Paper 7520-2016 Graphical Results in Clinical Oncology Studies

Nora H. Ruel, M.A., Paul H. Frankel, Ph.D., City of Hope Cancer Center

ABSTRACT

This paper will discuss various approaches to presenting oncologic data, and focus on ways each of the graphics can be customized and individualized based on the needs of a study. The topics will range from using SGPLOT and SGPANEL procedures — simple but visually effective ways to present stratified information — to using a complex integration of a series of HIGHLOW plots which depict duration of patient responses on a clinical trial. It will also introduce other graphs that are commonly used in oncology, and discuss how they can be tailored to fit the needs of each study and presentation.

INTRODUCTION

Graphical presentation of clinical data in oncology can quickly and effectively communicate the intended message. In addition to the commonly used survival curves that demonstrate progression-free survival, overall survival, failure-free survival and other survival endpoints, other types of graphics can be strategic in helping describe response in a drug study. Waterfall plots, for example, can help visualize tumor shrinkage or growth, while other plots can show a timeline of individual patient tumor changes and response changes over the course of treatment. Similarly, a simple panel of bar graphs can be a powerful way to demonstrate the difference in incidence of various toxicities between the two arms of a study. We describe our use of the ODS Graphics System to generate the specialized graphics.

GRAPHICAL GEARS

SAS® in version 9.3 or later can generate the graphics outlined in this paper. Plots are produced in either SGPLOT or SGPANEL, with the use of the TEMPLATE procedure and annotation module to make needed adjustments and enhancements. We will be using procedure syntax to create the graphs and, at times, templates and/or annotation to apply further customization.

PAINLESS PANELS

Bar charts are not first in line when one thinks of sophisticated statistical graphics. Although it is not an efficient use of space, a paneled bar chart can be a very effective tool for presentation. In this first example, a paneled bar graph was used to display the difference in post-operative pad-usage between patients who had a history of radiation and those who did not (summary data seen in **Table 1**).

One way to accomplish this is to output the frequencies desired by strata, and plot using PROC SGPLOT. The PROC FREQ code below outputs the desired data summary by XRT usage, followed by the DATA step, which calculates the *Percentage* variable needed as in step **①**. Then we use PROC SGPANEL to create our graph.

```
data statusfreq1;
  set statusfreq;
  percentage = percent/100;
  run;
```

| Obs | PREOPXRT | STATUS | PERCENTAGE | | |
|-----|-------------------|-----------|------------|--|--|
| 1 | No History of XRT | No Pads | 0.633 | | |
| 2 | No History of XRT | Improved | 0.267 | | |
| 3 | No History of XRT | No Change | 0.667 | | |
| 4 | No History of XRT | Worse | 0.333 | | |
| 5 | History of XRT | Improved | 0.286 | | |
| 6 | History of XRT | No Change | 0.286 | | |
| 7 | History of XRT | Worse | 0.429 | | |

Table 1. STATUSFREQ dataset output from FREQ procedure

The pre-defined SAS style selected here is a simple one, called LISTING. It has all the right style elements except for one: the data labels displayed for the percentages appear too small for publication. Prior to SASv9.3, one would have to use the TEMPLATE procedure to apply changes to the font size for such a plot. The syntax in ② shows us how PROC TEMPLATE can be used to created the style <code>BigDataLabel</code> from the parent style <code>Listing</code>. However, with more recent versions of SAS, the font size can be adjusted with the <code>datalabelattr=</code> option seen in ③.

```
proc template; @
define style styles.BiggerDataLabel;
   parent=styles.listing;
   style GraphDataText from GraphDataText/fontsize=11pt;
end;
run;
ods graphics on;
ods pdf body='\\MyNetwork\MyFolder\PadUsageHBar.pdf' style=Listing; 3
                                              /* style=BiggerDataLabel */
proc sgpanel data=statusfreq1;
   panelby preopxrt /columns=2 uniscale=all novarname; 4
   hbar status /response=percentage datalabel datalabelattrs=(size=11pt); •
   colaxis values=(0 to .70 by .10); 6
   footnote j=1 "p=0.001";
   format percentage percent7.1;
   label status = 'Pad usage'
        preopxrt = 'History of XRT'
       percentage = 'Percent';
   title 'Post-operative pad usage';
   title2 'by prior radiation treatment';
 run;
ods graphics off;
ods pdf close;
```

