

Breast Cancer: Wisconsin

Jose Tamez

2023-04-28

Contents

1 Showcasing RRPlots	1
1.1 Exploring Raw Features with RRPlot	2
1.2 Reporting the Metrics	17
1.3 Modeling	19
1.4 Cox Model Performance	20

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)
```

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

1.0.2 Wisconsin Data Set

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

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

```
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
```

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

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

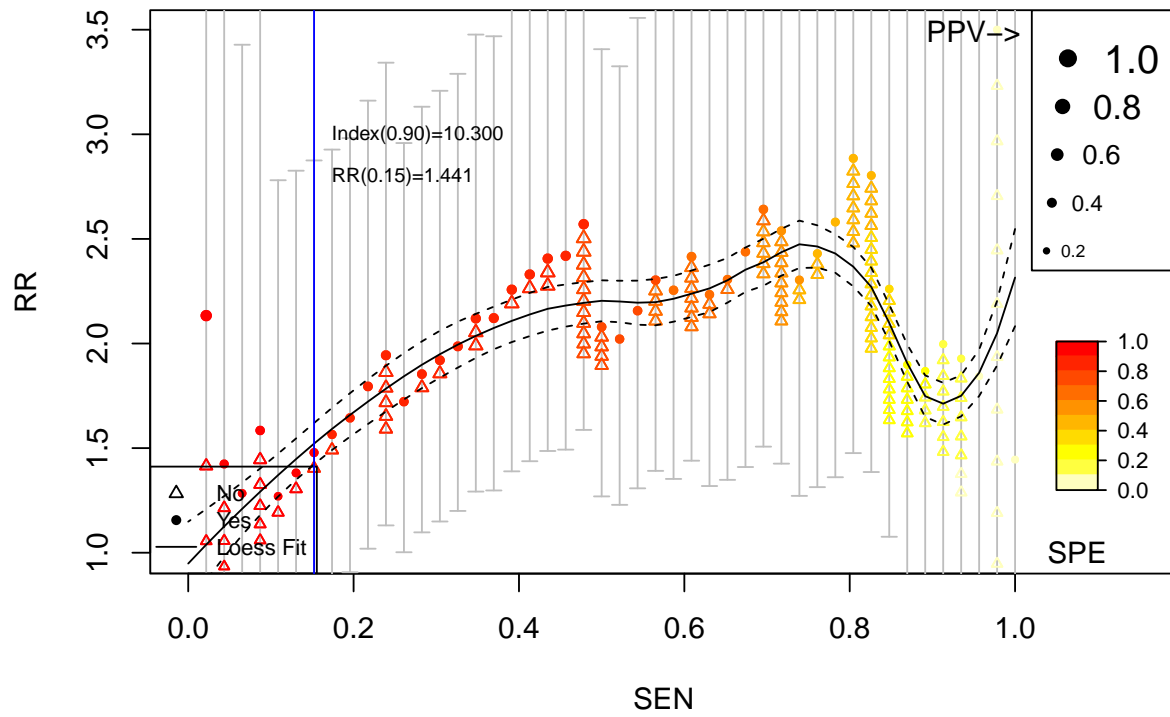
1.1 Exploring Raw Features with RRPlot

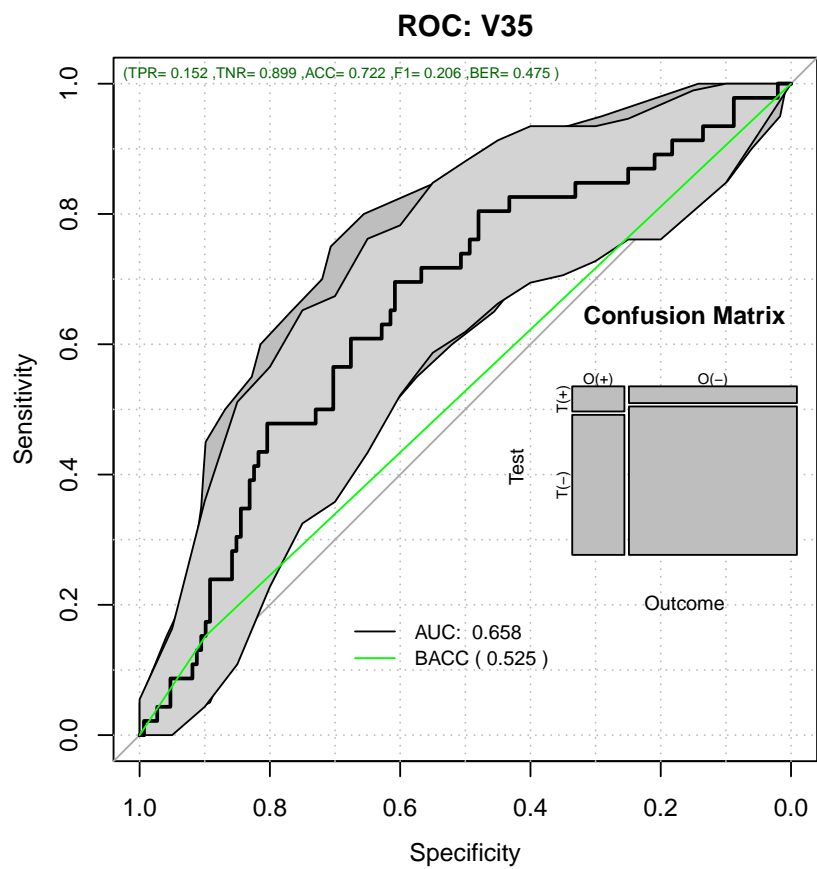
```
convar <- colnames(dataBreast)[lapply(apply(dataBreast,2,unique),length) > 10]
convar <- convar[convar != "time"]
topvar <- univariate_BinEnsemble(dataBreast[,c("status",convar)],"status")
