USING THE SAS® SYSTEM TO FACILITATE CLINICAL TRIALS RESEARCH AND NDA APPROVAL

Martin J. Rosenberg, M.J. Rosenberg Associates

ABSTRACT

The pharmaceutical industry, in conjunction with the Food and Drug Administration, has been experimenting with ways of using computer technology to facilitate and accelerate the process of introducing new drugs to the market. Some of these experimental approaches include "electronic" NDAs (new drug applications), remote data entry, and computer assisted monitoring of clinical trials.

The SAS system has long been the standard software used in the pharmaceutical industry for performing statistical analyses on clinical trials data. However, the SAS system has capabilities that go far beyond those of a statistical package, which make it an excellent choice for these new applications. Since most pharmaceutical companies already have considerable in-house expertise, the use of the SAS system is a particularly attractive alternative. This paper discusses these newer, nontraditional uses of the SAS System in the clinical research and NDA approval processes. Pertinent enhancements to the just released Version 6.03 of SAS for IBM PC compatibles are also discussed.

The Drug Approval Process

The pharmaceutical and biotechnology industries are unique in that before they can bring a product to market, they must first obtain government approval. The process of obtaining government approval is a long and complicated one. First, the company must use the proposed drug in animals to show that there are no gross toxic effects. With this knowledge, the sponsoring company can obtain an IND (Investigational New Drug) from the Food and Drug Administration (FDA). An IND permits the sponsor to ship the drug across interstate lines and to use it experimentally in humans.

Once an IND is obtained, the three premarketing phases of clinical research begin. Phase I starts by using small single doses in healthy human volunteers in order to look for toxic effects in man. As evidence of safety is accumulated, dosages are increased to therapeutic levels. In Phase II, the drug is tried for the first time in human volunteers with the targeted disease. Sample sizes are small, and information is gathered concerning the appropriate dose and regimen to use to obtain a therapeutic effect. Once sufficient information is gathered and hypotheses about the drug's efficacy in various indications have been formulated, large scale Phase III testing in human volunteers begins.

The research process takes five or more years. Throughout the process additional information is gathered concerning the safety of the drug in both humans and animals; the stability of the drug (how long it can remain on the shelf without degrading); and the ability of the company to manufacture the drug in production quantities.

When all the required research is completed, the drug company submits its information to the FDA along with a request for permission to market the drug in specified diseases. This registration process is known as a New Drug Application or NDA. The typical NDA is between 200 and 400 volumes long or in the hundreds of thousands of pages. Although information from many scientific disciplines is presented, frequently the largest single portion of the NDA is the clinical section. The FDA then reviews the NDA and can take three actions: approve the drug for all or some of the indications; not approve the drug; or request more information be gathered. The review process frequently takes another two or more years.

Evolution of Clinical Information Systems

Due to the volume of information which must be collected, traditional methods of data processing have long been a part of pharmaceutical R & D. Initially data was keypunched onto cards and then statistically analyzed using statistical packages as well as programs written in-house (Figure 1).

Figure 1: Early CIS included only data entry and statistical analysis.
This evolved into the current concept of a Clinical Information System or CIS (Figure 2). Physicians who conduct clinical trials of new drugs transmit the raw data to the sponsor on case report forms (CRFs) which are then entered into a database management system (DBMS). The data from the DBMS is transferred to a data analysis package for statistical analysis and computer generation of tables and graphs which are then incorporated into reports which become part of the NDA. Although no standard DBMS has emerged in the pharmaceutical industry, SAS has long been the standard software used for statistical analysis, report generation, and production of summary tables and graphs.

DBMS

STATISTICAL
ANALYSIS

AUTOMATED
TABLE AND GRAPH
GENERATION

Figure 2: Current CIS

In an effort to better manage the volume of information and to reduce the length of time required to introduce a drug in the United States, the pharmaceutical industry, in conjunction with the Food and Drug Administration, has been experimenting with new ways of using computer technology to facilitate and accelerate the process of introducing new drugs to the market. In this paper we talk about three of these experimental methods: remote data entry and monitoring; clinical data review; and electronic NDAs, which are also known as computer assisted NDAs (CANDA).

