

Safety Monitoring Methodology in the Premarketing Setting

American Statistical Association (ASA) Safety Workgroup on Safety Monitoring and Reporting*[†]

List of Authors*

Michael Fries¹, Karolyn Kracht², Judy Li³, Melvin Munsaka⁴, Matilde Sanchez-Kam⁵,
Krishan Singh⁶, William Wang⁷, Ed Whalen⁸, Kefei Zhou⁹

¹CSL Behring, 1020 First Av, King of Prussia, PA 19406

²AbbVie Inc., 1 North Waukegan Rd, North Chicago, IL 60064

³FDA, 10903 New Hampshire Av, Silver Spring, MD 20993

⁴Takeda Inc., One Takeda Pkwy, Deerfield, IL 60015

⁵SanchezKam LLC, 6331 River Downs Rd, Alexandria, VA 22312

⁶Glaxo Smith Kline Plc, 5 Crescent Dr, Philadelphia, PA 19112

⁷Merck Inc., 2000 Galloping Hill Rd, Kenilworth, NJ 07033

⁸Pfizer Inc., 235 East 42nd St, NY, NY 10017

⁹Amgen Inc., 1120 Veterans Blvd, South San Francisco, CA 94080

[†]Corresponding Authors: Melvin Munsaka, William Wang, Ed Whalen

Abstract

Safety monitoring and reporting has achieved a greater level of attention in the past 15 years. Statisticians play an important role in learning about a drug's safety profile. An ASA Safety Monitoring Working Group was established with a goal to empower the biostatistics community to play a proactive role and better enable quantification in safety monitoring. As part of its effort, this paper presents a systematic review and unique perspective on the existing methodology developments, which include Bayesian and frequentist, blinded versus unblinded safety monitoring, individual versus aggregate data meta-analyses, pre- and post-marketing methods, static versus dynamic safety reviews, and methods of visualization. These perspectives may serve as a background for future statistical work, both in methodology development and its application.

Key Words: Safety monitoring, Bayesian

1. Introduction

In June 1997 the Food and Drug Administration (FDA) approved a calcium channel blocker (mibefradil) for hypertension and angina, but with some concerns about drug-drug interactions. Safety monitoring continued after drug approval in the form of monitoring spontaneous reports and further drug-drug interaction studies with FDA requiring the addition of more drug-drug interactions to the product label. See, for instance, historical documents at:

http://www.fda.gov/ohrms/dockets/ac/98/briefingbook/1998-3454B1_03_WL32.pdf. In June 1998, only one year following approval, the FDA asked the sponsor to withdraw the drug. This extreme case illustrates the value of continual safety monitoring of biopharmaceutical products and opens the question of how safety monitoring prior to approval can help uncover risks sooner in a product's lifetime. Although the mibefradil example ended in voluntary withdrawal, many other products have labels amended as knowledge of their risks grows with accumulating data from many sources such as the

FDA Adverse Event Reporting System (FAERS) for spontaneous report events or FDA's evolving SENTINEL system that uses observational payer data and electronic health records. These systems and their counterparts around the world provide a perspective on safety risks but not with the level of control that the randomized controlled trial (RCT) brings to bear.

Quite a bit has been written on the advantages and disadvantages, as well as the complementary nature, of RCT data versus these typically observational data sources. This paper looks closer at how to use RCTs with respect to their advantages with the hope that better and earlier use of these data will improve the timing of identifying risks. It puts in perspective statistical methodologies for safety monitoring, with a focus during the clinical development stages. The paper reviews different categories of methodologies, such as Bayesian and frequentist, blinded vs unblinded safety reviews, individual vs aggregate data meta-analyses, pre- vs post-marketing methods, static vs dynamic safety reviews, and methods of visualization that can bring coherence to safety reviews.

The contents of this paper look at safety monitoring and reporting (SMR) from the statistician's perspective while keeping the larger SMR landscape in mind. The focus is on strategies to identify issues during the pre-marketing development phases, but will include descriptions of methods typically used in post-approval surveillance. Section 2 puts forth a working definition of safety monitoring to frame the subsequent material. Section 3 describes various aspects of the statistical methods used in SMR. In an effort to present a unique view in our summary, we chose to summarize the literature using the following six perspectives on data and analysis: Bayesian versus frequentist approaches; blinded versus unblinded assessments; post-marketing methods in pre-marketing evaluations; static versus dynamic evaluations; analyses from patient level data to meta-analytical evaluations; and visual analytics for improving comprehension. Section 4 discusses the state of SMR and some areas in which to improve on it. The passage, in 1965, of the FDA amendment act led to an emphasis on the proper study of efficacy claims in the drug industry. Events of the last 20 years, such as the previous example with mibefradil, have exposed the importance of the proper study and interpretation of drug safety. The ASA Biopharmaceutical Section recognizes this and has established the Safety Working Group. The Safety Monitoring and Reporting Group (SMRG) is one of the initiatives under the umbrella of the Safety Working Group. The SMRG consists of two workstreams (WS1 and WS2) with the following initial goals: WS1, review safety regulations and survey industry and interview thought-leaders; WS2, review statistical methodologies used across the spectrum of SMR. Figure 1 provides a list of our working group members for each working stream and the statistics advisors.

The scope of this paper is limited to current practices but many improvements have been made in the last two decades. Furthermore, the number of new data sources is growing with which we can examine drug safety. For example, ten years ago social media would have little to offer, but today it represents just one of many potential sources to look for trends or signals, see for example, Sarker, *et. al.* (2015). Others sources may be more familiar but also are evolving in their depth of information such as with electronic health records.

This article represents an initial step specifically into the area of SMR and, more generally, drug safety from the statistical point of view. The authors are part of a larger group of statisticians working through the ASA to raise awareness among statisticians of the need for greater input from our community in drug safety and to, hopefully, begin

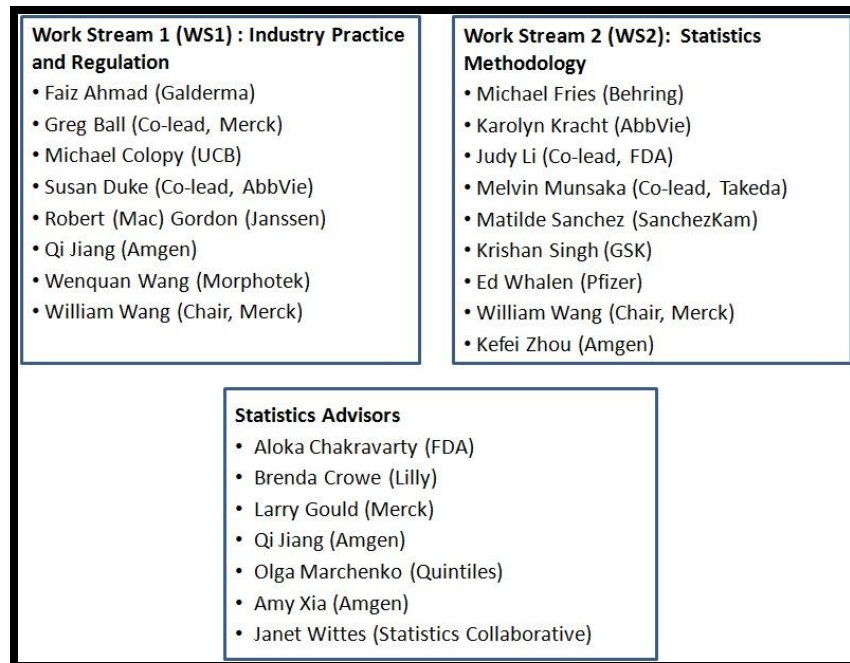


Figure 1: ASA SMWG members and advisors

exploring some of the many areas of statistical research that have been investigated and need further investigation. Finally, because of statisticians' unique mix of skills, the data rich environment in drug safety and the many open questions, statisticians have an opportunity to take both a partnership and leadership role in the field. We hope that with this paper, others will catch interest and the sense of urgency those opportunities offer.

