Plotting Against Cancer: Creating Oncology Plots Using SAS®

Debpriya Sarkar, SAS Institute Inc., Pune, India

ABSTRACT

Graphs in oncology studies are essential for getting more insight about the clinical data. This presentation demonstrates how ODS Graphics can be effectively and easily used to create graphs used in oncology studies. We discuss some examples and illustrate how to create plots like drug concentration versus time plots, waterfall charts, comparative survival plots, and other graphs using Graph Template Language and the ODS Graphics procedures. These can be easily incorporated into a clinical report.

INTRODUCTION

Graphical representation of oncology data is widely used for reporting purposes. Graphs are used in the course of cancer studies to visualize cancer statistics, response to treatment, survival rates, or the amount of concentration of the drug over a period of time. In this paper, we will see some graphical examples from oncology studies and discuss how we can create them using SAS®. We will use SAS® 9.3 ODS Statistical Graphics (SG) procedures and Graph Template Language (GTL), which are a part of the ODS Graphics System, to create the graphs. We will illustrate how different graph elements along with options can be combined to create effective and high quality graphs.

STATISTICAL GRAPHICS PROCEDURES

The ODS Statistical Graphics procedures were first released with SAS® 9.2 in the SAS/GRAPH® product. These procedures are an extension of the Graph Template Language. The procedures are easy to use and support a variety of plots, which can be combined to create your graph. The procedures are as follows:

- The SGPlot procedure – creates single-celled graphs with the facility of overlaying.

- The SGPanel procedure – creates data driven plots based on a classifier variable(s).

- The SGScatter procedure – creates scatter matrices and comparative scatter plots.

In this section, I will show how to create the graphs using the SAS 9.3 SG procedures. The SAS 9.3 versions of these procedures support new features that make it easier to create the graphs.

SINGLE-CELLED GRAPHS

This SGPLOT procedure can create a variety of plot types and can overlay them on a single set of axes to produce many different types of graphs. We will now see some examples of oncology graphs that can be easily created using the SGPLOT procedure in SAS 9.3. The graphs shown in this presentation do not follow any particular order in the course of a clinical study or analysis.

Figure 1 shows a waterfall chart for volumetric reduction of brain tumors as an indicator of the performance of a drug. The Y-axis of the chart displays the percent change in the tumor size from baseline while the X-axis is the observation number. For a relatively large patient population, the waterfall chart gives an indication about the drug activity.

Table 1 shows a partial data set that is used for the waterfall chart. The change from the baseline is recorded along with the observation number. A group variable GRP is also present in the data.

Obs	grp	change	n
1	Volumetric Reduction	-100.00	1
2	Volumetric Reduction	-95.00	2
3	Volumetric Reduction	-90.00	3
4	Volumetric Reduction	-80.00	4
5	Volumetric Reduction	-70.00	5
6	Volumetric Reduction	-68.00	6
7	Volumetric Reduction	-67.00	7
8	Volumetric Reduction	-66.00	8
9	Volumetric Reduction	-65.00	9
10	Volumetric Reduction	-64.10	10

Table 1. Snippet of the Data Set Used for Figure 1
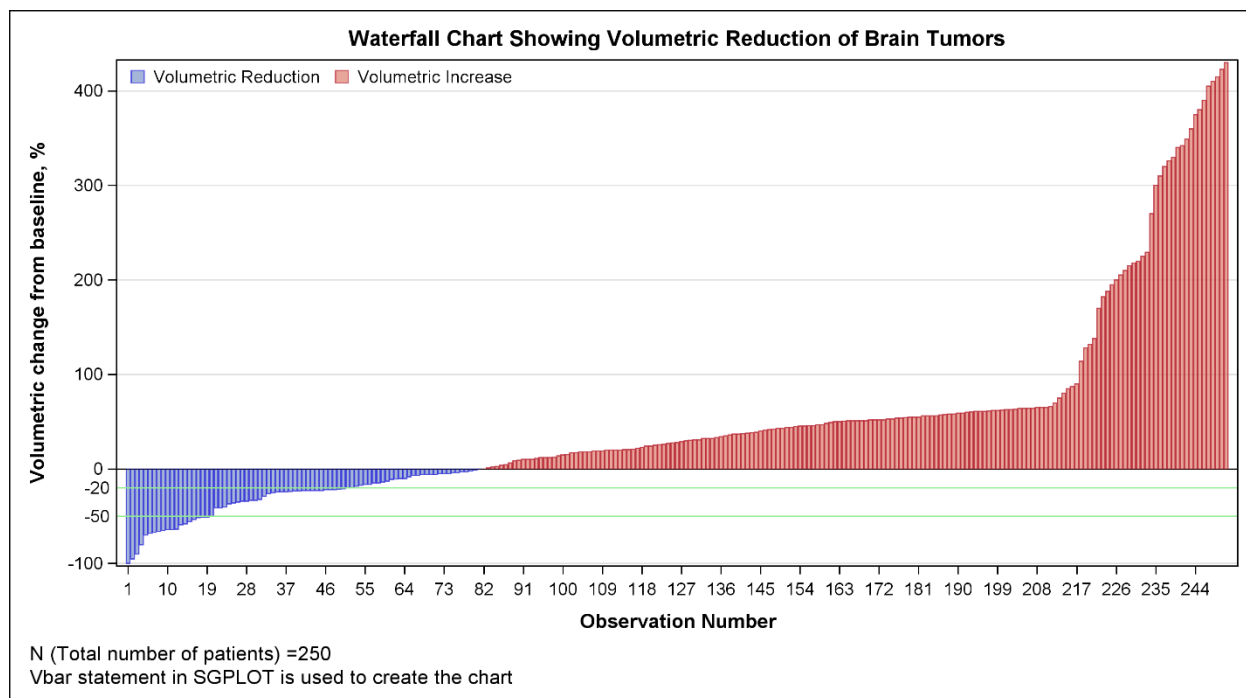

Figure 1. Waterfall Chart for Volumetric Reduction

CODE FOR FIGURE 1

```
title "Waterfall Chart Showing Volumetric Reduction of Brain Tumors";
footnote1 j=l "N (Total number of patients) =250";
footnote2 j=l "Vbar statement in SGPLOT is used to create the chart";

proc sgplot data=water3;
 vbar n / response=change group=grp transparency=0.3 name="vol";
```

```
 refline -50  / axis=y label="-50" labelloc=outside labelpos=min
lineattrs=(color=lightgreen);
 refline -20  / axis=y label="-20" labelloc=outside labelpos=min
lineattrs=(color=lightgreen);
 keylegend "vol" / noborder location=inside position=topleft;
 xaxis label="Observation Number" labelattrs=(weight=bold) fitpolicy=thin;
 yaxis label="Volumetric change from baseline, %" labelattrs=(weight=bold)
grid;
run;
```

Highlights of the graph shown in Figure 1:

- The SGPLOT procedure is used to create the chart.

- The input data is sorted in descending order by the response variable (CHANGE).

- A grouped bar chart is used to display the change from the baseline values in the tumors.

Figure 2 shows a similar waterfall chart for percent reduction of tumor measurements grouped by different response criteria. The Y-axis displays the percent change from baseline while the X-axis is the observation number.

Table 2 below is associated with the waterfall in Figure 2. The response criteria resides in the variable GROUP2, and the baseline changes are measured in variable CHNG2 along with the observation number in variable N.

Obs	group2	n	chng2
1	Progressive Disease	1	100.0
2	Progressive Disease	2	95.0
3	Progressive Disease	3	90.0
4	Progressive Disease	4	80.0
5	Progressive Disease	5	70.0
6	Progressive Disease	6	68.0
7	Progressive Disease	7	67.0
8	Progressive Disease	8	66.0
9	Progressive Disease	9	65.0
10	Progressive Disease	10	64.1

Table 2.1 Snippet of the Data Set Used for Figure 2

The attribute map feature enables you to control the visual attributes that are applied to specific data values in the graphs. The first step for attribute mapping is to create an SG attribute map data set, which associates data values with particular visual attributes. The second step is to modify the procedure and plot statements to use the data set.

Table 3.2 shows the SG attribute map data set that is consumed by the SGPLOT procedure. In this example, this data is used to assign different fill colors for the group values. Each observation defines the attribute for a particular response criteria. An observation uses reserved variable names for the attribute map identifier (ID) and the group value (VALUE) that contains the response criteria as shown in the variable GROUP2 of Table 3.1

FILLCOLOR, MARKERCOLOR, LINECOLOR, MARKERSUMBOL, LINEPATTERN are some other reserved variables in the data set that can be used to control the visual attributes of a graph.

