

Effectively Displaying Statistical Results in a High Volume Drug Research Environment using SAS/GRAPH®

Bruce Chinn, Roche Bioscience, Palo Alto, CA

ABSTRACT

In a drug discovery research environment, scientists sometimes perform many experiments in a short period of time to obtain information about drug compounds and its different concentrations. After analysis, they need a way to scan the results to confirm whether or not the compound has potential. Plotting graphs and summarizing the results enables the scientist to quickly review the information. This paper discusses the use of SAS/GRAPH and some of its features such as PROC GPLOT, PROC GPRINT, ANNOTATE, PROC GREPLAY and templates as a possible way to effectively present information. One other SAS® product mentioned in this paper is SAS/STAT. This paper is intended for an average skill level user of the SAS system.

INTRODUCTION

In a drug research environment where scientists need experiments analyzed and summarized at a high volume, time plays an important role in taking a compound forward to development. Once an experiment is completed and data is analyzed, results need to be returned to the scientist in a form in which they can quickly determine the potential or feasibility of a particular compound. A good way to do this is by using graphs side by side to present an overall picture of the results of the analysis. This paper will discuss the use of SAS/GRAPH and some of its features such as PROC GPLOT, ANNOTATE, PROC GREPLAY and templates as a possible way to present information.

Once data has been collected and stored in your data warehouse, the SAS system can then be invoked to retrieve the desired data to be analyzed and graphed. In this example, detail data representing the percent inhibition responses from each concentration are grouped by compound and retrieved. Each compound is then passed through a dose response analysis using either a nonlinear regression (PROC NLIN) or two-point linear regression ($Y=mX+b$) and the resulting curve for each compound is plotted on a graph along with the IC50 point estimate¹. The graph is printed using a print driver that allows multiple graphs to be printed on one side of the page.

Assuming that the regression analysis for the compound has been completed successfully and the graph is ready

¹ For purposes of this paper, an IC50 estimate is the dose or concentration (X coordinate) which yields 50% inhibition (Y coordinate) on the curve.

to be created, you can create the ANNOTATE data set which stores commands that allow you to draw the actual curve and reference lines on the graph before the data is actually plotted.

The following SAS code is an example of building the ANNOTATE data set:

```
*****
%MACRO ANNOT;

    DATA OUTANNO;

        LENGTH FUNCTION $8 TEXT $64;
        KEEP ASSAY FUNCTION X XSYS Y YSYS
            SIZE LINE STYLE TEXT
            CMDORDER POSITION;

        SET INP;

        POSITION=' '; TEXT=' ';
        * SET COORDINATES AND SPECIFICATIONS FOR
          LABELS ;
          CMDORDER=0;
          STYLE ='SWISS';
          LINE=21; SIZE=1.25;

        * DRAW IC50 LINE ;
          FUNCTION='MOVE';
          X= ICEST ; Y=0; XSYS=2';YSYS='1';
          CMDORDER+1; OUTPUT;

          FUNCTION='DRAW';
          X= ICEST ; Y=100;
          LINE=21;
          CMDORDER+1; OUTPUT;

        * DRAW 50% LINE;
          FUNCTION='MOVE';
          X= 0 ; Y=50; XSYS='1';YSYS='2';
          CMDORDER+1; OUTPUT;

          FUNCTION='DRAW';
          X= 100 ; Y=50;
          LINE=21;
          CMDORDER+1; OUTPUT;

          XSYS='2'; YSYS='2';

        * DRAW FITTED EQUATION ;

          X = 0.001;
          Y = 100 / (1+(ICEST/X)**SLOPE);
          FUNCTION='MOVE';
```

```

LINE=1;
OUTPUT;

DO X=
  %if "&_loconc" = "LO" %then %do;
    0.002 TO 0.009 BY 0.001,
  %end;
    0.01 TO 0.09 BY 0.01,
    0.10 TO 0.90 BY 0.1,
    1.00 TO 9.00 BY 1,
    10.00 TO 99.00 BY 10
  %if "&_hiconc" = "HI" %then %do;
    , 100.00 TO 999.00 BY 100
  %end;
;
Y = 100 / (1+(ICEST/X)**SLOPE);
FUNCTION='DRAW';
OUTPUT;
END;

RUN;

```

```

%MEND ANNOT;
*****

```

Notice that the ANNOTATE data set is a SAS data set that contains certain variables that are keywords and lets SAS know such things as type, position and font of the line that is to be drawn. You can also specify the type, position and font of any text to be written on the graph. The axes and/or curve line can also be lengthened based on the input data (by using macro variables &_loconc and &_hiconc which can be assigned earlier).

