Breast Cancer: Wisconsin

Jose Tamez

2023-05-02

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
```

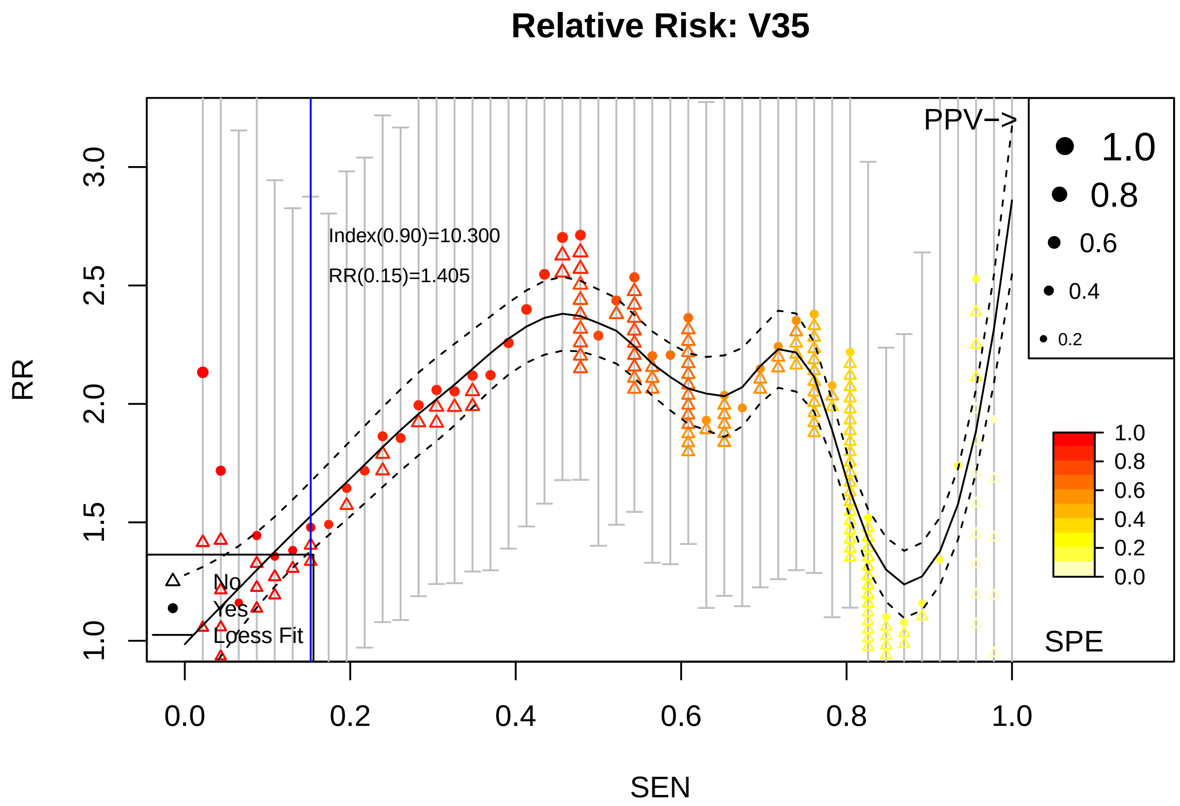
1.1 Exploring Raw Features with RRPlot

```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)
```

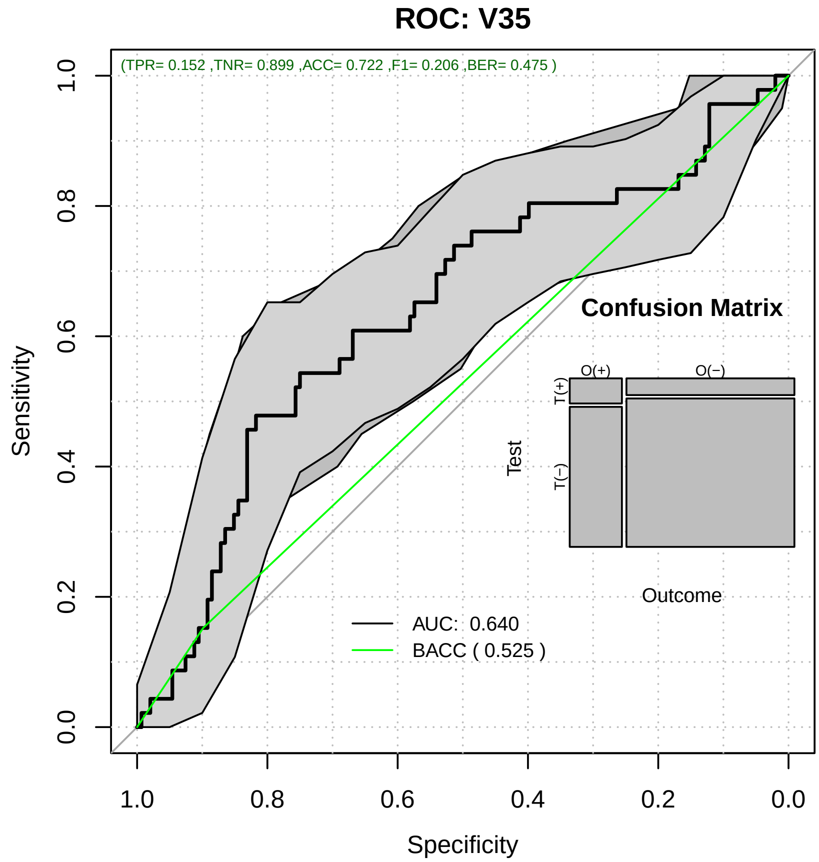
V35	V24	V34	V7	V16	V14	V17
0.0261	0.0261	0.0261	0.0623	0.126	0.126	0.126

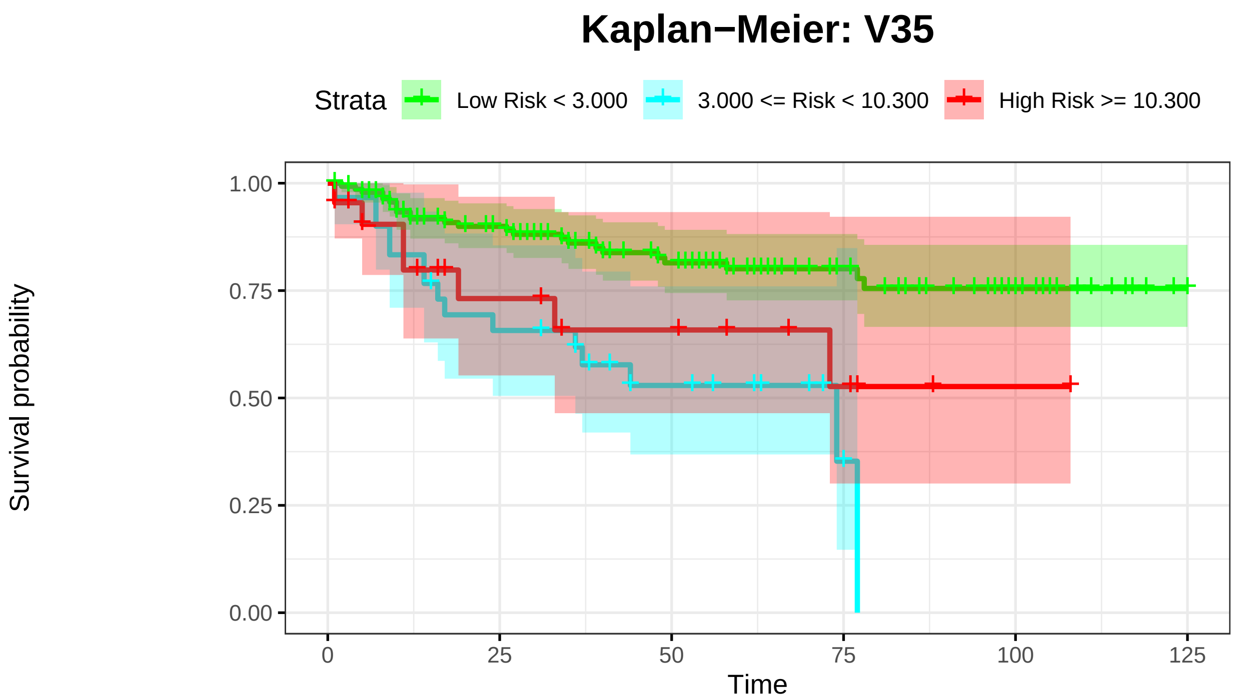
```r
topv <- min(5,length(topvar))
topFive <- names(topvar)[1:topv]
RRanalysis <- list();
idx <- 1
topf <- topFive[1]
for (topf in topFive)
{
  RRanalysis[[idx]] <- RRPlot(cbind(dataBreast$status,dataBreast[,topf]),
                              atProb=c(0.90,0.80),
                  timetoEvent=dataBreast$time,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

