Graphical Outputs for Efficacy Analyses Using SAS/GRAPH® SG Procedures

David A Collins, GlaxoSmithKline, Research Triangle Park, North Carolina

ABSTRACT

Graphical representation of treatment effect is typically easier to understand as compared to a statistical analysis table. The rollout of SAS® 9.2 has provided a variety of new options to create sophisticated analytical displays. Using PROC TEMPLATE and SAS/GRAPH® SG procedures, one can produce graphical outputs of efficacy data for decision making purposes. A simulated Proof-of-Concept (PoC) efficacy trial is analyzed for illustration purposes.

INTRODUCTION

With the enhancements in recent releases of SAS/GRAPH® procedures, there are many different types of graphical displays that can be used to show treatment effects rather than relying solely on descriptive summary information provided in a table format. Depending on your study design, you may want to use multiple types of displays to address your objective(s).

This paper highlights some types of plots that can be produced with using the SG procedures and applied to the analysis of an efficacy clinical trial. The outputs that you can produce are limited by only your imagination. All graphs in this paper were produced with the ODS PDF style.

STUDY DESIGN

For purposes of the paper, the following is a description of the study design and simulated data used for the subsequent graphical displays.

A single-blind phase 2a study design for a new compound being studied for dyslipidemia with 8 arms and 16 subjects per arm was simulated. The study duration was 2 weeks following a 4-week washout of all dyslipidemia medications. Study endpoints (LDL, HDL, and Triglycerides (TG)) were collected in the fasting state on Day 1 (Baseline), Day 7 and 15 (Day 14 24hrs post-dose). Subjects were stratified by gender and were randomized to one of the following treatment arms: 1mg, 2mg, 4mg, placebo, z+1mg, z+2mg, z+4mg, z+placebo, where z represents a current dyslipidemia treatment. Data were simulated on Days 1, 7, and 15, and then change from baseline and percent change from baseline values were computed for analysis purposes.

The objectives of this simulated study were as follows:

- To determine if the new compound shows benefit in a monotherapy regimen
- To determine if the new compound shows benefit as an add-on to a current dyslipidemia treatment

STATISTICAL ANALYSIS

Panel plots and box plots are used to show summary and individual results for LDL and LDL percent change from baseline.

An analysis of variance (ANOVA) on the percent change from baseline for LDL from Day 1 to Day 15 with treatment and gender as fixed effects was performed. The following SAS code depicts the model that was used for the analysis of the monotherapy data, similar analyses were conducted for the combination therapies and endpoints.

```
proc mixed data=lab_codose;
  class trtcode_codose gender;
  model pct_cfb=trtcode_codose gender;
  lsmeans trtcode_codose;
  estimate 'Monotherapy: 1mg vs Pbo' trtcode_codose -1 1 0 0/cl alpha=0.05;
  estimate 'Monotherapy: 2mg vs Pbo' trtcode_codose -1 0 1 0/cl alpha=0.05;
  estimate 'Monotherapy: 4mg vs Pbo' trtcode_codose -1 0 0 1/cl alpha=0.05;
  ods output lsmeans=lsm_pct_cfb_m;
  ods output estimates=est_pct_cfb_m;
run;
```

PANEL PLOTS

Panel plots are designed to produce graphs based on classification levels. A variety of plot types can be combined (overlayed) with the SGPANEL procedure. Data can easily be compared across your classification variable thus making your graph more effective.

Two figures (Figures 1 and 2) show the same data (LDL percent change from baseline over time by treatment) but differ in that Figure 1 combines each gender together whereas Figure 2 uses colors to depict the values for gender. In each figure, mean values (dotted lines) for each treatment and a reference line at 0 are included.

The following SAS code is used to compute the mean values by treatment and day.

```
proc univariate data=lab_codosex noprint;
  where trtcode_codose in (10 11 12 13);
  by lbtestcd labday trtcode_codose;
  var value_x1;
  output out=stat_x1 mean=mean_x1;
run;

proc univariate data=lab_codosex noprint;
  where trtcode_codose in (30 31 32 33);
  by lbtestcd labday trtcode_codose;
  var value_x3;
  output out=stat_x3 mean=mean_x3;
run;

data lab_codosex;
  merge stat_x1 stat_x3 lab_codosex;
  by lbtestcd labday trtcode_codose;
run;
```

