

Paper 174-2009

## Clinical Trial Reporting Using SAS/GRAPH® SG Procedures

Susan Schwartz, SAS Institute Inc., Cary, NC

### ABSTRACT

Graphics are a powerful way to display clinical trial data. By their very nature, clinical trials generate a large amount of information, and a concise visual presentation of the results is essential. Information about the patient population, drug dosages, clinical responses, and adverse events must be clear. Clinical laboratory results need to be presented within the context of acceptable limits, and subtle changes over time must be highlighted. This presentation will show, by example, how such graphs can easily be created using the SAS/GRAPH® SG procedures.

The techniques that will be emphasized in this presentation include:

- Creation of a dose response plot by overlaying multiple plots in one graph
- Construction of a hematology panel using treatment regimen and visit numbers as the classification variables
- Presentation of a matrix of liver function tests (LFTs) for at-risk patients
- Aggregation of data into on-the-fly classification variables using user-defined formats
- Getting the axis you want using built-in best fit algorithms
- Generation of publication-ready graphs in color and in black and white

### INTRODUCTION

The new SAS/GRAPH procedures—SGPLOT, SGPANEL, and SGSCATTER—provide new tools for viewing and reporting data collected during clinical trials. The SG procedures are an extension of the ODS Graphics framework, providing access to the Graph Template Language (GTL) in the familiar syntax of the SAS/GRAPH procedure.

The concept behind the SG procedures is simple in theory, yet powerful in execution. Each GRAPH contains one or more CELLS, and each CELL contains one or more overlaid PLOTS. Both the GRAPH and CELL can contain supporting elements such as titles and legends.

This paper highlights how to create single-cell and multi-cell graphs using examples commonly encountered in the reporting and analysis of clinical trial data. The output you can produce is limited only by your imagination. Unless otherwise noted, all graphs in this paper were produced with the ODS LISTING style.

### SINGLE-CELL GRAPHS

#### THE SGPLOT PROCEDURE

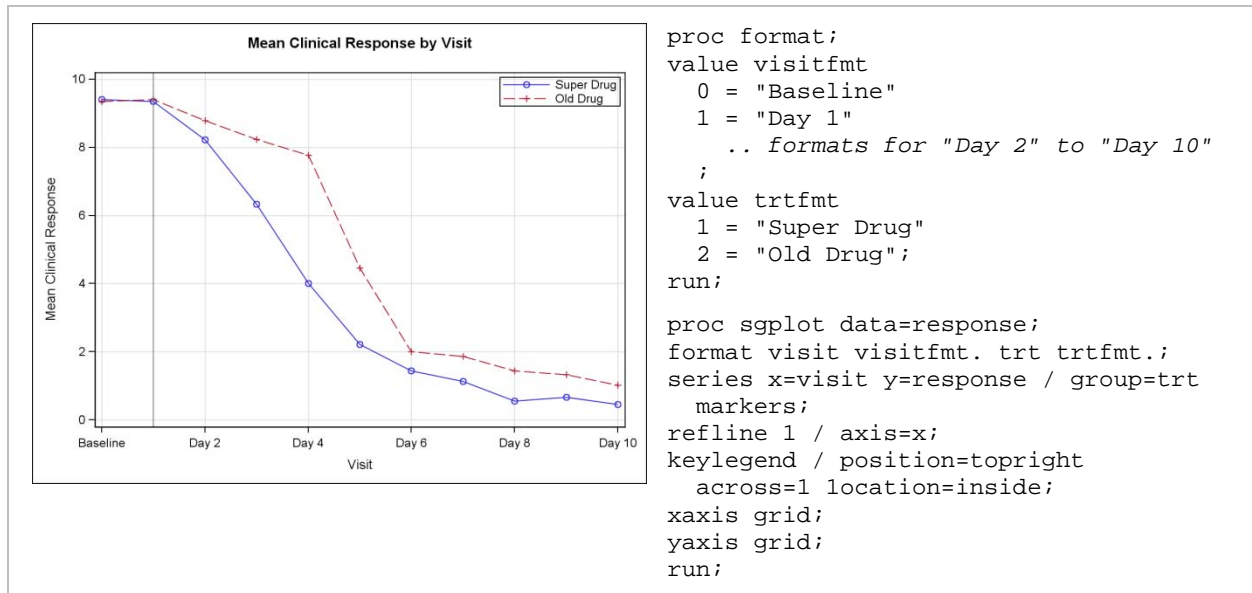
Single-cell graphs are produced by the SGPLOT procedure. Figure 1 shows a simple dose response plot that displays the mean clinical response of two treatment groups over a 10-day study. The SGPLOT procedure's cell is a composite of one PLOT statement and two supporting elements.

The input data set contains three numeric variables:

```
response    visit    trt
    9.40         0         1
    9.35         1         1
    ..... etc
```

- The PROC FORMAT step defines two user-defined formats, VISITFMT and TRTFMT, and the FORMAT statement associates the user-defined formats with their (numeric) variables.
- The SERIES statement produces a grouped series plot with marker symbols. The plot's colors, marker symbols, and line patterns are assigned automatically from the style elements GRAPHDATA1 through GRAPHDATA*n*.
- The REFLINE statement draws a reference line at X=1 to indicate day one of treatment.

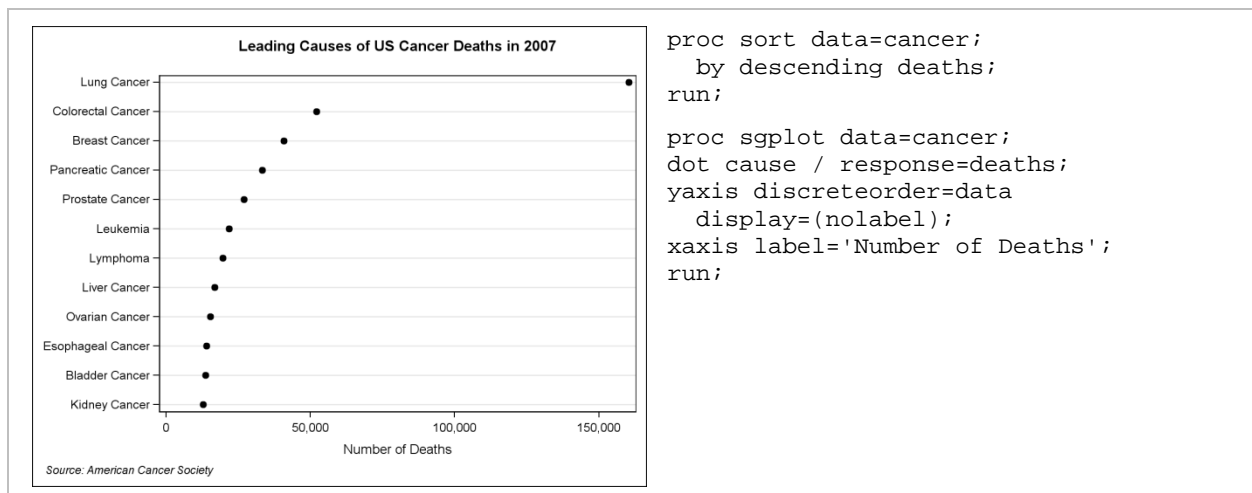
- The options in the KEYLEGEND statement position the legend in the upper-right, inside corner of the plot. By default, a legend is generated for grouped data, but it is placed at the bottom, outside of the plot area. The options in the KEYLEGEND statement customize the placement.
- The X and Y axes are drawn and labeled automatically.



**Figure 1. Dose Response Plot**

Figure 2 shows another example of a straightforward single-cell plot that conveys a lot of information with a minimum amount of code.

- The input data set is sorted in descending order by the response variable (DEATHS).
- The DOT plot displays the number of deaths by the category (CAUSE-OF-DEATH)
- The YAXIS statement DISCRETEORDER=DATA option stipulates that the category values be drawn as they appear in the data set. (The other choices for discrete axis order are FORMATTED and UNFORMATTED). Because the data set is in descending order by deaths, the effect is that of a waterfall showing the relative magnitude of cancer deaths in 2007.
- The XAXIS statement overrides the label of the variable DEATHS.

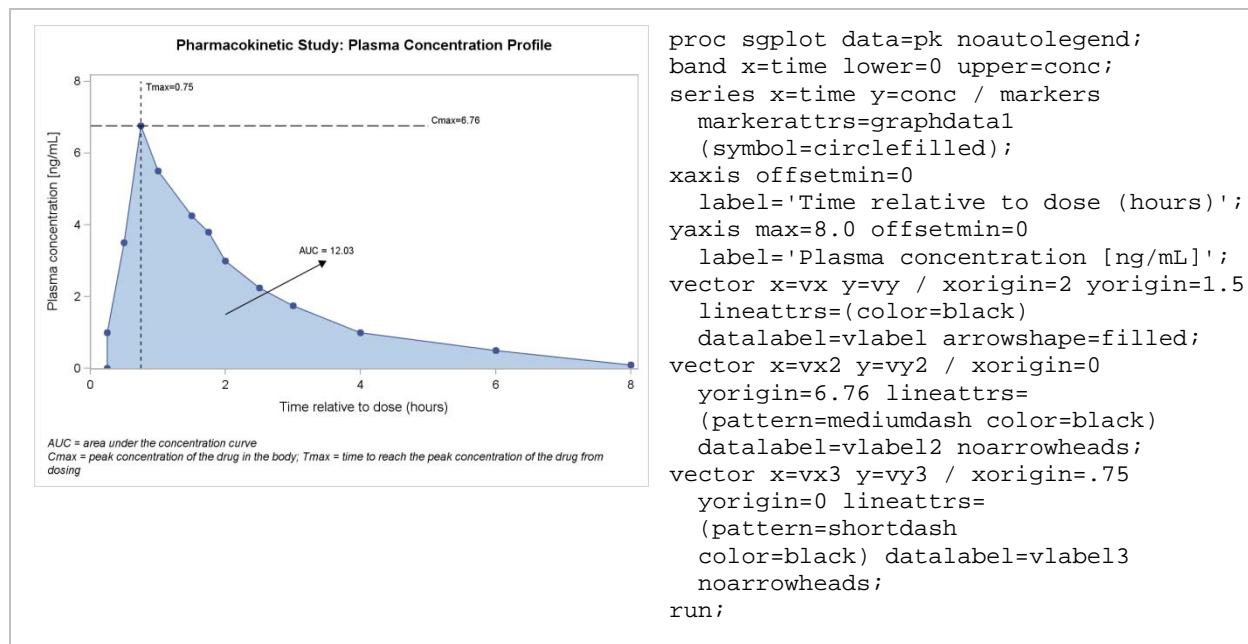


**Figure 2. Cancer Mortality Plot**

Figure 3 shows how the SGPLOT procedure can create a more complex single-cell plot that displays a Plasma Concentration Profile. The input data set contains the variables and values as shown. Note that only TIME and CONC are valued on every observation. The remaining variables are valued only on the first observation and have been added to the data set for the vector plots.

time	conc	vx	vy	vlabel	vx2	vy2	vlabel2	vx3	vy3	vlabel3
0.25	0.00	3.5	3	AUC = 12.03	5	6.76	Cmax=6.76	0.75	8	Tmax=0.75
0.25	1.00	.	.		.	.		.	.	
0.50	3.50	.	.		.	.		.	.	
etc.....										

