PharmaSUG 2016 - Paper DG03

Waterfall plot: two different approaches, one beautiful graph

Ting Ma, Pharmacyclics LLC, Sunnyvale, CA

ABSTRACT

This paper discusses two approaches to presenting oncologic data by use of a waterfall plot. One approach focuses on the statistical graphics (SG) family of procedure (SGPANEL) and the other utilizes the Graph Template Language (GTL). Both approaches are within the ODS Graphics system available with SAS 9.2 and later release. This paper will compare the specific SAS procedures to generate the visually effective waterfall plot. Basic plots will be built as the first step and then customized features will be added. These procedures should provide guidance to present stratified information from a complex clinical trial.

INTRODUCTION

Waterfall plots are gaining increasing popularity in oncology studies because it effectively demonstrates the best improvement in tumor size in relationship to the response to a drug seen in each subject. In such a plot, each patient in the trial is presented by a vertical bar on the plot and each bar represents the maximum change in the measurement of tumor. In trials with two arms, waterfall plots are often used to compare the outcome of active drug versus placebo.

Various techniques are available in SAS to produce waterfall plots. Here we describe two techniques. One utilizes SAS/GRAPH Statistical Graphics (SG) procedures, PROC SGPANEL. This procedure supports a simple and concise syntax for creating an attractive graph via the power of GTL behind the scenes. This procedure involves minimum coding, yet generates consistent quality of graph for reporting and publication. The other approach uses the GTL syntax directly, via PROC TEMPLATE and PROC SGRENDER. This system is powerful and flexible, ideal for creating complicated custom graphs with some learning curve. However, after a template is created, subsequent execution of the procedure can be easily carried out.

SNAP SHOTS OF INPUT DATASET

To get the effect of waterfall, the variable to be plotted on the vertical bar (*PCHG* in this dataset) has to be sorted by the descending order so that the largest value will show up on the left most bar and the smallest value on the right most bar. The input dataset should contain one observation per subject (**Figure1**). The dataset below is a fraction of a complete dataset used to generate figures in this paper. But the number of variables and the attributes of these variables need no change when we build individual plots in this paper.

	SUBJID	ORDER	PCHG	ARMN	BESTRIRC	BESTRIRCN	ARROW	TEXTNAME
1	511-002	1	-10	1	SD	6	.	.
2	512-001	2	-34.3	1	PR	4	.	.
3	513-001	3	-44.3	1	PR	4	.	.
4	514-004	4	-47.1	1	NPR	3	.	.
5	515-004	5	-49.9	1	CR	1	.	.
6	516-001	6	-50.6	1	PR	4	.	.
7	511-003	1	480	2	SD	6	175	480
8	512-003	2	152.2	2	SD	6	.	.
9	513-002	4	23.1	2	PD	7	.	.
10	566-001	5	21	2	PD	7	.	.

