

How to create an S-Plus graph from SAS

Ronan Knab, Novartis Pharma AG, Basel, Switzerland

ABSTRACT

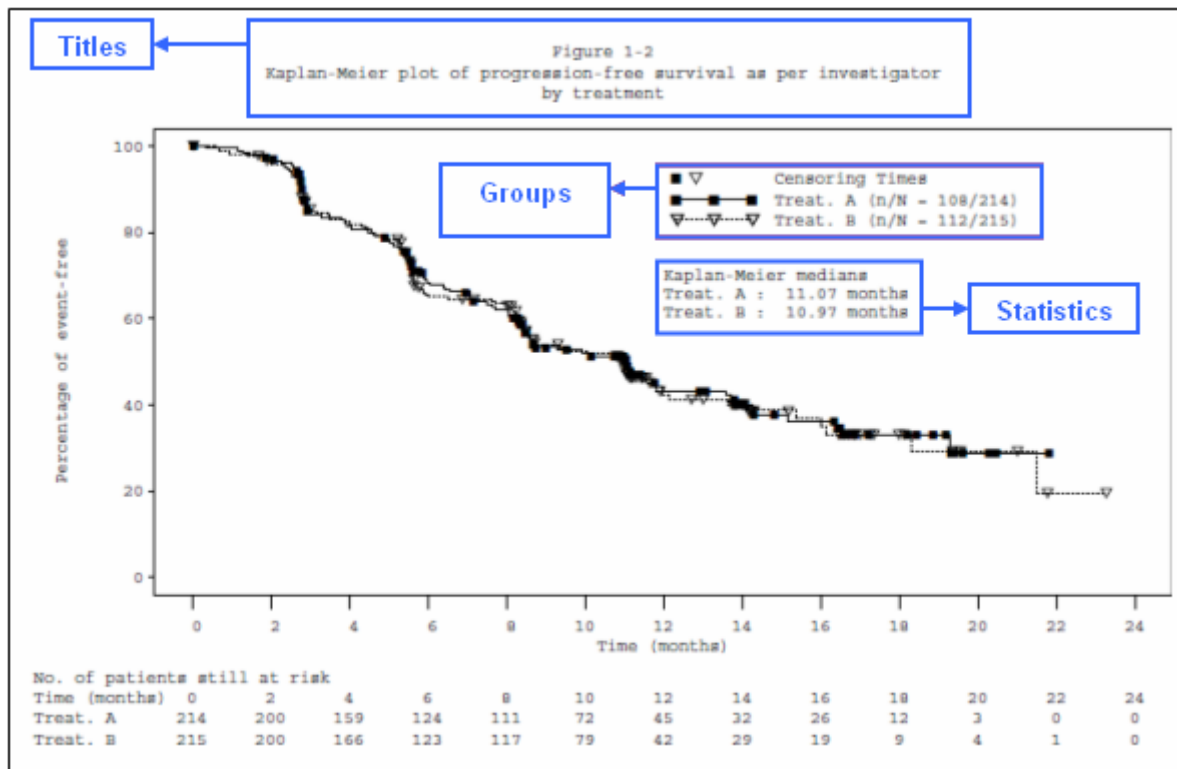
Figures are now more and more required by clinicians. Therefore, it could be useful to standardize the shells in order to save time for their creation. S-Plus® can be used to create the most frequently used graphs: Kaplan-Meier, Waterfalls or Scatter plots for example. However, as this language has not been mastered by all SAS® programmers, a programming environment has been set up in order to facilitate the graphs creation. The intention of this paper is to describe how the programmer doesn't need to use S-Plus, but only a calling program executed by SAS.

INTRODUCTION

The use of S-plus for graph creation has been decided upon by the clinician team a few years ago. SAS graphs were indeed difficult to program and the final results were not as nice as the S-Plus graphs. However, graph creation by S-Plus follows a complex process, no less than four S-Plus and two SAS files are needed to obtain the final graph. However, only one SAS program is needed to be created by the programmer, the rest of the process being indeed entirely automated. Therefore, the programmer does not have to master the S-Plus language because S-Plus is run only in the background.

