PharmaSUG 2025 - Paper DV-018

# Visually Exploring Kaplan-Meier Curves Using SAS GTL and R survminer

Yang Gao, Pfizer Inc.

## ABSTRACT

Kaplan-Meier curves are commonly used to visually summarize time-to-event data and present primary trial findings in clinical trials. A Kaplan-Meier curve shows the survival probability of an event at a certain time interval. In the SAS system, PROC LIFETEST and SGPLOT procedures are used to generate simple survival plots. SAS programmers can use SAS Graph Template Language (GTL) directly to design and customize their graphs using the TEMPLATE and SGRENDER procedures. In addition to SAS, R, a free programming language, is widely used for data manipulation, statistical modeling, and visualization. R has highly advanced graphical capabilities. This paper illustrates the procedures to generate Kaplan-Meier plots using two approaches of SAS GTL and survminer R package. By providing methods in both languages, readers have options when creating Kaplan-Meier curves for regulatory submission.

## INTRODUCTION

Kaplan-Meier is the most frequent survival analysis method used in oncology randomized clinical trials. The Kaplan-Meier plot offers a visual representation of predicted survival function of two or more treatment groups. It is not a smooth curve or line, but a series of downward steps occurring each time a patient has an event (e.g., death in an overall survival curve).

SAS and R are the most reliable statistical tools adopted in clinical data analysis and reporting. The SAS analytical procedures can produce graphs when SAS ODS Graphics is enabled. One survival analysis procedure is LIFETEST. Although LIFETEST is straightforward, there are limitations in the plot customization to support submission purposes. The SAS Statistical Graphics (SG) procedures are an alternative option. However, these two procedures do not provide many of the advanced layout capabilities as GTL. Customization beyond the procedure options still require the use of GTL. GTL syntax supports a variety of layout, plot, and other statements to create a custom and complicated graph which is required for submission. This paper discusses the creation of Kaplan-Meier plot using SAS GTL.

In addition to the SAS programming language, R has gained popularity and interest in statistical computing and data visualization. ggplot2 simplifies many of the graphics creation details while using the power of the underlying R graphics system to ensure those graphics are both beautiful and meaningful. It provides a more programmatic interface for specifying what variables to plot, how they are displayed, and general visual properties. survminer R package is built based on ggplot2. This paper illustrates constructing customized Kaplan-Meier plot using survminer R package.

## DUMMY DATASET IMPORTED TO SAS AND R

The Kaplan-Meier plot created in this paper uses a dummy SAS dataset as the input data frame.

| Variables | Variable Labels |
|---|---|
| TRT01P | Planned treatment for period 01 |
| TIME | Event or censored time |
| SURVP | Percentage of survival probability |
| CENSOR | Censoring indicator |
| TIMEC | Censoring timepoint |
| TABTIME | Selected time points |
| NBRISK | Number of subjects at risk |

**Table 1. Description of Dummy Dataset**

## KAPLAN-MEIER CURVES USING SAS GTL

### Defining a Graph Template

GTL uses a structured "building-block" approach to define a graph. The TEMPLATE procedure starts with a DEFINE statement. The first layer is the DEFINE block, indicating that a graphics template named “km” is to be created. The second layer is the BEGINGRAPH block, which defines the outermost container of the graph. A variety of statements can be used to define the content of the graph. The third layer is the LAYOUT block which defines the type of graphical layout to be used. In the figure shown below, the stepplot with survival time versus survival probability and the scatterplot with censoring time points versus survival probability are overlaid.

In addition to the customization of graph title, legends, and axes, an at-risk table is frequently included to indicate patient numbers at risk. An additional LAYOUT OVERLAY block can be built to display the number of at-risk subjects across different timepoints. The AXISTABLE statement is used to create an axis-aligned inset with an axis table. Since the primary plot and the at-risk table need to be aligned on all time points of the x-axis, the LAYOUT LATTICE statement can be used to define multi-cell grid of graphs that can align across cells or blocks. The two LAYOUT OVERLAY blocks are nested within the LAYOUT LATTICE. The below syntax is used to create an enhanced Kaplan-Meier plot shown in Figure 1.

```
proc template;
   define statgraph km;
      begingraph;
         entrytitle halign = center "Kaplan-Meier Estimates of Overall Survival";
         layout lattice / …;
            layout overlay / yaxisopts = (label = "Overall Survival (%)" griddisplay = on offsetmin = 0.05…)
                             xaxisopts = (label = "Time in Months" griddisplay = on offsetmin = 0.05…);
                             stepplot x = time y = survp / …;
                             scatterplot x = timec y = survp / …;
                             discretelegend 'step' / …;
             endlayout;
             layout overlay / …;
                entry halign = left "Number of Subjects at Risk"/ …;
                axistable x = tabtime value = nbrisk /…;
             endlayout;
          endlayout;
      endgraph;
   end;
run;
```

The Kaplan-Meier plot can be further customized in terms of font size, color, family, weight, and so on. In GTL, most statements provide options that enable users to specify attributes for the fills, lines, data markers, and text used in the display. Other attributes, such as labels, tick values, and markers can be controlled in a similar fashion.

