Breast Cancer: Wisconsin

Jose Tamez

2023-05-13

Contents

1 Showcasing RRPlots

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
```

```
#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)
```

```r
pander::panderOptions('keep.trailing.zeros',TRUE)
layout(matrix(1:1, nrow=1))
```

1.0.2 Wisconsin Data Set

```r
dataBreast <- read.csv("~/GitHub/RISKPLOTS/DATA/wpbc.data", header=FALSE)
table(dataBreast$V2)
```

```
##
##   N   R
## 151  47
```

```r
rownames(dataBreast) <- dataBreast$V1
dataBreast$V1 <- NULL
dataBreast$status <- 1*(dataBreast$V2=="R")
dataBreast$V2 <- NULL
dataBreast$time <- dataBreast$V3
dataBreast$V3 <- NULL
dataBreast <- sapply(dataBreast,as.numeric)
```

```
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
```

```r
dataBreast <- as.data.frame(dataBreast[complete.cases(dataBreast),])
table(dataBreast$status)
```

```
##
##   0   1
## 148  46
```

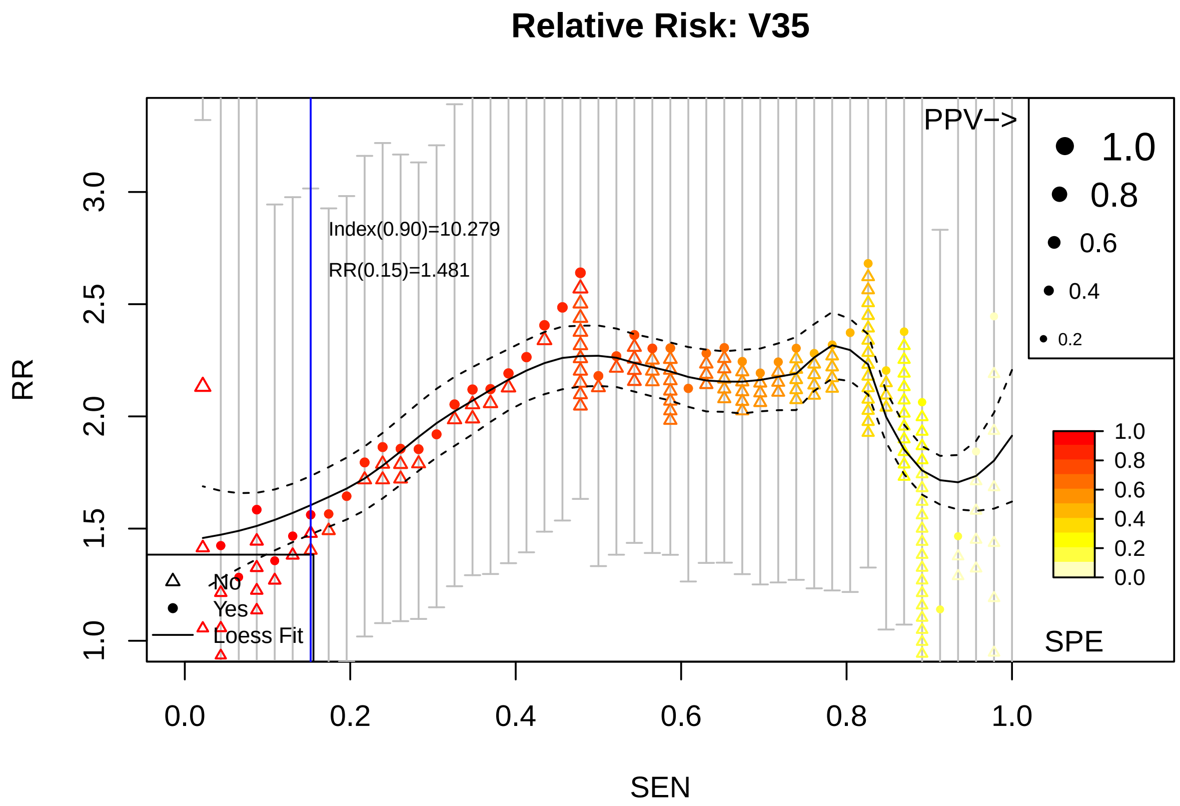
1.1 Exploring Raw Features with RRPlot

```r
convar <- colnames(dataBreast)[lapply(apply(dataBreast,2,unique),length) > 10]
convar <- convar[convar != "time"]
topvar <- univariate_BinEnsemble(dataBreast[,c("status",convar)],"status")
pander::pander(topvar)
```

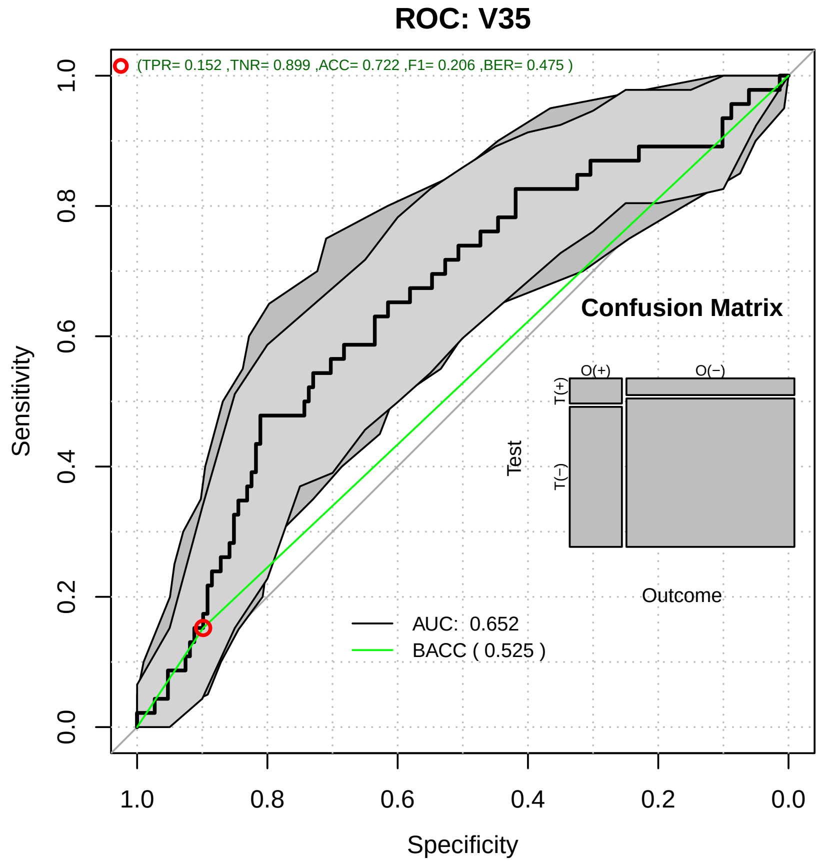
V35	V24	V34	V7	V16	V14	V17
0.0261	0.0261	0.0261	0.0623	0.126	0.126	0.126

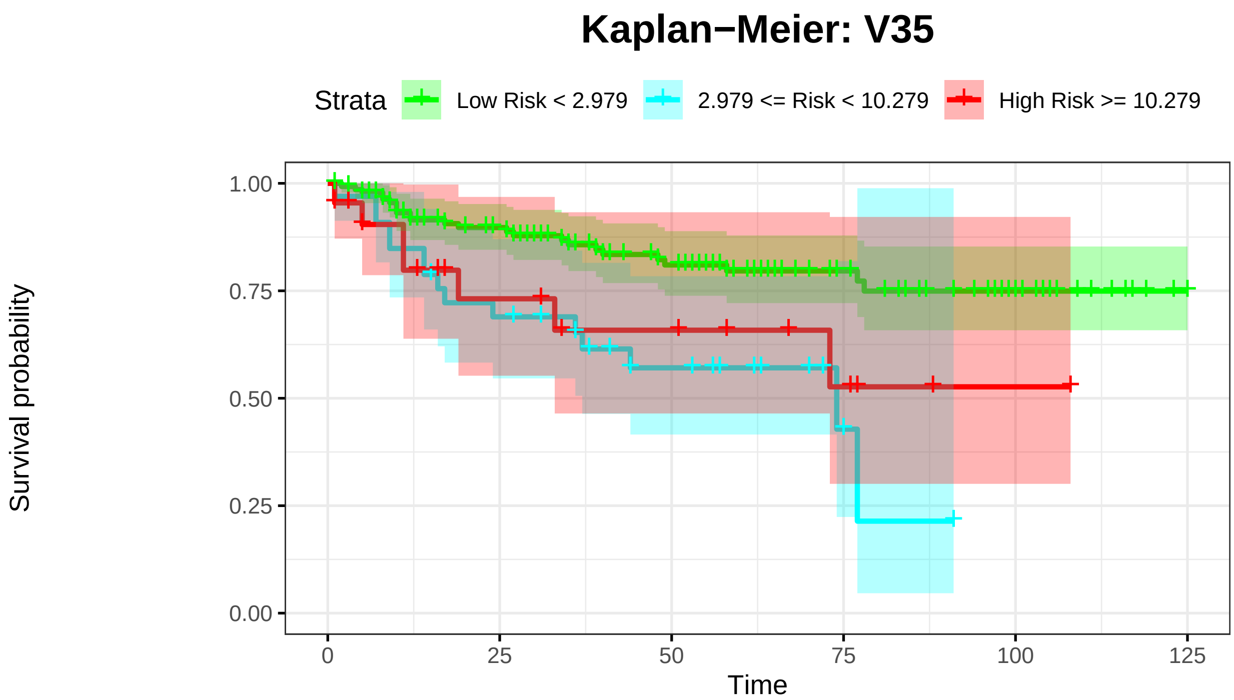
```r
topv <- min(5,length(topvar))
topFive <- names(topvar)[1:topv]
RRanalysis <- list();
idx <- 1
topf <- topFive[1]
for (topf in topFive)
{
  RRanalysis[[idx]] <- RRPlot(cbind(dataBreast$status,dataBreast[,topf]),
                              atRate=c(0.90,0.80),
                  timetoEvent=dataBreast$time,
                  title=topf,
#                  plotRR=FALSE
                  )
```

