

*A Framework for Safety Evaluation  
Throughout the Product Development Life-  
Cycle*

**Greg Ball, Tjark Reblin, James  
Buchanan, Barbara A. Hendrickson,  
Eric Lewis, Patrick M. Schnell & Frank  
W. Rockhold**

**Therapeutic Innovation & Regulatory  
Science**

ISSN 2168-4790

Ther Innov Regul Sci  
DOI 10.1007/s43441-019-00021-5



**Your article is protected by copyright and all rights are held exclusively by The Drug Information Association, Inc. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at [link.springer.com](http://link.springer.com)".**



# A Framework for Safety Evaluation Throughout the Product Development Life-Cycle

Greg Ball<sup>1</sup> · Tjark Reblin<sup>2</sup> · James Buchanan<sup>3</sup> · Barbara A. Hendrickson<sup>4</sup> · Eric Lewis<sup>5</sup> · Patrick M. Schnell<sup>6</sup> · Frank W. Rockhold<sup>7</sup>

Received: 5 September 2019 / Accepted: 18 November 2019  
© The Drug Information Association, Inc 2019

## Abstract

Evaluation of the safety profile of medicines is moving from a more reactive approach, where safety experts and statisticians have been primarily focusing on the review of clinical trial data and spontaneous reports, to a more proactive endeavor with cross-functional teams strategically evolving their understanding of the safety profile. They do this by anticipating the ultimate benefit–risk profile and its related risk management implications from the start of development. The proposed approach is based on assessments of integrated program-level safety data. These data stem from multiple sources such as preclinical information; clinical and spontaneous adverse event reports; epidemiological, real-world, and registry data; as well as, potentially, data from social media. Blended qualitative and quantitative evaluations allow integration of data from diverse sources. Adding to this, a collaborative multidisciplinary view, which is focused on continuous learning and decision-making via diverse safety management teams, ensures that companies look at their growing safety database and associated risk management implications from every relevant perspective. This multifaceted and iterative approach starts early in the development of a new medicine, continues into the post-marketing setting, and wanes as the product matures and the safety profile becomes more well understood. Not only does this satisfy regulatory requirements but, crucially, it provides the healthcare system and treated patients with a better understanding of the drug's safety profile.

**Keywords** Interdisciplinary safety evaluation · Aggregate safety assessment · Framework · Iterative process · Learning and decision-making

## Introduction

Interdisciplinary safety evaluation is needed for learning and decision-making, for understanding the evolving product safety profile, and for ensuring effective risk management strategies. While mature marketed products have been judged to have benefits that outweigh their risks, a continued positive balance is less certain for investigational and newly marketed products. Consequently, it is important to identify, evaluate, and assess potential risks starting early in development in order to ensure maintenance of a favorable benefit–risk balance. We propose a framework for safety evaluation throughout the product development life-cycle, an iterative process of learning and decision-making by a team of safety and data scientists working with evolving data sources, which engages critical team members, supports the iterative nature of the process, and recognizes different data sources used across the life-cycle of the product. This framework for

---

✉ Greg Ball  
Greg.Ball@Merck.com

<sup>1</sup> Clinical Safety Statistics, Merck & Co, Inc, RY 34-A318, 126 E Lincoln Ave, Rahway, NJ 07065-4607, USA

<sup>2</sup> Drug Safety and Risk Management, Vifor Pharma Management Ltd, Glattbrugg, Switzerland

<sup>3</sup> Covilance LLC, Belmont, CA, USA

<sup>4</sup> Pharmacovigilance and Patient Safety, AbbVie, North Chicago, IL, USA

<sup>5</sup> Global Clinical Safety and Pharmacovigilance, GlaxoSmithKline, Research Triangle Park, NC, USA

<sup>6</sup> Division of Biostatistics, The Ohio State University College of Public Health, Columbus, OH, USA

<sup>7</sup> Duke Clinical Research Institute, Duke University Medical Center, Durham, NC, USA

safety evaluation, leveraging interdisciplinary expertise, can more efficiently define the product's safety profile and avoid premature termination of development programs by effectively managing risks or directing treatment to those subpopulations who demonstrate greater benefit and lower risk. Application of this process ensures that an acceptable benefit–risk profile is maintained throughout the product life-cycle.

### Important Differences Between Efficacy Analyses and Safety Assessments

Randomized clinical trials have been the accepted standard for addressing important questions in medicine since Bradford Hill actuated randomization [1] into clinical research over 70 years ago. To address specific questions, a safety event can be the primary endpoint in a properly powered clinical trial; however, while the structure and process of conducting these trials are well documented and characterized, historically they have principally addressed efficacy endpoints. Multiple reasons for this include the rarity of some safety events, the desire to look at safety events on a case by case basis, the sheer volume and diversity of safety events, and the lack of ability to predict some important safety events prior to the conduct of a trial. Safety assessment entails an amalgamation of developing information. While there could be a conclusion of no efficacy effect, there will not be a conclusion of no safety issues. What has evolved is a recognition that, while pre-specification of safety events of interest can be appropriate under exceptional circumstances, signal detection and evaluation is fundamentally different from efficacy evaluation. Safety assessments are generally underpowered to reach statistical conclusions, and different processes and procedures are needed to judge whether or not a risk is indeed drug related. However, the scientific evaluation of efficacy and safety data should be carried out with similar rigor in order to construct quantitative benefit–risk analyses and comparisons. As an example, inhaled corticosteroid-long acting beta agonist (ICS-LABA) combination inhalers reduce exacerbations of chronic obstructive pulmonary disease (COPD) but increase the risk of pneumonia among those being treated [2]. A quantitative approach could measure the ratio of the number of reduced serious exacerbations to the number of serious pneumonias triggered. Owing to the large number and assortment of safety events, assessing signals requires some deviation from the traditional significance testing and multiplicity thought-process that is treated with extreme strictness in the efficacy setting. We still need to apply rigorous scientific methods, but we must be willing to weigh evidence based on a collection of clinical as well as quantitative considerations.

