

Clinical Data Management using SAS/AF®

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, when 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. 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.

INTRODUCTION

This paper describes a clinical data management system developed in SAS/AF. The sections of the paper correspond roughly to the primary system menu (see figure 1).

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 to administrate 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. These are 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.

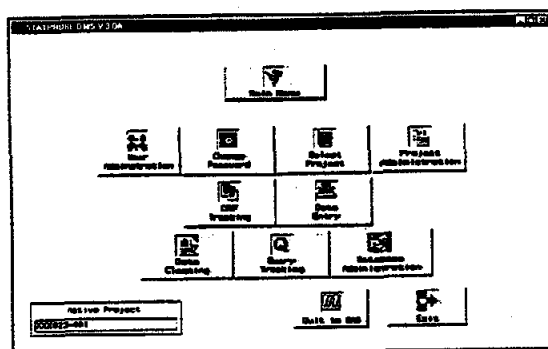


Figure 1. Main System Menu

The first section of the paper, Managing Clinical Data, details the specific tasks of the five main activities associated with managing data in clinical trials. The second section of the paper, Project Administration, discusses the project management aspects of the system. The third section of the paper, System Administration, discusses registering users and the log on system. The System Architecture section presents a brief background into the classes and templates that were used in building the system. The final section of the paper describes the STATPROBE Data Access application.

MANAGING CLINICAL DATA

CRF Tracking

The CRF tracking process involves printing CRF page registration, logging CRFs, and reporting the status of

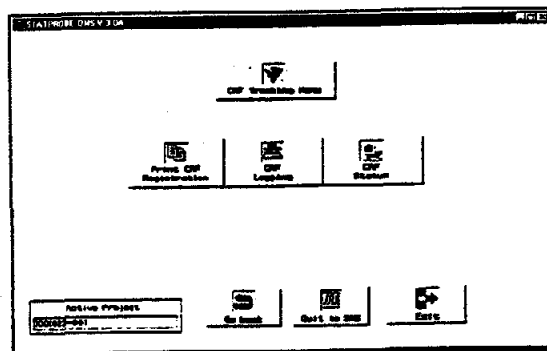


Figure 2. CRF Tracking Menu

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.

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 satisfies the list of criteria is printed to the output window or displayed in a viewer.

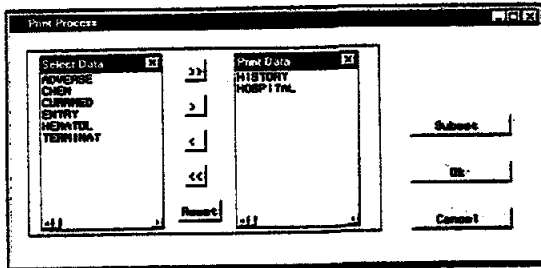


Figure 6. Selecting data for batch printing

For batch printing, the user selects the data they wish to print using mover lists (see figure 6) and the user is able to specify a set of criteria to subset all the data sets. The criteria is based on fields that occur in all data sets.

After data is double entered, it is routed from the entry slots to the PRIMARY and DOUBLE slots. Data is cleaned by following systematic steps to clean internal inconsistencies and to resolve inconsistencies in data between the two slots. Inconsistencies are resolved by referencing the original CRFs. See figure 7 for the data cleaning menu. Data cleaning options include:

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

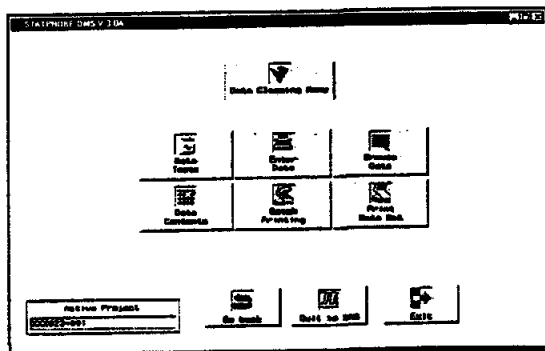


Figure 7. Data Cleaning Menu

- resolve duplicate PRIMARY and DOUBLE records
- resolve non-matching PRIMARY and DOUBLE records

- resolve PRIMARY and DOUBLE matching record inconsistencies at the field level

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

The non-matching observation process checks for one-to-one correspondence 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. Non-matching records are printed in a report for resolution.

