Paper 156-30

A SAS® Macro to Generate Enhanced Kaplan-Meier Plots

Heidi Christ-Schmidt and Matt Downs
Statistics Collaborative, Inc., Washington DC

ABSTRACT

Many clinical trials have time-to-event endpoints, and Kaplan-Meier estimates of time free of an event are commonly used to summarize these endpoints. Although plots of the Kaplan-Meier estimates are useful and easily generated through Proc Lifetest, reviewers often require additional information, such as the number of patients at risk or log rank statistic, to better interpret the Kaplan-Meier plots. In some cases the reviewers are only interested in estimates up to a specific timepoint of interest.

This paper presents a macro that allows the programmer to easily modify a Kaplan-Meier plot to meet most needs of clinical and statistical reviewers. In addition to allowing the programmer to add a variety of additional information to the figure, it also gives the programmer more control over the figure's appearance. Further, it provides flexibility in the structure of the input dataset needed to generate these enhanced Kaplan-Meier figures.

TYPICAL TIME-TO-EVENT DATASET

The input dataset has one record per subject that contains the following variables:

subject identifier	numeric or character variable
time to event or censoring	numeric
censoring/event flag	numeric
group variable	numeric or character variable, w/ or w/o format

Consider the following example dataset:

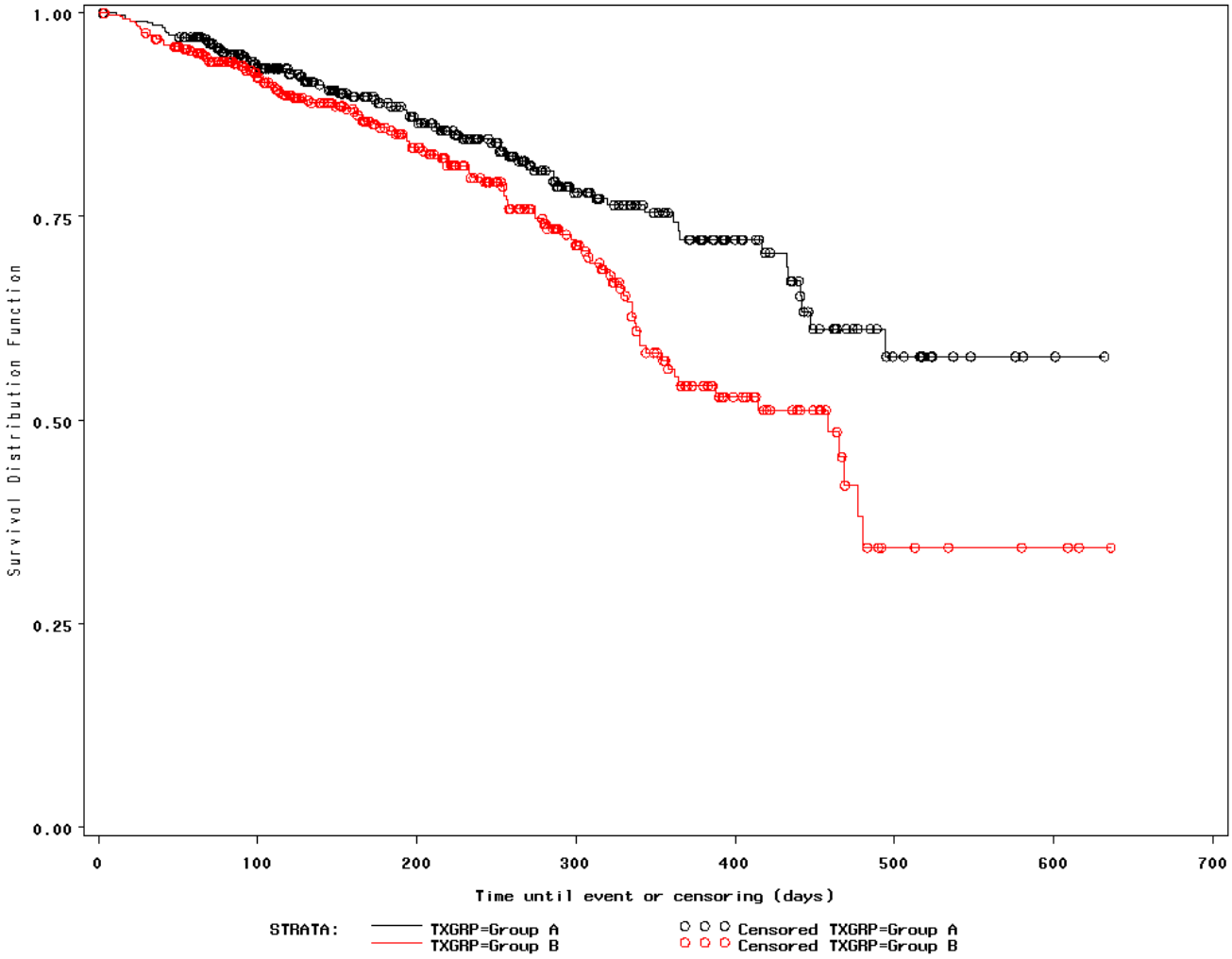
#	Variable	Type	Len	Pos	Format	Label
1	PATNUM	Num	8	0		Patient number
2	SURVCEN	Num	8	8		Flag: Patient censored? (1=cens 0=event)
3	SURVTIME	Num	8	16		Time until event or censoring (days)
4	TXGRP	Num	8	24	MASKF.	Treatment group

