Paper 195-2010

Creating Forest Plots from Pre-computed Data using PROC SGPLOT and Graph Template Language

Zoran Bursac, PhD, University of Arkansas for Medical Sciences, Little Rock, AR

ABSTRACT

Historically, forest plots graphically display the information from the individual studies that went into the meta-analysis, and more recently the results of observational studies (e.g. epidemiology). They show the estimates (e.g. odds ratios, hazard ratios, and log transformed hazard ratios) and the amount of variation (e.g. 95% confidence intervals) and in a case of meta-analysis an overall pooled estimate. Forest plots in various forms have been published for more than 20 years, but have gained identity and popularity in the past 15 years. During this time they have been improved but it is still not easy to draw them in most standard statistical packages. In this paper we introduce a custom approach in SAS PROC SGPLOT that creates a forest plot from pre-computed data based on the logistic regression results. Further we present a dynamic graph template for a forest plot that can be applied by researchers on their data. Finally we provide a dynamic graph template for meta-analysis.

Key words: forest plot, PROC SGPLOT, dynamic graph template, meta-analysis

BACKGROUND

Forest plot is typically a graphical representation of component studies within meta-analysis, in which squares represent point estimates for each study and horizontal lines that run through the square usually represent 95% confidence intervals. In addition, the overall estimate or the pooled estimate from the meta-analysis is placed at the bottom of the figure, commonly represented as a diamond. Significance is achieved if point and pooled estimates confidence limits are clear of the reference line of no effect. [1] More recently researchers have started to use this graphical display to present regression results, for instance individual odds ratios and their confidence intervals based on logistic regression.

The origin of forest plots dates back at least to the 1970s, when Freiman et al [2] displayed the results of multiple studies with mark depicting point estimates and horizontal lines showing the confidence intervals for each study. A decade later, Lewis and Ellis [3] produced a similar plot for meta-analysis along with the pooled estimate. This type of plot was not called a “forest plot” in print for some time. However, the phrase originates from the idea that the plot appears as a forest of lines and is first used in a publication in 1996. [1,4]

In the 1980s no standard computer packages could easily produce these plots and they came from specifically produced computer programs, which is for most part the case even today. [1]

In this paper we introduce a custom approach to creating these types of plots with SAS PROC SGPLOT from the pre-computed data. We expand the idea to dynamic graph templates to add flexibility for various researchers that engage in this type of research, and finally demonstrate a similar approach for meta-analysis.

EXAMPLE 1 – PROC SGPLOT and Pre-computed Data

In the first three examples we use the data set from Hosmer and Lemeshow. [5] We model the outcome of remaining drug free for subjects being randomized to short and long duration treatment, while adjusting for other covariates. We run the logistic regression model and save the odds ratios along with their confidence intervals using ODS OUPUT statement.

```
ods output "Odds Ratios"=orci;

proc logistic data=uis descending;
model dfree=age beck ivhx ndrugtx race treat site ;
run;


data orci;set orci;
effect=upcase(effect);
run;
```

Created data set called *ORCI* that contains the pre-computed estimates we need, is shown in Table 1.

Table 1. Data set *ORCI*.

Effect	OddsRatioEst	LowerCL	UpperCL
AGE	1.049419501	1.014598	1.085436
BEC	1.000419048	0.979467	1.021819
IVHX	0.690764015	0.537	0.888322
NDRUGTX	0.940114478	0.893912	0.988705
RACE	1.258126674	0.812708	1.947664
TREAT	1.536107014	1.040959	2.26678
SITE	1.128954001	0.740448	1.721306

Our next step is to use the pre-computed data along with SAS PROC SGPLOT to create a custom forest plot.

The idea is to produce a scatter plot of the odds ratio estimates on the x-axis against the matching variables on the y-axis using the SCATTER statement. As part of the options for that statement we further assign our lower and upper confidence interval variables using XERRORLOWER and XERRORUPPER commands. We can graphically define the type and size of the symbol we wish to use for our point estimates using SYMBOL and SIZE options. Finally we create a reference line of no effect using REFLINE statement. In this case it is set to 1 since our estimates of interest are odds ratios. In other situations it can appropriately be set to 0 or something else as needed. It is important to specify a minimum of 0 for the x-axis within the XAXIS LABEL statement. Otherwise it will be automatically set based on the lowest value of all lower confidence intervals.

```
title "SGPLOT: Forest Plot";
proc sgplot data=orci;

  scatter x=oddsratioest y=effect / xerrorlower=lowercl
                                    xerrorupper=uppercl
                                    markerattrs=or
                                    (symbol=DiamondFilled size=8);
  refline 1 / axis=x;

  xaxis label="OR and 95% CI " min=0;
  yaxis label="Covariates";


run;
```

Figure 1 below depicts the results of the SAS code described above.