pander::pander(topvar)
```

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

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

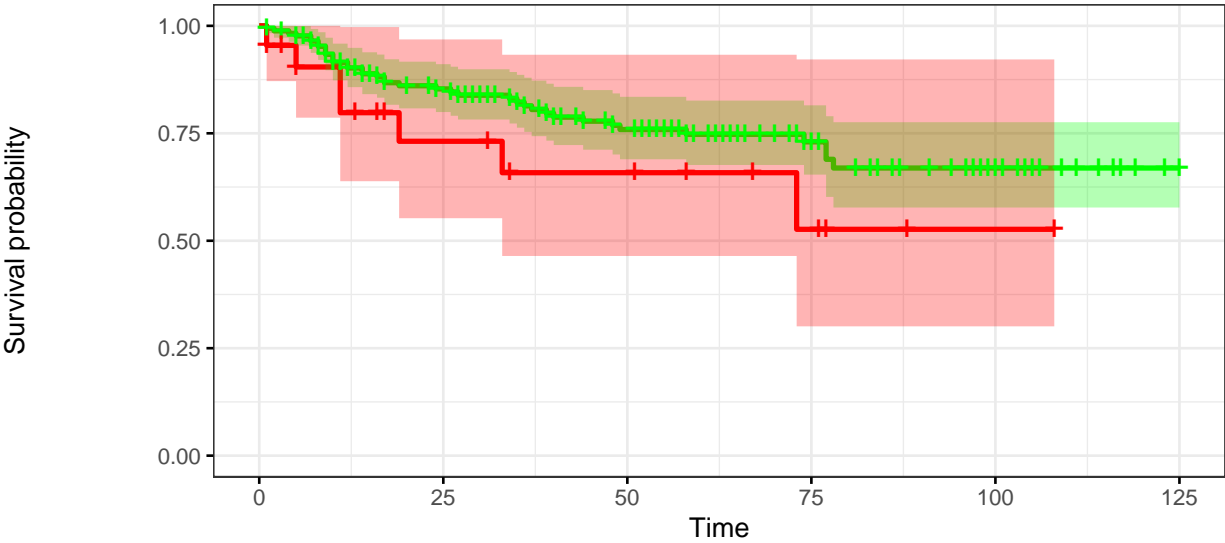
Relative Risk: V35





Kaplan–Meier: V35

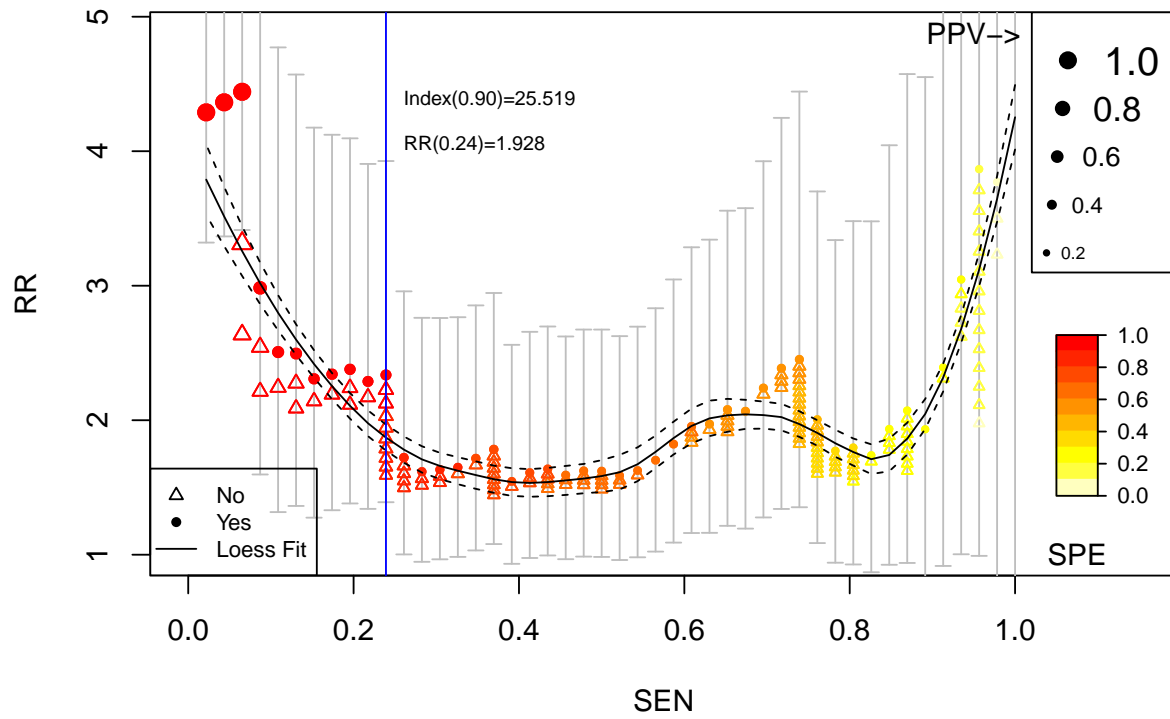
Strata + Low + At Risk > 10.300

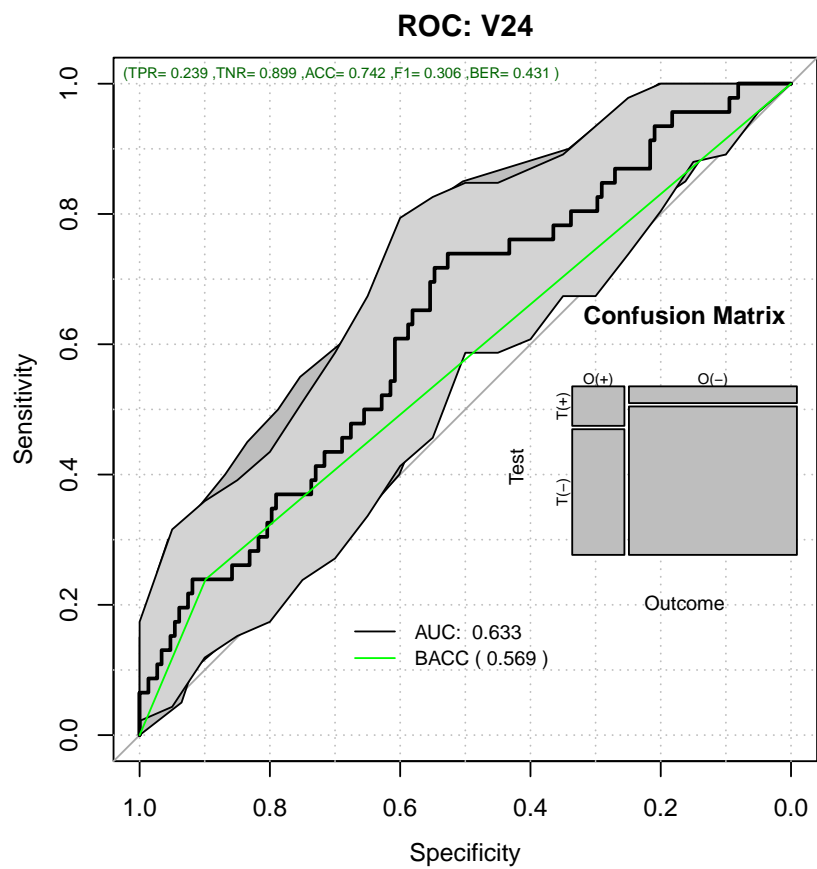


Number at risk

Low	172	116	78	40	19	1
At Risk > 10.300	22	11	8	4	1	0

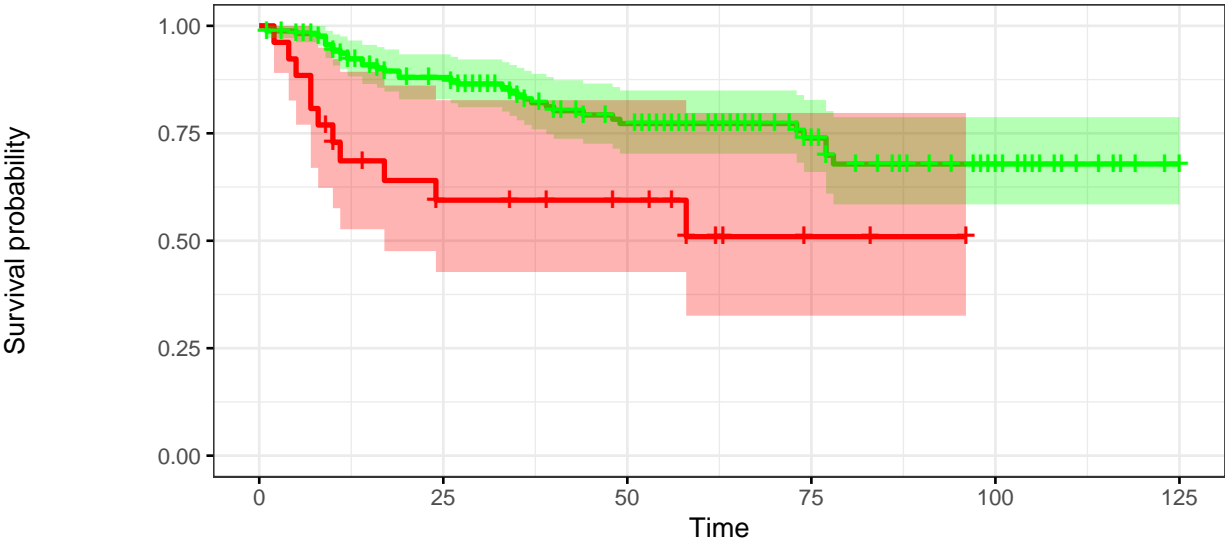
