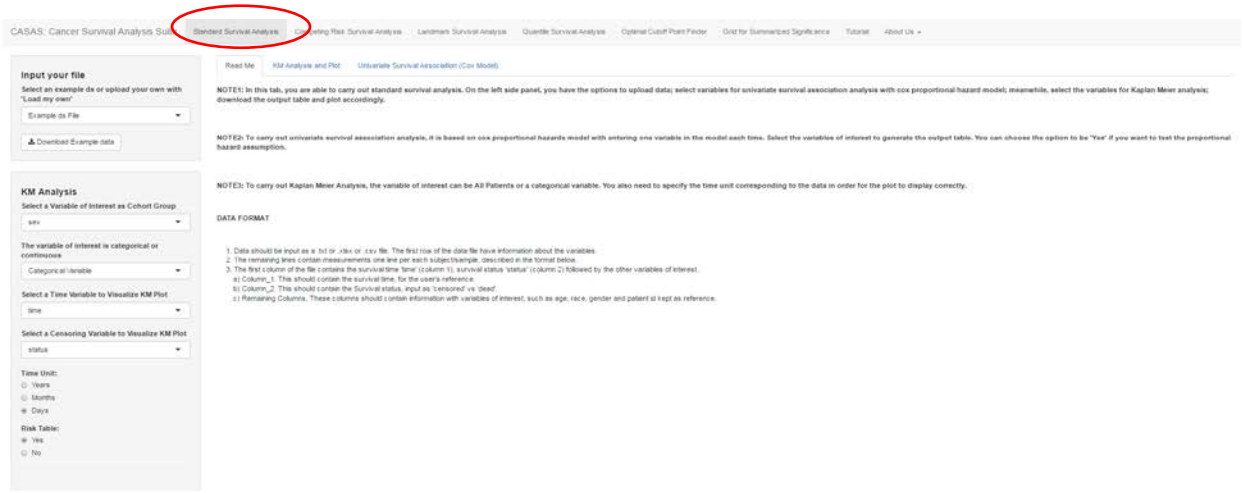
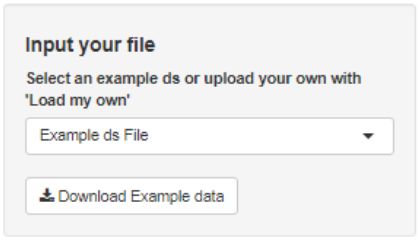
CASAS: Cancer Survival Analysis Suite Tutorial

Table of Contents

I. CASAS TUTORIAL

TAB 1. STANDARD SURVIVAL ANALYSIS

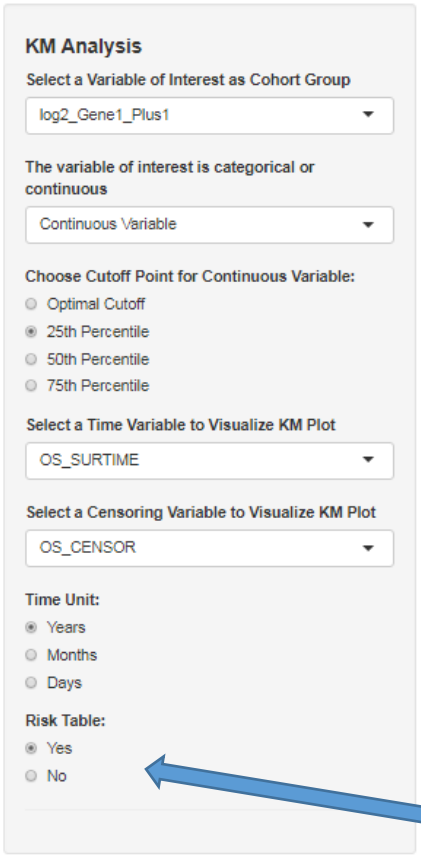




Step 1:

Select the example dataset or upload your own.

To view example data and format, use download button to view .csv file.



Step 2:

The variables from the input file will be available as drop down for selection.

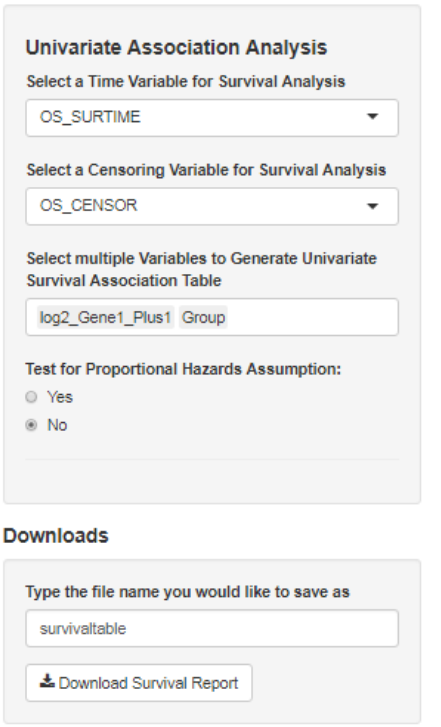
Choose variable of interest to divide the cohort. If continuous, a new set of parameters appear to dichotomize the continuous variable. The user can select an optimal cut-off based on the data, or either subset it by 25th or 50th or 75th percentile. Any sample with values above 25th percentile will be considered as “High” and remaining “Low”. Similarly for the other groupings.

If this variable is categorical, change drop down to categorical and the cutoff choice will disappear.

Also, choose appropriate variable for survival time, censor status (0- Censor, 1- Event) variables from your uploaded dataset to carry out Kaplan Meir analysis. User can also select the appropriate time unit (Years, Months or Days). Years being the default here.

Step 3:

User can opt to display the number at risk table under the KM curve in each categorical group.



Step 4:

Choose appropriate survival time, censor status (0- Censor, 1- Event) variables from your uploaded dataset for conducting the Univariate analysis.

