THE USE OF SAS/GRAPH™ SOFTWARE IN THE MANAGEMENT OF CLINICAL TRIALS

Mary Anne Hope
VA Cooperative Studies Program

ABSTRACT

At the Veterans Affairs Cooperative Studies Program Coordinating Center graphics are important tools in monitoring and presenting data from clinical trials. SAS/GRAPH™ is used to display a large amount of longitudinal data covering all phases of the trial from the first protocol to the final publications. Customized reports are created with the Annotate facility and the procedures available through SAS® and SAS/GRAPH software. A wide variety of graphic techniques and applications including line graphs, bar charts, and whisker plots are used to summarize and present data from the clinical trial on patient intake, drug compliance, primary endpoints and secondary endpoints. In addition, maps and pie charts are used to display data for administrative purposes including cost factors and participating sites for various studies.

INTRODUCTION

The Veterans Affairs Cooperative Studies Program manages large-scale longitudinal multicenter clinical trials. The responsibility of the center is to direct the planning, follow-up and analysis phases of studies. Clinical trials occur in diversified medical fields including cardiology, infectious disease, dentistry and gastroenterology.

Graphic presentations are an important aspect in all phases of the study. In the early phases of the planning and protocol development of the line graphs and charts are used to display the statistical power of various sample sizes and follow-up schedules. During the follow-up phase extensive reports are generated to allow a Data Monitoring Board Committee to review the progress and intermediate results. These reports are created on a bi-annual basis to track and monitor the trends of the data. The intent is to present complex longitudinal data in a concise, understandable and perhaps artistic form.

APPLICATIONS

To reduce the number of tabular presentations and to enhance interest and readability for the reviewers line graphs, bar charts, scatterplots and whisker plots are included in the report. In the analysis phase of the study, graphics are used to convey the primary and secondary endpoint results in slide or poster presentations and medical journal publications.

The first step is to create a SAS database for all of the key variables needed to produce the initially designed graphics for the trial. The database variables are efficiently used as input to a number of different statistical procedures as PROC FREQ, PROC MEANS, PROC SUMMARY and PROC LIFETEST. Secondly, the SAS/GRAPH procedure can easily use SAS database variables or statistical output variables from procedures to create an extensive collection of graphics. These graphics may be enhanced with the Annotate facility and repeated for numerous variables by implementing the Macro facility. This flexibility and power of the SAS language allows proficient production of customized pictures at all phases of the trial. The use of SAS/GRAPH is illustrated for applications in the following figures. Also included are portions of the Annotate and/or data step code used to accomplish the graphs.
visual picture of the progress is provided. The statistical variables created from a PROC FREQ output data set are the input data for the PROC GPLOT line graph. The Annotate facility enables the descriptive information, labels and arrows to be included on the graph for further clarification.

PROC GCHART is used in Figure 2 to compare overall compliance between treatments. Drug adherence is an important issue addressed in a trial and can be a complex variable to measure and present to reviewers. The histogram is an effective tool to show the exact numbers or percent of a multilevel variable within groups or subgroups. PROC GCHART allows considerable control of the bars pattern, width and depth descriptions. Also the legend options facilitate explanation of the variables chosen for analysis.

The survival analysis in Figure 3 shows the treatment effect on the predefined primary endpoint. Survival in the clinical trial is measured as the time from randomization (i.e. entry into the study) to a predefined endpoint such as death or heart attack. Patient withdrawal or lost-to-follow up are also considered endpoint events to calculate a patient's time in the study. A large number of cases over a long period of time can produce a tabular presentation that is lengthy and complex, especially when using the Kaplan-Meier method.

The survival time in this graph is presented as its inverse cumulative mortality. A minimal amount of SAS coding is required to create the two variables for the two treatment arms because of the straightforward use of SAS variables as input.

An Annotate data set enhances the plot by labelling the lines and placing text displaying the number of patients and statistical significance at different survival intervals. The Annotate facility allows text to be placed anywhere on the page by specifying the definition of the XSYS and YSYS area and coordinate system. Annotate functions were used to label the survival curves because a legend is not generated with the overlay option.

Figures 4 and 5 show the serial changes in a variable over the course of the study and also the relationship between two treatments at a given time in the study. These secondary endpoint variables can be the occurrence of certain events or a clinical measurement.

The whisker plots (Figure 4) presented here allow quick visual comparisons of the mean differences between group distributions. The whiskers show the standard error of each mean data point for each evaluated interval. The input data for the graphs were produced from the PROC SUMMARY procedure. The program written in Macro language generates a series of plots for related analysis variables. PROC GLOT is used to construct the graphs and PROC GSLIDE is utilized for the title on the bottom mean change plot. Rather than hardcode the values of the length of the whiskers, Annotate functions were used to calculate and draw the length from the SAS variables saved by the PROC SUMMARY procedure. Also the legend is placed outside of the axes using a XSYS and YSYS area definition.