### Creating the Graph

To produce a graph, the SGRENDER procedure is run to associate the appropriate data set with the compiled template.

```
proc sgrender data=kmdata template=km;
```

run;

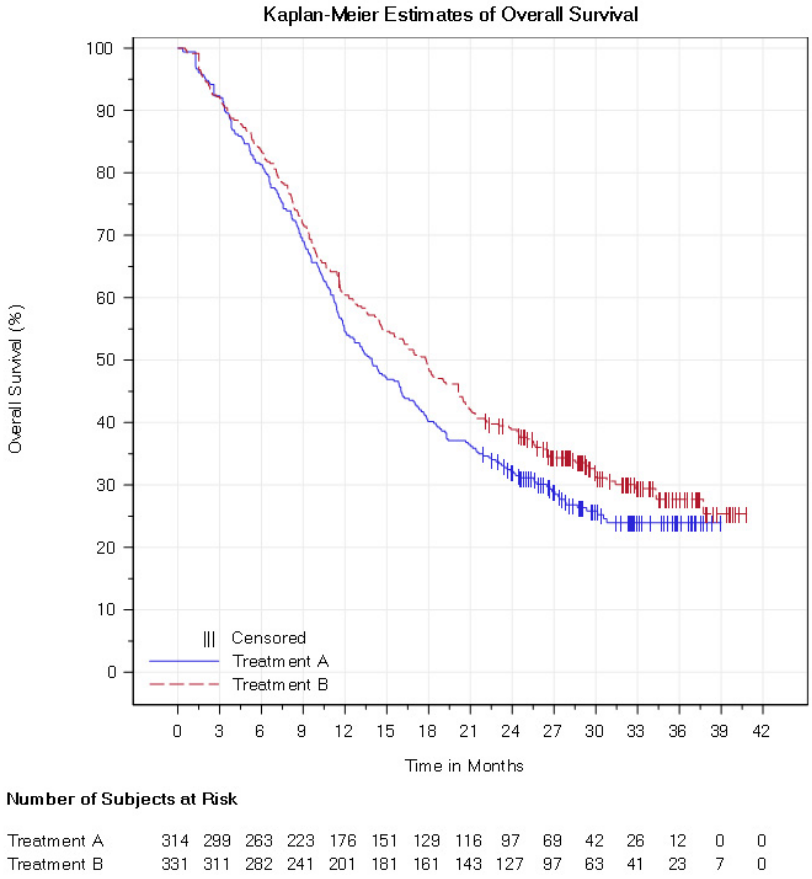


**Figure 1. An Enhanced Kaplan-Meier Plot using SAS GTL**

## KAPLAN-MEIER CURVE USING R SURVMINER PACKAGE

R is one of the most popular platforms for data analysis and visualization. The ggplot2 package provides a grammar-based system for generating graphs in a unified and coherent manner, allowing users to create new and innovative data visualizations.

The current version of survminer R package contains the function ggsurvplot() for drawing 'ready-to-publish' survival curves with the 'number at risk' table. The ggsurvplot function is built on ggplot2. Creating an informative, elegant and flexible survival plot using ggsurvplot function from the survminer R package is discussed below.

The main plot can be drawn using below R code. The ggsurvplot function can associate the plot with data, customize the x-axis and y-axis attributes, add the at-risk table and specify the legend location and contents.

```
Library ("survminer")
fit <- survfit (Surv (TIME, CENSOR) ~ TRT01P, data = kmdata)

ggsurv <- ggsurvplot (
```

```
 fit,
 data = kmdata,
 size = 0.5,
 palette = c ("blue", "red"),
 linetype = c (1,5),
 font.family = "Arial",
 xlab = "Time in Months",
 xlim = c (0, 42),
 break.x.by = 3,
 ylab = "Overall Survival (%)",
 axes.offset = TRUE,
 risk.table = TRUE,
 risk.table.y.text.col = FALSE,
 risk.table.height = 0.11,
 risk.table.y.text = TRUE,
 risk.table.fontsize = 4,
 legend = c (0.2, 0.2),
 legend.labs = c ("Treatment A",  "Treatment B"),
 legend.title = " |||    Censored")
```

