Colon Cancer

Jose Tamez

2023-04-28

Contents

1 RRPlot and the Colon data set

1.0.1 Libraries

```
library(survival)
library(FRESA.CAD)
```

```
## Loading required package: Rcpp
```

```
## Loading required package: stringr
```

```
## Loading required package: miscTools
```

```
## Loading required package: Hmisc
```

```
##
## Attaching package: 'Hmisc'
```

```
## The following objects are masked from 'package:base':
##
##     format.pval, units
```

```
## Loading required package: pROC
```

```
## Type 'citation("pROC")' for a citation.
```

```
##
## Attaching package: 'pROC'
```

```
## The following objects are masked from 'package:stats':
##
##     cov, smooth, var
```

```
#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)
```

1.1 The data set

```
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~.*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
313	308

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

0	1
129	138

1.2 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
age_nodes	0.000491	1.000	1.000	1.001	0.599
rxLev_5FU_age	-0.004668	0.993	0.995	0.997	0.576
rxLev_5FU	-0.095014	0.869	0.909	0.952	0.576
age_node4	0.001524	1.001	1.002	1.002	0.601
node4	0.243768	1.088	1.276	1.496	0.601
extent	0.218117	1.086	1.244	1.424	0.546
age	-0.006355	0.989	0.994	0.998	0.510
rxLev	0.015881	1.001	1.016	1.031	0.541
nodes	0.024687	1.003	1.025	1.047	0.609

Table 4: Table continues below

	r.Accuracy	full.Accuracy	u.AUC	r.AUC	full.AUC
age_nodes	0.541	0.614	0.598	0.541	0.614
rxLev_5FU_age	0.607	0.625	0.578	0.606	0.627
rxLev_5FU	0.607	0.610	0.578	0.606	0.610
age_node4	0.598	0.610	0.599	0.598	0.610
node4	0.616	0.625	0.599	0.617	0.627
extent	0.624	0.625	0.549	0.623	0.627
age	0.600	0.613	0.510	0.599	0.613
rxLev	0.614	0.618	0.540	0.613	0.618
nodes	0.626	0.625	0.608	0.627	0.626

	IDI	NRI	z.IDI	z.NRI	Delta.AUC	Frequency
age_nodes	0.02389	0.362	4.92	4.79	0.07320	1.0
rxLev_5FU_age	0.02439	0.311	4.61	4.18	0.02104	1.0
rxLev_5FU	0.01992	0.311	4.10	4.18	0.00379	0.6
age_node4	0.01222	0.329	3.34	4.89	0.01158	0.6
node4	0.01172	0.327	3.17	4.85	0.00998	1.0
extent	0.01252	0.169	3.16	2.98	0.00386	1.0
age	0.00911	0.129	2.79	1.63	0.01351	1.0
rxLev	0.00531	0.159	2.26	2.11	0.00500	0.9
nodes	0.00462	0.173	2.23	2.27	-0.00118	1.0

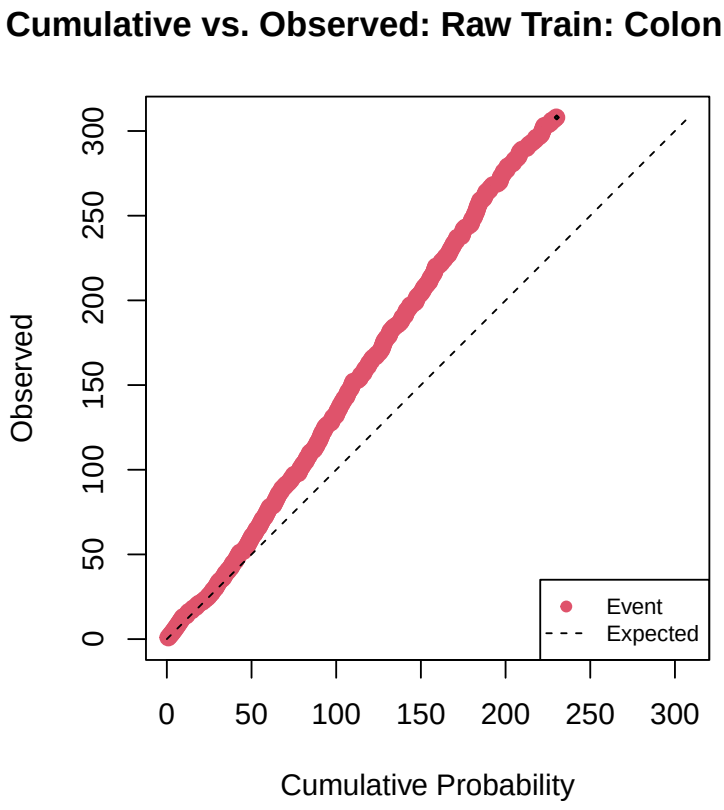
1.3 Cox Model Performance

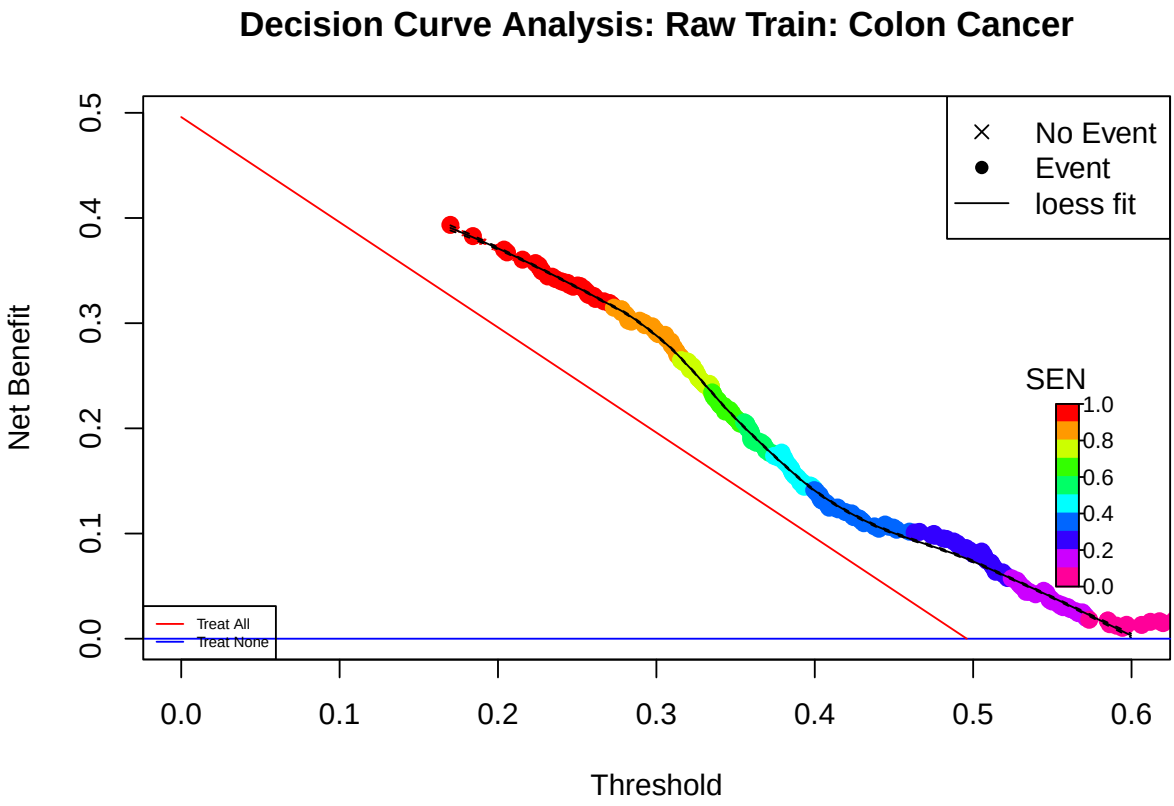
Here we evaluate the model using the RRPlot() function.

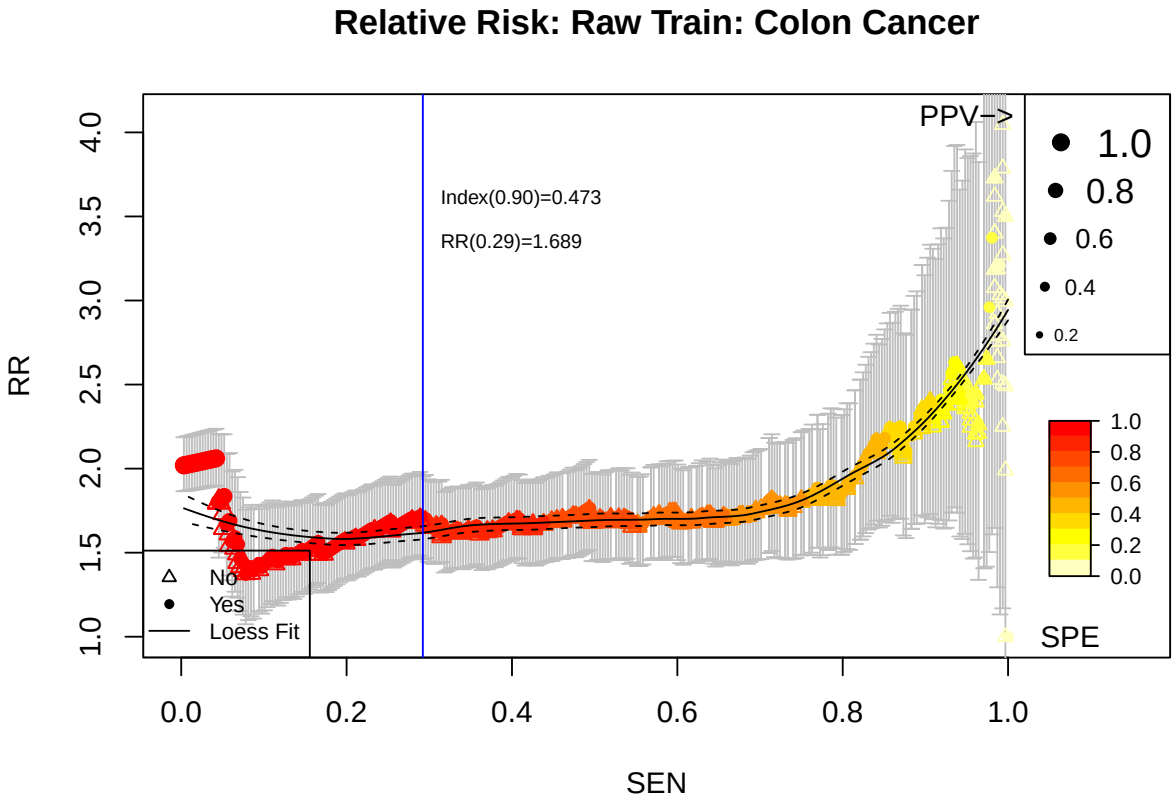
1.3.1 The evaluation of the raw Cox model with RRPlot()

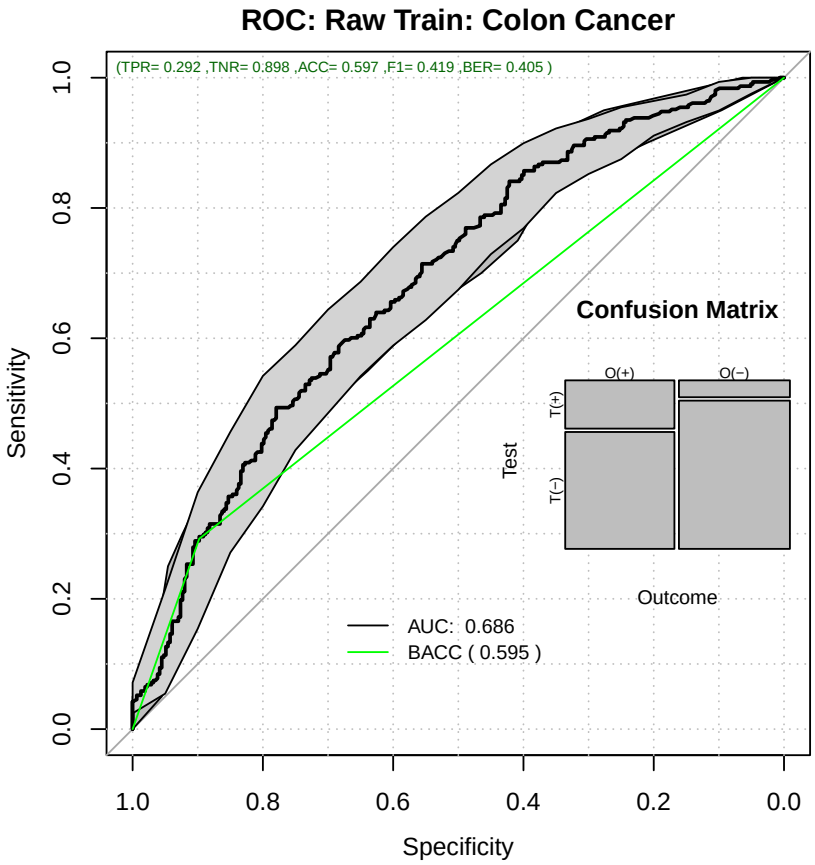
Here we will use the predicted event probability assuming a baseline hazard for events withing 5 years

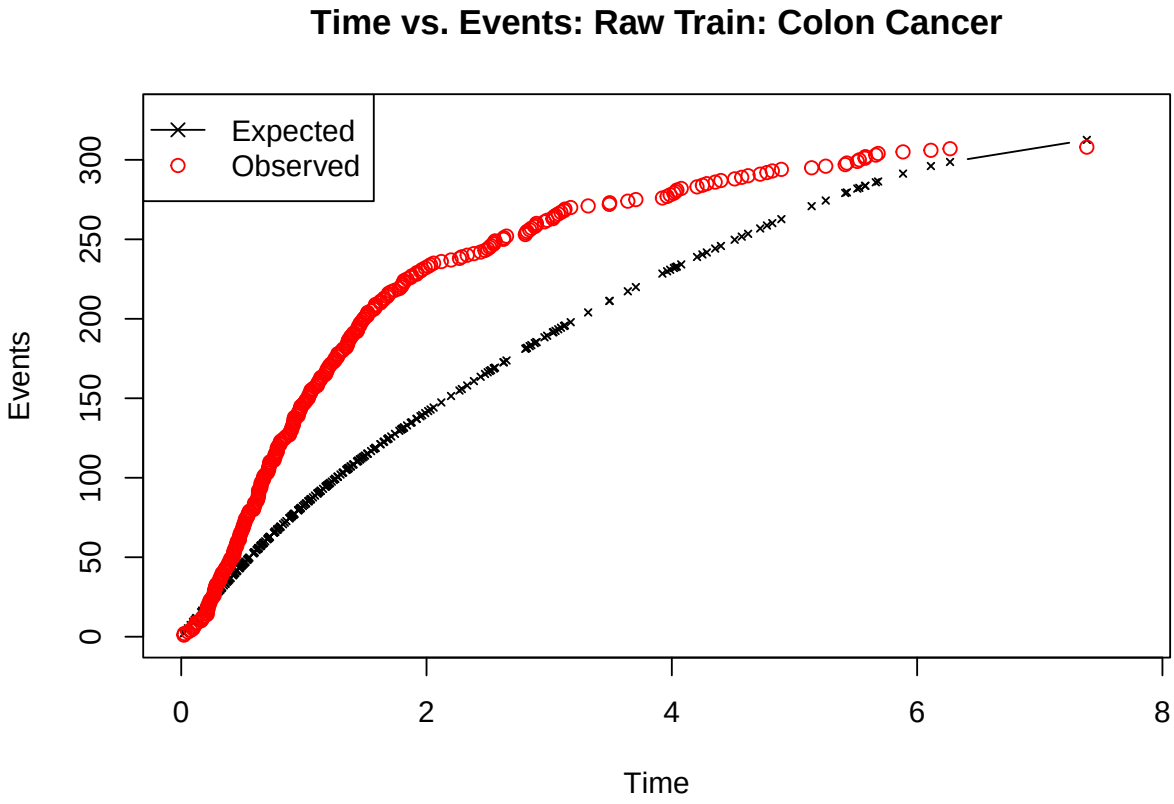
```
index <- predict(ml,dataColonTrain)
timeinterval <- 2*mean(subset(dataColonTrain,status==1)$time)

