

Survival Plots using SAS PROC LIFETEST, GPLOT, and SGPLOT:

What are their difference?

Ka Chun Chong and Benny Chung-Ying Zee

Clinical Trials and Biostatistics Laboratory, Shenzhen Research Institute,
The Chinese University of Hong Kong, China

ABSTRACT

Kaplan Meier survival curve is a useful non-parametric approach to summarizing the time-to-event data such as the overall survivals in cancer studies. In the SAS system, LIFETEST, GPLOT, and SGPLOT procedures are common ways to generate the survival curves. However, the logistics in SAS programming are different among these procedures. In this paper, the difference will be demonstrated by making use of some examples.

INTRODUCTION

Time-to-event data are usually outcomes of interest in many medical studies. For example, they can be death, recurrence of a tumor, discharge from hospital, or cessation of breast feeding. In cancer trials, the overall survivals are common to be primary endpoints for many therapeutic interventions. For these kinds of measurements, censoring of patients may occur in several ways such as lost to follow up.

One of the appropriate ways of summarizing time-to-event data is to use methods of survival analysis such as Kaplan Meier estimation. Kaplan Meier method is a widely used non-parametric approach to estimate and plot the survival probabilities as a function of time. It can also be used to obtain descriptive statistics such as the median survival time, and to compare the survival curves for two or more groups of subjects.

In the SAS system, several procedures: LIFETEST, GPLOT, and SGPLOT are useful tools to generate the Kaplan Meier survival curves. However, the logistics in SAS programming are different among these procedures. In this paper, the difference will be demonstrated by making use of some examples.

SURVIVAL DATASET

Suppose there is a placebo-controlled trial for patients with advanced hepatocellular carcinoma (HCC) (Llovet, 2008). Patients were equally randomized into placebo arm and intervention arm of sorafenib. The primary endpoint of this trial is the overall survival. Given the following settings, an analysis dataset named ADSURV with totally 100 patients is simulated for our demonstration purpose:

- USUBJID: Unique subject identifier
- TRTC: Treatment arm in character format
- SURV: Length of overall survival in months (assuming follow normal distribution with mean 10.7 and a standard deviation (SD) of 3 months for sorafenib arm and mean 7.9 and a SD of 3 months for placebo arm)
- SURVCEN: Censoring flag for overall survival (assuming 0.1 proportion for both groups)
- ECOGN: Baseline Eastern Cooperative Oncology Group (ECOG) scores (assuming follow a uniform distribution from 0 to 3 for sorafenib arm and from 1 to 4 for placebo arm)

The sample SAS code for generating the dataset is:

```
data adsurv;
  do usubjid = 1 to 50;
    trtc = "Sorafenib";
    surv = round(rand("normal",10.7,3),0.1);
    survcen = rand("bernoulli",0.1);
    ecogn = int(4*ranuni(0));
    output;
  end;
  do usubjid = 51 to 100;
    trtc = "Placebo";
    surv = round(rand("normal",7.9,3),0.1);
```

```
survcen = rand("bernoulli",0.1);  
ecogn = int(4*ranuni(0)+1);  
output;  
end;  
run;
```

KAPLAN MEIER SURVIVAL CURVES

LIFETEST PROCEDURE

LIFETEST procedure is the most common and convenient way in SAS system to generate a survival plot. The LIFETEST PROC can also be used to compute nonparametric estimate of the survivor function by the life table method. In the SAS PROC LIFETEST, the TIME statement is used to indicate the time-to-event variable and followed by an asterisk (*), the name of the censoring variable, and a value that correspond to right censoring. When there is a group for comparison, the STRATA statement is used to specify strata levels. The following is an example:

```
proc lifetest data = adsurv outsurv=survest;  
time surv*survcen(1);  
strata trtc;  
run;
```

An example of Kaplan Meier curve is as follow:

