

Colon Cancer

Jose Tamez

2023-05-02

Contents

1	RRPlot and the Colon data set	1
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
311	310

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

0	1
131	136

1.2 Modeling

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

$$[++++-+ + + - + + + - + + + - + + + + + - + + + + + - + + + - + + + -] \dots$$

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

Table 3: Table continues below

	Estimate	lower	HR	upper	u.Accuracy
age_nodes	0.000423	1.000	1.000	1.001	0.607
node4	0.355981	1.260	1.428	1.617	0.601
rxLev_5FU_age	-0.004353	0.994	0.996	0.998	0.583
rxLev_5FU	-0.131937	0.824	0.876	0.932	0.583
age_node4	0.002206	1.001	1.002	1.003	0.601
extent	0.175181	1.047	1.191	1.356	0.536
nodes	0.009276	1.002	1.009	1.017	0.605
adhere	0.067181	1.008	1.069	1.135	0.536

Table 4: Table continues below

	r.Accuracy	full.Accuracy	u.AUC	r.AUC	full.AUC
age_nodes	0.520	0.606	0.607	0.520	0.606
node4	0.599	0.615	0.600	0.599	0.615
rxLev_5FU_age	0.607	0.615	0.583	0.607	0.615
rxLev_5FU	0.601	0.616	0.583	0.600	0.615
age_node4	0.608	0.616	0.600	0.608	0.615
extent	0.601	0.615	0.537	0.600	0.615
nodes	0.596	0.616	0.605	0.596	0.615
adhere	0.609	0.606	0.536	0.608	0.606

	IDI	NRI	z.IDI	z.NRI	Delta.AUC	Frequency
age_nodes	0.02968	0.409	5.80	5.51	0.08602	1.0
node4	0.03732	0.401	5.60	5.89	0.01623	1.0
rxLev_5FU_age	0.02491	0.333	4.58	4.49	0.00867	1.0
rxLev_5FU	0.02160	0.333	4.24	4.49	0.01510	1.0
age_node4	0.01309	0.355	3.92	5.37	0.00764	1.0
extent	0.00926	0.147	2.66	2.84	0.01517	1.0
nodes	0.00420	0.207	2.44	2.73	0.01945	1.0
adhere	0.00470	0.143	2.22	2.49	-0.00260	0.9

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))
```

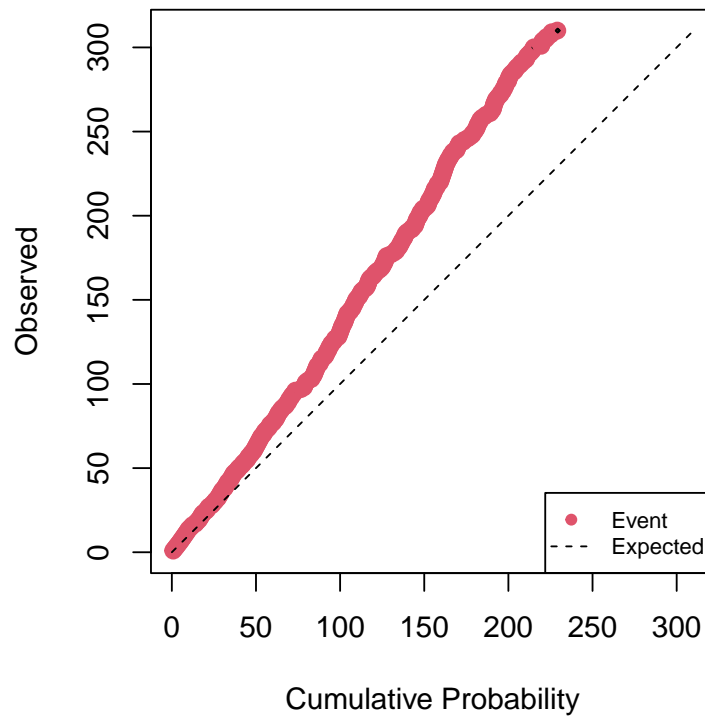
```

