

Lung Cancer

Jose Tamez

2023-04-26

Contents

0.1	Modeling	2
0.2	Cox Model Performance	3
0.3	Cross-Validation	19

```
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 <- model.frame(Surv(time,status)~*.,lung,na.action=NULL)
colnames(data) <-str_replace_all(colnames(data),":","_")
colnames(data) <-str_replace_all(colnames(data),"\\.", "_")

data$inst <- NULL
```

```
data$`Surv(time, status)` <- NULL
dataLung <- cbind(time=lung$time/365, status=lung$status-1, data)
dataLung <- dataLung[complete.cases(dataLung),]
pander::pander(table(dataLung$status))
```

0	1
47	121

```
pander::pander(summary(dataLung$time))
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.0137	0.479	0.736	0.849	1.14	2.8

0.1 Modeling

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

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

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

Table 3: Table continues below

	Estimate	lower	HR	upper	u.Accuracy	r.Accuracy
ph_ecog	4.32e-01	1.194	1.541	1.988	0.679	0.649
sex	-4.59e-01	0.456	0.632	0.876	0.649	0.679
pat_karno	-1.77e-03	0.997	0.998	1.000	0.506	0.720
ph_karno	-4.64e-07	1.000	1.000	1.000	0.577	0.720
age	4.57e-08	1.000	1.000	1.000	0.565	0.720

Table 4: Table continues below

	full.Accuracy	u.AUC	r.AUC	full.AUC	IDI	NRI
ph_ecog	0.601	0.601	0.620	0.600	0.0449	0.405
sex	0.601	0.620	0.601	0.600	0.0285	0.478
pat_karno	0.506	0.585	0.500	0.585	0.0292	0.342
ph_karno	0.577	0.570	0.500	0.570	0.0143	0.280
age	0.565	0.549	0.500	0.549	0.0162	0.195

	z.IDI	z.NRI	Delta.AUC	Frequency
ph_ecog	3.33	2.48	-0.02005	1.0
sex	2.76	2.85	-0.00167	1.0
pat_karno	2.44	2.24	0.08546	1.0
ph_karno	2.22	1.64	0.06998	0.8
age	1.97	1.14	0.04871	0.1

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

```
timeinterval <- 2*mean(subset(dataLung,status==1)$time)

h0 <- sum(dataLung$status & dataLung$time <= timeinterval)
h0 <- h0/sum((dataLung$time > timeinterval) | (dataLung$status==1))
pander::pander(t(c(h0=h0,timeinterval=timeinterval)),caption="Initial Parameters")
```

Table 6: Initial Parameters

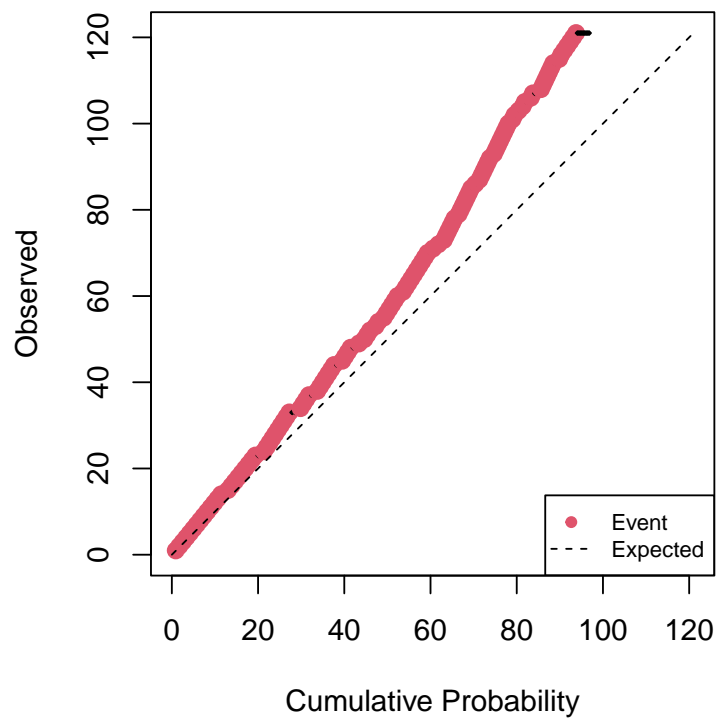
h0	timeinterval
0.85	1.58

```
index <- predict(ml,dataLung)

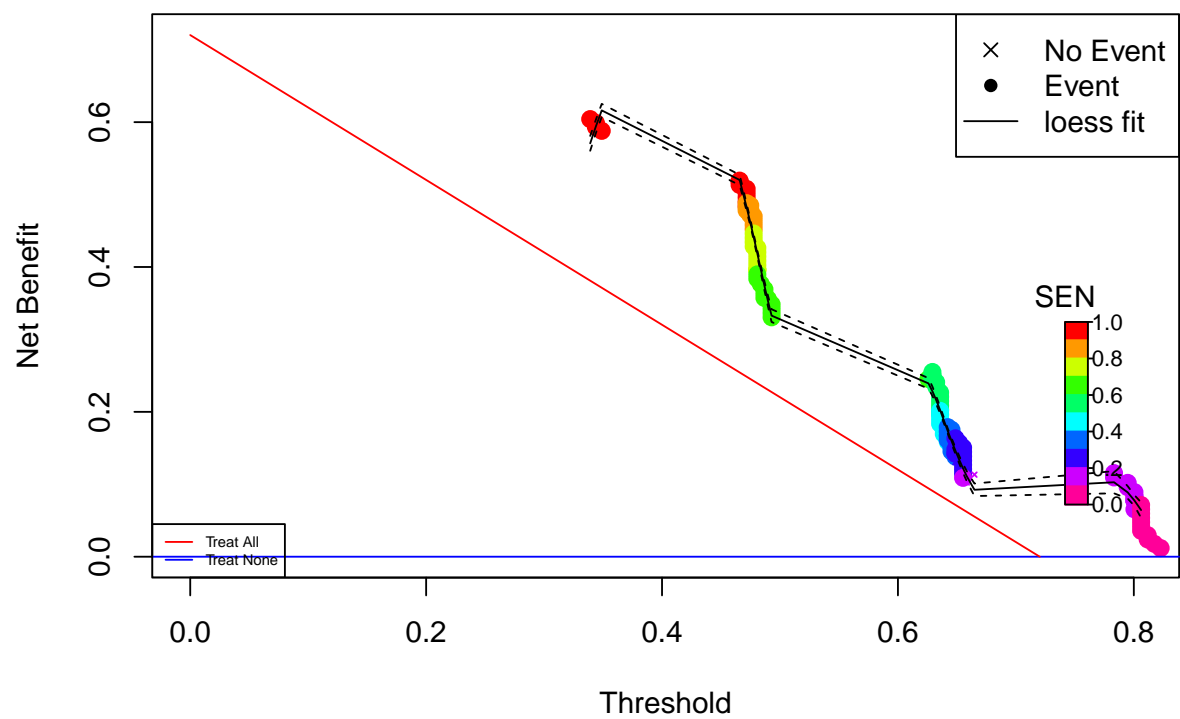
rdata <- cbind(dataLung$status,ppoisGzero(index,h0))

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

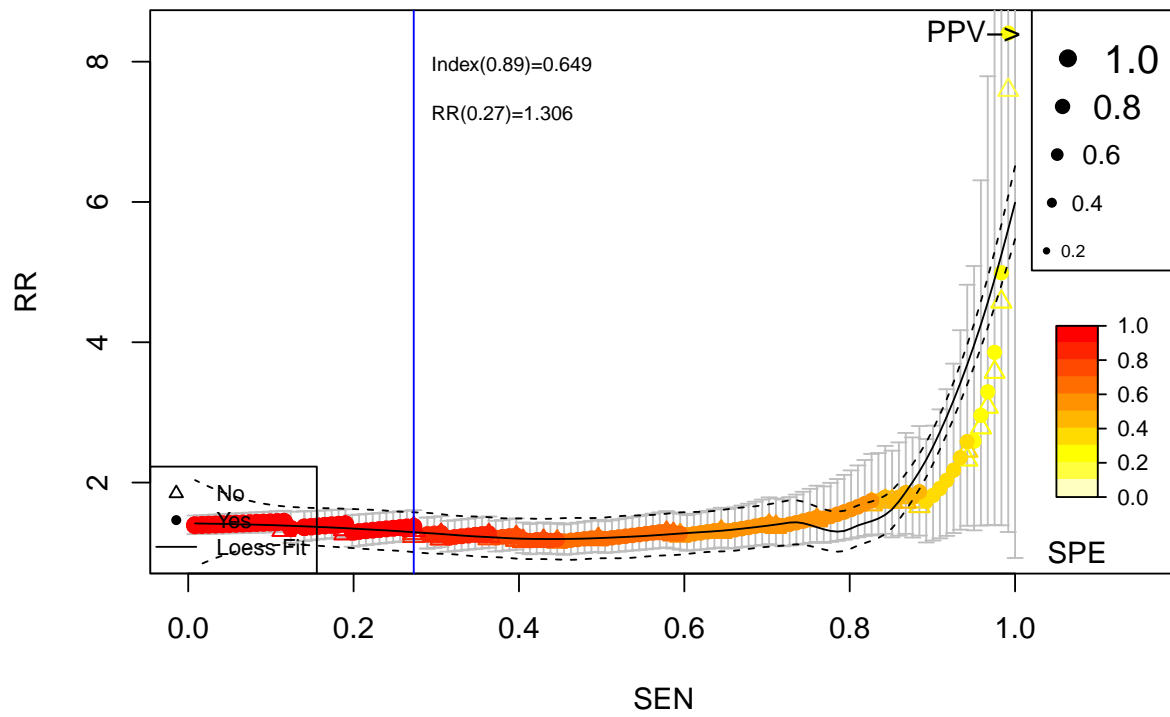
Cumulative vs. Observed: Raw Train: Lung Cancer

