

Colon Cancer

Jose Tamez

2023-04-28

Contents

0.1	Modeling	2
0.2	Cox Model Performance	3
0.3	Cross-Validation	27

```
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~.*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
322	299

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

0	1
120	147

0.1 Modeling

```
m1 <- BSWiMS.model(Surv(time,status)~1,data=dataColonTrain,NumberOfRepeats = 10)
```

```
[+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++]....
```

```
sm <- summary(m1)
pander::pander(sm$coefficients)
```

Table 3: Table continues below

	Estimate	lower	HR	upper	u.Accuracy
nodes	0.024655	1.016	1.025	1.034	0.614
node4	0.211991	1.154	1.236	1.325	0.618
age_nodes	0.000394	1.000	1.000	1.001	0.610
age_node4	0.002920	1.002	1.003	1.004	0.618
rxLev_5FU_age	-0.001931	0.997	0.998	0.999	0.554

	Estimate	lower	HR	upper	u.Accuracy
rxLev_5FU	-0.093619	0.860	0.911	0.964	0.554
age	-0.002773	0.995	0.997	0.999	0.530
extent	0.090985	1.024	1.095	1.172	0.536

Table 4: Table continues below

	r.Accuracy	full.Accuracy	u.AUC	r.AUC	full.AUC
nodes	0.496	0.614	0.608	0.512	0.609
node4	0.578	0.622	0.610	0.581	0.614
age_nodes	0.481	0.610	0.604	0.500	0.604
age_node4	0.577	0.619	0.610	0.580	0.615
rxLev_5FU_age	0.617	0.622	0.560	0.609	0.614
rxLev_5FU	0.612	0.619	0.560	0.605	0.615
age	0.616	0.619	0.527	0.607	0.615
extent	0.618	0.622	0.549	0.610	0.614

	IDI	NRI	z.IDI	z.NRI	Delta.AUC	Frequency
nodes	0.03209	0.393	6.10	5.24	0.09644	1
node4	0.04435	0.439	6.02	6.34	0.03213	1
age_nodes	0.03197	0.417	5.81	5.61	0.10428	1
age_node4	0.03972	0.427	5.76	6.13	0.03522	1
rxLev_5FU_age	0.01576	0.242	3.58	3.25	0.00430	1
rxLev_5FU	0.01306	0.241	3.26	3.23	0.01005	1
age	0.00882	0.180	2.91	2.26	0.00808	1
extent	0.00889	0.198	2.64	3.57	0.00382	1

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

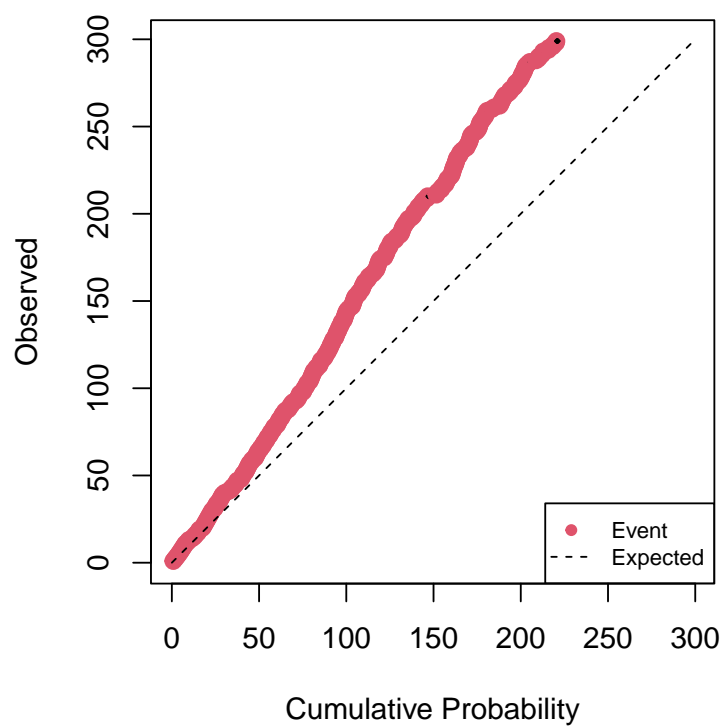
```
index <- predict(ml,dataColonTrain)
timeinterval <- 2*mean(subset(dataColonTrain,status==1)$time)

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

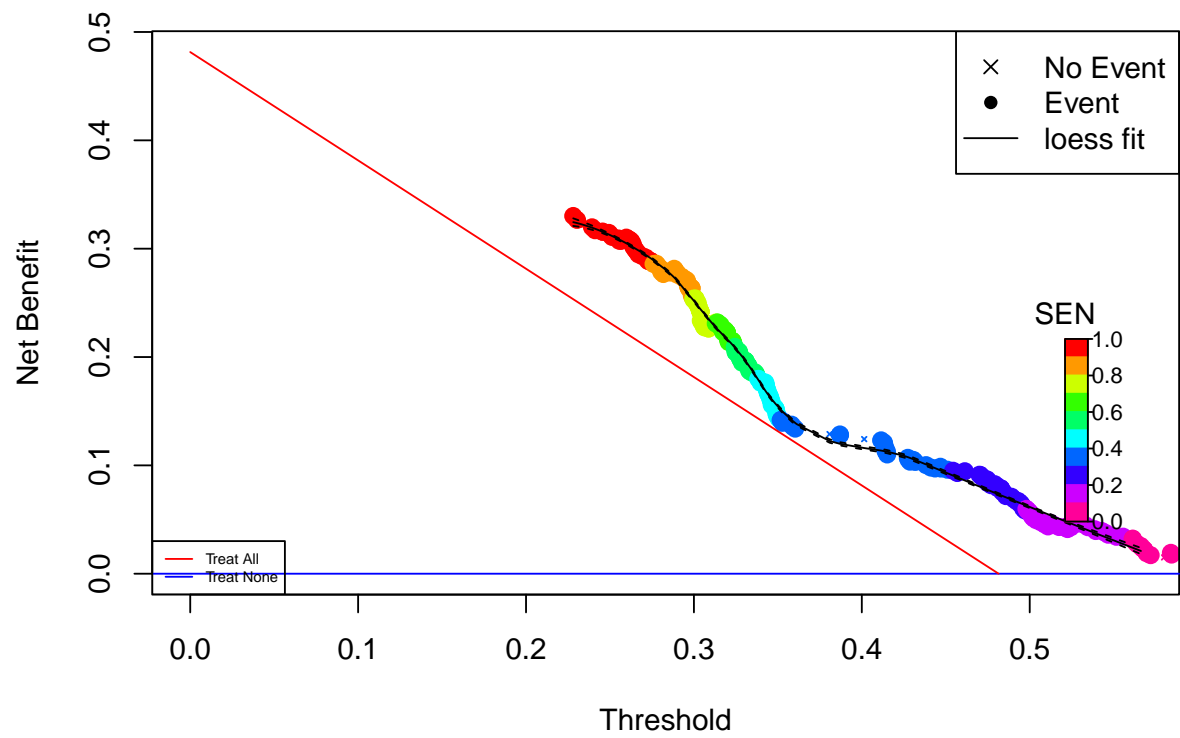
rdata <- cbind(dataColonTrain$status,ppoisGzero(index,h0))

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

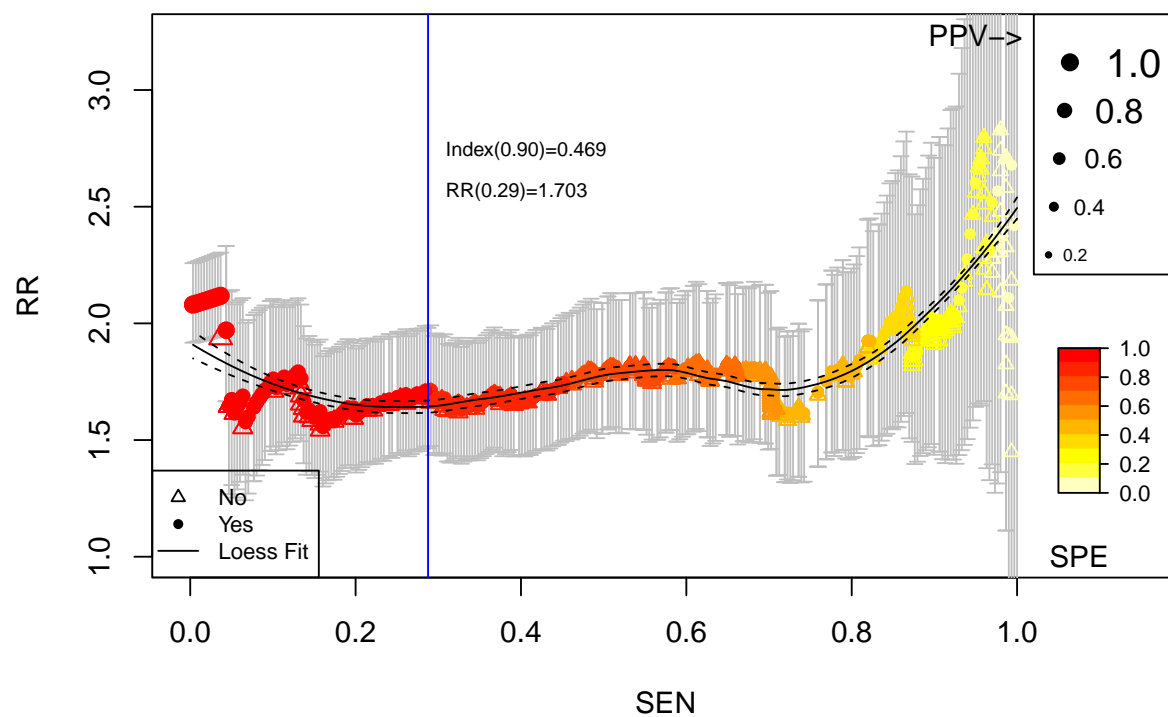
Cumulative vs. Observed: Raw Train: Colon Cancer