Clinical Data Review Systems

In the traditional data processing environment, only the data entry personnel, statisticians, statistical programmers, and other computer professionals have access to the data. A clinical data review system expands the concept of a CIS by providing medical monitors, clinical research associates, and other in-house clinical researchers access to the information stored on the computer. This capability can then be used to monitor the progress of ongoing trials. Three advantages arise:

- Reduced paperwork and greater efficiency for the medical staff
- Potential for detecting and correcting errors earlier, leading to fewer discrepancies in the final dataset and a greater percentage of usable patients
- Access to information useful for planning future clinical trials

Remote Data Entry and Monitoring

Remote data entry and monitoring (REDEM) refers to placing a computer in the investigator's office in order to allow him or her to enter data directly and transmit the data electronically to the sponsor. Although known sometimes as simply remote data entry, the real benefits derive from the sponsor's ability to monitor the trial remotely between site visits. These benefits accrue in several ways:

- Data entry programs that perform interactive edit checks and inform the investigator immediately about discrepancies.
- Programs that prompt the investigator for information, e.g. that a patient has missed a visit; the current status of an adverse reaction reported at a previous visit
- Nightly transmittal of information to the sponsor.
- Tools on the monitor's computer which can help the monitor detect and correct discrepancies.
- Improved communications between the sponsor and investigator through electronic mail and the ability to view data simultaneously in both locations.

The potential benefits derived from REDEM are:

- Reduced error rate in the final dataset.
- Reduced site visits.
- Reduced elapsed time from the end of enrollment to the completion of the final analysis.

It must be stressed that in those studies where REDEM has been used successfully, the benefits were derived primarily from the increased ability to monitor the study rather than the fact that data entry was performed at the investigator's site instead of at the sponsor. If this increased ability to monitor is not utilized by the sponsor, few benefits are generally derived and in fact REDEM may produce results that are worse than with traditional data processing.
"Electronic" NDAs or CANDAs

Electronic NDAs (also known as computer assisted NDAs or CANDAs) have generated considerable interest among the pharmaceutical companies; their trade association, the Pharmaceutical Manufacturers Association (PMA); and the FDA. Despite the space age name, an electronic NDA is simply an attempt to bring computers into the NDA review process. Electronic NDAs have several desired benefits:

- Reduction in the drudgery associated with handling so much paper
- Increased accuracy of reviews
- Ability of the reviewer to ask questions not directly covered by the NDA. Makes the reviewer more of a scientist.
- Acceleration of the review process.

A variety of approaches have been tried including: "dumb" terminals at the FDA tied into a database residing on a mini-computer at a third-party vendors site; terminals hooked into the sponsor's computer with the ability to view data and to send electronic messages to the sponsor, but not the ability to manipulate a dataset and perform original analyses; and systems based on microcomputers and standard PC software.

Our own previous efforts have involved using a system written entirely in Version 5 of SAS/AF on a mainframe, with the FDA communicating over phone lines via a microcomputer configured to emulate a 3270 style terminal.

Future Expectations of CIS

As these advances move from experiment to production, our concept of the Clinical Information System will evolve and, when appropriate, we will expect our CIS to support the entire process from data capture in the investigator's office through FDA review (Figure 3).

The Role of SAS in Expanding CIS

A number of features suggest using the SAS system as a way of implementing these new CIS applications. SAS is already the standard for analyzing clinical data. With the advent of the SAS/AF product, menu driven systems for clinical data review can be developed which harness the power of SAS without the need to know computer programming. In addition, PROC FSPROMS (available in SAS/FSP) provides the ability to view data and to perform ad hoc queries. These abilities have been enhanced with the addition of screen control language in Version 6.03.

