

# geneSurv: an interactive tool for survival analysis in genomics research

*Selcuk Korkmaz*

*February 3, 2017*

## Contents

Abstract . . . . .	1
1.Data upload . . . . .	1
2. Analysis Methods . . . . .	2

## Abstract

Survival analysis is often used in cancer studies. It has been shown that combination of clinical data with genomics increases predictive performance of survival analysis methods. This tool provides a wide range of survival analysis methods for genomics research, especially in cancer studies. The tool includes analysis methods including Kaplan-Meier, Cox regression, Penalized Cox regression and Random Survival Forests. It also offers methods for optimal cutoff point determination for continuous markers.

Each procedure includes following features:

Kaplan-Meier: descriptive statistics, survival table, mean and median life time, hazard ratios, comparison tests including Log-rank, Gehan-Breslow, Tarone-Ware, Peto-Peto, Modified Peto-Peto, Flemington-Harrington, and interactive plots such as Kaplan-Meier curves and hazard plots.

Cox regression: coefficient estimates, hazard ratios, goodness of fit test, analysis of deviance, save predictions, save residuals, save Martingale residuals, save Schoenfeld residuals, save dfBetas, proportional hazard assumption test, and interactive plots including Schoenfeld residual plot and Log-Minus-Log plot.

Penalized Cox regression: feature selection using ridge, elastic net and lasso penalization. A cross-validation to investigate the relationship between partial likelihood devaince and lambda values.

Random survival forests: overall survival predictions (Nelson-Aalen estimator, overall ensemble), individual survival predictions (with OOB), individual cumulative hazard predictions (with OOB), error rate, variable importance, and interactive plots including random survival plot, cumulative hazard plot, error rate plot, Cox vs RSF plot

Optimal cutoff: determination of optimal cutoff value by maxmizing test statistics, including log-rank, Gehan-Breslow, Tarone-Ware, Peto-Peto, modified Peto-Peto, Flemington-Harrington.

## 1.Data upload

This tool requires a dataset in `*.txt` format, which is seperated by `comma`, `semicolon`, `space` or `tab` delimiter. First row of dataset must include header. When the appropriate file is uploaded, the dataset will be appear immediately on the main page of the tool. Alternatively users can upload one of the example datasets provided within the tool for testing and understanding the operating logic of the tool.

## 2. Analysis Methods

### 2.1. Kaplan-Meier

#### Concept

Kaplan-Meier is a non-parametric statistical method that is used to estimate survival probabilities and hazard ratios for a cohort study group. In clinical trials, it is often used to measure the part of patients living for a certain period of time after a treatment.

#### Variables

- **Survival time:** Time until an event occurs (i.e. days, weeks, months, years)
- **Status variable:** The event (i.e. death, disease, remission, recovery)
- **Category value for status variable:** Category value of the event of interest (i.e. 1, yes)
- **Factor variable:** A categorical variable which indicates different study groups (i.e. treatment, gender)

#### Usage

A Kaplan-Meier analysis can be conducted by applying the following steps:

1. Select the analysis method as **Kaplan Meier** from **Analysis** tab.
2. Select suitable variables for the analysis, such as **survival time**, **status variable**, **category value for status variable** and **factor variable**, if exists.
3. In advanced options, one can change confidence interval type, as log, log-log or plain, variance estimation method, as Greenwood or Tsiatis, Fleming-Harrington weights, confidence level and reference category, as first or last.
4. Click **Run** button to run the analysis.

#### Outputs

*Desired outputs can be selected by clicking Outputs checkbox. Available outputs are;*

##### a. Case summary

Summary statistics, such as number and percent of observations, events and censored cases can be obtained.

<div>Copy</div>	<div>CSV</div>	<div>PDF</div>	<div>Print</div>	Search: <div></div>	
	<div>n</div>	<div>n of event</div>	<div>% of event</div>	<div>n of censor</div>	<div>% of censor</div>
0	77	27	35.065	50	64.935
1	209	80	38.278	129	61.722

Showing 1 to 2 of 2 entries

Previous

1

Next

•

##### b. Survival table

A survival table can be created. First column in the table represents factor group and number of time points (i.e. 1.2 means second time point in the first factor group, likewise 2.1 means first time point in the second group). Second column is survival time, third column gives number of subjects at risk, fourth column is the number of events, fifth column represents the cumulative probability of surviving, sixth, seventh and eighth columns are associated standard error, lower and upper limits, respectively.

Copy

CSV

PDF

Print

Search:

	Time	Number at risk	Number of event	Cumulative probability of surviving	S.E.	Lower limit	Upper limit
0.1	6	77	2	0.974	0.018	0.939	1
0.2	7	75	1	0.961	0.022	0.919	1
0.3	9	74	2	0.935	0.028	0.882	0.992
0.4	11	72	2	0.909	0.033	0.847	0.976
0.5	12	70	1	0.896	0.035	0.83	0.967
0.6	13	69	1	0.883	0.037	0.814	0.958
0.7	14	68	3	0.844	0.041	0.767	0.929
0.8	16	65	1	0.831	0.043	0.752	0.919
0.9	17	64	1	0.818	0.044	0.736	0.909
0.10	18	63	1	0.805	0.045	0.721	0.899

Showing 1 to 10 of 72 entries

Previous

1

2

3

4

5

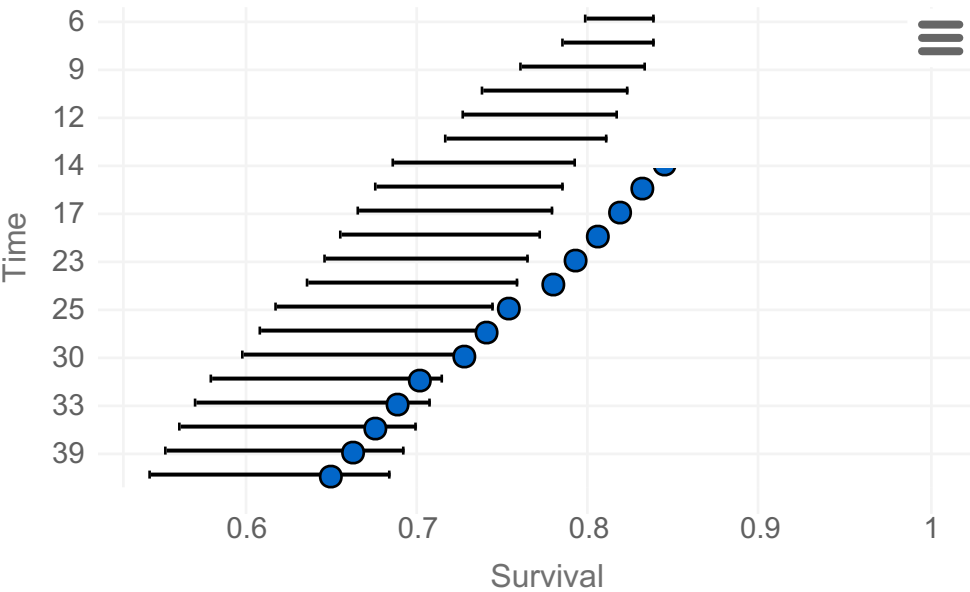
...

8

Next

c. Survival plot

A forest plot can be created for each level of factor group using survival probabilites at each end point.



.

d. Mean and Median life time

Mean and median life time and their associated confidence levels can be calculated for each level of factor group.