```
  idx <- idx + 1
}
```



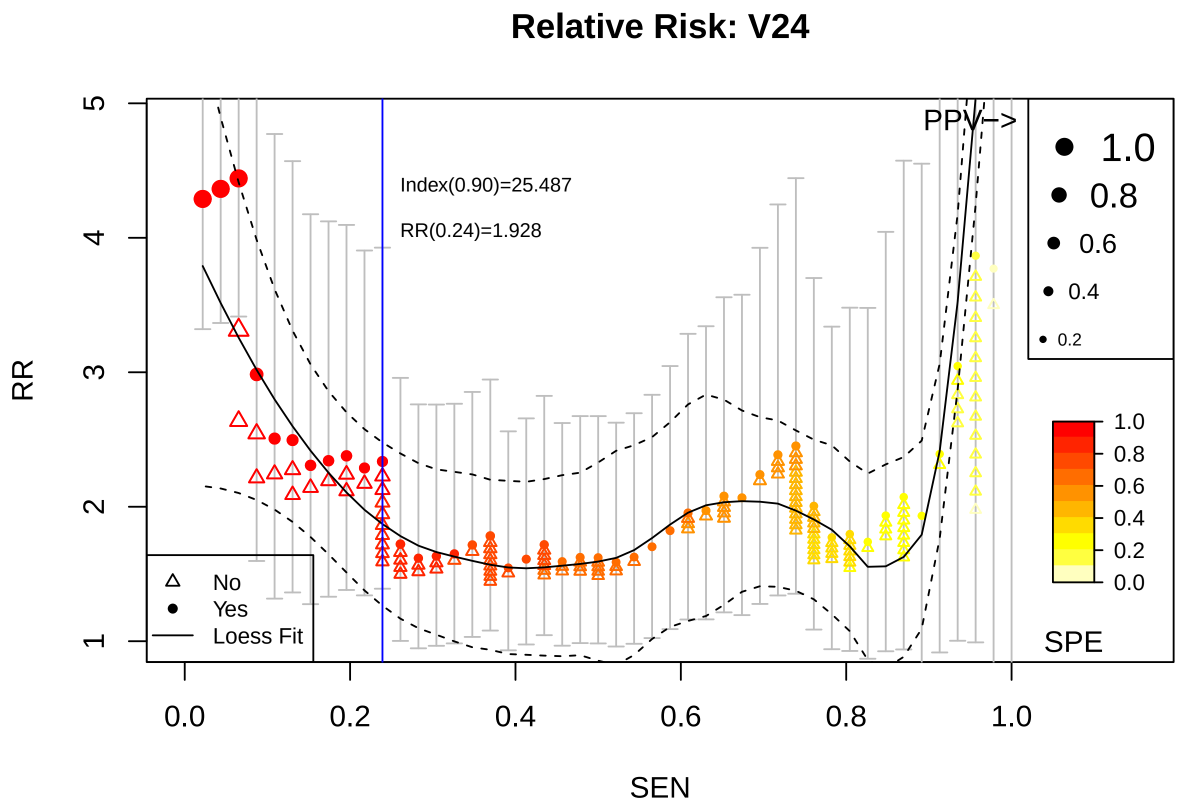
Relative Risk: V35
PPV->
1.0
0.8
0.6
0.4
0.2
Index(0.90)=10.279
RR(0.15)=1.481
RR
SEN
SPE
No
Yes
Loess Fit


ROC: V35
(TPR= 0.152 ,TNR= 0.899 ,ACC= 0.722 ,F1= 0.206 ,BER= 0.475)
Sensitivity
Specificity
1.0
0.8
0.6
0.4
0.2
0.0
Confusion Matrix
O(+)
O(−)
T(+)
T(−)
Test
Outcome
AUC: 0.652
BACC (0.525)

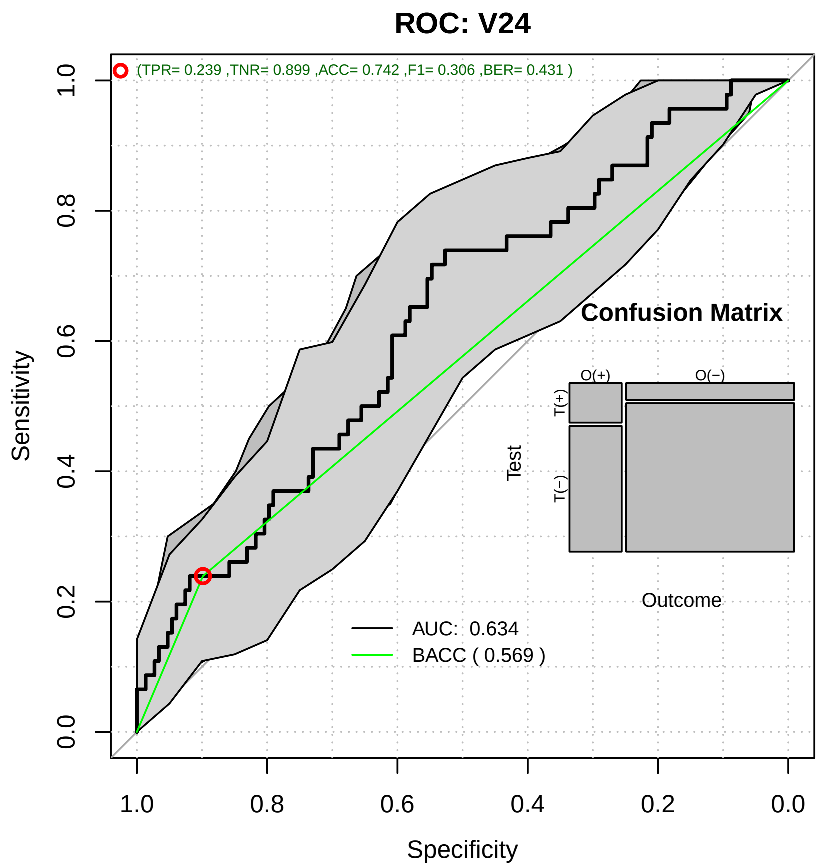


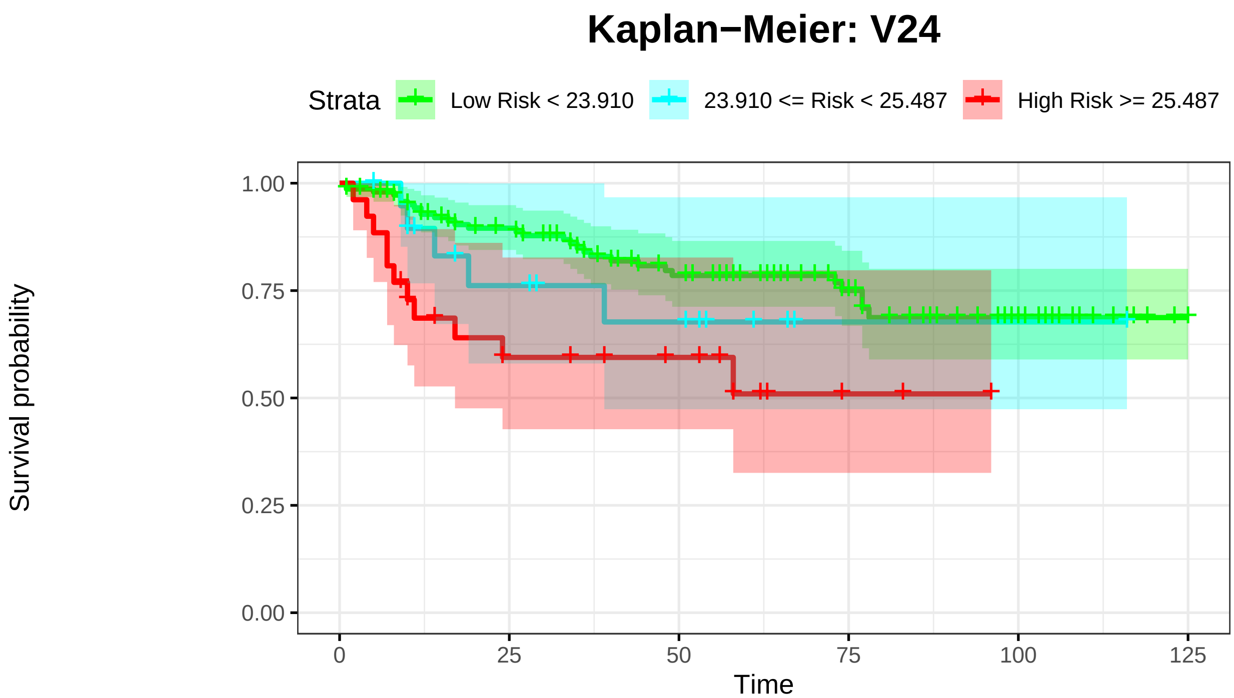
Number at risk

Low Risk < 2.979	139	95	66	37	19	1
2.979 <= Risk < 10.279	33	21	12	3	0	0
High Risk >= 10.279	22	11	8	4	1	0



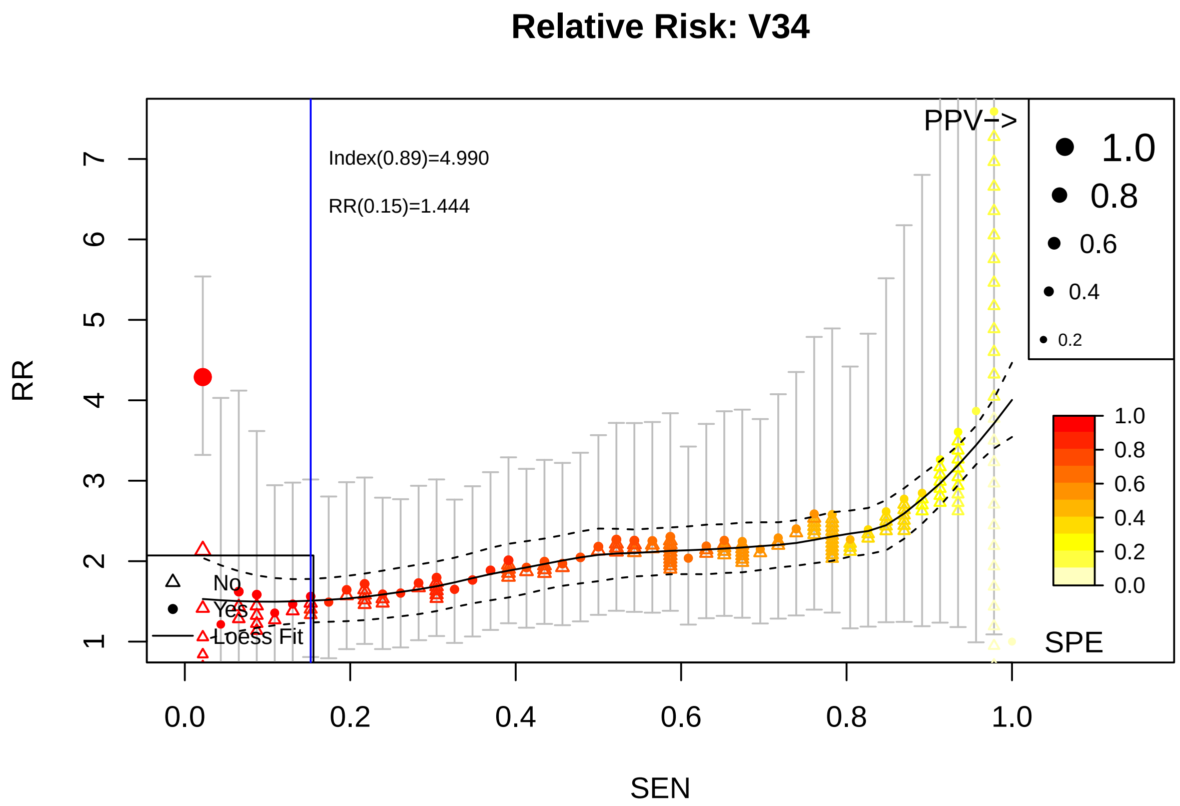
Relative Risk: V24
Index(0.90)=25.487
RR(0.24)=1.928
PPV-->
1.0
0.8
0.6
0.4