2. Working Definition of Safety Monitoring and Reporting

A one sentence definition of safety monitoring and reporting cannot cover SMR adequately given its multiple stakeholders, methods and purposes. A clinician working with an ongoing study may think of the individual patients he or she monitors as they progress through the study and the reporting of serious adverse events with an eye toward finding any trends of concern. A statistician may think of the many studies completed and/or ongoing, and look for overall trends as the drug continues in development. A regulator may have interest in studies as well as events of interests for a class of similar drugs. Although a simple definition may not capture the scope and many facets of SMR, one defining characteristic of SMR is the need for all stakeholders to examine available safety data for the drug to discover any trends that suggest side effects not expected or that are expected but of greater magnitude than anticipated. Monitoring of drug safety from the sponsor's perspective has several defining characteristics:

- Safety monitoring is a process and involves a wide range of stakeholders
- It assess a range of side effects across a spectrum of frequencies and magnitudes, ranging from easily detected side effects to those that rarely occur but have potentially high impact on patient well-being
- It serves to lay the foundation for an integrated analysis of safety or benefit-risk analysis in regulatory submissions such as a new drug application (NDA), or for a possible advisory committee meeting

- Post approval, SMR grows in terms of the available data sources and who needs to draw conclusions from them
- Monitoring can be dynamic as in pharmacovigilance or as outlined in the SPERT paper; or static such as for an NDA integrated analysis of safety

Figure 2 outlines the different dimension and stakeholders that define SMR.

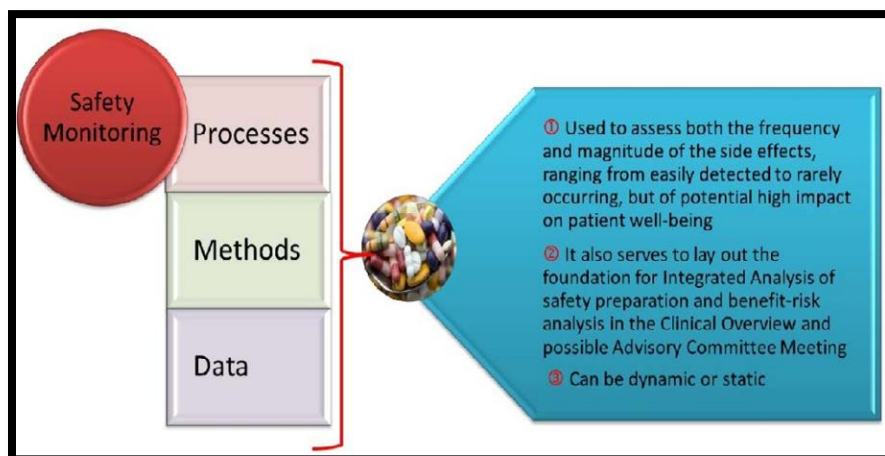


Figure 2: Multiple Dimensions and Associated Stakeholders in Safety Monitoring

Note that while the specifics here speak in terms of FDA, European Medicines Agency (EMA) and other regulators as the primary audience, the importance of SMR can positively address the concerns of many other stakeholders from patients to sponsors trying to decide the appropriate and optimal patient segments for the drug product.

3. A Perspective on Statistical Methods in Safety Monitoring

A Webster definition of “perspective” is “the capacity to view things in their true relations or relative importance” or “the interrelation in which a subject or its parts are mentally viewed”:

(<http://www.merriam-webster.com/dictionary/perspective> assed on 9/11/2016). Statistical thinking and methodologies can apply to drug safety data in many different ways. To provide some useful perspectives on the statistical methodologies, we would want to examine relevant methods in a meaningful relationship.

The following points outline one, but definitely not the only, way to break out some of those statistical aspects to provide a coherent framework for discussing drug safety statistics, especially in the development (pre-marketing) phases. The following points align with Figure 2.

Breaking down statistical methods and thinking for Safety Monitoring:

- Bayesian versus Frequentist Approaches
- Blinded versus Unblinded Assessment
- Post-marketing PV versus Pre-Marketing Evaluation
- Static versus Dynamic Evaluation
- Analyses from Patient Level Data to Meta-Analytical Evaluation
- Visual Analytics versus Tabular Approaches

Unlike a typical statistical application such as confirming a primary endpoint, safety monitoring requires multiple statistical views of safety data. Figure 3 illustrates further how these features form part of the development of an overall safety picture as the sponsors and regulators work to understand, track and fill out the safety profile of a drug.

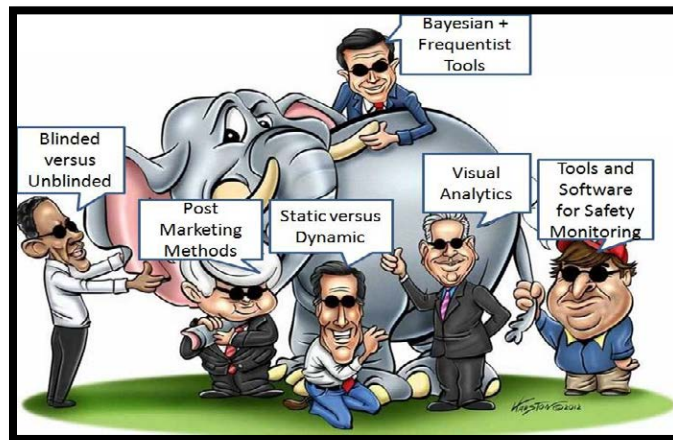


Figure 3: Safety Monitoring Involves Multiple Disciplines, Methods, and Interrelated Processes. Adapted from: <https://www.youtube.com/watch?v=ygT-UwgEaSo>

The following subsections cover some of the key specifics regarding each of these six general areas.

3.1 Bayesian and Frequentist Approaches

Comparing to other subsections, the methods in this subsection are covered in detail because many of them form the basis of some approaches described in later subsections. We will consider two distinct situations: 1) Bayesian stopping rule and dynamic monitoring for an event of interest (EOI) for an ongoing trial and 2) clinical trial signal detection.

3.1.1 Bayesian stopping rule and dynamic monitoring for an event of interest (EOI) for an ongoing trial

Some Bayesian methods for monitoring and stopping based on signal trends include beta-binomial and Poisson or gamma models as described in Tarone 1982, Thall and Simon 1994, Resnic *et al.* 2004, Yao *et al.* 2013, Xia *et al.* 2013 and Kashiwabara, 2014. Some of the key features of these methods allow for dynamic updates in monitoring, parameters updates as data accrue, and the setting of thresholds such that if posterior probability of exceeding a certain rate is greater than the threshold, then it indicates a need to stop or make an appropriate decision regarding further research.

Thall and Simon proposed a Bayesian approach to monitor phase IIB studies continuously based on a Bayesian beta-binomial model (Thall and Simon 1994). This is a direct application of the classic Bayesian construct with a Beta prior distribution for the probability parameter of the binomial distribution. During the trial, the parameters are updated as data are being collected. One of the concerns for the beta-binomial model is

the estimation of the unknown prior. Tarone recommended estimating the prior based on the historical data using maximum likelihood estimation (Tarone 1982). They proposed to estimate the priors based on the information from the animal model data which are collected in the preclinical studies.

Frequentist monitoring of clinical safety could be done under the group sequential paradigm. Jennison and Turnbull proposed a method for assessing both efficacy and safety under a group sequential paradigm (Jennison and Turnbull, 1993). Bolland and Whitehead (Bolland 2000) discuss formal sequential procedures for monitoring safety.

3.1.2 Clinical trial signal detection

In practice, the usual approach for safety data analysis involves the comparison of the proportion of subjects who experience an adverse event (AE) between treatment groups for each type of AE. This involves a large number of analyses with inadequate statistical power and no meaningful control of type 1 error. Unadjusted analyses may result in a large number of false positive results while using simple adjustments (e.g. Bonferroni procedures) are generally too conservative and counterproductive for considerations of safety. Thus, it is important to develop an approach to analyzing AE data that addresses concerns about multiplicity and the imprecision of data inherent in analysis of small number of events.

Several frequentist (Mehrotra and Heyse; Mehrotra and Adewale) and Bayesian (Berry and Berry; Xia, *et al*; Weaver *et al*) methods have been proposed to address this issue. There are four considerations when an AE should be flagged, namely: (1) actual significance levels; (2) total number of types of AEs being considered; (3) rates for those AEs not considered for flagging, including their similarity with those that are being considered; and (4) biological relationships among the various AEs (Berry and Berry, 2004). The first two are standard considerations in the frequentist approach to multiple comparisons. The second two are not, but they are relevant in the Bayesian approach.

The Bayesian approach for safety monitoring incorporates prior knowledge about the safety profile of the treatments and updates knowledge based on accumulating data. The potential advantages of the Bayesian approach relative to classical frequentist statistical methods also include the flexibility of incorporating the current knowledge of the safety profile originating from multiple sources into the decision-making process. However, it is sometimes difficult to specify the prior distributions reliably.

In early development most information will come from literature or early phase sponsor trials. Which events to follow come from knowledge of mechanisms of action or small amounts of trial data. This leads naturally to consideration of Bayesian methods. At later stages frequentists methods may be more helpful especially for events of the Tier 1 type based on the hierarchy in Crowe *et al* (2009).