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



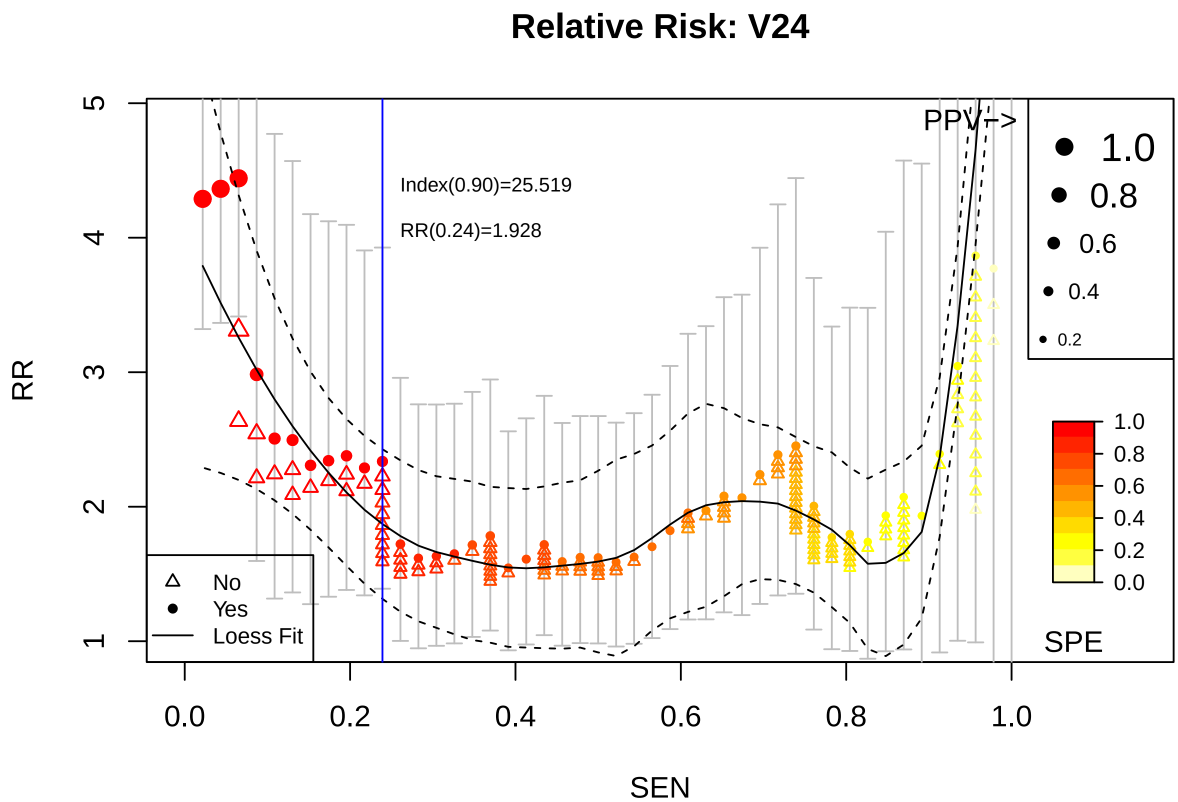
Relative Risk: V35
RR
SEN
Index(0.90)=10.300
RR(0.15)=1.405
PPV−>
1.0
0.8
0.6
0.4
0.2
No
Yes
Loess Fit
SPE


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: 0.640
BACC (0.525)

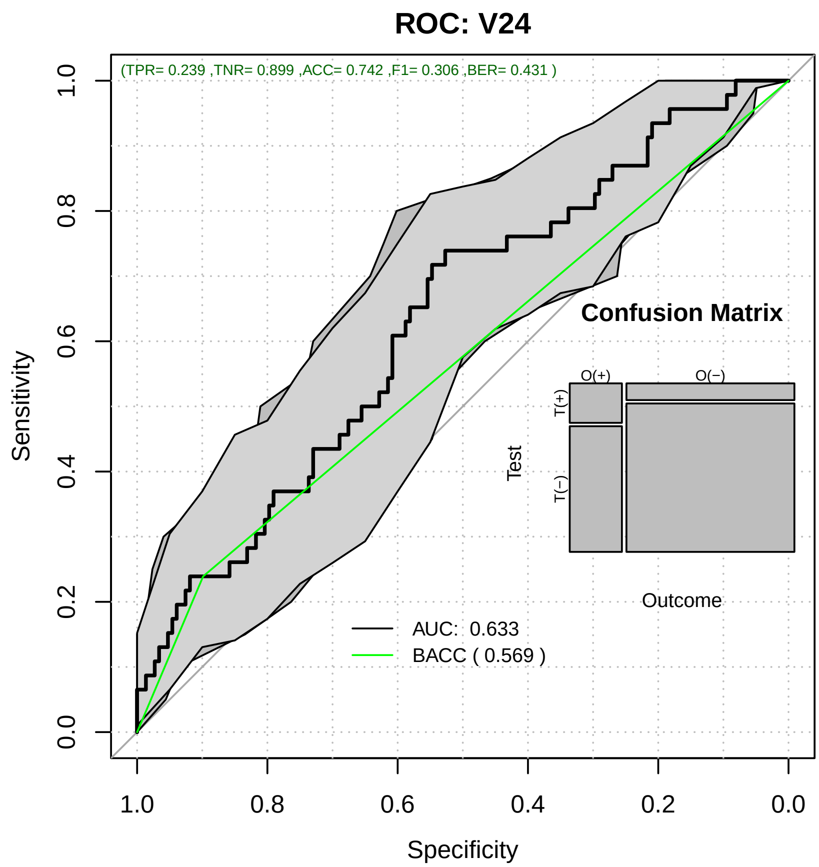


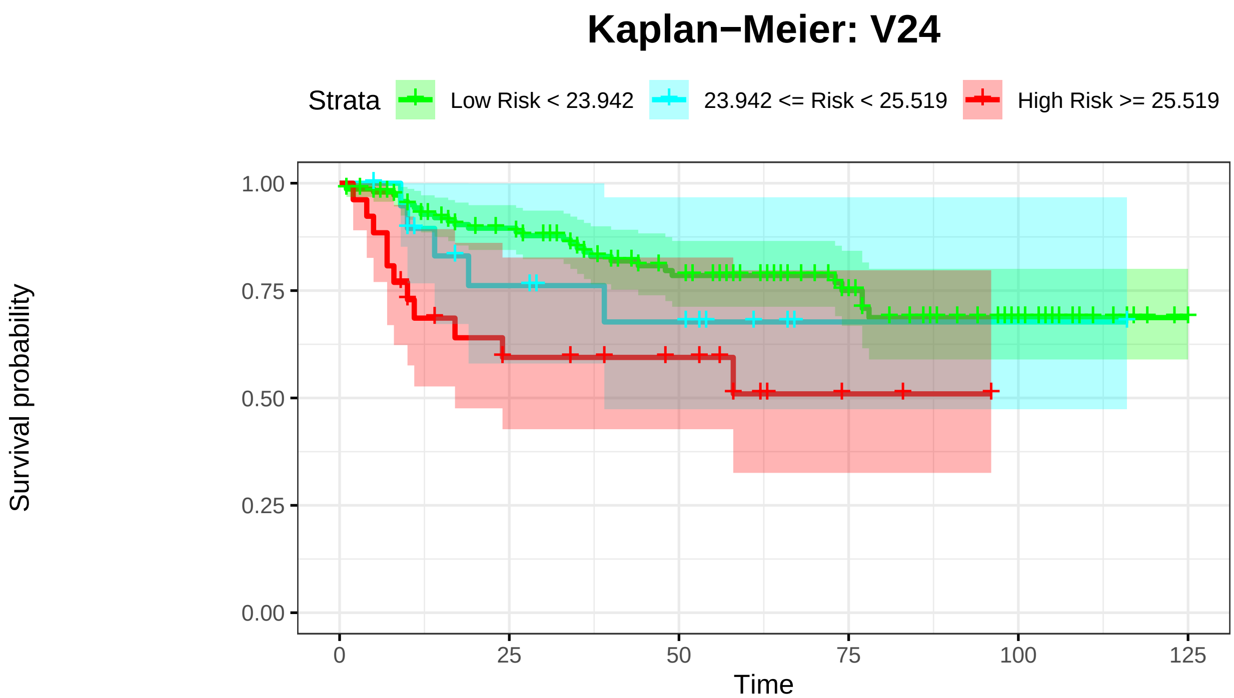
Number at risk

	0	25	50	75	100	125
Low Risk < 3.000	142	98	68	38	19	1
3.000 <= Risk < 10.300	30	18	10	2	0	0
High Risk >= 10.300	22	11	8	4	1	0