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)
```

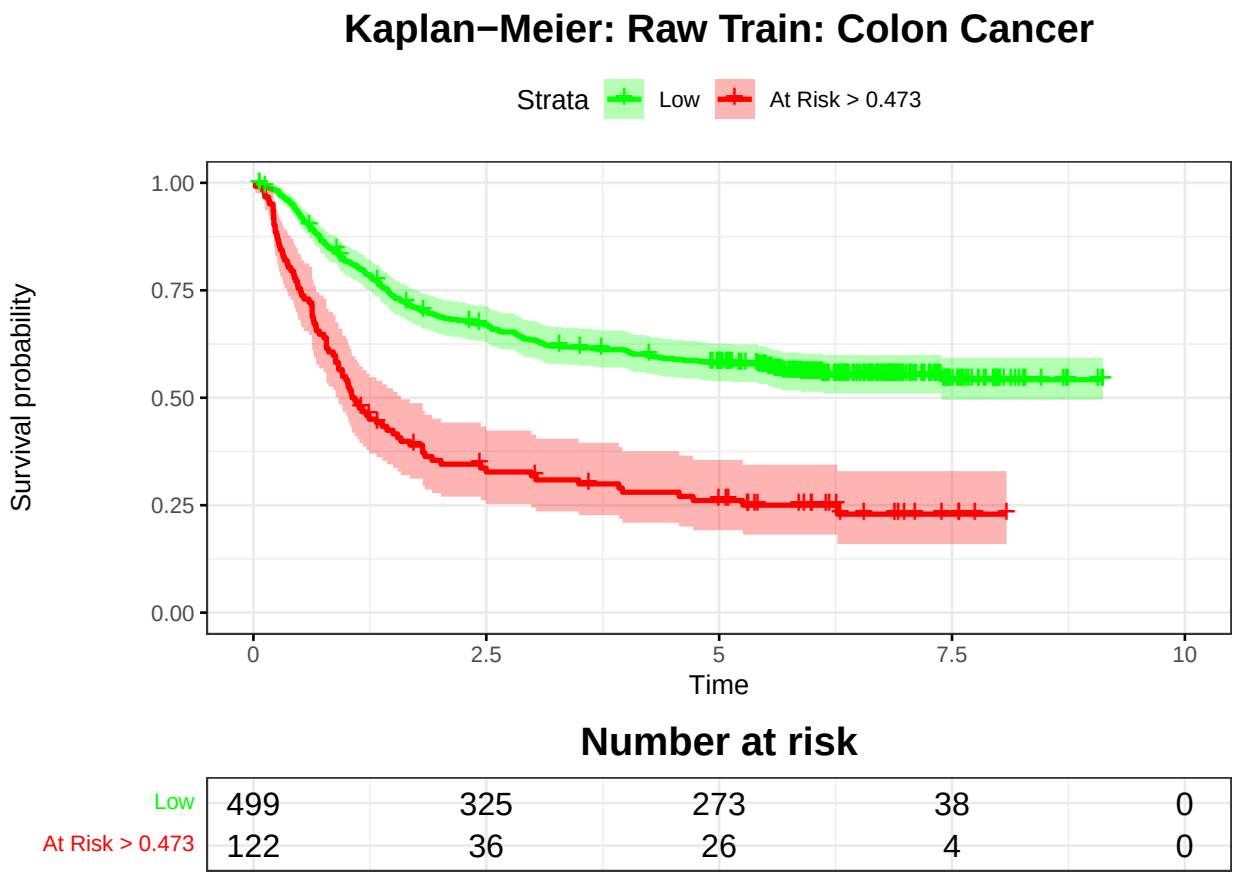



Decision Curve Analysis: Raw Train: Colon Cancer
Net Benefit
Threshold
0.0
0.1
0.2
0.3
0.4
0.5
0.6
No Event
Event
loess fit
SEN
1.0
0.8
0.6
0.4
0.2
0.0
Treat All
Treat None


Relative Risk: Raw Train: Colon Cancer
PPV–>
Index(0.90)=0.473
RR(0.29)=1.689
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: Raw Train: Colon Cancer
(TPR= 0.292 ,TNR= 0.898 ,ACC= 0.597 ,F1= 0.419 ,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: 0.686
BACC (0.595)


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



1.3.2 Uncalibrated Performance Report

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

Table 6: O/E Ratio

est	lower	upper
0.985	0.878	1.1

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

Table 7: O/E Ratio

mean	50%	2.5%	97.5%
1.52	1.52	1.49	1.55

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

Table 8: O/Acum Ratio

mean	50%	2.5%	97.5%
1.32	1.32	1.32	1.33

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

mean.C Index	median	lower	upper
0.652	0.652	0.622	0.684

