Using SAS/AF® for Managing Clinical Data

Carl R. Haske, Ph.D., STATPROBE, Inc., Ann Arbor, MI

ABSTRACT

Using SAS/AF as a software development platform permits rapid applications development. SAS supplies several classes of objects that the developer can use to quickly prototype and implement systems. However, in developing applications for use in the pharmaceutical industry, it is useful to augment these classes with additional classes that provide robust data management and reporting capabilities. Classes that provide generic tools to subset data, browse data, move data, and report data are useful in the clinical research environment.

STATPROBE, Inc., has developed an integrated data management system, using SAS/AF as the platform. The system incorporates security and tools that assist the valid production and analysis of the clinical data. Various modules in the system support numerous clinical data management activities. Data at the quality assurance stage is accessible remotely via the STATPROBE Data Access System. A final database may accumulate in phases or be generated at once when quality assurance is complete. The STATPROBE Data Management System (DMS) and STATPROBE Data Access System provide a systematic method of implementing the necessary tasks for the management of clinical data at STATPROBE, Inc. This paper describes these systems and the associated class libraries and templates used for the development of these systems.

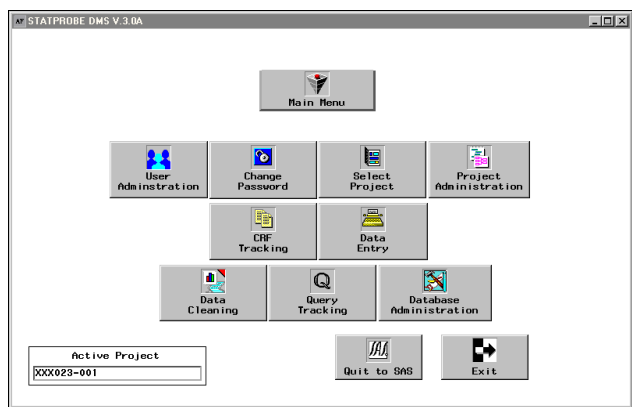## MANAGING CLINICAL DATA

Figure 1. Main system menu

Figure 1 shows the primary menu of the DMS. Pads on the primary menu are:

- User Registration
- Change Password
- Select Project
- Project Administration
- CRF Tracking
- Data Entry
- Data Cleaning
- Query Tracking
- Database Administration
- Quit to SAS
- Exit

The system architecture is logically divided into three sections: a core application section, a section for administration of multiple database management projects, and a section that addresses specific clinical data management tasks for a given project. The core application section handles user login, user registration, and changing passwords, functions that are common to many software applications. The project administration and project selection segment of the system is designed to allow for management of several databases for multiple clinical trials. Multiple project management is also a task common to many software applications. The five main activities specific to the management of clinical data are CRF tracking, data entry, data cleaning, query tracking, and database administration.
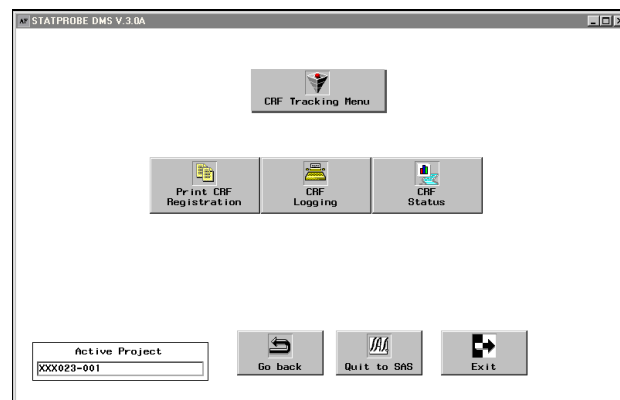
CRF Tracking

Figure 2. CRF tracking menu

The CRF tracking process involves printing CRF page registration, logging CRFs, and reporting the status of CRFs relative to the database (see figure 2). The actual registration of CRFs occurs via the Database Administration menu that is discussed later. The CRF tracking tasks access the CRF registration database. CRF pages are registered in the CRF dictionary (see table 1). CRF pages must be registered prior to the construction of the clinical database.