WHAT IS AVAILABLE DIRECTLY FROM PROC LIFETEST

To create a quick survival plot you simply need to use the following generic Proc Lifetest code.

```
proc lifetest  data    = SURVIVAL
               plots   = (s);
   time        SURVTIME*SURVCEN(1);
   strata      TXGRP;
   id          PATNUM;
run;
```

Figure 1. Example of standard Proc Lifetest figure



HOW TO IMPROVE ON THIS

This figure would be improved by the following enhancements:

1. Modify the y-axis so that it only displays the survival estimates between a specified range and add horizontal reference lines.
2. Control whether and how the censorings appear on the figure.
3. Supplement the survival curves with the number of subjects at risk at time points of interest.
4. Format the x-axis so that only times of interest are displayed in the format desired. That is, allow the figure to display x-axis ticks that are not evenly displayed and in a unit other than the original variable.
5. Annotate the figure with additional information, such as a statistic and p-value.
6. Allow for the possibility of displaying estimates of failure instead survival.

Figure 2. Example 1 using the KAPMEIER macro

```
%KAPMEIER ( DATA=           SURVIVAL,
            PATID=          PATNUM,
            GROUP=          TXGRP,
            TIME=           SURVTIME,
            CENSOR=         SURVCEN(1),
            CGMFILE=        %str(c:\new.cgm),
            Y_AXIS_ORDER=   %str(order = (0.3 to 1 by 0.1)),
            MAXTIME=        540,
            PRINT_REF=      %str(grey, 35),
            PRINT_STAT=     %str(Chisq),
            X_FORMAT=       MONTIMEF.
          );
```

Figure 3. Output of Example 1

Log-rank statistic 8.58 (p= 0.003)