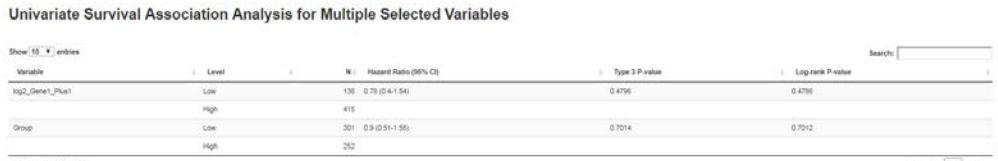
Enter variable(s) to generate the univariate analysis table to test association between variable and survival using the Cox PH Model.

User can also test for PH assumption. An additional column will be added with the PH assumption p-value.

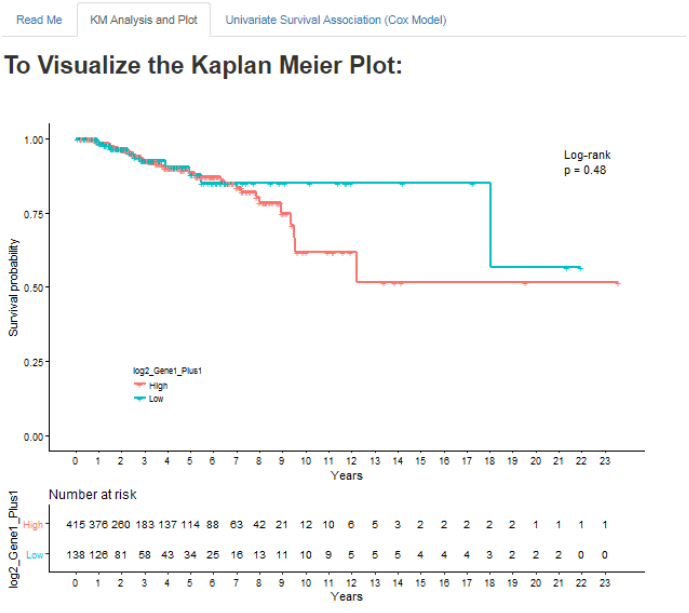
Step 6:

User can download the results from the univariate analysis association for the variable(s) of his/her interest.

Similar, download is available on the KM Analysis and Plot tab.

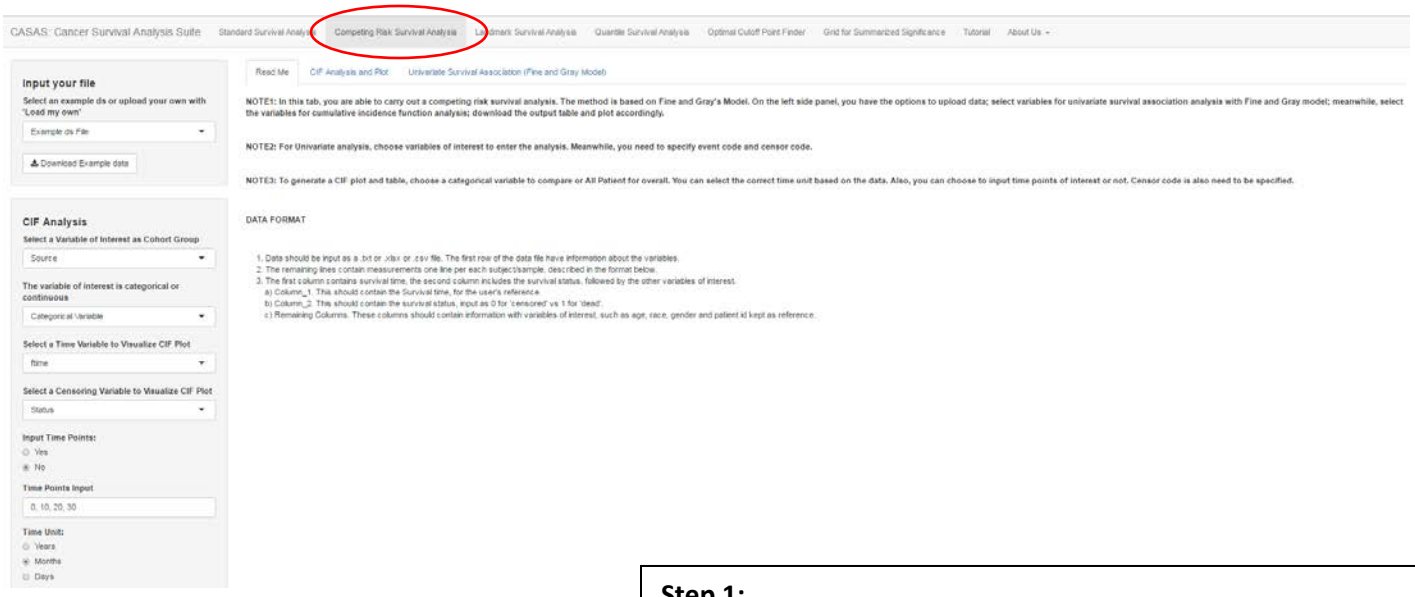


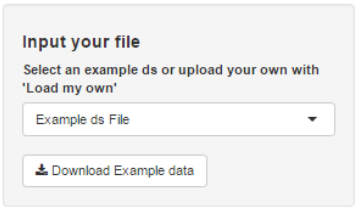
Univariate Survival Association analysis table is displayed for the single or multiple variables selected by the user. For example, here categorical variable Group and dichotomized continuous variable (based on 25th percentile cut point) are being tested.



KM plot stratified by the continuous variable log2_Gene1_Plus1. An at risk table is displayed under the KM plot. The time unit is years but can be changed using the options mentioned previously.

TAB 2. COMPETING RISK SURVIVAL ANALYSIS

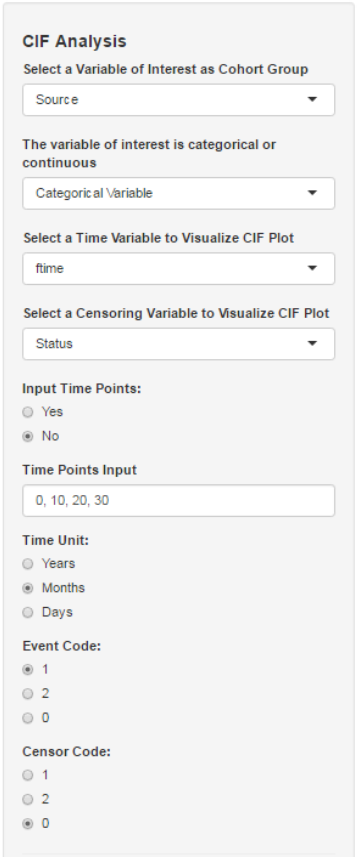




Step 1:

Select the example dataset or upload your own.