The following SAS code is used for the panel plot.

```
title1 "Plot of %CFB LDLc";
proc sgpanel data=lab_codosex noautolegend;
  where lbtestcd='LDL' and labday in (1 7 15) and
    trtcode_codose in (10 11 12 13 30 31 32 33);
  format trtcode_codose trtb_coy.;
  colaxis offsetmax=0.1 offsetmin=0.1 label='Day' discreteorder=data type=discrete;
  rowaxis label='%CFB' values=(-80 to 100 by 20) grid;

  /* defines panel plot with 8 columns (treatments) */
  panelby trtcode_codose/novarname columns=8 uniscale=column spacing=0;

  /* plots the values for the monotherapy treatments */
  scatter x=labday y=value_x1/markerattrs=(symbol=diamondfilled size=10 color=blue);
  series x=labday y=value_x1/group=subjid markerattrs=(size=0)
    lineattrs=(pattern=solid color=blue);

  /* plots the mean values for the monotherapy treatments */
  scatter x=labday y=mean_x1/markerattrs=(symbol=diamondfilled size=10 color=black);
  series x=labday y=mean_x1/markerattrs=(size=0) lineattrs=(pattern=dash color=black
    thickness=3);

  /* plots the values for the combination treatments */
  scatter x=labday y=value_x3/markerattrs=(symbol=diamondfilled size=10 color=green);
  series x=labday y=value_x3/group=subjid markerattrs=(size=0)
    lineattrs=(pattern=solid color=green);

  /* plots the means values for the combination treatments */
  scatter x=labday y=mean_x3/markerattrs=(symbol=diamondfilled size=10 color=black);
  series x=labday y=mean_x3/markerattrs=(size=0) lineattrs=(pattern=dash color=black
    thickness=3);
run;
```
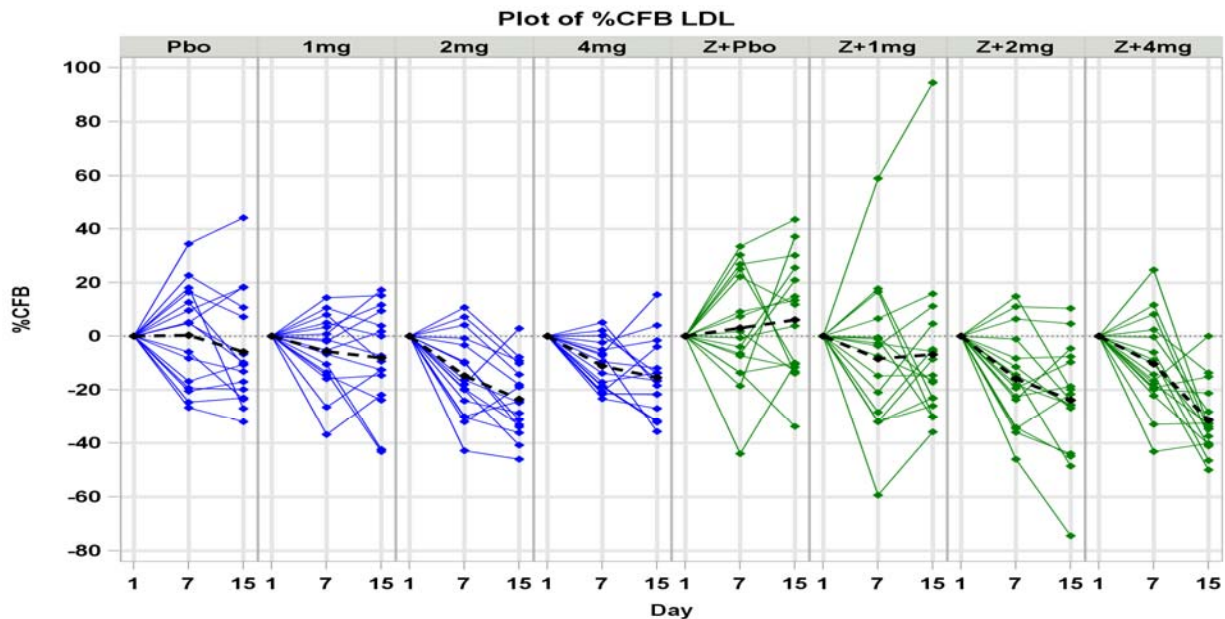
Figure 1: Panel Plot of Percent Change from Baseline - LDL

