

# Colon Cancer

Jose Tamez

2023-05-01

## Contents

<b>1</b>	<b>RRPlot and the Colon data set</b>	<b>1</b>
1.1	The data set . . . . .	2
1.2	Modeling . . . . .	2
1.3	Cox Model Performance . . . . .	3
1.4	Cross-Validation . . . . .	27

## 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
316	305

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

0	1
126	141

## 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
<b>age_node4</b>	6.05e-03	1.004	1.006	1.009	0.620
<b>age</b>	-2.07e-02	0.970	0.979	0.989	0.514
<b>rxLev_5FU_age</b>	-3.11e-03	0.995	0.997	0.998	0.570
<b>rxLev_5FU</b>	-2.36e-01	0.696	0.790	0.896	0.570
<b>node4</b>	1.81e-01	1.079	1.198	1.330	0.620
<b>age_extent</b>	4.46e-03	1.002	1.004	1.007	0.525
<b>age_nodes</b>	1.00e-04	1.000	1.000	1.000	0.622
<b>extent</b>	1.60e-01	1.061	1.174	1.298	0.548
<b>nodes</b>	4.24e-02	1.012	1.043	1.076	0.628
<b>differ</b>	7.37e-08	1.000	1.000	1.000	0.533

Table 4: Table continues below

	r.Accuracy	full.Accuracy	u.AUC	r.AUC	full.AUC
<b>age_node4</b>	0.595	0.621	0.616	0.596	0.620
<b>age</b>	0.612	0.625	0.513	0.609	0.624
<b>rxLev_5FU_age</b>	0.630	0.645	0.573	0.627	0.644
<b>rxLev_5FU</b>	0.623	0.621	0.573	0.621	0.619
<b>node4</b>	0.631	0.642	0.616	0.631	0.640
<b>age_extent</b>	0.625	0.625	0.526	0.623	0.624
<b>age_nodes</b>	0.609	0.634	0.619	0.608	0.633
<b>extent</b>	0.632	0.645	0.554	0.630	0.644
<b>nodes</b>	0.632	0.645	0.625	0.629	0.644
<b>differ</b>	0.491	0.533	0.527	0.500	0.527

	IDI	NRI	z.IDI	z.NRI	Delta.AUC	Frequency
<b>age_node4</b>	0.02955	0.4220	4.90	6.14	0.023108	1.0
<b>age</b>	0.02207	0.1873	4.23	2.35	0.015136	1.0
<b>rxLev_5FU_age</b>	0.01911	0.2930	4.08	3.97	0.016553	1.0
<b>rxLev_5FU</b>	0.01574	0.2930	3.67	3.97	-0.001565	1.0
<b>node4</b>	0.01328	0.3635	3.47	5.38	0.009490	1.0
<b>age_extent</b>	0.01388	0.1295	3.33	2.30	0.000977	1.0
<b>age_nodes</b>	0.01328	0.2245	3.23	2.93	0.024832	1.0
<b>extent</b>	0.01233	0.0673	3.12	1.07	0.013901	1.0
<b>nodes</b>	0.00827	0.1884	2.76	2.45	0.014472	1.0
<b>differ</b>	0.00656	0.1089	1.85	1.80	0.027215	0.4

## 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)

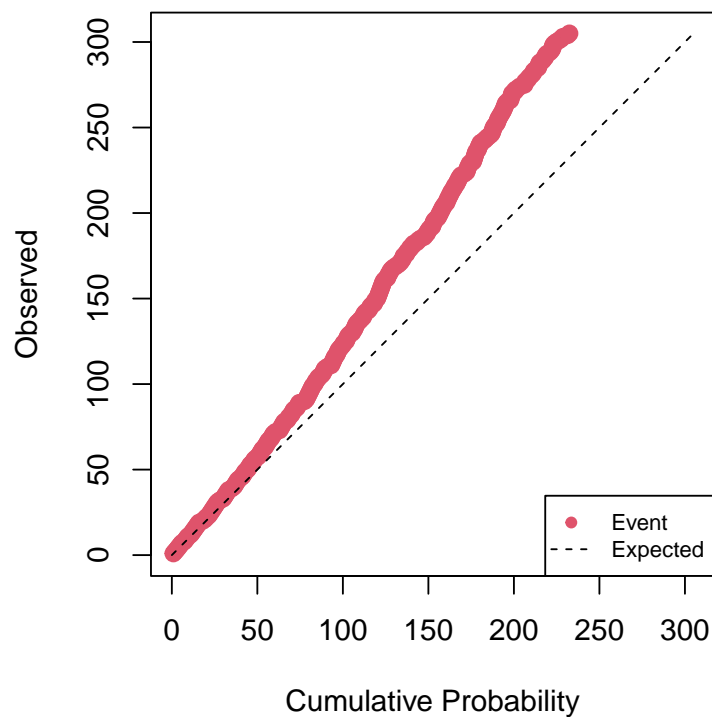
h0 <- sum(dataColonTrain$status & dataColonTrain$time <= timeinterval)
h0 <- h0/sum((dataColonTrain$time > timeinterval) | (dataColonTrain$status==1))

