A DATA MANAGEMENT SYSTEM FOR REMOTE MEDICAL STUDIES USING THE SAS® SYSTEM

Dwayne D. Oland, Timothy L. Cannon, David M. Warren, and Gene O. Nelson

U.S. Army Medical Research Institute of Infectious Diseases, Fort Detrick, Frederick, MD 21701-5011

ABSTRACT

The SAS® system is an integral part of an automated data management system currently being utilized to conduct clinical trials in foreign countries. The system, consisting of a portable microcomputer, an optical mark reader (OMR) and printer utilizes the SAS® system and dBASE III®+ to control the entry, management, analysis and graphing of medical research data. The SAS® system provides statistical analysis, graphics, the ability to perform on-site randomization, as well as an interface between the OMR and dBASE III®+. The wide variety of tools offered by the SAS® system give researchers and statisticians the capability to perform more extensive testing, monitoring and analysis of data at the actual site of the study.

BACKGROUND

Effective collection, verification, analysis and management of medical research data in foreign remote sites create a unique challenge for members of the research community. In the past, researchers collected data at remote sites on paper forms and returned them to a central computer facility where the data was keypunched, verified, and analyzed. More recently, researchers have deployed microcomputer systems at field sites to enhance the convenience and accuracy of data entry and to allow the local staff to better monitor the data collection effort.[1] Since most studies are conducted at smaller installations, clinics, and hospitals, the introduction of computer systems can become a strain on the small numbers of personnel at these institutions. The additional responsibilities of keyboard data entry, verification and management of large amounts of data often require that staff members work longer hours or direct time and effort from their regular duties. In an effort to alleviate the data entry burden and minimize the impact on local staff, we designed a data management system which includes a portable COMPAQ® microcomputer, a SCANTRON® full page optical mark reader (OMR), and an Okidata printer.[2] The OMR allows direct data entry from full page forms which are marked by physicians, health workers, or laboratory personnel. This greatly reduces the time that the local staff must spend on keyboard entry. It also improves the accuracy of the data by reducing the number of transcription errors and by providing a uniform data entry format. The SAS® system was used as the programming tool to create the software for the data entry, data analysis and graphics for the system, and functioned as an interface to dBASE III®+ (see Figure 1).

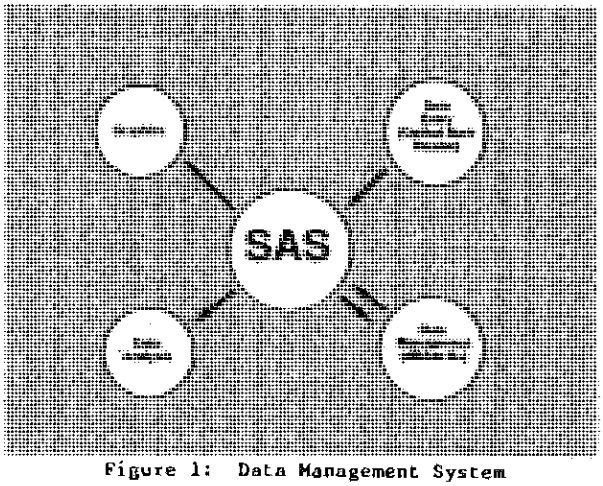

Figure 1: Data Management System

DATA ENTRY

Data entry is accomplished through the use of optical mark forms which are scored with a standard #2 pencil. Forms are designed so they closely approximate the paper forms that are currently in use at the remote site (see Figure 2). The content and number of forms vary for each study which necessitates creating unique decoding programs for each form. This process can be tedious and require large amounts of time to complete. In an effort to alleviate these problems, generalized template modules were written in the SAS® language to decode each of the most prevalent data types:

1. Vertical numbers
2. Horizontal numbers
3. Vertical alpha
4. Horizontal alpha
5. Vertical date
6. Horizontal date
7. Yes/no blocks

In order to generate the SAS® decoding program, one must assemble these prewritten SAS® modules, plus a module which converts the character string that is sent by the OMR into a 80 x 27 matrix of zeros and ones. Then one enters assignment statements identifying the type of field, starting row, starting column, size and variable name of each field on the form. A final module eliminates extraneous variables and creates the SAS® data set containing the values of all the fields on the form.

