Exploring Multi-dimensional Relationships with SAS/GRAPH® Software

Gail F. Kirk and Anne Horney
Cooperative Studies Program Coordinating Center
V A Medical Center, Perry Point, MD

ABSTRACT

When an outcome variable cannot be measured reliably, substitution of related variables for data analysis is often necessary. Graphs help us to evaluate proxies and to understand the relationships among the surrogates and the outcome measure. This poster will show a graphing technique that we have used to investigate possible relationships.

In a recent clinical trial four laboratory tests and measurements of an outcome variable X were obtained at baseline and follow-up. This outcome measure X is of questionable reliability. It is thought that changes in the laboratory test values may reflect changes in X. The strength of the relationships among the variables required assessment. We will show how we put four graphs and a table of background information on a single page by using the GPLOT and GSLIDE procedures, an Annotate data set, the GREPLAY procedure and a 5-panel template.

A similar approach might be used in many different environments --

- Exercise programs (miles, pounds, minutes, reps) and physical fitness
- Various test scores and student ability
- Music/sounds (jazz, classical, rock, waterfall) and the milking parlor (Do happy cows give more milk?)

INTRODUCTION

Sometimes an outcome cannot be measured directly, and we must rely on surrogate measurable variables for our estimation of the outcome. There are also times when the measurement of an outcome variable is suspect or unreliable, and we look to surrogates for a better understanding of what is occurring. In either case we need to explore the relationships among the outcome variable and its (proposed) proxies; usually a series of graphs are prepared for this purpose. We will show a graphing technique that puts data from five variables (one outcome and four possible proxies) onto a single sheet of paper for such exploration into multi-dimensional relationships.

THE PIECES OF THE PUZZLE

We take our example from a clinical trial concerning alcohol consumption. The outcome variable of interest was the amount of ethanol consumed, and we measured this by asking each subject how much alcohol he/she had drunk in the last week. Since the reliability of self reported alcohol consumption has been questioned, we also looked at several blood tests as possible biochemical markers of ethanol intake. The patients were divided into two groups (Control and Treatment); the blood tests and consumption reports were obtained at baseline and after six months in the study. It was thought that the important information was in the amount of change in consumption, not the actual starting or ending value; likewise, we wanted to look at the changes in the blood test values. Four blood tests were chosen: high density lipoprotein cholesterol (HDL), apolipoprotein components Apo A_1 and Apo A_2, and gamma glutamyltransferase (GGT). We wanted to prepare a chart showing consumption values in groups or categories and showing the blood test averages with some indication of variability. No actual study data are shown here; these data are simulated and may have little resemblance to actual results.

THE SOLUTION

DATA CHOICES

We look at the six-month changes in the ethanol consumption variable and the corresponding six-month changes in each of the blood tests. Changes in reported alcohol intake ranged from a decrease of about 500 grams per week to increases of more than 600 grams per week. We decided to look at quintiles of the intake change variable, and we developed the cut points from the FREQ option of the UNIVARIATE procedure. In a DATA step, each

patient was assigned to a quintile based on his/her change in intake value.

```
CP1=-162; CP2=-49; CP3=58; CP4=207;
IF ChEtOH NE . THEN SELECT;
   WHEN (ChEtOH LE CP1)          QN=1;
   WHEN (CP1 LT ChEtOH LE CP2)   QN=2;
   WHEN (CP2 LT ChEtOH LE CP3)   QN=3;
   WHEN (CP3 LT ChEtOH LE CP4)   QN=4;
      OTHERWISE   QN=5;
   END;
ELSE DELETE;
```

Means, Ns, and standard errors for intake change (by quintile) and for blood test values (by quintile, by group) were calculated in the MEANS procedure and placed in an output data set.

```
PROC MEANS  N MEAN STDERR;
  VAR ChEtOH CHDL CAA1 CAA2 CGGT;
  CLASS QN GROUP;
  OUTPUT  OUT=OUTMNS   N(ChEtOH)=NX
    MEAN=MEtOH MHDL MAA1 MAA2 MGGT
    STDERR=SEtOH SHDL SAA1 SAA2 SGGT;
```

Another DATA step placed the means and standard errors into the form to be used by the SAS/GRAPH® software. This step is not shown; instead, in this example, the INPUT statement reads the numbers from card images.