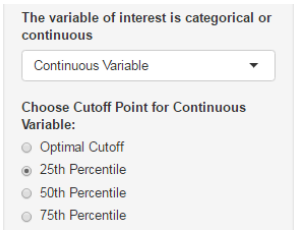
To view example data and format, use download button to view .csv file.



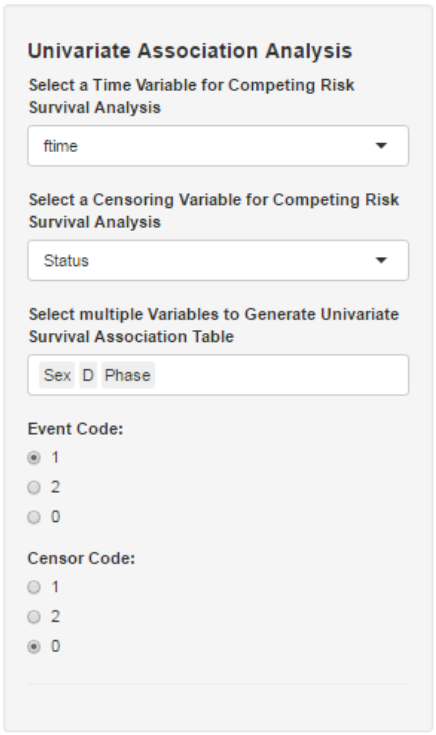
Step 2:

The variables from the input file will be available as drop down for selection.

Choose variable of interest to divide the cohort. If this variable is categorical, leave next drop down unchanged. If continuous, a new set of parameters will appear to dichotomize the continuous variable. The user can select an optimal cut-off based on the data, or either subset it by 25th or 50th or 75th percentile. Any sample with values above 25th percentile will be considered as "High" and remaining "Low". Similarly for the other groupings.



Choose appropriate survival time, censor status (0/1/2) variables from your uploaded dataset. For competing risk analysis, the user needs to select the code for the event and the censor. The third/last code would be the competing risk. User can also select the appropriate time unit (Years, Months or Days). Months being the default.



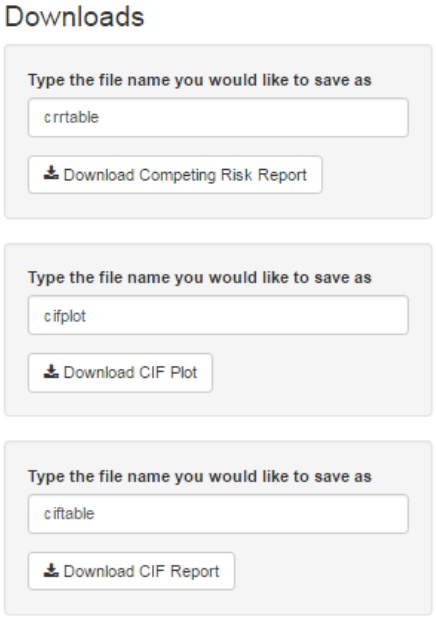
Step 4:

Choose the time points at which CIF Estimate is to be reported. User can input time points of their choice or use the predefined time points at 0, 10, 20, 30 days.

Step 5:

Choose appropriate survival time, censor status (0- Censor, 1- Event, 2- Competing Risk) variables from your uploaded dataset to carry out CIF analysis. The user can select different status codes based on their data.

Enter variable(s) to generate the univariate analysis table to test association between variable and survival using the Fine and Grey Model.



Step 6:

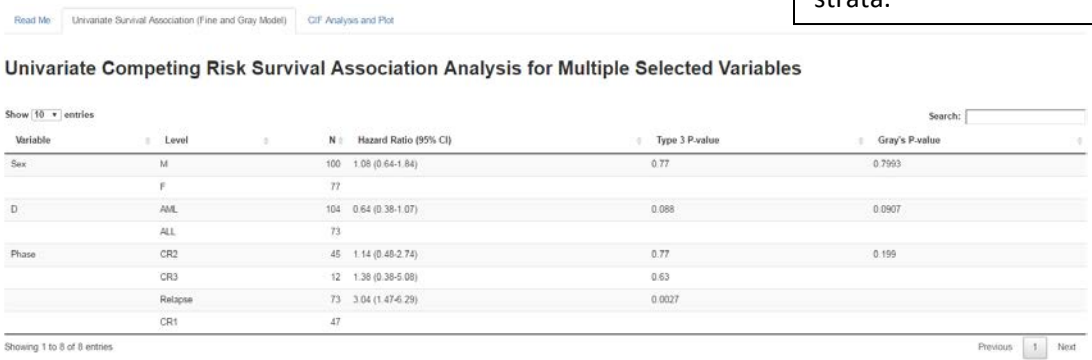
User can download the results from the univariate analysis association for the variable(s) of his/her interest.

Step 7:

User can download the CIF plot stratified by the categorical variable by event and competing risk. Gray's p-value is reported.

Step 8:

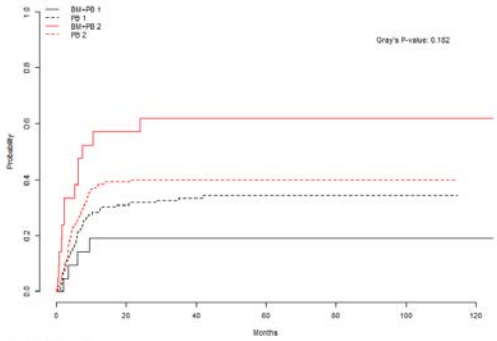
User can download the CIF result table for the time points of interest based on the strata.



Univariate Survival Association analysis table is displayed for the single or multiple variables selected by the user. For example, here association of Sex, Disease type and Phase of the disease are being tested.