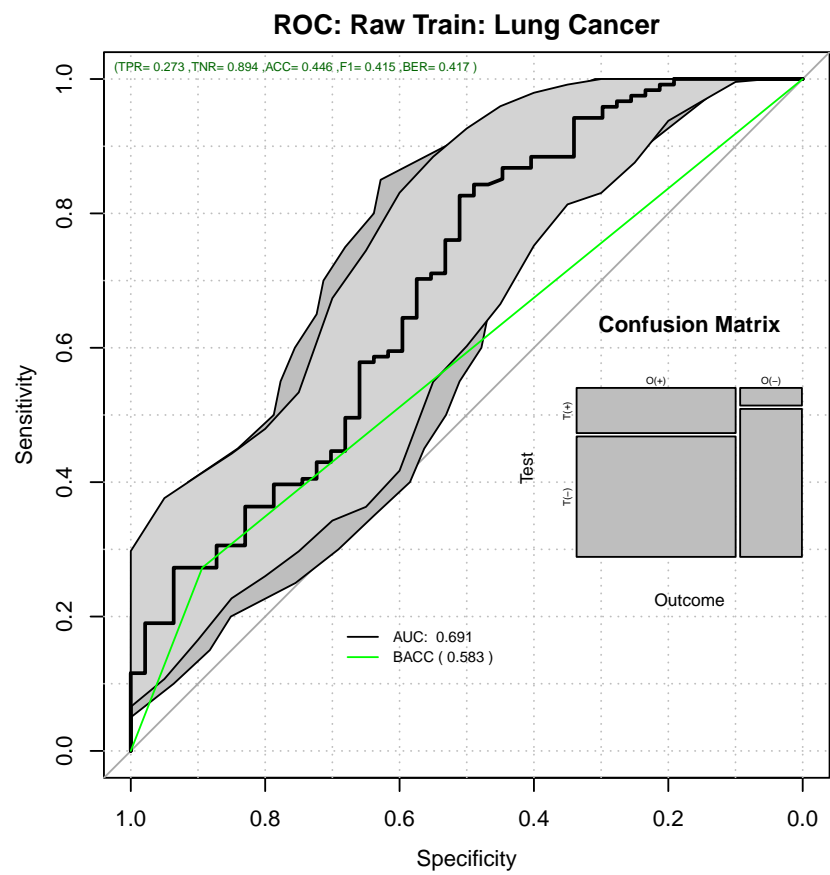


Decision Curve Analysis: Raw Train: Lung Cancer

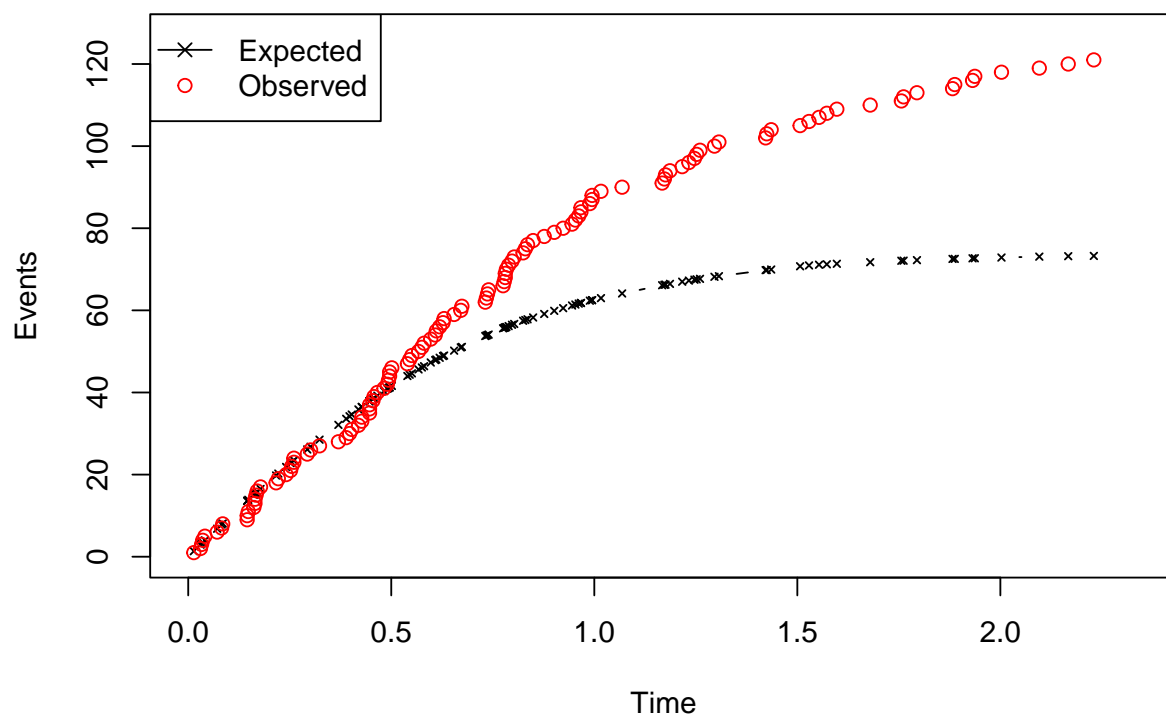


Relative Risk: Raw Train: Lung Cancer

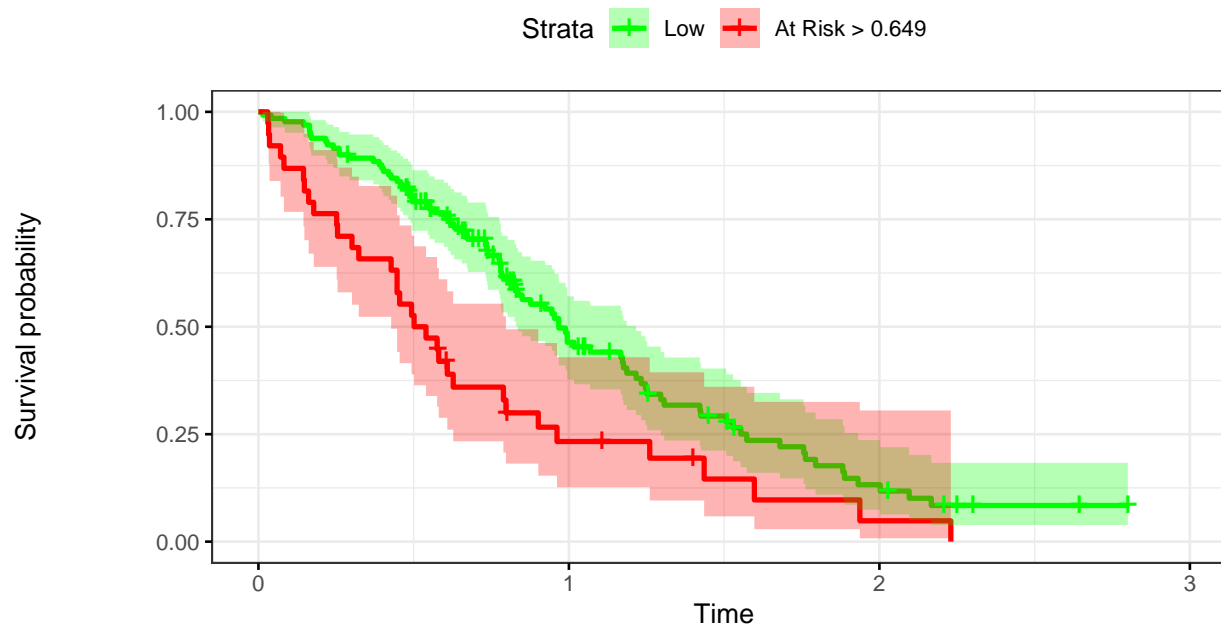




Time vs. Events: Raw Train: Lung Cancer



Kaplan–Meier: Raw Train: Lung Cancer



Number at risk

Low	130	42	9	0
At Risk > 0.649	38	7	1	0

As we can see the Observed probability as well as the Time vs. Events are not calibrated.

