Plotting Overall Mean and Individual Values in One Graph
Jenny Zhang, ICON Clinical Research, Inc. North Wales, PA

ABSTRACT

Using the SAS® PROC GPLOT procedure to display multiple plots onto one graph has been proven a very effective data presentation tool. One particular application of this kind of tool is often seen in graphically summarizing pharmacokinetic (PK) data. Obtaining PK data through PK studies is a key early step in the new drug development process. It is a quantitative measure of drug exposure that is needed for the sound interpretation of non-clinical efficacy. In PK studies, the drug (chemical) concentration values from individual subjects as well as summarized overall subject mean concentration by certain time points are often required to be plotted together for better observation. Since the GPLOT procedure can only open one dataset whereas in practice, the overall mean and individual values typically exist in different datasets, some programming intervention is needed for conveniently plot the data. This paper provides the basic introduction and a sample SAS program to solve this technical problem.

INTRODUCTION

In the pharmaceutical industry, we often use the GPLOT procedure to display the mean concentration by sample time (hour) for PK studies. The GPLOT procedure plots the values of two or more variables on a set of coordinate axes (X and Y). By default, the plot symbols in a plot are not connected. To emphasize the mean in the data, one can connect the plot symbols by using INTERPOL=join (i=join) in the SYMBOL statement. It is relatively straightforward to show mean concentration values in one graph. Three topics are discussed in detail: (1) PROC GPLOT for overall mean concentration (2) An approach combining overall subject mean concentration and individual subject concentration values in one graph, and (3) Programming tips.

PROC GPLOT FOR OVERALL MEAN

Below is the partial printout of the initial dataset (DS1) for illustration:

Dataset 1(Dataset DS1): Subject Concentration Data by Sample Time (HR)

<table>
<thead>
<tr>
<th>Obs</th>
<th>SUBJID</th>
<th>CONC</th>
<th>HR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>01001</td>
<td>0.000</td>
<td>0.0</td>
</tr>
<tr>
<td>2</td>
<td>01001</td>
<td>1.330</td>
<td>0.5</td>
</tr>
<tr>
<td>3</td>
<td>01001</td>
<td>3.990</td>
<td>1.0</td>
</tr>
<tr>
<td>4</td>
<td>01001</td>
<td>4.650</td>
<td>1.5</td>
</tr>
<tr>
<td>5</td>
<td>01001</td>
<td>4.800</td>
<td>2.0</td>
</tr>
<tr>
<td>6</td>
<td>01001</td>
<td>4.480</td>
<td>2.5</td>
</tr>
<tr>
<td>7</td>
<td>01001</td>
<td>3.620</td>
<td>3.0</td>
</tr>
<tr>
<td>8</td>
<td>01001</td>
<td>3.560</td>
<td>3.5</td>
</tr>
<tr>
<td>9</td>
<td>01001</td>
<td>3.550</td>
<td>4.0</td>
</tr>
<tr>
<td>10</td>
<td>01001</td>
<td>2.800</td>
<td>5.0</td>
</tr>
<tr>
<td>11</td>
<td>01001</td>
<td>1.940</td>
<td>6.0</td>
</tr>
<tr>
<td>12</td>
<td>01001</td>
<td>2.560</td>
<td>8.0</td>
</tr>
<tr>
<td>13</td>
<td>01001</td>
<td>1.580</td>
<td>10.0</td>
</tr>
<tr>
<td>14</td>
<td>01001</td>
<td>1.060</td>
<td>12.0</td>
</tr>
<tr>
<td>15</td>
<td>01001</td>
<td>0.000</td>
<td>18.0</td>
</tr>
<tr>
<td>16</td>
<td>01001</td>
<td>0.000</td>
<td>24.0</td>
</tr>
</tbody>
</table>
This example dataset lists 2 subjects’ concentration information for sample times 0 – 72.0 hours.

```
proc sort data=ds1;
  by hr;
run;

proc means data=ds1 noprint;
  by hr;
  var conc;
  output out=stat mean=concmean;
run;
```

Dataset 2(Dataset STAT): Overall Subject Mean Concentration (CONCMEAN) By Sample Time (HR)

