

## Forest Plots for Beginners

Savithri Jajam, Olesya Masucci, Covance Inc.

### ABSTRACT:

A forest plot is a common visualization for meta-analysis. The origin of forest plots were generated in at least to the 1970s by Freiman et al to display 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 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. 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.

### INTRODUCTION:

#### What is Forest plot and what does it show?

A **forest plot**, also known as a **blobbogram**, is a graphical display of estimated results from a number of scientific studies, number of end points and number of subgroups addressing the same question, along with the overall results. Although forest plots can take several forms, they are commonly presented with two columns:

- 1) The left-hand column lists the names of the studies or names of the end points or subgroups (frequently [randomized controlled trials](#) or [epidemiological studies](#) or subgroup), commonly in chronological order from the top downwards.
- 2) The right-hand column is a plot of the measure of effect (e.g. an [odds ratio](#)) for each of these studies or end points or subgroups (often represented by a square) incorporating confidence intervals represented by horizontal lines.

For each analysis label on right side, there will be a line with a box. The box represents the point effect estimate, that is, the mean effect estimate. The box size represents the weight given to the analysis, so eyes are drawn towards the studies that are given more importance. The diamond below or above the studies represents the overall effect in subgroup analysis plots. The width of the line shows the confidence intervals of the effect estimate of individual analysis. The width of the diamond shows the confidence intervals for the overall effect estimate.

The Forest plot also provides the summary data entered for each study. In addition, it provides the weight for each study; the effect measure, method and the model used to perform the meta-analysis; the confidence intervals used; the effect estimate from each study, the overall effect estimate, and the statistical significance of the analysis.

## Basic Details about Template:

There are four new SAS/GRAPH procedures for statistical graphics and each of them has their own purpose.

- 1) SGPLOT
- 2) SGPANEL
- 3) SGSCATTER
- 4) SGRENDER

Unlike the first three, the SGRENDER procedure allows you to create graphs based on templates you create. Since it uses GTL (Graph Template Language), it is the most flexible procedure. The ODS graphics statement controls the physical aspects of the graph, such as image size, name of the output, etc., and it affects all the graphs that are rendered in a SAS session until overridden by another ODS graphics statement. STATGRAPH and STYLE are two of many templates that PROC TEMPLATE supports. The STATGRAPH template defines the layout and detail of the output to be produced. It includes statements for graph layout (lattice or overlay,) plot type (scatter plots or histograms,) text elements (titles, footnotes and insets) etc. The STYLE template defines the format information for the visual SAS output.

## There are three basic steps to create a graph using GTL.

1. Use PROC TEMPLATE to define a STATGRAPH template
2. Define STYLE template
3. Create the graph by running the SGRENDER procedure with appropriate data

## Types of Forest plots:

We can create forest plots for 3 types of analysis.

- 1) Between Studies
- 2) Between Endpoints
- 3) Between Subgroups

## Dataset structure:

Dataset structure need to adjust based on what type of figure we need. For example, subgroup need leading and tailing spaces so labels can align correctly.

## EXAMPLE 1 For Subgroup analysis:

### Step 1:

Create dataset as below. Each variable details are as below.