Figure 1: A snap shot of the first 10 boservations of input dataset

GENERATION OF THE BASIC PLOT

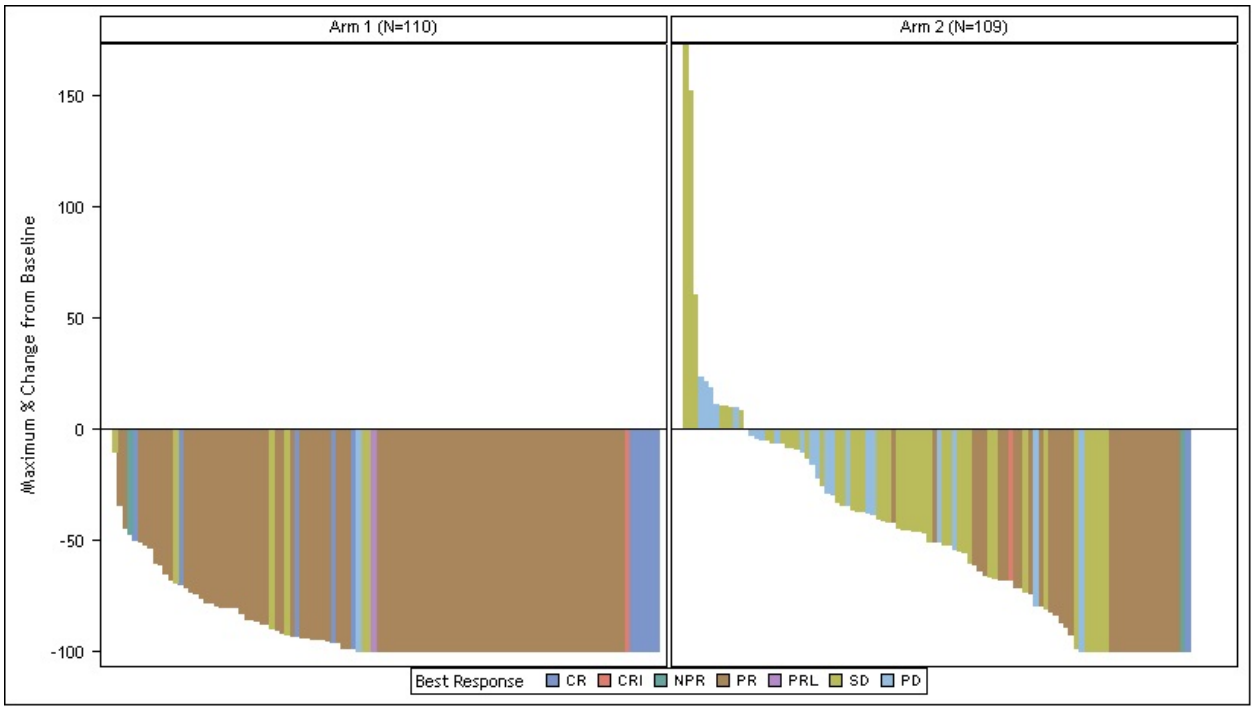

Figure 2: Basic water plot generated by PROC SGPANEL and GTL

- **BY PROC SGPANEL**

```
ods graphics on / width=1280 height=720;
ods rtf body='C:\Users\tma\Desktop\test\WaterFall\f_a1.rtf';

proc sgpanel data=final;
*========================*;
*== START OF  SECTION A  ==*;
*========================*;
   panelby armn /columns=2 novarname;
   vbar order /response=pchg GROUP=bestrircn BARWIDTH=0.6 legendlabel='Best
              Response' name='resp'  nooutline;    1
   colaxis display=none;
   rowaxis values=(-100 to 170 by 50) label="Maximum % Change from baseline";
   title "Maximum Percent Reduction from Baseline for Spleen Volume by Best
   Response" font="Courier";

   format armn armf.  bestrircn bestf.;           2
   label bestrircn = "Best Response ";
run;
ods graphics off;
*========================*;
*==  END OF  SECTION A   ==*;
*========================*;
```

1: The VBAR statement plots the variable *ORDER*. Subjects are grouped by variable *BESTRIRCN*. 2: The two formats here convert the numeric information in the arm and response variable to characters.

- **BY GTL**

```
ods graphics on /imagename="f_b1" imagefmt=emf width=1280 height=720;
ods rtf file="C:\Users\tma\Desktop\test\WaterFall\f_b1.rtf"
proc template;
   define statgraph waterfall;            3
   dynamic _ORDER _PCHG _ARM;
   dynamic _panelnumber_;
   begingraph / designwidth=960 designheight=720;

*=========================*;
*== START OF SECTION B ==*;
*=========================*;

   layout datalattice columnvar=_ARM / cellwidthmin=1 cellheightmin=1 rowgutter=3
                     columngutter=3 rowdatarange=unionall
                     headerlabeldisplay=value columnaxisopts=( display=none
                                                               altdisplay=none)
                     rowaxisopts=( label=('Maximum % Change from Baseline')
                     labelattrs=(family='Trebuchet MS' )
                     tickvalueattrs=(family='Trebuchet MS')
                     linearopts=(tickvaluesequence=(start=-100 end=200
                                                    increment=50)
                                                    viewmax=170));            4
*=========================*;
*==   END OF SECTION B   ==*;
*=========================*;

           layout prototype / ;
                 barchart x=_ORDER y=_PCHG / group=BESTRIRCN name='bar'
                 display=(FILL) stat=mean barwidth=0.95
                 groupdisplay=Cluster
                 clusterwidth=0.95;
           endlayout;                                          5

*=========================*;
*== START OF SECTION C   ==*;
*=========================*;
          sidebar / align=bottom spacefill=false;
                 discretelegend 'bar' / opaque=true border=true halign=center
                 valign=center  title='Bes Response ' displayclipped=true
                 order=rowmajor titleattrs=(family='Trebuchet MS' );
          endsidebar;          6
  endlayout;
  endgraph;
end;
run;

*=========================*;
*==   END OF SECTION C   ==*;
*=========================*;

*=========================*;
*== START OF SECTION D   ==*;
*=========================*;

proc sort data=final; by BESTRIRCN; run;
proc sgrender data=WORK.FINAL template=waterfall;                          7
     format armn armf. BESTRIRCN bestf.;
     dynamic _ORDER="ORDER" _PCHG="PCHG" _ARM="ARMN";
run;
```

```
 ods rtf close;
 ods graphics off;
*======================*;
*==   END OF SECTION D   ==*;
*======================*;
```

