Summary
Biases and heterogeneity in the data of spontaneous adverse event reporting systems can make it hard to gain the insights to inform decisions. Capturing all possible indications of an adverse event may overcome these limitations, but involves extensive validation of multiple signals that may prove to be inaccurate. To streamline these processes, detection strategies can be constructed to probe a more defined association between an event and a drug. This hypothesis-driven approach builds on multi-sourced information and requires an information framework that unifies that information and aligns signal detection and validation.
An Open Perspective for Post-Marketing Surveillance

Since implementation of the first post-marketing monitoring measures aimed at improving patient safety, pharmacovigilance has made enormous strides in capturing information to better characterize the risks and benefits of medicinal products. At the core of efforts to find evidence of unexpected drug effects are reports of adverse events voluntarily submitted by healthcare professionals, patients and other consumers.

The value of tapping into these spontaneous reports is indisputable. By definition, clinical trials are targeted in scope and limited in length. By contrast, the evaluation of spontaneous reports opens an unrestricted, real-world perspective of the patient population and indication landscape in which the life cycle of a drug evolves. This allows marketing authority holders (MAH) and regulatory agencies to examine conditions and patient populations that are not covered in clinical trials prior to the release of a drug. Beyond highlighting instances where the drug may be associated with an adverse event, probing large databases of spontaneous reports contributes to improved drug labeling and optimized positioning of a medication so that it remains available to patients without endangering their safety. Importantly, spontaneous reporting systems are a “free voice” for end users. They are a window into the needs of the people served by the pharmaceutical industry and thus, provide invaluable feedback.

Use of spontaneous reporting data for pharmacovigilance is not without limitations. Heterogeneity in what is reported and biases in when or why reports are made can mislead and mask true indications of adverse reactions. Typically, mining these databases seeks to pick up as many clues as possible about adverse events (referred to as signals). These are then meticulously scrutinized to determine their validity.

Joyce de Langen, PharmD, Senior Product Solution Manager for QUOSA & Pharmacovigilance at Elsevier, explains: “Validation of potential signals is a very resource-intensive component of pharmacovigilance workflows. Commonly, the bulk of detected signals leads to no action and validation is a bottleneck. More importantly, inefficiencies in the process can lead to incorrect and untimely action, which can have a big impact on patient well-being.”

However, the workflow can be streamlined. Injecting information into the design of signal detection strategies helps to clearly define what constitutes a signal and the context in which it is detected. This generates less ambiguous input for further processing. The full adverse events landscape surrounding a drug can then be examined by repeatedly interrogating the database with a range of hypothesis-driven detection strategies.

Success depends on implementing an information framework that supports informing hypothesis generation, tracking multiple detection strategies, and aligning signal detection with signal validation schemes.
Unrestricted, but Cluttered

Widespread underreporting is often cited as a factor that curtails the analytical power achievable with data from spontaneous reporting systems. Voluntary submission of reports is a core feature of these systems but means that not all events suspected to be associated with a medicinal product are documented and brought to the attention of authorities and MAHs. This is especially true in cases where events are not considered severe. There are ways to deal with underreporting. Owners and managers of spontaneous reporting systems work to educate and incentivize reporters. Those who leverage these systems build several information sources into their surveillance mechanisms, such that spontaneous reports are just one of multiple touchpoints that reveal signals of possible adverse reactions.

A larger issue with detecting signals in data from spontaneous reporting is that while they represent an unrestricted view of the adverse events landscape, that landscape also happens to be very cluttered by data heterogeneity and biases. Some examples of that clutter include submission of several reports about a single event, incomplete information regarding an event or a patient’s medical history, variable language used to describe events and medicines, or biases in submissions due to reporter behavior and knowledge. Also contributing to that clutter is a missing point of reference. Calculated risk of an adverse event associated with a drug has meaning only when the same risk can be calculated in the population not taking the drug. The latter is not represented in spontaneous reporting and thus, selecting data that can serve as a meaningful reference is difficult and can dramatically impact outcomes.

Tapping into this cluttered landscape can generate signals that do not reflect reality, causing resource investment into validating an inaccuracy. At the same time, clutter can mask signals that are relevant to the comprehensive assessment of a medicinal product. Missing such signals can be detrimental to the marketability of the drug, not to mention dangerous for patients. So, how can signal detection be improved to maximize insights based on data from spontaneous adverse events reporting?

Read more about the challenges of probing large databases of spontaneous adverse event reports in this white paper "Signal Detection in Spontaneous Reporting Databases—Sentinels in a Cluttered Landscape."
Inclusion of a priori knowledge | Signal detection strategy | Outcome
--- | --- | ---
No | A. Hypothesis-free approach | Reporting odds ratio of adverse event associated with drug 1 increased after time point 4. High background noise contributes to large confidence intervals, making the signal difficult to detect.

At time point 4, a publication and news reports on an association between drug and the adverse event were released.

The increase appears to be the result of stimulated reporting from lawyers. The activity returns to normal after time point 9 and the altered reporting does not lead to validation of a signal.

After time point 4, physicians prescribe drug 2 along with drug 1 more often. Reports of the adverse event associated with drug 1 may not contain information about co-medication, and vice versa.

