MWSUG 2017 - Paper DV02

Waterfall Plots in Oncology Studies in the Case of Multi-Arms Design

Ilya Krivelevich, Eisai Inc., Woodcliff Lake, NJ
Kalgi Mody, Eisai Inc., Woodcliff Lake, NJ
Simon Lin, Eisai Inc., Woodcliff Lake, NJ

ABSTRACT

Clinical data are easier to understand when presented in a visual format. In Oncology, in addition to the commonly used survival curves, other types of graphics can be helpful in describing response in a study. These plots are becoming more and more popular due to their easy-to-understand representation of data. Waterfall plots can help to visualize tumor shrinkage or growth; in such plots, each patient in the study is presented by a vertical bar on the plot and each bar represents the maximum change in the measurement of tumors. In the studies with two arms, waterfall plots are often used to compare the outcome between arms.

The excellent ground for understanding waterfall plots is proposed in the article of Theresa W. Gillespie, PhD, MA, RN: Understanding Waterfall Plots, Journal of the Advanced Practitioner in Oncology, 2012 Mar-Apr. This article claims that “A study using a randomization scheme other than 1:1 will not lend itself as well to a waterfall plot technique. As stated previously, since each vertical plot represents a single patient, waterfall plots limit the ability to portray different randomization schemes, e.g., 2:1 or 3:1”. This presentation shows how we can solve this problem by new techniques, using PROC SGPANEL and Graph Template Language.

INTRODUCTION

In Oncology, in addition to the commonly used survival curves, other types of graphics can be helpful in describing responses in a study. Waterfall Plots are gaining increasing popularity. These figures help to visualize tumor shrinkage or growth and effectively demonstrate the best improvement in tumor size in relationship to the response to a drug taken by each subject.

To get the effect of a waterfall in the presentation, the flow of the plot has to be from the worst value on the left to the best value on the right side of the plot among the subject population (descending order).The data should be sorted in the descending order of the variable, representing percentage of the best change from baseline. Each patient is also assigned to different categories based on the patient’s Best Overall Response. According to RECIST criteria, the Best Overall Response can have one of the values below:

- CR - Complete Response
- PR - Partial Response
- SD - Stable Disease
- PD - Progressive Disease
- NE - Not Evaluable

Plotting the maximum change in the sum of diameters of target tumors for each patient and coloring each patient’s bar based on the above mentioned categories for the Best Overall Response gives us a very clear picture of the performance of the experimental drug.

Various techniques, including Graph Template Language, are available in SAS® to get the required figures.

When a study compares two drugs for the responses measured, waterfall plots give a broader picture of the performance of both drugs. While previously proposed techniques work well on randomization schemes with 1:1 ratio between the two arms (with equal numbers of subjects in each arm), they have difficulties in portraying different randomization schemes, e.g., 2:1 or 3:1. Even in the case of 1:1 randomization schemes with unequal numbers of subjects in each arm, we may experience difficulties in portraying the data. In this article, we are going to show how this problem can be solved and how to create effective waterfall plots for visual comparison of two arms regardless of randomization ratio.

DATA USED IN PRESENTATION

Suppose that a randomization scheme of 2:1 was implemented in the study (the data used in this presentation were selectively copied from the PharmaSUG 2012 (Paper DG13) presentation and courtesy of Niraj J. Pandya: “Waterfall Charts in Oncology Trials - Ride the Wave” with subsequent fake randomization added for two arms).

Obs	usubjid	pchg	resp	Arm
1	1001	10	PD	Competitor
2	1002	12	PD	Study Drug
3	1003	-60	PR	Study Drug
4	1005	-72	PR	Competitor
5	1006	-42	PR	Study Drug
6	1010	-14	SD	Competitor
7	1012	-48	PR	Competitor
8	1013	-36	PR	Study Drug
9	1018	53	PD	Study Drug
10	1021	-37	PR	Competitor
11	1022	-43	PR	Study Drug
12	1023	50	PD	Study Drug
13	1024	-81	PR	Competitor
14	1026	-33	PR	Study Drug
15	1027	-83	PR	Competitor
16	1028	-34	PR	Study Drug

Obs	usubjid	pchg	resp	Arm
17	1029	-94	PR	Study Drug
18	1030	-14	SD	Study Drug
19	1031	-16	SD	Competitor
20	1032	-86	PR	Study Drug
21	1033	-89	PR	Competitor
22	1034	-13	SD	Study Drug
23	1036	13	SD	Study Drug
24	1037	-55	PR	Study Drug
25	1038	-86	CR	Study Drug
26	1039	-27	SD	Study Drug
27	1040	-20	SD	Study Drug
28	1043	-16	PD	Study Drug
29	1047	-83	PR	Competitor
30	1048	-41	PR	Study Drug
31	1049	-47	PR	Competitor
32	1052	0	SD	Study Drug

The total number of subjects was 32; 21 of them were randomized into one arm (named above as ‘Study Drug’) and 11 of them were randomized into another arm (named above as ‘Competitor’).

Before we start using PROC SGPLOT, we need to make an adjustment to the input data.