Obs	value	fillcolor	id
1	Progressive Disease	lightseagreen	myid1
2	Stable Disease	lightblue	myid1
3	Complete Response	orange	myid1
4	Partial Response	gray	myid1

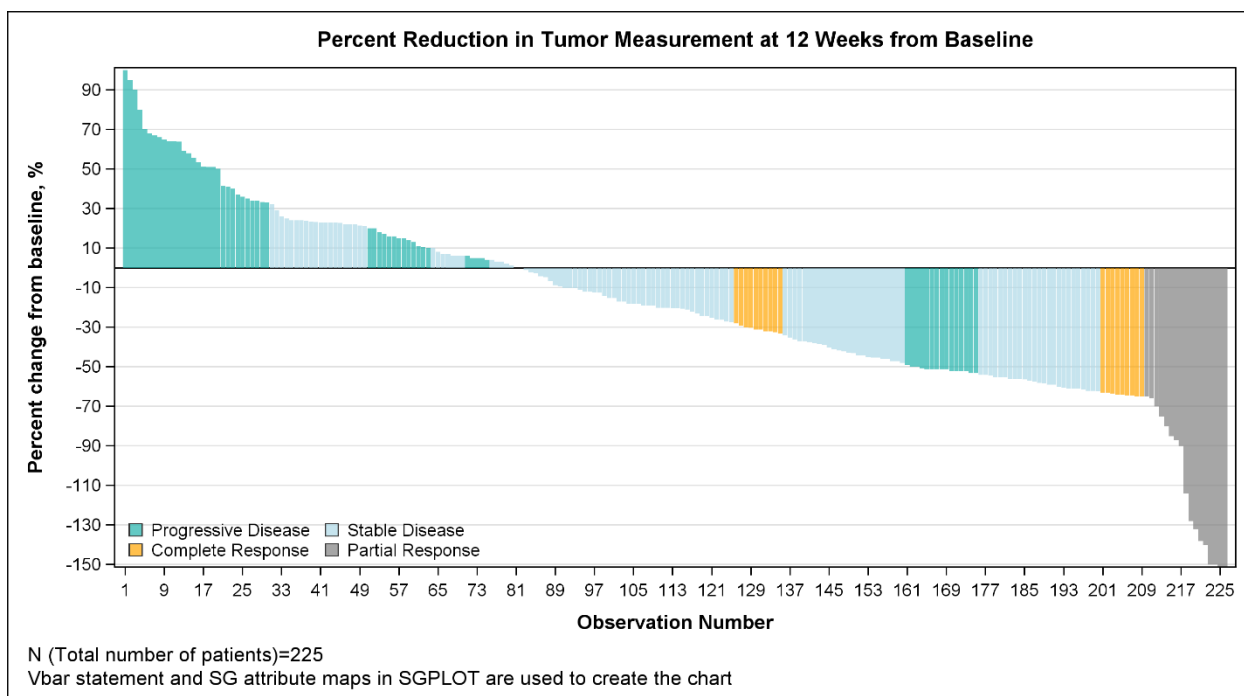**Table 2.2 Snippet of the SG Attribute Map Data Set Used for Figure 2**

Figure 2. Waterfall Chart for Percent Change by Response Criteria

CODE FOR FIGURE 2

```
title "Percent Reduction in Tumor Measurement at 12 Weeks from Baseline";
footnote1 j=l "N (Total number of patients)=225";
footnote2 j=l "Vbar statement and SG attribute maps in SGPLOT are used to
create the chart";
proc sgplot data=water4 dattrmap=attrmap;
 vbar n / response=chng2 group=group2 transparency=0.3 name="resp" nooutline
attrid=myid1;
 keylegend "resp"/ noborder title="" location=inside position=bottomleft
across=2;
 xaxis label="Observation Number" labelattrs=(weight=bold) fitpolicy=thin;
 yaxis label="Percent change from baseline, %" labelattrs=(weight=bold)
values=(-150 to 100 by 10) grid max=100;
run;
```

4

Highlights of the graph shown in Figure 2:

- An SG attribute map is used for customizing the fill colors of the grouped bar chart.
- The response (Y) axis of the bar chart is customized using the LINEAROPTS option in YAXISOPTS.

Figure 3 shows a skyline plot of hazard ratio of Time to Progression (TTP) over time along with the 95% confidence intervals as per the Independent Endpoint Review Committee (IRC).

Table 3 below is the partial data set used by Figure 8. The upper and lower confidence levels are given along with the hazard ratios, which are defined by the variables HIGH, LOW and HR respectively. The variables LABEL, X1, Y1, XORIGIN, and YORIGIN are created for the vector plot highlighting the time frame in Figure 8.

Obs	week	hr	high	low	label	x1	y1	xorigin	yorigin
1	1	0.30	1.50	0.10	November 15, 2005	85	-0.1	85	0.2
2	5	0.40	1.80	0.20	April 3, 2006	105	-0.1	105	0.2
3	10	0.25	1.10	0.08	.	.	.	.	.
4	15	0.20	0.90	0.12	.	.	.	.	.
5	20	0.18	0.88	0.05	.	.	.	.	.
6	25	0.30	1.10	0.15	.	.	.	.	.
7	25	0.32	1.20	0.16	.	.	.	.	.
8	30	0.35	1.25	0.15	.	.	.	.	.
9	35	0.40	1.10	0.22	.	.	.	.	.
10	40	0.37	1.23	0.20	.	.	.	.	.

Table 3 Snippet of the Data Set Used for Figure 3
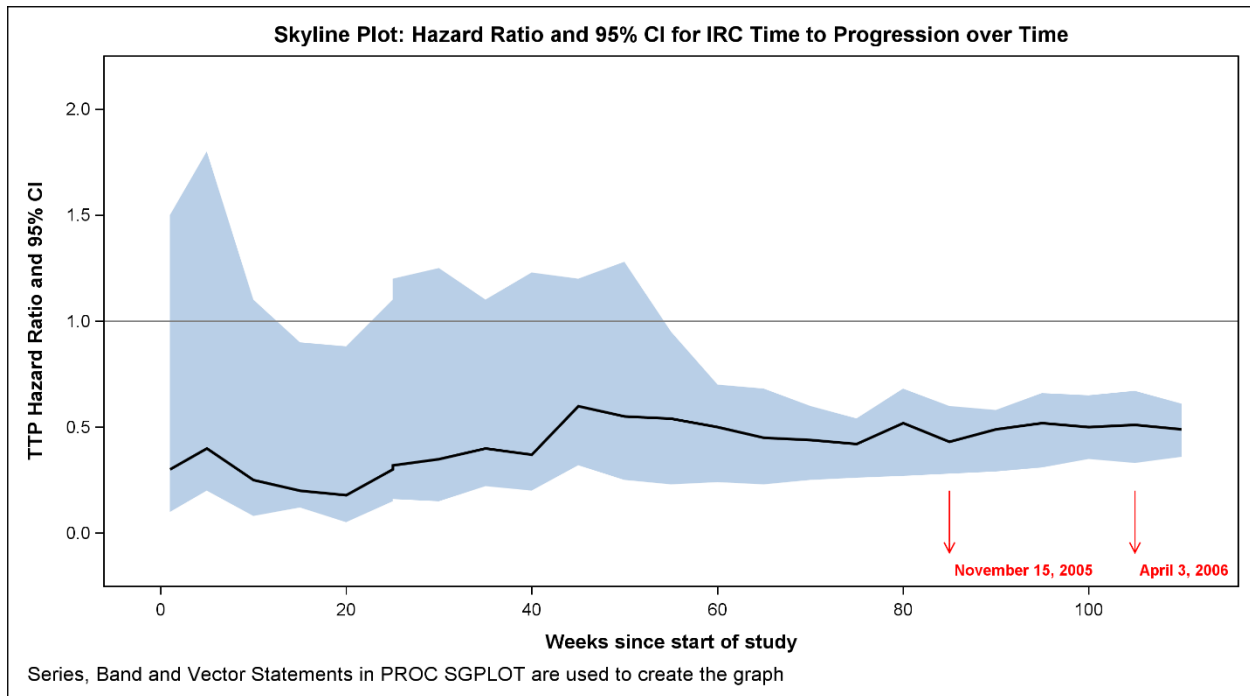
Figure 3. Skyline Plot

Highlights of the graph shown in Figure 3:

- The SGPLOT procedure is used to create the plot.
- A BAND statement is used to plot the 95% confidence intervals.
- A SERIES statement is used to plot the hazards ratio curve.
- A VECTOR statement is used to highlight the time frame.
- A REFLINE statement is used on the Y-axis to display hazard ratio=1.

CODE FOR FIGURE 3

```
title "Skyline Plot: Hazard Ratio and 95% CI for IRC Time to Progression
over Time";
footnote j=l "Series, Band and Vector Statements in PROC SGPLOT are used to
create the graph";
proc sgplot data=merged noautolegend;
      band x=week lower=low upper=high;
      series x=week y=hr / lineattrs=(thickness=2px);
      vector x=x1 y=y1   / xorigin=xorigin yorigin=yorigin datalabel=label
                           lineattrs=(color=red)
                           datalabelattrs=(color=red weight=bold size=8pt) ;
      refline 1 / axis=y;
      xaxis label="Weeks since start of study" labelattrs=(weight=bold)
offsetmin=0.05 offsetmax=0.05;
      yaxis label="TTP Hazard Ratio and 95% CI" labelattrs=(weight=bold)
values=(0 to 2 by 0.5) offsetmin=0.1 offsetmax=0.1;
run;
```

6

Figure 4 shows the relative odds that a clinical trial with the indicated development time will meet its accrual goals.

Table 9 below is the data set that is used to plot the likelihoods over time. The odds ratio and the confidence levels are contained in the numeric variables OR, HIGH, and LOW.