Relative Risk: V24





Kaplan–Meier: V24

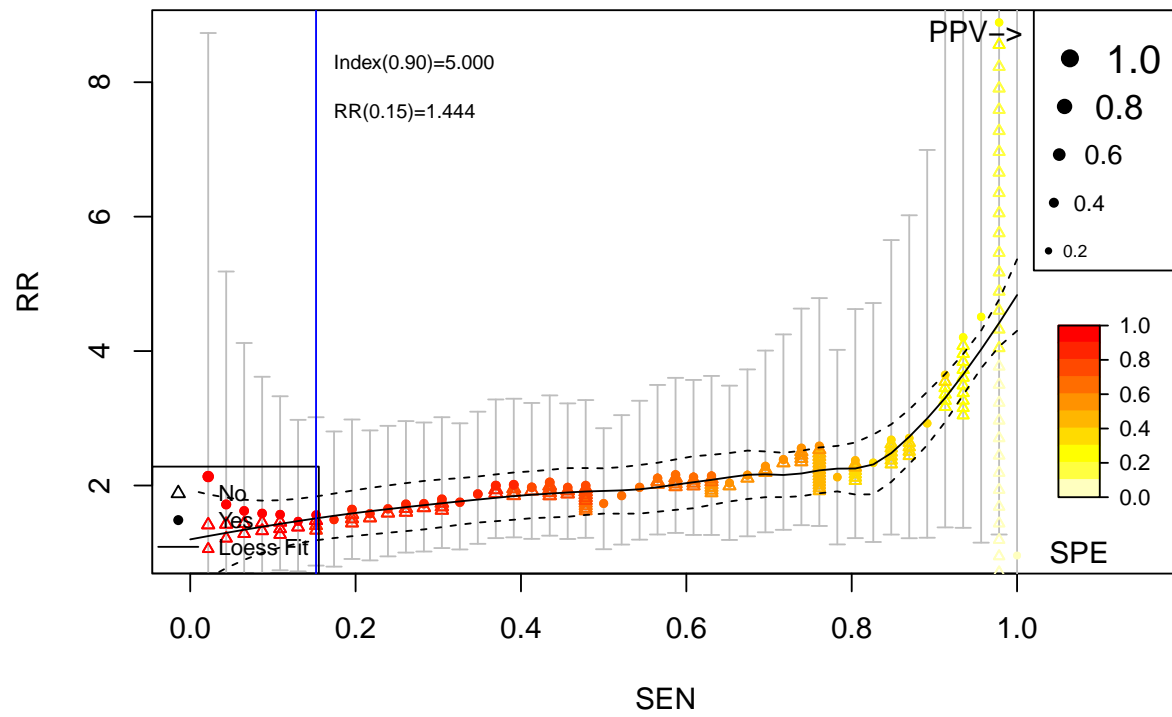
Strata + Low + At Risk > 25.519

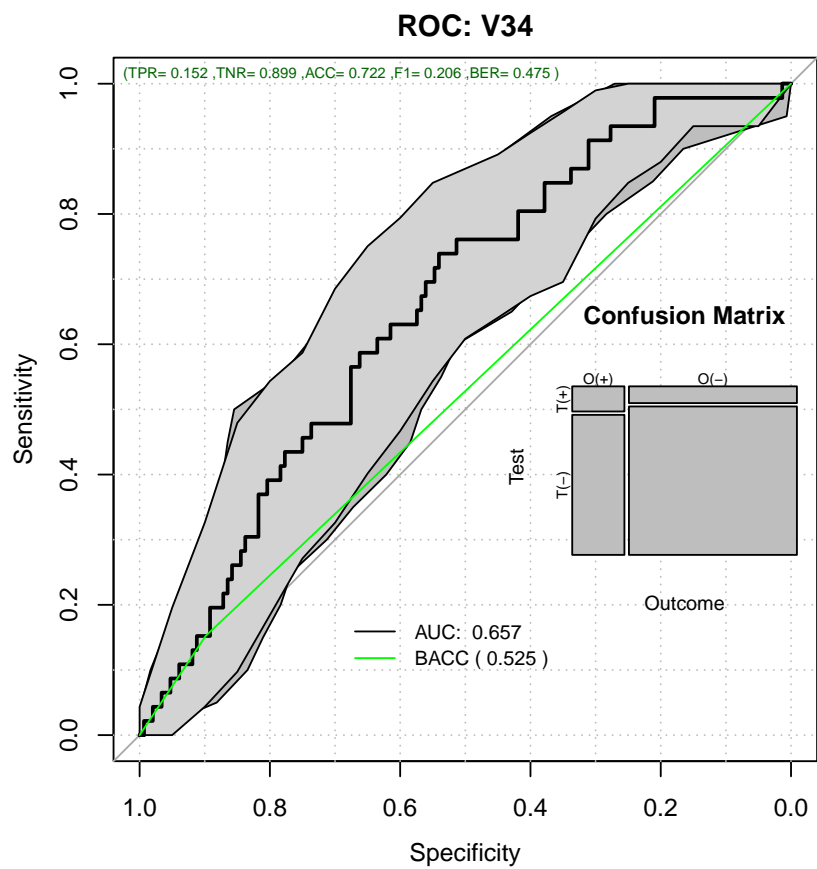


Number at risk

Low	168	115	77	42	20	1
At Risk > 25.519	26	12	9	2	0	0

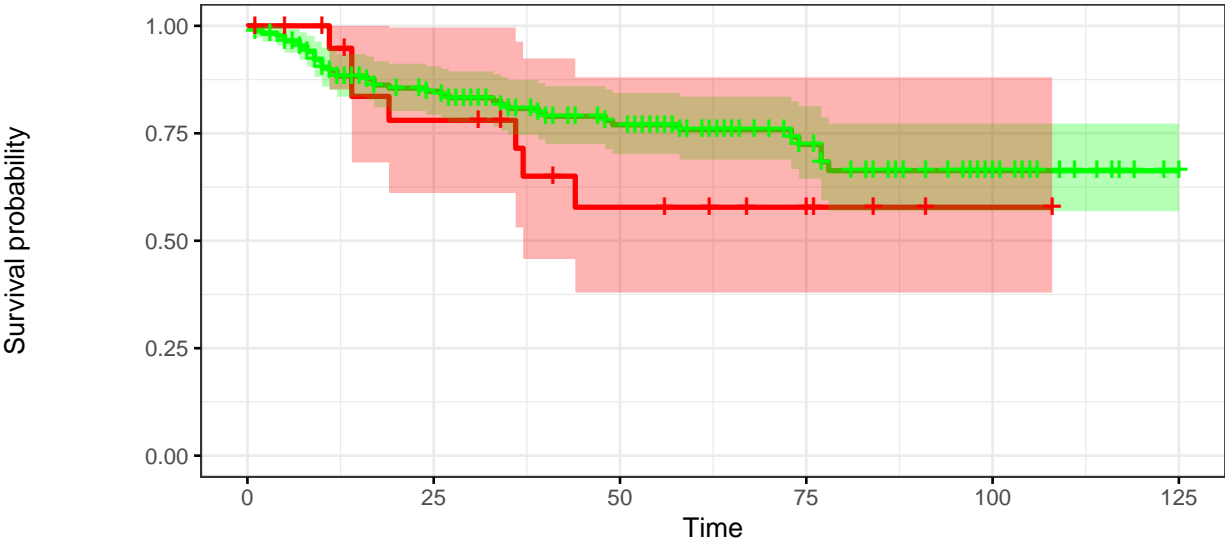
Relative Risk: V34





Kaplan–Meier: V34

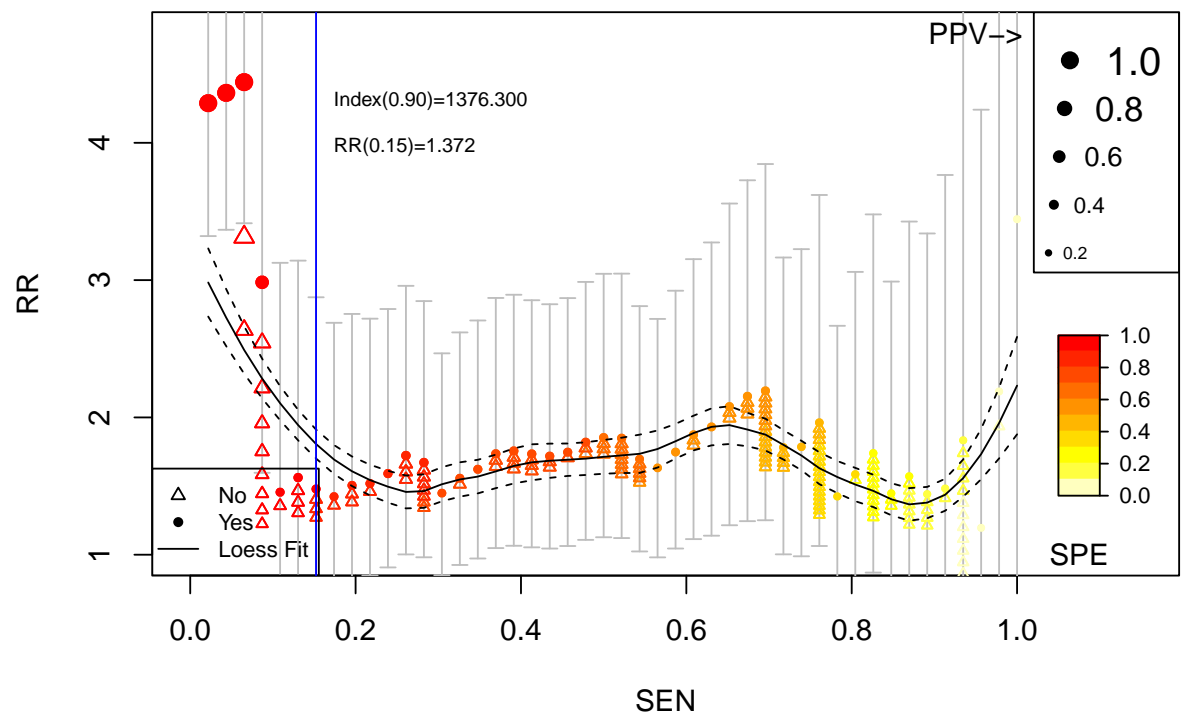
Strata + Low + At Risk > 5.000

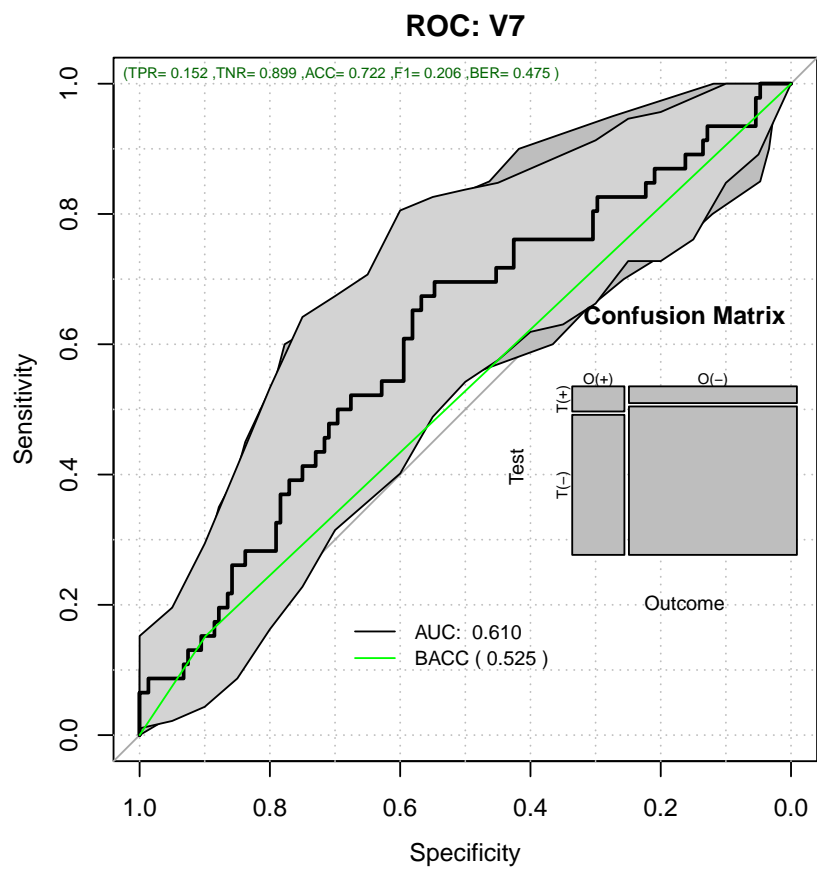


Number at risk

