PharmaSUG 2014 - Paper: DG01

Techniques of Preparing Datasets for Visualizing Clinical Adverse Events

Amos Shu, Endo Pharmaceuticals., Malvern, PA
Qinghua (Kathy) Chen, Exelixis Inc., South San Francisco, CA

ABSTRACT

With the introduction of SAS® Graph Template Language and SAS/GRAPH® SG Procedures, there have been papers [1-4] that have provided very useful codes to easily visualize clinical adverse events (AE). However, it is still not easy to create the suitable datasets that are used to generate the right graphs. This paper discusses techniques of preparing datasets that will be used to generate different types of figures for clinical adverse event.

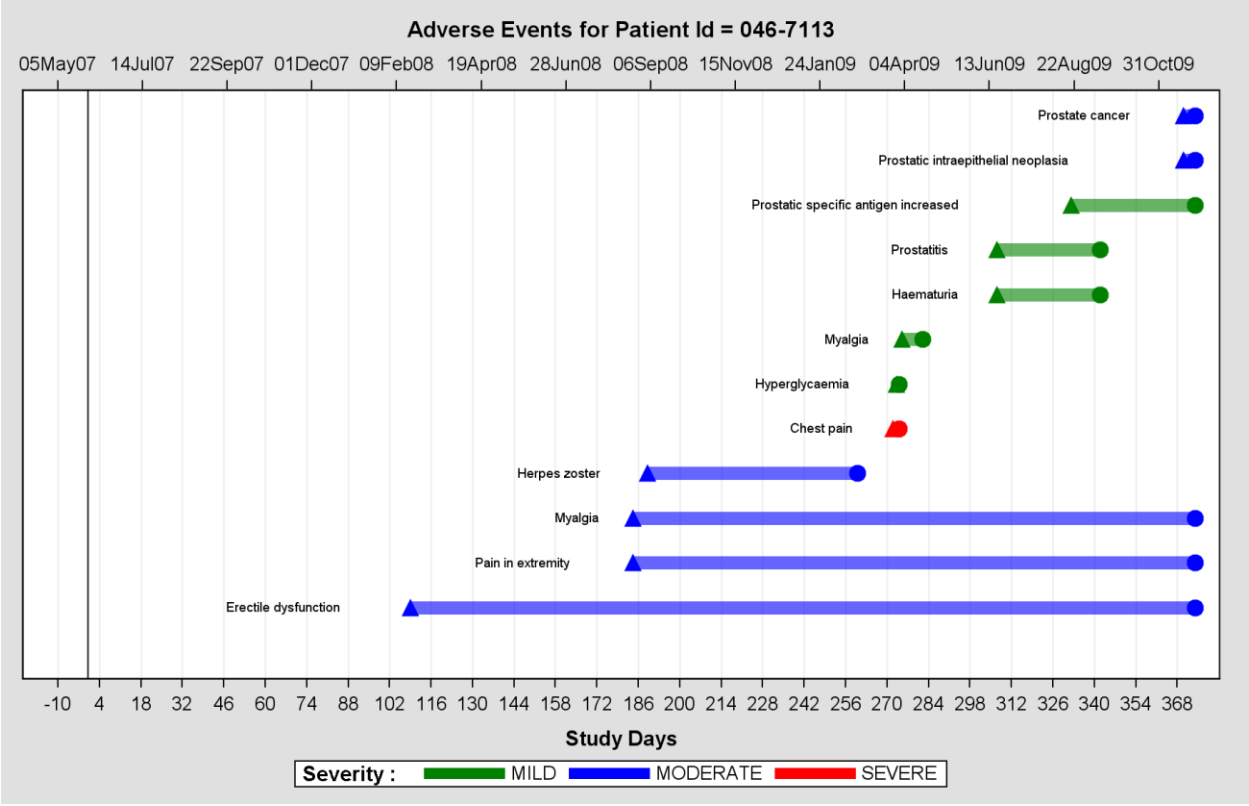
INTRODUCTION

Visualizing clinical adverse events (AE) is a great way to display clinical trial data and facilitate the evaluation of drug safety. SAS Graph Template Language and SAS/GRAPH SG Procedures are quite competitive to other languages such as R in graph making, and have provided simple approaches to generate graphs easily. However, how to generate the datasets that are used to generate graphs remains a difficult part in the process of making figures. This paper will focus on how to create suitable datasets that are used to visualize clinical adverse events.

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. VECTOR PLOT FOR ADVERSE EVENTS FOR INDIVIDUAL PATIENT

This graph is frequently used to track a specific individual patient in a study to understand the subject's adverse events. The following example is for subject 046 – 7113, who has prostate cancer in XYZ study. Information includes AE term, severity, start date, end date, study days of start and end of adverse events related to the first study dosing date, and the duration of each adverse event. VECTOR plot and SCATTER plot are used in the SGPLOT Procedure to generate this figure.



