

Automated Drug Safety Signal Detection with Guided Analysis

Eric Brinsfield, SAS Institute Inc., Cary, NC, USA

David Olaleye, SAS Institute Inc., Cary, NC, USA

ABSTRACT

Issues associated with drug safety continue to make headlines in the media. Combining SAS advanced analytics with new Data Integration and Business Intelligence capabilities yields an efficient way to detect drug safety signals and investigate their cause. The resulting software solution includes:

- data import from drug safety systems
- access to clinical or post-approval data
- metadata management
- updates of drug safety data marts
- execution of industry-standard and new signal detection algorithms
- ranked lists of possible product-related safety signals
- drill-down to supporting information and case details and profiles

During the presentation, Eric and David will discuss an approach for analyzing safety data from both clinical and post-approval sources. The presentation will focus on the automated signal detection process rather than the subsequent statistical analysis. Based on an actual solution, they will highlight the architecture, user interfaces, and analytical methods and review them with respect to success and applicability.

INTRODUCTION

SAS has provided statistical and reporting capabilities to pharmaceutical companies and regulators for drug safety analysis for many years. Recent developments have increased the need to expand how SAS software is utilized because of the following changes:

- Regulatory agencies and drug companies are under increasing scrutiny and criticism by the public, the media, and government
- More safety data is collected during clinical trials and after approval
- Systems for expedited reporting of ICSRs are commonly utilized and working well
- The volume of data is overwhelming
- Reviewers need help focusing on real issues

THE CHALLENGE

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 trends that may be emerging. To optimize the use of valuable medical and epidemiological resources, companies and regulators need ways to narrow the possibilities or find growing problems sooner.

As a step forward, safety reviewers enlisted programmers to create reports or run queries based on their instincts or hunches. Although better than manually scanning through reams of paper, the process is slow and still requires manual guidance and imagination.

Reviewers need an automated filtering mechanism that improves efficiency by helping to detect growing trends sooner and decreasing time wasted on false signals. Some pharmaceutical companies employ a team that performs regular data mining activities on safety data, but not all companies can justify creating a dedicated

PhUSE 2008

analytical team. As an alternative, automated signal detection provides a proactive, yet cost-effective approach for any risk management plan.

DETECTING SIGNALS

According to the *Report of CIOMS Working Group IV* (1998), a safety signal is defined as:

“A report or reports of an event with an unknown causal relationship to treatment that is recognized as worthy of further exploration and continued surveillance.”

Determining when a report is “worthy” is the challenge. Epidemiologists and medical practitioners make the final call as to “worthiness,” but automated safety signal detection can provide a preliminary prioritization of “worthiness” through advanced analytics and signal detection algorithms. Also, as Hauben and Bates (2007) state,

“Sensitivity is of fundamental importance in achieving these goals, but practical reality dictates that the search for truth in pharmacovigilance requires judicious limitations on the numbers of associations that we investigate.”

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 addition, signals must be managed. Automated systems should provide a way to flag and filter out unwanted signals and to limit drugs or compounds by investigator. Safety reviewers may want to filter out signals that:

- Represent known side effects
- Have already been investigated and deemed unimportant or “unworthy”
- Are not medically important

SOURCES OF SAFETY DATA

Signal detection can be performed on a variety of data sources. Each data source brings its own challenges. At a high level, we can group the data into several categories:

- Clinical data – collected during a clinical trial
- Sponsor-collected safety data – collected both during clinical trial and from spontaneous reporting
- Regulatory-collected safety data – collected through spontaneous reporting, such as Medwatch
- Health care data – owned by health care providers, health plans, and government agencies

Each of these data types are discussed in the sections below.

THE SOLUTION

In addition to the powerful analytics within SAS software, SAS Business Intelligence and SAS Data Integration provide new options for the automation, flexible development, and resulting usability needed for a patient safety signal detection system. SAS Enterprise Miner and SAS Text Miner are recommended optional components. In general, the system uses SAS Data Integration to:

- load new data from the customers safety system into a detailed data store
- clean, standardize, and transform the data
- run signal detection algorithms on predetermined combinations of products or product groups
- create data marts that contain summary results from the signal detection, as well as supporting data marts for exploratory reports

The system uses SAS Business Intelligence to display results, provide logical workflows, and enable exploration of signals. In most cases, SAS Stored Processes enable flexibility and full validation when generating results and displaying web pages and reports.