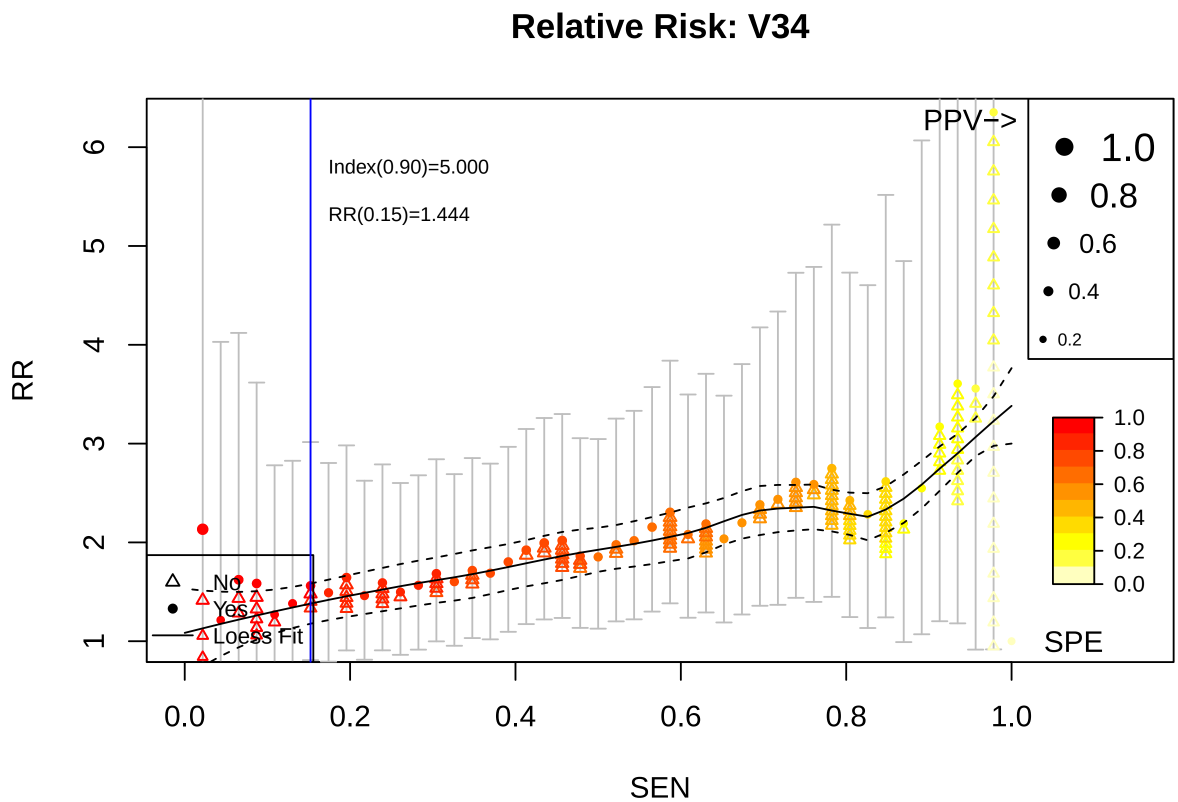
Relative Risk: V24
Index(0.90)=25.519
RR(0.24)=1.928
PPV-->
1.0
0.8
0.6
0.4
0.2
1.0
0.8
0.6
0.4
0.2
0.0
SPE
No
Yes
Loess Fit
RR
1
2
3
4
5
0.0
0.2
0.4
0.6
0.8
1.0
SEN


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: 0.633
BACC (0.569)

