Macro Based Reshaping of Pharmacokinetic Data for Crossover Studies

Doreen M. Turczyn-Toles

Abstract

During the early stages of drug development, study monitors need to assess the body's absorption, elimination and tolerance of study drug over time. To meet this need, programmers must often reshape pharmacokinetic data to produce both tables by time and graphs by subject. Programming quickly becomes complicated when time profiles need to be differentially formatted by treatment leg and by unequal number of subjects. The following program is designed to allow program flexibility while customizing the results for the individual study monitor by using SAS® macros and macro variables.

Introduction:
Clinical Pharmacology is often referred to as the preclinical to clinical link since it is usually the first opportunity for assessing laboratory findings in humans. Obtaining preliminary human pharmacokinetic (PK) data via plasma drug concentrations and time profiles is an integral part of the drug development process. This paper presents the programming issues involved in producing data displays of pharmacokinetic data relating to plasma drug concentrations in a crossover study.

Data Structure:
Data are assumed to be stored in a SAS® data set with one record per patient per scheduled collection time whether or not a value exists for that time. The identifying information consists of investigator number, subject number, sequence number (for sorting purposes), study phase, and treatment (including dose). These identifying values are repeated in each record of the data set. Actual and scheduled concentration times are recorded by 24-hour clock. Estimated concentration values are assumed to have been calculated by the biometrician and already merged into this data set.

Initial Setup:
The usual approach to calculating column positions for data layout is to mock up the layout on a printout design form and hard-code the results in the DATA _NULL_. Instead, this program uses the macro HEADCOL to center a complete string on a page, given the string and the page width. Moreover, this macro correctly determines the total string length for the condition where two or more variables need to be concatenated to create the string.

ZPRINT is a step-saving macro which allows PROC PRINT to work for entire data sets as well as subsets (i.e., OBS= option) by using only one line of code. This minimizes the overall number of keystrokes, thereby reducing the potential for typing errors.

The next step defines the global macro variables that will be used in the remainder of the program. The advantage of keeping all the macro variables for this program section in one place is that the programmer can avoid reading through many lines of code to find a particular item. The programmer can also compare the variables more easily to each other to avoid potential conflicts.

Macro variables allow maximum flexibility in data display generation because the incoming data set merely needs to be in the format described earlier. Defining the macro variables up front eliminates the potential for error by either globally renaming variables or by editing the variables one-by-one. Thus, the latter program code is self-contained.
Note, even the titles are passed as global macro variables.

The key macro variable, PTIMES, is the total number of scheduled concentration times for the entire study. For this study, we will identify four scheduled times: Day 1 morning and evening, Day 2 morning and evening. The PTIMES value is used extensively throughout the program to automate the calculation and printing of many other variables.

To further simplify the data display generation process, several SAS® formats are used for printing the final data displays. For example, SUMF is used to format all the summary statistics from PROC MEANS. CONF for concentration values. SUBF for subject number, and TIMEF for scheduled times. Again, the formats are located in the beginning of the program for ease of reading and comparison. Appropriately defining the formats provides maximum utility in a variety of situations because advanced SAS® features such as column pointer controls (+n) and line pointer controls (#n) can be used to customize the output.

Data Reshaping:
The incoming data consist of multiple records per subject which are reformatted to one record per subject per phase. Since this is a crossover study, the BY variables are phase, study drug, subject, and sequence number. The initial data are flagged according to the study-specific criteria for assay sensitivity, estimated values, diluted values, spurious values, and missing values. In so doing, the concentration values can be differentially handled later in the program (e.g., appropriate footnotes can be output on the final data display, erroneous values omitted from PK parameter calculations, etc.).

The values of the new variables are retained for the arrays STIM (scheduled time), ATIM (actual time), CONC (concentration value), and FLGS (concentration flags).

At the first occurrence of each subject, the pointer (PTR) is set to zero. The pointer is then incremented for each incoming record and represents the index value for the above arrays.

Thus, the array is only filled for the number of scheduled collection times for each subject during each treatment leg. The values are held in the above arrays until the last occurrence of the subject and are then output to the SUBJECT data set.

After each subject record is output, all the arrays are initialized to missing in preparation for the next subject.

At the end of the data step, all variables which are not needed (e.g., PTR, I) or no longer represent correct information (e.g., CONCA) are deleted.