OVERVIEW

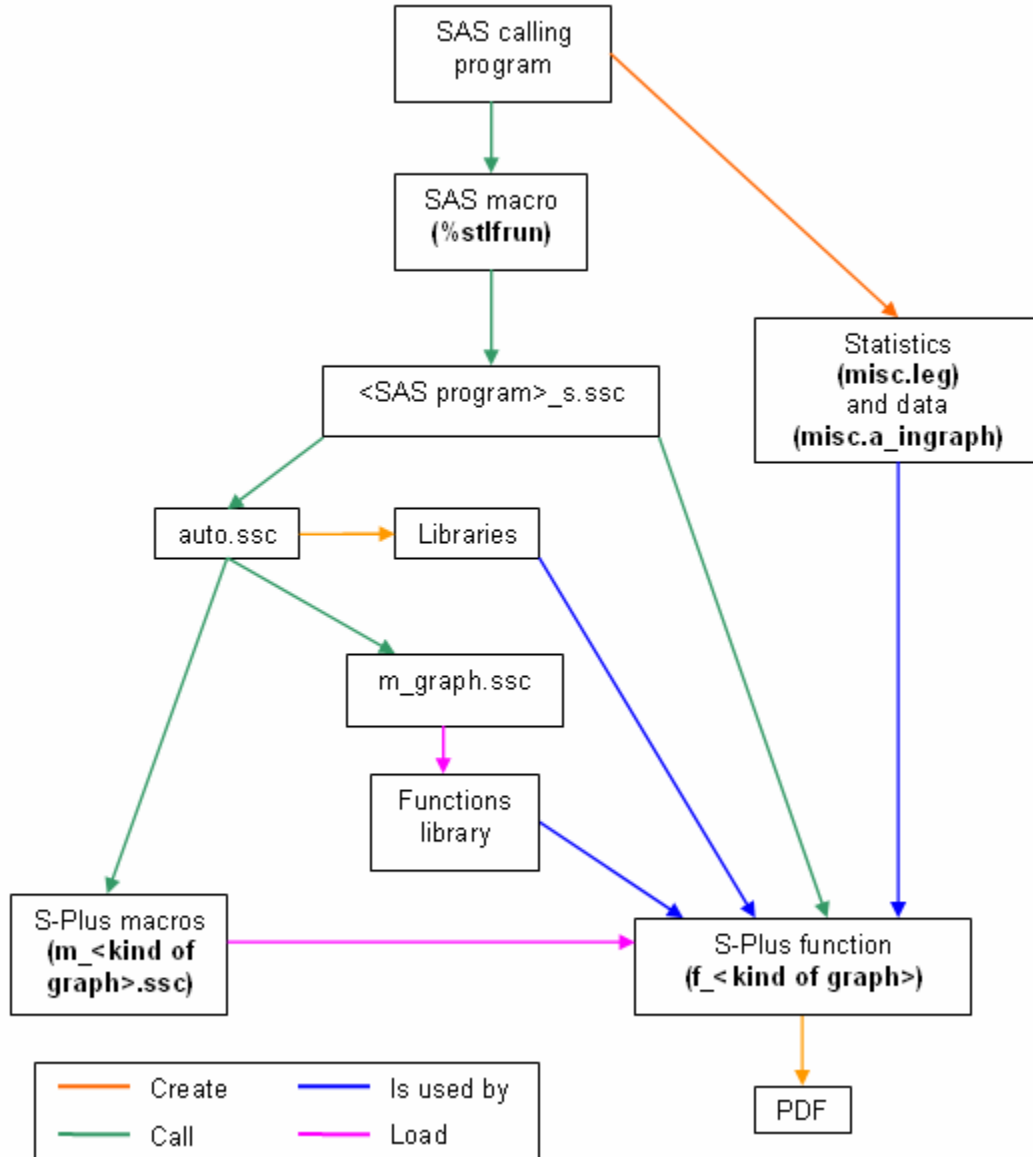
The graph below is an example of a Kaplan-Meier plot created by S-Plus:



However, the graphs are not produced entirely by S-Plus. The key point is that a SAS calling program is needed to obtain the data used for the graph and to calculate the statistics (here the Kaplan-Meier medians). Then the SAS

PhUSE 2009

program runs S-Plus via a SAS macro called **%STLFRUN**. The entire process is summarized on the diagram below and will be explained throughout this paper.