0.2
SPE
No
Yes
Loess Fit
RR
SEN
0.0
0.2
0.4
0.6
0.8
1.0
1
2
3
4
5


ROC: V24
(TPR= 0.239 ,TNR= 0.899 ,ACC= 0.742 ,F1= 0.306 ,BER= 0.431)
Sensitivity
Specificity
1.0
0.8
0.6
0.4
0.2
0.0
Confusion Matrix
O(+)
O(−)
T(+)
T(−)
Test
Outcome
AUC: 0.634
BACC (0.569)

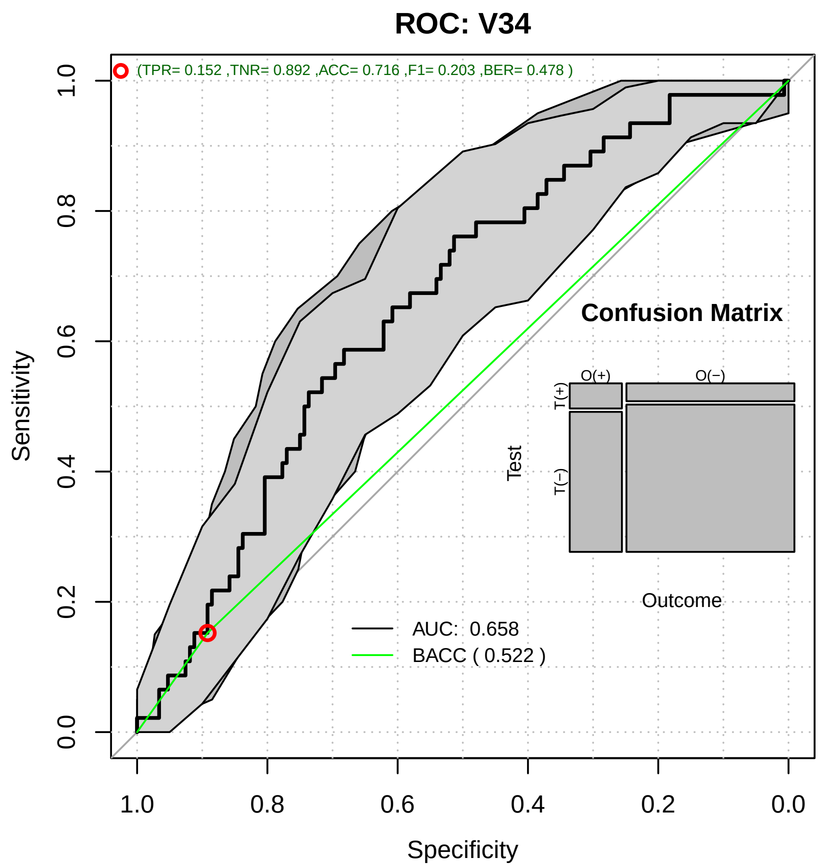


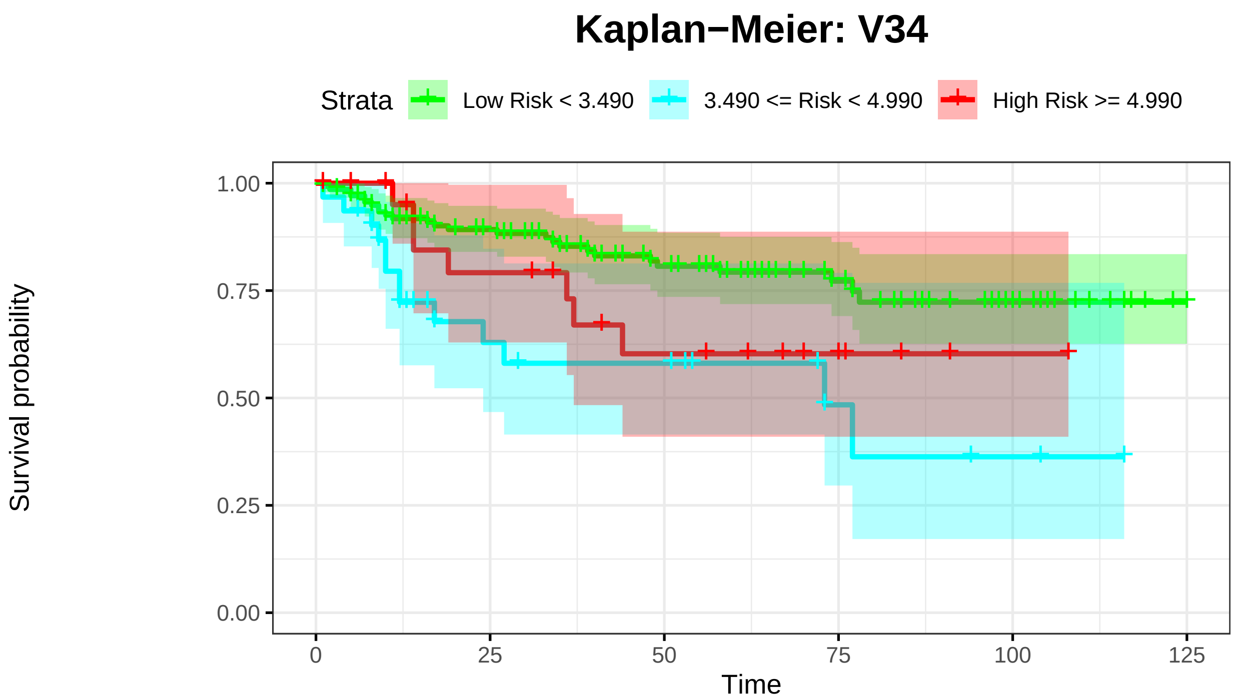
Number at risk

	0	25	50	75	100	125
Low Risk < 23.910	148	104	69	41	19	1
23.910 <= Risk < 25.487	20	11	8	1	1	0
High Risk >= 25.487	26	12	9	2	0	0



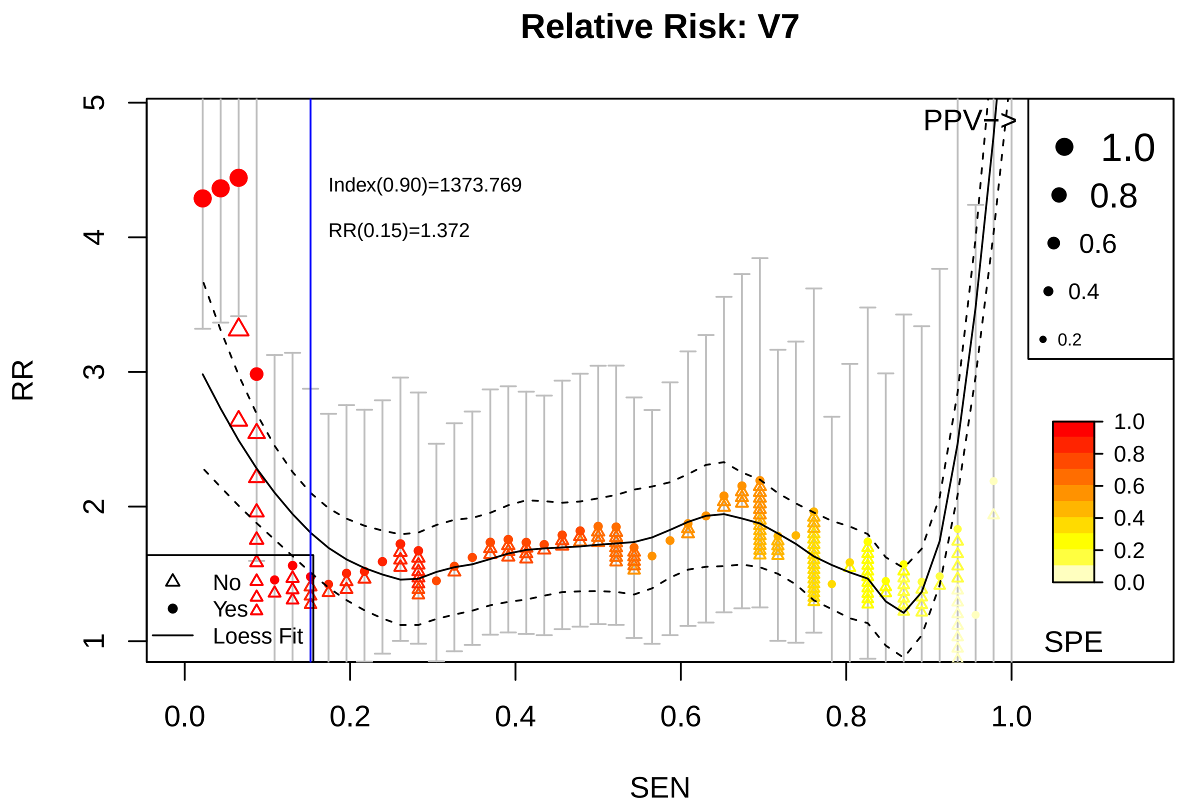
Relative Risk: V34
Index(0.89)=4.990
RR(0.15)=1.444
PPV–>
1.0
0.8
0.6
0.4
0.2
No
Yes
Loess Fit
SPE
1.0
0.8
0.6
0.4
0.2
0.0
RR
1
2
3
4
5
6
7
SEN
0.0
0.2
0.4
0.6
0.8
