Lost in the Forest Plot? Follow the GTL AXISTABLE Road!

Prashant Hebbar, SAS Institute Inc., Cary, NC

ABSTRACT

A forest plot is a common visualization for meta-analysis. Some popular versions that use subgroups with indented text and bold fonts can seem outright daunting to create. With SAS® 9.4, the Graph Template Language (GTL) has introduced the AXISTABLE statement, specifically designed for including text data columns into a graph. In this paper, we demonstrate the simplicity of creating various forest plots using AXISTABLE statements. Come and see how to create forest plots as clear as day!

INTRODUCTION

In this paper, we assume that you are familiar with GTL as well as forest plots. Let us look at some of the features of an example shown in Figure 1. This graph has three distinct areas: a left-side table of study names or labels, a plot in the middle, and a table of statistics on the right side.

If the plot is showing difference between means (standardized mean difference) then its X axis will be a linear axis with a reference line at zero representing no effect or the null hypothesis. If the plot is for ratios (odds ratio), then a linear or logarithmic axis may be used with a reference line at 1 for the null hypothesis.

Additionally, the study names or observation labels can be rendered with subgroup headings in bold-weighted font and indented subgroup values as rows in normal font. The subgroups can be additionally demarcated by reference bands. A diamond marker is also used at times to show the summary of all the observations. (A variant stretches the width of the diamond to show the range of the values.)
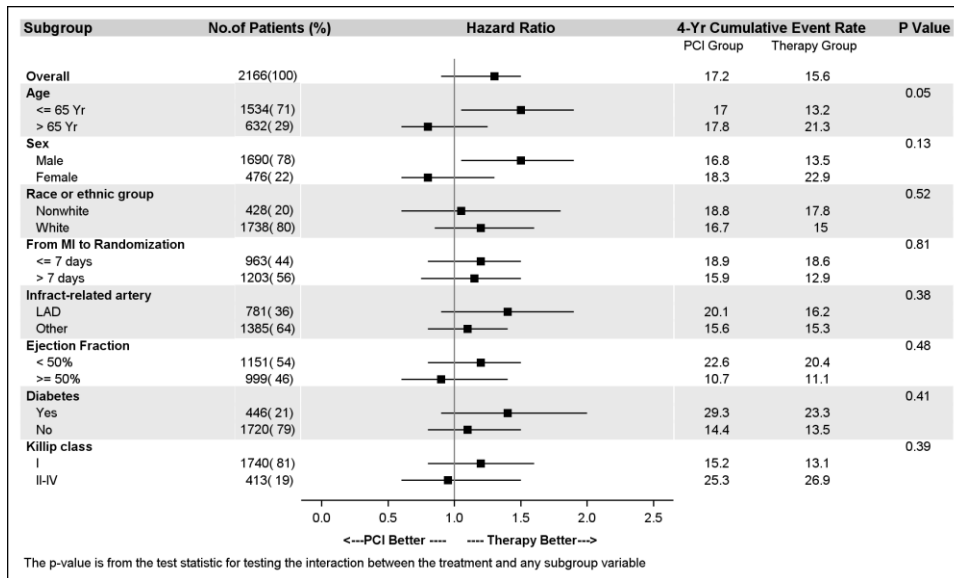

Subgroup	No.of Patients (%)	Hazard Ratio	4-Yr Cumulative Event Rate		P Value
			PCI Group	Therapy Group	
Overall	2166(100)		17.2	15.6	
Age					0.05
<= 65 Yr	1534(71)		17	13.2	
> 65 Yr	632(29)		17.8	21.3	
Sex					0.13
Male	1690(78)		16.8	13.5	
Female	476(22)		18.3	22.9	
Race or ethnic group					0.52
Nonwhite	428(20)		18.8	17.8	
White	1738(80)		16.7	15	
From MI to Randomization					0.81
<= 7 days	963(44)		18.9	18.6	
> 7 days	1203(56)		15.9	12.9	
Infract-related artery					0.38
LAD	781(36)		20.1	16.2	
Other	1385(64)		15.6	15.3	
Ejection Fraction					0.48
< 50%	1151(54)		22.6	20.4	
>= 50%	999(46)		10.7	11.1	
Diabetes					0.41
Yes	446(21)		29.3	23.3	
No	1720(79)		14.4	13.5	
Killip class					0.39
I	1740(81)		15.2	13.1	
II-IV	413(19)		25.3	26.9	