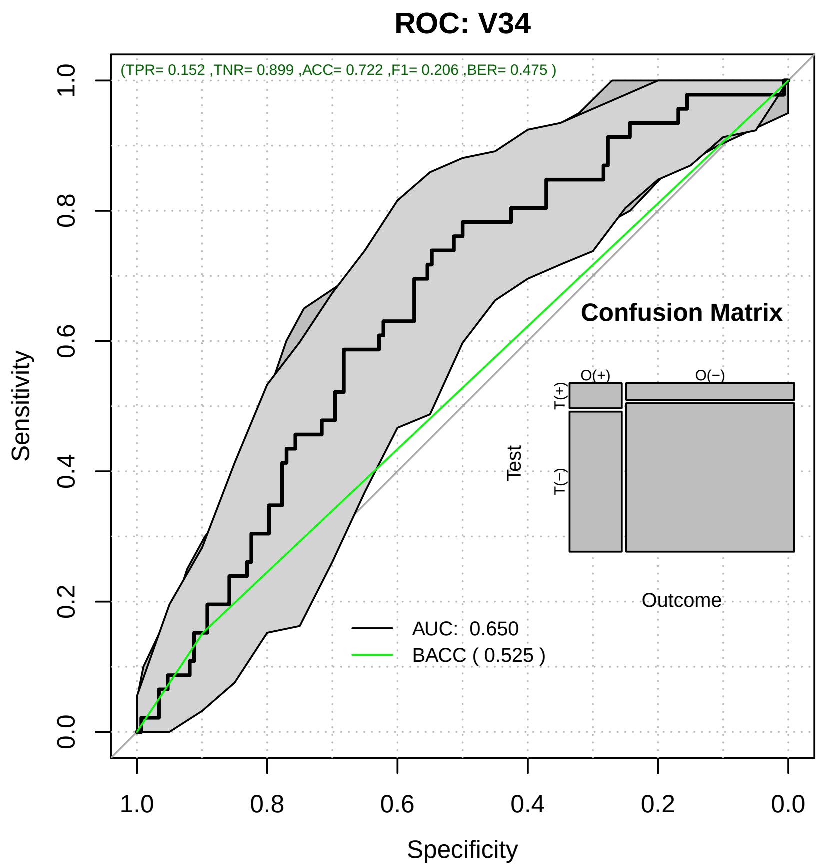


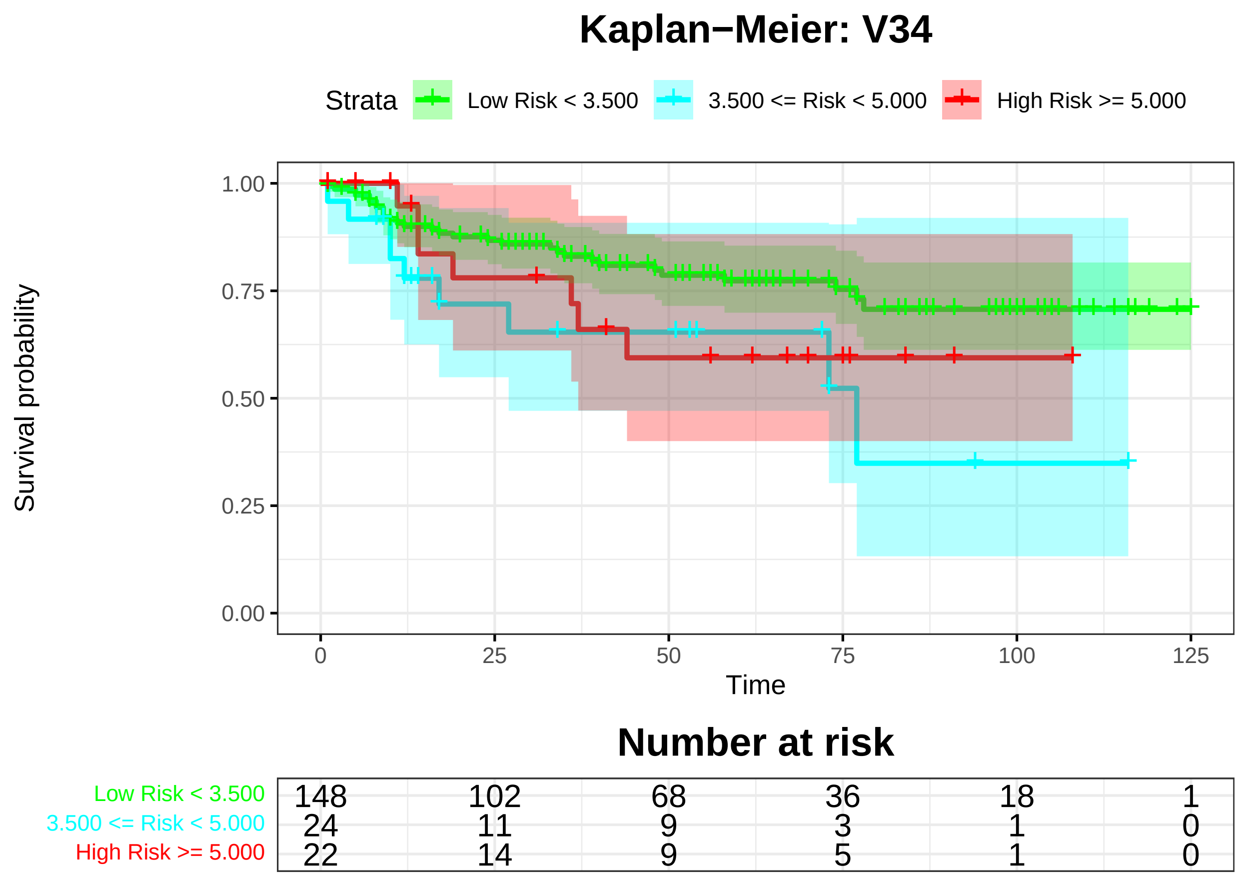
Number at risk

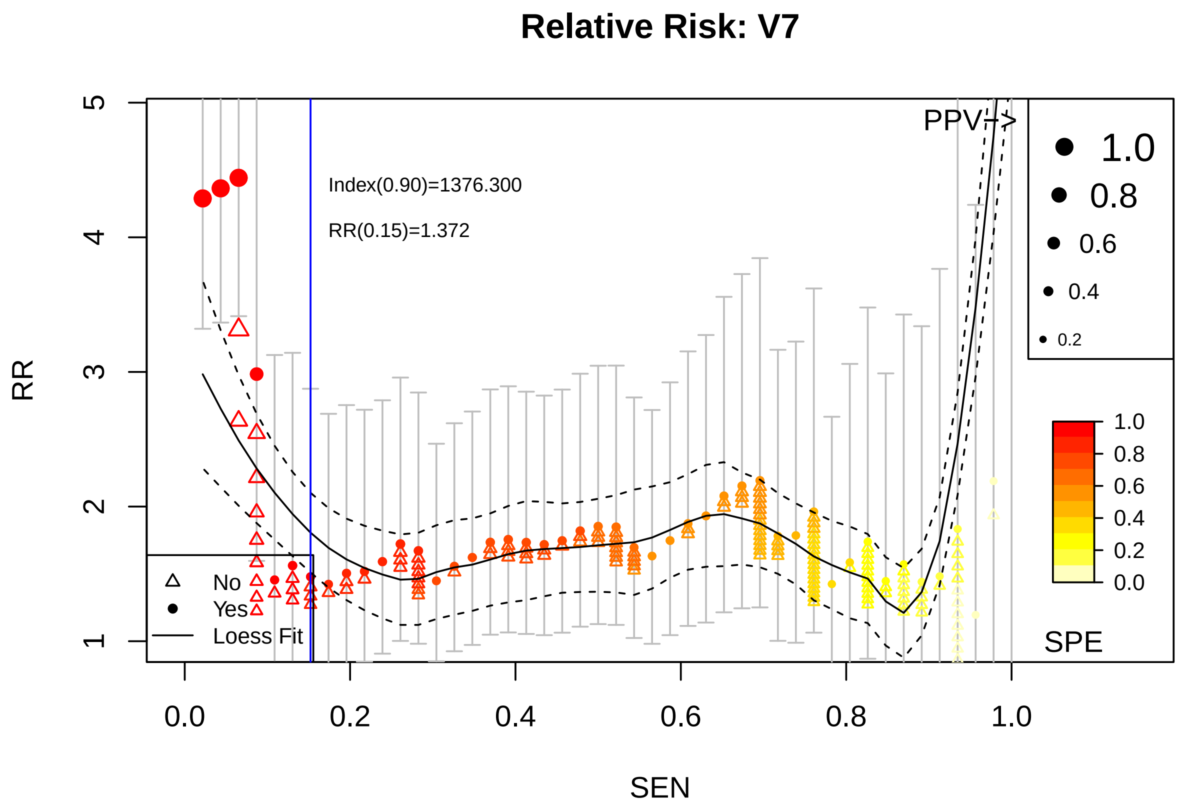
	0	25	50	75	100	125
Low Risk < 23.942	148	104	69	41	19	1
23.942 <= Risk < 25.519	20	11	8	1	1	0
High Risk >= 25.519	26	12	9	2	0	0



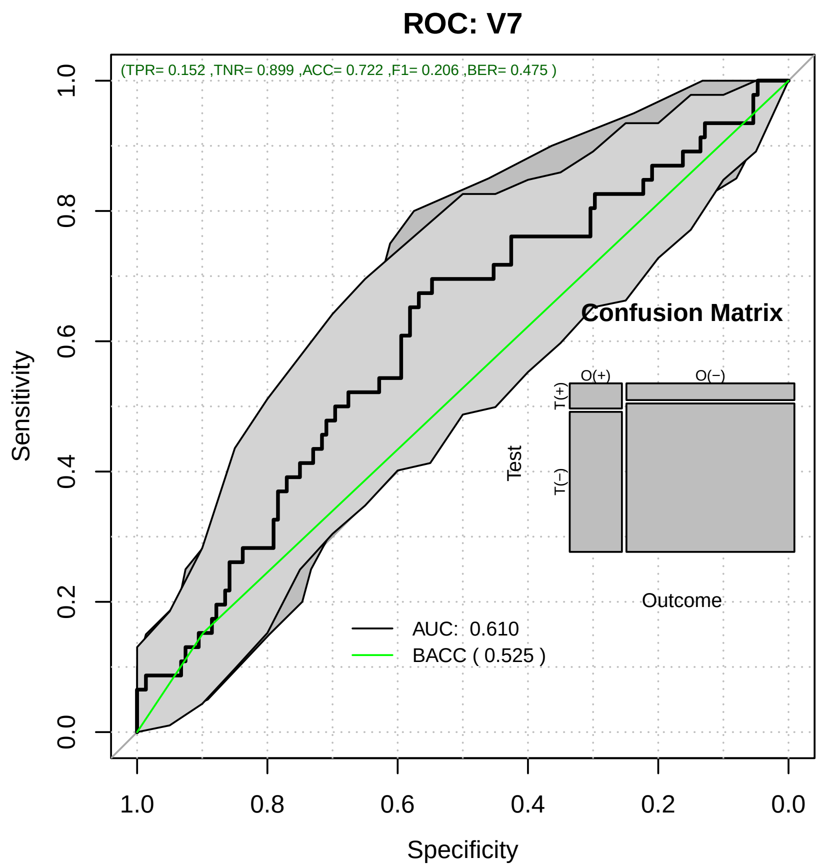
Relative Risk: V34
Index(0.90)=5.000
RR(0.15)=1.444
PPV−>
1.0
0.8
0.6
0.4
0.2
No
Yes
Loess Fit
SPE
RR
SEN

