Automating Pharmaceutical Safety Surveillance process

Pankaj Bhardwaj, Tata Consultancy Services Ltd, New Delhi, India
Chandramouli Raghuram, Tata Consultancy Services Ltd, Edison, NJ

ABSTRACT

Pharmaceutical companies invest huge amount of effort and cost in launching a drug into the market, keeping the two aspects in mind i.e. safety and efficacy. Post launching the drugs into the market, the organisations then need to monitor the adverse events from these drugs, and need to take the action accordingly. In this white paper few approaches are discussed for automating the post market safety surveillance processes in a cost effective manner.

The pharmaceutical companies collect the adverse events data from various heterogeneous sources, and this collected data need to be analysed for the safety surveillance. Generally in post market safety surveillance processes, each Drug-Event case is processed record by record which is causing the exponentially rise of the records and finally leading to the high computational complexity and Analytics performances issues in the system. This paper will also focus on handling these Performances issues by bringing Hadoop environment into the Solution Implementation. The technical SAS solution to be discussed in this white paper will include components for Data Extraction and Transformation, Analysis, Reporting and Automated Processes for Signals detection & Investigation.

INTRODUCTION

Recent changes in PV regulations and new legal requirements have increased the need to expand and implement a comprehensive signal detection and safety surveillance process for all pharmaceutical companies. Some of the primary reasons for establishing such a solution are:

1. The regulatory authorities (FDA, EMEA and MHRA) and drug companies are under increasing scrutiny and criticism by the public, the media, and government

   - The final GVP guidance further strengthened the need for such tools by providing details of signals management:

     ▪ GVP Module IX : The validation and assessment of signals, and any resulting actions should be tracked systematically

     ▪ GVP module VII: A tabular summary of safety signals will be included as an appendix to the PSUR for the period of the PSUR which can span between 6 months to 23 years. For all confirmed and non-confirmed closed signals evaluation should be documented to describe the basis upon which the signal was either rejected or considered to be a risk

2. Timely assessment of internal and external safety signals is vital to ensure effective risk management and ability to respond to queries from authorities and public.

3. Audits and inspections findings

KEY CHALLENGES

To monitor the safety of drugs, safety reviewers traditionally relied on their medical expertise and visual observation. Although this process is still critical, reviewers cannot absorb all of the data or see all subtle emerging trends. To optimize the use of valuable medical and epidemiological resources, companies and regulators need ways to narrow the possibilities or find growing problems sooner.

▪ **Computation of Effect Magnitude** - Existence of multiple diseases and numerous Drug combinations increases the probability of Drug-Protein and Drug-Drug interactions. Geography,

Food Habits, Genetic Factors, Lifestyle, Environment, Health Index etc. are few of the many other impacting factors.

- **Wide Data Sources & Accessibility** - Demand of increased quantity of data requires access to data sources available at various Locations, which have their independent Technological and Data Cleaning issues.

- **Incomplete or Unreliable Data** - Data available may not provide all necessary attributes. Frauds in Clinical enrolments often leads to False data capture and distinction of adverse Events occurring due to counterfeiting of Drugs are the another challenge.

- **No Gold Standard for comparison** - Bigger the data, bigger are data mining challenges and so is the analysis. The Traditional Methods are robust enough to handle Upto 1k GB of data, but handling above it has its own Technological and Statistical Limitations. In post market safety surveillance processes, each Drug-Event combination is processed record by record, which is causing the exponential rise of the records/data and finally leading to the high computational complexity and analytical performance issues in the system.

- **Reactive to Predictive Approach in Signal Detection** - The technical limitations and Traditional statistical methods are in vain. Industry is still in the process of developing and experimenting the better Algorithms for the Signal detection

Reviewers need an automated filtering mechanism that improves efficiency by helping to detect growing trends sooner and decreasing time wasted on false signals.

As an alternative, automated signal detection provides a proactive, yet cost-effective approach for any risk management plan. Given the number of products that regulators and safety reviewers must monitor, automation is critical for the safety of patients. Based on this fact, it is not the value of automated signal detection that should be evaluated, but the quality, effectiveness, and usability of a particular automated system.

