PharmaSUG 2017 - Paper DV08

Bird's eye view of the data, a graphical exploration!

Hrideep Antony, inVentiv Health, North Carolina, USA

ABSTRACT

Have you ever wondered if there is a better way to know who those outlier subjects are on your summary graphs without going through individual listings or without going back to your database filters to recognize them? Have you wished for a more efficacious means of identifying the outliers in your database and evaluating an individual subject's contribution to the overall database without the repetitious task of comparing numerous pages of patient listing outputs vs. summary tables? If you did, continue reading, as this paper may provide you with an optimal solution that you are looking for!

The number of listings and tables that are created for the purpose of understanding a subject's contribution or significance to an overall database could be reduced if we can "recognize" a subject or a group of subjects of interest on a broader summary picture. This paper will introduce some practical techniques using the SAS® SG procedures to have a broader understanding of the study databases and ways by which we can "spot" and "evaluate" the subjects of interest by having a "birds-eye view" (recognizing and analyzing from an elevated angle) of the subjects on high-quality summary graphics.

The primary focus here will be to generate summary graphs which recognize individual subjects of interest along with overall comparison statistics that are commonly used to analyze data in a clinical trial setting.

INTRODUCTION

Interpretation of individual subject's contribution is usually done by comparing the Patient Profile Listings and other Listings with the overall summary tables and analyzing the subject of interest based on that understanding.

The main disadvantage of using this technique is that we need multiple summary outputs and several cross comparison efforts even to have at least a minimal understanding of the subjects. Despite this effort, we may still fail to recognize other subjects with a similar profile of interest.

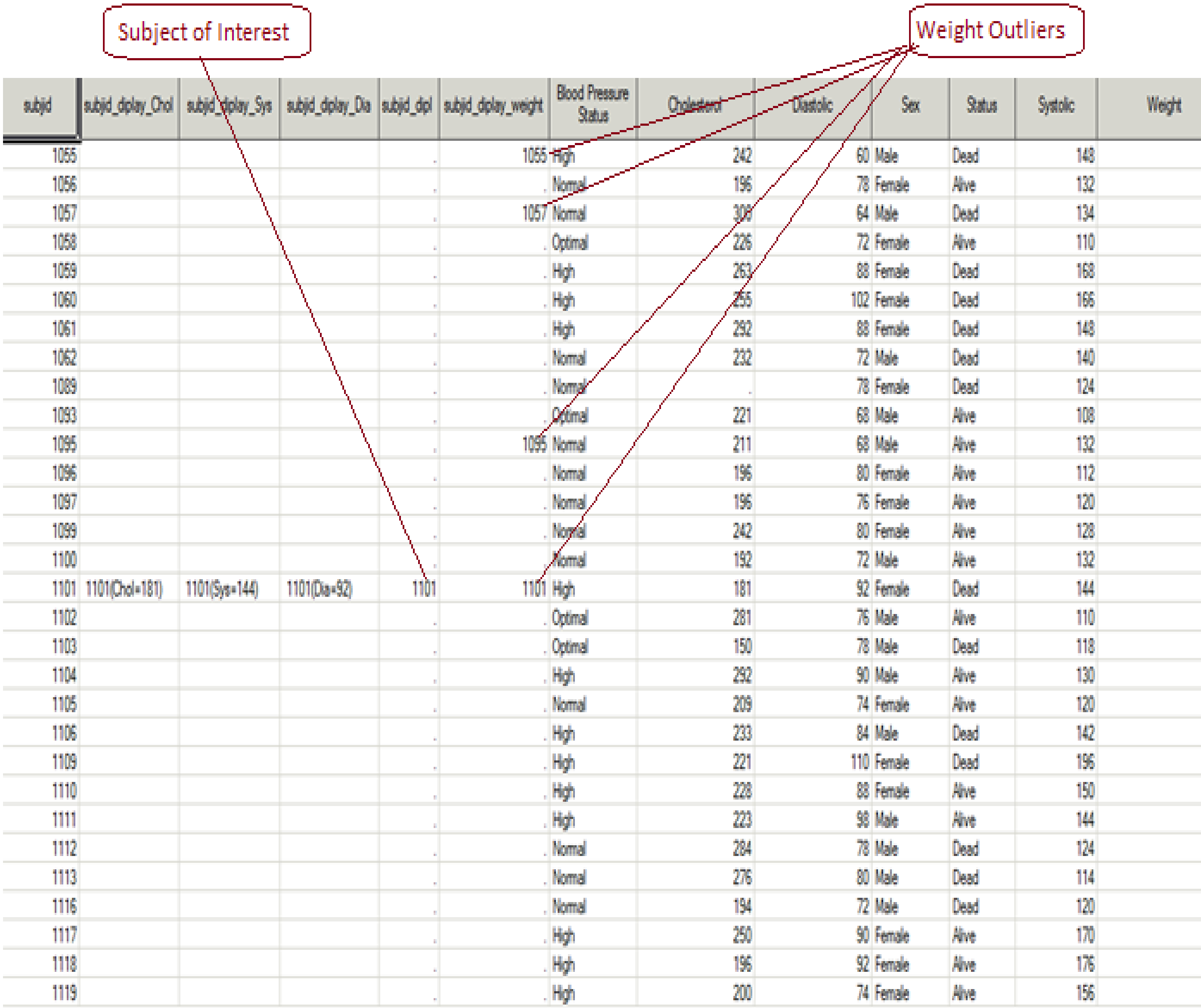
This paper uses procedures such as SGPLOT, SGSCATTER, and SGPANEL to generate multiple summary graphics, and introduces techniques to provide an elevated view of the subject's individual contribution to the overall database summary. This will also help us to understand the trends of the outliers. For example, if a subject of interest has an elevated weight, we could easily interpret this subject along with other subjects with similar weight profiles and understand the general impact of elevated weight on other vital parameters as shown in the examples below.

