Paper Session #300

Title: Two in one - a SAS macro that creates tables and graphs for Survival Analysis and combines results together

Author: Jianhua (Daniel) Huang – Celgene Corporation

Key words:

 Survival Analysis, SAS

Abstract:

Survival analysis is a commonly used statistical method in oncology clinical studies, where the comparison of overall survival becomes the key efficacy measurement. Most statistician and SAS programmers need to conduct survival analysis in their daily jobs, and they are facing following challenges: 1.They must know how to conduct survival analysis in SAS language (i.e. using PROC LIFETEST, PROC PHREG); 2. They need to understand the outputs from survival analysis and pick up the right parameters for the table outputs; 3. They need to create different types of tables; 4. In addition to tables, they are often required to generate graphs. This macro is aimed to resolve all above challenges and help conducting survival analysis in a much simpler and efficient way. It could create different types of tables and meanwhile it creates graphs as well. The most appealing part is that the macro could combine the table and the graph together by inserting table summary into the graph. This two-in-one approach gives reviewer a great convenience by reviewing the key study numbers and related graph on the same page.

Outline of this article:

1. Review of Survival Analysis (event, censoring, survival function, hazard function)
2. Review of Survival Analysis Test Models(**Proc LIFETEST, Proc PHREG**)
3. Introduction of Macro Program **-%STATSURV.SAS**
    3.1. General approach – create standard survival tables.
    3.2. Enhanced approach – deal with multiple (>2) treatment groups, generate survival plots.
    3.3. Special function – add reference group, title/footnotes, treatment group format, statistics outputs.
    3.4. Ensure accuracy, improve flexibility and make it user friendly
4. Discussion - what could be improved in future
5. Conclusion

1. Review of Survival Analysis

Survival analysis is a branch of statistics deal with deaths or other defined events (i.e. disease progression) in clinical trials. Survival analysis is also called 'time-to-event' analysis as it attempts to answer following questions: what is the possibility of a population which will survive past a certain time? Of those that survive, at what rate will they die or fail?

There are 3 key factors involved in survival analysis: Event, Time, and Censoring. Event is any type of interest in survival analysis, i.e. death, disease progression, response. Time is the time from beginning (i.e. randomization date) to event or censoring. If a subject does not have an event (or a qualified event) during the observation time period, it is described as censoring.

There are two general functions used in survival analysis: Survival Function and Hazard Function. Survival function: $S(t)=Pr(T>t)$ is defined as the probability of a subject survivals longer than time t. Hazard function: $\lambda(t)dt=Pr(t<=T<t+dt|T>=t)$ is defined as the event rate at time t conditional on survival until time t or later (that is, T ≥ t).

2. Survival Analysis Test Models

Two test models are frequently used in survival analysis: LIFE TEST and PHREG TEST.

The LIFETEST generates following statistics: the Kaplan–Meier estimator, also known as the product limit estimator, estimates the survival function from life-time data; the Survival rate indicates the percentage of people in a study or treatment group who are alive for a given period of time after diagnosis; the logrank test is a hypothesis test to compare the survival distributions of two samples. In addition, the plot of the Kaplan–Meier estimate could also be generated by LIFETEST. Below is the syntax of PROC LIFETEST in SAS programming:

PROC LIFETEST < options > **;**

       TIME variable < ***censor**(list) > **;**

       BY variables **;**

       STRATA variable < (list) > < ... variable < (list) > > **;**

       SURVIVAL options **;**

       TEST variables **;**

The PHREG TEST generates following statistics (based on proportional hazard assumptions): the Hazard Ratio (HR) and its 95% confidence interval. Below is the syntax of call PROC PHREG in SAS programming:

PROC PHREG < options > **;**

       MODEL response < ***censor**(list) > = variables < /options > **;**

       STRATA variable < (list) > < ...variable < (list) > > < /option > **;**

       FREQ variable **;**

       WEIGHT variable < /option >**;**

       OUTPUT < **OUT**=SAS-data-set > ;