Copy		CSV	PDF	Print	Search: <input type="text"/>			
	n	Mean	S.E.(mean)	Lower (mean)	Upper (mean)	Median	Lower (median)	Upper (median)
0	77	111.74	7.676	96.695	126.785			
1	209	118.739	4.657	109.612	127.865			

Showing 1 to 2 of 2 entries

Previous 1 Next

#### e. Hazard ratio

Hazard ratios and their respective lower and upper limits can be calculated for each factor group at each end point.

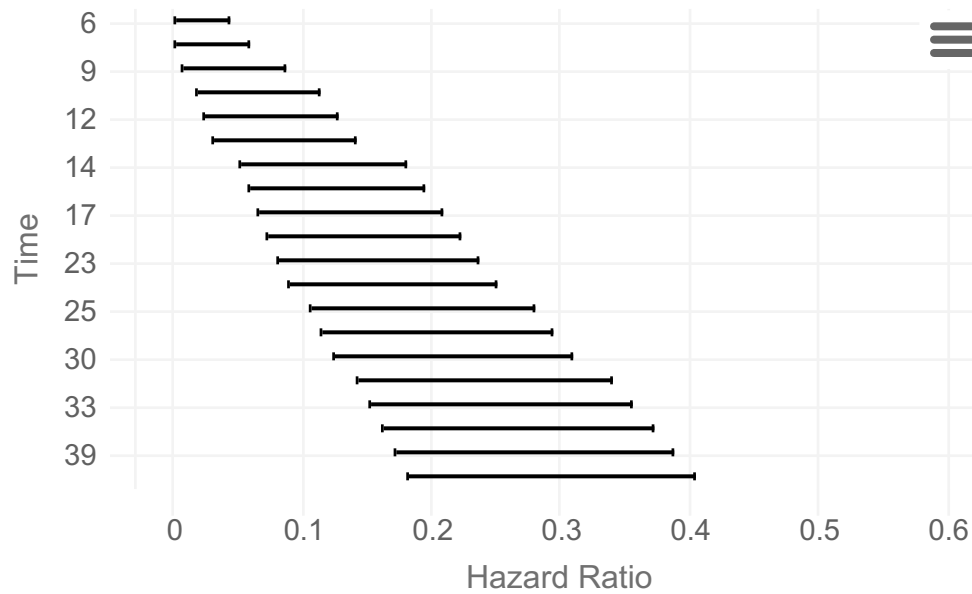
Copy		CSV	PDF	Print	Search: <input type="text"/>			
	Time	Hazard	Lower	Upper				
0.1	6	0.026	0	0.063				
0.2	7	0.04	0	0.084				
0.3	9	0.067	0.008	0.126				
0.4	11	0.095	0.024	0.166				
0.5	12	0.11	0.034	0.186				
0.6	13	0.124	0.043	0.206				
0.7	14	0.17	0.074	0.265				
0.8	16	0.185	0.084	0.285				
0.9	17	0.201	0.095	0.307				
0.10	18	0.217	0.106	0.327				

Showing 1 to 10 of 72 entries

Previous 1 2 3 4 5 ... 8 Next

#### f. Hazard plot

A forest plot can be created for each level of factor group using hazard ratios at each end point.



### g. Comparison tests

Six different comparison tests can be calculated for testing the differences in survival probability estimations between factor groups.

Copy CSV PDF Print

Search:

	Test	Chi square	df	p value
1	Log-rank	0	1	0.991
2	Gehan-Breslow	0.268	1	0.605
3	Tarone-Ware	0.074	1	0.785
4	Peto-Peto	0.207	1	0.649
5	Modified Peto-Peto	0.209	1	0.647
6	Flemington-Harnington $p = 1$ $q = 1$	2.197	1	0.138

Showing 1 to 6 of 6 entries

Previous

1

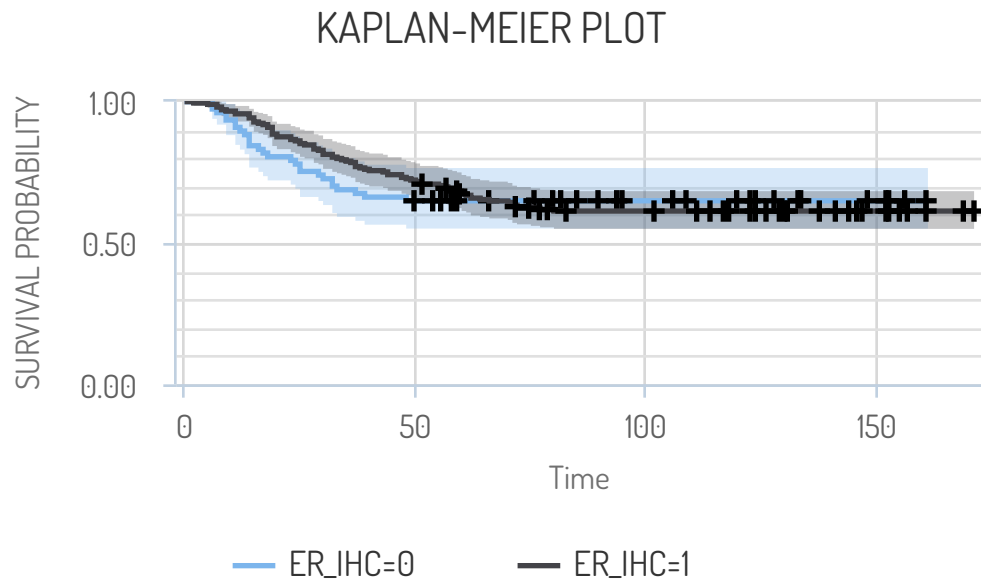
Next

•

### h. Plots

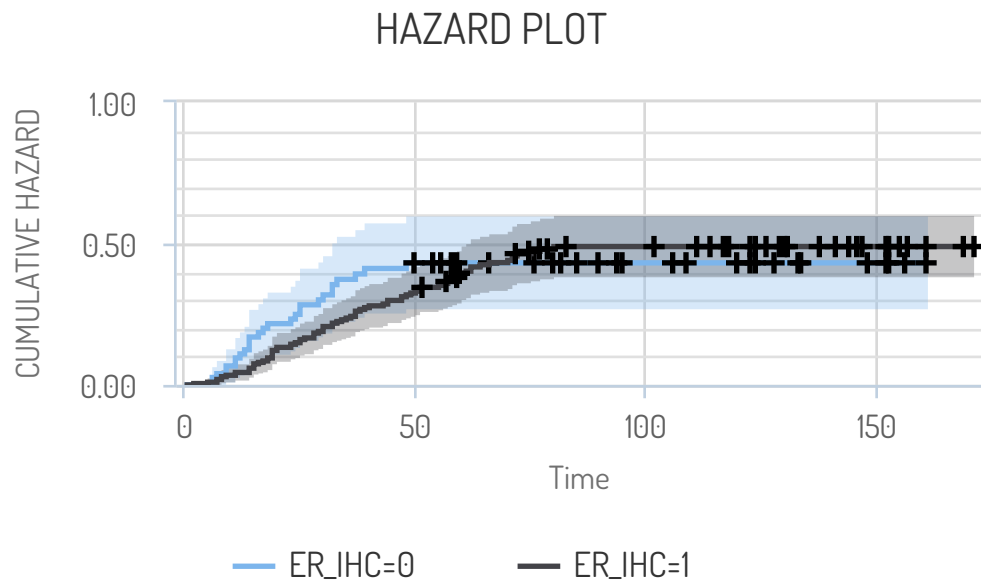
#### i. Kaplan-Meier curve

Kaplan-Meier curves can be created. A number of edit options is also available for plots.