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

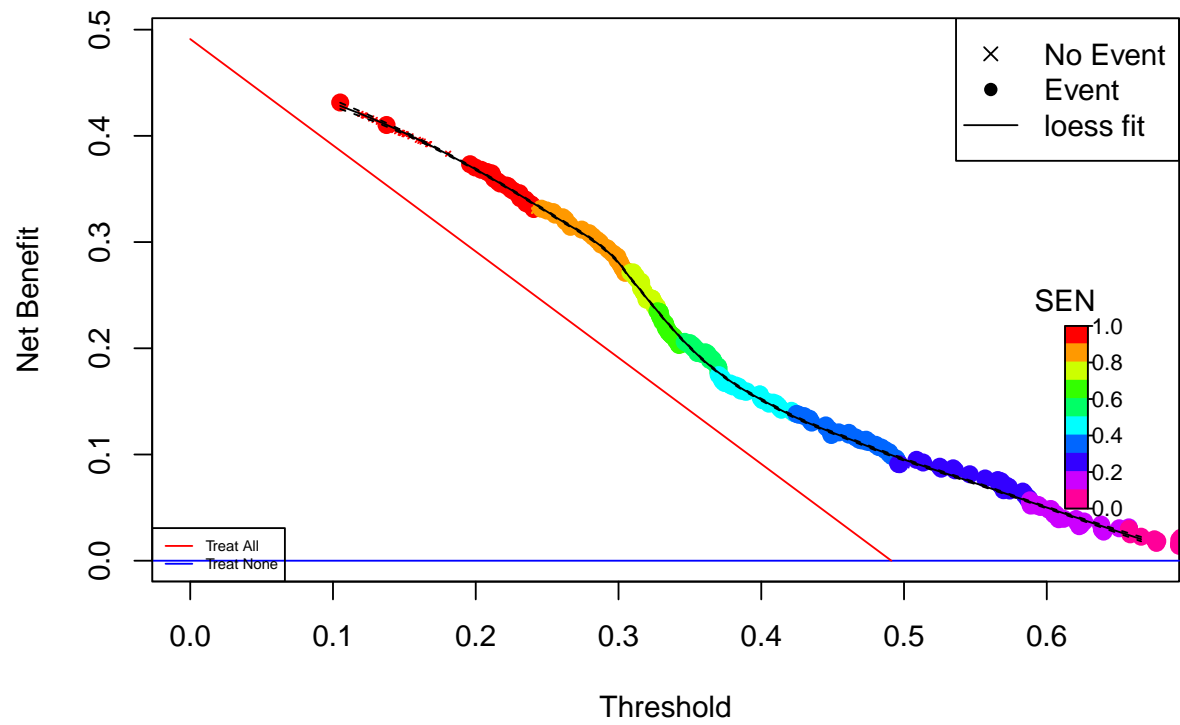
rrAnalysisTrain <- RRPlot(rdata,atProb=c(0.90),
  timetoEvent=dataColonTrain$time,
  title="Raw Train: Colon Cancer",
  ysurvlim=c(0.00,1.0),
  riskTimeInterval=timeinterval)

```

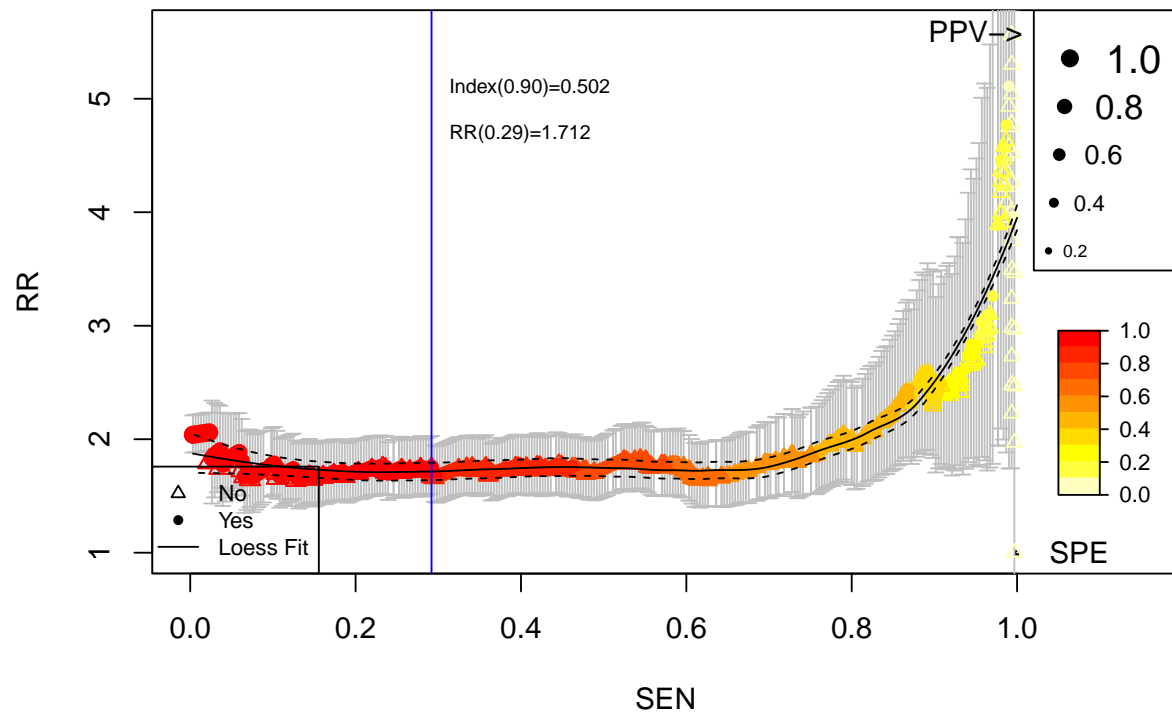
## Cumulative vs. Observed: Raw Train: Colon Cancer

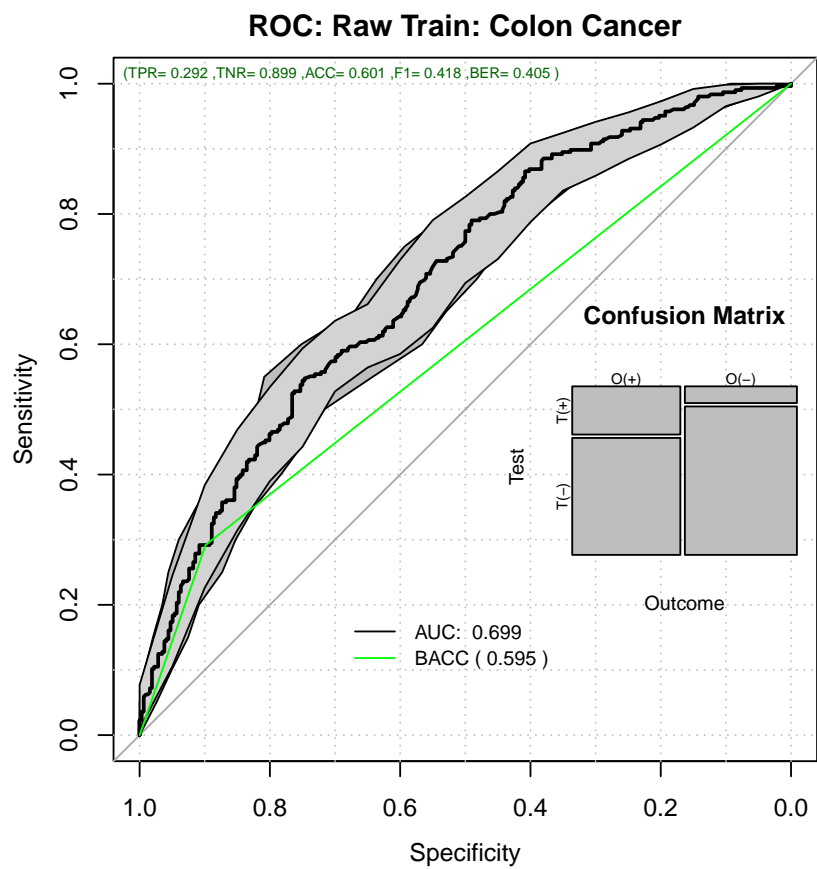


### Decision Curve Analysis: Raw Train: Colon Cancer

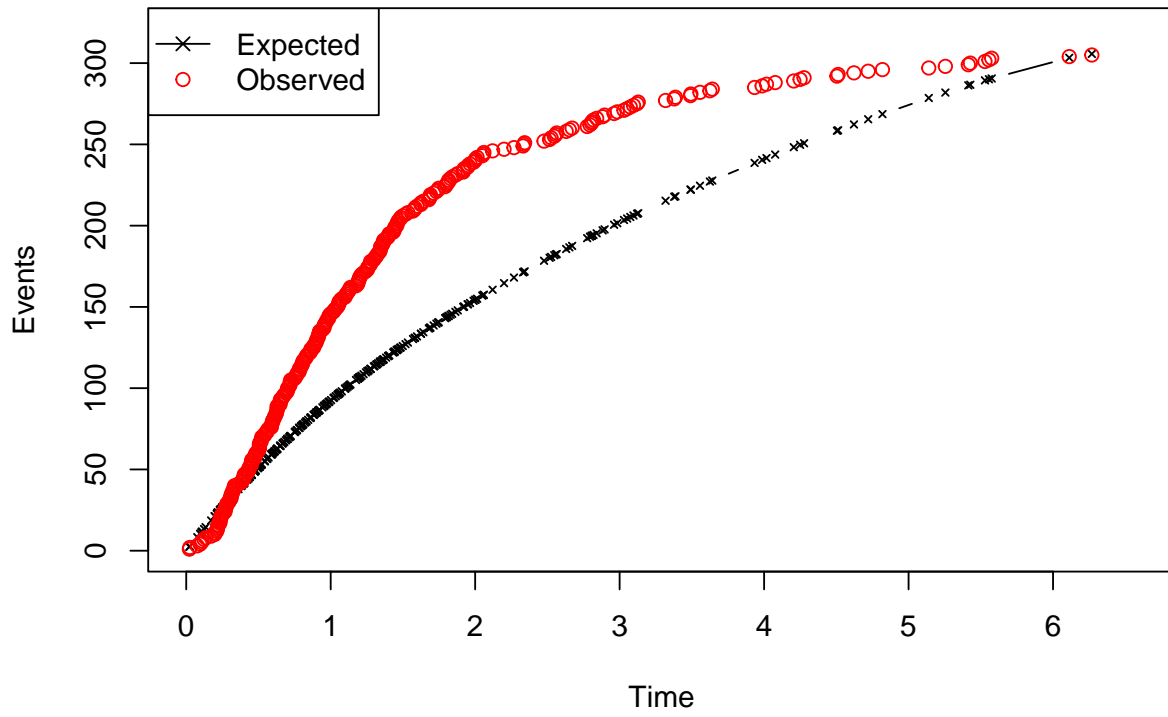


# Relative Risk: Raw Train: Colon Cancer



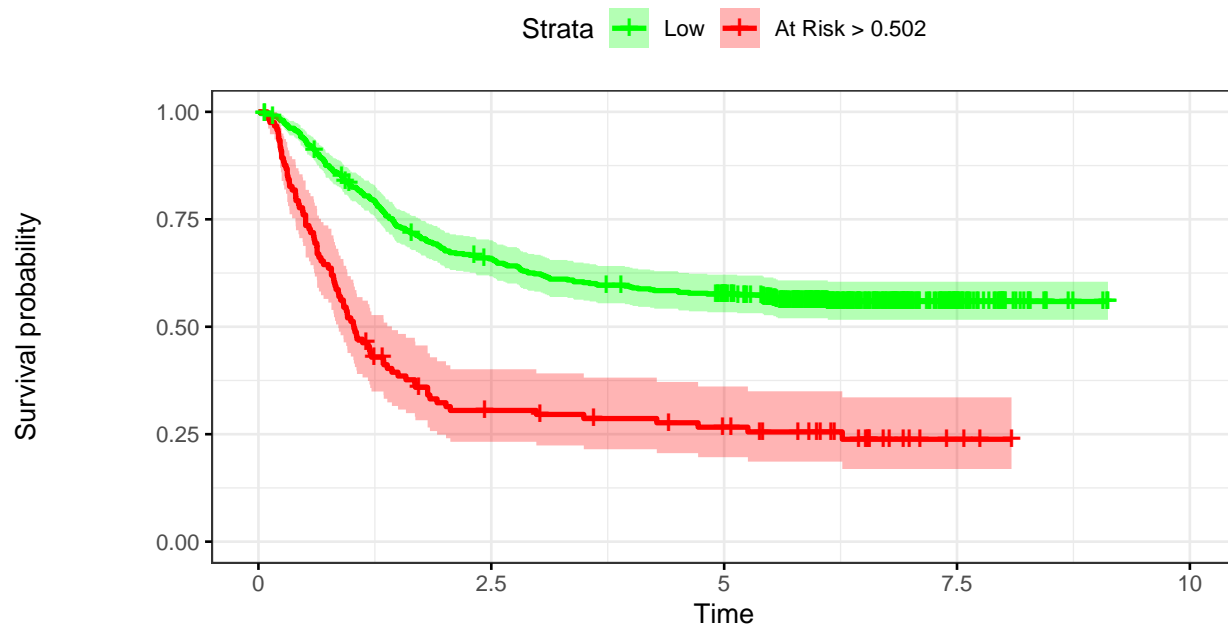


**Time vs. Events: Raw Train: Colon Cancer**





## Kaplan–Meier: Raw Train: Colon Cancer



### Number at risk

Low	500	321	271	43	0
At Risk > 0.502	121	33	25	3	0

### 1.3.2 Uncalibrated Performance Report