- The NOAUTOLEGEND option in PROC SGPLOT suppresses the legend.
- The BAND statement draws the solid blue area, with X=TIME and Y=0 to the CONC value. Thus, the band draws from X=.25, Y=(0 to 0), and then X=.25, Y=(0 to 1.0), and so on. The band represents the Area Under the Curve (AUC) value.
- The SERIES statement draws with X=TIME and Y=CONC, which is also the upper Y value of the BAND. Thus, the series is overlaid on the upper boundary of the band. By default, the marker symbols would inherit their attributes from the element GRAPHDATAdefault. Because this is also the default color of the BAND's fill color, the markers are assigned to GRAPHDATA1 to create a contrast. The series markers indicate the plasma concentrations at 30-minute intervals (for the first three hours), one-hour interval (at hour 4), and two-hour intervals (at hours 6 and 8).
- The XAXIS statement specifies the label and shifts the cell contents slightly to the left (OFFSET=0).
- The YAXIS statement specifies the label and again shifts the cell contents along the Y axis.
- The first VECTOR statement draws the arrowed line from (2, 1.5) to (vx, vy), and adds the label from VLABEL.
- The second VECTOR statement draws the horizontal line at Y=6.76, and adds the label from VLABEL2.
- The third VECTOR statement draws the vertical line at X=0.75, and adds the label from VLABEL3.



**Figure 3. Plasma Concentration Profile**

Figure 4 demonstrates how to use the SGPLOT procedure to display the results of a placebo-controlled double-blind trial with a three-week washout period. Note that the BAND is drawn first so that it is underneath the remainder of the PLOTS.

- The BAND draws with LOWER=12.1, UPPER=15, and Y=(0 to X1). X1 is set to 15 in the data set, so this has the effect of drawing a band from X=(12.1 to 15) and Y=(0 to 15). GRAPHDATA1 is the first group color, so the band will have the same fill color (blue, in this case) as the Placebo group, but with a lighter saturation. The band is used to highlight the study's washout period, weeks 12 to 15.
- The SCATTER statement draws the grouped markers (filled circles) with lower and upper error bars, with all other attributes from the GRAPHDATA1 through GRAPHDATA3 elements. The scatter plot displays each treatment group's mean percent change at three-week intervals, from week 0 to week 15.
- The first SERIES statement connects the markers in the SCATTER plot using the same group colors.
- The second SERIES statement draws the markers and series for the three-week washout period only. A dash is used to connect the markers. Again, the same group colors are applied by default.
- The XAXIS and YAXIS labels are specified and the XAXIS tick values are set manually.
- The first REFLINE statement draws a horizontal line at Y=0, and represents a zero percent change from the baseline.
- The second REFLINE statement specifies a line at X=13.5, which is the middle of the washout period. The line does not actually draw because the line thickness is set to 0. This is a trick to place the text "|--Washout--|" centered in the middle of the washout block.
- The KEYLEGEND statement draws a borderless legend from the SCATTER group values.

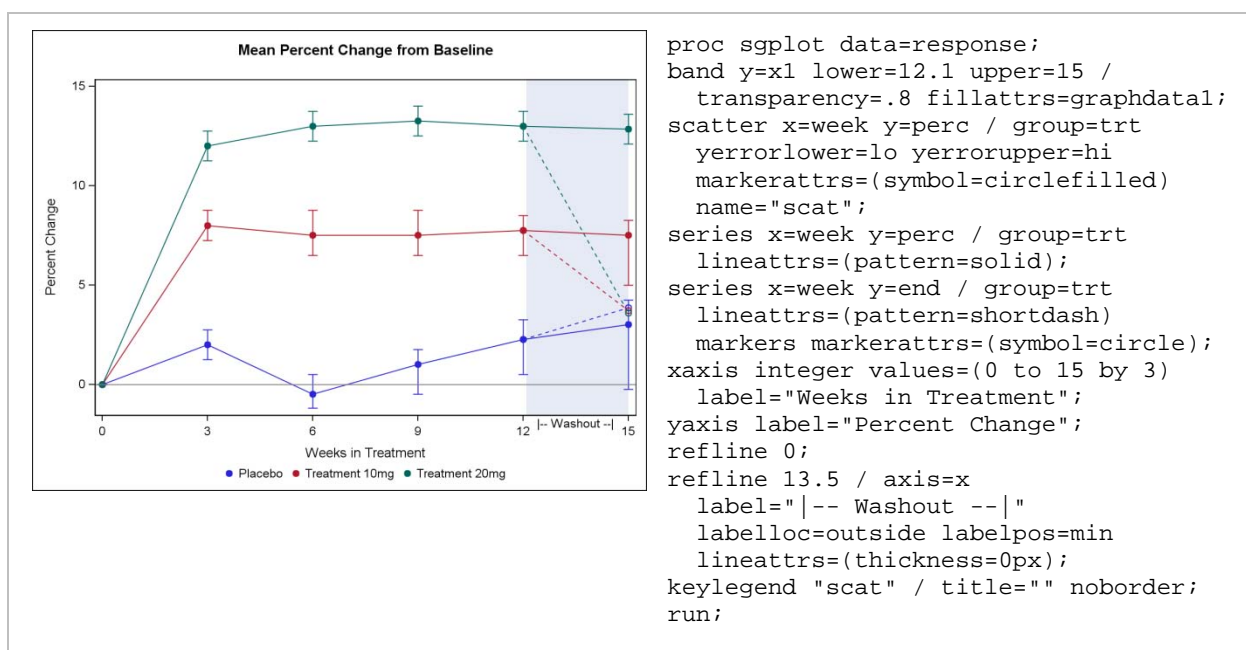


Figure 4. Mean Percent Change from Baseline

Figure 5 demonstrates the use of a built-in statistical function in an SG procedure. The input data set contains these values:

c500	t500	c100	t100	c50	t50	c25	t25
10%	5%	12%	10%	18%	18%	.	.
20%	7%	20%	12%	20%	19%	.	.
30%	9%	30%	17%	30%	22%	30%	29%
etc.....							

- The ODS escape character is set to ~. The default value is {ESC}—this will be used later in the code to draw Unicode glyphs for the alpha, beta, and delta characters.
- The NOAUTOLEGEND option suppresses the automatic legend.

- The first REG statement draws the bottom curve and labels it N=500.
  - The next three REG statements draw the other three curves and label them. Because there is more than one REG statement, PROC SGPLOT assigns each REG statement the attributes from the next unused GRAPHDATA $n$  element. Thus, the procedure “cycles” through GRAPHDATA1 through GRAPHDATA4. This behavior can be changed with the NOCYCLEATTRS option.
  - The REFLINE statement draws a reference line at X=.5, indicating a response rate in the control group equal to 50%.
  - The XAXIS statement specifies the axis label, tick values, and vertical grid lines.
  - The YAXIS statement specifies the axis label, tick values, and horizontal grid lines. The ~ is the ODS escape character, which provides access to the Unicode inline formatting of Unicode glyphs. The {UNICODE DELTA\_U} draws the uppercase delta.
- See the *SAS/GRAPH 9.2: Graph Template Language User's Guide* for more information about Unicode glyphs in GTL and the SG procedures.
- The INSET statement draws the label value pairs for the  $\alpha=5\%$  and  $\beta=20\%$  inset box. It is drawn in the top-right corner of the plot without a border.