### Proactive Safety Assessments can Enable Effective Risk Management

Complete elimination of risks will never be possible. Therefore, proactive evaluations of accumulating safety data are needed, so that decisions can be made based on benefit–risk assessments and the ability to manage important risks. Depending on the severity of the risk, one of the following three options can be considered. If a risk is unacceptable and cannot be mitigated with a black box warning, contraindications or other risk mitigation measures, the medicine would either need to be pulled from the market or, if still under development, the program would need to be terminated. If important risks can be attributed to distinct patient populations that are prospectively identified, for example, by testing for a specific biomarker, another option would be to contraindicate that medicine for those patients. A classic case is the testing requirement for HLA-B\*5701 in patients where treatment for human immunodeficiency virus (HIV) infection with abacavir is planned [3]. Patients testing positively are excluded from the treatment to avoid hypersensitivity reactions. A third option is to put risk minimization measures in place to safeguard patients while taking the drug. This can happen through regularly checking certain laboratory parameters before important clinical consequences occur (e.g., neutrophil count measurements before serious infections are acquired). Alternatively, the prescribing information can provide specific guidance on dealing with certain risks as part of the risk management approach. The newly approved treatment for hemophilia A, emicizumab, carries a warning of the risk for thrombotic microangiopathy and thromboembolism. Cases had been observed when cumulative doses above 100 U/kg/24 h of activated prothrombin complex concentrate (aPCC) were administered for 24 h or more. A warning highlights these risks in order to avoid such complications in the future by not dosing above that threshold.

### Aggregate Safety Monitoring and Scientific Evaluation of Integrated Safety Data

The evolving regulatory and drug development landscape is increasingly compelling aggregate safety monitoring and scientific evaluation of integrated safety data earlier in the development program as well as after marketing authorization. These ongoing aggregate safety evaluations are challenging for many reasons. Aggregate safety data analysis takes a holistic approach towards the evaluation of the safety profile of a medicine taking into consideration preclinical data, data from clinical trials and from

spontaneous reporting, as well as epidemiological data, and, increasingly, from new sources in clinical care settings and directly from patients. Data from a single program generated in multiple studies and real-world data sources can be stored in different data formats and in different databases, making combined analysis of integrated datasets difficult. This limits the utility of the data and, since regulators expect analyses of the complete database for a product, companies are putting increased effort into accessing their data in a much more integrated fashion. For example, Novartis employs digital technology and data analytics on a large scale to review clinical results of hundreds of studies completed over the past 20 years [4]. Roche, similarly, is leveraging its broad array of databases to explore for new scientific insights [5]. Although the primary focus of such initiatives has been on finding new and more efficacious treatment options for patients, these initiatives will, at the same time, help to better understand the evolving safety profiles of medications and to better target risk management strategies.

## Background

### Resolving Ambiguity in Productive Ways

Safety clinicians and statisticians, when confronted with uncertainty inherent in evaluation of safety data, approach the challenge in different ways. Drawing from the diagnostic work-up paradigm, clinicians look at data through the lens of biologic plausibility and personal clinical experience. However, the inherently limited dataset of personal experience introduces bias. A higher probability of a causal association is placed on events that match our expectations, known as the representativeness fallacy [6]. The human tendency to discern patterns where none exist and invent causal associations, the narrative fallacy [6], similarly biases data interpretation. There is also an underappreciation that statistical analyses in the context of safety data are an exercise in exploration and decision-making [7], as described by Tukey and colleagues, rather than the hypothesis testing that typifies efficacy endpoint analysis.

When interacting with clinicians and other decision-makers, statisticians are susceptible to oversimplifying problems to be mathematically convenient. For instance,  $p$  values, confidence intervals, and their counterparts in other inferential paradigms are founded on a set of assumptions, distributional and/or structural, that are often presented minimally or not at all. In underpowered situations, commonly encountered in safety monitoring, distributional assumptions carry heavy weight and usually cannot be adequately evaluated. This is not a problem that can be solved purely through more sophisticated statistical techniques: when a study is underpowered,

we cannot rely on sufficient events (or information) to provide robustness. Regardless, evaluations of structural assumptions, such as study design issues and subgroup comparisons, need to be made in collaboration with subject-matter experts.