In Figure 1, the black dotted line represents the mean value across time for each treatment. As an extension of Figure 1, Figure 2 shows by-gender values and means by use of colors to represent each gender and a dotted line to represent the mean values by gender.
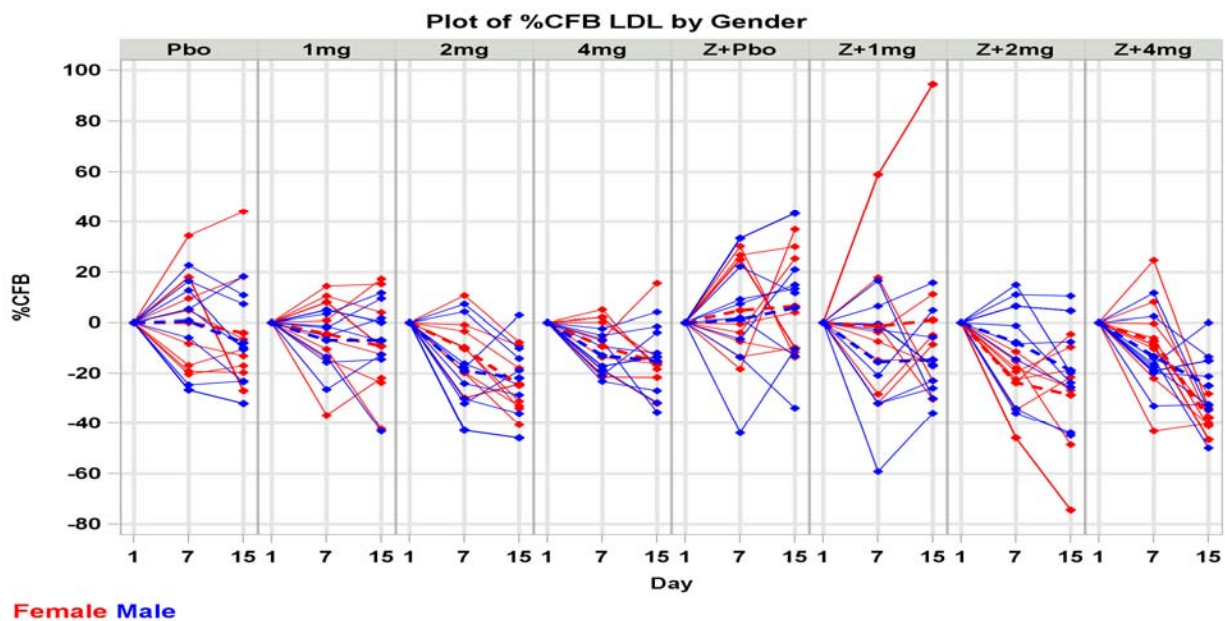

Figure 2: Panel Plot of Percent Change from Baseline by Gender - LDL

BOX PLOTS

Box plots are a way of graphically depicting groups of numerical data through their summaries (minimum, lower quartile, median, upper quartile, and maximum). Box plots are non-parametric in nature, as they show differences between populations without making any assumptions of the underlying distribution. The length of the box itself can help indicate the degree of spread and skewness in the data, and indentify outliers. By default, box plots generated

with either PROC SGPANEL or PROC SGPLOT provides the median value (horizontal line) and the mean value (diamond). There are options associated with box plots for the ends of the whiskers (different percentiles, one standard deviation above and below the mean of the data, 1.5 times the interquartile range, etc.), for outlier identification, and for median differences (notched box plots).

In Figure 3 below, PROC TEMPLATE is used to create a template that includes two Boxplot statements to account for the two combination therapies (monotherapy and combination with drug z) and to enhance the plot. The following are some of the options that have been included: Using color for the two different combination therapies, addition of connecting means across combination therapies, changing display of mean value to a '+' symbol, inclusion of a reference line at 0. PROC SGRENDER is used to produce the output from the template using the specified dataset.