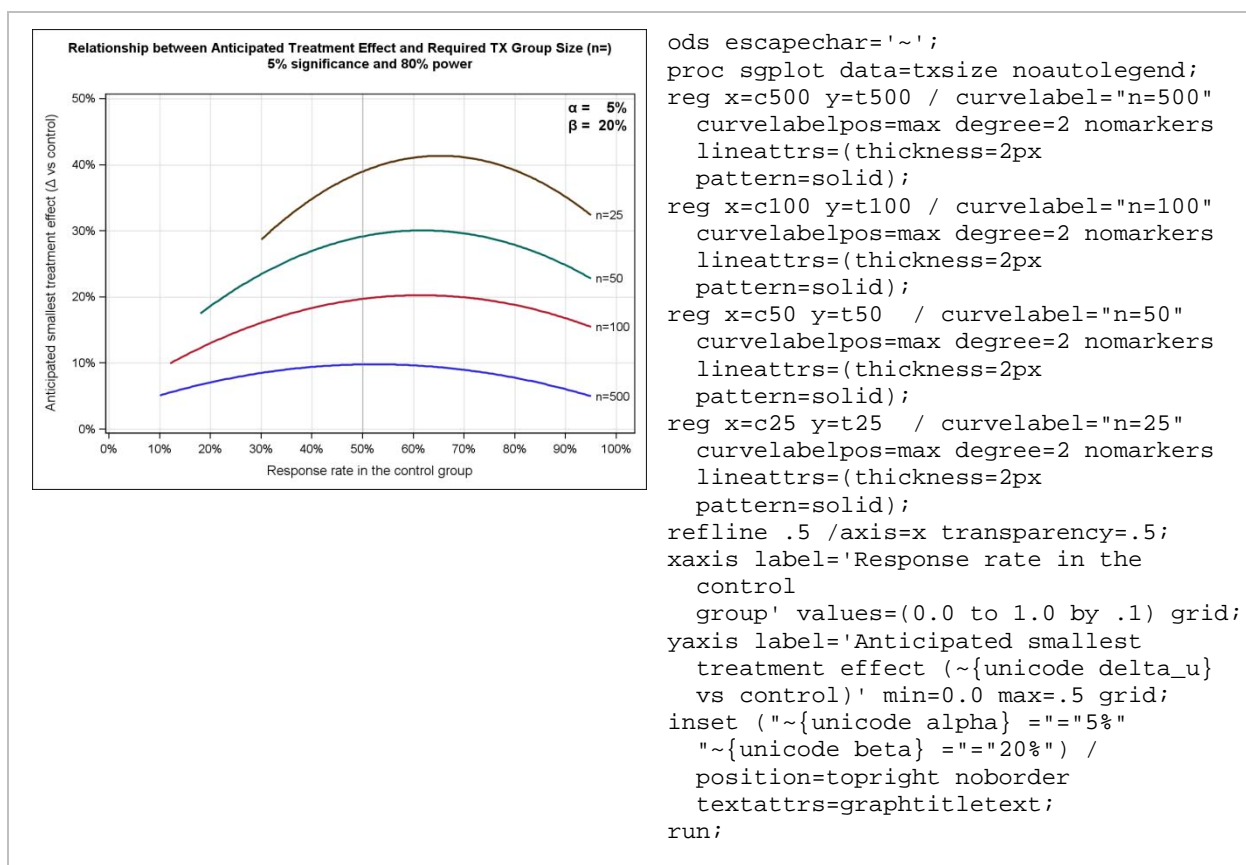


Figure 5. Clinical Study Design Plot

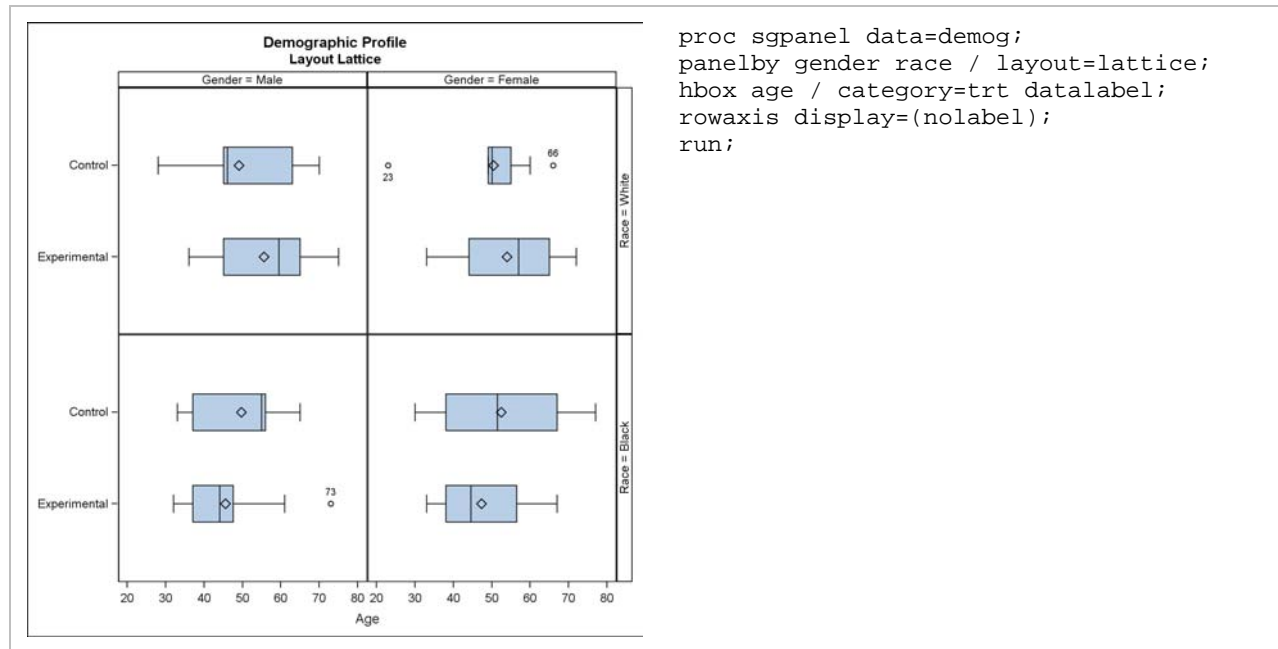
## MULTI-CELL GRAPHS

### THE SGPPANEL PROCEDURE

The SGPPANEL procedure creates a panel of single-cell plots within one or more classification variables. Panels can be arranged in a grid, in a series of cells, in a single column, or in a single row. The panel type is specified by the LAYOUT= option in the PANELBY statement. The axis data ranges can be set as uniform across all columns, across all rows, or across both (the default). PROC SGPPANEL determines the axis ranges before it draws any panels, thus assuring uniformity across all panels.

Figure 6 displays a basic panel of trial demographics, arranged in a LATTICE layout.

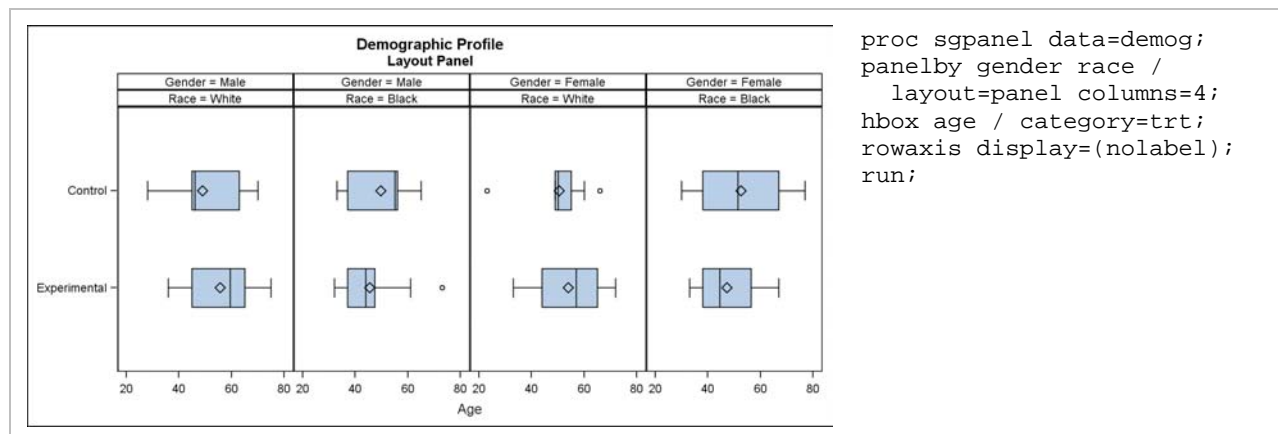
- There are two PANELBY variables, the first for the column (GENDER) and the second for the row (RACE).
- Each cell contains an HBOX (horizontal box plot) of AGE (the response) by TRT (the category). The DATALABEL option displays the value of each outlier.
- The ROWAXIS statement suppresses display of the row label.



**Figure 6. Demographic Profile in a LATTICE Layout**

Figure 7 displays the same demographic information, but in a PANEL layout.

- The classification values appear in one series of cells, ordered by GENDER, then RACE.
- The COLUMNS=4 option explicitly requests that all four cells appear in one row. The ONEPANEL option would do the same thing. (See Figure 9.)
- Each cell contains an HBOX (horizontal box plot), this time without the DATALABEL option.
- The ROWAXIS statement suppresses display of the row label.



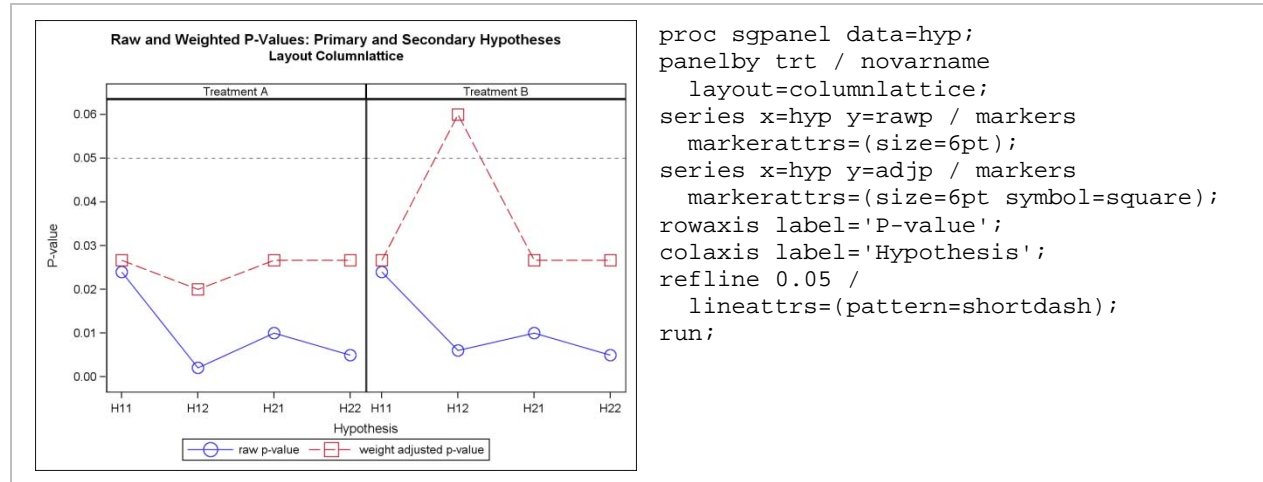
**Figure 7. Demographic Profile in a PANEL Layout**

Figure 8 displays a comparison of the raw and weighted  $p$ -values in a clinical trial that has both primary and secondary endpoints. The results are arranged in a COLUMNLATTICE layout.

- There is only one classification variable, TRT. The NOVARNAM option suppresses display of the text "TRT=" in the cell header.

The contents of each cell are then specified:

- The first SERIES statement draws the plot of the raw  $p$ -values.
- The second SERIES statement draws the plot of the weight-adjusted  $p$ -values.
- The ROWAXIS and COLAXIS statements specify the labels of the row (Y) and column (X) axes.
- The REFLINE statement draws a horizontal line at  $Y=.05$  (to indicate 95% confidence).



**Figure 8. Comparison of  $p$ -values in a COLUMNLATTICE Layout**

Finally, figure 9 displays an HBAR (horizontal bar chart), arranged in a ROWLATTICE layout.

- The PANELBY statement contains one classification variable, SEX. The NOVARNAM option suppresses the variable name in the cell header, and the ONEPANEL option forces all cells to appear in one panel. Specifying UNISCALE=COLUMN forces common data values in the column (X) axis, but not in the row (Y) axis.
- The HBAR requests a bar chart of CASES (response) by CAUSE (category), with the response values summed by category.
- Just as in figure 2, the input data set has been sorted in descending order by the response variable (CASES). The ROWAXIS statement DISCRETEORDER=DATA option forces the category values to be drawn as they appear in the data set. As before, this creates the effect of a waterfall showing the relative number of cancer diagnoses. This time, the effect is shown independently for each sex because it is used with the SG PANEL procedure, not the SG PLOT procedure.
- The UNISCALE=COLUMN option prevents a PROSTATE category in the FEMALE row, and prevents an OVARIAN category in the MALE row.



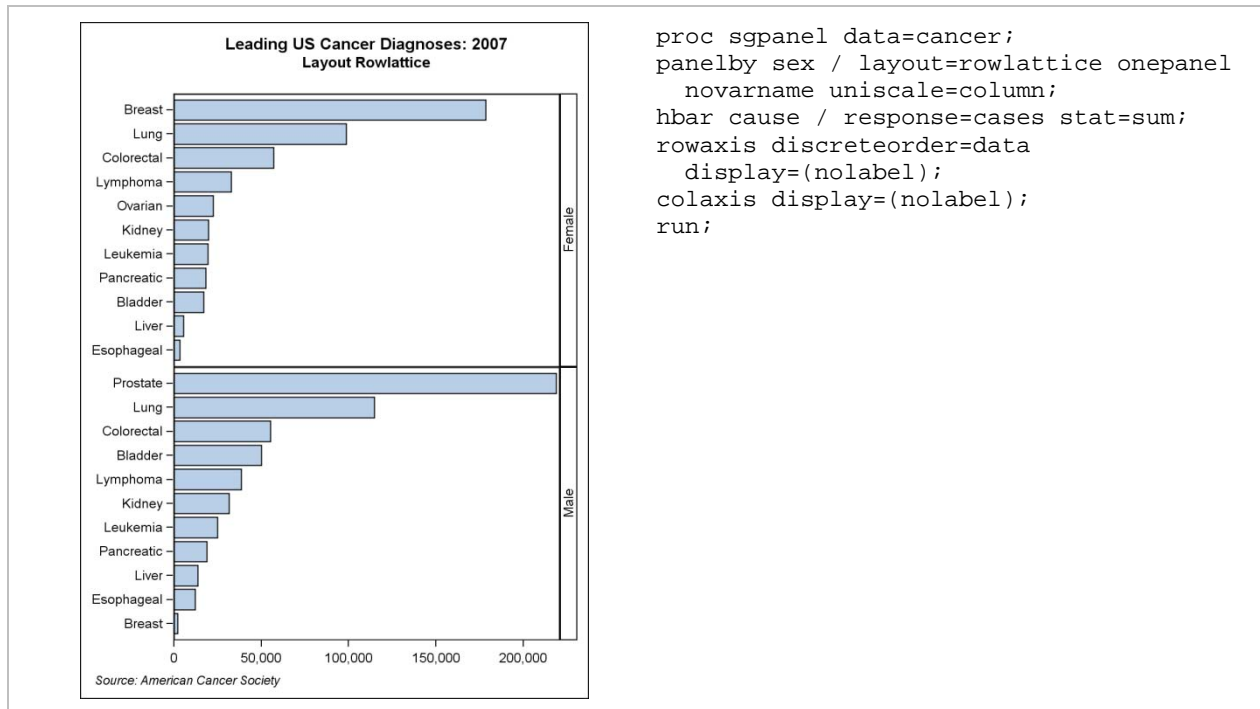


Figure 9. Horizontal Bar Chart in a ROWLATTICE Layout

## THE SGSCATTER PROCEDURE

The SGSCATTER procedure creates a panel of one or more scatter plots. The scatter plots can be displayed individually or in a matrix.

