Paper 239-28

Enhancement of Survival Graphs

**Joanne C. Zhou, GlaxoSmithKline Pharmaceuticals,
Collegeville, PA**

ABSTRACT

As use of time-to-event data in clinical trials is becoming more frequent, survival analysis and the graphical presentation of the survival analysis are more and more popular. In order to aid correct interpretation, overall summary statistics and some reflection of statistical uncertainty should be added in the survival plots. In this paper, Proc Lifetest, Proc Gplot and Proc Greplay are used to present the results of survival analysis with annotated +/-standard errors on the curve for selected time points on the Kaplan-Meier estimates; the numbers of patients at risk are displayed under the time axis.

SURVIVAL ANALYSIS AND PRESENTATION

Simulated Survival Data

Following code is used to generate the survival data. In these data, we assume there are 2 treatment groups, namely study drug and placebo. And there are 250 patients in each treatment group. The survival time is exponentially distributed and the treatment effect is 20 units for study drug and 10 units for placebo. The censoring indicator variable CENSOR is coded 1 for the patients who were censored and 0 for the patients who had event. In addition, we assume the follow up time for the trial is 36 months. So the patients' survival time is set to 36 months if the survival time is longer than 36 months.

```
options nocenter;

proc format;
```

```
  value trt
    1 = 'Study Drug'
    2 = 'Placebo';
run;

data surv;
  do trt = 1 to 2;
    do pat = 1 to 250;
      if trt = 1 then do;
        surv = -log(ranuni(0))*20;
        if mod(pat, 5) then censor = 0;
        else censor = 1;
      end;
      else do;
        if mod(pat, 7) then censor = 0;
        else censor = 1;
        surv = -log(ranuni(0))*10;
      end;
      drop pat;
      patient = put(pat, z5.);
      output;
    end;
  end;
run;

data surv;
  set surv;
  if surv > 36 then do;
    if censor = 1 then surv = 36;
    else surv = surv - int(surv/36)*36;
  end;
run;
```

Sample data:

patient	trt	surv	censor
00001	1	17.6378	0
00002	1	13.8545	0
00003	1	3.2900	0
00004	1	9.3617	0
00005	1	36.0000	1
00006	1	18.4815	0
00007	1	21.3137	0
00008	1	16.1309	0

```
 00009      1     32.1653       0
 00010      1      3.5169       1
.
.
```

Survival Analysis

Log-Rank test is used to analyze the
simulated survival data. For the
graphical presentation, we want to plot
the Kaplan-Meier survival estimates
along with overall summary statistics,
standard error bars on the curve and
the numbers of patients at risk
displayed under the time axis for every
6 months up to 36 months. The survival
analysis is done by calling Proc
Lifetest procedure.

```
ods listing close;
ods output  ProductLimitEstimates  = s_est
            CensoredSummary = cen_sum
            HomTests = homtest
            Quartiles = quart;
  proc lifetest data = surv;
    time surv*censor(1);
    strata trt;
    id patient;
    format trt trt.;
  run;
ods output close;
ods listing;
```