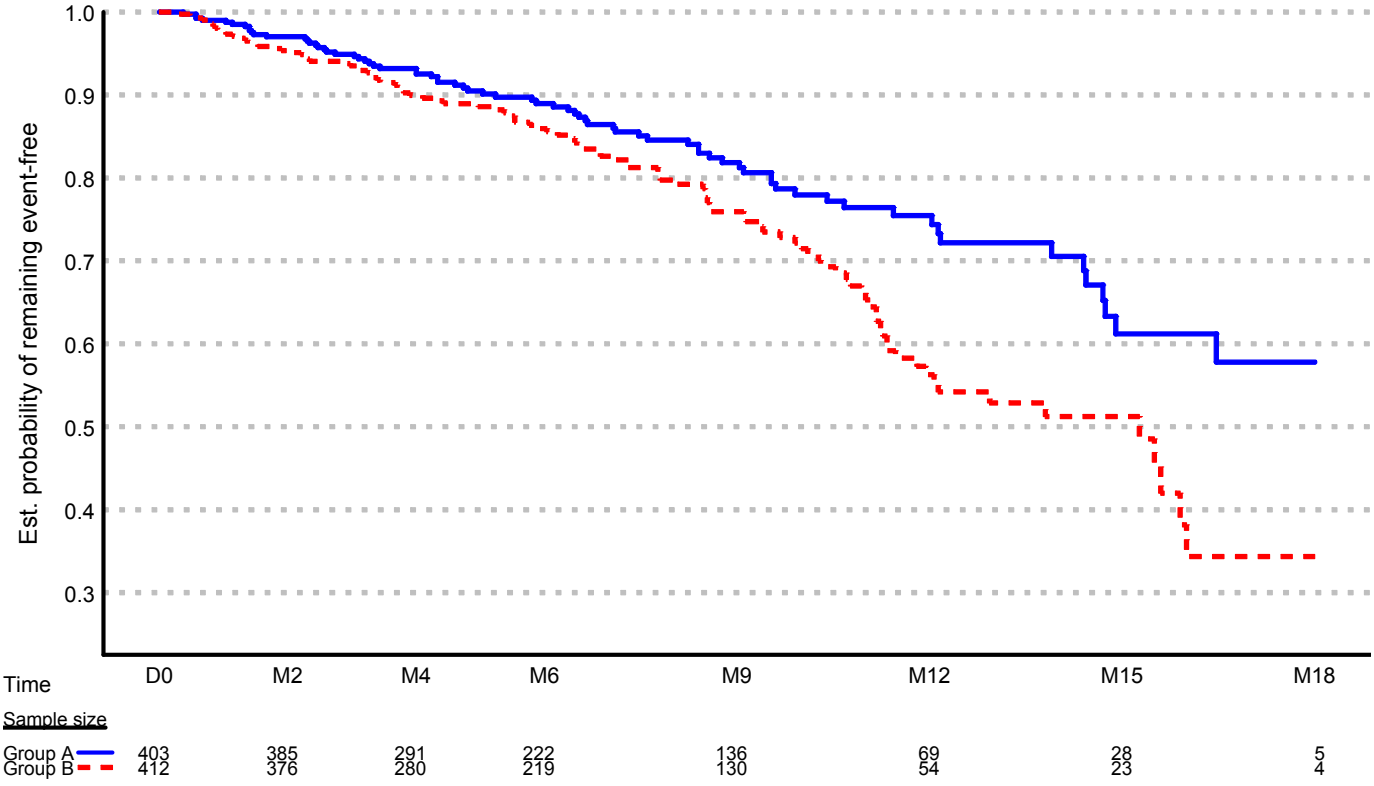
