Paper 166-2013

Identifying and Addressing Post-Marketing Pharmaceutical Safety Surveillance and Spontaneous Reported Events

Carrie Boorse, Stuart Levine, and Kathy Schaan

SAS Institute Inc., Cary, NC

ABSTRACT

While pharmaceutical medications and medical devices must undergo clinical trials to determine their safety, often their long-term side effects will not be recognized until the medication or device has been approved and consumed by a larger population of patients for a longer period of time. Signals that an adverse event is brewing need to be identified by using post-marketing data as expeditiously as possible. Using SAS® analytical platform and foundation products, a solution was implemented that incorporates recognized statistical procedures and enables users to incorporate new and enhanced procedures for signal determination. The solution also alerts users that a signal has been triggered, while also managing the workflow and results of the signal's investigation.

INTRODUCTION

Clinical trials are used to determine a drug or device's effectiveness and safety. Although the trials play an important role in bringing a drug or device to market, there are limitations regarding the types of participants included in the sample population and also the length of time in which the product is consumed by those participants. Because of these limitations, once a product is released to market it will be consumed by a portion of the population that were not included in the sample, such as children, the elderly, and women of childbearing age. There are multiple post-marketing factors that were previously controlled during the clinical trials that are now able to impact the product's safety, and this might result in conditions and reactions not previously documented; those reactions are also known as adverse reactions. The science and activities related to the detection, assessment, understanding, and prevention of the adverse reactions are defined by the World Health Organization (WHO) as pharmacovigilance (Uppsala Monitoring Centre 2011).

In 2004, Merck, the manufacturer of Vioxx, removed the drug from the market due to an increase in the incidence of heart attacks and strokes in patients who took it regularly. During 2006 and 2007, Bausch & Lomb and Advanced Medical Optics recalled contact lens solution because it was associated with corneal infections that could lead to blindness. In 2007, Listerine recalled four million bottles of mouthwash because added preservatives did not provide a sufficient level of protection against microorganisms that could lead to serious health risks for people with compromised immune systems. Again in 2007, several infant-strength cold medicines were recalled due to infants being injured or killed because of accidental overdosing. And in 2010, McNeil Consumer Healthcare recalled several of their over-the-counter medications because of the packaging being contaminated by a chemical used to treat the wood pallets that were used for its transportation and storage (iVillage 2012, 5, 8, 9, 11, 4). While these are some of the largest product recalls that have taken place over the last decade, it is far from a complete list. It is because of these types of activities that pharmacovigilance and post-marketing product safety have become a hot topic and has led to the partnering of an industry leader with SAS® to create a safety surveillance solution to aid in the recognition and documentation of adverse events.

The system that is currently being implemented within this partnership will incorporate recognized statistical procedures, enable the users to incorporate new or enhanced procedures for signal detection, and alert users that a signal has been triggered as well as manage the workflow and results of the signal's investigation.

Some of the business problems the customer is looking to improve upon are: their data preparation is manual and time-consuming, their process is manual and might not be repeatable, the data is not integrated, and the current process has not been validated. From an analytical perspective there are limitations to the root-cause analysis that can be done effectively, the signals are not being detected early enough, and the algorithms being leveraged are not as sophisticated as the client would like them to be. With all of this defined, SAS and the customer are developing a system that will address these critical issues.

The signal detection solution is implemented using SAS® Quality Lifecycle Analysis, which is a SAS software solution that focuses on bringing supply chain intelligence to enterprises. This is important to life science companies that develop drugs or medical devices and that need to monitor safety issues once their drugs or medical devices are in market. SAS Quality Lifecycle Analysis provides a number of unique capabilities for this task, including an analytics-based engine for integrating and analyzing disparate and isolated data sources relevant to adverse event details, product information, and potential safety signals. It also provides best-practice workflows and a case management

feature to assist in documenting findings and problem-resolution measures, while promoting collaboration and knowledge sharing.

INTEGRATION OF SOLUTION COMPONENTS

The signal detection solution consists of system components that include:

- multiple source systems

- a data layer that incorporates both SAS processing and data preparation

- a robust metadata layer

- a set of web interfaces for reviewing analytics and case status

- a number of SAS clients used for solution implementation and maintenance tasks.

These system components are pictured in Figure 1.