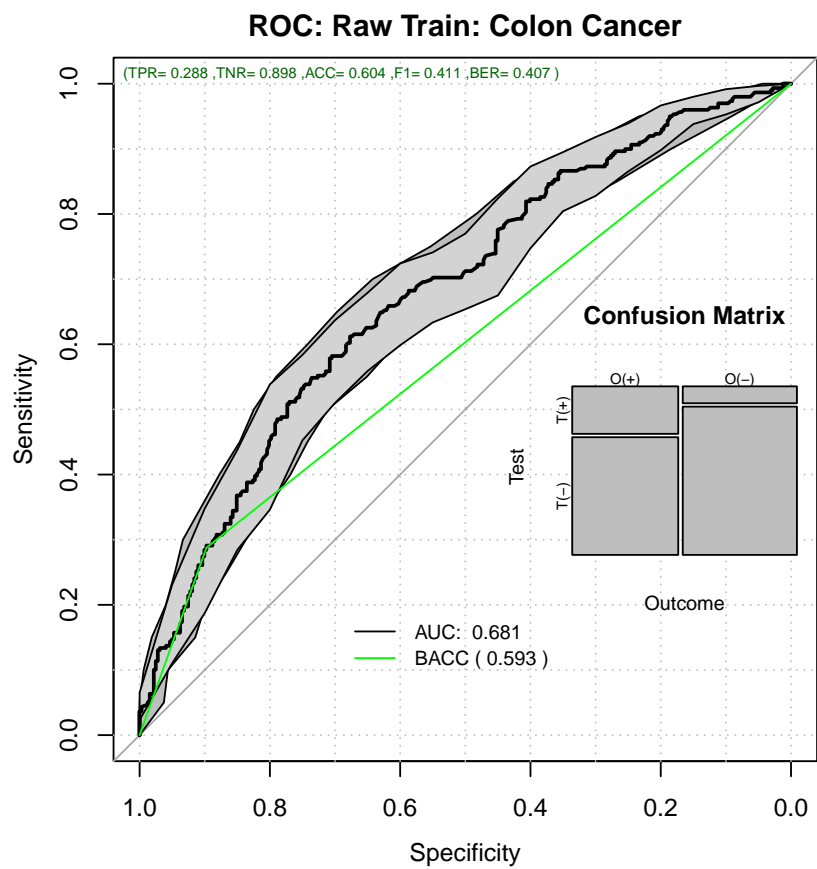


Decision Curve Analysis: Raw Train: Colon Cancer

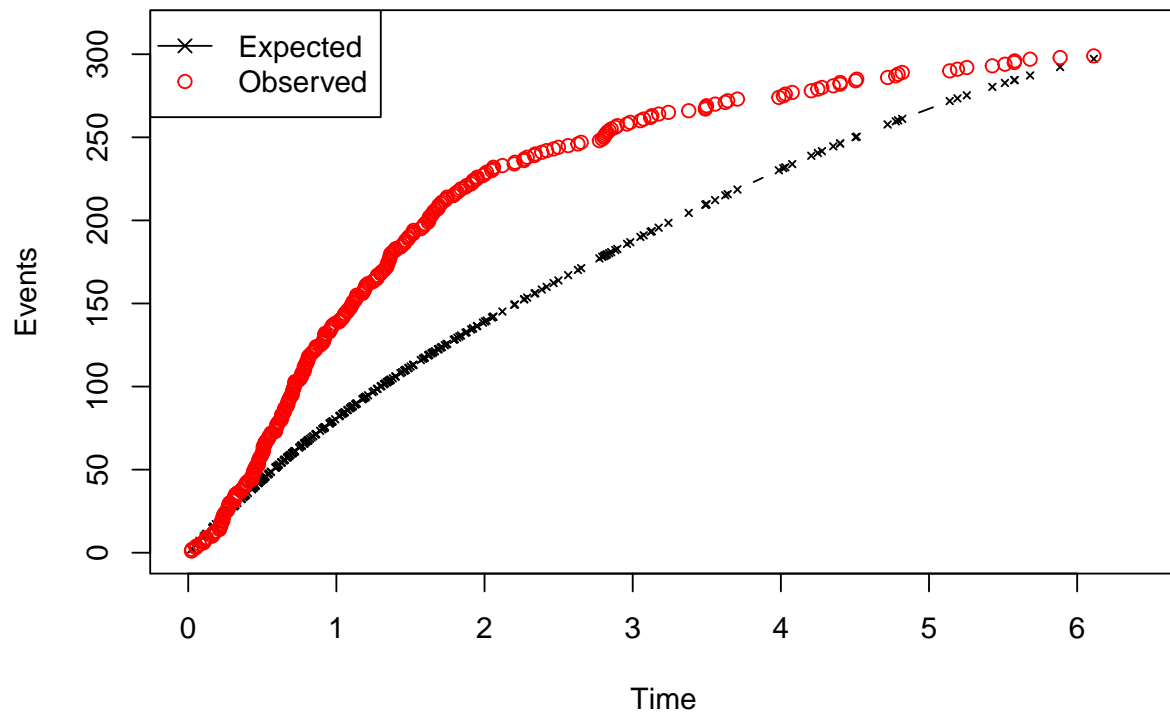


Relative Risk: Raw Train: Colon Cancer

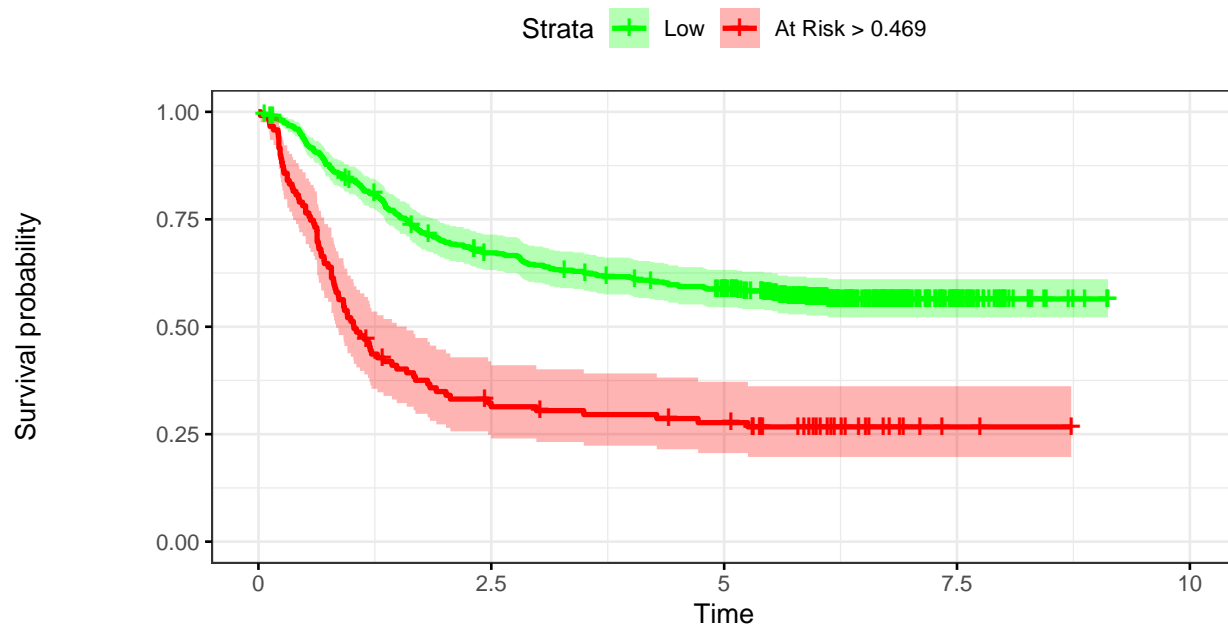




Time vs. Events: Raw Train: Colon Cancer



Kaplan–Meier: Raw Train: Colon Cancer



Number at risk

Low	502	328	276	43	0
At Risk > 0.469	119	35	29	2	0

0.2.2 Uncalibrated Performance Report

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

Table 6: O/E Ratio

est	lower	upper
1.01	0.895	1.13

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

Table 7: O/E Ratio

mean	50%	2.5%	97.5%
1.49	1.49	1.46	1.51

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

Table 8: O/Acum Ratio

mean	50%	2.5%	97.5%
1.38	1.38	1.37	1.38

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

mean.C Index	median	lower	upper
0.653	0.653	0.622	0.682

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

Table 10: ROC AUC

est	lower	upper
0.681	0.639	0.723

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

Table 11: Sensitivity