Table 1. CRF Registration Data

ATTRIBUTE	DESCRIPTION
Crf_page	Unique CRF page number or id
Crf_desc	CRF description
Crf_keys	Key identifying fields
Num_flds	Number of entry fields on CRF
Crf_req	Logical - Is CRF required?

Table 2. Relation of CRF Entry Field to Database

ATTRIBUTE	DESCRIPTION
Crf_page	Unique CRF page number or id
Crf_fld	Entry field name
Tablname	SAS data set name
Varname	SAS variable name

The CRF registration data are linked to the clinical database via a relation table (see table 2). This relation table acts as an electronic annotation of the CRFs. The registration table for the clinical database is described later in the "Database Administration" portion of this section. The CRF dictionary and relation to the clinical database allow benchmark information and tracking reports to be extracted from the system.

Data Entry

An essential component of the DMS is the ability to enter and modify clinical data. Clinical data at STATPROBE are entered into several

physical slots. Data entry personnel are assigned a logical entry level of PRIMARY or DOUBLE to support duplicate entry of data. Users are allocated a slot and are owners of that slot until their data are routed to a cleaning platform. After data have been entered twice, they pass through a cleaning phase described later.
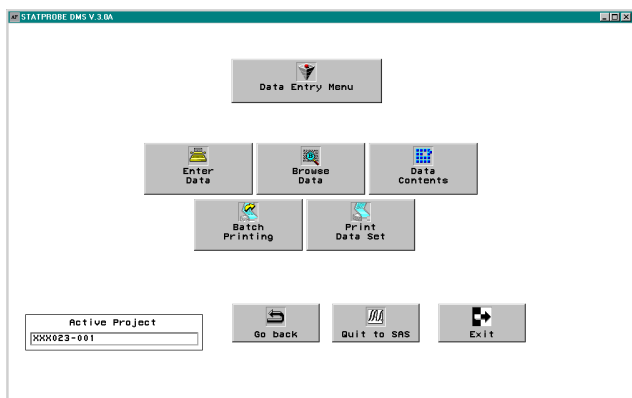

Figure 3. Data entry menu

To support the data entry process, users have the option of entering data, browsing data, producing a PROC CONTENTS of the data, or producing a printed copy of existing data either in batch or by specifying a single data set (see figure 3).
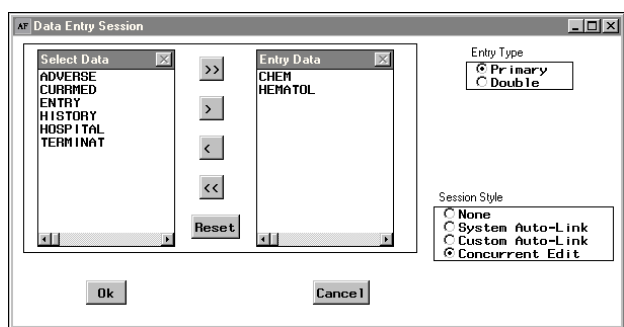

Figure 4. Specifying a data entry session

The system allows the user to select from four types of data entry session styles (see figure 4). To edit data, the user selects the data sets and entry session type, and then performs all the standard commands to add, delete, and modify data using an FSEDIT window. By using a mover list object, the user can list the data sets in order and have the system prompt prior to entry in each data set or, when a data set is closed, automatically open the next data set in the list. The user can also select system auto-link. This option links data entry via an order specified by the database administrator. Finally, the user can select concurrent edit. This allows the user to have multiple data sets open at the same time.

