Paper 1342-2017

A SAS® Macro to Generate Kaplan-Meier Plot and Optional Estimates

Chia-Ling (Ally) Wu, University of Southern California

ABSTRACT

This paper aims to provide a SAS® macro which hopes to help users to have a Kaplan-Meier plot with optional estimates, which includes the number of subjects at risk, the number of events and total subjects, median survival rate and Cox proportional hazard ratio. This macro supports customized header, labels for time interval (X-axis) and survival rate (Y-axis), and tick marks for time interval (X-axis) and subject at risk table, as well as adding filter to the data. The output destination can be PDF, RTF or html files.

INTRODUCTION

We have a lot of chances to use time-to-event (survival) analysis, especially in biomedical and pharmaceutical field. In time-to-event data, investigators follow subjects until they reach a prespecified endpoint (for example, death). However, subjects sometimes withdraw from a study, or the study is completed before the endpoint is reached. In these cases, the survival times are censored. The uncensored survival times are referred to as event times.

In SAS®, the PROC LIFETEST procedure uses to delineate Kaplan-Meier survival plot and computes the survivor function from lifetime data, compare the survivor function by log-rank test and Wilcoxon test. The PROC PHREG procedure is used in Cox proportional regression model to estimate the effect of predictors in hazard rates. With SAS ODS SELECT tables produced by PROC LIFETEST and PROC PHREG can create output dataset which contains statistical information for subsequent use.

All the examples in this macro use the Framingham Heart Study dataset. Framingham Heart Study is a cohort study which began in 1948 by recruiting men and women between the ages of 30 and 62 from the town of Framingham, Massachusetts, who had not yet developed overt symptoms of cardiovascular disease or suffered a heart attack or stroke.

OVERVIEW OF MACRO

This SAS® macro, *mKMplot*, in this paper aims to create a customized Kaplan-Meier plot with a table for subject at risk and optional estimates which is including survival rate with 95% CI, Cox proportional hazard ratio with 95% CI and corresponding p-value. This macro also wants to provide a customized output, such as header of figure, label for time interval (X-axis) and survival rate (Y-axis), and tick marks at desired time points for time interval (X-axis) and subject at risk table, as well as adding filter to the data. The output destination can be PDF, RTF or html files from in SAS ODS.

Figure 1 is an example output with all estimates, it can be considered as four parts: Kaplan-Meier plot, the table for number of subjects at risk, the event numbers, the Cox proportional hazard ratio. These macro parameters will be described in Table 1. The number captions are the optional parameters in the macro, it will be blank or default value if user does not provide values.

1. Kaplan-Meier Plot: User can define labels of X-axis and Y-axis, and tick marks at desired time point. The Kaplan-Meier plot is delineated by PROC LIFETEST.
2. Subjects at Risk: User can define label of table header and subject numbers calculated at desired time point. The table for number of subjects at risk is provided by option ATRISK from PROC LIFETEST.
3. Subject/ Event Numbers: The legends for strata, events number and total subject number are essential outputs, whereas median survival rate with 95% CI are optional outputs.
4. The Hazard Ratio: The Cox proportional hazard ratio estimates along with 95% CI and p-value are optional outputs. User can output hazard ratio (with 95% CI) only, hazard ratio with p-value, or neither of these estimates. These Cox proportional regression estimates are provided by PROC PHREG in EXACT method. The hazard ratio is based on the reference group specified in the macro, or the default

value is the last group if not specified.