LCL: lower CI or Min, UCL: upper CI or Max, Per : Mean, Res1: TRT A, Res2: TRT B

Res3: treatment difference

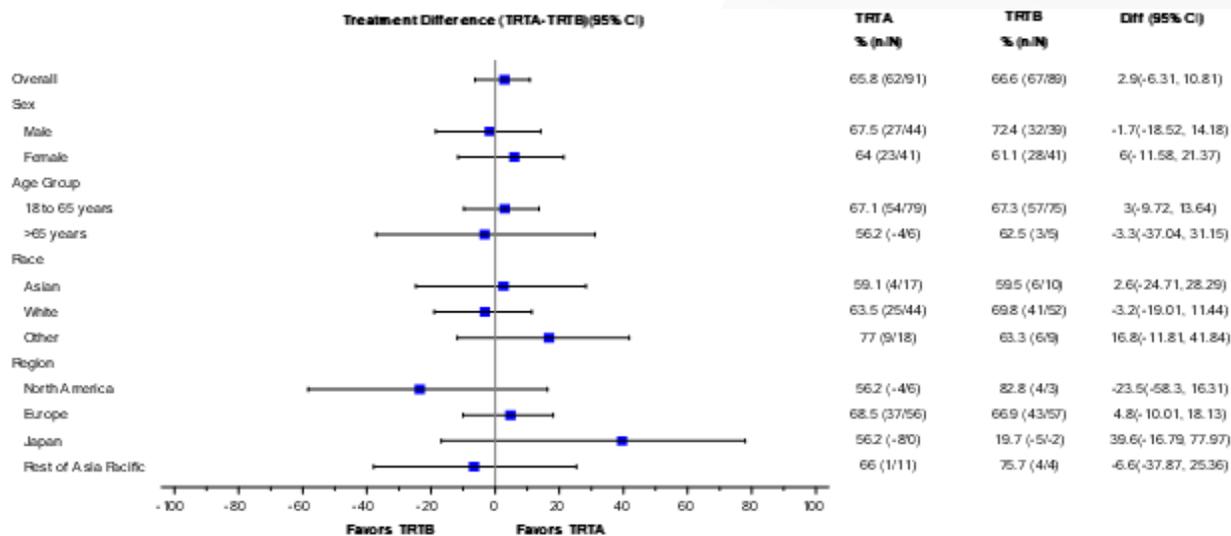
Study	res1	res2	res3	per	LCL	UCL
Overall	65.8 (62/91)	66.6 (67/89)	2.9(-6.31,10.81)	2.9	-6.31	10.81
<b>Sex</b>						
Male	67.5 (27/44)	72.4 (32/39)	-1.7(-18.52,14.18)	-1.7	-18.52	14.18
Female	64 (23/41)	61.1 (28/41)	6(-11.58,21.37)	6	-11.58	21.37
<b>Age Group</b>						
18 to 65 years	67.1 (54/79)	67.3 (57/75)	3(-9.72,13.64)	3	-9.72	13.64
>65 years	56.2 (-4/6)	62.5 (3/5)	-3.3(-37.04,31.15)	-3.3	-37.04	31.15
<b>Race</b>						
Asian	59.1 (4/17)	59.5 (6/10)	2.6(-24.71,28.29)	2.6	-24.71	28.29
White	63.5 (25/44)	69.8 (41/52)	-3.2(-19.01,11.44)	-3.2	-19.01	11.44
Other	77 (9/18)	63.3 (6/9)	16.8(-11.81, 41.84)	16.8	-11.81	41.84
<b>Region</b>						
North America	56.2 (-4/6)	82.8 (4/3)	-23.5(-58.3, 16.31)	-23.5	-58.3	16.31
Europe	68.5 (37/56)	66.9 (43/57)	4.8(-10.01, 18.13)	4.8	-10.01	18.13
Japan	56.2 (-8/0)	19.7 (-5/-2)	39.6(-16.79, 77.97)	39.6	-16.79	77.97
Rest of Asia Pacific	66 (1/11)	75.7 (4/4)	-6.6(-37.87, 25.36)	-6.6	-37.87	25.36