```
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
0.292	0.242	0.346

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

Table 12: Specificity

est	lower	upper
0.898	0.859	0.929

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

Table 13: Probability Thresholds

90%
0.473

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

Table 14: Risk Ratio

est	lower	upper
1.69	1.46	1.95

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

Table 15: Logrank test Chisq = 64.689536 on 1 degrees of freedom, p = 0.000000

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	499	218	266.1	8.71	64.7
class=1	122	90	41.9	55.35	64.7

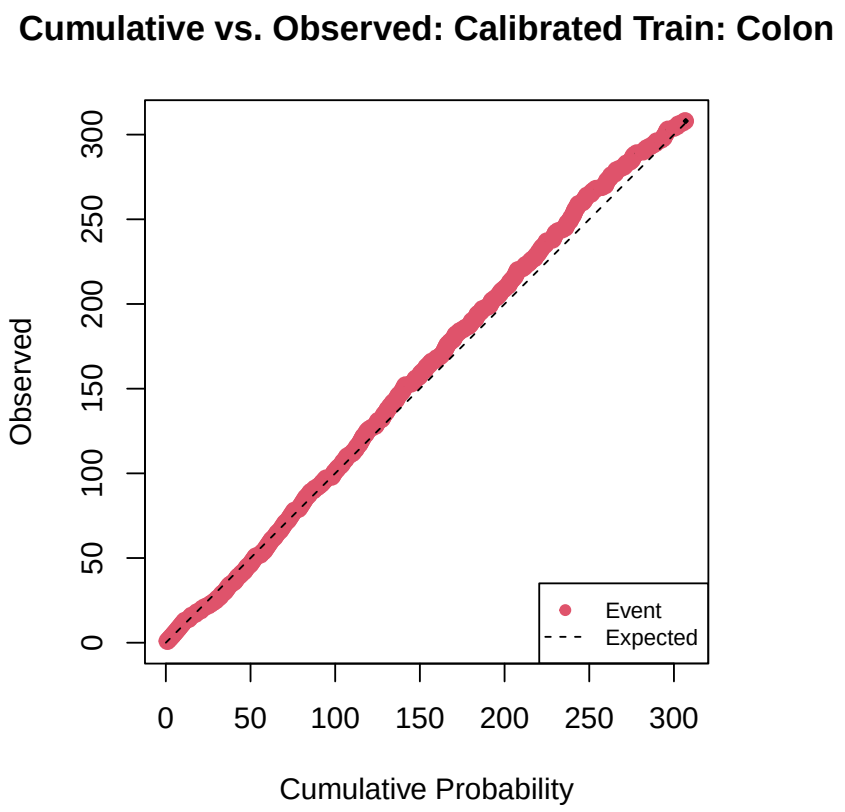
1.3.3 Cox Calibration

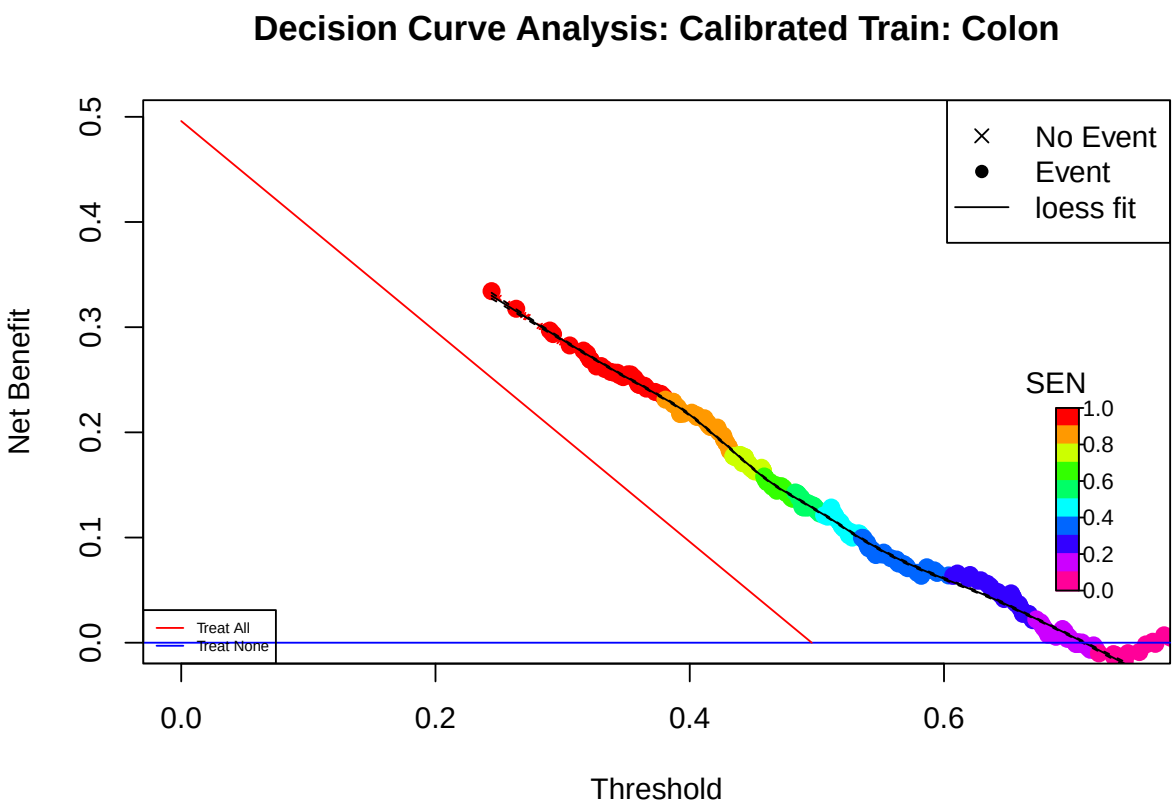
```
op <- par(no.readonly = TRUE)


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

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

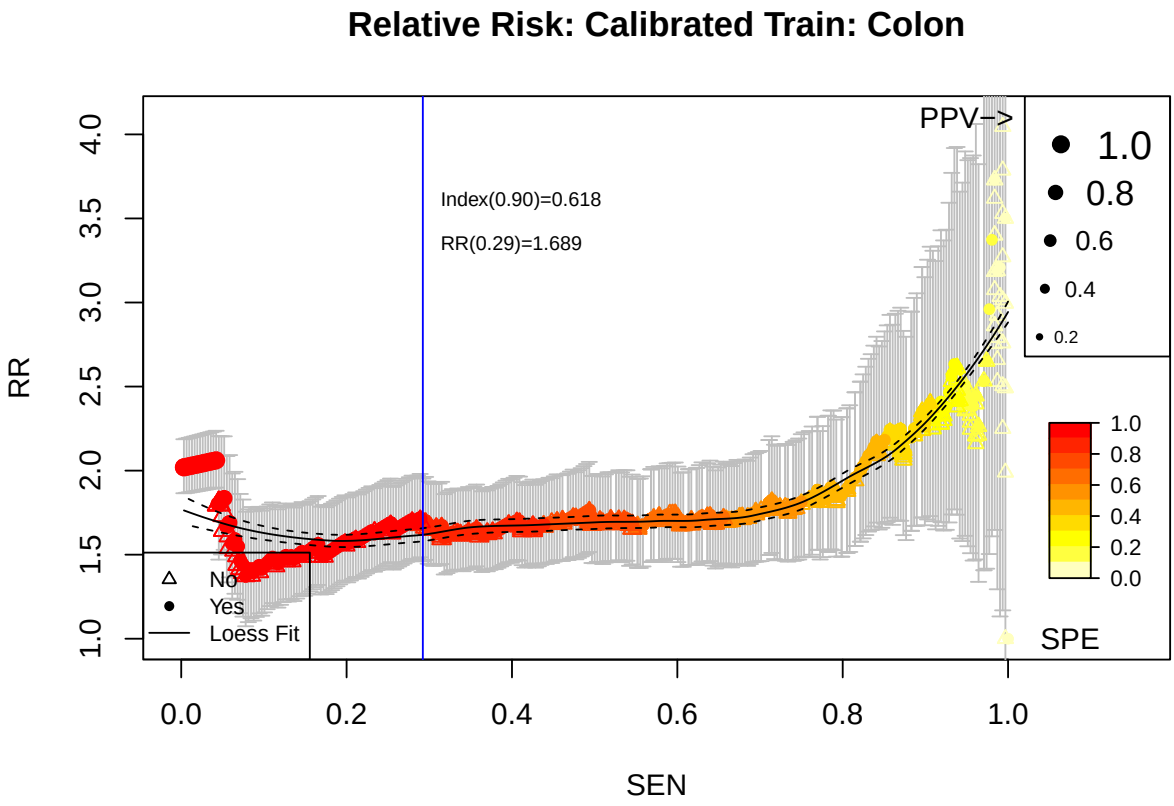
h0	Gain	DeltaTime
0.669	1.5	2.96

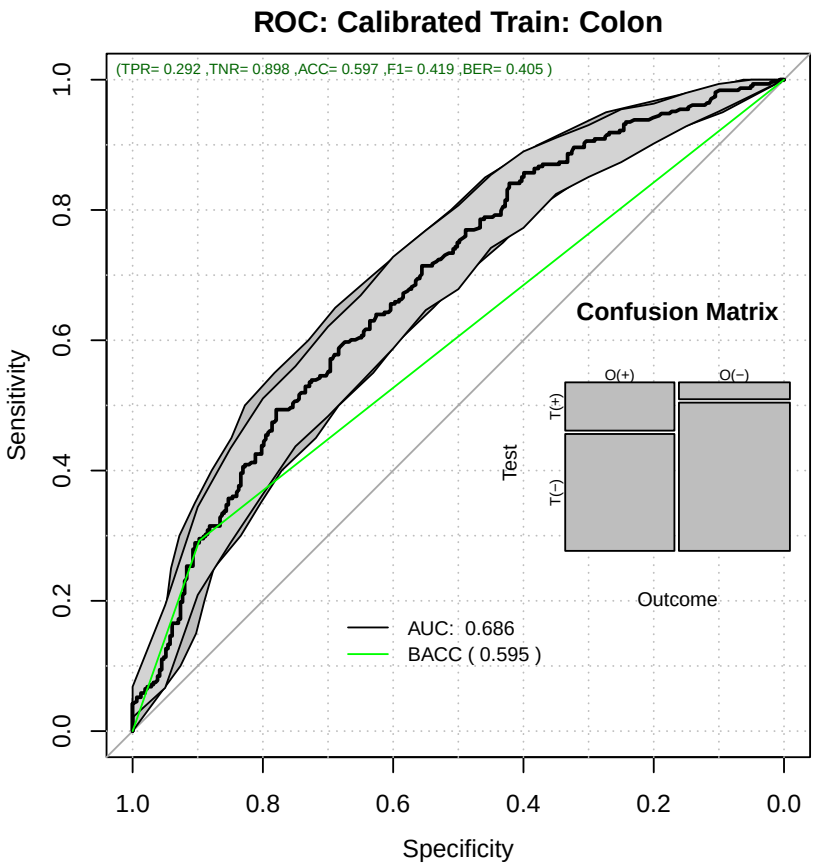
1.3.4 The RRplot() of the calibrated model

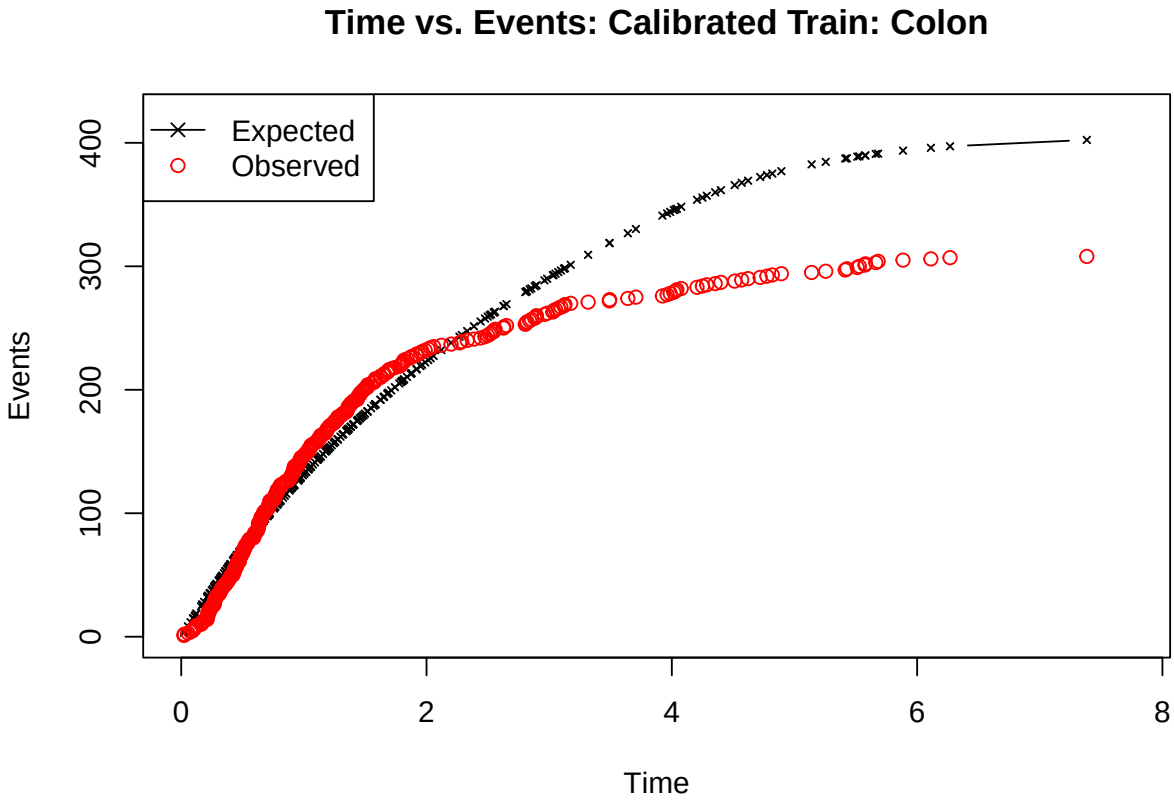
```
h0 <- calprob$h0
timeinterval <- calprob$timeInterval;