<table>
<thead>
<tr>
<th>Obs</th>
<th>HR</th>
<th>CONCMEAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0</td>
<td>0.00000</td>
</tr>
<tr>
<td>2</td>
<td>0.5</td>
<td>3.50250</td>
</tr>
<tr>
<td>3</td>
<td>1.0</td>
<td>4.48688</td>
</tr>
<tr>
<td>4</td>
<td>1.5</td>
<td>4.76250</td>
</tr>
<tr>
<td>5</td>
<td>2.0</td>
<td>5.07063</td>
</tr>
<tr>
<td>6</td>
<td>2.5</td>
<td>4.88875</td>
</tr>
<tr>
<td>7</td>
<td>3.0</td>
<td>4.47438</td>
</tr>
<tr>
<td>8</td>
<td>3.5</td>
<td>4.35313</td>
</tr>
<tr>
<td>9</td>
<td>4.0</td>
<td>4.27750</td>
</tr>
<tr>
<td>10</td>
<td>5.0</td>
<td>3.83750</td>
</tr>
<tr>
<td>11</td>
<td>6.0</td>
<td>2.94188</td>
</tr>
<tr>
<td>12</td>
<td>8.0</td>
<td>2.13250</td>
</tr>
<tr>
<td>13</td>
<td>10.0</td>
<td>1.51563</td>
</tr>
<tr>
<td>14</td>
<td>12.0</td>
<td>1.26169</td>
</tr>
<tr>
<td>15</td>
<td>18.0</td>
<td>0.89306</td>
</tr>
<tr>
<td>16</td>
<td>24.0</td>
<td>0.63600</td>
</tr>
<tr>
<td>17</td>
<td>36.0</td>
<td>0.00000</td>
</tr>
<tr>
<td>18</td>
<td>48.0</td>
<td>0.00000</td>
</tr>
<tr>
<td>19</td>
<td>72.0</td>
<td>0.00000</td>
</tr>
</tbody>
</table>
For summarizing drug concentration information from DS1, PROC MEANS was used to calculate the mean concentration for 16 subjects by sample time. Typically, each variable value on the horizontal axis (HR) has a corresponding value on the vertical axis (CONCMEAN)).

goptions
reset = all
device = PSLAND
gsfmode = replace
gsfname = _lstperm
hsize = 9.00 in
horigin = 0.00 in
vsize = 6.50 in
vorigin = 0.00 in

axis1 value=(font=swissb justify=right h=0.3000 cm)
   offset=(2 pct,)
   minor=none
   order=(0 to 6 by 1)
   label=(font=swissb angle=90 h=0.4800 cm "Mean Concentration (ng/mL)");

axis2 value=(font=swissb h=0.2800 cm)
   label=(font=swissb h=0.2800 cm "Time (hr)"
   offset=(2 pct,)
   minor=none
   order=(0 to 72 by 6 );

symbol1 height=1.2 i=join font=marker, value=C color=black width=1.2;

The OFFSET = (2 pct) option places the first major tick mark 2 units from the origin and the last major tick mark 2 units from the end of the axis. (The unit of measurement is a percent of the display area).

The order = option specifies the data values in the order that they are to appear on the axis. For AXIS1, it represents a mean concentration range from 0 ng/mL to 6 ng/mL. For AXIS2, it represents the time range from 0 hr to 72 hr.

In the SYMBOL statement, by default, no font is specified, and the symbol specified by the VALUE = option is taken from the special symbol table. The ‘marker’ font and ‘C’ value were selected to produce the triangles shown in Figure 1.

proc gplot data=stat;
   plot concmean * hr /frame nolegend vaxis=axis1 haxis=axis2 ;
run;

The HAXIS= option in the plot statement specifies the AXIS2 definition. The VAXIS= option in the plot statement specifies the AXIS1 definition.
Figure 1: Plots of Two Variables (Mean Concentration by Sample Time)

You can see it is not difficult to use GPLOT procedure to produce two-dimensional graphs that plot one dependent variable against an independent variable within a set of coordinate axes. We used one SYMBOL statement to join the points.

**PLOTTING OVERALL SUBJECT MEAN CONCENTRATION AND INDIVIDUAL SUBJECT CONCENTRATION VALUES IN ONE GRAPH**

In some situations, we need to show the overall mean concentrations as well as individual subject concentrations in one graph. However, the GPlot procedure can only open one dataset at a time. How can we use the SAS® GPlot procedure to show the overall means and individual values in one graph?