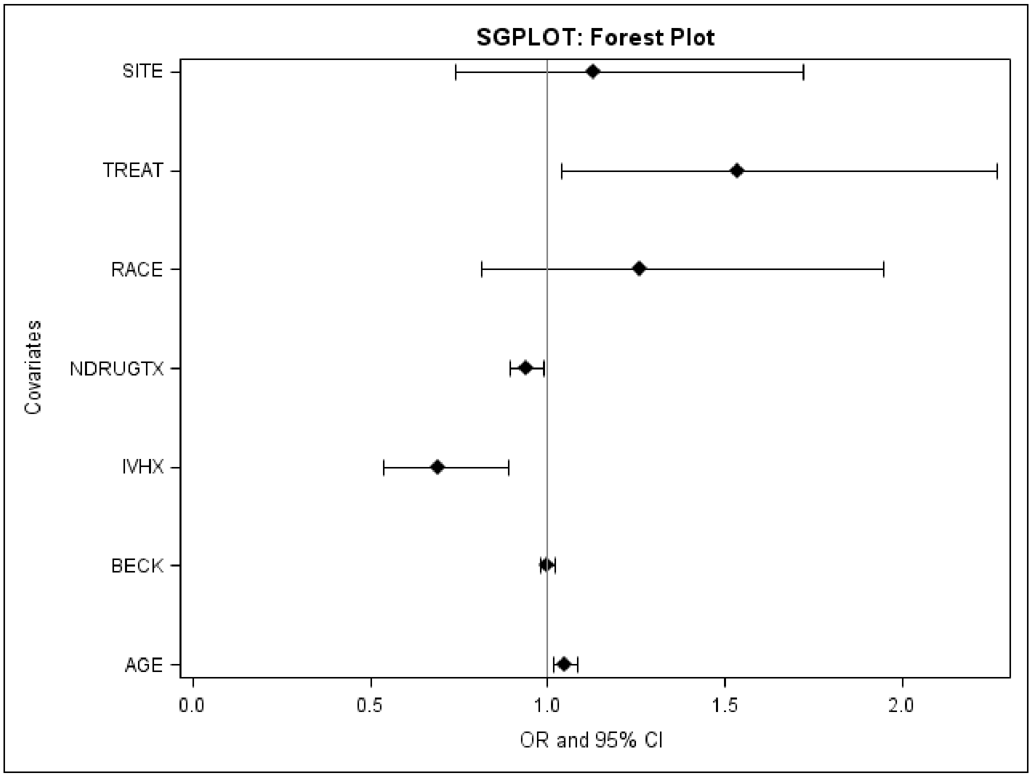

Figure 1. Forest plot of the odds ratios and their 95% confidence intervals created using SAS PROC SGPLOT and pre-computed data

EXAMPLE 2 – Graph Template Language and Pre-computed Data

We use similar approach with SAS PROC TEMPLATE and utilize graph template language in order to create identical figure to the one above. One has to be aware of the fact that graph template language and syntax used within SAS SG procedures are not identical. The settle differences in SAS can mean the difference between program running or not. For instance XAXISOPTS and YAXISOPTS are different as well as specified within LAYOUT statement and VIEWMIN option is used to create 0 minimum on x-axis. SCATTER statement is replaced with SCATTERPLOT statement, and REFERENCELINE instead of REFLINE is used to set the no effect line at 1. DEFINE STATGRAPH statement is used to assign a name to the template.

```
proc template;

define statgraph forest;

 begingraph;

 entrytitle "TEMPLATE: Forest Plot";

      layout overlay / xaxisopts=(label="OR and 95% CI"
                       linearopts=(viewmin=0))
                       yaxisoopts=(label="Covariates");

            scatterplot x=oddsratioest y=effect / xerrorlower=lowercl
                         xerrorupper=uppercl
                         markerattrs=or (symbol=DiamondFilled size=8);

            referenceline x=1 ;

      endlayout;

 endgraph;
end;
run;
```