④ **Figure 1: Kaplan-Meier Plot for Hypertension**

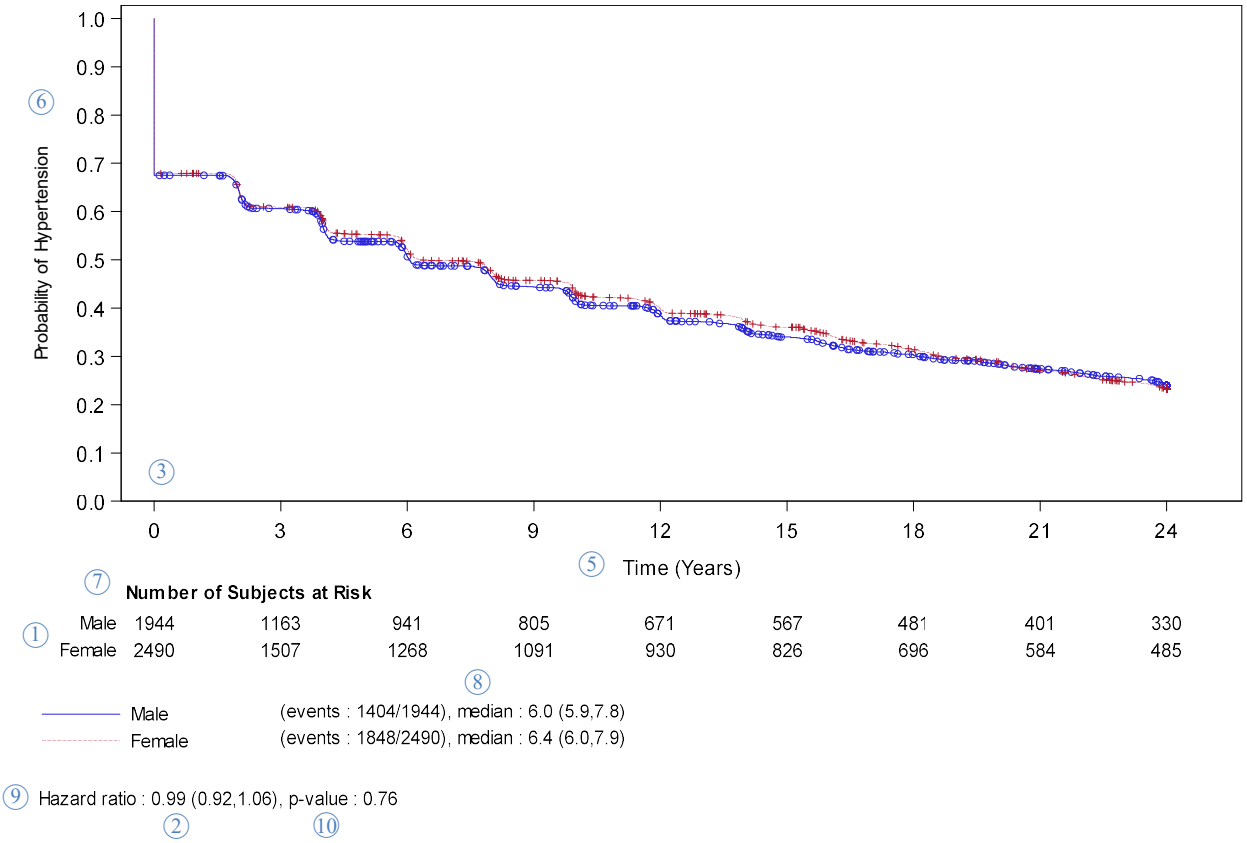

Figure 1 is hypertension survival rate among men and women. Output with customized output and optional estimates. It uses censor variable HYPERTEN, the time variable TIMEHYP, class (categorical) variable SEX and male (1 in &_REF macro variable) as reference group, presents subjects number for every three years. Optional estimates of median survival rate, hazard ratio and p-value are presented. Figure 1 does not use macro variable &_STRATALIS which is applied in STRATA statement in PROC PHREG. If we add, for example DEATH, the hazard ratio will be stratified calculated by death status.

```
%mKMplot (_indata= sas.EXAMPLE,
          _timevar= TIMEHYP,
          _censorvar= HYPERTEN,
          _stratavar= SEX_N,
          _tinterval= 3,
          _ref= "1",
          _stratalis= ,
          _datafl= ^missing(RANDID),
          _header= "Figure 1: Kaplan-Meier Plot for Hypertension",
          _Xlabel= "Time (Years)",
          _Ylabel= "Probability of Hypertension",
          _ARtitle="Number of Subjects at Risk",
          _fnmed= Y,
          _fnhr= Y,
          _fnp= Y);
```

MACRO VARIALES

Table 1. Introduction of the parameters and optional estimates.

