

# **SAS®: The Complete Solution For Medical Data Analysis**

**Diane E. Brown**  
**AdminaStar Solutions, Indianapolis, IN**

## **ABSTRACT**

With health care costs out of control, the importance of health care information is crucial to making appropriate revisions in the health care system. A wealth of information is held by large insurance companies, but it is often stored in large databases in a format that is hard to use for analysis and reporting. This paper discusses an application that was built for employer groups, physician/hospital organizations, and insurance companies to analyze health care data for purposes of fraud and abuse, benefit reengineering, identification of preventive health care programs, identification of case management areas, competitive pricing, and provider profiling. Technical features of this system include a relational SAS data base, SAS reporting and analysis, and a SAS/AF front end. The SAS data base contains claims tables which are indexed and several look up files to support medical coding systems and provider information. Standard SAS reports are generated by the user through parameter input including subsetting claims data prior to report generation. SAS/ASSIST and SAS/INSIGHT are available through the system for more sophisticated users. The SAS/AF front end, developed with the help of SAS Consulting Services, is entirely data driven enabling new users and new reports to be added by adding a new observation in a SAS data set. The system is Windows based and can run on a standalone PC, PC LAN, or client/server, depending on data base sizing.

## **INTRODUCTION**

AdminaStar Solutions is a subsidiary of Associated Insurance Companies, the third largest health insurer in the United States. Associated Insurance Companies holds the Blue Cross Blue Shield license in the states of Indiana and Kentucky. AdminaStar Solutions provides health data analysis for six Medicare B carriers and three Medicare A intermediaries in eight states. We have now entered into the private health insurance market. Our employer group clients consist of human resource professionals using the system primarily for benefit reengineering. Our physician/hospital clients use our system primarily for provider profiling against a large normative data base. The product being presented in this paper is our private product, currently named AMDS (AdminaStar Medical Data Systems).

## **SAS INFORMATION DATA BASE**

The AMDS system is supported by a SAS information data base. This data base provides a common format for all health claims regardless of the claims processing system. It is designed to protect the user from misinterpreting or misusing the data through data reformatting, data cleansing, and the creation of analytical and categorical variables.

The SAS information data base is built from flat files and DB2 tables within Associated

Insurance Companies for clients for which we currently process their health insurance claims. We also have a foreign flat file design to accept data from outside sources. SAS data sets were selected as the data base of choice because:

- Features of a conventional DBMS such as rollback and recovery, audit trails, transaction processing, and maintenance of referential integrity are not required.
- Complex data manipulation is required to build the SAS files from the raw data.
- Internally created variables require complex interdependencies between the SAS files and are beyond the scope of relational dependencies such as primary and foreign key relationships.
- Inserts into tables in order to maintain referential integrity are not desired because this creates rows with missing values for non-keyed variables. It is our desire to research these mismatch situations and complete the insert manually with the appropriate variable values.
- All report writing and data analysis uses the SAS System so performance is increased with a SAS data base.

The SAS information data base contains the following files:

#### 1. Global Data Files

These files are used by all clients and include the following:

- **Procedure Codes:** This file translates CPT/HCPC procedure codes into English descriptions for reporting.
- **ICD9 Procedure Codes:** This file translates ICD9 procedure codes into English descriptions for reporting.
- **Diagnosis Codes:** This file translates ICD9 diagnosis codes into English descriptions for reporting.
- **Comprehensive Procedure Groups:** These are medically assigned groups of CPT/HCPC procedures for physician claims and collapse approximately 12,000 codes into 100+ groups. For institutional claims, these same groups are formed using revenue codes.
- **Comprehensive Diagnosis Groups:** These are medically assigned groups of ICD9 diagnosis codes which collapse approximately 15,000 codes into 100 + groups.
- **Format Library:** This is a SAS format library containing descriptions for variables such as provider specialty, place of service, type of service, relationship code, rate structure, and submission type.
- **Non-institutional Normative Summaries:** These files are built annually from Associated Insurance Company internal claims files and can accommodate external customer data as foreign system data is made available with sales to customers for which we do not process their health

insurance claims. These files are used for Provider Profiling to compare providers to their peers and can also be used for independent analysis.

