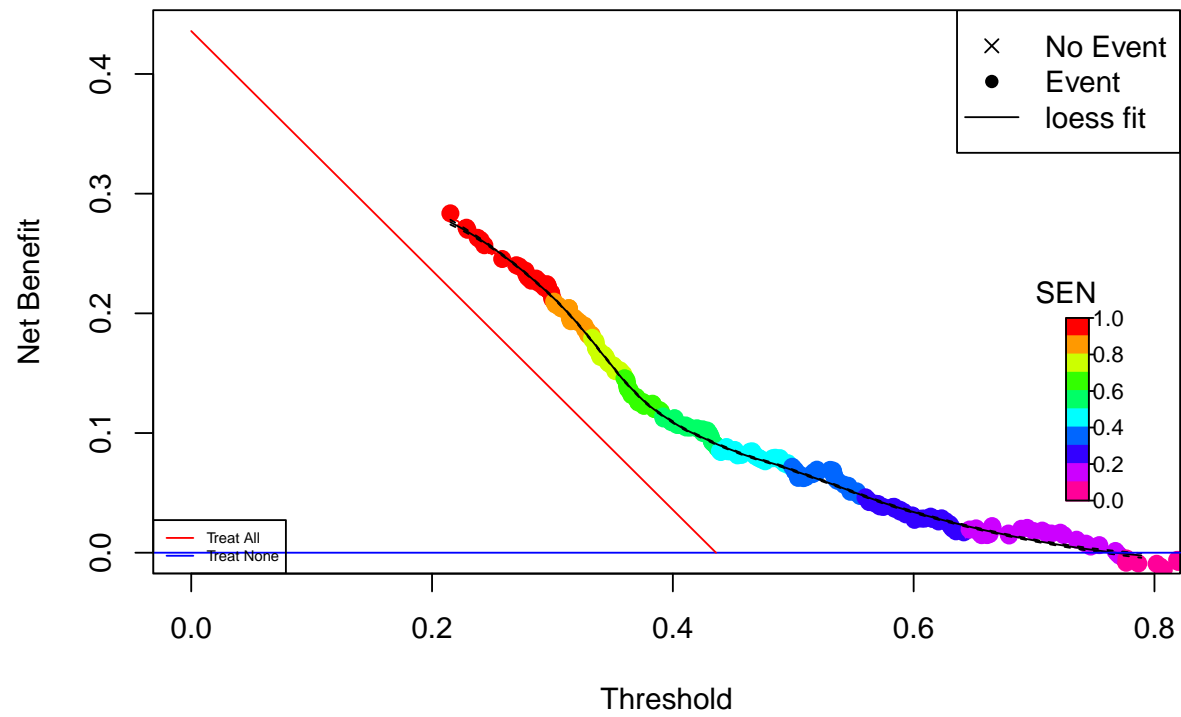
```
rdata <- cbind(dataBrestCancerTest$status,calprob$prob)

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)
```

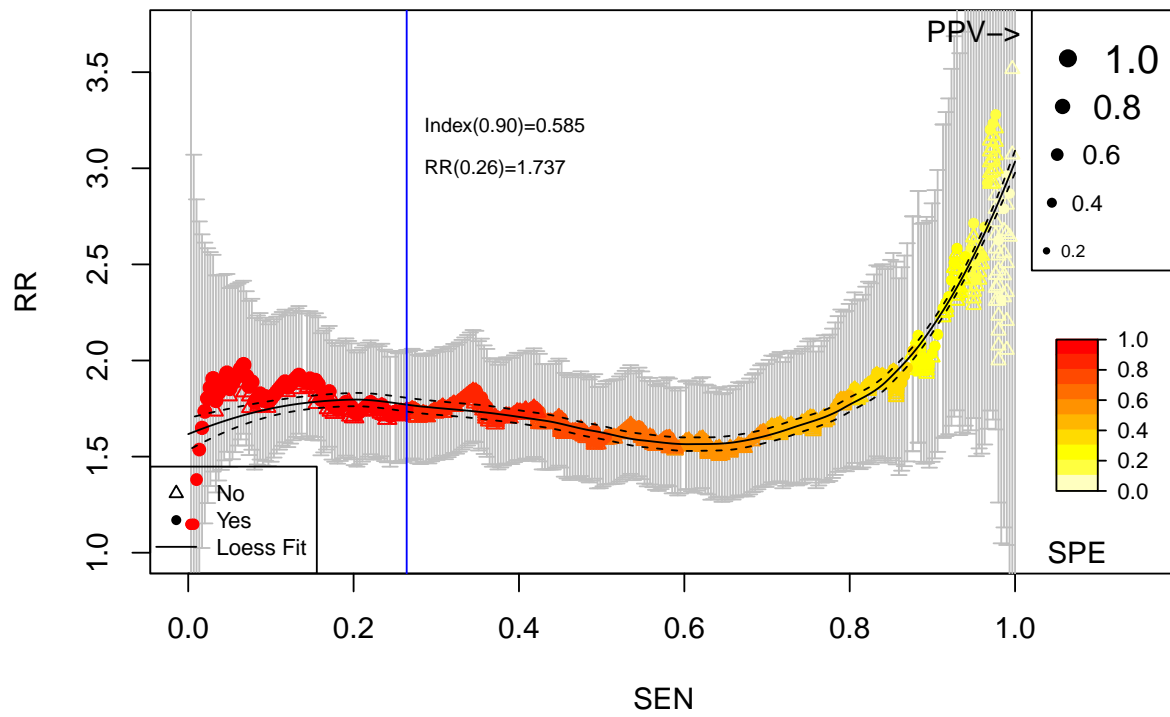
Cumulative vs. Observed: Cal. Test: Breast Cancer