	Macro Variable	Required	Description of Function
	_INDATA	Y	Input dataset assumed to be one record per subject.
	_TIMEVAR	Y	Time at risk variable.
	_CENSORVAR	Y	Censoring variable, should include the value to identify a censored observation. It identifies 0 as censored data and 1 as the event of interest.
①	_STRATAVAR	Y	Analysis stratification for KM analysis. For example, gender in this case, requires numeric format.
②	_REF		Study reference group. For example, “Male” in this case. Default is the last in the dataset.
②	_STRATALIS		Analysis stratification for Cox regression. For example, death in this case.
	_DATAFL		Select partial data to do survival analysis. Use in the WHERE statement in the LIFETEST and PHREG procedure, should be used in SAS logic, for example, SEX= ”Male”.
③	_TINTERVAL	Y	Time interval for displaying tick marks at desired time point for X-axis and subject at risk table.
④	_HEADER		The header of the figure.
⑤	_XLABEL		The label of X-axis (Time interval).
⑥	_YLABEL		The label of Y-axis (Survival Rate).
⑦	_ARTITLE		The label of Number of subjects at risk.
⑧	_FNMED		The footnote shows survival median or not. Default value is N.
⑨	_FNHR		The footnote shows hazard ratio or not. Default value is N.
⑩	_FNP		The footnote shows p-value for hazard ratio or not. Only functions when hazard ratio presents. Default value is N.

The dataset requires macro variable &_STRATVAR in a numeric variable, and requires a variable contains corresponding characteristics for that and names as suffix “_C”. In this case, we used variable SEX_N for strata, which is a numeric variable and contains 1 for male and 2 for female, hence it is required to have a variable SEX_C, which is a character variable and contains male, female values that this macro will capture and create the format for SEX_N from here.