- **Ambulatory Care Groups (ACGS):** This file contains case-mix relative value indexes based on medically created groups and normative claims statistics.

## 2. Client Specific Data Files

These files contain client specific data and are built for each client separately.

Claims files typically contain a rolling twelve months of data refreshed quarterly, but can be adjusted based on client requirements. Claims files contain 40 to 50 variables depending on whether the claim is Institutional Inpatient, Institutional Outpatient, or Non-institutional. The three client specific claims files are indexed by Comprehensive Procedure Group, Comprehensive Diagnosis Group, Provider Specialty, and Provider Id. These variables are the most often used for subsetting the claims files prior to reporting. Run time is much faster when subsetting by an indexed variable.

- **Institutional Inpatient Claims:** This file contains hospital inpatient claims.
- **Institutional Outpatient Claims:** This file contains hospital outpatient claims.

- **Non-Institutional Claims:** This file contains physician and supplier claims.

- **Provider Master:** This file contains provider information such as specialty, name, and address for those providers which had at least one claim during the period covered by the claims files. This file also includes each provider's case mix relative value index.

- **Group:** This file contains the groups within a client's business, such as active/retiree or separate locations, along with a description of each group.

- **Membership:** This file contains a record of each member and their dependents, along with demographic information for each.

Several types of data cleansing and reformatting occur to build the SAS information data base from raw data files. This process standardizes all client specific data from any claims processing system into a common SAS data base so that all reporting and analysis tools in the AMDS system can be utilized, along with a common SAS/AF front end. It also simplifies the raw data by completing complex processing which must occur prior to reporting and analysis. This not only makes the data easier for the client to use, but avoids misuse and misinterpretation. The following items are addressed in this process:

- **Transform data values:** Claims from different claims processing systems contain different coding systems for variables such as provider specialty,

place of service, type of service, etc.. We have developed standard coding systems for all coded variables and map raw data to those values.

- **Resolve data inconsistencies:** Variable relationships have been defined and are validated. If problems are identified, they are corrected when possible.
- **Create analysis variables:** Several variables are created to support reporting and analysis which are not contained in the raw data. This includes categorical variables such as Comprehensive Procedure Group and submission type (electronic/paper), and analysis variables such as the allowed charge. The allowed charge is an important statistic in the analysis process. It is the amount of money that someone paid for the medical services rendered, including both the insurance company and the individual receiving the service. This makes an apples to apples comparison when reviewing similar claims.

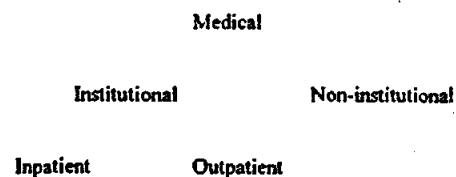
**Adjust claims:** Processed claims are sometimes adjusted in the future, typically to pay additional monies based on more medical information being supplied. Both the original and adjusted claim appear in the raw data. For the SAS file, only the final claim resolution is saved. This may mean adding the original and the adjustment or deleting the original and keeping only the adjustment, depending on how the claims processing system handles claim adjustments.

## ANALYSIS AND REPORTING

Approximately 150 standard reports have been developed to support user data analysis. These reports fall into the following categories:

- Comprehensive Procedure Groups
- Comprehensive ICD9 Procedure Groups
- CPT/HCPC Procedures
- Revenue Codes
- Provider Ranking
- Provider Profiling
- Diagnosis
- Employer
- Operational
- Hedis
- Detail Claim Listings
- Membership

In addition to the report categories, different levels of reporting are available as shown in the following diagram:



Medical reporting includes all medical claims; Institutional Inpatient, Institutional Outpatient, and Non-Institutional. Institutional reporting includes both Inpatient and Outpatient. Institutional Inpatient, Institutional Outpatient, and Non-Institutional levels include only those specific type of claims. Only those report categories which are appropriate for each report level are available. This is another way in which the system directs the user in the appropriate use of the data. Many of the reports exist at multiple report levels.