3: A new statgraph template called waterfall is defined. 4: DATALATTICE layout is chosen to create paneled graphs and the attributes of Xaxis and Yaxis are set. 5: LAYOUT PROTOTYPE is a special layout used by DATALATTICE to define the cell.The variable *PCHG* is plotted for each subject. *BESTRIRCN* is chosen as the grouping variable. 6: The legend is placed in the SIDEBAR statement. 7: The defined template WATERFALL is associated with the final input dataset.

CUSTOMIZATION OF THE COLOR ATTRIBUTES

One attribute we would like to modify on this basic plot is the color scheme of the bars. We would like to color the bars representing complete response (CR, CRi) using the range of blue while the PD using the range of red to make the visual effect more natural.

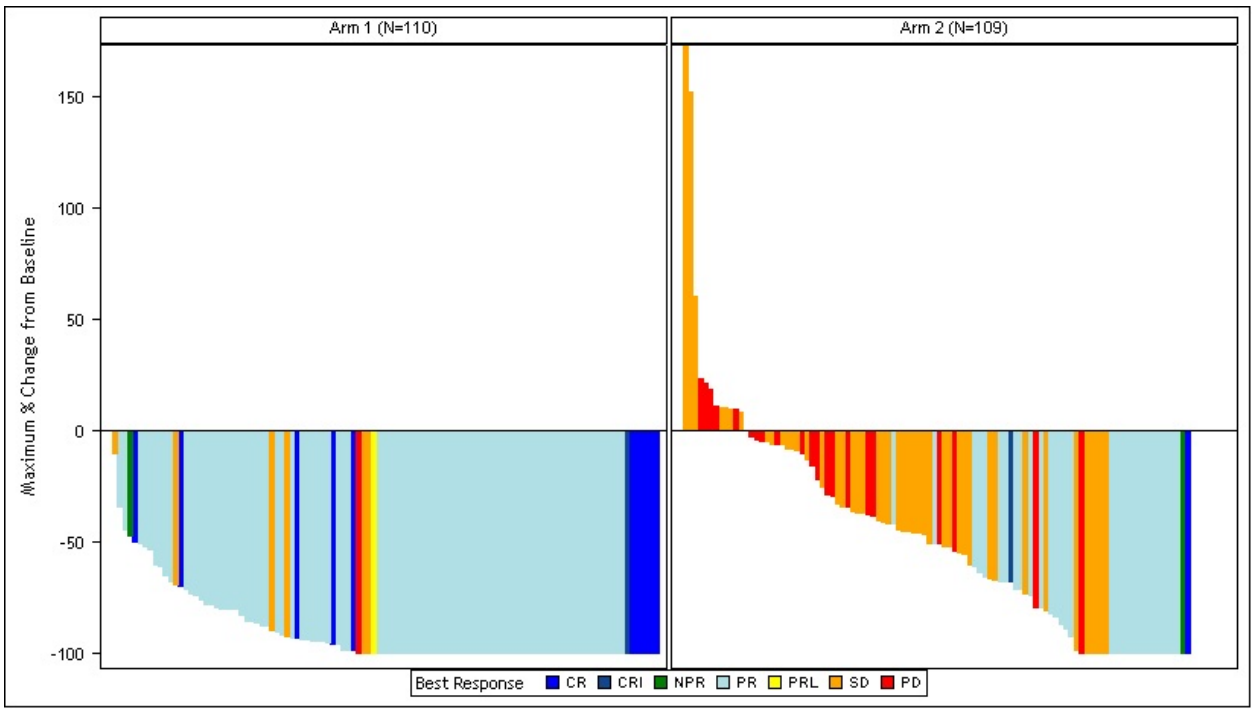

Figure 3: Waterfall plot with customized color attributes

- **BY PROC SGPANEL**

The use of GROUP option in VBAR statement led to the default grouping of the patients by the grouping variable, in this example, the variable *BESTRIRCN*. Group colors were then assigned by the GraphData1 – GraphData12 style elements from the active style in the order of their occurrence. To over write the default color attributes, we need to use PROC TEMPLATE to define a new style and apply the newly defined style in the PROC SGPANEL.

```
/*assign new color scheme */
proc template;
       define style MySurvStyle_a2;     8
       parent = styles.analysis;
       style graphdata1 from graphdata1/color=orange;
       style graphdata2 from graphdata2/color=powderblue;
       style graphdata3 from graphdata3/color=green;
       style graphdata4 from graphdata4/color=blue;
       style graphdata5 from graphdata5/color=red;
       style graphdata6 from graphdata6/color=yellow;
       style graphdata7 from graphdata7/color=vigb;
end;

ods graphics on / imagename="f_a2" imagefmt=emf width=1280 height=720;
ods rtf body='C:\Users\tma\Desktop\test\WaterFall\f_a2.rtf' style=MySurvStyle_a2;

proc sgpanel data=final;
**-------------------------------------------------------------**;
* Codes identical to SECTION A are omitted here */
**-------------------------------------------------------------**;
```

8: The predefined color attributes in STYLES.ANALYSIS are modified to fit our needs.

- **BY GTL**

Same as the PROC SGPANEL procedure, the default attributes are assigned to the groups depending on the occurrence of the data. To assign more visually friendly attributes, discrete attribute map is used to refine the graph and specify the color you want to associate with the group.

```
proc template;
       define statgraph waterfall;
