

The SAS KMPLLOT9 Macro

Ellen Hertzmark and Donna Spiegelman

July 13, 2011

Abstract

The %KMPLLOT9 macro makes publication-quality Kaplan-Meier curves for a whole sample or for subgroups/strata. If there are subgroups/strata, it does the log-rank test.

Keywords: SAS, macro, cumulative incidence plot, survival plot, Kaplan-Meier, number-at-risk table

Contents

1	Description	2
2	Invocation and Details	2
3	Examples	7
3.1	Example 1. Errors in call to KMPLLOT9	7
3.2	Example 2. A macro call with only the required parameters	7
3.3	Example 3. A stratified incidence graph with lots of customization	9
3.4	Example 4. A complex graph and analysis using only a subset of the data	12
3.4.1	One version	12
3.4.2	Another version	13
3.5	Example 5. Output a text file	14
3.6	Example 6. Using proc plot	14
3.7	Example 7. A plot with 5 strata	15
4	Importing the Graph into a MS/WORD Document	17
5	Warnings	18
6	Frequently asked Questions	18

6.1	Q: I ran KM PLOT9 with no STRATA parameter and the default PICTNAME, and the pictname contains _MVAR_ where the STRATA variable name should be.	18
6.2	Q: All I see is shades of gray.	18
6.3	Q: One of my curves doesn't show.	18
6.4	. Q: I am making 2 curves and want the confidence bands for both to show. Can KM PLOT9 do this?	18
7	Credits	18
8	See Also	19

1 Description

%KM PLOT9 is a SAS macro that makes publication-quality Kaplan-Meier curves. It can also accommodate covariates by running proc phreg with the baseline option. The defaults are set to conform to the requirements of JAMA (cumulative incidence, solid lines of different colors, table of number-at-risk for selected times), but the user can control color, line type, and cumulative incidence vs. survival. The user can also control font and font size, as well as whether the label of the vertical axis prints horizontally or vertically. Curves for a whole sample show the 95% confidence bounds, while those for two or more subgroups show only the point estimates. The macro also gives the output of the tests for inhomogeneity of strata when there are two or more subgroups. A text file suitable for importing to PC graphics programs can be made.

Here is an excerpt from the letter Ronna Siegel of JAMA wrote to Michelle Holmes about a Kaplan-Meier curve in one of her papers.

Please revise the Kaplan-Meier plot to show mortality rather than survival, with the y-axis extending from 0% to 25% (see Pocock et al. Lancet, 2002; 359:1686-1689) and each curve as a thin, solid line in a distinctive color. Please show the numbers at risk at intervals along the x-axis.

2 Invocation and Details

In order to run this macro, your program must know where to look for it. You can tell SAS where to look for macros by using the options

```
mautosource sasautos= <directories where macros are located>};
```

For example, an options statement might be

```
options nocenter ps=78 ls=125 replace
mautosource
sasautos=(' /usr/local/channing/sasautos',
          '/proj/nhsass/nhsas00/nhstools/sasautos');
```

This will allow you to use all the SAS read macros for the data sets
 (/proj/nhsass/nhsas00/nhstools/sasautos), as well as other public SAS macros, such as %PM,
 %INDIC3, %EXCLUDE, and %MPHREG9
 (/usr/local/channing/sasautos).

NOTE: With this and all other macros, DO NOT include optional parameters in your macro call
 unless you want to give them non-default values. For example, giving

```
strata=,
```

will override the default and cause problems for the running of the macro.

