

PharmaSUG 2021 - Paper DV-114

Customize and Beautify Your Kaplan-Meier Curves in Simple Steps

Using PROC SGPLOT to Generate Kaplan-Meier Curves with Confidence Interval Brackets at Time-at-Risk

John Henderson and Lili Li

ABSTRACT

Kaplan-Meier curves are among the most frequently employed graphs to visualize and demonstrate the estimate of survival patterns in clinical trials. SAS® software offers simple tools for generating survival analysis results using PROC LIFETEST. Default graphical outputs by PROC LIFETEST are adequate for assessing simple outcomes; however, presentation of analysis results often requires various and complex customizations. Specifically, confidence interval is a very commonly requested addition to aid the interpretation of survival analysis data. Among different ways of displaying confidence intervals, confidence interval bracket is one particularly helpful feature that can provide a better representation of analysis at specific increments of interest. Yet adding confidence interval brackets usually requires advanced and complex coding as PROC LIFETEST default outputs incorporate background shading representing the pointwise confidence intervals not brackets only at specified time points of interest. This paper provides a straightforward and flexible approach by applying tools provided in the PROC SGPLOT procedure, together with implementation of annotation datasets, that allow sophisticated customizations of confidence interval brackets using survival data generated from PROC LIFETEST.

INTRODUCTION

A common endpoint investigated in clinical trials is survival or failure probability over time. These analyses allow for evaluation and often comparison of interventions impacting events of interest, such as death, major adverse events, and other endpoints (In & Lee, 2018). One of the most common statistical techniques used in evaluating survival or failure over time is a non-parametric method developed by Edward L. Kaplan and Paul Meier in 1958 (Kaplan & Meier, 1958). The Kaplan-Meier approach provides survival estimates while accounting for missing data due to censoring and is regarded as, “the simplest way of computing the survival over time,” (Goel, Khanna, & Kishore, 2010).

SAS software provides multiple ready-to-use procedures for estimating Kaplan-Meier survival analysis and creating related plots. PROC LIFETEST is the most commonly used procedure for generating survival estimates over time using the Kaplan-Meier method while also providing analytical outputs testing estimates across strata and providing basic survival curve plots. However, PROC LIFETEST is limited in the capacity of customizing plots, so additional procedures must be utilized in order to produce more customized plots. PROC GPLOT and PROC SGPLOT are two alternative tools that would provide more flexibility for plotting Kaplan-Meier curves. Based on comparison conducted by Amruthnath in 2013, PROC SGPLOT has greater functionality in producing statistical plots than does PROC GPLOT (Amruthnath, 2013). This paper will focus on introducing features in PROC SGPLOT to construct Kaplan-Meier curves with a particular customization of adding confidence interval brackets and plot annotations.

DATA PREPARATION

The BMT dataset from the default SASHELP library is used as sample data in this paper. This is a survival dataset where each observation represents one subject with three variables: GROUP, T, and STATUS. The GROUP variable has three categorical values: ALL, AML-High Risk, and AML-Low Risk. The T variable denotes the survival time of each subject in days, and the STATUS variable is the survival status (0=Alive, 1=Dead). This demonstration will subset the dataset to include only one group, AML-Low Risk. Subject identifiers are added for each observation. The following code is used as the first step of data preparation for the analysis.

```
data adkm;
  set sashelp.bmt(where=(group="AML-Low Risk"));
  SUBJID = strip(put(monotonic(), z3.));
run;
```

KAPLAN MEIER CURVE CUSTOMIZATION USING PROC SGPLOT

Following the data preparation step above, the survival analysis dataset is created using the PROC LIFETEST procedure. The sample codes below demonstrate analysis to plot Kaplan-Meier survival estimates through 2500 days. In order to output a dataset of survival estimates, the PLOTS=SURVIVAL option is specified in the procedure statement. The ATRISK option sets the number of subjects at risk at a desired interval for the plot, in this example, every 500 days. The CL option is added to produce the pointwise confidence limits. The TIME statement is where the time and event status variables are specified in the Kaplan-Meier estimation. And finally, the ODS OUTPUT statement with the SURVIVALPLOT dataset naming option will output the dataset needed to construct and customize the Kaplan-Meier curve later.

```
proc lifetest data = adkm
  plots=survival(atrisk=(0 to 2500 by 500) cl);
  time T*STATUS(0);
  ods output survivalplot=plot;
run;
```

