Colon Cancer

Jose Tamez

2023-04-26

Contents


```
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
```

```
#library(corrplot)
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)
```

```
data(cancer)
colon <- subset(colon,etype==1)
colon$etype <- NULL
rownames(colon) <- colon$id
colon$id <- NULL
```

```
colon <- colon[complete.cases(colon),]
time <- colon$time
status <- colon$status
data <- colon
data$time <- NULL
data$study <- NULL
table(data$status)
```

0 1 442 446

```
#dataColon <- as.data.frame(model.matrix(status~.*.,data))
dataColon <- as.data.frame(model.matrix(status~.*age,data))
dataColon$`(Intercept)` <- NULL
dataColon$time <- time/365
dataColon$status <- status
colnames(dataColon) <-str_replace_all(colnames(dataColon),":","_")
colnames(dataColon) <-str_replace_all(colnames(dataColon),"\\.","_")
colnames(dataColon) <-str_replace_all(colnames(dataColon),"\\+","_")
data <- NULL

trainsamples <- sample(nrow(dataColon),0.7*nrow(dataColon))
dataColonTrain <- dataColon[trainsamples,]
dataColonTest <- dataColon[-trainsamples,]


pander::pander(table(dataColonTrain$status))
```

0	1
320	301

```
pander::pander(table(dataColonTest$status))
```

0	1
122	145

0.1 Modeling

```
ml <- BSWiMS.model(Surv(time,status)~1,data=dataColonTrain,NumberofRepeats = 10)
```

[++]....

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

Table 3: Table continues below

	Estimate	lower	HR	upper	u.Accuracy
nodes	0.018702	1.013	1.019	1.025	0.602
node4	0.205428	1.142	1.228	1.321	0.612
age_node4	0.005120	1.003	1.005	1.007	0.612
age_nodes	0.000160	1.000	1.000	1.000	0.593

	Estimate	lower	HR	upper	u.Accuracy
age	-0.016078	0.978	0.984	0.991	0.517
age_extent	0.003637	1.002	1.004	1.006	0.517
extent	0.130698	1.062	1.140	1.223	0.543
rxLev_5FU_age	-0.001709	0.997	0.998	0.999	0.552
rxLev_5FU	-0.142480	0.782	0.867	0.962	0.552
surg	0.089631	1.016	1.094	1.177	0.546
age_surg	0.000302	1.000	1.000	1.001	0.546

Table 4: Table continues below

	r.Accuracy	full.Accuracy	u.AUC	r.AUC	full.AUC
nodes	0.512	0.607	0.597	0.518	0.603
node4	0.583	0.620	0.604	0.586	0.613
age_node4	0.590	0.628	0.604	0.594	0.627
age_nodes	0.498	0.594	0.588	0.508	0.589
age	0.619	0.628	0.515	0.614	0.627
age_extent	0.615	0.628	0.518	0.611	0.627
extent	0.612	0.620	0.554	0.604	0.613
rxLev_5FU_age	0.620	0.620	0.558	0.613	0.613
rxLev_5FU	0.632	0.628	0.558	0.629	0.627
surg	0.605	0.628	0.539	0.602	0.627
age_surg	0.603	0.628	0.539	0.601	0.626

	IDI	NRI	z.IDI	z.NRI	Delta.AUC	Frequency
nodes	0.02913	0.390	5.87	5.21	0.085718	1.0
node4	0.03800	0.418	5.51	6.12	0.027357	1.0
age_node4	0.03330	0.418	5.14	6.12	0.033102	1.0
age_nodes	0.02376	0.350	5.03	4.69	0.080486	1.0
age	0.02789	0.220	4.77	2.76	0.012935	1.0
age_extent	0.01884	0.215	3.86	3.92	0.015303	1.0
extent	0.01718	0.215	3.63	3.92	0.008799	1.0
rxLev_5FU_age	0.01209	0.231	3.18	3.10	-0.000197	1.0
rxLev_5FU	0.00879	0.231	2.68	3.10	-0.002800	1.0
surg	0.00576	0.156	2.37	2.18	0.024751	1.0
age_surg	0.00513	0.156	2.23	2.18	0.025497	0.5

0.2 Cox Model Performance

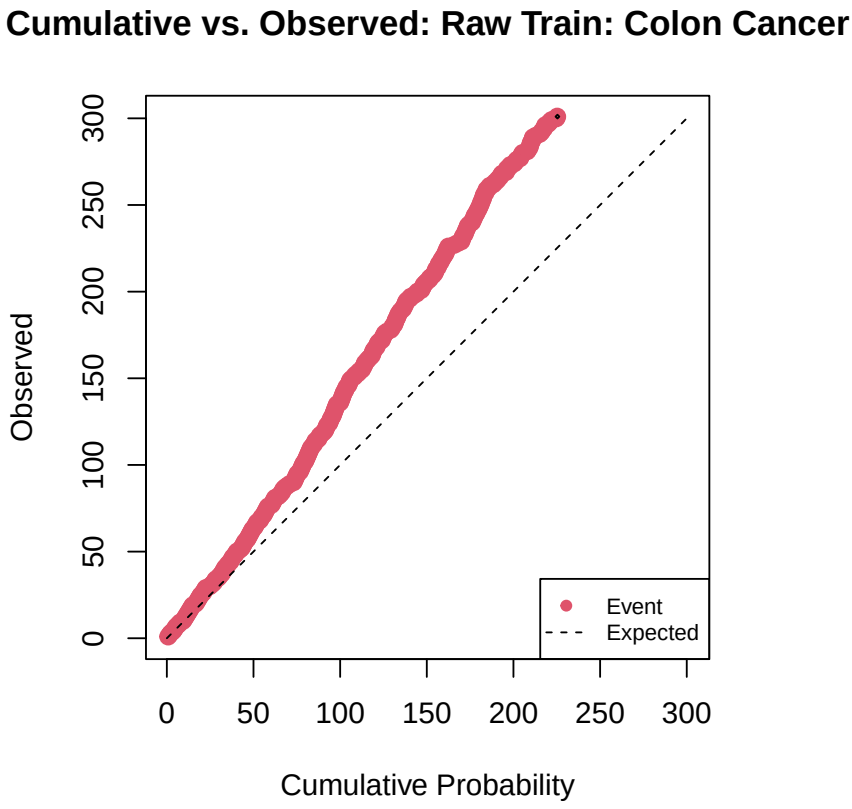
Here we evaluate the model using the RRPlot() function.

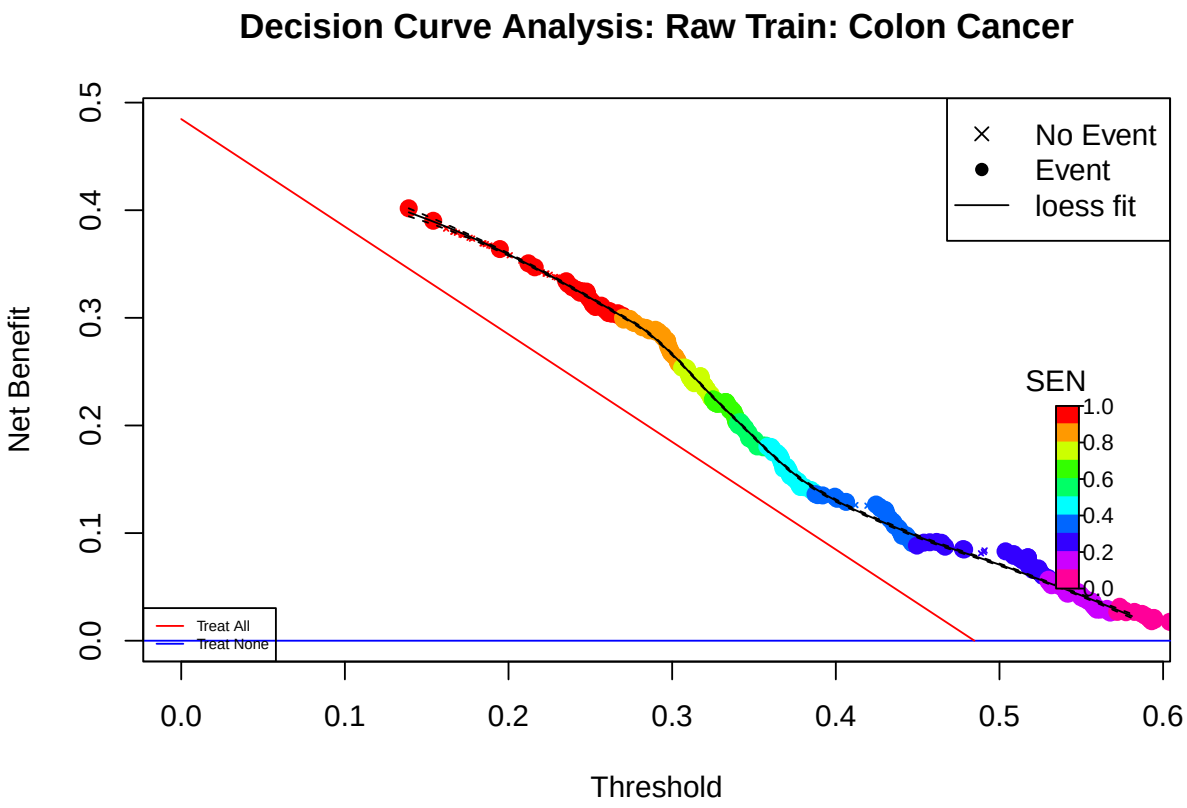
0.2.1 The evaluation of the raw Cox model with RRPlot()