At its core, the solution contains over seven hundred thousand historical records from its source systems, and it receives approximately one thousand incremental record updates per week. The data layers perform an extract, transform, and load (ETL) process on the events and product dictionary to create a safety database of over three million product and adverse event combinations. Analytical metrics to aid in signal detection are then calculated against the product- event combinations. The analytical results are then presented to users in a series of interactive web reports and dashboards. A case management database stores records that drive a web-based workflow providing detection, triage, and validation for those product-event combinations that require attention. Throughout the workflow processing of the cases, SAS metadata enables system-wide auditing, data security that includes user authentication and authorization, and the tracking, monitoring, and documentation of data movement and actions at appropriate levels of granularity.
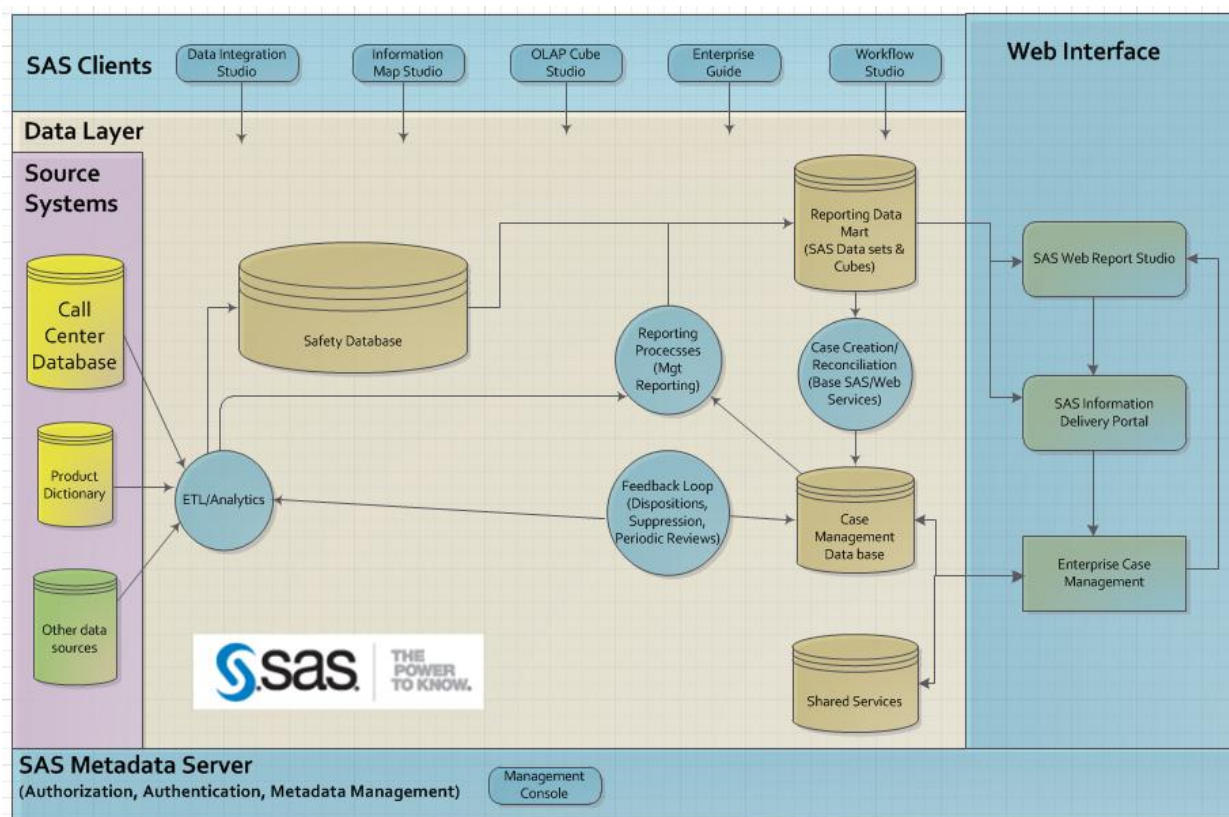
Figure 1. Signal Detection Data Flow and Processing

SAS SOFTWARE USED

The signal detection solution includes a number of SAS components, including both basic tools and more advanced solution packages.

The following is the list of software used in this solution:

- SAS® Data Integration Server pulls the various data sources together (the primary call center data along with the ancillary tables that might exist) and performs the transformations. Also deploys Data Integration Server jobs to the SAS® Metadata Server for scheduling through platform services.

    o Clients:

        ▪ Data® Integration Studio

- SAS® Enterprise BI Server is used for signal reporting, analysis (through OLAP reporting), and ad-hoc analysis (through SAS® Enterprise Guide®).

    o Clients:

        ▪ SAS® Information Map Studio
        ▪ SAS® OLAP Cube Studio
        ▪ SAS® Enterprise Guide®
        ▪ SAS® Add-In for Microsoft Office
        ▪ SAS® Web Report Studio (web-based)
        ▪ SAS® BI Dashboard (web-based)
        ▪ SAS® Information Delivery Portal (web-based)

- SAS® Foundation creates the analytic routines that calculate the statistics such as empirical Bayes geometric mean (EBGM), reporting odds ratio (ROR), and so on, used to determine disproportionality that ultimately leads to detecting a signal.

    o Clients:

        ▪ SAS windowing environment

- SAS® Enterprise Case Management is used as a signal management system to track the progress of the investigation of a signal through the triage, validation, evaluation, possible escalation, and disposition stages.

    o Clients:

        ▪ SAS Enterprise Case Management (web-based)

- SAS® Workflow Administrator creates and maintains a workflow diagram of how a signal will be handled through the steps described above.

    o Clients:

        ▪ SAS® Workflow Studio

Other Components:

    o Server Components:

        ▪ SAS Enterprise Case Management safety database
        ▪ Web application server (JBoss, WebLogic, or WebSphere)

    o Clients:

        ▪ SAS® Management Console
        ▪ SAS® Personal Login Manager

Below is a SAS technical architecture diagram showing the various components and the tier they are installed on.
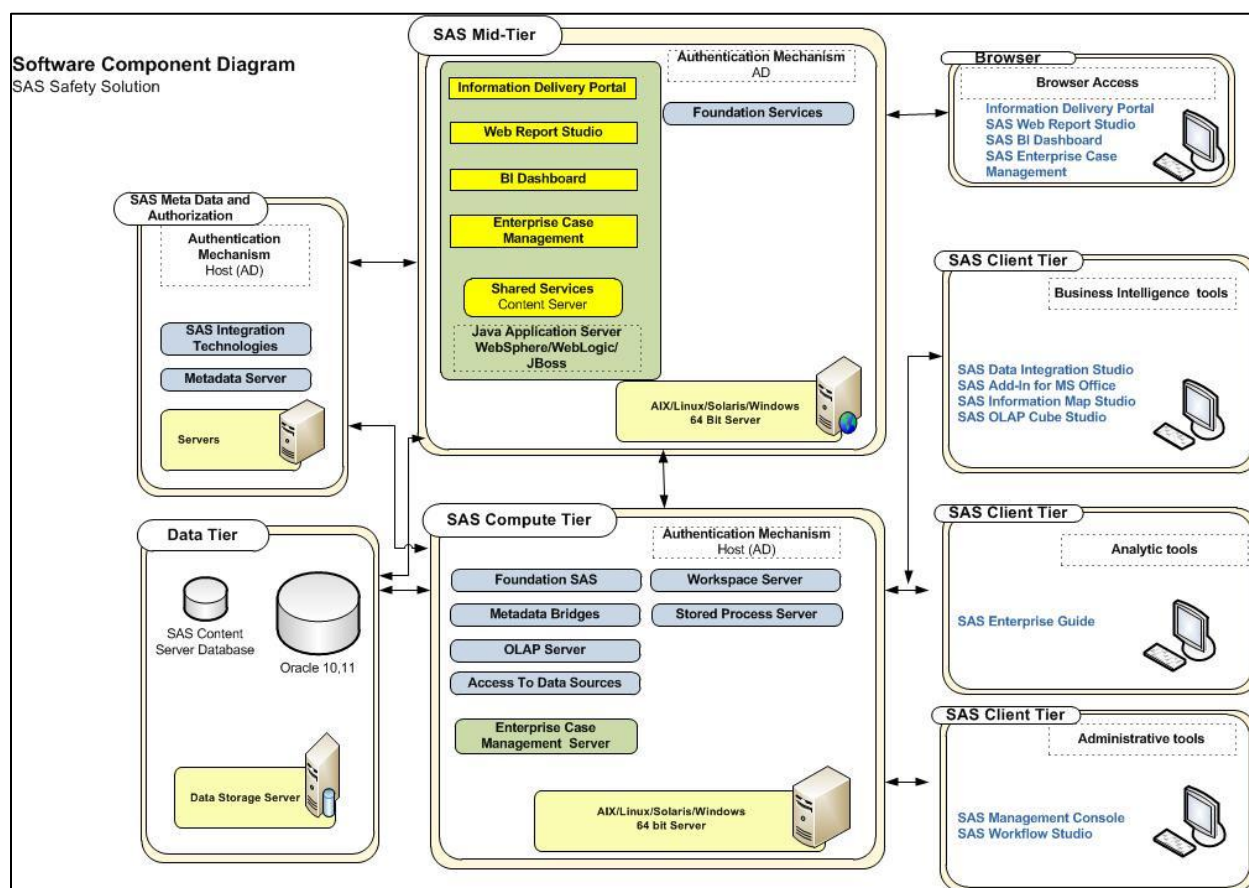
Figure 2. SAS Technical Architecture for Safety Detection

COLLABORATION ACROSS THE SYSTEM

The customer decided to collaborate with SAS for several reasons.

1. SAS brings the appropriate skill sets to the project that will enable the customer to leverage their internal knowledge and produce a white-box solution that the customer ultimately owns and can change as needed in the future.

2. SAS is approaching this project in phases. This allows a project of this size to move forward in an efficient way. In the early phases, the project will address and implement the parts of the system that are most important to the customer and then allow the customer to use and acclimate to the system. This approach works well for both the customer and SAS because as the users become accustomed to the system they are able to make decisions on the next phases in a more educated fashion.

3. The initial work done can be leveraged to build a standard safety detection system across the various divisions of the corporation.

Rather than buying a black-box system, the customer was looking to work with SAS to build a system that would be customized to their specifications. They also wanted a system that could grow with their needs and ultimately they could change and modify. A black-box system just simply was not going to fit their needs. Through several workshops, the team, consisting of both the customer and SAS, worked to identify the business problems and resolutions needed to be housed in this system. By bringing the right people and skill sets to the table, the well-rounded team was able to work together and move forward quickly to design a pilot system.

The skills brought to the project from the customer are:

- Physicians and scientists. This team is the end user of the system. They are tasked with identifying their daily workflow and conveying their needs. They are also needed to test and socialize the product at the end of each phase in an effort to fully define the requirements going into the latter phases. SAS is working closely with this team throughout the life of this project. This team of physicians and scientists are shaping the workflow in SAS Enterprise Case Management. They are validating the metrics used in the safety detection system and are offering insight into how they currently use the data in their decision-making process.

- Statistician. The customer's statistician needs to identify the metrics that are appropriate to be built into the system. He or she identifies the analysis that needs to be done and how to process through and determine the root cause of the signal. The statistician also has access to SAS® Enterprise Miner to develop additional metrics or models if necessary. These metrics and models can be included in the system as needed by the customer.

- Information technology (IT) administrator. The IT administrator has multiple tasks within the project. He or she is the owner of the data and are tasked with preparing the data and making it available to the SAS team. This person has prepared the environment needed for the system.

- Qualified Person Responsible for Pharmacovigilance (QPPV). The QPPV role is to aid is ensuring the system adheres to regulatory requirements.

The SAS team consists of the following:

- Industry domain consultant. This consultant works with the customer to understand their business problems. He or she is a liaison between the customer and the SAS team of consultants to help decipher the needs of the customer and the functionality of the software and how the two can be orchestrated to develop a unified system.

- Systems architect. This consultant helps define the server specifications needed to build the system. This consultant works with the customer to install the software and prepare the environment.

- Data integration consultant. This consultant is tasked with understanding the anomalies of the data and how it will impact the joining of the multiple data sources. This person designs and creates a data model and prepares the data to be consumed by the analytics.

- Analytic consultant. This consultant is tasked with understanding the metrics needed in the system. This person will run, modify, and create code to populate the metrics throughout the system.

- Business intelligence consultant. This consultant will take the analytic results and create and populate the reports needed in the system.

- Enterprise case management consultant. This consultant will understand the workflow used by the physicians and scientists and build that into the software. He or she will also build the customized SAS Enterprise Case Management system that will house the audit trail.

