**GRAPHICAL PRESENTATION OF CLINICAL DATA USING SAS MACROS**

Pan-Yu Lai, The Upjohn Company

Thomas E. Schneider, The Upjohn Company

**ABSTRACT**

The adage, "a picture is worth a thousand words" is especially true when graphically displaying clinical safety data such as vital signs, laboratory assays, and ECGs. Since clinical abnormalities generally depend on normal ranges and/or the magnitudes of changes from baseline values, a data plot which is able to clearly reveal these kinds of information and also label the "abnormal patients" by their identification numbers is very helpful to medical reviewers. This paper presents a SAS macro which performs this task on an X-Y plot using the X-axis for baseline values and the Y-axis for follow-up values. This macro automatically extracts the necessary information from the input data set to produce the desired graphs. It also handles points with missing baseline values without excluding these points or distorting the graphs.

**INTRODUCTION**

Laboratory tests and vital signs of a clinical trial are not only used to screen patients before a trial but also to protect patients and to assess drug effects during the trial. An assay's result is said to be abnormal if it falls outside the predetermined normal range and/or has a significant change from its baseline value. Patients with abnormal values in any assays during or after the trial should be identified, studied, and discussed in the report. The magnitudes of changes from baseline values are often used to justify the drug effect on an assay.

For these purposes, a SAS macro was developed to produce an X-Y plot for each assay to help physicians review the whole data set in one glance. This well-designed X-Y plot reveals a lot of information which are imbedded within the data and are hard to be shown by tabulations and listings. Reviewers can then easily identify outliers and abnormal values, and evaluate the drug effect.

**OUTPUT**

The output of the macro is a two-dimensional plot using the abscissa (X-axis) for baseline values and the ordinate (Y-axis) for follow-up values. Therefore, each baseline, followup pair of a patient constitutes a point in the graph. A 45 degree (diagonal) line is also included to reflect the direction and magnitude of the changes from the baseline. For example, a point below the diagonal line indicates a decrease from the baseline value. The identification number of the patient who has the abnormal value is printed beside the point so that reviewers can associate this abnormal value with other assays, concomitant therapy, and medical events of the same patient. The ID is to be printed at the right hand side of the corresponding point if possible. When there is no space at the right hand side, it is printed either at the left hand side, above the point, below the point, or searching continues all the way down until a space is found.

Patients with only a baseline value are not included in the plot because baseline values alone do not provide information on drug effect. But patients with only a follow-up value are included in the plot because abnormal follow-up may be drug related and should be presented. They are plotted along a dashed vertical line appearing on the right hand side of the graph.

A sample graph with all of the features mentioned above is given in Figure 1. It is a graphic presentation of supine diastolic blood pressure of patients in study XXXXXX conducted by Dr. A. Four drugs were used in the study.

Another interesting feature of the program is that it can automatically kick extreme outliers out of the X-Y plot and print the information about these points in a box which appears in the lower right hand corner of the page. Figure 2 shows how the outlier can ruin the plot while Figure 3 is the plot produced by our macro which automatically detects the outlier and moves it outside the X-Y plot region.

**INPUT DATA**

Each observation should contain both baseline and follow-up values to make up a point in the X-Y plot. Therefore, in addition to BY variables, the data set should contain the following 11 variables (lower case indicates that the variable name could be changed by means of passing a parameter to the macro):

- **BASELINE**: baseline value
- **FOLLOWUP**: follow-up value
- **PROT**: protocol number to identify the study
- **pat**: patient identification number
- **med**: medication code
- **qname**: name of assay
- **OUTLIER**: =1 if the patient's ID needs to be printed

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BASELINE</td>
<td>baseline value</td>
</tr>
<tr>
<td>FOLLOWUP</td>
<td>follow-up value</td>
</tr>
<tr>
<td>PROT</td>
<td>protocol number to identify the study</td>
</tr>
<tr>
<td>pat</td>
<td>patient identification number</td>
</tr>
<tr>
<td>med</td>
<td>medication code</td>
</tr>
<tr>
<td>qname</td>
<td>name of assay</td>
</tr>
<tr>
<td>OUTLIER</td>
<td>=1 if the patient's ID needs to be printed</td>
</tr>
</tbody>
</table>
LOW, HIGH: investigator's normal range for that assay
LOW, OHIGH: clinical normal range

We allow two sets of normal ranges to be included in the program because most of the time the normal ranges given by the investigators are different from the ranges used to flag the clinical abnormalities.