PROC SGSCATTER offers three different types of panels:

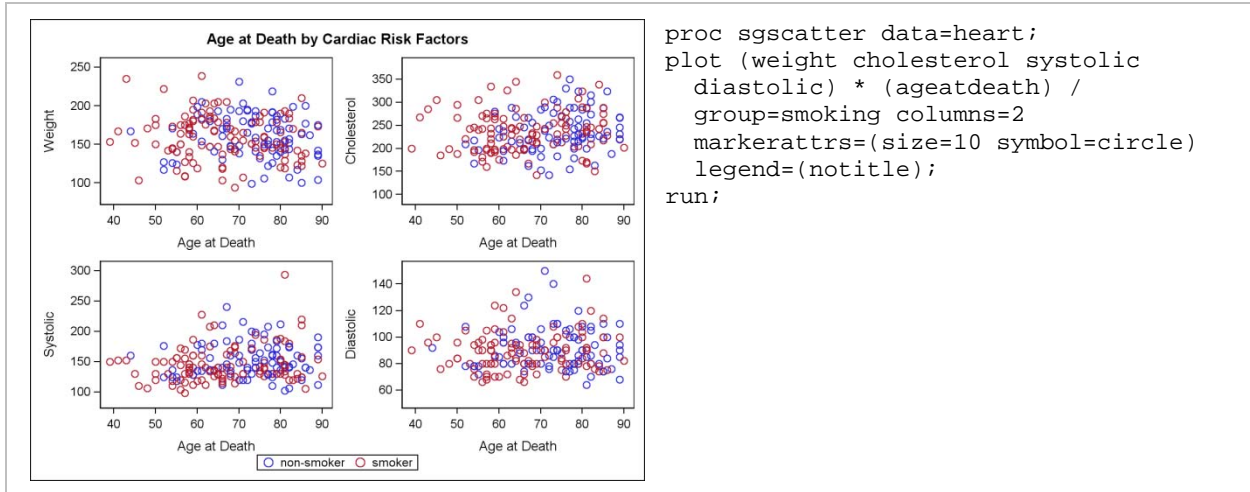
- a paneled graph of scatter plots where each plot cell has its own independent set of axes
- an MxN rectangular matrix, in which all cells in a row have the same (row) axis values, and all cells in a column have the same (column) axis values
- a scatter plot matrix, in which each of the variables is plotted against the others

Figure 10 shows a paneled graph of independent scatter plots.

- The variables WEIGHT, CHOLESTEROL, SYSTOLIC, and DIASTOLIC are the independent variables, the variable AGEATDEATH is the dependent variable, and the variable SMOKING is the group. Two columns are requested.

The result is four scatter plots of 1) WEIGHT x AGEATDEATH, 2) CHOLESTEROL x AGEATDEATH, 3) SYSTOLIC x AGEATDEATH, and 4) DIASTOLIC x AGEATDEATH. The group legend is displayed automatically.

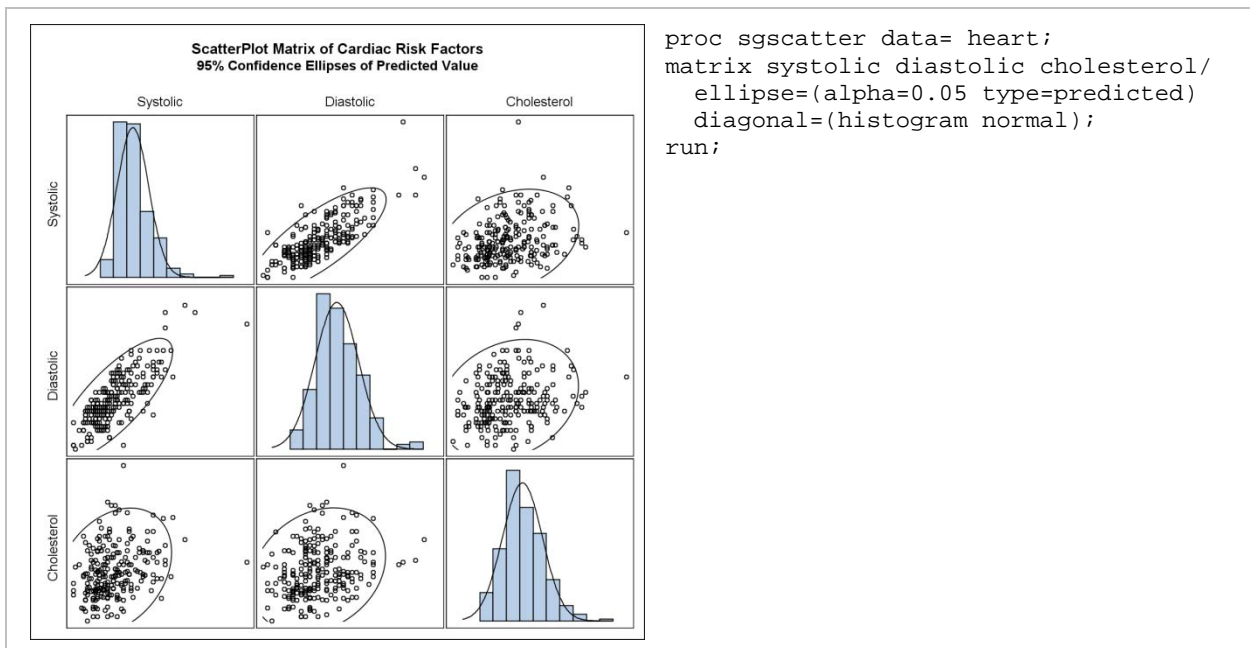




**Figure 10. Paned Graph of Independent Scatter Plots**

By contrast, Figure 11 shows a scatter plot matrix.

- The rows and columns of the matrix contain the variables SYSTOLIC, DIASTOLIC, and CHOLESTEROL.
- Each scatter plot is overlaid with a 95% confidence ellipse of the predicted value.
- Each diagonal contains a histogram and normal density curve of its corresponding variable.



**Figure 11. Scatter Plot Matrix**

## LABORATORY PANELS

There are several ways to construct multi-cell graphs using the SG PANEL and SG SCATTER procedures. This section demonstrates how to use these tools to construct panels of laboratory results. All laboratory panels in this section (figures 12 through 14) were generated with the STATISTICAL style.

Figure 12 demonstrates how to generate a high-level overview of the White Blood Count (WBC) results for all patients during a six-week clinical trial. The input data set contains these values:

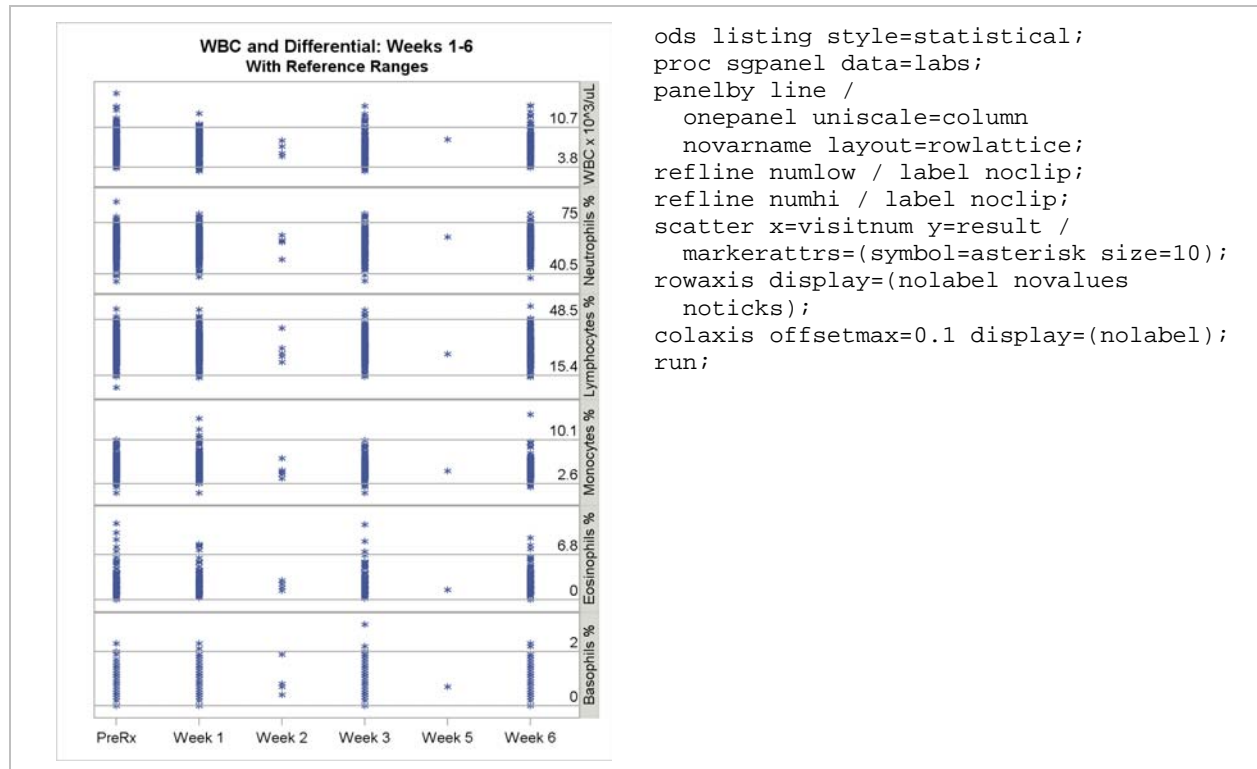
Testname	Numlow	Numhi	Result	Visitnum	PatID	Panel
Neutrophils (%)	40.5	75.0	48.60	PreRx	102	Hematology
Lymphocytes (%)	15.4	48.5	41.60	PreRx	102	Hematology
etc.....						

Note that each observation contains the appropriate high and low values of its normal range.

- There is one PANELBY variable, LINE. LINE was obtained from the original test name and contains a number from 1 to 6. LINE is used to control the order of the laboratory test names in the panel. A format has been applied to transform LINE back to the original test name. The PANELBY statement specifies the NOVARNAME and ONEPANEL options. The UNISCALE=COLUMN option requests common data values on the column (X) axis, but not on the row (Y) axis.