The title, label, theme, title attributes and axis attributes can be further customized using the below R code.

```
ggsurv$table <- ggsurv$table +
 labs (title = "Number of Subjects at Risk") +
 theme_cleantable () +
 theme (
  plot.title = element_text (size = 13, color = "black", face = "bold"))
```

The primary survival plot can be customized in terms of the label, theme, legend title and axis text as well as grid lines using below R code. The theme selected is to ensure all text size and colors are consistent. The grid lines for x-axis are to assist with the readability of the survival plot. The customized Kaplan-Meier plot for regulatory submission is generated shown in Figure 2.

```
ggsurv$plot <- ggsurv$plot +
 labs (title = "Kaplan-Meier Estimates of Overall Survival") +
 scale_y_continuous (breaks = seq (0,1, by = 0.1), labels = seq (0, 100, by = 10)) +
 theme (
  plot.title=element_text (size = 13, color = "black", hjust = 0.5),…)
```

## CONCLUSION

Both SAS and R have a strong capacity for creating submission ready Kaplan-Meier plots. Depending on the situation and analysis needs, users may choose one package over the other. This paper demonstrates the flexibility of graphing in R, while showing the capability of SAS GTL supporting a variety of layouts.

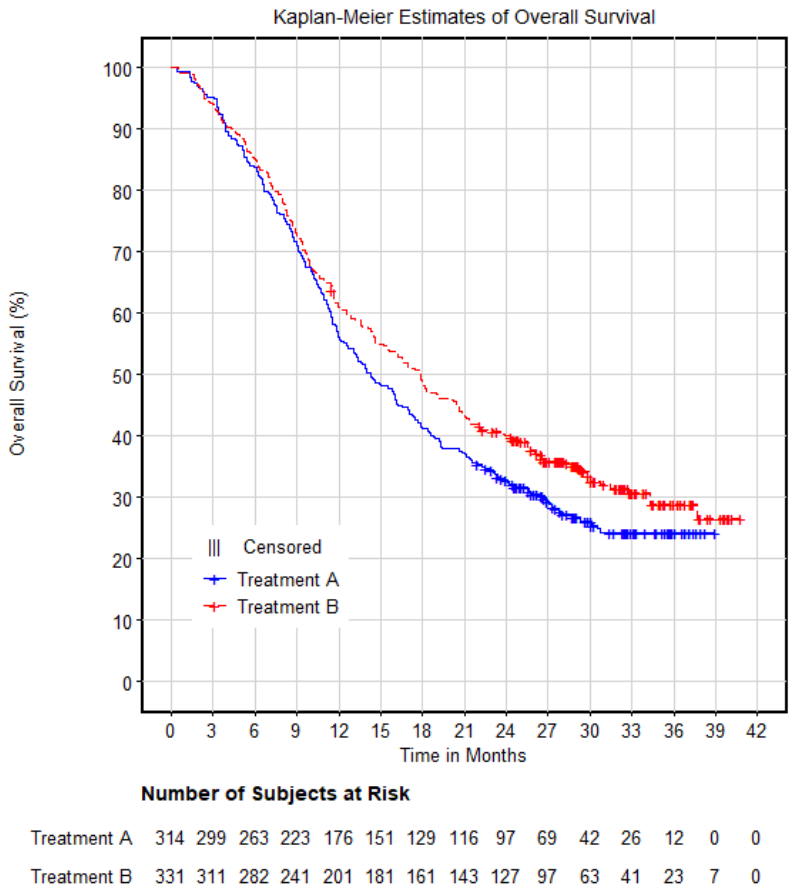


**Figure 2. A Customized Kaplan-Meier Plot using R survminer package**

## REFERENCES

SAS®9.4 Graph Template Language: User's Guide, Fifth Edition
https://documentation.sas.com/api/docsets/grstatug/9.4/content/grstatug.pdf
survminer: Drawing Survival Curves using 'ggplot2'. https://cran.r-project.org/web/packages/survminer/index.html
Barraclough, H. etc. 2011. "Biostatistics Primer: What a Clinician Ought to Know: Hazard Ratios". *Journal of Thoracic Oncology, 6:978-982*
Kabacoff R. 2015. *R in Action.* 2nd ed.

## CONTACT INFORMATION

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

Name: Yang Gao
Enterprise: Pfizer Inc.
Address: 401 N Middletown Rd, Pearl River, NY 10965
E-mail: yang.gao2@pfizer.com
Web: www.pfizer.com