This publication further discusses four types of summary plots, along with the corresponding SAS® Code used to generate them, which can be further enhanced as per the specific needs of the programmers.

DATA PREPARATION

Table 1 shows the input data format that has been used in the summary plots.

Note that this data has maximum post-baseline values for each of the subjects, and has one record per subject with each of the corresponding maximum post-baseline vital parameters. Notice that there are few derived variables with Subjid_ prefix that are only populated for weight outliers and for subjects of interest (subject 1101, in this case is an example).



subjid	subjid_diplay_Chol	subjid_diplay_Sys	subjid_diplay_Dia	subjid_dipl	subjid_diplay_weight	Blood Pressure Status	Cholesterol	Diastolic	Sex	Status	Systolic	Weight
1055				.	1055	High	242	60	Male	Dead	148	
1056				.	.	Normal	196	78	Female	Alive	132	
1057				.	1057	Normal	308	64	Male	Dead	134	
1058				.	.	Optimal	226	72	Female	Alive	110	
1059				.	.	High	263	88	Female	Dead	168	
1060				.	.	High	255	102	Female	Dead	166	
1061				.	.	High	292	88	Female	Dead	148	
1062				.	.	Normal	232	72	Male	Dead	140	
1089				.	.	Normal	.	78	Female	Dead	124	
1093				.	.	Optimal	221	68	Male	Alive	108	
1095				.	1095	Normal	211	68	Male	Alive	132	
1096				.	.	Normal	196	80	Female	Alive	112	
1097				.	.	Normal	196	76	Female	Alive	120	
1099				.	.	Normal	242	80	Female	Alive	128	
1100				.	.	Normal	192	72	Male	Alive	132	
1101	1101(Chol=181)	1101(Sys=144)	1101(Dia=92)	1101	1101	High	181	92	Female	Dead	144	
1102				.	.	Optimal	281	76	Male	Alive	110	
1103				.	.	Optimal	150	78	Male	Dead	118	
1104				.	.	High	292	90	Male	Alive	130	
1105				.	.	Normal	209	74	Female	Alive	120	
1106				.	.	High	233	84	Male	Dead	142	
1109				.	.	High	221	110	Female	Dead	196	
1110				.	.	High	228	88	Female	Alive	150	
1111				.	.	High	223	98	Male	Alive	144	
1112				.	.	Normal	284	78	Male	Dead	124	
1113				.	.	Normal	276	80	Male	Dead	114	
1116				.	.	Normal	194	72	Male	Dead	120	
1117				.	.	High	250	90	Female	Alive	170	
1118				.	.	High	196	92	Female	Alive	176	
1119				.	.	High	200	74	Female	Alive	156	

Table 1, input data format

TYPE 1: ANALYZING MULTIPLE VITAL PARAMETERS ON THE SAME PLOT WITH REGRESSION LINES AND 95% CI.

Notice that Figure 1 below provides a correlation of weight vs. Systolic, Diastolic and Cholesterol values. It also displays the result values and indicates where the subject 1101 (who is our subject of interest) lies in this correlation plot, hence giving a broader view of subject's results in comparison with the overall data.