Extraction of Overall Summary Statistics

SAS ODS provides great convenience for
extraction of Log-Rank p-value, median
survival time and other statistics.
Here these statistics are converted
into macro variables so that they can
be annotated onto the survival curves
later on.

```
/****************************************
   Populate the censored patients
   with appropriate KM estimates.
****************************************/

data s_est;
  set s_est;
  retain pr_s;
  if censor = 0 then pr_s = survival;
  surviva = .;
  if censor = 1 then do;
```

```
    surviva = pr_s;
    survival = .;
  end;
run;
```

```
/****************************************
   Convert the total number of
   patients, number of patient with
   event, number of patients with
   censoring into macro variables.
****************************************/

proc sort data = cen_sum;
  by trt;
  where control_var is missing;
run;

data _null_;
  set cen_sum end = last;
  pctevt = 100 - pctcens;
  call symput('groupf'||compress(_n_),
              compress(put(trt, trt.)));
  call symput('ng'||compress(_n_),
              compress(put(total, 8.)));
  call symput('ngevt'||compress(_n_),
              compress(put(failed, 8.)));
  call symput('pgevt'||compress(_n_),
              compress(put(pctevt, 8.2)));
  call symput('ngcen'||compress(_n_),
              compress(put(censored, 8.)));
  call symput('pgcen'||compress(_n_),
              compress(put(pctcens, 8.2)));
run;
```

```
/****************************************
   Convert Log-Rank test p-value
   into macro variable.
****************************************/

proc sort data = homtest;
  where test = 'Log-Rank';
  by test;
run;

data _null_;
  set homtest;
  call symput('logrp', compress(put(probchisq,
              pvalue6.4)));
run;
```

```
/****************************************
   Convert median survival estimates
   and its 95% CI into variables.
****************************************/
```

```
proc sort data = quart;
  by trt;
  where percent = 50;
run;

data _null_;
  set quart;
  call symput('gmed'||compress(_n_),
              compress(put(estimate, 8.2)));
  call symput('llmit'||compress(_n_),
              compress(put(lowerlimit, 8.2)));
  call symput('ulmit'||compress(_n_),
              compress(put(upperlimit, 8.2)));
run;
```

Derive Number of Patients at Risk at Selected Time Points

Derive the number of patients at risk at baseline and every 6 months up to 36 months. Then convert the risk set into macro variables.

```
/*************************************
   Derive the numbers of patients at
risk from dataset surv.
*************************************/

data riskset;
  set surv;
  array rsk x0-x6;
  retain x0 0 x1 6 x2 12 x3 18 x4 24 x5 30 x6 36;
  do i = 1 to dim(rsk);
    if surv >= rsk{i} then do;
      riskat = rsk{i};
      output;
    end;
  end;
  drop x0 x1 x2 x3 x4 x5 x6 i;
run;

proc freq data = riskset noprint;
  table trt*riskat/list out = orisk;
  format trt trt.;
run;


/*************************************
   Convert the risk set into macro
variables.
*************************************/

data _null_;
  set orisk;
  by trt;
  call symput('atrisk'||compress(trt||riskat),
```

```
              count);
run;
```

Create Annotate Dataset Using the Macro Variables for Numbers of Patients at Risk

```
%macro getantt;
  data anno;
  length function style $8 text $20 color $8;
  retain style 'SIMPLEX'; xsys = '3'; ysys = '3';
  x = 10; y = 9.5; size = 1.5; position = '5';
  function = 'label'; text = 'Patients at Risk';
  color = 'black';  output;

  %do i = 1 %to &groupn;
    xsys = '3';  ysys = '3';  x = 2;
    %if &i = 1 %then %do;
      y = 5;  color = 'black';
    %end;
    %else %do;
      y = 2;  color = 'black';
    %end;
    size = 1.5;  function = 'label';
    position = '6';
    text = left(trim("&&groupf&i"));
    output;
    %do j = 0 %to 36 %by 6;
      position = '5';
      xsys = '2';  ysys = '3';
      function = 'label';
      x = &j;
      %if &i = 1 %then %do;
        color = 'blue'; y = 5;
      %end;
      %else %do;
        y = 2;  color = 'red';
      %end;
      text = compress("&&atrisk&i&j");
      size = 1.5;  output;
    %end;
  %end;
  run;
%mend;

%getantt;
```

Create Dataset to Contain Kaplan-Meier Estimates and Plus/Minus their Standard Errors.