Seven variables in the following dataset are used to generate this figure.

AEDECOD	AESTDY	AEENDY	Y_Value	AESEV	AESTDT	RFSTDTC	AE_Psn
Erectile dysfunction	109	374	0.8	MODERATE	01Sep2007	2007-05-16T07:29	74.5
Chest pain	272	274	4.2	SEVERE	11Feb2008	2007-05-16T07:29	252.5
Hyperglycaemia	273	274	5.0	MILD	12Feb2008	2007-05-16T07:29	247.5
……					14Feb2008	2007-05-16T07:29	260

In VECTOR statements, three variables – AESTDY, AEENDY, and Y_Value (for vertical axis) are used to draw the start point, end point, and duration of each AE, respectively.

```
VECTOR   X=AEENDY   Y=Y_Value / XORIGIN=AESTDY   YORIGIN =Y_Value ;
```

AESTDY and AEENDY are the study day of start and end of adverse event relative to the sponsor-defined RFSTDTC (Reference Start Date), respectively. The sponsor-defined RFSTDTC is designated as Study Day 1, which is usually the first study drug dosing date.

- AESTDY = AESTDTC - RFSTDTC +1;
- AEENDY = AEENDTC - RFSTDTC +1; If AEENDTC is missing, the date of EOS (End of Study) in DSSTDTC can be used instead.
- Y_Value is used for separating each AE in the figure. You can assign such a value to each AE in your own way. Values assigned should give each AE enough space in the figure.

AESTDY, AEENDY, and Y_Value are also used in SCATTER statements to draw the start and end point of each AE, respectively.

```
SCATTER X=AESTDY  Y=Y_Value / MARKERATTRS=(SIZE=11px SYMBOL=trianglefilled);
SCATTER X=AEENDY  Y=Y_Value / MARKERATTRS=(SIZE=11px SYMBOL=circlefilled);
```

To label each AE, you need to use AEDECOD and create another variable (AE_Psn) that positions each AE name. Since each AE name has a different length, you can use the following formula to calculate AE_Psn:

AE_Psn = AESTDY – 1.5 * (the length of AE name) ;

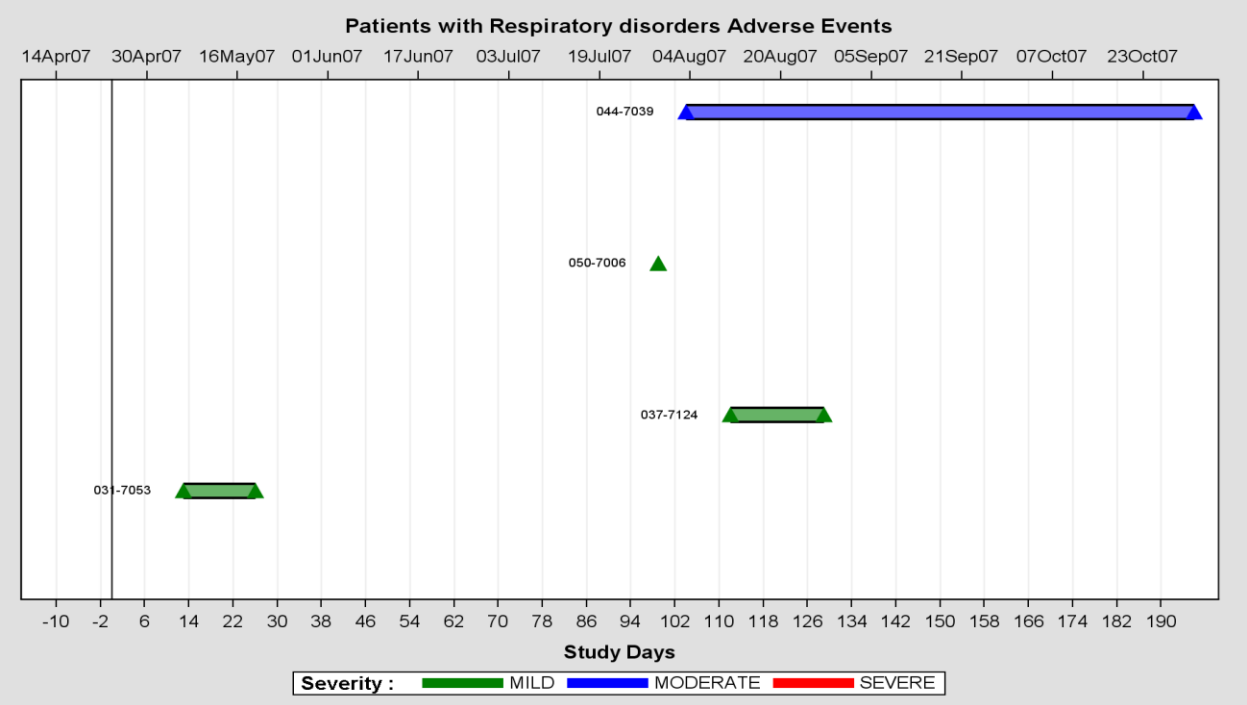
```
SCATTER X=AE_Psn  Y= Y_Value/ MARKERATTRS= AEDECOD  MARKERCHARATTRS=(SIZE=6);
```