In the figure to be produced, the legend is supposed to specify the Best Overall Response. As was mentioned above, according to RECIST criteria, the Best Overall Response can have one of the values below:

- CR - Complete Response
- PR - Partial Response
- SD - Stable Disease
- PD - Progressive Disease
- NE - Not Evaluable

To display possible values in the required order (‘CR’, ‘PR’, ‘SD’, ‘PD’, ‘NE’) in the legend, the new formatted variable num_resp (the numeric equivalent of variable resp) was introduced:

```
proc format;
  value responses
  1 = 'CR'
  2 = 'PR'
  3 = 'SD'
  4 = 'PD'
```

```
    5 = 'NE'
    ;
  run;

  data two_arms;
    set two_arms;
    if resp = 'CR' then num_resp  = 1;
    if resp = 'PR' then num_resp  = 2;
    if resp = 'SD' then num_resp  = 3;
    if resp = 'PD' then num_resp  = 4;
    if resp = 'NE' then num_resp  = 5;
  run;

  data two_arms;
    set two_arms;
    format num_resp responses. ;
  run;
```

DRAWING WATERFALL PLOTS OVER POOLED POPULATION (REGARDLESS OF ARM)

As was mentioned above, the displayed bars should be sorted in the descending order by maximum percent change from baseline (by variable pchg). The new variable order_id was introduced as:

```
  proc sort data = two_arms out = two_arms_1;
    by descending pchg;
  run;

  data two_arms_1;
    set two_arms_1;
    order_id = _n_;
  run;
```

To control the appearance of the colors, we need to create an attribute map:

```
  data waterfall_attrmap_a;
    length id $11;
    input id $ value $ fillcolor $ linecolor $;
    cards;
    waterfall_a CR red red
    waterfall_a PR green green
    waterfall_a SD blue blue
    waterfall_a PD gray gray
    waterfall_a NE orange orange
    ;
  run;
```

The SAS code below will produce the required figure, Figure 1:

```
  ods graphics
    /
    noborder
    height = 6.5 in
    width = 9 in
    ;

  ods pdf file = 'R:\projects\...\presentation_a.pdf' notoc;
    proc sgplot data = two_arms_1 dattrmap = waterfall_attrmap_a;
```

```
    title1
      font = 'SAS Monospace' color = black height = 9 pt justify = left
      'Eisai Protocol: Exxxx-yyyy-zzz' justify = right 'Page 1 of 1';
    title2
      font = 'SAS Monospace' color = black height = 10 pt
      justify = center 'Figure 14.x.y.z';
    title3
      font = 'SAS Monospace' color = black height = 10 pt
      justify = center
      'Percentage Change from Baseline in Sums of Diameters';
    title4
      font = 'SAS Monospace' color = black height = 10 pt
      justify = center 'xxx Analysis Set';
    footnote1
      font = 'SAS Monospace' color = black height = 9 pt justify = left
      'First footnote';
    footnote2
      font = 'SAS Monospace' color = black height = 9 pt justify = left
      'Second footnote';
    vbarparm category = order_id response = pchg
      /
      barwidth = 1
      group = num_resp
      grouporder = ascending
      attrid = waterfall_a
      ;
    Keylegend
      /
      location = outside
      position = bottom
      noborder
      title = 'Best Overall Response'
      down = 1
      titleattrs = (size = 10 pt weight = bold)
      valueattrs = (size = 10 pt)
      ;
    refline 20 -30 / lineattrs = (color = black pattern = dot);
    refline 0 / lineattrs = (color = black thickness = 1 pattern = 1);
    xaxis display = none;
    yaxis
      label = 'Best Change from Baseline (%)' labelattrs = (size = 10 pt)
      values = (-100 to 100 by 20) valueattrs = (size = 9 pt);
  run;
ods pdf close;
```