Below, any values given to the right of the “=” are the defaults.

```
=====
REQUIRED PARAMETERS
=====
```

```
data      Name of data set to use ,
          REQUIRED
time      Name of the survival time variable ,
          REQUIRED
censor    Name of the censoring variable.
          This variable should be coded so that
          0 means censored and 1 means failure.
          REQUIRED
```

```
=====
OPTIONAL PARAMETERS RELATING TO DATA, MODEL, OR OUTPUT
=====
```

```
where     A subsetting clause, in the form of a 'where' or 'if' clause.
          NOTE: Use 'eq', 'ne', 'gt', etc., rather than '=', '^=', '>'.
          OPTIONAL
strata     Name of the strata variable, if you want 2 or more curves.
          OPTIONAL
timelist  A list of times for which you want the survival table.
          This is also the list of times for which the number at risk
          will be shown.
          If you do not give a list, no survival table will be shown.
          If atrisktab (see below) is T (the default) and timelist is not
          given, the macro will change atrisktab to F and continue
          after putting out a warning message.
          OPTIONAL
```

```
=====
BASIC PARAMETERS RELATING TO THE GRAPHIC OUTPUT
=====
```

```
plot=2    The type of output you want for your graph:
          0 = no plot
          1 = proc plot
          2 = proc gplot (the default)
          4 = text file
          OPTIONAL
outplot=PS If PLOT=2, the type of format you wish to make the plot in:
```

PS = postscript (the default)
 JPEG = JPEG
 HTML = HTML
 CGM = computer graphics metafile
 PDF = PDF (Adobe)
 JPEG, PDF, and HTML can be imported into MS/WORD.
 OPTIONAL
 NOTE: Postscript (PS) does not produce different colors (except shades of gray).
 If you need to see the colors, JPEG and PDF work fine.
 You can view the JPEG file using Netscape or StarOffice.
 You can view a PDF using acroread (but it's slow).
 Otherwise, email the file to yourself as an attachment and view it on the PC.

plotdata=DATA.TIME.CENSOR.STRATA.txt
 If PLOT=4, the name of the text file to which the plotting points will be output.
 The file will be 'pipe' (|) delimited, and will be in the following order:
 STRATA TIME SURVIVAL SLCL SUCL INC ILCL IUCL,
 where 'LCL' means lower 95% confidence limit, and 'UCL' means upper 95% confidence limit.
 OPTIONAL

pictname=DATA.TIME.CENSOR.STRATA.OUTPLOT
 If PLOT=2, the name of the graphics file.
 OPTIONAL, but we strongly suggest that you use mnemonic names.

pictdirec= The directory in which the graph is to be stored.
 If you are running on the Channing system and want the graph to be in the same directory as the program, leave this parameter blank. If you want the graph somewhere else, give the full path name of the directory, INCLUDING THE FINAL SLASH (/).
 If you are running on a PC, you should give the full path name of the directory, INCLUDING THE FINAL BACKSLASH (\).

pwhich=inc Whether you want to plot survival or cumulative incidence.
 Options are surv and inc (the default).
 OPTIONAL

vlabel Label of the vertical axis.
 If you do not give a value for this parameter, the macro will check PWHICH and label the axis either
 Fraction not Failed or
 Cumulative Incidence
 Be specific. "Cumulative Incidence of Death," rather than
 "Cumulative Incidence."
 OPTIONAL

axordv0 to 1 by .1 Limits and major tick marks for the vertical axis, in the form
 LOW to HIGH by INCREMENT.
 OPTIONAL

tlabel=Time Label of the horizontal (time) axis.

In general, it is desirable to make TLABEL as specific as possible, For example, ‘‘Years since diagnosis of Breast Cancer,’’ rather than ‘‘Time (years)’’.

OPTIONAL

axordt Limits and major tick marks for the time (horizontal) axis.
If you do not give this parameter, the macro will find the highest value of TIME and make the axis
0 to MAXTIME by (maxtime/10).
This is unlikely to come out looking nice.
In general, however, you want around 10 major tick marks.
OPTIONAL

=====

OTHER GRAPH OPTIONS

=====

landscape=F Whether you want the plot to be in landscape, rather than portrait.

font=swiss Name of font to use for the graph.
Other fonts for which the macro has been tested are cent (century) and zapf.
NOTE: The font should exist in both regular and BOLD forms.
OPTIONAL

header1 Top title on graph (if any).
OPTIONAL

header2 Second title on graph (if any).
OPTIONAL

header3 Third title on graph (if any).
OPTIONAL

hsize1=2 Print size for header1.

hsize2=1.7 Print size for header2.

hsize3=1.5 Print size for header3.
You may wish to vary these print sizes depending on the font, the length of the header, and OUTPLOT (Things come out differently in JPEG from PS).

fontmult=1 a multiplier for the standard character size everywhere but the headers, to make the characters the size you want them.
Since the actual size of the characters varies depending on the font and on the output device, this allows the user to customize the output. This will require some experimentation.