Read Me | Univariate Survival Association (Fine and Gray Model) | CIF Analysis and Plot

To Visualize the Cumulative Incidence Function Plot:



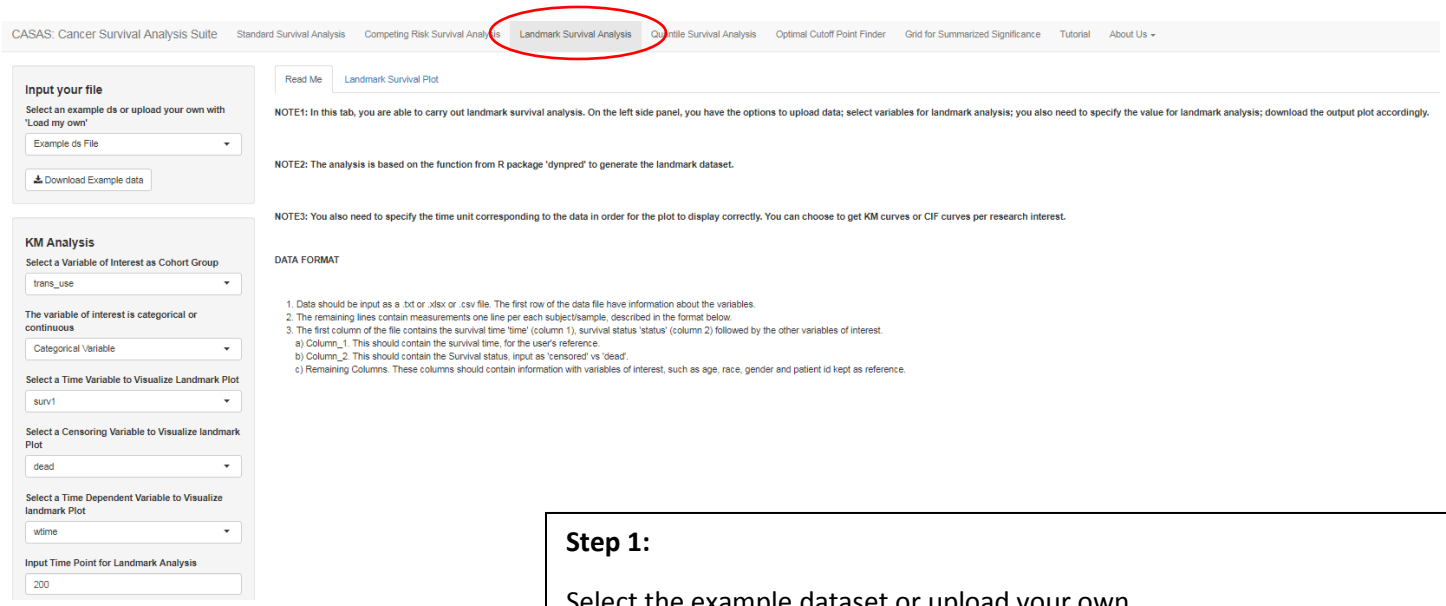
CIF plot stratified by the categorical variable source and further by the event and the

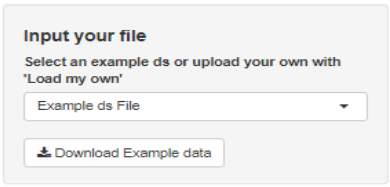
Show 100 entries Search:

Source	Time (Months)	CIF Estimate (95% CI)
BM+PB 1	0	0 (NaN, NaN)
	20	0.1905 (0.055, 0.3874)
	40	0.1905 (0.055, 0.3874)
	60	0.1905 (0.055, 0.3874)
	80	0.1905 (0.055, 0.3874)
	100	0.1905 (0.055, 0.3874)
	120	0.1905 (0.055, 0.3874)

CIF estimate table at 10, 20, 30 days (Time points) chosen by the user.

TAB 3. LANKMARK SURVIVAL ANALYSIS

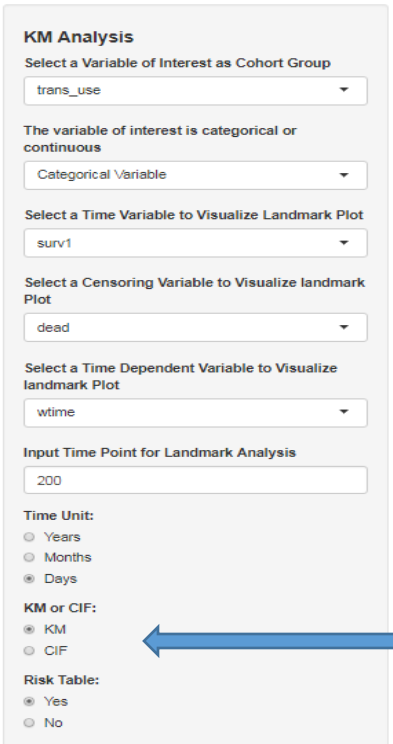




Step 1:

Select the example dataset or upload your own.

To view example data and format, use download button to view .csv file.



Step 2:

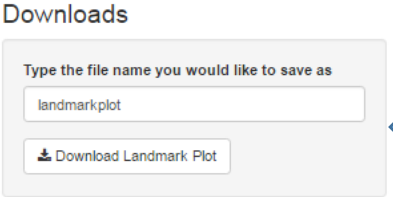
The variables from the input file will be available as drop down for selection.

Choose variable of interest to divide the cohort. If this variable is categorical, leave next drop down unchanged. If continuous, a new set of parameters will appear to dichotomize the continuous variable. The user can select an optimal cut-off based on the data, or either subset it by 25th or 50th or 75th percentile. Any sample with values above 25th percentile will be considered as "High" and remaining "Low". Similarly for the other groupings.

For Landmark analysis, additionally user needs to input the time point for landmark analysis and choose the time dependent variable. This should be in the same time unit as the survival time. Survival time unit also needs to be specified. By default, it is days.

Step 3:

Choose to display either a KM curve or a CIF curve for the landmark analysis. If user selects a KM curve option, a combined KM curve for overall survival until the landmark time followed by a landmark KM curve is generated with their respective p-values.