The standard reports can be produced in a variety of ways. They can be ran based on the entire claim data base or the data base can be subsetted prior to running the report. For each report, the user can click on SUBSET in the SAS/AF system and a list of appropriate subsetting variables will be displayed. Note that only those variable which are appropriate to subset for a specific report will be displayed to the user. This is another way in which the system protects the user from misuse of the data. This opens up the report possibilities available to the user in a standard reporting mode to millions of combinations.

## **HARDWARE ENVIRONMENT**

The system may be ran on a standalone PC, on a LAN, or in client/server mode, depending primarily on the size of the claims files. We currently have applications running under all three senarios. The client/server approach uses an IBM 3090 mainframe as the server.

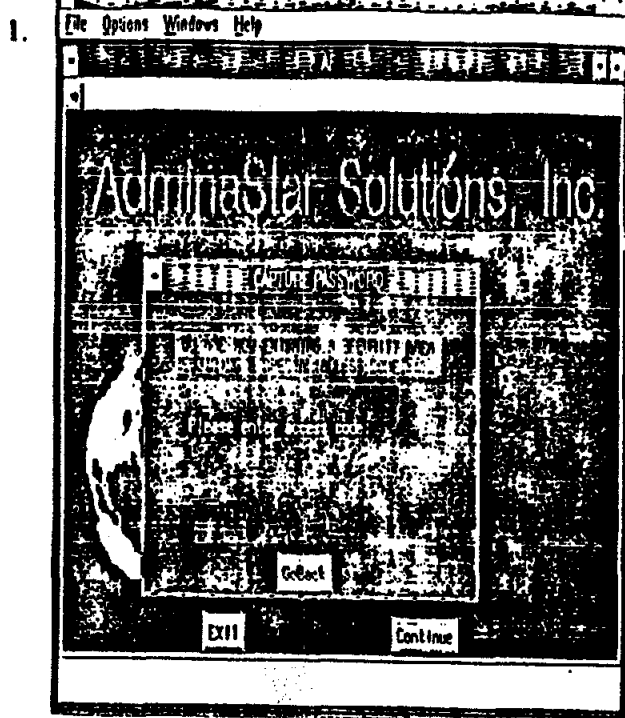
The standalone PC or LAN approach can easily support 500,000 claim line SAS files with a report response time of 2 minutes on the average when subsetted by indexed

variables. This would equate to a client with approximately 300,000 claims over the desired time frame. Larger data bases can be built depending on disk availability and user satisfaction with response time. As an alternative, multiple data bases could be built for independent analysis. This may be a good alternative for arrangements such as employer group benefit years or regional separation.

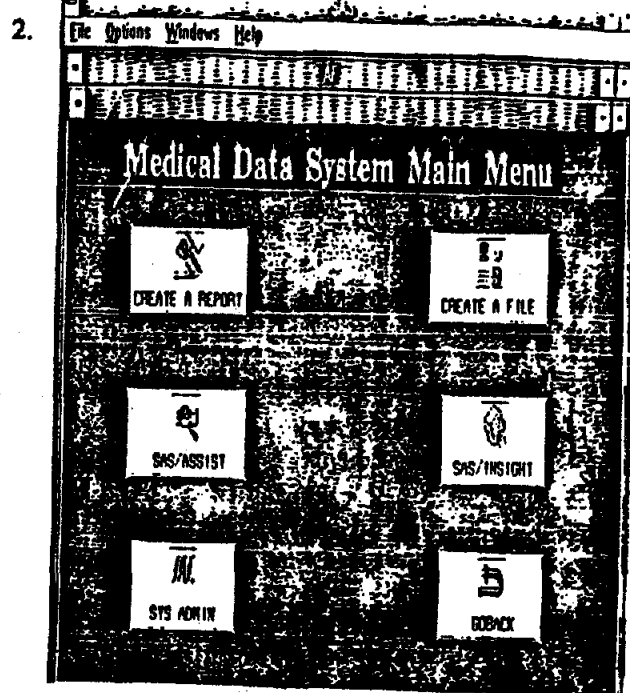
PC recommendations for the PC or LAN server include a 486 machine, the faster the better, 12+ meg memory, VGA monitor, mouse, 40-70 meg hard disk for the SAS System under Windows (depending on the SAS products selected), and 150 meg hard disk for the client with 300,000 claims (500,000 claim lines). Software requirements include Windows 3.1, DOS 5.0, Base SAS, SAS/GRAPH, and SAS/FSP. SAS/ASSIST is recommended and all other SAS products are optional.