To browse or print a single data set, the user selects the data set and specifies a filter condition using an object instantiated from the filter class (see figure 5). The data that satisfy the list of criteria are printed to the output window or displayed in a viewer.
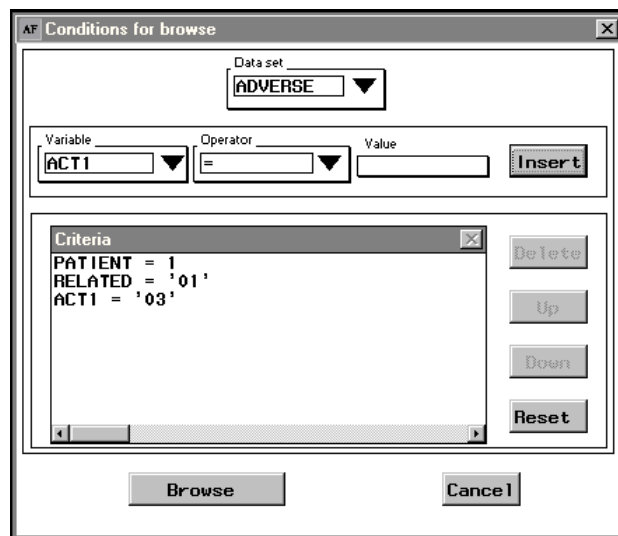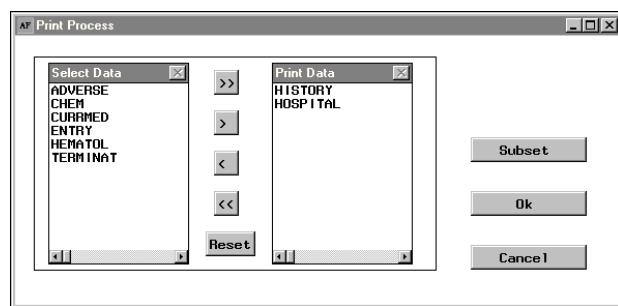

Figure 5. Filtering object

Figure 6. Selecting data for batch printing

For batch printing, users select the data they wish to print using mover lists (see figure 6) and may specify a set of criteria to subset all the data sets.

Data Cleaning

After data are double entered, they are routed from the entry slots to the PRIMARY and DOUBLE slots for cleaning. See figure 7 for the data cleaning menu.
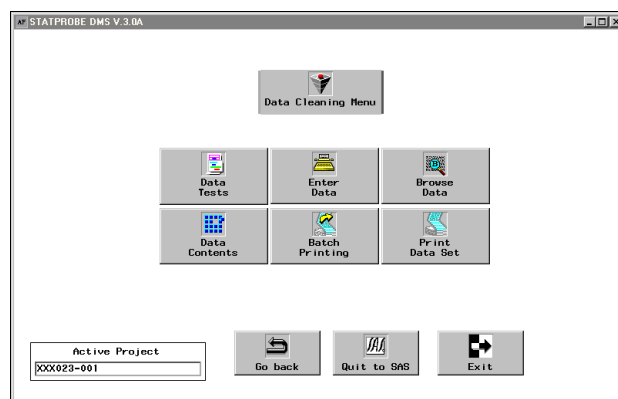

Figure 7. Data cleaning menu

Data are cleaned by following systematic steps to clean internal inconsistencies and to resolve inconsistencies in data between the

two slots. Inconsistencies are resolved by reference to the original CRFs. Data cleaning options include

- Data cleaning tests
- Enter data
- Browse data
- Proc contents of data
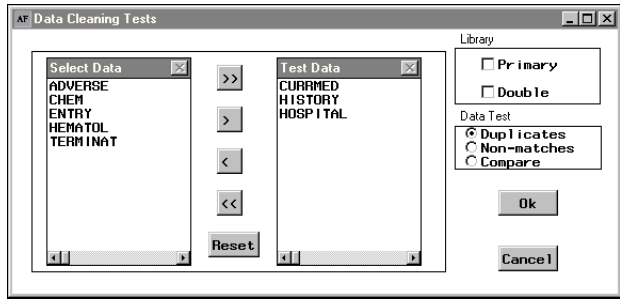- Batch printing of data
- Print a single data set

Figure 8. Data cleaning tests

Testing the data consists of the following three tasks executed in the order listed (see figure 8):

- Resolve duplicate PRIMARY and DOUBLE records
- Resolve nonmatching PRIMARY and DOUBLE records
- Resolve PRIMARY and DOUBLE matching record inconsistencies at the field level

Duplicate records are internal inconsistencies. These are located by key variables for each data set as defined in the database dictionary section of the DMS when a database is registered.