The solution is to merge two data sets prior to GPlot. The following sample program merges 2 data sets (DS1 and STAT) by the variable HR.

```sas
data ds2;
  merge ds1 stat;
  by hr;
run;

proc sort data=ds2;
  by hr subjid;
run;
```
Four variables (SUBJECT, CONC, HR, and CONCMEAN) were merged into one dataset (DS2).

If your data contain two sets of values for the same type of data, you can use the PLOT2 statement to generate a right axis that is calibrated the same as the left axis. PLOT and PLOT2 statements generate multiple graphs. You can specify plot requests of the type vertical (CONC)*horizontal (HR) = third variable (SUBJECT) in PLOT2 statements to produce multiple plots on one graph.

axis1 value=(font=swissb justify=right h=0.3000 cm) offset=(2 pct,) minor=none order=(0 to 11 by 1) label=(font=swissb angle=90 h=0.4800 cm "Concentration (ng/mL)"

axis2 value=(font=swissb h=0.2800 cm) label=(font=swissb h=0.4800 cm "Time (hr)"

axis3 label=none order=(0 to 11 by 1) value=none major=none minor=none;

 AXIS, LEGEND, and SYMBOL definitions can be used with the PLOT2 statement in the same way that they are used with the PLOT statement.

R specifies the number of times that a symbol definition is applied before the next SYMBOL definition is used. By default, R=1. R=16 Indicates “Repeat = Number of Times”.

We would need 16 SYMBOL statements if we didn’t use R=16 in the SYMBOL2 statement. The code would be as follows:

symbol1 l=1 i=join font=marker, value=C color=blue width=1.2;
symbol2 l=1 i=join height=1.2 value=none color=black width=1.2 r=16;

symbol12 l=1 i=join height=1.2 value=none color=black width=1.2 ;
symbol13 l=1 i=join height=1.2 value=none color=black width=1.2 ;
...
symbol17 l=1 i=join height=1.2 value=none color=black width=1.2 ;

Finally, we produce the multiple graphs using the code below.

proc gplot data=ds2;
  plot conc * hr = subjid/frame nolegend vaxis=axis1 haxis=axis2;
  plot2 concmean * hr /frame nolegend vaxis=axis3 haxis=axis2 ;
run;
quit;

The VAXIS = option specifies AXIS1 and AXIS3 definition in the PLOT and PLOT2 statements, respectively.

The PLOT2 statement generates a right axis (AXIS3) that displays the values in a same scale (0-11) from the scale used in the PLOT statement for the left axis (AXIS1).
Figure 2: Overall Mean Concentrations and Individual Concentrations in One Graph

PROGRAMMING TIPS

(1) Merging two datasets prior to PROC GPLOT

Careful data preparation is an important step. After you obtain the mean PK concentration for each time point HR (that is, the mean across subjects for each PK sample time), use HR to merge the mean PK concentration data back with the original PK data so each subject record contains the value of the mean concentration for that time point, as well. Check the resulting, merged data for problems.

(2) Using the PLOT2 statement

You can use the same type of plot request with a PLOT2 statement that you can with a PLOT statement, but a PLOT2 statement always plots the values of the dependent variable on the right vertical axis.

(3) Requesting plots that assign a SYMBOL definition
In the example, SYMBOL2 corresponds to the PLOT2 statement. The “R=16” option on the statement SYMBOL2 will repeat the SYMPOLE2 statement 16 times. If there are fewer than 16 values for the SUBJECT variable, the PLOT2 statement will not display all the subjects (16 subjects) in this graph.

CONCLUSION

The GPLOT procedure can only open one dataset. Overall subject concentration means and individual concentration values typically exist in different datasets. In order to plot the individual values and overall summary values together, merging the original dataset and the dataset that contains the summarized values (such as means) by a key variable is a required step to prepare for such a plot. Understanding the concept of “R=XX” is also critical.

REFERENCE

ACKNOWLEDGMENT
SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.

CONTACT INFORMATION
Jenny Zhang
ICON Clinical Research, Inc.
1700 Pennbrook Parkway
North Wales, PA 19454
215-616-3142
Email: zangh@iconus.com