0.2.2 Uncalibrated Performance Report

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

Table 7: O/E Ratio

est	lower	upper
1.65	1.37	1.97

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

Table 8: O/E Ratio

mean	50%	2.5%	97.5%
1.21	1.21	1.16	1.25

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

Table 9: O/Acum Ratio

mean	50%	2.5%	97.5%
1.2	1.2	1.19	1.21

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

mean.C Index	median	lower	upper
0.651	0.652	0.592	0.712

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

Table 11: ROC AUC

est	lower	upper
0.691	0.598	0.784

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

Table 12: Sensitivity

est	lower	upper
0.273	0.196	0.361

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

Table 13: Specificity

est	lower	upper
0.894	0.769	0.965

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

Table 14: Probability Thresholds

90%
0.649

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

Table 15: Risk Ratio

est	lower	upper
1.31	1.11	1.54

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

Table 16: Logrank test Chisq = 10.879375 on 1 degrees of freedom,
p = 0.000972

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	130	88	101.3	1.76	10.9
class=1	38	33	19.7	9.05	10.9

0.2.3 Cox Calibration

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

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

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

h0	Gain	DeltaTime
1.29	1.52	2.05

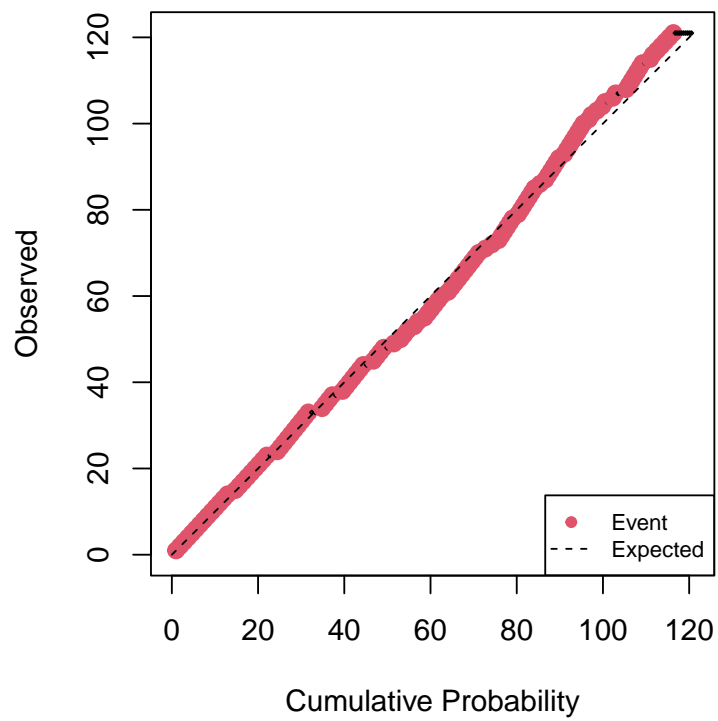
0.2.4 The RRplot() of the calibrated model

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

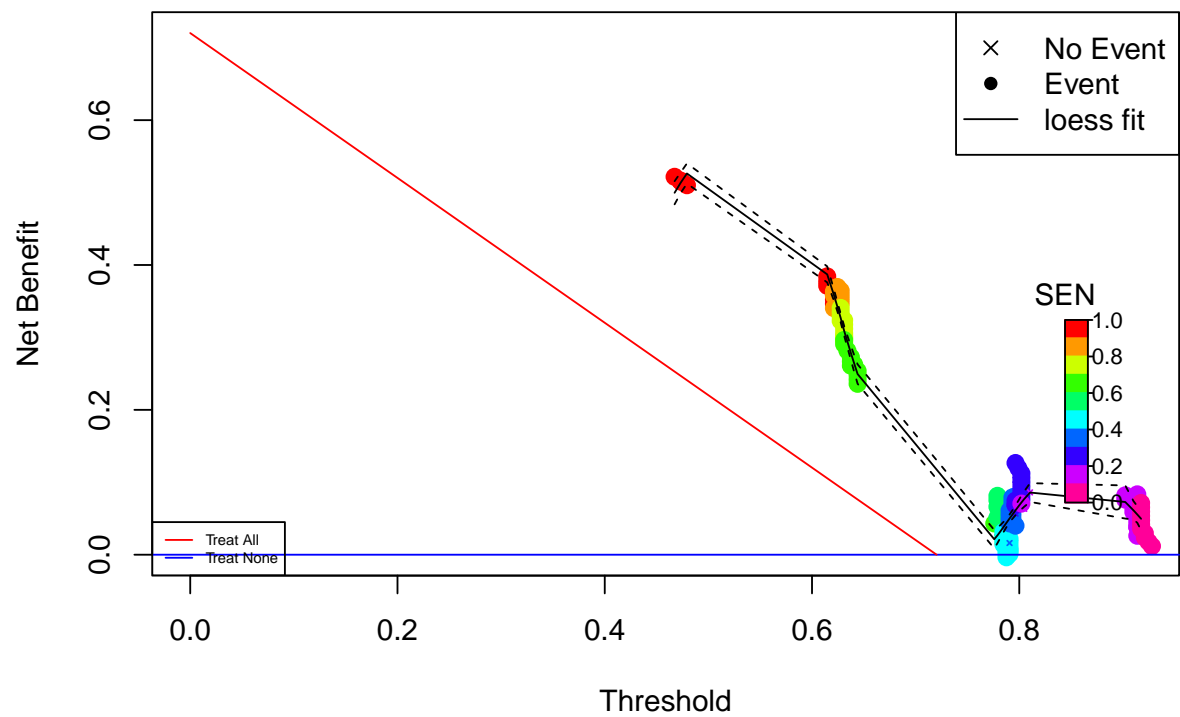
rdata <- cbind(dataLung$status, calprob$prob)