0.0   0.5   1.0   1.5   2.0   2.5

<---PCI Better ----   ---- Therapy Better--->

The p-value is from the test statistic for testing the interaction between the treatment and any subgroup variable

Figure 1: A Typical Forest Plot

In this paper we will go over some ways to make forest plots in GTL with a focus on using the new SAS® 9.4 AXISTABLE statement.

Please see the Resources section towards the end of this paper for links to the PDF file of this paper and the programs used here.

FOREST PLOTS WITHOUT THE AXISTABLE STATEMENT

In earlier SAS releases, we had several ways to create a forest plot, each with its own limitations. In all these implementations, the outermost layout can be one of the following:

- an overlay layout that is partitioned into virtual cells using the OFFSETMIN= and OFFSETMAX= options on the X axis
- a lattice layout with additional nested layouts

SAS 9.2

An example of a forest plot created using SAS 9.2 is shown in Figure 2. It happens to be an SGPLOT procedure implementation with a single layout overlay partitioned into virtual cells using axis offsets, but it can be easily implemented in GTL with the same structure or with a lattice layout as we will see in the next section.
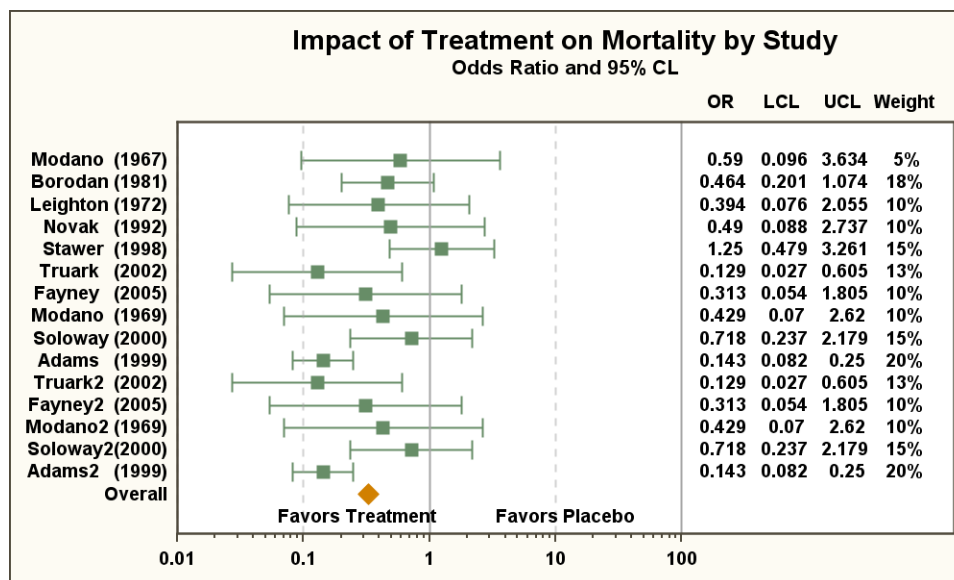

Figure 2: SAS 9.2 Forest Plot Using SGPLOT Procedure

The left-side study names are just the axis tick values for the plot in the center.

The plot itself is showing odds ratio values (with a log scale axis). The individual study values are plotted using a scatter plot with XERRORUPPER= and XERRORLOWER= options to represent the confidence intervals. The overall observation is plotted using a separate scatter plot with a diamond marker symbol.

The right-hand side table of statistics comes from four scatter plots with the MARKERCHARACTER= option.

More details and full code for this example are available in the Graphically Speaking blog post at http://blogs.sas.com/content/graphicallyspeaking/2012/02/01/forest-plot-using-sgplot-procedure/.