The scatter plot (Figure 5) effectively shows the correlation between two distributions at a given point in time in the trial. The variables from a previously created SAS data base are used for direct input to the PROC GPLOT procedure. The overlay option is used to plot the regression line and the upper and lower 95th confidence level lines. Annotate statements were used for placement of the sample size text on the graph to complete the picture.

The whisker plots (Figure 4) presented here allow quick visual comparisons of the mean differences between group distributions. The whiskers show the standard error of each mean data point for each evaluated interval. The input data for the graphs were produced from the PROC SUMMARY procedure. The program written in Macro language generates a series of plots for related analysis variables. PROC GLOT is used to construct the graphs and PROC GSLIDE is utilized for the title on the bottom mean change plot. Rather than hardcode the values of the length of the whiskers, Annotate functions were used to calculate and draw the length from the SAS variables saved by the PROC SUMMARY procedure. Also the legend is placed outside of the axes using a XSYS and YSYS area definition.

The scatter plot (Figure 5) effectively shows the correlation between two distributions at a given point in time in the trial. The variables from a previously created SAS data base are used for direct input to the PROC GPLOT procedure. The overlay option is used to plot the regression line and the upper and lower 95th confidence level lines. Annotate statements were used for placement of the sample size text on the graph to complete the picture.

SAS/GRAPH is also useful in managing other areas of a clinical trial. In Figure 6 a pie chart was created to display the distribution of the parts of the variable in relation to the whole. PROC GCHART was generated to present a breakdown of the types of trials conducted at the Coordinating Center.

Figure 7, a map of the United States with specific cities denoted is used as a visual tool for easy identification of the sites participating in the trial. The special SAS data sets associated with PROC GMAP and also the use of an annotate data set quickly produces the map for a slide or poster presentation.
ADVANTAGES AND DISADVANTAGES

The large volume of data in a clinical trial is managed easily and effectively with SAS and SAS/GRAPH software. Other graphic packages previously used to create similar graphs had required either conversion of SAS data to ASCII format or manual entry of numbers from printed SAS output. Using SAS databases and SAS/GRAPH procedure has eliminated the conversion step by direct input of SAS variables either from analysis files or statistical output variables from procedures. This saves production time and eliminates any manual entry error.

Usually a large number of cases are needed to demonstrate the relationship of the variables in a scatterplot. In SAS/GRAPH these scatterplots can be generated for variables from very large databases and save production time by plotting the graphs quickly.

The power of the SAS Macro facility enables code to be stacked for repeated production runs. Similar plots generated for different analysis variables are created by macro variable substitution for the X and Y axes. This allows for easy generation of a multitude of graphs which comprise a great proportion of the bi-annual reports produced for each study.

The Annotate facility allows the user to compose a special presentation by adding important statistical, legends or lines as whiskers to a graph. Full page manipulation is provided with the Annotate facility to create detailed graphs.

The improved capabilities of SAS/GRAPH over the years has also involved an increase in the complexity of the code written to achieve the desired visual picture. The use of Macro language and annotate instructions can simplify the production of multiple graphs for different sets of variables but also be troublesome in revising and debugging the detailed coding.

CONCLUSION

In the future, the goal is to develop more generalized plotting routines to be used across studies. Also, the development of Annotate macros to streamline coding for generalized programs is being explored. The SAS/GRAPH graphics editor available in Release 6.07 has the ability to modify graphics output interactively. The ability to incorporate one graph into another and edit will allow greater power to create more complex customized graphics interactively. Other possibilities for future development in the management of clinical trials are three-dimensional graphics and the utilization of more color in report generation.

SAS and SAS/GRAPH are registered trademarks of the SAS Institute, Inc., Cary, NC, USA
Figure 1

PATIENT INTAKE GRAPH
(Line Plot)

Figure 2

Monitoring Drug Compliance
(Vertical bar chart)

Figure 3

MORTALITY GRAPH
(Line plot)
Figure 4

**Figure 4**

**MEAN OF VARIABLE 1**

<table>
<thead>
<tr>
<th>MONTH</th>
<th>VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
</tr>
<tr>
<td>3</td>
<td>30</td>
</tr>
<tr>
<td>4</td>
<td>40</td>
</tr>
<tr>
<td>5</td>
<td>50</td>
</tr>
</tbody>
</table>

**MEAN CHANGE IN VARIABLE 1**

<table>
<thead>
<tr>
<th>MONTH</th>
<th>VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
</tr>
<tr>
<td>3</td>
<td>30</td>
</tr>
<tr>
<td>4</td>
<td>40</td>
</tr>
<tr>
<td>5</td>
<td>50</td>
</tr>
</tbody>
</table>

![Figure 4 Diagram](image)

**Figure 5**

**SECONDARY ENDPOINT VARIABLE**

(Scatter plot with Regression line)

![Figure 5 Diagram](image)
Figure 6

Distribution of Active and Non-active Studies
(Fig chart)

Figure 7

LOCATION OF PARTICIPATING SITES
(Map of United States)