Obs	or	high	low	time
1	1.10	2.5	0.50	7
2	2.00	3.6	1.10	9
3	0.95	1.6	0.40	12
4	1.00	.	.	15
5	0.50	1.1	0.25	18
6	0.75	1.7	0.50	21
7	0.50	1.5	0.25	24
8	0.10	0.5	0.05	27
9	0.05	0.4	0.02	30

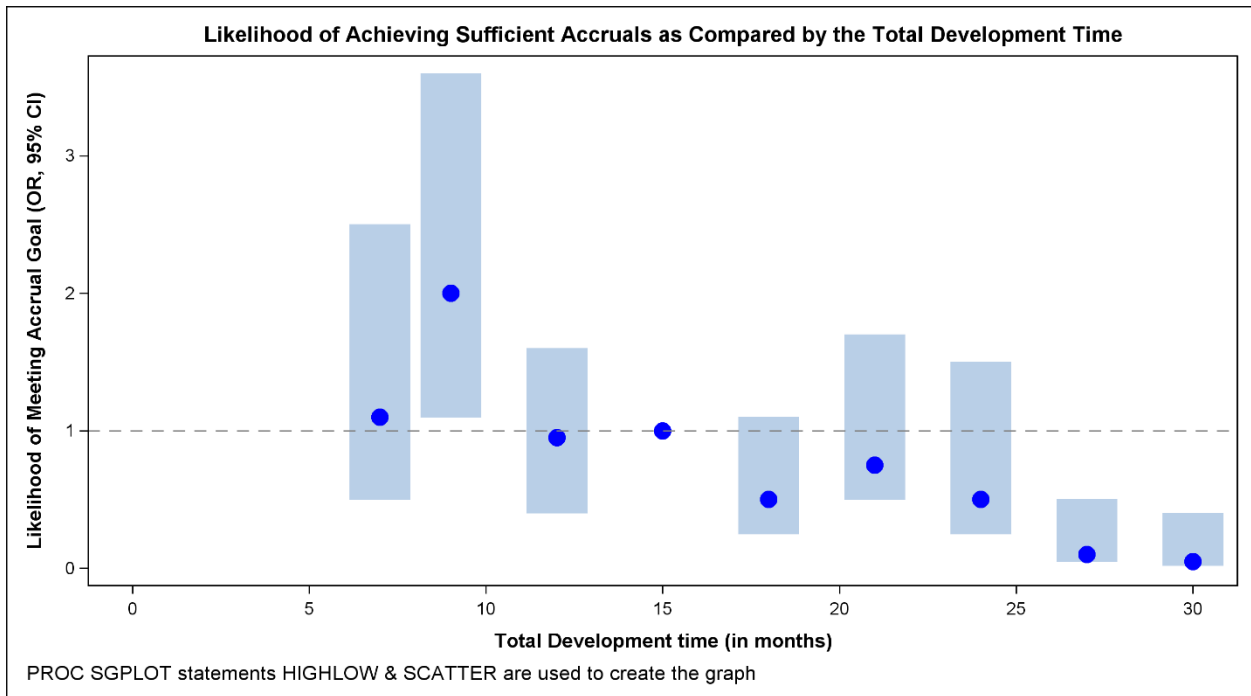**Table 4 Snippet of the Data Set Used for Figure 4**

Figure 4. Plot for Likelihood of Achieving Accruals with 95% Confidence Intervals

Highlights of the graph shown in Figure 4:

- The SGPLOT procedure is used to create the plot.
- A HIGHLOW statement is used to plot the 95% confidence intervals.

- A SCATTER statement is used to plot the odds ratio for the months.

CODE FOR FIGURE 4

```
title "Likelihood of Achieving Sufficient Accruals as Compared by the Total
Development Time";
footnote j=l "PROC SGPLOT statements HIGHLOW & SCATTER are used to create
the graph";

proc sgplot data=or noautolegend;
     highlow x=time high=high low=low / type=bar barwidth=0.05nooutline;
     scatter x=time y=or / markerattrs=(symbol=circlefilled
                                        size=12 color=blue);
     refline 1 / axis=y lineattrs=(pattern=dash );
     xaxis label="Total Development time (in months)"
               labelattrs=(weight=bold)  min=0 ;
     yaxis label="Likelihood of Meeting Accrual Goal (OR, 95% CI)"
               labelattrs=(size=8);
run;
```

Figure 5 shows the probability of survival for two different treatments.

Table 5 shows the partial data set that is used for the concentration plot in Figure 5. This data set contains 3 numeric variables.

Obs	MIDPOINT	SURVIVAL	trt
5	4.5	0.50510	1
6	5.5	0.32378	1
7	6.5	0.32378	1
8	7.5	0.25183	1
9	8.5	0.21585	1
10	9.5	0.21585	1
11	10.5	0.17988	1
12	11.5	0.10793	1
13	12.5	0.03598	1
14	13.5	0.03598	1

Table 5 Snippet of the Data Set Used for Figure 5
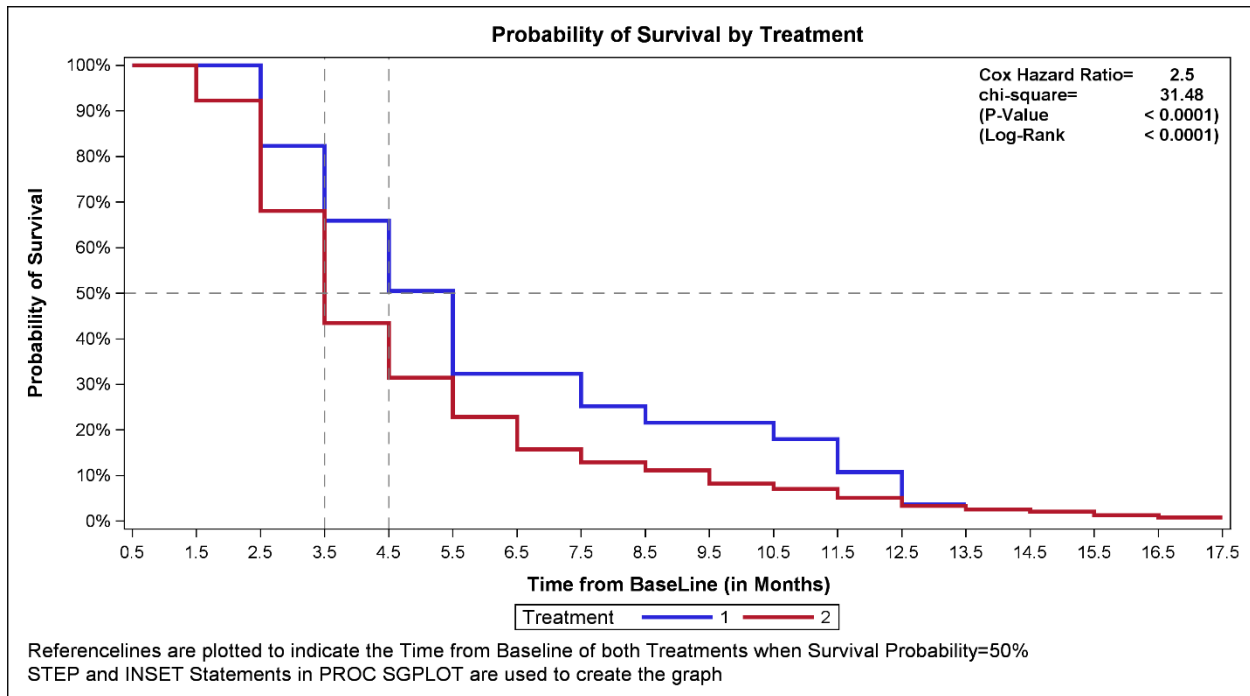
Figure 5. Survival Rates of Two Treatments

Highlights of the graph shown in Figure 5:

- The SGPLOT procedure is used to create the plot.

- A STEP statement is used to plot the survival rates. The GROUP option is used for the treatments.

- An INSET statement is used to display the statistics on the graph. These statistics are essentially label value pairs that are drawn at the top right corner of the graph.

CODE FOR FIGURE 5

```
title "Probability of Survival by Treatment";
footnote1 j=l "Referencelines are plotted to indicate the Time from Baseline
of both Treatments when Survival Probability=50%";
footnote2 j=l "STEP and INSET Statements in PROC SGPLOT are used to create
the graph";

proc sgplot data=work.survstrt dattrmap=attrmap;
     step x=MIDPOINT y=survival / group=trt name="step" attrid=myid1
                                     lineattrs=(thickness=3 pattern=solid);
     keylegend / title="Treatment"location=outside position=bottom;
     inset ("Cox Hazard Ratio="="2.5""chi-square="="31.48""(P-Value"="<
0.0001)""(Log-Rank"="< 0.0001)") /
                 valuealign=center textattrs=(weight=bold);
     refline 3.5 /axis=x lineattrs=(pattern=dash);
     refline 0.5 / axis=y lineattrs=(pattern=dash);
     refline 4.5 / axis=x lineattrs=(pattern=dash);
     xaxis label="Time from BaseLine (in Months)" type=discrete
              labelattrs=(weight=bold);
     yaxis label="Probability of Survival" values=(0 to 1 by 0.1)
           labelattrs=(weight=bold);
```

```
        format survival percent10.;
run;
```

Figure 6 shows the cancer survival rates for a group of 95 metastatic prostate cancer patients who were diagnosed between 2000 and 2009. Each patient in the group was first diagnosed at Cancer Treatment Center of America (CTCA) and/or received at least part of their initial course of treatment at CTCA. The CTCA survival rates of the cancer patients were compared with the rates reported in the Surveillance, Epidemiology, and End Results (SEER) database of the National Cancer Institute.

Table 6 below is the data set that is used for figure 6.

Obs	diff	time	survival_rate1	survival_rate2	y3	center1	center2
1	9%	0.5	0.92	0.83	1	CTCA	SEER
2	13%	1.0	0.83	0.70	1	CTCA	SEER
3	19%	1.5	0.78	0.59	1	CTCA	SEER
4	23%	2.0	0.73	0.50	1	CTCA	SEER
5	26%	2.5	0.69	0.43	1	CTCA	SEER
6	26%	3.0	0.64	0.38	1	CTCA	SEER
7	21%	3.5	0.54	0.33	1	CTCA	SEER
8	20%	4.0	0.49	0.29	1	CTCA	SEER

