

Multidimensional Data Visualization for Generating Patient Profile

Rajavel Ganesan, Merck and Co., Upper Gwynedd, PA

Suhas R. Sanjee, Merck and Co., Upper Gwynedd, PA

Jason Clark, Merck and Co., Upper Gwynedd, PA

ABSTRACT

Traditionally, patient profile information is displayed in multiple tables and graphs in a disconnected fashion. A methodology is presented here to enable multidimensional data visualization of a patient's profile. Data from multiple SDTM domains can be presented graphically without compromising any details by adding text annotations wherever needed. These profiles can be very useful in evaluating a patient's safety and disposition. It can also help with medical monitoring and serve as reference for medical writing. Medical Monitoring is an on-going consolidated review conducted by clinical in life by reviewing the safety, efficacy, and other relevant data in an effort to assess for specific trends, outliers, and to ensure protocol compliance. We have effectively used this approach in oncology therapeutic area. We are proposing an approach in which the tables and graphs are presented in a composite form which enhances user friendliness especially vis-à-vis non-quantitatively oriented clinical personnel who are the main consumers of this resource. This way, the clinical personnel can look at information about all important domains at one place without having to look at multiple tables and graphs. Additionally, this approach has the distinct advantage of using SDTM datasets which makes it portable across protocols.

INTRODUCTION

Patient profile includes data from different domains. A sample patient profile is shown in Figure 1. This section will provide more details on the methodology and brief explanation of various Base SAS and SAS/Graph procedures used.

A patient's adverse experiences, laboratory tests, study medication, disposition, death and related scans can be displayed in the graph. Adverse experiences can be limited to specific intensity or toxicity and laboratory tests are user specified.

The composite patient profile approach has been especially effective in Oncology area in the context of all important AE toxicity grades and tumor scans.

ANNOTATE FACILITY

The Annotate facility enables generation of a special data set of graphics commands from which you can produce graphics output. This data set is referred to as an Annotate data set. It can be used to generate graphical annotations and to enhance graphics output from many SAS/GRAPH procedures, including GCHART, GCONTOUR, GMAP, GPLOT, GPRINT, GRADAR, GSLIDE, and G3D. Annotate facility can be used to generate line and text annotations on a graph generated by SAS graph procedures at desired location.

GPLOT PROCEDURE

The GPLOT procedure plots two or more variables in the input data set on a set of coordinate axes. The coordinates of each point on the plot correspond to two variable values in an observation of the input data set. In conjunction with the SYMBOL statement the GPLOT procedure can produce join plots, high-low plots, needle plots, and plots with simple or spline-interpolated lines. The SYMBOL statement can also display regression lines on scatter plots. The syntax for GPLOT is as follows

```
PROC GPLOT <DATA=input-data-set>  
<ANNOTATE=Annotate-data-set>  
<GOUT=<libref.>output-catalog>  
<IMAGEMAP=output-data-set >  
<UNIFORM>;  
RUN;
```

The GPLOT procedure is used with ANNOTATE= option to generate the composite patient profile.

SAS PROGRAMMING OVERVIEW

The basic programming approach is to keep it as simple as possible. The GPLOT procedure is used to generate the patient profile. X axis has the relative study day. The lines corresponding to start and end of AE, text corresponding to lab tests names / values, AEs etc. are displayed using the graph annotate facility. The annotate facility, being data driven allows for greater flexibility. Each observation in the annotated dataset contains complete instructions for drawing a graphic or moving to a position to draw a graphic. The value of the FUNCTION variable determines the behavior of the observation. The values of the coordinate variables X and Y combined with the values of the coordinate system variables HSYS, XSYS, and YSYS specifies the location. The values of the attribute variables STYLE, COLOR, TEXT, POSITION, and SIZE specifies the attributes.

METHODOLOGY:

Read and combine data from different domains (AE, LB, EX, DS and RS). RS is the response domain that has tumor scans results. Use parameters to make the macro flexible. A flag is created to differentiate between different domains and an ordering variable is created to help with the alignment of the plots.

1) Create annotated datasets

AE annotated dataset

- Get toxicity grade, preferred term and relative study day of start and relative study day of end for each AE
- Horizontal lines representing the start and end relative study day of each AE. These lines are drawn using MOVE and DRAW functions. The LABEL function is used to represent the AE term. POSITION is a character variable used to control placement and alignment of a text string. Note that Relative study day is on the X axis.
- Different colors are assigned based on the toxicity grade. System variable LINE specifies the line type.
- The MOVE and DRAW functions are used to draw lines that separate different parts of the graph.
- The table below shows how the annotated dataset for AE looks like.