       BY variables **;**

3. Introduction of Macro-%STATSURV.sas (Statistical Summary for Survival Analysis)

3.1 General Approach – create standard table outputs

Template 1- Summary of Overall Survival (Unstratified Log-Rank Test)

Table 14.2.1.1
Summary of Overall Survival
Unstratified Log-Rank Test[1]
Intent-to-treat Population[2]

		Treatment (N = xxx)	Placebo (N = xxx)
Died	n(%)	XXX (XX.X)	XXX (XX.X)
Censored	n(%)	XXX (XX.X)	XXX (XX.X)
Survival Time (months)	Median (95% CI)	XX (XX.X, XX.X)	XX (XX.X, XX.X)
	Q1 (95% CI)	XX (XX.X, XX.X)	XX (XX.X, XX.X)
	Q3 (95% CI)	XX (XX.X, XX.X)	XX (XX.X, XX.X)
Hazard Ratio (Treatment:Placebo)	HR (95% CI)	X.XX (X.XXX, X.XXX)	
	p-value[3]	X.XXX	

Sample code for Overall Survival:

```
%STATSURV  (datain=adtte,  /*-input dataset*/
            trt=newtrt,    /*-treatment group*/
            aval=aval,     /*-survival time*/
            cnsr=cnsr(1));  /*-censoring flag, 1=censor*/
```

Comments: Once the input dataset, treatment group, time variable (aval) and censoring variable and its value (cnsr(1)) are defined, the macro can generate standard overall survival tables as shown in above table shell.

Template2- Summary of Overall Survival Rates

Table 14.2.1.3
Summary of Overall Survival Rates
Intent-to-treat Population[1]

	Treatment (N = xxx)		Placebo (N = xxx)	
Time Points	At Risk (n)	KM Est. (95% CI)	At Risk (n)	KM Est. (95% CI)
Month 3	XXX	XX.X% (XX.X%, XX.X%)	XXX	XX.X% (XX.X%, XX.X%)
Month 6	XXX	XX.X% (XX.X%, XX.X%)	XXX	XX.X% (XX.X%, XX.X%)
Month 9	XXX	XX.X% (XX.X%, XX.X%)	XXX	XX.X% (XX.X%, XX.X%)
Month 12	XX	XX.X% (XX.X%, XX.X%)	XX	XX.X% (XX.X%, XX.X%)
Month 18	XX	XX.X% (XX.X%, XX.X%)	XX	XX.X% (XX.X%, XX.X%)
Month 24	X	X.X% (X.X%, X.X%)	X	X.X% (X.X%, X.X%)

Sample Code for KM rates:

```
%STATSURV  (datain=adtte,
            trt=newtrt,
            aval=aval,
            cnsr=cnsr(1),
            timelist=%str(0 to 24 by 3), /*- define time points  for the KM estimators*/
            keepdata=km_param ,/*-select intermediate SAS data related to KM rates*/
            keepvar= %str(at_risk: km_estci:));/*-select variables related to KM rates*/
```

Comments: Once the time points (timelist) , the output datasets (keepdata) and output variables (keepvar) are defined. The macro can generate another type of survival table-summary of overall survival rates.

Template3- Summary of Overall Survival –Cox Proportional Hazards Model with Covariates

Table 14.2.2.3.1
Summary of Overall Survival
Cox Proportional Hazards Model with Covariates
_____

Covariate	Parameter Estimate[2]	p-value	95% Confidence Interval[2]
ECOG Performance Score	X.XXX	X.XXXX	
Age	X.XXX	X.XXXX	
Previous Response	X.XXX	X.XXXX	
Disease Stage	X.XXX	X.XXXX	
Treatment	X.XXX	X.XXXX	
Hazard Ratio (Treatment:Placebo)	X.XXX		(X.XXX, X.XXX)

Sample Code for Cox Proportional Hazards Model with Covariates

```
%STATSURV  (datain=adtte,
              trt=newtrt,
              aval=aval,
              cnsr=cnsr(1),
              covar=ecoggrn age response stage, /*-define co-variables for Cox PHREG*/
              keepdata=hr ,
              keepvar= %str(_covariate_ hr:);
```

Comments: If the co-variables (covar) are defined, the macro can conduct PHREG test and generate hazard ratio statistical summary for the treatment group and each covariate.