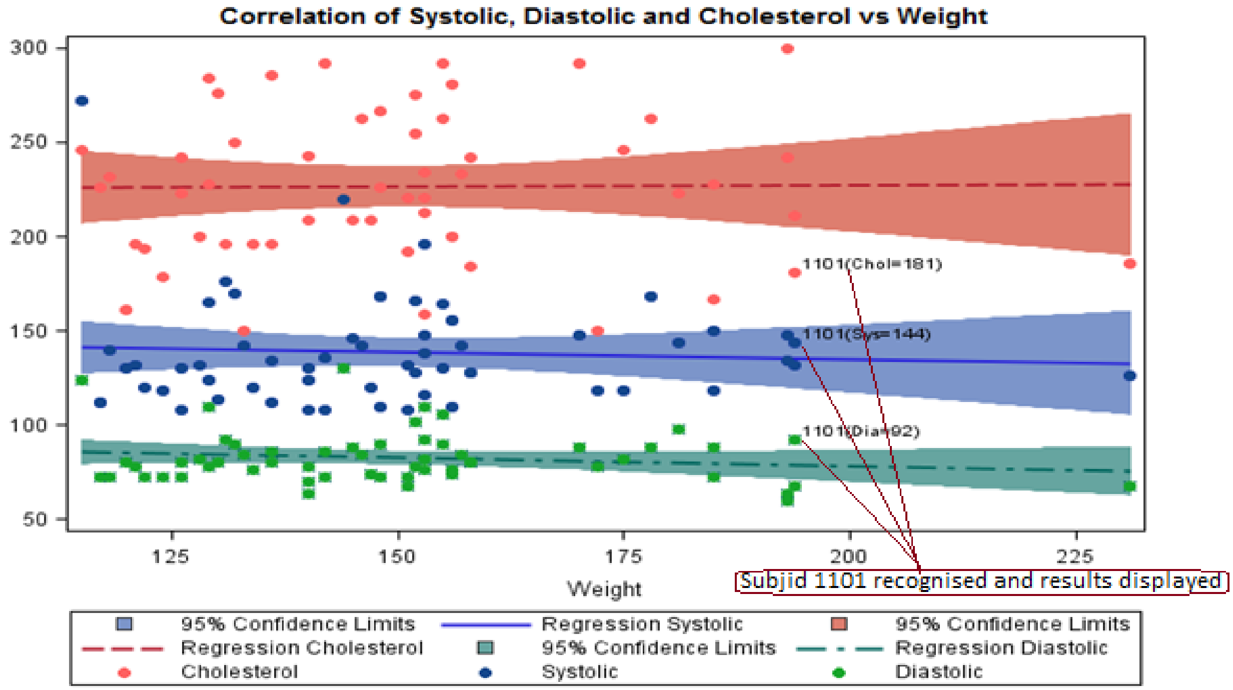


Figure 2, Correlation of Systolic, Diastolic and Cholesterol vs Weight values

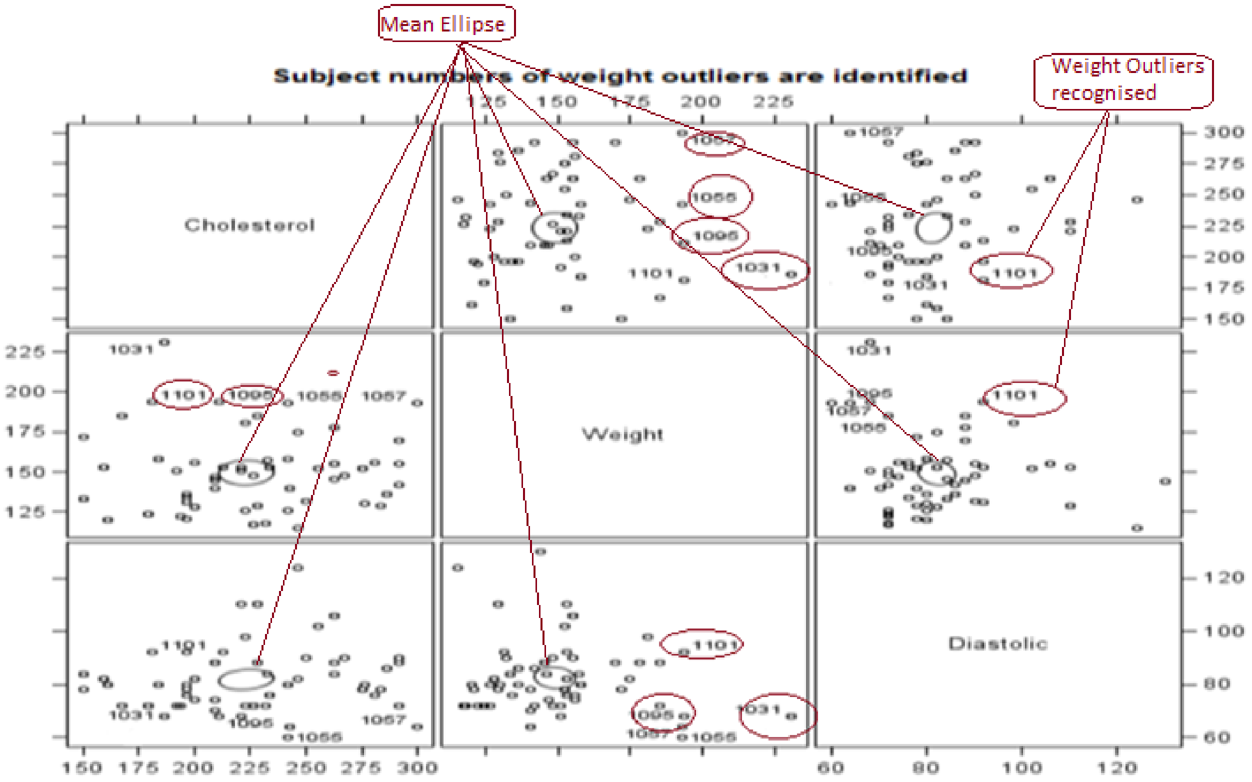
SAS® Code:

```
proc sgplot data=vital_data;
  Title "Correlation of Systolic, Diastolic and Cholesterol vs Weight ";
/********************************************************************************/
 Creating regression lines and 95% CI. The Alpha values can be adjusted
 based on the CI of interest
/********************************************************************************/

  REG X=weight Y=Systolic/alpha=0.05 clm  legendlabel="Regression Systolic";
  REG X=weight Y=Cholesterol/alpha=0.05 clm legendlabel="Regression Cholesterol";
  REG X=weight Y=Diastolic/alpha=0.05 clm legendlabel="Regression Diastolic";
/********************************************************************************/
Notice that here we are creating correlation scatter plots for Cholesterol Systolic
and Diastolic in the same plot.
The data label option uses three different variables with information on the subjid
and result values for each parameter.
The Symbol and color statement are used to define the type of symbol and color options
/********************************************************************************/
scatter x=weight y=Cholesterol/datalabel=subjid_diplay_Chol
 markerattrs=(symbol=CircleFilled color=CXFF6060 );
scatter x=weight y=Systolic /datalabel=subjid_diplay_sys
 markerattrs=(symbol=CircleFilled color=CX13478C );
scatter x=weight y=Diastolic / datalabel=subjid_diplay_dia
 markerattrs=(symbol=CircleFilled color=CX16A629);
 yaxis display=(nolabel);

run;
```