FUNCTION	COLOR	XSYS	YSYS	SIZE	AEDECOD	AESTDY	AEENDY	AETOXGR	AENUM
label	black	2	2	1	Anorexia	71	87	1	32
move	black	2	2	1	Anorexia	71	87	1	32
draw	maroon	2	2	1	Anorexia	71	87	1	32
move	maroon	2	2	1	Anorexia	71	87	1	32
draw	black	2	2	1	Anorexia	71	87	1	32

DOMAIN	AN	SITEID	SITE	ECOG	X	Y	TEXT	POSITION	LINE
AE	0001	(0040)	Smith Medical Center	0	-38	32	Anorexia	6	
AE	0001	(0040)	Smith Medical Center	0	71	32	Anorexia	6	
AE	0001	(0040)	Smith Medical Center	0	87	32	Anorexia	6	
AE	0001	(0040)	Smith Medical Center	0	-40	30	Anorexia	6	
AE	0001	(0040)	Smith Medical Center	0	95	30	Anorexia	6	1

Lab, Disposition, Study Medication and Scans Annotated Dataset

- Get lab test, lab result, disposition status, Study medication, all applicable scans and the related relative study days from corresponding domains.
- The MOVE and LABEL functions are used to represent the actual values of the result aligned with the relative study.
- For the labs, the abnormal results are shown in red.

- The table below shows how the data is setup for Study medication.

FUNCTION	COLOR	XSYS	YSYS	SIZE	AENUM	DOMAIN	EXTRT	EXSTDY
label	black	2	2	1	13	ex	Trt C	1
move	black	2	2	1	13	ex	Trt C	1
label	black	2	2	1	13	ex	Trt C	1
move	black	2	2	1	13	ex	Trt C	1
draw	black	2	2	1	13	ex	Trt C	1

EXDOSTOT	EXDOS	AN	SITEID	SITE	ECOG	X	Y	TEXT	POSITION	LINE
400	400	0001	(0040)	Smith Medical Center	0	-38	13	Trt C	6	
400	400	0001	(0040)	Smith Medical Center	0	1	13	Trt C	6	
400	400	0001	(0040)	Smith Medical Center	0	1	13	400	6	
400	400	0001	(0040)	Smith Medical Center	0	-40	11	400	6	
400	400	0001	(0040)	Smith Medical Center	0	95	11	400	6	1

- 2) Annotated dataset created in step 2 is passed on to the PROC GPLOT to produce the output. A BY clause is added to display one patient per page. To add legends for the colors used in the horizontal lines and the disposition status, a LEGEND option is added to the PROC GPLOT. Sample output is shown in Figure 1.
- 3) By using the DEVICE=PDFC in the GOPTIONS statement, the graph is directly created in PDF format. Choosing the "Recognize Text using OCR" feature in Adobe Acrobat the graph can be made searchable. This feature can be accessed by selecting on the "Document" and then "OCR Text Recognition".
- 4) Another exciting feature is the ability to create a compound document i.e., graphs with hyperlinks to provide drill down into details while not cluttering the main graph. This is especially significant if lot of information needs to be presented in a single graph. Hyperlink also affords the ability to invoke spreadsheets, documents and other graphs. This is accomplished by using the DEVICE=HTML in the GOPTIONS statement and GANNO procedure, the graph can be embedded in a HTML document with drill down feature.

GENERATION OF PATIENT PROFILE

Figure 1 gives composite graphic and table related information (lab values, AE, dosing and disposition etc.) for a patient over time. The graphic component corresponding to AE portion (at the top) shows the lines representing the start and end of different AEs (specified on the left). Colors of lines represent toxicity grades (refer to the legend; bottom left). The lab portion of the output has the lab values taken on relative study days represented on X axis.

SEARCHABLE PDF

The GPLOT procedure in conjunction with annotate facility is used to generate patient profiles (Figure 1). X axis represents relative day of visit. The lines representing AE start and stop dates are displayed using line annotations. Laboratory values, dosing schedule and disposition information are displayed as text annotations.