In the final SGPANEL code, first, ODS PDF is chosen to export the graph into a PDF file, and the style is indicated as the newly created BiggerDataLabel style **②**. For a panel graph, we must first specify how the panels will be determined. In this case, we use the PreOpXRT status to divide the patients into two groups. The options within the PANEL statement are chosen so that the data is displayed in columns (columns=2), the scale remains the same for both panels (uniscale=all), and the variable names are left off the panel headers (novarname) **③**. Within the HBAR statement, we specify that our response variable is the Percentage variable calculated from the FREQ procedure and DATA step, and to include data labels, which displays the percentages at the end of the horizontal bars **⑤**. The COLAXIS statement tells the procedure to extend the horizontal axis to 0.7 in increments of 0.1 **⑥**, and the p-value from the Fisher's Exact test is shown with the use of a footnote. Overall, the resulting graph is a simple but effective way to see differences in distribution, as seen in **Figure 1**.

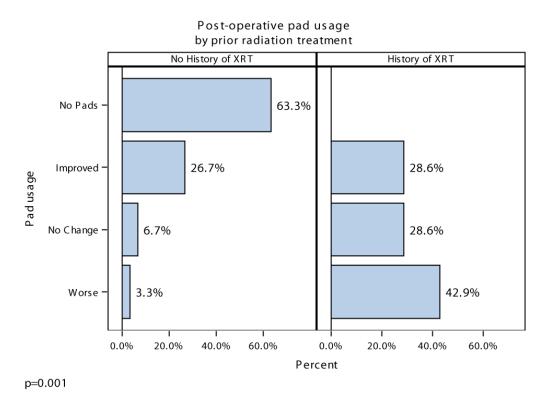


Figure 1. Paneled Bar Chart shows stratified results

TARGETING TUMORS

There are many ways to look at changes in tumor size in a clinical study. Here we will look at two of those methods – waterfall plots, and a line plot, for patients on a single-arm lung cancer drug trial. The idea of a waterfall plot is to display the best improvement in tumor size seen in each of the participating subjects. Each of the bars in the plot describes the percentage of change in the target lesions as compared to the baseline measurements for each subject. This optimal change usually corresponds to what is referred to as the 'best response' to the drug for each subject, with the most common exception being new lesions which suggest progression regardless of tumor size calculations. Here, we have color-coded the best responses. We further use the datalabel option to denote the results of a biomarker test.

The data from the study, once the optimal percentage of change in the sum of target lesion dimensions is calculated and sorted by decreasing values of percent change, is listed partially in **Table 2**.

| Obs | Subject | Percent Change | Best Response | Biomarker |
|-----|---------|----------------|---------------|-----------|
| 1 | XXX | 32.5% | PD (N=10) | + |
| 2 | xxy | 23.5% | PD (N=10) | |
| 3 | xyx | 19.6% | PD (N=10) | * |
| 4 | yxx | 13.0% | SD (N=22) | |
| 5 | XXZ | 5.4% | SD (N=22) | + |
| 6 | XZX | 5.3% | SD (N=22) | - |
| 7 | уух | 4.3% | PD (N=10) | * |
| 8 | уху | 0.9% | SD (N=22) | |

Table 2. Waterfall dataset (partially displayed)

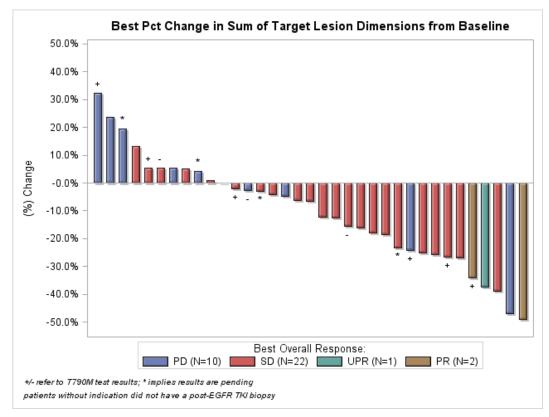


Figure 2. Waterfall Plot with data labels

Notice that the data Waterfall has been sorted in descending order of the variable *Percent_change*. This allows us to see the responses with respect to tumor growth or shrinkage in the order of magnitude of drug response. A quick sort and DATA step help achieve this, and no other special features are needed to make this graph (seen in **Figure 2**) fully functional.