Template4- Summary of Overall Survival –by Subgroup

Table 14.2.2.16 (AD HOC)
Summary of Overall Survival
Log-Rank Test for Subgroups[1]

_____

		Treatment (N = xxx)	Placebo (N = xxx)
Subgroups			

_____

Sensitive	Died n(%)	XX (XX.X)	XX (XX.X)
	Censored n(%)	XX (XX.X)	XX (XX.X)
	Median (months)	XX (XX.X, XX.X)	XX (XX.X, XX.X)
	HR (95% CI)	X.XXX (X.XXX, X.XXX)	
	p-value (LR)	X.XXXX	
Refractory	Died n(%)	XX (XX.X)	XX (XX.X)
	Censored n(%)	XX (XX.X)	XX (XX.X)
	Median (months)	XX (XX.X, XX.X)	XX (XX.X, XX.X)
	HR (95% CI)	X.XXX (X.XXX, X.XXX)	
	p-value (LR)	X.XXXX	

Sample code for Subgroup Analysis.

```
%STATSURV  (datain=adtte,
           trt=newtrt,
           aval=aval,
           cnsr=cnsr(1),
           bygrp=response, /*-define by group variable */
           bygrpfmt=responsef.); /*-define format of by group */
```

Comments: The macro can conduct survival analysis by subgroups once the subgroup variable (bygrp) is defined.

3.2 Enhanced Approach

1. Conduct Survival analysis for multiple (>2) treatment groups.

RESULT - all summary combined

Obs	page	skip	row	variable	stats	c0	c1	c2	c88
1	1	1	1	Total		XXX	XXX	XXX	
2	1	1	2	Event	n(%)	XX(XX.X%)	XX(XX.X%)	XX(XX.X%)	
3	1	1	3	Censor	n(%)	XX(XX.X%)	XX(XX.X%)	XX(XX.X%)	
4	1	2	1	Median(month)	Median	XX.XX	XX.XX	XX.XX	
5	1	2	2	Q1	Q1	XX.XX	XX.XX	XX.XX	
6	1	2	3	Q3	Q3	NA	NA	NA	
7	1	3	1	HR(Trt1:Trt0)	HR (95% CI)				X.XX (X.XX, X.XX)
8	1	3	2	pvalue	pvalue(logrank)				X.XXXX
9	1	3	3	HR(Trt2:Trt0)	HR (95% CI)				X.XX (X.XX, X.XX)
10	1	3	4	pvalue	pvalue(logrank)				X.XXXX

Sample Code of Overall Survival (with > 2 treatment groups)

```
%STATSURV  (datain=adtte,
             trt=newtrt, /*-here treatment has 3 groups: c0, c1, c2*/
             aval=aval,
             cnsr=cnsr(1));
```

Comments: The macro has an enhanced capability of dealing with >2 treatment groups. Above sample demonstrates an overall survival summary with 3 treatment groups and correspondingly 2 set of hazard ratio and p-values.

Another enhanced capability is to generate both survival table and graphs and put them together. This gives a great convenience for reviewer to review the graph linked with key efficacy numbers. See below example.

2. 2 in 1- generate both survival table and plot, and insert table output into the plot.