The signal detection components described here fit within a larger SAS patient safety solution, whose other features will not be discussed.

WORKFLOW FOR GUIDED ANALYSIS

The solution provides several workflow pathways for reviewing and exploring signals. Figure 1 shows one workflow that provides the reviewer with an overall summary of products and product groupings ranked in descending order of perceived risk. The first page that the reviewer sees is referred to as the Product Signal Summary, whose information is based on a compilation of results from the signal detection algorithms for all products and product groupings. The risk index is based on a weighted composite of multiple signal detection algorithms.

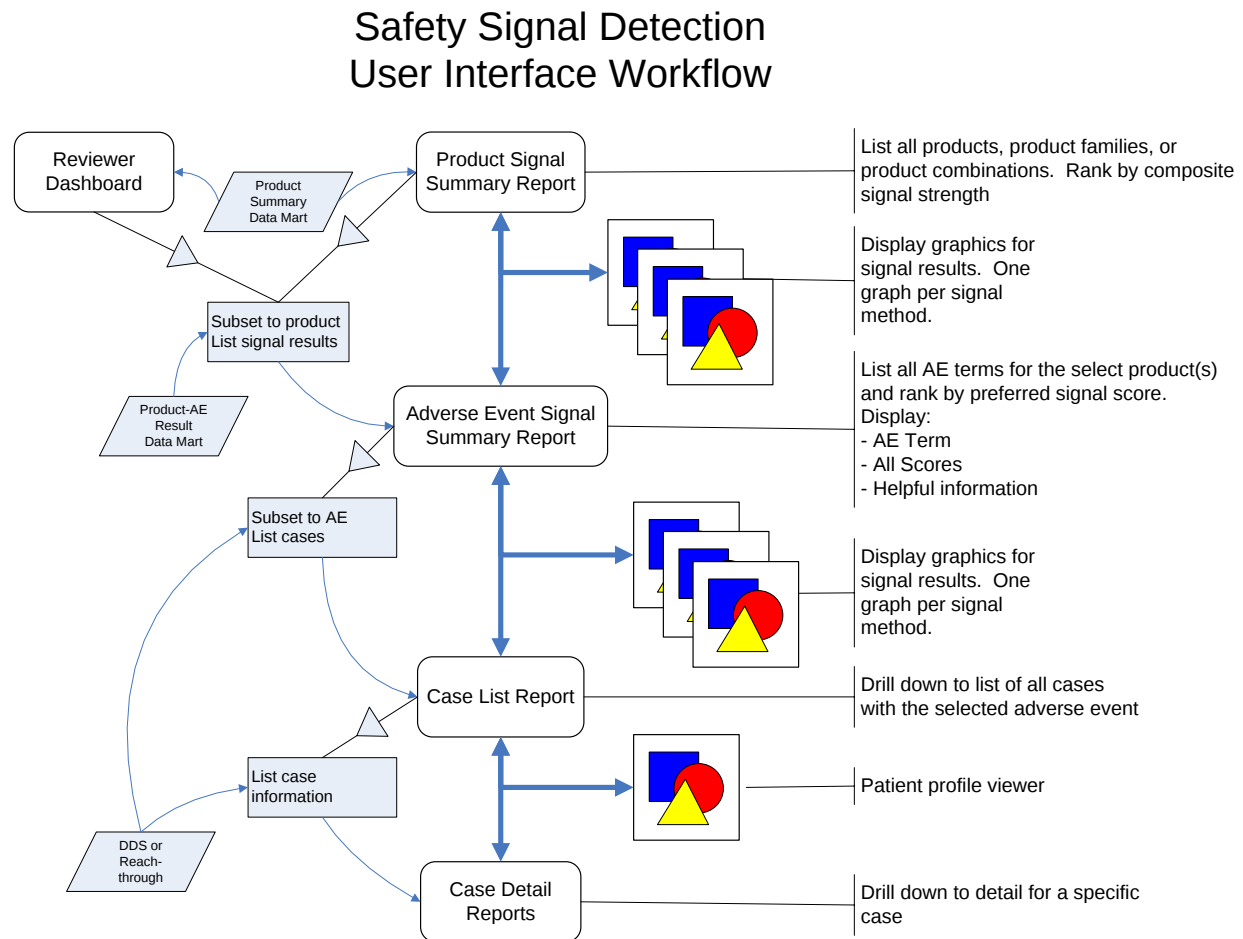


Figure 1: Sample workflow

From the Product Signal Summary page, users drill-down by clicking on various fields to get more detail about signals. In the central flow, users