The nonmatching observation process checks for one-to-one cor-respondence between observations in the PRIMARY and DOUBLE databases. Each record in a given database must have a matching record for the key variables in the other database. Nonmatching records are printed in a report for resolution.

In the final cleaning step, data sets are compared for an *exact* match. A PROC COMPARE report is generated that lists linked records in PRIMARY and DOUBLE slots that disagree in non-key fields. When all fields in a record agree in PRIMARY and DOUBLE, the record is considered clean and stamped with a clean date. Data are now ready to be locked and moved to the QA slot for quality assurance and query resolution. At this point, data are available to remote clients via the STATPROBE Data Access application.

Database Administration

Database administration incorporates all the high-level operations involved in managing a study database. The database administrator works closely with the other data management professionals working to build the clinical database. The database administrator is responsible for making time-critical decisions and must be cognizant of the status of data in various slots. The tasks captured on the Database Administration menu are as follows:

- System Setup
- Register database
- Register CRFs
- Auto-link specification
- Batch routing of data
- Route a single data set
- Administration utilities

The Administration Utilities menu branches to another menu offering the following tasks:
- Entering data
- Browsing data

- Contents of database
- Batch printing of data
- Printing a single data set
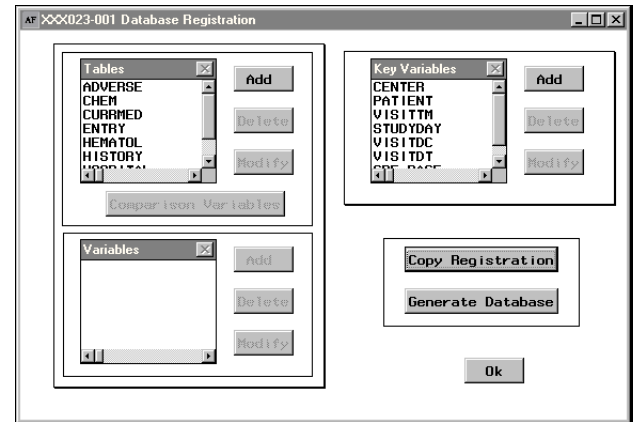- Allocation of entry slots
- Format designer

Figure 9. Database registration

The *register data sets* task is used to design a database dictionary and maintain the database structure (see figure 9). The database dictionary consists of two tables (see tables 3 and 4). When the database administrator designs or updates the database, the database dictionary is modified and the file structures are updated.

Table 3. Database Table Dictionary

ATTRIBUTE	DESCRIPTION
Tablname	SAS data set name
Tabldesc	SAS data set label
Key	Key table attributes
Numattrb	Number of attributes

Table 4. Database Column Dictionary

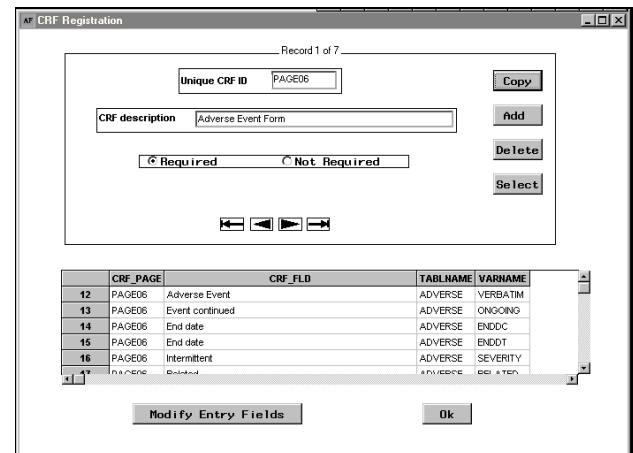| ATTRIBUTE	DESCRIPTION
Tablname	SAS data set name
Varname	Variable name
Varlabel	Variable label
Varlen	Variable length
Vartype	Variable type
Varfmt	Variable format
Varinfmt	Variable informat