SAS 9.2 WITH SUBGROUPS

If your data contains subgroups that need to be displayed in the forest plot, we need to employ a slightly different variation. An example of this implementation is shown in Figure 3.
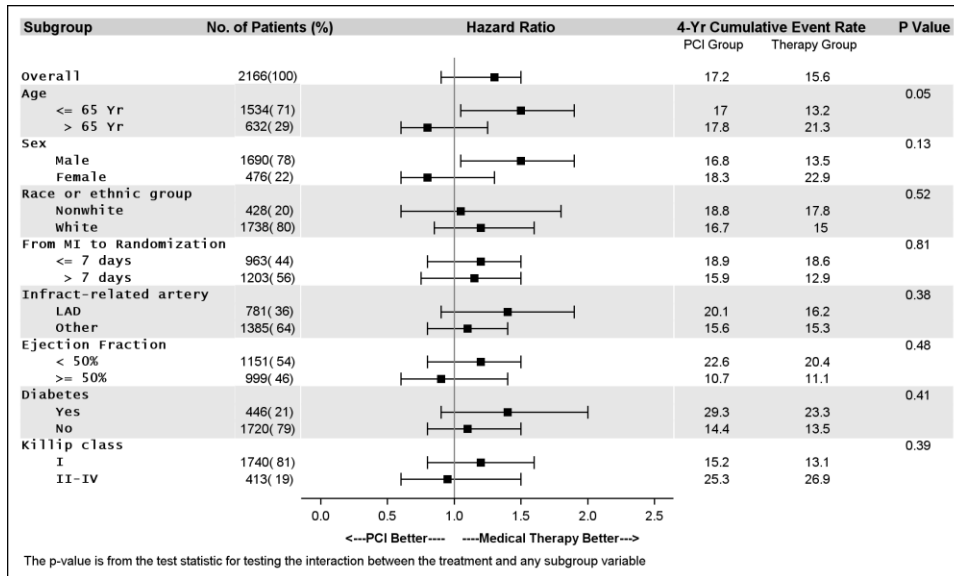
Figure 3: SAS 9.2 Forest Plot with Subgroups

Here we have a GTL implementation with a four-celled lattice layout. Instead of the Y-axis labels on the left side, we use a scatter plot, once again with the MARKERCHARACTER= option to position the subgroups and the indented individual labels. Note that the marker-character text in scatter plots is center-aligned on its data coordinate. Left-aligning marker-character text needs some more steps. We have to calculate the string length and prefix an equal number of nonbreaking space characters ('A0'x in ASCII and UTF-based encodings, '41'x in EBCDIC) to push the actual text rightwards to the center of the new padded string. We also have to specify a fixed-width font family. *The GTL renderer strips leading and trailing whitespace in marker characters!*

The odds ratio plot and the right-side table are implemented similarly to previous versions. Details and full code for this example are available in the Graphically Speaking blog post at http://blogs.sas.com/content/graphicallyspeaking/2012/09/30/forest-plot-with-subgroups.

The **third maintenance release of SAS 9.2** introduced discrete offsets. This enables an alternate implementation for the indention of subgroups by using separate scatter plots for the headings and observation labels, each with its own discrete offset.