TYPE 2: USING THE MATRIX STATEMENT TO CROSS COMPARE MULTIPLE PARAMETERS ON THE SAME PANEL.

Figure 2 below identifies weight outlier subjects (weight >190) and the subjid's are displayed only for the outliers.

The Matrix statement creates a Correlation matrix of all possible parameter pair combinations to be analyzed. Notice that the datalabel=subjid_diplay_weight statement displays the weight outliers subjid (Some circled in brown for quick reference). The Vital parameters for these outliers can easily be compared and analyzed with the rest of the study population, which will provide us with an overall understanding of the impact of weight on other vital parameters in this study database.


Mean Ellipse

Weight Outliers recognised

Figure 2, Correlation of Weight vs. Diastolic and Cholesterol values using matrix option

SAS® Code:

```
proc sgscatter data= vital_data;
  matrix Cholesterol Weight Diastolic
/*************************************************************************/
The ellipse statement with type=mean creates the mean ellipse as shown in the
figure 2. Type=prediction could be used if we need the prediction ellipse
instead of mean.
Notice that the datalabel=subjid_diplay_weight statement displays the weight
outliers subjid values in the figure
/*************************************************************************/
/ Ellipse = (type=mean) datalabel=subjid_diplay_weight;
run;
```

Now let us take this a step further and try to recognize the subject 1101(our subject of interest) and use the diagonal option to add histogram and normal distribution curve as shown below in Figure 3.

This plot now provides us with a better understanding of the data at the same time, allowing us to recognize the subject 1101 and understand this subject based on the remaining subjects in the database from an elevated angle.

