ABSTRACT

Data at the physician level plays an integral part in the Decision Support work of the pharmaceutical industry. The sheer volume of data available and the often disparity between data sources provides many challenges for processing and analyzing these data. Historically, physician level data have resided on multiple platforms, further complicating processing.

Both production applications and ad hoc systems have been developed using the SAS system to process and analyze physician level data. This paper will review the purposes of these systems, the functionality that they provide, and the capabilities of SAS in development and support.

INTRODUCTION

The genesis of behavioral physician data is fairly recent (within the last 4-5 years). Previous survey data, while useful has all but been replaced by these powerful data sources. On-line pharmacy systems, required for prescription adjudication created immense databases which were identified by clever entrepreneurs as having potential for use by many medically-oriented groups such as managed care organizations and research groups. The pharmaceutical industry has seen this data explosion as an opportunity to implement focused programs to benefit both the customer (physician and consumer) and the marketer.

The primary categories of physician-level data include demographic information, prescribing information and sales activity data.

APPLICATIONS

One of the earliest implementations of physician-level prescription data was promotion response testing. Any form of product promotion is a costly undertaking, the objective of which is to increase market share. Due to the many and varied products available in the pharmaceutical industry there are additional objectives such as increasing product awareness and knowledge and appropriate product usage. The goal of promotion research is to identify what promotions work and the ROI achieved as result.

The second major application developed for the physician data was targeting and micromarketing. By developing specialized product messages through factor analysis and segmentation techniques, it is possible to identify key product users and advocates and determine whether these segments offer sufficient potential to warrant development of
marketing programs. SAS procedures used include FACTOR, CLUSTER and DISCRIM.

Once target lists are developed, strategies to optimize message delivery can be implemented. With the multitude of media available, including print, personal selling and a host of new options, it is important to determine the appropriate level and mix of media to obtain necessary ROI levels. This is accomplished through segmentation and post-activity testing using the physician-level data. After program implementation, response testing and primary research is performed to assess the program effects.

The largest portion of activity against the physician data comes in the form of ad hoc requests. Access to the data through SAS has created an overwhelming number of information requests. These range from very simple data counts to complex models. Ad hoc requests accessing the physician data constitute approximately 70-80% of the staff resources involved in analyzing the physician data. One goal we have for 1995 is to shift the simpler requests to the occasional user thus allowing the more technical staff additional time to construct and test the more complicated analyses.

Since we have physician data across multiple therapeutic categories in our SAS systems, we have been able to focus on areas where a shift in prescribing activity has occurred and take advantage of these trends. It is then possible to design specific messages and programs of interest.

National audit data provide very important benchmarks in the pharmaceutical industry and are closely monitored by management and investors alike. With the major shifts in U.S. demographics and implementation of health care options (e.g., managed care), a definite need to monitor marketing opportunities and trends at the regional level has emerged. While fully projectable at the national level, audit data does not allow regional analyses but physician-level gives us this option. In addition to evaluation of activity in the current alignment, we use SAS to model and test possible changes in alignments to better organize selling activities.

The most challenging use of physician data is development of models to predict product use. We have been using SAS and physician level data to develop, test and improve various response models. Most of the models we have been working on use GLM, however, some of the newer options look like exciting enhancements. Modeling will be worked on continuously in efforts to find the best options.

Development of primary research recruitment lists is a very straight-forward application in which physician data and SAS have been used. Use of the data reduces the costs for primary research by pre-identifying research candidates and providing randomization techniques, especially important for quantitative market research studies.

Overall, movement from the mainframe into the DBMS environment and the use of SAS/ASSIST® and SAS/ACCESS® views to access the data has increased interest in and availability of physician level data to many researchers. We predict that this will lead to identification of competitive opportunities in Marketing and Sales Analysis. In addition to ease of access through user-friendly tools, SAS support provided by our IS groups is key to additional system use.

CAPABILITIES

Because physician level data is derived from many sources, a considerable amount of effort is required to process and validate these data. A recent effort has been started to consolidate all physician level data to a single platform to reduce redundancy and to facilitate access. SAS plays a significant role in these activities: The ability to perform extensive data manipulation (at both the file and attribute levels) provides the mechanism to efficiently integrate physician level data. Because data is often derived from different operating system platforms, functionality across platforms has made much of the integration possible. SAS also provides interfaces to the data, including batch processing, SAS/ASSIST, and SAS/ACCESS.

