Using SAS® Macro and SAS/GRAPH® To Produce Data Driven Laboratory Data Plots

ABSTRACT
Data driven programming is a method of programming where the values of the data "drive" or create the SAS code. This technique is illustrated using SAS macro to read a dataset and create code to produce customized reports and plots. The data driven macro produces SAS code to subset data, sort data, create annotate datasets, create proc statements, create formats, create titles and create footnotes. This code is created and executed for each similar subgroup of data contained in the dataset. Source code is provided.

INTRODUCTION
The review and analysis of laboratory data is an important part of most clinical studies. In the past the Sterling Winthrop Clinical Biostatistics (CB) Department's programmers have written SAS programs to produce customized reports and plots for each Clinical Study Report (CSR). The process had to be streamlined to shorten the time needed to produce a CSR. A set of reports and plots that are used in most CSR's was identified and a flexible data driven macro was written to produce each of these data displays. The term "data driven" means the macro reads the data and creates individual PROC statements with specific titles, footnotes, labels, formats, plot axes and by group for each laboratory test. This method of programming produces customized tables and plots for each CSR. This paper will discuss one of these macros.

MACROS
CB staff identified reports and plots which are commonly used in a CSR. Macros were created to produce these data displays. The macros use common input parameters such as dataset name, optional BY variables, format name for test and first title number. Additionally, each macro contains input parameters which are specific to that macro. I will discuss one of these, the %LAB4 macro.

The %LAB4 macro produces a plot of post value versus baseline value with a solid line representing no change. Dotted lines representing increases or decreases from baseline which are a specified percent (eg. 40%) of the reference range are optional. Figures 1 - 4 are the four possible plots which the macro will produce.

DATA DRIVEN PROGRAMMING
Data driven programming is not a new idea, the concept is to read the dataset and produce code which incorporates values and limits of the data. This approach works well with laboratory data from clinical studies for two reasons. First, no two studies are alike, dose regimens are unique to the study and laboratory data is collected at different time points and frequencies depending on the study requirements. Secondly, each study can contain up to 100 different laboratory tests with varying ranges of data. These two factors combine in creating a dataset containing many dissimilar subgroups of data.

This macro incorporates data driven code creation by reading the dataset and creating PROC PLOT code for each laboratory test. The code includes a TITLE statement containing the laboratory test name and reference range, different footnote statements depending on options used, axis statements which create equal distant integer labeled axes, formatted LEGEND statement containing labeled dose information and annotated lines for displaying ranges. This approach allows us to create plots tailored specifically to each study and laboratory test.

INPUT DATASET
The macros run using a standard input dataset. This dataset is similar to the raw data set, so a programmer can create it in a few hours. An important factor in allowing these macros to work on many different types of studies is including the dose sort order (dose), dose format (dose_fmt), time sort order (time) and the time format (time_fmt) variables in the dataset. These variables are used in producing the sorting order, axis, formats and labels. Additional variables which can be used as optional BY variables in the macros can be added to the dataset. The dataset structure is listed below and Figure 5 is an example of the dataset.
MACRO LOGIC

This section will discuss the logic used in the macro.

The logic for all the macros is similar so reviewing one should provide an understanding of the algorithms used.

The %LAB4 macro contains three steps. Step one produces the format to label the doses in the PROC GPLOT legend. The input data set is used to produce an input control data set which PROC FORMAT then uses to create formats. Step two utilizes the input data set to create a set of macro variables containing information which is specific to each laboratory test. These macro variables are then used in step three. Step three is a call to %PLOTIT macro for each laboratory test.

The %PLOTIT macro produces and runs PROC GPLOT code for the individual laboratory test. %PLOTIT uses the macro values passed in from %LAB4 to create PROC GPLOT code with axes, titles and footnotes tailored specifically for a laboratory test. The first step is to subset the input data set for one test. Minimum and maximum values for the test result and baseline are calculated and then used to create equal length horizontal and vertical axes. An algorithm was developed to calculate a best fit integer axis. An annotate dataset is created to produce the line of no change and the dashed lines representing a specified percent of the reference range shift. Titles and footnotes tailored to the laboratory test description and units are produced. Finally the PROC PLOT code is created with by variables and axes tailored to the data.