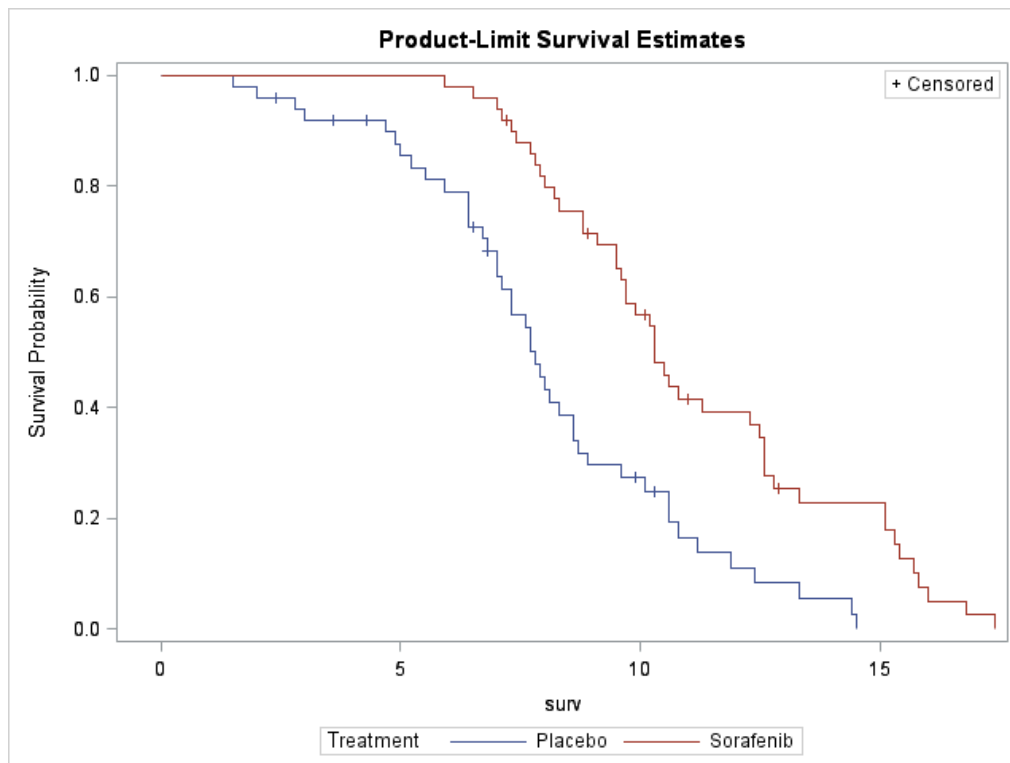


Figure 1. Kaplan Meier curve by LIFETEST PROC

GPLOT PROCEDURE

Although using LIFETEST PROC is straight forward, directly customizing the survival plots would require the use of other procedures, for example, creating a SAS style template by using TEMPLATE PROC (Smoak, 2006). Typically, SAS/GRAPH procedures as well as GPLOT PROC is recommended for the figure customizing convenience on the dataset of estimates (Liu, 2006).

In the previous LIFETEST PROC, the Kaplan Meier estimates are outputted in SURVEST dataset by using the OUTSURV= option in the procedure step. Before generating the survival curve from SURVEST, several data steps have to be done for ensuring the 'tail' probability can be drawn due to last censoring observations:

```
/* To obtain tail probab */
data survtail;
  set survest;
  by trtc;
  if last.trtc;
  keep trtc survival;
  rename survival=tailp;
run;

/* Identify whether the last obs(s) are censored or not */
data survp;
  merge survest survtail;
  by trtc;
  if _censor_ = 1 and not missing (survival) then cenp = survival;
  else if _censor_ = 1 and missing (survival) then do;
    cenp = tailp;
    survival = tailp;
  end;
  else cenp = .;
run;
```

The first data step is to obtain the tail probability in case last several observations are censored. As the dataset SURVEST does not assign a probability for last censored observations for the plotting, the second data step is to develop probability points on plots given they are censoring data.