GRAPHS WITH HYPERLINKS

The GANNO procedure is used to generate the patient profile graph with the drill down capability. An additional variable (HTML) has to be added to the annotated dataset to specify hyperlink. In the figure below, "AE Duration" is the text on the graph associated with the hyperlink to the MS WORD document *AE.doc*

FUNCTION	COLOR	TEXT	YSYS	SIZE	HTML
label	black	AE Duration	2	1	href=AE.doc
label	black	Vomitting	2	1	href=VOM.xls

Patient Profile

AN=0001 SITEID=(0040) SITE=Smith Medical Center ECOG=0

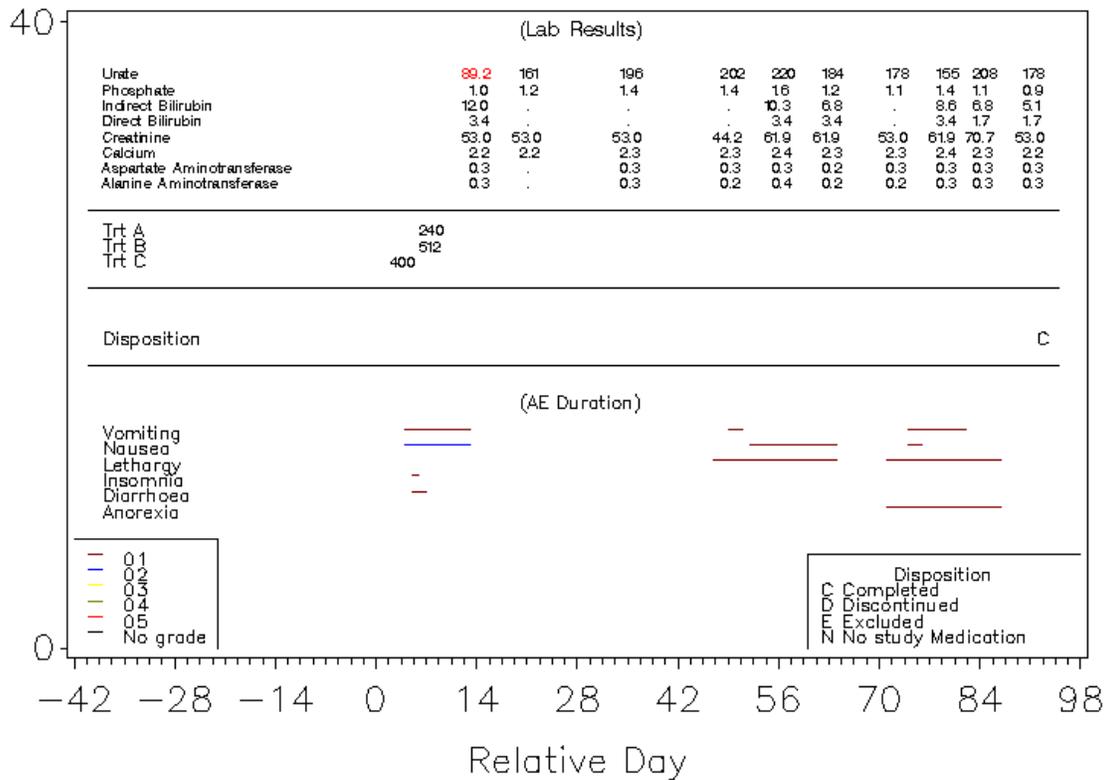


Figure 1: Composite Patient Profile

CONCLUSION AND SUMMARY

A methodology to generate patient profile containing information from multiple domains is presented. Other possibilities include hyperlinks to tables for drill down into more details.

ACKNOWLEDGEMENTS

We would like to thank our manager Mary Varughese, Associate Director, Merck & Co., Inc. for her encouragement and support. We also would like to thank Gopal Rajagopal and Sekhar Bhamidipati of Merck & Co., Inc. for their valuable suggestions.

Author Name: Rajavel Ganesan
 Title : Senior Scientific Programmer
 Company : Merck & Co., Inc.
 Address : 351 Sumneytown Pike
 City: North Wales State: PA Zip: 19454
 Email: rajavel_ganesan@merck.com

Author Name: Suhas R. Sanjee
 Title : Scientific Programming Analyst
 Company : Merck & Co., Inc.
 Address : 351 Sumneytown Pike
 City: North Wales State: PA Zip: 19454
 Email: suhas_sanjee@merck.com

Author Name: Jason Clark, PhD
 Title : Associate Director (Statistics)
 Company : Merck & Co., Inc.
 Address : 351 Sumneytown Pike
 City: North Wales State: PA Zip: 19454
 Email: jason_clark4@merck.com