Weaver *et al* (2016) introduced three approaches using simple Bayesian methods to access pre-specified adverse event: Single conjugate prior; meta-analytic predictive (MAP) prior, which comprises a mixture of conjugate priors; and a robust mixture prior that incorporates a robust parameter and a weakly informative component to the MAP prior so that the prior-data conflict could be handled more flexibly. Berry and Berry (2004) proposed a three-level hierarchical mixed model to account for multiplicities in AE assessment. The basic level is the type of AE. The second level refers to body system

which contains a number of types of possibly related AEs. The highest level is the collection of all body systems. The proposed three-level hierarchical mixed model provides an explicit method for borrowing information across types of AEs. The probability that a drug caused a type of AE is greater if its rate is elevated for several types of AEs within the same body system than if the AEs with elevated rates were in different body systems. AEs in the same body system are modeled as being exchangeable. The hierarchical nature of the model gives rise to regression effect which is appealing in the context of multiplicities because it modulates extremes. In their model, the conclusion that 1 type of AE is affected by treatment depends on data from other AE types, especially from within the same body system. More traditional multiple comparison methods base conclusions only on the number of AE types under consideration.

Xia *et al* (2011) expanded Berry and Berry's method into a hierarchical Poisson mixture model which accounts for the length of the observation of subjects and improves the characteristics of the analysis for rare events. This makes the applicability of this method reach beyond Tier 2 events (see Crowe *et al* for the Tier definitions). They also considered 5 different approaches:

- Model 1a: three-stage model with normal prior on log-OR (logarithm of odds ratio).
- Model 1b: three-stage model with mixture prior on log-OR.
- Model 1c: nonhierarchical one-stage Bayesian mixture model.
- Model 2a: three-stage model with normal prior on log-RR (logarithm of relative risk).
- Model 2b: three-stage model with mixture prior on log-RR.

DuMouchel (2012) proposed a multivariate Bayesian logistic regression (MBLR) method to analyze safety data when there are rare events and sparse data from a pool of clinical studies. It is designed to be a compromise between performing separate analyses of each event and a single analysis of a pooled event. It requires the selection of a set of medically related issues, potentially exchangeable with respect to their dependence on treatment and covariates. As with the Berry and Berry method, MBLR assumes that the events are classified into similar medical groupings in order to use a shrinkage model to allow borrowing strength across similar events. However, Berry and Berry do not consider covariates or the use of logistic regression.

The MBLR method is exploratory in nature and examines the relationship of the adverse event frequencies to multiple covariates and to treatment by covariate interactions. The rationale for including covariates in MBLR is not so much to adjust for potential biases in the treatment main effect, but to be able to include treatment-by-covariate interactions in order to detect possibly vulnerable subgroups that might react differently to the treatment. Note that this more complicated model may not be estimable by a standard logistic regression algorithm because the data are often too sparse for the number of parameters being estimated. The proposed Bayesian method allows the analysis of each event to borrow strength from the other events, assumed medically related. There is also a tendency for the treatment-by-covariate interaction coefficients to be shrunk toward the null value of 0. This shrinkage is intended to offset the tendency of exploratory methods to find "significant" subgroup effects purely by chance.

Gould (2008, 2013) proposed an alternative Bayesian screening approach to detect potential safety issues when event counts arise from binomial and Poisson distributions. The method assumes that the adverse event incidences are realizations from a mixture of

distributions and seeks to identify the element of the mixture corresponding to each adverse event. One of these distributions applies to events where there is no true difference between the treatments and one applies to events where there is a treatment effect. The components of the mixture are specified a priori, and the calculations then determine the posterior probability that the incidences for each adverse event are generated from one or the other of the mixture components. It directly incorporates clinical judgment in the determination of the criteria for treatment association.

The Bayesian approaches can be contrasted with the New Double FDR method from Mehrotra and Adewale (2012), which is a frequentist method for screening frequent (Tier 2) adverse events, controlling for multiplicity. The double FDR method (Mehrotra and Heyse, 2004) involves a two-step application of adjusted P values based on the Benjamini and Hochberg false discovery rate (Benjamini, Hochberg, 1995; available in SAS proc multtest). In the New Double FDR method, they proposed a new flagging mechanism that significantly lowers the false discovery rate without materially compromising the power for detecting true signals, relative to the common no-adjustment approach. The R Package c212 is available to apply the Bayesian hierarchical model and the New Double FDR along with other methods for error control when testing multiple hypotheses: <http://personal.strath.ac.uk/raymond.carragher/files/c212/c212-manual.pdf>

Comparing the frequentist and Bayesian methods, one key advantage of the Bayesian methods is the ability to interpret results for the nature of the adverse event, it is possible to even pre-specify probability criteria for interpreting the posterior distributions, while the frequentist method is not adjusted for nature of AE – all events are treated equal. The Bayesian method can also incorporate historical data, and provide posterior based credibility intervals for each adverse event. The ability to continuously monitor a trial under the Bayesian philosophy is also a clear advantage. The recent work by Xia *et. al.* also extends beyond the restriction of Tier 2 events. On the other hand, the frequentist method is easy to implement, while in the Bayesian methodology estimation is done using MCMC methods, and convergence of the chains must be considered across each parameter of the model. This can be daunting for the novice user of these techniques. Frequentist methods fit well with situations involving a fixed number of looks at the data and where well defined notions of type I error control are needed. Exploring the sensitivity to the choice of prior is suggested for Bayesian analysis (Spiegelhalter, 2004) and adds additional complexity and effort.

3.2 Blinded vs Unblinded Safety Assessments

Before we go into the question of blinded versus unblinded assessment of safety data, we will begin by first defining what we mean by blinded assessment of safety data as opposed to unblinded assessment. We define blinded safety monitoring to be the setting in which a clinical trial is ongoing and there is some interest in assessing safety outcome(s) of interest in a blinded fashion in the sense that no treatment information is made available. This also includes the setting in which the treatment allocation is provided but without knowledge of the actual treatment, for example, Treatment A versus Treatment B. One immediate point to note is that blinded looks at safety data will not be as informative and as efficient as in an unblinded look at the data where the treatments are known.

During drug development many ongoing studies may hold important clues to the safety profile. Upon study completion, the data typically get unblinded for assessment of efficacy and safety. Any unblinding prior to study completion is either part of a planned or unplanned set of interim analyses or the result of an unblinding specific patients who have experienced serious adverse events that require knowledge of the blinded treatment for deciding on countermeasures to the event. However, a safety data monitoring committee (DMC) can make much better judgments about a drug program if they have some level of actual treatments received, but at the expense of no longer having a fully blinded study, at least for some involved with it. Importantly and finally, there is an ethical and moral responsibility to proactively assess safety interest of patients.

Assessing safety data in a blinded fashion comes with various challenges and issues, including ethical considerations, statistical considerations, and design considerations. Ball (2011), Ball and Piller (2011), Ball, Piller, and Silverman (2011), Ball and Silverman (2011), and Gould and Wang (2016) has discussed these issues and challenges as they pertain to blinded safety monitoring. For example, proactive safety signal monitoring using blinded data in an on-going clinical trial will inevitably raise logistical questions regarding monitoring patient safety while at the same time maintaining the study blind. Another consideration that comes up is with regard to available information on the safety concern as well as how to address blinded safety monitoring in the case where some studies are still ongoing. In particular, how does one harness from the completed studies and those that are ongoing?

All the above considerations raise the question: how much information and value can blinded safety monitoring provide? As it turns out, a good deal of safety information can be gained from blinded data as suggested in the definition itself and some consequences of defining it this way. Table 1 lists the merits of using blinded assessments along with the constraints.

When conducting blinded SMRs the choices of the preceding subsection on Bayesian and frequentists, methods have been explored by several authors. For both the Bayesian and frequentist methods, the general idea is to make some inference about the rate (or an exposure adjusted-rate) θ of a safety concern, for example, that of an adverse event of special interest. Note that θ can be a derived metric, such as the relative risk, risk difference, or odds ratio. Further, in both the Bayesian and frequentist settings, the ultimate objective is to make a decision on the basis of results from the accrued blinded data. The decision can also be made on the basis of confidence intervals or credible intervals in the case of Bayesian methods.

Table 1: Advantages and Disadvantages of Maintaining Study Blind in SMR Setting	
<i>Advantages</i>	<i>Disadvantages</i>
<ul style="list-style-type: none"> - Identify potential safety issues ahead of scheduled DMC meetings - Identify safety issues that are, or have potential to become a key concern - Drive decisions regarding an unblinded analysis or a decision to setup a DMC, or even stopping a trial or development altogether 	<ul style="list-style-type: none"> - It may not be as informative and efficient as in an unblinded analysis - It will inevitably raise logistical questions regarding monitoring patient safety while at the same time maintaining the study blind

As noted in the previous section, Bayesian methods make use of prior information and likelihood leading to a posterior distribution from which inference is made and frequentist methods generally revert to large sample theory to make inference. Both the Bayesian and frequentist approaches may also utilize historical control data relative to the estimate obtained from the blinded data as part of the decision making process. The decision itself will require careful consideration from various stakeholders, but with patient safety being the ultimate objective. Table 2 lists some and the key features Bayesian and frequentist methods in the context of blinded SMR.