MACRO CODE

```sas
%macro lab4(in=,by=,test_fmt=,cutoff=,percent=,title=);
  .......
  %macro plotit(code,parm,unlt,shift,low,high,range);
  .......
```

```sas
proc sort data=&in out=a; by &by sub;
  %if &percent=NO %then %do;
    where test ne &code and unit ne &unlt and time>&cutoff;
  %end;
  %if &percent ne NO %then %do;
    where test = &code and low = &low and high = &high and unit ne &unlt and time>&cutoff;
  %end;
```

```sas
  data _null_; set b;
    maxx=max(maxb,maxp);
    minx=min(minb,minp);
    rra=maxx-minx;
    if rra = 0 then do;
      if maxx ne 0 then rra=max(maxx,0.1);
      else if maxx=0 then rra=1;
      rra=round(rra,10);
    end;
  ...```

```sas
    x=round(rra*0.01);
    by=round(x,0.05);
    max=min(by,100);
    min=max-(by*round(rra/10));
```

```sas
  end;
  .......
```

* in=(required) input dataset,
* by=(optional) additional by variables.
* test_fmt=(required) format for test,
* cutoff=(required) time value which corresponds to baseline.
* percent=(required) NO or percent of range for dashed lines.
* title=(required) title number.

The input dataset must contain the following variables:

* sub=(subject id)
* test=(char) test code
* dose=(num) dose code in sort order.
* dose_fmt=(char) character description of dose.
* time=(num) time code in sort order.
* time_fmt (char) time format
* result=(num) test value
* baseline=(num) baseline value
* low=(num) low reference range
* high=(num) high reference range
* unit=(char) reference unit
* eval=(char) reference range evaluation

MACRO CODE