QUALIFYING SIGNALS

Because their workload is already heavy, reviewers aim to minimize the number of less-qualified signals that they must pursue. They look for higher quality signals that lead them to real issues faster. Part of the automation, therefore, must find a signal and perform as much refinement and qualification as possible prior to creating an alert. At a minimum, the signals should be prioritized, so nothing is omitted incorrectly.

In this paper, we propose an intelligent safety surveillance system, based on SAS suite of products; to generate different data combinations; data mining statistics to discover trends and search for prospective safety signals. This paper will also focus on handling some of the inherent performance issues by bringing Hadoop / BIG DATA environment into the Solution Implementation. The technical SAS solution to be discussed in this white paper will include components for Data Extraction and Transformation, Analysis, Reporting and Automated Process for Signals detection & Investigation.

THE SOLUTION

The Adverse Event Data coming from various sources need to me merged with other Data sources like Medical Dictionaries or Designated Medical Files and then then after deriving the analysis variables, Data is finally loaded into the Database. The Data processing via this solution is completely automated, and these jobs/programs are scheduled to run either on a daily or weekly basis, depending on the needs of the business and organization. These jobs/programs can be done either through Base SAS or SAS DI Studio.

Considering the Data criticality, the organisations create one record for each Drug-Event combination and process very carefully. In the Data analysis phase, if some threshold values are crossed then signals are generated which are then sent to Case Management tools for processing. These case management tools can be SAS ECM or a custom solution based on MS .NET – SharePoint, built for handling the entire case

management process, including workflows, audit trail, triaging and escalation.. The Scientists and Data Analysts process those signal and this entire process needs to comply governing body regulations.

Apart from the signal management, at every stage of this data journey various sets of reports can be created. These reports can be created either in Base SAS or in a SAS BI environment consisting of SAS Information Map Studio and SAS Web Report Studio. For Adhoc Reporting OLAP Cubes can be used to derive important information using various dimensions.

Tools like SAS E-Miner can also be leveraged to derive some important results which can play key contribution to the surveillance.
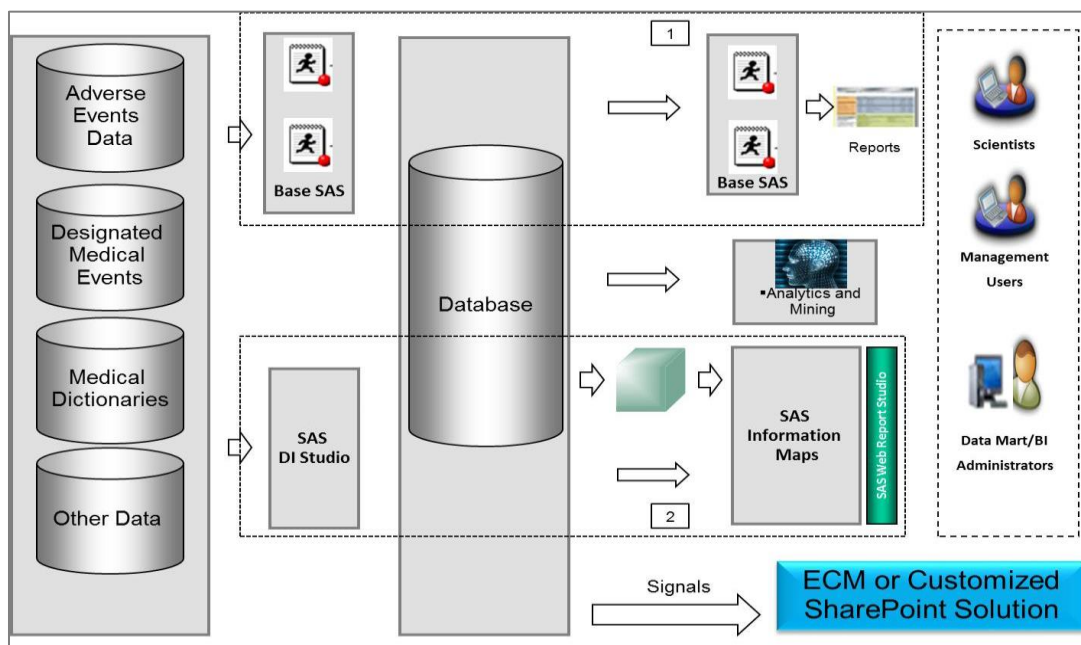