Figure 10. CRF registration

The CRF registration provides an easy-to-use interface (see figure 10) for the database administrator to link fields on each CRF page

with variables in the data sets. This relation is tracked in the table described in table 2, located in the "CRF Tracking" section above. The CRF registration allows reporting facilities to access the database and scrutinize the status of CRF entry.
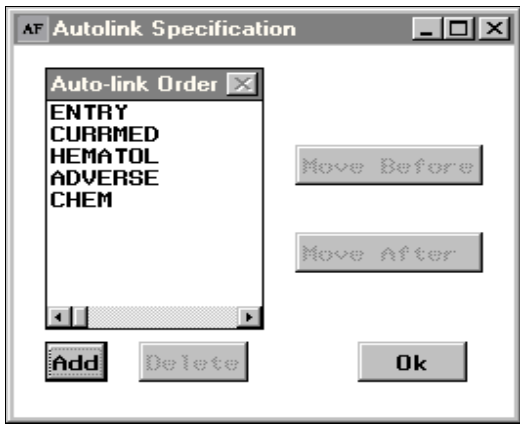

Figure 11. Auto-link specification

The auto-link specification allows the database administrator to order the data sets in a manner corresponding to the natural flow of the CRFs (see figure 11). This is saves time for data entry persons by not requiring them to select data sets and automatically opening data sets for entry in the order specified by the auto-link sequence.
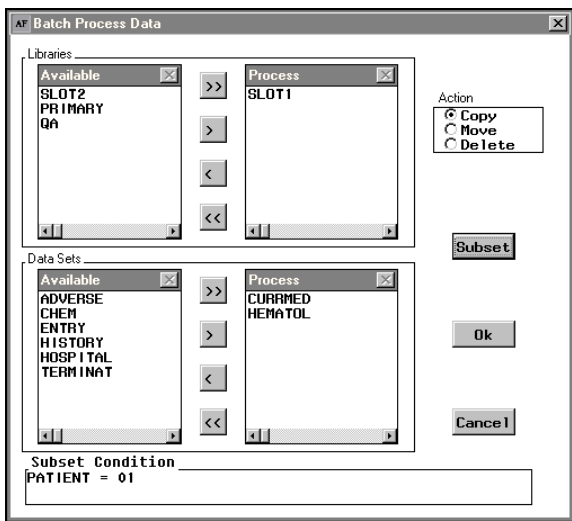

Figure 12. Batch routing

Each time data are routed between physical slots, the database administrator has the option to route an entire data set or subset the data, using a filter tool similar to the filter object in figure 4 to define conditions. There is an option to route a single data set and subset on variables that are specific to that data set or perform batch routing and subset on variables that are in all data sets. Batch routing is handled by the frame displayed in figure 12.

Even when data are frozen, it is likely that they will need to be modified. The database administrator is the only individual with access to "thaw" data and make modifications. An audit trail of all modifications is maintained.

ADMINISTRATION

Project administration is accessible only by a user who has system administration security clearance. The user can manage projects via a composite object (see figure 13). The table at the top of this frame is used for managing the project data. The information in this table is stored in a SAS data set. The four fields for client, product, project, and prefix are concatenated to form the path specification for the standard project file structure. Details of the standard physical layout are omitted.