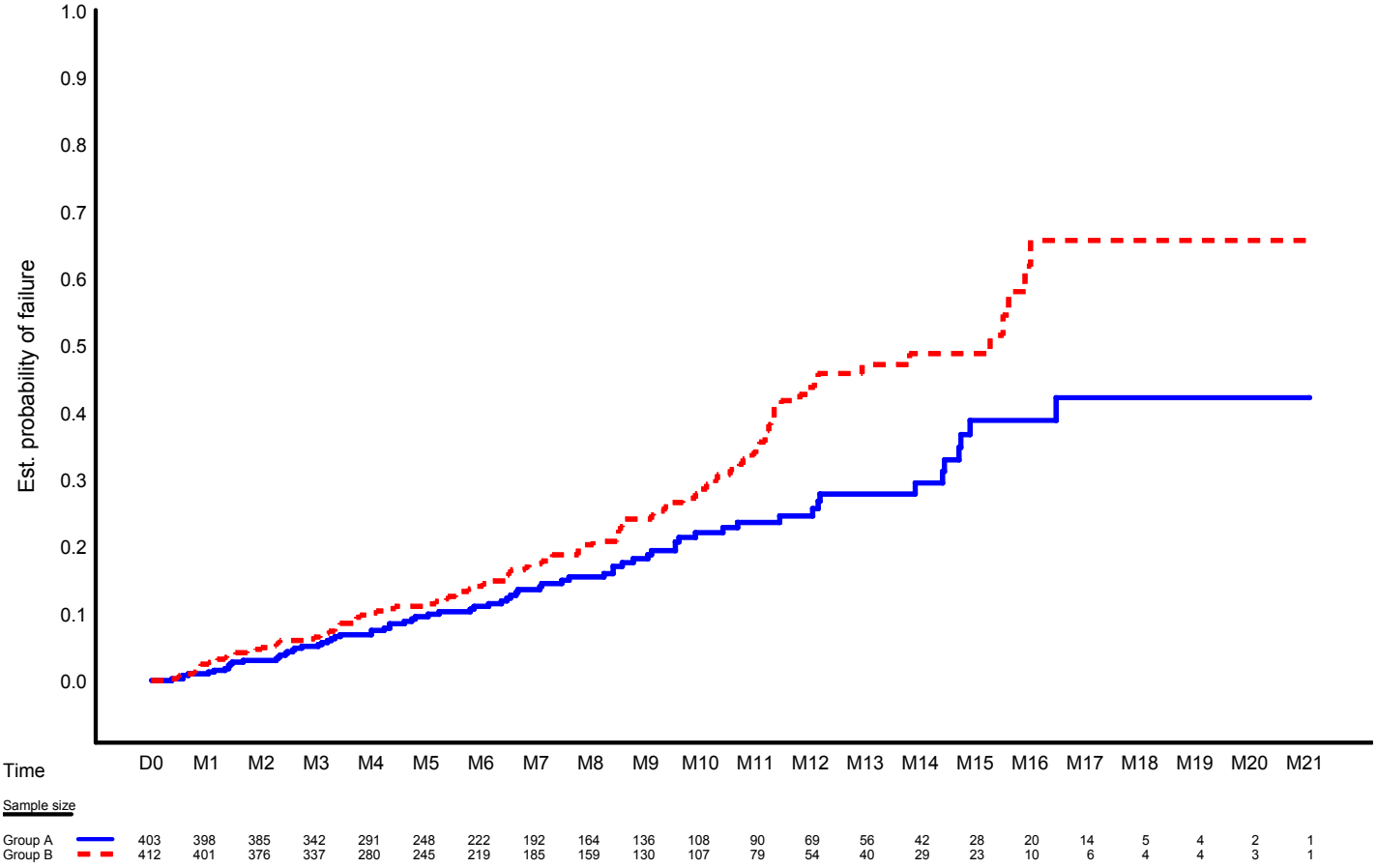


