Paper CC14
Single Page Graphical Display of Multiple Subjects' Time Pattern on Studies

Louie Huang, Baxter Healthcare Cooperation, Baxter BioScience, Westlake Village, California 91362, USA

ABSTRACT

SAS/GRAPH® is a very powerful tool that allows you to efficiently visualize your clinical data. This presentation introduces some useful skills that create a meaningful and self-explanatory patient profile displaying study discontinuation information of several studies for multiple subjects. The paper provides a concrete programming example for programmers’ reference in their clinical application. The techniques include creating efficient macro variables with PROC SQL, creating and using formats, annotation, and the presenter’s own tips and tricks.

INTRODUCTION

It is easy to create a basic graphical output using procedures, options, graphical options and statements of SAS/GRAPH®. The GPLOT and GCHART procedures are common tools to generate scatter plots and bar charts for displays of clinical data. However, it is a bit challenging to generate informative and demanded graphs (Figure 1). It is not possible to generate this kind of graphs without some special tricks and tips. The objective of this paper is to demonstrate how to use the PROC GLOT, tricks in symbol and axis statements, Annotate Facility, formats and macro variables to achieve the programming task.

Figure 1. Subject Study Path
MANIPULATION OF GRAPH APPEARANCE

Figure 1 is straightforward, efficient, informative and compact. The whole process is
displayed by several graphical steps in order to facilitate the description. The SAS code
in Appendix I is numbered for your convenience.

A layout of subject identifications over study time is presented in Figure 2. The plot
presented in Figure 2 is absolutely undesirable. The undesirable features are:
the symbols and curves for subject over study date do not make any sense; the actual
subject Ids are not displayed along the y-axis; and the vertical display of study date on the
x-axis takes too much space from the plot.

Figure 2. Scatter plot created by PROC GPLOT

The following modifications make the plot look a little nicer (Figure 3):
1. The visible lines and symbols were made to invisible by manipulating the symbol
   statements.

   ```
   symbol1 color=white;
   symbol2 color=white;
   ```

   What you need to do is changing the original color (red and blue in this case) to white or
the background color of your graph. The purposes of creating a plot as shown in Figure 3
are to line up the subject identifications along the y-axis by total study days in descending
order, and create an empty space for superimposing the time pattern on the plot by
annotation.
2. A dummy variable `subnum` is created by a short data step (line 29 to 34, APPENDIX I). This variable contains a unique numeric number for each subject ID. Unique subject Identifications are converted to a macro variable `&sublist2` (line 15, APPENDIX I) by a SQL procedure. A format value `subj` is created by a FORMAT procedure (line 83 to 92, APPENDIX I). Variable `subnum` is formatted back to subject Identifications in the GPLOT procedure (line 104, APPENDIX I).

3. In order to horizontally display the study dates a `date7` format is used to format the study dates in Figure 3.

Figure 3. Displays of subject identifications and study dates

![Figure 3](image)

After the undesirable symbols and lines are made to disappear, the time patterns by subject and study date are able to clearly display in the plot by annotation. An annotate data set is created by a few lines of SAS code (line 50 to 61, APPENDIX I). There are three studies in this case. In Figure 4, different studies are differentiated with distinct colors. Now, the study time patterns are unique to each subject, and comparable between subjects. The total study days are labeled at the end of the last study in order to make the plot informative. This is accomplished by an annotate data set (line 36 to 49, APPENDIX I, Figure 5). Finally, a legend for study names is created by annotation (line 55 to 82, APPENDIX I, Figure 6).

Generation of Figure 1 is automated by 8 macro variables, which are created by a SQL procedure (line 8 to 28, APPENDIX I). The values of these macros are listed in Table 1.
Figure 4. Display of time pattern by annotation

Figure 5. Time patterns labeled with total study days
Figure 6. Time patterns labeled with total study days and study names