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)
```

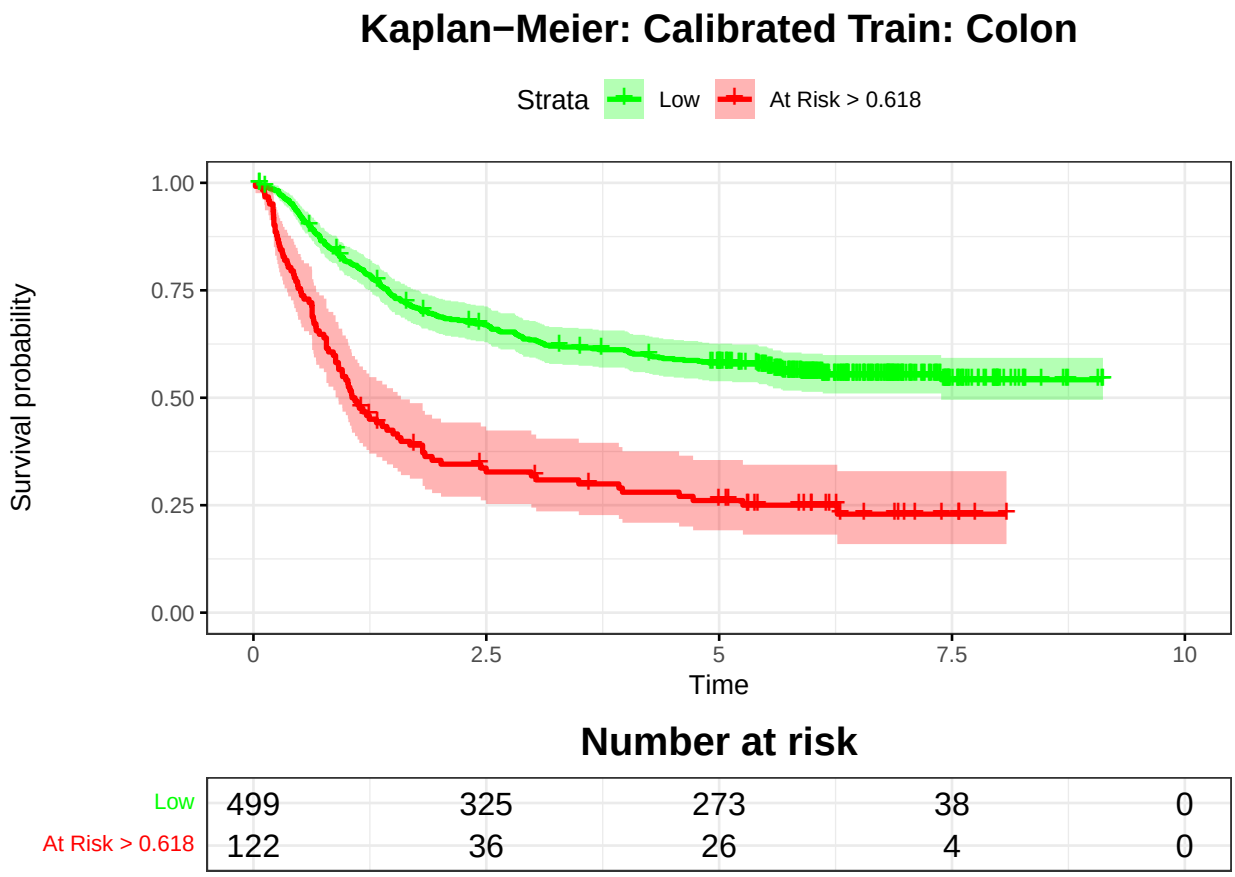

Cumulative vs. Observed: Calibrated Train: Colon
Observed
0
50
100
150
200
250
300
Event
Expected
Cumulative Probability


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
PPV–>
Index(0.90)=0.618
RR(0.29)=1.689
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
1.0
1.5
2.0
2.5
3.0
3.5
4.0
0.0
0.2
0.4
0.6
0.8
1.0


ROC: Calibrated Train: Colon
(TPR= 0.292 ,TNR= 0.898 ,ACC= 0.597 ,F1= 0.419 ,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: 0.686
BACC (0.595)


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



1.3.5 Calibrated Train Performance

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

Table 17: O/E Ratio

est	lower	upper
0.765	0.682	0.856

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

Table 18: O/E Ratio

mean	50%	2.5%	97.5%
0.969	0.969	0.951	0.985

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

Table 19: O/Acum Ratio

mean	50%	2.5%	97.5%
1.02	1.02	1.02	1.03

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

mean.C Index	median	lower	upper
0.652	0.652	0.624	0.68