click on a product to see a list of Adverse Events (AEs) for that product or product group along with the associated signal scores (Figure 2). From that page, users either click on scores to see explanatory graphics or continue to drill down to a list of cases for a selected adverse event. For large numbers of cases, the system shows clusters of cases that may be related.

Alternatively, users may prefer to start the workflow with a customized graphical dashboard that only displays products for which a specific reviewer is responsible. Alerts and key metrics provide links into the workflow.

PhUSE 2008

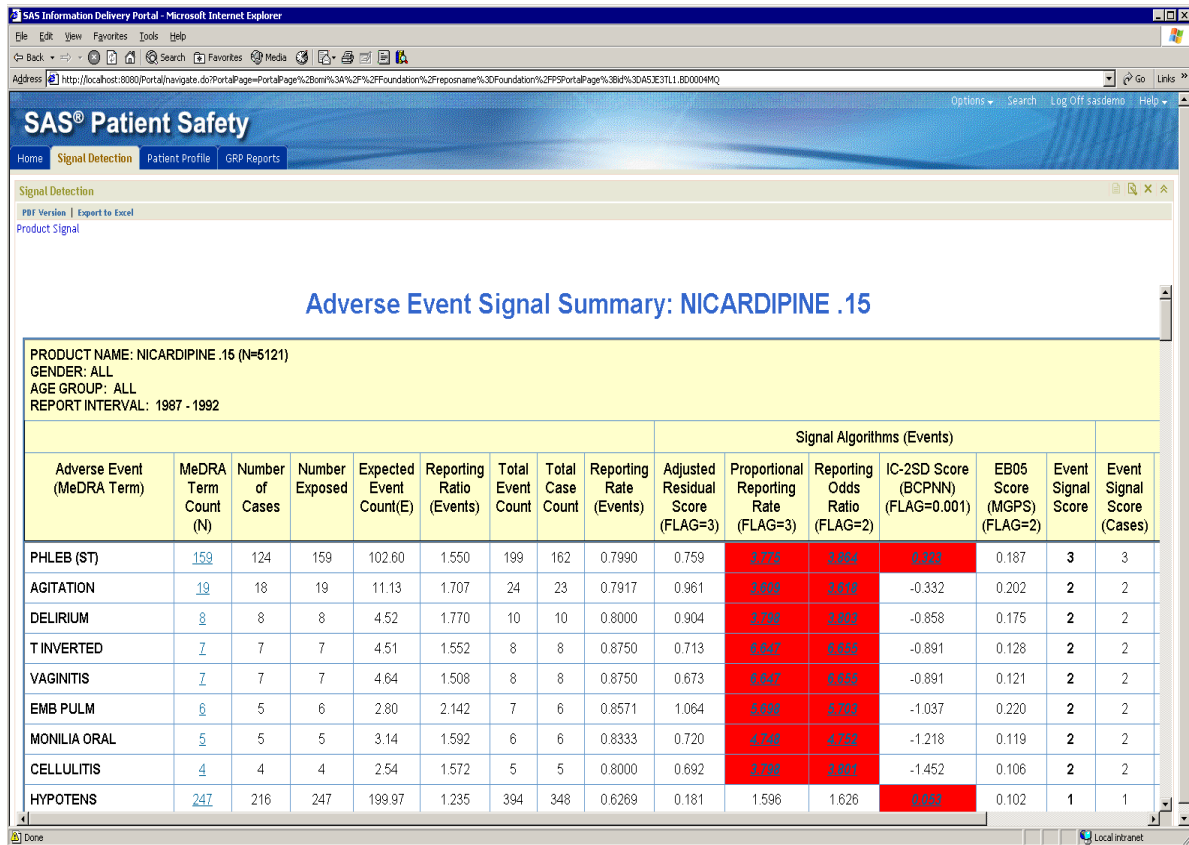


Figure 2: Adverse Event Signal Summary with highlighted signal scores

If a user clicks on a case number, the system displays a detailed view of the case information, including any narratives provided by the reporter. Graphic representations are available at all levels of the workflow.