- Project manager. The project manager is tasked with keeping the project on task. He or she will create the project plan, orchestrate the meetings, and facilitate the meeting needed to answer questions and ensure the customer and the consultants all have the information they need to efficiently complete the tasks needed to complete the project.

By bringing these various skills together there is now a complete team needed to successfully implement the safety detection system. But this system has many facades and will likely continue to grow. It is for that reason SAS approached the full project by breaking it into phases. By doing this, the system can be implemented in a variety of ways, addressing each customer's individual needs. The goal is to implement the portion of the system that is most-needed first and build on that over time. Below is the path that was followed for the customer.
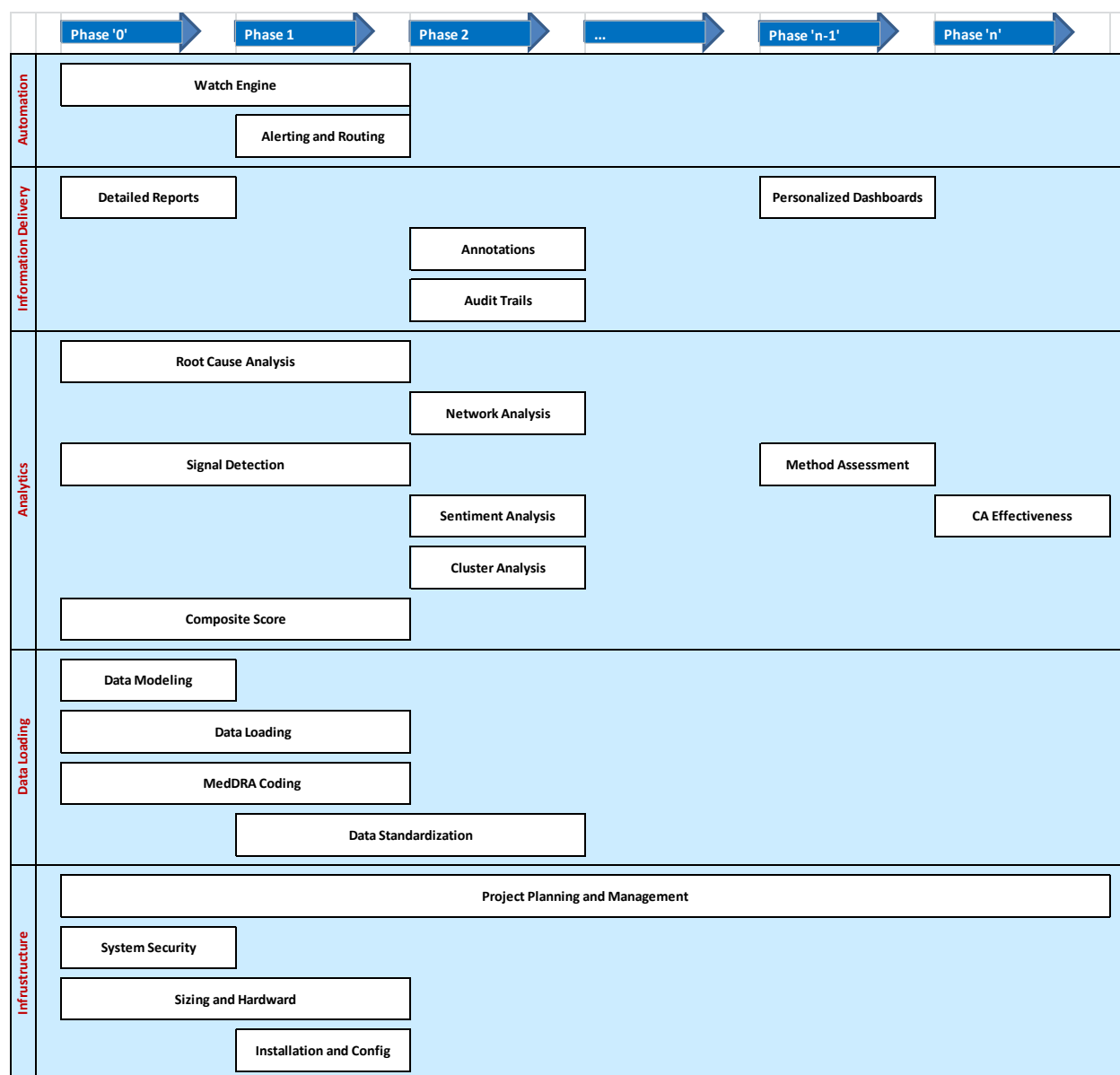
Figure 3. Phases of Implementation

As each phase completes, the customer will begin socializing the system. This allows the end users time to digest the functionality and start to postulate requirements for the next phase of the project. Those new requirements will be recorded by SAS and documented. For this customer, we found that the system developed in the pilot phase allowed the team to move into phase 1, the production system, efficiently. The pilot phase allowed the team to have a starting point to clearly define what was going to be kept moving forward and what was going to be changed or added. We found that many of the custom metrics defined in phase 0 were removed or changed in phase 1 while new metrics were added.

In addition to socializing within the organization, our customer will socialize the system throughout the corporation in an attempt to create a standardized safety detection system across all divisions. This system has the flexibility to be customized for each division. Where there is overlap between divisions, the work done for the other divisions can be leveraged and reutilized. It is this flexibility that is appealing to the customer.

Because of the collaboration on many levels, across the customer and SAS team, across the customer's corporation and across the phases of the project, this project will be successful for all those involved.

