

Techniques of Preparing Datasets for Visualizing Clinical Laboratory Data

Amos Shu, MedImmune, Gaithersburg, MD

Victor Sun, MedImmune, Gaithersburg, MD

ABSTRACT

Visualizing clinical laboratory data helps clinicians and statisticians quickly understand the results of lab tests. Some people^[1-4] have been able to easily generate figures for clinical laboratory data by using SAS[®] Graph Template Language and SAS/GRAPH[®] SG Procedures. The most difficult part of generating graphs, however, is not the SAS/GRAPH[®] Language, but the process of creating the datasets that are used to generate the graphs. This paper discusses the techniques of preparing datasets that will be used to generate five types of figures for clinical laboratory data. These same techniques are applicable to visualizing electrocardiography and vital signs data as well.

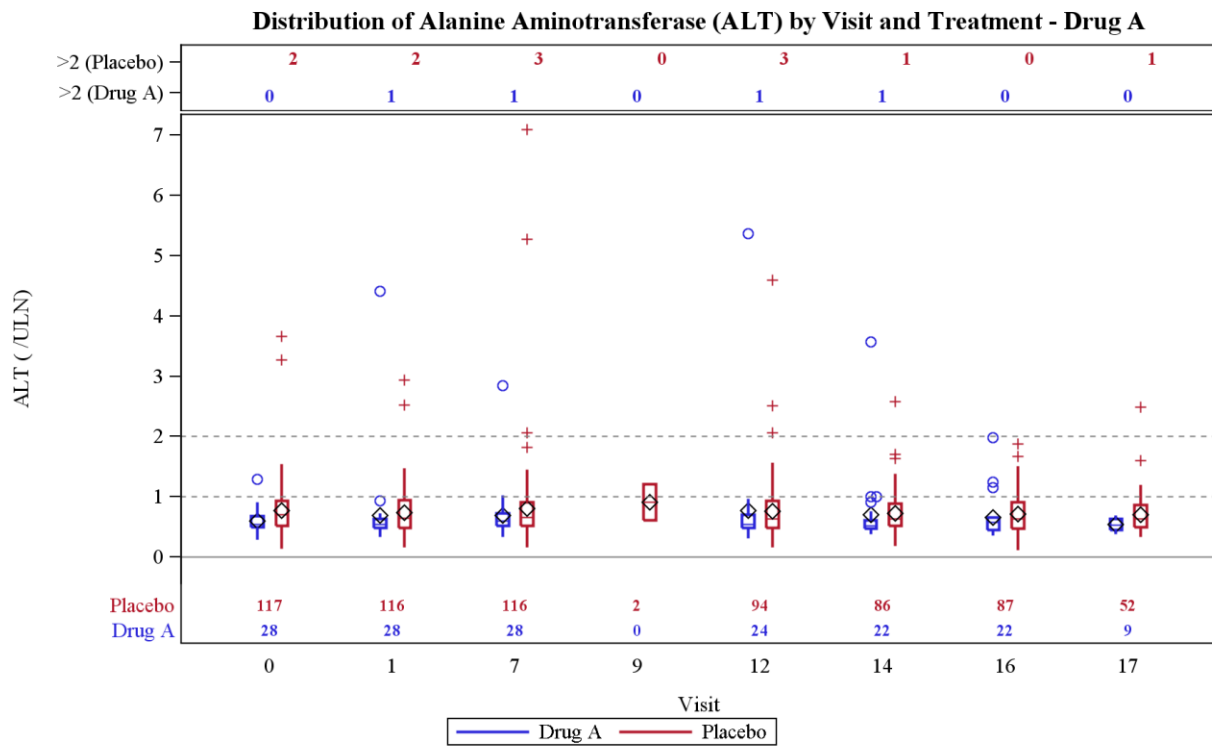
INTRODUCTION

SAS Graph Template Language and SAS/GRAPH SG Procedures provide simple approaches to generate graphs easily. The most difficult part in the whole graphing process is to create the datasets that are suitable for the graph that you are trying to create. Visualizing clinical laboratory data is a great way to help clinicians and statisticians quickly understand the results of lab tests. This paper will focus on how to create datasets that are used to visualize clinical laboratory data in five different ways. Those techniques can also be used to visualize electrocardiography and vital signs data.

Most variable names used in this paper are from the CDISC SDTM Implementation Guide (Version 3.1.2). All programs presented in this paper were developed in Server SAS[®] 9.2 in the Windows environment.