SAS PROC TEMPLATE is used jointly with PROC SGRENDER. PROC SGRENDER invokes the code or the graphic created within given template. [6] Template example above is not a dynamic template, and therefore pre-computed variables are directly assigned, for instance *ODDSRATIOEST* variable is directly assigned to *X*-axis.

```
proc sgrender data=orci template="forest";
run;
```

EXAMPLE 3 – Dynamic Template and Pre-computed Data

In this example we generalize the template to similar situations, so that it can be applied to variety of appropriately constructed data sets. The difference here is that we create and define a set of dynamic variables using a DYNAMIC statement. One can also think of those as something similar to the macro variables. In the following steps we assign these dynamic or general variables at the appropriate place.

```
proc template;

define statgraph forest;
 dynamic VARX VARY TITLE ORLABEL VARLABEL LCL UCL SYMBOL;

  begingraph;

      entrytitle TITLE;

      layout overlay / xaxisopts=(label=ORLABEL
                       linearopts=(viewmin=0))
                       yaxisoopts=(label=VARLABEL);

            scatterplot x=VARX y=VARY / xerrorlower=LCL xerrorupper=UCL
                         markerattrs=or (symbol=SYMBOL size=8);

            referenceline x=1 ;

      endlayout;

   endgraph;
end;

run;
```

In order to be able to use the template with a specific data set we use PROC SGRENDER, and we assign our own pre-created or pre-generated variables to the dynamic variables we defined within graph template. So for example we assign *ODDSRATIOEST* to *VARX*, which is then “sent” and/or assigned to *X*-axis within the template.

```
proc sgrender data=orci template="forest";
dynamic varx="oddsratioest" vary="effect"
        title="PROC TEMPLATE: Dynamic Forest Plot"
        orlabel="OR & 95% CI" varlabel="Variables"
        lcl="lowercl" ucl="uppercl" symbol="SquareFilled";
run;
```

Dynamic template defined above produces Figure 2, in which the only difference is the symbol representing the point estimate. Within the PROC SGRENDER we assigned *SQUAREFILLED* symbol type to our dynamic variable *SYMBOL*.

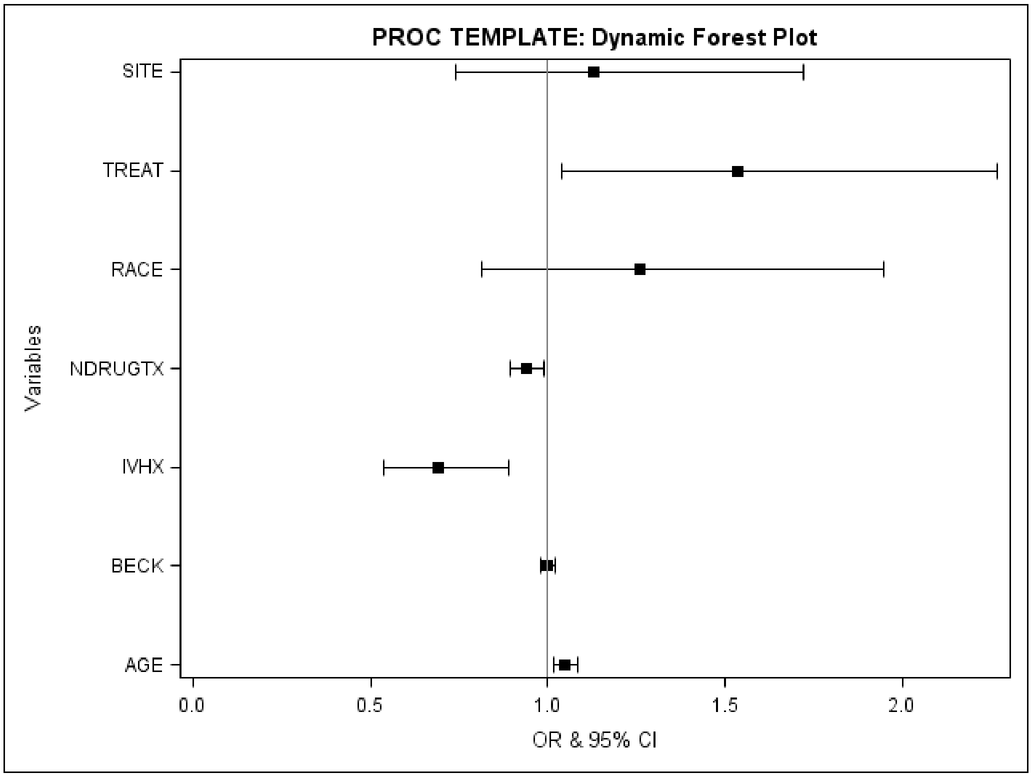

Figure 2. Forest plot of the odds ratios and their 95% confidence intervals created using dynamic template and pre-computed data.