est	lower	upper
0.288	0.237	0.343

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

Table 12: Specificity

est	lower	upper
0.898	0.859	0.928

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

Table 13: Probability Thresholds

90%
0.469

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

Table 14: Risk Ratio

est	lower	upper
1.7	1.46	1.98

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

Table 15: Logrank test Chisq = 64.815888 on 1 degrees of freedom,
p = 0.000000

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	502	213	259.8	8.42	64.8
class=1	119	86	39.2	55.81	64.8

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
0.645	1.51	2.97

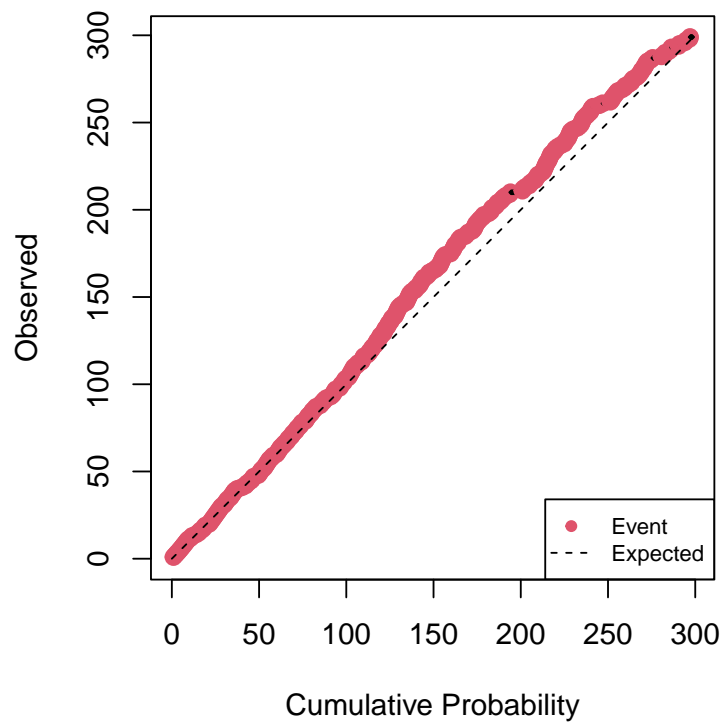
0.2.4 The RRplot() of the calibrated model

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

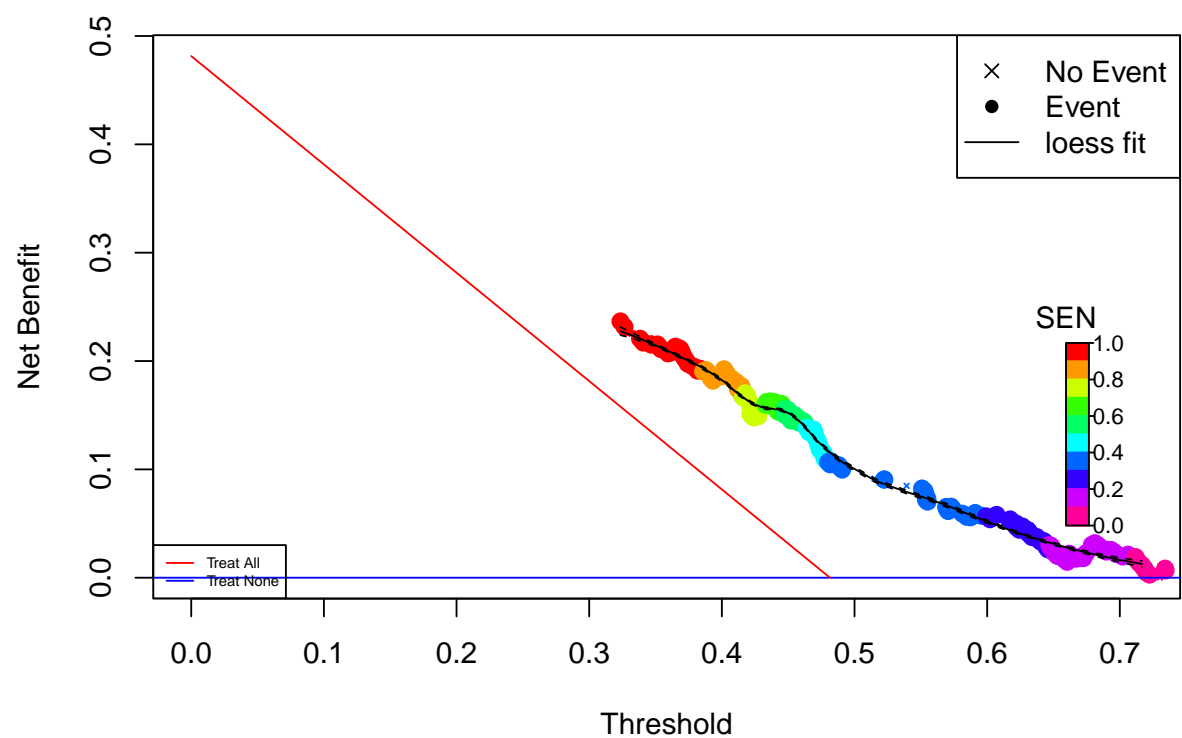
rdata <- cbind(dataColonTrain$status, calprob$prob)

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

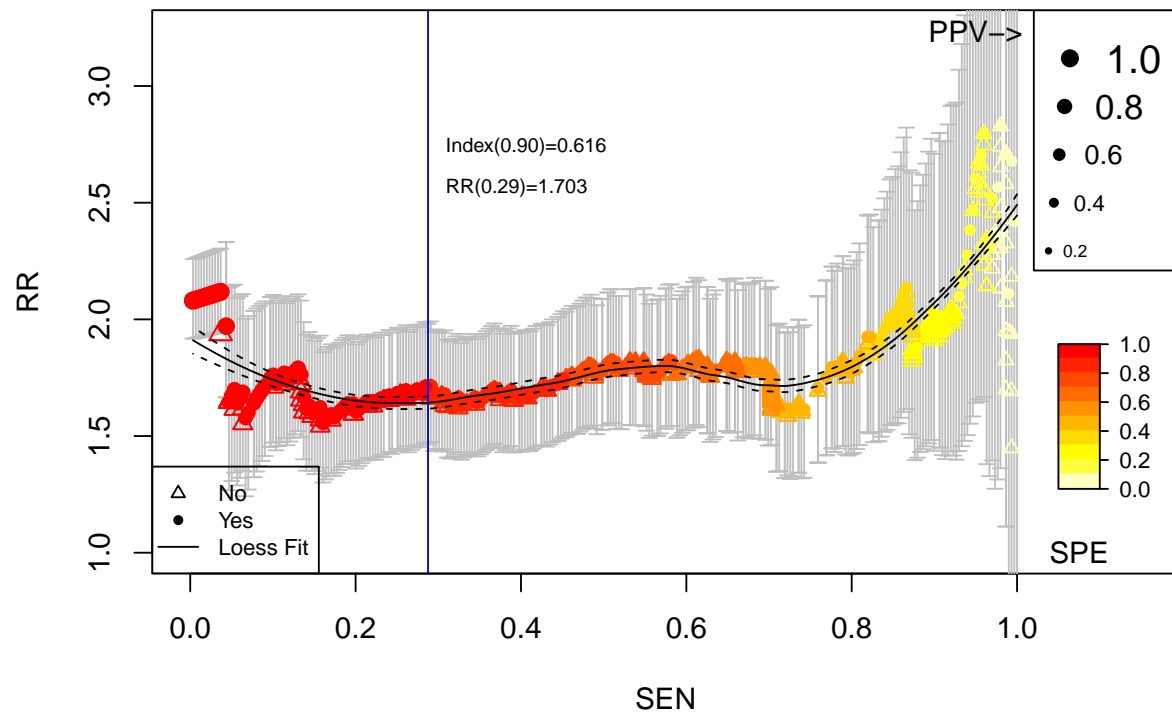
Cumulative vs. Observed: Calibrated Train: Colon