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

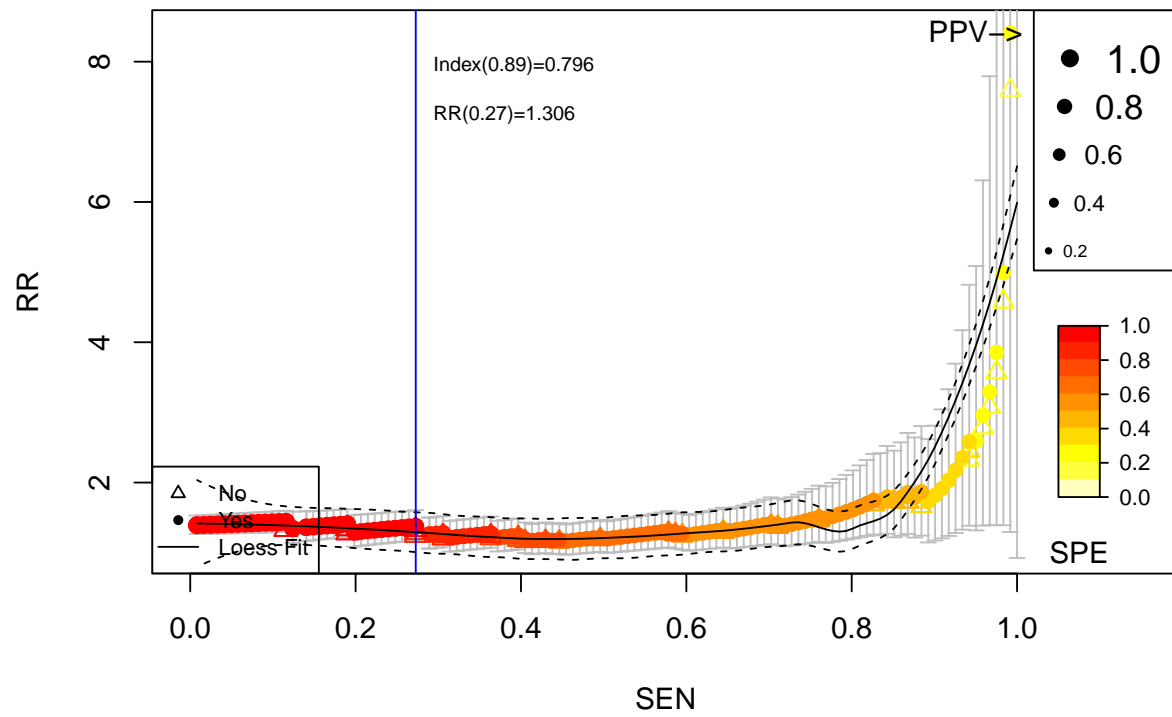
Cumulative vs. Observed: Train: Lung

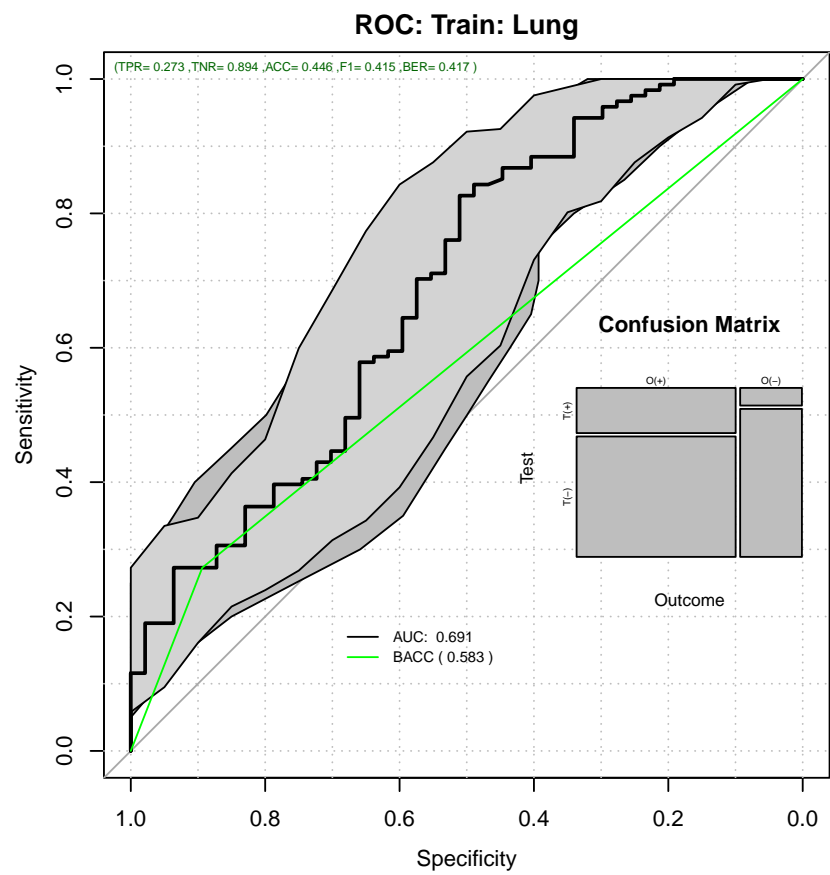


Decision Curve Analysis: Train: Lung

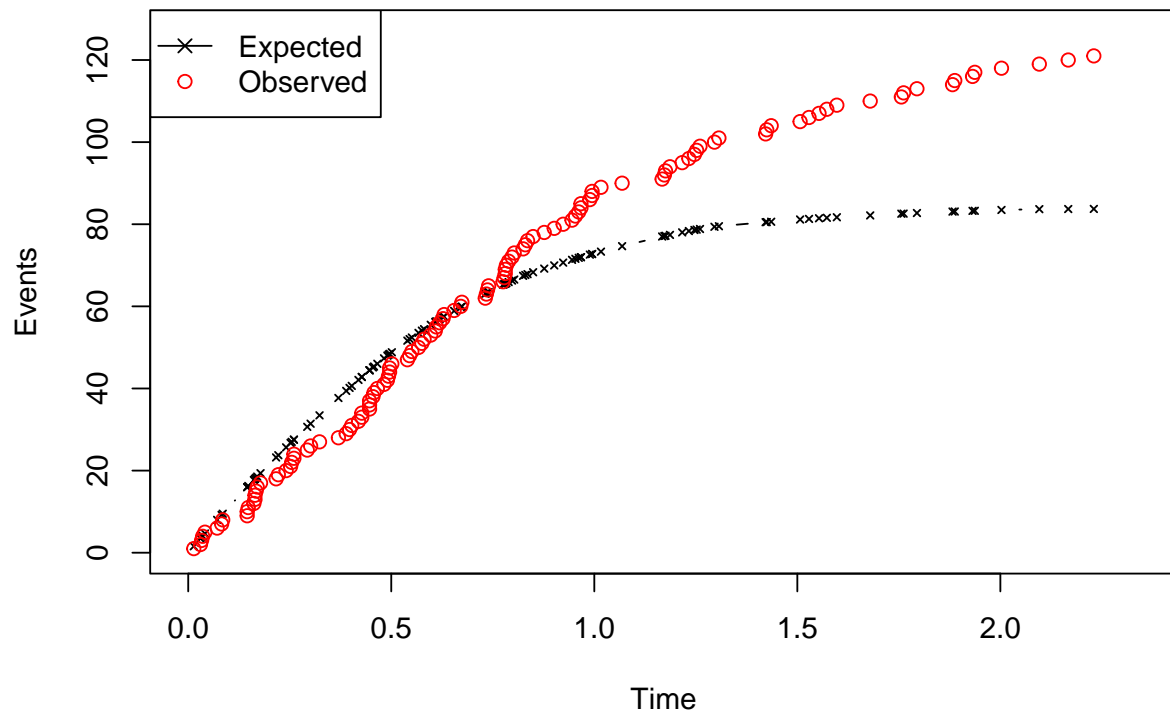


Relative Risk: Train: Lung

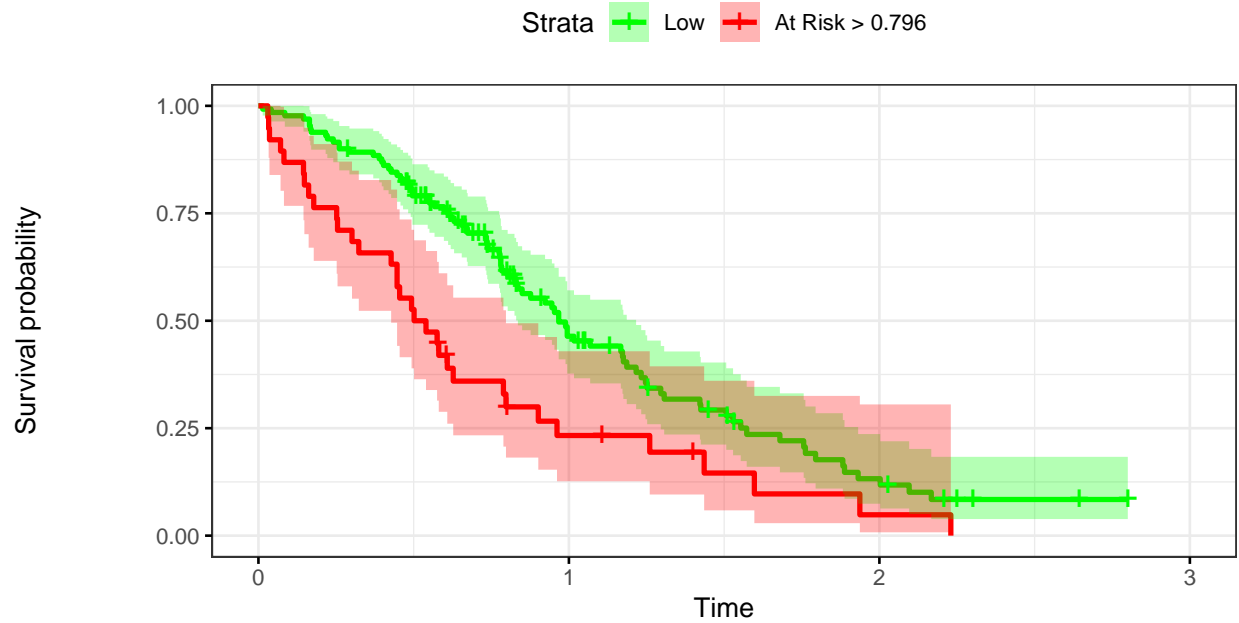




Time vs. Events: Train: Lung



Kaplan–Meier: Train: Lung



Number at risk

Low	130	42	9	0
At Risk > 0.796	38	7	1	0

0.2.5 Calibrated Train Performance

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

Table 18: O/E Ratio

est	lower	upper
1.45	1.2	1.73

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

Table 19: O/E Ratio

mean	50%	2.5%	97.5%
1.04	1.04	0.998	1.08

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

Table 20: O/Acum Ratio

mean	50%	2.5%	97.5%
1.01	1.01	0.999	1.01

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

mean.C Index	median	lower	upper
0.651	0.653	0.59	0.712

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

Table 22: ROC AUC

est	lower	upper
0.691	0.598	0.784

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

Table 23: Sensitivity

est	lower	upper
0.273	0.196	0.361

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

Table 24: Specificity

est	lower	upper
0.894	0.769	0.965

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

Table 25: Probability Thresholds

90%
0.796

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

Table 26: Risk Ratio

est	lower	upper
1.31	1.11	1.54

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

Table 27: Logrank test Chisq = 10.879375 on 1 degrees of freedom,
p = 0.000972

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	130	88	101.3	1.76	10.9
class=1	38	33	19.7	9.05	10.9