```
pander::pander(t(rrAnalysisTrain$keyPoints),caption="Threshold values")
```

Table 6: Threshold values

	@:0.9	@MAX_BACC	@MAX_RR	@SPE100
<b>Thr</b>	0.502	0.361	0.204	0.105
<b>RR</b>	1.685	1.832	4.768	1.000
<b>SEN</b>	0.292	0.548	0.987	1.000
<b>SPE</b>	0.896	0.747	0.104	0.000
<b>BACC</b>	0.594	0.647	0.546	0.500

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

Table 7: O/E Ratio

est	lower	upper
0.998	0.889	1.12

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

Table 8: O/E Ratio

mean	50%	2.5%	97.5%
1.39	1.39	1.36	1.41

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

Table 9: O/Acum Ratio

mean	50%	2.5%	97.5%
1.27	1.27	1.26	1.27

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

mean.C Index	median	lower	upper
0.67	0.67	0.638	0.699

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

Table 11: ROC AUC

est	lower	upper
0.699	0.659	0.74

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

Table 12: Sensitivity

est	lower	upper
0.292	0.241	0.346

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

Table 13: Specificity

est	lower	upper
0.899	0.86	0.93

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

Table 14: Probability Thresholds

90%	at_max_BACC	at_max_RR	atSPE100
0.502	0.361	0.204	0.105

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

Table 15: Risk Ratio

est	lower	upper
1.71	1.48	1.98

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

Table 16: Logrank test Chisq = 69.007475 on 1 degrees of freedom,  
p = 0.000000

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
<b>class=0</b>	500	216	264.8	9.0	69
<b>class=1</b>	121	89	40.2	59.3	69

### 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
0.651	1.48	2.97

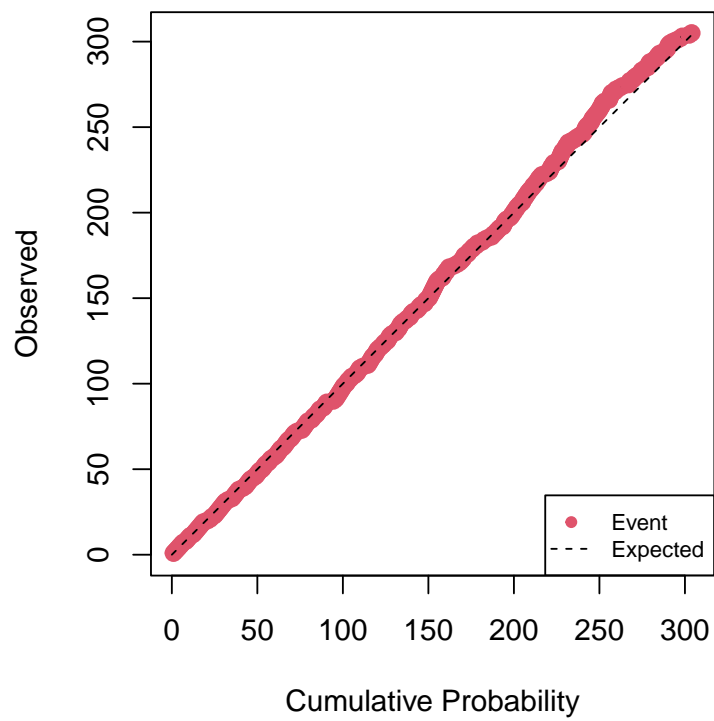
### 1.3.4 The RRplot() of the calibrated model

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

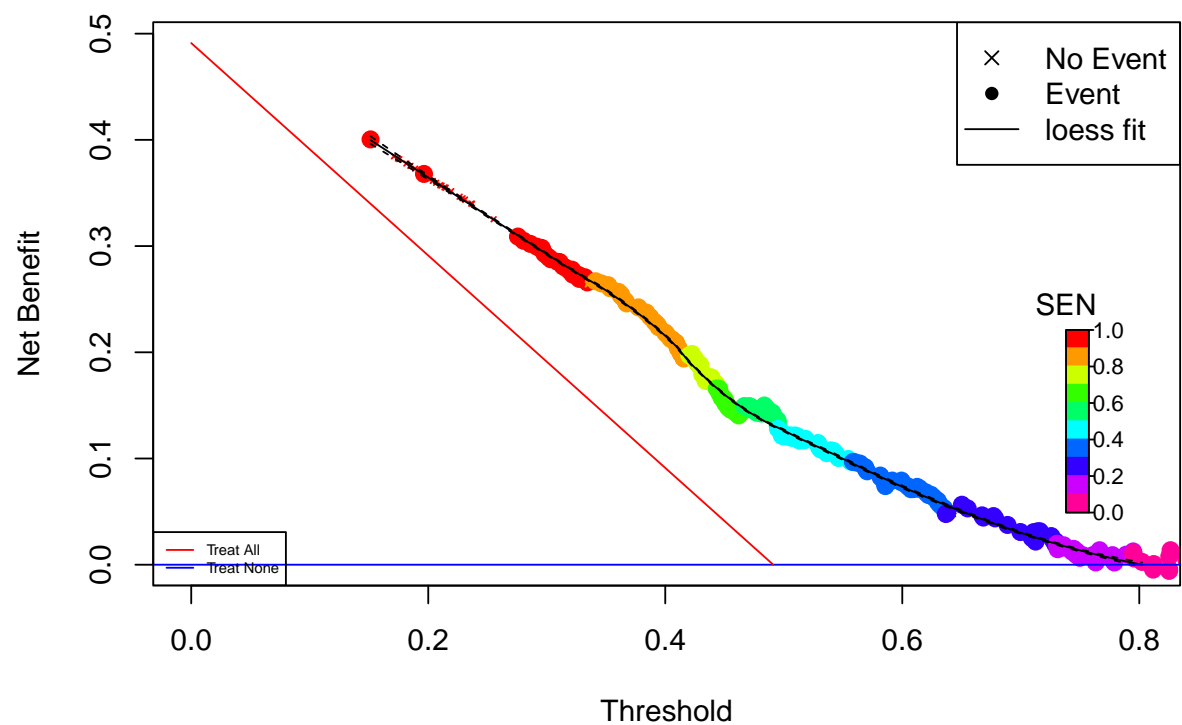
rdata <- cbind(dataColonTrain$status,calprob$prob)

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

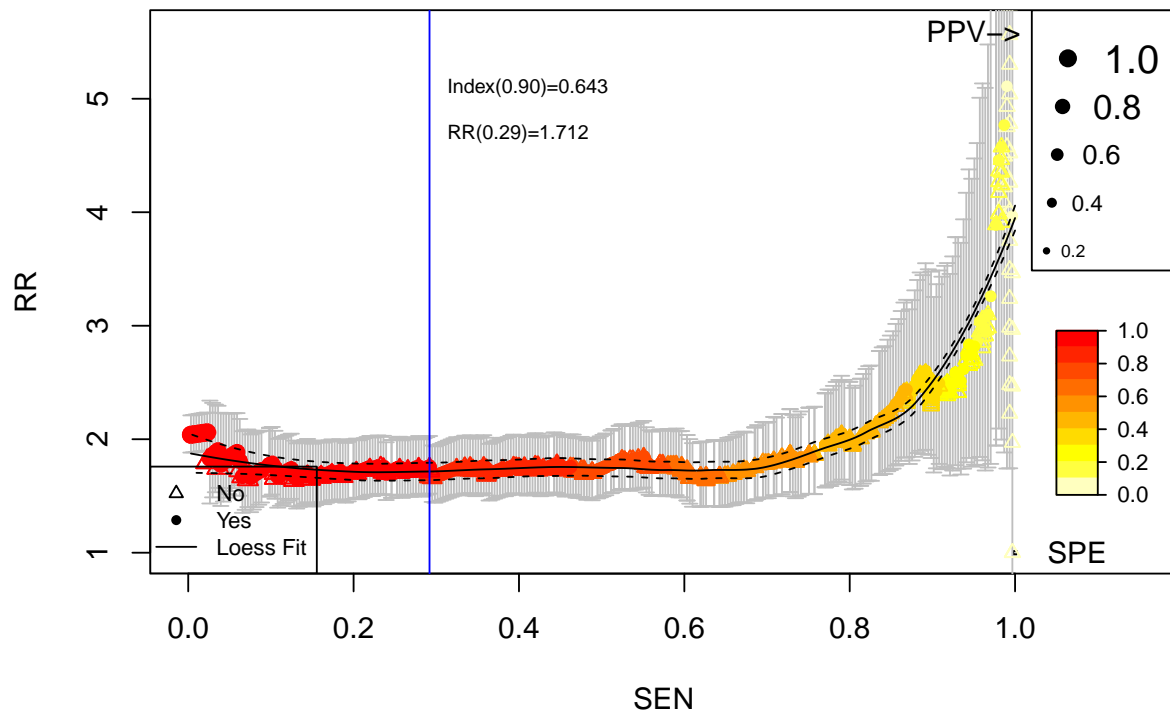
### Cumulative vs. Observed: Calibrated Train: Colon

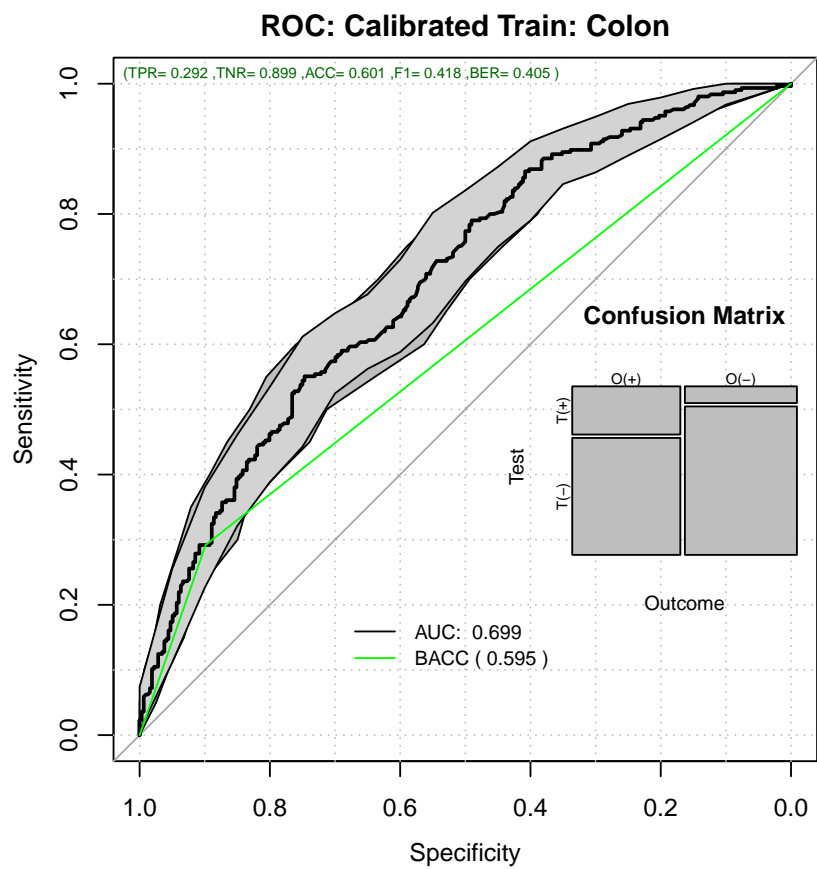


Decision Curve Analysis: Calibrated Train: Colon

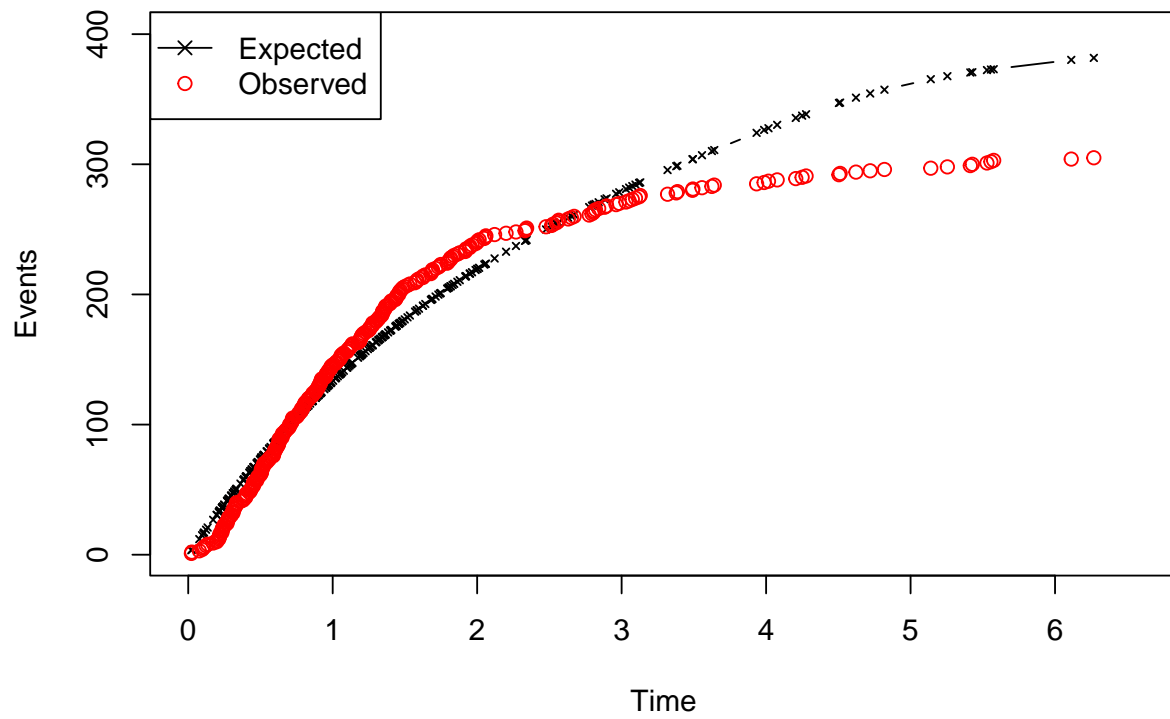


# Relative Risk: Calibrated Train: Colon



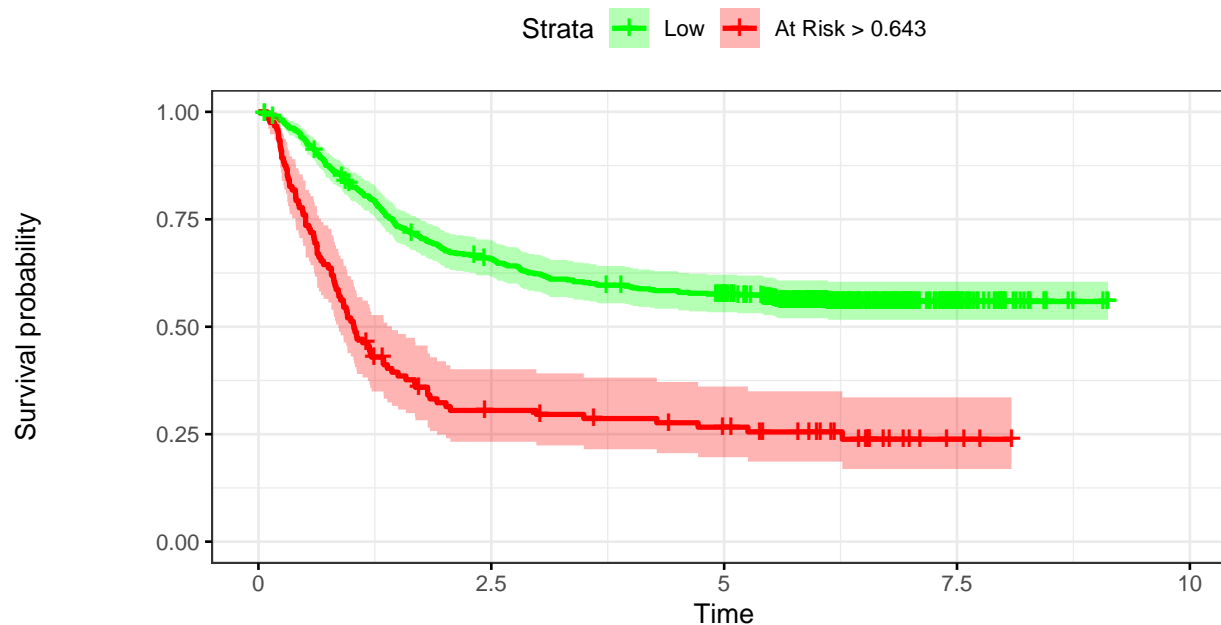


**Time vs. Events: Calibrated Train: Colon**





## Kaplan–Meier: Calibrated Train: Colon



### Number at risk

Low	500	321	271	43	0
At Risk > 0.643	121	33	25	3	0

### 1.3.5 Calibrated Train Performance