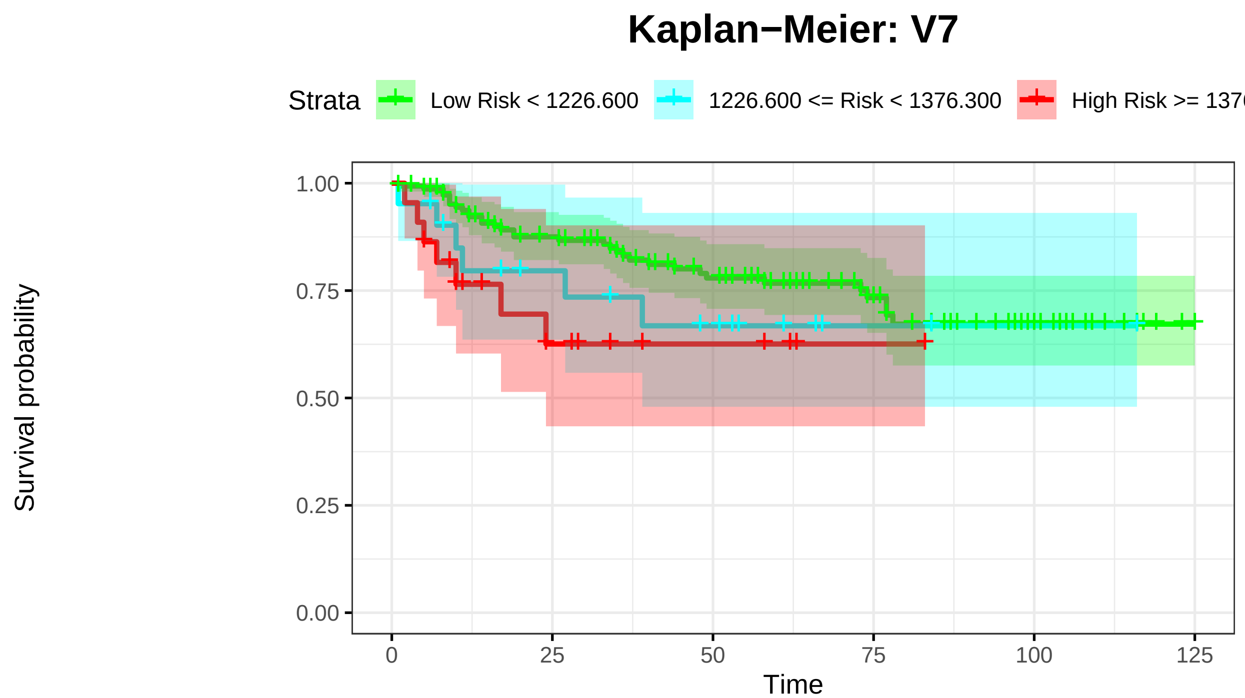

ROC: V34
(TPR= 0.152 ,TNR= 0.899 ,ACC= 0.722 ,F1= 0.206 ,BER= 0.475)
Sensitivity
Specificity
1.0
0.8
0.6
0.4
0.2
0.0
Confusion Matrix
O(+)
O(−)
T(+)
T(−)
Test
Outcome
AUC: 0.650
BACC (0.525)


Kaplan−Meier: V34
Strata
Low Risk < 3.500
3.500 <= Risk < 5.000
High Risk >= 5.000
Survival probability
1.00
0.75
0.50
0.25
0.00
0
25
50
75
100
125
Time
Number at risk
Low Risk < 3.500 148 102 68 36 18 1
3.500 <= Risk < 5.000 24 11 9 3 1 0
High Risk >= 5.000 22 14 9 5 1 0



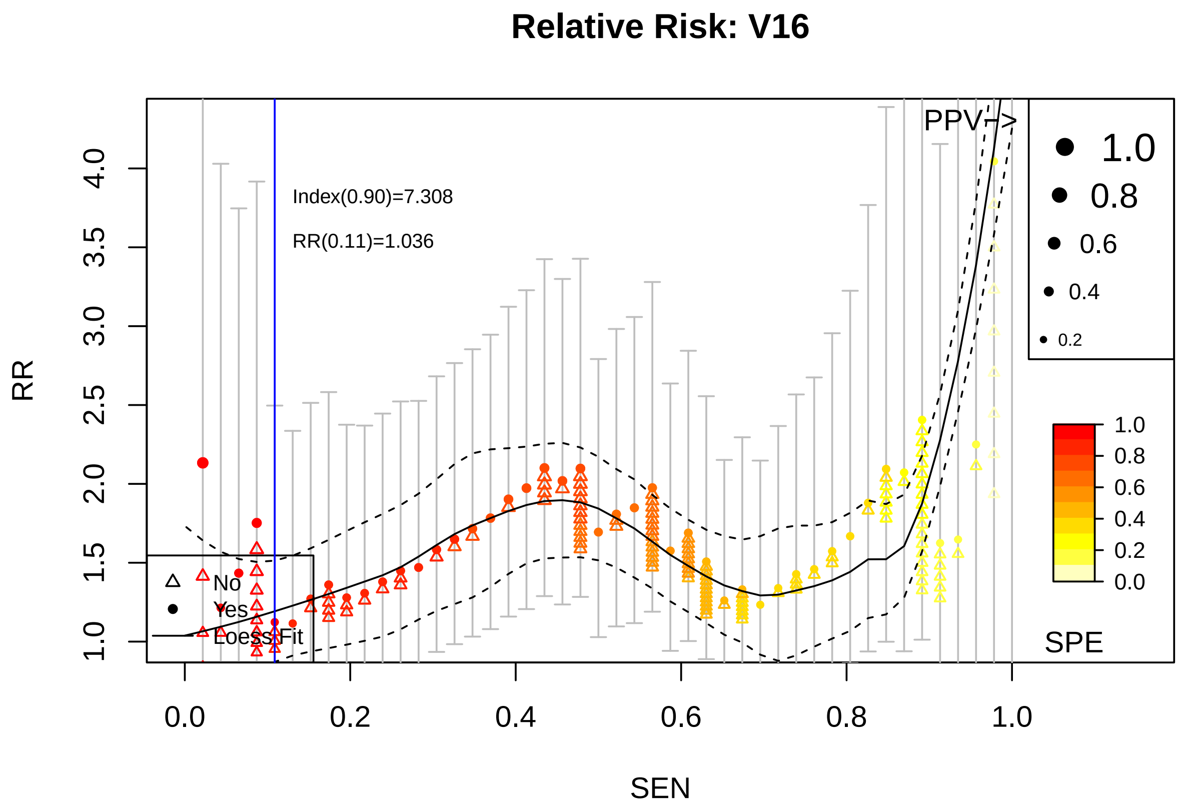
Relative Risk: V7
RR
SEN
5
4
3
2
1
0.0
0.2
0.4
0.6
0.8
1.0
Index(0.90)=1376.300
RR(0.15)=1.372
PPV-->
1.0
0.8
0.6
0.4
0.2
1.0
0.8
0.6
0.4
0.2
0.0
SPE
No
Yes
Loess Fit