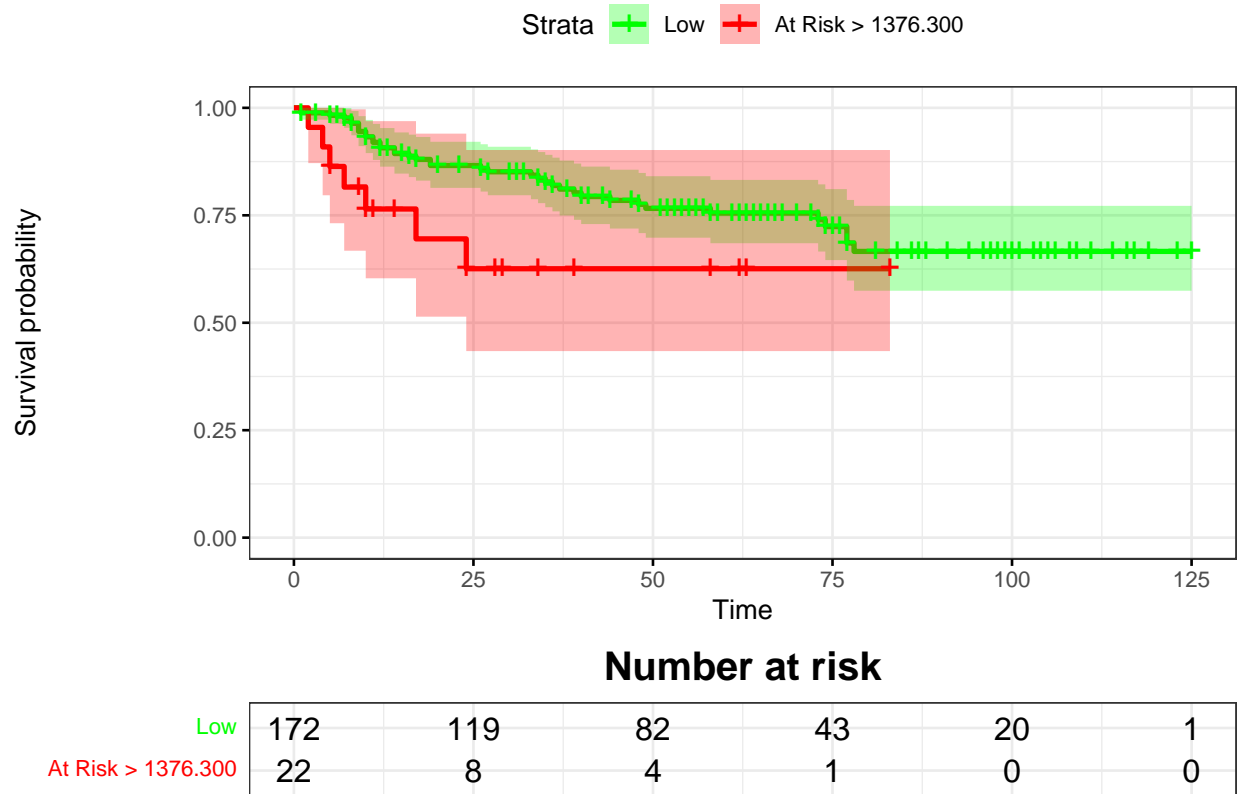
Low	172	113	78	39	19	1
At Risk > 5.000	22	14	8	5	1	0

Relative Risk: V7

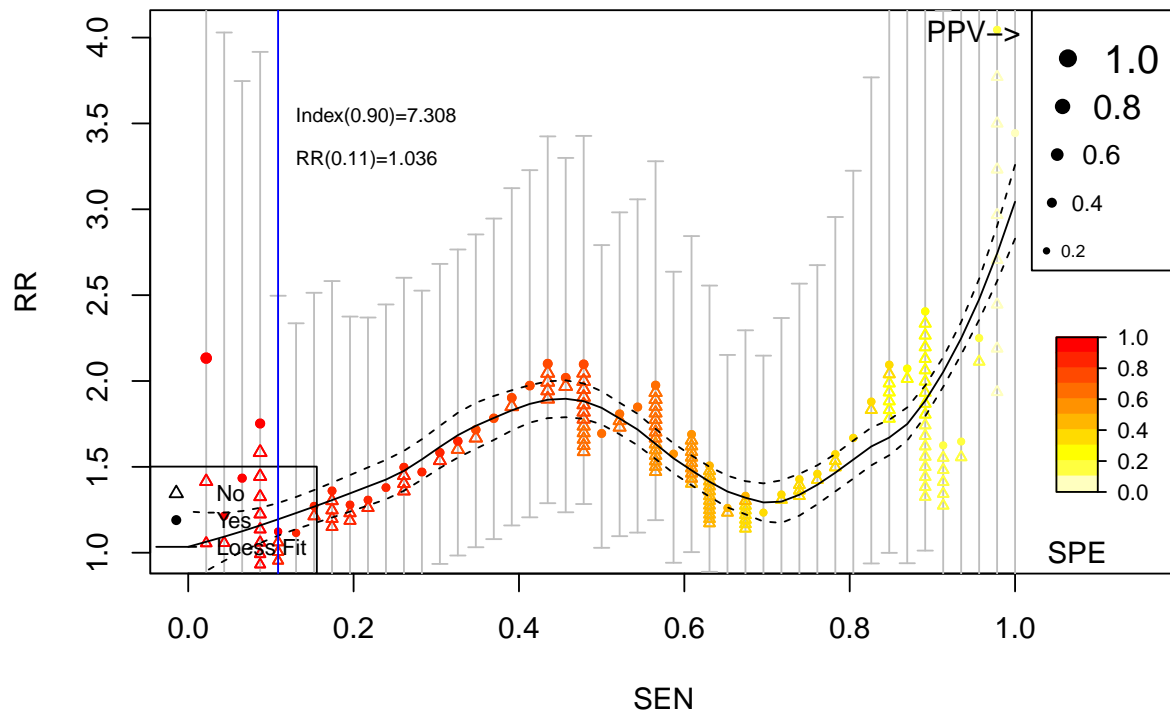


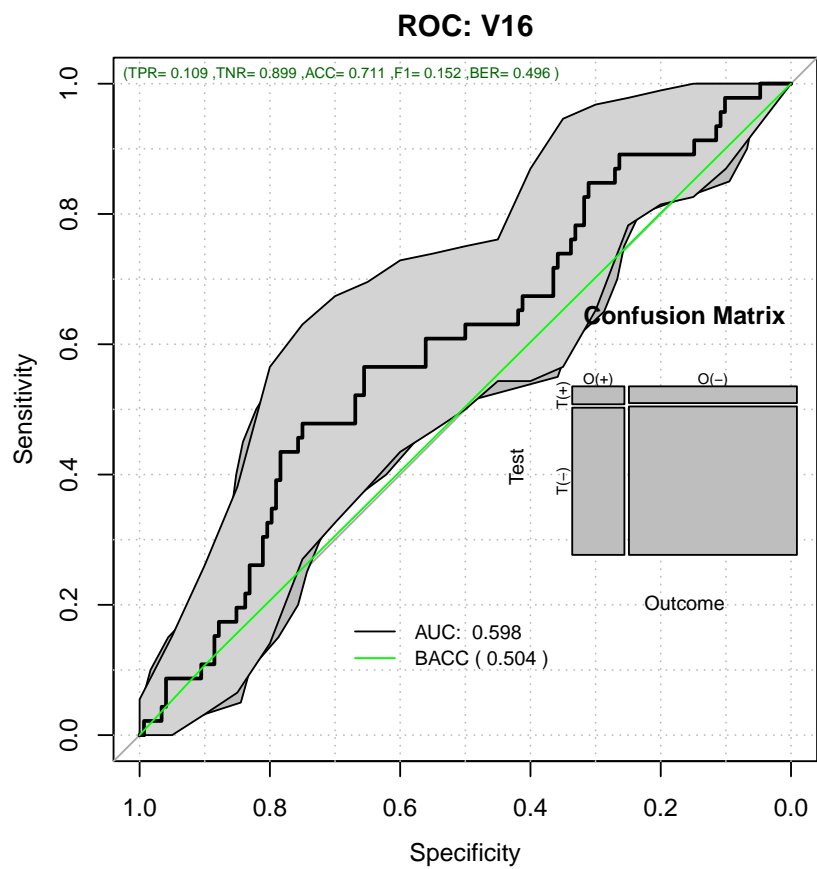


Kaplan–Meier: V7

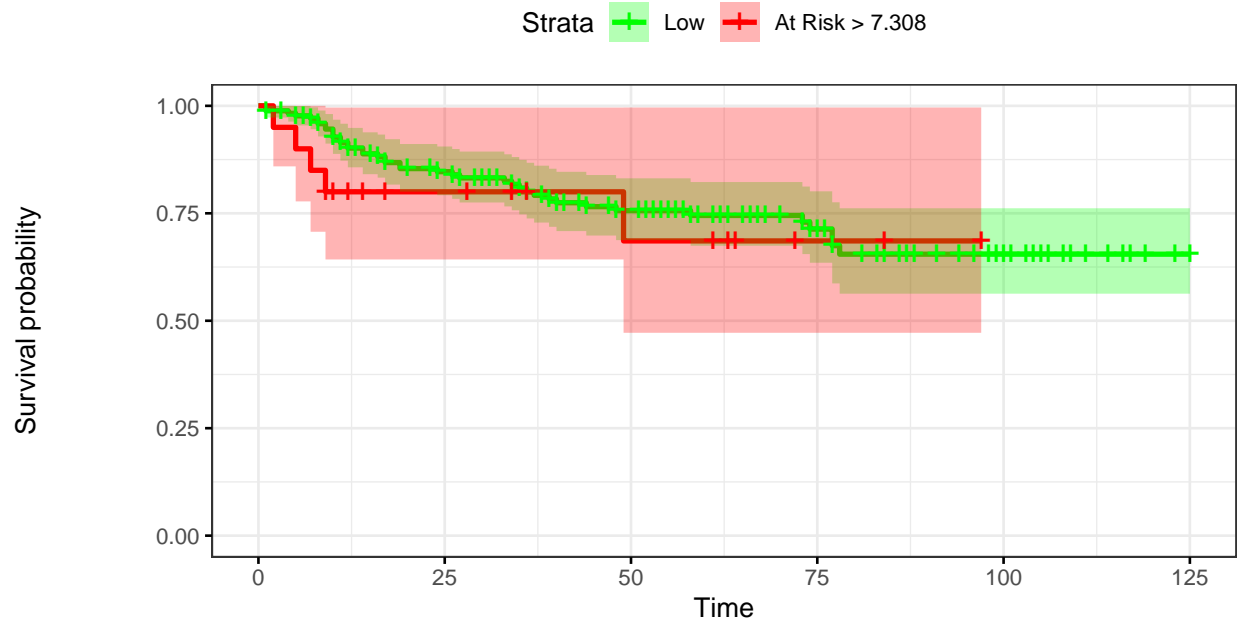


Relative Risk: V16





Kaplan–Meier: V16



Number at risk

Low	174	117	80	42	20	1
At Risk > 7.308	20	10	6	2	0	0

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

1.2 Reporting the Metrics