One example of SAS's ability to manipulate data efficiently is the need to merge different
sources of physician demographic information onto a single record for each physician. Most data sources have common information (e.g. last name, address) and each individual source may contain information that is unique (e.g. practice type, status of targeting). One single merge statement allows multiple sources to be integrated.

When these data sources are merged with SAS, it is also possible to create an attribute that details which sources contain information on a particular physician. This is accomplished by setting the IN= option for each input file. SAS automatically will update the defined variable and this information can be used in a single SAS statement to generate a code that identifies the sources where a physician comes from. The usefulness of this attribute is in tracing back information on a given physician and also for subsetting the data to include only physicians from a given data source.

An additional feature is the ability of SAS to write out multiple records from a single read. This is very useful in applying validation logic and then segregating data into production files or exception files. For example, data on sampling has a product code that is compared against table of known product codes. Rather than loading data with an invalid code, these data can be easily segregated into an exception file for review and later processing.

Physician-level data is being loaded into a third-party DBMS. The structure of the input files must meet prescribed formats and SAS is very useful in customizing data to meet these formats.

Some data, such as information on call and sampling activity, are available on UNIX. Other data, such as prescribing information, are available in tape formats that are on the MVS® platform.

Data on MVS is processed in flat file format and loaded into SAS databases. These data are then transferred across a network link to the UNIX environment. By using the XPORT Engine option when building the SAS databases, several benefits are derived: One, a single pass is made from the source tape into the SAS database. Two, once migrated across the network, the data is in a compatible format with other physician-level data.

As source data becomes native to the UNIX environment, the SAS code that is used on MVS to extract and validate the tape data can be migrated to the UNIX platform. Migration of SAS code between MVS and UNIX has been, overall, very straight-forward.

SAS provides several mechanisms to access physician-level data. Historically, batch processing has been used in both the MVS and UNIX environments. This is very applicable for those comfortable and capable of writing computer programs. Programs can easily be stored for re-use and can be modified to handle ad hoc requests. Batch also provides the benefit of background processing: a terminal or workstation does not have to be tied up while the program is executing.

An application of batch processing was developed to select information from multiple files for a given physician. SAS’s macro facility was used to pass a string of physician identification numbers between multiple data steps that accessed different levels of physician data (e.g. demographics, market share).

SAS/ASSIST is a tool that is more applicable for clients who have had limited exposure to computer programming. SAS/ASSIST has been used to extract information for more tactical uses, such as finding demographic information for an individual physician. As physician-level data is becoming more centralized, it also is becoming more normalized (i.e. less duplication, more unique files) that require the client to merge files. SAS/ASSIST provides this mechanism, but requires more in-depth understanding of the data and its relationships.

SAS/ASSIST is very interactive in nature and an expectation exists that queries made in an interactive mode will be completed in a timely manner. Given that many of the physician-level files are large, indexing these SAS databases greatly improves access to the data.

A benefit of using SAS/ASSIST is the ability to create subset files using the GUI front-end (i.e. pull-down lists to select variables to keep) and then be able to drop back into SAS’s Display...
Manager to perform more complicated manipulation or analysis.

SAS/ACCESS is beginning to be used as a means to access physician-level data in a third-party DBMS. Views to files, or tables, within the DBMS are created and stored in SAS files. DBMS data can be accessed through standard SAS routines (i.e. Data steps, Proc Prints), or more directly through native SQL code.

SAS/ACCESS can also be used to load data into the DBMS and to control certain operational features of the DBMS, such as security access.

NOTES

SAS, SAS/ACCESS, and SAS/ASSIST are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries.

MVS is a registered trademark or trademark of International Business Machines Corporation.

® indicates USA registration.

Other brand and product names are registered trademarks of their respective companies.

REFERENCES


AUTHORS

Gail Kirby
G.D. Searle
5200 Old Orchard Road
Skokie, IL 60077
(708) 470-6938

Daniel M. Plath
G.D. Searle
5200 Old Orchard Road
Skokie, IL 60077
(708) 470-6790