Reduced background no longer obscures increased reporting of adverse event associated with drug 1. A signal is detected and the hypothesis that co-medication may underlie the adverse event can guide subsequent signal validation.

Figure 1. Hypothesis-free versus hypothesis-driven interrogation of spontaneous adverse event reports. A hypothesis-free approach to signal detection (A) scans a database at regular intervals with broad search criteria and low detection threshold. The hypothesis-driven approach assimilates a priori knowledge and generates detection strategies (B and C) with narrower criteria and/or a defined background to probe specific aspects of the association between an event and drug. This can reveal confounding factors, decrease noise in the data used, and inform subsequent signal validation. Graphs show reporting odds ratio (ROR) plus 95% confidence intervals and represent an increase in reporting frequency after time point 4. An ROR with a 95% confidence interval above the neutral ratio 1 is considered a signal.
Different Measures of Disproportionality

Disproportionality is measured in various ways. The following ratios use as framework a two-by-two contingency table for a drug of interest and adverse event of interest. Each measure compares different cells of that table or calculates a different parameter.

Relative Reporting Ratio (RRR)
This measures the probability of an event in a defined group and compares it to the probability of the event in the total population.

Proportional Reporting Ratio (PRR)
This calculates the probability of an event in an exposed population and compares it to the probability in an unexposed population.

Reporting Odds Ratio (ROR)
This compares the odds of an event in an exposed population to the odds of the event in an unexposed population.

Making the Most of Invaluable Data

One approach to deal with the cluttered landscape is to guarantee a high chance of signal detection, i.e., search and capture any possible signal. Then, the resulting large bulk of signals is validated through a careful examination of evidence from an exhaustive assessment of medical literature, clinical trial and regulatory reports, internal and third party databases, and more. While this open, wholesale detection has the potential to pick up as many relevant signals as possible, subsequent validation is time- and resource-consuming. Conversely, this can cause delays and missteps in an environment where accurate, quick and informed decisions are essential to protect patients and safeguard useful medications.

Signal detection, however, can be enhanced via a secondary approach. With current advances in automation, statistics, informatics and machine learning, detection can be enriched through hypothesis-driven detection strategies that capture signals from spontaneous reports in a more precise manner. The hypotheses must emerge from facts; from accurate insights based on the integration of data from multiple and diverse sources. In contrast to sweeping the complete database in search for any new or different drug-event parings, this hypothesis-driven approach guides decisions about what qualifies as a signal and what constitutes the background in which it is detected (Figure 1). This approach increases specificity of detection and eliminates noise (or clutter) that can obscure signals in a less discriminating examination of data. However, it also narrows the scope of detection. Therefore, databases are probed repeatedly, each time with a different hypothesis-driven strategy. The sum of all interrogations, then, amounts to a broad yet still nuanced surveillance of post-marketing information on a drug.

Running the Specificity Spectrum

A hypothesis-driven approach that enables informed signal detection strategies is not necessarily meant to replace a sweeping, hypothesis-free approach. In fact, in many ways the two approaches are similar. Both require repeated data probing. Both improve data interrogation through iterative re-formulation of processes and detection criteria based on new information collected from multiple sources. Where the approaches differ is in the time point at which knowledge feeds into the workflow and in the objective of signal detection. While the hypothesis-free approach attempts to capture the gamut of potentially relevant signals and then injects information from multiple sources at a later stage to assess if a detected signal is real, the hypothesis-driven approach uses information from multiple sources to define a priori what constitutes a signal.

Implementing multiple hypothesis-driven strategies to detect signals builds flexibility into the monitoring process in terms of specificity. One can think of each detection strategy in the hypothesis-driven approach as a tool to zoom in on a specific portion of the monitored landscape with the goal to answer a defined question or explore a specific association. Used together, the hypothesis-free and hypothesis-driven approaches are complementary and enable covering a complete spectrum of detection specificities. At one end of the spectrum is the open detection afforded by the hypothesis-free approach: unspecific but highly inclusive. At the other end, targeted probing is defined by detection strategies of the hypothesis-driven approach that have tight signal inclusion criteria.

The Right Background Boosts Sensitivity

Background is essentially the noise against which a signal is detected. In disproportionality analyses, that background is commonly the reporting rate of adverse events for all drugs represented in a database, other than the one of interest. A signal is detected when reporting for the drug of interest is disproportionately higher than the background rate. “The data used to calculate the background plays a significant role in the sensitivity of detection,” says de Langen. “High background noise can mask an important disproportional signal; too low a background noise can flood the subsequent validation workflows with irrelevant ones. And yet, there is currently little consensus on this topic among authorities and MAHs.”
One benefit of the hypothesis-driven approach is that it not only primes detection of specific signals, it also lays out a meaningful definition of background. Each detection strategy builds on an informed hypothesis that defines a signal and a context for detection. That context is the background. Calculating a clearly defined background can increase sensitivity of detection. It decreases the proportion of false positive signals by focusing on a specific aspect of the relationship between medicinal product and adverse event, thereby uncluttering the landscape in the specific context of the tested hypothesis.