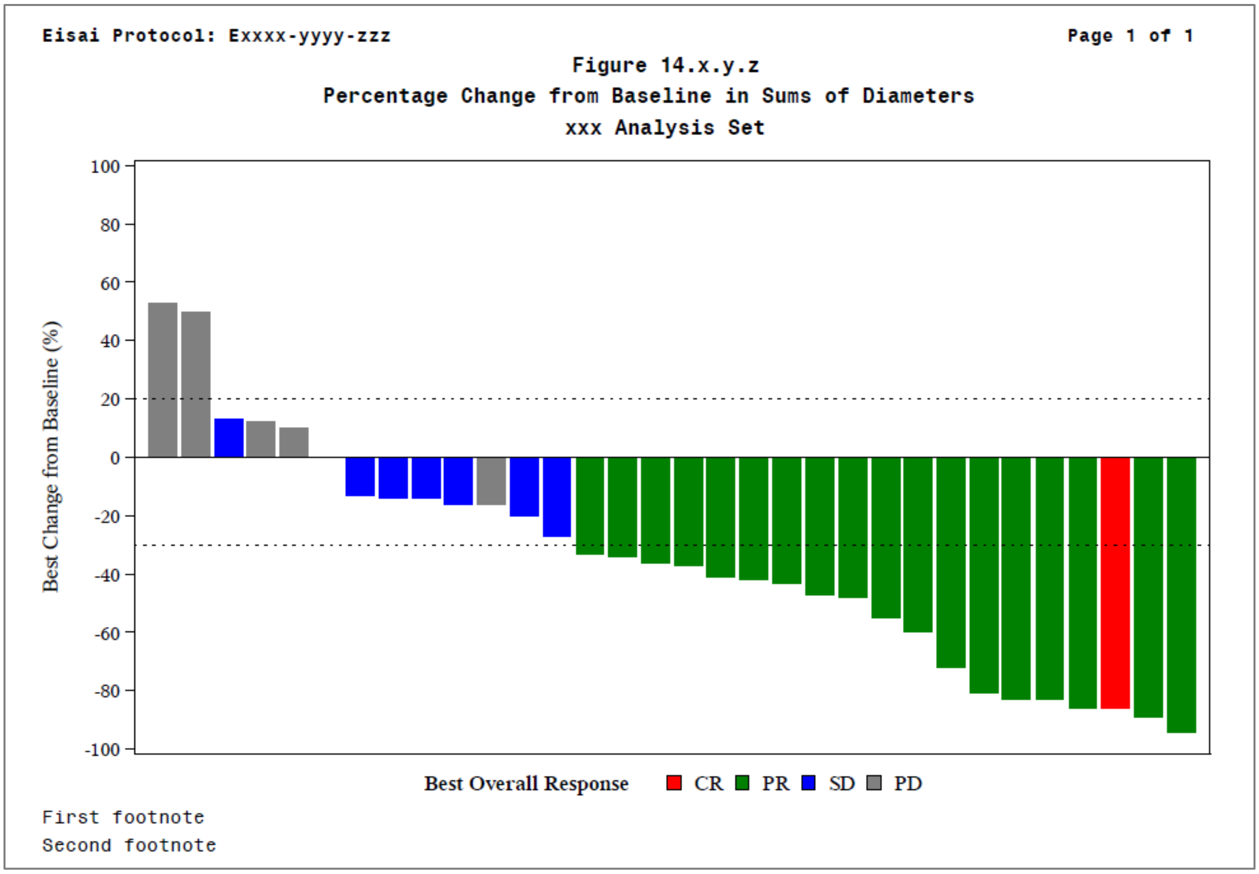

Figure 1: Drawing waterfall plots over pooled population (regardless of arm)

DRAWING WATERFALL PLOTS OVER POOLED POPULATIONS WITH DISPLAYED SUBJECTS' IDS

Sometimes, when the number of subjects in a study is reasonably small, we might be required to identify each bar with a subject ID. To make this possible, we might use techniques described by Stan Li in the presentation "Some Useful Techniques of Proc Format" (PharmaSUG 2011 - Paper CC19). We will format each value of the variable order_in with the value of its subject ID::

```
proc sort data = two_arms out = two_arms_2;
  by descending pchg;
run;

data two_arms_2;
  set two_arms_2;
  order_id = _n_;
run;

data new_format;
  set two_arms_2;
  retain fmtname 'x_subj';
  length label $9;
  start = order_id;
  label = 'ID = ' || strip (put (usubjid, best.));
run;

proc format cntlin = new_format;
run;
```

```
data two_arms_2;
  set two_arms_2;
  format order_id x_subj.;
run;
```

Only one statement needs to be changed from the previous section:

```
xaxis display = (nolabel) fitpolicy = rotate;
```

As a result of this minor change, we will get the output below (Figure 2):