Table 6. Snippet of the Data Set Used for Figure 6
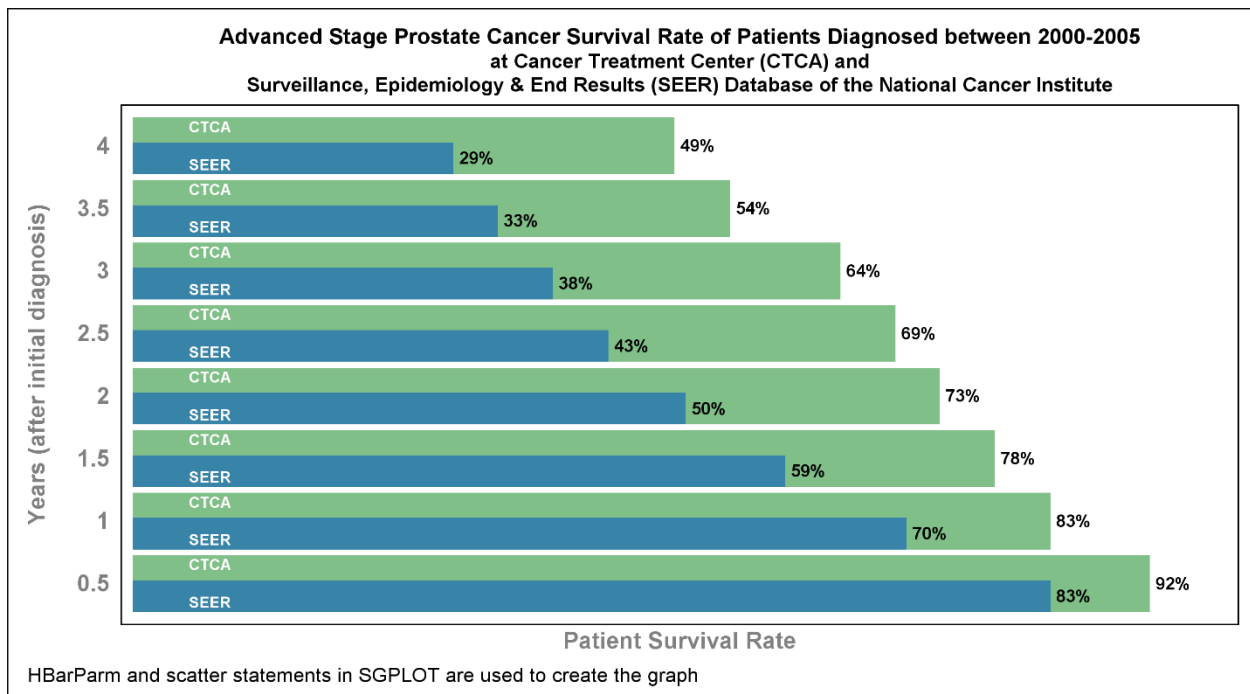

Figure 6. Survival Rates of Patients with Advanced Prostate Cancer at CTCA and SEER Databases

10

CODE FOR FIGURE 6

```
title1 "Advanced Stage Prostate Cancer Survival Rate of Patients Diagnosed
between 2000-2005";
title2 "at Cancer Treatment Center (CTCA) and";
title3 "Surveillance, Epidemiology & End Results (SEER) Database of the
National Cancer Institute";
footnote j=l "HBarParm and scatter statements in SGPLOT are used to create
the graph";

proc sgplot data=prostate3 noautolegend ;
 yaxis display=(noticks noline) valueattrs=(color=gray size=12 weight=bold)
 label="Years (after initial diagnosis)" labelattrs=(color=gray size=12
weight=bold)
 reverse;
 xaxis display=(noticks novalues noline) label="Patient Survival Rate"
labelattrs=(color=gray size=12 weight=bold)
 offsetmin=0.01 offsetmax=0 max=1;
 hbarparm category=time  response=survival_rate1/ barwidth=0.9
 nooutline fillattrs=(color=CX80BF88) datalabel datalabelattrs=(weight=bold
size=9pt);
 hbarparm category=time  response=survival_rate2 /barwidth=0.5
 nooutline fillattrs=(color=CX3883A8) discreteoffset=-0.2
 datalabel  datalabelattrs=(weight=bold size=9pt);
 scatter x=x y=time /markerchar=center1 discreteoffset=0.3
markercharattrs=(color=white size=8 weight=bold) ;
 scatter x=x y=time / markerchar=center2 discreteoffset=-0.3
markercharattrs=(color=white size=8 weight=bold) ;
 format survival_rate1 survival_rate2 percent9.;
run;
```

Highlights of the graph shown in Figure 6:

- The SGPLOT procedure is used to create the plot.

- Two parameterized horizontal bar charts (HBARPARM) are overlaid on top of each other. The BARWIDTH option is set to different values for distinguishing both bar charts.

- For the information about the cancer centers, two SCATTER statements are overlaid on the barcharts. Further, MARKERCHAR option is used to get the textual information instead of displaying the markers.

MULTI-CELL PANELED GRAPHS

The SGPANEL procedure creates a panel of graph cells based on the values of one or more classification variables. The layout of the panel can be controlled using the LAYOUT option in the PANELBY statement. We will now see some examples that can be easily created using the SGPANEL procedure in SAS 9.3.

Figure 7 shows a paneled plot for drug concentration over a 3-month period.

Table 7.1 shows the partial data set that is used for the concentration plot in Figure 7.

Obs	Day	conc	pat	Period
1	1	0.90	1	1
2	2	1.10	1	1
3	3	3.00	1	1
4	1	0.30	2	1
5	2	0.05	2	1
6	3	1.00	2	1
7	1	0.40	3	1
8	2	0.60	3	1
9	3	0.80	3	1
10	1	0.25	4	1

Table 7.1 Snippet of the Data Set Used for Figure 7

Table 7.2 shows the SG attribute map data set that is consumed by the SGPANEL procedure. The data set is used to associate data values with visual attributes. In this example, to assign different line and marker colors and different marker symbols for the patients. Each observation defines the attribute for a particular patient. An observation uses reserved variable names for the attribute map identifier (ID) and the group value (VALUE) that contains the patient-number as given in the PAT variable of Table 7.1.

Obs	value	fillcolor	markercolor	linecolor	markersymbol	linepattern	id
1	1	lightseagreen	lightseagreen	lightseagreen	asterisk	solid	myid1
2	2	lightgreen	lightgreen	lightgreen	trianglefilled	solid	myid1
3	3	orange	orange	orange	circlefilled	solid	myid1
4	4	maroon	maroon	maroon	squarefilled	solid	myid1
5	5	CX73545A	CX73545A	CX73545A	starfilled	solid	myid1
6	6	lightblue	lightblue	lightblue	diamondfilled	solid	myid1
7	7	gray	gray	gray	homedownfilled	solid	myid1

Table 7.2 SG Attribute Map Data Set Used for Figure 7
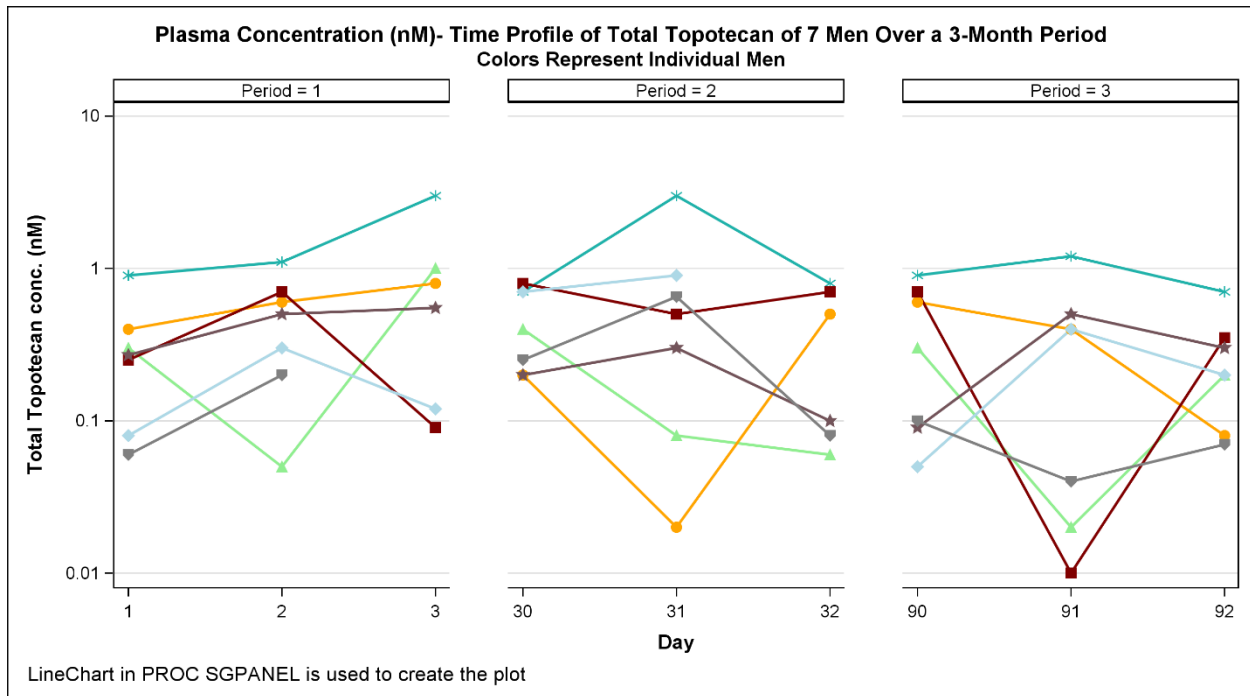
Figure 7. Plot for Drug Concentration – Time Profile of 7 Men Over a 3-Month Period

Highlights of the graph shown in Figure 7:

- The SGPANEL procedure is used to create the plot.

- An attribute map is associated with the procedure to assign custom colors for the 7 men in the study. The DATTRMAP option is used in the PROC SGPANEL statement to reference the attribute map data set that is already created.

- The time period is used as the classification variable in the procedure.

- The border around each cell is turned off using the NOBORDER option in PANELBY statement.

- A VLINE (vertical linechart) statement is used to plot the concentration.

- A logarithmic scale of base 10 is used for the concentrations on the rowaxis. This is done by using the options TYPE=LOG and LOGBASE=10 in the ROWAXIS statement.