After making the analysis dataset SURVP, some graphical statements such as SYMBOL, AXIS, and LEGEND can be adopted to make your own survival plot. In many journal paper, authors are usually recommended using black and white figures. In the following example, solid line and dash line represent the survivals of placebo and intervention groups respectively by using SYMBOL1 and SYMBOL2 statements. The SYMBOL3 statement specifies the censoring mark ('+'). AXIS1 and AXIS2 statement are used to control the details of x and y axis such as ranges and labels. AXIS3 statement is an overlap axis to control the plot of censored points. The SAS code is as follow:

```
* Solid line for placebo;
symbol1 line = 1 color = black i = stepj;
* Dash line for intervention;
symbol2 line = 2 color = black i = stepj;
* Cross mark for censored points;
symbol3 color = black;

* Axis control;
axis1 label = (angle=90 height=1.5 'Overall Survival (Probability)')
  major = (height = 1) minor = none order = (0 to 1 by 0.2)
  value = (height=1.5);
axis2 label = (height = 1.5 'Time (Months)') major = (height = 1) minor = none
  order = (0 to 20 by 4) value = (height=1.5);
axis3 label = none major = none minor = none value = none order = (0 to 1 by 0.2);

* Legends control;
legend1 label = none value = (height = 1.5) position = (top right inside) down = 2;
legend2 label = none value = (height = 1.5) position = (bottom right inside);
```

After declaring the plot style, we can make use of GPLOT PROC to draw a customized survival plot:

```
proc gplot data = survp;
  plot survival*surv = trtc / vaxis=axis1 haxis=axis2 legend=legend1;
  plot2 cenp*surv / vaxis=axis3 legend=legend2;
run;
```

Within the GPLOT, PLOT2 statement is employed to specify the censoring points (SYMBOL3) in addition to the PLOT statement of the survival probability for two treatment groups (=TRTC). The customized figure is as follow:

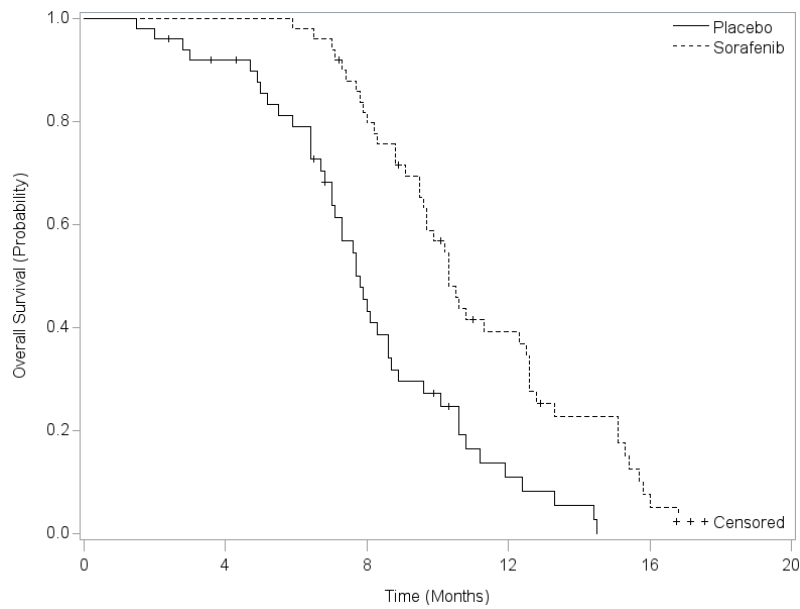


Figure 2. Kaplan Meier curve by GPLOT PROC

SGPLOT PROCEDURE

Apart from GPLOT, SGPLOT is another procedure to create statistical graphics such as histograms and regression plots, in addition to simple graphics such as scatter plots and line plots. The procedure offers a wide variety of graphs using simple syntax – it may be easier to locate, update, modify and maintain programs than GPLOT (Amruthnath, 2013). However, this procedure would sometimes not be flexible enough when creating the survival plots. For example, producing different line patterns for treatment groups still requires the use of graph template language (GTL) (Phillips, 2012). The following is an example using SGPLOT:

```
proc sgplot data = survp;
    step x = surv y = survival / group = trtc lineattrs = (pattern = solid)
    name = "Survival";
    scatter x = surv y = cenp / markerattrs = (symbol = plus) name = "Censored";
    scatter x = surv y = cenp / group = trtc markerattrs = (symbol = plus);
    yaxis label = "Overall Survival (Probability)" values = (0 to 1 by 0.2);
    xaxis label = "Time (Months)" values = (0 to 20 by 4);
    keylegend "Survival";
    keylegend "Censored" / location = inside position = topright;
run;
```