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

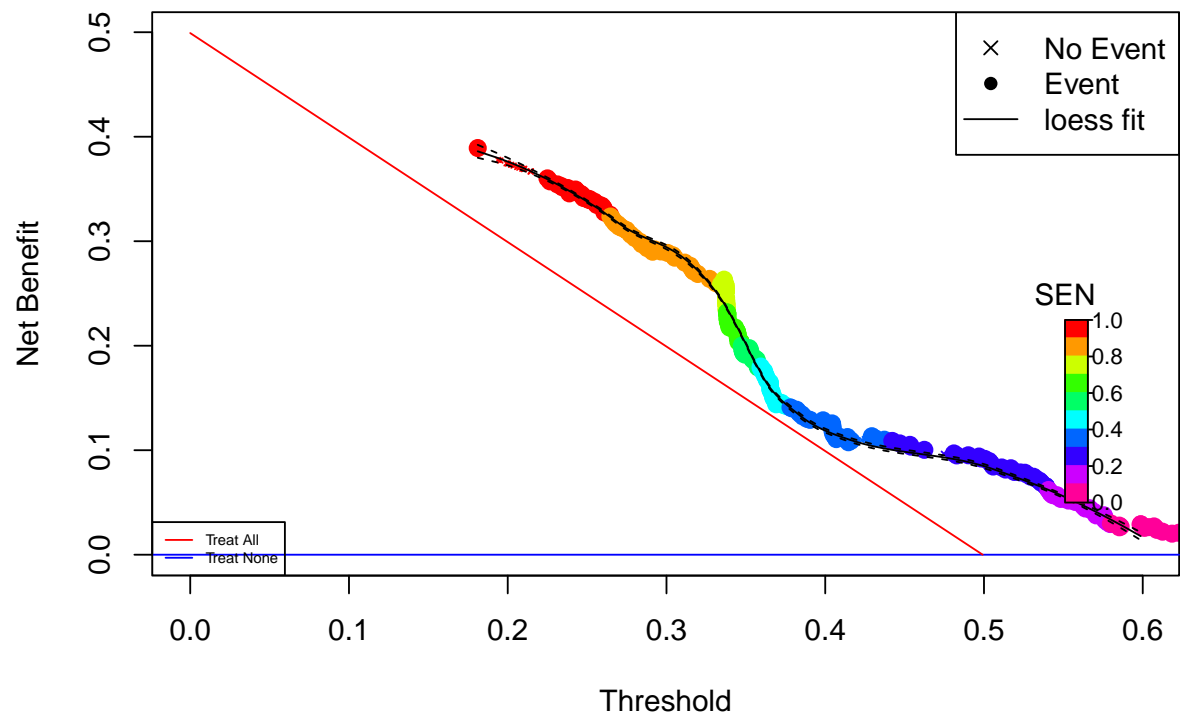
rrAnalysisTrain <- RRPlot(rdata, atProb=c(0.90),
  timetoEvent=dataColonTrain$time,
  title="Raw Train: Colon Cancer",
  ysurvlim=c(0.00, 1.0),
  riskTimeInterval=timeinterval)

```

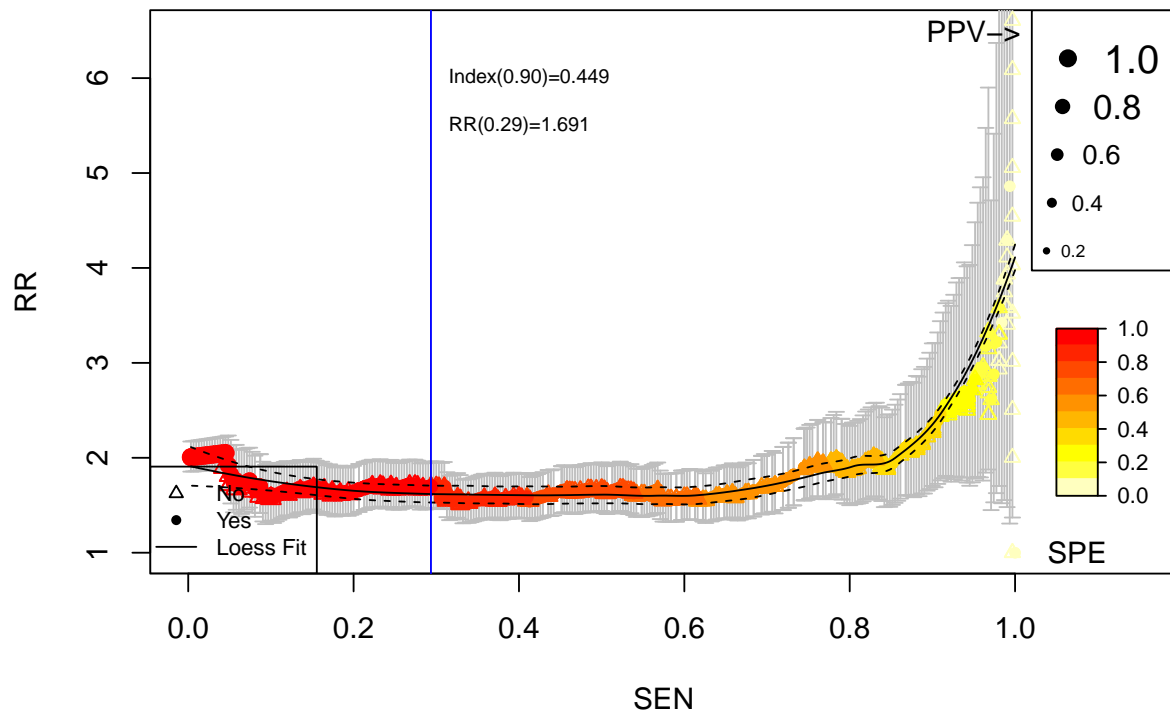
Cumulative vs. Observed: Raw Train: Colon Cancer

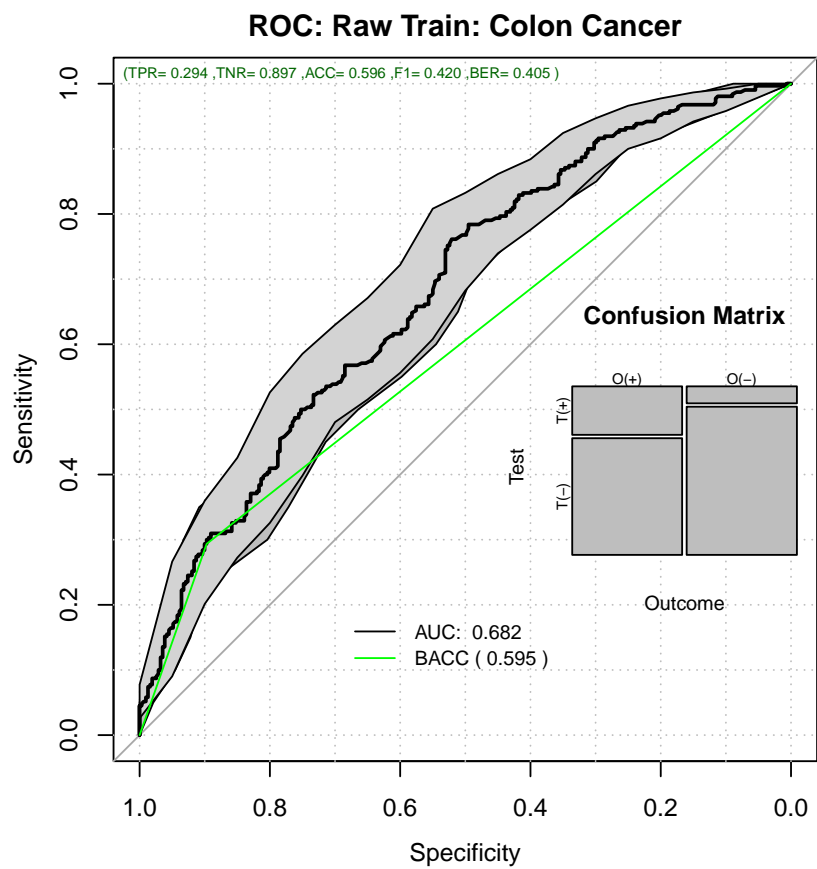


Decision Curve Analysis: Raw Train: Colon Cancer

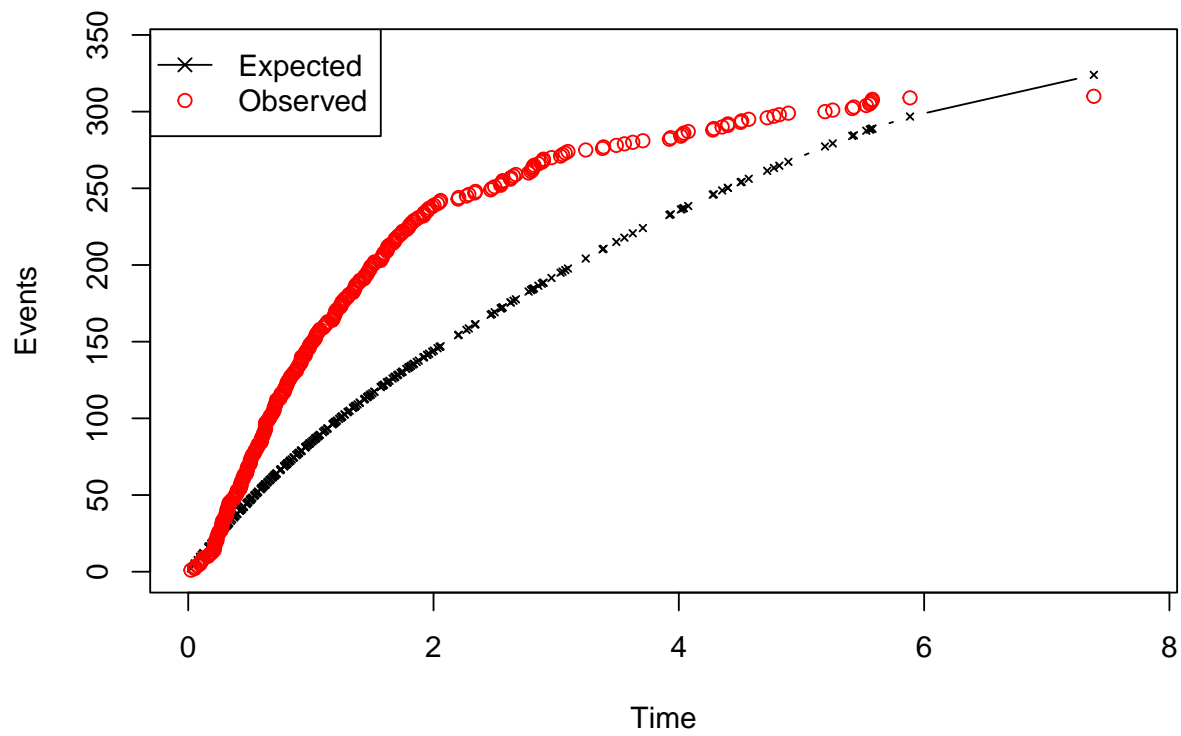


Relative Risk: Raw Train: Colon Cancer

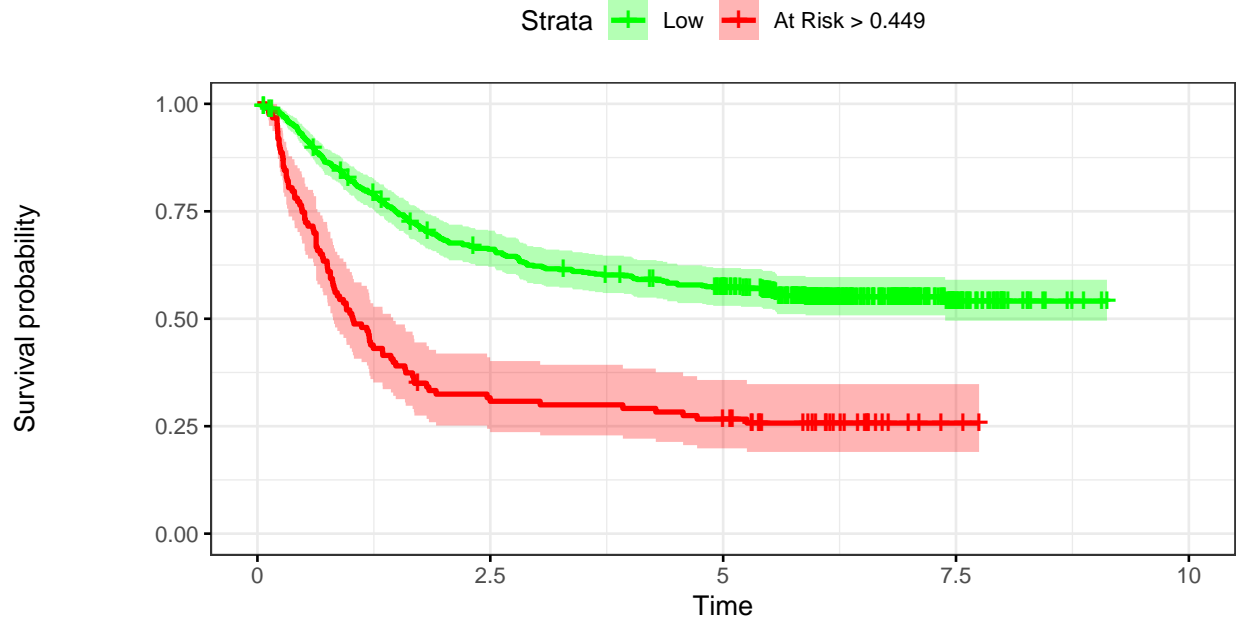




Time vs. Events: Raw Train: Colon Cancer



Kaplan–Meier: Raw Train: Colon Cancer



Number at risk

Low	498	320	266	42	0
At Risk > 0.449	123	37	31	2	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	p(0.5)
Thr	0.449	0.337	0.238	0.18110	0.501
RR	1.682	1.955	3.582	1.00000	1.717
SEN	0.294	0.761	0.981	1.00000	0.268
SPE	0.897	0.521	0.113	0.00643	0.916
BACC	0.595	0.641	0.547	0.50322	0.592

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

Table 7: O/E Ratio

O/E	Low	Upper	p.value
0.957	0.853	1.07	0.453

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

Table 8: O/E Mean

mean	50%	2.5%	97.5%
1.57	1.57	1.54	1.59

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

Table 9: O/Acum Mean

mean	50%	2.5%	97.5%
1.35	1.35	1.35	1.36

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

mean.C Index	median	lower	upper
0.656	0.656	0.627	0.69

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

Table 11: ROC AUC

est	lower	upper
0.682	0.641	0.724

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

Table 12: Sensitivity

est	lower	upper