A Spider Plot is yet another way to visualize changes in tumor size. With a Spider Plot, pre-treatment and post-treatment measurements are compared to the baseline measurement (time 0). Thus, the data plotted is the percent change from baseline over the period of subject evaluation. Each leg of the "spider" represents a unique subject, and a horizontal reference at 1.0 is drawn to emphasize the baseline relative value.

The data in **Table 3** are a partial listing of what is plotted in Graph 3. The *Days* variable represents days from baseline (some before, mostly after), and *relsize* (relative size) is calculated with the baseline value as a reference.

| Obs | Subject | days | relsize | | | |
|-----|---------|------|--------------------|--|--|--|
| 1 | XXX | -74 | 0.34579 | | | |
| 2 | XXX | 0 | 1.00000 1.07477 | | | |
| 3 | XXX | 59 | | | | |
| 4 | XXX | 116 | 1.00935 | | | |
| 5 | XXX | 171 | 1.25234 | | | |
| 6 | ууу | -75 | 0.89063 | | | |
| 7 | ууу | 0 | 1.00000 | | | |
| 8 | ууу | 55 | 0.87500 | | | |
| 9 | ууу | 112 | 1.21875 | | | |

Table 3. Partial listing of data TumorTrend

```
proc sgplot data=tumortrend noautolegend;
  series x=days y=relsize / lineattrs=(thickness=2)
        group=subject markers markerattrs=(symbol=circlefilled size=7);
  refline 1.0;
  label relsize = 'Tumor Size Relative to Baseline';
  title "Relative Change in Tumor Size";
  xaxis label="Days from Baseline Scan (time 0)" values=(-320 to 400 by 20)
        valueshint;
  run;
```

This graph in **Figure 3** is accomplished using a SERIES statement within the SGPLOT procedure. To start, we use the *noautolegend* option to exclude a legend that outlines patient ID's. The SERIES statement asks to plot the *Days* variable against *RelSize*, grouped by *Subject*. Hence each subject is plotted with a different line. Marker symbols are used to indicate time-point values, and a less known option *valueshint* is requested in the XASIS statement to extend the x-axis to include any values that fall outside of the designated values range. This will help avoid the scenario of values cut-off the graph. The usefulness of this graph stems from the fact that it tells a story of what happens to each of these tumors over time, while the waterfall plot focuses on the best response measure only. This graphical representation is often most appropriate when the treatment is a cytostatic agent, where response is not necessarily the primary objective.

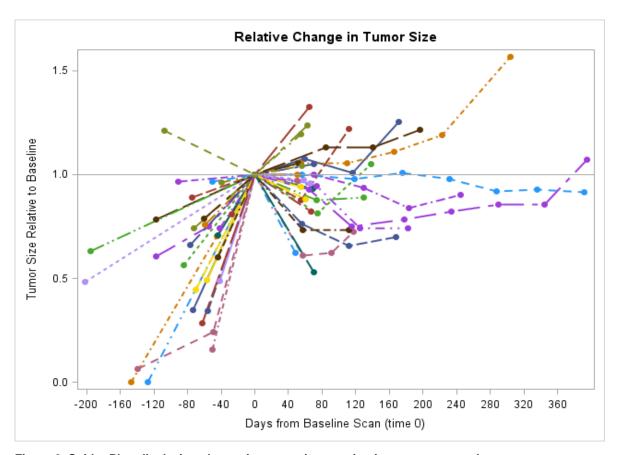


Figure 3. Spider Plot, displaying change in tumor size, starting in pre-treatment phase

STRATIFIED WATERFALL

A study with multiple arms or phases poses additional challenges for graphical representation. We use a two-armed randomized study to treat soft tissue sarcomas as an example. The response of interest once again involves change in tumor size, and we strive for a waterfall plot that is stratified by each of the two arms of the study. For this, SGPANEL is the preferred procedure. However, we notice the data contains an outlier – a patient with a significant tumor growth of 220%. It would be nicer to extend the y-axis to just 100% – which accommodates all but one subject on the study – and use the annotation facility to denote the outlier subject. The DATA step below allows us to put an asterisk on the bar that extends beyond the y-axis.