XXXXXX Corporation
Protocol: XX-XXXX-XX-XXX

Figure 6.3
Summary of Overall Survival Analysis (ITT Population)
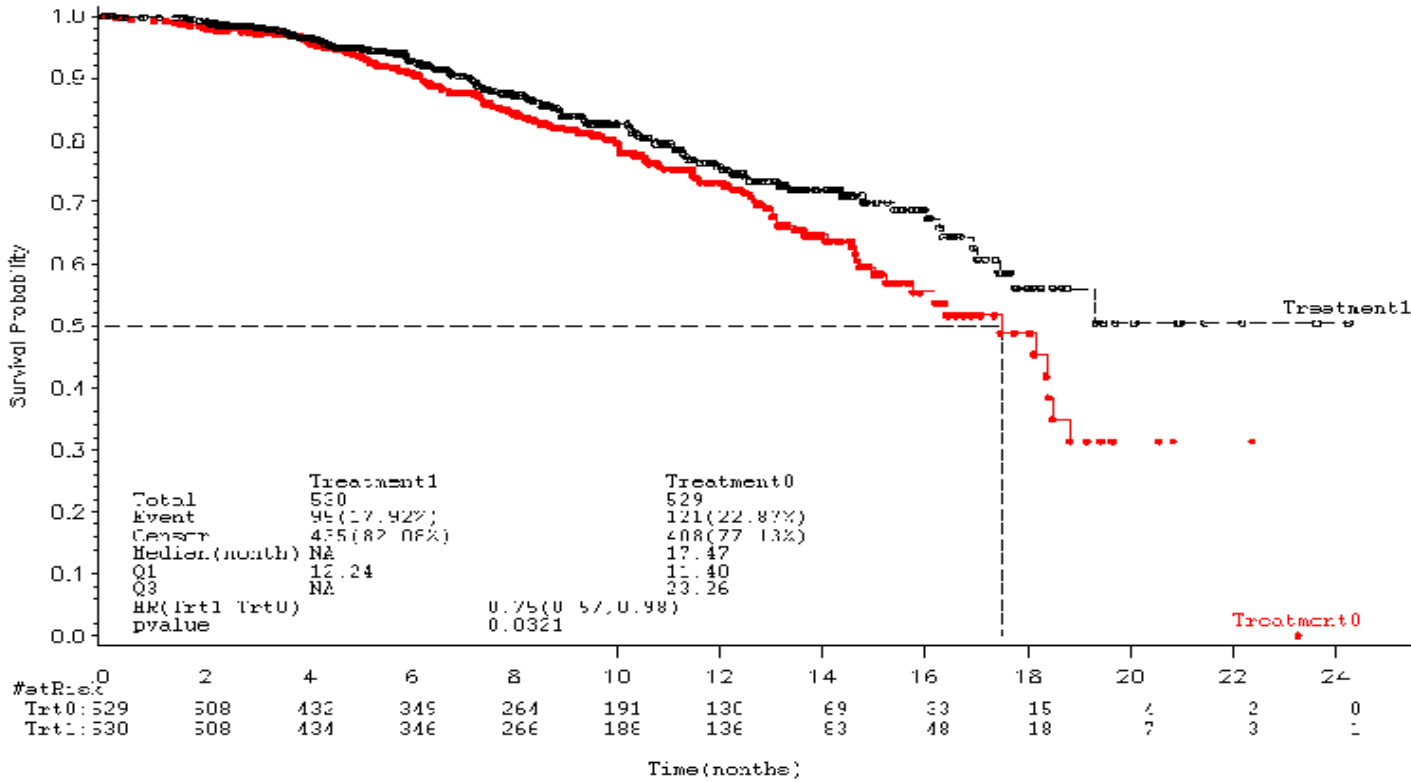
Path:S:\DEV\projects\CC-xxxx\CC-xxxx-xx-xxx\Programs\jhuang\Testing\test_statsurv.sas, Executed: 25MAR12/21:32

Sample Code: %STATSURV (datain=adtte,
                          trt=newtrt,
                          aval=aval,
                          cnsr=cnsr(1)
                          Plot=1, /*-to create survival plot */
                          graphout=%str(S:\DEV\projects\xxxx\output\graphs\testing\));
                          /*-define output folder for graph*/

3. 2 in 1- generate both survival table and plot, and insert table output into the plot - by Subgroup.