Derive the Kaplan-Meier estimates and plus/minus their standard errors for month 6 and every 6 months up to 36 months.

```
proc sort data = s_est out = hilo;
  by trt surv;
run;

data hilo;
  set hilo;
  by trt surv;
  ck = 0;
   if survival = . then survival = surviva;
  if stderr = . then stderr = stdera;
  array rsk x1-x6;
  retain x1 6 x2 12 x3 18 x4 24 x5 30 x6 36;
  do i = 1 to dim(rsk);
    if surv <= rsk{i} and ck = 0 then do;
      riskat = rsk{i}; ck = 1;
    end;
  end;
  drop x1 x2 x3 x4 x5 x6 i ck;
run;

proc sort data = hilo;
  by trt riskat surv;
run;

data hilo;
  set hilo;
  by trt riskat;
  if last.riskat then do;
    survival = survival - stderr;
    surv = riskat; output;
    survival = survival + 2*stderr; output;
  end;
run;
```

Plot the Kaplan-Meier Estimates

Proc Gplot is used to plot the survival
estimates against the observed survival
time.   Notice that the Proc Gplot is
called three times.   One is for the
event curve, one for the censored curve
and one for the +/- standard errors and
number of patients at risk at selected
time points.   The survival statistics
including log-rank p-value are put on
the graph by note statements. The three
graphs are stored in three graphic
entries named G1, G2 and G3 in order to
replay the three graphs onto one graph
that will have both event and censored
curves,   overall   survival   analysis
results, the standard error and number
of patients at selected time points.

```
/********
   Gplot
********/

title1 h = 3 "Kaplan-Meier Survival Estimate (+/-
Standard Error)";

filename event " c:/SugiPapers/life_event.cgm";
filename censor "c:/SugiPapers/life_censo.cgm";
filename hilo   "c:/SugiPapers/life_hilo.cgm";

axis1
  c = black value = (h = 2.2) minor = none
  order = (0 to 1 by 0.1) width = 2
  label=(angle=90 h=2.4 'Kaplan-Meier Estimate');

axis2
   c = black value = (h = 2.2) minor = none
   order = (0 to 80 by 10) width = 2
   label = (h=2.4 "Survival Time (Months)");

legend value = (height = 1.8) cborder  = black
       label    = none mode     = share
       position = (TOP inside RIGHT);

goptions
   reset=symbol rotate=landscape device=cgmmwwc
   gsfname=event ctext=black cback=white
   gsfmode=replace ftext=swissb gsflen=80
   display autofeed gunit=pct;

proc gplot data = s_est gout = work.mygraf;
  note height = 1.3 j = left font = swissbu
  move = (64, 76) "&groupf1"
  move = (50, 73) "No of Patients"
  move = (50, 71) "Death"
  move = (50, 69) "Censored"
  move = (50, 67) "Median Survival"
  move = (50, 65) "95% CI"
  move = (50, 62) "Log Rank P-Value : &logrp"
  move = (66, 73) "&ng1"
  move = (66, 71) "&ngevt1 (&pgevt1.%)"
  move = (66, 69) "&ngcen1 (&pgcen1.%)"
  move = (66, 67) "&gmed1 Mos."
  move = (66, 65)  "[&llmit1, &ulmit1]"
  move = (78, 76) "&groupf2"
  move = (80, 73)  "&ng2"
  move = (80, 71)  "&ngevt2 (&pgevt2..%)"
  move = (80, 69)  "&ngcen2 (&pgcen2..%)"
  move = (80, 67)  "&gmed2 Mos."
  move = (80, 65)  "[&llmit2, &ulmit2]";

  symbol1 V = none I = steplj L = 1 c = black
        H = 1.5 W = 3.5;
  symbol2 V = none I = steplj L = 2 c = black
        H = 1.5 W = 3.5;

  plot survival*surv = trt
        / name = 'g1' haxis=axis2
```

```
          vaxis=axis1 legend = legend;
          format survival 8.1 surv 8.;
run;

goptions reset = symbol gsfname = censor
         gsfmode = replace;

proc gplot data = s_est gout = work.mygraf;
   symbol1 V = dot I = none c = black
          H = 1.5 W = 3.5;
   symbol2 V = circle I = none c = black
          H = 1.5 W = 3.5;
   plot surviva*surv = trt
        / name = 'g2' haxis=axis2
          vaxis=axis1 legend = legend;
          format surviva 8.1 surv 8.;
run;

goptions reset=symbol gsfmode = replace
         gsfname=hilo;

proc gplot data = hilo gout = work.mygraf
          annotate = anno;
   symbol1 V = none I = hilot c = blue
          H = 1.5  W = 1;
   symbol2 V = none I = hilot c = red
          H = 1.5  W = 1;
   plot survival*surv = trt
           / name = 'g3' haxis=axis2
             vaxis=axis1 legend = legend;
             format survival 8.1 surv 8.;
run;
```