       dynamic _ORDER _PCHG _ARM;
       dynamic _panelnumber_;
       begingraph / designwidth=960 designheight=720;

          /*assign new color scheme */
         discreteattrmap name='colors' / ignorecase=true;          9
             value "CR" / fillattrs=(color=blue);
             value "CRI" / fillattrs=(color=vigb);
             value "NPR" / fillattrs=(color=green);
             value "PR" / fillattrs=(color=powderblue);
             value "PRL" / fillattrs=(color=yellow);
             value "SD" / fillattrs=(color=lilac );
             value "PD" / fillattrs=(color=red);
             value "UNK/NE" / fillattrs=(color=purple);
        enddiscreteattrmap;

       discreteattrvar attrvar=barcolors var=bestrirc attrmap='colors';

**-------------------------------------------------------------**;
/* Codes identical to SECTION B are omitted here */
**-------------------------------------------------------------**;
     layout prototype / ;
             barchart x=_ORDER y=_PCHG / group= barcolors  name='bar'
             display=(FILL) stat=mean barwidth=0.95
             groupdisplay=Cluster
             clusterwidth=0.95;
     endlayout;                                                  10

**-------------------------------------------------------------------**;
 /* Codes identical to SECTION C and D are omitted here */
**-------------------------------------------------------------------**;
```

9: A discrete attributes map is used to define the preferred color. The association between the grouping variable *BESTRIRC* and the intermediate mapping variable *BARCOLORS* is established. 10 The intermediate mapping variable *BARCOLORS* replaces the original grouping variable for the new color assignment to take effect.

ANNOTATION

We notice that the data contains an outlier – a patient with a significant spleen volume enlargement reaching the maximum view point of the Y axis. In fact the value of *PCHG* for this patient is 480. Instead of accommodating this data point by extending the Y axis to a significantly large number for one subject, we decide to use the annotation facility to denote the outlier subject. The DATA step below marks the specific bar with a plus sign at the top of the bar and exhibits the actual value next to the bar.

- **BY PROC SGPANEL**

--Modification of the original input dataset:

```
data final;                                                    11

      set final;
      if subjid = “511-003” then do;
            pchg = 170;
      end;
run;
```

11: The original dataset is modified so that the bar for the outlier is cut short at the value of 170.