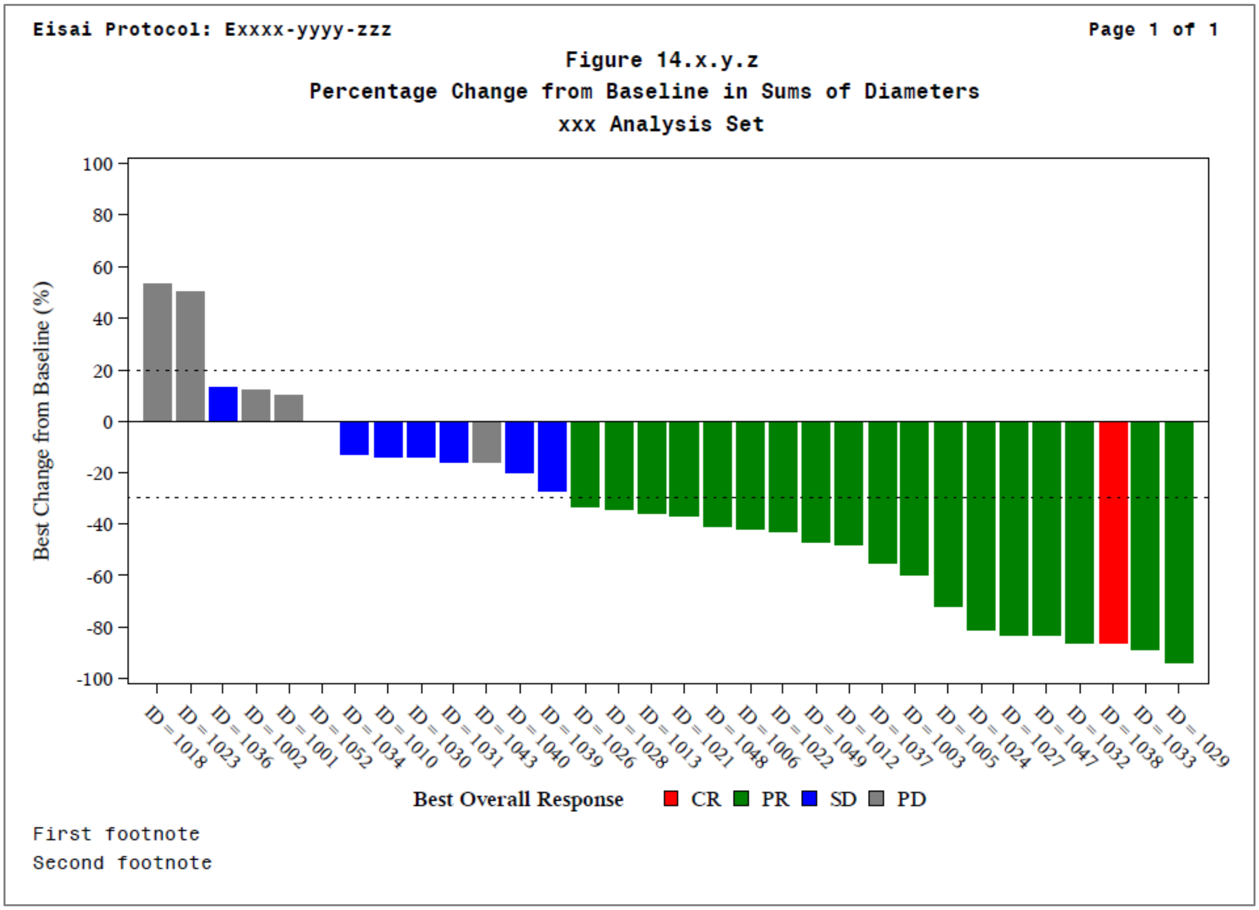

Figure 2: Drawing waterfall plots over pooled populations with displayed subjects' IDs

DRAWING WATERFALL PLOTS BY ARM (TWO COLUMNS, ONE ROW) USING PROC SGPANEL

PROC SGPANEL provides us with an excellent way to display (as a unique figure) Waterfall Plots by arm.

First, we need to sort the input data by arm with descending pchg and create a new variable order_id:

```
proc sort data = two_arms out = two_arms_3;
  by arm descending pchg;
run;

data two_arms_3;
  set two_arms_3;
  by arm descending pchg;
  retain order_id 0;
```

```
    if first.arm then order_id = 0;
    order_id = order_id + 1;
  run;
```

After the above, we can apply the following SAS code, using PROC SGPANEL:

```
  ods pdf file = 'R:\projects\...\presentation_c.pdf' notoc;
    proc sgpanel data = two_arms_3 dattrmap = waterfall_attrmap_a;
      title1 font = 'SAS Monospace' color = black height = 9 pt
        justify = left
        'Eisai Protocol: Exxxx-yyyy-zzz' justify = right 'Page 1 of 1';
      title2
        font = 'SAS Monospace' color = black height = 10 pt
        justify = center 'Figure 14.x.y.z';
      title3
        font = 'SAS Monospace' color = black height = 10 pt
        justify = center
        'Percentage Change from Baseline in Sums of Diameters';
      title4
        font = 'SAS Monospace' color = black height = 10 pt
        justify = center 'xxx Analysis Set';
      footnote1
        font = 'SAS Monospace' color = black height = 9 pt justify = left
        'First footnote';
      footnote2
        font = 'SAS Monospace' color = black height = 9 pt justify = left
        'Second footnote';
      panelby arm
        /
        columns = 2
        rows = 1
        spacing = 10
        novarname
        ;
      vbarparm category = order_id response = pchg
        /
        barwidth = 1
        group = num_resp
        grouporder = ascending
        attrid = waterfall_a
        ;
      keylegend
        /
        position = bottom
        noborder
        title = 'Best Overall Response'
       down = 1
       titleattrs = (size = 10 pt weight = bold)
       valueattrs = (size = 10 pt)
        ;
      refline 20 -30 / lineattrs = (color = black pattern = dot);
      refline 0 / lineattrs = (color = black thickness = 1 pattern = 1);
      colaxis display = none;
      rowaxis
        label = 'Best Change from Baseline (%)' labelattrs = (size = 10 pt)
        values = (-100 to 100 by 20) valueattrs = (size = 9 pt);
    run;
```

```
ods pdf close;
```

As a result of the above code, we will get the following figure (Figure 3):

