Paper 2082-2014

Leveraging SAS Visual Analytics for Healthcare Research

Al Cordoba, Truven Health Analytics; Jim Fenton, SAS Institute, Inc.

William D. Marder, PhD, Truven Health Analytics; Anthony Pepitone, Truven Health Analytics;

ABSTRACT

There is increasing interest in exploring healthcare practices and costs for the US population to improve the quality, efficiency and outcomes of care. Fortunately, comparative data derived from inpatient, outpatient and prescription drug claims is available to evaluate and benchmark financial and clinical performance. The data source described in this paper represents the collective experience of tens of millions of people over many years. While it is highly useful to visualize trends and relationships in the data, this has always been difficult because of the scale. In this paper, the authors describe their experience extracting data from Truven Health MarketScan® Research Databases, loading the data into SAS® Visual Analytics and exploring the data to better understand drug adherence patterns for a population of diabetes patients.

INTRODUCTION

There are four main components that contribute to effective research: analytical software, research methods, domain expertise and data. In this paper we discuss how SAS Visual Analytics, a software tool for visualization and analytics, the knowledge of pharmaceutical drugs and analytical methods supplied by Truven Health experts and the Truven Health MarketScan Research Databases help improve healthcare research.

SAS Visual Analytics is a complete platform for analytics visualization that enables a user to identify patterns and relationships in data that weren't evident using traditional methods. It combines powerful in-memory technologies with an easy to use drag and drop interface to interactively explore data, perform complex analytics and get extremely fast results without the need to write code. The insights gained from these discoveries can be shared across the organization using web interfaces and mobile devices. SAS Visual Analytics platform easily scales to provide high-performance, big data analytics using the Truven Health MarketScan Research Databases.

The Truven Health Analytics MarketScan Research Databases – containing the healthcare experience of over 10 million people since 1995 – are widely used by a variety of stakeholders to analyze healthcare utilization, clinical practice and cost. The data has fully integrated medical, prescription drug, and in many cases laboratory, dental, and vision claims at the patient level. The MarketScan databases have broad geographic coverage, strong longitudinal reach over many years, are detailed at the patient level and reflect the true continuum of care. MarketScan data are derived from multiple sources (employers, health plans, Medicaid, and other carriers) and generally provide the ability to track patients even when they switch health plans and stay with a single one of the Truven customers. While the data have been comprehensively de-identified, they retain full clinical content including exact dates of all inpatient, outpatient and prescription drug events by patient. (The data are statistically de-identified under the HIPAA regulations – not using the safe harbor approach.) In addition, the large number of patients and detailed patient-level information allows for assessment of specific constellations of conditions and treatments that might otherwise be rare or missing in smaller databases.

As a result of their strengths, the MarketScan databases enable accurate comparative analysis with data tracking of patients over a number of years and across multiple health plans. Consistency across services is achieved by in-depth, patient-level detail linked by unique identifiers. The information provided by the MarketScan databases are utilized by multiple departments including: network development/provider contracting, medical informatics/economics, data management, group account management, and care management.

Truven Health MarketScan Research Databases are segmented into categories to help a researcher find the specific claims data they need. We looked into two categories: The Truven Health MarketScan Commercial Claims and Encounters Database which provides in-depth, cross-sectional, and longitudinal views of healthcare. Fully adjudicated, patient-level claims are linked with other MarketScan databases. Linking data at the patient level significantly enriches the insights derived from claims-based research. We also looked into the Truven Health MarketScan Medicare Supplemental and Coordination of Benefits Database which includes both the Medicare-covered and employer-paid portions of the healthcare encounter. Detailed cost, utilization, and other information from inpatient and outpatient settings allows tracking of employees and dependents moving through their Medicare years.

In medicine, adherence describes the degree to which a patient correctly follows medical advice. Most commonly, it refers to medication or prescription drug compliance. Non-compliance is a major obstacle to the effective delivery of

health care. Both the patient and the health-care provider affect compliance, and a positive physician-patient relationship is the most important factor in improving compliance, although it has been reported the high cost of prescription medication also plays a major role.

Estimates from the World Health Organization (2003) indicate that only about 50% of patients with chronic diseases living in developed countries follow treatment recommendations. In particular, low rates of adherence to therapies for asthma, diabetes, and hypertension are thought to contribute substantially to the human and economic burden of those conditions.