XXXXXX Corporation
Protocol: XX-XXXX-XX-XXX

Figure 6.3
Summary of Overall Survival Analysis (ITT Population)

Note: Region includes US/Canada and EU/Australia
Path:S:\DEV\projects\CC-xxxx\CC-xxxx-xx-xxx\Programs\jihuang\Testing\test_statsurv.sas, Executed: 14MAY12/2
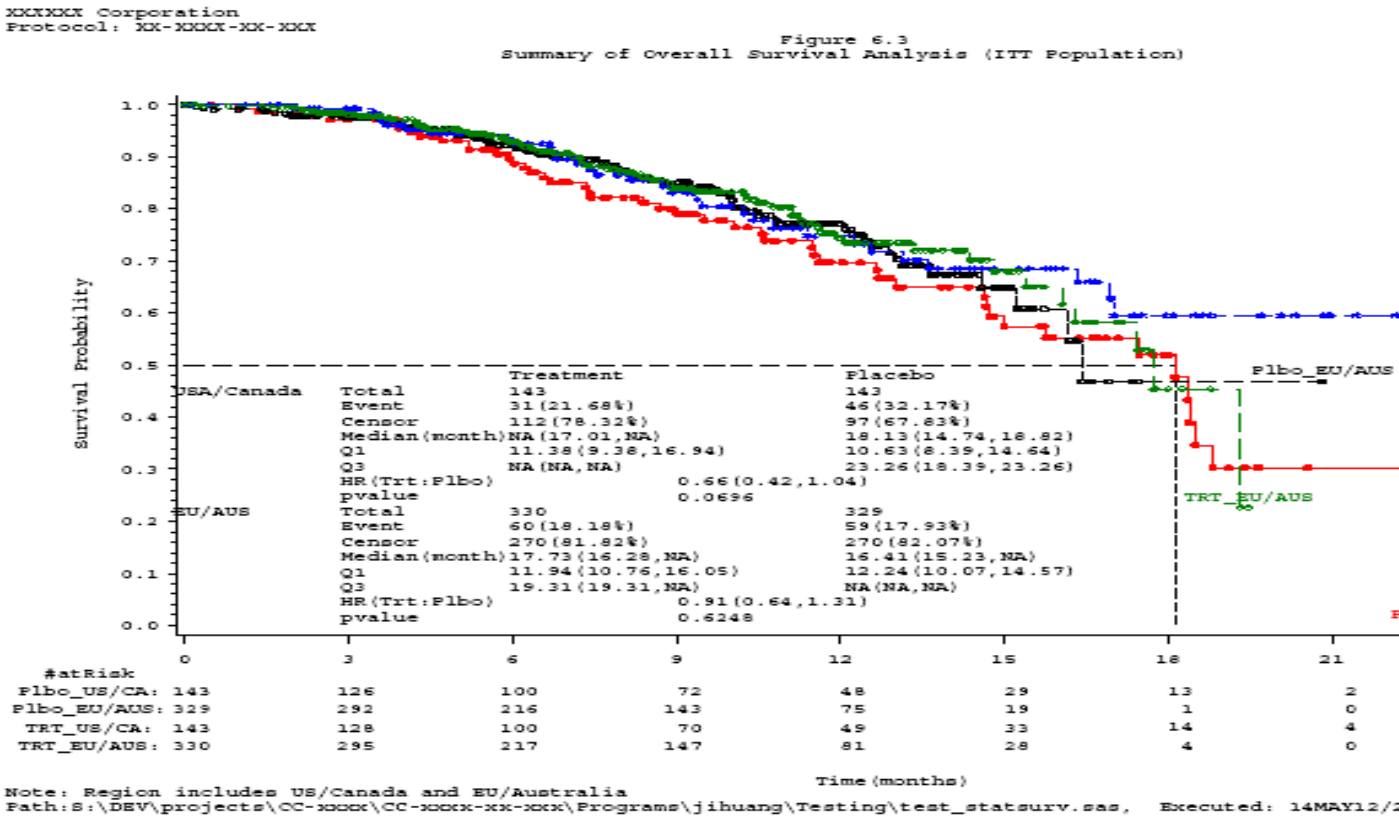
Sample Code: %STATSURV (datain=adtte,
                    trt=newtrt,
                    aval=aval,
                    cnsr=cnsr(1)
                    bygrp=regionn,
                    bygrpfmt=regionf.,
                    Plot=1, /*-to create survival plot */
                    graphout=%str(S:\DEV\projects\xxxx\output\graphs\testing\));

3.3 Special Function

In addition to above standard and enhanced approaches, the macro is also built with some special functions such as: defining reference group, specifying treatment group formats, adding title and footnote and making customized K-M plot.