For the client/server environment, PC recommendations include everything shown for the PC/LAN version with the exception of the hard disk space required for the client data. In addition a modem may be required, depending on the user environment. Software for the PC include all shown for the PC/LAN version, but in addition, SAS/CONNECT is required and Extra for Windows is required for access through a modem. On the server side, SAS software requirements include Base SAS and SAS/CONNECT.

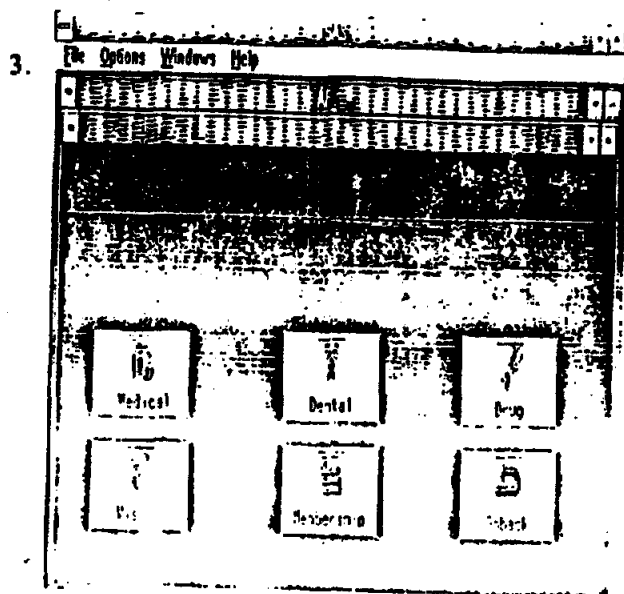
## THE AMDS SYSTEM



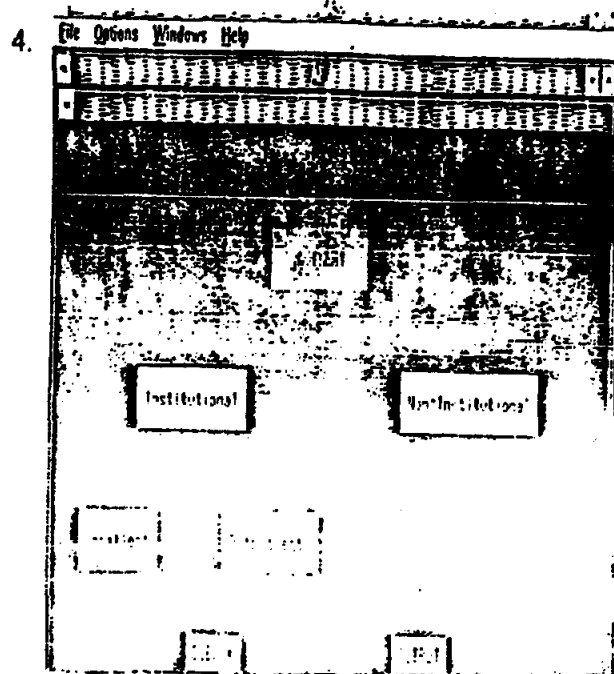
Enter password.



Select 'Create a Report'.