Different workflows are provided so that users can choose how they want to pursue the investigation. Providing “guided analysis,” the system provides high-level indicators backed by supporting evidence and explanation. Because the detailed data store and data marts are permanent structures, reviewers also have the option of performing ad hoc queries directly against the data, outside of the workflow.

In addition, the stored processes that are listed in Figure 1 can be executed from within SAS Enterprise Guide, Microsoft Excel, or other components of SAS.

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 (for example, adverse events syndromes and polytherapy), 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:

- the 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. For clinical trials signal detection, the screening process compares the degree to which patient population on a drug regimen experience adverse events at a rate higher than the patient population on placebo or a comparator drug.

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 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").

ARCHITECTURE AND DATA FLOW

The system is based on the multi-tier SAS Intelligence Platform. Most of the heavy lifting is performed with SAS Data Integration, while the presentation relies on SAS Business Intelligence, as shown in Figure 3.

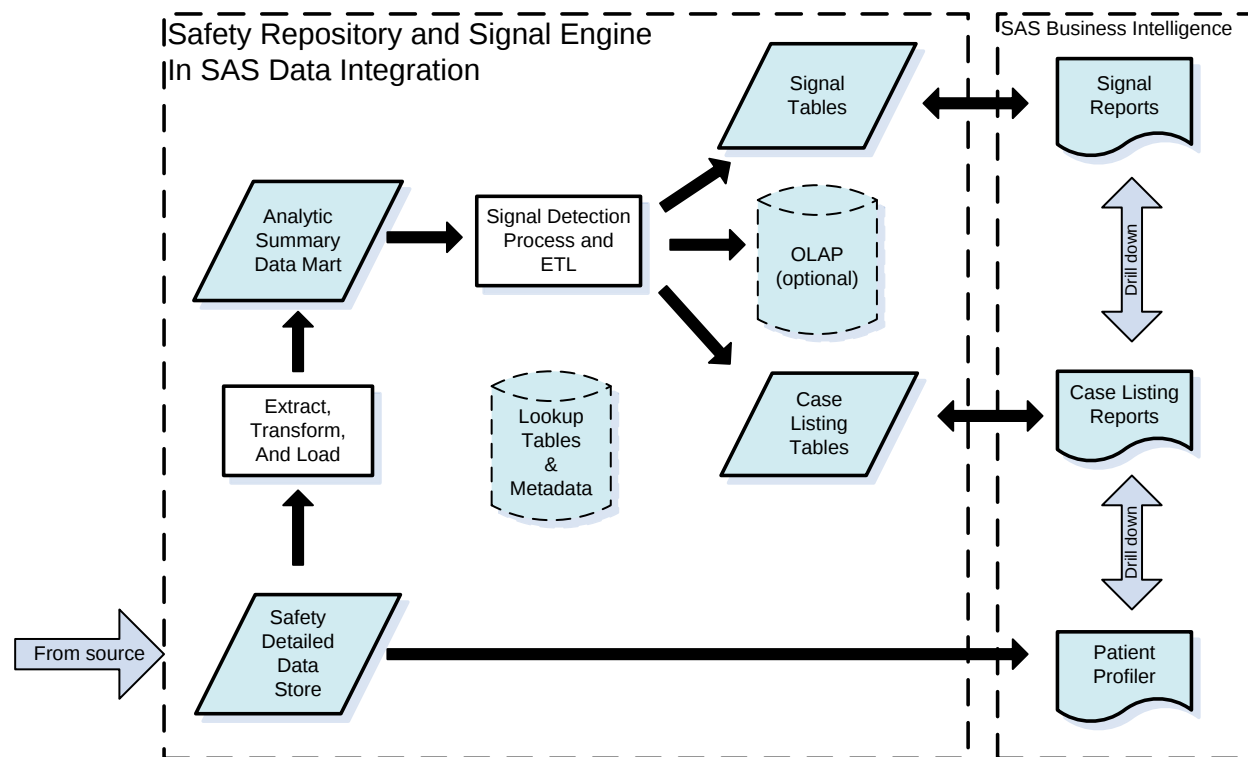


Figure 3: High-level architecture and data flow

LIMITATIONS OF DATA

The source of safety data will constrain or determine the types of analyses that can be performed. Each category of has its own unique idiosyncrasies. SAS Data Integration offers the flexibility to manage the processing of each data source, while SAS Business Intelligence provides the intelligence user interface layer to control which is accessed or viewed.