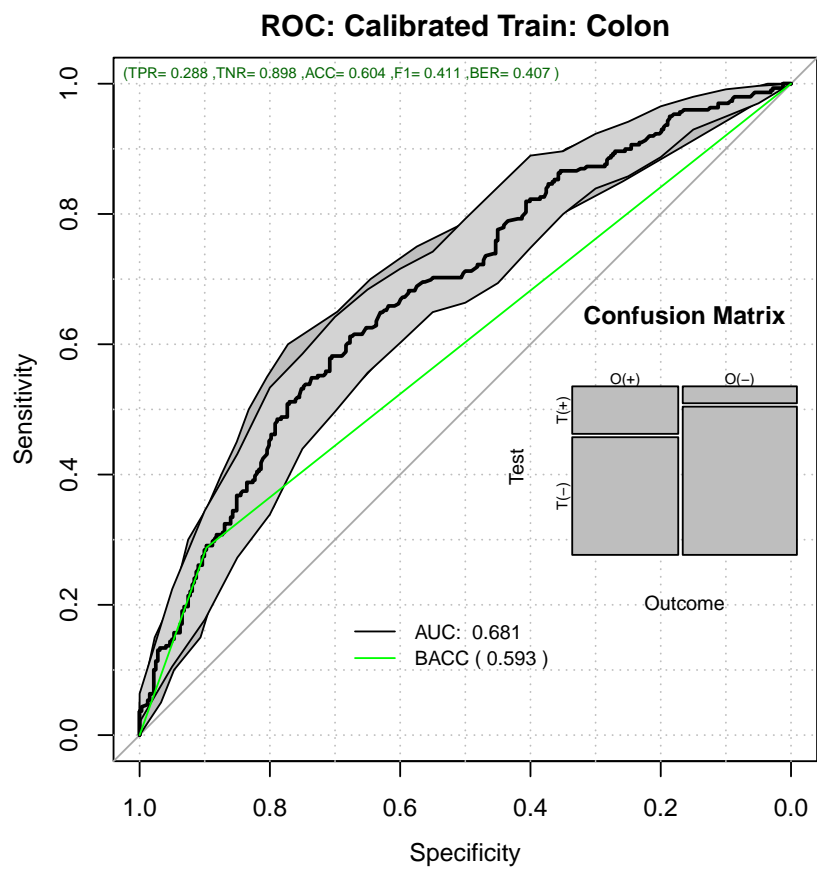


Decision Curve Analysis: Calibrated Train: Colon

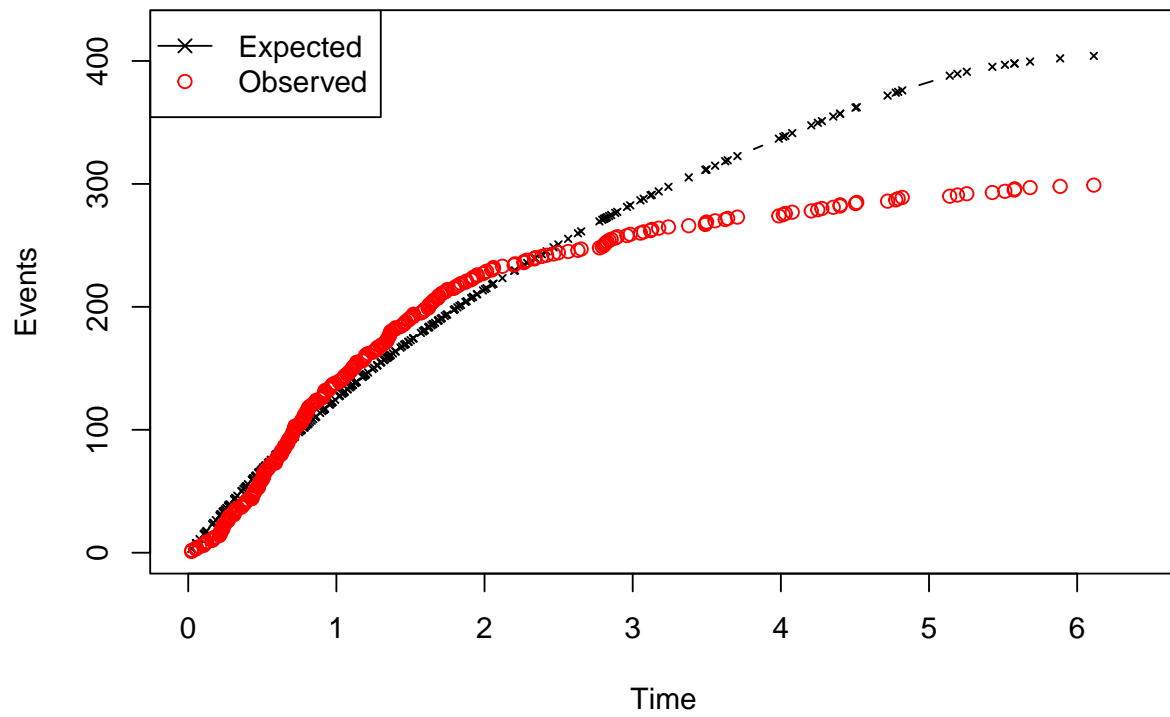


Relative Risk: Calibrated Train: Colon

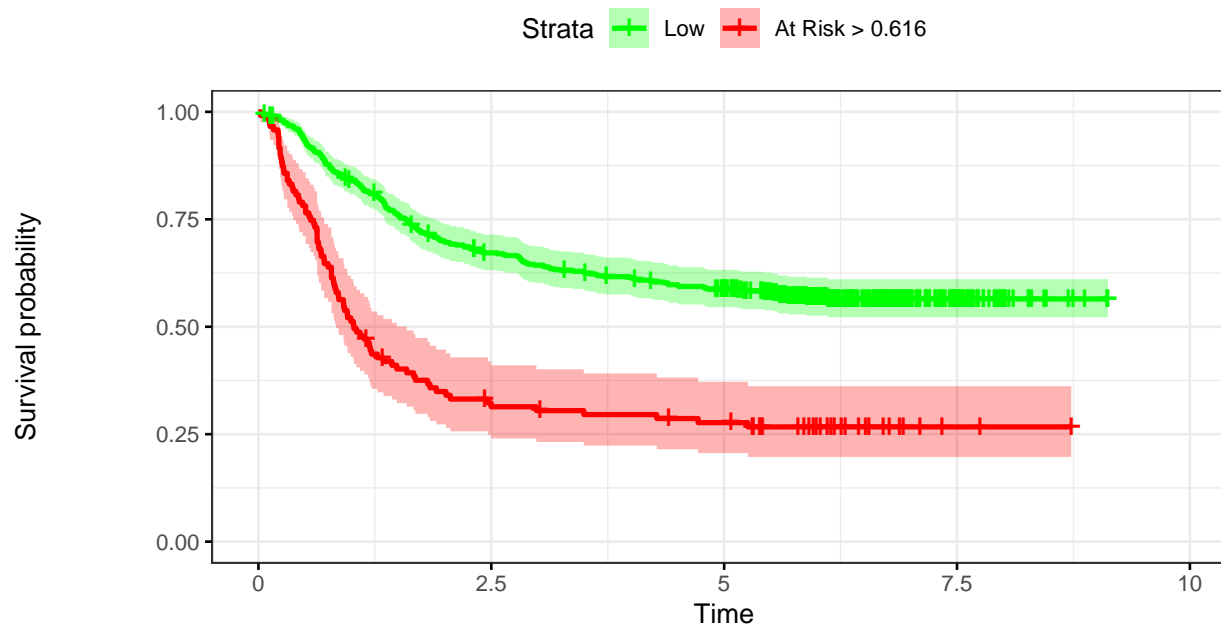




Time vs. Events: Calibrated Train: Colon



Kaplan–Meier: Calibrated Train: Colon



Number at risk

Low	502	328	276	43	0
At Risk > 0.616	119	35	29	2	0

0.2.5 Calibrated Train Performance

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

Table 17: O/E Ratio

est	lower	upper
0.74	0.658	0.829

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

Table 18: O/E Ratio

mean	50%	2.5%	97.5%
0.968	0.969	0.95	0.986

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

Table 19: O/Acum Ratio

mean	50%	2.5%	97.5%
1.05	1.05	1.05	1.06

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

mean.C Index	median	lower	upper
0.653	0.653	0.622	0.685

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

Table 21: ROC AUC

est	lower	upper
0.681	0.639	0.723

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

Table 22: Sensitivity

est	lower	upper
0.288	0.237	0.343

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

Table 23: Specificity

est	lower	upper
0.898	0.859	0.928

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

Table 24: Probability Thresholds

90%
0.616

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

Table 25: Risk Ratio

est	lower	upper
1.7	1.46	1.98

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

Table 26: Logrank test Chisq = 64.815888 on 1 degrees of freedom,
p = 0.000000

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	502	213	259.8	8.42	64.8
class=1	119	86	39.2	55.81	64.8