0.294	0.243	0.348

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

Table 13: Specificity

est	lower	upper
0.897	0.858	0.929

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

Table 14: Probability Thresholds

90%	at_max_BACC	at_max_RR	atSPE100	at_0.5
0.449	0.337	0.238	0.181	0.5

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

Table 15: Risk Ratio

est	lower	upper
1.69	1.46	1.95

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

Table 16: Logrank test Chisq = 63.453880 on 1 degrees of freedom,
p = 0.000000

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	498	219	267.2	8.69	63.5
class=1	123	91	42.8	54.21	63.5

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
0.677	1.52	2.87

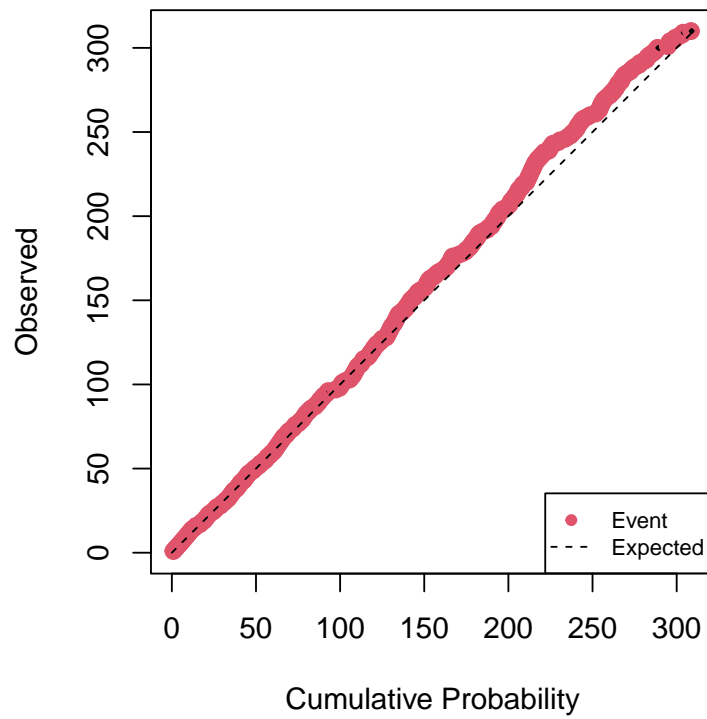
1.3.4 The RRplot() of the calibrated model

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

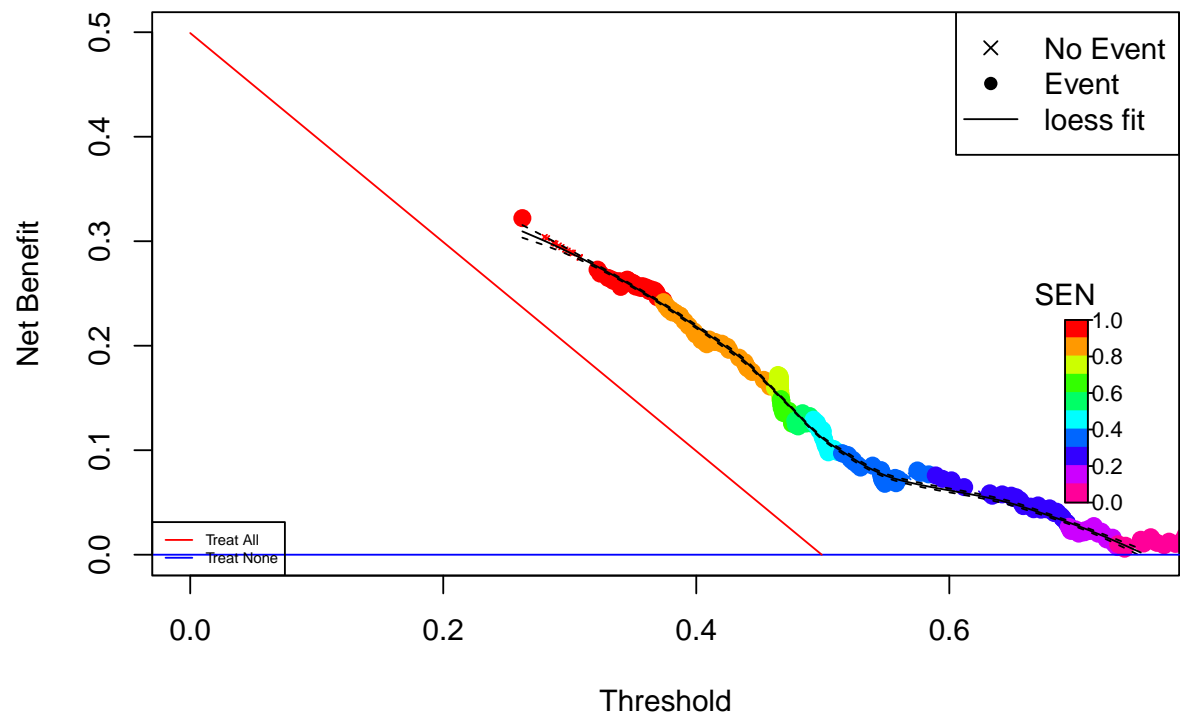
rdata <- cbind(dataColonTrain$status,calprob$prob)

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

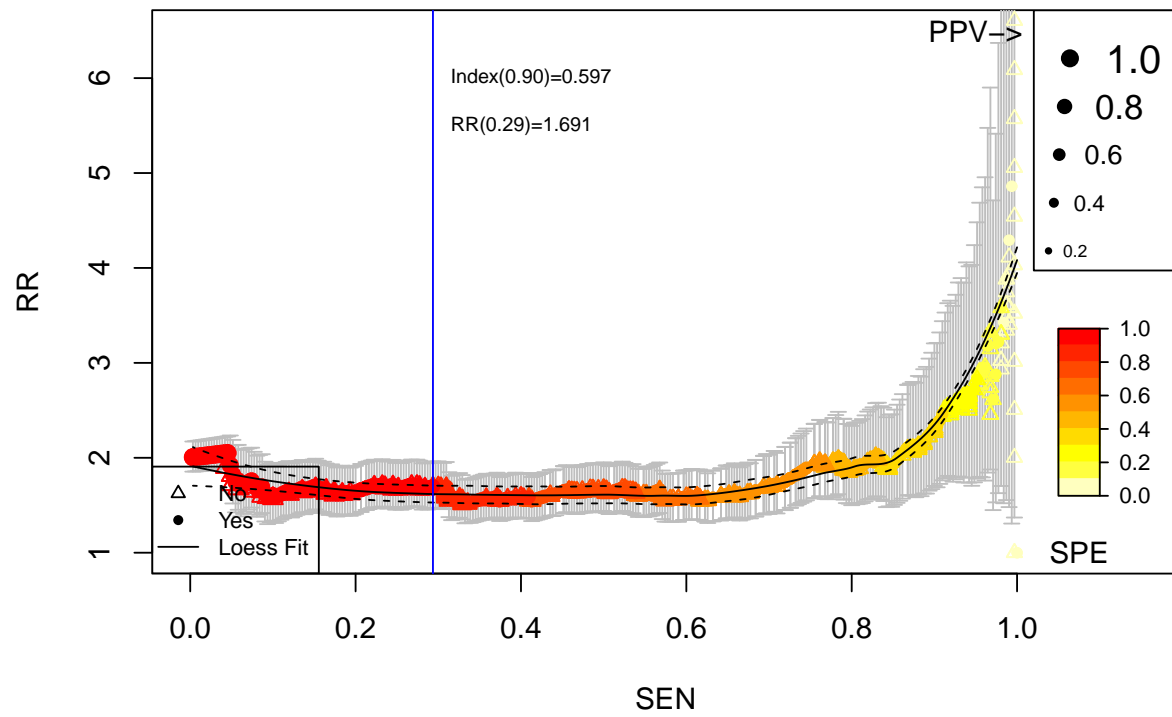
Cumulative vs. Observed: Calibrated Train: Colon

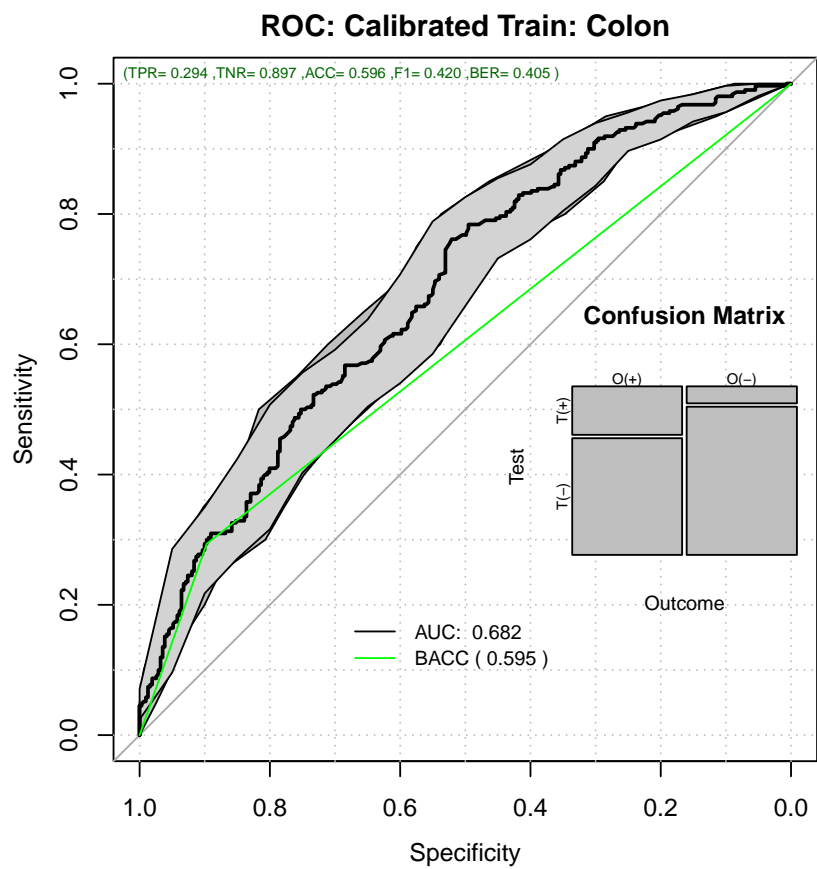


Decision Curve Analysis: Calibrated Train: Colon

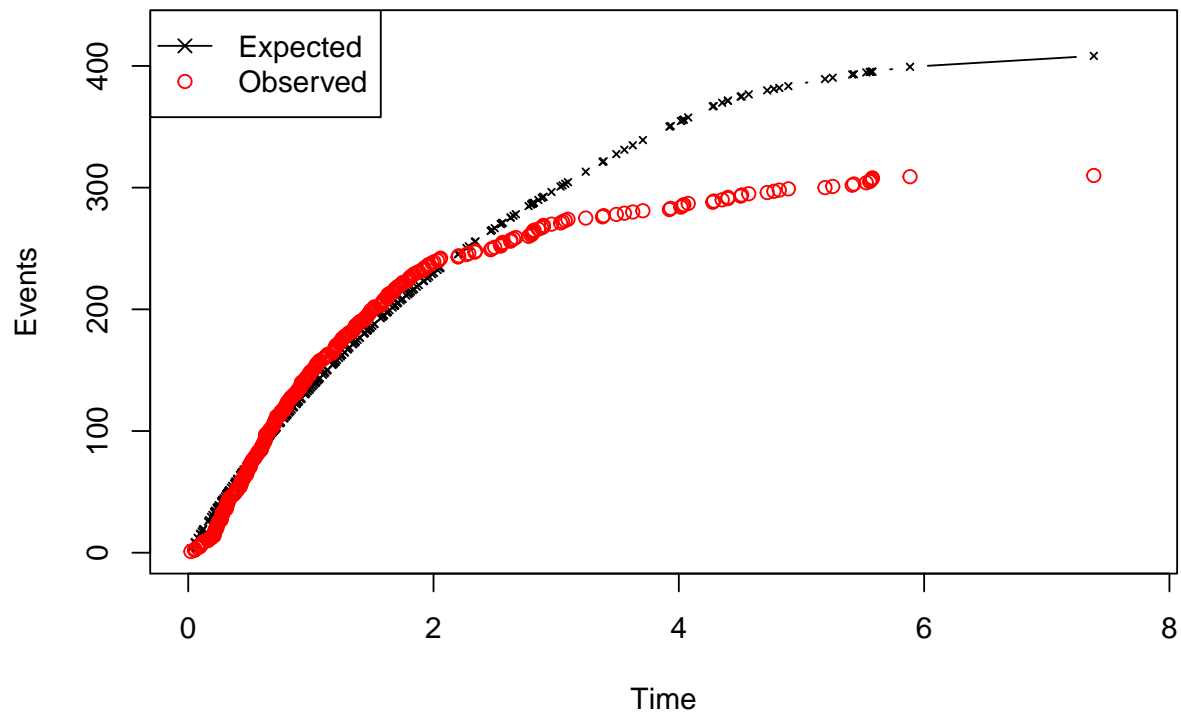


Relative Risk: Calibrated Train: Colon

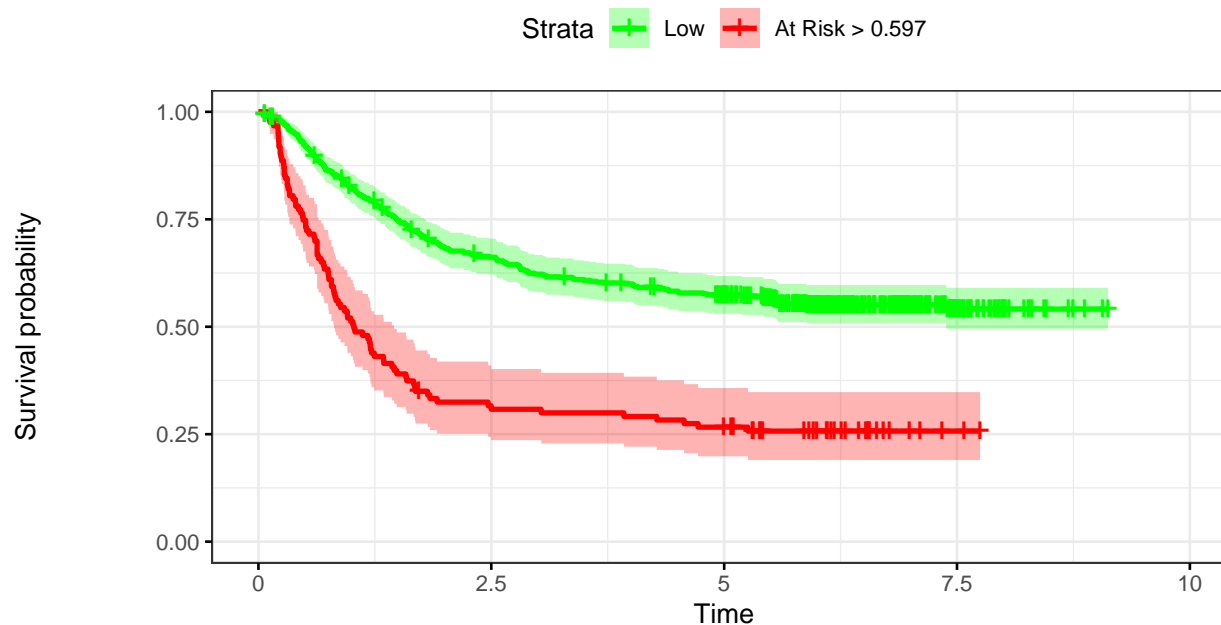




Time vs. Events: Calibrated Train: Colon



Kaplan–Meier: Calibrated Train: Colon



Number at risk

Low	498	320	266	42	0
At Risk > 0.597	123	37	31	2	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	p(0.5)
Thr	0.597	0.466	0.339	0.26253	0.500
RR	1.682	1.955	3.582	1.00000	1.619
SEN	0.294	0.761	0.981	1.00000	0.442
SPE	0.897	0.521	0.113	0.00643	0.785