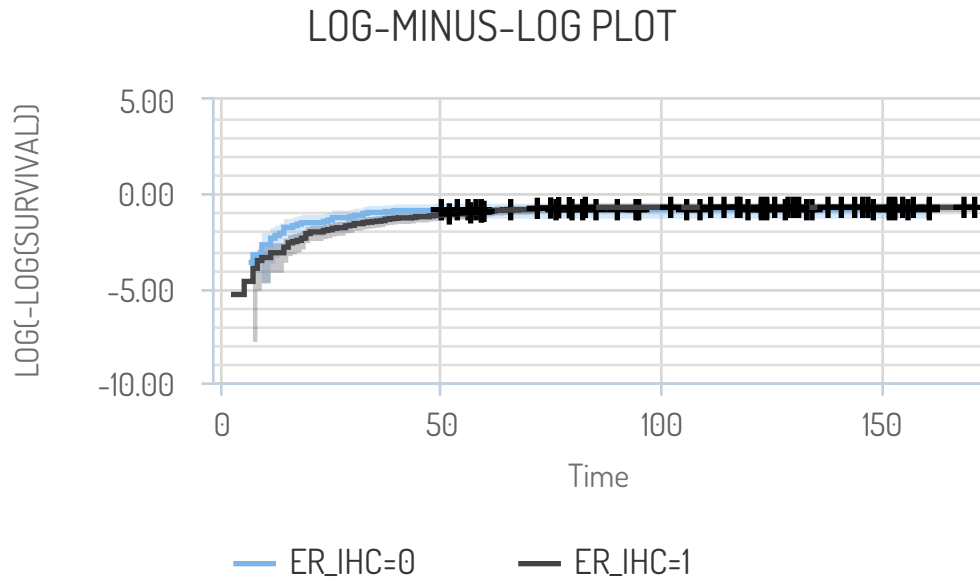
#### j. Hazard plot

Hazard plot can be created. A number of edit options is also available for plots.



#### k. Log-Minus-Log plot

Log-Minus-Log plot can be created. A number of edit options is also available for plots.



## 2.2. Cox Regression

### Concept

Cox regression, also known as proportional hazard regression, is a method to investigate the effect of one or multiple factors (i.e. gene expressions) upon the time an event of interest occurs. In this model, the effect of a unit increase in a factor is multiplicative with respect to the hazard rate.

### Usage

A Cox regression analysis can be conducted by applying the following steps:

1. Select the analysis method as **Cox Regression** from **Analysis** tab.
2. Select suitable variables for the analysis, such as **survival time**, **status variable**, **category value for status variable**, and categorical and continuous predictors for the model.
3. In advanced options, **interaction terms**, **strata terms** and **time dependent covariates** can be added to the model. Moreover, if there are multiple records for observations, users can specify it by clicking **Multiple ID** checkbox. Furthermore, one can choose model selection criteria, as **AIC** or **p-value**, model selection method, as **backward**, **forward** or **stepwise**, reference category, as **first** or **last**, and ties method, as **Efron**, **Breslow** or **exact** and change the **confidence level**.
4. Click **Run** button to run the analysis.

### Outputs

Desired outputs can be selected by clicking **Outputs** checkbox. Available outputs are coefficient estimates, hazard ratio, goodness of fit tests, analysis of deviance, predictions, residuals, Martingale residuals, Schoenfeld residuals and DfBetas.

#### a. Coefficient Estimates

A coefficient estimation table, which includes variable names, coefficient estimates and their associated standard errors, z statistics and p values, can be created.

<div>Copy CSV PDF Print</div>					Search: <input type="text"/>
	Variable	Coefficient estimate	Standard error	z statistic	p value
1	ER_IHC1	0.299	0.237	1.260	0.208
2	ESR1.205225_at	NA	0.000	NA	NA
3	PGR.208305_at	-0.155	0.046	-3.340	0.001

Showing 1 to 3 of 3 entries

Previous 1 Next

## b. Hazard ratio

A hazard ratio table, which includes variable names, hazard ratios and their associated lower and upper limits, can be created.

Copy

CSV

PDF

Print

Search:

	Variable	Hazard ratio	Lower limit (95%)	Upper limit (95%)
1	ER_IHC1	1.349	0.847	2.148
2	ESR1.205225_at	NA	NA	NA
3	PGR.208305_at	0.856	0.782	0.938

Showing 1 to 3 of 3 entries

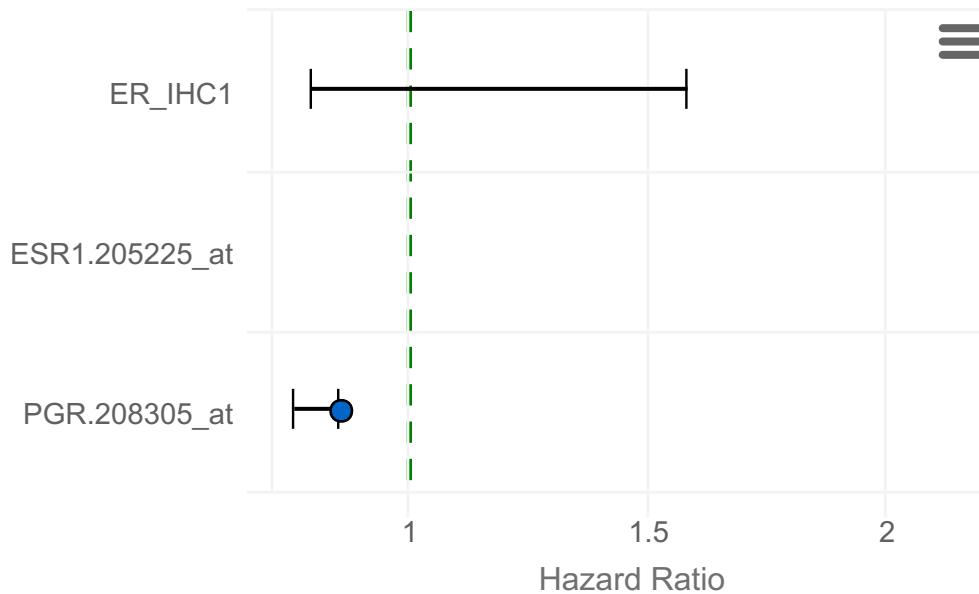
Previous

1

Next

## c. Hazard plot

A forest plot can be created for hazard ratios to give them a visual inspection.



## d. Goodness of Fit Tests

Fitted Cox regression model can be tested with three tests: Likelihood ratio, Wald, Score.



Copy	CSV	PDF	Print	Search: <input type="text"/>
Test	Statistic	df	p value	
Likelihood ratio	11.353	2	0.003	
Wald	11.170	2	0.004	
Score (logrank)	11.399	2	0.003	

Showing 1 to 3 of 3 entries

Previous 1 Next

### e. Analysis of Deviance

A deviance analysis can be conducted for each variable in the fitted model.