```sas
%macro lab4(in=,by=,test_fmt=,cutoff=,percent=,title=);
  .......
```
if byv=0 then do;
x=.001;
byv=round(ma(10,x));
end;
min=round(min(1));
max=max+byv*round;
end;
call symput('minval',min);
call symput('maxval',max);
call symput('byval',byv);

* Create an annotate dataset for line of no change and optional lines of x percent of range shift.
******************************************************
data cutoff;
length function $8. color $8.;
function='MOVE'; X=&minval; Y=&minval; XSYS='2'; YSYS='2';
output;
function='DRAW'; X=&maxval; Y=&maxval; color='BLACK'; line=1; XSYS='2'; YSYS='2';
output;
%if &percent=NO %then %do;
function='MOVE'; X=&minval; Y=&minval + &shift;
XSYS='2'; YSYS='2';
output;
function='DRAW'; X=&maxval + &shift; Y=&maxval;
color='BLACK'; line=2; XSYS='2'; YSYS='2';
output;
%end;
end;
run;

* Create titles and footnotes depending on request.
*************************************************************
%if &percent=NO %then %do;
title1='PLOT OF POST VS BASELINE FOR &parm (&uni);
footnote1='THE SOLID DIAGONAL LINE REPRESENTS THE LINE OF NO CHANGE';
%end;

%if &percent ne NO %then %do;
title1='PLOT OF POST VS BASELINE FOR &parm (&range)';
footnote1='THE SOLID DIAGONAL LINE REPRESENTS THE LINE OF NO CHANGE';
%end;

* Create annotate code.
******************************************************************
proc gplot data=a annotate=cutoff uniform;
%if &by=NO %then %do;
by &aby;
%end;
plot result*baseline=dose
/ haxis=axis1 vaxis=axis2
legend=legend1;
symbol1='none v=star c=black';
symbol2='none v=square c=black';
symbol3='none v=diamond c=black';
symbol4='none v=circle c=black';
symbol5='none v=triangle c=black';
symbol6='none v= c=black';
symbol7='none v= c=black';
symbol8='none v= c=black';
axis1 order=(&minval to &maxval by &byval)
length=6 in labels="BASELINE";
axis2 order=(&minval to &maxval by &byval)
length=6 in labels="POST";
Legend1 label="DOSE";
run;
%end;
run;

* Create format for dose from the input dataset.
***************************************************************
proc sort data=&in(keep=dose dose_tmt)
out=dose(rename=(dose=st dose_fmt=label)
nodupkey;
by dose dose_fmt;
data dose; set dose;
fmtname::>'ds';
type='N';
data dose; set dose; by fmtname;
length start $5;
start='other';
output;
if first.fmtname then do;
start='other';
label='O';
hs=O';
output;
end;
proc format cntlin=dose;
******************************************************************
* Step 1
* Create format for dose from the input dataset.
******************************************************************
proc sort data=keepdose dose_fmt)
out=dose(rename=(dose=st dose_fmt=label))
nodisplay;
by dose dose_fmt;
data dose; set dose; by fmtname;
length start $5;
start='O';
output;
if first.fmtname then do;
start='other';
label='O';
hs=O';
output;
end;
proc format cntlin=dose;

******************************************************************
* Step 2
* Create macro variables for each test. Do freq by test and unit if x % of range shift is not requested. Do freq by test, low, high and unit if x % of the range shift is requested.
******************************************************************
%if &percent=NO %then %do;
proc freq data=&in; tables test*unit/noprint out=parms;
%end;

%if &percent ne NO %then %do;