0.3 Cross-Validation

```
rcv <- randomCV(theData=dataLung,
  theOutcome = Surv(time,status)~1,
  fittingFunction=BSWiMS.model,
  trainFraction = 0.95,
  repetitions=200,
  classSamplingType = "Pro"
)
```

```
## .[+].[+].[+++].[+++].[+].[+++].[+++].[+].[+].[+++]10 Tested: 79 Avg. Selected: 3.4 Min Tests: 1
## .[+].[+++].[+].[+++].[+].[+].[+].[+++].[+++].[+++]20 Tested: 110 Avg. Selected: 3.35 Min Tests: 1
## .[++++].[+].[+++].[+++].[+].[+].[+++].[+].[+].[+++]30 Tested: 136 Avg. Selected: 3.4 Min Tests: 1
## .[+++].[+].[++++].[+++].[+].[+].[+++].[+++].[+++].[+++]40 Tested: 148 Avg. Selected: 3.475 Min Tests: 1
## .[+++].[+++].[++++].[+++].[+++].[+++].[+++].[+].[+].[+]50 Tested: 153 Avg. Selected: 3.56 Min Tests: 1
## .[+++].[+++].[+++].[+++].[+++].[+].[+-].[+++].[+++].[+]60 Tested: 160 Avg. Selected: 3.6 Min Tests: 1
## .[+++].[+-].[+++].[+].[+++].[+].[+++].[+++].[+++].[+]70 Tested: 164 Avg. Selected: 3.6 Min Tests: 1
## .[+++].[+++].[+].[+++].[+++].[+].[+++].[+++].[+++].[+]80 Tested: 167 Avg. Selected: 3.6 Min Tests: 1
## .[+++].[+].[+-].[+++].[+++].[+++].[+++].[+].[+].[+]90 Tested: 168 Avg. Selected: 3.6 Min Tests: 1
## .[+].[+++].[+++].[+++].[+++].[+].[+].[+-].[+++].[+]100 Tested: 168 Avg. Selected: 3.59 Min Tests: 1
## .[+++].[++++].[+].[+].[+].[+++].[+-].[+++].[+].[+]110 Tested: 168 Avg. Selected: 3.581818 Min Tests: 1
## .[+++].[+].[+++].[+++].[++++].[+].[+-].[+++].[++++].[+]120 Tested: 168 Avg. Selected: 3.6 Min Tests: 1
## .[++++].[+++].[+++].[+].[+++].[+++].[+++].[+++].[+++].[+-]130 Tested: 168 Avg. Selected: 3.623077 Min Tests: 1
## .[+++].[+].[+++].[++++].[+++].[+].[+++].[+-].[+].[+-]140 Tested: 168 Avg. Selected: 3.614286 Min Tests: 1
## .[+].[+].[+++].[+-].[+++].[+++].[+++].[+++].[+++].[+]150 Tested: 168 Avg. Selected: 3.606667 Min Tests: 1
## .[+].[+++].[+++].[+++].[+++].[+++].[+++].[+].[+].[+]160 Tested: 168 Avg. Selected: 3.60625 Min Tests: 1
## .[++++].[+].[+++].[+++].[+++].[+++].[+].[++++].[+++].[+]170 Tested: 168 Avg. Selected: 3.617647 Min Tests: 1
## .[+++].[+].[+].[+++].[+++].[+++].[+].[+-].[+++].[+]180 Tested: 168 Avg. Selected: 3.611111 Min Tests: 1
## .[+++].[+++].[+++].[++++].[+++].[+++].[+++].[+++].[+++].[+]190 Tested: 168 Avg. Selected: 3.636842 Min Tests: 1
## .[+].[+++].[++++].[+].[+++].[+++].[+].[+++].[+++].[+]200 Tested: 168 Avg. Selected: 3.64 Min Tests: 1
##
```

```
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,]
```

```

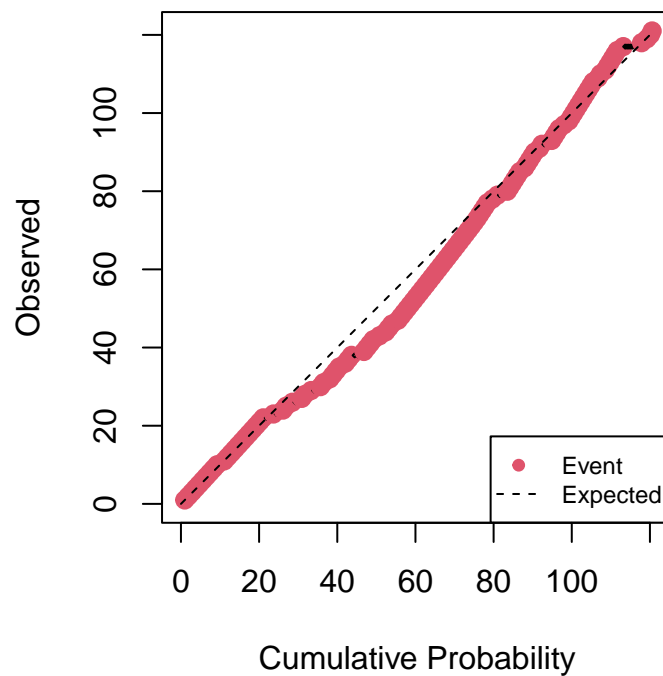
prob <- ppoisGzero(boxplot(unlist(stp[,4])~rownames(stp),plot=FALSE)$stats[3,],h0)

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

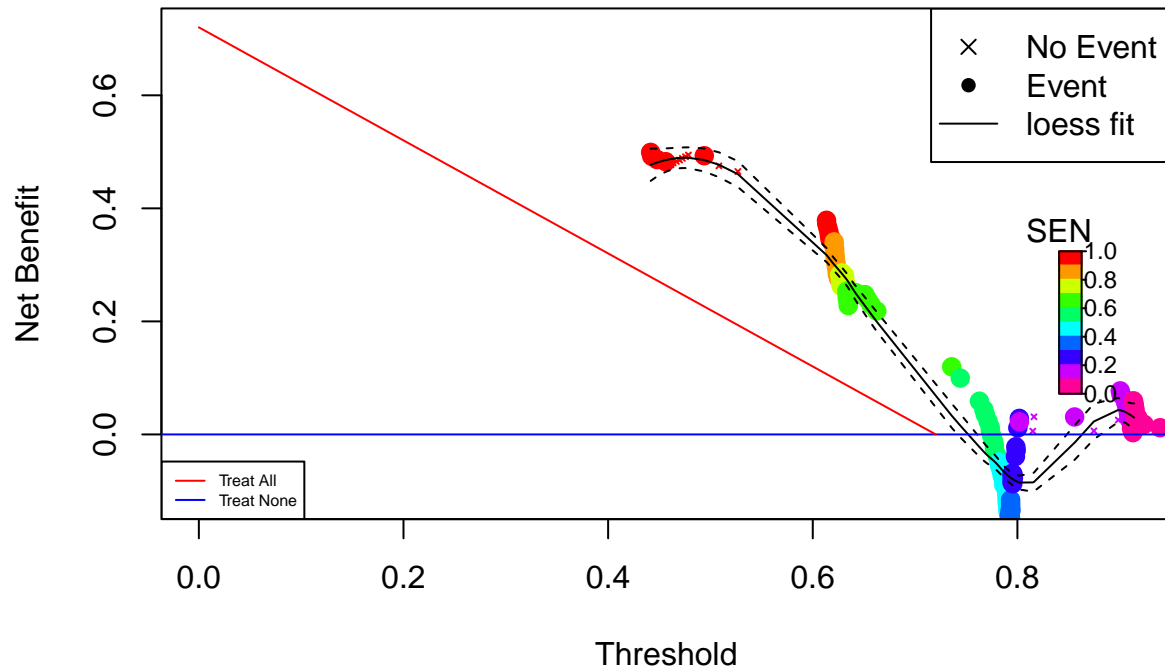
rrAnalysisTest <- RRPlot(rdatacv,atProb=c(0.90),
                          timetoEvent=times,
                          title="Test: Lung Cancer",
                          ysurvlim=c(0.00,1.0),
                          riskTimeInterval=timeinterval)

```

Cumulative vs. Observed: Test: Lung Cancer



Decision Curve Analysis: Test: Lung Cancer



```
## Warning in plot.xy(xy.coords(x, y), type = type, ...): font width unknown for
## character 0x1
## Warning in plot.xy(xy.coords(x, y), type = type, ...): font metrics unknown for
## character 0x1
## Warning in plot.xy(xy.coords(x, y), type = type, ...): font width unknown for
## character 0x1
## Warning in plot.xy(xy.coords(x, y), type = type, ...): font metrics unknown for
## character 0x1
## Warning in plot.xy(xy.coords(x, y), type = type, ...): font width unknown for
## character 0x1
## Warning in plot.xy(xy.coords(x, y), type = type, ...): font metrics unknown for
## character 0x1
## Warning in plot.xy(xy.coords(x, y), type = type, ...): font width unknown for
## character 0x1
## Warning in plot.xy(xy.coords(x, y), type = type, ...): font metrics unknown for
## character 0x1
## Warning in plot.xy(xy.coords(x, y), type = type, ...): font width unknown for
## character 0x1
## Warning in plot.xy(xy.coords(x, y), type = type, ...): font metrics unknown for
## character 0x1
## Warning in plot.xy(xy.coords(x, y), type = type, ...): font width unknown for
## character 0x1
```

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

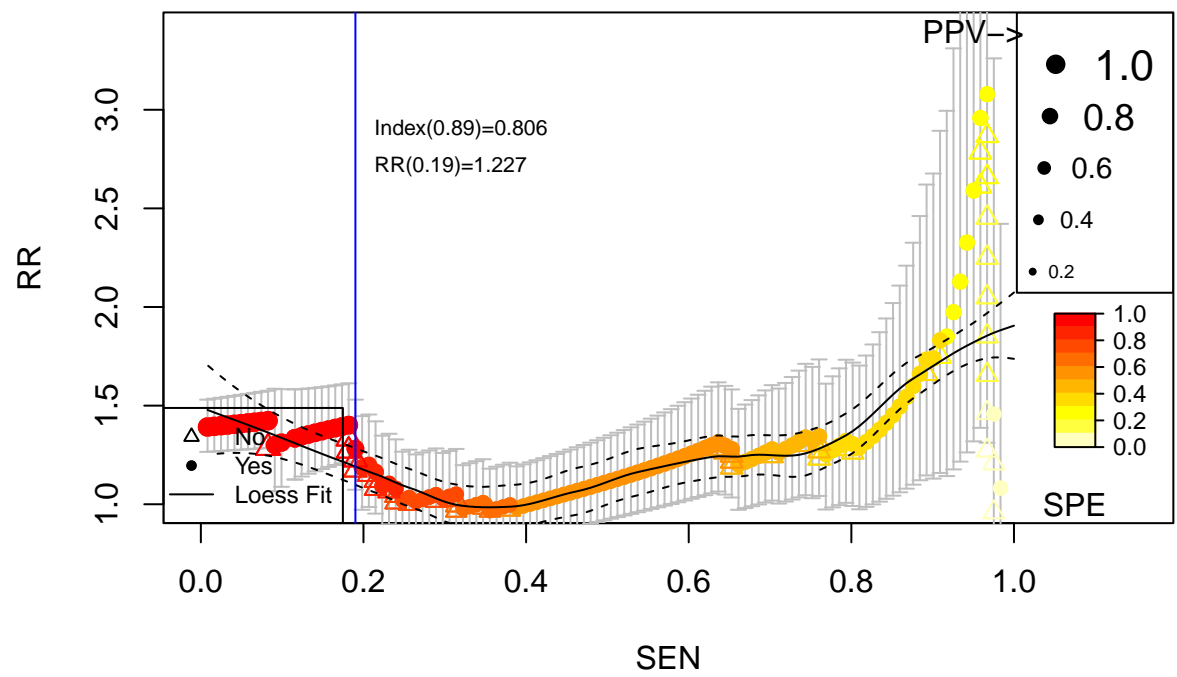
[illegible]

