

Data Driven Plots in Clinical Research... What Next?

George A. Nestory, ASG, Inc., Cary, NC

ABSTRACT

Analyzing the effect a drug has on an efficacy variable (e.g. liver function or blood parameter level) is one of the key components in the design of a drug study. Reporting the response to treatment a drug has on this variable over time is an important factor in the evaluation of a clinical trial. A basic X/Y plot of the values with respect to time is an effective way to represent this data. Without much effort, text, labels, legends, and other 'bells and whistles' can increase the usefulness of an X/Y plot. This paper demonstrates how to manipulate the data with the use of annotate options and macros, satisfying the whims of our graphics-loving clinical brethren.

INTRODUCTION

Using the GPLOT procedure is an easy way to respresent data values that are taken over a period of time. This basic X/Y plot can become more useful if additional information is included on the page. Including reference lines to display the criteria used to analyze whether the subject responded to treatment or not can make the plot a more useful tool. From the plot it will be easy to visualize the baseline criteria, treatment duration, outlying values, interval of response, and the day of relapse. A detailed legend can describe the many lines that are displayed on the plot. Additionally placing a bar graph beneath the x-axis of the plot can be an alternate means of displaying the same information as the plot giving an overview of the analysis for the particular subject. Displaying all of this information on one graph can effectively be accomplished through the use of SAS/GRAPH® and the annotation facility.

Most of the graph will be driven by the data itself. A few of the reference lines are constant across the subjects in the analysis and can be hardcoded within the program. The rest of the display will be created through the GPLOT procedure and use of annotate data sets. Two annotate data sets will create the bar graph and

label the outlying values in the plot.

A stepwise approach was used beginning with the basic plot. The reference lines, annotations, and legend were then added.

DATA

The analysis data set consists of 5 variables and contains the values from the liver function parameter SGOT. SUBJ will hold the number of the patient in the study, 3 patients total in this example. STUDYDAY holds the actual day with respect to the first day of treatment (baseline) that the SGOT reading was taken. For example, a STUDYDAY value of -75 is a reading taken 75 days prior to the first day of treatment. SGOTVAL is the value of the SGOT reading. TRTEND will hold the last day of treatment per subject. FLAG will hold the key readings from the analysis of the data and can have 3 possible values aside from missing. Values of 'RESPBEG' and 'RESPEND' will flag the start and end of the response interval, if any, and 'RELAPSE' will flag the day of relapse, if any. Table 1 will illustrate an example of the data for subject 3003. This example shows the flags for the response interval and relapse:

TABLE 1: Example Data Set GROUPS

SUBJ	STUDYDAY	SGOTVAL	TRTEND	FLAG
3003	-75	3.20	133	
3003	-12	2.10	133	
3003	0	2.65	133	
3003	7	3.10	133	
3003	14	2.60	133	
3003	21	2.20	133	
3003	28	1.45	133	
3003	35	1.30	133	
3003	42	1.00	133	RESPBEG
3003	49	0.80	133	
3003	56	0.60	133	
3003	84	0.70	133	
3003	112	1.00	133	RESPEND
3003	140	2.30	133	
3003	168	0.50	133	
3003	196	1.80	133	RELAPSE
3003	224	2.10	133	

CODING

There were many details to keep in mind when preparing the plot. Several reference lines were needed with a detailed legend describing these lines along with the SGOT plot. Outlier values were to be labeled on the graph. A bar graph displaying all the data in the above plot to be placed beneath the x-axis, relevant to the x-axis scale. This required a systematic approach.

X/Y Plot of SGOT Values

This was accomplished straight from the GROUPS data set plotting SGOTVAL by STUDYDAY.

Reference Lines

Since all reference lines were to be included in the legend a key variable GROUP was created needed to make each plotted line unique. Therefore, each reference line required values for STUDYDAY and SGOTVAL. For the reference lines that were constant across subjects, STUDYDAY and SGOTVAL did not need to be data dependent. One example is the reference for day=0. Using GROUP as the key variable, the code to create this line follows:

```
data ref;
  set ...;
  by subj;
  if first.subj; ** plot once per subj ;
  ** vertical line at day=0 ;
  do i=0,3;
    studyday=0;
    sgotval=i;
    group=1;
    output;
  end;
```

For the reference lines that were data dependent the same code is used within the same data step, but the dependent variable is substituted for studyday. One example is a reference line at treatment end using TRTEND as the dependent variable creating another GROUP variable:

```
** vertical line at end of treatment ;
do i=0,3;
  studyday=trtend;
  sgotval=i;
  group=2;
```

```
output;
end;
```

Note that in order to plot these lines as reference lines I used the symbol for 'point' in the symbol statement. Therefore, no readable symbol will be created when plotting the line.