--Creation of annotation dataset:

```
*========================*;
*== START OF SECTION F   ==*;
*========================*;
data anno;                                                    12
retain drawspace "datapercent" x1space “graphpercent” y1space “graphpercent”  textsize
11 textcolor "black" ;
length function $ 9 ;
input function $ x1 y1 textsize label$;
datalines;
text 55.5 85 9 +
text 58 85 9 480;
run;
*========================*;
*== END OF SECTION F        ==*;
*========================*;
```

12: A plus sign and the number “480” will be displayed as texts at the specified position.

--Proc sgpanel procedure:

```
ods graphics on / width=1280 height=720;
ods rtf body= 'C:\Users\tma\Desktop\test\WaterFall\f_a3.rtf';
proc sgpanel data=final sganno=anno;


**-----------------------------------------------------------------**;
/* Codes identical to SECTION A are omitted here */
**-----------------------------------------------------------------**;
```

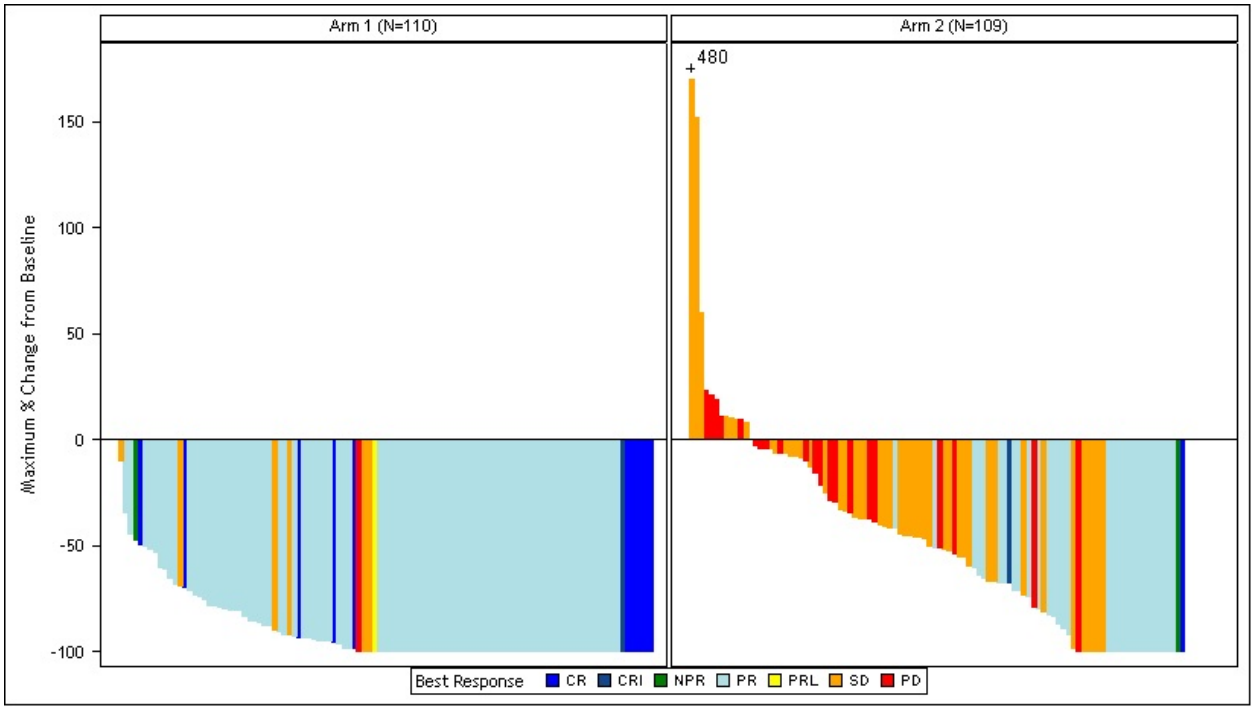

Figure 4: Waterfall plot with annotation

- **BY GTL**

Method 1:

The annotation can be achieved by creating two new variables in the input dataset for the outlier and its label (variable *ARROW* and *TEXTNAME* in **Figure 1**). A scatter plot is then overlaid with the original bar chart in the same layout container. This method does not need an annotation dataset and works well for SAS9.3 and earlier releases.