Copy	CSV	PDF	Print	Search: <input type="text"/>
Variable	df	Chisquare	p value	
ER_IHC	1	0.000	0.986	
ESR1.205225_at	0	0.000	1.000	
PGR.208305_at	1	11.353	0.001	

Showing 1 to 3 of 3 entries

Previous 1 Next

### f. Predictions

Predictions from the fitted model can be obtained.

Copy	CSV	PDF	Print	Search: <input type="text"/>
Predictions				
1	-0.247			
2	-0.247			
3	-0.247			
4	-0.247			
5	-0.247			
6	-0.247			
7	-0.247			
8	-0.247			
9	-0.247			
10	-0.247			

Showing 1 to 10 of 12,363 entries

Previous 1 2 3 4 5 ... 1237 Next

### g. Residuals

Residuals from the fitted model can be obtained.

Copy CSV PDF Print

Search:

### Residuals

1	-0.071
2	-0.071
3	-0.101
4	-0.124
5	-0.102
6	-0.126
7	-0.146
8	-0.073
9	-0.074
10	-0.181

Showing 1 to 10 of 12,363 entries

Previous **1** 2 3 4 5 ... 1237 Next

### h. Martingale Residuals

Martingale residuals from the fitted model can be obtained.

Copy CSV PDF Print

Search:

### Martingale Residuals

1	-0.003
2	-0.003
3	-0.005
4	-0.008
5	-0.005
6	-0.008
7	-0.011
8	-0.003
9	-0.003
10	-0.016

Showing 1 to 10 of 12,363 entries

Previous **1** 2 3 4 5 ... 1237 Next

### i. Schoenfeld Residuals

Schoenfeld residuals from the fitted model can be obtained.

Copy	CSV	PDF	Print	Search:	
	ER_IHC1	ESR1.205225_at	PGR.208305_at		
1	0.266	0	1.904		
2	0.267	0	3.261		
3	-0.734	0	-1.975		
4	-0.734	0	0.888		
5	0.263	0	-0.853		
6	-0.737	0	1.503		
7	0.263	0	0.887		
8	0.264	0	-2.194		
9	0.264	0	-2.144		
10	-0.735	0	-1.936		

Showing 1 to 10 of 107 entries

Previous 1 2 3 4 5 ... 11 Next

## j. DfBetas

DfBetas residuals from the fitted model can be obtained.

Copy	CSV	PDF	Print	Search:	
	ER_IHC1	ESR1.205225_at	PGR.208305_at		
1	-0.000	NaN	-0.000		
2	-0.000	NaN	-0.000		
3	-0.000	NaN	-0.001		
4	-0.000	NaN	-0.001		
5	-0.000	NaN	-0.001		
6	-0.000	NaN	-0.001		
7	-0.000	NaN	-0.001		
8	-0.000	NaN	-0.000		
9	-0.000	NaN	-0.000		
10	-0.000	NaN	-0.002		

Showing 1 to 10 of 12,363 entries

Previous 1 2 3 4 5 ... 1237 Next

## k. Proportional Hazard Assumption

### l. Proportional Hazard Test

To check the proportionality assumption of Cox regression model, a proportional hazard test can be conducted both globally and for each variable in the fitted model.

Copy

CSV

PDF

Print

Search:

	Rho	Chi-square statistic	p value
ER_IHC1	0.310	10.891	0.001
ESR1.205225_at	NA	NaN	NaN
PGR.208305_at	-0.014	0.023	0.880
GLOBAL	NA	12.017	0.007

Showing 1 to 4 of 4 entries

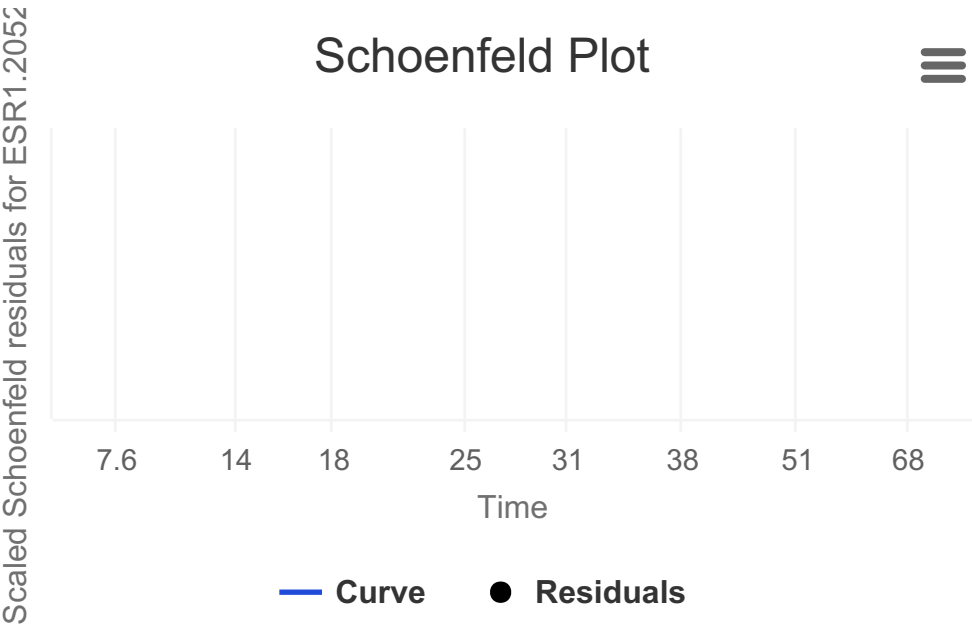
Previous

1

Next

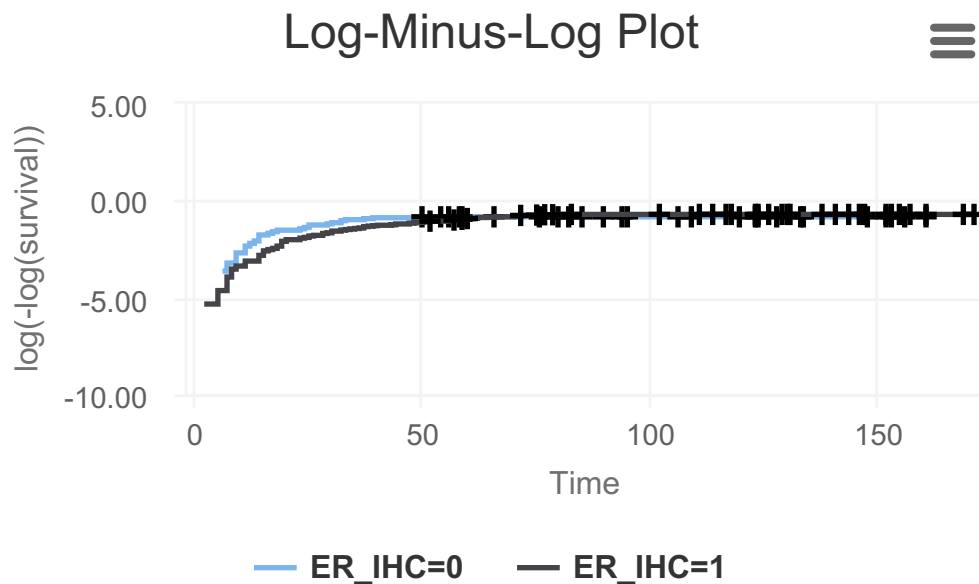
m. Schoenfeld Plot

Beside a formal test for proportionality assumption, a Schoenfeld plot can be created to check the assumption visually.



n. Log-Minus-Log Plot

Another useful plot for checking proportionality assumption is log-minus-log plot. Lines should be parallel to each other to satisfy proportionality.