Figure 2: Kaplan-Meier Plot of CHD

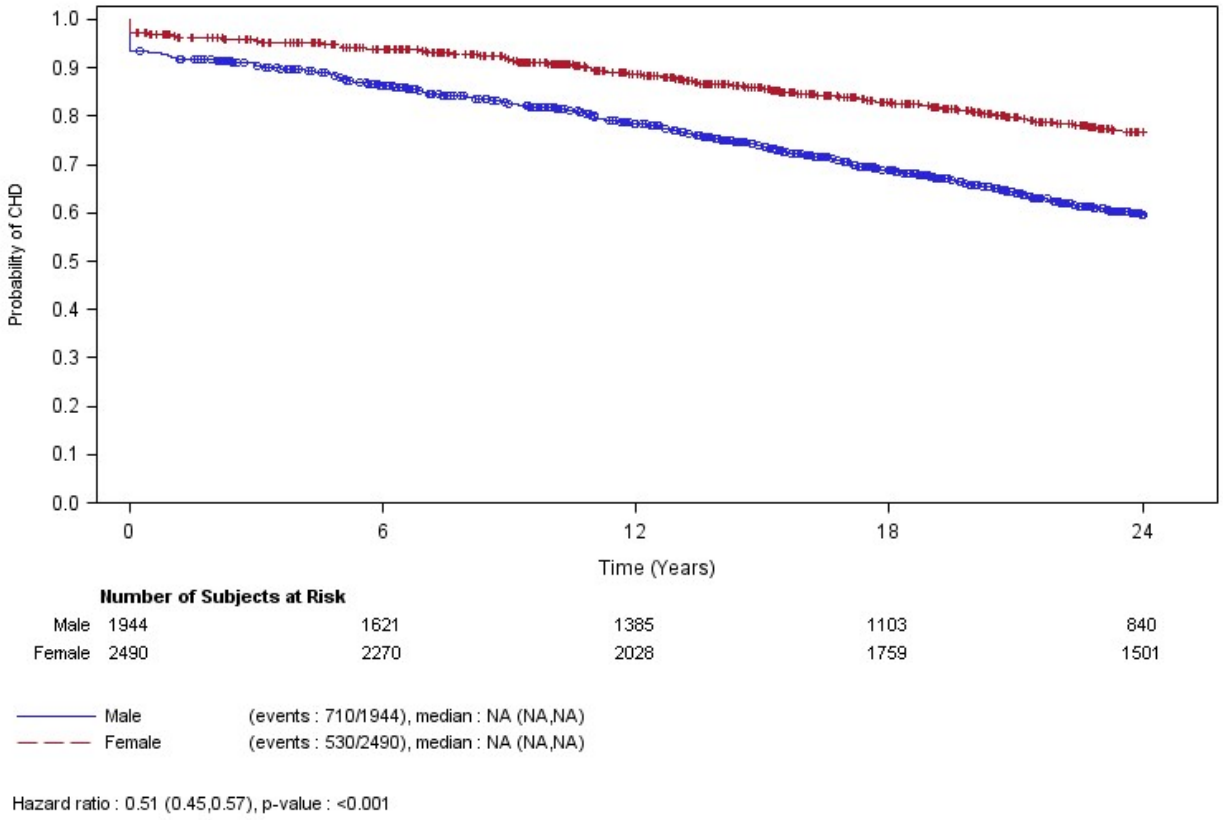

Figure 2 is an example of hypertension survival rate among men and women, with specified output and optional estimates. The time interval is presented in every six years. Since over than 50% of subject is free from cholesterol hospitalization, therefore median survival rate is presented in “NA”.

```
%mKMplot (_indata= sas.EXAMPLE,
          _timevar= TIMECHD,
          _censorvar= ANYCHD,
          _stratavar= SEX_N,
          _tinterval= 6,
          _ref= "1",
          _stratalis= ,
          _header= "Figure 2: Kaplan-Meier Plot of CHD",
          _Xlabel= "Time (Years)",
          _Ylabel= "Probability of CHD",
          _ARtitle="Number of Subjects at Risk",
          _fnmed= Y,
          _fnhr= Y,
          _fnp= Y);
```