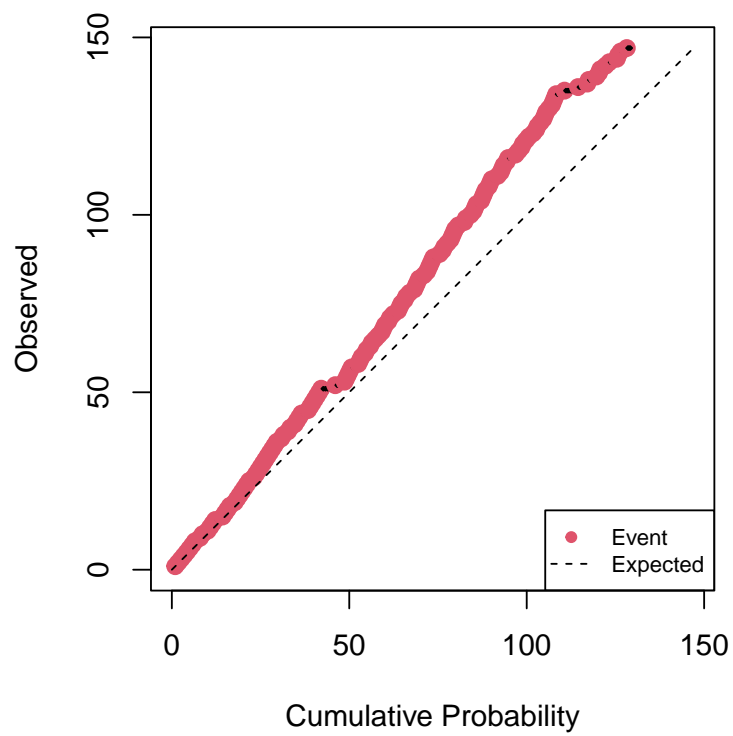
0.2.6 Evaluating on the test set

The calibrated h0 and timeinterval were estimated on the training set

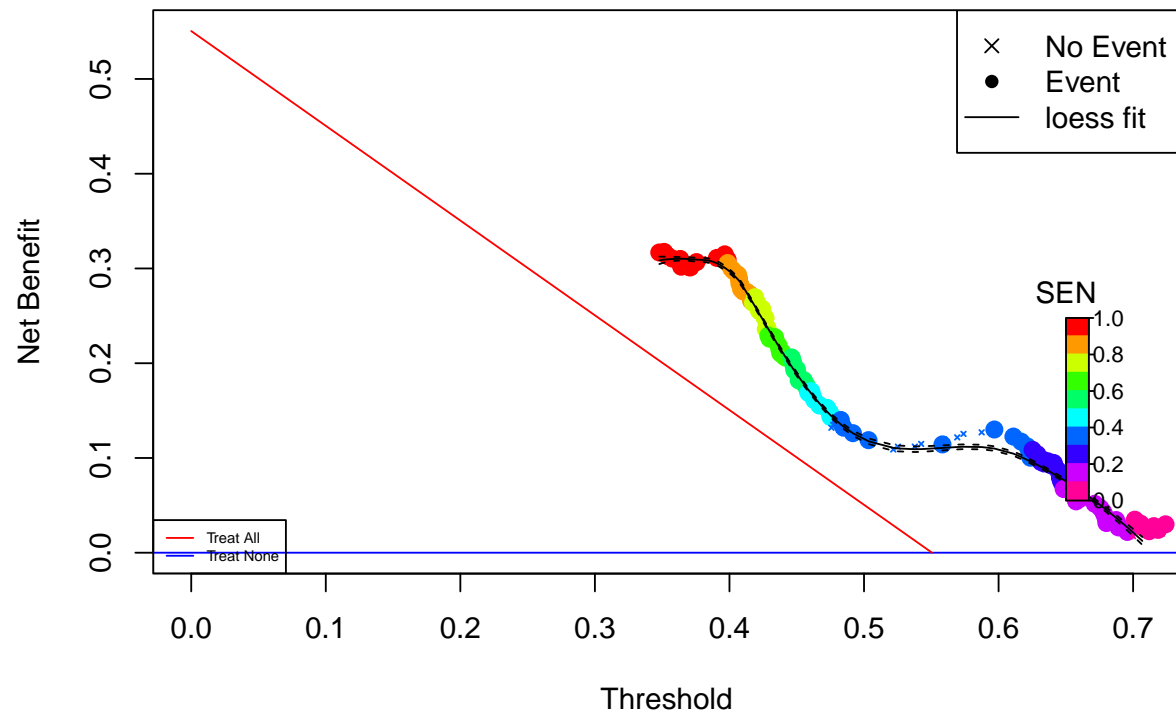
```
index <- predict(ml,dataColonTest)
rdata <- cbind(dataColonTest$status,ppoisGzero(index,h0))

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

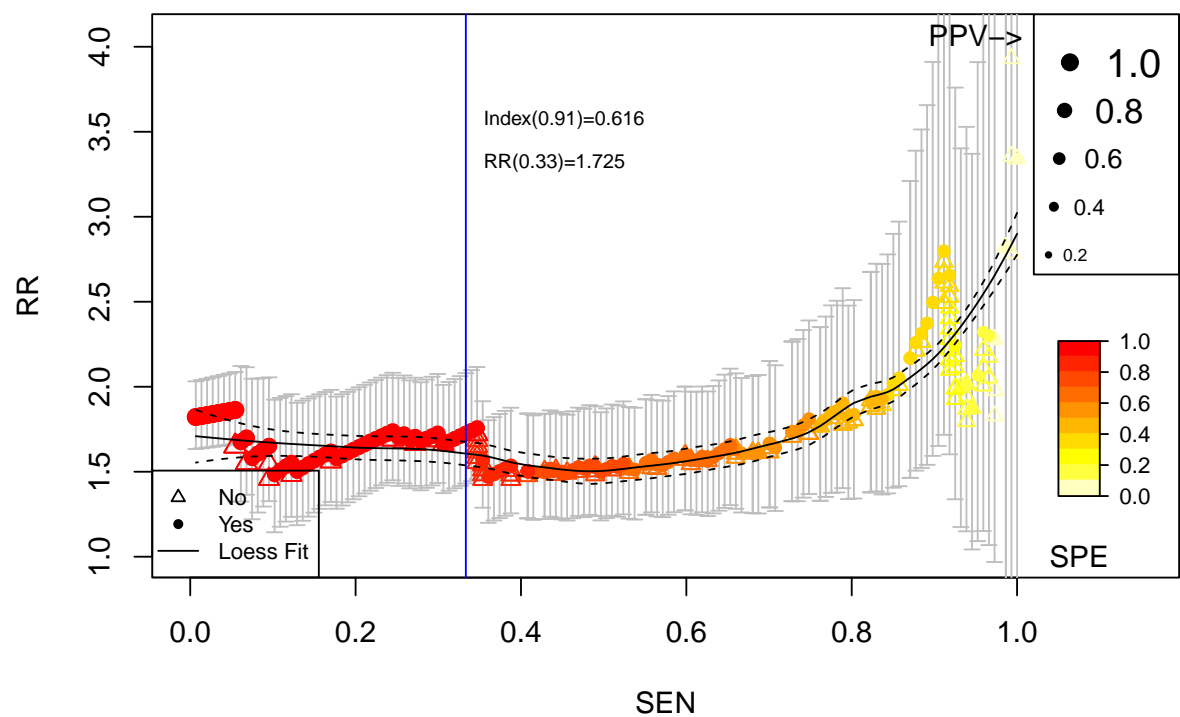
Cumulative vs. Observed: Test: Colon Cancer

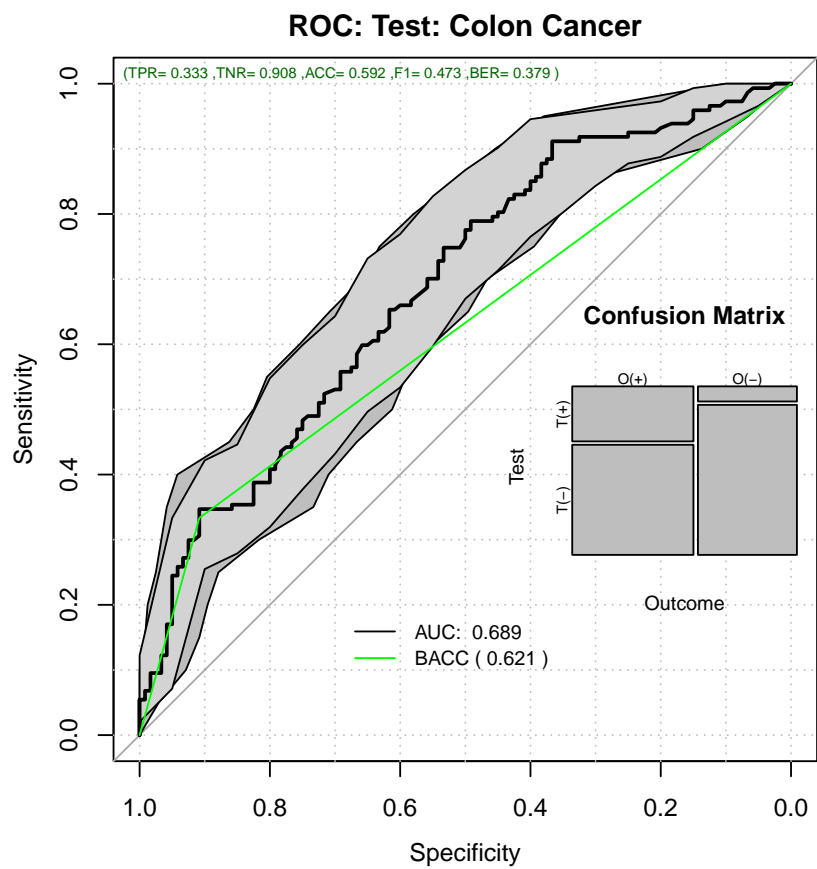


Decision Curve Analysis: Test: Colon Cancer

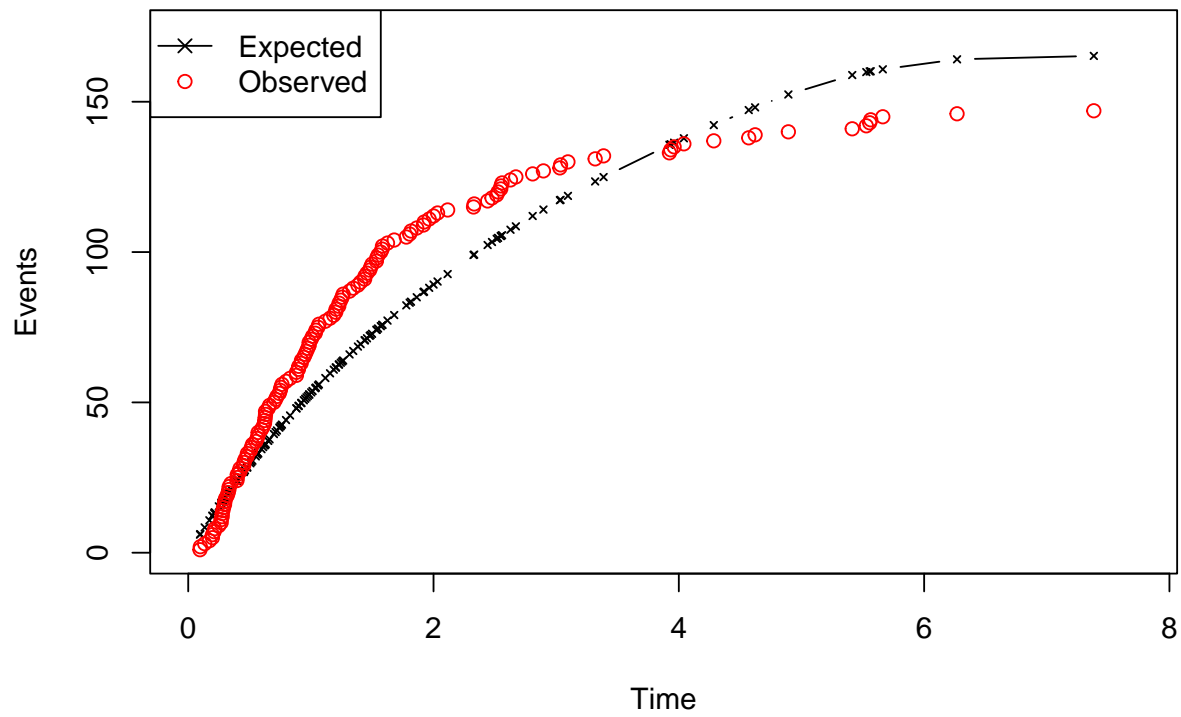


Relative Risk: Test: Colon Cancer

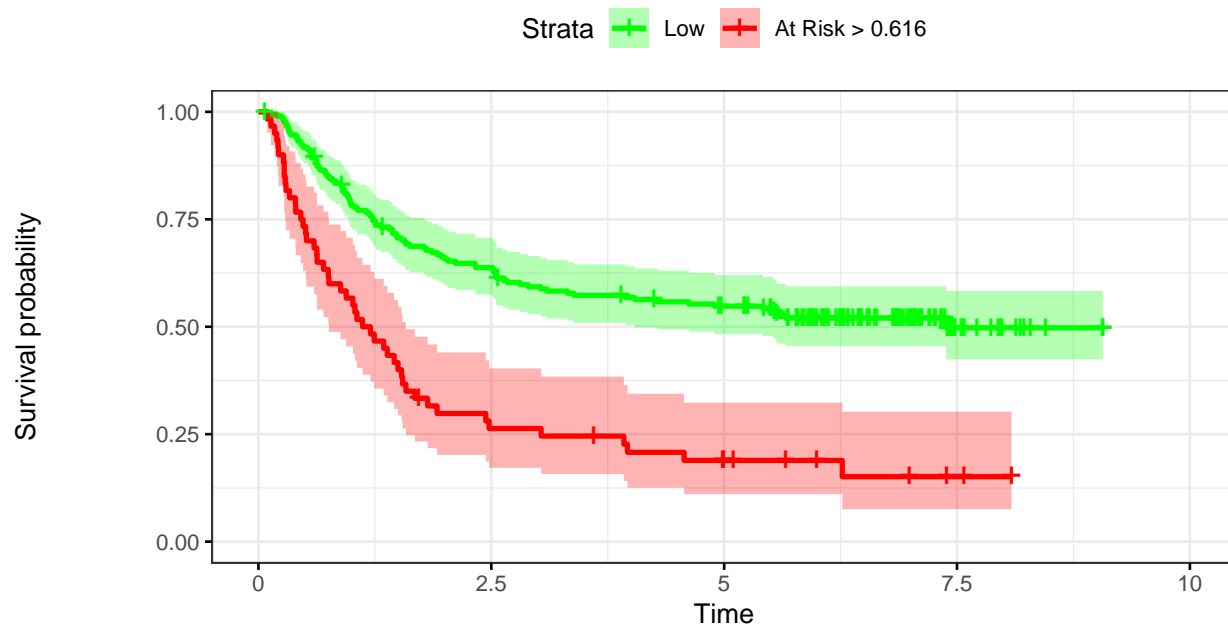




Time vs. Events: Test: Colon Cancer



Kaplan–Meier: Test: Colon Cancer



Number at risk

Low	207	129	106	15	0
At Risk > 0.616	60	15	8	2	0

0.2.7 Test Performance

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

Table 27: O/E Ratio

est	lower	upper
0.89	0.752	1.05

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

Table 28: O/E Ratio

mean	50%	2.5%	97.5%
1.16	1.16	1.13	1.19

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

Table 29: O/Acum Ratio

mean	50%	2.5%	97.5%
1.17	1.17	1.16	1.17

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

mean.C Index	median	lower	upper
0.656	0.656	0.611	0.699

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

Table 31: ROC AUC

est	lower	upper
0.689	0.626	0.753

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

Table 32: Sensitivity