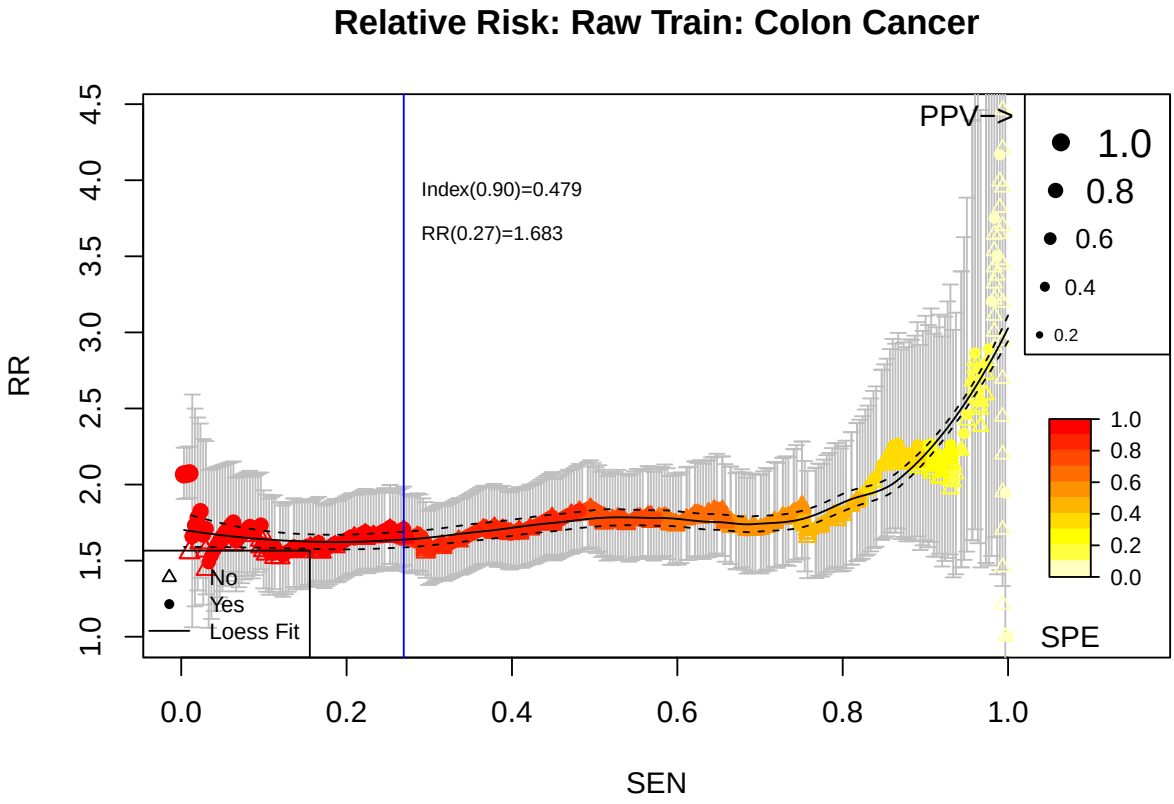
Here we will use the predicted event probability assuming a baseline hazard for events withing 5 years

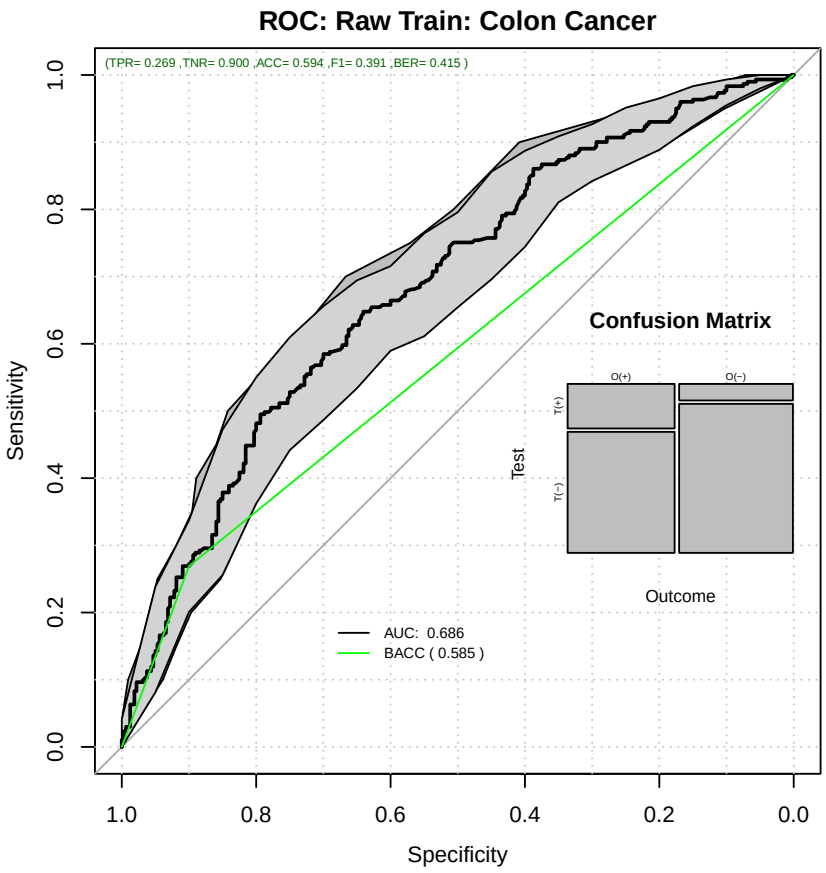
```
index <- predict(ml,dataColonTrain)
timeinterval <- 2*mean(subset(dataColonTrain,status==1)$time)

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

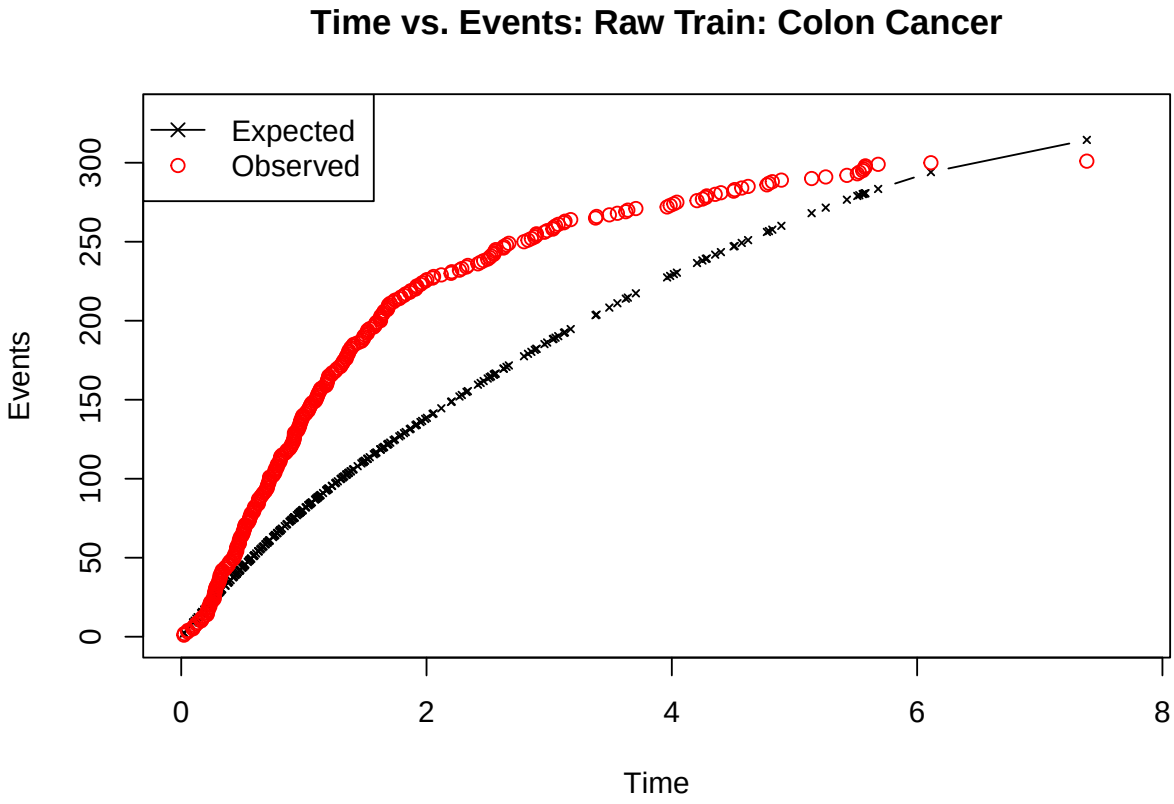
```
rdata <- cbind(dataColonTrain$status,ppoisGzero(index,h0))

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

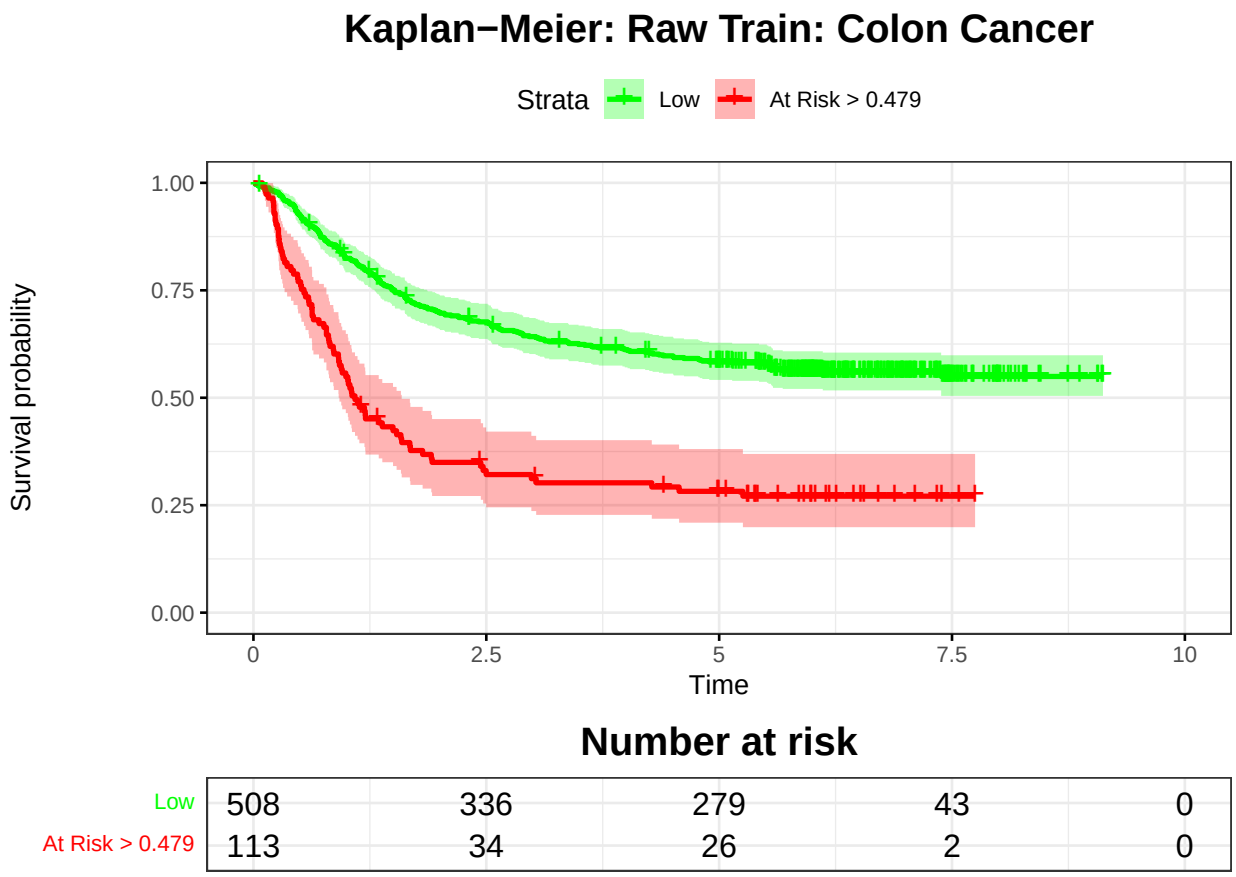



Decision Curve Analysis: Raw Train: Colon Cancer
Net Benefit
0.0
0.1
0.2
0.3
0.4
0.5
Threshold
0.0
0.1
0.2
0.3
0.4
0.5
0.6
No Event
Event
loess fit
SEN
1.0
0.8
0.6
0.4
0.2
0.0
Treat All
Treat None