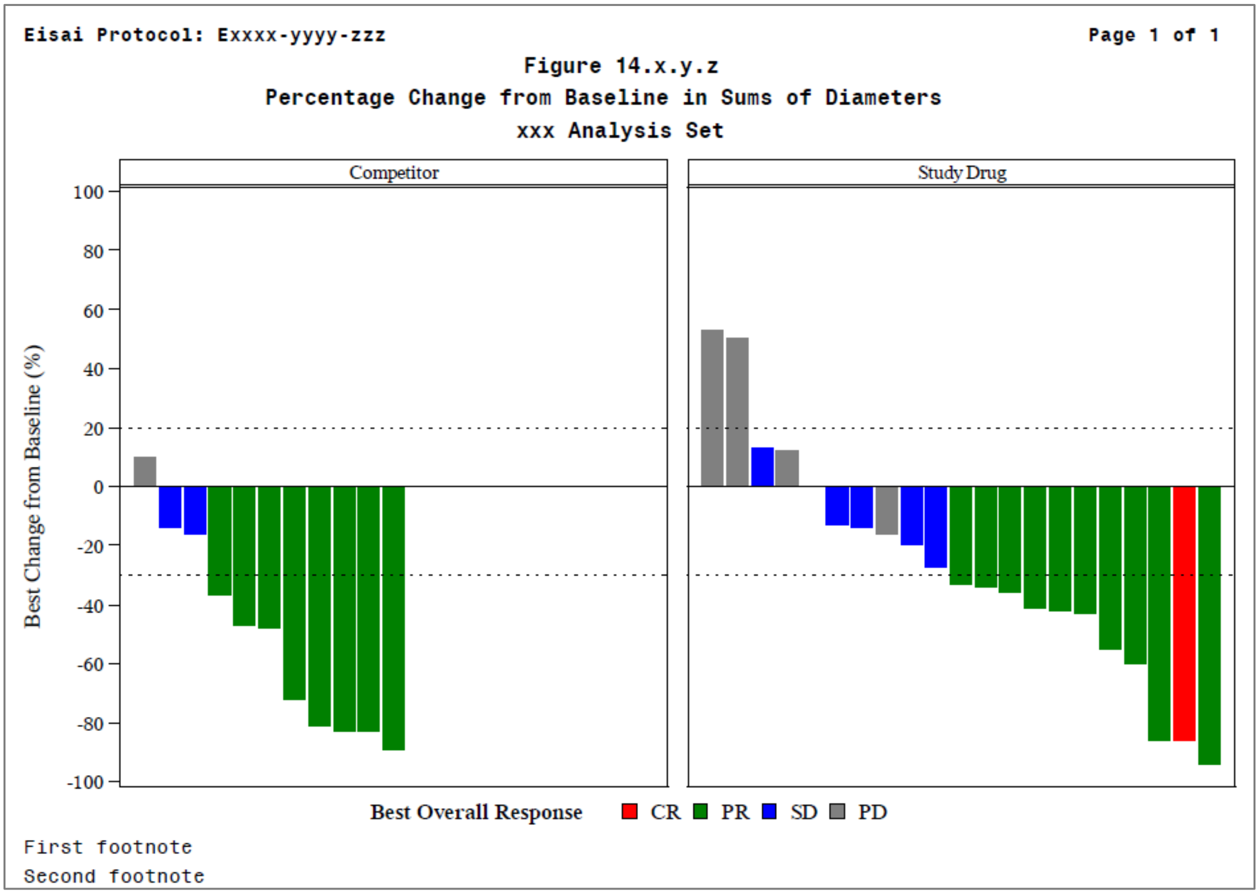

Figure 3: Drawing waterfall plots by arm (two columns, one row) using proc sgpanel

As presented above, this figure looks awkward, because the left panel has huge amounts of unused space. Such output also does not help in a visual comparison of the two arms.

DRAWING WATERFALL PLOTS BY ARM (ONE COLUMN, TWO ROWS) USING PROC SGPANEL

To produce Waterfall Plots by arms with one column / two rows, we would recommend using paper portrait layout (although the SAS code will work for landscape layout as well). From the previous section, we need to change the panelby statement to:

```
panelby arm
  /
  columns = 1
  rows = 2
  spacing = 10
  novarname
  ;
```

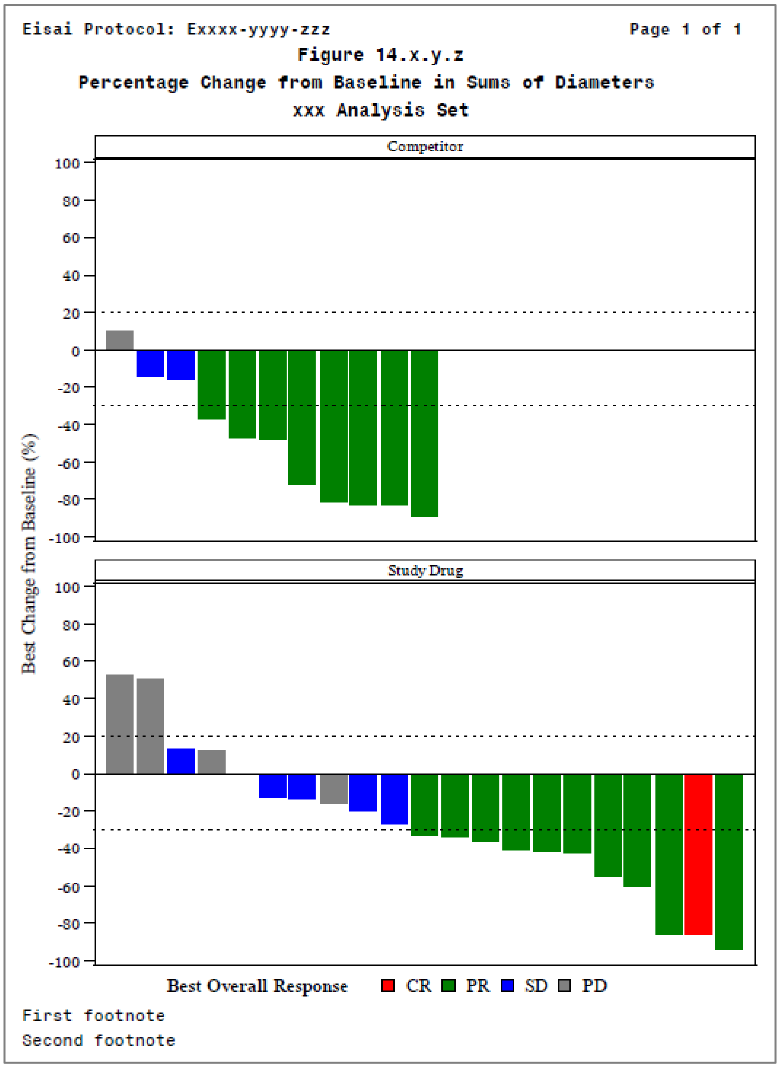

Figure 4: Drawing waterfall plots by arm (one column, two rows) using proc sgpanel

We can repeat what we said in the previous section: This figure (Figure 4) looks awkward, because the top panel has huge amounts of unused space. Such output does not help in a visual comparison of two arms.