BACC	0.595	0.641	0.547	0.50322	0.613

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

Table 19: O/E Ratio

O/E	Low	Upper	p.value
0.759	0.677	0.849	4.37e-07

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

Table 20: O/E Mean

mean	50%	2.5%	97.5%
0.982	0.982	0.969	0.995

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

Table 21: O/Acum Mean

mean	50%	2.5%	97.5%
1.03	1.03	1.03	1.03

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

mean.C Index	median	lower	upper
0.656	0.656	0.625	0.686

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

Table 23: ROC AUC

est	lower	upper
0.682	0.641	0.724

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

Table 24: Sensitivity

est	lower	upper
0.294	0.243	0.348

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

Table 25: Specificity

est	lower	upper
0.897	0.858	0.929

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

Table 26: Probability Thresholds

90%	at_max_BACC	at_max_RR	atSPE100	at_0.5
0.597	0.466	0.339	0.263	0.5

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

Table 27: Risk Ratio

est	lower	upper
1.69	1.46	1.95

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

Table 28: Logrank test Chisq = 63.453880 on 1 degrees of freedom,
p = 0.000000

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	498	219	267.2	8.69	63.5
class=1	123	91	42.8	54.21	63.5

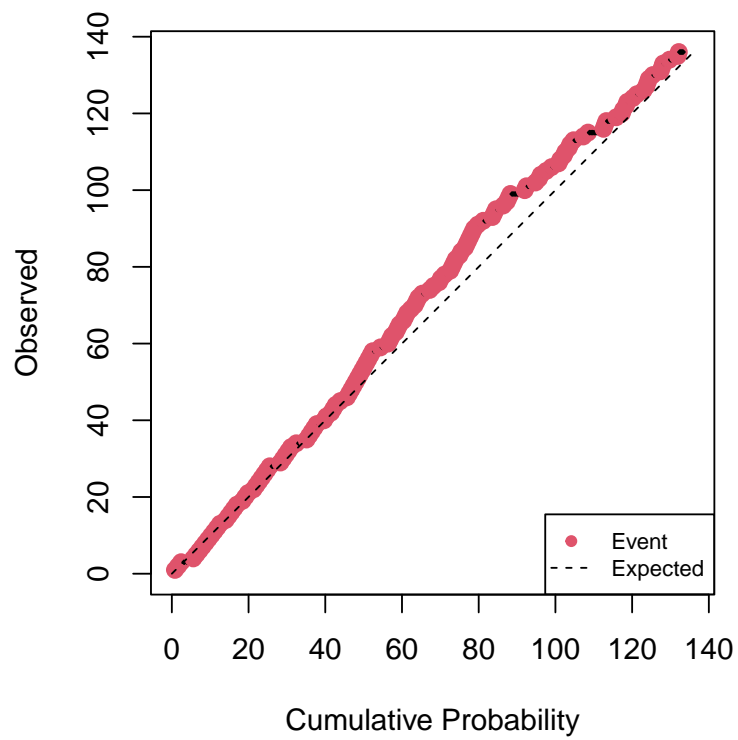
1.3.6 Evaluating on the test set

The calibrated h0 and timeinterval were estimated on the training set

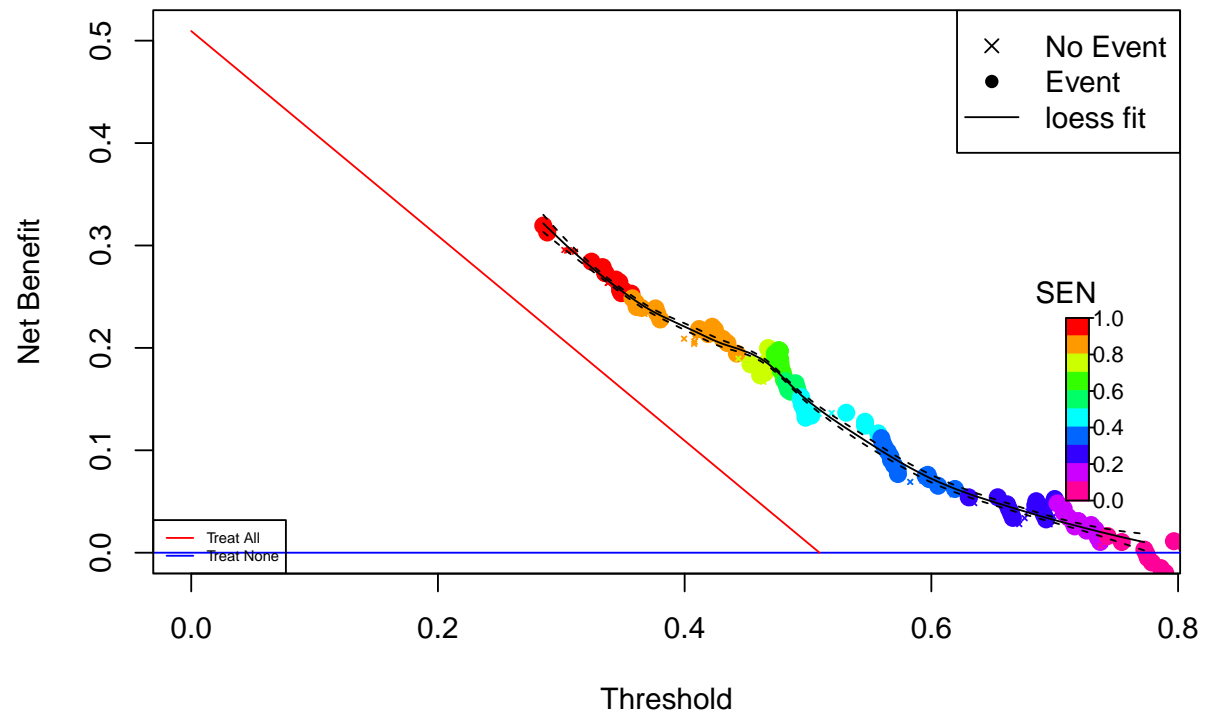
```
index <- predict(ml,dataColonTest)
rdata <- cbind(dataColonTest$status,ppoisGzero(index,h0))