--Modification of the original input dataset:

```
data final;                                                    13
       set final;
       if subjid = “511-003” then do;
              pchg = 170;
              arrow = 175;              /*The position of the marker*/
              textname = "480";   /* The value to show in label */
       end;
run;
```

13: The original dataset is modified so that the bar for the outlier is cut short at the value of 170 and the actual value is displayed next to the bar (*TEXTNAME* in **Figure 1**).

--Modification of the original template:

```
ods graphics on /imagename="f_b3" imagefmt=emf width=1280 height=720;
ods rtf file="C:\Users\tma\Desktop\test\WaterFall\f_b3.rtf"
proc template;
       define statgraph waterfall;

       dynamic _ORDER _PCHG _ARM;
       dynamic _panelnumber_;
```

```
      begingraph / designwidth=960 designheight=720;
**------------------------------------------------------------**;
/* Codes identical to SECTION B are omitted here */
**------------------------------------------------------------**;
      layout prototype ;
            barchart x=_ORDER y=_PCHG / group=barcolors name='bar' display=(FILL)
            stat=mean barwidth=0.95 groupdisplay=Cluster
            clusterwidth=0.95 index=bestrircn;
            scatterplot x = _ORDER y = arrow/   markerattrs=(size=10 symbol=plus
            color=black )  datalabel=textname datalabelattrs=(color=black size=10pt);
                                                                              14
      endlayout;

**------------------------------------------------------------------**;
/* Codes identical to SECTION C and D are omitted here */
**------------------------------------------------------------------**;
```

14: The plus sign and actual value are drawn on the plot by scatterplot statement.

Method 2:

An annotation dataset is created and specified in the relevant plotting layout block in the template. The graph is then rendered with the annotation in the SGRENDER procedure.

--Creation of annotation dataset:

```
**------------------------------------------------------------**;
/* Code identical to SECTION F are omitted here */
**------------------------------------------------------------**;
```

--Modification of the original template:

```
ods graphics on /imagename="f_b4" imagefmt=emf width=1280 height=720;
ods rtf file="C:\Users\tma\Desktop\test\WaterFall\f_b4.rtf"
proc template;
      define statgraph waterfall;

      dynamic _ORDER _PCHG _ARM;
      dynamic _panelnumber_;
      begingraph / designwidth=960 designheight=720;
**------------------------------------------------------------**;
 /* Code identical to SECTION B are omitted here */
 **------------------------------------------------------------**;

layout prototype / ;
        barchart x=_ORDER y=_PCHG / group=barcolors name='bar' display=(FILL)
                                    stat=mean barwidth=0.95 groupdisplay=Cluster
                                   clusterwidth=0.95 index=bestrircn;
        annotate;                                                          15
endlayout;
**------------------------------------------------------------**;
/* Codes identical to SECTION D are omitted here */
**------------------------------------------------------------**;
proc sort data=final; by BESTRIRCN; run;
proc sgrender data=WORK.FINAL template=waterfall sganno=anno;              16
            format armn armf. BESTRIRCN bestf.;
            dynamic _ORDER="ORDER" _PCHG="PCHG" _ARM="ARMN";

ods rtf close;
ods graphics off;
```

15: Annotate statement is placed in the same layout container as the barchart statement.
16: Include the sganno=anno option to render the graph with the annotat

CONCLUSION:

SAS provides the flexibility to produce waterfall plots using different techniques. Custom waterfall plots can be created with PROC SGPANEL in combination with annotation dataset and modification of template. When custom graph gets detailed and complex GTL syntax provides an alternative tool to accomplish the goal.

REFERENCES:

Sanjay Matange. “Introduction to the Graph Template Language”. Paper presented at SAS Global Forum 2008.
Heath, Dan. “Effective Graphics Made Simple Using SAS/GRAPH SG Procedures.” Paper presented at SAS Global Forum 2008.
Nora H. Ruel. “Graphical Results in Clinical Studies A focus on graphics used in oncology”. Paper presented at WUSS2015.

CONTACT INFORMATION

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

Name: Ting Ma
Enterprise: Pharmacyclics, LLC
Address: 999 East Arques Ave, Sunnyvale, CA 94085
Work Phone: (408) 215-3577
E-mail: tma@pcyc.com