1. Define reference group

    Sample Code :

        %STATSURV (datain=adtte,
                        Refgrp=2, /*-define group=2 as the reference group*/
                        trt=newtrt,
                        aval=aval,
                        cnsr=cnsr(1));

 (* By default, the treatment group with the lowest value will be set as reference group if 'Refgrp=' is missing)

    Output:

               List of Treatment Group Parameters

Obs	Group Type	Short Full Name	Numeric Name	Column Code	Compare Name	Compare Group
1	Reference Group	Treatment2	Trt2	2	c2	
2	Compare Group	Treatment0	Trt0	0	c0	Trt0:Trt2
3	Compare Group	Treatment1	Trt1	1	c1	Trt1:Trt2

Comments: The macro can define any treatment group as reference group by specifying value in the macro variable (refgrp). By default, the treatment group with lowest value will be set as reference group if (refgrp) is leaving as missing. When the reference group is defined, the macro also lists which groups are compared to reference group (see the last column of above listing output).

2. Specify formats of treatment group- i.e. the full name, short name of treatment groups.

Sample Code :

```
proc format;
     value trtf
         2 = 'Drug2  XXXX'
         1 = 'Drug1  XXXX'
         0 = 'Placebo';
     value trtcdf
         2 = 'Drug2'
         1 = 'Drug1'
         0 = 'Plbo';
   run;
   %STATSURV  (datain=adtte,
                Refgrp=0,
                trt=newtrt,
                trtfmt=trtf.,
                trtfcdmt=trtcdf.,
                aval=aval,
                cnsr=cnsr(1));
```

(*If the format of treatment groups are not created, by default, the full name will be set as: *Treatment0 , Treatment1*… and the short name will be set as: *Trt0, Trt1…*)

Output:

List of Treatment Group Parameters

Obs	Group Type	Full Name	Short Name	Numeric Code	Column Name	Compare Group
1	Reference Group	Placebo	Plbo	0	c0	
2	Compare Group	Drug1 xxxx	Drug1	1	c1	Drug1:Plbo
3	Compare Group	Drug2 xxxx	Drug2	2	c2	Drug2:Plbo