The STEP statement is used to specify the step plot for survival probability against its overall survival time. The options GROUP is used to declare the treatment groups. The first SCATTER statement is adopted to specify the legend of censoring points in case they appear twice in the legend. The second SCATTER statement is to specify the censoring marks ('+') by treatment groups. The XAXIS, YAXIS, and KEYLEGEND statements are used to control the layout of the figures in order to produce a similar figure to that from GPLOT PROC. The figure is as follows:

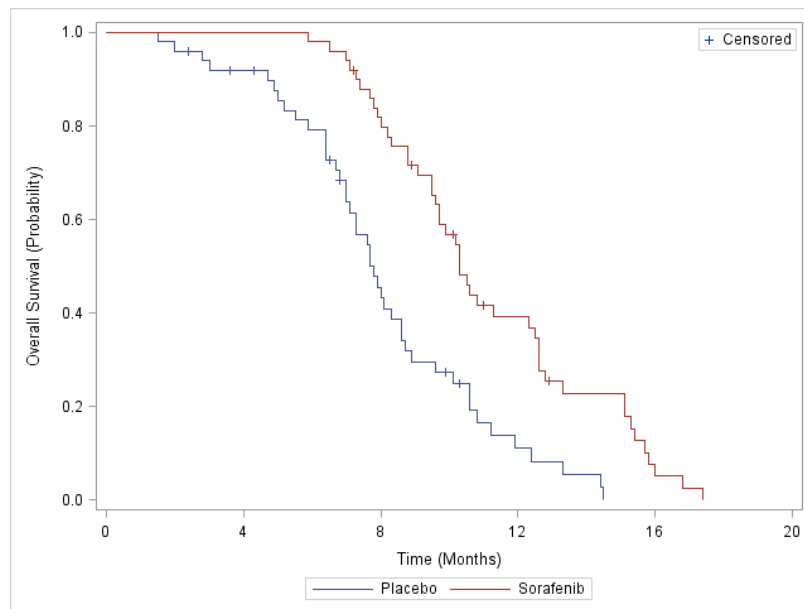


Figure 3. Kaplan Meier curve by SGPLOT PROC

COVARIATE-ADJUSTED SURVIVAL CURVES

PHREG PROCEDURE

In small trials or trials with complex design, covariates could affect the significance on survivals. Therefore, covariates-adjusted survival analysis is a common task for treatment effect investigation. In SAS system, the PHREG procedure is able to conduct a regression analysis of survival data based on a Cox proportional hazards model. Cox's model has been widely used in survival analysis for explaining the effect of covariates on hazard rates. In graphical presentation, as similar to LIFETEST PROC, PHREG can also produce the covariate-adjusted survival curve by using the BASELINE statement. Suppose we want to obtain a mean ECOG score adjusted plot, we first compute the mean value of ECOG score as to make a dataset that contains the set of covariate of interest. MEANS procedure and data step are used:

```
* Overall mean value of ECOG;
proc means data = adsurv noprint;
  var ecogn;
  output out = mecogdat(drop = _TYPE_ _FREQ_) mean = ;
run;

* Dataset of covariate value;
data covecog;
  set mecogdat;
  length trtc $9.;
  do trtc = "Placebo", "Sorafenib";
    output;
  end;
run;
```

The output dataset MECOGDAT contains the overall mean value of ECOG and it is further developed to dataset COVECOG. The dataset is specified in the below COVARIATES option of BASELINE statement in PHREG procedure:

```
proc phreg data = adsurv plots(overlay) = survival;
  class trtc;
  model surv*survcen(1) = trtc ecogn;
  baseline out = adjsurv survival = _all_ covariates = covecog / rowid = trtc;
run;
```