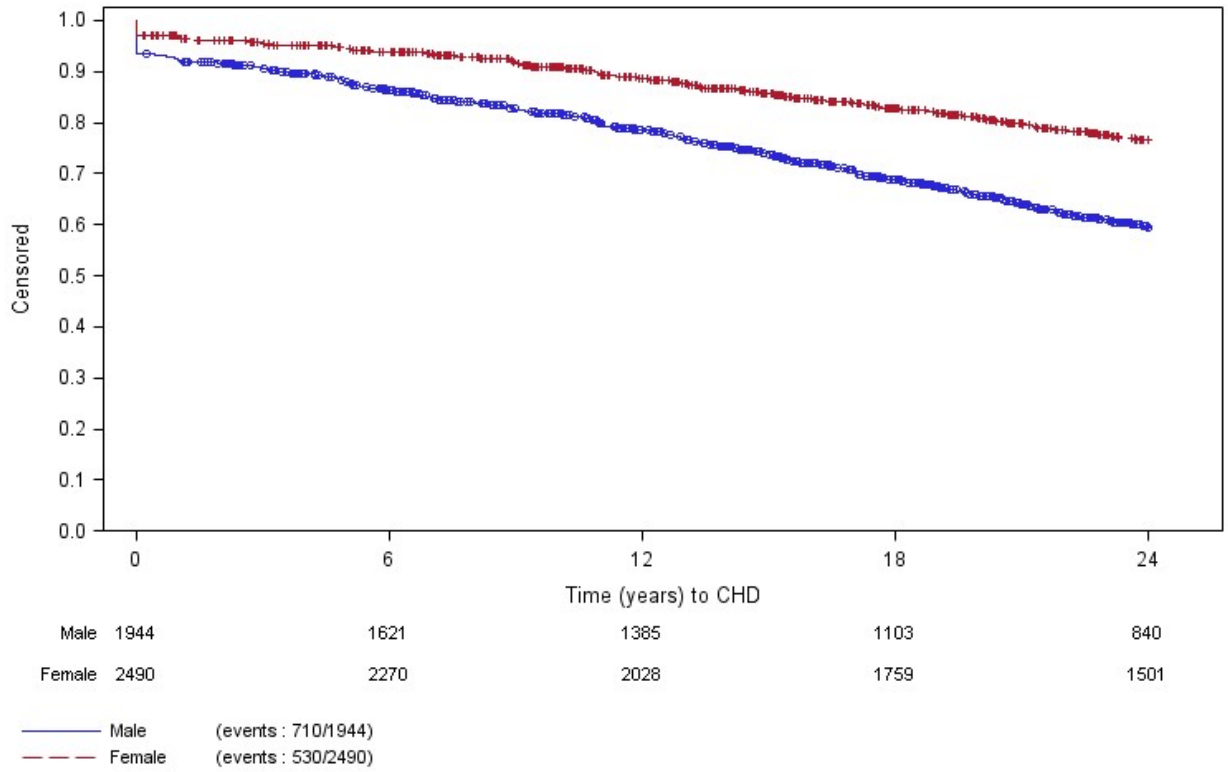

Figure 3 is cholesterol hospitalization survival rate among men and women, without specified output and optional estimates. Therefore, the label of X-axis and Y-axis are based on label of TIMEHYP and HYPETEN variables. The footnote for median survival rate, Cox proportional hazard ratio and p-value are not showed in the plot.

```
%mKMplot (_indata= sas.EXAMPLE,
          _timevar= TIMECHD,
          _censorvar= ANYCHD,
          _stratavar= SEX_N,
          _tinterval= 6);
```