GRAPHING THE PROXIES

We used the procedure GPLOT to graph for each group the mean change in each of the blood tests against the quintiles of intake change. Also bars of length ±2 standard errors were drawn around each mean by use of the SYMBOL statement interpolation option INTERPOL=HILO. Treatment group data were plotted .03% to the left of the vertical lines for quintiles 2, 3, and 4 and .06% to the left of the fifth quintile line; control group data were plotted .03% to the right of the vertical lines for quintiles 2, 3, and 4 and .06% to the right of the first quintile line. This usually prevented the asterisk from overprinting the circle. We set the left axis ranges to be symmetric around zero. Although the four blood test graphs are identical in design, each graph is programmed to be drawn independently; identical design gives a nice balance but is not necessary.

The SAS programming to produce the graph for the HDL test is shown in Appendix A.

TABLE and TITLES

Average change in intake was tabulated along with the group sample sizes in each quintile, and the table was enclosed in a box. An Annotate data set included the commands to draw this table box and also the boxed legend of group symbols. TITLEs and a FOOTNOTE were added, and the GSLIDE procedure was invoked.

VIEWING THE RESULT

After the four blood test (the proposed proxies) graphs were produced and the data table of the target variable (the measure of interest) was prepared, the next step was to plan for displaying all of them on a single piece of paper or a single screen. The plan, called the template, was designed in the GREPLAY procedure:

EDIT *templatename*.TEMPLATE

The proxies graphs were placed in 4 landscape panels, two rows by two columns, and a larger portrait panel was placed in the center to display titles and the table of target information. See figure 1 for the panel outlines and Appendix B for the template design screen.

GRAPH POSTER.T4PLUS.TEMPLATE

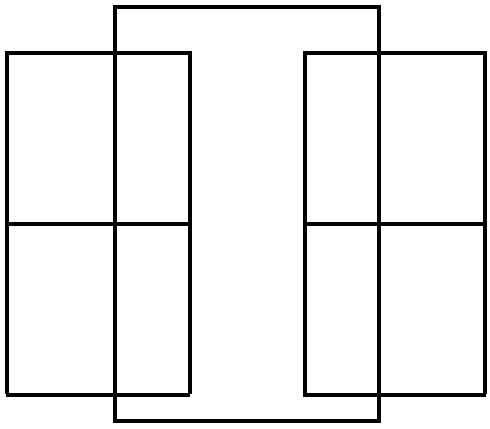

Figure 1. A five panel template

The template was designed for landscape viewing. An example of the commands submitted to view graphs/slides/charts stored in the graphics catalog

CAT_AX on the template T4PLUS from the TMPLT template catalog is:

```
LIBNAME  POSTER  'C:\SUGI98\POSTER';
PROC GREPLAY  IGOUT=POSTER.CAT_AX
              GOUT=POSTER.CAT_AX
              TC=POSTER.TMPLT
              TEMPLATE=T4PLUS;
```

On the GREPLAY screen the content of each of the five panels is selected by panel number. (See figure 2.)

Discussion about how well the blood tests in this example mirror self reported ethanol intake is left for another time. However, it can be observed in these simulated data that HDL and GGT changes increase as intake changes increase while Apo A_1 changes decrease and Apo A_2 seems to vary in no particular direction. No systematic difference can be seen here between the treatment groups. Figure 3 is the completed graph.

PROC GREPLAY

IGOUT: POSTER.CAT_AX GOUT: POSTER.CAT_AX Device: WIN

TC: POSTER.TMPLT Template: T4PLUS Scroll: PAGE

CC: _____ Cmap: _____

Sel	Name	Type	Description	Created
5	CENTER	I	ALC + TITLES GSLIDE	06/12/97
1	PANEL1	I	HDL PLOT, PLOT2	06/12/97
2	PANEL2	I	AA1 PLOT, PLOT2	06/12/97
3	PANEL3	I	AA2 PLOT, PLOT2	06/12/97
4	PANEL4	I	GGT PLOT, PLOT2	06/12/97

Figure 2. GREPLAY Screen with panels selected

SOFTWARE/HARDWARE

This work was done using the SAS System for Windows 3.10, release 6.08, and Microsoft Windows for Workgroups, version 3.11, on a Gateway2000, model P5-75, computer system.

