

A CONVERSATIONAL FRONT END TO SAS

Julian Horwich and Mark Phelps
Abbott Laboratories

ABSTRACT

A conversational data entry and data management system for Drug Clinical Trial data has been operational at Abbott for the past year. It provides a direct interface into S.A.S., even though it runs on a DECsystem-10 computer. Data may be entered via various CRT or hard-copy keyboard terminals.

INTRODUCTION

Before a drug may be approved for entry into the marketplace, it must go through extensive animal and human testing. The latter is accomplished through what is called a clinical trial. Volunteers are given the new drug by physicians who record the data on "Case Report Forms". These forms are then forwarded to Abbott. The data is then entered into a digital computer to produce formatted reports, statistical analyses, and graphs. The system described below has been written to handle the entry and management of this data prior to its use in SAS, EDMP and other statistical packages. Its aim is to eliminate much of the problems and human time involved in handling this data.

DICTIONARY BUILDING

While the most efficient system (from the computer standpoint) would be one that was coded by a programmer for a particular drug study, this is not practical from a human standpoint. Our aim is to eliminate the need for computer programmers in both the data entry and the analysis. In the case of the analysis, much of the work is now done by statisticians or data technicians with very little aid from computer specialists. The aim of the data entry system is to provide the same operating philosophy.

A program was written to, conversationally, build a dictionary for each drug study. This is used to drive the data entry, control the data management and provide the dictionary input to S.A.S. The data technicians are asked various questions by the program about their data. Questions asked are based on the previous answers. The questions are designed so that the most common answer is simply to depress the carriage return key.

Figure 1 indicates the attributes of each data item. After a study was made of past case report forms it was found that all data could be broken down into the three types shown.

Every piece of data is given a unique six character name to be used for entry, corrections, verifications, and statistical analysis. If a longer name (up to 70 characters) is desired for the entry of data an optional prompt name is given. An optional coding prompt is provided to display coding information during entry (e.g. 1=MALE, 2=FEMALE). For character items only, a size is given of 1 to 70 characters. Optionally, editing information may be provided.

ACCURACY CHECKS

Rather than relying completely on visual verification of input, the system allows for editing as each item is entered. Error messages are displayed if an edit check occurs and the user may either retype the data or ignore the warning. A second accuracy check is key verification. The data is keyed a second (or third or...) time and compared against what was previously

ACCURACY CHECKS, cont.

entered. If a difference occurs, the user may correct the file at that point. If errors are discovered at a later time, corrections may be applied instantly. Corrections may also be key verified. In addition, listings may be requested of all or selected portions of the data for further checking.

DICTIONARY MODIFICATION

During a drug study it may be necessary to change information in the dictionary to reflect changes in the data required for a given study. A program named MODIFY will permit changes to the dictionary at any point in a study. Specific characteristics of a data item can be altered, an existing item can be deleted, or a new item can be added to the dictionary.

For all data items the name used to access the data can be replaced by a new name. The prompts used to enter the data can be changed, deleted, or added. The editing information used in entering the data can be modified, added to or deleted from the dictionary.

For numeric data items the data type can be changed such that a one-digit item can be converted to either a 2-7 digit item or a character item and a 2-7 digit item can be converted to a character item.

For character items the length of the item can be changed to allow more characters or less characters in entering the data.

Most changes allowed in the MODIFY program result in modifying only the dictionary and related files. When changing either the type of numeric items, the length of character items, or adding a data item, it is necessary for the data file, if data has been entered, to be restructured. This permits space to be allocated for converted data items or new items to be subsequently entered.

The MODIFY Program saves all original files for recovery procedures and initiates an automatic compress of the data file to delete unused data items. It also produces a printed report of the changes.

CARD INPUT

The system may also accept card input produced by keypunches or Abbott's OPSCAN Mark-sense device. Data may then be key-verified or corrected via CRT.

DOCUMENTATION

The dictionary provides automatic documentation of the data file. Every point of data carries flags that indicate whether data has been entered, edit error flagged, verified and/or corrected. The data last changed is stored for each logical grouping of data. Every function provides a printed audit trail of what was done by the data technician.

ARCHIVES

The ARCHIVE program allows the data technician to move the data from disk to two magnetic tape copies with only three type-ins. The data may be brought back to disk with the same procedure.

HUMAN INTERFACE

Much time has been spent working on the human-computer interface. Many times machine efficiency has been sacrificed to improve human efficiency. Many changes have been made based on suggestions by data technicians who have used the system. One example is an option to allow operators with bifocals to have the data displayed at the bottom of the screen rather than the top.

STATISTICAL ANALYSIS

The major objective of the system is statistics. So, at any point during or after the entry, the data may be transferred to an SAS internal format data set. After three type-ins a copy of the data file is sent down a high speed link between Abbott's DECSYSTEM-10 and Abbott's IBM370/158's. Assuming a restructure hasn't been done, only the data entered or corrected since the last transfer is sent. The data may be sent to a SAS-72 or SAS75 data set. Following this, the data may be analyzed statistically. Figures 2 and 3 show an example of a frequency distribution written against a data entry file. If the use of BMDP is required, PROC BMDP is used to interface to this package. The PUT statement may be used to interface to other packages. Various Abbott written extensions to SAS (procs and functions) may also be used. Transfers may be done at the time of request or the data technician may tell the computer to defer the transfer to midnight.

SUMMARY

CRT data entry and management, combined with SAS provides an effective method of handling data prior to its use by statistical software.

Data type (1=1 digit numeric, 2=2-6 digit numeric, 3=character item)

Data name

Prompt name

Coding prompt

Size if character item

Editing

Check against specific values

Check within range

Flag if not:

date

negative

positive

zero

blank

numeric

alphabetic

Figure 1 Data Attributes

Sample run made against a Data Entry file

```
// EXEC ABSAS75
//SYSIN DD *
DATA; SET DDE,M1091;
IF SEX = 1 THEN CSEX = 'MALE';
IF SEX = 2 THEN CSEX = 'FEMALE';
PROC FREQ; TABLES AGE*CSEX;
/*
//DDE DD DSN=ABT.RD.DJE.E997DE(0),DISP=SHR
```

Figure 2 SAS program

TABLE OF AGE BY CSEX

| AGE | CSEX | | TOTAL |
|-----------|--------|--------|--------|
| FREQUENCY | FEMALE | MALE | |
| PERCENT | | | |
| ROW PCT | | | |
| COL PCT | | | |
| 23 | 1 | 2 | 3 |
| | 6.25 | 12.50 | 18.75 |
| | 33.33 | 66.67 | |
| | 20.00 | 18.18 | |
| 24 | 1 | 5 | 6 |
| | 6.25 | 31.25 | 37.50 |
| | 16.67 | 83.33 | |
| | 20.00 | 45.45 | |
| 25 | 1 | 2 | 3 |
| | 6.25 | 12.50 | 18.75 |
| | 33.33 | 66.67 | |
| | 20.00 | 18.18 | |
| 27 | 2 | 0 | 2 |
| | 12.50 | 0.00 | 12.50 |
| | 100.00 | 0.00 | |
| | 40.00 | 0.00 | |
| 28 | 0 | 2 | 2 |
| | 0.00 | 12.50 | 12.50 |
| | 0.00 | 100.00 | |
| | 0.00 | 18.18 | |
| TOTAL | 5 | 11 | 16 |
| | 31.25 | 68.75 | 100.00 |

Figure 3 Output of above program