Efforts to improve compliance have been aimed at simplifying medication packaging, providing effective medication reminders, improving patient education, and limiting the number of medications prescribed simultaneously.

DATA MANAGEMENT

During the last quarter of 2013, a group of SAS and Truven Health experts worked to develop a prototype to better understand how data, domain expertise, and technical expertise could come together to enrich the data discovery process. Using SAS, we identified and extracted a subset of data for diabetes patients over a two year period from the MarketScan Databases. Medication adherence is critical for diabetes care, and we wanted to evaluate and compare adherence patterns. The initial file provided insight on the subsequent direction to take. To begin, we developed indicators to identify the number of medications prescribed simultaneously. We quickly discovered during initial visualization explorations that the in-memory SAS Visual Analytics platform eliminates the need to code data and apply display formats to reduce the data volume for processing, which resulted in a V2 update to provide more descriptive detail to the data. A description of the initial file is included in Table 1 below. In the file layout below for SAS VA diabetes file #1, the data transformations or additions from version 1 to version 2 are highlighted. The file ended containing 11 variables.

Variable Name	Type, V1	Type, V2	Variable Label	ETL Comment
ENROLID	Num	same	Enrollee ID	
AGE	Num	same	Age of patient	
GENDER	Char	same	Gender	Change M to Male, F to Female
INDEX_DT	Num	same	Date Service Incurred	
INDEX_DIABETES_DRUG	Char	same	Name of the index diabetes drug	
FILLDATE	Num	same	Dispensing Date	
THERCLS	Num	Char	Therapeutic Class	Change ID # to name of class
GENNME	Char	same	Generic Drug Name	
QTY	Num	same	Quantity of Services	
DAYSUPP	Num	same	Days Supply	
STRNGTH	Char	same	Strength	

Table 1. SAS VA Diabetes File #1 -- 11 Variables

The second version of this file provided more descriptive detail in the data, but we quickly exhausted everything the data had to tell us, which compelled the need for more data elements to expand the exploratory analytics discovery process.

Table 2 below is a table showing modifications or additions of new variables from File version #1 to File version #2. These changes are also highlighted.

Some variable were re-coded to change their format from numeric to character. New flags were developed to mark specific diagnostics, for instance, a congestive heart failure. The rationale was that these conditions could help explain the lack of adherence to a treatment. Also, the researchers developed indicators to show the use of a particular drug during the period, like a Metformin indicator. We captured all costs for the patient as well as the drug costs.

Type of Information	Variable Name	Type, V1	Type , V2	Variable Label
Demographic and clinical patient profile at Index	ENROLID	Num	same	Enrollee ID
	AGE	Num	same	Age of patient
	SEX	Char	same	Gender of patient (M to Male, F to Female)
	MSA	Num	same	Metropolitan Statistical Area **(ID #)**
	MSA2	absent	Char	Metropolitan Statistical Area **(name)**
	State	absent	Char	Two-character postal code
	Region	absent	Char	Name of geographic region
	PLANTYP	Num	Char	Name of insurance plan type
	MEAN_COSTSHARE	Num	same	Plan-level flag, cost sharing
	PAYER	Char	same	1 to Commercial, 2 to Medicare
	CCI	Num	same	Charlson Comorbidity Index
	ICD9_N	Num	same	Number of unique 3-digit ICD-9 DX
	PRE_ENDO_FLG	Char	Num	Patient visited an endocrinologist in the baseline
	PRE_CAR_FLG	Char	Num	Patient visited an cardiologist in the baseline
	PRE_REN_IMP_FLG	Char	Num	Patient has renal impairment in the baseline
	PRE_MACRO_FLG	Char	Num	Patient has macrovascular disease in the baseline
	Acute myocardial infarction	absent	Num	a new indicator
	Other ischemic heart disease	absent	Num	a new indicator
	Congestive heart failure	absent	Num	a new indicator
	Cerebrovascular accident	absent	Num	a new indicator
	Peripheral vascular disease	absent	Num	a new indicator
	PRE_MICRO_FLG	Char	Num	Patient has microvascular disease in the baseline
	Anxiety	absent	Num	a new indicator
	Depression	absent	Num	a new indicator