CONCLUSION

We have shown a graphing technique we have found useful in exploring relationships among a group of variables, one of which is difficult to measure exactly. The need to bring several kinds of information together on one graph occurs not only in clinical trials, where our example was developed, but also in education, manufacturing, agriculture, and many other fields. SAS/GRAPH software can help analysts develop solutions to meet these needs.

ACKNOWLEDGMENTS

We would like to thank Dr. Joe Collins for his helpful comments, and our thanks go to Dr. Steve Bingham, Dr. Bill Cushman and Dr. Dean Follman for the sharing of their insights during the development of this technique.

THE AUTHORS

Gail F. Kirk
(410) 642-2411 ext. 5296

Anne Horney
(410) 642-2411 ext. 5298

Mailing address for both:
CSPCC (151E)
VA Medical Center
P. O. Box 1010
Perry Point, MD 21902

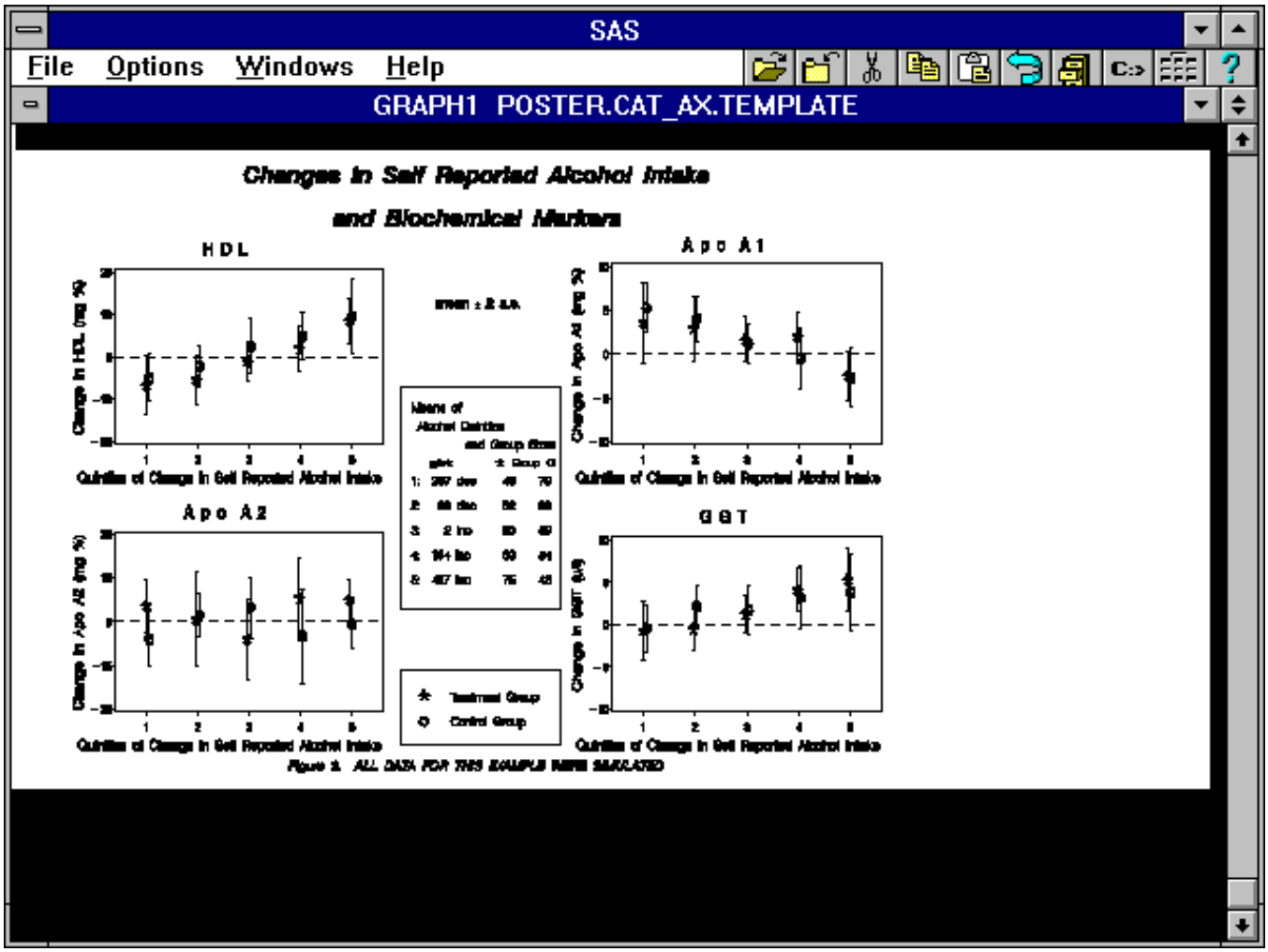

Figure 3. ALL DATA FOR THE EXAMPLE WERE SIMULATED

APPENDIX A. SAS CODE

A blood test results panel:

```
LIBNAME POSTER 'C:\SUGI98\POSTER';

GOPTIONS  RESET=ALL
          DEVICE=WIN    TARGET=WINPRTG
          PROMPT        ROTATE=LANDSCAPE
          FTEXT=SWISS   HTEXT=4
          CBACK=WHITE   COLORS=(BLACK)
          GUNIT=PCT;

DATA INHDL;
  INPUT GROUP QN MN SE;
  CARDS;
        1   1    -6.8   3.5
        2   1    -4.9   2.8
        1   2    -5.4   2.9
        2   2    -2.2   2.5
        1   3    -1     2.3
        2   3     2.6   3.3
        1   4     1.9   2.7