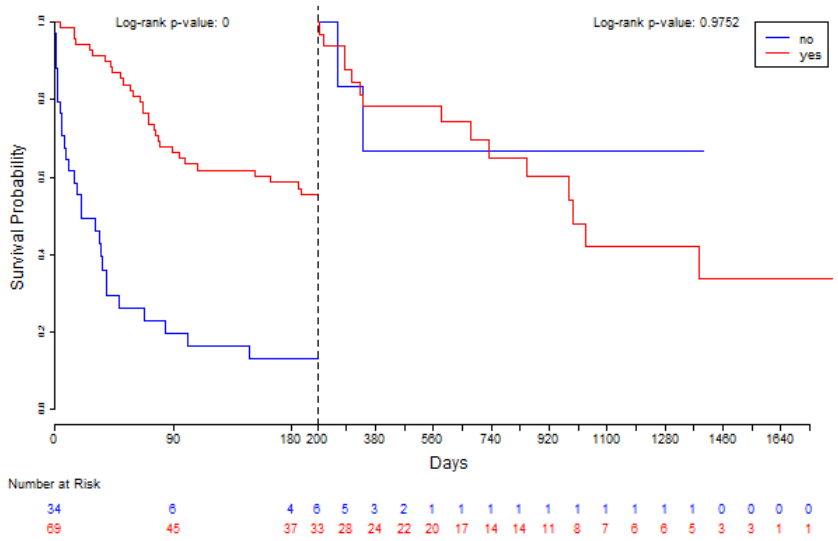


Step 4:

User can download the Landmark plot stratified by the categorical variable. Log rank p-values are also reported.

Read Me | Landmark Survival Plot

To Visualize the Landmark Survival Plot:

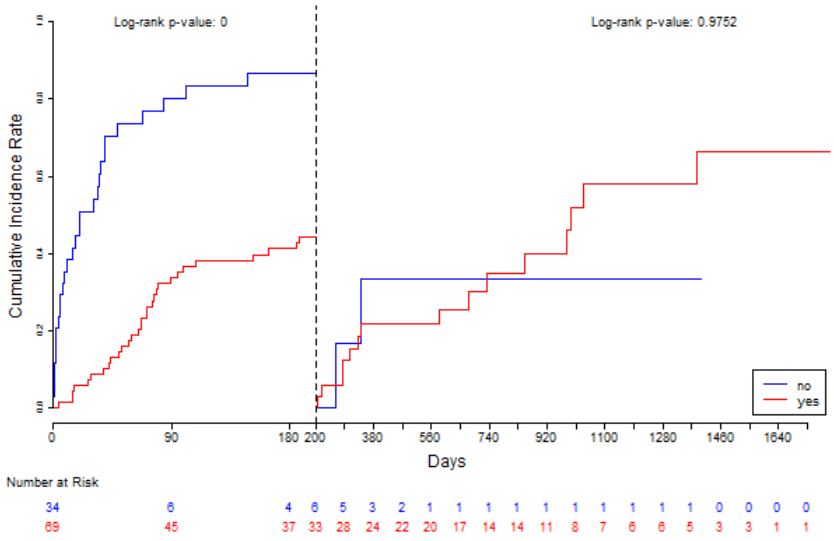


A landmark analysis KM plot stratified by the time dependent categorical variable 'trans_use'. Left side shows the overall survival curve and the right side represents the landmark survival plot.

User can optionally have the numbers of risk table under the plot.

Read Me | Landmark Survival Plot

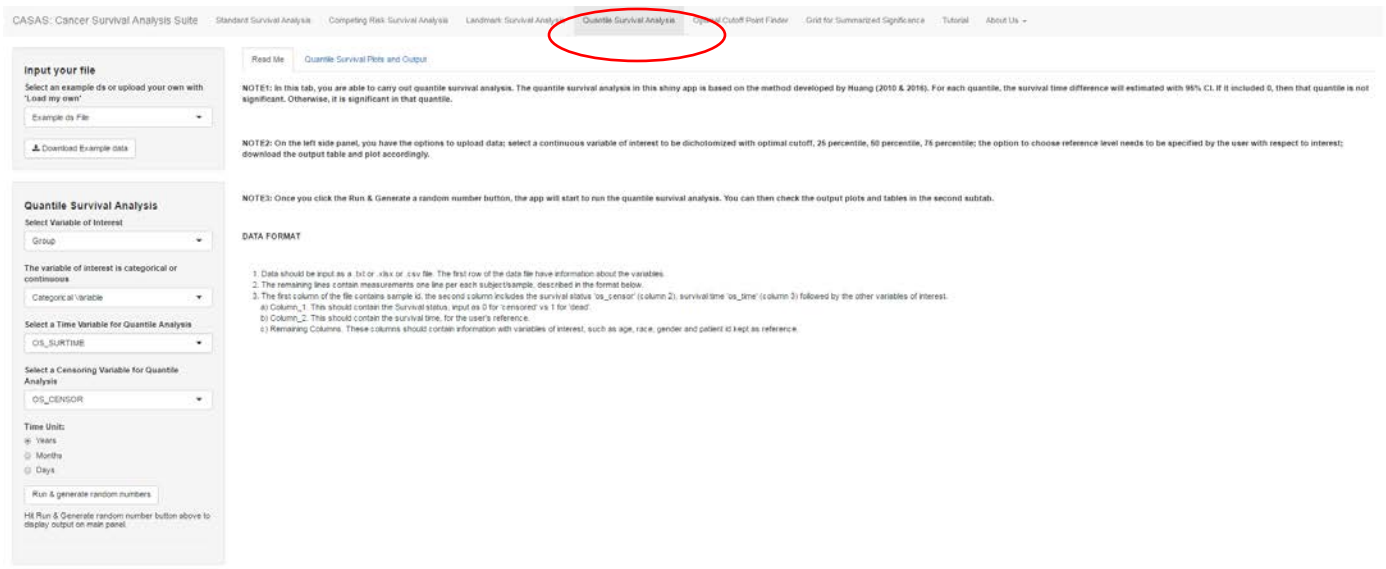
To Visualize the Landmark Survival Plot:

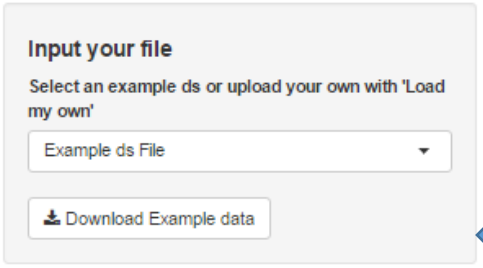


Alternatively, a landmark analysis CIF plot stratified by the categorical variable.