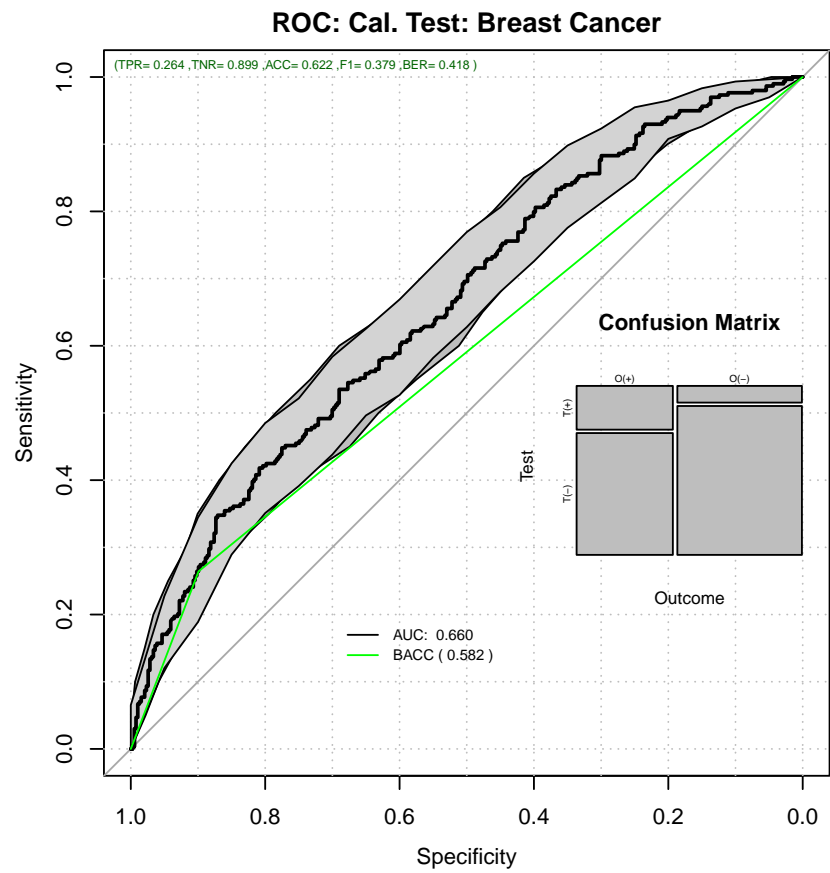


Decision Curve Analysis: Cal. Test: Breast Cancer

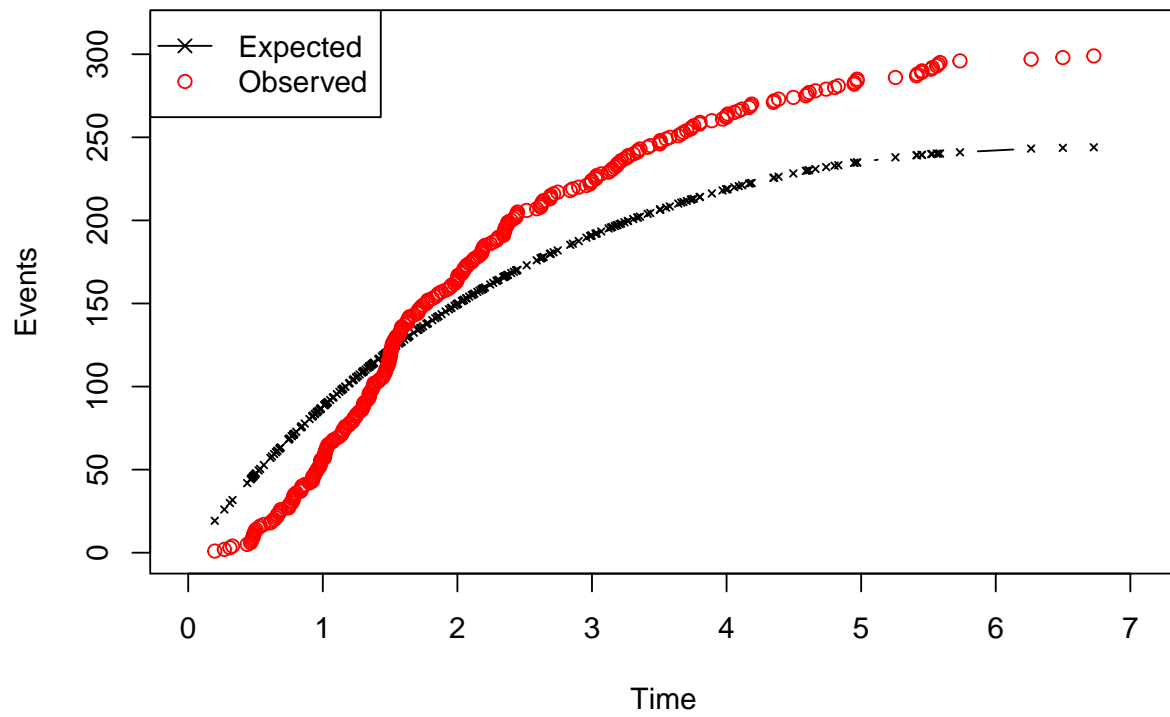


Relative Risk: Cal. Test: Breast Cancer

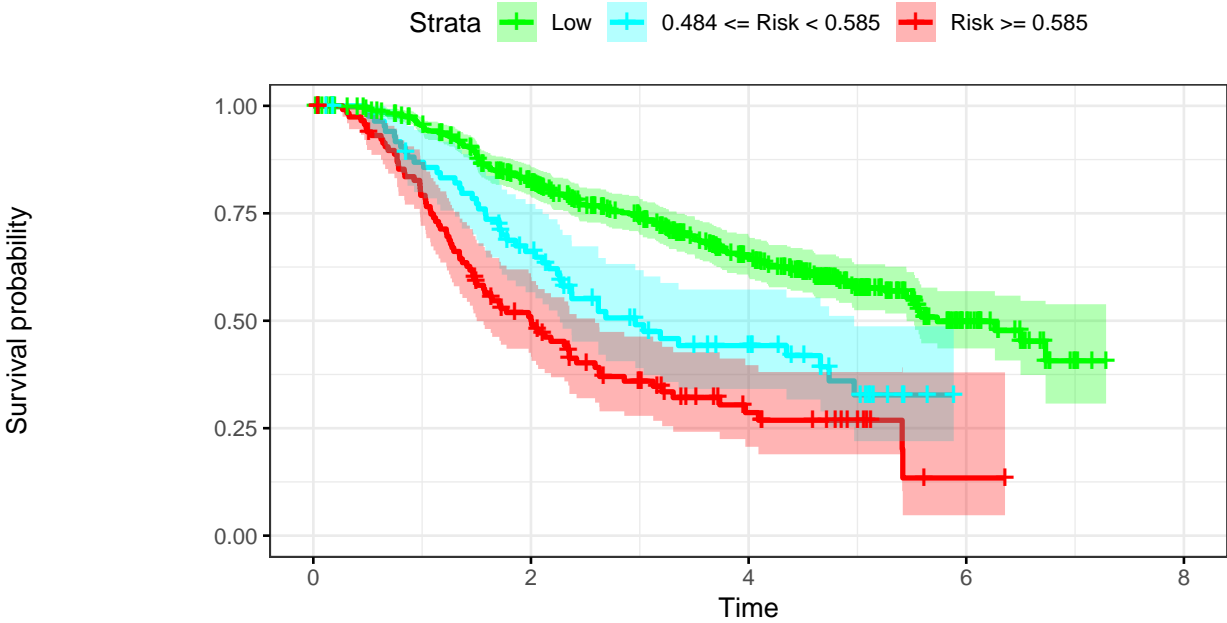




Time vs. Events: Cal. Test: Breast Cancer



Kaplan–Meier: Cal. Test: Breast Cancer



Number at risk

Low	482	354	190	35	0
0.484 <= Risk < 0.585	86	51	23	0	0
Risk >= 0.585	118	54	16	1	0

1.5.3 After Calibration Report

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

Table 35: 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	0.695

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

Table 37: ROC AUC

est	lower	upper
0.66	0.619	0.7

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

Table 38: Sensitivity

est	lower	upper
0.264	0.215	0.318

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

Table 39: Specificity

est	lower	upper
0.899	0.865	0.927

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

Table 40: Probability Thresholds

90%	80%
0.585	0.484

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

Table 41: Risk Ratio

est	lower	upper
1.74	1.48	2.05

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

Table 42: 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)
```

	Estimate	lower OR	upper	u.Accuracy	Accuracy	Accuracy	AUC	Full.AUC	AUC	NRI	z.IDI	z.NRI	Delta.AUC	Frequency		
size_nodes	1.001	1.001	1.001	0.669	0.571	0.668	0.627	0.500	0.628	0.1123	0.6365	17.86	18.87	0.1284	901	
nodes	4.33e-02	1.040	1.044	1.048	0.676	0.634	0.690	0.639	0.621	0.662	0.0711	10.57	10.64	13.16	1790	
grade_nodes	1.014	1.015	1.016	0.682	0.637	0.686	0.649	0.624	0.655	0.0658	0.5486	13.66	15.65	0.0310	871	
age_nodes	1.001	1.001	1.001	0.678	0.653	0.686	0.642	0.621	0.657	0.0334	0.2131	2.39	5.71	0.0358	961	
size_grade	1.001	1.002	1.002	0.632	0.682	0.686	0.626	0.646	0.655	0.0178	0.2941	6.74	7.72	0.0086	481	
age_size	8.73e-05	1.000	1.000	1.000	0.608	0.682	0.686	0.577	0.649	0.657	0.0153	0.2915	3.41	7.65	0.0076	001
grade	2.27e-01	1.168	1.254	1.347	0.571	0.683	0.690	0.500	0.653	0.662	0.0134	0.1903	6.20	4.98	0.0084	611
age_meno	0.992	0.994	0.996	0.571	0.676	0.686	0.500	0.645	0.657	0.0078	0.0805	1.76	2.33	0.0120	651	
age_pgr	1.000	1.000	1.000	0.571	0.686	0.686	0.500	0.656	0.657	0.0051	0.0074	1.11	0.19	0.0004	171	
age_grade	0.997	0.998	0.999	0.574	0.690	0.690	0.507	0.661	0.662	0.0045	0.1137	2.60	2.96	0.0003	151	
meno_grade	1.045	1.107	1.173	0.571	0.683	0.686	0.500	0.652	0.657	0.0042	0.2042	3.47	5.34	0.0044	111	

	Estimate	lower OR	upper	u.Accuracy	Accuracy	Accuracy	AUC	AUC	AUC	DI	NRI	z.IDI	z.NRI	Delta	Frequency	
nodes_hormon	0.979	0.986	0.994	0.587	0.688	0.686	0.526	0.658	0.655	0.0028	0.455	2.44	12.15	-	1	
1.38e-02														0.002853		
size	3.94e-03	1.002	1.004	1.006	0.611	0.693	0.690	0.618	0.663	0.662	0.0050	0.2105	3.42	5.600	-	1
														0.001075		
meno_pgr	3.19e-04	1.000	1.000	1.001	0.571	0.687	0.686	0.500	0.657	0.657	0.0031	0.0597	3.35	1.558	-	1
														0.000429		
pgr	-	1.000	1.000	1.000	0.571	0.689	0.686	0.500	0.659	0.655	0.0025	0.1975	2.64	5.745	-	1
1.07e-04														0.004123		
meno_nodes	0.955	0.974	0.994	0.640	0.686	0.686	0.595	0.656	0.657	0.0026	-	2.59	-	0.0006	311	
2.60e-02										0.06329			1.645			
grade_pgr	1.000	1.000	1.000	0.571	0.669	0.668	0.500	0.627	0.628	0.0024	0.1747	2.55	5.058	0.0012	521	
3.51e-05																
meno_size	1.34e-03	1.000	1.002	1.004	0.604	0.691	0.690	0.578	0.663	0.662	0.0018	0.1022	2.43	2.662	-	1
														0.001378		

1.7 Logistic Model Performance

```

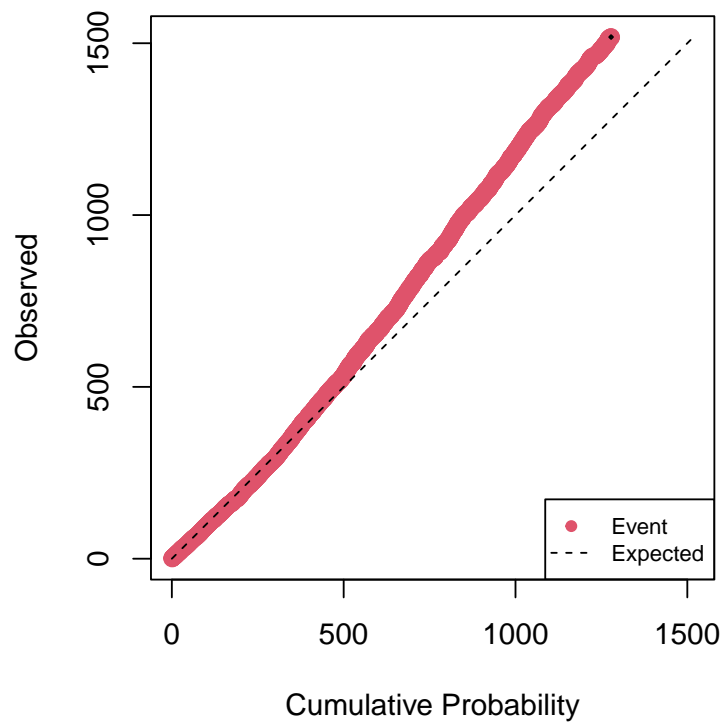
op <- par(no.readonly = TRUE)

cprob <- predict(mlog,dataBrestCancerTrain)