3. Add title and footnotes for the plots
   Sample Code:

```
proc format CNTLOUT=title_footnote;
        value title
            1="XXXXXX Corporation"
            2="Protocol: XX-XXXX-XX-XXX"
            3="Figure 6.3"
            4="Summary of Overall Survival Analysis (ITT Population)";
        value footnote
            1="Path:S:\DEV\projects\CC-xxxx\CC-xxxx-xx-
                xxx\Programs\jihuang\Testing\test_statsurv.sas,  Executed: &sysdate/&systime" ;
run;

%STATSURV (datain=adtte,
                trt=newtrt,
                aval=aval,
                cnsr=cnsr(1)
                Plot=1
                graphout=%str(S:\DEV\projects\xxxx\output\graphs\testing\)
                TLGfmt=title_footnote); /*-specify data name which contains plot title and
                                            footnote*/
```
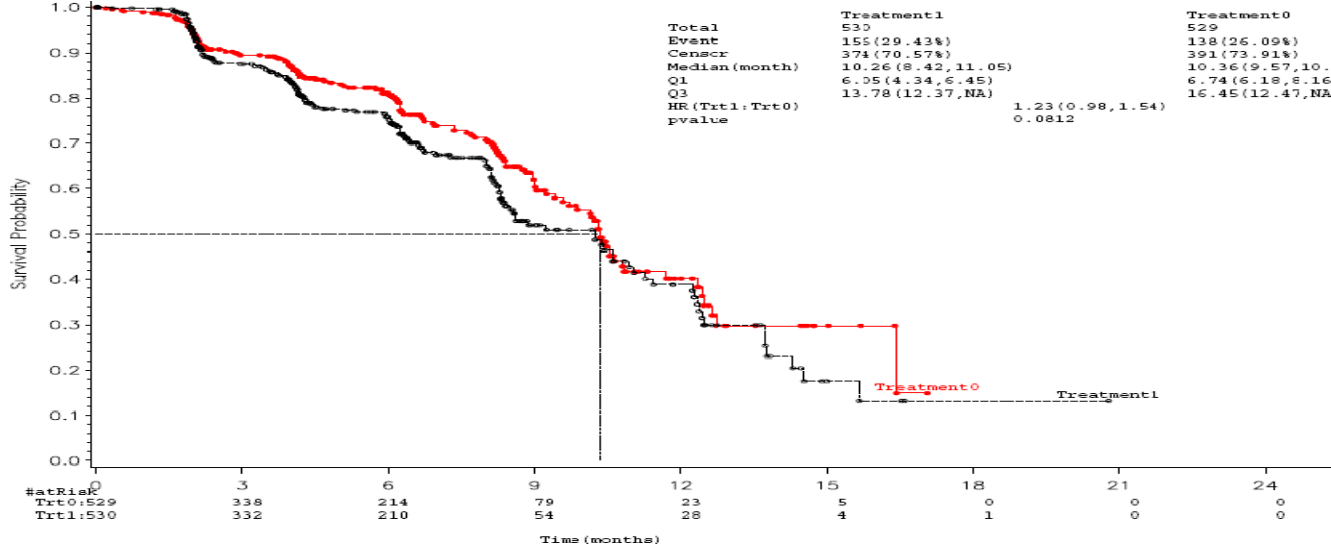
Comments: When creating K-M plot, the title and footnote of plot can be specified in a format and then saved into a temporary SAS dataset (i.e. *title_footnote*). Once the macro variable (TLGfmt) is specified, the macro can pull the title and footnote into the plot.

4. Customized the K-M plot

XXXXXX Corporation
Protocol: XX—XXXX—XX—XXX

Figure 6.3
Summary of Overall Survival Analysis (ITT Population)

	Treatment1	Treatment0
Total	530	529
Event	155(29.43%)	138(26.09%)
Censor	374(70.57%)	391(73.91%)
Median(month)	10.26(8.42,11.05)	10.36(9.57,10.
Q1	6.05(4.34,6.45)	6.74(6.18,6.16
Q3	13.78(12.37,NA)	16.45(12.47,NA
HR(Trt1:Trt0)	1.23(0.98,1.54)	
pvalue	0.0812	

Path:S:\DEV\projects\CC—xxxx\CC—xxxx—xx—xxx\Programs\jihuang\Testing\test_statsurv.sas, Executed: 27MAR12/13:46

```
%STATSURV (datain=adtte,
          trt=newtrt,
          aval=aval,
          cnsr=cnsr(1)
          Plot=1
          graphout=%str(S:\DEV\projects\xxxx\output\graphs\testing\)
          KMtable=right); /*-inset KM table into upper right corner of the plot */
```

Comments: Depends what shape of K-M plot looks like, the built-in table can be inserted either in lower-left or upper-right corner, so that the table output will not overwrite the original K-M curves.

3.4 Accuracy, Flexibility, User-friendly

Beside its robust functions, the macro is also built with efforts of ensure its accuracy, increase its flexibility and make it user-friendly.