Left side shows the overall survival curve and the right side represents the landmark survival plot.

TAB 4. QUANTILE SURVIVAL ANALYSIS

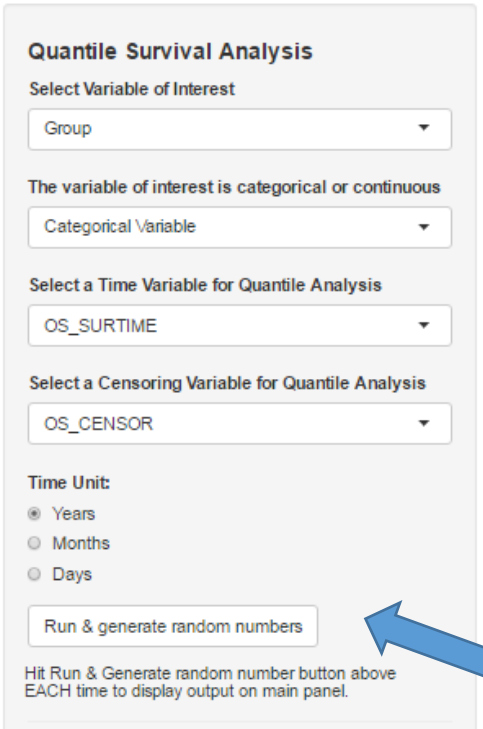




Step 1:

Select the example dataset or upload your own.

To view example data and format, use download button to view .csv file.



Step 2:

The variables from the input file will be available as drop down for selection.

Choose variable of interest to divide the cohort. If this variable is categorical, leave next drop down unchanged. If continuous, a new set of parameters will appear to dichotomize the continuous variable. The user can select an optimal cut-off based on the data, or either subset it by 25^{th} or 50^{th} or 75^{th} percentile. Any sample with values above 25^{th} percentile will be considered as "High" and remaining "Low". Similarly for the other groupings.

Choose appropriate survival time, censor status (0/1) variables from your uploaded dataset. User can also select the appropriate time unit (Years, Months or Days). Years being the default.

Step 3:

In order to run the analysis, user needs to hit 'Run & generate a random number' each time they select a new variable of interest.

Downloads

Type the file name you would like to save as

QAplot

Download Quantile Survival Plot

Step 4:

User can download the Quantile Survival plot. This includes the overall KM analysis plot and the plot of mean survival differences between the two groups for each quantile.

Type the file name you would like to save as

forestplot

Download Forest Plot

Step 5:

User can download the Forest plot summary. This includes the overall KM analysis plot and the plot of mean survival differences between the two groups for each quantile.

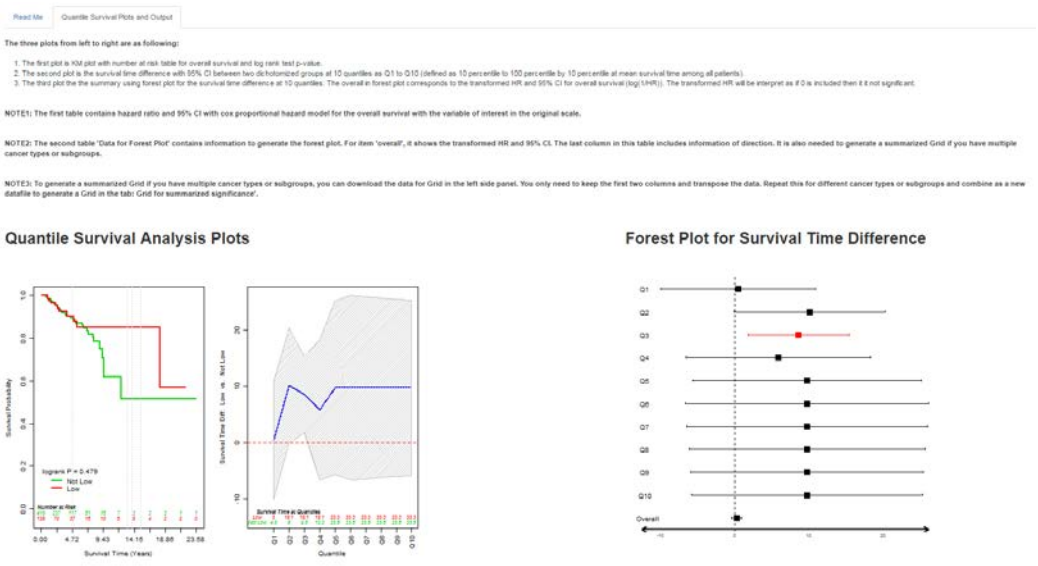
Type the file name you would like to save as

Data for grid

Download grid Data

Step 6:

If a user is interested in multiple genes (variables) in the same dataset or multiple datasets (for example, numerous cancer types), they can download the table of mean difference estimates for a comparative grid in tab 6.



Left to Right: Quantile Survival analysis plots stratified by the dichotomous groups, survival time difference with 95% CI between two dichotomized groups at 10 quantiles as Q1 to Q10 (defined as 10 percentile to 100 percentile by 10 percentile at mean survival time among all patients) based on method developed by Huang (2010 &2016). The third plot is the summary using forest plot for the survival time difference at 10 quantiles. The overall in forest plot corresponds to the transformed HR and 95% CI for overall survival (log(1/HR)). The transformed HR will be interpret as if 0 is included then it is not significant.

Univariate Cox Survival Analysis Table

Show 10 entries Search:

	Hazard Ratio	95% CI Lower Limit	95% CI Upper Limit	Z Score	P-value
1	1.2757	0.399	1.5398	-0.707	0.4796

Showing 1 to 1 of 1 entries Previous 1 Next

Univariate Survival Association analysis table is displayed for the variable of interest.