DRAWING WATERFALL PLOTS BY ARM (TWO COLUMNS, ONE ROW) USING PROC SGPANEL WITH OPTION UNISCALE

The graph above, produced with two columns / one row can be easily improved, using uniscale = row option in panelby statement. The default value for this option is uniscale = all. The uniscale = row option specifies that only the shared row axes are identical. The column axes vary based on the values of order_id for the respective arms:

```
panelby arm
  /
  columns = 2
  rows = 1
  spacing = 10
  uniscale = row
  novarname
  ;
```

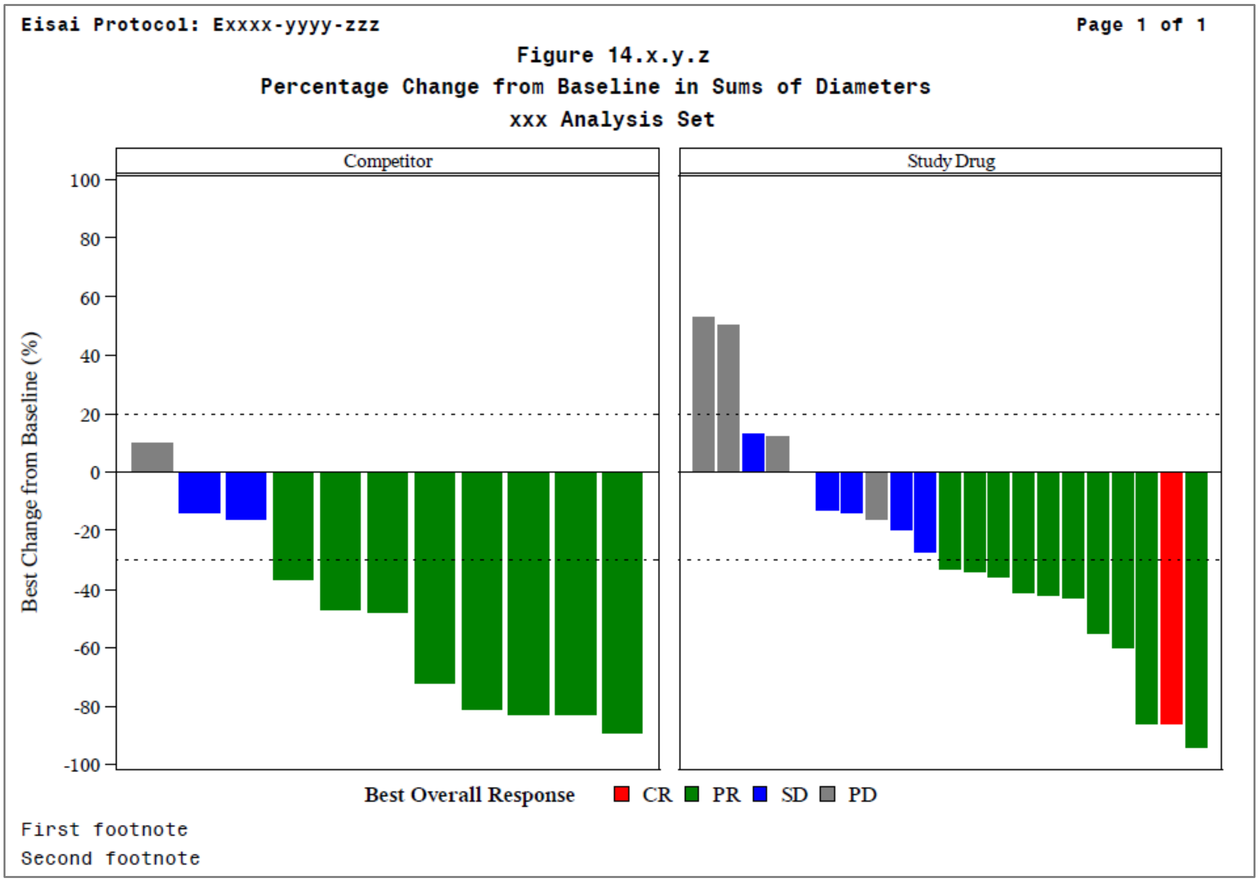

Figure 5: Drawing waterfall plots by arm (two columns, one row) using proc sgpanel with option uniscale

As presented above in Figure 5, the left panel now uses the available space; however, this output is still not very convenient for comparing results between the two arms. It is much more convenient to perform visual comparisons if we use a "One Column Two Rows" design for the output.