DATA PREPARATION

As expected, the data preparation work depends on the format of the data being accessed or available. However, in many cases, data being used for safety detection in the pharmaceutical industry will have many pieces of information in a single field or column. For example, a list of adverse events, drugs taken, or lot numbers involved might be stored in single columns. This makes sense for the data storage model, as there could be a large number of any of these for any reported event and it is not feasible to store each in a separate column.

However, for detecting that there might be an issue with a product, the analysis should be done one drug, one event, and one lot at a time. So, the stored data needs to be transformed into a format that can more easily be analyzed. The easiest way to analyze this is to create a single row in a table for each combination of drug, adverse event, and lot number. For example, if a single event report includes the primary drug, two other drugs being taken along with it, five separate adverse events, and two lot numbers for those drugs, the single report records becomes thirty records in the new table. From there, the disproportionality can be determined for each of the combinations.

ANALYTICS

There are several analytic metrics used for signal detection, but for this customer we focused on the most widely used forms of analytics in signal detection: disproportionality analysis. Disproportionality analysis is a way of identifying if a population of consumers using a particular product is experiencing an adverse reaction at a rate higher than expected—in other words, calculating the ratio of observed to expected.

The post-marketing data is unique in that it contains records that have been reported only after experiencing an adverse reaction. These reports can be submitted by a health care professional or the consumer. Because of this type of data collection methodology, there is no control group contained in the data and there is also no way to estimate the volume of consumers using the product. It is because of these limitations that there are several ways to estimate the expected number of consumers experiencing an adverse event after consuming a product. The four methods that we used were:

- proportional reporting ratios (PRR)

- reporting odds ratios (ROR)

- multi-item gamma Poisson shrinker (MGPS)

- Bayesian confidence propagation neural network (BCPNN).

All four of these calculations are based on a 2x2 contingency table as shown below.

	Event of Interest (Yes)	Event of Interest (No)	Total Events
Product of Interest (Yes)	n_{00}	n_{01}	$n_{00}+n_{01}$
Product of Interest (No)	n_{10}	n_{11}	$n_{10}+n_{11}$
Total Products	$n_{00}+n_{10}$	$n_{01}+n_{11}$	$N=n_{00}+n_{01}+n_{10}+n_{11}$

Table 1. 2x2 Contingency Table

The PRR methodology is based on the more widely known calculation for relative risk (Evans et al.). It uses the other products in the database as a control. The metric calculates the ratio of the frequency of calls associated with the product of interest and event of interest combination and compares that to those frequency counts for all other products in the data. PRR is computed as follows:

$$PRR = \frac{n_{00}/(n_{00} + n_{01})}{n_{10}/(n_{10} + n_{11})}$$

The ROR methodology is very similar to that of PRR except it attempts to control for certain types of under-reporting (García et al. 2012). The ROR is computed as follows:

$$ROR = \frac{n_{00}n_{11}}{n_{10}n_{01}}$$

While the two Bayesian calculations are more complicated, they both attempt to determine the relationship between the reporting of events associated with a product that is observed and the expected frequency of reporting for product (i) and the adverse event (j) (García et al. 2012). The expected counts E_{ij} are computed as follows:

$$E_{ij} = \frac{n_{i.}n_{.j}}{n_{..}}$$

where

$$n_{i.} = \sum_{j=1}^{J} n_{ij}, \; n_{.j} = \sum_{1=1}^{I} n_{ij}, \; n_{i.} = \sum_{i=1}^{I} \sum_{j=1}^{J} n_{ij}$$

are the total counts corresponding to drug *i*, the total counts corresponding to adverse event *j*, and the total number of reports, respectively.

Now the measure of disproportionality can be calculated as follows:

$$IC_{ij} = log_2 \left(\frac{n_{ij}}{E_{ij}} \right)$$

For BCPNN, the thresholds rely on the IC calculation above and the 95% confidence interval as criteria for its signal detection methodology (Bates et al.).

The MGPS methodology uses the Poisson distribution to model the observed count for each product-event combination (DuMouchel 1999). The model is:

$$n_{ij} \sim \text{Poisson}(\mu_{ij})$$

$$\mu_{ij} = \lambda_{ij} E_{ij}$$

$$\lambda_{ij} \sim \text{Mixture Gamma}$$

The posterior distribution of λ_{ij} is obtained and the Empirical Bayes geometric mean, EBGM, is defined (García et al. 2012).

In addition to the standard metrics, custom metrics are embedded into the system. These metrics, defined by the client, are as follows:

- Frequency of death associated with a product and event combination in the most recent period of data

- Frequency of serious events, using a variable defined in the customer's data, associated with a product and event combination in the most recent period of data

- Frequency of serious event divided by the frequency of non-serious events in the most recent period of data.

With the calculation in place, the client identified the appropriate thresholds to implement into the system. While the calculations for ROR and BCPNN are being produced within the system, they will not be used in the first phase of the system as indicators of a signal. As the first phase of the project, the client decided it was best to limit the number of metrics visible to the end user. This allowed them to focus on the metrics that they felt were most prudent to the investigation process. The following are the thresholds implemented in the system.

Metric	Thresholds for RED	Thresholds for YELLOW
EBGM	EBGM > 2 and EBGM05 >1	EBGM > 1.8 and EBGM05 > 0.75
PRR	PRR > 2 and N > 3 and PRR_CHISQ > 4	PRR > 1.8 and N > 2 and PRR_CHISQ > 3
Death Frequency	Death Frequency => 1	
Serious Frequency	Serious Frequency => 2	Serious frequency =>1
Proportion of Serious	Chi-square significant (p<0.05) compared to background proportion and PRR_CHISQ > 3.84	3< PRR_CHISQ <= 3.84

Table 2. Signal Detection Thresholds

BUSINESS INTELLIGENCE REPORTS

Once all of the data has been processed, the analytic routines have been performed, and the disproportionalities have been detected programmatically, the next step is to present this information to the users who need to investigate these potential issues.