| Obs | trtarm | n | percent_change | Isresp | flag |
|-----|---------------------------------|---|----------------|--------|------|
| 1 | ARM A: Selumetinib 75 mg PO BID | 1 | 0.0% | PD | Р |
| 2 | ARM A: Selumetinib 75 mg PO BID | 2 | 80.0% | PD | Р |
| 3 | ARM A: Selumetinib 75 mg PO BID | 3 | 76.6% | PD | Р |
| 4 | ARM A: Selumetinib 75 mg PO BID | 4 | 65.1% | PD | Р |
| 5 | ARM B: Selumetinib 50 mg PO BID | 1 | 0.0% | PD | Р |
| 6 | ARM B: Selumetinib 50 mg PO BID | 2 | 220.4% | PD | Р |
| 7 | ARM B: Selumetinib 50 mg PO BID | 3 | 75.0% | PD | Р |

Table 4. WaterfallDivide data (partial listing)

Table 4 shows us the structure of the data that is plotted for our figure. The variable *trtarm* will be the main stratification variable; *n* orders the subjects within each strata or treatment arm by the percentage of change they were observed to have in tumor size from baseline, *Isresp* is the response for each subject, and *flag* indicates the reason the subject came off study.

```
ods graphics on
      /outputfmt=wmf imagename="waterfallplot1" noscale border=off;
ods listing style=statistical file='waterfallplot.wmf' image_dpi=200; •
proc sgpanel data=waterfall sganno=anno;
  panelby trtarm /novarname columns=1; 2
  vbar n / response=percent_change GROUP=lsresp BARWIDTH=0.6
           legendlabel='Best Recist Response' name='resp' datalabel=flag
           dataskin=crisp datalabelattrs=(size=7pt); 3
  colaxis display=none;
  rowaxis values=(-0.6 to 1 by .2) label="Percent Change";
  refline -0.30 /axis=y label='30% decr' lineattrs=(color='gray')
                 transparency=0.70 labelpos=min; 4
  refline 0.20 /axis=y label='20% incr' lineattrs=(color='red')
                transparency=0.70; 4
  label lsresp='Best Overall Recist Response';
  format percent_change percent8.1;
  title h=1 "Best Percent Change in Sum of Target Lesion Dimensions (from
             Baseline)"; title2 h=0.8 'All patients';
  footnote j=l "P: off-tx due to progression; T: off-tx due to toxicity; O:
                off tx other reason";
  footnote2 j=1 h=0.8 color='red' "* patient had 220% increase"; 6
 run;
ods listing close;
ods graphics off;
```

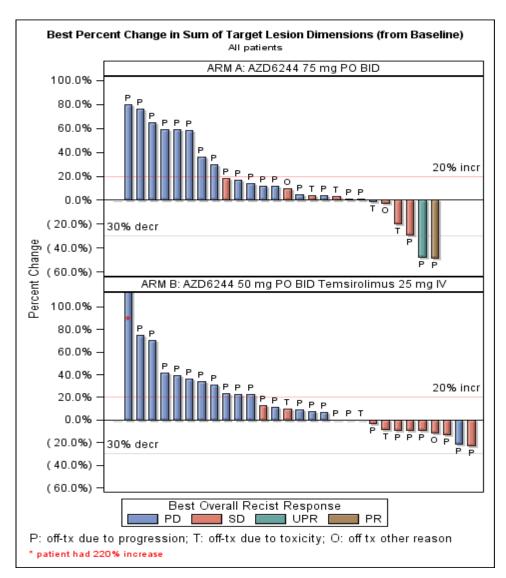


Figure 4. Panel of Waterfall Plots, with annotated comment

RESPONSE TIMELINE

To best summarize response data for a clinical trial in hematology, a timeline plot was desired to show duration of each type of response and retreatment for each of the subjects. Further, annotation was provided in the plot to indicate the type of disease histology, prior treatment, etc., and add arrows at the end of the bars to identify the patients who were continuing treatment.