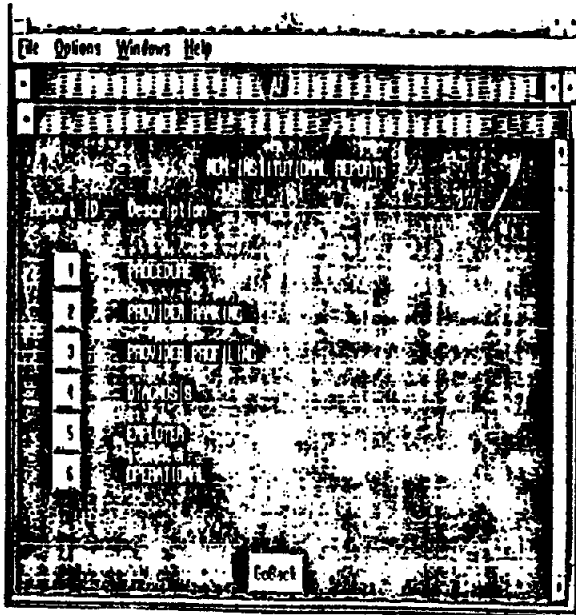
Select 'Medical'



Select 'Non-Institutional'.

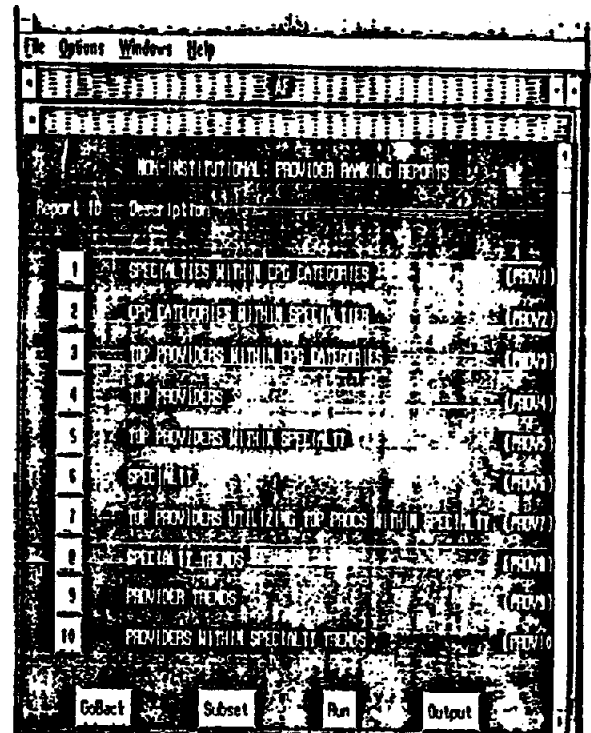
## THE AMDS SYSTEM

5.



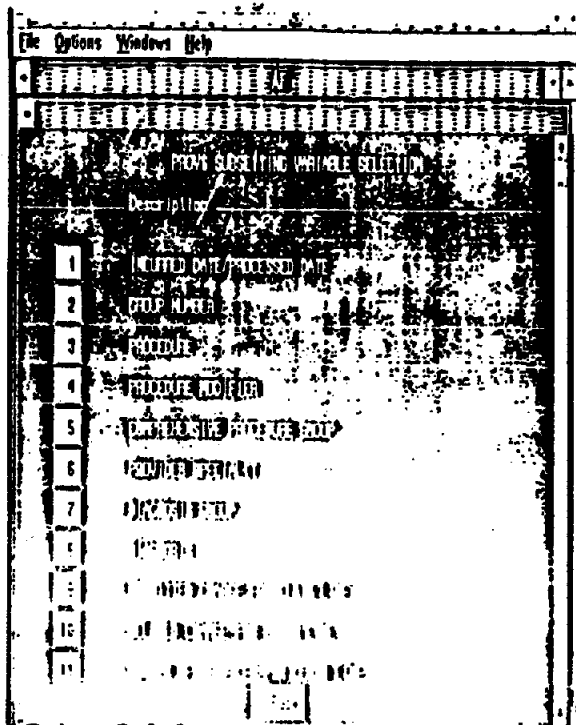
Select 'Provider Ranking'.

6.