For REDEM, screen control language provides the ability to write the necessary data entry programs that will prompt the investigator and perform interactive edit checks. The micro to host link, available as part of the base SAS product on microcomputers, provides the necessary ability to easily transfer data from the investigators office to the sponsor's computer over telephone lines.

In CANDAs, once again the powerful report writing, data analysis, browsing, and ad hoc query tools can be harnessed for the FDA reviewer through SAS/AF. The ability to construct computer based tutorials with SAS/AF, allows the reviewer to conveniently learn the system by actually using it.

Version 6 and Screen Control Language

Version 6 of SAS enjoys substantially enhanced data management features through the addition of Screen Control Language (SCL) to the SAS/AF and SAS/FSP products. SCL is a programming language for controlling fields on a screen and for managing the screens that comprise an application. Users familiar with the DATA step will find SCL to be syntactically similar, but with a new assortment
of functions for manipulating fields on a screen which can, but need not, correspond to variables in a SAS dataset. SCL is available in PROCs FSEDIT and FSBROWSE within SAS/FSP; and in program screens created with SAS/AF.

PROC FSEDIT has always had the ability to perform data entry and ad hoc queries while displaying data in a format which could be customized to resemble CRFs. The addition of SCL allows a program to be attached to the customized screen. SCL programs can be used for such things as: branching to other SAS datasets; computing new variables from existing variables; performing double-key entry (verification); and performing cross-validations with other variables in the same or a different SAS dataset.

SCL programs have up to five separate blocks of code which are designated by the following reserved labels:

FSEINIT: Code which is executed when the FSEDIT application is first entered. Used for opening datasets and setting parameters.

FSETERM: Executed when the FSEDIT or FSBROWSE application is ended. Generally used for closing datasets.

INIT: Executed whenever one initially displays an observation.

MAIN: Executed when an observation is added or whenever the enter key or a function key is pressed if a screen field has been modified and no errors are detected.

TERM: Executed whenever one leaves an observation if a screen field on that observation has been modified.

In SAS/AF only the INIT, MAIN, and TERM labels are available and these blocks of code are executed whenever one enters, modifies, or leaves a program screen, respectively.

A Sample CANADA System

We'll illustrate the ideas presented in this paper, by means of a sample CANADA system constructed using the following Version 6.03 SAS software: SAS/AF, SAS/FSP, and the base product. Figure 1 displays a typical menu screen. The user selects an option and enters its number on the "Select Option" line. In this system, the FDA reviewer can browse prepared text, graphics, and tabular information about the NDA; view data in case report or tabular format; rearrange and transform data; compute descriptive statistics; generate high resolution color graphics; and access utilities such as electronic mail to the sponsor.

One of the most useful features of a CANADA is the ability to view data collected during the clinical trials. SAS provides two formats for viewing data. The first displays data in the form of a rectangular table where the columns are variables and the rows observations on those variables. This format is particularly useful for viewing data across patients or visits.

Figure 2 displays a fill in the blank program screen for displaying data as a rectangular table. The FDA reviewer fills in blanks to specify the action to be taken. In this instance, the user need only specify the dataset containing the information to be viewed.

A Sample CANADA System

We'll illustrate the ideas presented in this paper, by means of a sample CANADA system constructed using the following Version 6.03 SAS software: SAS/AF, SAS/FSP, and the base product. Figure 1 displays a typical menu screen. The user selects an option and enters its number on the "Select Option" line. In this system, the FDA reviewer can browse prepared text, graphics, and tabular information about the NDA; view data in case report or tabular format; rearrange and transform data; compute descriptive statistics; generate high resolution color graphics; and access utilities such as electronic mail to the sponsor.

One of the most useful features of a CANADA is the ability to view data collected during the clinical trials. SAS provides two formats for viewing data. The first displays data in the form of a rectangular table where the columns are variables and the rows observations on those variables. This format is particularly useful for viewing data across patients or visits.