Type of Information	Variable Name	Type, V1	Type , V2	Variable Label
	Cardiac arrythmia	absent	Num	a new indicator
	Cerebrovascular disease	absent	Num	a new indicator
	Coronary artery disease	absent	Num	a new indicator
	COPD	absent	Num	a new indicator
	Essential hypertension	absent	Num	a new indicator
	Cancer	absent	Num	a new indicator
	HIV	absent	Num	a new indicator
	PRE_MET_FLG	Char	Num	Had a prescription for metformin in the pre-period
	PRE_INS_FLG	Char	Num	Had a prescription for insulin in the pre-period
	PST_PREGNANT	Char	Num	Pregnant flag in the follow-up period
	INDEX_DT	Num	same	Date Service Incurred
	MAIL_FLG	Char	Num	Index drug was mail order flag
	FIXED_MET_FLG	Char	Num	Index drug was a fixed dose metformin combo
Day number	DAY_POST_INDEX	Num	same	(1,2,3,...,730) Please Note: Day 1 = index_dt
Type of anti-diabetic medication	TZD	Char	Num	TZDs indicator
	SU	Char	Num	Sulfonylureas indicator
	DPP	Char	Num	DPP4 indicator
	METF	Char	Num	Metformin indicator
	INSL	Char	Num	Insulins indicator
	INCRETIN_MIMETICS	Char	Num	Incretin mimetics indicator
	OTHER_DIAB_RX	Char	Num	All other antidiabetic drugs indicator
Occurrence of utilization events	IP_ADMISSION	Num	same	inpatient admission indicator
	LOS	Num	same	inpatient admission length of stay
	ER	Num	same	Indicator for Emergency Department Visit
	OFFICE_VISIT	Num	same	Indicator for OP Office Visit
	OTHER_OP	Num	same	Indicator for all other OP Care Visit
All-cause cost	TOTAL_COST	Num	same	All-cause total Cost
	IP_COST	Num	same	All-cause Inpatient Costs

Type of Information	Variable Name	Type, V1	Type , V2	Variable Label
	ER_COST	Num	same	All-cause Emergency Department Cost
	OFFICE_VISIT_COST	Num	same	All-cause Outpatient Office and Clinic Cost
	OTHER_OP_COST	Num	same	All-cause Other Outpatient Care Cost
	RX_COST	Num	same	All-cause RX Cost
Diabetes-related cost	DIAB_TOTAL_COST	Num	same	total diabetes-related Cost
	DIAB_IP_COST	Num	same	Diabetes Related Inpatient Costs
	DIAB_ER_COST	Num	same	Diabetes-related Emergency Department Cost
	DIAB_OFFICE_VISIT_COST	Num	same	Diabetes-related Outpatient Office and Clinic Cost
	DIAB_OTHER_OP_COST	Num	same	Diabetes-related Other Outpatient Care Cost

Table 2. SAS VA diabetes file #2 – 61 Variables

VISUALIZATION

Data visualization is the creation and study of the visual representation of data, meaning "information that has been abstracted in some schematic form, including attributes or variables for the units of information" Friendly (2008).

According to Henschen (2012), visualization is very useful today because it makes the analysis of large collections of data easier. Analysis with more conventional BI query and analysis tools isn't so easy, according to the 2012 InformationWeek Business Intelligence, Analytics and Information Management Survey. Nearly half (45%) of the 414 respondents to the poll, which was conducted late 2011, cited "ease-of-use challenges with complex software/less-technically savvy employees" as the second-biggest barrier to adopting BI/analytics products. That was just behind the biggest barrier, "data quality problems," cited by 46% or respondents.

Our team thought that, using SAS Visual Analytics, we could explore the MarketScan Research Databases and create visualizations to uncover relevant compliance and non-compliance patterns for certain drugs used to treat diabetes, and share those visualizations in reports. Traditional reporting is descriptive. You know what you are looking for and what you need to convey. However, data discovery allows users to explore data, its characteristics, and relationships. Then, when useful visualizations are produced, those visualizations can be incorporated into reports that are disseminated in a variety of ways.

We used SAS Visual Analytics because it enables users to apply the power of SAS analytics to massive amounts of data like the MarketScan Research Databases, empowers researchers to visually explore large collections of healthcare data, based on any variety of measures, at amazingly fast speeds, and enables healthcare researchers to share insights with other healthcare researchers via the Web or a mobile device.