1. DISTRIBUTION OF LAB TESTS BY STUDY TIME AND TREATMENT

A typical figure of distribution by visit and treatment for ALT (Alanine Aminotransferase) looks like the following . It includes four parts – the number of subjects who have the test results above the clinical concern level, the clinical concern level scatter plot, the box plot, and the number of subjects having the test at each visit. BOXPLOTPARM plot, BLOCKPLOT, SCATTER plot, and REFERENCE plot are used in the TEMPLATE procedure to generate this figure.



The dataset has to contain data for the above four parts:

stat	VISITNUM	Drug_A	Placebo	NT	NP	NTccl	NPccl	labelT	labelP
MIN	1	0.333	0.159	28	116	1	2	>2 (Drug A)	>2 (Placebo)
Q1	1	0.485	0.485	0	0	0	0		
MEDIAN	1	0.545	0.648	0	0	0	0		
MEAN	1	0.699	0.743	0	0	0	0		
Q3	1	0.636	0.947	0	0	0	0		
MAX	1	0.727	1.477	0	0	0	0		
OUTLIER	1	0.939	0	0	0	0	0		
FAROUTLIER	1	4.409	0	0	0	0	0		
FAROUTLIER	1	0	2.932	0	0	0	0		
FAROUTLIER	1	0	2.523	0	0	0	0		
.....									

The clinical concern level (CCL) is the key variable that has to be created first:

```
CCL = LBSTRESN/LBSTNRHI ;
```

For ALT, the Clinical Concern Level (CCL) is 2 ULN; ULN is the upper level of normal range.

Five variables – visitnum, labelT, labelP, NTccl, and NPccl within SCATTERPLOT statements will be used to display the number of subjects who have the test results above the clinical concern level at each visit, i.e.

```
SCATTERPLOT X=visitnum Y=labelT / MARKERCHARACTER = ntCCL;
```

Other variables such as FAROUTLIER, OUTLIER, Q1,Q3, MEAN, MEDIAN, MIN, and MAX can be calculated following page 215 in the pdf document of Graph Template Language Reference (GTL) 9.3. Then transpose these variables into three variables – stat, Drug_A and Placebo, which will be used in BOXPLOTPARM statements to draw the box plot.

```
BOXPLOTPARM x=visitnum y=Drug_A stat=stat /BOXWIDTH=0.1 CAPSHAPE=bracket
DISPLAY=(outliers mean median) ;
```

The REFERENCELINE statements will draw the normal range values of the ALT test.

```
REFERENCELINE Y= 2 / LINEATTRS=(pattern=shortdash) ;
```

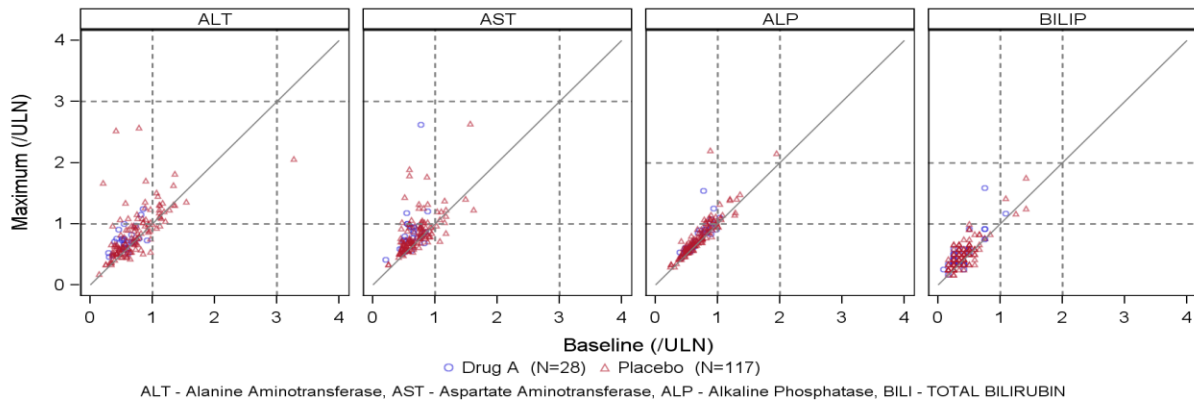
BLOCKPLOT statements with two variables – NT and NP will put the number of subjects having the test in Drug A and placebo at each visit into the final part of this figure, respectively.

```
BLOCKPLOT X=visitnum BLOCK=NT / DISPLAY=(label values) ;
```