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

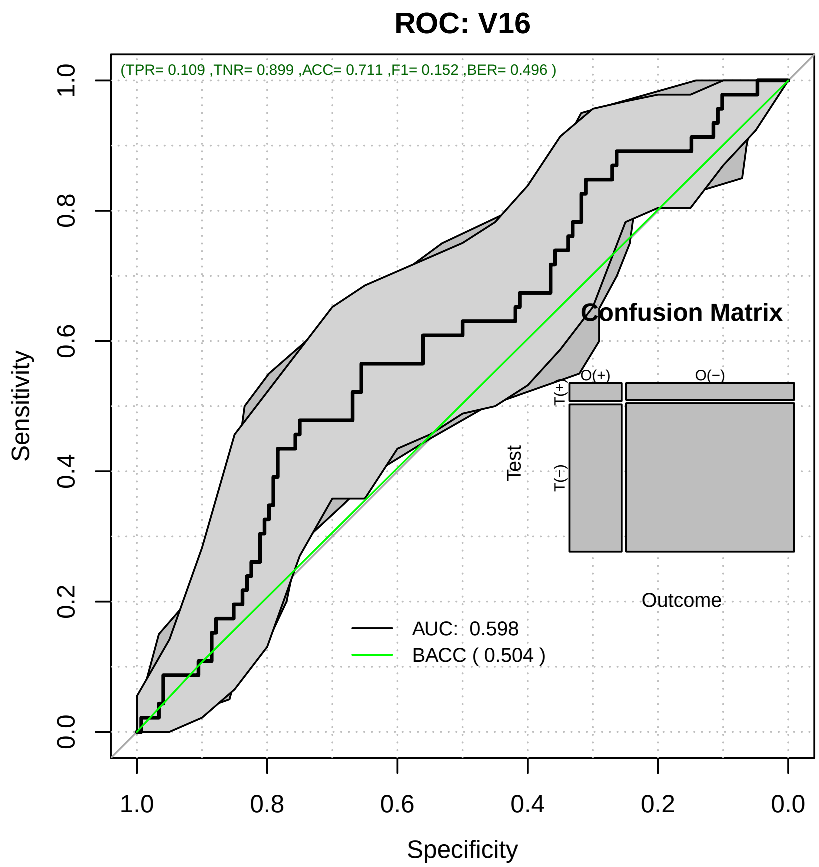


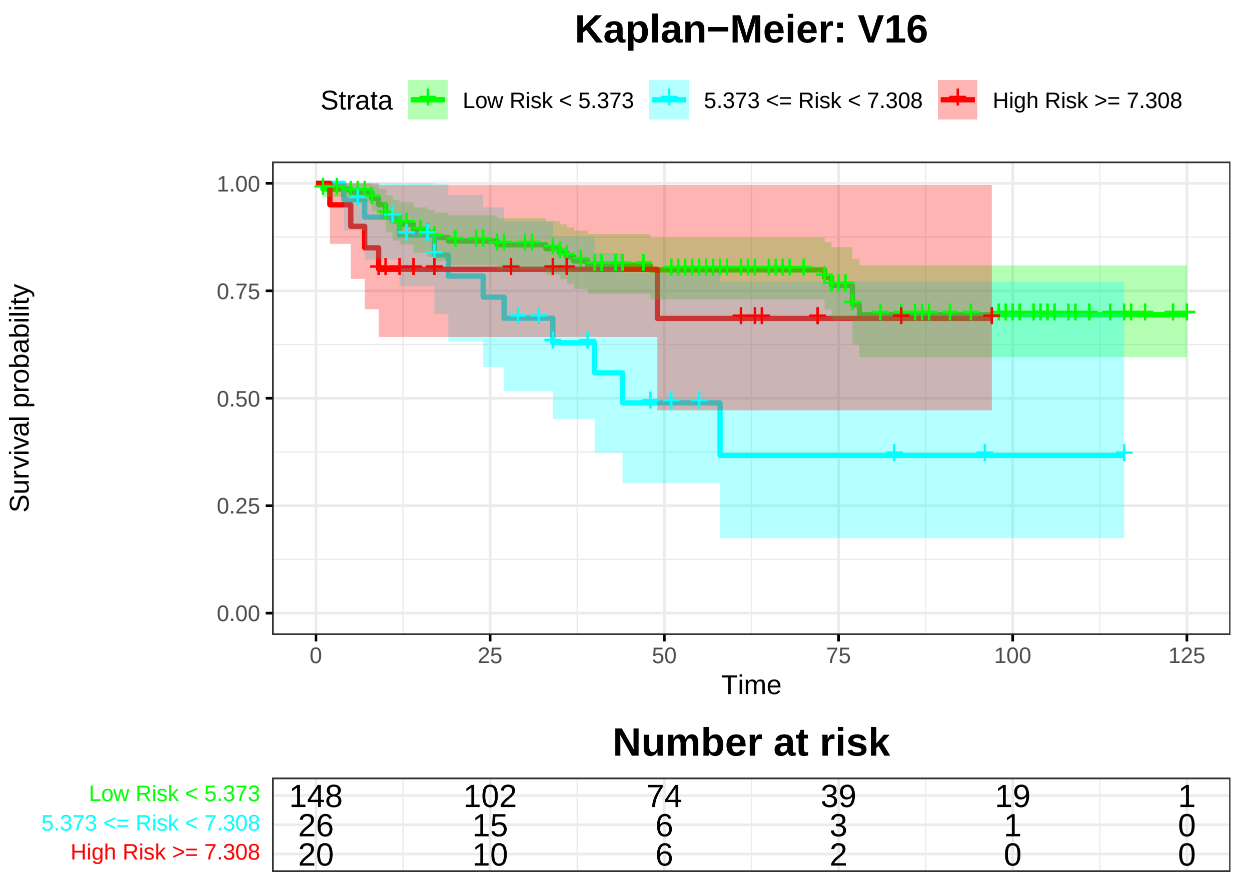
Number at risk

	0	25	50	75	100	125
Low Risk < 1226.600	151	106	73	41	19	1
1226.600 <= Risk < 1376.300	21	13	9	2	1	0
High Risk >= 1376.300	22	8	4	1	0	0



Relative Risk: V16
Index(0.90)=7.308
RR(0.11)=1.036
PPV->
1.0
0.8
0.6
0.4
0.2
SPE
No
Yes
Loess Fit
RR
SEN


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	SEN	SPE	BACC
@:0.9	1.00e+01	1.33	0.152	0.8919	0.522
@:0.8	3.00e+00	2.50	0.478	0.7973	0.638
@MAX_BACC	4.00e+00	2.71	0.478	0.8176	0.648
@MAX_RR	4.00e+00	2.71	0.478	0.8176	0.648
@SPE100	-8.41e-09	7.23	1.000	0.0203	0.510

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

Table 3: V24

	Thr	RR	SEN	SPE	BACC
@:0.9	25.4	1.94	0.239	0.8919	0.566
@:0.8	24.0	1.72	0.348	0.7973	0.573
@MAX_BACC	20.3	2.45	0.739	0.5270	0.633

	Thr	RR	SEN	SPE	BACC
@MAX_RR	16.6	3.87	0.957	0.1824	0.569
@SPE100	15.5	30.33	1.000	0.0811	0.541

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

```
           BACC        RR
```

@MAX_BACC 0.6330787 2.451923

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

for (topf in topFive)
{
  CstatCI <- rbind(CstatCI,RRanalysis[[topf]]$c.index$cstatCI)
  RRatios <- rbind(RRatios,RRanalysis[[topf]]$RR_atP)
  LogRangp <- rbind(LogRangp,RRanalysis[[topf]]$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")])
}
rownames(CstatCI) <- topFive
rownames(RRatios) <- topFive
rownames(LogRangp) <- topFive
rownames(Sensitivity) <- topFive
rownames(Specificity) <- topFive
rownames(ROCAUC) <- topFive
rownames(MAXBACC) <- topFive

pander::pander(ROCAUC)
```

	est	lower	upper
V35	0.640	0.544	0.736
V24	0.633	0.542	0.725
V34	0.650	0.563	0.736
V7	0.610	0.515	0.705
V16	0.598	0.504	0.692

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

	mean.C Index	median	lower	upper
V35	0.623	0.623	0.525	0.709
V24	0.677	0.676	0.597	0.753
V34	0.651	0.650	0.578	0.718

	mean.C Index	median	lower	upper
V7	0.667	0.664	0.579	0.740
V16	0.614	0.615	0.527	0.702

```
pander::pander(RRatios)
```

	est	lower	upper
V35	1.41	0.719	2.75
V24	1.93	1.122	3.31
V34	1.44	0.741	2.82
V7	1.37	0.700	2.69
V16	1.04	0.462	2.32

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

V35	0.000128
V24	0.009376
V34	0.055324
V7	0.073322
V16	0.021345

```
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	0.899	0.838	0.942
V7	0.899	0.838	0.942
V16	0.899	0.838	0.942

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

V35	0.648
V24	0.633
V34	0.643
V7	0.621

V16	0.614

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

	ROCAUC	C-Stat	Sen	Spe	RR	MAX_BACC
V35	0.640	0.623	0.152	0.899	1.41	0.648
V24	0.633	0.677	0.239	0.899	1.93	0.633
V34	0.650	0.651	0.152	0.899	1.44	0.643
V7	0.610	0.667	0.152	0.899	1.37	0.621
V16	0.598	0.614	0.109	0.899	1.04	0.614

1.3 Modeling

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

[++]....

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

Table 12: Table continues below

	Estimate	lower	HR	upper	u.Accuracy	r.Accuracy
V24	5.85e-02	1.02	1.06	1.10	0.598	0.237
V27	2.06e-04	1.00	1.00	1.00	0.608	0.293
V26	4.16e-03	1.00	1.00	1.01	0.593	0.376
V34	1.07e-02	1.00	1.01	1.02	0.634	0.305
V7	5.28e-08	1.00	1.00	1.00	0.588	0.237
V35	4.63e-03	1.00	1.00	1.01	0.727	0.597
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	0.500	0.609	0.0619	0.437	2.87
V27	0.609	0.608	0.516	0.607	0.0561	0.434	2.75
V26	0.600	0.598	0.540	0.603	0.0621	0.400	2.75
V34	0.630	0.618	0.519	0.615	0.0307	0.463	2.38
V7	0.588	0.595	0.500	0.595	0.0487	0.380	2.30
V35	0.617	0.641	0.601	0.610	0.0279	0.551	2.25
V6	0.577	0.588	0.500	0.588	0.0459	0.353	2.19

	z.NRI	Delta.AUC	Frequency
V24	2.67	0.10914	1.0
V27	2.63	0.09082	1.0

	z.NRI	Delta.AUC	Frequency
V26	2.43	0.06298	1.0
V34	2.80	0.09612	0.9
V7	2.30	0.09489	0.7
V35	3.41	0.00877	0.9
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()

