How to produce Forest Plots with Proc GPLOT and Annotate Facility

Laetitia Lemoine Le Halpert, Quintiles, Paris

ABSTRACT

The purpose of this paper is to describe how to produce Forest-Plots by using the SAS® procedure GPLOT and Annotate facility. The Forest-Plot can be used for various analyses to display statistical data such as Odds-Ratios, Hazard Rates, Risk ratios and many other statistical data.

INTRODUCTION

There is no SAS procedure which can directly display a Forest plot. Using an annotate dataset containing all the elements of the graphics and then using a SAS® proc GPLOT allow to generate this graph. This paper will be illustrated by an example of a Forest-Plot (see Section 4) produced to show the benefit of one medical treatment compared to another using Odds-Ratios. Odds-Ratios and 95%CIs will also be displayed on the same graph for several sub-groups of patients.

The forest-plot allows the reader to quickly assess if an odds-ratio is significant by showing the value of the OR and its 95%CI. An odds-ratio is significant when the line joining the confidence limits does not cross the vertical axis, i.e. if the 95%CI does not contain the value 0.

Four steps are needed to produce the Forest-Plot.
1- Preparation of the main SAS® dataset
2- Creation of annotate dataset containing all information to be displayed in the plot
   This part will explain how and in which areas to put text (label of variables, of sub-groups, number of events in each sub-group, odds-ratios and 95%CIs, draw horizontal and vertical lines, how to put ticks on the horizontal axis since a logarithmic scale has to be used etc.).
3- Procedure GPLOT
4- Forest-plot results

1-PREPARATION OF THE MAIN SAS® DATASET

The example used is data from 1000 patients suffering from cancer, two treatment groups X and Y, several variables such as the Best Overall Response to treatment, and sub-groups variables (age, sex, ethnic origin, region). Frequencies, common OR and CI values are obtained by performing a PROC FREQ between the response variable and treatment group, for each sub-group variable.

ods output commonrelrisks=commonOR crosstabfreqs=freqs ;
proc freq data=table;
tables trtgrp*response /all ;
by subgroup ;
run;
RESULTING DATASET ‘FINAL’

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>VAR</td>
<td>CAT</td>
<td>N1</td>
<td>N2</td>
<td>NBEV1</td>
<td>NBEV2</td>
<td>OR</td>
<td>ORLCL</td>
<td>ORUCL</td>
<td>NTEXT</td>
<td>NBEV</td>
<td>ORCI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>1091</td>
<td>544</td>
<td>279 vs 544</td>
<td>279 vs 231</td>
<td>1.408</td>
<td>1.109</td>
<td>1.789</td>
<td>(547 vs 544)</td>
<td>279 vs 231</td>
<td>1.408</td>
<td>[1.109, 1.789]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>9</td>
<td>691</td>
<td>344</td>
<td>183 vs 344</td>
<td>183 vs 139</td>
<td>1.710</td>
<td>1.265</td>
<td>2.310</td>
<td>(175 vs 145)</td>
<td>175 vs 145</td>
<td>1.277</td>
<td>[0.941, 1.733]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>423</td>
<td>220</td>
<td>104 vs 220</td>
<td>104 vs 86</td>
<td>1.591</td>
<td>1.079</td>
<td>2.347</td>
<td>(203 vs 196)</td>
<td>104 vs 86</td>
<td>1.591</td>
<td>[1.079, 2.347]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>372</td>
<td>188</td>
<td>100 vs 188</td>
<td>100 vs 64</td>
<td>2.113</td>
<td>1.387</td>
<td>3.218</td>
<td>(188 vs 184)</td>
<td>100 vs 64</td>
<td>2.113</td>
<td>[1.387, 3.218]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>243</td>
<td>123</td>
<td>60 vs 123</td>
<td>60 vs 53</td>
<td>1.205</td>
<td>0.729</td>
<td>1.994</td>
<td>(123 vs 120)</td>
<td>60 vs 53</td>
<td>1.205</td>
<td>[0.729, 1.994]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

DESCRIPTION OF VARIABLES

VAR : 1 TO 5, all subjects, Age, Sex, Ethnic Origin, Region
CAT : 1 TO 10, subgroups
N : Total number of patients
N1 : Number of patients in Treatment X group
N2 : Number of patients in Treatment Y group
NBEVE1 : Number of events in Treatment X group
NBEVE2 : Number of events in Treatment Y group
OR : Odds-ratio
ORLCL : Lower limit of odds-ratio
ORUCL : Upper limit of odds-ratio
NTEXT : Number of patients in Treatment X group versus Number of patients in Treatment Y group
NBEV : Number of events in Treatment X group versus Number of events in Treatment Y group
ORCI : Odds-ratio and Confidence interval