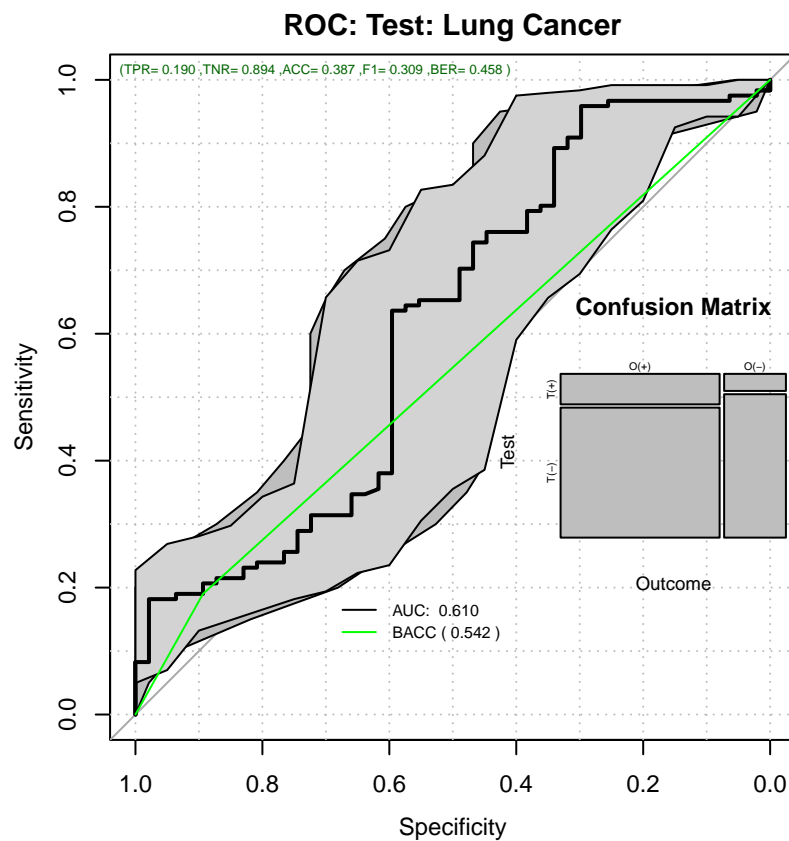
[illegible]

[illegible]

[illegible]

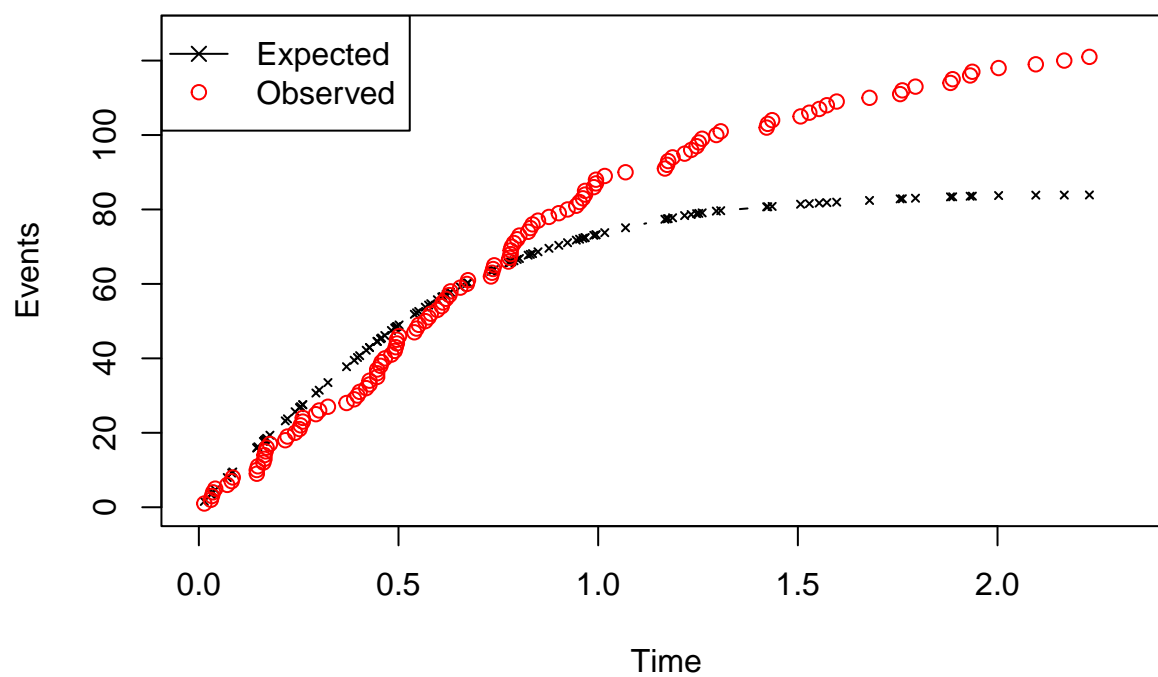
Relative Risk: Test: Lung Cancer



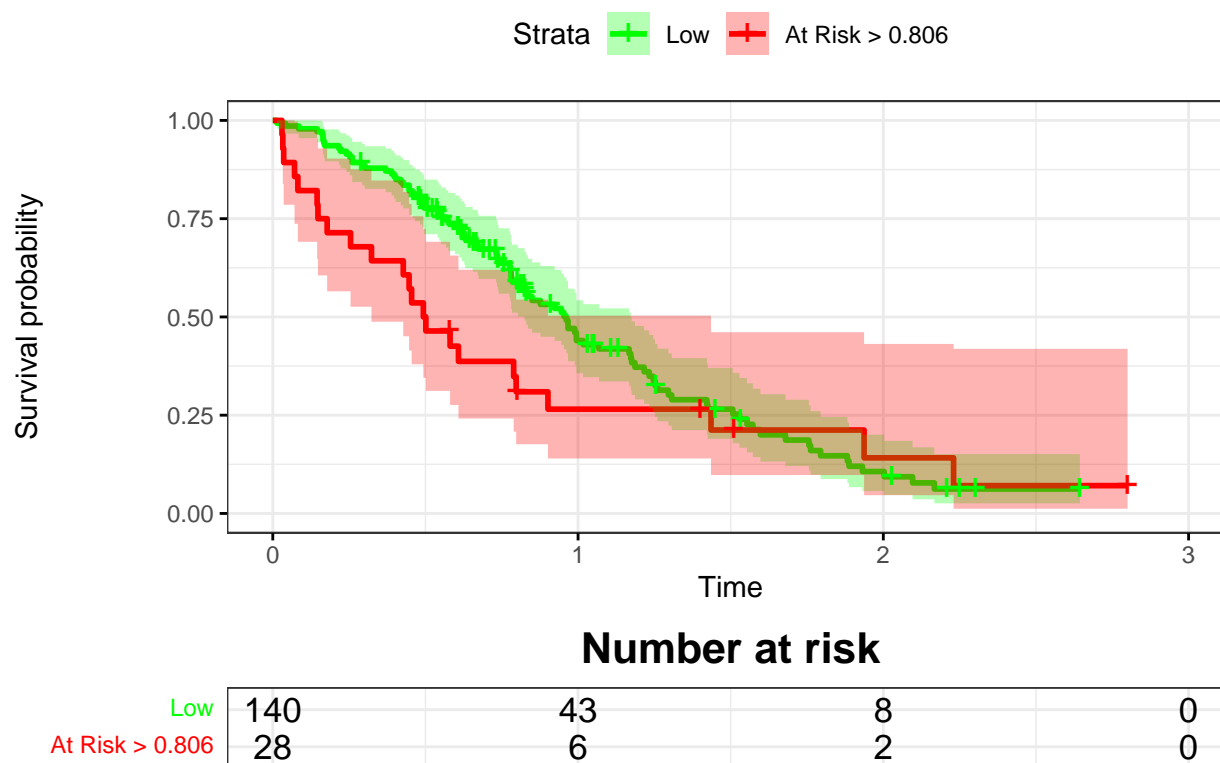


```
## Warning in cstat$cstatCI <- cstatCI: Coercing LHS to a list
```