Figure 1 - SAS based solution

COMBINING THE POWER OF "BIG DATA"

Due to the large number of reported drugs and adverse effects, Data Mining had become a challenging task in terms of computational complexity. For example, in the ADE reports collect from FDA, the number of reported drugs and adverse effects are 237,579 and 14,401 respectively. Theoretically, we have 3,421,375,179 number of 1-drug and 1- effect combinations. This number will increase exponentially if the combinations of arbitrary number of drugs and effects are considered.

For addressing this issue the Data can be resided in the Hadoop Distributed File System and SAS Access to Hadoop can be leveraged for performing the advance Analytics.
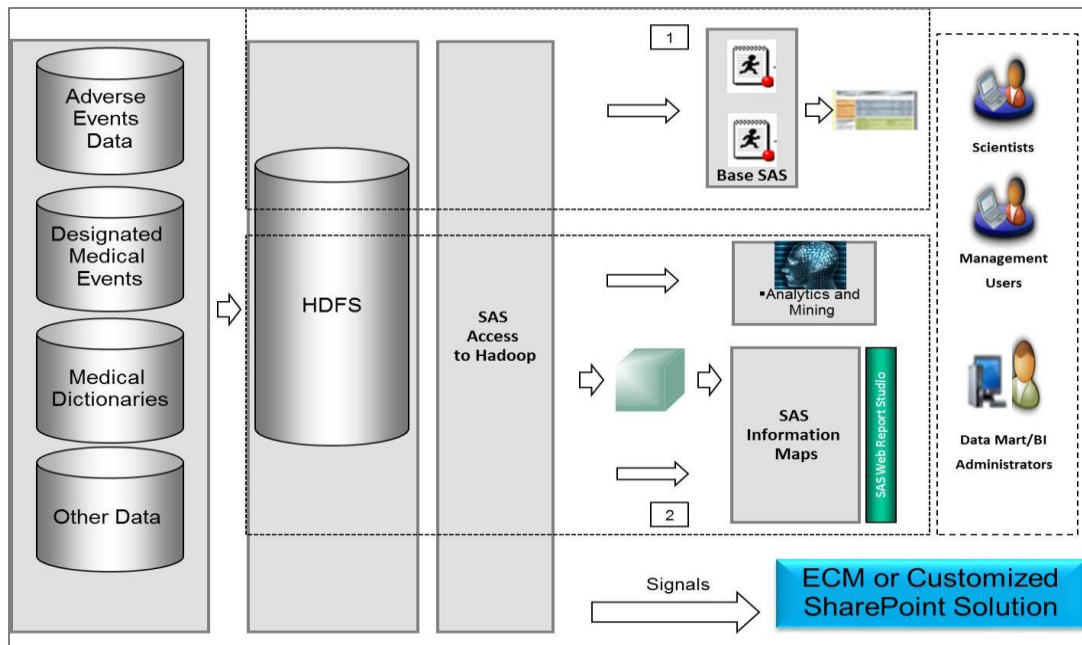
Figure 2 - Hadoop - SAS based solution

DATA FLOW

The Data from various source systems will be brought into the HDFS staging area and after Data Cleansing; the Data will be kept into Hive warehouse. SAS Access to Hadoop will be used to pull the data from the warehouse into SAS Environment, and this data will be ready for performing any kind of Analysis.