Relative Risk: Raw Train: Colon Cancer
RR
SEN
PPV–>
Index(0.90)=0.479
RR(0.27)=1.683
No
Yes
Loess Fit
SPE
1.0
0.8
0.6
0.4
0.2
0.0


ROC: Raw Train: Colon Cancer
(TPR= 0.269 ,TNR= 0.900 ,ACC= 0.594 ,F1= 0.391 ,BER= 0.415)
Sensitivity
Specificity
1.0
0.8
0.6
0.4
0.2
0.0
AUC: 0.686
BACC (0.585)
Confusion Matrix
O(+)
O(−)
T(+)
T(−)
Test
Outcome


Time vs. Events: Raw Train: Colon Cancer
Expected
Observed
Events
0
50
100
150
200
250
300
0
2
4
6
8
Time



0.2.2 Uncalibrated Performance Report

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

Table 6: O/E Ratio

est	lower	upper
0.957	0.852	1.07

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

Table 7: O/E Ratio

mean	50%	2.5%	97.5%
1.5	1.5	1.47	1.53

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

Table 8: O/Acum Ratio

mean	50%	2.5%	97.5%
1.34	1.34	1.34	1.35

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

mean.C Index	median	lower	upper
0.66	0.66	0.629	0.69

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

Table 10: ROC AUC

est	lower	upper
0.686	0.644	0.727

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

Table 11: Sensitivity

est	lower	upper
0.269	0.22	0.323

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

Table 12: Specificity

est	lower	upper
0.9	0.862	0.931

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

Table 13: Probability Thresholds

90%
0.479

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

Table 14: Risk Ratio

est	lower	upper
1.68	1.45	1.96

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

Table 15: Logrank test Chisq = 56.555705 on 1 degrees of freedom, p = 0.000000

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	508	220	263.1	7.06	56.6
class=1	113	81	37.9	49.00	56.6