## 2.3. Penalized Cox Regression

### Concept

Feature selection is an useful strategy to avoid over-fitting, to obtain more reliable predictive results, and to provide more insights into the underlying casual relationships (Ma and Huang, 2008). In this section, a feature selection can be performed using ridge, elastic net or lasso penalty, especially when there are too many predictors (e.g.  $n < p$ ). More information can be found in Zou and Hastie, 2005, Freidman et al, 2008 and Simon et al, 2011.

### Usage

A Penalized Cox regression analysis can be conducted by applying the following steps:

1. Select the analysis method as **Penalized Cox Regression** from **Analysis** tab.
2. Select suitable variables for the analysis, such as **survival time**, **status variable**
3. If all predictors are continious then one can check the **Select All Variables** option to include all variables in dataset to the feature selection process. If some predictors categorical and others are continious, then uncheck the **Select All Variables** option and select categorical and continuous variables seperately.
4. Define the penalty term using the **Penalty term** slider as follow:

**Penalty term = 0:** ridge penalty **0 < Penalty term < 1:** elastic net penalty **Penalty term = 1:** lasso penalty

5. Select the number of folds for cross-validation. Note that number of folds must be greater than 3.
6. Click **Run** button to run the analysis.

### Outputs

#### a) Variables in the model

Variable selection is conducted with the selected penalized method (i.e. ridge, elasticnet, lasso) and results will be displayed as a table, which includes selected variables and their associated coefficient estimates.

Copy

CSV

PDF

Print

Search:

	Variable	Coefficient estimate
10	PGR.208305_at	-0.033

Showing 1 to 1 of 1 entries

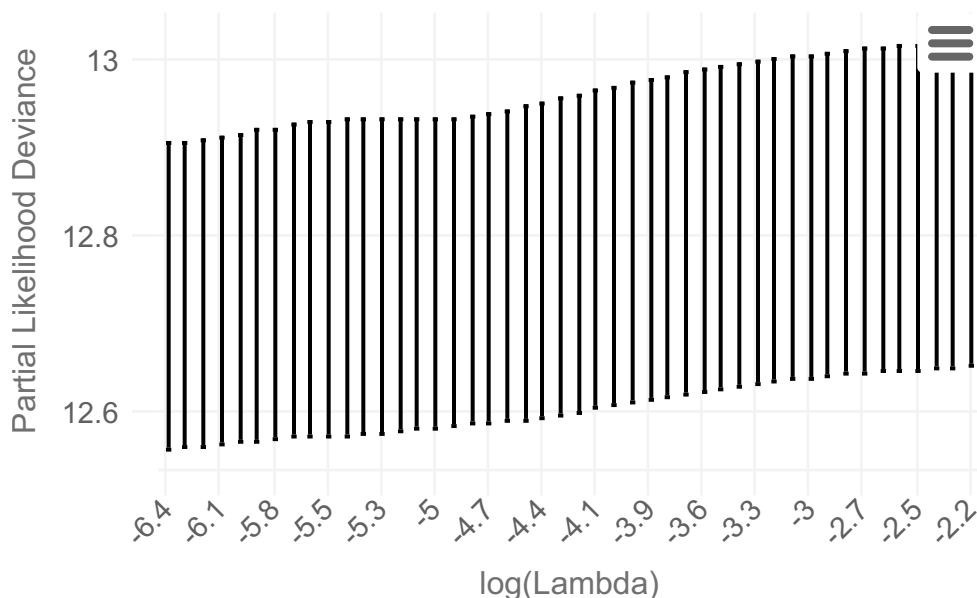
Previous

1

Next

## b) Cross-validation curve

A cross-validation curve can be created to investigate the relationship between partial likelihood devaince and lambda values.



## 2.4. Random Survival Forests

### Concept

Random survival forests, an ensemble method for analysing right censored data, first introduced by Ishwaran et al, 2008. RSF has several advantages over Cox regression: (i) Unlike Cox regression, RSF does not rely on proportional hazard assumption. (ii) RSF accounts for nonlinear effects and interactions for factor variables.

### Usage

A random survival forests analysis can be conducted by applying the following steps:

1. Select the analysis method as Random Survival Forests from Analysis tab.
2. Select suitable variables for the analysis, such as survival time, status variable, category value for status variable, and categorical and continuous predictors for the model.
3. In advanced options, interaction terms, strata terms and time dependent covariates can be added to the model. Moreover, if there are multiple records for observations, users can specify it by clicking Multiple ID checkbox. From RSF options, number of tree, bootstrap method, randomly selected number of variable, minimum number of cases in terminal node, maximum depth for a tree, splitting rule, number of split, missing values, number of iterations of the missing data algorithm, proximity of cases, size of bootstrap and type of bootstrap can be adjusted.

4. Click **Run** button to run the analysis.

## Outputs

##

## Trees Grown: 682, Time Remaining (sec): 1

### a. Individual Survival Predictions

Survival predictions for each observation can be obtained. In this table, rows represent observations whereas columns represent time endpoints.

Copy	CSV	PDF	Print	Search: <input type="text"/>													
	2	5	6	7	8	9	11	12	13	14	15	16	17	18	19	20	
1	1	1	1	1	1	0.998	0.998	0.998	0.998	0.998	0.951	0.951	0.951	0.951	0.951	0.875	
2	1	1	1	0.98	0.979	0.979	0.979	0.979	0.956	0.93	0.93	0.93	0.929	0.818	0.711	0.711	
3	1	1	1	1	1	1	1	1	1	0.998	0.998	0.998	0.998	0.998	0.998	0.998	
4	1	1	0.995	0.995	0.995	0.995	0.994	0.994	0.994	0.994	0.994	0.95	0.95	0.899	0.877	0.877	
5	1	1	1	1	1	1	0.998	0.998	0.998	0.998	0.991	0.98	0.98	0.98	0.967	0.967	
6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
8	1	1	1	1	1	1	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.979	0.963	0.962	
9	1	1	0.981	0.981	0.975	0.971	0.971	0.971	0.971	0.971	0.912	0.911	0.911	0.911	0.768	0.768	
10	1	1	1	1	1	1	1	1	1	1	0.991	0.991	0.991	0.991	0.991	0.991	
Showing 1 to 10 of 286 entries																	
Previous										1	2	3	4	5	...	29	Next

•

### b. Individual Survival Predictions OOB

Out of bag (OOB) survival predictions for each observation can be obtained. In this table, rows represent observations whereas columns represent time endpoints.