```
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
0.292	0.242	0.346

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

Table 23: Specificity

est	lower	upper
0.898	0.859	0.929

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

Table 24: Probability Thresholds

90%
0.618

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

Table 25: Risk Ratio

est	lower	upper
1.69	1.46	1.95

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

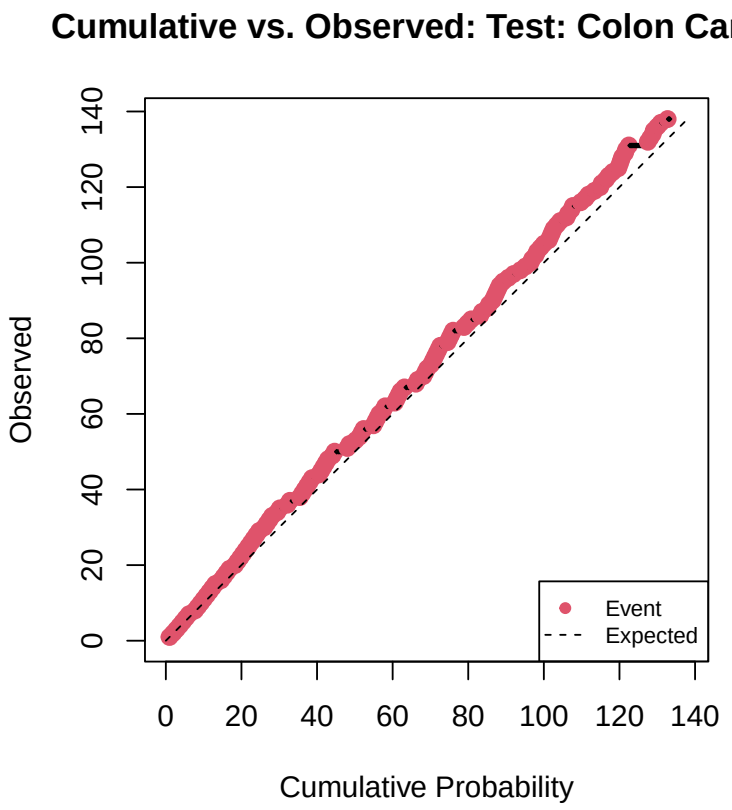
Table 26: Logrank test Chisq = 64.689536 on 1 degrees of freedom, p = 0.000000

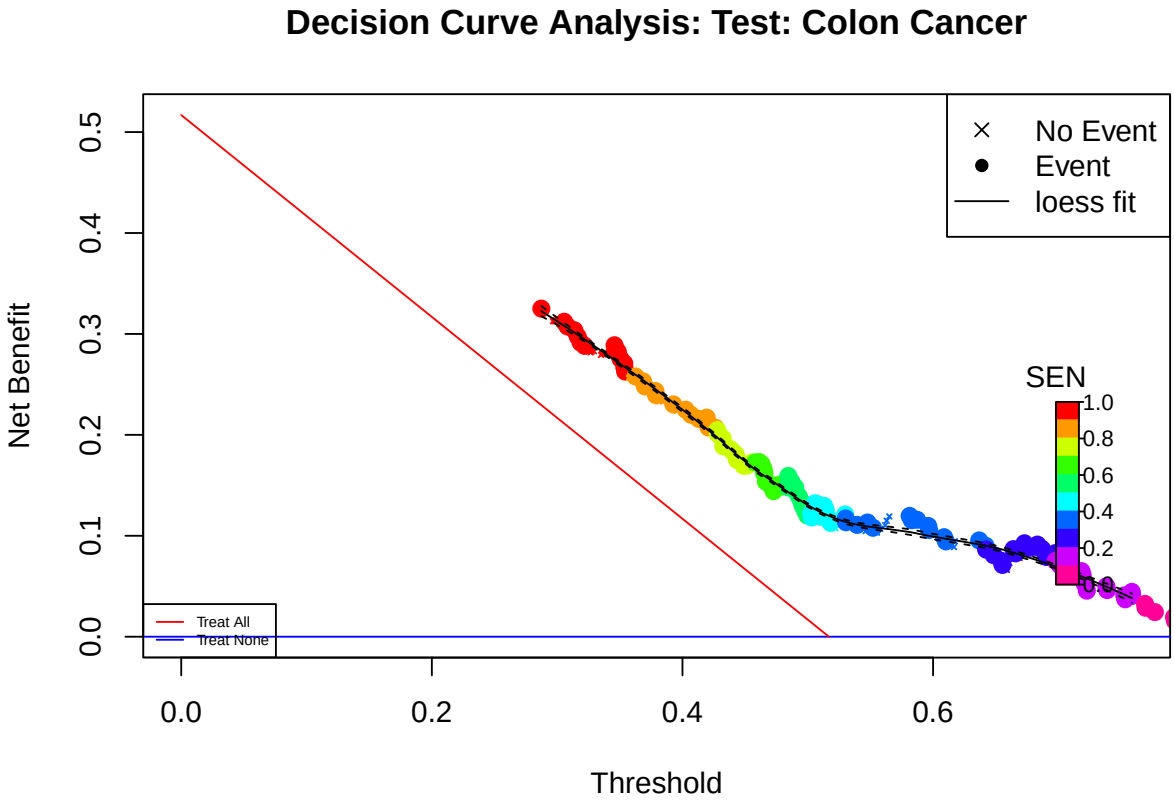
	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	499	218	266.1	8.71	64.7
class=1	122	90	41.9	55.35	64.7

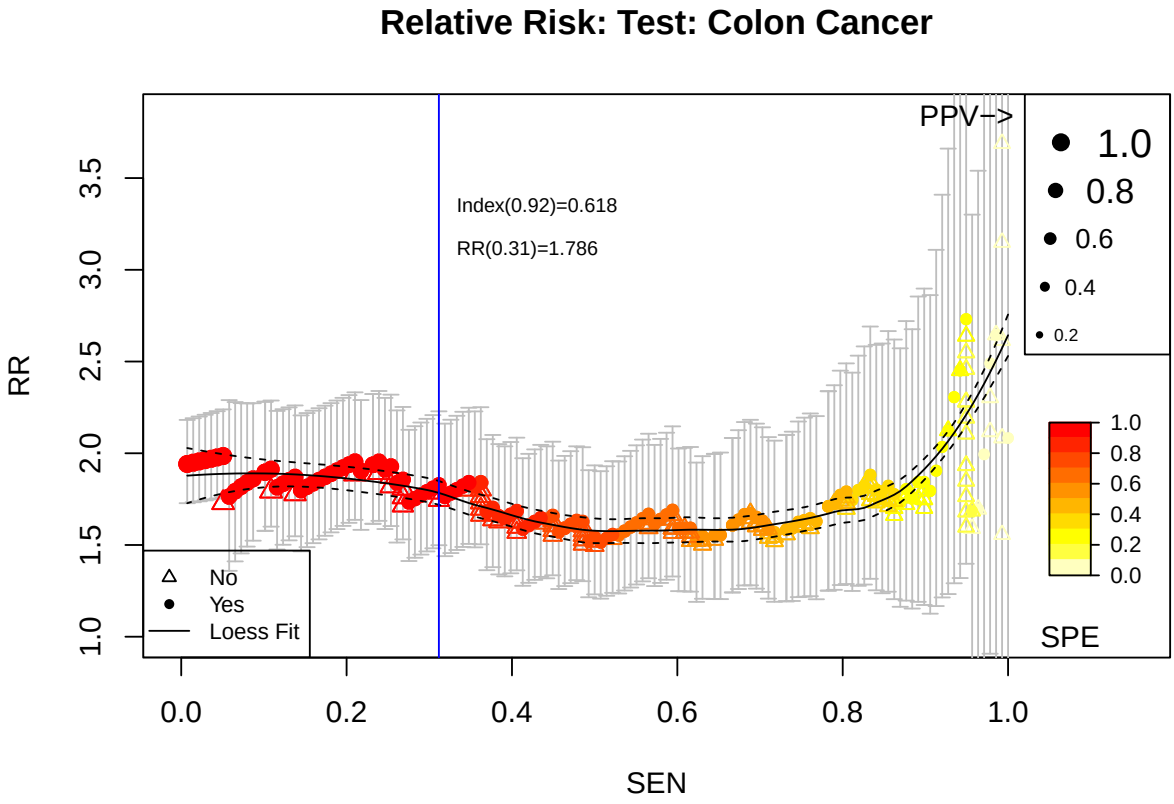
1.3.6 Evaluating on the test set

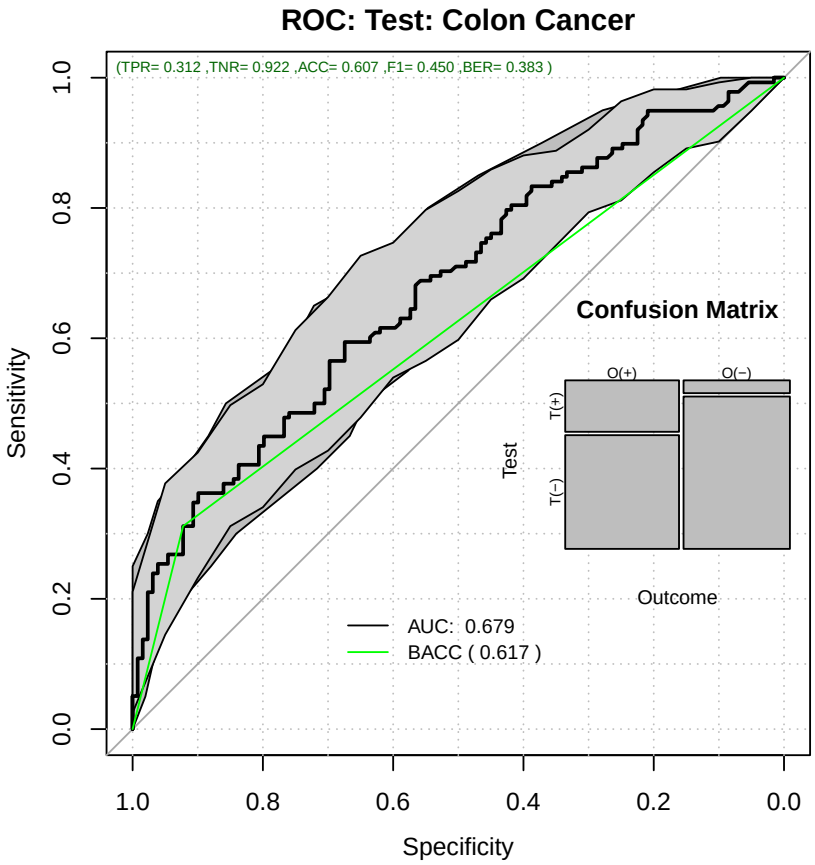
The calibrated h0 and timeinterval were estimated on the training set