atrisktab=T Whether you want a table at the bottom showing the number remaining at risk at various times during followup.
The times shown will be those given in TIMELIST (see above).
Since the table will be printed in the same colors as the graph, if you have the table and are using color, you do not need the legend. However, the titles for the lines in the table will be what you give in leglab1,

widelabel=F Whether the strata labels are excessively wide.
This is useful only if ATRISKTAB is T.
If the labels are wide (e.g. more than 13 characters) and your TIMELIST includes 0, WIDELIST=T will move the graph origin to the right, so the labels have more room to print.

href List of values of TIME for which you want vertical lines.
OPTIONAL

vref List of values of survival or incidence (depending on pwhich)
for which you want horizontal lines.
This may be helpful for estimating when survival gets
below a specified level (or cumulative incidence gets
above a specified level).
OPTIONAL

nolegend=F Whether you want to prevent the graph from having a legend.
OPTIONAL

legloc=center bottom Location of legend.
center bottom: prints at the bottom outside the graph.
other options are any combination of
left/center/right with top/middle/bottom.
if you add 'inside' the legend will be inside
the axes.
OPTIONAL

legacross=1 Number of columns in legend.
OPTIONAL

legframe=F Whether you want a frame for the legend (sometimes useful
if you put the legend inside the axes).
OPTIONAL

leglabel A name for the STRATA variable suitable for the title of the
legend.
OPTIONAL

leglab1 A description of the first level of the STRATA variable
for the legend.
OPTIONAL

.

.

.

leglab7 A description of the 7th level of the STRATA variable

color1 The color for the first survival/incidence curve.
default=black

color2 default=red

color3 default=tan

color4 default=lib (light blue)

color5 default=violet

color6 default=gold

color7 default=pink

linetype1 default=1 (solid)

.

.

linetype7 default=1 (solid)
for the legend.
If you are not using color, a reasonable list of linetypes is
1, 4, 3, 2, 35, 37, 43 .
A diagram of all the line types is shown in the SAS/Graph manual,
version 8, page 249.
OPTIONAL

linewidth=6 The width of the incidence or survival lines.
 For any given value, the actual width will depend on the value of OUTPLOT, so some experimentation may be necessary.
 Note that you want reasonably thick lines so they will be visible when the graph is shown in reduced form in a paper.

extend whether the final values of the survival (or incidence) curve(s) should be extended beyond the last event.
 default=F
 OPTIONAL

showci=T Whether to show the confidence band of the incidence or survival curve.
 NOTE: This option only applies when there is one curve. The macro automatically changes showci to F if STRATA is given.

vlabelstyle=V Whether you want the label for the vertical axis to print parallel to the axis or horizontally (H).

3 Examples

The examples below use data from the Trial of Vitamins in the Progression and Transmission of HIV-1 in Dar es Salaam, Tanzania (TOV).

3.1 Example 1. Errors in call to KMPLLOT9

%KMPLLOT9 checks that the user has passed the required parameters. In the case of the data set name, it also checks to see that the data set exists. To illustrate the error messages %KMPLLOT9 gives, we called %KMPLLOT9 with a mistake in the data set name and leaving out the required TIME and CENSOR parameters.

```
%kmplot9(data=chid);
```

The following ERROR messages were printed in both the .log and the .lst parts of the .saslog.

```

[

/udd/stleh/helpme/ruilan/bwtkmex.sas    15:03 Tuesday, November 20, 2007    1
example with errors in macro call
example 1
ERROR in MACRO call:  Data set named (chid ) does not exist
  Check your spelling.
ERROR in MACRO call:  No time variable given.
ERROR in MACRO call:  No censoring variable given
The KMPLLOT9 macro will stop.

[

```

Here is the bare-bones macro call.

As a result of this call, the macro made a incidence graph with vertical axis 0 to 1 and the default axis labels.

MVAR is the dummy strata variable when there are no strata. It is 1 for all observations.

Since no *TIMELIST* was given and the analysis is not stratified, the printed output is only the censoring summary.

```
/udd/stleh/helpme/ruilan/bwtkmex.sas    15:03 Tuesday, November 20, 2007    2
example with only required parameters
example 2
```

8

The problem here is that the default value for *ATRISKTAB* is T.