Time vs. Events: Test: Lung Cancer



Kaplan–Meier: Test: Lung Cancer



0.3.1 Cross-Validation Test Performance

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

Table 28: O/E Ratio

est	lower	upper
1.44	1.2	1.72

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

Table 29: O/E Ratio

mean	50%	2.5%	97.5%
1.03	1.03	0.996	1.07

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

Table 30: O/Acum Ratio

mean	50%	2.5%	97.5%
0.953	0.953	0.943	0.964

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

mean.C Index	median	lower	upper
0.602	0.602	0.537	0.669

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

Table 32: ROC AUC

est	lower	upper
0.61	0.508	0.711

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

Table 33: Sensitivity

est	lower	upper
0.19	0.124	0.271

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

Table 34: Specificity

est	lower	upper
0.894	0.769	0.965

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

Table 35: Probability Thresholds

90%
0.806

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

Table 36: Risk Ratio

est	lower	upper
1.23	1.01	1.48

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

Table 37: Logrank test Chisq = 2.346444 on 1 degrees of freedom,
p = 0.125569

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	140	98	103.8	0.328	2.35
class=1	28	23	17.2	1.981	2.35

0.3.2 Calibrating the test results

```
rdatacv <- cbind(status,prob,times)
calprob <- CalibrationProbPoissonRisk(rdatacv)

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

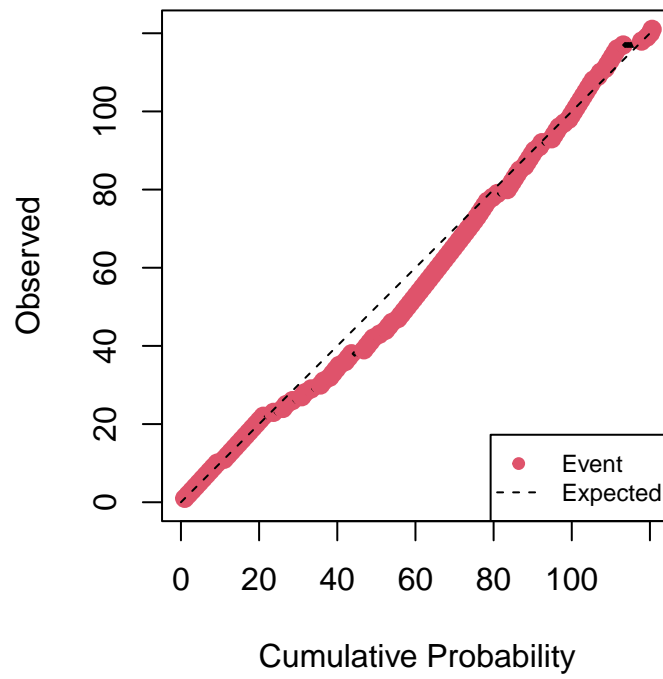
h0	Gain	DeltaTime
0.85	1	2.06

```
timeinterval <- calprob$timeInterval;

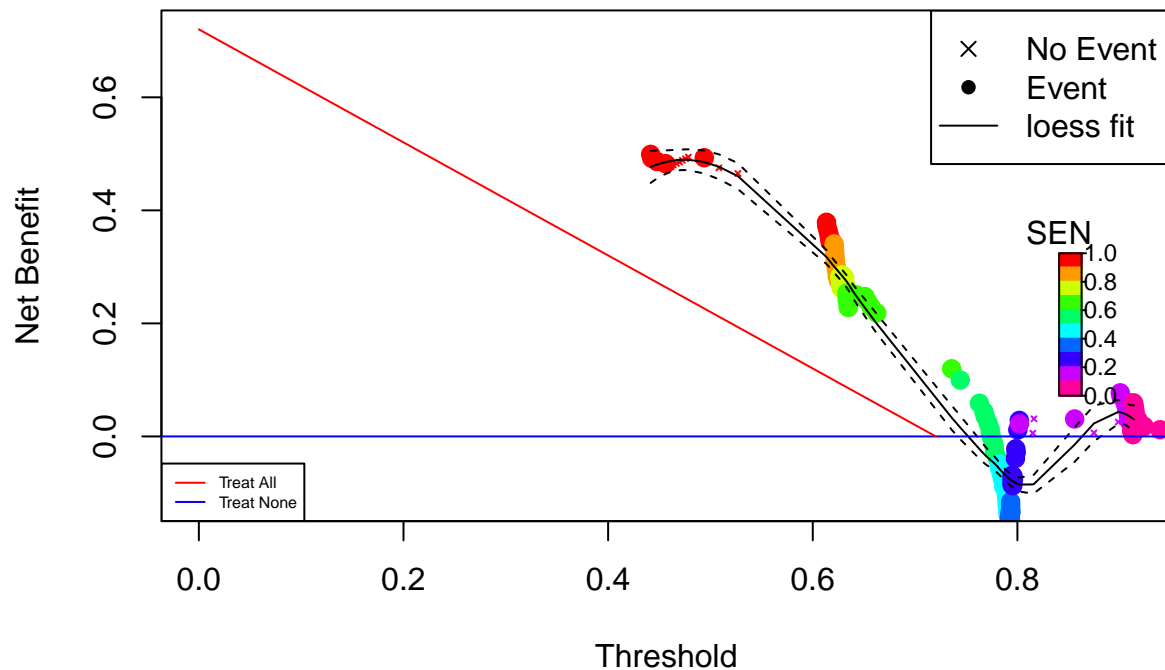
rdata <- cbind(status,calprob$prob)

rrAnalysisTest <- RRPlot(rdata,atProb=c(0.90),
  timetoEvent=times,
  title="Calibrated Test: Lung",
  ysurvlim=c(0.00,1.0),
  riskTimeInterval=timeinterval)
```

Cumulative vs. Observed: Calibrated Test: Lung



Decision Curve Analysis: Calibrated Test: Lung



```
## Warning in plot.xy(xy.coords(x, y), type = type, ...): font width unknown for
## character 0x1
```

```
## Warning in plot.xy(xy.coords(x, y), type = type, ...): font metrics unknown for
## character 0x1
```

```
## Warning in plot.xy(xy.coords(x, y), type = type, ...): font width unknown for
## character 0x1
```

```
## Warning in plot.xy(xy.coords(x, y), type = type, ...): font metrics unknown for
## character 0x1
```

```
## Warning in plot.xy(xy.coords(x, y), type = type, ...): font width unknown for
## character 0x1
```

```
## Warning in plot.xy(xy.coords(x, y), type = type, ...): font metrics unknown for
## character 0x1
```

```
## Warning in plot.xy(xy.coords(x, y), type = type, ...): font width unknown for
## character 0x1
```

```
## Warning in plot.xy(xy.coords(x, y), type = type, ...): font metrics unknown for
## character 0x1
```

```
## Warning in plot.xy(xy.coords(x, y), type = type, ...): font width unknown for
## character 0x1
```

```
## Warning in plot.xy(xy.coords(x, y), type = type, ...): font metrics unknown for
## character 0x1
```