Finally, to display the corresponding AE start and end date values to the secondary horizontal axis (X2 axis), another SCATTER statement will be used with AESTDT (the numeric value of AESTDTC):

```
SCATTER X= AESTDT  Y= Y_Value / MARKERATTRS=(SIZE=0)  X2AXIS;
```

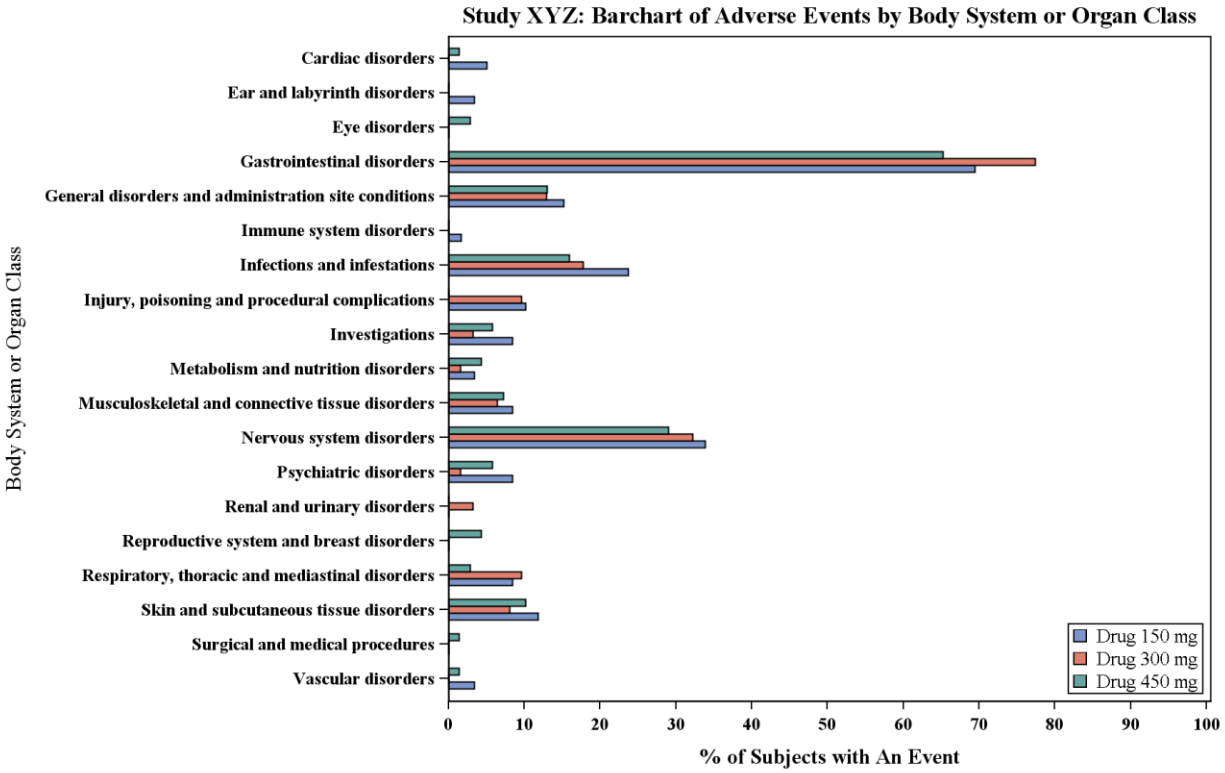
2. VECTOR PLOT FOR SUBJECTS WHO HAVE THE SAME ADVERSE EVENT

