PhUSE 2010

Paper SP08

Waterfall plots: A beautiful and easy way of showing a whole picture of an interesting outcome

Antonio Nieto, PharmaMar, Madrid, Spain
Javier Gómez, PharmaMar, Madrid, Spain

ABSTRACT

Waterfall plots are becoming very useful nowadays in the oncology field due to their simplicity in understanding and their capability to show interesting outcomes in the whole population. Waterfall plots can be used to display outcomes such as: variation during the treatment in the sum of the longest diameters of the lesions in solid tumours trials, variation in tumor markers like the prostate-specific antigen (PSA) in prostate trials or CA125 in ovarian trials, etc.
This paper describes how to produce a waterfall plots by using the SAS® v8.2 or v9.x system.

INTRODUCTION

A waterfall plot is an ordered chart where each subject is symbolized by a vertical bar, which represents the maximum change with respect to a reference evaluation obtained during a specific period. In our example, the waterfall plot shows the largest variation in the PSA marker during treatment for each subject enrolled in a prostate cancer trial. Usually, the waterfall plots go from the worst value (e.g., highest growth) on the left side to the best value (e.g., highest reduction) on the right side.

OVERVIEW OF %WATERFALL MACRO

The waterfall plot macro will be explained using an example. The outcome of subjects who were treated with a marine-derived anticancer agent, in a prostate cancer trial will be summarized. For each patient, the outcome of interest was best PSA variation during treatment defined as the variation in % between the lowest PSA value obtained during treatment vs. the PSA value at baseline.

The first step should be done out of the macro and consists of the preparation of the SAS® dataset. The SAS® dataset only needs a row for each subject and the variable which contains the value to be plotted in the vertical bar.

To maintain the simplicity and the flow, an ideal example is followed but slight adjustments to the code may be necessary to fulfill other potential needs or preferences, such as applying different colors to the bars according to subgroups or dealing with a situation where all the patients may have had negative or positive values, etc.

Macro parameters are: *dsn*; SAS dataset name, *var*; variable to be printed, *outfile*; graphic output name including the path, and *title*; the chart title name.

Our macro can be splitted in four parts:

1-. Edition process.

2-. Annotate dataset.

3-. Plot.

4-. Removal of the auxiliary datasets.

%WATERFALL MACRO

Once the dataset containing one row per subject has been created then we are ready to submit the macro program.

%waterfall(dsn=PSA_variation, var=min_dif_baseline, outfile="PSA waterfall.gif", title="Best PSA variation (%) during the treatment");

1-.Edition process.

First all of the processes of sorting, editing and creation of auxiliary variables are done. The macro puts in descending order the original dataset by the values to be plotted. Auxiliary variables to count the total number of subjects and the number and percentage of subjects with increment/decrement are created. Axis y format values are defined to have in the vertical axis the absolute value.

```
* The original dataset is sorted. Auxiliary variables are added;;

Proc sort data=&dsn;by descending &var;run;
Data aux_waterfall1;
set &dsn;
flag=1              ;/*Auxiliary variable to create the plot, the bars will be added using the anno dataset*/
count=_n_;
If &var>0 then variation=0;                 /*Auxiliary variable for bars with positive values*/
else If &var<=0 then variation=1;           /*Auxiliary variable for bars with negative values*/
run;

* Calculation of total number of subjects;

PROC SQL NOPRINT;
SELECT max(count) INTO:n_subjid FROM aux_waterfall1;
quit;

*Calculation of the number of subjects/percentage by each category (increase and decrease).
* The total number of patients is added to the graph title;

ods exclude all;
Proc freq data=aux_waterfall1;
table variation/missing out=aux_waterfall2 sparse;
run;

Data aux_waterfall3;
length title1 $55.;
set aux_waterfall2;
part1=trim(&title);
part2=trim(compbl(" (n="||(&n_subjid)||")"));
title1=part1||part2;                                ;/*To add the number of patients to the title*/
  call symput('title2',title1);
PERCENT1=compbl(round(PERCENT,.1)||"%");
If variation=0 then do;
  call symput('n_increase',COUNT);
  call symput('n_increaseb',COUNT+0.5)              ;/*X coordinate to plot a href to split each group*/
  call symput('increase_percent',PERCENT1);
end;
If variation=1 then do;
  call symput('n_decrease',COUNT);
  call symput('decrease_percent',PERCENT1);
end;run;
```

```
*Format of the Y axis;
*A discrete axis with formats is created to have increases and decreases from 0 to 100 by 10  in the Y axis;

Data aux1;
length count $3.;
do i=-120 to 120 by 10;                    /*20% offset  to have enough space for  annotations in the graph */
If -120<=i<-100 then count=compress('');
else If -100<=i<0 then count=compress(-i);
else If 0<=i<=100 then count=compress(i);
else If 100<i<=120 then count=compress('');
output;
end;
run;

Data aux2 (rename=(i=start count=label));
set aux1;
retain fmtname "yformatb";
run;

proc format library=work cntlin=aux2;
run;
```

2-. Annotate dataset.