A table that shows project assignments is displayed at the bottom of the frame. The table displays the data set that defines the relation between the project data set and the user data set. The user data set is described later.
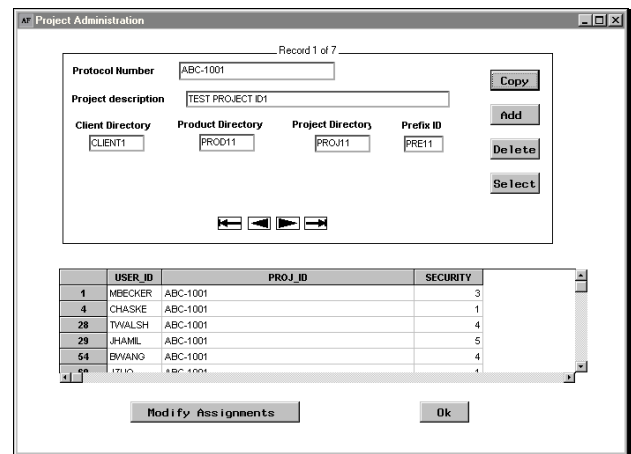

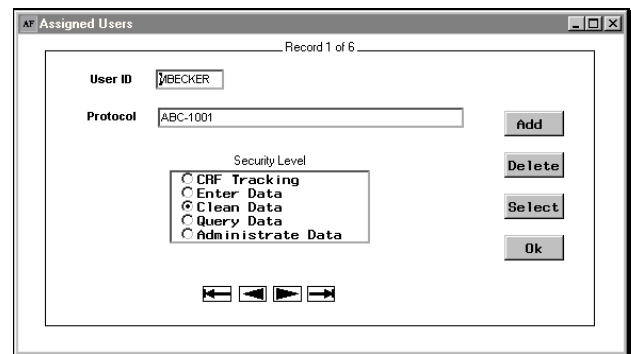**Figure 13.** Project management

Figure 14. Assignment modification

To modify assignments on a project, the administrator selects the Modify Assignments button, and a form allows additions, deletions, or modifications of personnel who are assigned to the project (see figure 14). When the Add button is selected, a list of available users is displayed (see figure 15). This list contains only users that have not been assigned to the project. Project protocol numbers are displayed in a list box for selection when the user chooses a project from the main menu (see figure 16). The only projects available to a user are those that have a security level assigned for that user. The system accesses the project path to assign standard library references to the physical locations of data files.
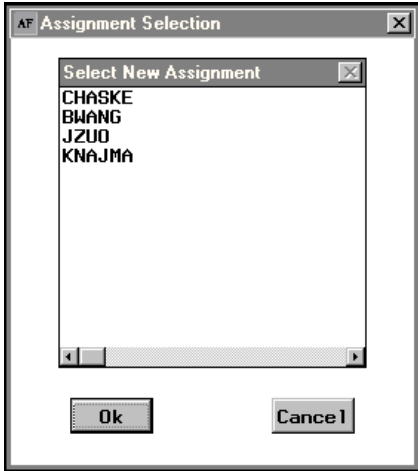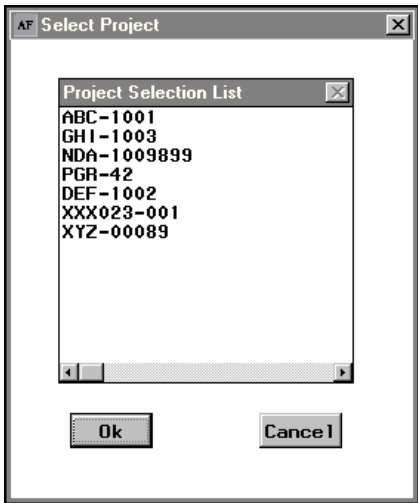
Figure 15. Assignment selection

Figure 16. Project selection

After selection of a project, the user has a choice on the main menu of the various data management processes: CRF Tracking, Data Entry, Data Cleaning, Query Tracking, and Database Administration. If a user does not have security clearance for a given process, the process button is turned invisible. The SAS data set that stores project information is displayed in table 5.