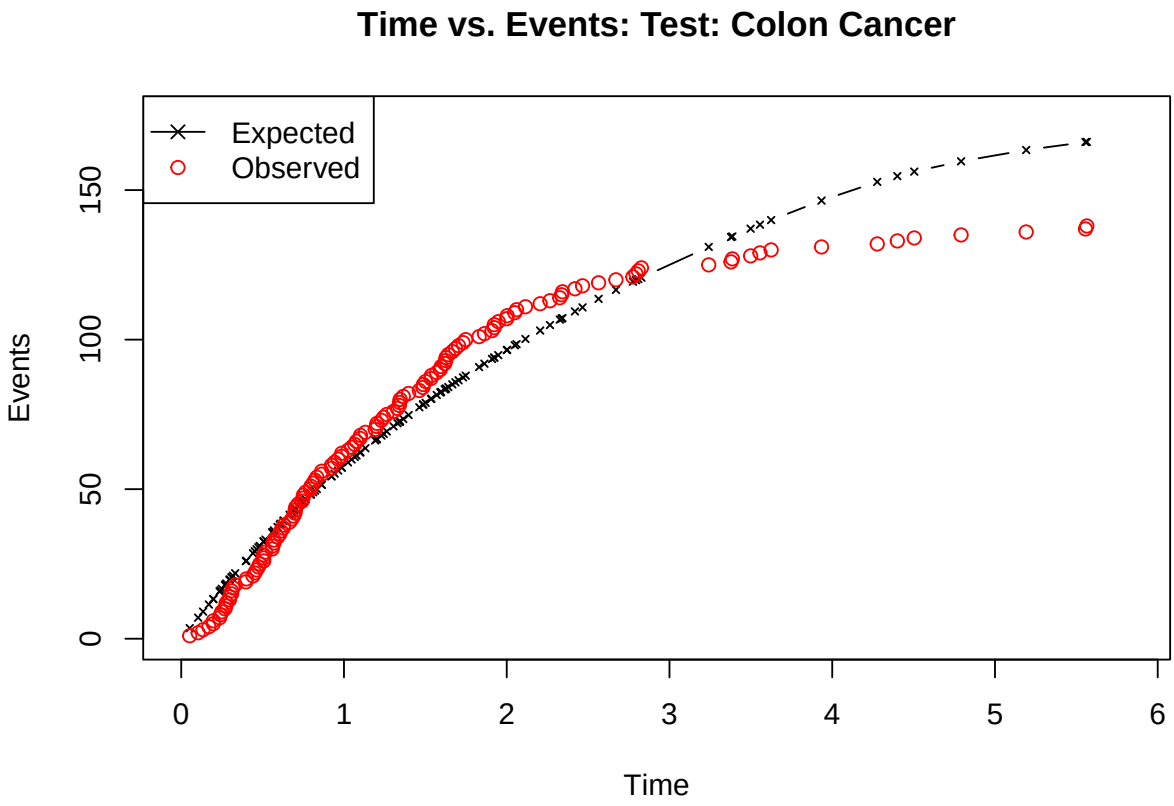
```
index <- predict(ml,dataColonTest)
rdata <- cbind(dataColonTest$status,ppoisGzero(index,h0))

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

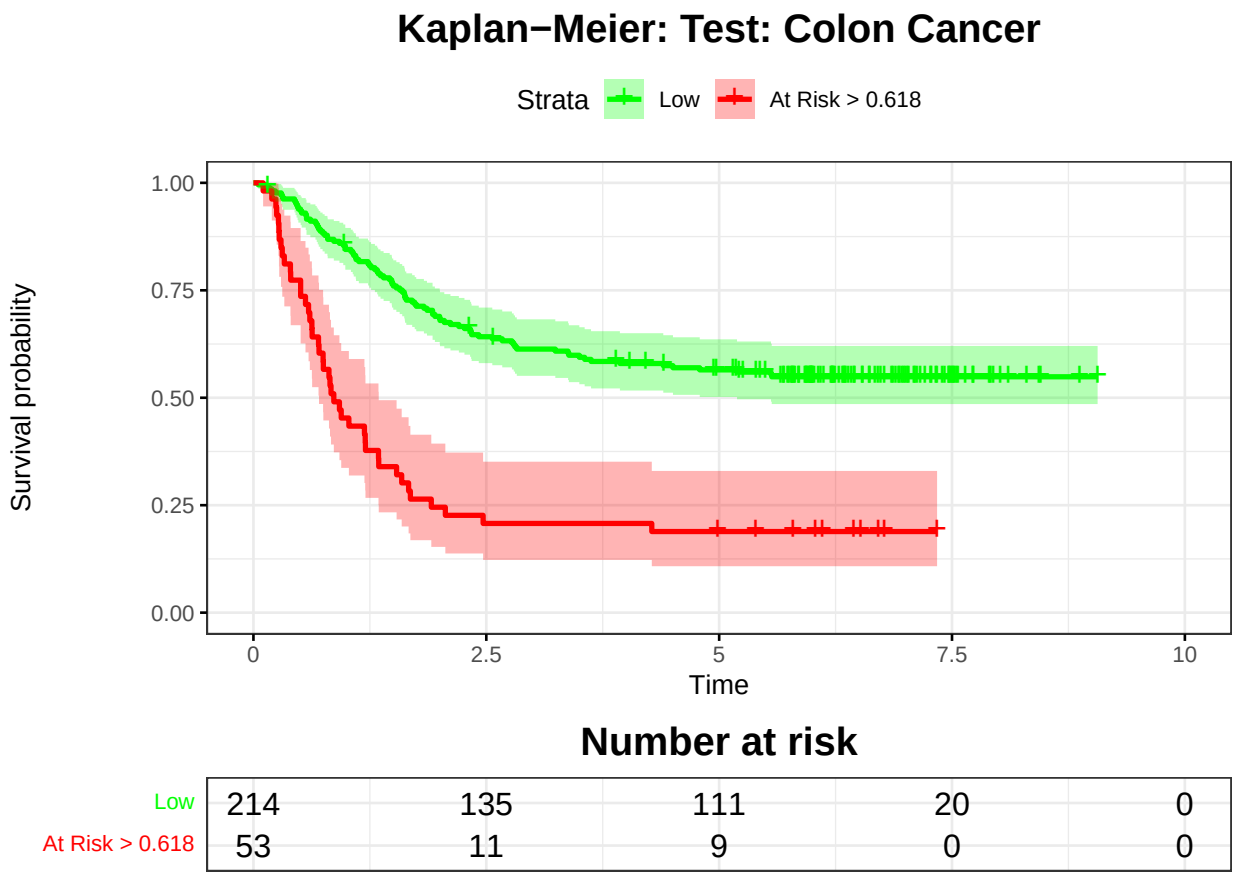

Cumulative vs. Observed: Test: Colon Cancer
Observed
0
20
40
60
80
100
120
140
Cumulative Probability
Event
Expected


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
0.0
0.2
0.4
0.6
Threshold


Relative Risk: Test: Colon Cancer
PPV–>
1.0
0.8
0.6
0.4
0.2
Index(0.92)=0.618
RR(0.31)=1.786
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: Test: Colon Cancer
(TPR= 0.312 ,TNR= 0.922 ,ACC= 0.607 ,F1= 0.450 ,BER= 0.383)
Sensitivity
Specificity
1.0
0.8
0.6
0.4
0.2
0.0
Confusion Matrix
O(+)
O(−)
T(+)
T(−)
Test
Outcome
AUC: 0.679
BACC (0.617)


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



1.3.7 Test Performance

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

Table 27: O/E Ratio

est	lower	upper
0.831	0.698	0.981

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

Table 28: O/E Ratio

mean	50%	2.5%	97.5%
0.98	0.98	0.954	1.01

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

Table 29: O/Acum Ratio

mean	50%	2.5%	97.5%
1.06	1.06	1.06	1.07

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

mean.C Index	median	lower	upper
0.656	0.656	0.612	0.699

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

Table 31: ROC AUC

est	lower	upper
0.679	0.616	0.742

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

Table 32: Sensitivity

est	lower	upper
0.312	0.236	0.396

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

Table 33: Specificity

est	lower	upper
0.922	0.862	0.962

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

Table 34: Probability Thresholds

90%
0.618

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

Table 35: Risk Ratio

est	lower	upper
1.79	1.46	2.18

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

Table 36: Logrank test Chisq = 46.790198 on 1 degrees of freedom, p = 0.000000

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	214	95	121.1	5.63	46.8
class=1	53	43	16.9	40.44	46.8

1.4 Cross-Validation

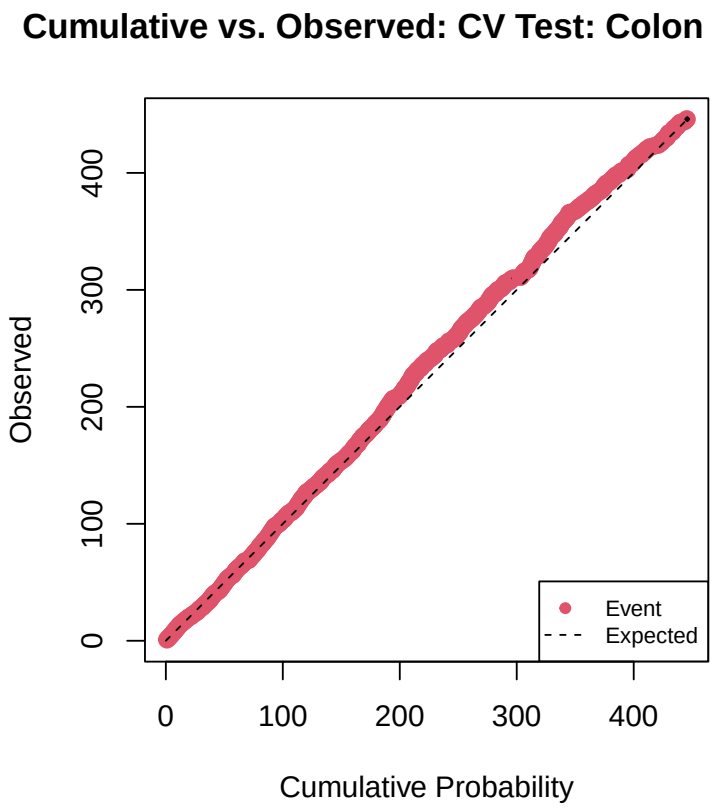
Here we will cross validate the training set and evaluate also on the testing set. The h0 and the timeinterval are the ones estimated on the calibration process

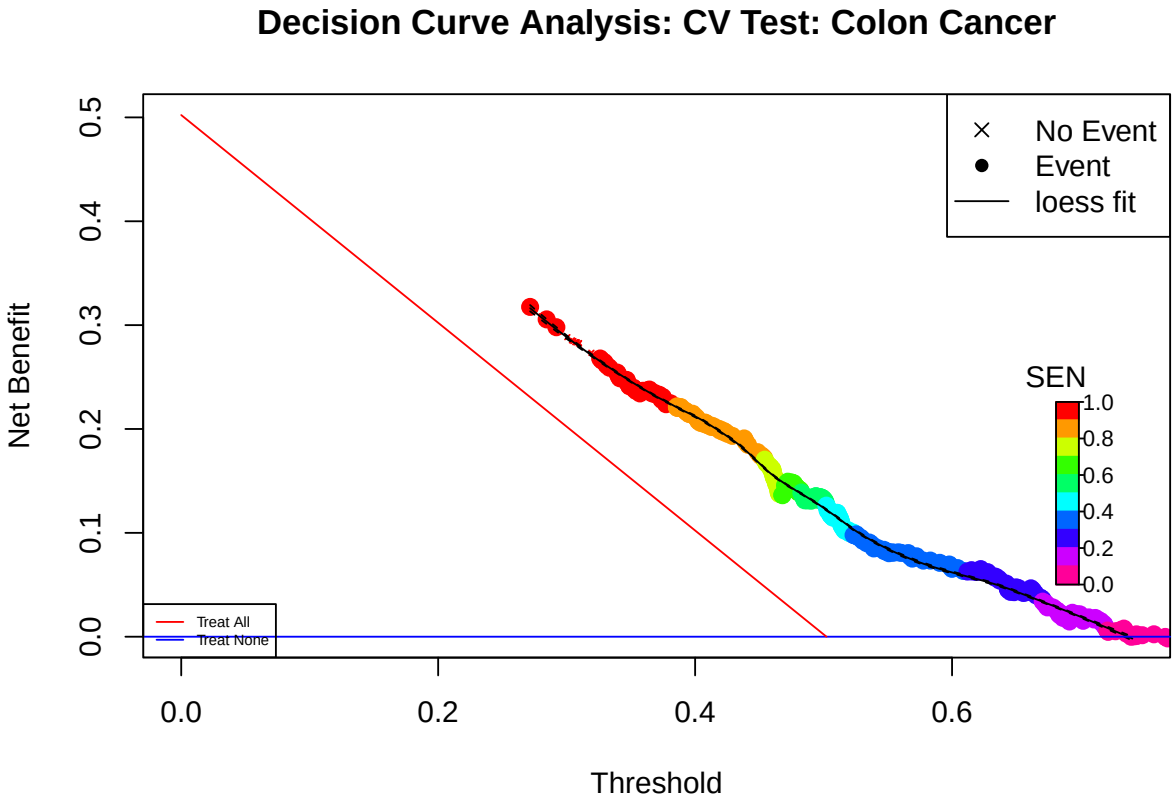
```
rcv <- randomCV(theData=dataColonTrain,
                theOutcome = Surv(time,status)~1,
                fittingFunction=BSWiMS.model,
                trainFraction = 0.75,
                repetitions=50,
                classSamplingType = "Pro",
                testingSet=dataColonTest
      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