Data for Forest Plot

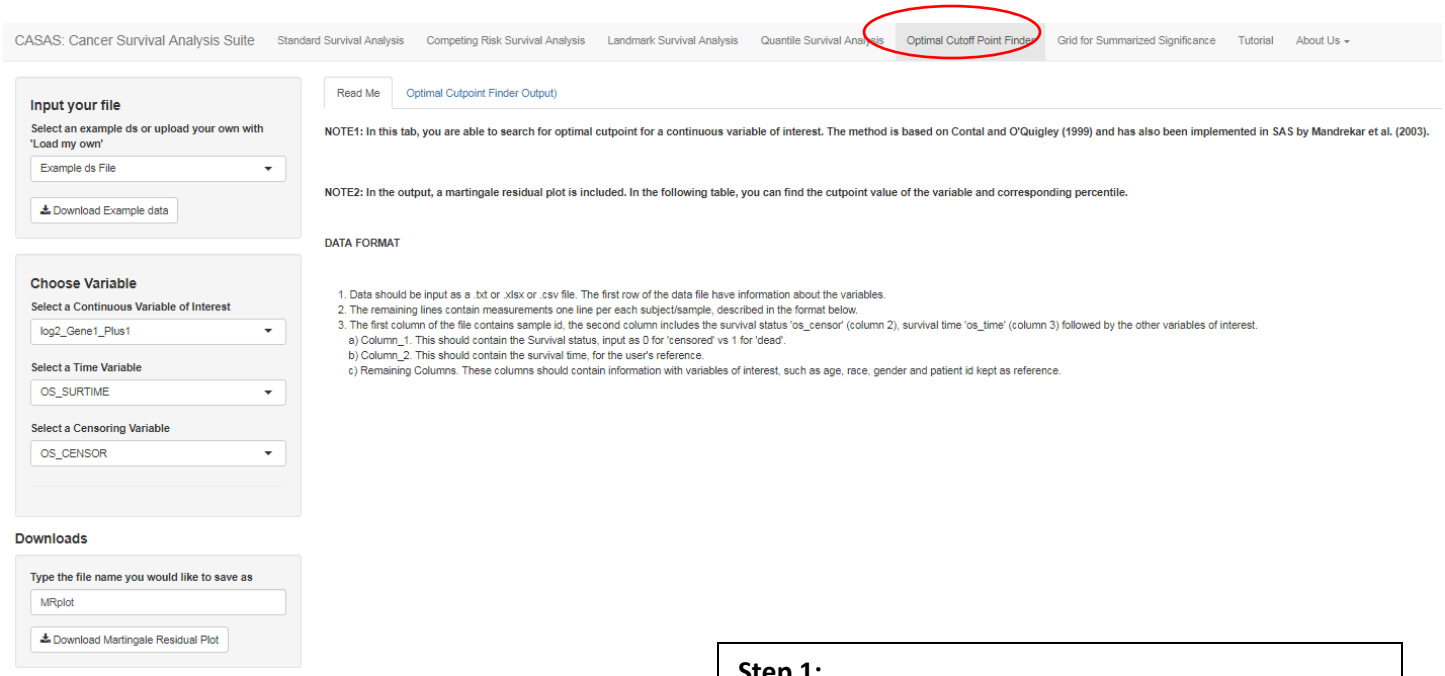
Show entries Search:

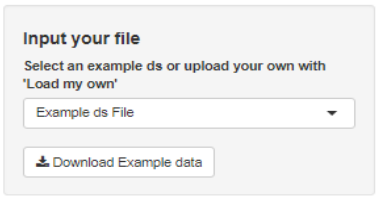
	Quantiles	Mean Time Difference/Transformed Hazard Ratio(log(1/HR))	CI Lower Limit	CI Upper Limit	Significance(0: Non-sig, 1:Not Low Better, 2:Low Better, 3:Non-estimable)
1	Q1	0.4466	-10.0049	10.8981	0
2	Q2	10.0877	-0.1158	20.2911	0
3	Q3	6.5806	1.7771	15.3645	2
4	Q4	5.8548	-6.6139	18.3235	0
5	Q5	9.7315	-5.7029	25.1659	0
6	Q6	9.7315	-6.6398	26.1028	0
7	Q7	9.7315	-6.4635	25.9466	0
8	Q8	9.7315	-6.1721	25.6351	0
9	Q9	9.7315	-5.9719	25.435	0
10	Q10	9.7315	-5.8656	25.3487	0
11	Overall	0.2435	0.9187	-0.4316	0

Showing 1 to 11 of 11 entries Previous 1 Next

Quantile wise mean survival time estimates with 95% CI limits with significance. This is useful for combining data from different cancer types or different variables (genes) in the same dataset.

TAB 5. OPTIMAL CUT-OFF FINDER

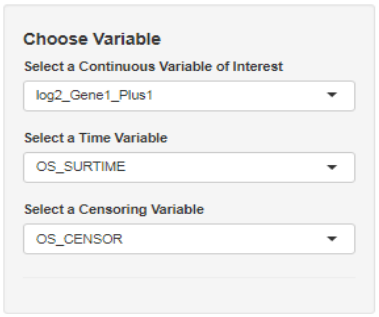




Step 1:

Select the example dataset or upload your own.

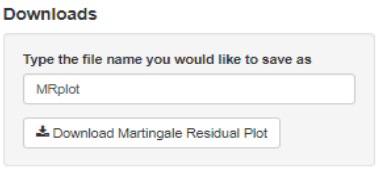
To view example data and format, use download button to view .csv file.