        2   4     4.9   2.8
        1   5     8.6   2.7
        2   5     9.6   4.4
;
RUN;
DATA HDLHL1(KEEP=GROUP QN M)
     HDLHL2(KEEP=GROUP QN M)
     HDLMN1(KEEP=GROUP QN MEAN)
     HDLMN2(KEEP=GROUP QN MEAN);
  SET INHDL;
  IF GROUP=1 THEN DO;
     IF QN=5 THEN QN=QN-0.06;
     IF QN IN (2,3,4) THEN QN=QN-0.03;
     M=MN-2*SE; OUTPUT HDLHL1;
     M=MN; OUTPUT HDLHL1;
     M=MN+2*SE; OUTPUT HDLHL1;
     MEAN=MN; OUTPUT HDLMN1;
     END;
  IF GROUP=2 THEN DO;
     IF QN=1 THEN QN=QN+0.06;
     IF QN IN (2,3,4) THEN QN=QN+0.03;
```

```
     M=MN-2*SE; OUTPUT HDLHL2;
     M=MN; OUTPUT HDLHL2;
     M=MN+2*SE; OUTPUT HDLHL2;
     MEAN=MN; OUTPUT HDLMN2;
     END;
RUN;
DATA HDL; SET HDLHL1 HDLMN1 HDLHL2 HDLMN2;
RUN;

SYMBOL1 I=HILOT;
SYMBOL2 I=HILOT;
SYMBOL3 H=9.5  F=SWISS  V=*;
SYMBOL4 H=6.5  F=SWISS  V=o;

AXIS1  MINOR=NONE  OFFSET=(10,10)
       ORDER=(1 TO 5)
       LABEL=(H=5 'Quintiles of Change in'
' Self Reported Alcohol Intake');

AXIS2  MINOR=NONE          /*  HDL  */
       ORDER=(-20 TO 20 BY 10)
       LABEL=(H=5  ANGLE=90  'Change in'
' HDL (mg %)');

AXIS6  MAJOR=NONE  MINOR=NONE
       ORDER=(-20 TO 20 BY 10)
       LABEL=NONE  VALUE=NONE;

TITLE1 ' ';  TITLE2  H=6 'H D L';
PROC GPLOT DATA=HDL    GOUT=POSTER.CAT_AX;
   PLOT M*QN=GROUP
       / NAME='PANEL1' DES='HDL PLOT,PLOT2'
         HAXIS=AXIS1
         VAXIS=AXIS2   VREF=0   LVREF=2
         NOLEGEND      FRAME;
   PLOT2 MEAN*QN=GROUP
       / VAXIS=AXIS6
         NOLEGEND;
RUN;
QUIT;
```