DRAWING WATERFALL PLOTS BY ARM (ONE COLUMNS TWO ROWS) USING GRAPH TEMPLATE LANGUAGE AND PROC SGRENDER

In order to produce this graph, we need to have another data adjustment: Two new variables order_id_top and order_id_bottom will be created in the input data set. We also need 2 macro variables, top_cell and bottom_cell to keep the names of both arms:

```
proc sort data = two_arms out = two_arms_6;
  by arm descending pchg;
run;

data two_arms_6;
  set two_arms_6;
  by arm descending pchg;
  retain order_id 0;
  if first.arm then
    do;
      order_id = 0;
      if _n_ = 1 then call symput ('top_cell', arm);
      if _n_ > 1 then call symput ('bottom_cell', arm);
    end;
  order_id = order_id + 1;
run;
```

```
data two_arms_6;
  set two_arms_6;
  if arm = strip (symget ('top_cell')) then
    do;
      order_id_top = order_id;
      order_id_bottom = . ;
    end;
  if arm = strip (symget ('bottom_cell')) then
    do;
      order_id_top = . ;
      order_id_bottom = order_id ;
    end;
run;
```

Again, similar to the previously produced figure with one column and two rows, the portrait page layout is preferable.

Now we can use the following SAS code:

```
ods pdf file = 'R:\projects\...\presentation_f.pdf' notoc;

  proc template;
    define statgraph one_page_two_cells_template;

      begingraph;

        entrytitle
          halign = left
          textattrs =
            (family = 'SAS Monospace' color = black size = 9 pt)
          'Eisai Protocol: Exxxx-yyyy-zzz'
          halign = right
          textattrs =
            (family = 'SAS Monospace' color = black size = 9 pt)
          'Page 1 of 1';
        entrytitle
          halign = center
          textattrs =
            (family = 'SAS Monospace' color = black size = 10 pt)
          'Figure 14.x.y.x';
        entrytitle
          halign = center
          textattrs =
            (family = 'SAS Monospace' color = black size = 10 pt)
          'Percentage Change from Baseline in Sums of Diameters';
        entrytitle
          halign = center
          textattrs =
            (family = 'SAS Monospace' color = black size = 10 pt)
          'xxx Analysis Set';
        entryfootnote
          halign = left
          textattrs =
            (family = 'SAS Monospace' color = black size = 9 pt)
          'First footnote';
        entryfootnote
          halign = left
```

```
  textattrs =
    (family = 'SAS Monospace' color = black size = 9 pt)
  'Second footnote';

discreteattrmap name = 'gtl_attribute_map_two_cells';
  value
    'CR' / fillattrs = (color = red) lineattrs = (color = red);
  value
    'PR' / fillattrs = (color = green)
    lineattrs = (color = green);
  value
    'SD' / fillattrs = (color = blue) lineattrs = (color = blue);
  value
    'PD' / fillattrs = (color = gray) lineattrs = (color = gray);
  value
    'NE' / fillattrs = (color = orange)
    lineattrs = (color = orange);
enddiscreteattrmap;

discreteattrvar
  attrvar = w_f_group
  var = num_resp
  attrmap = 'gtl_attribute_map_two_cells';

layout lattice / rows = 2 rowgutter = 0;

  layout overlay
    /
    xaxisopts = (display = none)
    yaxisopts =
      (
      label = "Best Change from Baseline (%)"
      labelattrs = (family = 'SAS Monospace' size = 10 pt)
      tickvalueattrs = (family = 'SAS Monospace' size = 9 pt)
      linearopts =
        (
        tickvaluesequence =
          (start = -100 end = 100 increment = 20)
        viewmin = -100 viewmax = 100
        )
      );
    entry
      halign = center textattrs = (size = 10 pt weight = bold)
        "&top_cell"
        /
        valign = top border = false;
    barchartparm x = order_id_top y = pchg
      /
      group = w_f_group
      name = 'for_legend'
      barwidth = 1
      ;
    referenceline
      y = -30 / lineattrs = (color = black pattern = dot);
    referenceline
      y = 0 / lineattrs = (color = black pattern =  1);
    referenceline
```

```
          y = 20 / lineattrs = (color = black pattern = dot);
      endlayout;

      layout overlay
        /
        xaxisopts = (display = none)
        yaxisopts =
          (
          label = "Best Change from Baseline (%)"
          labelattrs = (family = 'SAS Monospace' size = 10 pt)
          tickvalueattrs = (family = 'SAS Monospace' size = 9
          linearopts =
            (
            tickvaluesequence =
              (start = -100 end = 100 increment = 20)
            viewmin = -100 viewmax = 100
            )
          );
       entry
          halign = center textattrs = (size = 10 pt weight = bold)
          "&bottom_cell"
          /
          valign = top border = false;
        barchartparm x = order_id_bottom y = pchg
          /
          group = w_f_group
          name = 'for_legend'
          barwidth = 1
          ;
        referenceline
          y = -30 / lineattrs = (color = black pattern = dot);
        referenceline
          y = 0 / lineattrs = (color = black pattern = 1);
        referenceline
          y = 20 / lineattrs = (color = black pattern = dot);
      endlayout;

      layout globallegend
        /
        legendtitleposition = top
        border = false
        ;
        discretelegend "for_legend"
          /
         location = outside
         autoalign = (bottom)
          title = "Best Overall Response"
          titleattrs = (size = 10 pt weight = bold)
          valueattrs = (size = 10 pt)
          ;
      endlayout;

    endgraph;
  end;
run;

proc sgrender data = two_arms_6 template = one_page_two_cells_template;
```

```
  run;

ods pdf close;

/* deleting created template 'one_page_two_cells_template' */
proc template;
  list / store = sasuser.templat;
run;

proc template;
  delete one_page_two_cells_template / store = sasuser.templat;
run;

proc template;
  list / store = sasuser.templat;
run;
```