Table 1. List of Macro Variables

<table>
<thead>
<tr>
<th>Macro name</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>&amp;sublist</td>
<td>090226 090227 020226 040226 100226 010226 080226 070226 100227 010228 010227 030226 060226 110226</td>
</tr>
<tr>
<td>&amp;sublist2</td>
<td>&quot;090226&quot; &quot;090227&quot; &quot;020226&quot; &quot;040226&quot; &quot;100226&quot; &quot;010226&quot; &quot;080226&quot; &quot;070226&quot; &quot;100227&quot; &quot;010228&quot; &quot;010227&quot; &quot;030226&quot; &quot;060226&quot; &quot;110226&quot;</td>
</tr>
<tr>
<td>&amp;totday</td>
<td>418 418 376 352 349 313 291 217 124 108 92 92 92 91</td>
</tr>
<tr>
<td>&amp;mdtlist</td>
<td>16467 16467 16354 16406 16397 16399 16249 16416 16512 16381 16087 16225 16176 16389</td>
</tr>
<tr>
<td>&amp;cnt</td>
<td>14</td>
</tr>
<tr>
<td>&amp;div</td>
<td>70</td>
</tr>
<tr>
<td>&amp;lovalue</td>
<td>15890</td>
</tr>
<tr>
<td>&amp;upvalue</td>
<td>16520</td>
</tr>
</tbody>
</table>

INTERPRETATION OF MACRO VARIABLES

&sublist: list of unique subject identifications.
&sublist2: list of unique subject identifications, each quoted by a double quotation mark.
&totday: list of total study days by subject.
&mdtlist: list of numeric values of last study date by subject.
&cnt: counts of unique subjects.

The values of macro variables &sublist, &sublist2, &totday and &mdtlist are arranged in a descending order of total study days. Macro variables &sublist and &cnt function to correctly assign a numeric number to the dummy variable subnum in a Data Step (line 29 to 34, APPENDIX I). Macro variable &sublist2 is used in FORMAT procedure to correctly return the format value fmt to subject identifications (line 83 to 92, APPENDIX I). Macro variable &totday is used to annotate the total study day for each subject, and
macro variable &mdtlist is used to specify the position of annotated text at the end of the last study for each subject (line 36 to 49, APPENDIX I). Macro variable &div determines the interval value between two major tick marks. This macro value is calculated by the following formula:

\[
\text{ceil}((\text{max}(\text{date\_out}) - \text{min}(\text{date\_in}))/8)
\]

Macro variables &lovaue and &upvalue determine the start and end major tick marks, respectively. These two macro values are calculated by the following formulae:

\[
\begin{align*}
\text{min}(\text{date\_in}) - \text{mod}(\text{min}(\text{date\_in}),&\text{div}) \\
\text{max}(\text{date\_out}) - \text{mod}(\text{max}(\text{date\_out}),&\text{div}) + &\text{div}
\end{align*}
\]

CONCLUSION

The above interpretation and discussion conclude that SAS/GRAPH offers great flexibility and power for display of clinical data. This paper demonstrates how to create an efficient and informative graph to track and compare time patterns for multiple subjects on several studies. Hopefully, the skills offered in this paper will provide some helpful hints for your work.

CONTACT INFORMATION

The author appreciates any comments, questions, and suggestions for improvement.

Louie Huang
Baxter Healthcare Corporation, Baxter BioScience
One Baxter Way
Westlake Village, CA 91362
Tel: 805-372-3487
Fax: 805-372-3462
Email: louie_huang@baxter.com

Disclaimer

The contents of this paper are purely the work of the author and do not necessarily represent the opinions, or recommendations of Baxter BioScience. 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.
APPENDIX I: SAS Code

1. libname sugi "C:\huang\temp\sugi31\saedate";

2. data time;
3. set sugi.time(keep=subjid part date_in date_out);
4. where part ne .;
5. days=date_out-date_in;
6. if days ne .;
7. run;

8. proc sql noprint;
9. select subjid
10. ,quote(put(subjid,z6.))
11. ,totdays
12. ,date_out
13. ,count(subjid)
14. into :sublist separated by ' ' /* For creating annotate data set */
15. ,sublist2 separated by ' ' /* For formatting Subject IDs in the graph */
16. ,:totdays separated by ' ' /* For labelling total study days by subject */
17. ,:mdtlist separated by ' ' /* the latest study date by subject */
18. ,,:cnt                        /* number of subjects */
19. from
20. (select subjid,sum(days) as totdays,max(date_out) as date_out
21. from
22. time
23. group by 1)
24. order by totdays desc,subjid;