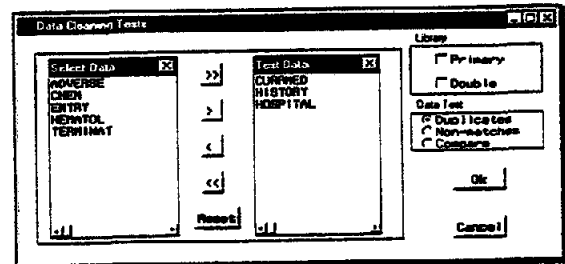


Figure 8. Data Cleaning Tests

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

When discrepancies are found during the cleaning process, the data cleaner selects the Enter Data option to modify the appropriate data. Additional tools available for the cleaning process are browsing, data contents, batch printing, and printing a single data set. The functionality for these additional data cleaning tools is similar to the analogous tools for data entry.

After data has moved to QA, data queries are resolved. Queries are generated when data is entered inconsistently on a CRF. For example, if a male subject had height recorded as 6 feet and weight recorded as 95 pounds, a data query would be generated. Queries are usually resolved at the site where the data was collected and entered on the form. Query resolution is the final step in database quality assurance. The data set that tracks queries is described in table 3.

ATTRIBUTE	DESCRIPTION
Crf_page	Unique CRF page number or id
Crf_keys	Key identifying fields
Crf_fid	Entry field name
Q_val	Query value
R_date	Resolution date
R_val	Resolved value

Table 3. CRF query tracking

The two primary tasks involved in tracking queries are logging queries to the query data set and modifying data according to a query resolution. When a query is resolved, the data in the QA slot is modified. In order to modify data in the QA slot, the user must have Database Administration clearance and access the data via the Database Administration Utilities menu explained in the next section. Reporting and printing operations on data are available to help with query resolutions.

An audit trail of all modifications to the QA data is generated. The audit trail tracks the person making the change, date of the change, the name of the variable changed, the old variable value, the new variable value, and the reason for the change. After all queries are resolved, data is ready to be frozen and moved to the FINAL slot. Data in the FINAL slot is validated and ready for analysis.

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

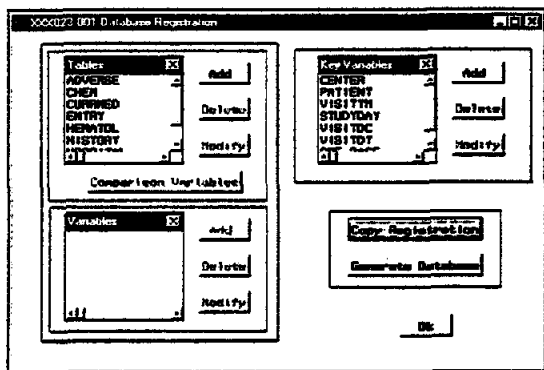


Figure 9. Database Registration

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 4 and 5).

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

Table 4. Database table dictionary

When the database administrator designs or updates the database, the database dictionary is modified and the file structures are updated.

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

Table 5. Database column dictionary

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 discussion section.