Step 2:

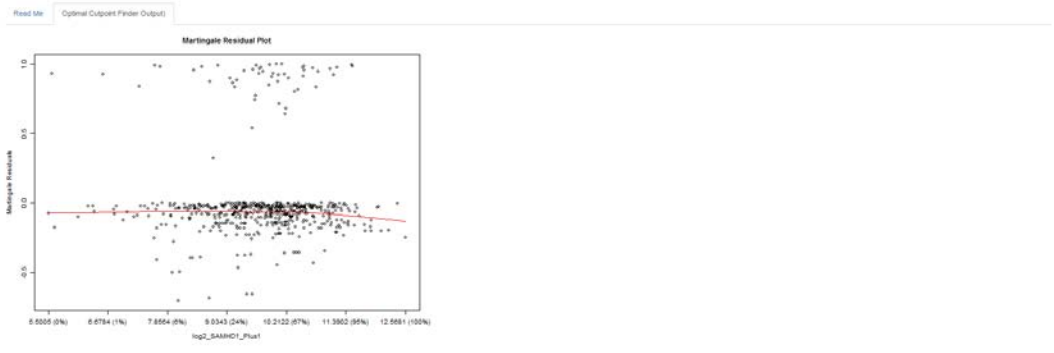
The variables from the input file will be available as drop down for selection.

Choose appropriate continuous variable (eg. gene expression) for which you wish to determine optimal cut – point based on Contal and O'Quigley (1999). Also select the appropriate time and the censor variable.



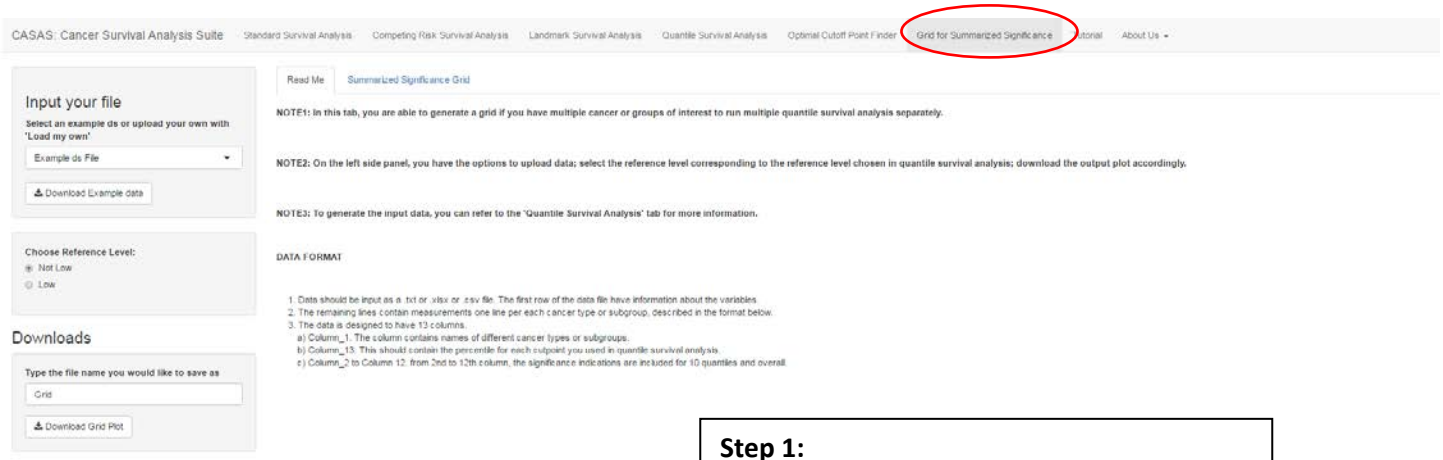
Step 3:

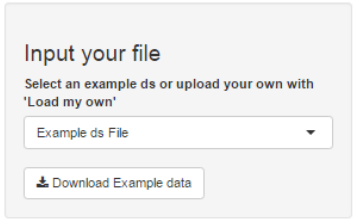
User can download the Martingale residual plot.



Martingale residual plot along with cut point value of the variable and corresponding percentile.

TAB 6. GRID FOR SUMMARIZED SIGNIFICANCE

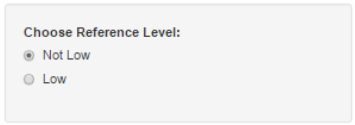




Step 1:

Select the example dataset or upload your own.

To view example data and format, use download button to view .csv file. This tab can combined results from various cancers of genes in the same cancer types.

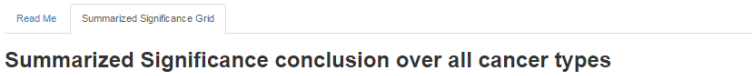


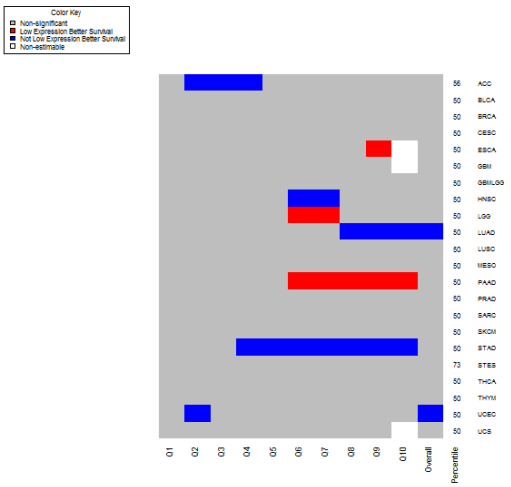
Step 2:

Based on all the combined data, select the reference level. All data should have the same reference.

Step 3:

User can download the comparison grid to access the different cancers or genes within the same cancer type.





Comparison grid for various cancer types along with the percentiles used as cut-off in each case provided in the example data.

II. TROUBLESHOOTING GUIDE

- Error: An error has occurred. Check your logs or contact the app author for clarification.

  This error is likely when:
  - User has selected a categorical variable where a continuous variable was required
  - User has selected a continuous variable where a categorical variable was required
  - The appropriate time variable in the data was not selected
  - The appropriate censor variable in the data was not selected
  - When changing/ entering new time points for CIF analysis, a ", " a space was not entered after the comma
  - In appropriate symbols are present in the data such as: ./#@!$%^&*;:,
  - Censor variable has missing values or in appropriate symbols
  - For landmark analysis, the time point is not within the range of the survival time variable.

- Images are not updating
  - For Quantile analysis, make sure to hit the "Run and generate Random numbers" button each time you choose a new variable.

- Images are not displaying
  - User chose the variable of interest that was continuous but forgot to update the dropdown under from categorical to continuous.