An annotate dataset is created by combining five auxiliary datasets. The first auxiliary dataset captures the creation of the vertical bars for each subject. The second and third datasets are the summaries of the percentage of subjects with increment/decrement. The fourth dataset creates a flag for the subjects who have an increment higher than one hundred percent. The last one is an editing dataset to draw the ticks out of the range (-100, 100).

```
* (1) Waterfall bars;

data anno1;
length function color $8;
retain xsys ysys '2' size 4 color '' line 1;
set aux_waterfall1;
color='red';line=1;                /* Create solid lines in red color */
If &var<=100 then do;              /* Create bars for negative values and positive values <=100%*/
function='move';
y=0;
x=count;
output;
function='draw';
y=&var;
x=count;
output;
end;
If &var>100 then do;               /* For positive values >100%, the bar is drawn until 100%*/
function='move';                   /* The arrows will be drawn later */
y=0;
x=count;
output;
function='draw';
y=100;
x=count;
output;
end;run;
```

```
* (2) Labels with the percentage of total variation and arrowheads to be drawn in the bottom of the graph;

data anno2;
length color function style $50 ;
length text $ 100;
retain function 'label' xsys ysys '2';

y = -107;                          /* Label for percentage of subjects with increase */
X = &n_increase/2;
style='swiss';
text="&increase_percent";
color='black';
size=1.5;
output;

y = -107;                          /* Label for percentage of subjects with decrease */
X = &n_increase+&n_decrease/2;
style='swiss';
text="&decrease_percent";
color='black';
size=1.5;
output;

y = -112;                          /* Arrowheads */
X = 0.5;
style='swiss';
text="<";
color='black';
size=1.5;
output;

y = -112;
X = &n_increase;
style='swiss';
text=">";
color='black';
size=1.5;
output;

y = -112;
X = &n_increase+1.5;
style='swiss';
text="<";
color='black';
size=1.5;
output;

y = -112;
X = &n_subjid;
style='swiss';
text=">";
color='black';
size=1.5;
output;

run;
```

```
* (3) Lines for the  arrows to be drawn in the bottom of the graph;

data anno3;
length function color $8;
retain xsys ysys '2' size 1 color '' line 1;

color='black';

line=1;                                 /* Arrow line for subjects with increase */
function='move';
y=-115;
x=0.5;
output;
function='draw';
y=-115;
x=&n_increase-0.5;
output;

line=1;                                 /* Arrow line for subjects with decrease */
function='move';
y=-115;
x=&n_increase+1.5;
output;
function='draw';
y=-115;
x=&n_subjid-0.5;
output;

run;

* (4) Arrows for bars with Increase>100%;

data anno4;
length color function style $50 ;
length text $ 10;
retain function 'label' xsys ysys '2';
set aux_waterfall1;

If &var>100 then do;
y = 105;
X = count;
style='MARKER';
text='K';
color='black';
size=1;
output;
end;

run;

* (5) Hide ticks higher than |100| (white lines of size 0.12 are overlapped) ;

data anno5;
length function style color $8;
retain xsys '5' ysys '2'  when 'a' style 'solid';

color='white'; size=1;
```

```
function='move'; xsys='1'; x=-0.12;  ysys='2'; y=110; output;
function='draw'; xsys='B'; ysys='B'; x=-3; y=0; output;

function='move'; xsys='1'; x=-0.12;  ysys='2'; y=105; output;
function='draw'; xsys='B'; ysys='B'; x=-3; y=0; output;

function='move'; xsys='1'; x=-0.12;  ysys='2'; y=-105; output;
function='draw'; xsys='B'; ysys='B'; x=-3; y=0; output;

function='move'; xsys='1'; x=-0.12;  ysys='2'; y=-110; output;
function='draw'; xsys='B'; ysys='B'; x=-3; y=0; output;

function='move'; xsys='1'; x=-0.12;  ysys='2'; y=-115; output;
function='draw'; xsys='B'; ysys='B'; x=-3; y=0; output;

function='move'; xsys='1'; x=-0.12;  ysys='2'; y=-120; output;
function='draw'; xsys='B'; ysys='B'; x=-3; y=0; output;

run;

* (1-5) All annotate datasets are joined;

Data anno;
format function style color $8. text $100.;
set anno1 anno2 anno3 anno4 anno5;
run;
```

3-. Plot.