CODE FOR FIGURE 7

```
title1 "Plasma Concentration (nM)- Time Profile of Total Topotecan of 7 Men
Over a 3-Month Period";
title2 "Colors Represent Individual Men";
footnote j=l "LineChart in PROC SGPANEL is used to create the plot";
proc sgpanel data=conc dattrmap=attrmap2 noautolegend;
      panelby period / noborder layout=columnlattice uniscale=row
                       spacing=40;
      vline day       / response=conc group=pat attrid=myid1 markers
                        lineattrs=(thickness=2) markerattrs=(size=10);
      rowaxis type=log logbase=10 max=10 label="Total Topotecan conc (nM)"
                        labelattrs=(weight=bold) grid;
      colaxis labelattrs=(weight=bold);
run;
```

13

Figure 8 shows a paneled plot for deviation from mean of drug concentration across different cancer types and cell lines and for different treatments. The graph displays the concentration that inhibits cell growth by 50% (GI50) in the 60 cell lines of the NCI Anticancer Drug Screen, representing different tumor types like leukemia, non-small cell lung cancer, colon cancer, CNS cancer, melanoma, ovarian cancer, renal cancer, prostate cancer, and breast cancer.

Table 8 shows the partial data set that is used for the concentration plot in Figure 8.

Obs	Drug	cancer	Cline	Conc
1	Topotecan	Leukemia	1	50
2	Topotecan	Leukemia	2	40
3	Topotecan	Leukemia	3	30
4	Topotecan	Leukemia	4	60
5	Topotecan	Leukemia	5	20
6	Topotecan	Leukemia	6	50
7	Topotecan	Lung	7	20
8	Topotecan	Lung	8	5
9	Topotecan	Lung	9	50
10	Topotecan	Lung	10	20

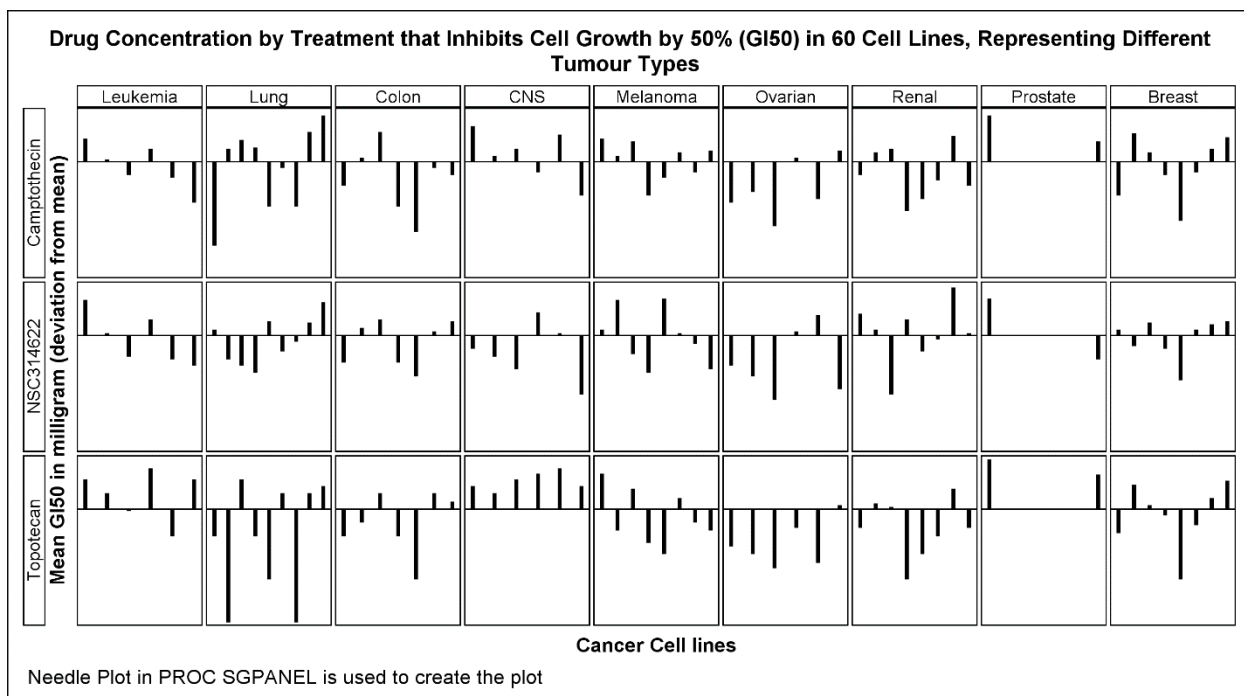**Table 8 Snippet of the Data Set Used for Figure 8**

Figure 8. Plot for Deviation of Drug Concentration from Mean by Different Cancer Types and Treatment

Highlights of the graph shown in Figure 8:

- The SGPANEL procedure is used to create the plot.
- The variables DRUG and CANCER are used as the classification variable in the procedure.
- LAYOUT=LATTICE is used in the PANELBY statement of the procedure.
- A NEEDLE statement is used to plot the concentration with the baseline value being the mean concentration.
- The display for ticks and tick values is turned off for the row and column axes.

CODE FOR FIGURE 8

```
title "Drug Concentration by Treatment that Inhibits Cell Growth by 50%
(GI50) in 60 Cell Lines, Representing Different Tumor Types";
footnote j=l "Needle Plot in PROC SGPANEL is used to create the plot";

proc sgpanel data=conc2;
 panelby  cancer drug / layout=lattice uniscale=row onepanel novarname
                        colheaderpos=top rowheaderpos=left spacing=2;
 needle x=cline y=conc / baseline=&mean lineattrs=(thickness=3) ;
 colaxis display=(novalues noticks) label="Cancer Cell lines"
                        labelattrs=(weight=bold);
 rowaxis display=(novalues noticks) label="Mean GI50 in milligram (deviation
                                     from mean)"
                        labelattrs=(weight=bold);
 run;
```

GRAPH TEMPLATE LANGUAGE

The Graph Template Language (GTL), on the other hand, offers a wide range of plot statements, layouts, and other options. In the next set of examples, we will see graphs that are relatively complex and may not be accomplished using the Statistical Graphics procedures. Hence, we will use GTL to create them.

Creating a graph using GTL is a two-step process.

The first step is to define a statgraph template that describes the structure and appearance of a graph to be produced. Here is a simple example of a statgraph template that is used for creating a scatter plot:

```
proc template;
  define statgraph scatter;
    begingraph;
     entrytitle "Scatter Plot of Weight v/s Height";
     layoutoverlay;
       scatterplot x=height y=weight;
     endlayout;
    endgraph;
  end;
run;
```

The second step is to use the SGRENDER procedure to associate the template and the data object to create the graph.

```
proc sgrender data=sashelp.class template=scatter; run;
```

Various layout, plot, text, legend, and other statgraph statements can be combined to create the template for a graph.

Figure 9 shows patient history and response to treatment over a 20-month period. The primary Y-Axis (Y) displays the baseline circulating tumor cells (CTC) count and the secondary Y-axis (Y2) displays the baseline carcinoembryonic

antigen (CEA). The blocks at the top of the graph represent the response criteria, and the blocks at the bottom is the timeline for chemotherapy regimen.

Table 9 shows the partial data set used to create the graph.

Obs	date	therapy	ctc	cea	blk
1	JAN07	Chemotherapy Regimen #1	7	12	Imaging
2	FEB07	Chemotherapy Regimen #1	2	10	Imaging
3	MAR07	Chemotherapy Regimen #1	.	.	Imaging
4	APR07	Chemotherapy Regimen #1	.	.	Imaging
5	MAY07	Chemotherapy Regimen #1	0	.	PR
6	JUN07	Chemotherapy Regimen #1	.	.	PR
7	JUL07	Chemotherapy Regimen #1	.	.	SD
8	AUG07	Chemotherapy Regimen #2	0	.	SD
9	SEP07	Chemotherapy Regimen #2	0	.	SD
10	OCT07	Chemotherapy Regimen #2	0	.	SD

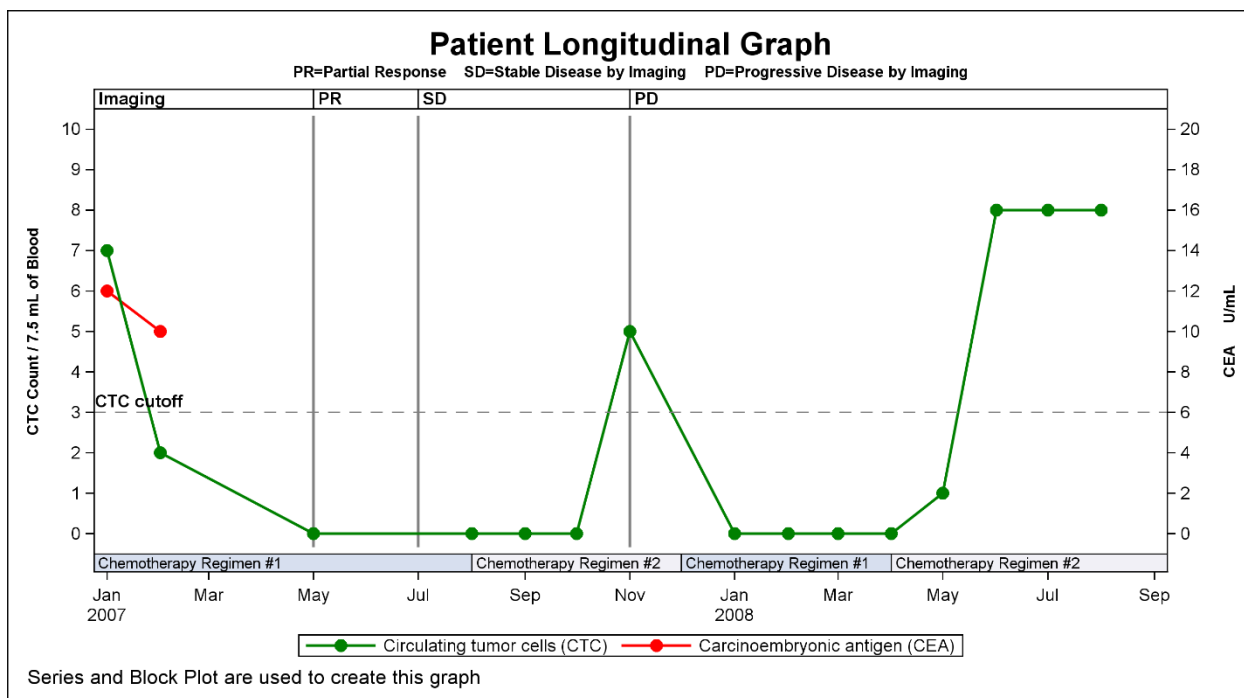**Table 9. Snippet of the Data Set Used for Figure 9**