Using the follow-up information, we assemble a dataset that contains the duration for each phase of response for each of the responding study subjects. Variable *time1* indicates duration of Stable Disease (SD), *time2* indicates duration of Partial Response (PR), etc. Further, we add a variable that will indicate whether the subject is continuing on the last phase. A datalabel is added, which includes descriptive text regarding disease and prior treatment. The solution for generating the desired graphic was determined to be a HIGHLOW graph in PROC SGPLOT, since that was the only graph that allows a special CAP (such as an arrow) to be tacked on the end of the bar. HIGHLOW graphs require that a low point, a high point, a low cap, and a high cap be identified. The data in **Table 5** below is assembled with all the components of the graph in mind.

| Obs | patid | Time 0 | Time 1 | Time 2 | Time 3 | Duration orange | Duration dagr | Duration ligr | datalabel | Low cap | High cap |
|-----|-------|-----------|-----------|-----------|-----------|-----------------|---------------|---------------|-----------------------|------------|-------------|
| 1 | 1 | 0 | 8.3 | 23.6 | 38.3 | 48.7 | | | OFF TREATMENT, | none | NONE |
| 2 | 2 | 0 | 6.2 | 15.7 | 29.2 | 61.3 | | | Still On, Follicular, | none | FILLEDARROW |
| 3 | 3 | 0 | 2.1 | 11.3 | 52.8 | 57.2 | 57.2 | | Still On, Follicular, | none | FILLEDARROW |
| 4 | 4 | 0 | 6.1 | 6.1 | 14.4 | 46.8 | | | OFF TREATMENT, | none | NONE |
| 5 | 5 | 0 | 2.1 | 12.4 | 37.9 | 42.3 | 42.3 | | Still On, Follicular, | none | FILLEDARROW |
| 6 | 6 | 0 | 4.1 | 23.5 | 23.5 | 27.9 | | 27.9 | Still On, Follicular, | none | FILLEDARROW |

Table 5. Partial listing of Responders data

As this is a custom graph, the annotation facility was used to generate a legend. This involves drawing a box somewhere on the graph, inside which rectangles are inserted with corresponding colors and appropriate labels.

```
data anno;
 retain drawspace "wallpercent" widthunit "pixel" heightunit "pixel"
         linecolor "black" linethickness 1 textsize 8 color "black";
  length function $ 9 ;
  input function $ x1 y1 width height textsize anchor $ 30-39 label $ 42-60
display $ 62-66 fillcolor $ 70-77 justify $ 79-83;
 datalines;
Rectangle 72 78.5 12 12 12 bottom
                                                          all CX4D7EBF left
                                        SD
Rectangle 72 73.5 12 12 12 bottom
                                      PR
                                                         all CXABBF66 left
Rectangle 72 68.5 12 12 12 bottom
                                      CR
                                                         all CX2A8307 left
Rectangle 72 68.5 12 12 12 bottom
Rectangle 72 63.5 12 12 12 bottom
                                                         all CXFF8000 left
                                       Retreatment
Text 74 77.2 100 . . bottomleft SD duration
                                                                      left2
Text
         74 72.2 100 . . bottomleft PR duration
                                                                      left
                          . bottomleft CR duration
                                                                      left
Text
         74 67.2 100
Text 74 62.2 135 . . bottomleft Retreatmnt post-PD
                                                                       left
Rectangle 70 61.5 155 80 12 bottomleft
                                                         all white
                                                                      left3
run;
```

The data step DATA ANNO contains instructions to perform three main functions. The first four rows of data provide instructions for drawing small individual rectangles with colors corresponding to the type of duration **1** (e.g., the legend colors). The next final four lines of data provide labels for those four boxes as appropriate **2**. Then, the final row of data tells PROC SGPLOT to draw one large rectangle around the legend contents **3**.

```
thickness=0) lowcap=lowcap highcap=highcap; *ltgreen base/arrow;
  *** next plot the various intervals defining sd, pr, and cr ***;
 highlow y=patid low=time2 high=time3 /type=bar barwidth=0.8 lowcap=lowcap 2
         highcap=highflatcap fillattrs=(color=CX2A8307)
          lineattrs=(thickness=0 color=CX2A8307); *dkgreen;
 highlow y=patid low=time1 high=time2 /type=bar barwidth=0.8 lowcap=lowcap
         highcap=highflatcap fillattrs=(color=CXABBF66)
          lineattrs=(thickness=0 color=CXABBF66); *ltgrn;
 highlow y=patid low=time0 high=time1 /type=bar barwidth=0.8 lowcap=lowcap
         highcap=highflatcap fillattrs=(color=CX4D7EBF)
          lineattrs=(thickness=0 color=CX4D7EBF); *blue;
 yaxis values=(1 to 13) display=(noticks) discreteorder=data
        labelattrs=(size=12 weight=bold);
 y2axis display=none min=0;
 xaxis grid values=(0 to 60 by 10) labelattrs=(size=12 weight=bold)
        label='Months after Initial Treatment';
 title h=2 'Responding Patients';
ods listing close; ods graphics off;
```