1.0

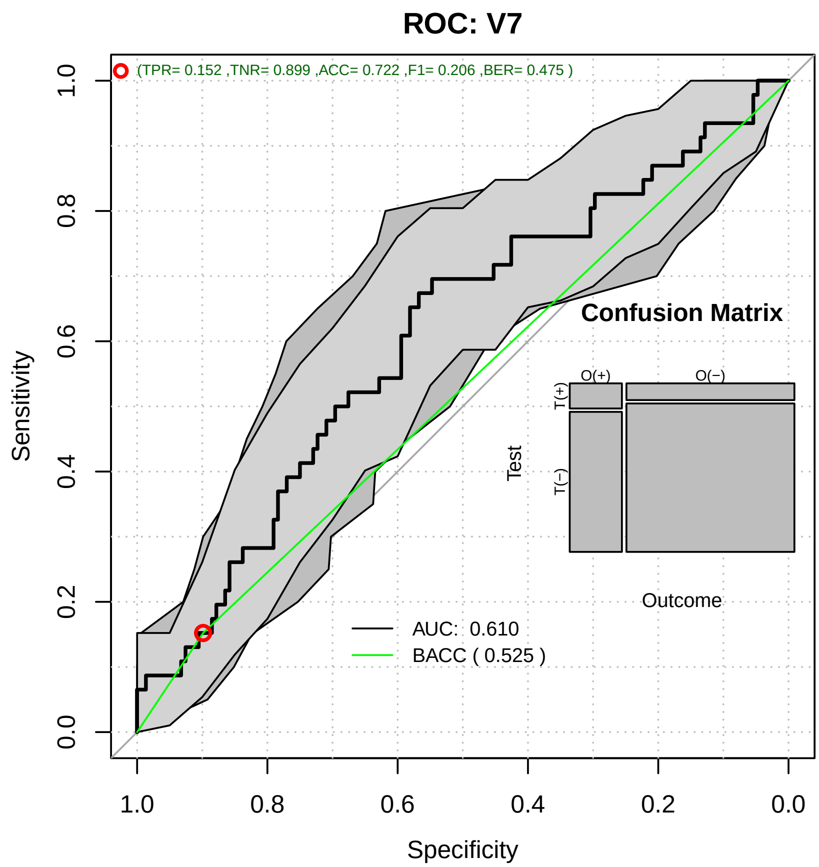

ROC: V34
(TPR= 0.152 ,TNR= 0.892 ,ACC= 0.716 ,F1= 0.203 ,BER= 0.478)
Confusion Matrix
O(+)
O(−)
T(+)
T(−)
Test
Outcome
AUC: 0.658
BACC (0.522)
Sensitivity
Specificity
1.0
0.8
0.6
0.4
0.2
0.0

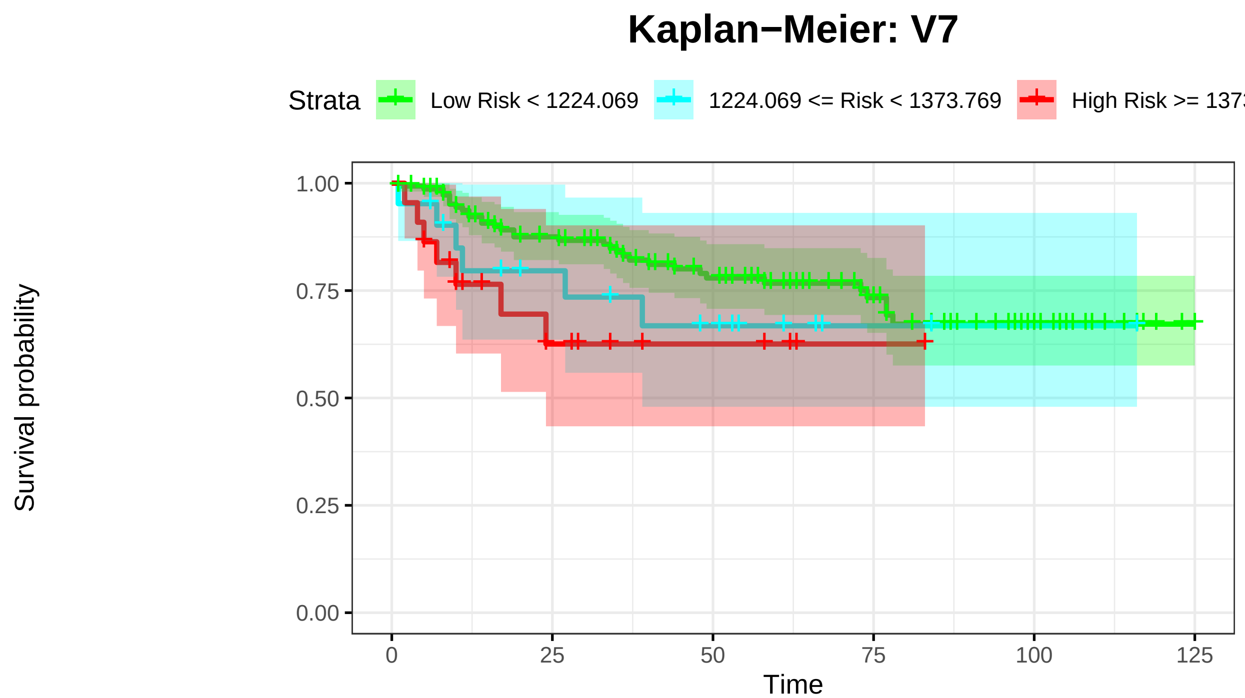


Number at risk

	0	25	50	75	100	125
Low Risk < 3.490	140	99	66	35	17	1
3.490 <= Risk < 4.990	31	13	11	4	2	0
High Risk >= 4.990	23	15	9	5	1	0

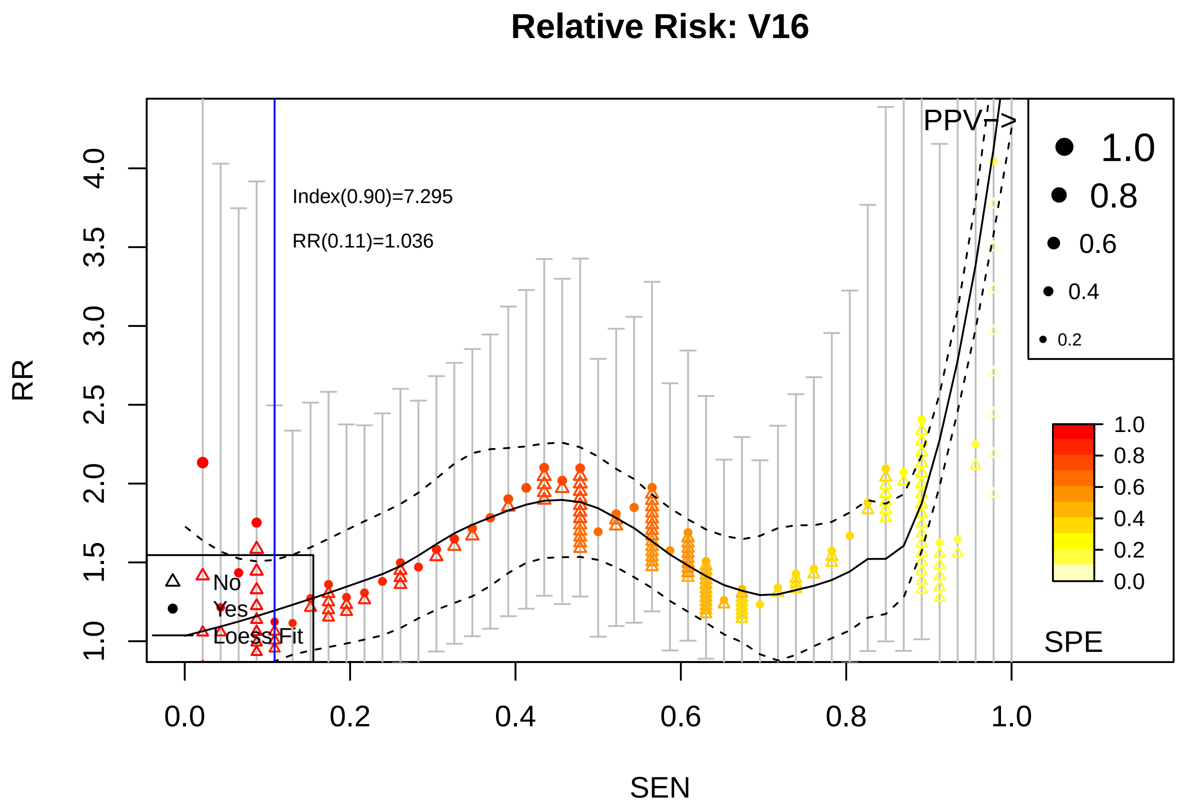

Relative Risk: V7
RR
SEN
Index(0.90)=1373.769
RR(0.15)=1.372
PPV-->
No
Yes
Loess Fit
SPE
1.0
0.8
0.6
0.4
0.2
0.0
1
2
3
4
5

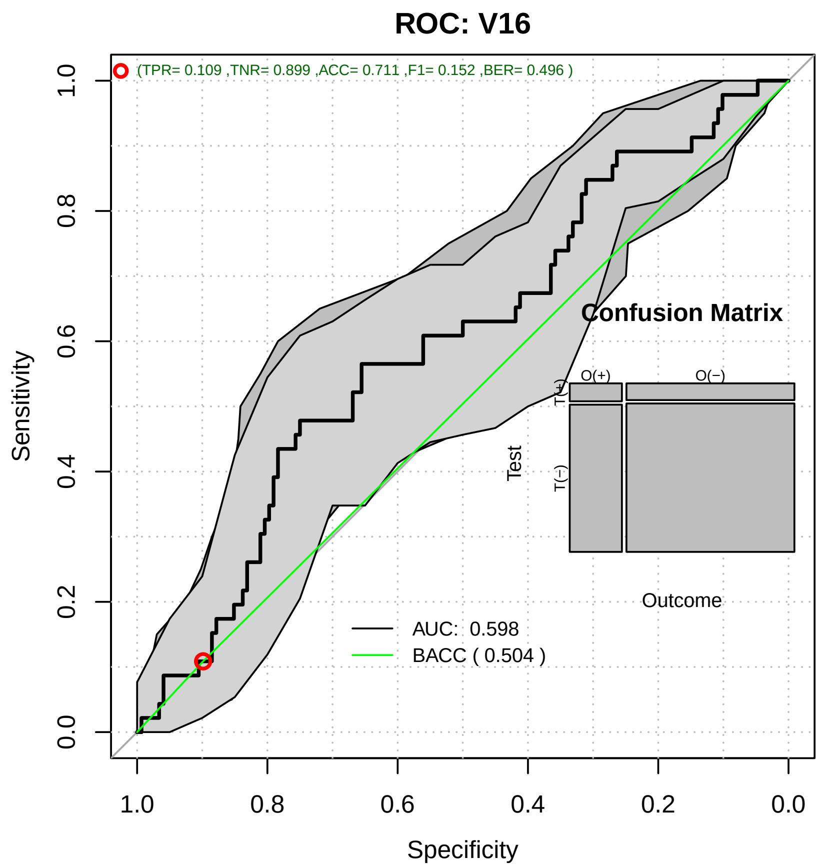

ROC: V7
(TPR= 0.152 ,TNR= 0.899 ,ACC= 0.722 ,F1= 0.206 ,BER= 0.475)