THE S-PLUS MACROS

One S-Plus macro has been created for each kind of graph following a naming convention: `m_<kind of graph>.ssc`. That means the macro creating the Kaplan-Meier plots is named `m_kapl.ssc`, `m_scat.ssc` creates the Scatter plots and so on. Inside the file, there is only one S-Plus function also following the naming convention: `f_<kind of graph>`. There are some default parameters which could be changed in the SAS calling program, as mentioned in the section **'Set the parameters for the %STLFRUN'**.

```

#Kaplan Meier function
f_kapl <- function(
    #Min and Max of X axis
    hglox      = c(0,30) ,
    #Number of interval on X axis
    nbinterx   = 3,
    #Label of Y axis

```

PhUSE 2009

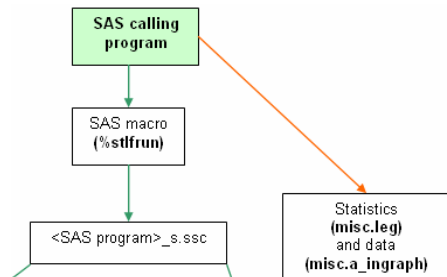
```

labely      =      "Probability (%)",
#Label of X axis
labelx      =      "Time (months)" ,
#Y coordinate of legend
ystart1     =      0,
#X coordinate of legend
xstart1     =      0,
#Groups to use
grpSORT     =      c("A","B","C"),
#Group labels
grptext     =      c("Treat. A","Treat. B","Third")
)

{
#Import data & statistics
#Plot the graph and export it into a PDF file
}

```

THE SAS CALLING PROGRAM



GET THE DATA

The SAS calling program is used to create a dataset of standard structure which will be used by S-Plus. For example, all Kaplan-Meier plot SAS programs will begin by creating a frame dataset called **a_ingraph**:

```

proc sql;
  create table misc.a_ingraph(
    stysid1a   Char(28)   Format=$28.   Label="Study + SID",
    group      Char(9)    Format=$9.    Label="Group",
    strata      Num(8)     Format=8.     Label="Strata",
    time_cs     Num(8)     Format=8.     Label="Censored?",
    time_1n     Num(8)     Format=8.     Label="Time",
    pop         Num(8)     Format=8.     Label="Population"
  );
quit;

```

Hereafter a short description of these variables:

Stysid1a	Unique patient ID
Group	Groups variable (e.g. Treatment variable)
Strata	Stratification variable (if needed)
Time_cs	Is the patient censored? (Yes / No)
Time_1n	Study days from randomization to event (Disease progression, adverse event...)
Pop	Does the patient have to be included in the analysis? (Yes / No)

This standardization is needed in order to minimize the number of parameters for the S-Plus macro call. Indeed, if the variable names were kept as they were, we would have to specify in the macro call which variable is the time variable etc. The dataset should not be created in the SAS work library, but has to be stored in a folder where S-Plus will be able to access (here the misc library).

Then, following the graph given in the above example, the programmer has to import the Progression-free survival data and put this in the correct data structure and finally set the two datasets together:

```

proc append base=misc.a_ingraph data=<data in the correct shape>;
run;

```

The use of a proc append is strongly recommended to ensure the correct structure of the data before S-Plus uses this.

CALCULATE THE STATISTICS

The statistics are calculated from SAS. They are then stored in a standard dataset which will be stored in the same folder as **misc.a_ingraph**. This dataset will then be imported into S-Plus and directly displayed inside the graphical area. That is why it needs to be clean before the S-Plus execution. Here is an example of a statistics dataset:

	LEG
1	Kaplan-Meier medians
2	Treat. A : 11.07 months
3	Treat. B : 10.97 months

The reason to use SAS for the statistics calculation is that is critical to obtain the same results in the graphs and in the tables. So it has been chosen to use the same SAS macros to calculate statistics in the tables and in the graphs.