The contents of each cell are then specified:

- The first REFLINE statement draws a horizontal line for the low normal range, with Y=NUMLOW, as specified in the data set. The NOCLIP option protects against the situation in which the low normal range is outside the range of all data values. Specifying NOCLIP forces the data range to expand, if necessary, to retain the reference line.
- The second REFLINE statement draws a horizontal line for the high normal range, with Y=NUMHI, as specified in the data set. The NOCLIP option is specified.
- The SCATTER statement draws one marker for each X=VISITNUM and Y=RESULT.
- The ROWAXIS statement turns off the row (Y axis) label, tick marks, and tick values.
- The COLAXIS statement turns off the column (X axis) label, and sets the OFFSETMAX option to 0.1. This has the effect of moving the columns slightly to the left and creating more space for the REFLINE labels.



**Figure 12. Laboratory Panel of WBC Results**

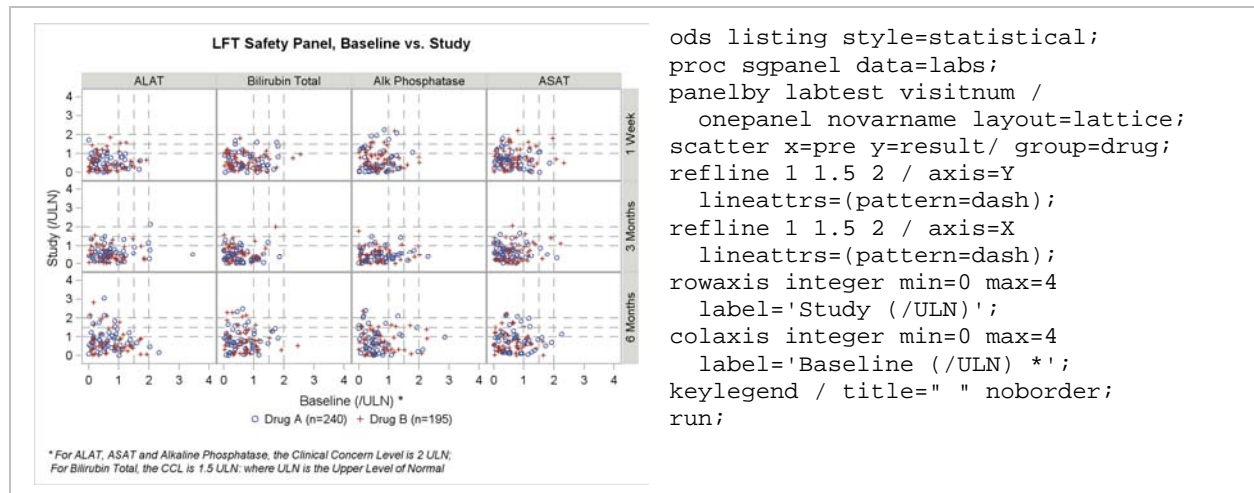
Figure 13 shows how to construct a panel of Liver Function Tests (LFTs), comparing each patient's pretreatment value with their one week, three-month, and 6-month results. Instead of displaying actual laboratory results, the panel displays the results as the ratio of the Upper Limit of Normal (ULN). The input data set contains these values:

Visitnum	Labtest	Drug	Pre	Result
1 Week	ALAT	Drug A (n=240)	0.27216	0.64893
1 Week	Bilirubin Total	Drug A (n=240)	1.28536	0.54600
etc.....				

- There are two PANELBY variables, LABTEST and VISITNUM. The PANELBY statement specifies the NOVARNAME and ONEPANEL options and a LATTICE layout. Thus, each column contains one of the four LFTs, each row contains one of the study's three visits, and each cell contains each patient's baseline by visit results.

The contents of each cell are then specified:

- The SCATTER statement draws a marker for each X=PRE and Y=RESULT pair. The group is set to the DRUG (treatment group) variable. The markers use the attributes of the GRAPHDATA1 and GRAPHDATA2 elements.
- The first REFLINE statement draws horizontal lines at Y=(1, 1.5, and 2) to indicate study values of 1 ULN, 1.5 ULN, and 2 ULN.
- The second REFLINE statement draws vertical lines at X=(1, 1.5, and 2) to indicate baseline values of 1 ULN, 1.5 ULN, and 2 ULN.
- The ROWAXIS and COLAXIS values are explicitly set to be integers from 0 to 4.
- The group KEYLEGEND is drawn without a border and displays at the bottom by default.



**Figure 13. Panel of LFT Results**

Figure 14 also shows a panel of LFT results, but only for at-risk patients. The data values are the same as in figure 13, but the data set is organized differently:

Patient	alat	biltot	alkph	asat	days	dval	sdays
Patient 5152: White Male Age 48; Drug: A	0.50000	0.40000	0.40000	0.50000	-25	.	.
Patient 5152: White Male Age 48; Drug: A	1.36533	0.64286	0.68972	1.08498	0	-0.5	0
Patient 5152: White Male Age 48; Drug: A	2.18967	0.87118	0.97332	1.64750	25	-0.5	25
etc.....							

- There is one PANELBY variable, PATIENT, and the NOVARNAME option is used. LAYOUT= is not specified, so the default PANEL layout is used. Each cell displays one patient.

The contents of each cell contain:

- The first SERIES statement draws the ALAT values with X=RELDAY and Y=ALAT.
- The next three SERIES statements draw the ASAT, ALKPHOS, and BILTOT values for each X=RELDAY and Y=<testname>.
- The BAND statement draws with X=(0 to SDAY) and Y=(MINY to 4.5), where both SDAY and MINY are defined in the data set. This has the effect of drawing a band from (X, Y)=(0, -0.5) to (150, 4.5) in the first

cell. The band is different in the second cell because the SDAY values are different. The band represents the time that the patient remained in the trial.

- PROC SGPANEL draws the four SERIES statements with the attributes of GRAPHDATA1 through GRAPHDATA4, and the BAND with the attributes of GRAPHDATA5.
- The REFLINE statement draws horizontal lines at Y=(1, 1.5, and 2) to indicate LFT values of 1 ULN, 1.5 ULN, and 2 ULN.
- The COLAXIS minimum and maximum values are set explicitly.
- The legend for the four SERIES plots and the one BAND plot is automatic.

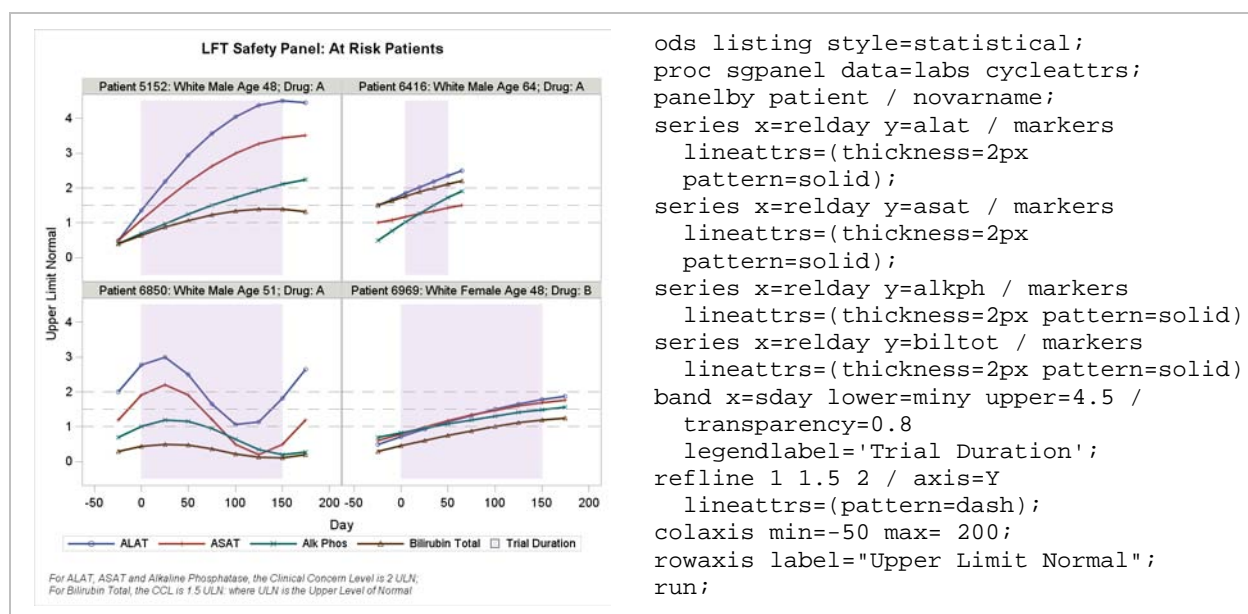


Figure 14. Panel of LFT Results for At-Risk Patients

## AGGREGATING FORMATS

The classification variables used in the previous PROC SGPANEL examples have all contained discrete values. You can create your own classification variables from continuous data by using user-defined formats to map continuous values into discrete buckets. This is referred to as creating classifiers on the fly. Figure 15 contains a LATTICE layout with two such variables, one for the column and one for the row.

- PROC FORMAT creates user-defined formats for SMOKFMT (the column classifier) and DIASFMT (the row classifier).
- The FORMAT statement assigns the user-defined formats to the continuous variables.
- The PANELBY statement uses the formatted variables as the classifiers. Because the variable SMOKING contains missing values (that indicate non-smoker), the PANELBY statement contains the option MISSING so that the missing values are retained. The user-defined format puts these missing values in the NON-SMOKER bucket. The PANELBY statement also specifies the NOVARNAM and ONEPANEL options.

The contents of each cell are then defined:

- a histogram of WEIGHT overlaid with a normal density curve
- a ROWAXIS statement with integer tick values and horizontal grid lines, and a COLAXIS statement with vertical grid lines

The user-defined classifiers are independent of each other. Every observation is assigned to one of the six (2x3) cells, and then the histogram and normal density curves are computed for the values assigned to the cells.