Figure 1 below is a model representing the data exploration activities we conducted. Our process was as follows. We identified and extracted relevant data from the MarketScan Research Databases based on our data and research domain expertise. Next, Truven researcher would examine the data, discuss findings within the team, and typically uncover a number of interesting trends or events. At that point, there were either discoveries which merited immediate action (creation of new indicators or new data model designs) or discoveries that generated new questions. New questions prompted the needed for additional indicators and derived data, and the cycle continued. That's what the diagram below shows, the iterative and cyclical nature of data-based discovery.

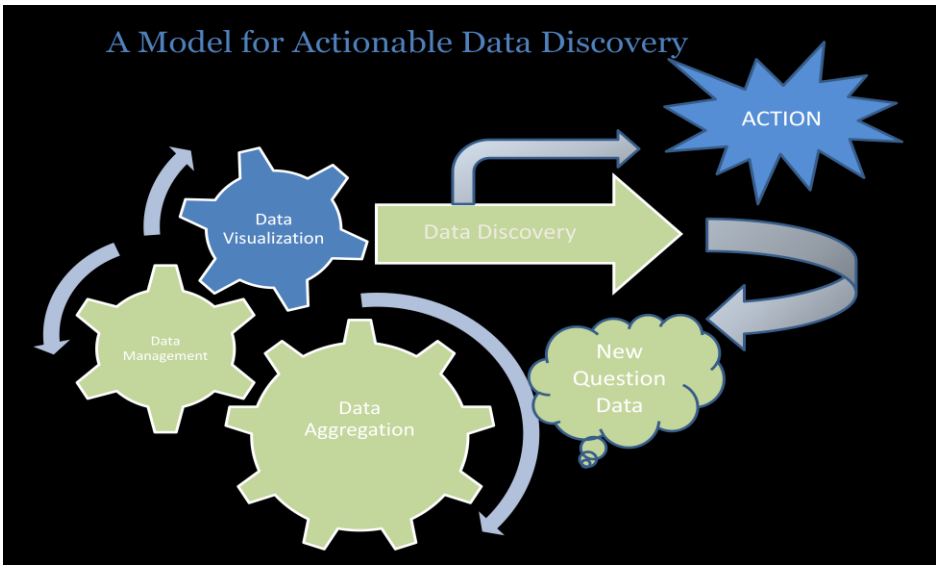

Figure 1 Data Discovery Process

Once a dataset was created, a Truven Health researcher, a pharmaceutical expert, worked with a SAS VA expert to load the data into the SAS Visual Analytics server, create on-the-fly simple data calculations, drill down and explore data without help from IT. We produced visualizations very quickly, and shared results via the Web. Once we loaded the data, our team concentrated in finding interesting patterns for drug adherence related to the diabetic population. As we were finding trends, a need to go back and reformulate new questions as well as gathering and transforming some additional data was apparent.

Figure 2 below shows the kind of dashboards built to discover new trends and patterns in the data.

Figure 2 Example Visual Analytics Drug Adherence Dashboard for Diabetes

CONCLUSIONS

In conclusion, we found:

1. The Truven Health MarketScan Research Databases provided valuable data for study. In spite of the narrow clinical specificity that much of our exploratory research required, the database was able to supply sufficient cases to support statistically significant results.
2. SAS Visual Analytics is a valuable tool to simplify healthcare data discovery. The intuitive web browser interface enables the user to interactively explore data and perform complex analytics, and the powerful in-memory platform provides extremely fast results on big data without the need to write code.
3. The number of variables increased considerably as we worked through the discovery process from an initial count of 11 variables in File #1 to 61 variables in File #2. This situation created a need to scale storage as data volume grew.
4. SAS Visual Analytics Integrates with Hadoop for performance optimization, scalability and cheaper storage and can be used on commodity hardware (we used HP blades).
5. We identified a consistent and recurring need to create new analytic variables as the discovery process moved forward. Some of this augmentation was performed inside Visual Analytics and some of it required returning to the source data for manipulation with Base SAS and SQL. Tools that support cohort creation and analytic variable construction, including SAS PROCS, SAS macros and tools such as Truven Health Analytics Treatment Pathways, add significant value in the data exploration process.