SET THE PARAMETERS FOR %STLFRUN

Some parameters are needed for a correct execution of the S-Plus macro. They will be different according to the type of graph to produce (Kaplan-Meier, Waterfall, Scatter plot...) but some of them have to be defined for all types: titles, footnotes, file name, groups label or groups to use for the graph. For example, in the graph given in the example, the parameters were defined in several SAS macro variables.

These following SAS macro variables will be converted later into S-Plus global macro variables:

```
%let rtfname = f1_02;
%let title1 = "Figure 1-2";
%let title2 = "Kaplan-Meier plot of progression-free survival as per investigator";
%let title3 = "by treatment";
```

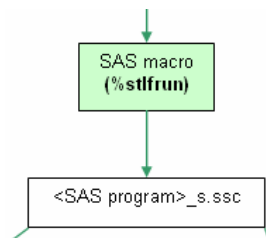
Then, some other graph-specific parameters are requested. For example the axis labels, the position of the legend... These specific parameters are defined by default in the S-Plus macros (see section '**The S-Plus macros**'), but if they have to be changed, they need to be defined in a SAS macro variable called PARA:

```
%let PARA = labely = "Percentage of event-free", /*Label of the Y axis */
,grpsort = c("A","B"), /*Groups to use*/
,grptext = c("Treat. A","Treat. B"); /*Groups labels*/
```

The use of these SAS macro variables will be defined in the section '**<SAS program name>_s.ssc: Run the graphical function**'.

Then, the macro STLFRUN needs to be called, with one parameter: the kind of graph to produce (kapl, scate...)

%STLFRUN



The SAS macro %STLFRUN has two roles: create three temporary S-Plus programs (auto.ssc, m_graph.ssc, <SAS program name>_s.ssc) and run S-Plus. They are created within a data step as mentioned in the piece of code below, which is a part of the program auto.ssc.

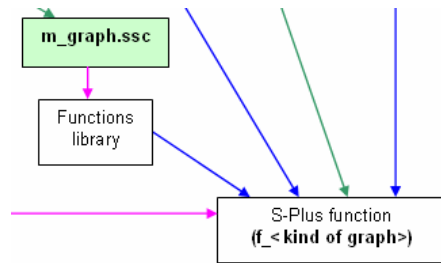
```
%MACRO stlfrun(STLTYPE=);

DATA _NULL_;
FILE "<path to the SAS calling program folder>./auto.ssc" LRECL=1024;
PUT "#####";
PUT "### FILENAME:          auto.ssc";
PUT " #Raw data";
PUT " assign(x=""data.s"",
PUT " value= paste(.root, .proj, .prot, ""report/data_s/"" , sep=""/"", frame=0) ";
PUT " print(paste(""data.s"", "" has been assigned to "" ,data.s,sep= """))";
... ..
RUN;
```

PhUSE 2009

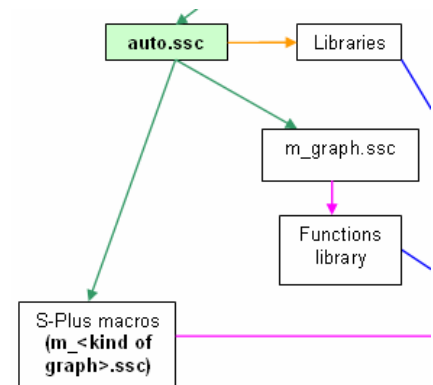
Three programs are created by this macro, each in one data step. The S-Plus code is written within the PUT statement and the program is exported into the SAS calling program folder with the FILE statement. Finally, S-Plus is called at the end of the macro (see section 'Execution')

M_GRAPH.SSC : THE FUNCTIONS LIBRARY



This file contains some standard functions which are used by the S-Plus macro. For example, a function which imports the data from the SAS dataset misc.a_ingraph to S-Plus has been created, another to define the graphical areas, import the titles and footnotes... This file is independent of the kind of graph to produce.