Although this original strategy has worked well in two large clinical trials, we are adding some enhancements to make the system faster and easier to use. We are currently developing an interactive menu-driven system, written in the C language, which will prompt and help the end user create the SAS® decoding program. Another concern that is being addressed, is the speed of the SAS® decoding program. It is felt that the newer 80386-based microcomputers will overcome

many of the speed problems. We are also evaluating the possibility of writing a decoding system based solely on the C language which would decode forms at a faster rate.

DATA MANAGEMENT

Once the form has been decoded, the SAS® system converts SAS® data sets into dBASE III®+ files using PROC DBF. A menu-driven dBASE system then provides facilities for verifying data, editing data and generating simple management reports. The data manager uses this system to keep track of all of the records associated with the individual patient. In addition it allows the manager to check for patient compliance with the protocol.

The ability to do most of the data screening at the field site is extremely advantageous. We have found on-site data quality control and quality assurance (QC/QA) by a database supervisor to be essential in reducing occurrence of outliers and missing values which can seriously compromise large field studies. During the data management stage, SAS® programs are used to monitor continuous variables collected throughout the study. SAS® programs, which utilize PROC FREQ, MEANS, CHART, and TABULATE, look for outliers or radical changes of parameters over time. They are also used to establish normal ranges for patient parameters, which may be drastically different due to varying environmental and dietary conditions. If data problems are not corrected as the data is being collected, the amount of time and effort required to correct these problems during later stages of the project is much greater. In some cases it becomes impossible to resolve the inconsistencies and missing values in the database. The on-site database administrator does not necessarily need to be a computer professional but must be absolutely committed to data integrity and have not only the authority but the tools to do the job.

DATA ANALYSIS

During the data collection phase of the study, the SAS® system is used to produce descriptive statistics, graphs, and frequency counts on a real-time basis which aid researchers in quality control and project management. Once data collection has been completed, the SAS® system is used to screen the data before the analysis phase begins. We have found that much of the statistical analysis can be performed on 80286-based microcomputers equipped with a hard disk, a math coprocessor, and additional random access memory (RAM) which is utilized by the SAS® system When larger data sets are processed or when more memory intensive analyses are required, the SAS® system data sets may be easily transferred to larger computer systems for further analysis. On

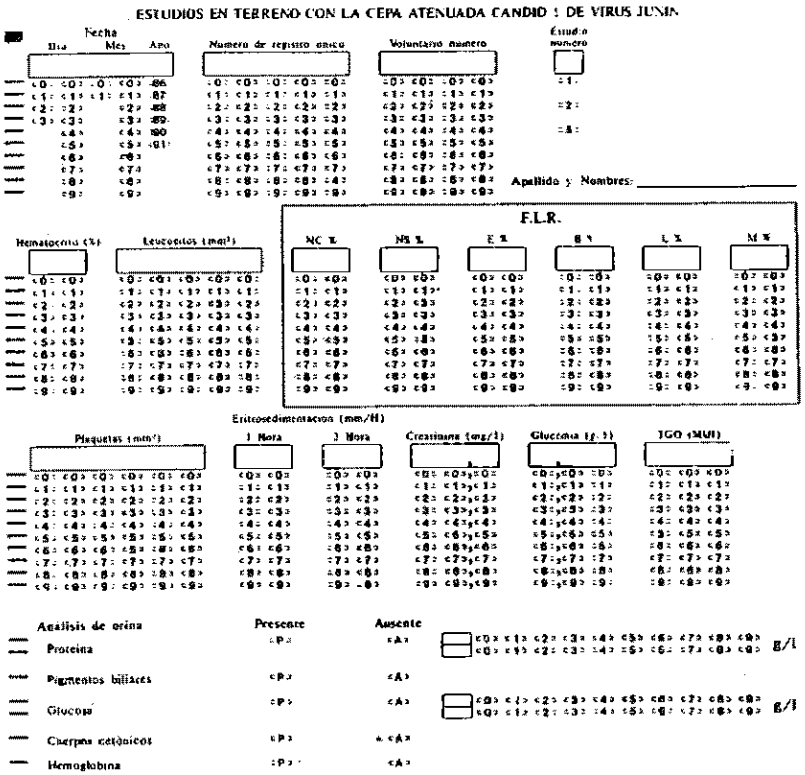

Figure 2: Portion of Clinical Laboratory Optical Mark Reader Form

the larger systems, other statistical analysis programs such as BMDP™ are sometimes utilized in conjunction with the SAS® system.