Table 2: Bayesian and Frequentist Methods in the SMR Setting
<i>Bayesian Example Setting</i>
<ul style="list-style-type: none"> - Set criteria for decision making, for example, set cut-off point, $\theta_{\text{Cut-off}}$ which if exceeded by a particular percentage, say, 90% would lead to a decision being made, e.g., unblind study - So, want to estimate a probability associated with θ, e.g., $P(\theta > \theta_{\text{Cut-off}}) > 90\%$ - Via Bayes theorem, use blinded data and appropriate prior to get posterior and which is used to get $P(\theta > \theta_{\text{Cut-off}})$
References: Ball, 2011; Wen, Ball, Dey, 2015; Gould, Wang, 2015; Schnell, Ball, 2016
<i>Frequentist Example Setting</i>
<ul style="list-style-type: none"> - Set criteria for decision making, for example, upper bound θ^U for θ and a cut-off point $\theta_{\text{Cut-off}}$ which if exceeded by θ^U would lead to a decision being made, e.g., unblind study - So want to check if $\theta^U > \theta_{\text{Cut-off}}$ which would lead to a decision being made, e.g., unblind study - Use large sample theory to estimate θ^U using blinded data
References: Chen and Peace, 1987, 2014; Yao, Zhu, Jiang, Xia 2013; Zhu, Yao, Xia, Jiang, 2016; Moye, 2006; Herson, 2015

In summary, a good deal of safety information can be gained from blinded data. This includes identifying potential safety issues ahead of scheduled DMC meetings, identifying safety issues that are, or have potential to become a key concern, and there is ethical and moral responsibility to proactively assess safety interest of patients and blinded safety monitoring. Additionally, it important to acknowledge some of the challenges of blinded SMR, such as logistical questions regarding monitoring patient safety while at the same time maintaining the study blind. It is recommended that when using blinded SMR, the following should be considered carefully: criteria and decisions should be made a priori preferably cross-functionally and there should be some careful thought as to the actual method to use between Bayesian versus frequentist approaches.

3.3 Post-marketing PV versus Pre-Marketing Evaluation

Much progress has been made to develop statistical methods that handle spontaneous report safety data. Among the more prominent methods for analyzing spontaneous reports are the Multi-item Gamma Poisson Shrinker (MGPS) from DuMouchel (1999) and Szarfman, Machado, and O'Neil (2002). These and related methods developed to handle the fact that such data did not provide information to form a denominator any of the observed adverse event and drug combinations. However, such methods can still be applied to randomized controlled data from drug development. Kajungu *et al* (2014) perform such analyses in the setting of a malaria RCT.

In general there are many methods developed for spontaneous report and observational data set such as electronic health record data or claims data. Figure 3 provides an outline of the level of data from development level experiments to later drug lifecycle data source such the spontaneous and observational types. The long, straight arrow indicates that more data sources become available as the drug moves into post-approval phase. Certainly RCTs still get done and can provide critical safety data but questions that require much larger exposures or longer treatment duration may not be doable in an RCT. Therefore claims data, EHRs, and spontaneous report data fill an information gap than RCTs cannot. The shorter arrow refers to the limits of pre-approval data to primarily RCTs and other clinical trial data. Some epidemiological data may help describe general safety concerns for the indicated patient population or for drugs in the same class as the drug under study, but it will not extend to investigational drug specific concerns.

Figure 4 will help in the next subsection as well to understand the history and current use of methods used for the dynamic SMR strategy – such as sequential probability ratio test (SPRT); proportional reporting ratio (PRR); reporting odds ratio (ROR); Bayesian confidence neural network (BCPNN); longitudinal likelihood ratio test (LongLRT). The particular methods and a few others are discussed with an emphasis on their one time analyses (static) uses and ongoing (dynamic) reviews of accumulating safety data.

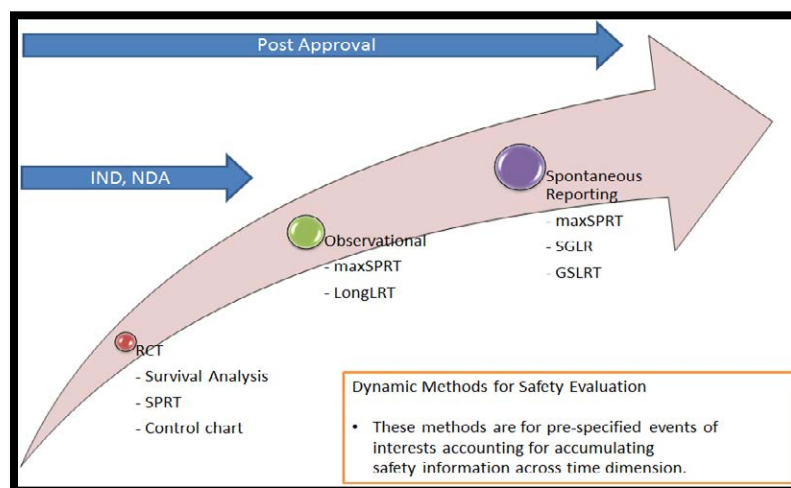


Figure 4: Safety Evaluation During the Life Cycle

3.4 Static versus Dynamic Evaluation

With Figure 3 in mind, SMR activities can be thought of as an ongoing set of reviews of the safety data that get updated as new data become available either within a given study, across a growing set of studies, or over the passage of time as in the case of many observational data sources and spontaneous report sources. Some of the methods in the preceding section are better suited for a dynamic evaluation of data while others work nicely for a one time review as in the production of summaries of clinical safety at the filing of an NDA.

In general, the focus here is on the dynamic approach because it better fulfills the needs for detecting side effect issues as early as possible in a drug development program.

Crowe *et al* have such a paradigm in the setup for repeated analyses of clinical trial data in development. Cumulative meta-analyses also aid in the pre-approval setting with a dynamic framework in which studies get put into a meta-analysis after the completion of each new study. This is discussed further in the next subsection. Table 3 shows methods based on whether they typically get used in a static or dynamic SMR setting. The distinction between methods better suited for one strategy or the other is in part to make easier the discussion of the various methods. Some methods, e.g. Cochran-Mantel-Haenszel (CMH), were designed for a single analysis with specific operating characteristics in mind to control type I error and maintain a prescribed level of power. Methods such as the LongLRT are more general and can accommodate a series of analyses as data accrue – making them better suited for a dynamic data setting.

Table 3: Statistical methods categorized by setting in which they typically get used.	
<i>Dynamic Safety Review</i>	
SPRT	<ul style="list-style-type: none"> - Continuous sequential monitoring test where an adverse event signal is generated if the likelihood ratio exceeds a certain predetermined value, and observation ends if likelihood falls below another predetermined lower bound - For testing a simple null against a simple alternate hypothesis - Allows looking at data in a continuous fashion or as often as needed while controlling type I error - Not suited for surveillance to detect signal, but to monitor a known safety problem - Sensitive to the choice of relative risk for specification of alternate hypothesis
maxSPRT	<ul style="list-style-type: none"> - Maximized SPRT based on truncated MLE is used where the alternate hypothesis is a composite rather than simple. - Controls type I error, and uses only one critical value boundary to reject null hypothesis when excess risk is found. No reason to stop the study if the drug has beneficial effect. - Since test uses acceptance and rejection boundaries that remain unchanged over time, MaxSPRT is a ‘generalized sequential probability ratio test’ (Weiss, 1953)
SGLRT	<ul style="list-style-type: none"> - Sequential generalized likelihood ratio test (Shih et al, 2010) - Continuous monitoring test procedure to detect increased risk of targeted adverse events as soon as possible - Used for evaluation of adverse events in two-armed pre-licensure clinical trials and one armed post-licensure studies - Developed for safety surveillance in vaccine trials
LongLRT	<ul style="list-style-type: none"> - Longitudinal likelihood ratio test (Huang et al 2014) - For active safety surveillance for large databases when exposure information is available - Exposure can be defined as event-time, person-time or exposure-time - Covers recurrent or single cases of one or more AEs for one or more drugs - Uses spending function to control family-wise type I error, decreasing spending function is preferred if maximum number of looks is not specified - SeqLRT a special case of LongLRT where process stops at a look when a success (ie, signal) is found for drug of interest vs placebo or comparator with a single AE of interest