### Step 2:

Create variables STAT='MIN' , Value=LCL, STAT='MAX' , Value=UCL and STAT='MEAN', Value=PER from step 1. Add variable study3 with non-printable text, so first column have extra spaces in subgroups and align correctly as shown in figure 1. Final data structure is below and used in PROC TEMPLATE:

study1	Stat	Value	per	LCL	UCL	Study1	res1	res2	res3
Overall	MIN	-6.31	2.900	-6.31	10.81	Overall	65.8 (62/91)	66.8 (67/89)	2.9(-6.31, 10.81)
Overall	MAX	10.81				Sex			
Overall	Q1	-2.9	-1.70	-18.5	14.18	Male	67.5 (27/44)	72.4 (32/39)	-1.7(-18.52, 14.18)
Overall	Q3	2.9	6.000	-11.0	21.37	Female	64 (23/41)	61.1 (28/41)	6(-11.58, 21.37)
Overall	Mean	2.9				Race			
Overall	BOX/WDTH	0.15	2.600	-24.7	28.29	Asian	59.1 (4/17)	59.5 (6/10)	2.6(-24.71, 28.29)
Sex	MIN	-3.20	-19.0	11.44		White	63.5 (25/44)	69.8 (41/52)	-3.2(-19.01, 11.44)
Sex	MAX	16.80	-11.0	41.94		Other	77 (9/18)	63.3 (6/9)	16.8(-11.81, 41.94)
Sex	Q1					Region			
Sex	Q3		-23.5	-58.3	16.31	North America	56.2 (-4/6)	82.8 (4/3)	-23.5(-58.3, 16.31)
Sex	Mean	4.800	-10.0	18.13		Europe	68.5 (37/56)	66.9 (43/57)	4.8(-10.01, 18.13)
Sex	BOX/WDTH	0.15	39.60	-18.8	77.97	Japan	56.2 (-8/0)	19.7 (-5/2)	39.6(-18.79, 77.97)
Male	MIN	-18.52	-6.60	-37.9	25.36	Rest of Asia Pa	66 (1/11)	75.7 (4/4)	-6.6(-37.87, 25.36)
Male	MAX	14.18							
Male	Q1	-1.7							
Male	Q3	-1.7							
Male	Mean	-1.7							
Male	BOX/WDTH	0.15							
Female	MIN	-11.58							
Female	MAX	21.37							
Female	Q1	6							
Female	Q3	6							
Female	Mean	6							
Female	BOX/WDTH	0.15							

Figure 1:



EXAMPLE 2 Overall analysis:

Step 1:

Create dataset as below. Each variable details are as below.

