ABSTRACT
In clinical trials, the Kaplan-Meier (KM) estimator is the most widely used method for estimating the distribution functions of time-to-event endpoints. KM estimation is often performed through the SAS PROC LIFETEST procedure, which is also used to create plots of survivor curves. However, this procedure is limited in terms of its graphical displays and its inability to calculate the number of patients at risk.

To improve upon PROC LIFETEST, the Survival Analysis Macro (SAM) Committee at Merck developed the %KMPlot macro. Using the Kaplan-Meier (KM) Product Limit Method, this macro provides estimates of the cumulative event rate percentage or the survival rate percentage over a study period. %KMPlot also enhances the SAS LIFETEST procedure by using PROC GPLOT to produce report-ready graphs, as well as output SAS datasets that can be used for further analysis or graph customization.

KEYWORDS: SAS PROC LIFETEST, Kaplan-Meier, Survival Analysis, SAS Macro

INTRODUCTION
The %KMPlot macro was developed at Merck as a tool for estimating survivor functions. It provides estimates of either the cumulative event rate percentage or the survival rate percentage (1 minus the cumulative event rate) over a study period. It calculates these estimates by using the Kaplan-Meier (KM) Product Limit Method (Kaplan and Meier, 1958). The KM Product Limit estimator yields a consistent estimate of the cumulative percentage of patients experiencing an event at each event time by accounting for the decreasing number at risk over time.

In addition, the %KMPlot macro reduces many of the limitations of the current PROC LIFETEST procedure. For example, PROC LIFETEST does not allow the user to control the graph's legend or the range of its axes. There is also no control over the range of the curves and no way to reflect variability in the KM estimate through plots on a graph. Finally, PROC LIFETEST is unable to calculate the number of patients at risk, which is used in many papers regarding survival analyses.

In contrast, the %KMPlot macro provides the user with much greater control and flexibility. This paper will describe the basic features and structure of this macro and illustrate its usage through some examples.

BASIC FEATURES
The %KMPlot macro offers many useful options for tailoring graphs to the needs of the user. It produces report-ready graphs with flexible graphical manipulation. It provides options for selecting either the survival or failure curve and for controlling the range of the x- and y-axes, in addition to the range of the curve itself. If more information is needed in the graph, the macro has functions for adding a text box, as well as for adding titles and footnotes.

%KMPlot also enables a user to indicate the number of patients at risk. Users can pre-select certain time points to display risk sets (number at risk) by strata group. They can even terminate the graph when the proportion of patients that is free of an event but still in follow-up becomes small.

In addition, this macro has the option of providing point-wise CI/SE bars plotted at specific time points at the user-specified confidence level \( \alpha \). Alternatively, it can provide continuous upper-lower bound CI/SE curves.

Finally, macro %KMPlot offers some other conveniences to the user. The graphics produced by this macro can be exported into an image file, which can be easily imported into MS Word or can be created in an RTF formatting file. This macro also makes two additional SAS output datasets available for further statistical analyses or plotting.
MACRO MINIMUM REQUIREMENTS
Macro %KMPlot was developed in SAS Version 8 and SAS Base, Graph, and Stat. An MS Windows NT or XP operating system is required for running this macro.

This macro also requires a user to fulfill the following four minimum parameters:

- **Dsin** (input dataset)
- **SurvTime** (relative time variable)
- **Cen_Var** (censoring variable)
- **Cen_Ind** (censoring indicator)

If any of the subjects are missing data for any of these variables in a macro call, these subjects will be excluded from the analyses.

MACRO STRUCTURES
There are 28 total parameters in the %KMPlot macro. The chart below lists all of these parameters and includes a brief description of each one. Examples showing how key parameters are used follow this chart.