proc freq data=&in; tables test*low*high*unit/noprint out=parms;
%end;
data parms; set parms;
length x $3;
parm=put(test, &testFmt.);
parm=translate(parm, ".");
x=left(x);
call symput("n", n);
call symput("code", test);
call symput("parm", parm);
call symput("unit", unit);
/* calculate x % of range shift if requested */
%if &percent ne NO %then %do;
shift=(high-low)*(.01*percent);
length range $20;
range=trim(left(low))"-"trim(left(high))" "unit;
call symput("low", low);
call symput("high", high);
call symput("shift", shift);
call symput("range", range);
%end;
%end lab4;
run;

CONCLUSION

Data driven programming is one method of creating customized output. By combining SAS macro and SASGRAPH we have created data driven macros which allow us to produce various laboratory reports and plots with minimal programming. This reduces time spent on customized programming and ultimately will reduce the time needed to produce a Clinical Study Report.

REFERENCES


Paul Gilbert
Sterling Winthrop Inc.
9 Great Valley Parkway
Malvern, PA 19355
215-889-6542

Figure 1
%lab4(in=lab1, by=test, fmt=parm, cutoff=0, percent=NO, title=5)
PLDT OF POST VS BASELINE FOR ALKALINE PHOSPHATASE (LUL)

Figure 2
%lab4(in=lab2, by=center, fmt=parm, cutoff=0, percent=NO, title=5)
PLDT OF POST VS BASELINE FOR ALKALINE PHOSPHATASE (LUL) CENTER=A
Figure 3

PLOT OF POST VS BASELINE FOR ALKALINE PHOSPHATASE (39–132 IU/L)

The solid diagonal line represents the line of no change. The dashed lines represent increases or decreases from baseline which are 40% of the span of the reference range.

Figure 4

PLOT OF POST VS BASELINE FOR ALKALINE PHOSPHATASE (39–132 IU/L)

The solid diagonal line represents the line of no change. The dashed lines represent increases or decreases from baseline which are 40% of the span of the reference range.

Figure 5

INPUT DATASET

<table>
<thead>
<tr>
<th>CENTER</th>
<th>SUB</th>
<th>TEST</th>
<th>DOSE</th>
<th>DOSE_FMT</th>
<th>TIME</th>
<th>TIME_FMT</th>
<th>BASELINE</th>
<th>RESULT</th>
<th>EVAL</th>
<th>LOW</th>
<th>HIGH</th>
<th>UNIT</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>001</td>
<td>CALA</td>
<td>4</td>
<td>15 MG</td>
<td>0</td>
<td>BASE</td>
<td>3.35</td>
<td>3.300</td>
<td>N</td>
<td>3.3</td>
<td>5.5</td>
<td>GM/DL</td>
</tr>
<tr>
<td>A</td>
<td>001</td>
<td>CALB</td>
<td>4</td>
<td>15 MG</td>
<td>1</td>
<td>24 HR</td>
<td>3.35</td>
<td>3.300</td>
<td>N</td>
<td>3.3</td>
<td>5.5</td>
<td>GM/DL</td>
</tr>
<tr>
<td>A</td>
<td>001</td>
<td>CALC</td>
<td>4</td>
<td>15 MG</td>
<td>2</td>
<td>48 HR</td>
<td>3.35</td>
<td>3.300</td>
<td>N</td>
<td>3.3</td>
<td>5.5</td>
<td>GM/DL</td>
</tr>
<tr>
<td>A</td>
<td>001</td>
<td>CALD</td>
<td>4</td>
<td>15 MG</td>
<td>3</td>
<td>72 HR</td>
<td>3.35</td>
<td>3.300</td>
<td>N</td>
<td>3.3</td>
<td>5.5</td>
<td>GM/DL</td>
</tr>
<tr>
<td>A</td>
<td>001</td>
<td>CAME</td>
<td>4</td>
<td>15 MG</td>
<td>4</td>
<td>BASE</td>
<td>64.50</td>
<td>59.000</td>
<td>N</td>
<td>39.0</td>
<td>132.0</td>
<td>IU/L</td>
</tr>
<tr>
<td>A</td>
<td>001</td>
<td>CAMF</td>
<td>4</td>
<td>15 MG</td>
<td>5</td>
<td>24 HR</td>
<td>64.50</td>
<td>62.000</td>
<td>N</td>
<td>39.0</td>
<td>132.0</td>
<td>IU/L</td>
</tr>
<tr>
<td>A</td>
<td>001</td>
<td>CAMG</td>
<td>4</td>
<td>15 MG</td>