<i>Static Safety Review</i>	
CMH	<ul style="list-style-type: none"> - Cochran-Mantel-Haenszel methods allow stratified analyses of categorical data - With binary data and multiple RCTs the stratifications can prevent Simpson paradox (Chuang-Stein and Beltangady 2011) - Allows for estimation and testing on the odds-ratio, risk-ratio and risk difference scales - Type I error controlled only in a single analysis setting. Additional Type I error controls needed in cases of repeated use
Logistic/Poisson	<ul style="list-style-type: none"> - Logistic regression for binary data with sample unit counts as denominator - Poisson regression also for event counts and with exposures as denominator - Both can estimate and testing on the ratio based and difference based comparisons - Type I error controlled only in a single analysis setting. Additional Type I error controls needed in cases of repeated use
Disproportion Analysis	<ul style="list-style-type: none"> - A general set of comparative methods that were developed for spontaneous reporting data in which no exposure/sample unit counts are available - Proportional reporting ratio (PRR) is one example of such a method based on 2X2 tables as described in Chapter 12, Poluzzi et al. - The four cells of the table are counts of records with A=combination of drug and event of interest, B=other drugs besides drug of interest and event of interest, C=drug of interest but other events, and D=other drugs and other events of interest <ul style="list-style-type: none"> • $PRR = (A/(A+B)) / (C/(C+D))$ • Reporting odds ratio, $ROR = (A/C) / (B/D)$ - Confidence intervals and testing are further described Poluzzi et al - Both PRR and ROR are frequentist methods, Bayesian methods are described below under MGPS and BCPNN
LGPS	<ul style="list-style-type: none"> - In the spontaneous reporting area, DuMouchel (1999) published a Gamma Poisson Shrinkage (GPS DuMouchel 1999), which is based on empirical Bayesian methodology. This has been used and enhanced by the US FDA (Szarfman and O'neil 2002). - Schuemie et al (Pharmacoeconomics Drug Safety 2011) extended the GPS into a Longitudinal GPS (LGPS), which uses person time rather than case counts for the estimation of the expected number of events. In the mini-sentinel and medical claim observational database, drug exposure information is available at the individual subject level. - Longitudinal Evaluation of Observational Profiles of Adverse Events Related to Drugs (LEOPARD) is a method that can be used to automatically discard false drug-event associations caused by protopathic bias or misclassification of the dates of the adverse events by comparing prior event prescription rates to post event prescription rates.
MGPS	<ul style="list-style-type: none"> - Multi-item gamma Poisson shrinker, described in DuMouchel (1999), Szarfman et al (2002) - $MGPS = A(A+B+C+D) / ((A+C)(A+B))$ where A, B, C, and D defined under disproportion analyses - Measures of association based on empirical Bayes bounds
BCPNN	<ul style="list-style-type: none"> - Bayesian Confidence Propagation Neural Network, described in Bate et al (1998) - $BCPNN = \log_2(MGPS)$ - Measures of association based on information criteria

Mixture Bayesian	<ul style="list-style-type: none"> - Gould (JBS, 2013) has extended DouMohcel's GPS setup to detecting Potential Safety Issues in Large Clinical or Observational Trials by Bayesian Screening When Event Counts Arise from Poisson Distributions. - The method assumes that the adverse event incidences are realizations from a mixture of two distributions, one coming from the same distribution as the control group, one coming from an alternative distribution - The Bayesian setup provides the posterior probability that the test group event rate was generated by the same process that generated the control group event rate. Low values of this probability identify events whose risk may be elevated in the test group relative to the control group.
------------------	---

The descriptions point to a tendency for LRT based methods to be used in dynamic settings. Most methods in the 'static' rows can be used in a dynamic setting but would require some method for false signal adjustment such as an FDR method or the method outlined in Hu, Cappelleri, Lan (2003, 2007).

3.5 Analyses from Patient Level Data to Meta-Analytical Evaluation

Meta-analyses play an important role in submissions for approval. In the past, and to an extent today, most involved patient-level data that were pooled across several studies with no adjustments for study and producing simple summary statistics such as the proportion of patients with a given adverse event or a similar rate using treatment exposure as the denominator. In recent years important safety concerns have emerged – e.g. cardiovascular side effects in the development of anti-diabetic treatments – in which sponsors and regulators need to look deeper as certain side effect questions. Some questions can be approached prospectively while other may arise from the SMR process. In either case, simple statistics will not suffice, methods that may be thought of meta-analytic come into play, see Crowe *et al* (2016), Chuang-Stein, C., Beltangady, M. (2011).

Prior to a drug application filing, cumulative meta-analysis can be used to update data from new trials every time there is a new trial available for the evaluation benefit or harm. Crowe *et al* (2011) provide a broad description of the value and strategies for incorporating these meta-analyses into a drug's development. One issue that may arise when doing cumulative meta-analyses stems from the use of multiple inspections of data and their effect on type 1 error. Hu *et al* (2007) propose an approach motivated the law of iterated logarithm (LIL) that penalizes the Z-value of the test statistic to account for multiple tests across time in a cumulative meta-analysis for a binary endpoint planned prospectively or examined retrospectively. This paper extends the method to the analysis of parameter of interest to relative risk, odds ratio, or risk difference. When applied to random-effect model, the method can also account for heterogeneity in treatment effects across studies. The LIL method involves estimating an adjustment factor which is directly related to the control of type 1 error. A limitation of the LIL method is that since there is no closed form solution for estimation of the adjustment factor, it must be determined through extensive simulations under various conditions.

3.6 Visual Analytics versus Tabular Approaches

Safety data present many challenges with regard to analysis and interpretation. The very nature of safety data makes it challenging to analyze using conventional statistical methods because many standard assumptions may not be fulfilled. Additionally, a typical clinical trial is generally not sufficient to detect safety signals, unless a study is specifically powered for safety. Further, the pathological features of diseases lead to asymmetric non-normal distributions and heterogeneous subpopulations. Safety outcomes also have high variability in measurements and safety data are multidimensional and interrelated in nature with some safety endpoints of concern unknown prior to trial. The use of tabular outputs for safety data often results in large volumes of output leading to problems in generation, assessment, validation, assembly and last and worst of all comprehension and communication of key safety findings, leading to challenges in interpretation. As pointed by Wittes (1996): *A plethora of tables and graphs that describe safety may bury some true signal in a cacophony of numbers.* The simple descriptive summary tabular outputs and the review of individual patient data are rarely analytical.

It is well recognized that visual analytics present a useful alternative to tabular outputs for exploring safety data and present a great opportunity to enhance evaluation of drug safety. For example, Harrell (2005) pointed out that it is difficult to see patterns in tables and substituting graphs for tables can help increase efficiency of review. McKain, Jackson, Elko-Simms (2015) argued traditional case reviews and use of tables and listings is not sufficient for safety surveillance principles. Vlachos (2015) argued that despite their potential, graphics are an underutilized resource in safety analysis. Graphs can be used to aid in inference and communicating safety results and to help display large amounts of safety data coherently and maximize the ability to detect unusual features or patterns. They can also play a big role in facilitating communication of safety results with regulators, investigators, DMC, and other stakeholders. Visualization of safety data can help convey multiple pieces of information concisely and more effectively than tables. Graphical exploration can substantially improve information gain from safety data.

In the context of SMR, many variables are considered and many safety analyses are performed and as result it can be challenging to come up with a coherent understanding of the safety profile without the help of visualizations techniques. With this mix of data, visual analytics arguably play a central role in the ongoing evaluation of a drug both during study conduct in a blinded fashion and in surveying a set of studies – completed or ongoing. In order to get maximum gain from using visual analytics in SMR, a couple of considerations must be borne in mind. These include taking into account principles for data visualization, pausing to frame the right questions to interrogate safety data, and the use of visualization tools.

Principles for construction of graphs to aid safety interpretation of data have been discussed in the literature, for example, Duke, 2014, Duke *et al*, 2015. Some of these principles include considerations for graph content, communication, information, annotation, axes, and style. All these are important to ensure that we have good visualization and successful visualization of the data and especially in the context of safety monitoring. Successful visualization of data can best be summarized in the context of information, a story, goal, and a visual form, see for example,

<http://www.informationisbeautiful.net/2015/workshops-are-beautiful-learn-our-dataviz-process/>. Along the same lines of thought, data needs to be put into context with the concepts (safety profiles). Likewise the goal-detecting safety issues early and accurately and the form of visualization should match with the data and the idea of what an SMR does.