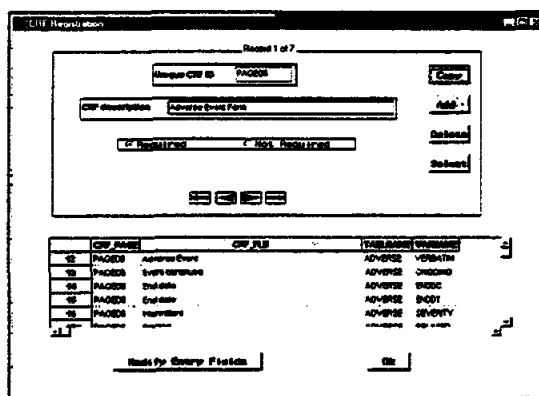


Figure 10. CRF Registration

The auto-link specification allows the database administrator to order the data sets in a manner

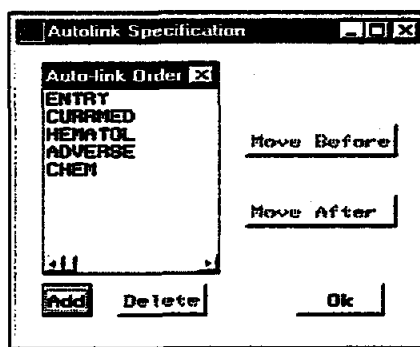


Figure 11. Auto-link Specification

corresponding to the natural flow of the CRFs (see figure 11). This saves time for the data entry person by not requiring them to select data sets and automatically opening data sets for entry in the order specified by the auto-link sequence.

Each time data is 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.

Figure 12. Batch routing

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

Status reports are also available to the database administrator. The system tracks key parameters such as entry personnel, cleaning personnel, entry dates, clean dates, lock dates, and freeze dates. The system can scan the data in each slot and generate reports on the status of the clinical database, entry logs, cleaning logs, and audit trails.

PROJECT ADMINISTRATION

Project administration is accessible only by a user that 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

To modify assignments on a project, the administrator selects the modify assignments button and a form allows additions, deletions, or modifications of personnel that are assigned to the project (see figure 14).

Figure 14. Assignment modification

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.

Figure 15. Assignment selection

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 available projects 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.

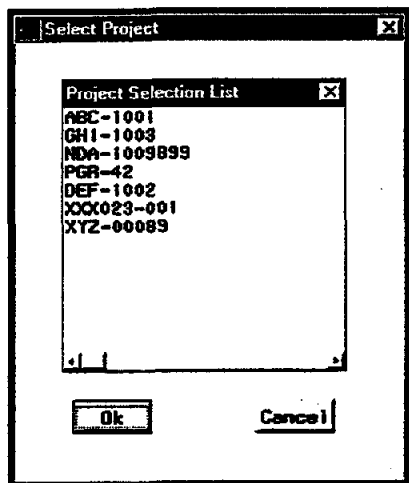


Figure 16. Project selection

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 6.

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

Table 6. Project Profile

Security for the high level processes of the data management system include 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, meaning that 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 7 displays the data that is stored by the form in figure 14.

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

Table 7. Project assignment table

SYSTEM ADMINISTRATION

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 only login in to the system if the system administrator has added their user profile to the user data set. Table 8 shows the structure of the user data set.

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

Table 8. User Table

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 utilized.

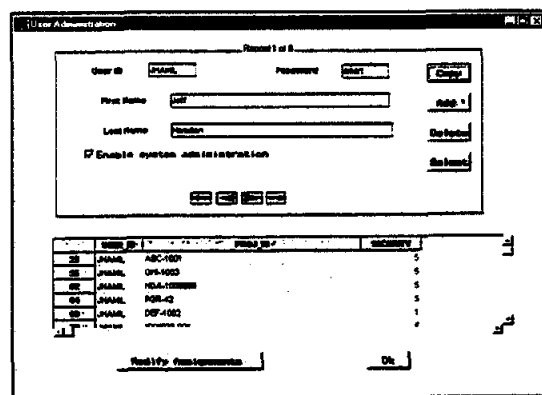


Figure 17. User Registration

Another system administration task allows a supervisor to set the path to the system database. This option is typically set at application installation and is rarely used thereafter. This option can be used to copy or backup the entire application to a different physical location. It also allows for multiple data management systems to be installed, each with a different user base and project base. The physical location of the system database is typically the same as that of the application.

SYSTEM ARCHITECTURE

The foundation of the application is STATPROBE's generic SAS/AF classes and templates. The class library and template library development was independent from the development of the clinical aspects of the system.