SAS 9.3

The high-low plot debuted in this release of SAS. This plot enables you to draw varying length bars (or lines), each with its own start (low) and end (high). See Figure 4 for a forest plot created using high-low plots.
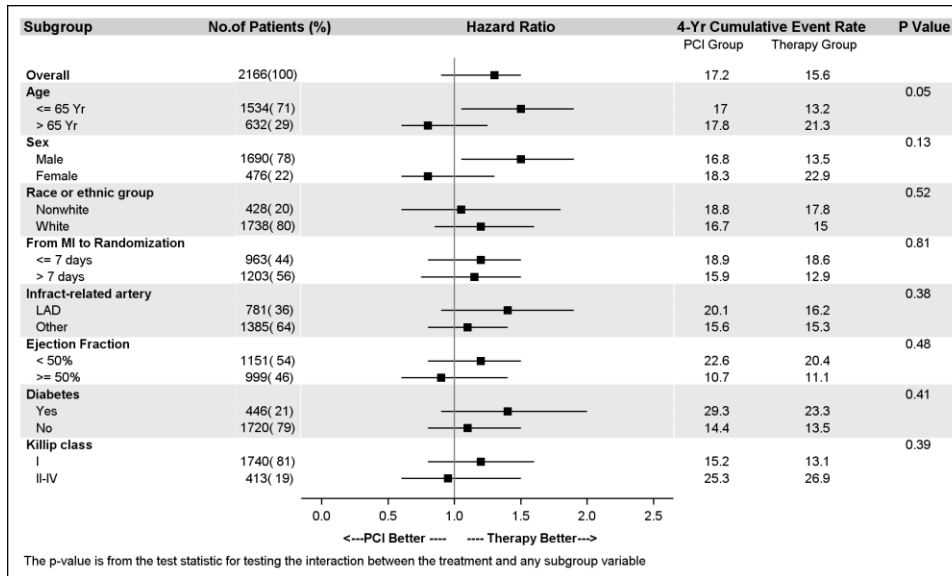
Subgroup	No.of Patients (%)	Hazard Ratio	4-Yr Cumulative Event Rate		P Value
			PCI Group	Therapy Group	
Overall	2166(100)		17.2	15.6	
Age					0.05
<= 65 Yr	1534(71)		17	13.2	
> 65 Yr	632(29)		17.8	21.3	
Sex					0.13
Male	1690(78)		16.8	13.5	
Female	476(22)		18.3	22.9	
Race or ethnic group					0.52
Nonwhite	428(20)		18.8	17.8	
White	1738(80)		16.7	15	
From MI to Randomization					0.81
<= 7 days	963(44)		18.9	18.6	
> 7 days	1203(56)		15.9	12.9	
Infract-related artery					0.38
LAD	781(36)		20.1	16.2	
Other	1385(64)		15.6	15.3	
Ejection Fraction					0.48
< 50%	1151(54)		22.6	20.4	
>= 50%	999(46)		10.7	11.1	
Diabetes					0.41
Yes	446(21)		29.3	23.3	
No	1720(79)		14.4	13.5	
Killip class					0.39
I	1740(81)		15.2	13.1	
II-IV	413(19)		25.3	26.9	

0.0   0.5   1.0   1.5   2.0   2.5

<---PCI Better ----   ---- Therapy Better--->

The p-value is from the test statistic for testing the interaction between the treatment and any subgroup variable

Figure 4: SAS 9.3 Forest Plot Using High-Low Plots

The main layout is once again a four-celled lattice layout. The top-level headers were created by nesting another four-celled lattice layout inside the top sidebar, each containing an entry statement.

The left-hand subgrouped table was created using two high-low plots: one for the bold headings and another for the values in regular font drawn with a high of 1 unit. Both plots used a line thickness of zero to suppress the drawing of the low-to-high line. The data for these plots can be set up as two separate columns of zero and one unit for HIGH=. They can also be generated from the same column by using the *eval(ifn())* expression to partition the data into two sets based on a subgroup "flag" column. Note that the font no longer needs to be fixed-width: this works with proportional fonts!

A scatter plot with marker character was used for all other text columns, including the right-hand side table.

The odds ratio plot in the middle was implemented using a high-low plot for the confidence limits and a scatter plot overlay for the odds-ratio value. Notice that the lines for the confidence limits no longer have serifs, thanks to the high-low plot!

You can read more about this implementation in the Graphically Speaking blog post at
http://blogs.sas.com/content/graphicallyspeaking/2013/01/06/forest-plot-with-sas-9-3/.

THE AXISTABLE STATEMENT

As you saw in the previous sections, creating left or right aligned text at specific data coordinates required either padding the text with non-breaking space and using fixed-width font, or using the high-low plot and suppressing the high-low line. Rendering text with indentations required adjusting the text element's non-breaking space padding or the modifying the data for the high-low columns. Creating bold subgroup headings also required some extra work.

