AN INTEGRATED CLINICAL DATA MANAGEMENT SYSTEM AT ALLERGAN, INC.

Jeff S. Kouba, Allergan, Inc

INTRODUCTION

Allergan, Inc., a subsidiary of SmithKline Beckman Corporation, located in Irvine, California is involved in the research, development, manufacture and sale of ethical drugs and devices. Allergan specializes in ophthalmologicals, contact lenses, contact lens solutions and dermatologicals.

As manager of a programming support group for Allergan Research and Development efforts (deemed R&D Scientific Information Services), my responsibilities include support of the R&D Data Management department which ensures the integrity of clinical data, and of the Biostatistics department responsible for the analysis and reporting of that data. Over the course of several years in support of these activities, we have developed a series of integrated SAS* programs on our IBM 3084 under VM/CMS to manage the ever-increasing volume of clinical data and speed up the process which brings that clinical data from the Case Report Form (CRF) to a New Drug Application (NDA) or device Pre-Market Approval (PMA).

These programs are divided into two major areas:

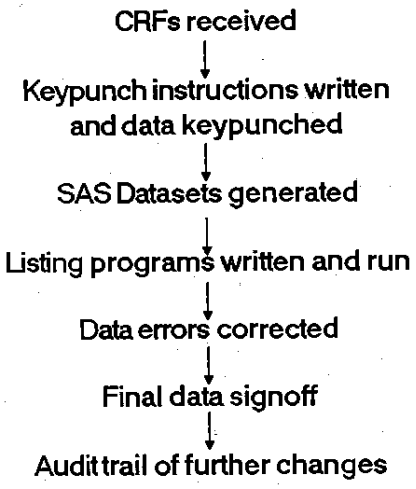
1. Clinical data entry and editing
2. Standardized Table Generation

CLINICAL DATA ENTRY AND EDITING

My discussion of clinical data entry will begin with a short description of the data flow from the raw CRF to the SAS dataset containing "clean" (accurate) data used by the Biostatistician for analysis. I will then discuss the various programs we have set up to speed this process. Finally, I will discuss some problems we have encountered in using and managing the system. As you will see, the concept of clean, accurate data and a standardized, organized method of storing that data are central to the process and the driving forces behind the programs we developed. In addition, we have been able to maximize our programmers' efficiency by freeing them from dull, repetitive programming and allowing them to devote their talents to custom, original projects.

The Process

A diagram of the Clinical data entry process and a description of each step is provided below:



1. CRF's are received by R&D Data Management after a preliminary visual check for errors and omissions by the Clinical Research Associate (CRA). This check removes most obvious data omissions and deficiencies on the CRFs.

2. CRF keypunch instructions are written and data sent to an outside service for keypunching.

3. SAS datasets are generated from the keypunched data. Data is maintained in CMS (operating system) files for easy editing and to protect the data from software changes.

4. Listing programs are written and run. Two kinds of listings are generated.
   1) CRF duplicates used by the Clinical Coordinators
   2) Line-by-line listings used by Clinical Data Management

5. Typographical errors entered into the system from keypunch are corrected in the CMS file and SAS datasets are regenerated.

6. Once the data have been deemed "clean", it is signed off by the Clinical Coordinator. It then becomes the 'source data' and all changes to it must be audited. As stated before, this 'source data' is maintained in CMS files.

7. Additional errors found during analysis are corrected through an audit trail. These errors include logical data errors (those occurring across subject-visits), and subject disqualifications and are usually discovered by the Biostatistician.

The Programs

Here are the SAS programs and datasets which comprise the Clinical Data Entry System and a discussion of how they aid the process.

1. The Data Dictionary

The Data Dictionary is our central control point; the Dictionary ties all the parts of the system together and provides an organized standardization of clinical data variables.

The Data Dictionary is a SAS dataset containing one observation for each different variable used in our clinical trials (currently about 4000 observations). The Dictionary contains such items as:

a. Variable name
b. Description
c. Data format
d. Value ranges (if appropriate)
e. Notes and comments

As soon as we receive approved Case Report Forms from the Clinical Research Department for a study (they may come with the completed CRF's at the end of the study), the study variables are checked against existing variables in the Dictionary. Any new variables are added to the Dictionary as required. All of the other programs rely strongly on the information in the centralized Data Dictionary.

2. The Keypunch generation program

The keypunch instructions are automatically written by this program. The program takes a list of variables defined from the CRF as input and, working with the Data Dictionary for formatting length and special keypunch directions, outputs the keypunch instructions for the data. The keypunch instructions are sent to the keypunch service with the data.

3. SAS <-> CMS program code generator

As I mentioned before, all of our clinical data are permanently stored in CMS (operating system) files to promote easy editing and to protect the data from software changes. Because the data are analyzed in SAS and all editing is done to the CMS data file, we must be able to generate SAS datasets quickly and easily from CMS files and CMS files from SAS datasets. This code generator writes the SAS program code (from information in the Data Dictionary) which translates the data back and forth between the two file formats. This code generation program saves approximately 40 programmer hours per month over custom programming and is significantly more error-free.

4. Listings program code generator

This system, currently under development, will automatically generate program code which outputs data in both CRF duplicate format and line-by-line listing format. It is somewhat complex since it must allow the user to "paint" the output format to match the CRF for the CRF duplicate format. Variable output formats will be checked against the Data Dictionary.