The Analysis Datasets will be kept in the Reporting Data mart. The Semantic Layer in SAS Environment will help end users to access the data autonomously using common business terms. Based on the reporting requirements, all the essential data elements will be created with business terms in this layer. Business meaningful abstract layer created will be used as the data source for creating reports.

OLAP cubes can also be built in the semantic layer to aid OLAP analysis. Key reporting modules will be developed from semantic layer. The Reporting Layer will also provide users to create/view Ad-hoc reports, Standard reports and detailed reports through web interface or/and desktop client interface.

The data from the Reporting Data Marts will be fed to SAS Web Report Studio and also to the Case Management solution i.e. either ECM or custom MS .Net SharePoint solutions. The Case Management solution will be used for managing investigation workflows, attaching documentation and escalating cases to management.

ANALYTICAL METHODS

The system offers pre-built analytical modules and screening algorithms. These modules and algorithms can systematically and independently detect potential adverse drug event signals out of several millions of drug-event combination pairs generated from a spontaneous adverse event reporting database or proprietary safety and clinical trial safety databases. Depending on the entry point your company has implemented, some of the routines can also be used to investigate associations involving multiple adverse events, as well as adverse events that may be drug-induced or drug-drug interactions induced.

Analytical methods for screening drug-event associations and for disproportional reporting include industry standard Signal detection routines, such as:

- Proportional Reporting Ratio (PRR) method

- Reporting Odds Ratio (ROR) method

- Multi-Gamma Poisson Shrinker - Empirical Bayes Geometric Mean (MGPS-EBGM)

- Bayesian Confidence Propagation Neural Network – information component score method

In the case of spontaneous reporting systems, these algorithms produce statistical scores that quantify the degree or frequency with which a drug occurs with a particular event relative to the expected frequency based on independence model.

In addition to these methods, the system also employs a statistical detection outlier measure, known as an adjusted residual score, for detecting drug-event pairs with unusually large values that may appear as potential safety signals. The utility of this algorithm is the ability to identify outliers that help point reviewers to the source and the underlying patient reporting population for further investigation.

The signal detection algorithms and process framework implemented in the SAS solution platform also extend the utility of current and industry standard safety algorithms to move beyond signal detection to signal prediction. With the advanced entry point platform, safety professionals are presented with opportunities to build a variety of predictive and risk models for testing different hypotheses involving one or more drugs with specific adverse events and their outcomes. Such models can be deployed to proactively monitor trend and safety profile of the drug product while the drug is still being marketed, as opposed to reacting to unexpected or unanticipated safety issues that may surface ("reacting to the past").

PRE-DEFINED REPORTS

Display of disproportionalities exceeding defined thresholds, along with drillable options to pre-defined tables is one of the key features that this kind of a solution can provide and support, ensuring complete insight of the data and its underlying patterns for the reviewers. Few examples of pre-defined reports that can be built via such a solution – the alerts report and adverse event analysis report are shown below.

Reviewer	Durg	Level/Adverse Event	Number Of Events	EBGM	EBGM Signal Fired	EBGM Lower 5% Confidence Level	PRR	PRR Signal Fired	FRR Current Month Compared To Previous Month
ABC	Paracetamol	HLT:Rashes, Eruptions and Exanthems NEC	2570	1.20		1.17	1.24		0.29
ABC	Paracetamol	HLT: Neurological Signs and Symptoms NEC	2352	1.13		1.09	1.15		0.08
ABC	Paracetamol	PT: Rash	2262	1.19		1.15	1.22		0.46
ABC	Paracetamol	PT: Dizziness	2195	1.10		1.06	1.11		0.10

Workflow

	* Summary	* Investigation	Escalation	Signal History	

Signal Metrics
Add incident to signal

Incident ID △	Description	RR	PRR	EBGM	Serious Event	Date Created
✖ 201359623232924_INCI59623483			8.6557573573	6.135653	0.0	06/10/2013