Next, you can plot the graph using PROC GPLOT and the ANNOTATE data set. Use a PLOT statement with the X and Y variables and specify what type of symbol to use for the points on the plot. You can also issue specifications for the title and axes and that the graph is to be framed. Immediately following the graph, the two tables are printed which show the results of the regression. A PROC PRINTTO is issued to save the output from two PROC PRINT steps into a file. A PROC GPRINT is then used to copy the saved file into the graph catalog OUTPLOTZ. The font and text attributes of the text in the tables are also adjusted for better readability.

The following is the SAS code for the PROC GPLOT to create the graph and the tables:

```

*****
%MACRO PLOT;

GOPTIONS HSIZE=0 VSIZE=0 HTEXT=2 VPOS=0
  HPOS=0 ROTATE=PORTRAIT DEVICE=PS
  COLORS=(BLACK) FBY=SWISS HBY=2
  FTEXT=SWISS NODISPLAY NOPROMPT;

```

* Grab individual data from input file;

```

DATA ANAL;
SET INP;
IF _N_=1 THEN DO;
CALL SYMPUT('_ENDDT',COMPRESS(ENDDT));
CALL SYMPUT('_NB',COMPRESS(NBNUM));
CALL SYMPUT('_RSNUM',COMPRESS(RS_NUM));
CALL SYMPUT('_ASSAY',
  COMPRESS(PUT(ASSAY,8.)));
END;
RUN;

* Grab IC50 estimate from summary file;
DATA _NULL_;
SET EST;
* Use Exponential notation ;
CALL SYMPUT('_ICEST',
  COMPRESS(GTLTIIPUT(ICEST,E9.)));
RUN;

```

* PLOT THE DATA ;

```

PROC GPLOT DATA=ANAL
  ANNOTATE=OUTANNO GOUT=OUTPLOTZ;
PLOT RESP*CONC / HAXIS=AXIS1
  VAXIS=AXIS2 FRAME;

```

* Define the symbol for the data points (SYMBOL1).;

```

SYMBOL1 MODE=INCLUDE V=SQUARE W=2
  I=NONE R=1;

```

```

AXIS1 LABEL=(H=2 J=C F=SWISS
  "CONCENTRATION (" F=CGREEK "m"
  F=SWISS "M)")
LOGBASE=10 LOGSTYLE=EXPAND
ORDER=
%if "&_loconc" = "LO" %then %do;
.0001 .001
%end;
.01 .1 1 10 100
%if "&_hiconc" = "HI" %then %do;
1000
%end;
VALUE=(F=SWISS H=2);

```

```

AXIS2 LABEL=(H=2 J=C F=SWISS R=0 A=90
  "PERCENT INHIBITION")
ORDER=(-10 TO 100 BY 10)
VALUE=(F=SWISS H=2);

```

```

TITLE1 J=LEFT H=2.5 "ASSAY &_ASSAY" J=RIGHT
H=2.5 "IC50 (uM) = &_ICEST";
TITLE2 J=LEFT H=2.5 "PROTOCOL: XYZ" J=RIGHT
H=2.5 "COMPOUND: &_RSNUM";
TITLE3 J=LEFT H=2.5 "END DATE: &_ENDDT";
RUN;

```

```

TITLE2; TITLE3;

```

```

* Print tables to this file ;
FILENAME P "DETAIL.GPR";