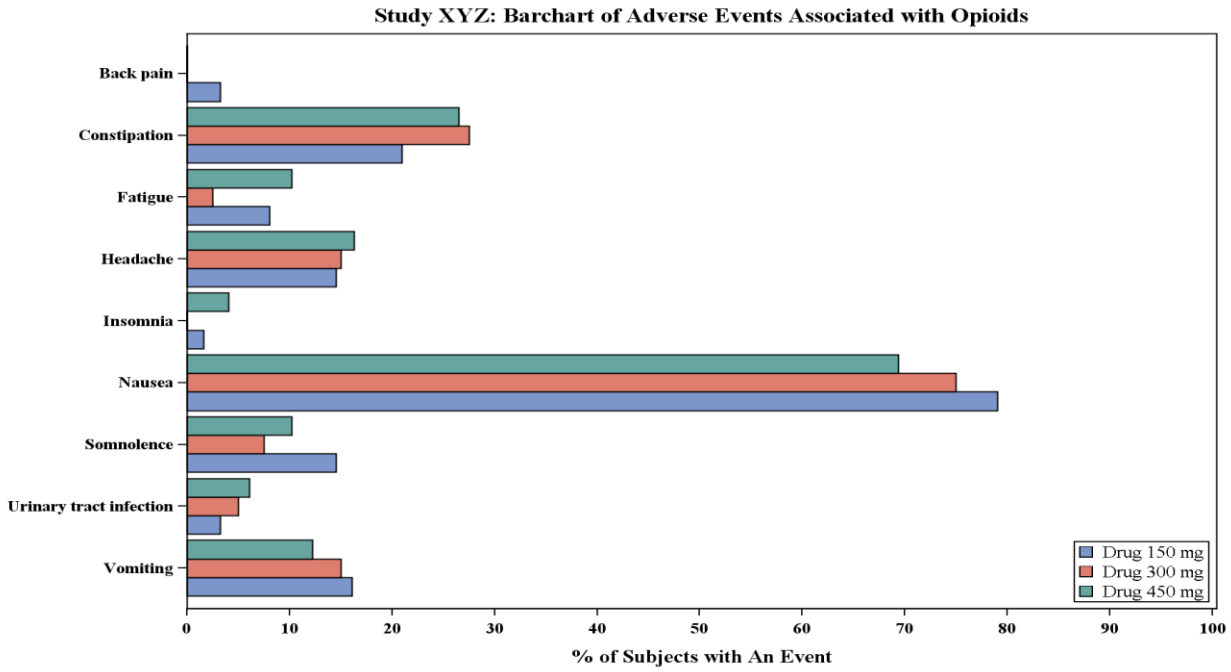
The above figure can also be used for subjects that have the same adverse event in a clinical study. This could be useful when a particular adverse event is getting the attention of the clinicians and/or FDA reviewers. The following figure shows all subjects that have respiratory disorders.



3. BARCHART OF ADVERSE EVENTS BY BODY SYSTEM OR ORGAN CLASS

The bar chart summarizes a set of categorical adverse events by system organ class or preferred term. The size of each bar is proportional to the number of unique subjects in the category it represents. This figure also can be used to display specifically selected adverse events.





The dataset for this type of figure will look like this:

AEBODSYS	Pcnt1	Pcnt2	Pcnt3
Psychiatric disorders	8.475	1.613	5.797
Nervous system disorders	33.9	32.26	28.99
Musculoskeletal and connective tissue disorders	8.475	6.452	7.246
……			

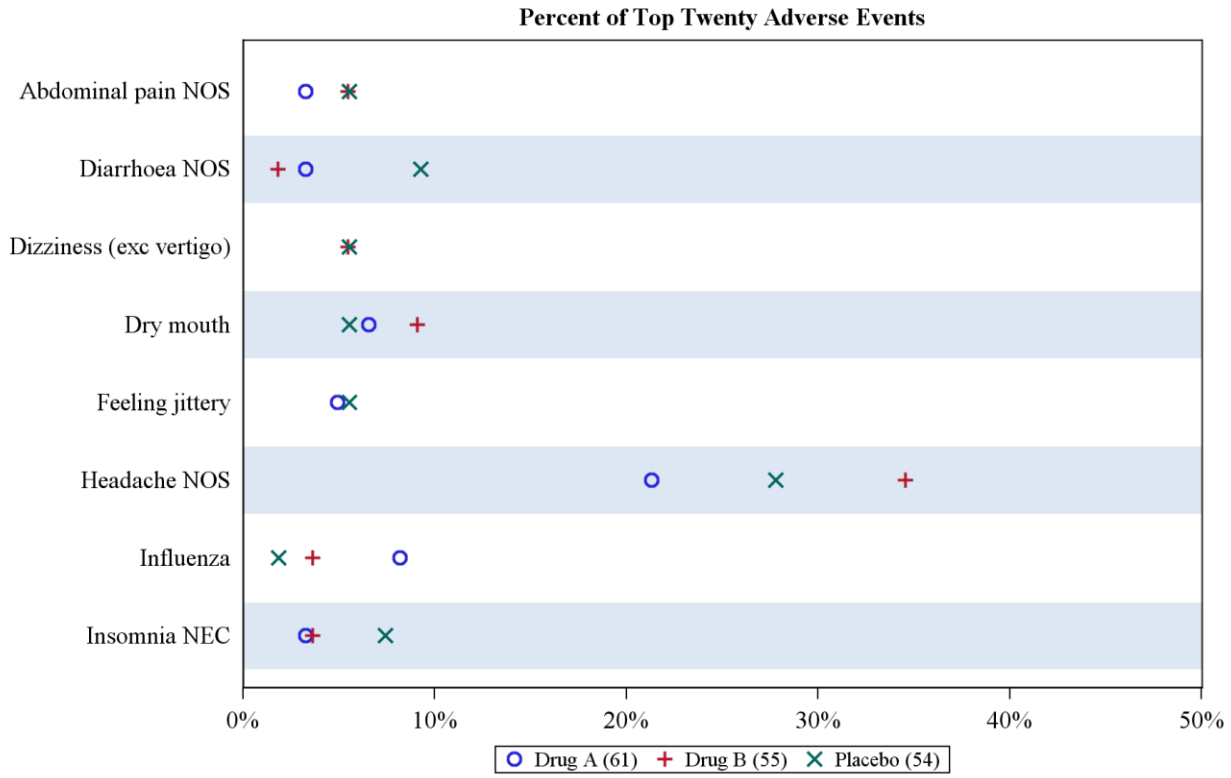
The dataset for this figure is relatively easy to create. In BARCHART statements of the TEMPLATE Procedure, variables AEBODSYS or AEDECOD will display at X axis. The percentage (variables Pcnt1, Pcnt2, Pcnt3) of each categorical adverse event will be equal to the count of unique subjects in the particular categorical adverse event divided by the count of total unique subjects in that treatment group. Variables Pcnt1, Pcnt2, Pcnt3 are for the Y axis:

```
BARCHART X=AEDECOD  Y=Pcnt1 / ORIENT= horizontal;
```