Sensitivity
Specificity
0.0
0.2
0.4
0.6
0.8
1.0
AUC: 0.610
BACC (0.525)
Confusion Matrix
O(+)
O(−)
T(+)
T(−)
Test
Outcome

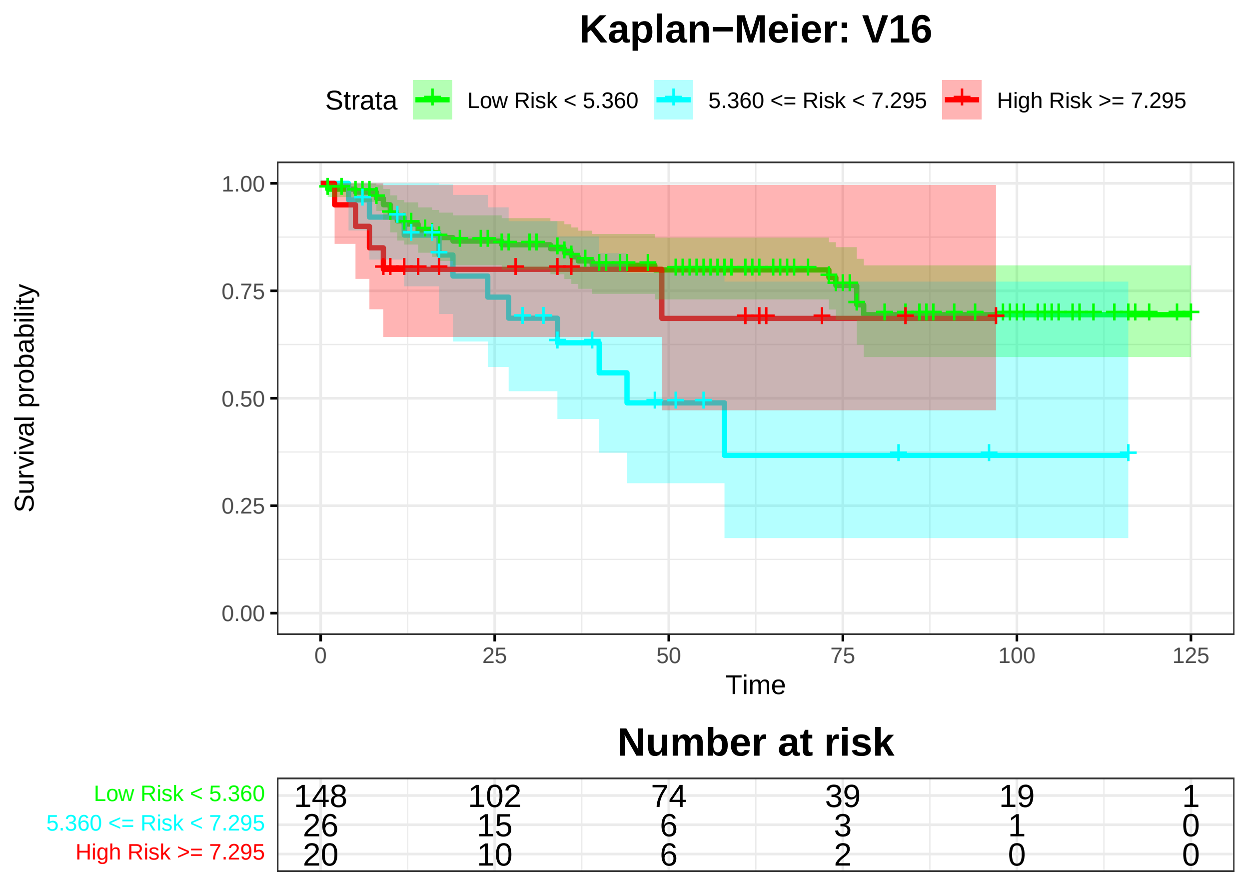


Number at risk

	0	25	50	75	100	125
Low Risk < 1224.069	151	106	73	41	19	1
1224.069 <= Risk < 1373.769	21	13	9	2	1	0
High Risk >= 1373.769	22	8	4	1	0	0


Relative Risk: V16
Index(0.90)=7.295
RR(0.11)=1.036
PPV-->
1.0
0.8
0.6
0.4
0.2
No
Yes
Loess Fit
SPE
1.0
0.8
0.6
0.4
0.2
0.0
RR
SEN
0.0
0.2
0.4
0.6
0.8
1.0
1.0
1.5
2.0
2.5
3.0
3.5
4.0


ROC: V16
(TPR= 0.109 ,TNR= 0.899 ,ACC= 0.711 ,F1= 0.152 ,BER= 0.496)
Sensitivity
Specificity
1.0
0.8
0.6
0.4
0.2
0.0
Confusion Matrix
O(+)
O(−)
T(+)
T(−)
Test
Outcome
AUC: 0.598
BACC (0.504)