However, the SURVIVALPLOT dataset generated from the ODS OUTPUT statement must be modified prior to implementing it with PROC SGPLOT for generation and customization of the Kaplan-Meier curves. The reason for that is because survival estimates and related confidence limits are generated only at times when subjects experience an event or are censored. These often do not fall in exact alignment with the time intervals of interest. Therefore, the last estimate produced prior to the interval of interest must be retained for proper plotting. The code shown below demonstrates this process.

```
data plotds_;
  set plot;
  by stratumnum time;
  retain surv lcl ucl;
  if first.startumnum then do;
    surv=.;
    lcl=.;
    ucl=.;
  end;
  if survival ne . then surv=survival;
  if sdf_lcl ne . then lcl=sdf_lcl;
  if sdf_ucl ne . then ucl=sdf_ucl;
run;
```

Confidence interval data is limited to the intervals of interest.

```
data plotds;
  set plotds_(where=(time le 2500));
  if tatrisk = . then do;
    lcl=.;
    ucl=.;
  end;
run;
```

The syntax of PROC SGPLOT below aligns plotting options into a block format. The XAXIS and YAXIS options shown below are used to customize the X and Y axes: The VALUES and VALUEATTRS options set up parameters for the axes range, intervals, and appearance of the tick values shown on the axes; The LABEL and LABELATTRS options provide customizable labels and label appearance.

The STEP block is where Kaplan-Meier curves are plotted. The LINEATTRS option allows for adjustment of the plot line style, thickness, and color. The YERRORLOWER and YERRORUPPER options are for specifying the confidence intervals brackets. The ERRORBARATTRS option customizes the error bracket appearance. The NAME option provides an indicator for the name of the plot for use in the KEYLEGEND block, and the LEGENDLABEL option determines the label shown in the plot legend for the step plot.

In order to add the At-Risk values to the bottom of the plot, the XAXISTABLE block is used. This block specifies the variables used for the plotting interval and values plotted. Appearance options are provided for the XAXISTABLE similarly to the options used in the XAXIS and YAXIS blocks.

The KEYLEGEND block identifies the plots to be included in the plot legend, position and location of the legend as well as appearance options.

```
proc sgplot data=plotds;
  xaxis
    values=(0 to 2500 by 500)
    valueattrs=(color=black size=10pt weight=bold family=arial)
    label="Time After Index Procedure (Days)"
    labelattrs=(color=black size=12pt weight=bold family=arial)
  ;

  yaxis
    values=(0 to 1 by 0.2)
    valueattrs=(color=black size=10pt weight=bold family=arial)
    label="Survival Probability (%)"
    labelattrs=(color=black size=12pt weight=bold family=arial)
  ;

  step x=time y=surv /
    lineattrs=(pattern=solid thickness=2 color=blue)
    yerrorlower=lcl
    yerrorupper=ucl
    errorbarattrs=(pattern=solid thickness=2 color=blue)
    name='step'
    legendlabel="AML-Low Risk"
  ;

  xaxistable atrisk /
    x=tatrisk
    location=outside
    pad=(top=10)
    title="No. At Risk"
    titleattrs=(color=black size=10pt weight=bold)
    label="AML-Low Risk"
    labelpos=left
  ;
```

```

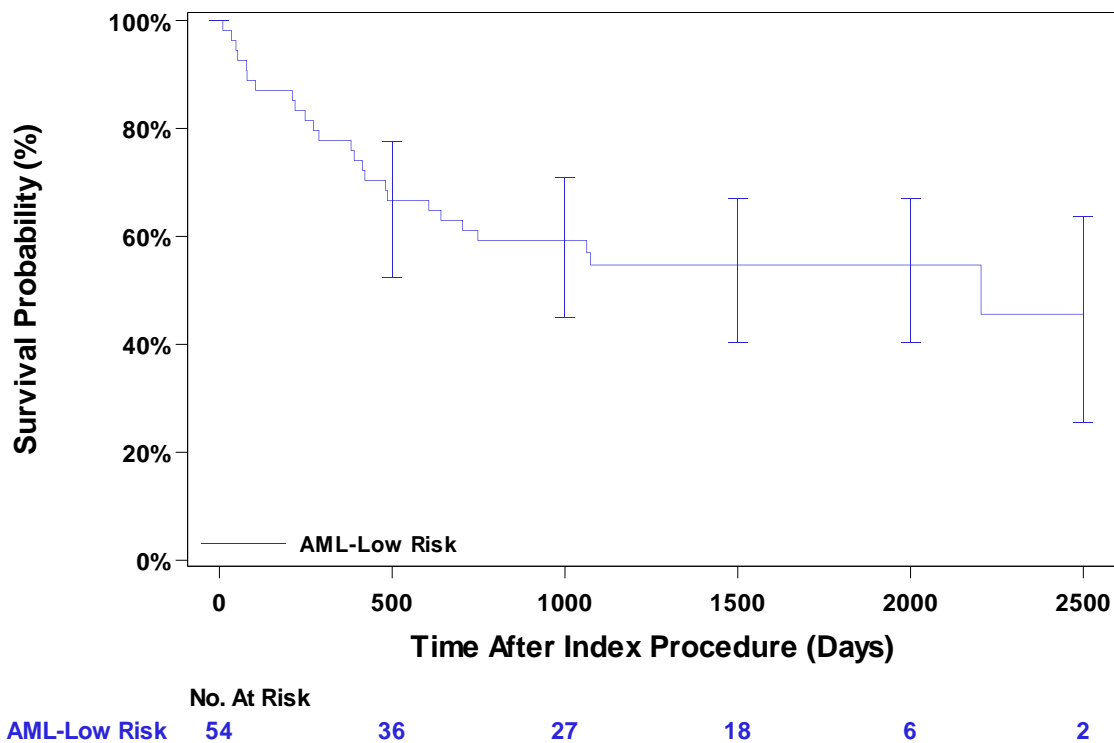
labelattrs=(color=blue size=10pt weight=bold family=arial)
valueattrs=(color=blue size=10pt weight=bold family=arial)
;

keylegend 'step' /
location=inside
position=bottomleft
noborder
down=2
valueattrs=(color=black size=10pt weight=bold family=arial)
;

format surv percent11.1;
run;