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

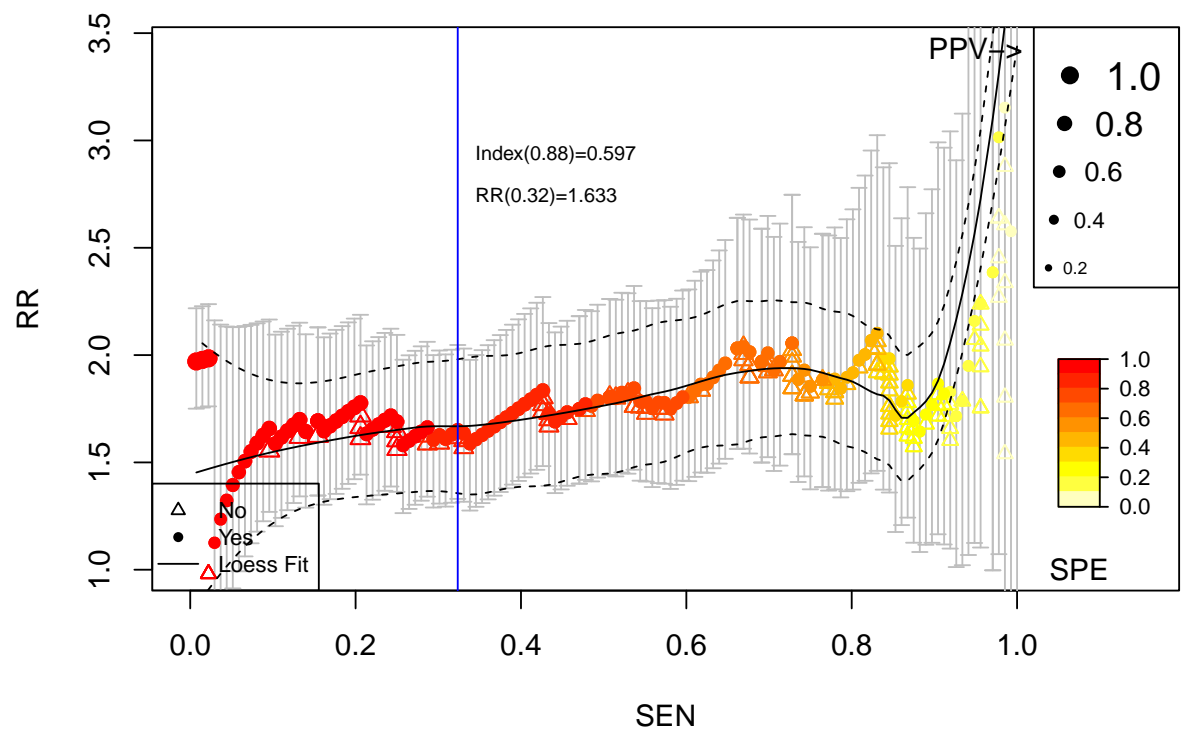
Cumulative vs. Observed: Test: Colon Cancer