To solve these issues and more, in SAS 9.4 we expressly designed the axis table. *The forest plot was one of the prime drivers for the AXISTABLE statement*. This statement offers a convenient way to place text values at specific coordinates along an axis. The table values can be aligned along the X or the Y axis. This statement supports data-driven indentation for each text element. It also allows you to control the text attributes with a text group. A simple example of a Y axis table on the left, aligned with a horizontal bar chart on its right is shown in Figure 5.
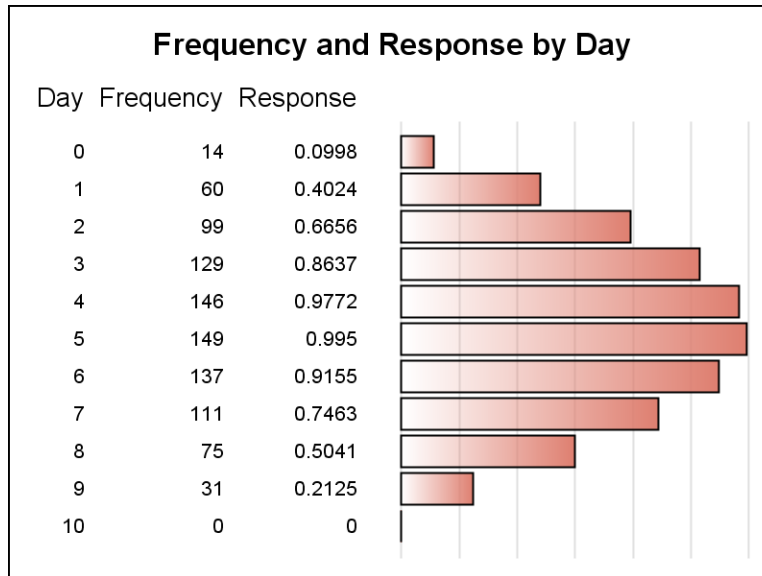
Figure 5: A Y Axis Table Example

Its basic syntax is:

```
AXISTABLE X=column VALUE=column < / options >;
-OR-
AXISTABLE Y=column VALUE=column < / options>;
```

Here are some of options:

- CLASS=column - creates multiple rows (for the X axis table) or columns (for the Y axis table) for each class value.
- COLORGROUP=column - assigns a color for the text values based on the values of the column.
- CLASSDISPLAY=STACK | CLUSTER - positions the class values as a stack (default) or clustered around the coordinate value.
- STAT=AUTO | SUM | MEAN - specifies how to summarize the (numeric) columns.
- TEXTGROUP=column - specifies a classifier for controlling text attributes such as size, font and weight.
- SHOWMISSING=boolean - suppresses the display of numeric and character missing values if set to true.

By now you can probably guess how these features play into making a forest plot!

SAS 9.4 FOREST PLOTS USING THE AXISTABLE STATEMENT

The AXISTABLE statement is a versatile statement for aligning text along axis coordinates. It works well for creating rows (or columns) of text aligned with the X (or Y) axis. Let us now look at how you can use the axis table to create forest plots. We start with the data as shown in Table 1.

ID	Subgroup	Count	Percent	Mean	Low	High	PCIGroup	Group	PValue
1	Overall	2166	100	1.30	0.90	1.50	17.2	15.6	.
1	Age	.	.	.	.	.	.	.	0.05
2	<= 65 Yr	1534	71	1.50	1.05	1.90	17.0	13.2	.

ID	Subgroup	Count	Percent	Mean	Low	High	PCIGroup	Group	PValue
2	> 65 Yr	632	29	0.80	0.60	1.25	17.8	21.3	.
1	Sex	.	.	.	.	.	.	.	0.13
2	Male	1690	78	1.50	1.05	1.90	16.8	13.5	.
2	Female	476	22	0.80	0.60	1.30	18.3	22.9	.
	. . .								

Table 1: Data for Forest Plot

The ID variable determines subgroup headings. It is set to 1 for headings and 2 otherwise. Subgroup, Count and Percent variables are for the left-side table. Low, High and Mean variables are for the odds-ratio plot. PCIGroup, Group and PValue variables are for the right-side table.