The following SAS code is used to create the box plots.

```
proc template;
  define statgraph twoboxes;
  begingraph;
    entrytitle halign=center 'Plot of %CFB LDL';
    layout overlay/xaxisopts=(label="Treatment"
                  linearopts=(tickvaluelist=(10 11 12 13 30 31 32 33)))
                  yaxisopts=(label="%CFB" linearopts=(tickvaluesequence=(start=-80
                  end=100 increment=20) viewmin=-80 viewmax=100)) cycleattrs=true;
      boxplot x=trtcode_codose1 y=pct_cfb1/meanattrs=(symbol=plus) boxwidth=0.8
            connect=mean display=(connect fill mean median outliers caps);
      boxplot x=trtcode_codose3 y=pct_cfb3/meanattrs=(symbol=plus) boxwidth=0.8
            connect=mean display=(connect fill mean median outliers caps);
      referenceline y=0/lineattrs=(color=black pattern=dot);
    endlayout;
  endgraph;
end;
quit;

proc sgrender data=lab_codose_plotz template=twoboxes;
run;
```
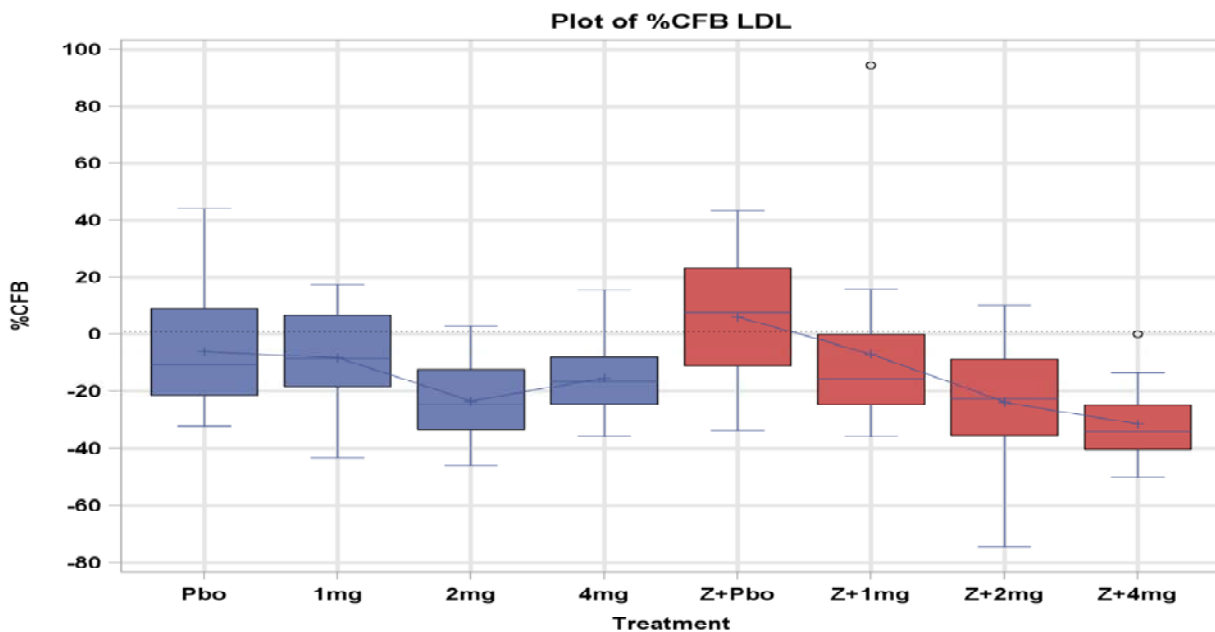

Figure 3: Box Plot of Percent Change from Baseline Day 15 - LDL

From Figure 3, you can see that both the monotherapy and combination treatments reduce LDL by Day 15, but only for the combination therapies is there a clear dose-dependent reduction and are different from placebo.

FOREST PLOTS

Forest plots were developed for use in medical research as a means of graphically representing a meta-analysis of the results of randomized controlled trials. Interest in forest plots has increased in recent years as those in the pharmaceutical industry have started to recognize their benefit when assessing results across multiple studies or treatments. The graphical representation allows for an easier review and comparison across studies or treatments as compared to review of summary tables. Forest plots can take several forms, though they are commonly reported in a similar fashion. On the left-hand side, descriptions of the studies or treatments are commonly presented. On the right-hand side, is a plot of the measure of effect (e.g. odds ratio or point estimate) for each study or treatment incorporating confidence intervals represented by horizontal lines.

A vertical line representing no effect is commonly plotted. If the confidence intervals for individual studies or treatments overlap, it demonstrates that at the given level of confidence their effect sizes no not differ from no effect for the individual study or from a comparator treatment (e.g. placebo).

In Figure 4 below, results of the ANOVA for LDL percent change from baseline is reported. Besides the forest plot, the actual point estimates and the 95% confidence intervals are reported numerically. From the SAS code shown below, some formatting is done prior to using PROC SGPLOT to produce the forest plot.