```

Using PROC SGPLOT syntax above, the following Kaplan-Meier survival curve is produced.



As frequently requested in actual clinical trial analysis, the annotation of actual confidence interval values on the plot is an additional customization that we will discuss and demonstrate next. This additional feature on a Kaplan-Meier curve is best facilitated using an annotate dataset. By utilizing the SGANNO option in the PROC SGPLOT statement, it allows with great flexibility for text and other items to be positioned on the plot. The following codes create an sample annotation dataset for adding a text caption and text confidence interval values:

```

data caption;
length ID FUNCTION LABEL ANCHOR JUSTIFY TEXTWEIGHT X1SPACE Y1SPACE $100
width X1 Y1 8;
retain ID "EST" FUNCTION "TEXT" ANCHOR "TOPRIGHT" JUSTIFY "RIGHT"
TEXTWEIGHT "BOLD" X1SPACE "DATAVALUE" Y1SPACE "DATAVALUE" WIDTH 60;
label = "*95% Confidence Interval";

```

```

X1=2500;
Y1=0.08;
run;

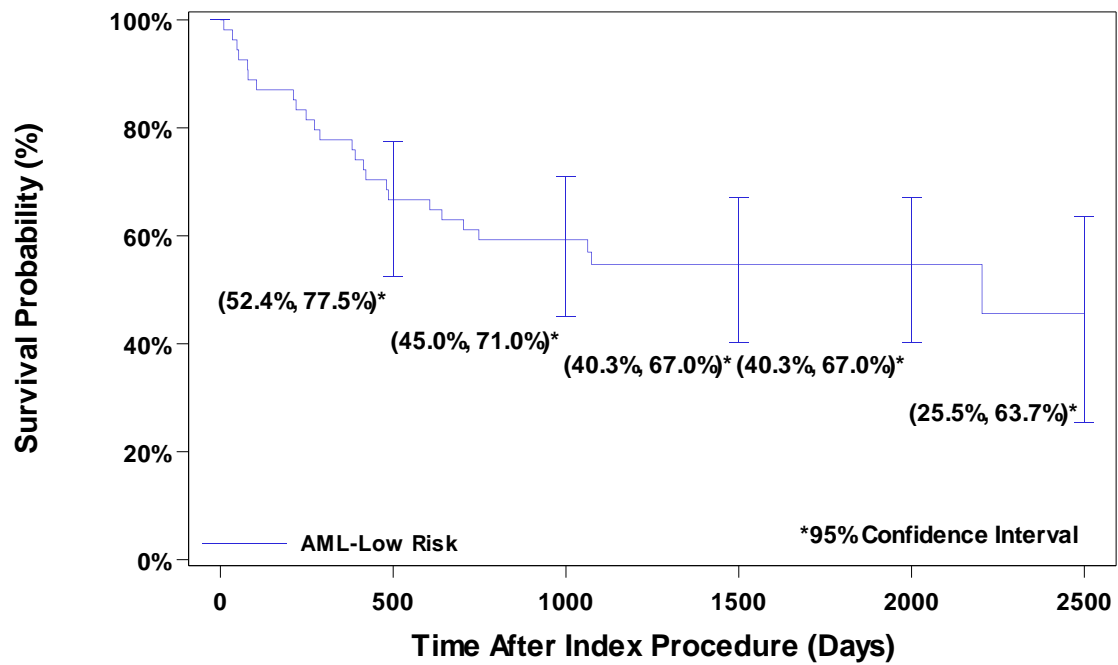
data survpct; set plotds(where=(tAtRisk>0));
  length ID FUNCTION LABEL ANCHOR JUSTIFY TEXTWEIGHT X1SPACE Y1SPACE $100
  WIDTH X1 Y1 8;
  retain ID "EST" FUNCTION "TEXT" ANCHOR "TOPRIGHT" JUSTIFY "RIGHT"
  TEXTWEIGHT "BOLD" X1SPACE "DATAVALUE" Y1SPACE "DATAVALUE" WIDTH 60;
  label = " (||strip(put(lcl, percent11.1))||", "||strip(put(ucl,
  percent11.1))||")*";
  x1 = tatrisk;
  Y1= surv-0.15;
run;

data anno; set survpct caption;
run;

proc sgplot data=plotds sganno=anno;
  [SGPLOT BLOCKS PRESENTED EARLIER ADDED HERE];
run;

```

Using the syntax above, the following output is created.



	No. At Risk					
AML-Low Risk	54	36	27	18	6	2

ADDITIONAL NOTES

As discussed in the introduction, the Kaplan-Meier method is very frequently used for estimating survival as well as failure patterns over time. The examples in this paper focused on survival. However, if failure estimates are needed, the PROC LIFETEST procedure codes included in this paper can be adapted accordingly with simple modification:

```
proc lifetest data = adkm
  plots=survival(failure atrisk=(0 to 2500 by 500) cl);
  time T*STATUS(0);
  ods output failureplot=plot;
run;
```

In addition, one useful feature of PROC SGPLOT is that it can output PROC TEMPLATE from the plot into a file using the TMPLOUT= option in the procedure statement. These files can be imported later and used with PROC SGRENDER to produce plots of the same style without needing to reproduce PROC SGPLOT code.

CONCLUSION

This paper introduces the flexibility of adding additional features and annotations into Kaplan-Meier curves using PROC SGPLOT procedure. PROC SGPLOT is an effective SAS procedure that allows more efficient customization while keeping code simple and straightforward. With the additional employment of an annotation dataset, the PROC SGPLOT procedure is powerful in further adaption and adjustment of survival plots.

REFERENCES

- Amruthnath, S. 2019. "PROC SGPLOT over PROC GPLOT." *MWSUG Conference*, DV01-2013. Columbus, OH: Midwest SAS® Users Group.
- Goel, M. K., Khanna, P., and Kishore, J. 2010. "Understanding survival analysis: Kaplan-Meier estimate". *International journal of Ayurveda research*, 1(4), 274–278. <https://doi.org/10.4103/0974-7788.76794>
- In, J., and Lee, D. K. 2018. "Survival analysis: Part I - Analysis of time-to-event." *Korean journal of anesthesiology*, 71(3), 182-191. <https://doi.org/10.4097/kja.d.18.00067>
- Kaplan, E. L. and Meier, P. 1958. "Nonparametric estimation from incomplete observations." *Journal of the American Statistical Association*, 53:457–81.

CONTACT INFORMATION

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

Name: John Henderson

Enterprise: Abbott

Address: 387 Technology Circle NW, Suite 500, Atlanta, GA 30313

E-mail: john.henderson5@abbott.com

Name: Lili Li

Enterprise: Abbott

Address: 3200 Lakeside Drive, Santa Clara, CA 95054

E-mail: lili.li4@abbott.com