The software chosen to display the signal information to users is SAS Web Report Studio. The report created contains a series of tables displaying the list of adverse events by various combinations such as product/event, product/lot number/event, and so on. This data is presented at different levels, as the analytics are generated at multiple levels. These levels are, at a minimum, the adverse event and the system organ class, but there could be others.

The tables are then sorted in descending order by what is selected as the most important statistic. This is done to highlight the most serious of the signals based on that statistic. Also, conditional highlighting is added to further emphasize the more serious events.

If a user wants to get more information about a particular signal, they can click on the row they are interested in to get summarized data about the selected product/event combination. This table will show how many reports have been received by:

- Gender
- Age group
- Lot number.

These numbers are broken down by geography and by seriousness of the event.

If more detailed information is desired, the user can click on the single row table at the top of the report to link to the next screen. This link leads to a simple table displaying all of the columns from the call center data for the chosen product/event.

The following displays depict a series of linked reports. Display 1 depicts the top ten products with the most events in the most recent time period, sorted by descending PRR. Display 2 depicts a first-level summary table for a selected product. Display 3 depicts the call-level detail behind the data that is summarized in the previous two reports.

Top 10 Products by Adverse Events – Most Recent Period by Descending PRR

Product	Event	Number of Adverse Events	RR	PRR	ROR	BCPNN	EBGM	EB05	EB95	Composite Score (EM)	Priority Score	Not Prev Report	Rare	Serious Event	Off-label instances	AE Volume	Degree of Disproportionality	Health Care Professional	Consumer
PRODUCT 135	PT: THROAT IRRITATION	1	3.20	12.20	22.81	3.30	6.86	5.17	8.97	0.58	6	0	0	5	0	0	0	0	1
PRODUCT 60	PT: MULTIPLE DRUG OVERDOSE INTENTIONAL	1	0.14	6.93	8.12	1.22	0.27	0.03	1.30	0.42	14	3	3	5	0	0	0	3	0
PRODUCT 138	PT: TACHYCARDIA	2	0.17	5.97	8.45	1.17	0.26	0.03	1.28	0.40	12	3	3	5	0	0	0	0	1
PRODUCT 135	SOC: RESPIRATORY, THORACIC AND MEDIASTINAL DISORDERS	1	11.90	4.42	5.15	1.97	3.94	3.11	4.94	0.57	6	0	0	5	0	0	0	0	1
PRODUCT 143	PT: SOMNOLENCE	2	29.94	2.35	2.88	1.17	2.12	1.73	2.58	0.23	9	0	0	5	0	0	0	3	1
PRODUCT 60	SOC: INJURY, POISONING AND PROCEDURAL COMPLICATIONS	1	54.69	2.24	2.29	0.79	1.70	1.43	2.00	0.23	14	0	0	5	5	0	0	3	1
PRODUCT 138	SOC: CARDIAC DISORDERS	2	2.85	2.12	2.26	0.95	1.67	0.98	2.86	0.35	6	0	0	5	0	0	0	0	1
PRODUCT 49	PT: SOMNOLENCE	2	15.43	1.99	2.13	0.94	1.70	1.24	2.27	0.19	6	0	0	5	0	0	0	0	1
PRODUCT 143	SOC: NERVOUS SYSTEM DISORDERS	2	109.89	1.65	1.80	0.64	1.57	1.38	1.77	0.15	14	0	0	5	0	0	5	3	1
PRODUCT 131	PT: FAECES DISCOLOURED	1	3.20	1.56	1.73	0.57	0.86	0.38	1.73	0.15	6	0	0	0	0	0	5	0	1
PRODUCT 49	SOC: NERVOUS SYSTEM DISORDERS	2	57.13	1.50	1.55	0.52	1.41	1.17	1.67	0.13	6	0	0	5	0	0	0	0	1
PRODUCT 68	PT: VOMITING	2	6.57	1.38	1.40	0.43	0.99	0.55	1.67	0.11	6	0	0	0	0	0	5	0	1
PRODUCT 131	SOC: GASTROINTESTINAL DISORDERS	1	234.61	0.98	0.98	-0.02	0.99	0.89	1.10	0.03	14	0	0	5	0	0	5	3	1
PRODUCT 68	SOC: GASTROINTESTINAL DISORDERS	2	58.82	0.90	0.90	-0.12	0.92	0.74	1.14	0.02	6	0	0	5	0	0	0	0	1

Display 1. System Main Report

Display 2. System First Summary Table

Display 3. System Final View of the Call Level Data

ENTERPRISE CASE MANAGEMENT

The project presents a unique opportunity because the client has a vision of what they want but needs the technical expertise that SAS can offer to bring it to fruition.   Together, we are translating their requirements into a system that will address the multiple stages associated with the customer's workflow and document the results within each stage.

10

At a simplistic view of the safety detection process, there are three stages that need to be completed: detection, triage, and validation. The signal management process is documented and audited throughout these stages. This audit trail is captured using SAS Enterprise Case Management.

The detection stage is the first stage of the process. As discussed above in this paper, there is much that needs to be completed before the system can process new information and detect a possible risk with a product, otherwise known as a signal (European Medicines Agency 2011). It is during this preparation that SAS Data Integration Studio is leveraged to integrate and prepare the data. Base SAS® and SAS/STAT® software are used to calculate the industry-recognized statistics used to determine if a signal should be generated. In addition, business rules are built into the system using Base SAS to automatically generate metrics that the analyst would otherwise do manually. It is because of the output of this preparation the system is able to detect and identify the signals that the investigator will begin further tracking. All signals that have been identified during the detection stage will be tracked in SAS Enterprise Case Management and will move forward into the next stage.

