

Multiple Forest Plots and the SAS® System

Poster 10

Anne Barfield, Quanticate, Manchester, United Kingdom

ABSTRACT

This paper is the accompanying paper to the poster entitled 'Multiple Forest Plots and the SAS® System'. It discusses the SAS/Graph® Annotation option and how this can be used in combination with SAS® Macros to allow the creation of multiple Forest Plots, giving details on what can and cannot be controlled as part of the macro call.

The paper will also cover some basic introductory information on creating annotations and the background to Forest Plots. A simple example of how to use annotation datasets will be presented – further more detailed examples are given on the Poster itself.

INTRODUCTION

The purpose of this poster and accompanying paper are to highlight the methods of using the ANNOTATE Option available with the SAS/GRAPH® procedures for producing a Forest Plot using the SAS® system and how this can be adapted to allow multiple plots to be created using SAS® Macros.

SAS® Programmers are frequently asked to produce graphical representations of data, from simple pie charts and box plots to more statistical graphics such as the Forest Plot. SAS/Graph® provides graphing tools to create plots and charts with only a few SAS statements and by using the built in options provided it is possible to produce graphical summaries of almost any need. However the default graphs do not always suffice and it is occasionally necessary to 'add' to the default outputs.

In these cases, SAS/Graph's Annotate data set can be used to overcome the limitations of SAS/Graph's procedures to create a new type of graph from an existing one or to enhance the normal output of a graph. Forest Plots are an example of this in that there is no single method of creating such a plot using the SAS® system and the use of the Annotate dataset is necessary.

The Annotate data set can be used to add additional information to an existing graphical output or to create a completely new type of graph. The Annotate dataset allows the user to place lines, bars, text or symbols on a graph with the positioning of these additions given using either the data from the graphed data set itself as a reference, or with coordinates provided by the programmer. Usually, the annotation uses a combination of the two.

THE FOREST PLOT

A Forest Plot is a graphical display designed to illustrate the strength of treatment effects across treatments groups, subgroups of a study and multiple studies addressing the same question. The Plots were initially developed as a means of graphically representing a meta-analysis of the results of randomized controlled trials.

Although Forest Plots can take several forms, they are commonly presented with two columns. The left-hand column lists the names of the analysis groups and the right-hand column is a plot of the measure of effect (e.g. an odds ratio) for each of these groups (often represented by a circle) incorporating confidence intervals represented by horizontal lines.

The advantage of using a forest plot to show such data is that the reviewer can see at a glance any differences between several analysis groups without having to look at the results in detail – saving time.

An example of a Forest Plot is given below (*Figure 1*) – Subgroup analysis names have been removed but you can see that the confidence Intervals and the hazard ratios are clearly shown and allow a quick comparison of the subgroups. In this example some Hazard Ratios and Confidence Intervals are not shown due to the number of events for that subgroup being too low to calculate an accurate Hazard ratio.