2. PANEL OF LIVE FUNCTION TEST SHIFT FROM BASELINE TO MAXIMUM BY TREATMENT

The best way to display the results of live function tests is to put each test into a panel on one page as follows:

Panel of Live Function Test Shift from Baseline to Maximum by Treatment



For ALT, AST, and ALP, the Clinical Concern Level (CCL) is 2 x ULN: where ULN is Upper Level of Normal Range
 The Clinical Concern Level for BILIP is 1.5 x ULN

The SGPPANEL procedure uses SCATTER, SERIES, and REFLINE statements to generate this figure. The dataset for this figure is not difficult to create. The clinical concern level (CCL) is the key variable. For ALT, AST, and ALP, the Clinical Concern Level (CCL) is 2 ULN; ULN is the upper level of normal range. The clinical concern level for BILIP is 1.5 ULN. Here variable Baseuln is the clinical concern level at baseline, and variable maxuln is the maximum clinical concern level among post visits. trt is the treatment group variable.

```
SCATTER X=baseuln Y=maxuln / group=trt ;
```

The line of CCL would help clinicians focus on the upper left-hand quadrant of the figure. Variable Ser is used to draw the line of CCL in the SERIES statement.

```
SERIES X=Ser Y=Ser /LINEATTRS=graphreference ;
```

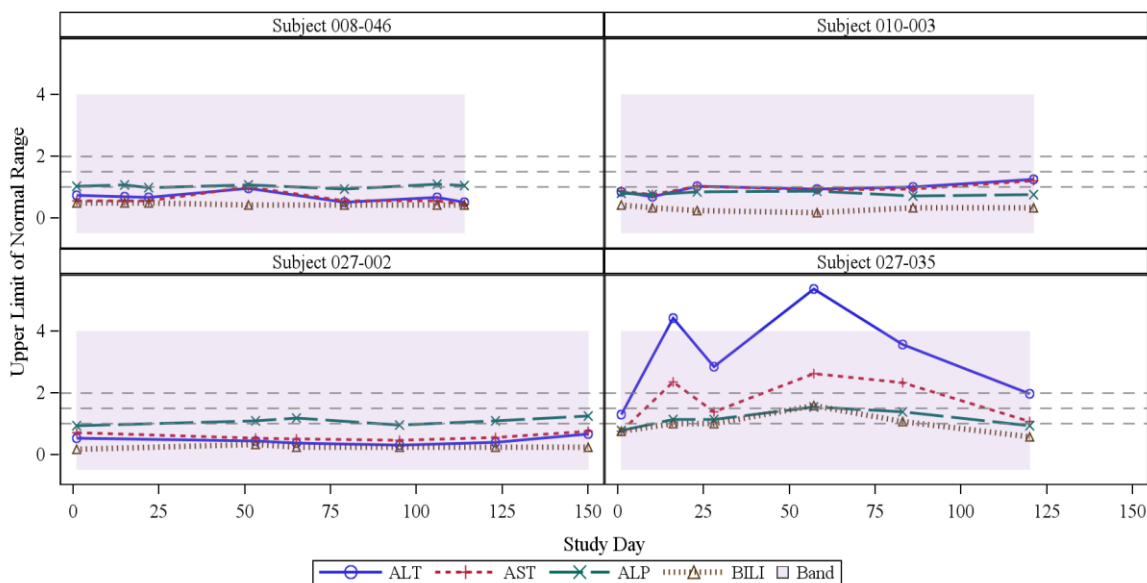
The difficult part for this figure is the reference lines, which may vary by tests. Variable Ref is used to draw the reference lines in the REFLINE statements.

trt	USUBJID	LBTESTCD	baseuln	maxuln	Ser	Ref
Drug A (N=28)	006-022	ALT	0.66	0.68		
Drug A (N=28)	006-029	ALT	0.61	0.52		
Drug A (N=28)	008-013	ALT	0.36	0.48		
Drug A (N=28)	008-022	ALT	0.55	0.61		
Drug A (N=28)	008-046	ALT	0.73	0.97		
Placebo (N=117)	028-022	ALT	0.42	0.52	0	
Placebo (N=117)	028-022	ALT	0.42	0.52	4	
Placebo (N=117)	028-022	ALT	0.42	0.52		1
Placebo (N=117)	028-022	ALT	0.42	0.52		3

3. RISK PATIENTS PROFILES OF LIVE FUNCTION TESTS BY TRIAL DAYS

For patients whose test results contain abnormal values, you can display each patient in a panel on a page as the following figure. This will be helpful for exploring individual patient.