Copy	CSV	PDF	Print	Search: <input type="text"/>													
	2	5	6	7	8	9	11	12	13	14	15	16	17	18	19	20	
1	1	1	1	1	1	0.998	0.998	0.998	0.998	0.998	0.913	0.913	0.913	0.913	0.913	0.757	
2	1	1	1	0.973	0.973	0.973	0.973	0.973	0.946	0.903	0.903	0.903	0.902	0.807	0.661	0.661	
3	1	1	1	1	1	1	1	1	1	0.996	0.996	0.996	0.996	0.996	0.996	0.996	
4	1	1	0.991	0.991	0.991	0.99	0.988	0.988	0.988	0.988	0.988	0.887	0.887	0.777	0.737	0.737	
5	1	1	1	1	1	1	0.996	0.996	0.996	0.996	0.983	0.962	0.962	0.962	0.94	0.94	
6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
8	1	1	1	1	1	1	1	1	1	1	0.999	0.999	0.999	0.998	0.997	0.995	
9	1	1	0.957	0.957	0.942	0.935	0.933	0.933	0.933	0.933	0.799	0.799	0.799	0.799	0.497	0.497	
10	1	1	1	1	1	1	1	1	1	1	0.983	0.983	0.983	0.983	0.983	0.983	
Showing 1 to 10 of 286 entries																	
Previous										1	2	3	4	5	...	29	Next

c. Individual Cumulative Hazard Predictions

Cumulative hazard predictions for each observation can be obtained. In this table, rows represent observations whereas columns represent time endpoints.

Copy CSV PDF Print				Search: <input type="text"/>						
	2	5	6	7	8	9	11	12	13	14
1	0.00000725	0.0000429	0.0000429	0.0000429	0.0000429	0.00212	0.00212	0.00212	0.00212	0.00215
2	0	0	0	0.0205	0.0205	0.0205	0.0206	0.0206	0.0443	0.0746
3	0.00000725	0.0000429	0.0000429	0.0000429	0.0000429	0.0000503	0.0000503	0.0000503	0.0000503	0.00236
4	0	0	0.00453	0.00453	0.00453	0.00506	0.00574	0.00574	0.0058	0.00589
5	0	0	0	0.0000945	0.000245	0.000495	0.00243	0.00243	0.00243	0.00251
6	0.00000725	0.0000429	0.0000429	0.0000429	0.0000429	0.0000503	0.0000503	0.0000503	0.0000503	0.0000793
7	0.00000725	0.0000429	0.0000429	0.0000429	0.0000429	0.0000503	0.0000503	0.0000503	0.0000503	0.0000793
8	0.00000725	0.0000218	0.0000218	0.000116	0.000183	0.000191	0.00163	0.00163	0.00163	0.00171
9	0	0	0.0189	0.019	0.0254	0.0287	0.0295	0.0295	0.0295	0.0296
10	0.00000725	0.0000429	0.0000429	0.0000429	0.0000429	0.0000503	0.0000503	0.0000503	0.0000503	0.0000793

Showing 1 to 10 of 286 entries

Previous 1 2 3 4 5 ... 29 Next

d. Individual Cumulative Hazard Predictions OOB

Out of bag (OOB) cumulative hazard predictions for each observation can be obtained. In this table, rows represent observations whereas columns represent time endpoints.



Copy	CSV	PDF	Print	Search: <input type="text"/>						
	2	5	6	7	8	9	11	12	13	14
1	0	0	0	0	0	0.00246	0.00246	0.00246	0.00246	0.00246
2	0	0	0	0.0266	0.0266	0.0266	0.0267	0.0267	0.0544	0.108
3	0.0000201	0.0000607	0.0000607	0.0000607	0.0000607	0.0000813	0.0000813	0.0000813	0.0000813	0.00371
4	0	0	0.00927	0.00927	0.00927	0.00998	0.0118	0.0118	0.0118	0.0118
5	0	0	0	0.0000459	0.000116	0.000116	0.00383	0.00383	0.00383	0.00388
6	0	0.0000575	0.0000575	0.0000575	0.0000575	0.0000575	0.0000575	0.0000575	0.0000575	0.000116
7	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0.000123	0.000236	0.000236	0.000442	0.000442	0.000442	0.000494
9	0	0	0.0425	0.0426	0.0582	0.0658	0.0673	0.0673	0.0673	0.0673
10	0	0	0	0	0	0	0	0	0	0

Showing 1 to 10 of 286 entries

Previous 1 2 3 4 5 ... 29 Next

### e. Error Rate

An error rate table, which shows error rate estimations for each tree, can be obtained.

Copy	CSV	PDF	Print	Search: <input type="text"/>						
	Number of tree							Error rate		
1	1							0.474		
2	2							0.482		
3	3							0.524		
4	4							0.52		
5	5							0.503		
6	6							0.521		
7	7							0.505		
8	8							0.513		
9	9							0.513		
10	10							0.501		

Showing 1 to 10 of 1,000 entries

Previous 1 2 3 4 5 ... 100 Next

### f. Variable Importance

A variable importance table as well as an interactive plot, which shows relative importance of variables in fitted model, can be obtained.

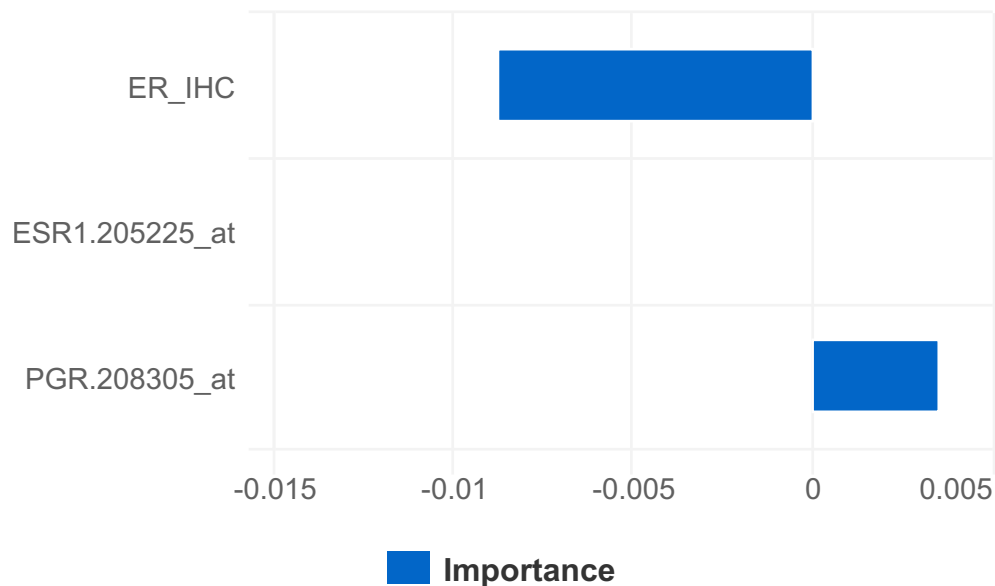
	variableImp
ER_IHC	-0.00995
ESR1.205225_at	0
PGR.208305_at	0.00441