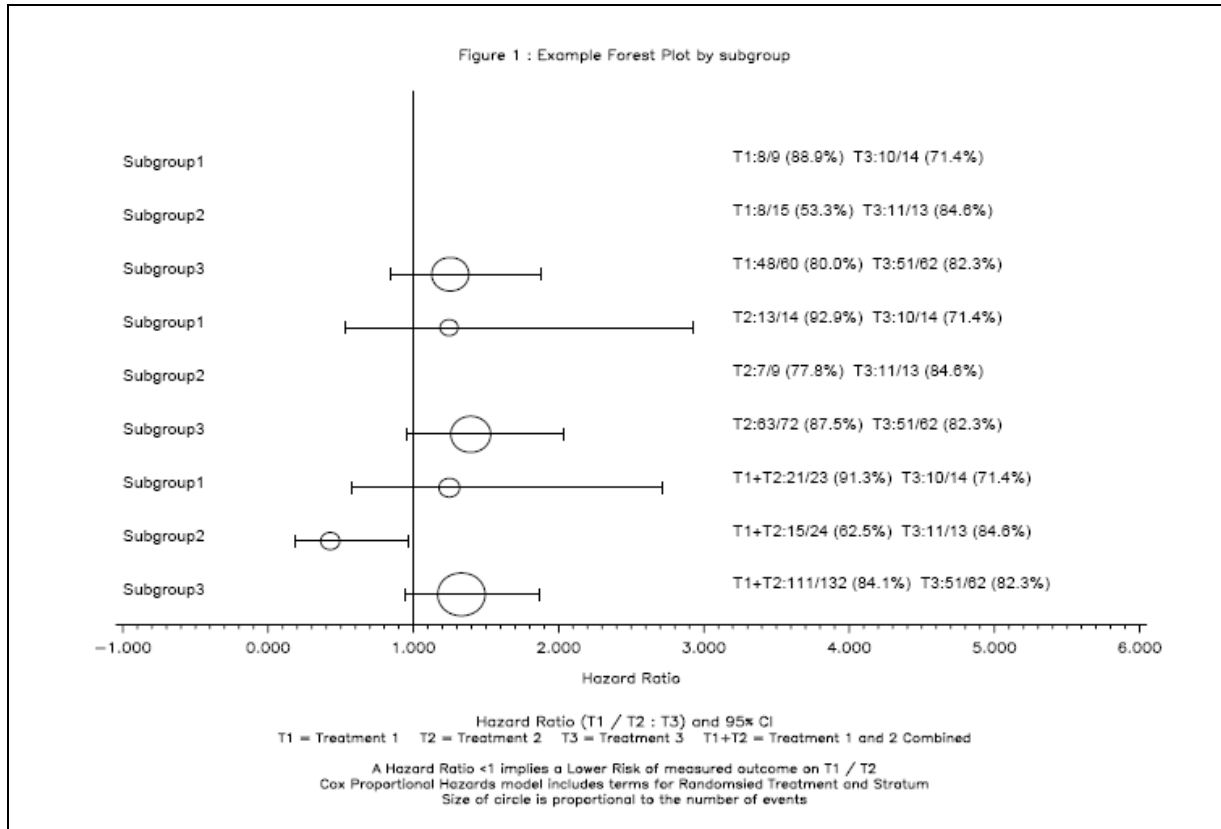


FIGURE 1 : EXAMPLE FOREST PLOT

Figure 1 has been created using the bubble option of PROC Gplot and as a result the circles representing the Hazard Ratios are part of the default output. However the Subgroup labels, the Confidence Intervals and the Text showing the number of events within each subgroup have all been added to the default output using the Annotate option and an Annotate Dataset.

ANNOTATE DATASET

The Annotate option requires a dataset specifically designed to tell SAS/GRAPH® how to enhance your current output. There are two basic functions which allow for moving and drawing, and other functions can be used to position labels, create bars, pies and polygons. The ANNOTATE dataset itself must have one observation per function and essentially tells SAS® what and where to draw or place text.

The FUNCTION variable created in the dataset determines the graphics element that is drawn. The elements available are given in Table 1. Any of the elements given in Table 1 can be used to create a Figure. For the Forest Plot shown in Figure 1 we need only use the Move, Draw and Label elements to create the Confidence intervals and the text that is shown on the graph. The Hazard ratio bubbles are created as default and so no annotation is required for this.

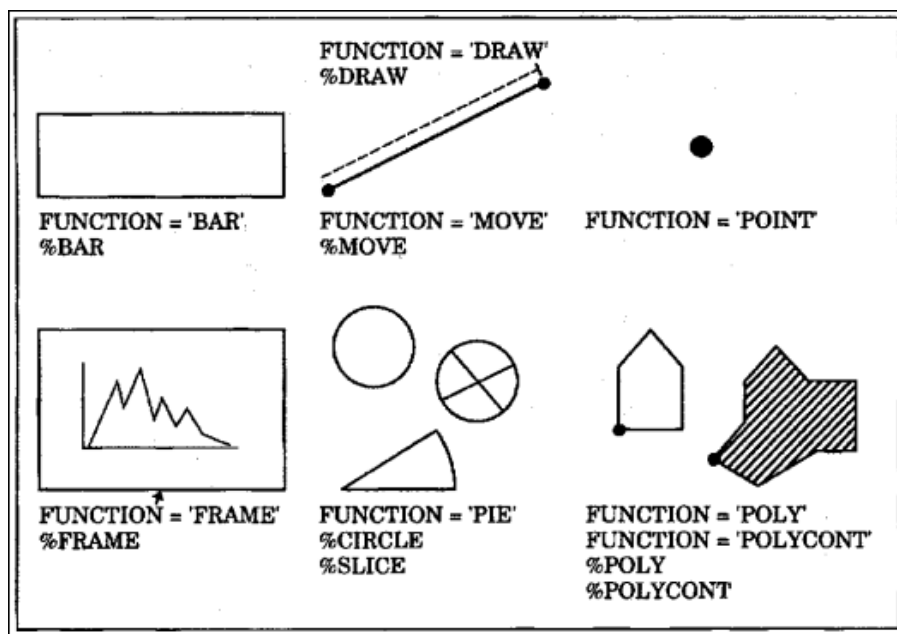


TABLE 1 : Graphics Elements available for use in the ANNOTATE Dataset
Reference 1