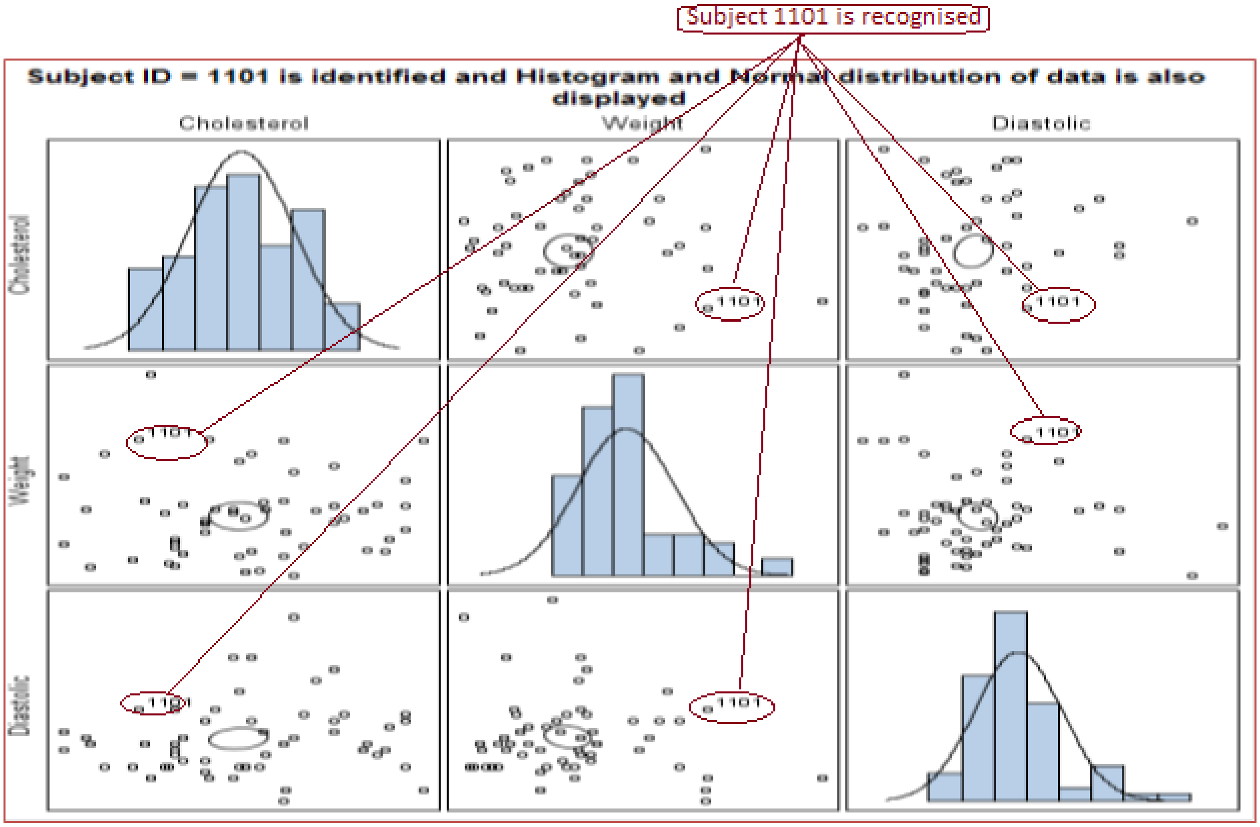


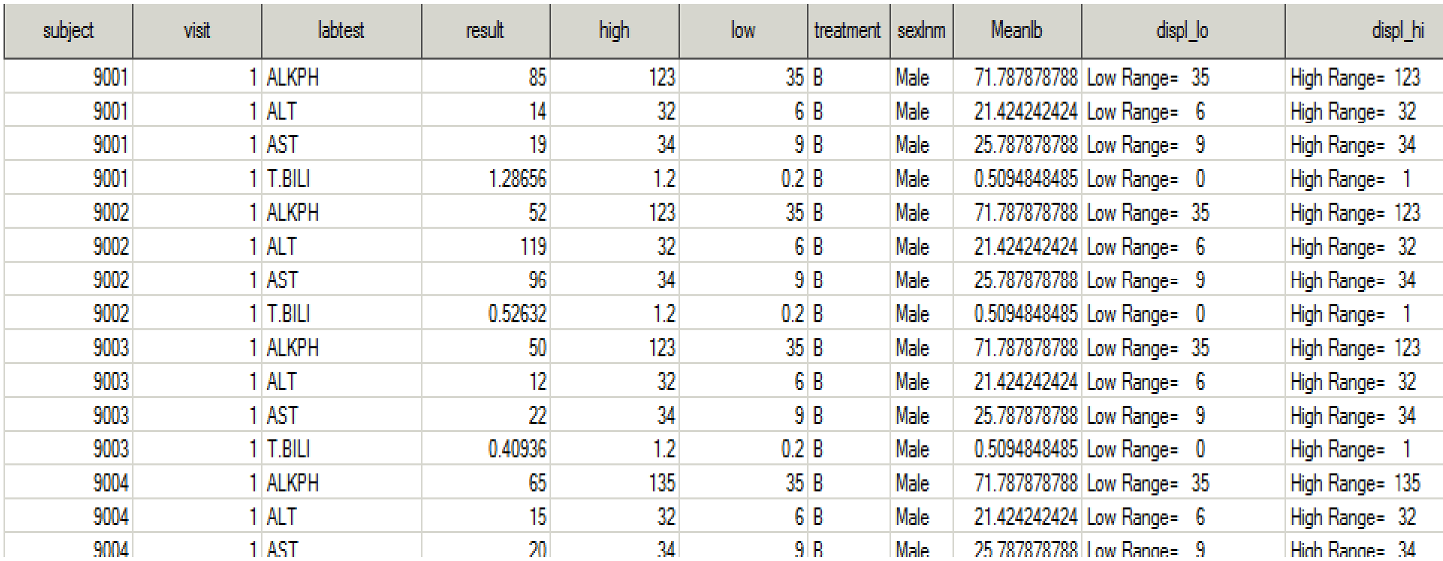
Figure 3, Histogram and normal distribution curves