Overlay the Survival Curves

The Proc Greplay is used to overlay the three survival curves together to illustrate the survival plot with overall survival analysis results, standard errors and number of patients at risk at selected time points.

```
filename wordfile "c:/SugiPapers/Sugi2003.cgm";

goptions reset = symbol gsfname = wordfile
         gsfmode = replace;

proc greplay  igout = work.mygraf
              tc  = work.newtwo nofs;
  tdef newtwo  des = 'Survival Curves'
     1/llx = 0   lly = 0   ulx = 0   uly = 100
       urx = 100 ury = 100 lrx = 100 lry = 0;
       template newtwo;
       treplay 1:g1 1:g2 1:g3;
run;
```
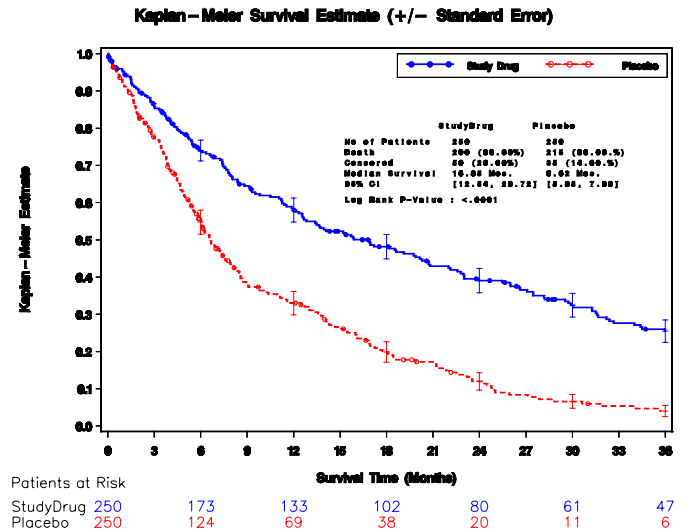
```
proc catalog cat = work.mygraf;
   delete g1 g2 g3/entrytype = grseg;
run;
```


CONCLUSIONS

Since survival graphs without summary statistics and indication of statistical uncertainty could be less informative and misleading, it is essential to present graphs with those elements. In this paper we successfully generated the survival graphs along with overall summary statistics and some statistical uncertainty. With Lifetest, Gplot, Greplay procedures and annotation techniques, we can create accurate survival presentations and help readers to draw correct conclusions.

REFERENCES

SAS Institute Inc. (1990), SAS/GRAPH Software: Reference, Version 6, First Edition, Volume 1, Volume 2, Cary, NC: SAS Institute Inc.

SAS Institute Inc. (1996), SAS/STAT Software: Changes and Enhancements through Release 6.11, Cary, NC: SAS Institute Inc.

S. Pocock, T. Clayton and D. Altman, Survival Plots of time-to-event

outcomes in clinical trials: good
practice and pitfalls, The Lancet 2002;
359:1686-1689.

J. Zhou, Graphical Presentation of
Survival Analysis, SUGI 27:Paper 321-
27.

SAS is registered trademark of SAS
Institute Inc. in the USA and other
countries.

AUTHOR INFORMATION

Joanne C. Zhou
GlaxoSmithKline Pharmaceuticals
1250 South Collegeville Road
PO Box 5089
Collegeville, PA. 19426-0989
Work Phone: (610) 917-6962
Fax: (610) 917-4107
Email: Joanne_c_zhou@gsk.com