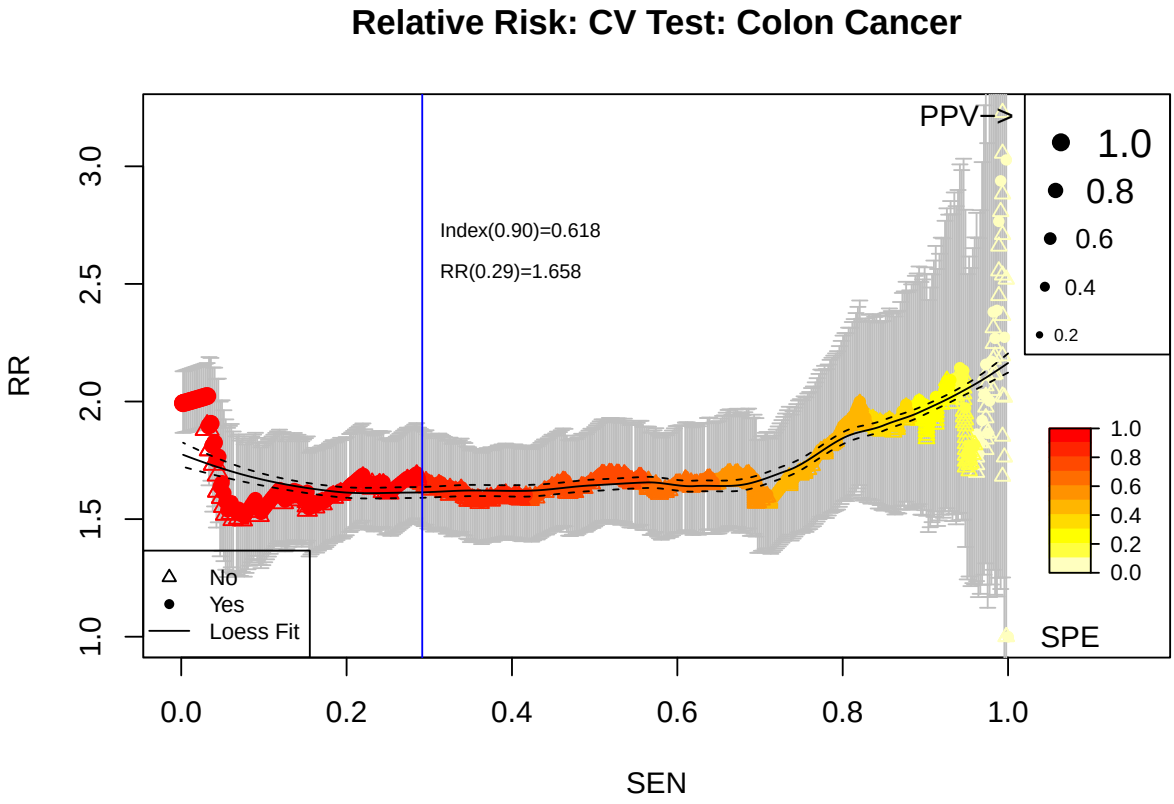
.[++-].[++++++].[++++].[++++].[++++].[++++].[++++++].[++++-].[+++++].[+++++]10 Tested: 847 Avg. Selected: 8.4 Min Tests: 1 Max Tests: 10 Mean Tests: 4.994097 . MAD: 0.4731862 .[+-++].[++++].[++++-].[++++].[+++].[+++++].[++++].[++].[+++++].[+++]20 Tested: 882 Avg. Selected: 8 Min Tests: 1 Max Tests: 20 Mean Tests: 9.591837 . MAD: 0.4731669 .[++++-].[++++].[++++].[++++].[++++-].[+++-].[++++].[+++].[+++++].[++++-]30 Tested: 888 Avg. Selected: 8.033333 Min Tests: 1 Max Tests: 30 Mean Tests: 14.29054 . MAD: 0.4735136 .[++++-].[+++-].[+–+].[++++].[+++++].[+++++].[+++++].[++-].[++++].[++++]40 Tested: 888 Avg. Selected: 8.05 Min Tests: 3 Max Tests: 40 Mean Tests: 19.05405 . MAD: 0.4735561 .[++].[+++++-].[++++].[+++-].[++++].[++++–].[++++].[+++].[++++++].[+++++]50 Tested: 888 Avg. Selected: 8.14 Min Tests: 4 Max Tests: 50 Mean Tests: 23.81757 . MAD: 0.4731824