Graph definition is outlined including graphic options, output file, graphic device, axes definition, title, and the final gplot procedure.

```
ods select all;filename outfile &outfile;
goptions RESET=ALL gsfmode=replace gsfname=outfile device=gif ftext="Arial" htext=1;

SYMBOL1 value=none;
AXIS1
   MAJOR=(number=1)
   MINOR=(number=1)
   label= (angle=90 h=1.5 "Decrease                    Increase")
          offset=(1,1)
          order= (-120 to 110 by 10) ;
AXIS2
          label=none
          offset=(0.25,0.25)
          value=none
   MAJOR=none
   MINOR=none
          order= 0 to &n_subjid+1;

TITLE JUSTIFY=CENTER HEIGHT=14pt "&title2";

Proc gplot data=aux_waterfall1;
plot flag*count/
           haxis = axis2 href=&n_increaseb lhref=34
           vaxis = axis1 vref=0 anno=anno;
format flag yformatb.;
run;quit;title;
goptions RESET=ALL;
```

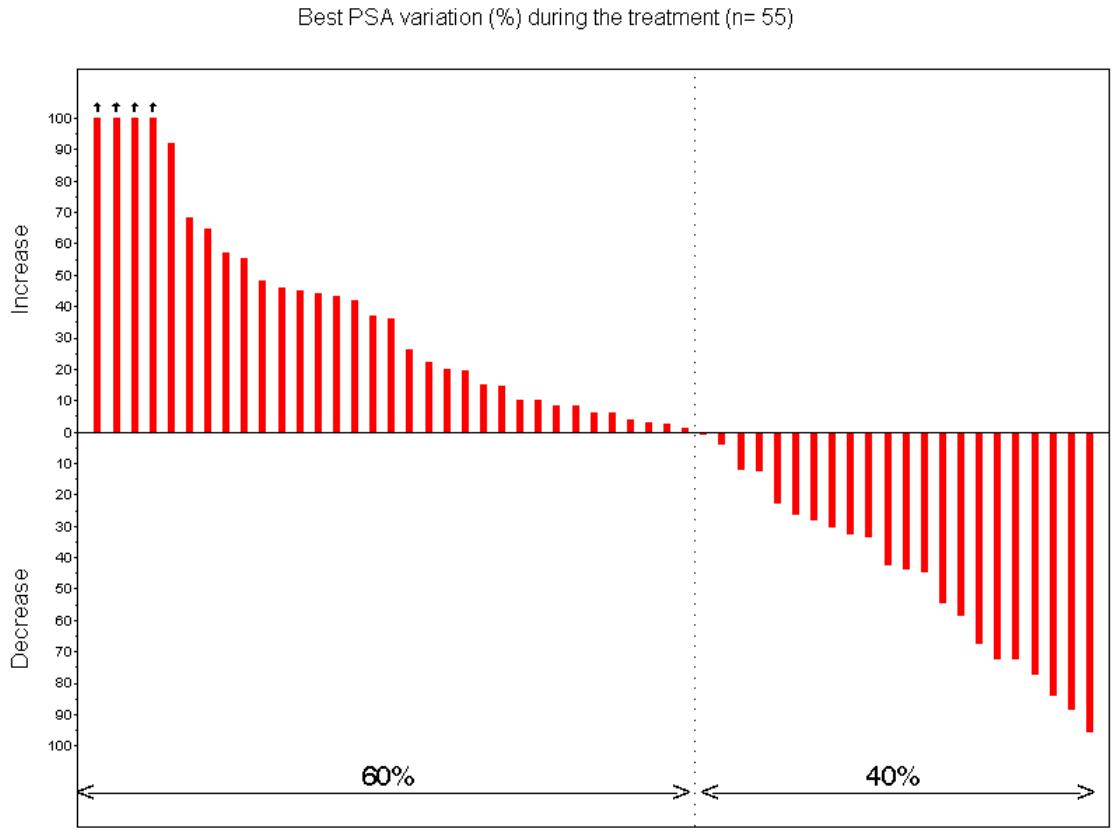
4-. Removal of the auxiliary datasets.

Elimination of the auxiliary datasets created during the macro submission to clean the working library.

```
proc datasets library=work nodetails;
delete aux_waterfall1 aux_waterfall2 aux_waterfall3 aux1 aux2 anno anno1 anno2 anno3 anno4 anno5;
run;
quit;

%mend waterfall;
```

PLOT



CONCLUSION

Waterfall plots are easy to generate and interpret. They represent a useful tool to fully capture the magnitude of the intrasubject variation of a parameter of interest and provide an attractive way to illustrate the global picture of the efficacy/ safety outcomes examined in a clinical trial in the overall population.

REFERENCES

-SAS online doc.

CONTACT INFORMATION

Antonio Nieto Archilla
Clinical Development. PharmaMar S.A.
Avda. de los Reyes, 1
Polígono Industrial La Mina
28770 Colmenar Viejo. Madrid (SPAIN)
anieto@pharmamar.com

Javier Gómez García
Clinical Development. PharmaMar S.A.
Avda. de los Reyes, 1
Polígono Industrial de la Mina
28770 Colmenar Viejo. Madrid (SPAIN)
jgomez@pharmamar.com