Some examples of visualizations of safety data that meet the definition of good visualization of safety data and that can be used for SMR can be seen in the following resources:

- CTSpedia website: www.ctspedia.org
- Gersonides: <http://www.gersonides.com/r/>
- Picture is worth a thousand tables: <http://www.elmo.ch/doc/life-science-graphics/>
- PhUSE: <https://github.com/phuse-org/phuse-scripts/wiki/Standard-Script-Index>

In order to effectively use visual analytics in SMR, it is a good idea to begin with some questions with regards to safety data under consideration. More specifically, SMR should be driven by asking the right question(s) of safety data. Some examples of questions associated with adverse events may include some of the following:

- What is the temporal relation of drug experience and exposure?
- Which AEs are elevated in treatment versus control?
- What is the constellation of AEs that come with the drug?
- Is there any evidence of a dose-response-relationship?
- Is the potential AE of interest increasing over time?
- Is there a difference in the time to the first event across treatment groups?
- What are the trends of time to the first event among different AEs?
- Which AEs are elevated in patient subgroups?
- What are the risk factors of the AE?
- Are there withdraws and/or interruption due to AE of interest?
- Is there a relationship with other AEs?
- What is the severity of the AEs?
- Are the most prevalent AEs suggestive of more serious events or medical concern?
- Is there a relationship with use of concomitant medications?
- For multiple studies, does a meta-analysis reveal a degree of heterogeneity of event across studies?
- Which AEs could be a safety signal? Are there any surprises in the data?
- Under what circumstances is the event most important to the patient?

When one considers the many questions that one can ask in the SMR setting to help in effective visualization of data and hence identify potential concerns, the more evident it becomes to see that visualization types and settings can fit many subcategories. These subcategories range from graph types, graph complexity, graph usage, graph information type, and static and dynamic aspects for the graphs. Ultimately, the safety question and graph type will dictate the right tool to use for SMR. There are many tools available that can be used to aid in visual analytics in SMR. These include R, R Shiny, Splus, Spotfire, SAS, JMP, JMP Clinical, Tableau, and J-Review, to mention a few. All these visual analytics tools have different functionality, and hence the choice of tool should consider tool functionality, for example, static versus interactive and/or dynamic visualization, drill down to patient level data, and the graph types most effective for SMR question.

Clearly, the safety question and ultimately the graph type will determine the choice of the visual analytics that will be used. Selection of the visual type or graph type may also be driven by the nature of the event in terms of AE Tier categories as defined in Crowe *et al* (2009). Figures 5.1 and 5.2 show graphs that may be used for Tier 1 and Tier 2 events.

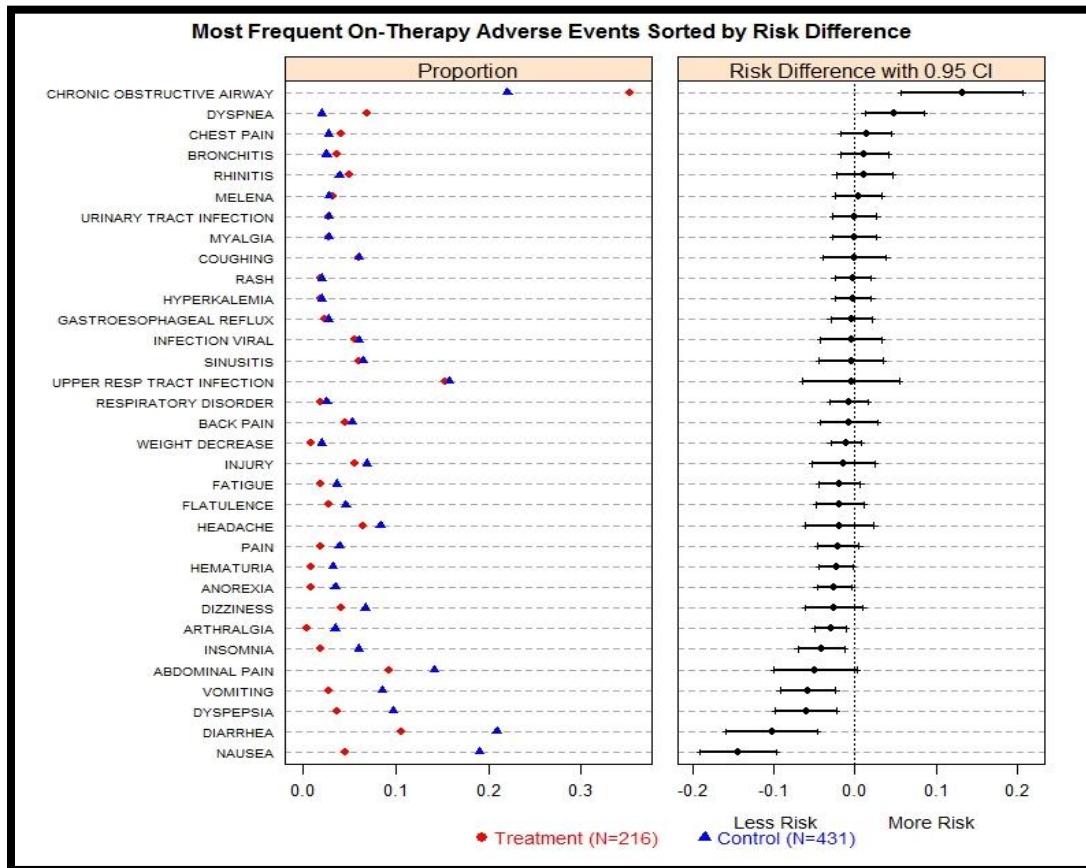


Figure 5.1: Graphical Presentation of Tier 1 AEs: Source: <https://www.ctspedia.org/do/view/CTSpedia/ClinAEGraph000>

In summary, visual analytics can help in SMR and in safety data analysis in general. Utilizing visualization tools can help exploration and can substantially improve information gain from SMR activities. Visualization should however take into consideration important principles of graph construction in order to render them useful in SMR. Ultimately, the visual type and tool used will depend on the question or questions under consideration in the SMR activity.

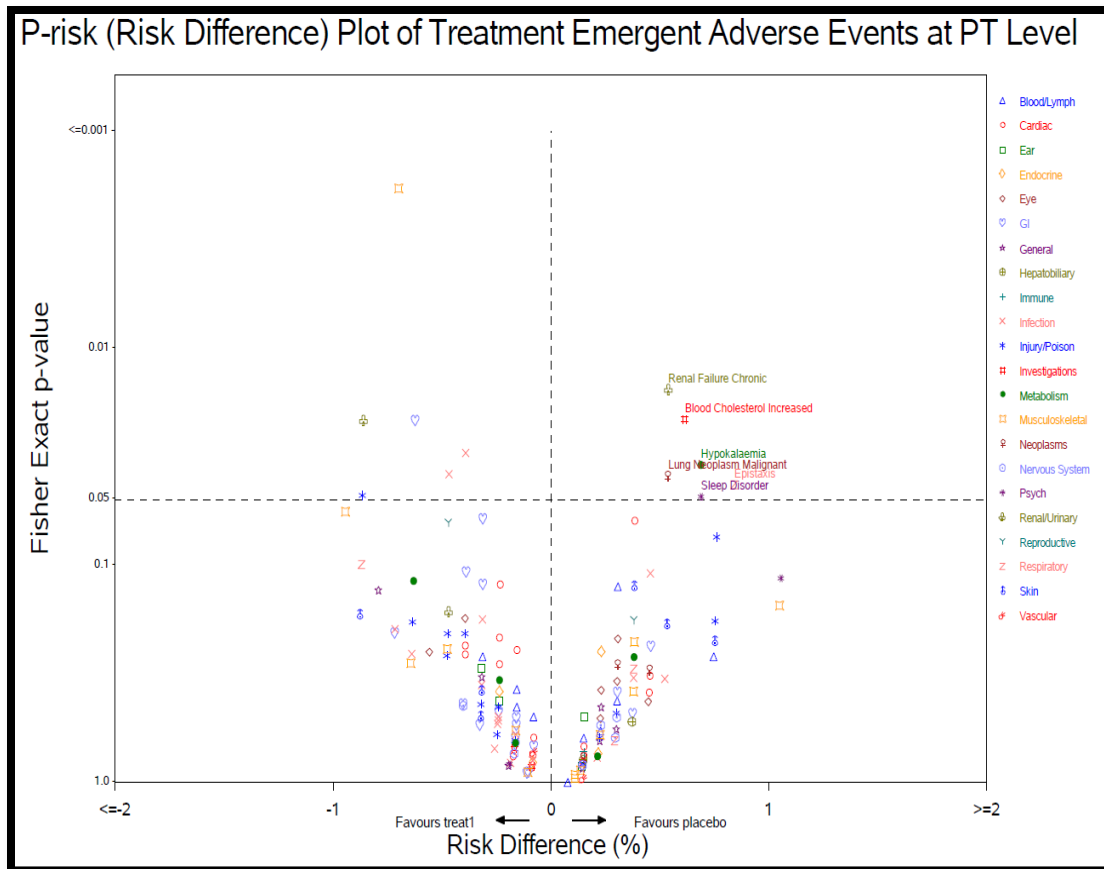


Figure 5.2: Graphical Presentation of Tier 2 AEs. Source: <https://www.ctspedia.org/do/view/CTSpedia/ClinAEGraph003>

4. Conclusions on SMR and Future Work

This paper provides an overview of statistical methods and strategies for answering drug safety questions in the pre-approval setting. The statistics community has developed many methods to address such questions in the post-approval setting and made significant headway in the development phases as well. However, much more remains to be done. The importance of safety monitoring reporting continues to increase as evidenced by the 2010 FDA IND safety reporting final rule and its subsequent 2012, 2015 guidance.