These classes and templates are used so that all the database applications developed at STATPROBE will have similar characteristics. Using the classes and templates also allows for a rapid application development environment. Figure 18 shows some entries in the class library.

[illegible]

One template that is reusable in many applications is an error log. The allows the system maintains an error log to help in debugging and validation. If a system error occurs, a record is written to the error log. The user can post a note on this record describing the circumstances of the error. This helps the application developers track errors and maintain the system. Table 9 shows the structure of the error log data. This data is written to a SAS catalog and stored using SLIST structures.

ATTRIBUTE	DESCRIPTION
Err_date	Date of the error
Err_time	Time of the error
User_id	User identification
Err_msg	Error code
Err_src	Source location of the error
Err_info	Processing information
User msg	User posted message

The class library provides tools for the user to construct, modify, copy, and view data. In addition, typical actions like appending, sorting, and printing data are handled by classes.

0124-0125									
NAME	SSN	DOB	AGE	STATE	RELIGION	EDUCATION	EMPLOYMENT	RESIDENCE	REMARKS
1	PAQUEN	030604	2	3	22		402 05-01-60	000000 AMERICA	
2	PAQUEN	030604	2	3	22		403 05-02-60	000000 AMERICA	
3	PAQUEN	030604	2	3	22		404 05-03-60	000000 AMERICA	
4	PAQUEN	030604	2	3	22		405 05-04-60	000000 AMERICA	
5	PAQUEN	030604	2	3	22		406 05-05-60	000000 AMERICA	
6	PAQUEN	030604	2	3	22		407 05-06-60	000000 AMERICA	
7	PAQUEN	030604	2	3	22		408 05-07-60	000000 AMERICA	
8	PAQUEN	030604	2	3	22		409 05-08-60	000000 AMERICA	
9	PAQUEN	030604	2	3	22		410 05-09-60	000000 AMERICA	
10	PAQUEN	030604	2	3	22		411 05-10-60	000000 AMERICA	
11	PAQUEN	030604	2	3	22		412 05-11-60	000000 AMERICA	
12	PAQUEN	030604	2	3	22		413 05-12-60	000000 AMERICA	
13	PAQUEN	030604	2	3	22		414 05-13-60	000000 AMERICA	
14	PAQUEN	030604	2	3	22		415 05-14-60	000000 AMERICA	
15	PAQUEN	030604	2	3	22		416 05-15-60	000000 AMERICA	
16	PAQUEN	030604	2	3	22		417 05-16-60	000000 AMERICA	
17	PAQUEN	030604	2	3	22		418 05-17-60	000000 AMERICA	
18	PAQUEN	030604	2	3	22		419 05-18-60	000000 AMERICA	
19	PAQUEN	030604	2	3	22		420 05-19-60	000000 AMERICA	
20	PAQUEN	030604	2	3	22		421 05-20-60	000000 AMERICA	
21	PAQUEN	030604	2	3	22		422 05-21-60	000000 AMERICA	
22	PAQUEN	030604	2	3	22		423 05-22-60	000000 AMERICA	
23	PAQUEN	030604	2	3	22		424 05-23-60	000000 AMERICA	
24	PAQUEN	030604	2	3	22		425 05-24-60	000000 AMERICA	
25	PAQUEN	030604	2	3	22		426 05-25-60	000000 AMERICA	
26	PAQUEN	030604	2	3	22		427 05-26-60	000000 AMERICA	
27	PAQUEN	030604	2	3	22		428 05-27-60	000000 AMERICA	
28	PAQUEN	030604	2	3	22		429 05-28-60	000000 AMERICA	
29	PAQUEN	030604	2	3	22		430 05-29-60	000000 AMERICA	
30	PAQUEN	030604	2	3	22		431 05-30-60	000000 AMERICA	
31	PAQUEN	030604	2	3	22		432 05-31-60	000000 AMERICA	
32	PAQUEN	030604	2	3	22		433 05-32-60	000000 AMERICA	
33	PAQUEN	030604	2	3	22		434 05-33-60	000000 AMERICA	
34	PAQUEN	030604	2	3	22		435 05-34-60	000000 AMERICA	
35	PAQUEN	030604	2	3	22		436 05-35-60	000000 AMERICA	
36	PAQUEN	030604	2	3	22		437 05-36-60	000000 AMERICA	
37	PAQUEN	030604	2	3	22		438 05-37-60	000000 AMERICA	
38	PAQUEN	030604	2	3	22		439 05-38-60	000000 AMERICA	
39	PAQUEN	030604	2	3	22		440 05-39-60	000000 AMERICA	
40	PAQUEN	030604	2	3	22		441 05-40-60	000000 AMERICA	