Figure 4. Example 2 using the KAPMEIER macro

```
%KAPMEIER ( DATA=          SURVIVAL,
            PATID=         PATNUM,
            GROUP=         TXGRP,
            TIME=          SURVTIME,
            CENSOR=        SURVCEN(1),
            CGMFILE=       %str(c:\failure.cgm),
            X_FORMAT=      MONNUMF.,
            PLOT_TYPE=     failure,
            Y_AXIS_LABEL=  %str(label = (a=90 h=0.8 j=l "Est. probability of failure"))
           );
```

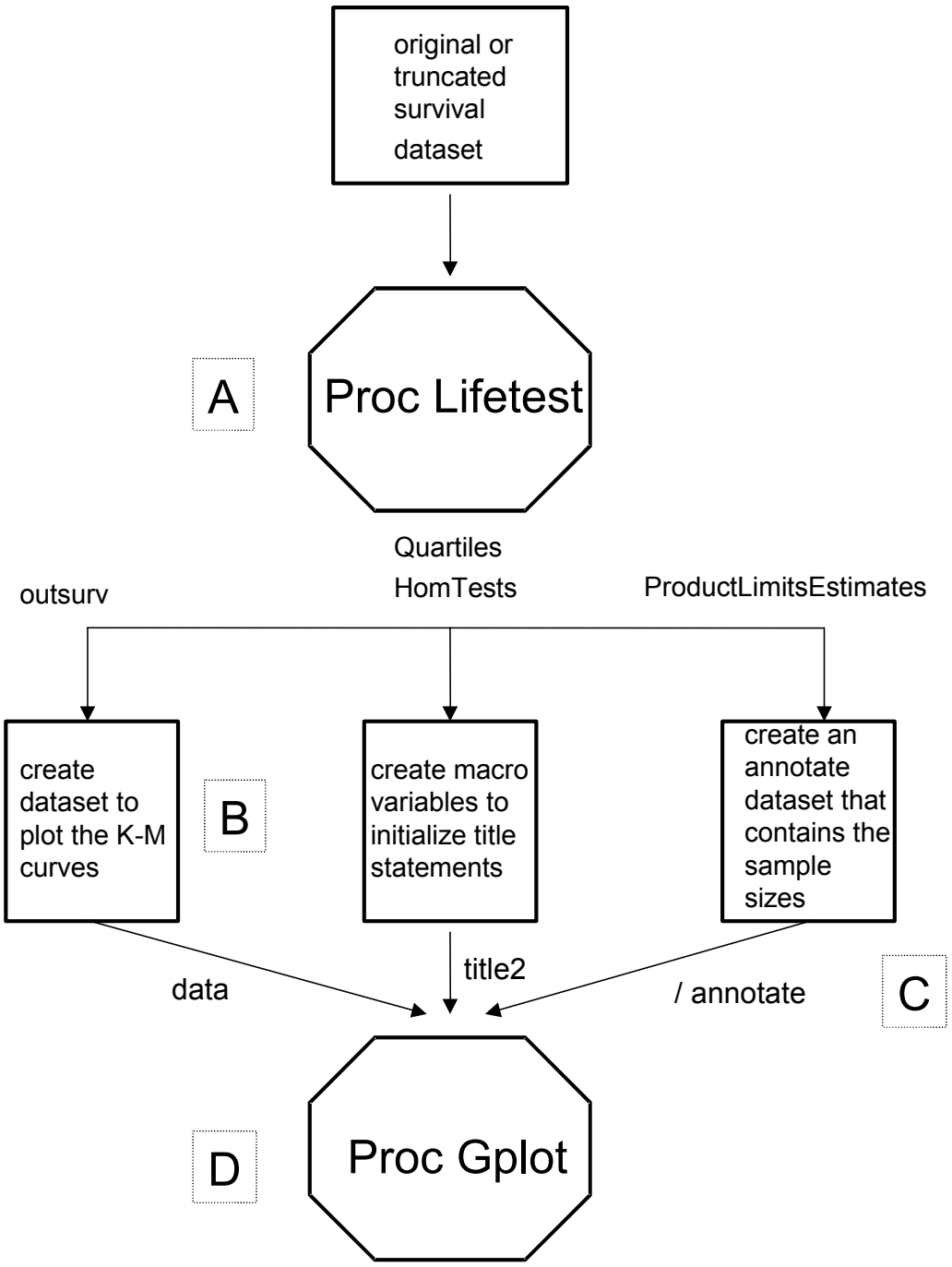
Figure 5. Output of Example 2



ORGANIZATION OF KAPMEIER MACRO CODE

The figure below shows the general organization of the KAPMEIER code. On the next few pages, we provide further details for sections of the macro code

Figure 6: KAPMEIER algorithm



ADDITIONAL DETAILS ABOUT THE KAPMEIER CODE

A. Proc Lifetest call. Note the three boxed in sections of code correspond to the three boxes in the diagram on the previous page.

```
proc lifetest  data    = KAPMEIER_TIME_TO_EVENT_TRUNCATED
               plots   = (s)
               outsurv = KAPMEIER_CRUPLT  /* Contains estimates of the survival function */
               alphaqt = &ALPHAQT;

   format      KM_GROUP;          /* Note: Because of a bug in SAS 8.2 with user-defined     */
                                  /*       formats, we need to remove this temporary format  */
                                  /*       and then re-apply it after the proc lifetest call.*/
   time        &TIME.*&CENSOR.;

   strata      KM_GROUP &GROUP;   /* Note: The inclusion of both KM_GROUP and &GROUP is for  */
                                  /*       printing purposes only. */

   id          &PATID;

   ods         output Quartiles = KAPMEIER_OUTQUART   /* Contains quartiles              */
                      %if     &&NUMBER_OF_GROUPS > 1  /* If there is only one group,     */
                                                      /* KAPMEIER_OUTSTAT is not created */
                      %then   %do;
                                 HomTests  = KAPMEIER_OUTSTAT /* Contains statistics */
                              %end;
                      ProductLimitEstimates = KAPMEIER_PLE (keep    = LEFT KM_GROUP &TIME);
                                                   /* Contains survival estimates         */
                                                   /* and samples sizes at each timepoint */

run;
```