Although most of the analysis work is currently done by statisticians who are experienced SAS® users, we are formulating ways to make the analysis process easier for the researcher. Two methods currently being evaluated are the creation of menu-driven systems using SAS/AF® or by using SAS/ASSIST™.

GRAPHICS

Graphics play an important role in the evaluation of any scientific problem. The ability to transform tables of numbers into meaningful graphs and diagrams simplifies the analysis process. SAS/GRAPH® is used to create a wide variety of histograms, tables and graphs. Discrete data is compiled into histograms and tables for easy evaluation. Continuous time-oriented medical data on individual patients is sometimes displayed as a series of stacked charts which allow doctors to monitor several patient variables on the same screen[3]. These graphs are designed to pinpoint safety problems such as adverse reactions and trends. During the analysis phase the study, all variables are graphed according to treatment groups using SAS® macros and more sophisticated graphs are created to show interrelationships between variables.

All graphs at the field site are printed on the Okidata printer. Although graphs produced on a line plotter would be superior to the graphs on the printer, we have refrained from including a plotter with the system for logistical reasons. We have attempted to limit the amount of equipment in order to make the system portable, inexpensive and as durable as possible.

DISCUSSION

The first prototype of this system was deployed for use in the second phase of a clinical trial in the Peoples Republic of China[4]. This system proved to be more successful than a dBASE III®+ keyboard entry system used during the first phase of the study. Data entry proved to be faster and error rates were reduced. However, the system was operated by researchers from the United States who were experienced microcomputer users. It has also been deployed for all three phases of an ongoing clinical trial in Argentina. In this case, inexperienced Argentine personnel who were given a two-week training course, have successfully operated the system for over two years. This system is currently monitoring 5600 individuals who volunteered to participate in the efficacy phase of this vaccine protocol. It also catalogs data from an associated rodent capture project. While the field site personnel have become proficient in the data entry and management areas, most of the analysis and graphics have been done by statisticians and computer specialists from the United States. We are currently evaluating methods to correct this situation and make analysis and graphics easier to learn.

The SAS® system has proven to be a versatile tool at the remote site. It's various procedures have allowed us to randomize volunteers for double-blinded placebo controlled projects. It has been useful for ad hoc requests and for solving unanticipated problems. The ability to quickly respond to information requests from governmental and regulatory agencies has also been enhanced through the use of the SAS® system.

SUMMARY

The deployment of this microcomputer-based data management system has proven successful in the data entry, data management, descriptive analysis and graphics of clinical trials in foreign countries. Much of the success can be attributed to the power and versatility of the SAS® system. The use of the SAS® system on today's highly powerful portable microcomputers provides the computing power and resources needed to conduct medical research in remote locations.

REFERENCES

(1) Bertrand, W. "Use of Microcomputers in Health and Social Service Applications in Developing Nations", CRC Critical Reviews in Medical Informatics, Volume 1, Issue 3.

(2) Oland, D., Feuillade, M., Warren D., Cannon T. "An Automated Data Management System For Remote Medical Studies", IEEE Proceedings of the Eleventh Annual Symposium on Computer Applications in Medical Care, IEEE Computer Society, 1987.

(3) Remsburg, M., "A Condensed Picture - Twenty-One Graphs On Two Pages", Proceedings of the SUGI Thirteenth Annual Conference, SAS® Institute, Inc., 1988.

(4) Stoll, M., "Scanners Play Key Role in Epidemic Research," PC Week, December 5, 1988.

The views of the authors do not purport to reflect the positions of the Department of the Army or the Department of Defense.

Approved for public release, distribution unlimited.

Dwayne D. Oland
Chief, Department of Applications Development
Biometrics and Information Management Division
Fort Detrick, MD 21701-5011
(301)663-7514