Statistical education, best practice, and regulatory guidance dealing with pharmaceutical studies emphasize strict adherence to pre-defined analyses and decision rules, such as for stopping rules in group sequential trials. These requirements make evaluations of operating characteristics more feasible and are critical for efficacy analyses for which regulatory agencies insist on strong control of the Type 1 error rate for concluding superiority. This insistence is necessary to control the probability that an ineffective treatment reaches market, and feasible because efficacy endpoints are well-defined a priori. However, neither of these conditions applies to safety monitoring. While a Type 1 error (falsely identifying an adverse reaction) could have consequences (such as increased scrutiny of an adverse event that is not actually related to treatment), a Type 2 error (missing a real adverse reaction) could be more consequential (such as increased risk for a serious event). Additionally, it is not feasible to anticipate every possible safety concern, as they can be numerous, diverse, and often unknown. This is not a formal hypothesis testing paradigm (except, as mentioned before, when safety is the primary endpoint); however, the approach cannot be completely ad hoc either. It needs to be planned on an ongoing basis, as understanding of the safety profile grows, in order to have appropriate software, datasets, and programs available to suitably characterize the evolving safety topics of interest for robust evaluation, assessment, and decision-making. Thus, although any approach for aggregate safety monitoring should be systematic and quantitative, strict adherence to pre-defined analysis plans and decision rules is neither necessary nor practical.

When safety clinicians and statisticians collaborate on aggregate safety data evaluation, each can alleviate the limitations of the other and, instead, allow their collective strengths to make better decisions when assessing potential drug-related risks. While medical experts can identify the most important clinical questions, statisticians are poised to help frame the context of the analyses and interpret the results. Statistical methods impart objectivity to the exercise. However, the ascertainment of a causal association, for example as promulgated by Bradford Hill [8] or the CIOMS Working Groups III and V report [9], employs evaluation of factors beyond quantitative analysis to which clinicians can best contribute.

### Management and Reporting of Serious Adverse Events

Consistency in labeling information across International Conference on Harmonization (ICH) regions has improved

compared to labeling prior to the ICH harmonization process. For example, the first nonsedating antihistamine, terfenadine, was launched in the United States (US) in 1985. In June of 1990, the Food and Drug Administration (FDA) held its first advisory panel to review risk factors associated with the development of torsade de pointes due to the inhibition of CYP3A4. By 1992, the FDA had placed a Black Box Warning on terfenadine. Meanwhile in Japan, terfenadine was undergoing approval; sales began in 1991, but without a similar warning. Such a disparity is not likely to occur again, given the increased communication and knowledge sharing between the major drug development regions. Many pharmaceutical companies are currently filing in multiple countries and regions, basically simultaneously, and thus taking a much more global approach to describing the safety profiles of their products, including the associated risk management approaches.

In spite of harmonization, important differences among ICH regions remain for the management and reporting of serious adverse events to the respective regulatory authorities. Although both the US and the European Union (EU) follow the Suspected Unexpected Serious Adverse Reaction (SUSAR) reporting process [10], the US also operates under the Investigational New Drug (IND) safety reporting process [11, 12]. The primary difference between these two reporting structures is encountered during the sponsor causality assessment and reporting process. In the EU, all serious, unexpected, and possibly related adverse events (as determined by the investigator or the sponsor) are reported as SUSARs within the reporting timeframes outlined in ICH E2F (Development Safety Update Report) [13]. Expected events should only include those events where evidence of a causal relationship with study drug exists, based on aggregate safety data review from all previous and ongoing trials, which have been properly documented (for example, the Reference Safety Information (RSI) [14] in the current Investigator's Brochure (IB)). In the US, the FDA has instructed the sponsor to review all potential IND safety reports but to report only those that the sponsor considers possibly related to the drug [12, 15], i.e., there is evidence to suggest a causal relationship between the drug and the adverse event. The FDA distinguishes between "anticipated" events, that is, those that occur within the population being studied regardless of investigational intervention, and "expected" events, those that are known to be associated with investigational drug use. The FDA highlights limitations of single-case analysis and emphasizes the importance of aggregate analysis for determining "reasonable possibility" of an association with study drug for anticipated events. Events, such as myocardial infarction, stroke, and various cardiac arrhythmias, can have a background rate that is independent of study treatment. The more important question is not if such events occur, but if their rate of occurrence has been

increased because of the experimental treatment. A classic example of this occurred with the thiazolidinedione, rosiglitazone. Cardiovascular events are expected in patients with type 2 diabetes. At issue was whether an increase in the rate of these events was due to treatment with rosiglitazone [16, 17]. Sponsors must make judgments, based on accumulating results, about the threshold for IND safety reporting. The FDA expects a substantial reduction in the number of reports containing nonmeaningful safety information. Concerned about missing serious adverse reactions, EU regulators and regulators in other regions prefer to receive all SUSAR reports from sponsors, including those where the sponsor's causality assessment differs from the reporting investigator.

### Medical Judgment Within a Quantitative Framework

A paradigm shift has been occurring in the safety landscape: sponsors and regulators are placing more emphasis on aggregate safety assessments for determining association between a medicinal product and an adverse event [18]. An adverse drug reaction must have sufficient evidence to establish a reasonably possible causal association with study drug. Oftentimes, a single or a small cluster of events will not be sufficiently informative, with most reactions requiring an aggregate analysis to judge association. Sponsors need to supplement medical safety reviews of individual cases with regular reviews of accumulating aggregate safety data collected across the development program. As the database grows, medical review of individual cases can become overwhelming, but aggregate analysis will become more informative.