Figure 3 - Alerts Report
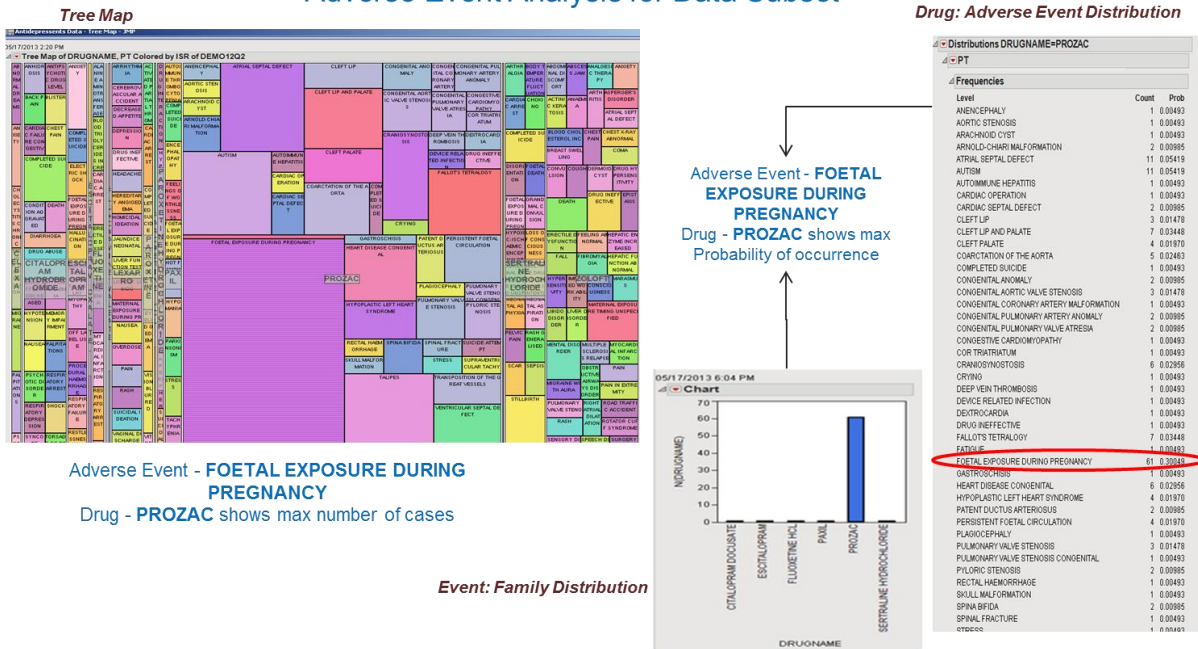
Figure 4 - Adverse Event Analysis

WORKFLOW FOR GUIDED ANALYSIS

The solution provides several workflow pathways for reviewing, exploring and evaluating signals. Figure 3 shows the workflow of all the alerts that are generated after the analytics data is generated by applying various signal detection algorithms, and that are provided to the reviewer for further processing.
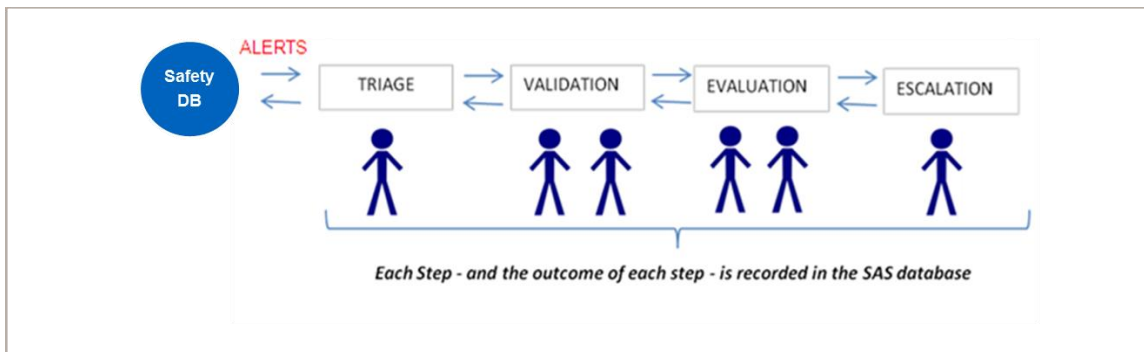

