SAS AS A MANAGEMENT SYSTEM FOR ROUTINE ECHOCARDIOGRAMS

Glenn A. Augustine, Kenneth P. Offord, Vernon P. Weber, Emilio R. Giuliani, M.D.

Mayo Clinic, Rochester, Minnesota

This paper will discuss how SAS is used in a batch processing mode as a data management and retrieval system to handle the collecting of data from echocardiograms done at the Mayo Clinic. A discussion of the edit/update program written in SAS will be presented, as well as why SAS was selected, and the benefits it has provided.

Echocardiography is a non-invasive technique for examination of the heart by use of ultrasound. It was used sparingly in the late 1960's at Mayo, but in 1970 when the ultrasound unit was interfaced with a strip chart recorder, and new techniques were mastered, it gained in popularity and usefulness. In 1971, only 158 echos were done, but in 1976, 4534 were done. The monthly average in 1977 through November was slightly over 400.

As echocardiography became more reliable and echo activity increased, the echo lab decided that there was a definite need to edit the data thoroughly and be able to reference and retrieve it easily.

In 1974 work began on this system. The flow of the system is as follows and is flowcharted in Figure 1. The information is abstracted onto a worksheet and then keypunched onto three punch cards. These cards are then submitted to the edit/update program. All of the raw information on the cards is listed, and then it is edited using SAS. Only records that error free are added to the master SAS data set, thus keeping the master clean. A list of errors is printed for each record and sent to the echo lab for corrections. Once completed, the data cards are corrected and resubmitted to the edit/update program. Those patients whose data pass the edit are also screened by the program for possible inclusion in over twenty echocardiographic abnormality groups, and a subset of the data is printed for each.

Several considerations were taken into account when deciding which software package or computer language should be selected for the edit/update program. First, the data would be processed in a batch mode, not real/time. Since the data was already being collected, it was desirable to have the routine written as soon as possible. Minimal interaction with the programmer was desirable for updating and retrieval. A meaningful name for each variable was desired to make requests for analysis easily understood by investigator, statistician, and programmer. The system should support diversified retrievals and statistical analyses. Finally, the ability to access the data by other systems in the future was a necessity.

To achieve these goals, SAS was selected. This system, originally written in SAS/72, afforded ease in debugging, as well as the necessary sort, merge, and update facilities. Only 35 programmer hours were required to get the 700 statement edit/update program operational, and a data clerk now spends about one hour each month handling the procedure with very little interaction with the original programmer. The SAS data set provides names for each variable understandable to everyone involved. Each eight-character name consists of 5 characters and 3 numbers. The 5 characters are an alphabetic

Figure 1. MONTHLY FLOW OF THE SYSTEM
abbreviation of the variable. The first number denotes on which of the original three cards the variable is located, and the last two numbers denote the beginning column number on that card (e.g., C.RHY130 is the variable name referring to cardiac rhythm from card 1 beginning in column 30).

Retrieval of the data set is a simple task with SAS. Listings and numerous statistical analyses can be performed easily on any subset desired, and the output can be thoroughly documented within the program. To allow for portability to other systems in the future, copies of the original punch cards are also stored as variables on the SAS data set.

With the advent of SAS/76, the system was converted in October, 1976. The conversion required only 11 programmer hours and incorporated the SAS/76 feature of multiple creation of data sets in a DATA statement, thus increasing the efficiency of the program. At present about 10 minutes of CPU time on an IBM 370/145 is used each month to edit over 400 new cases and update a file in excess of 11,000 echocardiographs with 98 variables.

There are a variety of edit checks made by the edit/update program. Since the data is keypunched, the cards are checked for matching Mayo Clinic number and date on all three cards. All items are checked for legitimate values. Over 90 checks are made in all. If an error is detected, it is printed with an associated error number for that check, the Mayo Clinic identification number, date, the items in error, and their values. In some cases an informative comment is also printed. The error number is used by the echo lab as a cross-reference with a dictionary of legitimate values for that number.

The benefits that the echo lab has received from the system have been numerous. At any given time, the master echocardiographic data set is error-free. Analysis of the data or retrieval of a subset of the data is a simple task. Many of these retrievals have been done by data clerks, and even a few have been done by secretaries. About 10 percent of the analyses and retrievals have been done by a statistician, 80 percent by data clerks under proper supervision, and 10 percent by programmers. The echocardiographic data set has also been interfaced with Mayo's surgical index and diagnostic reference files. Fourteen clinical research projects in various stages of completion have utilized this system for retrieval.

Having all of this information available has greatly reduced the number of medical charts that have to be reviewed. It is a simple matter with SAS to find out if there are enough patients with a given condition to merit investigation.

Another benefit of the system is the error listings generated by the edit/update routine. It provides feedback to the supervisors as to what errors are occurring most frequently in filling out the worksheet, and efforts can be

made to quickly eliminate these common errors. Initially, a large number of errors were made. Now less than .05 percent of the data items coded have errors.

This system has worked extremely well. The initial cost was low, the cost of the edit/update program is reasonable, and even though the data is on cards, very little time is required each month to process that data. It has proven to be a valuable tool for echocardiographic research.