B. Creation of the dataset used to plot the Kaplan-Meier curves.

```
%sort (KAPMEIER_CRUPLT, KM_GROUP &TIME);

data  KAPMEIER_FINALPLOT;

   retain   PROB;                 /* Retain so that the line extends to last    */
                                  /* observation, regardless of whether it is a */
                                  /* censoring.                                 */

   set      KAPMEIER_CRUPLT;
   by       KM_GROUP;

   if       _CENSOR_ = 1          /* Create a new group variable for distinguishing */
   then     PLOT_GROUP = 99;      /* between censored and non-censored subjects.    */
   else     PLOT_GROUP = KM_GROUP;

   if       .Z < SURVIVAL
   then     PROB = SURVIVAL;      /* If the obs. is censored, carry forward    */
                                  /* the survival estimate carry forward the   */
                                  /* survival estimate of the previous record. */

   %if      &PLOT_TYPE = survival
   %then    %do;
```

```
                PLOT_PROB = PROB;
             %end;

   %else %if   &PLOT_TYPE = failure
   %then       %do;
                PLOT_PROB = 1 - PROB;
             %end;

   %else       %do;
                put "ERROR: You have reqeusted an invalid plot!";
                PLOT_PROB = PROB;
             %end;

   output;
                              /* Keep an additional last record for each      */
                              /* group. This will ensure that the survival    */
   if       last.KM_GROUP     /* curves will continue to the last observation */
   then     do;               /* regardless of whether it is a censoring.     */
             PLOT_GROUP = KM_GROUP;
             output;
            end;

   format   PLOT_GROUP  KM_FMT.; /* Need to reapply the group format. */

   keep     PLOT_GROUP
            &TIME
            PROB
            PLOT_PROB;

   rename   PLOT_GROUP  =  KM_GROUP;

run;
```

C. Create an annotate dataset that contains the sample sizes for each timepoint of interest using the following steps.

1. Capture from KAPMEIER_PLE for each group and timepoint the number of people left after all of the events and censorings have been accounted for (KAPMEIER_REMAINING).
2. Capture from KAPMEIER_PLE for each group the number of people with any non-missing survival time information (KAPMEIER_STARTING).
3. Combine datasets KAPMEIER_REMAINING and KAPMEIER_STARTING to determine the maximum &TIME value for each group (KAPMEIER_REMAINING_OVERALL with F_START = 1 for observations that came from KAPMEIER_STARTING).
4. Create a frame dataset KAPMEIER_TIMEFRAME that contains for each group an observation for all timepoints of interest.
5. Merge KAPMEIER_REMAINING_OVERALL and KAPMEIER_TIMEFRAME together as shown in the following KAPMEIER_ATRISK data step.

```
data     KAPMEIER_ATRISK;

   retain   STILL_REMAINING;

   merge    KAPMEIER_TIMEFRAME
            KAPMEIER_REMAINING_OVERALL;

   by       KM_GROUP
            TIME;

   if       .Z < REMAINING              /* If there were no events or censorings at a */
   then     STILL_REMAINING = REMAINING;/* specific time than the number remaining is */
                                        /* same as for the time before.               */

   ATRISK   = lag(STILL_REMAINING);     /* The number at risk at a specific time are the */
                                        /* number remaining from the time before         */

   if       not(F_START);               /* By removing these observations, we are only the  */
                                        /* &TIME = 0 records that contain (because of the   */
                                        /* lag) the number at risk at the beginning of time.*/

   if       put(TIME, &X_FORMAT) ^= " "; /* Keep only those timepoints that will be printed */
                                         /* on the graph.                                    */
run;
```

6. Create the annotate dataset as shown below. Note that you need to have invoked %annomac earlier in the program to take advantage of the annotate macros.

```
%sort (KAPMEIER_ATRISK, KM_GROUP);

data  KAPMEIER_ANNO;

   length   STYLE $8.;
   format   TEXT $60.;

   set      KAPMEIER_ATRISK
            end = LASTOBS;

   X_VALUE = TIME + 0.3
   %system(2,3);      /* Coordinate system: data values for x-axis and percentage of */
                      /* graphics area for y-axis.                                   */
```

For each of the groups to be plotted use a %label statement to contain the sample size at each timepoint of interest (see example below for a single group):

```
 %label(X_VALUE, 15.5, put(ATRISK, &COUNT_FORMAT), black, 0, 0, 0.5, hwcgm001, 4);

/* Now add the group labels, line types, and censoring key if necessary. */

   if       LASTOBS
   then     do;

               %system (3,3);    /* Coordinate system: both as % of graphics area.   */
               &CORNER_LABEL;    /* Add the corner label = default is "Time"         */

               %label(0, 19, "Sample size", black, 0, 0, 0.5, hwcgm001, 6);
               %line (0, 18, 5, 18, black, 1, 1);     /* Underline of the label */
```