The second stage, triage, is when the analyst will review and prioritize all detected signals and determine the appropriate workflow transitions of that signal. The analyst will use the information housed in the system to determine if a signal will be:

- Moved through the process to the validation phase because further investigation is necessary

- Tabled but continually investigated on a periodic basis

- Closed because no further investigation is warranted.

Once all signals have moved to the appropriate category, those that will continue through to the third phase, validation, must be prioritized. This will ensure those most likely to have the greatest impact on the general public are investigated in the most expeditious fashion.

The validation stage is where the majority of the investigator's work will occur. During this stage, the investigator will review the signal for several factors. Those factors include but are not limited to:

- Is the reaction new or never seen before?

- What other concomitant products was the consumer taking during the same time period?

- Is it biologically plausible to experience this event after consuming the product?

During this validation stage, the signal might be escalated if there is a possible significant impact to public health. Appropriate measures are taken to minimize the negative impact on the public.

SAS Enterprise Case Management is an integral part of the signal management from stage to stage. It allows the system to assign the signals to the appropriate investigator and contacts the investigator when a new signal is available for them to investigate. SAS Enterprise Case Management also allows the assigned investigator to document results at each stage. Finally, it allows the investigator to easily search and identify previously documented investigations.

FUTURE DIRECTIONS

Incorporating text analytics is a possible future direction to consider including in the signal detection system. There is a large volume of unstructured textual data stored in call center notes, social media conversations, blog text, medical publications, and health practitioners' notes that safety experts do not currently analyze. These data sources might contain additional hidden insight about products under surveillance. Applying advanced text analytics and sentiment analysis to these data sources could provide a valuable signal about product safety issues in the future (Brinsfield 2009).

Applying data mining techniques to signal detection is another plausible future direction. Physicians, scientists, or statisticians might generate new hypotheses using predictive analytics modeling techniques such as those available within SAS Enterprise Miner. Some of these techniques include:

- Exploratory and unsupervised techniques, such as clustering, self-organizing map (SOM) or Kohonen map, association analysis, and path analysis. These might detect related groups of products or events for further study.

- Regression- or classification-based techniques. These can reveal relationships between dependent variables and independent variables.

- Decision tree techniques. These can be used for obtaining interpretable English rules or logic statements that map attributes of a product-adverse event combination to probabilities of whether the product-adverse event combination contains a safety signal (Brinsfield 2009).

AN ARCHITECTURE FOR SCALABILITY

The signal detection system is well equipped to accommodate exponential growth in data or computing workloads. The SAS architecture is designed to efficiently access large amounts of data, while simultaneously providing timely analytical intelligence to a large number of users. Because the platform uses an n-tier architecture that enables functionality to be distributed across computer resources, each type of work is performed by the SAS resources that are most suitable for the job.

In addition, the system can be scaled by incorporating SAS® High-Performance Analytics as the data and analytical computing needs driven by the business requirements dictate. Customers can enhance their SAS analytical lifecycle with high-performance-enabled data exploration, model development, or model deployment capabilities, as seen in Figure 4. Using this approach, customers employing SAS for signal detection can adapt their architecture to solve big data problems with big analytics. They can explore large quantities of granular data, perform more sophisticated analyses, spread the statistical processing across a grid of available SAS servers, and reduce their time to reviewing signal metrics, all of which will enable faster, more confident decision-making when evaluating potential safety signals.
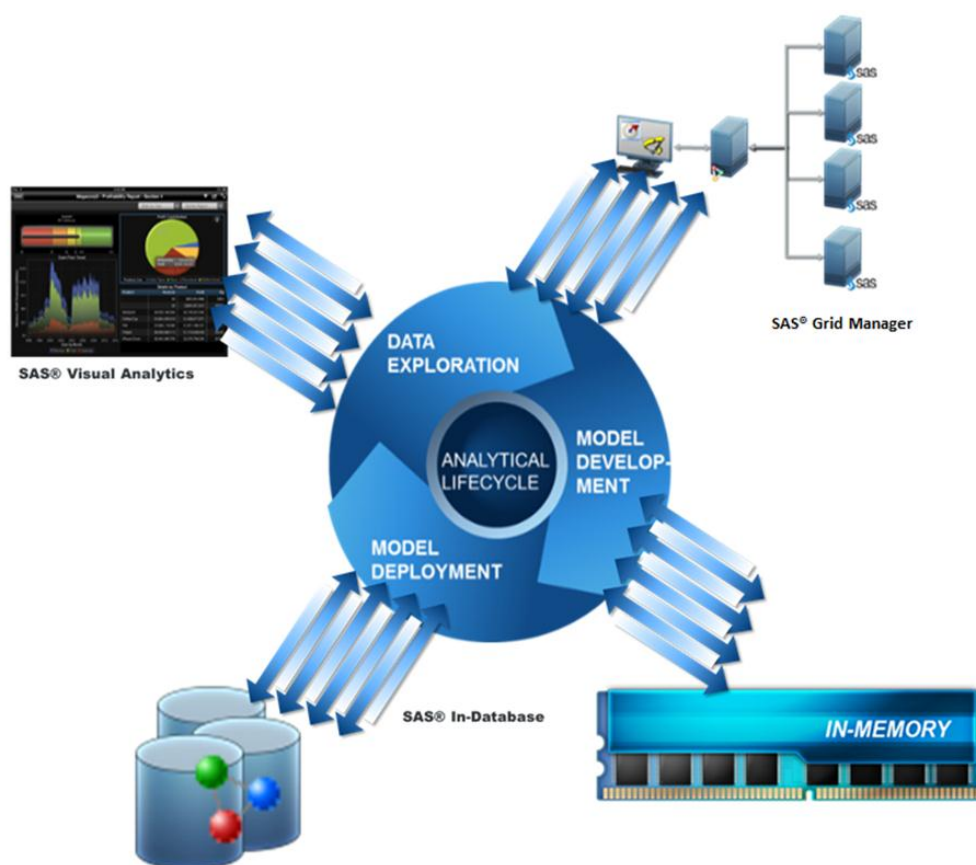