```

```

PROC PRINTTO PRINT=P NEW;
RUN;

DATA PRTEST (KEEP=PARAM IC50 SE LO HI);
  SET EST;
  LENGTH PARM $4 IC50 SE HI LO $11;
  IC50 = "&_ICEST";
  PARM = 'IC50';
  * Use Exponential notation;
  SE = COMPRESS(PUT(ICSE,E9.));
  HI = COMPRESS(PUT(ICH1,E9.));
  LO = COMPRESS(PUT(ICLO,E9.));
  OUTPUT;
  IC50 = COMPRESS(PUT(SLOPE,E9.));
  PARM = 'HILL';
  SE = COMPRESS(PUT(SLOPESE,E9.));
  HI = COMPRESS(PUT(SLOPEHI,E9.));
  LO = COMPRESS(PUT(SLOPELO,E9.));
  OUTPUT;
RUN;

OPTIONS LS=80;

PROC PRINT DATA=PRTEST NOOBS SPLIT='_';
VAR PARM IC50 SE LO HI;
LABEL PARM = 'PARAM'
      IC50 = 'EST'
      SE = 'A.S.E.'
      LO = 'LOWER'
      HI = 'UPPER'
;
TITLE "SUMMARY VALUES";
RUN;

* No page break ;
OPTIONS FORMDLIM=' ';

PROC PRINT DATA=ANAL NOOBS SPLIT='_';
VAR CONC RESP;
%if "&_loconc" = "LO" %then %do;
  FORMAT CONC 7.4;
%end;
%else %do;
  FORMAT CONC 7.3;
%end;
LABEL CONC = 'CONC (uM)'
      RESP = '% INHIB'
;
TITLE "DETAIL VALUES";
RUN;

PROC PRINTTO;
RUN;

OPTIONS FORMDLIM=" LS=132;
TITLE;

* Copy print file to SASGRAPH catalog to print after
graphs ;
GOPTIONS HTEXT=1.6 FTEXT=SWISSL;

```

```

PROC GPRINT FILEREF=P GOUT=OUTPLOTZ;
RUN;

```

```

* Reset options and deallocate filename ;
GOPTIONS HTEXT=2 FTEXT=SWISS;
FILENAME P;

```

```

%MEND PLOT;
*****

```

The graph catalog called OUTPLOTZ is used later to generate all graphs into one file using PROC GREPLAY with a template. The graph shows the percent inhibition values (the Y axis with attributes specified in the AXIS1 statement) for each concentration tested (the X axis with attributes specified in the AXIS2 statement). The points are plotted using attributes in the SYMBOL1 statement. Each title line on the graph is split into two sections (left and right justification) to display different information.

After all compounds are processed and graphs and tables are saved in the graph catalog, a PROC GREPLAY with a "two-up" per page template is then used to create the total graph package. In this example, the macro variable &_iter contains the number of compounds to be graphed and each compound has two entries saved in the graph catalog.

The following is the SAS code for the PROC GREPLAY:

```

*****

```

```

%MACRO REPLAY;

```

```

FILENAME GSF "graph.gsf" ;

```

```

GOPTIONS ROTATE=PORTRAIT
      DEVICE=PS COLORS=(BLACK) FBY=SWISS
      GSFMODE=REPLACE GSFNAME=GSF
      HSIZE=8 VSIZE=10 VPOS=0 HPOS=0
      NODISPLAY NOPROMPT;

```

```

* TWO-UP (V2S);
PROC GREPLAY TC=SASHELP.TEMPLT
  TEMPLATE=V2S
  IGOUT=OUTPLOTZ NOFS;

```

```

%do i = 1 %to &_iter;

```

```

  %let j = %eval(&i*2-1);
  %let k = %eval(&i*2);

```

```

  TREPLAY 1:&j 2:&k;

```

```

%end;

```

```

QUIT;

```

```

%MEND REPLAY;

```

```

*****

```

The output graph file "graph.gsf" can now be printed. In this case, you want to put as many graphs on a page as possible so that the scientist can quickly review the results. In this example, the scientists have decided that printing six graph/table combinations per page is still readable without having to use a magnifying glass. The output graph file is then sent to a VAX® system print queue which allows "six-up" printing and looks like the attached graph.

SUMMARY

By scanning the shapes of the graphs' curve, the scientist can quickly gain insight into the compound's binding mechanism. The fit of the estimated curve to the raw data also indicates whether the statistical fit was good. The resulting IC50 point estimates help the scientist to determine whether to take the compound forward or return to the drawing board. In terms of productivity, time is saved by not having to manually graph each compound, therefore allowing the scientist to go on to other important tasks.

In general, using graphs to present information is a good idea, but in particular, when used in an environment where data is analyzed in high volume, graphs can be a very powerful and effective tool.

TRADEMARK INFORMATION

SAS, SAS/GRAPH and SAS/STAT are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. © indicates USA registration.

VAX is a registered trademark of Digital Equipment Corporation, Maynard, MA.

Other brand and product names are registered trademarks or trademarks of their respective companies.

AUTHOR

Bruce Chinn
Roche Bioscience
M/S S1-137
3401 Hillview Avenue
Palo Alto, CA 94304
Phone: (415) 354-7314
Fax: (415) 855-5627
E-mail: bruce.chinn@roche.com