We need to add some more secondary variables to this data as shown below:

```
data forest2;
  set forest;

    /* reference bands */
  if mod(_N_-1, 6) in (1, 2, 3) then ref=obsId;

   /* Reduce indent for subGroup heading */
  if ID=1 then indentWt=0;

run;
```

The ref variable is for the reference bands. indentWt is computed from the ID subgroup flag.

Let us focus on a forest plot with the left-side table and an odds-ratio plot first. One such example is shown in Figure 6. The outermost layout in this example is a lattice layout with a single overlay layout cell and a top sidebar for the headers.
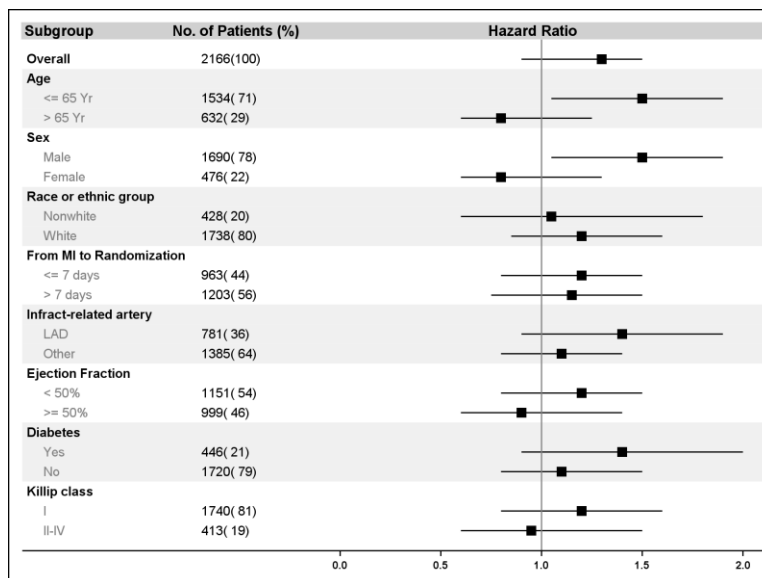

Figure 6: Forest Plot Using SAS 9.4 AXISTABLE Statement

The left-side table containing Subgroup and the No. of Patients was created using two AXISTABLE statements that were nested inside the overlay layout's inner margin. Reference bands were created

using a REFERENCELINE statement with a judiciously chosen line thickness. Here is a snippet of the GTL template for the left side:

```
...
layout lattice / columns=1;

    /*--Column headers--*/
    sidebar / align=top;
       <Entry statements>
    endSidebar;

    /* Single cell with inner margin for left table */
    layout overlay / yAxisOpts=(reverse=true display=none)
                     xAxisOpts=( ... ) wallDisplay=none ... ;

       /* Left-side table */
       innerMargin / align=left;
            axisTable y=obsId value=subGroup / indentWeight=indent
                                               textGroup=grp;
            axisTable y=obsId value=countPct / labelattrs=(size=7);
       endInnerMargin;

       /* Odds Ratio plot */
       referenceLine y=ref / lineAttrs=(thickness=15 color=_bandcolor);
       scatterPlot y=obsId x=mean / markerAttrs=(symbol=squareFilled)
                   xErrorLower=low xErrorUpper=high errorBarCapShape=none;
       referenceLine x=1;

    endLayout; /* overlay */

endLayout; /* lattice */
...
```

The plot in the middle is composed of a scatter plot for the odds-ratio values with error bars for representing the confidence interval. The serifs on the error bars have been turned off by setting the new SAS 9.4 option ERROBARCAPSHAPE to *none*. The Y reference line is for creating the reference 'bands'. By now, you can see how we can easily extend this to add a right-side statistics table using more Y axis tables!