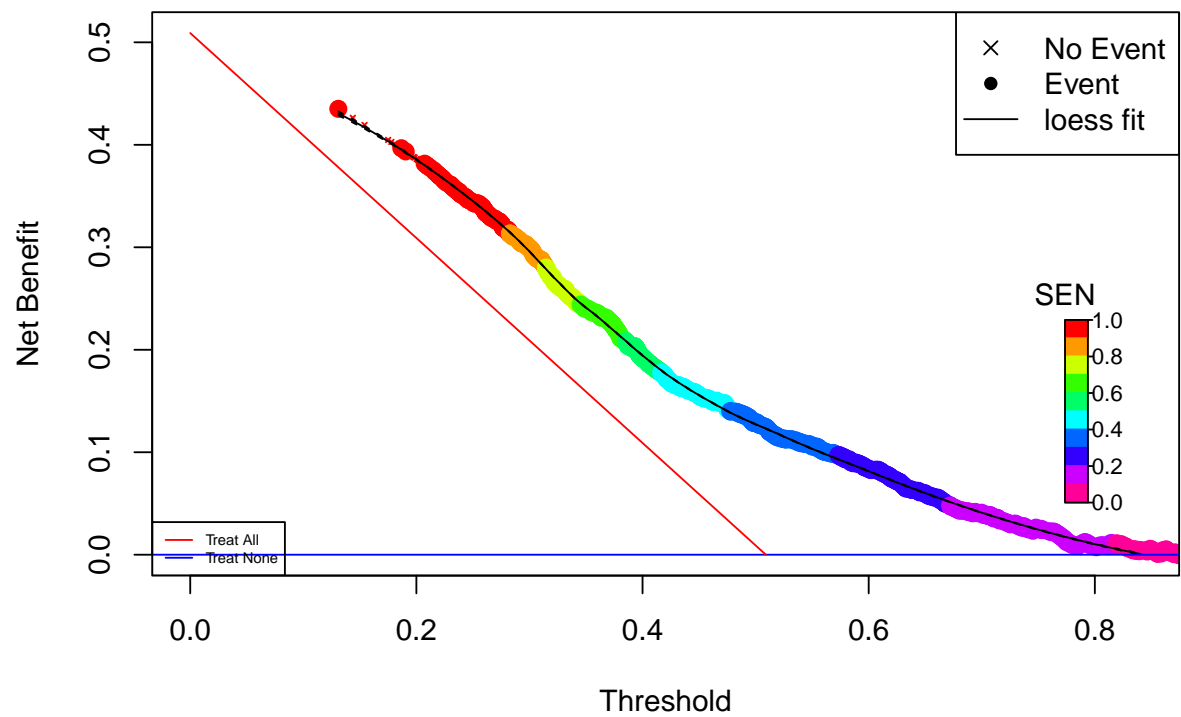
rdata <- cbind(dataBrestCancerTrain$status,cprob)
rrAnalysisTrain <- RRPlot(rdata,atProb=c(0.90,0.80),
                           timetoEvent=dataBrestCancerTrain$time,
                           title="Logistic Train: Breast Cancer",
                           ysurvlim=c(0.00,1.0),
                           riskTimeInterval=5.0)

```

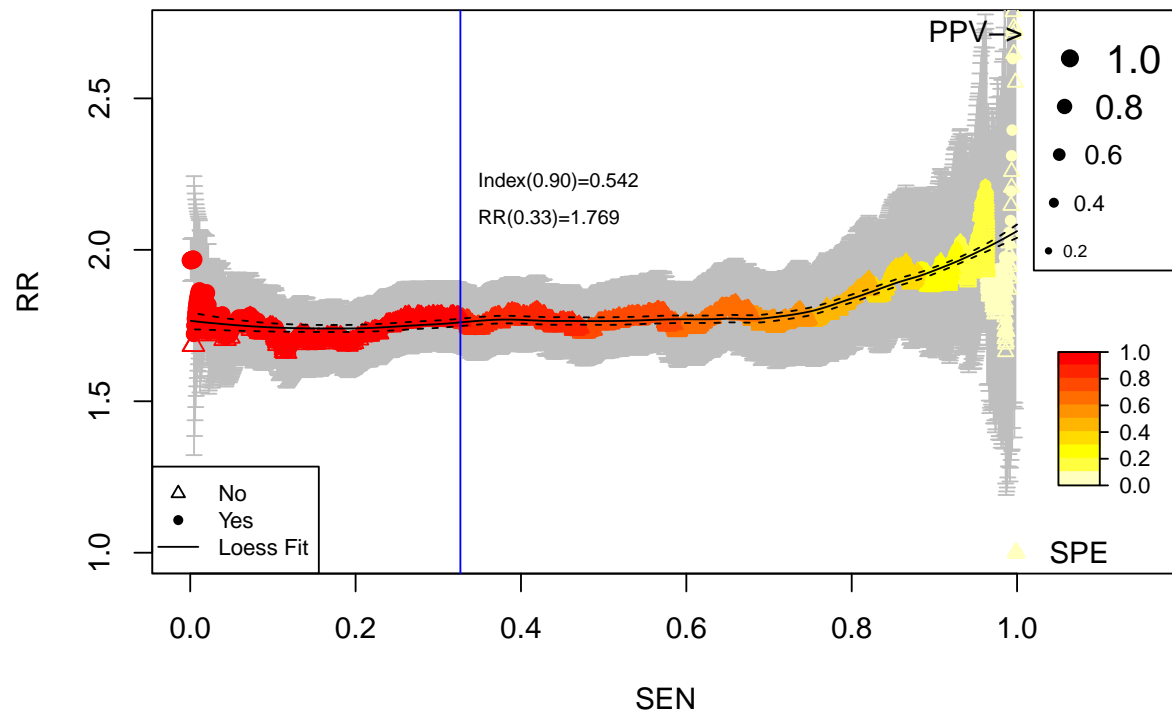
Cumulative vs. Observed: Logistic Train: Breast Cancer

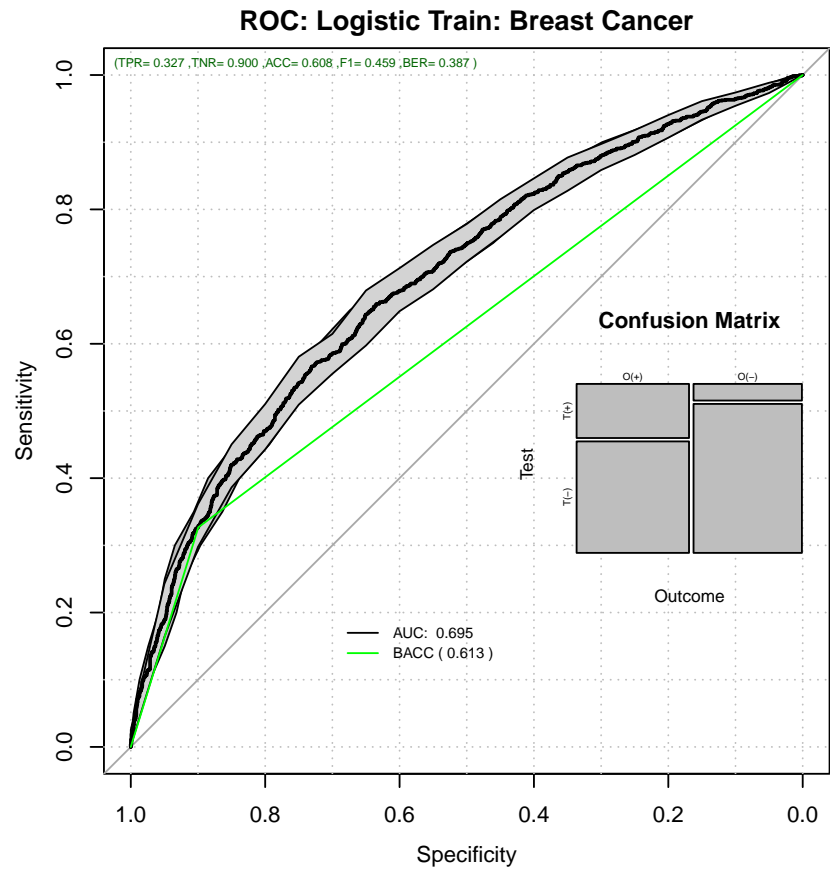


Decision Curve Analysis: Logistic Train: Breast Cancer

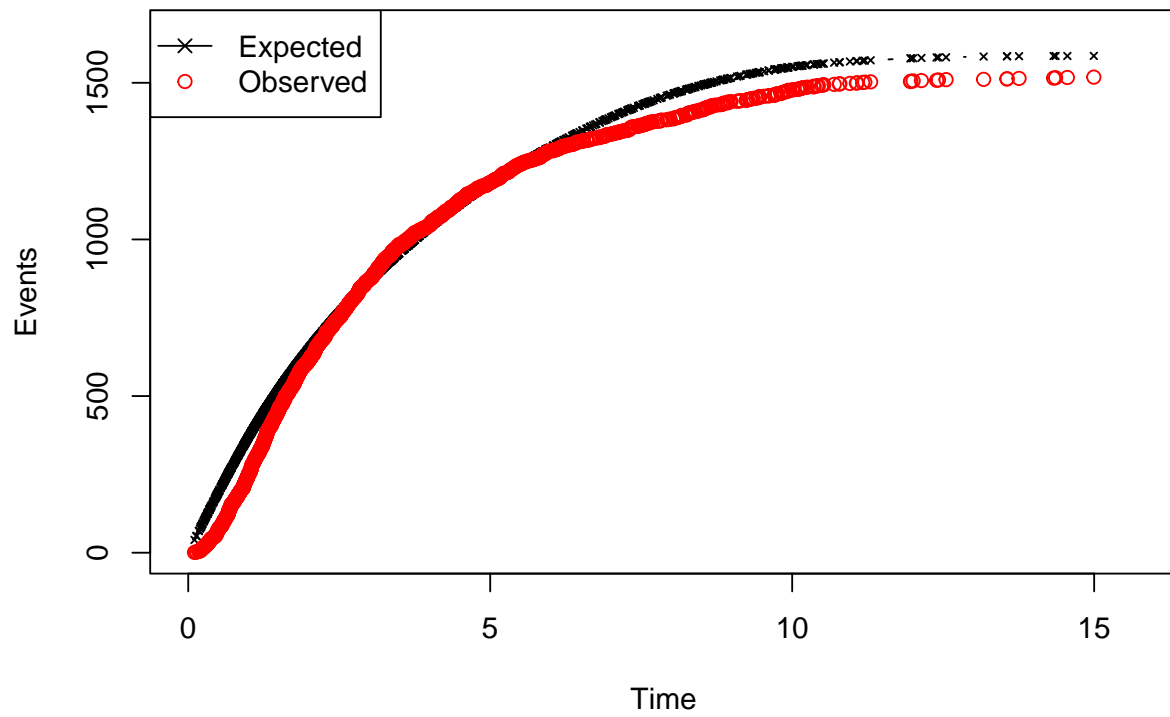


Relative Risk: Logistic Train: Breast Cancer

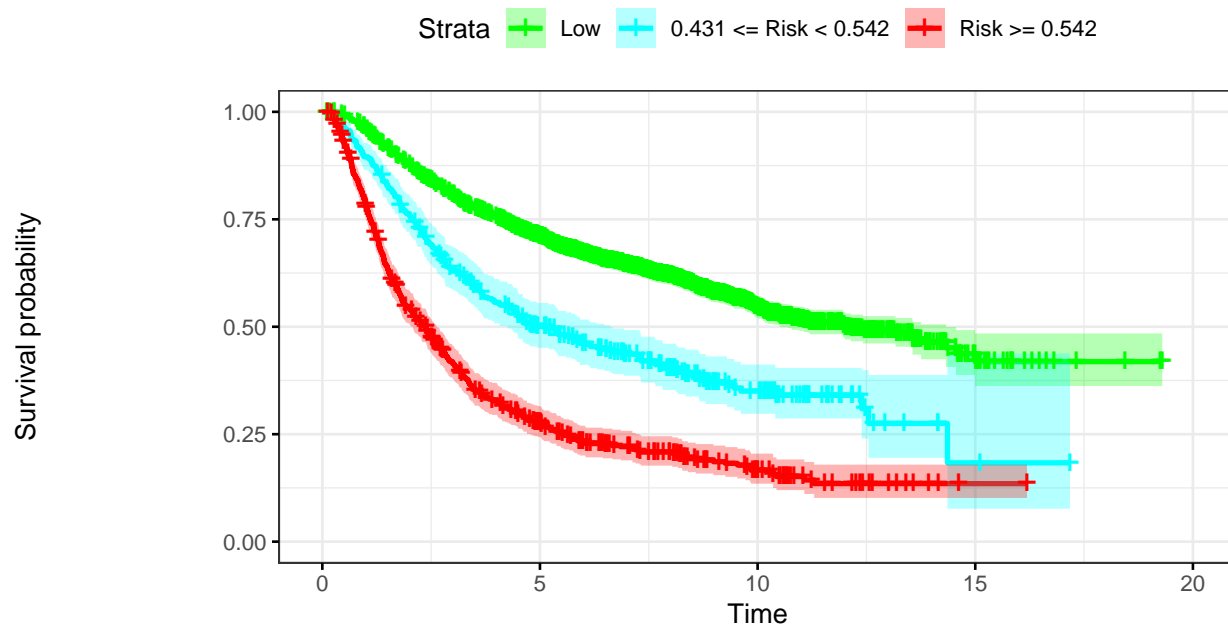




Time vs. Events: Logistic Train: Breast Cancer



Kaplan–Meier: Logistic Train: Breast Cancer



Number at risk

Low	1975	1268	399	23	0
0.431 <= Risk < 0.542	364	160	47	2	0
Risk >= 0.542	643	143	37	1	0