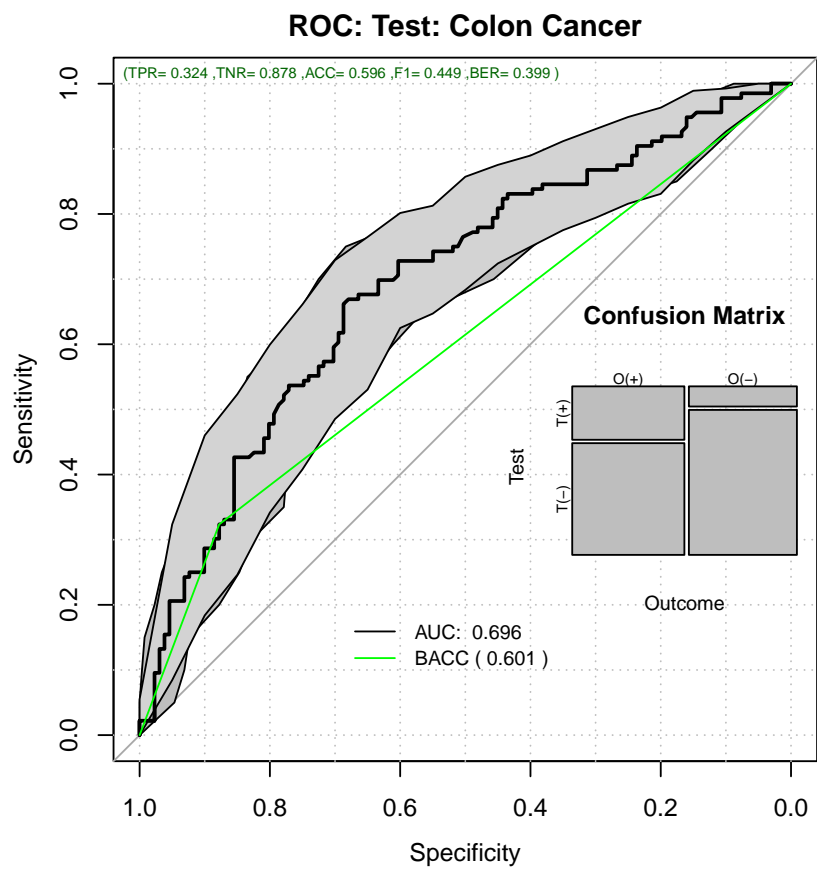


Decision Curve Analysis: Test: Colon Cancer

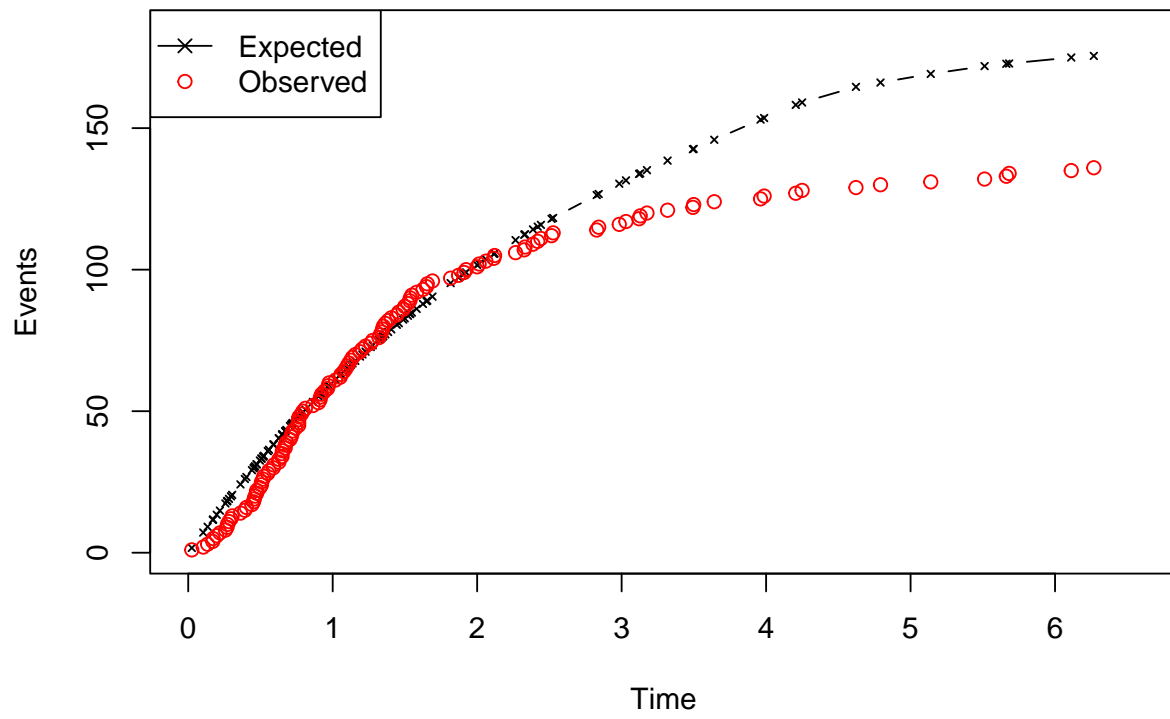


Relative Risk: Test: Colon Cancer

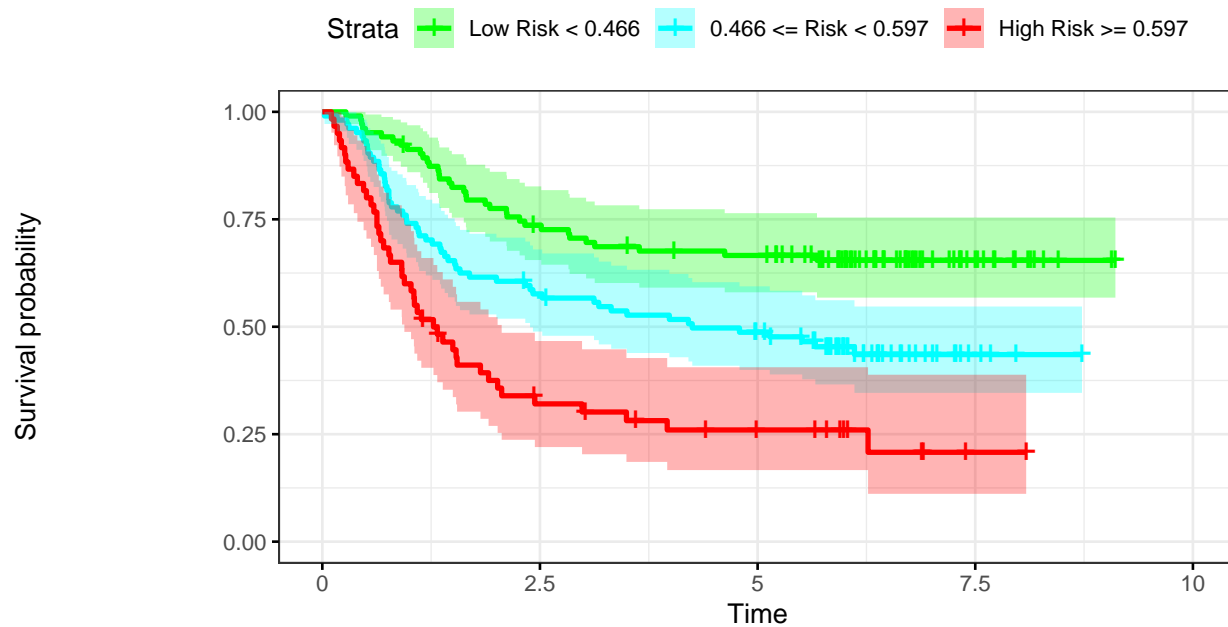




Time vs. Events: Test: Colon Cancer



Kaplan–Meier: Test: Colon Cancer



Number at risk

Low Risk < 0.466	103	74	65	13	0
0.466 <= Risk < 0.597	104	59	47	4	0
High Risk >= 0.597	60	17	10	1	0

1.3.7 Test Performance

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

Table 29: Threshold values (continued below)

	@:0.596543671607801	@:0.465714066985802	@:0.338856488077808
Thr	0.597	0.466	0.339
RR	1.615	1.812	1.946
SEN	0.324	0.743	0.956
SPE	0.870	0.519	0.122
BACC	0.597	0.631	0.539

	@:0.262529412124052	@:0.5	@MAX_BACC	@MAX_RR	@SPE100	p(0.5)
Thr	0.2854	0.499	0.477	0.334	0.2854	0.499
RR	20.6844	1.669	2.031	3.015	20.6844	1.669
SEN	1.0000	0.434	0.662	0.978	1.0000	0.434
SPE	0.0305	0.809	0.687	0.107	0.0305	0.809
BACC	0.5153	0.621	0.674	0.542	0.5153	0.621

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

Table 31: O/E Ratio

O/E	Low	Upper	p.value
0.775	0.65	0.917	0.00222

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

Table 32: O/E Mean

mean	50%	2.5%	97.5%
0.916	0.916	0.895	0.937

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

Table 33: O/Acum Mean

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	0.614	0.698

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

Table 35: ROC AUC

est	lower	upper
0.696	0.632	0.759

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

Table 36: Sensitivity

est	lower	upper
0.324	0.246	0.409

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

Table 37: Specificity

est	lower	upper
0.878	0.809	0.929

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

Table 38: Probability Thresholds (continued below)

90%	at_max_BACC	at_max_RR	atSPE100	at_0.5	at_max_BACC	at_max_RR
0.597	0.466	0.339	0.263	0.5	0.477	0.334

atSPE100	at_0.5
0.285	0.5

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

Table 40: Risk Ratio

est	lower	upper
1.63	1.31	2.03

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

Table 41: Logrank test Chisq = 36.562917 on 2 degrees of freedom,
p = 0.000000

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	103	35	62.1	11.830	21.942
class=1	104	57	52.5	0.389	0.634
class=2	60	44	21.4	23.831	28.634