LCL: lower CI or Min, UCL: upper CI or Max, Per : Mean, Res1: TRT A, Res2: TRTB

### Res3: treatment difference

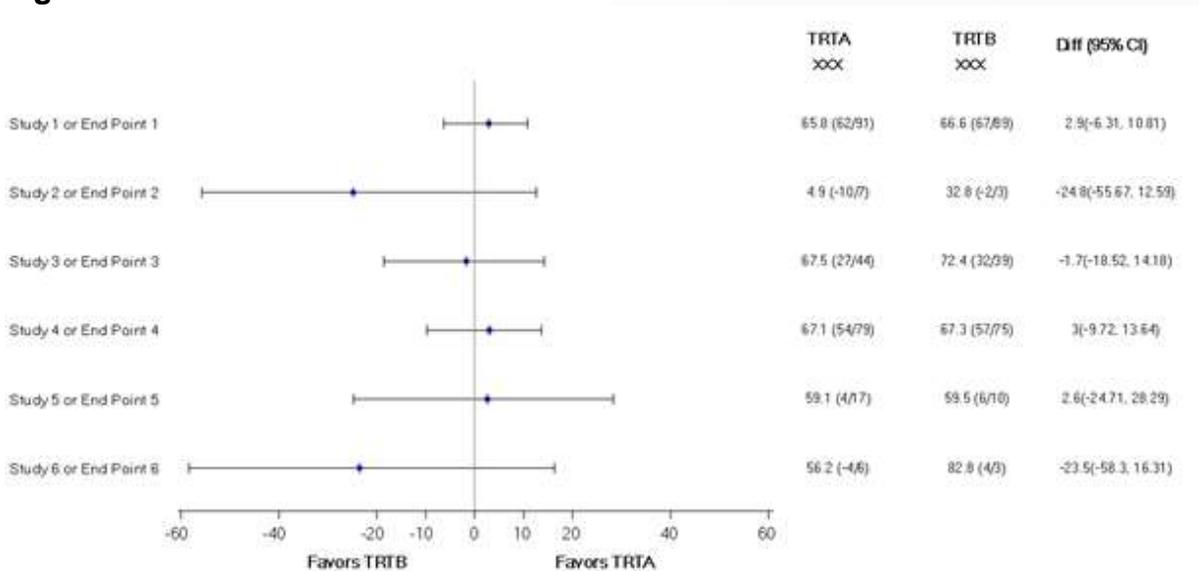
Study	LCL	UCL	per	Weight	res1	res2	res3
Study 1 or End Point 1	-6.31	10.81	2.9	1	65.8 (62/91)	66.6 (67/89)	2.9(-6.31, 10.81)
Study 2 or End Point 2	-55.67	12.59	-24.8	1	4.9 (-10/7)	32.8 (-2/3)	-24.8(-55.67, 12.59)
Study 3 or End Point 3	-18.52	14.16	-1.7	1	67.5 (27/44)	72.4 (32/39)	-1.7(-18.52, 14.16)
Study 4 or End Point 4	-9.72	13.64	3	1	67.1 (54/79)	67.3 (57/75)	3(-9.72, 13.64)
Study 5 or End Point 5	-24.71	28.29	2.6	1	59.1 (4/17)	59.5 (6/10)	2.6(-24.71, 28.29)
Study 6 or End Point 6	-58.3	16.31	-23.5	1	56.2 (-4/6)	82.8 (4/3)	-23.5(-58.3, 16.31)

### Step 2:

Create variables STAT='MIN', Value=LCL, STAT='MAX', Value=UCL and STAT='MEAN', Value=PER from step 1. Final data structure is below and used in PROC TEMPLATE.

Study	Stat	Value	LCL	UCL	per	Study2	res1	res2	res3
Study 1 or End Point 1	MIN	-6.31	-6.31	10.81	2.900	Study 1 or End Point 1	65.8 (62/91)	66.6 (67/89)	2.9(-6.31, 10.81)
Study 1 or End Point 1	MAX	10.81	-55.7	12.59	-24.8	Study 2 or End Point 2	4.9 (-10/7)	32.8 (-2/3)	-24.8(-55.67, 12.59)
Study 1 or End Point 1	Q1	2.9	-18.5	14.16	-1.70	Study 3 or End Point 3	67.5 (27/44)	72.4 (32/39)	-1.7(-18.52, 14.16)
Study 1 or End Point 1	Q3	2.9	-9.72	13.64	3.000	Study 4 or End Point 4	67.1 (54/79)	67.3 (57/75)	3(-9.72, 13.64)
Study 1 or End Point 1	MEAN	2.9	-24.7	28.29	2.600	Study 5 or End Point 5	59.1 (4/17)	59.5 (6/10)	2.6(-24.71, 28.29)
Study 1 or End Point 1	BOXWIDTH	0.15	-58.3	16.31	-23.5	Study 6 or End Point 6	56.2 (-4/6)	82.8 (4/3)	-23.5(-58.3, 16.31)
Study 2 or End Point 2	MIN	-55.67							
Study 2 or End Point 2	MAX	12.59							
Study 2 or End Point 2	Q1	-24.8							
Study 2 or End Point 2	Q3	-24.8							
Study 2 or End Point 2	MEAN	-24.8							
Study 2 or End Point 2	BOXWIDTH	0.15							
Study 3 or End Point 3	MIN	-18.52							
Study 3 or End Point 3	MAX	14.16							
Study 3 or End Point 3	Q1	-1.7							
Study 3 or End Point 3	Q3	-1.7							
Study 3 or End Point 3	MEAN	-1.7							
Study 3 or End Point 3	BOXWIDTH	0.15							

Figure 2:



## TEMPLATE AND FULL CODE:

We used proc template with layout lattice and separated into 2 columns Column weight describes which column should occupy how much space in the plot area.