**Macro Parameters for %KMPlot**

<table>
<thead>
<tr>
<th>Macro Parameter</th>
<th>Description</th>
<th>Default</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dsin</strong></td>
<td>Input SAS dataset name</td>
<td>*Required</td>
</tr>
<tr>
<td>Popcond</td>
<td>Condition of a subset of the input dataset (e.g., POPCOND=%str(endpoint=&quot;DEATH&quot;))</td>
<td></td>
</tr>
<tr>
<td>AN</td>
<td>Subject identifier (e.g., AN=alloc)</td>
<td></td>
</tr>
<tr>
<td>Macro Parameter</td>
<td>Description</td>
<td>Default</td>
</tr>
<tr>
<td>-----------------</td>
<td>-------------</td>
<td>---------</td>
</tr>
<tr>
<td>Survtime</td>
<td>Relative time</td>
<td>*Required</td>
</tr>
<tr>
<td>Cen_var</td>
<td>Censor variable. Must be numeric.</td>
<td>*Required</td>
</tr>
<tr>
<td>Cen_ind</td>
<td>Censor indicator to indicate whether values of Survtime are complete or censored</td>
<td>*Required</td>
</tr>
<tr>
<td>Alpha</td>
<td>Alpha Level for the confidence interval</td>
<td>$\alpha=0.05$, if CI=SE=CI</td>
</tr>
<tr>
<td>DocLibr</td>
<td>Doc file path (required if GOUTKM is specified)</td>
<td></td>
</tr>
<tr>
<td>TxtDim</td>
<td>Delimiter for text box, title, and footnote</td>
<td></td>
</tr>
<tr>
<td><strong>Survival Curve Parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sur_fail</td>
<td>Plot type: S (survival) or F (failure).</td>
<td>F</td>
</tr>
<tr>
<td>DsoutKM</td>
<td>Output dataset name from %KMPPlot (for AXUM plot). NOTE: Allows user to specify either temporary (filename) or permanent datasets (libname.filename).</td>
<td></td>
</tr>
<tr>
<td>DsoutStat</td>
<td>output dataset (for statistics)</td>
<td></td>
</tr>
<tr>
<td>TmFactor</td>
<td>Time Unit converter. A positive numeric value which scales the follow-up time (x-axis). If specified, the x-axis will be displayed in the converted scale and the specifications of X_AXIS and TIMELIST should be specified in the converted scale. e.g., if SURVTIME is in days and TMFACTOR=365.25, then the x-axis will be displayed in the scale of SURVTIME/365.25 (years). Therefore, X_AXIS and TIMELIST should be specified in a scale of years. (e.g., X_AXIS=0 to 6 by 1; TIMELIST=0 1 2 3 4)</td>
<td>1</td>
</tr>
<tr>
<td>TimeList</td>
<td>List of time points at which the numbers at risk are displayed. NOTE: user should specify TIMELIST using the converted scale when TMFACTOR is specified. e.g., If SURVTIME is in days and TMFACTOR=365.25, then the TIMELIST should be specified in years instead of days.</td>
<td></td>
</tr>
<tr>
<td>KMStrata</td>
<td>Stratification variable (e.g., treatment group). One KM curve is produced for each stratum level.</td>
<td></td>
</tr>
<tr>
<td>Stratfmt</td>
<td>Format for stratification. Formatted stratum values are used in the graph legend.</td>
<td></td>
</tr>
<tr>
<td>X_axis</td>