PROC FORMAT

```
proc format;
VALUE VAR 1 = "All Subjects"
   2 = "Age"
   3 = "Sex"
   4 = "Ethnic Origin"
   5 = "Region"
;
VALUE CAT 10 = "All Subjects"
   9 = "< 65 years"
   8 = "\geq 65 years"
   7 = "Male"
   6 = "Female"
   5 = "Caucasian"
   4 = "Non caucasian"
   3 = "Western Europe"
   2 = "Eastern Europe"
   1 = "Other"
;
run;
quit;
```

data final ;
set final ;
format var var. cat cat. ;
logOR = log(OR) ;
logORLCL = log(ORLCL) ;
logORUCL = log(ORUCL) ;
run ;
2- CREATION OF THE ANNOTATE DATASET

Annotate SAS Facility will be used to display text as well to draw lines in the required area of the graph. Some macros variables will be defined at first to fix the size and format of text.

* Initialization of annotate macros ;
%annomac ;

* Macro-variables for size and font of text : 
%let size=1.2 ;
%let tsize=1.5 ;
%let ftitle='Arial' ;
%let ftext ='Arial' ;

* Macro for ticks on the horizontal axis ;
%macro tick(val=,VALlbl=_NO_,fontVAL=simplex,origin=0,tickl=-0.8,ticks=0.15);

* Area (Data): Absolute, values 
xsys='2';ysys='2';
%move(&val,&origin);

* Area (Procedure output area): Relative, %.
xsys='B';ysys='B';

* Draw each tick from origin.
%draw(&origin,&tickl.,black,1,&ticks.);
%cntl2txt(); (Copies the values of the internal coordinates (XLAST, YLAST) to the text coordinate (XLASTT, YLASTT).
%if "&VALlbl" eq "blank" %then %let val=" ";
%else %if "&VALlbl" ne "_NO_" %then %let val="&VALlbl.";
%else %let val="&val.";
%label(0,0,&val,black,0,0,&size.,&ftext.,6);
%mend tick;

proc sort data=final;
  by var descending cat;
run ;

* Annotate dataset ;

DATA anno;
  length text catnobs varc. $100. ;
  SET final ;
  by var descending cat catgrpc;
  catnobs=compbl(put(cat,catg.)!!' '!!ntext);
  varc=trim(left(put(var,var.)));'

1- HOW TO PUT TITLE AT THE TOP OF THE GRAPH AND TEXT AT THE BOTTOM OF THE GRAPH

* Area (Graphic output area): Absolute, %.
%system(3,3);

Draw the text by using the SAS function %label and the two arrows by using the SAS function %line. Several variables will be created in the annotate dataset: function variable, positional variables and attribute variables.

%label(2,88,'Subgroup (number of patients in Treatment X vs Treatment Y)' ,black,0,0,&size.,&ftext.,6) ;
%label(63,88,'Number of responses' ,black,0,0,&size.,&ftext.,6) ;
PhUSE 2008

%label(60,86.5,'Treatment X vs Treatment Y' ,black,0,0,&size.,&ftext.,6) ;
%label(93,88,'OR [95% CI]' ,black,0,0,&size.,&ftext.,6) ;
%label(20,3.5,'No Benefit under Treatment X' ,black,0,0,&size.,&ftext.,6) ;
%line(30,1.8,50,1.8,black,1,0.5) ;
%line(30,1.8,32,2.1,black,1,0.5) ;
%line(30,1.8,32,1.5,black,1,0.5) ;
%label(65,3.5,'Benefit under Treatment X' ,black,0,0,&size.,&ftext.,6) ;
%line(54,1.8,74,1.8,black,1,0.5) ;
%line(74,1.8,72,2.1,black,1,0.5) ;
%line(74,1.8,72,1.5,black,1,0.5) ;
%label(54,3.7,'OR and 95% CI ' ,black,0,0,&size.,&ftext.,5) ;

2- HOW TO PUT TICS ON THE HORIZONTAL AXIS USING A LOGARITHMIC SCALE

In this example, each tick will be drawn from OR=-2.99 to 2.99 (=log(0.05) to log(20))

%dclanno;
hsys='3';