```
names(RRanalysis) <- topFive
```

1.2 Reporting the Metrics

```
pander::pander(RRanalysis[[1]]$keyPoints,caption=topFive[1])
```

Table 2: V35

	Thr	RR	RR_LCI	RR_UCI	SEN	SPE	BACC
@:0.9	1.00e+01	1.57	0.8370	2.93	0.174	0.8986	0.536
@:0.8	3.00e+00	2.32	1.4235	3.77	0.478	0.7770	0.628
@MAX_BACC	4.00e+00	2.64	1.6323	4.27	0.478	0.8108	0.645
@MAX_RR	6.59e-09	2.68	1.3264	5.42	0.826	0.4189	0.623
@SPE100	-8.34e-09	4.79	0.0113	2024.86	1.000	0.0135	0.507

```
pander::pander(RRanalysis[[2]]$keyPoints,caption=topFive[2])
```

Table 3: V24

	Thr	RR	RR_LCI	RR_UCI	SEN	SPE	BACC
@:0.9	25.4	1.94	1.1306	3.34	0.239	0.8919	0.566
@:0.8	23.9	1.67	1.0015	2.78	0.348	0.7905	0.569
@MAX_BACC	20.3	2.45	1.3530	4.44	0.739	0.5270	0.633

	Thr	RR	RR_LCI	RR_UCI	SEN	SPE	BACC
@MAX_RR	16.6	3.87	0.9914	15.08	0.957	0.1824	0.569
@SPE100	15.5	33.04	0.0685	15945.00	1.000	0.0878	0.544

```
RRanalysis[[2]]$keyPoints["@MAX_BACC",c("BACC","RR")]
```

```
           BACC       RR
```

@MAX_BACC 0.6330787 2.451923

```
ROCAUC <- NULL
CstatCI <- NULL
LogRangp <- NULL
Sensitivity <- NULL
Specificity <- NULL
MAXBACC <- NULL
RREst <- NULL

for (topf in topFive)
{
  CstatCI <- rbind(CstatCI,RRanalysis[[topf]]$c.index$cstatCI)
  LogRangp <- rbind(LogRangp,RRanalysis[[topf]]$surdif$pvalue)
  Sensitivity <- rbind(Sensitivity,RRanalysis[[topf]]$ROCAnalysis$sensitivity)
  Specificity <- rbind(Specificity,RRanalysis[[topf]]$ROCAnalysis$specificity)
  ROCAUC <- rbind(ROCAUC,RRanalysis[[topf]]$ROCAnalysis$aucs)
  MAXBACC <- rbind(MAXBACC,RRanalysis[[topf]]$keyPoints["@MAX_BACC",c("BACC")])
  RREst <- rbind(RREst,RRanalysis[[topf]]$keyPoints[1,c("RR")])
}
rownames(CstatCI) <- topFive
rownames(LogRangp) <- topFive
rownames(Sensitivity) <- topFive
rownames(Specificity) <- topFive
rownames(ROCAUC) <- topFive
rownames(MAXBACC) <- topFive
rownames(RREst) <- topFive

pander::pander(ROCAUC)
```

	est	lower	upper
V35	0.652	0.560	0.745
V24	0.634	0.542	0.725
V34	0.658	0.571	0.744
V7	0.610	0.515	0.705
V16	0.598	0.504	0.692

```
pander::pander(CstatCI)
```

	mean.C Index	median	lower	upper
V35	0.629	0.631	0.538	0.720
V24	0.677	0.680	0.594	0.762
V34	0.656	0.655	0.579	0.729

	mean.C Index	median	lower	upper
V7	0.667	0.667	0.584	0.749
V16	0.614	0.615	0.525	0.700

```
pander::pander(LogRangp)
```

V35	0.00104
V24	0.00938
V34	0.00282
V7	0.07332
V16	0.02135

```
pander::pander(Sensitivity)
```

	est	lower	upper
V35	0.152	0.0634	0.289
V24	0.239	0.1259	0.388
V34	0.152	0.0634	0.289
V7	0.152	0.0634	0.289
V16	0.109	0.0362	0.236

```
pander::pander(Specificity)
```

	est	lower	upper
V35	0.899	0.838	0.942
V24	0.899	0.838	0.942
V34	0.892	0.830	0.937
V7	0.899	0.838	0.942
V16	0.899	0.838	0.942

```
pander::pander(MAXBACC)
```

V35	0.645
V24	0.633
V34	0.637
V7	0.621
V16	0.614