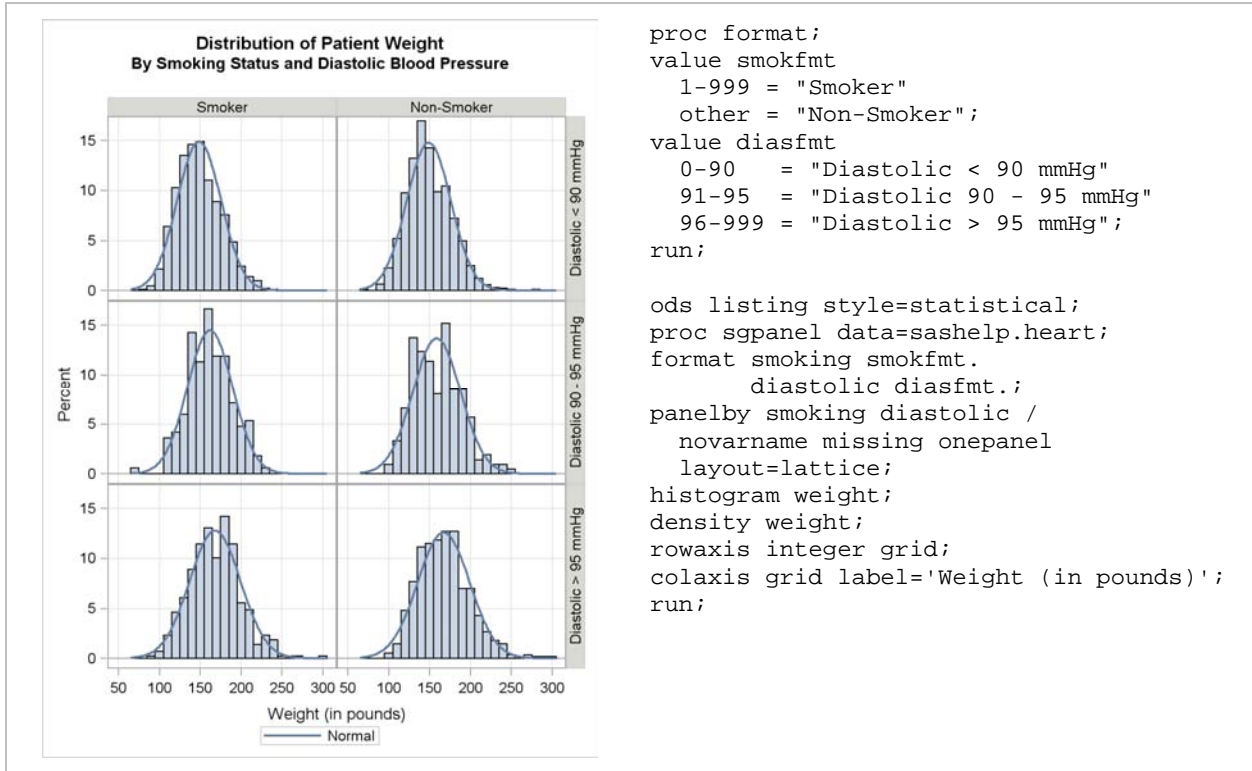


Figure 15. Two Aggregating Formats Used in a LATTICE Layout

### GROUPED BAR CHARTS

Bar charts, especially grouped bar charts, are common in the reporting of clinical trial data. The SGPLOT and SGPANEL procedures contain several options that you can use to generate bar charts. Figures 16 through 18 were generated with the ANALYSIS style and use the same input data set, but assign the CATEGORY, RESPONSE, and GROUP variables differently. The input data set contains three variables with these values:

```

trt      pain    percent
Placebo  0         12%
Placebo  1-2      14%
Placebo  3-4      16%
etc.....
    
```

The variables are assigned as shown:

<u>Bar Chart Type</u>	<u>PANELBY</u>	<u>CATEGORY</u>	<u>RESPONSE</u>	<u>GROUP</u>
Stacked Group Bar Chart	—	TRT	PERCENT	PAIN
Adjacent Group Bar Chart	PAIN	TRT	PERCENT	TRT
Latticed Group Bar Chart	TRT	PAIN	PERCENT	TRT

### STACKED GROUP BAR CHART

Figure 16 shows a stacked group bar chart, in which each group represents a proportionate share of the total. The bars change appearance for each group value, cycling through the GRAPHDATA1 to GRAPHDATA5 style elements.

- The stacked bar chart is generated with the SGPLOT procedure.
- The VBAR (vertical bar chart) statement has CATEGORY=TRT, RESPONSE=PERCENT, and GROUP=PAIN.
- The XAXIS statement suppresses display of the X axis label, and the YAXIS statement explicitly sets the Y axis label.

- The KEYLEGEND statement sets the legend title explicitly and displays a chicklet for each of the five group values.

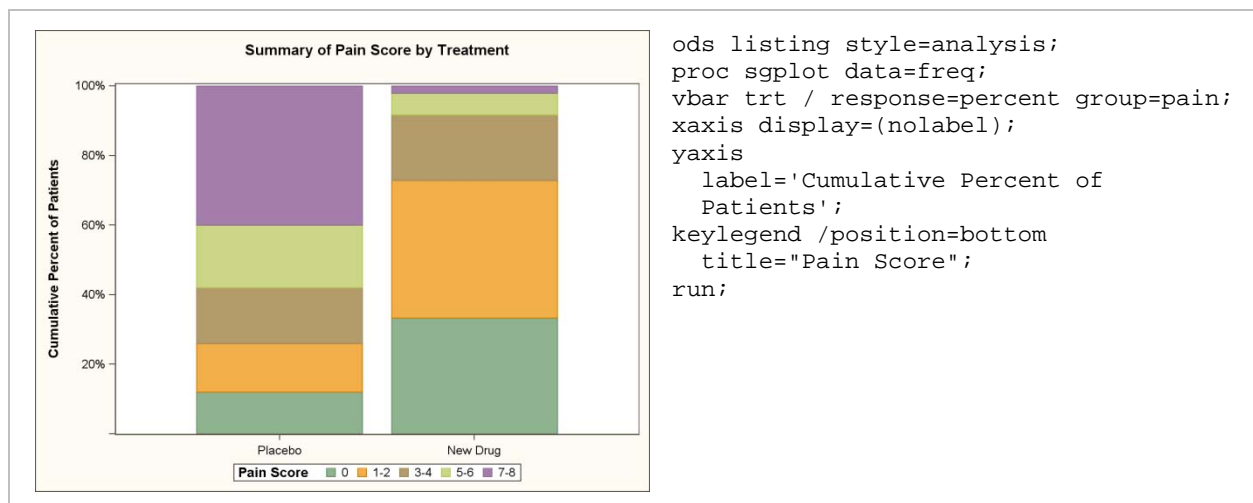


Figure 16. Stacked Group Bar Chart

## ADJACENT GROUP BAR CHART

Figure 17 shows an adjacent group bar chart. Each group value has its own column, with the categories within the group adjacent to each other. The bars change appearance for the two category values, cycling through the GRAPHDATA1 and GRAPHDATA2 style elements.

- The adjacent group bar chart is generated with PROC SG PANEL.
- The PANELBY statement uses PAIN as the column variable, with the PAIN values displayed at the bottom of the column.
- The VBAR statement uses PERCENT as the response variable and TRT as both the category and the group variables. Using TRT as the group variable is done to cycle the group colors.
- The ROWAXIS statement draws horizontal grid lines and sets OFFSETMAX=.05. This adjusts the tick values down slightly and prevents clipping the top tick value.
- The COLAXIS statement suppresses display of the column axis label, tick marks, and values. If present, they would display above the column headings, which would compromise the appearance of the graph. Instead, the string **Pain Score** is specified with a FOOTNOTE.
- The KEYLEGEND statement specifies a one-column legend on the right.

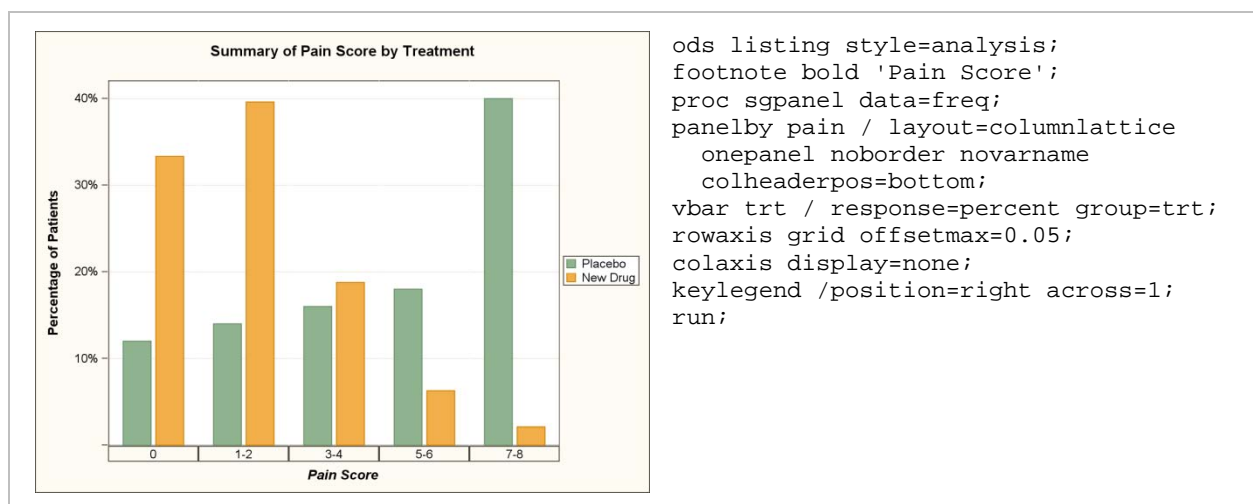


Figure 17. Adjacent Group Bar Chart



## LATTICED GROUP BAR CHART

Figure 18 shows a latticed group bar chart. This example uses a COLUMNLATTICE layout; however, a ROWLATTICE layout would work as well. Each group has its own column, and each column contains all of the category values. The bars change appearance for the two group values, cycling through the GRAPHDATA1 and GRAPHDATA2 style elements.

- The latticed group bar chart is generated with PROC SG PANEL.
- The PANELBY statement uses TRT as the column variable and specifies the COLUMNLATTICE layout.
- The VBAR statement uses PERCENT as the response variable, PAIN as the category variable, and TRT as the group variable. Using TRT as both the PANELBY variable and the group variable is done to cycle the group colors.
- The ROWAXIS statement draws horizontal grid lines.

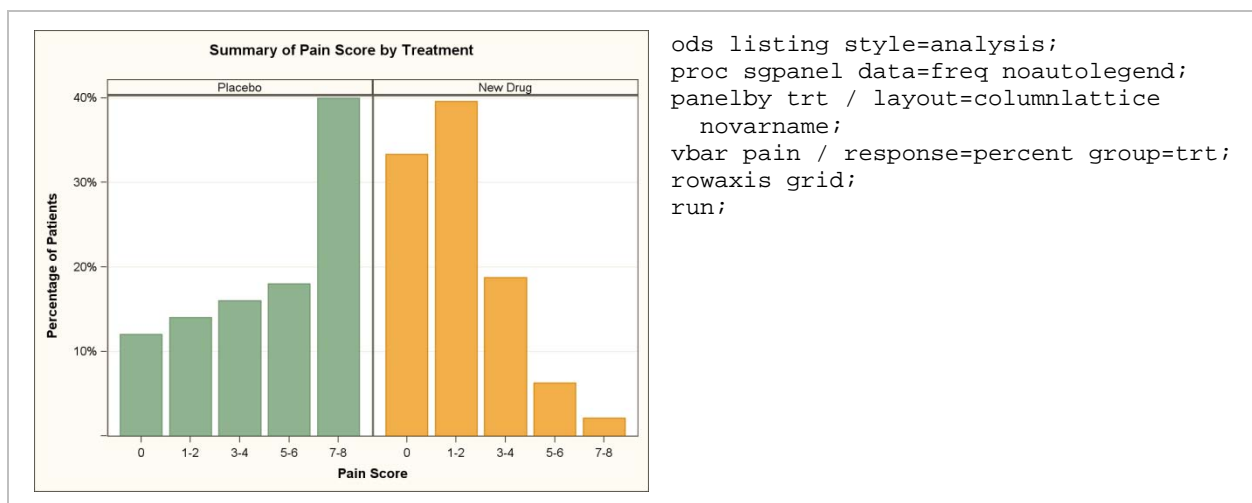


Figure 18. Latticed Group Bar Chart

## BUTTERFLY GROUP BAR CHART

A butterfly plot is a plot drawn across a centered axis. The butterfly plot is particularly useful when comparing two halves of a whole. Figure 19 shows an example of a butterfly grouped bar chart that is generated with PROC SG PLOT and the LISTING style.