SAS® Code:

```
Title "Subject ID = 1101 is identified and Histogram and Normal distribution
of data is also displayed";
proc sgscatter data= vital_data;
matrix Cholesterol Weight Diastolic
/ ellipse=(type=mean) datalabel=subjid_diplay_sub
/*Diagonal statement used generates the histogram and distribution curve */
  diagonal=(histogram normal) ;
run;
```

TYPE 3: MULTIPLE PARAMETER PLOTS WITH MEAN VALUE AND REFERENCE VALUES AT EACH VISIT USING SGPANEL.

In the next series of plots, let us consider labs data that has values at each visit as shown in Table 2 below. Note that the meanlb variable is a derived variable that has the overall mean value at each visit, which will be used in our summary plots. Meanlb variable will be used in our plots to determine the general mean values at each visit for all subjects combined at each visit and to compare with the values of the subject of interest.



subject	visit	labtest	result	high	low	treatment	sexlnm	Meanlb	displ_lo	displ_hi
9001	1	ALKPH	85	123	35	B	Male	71.787878788	Low Range= 35	High Range= 123
9001	1	ALT	14	32	6	B	Male	21.424242424	Low Range= 6	High Range= 32
9001	1	AST	19	34	9	B	Male	25.787878788	Low Range= 9	High Range= 34
9001	1	T.BILI	1.28656	1.2	0.2	B	Male	0.5094848485	Low Range= 0	High Range= 1
9002	1	ALKPH	52	123	35	B	Male	71.787878788	Low Range= 35	High Range= 123
9002	1	ALT	119	32	6	B	Male	21.424242424	Low Range= 6	High Range= 32
9002	1	AST	96	34	9	B	Male	25.787878788	Low Range= 9	High Range= 34
9002	1	T.BILI	0.52632	1.2	0.2	B	Male	0.5094848485	Low Range= 0	High Range= 1
9003	1	ALKPH	50	123	35	B	Male	71.787878788	Low Range= 35	High Range= 123
9003	1	ALT	12	32	6	B	Male	21.424242424	Low Range= 6	High Range= 32
9003	1	AST	22	34	9	B	Male	25.787878788	Low Range= 9	High Range= 34
9003	1	T.BILI	0.40936	1.2	0.2	B	Male	0.5094848485	Low Range= 0	High Range= 1
9004	1	ALKPH	65	135	35	B	Male	71.787878788	Low Range= 35	High Range= 135
9004	1	ALT	15	32	6	B	Male	21.424242424	Low Range= 6	High Range= 32
9004	1	AST	20	34	9	B	Male	25.787878788	Low Range= 9	High Range= 34

Table 2, Input data format

Notice that the SGPANEL procedure here plots the Alkaline phosphatase (ALKPH), Alanine aminotransferase (ALT), Aspartate aminotransferase (AST) and Total bilirubin (T.BILI) values for the subject of interest in the same panel along with the reference ranges and mean result values of all subjects combined(which is used as a reference). This clustered set of plots provides a clearer understanding about this subjects lab values compared to normal ranges and with the overall study database as the subjects progressed through the study.

The transparency option on the reference plots causes the reference plots to be less transparent (transparency can be adjusted as per need) providing a better overall visualization. Here again, the datalabel option is used to display the result values at each visit.