```
pander::pander(t(rrAnalysisTrain$keyPoints),caption="Threshold values")
```

Table 18: Threshold values

	@:0.9	@MAX_BACC	@MAX_RR	@SPE100
<b>Thr</b>	0.643	0.484	0.286	0.151
<b>RR</b>	1.685	1.844	4.768	1.000
<b>SEN</b>	0.292	0.551	0.987	1.000
<b>SPE</b>	0.896	0.747	0.104	0.000
<b>BACC</b>	0.594	0.649	0.546	0.500

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

Table 19: O/E Ratio

est	lower	upper
0.799	0.712	0.894

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

Table 20: O/E Ratio

mean	50%	2.5%	97.5%
0.973	0.973	0.956	0.991

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

Table 21: O/Acum Ratio

mean	50%	2.5%	97.5%
1.01	1.01	1	1.01

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

mean.C Index	median	lower	upper
0.67	0.67	0.641	0.698

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

Table 23: ROC AUC

est	lower	upper
0.699	0.659	0.74

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

Table 24: Sensitivity

est	lower	upper
0.292	0.241	0.346

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

Table 25: Specificity

est	lower	upper
0.899	0.86	0.93

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

Table 26: Probability Thresholds

90%	at_max_BACC	at_max_RR	atSPE100
0.643	0.484	0.286	0.151

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

Table 27: Risk Ratio

est	lower	upper
1.71	1.48	1.98

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

Table 28: Logrank test Chisq = 69.007475 on 1 degrees of freedom,  
p = 0.000000

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
<b>class=0</b>	500	216	264.8	9.0	69
<b>class=1</b>	121	89	40.2	59.3	69

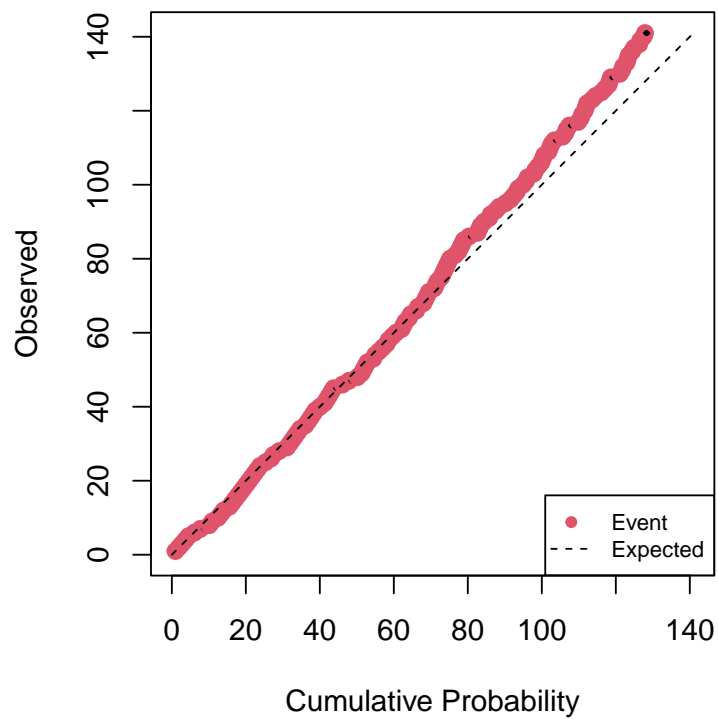
### 1.3.6 Evaluating on the test set

The calibrated h0 and timeinterval were estimated on the training set

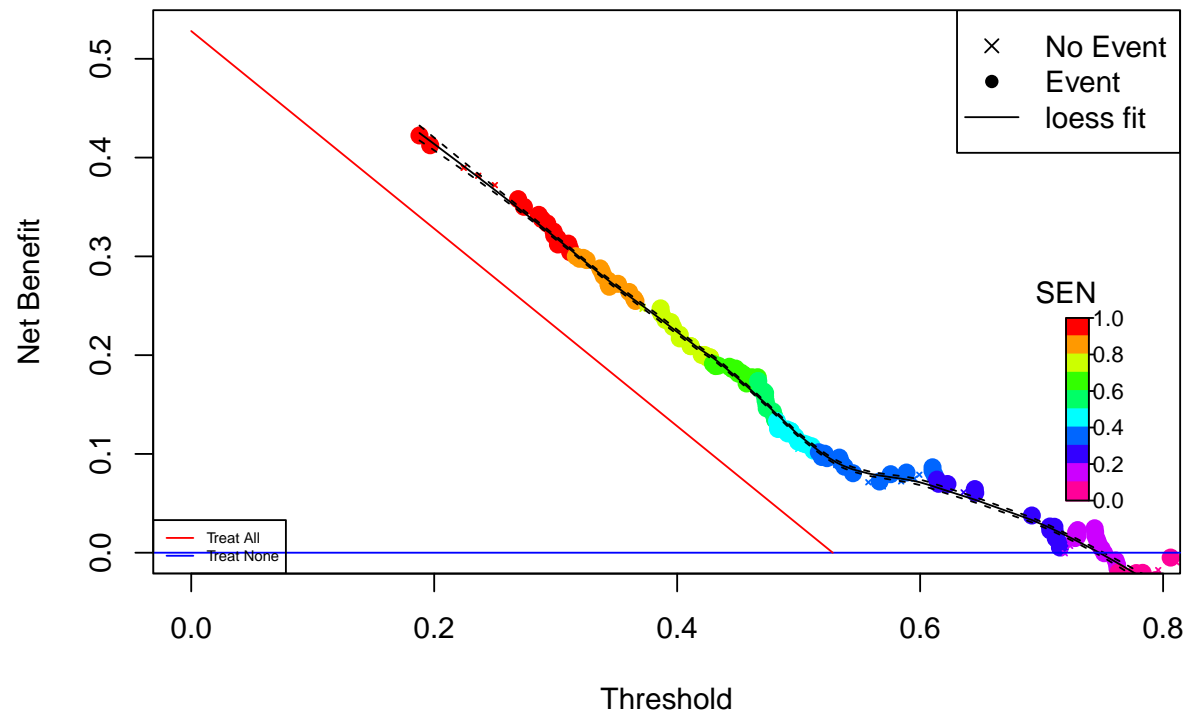
```
index <- predict(ml,dataColonTest)
rdata <- cbind(dataColonTest$status,ppoisGzero(index,h0))