Figure 4. SAS High-Performance Analytics

Because of its architecture, the SAS platform can be scaled to meet the increasing demands of an organization's post-market surveillance over time. For example, as the number of medical products (drugs or devices) increases, either due to new in-house development or acquisition, the number of product and event combinations increases quickly, especially when additional groupings by gender, age group, or lot number are added. Table 3 illustrates multiple scenarios where the number of total metrics requiring analysis grows exponentially.

The total number of times a signal metric must be executed is defined as:

$$Total\ signal\ metrics\ = Products\ \times Events\ per\ Product\ \times Age\ groups\ \times Genders\ \times Metrics\ per\ Product$$

The total execution time then becomes:

$$Total\ processing\ time\ = Total\ signal\ metrics\ \times Average\ execution\ time\ per\ metric$$

Number of Products Analyzed	Average Number of Possible Events per Product	Age Groups	Genders	Number of Signal Metrics per Product-Event Combination	Average Execution Time per Metric (in seconds)	Number of times a signal metric is executed	Total Execution Seconds	Total Execution Hours	Total Execution Days
15	50	7	3	5	0.1	78,750	7,875.0	2.2	0.1
20	100	7	3	5	0.1	210,000	21,000.0	5.7	0.2
100	250	7	3	5	0.1	2,625,000	262,500.0	71.7	3.0
1000	500	7	3	5	0.1	52,500,000	5,250,000.0	1,434.4	60.8
3000	750	7	3	5	0.1	236,250,000	23,625,000.0	6,454.9	273.4

Table 3. Analytical Processing Increases Exponentially as Drug-Event Combinations Increase

Even assuming a sub-second response time for each metric, and that each metric is executed sequentially, we see that the execution time to complete the analyses can quickly exceed the required completion time or batch window size of organizations using SAS in a multi-core processing architecture. Adding memory, processing cores, and multi-threaded processing can help reduce the processing completion time. The advantage of using SAS in this architecture is that SAS High-Performance Analytics can bring order-of-magnitude reductions in execution times needed to post-market surveillance in these scenarios.

CONCLUSION

The signal detection components described in this paper provide the customer with a flexible, scalable solution for integrating and preparing disparate safety data sources for analysis. It also automates their current business processes while providing a repeatable, auditable process to manage potential safety signals as they move through the phases of detection, triage, and validation. Today, the SAS system architecture plays a key role in enabling sophisticated statistical analyses on large volumes of product-adverse event combinations. SAS High-Performance Analytics can be employed as the system's computing requirements grow due to new product developments and/or corporate acquisitions. In addition, the wide variety of text analytics and predictive analytics techniques in SAS can be incorporated into the system in future phases. By leveraging SAS as a strategic piece of its safety detection platform, the customer is positioned to not only detect signals earlier in their products' post-market lifecycles, but with better accuracy and analytical rigor than is possible with other approaches.

REFERENCES

Uppsala Monitoring Centre. 2011. "Glossary of Terms Used in Pharmacovigilance." Available at http://who-umc.org/Graphics/24729.pdf. Accessed on January 21, 2013.

iVillage Inc. 2012. "Biggest Drug Recalls of the Past Decade." Available at http://www.ivillage.com/biggest-drug-recalls-past-decade/4-b-156084. Accessed on January 21, 2013.

European Medecines Agency. 2011. "Guide on the Interpretation of Spontaneous Case Reports of Suspected Adverse Reactions to Medicines." Available at http://www.ema.europa.eu/docs/en_GB/document_library/Report/2011/07/WC500109582.pdf. Accessed on January 21, 2013.

García, Viviana, Peng He, Xuefei Jian, Andrea Knezevic, Ying Lu, Mahmoud Shehadeh, and Ranye Sun. 2012. Problem 6: Signal Detection for Drug Safety. In *Eighteenth Mathematical and Statistical Modeling Workshop for*

Graduate Students. Proceedings of the Industrial Mathematical and Statistical Modeling Workshop for Graduate Students, Raleigh, NC, 2012. Available at http://www.samsi.info/sites/default/files/IMSM12_report_0.pdf.

Brinsfield, Eric. 2009. "Signal Detection for Drug Safety: A Biomedical Informatics Challenge. *DIA Global Forum* 9, no. 6:29–30. Available at http://www.nxtbook.com/nxtbooks/dia/globalforum_200912/index.php?startid=27#/29/OnePage.

Evans, S.J. et al. Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. Pharmacoepidemiol. Drug Saf 2001; 10: 483–486

DuMouchel W. Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. Am. Stat 1999; 53:177–190

Bate A. *et al.* A Bayesian neural network method for adverse drug reaction signal generation. Eur. J. Clin. Pharmacol. 1998; 54: 315-321

ACKNOWLEDGMENTS

CONTACT INFORMATION

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

Carrie Boorse
43 Firebush Rd.
Levittown, PA 19056
919-531-9868
carrie.boorse@sas.com

Stuart Levine
SAS Professional Services BI Practice
111 Rockville Pike, Suite 1000
Rockville, MD 20850
301-838-0453, ext. 51163
stuart.levine@sas.com

Kathy Schaan
SAS Health Analytics Practice
SAS Campus Drive
Cary, NC 27513
919-531-9737
Kathy.schaan@sas.com