Showing 1 to 3 of 3 entries

Previous

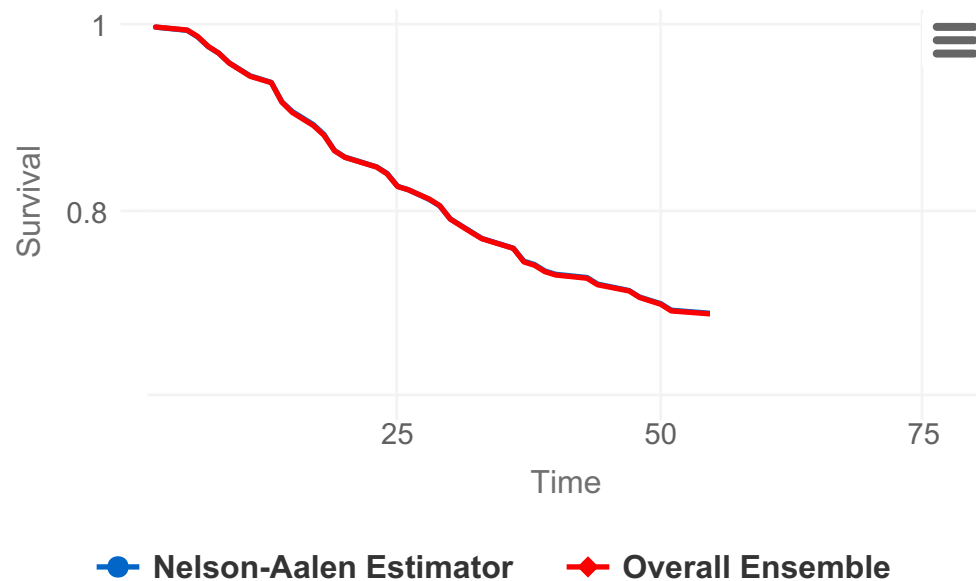
1

Next



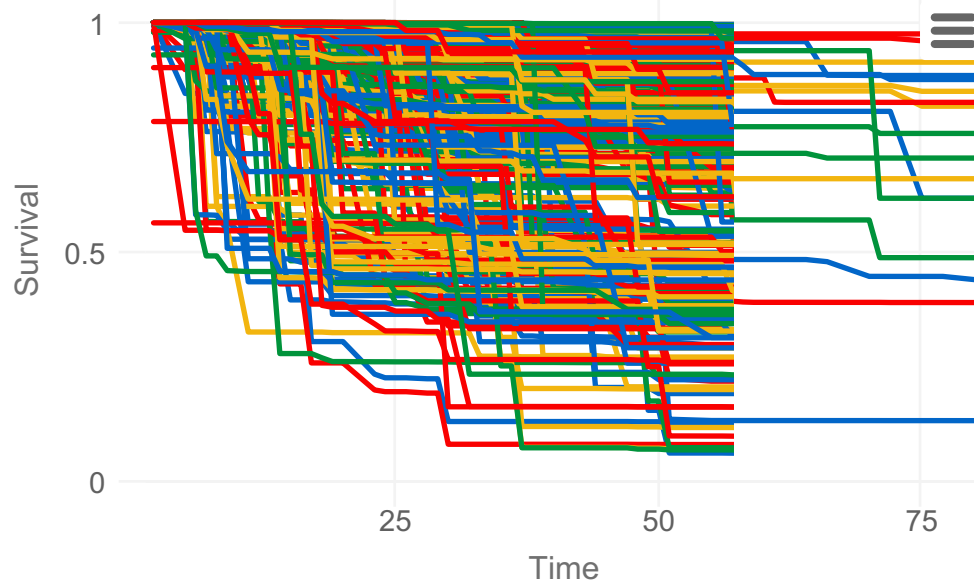
### g. Overall Survival Plot

A survival plot can be created based on Nelson-Aalen estimator and overall ensemble predictions.



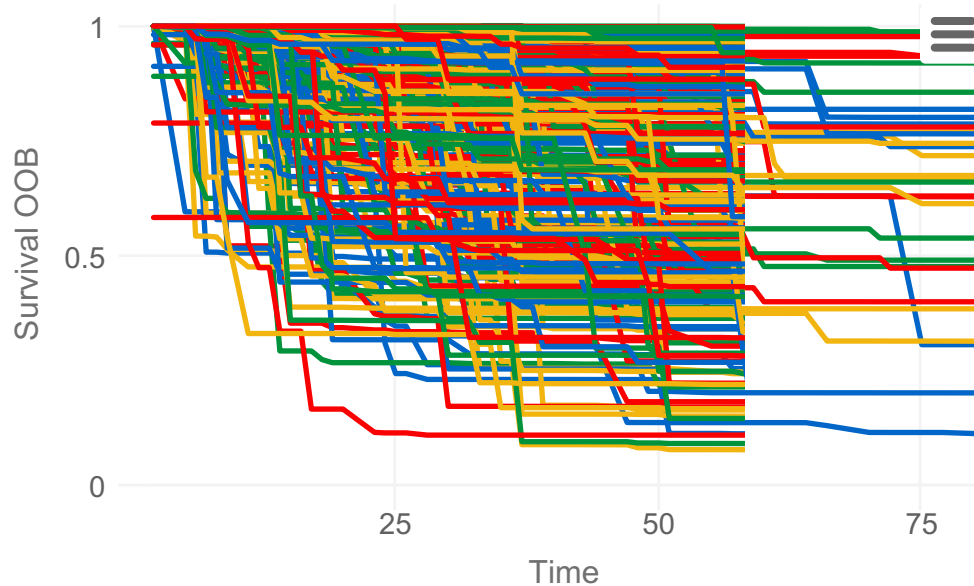
### h. Individual Random Survival Plot

A survival plot can be drawn for survival predictions from random survival forests model. Each line represents a survival curve for each observation.



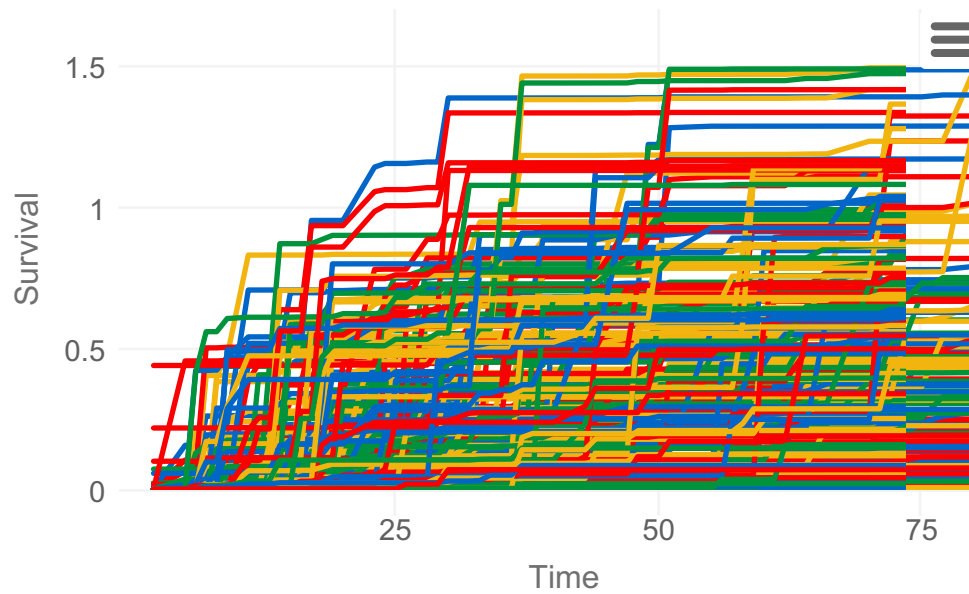
#### i. Individual Survival OOB Plot

A survival plot can be drawn for OOB survival predictions from random survival forests model. Each line represents a survival curve for each observation.