Figure 9. Patient History and Response to Treatment

TEMPLATE CODE FOR FIGURE 9

In order to create the patient longitudinal graph in GTL, we will superimpose various elements of the graph using the

16

layout overlay. Here is a snippet of the program:

```
proc template;
   define  statgraph patlong;
     begingraph;

...

       layout overlay /
                       yaxisopts=(label="CTC Count / 7.5 mL of Blood"
                       labelattrs=(weight=bold size=8)
                       linearopts=(viewmax=10 viewmin=0
                       tickvaluesequence=(start=0 end=10 increment=1)))
                       y2axisopts=(label="CEA      U/mL"
                       labelattrs=(weight=bold size=8)
                       linearopts=(viewmax=20 viewmin=0
                       tickvaluesequence=(start=0 end=20 increment=2)))

                       xaxisopts=(tickstyle=outside display=(ticks
tickvalues));

           innermargin / align=top;
              blockplot x=date block=blk /
                       filltype=alternate fillattrs=(color=lightgray)
                       altfillattrs=(color=lightgray)
                       valueattrs=(weight=bold)
                       display=(values outline);
           endinnermargin;

           innermargin / align=bottom;
              blockplot x=date block=chem /
                       filltype=alternate
                       valueattrs=(weight=normal size=8)
                       display=(values fill outline);
           endinnermargin;

...
```

To display the information about the response criteria and timeframes for chemotherapy regimen, two innermargin blocks are used within the overlay layout. The first innermargin is positioned at the top of the graph and contains a blockplot for the response criteria. Similarly, the second innermargin block is positioned at the bottom and contains another blockplot for the chemotherapy regimen.

Now let us look at the how the curves were constructed within the same template.

```
...
           seriesplot x=date y=ctc /
                       lineattrs=(thickness=2 color=green)
                       display=all
                       markerattrs=(size=9 color=green symbol=circlefilled)
                       name="ser1" legendlabel="Circulating tumor cells
(CTC)";

           seriesplot x=date y=cea /
                       lineattrs=(thickness=2 color=red)
                       display=all
```

```
                        markerattrs=(size=9 color=red symbol=circlefilled)
                        yaxis=y2name="ser2"
                        legendlabel="Carcinoembryonic antigen (CEA)";


            discretelegend "ser1""ser2" /
                        location=outside border=true;

            referenceline y=3 / curvelabel="CTC cutoff"
                        lineattrs=(pattern=dash color=gray)
                        curvelabelattrs=(weight=bold)
                        curvelabelposition=min curvelabellocation=inside;

            referenceline x='01MAY07'd / lineattrs=(thickness=2) ;

            referenceline x='01JUL07'd /lineattrs=(thickness=2) ;

            referenceline x='01NOV07'd /lineattrs=(thickness=2 ) ;
        endlayout;
      endgraph;
    end;
run;
```

Two seriesplot statements were overlaid with the variables CTC and CEA being on the primary Y-axis and the secondary Y-axis (Y2) respectively. A discretelegend with the LOCATION option is used to get the legend outside the graph. Also note that the axis options are used at the layout overlay level to control the appearance of the axes. Reference lines are plotted on the axes appropriately to display the cut-offs and the response criteria.

After the template is constructed and compiled, the next step is to render the graph using the SGRENDER procedure.

```
proc sgrender data=patlong template=patlong; run;
```

Summary:

- An OVERLAY layout is used as the parent layout for the graph.

- Two individual SERIESPLOT statements are used to plot the CTC and CEA counts. The SERIESPLOT for CTC uses the primary Y-axis while CEA count is plotted on the secondary Y-axis.

- To display the information about the Response Criteria, a BLOCK plot is used within the INNERMARGIN block that is positioned at the top.

- Similarly, a BLOCK plot is used at the bottom for Chemotherapy Regimen timelines.

- Reference lines are plotted on the X and primary Y axes for response criteria and CTC cut-off respectively.

Figure 10 shows the distribution of p-value for testing the hypotheses whether the new drug helps in reducing the mortality rates. The study has a total of 2200 patients out of which 700 patients receive the old drug while 1500 patients are administered the new drug.

Table 10 shows the partial data set for Figure 10.

Obs	density1	pvalue1	density2	pvalue2	density3	pvalue3	group
1	10	0.0001	99.7771	0.0001	99.6440	0.0001	pval=0.05
2	10	0.0101	79.8217	0.0101	69.7508	0.0101	pval=0.05
3	10	0.0201	63.8573	0.0201	48.8255	0.0201	pval=0.05
4	10	0.0301	51.0859	0.0301	34.1779	0.0301	pval=0.05
5	10	0.0401	40.8687	0.0401	23.9245	0.0401	pval=0.05
6	10	0.0501	32.6950	0.0501	16.7472	0.0501	pval=0.2
7	10	0.0601	26.1560	0.0601	11.7230	0.0601	pval=0.2
8	10	0.0701	20.9248	0.0701	8.2061	0.0701	pval=0.2
9	10	0.0801	16.7398	0.0801	5.7443	0.0801	pval=0.2
10	10	0.0901	13.3919	0.0901	4.0210	0.0901	pval=0.2

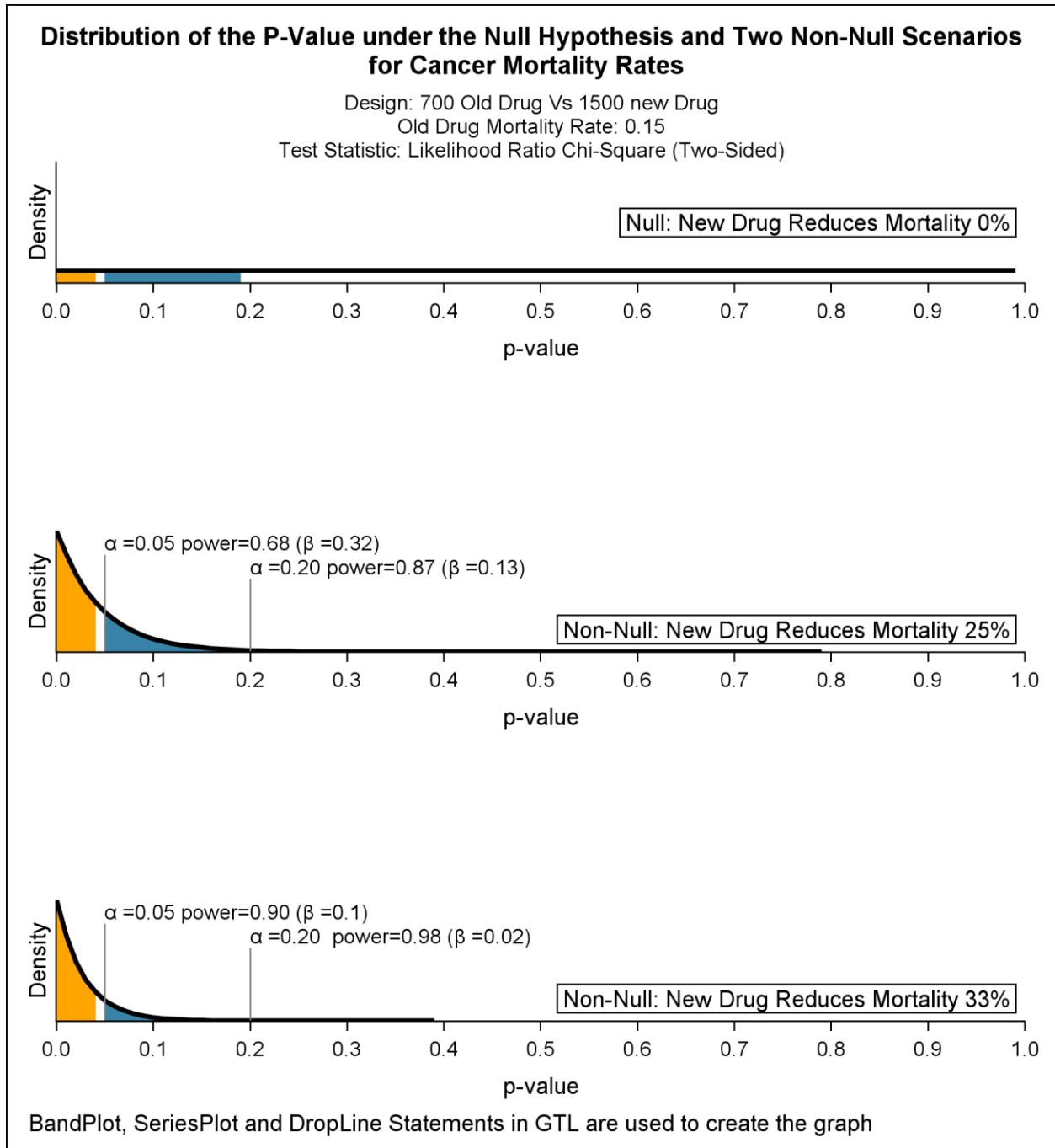**Table 10. Snippet of the Data Set Used for Figure 10**

Figure 10. Distribution of P-Value for Testing the Reduction in Mortality Rates

TEMPLATE CODE FOR FIGURE 10

We will use the gridded layout to create the template with one column. The entry statements are specified for the text that needs to appear on the graph.

```
proc template;
...
        layout gridded;
```

```
            entry "Design: 700 Old Drug Vs 1500 new Drug";
            entry "Old Drug Mortality Rate: 0.15";
            entry "Test Statistic: Likelihood Ratio Chi-Square (Two-Sided)";
...
```