For the **third maintenance release of SAS 9.4** we will add support for controlling the extents of an axis line to just the data area, excluding any offsets or inner margins. This will enable you to improve this graph as shown in Figure 7. In this implementation, we have added the right-side table using additional axis tables placed inside another inner margin aligned to the right. Notice that the X axis line only spans the odds-ratio plot and does not intrude into the tables on either side.
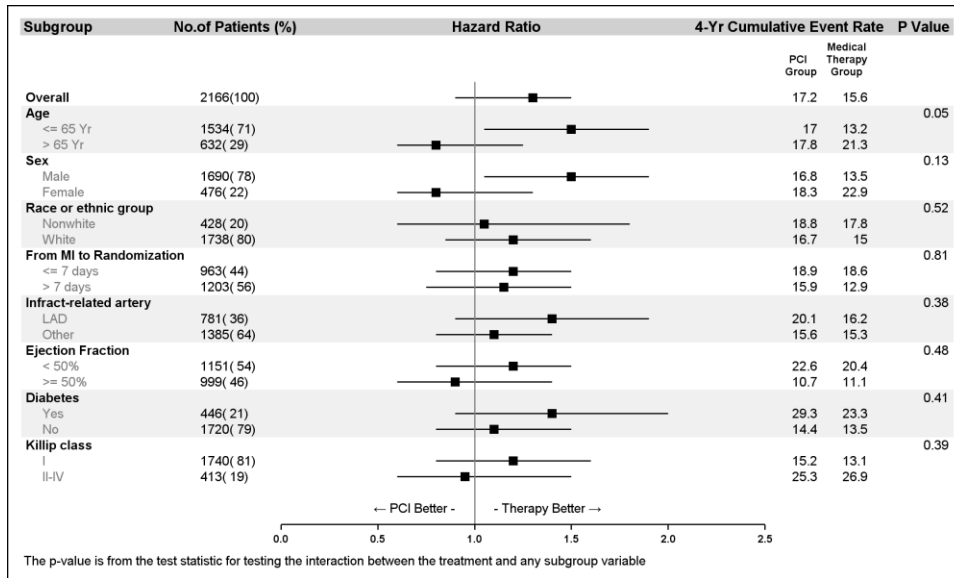
Subgroup	No.of Patients (%)	Hazard Ratio	4-Yr Cumulative Event Rate		P Value
			PCI Group	Medical Therapy Group	
Overall	2166(100)		17.2	15.6	
Age					0.05
<= 65 Yr	1534(71)		17	13.2	
> 65 Yr	632(29)		17.8	21.3	
Sex					0.13
Male	1690(78)		16.8	13.5	
Female	476(22)		18.3	22.9	
Race or ethnic group					0.52
Nonwhite	428(20)		18.8	17.8	
White	1738(80)		16.7	15	
From MI to Randomization					0.81
<= 7 days	963(44)		18.9	18.6	
> 7 days	1203(56)		15.9	12.9	
Infract-related artery					0.38
LAD	781(36)		20.1	16.2	
Other	1385(64)		15.6	15.3	
Ejection Fraction					0.48
< 50%	1151(54)		22.6	20.4	
>= 50%	999(46)		10.7	11.1	
Diabetes					0.41
Yes	446(21)		29.3	23.3	
No	1720(79)		14.4	13.5	
Killip class					0.39
I	1740(81)		15.2	13.1	
II-IV	413(19)		25.3	26.9	

← PCI Better -   - Therapy Better →

0.0   0.5   1.0   1.5   2.0   2.5

The p-value is from the test statistic for testing the interaction between the treatment and any subgroup variable

Figure 7: Forest Plot with Axis Line Extent Set to Data (SAS 9.4 M3)

Also note that we have shown explanatory labels containing Unicode left and right arrows just above the X axis. These labels were implemented using the TEXTPLOT statement which was introduced in the **second maintenance release of SAS 9.4** in conjunction with *Unicode support in user-defined formats*, which will be a powerful new feature in the **third maintenance release of SAS 9.4**.

AXIS TABLES IN SAS 9.4 SG PROCEDURES

We have discussed GTL so far, but you can also use the SGPLOT procedure to create forest plots with this new feature. In SAS 9.4, the SGPLOT procedure supports XAXISTABLE and YAXISTABLE statements. The **third maintenance release of SAS 9.4** enhances the SGPANEL procedure with the COLAXISTABLE and ROWAXISTABLE statements and also supports control of axis line extent via the STYLEATTRS statement's AXISEXTENT option.