For each of the groups create similar label and line statements as noted below. These will used to create a key for the line type used to plot each group's survival curve (see example below for a single group):

```
                %label(0, 15.5, put(1, KM_FMT&GROUP_LENGTH..), black, 0, 0, 0.5, hwcgm001, 6);
                %line (4.5, 15.5, 7, 15.5, 2, blue, 1);

                %if   &PRINT_CENSOR ^= 0       /* Only print this censoring key if the     */
                                               /* censorings will be included on the plot. */
                %then %do;
                         %label(0.5,  12.5, "J",               black, 0, 0, 0.5, special,  6);
                         %label(3.5,  12.5, "Censored subject", black, 0, 0, 0.5, hwcgm001, 6);
                      %end;
             end;
    run;
```

D. Proc Gplot call.

```
%sort(KAPMEIER_FINALPLOT, KM_GROUP &TIME descending PLOT_PROB);

%if      &PRINT_CENSOR = 0      /* If the censorings are to be suppressed, remove those */
%then    %do;                   /* observations from the dataset.                       */
            proc gplot    data = KAPMEIER_FINALPLOT(where=(KM_GROUP < 99));
         %end;
%else    %do;
            proc gplot    data = KAPMEIER_FINALPLOT;  /* Otherwise, use complete dataset. */
         %end;

   format      &TIME &X_FORMAT  /* By applying the &X_FORMAT we are ensuring that only the */
                                /* labels of the timepoints of interest are printed.       */
               KM_GROUP;        /* Removing the group format ensures that the groups and   */
                                /* symbol statements are correctly paired, because the     */
                                /* internal values (numeric) of the groups are used.       */

   /* Set title statements according to the programmer requests */

   plot        PLOT_PROB*&TIME = KM_GROUP / annotate = KAPMEIER_EXTRA_ANNO
                                    %if   &PRINT_REF ^= %str()/* Switch for ref. lines. */
                                    %then %do;
                                             autovref
                                             cautovref    = %scan(&PRINT_REF, 1, ',')
                                             lautovref    = %scan(&PRINT_REF, 2, ',')
                                          %end;
                                    haxis    = axis1
                                    vaxis    = axis2
                                    nolegend
                                    noframe;
   &FOOTNOTE;

run;
quit;
```

CONTACT INFORMATION

Heidi Christ-Schmidt and Matt Downs

Statistics Collaborative, Inc.

1710 Rhode Island Ave., NW Ste 200

Washington, DC 20036

Fax: (202) 429-9268

E-mail: heidi@statcollab.com; matt@statcollab.com

APPENDIX: PARAMETERS OF THE KAPMEIER MACRO

```
%macro KAPMEIER ( /* Required parameters */

DATA= ,         /* Time to event dataset */
PATID= ,        /* Patient identifier for output*/
GROUP= ,        /* Group variable (numeric or character) */
TIME= ,         /* Time variable (numeric) */
CENSOR= ,       /* Censoring expression indicating the variable and censoring value,
                   e.g., F_CENSOR(1) or F_EVENT(0)*/
CGMFILE= ,      /* Full pathname of the cgm file to be created */

/* Optional parameters */

   /* Type of plot */

PLOT_TYPE    =      survival, /* Type of plot to be generated */
                              /*    DEFAULT     = survival */
                              /*    ALTERNATIVE = failure  */
   /* Group */

GROUP_FORMAT=       %str(),   /* Format to be applied to the group */
                              /*      DEFAULT     = format already applied (%str()) */
                                     ALTERNATIVE = format name with period
                                                   (eg., GROUPF. or $GROUPF.) */

GROUP_LENGTH=       8,        /* Length of group labels used in the annotations */
                              /*      DEFAULT     = length 8 (8)
                                     ALTERNATIVE = any number between 1 and 8, greater
                                                   than 8 will not look nice  */

   /* Time and x-axis parameters */

MAXTIME=            .A,      /* Maximum time included in the figure */
                              /*       DEFAULT     = automatically determined (.A)
                                     ALTERNATIVE = any number > 0 or <= .Z, note for
                                                   extremely large numbers the program
                                                   will take a long time to run */
```