<td>6</td>
<td>48 HR</td>
<td>64.50</td>
<td>60.000</td>
<td>N</td>
<td>39.0</td>
<td>132.0</td>
<td>IU/L</td>
</tr>
<tr>
<td>A</td>
<td>001</td>
<td>CAMH</td>
<td>4</td>
<td>15 MG</td>
<td>7</td>
<td>72 HR</td>
<td>64.50</td>
<td>60.000</td>
<td>N</td>
<td>39.0</td>
<td>132.0</td>
<td>IU/L</td>
</tr>
<tr>
<td>A</td>
<td>002</td>
<td>CALA</td>
<td>4</td>
<td>15 MG</td>
<td>0</td>
<td>BASE</td>
<td>4.05</td>
<td>4.020</td>
<td>N</td>
<td>3.3</td>
<td>5.5</td>
<td>GM/DL</td>
</tr>
<tr>
<td>A</td>
<td>002</td>
<td>CALB</td>
<td>4</td>
<td>15 MG</td>
<td>1</td>
<td>24 HR</td>
<td>4.05</td>
<td>4.020</td>
<td>N</td>
<td>3.3</td>
<td>5.5</td>
<td>GM/DL</td>
</tr>
<tr>
<td>A</td>
<td>002</td>
<td>CALC</td>
<td>4</td>
<td>15 MG</td>
<td>2</td>
<td>48 HR</td>
<td>4.05</td>
<td>4.020</td>
<td>N</td>
<td>3.3</td>
<td>5.5</td>
<td>GM/DL</td>
</tr>
<tr>
<td>A</td>
<td>002</td>
<td>CALD</td>
<td>4</td>
<td>15 MG</td>
<td>3</td>
<td>72 HR</td>
<td>4.05</td>
<td>4.020</td>
<td>N</td>
<td>3.3</td>
<td>5.5</td>
<td>GM/DL</td>
</tr>
<tr>
<td>A</td>
<td>002</td>
<td>CAME</td>
<td>4</td>
<td>15 MG</td>
<td>4</td>
<td>BASE</td>
<td>63.50</td>
<td>65.000</td>
<td>N</td>
<td>39.0</td>
<td>132.0</td>
<td>IU/L</td>
</tr>
<tr>
<td>A</td>
<td>002</td>
<td>CAMF</td>
<td>4</td>
<td>15 MG</td>
<td>5</td>
<td>24 HR</td>
<td>63.50</td>
<td>67.000</td>
<td>N</td>
<td>39.0</td>
<td>132.0</td>
<td>IU/L</td>
</tr>
<tr>
<td>A</td>
<td>002</td>
<td>CAMG</td>
<td>4</td>
<td>15 MG</td>
<td>6</td>
<td>48 HR</td>
<td>63.50</td>
<td>63.000</td>
<td>N</td>
<td>39.0</td>
<td>132.0</td>
<td>IU/L</td>
</tr>
<tr>
<td>A</td>
<td>002</td>
<td>CAMH</td>
<td>4</td>
<td>15 MG</td>
<td>7</td>
<td>72 HR</td>
<td>63.50</td>
<td>63.000</td>
<td>N</td>
<td>39.0</td>
<td>132.0</td>
<td>IU/L</td>
</tr>
<tr>
<td>A</td>
<td>003</td>
<td>CALA</td>
<td>4</td>
<td>15 MG</td>
<td>0</td>
<td>BASE</td>
<td>3.70</td>
<td>3.700</td>
<td>N</td>
<td>3.3</td>
<td>5.5</td>
<td>GM/DL</td>
</tr>
<tr>
<td>A</td>
<td>003</td>
<td>CALB</td>
<td>4</td>
<td>15 MG</td>
<td>1</td>
<td>24 HR</td>
<td>3.70</td>
<td>4.220</td>
<td>N</td>
<td>3.3</td>
<td>5.5</td>
<td>GM/DL</td>
</tr>
<tr>
<td>A</td>
<td>003</td>
<td>CALC</td>
<td>4</td>
<td>15 MG</td>
<td>2</td>
<td>48 HR</td>
<td>3.70</td>
<td>4.220</td>
<td>N</td>
<td>3.3</td>
<td>5.5</td>
<td>GM/DL</td>
</tr>
<tr>
<td>A</td>
<td>003</td>
<td>CALD</td>
<td>4</td>
<td>15 MG</td>
<td>3</td>
<td>72 HR</td>
<td>3.70</td>
<td>5.800</td>
<td>N</td>
<td>3.3</td>
<td>5.5</td>
<td>GM/DL</td>
</tr>
<tr>
<td>A</td>
<td>003</td>
<td>CAME</td>
<td>4</td>
<td>15 MG</td>
<td>4</td>
<td>BASE</td>
<td>72.80</td>
<td>72.800</td>
<td>N</td>
<td>3.3</td>
<td>5.5</td>
<td>GAM/L</td>
</tr>
<tr>
<td>A</td>
<td>003</td>
<td>CAMF</td>
<td>4</td>
<td>15 MG</td>
<td>5</td>
<td>24 HR</td>
<td>72.80</td>
<td>112.000</td>
<td>N</td>
<td>39.0</td>
<td>132.0</td>
<td>IU/L</td>
</tr>
<tr>
<td>A</td>
<td>003</td>
<td>CAMG</td>
<td>4</td>
<td>15 MG</td>
<td>6</td>
<td>48 HR</td>
<td>72.80</td>
<td>65.000</td>
<td>N</td>
<td>39.0</td>
<td>132.0</td>
<td>IU/L</td>
</tr>
<tr>
<td>A</td>
<td>003</td>
<td>CAMH</td>
<td>4</td>
<td>15 MG</td>
<td>7</td>
<td>72 HR</td>
<td>72.80</td>
<td>90.000</td>
<td>N</td>
<td>39.0</td>
<td>132.0</td>
<td>IU/L</td>
</tr>
</tbody>
</table>

1144