The GROUPDISPLAY = cluster option is available in SAS 9.3, and will make the code much simple.

4. SCATTER OF PERCENT OF TOP TWENTY OR MOST CRITICAL ADVERSE EVENTS

To get a rough picture of the top twenty percent or the most critical adverse events in a study, the following figure is useful.



This type of figure will use a dataset that looks like the following.

treat	AEDECOD	Pcnt	BarH
Drug A (61)	Diarrhoea NOS	3.28%	
Drug B (55)	Diarrhoea NOS	1.82%	
Placebo (54)	Diarrhoea NOS	9.26%	0.5
……			

In the TEMPLATE Procedure, a SCATTERPLOT statement will be used to draw each plot. AEDECOD will display at Y axis. The percentage (variable Pcnt) of each categorical adverse event will be equal to the count of unique subjects

in the particular adverse event divided by the count of total unique subjects in that treatment group. Variable Pcnt is for the X axis:

```
SCATTERPLOT  X=Pcnt  Y= AEDECOD / MARKERATTRS=(size=9 weight=bold);
```

To shade or band AEs for easier reading, you have to create a numeric variable (BarH) and use a BARCHART statement as follows. Barht is 0.5 here.

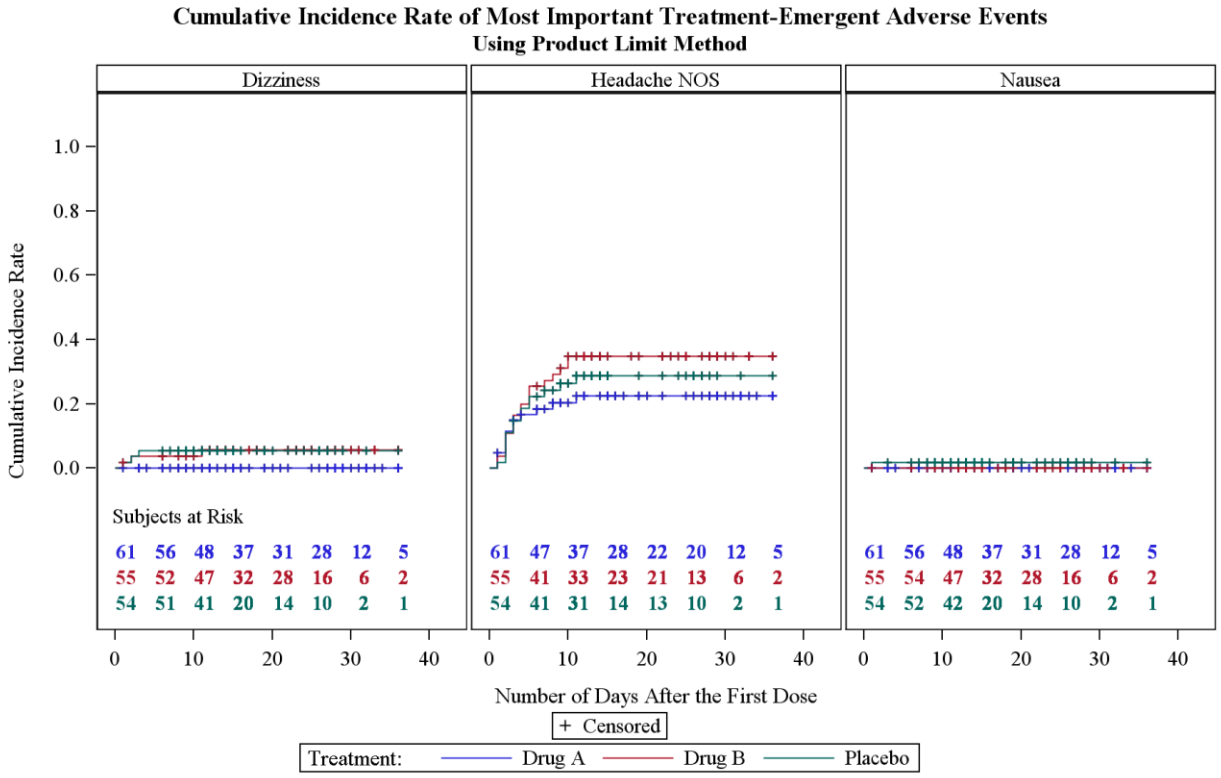
```
BARCHART X=AEDECOD  Y=BarH/ORIENT= horizontal DISPLAY=(fill)
                                  DATATRANSPARENCY=0.5;
```