15:03 Tuesday, November 20, 2007 3

Stratum	Total	Number Failed	Percent Failed	Number Censored	Percent Censored
1	823	132	16.0389	691	83.96

Note that the output tells you the name of the corresponding graph.

The macro call is

The use of `%quote` for `HEADER2` allows us to use the comma inside a macro parameter.

9

[illegible]

```
makes graph death.bwt.ps
```

[illegible]

```

makes graph death.bwt.ps

```

[illegible]

Note that the three 'HEADER' parameters have become the titles of the output.

3.4 Example 4. A complex graph and analysis using only a subset of the data

In this example we stratify by two variables (`lowb25` and `hiv6wk`), recoded to be one categorical variable (`lbhiv`).

The code to make `lbhiv` is

```
lbhiv=1 + 2*lowb25 + hiv6wk ;
/* the 1 makes the numbers go from 1 to 4 instead of 0 to 3.
   not necessary, but it may save confusion */
/* 1 and 2 are normal birth weight, negative and positive, respectively.
   3 and 4 are low birth weight, negative and positive, respectively */
```

Because one of our exposures (HIV status at 6 weeks) is not known till 6 weeks of age, we limit the data to children who survived at least 6 weeks (here taken as 1.5 months). We do this with the *WHERE* parameter. In order to keep the marked times on the time axis at 'nice' levels, we still begin the time axis (parameter *AXORDT*) at 0. No deaths occur till 1.5 months.

3.4.1 One version

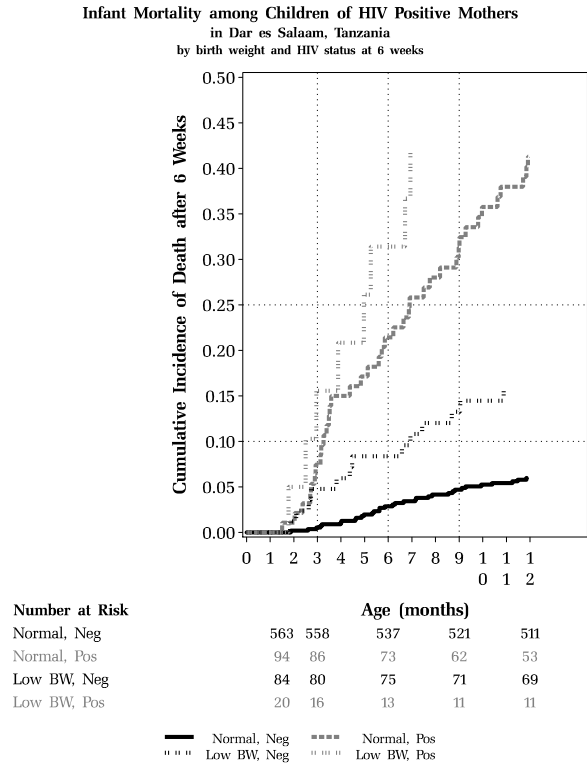
Here is one macro call.

```
%kmplot9(data=child, time=timem, censor=dead, strata=lbhiv,
where= timem ge 1.5,
timelist=1.5 3 6 9 12,
header1=Infant Mortality among Children of HIV Positive Mothers,
header2=%quote(in Dar es Salaam, Tanzania),
header3=by birth weight and HIV status at 6 weeks,
pictname=death.bwt.hiv.ps, widelabel=t,
axordv=0 to .5 by .05, vref=.1 .25,
axordt=0 to 12 by 1, href=3 6 9 ,
leglabel=Birth weight and HIV status,
leglab1=%quote(Normal, Neg), leglab2=%quote(Normal, Pos),
leglab3=%quote(Low BW, Neg), leglab4=%quote(Low BW, Pos),
legacross=2,
font=zapf, fontmult=1.1,
linewidth=4,
strname=(Birth weight, HIV status),
color2=gray, linetype2=3,
color3=black, linetype3=36,
color4=gray, linetype4=40,
atrisktab=t,
tlabel=Age (months), vlabel=Cumulative Incidence of Death after 6 Weeks);
```

NOTE: The "%quote" in *HEADER2* and *LEGLAB1*, *LEGLAB2*, *LEGLAB3*, and *LEGLAB4* allows the comma to be taken as text, rather than as a delimiter between macro parameters.