<td>Range (i.e., minimum and maximum limit) and tick labels for the x-axis. e.g., 0 2 4 6, or 0 to 6 by 2, or 0 to Tpct10 (where Tpct10 stops the x-axis when the risk set is &lt; 10% in any of the stratification/treatment groups). NOTE: User should specify x-axis range using the converted scale when specifying the TMFACTOR (time factor) macro parameter (e.g., if the SURVTIME is in days and TMFACTOR=365.25, then the x-axis should be specified in years instead of days). ALSO NOTE: The maximum of X_AXIS takes priority over the maximum of TIMELIST (e.g., if the maximum of X_AXIS is 6 years but the maximum of TIMELIST is 8 years, the graph will be cut off at 6 years). The x-axis is equally spaced. If we specify X_AXIS=0 10 40 50 80 100 130, the x-axis will not be proportionally scaled. Options: Tmaxflup = Maximum follow-up time (default) Tmaxevent = Maximum event time Tpctbx = User specified minimum proportion remaining at risk (e.g., Tpct10 = stop curves when the risk set is &lt; 10% in any of the stratification/treatment groups) XX = User-specified time to stop curves (e.g., 8)</td>
<td></td>
</tr>
<tr>
<td>Macro Parameter</td>
<td>Description</td>
<td>Default</td>
</tr>
<tr>
<td>-----------------</td>
<td>-------------</td>
<td>---------</td>
</tr>
<tr>
<td>X_curstp</td>
<td>The maximum time point on the x-axis where the KM curves stop. It should be specified in the converted scale when TMFACTOR is specified.</td>
<td></td>
</tr>
</tbody>
</table>
| NOTE:          | 1. X_CURSTP determines the extent of all curves (i.e., where KM curves stop), while X_AXIS determines the extent of the graph (i.e., where the x-axis stops)  
2. If X_CURSTP=Tmaxevent (maximum event time), and there are no events after the maximum time point in a stratum, the curve for that stratum will stop before the specified X_CURSTP. | |
| Options:       | Tmaxflup = Maximum follow-up time (default)  
Tmaxevent = Maximum event time  
Tpctbx = User specified minimum proportion remaining at risk (e.g., Tpct10 = stop curves when the risk set is < 10% in any of the stratification/treatment groups)  
XX = User-specified time to stop curves (e.g., 8) | |
| Y_axis         | Range and tick labels for y-axis | no label |
| X_label        | Label for x-axis | If Suv_Fail=F then default is “Cumulative percentage with event.”  
If Suv_Fail=S then default is “Percentage event-free.” |
| Y_label        | Label for y-axis | |
| Debug          | Y or N | N |
| CI_SE          | CI (confidence interval), SE (standard error), or BLANK. Specifies whether to compute and plot confidence intervals or standard error bars.  
NOTE: If CI or SE is specified and TIMELIST is not specified, the macro will generate a CI or SE band over all the event time points. If CI or SE and TIMELIST are specified, the CI or SE bar will be generated at each TIMELIST point. | |
| GoutKM          | Output graphics file | |
| GStps=N        | Suppress graph. Y: suppress graph, N: produce graph | N |
| TxtBox          | Text for a Text Box (e.g., Log-Rank=xxx, p-value=x.xx, Hazard ratio=xx.x). Default delimiter is | |
| TitleKM        | Title for the KM plot. Default delimiter is | |
| FootKM         | Footnote for the KM plot. Default delimiter is | |