5. TIME-TO-EVENT PLOT FOR ADVERSE EVENTS

Adverse events are commonly treated as time-to-event in clinical statistics. There are two ways to draw the time-to-event plot, which shows the survival estimates over time by treatment. One is the product-limit method (Kaplan-Meier plot), and the other one is the Hazard function for each treatment.

1). Product-Limit Method (Kaplan-Meier plot)

The following Kaplan-Meier survival plot for adverse events includes three parts – the graph with the survival step plot, the "Censored" scatter plot, and the number of subjects at risk. STEP plot and SCATTER plot in SGPANEL procedure are used to generate this figure.



This type of figure will need a complex dataset as follows.

AEDECOD	Treat	TTE	TTEcat	_CENSOR_	Survival	Censored	ARn	Y_Value	riskname	Y_risk	X_risk
Dizziness	Drug A	0	0		0		61	-0.26	Subjects at Risk	-0.15	8
Dizziness	Drug B	1	5	1	0.01818	0.01818	52	-0.34			
Dizziness	Drug B	31	35	1	0.05720	0.05720	2	-0.34			
……											

In STEP statements, three variables – TTE (time to event), survival, and treat will be used to draw the step plot. TTE actually is equal to AESTDY. If a subject has the same AE for multiple times, only the first occurrence will be picked. If a subject does not have the plotted AE, the TTE will be calculated between the RFSTDTC and date of EOS (End of Study). The variable survival will be calculated by TTE and censor from the LIFETEST procedure. If a subject has a Plotted AE, it means this subject had an event (i.e. censor=1), otherwise, he/she had a censor (i.e. censor=0). The variable treat contains the treatment group values.

```
STEP X=TTE  Y=survival /GROUP= treat  LINEATTRS=(PATTERN=1 );
```

If the subject has a censor, then variable censored will be equal to variable survival. Variable censored and variable TTE will be used in the SCATTER statements to draw the "Censored" scatter plot.

```
SCATTER X=TTE  Y=censored /GROUP= treat  MARKERATTRS=(SYMBOL=plus );
```

To add the words "Subjects at Risk", you need to add three variables to the dataset – riskname, x_risk, and y_ risk. Riskname has "Subjects at Risk" only. X_ risk and y_ risk are numeric variables to specify the position of "Subjects at Risk" in the figure.

```
SCATTER X=x_risk Y=y_risk/ MARKERCHAR=riskname;
```

Beneath the words "Subjects at Risk", you need another SCATTER statement (below) to display the number of subjects at risk by treatment group. To do so, you have to create three variables – TTEcat (the category of time to event), Y_Value (values on Y axis), and ARn (the number of subjects at each category of time to event).