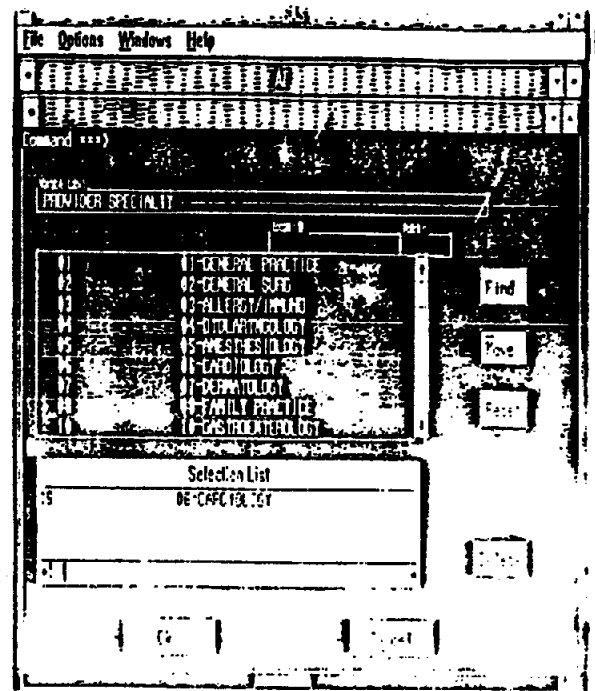
Select report 3 and click on Subset.

7.



Select 6, Provider Specialty.

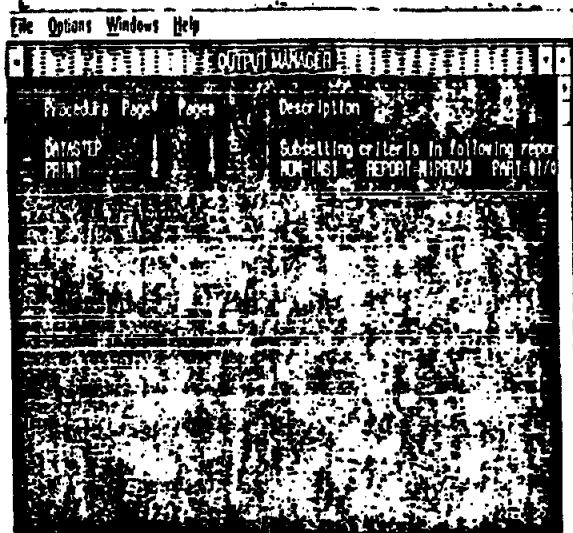
8.



Select on 06 Cardiology, click on Move, and click on Ok. You return to screen 6 and click on Run.

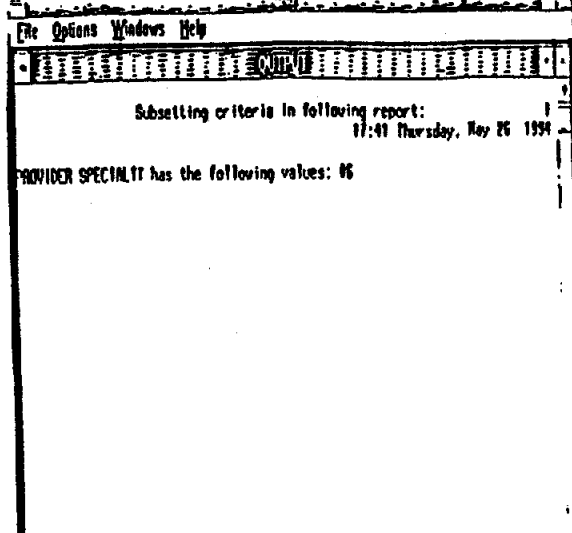
## THE AMDS SYSTEM

9.