est	lower	upper
0.333	0.258	0.416

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

Table 33: Specificity

est	lower	upper
0.908	0.842	0.953

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

Table 34: Probability Thresholds

90%
0.616

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

Table 35: Risk Ratio

est	lower	upper
1.73	1.43	2.08

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

Table 36: Logrank test Chisq = 35.640204 on 1 degrees of freedom,
p = 0.000000

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	207	98	124.1	5.48	35.6
class=1	60	49	22.9	29.69	35.6

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:
852 Avg. Selected: 7.9 Min Tests: 1 Max Tests: 10 Mean Tests: 4.964789 . MAD: 0.4738661
.[++++].[++++].[++++].[+++].[++++].[++++].[++++].[++++].[++++].[++++]20 Tested: 886 Avg.
Selected: 7.75 Min Tests: 1 Max Tests: 20 Mean Tests: 9.548533 . MAD: 0.4738539 .[++++].[++-
].[++++].[++++].[+++].[+++].[++++].[++++].[++++].[++++]30 Tested: 888 Avg. Selected:
7.833333 Min Tests: 1 Max Tests: 30 Mean Tests: 14.29054 . MAD: 0.4741594 .[+++].[++++-
].[++++].[++++].[++++].[+++].[++++].[++++].[++++].[++++]40 Tested: 888 Avg. Selected: 7.65 Min
Tests: 2 Max Tests: 40 Mean Tests: 19.05405 . MAD: 0.4747553 .[++-].[++++].[++++].[++++].[++++-
].[++++-].[++++].[++++].[++++].[++++]50 Tested: 888 Avg. Selected: 7.72 Min Tests: 2 Max Tests: 50
Mean Tests: 23.81757 . MAD: 0.4748956
```