There are many methods for safety monitoring in the literature and we have highlighted only a sampling of the methods from one perspective among the many possible approaches. It is by no means an exhaustive list of what we have looked at or what is out there, but it portrays the thinking so far and the limits of beyond which the statistics community needs go beyond through greater partnering with the many other stakeholders in drug safety.

Within the field of statistics the multifaceted nature of SMR can be expressed by the analogy with the blind trying to understand the elephant. Bearing this in mind we:

- can proactively articulate the need to address different perspectives and stakeholders

- hope these different views provided provide some guidance in our own work on safety monitoring and reporting
- recognize that the ultimate pick, method-wise, will depend on the question at hand, or safety issue(s) focus and the context

This work in this paper represents a first phase of the ASA safety monitoring working group towards the goal of developing a systematic approach to safety monitoring and reporting. In the years ahead the biopharmaceutical industry will see an exponential increase in the need for clear and in depth understandings of the risk side of benefit-risk. Statisticians must play a key role in that growth of knowledge development but will do so only if fully committed and motivated to engage in the research and application of our field to safety and our products. The systematic review and perspective in this paper will lay a solid foundation for ASA safety monitoring working group's future efforts, both in method development and its applications.

Acknowledgements

We wish to thank our collaborators in the ASA Safety Monitoring Working Group from W1 on Safety Regulation, Thought-leader Interview and Industry Survey: Faiz Ahmad (Galderma), Greg Ball (Merck), Michael Colopy (UCB), Susan Duke (AbbVie), Robert Gordon (Janssen), Qi Jiang (Amgen), and Wenquan Wang (Morphotek). We also want to thank the Safety Monitoring Statistical Advisors: Aloka Chakravarty (FDA), Brenda Crowe (Lilly), Larry Gould (Merck), Qi Jiang (Amgen), Olga Marchenko (Quintiles), Janet Wittes (Statistics Collaborative), and Amy Xia (Amgen).

References

Introduction

Sarker, A. Ginn, R. Nikfarjam, A., O'Connor, K., Smith, K., Jayaraman, J., Upadhaya, T., Gonzalez, G. (2015), Utilizing social media data for pharmacovigilance: A review, *Journal of Biomedical Informatics*, **54**,202-212.

Bayesian and Frequentist Approaches

Benjamini, Y., Hochberg, Y. (1995), Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J RSS Ser B*, **57**, 289-300.

Berry, S., and Berry, D., (2004), Accounting for Multiplicities in Assessing Drug Safety: A Three-Level Hierarchical Mixture Model, *Biometrics*, **60**, 418-426.

Boland, K., and Whitehead, J., (2000) Formal approaches to safety monitoring of clinical trials in life-threatening conditions, *Statistics in Medicine*, **19**, 2899-2917.

Chen, W., Zhao, N., Qin, G., Chen, J. (2013), A Bayesian group sequential approach to safety signal detection, *J. Biopharm. Statist.*, **23**, 213-230.

DuMouchel, W. (2012), Multivariate Bayesian logistic regression for analysis of clinical study safety issues, *Statistical Science*, **27**, 319-339.

DuMouchel, W. (1999), Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. *The American Statistician*, **53**, 177-190.

Fayers, P., Ashby, D., Parmar, M. (1997), Tutorial in Biostatistics: Bayesian Data Monitoring in Clinical Trials. *Statistics in Medicine*, **16**, 1413-1430.

- Gould, L., (2008), Detecting potential safety issues in clinical trials by Bayesian screening. *Biometrics Journal*, **50**, 837-851.
- Gould, L. (2013), Detecting potential safety issues in large clinical or observational trials by Bayesian screening when event counts arise from Poisson distributions, *J. Biopharm. Statist.*, **23**, 829-847.
- Hu, N.; Huang, L. and Tiwari, R. (2015) Signal Detection in FDA AERS database Using Dirichlet Process, *Statistics in Medicine*, **34**, 2725-2742.
- Jennison, C. and Turnbull, B., (1993), Group sequential tests for bivariate response: Interim analyses of clinical trials with both efficacy and safety endpoints. *Biometrics*, **49**, 741-752.
- Kashiwabara, K., Matsuyama, Y., Ohashi, Y. (2014), A Bayesian stopping rule for sequential monitoring of serious adverse events, *Therapeutic Innovation and Regulatory Science*, **48**, 444-452.
- Mehrotra, D. V., Heyse, J. F. (2004), Use of the false discovery rate for evaluating clinical safety data, *Statistical Methods in Medical Research*, **13**, 227-238.
- Mehrotra, D. V., Adewale, A. J. (2012), Flagging clinical adverse experiences: Reducing false discoveries without materially compromising power for detecting true signals, *Statistics in Medicine*, **31**, 1918-1930.
- Resnic, F. S., Zou, K. H., Do, D. V., Apostolakis, G., Ohno-Machado (2004), Exploration of a Bayesian updating methodology to monitor the safety of interventional cardiovascular procedures, *Med. Decis. Making*, **24**, 399-407.
- Spiegelhalter, D., (2004), Incorporating Bayesian ideas into health-care evaluation, *Statistical Science*, **19**, 156-174.
- Tarone R. E. (1982), The use of historical control information in testing for a trend in proportions, *Biometrics*, **38**, 215-220.
- Thall P., Simon R. (1994), Practical Bayesian guidelines for Phase IIB clinical trials. *Biometrics*, **50**, 337-349.
- Weaver, J., Ohlssen, D., and Li, J., (2016), Strategies on using prior information when assessing adverse events, *Statistics in Biopharmaceutical Research*, **8**, 106-115.
- Xia, A. H.; Ma, H., Carlin, B. P. (2011), Bayesian hierarchical modeling for detecting safety signals in clinical trials, *J. Biopharm. Statist.*, **21**, 1006-1029.
- Yao, B, Zhu, L., Jiang, Q., Xia, H .A. (2013), Safety Monitoring in Clinical Trials, *Pharmaceutics*, **5**, 94-106.
- Zhu, L., Yao, B., Xia, H. A., Jiang, Q., (2016), Statistical monitoring of safety in clinical trials, *Statistics in Biopharmaceutical Research* (online).

Blinded versus Unblinded

- Ball, G., (2011), Continuous safety monitoring for randomized controlled clinical trials with blinded treatment information. Part 4: One method, *Contemp. Clin. Trials*, **32**, S11-S17.
- Ball, G, Piller, L. B. (2011), Continuous safety monitoring for randomized controlled clinical trials with blinded treatment information - Part 2: Statistical considerations, *Contemporary Clinical Trials*, **32**, S5-S7.
- Ball, G., Piller, L. B., Silverman, M. H. (2011), Continuous safety monitoring for randomized controlled clinical trials with blinded treatment information - Part 1: Ethical considerations, *Contemporary Clinical Trials*, **32**, S2-S4.
- Ball, G., Silverman, M. H., (2011), Continuous safety monitoring for randomized controlled clinical trials with blinded treatment information - Part 3: Design considerations, *Contemporary Clinical Trials*, **32**, S8-S10.

- Gould, A. L., Wang, W. W. B. (2015), Estimating event rate differences using data from blinded trials: The canary in the coal mine, *JSM* (2015).
- Gould, A. Lawrence, Wang, W. (2016). Monitoring potential adverse event rate differences using data from blinded trials: the canary in the coal mine, To appear *Statistics in Medicine*.
- Peace, K. (1987), Design, monitoring, and analysis issues relative to adverse events, *Drug Information Journal*, **21**, 21-28.
- Herson, J., Safety Monitoring, Chapter 11. In *Statistical Methods for Evaluating Safety in Medical Product Development*, by A. Lawrence Gould. New York; Wiley.
- P. M. Schnell, G. Ball. A Bayesian exposure-time method for clinical trial safety monitoring with blinded data. To appear in *Therapeutic Innovation and Regulatory Science*.
- Wen, S., G. Ball, G., and Dey, J., (2015), Bayesian monitoring of safety signals in blinded clinical trials, *Ann. Public Health Res.*, **2**, 1019.
- Zhu, L., Yao, B., Xia, H. A., Jiang, Q., (2016) Statistical monitoring of safety in clinical trials, *Statistics in Biopharmaceutical Research* (online).