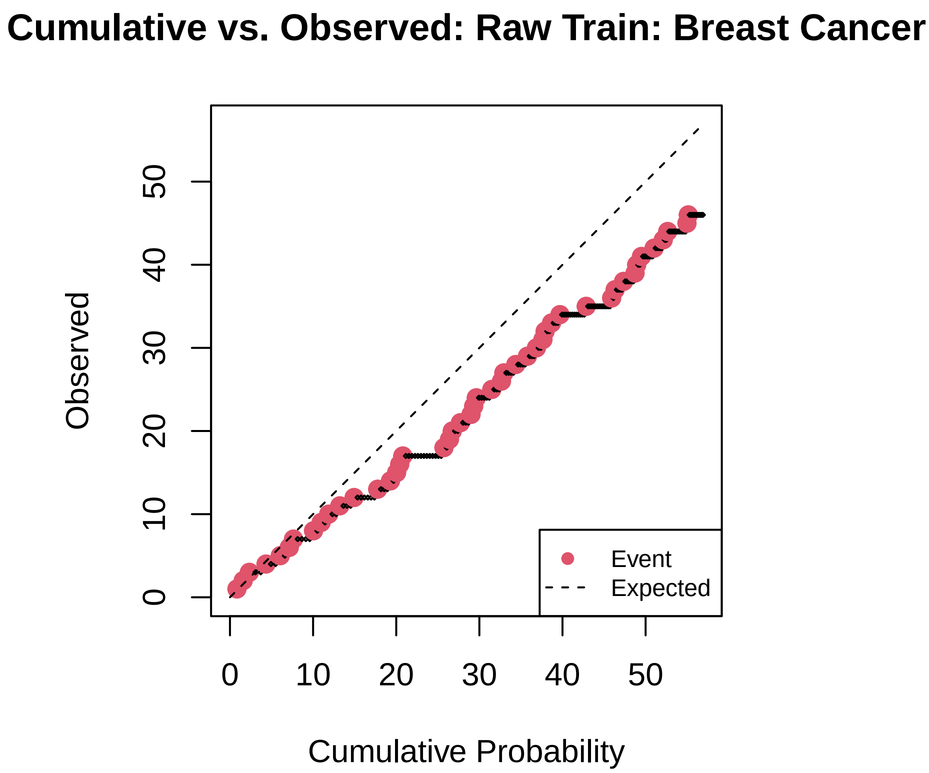
Here we will use the predicted event probability assuming a baseline hazard for events withing 5 years

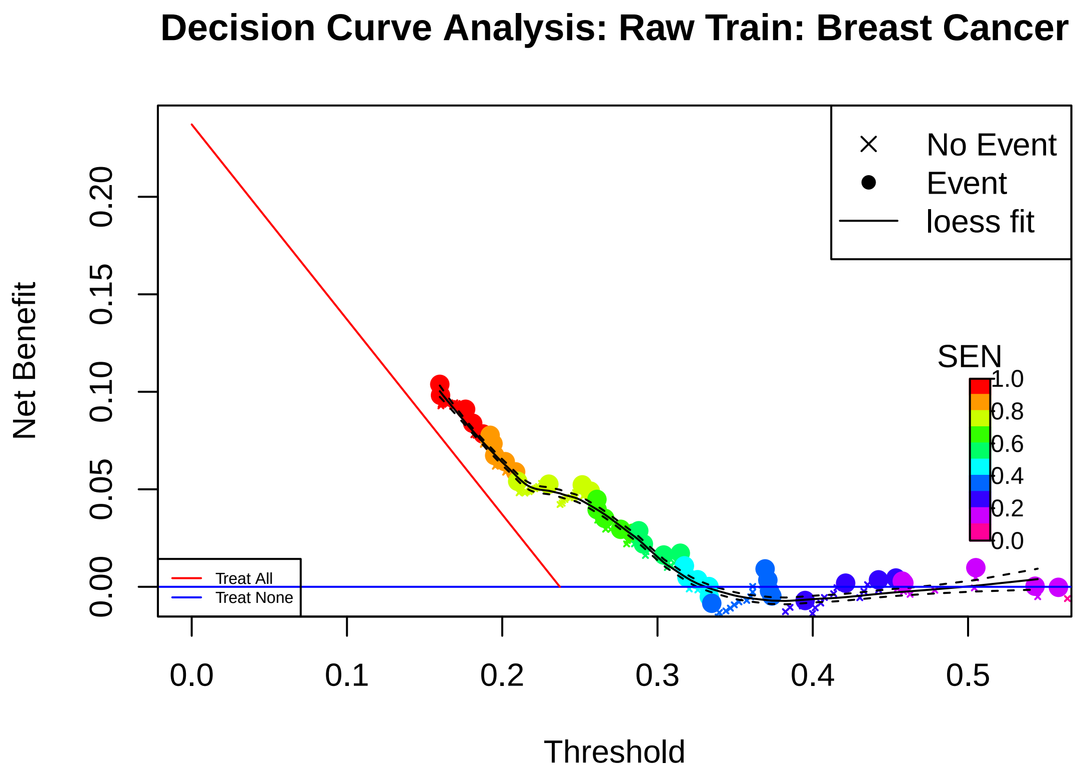
```
index <- predict(ml,dataBreast)
timeinterval <- 2*mean(subset(dataBreast,status==1)$time)

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

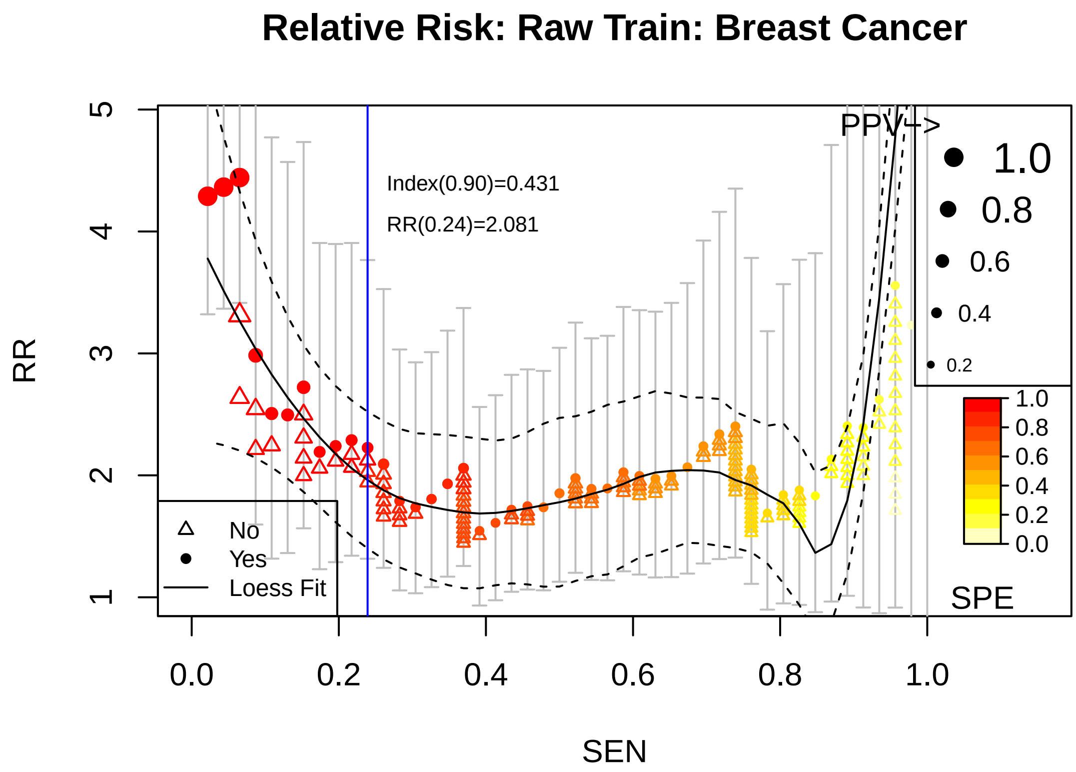
h0	timeinterval
0.323	51.1

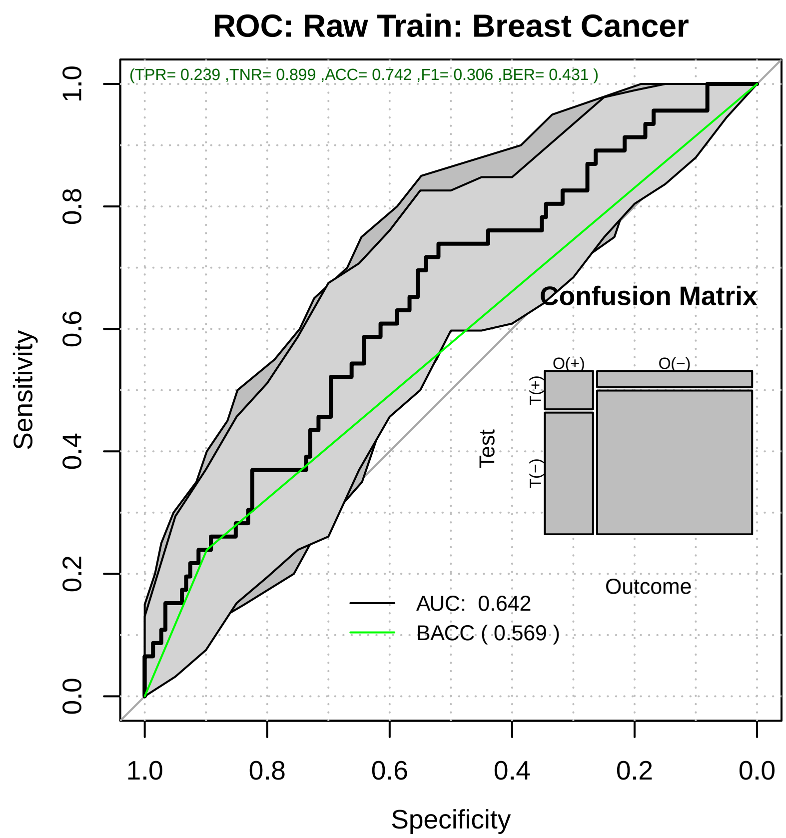
```
rdata <- cbind(dataBreast$status,ppoisGzero(index,h0))
rownames(rdata) <- rownames(dataBreast)