Table 5. Project Profile

ATTRIBUTE	DESCRIPTION
Proj_id	Unique project system identification
Protocol	Protocol number
Desc	Project description
Path	Physical path to project files

Security for the high-level processes of the data management system includes CRF Tracking (security level 1), Data Entry (security level 2), Data Cleaning (security level 3), Query Tracking (security level 4) and System Setup and Database Administration (security level 5). The security level system is hierarchical: a user with a security level of 3 will be allowed to access processes of security levels 1, 2, and 3. Security task levels are assigned to personnel for each project (see figure 14). Table 6 displays the data that are stored by the form in figure 14.

Table 6. Project Assignment Table

ATTRIBUTE	DESCRIPTION
User_id	Unique user system identification
Proj_id	Unique project system identification
Security	Security level

All users are required to specify a login id and password. The purpose of logging into the DMS is to provide security for accessing the data management system. A user can login to the system only if the system administrator has added the user's user profile to the user data set. Table 7 shows the structure of the user data set.

Table 7. User Table

ATTRIBUTE	DESCRIPTION
User_id	Unique user system identification
Password	User logon password
Username	Full user name
Syslevel	System level access
Prefs	SCL list of user preferences

The system-level security determines if a user has system administration rights. The registering of users employs a composite object similar to that used for registering projects displayed in figure 13. For adding or editing a user, the frame in figure 17 is used.
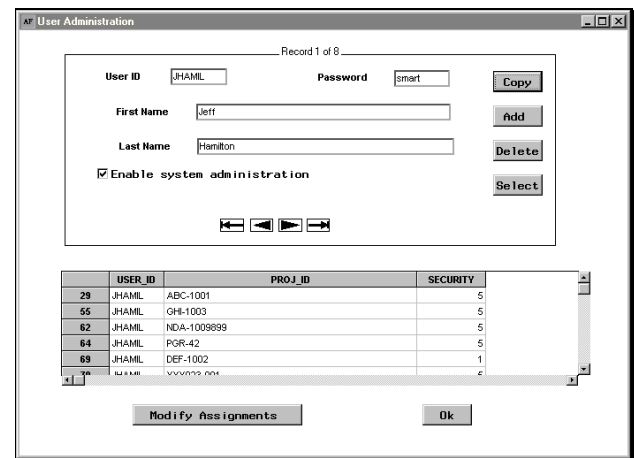

Figure 17. User registration

STATPROBE DATA ACCESS

The STATPROBE Data Access system is a client system that is integrated into the STATPROBE DMS. This system consists of hardware and software. The system was developed on OS/2®. Since initial development of the system in 1994, it has been ported to Windows 95®. Currently the system is at version 1.6.

The Data Access system is placed at a remote site and allows the client access to a clinical database as it is under development. The client has access to data in the QA stage of development.
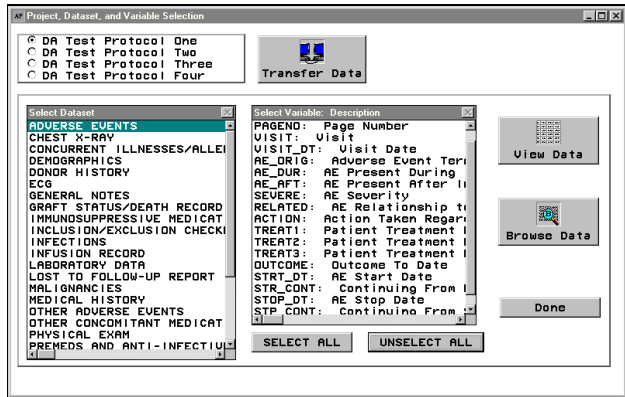
Figure 18. Data access main frame

Figure 18 displays the main selection frame of the system. The system supports multiple protocols displayed in a radio box at the upper left of the frame. The list box for the data sets at the left refreshes with data for a protocol whenever the protocol is selected in the radio box. As the user selects a specific data set, the variable list to the left refreshes with the variables in that data set.