```
par(op)
```

1.7.1 Training Report

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

Table 44: 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:** *4206588*
- **Uncertain:** *2703838*
- **cstatCI:**

mean.C Index	median	lower	upper
0.68	0.68	0.667	0.694

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

Table 46: ROC AUC

est	lower	upper
0.695	0.677	0.714

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

Table 47: Sensitivity

est	lower	upper
0.327	0.303	0.351

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

Table 48: Specificity

est	lower	upper
0.9	0.883	0.915

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

Table 49: Probability Thresholds

90%	80%
0.542	0.431

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

Table 50: Risk Ratio

est	lower	upper
1.77	1.66	1.88

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

Table 51: Logrank test $\text{Chisq} = 543.347175$ on 2 degrees of freedom,
 $p = 0.000000$

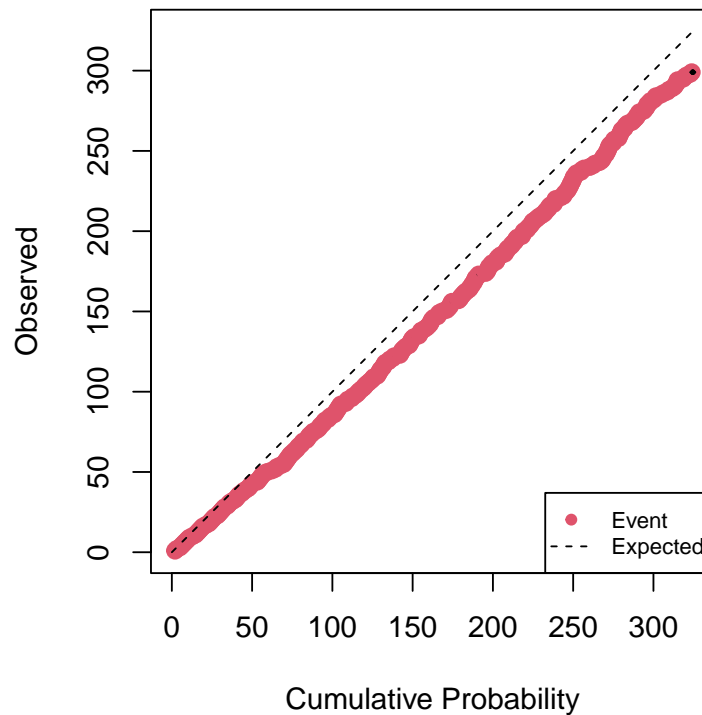
	N	Observed	Expected	$(O-E)^2/E$	$(O-E)^2/V$
class=0	1975	804	1145	101.5	418.9
class=1	364	218	169	14.1	15.9
class=2	643	496	204	418.2	490.7

1.7.2 Results on the validation set using Logistic model

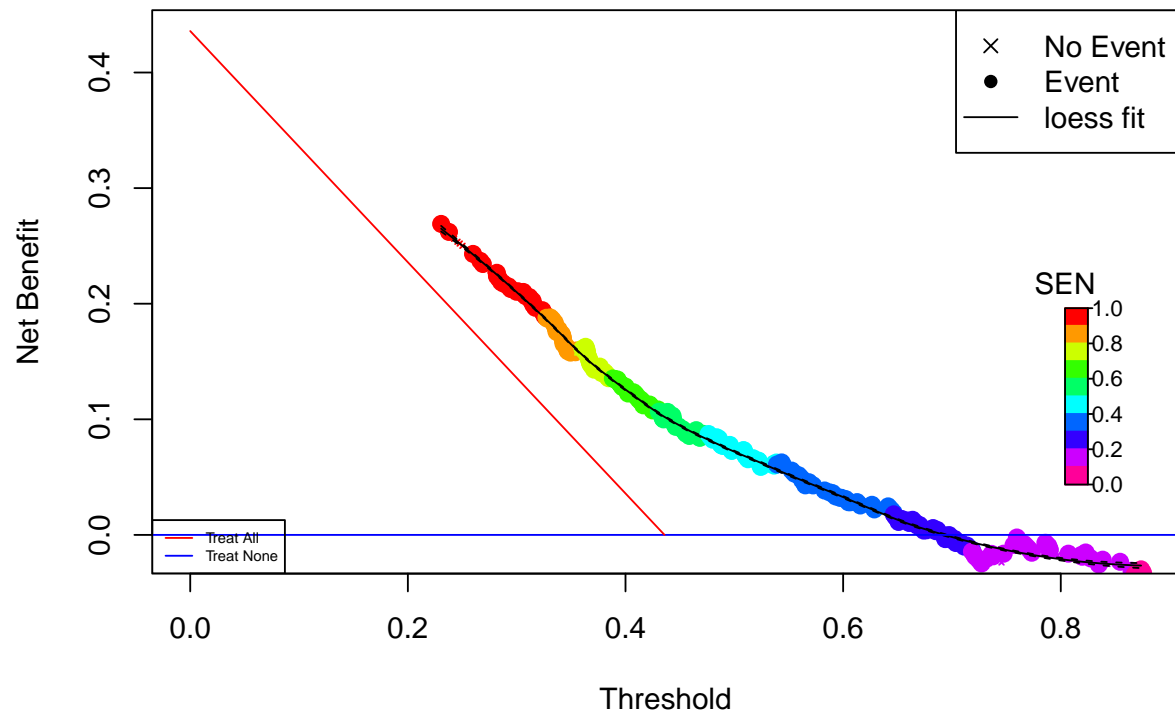
```
pre <- predict(mlog,dataBrestCancerTest)
rdata <- cbind(dataBrestCancerTest$status,pre)

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

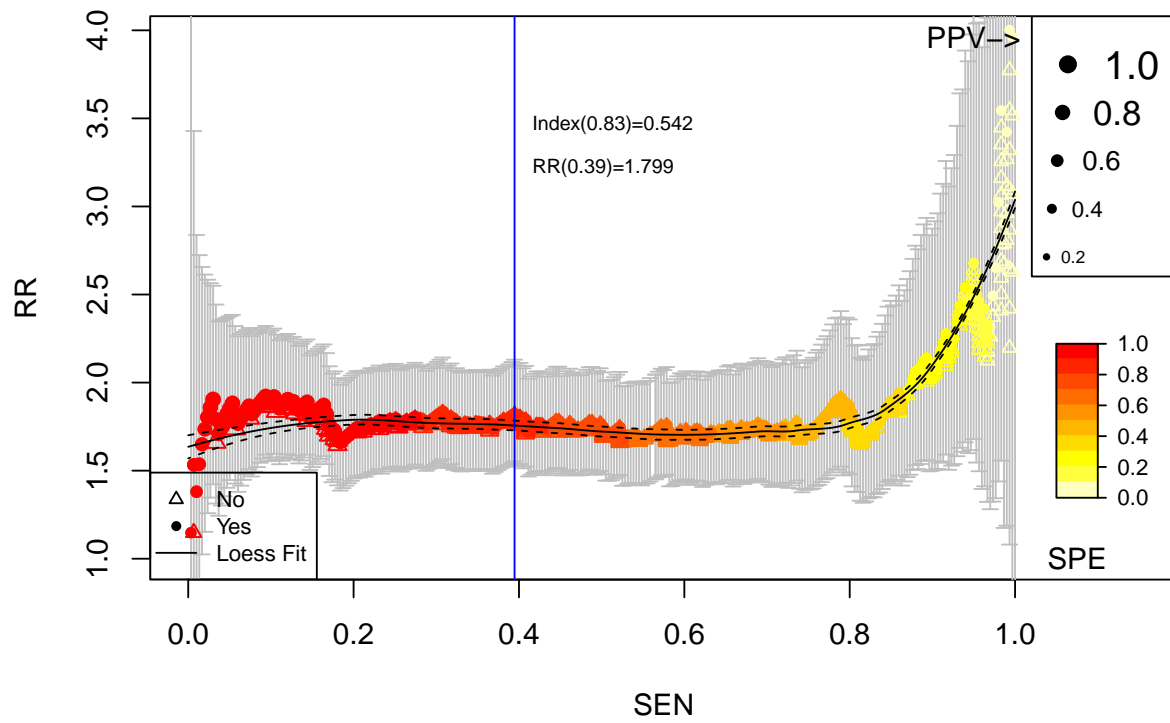
Cumulative vs. Observed: Logistic Test: Breast Cancer