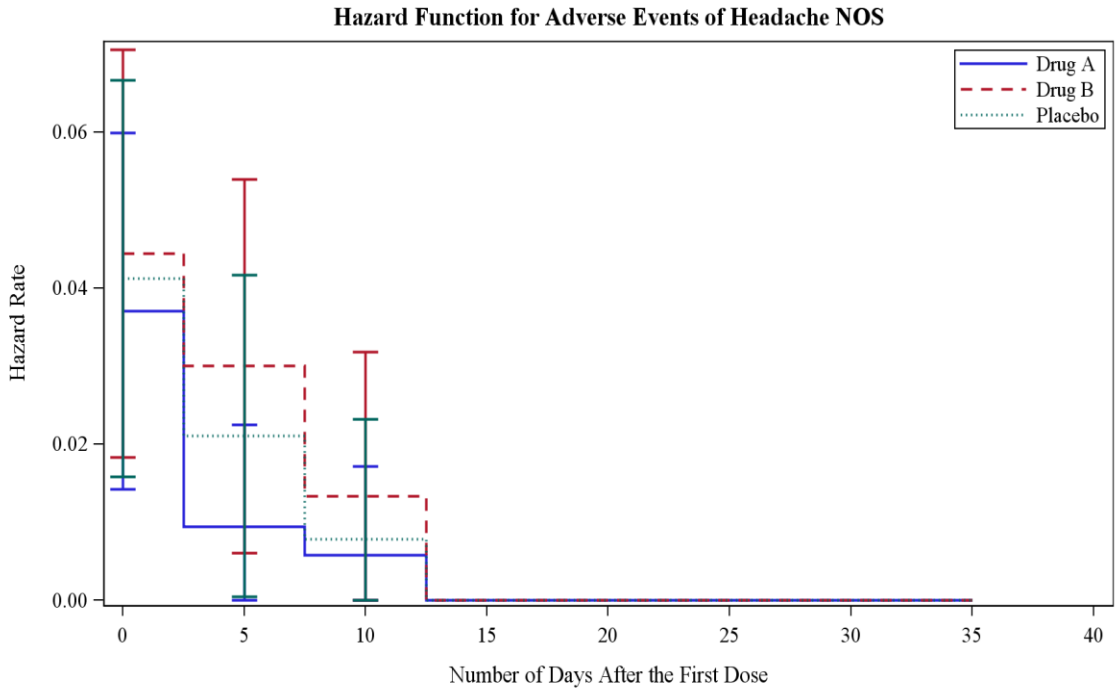
```
SCATTER X= TTEcat  Y= Y_Value / GROUP= treat  MARKERCHAR=ARn ;
```

TTEcat – category value (i.e. 10, 20, 30 days, etc.) of the continuous variable TTE

Y_Value – numeric values (i.e. 0.25, 0.5, etc.) assigned by you for each treatment group based on your figure.

2). Hazard Function

The hazard function gives the instantaneous rate of occurrence of the adverse event during the treatment period. Displaying the hazard function requires a relative simple dataset, which can be obtained from the LIFETEST procedure as well.



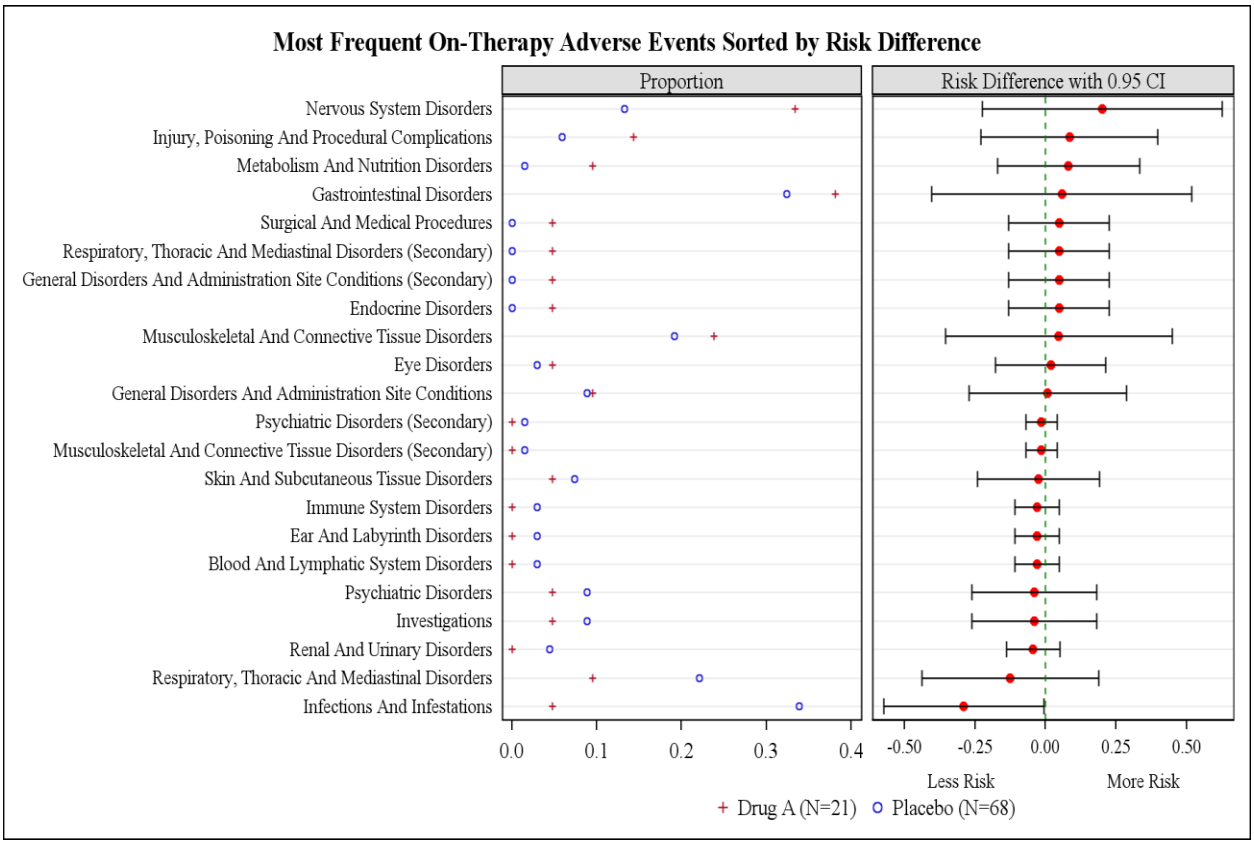
TTE	A	ALow	AHigh	B	BLow	BHigh	C	CLow	CHigh
0	0.037037	0.01418	0.059894	0.044444	0.018343	0.070546	0.041237	0.015815	0.066659
5	0.009412	0	0.022452	0.03	0.006063	0.053937	0.021053	0.00045	0.041655
10	0.005797	0	0.017158	0.013333	0	0.031802	0.007843	0	0.023212
15	0			0			0		
20	0			0			0		
25	0			0			0		
30	0			0			0		
35	0			0			0		
40									