Figure 8 shows one such example created using YAXISTABLE statements inside the SGPLOT procedure. Once again, this graph uses axis offsets to create virtual cells on the left and right side of the overlay layout.
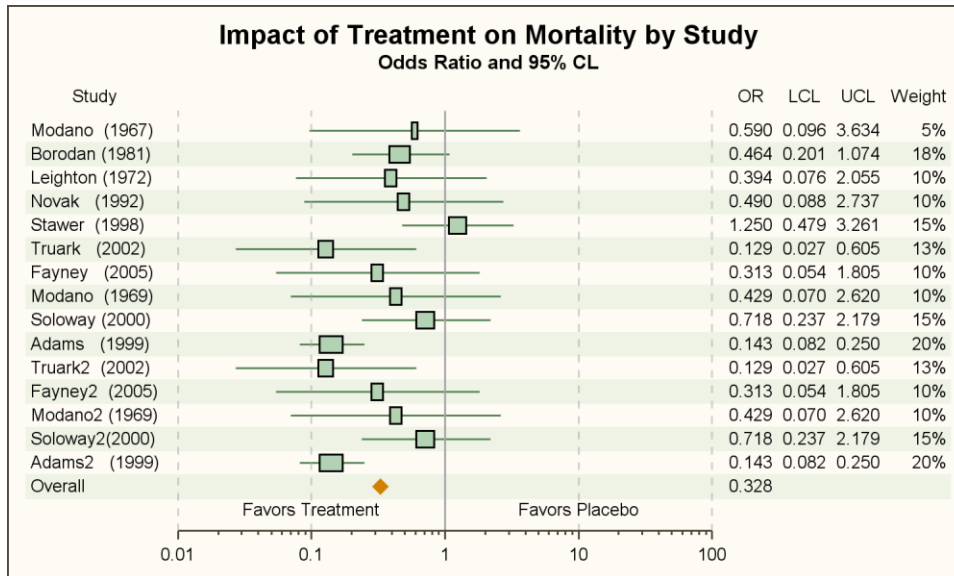
Figure 8: SAS 9.4M3 SGPLOT Forest Plot using YAXISTABLE and AXISEXTENT

CONCLUSION

The forest plot is a complex graph with many variations. You can create these graphs in GTL as well as the SGPLOT procedure in previous versions of SAS. However, with the SAS 9.4 AXISTABLE statement, constructing forest plots is a lot easier! Likewise, you can also use the SAS 9.4 XAXISABLE and YAXISTABLE statements in the SGPLOT procedure to craft forest plots with ease.

RESOURCES

The PDF file of this paper and the SAS code for all the programs is available at:
Paper: http://support.sas.com/resources/papers/proceedings15/SAS1748-2015.pdf
Code: http://support.sas.com/resources/papers/proceedings15/SAS1748-2015.zip

REFERENCES

Matange, S., and D. Heath. 2011. *Statistical Graphics Procedures by Example: Effective Graphs Using SAS®*. Cary, NC: SAS Institute Inc.

Matange, Sanjay. 2013. *Getting Started with the Graph Template Language in SAS®: Examples, Tips, and Techniques for Creating Custom Graphs*. Cary, NC: SAS Institute Inc.

Matange, Sanjay. 2013. "Getting Started with SAS® Graph Template Language." *SAS Talks Webinar Series*. Available at http://www.sas.com/reg/web/corp/2260494.

Matange, Sanjay. "Graphically Speaking." Available at http://blogs.sas.com/content/graphicallyspeaking . Accessed on April 10, 2015.

ACKNOWLEDGMENTS

Many thanks to Sanjay Matange for his guidance and constructive comments.

RECOMMENDED READING

- *SAS® 9.4 Graph Template Language: Reference*

- *SAS® 9.4 Graph Template Language: User's Guide*

- *SAS® 9.4 ODS Graphics: Procedures Guide*

CONTACT INFORMATION

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

> Prashant Hebbar
> 100 SAS Campus Drive
> Cary, NC 27513
> SAS Institute Inc.
> prashant.hebbar@sas.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.