The following code is used to create the order of outputs in the forest plot.

```
data estall_pct_cfb_m;
  set est_pct_cfb_m;
  if upcase(label)='MONOTHERAPY: 1MG VS PBO' then trt_order=1;
  if upcase(label)='MONOTHERAPY: 2MG VS PBO' then trt_order=2;
  if upcase(label)='MONOTHERAPY: 4MG VS PBO' then trt_order=3;
run;

data estall_pct_cfb_c;
  set est_pct_cfb_c;
  if upcase(label)='COMBINATION +Z: 1MG VS PBO' then trt_order=5;
  if upcase(label)='COMBINATION +Z: 2MG VS PBO' then trt_order=6;
  if upcase(label)='COMBINATION +Z: 4MG VS PBO' then trt_order=7;
run;
```

The following code is used to create additional records that PROC SGPLOT will use as reference lines for readability.

```
data estall_extra;
  set est_pct_cfb_m;
  if trt_order=4 then do; trt_order=4; lower_ci=.; upper_ci=.; estimate=.; end;
  if trt_order=8 then do; trt_order=8; lower_ci=.; upper_ci=.; estimate=.; end;
  format trt_order trtfmtw.;
run;
```

The following code is used to merge all data together for use within PROC SGPLOT.

```
data estall_pct_cfb;
  set estall_pct_cfb_m estall_pct_cfb_c estall_extra;
  format trt_order trtfmtw.;
  if lbtestcd='HDL' then labtest='HDL';
  if lbtestcd='LDL' then labtest='LDL';
  if lbtestcd='TG' then labtest='Triglycerides';
  lower_ci=lower;
  upper_ci=upper;
run;
```

The PROC SGPLOT code this is used to produce the forest plot.

```
proc sgplot data=estall_pct_cfb noautolegend;
  where labtest ne '' and lbtestcd='LDL' and labday=15;
  title1 "Plot of LS Mean (95% CI) ^{unicode delta_u}^{unicode delta_u} LDLc %CFB";
  title2 h=1.5 "Day = 15";
  title3 h=1.5 "MODEL: %CFB = Treatment";
  format trt_order trtfmtw.;
  label labday='Day' trt_order='Treatment' estimate='LS Mean Double-Delta (95% CI)';
  yaxis type=linear values=(1 to 8 by 1) display=(noticks nolabel)
```

5

```
        offsetmin=0.1 offsetmax=0.05;
  xaxis grid type=linear values=(-60 to 20 by 10) offsetmin=0 offsetmax=0.35 minor
        display=(nolabel);

  /* Forest plot */
  scatter y=trt_order x=estimate /markerattrs=(symbol=circlefilled)
          errorbarattrs=(color=black) xerrorlower=lower_ci xerrorupper=upper_ci;

  /* Prints point estimate and 95% confidence interval to the right side */
  scatter y=trt_order x=pe/markerchar=estimate x2axis;
  scatter y=trt_order x=lcl/markerchar=lower_ci x2axis;
  scatter y=trt_order x=ucl/markerchar=upper_ci x2axis;
  refline 20/axis=x;
  x2axis offsetmin=0.7 display=(noticks nolabel);

  /* Reference lines to separate the monotherapy and the combination treatments */
  refline 4 8/axis=y lineattrs=(color=blue pattern=2 thickness=1) noclip;

  /* Vertical reference line */
  refline 0/axis=x lineattrs=(color=red pattern=2 thickness=5) noclip;
run;
```
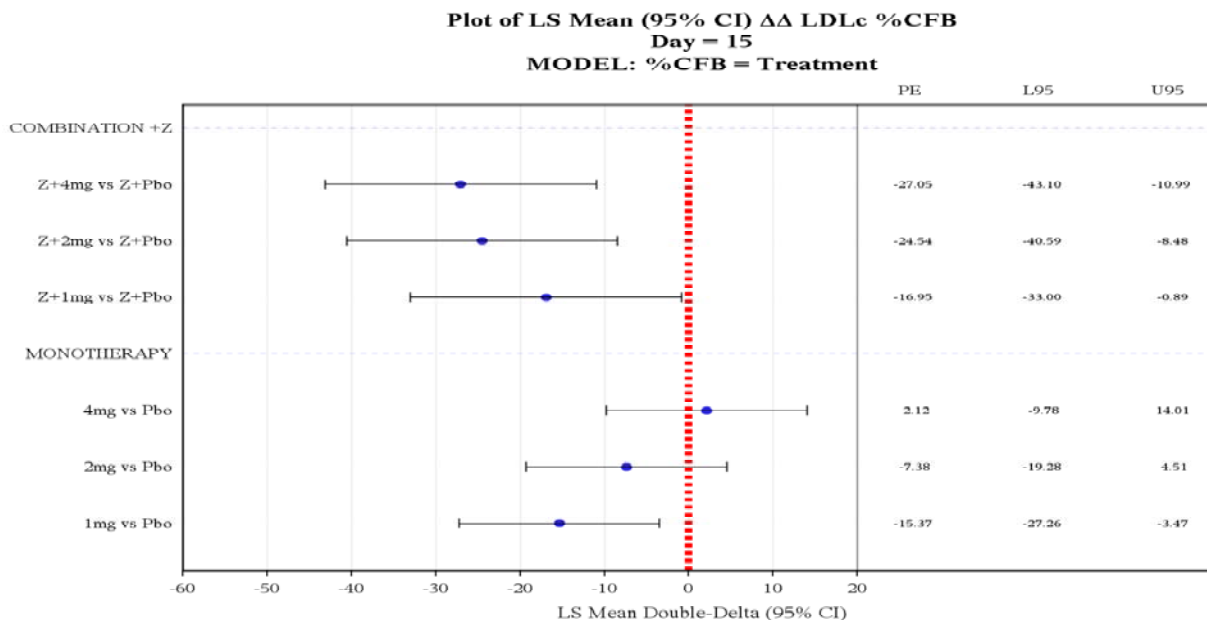