%tick(val=-2.99,VALlbl=0.05 ,fontVAL=&ftext);
%tick(val=-2.81,VALlbl=blank,fontVAL=&ftext,tickl=-0.5,ticks=0.15);
%tick(val=-2.66,VALlbl=blank,fontVAL=&ftext,tickl=-0.5,ticks=0.15);
%tick(val=-2.53,VALlbl=blank,fontVAL=&ftext,tickl=-0.5,ticks=0.15);
%tick(val=-2.40,VALlbl=blank,fontVAL=&ftext,tickl=-0.5,ticks=0.15);
%tick(val=-2.30,VALlbl=0.1 ,fontVAL=&ftext);
%tick(val=-1.61,VALlbl=0.2 ,fontVAL=&ftext);
%tick(val=-1.20,VALlbl=blank,fontVAL=&ftext,tickl=-0.5,ticks=0.15);
%tick(val=-0.92,VALlbl=blank,fontVAL=&ftext,tickl=-0.5,ticks=0.15);
%tick(val=-0.69,VALlbl=blanck,fontVAL=&ftext);
%tick(val=-0.51,VALlbl=blank,fontVAL=&ftext,tickl=-0.5,ticks=0.15);
%tick(val=-0.36,VALlbl=blank,fontVAL=&ftext,tickl=-0.5,ticks=0.15);
%tick(val=-0.22,VALlbl=blank,fontVAL=&ftext,tickl=-0.5,ticks=0.15);
%tick(val=-0.11,VALlbl=blank,fontVAL=&ftext,tickl=-0.5,ticks=0.15);
%tick(val= 0 ,VALlbl=1 ,fontVAL=&ftext);
%tick(val= 0.69,VALlbl=2 ,fontVAL=&ftext);
%tick(val= 1.1 ,VALlbl=blank,fontVAL=&ftext,tickl=-0.5,ticks=0.15);
%tick(val= 1.39,VALlbl=blank,fontVAL=&ftext,tickl=-0.5,ticks=0.15);
%tick(val= 1.61,VALlbl=5 ,fontVAL=&ftext);
%tick(val= 1.79,VALlbl=blank,fontVAL=&ftext,tickl=-0.5,ticks=0.15);
%tick(val= 1.95,VALlbl=blank,fontVAL=&ftext,tickl=-0.5,ticks=0.15);
%tick(val= 2.08,VALlbl=blank,fontVAL=&ftext,tickl=-0.5,ticks=0.15);
%tick(val= 2.2 ,VALlbl=blank,fontVAL=&ftext,tickl=-0.5,ticks=0.15);
%tick(val= 2.3 ,VALlbl=blank,fontVAL=&ftext);
%tick(val= 3.0 ,VALlbl=20 ,fontVAL=&ftext);

3- HOW TO DISPLAY THE SUB-GROUPS AND VARIABLES LABELS, HOW TO DRAW LINES BETWEEN LOWER AND UPPER CONFIDENCE LIMITS

* To put sub-groups and variables labels ;

Area :
%system(3,2);
if var^=1 then do ;
   if first.var then do;
      %label(5,cat+0.5,vardc,black,0,0,&size.,&ftext.,6) ;
   end;
end ;
if ^(type=2 & cat in(8)) then do ;
   %label(10,cat,catsnobs,black,0,0,&size.,&ftext.,6) ;
end ;

Math Font has to be used to put the symbol “<” and not “<=".
else if (type=2 & cat=8) then do;
  \%label(10,cat,'M' ,black,0,0,\&size.,'Math',6) ;
  \%label(12,cat,'65 years',black,0,0,\&size.,\&ftext.,6) ;
  \%label(20,cat,\ntext     ,black,0,0,\&size.,\&ftext.,6) ;
end ;

* To draw a line between the Lower and Upper limit ;

Area= Data, values

\%system(2,2) ;
if OR^=. then do ;
  \%line(logORLCL,cat,logORUCL,cat,black,1,0.5);
  \%line(logORLCL,cat-0.05,logORLCL,cat+0.05,black,1,0.5);
  \%line(logORUCL,cat-0.05,logORUCL,cat+0.05,black,1,0.5);
end ;

* Put the number of events (ev1 vs ev2) and OR [CI] for each sub-group;

\%system(3,2);
\%label(70,cat,nbev,black,0,0,\&size.,\&ftext.,6) ;
\%label(82,cat,ORCI,black,0,0,\&size.,\&ftext.,6) ;
run ;

3- PROC GPLOT

The Annotate macro facility will be used to display text as well as to draw lines in the required area of the graph. Some macros variables will be defined at first to fix the size and format of text.