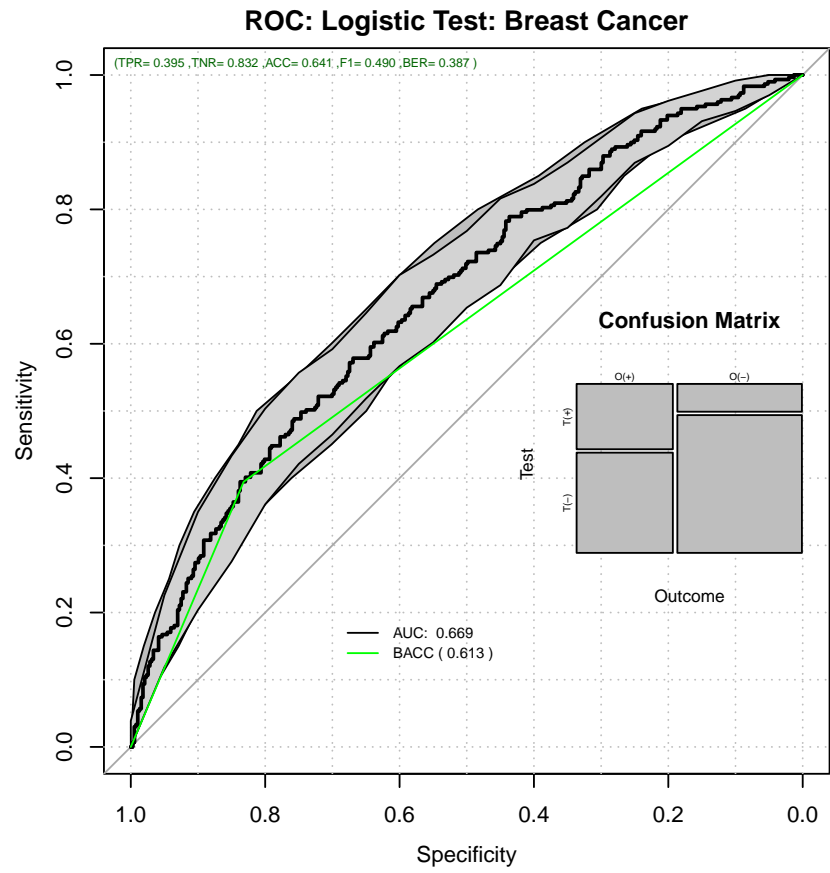


Decision Curve Analysis: Logistic Test: Breast Cancer

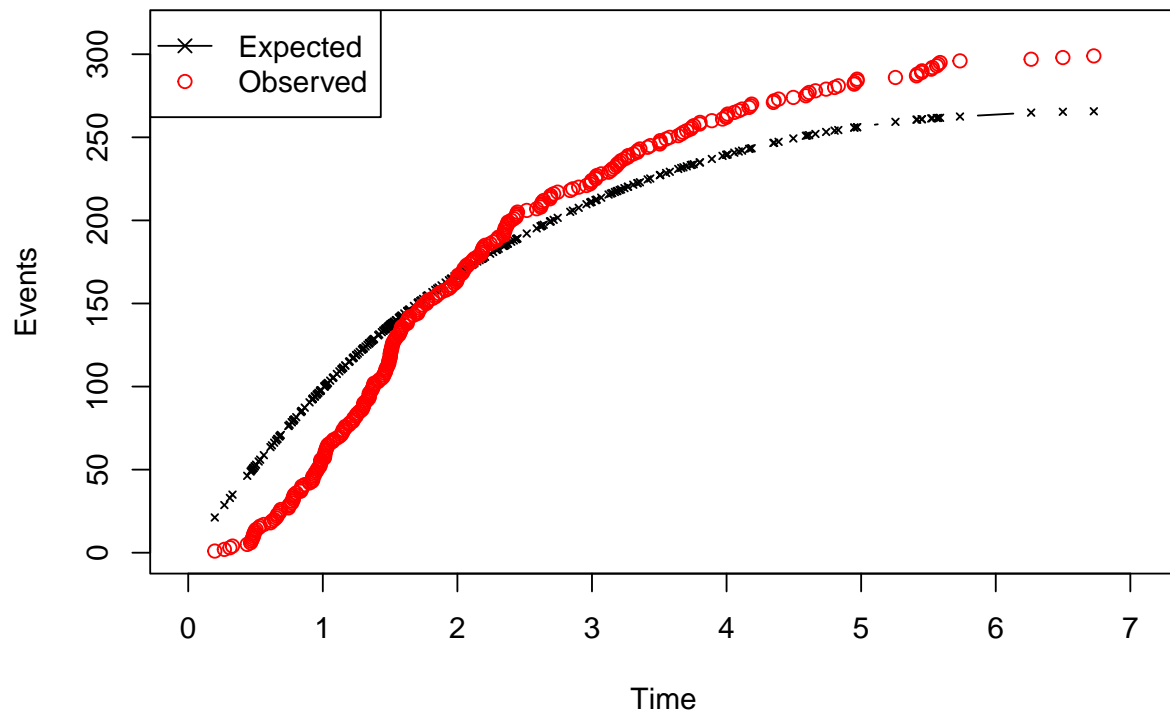


Relative Risk: Logistic Test: Breast Cancer

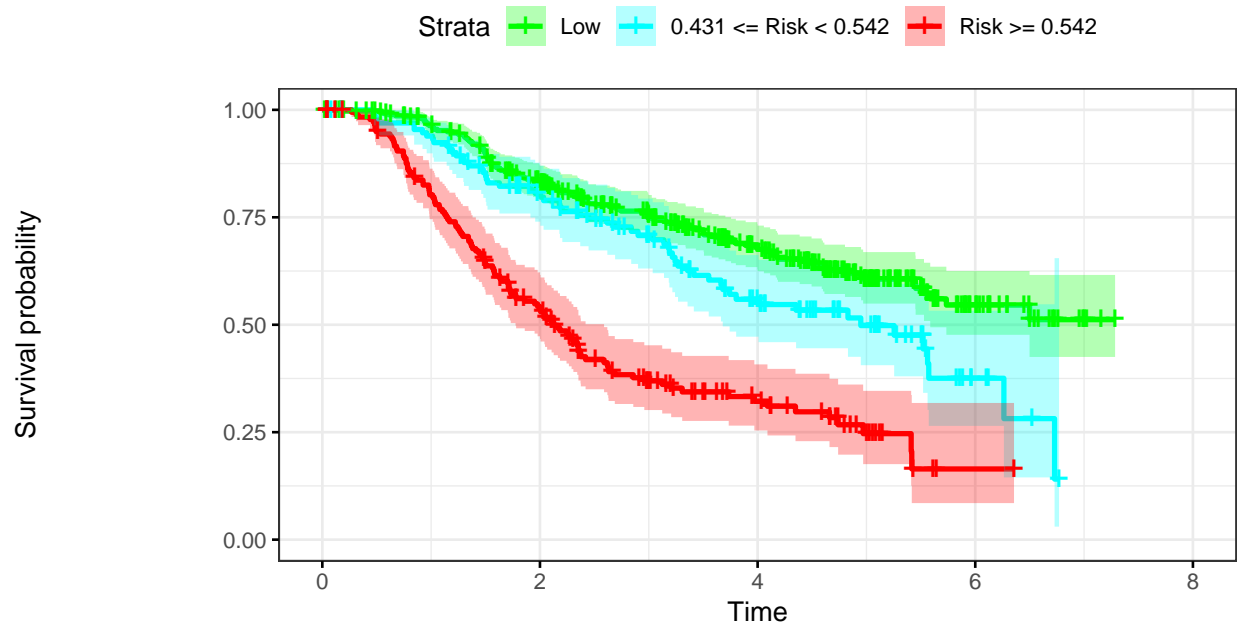




Time vs. Events: Logistic Test: Breast Cancer



Kaplan–Meier: Logistic Test: Breast Cancer



Number at risk

Low	369	274	154	29	0
0.431 <= Risk < 0.542	134	96	46	6	0
Risk >= 0.542	183	89	29	1	0

```
par(op)
```

1.7.3 Validation Report

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

Table 52: 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$ROCAAnalysis$aucs),caption="ROC AUC")
```

Table 54: ROC AUC

est	lower	upper
0.669	0.628	0.709

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

Table 55: Sensitivity

est	lower	upper
0.395	0.339	0.453

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

Table 56: Specificity

est	lower	upper
0.832	0.791	0.868

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

Table 57: Probability Thresholds

90%	80%
0.542	0.431

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

Table 58: Risk Ratio

est	lower	upper
1.8	1.54	2.11

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

Table 59: Logrank test $\text{Chisq} = 92.507991$ 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"), dataBrestC
calprob <- CalibrationProbPoissonRisk(riskdata)
```

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

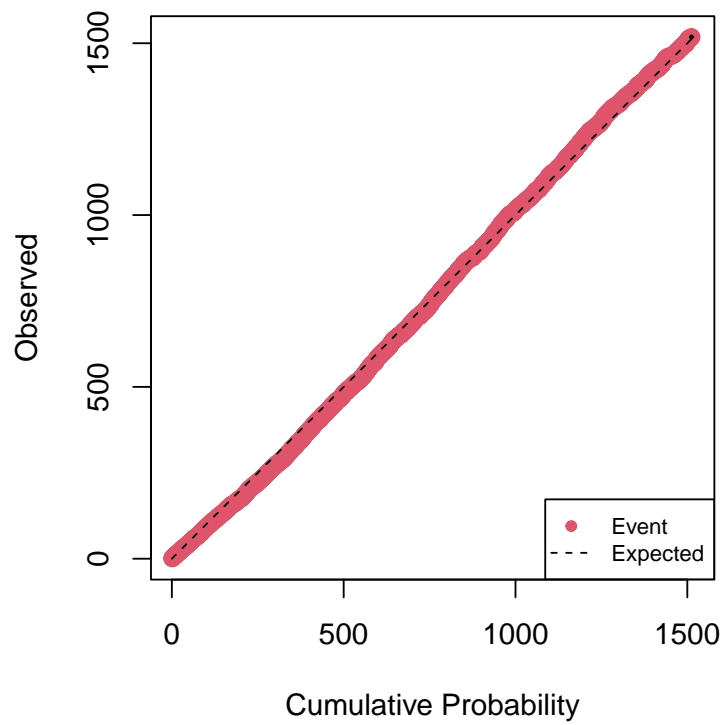
h0	Gain	DeltaTime
0.676	1.31	7.14

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

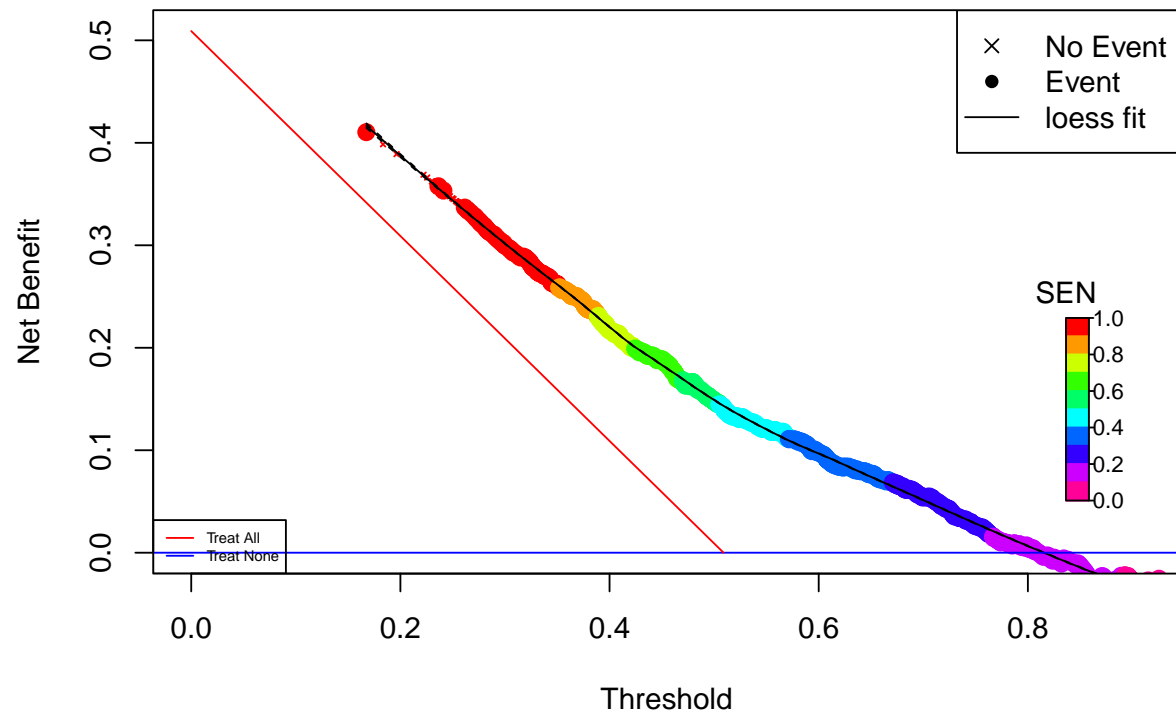
```
rdata <- cbind(dataBrestCancerTrain$status, calprob$prob)
```

```
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)
```