Figure 4: Forest Plot of ANOVA Percent Change from Baseline Day 15 - LDL

Figure 4 provides a lot of information for the reviewers to see that combination therapy shows greater efficacy in a dose dependent manner that monotherapy at the same doses. In addition, having the actual point estimates and associated 95% confidence interval, one can state the values rather than trying to estimate from the figure. With the use of the offsetmin and offsetmax options in the x-axis statements, one can enhance the forest plot by including other supporting data (e.g. point estimates and confidence intervals, horizontal box plot for test and reference treatments).

With PROC SGPLOT, you can see that there is a greater amount of flexibility to design your graphics the way you want them, whether it be color, combination of different graphs, addition of numerical data, etc.

CONCLUSIONS

The addition of the SG procedures in SAS 9.2 has provided the user with many new procedures for producing statistical graphics. The flexibility that is now available gives the user the ability to customize outputs for meeting ones needs. Minimal SAS code is now needed to create outputs as compared to when SG procedures did not exist.

The ability to overlay plots with the SG procedures is a powerful tool that its ease of use will please all users.

REFERENCES

SAS Institute Inc. (2008). "SAS/GRAPH 9.2: Statistical Graphics Procedures Guide." Available at
http://support.sas.com/documentation/onlinedoc/graph/index.html.

Cartier, J. (2006), "A Programmer's Introduction to the Graphics Template Language," Proceedings of the Thirty-first Annual SAS Users Group International Conference. Cary, NC: SAS Institute Inc. Available at http://www2.sas.com/proceedings/sugi31/toc.html.

Rodriguez, R. 2006. "Creating Statistical Graphics in SAS® 9.2 – What Every Statistical User Should Know". *Proceedings of the Thirty-First Annual SAS Users Group International Conference*, San Francisco, CA. Available at http://support.sas.com/rnd/base/topics/statgraph/192-31-updated.pdf.

ACKNOWLEDGMENTS

RECOMMENDED READING

Heath, Dan. 2007. "New SAS/GRAPH Procedures for Creating Statistical Graphics in Data Analysis." *Proceedings of the SAS Global Forum 2007 Conference.* Cary, NC: SAS Institute Inc. Available at http://www2.sas.com/proceedings/forum2007 .

Heath, Dan. 2008. "Effective Graphics Made Simple Using SAS/GRAPH 'SG' Procedures." *Proceedings of the SAS Global Forum 2008 Conference.* Available at http://www2.sas.com/proceedings/forum2008 .

Schwartz, Susan. 2008. "Butterflies, Heat Maps, and More: Explore the New Power of SAS/GRAPH." *Proceedings of the SAS Global Forum 2008 Conference.* Available at http://www2.sas.com/proceedings/forum2008 .

Heath, Dan. 2009. "Secrets of the SG Procedures." *Proceedings of the SAS Global Forum 2009 Conference.* Available at http://www2.sas.com/proceedings/forum2009.

CONTACT INFORMATION

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

David A Collins
GlaxoSmithKline
Five Moore Drive
Research Triangle Park, North Carolina 27709
(919) 483-0828
david.a.collins@gsk.com