The table and titles panel:

```
LIBNAME POSTER 'C:\SUGI98\POSTER';
GOPTIONS  RESET=ALL
          DEVICE=WIN   TARGET=WINPRTG
          PROMPT       ROTATE=PORTRAIT
          FTEXT=SWISS  HTEXT=2
          CBACK=WHITE  COLORS=(BLACK)
          GUNIT=PCT;

TITLE1  H=3.55  F=SWISSBI  'Changes in Self'
 ' Reported Alcohol Intake';
TITLE3  H=3.55  F=SWISSBI  'and Biochemical'
 ' Markers';
TITLE9  H=2  F=SWISS 'mean'  F=MATH ' G '
  F=SWISS '2 s.e.';
FOOTNOTE H=2 F=SWISSI 'Figure 3.  ALL DATA'
 ' FOR THIS EXAMPLE WERE SIMULATED';

DATA ANNX;
  LENGTH FUNCTION STYLE $ 8  TEXT $ 22
      HSYS XSYS YSYS WHEN POSITION $ 1;
  RETAIN  WHEN 'a'  HSYS XSYS YSYS '3';

/* THE BOX AROUND THE TABLE */
   FUNCTION='MOVE';  X=34; Y=27;  OUTPUT;
   FUNCTION='DRAW';  X=34; Y=63;  SIZE=.06;
      LINE=1;  OUTPUT;
   FUNCTION='DRAW';  X=67; Y=63;  SIZE=.06;
      LINE=1;  OUTPUT;
   FUNCTION='DRAW';  X=67; Y=27;  SIZE=.06;
      LINE=1;  OUTPUT;
   FUNCTION='DRAW';  X=34; Y=27;  SIZE=.06;
      LINE=1;  OUTPUT;

/* TABLE HEADINGS */
   FUNCTION='LABEL';  POSITION=6;
     STYLE='SWISSL';  X=36; Y=60;  SIZE=1.8;
     TEXT='Means of';  OUTPUT;
   FUNCTION='LABEL';
     STYLE='SWISSL';  X=36; Y=57;  SIZE=1.8;
     TEXT=' Alcohol Quintiles';  OUTPUT;
   FUNCTION='LABEL';
     STYLE='SWISSL';  X=46; Y=54;  SIZE=1.8;
     TEXT=' and Group Sizes ';  OUTPUT;

   FUNCTION='LABEL';
     STYLE='SWISS'; SIZE=1.5; X=39; Y=51;
     TEXT=' g/wk ';  OUTPUT;
   FUNCTION='LABEL';
     STYLE='SWISS'; SIZE=3; X=53.5; Y=50.8;
     TEXT='*';  OUTPUT;
   FUNCTION='LABEL'; POSITION='0';
     STYLE='SWISSL'; SIZE=1.5; X=56; Y=50;
     TEXT=' Group';  OUTPUT;
   FUNCTION='LABEL'; POSITION='0';
     STYLE='SWISS'; SIZE=2.1; X=.; Y=.;
     TEXT=' o';  OUTPUT;

/* TABLE OF QUINTILE MEANS, Ns */
   FUNCTION='LABEL';
     STYLE='SWISSL'; SIZE=1.8; X=36.2; Y=47;
     TEXT='1:  307 dec';  OUTPUT;
   FUNCTION='LABEL';
     STYLE='SWISSL'; SIZE=1.8;  X=55; Y=47;
     TEXT='46    79';  OUTPUT;

   FUNCTION='LABEL';
```

```
    STYLE='SWISSL'; SIZE=1.8; X=36; Y=43;
    TEXT='2:   99 dec';  OUTPUT;
  FUNCTION='LABEL';
    STYLE='SWISSL'; SIZE=1.8; X=55; Y=43;
    TEXT='52    68';  OUTPUT;
   FUNCTION='LABEL';
    STYLE='SWISSL'; SIZE=1.8; X=36; Y=39;
    TEXT='3:    2 inc';  OUTPUT;
  FUNCTION='LABEL';
    STYLE='SWISSL'; SIZE=1.8; X=55; Y=39;
    TEXT='60    59';  OUTPUT;

  FUNCTION='LABEL';
    STYLE='SWISSL'; SIZE=1.8; X=36; Y=35;
    TEXT='4:  134 inc';  OUTPUT;
  FUNCTION='LABEL';
    STYLE='SWISSL'; SIZE=1.8; X=55; Y=35;
    TEXT='69    54';  OUTPUT;

  FUNCTION='LABEL';
    STYLE='SWISSL'; SIZE=1.8; X=36; Y=31;
    TEXT='5:  417 inc';  OUTPUT;
  FUNCTION='LABEL';
    STYLE='SWISSL'; SIZE=1.8; X=55; Y=31;
    TEXT='75    43';  OUTPUT;

/* BOX AROUND THE LEGEND */
  FUNCTION='MOVE';  X=34; Y=5;   OUTPUT;
  FUNCTION='DRAW';  X=34; Y=17;  SIZE=.06;
     LINE=1;  OUTPUT;
  FUNCTION='DRAW';  X=67; Y=17;  SIZE=.06;
     LINE=1;  OUTPUT;
  FUNCTION='DRAW';  X=67; Y=5;   SIZE=.06;
     LINE=1;  OUTPUT;
  FUNCTION='DRAW';  X=34; Y=5;   SIZE=.06;
     LINE=1;  OUTPUT;

/* LEGEND */
     FUNCTION='LABEL';
       STYLE='SWISS'; SIZE=4; X=37.4; Y=10.7;
       TEXT='*';  OUTPUT;
     FUNCTION='LABEL';
       STYLE='SWISSL'; SIZE=1.8; X=40.5; Y=12;
       TEXT='   Treatment Group';  OUTPUT;

     FUNCTION='LABEL';
       STYLE='SWISS'; SIZE=2.5; X=37.6; Y=8;
       TEXT='o';  OUTPUT;
     FUNCTION='LABEL';
       STYLE='SWISSL'; SIZE=1.8; X=40.5; Y=8;
       TEXT='   Control Group';  OUTPUT;
RUN;

PROC GSLIDE  GOUT=POSTER.CAT_AX
             ANNOTATE=ANNX
             NAME='CENTER'
             DES='ALC + TITLES GSLIDE';
RUN;
QUIT;
```