As a result of applying the above SAS code, we will get an output that can be used for comparing results visually between two arms regardless of the randomization ratio (Figure 6):

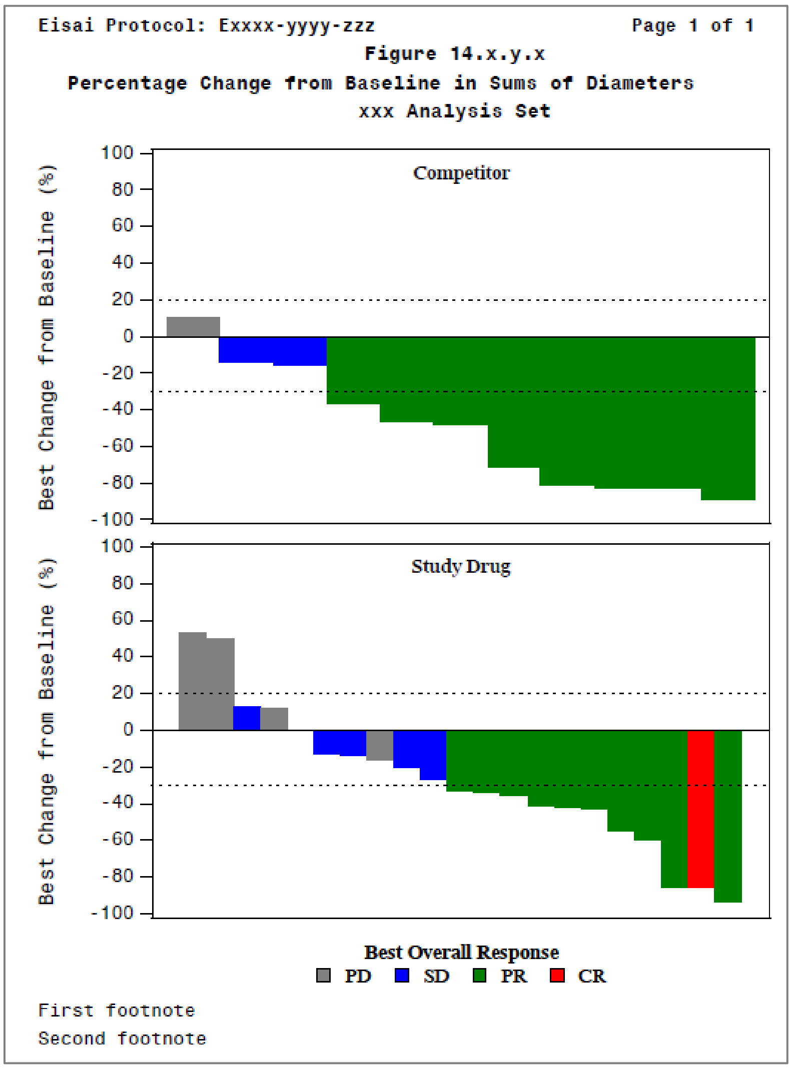

Figure 6: Drawing waterfall plots by arm (one columns two rows) using GTL and proc sgrender

CONCLUSION

SAS has provided us with powerful tools to generate figures using ODS Statistical Graphics and Graph Template Language techniques. This paper has shown that, regardless of randomization scheme, waterfall plots can be a means of visual comparison for efficacy results such as the Best Change from Baseline and the Best Overall Response, in the case of two or more arms. A similar technique, using Graph Template Language, can easily be employed in the case of two or more arms.

REFERENCES

SAS® 9.3 Graph Template Language: Reference, Third Edition

SAS® 9.3 Graph Template Language: User's Guide

SAS® 9.3 ODS Graphics: Procedures Guide, Third Edition

Matange, Sanjay. 2016. Clinical Graphs Using SAS®. Cary, NC: SAS Institute Inc.

Ting Ma, Pharmacyclics LLC. Waterfall plot: two different approaches, one beautiful graph. PharmaSUG 2016 - Paper DG03.

Nora H. Ruel, M.A., City of Hope Cancer Center Paul H. Frankel, Ph.D., City of Hope Cancer Center. Graphical Results in Clinical Studies. A focus on graphics used in oncology. WUSS 2015.

Theresa W. Gillespie, PhD, MA, RN - Understanding Waterfall Plots - Journal of the Advanced Practitioner in Oncology - 2012 Mar-Apr; 3(2): 106–111.

Niraj J. Pandya, Element Technologies Inc. Waterfall Charts in Oncology Trials - Ride the Wave. PharmaSUG 2012 - Paper DG13

Stacey D. Phillips, PharmaNet/i3. A Multifaceted Approach to Generating Kaplan-Meier and Waterfall Plots in Oncology Studies. MWSUG 2012 - Paper PH11

Stan Li, Minimax Information Services. Some Useful Techniques of Proc Format. PharmaSUG2011 - Paper CC19

CONTACT INFORMATION

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

Name: **Ilya Krivelevich**
Enterprise: Eisai Inc.
Address: 155 Tice Blvd.
City, State ZIP: Woodcliff Lake, NJ 07677
Work Phone: 201-949-4204
E-mail: ilya_krivelevich@eisai.com