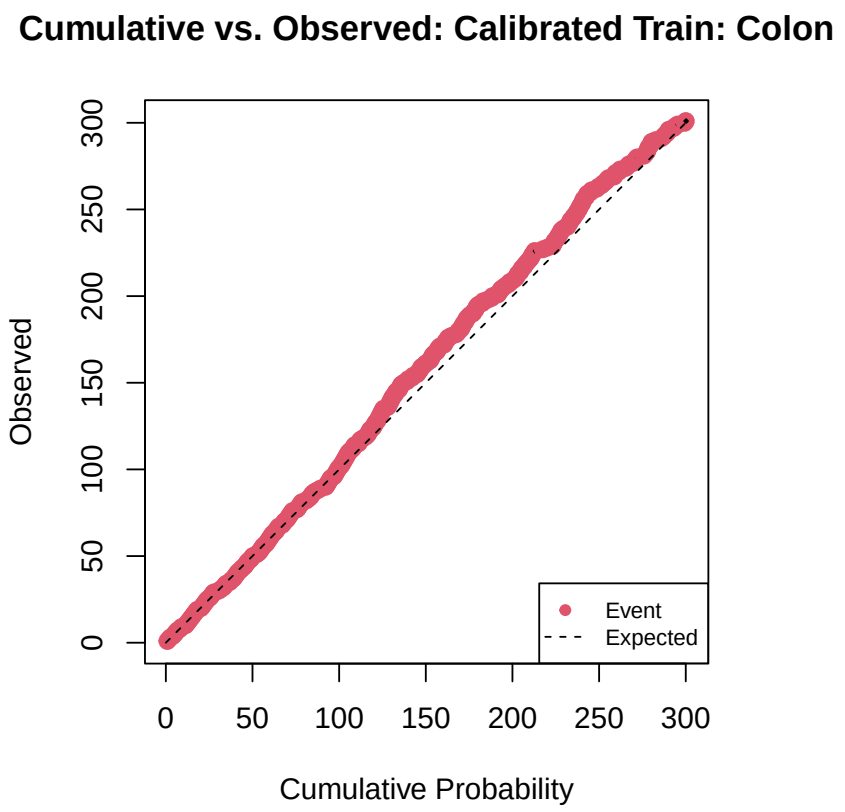
0.2.3 Cox Calibration

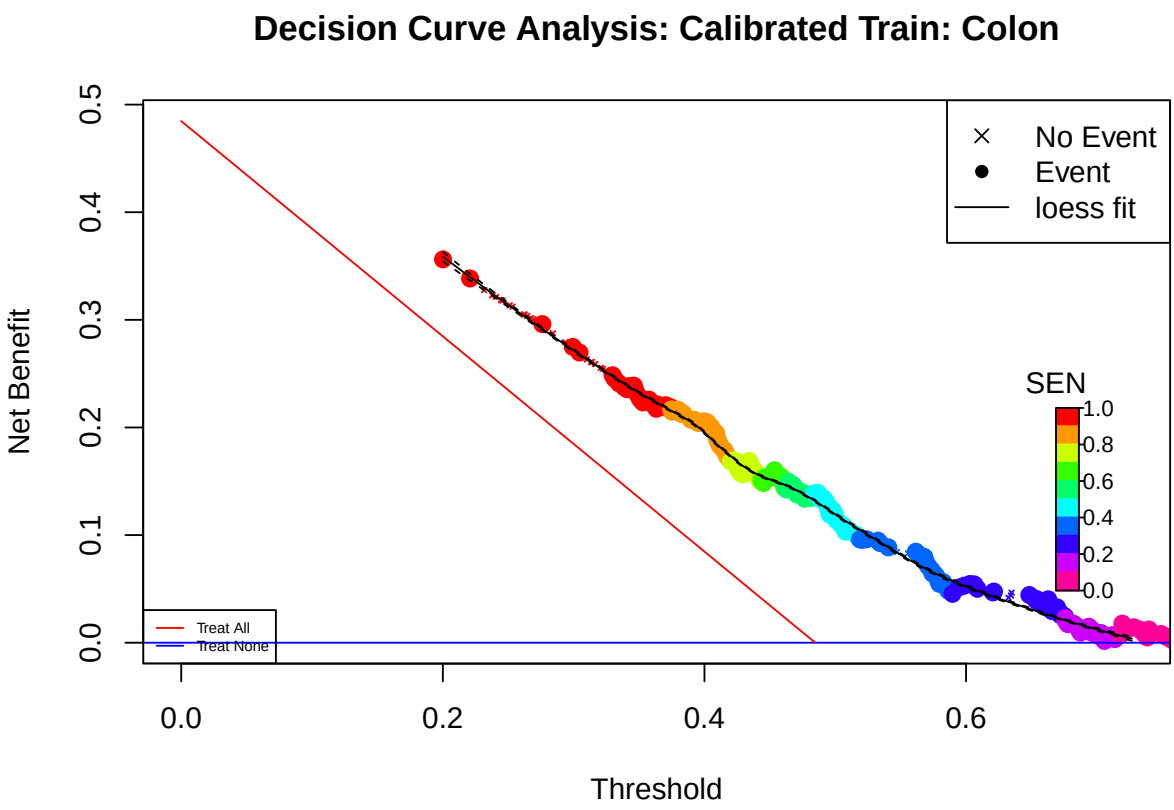
```
op <- par(no.readonly = TRUE)


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

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

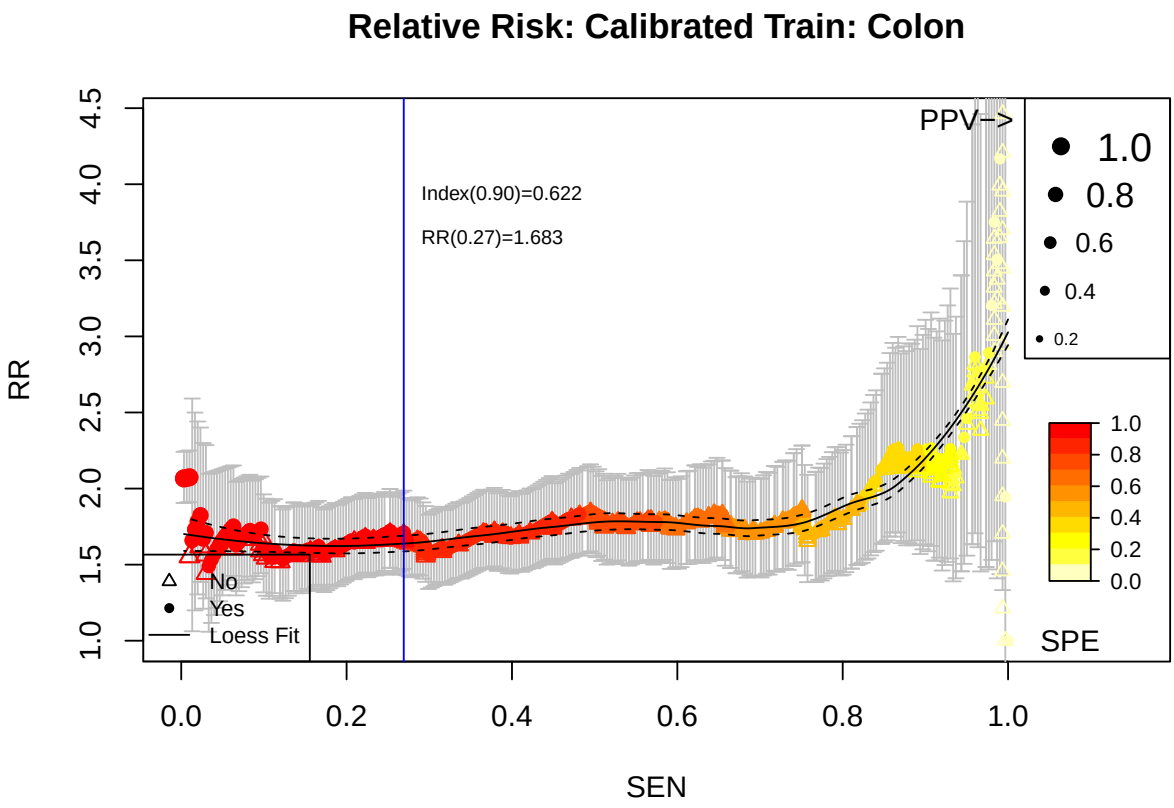
h0	Gain	DeltaTime
0.646	1.49	2.98

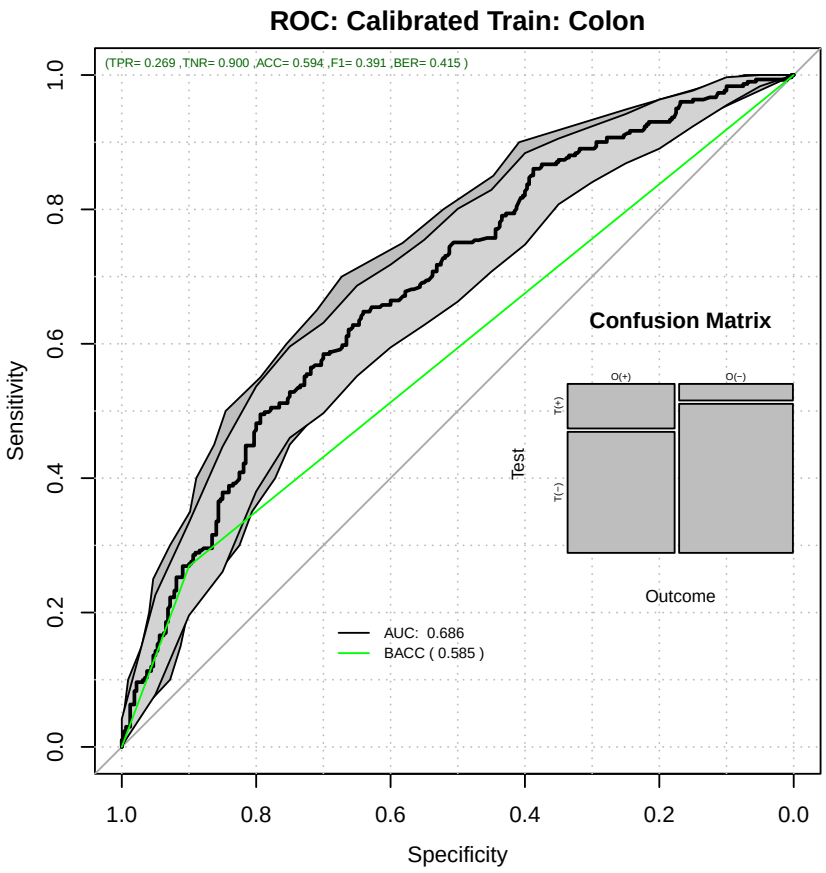
0.2.4 The RRplot() of the calibrated model

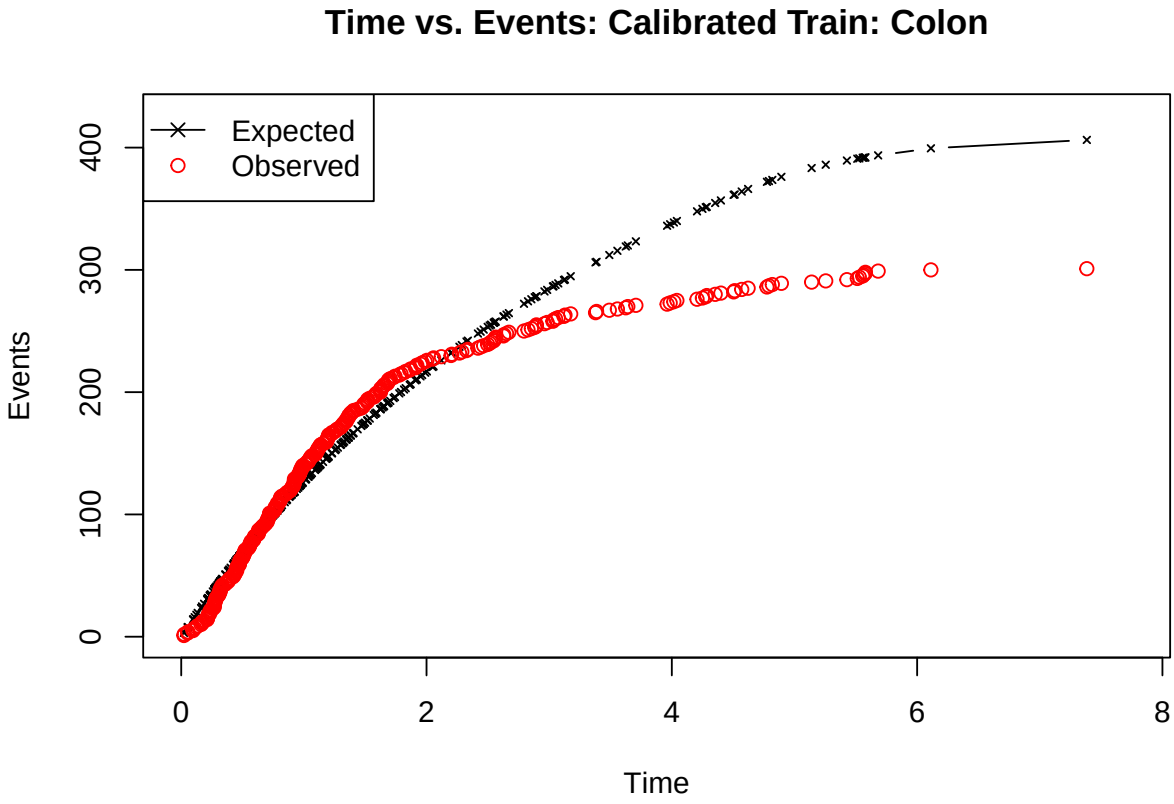
```
h0 <- calprob$h0
timeinterval <- calprob$timeInterval;

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


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

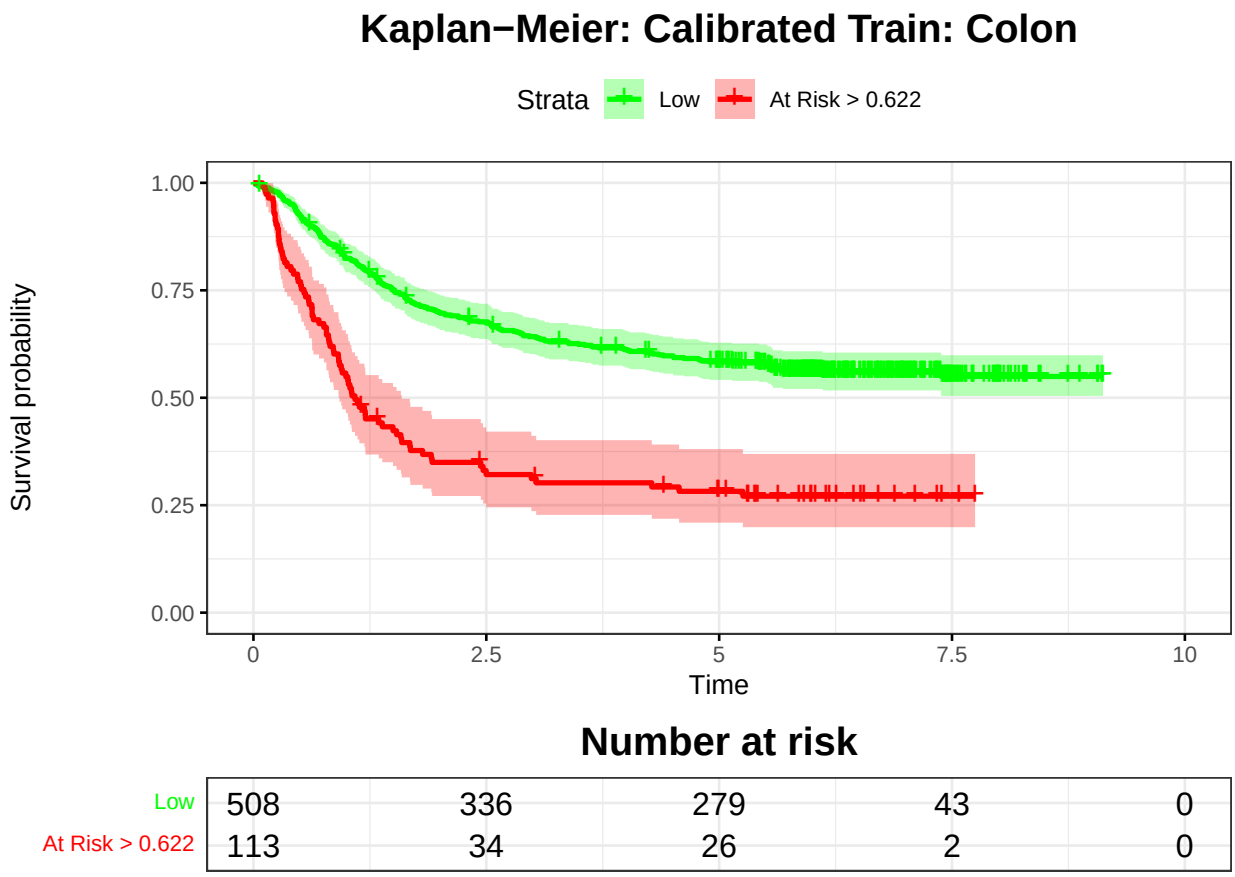

Cumulative vs. Observed: Calibrated Train: Colon
Observed
0
50
100
150
200
250
300
Event
Expected
Cumulative Probability


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


Relative Risk: Calibrated Train: Colon
RR
SEN
PPV–>
Index(0.90)=0.622
RR(0.27)=1.683
No
Yes
Loess Fit
SPE
1.0
0.8
0.6
0.4
0.2
0.0
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
0.0
0.2
0.4
0.6
0.8
1.0