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

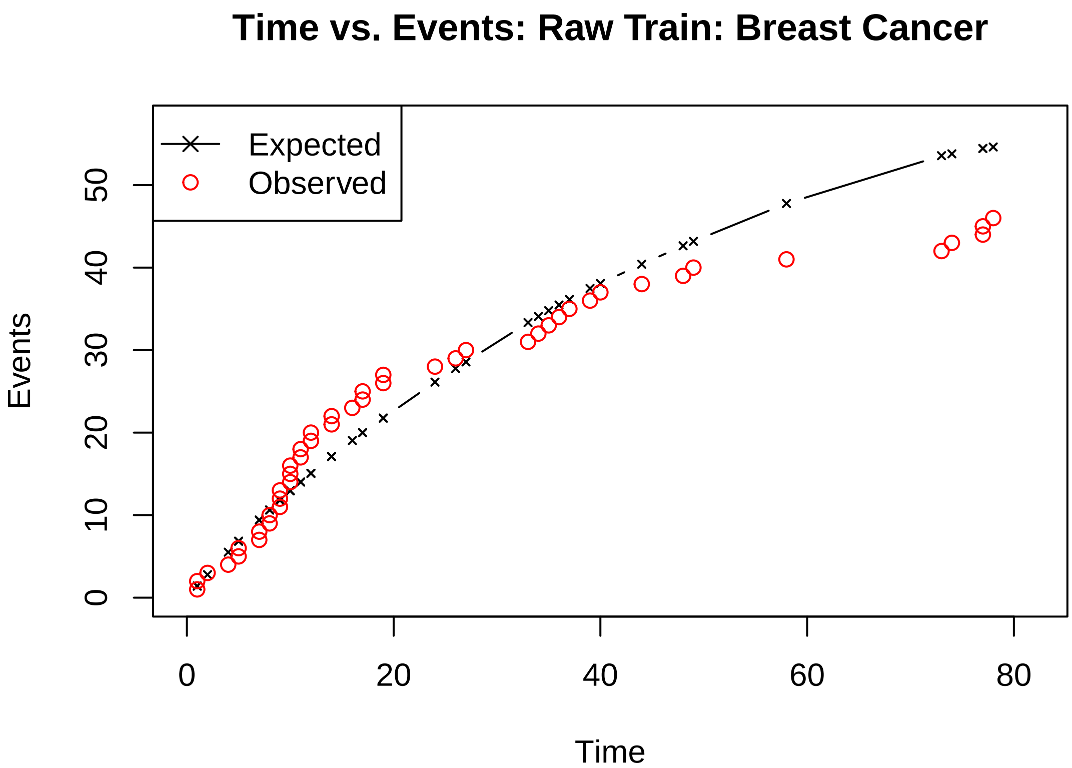

Cumulative vs. Observed: Raw Train: Breast Cancer
Observed
0
10
20
30
40
50
Event
Expected
0
10
20
30
40
50
Cumulative Probability

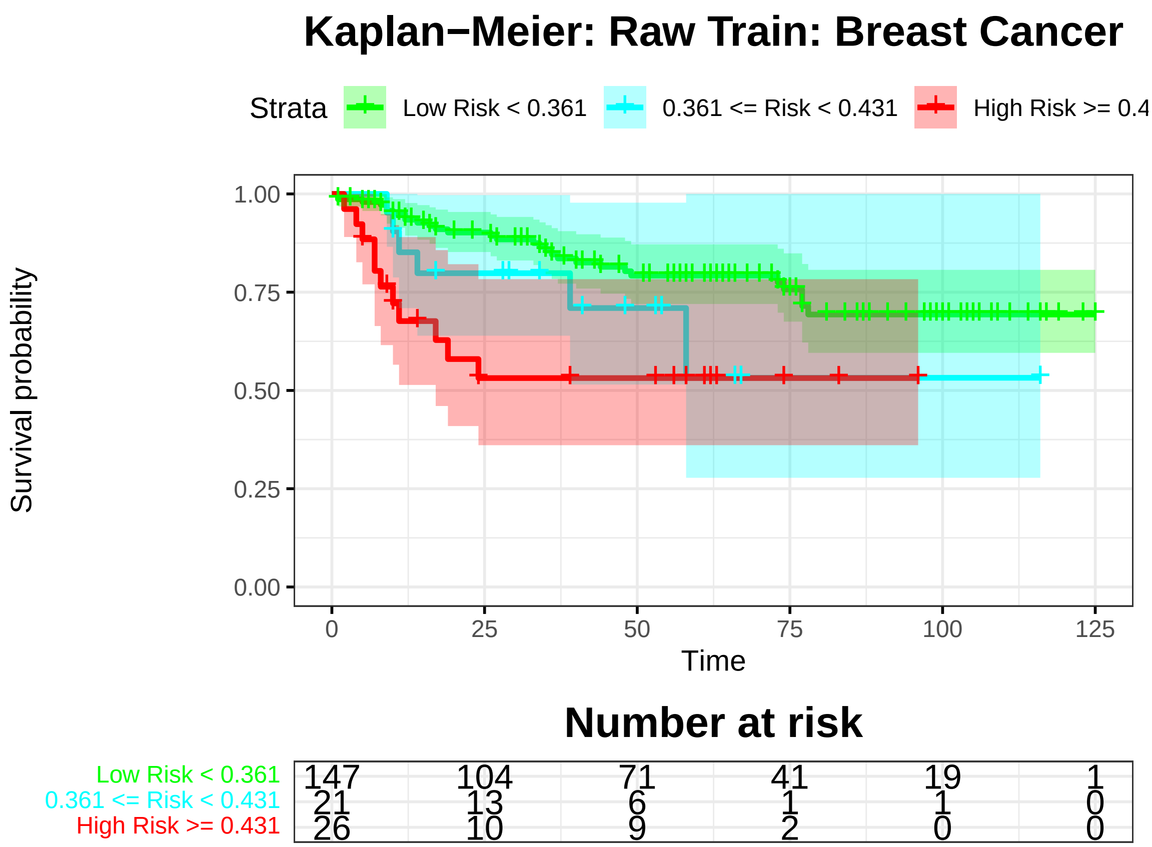


Decision Curve Analysis: Raw Train: Breast Cancer
No Event
Event
loess fit
SEN
1.0
0.8
0.6
0.4
0.2
0.0
Treat All
Treat None
Net Benefit
0.00
0.05
0.10
0.15
0.20
0.0
0.1
0.2
0.3
0.4
0.5
Threshold


Relative Risk: Raw Train: Breast Cancer
RR
SEN
Index(0.90)=0.431
RR(0.24)=2.081
PPV-->
SPE
No
Yes
Loess Fit
0.0
0.2
0.4
0.6
0.8
1.0
1
2
3
4
5


ROC: Raw Train: Breast Cancer
(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: 0.642
BACC (0.569)


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



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	0.430	0.361	0.252	0.176	0.1598	0.504
RR	1.944	1.833	2.402	3.557	30.3297	2.308
SEN	0.239	0.370	0.739	0.957	1.0000	0.152
SPE	0.892	0.797	0.520	0.169	0.0811	0.953
BACC	0.566	0.583	0.630	0.563	0.5405	0.552

```
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

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

Table 18: O/E Mean

mean	50%	2.5%	97.5%
1.02	1.02	0.969	1.07

```
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

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

Table 20: O/Acum Mean

mean	50%	2.5%	97.5%
0.793	0.793	0.787	0.799

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

Table 21: C. Index

mean.C Index	median	lower	upper
0.684	0.685	0.605	0.76

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

Table 22: ROC AUC

est	lower	upper
0.642	0.551	0.733

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

Table 23: Sensitivity

est	lower	upper
0.239	0.126	0.388

```
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%	at_max_BACC	at_max_RR	atSPE100	at_0.5
0.431	0.361	0.252	0.176	0.16	0.5

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

Table 26: Risk Ratio

est	lower	upper
2.08	1.22	3.55

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

Table 27: Logrank test Chisq = 11.441480 on 2 degrees of freedom, p = 0.003277

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	147	29	37.05	1.75	9.173
class=1	21	6	4.35	0.63	0.706
class=2	26	11	4.60	8.90	9.982