The input data set contains six variables, where:

MCASES = -1 x (# of male diagnoses)      MDEATHS = -1 x (# of male deaths)  
 FCASES = # of female diagnoses          FDEATHS = # of female deaths  
 CAUSE = cancer diagnosis                DEATHS = [abs(mdeaths) + fdeaths]

cause	mcases	fcases	mdeaths	fdeaths	deaths
Lung Cancer	-114760	98620	-89510	70880	160390
Colorectal Cancer	-55290	57050	-26000	26180	52180
Breast Cancer	-2030	178480	-450	40460	40910
etc.....					

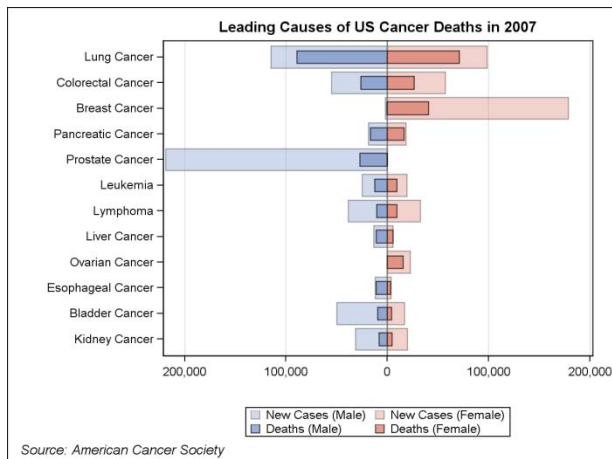
A trick is involved in the construction of the butterfly plot: The values to the left of the centered axis are actually negative numbers (and thus are negative bars), while the values to the right are positive numbers. The negative range of the axis displays with positive tick values by using a user-defined picture format. The user-defined picture format is named POSITIVE. POSITIVE displays numbers greater than 0 in a comma-separated format, numbers less than 0 in a comma-separated format without a negative sign, and the number 0 as itself. (Because 0 is not included in the range of values in the picture format, it is not formatted.)

In figure 19, the left (negative) side displays horizontal bars for males, and the right (positive) side displays horizontal bars for females. The larger, more transparent bars show the number of new diagnoses, while the smaller, more opaque bars show the number of deaths.



In PROC SGPLOT:

- The user-defined format POSITIVE is assigned to all four response variables MCASES, MDEATHS, FCASES, and FDEATHS.
- The first HBAR statement displays MCASES (the response) by CAUSE (the category). MCASES is negative, so the bars display to the left of the centered axis. The bar uses the fill color from GRAPHDATA1 with TRANSPARENCY=.65, which lightens the saturation. The LEGENDLABEL text and a NAME= are assigned. These settings will be used later to construct the butterfly legend.
- The second HBAR statement displays MDEATHS by CAUSE. MDEATHS is also negative, so it displays to the left of the centered axis. The bars use GRAPHDATA1 for the fill color, but with more saturation (TRANSPARENCY=.25). The bar width is set to .5, half of the bar width in the first HBAR, and appears on top of the first HBAR. Again, the LEGENDLABEL text and a NAME= are assigned for later use.
- The third and fourth HBAR statements repeat the process for FCASES and FDEATHS by CAUSE. FCASES and FDEATHS are positive, so they display to the right of the centered axis. These two bars use GRAPHDATA2 for fill color.
- The KEYLEGEND statement explicitly states the order of the chicklets in the legend. The default order is to follow the order of the plot statements. In the KEYLEGEND statement, each chicklet is added explicitly by NAME=, with the chicklet text from LEGENDLABEL=.
- The XAXIS statement sets a blank axis label and vertical grid lines.
- The YAXIS statement sets a blank axis label and stipulates that the category values be drawn as they appear in the data set. The input data set is sorted in descending order by deaths, so the butterfly plot is also showing that lung cancer is the most deadly cancer, with colorectal cancer the second most deadly. Note that breast cancer causes more female deaths, but colorectal cancer causes more total deaths.



```
proc format;
picture positive low-<0='000,000'
                0<-high='000,000';
run;

proc sort data=cancer;
  by descending deaths;
run;

proc sgplot data=cancer;
format mcases mdeaths
       fcases fdeaths positive.;
hbar cause / response=mcases
  fillattrs=graphdata1 transparency=.65
  legendlabel="New Cases (Male)"
  name="mcases" ;
hbar cause / response=mdeaths
  barwidth=.5
  fillattrs=graphdata1 transparency=.25
  legendlabel="Deaths (Male)"
  name="mdeaths" ;
hbar cause / response=fcases
  fillattrs=graphdata2 transparency=.65
  legendlabel="New Cases (Female)"
  name="fcases" ;
hbar cause / response=fdeaths
  barwidth=.5
  fillattrs=graphdata2 transparency=.25
  legendlabel="Deaths (Female)"
  name="fdeaths" ;
keylegend "mcases" "fcases" "mdeaths"
         "fdeaths" / across=2;
xaxis label=" " grid;
yaxis label=" " discreteorder=data;
run;
```

Figure 19. Butterfly Plot

## AXES

An axis always has an axis type. The underlying GTL determines the axis type from the plot statements that are used. Axes can be interval (numeric data where the distance between ticks is meaningful), or non-interval (discrete data where the distance between ticks is not meaningful). The XAXIS and YAXIS statements in PROC SGPLOT and the ROWAXIS and COLAXIS statements in PROC SGPANEL contain a TYPE= option that allows you to assign your own axis type, but you cannot assign an axis type incompatible with the axis variable. The axis types available are:

LINEAR	Use a standard interval axis if possible. Axis values must be numeric.
LOG	Use a logarithmic axis if possible. Axis values must be numeric.
TIME	Use a time axis if possible. All axis variables must be numeric SAS date, time, or datetime values.
DISCRETE	Use a non-interval axis if possible. Axis values can be character or numeric.

There are two situations in which you might want to override the automatic assignment of axis types. One is to create a log axis when the default axis is linear, and the other is to create a discrete axis when the default axis is linear. GTL uses a best fit algorithm to choose and format axis tick values. You can override part, or all, of the algorithm with the VALUES= and FITPOLICY= options.

Figure 20 shows an example of a logarithmic scale on the Y axis. The input data set contains these data values:

```
age    Cpeptide
5.2    4.8
8.8    4.1
10.5   5.2
10.6   5.5
etc.....
```

- The PBSPLINE statement creates a scatter plot and adds a penalized B-spline model fit line.
- The REFLINE statement draws a horizontal line at 4.4817 ( $e^{1.5} = 4.4817$ ). The line is labeled with the exponent.
- The YAXIS statement specifies the LOG axis, LOGBASE=e, and LOGSTYLE=LOGEXPONENT.

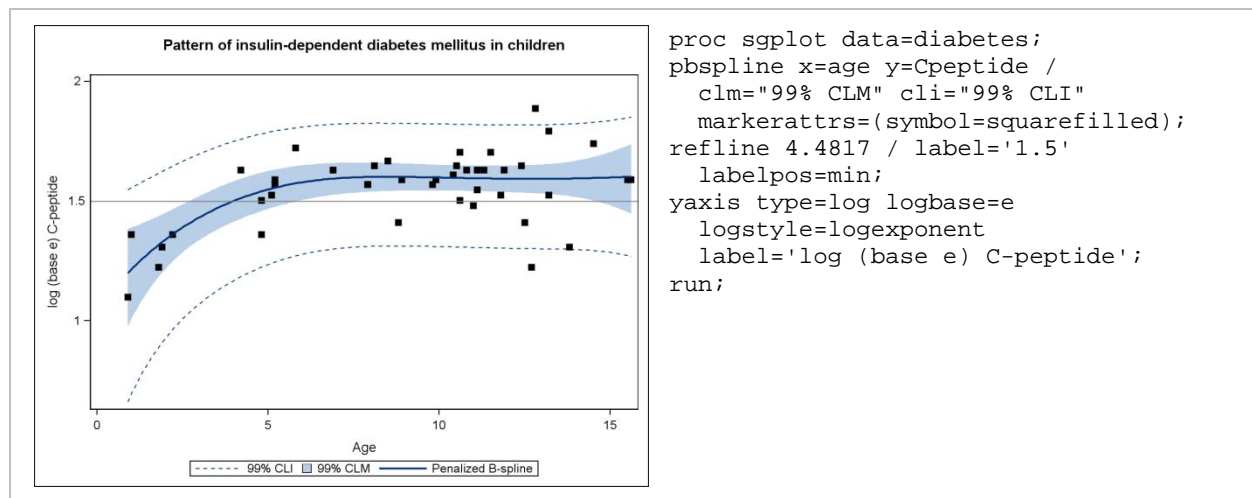


Figure 20. Logarithmic Scale on the Y Axis

## GRAPH TEMPLATE LANGUAGE

GTL is the foundation upon which the SG procedures are built. While the SG procedures are powerful, they do not surface all of the options in GTL. You can, however, use the SGPLOT and SGSCATTER procedures to generate GTL code, which you can then customize as needed. To generate GTL, use the TMPLOUT=FILENAME option on

the procedure statement. FILENAME will contain the GTL template. Update the template with your changes, and add a PROC SGRENDER to render the graph. As an example, consider the following simple PROC SGPLOT:

```
proc sgplot data=sashelp.class;
  scatter x=weight y=height;
  refline 100 / axis=x labelloc=inside label='Weight=100 lb' ;
run;
```

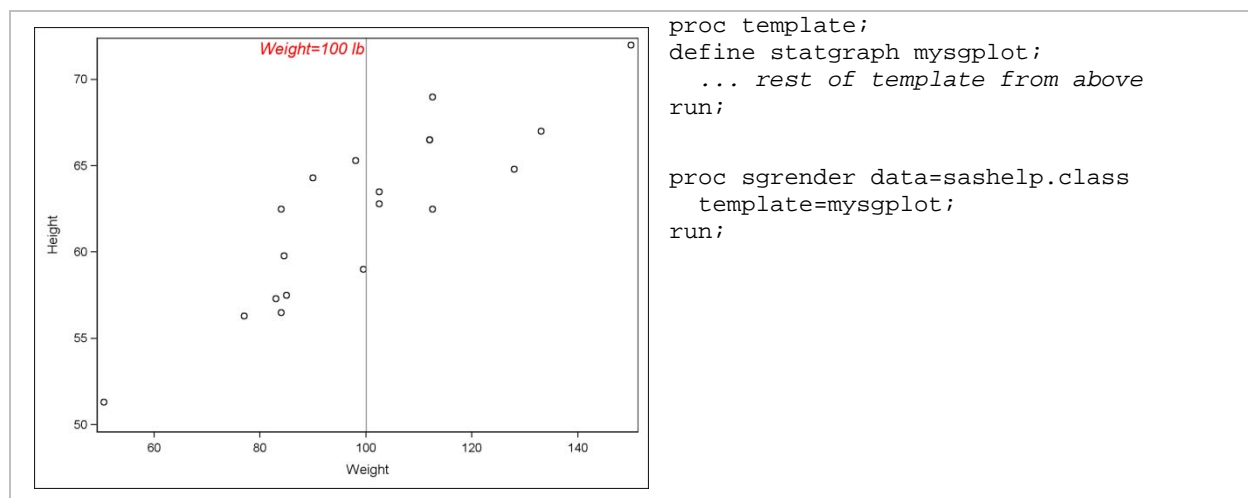
The SGPLOT procedure does not have an option to change the attributes of the REFLINE label. Next, generate the GTL template and add the text attributes:

```
proc sgplot data=sashelp.class tmlout="tmpl.sas"; ... run;
```

Add the attributes in the REFERENCELINE statement in the tmpl.sas file. For clarity, change the name of the template from SGPLOT (the procedure name is the default) to MYSGPLOT:

```
proc template;
  define statgraph mysgplot;
  begingraph;
  layout overlay;
  ScatterPlot X=Weight Y=Height / primary=true LegendLabel="Height" name="scatter";
  ReferenceLine x=100 / clip=true CurveLabel="Weight=100 lb"
  CurveLabelLocation=Inside curveLabelAttrs=(color=red style=italic size=12pt);
  endlayout;
  endgraph;
end;
run;
```

Submit PROC TEMPLATE, and then submit PROC SGRENDER, as shown in Figure 21.