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
)
```

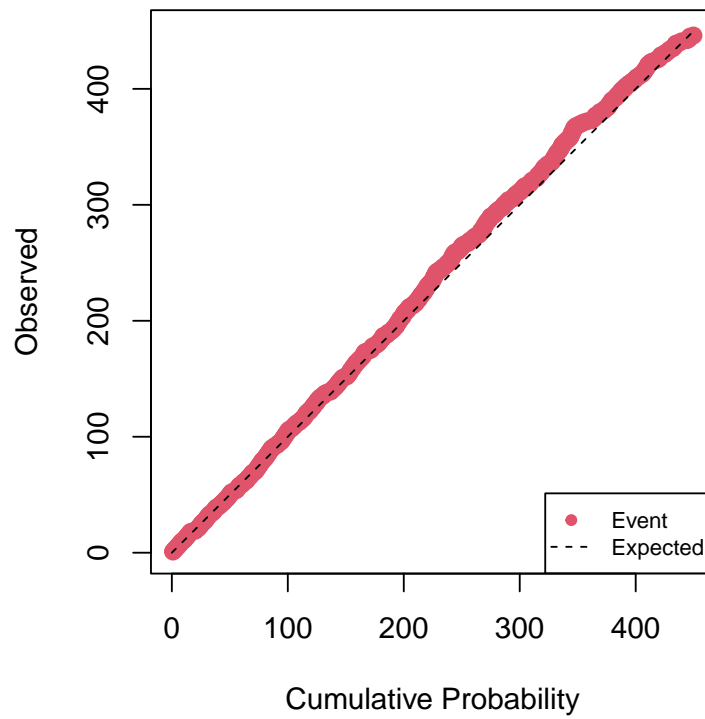
```
stp <- rcv$survTestPredictions
stp <- stp[!is.na(stp[,4]),]

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)

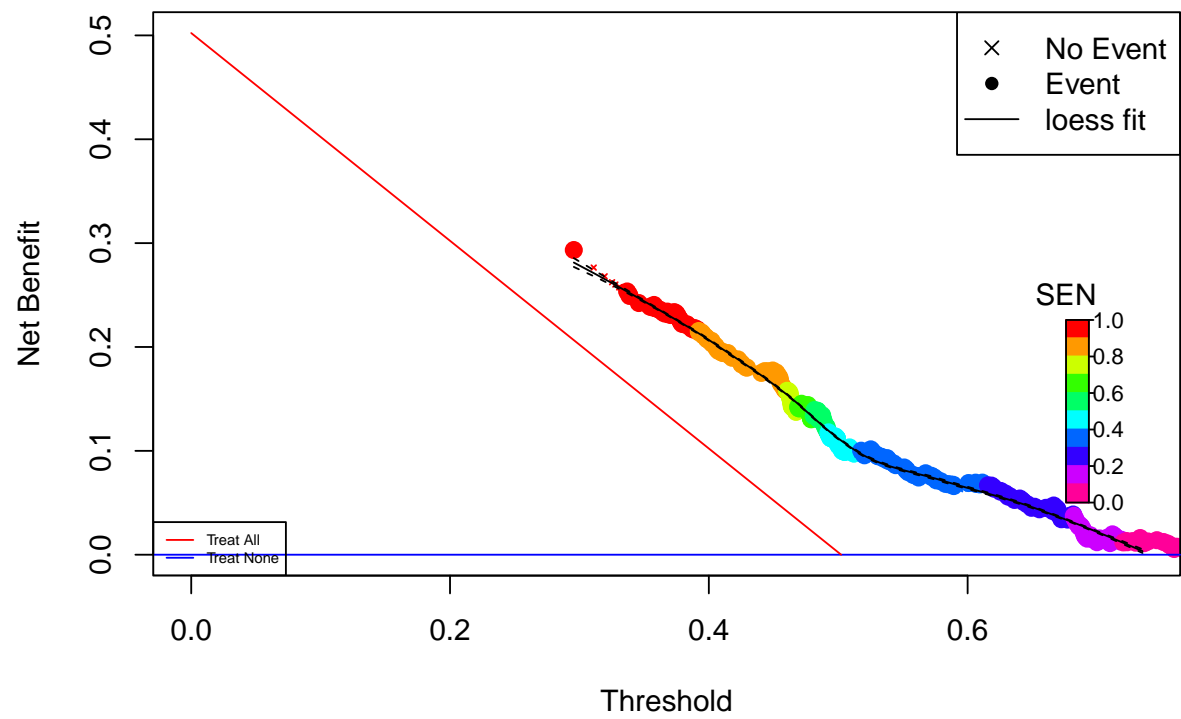
rdatacv <- cbind(status,prob)
rownames(rdatacv) <- bbx$names
names(times) <- bbx$names

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

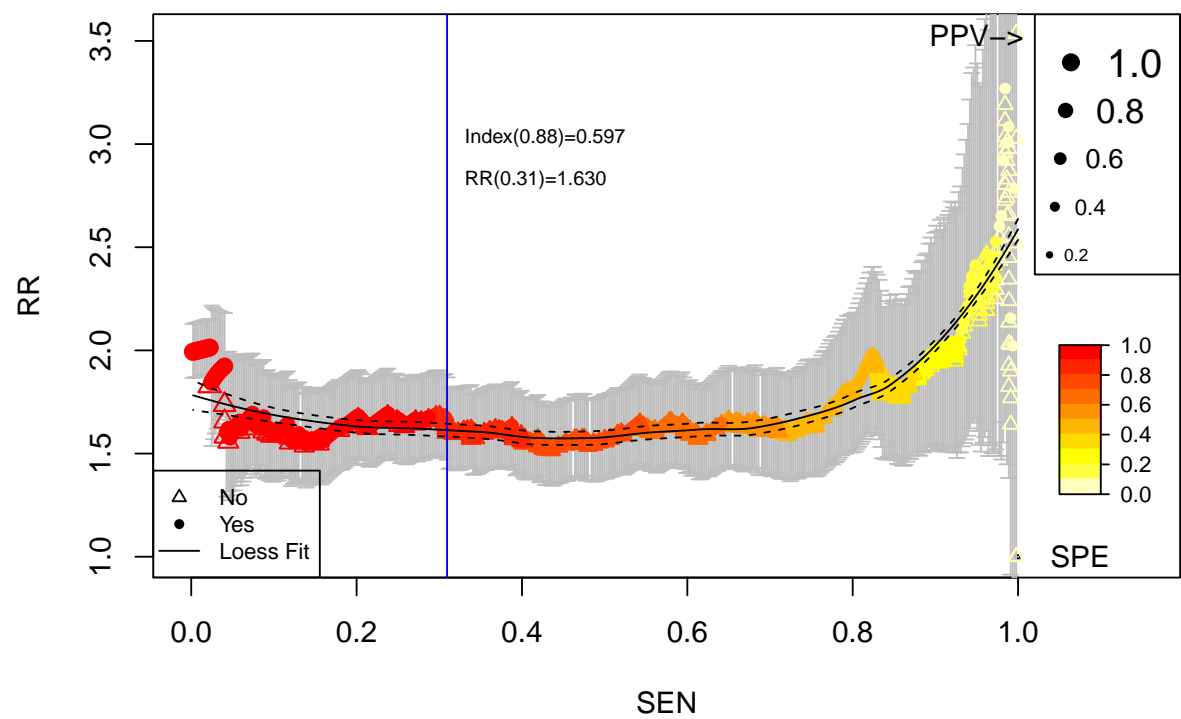
Cumulative vs. Observed: CV Test: Colon Cancer