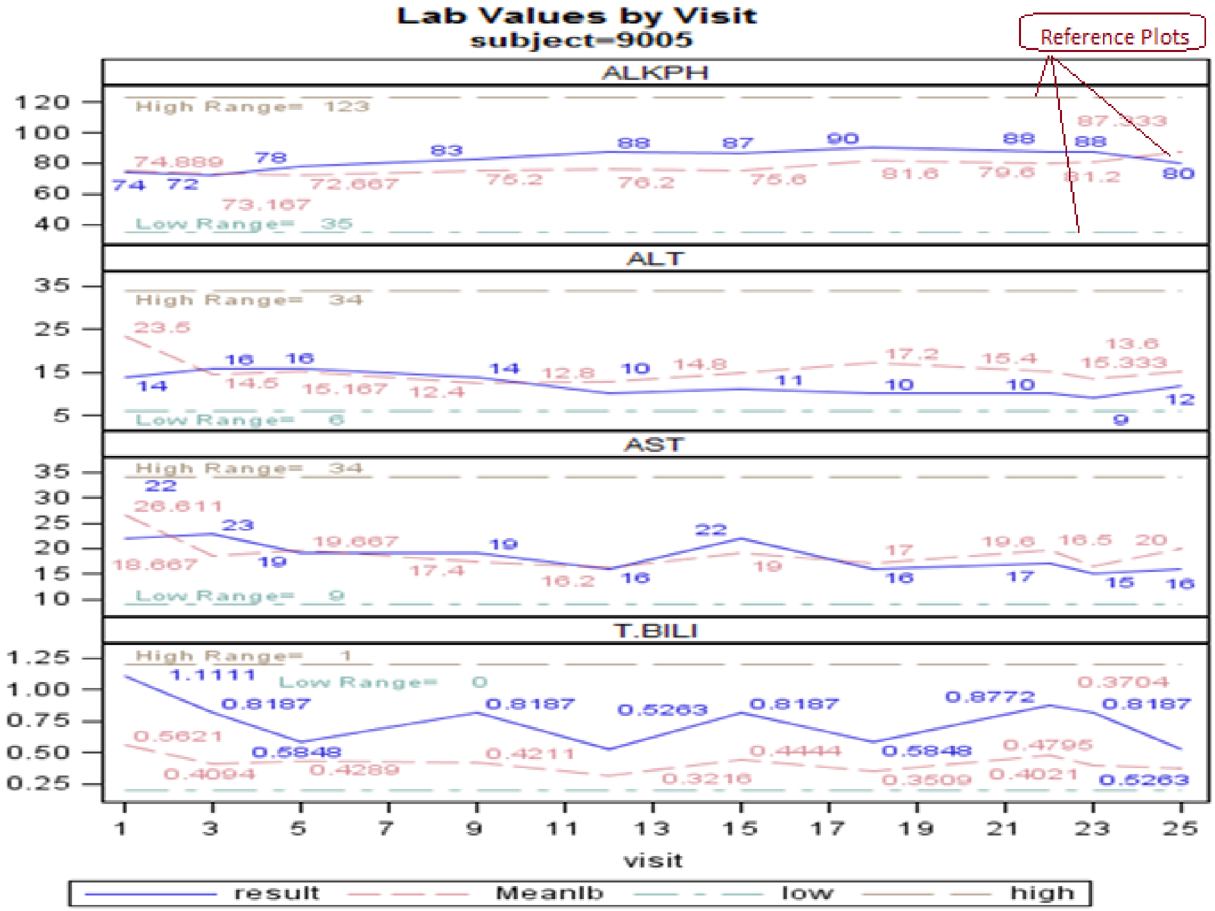


Figure 4, SGPANEL plots the mean values along with reference ranges on the same panel for multiple lab values

SAS® Code:

```
TITLE 'Lab Values by Visit ';

proc sgpanel data = labs_(where=( subject=9005 and labtest in("ALT" "AST"
"T.BILI" "ALKPH") ));;
/*Specifies the number of columns and rows in the panel*/
  panelby labtest / columns = 1 rows = 4 uniscale = column novarname ;
  series x = visit y = result/ datalabel = result  ;
/*Transparency statement is used only for the reference plots*/

  series x = visit y = meanlb / datalabel = meanlb transparency = 0.5 ;
  series x = visit y = low/ datalabel = displ_lo transparency = 0.5;
  series x = visit y = high / datalabel = displ_hi transparency = 0.5;
 by subject;
 colaxis values = (1 to 25 by 1);
run;
```

TYPE 4: MULTIPLE AXIS (Y AND Y2) PLOTS USING THE Y2AXIS STATEMENTS ALONG WITH MEAN AND REFERENCE RANGES

In this section, we are utilizing Multiple Axis (Y and Y2) along with the overall mean and reference range lines to analyze the subject 9005. The ALT result values are plotted on the Y axis and AST on the Y2 axis at each visit as shown below. Notice that overall ALT/AST mean values of the subjects in this database are elevated at the first few visits and then reaches normal levels as subjects progressed through the study. However, subject 9005 has normal ALT/AST at the initial visit and appears to follow the trend of the overall population at the later visits. This is the type of details that would have been overlooked if these plots were not put together.