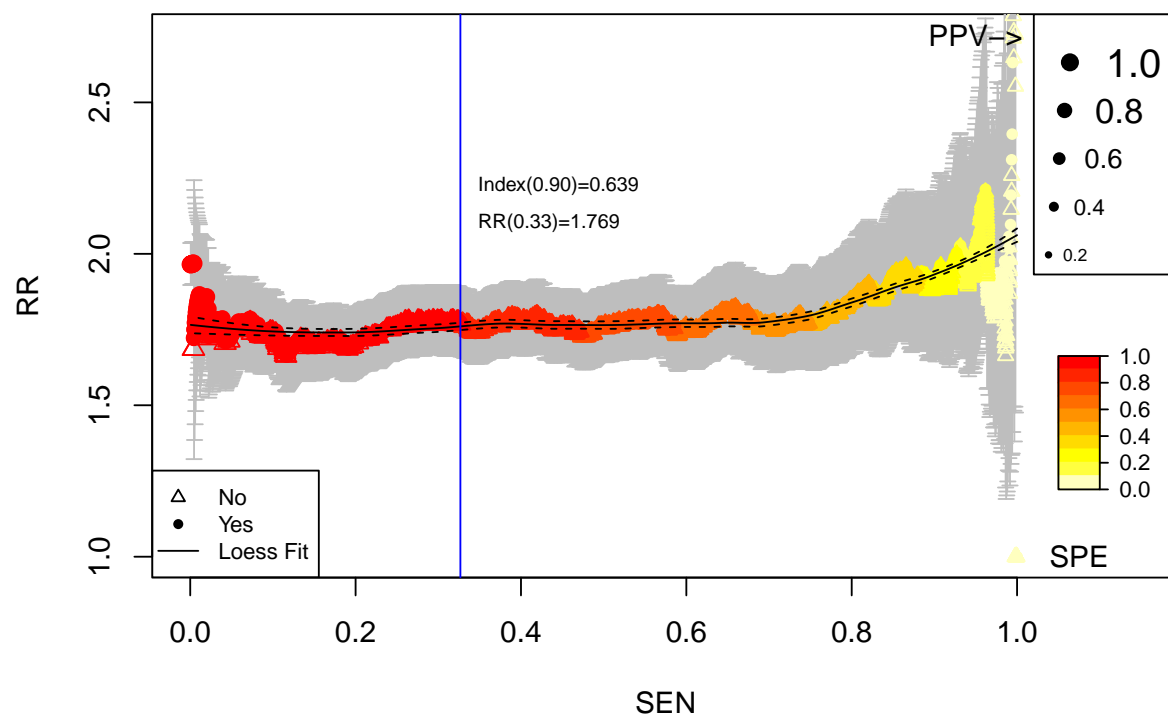
Cumulative vs. Observed: Cal. Logistic Train: Breast Cancer

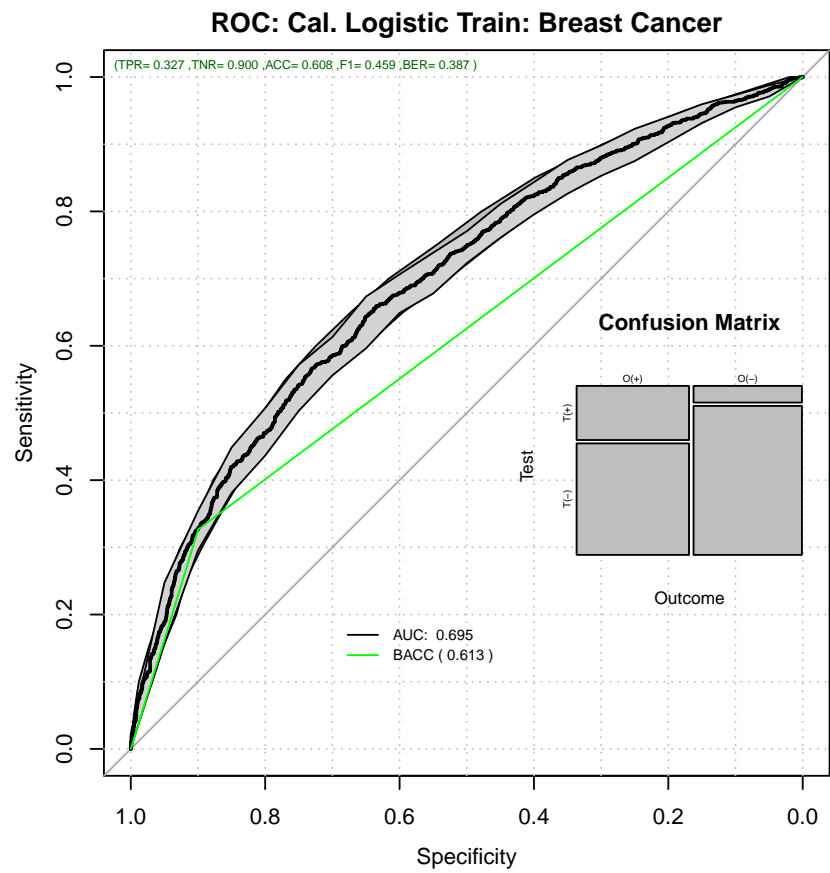


Decision Curve Analysis: Cal. Logistic Train: Breast Cancer

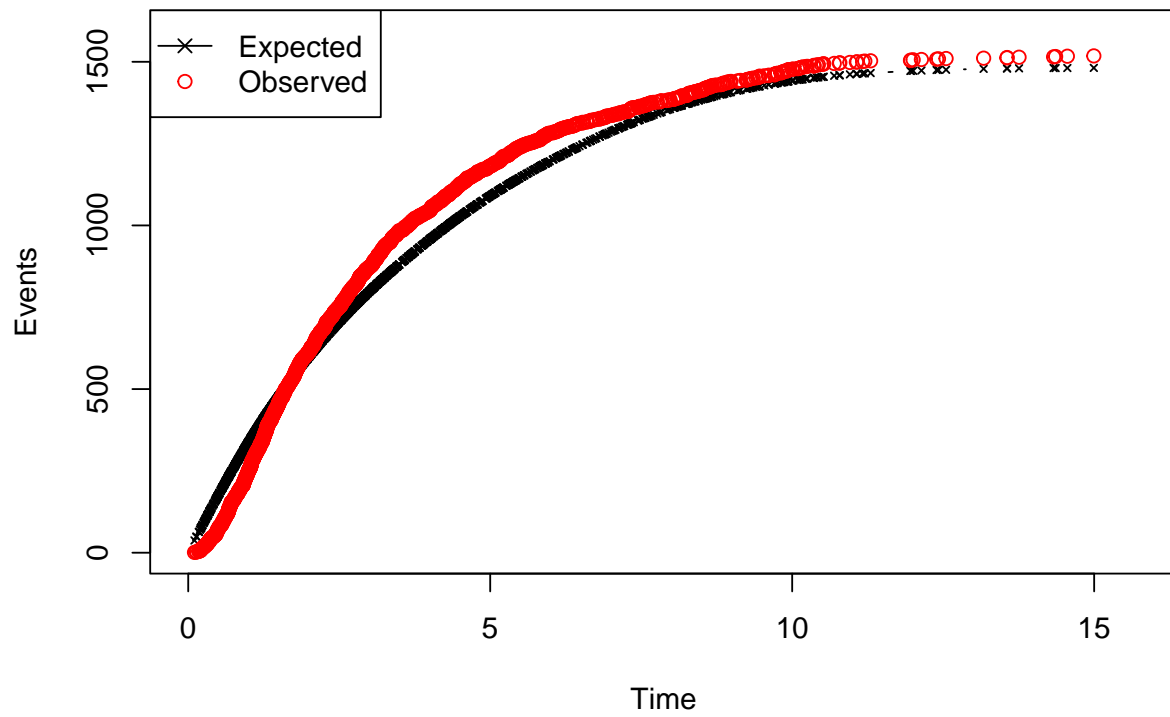


Relative Risk: Cal. Logistic Train: Breast Cancer

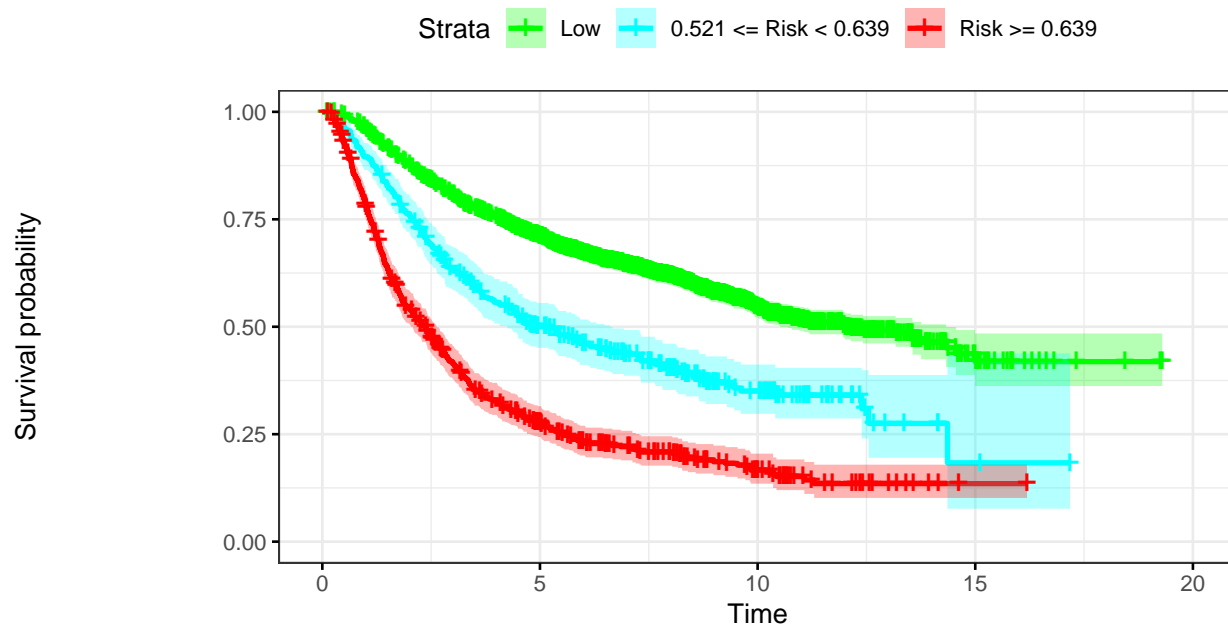




Time vs. Events: Cal. Logistic Train: Breast Cancer



Kaplan–Meier: Cal. Logistic Train: Breast Cancer



Number at risk

Low	1975	1268	399	23	0
0.521 <= Risk < 0.639	364	160	47	2	0
Risk >= 0.639	643	143	37	1	0

```
par(op)
```