Global regulators have been strengthening requirements for aggregate safety monitoring and scientific evaluation of integrated safety data, requiring sponsors to regularly conduct program-level reviews. In Europe, RSI should be based on the comparative incidence of suspected adverse reactions in all previous and ongoing trials to establish reasonable evidence of a causal relationship for drug-event pairs [13]. In the US, to meet the spirit of the FDA IND Safety Reporting final rule [11, 12, 15], sponsors have developed methodologies and processes for evaluating and assessing accumulating safety information during development on an ongoing basis to determine causality. Sponsors have also been developing, evaluating, and implementing innovative procedures for review of aggregate blinded clinical trial data [19–22], minimizing the need to intentionally unblind data in ongoing studies. Regulators in both the US and EU currently support a more systematic approach to recognize, assess, and characterize important safety information, which will inform and improve risk management.

Methods for measuring evidence in the data need to be combined with a dynamic collaborative process for engaging with scientists from safety, clinical, epidemiology, and other

disciplines [6, 23]. To review accumulating program-level safety information across the drug development continuum, incorporating multiple data sources, sponsors have implemented product-specific safety management teams (SMTs) [24], enabling (1) a multidisciplinary approach, (2) quantitative frameworks for measuring level of evidence (based on rigorous data analysis, including artificial intelligence and machine learning) [25], and (3) decisions that are product-specific and driven by medical judgment [18]. Emphasis shifts from analyses with statistical testing and confirming (strict or implicit) to assessments with medical judgment and decision-making within a quantitative framework. A special challenge is the application of instructive statistical methodologies that employ proper meta-analytic techniques [26, 27] and appropriately incorporate subject-matter expertise [28, 29]. A systematic approach that combines state-of-the-art analysis of data, integrating clinical and quantitative perspectives, enables cross-disciplinary teams to develop clinical as well as statistical understanding of the safety profile and to drive the safety decision-making process with informed medical judgment. As John Tukey said, “far better an approximate answer to the right question, which is often vague, than an exact answer to the wrong question, which can always be made precise [30].”

## Discussion

### Quantitative Approaches for Analysis of Safety Data

Apart from review of individual case safety reports, safety clinicians seek evidence of emerging safety risks from the review of aggregate data with the assistance of statisticians. This task is less straightforward than statistical analyses of efficacy, which are predicated on hypothesis testing. Many aspects of a drug are known from preclinical and early phase studies and as such these targeted events can be pre-specified as efficacy endpoints (or possibly safety endpoints) in the protocol. For aspects of a drug's risk profile that are less well known or understood, a priori hypotheses are not available, and statistical evaluations take on an exploratory nature. The Pharmaceutical Research and Manufacturers of America (PhRMA) Safety Planning Evaluation and Reporting Team (SPERT) [31] have proposed three tiers for analyzing adverse events: (1) pre-specified events with specific hypotheses, (2) more frequent events that are not pre-specified, and (3) less frequent events that are not pre-specified. An alternative approach is to think of two tiers: (1) for learning—all events are evaluated in the general aggregate safety data review and (2) for decision-making—identified safety topics of interest, which can emerge or submerge throughout product development, are evaluated more closely. Except for studies designed to answer specific safety questions, safety assessment calls for a learning and

decision-making approach (scientific evaluation), rather than a testing and confirming approach (hypothesis testing).

Both individual clinical trial data and program-level safety information must be considered in assessing the emerging safety profile of the product. For statistical approaches with a safety mindset (with the understanding that assessment of safety differs from analysis of efficacy), we focus on two main points. First, constraints and objectives of safety evaluation procedures that make strict adherence to pre-specified analysis plans and decision rules are neither necessary nor feasible. Second, compared to information about efficacy, information about safety is sparse within any one trial, or even program, but is potentially more abundant in the totality of epidemiological and clinical studies.

The reduced emphasis on strict adherence to pre-specified analysis plans and decision rules facilitates the use of blended qualitative and quantitative approaches that may be less appropriate in efficacy analyses, but may offer a route to incorporate safety information from diverse sources in an assessment. For instance, information from pharmacovigilance databases such as the US FDA's Adverse Event Reporting System (FAERS), lacking reliable denominators for adverse event counts, cannot be directly incorporated into a meta-analysis of randomized trials (in any clear and obvious way) to produce a statistical estimate or test statistic. However, the discovery of a possible safety signal in a pharmacovigilance database through data mining techniques, such as described by Ibrahim et al. [32], qualitatively could add support to a similar signal under examination in an experimental program.

Even for quantitative methods of combining information from multiple sources, such as with proper meta-analyses of data from multiple studies, implementation requires decisions about the validity of studies and comparability of populations, interventions, comparators, outcomes, and timeframes. When performing stand-alone meta-analyses, analysts generally make better decisions as part of a collaborative process. As noted above, the same holds for reviews of safety information across an ongoing development program. Of course, the time and effort required to perform meta-analyses covering all possible safety issues before all trials begin would be prohibitive and overwhelming. However, while ad hoc meta-analyses confirming efficacy would not generally be permissible to perform as information accumulates, such meta-analyses are necessitated in response to possible safety signals.

### Partnership Between Clinical and Statistical Scientists

The Council for International Organizations of Medical Sciences (CIOMS) VI Report (Management of Safety

Information from Clinical Trials) [27] focuses on the collection, monitoring, analysis, interpretation, and communication of safety information from clinical trials. This report has been crucial for highlighting key aspects of clinical trial safety including the establishment of multidisciplinary SMTs, which are responsible for the timely review and assessment of accumulating, program-wide safety data. SMTs should include representatives from functional areas that are involved in evaluating the safety of a product such as pharmacovigilance, clinical development, statistics and regulatory affairs, as well as personnel with expertise in epidemiology, toxicology, and pharmacology. The focus of the SMT is on assessing safety data generated from ongoing and completed clinical trials in a program, evaluating safety signals that arise, and communicating and minimizing risk to clinical trial subjects.