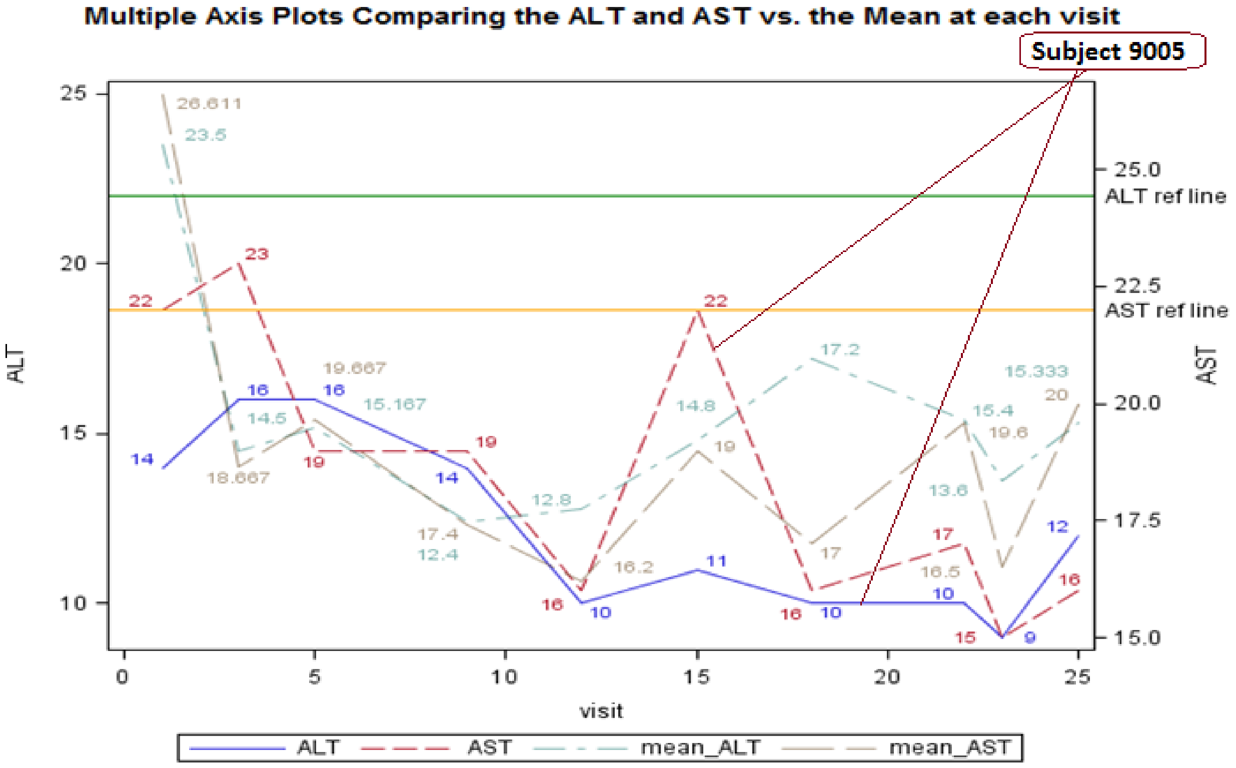


Figure 5, Multiple "Y" axis along with the overall mean and reference range lines

SAS® Code:

```
proc sgplot data = merged(where=( subject=9005));
  series x = visit y = alt/ datalabel = alt;;
  series x = visit y = ast / y2axis datalabel = ast;
  series x = visit y = mean_alt / transparency = 0.5 datalabel = mean_alt;
  series x = visit y = mean_ast / y2axis transparency = 0.5 datalabel =
mean_ast ;
  refline 22 /axis=y label="ALT ref line" lineattrs=(color=green)  ;
  refline 22/axis=y2 label="AST ref line" lineattrs=(color=orange)  ;
 title1 "Multiple Axis Plots Comparing the ALT and AST vs. the Mean at each
visit";
run;
```

CONCLUSION

The plots and summaries discussed above provide a methodology for a graphical overall understanding of the database, along with a strong understanding of the contribution of subjects of interest. There are numerous enhancements that can be incorporated into these plots using several statistics since every data speaks differently.

With the introduction of SG procedures, incorporating summary statistics within the graphics is made simpler and provides a broader understand the subjects. Having a "Birds-eye view" on the database is the key to an efficient analysis!

REFERENCES

Susan J. Slaughter, Avocet Solutions, Davis, Lora D. Delwiche, University of California,(2010).

"Using PROC SGPLOT for Quick High-Quality Graphs" SAS Global Forum,

http://support.sas.com/resources/papers/proceedings10/154-2010.pdf

Mina Chen, Roche Product Development in Asia Pacific, Shanghai, China (2015).

"Handling multiple y-axes using SAS® Graphs" PharmaSUG China

http://www.lexjansen.com/pharmasug-cn/2015/PT/PharmaSUG-China-2015-PT47.pdf

Sanjay Matange, SAS Institute Inc. (2016) "Clinical Graphs Using SAS®" SAS Institute.

http://support.sas.com/resources/papers/proceedings16/SAS4321-2016.pdf

ACKNOWLEDGMENTS

Thanks to Aman Bahl, Senior Manager of Statistical Programming at inVentiv Health for providing valuable feedback and for encouraging this publication.

CONTACT INFORMATION

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

Hrideep Antony
inVentiv Health
1001 Winstead Drive, Suite 200
Cary, NC 27513
Phone: 919-337-1415
E-mail:Hrideep.antony@inventivhealth.com
Web: http://www.inventivhealth.com