ROC: Calibrated Train: Colon
(TPR= 0.269 ,TNR= 0.900 ,ACC= 0.594 ,F1= 0.391 ,BER= 0.415)
Sensitivity
Specificity
1.0
0.8
0.6
0.4
0.2
0.0
AUC: 0.686
BACC (0.585)
Confusion Matrix
O(+)
O(−)
T(+)
T(−)
Test
Outcome


Time vs. Events: Calibrated Train: Colon
Expected
Observed
Events
0
100
200
300
400
0
2
4
6
8
Time



0.2.5 Calibrated Train Performance

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

Table 17: O/E Ratio

est	lower	upper
0.741	0.659	0.829

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

Table 18: O/E Ratio

mean	50%	2.5%	97.5%
0.964	0.964	0.948	0.981

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

Table 19: O/Acum Ratio

mean	50%	2.5%	97.5%
1.04	1.04	1.03	1.04

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

mean.C Index	median	lower	upper
0.66	0.659	0.629	0.688

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

Table 21: ROC AUC

est	lower	upper
0.686	0.644	0.727

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

Table 22: Sensitivity

est	lower	upper
0.269	0.22	0.323

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

Table 23: Specificity

est	lower	upper
0.9	0.862	0.931

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

Table 24: Probability Thresholds

90%
0.622

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

Table 25: Risk Ratio

est	lower	upper
1.68	1.45	1.96

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

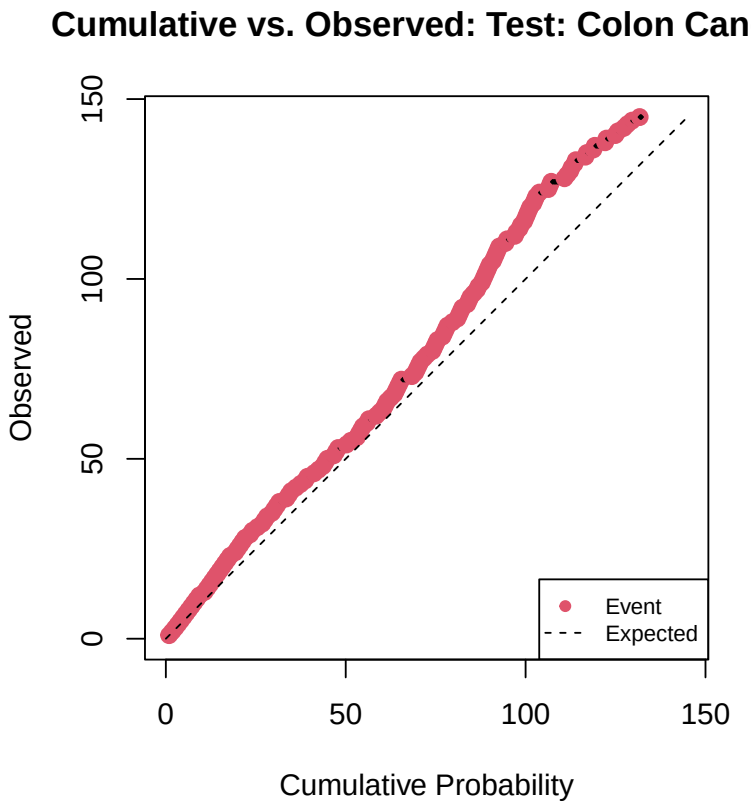
Table 26: Logrank test Chisq = 56.555705 on 1 degrees of freedom, p = 0.000000

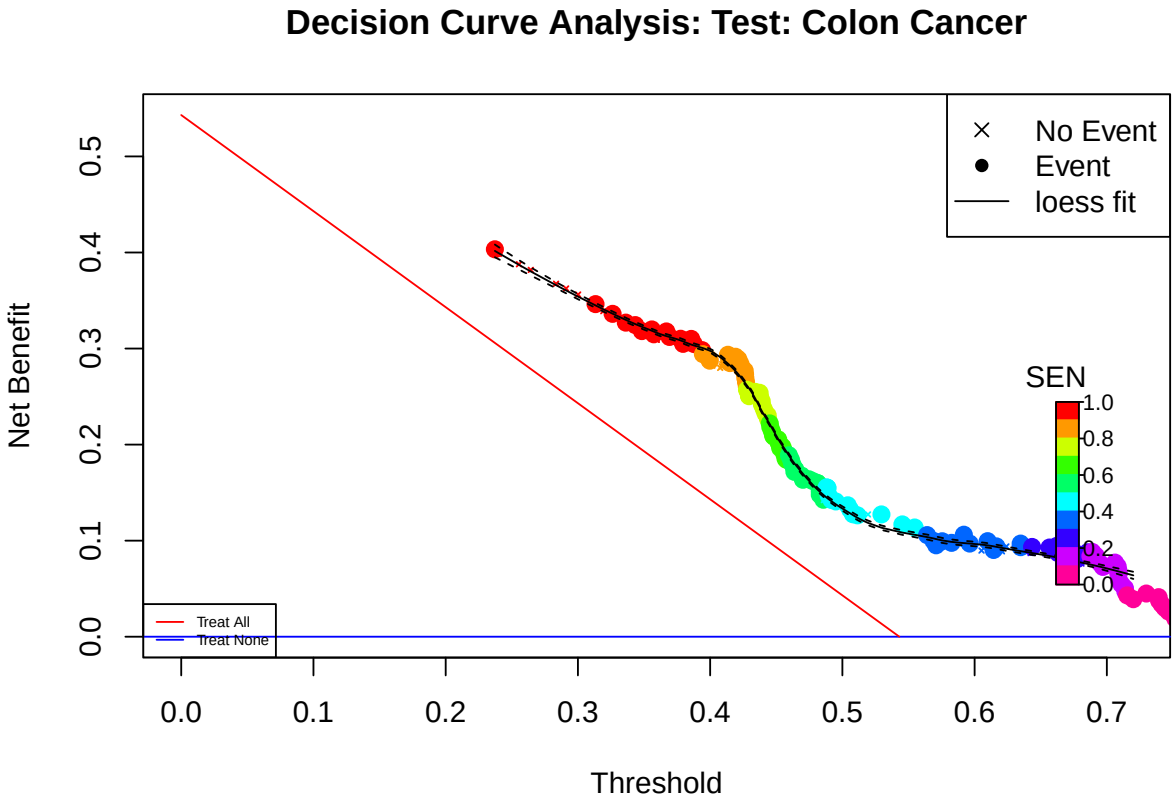
	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	508	220	263.1	7.06	56.6
class=1	113	81	37.9	49.00	56.6

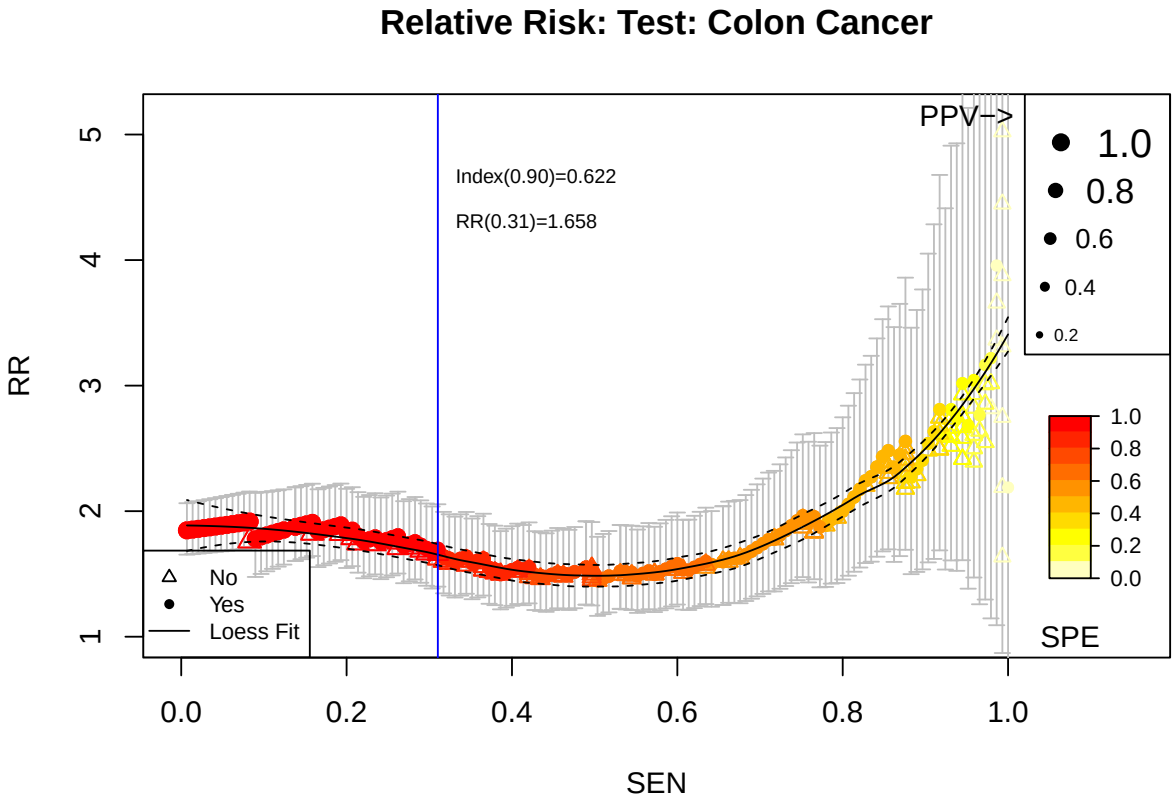
0.2.6 Evaluating on the test set

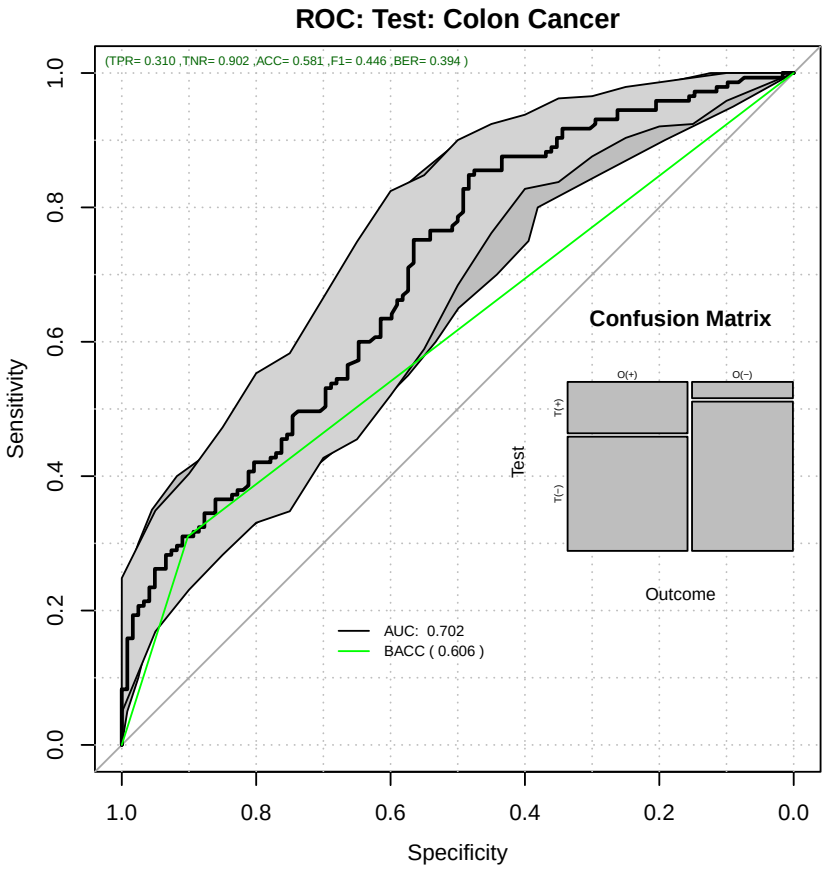
The calibrated h0 and timeinterval were estimated on the training set