PARAMETERS

Users should pass some variable names and options to the macro through seven key-word parameters while calling the macro. They are:

**INDATA** = name of data set to be plotted

**PAT** = macro variable containing name for patient ID in the input data set. The default is **PAT**.

**QNAME** = macro variable containing the name of the assay variable in the input data set. The default is **QNAME**.

**MED** = macro variable containing the name of the medication code in the input data set. The default is **MED**.

**BYVARS** = list of BY variables. In order to produce a better output, users can specify a format for each BY variable. For example, **BYVARS=INV<INV.> LOATE SEX<$SEX.>** requests the macro to produce a plot for each INV*LOATE*SEX*QNAME combination with the formats INV. and $SEX. associated with INV and SEX, respectively.

**FIGURE** = If this is left blank, no figure number will be printed. If SEQ is used, consecutive figure numbers will be printed for the plots starting from the number given by the parameter **FIGCNT**. If a format name is used, then this format will be associated with the variable name given by the parameter **FIGCNT**.

**FIGCNT** = If this is blank, no number is printed. If **FIGURE=SEQ** then numbering will begin with the value of **FIGCNT**. If a variable name is used the value of the variable is printed

LOGIC

The program starts with parameter initialization and symbol definitions and then goes into a big %DO loop which contains the major portion of the macro. Each time the loop is executed, the first group of observations from the remaining input data set are grabbed to extract information and the extreme outliers are removed into another data set. The information on these outliers will then be printed in the lower right hand corner of the plot by using annotation facilities. The program then goes through this sub data set once more to generate ad hoc observations for printing abnormal patients, normal range boxes and the diagonal line.

Sometimes abnormal points are very close to each other causing their IDs to be partially or completely overlapped. In this case, some IDs need to be moved around to avoid the overlapping. The first version of this program used PROC MATRIX to handle this problem. But because PROC MATRIX has several disadvantages (heavier use of CPU time and memory, less familiarity with PROC MATRIX by programmers and the replacement of PROC MATRIX by a product independent of the base SAS system) we switched to using a DATA step approach. The idea is to grid the whole page of the plot into an 80 by 80 matrix by setting up an array of 80 elements with each element of the array being an 80-byte character variable. The observations which give the coordinates of patient IDs with abnormal points are then filtered through these variables. If something is to be printed in a square of the graph, the corresponding byte in the character variable is marked. If an ID happens to fall into squares which are already marked, the program will search for unmarked squares over the surrounding area and then put this ID into the first available squares.

After all the information is extracted, properly modified, and stored in macro variables or moved into the annotation data set, PROC GPLOT is then used to produce the desired output graphs.

CONCLUSION

There are many forms in which we summarize and present data, such as tables, listings, and graphs. Because of their ability to show trends, directions and outliers of data sets at a glance, graphs are often preferred over other forms for presenting data. The X-Y plot generated by our macro not only plots the patients' actual data but also depicts the magnitudes of changes from baseline values. It clearly identifies both abnormal patients and extreme outliers without cluttering or overscaling. This program has proven to be very useful in presenting clinical laboratory and vital sign data, especially for monitors and FDA reviewers.

SAS is a registered trademark of SAS Institute Inc., Cary, NC, USA
Figure 1

SUPINE DIASTOLIC BLOOD PRESSURE -- Baseline vs. Follow-up
PROTOCOL XXXXXX -- INVESTIGATOR DR. A

Patients without baseline appeared on dashed vertical line
Patients without baseline appeared on dashed vertical line

Patients without baseline appeared on dashed vertical line