The CIOMS VI report also draws attention to the overall development program and design of each clinical trial regarding the collection of safety information. Regulatory affairs representatives can provide information about regulatory expectations for the size of the safety database concerning subject numbers and duration of drug exposure to help enable a successful product filing [33, 34]; however, clinical trials for developmental products are generally focused on and powered for efficacy endpoints. Limitations of the safety database at the time of an initial product indication submission are well known and include limited numbers of exposed patients and constrained long-term follow-up for chronically administered drugs. This point is becoming increasingly important in the oncology space (especially immunoncology), where genetic and other biomarkers allow more personalized and targeted therapeutic approaches with better defined but, at the same time, smaller patient subpopulations. Also, shortening development timelines with condensing, for example, Phase 2 and Phase 3 studies into one set of studies, lessens patient exposure and reduces the amount of available safety data at launch. Issues may be raised about generalizing the data collected in the clinical trial population to patients for whom the product may be used in the post-marketing setting. In addition, study designs that do not include a comparator arm, or that involve switching treatment from control to study drug during an open-label extension or imbalances in the randomization ratio between experimental drug and comparator groups, all present challenges in the interpretation of the safety data subsequently collected. More systematic consideration is needed on the impact of clinical trial design, program-wide development, and high-level pooling strategy in the subsequent interpretation of the safety data collected.

To meet these challenges, the SMT needs to be proactive in developing an understanding of the epidemiology and safety topics of interest for the patient population under study. Critical epidemiological aspects include the

demographics of the study population, comorbidities of the disease under study, expected use of concomitant medications and their potential side effects, and anticipated background rates of events of interest, hospitalization, and death. This effort underscores the importance of the multidisciplinary nature of the SMT regarding the combined expertise of the clinicians, epidemiologists, and statisticians in interpretation of the clinical trial safety data. Furthermore, the SMT should ensure the consistent and comprehensive collection of safety data across studies, particularly for safety topics of interest. The objective of the analysis is to assess the causality of product administration in adverse events observed during the trials, to identify safety signals early, and to answer important safety questions as completely and comprehensively as possible at product filing. Clinical scientists are essential here for conveying the medical information that should be collected in order to adequately assess the emerging safety data and judge product causality. Moreover, they can provide perspective on the advisability of expert event adjudication against pre-specified medical concept definitions to enable appropriate classification of reported adverse events.

Without careful consideration for high quality and systematic safety data collection, in addition to a well-developed understanding of the impact of clinical study designs and study population epidemiology on safety data interpretation, any subsequent statistical analyses of the safety data would lack critical context. These qualitative considerations are indispensable for assessing the accumulating safety data. In addition, quantitative statistical techniques need to be applied with a safety rather than efficacy mindset. How measurements are being altered by the accumulating data (estimands) should also be part of the aggregate safety assessment planning process [35–37].

When interpreting safety data from ongoing and newly completed clinical trials, learning from available prior information is important. Data from clinical trials previously conducted in a similar patient population or related class of product may inform the assessment of the emerging safety data for the product under investigation. Also, data from earlier clinical trials in the development program may provide useful information in assessing newly generated product safety data. Combining this prior learning and statistical approaches to data assessment along with epidemiologic and clinical disease knowledge enables medical judgment and decision-making within a quantitative framework.

### Advantages/Disadvantages of Different Data Sources

The ICH E19 Concept Paper (Optimization of Safety Data Collection) [38] and FDA guidance (Determining the Extent of Safety Data Collection Needed in Late-Stage Premarket

and Postapproval Clinical Investigations) [39] discuss a more targeted approach to safety data collection in later stages of drug development to reduce clinical trial participant burden while still ensuring that patient welfare is not compromised. In order to achieve this objective, the safety profile of the drug should be characterized well during clinical development for identifying the important safety risks of the product and developing associated risk minimization measures. In addition, any remaining safety questions, which should drive additional data collection in the post-marketing phase to further characterize the product risk profile, need to be defined. More specifically, in light of target populations for drugs, particularly in the oncology space, becoming smaller and smaller owing to more targeted precision medicine approaches, strong emphasis must be placed early on to develop ideas and strategies that feed into solid risk management plans and that allow a continuation of the strategic evaluation of the safety profiles of products in the post-marketing phase. This is consistent with good pharmacovigilance practices (GVP) Module 5 (Risk Management Systems) [40].

Regulatory authorities often require post-marketing safety studies to be included in the risk management plan at product approval, as questions inevitably remain regarding the safety profile of the drug. Further characterization of the long-term safety of chronically administered products is a frequent requirement. Moreover, the incidence rate and risk factors for uncommon events may not have been adequately established at the time of approval. Rare adverse reactions may not have occurred in the drug development program, which includes a limited number of patients enrolled into the studies under specific inclusion and exclusion criteria. Different data sources can be used for clinical evidence regarding potential risks of a medicinal product, for regulatory decision-making, depending on whether the data are representative of the patient population and accurately capture the safety topics of interest [41].