```
ROCAUC <- NULL
CstatCI <- NULL
RRratios <- NULL
LogRangp <- NULL
Sensitivity <- NULL
Specificity <- NULL

for (topf in topFive)
{
  CstatCI <- rbind(CstatCI, RRanalysis[[topf]]$c.index$cstatCI)
  RRratios <- rbind(RRratios, RRanalysis[[topf]]$RR_atP)
  LogRangp <- rbind(LogRangp, RRanalysis[[topf]]$surdif$pvalue)
  Sensitivity <- rbind(Sensitivity, RRanalysis[[topf]]$ROCAanalysis$sensitivity)
  Specificity <- rbind(Specificity, RRanalysis[[topf]]$ROCAanalysis$specificity)
  ROCAUC <- rbind(ROCAUC, RRanalysis[[topf]]$ROCAanalysis$aucs)
}

rownames(CstatCI) <- topFive
rownames(RRratios) <- topFive
rownames(LogRangp) <- topFive
rownames(Sensitivity) <- topFive
rownames(Specificity) <- topFive
```

```
rownames(ROCAUC) <- topFive
```

```
pander::pander(ROCAUC)
```

	est	lower	upper
V35	0.658	0.567	0.749
V24	0.633	0.542	0.724
V34	0.657	0.571	0.743
V7	0.610	0.515	0.705
V16	0.598	0.504	0.692

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

	mean.C Index	median	lower	upper
V35	0.632	0.633	0.543	0.724
V24	0.677	0.676	0.586	0.751
V34	0.658	0.659	0.590	0.728
V7	0.666	0.665	0.577	0.748
V16	0.614	0.614	0.530	0.701

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

	est	lower	upper
V35	1.44	0.739	2.81
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	0.16213
V24	0.00331
V34	0.34023
V7	0.03818
V16	0.50639

```
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

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

	ROCAUC	C-Stat	Sen	Spe	RR
V35	0.658	0.632	0.152	0.899	1.44
V24	0.633	0.677	0.239	0.899	1.93
V34	0.657	0.658	0.152	0.899	1.44
V7	0.610	0.666	0.152	0.899	1.37
V16	0.598	0.614	0.109	0.899	1.04

1.3 Modeling

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

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

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

Table 9: Table continues below

	Estimate	lower	HR	upper	u.Accuracy	r.Accuracy
V24	5.07e-02	1.02	1.05	1.09	0.598	0.241
V27	2.61e-04	1.00	1.00	1.00	0.608	0.241
V26	4.09e-03	1.00	1.00	1.01	0.593	0.319
V34	1.11e-02	1.00	1.01	1.02	0.634	0.331
V7	4.31e-08	1.00	1.00	1.00	0.588	0.237
V35	1.57e-03	1.00	1.00	1.00	0.727	0.592
V6	1.02e-07	1.00	1.00	1.00	0.577	0.237

Table 10: 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
V27	0.608	0.608	0.501	0.608	0.0563	0.433	2.76
V26	0.597	0.598	0.524	0.601	0.0621	0.395	2.75
V34	0.627	0.618	0.527	0.616	0.0300	0.457	2.36
V7	0.588	0.595	0.500	0.595	0.0487	0.380	2.30
V35	0.619	0.641	0.598	0.615	0.0275	0.551	2.24

	full.Accuracy	u.AUC	r.AUC	full.AUC	IDI	NRI	z.IDI
V6	0.577	0.588	0.500	0.588	0.0459	0.353	2.19