* Initialization of annotate macros ;

\%let vsize=27.7 cm;
\%let hsize=20.5 cm;
\%let papersize=A4 ;

options orientation=portrait papersize="\&papersize.";
title h=0.5 cm "options landscape goptions rotate=portrait" ;
goptions reset=all
gunit=pct
noborder
htitle=2
htext=\&size.
cback=w
ftitle=\&ftitle.
ftext=\&ftext.
offshadow=(2,2)
rotate=portrait
hsize=\&hsize.
vsize=\&vsize.
autofeed;

* Set-up Axes and Plots

AXIS1 MINOR=NONE
MAJOR=NONE
ORDER=(0 TO 10 BY 1)
LABEL=NONE
value=NONE
offset=(0,7);

AXIS2 MINOR=NONE
MAJOR=NONE
ORDER=(-4.38 to 4.38)
LABEL=(h=3 c=white ' .')
VALUE=none
WIDTH=1 offset=(30,30) ;
* Create symbol for odds-ratio

symbol1 i=none value=dot color=black ;
filename gr "&flst_root.\Figure_X.emf ";
goptions gsfname=gr
device=emf
gsfmode=replace;

* Set-up titles and footnotes

*title1 j=l h=&tsize. "Title 1" ;
*title2 j=l h=&tsize. "Title 2" ;
*title3 j=l h=&tsize. "Title 3" ;
*title4 ;
title5 j=l h=&tsize. "Figure XXX: Odds-Ratios for Best Overall Response by subgroup - xxx Population";
footnote1 h=1 " ";

proc gplot data=final;
   plot cat*logOR/ VAXIS=AXIS1
            HAXIS=AXIS2
            HREF=0
            NOLEGEND
            ANNOTATE=anno;
run ;
QUIT;
## 4- FOREST- PLOT RESULTS

Figure XXX: Odds-Ratios for Best Overall Response by subgroup

<table>
<thead>
<tr>
<th>Subgroup (number of patients in Treatment X vs Treatment Y)</th>
<th>Number of responses</th>
<th>OR [95% CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Subjects (547 vs 544)</td>
<td>279 vs 231</td>
<td>1.408 [1.109, 1.789]</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 65 years (343 vs 348)</td>
<td>183 vs 139</td>
<td>1.710 [1.265, 2.310]</td>
</tr>
<tr>
<td>≥ 65 years (203 vs 196)</td>
<td>96 vs 92</td>
<td>1.011 [0.679, 1.505]</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (344 vs 324)</td>
<td>175 vs 145</td>
<td>1.277 [0.941, 1.733]</td>
</tr>
<tr>
<td>Female (203 vs 220)</td>
<td>104 vs 86</td>
<td>1.591 [1.079, 2.347]</td>
</tr>
<tr>
<td><strong>Ethnic Origin</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caucasian (469 vs 469)</td>
<td>238 vs 197</td>
<td>1.419 [1.096, 1.837]</td>
</tr>
<tr>
<td>Non caucasian (78 vs 75)</td>
<td>41 vs 34</td>
<td>1.336 [0.707, 2.527]</td>
</tr>
<tr>
<td><strong>Region</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Western Europe (236 vs 240)</td>
<td>119 vs 114</td>
<td>1.130 [0.789, 1.621]</td>
</tr>
<tr>
<td>Eastern Europe (188 vs 184)</td>
<td>100 vs 64</td>
<td>2.113 [1.387, 3.218]</td>
</tr>
<tr>
<td>Other (123 vs 120)</td>
<td>60 vs 53</td>
<td>1.205 [0.729, 1.994]</td>
</tr>
</tbody>
</table>
CONCLUSION

Building a Forest–plot with the Annotate facility is not an easy task, fortunately the macro facility helps a lot and use of a logarithmic scale enables the display of extreme values for 95% confidence interval values. The figure obtained allows the reader to visualize which odds-ratio is significant and hence determine which of the two treatments is better overall and also within each sub-group.

REFERENCES

SAS online V9.1.3

“How to annotate Graphics” PhUSE 2006 – Sandrine Stepien, Quintiles, Sidney, Australia

CONTACT INFORMATION (HEADER 1)
Your comments and questions are valued and encouraged. Contact the author at:
Laetitia Lemoine Le Halpert
Quintiles Paris
3-5 rue Maurice Ravel
92594 Levallois-Perret Cedex
Work Phone: +33 (0) 1 41 27 73 03
Fax: +33 (0) 1 41 27 72 72
Email: laetitia.lehalpert@quintiles.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.

Brand and product names are trademarks of their respective companies.