Post-marketing safety studies using real-world data from population-based registries, claims databases, and electronic health records are important tools to further characterize the safety profile of the drug after approval [42–44]. Data sources must be selected that are appropriate to address the study objectives. Data quality must be assessed, potential biases investigated, and proof that these data sources are fitting, relevant, and useful should be in place before use in characterizing the safety profile. Study proposals must be evaluated to determine the validity of the obtainable outcome assessment and ability to control for significant confounding factors. Sufficient numbers of patients in the target population are needed to adequately assess the safety topics of interest. To put the generated product data in context, the availability of a suitable comparator group is crucial. For events of interest with long latency periods, retention

of patients in the database is needed to achieve the required follow-up. Methodologic considerations are important to manage bias such as confounding by indication or channeling [45, 46], in which the product is used more often in high-risk patient populations.

In contrast, with disease registries, much of the data is specifically gathered from healthcare providers and patients for the specific purpose of the registry rather than accessed from sources where the data are collected for other reasons (such as for billing purposes) [47, 48]. Consequently, disease registries offer the advantage of collecting data in a systematic manner for each patient. In addition, registries typically provide more detailed information about disease severity and employ targeted questionnaires for adverse events of interest in the relevant patient population. However, participation in such registries is voluntary and adequate numbers of patients need to be recruited to address the desired scientific inquiries. Also, data collected to answer a specific regulatory question in a particular clinical context may not be suitable for answering other questions.

Complex challenges exist for evaluating the relationship between a study drug and adverse events; however, the same principles that apply for assessment of accumulating safety data with randomized controlled trials apply for real-world data sources. Randomized clinical designs have moved medical science towards greater scientific rigor; at the same time, populations studied in clinical trials have diverged from patients in clinical practice. Randomization implemented in a clinical care setting could help bridge the evidentiary gap between clinical research and practice. This is a great opportunity for sponsors and academia to partner with regulatory authorities for developing “high-quality interoperable networks of data that can be seamlessly leveraged for clinical and research purposes” [49]. However, just as deriving evidence supportive of a causal association between a drug and event differs between clinical and post-marketing data sources, so too methodologies to support causal associations based on real-world data will need to be developed.

### **Continuum of Clinical Trial Safety Monitoring and Post-marketing Safety Surveillance**

A framework for the safety evaluation and assessment of a product is most useful when it engages critical team members, supports the iterative nature of the process, and recognizes the different data sources utilized across the life-cycle of the product. The success of the framework depends on recognizing and addressing the features of each of these three elements.

The SMT is best poised to identify safety signals, evaluate those signals for evidence of a causal relationship to the product, and develop risk management strategies, when there is close collaboration between the safety clinicians and data

scientists. Both qualitative and quantitative analyses are needed in this process, which is inherently exploratory in nature, differing substantially from the hypothesis testing techniques used in efficacy evaluation. Patterns in safety data that are due to random variation must be separated from actual causal associations. Accounting for confounding factors becomes an important exercise to isolate causal factors, which can be best performed with the assistance of knowledgeable biostatistical colleagues. However, until such time that an algorithm can reliably establish a causal relationship between a product and an adverse event, the decision process will require the expertise of both clinical and statistical scientists.

The assessment of safety data is an iterative process of learning and decision-making. The team of clinical and statistical scientists evaluates accumulating safety data and proposes identified and potential risks for the product, which may be subject to confirmation by a separate adjudication body, such as a DMC or SAC, which is optimally also staffed with experts in clinical and statistical science. As this process is repeated over the development period, the safety profile of the product emerges and is captured in living documents, such as the Development Core Safety Information and the Investigator's Brochure. At the time of marketing application, the process is repeated by reviewers at the regulatory authority who separately evaluate and, presumably, confirm the risks the sponsor has identified for the product. Upon marketing authorization, the iterative process of learning and decision-making continues with the sponsor as post-marketing data accumulate, leading to modifications of the Company Core Safety Information and product label, also with evaluation and confirmation by the regulatory authorities. This framework for an iterative process of learning and decision-making by a team of safety and data scientists working with evolving data sources can be supported by effective planning tools such as the Aggregate Safety Assessment Planning (ASAP) process, a product of the Drug Information Association (DIA)-American Statistical Association (ASA) Interdisciplinary Safety Evaluation (DAISE) scientific working group [50]. The ASAP process, building on the Program Safety Analysis Plan (PSAP) [31], recognizes the importance of the collaboration of clinicians and statisticians and the application of medical judgment in a quantitative framework.

Throughout the development and post-marketing periods, the sources of data will evolve, requiring the development of appropriate strategies and tools for their evaluation. Datasets from clinical studies, although often small, can be rich sources of information. In contrast, post-marketing data sources can be large in size, but with limited data elements available and potential biases. Actual frequency estimates from clinical trial data are replaced by reporting rates in

post-marketing databases. Additional data sources from registries, electronic healthcare records, and other post-marketing sources can provide real-world data but are often uncontrolled and require tailored analysis techniques. Across the life-cycle of the product, the range of data sources necessitates a variety of quantitative approaches so that the information can be appropriately viewed with the lens of clinical judgment.

## Conclusion

When a new product becomes widely accessible to patients, a core concern is whether that product's safety profile is adequately described in the label based on clinical trials performed during development and maintains its intended effects when in use. While mature marketed products have been judged to have benefits that outweigh their risks; a continued positive balance is less certain for investigational and newly marketed products. Consequently, it is important to identify, evaluate, and assess potential risks starting early in development in order to ensure maintenance of a favorable benefit–risk balance.

