Use of SAS® Software to Process and Analyze Skin Tumor Response Data From a Data Acquisition System.

R. Dana Laurie, Judy Stober, Thomas Woennendorp and Freda Truman,
Toxicology and Microbiology Division, Health Effects Research Laboratory
U.S. Environmental Protection Agency
Cincinnati, Ohio 45268

ABSTRACT

SAS software is being used to check and summarize vast quantities of data. The data comes from computerized collection of data in an animal housing facility by a Hewlett Packard 9825 "Calculator". The data is collected weekly for a period of one year and consists of location, size and type of each skin tumor observed. In addition, study number, treatment number, cage number, animal number, week of the study, person responsible for the data, date that each observation was made, body weight of the mouse and size, location and type of each observed skin tumor. A typical study would contain 600 mice in 15 to 20 groups.

In addition to 9825s, we also had a hardwire line from an HP 9825 to a PDP 11/70 mainframe located in our Cincinnati facility. In turn, the PDP 11/70 had communication capability with the Agency IBM 370 located in North Carolina. The point is that with available equipment and a fair amount of effort, a system was designed in a rather simple manner that is currently handling 8000 mice and producing over 4 million pieces of information per year. SAS software became an integral part of the system because it could process the data in a manner that was acceptable to both immediate study management personnel and upper management people responsible for making higher level decisions based upon the most recent data.

One of the major difficulties we thought we would encounter is that the HP 9825 program would need to be debugged in a manner that would assure the quality of the data that was being collected. The program on the following page (PROGRAM CHECK) was written to be used to check the collected data. The basic program was developed to be compatible with any of the ongoing individual study data bases. This is accomplished by simply changing 'XX' to the study number for which one needs the report and 'YY' to the maximum number of tumors observed on any one individual mouse.

Basically, PROGRAM CHECK examines the number of occurrences of each variable. In addition, it checks for two specific problems that we anticipated with the HP 9825 system. First, we knew that, at times, a mouse would be in a study, but for some reason, did not get weighed. Second, after an animal died, subsequent tumor size should appear as zeroes. However, in some instances, tumor size would be greater than zero. By using this program and spot checking data by comparing HP 9825 printouts with individual data in the SAS database, we were able to assure the quality of the collected data. (There were examples of printouts at the Foster Session.)

The second program, PROGRAM SUMMARIZE, was also designed to be used with any of the many studies in progress. It generates a summary for each date of observation for each treatment group in the study of interest. A toxicologist can examine these reports and be quickly apprised of the status of a particular experiment. (There were examples of printouts at the Foster Session.)
PROGRAM CHECK

/*FDP11: RO=DOB:[330,13]CHECKXX  PRI=BAN=CHECKXX 
/*JOBPARM LINES=50
/*STEPE EXEC SAS,REGION=1500K
/*# PUT ANY I/O FILES HERE
/*STEPE DD DSN=LDRA24.AXXALLA,DISP=OLD,UNIT=DISK
/*SYSIN DD *

*******************************************************************

** THIS PROGRAM IS SAVED AS 

*******************************************************************

* INSTRUCTIONS FOR RUNNING WEEKLY CHECK PROGRAM. 

* 1) CHANGE 'XX' TO 'STUDY NUMBER'.

* 2) CHANGE 'YY' TO 'MAXIMUM NUMBER OF TUMORS IN CURRENT STUDY'.

*******************************************************************

* BRING IN RAW DATA AND SORT.

*******************************************************************

DATA RAW; SET STEPB.AXXTRY;
PROC SORT; BY ID WEEK;

*******************************************************************

* PRINT RAW DATA AS A REFERENCE.

*******************************************************************

PROC PRINT; BY ID;
TITLE SENDAR MOUSE SKIN DATA CHECKS - STUDY XX;

*******************************************************************

* CALCULATE FREQUENCIES OF VARIABLES.

*******************************************************************

PROC FREQ; TABLES STUDY GROUP CAGE WEEK DATE LOC1-LOCYY T1-TYY;

*******************************************************************

* CHECK THE NUMBER OF OBSERVATIONS FOR DATES AND WEEKS.