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

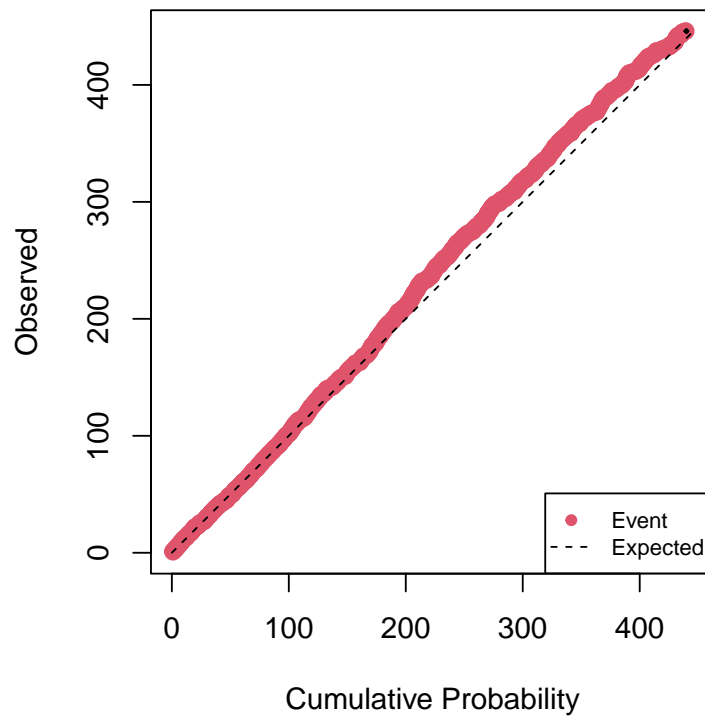
bbx <- boxplot(unlist(stp[,1])~rownames(stp), plot=FALSE)
times <- bbx$stats[3,]
status <- boxplot(unlist(stp[,2])~rownames(stp), plot=FALSE)$stats[3,]
prob <- ppoisGzero(boxplot(unlist(stp[,4])~rownames(stp), plot=FALSE)$stats[3,], h0)