Only STEPPLOT statements are needed in the TEMPLATE Procedure if you do not want to show the number of subjects at risk. Otherwise, you need to use the BLOCKPLOT statements to add the number of subjects at risk like the Product-Limit method mentioned above.

```
STEPPLOT X=TTE  Y=A /ERRORUPPER=AHigh ERRORLOWER=ALow  JUSTIFY=CENTER
                    LINEATTRS=GRAPHDATA1(PATTERN=SOLID  THICKNESS=2)
                    ERRORBARATTRS= GRAPHDATA1(PATTERN=SOLID  THICKNESS=2)
                    LEGENDLABEL='Drug A';
```

6. Forest Plot

The figure below is a typical Forest plot, which usually consists of two panels – the one on the left displays the incidence by treatment group, and the one on the right shows the relative risk of an AE on the active arm relative to the placebo arm with a 95% confidence interval.



The dataset for this type figures looks like the following:

AEBODSYS	Pcnt_T	Pcnt_B	factor	LCL	UCL	mean
Infections And Infestations	0.0476	0.3382	0.144712	-0.574	-0.007	-0.291
Investigations	0.0476	0.0882	0.113319	-0.263	0.1815	-0.041
Psychiatric Disorders	0.0476	0.0882	0.113319	-0.263	0.1815	-0.041
……						

The left panel is almost identical to the one discussed in section 4 above except that you use variables AEBODSYS to display at Y axis.

To fill out the right panel, you need to calculate the mean value (mean), the lower limit (LCL) and the upper limit (UCL) of 95% confidence interval for the predicted probability.

Pcnt_T – equal to the count of unique subjects in the particular body system or organ class divided by the count of total unique subjects in the treatment group (NT)

Pcnt_B – equal to the count of unique subjects in the particular body system or organ class divided by the count of total unique subjects in placebo (NB)

Factor = 1.96*sqrt(Pcnt_T*(1- Pcnt_T)/NT + Pcnt_B*(1- Pcnt_B)/NB);

LCL = (Pcnt_T - Pcnt_B) – 1.96*factor;

UCL = (Pcnt_T - Pcnt_B) + 1.96*factor;

mean = 0.5*(LCL + UCL);

A SCATTERPLOT statement and a REFERTENCE statement will be used to draw the right panel.

```
SCATTERPLOT Y= AEBODSYS X=mean / XERRORLOWER=LCL  XERRORUPPER=UCL
                MARKERATTRS=(SYMBOL=circlefilled size=8  color=red);
REFERENCELINE X=0 /LINEATTRS=graphdatadefault(PATTERN=shortdash  color= green);
```

CONCLUSION

SAS Graph Template Language and SAS/GRAPH SG Procedures have made significant improvement over the traditional SAS Graph Language for graph generation. Visualizing drug safety data is more and more common across the entire clinical research field. The seven types of figures discussed in this paper are the most commonly used for displaying adverse events data. It is very important to understand the targeted/expected figure, identify the right variables, and then determine the final structure of the input dataset. This is a reverse approach but the best way to build the right input dataset.

ACKNOWLEDGEMENTS

Special thanks to Thomas Li, PhD, for his help on time-to-event plots for adverse events, and Teng-Chiao Chu, PhD, for his help on hazard function plot.

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 Adverse Events Data 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
Endo Pharmaceuticals
1400 Atwater Dr.
Malvern, PA 19355
Email: shu.amos@endo.com

TRADEMARK INFORMATION