	z.NRI	Delta.AUC	Frequency
V24	2.66	0.1079	1.0
V27	2.63	0.1071	1.0
V26	2.39	0.0776	1.0
V34	2.76	0.0893	1.0
V7	2.30	0.0949	0.6
V35	3.41	0.0172	0.8
V6	2.13	0.0881	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")
```

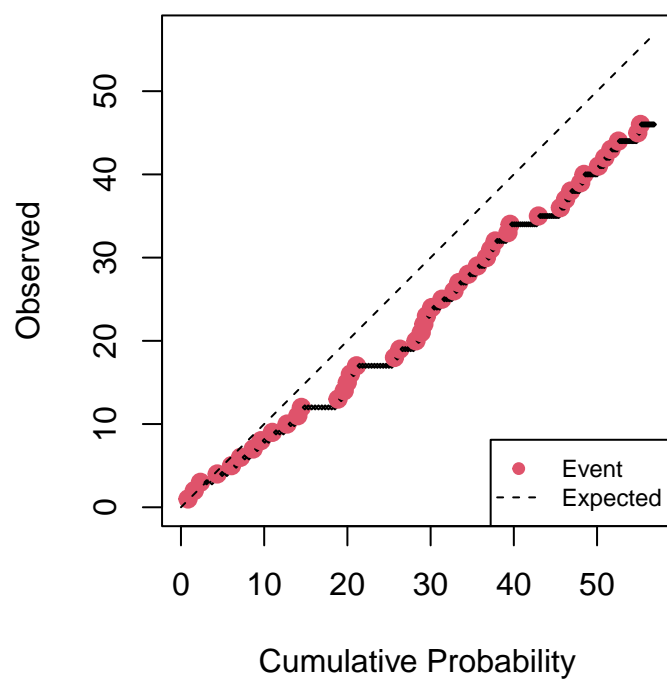
Table 12: Initial Parameters

h0	timeinterval
0.323	51.1

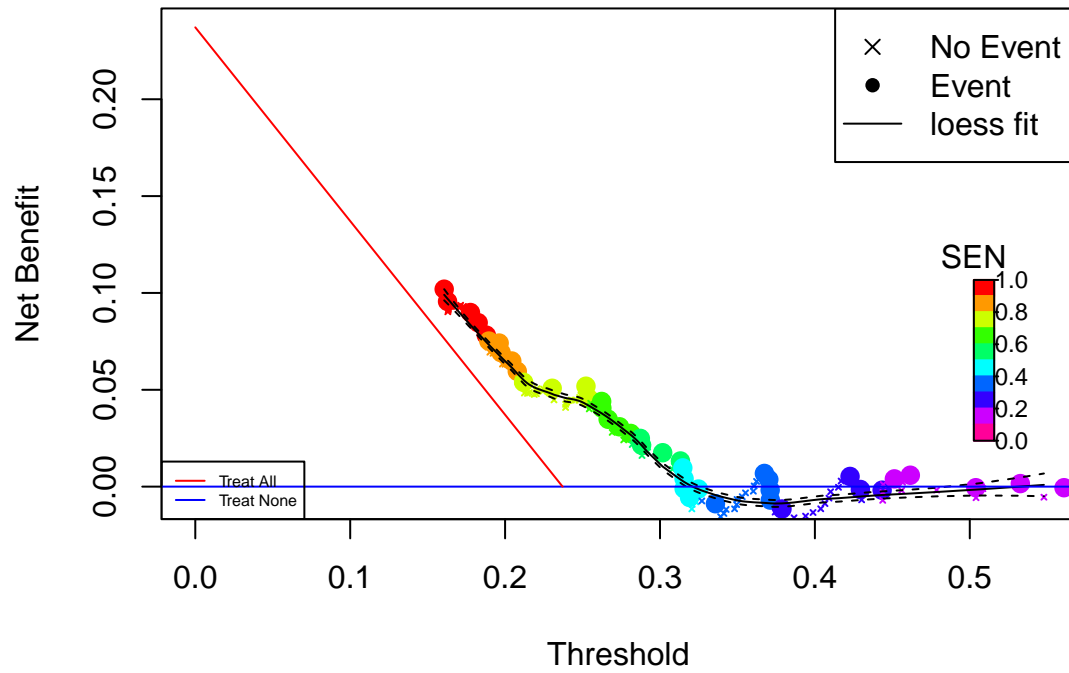
```
rdata <- cbind(dataBreast$status,ppoisGzero(index,h0))
rownames(rdata) <- rownames(dataBreast)