The PLOTS and OVERLAY option are to create one overlay survivor plot for two treatment groups. The ROWID option is to identify the curves in the plot. The mean ECOG adjusted survival plot is as follow:

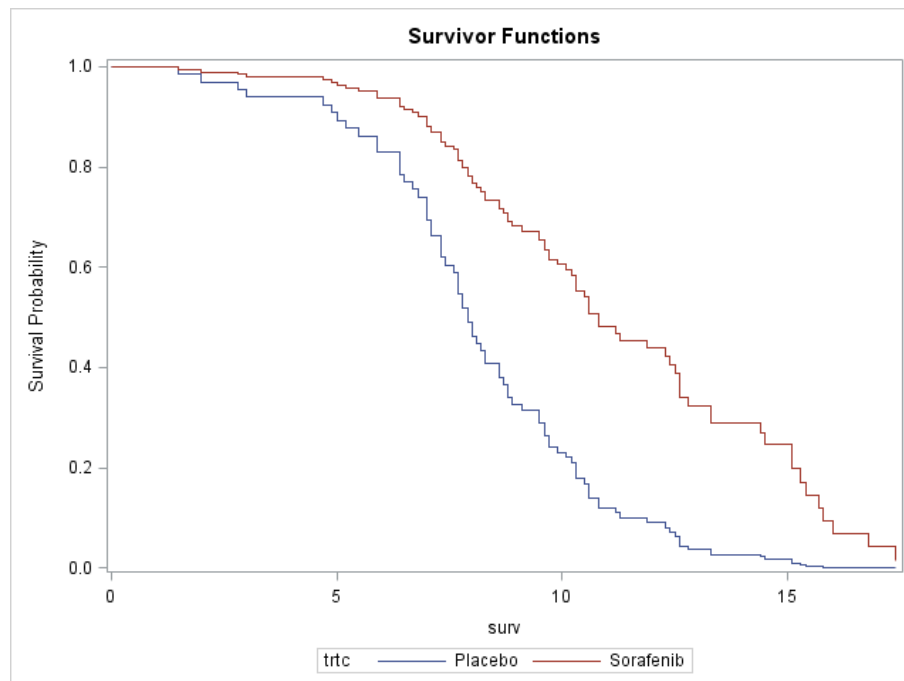


Figure 4. ECOG-adjusted survival curves by PHREG PROC

GPLOT PROCEDURE

Nevertheless, as similar to that using LIFETEST PROC, directly customizing the survival plots would not be an easy way. We employed GPLOT PROC to further customization of survival plots. In the previous PHREG PROC, the adjusted survival estimates are outputted in ADSURV dataset by using the OUT= and SURVIVAL= options in the BASELINE statement. Followed the previous example, we develop a black and white figure styles by using some SAS/GRAPH procedures:

```
symbol1 line = 1 color = black i = stepj;
symbol2 line = 2 color = black i = stepj;
axis1 label = (angle = 90 height = 1.5 'Overall Survival (Probability)')
major = (height = 1) minor = none order = (0 to 1 by 0.2)
value = (height = 1.5);
axis2 label = (height = 1.5 'Time (Months)') major = (height = 1)
minor = none order = (0 to 18 by 3) value = (height = 1.5);
legend1 label = none value = (height = 1.5) position = (top right inside) down = 2;
```

As the estimates from Cox model do not contain censored values, no additional SYMBOL and LEGEND have to be created. The SAS code of GPLOT PROC is as follow:

```
proc gplot data = adjsurv;
    plot survival*surv = trtc
        / vaxis = axis1 haxis = axis2 legend = legend1;
run;
```

and the results:

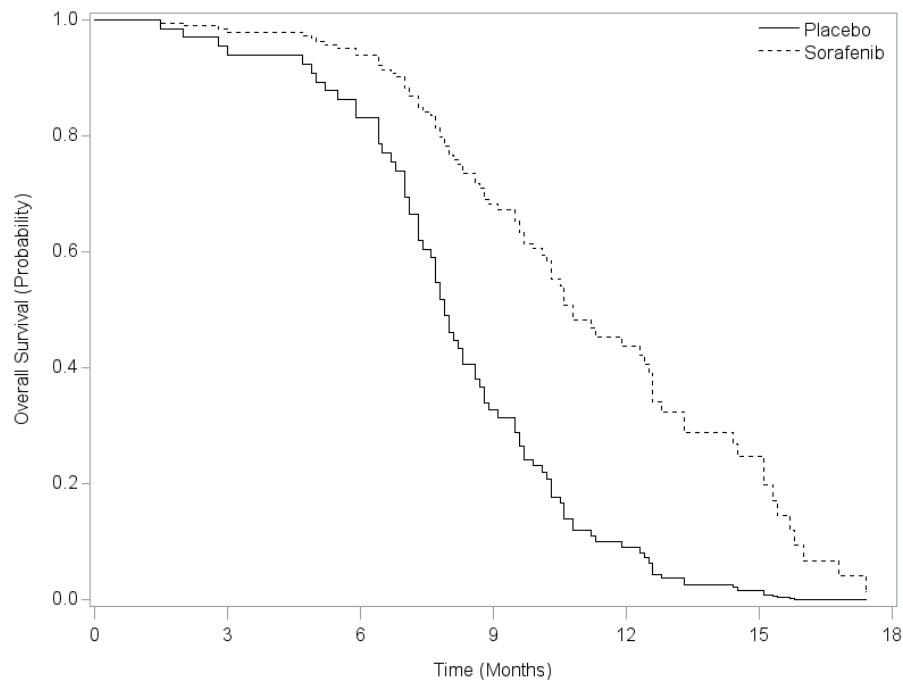


Figure 5. ECOG-adjusted survival curves by GPLOT PROC

CONCLUSION

This paper presents some programming differences among procedures in SAS system to generate survival curves. A brief summary is drawn in the following table:

Procedures	Characteristics
LIFETEST / PHREG	Easy and direct way to generate a survival plot. Customizing the figure require the use of Graph Template Language (GTL).
GPLOT	Customizing the figure can use traditional SAS\GRAPH options and produce more flexible output. It can also customize the covariate-adjusted survival curves.
SGPLOT	Provide more extensive and flexible coding than LIFETEST PROC. Customization beyond the options in procedure still require the use of GTL.

Table 1. Summary on using different SAS procedures to produce survival plots

In brief, the paper would assist SAS users to produce the survival plots based on different characteristics. For example, researcher is recommended using GPLOT procedure to produce black and white figure for their journal submission.

REFERENCES

- Llovet, JM, et al. 2008. "Sorafenib in advanced hepatocellular carcinoma". *New England Journal of Medicine*. 359:378-390.
- Liu, N. Coca-Perrailon, M. 2006. Customizing Survival Plots by Example. *NESUG*.
- Smoak, C. 2006. Survival (Kaplan-Meier) Curves Made Easy. *PharmaSUG*. Paper TT07.
- Amruthnath, S. 2013. PROC SGPLOT over PROC GPLOT. *MWSUG*. Paper DV01-2013.
- Phillips, SD. 2012. A Multifaceted Approach to Generating Kaplan-Meier and Waterfall Plots in Oncology Studies. *PharmaSUG*. Paper DG05.

ACKNOWLEDGMENTS

The work is supported by NSFC [81201817, 31401124, 81473035] and Funding of New Drug Clinical Evaluation Research Technology Platform [2012ZX090312-002].

CONTACT INFORMATION

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

Name: Marc Chong

Enterprise: Centre for Clinical Research and Biostatistics, The Chinese University of Hong Kong

Address: Rm502, JC School of Public Health and Primary Care, The Chinese University of Hong Kong, Shatin, NT, Hong Kong, China

Work Phone: +852 2252 8715

Fax: +852 2646 7297

E-mail: marc@cuhk.edu.hk

Web: www.ccrb.cuhk.edu.hk

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.

Other brand and product names are trademarks of their respective companies.