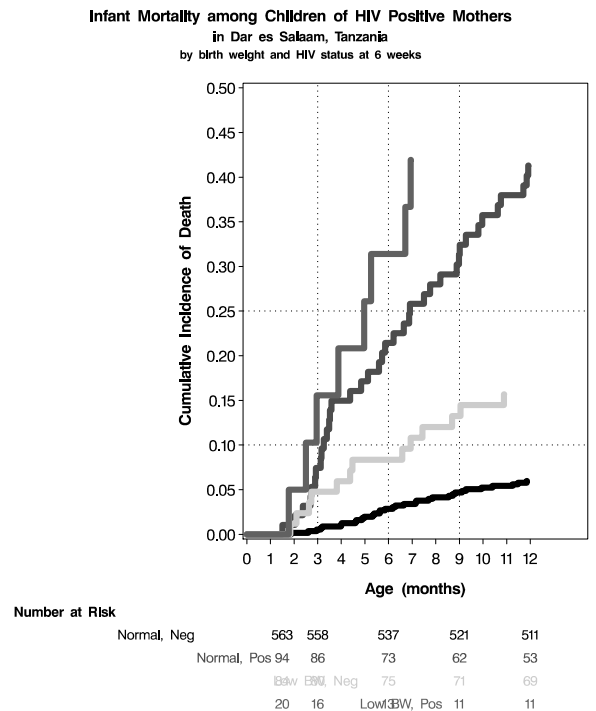
Even though this is a black-and-white graph, we can vary the color somewhat by using gray for the HIV positive strata.

The graph is



After the first 6 weeks, low birth weight only predicts survival in the HIV-negative children. This macro call also demonstrates the use of *WIDELABEL*, *VLABELSTYLE*, *FONT* and *FONTMULT*.

3.4.2 Another version



In this macro call, we have used xx.... The resulting graph is

3.5 Example 5. Output a text file

The following macro call outputs a text file.

```
title2 'example making points for plotting';
%kmpplot9(data=child, time=timem, censor=dead, strata=cd4lt350,
where=lowb25 eq 0,
timelist=0 3 6 9 12 ,
plot=4, plotdata=normbw.cd4.txt);
```

The first 10 lines of the text file are shown below.

The order of the variables is STRATA, TIME, survival, lower-95%-CL-survival, upper-95%-CL-survival, incidence, lower-and-upper-95%-CL-incidence.

0	0	1	1	1	0	0	0
0	0	0.9876237624	0.9768430711	0.9984044537	0.0123762376	0.0015955463	0.023156928
0	0.0657894737	0.985104314	0.9732741448	0.9969344832	0.014895686	0.0030655168	0.0
0	0.0986842105	0.9825848656	0.9697958232	0.995373908	0.0174151344	0.004626092	0.0
0	0.2960526316	0.9800654173	0.9663890357	0.9937417988	0.0199345827	0.0062582012	0.0
0	0.5263157895	0.9775459689	0.9630406483	0.9920512894	0.0224540311	0.0079487106	0.0
0	2.8947368421	0.9750134664	0.9597200334	0.9903068994	0.0249865336	0.0096931006	0.0
0	2.9605263158	0.9724809639	0.9564419792	0.9885199485	0.0275190361	0.0114800515	0.0
0	3.125	0.9699418491	0.9531904621	0.9866932361	0.0300581509	0.0133067639	0.04680
0	3.1578947368	0.9673960699	0.949961366	0.9848307739	0.0326039301	0.0151692261	0.0

Note: The lines have been truncated because of the page width. To read this file into excel,

```
open
file
delimited
other (specify pipe), and unmark tab .
```

3.6 Example 6. Using proc plot

Although the main purpose of the %KMPLLOT macro is to produce publication-quality graphs a user might want to do a simple proc plot just to see what axis definitions would be appropriate. Below is the graph comparable to Example 3.