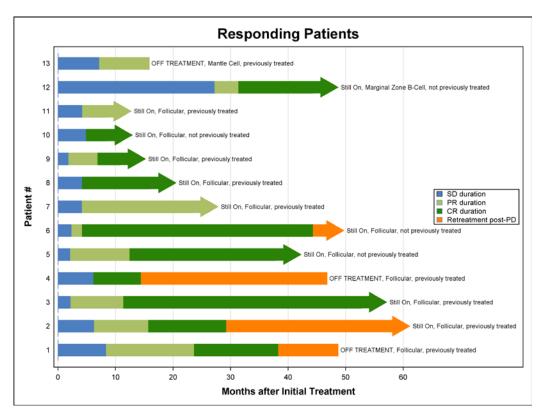


Figure 5. Timeline plot of response and retreatment durations

The procedure syntax for SGPLOT used to generate the desired graph (**Figure 5**) is then put together in several layered pieces. The first three HIGHLOW plot statements **①** draw either a bar with a flat cap or a bar with an arrow cap (defined by the *highcap* option), using the full duration of the follow-up period. The order of the statements follow the chronological progression of response/retreatment phases, as the layers that come last will appear on top, covering up previous layers. In the next three HIGHLOW

statements **②**, the start and stop times for each response or retreatment duration are used to draw rectangular segments of a key color (*time0-time1*, *time1-time2*, *then time2-time3*). In each HIGHLOW statement, the color is specified that corresponds to the response type. The id's of the 13 patients are used for the y-axis, and plotted, in this case, without tick marks. The x-axis has values up to 60, but is extended further out due to the lengthy data labels which are assigned to appear at the end of the bars. SGPLOT produces a legend that occupies an otherwise empty space, and the final graph displays a nice timeline of response and retreatment events and durations.

SIMPLIFLIED SURVIVAL

While survival curves are standard, the graphical approaches to producing them can vary significantly. We use PROC SGPLOT to generate Kaplan-Meier graphs in a poster or manuscript, with some customization to integrate with the presentation. Below, the goal is to generate a survival graph that matches the colors of the poster. The approach we take involves outputting the survival data with the use of the LIFETEST procedure, then graphing the outputted data with the tools and options of our choice. The first step in this process is to refer to ODS graphics to output the PROC LIFETEST survival dataset. Below, the ODS statement outputs the plot data into the dataset work. Survival Plot **①**.

```
ods graphics on ;
ods output Survivalplot=work.SurvivalPlot;  
proc lifetest data=all plots=survival outsurv=pfs_all;
   time pfs*pfs_censor(1);
   strata group/ test=logrank adjust=sidak;
   format group $group.;
run;
ods graphics off;
```

Next, an annotation dataset is needed to place various text elements on the graph. In this case, text on the graph is used to delineate between subjects in Group A and Group B ②. It also allows us to place the p-value (or other desirable information) on the graph ③. Oftentimes, researchers prefer this method to using a legend. Note that this annotation method can easily be extended to include hazard ratios, survival rates at fixed time points, and other desired text in the output. Note that with SASv9.4, the option CURVELABEL is available for use along with most SGPLOT statements (such as the STEP statement used here), and is a conveniently easy way to label your curves. However, the labels must appear either at the beginning of the curve(s) (in which case they would overlap at the vertical axis), or at the end of the curve (extending the axis). Currently, it is not possible to label curves in the middle.

```
define style MySurvStyle_Grn; 4
 parent = styles.analysis;
 style graphdata1 from graphdata1/markersymbol="plus" linestyle=1
                        contrastcolor=CX388C4C color=CX388C4C; 6
 style graphdata2 from graphdata2/markersymbol="plus" linestyle=1
                        contrastcolor=CX138C89 color=CX138C89;
 style GraphDataText from GraphDataText /fontsize=11pt; 6
 style GraphLabelText from GraphLabelText/fontsize=12pt;
 style GraphTitleText from GraphTitleText/fontsize=14pt;
 end;
run;
ods graphics on;
ods listing file='pfsplot.emf' style=MySurvStyle_Grn
gpath='\\MyNetwork\MyFolder\PFS_SurvStyle\PFSPlot_GRN.emf'; •
proc sgplot data=survivalplot1 noautolegend sganno=anno;
  step x=time y=survival
    /group=stratum markerattrs=(symbol=plus size=12)
                     lineattrs=(thickness=3); 3
  scatter x=time y=censored
    /markerattrs=(symbol=plus size=12) group=stratum; 9
  xaxis values = (0 to 60 by 12) label='Months from Treatment';
 yaxis min=0 label='Progression-free Survival' ;
run;
ods listing close;
ods graphics off;
```