Outlier SGOT values

The maximum value to be plotted vertically is 3. All values greater than 3 are plotted at the maximum vertical value but the point is labeled with the actual value. This requires an annotate data set (OUTLIER) to be included in the GPLOT statement.

```
data outlier;
  length function style $ 8 text $ 4;
  set ...;
  retain hsys '1' xsys ysys '2';
  if sgotval gt 3 then do;
    function='label';
    x=studyday;
    y=3;
    position='8'; ** place value below point ;
    text=left(put(sgotval,5.2));
    style='simplex';
    output;
  end;
```

In order to plot these values at y=3 the original data needed to be altered after this annotate data set was created. Also, the group variable for the SGOT plot is created.

```
data sgot(rename=(_val=sgotval));
  set sgot;
  if sgotval gt 3 then _val=3;
  else _val=sgotval;
  group=6;
  drop sgotval;
```

This data set is appended to the REF data set containing the reference lines and used as the input data set for the GPLOT procedure. This final data set is named GROUPS.

Legend

The detailed legend is straightforward except for one distinction. It will display the plotted lines by their lines types as opposed to the plotted symbols. This uses the GROUP variable in the

GPLOT statement to automatically create the legend. The legend, which is placed in the upper inside corner of the graph, will have a frame drawn and will be shadowed (only one line definition is displayed).

```
legend1 across=1;
position=(top right inside);
mode=share;
frame;
cblock=black;
offset=(-5,);
label=none;
shape=line(5);
value=(tick=1 justify=l
' SGOT BASELINE AVE (studyday=0)'
...);
```

Bar Beneath Plot

Placing the bar graph beneath the x-axis is accomplished with the use of the built in annotation macros. To compile and prepare these macros for use the following statement is required:

```
%annomac;
```

Separate annotation data sets are created to create the bar (created in sections) and lines corresponding to the end of treatment and relapse. To create the bar sections 2 situations need to be addressed; blank sections for treatment with no response and a solid section displaying the response interval. The main points to consider are the 2 end points of the bar for each section. If a response follows a section without response the end point of the blank section needs to correspond with the start of the solid section. The same is true if a blank section follows a response. Therefore, when the annotate data set is output the separate sections will connect to create one bar. This processing is accomplished in a previous data step and the variables are named BEGDAY and ENDDAY. The annotate macro used to create the bar is %BAR. The horizontal section of the bar uses the absolute value of STUDYDAY, the vertical section will use the percentage of the output area. This is set by the XSYS and YSYS definitions. Using the ORIGIN statement in the AXIS2 definitions and moving the graph up 15 percent space is provided below the graph for the bar. The FLAG variable is used to key whether

solid or empty sections need to be created.

```
...
retain xsys '2' ysys '3';
if upcase(left(trim(flag)))='RESPEND' then do;
  * fill for response ;
  %bar(begday,0,endday,5,gray4c,0,r2);
end;
if upcase(left(trim(flag))) ne 'RESPEND'
then do;
  * empty otherwise ;
  %bar(begday,0,endday,gray4c,0,empty);
end;
```

The bar will have a height of 5 percent of the output area.

The reference lines within the bar are created with the annotation macros %DRAW and %LINE. This creates another annotate data set.

```
retain xsys '2' ysys hsys '3';
if last.subj then do;
  %move(trtend,0); * point to begin line ;
  %draw(trtend,5,gray4c,3,.25); * draw line ;
end;
if upcase(left(trim(flag)))='RELAPSE' then do;
  %move(studyday,0);
  %draw(studyday,5,gray4c,1,.25);
end;
```

Final Steps

Once the annotate data sets are created for the outliers and bar graph they need to be concatenated with each other to create one data set. A GOPTIONS statements is created to set up graphics parameters and the offshadow for the legend. Some important options are defined:

```
goptoins norotate offshadow=(.5,.5) nocharacters
nocell gunit=pct gsfmodes=replace...;
```

OFFSHADOW builds the shadow around the legend box in the plot. The entire GPLOT statement will include the axis definitions, the legend, and the complete annotate data set.

```
proc gplot data=groups;
by subj;
plot sgotval*studyday=group / vaxis=axis1
haxis=axis2 hminor=3 anno=extras
legend=legend1;
run;
```

This creates a separate plot for each subject in the analysis dataset. The option GSFMODE=APPEND will prevent succeeding plots from overwriting previously created plots. Figure 1 illustrates the complete display of annotated data. Outliers are potted and labelled, the subject has both a response interval and a relapse, which are displayed in both the plot and the corresponding bar graph. Figure 2 is a listing using PROC GREPLAY that displays all 3 subjects. Subject 3003 is included from Figure 1. Subject 1001 has no flags to display, but the bar graph shows the length of treatment and the end of treatment window (end of treatment plus 7 days). Subject 2002 has a response without a relapse. These 3 subjects display all possible situations that are created from the one program.

CONCLUSION

Basic plots of values over time are used throughout the pharmaceutical industry. It is a way to display data that is easy to comprehend. Through the use of the annotation facility these graphs can become more meaningful and useful in the analyzation of clinical data by displaying a wide range of important information on one page.