The transfer data button at the top initializes a modem that connects with the data management site and transfers the database of the selected protocol. This allows users to periodically update their system with recent data. Data may be viewed either in tabular format or in record format. The view data button displays the data in tabular format (figure 19). The browse data launches a FSEDIT session in browse mode.
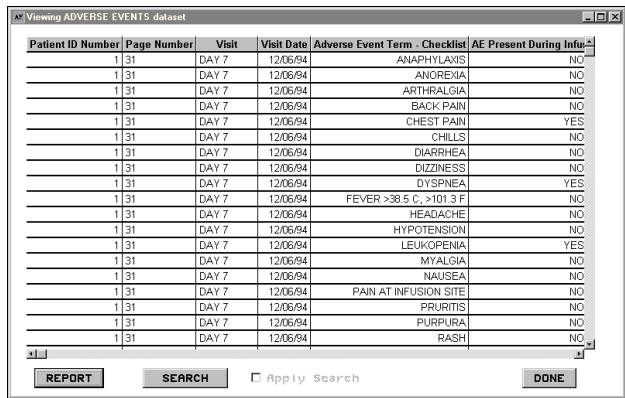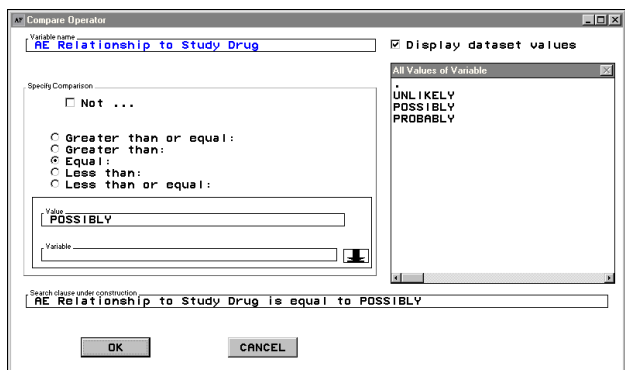

Figure 19. Viewing data in tabular format

Figure 20. Constructing a query

When viewing data, the user has access to a very user-friendly query procedure referred to in the system as a "Search." Figure 19 shows a search button at the bottom of the screen. This system allows the novice user to construct a query that is displayed to the user in plain English.

Figure 20 shows one of the dialogs used in constructing a query. The English text of the query is displayed in a text box at the bottom of the screen. An actual query can contain several conditions like that displayed in figure 20 and these lines can be connect with logical operators AND and OR.

CONCLUSION

SAS/AF provides many classes for developing solid clinical applications. More important is the ability to derive subclasses. In particular, the composite class allows classes to be combined to create very complex classes that can be reused in many applications. Systems developed in SAS/AF also have the advantage of all the analytical tools available in SAS.

REFERENCES

Haske, Carl R. (1995), "Using SAS/AF and Frame Entry to Access Data," *Proceedings of the Twentieth Annual SAS Users Group International Conference,* 647-651.

Haske, Carl R. (1995), "Developing SAS/AF® Applications for Reviewing Clinical Data," *Proceedings of the 1995 Midwest SAS Users Group Conference,* 5-9.

Haske, Carl R. (1996), "Taking Advantage of Inheritance in Developing SAS/AF® Applications," *Proceedings of the Twenty-First Annual SAS Users Group International Conference,* 11-16.

Haske, Carl R. (1996), "A Clinical Data Management System in SAS®," *Proceedings of the Twenty-First Annual SAS Users Group International Conference,* 1217-1222.

Haske, Carl R. (1996), "Developing Clinical Software Systems in SAS/AF®," *Proceedings of the 1996 Southeast SAS Users Group Conference,* 88-97.

ACKNOWLEDGMENTS

AUTHOR'S ADDRESS

Carl R. Haske, Ph.D.
STATPROBE, Inc.
3885 Research Park Drive
Ann Arbor, MI 48108
(313) 769-5000 x115
E-Mail: chaske@statprobe.com