**EXAMPLES**
This section offers three hypothetical examples to illustrate how the %KMPlot macro works with different sets of parameters. Example 1 illustrates the basic functionality of %KMPlot. The second example builds on the basic functionality shown in example 1 and allows for more precise graphics capabilities that PROC LIFETEST, particularly in the depiction of the confidence interval and beyond PROC LIFETEST capabilities in terms of calculating the number of patients at risk and producing SAS datasets for further analysis or graphics. Finally, example 3 is similar to example 2 in terms of providing greater graphics flexibility than PROC LIFETEST. Graphically, example 3 depicts the standard error in the graph.

**EXAMPLE 1**
The first example is the simplest of the three. It uses only the minimum parameters required by the macro. Since no
time factor is specified, the graph produced simply indicates a raw time unit. There is no treatment group, and the curve displayed in this case is the default failure curve.

```sas
%KMPlot(
    /***** General Parameters *****/
    dsin=kmdtdir.evnttime01  /* input data name */
    ,survtime=time2       /* relative time variable */
    ,cen_var=censor1      /* censor variable */
    ,Cen_Ind=0             /* censor indication */
);  
```

EXAMPLE 2
The second example is much more complicated. It has a number of very precise specifications, which involve the application of numerous parameters. The following is a list of the specifications required by this hypothetical:

1. There are two treatment groups that must be formatted;
2. A subset of input data is used;
3. The time unit must be converted from days to months;
4. The time tick must be displayed from Month 0 to Month 50 in intervals of 10 months with the curve stopped at Month 45;
5. The x-axis time range must be displayed from Month 0 to the maximum event time;
6. The y-axis ticks must be displayed from 0% to 40% in intervals of 5%;
7. The x-axis and y-axis labels must be customized; and
8. A failure curve with a 95% confidence interval is required.

```sas
proc format;
    value trtfmt
       0='Placebo'
       1='Treatment';
run;

%KMPlot(
    /***** General Parameters *****/
    dsin=kmdtdir.renaal  /* input data name */
    ,PopCond=%str(endpoint='ESRD')
    ,an=an
    ,survtime=relday
    ,cen_var=censor
    ,Cen_Ind=0
    ,DocLibr=&graphdir

    /***** Kaplan-Meier Curve Parameters *****/
```
EXAMPLE 3
The third example is very similar to the second one with the exception of two differences. One difference is in the
 time ranges that are used in the two examples. The other difference is that Example 2 displays a failure curve with a
 confidence interval, while Example 3 displays a survival curve with standard error.

%KMPlot(
  /***** General Parameters *****/
  dsin=test1 /* input data name */
  ,PopCond= /* subset of input dataset */
  ,an=allocat /* allocation number variable */
  ,survtime=time1 /* relative time variable */
  ,cen_var=censor1 /* censor variable */
  ,Cen_Ind=0 /* censor indication */
  ,Alpha=0.05 /* if missing, set to 0.05 */
  ,DocLibr=&graphdir /* doc file path */
  ,TxtDlim=| /* delimiter for text */

  /***** Kaplan-Meier Curve Parameters *****/
  ,sur_fail=s /* S|F plots, default is F */
  ,dsoutKM=dsoutkm /* output dataset of KM analysis */
  ,DsoutStat=dsoutstat /* output dataset of KM analysis */
  ,TimeFactor= /* relative time convert factor */
  ,TimeList=0 to 1600 by 200 /* ticks of at-risk number */
  ,KMstrata=treatment /* stratification for KM curve */
  ,straFmt=ftrt /* format corresponds to KMstrata*/
  ,x_axis=0 to 1600 by 200 /* x axial ticks, def 0 to max */
);
CONCLUSION
The Kaplan-Meier (KM) estimator is the primary method for estimating survivor functions in clinical trials. The current version of SAS PROC LIFETEST can be used to obtain the KM estimator, but not without limitations in terms of KM estimation and graph production. The %KMPlot macro, however, is able to exceed these limitations and provide many significant capabilities to users performing survivor analyses. As a result, after being validated and internally approved, the %KMPlot macro has been widely used in statistical analysis in Merck's Phase II and Phase III clinical trial studies.

REFERENCES


ACKNOWLEDGMENTS
We would like to thank each member of the SAM Committee whose untiring efforts resulted in making %KMPlot a successful product. We would also like to thank our managers since we could not have produced this macro without their support and encouragement. Finally, we are grateful to Margaret Coughlin for her careful review of this paper and for her helpful comments and suggestions.

TRADEMARKS
SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS
Institute Inc. in the USA and other countries. ® indicates USA registration.

CONTACT INFORMATION
Weifeng Xu
Merck & Co., Inc.
BL 3-7
West Point, PA 19486-0004
(484) 344-2896
Email Weifeng_xu@merck.com

Bill Wang
Merck & Co., Inc.
UNA-102
West Point, PA 19486-0004
(484) 344-3005
Email William_wang@merck.com