```
pander::pander(RREst)
```

V35	1.57
V24	1.94
V34	1.33
V7	1.33
V16	1.00

```r
meanMatrix <- cbind(ROCAUC[,1],CstatCI[,1],RREst,Sensitivity[,1],Specificity[,1],MAXBACC)
colnames(meanMatrix) <- c("ROCAUC","C-Stat","RR","Sen","Spe","MAX_BACC")
pander::pander(meanMatrix)
```

	ROCAUC	C-Stat	RR	Sen	Spe	MAX_BACC
V35	0.652	0.629	1.57	0.152	0.899	0.645
V24	0.634	0.677	1.94	0.239	0.899	0.633
V34	0.658	0.656	1.33	0.152	0.892	0.637
V7	0.610	0.667	1.33	0.152	0.899	0.621
V16	0.598	0.614	1.00	0.109	0.899	0.614

1.3 Modeling

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

[++]. . . . .

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

Table 12: Table continues below

	Estimate	lower	HR	upper	u.Accuracy	r.Accuracy
V24	5.28e-02	1.02	1.05	1.09	0.598	0.241
V26	5.05e-03	1.00	1.01	1.01	0.593	0.271
V27	2.04e-04	1.00	1.00	1.00	0.608	0.293
V34	1.13e-02	1.00	1.01	1.02	0.634	0.324
V7	6.44e-08	1.00	1.00	1.00	0.588	0.272
V35	2.14e-03	1.00	1.00	1.00	0.727	0.602
V6	1.07e-07	1.00	1.00	1.00	0.577	0.237

Table 13: Table continues below

	full.Accuracy	u.AUC	r.AUC	full.AUC	IDI	NRI	z.IDI
V24	0.598	0.609	0.501	0.609	0.0618	0.435	2.86
V26	0.594	0.598	0.510	0.600	0.0624	0.394	2.76
V27	0.609	0.608	0.516	0.607	0.0561	0.434	2.75
V34	0.628	0.618	0.525	0.615	0.0302	0.460	2.36
V7	0.589	0.595	0.510	0.596	0.0483	0.380	2.28
V35	0.616	0.641	0.604	0.604	0.0283	0.551	2.26
V6	0.577	0.588	0.500	0.588	0.0459	0.353	2.19

	z.NRI	Delta.AUC	Frequency
V24	2.66	0.10788	1.0
V26	2.39	0.08956	1.0
V27	2.63	0.09082	1.0
V34	2.78	0.08993	1.0
V7	2.30	0.08552	0.9

	z.NRI	Delta.AUC	Frequency
V35	3.41	-0.00042	1.0
V6	2.13	0.08813	0.1

1.4 Cox Model Performance

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

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

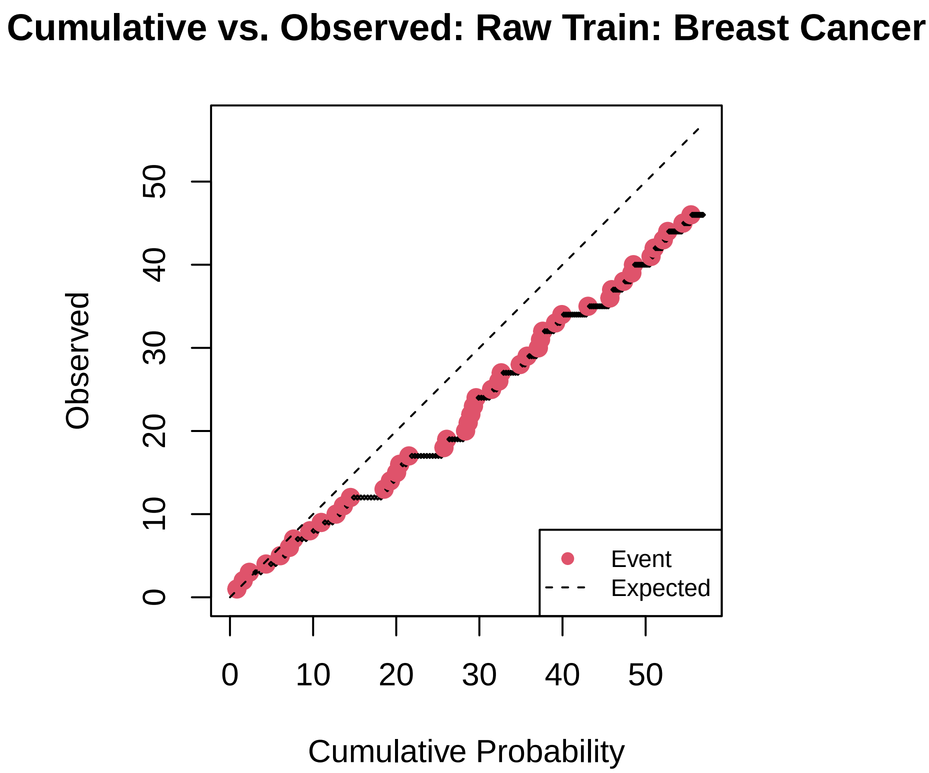
Here we will use the predicted event probability assuming a baseline hazard for events withing 5 years

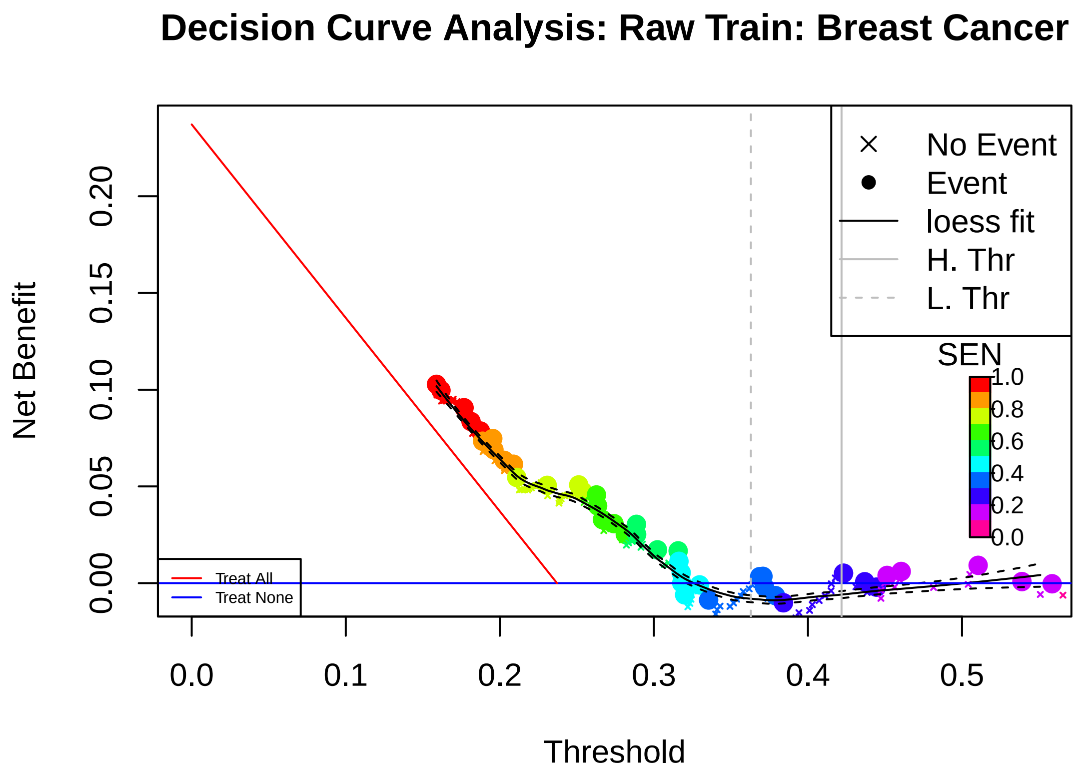
```
index <- predict(ml,dataBreast)
timeinterval <- 2*mean(subset(dataBreast,status==1)$time)

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

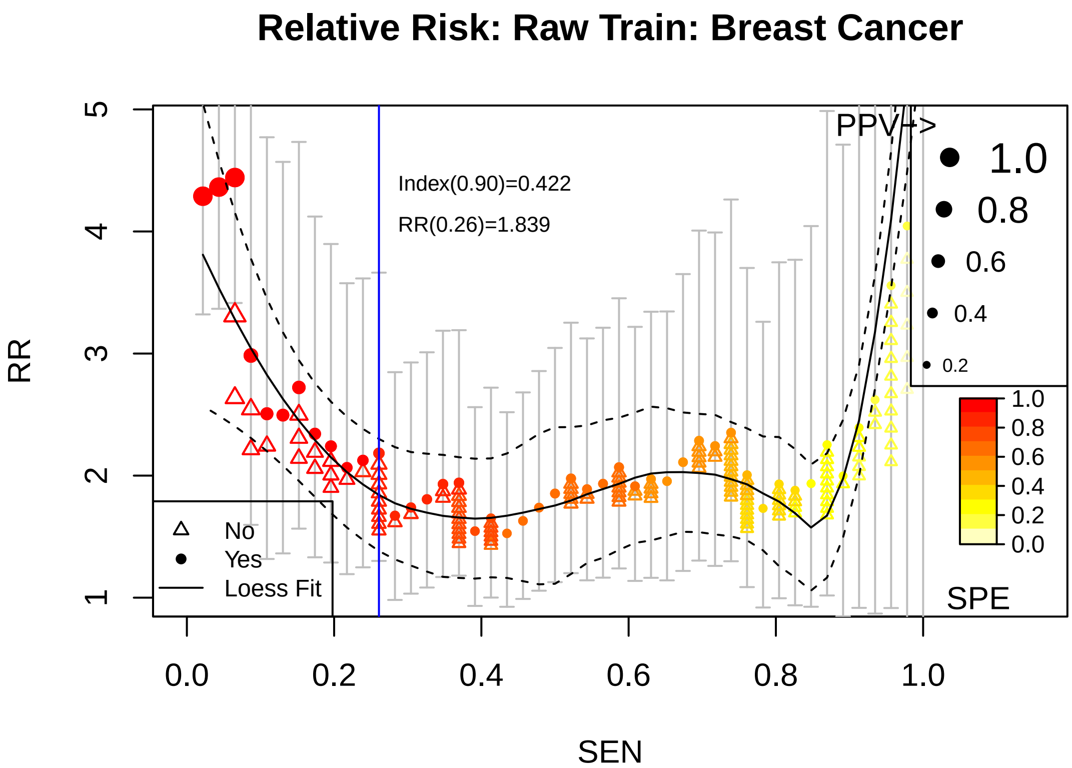
Table 15: Initial Parameters

h0	timeinterval
0.323	51.1

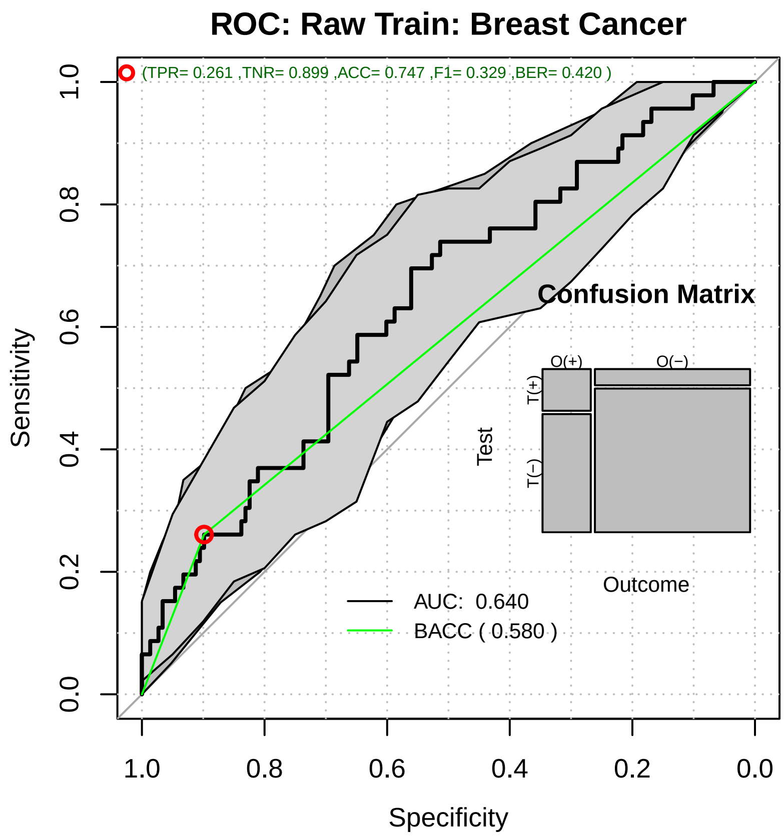
```
rdata <- cbind(dataBreast$status,ppoisGzero(index,h0))
rownames(rdata) <- rownames(dataBreast)

rrAnalysisTrain <- RRPlot(rdata,atRate=c(0.90,0.80),
                     timetoEvent=dataBreast$time,
                     title="Raw Train: Breast Cancer",
                     ysurvlim=c(0.00,1.0),
                     riskTimeInterval=timeinterval)
```