APPENDIX B. FIVE PANEL TEMPLATE DESIGN SCREEN

PROC GREPLAY: TEMPLATE DESIGN

TEMPLATE: T4PLUS — TC: POSTER.TMPLT
DESC: 4 panels LANDSCAPE, Center panel PORTRAIT — Scroll: PAGE
Device: WIN

Panel	Clp	Color		L-left	U-left	U-right	L-right		Scale	Xlate	Rotate
1	_	____	X:	3.0	3.0	40.0	40.0	X:	_____	_____	_____
			Y:	47.0	89.0	89.0	47.0	Y:	_____	_____	
2	_	____	X:	61.0	61.0	98.0	98.0	X:	_____	_____	_____
			Y:	47.0	89.0	89.0	47.0	Y:	_____	_____	
3	_	____	X:	3.0	3.0	40.0	40.0	X:	_____	_____	_____
			Y:	5.0	47.0	47.0	5.0	Y:	_____	_____	
4	_	____	X:	61.0	61.0	98.0	98.0	X:	_____	_____	_____
			Y:	5.0	47.0	47.0	5.0	Y:	_____	_____	
5	_	____	X:	23.0	23.0	79.0	79.0	X:	_____	_____	_____
			Y:	2.0	99.0	99.0	2.0	Y:	_____	_____	