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

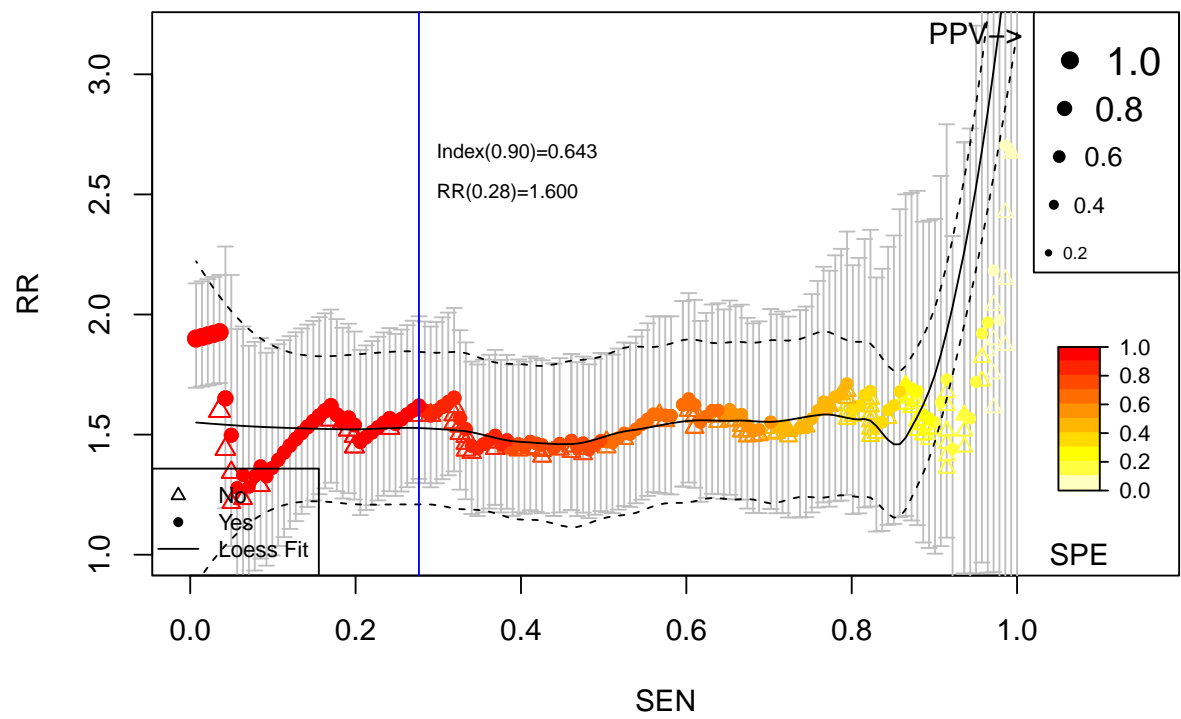
### Cumulative vs. Observed: Test: Colon Cancer