5. Automated Computerized Data Editing System (ACDES)

This system does logical and range checks of the data based upon criteria entered into the Dictionary. It will find errors not only on a particular subject-visit form but across all forms and all visits. In addition a "bubble-sheet" report is generated which inventories the CRFs entered into the system. This system has been very useful in finding errors and omissions that used to appear only during the analysis.

6. The Audit Trail System

After the Clinical Coordinator signs off on the data, they are considered 'source data' and any errors found in the data must be followed by an audit trail. This system maintains that audit trail and generates printed reports of changes made to the data.

Problems

No system designed is without its problems and we have had our share with this system. I will present here the two major problems we have encountered with the Clinical Data Entry and Editing System.

1. Case Report Form Standardization

When working with a centralized Data Dictionary, as we have here, great efficiency can be realized through standardization of the Case Report Forms. Sharing variables between CRFs as much as possible decreases the size of the Data Dictionary and makes the process of incorporating new CRFs into the system much easier. Unfortunately, in our organization, this is not a reality. The R&D Data Management group is a central service function for the Clinical Research groups without any formal organizational ties. There is no central authority ensuring that both groups would comply with standardization. Changes in the Clinical Research process can produce significant changes in the way data are collected on the CRF. Attempts in the past to create more standardized CRFs have been laudable but lacked the necessary support to succeed.

2. Data Dictionary Maintenance

Currently there are approximately 4000 different variables in the Data Dictionary. Since it is the central control for the Clinical Data Management System, it is critical that it be maintained properly. Although the SAS Full Screen Product helps us maintain the Dictionary, it still requires a maintenance operator who is familiar enough with all of the data to recognize potential problems and be able to solve them quickly. This expertise is acquired only after a significant investment in time and training which can be lost if the person leaves the department or company. Fortunately, this has not happened yet, but we recognize the possibility as a potential problem.

STANDARDIZED TABLE GENERATION

The Standardized Table Generation System is another system which is built upon the concept of a centralized Data Dictionary. I will now spend some time defining the Standardized Table Generation System (STGS), describing the process used for its development and associated problems encountered in its installation and implementation.

What is STGS?

The statistical and clinical final reports which become part of the NDA or PMA include numerous data tables which list both raw and derived study data in a variety of formats and breakdowns. Much time and effort is expended in generating these tables for each study report.

If studies are sufficiently similar in design, significant time savings and effort can be realized by reusing the program code written to generate these tables.

This is a fairly simple concept involving a limited library of programs. Allowing programmers to share common code for formatting a specific table type. With a standardized, organized method of storing the data, an additional step can be taken to further simplify the table generation process: write generalized programs which can work for ANY study of similar design and place those programs into a menu-driven system allowing non-programmers to generate the required tables quickly and efficiently.

This was the goal when we (in coordination with the Biostatistics group) started the development of the Standardized Table Generation System at Allergan. We had an initial group of studies within the Contact Lens Clinical Research area which were very similar in design. Our centralized Data Dictionary provided the necessary data organization. The code for the tables was, for the most part, already written. It seemed at the outset to be a fairly straight-forward development process.

The Development Process

The development process was planned to include the following:

1. Determine which tables would be included in the system.

Discussions with the Clinical Research Coordinators and review of previous reports would result in a list of 'standard' tables to be used in Contact Lens solution study reports.

2. Organize the existing programs used to generate those tables and write any additional required code.

This would require the gathering of all the programs used in the past to generate tables. Any new tables that resulted from the discussions in Step 1 would have to be programmed from scratch.

3. Generalize the programs to work for 'any' study and place them into a menu-driven system.

This would entail integrating the programs with the Data Dictionary and assembling the menu-driven system.

Problems with STGS

We ran into two rather serious problems in developing STGS.

Changes to the Tables

After working with the table generation programs for quite some time, we found that the Case Report Forms (CRFs) which are the source of the data in the tables, change about every six months. These changes result from changes in the Clinical Research process and cannot be avoided. In addition, the design of the tables must be altered periodically to meet FDA requirements. Unfortunately, these changes are inevitable drawbacks to the standardization required for STGS to function efficiently. They necessitate modification of the table programs. The resultant table programs would no longer work for 'all' studies. STGS requires additional flexibility to overcome this problem.

Resource Restrictions

The main purpose of the Biostatistics group is to provide statistical support in the design, analysis and interpretation of data for clinical studies. STGS is an 'extra' project which is worked on during 'spare' hours. Unfortunately, time is at a premium Hence, the project has had insufficient manpower inhibiting a quick completion.

In spite of these problems, STGS continues to be an important system to us. We are hoping to bring on additional programming support this year to help bring STGS to completion.

CONCLUSIONS

I have discussed two major systems that we have developed under SAS which help us to manage the ever-increasing volumes of clinical data processing that we are faced with each day. Central to these systems is an organized, standardized Data Dictionary without which these systems would be impossible to develop. Problems relating to the standardization of diverse clinical studies have plagued us but not deterred us from our overall goal of clean clinical data and reports created in an efficient, computerized environment.

Jeff S. Kouba
Mgr,R&D Scientific Information Services
Allergan, Inc.
2525 Dupont Drive
Irvine, California 92715
(714) 955-6819