SAS CODE

```
%macro mKMplot (_indata= ,     /*Required. Input dataset*/
                _timevar= ,    /*Required. Time*/
                _censorvar=,   /*Required. Censoring variable*/
                _stratavar=,   /*Required. KM strata in PROC LIFETEST*/
                _ref= ,        /*Permissible. Study reference group*/
                _stratalis= ,  /*Permissible. HR strata in PROC PHREG*/
                _datafl=,      /*Permissible. Data filter for PROC LIFETEST
                                 and PHREG (where=(&_datafl))*/
                _tinterval= ,  /*Required. Time interval- ticks and table*/
                _header= ,     /*Permissible. The header of Figure*/
                _Xlabel= ,     /*Permissible. The label of X-axis (Time)*/
                _Ylabel= ,     /*Permissible. The label of Y-axis (Survival)*/
                _ARtitle=,     /*Permissible. The label of subjects at risk*/
                _fnmed=,       /*Permissible. Whether shows survival rate*/
                _fnhr=,        /*Permissible. Whether shows hazard ratio*/
                _fnp=,         /*Permissible. Whether shows p-value for HR*/);
```

```
/*ticks for X-axis and At Risk table*/
proc sql noprint;
    select &_TINTERVAL.*ceil((max(&_TIMEVAR.)/&_TINTERVAL.)) into: _maxi
    from &_INDATA.
    ;
run;
quit;

/*label for strata*/
proc sort data= &_INDATA out = fmt1 nodupkey;
    by &_STRATAVAR.;
run;


%let _STRATAVAR_C = %sysfunc(cats(%scan("&_STRATAVAR.",1, "_"),_C));

data fmt1 (keep=fmtname type start label);
    set fmt1;
    retain fmtname "STRATA"  type "N";
    length start 8. label $20.;
    start = &_STRATAVAR.;
    label = &_STRATAVAR_C.;
run;

proc format cntlin=fmt1;
run;


/*PROC LIFETEST for KM-plot, subject at risk, median survival rate, event
numbers*/

ods graphics / reset;
ods exclude all;
ods trace on;

ods output survivalplot=zz_plotdata HomTests=zz_htest CensoredSummary=zz_esum
Quartiles=zz_qtp;
proc lifetest data=&_INDATA %if %bquote(&_DATAFL.) ne %then %do;
%str((where=(&_DATAFL.))) %end;
    method=pl
    plots=survival(atrisk(outside)=0 to &_maxi. by &_TINTERVAL.);;
    ods select SurvivalPlot Homtests CensoredSummary Quartiles;
    time &_TIMEVAR. * &_CENSORVAR.(0);
    strata &_STRATAVAR.;
run;


data zz_plotdata;
    set zz_plotdata;
    by STRATUM;
    retain CUMEVENT SURVPROB;
    if first.STRATUM then do;
        CUMEVENT=0;
        SURVPROB=1;
    end;
    else do;
```

```
        if EVENT ne . then CUMEVENT=CUMEVENT+EVENT;
        if SURVIVAL ne . then SURVPROB=SURVIVAL;
    end;
run;


data fOS;
    set zz_plotdata;
    format STRATUMNUM strata.;
run;

/*footnote event/ median*/
proc sql;
    create table fn1_0 as
    select a.&_STRATAVAR., a.STRATUM, a.TOTAL, a.FAILED, b.ESTIMATE,
           b.LOWERLIMIT, b.UPPERLIMIT
    from zz_esum as a
        left join zz_qtp (where=(PERCENT=50)) as b
        on a.&_STRATAVAR.=b.&_STRATAVAR.
    having ^missing(&_STRATAVAR.)
    order by STRATUM
    ;
quit;

%if &_REF= %then %do;
    data aa;
        set fn1_0 end=eos;
        by STRATUM;
        if eos then call symput ('_REF', '"' || STRATUM || '"');
    run;
%end;

data fn1;
    set fn1_0 end=eos;
    by STRATUM;

    /*length setting- for output space*/
    %if &_fnmed=Y %then %do; length x $120.; %end;
    %else %do; length x $150.; %end;

    if missing(TOTAL) then TOTAL=0;
    if missing(FAILED) then FAILED=0;

    array zero1(*) ESTIMATE LOWERLIMIT UPPERLIMIT;
    array zero2(*) $8. ESTIMATE_ LOWERLIMIT_ UPPERLIMIT_;
    do i=1 to dim(zero1);
        if missing(zero1(i)) then zero2(i)="NA";
        else zero2(i) = strip(put(zero1(i),5.1));
    end;

    x1 = cat("(events : ", strip(put(FAILED,best.)) , "/",
             strip(put(TOTAL,best.)), ")" );
    x2 = cat("median : ", strip(ESTIMATE_), " (", strip(LOWERLIMIT_), ",",
             strip(UPPERLIMIT_), ")" );
    %if &_fnmed=Y %then %do; x = catx(", ", x1, x2); %end;
    %else %do; x = x1; %end;
```

```
    if eos then call symputx ('_Nstrata', STRATUM);
    if &_stratavar = &_ref then call symputx ('_FAILED', FAILED);
run;
%put &_Nstrata;
%put &_FAILED;

%macro strata;

%do i=1 %to &_Nstrata.;
    %global _event_&i;
    data _null_;
        set fn1;
        if _n_=&i. then call symput("_event_&i.",x);
    run;
%end;

%mend;
%strata;
%put &&_event_1;
%put &&_event_2;


/*footnote HR and P-value*/
ods output  ParameterEstimates=zz_hazard;
proc phreg data=&_INDATA. %if %bquote(&_DATAFL.) ne %then
%str((where=(&_DATAFL.)));;
    ods select  ParameterEstimates;
    class &_STRATAVAR. (ref=&_REF.) &_STRATALIS.;
    model &_TIMEVAR.* &_CENSORVAR.(0)=&_STRATAVAR.
            /ties=EXACT risklimits alpha=0.05 ;
    strata &_STRATALIS.;
run;