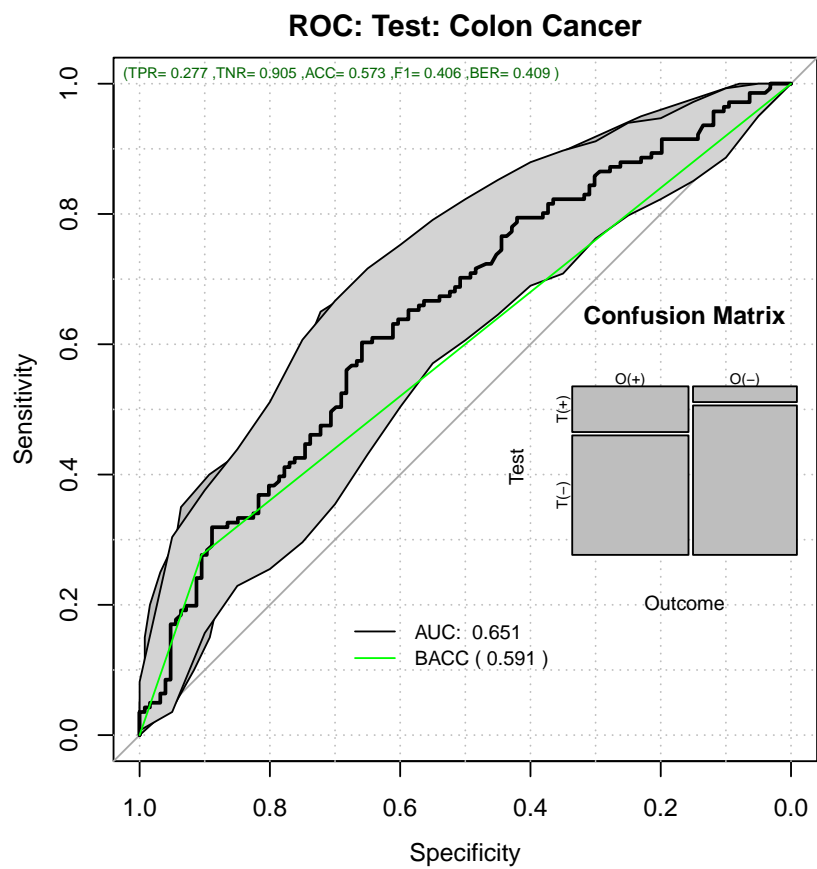


## Decision Curve Analysis: Test: Colon Cancer

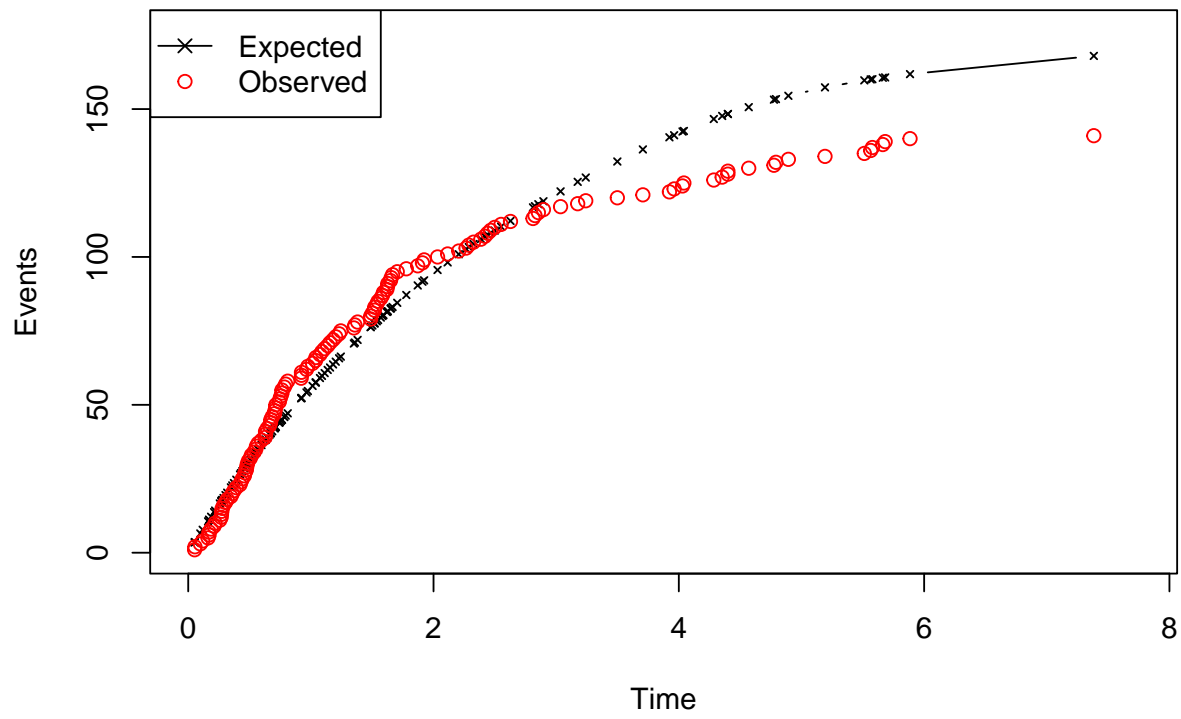


Relative Risk: Test: Colon Cancer



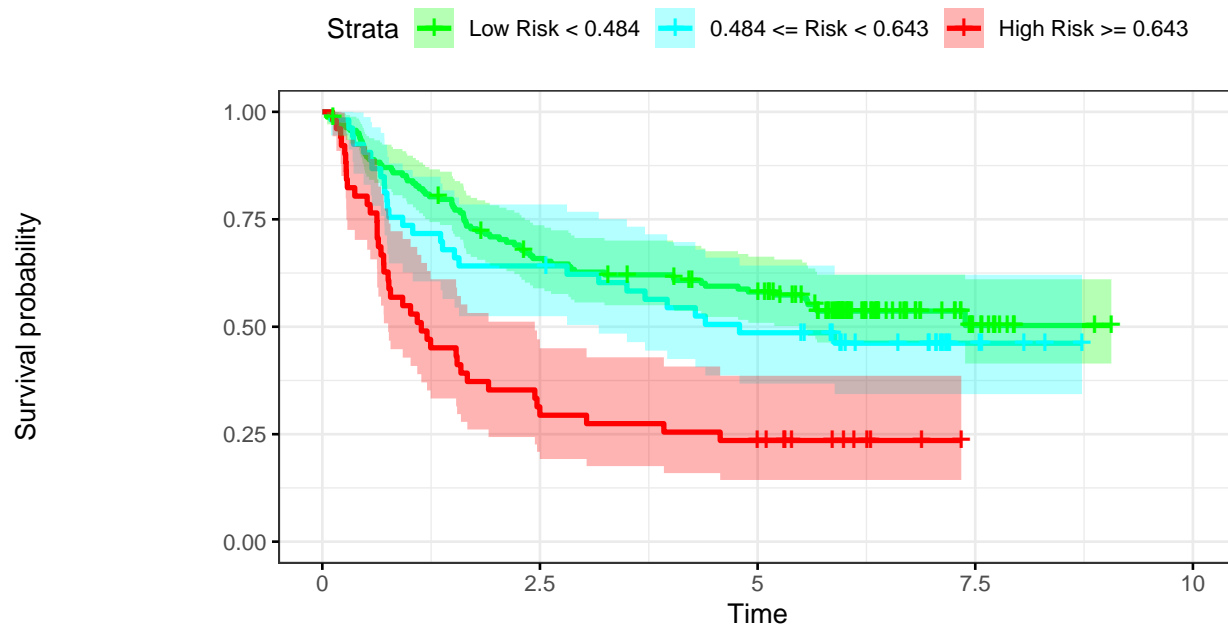


Time vs. Events: Test: Colon Cancer





## Kaplan–Meier: Test: Colon Cancer



### Number at risk

Low Risk < 0.484	163	104	87	11	0
0.484 <= Risk < 0.643	53	34	25	5	0
High Risk >= 0.643	51	15	11	0	0

### 1.3.7 Test Performance

```
pander::pander(t(rrAnalysisTest$keyPoints),caption="Threshold values")
```

Table 29: Threshold values (continued below)

	@:0.64303803374701	@:0.483598597950569	@:0.286109857686596
<b>Thr</b>	0.645	0.483	0.2861
<b>RR</b>	1.619	1.437	2.1833
<b>SEN</b>	0.277	0.482	0.9716
<b>SPE</b>	0.905	0.706	0.0952
<b>BACC</b>	0.591	0.594	0.5334

	@:0.15130951853827	@MAX_BACC	@MAX_RR	@SPE100
<b>Thr</b>	0.1879	0.466	0.289	0.1879
<b>RR</b>	21.4449	1.648	1.966	21.4449
<b>SEN</b>	1.0000	0.603	0.965	1.0000
<b>SPE</b>	0.0317	0.659	0.103	0.0317
<b>BACC</b>	0.5159	0.631	0.534	0.5159

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

Table 31: O/E Ratio

est	lower	upper
0.839	0.707	0.99

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

Table 32: O/E Ratio

mean	50%	2.5%	97.5%
1.01	1.01	0.984	1.03

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

Table 33: O/Acum Ratio

mean	50%	2.5%	97.5%
1.03	1.03	1.02	1.04

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

mean.C Index	median	lower	upper
0.619	0.618	0.573	0.663

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

Table 35: ROC AUC

est	lower	upper
0.651	0.585	0.716

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

Table 36: Sensitivity

est	lower	upper
0.277	0.205	0.358

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

Table 37: Specificity

est	lower	upper
0.905	0.84	0.95

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

Table 38: Probability Thresholds

90%	at_max_BACC	at_max_RR	atSPE100	at_max_BACC	at_max_RR	atSPE100
0.643	0.484	0.286	0.151	0.466	0.289	0.188

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

Table 39: Risk Ratio

est	lower	upper
1.6	1.3	1.97

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

Table 40: Logrank test Chisq = 25.920753 on 2 degrees of freedom,  
p = 0.000002

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
<b>class=0</b>	163	74	93.5	4.0691	12.1407
<b>class=1</b>	53	28	28.6	0.0147	0.0184
<b>class=2</b>	51	39	18.8	21.5536	25.1518