* THE NUMBER OF DATES AND WEEKS PER ID SHOULD BE EQUAL FOR

* ALL IDS.

*******************************************************************

PROC SUMMARY; BY ID; CLASS WEEK; VAR DATE; OUTPUT OUT=CK1 N=NUM;
DATA; SET CK1; IF NUM>1;
PROC PRINT;
65. PROC SUMMARY DATA=A; BY ID; CLASS DATE; VAR DATE;
66. OUTPUT OUT=CK2 N=NUM;
67. DATA; SET CK2; IF NUM>1;
68. PROC PRINT;
69.*******************************************************************
70. * CHECK THE NUMBER OF LIVE ANIMALS THAT ARE MISSING WEIGHTS. *
71.*******************************************************************
72. PROC SORT DATA=A; BY ID DESCENDING WEEK;
73. DATA WTJSET A; BY ID;
74. WTCK=LAG(WEIGHT);
75. IF FIRST.ID THEN RETURN;
76. IF WEIGHT=. AND WTCK NE . THEN OUTPUT;
77. PROC PRINT;
78. TITLE2 ANIMALS ALIVE WITH MISSING WEIGHTS;
79.*******************************************************************
80.*******************************************************************
81. DATA SIZE; SET A;
82. IF WEIGHT = .;
83. IF SIZE1>0 OR SIZE2>0 OR SIZE3>0 OR SIZE4>0 OR SIZE5>0 OR SIZE6>0;
84. PROC PRINT;
85. TITLE2 ANIMALS WITH MISSING WEIGHTS AND TUMOR SIZE OBSERVATIONS;
86.------------------------------------------
87. PROGRAM SUMMARIZE
88.------------------------------------------
89. 1. //ELN JOB (HERLBDSP, MELN), LAURIE, TIME=5
90. 2. //PDP11: RO=DB0; [330,13] SUMMX.PB1/BAN=SUMMX
91. 3. //STEPA EXEC SAS
92. 4. //* PUT ANY I/O FILES HERE
93. 5. //STEPB DD DSN=LDRA424.AXXALLA,INIT=DISK, DISP=OLD
94. 6. //STEPC DD DSN=LDRA424.AXXSUMA, INIT=DISK, DISP=OLD
95. 7. //SYSIN DD *
96. 8. OPTIONS LINESIZE=79;
97. 9.*******************************************************************
98. * THIS PROGRAM IS SAVED AS 
99.*******************************************************************
100.*******************************************************************
101.*******************************************************************
102.*******************************************************************
103.*******************************************************************
104.*******************************************************************
105.*******************************************************************
106.*******************************************************************
107.*******************************************************************
108.*******************************************************************
109.*******************************************************************
110.*******************************************************************
111.*******************************************************************
112.*******************************************************************
113.*******************************************************************
114.*******************************************************************
115.*******************************************************************
116.*******************************************************************
117.*******************************************************************
118.*******************************************************************
119.*******************************************************************
120.*******************************************************************
DATA RAW; SET STEPB.AXXTRY; 
IF WEEK<ZZ-12 THEN DELETE; 
PROC SORT; BY GROUP ID; 
*****************************A************************************* 
DATA NUMBER; SET RAW; BY GROUP ID; 
KEEP GROUP ID; 
IF FIRST.ID THEN OUTPUT; 
PROC FREQ; TABLES GROUP/NOCUM OUT=GROUPCT; 
TITLE NUMBER OF ANIMALS PER GROUP; 
**************************************************C**************; 
DATA INDIVWK; MERGE RAW GROUPCT; BY GROUP; 
KEEP ID GROUP WEEK CURCNT CUMCNT DEAD INITN CURCAR CUMCAR; 
INITN=COUNT; 
*******************************************************************; 
ARRAY TYPE (I) T1-TYY; 
ARRAY SIZE (I) SIZE1-SIZEYY; 
CURCNT=0; CURCAR=0; 
CUMCNT=0; CUMCAR=0; 
DO OVER TYPE; 
IF TYPE=2 OR TYPE=4 OR TYPE=6 OR TYPE=8 
THEN CURCNT=CURCNT+1; 
IF TYPE=1 OR TYPE=2 OR TYPE=6 
THEN CURCNT=CURCNT+1; 
IF TYPE=6 
THEN CURCAR=CURCAR+1; 
IF TYPE=6 OR TYPE=8 
THEN CUMCAR=CUMCAR+1; 
END; 
DEAD=0;
IF WEIGHT=. THEN DEAD=1;