APPENDIX: PARAMETERS OF THE KAPMEIER MACRO (CONTINUED)

```
X_FORMAT=          3.0,      /* Format for the x-axis tick labels */
                             /*       DEFAULT     = 3.0 (NOT RECOMMENDED this will print
                                                     every number between 0 and the
                                                     maximum time by 0.01)
                                     ALTERNATIVE = any numeric format name with period.
                                                     It is RECOMMENDED to have the
                                                     following type structure
                                                     value    KMTIMEF      0      =  "D0"
                                                                          28      =  "W4"
                                                                          56      =  "W8"
                                                                          84      =  "W12"
                                                                          other    =  " "; */

CORNER_LABEL=      %quote(%label(0, 22.5, "Time", black, 0, 0, 0.7, hwcgm001, 6);),
                             /* Label to appear at the corner of the figure */
                             /*       DEFAULT     = "Time" as indicated above
                                     ALTERVATIVE = %label statements for the annotate
                                                     dataset*/

COUNT_FORMAT=      3.0,     /* Format for sample sizes that appear below the x-axis */
                            /*       DEFAULT     = 3.0
                                    ALTERVATIVE = any numeric format name w/ period */

   /* Probability estimates and y-axis parameters */

Y_AXIS_ORDER=      %str(order = (0 to 1 by 0.1)),
                           /* Order statement for the y-axis statement */
                           /*       DEFAULT     = y-axis goes from 0 to 1
                                   ALTERNATIVE = a string of the format
                                                   %str(order = (X to Y by Z))
                                                 or 'AUTOMATIC', which results in the
                                                     macro automatically setting the
                                                     order statement & tick marks. */

Y_AXIS_LABEL=      %str(label = (a=90 h=0.8 j=l "Est. probability of remaining event-free")),
                           /* y-axis label */
                           /*       DEFAULT     = as noted above
                                   ALTERNATIVE = any label statement in a %str(). */

   /* Quartiles */

PRINT_QUARTILE=    .N,      /* Switch for printing quartiles */
                            /*       DEFAULT     = suppressed (.N)
                                    ALTERNATIVE = 25, 50, or 75 */

QUARTILE_UNITS=    %str(), /* Time label associated with quartiles */
                           /*       DEFAULT     = blank (%str())
                                   ALTERNATIVE = a string indicating a time label

QUARTILE_FORMAT=   4.1,     /* Format for printing the quartiles estimates */
                            /*       DEFAULT     = 4.1
                                    ALTERVATIVE = any numeric format name w/ period */

ALPHAQT=           0.05,    /* Parameter to set CI for survival time quartiles */
                            /*       DEFAULT     = 95% (0.05)
                                    ALTERNATIVE = any number < 0.50 */
```

APPENDIX: PARAMETERS OF THE KAPMEIER MACRO (CONTINUED)

```
   /* Other annotation parameters */

PRINT_REF=     %str(),  /* Switch for printing horizontal reference lines at the tick marks */
                        /*       DEFAULT     = none (%str()) */
                        /*       ALTERNATIVE = a string indicating a color and line
                                                type, e.g., %str(grey, 35) */
PRINT_CENSOR=  0,       /* Switch for printing the censorings on the figure */
                        /*       DEFAULT     = suppress (0) */
                        /*       ALTERNATIVE = print (1) */

PRINT_STAT=    %str(),  /* Statistic to be printed at the top of the figure */
                        /*       DEFAULT     = none (%str()) */
                        /*       ALTERNATIVE = one of the following (no quotes)
                              Zscore =>  both the abs.value of the Log-rank and p-value
                              Chisq =>   both the Log-rank stat. and 2-sided p-value
                              Pvalue =>  only the 2-sided Log-rank p-value
                              Wilcox =>  both the Wilcoxon stat. and 2-sided p-value
                              WPvalue => only the Wilcoxon 2-sided p-value */

LIST_OF_COLORS=%str(blue, 1, red, 2, green, 3, orange, 4, violet, 5, brown, 6, black, 7, pink, 8,
                    yellow, 9, grey, 10),
                        /* List of colors and line types for the Kaplan-Meier curves */
                        /*       DEFAULT     = as noted above */
                        /*       ALTERNATIVE = a string of colors & line types */

FOOTNOTE=      %str(footnote6 h=0.55 j=l "&pfold.\&pname..sas v.&vers_num
                                    (last run: &RUNDATE, &RUNTIME): &TITLE_CGM";
                    footnote7 h=0.55 j=l " ";),
                        /* Figure footnote statement */
                        /*       DEFAULT     = as noted above */
                        /*       ALTERNATIVE = a string containing valid footnote */

ANNO=,                  /* Additional annotation dataset */
                        /*       DEFAULT     = none */
                        /*       ALTERNATIVE = an annotate dataset. To completely
                                                override the automatic
                                                annotations, you can use a string like:
                                                %str(EXTRA_ANNO(in = KEEP); if KEEP;)*/
                );
```