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

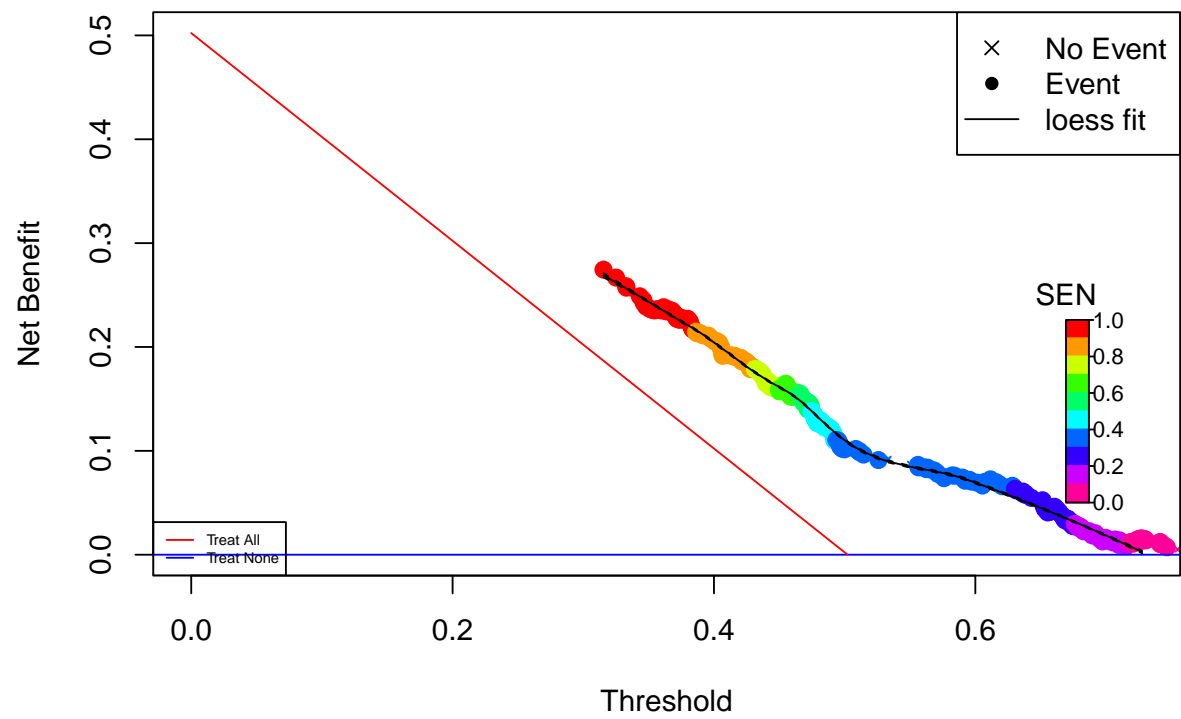
rrAnalysisCVTest <- RRPlot(rdatacv, atThr = rrAnalysisTrain$thr_atP,
  timetoEvent=times,
```

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

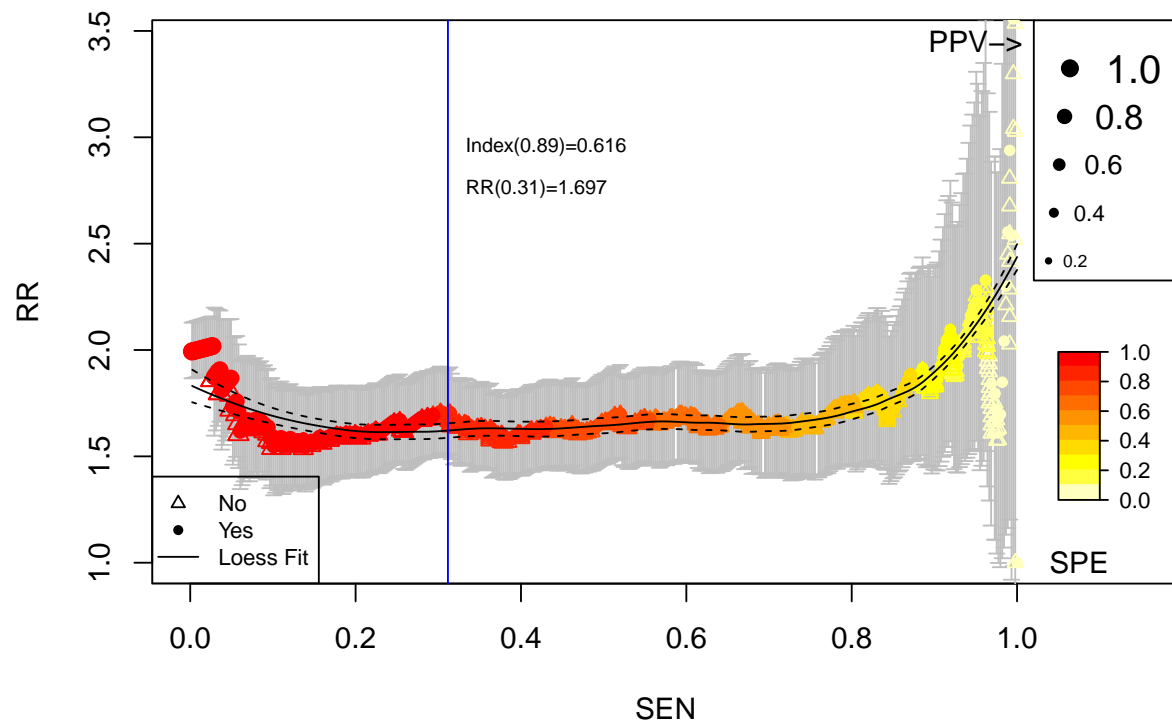
Cumulative vs. Observed: CV Test: Colon Cancer

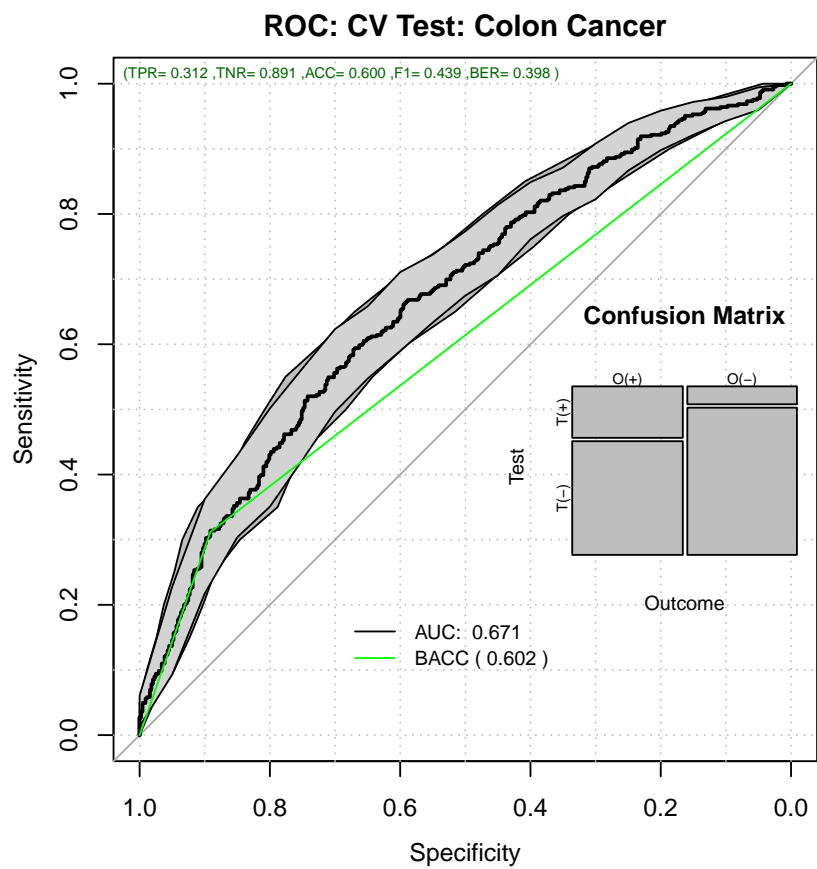


Decision Curve Analysis: CV Test: Colon Cancer

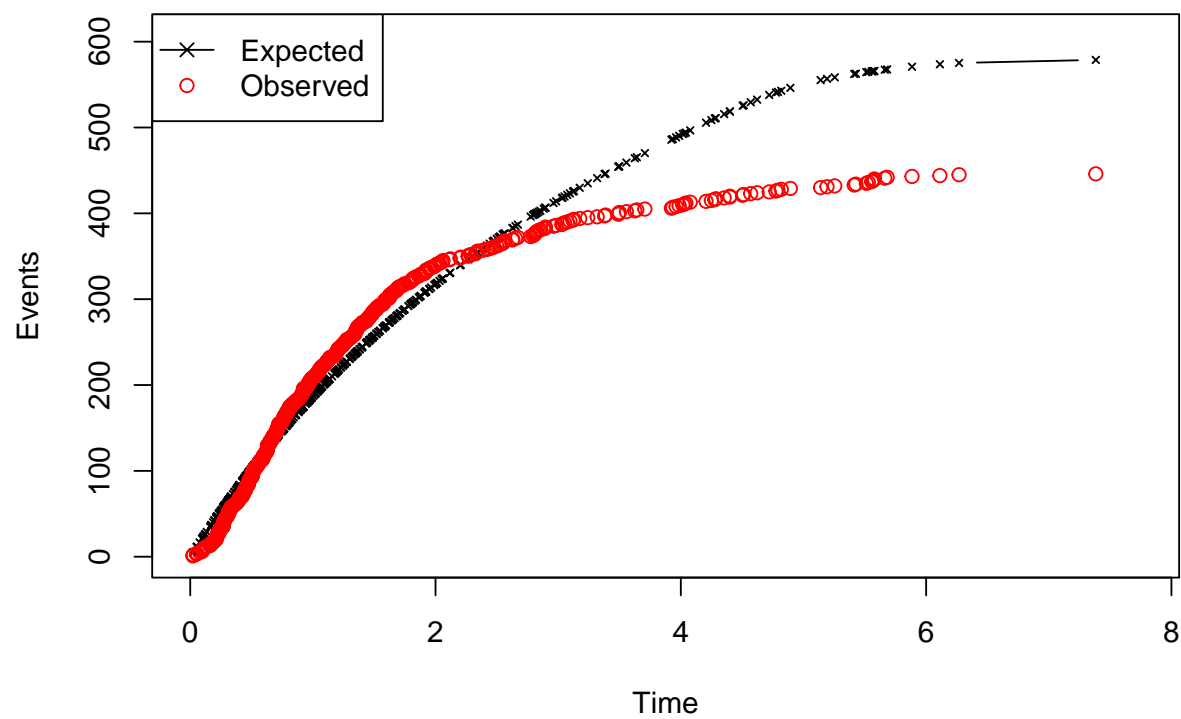


Relative Risk: CV Test: Colon Cancer

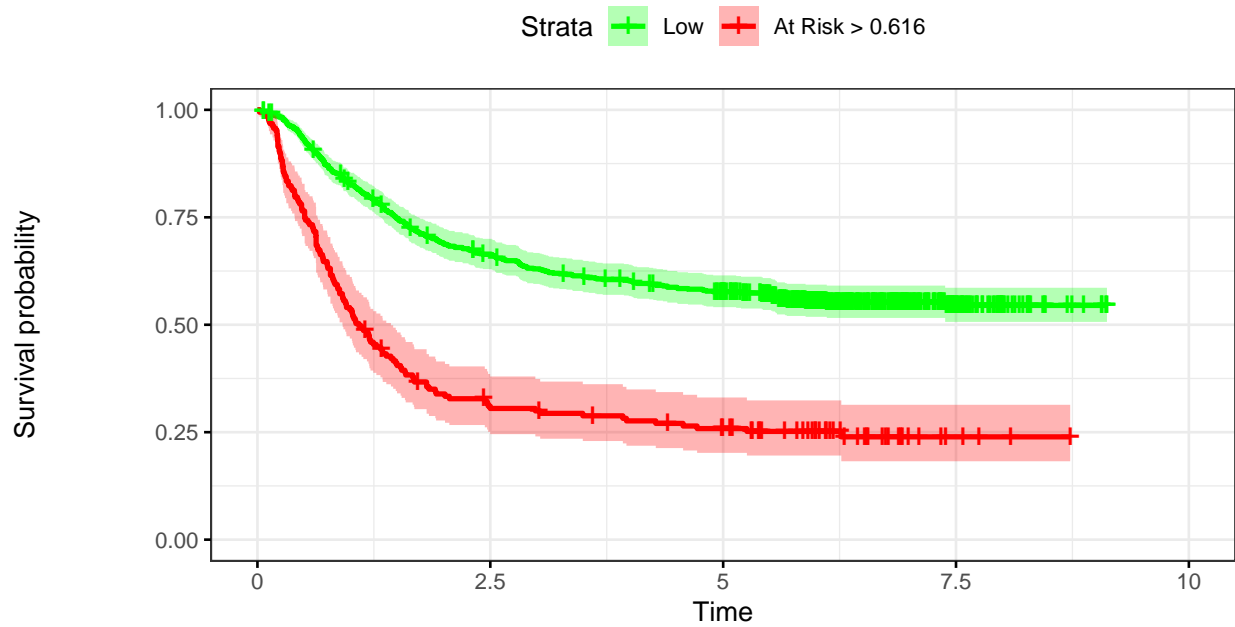




Time vs. Events: CV Test: Colon Cancer



Kaplan–Meier: CV Test: Colon Cancer



Number at risk

Low	701	453	378	58	0
At Risk > 0.616	187	54	41	4	0

0.3.1 CV Test Performance

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

Table 37: O/E Ratio

est	lower	upper
0.771	0.701	0.846

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

Table 38: O/E Ratio

mean	50%	2.5%	97.5%
0.979	0.979	0.965	0.992

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

Table 39: O/Acum Ratio

mean	50%	2.5%	97.5%
1.04	1.04	1.04	1.04

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

mean.C Index	median	lower	upper
0.645	0.645	0.62	0.67

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

Table 41: ROC AUC

est	lower	upper
0.671	0.636	0.706

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

Table 42: Sensitivity

est	lower	upper
0.312	0.269	0.357

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

Table 43: Specificity

est	lower	upper
0.891	0.859	0.919

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

Table 44: Probability Thresholds

90%
0.616

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

Table 45: Risk Ratio

est	lower	upper
1.7	1.51	1.91

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

Table 46: Logrank test Chisq = 98.948116 on 1 degrees of freedom,
p = 0.000000

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	701	307	380.8	14.3	98.9
class=1	187	139	65.2	83.7	98.9