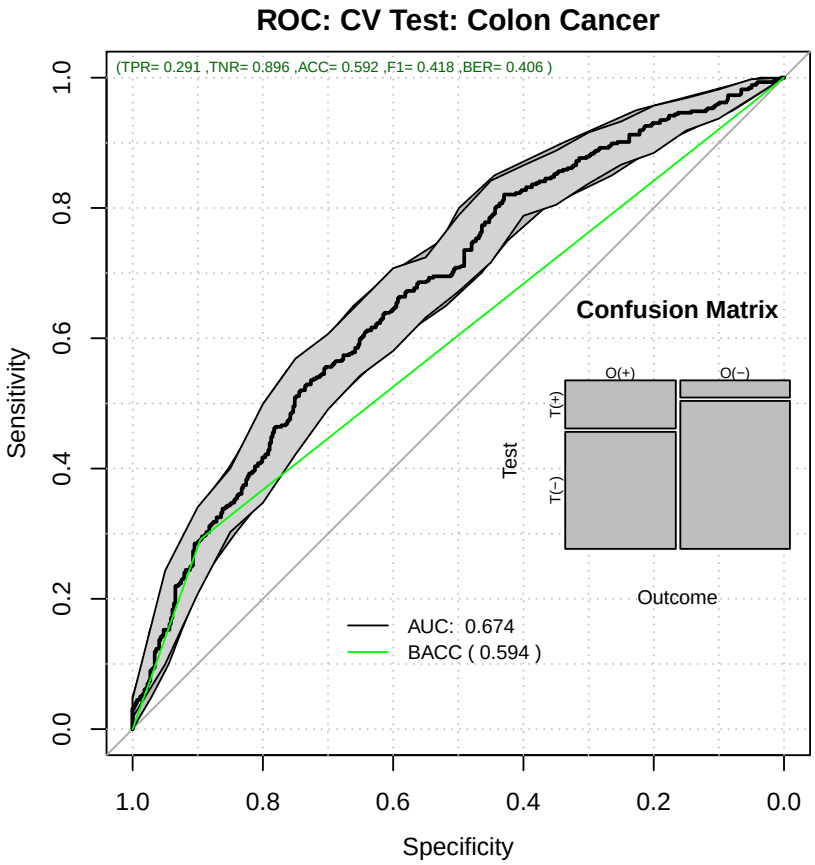
```
stp <- rcv$survTestPredictions
stp <- stp[!is.na(stp[,4]),]

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,
```

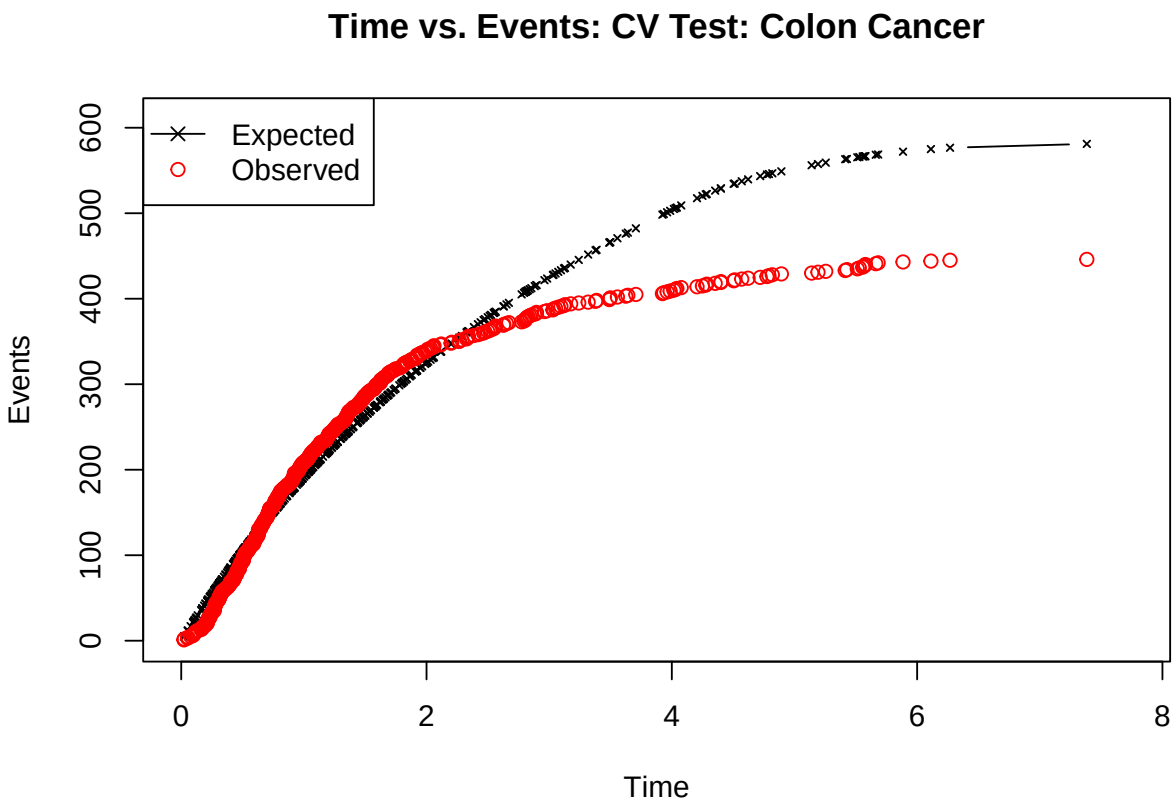
```
                    title="CV Test: Colon Cancer",
                    ysurvlim=c(0.00,1.0),
                    riskTimeInterval=timeinterval)
```

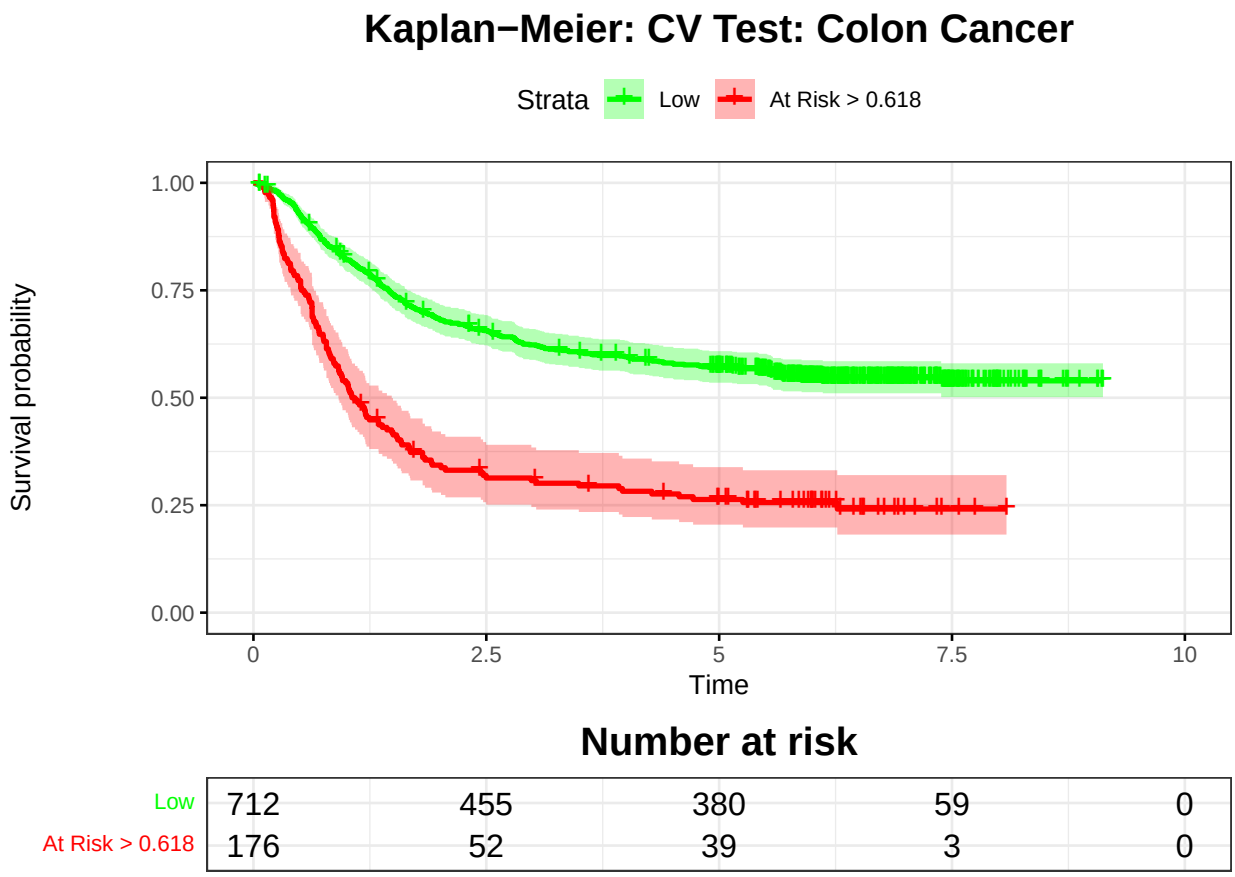



Decision Curve Analysis: CV Test: Colon Cancer
No Event
Event
loess fit
SEN
1.0
0.8
0.6
0.4
0.2
0.0
Treat All
Treat None
Net Benefit
0.0
0.1
0.2
0.3
0.4
0.5
0.0
0.2
0.4
0.6
Threshold


Relative Risk: CV Test: Colon Cancer
PPV–>
1.0
0.8
0.6
0.4
0.2
Index(0.90)=0.618
RR(0.29)=1.658
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
0.0
0.2
0.4
0.6
0.8
1.0
SEN


ROC: CV Test: Colon Cancer
(TPR= 0.291 ,TNR= 0.896 ,ACC= 0.592 ,F1= 0.418 ,BER= 0.406)
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: 0.674
BACC (0.594)


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



1.4.1 CV Test Performance

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

Table 37: O/E Ratio

est	lower	upper
0.768	0.698	0.842

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

Table 38: O/E Ratio

mean	50%	2.5%	97.5%
0.957	0.957	0.943	0.97

```
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
0.648	0.648	0.622	0.674

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

Table 41: ROC AUC

est	lower	upper
0.674	0.639	0.709

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

Table 42: Sensitivity

est	lower	upper
0.291	0.25	0.336

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

Table 43: Specificity

est	lower	upper
0.896	0.864	0.923

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

Table 44: Probability Thresholds

90%
0.618

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

Table 45: Risk Ratio

est	lower	upper
1.66	1.47	1.87

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

Table 46: Logrank test Chisq = 88.579988 on 1 degrees of freedom, p = 0.000000

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	712	316	384.3	12.1	88.6
class=1	176	130	61.7	75.6	88.6