EXAMPLE 4 – Dynamic Template for Meta-analysis

The last example is based on the hypothetical meta-analysis in which we wish to present the results of sixteen studies along with the pooled estimate. In addition to the variables we used in the previous examples we add a classification or grouping variable *C* which we use to distinguish between pooled estimate and individual odds ratios.

First we sort the data set *METAHR* presented in the Table 2 by the grouping variable *C,* and individual log hazard ratios in a descending fashion this time. Results of individual studies within meta-analysis are usually presented in sorted order so that the reader can easily visualize the effects and the confidence bands from the largest one to the smallest, or vice versa.

```
proc sort data=metahr;
by c descending loghr;
run;
```

Table 2. Data set *METAHR*.

LogHR	SELogHR	LowerCL	UpperCL	Study	C
-0.135	0.0799	-0.291604	0.021604	S1	2
-0.257	0.0734	-0.400864	-0.113136	S2	2
-0.461	0.0492	-0.557432	-0.364568	S3	2
0.203	0.0401	0.124404	0.281596	S4	2
-0.798	0.1203	-1.033788	-0.562212	S5	2
-0.324	0.0933	-0.506868	-0.141132	S6	2
-0.551	0.0577	-0.664092	-0.437908	S7	2
-0.682	0.1084	-0.894464	-0.469536	S8	2
-0.334	0.1385	-0.60546	-0.06254	S9	2
-0.384	0.0472	-0.476512	-0.291488	S10	2
0.0564	0.0671	-0.075116	0.187916	S11	2
-0.991	0.0528	-1.094488	-0.887512	S12	2
-0.723	0.0319	-0.785524	-0.660476	S13	2
-0.424	0.0289	-0.480644	-0.367356	S14	2
0.0178	0.0817	-0.142332	0.177932	S15	2
-0.187	0.0203	-0.226788	-0.147212	S16	2
-0.3712	0.08	-0.528	-0.2144	Combined	1

In addition to the previous version of the template we define two more dynamic variables, one for minimum value of the x-axis (*MIN*) and one for the reference line of no effect (*REF*).

```
 proc template;

define statgraph forestmeta;
dynamic VARX VARY TITLE ORLABEL VARLABEL MIN LCL UCL SYMBOL GROUP
        REF;

begingraph;

 entrytitle TITLE;

       layout overlay / xaxisopts=(label=ORLABEL
                        linearopts=(viewmin=MIN))
                        yaxisopts=(label=VARLABEL);

       scatterplot x=VARX y=VARY / xerrorlower=LCL xerrorupper=UCL
                                   group=GROUP ;
       referenceline x=REF ;

       endlayout;

endgraph;
end;

run;
```

Since hazard ratios were converted to the log scale the reference line of no effect becomes 0, therefore we cannot keep it 1 any longer so we add more flexibility by making it a dynamic variable. Similarly for the minimum value of the x-axis, we need added flexibility due to the fact that we are dealing with log scale and cannon settle for automatic 0 like we did with the odds ratios.

```
proc sgrender data=metahr template="forestmeta";
dynamic varx="loghr" vary="study"
title="PROC TEMPLATE: Dynamic Forest Plot for Meta-Analysis"
orlabel="Log(HR) & 95% CI" varlabel="Study"  min="-1.25"
lcl="lowercl" ucl="uppercl" group="c" ref="0";
run;
```

Arguably, it makes sense to convert either odds ratios or hazard ratios to log scale in a case of meta-analysis because log scale estimates are equidistant from 1, whereas on a linear scale they are not. Also unit change for ratio statistics does not have the same interpretation at all points on the scale, while a unit change on the log scale corresponds to multiplying the ratio statistic by the same factor. [7]