## SAMPLE FULL CODE FOR BY GROUP ANALYSIS:

```
proc template;
  define statgraph ForestPlot_2Col;
  dynamic _pct;
  begingraph / designwidth=600px designheight=520px;
  layout lattice / columns=2 columngutter=0 columnweights=(.67 .33 );
  /**FIRST COLUMN**/
  layout overlay / walldisplay=(fill)
    yaxisopts=(display=(tickvalues) type=DISCRETE
  discreteopts=(tickvaluefitpolicy=THIN)
  tickvaluehalign=left reverse=true offsetmin=1 offsetmax=1 TICKVALUEATTRS=(SIZE=8pt
  family="Courier New"))
    xaxisopts=( offsetmin=1 offsetmax=1
  label="Favors TRTB          Favors TRTA"
    linearopts=(minorticks=true tickvaluelist=(- 100 -80 -60 -40 -20 0 20 40 60 80 100)
  viewmin=- 100 viewmax=100) TICKVALUEATTRS=(SIZE=8pt family="Courier New")
  labelattrs=(family="Courier New" weight=bold size=8pt));
  entry halign=left "          "
    halign=center textattrs=(color=black weight=bold size=8pt family="Courier New") "
  Treatment Difference (TRTA-TRTB)(95% CI)" / location=outside valign=top ;
  boxplotparm x=Study3 y=value stat=stat / orient=horizontal capshape=line extreme=true
  meanattrs=(symbol=squarefilled color=blue );
  referenceline x=0 / lineattrs=(pattern=solid);
  endlayout; /**END OF FIRST COLUMN**/
  /**SECOND COLUMN**/
  layout overlay / walldisplay=none border=false
    yaxisopts=(reverse=true type=discrete display=none labelattrs=(family="Courier
  New" size=8pt))
    xaxisopts=(display=none offsetmin=0.15 offsetmax=0.15);
  entry halign=left textattrs=(color=black weight=bold size=8pt family="Courier New") "  %
  (n/N) "
    halign=center textattrs=(color=black weight=bold size=8pt family="Courier New") "  %
```

```

(n/N)"
      halign=right " " / pad=(right=2.75% TOP = 0 BOTTOM = 0) location=outside valign=top;
      entry halign=left textattrs=(color=black weight=bold size=8pt family="Courier New") " TRTA"
      halign=center textattrs=(color=black weight=bold size=8pt family="Courier New") "
TRTB"
      halign=right textattrs=(color=black weight=bold size=8pt family="Courier New") " Diff
(95% CI)" / pad=(right=2.75% TOP = 0.1% BOTTOM = 0.1%) location=outside valign=top;
      scatterplot y=Study2 x=eval(constant*1) / markercharacter=res2;
      scatterplot y=Study2 x=eval(constant*2) / markercharacter=res3;
      scatterplot y=Study2 x=eval(constant*3) / markercharacter=res4;
      endlayout; /**END OF SECOND COLUMN**/

endlayout; /**END OF LATICE**/
endgraph; /**END OF GRAPH**/
end;
run;

```

### **SAMPLE FULL CODE FOR OVERALL ANALYSIS:**

```

proc template;
  define statgraph ForestPlot_2Col;
    dynamic _pct;
    begingraph / designwidth=600px designheight=520px;
    layout lattice / columns=2 columngutter=0 columnweights=(.66 .34 );

    layout overlay / walldisplay=(fill)
      yaxisopts=(display=(tickvalues) type=DISCRETE
discreteopts=(tickvaluefitpolicy=THIN)
tickvaluehalign=left reverse=true offsetmin=1 offsetmax=1 TICKVALUEATTRS=(SIZE=9Pt
family="Courier New") )
xaxisopts=(offsetmin=1 offsetmax=1
      label="Favors TRTB          Favors TRTA"

      linearopts=(minorticks=true tickvaluelist=( -20 -15 -10 -5 0 5 10 15 20) viewmin=-20
viewmax=20 ) TICKVALUEATTRS=(SIZE=10pt family="Courier New")
labelattrs=(family="Courier New" weight=bold size=10pt));