Post-Marketing References

- Bate, A., Lindquist, M., Edwards, I. R., Olsson S., Orre, R., Lansner, A., De Freitas, R. M., (1998), A Bayesian neural network method for adverse drug reaction signal generation, *European J. Clin. Pharmacol*, **54**, 315-321.
- Chen, W., Zhao, N, Qin, G., Chen, J., (2013), A Bayesian group sequential approach to safety signal detection, *J. Biopharm. Statist.*, **23**, 213-230.
- Davis, r. I., Kolczak, m., Lewis, E., Nordin. J., Goodman, M., Shay, D. K., Platt, R., Black, S., Shinefield, H., Chen, R. T., (2005), Active surveillance of vaccine safety: A system to detect early signs of adverse events, *Epidemiology*, **16**, 336-441.
- DuMouchel, W., (2012), Multivariate Bayesian logistic regression for analysis of clinical studysSafety issues, *Statistical Science*, **27**, 319-339.
- DuMouchel, W., (1999), Bayesian data mining in large frequency tables, with an application to the FDA Spontaneous Reporting System, *The American Statistician*, **3**, 177-190.
- Gould, A. Lawrence (2008), Detecting potential safety issues in clinical trials by Bayesian screening, *Biometrical Journal*, **50**, 1521-4036.
- Gould, A. Lawrence (2008), Detecting potential safety issues in clinical trials by Bayesian Screening, *Biometrical Journal*, **50**, 837-851.
- Gould, A. Lawrence (2013), Detecting potential safety issues in large clinical or observational trials by Bayesian screening when event counts arise from Poisson distributions, *J. Biopharm Statist.*, **23**, 829-247.
- Hauben, M., Zhou X., (2003), Quantitative methods in pharmacovigilance: focus on signal detection, *Drug Safety*, **26**, 159-186.
- Hu, N., Huang, L., Tiwari, R. (2015), Signal detection in FDA AERS database using Dirichlet process, *Stat. Meth.*, **34**, 2725-2742.
- Huang, L., Zalkikar, J., Tiwari, R. C., (2011), A likelihood ratio test based method for signal detection with application to FDA's drug safety data, *JASA*, **106**, 1230-1241.
- Huang, L., Zalkikar, J, Tiwari, R. C., (2013), Likelihood ratio test-based method for signal detection in drug classes using FDA's AERS database, *J. Biopharm Statist.*, **23**, 178-200.
- Huang, L., Zalkikar, J., Tiwari, R. T., (2014), Likelihood ratio based tests for longitudinal drug safety data, *Statist. Medic.*, **33**, 2408-2424.
- Kajungu, D., Erhart, A., Talisuna, A., Bassat, Q., Karema, C., Nabasumba, C., Nambozi, M., Tinto, H., Kremsner, P., Meremikwu, M., D'Alessandro, U., Speybroeck, N.,

- (2014), Paediatric pharmacovigilance: Use of pharmacovigilance data mining algorithms for signal detection in a safety dataset of a paediatric clinical study conducted in seven African countries, *PLoS ONE*, **9**, e96388. doi:10.1371/journal.pone.0096388.
- Kulldorff, M., Davis, R. L., Kolczak, M., Lewis, E., Lieu, T., Platt, R., (2011), A maximized sequential probability ratio test for drug and vaccine safety surveillance, *Sequential Analysis*, **30**, 58-78.
- Li, L., Kulldorff, M., (2010), A conditional maximized sequential probability ratio test for pharmacovigilance, *Statist. Med.*, **29**, 284- 295.
- Poluzzi et al, Chapter 12 “Data Mining Techniques in Pharmacovigilance: Analysis of the Publicly Accessible FDA Adverse Event Reporting System (AERS).
- Shih, M. L., Lai, T. L., Heyse, J. F., Chen, (2010), Sequential generalized likelihood ratio tests for vaccine safety evaluation, *Statist. Medic.*, **29**, 2698-2708.
- Szarfman, A., Machado, S. G., O'Neill, R. T., (2002), Use of screening algorithms and computer systems to efficiently signal higher-than-expected combinations of drugs and events in the US FDA's spontaneous reports database, *Drug Safety*, **25**, 381-392.

Analyses from Patient Level Data to Meta-Analytic Evaluation

- Chuang-Stein, C., Beltangady, M., (2011), Reporting cumulative proportion of subjects with an adverse event based on data from multiple studies, *Pharmaceutical Statistics*, **10**, 3-7.
- Crowe, B., Chuang-Stein, C., Lettis, S., Brueckner, A., (2016), Reporting adverse drug reactions in product labels, *Therapeutic Innovation and Regulatory Science*.
- Crowe B. J., Xia, H.A, Berlin, J. A., Watson, D. J., Shi, H., Lin, S. L., et al (2009), Recommendations for safety planning, data collection, evaluation and reporting during drug, biologic and vaccine development: a report of the safety planning, evaluation, and reporting team, *Clinical Trials*, **6**, 430-440.
- Hu, M., Cappelleri, KJ. C., Lan, G. K. K. (2003). Applying the law of iterated logarithm to cumulative meta-analysis of a continuous endpoint, *Statistica Sinica*, **13**, 1135-1145.
- Hu, M., Cappelleri, KJ. C., Lan, G. K. K. (2007). Applying the law of iterated logarithm to control type I error in cumulative meta-analysis of binary outcomes, *Clinical Trials*, **4**, 329-340.

Visual Analytics

- Amit, O., Heiberger, R. M., Lane, P. W., (2008), Graphical approaches to the analysis of safety data from clinical trials, *Pharm Stat.*, **7**, 20-35.
- Amit, O., Heiberger, R. M., Lane, P. W., Graphical Approaches to the Analysis of Safety Data from Clinical Trials,
http://www.mbswonline.com/upload/presentation_Richard7-26-2009-21-42-29.pdf
- Duke, S. P., Bancken, F., Crowe, B., Soukup, M., Botsis, T., Forshee, R., (2015), Seeing is believing: good graphic design principles for medical research, *Statist. Medic.*, **34**, 3040-3059.
- Duke, S. P., Crowe, B., Forshee, R. F., Seeing is believing! Good graphic design principles for medical research:
https://www.ctspedia.org/wiki/pub/CTSpedia/GraphicsPresentationArchive/DIA2014_Susan_Duke_Graphics.pdf.

- Duke, S. P., Decision making and safety in clinical trials – graphs make a Difference!, https://www.ctspedia.org/wiki/pub/CTSpedia/GraphicsPresentationArchive/Duke_Susan_DIA11_v3.pdf.
- Duke, S. P., Bancken, F., Crowe, B., Soukup, M., Botsis, T., Forshee, R., Translating complex clinical trial and post-market safety data into visual stories: Seeing is believing: Good graphic design principles for medical research, <http://www.fda.gov/downloads/BiologicsBloodVaccines/ScienceResearch/UCM490777.pdf>.
- Harrell, F. (2005). Exploratory Analysis of Clinical Safety Data to Detect Safety Signals. <http://biostat.mc.vanderbilt.edu/wiki/pub/Main/FHHandouts/gksafety.pdf>.
- Krause, A., O'Connell, M. (eds), (2012), *A Picture is Worth a Thousand Tables: Graphics in Life Sciences*, Springer, New York.
- McKain, Jackson, T, Elko-Simms, C. (2016), Optimizing Safety Surveillance During Clinical Trials Using Data Visualization Tools. <http://www.dddmag.com/articles/2015/10/optimizing-safety-surveillance-duringclinical-trials-using-data-visualization-tools>.
- Vlachos, P. Graphical (and Analytical) Tools for the Systematic Analysis of Safety Data in Clinical Trials. <http://www.ssfa.it/allegati/3%20-%20SafetyVisualizationFINALVlachos.pdf>
- Wittes, J., (1996), A statistical perspective on adverse event reporting in clinical trials. *Biopharm. Report*.
- Zink, R. C., Wolfinger, R. D., Mann, G., (2013), Summarizing the incidence of adverse events using volcano plots and time intervals, *Clinical Trials*, **10**, 398-406.
- Zink, R. C., Huang, Q., Zhang, L. Y., Bao, W. J., (2013), Statistical and graphical approaches for disproportionality analysis of spontaneously-reported adverse events in pharmacovigilance, *Chin. J. Nat. Med.*, **11**, 314-320.