## 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: 850 Avg. Selected: 7.9 Min
Tests: 1 Max Tests: 10 Mean Tests: 4.976471 . MAD: 0.4681416 .[+++].[++].[+++].[+++].[++].[+++].[+++
].[+++].[++].[++++]20 Tested: 888 Avg. Selected: 8 Min Tests: 1 Max Tests: 20 Mean Tests: 9.527027
. MAD: 0.4671251 .[+++].[+++++].[+++++].[+++++].[++].[++].[+++].[++++].[+++++].[++++]30
Tested: 888 Avg. Selected: 7.966667 Min Tests: 1 Max Tests: 30 Mean Tests: 14.29054 . MAD: 0.4672746
.[++++-].[++-].[+++-].[+++++].[++].[++-].[+++].[++].[++].[++].40 Tested: 888 Avg. Selected: 7.95
```

Min Tests: 2 Max Tests: 40 Mean Tests: 19.05405 . MAD: 0.4669155 .[+++].[+++].[+++].[+++].[+++-.  
 ].[+++].[+++].[+++].[+-].[++++-]50 Tested: 888 Avg. Selected: 7.88 Min Tests: 5 Max Tests: 50 Mean  
 Tests: 23.81757 . MAD: 0.4669755

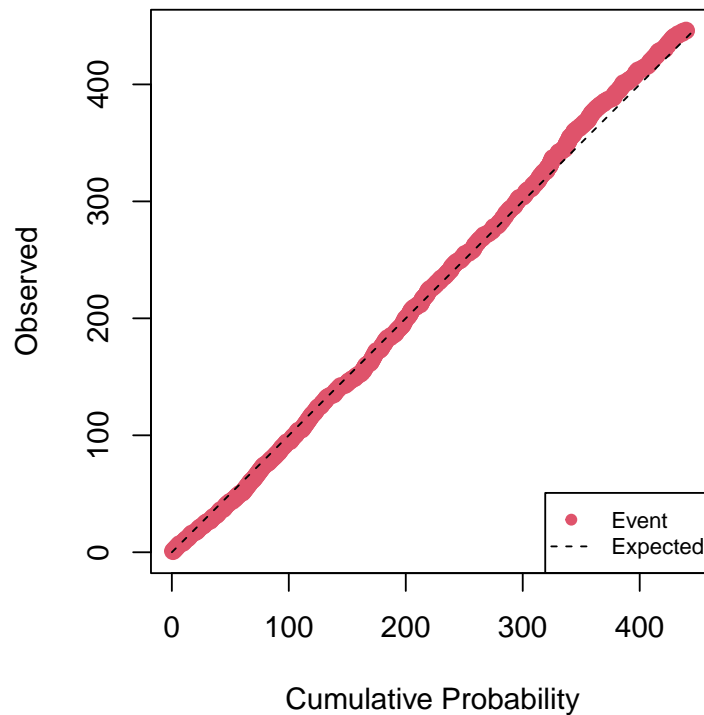
```
stp <- rcv$survTestPredictions
stp <- stp[!is.na(stp[,4]),]

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)

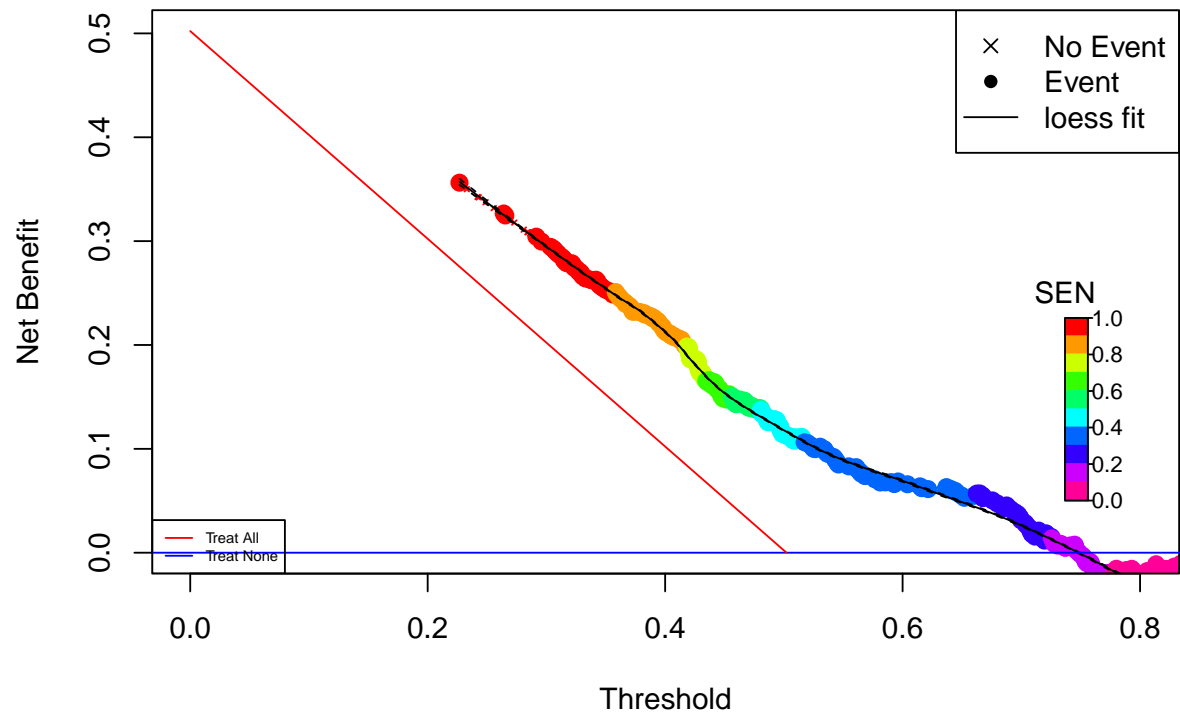
rdatacv <- cbind(status,prob)
rownames(rdatacv) <- bbx$names
names(times) <- bbx$names