Liver Function Tests by Trial Day: At Risk Subjects



ALT - Alanine Aminotransferase, AST - Aspartate Aminotransferase, ALP - Alkaline Phosphatase, BILI - TOTAL BILIRUBIN

For ALT, AST, and ALP, the Clinical Concern Level (CCL) is 2 x ULN: where ULN is Upper Level of Normal Range

The Clinical Concern Level for BILIP is 1.5 x ULN

Shaded region indicates time on treatment.

SERIES, BAND, and REFLINE statements will be used in SGPPANEL procedure. The dataset for this figure contains the results of each test. LBDY is the Study Day of Specimen Collection.

```
SERIES X=LBDY Y=ALT /MARKERS MARKERATTRS=(size=10)
LINEATTRS=(THICKNESS=2px PATTERN=solid) ;
```

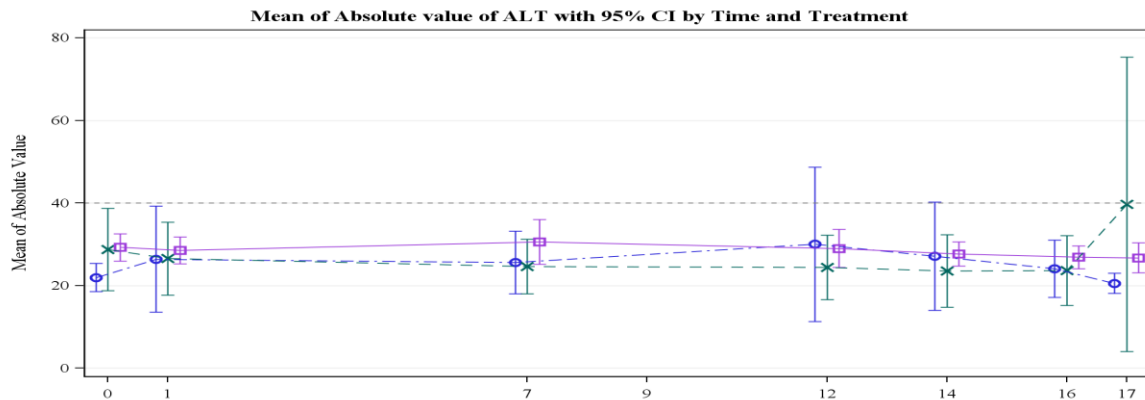
If you use BAND statement to shade treatment region, you will have to create a variable YMin, which is for the lower boundary of the treatment region.

```
BAND X=LBDY LOWER=YMin UPPER=4 /TRANSPARENCY=0.8;
```

USUBJID	ALP	ALT	AST	BILI	LBDY	YMin
Subject 008-046	1.030612	0.727273	0.558824	0.5	1	-0.5
Subject 008-046	1.071429	0.69697	0.558824	0.5	15	-0.5
Subject 008-046	0.979592	0.666667	0.558824	0.5	22	-0.5
Subject 008-046	1.081633	0.969697	1	0.416667	51	-0.5
Subject 008-046	0.938776	0.515152	0.558824	0.416667	79	-0.5
Subject 008-046	1.102041	0.666667	0.558824	0.416667	106	-0.5
.....						

4. MEAN AND MEAN OF CHANGE FROM BASELINE OF LAB TESTS WITH 95% CI OR STANDARD DEVIATION BY STUDY TIME AND TREATMENT

Another typical figure (below) of Lab tests is the mean or mean of change from baseline of lab tests with 95% CI or standard deviation. This type of figure is relatively simple. It only contains two parts – the series plot and the number of subjects having the test at each visit.