%put &_stratalis;

ods exclude none;

%let _hr1=; %let _hr2=; %let _hr3=; %let _hr4=;

data fn2;
    set zz_hazard;
    array zero1(*) HAZARDRATIO HRLOWERCL HRUPPERCL PROBCHISQ;
    array zero2(*) $8. HAZARDRATIO_ HRLOWERCL_ HRUPPERCL_ PROBCHISQ_;
    do i=1 to dim(zero1);
        %if &_FAILED. ^= 0 %then %do;
            if missing(zero1(i)) then zero2(i)="NA";
            else zero2(i) = strip(put(zero1(i),5.2));
        %end;
    end;

    if ^missing(PROBCHISQ) and PROBCHISQ < 0.001 then PROBCHISQ_= "<0.001";
    else if  PROBCHISQ > 0.999 then  PROBCHISQ_= ">0.999";

    %if &_fnhr=Y %then %do;
        %let _hr1 = %str(Hazard ratio : );
        %if &_FAILED. = 0 %then %do;
            %let _hr2 = {unicode '221e'x};
```

```
            call symput ("_hr3" , " (NA,NA)");
        %end;
        %else %do;
            call symput ("_hr2" , strip(HAZARDRATIO_));
            call symput ("_hr3" , " (" || strip(HRLOWERCL_) || "," ||
                        strip(HRUPPERCL_) || ")");
        %end;
        %if &_fnp=Y %then
            call symput ("_hr4" , ", p-value : " || strip(PROBCHISQ_));;
        %end;
run;

/*for output*/
%let _OutType = rtf; /*can be rtf, pdf, html*/

%let _lev2=%sysevalf(%sysevalf(&_Nstrata)*.05);
%let _lev3=%sysevalf(&_lev2*1.2);
%let _lev4=%sysevalf(0.06);
%let _lev1=%sysevalf(1-&_lev2-&_lev3-&_lev4);

%macro temp;

proc template;
    define statgraph fOS_SURVIVAL;
    dynamic TOTAL FAILED ;
    mvar _Nstrata _ARtitle _Ylabel _Xlabel _header  _lev1 _lev2 _lev3 _lev4;

    /*output setup- 4 cells*/
    begingraph / border=FALSE designwidth=20cm designheight=14cm;
    layout lattice / rows=4 rowweights=(&_lev1 &_lev2 &_lev3 &_lev4)
                     rowgutter=0 columndatarange=union ;

    /*1st cell- KM plot*/
    cell;
    layout overlay / border=false
    yaxisopts=(offsetmin=0 linearopts=(thresholdmin=0 viewmin=0
               tickvaluesequence=(start=0 end=1 increment=.1))
               %if &_Ylabel^= %then %do;
               label=&_Ylabel. labelattrs=(size=8pt)
               %end;)
    xaxisopts=(offsetmin=0 display=(line label ticks tickvalues)
               %if &_Xlabel^= %then %do;
               label=&_Xlabel. labelattrs=(size=9pt)
               %end;
               linearopts=(thresholdmin=0
                           tickvaluesequence=(start=0 end=%eval(&_maxi)
                           increment=%eval(&_tinterval) )));
    scatterplot x=TIME y=CENSORED / group=STRATUMNUM name="scat";
    stepplot x=TIME y=SURVIVAL / group=STRATUMNUM name="step";
    endlayout;
    endcell;

    /*2nd cell- AT Risk*/
    cell;
    %if &_ARtitle^= %then %do;
    cellheader;
        entry halign=left textattrs=(size=8pt weight=bold) &_ARtitle.
```

```
            / border=FALSE;
    endcellheader;
    %end;
    layout overlay/ walldisplay=none xaxisopts=(display=none);
        blockplot x=TATRISK block=ATRISK
            / class=STRATUMNUM blockindex=STRATUMNUM
              repeatedvalues=true display=(label values)
              valuehalign=start valuefitpolicy=truncate
              valueattrs=graphdatatext(size=8pt)
              labelposition=left labelattrs=graphvaluetext
              labelattrs=(size=8pt) includemissingclass=false;
    endlayout;
    endcell;

    /*3rd cell- Event Number*/
    cell;
    layout overlay/ walldisplay=none xaxisopts=(display=none)
                    pad=(top=0 left=5 bottom=0);
        discretelegend "step"
            / valueattrs=(size=8pt) border=false displayclipped=true
              halign=left;
        layout gridded / rows=2 border=false;
        %do i=1 %to &_Nstrata.;
            entry halign=left textattrs=(size=8pt)"%nrbquote(&&&_event_&i)";
        %end;
    endlayout;
    endlayout;
    endcell;

    /*4th cell- Hazard Ratio*/
    cell;
    layout overlay/ walldisplay=none xaxisopts=(display=none);
    %if &_FAILED. ne 0 %then %do;
        entry halign=left textattrs=(size=8pt)
        " " "&_hr1" "&_hr2" "&_hr3" "&_hr4";
    %end;
    %else %if &_FAILED. eq 0 %then %do;
        entry halign=left textattrs=(size=8pt)
        " " "&_hr1" &_hr2 "&_hr3" "&_hr4";
    %end;
    endlayout;
    endcell;

    endlayout;
    endgraph;
    end;
run;
quit;

%mend temp;
%temp;

options orientation=landscape papersize=letter;
ods noproctitle escapechar="~";

ods listing style=listing;
ods graphics / reset noborder imagename="fOS" ;
```

```
ods listing gpath="&_outputPath\";
ods &_OutType file ="&_outputPath\fOS.&_OutType"
              style=listing nogtitle nogfootnote;

title1 %str(&_HEADER);
proc sgrender data=fOS template=fOS_SURVIVAL
    %if &_HEADER^= %then description=&_HEADER.;;
run;

ods &_OutType close;
ods listing ;

%mend mKMplot;
```