1.8.1 Report of the calibrated logistic: training

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

Table 61: 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:** 4206588
- **Uncertain:** 2703838
- **cstatCI:**

mean.C Index	median	lower	upper
0.68	0.68	0.666	0.693

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

Table 63: ROC AUC

est	lower	upper
0.695	0.677	0.714

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

Table 64: Sensitivity

est	lower	upper
0.327	0.303	0.351

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

Table 65: Specificity

est	lower	upper
0.9	0.883	0.915

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

Table 66: Probability Thresholds

90%	80%
0.639	0.521

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

Table 67: Risk Ratio

est	lower	upper
1.77	1.66	1.88

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

Table 68: 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

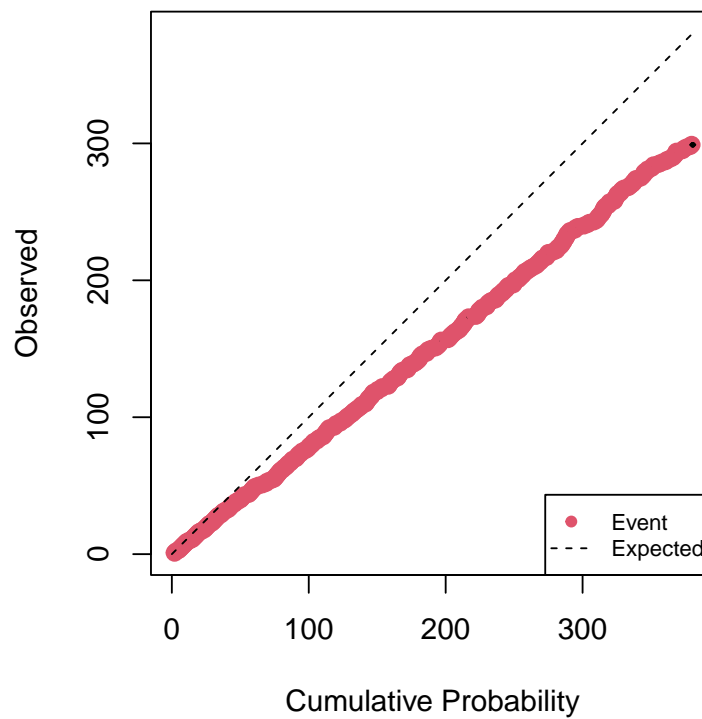
```

probLog <- predict(mlog,dataBrestCancerTest)
aprof <- adjustProb(probLog,gain)

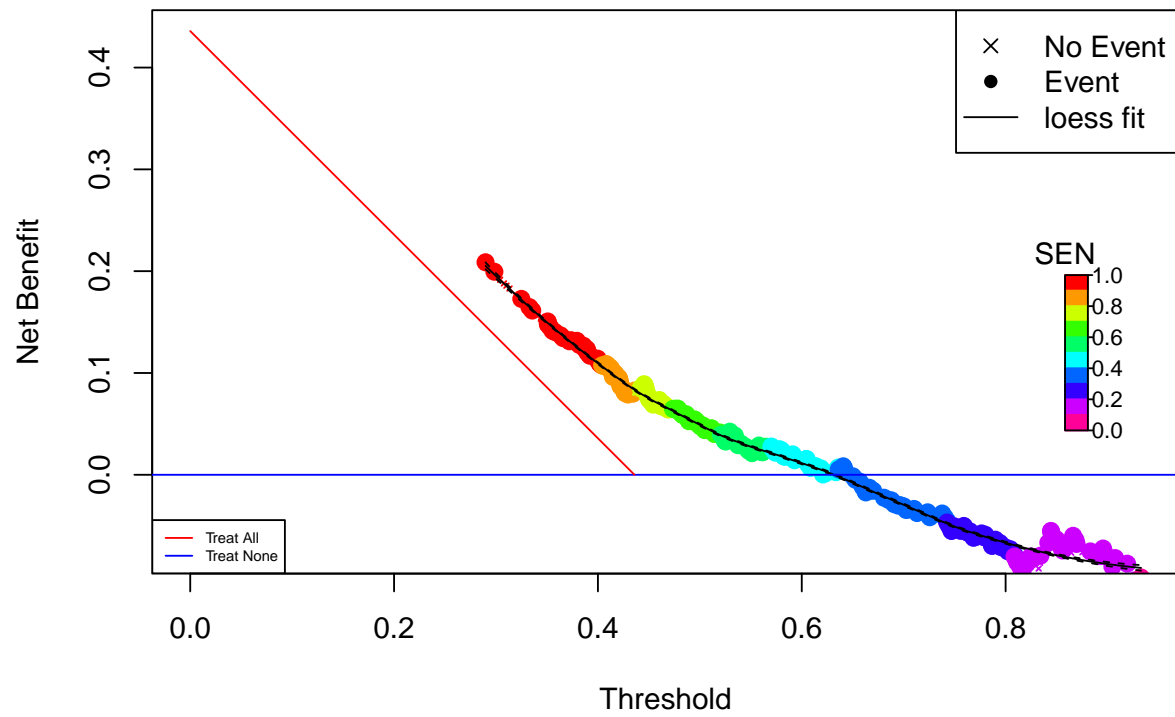
rdata <- cbind(dataBrestCancerTest$status,aprof)
rrAnalysisTestLogistic <- RRPlot(rdata,atThr=rrAnalysisTrain$thr_atP,
  timetoEvent=dataBrestCancerTest$time,
  title="Cal. Logistic Test: Breast Cancer",
  ysurvlim=c(0.00,1.0),
  riskTimeInterval=timeinterval)

```

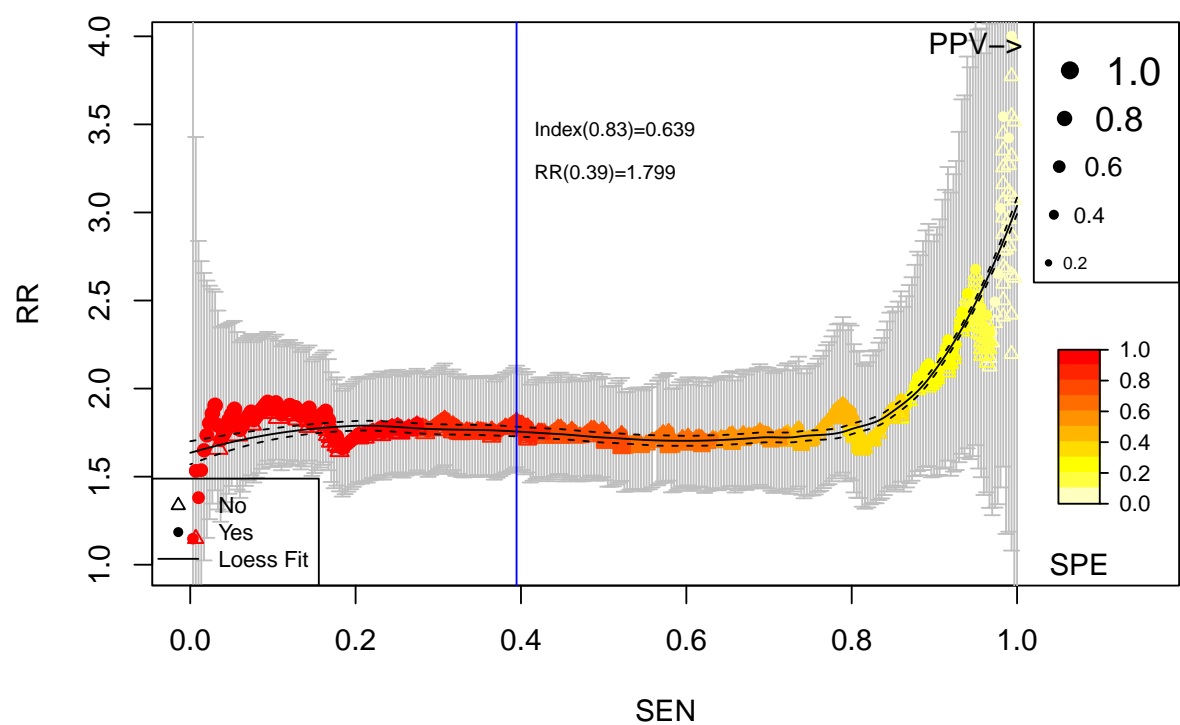
Cumulative vs. Observed: Cal. Logistic Test: Breast Cancer