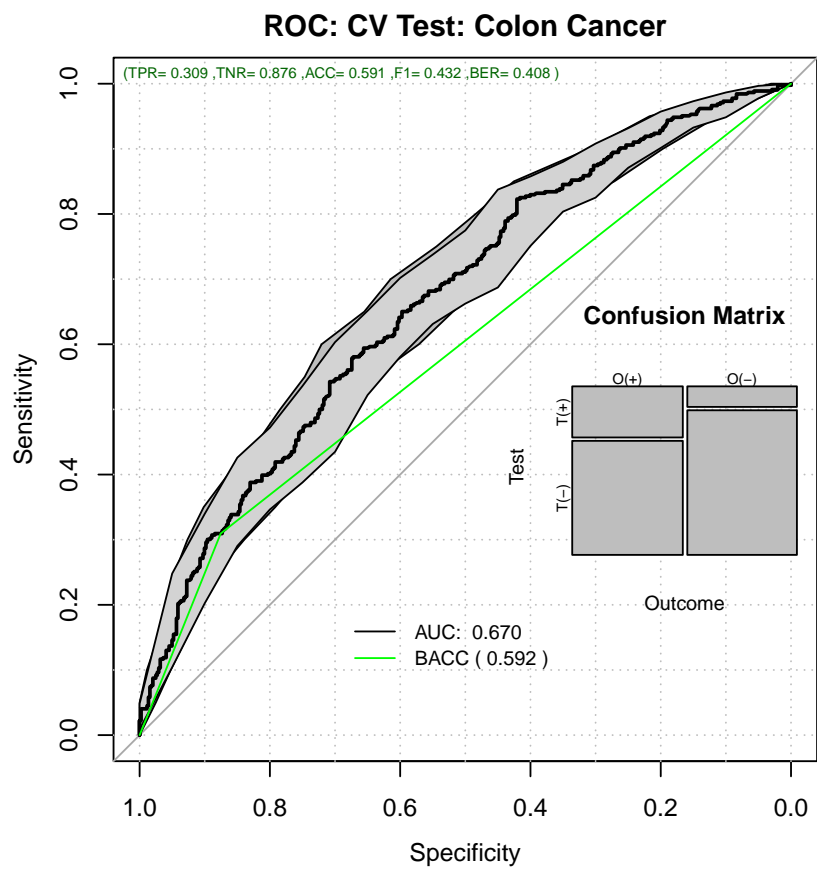


Decision Curve Analysis: CV Test: Colon Cancer

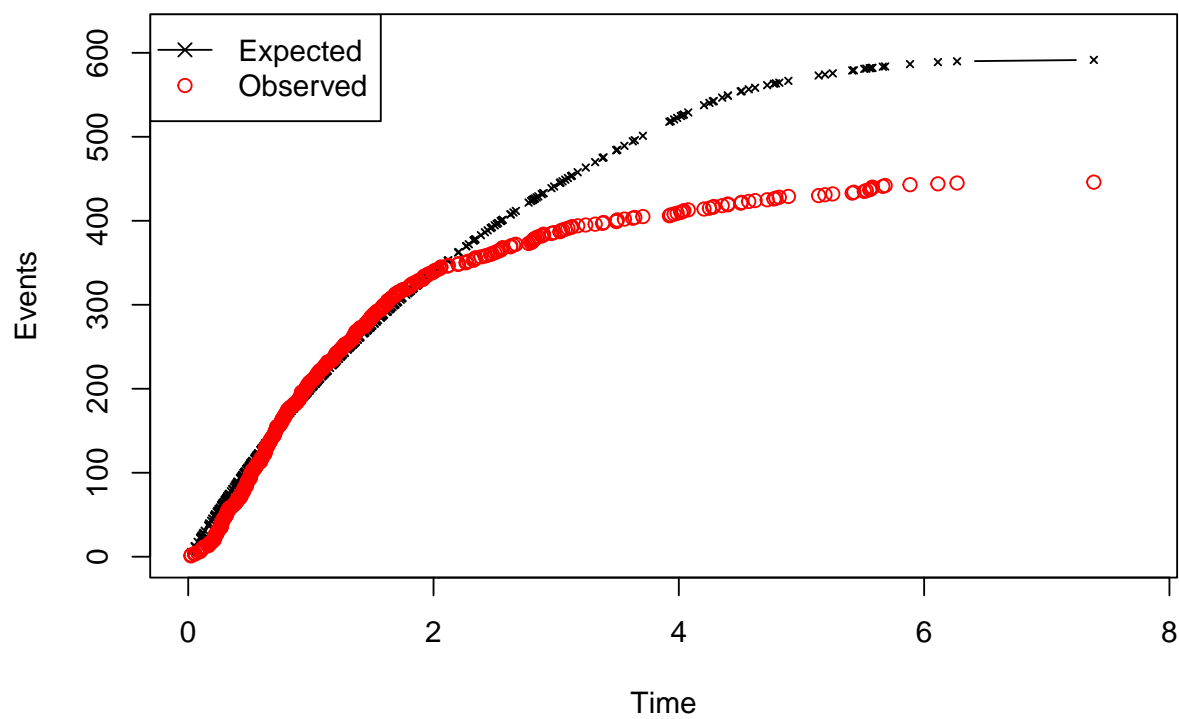


Relative Risk: CV Test: Colon Cancer

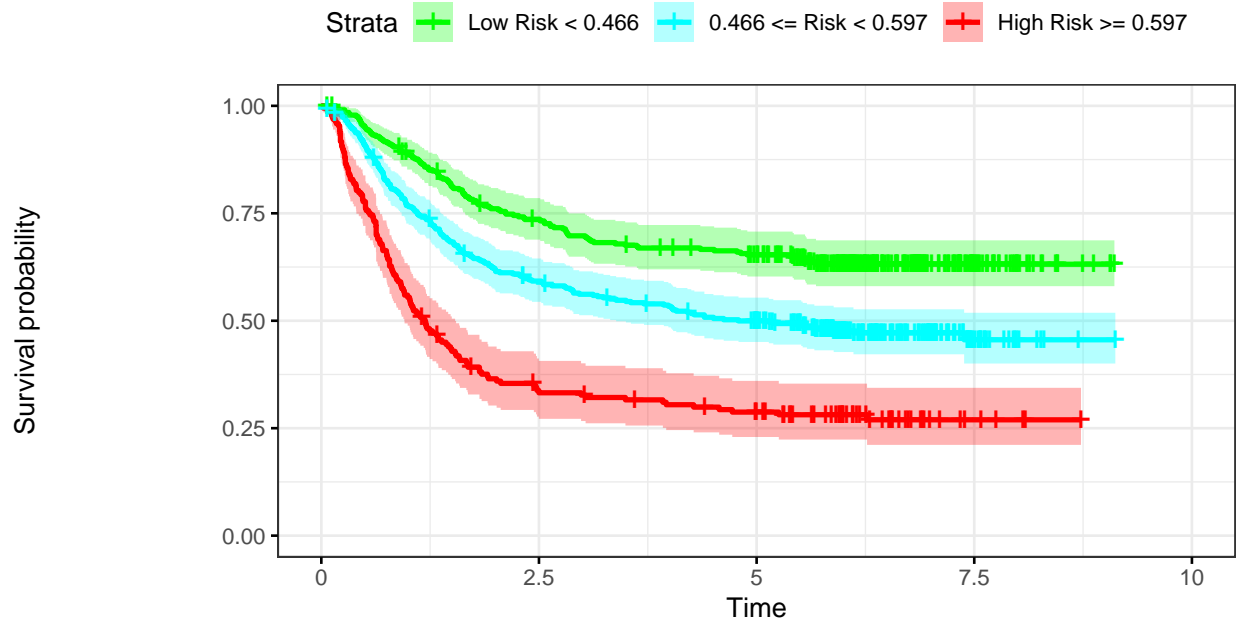




Time vs. Events: CV Test: Colon Cancer



Kaplan–Meier: CV Test: Colon Cancer



Number at risk

Low Risk < 0.466	326	233	200	36	0
0.466 <= Risk < 0.597	369	213	171	21	0
High Risk >= 0.597	193	61	48	5	0

1.4.1 CV Test Performance

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

Table 42: Threshold values (continued below)

	@:0.596543671607801	@:0.465714066985802	@:0.338856488077808
Thr	0.597	0.466	0.3387
RR	1.613	1.623	2.0228
SEN	0.309	0.738	0.9933
SPE	0.876	0.471	0.0204
BACC	0.592	0.604	0.5068

	@:0.262529412124052	@:0.5	@MAX_BACC	@MAX_RR	@SPE100	p(0.5)
Thr	0.296	0.500	0.484	0.362	0.296	0.500
RR	1.000	1.553	1.659	2.527	1.000	1.553
SEN	1.000	0.451	0.578	0.973	1.000	0.451
SPE	0.000	0.760	0.674	0.104	0.000	0.760
BACC	0.500	0.605	0.626	0.539	0.500	0.605

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

Table 44: O/E Ratio

O/E	Low	Upper	p.value
0.754	0.686	0.827	4.53e-10

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

Table 45: O/E Mean

mean	50%	2.5%	97.5%
0.937	0.937	0.926	0.948

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

Table 46: O/Acum Mean

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
0.644	0.644	0.618	0.669

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

Table 48: ROC AUC

est	lower	upper
0.67	0.634	0.705

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

Table 49: Sensitivity

est	lower	upper
0.309	0.267	0.355

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

Table 50: Specificity

est	lower	upper
0.876	0.841	0.905

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

Table 51: Probability Thresholds (continued below)

90%	at_max_BACC	at_max_RR	atSPE100	at_0.5	at_max_BACC	at_max_RR
0.597	0.466	0.339	0.263	0.5	0.484	0.362

atSPE100	at_0.5
0.296	0.5

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

Table 53: Risk Ratio

est	lower	upper
1.63	1.44	1.84

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

Table 54: Logrank test Chisq = 94.306437 on 2 degrees of freedom,
p = 0.000000

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	326	117	189.2	27.5255	48.037
class=1	369	191	186.8	0.0963	0.166
class=2	193	138	70.1	65.8142	78.707