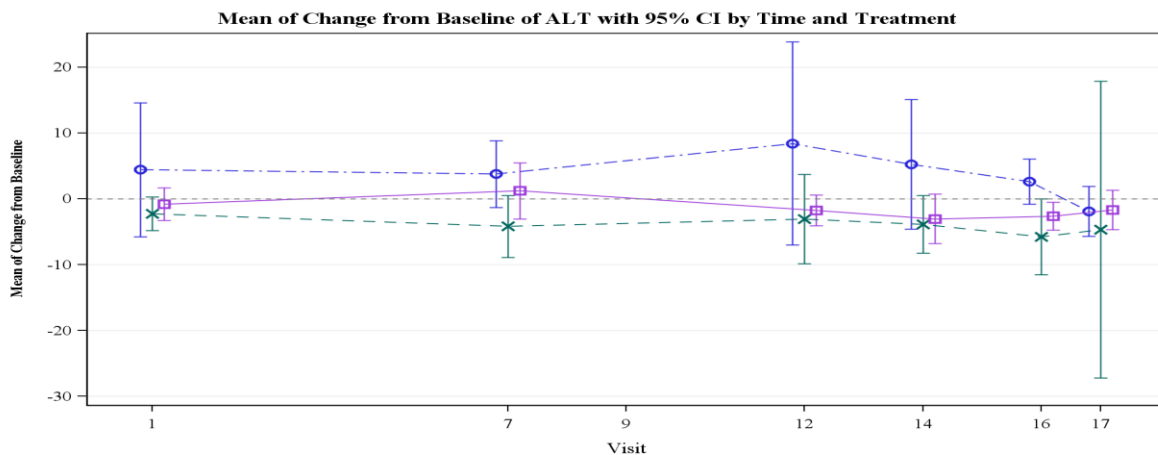


Subjects at Visits

Drug A	28	28
Drug B	21	21
Placebo	118	116

	28		24	22	22	9
	21		18	18	17	6
	116		94	86	87	52

○ Drug A × Drug B — Placebo



Subjects at Visits

Drug A	28
Drug B	21
Placebo	116

	28		24	22	22	9
	21		18	18	17	6
	116		94	86	87	52

--- Drug A --- Drug B --- Placebo

Five key variables – visit, mean, lower 95% confidence interval for the mean (LCLM), upper 95% confidence interval for the mean (UCLM), and the number of subjects (N) with SERIES plot, BLOCKPLOT, SCATTER plot, and REFERENCE plot are used in the TEMPLATE procedure to generate this figure.

VISITNUM	Mean1	Mean2	Mean3	LCLM1	LCLM2	LCLM3	UCLM1	UCLM2	UCLM3	N1	N2	N5
0	21.93	28.73	29.23	18.53	18.72	25.87	25.33	38.73	32.58	28	21	118
1	26.36	26.52	28.55	13.51	17.71	25.31	39.20	35.33	31.79	28	21	116
7	25.55	24.62	30.58	17.96	18.01	25.15	33.14	31.23	36.01	28	21	116
12	30.00	24.39	29.00	11.28	16.59	24.42	48.72	32.19	33.58	24	18	94
14	27.09	23.56	27.66	13.99	14.80	24.71	40.19	32.31	30.62	22	18	86
16	24.09	23.65	26.84	17.15	15.20	24.05	31.03	32.10	29.63	22	17	87
17	20.50	39.67	26.72	18.06	4.02	23.13	22.94	75.31	30.31	9	6	52

