The Selling of SAS® Software: Toward a Unified Analysis Environment in Basic Research

Mark H. Bradshaw, Syntex Research

BACKGROUND:

The mission of Basic Research at Syntex is to identify new compounds with marketable biological activity for potential therapeutic use in humans. A compound's biological activity is determined through a series in vitro, ex vivo and/or in vivo experiments conducted by a large number of biologists and experimental pharmacologists. Recently a statistical support group was formed specifically to work with scientists in Basic Research at Syntex on experimental design and statistical analysis.

COMPUTING ENVIRONMENT:

When the statistical support group began to work with scientists in Basic Research a wide variety of approaches to statistical computing were discovered. Hardware in use included IBM-compatible PCs, non-ICM-compatible PCs, hand calculators, mini-computers and limited use of the corporate IBM mainframes. IBM-compatible PCs were predominant.

Software for data summarization and analysis on PCs included several different limited statistical packages, spreadsheets and database managers with statistical functions, and programs written in-house in BASIC, FORTRAN, PL-1, PASCAL, APL and C for on-line data collection, summarization and/or statistical analysis. In-house programs were generally written for specific studies.

The SAS® System was available on the corporate mainframes but had seldom been used for Basic Research. However, a VAX 8700 was to be installed, and a biological data system on the VAX was being planned which would include SAS software in the future. Preparation for this system, along with the current availability of a large number of IBM-compatible PCs in various laboratories were two important factors in the decision to begin to standardize on the SAS System for data analysis in Basic Research.

GOALS OF THE STATISTICAL SUPPORT GROUP:

The mission of the statistical support group is three-fold:

- To work as scientific colleagues with biologists and pharmacologists, providing statistical consulting in the design of experiments and training in statistical concepts.
- To perform statistical analyses on the results of various experiments and pilot studies, assuring the validity of all analyses used for internal decision-making, submitted for publication, and/or presented to the FDA.
- To provide valid statistical application software to scientists which will allow them to enter and analyze their own data from drug screening studies which become routine, while encouraging better standardization between laboratories in statistical techniques. Analysis applications provided must allow for migration to the VAX environment in the future.

ADVANTAGES OF THE SAS SYSTEM IN ACHIEVING THE GOALS:

SAS software for the PC was chosen to accomplish the goals of the statistical support group in three main areas: 1) SAS software is being used to provide a quick and convenient means of entering raw data into SAS datasets using SAS/FSP®; 2) statisticians in the group are performing most of the requested statistical analyses of experiments and pilot studies using SAS software; and 3) end-user analysis applications for routine screening studies are being built with SAS software for the PC by modifying data entry screens and statistical analysis programs already written in 1) and 2).

1) PC Data Entry System for Statisticians' Analyses: SAS/FSP is being used to build simple data entry screens with simple validations. An analyst can build a screen quickly and provide a copy to Data Entry along with the laboratory data sheets from an experiment. This procedure yields a SAS dataset directly, it eliminates the need to extract or convert data from another database or data file, and it gives the analyst control over the structure of the dataset so that statistical analyses can proceed with a minimum of data manipulation. Typical turnaround time for a small study is less than one day.

2) Analyses Performed by Statisticians on PCs: We have found that most analyses required for Basic Research can be performed using SAS software for the PC. A good range of descriptive and inferential statistics, parametric and non-parametric procedures, as well as listings, plots and graphs are available directly from SAS PROCs, and additional statistical procedures can be written in data steps or using SAS/IML.™ In addition to the variety of statistics available as PROCs, the data manipulation capabilities are very powerful.

One important improvement this approach offers over the use of many different software packages and programs is that SAS code can be easily documented and validated, and we can assume that SAS PROCs have been carefully validated by SAS Institute. This is especially important in our environment because many analyses become a part of submissions to the FDA. It has also been relatively easy to utilize the power of the mainframe when necessary simple by using PROC UPLOAD/DOWNLOAD and RSUBMIT. Our analysis code is portable to both the IBM mainframe and the VAX environments.

3) Study-Specific PC Data Entry/Analysis Software for End-Users: Data entry screens and analysis code are often prepared to analyze sample or pilot data from a study which will be run on a routine basis in the future. Once data formats and analysis strategies are finalized for one of these studies, we must provide user-friendly programs so that the scientist can enter, summarize and statistically analyze his own data in the future.

The SAS System for PCs has become our standard approach to solving this problem. Because data entry screens and analysis code have already been generated for sample or pilot data, end-user applications can evolve from the pilot analyses. A modular approach is used, where an application typically has at least four modules--DATA ENTRY, DATA EDIT, DATA DISPLAYS, and ANALYSES. Currently these modules are separate programs which the user must run individually but in the future we

intend to link them together under a SAS/AF^{tm} menu system.
The first step is to make the existing SAS/FSP data entry screens somewhat more user-friendly in appearance, sometimes adding more validations. A general approach for creating and storing permanent datasets is then worked out with the user. The other modules evolve from the analysis code that has already been written. Data listings and summary reports are designed around the user's needs. After testing, the use of the system is documented and the intended user is trained to run the programs and interpret the output.
This approach meets several important criteria necessary for success in our situation. It allows us to provide valid, powerful and versatile statistical analysis programs in the predominant hardware environment--PCs. It encourages standardization of data formats and analyses within and across laboratories. End-user analysis programs have a modular design so that SAS/AF menu systems can easily be added in the future. End-user programs evolve from statisticians' pilot data analysis programs, using the same analysis code, so the process is very efficient.
When an end-user has already developed a data management system using other PC software, is very familiar with that environment, and prefers to continue to enter data using his existing system, we have been successful in providing programs and/or training necessary to transfer the data to our SAS systems for analysis. This flexibility has been an important factor in convincing some investigators to standardize on the SAS System for their inferential statistical analyses.
A final, key advantage to our approach is the ability to port end-user applications to the VAX systm in the future with a minimum of re-coding. For some routine studies, we believe our SAS/FSP and AF systems can become a standard method of entering raw data into the VAX. Analysis results can then be transferred routinely to the biological database.
The advantages we see of using SAS as a standard analysis environment for Basic Research at Syntex can be summarized as follows:

* The SAS System is POWERFUL.
* Our SAS code is PORTABLE.
* The results are PUBLISHABLE.

Mark H. Bradshaw
Syntex Research, A4-100
P.O. Box 10850
Palo Alto, CA. 94304
(415)852-1515