The second part of the graph is the plot for null hypothesis. bandplot with GROUP option and seriesplot statements are overlaid within a layout overlay block to achieve the desired output.

```
...
        layout overlay / ...;

            entry halign=right textattrs=(size=10) "Null: New Drug Reduces
                        Mortality  0%" / border=true ;
          bandplot x=pvalue1 limitlower=0 limitupper=density1/ group=gpvar
... ;
          seriesplot x=pvalue1 y=density1 / ... ;

       endlayout;
...
```

The third and fourth part of the graph contain the plots for non-null hypotheses. Similar to the second part, a bandplot statement with GROUP option and seriesplot statement are overlaid within layout overlay blocks. A dropline statement is additionally used in third and fourth part to draw the reference lines. The LABEL option with Unicode values is used in the dropline statement to display the type I error (alpha) and type II (beta) values of the test. Also note that empty cells are created (with the help of entry statements that do not contain any text) to allow some space between the plots.

```
...

       layout overlay;
         entry "";
       endlayout;

       layout overlay / ...;
          entry ...;
         bandplot x=pvalue2 limitlower=0 limitupper=density2 /  ... ;
         dropline x=0.05 y=80 / label="(*ESC*){unicode alpha} =0.05
                               power=0.68
                               ((*ESC*){unicode beta} =0.32) ";
          dropline x=0.2 y=60 / label="(*ESC*){unicode alpha} =0.20
                               power=0.87
                               ((*ESC*){unicode beta} =0.13) ";
       endlayout;

       layout overlay;
         entry "";
       endlayout;

       layout overlay / ...;
          entry ... ;
         bandplot x=pvalue3 limitlower=0 limitupper=density3 / ...;

        dropline x=0.05 y=80 /  label="(*ESC*){unicode alpha} =0.05
                               power=0.90
                               ((*ESC*){unicode beta} =0.1) ";
```

```
        dropline x=0.2 y=60 /    label="(*ESC*){unicode alpha} =0.20
                                 power=0.98
                                 ((*ESC*){unicode beta} =0.02) ";
     endlayout;

...
```

Summary:

- GRIDDED is used as the parent layout of the graph within which blocks of OVERLAY are used that display individual plots for different hypotheses.

- ENTRY statements are used for the text that appear on the graph area.

- SERIESPLOT statements are used for plotting the curve.

- BANDPLOT statements are used for displaying the filled regions.

- DROPLINE statements are used for referencing the regions.

- Unicode values are used as part of the dropline labels for the symbols and notations.

Figure 11 shows the survival rates of different cancer types. These rates are derived for both sexes all different ethnic groups.

Table 11 shows the partial data set used for Figure 11 for 10 different cancer types.

Obs	time	cancer1	cancer2	cancer_type1	cancer_type2
1	5 Year	99	96	Prostate	Thyroid
2	10 Year	95	96	Prostate	Thyroid
3	15 Year	87	94	Prostate	Thyroid
4	20 Year	81	95	Prostate	Thyroid

Table 11. Snippet of the Data Set Used for Figure 11

Relative Survival Rate Gradients (in %) for Different Cancer Types

| 5 Year | 10 Year | 15 Year | 20 Year |

Prostate
99   95   87   81

Thyroid
96   96   94   95

Testis
95   95   91   88

Melanomas
89   87   84   83

Breast
86   78   71   65

Stomach
24   19   19   15

Lung and bronchus
15   11   8   6

Esophagus
14   8   8   5

Liver
8   6   6   8

Pancreas
4   3   3   3

SeriesPlot and BlockPlot in GTL are used to create this plot
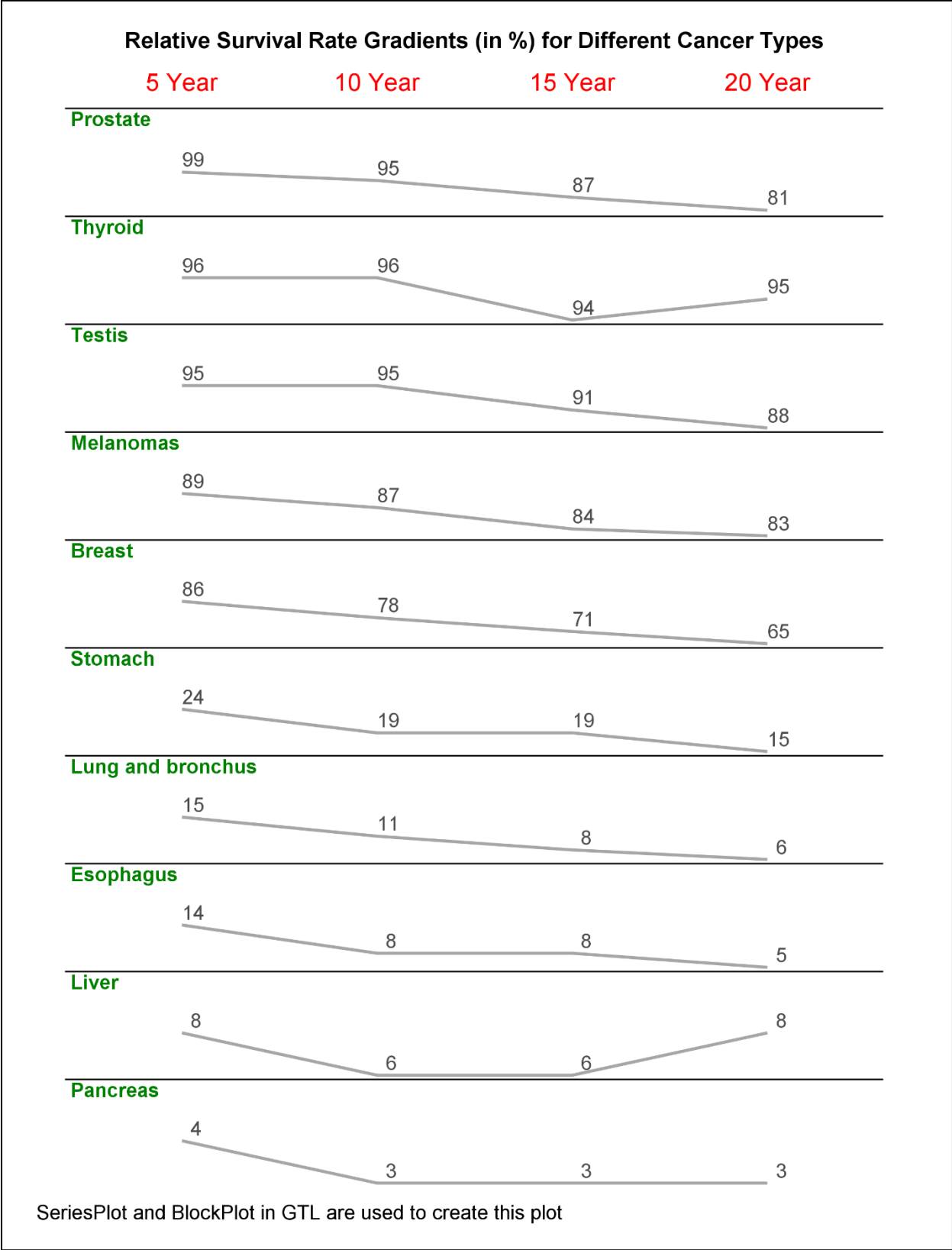
Figure 11. Relative Survival Rates of Patients with Different Cancer Types

TEMPLATE CODE FOR FIGURE 11

We will use the overlay layout to create the template. Each of the survival rates for the 10 cancer types will be plotted in a different cell. We will run a do loop to create these cells with the layout. We will also externalize the axes with the help of UNIONALL option. Below is a snippet:

```
proc template;

  ...

   layout lattice / columns=1 column2datarange=unionall ;

...
```

The next step is to define the axes for the plot. We will use a secondary columnaxes block within the layout. The display for the tickvalue and the line are turned on.

```
...

column2axes;

    %do i=1 %to &n;
     columnaxis / display=(tickvalues line) ... ;
    %end;

endcolumn2axes;

...
```

The contents of each cell will use a seriesplot statement to plot the curve. We will also use a blockplot within an innermargin block for plotting the names of the cancer types. Both seriesplot and blockplot use the secondary x-axis and will be overlaid within a nested overlay layout.

```
...

cell;

   layout overlay / ...;

      innermargin ... ;
       blockplot x=time block=cancer_type&i/ xaxis=x2   ... ;
      endinnermargin;

      seriesplot x=time y=cancer&i / xaxis=x2  ... ;

   endlayout;

endcell;

...
```

Now, we need to create all the cells in the same fashion for different cancer types. Hence, we will run the do loop to achieve this.

```
...

%do i=1 %to &n;

   cell;
    layout overlay ... ;

        innermargin;
          blockplot  ... ;
        endinnermargin;

        seriesplot  ...;
      endlayout;
   endcell;

%end;

...
```

Summary:

- LATTICE is used as the parent layout to divide the survival rates of each cancer type into each cell in the arrangement.
- A DO loop within the TEMPLATE procedure is run to create cells for each of these survival rates.
- The secondary X-axis of the plot is externalized.
- The Y-axis display is suppressed.
- A SERIESPLOT statement is used for plot the survival rates.
- A BLOCKPLOT statement is used to display the cancer type.

Figure 12 shows the correlation of gene expression and drug activity patterns in numerous human cancer cell types and the clustering of genes on the basis of their correlations with drugs. The original data set has about 1376 genes. For simplicity, we have taken a subset of 30 genes.

Table 12.1 shows the partial data set used for Figure 12. This data set contains the variables DRUG, GENE, and CORR. This part of the data set is consumed by the heat map. The correlations have been obtained using PROC CORR.

Table 12.2 shows the partial data set used for Figure 12. This data set is used for the dendrogram and is obtained from PROC CLUSTER.

Both the data sets mentioned above are merged into one single data set and is then consumed by the graph.