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)
```

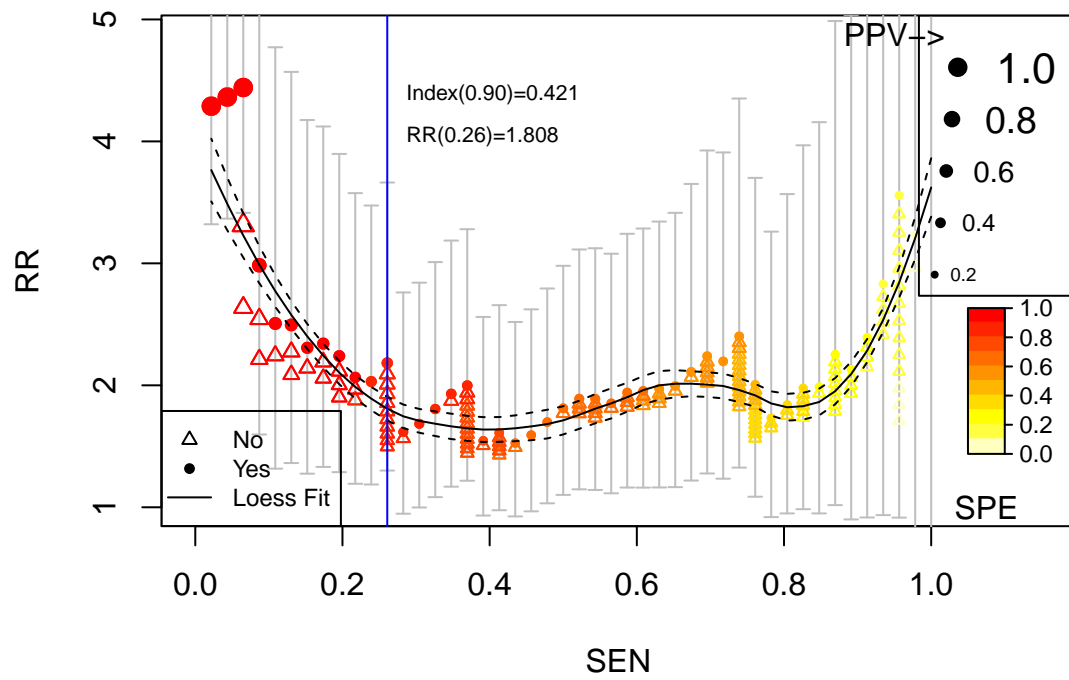
Cumulative vs. Observed: Raw Train: Breast Cancer

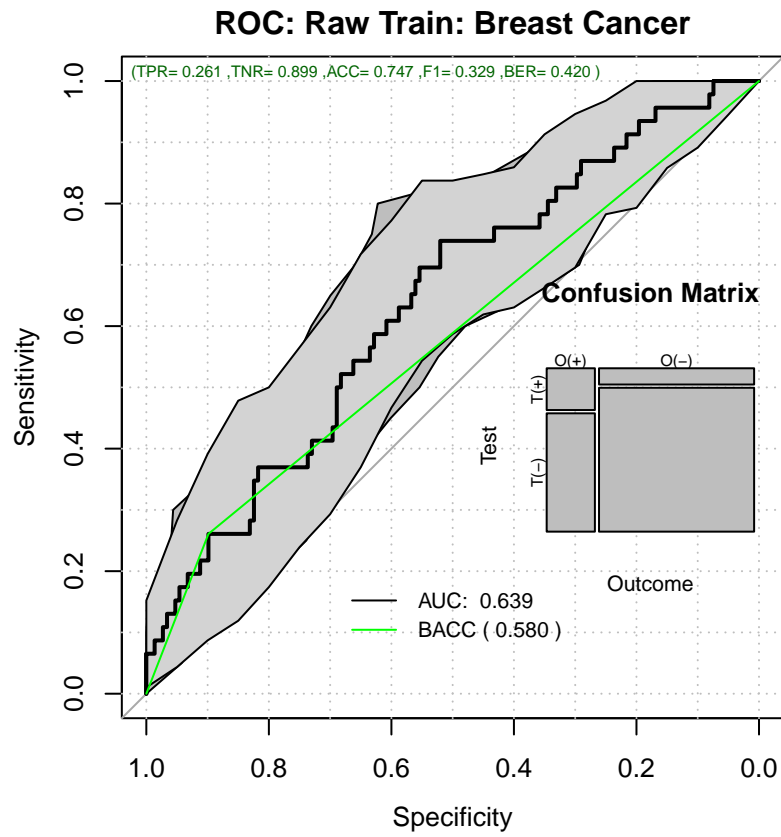


Decision Curve Analysis: Raw Train: Breast Cancer

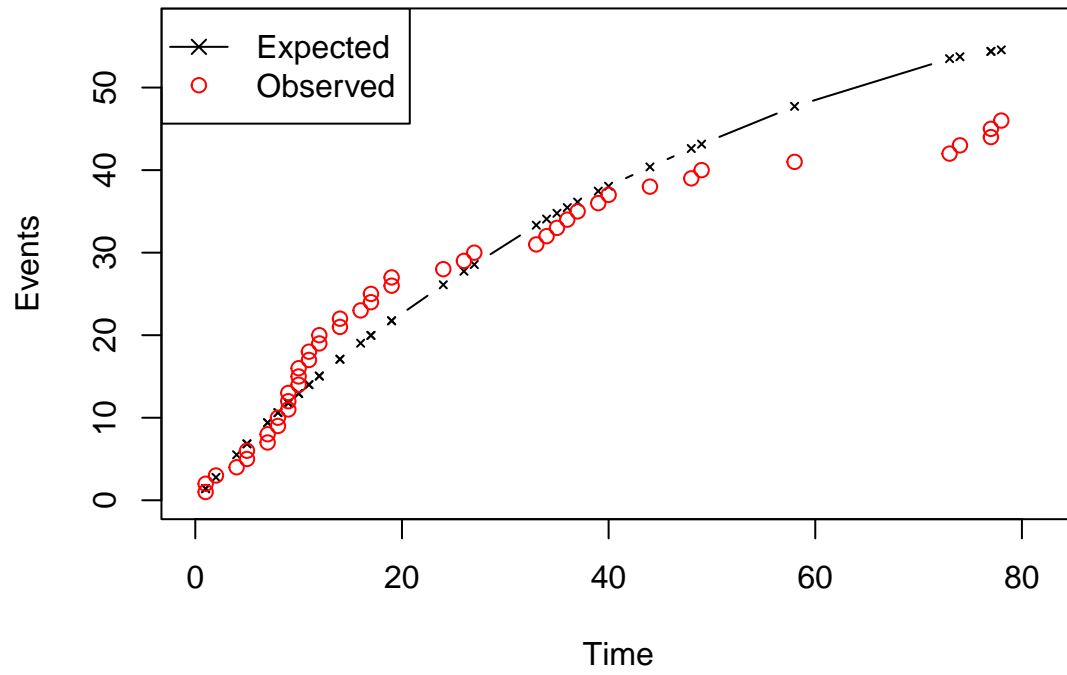


Relative Risk: Raw Train: Breast Cancer

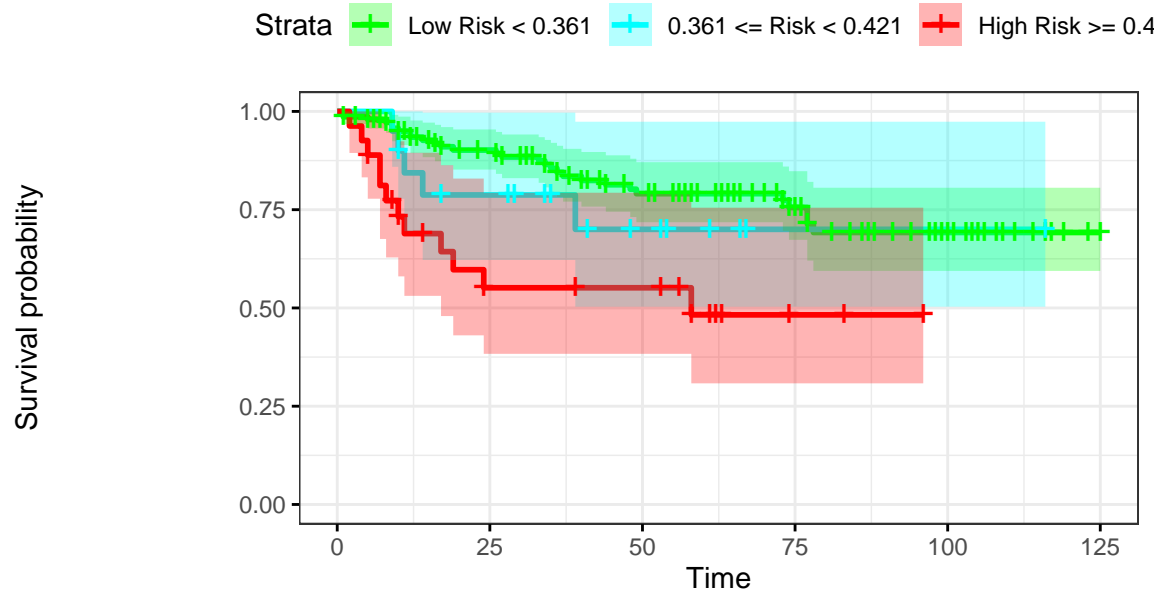




Time vs. Events: Raw Train: Breast Cancer



Kaplan–Meier: Raw Train: Breast Cancer



Number at risk

Low Risk < 0.361	147	103	70	41	19	1
0.361 <= Risk < 0.421	20	13	6	1	1	0
High Risk >= 0.421	27	11	10	2	0	0

1.4.2 Uncalibrated Performance Report

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

Table 13: O/E Ratio

est	lower	upper
0.843	0.617	1.12

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

Table 14: O/E Ratio

mean	50%	2.5%	97.5%
1.01	1.01	0.964	1.06

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

Table 15: O/Acum Ratio

mean	50%	2.5%	97.5%
0.793	0.792	0.785	0.801

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

mean.C Index	median	lower	upper
0.68	0.679	0.596	0.758

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

Table 17: ROC AUC

est	lower	upper
0.639	0.547	0.73

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

Table 18: Sensitivity

est	lower	upper
0.261	0.143	0.411

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

Table 19: Specificity

est	lower	upper
0.899	0.838	0.942

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

Table 20: Probability Thresholds

90%	80%
0.421	0.361

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

Table 21: Risk Ratio

est	lower	upper
1.81	1.06	3.09

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

Table 22: Logrank test Chisq = 12.263179 on 2 degrees of freedom,
p = 0.002173

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	147	29	36.9	1.688	8.685
class=1	20	5	4.2	0.151	0.168
class=2	27	12	4.9	10.269	11.609