PROC SORT; BY GROUP WEEK;
TITLE THIS IS THE WEEKLY SUMMARY REPORT FOR STUDY XX.;

******************************************************************
* CALCULATE WEEKLY SUMMARY COUNTS.
******************************************************************;

DATA GROUPWK; SET INDIVWK END=EOF; BY GROUP WEEK;
RETAIN TWEEK 1 SUMCUR SUMCUM NTUM CUMNUM NCAR TCURCAR TCURCAR TGROUP NALIVE;
DROP ID WEEK GROUP CURREN CUMCUM DEAD INITN ;
IF WEEK NE TWEEK OR GROUP NE TGROUP THEN DO;
  IF GROUP HE . THEN OUTPUT;
  TWEEK=WEEK; TGROUP=GROUP; NALIVE=INITN;
  SUMCUR=CURCNT;
  SUMCUM=CUMCNT;
  TCURCAR=CURCAR;
  TCURCAR=CURCAR;
  IF CURCNT>0 THEN NTUM=1; ELSE NTUM=0;
  IF CUMCNT>0 THEN CUMNUM=1; ELSE CUMNUM=0;
  IF CURCAR>0 THEN NCAR=1; ELSE NCAR=0;
  IF CUMCAR>0 THEN TWCAR=1; ELSE TWCAR=0;
END;
ELSE DO;
  SUMCUR=SUMCUR+CURCNT;
  SUMCUM=SUMCUM+CUMCNT;
  IF CURCNT>0 THEN NTUM=NTUM+1;
  IF CUMCNT>0 THEN CUMNUM=CUMNUM+1;
  TCURCAR=TCURCAR+CURCAR;
  TCURCAR=TCURCAR+CUMCAR;
  IF CURCAR>0 THEN NCAR=NCAR+1;
  IF CUMCAR>0 THEN TWCAR=TWCAR+1;
END:
NALIVE=NALIVE-DEAD;
IF EOF THEN OUTPUT;
LABEL NALIVE='NUMBER ALIVE'
NTUM='CURRENT*NO. WITH TUMOR'
SUMCUR='CURRENT*TUMOR COUNT'
SUMCUM='CUMULATIVE*NO. WITH TUMOR'
TGROUP='GROUP'
TCURCAR='CURRENT*CARCINOMA*COUNT'
TCURCAR='CUMULATIVE*CARCINOMA*COUNT'
NCAR='CURRENT*NO. WITH*CARCINOMA'
TWCAR='CUMULATIVE*NO. WITH*CARCINOMA';

PROC SORT; BY TGROUP TWEEK;
DATA SUM; SET STEPSUM;
PROC SORT; BY TGROUP TWEEK;
DATA STEPCSUM;MERGE SUM GROUPWK; BY TGROUP TWEEK;

******************************************************************
* PRINT SUMMARY TABLES.
******************************************************************;
A third program has been written for all studies, but because of space limitations is not presented within the Proceedings. However, it was presented at the Poster Session. The third program is generated for each particular study and includes a small data base containing the study's protocol (e.g. test compound, route of exposure, dose administered, initial group size). In addition to the protocol, the most frequently requested data (cumulative tumors) are presented in the report right after the protocol. Next, graphs of important relationships between groups are presented, followed by each group's specific data in detail (similar to PROGRAM SUMMARIZE format). These reports serve to keep upper management up-to-date on study progress and results. Importantly, they essentially eliminate lag time between data collection and reports in a brief and consistent format which include graphical presentation of the data. The reports from this third set of programs are transferred to a Wang computer system.

SUMMARY

This poster presentation outlined programs which form a complete data management system that enables us to follow 8000 mice per week in terms of data collection, verification and report production. The programs were designed to be used for similar databases and can be easily altered to accommodate different experiments. Output examples were on display at the conference to show what type of reports the programs produce.