[illegible]

18:04 Monday, November 17, 2003 6

Plot of inc*timem. Symbol is value of STRATUM.

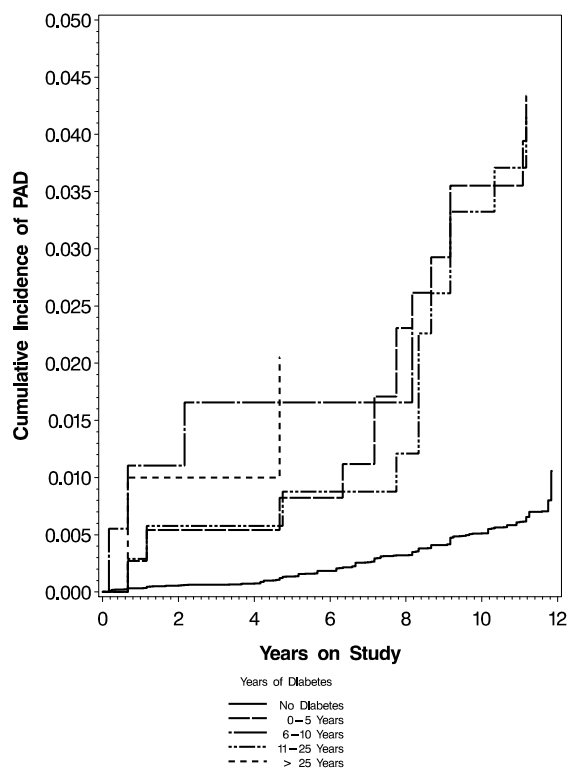
inc				
0.50 +				


```

leglab3=6-10 Years, leglab4=11-25 Years, leglab5=> 25 Years,
axordt=0 to 12 by 2, tlabel=Years on Study,
axordv=0 to .05 by .005, vlabel=Cumulative Incidence of PAD);

```

and the graph is



Since no TIMELIST was given, there is no number-at-risk table. The macro printed the following diagnostic to the .log and .lst files.

```

WARNING: Problem in MACRO call: ATRISKTAB=T, but there is no TIMELIST
The macro will continue, setting ATRISKTAB=F

```

Even though the macro call does not include the parameter ATRISKTAB, the default value is T. The printed output is

```

[

```

makes graph fordoctex1.ps

Stratum	YRDEGP	Total	Number Failed	Percent Failed	Number Censored	Percent Censored
1		45083	319	0.70758	44764	99.29
2		369	14	3.79404	355	96.21

5 Warnings

1. The %KMPLLOT9 macro will not plot values for missing STRATA. If you want to treat STRATA=. as just another value, give it a non-missing value.
2. Usually program titles (i.e. title1 'ellens work'; title2 'very silly';) carry over to graphics. To prevent unwanted titles from appearing in graphs, the macro deletes titles. We are sorry for the inconvenience. Titles from the graphs will also continue to later output, unless you restate the titles you want.

6 Frequently asked Questions

6.1 Q: I ran KMPLLOT9 with no STRATA parameter and the default PICT-NAME, and the pictname contains _MVAR_ where the STRATA variable name should be.

A: The default value of STRATA is _MVAR_, which, for the convenience of the programmer is set to 1 for every observation in the original data set.

6.2 Q: All I see is shades of gray.

A: Try JPEG to see the colors.

6.3 Q: One of my curves doesn't show.

A: If you are looking at a postscript file in ghostview, the 'tan' curve is very light colored and looks as if it is not there (though if you look closely, you may be able to see it).

6.4 . Q: I am making 2 curves and want the confidence bands for both to show. Can KMPLLOT9 do this?

A: Not directly. What you can do is make each of the curves separately using plot=4, then combine and do the proc gplot yourself. Unless the curves are very different, we do not recommend showing both confidence bands.

7 Credits

Written by Ellen Hertzmark and Donna Spiegelman for the Channing Laboratory. Thanks to Rong Chen for the code that produces the table of number still at risk. Questions can be directed to Ellen Hertzmark,
stleh@channing.harvard.edu, (617) 432-4597 or 432-1200.

8 See Also

Other Channing graphics macros are %LGTPHCURV8, %MIXCURV8, %GLMCURV8.