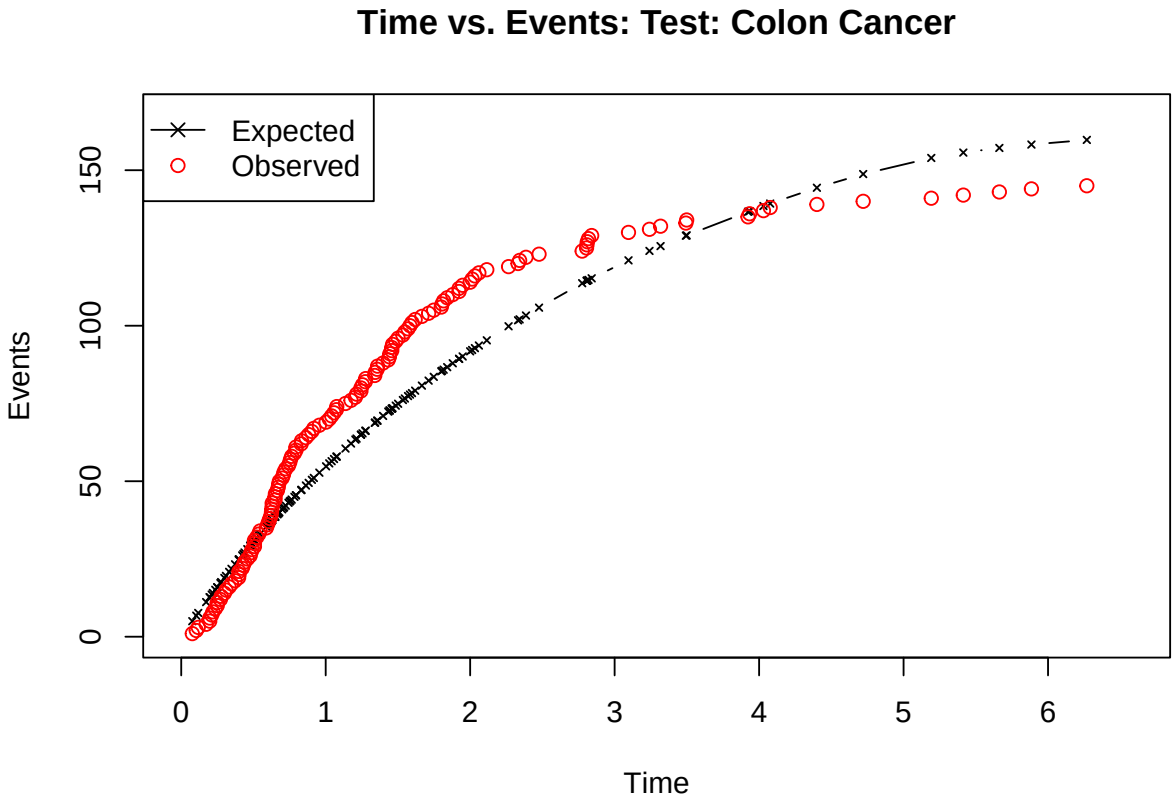
```
index <- predict(ml,dataColonTest)
rdata <- cbind(dataColonTest$status,ppoisGzero(index,h0))

rrAnalysisTest <- RRPlot(rdata,atThr = rrAnalysisTrain$thr_atP,
                     timetoEvent=dataColonTest$time,
                     title="Test: Colon Cancer",
                     ysurvlim=c(0.00,1.0),
                     riskTimeInterval=timeinterval)
```

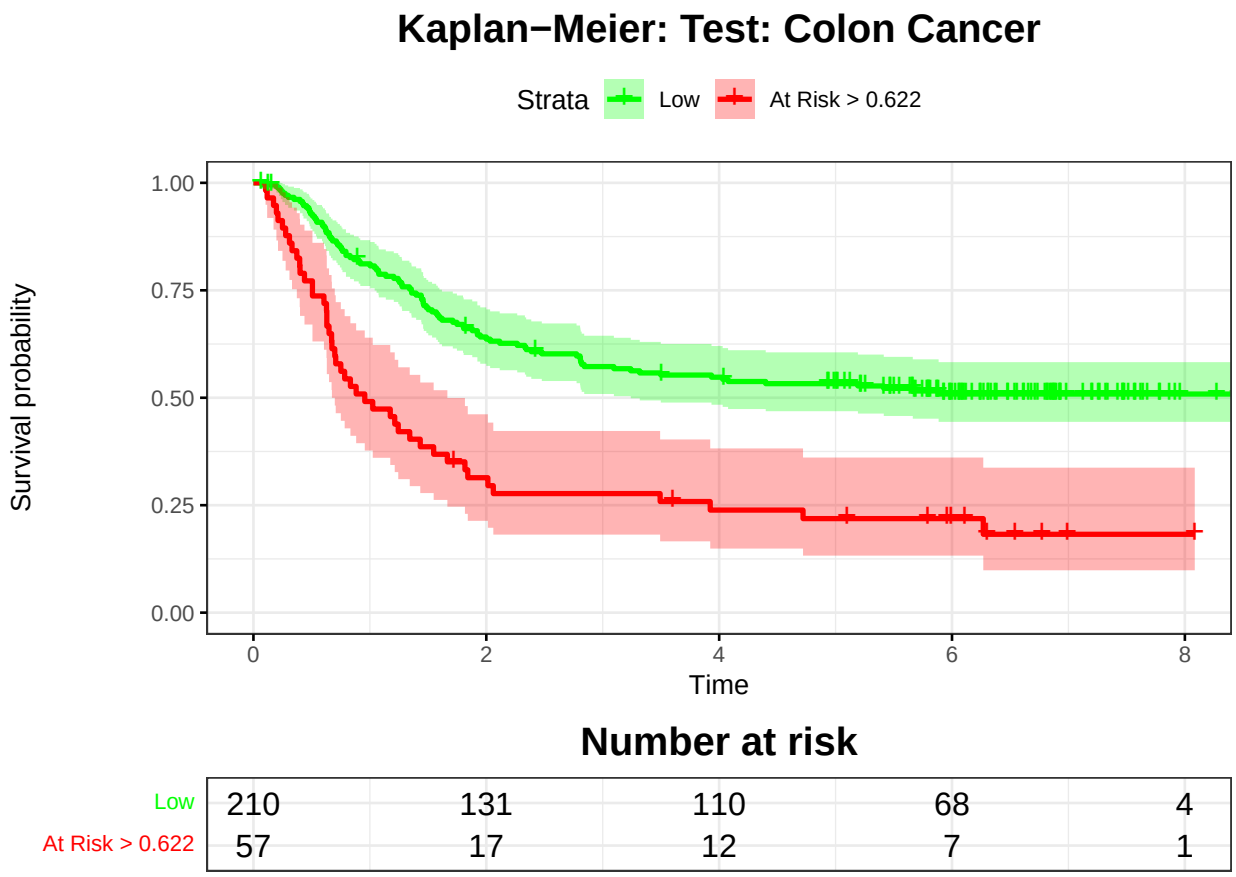

Cumulative vs. Observed: Test: Colon Cancer
Observed
0
50
100
150
Cumulative Probability
0
50
100
150
Event
Expected


Decision Curve Analysis: Test: Colon 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
0.5
0.6
0.7
Threshold


Relative Risk: Test: Colon Cancer
PPV–>
Index(0.90)=0.622
RR(0.31)=1.658
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
SEN
1
2
3
4
5
0.0
0.2
0.4
0.6
0.8
1.0


ROC: Test: Colon Cancer
(TPR= 0.310 ,TNR= 0.902 ,ACC= 0.581 ,F1= 0.446 ,BER= 0.394)
Sensitivity
Specificity
1.0
0.8
0.6
0.4
0.2
0.0
AUC: 0.702
BACC (0.606)
Confusion Matrix
O(+)
O(−)
T(+)
T(−)
Test
Outcome


Time vs. Events: Test: Colon Cancer
Expected
Observed
Events
0
50
100
150
0
1
2
3
4
5
6
Time



0.2.7 Test Performance

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

Table 27: O/E Ratio

est	lower	upper
0.908	0.766	1.07

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

Table 28: O/E Ratio

mean	50%	2.5%	97.5%
1.12	1.12	1.09	1.15

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

Table 29: O/Acum Ratio

mean	50%	2.5%	97.5%
1.14	1.14	1.13	1.14

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

mean.C Index	median	lower	upper
0.653	0.654	0.609	0.698

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

Table 31: ROC AUC

est	lower	upper
0.702	0.64	0.764

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

Table 32: Sensitivity

est	lower	upper
0.31	0.236	0.392

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

Table 33: Specificity

est	lower	upper
0.902	0.834	0.948

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

Table 34: Probability Thresholds

90%
0.622

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

Table 35: Risk Ratio

est	lower	upper
1.66	1.36	2.02

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

Table 36: Logrank test Chisq = 30.415807 on 1 degrees of freedom, p = 0.000000

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	210	100	123.5	4.47	30.4
class=1	57	45	21.5	25.65	30.4