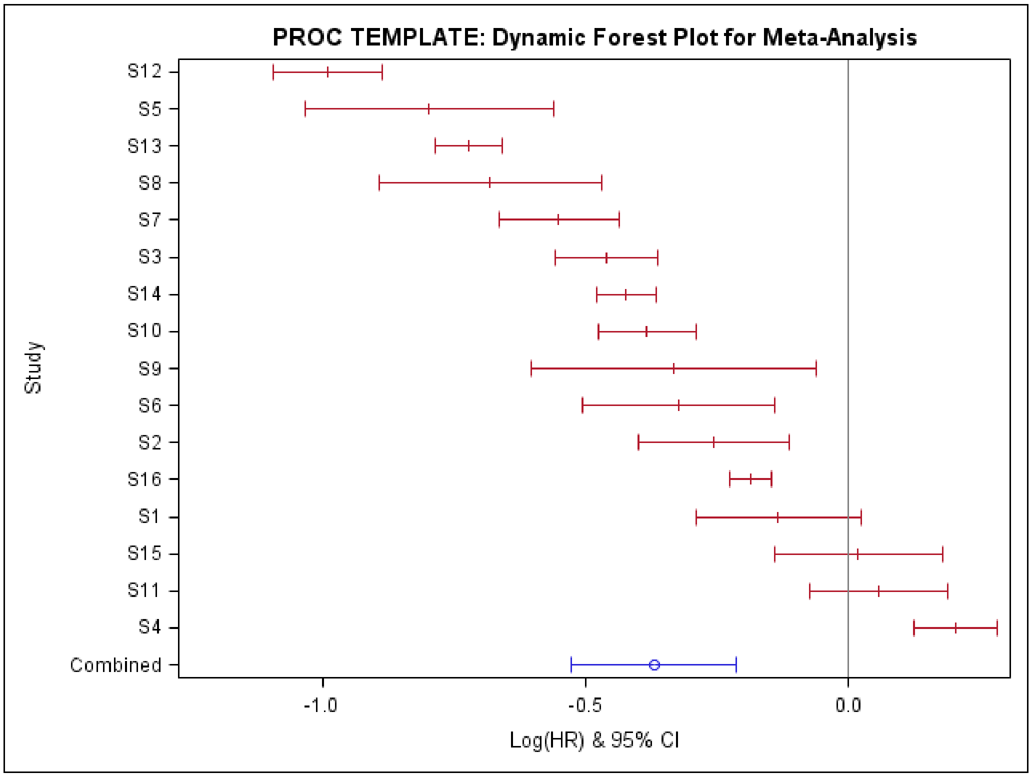

Figure 3. Forest plot of the individual log hazard ratio estimates and their 95% confidence intervals created using dynamic graph template.

REMARKS

While several specialized software packages for meta-analysis can produce forest plots for their purposes, other commonly used packages still seem to lack that capability. In this paper we offer a relatively simple and flexible alternative to produce forest plots whether it is for meta-analysis or some other analytical result, for instance various types of regression models. New SAS SG procedures have the capability of producing high quality graphics with relatively little code, compared to fairly complex coding schemes needed in the past. Graph template language further allows users to create highly customized graphics, or flexible templates for anyone to use. In our examples we demonstrate both approaches in order to create forest plots. This gives SAS users another graphical tool to better present and visualize the results of some statistical model or analysis.

REFERENCES

1. Lewis S and Clarke M. (2001). Forest plots: trying to see the woods and the trees. *BMJ*, 322: 1479-1480.
2. Freiman JA, Chalmers TC, Smith H, Kuebler RR. (1978). The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial: survey of 71 “negative trials”. *N Engl J Med*, 299: 690-694.
3. Lewis JA and Ellis SH. (1982). A statistical appraisal of post-infarction beta-blocker trials. *Prim Cardiol*, S1: 31-37.
4. Sindhu F. (1996). Are non-pharmacological nursing interventions for the management of pain effective? A meta-analysis. *J Adv Nurs*, 24: 1152-1159.
5. Hosmer DW and Lemeshow S. (2000). Applied Logistic Regression, 2nd ed. Wiley, NY.
6. SAS v 9.2 Documentation. SAS Institute Inc. Cary, NC.
7. Galbraith RF. (1988). A note on graphical presentation of estimated odds ratios from several clinical trials. *Stat in Med*, 7:889-894.

CONTACT INFORMATION

Zoran Bursac, PhD, MPH
Associate Professor
Biostatistics
Fay W. Boozman College of Public Health
University of Arkansas for Medical Sciences
4301 W. Markham, Slot 781
Little Rock, AR 72205
Work Phone: (501) 526-6723
Fax: (501) 526-6729
E-mail: zbursac@uams.edu
Web: www.uams.edu/biostat/bursac/