Notice the *textcolor* option used in the first two rows of the annotation dataset. These are assigned to match the colors chosen for the graph, and are repeated in PROC TEMPLATE which follows. Within the TEMPLATE procedure, we define a custom style named MySurvStyle_Grn ②, using the pre-defined style ANALYSIS as the parent. The first two *style* statements defined in the procedure ③ specify the line styles that will be used in the graph, with regards to color, symbols, and line types, while the last three *style* statements ③ simply modify font sizes to fit our needs.

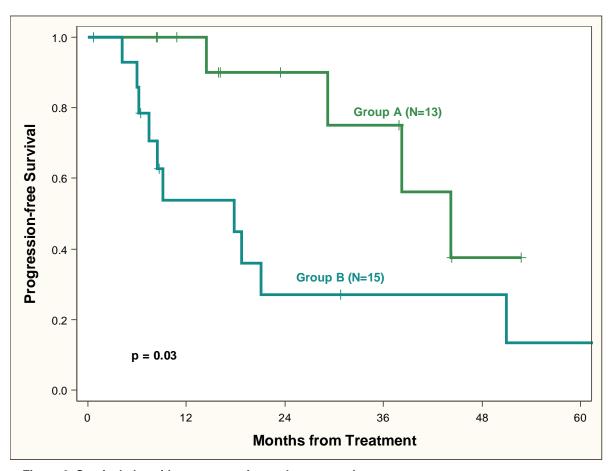


Figure 6. Survival plot with text annotation and custom colors

Prior to calling on the SGPLOT procedure, an ODS statement is refers to MySurvStyle_Grn as the style for the graphic **②**. Finally, PROC SGPLOT is executed, with a STEP statement to plot the survival curve **③**, paired with a SCATTER statement that denotes the censored observations **④** (**Figure 6**). The annotations appear alongside the survival curves and within the graph to put the final touches on the custom graph. Although this is a very simplified version of a survival curve, it is meant to provide the basis on which one can build many more complex elements as needed.

CONCLUSION

Each clinical study comes with its own needs and unique reporting challenges. Some situations call for solutions that can be coded using only procedure syntax, while others rely on the use of templates, annotation, and some creative outside-the-box thinking. The methods shown in this paper will hopefully offer some technical solutions to handling graphical challenges in oncologic studies.

REFERENCES

Sangay Matange, 2011, "Tips and Tricks for Clinical Graphs using ODS Graphics," Proceedings of the SAS Global Forum 2011 Conference, Las Vegas, NV, Cary, NC: SAS Institute Inc.

Schwartz, Susan. 2009 "Clinical Trial Reporting Using SAS/GRAPH® SG Procedures." Proceedings of the SAS Global Forum 2009 Conference. Cary, NC: SAS Institute Inc. Available at http://support.sas.com/resources/papers/proceedings09/174-2009.pdf.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Feel free to contact the authors at:

Nora H. Ruel City of Hope Cancer Center 1500 East Duarte Road Duarte, CA 91010 (626) 256-4673 X63852

nruel@coh.org

https://www.researchgate.net/profile/Nora_Ruel

Paul H. Frankel City of Hope Cancer Center 1500 East Duarte Road Duarte, CA 91010 (626) 256-4673 X65265

pfrankel@coh.org

https://www.researchgate.net/profile/Paul_Frankel

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. Other brand and product names are trademarks of their respective companies.