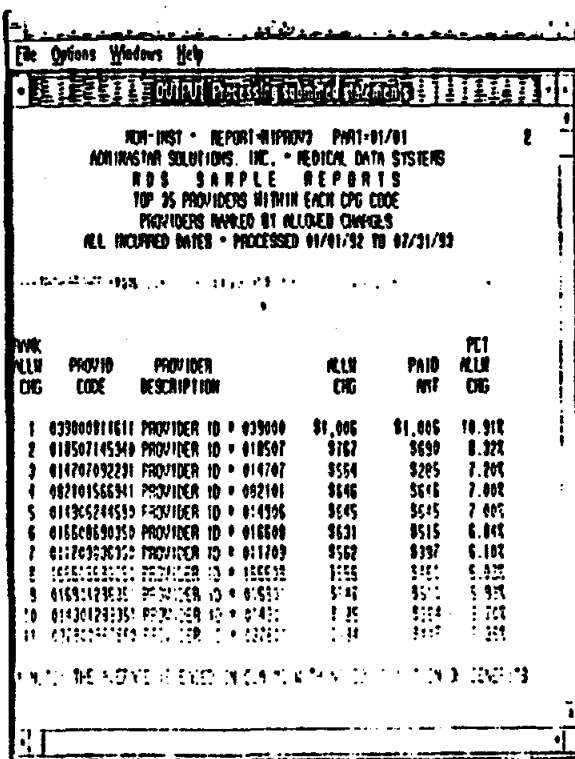
Select 'Subsetting criteria'.

10.



Scroll down.

11.



## **SAS/AF INTERNALS**

SAS/AF was used to build a user front end for the generation of standard reports. SAS/ASSIST and SAS/INSIGHT are available for the more sophisticated user and can be accessed through the SAS/AF front end.

SAS Consulting Services was used to develop the SAS/AF system due to limited resources and availability of local SCL programmers. Their data driven approach is a unique and superior approach for a front end system. It has made system maintenance very straightforward. After seeing this approach, it is hard to imagine developing such a system any other way. Not only were they instrumental in the development of our system, but they have educated us in the development of applications of this type. This will prove to be valuable in our future development.

The system is Windows based and incorporates features which are typical of Windows applications. It was designed to take advantage of GUI point and click features including pull downs, icons, scroll bars. Many of these features were built using FRAME entries.

The following types of files were built to support the SAS/AF component of the system:

- Global Support Files; consisting of include libraries, an autoexec, and Base SAS source code used to produce the AMDS standard reports.
- Client Specific Support Files; consisting of include libraries, control files, and autoexecs.

The SAS/AF system files include the SAS/AF catalog, stored macros, parameter control files, and PMENU source code. The data driven approach is based on the parameter control files and include the following files:

- Report Control File: Contains all reports available for each report level.
- Valid Subsetting Variables: Contains the valid subsetting variables for each report.
- Subsetting Variables: Contains each subsetting variable and the label displayed to the user. It is used to build a WHERE clause for subsetting and contains control information for this such as numeric/character and the type of selection operator (IN, OR, AND, comparison operators, etc.).
- Subsetting Variable Values: Several of these types of parameter files exist and allow the application programmer to update data values such as add a new provider specialty, or change a Comprehensive Procedure Group description.
- Report Groups: Contains report level and category combinations.
- Report Titles: Contains menu title for each menu displayed to the user.
- User and Security: This file controls the password and the application environment with information such as the PC drive, the client directory and whether the application is PC standalone, LAN, or client/server.

- **Application Menulist Selections:** For preliminary menus for which icons are used, this file points to the selected icon.
- **Parm File Maintenance Control Dataset:** This file controls all the above parameter files as far as sort order, whether a format is to be created, and whether the data set is indexed. When any of the above files are edited, this file automatically updates based on this control information.

## ACKNOWLEDGEMENTS

This system would not have been possible without the contributions of the following people:

- Lisa Aberson's expertise in the raw data files and knowledge of health insurance claims processing was crucial in the data base design and identification of data issues.
- Trina Houston wrote the complex code required to build the SAS files from the raw data including the data reformatting and cleansing required for a true information system..
- Debbie Tinsley is REPORTS R US. Her knowledge of the required analysis and fluent report writing was invaluable.
- Al Toney and others from SAS Consulting Services, who built and taught us how to build a data driven SAS/AF application. The system development was on a tight schedule and Al was my body guard!

You can contact the author at:

Diane E. Brown  
AdminaStar Solutions  
9525 Delegates Row  
Indianapolis, IN 46240

(317) 581-4159

SAS, SAS/AF, SAS/FSP, SAS/CONNECT, SAS/GRAPH, SAS/ASSIST, and SAS/INSIGHT are registered trademarks or trademarks of SAS Institute, Inc, Cary, NC, USA in the USA and other countries.