Ensure accuracy:

1. Output PROC LIFETEST and PROC PHREG decode models for programmer and statistician review.

```
/*******************************************************************************/
/*-Below PROC LIFETEST model is used for mutiple-comarisons, please check if it is correct!!*/
/*******************************************************************************/
PROC LIFETEST data=adtte method=km alpha=0.05 outsurv=survival timelist=(0 to 26 by 2) reduceout;
 time aval*cnsr(1);
 strata newtrt/TEST=(logrank) diff=control('0') adjust=dunnett;
 ods output quartiles=quart censoredsummary=cnsr survdiff=km_pval productlimitestimates=km_est;
run;


/*******************************************************************************/
/*- Below PROC PHREG model is used for mutiple-comarisons, please check if it is correct!!! */
/*******************************************************************************/
PROC PHREG data=adtte;
 class newtrt(ref='0');
 model aval*cnsr= newtrt/rl alpha=0.05 ties=efron;
 ods output ParameterEstimates=hr_est;
run;
```

Improve flexibility:

The macro produced different types of statistical summaries and saved each of them into intermediate SAS dataset. So the users can just pickup whichever dataset and variable they need and keep them in the final result dataset.

CENSOR - event and censoring

Obs	variable	stats	c0	c1	c2
1	Total		XXX	XXX	XXX
2	Event	n(%)	XX(XX.X%)	XX(XX.X%)	XX(XX.X%)
3	Censor	n(%)	XX(XX.X%)	XX(XX.X%)	XX(XX.X%)

QUARTILE - Survival Time

Obs	variable	stats	c0	c1	c2	level
1	Median(month)	Median	XXX	XXX	XXX	1
2	Q1	Q1	XXX	XXX	XXX	2
3	Q3	Q3	XXX	XXX	XXX	3

4. Discussion-what could be improved in future?

In addition of generating K-M plot, the macro could extend its function to generate forest tree plot as below example.
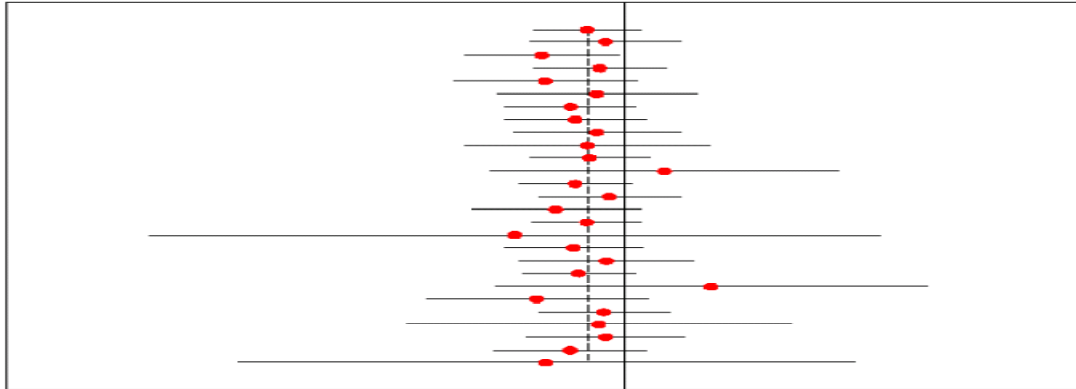
Also, in the K-M plot, add an option of inserting median survival time and 95% CI (instead of inserting the whole table) in case if there is no enough space there for the whole table output.

Finally, the current macro can generate K-M table for subgroups, but it can not create K-M plots for subgroups. This function could be added in future as another enhanced approach.

5. Conclusion

In conclusion, the macro has strong function of conducting survival analysis. It generates both standard tables and enhanced output which combines both graph and table together. Many special functions have been added in the macro to increase its capabilities. In addition, the macro makes efforts to ensure its accuracy, increase its feasibility and make it more user-friendly. Therefore, this new macro would be a useful tool for SAS programmer and statistician to conduct survival analysis.

Contact Information:

Your comments and questions are valued and encouraged. Contact author at:

Daniel Huang, Principle SAS Programmer

Celgene Corporation

110 Allen Road, Basking Ridge, NJ 07920

Tel: 908-860-4347

Email: jihuang@celgene.com