Descriptive Statistics:
PROC MEANS is used on the SUBJECT data set by phase and treatment to produce the statistics requested by the individual study monitor. This has been made generic by using the global macro variable PTIMES. Placing each statistic on a separate line also allows it to become a comment if it is not required for a particular study. The macro variable calculation for ENDPT is adjusted for the total number of statistics requested and then becomes the dynamic allocation of the explicit array subscript for rounding purposes. For example, a request for the mean, standard deviation, and coefficient of variation gives us three times the total number of expected concentrations. four, for the study, for a total of 12 array elements. Thus, the program code is easy to change and maintain.

Final Data Display Production:
The data can now be printed by concentration time by treatment phase. In order to prepare the concentration values for printing, the data are rounded before output to the page utilizing the PTIMES macro variable.

This report will be paged by treatment and the PAGE flag is initially set to zero. The column header information will only print at the top of each new page. Utilizing the global variables and the formats defined earlier (e.g., PTIMES), the PUT statement is merely five lines long. This is more efficient for SAS® to process and for the programmer to change as necessary.

Instead of merging the data set output from the PROC MEANS as some programmers do, the...
MNDAT data set is merely set after the last concentration value for each treatment phase. This method avoids the extra processing of merging the two data sets.

Any applicable footnotes are printed after the last subject in a treatment group but before the page is released for the next treatment group. The footnotes can be printed all at once or conditionally according to the preference of the individual study monitor or biometrician.

The data display header is defined using the HEADER label instead of a LINK command. This eliminates the problem of retaining or lagging the values of any variables used to create the header because SAS® creates the header before creating the rest of the page. Here the macros defined earlier are used to customize the data display header and page placement.

Graph Production:
In order to ease the production of graphs, global macro variables are defined for the X and Y labels, intervals, minimums, and maximums. As was done for the data display program, the global variables are defined at the beginning of each major program section. This allows the graphic portion of the program to be self-contained and to be run as a separate program as necessary.

Since these graphs will be produced for plasma drug concentrations for each subject using PROC GPlot®, the original PLASMA data set is used. For the purposes of this study's biometrician, scheduled concentration times are converted from 24-hour clock to minutes and missing values are skipped.

Next, graphs of the treatment means for the plasma drug concentrations are produced using the same data set as the subject graphs. PROC GPlot® has an interpolation option that will produce means, standard deviations, and standard errors of the mean when the entire SAS® data set is passed to the procedure. Allowing PROC GPlot® to calculate the means is the only way that PROC GPlot® will display the standard error bars. In order to prevent the display of all data points used in calculating the means, the V option is set to NONE. If shortening the axis length is desired, then the MODE must be set to INCLUDE so that data points outside the current axis length are still included in the calculation of the means.

Conclusion:
Data display production can be made simpler and easier to maintain using global macro variables, formats, and calculated string placements.

Acknowledgements:
SAS is the registered trademark of SAS Institute, Inc., Cary, NC, USA.
MACROS FOR PRINTING

%MACRO HEADCOL(HCOL=,HEADER=,LSIZE=132);
%* THIS MACRO WILL CALCULATE THE STARTING COLUMN POSITION OF
A TEXT STRING GIVEN THE SIZE OF THE LINE (PASSED AS LSIZE)
AND THE STRING (PASSED AS PTI,PT2) SO THAT IT IS CENTERED ON
THE LINE AT POSITION HCOL WITH AN @ IN FRONT OF IT AND THE
TEXT STRING IN QUOTES;
%*** THE HEADING TO BE PRINTED IS PASSED WITHOUT QUOTES;
&HCOL=INT((&LSIZE - LENGTH(&HEADER)) / 2));
%MEND HEADCOL;

%MACRO ZPRINT(SIZE=MAX,DSN=,TITL=);
%* THIS MACRO WILL PRINT A SAS DATA SET GIVEN THE NAME OF THE
DATA SET (PASSED AS DSN) AND THE NUMBER OF OBSERVATIONS (PASSED
AS SIZE) AND A TITLE (PASSED AS TITL);
%*** THE TITLE IS PASSED WITH QUOTES;
OPTIONS OBS=&SIZE;
PROC PRINT DATA =&DSN;
   TITLE3 &TITL;
   OPTIONS OBS=MAX;
RUN;
%MEND ZPRINT;
MACRO VARIABLES FOR DATA MANIPULATION