Mean, LCLM, UCLM can be calculated by PROC MEAM or PROC SUMMARY procedure.

```

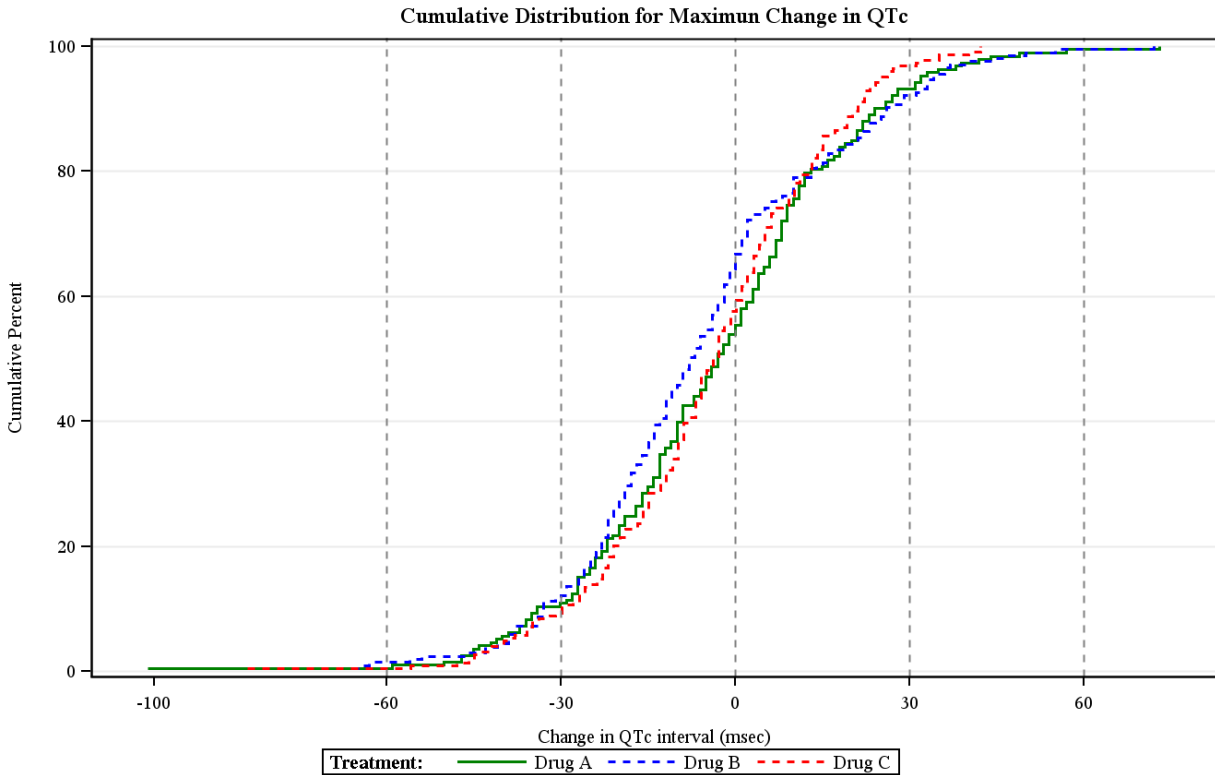
SCATTERPLOT X=visitnum Y=mean1 / YERRORLOWER=LCLM1 YERRORUPPER =UCLM1
            MARKERATTRS=graphdata1(size=9px weight=bold )
            ERRORBARATTRS=graphdata1(pattern=solid thickness=2);
SERIESPLOT X=visitnum Y=mean1 / LINEATTRS=graphdata1(pattern=DashDot
            thickness=2);
BLOCKPLOT X=visitnum BLOCK=N1 / DISPLAY=(label values) ;

```

5. CUMULATIVE DISTRIBUTION FOR MAXIMUM CHANGE IN ECG TESTS

This type of figure is often used to display the results of electrocardiography (ECG) tests. The dataset for the following graph contains four variables. Drug_A, Drug_B, and Drug_C are cumulative percentage of the number of subjects for drug A, B, and C, respectively.

EGCHGBL	Drug_A	Drug_B	Drug_C
-5	47.15	54.63	48.21
-4	48.70	57.07	50.45
-3	50.78	59.02	54.46
-2	52.33	61.95	55.36
-1	53.89	64.39	57.59
0	55.44	66.83	59.38
1	58.03	69.27	61.61
2	59.07	72.20	63.84
3	61.14	73.17	66.52
4	63.73		68.30
5	64.77	74.15	70.98



Only STEP plot and REFERENCE plot are needed in the TEMPLATE procedure to draw this figure.

```
STEPLOT X= EGCHGBL Y=Drug_B /LINEATTRS=graphdata2(PATTERN=shortdash
color=blue THICKNESS=2.0px);
```

CONCLUSION

Visualize clinical laboratory measurements is usually not straightforward. Data transformations, such as summarizing data, transposing data, scaling the data values, and creating new variables, are often necessary for generating right graphs. SAS Graph Template Language and SAS/GRAPH SG Procedures have been very helpful to generate typical figures for lab results. Creating the right datasets and having a good understanding of SAS Graph Template Language and SAS/GRAPH SG Procedures are keys to success in the lab data visualization process.

REFERENCES

- [1]. Amit, Ohad, *Graphical approaches to the analysis of safety data from clinical trials*, Pharmaceutical Statistics, January/March 2008
- [2]. Shi-Tao Yeh, *Clinical Laboratory dataData Analysis and Visualization*, PharmaSUG, 2007
- [3]. Sanjay Matange, *Tips and Tricks for Clinical Graphs using ODS Graphics*, SAS Global Forum, 2011
- [4]. Sanjay Matange, *Clinical Graphs using SAS 9.2*, SAS Global Forum, 2012

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Please contact the author at:
 Amos Shu
 MedImmune, LLC
 One MedImmune Way
 Gaithersburg, MD 20878
 Email: shua@medimmune.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.