0.3 Cross-Validation

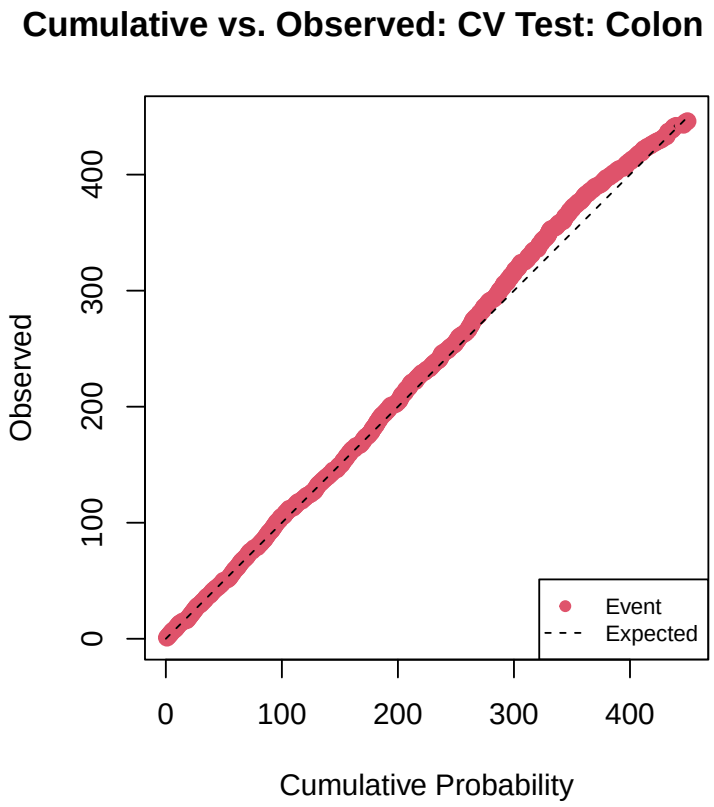
Here we will cross validate the training set and evaluate also on the testing set. The h0 and the timeinterval are the ones estimated on the calibration process

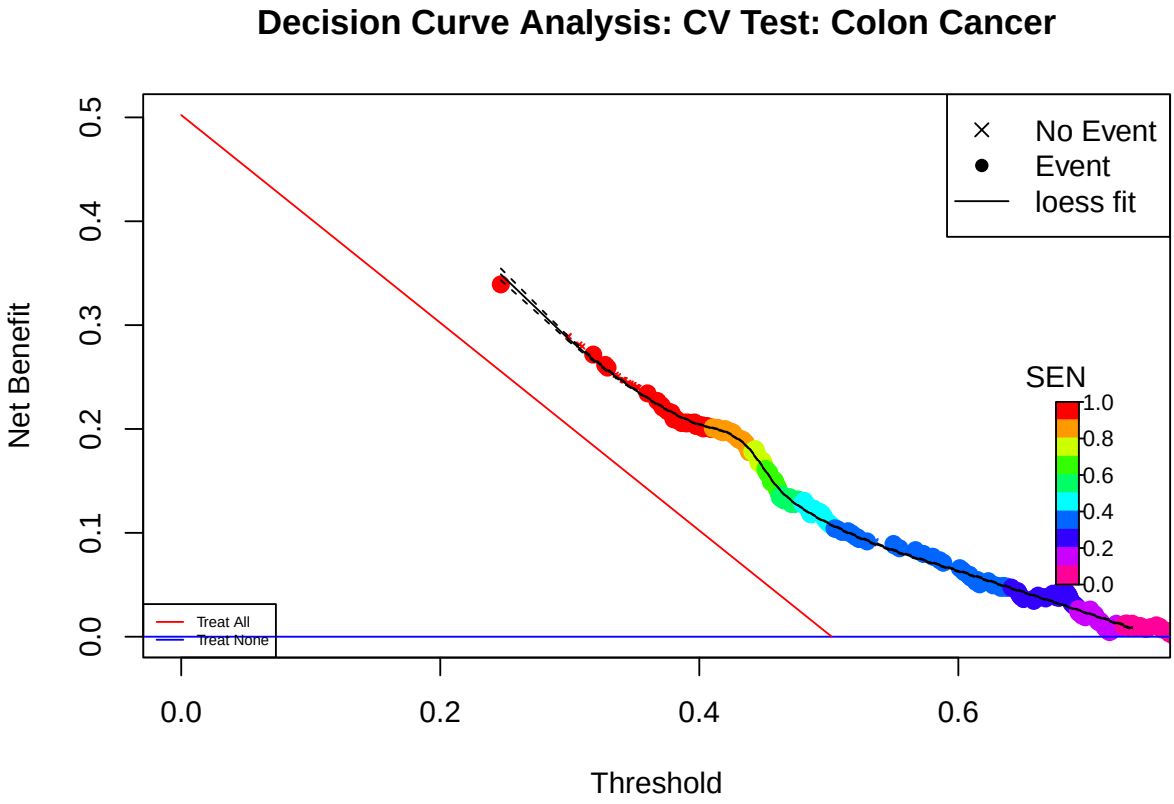
```
rcv <- randomCV(theData=dataColonTrain,
                theOutcome = Surv(time,status)~1,
                fittingFunction=BSWiMS.model,
                trainFraction = 0.75,
                repetitions=50,
                classSamplingType = "Pro",
                testingSet=dataColonTest
         )
```

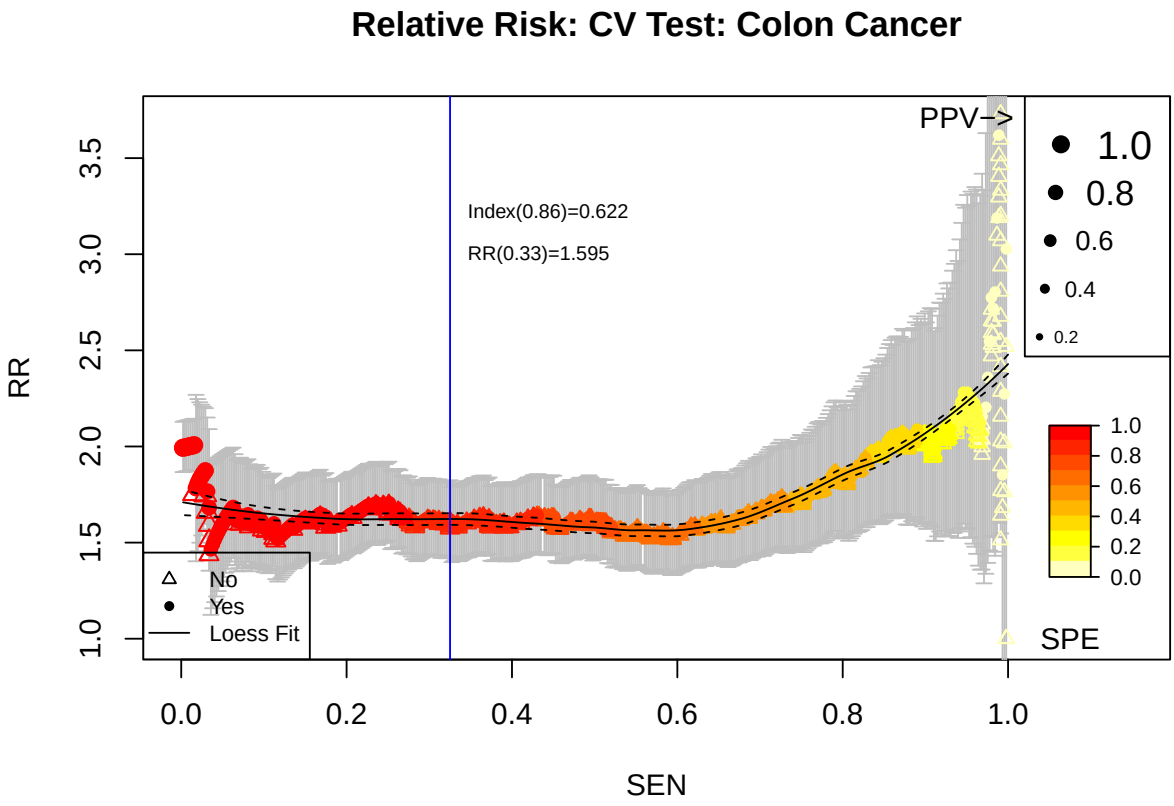
.[++++-].[++++].[++++].[++++–].[+++++].[++++].[+++++].[++++].[+++++].[++++++]10 Tested: 858 Avg. Selected: 8 Min Tests: 1 Max Tests: 10 Mean Tests: 4.93007 . MAD: 0.4773942 .[+++].[++++].[++++].[++++].[+++++].[++++-].[++++].[+++++].[++++-].[++-+-]20 Tested: 887 Avg. Selected: 8.4 Min Tests: 1 Max Tests: 20 Mean Tests: 9.537768 . MAD: 0.4757449 .[+++–].[++++].[++++++++++].[+++-].[++++].[++++].[++++].[++++++].[++++++].[+++++]30 Tested: 888 Avg. Selected: 8.333333 Min Tests: 1 Max Tests: 30 Mean Tests: 14.29054 . MAD: 0.4755663 .[++++].[++++++++].[+++-].[++++].[+++-].[++++].[++++].[++++-].[+++].[++++++]40 Tested: 888 Avg. Selected: 8.25 Min Tests: 3 Max Tests: 40 Mean Tests: 19.05405 . MAD: 0.4757152 .[+++-].[++++].[+++-].[++++].[+++–].[+++++].[++++].[++++].[+++-].[++++]50 Tested: 888 Avg. Selected: 8.16 Min Tests: 4 Max Tests: 50 Mean Tests: 23.81757 . MAD: 0.4752228