```

```

    boxplotparm x=Study y=value stat=stat / orient=horizontal capshape=line extreme=true
meanattrs=(symbol=squarefilled size=4pt color=blue );
    referenceline x=0 / lineattrs=(pattern=solid);
endlayout;

layout overlay / walldisplay=none border=false
    yaxisopts=(reverse=true type=discrete display=none labelattrs=(family="Courier
New" size=9pt))
    xaxisopts=(display=none offsetmin=0.15 offsetmax=0.15);

    entry halign=left textattrs=(color=black weight=bold size=9pt family="Courier New") "
&lab1b"
        halign=center textattrs=(color=black weight=bold size=9pt family="Courier New") "
&lab2b"

        halign=right textattrs=(color=black weight=bold size=9pt family="Courier New") " " /
pad=(right=2.75% TOP = 0 BOTTOM = 0) location=outside valign=top;

    entry halign=left textattrs=(color=black weight=bold size=9pt family="Courier New") "
TRTA "
        halign=center textattrs=(color=black weight=bold size=9pt family="Courier New") "
TRTB          Diff (95% CI)"
        halign=right textattrs=(color=black weight=bold size=9pt family="Courier New") " " /
pad=(right=2.75% TOP = 0.1% BOTTOM = 0.1%) location=outside valign=top;

    scatterplot y=Study2 x=eval(constant*0) / markercharacter=res1
MARKERCHARACTERPOSITION=center;
    scatterplot y=Study2 x=eval(constant*1) / markercharacter=res2
MARKERCHARACTERPOSITION=center;
    scatterplot y=Study2 x=eval(constant*2) / markercharacter=res3
MARKERCHARACTERPOSITION=center;
endlayout;
endlayout;
endgraph;
end;
run;

```

## CONCLUSION:

Until we were asked to create forest plots, we didn't know that the forest plot is a complex graph with many variations. In this paper we offer a relatively flexible code to produce forest plots that can be customized easily. We used same approach for both subgroup analysis plots and metadata analysis.

## REFERENCES:

1. Prashant, S. 2015. Lost in the Forest Plot? Follow the GTL AXISTABLE Road. Cary, NC: SAS Institute Inc
2. Matange, Sanjay. (2016) CTSPedia Clinical Graphs – Subgrouped Forest Plot. “Graphically Speaking Blog.” Available at <http://blogs.sas.com/content/graphicallyspeaking/2016/05/30/ctspedia-clinical-graphssubgrouped-forest-plot/>. Accessed on March 11, 2017.
3. Matange, Sanjay. (2012) Forest Plot with Subgroups. “Graphically Speaking Blog.” Available at <http://blogs.sas.com/content/graphicallyspeaking/2012/09/30/forest-plot-with-subgroups/>. Accessed on March 11, 2017.
4. Zoran B. 2015. Creating Forest Plots from Pre-computed Data using PROC SGPLOT and Graph Template Language. Little Rock, AR, University of Arkansas for Medical Sciences.
5. SAS/GRAPH 9.2: Graph Template Language User's Guide, Second Edition <http://support.sas.com/kb/35/143.html>

## ACKNOWLEDGMENTS:

We would like to thank Michael Stackhouse and Anbu Damodaran for their guidance and support.

## CONTACT INFORMATION:

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

Savithri Jajam  
[savithri.jajam@covance.com](mailto:savithri.jajam@covance.com)

Olesya Masucci  
[Olesya.Masucci@covance.com](mailto:Olesya.Masucci@covance.com)

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.