The filter object used in the frame shown in figure 4 is an example of a generic tool derived from the class library. When this object is instantiated, the developer specifies the action to be taken when filter criteria are complete. For example, the action might be browse data, print data, copy data, or some other database operation. The button labeled "Browse" in figure 4 is set dynamically to describe the action that is executed when the button is clicked. The browser class is interactive with a custom browse of the data. Figure 19 shows an example of a data browse. The browsing tool is implemented as a composite class that contains an extended table.

Select Data

ADVERSE
CHEM
CURRMED
ENTRY
HEMATOL
HISTORY
HOSPITAL
TERMINAT

>>>
>
<
<<<
Reset

Entry Data

STATPROBE Data Access

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.

Print Patient and Variant Summary

Transfer Data

On Test Protocol One
On Test Protocol Two
On Test Protocol Three
On Test Protocol Four

Patient Information

Patient Name
Patient Address
Patient Date of Birth
Patient Sex
Patient Race
Patient Religion
Patient Marital Status
Patient Education
Patient Occupation
Patient Current Address
Patient Current Phone
Patient Current Fax
Patient Current Email
Patient Current Mail
Patient Current Telex
Patient Current Cable
Patient Current Radio
Patient Current TV
Patient Current Other

SELECT ALL UNSELECT ALL

Figure 21. Data Access Main Frame

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 the user to periodically update their system with recent data.

[illegible]

When viewing data, the user has access to a very user friendly query procedure referred to in the system as a "Search". Figure 22 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.

Relationship to Study Book

☒ Display default values

Relationship of Student

- IMPLY
- POSSIBLY
- PROBABLY

Relationship

- ☒ Greater than or equal:
- ☐ Greater than:
- ☐ Equal:
- ☐ Less than:
- ☐ Less than or equal:

Possibly

OK

Figure 23 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 23 and these lines can be connect with logical operators AND and OR.

SAS/AF provides many classes for developing solid applications. More important is the ability to derive sub-classes. In particular, the composite class allows classes to be combined to create very complex classes that can be re-used in many applications.

REFERENCES

SAS Institute, Inc. (1994), *SAS Screen Control Language: Reference, Version 6, Second Edition*, Cary, NC: SAS Institute Inc.

Haske, Carl R. (1995), *Developing SAS/AF® Applications for Reviewing Clinical Data*, Proceedings of MWSUG '95, 5-9, Cleveland, OH.

Haske, Carl R. (1995), *Taking Advantage of inheritance in Developing SAS/AF® Applications*, Proceedings of MWSUG '95, 32-37, Cleveland, OH.

Haske, Carl R. (1996), *Taking Advantage of inheritance in Developing SAS/AF® Applications*, SUGI 21 Proceedings, Cary, NC SAS Institute Inc.

Haske, Carl R. (1996), *A Clinical Data Management System in SAS®*, SUGI 21 Proceedings, Cary, NC SAS Institute Inc.

SAS and SAS/AF are registered trademarks of SAS Institute Inc. in the USA and other countries. IBM and OS/2 are registered trademarks of International Business Machines Corporation. Windows 95 is a registered trademark of Microsoft Corporation. ® indicates USA registration.

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

- 72700,1075@compuserve.com
- statprob@oeonline.com