Obs	DRUG	GENE	corr
60	d1	Protein: mdr1,mrp,p-glyco	0.107
61	d1	SID 310849, Human mRNA fo	-0.084
62	d1	Human brain mRNA for phot	-0.089
63	d1	Protein: mdr-1, mrp, p-gl	0.051
64	d1	SID W 28573, MULTIDRUG RE	0.005
65	d1	ESTs Chr.10 [278375, (D),	0.124
66	d1	ESTs Chr.14 [307115, (IW)	-0.199
67	d1	Human ATPase, DNA-binding	-0.090
68	d1	GALC Galactocerebrosidase	-0.357
69	d1	SID W 487092, INTERFERON	-0.142
70	d1	INTERFERON-INDUCIBLE PRO	-0.187

Table 12.1 Snippet of the Data Set Used for Figure 12

Obs	_NAME_	_PARENT_	_HEIGHT_
40	29	CL10	0.00000
41	9	CL9	0.00000
42	CL15	CL9	0.64390
43	CL10	CL8	0.68902
44	28	CL8	0.00000
45	CL14	CL7	0.66225
46	CL13	CL7	0.66705
47	CL16	CL6	0.64261
48	CL9	CL6	0.72827
49	1	CL5	0.00000
50	CL12	CL5	0.67828

Table 12.2 Snippet of the Data Set Used for Figure 12
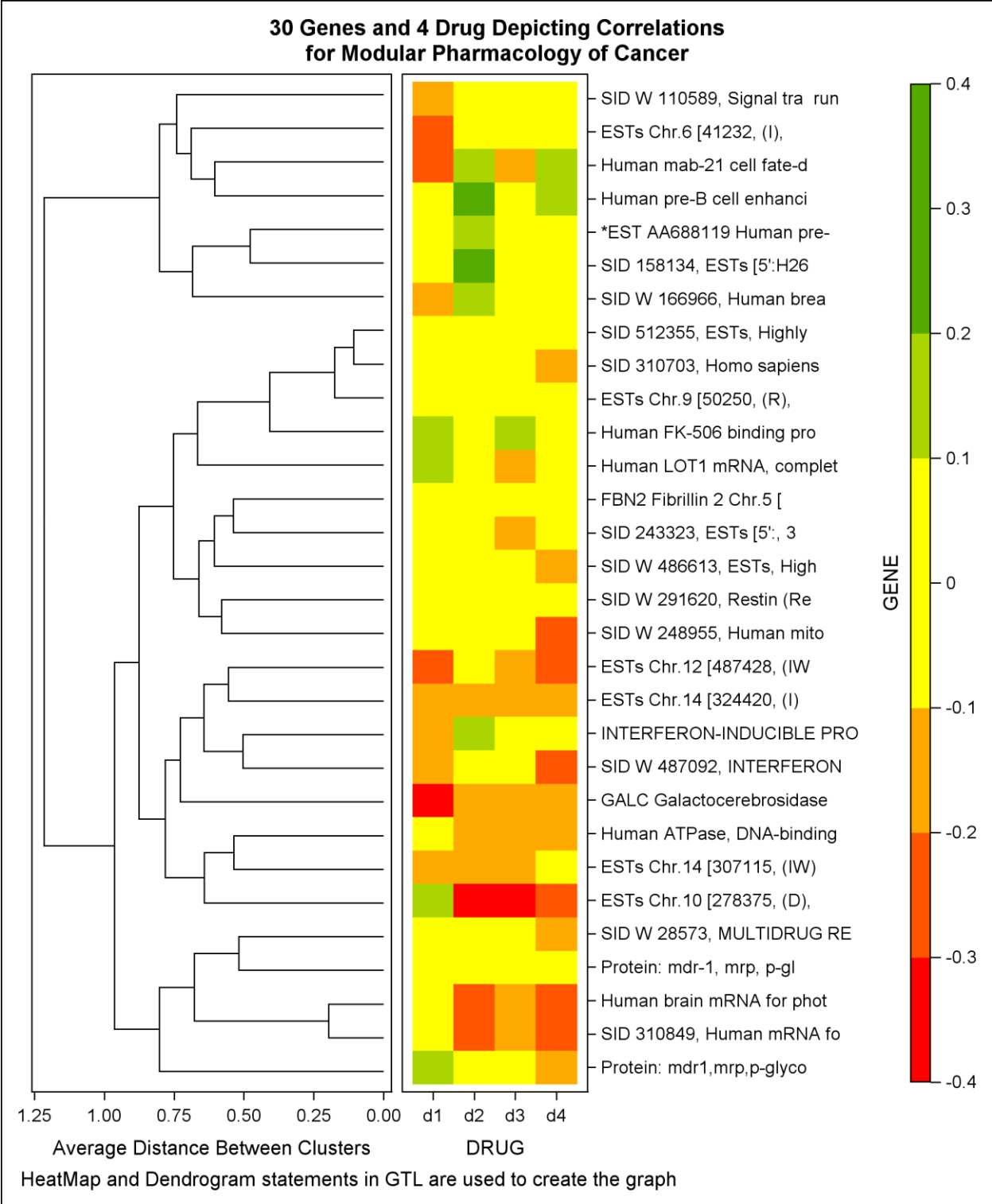
Figure 12. 30 Genes and 4 Drug Showing Correlations

TEMPLATE CODE FOR FIGURE 12

We will use the lattice layout and a nested overlay layout to create the template.

```
proc template;

...
      layout lattice /  columns=2columnweights=(0.40.6);

            layout overlay / ... ;

                  dendrogram nodeid=_name_ parentid=_parent_
                  clusterheight=_height_ /
                                    name="dendrogram"
                                    orient=horizontal;
            endlayout;

            layout overlay   ;

                  heatmapparm x=drug y=GENE   colorresponse=foo /
                                        name="heat1" yaxis=y2  ;

                  continuouslegend "heat1" / location=outside  ;

            endlayout;

      endlayout;
...
```

We have specified the COLUMNS and COLUMNWEIGHTS option in the lattice layout in order to achieve the required arrangement. The first column in the graph contains a dendrogram statement and the second columns contains a parameterized heatmapparm statement. Both the statements are included in a nested overlay layout. The ORIENT option is used in dendrogram to generate horizontal plot. Note that a continuouslegend is associated with the heat map to display the correlations.

In order to customize the colors of the heat map, a range attribute map is also used before the parent layout level.

Here is the snippet of the range attribute map.

```
proc template;

...

      rangeattrmap name="foo";

         range -0.4 -< -0.3 / rangecolor=gradientstepper(red,yellow,4,1);
         range -0.3 <- -0.2 / rangecolor=gradientstepper(red,yellow,4,2);
         range -0.2 <- -0.1 / rangecolor=gradientstepper(red,yellow,4,3);
         range -0.1 <- 0    / rangecolor=gradientstepper(red,yellow,4,4);
         range 0 <- 0.1     / rangecolor=gradientstepper(yellow,green,4,1);
         range 0.1 <- 0.2   / rangecolor=gradientstepper(yellow,green,4,2);
         range 0.2 <- 0.3   / rangecolor=gradientstepper(yellow,green,4,3);
         range 0.3 <- 0.4   / rangecolor=gradientstepper(yellow,green,4,3);

      endrangeattrmap;
```

```
        rangeattrvar attrvar=foo var=corr attrmap="foo";

        layout  lattice  / ... ;

        ...

        endlayout ;

...
run;
```

The gradient colors are specified for each of the ranges in the attribute map. The attribute variable created in the rangeattrvar statement is specified in the colorresponse argument of the heat map.

Summary:

- LATTICE is the parent layout with COLUMNS=2. Additionally, the COLUMNWEIGHTS option is specified to assign applicable weights to both the columns.

- A parameterized HEATMAPPARM statement is used for creating the correlation heat map of drug and genes.

- The colors of the heat map are mapped to the correlations.

- A CONTINUOUSLEGEND statement is used to display the calibration of the correlation in the heat map.

- A Range attribute map is used to assign the colors of the different ranges in the continuouslegend.

- A DENDROGRAMstatement is used for displaying the clusters.

CONCLUSION

Cancer is a global issue that affects many lives across geography and all demographics. Visualization of cancer data, in the context of clinical trials or otherwise, is crucial in order to understand the disease and the potential treatment. Graphs related to various types of cancer are used to create reports that are useful in oncology trials. GTL and ODS Statistical Graphics procedures provide the essential tools needed to create quality graphs. In this presentation, we have used SAS 9.3 SGPLOT, SGPANEL, and GTL to illustrate some examples in cancer studies. While the ODS Statistical Graphics procedures are relatively simple to use, GTL offers extensive customization and can help create complex graphs. Together, they are a powerful tool for data visualization.

REFERENCES

- *Past PharmaSUG Conference Proceedings*. Available at http://www.pharmasug.org/proceedings.html.

- Clinical Trial Reporting using SG Procedures. Available at http://support.sas.com/rnd/datavisualization/papers/sgf2009/174-2009.pdf.

- Bushnell, Will. 2007. "Understanding Clinical Trial Data Through Use of Statistical Graphics." *ASA section on Statistical Computing Statistical Graphics*. GlaxoSmithKline. Available at http://stat-computing.org/events/2007-jsm/bushnell.pdf.

- Dmitrienko, Alex, Christy Chuang-Stein, and Ralph D'Agostino. 2007. *Pharmaceutical Statistics Using SAS A Practical Guide*, 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/rnd/datavisualization/papers/sgf2009/174-2009.pdf.

ACKNOWLEDGMENTS

RECOMMENDED READING

- *SAS® 9.3 Graph Template Language: User's Guide. Cary, NC.*

- *SAS® 9.3 ODS Graphics: Procedures Guide, Cary, NC.*

CONTACT INFORMATION

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

Debpriya Sarkar
SAS Institute Inc.
Level 1, Cybercity, Tower 5, Magarpattta City, Hadapsar
Pune, Maharashtra, India -411013
debpriya.sarkar@sas.com