**Figure 21. Customized GTL Code**

See the *SAS/GRAPH 9.2: Graph Template Language User's Guide* for more information.

## INTEGRATION WITH ODS

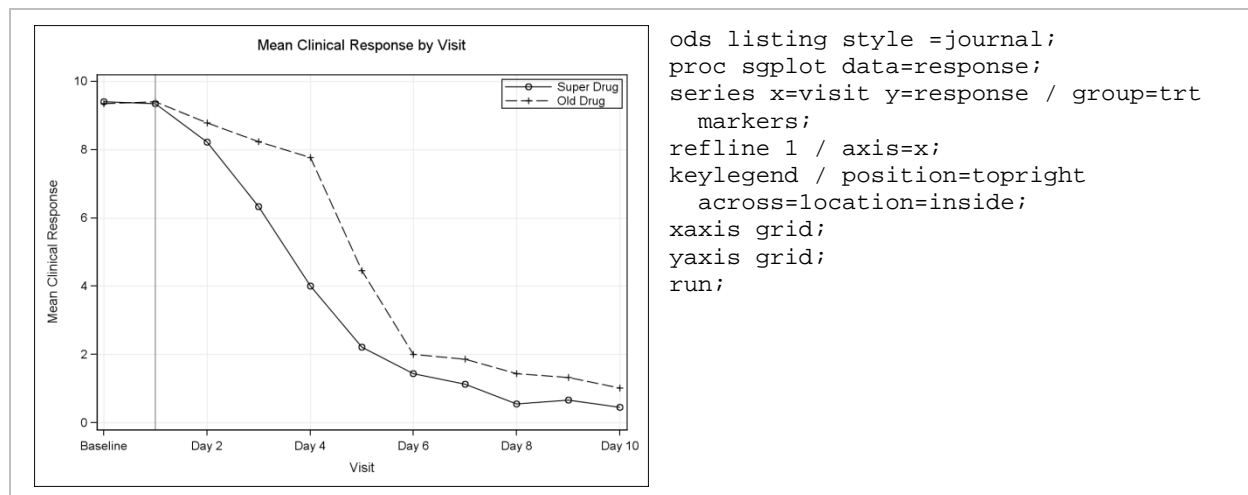
The SAS/GRAPH SG procedures (and GTL) are fully integrated with ODS. You can use ODS statements to customize your graph's appearance, and you can extract ODS objects directly to use as input to the SG procedures. The following techniques are explained.

```
ODS destination STYLE=style-definition IMAGE_DPI/DPI=dpi-setting SGE=on|off;
```

DESTINATION can be LISTING, HTML, PDF, RTF, and so on. STYLE-DEFINITION can be any of the ODS styles.

As noted earlier, the graphs in this paper were generated with the LISTING, ANALYSIS, or STATISTICAL style. These three styles, plus the JOURNAL (gray scale) and JOURNAL2 (black and white) styles, were designed for the effective presentation of graphics.

Figure 22 was generated with the same SG procedure code as in figure 1, but with the JOURNAL style instead of the LISTING style.



**Figure 22. Dose Response Plot with JOURNAL Style**

DPI-SETTING stands for dots per inch and indicates the requested resolution of the graph. Each ODS destination has its own default DPI setting.

SGE=ON|OFF—SGE=ON requests that a second output file be generated (an \*.sge file) along with the image. This option is available only with the LISTING destination.

The .sge file is used by the SAS/GRAPH ODS Graphics Editor, which is an interactive graphical application used to edit and annotate ODS graphics files.

See the *SAS/GRAPH 9.2: ODS Graphics Editor User's Guide* for more information.

```

ODS GRAPHICS / HEIGHT=dimension WIDTH=dimension IMAGEFMT=image-file-type | STATIC
IMAGENAME='filename';

```

DIMENSION is the height and width of any graph. The SG procedures determine a best use image size, but use the HEIGHT= and WIDTH= dimensions when available.

IMAGE-FILE-TYPE is the image format (PNG, GIF, and so on). STATIC is the best quality static image format for the requested ODS destination.

FILENAME is the base name. If this is not specified, the name of the SG procedure is used. The filename is incremented automatically.

See the *SAS 9.2 Output Delivery System User's Guide* for more information.

```

ODS OUTPUT object = filename;

```

You can ask ODS to produce a SAS data set from an ODS output object. Figure 23 shows an example from the *SAS/STAT User's Guide*. The NLMIXED procedure generates an ODS OBJECT named PARAMETERESTIMATES. PARAMETERESTIMATES is written to a data set named EST. EST is used to generate a second data set named PRED, and PRED is used to generate a PROC SGPLOT.

See the *SAS/STAT 9.2 User's Guide* for more information.

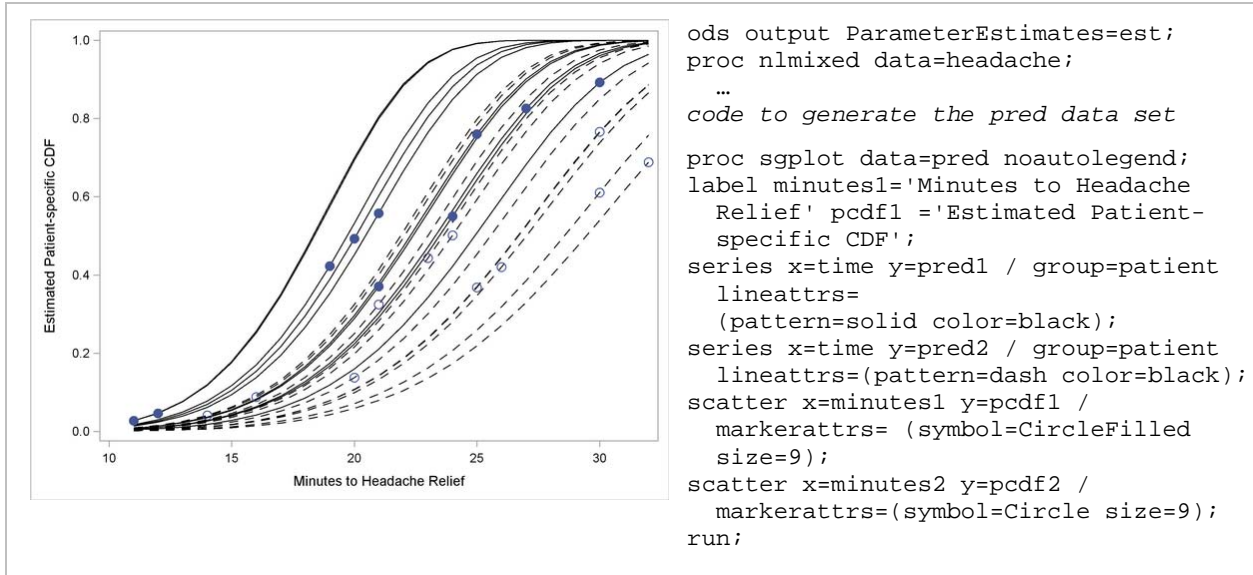


Figure 23. PROC NLMIXED and PROC SGPLOT

## USING SAS STATEMENTS

The SG procedures support TITLE, FOOTNOTE, BY (GROUP), FORMAT, and LABEL statements. PROC SG PANEL can contain both a BY variable and a PANELBY variable. The BY variable takes precedence, and the PANELBY variable generates one or more panels within each BY (GROUP) value.

The SG procedures do not support the SAS/GRAPH statements GOPTIONS, LEGEND, AXIS, PATTERN, and SYMBOL. GTL (and the SG procedures) derive their appearance options from the active ODS style, with overrides specified with the MARKERATTRS=, LINEATTRS=, and FILLATTRS= options.

## SG PROCEDURE STATEMENTS

Many plot statements are common to both the SG PLOT and SG PANEL procedures. These statements are:

Basic Plots	Categorical Plots	Fit Plots	Distribution Plots	Other
BAND NEEDLE SCATTER SERIES STEP	DOT HBAR HBOX HLINE VBOX VBAR VECTOR VLINE	LOESS PBSPLINE REG	DENSITY HISTOGRAM	KEYLEGEND REFLINE

Other statements are specific to only one SG procedure:

SGPLOT	SGPANEL	SGSCATTER
ELLIPSE INSET XAXIS X2AXIS YAXIS Y2AXIS	PANELBY COLAXIS ROWAXIS	COMPARE MATRIX PLOT

## CONCLUSION

The SGPLOT, SGPANEL, and SGSCATTER procedures provide new tools to display and summarize results from clinical trials. The ability to overlay plots gives you the flexibility to create graphs specific to your needs. ODS styles enable you to simply and effectively use colors. The paneling features provide a concise way to view data by multiple levels of classification. And, best of all, the procedure syntax is straightforward and easy to use.

## REFERENCES

Wang, Duolao, and Ameet Bakhai. 2006. *Clinical Trials: A Practical Guide to Design, Analysis, and Reporting*. London: Remedica Publishing.

Dmitrienko, Alex, et al. 2005. *Analysis of Clinical Trials Using SAS: A Practical Guide*. Cary, NC: SAS Institute Inc.

Shostak, Jack. 2005. *SAS Programming in the Pharmaceutical Industry*. Cary, NC: SAS Institute Inc.

Matthews, Carol, and Shilling, Brian. 2008. *Validating Clinical Trial Data Reporting with SAS*. Cary, NC: SAS Institute Inc.

The data for the butterfly plot are available from the American Cancer Society Web site: [www.cancer.org](http://www.cancer.org).

## ACKNOWLEDGMENTS

The author would like to thank Melisa Turner, Scott Singer, Dan Heath, and Sanjay Matange for their invaluable input.

## RECOMMENDED READING

*SAS/GRAPH 9.2: Statistical Graphics Procedures Guide*

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>.

Matange, Sanjay. 2008. "ODS Graphics Editor." *Proceedings of the SAS Global Forum 2008 Conference*. Available at <http://www2.sas.com/proceedings/forum2008>.

Matange, Sanjay. 2008. "Introduction to Graph Template Language." *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*.

## CONTACT INFORMATION

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

Susan Schwartz  
SAS Institute  
SAS Campus Drive  
Cary, NC 27513  
[susan.schwartz@sas.com](mailto:susan.schwartz@sas.com)

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.