REFERENCES

SAS Institute, Inc. (1990), *SAS/GRAPH Software Reference Volumes 1+2, Version 6, First Edition*, Cary, NC: SAS Institute Inc.

ACKNOWLEDGEMENTS

I would like to thank Stephen Black for his assistance in developing this application.

For Further Contact

The author can be reached for comments or questions by mail at:

George A. Nestory
2 Sabo Court
Elmwood Park, NJ 07407.

By phone: (201) 503-5742
Or through ASG: 1(800) 279-8082

Bye-mail: nestory.george@ehccgate.sandoz.com

SAS/GRAPH is a registered trademark of SAS Institute Inc. in the USA and other countries. ® Indicates USA registration.

Figure 1

SESUG '95 Proceedings

RESPONSE TO TREATMENT (SGOT over Time) SUBJECT = 3003

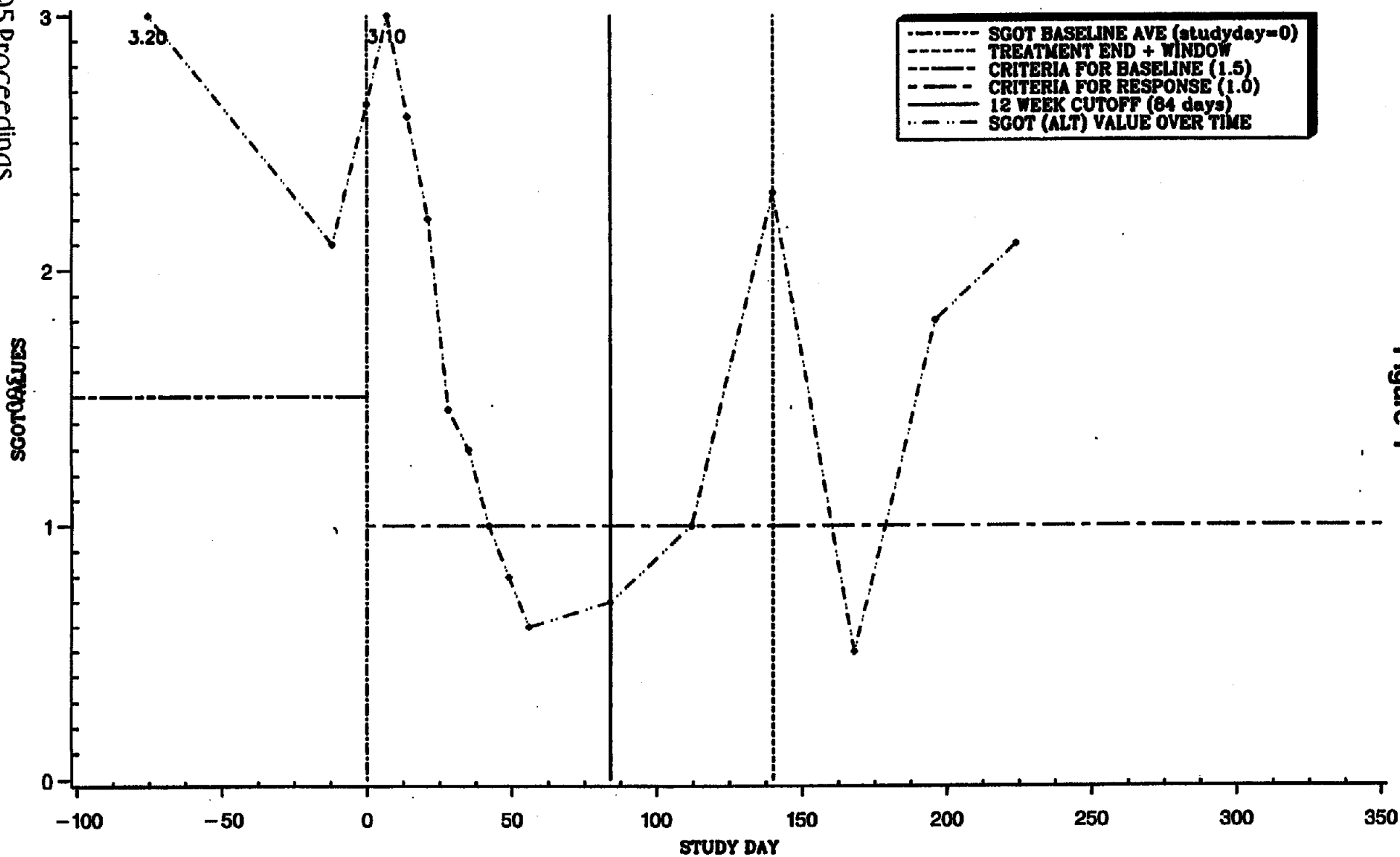


Figure2