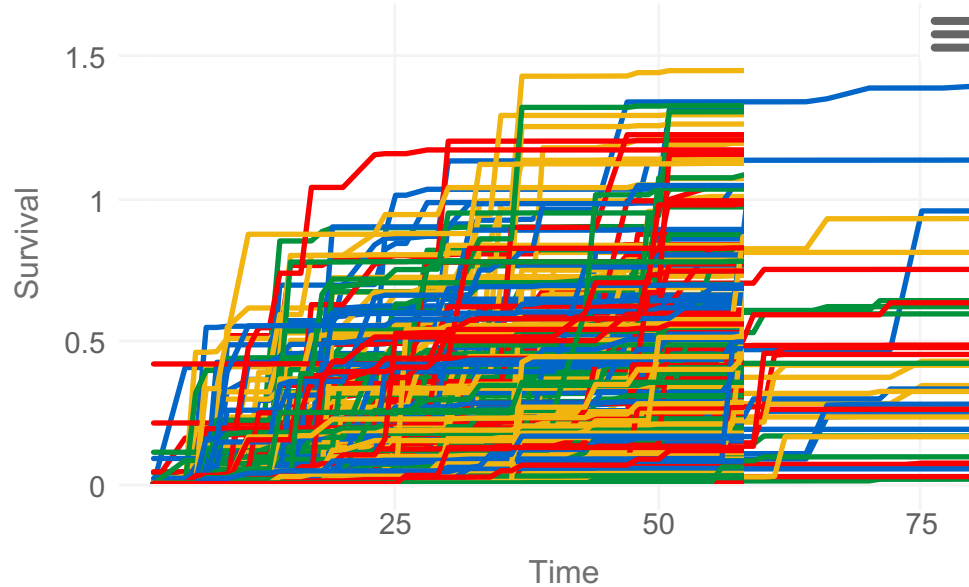
#### j. Individual Cumulative Hazard Plot

A cumulative hazard plot can be drawn for hazard predictions from random survival forests model. Each line represents a survival curve for each observation.



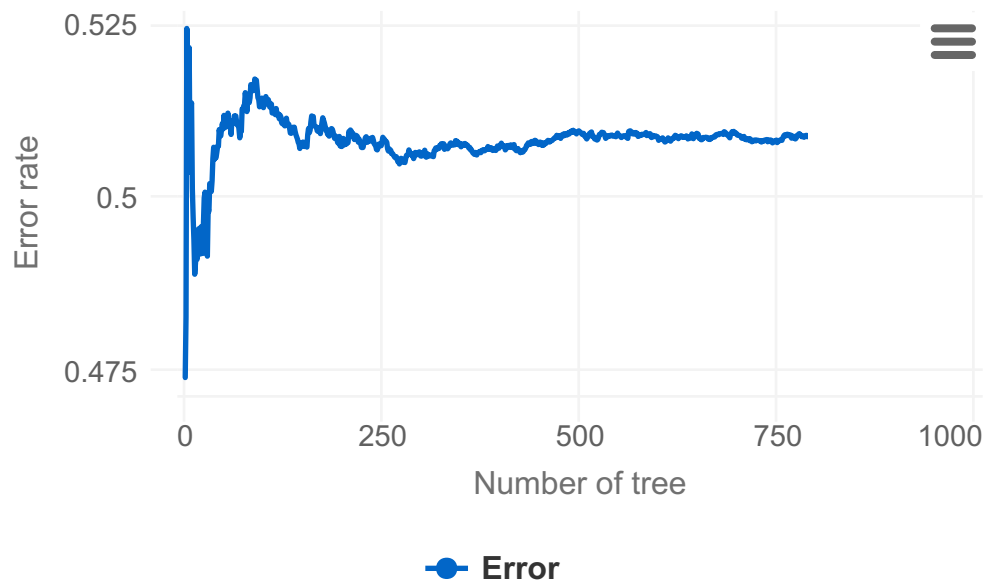
#### k. Individual Cumulative Hazard OOB Plot

A cumulative hazard plot can be drawn for OOB cumulative hazard predictions from random survival forests model. Each line represents a survival curve for each observation.



#### l. Error Rate Plot

An interactive error rate plot, which shows error rate alterations when number of trees increased, can be drawn.

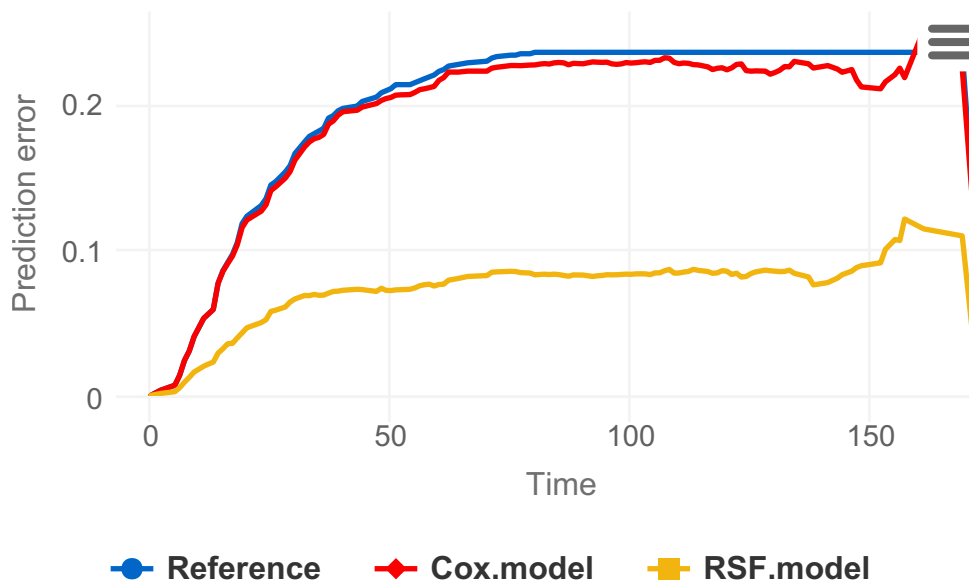


#### m. Cox vs RSF

A Cox model can be compared to random survival forests model through an interactive plot for visual inspection of both models.

##

## Trees Grown: 794, Time Remaining (sec): 1



#### 2.5. Optimal Cutoff

Concept

To investigate whether the higher or lower expressions of differentially expressed genes lead to more survival risks for patients, expression levels of genes can be dichotomized based on certain cutoff values by maximizing certain test statistics.

Usage

An optimal cutoff value can be determined by applying the following steps:

- 1. Select the analysis method as **Optimal Cutoff** from **Analysis** tab.
- 2. Select one or more markers from **Select marker(s)** box
- 3. Select appropriate **Survival time** and **Status variable**
- 4. Select a category for status variable
- 5. Select a test statistic (log-rank, Gehan-Breslow, Tarone-Ware, Petp-Peto, Modified Petp-Peto, Flemington-Harrington) for optimal cutoff determination.
- 6. In advanced options, one can change confidence interval type, as log, log-log or plain, variance estimation method, as Greenwood or Tsiatis, Flemington-Harrington weights and confidence level
- 7. Click **Run** button to run the analysis.

Outputs

a) Optimal cutoff value(s)

An optimal cutoff value can be obtained as well as hazard ratio (HR) with confidence interval, mean survival time for low and high gene expression levels, and p value for selected significance test.

Copy

CSV

PDF

Print

Search:

	Optimal cut-off	HR	HR (lower)	HR (upper)	Mean survival time (low expression)	Mean survival time (high expression)	p value
PGR.208305_at	2.544	0.462	0.301	0.708	88.794	124.791	0.000

Showing 1 to 1 of 1 entries

Previous1Next

#####b) Optimal cutoff value(s) A Kaplan-Meier plot can be created after dichotomize the gene expression level as high and low.