We have proposed a framework for safety evaluation throughout the product development life-cycle that is an integrated process of learning and decision-making by a team of experts, including clinical safety and quantitative data scientists, working with evolving data sources. This framework relies on a well-functioning team with diverse perspectives and expertise that supports the iterative nature and challenges of safety assessment across the life-cycle of the product. The focus of our approach is based on aggregate data assessment, as opposed to individual case evaluation, to determine if a causal association exists. This approach also allows identification of distinct patient populations to realize important health benefits of effective drugs (from the patient perspective). The quantitative framework involves measures of evidence in the data rather than hypothesis tests or decision rules. We believe this approach will produce quantitative assessments that are product-specific and decisions that are driven by medical judgment, facilitating consistent and authoritative communication of the safety story in scientific evaluations and public disclosures to healthcare providers and patients.

## Acknowledgements

We would like to acknowledge Jacqueline Corrigan-Curay, Bruce Binkowitz, Jay Herson, and Janet Wittes for helping us to refine our framework.

## Disclaimer

This article reflects the views of the individual authors and should not be construed to represent the views or policies of their companies.

## Funding

No funding sources.

## Compliance with Ethical Standards

### Conflict of interest

Dr. Schnell reports personal fees from Merck & Co, Inc, outside the submitted work. Dr. Rockhold reports grants from AstraZeneca, grants and personal fees from Janssen, personal fees from Merck Research Labs, personal fees from Merck Healthcare KGaA, personal fees from Novo Nordisk, personal fees from Rhythm, personal fees from KLSMC, personal fees from Aldeyra, personal fees from Complexa, personal fees from Sarepta, grants and personal fees from Eidos, grants from American Regent, grants from Reneuron, personal fees from Phathom, other from Athira, other from Spencer Healthcare, other from Datavant, grants from BMS, outside the submitted work.

## References

1. Yoshioka A. Use of randomization in the Medical Research Council's clinical trial of streptomycin in pulmonary tuberculosis in the 1940's. *BMJ*. 1998;317:1220–3.
2. Calverley PMA, Anderson JA, Celli B, et al. Salmeterol and fluticasone propionate and survival in chronic obstructive pulmonary disease. *N Engl J Med*. 2007;356:775–89.
3. Ziagen 300 mg film-coated tablets. Summary of product characteristics. [https://www.ema.europa.eu/en/documents/product-information/ziagen-epar-product-information\\_en.pdf](https://www.ema.europa.eu/en/documents/product-information/ziagen-epar-product-information_en.pdf). Accessed 27 Aug 2019
4. Novartis. 2018 Annual review: go big on data and digital. <https://www.annualreview.novartis.com/at-a-glance/what-we-do/strategic-priorities/go-big-on-data-and-digital.html>. Accessed 27 Aug 2019
5. Chainey J. CDISC US interchange 2018: CDISC standards in the real world. [https://www.cdisc.org/system/files/all/event/restricted/2018\\_US/1\\_Keynote\\_Chainey.pdf](https://www.cdisc.org/system/files/all/event/restricted/2018_US/1_Keynote_Chainey.pdf). Accessed 27 Aug 2019
6. Weisberg HI. *Willful ignorance: the mismeasure of uncertainty*. Hoboken: Wiley; 2014.
7. Tukey JW. *Exploratory data analysis*. Boston: Addison Wesley; 1977.
8. Hill B. The environment and disease: association or causation? *Proc R Soc Med*. 1965;58:295–300.
9. Council for International Organizations of Medical Sciences. Guidelines for preparing core clinical-safety information on drugs (second edition): report of CIOMS working groups III and V. 1999.
10. European Commission. Communication from the commission—detailed guidance on the collection, verification and presentation of adverse event/reaction reports arising from clinical trials on medicinal products for human use ('CT-3'). *Off J Eur Union*. 2011;172:01–2011.
11. US Food and Drug Administration. Final rule: investigational new drug safety reporting requirements for human drug and biological products and safety reporting requirements for bioavailability and bioequivalence studies in humans. September 2010.
12. US Food and Drug Administration. Safety reporting requirements for INDs and BA/BE studies. December 2012.
13. European Medicines Agency. ICH guideline E2F on development safety update report. September 2011.
14. Clinical Trial Facilitation Group CTFG. Q&A document—reference safety information. November 2017.
15. US Food and Drug Administration. Guidance for industry: safety assessment for IND safety reporting (draft). December 2015.
16. Rosen CJ. The rosiglitazone story—lessons from an FDA advisory committee meeting. *N Engl J Med*. 2007;357:844–6.
17. US Food and Drug Administration. Drug safety communication: FDA requires removal of some prescribing and dispensing restrictions for rosiglitazone-containing diabetes medicines. <https://www.fda.gov/drugs/drug-safety-and-availability/fda-drug-safety-communication-fda-requires-removal-some-prescribing-and-dispensing-restrictions>. Accessed 29 July 2019
18. Ball G, Kurek R, Hendrickson BA, et al. Global regulatory landscape for aggregate safety assessments: recent developments and future directions. *Ther Innov Regul Sci*.
19. Ball G, Piller L, Silverman M. Continuous safety monitoring for randomized controlled clinical trials with blinded treatment information. *Contemp Clin Trials*. 2011;32:2–10.
20. Schnell P, Ball G. A Bayesian exposure-time method for clinical trial safety monitoring with blinded data. *Ther Innov Regul Sci*. 2016;50(6):833–45.
21. Gould AL, Wang WB. Monitoring potential adverse event rate differences using data from blinded trials: the canary in the coal mine. *Stat Med*. 2017;36(1):92–104.
22. Mukhopadhyay S, Waterhouse B, Hartford A. Bayesian detection of potential risk using inference on blinded safety data. *Pharm Stat*. 2018;17(6):823–34.
23. Ball G, Lievano F. The importance of cross-disciplinary scientific engagement in the development of quantitative procedures for aggregate safety assessments. *Pharm Stat*. (online).
24. Council for International Organizations of Medical Sciences. Management of safety information from clinical trials: report of CIOMS working group VI. 2005.
25. Comfort S, Dorrell D, Meiris S, Fine J. Modified NARanjo causality scale for ICSRs (MONARCSI): a decision support tool for safety scientists. *Drug Saf*. 2018;41(11):1073–85.
26. Chuang-Stein C, Beltangady M. Reporting cumulative proportion of subjects with an adverse event based on data from multiple studies. *Pharm Stat*. 2011;10(1):3–7.
27. Crowe B, Chuang-Stein C, Lettis S, Brueckner A. Reporting adverse drug reactions in product labels. *Ther Innov Regul Sci*. 2016;50(4):455–63.
28. Wittes J, Crowe B, Chuang-Stein C, et al. The FDA's final rule on expedited safety reporting: statistical considerations. *Stat Biopharm Res*. 2015;7(3):174–90.
29. Duke SP, Kleoudis C, Polinkovsky M, et al. Quantitative methods for safety monitoring of rare serious adverse events. *Pharm Med*. 2017;31(2):113–8.
30. Tukey JW. The future of data analysis. *Ann Math Stat*. 1962;33(1):13.
31. Crowe BJ, Xia HA, Berlin JA, et al. 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. *Clin Trials*. 2009;6(5):430–40.
32. Ibrahim H, Saad A, Abdo A, Eldin AS. Mining association patterns of drug-interactions using post marketing FDA's spontaneous reporting data. *J Biomed Inf*. 2016;60:294–308.
33. International Conference on Harmonization. E1: the extent of population exposure to assess clinical safety for drugs intended for long-term treatment of non-life threatening conditions. October 1994.