RECOMMENDATIONS

There are a few basic visualization concepts that can help you generate the best visuals for displaying data:

- Make sure you understand the data you are trying to visualize, including its size and cardinality (the uniqueness of data values in a column).
- Determine your research questions and what you are trying to visualize and what kind of information you want to communicate. Be prepared to discover something new.
- Know your audience and understand how it processes visual information. Using SAS Visual Analytics you can pack a lot of information in one single screen, so don't go to fast.
- Choose a visual that conveys the information in the best and simplest form for your audience.

If you are considering using SAS Visual Analytics to work research questions, you may want to review the following checklist:

Visual Analytics Readiness Questions

- How often is it important for someone with your role in your industry to go from question to data-driven answer in five minutes or less? An hour? A day? A week?
- How much more valuable would you find a tool that is really visual and interactive?
- Do you think you could find out new things that wouldn't be identified without such a tool? Can you give some examples of problems you would be interested in approaching?
- Suppose there was an initial implementation period involving a week of back and forth getting the data ready. How would that affect the value of the tool?
- We spend time linking disparate data sources. One of the challenges we have with potential partners who have clinical data is that they are concerned about data release of individual patient data. An interactive tool can protect against that kind of release. How important is it to you to have access to the underlying individual patient data in all of its detail?
- Can you provide the educational background of the analysts in your organization who would be interested in analyzing big data streams? Do they have programming experience or programming support available?
- Do your information technology professionals participate in all licensing of web accessible tools? All that is needed locally to use Visual Analytics is a browser.

REFERENCES

Friendly, Michael (2008). Milestones in the history of thematic cartography, statistical graphics, and data

visualization.

Henschen, Doug -- http://www.informationweek.com/software/information-management/how-to-choose-advanced-data-visualization-tools/d/d-id/1105480?piddl_msgid=198941#msg_198941. 7/24/2012

Forrester's report, The Forrester Wave: Advanced Data Visualization Platforms, Q3 2012,"www.sas.com

World Health Organization (2003). Adherence to long-term therapies: evidence for action (PDF). Geneva: World Health Organisation. ISBN 92-4-154599-2.

Friendly, Michael (2008). "Milestones in the history of thematic cartography, statistical graphics, and data visualization"

Henschen, Doug -- http://www.informationweek.com/software/information-management/how-to-choose-advanced-data-visualization-tools/d/d-id/1105480?piddl_msgid=198941#msg_198941. 7/24/2012

ACKNOWLEDGMENTS

The authors would like to acknowledge the work Dan Huse and Greg Lenhard from Truven Health Analytics who contributed to this effort. Also, the team appreciated the work that Derek Fugh conducted as the SAS statistical programmer.

RECOMMENDED READING

- *SAS Visual Analytics Guide*
- *SAS Visual Analytics Installation*

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Al Cordoba, M.S.
Truven Health Analytics Inc.
5425 Hollister Avenue, Suite 140
Santa Barbara, CA 93111
858-252-5220
Alberto.Cordoba@truvenhealth.com
Web: www.truven.com

Jim Fenton
SAS Institute, Inc.
100 SAS Campus Drive
Cary, NC 27513
919-677-8000
Jim.Fenton@sas.com
Web: www.sas.com

William Marder, PhD.
Truven Health Analytics Inc.
150 Cambridge Pk Dr
Cambridge, MA 02140
617.492.9300
William. Marder@truvenhealth.com
Web: www.truven.com

Anthony Pepitone
Truven Health Analytics Inc.
6200 S. Syracuse Way, #300
Greenwood Village, CO 80111
800.525.9083
Anthony. Pepitone@truvenhealth.com
Web: www.truven.com

ABOUT TRUVEN HEALTH ANALYTICS

Truven Health Analytics (formerly Thomson Reuters Healthcare) delivers unbiased information, analytic tools, benchmarks, and services to the healthcare industry. Hospitals, government agencies, employers, health plans, clinicians, pharmaceutical, and medical device companies have relied on Truven Health for more than 30 years. The firm provides deep clinical, financial, and healthcare management expertise with innovative technology platforms and information assets to make healthcare better. Truven Health collaborates with customers to uncover and realize opportunities for improving quality, efficiency, and outcomes. With more than 2,000 employees globally, Truven Health has major offices in Ann Arbor, Mich.; Chicago; and Denver.

TRADEMARKS