Figure 5 – Signal Workflow

COMPARISON OF OLD VS NEW PROCESS

With the availability of such an automated solution, the Safety database which is the integral piece of the solution becomes the one stop-shop for all queries and data, related to alerts of all the drugs in question within an organization. A visual comparison of the old manual versus the automated process before deploying such a solution is provided in figure 5.
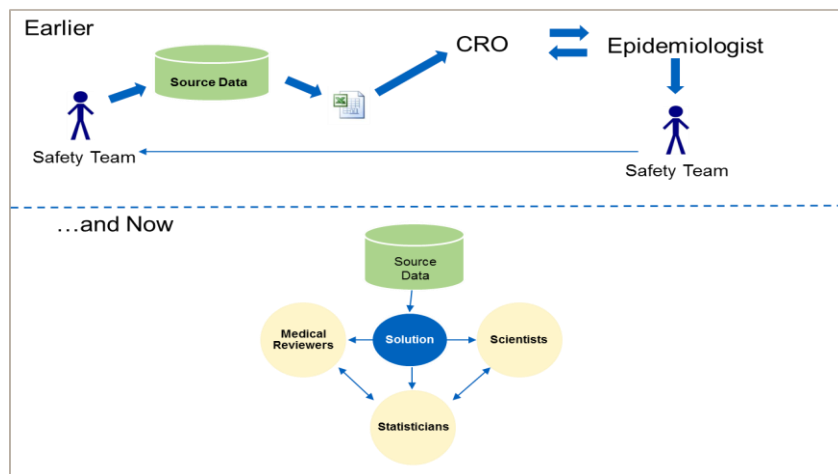
Figure 6 - Earlier vs New Process

FUTURE DIRECTIONS

The pharmaceutical industry is currently facing issues with the traditional statistical methods as they have some limitations and also they are not robust enough to handle large volumes of data greater than 1k GB. Considering this, the industry is trying to get away from a reactive mode to a predictive mode, and is in the process of developing and experimenting better algorithms for signal detection.

The overall solution discussed in this paper could be scaled up for performing some of the next generation methods like

- Temporal Analysis
- Bagging
- Clustering of data
- Heterogeneous ensemble
- Sequence pattern etc.

In addition to this, the solution discussed could incorporate "Text Analytics" leveraging the Hadoop/BIG Data components, considering the large volume of unstructured textual data stored in call center notes, social media, blogs, and medical publications etc. which are currently not analyzed. These additional sources of data could contain some hidden insights and could provide valuable signals about product safety issues in future.

CONCLUSION

We observed various approaches for implementing the surveillance solution. Because of the huge Data volume there are challenges in performing the analysis. Thus leveraging the SAS Access to Hadoop Capabilities can bring a lot of value in terms of performance and make the Safety Surveillance; process very effective.

REFERENCES

Lesley Wise, John Parkinson, June Raine and Alasdair Breckenridge. New approaches to drug safety: a pharmacovigilance tool kit. Nat Rev Drug Discov. 2009; 8:779-782.

Almenoff JS, LaCroix KK, Yuen NA, Fram D, and DuMouchel W. Comparative performance of two quantitative safety signalling methods: implications for use in a pharmacovigilance department. Drug Saf. 2006;29(10):875-87.

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

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

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

CONTACT INFORMATION

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

Pankaj Bhardwaj
Technical Architect
Tata Consultancy Services Limited,
PTI Building, Parliament Street,
New Delhi - 110001
Cell # +91 9891877776
pankaj.bhardwaj@tcs.com

Chandramouli Raghuram
Enterprise Architect (SAS)
Life Sciences Industry Solutions Unit
Tata Consultancy Services Limited
379 Thornall Street,
Edison, New Jersey 08869
Cell # +1 (267) 474 - 7342
chandra.mouli@tcs.com