A simple example (showing the creation of the Confidence Intervals for Figure 1 above) is shown below. You can see that the code uses the data set that is to be graphed as a basis and adds the functions move and draw to tell SAS® where to put the Confidence Intervals. The Confidence Intervals and hazard ratios have already been calculated and are stored in the dataset *data1*. The bars at the end of the Confidence Intervals are created separately to the actual Confidence Intervals.

```
DATA BARS;
  SET DATA1;
  IF HAZARD NE .; /* ONLY FOR OBSERVATIONS WHERE THE HAZARD RATIO HAS BEEN CALCULATED.
  LENGTH FUNCTION $8. TEXT $40.; XSYS = '2'; YSYS = '2';
  Y=ID; X=HAZARD; FUNCTION='MOVE'; OUTPUT;

  * LOWER CONFIDENCE INTERVAL;
  IF LCL NE . THEN DO;
    X=LCL; FUNCTION='DRAW'; OUTPUT;
    LINK TIPS;
  END;

  * UPPER CONFIDENCE INTERVALS;
  IF UCL NE . THEN DO;
    X=UCL; FUNCTION='DRAW'; OUTPUT;
    LINK TIPS;
  END;
  RETURN;

  TIPS: *** DRAW TOP AND BOTTOM OF BARS ***;
  Y=ID-0.1; FUNCTION = 'MOVE'; OUTPUT;
  Y=ID+0.1; FUNCTION= 'DRAW'; OUTPUT;
  Y=ID; FUNCTION='MOVE'; OUTPUT;
  RETURN;
RUN;
```

PhUSE 2010

The text on the graph is created using the label element function an example of which is given below.

```
DATA LABELS;
  SET DATA1;
  IF HAZARD > . THEN DO;
    LENGTH FUNCTION $8. TEXT $50.;

    * LABELS FOR THE SUBGROUPS.;
    Y=ID; X=6; STYLE="'ARIAL'";
    FUNCTION = 'LABEL'; TEXT='SUBGROUP' || TRIM(LEFT(PUT(INTNSTAT, 1.0)))
      SIZE=0.8; POSITION='0';
    OUTPUT;

    * NUMBER OF EVENTS;
    POSITION='0'; FUNCTION='LABEL';
    IF VARIABLE='TX1' THEN TEXT =
      'T1:' || TRIM(LEFT(EG1)) || '/' || TRIM(LEFT(NG1)) || ' (' || TRIM(LEFT(PUT(PG1, 4.1))) || '%' )
      || ' T3:' || TRIM(LEFT(EM1)) || '/' || TRIM(LEFT(NM1)) || ' (' || TRIM(LEFT(PUT(PM1, 4.1)))
      || '%' )';
    OUTPUT;
  END;
RUN;
```

Each subgroup requires its own label function and so the code above would have several more label statements for the number of events.

The Confidence Interval and Label annotation datasets can then be set together to create a single annotation dataset which is then referenced within the graphical procedure as ANNOTATE= *dataset name*.

```
DATA ANNO1;
  SET BARS LABELS;
RUN;

PROC GPLOT DATA=DATA1 ANNOTATE=ANNO1;
  BUBBLE ID*HAZARDRA=EVENT;
RUN;
```

Axis, titles and footnote statements can be added to the GPLOT procedure as normal.

MACROS

Due to the amount of detail required in the Annotate dataset programmers tend to prefer to modify the code for each output required. However sometimes ten's maybe hundred's of figures are required and so a macro would be the most efficient method to use. The poster I have created details the changes that would need to be made in order for the program to be used to create multiple forest plots.

The changes take place throughout the code including the code which creates the annotate dataset and dependant upon the changes across figures are quite minimal. All the cosmetics of the figure are controllable and the annotate dataset is created from your actual data and so can be adapted within a macro taking away the need for separate codes.

The poster gives examples of these changes and shows which elements of the outputs can be altered using this method.

REFERENCES

Reference 1 – www.sas.com

CONTACT INFORMATION

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

Anne Barfield
Quanticate
BlackBox Business Centre (Suite G.4)
Beech Lane
WILMSLOW
Cheshire
SK9 5ER

Work Phone: 01625 544825
Email: anne.barfield@quanticate.com
Web: www.quanticate.com

Brand and product names are trademarks of their respective companies.