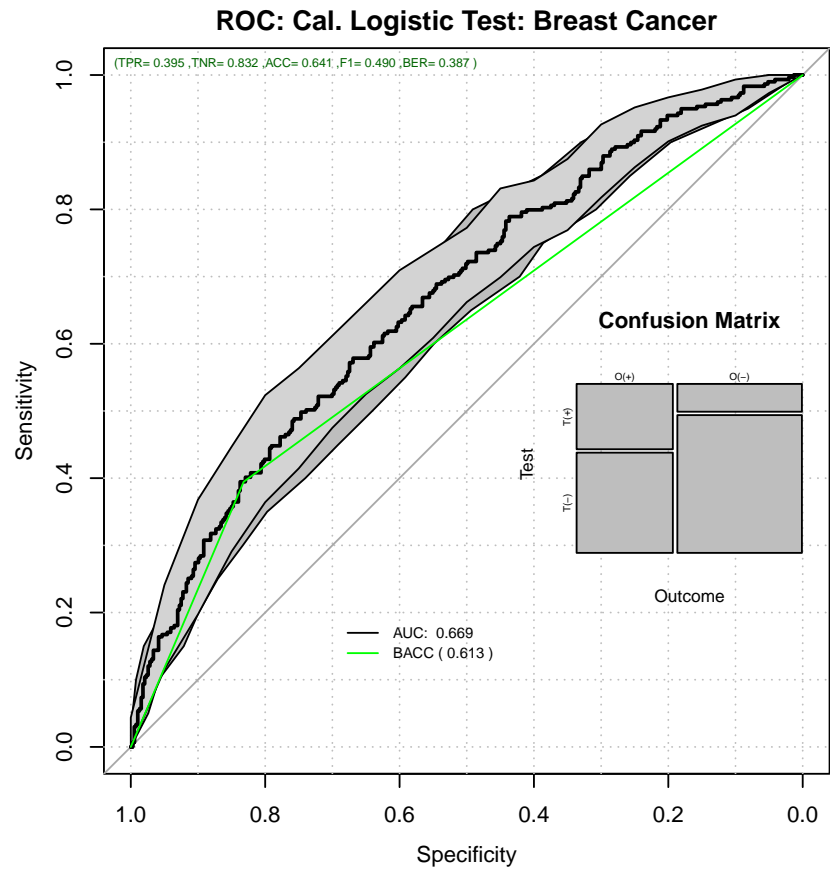


Decision Curve Analysis: Cal. Logistic Test: Breast Cancer

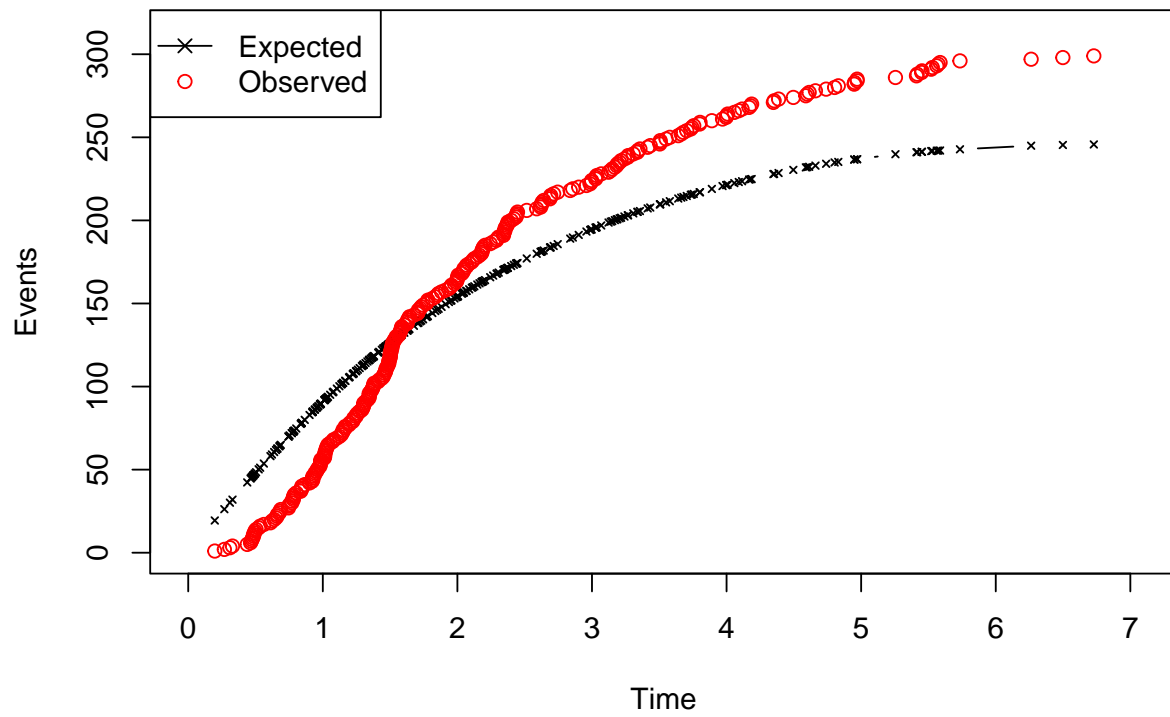


Relative Risk: Cal. Logistic Test: Breast Cancer

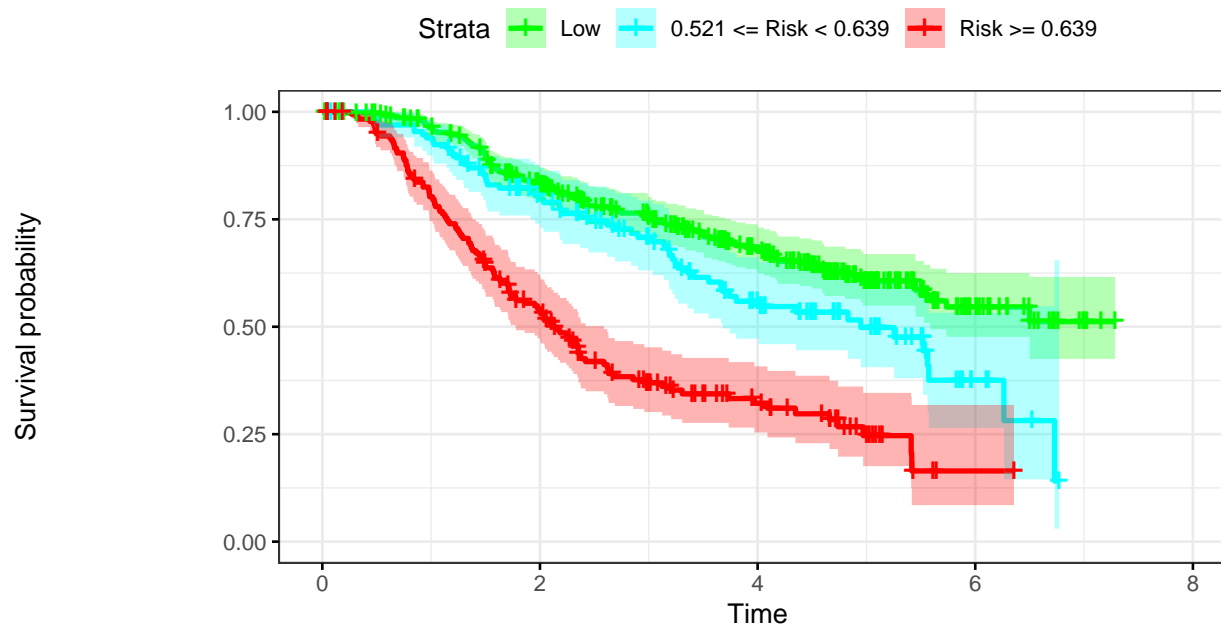




Time vs. Events: Cal. Logistic Test: Breast Cancer



Kaplan–Meier: Cal. Logistic Test: Breast Cancer



Number at risk

Low	369	274	154	29	0
0.521 <= Risk < 0.639	134	96	46	6	0
Risk >= 0.639	183	89	29	1	0

```
par(op)
```

1.8.2 Report of the calibrated validation

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

Table 69: 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	0.637	0.698

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

Table 71: ROC AUC

est	lower	upper
0.669	0.628	0.709

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

Table 72: Sensitivity

est	lower	upper
0.395	0.339	0.453

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

Table 73: Specificity

est	lower	upper
0.832	0.791	0.868

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

Table 74: Probability Thresholds

90%	80%
0.639	0.521

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

Table 75: Risk Ratio

est	lower	upper
1.8	1.54	2.11

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

Table 76: 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	0.923	0.986

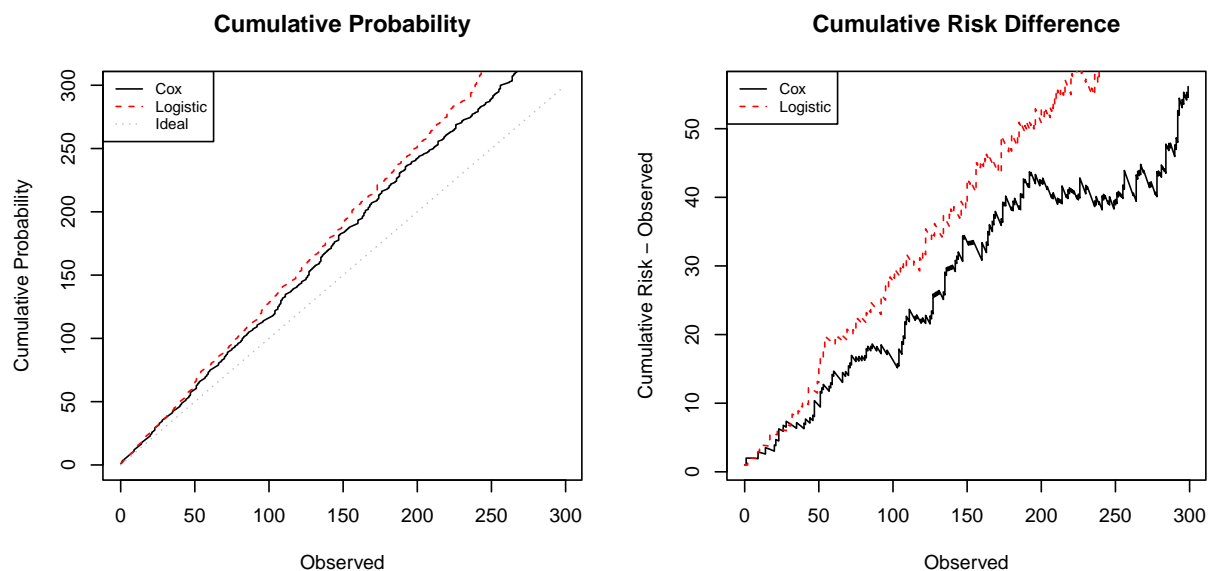
```
maxobs <- sum(dataBrestCancerTest$status)

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),
```

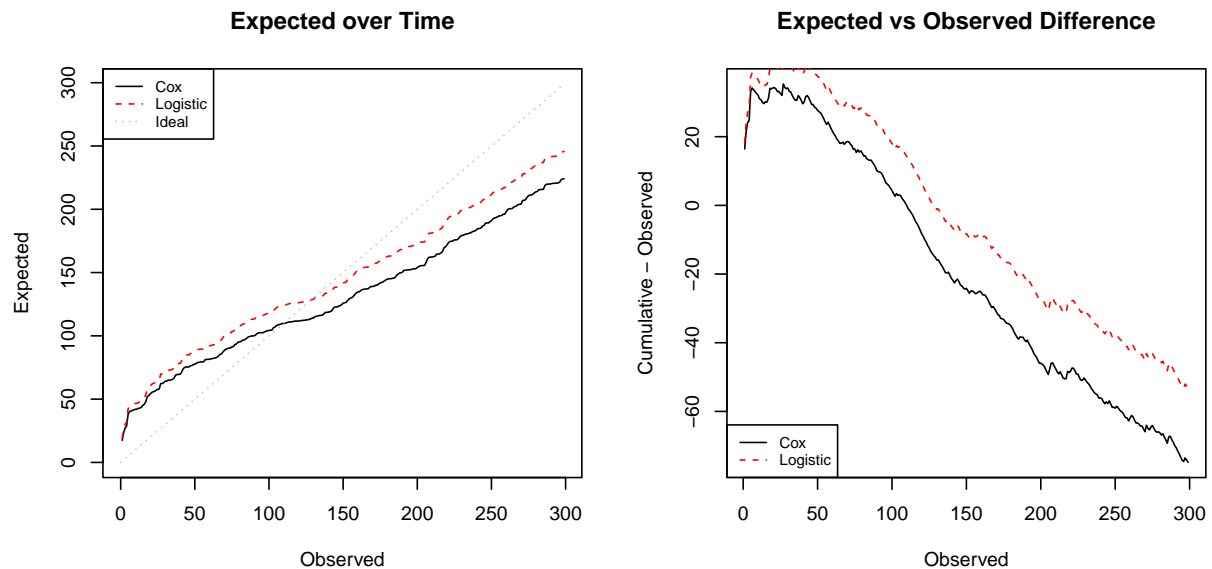
```

    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)

```