```
## Warning in plot.xy(xy.coords(x, y), type = type, ...): font width unknown for
```


[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

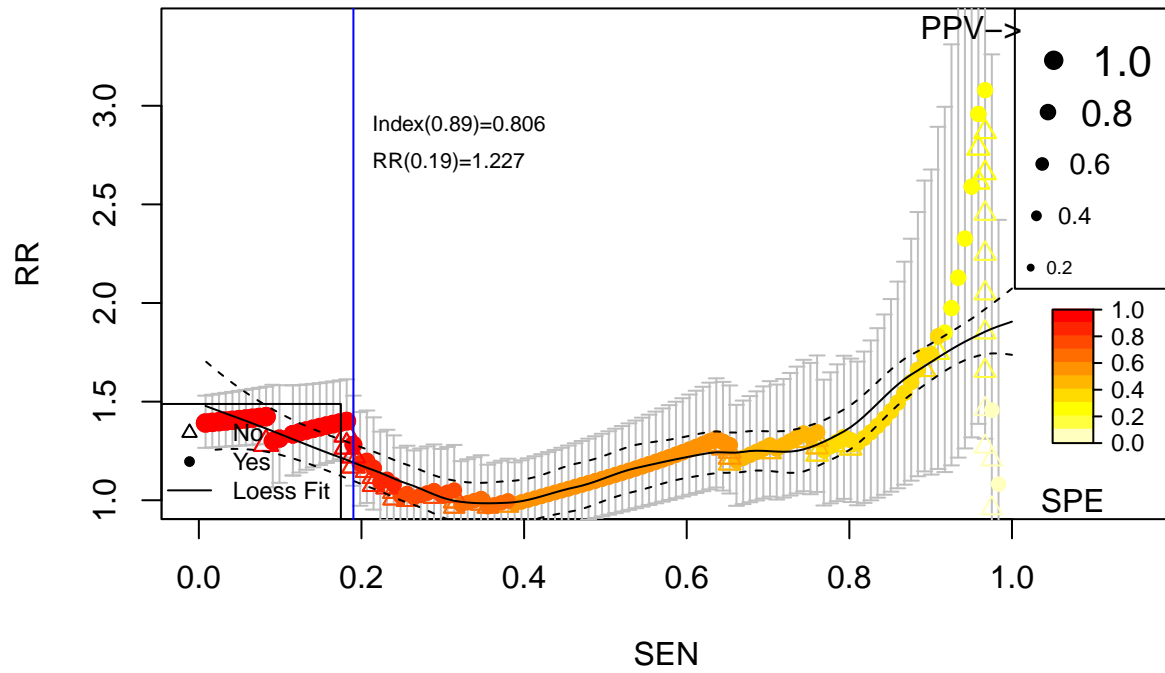
[illegible]

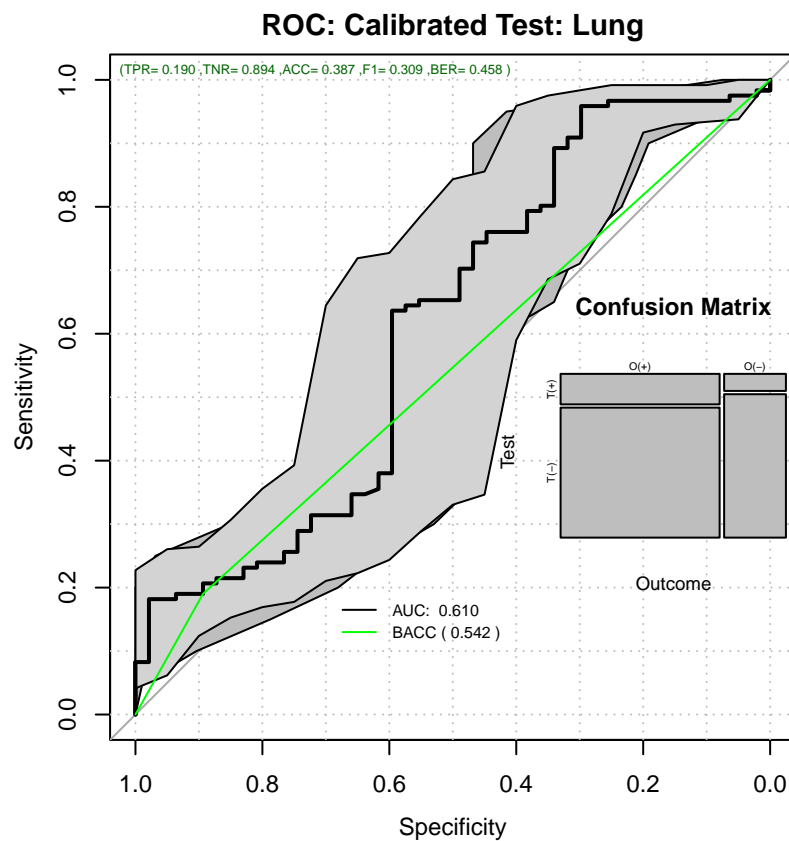
[illegible]

[illegible]

[illegible]

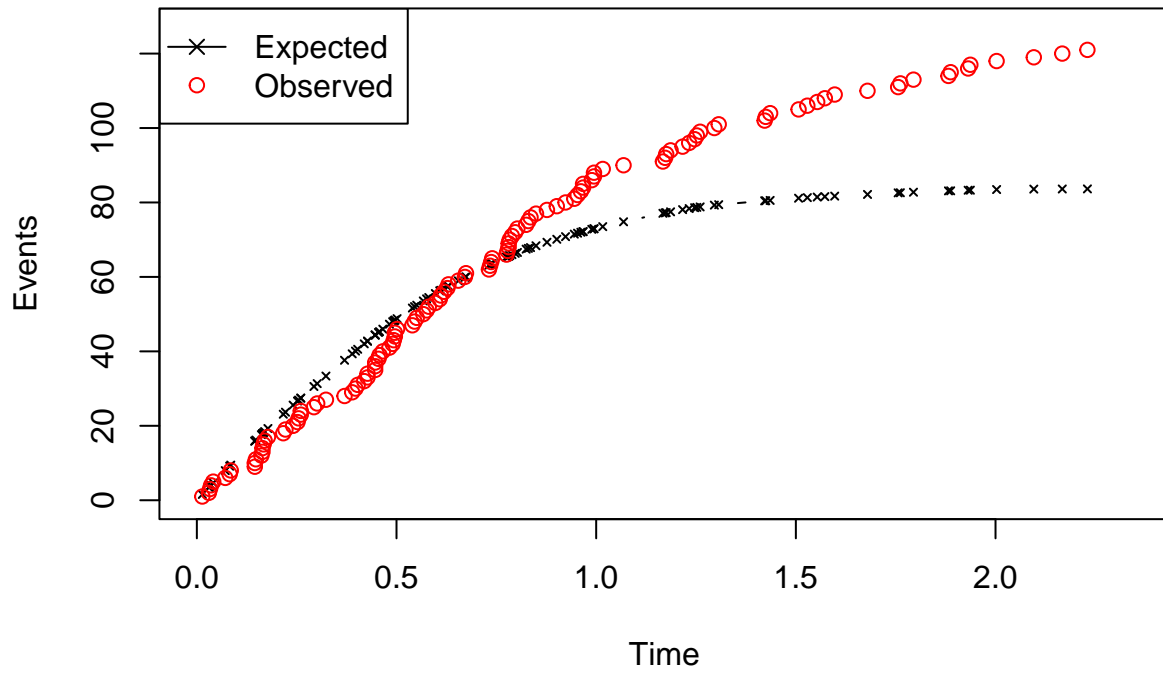
Relative Risk: Calibrated Test: Lung



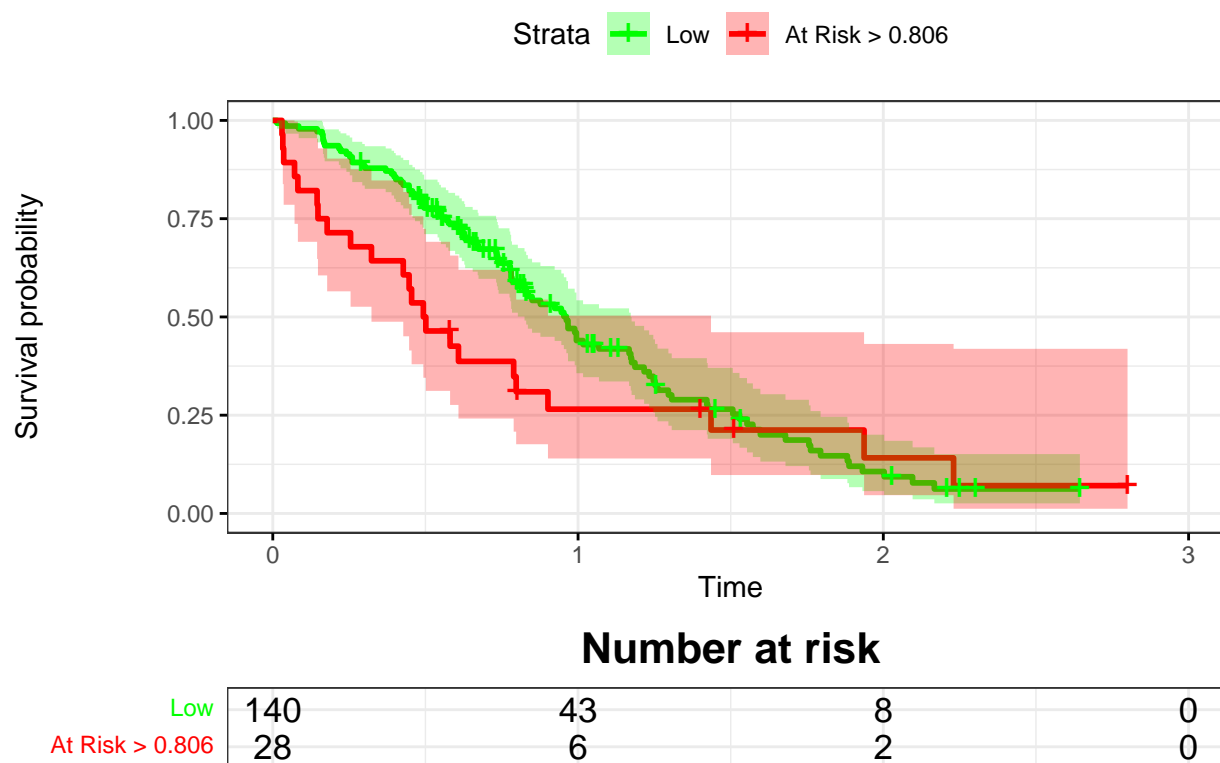


```
## Warning in cstat$cstatCI <- cstatCI: Coercing LHS to a list
```

Time vs. Events: Calibrated Test: Lung



Kaplan–Meier: Calibrated Test: Lung



0.3.3 Calibrated Test Performance

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

Table 39: O/E Ratio

est	lower	upper
1.45	1.2	1.73

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

Table 40: O/E Ratio

mean	50%	2.5%	97.5%
1.04	1.04	0.998	1.08

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

Table 41: O/Acum Ratio

mean	50%	2.5%	97.5%
0.953	0.953	0.942	0.964

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

mean.C Index	median	lower	upper
0.602	0.602	0.532	0.667

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

Table 43: ROC AUC

est	lower	upper
0.61	0.508	0.711

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

Table 44: Sensitivity

est	lower	upper
0.19	0.124	0.271

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

Table 45: Specificity

est	lower	upper
0.894	0.769	0.965

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

Table 46: Probability Thresholds

90%
0.806

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


Table 47: Risk Ratio

est	lower	upper
1.23	1.01	1.48

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

Table 48: Logrank test Chisq = 2.346444 on 1 degrees of freedom,
p = 0.125569

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	140	98	103.8	0.328	2.35
class=1	28	23	17.2	1.981	2.35