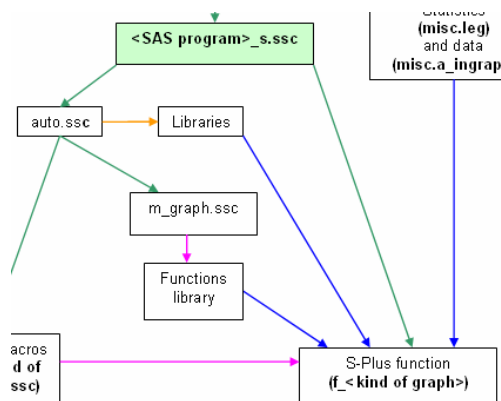
AUTO.SSC : THE S-PLUS AUTORUN



Auto.ssc creates the libraries for S-Plus. It is the equivalent of the SAS libnames. It also loads the functions library by calling m_graph.ssc as well as the S-Plus graphical function (f_kapl, f_scate...) by calling the S-Plus macro (m_kapl.ssc, m_scate.ssc...)

```
source(paste(getenv("PWD"), "m_graph.ssc" ,
sep="/"))
source(paste(.compmac, 'm_&STLTYPE..ssc' , sep="" ))
```

<SAS PROGRAM NAME>_S.SSC: RUN THE GRAPHICAL FUNCTION



This last key program is created by %STLFRUN. Here, it is needed to mix SAS and S-Plus code, in particular to create the global S-Plus macro-variables containing the titles or footnotes (see 2nd bullet point) and to run the S-Plus function with the parameters defined in the SAS macro variable PARA (see 3rd bullet point).

It has several roles :

- runs the S-Plus autorun:

```
source(paste(getenv("PWD"), "auto.ssc" ,
sep=""/"))
```

- converts the SAS macro variables for the titles and footnotes into S-Plus macro variables :

```
PUT " glob(                                     ";
%do i = 1 %to 7;
  %if %length(&&title&i) >0 %then %do;
    PUT " tit&i      = "&&title&i",             ";
  %end;
PUT " )                                           ";
```

The result of this piece of code in the S-Plus file will be :

```
glob(
  tit1      = "Figure 1-2",
  tit2      = "Kaplan-Meier plot of progression-free survival as per investigator",
  tit3      = "by treatment")
```

- runs the graphical function (f_kapl...) with the parameters defined in the SAS macro variable PARA (see section **'Set the parameters for the %STLFRUN'**).

```
PUT " f_&STLTYPE.(                                     ";
%if %length(&PARA.) >0 %then %do;
  PUT " &PARA.                                           ";
%end;
```

The result of this piece of code in the S-Plus file will be :

```
f_kapl(grpsort   = c("A", "B"),
      grptext    = c("Treat. A", "Treat. B"),
      labely     = "Percentage of event-free")
```

EXECUTION

Once all S-Plus programs have been created, <SAS program>_s.ssc is launched by the macro %STLFRUN with the following UNIX command:

```
x "Splus SBATCH -input <path/SAS program>_s.ssc -output <path/SAS program>_s.log";
```

CONCLUSION

This complex process permits the programmer to obtain S-Plus graphs without any knowledge in S-Plus. However, it needs support from S-Plus programmer: some changes on the macro could be requested by the statisticians in order to manage some particular cases. Nevertheless, once the process is in place correctly and the documentation on how to create the SAS calling programs is available, the graph creation is simple, which will save time for the programmers.

ACKNOWLEDGMENTS

Thanks to Annabelle Millischer-Foulon, Troy Dilworth and Sebastien Jolivet for their help and their review of this paper.

CONTACT INFORMATION

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

Ronan Knab
Novartis PHARMA AG
WSJ-1.3.2.09.17
Postfach
CH-4056 Basel / Switzerland
Phone : +41 61 32 41413 (direct)
ronan.knab@novartis.com:

Brand and product names are trademarks of their respective companies.