/* ASSIGN VARIABLE NAMES FOR PARTICULAR STUDY TO GLOBAL VARIABLES */

%LET TITLE1 = "ANY PHARMACOKINETICS PROTOCOL";
%LET TITLE2 = "EXAMPLE FOR NESUG";
%LET INVEST = INVEST; /* INVESTIGATOR IDENTIFIER */
%LET SUBNO = SUBNO ; /* SUBJECT NUMBER */
%LET SEQNO = SEQNO ; /* SEQUENCE NUMBER */
%LET PHASE = PHASE ; /* STUDY PHASE */
%LET TREATC = TREATC; /* TREATMENT, CHARACTER */
%LET DOSE = DOSE ; /* DRUG DOSE OR PLACEBO */
%LET TIME = TIME ; /* SCHEDULED TIME OF ASSAY */
%LET ATIME = ATIME ; /* ACTUAL TIME OF ASSAY */
%LET FLAG = FLAG ; /* CONCENTRATION FLAG */
%LET CONCA = CONCA ; /* CONCENTRATION, FIRST */
%LET PTIMES = 004 ; /* NUMBER OF PLASMA CONCENTRATION PER STUDY */
%LET LIMIT = 0.0 ; /* LIMIT OF ASSAY SENSITIVITY FOR THIS DRUG */
%LET ENDP = %EVAL(3*&PTIMES); /* NUMBER OF STATS USED PER CONC */
%LET FCOL = 10 ; /* STARTING COLUMN FOR FOOTNOTES */

/*========================================*/
/* CREATE FORMATS - THESE WILL BE STUDY DRUG SPECIFIC */
/*========================================*/

PROC FORMAT;
/* PRINTING FORMAT FOR SUBJECT VARIABLE */
PICTURE SUBF 0-HIGH = "0009"
;
/* PRINTING FORMAT FOR 24 HOUR CLOCK TIME */
PICTURE TIMEF 0800 = "8:00 AM" (NOEDIT)
  2000 = "8:00 PM" (NOEDIT)
;
/* PRINTING FORMAT FOR ALL PLASMA CONCENTRATION VARIABLES */
PICTURE CONF 0-HIGH = "0009.9"
  . = " NAV"
;
/* PRINTING FORMAT FOR SUMMARY STATISTICS */
PICTURE NF 0-HIGH = "0000009"
  . = " NA"
;
PICTURE SUMF 0-HIGH = "0009.99"
  . = " NA"
;
/*========================================*/
/* RESHAPE PLASMA DATA BY SUBJECT TO PRODUCE DESCRIPTIVE STATS */

DATA SUBJECT:
    SET PLASMA:
    BY &PHASE &TREATC &SUBNO &SEQNO;
    RETAIN
        STM001-STM&PTIMES /* SCHEDULED TIMES FOR CONCENTRATIONS */
        ATM001-ATM&PTIMES /* ACTUAL TIMES FOR CONCENTRATIONS */
        CON001-CON&PTIMES /* PLASMA CONCENTRATIONS */
        FLG001-FLG&PTIMES /* CONCENTRATION FLAGS */
    ARRAY STIM (&PTIMES) STM001-STM&PTIMES;
    ARRAY ATIM (&PTIMES) ATM001-ATM&PTIMES;
    ARRAY CONC (&PTIMES) CON001-CON&PTIMES;
    ARRAY FLGS (&PTIMES) S01 FLG001-FLG&PTIMES;
    /* RESHAPE DATA BY SUBJECT */
    IF FIRST.&SUBNO THEN PTR = 0:
    PTR+l:
        STIM{PTR} = &TIME:
        ATIM{PTR} = &ATIME:
        CONC{PTR} = &CONCA:
        FLGS{PTR} = &FLAG:
    IF LAST.&SUBNO THEN DO:
        OUTPUT:
        /* AFTER EACH SUBJECT (ROW), RE-INITIALIZE ARRAYS */
        DO I = 1 TO &PTIMES :
            STIM {I} = .:
            ATIM {I} = .:
            CONC {I} = .:
            FLGS {I} = ".":
        END:
    END:
    DROP /* REMOVE UNNEEDED VARIABLES */
        PTR I &TIME &ATIME &CONCA &FLAG &SEQNO ;

/* STATISTICS */

PROC MEANS DATA = SUBJECT NOPRINT;
    BY &PHASE &TREATC ;
    OUTPUT OUT = MNDAT
        N = N001-N&PTIMES
        MEAN = MN001-MN&PTIMES
        STD = SD001-SD&PTIMES
        CV = CV001-CV&PTIMES
        VAR CON001-CON&PTIMES:

Pharmaceutical Industry
/*
CREATE PLASMA CONCENTRATION TABLE */

DATA _NULL_:;
SET SUBJECT END=EOF;
BY &PHASE &TREATC &SUBNO ;
FILE PRINT NOTITLES PAGESIZE=58 N=PS HEADER=PART1 LINESLEFT=LL;
/* PAGE REPORT BY TREATMENT */
PAGE = 0;
IF FIRST.&TREATC OR LL < 8 THEN DO;
   PAGE = 1;
   PUT _PAGE_ @;
END;
/* INVOKE MACRO TO CALCULATE STARTING COLUMN */
%STARTCOL(MAXCT=4):*MAXCT IS THE TOTAL NUMBER OF COLUMNS;
IF PAGE THEN PUT
//@SCOL+10 "ANY DRUG PLASMA CONCENTRATION (NG/ML)"
//@SCOL "SUBJECT"
//@SCOL+10 STM001 TIMEF. (STM002-STM&PTTIMES) (+2 TIMEF.)
/:;
PUT @SCOL &SUBNO SUBF.
@SCOL+11 CON001 CONF. (CON002-CON&PTTIMES) (+3 CONF.)
@SCOL+17 FLG001 $1. (FLG002-FLG&PTTIMES) (+8 $1.);
/* AT THE END OF EACH TREATMENT, BRING IN STATISTICS DATA SET */
IF LAST.&TREATC THEN DO;
   SET MNDAT;
<<<<<<<<<<< Code inbetween >>>>>>>>>>>>
/* >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
/* DEFINE TABLE HEADER */
/* >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
PART1:
/* INVOKE MACRO TO CALCULATE STARTING COLUMN */
%MAKEHEAD;
PUT
  # 7 @HCOL1 HEAD1
  # 8 @HCOL2 HEAD2
  #10 @HCOL3 HEAD3
  #11 @HCOL4 HEAD4
;//

-------------------- Code inbetween ---------------------
/* CUSTOMIZE DATA FOR GRAPHIC OUTPUT */
/* GRAPHS OF SUBJECT DATA */
/* DEFINE GLOBAL VARIABLES AND VALUES FOR SUBSEQUENT PROGRAM USE */

%GLOBAL
YLABEL YMAX YMIN YINT XLABEL XMAX XMIN XINT :
/* DEFINE MACROS FOR LABEL, MINIMUM VALUE, MAXIMUM VALUE */
/* AND INTERVAL FOR X AND Y AXIS. */
%LET YLABEL = "PLASMA CONCENTRATION (NG/ML)"
%LET YMAX = 5.5;
%LET YMIN = 0;
%LET YINT = .2;
%LET XLABEL = "TIME IN MINUTES"
%LET XMAX = 2100;
%LET XMIN = 750;
%LET XINT = 100;
/* DEFINE TITLES */
TITLE1 H=.6 CM FONT=DUPLEX &TITLE1;
TITLE2 H=.6 CM FONT=DUPLEX &TITLE2:
TITLE3 H=.6 CM FONT=DUPLEX &TITLE3:

/* AXIS = LINE, TICK MARKS, VALUES AND LABELS ASSOCIATED WITH A VARIABLE
   LEGEND = SHAPES, VALUES AND LABEL ASSOCIATED WITH A LEGEND VARIABLE IN A PLOT
   LABEL = TEXT AND ATTRIBUTES ASSOCIATED WITH A VARIABLE IN A PLOT
   VALUE = TEXT AND ATTRIBUTES ASSOCIATED WITH THE VALUES OF A VARIABLE IN A PLOT */
PROC  GPLOT DATA = PLOTS:
LEGEND1 LABEL = NONE
   SHAPE = SYMBOL(2.2):
   AXIS1 ORDER = (&YMIN TO &YMAX BY &YINT)
   VALUE = (FONT=DUPLEX HEIGHT=1)
   LABEL = (FONT=DUPLEX A=90 R=0 HEIGHT=1.5 &YLABEL)
   MINOR = NONE ;
   AXIS2 ORDER = (&XMIN TO &XMAX BY &XINT)
   VALUE = (FONT=DUPLEX HEIGHT=1)
   LABEL = (FONT=DUPLEX HEIGHT=1.5 &XLABEL)
   MINOR = NONE ;
PLOT &CONCA * &TIME = &TREATC/ SKIPMISS HAXIS=AXIS2 VAXIS=AXIS1;
   BY &SUBNO:
/* W= WIDTH OF LINE
   MODE=INCLUDE OBSERVATIONS OUTSIDE AXIS RANGE FOR INTERPOLATION
   I=INTERPOLATION (LINES CONNECTED, SMOOTHED, ETC.),
   STDKXXX= CALCULATES MEANS AND STANDARD DEVIATIONS OF Y VALUES
   K = NUMBER OF STANDARD DEVIATIONS, X = WHERE M IS THE STANDARD ERROR OF THE
   WHERE J CONNECTS MEANS FROM BAR TO BAR, WHERE T PUTS TOPS/BOTTOMS ON BARS
   V= PLOT CHARACTER TO BE USED WITH EACH UNIQUE SYMBOL, L=SOLID OR DASHED LINE, I=
   SOLID */
   SYMBOL1 WIDTH=1 MODE=INCLUDE I=JOIN V=CIRCLE L=1;
   SYMBOL2 WIDTH=2 MODE=INCLUDE I=JOIN V=SQUARE L=2;
   SYMBOL3 WIDTH=3 MODE=INCLUDE I=JOIN V=TRIANGLE L=3;
   SYMBOL4 WIDTH=4 MODE=INCLUDE I=JOIN V=STAR L=4;

---------------------------------------------------
A:\WP5\PAPERS\INESUG.PAP  10/10/95
/* GRAPHS OF TREATMENT MEANS */

PROC SORT DATA = PLOTS;
   BY &TREATC &PHASE &TIME;
RUN;

/* DEFINE MACROS FOR LABEL, MINIMUM VALUE, MAXIMUM VALUE */
/* AND INTERVAL FOR X AND Y AXIS. */
%LET YLABEL = "MEAN PLASMA CONCENTRATION (NG/ML)";
%LET YMAX = 4.0;
%LET YMIN = 0 ;
%LET YINT = .2;
%LET XLABEL = "MINUTES AFTER DOSING";
%LET XMAX = 2100;
%LET XMIN = 750;
%LET XINT = 100;
RUN;

/* RUN PLOTS */

PROC GPLOT DATA = PLOTS;
   LEGEND1 LABEL = NONE
      SHAPE = LINE(4);
      AXIS1 ORDER =(&YMIN TO &YMAX BY &YINT)
      VALUE = (FONT=DUPLEX HEIGHT=1)
      LABEL = (FONT=DUPLEX A=90 R=0 HEIGHT=1.5 &YLABEL)
      MINOR = NONE;
      AXIS2 ORDER =(&XMIN TO &XMAX BY &XINT)
      VALUE = (FONT=DUPLEX HEIGHT=1)
      LABEL = (FONT=DUPLEX HEIGHT=1.5 &XLABEL)
      MINOR = NONE;
      PLOT &CONCA * &TIME = &TREATC/ LEGEND=LEGEND1
      SKIPMISS HAXIS=AXIS2 VAXIS=AXIS1;
      W= WIDTH OF LINE
      MODE=INCLUDE OBSERVATIONS OUTSIDE AXIS RANGE FOR INTERPOLATION
      I=INTERPOLATION (LINES CONNECTED, SMOOTHED, ETC.),
      STDKXXX = CALCULATES MEANS AND STANDARD DEVIATIONS OF Y VALUES
      K = NUMBER OF STANDARD DEVIATIONS
      X = WHERE M IS THE STANDARD ERROR OF THE MEAN,
      WHERE J CONNECTS MEANS FROM BAR TO BAR,
      WHERE T PUTS TOPS/BOTTOMS ON BARS
      V=PLOT CHARACTER TO BE USED WITH EACH UNIQUE SYMBOL,
      WHERE NONE PREVENTS DISPLAY OF ALL DATA POINTS USED IN
      CALCULATING THE MEAN
      L=SOLID OR DASHED LINE, 1= SOLID */
      SYMBOL1 WIDTH=1 MODE=INCLUDE I=STD1JMT V=NONE L=1;
      SYMBOL2 WIDTH=2 MODE=INCLUDE I=STD1JMT V=NONE L=2;
      SYMBOL3 WIDTH=3 MODE=INCLUDE I=STD1JMT V=NONE L=3;
      SYMBOL4 WIDTH=4 MODE=INCLUDE I=STD1JMT V=NONE L=4;