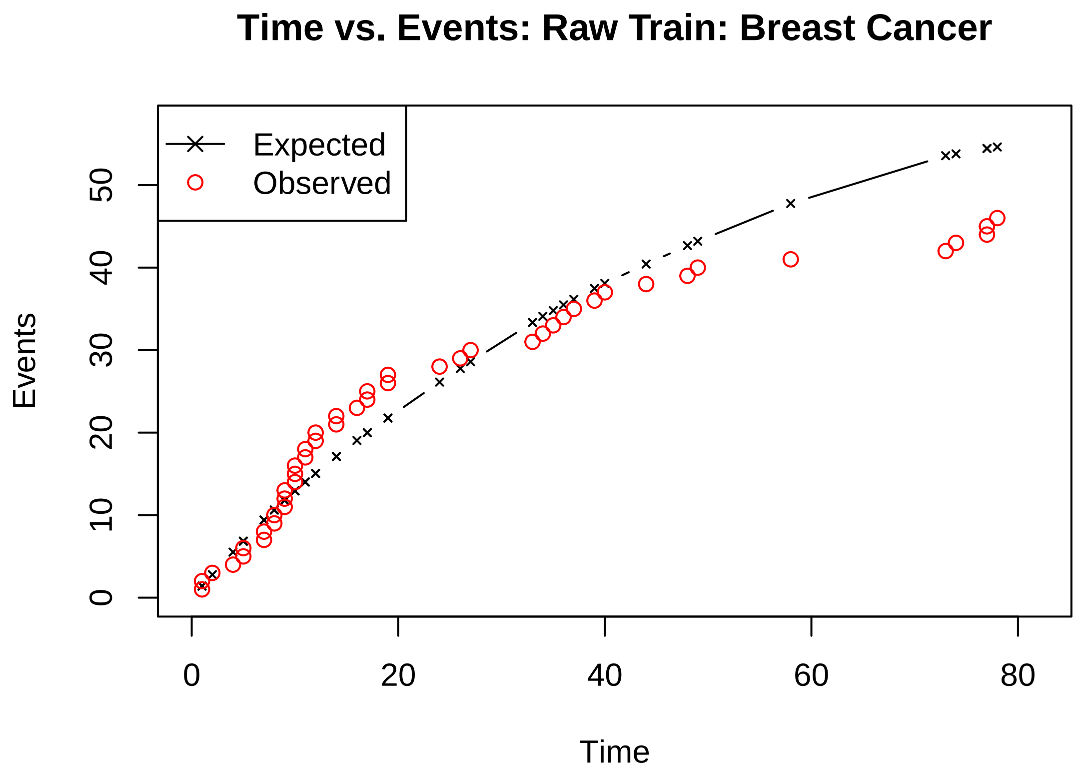

Cumulative vs. Observed: Raw Train: Breast Cancer
Observed
Cumulative Probability
0
10
20
30
40
50
0
10
20
30
40
50
Event
Expected

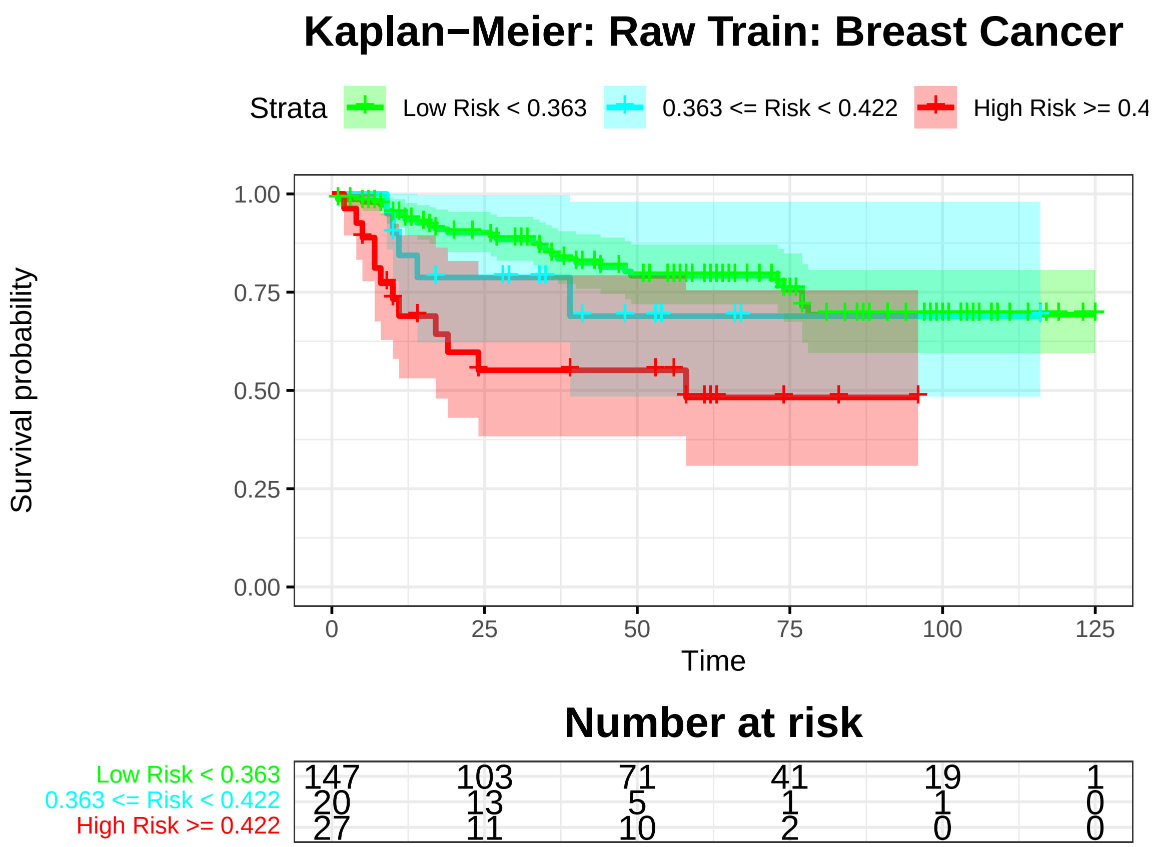

Decision Curve Analysis: Raw Train: Breast Cancer
Net Benefit
0.00
0.05
0.10
0.15
0.20
0.0
0.1
0.2
0.3
0.4
0.5
Threshold
No Event
Event
loess fit
H. Thr
L. Thr
SEN
1.0
0.8
0.6
0.4
0.2
0.0
Treat All
Treat None



Relative Risk: Raw Train: Breast Cancer
Index(0.90)=0.422
RR(0.26)=1.839
PPV->
1.0
0.8
0.6
0.4
0.2
SPE
No
Yes
Loess Fit
RR
SEN


ROC: Raw Train: Breast Cancer
(TPR= 0.261 ,TNR= 0.899 ,ACC= 0.747 ,F1= 0.329 ,BER= 0.420)
Confusion Matrix
O(+)
O(−)
T(+)
T(−)
Test
Outcome
AUC: 0.640
BACC (0.580)
Sensitivity
Specificity
1.0
0.8
0.6
0.4
0.2
0.0


Time vs. Events: Raw Train: Breast Cancer
Expected
Observed
Events
0
10
20
30
40
50
0
20
40
60
80
Time



1.4.2 Uncalibrated Performance Report

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

Table 16: Threshold values

	@:0.9	@:0.8	@MAX_BACC	@MAX_RR	@SPE100	p(0.5)
Thr	0.42310	0.363649	0.2627	0.1618	1.59e-01	0.504013
RR	2.18301	1.833456	2.2857	4.0449	2.50e+01	2.307692
RR_LCI	1.30105	1.111407	1.3037	0.5963	5.22e-02	1.275480
RR_UCI	3.66282	3.024598	4.0074	27.4389	1.20e+04	4.175246
SEN	0.26087	0.369565	0.6957	0.9783	1.00e+00	0.152174
SPE	0.89865	0.797297	0.5608	0.1014	6.76e-02	0.952703
BACC	0.57976	0.583431	0.6282	0.5398	5.34e-01	0.552438
NetBenefit	0.00516	-0.000727	0.0456	0.0997	1.03e-01	-0.000577

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

Table 17: O/E Test

O/E	Low	Upper	p.value
0.842	0.617	1.12	0.278

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

Table 18: O/E Mean

mean	50%	2.5%	97.5%
1.02	1.02	0.97	1.07

```r
pander::pander(t(rrAnalysisTrain$OARatio$estimate),caption="O/Acum Test")
```

Table 19: O/Acum Test

O/A	Low	Upper	p.value
0.809	0.592	1.08	0.163

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

Table 20: O/Acum Mean

mean	50%	2.5%	97.5%
0.79	0.79	0.784	0.796

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

Table 21: C. Index

mean.C Index	median	lower	upper
0.682	0.683	0.605	0.761

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

Table 22: ROC AUC

est	lower	upper
0.64	0.549	0.732

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

Table 23: Sensitivity

est	lower	upper
0.261	0.143	0.411

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

Table 24: Specificity

est	lower	upper
0.899	0.838	0.942

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

Table 25: Probability Thresholds

90%	80%
0.422	0.363

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

Table 26: Logrank test Chisq = 12.346960 on 2 degrees of freedom, p = 0.002084

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	147	29	36.99	1.725	8.974
class=1	20	5	4.11	0.193	0.216
class=2	27	12	4.90	10.269	11.609