25. select (max(date_out)-min(date_in))/8-mod((max(date_out)-min(date_in))/8,10)+10
26. into :div /* interval value of major tick-marks */
27. from time;
28. select min(date_in)
29. ,max(date_out)-mod((max(date_out)-min(date_in))/&div)+&div
30. into :itval /* first tickmark value of the x-axis */
31. ,:ltval /* last tickmark value of the x-axis */
32. from time;
33. quit;

34. data sorted(drop=i); /*Insert a numeric variable (subject ID) into the dataset */
35. set sugi.time(keep=subjid part date_in date_out);
36. length color $8.;
37. retain xsys ysys '2' hsys '1' when 'a' position '6' size 3 function 'label' color 'blue';
38. %macro lbt;
39. %do i=1 %to &cnt;
40. %let mdt=%scan(&mdtlist,&i);
41. %let days=%scan(&totday,&i);
42. %let num=%scan(&sublist,&i);
43. %end;
44. %put subjid: &num &mdt &days;
45. %end;
46. %stop;
47. %stop;
48. %stop;
43. x=mdt+6; y=i;
44. text=catx( ' ', "days", '\', days );
45. output;
46. %end;
47. %mend;
48. %lbt
49. run;

50. data _anot(drop=subnum date_in date_out); /* draw bars for each study part */
51. length color $6. ;
52. retain size 1.5 xys yys '2' hys '1';
53. set sorted(keep=subjid subnum part date_in date_out);
54. by subnum part;
55. function='move'; x=date_in; y=subnum; output;
56. function='draw'; x=date_out; y=subnum;
57. if part=1 then color='red';
58. else if part=2 then color='black';
59. else color='blue';
60. output;
61. run;

62. data anot;
63. set anot _anot;
64. run;

65. %annomac;
66. data anot2; /* Assign a label to each study part */
67. %dlanno
68. xys='1'; yys='1'; hys='1'; when='a';
69. x=93; function='label'; color='black'; size=3.5; style='"TimesRoman"';
70. y=96.5; text='Study 1'; output;
71. y=92.5; text='Study 2'; output;
72. y=88.5; text='Study 3'; output;

73. /* Assign a distinct color to each study */
74. %line(82,96.5,88,96.5,red,1,1)
75. %line(82,92.5,88,92.5,black,1,1)
76. %line(82,88.5,88,88.5,blue,1,1)

77. /* draw a legend box */
78. %line(81,85.5,81,99,green,1,0.2)
79. %line(81,85.5,97.5,85.5,green,1,0.2)
80. %line(97.5,85.5,97.5,99,green,1,0.2)
81. %line(97.5,99,81,99,green,1,0.2)
82. run;

83. proc format; /* format the assigned subject number to subject ID */
84. value subj
85. %macro fmt;
86. %do i=1 %to &cnt;
87. &i=%scan(&sublist2,&i)
88. %end;
89. ;
90. %mend;
91. %fmt
92. run;
93. goptions reset=global gunit=pct cback=white 
    hsize=8 in vsize=5.2 in 
    htitle=4.5 htext=2 
    ctext=black 
    colors=(blue black green red) 
    ftext="TimesRomanBold" ftitle="TimesRomanBold";

94. filename grafout "C:\huang\temp\profile.cgm";
95. goptions device=cgmofmp gsfname=grafout gsfmode=replace;

96. title 'Subject Study Path';
97. symbol1 color=white interpol=none line=1 
    width=2 value=dot height=0.01;

98. axis1 label=(h=3.5 color=black 'Study Date')
99. order=(&ftmval to &ltmval by &div)
100. width=4 length=88 
101. major=(color=black height=2) offset=(1,8);

102. axis2 label=(h=3.5 angle=90 color=black 'Subject')
    order=(1 to 14 by 1)
    width=4 
    major=(color=black height=0.8) length=75 offset=(4);

103. proc gplot data=sorted anno=anot2;
104. format date_out date7. subnum subj.;
105. plot subnum*date_out / anno=anot noframe 
    haxis=axis1 hminor=9 
    vaxis=axis2 vminor=0 
    caxis=black nolegend;

106. run;
    quit;