34. US Food and Drug Administration. Premarketing risk assessment. March 2005.
35. Yang F, Wittes J, Pitt B. Beware of on-treatment safety analyses. *Clin Trials*. 2019;16(1):63–70.
36. Unkel S, Amiri M, Benda N, et al. On estimands and the analysis of adverse events in the presence of varying follow-up times within the benefit assessment of therapies. *Pharm Stats*. 2019;18:166–83.
37. Rockhold FW, Lindblad A, Siegel JP, Molenberghs G. University of Pennsylvania 11th annual conference on statistical issues in clinical trials: estimands, missing data and sensitivity analysis (morning panel session). *Clin Trials*. 2019;16:350–62.
38. International Conference on Harmonization. Final concept paper E19: optimization of safety data collection. July 2017.
39. US Food and Drug Administration. Determining the extent of safety data collection needed in late-stage premarket and postapproval clinical investigations. February 2016.
40. European Medicines Agency. Guideline on good pharmacovigilance practices (GVP) Module V: risk management systems. March 2017.
41. US Food and Drug Administration. Framework for FDA's real-world evidence program. December 2018.
42. Haque A, Daniel S, Maxwell T, Boerstol M. Postmarketing surveillance studies—an industry perspective on changing global requirements and implications. *Clin Ther*. 2017;39(4):675–85.
43. US Food and Drug Administration. Best practices for conducting and reporting pharmacoepidemiologic safety studies using electronic healthcare data. May 2013.
44. Ball R, Robb M, Anderson SA, Dal Pan G. The FDA's sentinel initiative: a comprehensive approach to medical product surveillance. *Clin Pharmacol Ther*. 2016;99(3):265–8.
45. Franklin JM, Eddings W, Austin PC, et al. Comparing the performance of propensity score methods in healthcare database studies with rare outcomes. *Stats Med*. 2017;36(12):1946–63.
46. Benedetto U, Head SJ, Angelini GD, Blackstone EH. Statistical primer: propensity score matching and its alternatives. *Eur J Cardio-Thoracic Surg*. 2018;53(6):1112–7.
47. Gliklich RE, Dreyer NA, Leavy MB, editors. *Registries for evaluating patient outcomes: a user's guide*. 3rd ed. Rockville: Agency for Healthcare Research and Quality; 2014.
48. Jonker CJ, van den Berg HM, Kwa MSG, Hoes AW, Mol PGM. Registries supporting new drug applications. *Pharmacoepidemiol Drug Saf*. 2017;26:1451–7.
49. Corrigan-Curay J, Sacks L, Woodcock J. Real-world evidence and real-world data for evaluating drug safety and effectiveness. *JAMA*. 2018;320(9):867–8.
50. Tremmel L, Hendrickson B, Buchanan J, et al. Joint statistical meetings (JSM) 2018: drug safety monitoring in a complex worldwide regulatory environment—strategy meets methodologies. <https://ww2.amstat.org/meetings/jsm/2018/onlineprogram/ActivityDetails.cfm?SessionID=215471>. Accessed 27 Aug 2019

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.