Figure 2 displays a fill in the blank program screen for displaying data as a rectangular table. The FDA reviewer fills in blanks to specify the action to be taken. In this instance, the user need only specify the dataset containing the information to be viewed.
A pop-up window then appears (Figure 3) which displays the names of the variables in the dataset. The user can optionally select up to three variables to be used to identify each row by positioning the cursor on a variable name and pressing enter. These ID variables remain in place as the user scrolls left or right on the data display.

Another new feature of Version 6.03 is the availability of command menus within the full screen procedures (Figure 4). Without having to memorize commands, the menus allow the user to scroll left or right and up or down; add or drop variables from the display; and rearrange the order of the variables. This last ability is particularly useful. By re-arranging their order, the user can position variables adjacent to each other, so that they can be viewed simultaneously.

The second way of viewing data is with PROC FSBROWSE. This procedure permits the data display to resemble the original case report forms. Additionally, the user can query the database. In the example shown in Figure 5, the reviewer has used the FIND command to request to see female patients with a high diastolic blood pressure.

In addition to the commands supplied by the SAS Institute, screen control language permits the programmer to create new commands. In the example shown in Figure 6, the customized command "OTHER" displays a list of other datasets containing information about the displayed patient. The reviewer selects the desired dataset and the requested data for the patient appears. The reviewer can continue to browse and query the new dataset. When through, a single keystroke returns the reviewer to the original dataset.
A major problem in clinical trials is how to handle comments which investigators write on case report forms. The comments can be stored conveniently and compactly in a separate SAS dataset. Then with screen control language, the reviewer can choose to view the comment when desired. In Figure 7, the box in the lower right corner indicates a comment is available for this patient. The reviewer responds "y" for yes and is taken to the comment shown in Figure 8. Once again, the reviewer can continue to browse or query the comment dataset and with a single keystroke can return to the demography dataset.

Another desirable feature of a CANDA system is the ability to compute descriptive statistics such as means, frequencies, and percentages, and the ability to generate graphics. With SAS/AF the CANDA programmer can create customized front ends to SAS procedures. For example, Figure 9 displays a program screen which creates a hierarchical table of descriptive statistics using PRCC TABULATE. The reviewer fills in the names of the dataset, the analysis and classification variables, and the desired statistics in order to generate the table shown in Figure 10.
Other features of the SAS system allow the CANDA programmer to design a system in which on-line documentation is readily available. Context sensitive HELP screens, such as the one shown in Figure 11, can be called up with a single keystroke. HELP screens can be made available for each menu and program screen as well as each blank on a program screen. Another new feature of Version 6.03 is the LEGEND window. LEGEND windows provide the user with detailed instructions and feedback when they are filling in a program screen. Other windows are available to display the definitions of the function keys, and the names of datasets and variables.

Finally, an extremely important feature of SAS/AF is the CBT screen. With this feature, the CANDA programmer can create computer based training tutorials which teach the FDA reviewer to use the system by actually using the system. Given the physical distances that generally exist between the FDA and the sponsoring pharmaceutical firms, a well thought out CBT tutorial could determine the success of a computer based NDA system.

The recent introduction of Version 6.02 is of particular interest for those wishing to develop clinical data review, remote data entry, or computer assisted NDA systems. With the addition of screen control language, it is now possible to do many things that previously required a sophisticated (and usually expensive) database management system. While we don't suggest that the SAS system is the one right tool for everything, perhaps it is time to reevaluate your company's use of SAS to ensure that it is reaping the full benefits of this powerful tool.

For further information, please contact:

Martin J. Rosenberg, Ph.D.
M.J. Rosenberg Associates
99 East Middlefield Road, Suite 31
Mountain View, CA 94043
tel. (415) 961-8432
(415) 961-9260

*SAS, SAS/AF, and SAS/FSP are the registered trademarks of SAS Institute Inc., Cary, NC, USA.