CONCLUSION

This paper presents a SAS® macro to generate a survival plot, with customized estimates and output setting. We hope to provide an efficient creation of Kaplan-Meier output for time-to-event analysis.

REFERENCES

Berglund, P. A. 2011. “An Overview of Survival Analysis using Complex Sample Data”, presented at SAS Global Forum, Las Vegas, NV, 2011.

Gharibvand, L. 2008. “A Step-by-Step Guide to Survival Analysis”, presented at SAS WUSS Forum, Universal City, CA, 2008.

Kaplan, E. L. and Meier, P. (1958), “Nonparametric Estimation from Incomplete Observations,” Journal of the American Statistical Association, 53, 457–481.

Kuhfeld, W. F. and So, Y. 2013. “Creating and Customizing the Kaplan-Meier Survival Plot in PROC LIFETEST”, presented at SAS Global Forum, San Francisco, CA, 2013.

UCLA Institute for Digital Research and Education. http://stats.idre.ucla.edu/sas/seminars/sas-survival/

ACKNOWLEDGMENTS

I would like to acknowledge Grace Ng, David Prince, Bill Coar and Nina Pashova at Axio Research for their valuable consultant and assistance in the development of this macro.

CONTACT INFORMATION

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

Chia-Ling (Ally) Wu
University of Southern California
Chialing.ally.wu@gmail.com