PhUSE 2008

CLINICAL DATA

Clinical data usually provides detailed information about a patient, including accurate estimates of exposure to a drug. The data includes demographic information as well as concurrent medications and other aspects of medical history.

Due to the nature of clinical trials, the data is very controlled or restricted in scope, which means that safety signals may not show up until the drug is released to the broader population. As part of a good risk management plan, researchers should still include clinical data in their signal detection activities in order to catch unexpected trends that may not be revealed during traditional safety analysis.

SPONSOR-COLLECTED SAFETY DATA

Sponsor-collected safety data include adverse event reports collected from clinical trials as well as post-approval external reports. Although the data does not usually contain accurate accounts of dose exposure and does not include patient history, the data includes a broader audience of patients using the products in the field. Thus, the data covers subpopulations that may not have been included in a clinical trial.

This data can provide valuable insights into product safety based on the reactions of the general population. Unfortunately, without detailed background information, including concomitant medications or medical history, signal detection results are difficult to confirm or trust without additional study or different data. Disproportionality algorithms are required for this data due to the missing estimate of exposure. Some safety teams have acquired drug sales data to improve their estimates of exposure.

REGULATORY-COLLECTED SAFETY DATA

Many non-commercial agencies and organizations collect adverse event reports from patients and health care providers. The quality and richness of the data varies between each source, but most still lack good estimates of exposure or provide medical history. For example, the FDA collects information through the MedWatch program, whose data is provided to researchers in the AERS database.

As with sponsor-collected safety data, safety researchers must use disproportionality methods for automatic signal detection. Sales data can also be merged into these databases to provide a better estimate of exposure.

HEALTH CARE DATA

Health care provider data and health insurance claims data provide new hope for achieving better analytical insight. This type of data provides a greater opportunity to extract patient history, concomitant medications, and dose exposures. Health care data has provided a new frontier in signal detection.

Health care data should enable many new benefits, including:

- Obtaining better estimates of exposure
- Using better methods of analysis once better estimates of exposure are obtained
- Analyzing interactions between a drug and other drugs or other events or procedures
- Analyzing longitudinal data

Unfortunately, health care data also brings some new challenges. For example:

- Patients tend not to stay with the same health plan or visit the same hospital throughout life, so full medical history is someone elusive.
- Health care data contains much more information than just adverse events. Consequently, researchers need to spend more effort separating normal events (e.g., child-birth) from non-drug related adverse events (e.g., broken bones) from adverse drug reactions.
- Data may need even more cleaning for analytical purposes and other sources.

PhUSE 2008

TEXTUAL DATA

To cover all possible opportunities to detect trends or signals earlier, a thorough risk management plan should include proactive analysis of textual data, such as literature, narratives, and other documents.

CONCLUSION

Combining SAS analytics with SAS Data Integration and SAS Business Intelligence offers a new opportunity to improve patient safety. Through automation and guided analysis, reviewers can increase efficiency by focusing on real problems rather than false alarms. In addition, reviewers can investigate signals more easily, thereby saving time and increasing the understanding of drug-AE relationships. To take advantage of health care data, SAS' flexibility to support different data models and add new analyses are essential.

REFERENCES

1. CIOMS (1998). *Benefit-Risk Balance for Marketed Drugs, Report of CIOMS Working Group IV*. Geneva.
2. Hauben, M., Bate, A. (2007). *Data Mining in Drug Safety. An Essay in Side Effects of Drugs Annual 29*. Elsevier.

CONTACT INFORMATION

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

Eric Brinsfield and Dr. David Olaleye
SAS Institute Inc.
SAS Campus Drive, Bldg T
Cary, NC 27513
Work Phone: (919) 531-0213 or (919) 531-1645
E-mail: eric.brinsfield@sas.com or david.olaleye@sas.com
Web: www.sas.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.
Other brand and product names are trademarks of their respective companies.