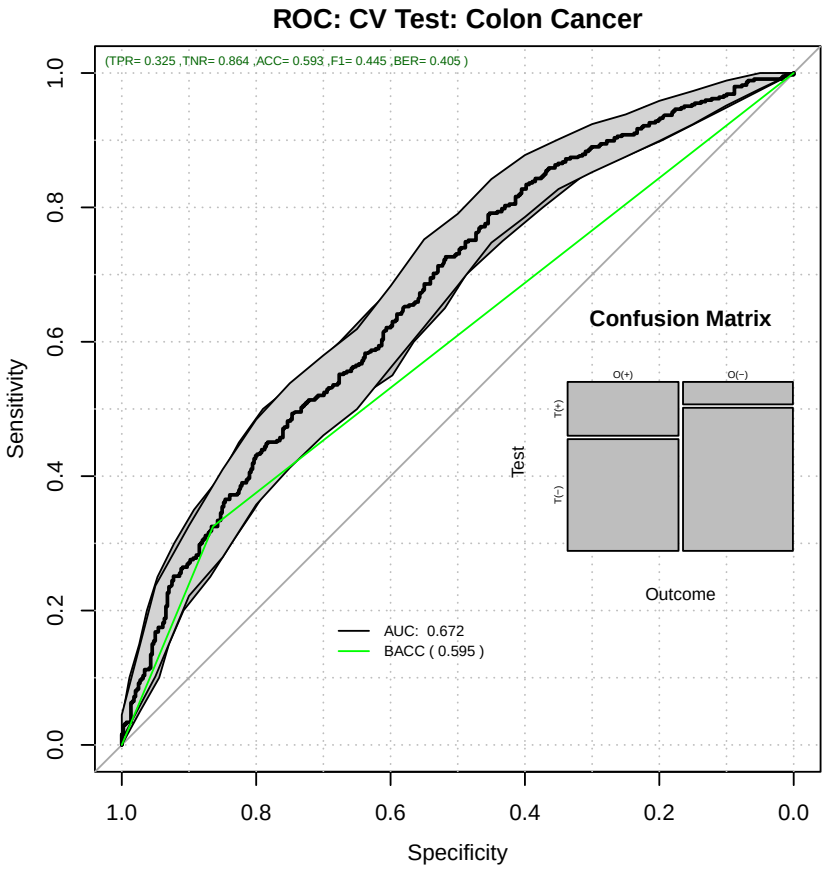
```
stp <- rcv$survTestPredictions
stp <- stp[!is.na(stp[,4]),]

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

rrAnalysisCVTest <- RRPlot(rdatacv,atThr = rrAnalysisTrain$thr_atP,
                     timetoEvent=times,
```

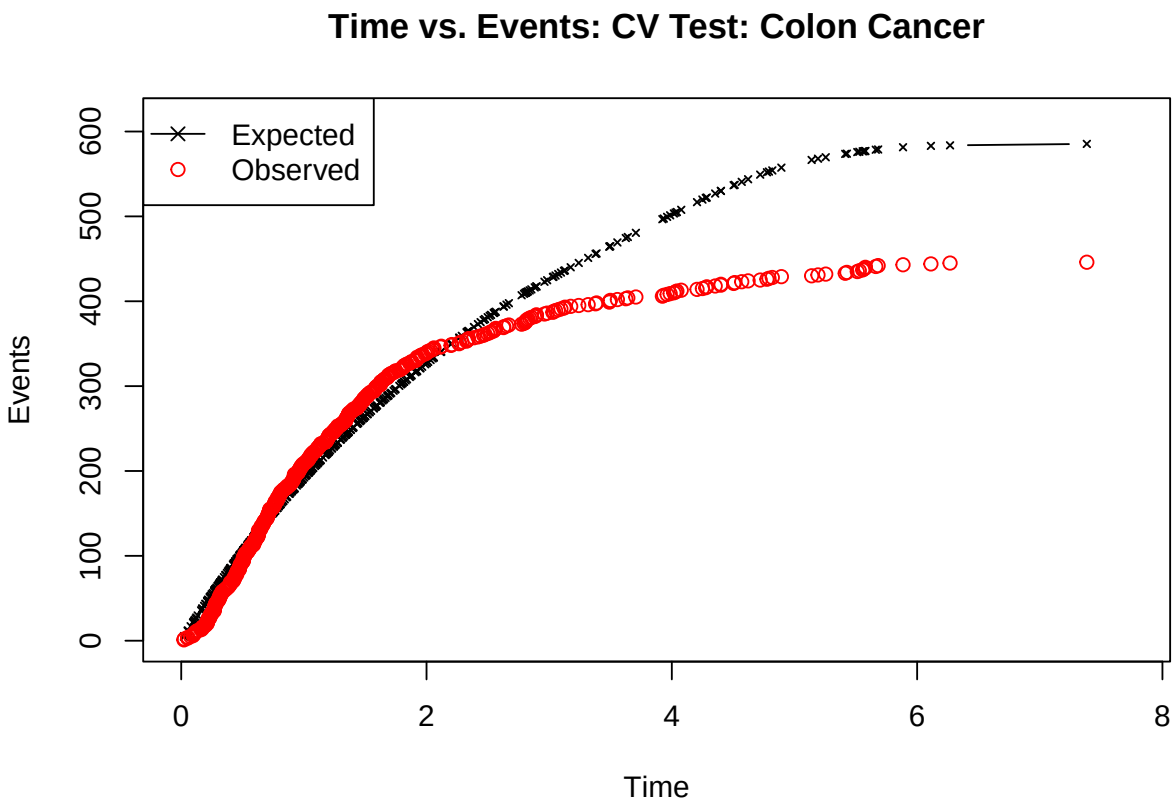
```
                    title="CV Test: Colon Cancer",
                    ysurvlim=c(0.00,1.0),
                    riskTimeInterval=timeinterval)
```

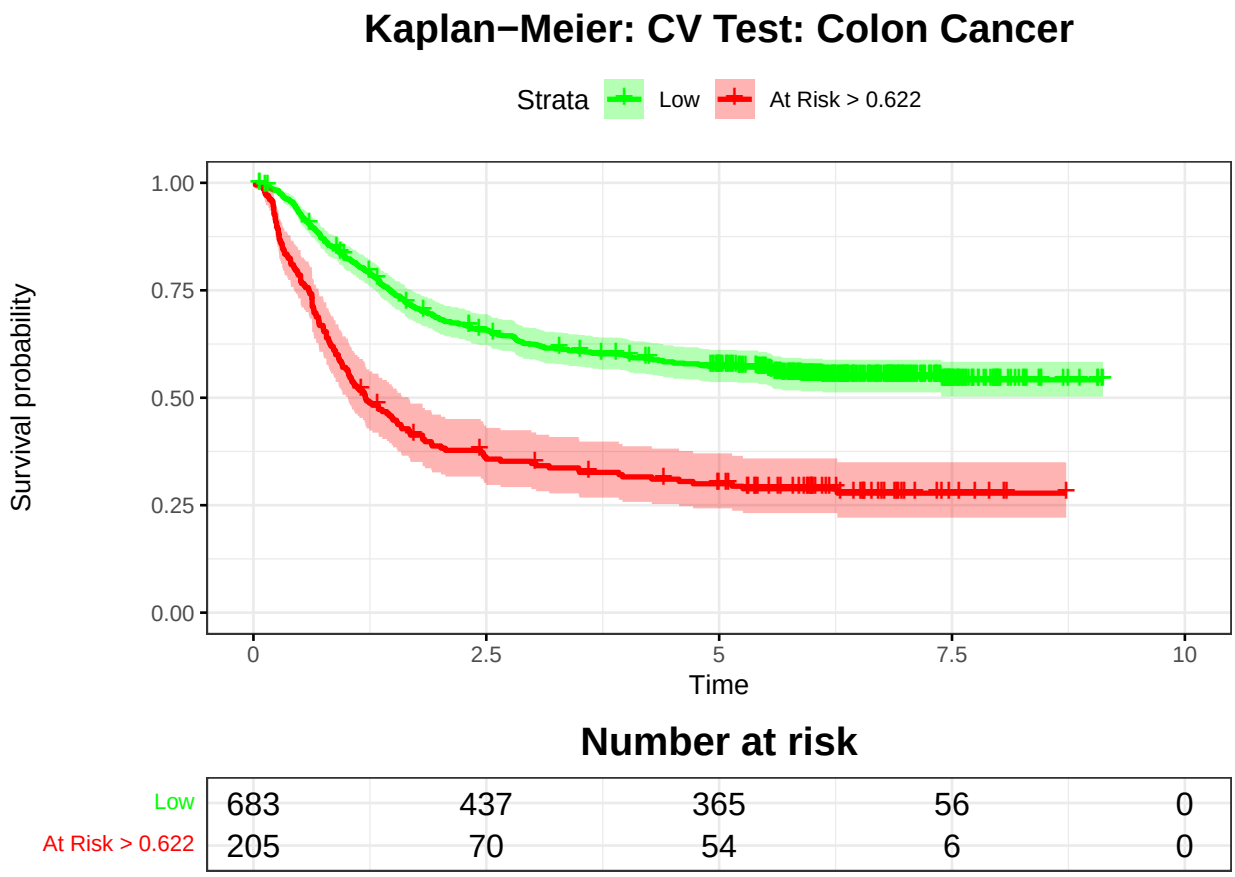



Decision Curve Analysis: CV Test: Colon Cancer
No Event
Event
loess fit
SEN
1.0
0.8
0.6
0.4
0.2
0.0
Treat All
Treat None
Net Benefit
0.0
0.1
0.2
0.3
0.4
0.5
0.0
0.2
0.4
0.6
Threshold


Relative Risk: CV Test: Colon Cancer
PPV–>
Index(0.86)=0.622
RR(0.33)=1.595
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
0.0
0.2
0.4
0.6
0.8
1.0
SEN


ROC: CV Test: Colon Cancer
(TPR= 0.325 ,TNR= 0.864 ,ACC= 0.593 ,F1= 0.445 ,BER= 0.405)
Sensitivity
Specificity
1.0
0.8
0.6
0.4
0.2
0.0
Confusion Matrix
O(+)
O(−)
T(+)
T(−)
Test
Outcome
AUC: 0.672
BACC (0.595)


Time vs. Events: CV Test: Colon Cancer
Expected
Observed
Events
0
100
200
300
400
500
600
0
2
4
6
8
Time



0.3.1 CV Test Performance

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

Table 37: O/E Ratio

est	lower	upper
0.762	0.693	0.836

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

Table 38: O/E Ratio

mean	50%	2.5%	97.5%
0.954	0.954	0.942	0.967

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

Table 39: O/Acum Ratio

mean	50%	2.5%	97.5%
1.03	1.03	1.03	1.03

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

mean.C Index	median	lower	upper
0.645	0.646	0.62	0.67

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

Table 41: ROC AUC

est	lower	upper
0.672	0.637	0.707

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

Table 42: Sensitivity

est	lower	upper
0.325	0.282	0.371

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

Table 43: Specificity

est	lower	upper
0.864	0.829	0.895

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

Table 44: Probability Thresholds

90%
0.622

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

Table 45: Risk Ratio

est	lower	upper
1.59	1.41	1.8

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

Table 46: Logrank test Chisq = 74.685264 on 1 degrees of freedom, p = 0.000000

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	683	301	369.5	12.7	74.7
class=1	205	145	76.5	61.4	74.7