rrAnalysisCVTest <- RRPlot(rdatacv,atThr = rrAnalysisTrain$thr_atP,
  timetoEvent=times,
  title="CV Test: Colon Cancer",
  ysurvlim=c(0.00,1.0),
  riskTimeInterval=timeinterval)
```

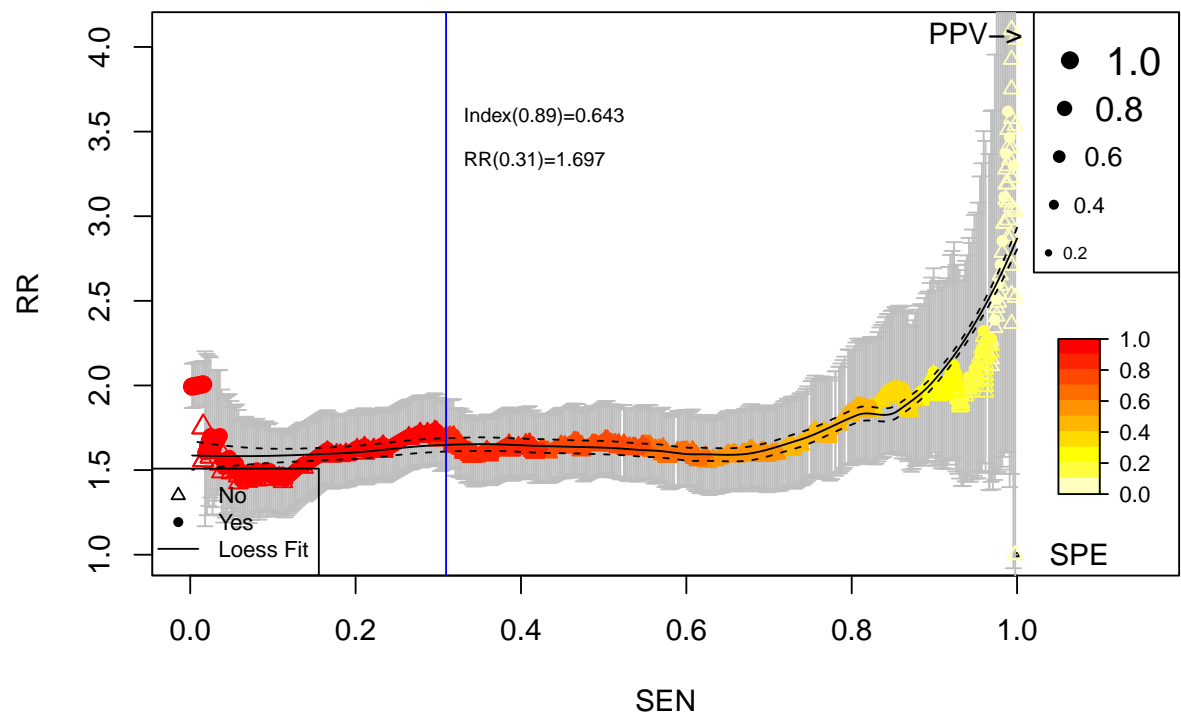
## Cumulative vs. Observed: CV Test: Colon Cancer

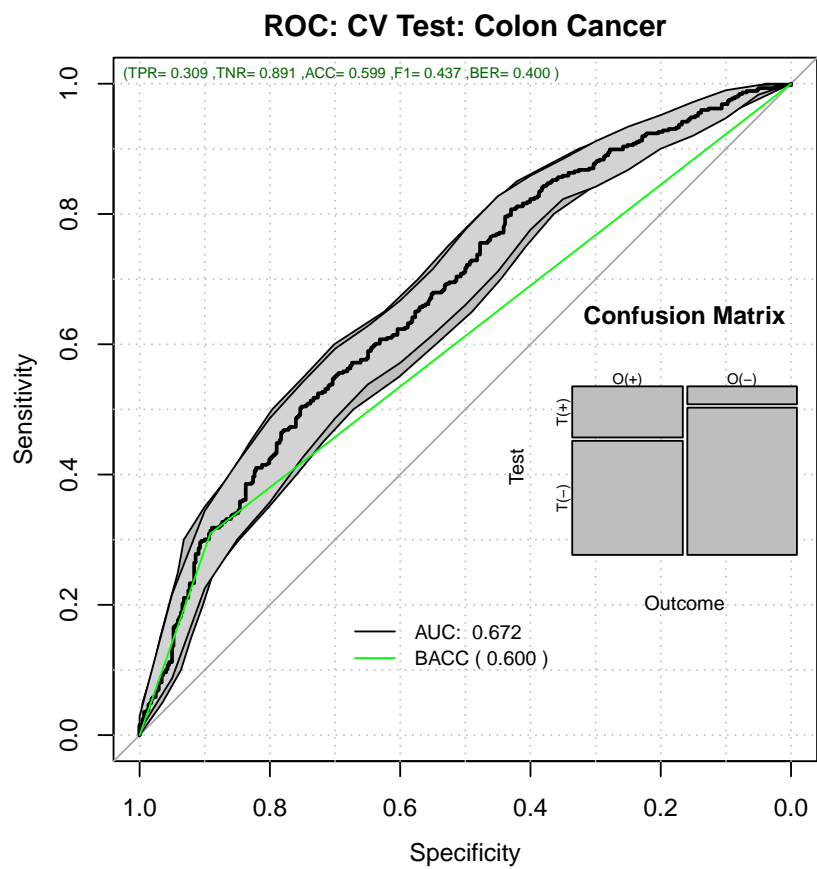


### Decision Curve Analysis: CV Test: Colon Cancer

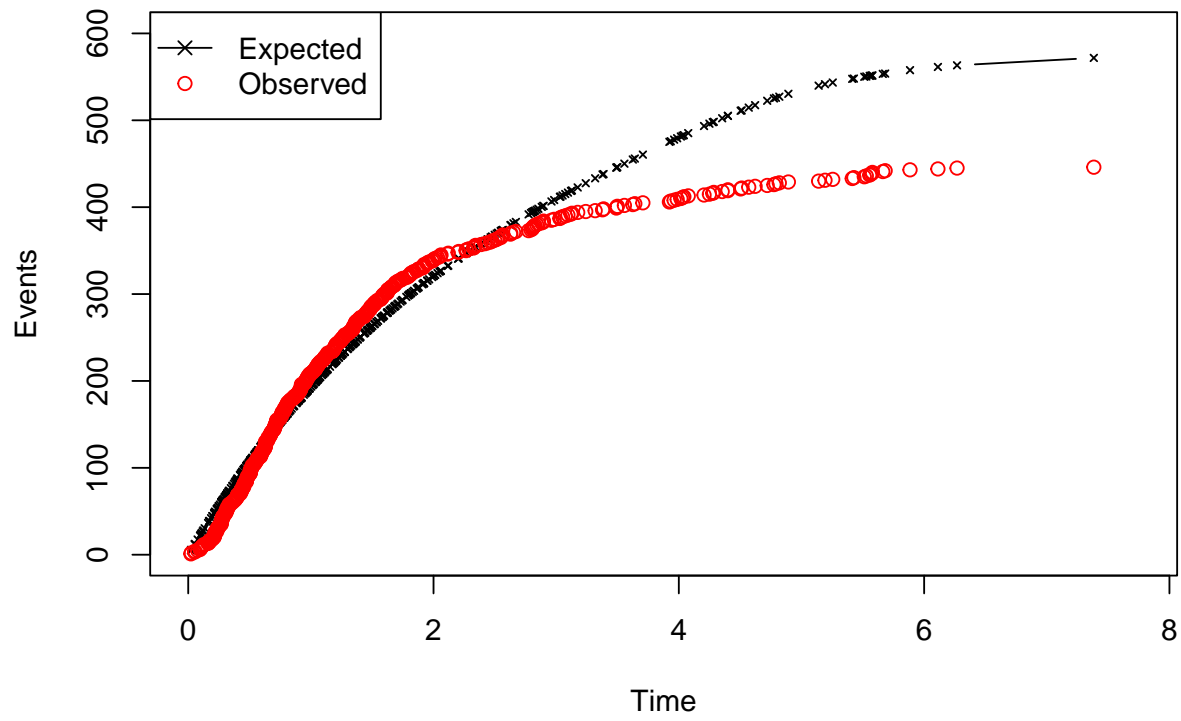


Relative Risk: CV Test: Colon Cancer



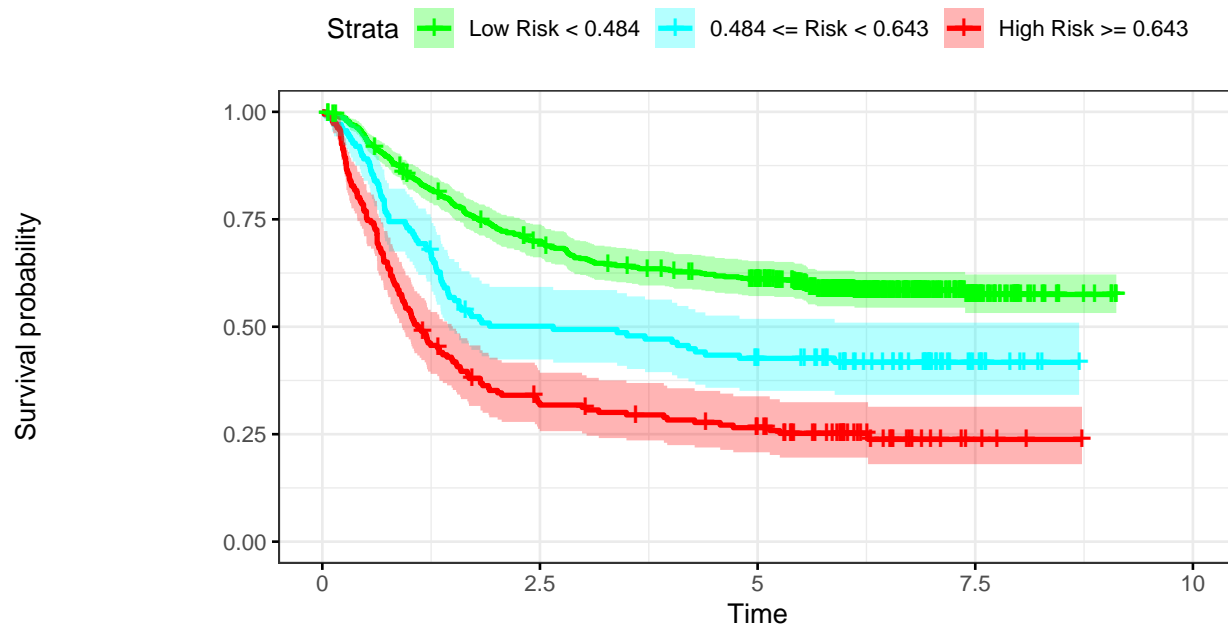


Time vs. Events: CV Test: Colon Cancer





## Kaplan–Meier: CV Test: Colon Cancer



### Number at risk

Low Risk < 0.484	565	384	322	47	0
0.484 <= Risk < 0.643	137	67	55	11	0
High Risk >= 0.643	186	56	42	4	0

### 1.4.1 CV Test Performance

```
pander::pander(t(rrAnalysisCVTest$keyPoints),caption="Threshold values")
```

Table 41: Threshold values (continued below)

	@:0.64303803374701	@:0.483598597950569	@:0.286109857686596
<b>Thr</b>	0.643	0.483	0.2864
<b>RR</b>	1.691	1.650	3.5767
<b>SEN</b>	0.309	0.487	0.9933
<b>SPE</b>	0.891	0.758	0.0407
<b>BACC</b>	0.600	0.622	0.5170

	@:0.15130951853827	@MAX_BACC	@MAX_RR	@SPE100
<b>Thr</b>	0.227	0.481	0.322	0.227
<b>RR</b>	1.000	1.682	2.322	1.000
<b>SEN</b>	1.000	0.502	0.960	1.000
<b>SPE</b>	0.000	0.753	0.138	0.000
<b>BACC</b>	0.500	0.628	0.549	0.500

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

Table 43: O/E Ratio

est	lower	upper
0.78	0.709	0.856

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

Table 44: O/E Ratio

mean	50%	2.5%	97.5%
0.958	0.958	0.944	0.972

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

Table 45: O/Acum Ratio

mean	50%	2.5%	97.5%
0.997	0.997	0.994	1

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

mean.C Index	median	lower	upper
0.645	0.644	0.619	0.67

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

Table 47: ROC AUC

est	lower	upper
0.672	0.637	0.707

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

Table 48: Sensitivity

est	lower	upper
0.309	0.267	0.355

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

Table 49: Specificity

est	lower	upper
0.891	0.859	0.919

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

Table 50: Probability Thresholds

90%	at_max_BACC	at_max_RR	atSPE100	at_max_BACC	at_max_RR	atSPE100
0.643	0.484	0.286	0.151	0.481	0.322	0.227

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

Table 51: Risk Ratio

est	lower	upper
1.7	1.51	1.91

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

Table 52: Logrank test Chisq = 108.473827 on 2 degrees of freedom,  
p = 0.000000

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
<b>class=0</b>	565	229	316.6	24.25	84.33
<b>class=1</b>	137	79	63.6	3.72	4.34
<b>class=2</b>	186	138	65.8	79.35	93.93