Another benefit is that background measures need not emerge from spontaneous adverse events reports. Other sources contain data that can be leveraged to disambiguate or fill in gaps in spontaneous reports data. For example, the U.S. Center for Disease Control and the U.S. Environmental Protection Agency generate large amounts of statistics that may be a more accurate picture of the background against which a signal is compared. By informing detection strategies with data from such sources a priori, signals detected from spontaneous reports can be better matched to a more relevant and accurate background.

Some questions to consider when defining background are:

- Is the examined event symptomatic of or correlated with other adverse events?
- Are there co-medications to consider in calculating background?
- Are there patient subpopulations that should be examined separately because event incidence correlates with age, location, or other parameters?
- Can data from a different source be used to calculate incidence of the event for the population not taking the examined drug?
- Do the data to calculate background take language variability into account?

The Devil is in the Details

The relationship between an adverse event and a medicinal product is often complex. Causality can be influenced or confound by various factors. Teasing apart this complexity is typically relegated to validation steps, where signals are scored based on evidence from as many sources as possible to support or negate a causal association. Aspects of this disentanglement can be shifted to the front end of signal detection. Building details into detection strategies allows, for example, examining co-medications as a possible confounding factor even though these are not always communicated in adverse event reports. Information about standard treatment protocols and co-occurrence of diseases gleaned from drug labeling, clinical studies and primary literature can be used to construct auxiliary detection strategies that detect signals of the same adverse event in patients taking commonly co-medicated drugs. Then, comparing results with signals connected with the drug of interest helps assess the true nature of the presumed connection between event and medication.

The same attention to detail can help disentangle complexity in the data itself. Redundancy in drug names, for example, can impact signal detection. The omission of synonyms may uncover only a portion of relevant cases which might render a signal insignificant. The opposite may also happen. Including only a subset of cases in proportionality calculations can decrease the number of patients that did not experience the event. As a result, a small number of adverse events may suddenly become significant. Systematically including drug name synonyms in signal detection overcomes data variability. Information systems that include synonyms and map adverse events to a defined taxonomy like MedDRA (1) are instrumental to comprehensively inform signal detection. Going one step further enables more nuanced insights: knowledge about the structure and mechanism of action of active ingredients in a drug allows probing other medications that operate on the same targets or have the same outcome to see if they exhibit similar adverse event patterns.
A Unifying Information Framework

The hypothesis-driven approach takes advantage of the accumulating abundance and diversity of information that has become easier to access through increasingly sophisticated data systems. This iterative interrogation of spontaneous reports databases, where each probe seeks signals that meet specific criteria within a well-defined context, must integrate accurate and reliable information from a broad disciplinary scope. The action of a drug, the biology of the disease it treats, the documented pharmacology and toxicology of its active ingredient and similar compounds, patterns of adverse events mined from preclinical, clinical and post-marketing safety data in regulatory approval documents are all examples of knowledge that should inform hypothesis-driven detection strategies. Systems that mine medical, regulatory and research literature, knowledge bases that predict drug-drug interactions, and platforms that facilitate the exchanging, handling and processing of information across multiple and diverse sources are an imperative.

Technology-based solutions already play an essential role in supporting the information needs of pharmacovigilance workflows. Diverse tools facilitate periodic report submissions, enable management of regulatory information, grant access to regulatory authority or large-scale public safety databases (like spontaneous reporting systems), support risk management activities, simplify literature monitoring, expedite finding chemical and biological knowledge, and streamline processes with mobile apps and cloud computing. More unorthodox sources, like digital and social media, information exchange platforms for patients and physicians, and anonymized electronic health records are increasingly being mined to gain insights and will also become standard in pharmacovigilance.

While access to information is essential, success of the hypothesis-driven approach to signal detection rides on three additional aspects. First, the process is an iterative and repeated interrogation of databases. Second, each detection strategy leverages knowledge garnered from multiple sources and integrated into a single, meaningful hypothesis. Finally, each strategy is adapted and improved as new relevant information emerges. This means that above and beyond getting information, implementing the hypothesis-driven approach demands aligning synergistic information from different sources.

This alignment is two-tiered. It begins with constructing intelligent hypotheses to drive signal detection strategies. Then, at a second level, what informs detection must also align with the queries of signal validation. Only systematic examination, gathering, processing and integration of information can accomplish this alignment. de Langen describes what technology must accomplish to better support systematic, hypothesis-driven signal detection in pharmacovigilance: “Identifying significant information connections in a diverse landscape of data types and sources requires a comprehensive view and understanding of that landscape. But that is only half the picture. Leveraging those connections requires the means to maintain an overview of what information is relevant and of how it is implemented.”

The information framework that supports a hypothesis-driven approach is an overarching platform that unifies information sources, but also:

- Powers the development of complex and cross-source search strategies
- Tracks signal detection criteria, supporting documentation, outcomes and iterations
- Facilitates the triage of raw data entering the pipeline and automatically advises of new relevant information
- Coordinates information exchange along the pharmacovigilance pipeline so that the right questions drive the identification of safety risks
“Information frameworks that support hypothesis-driven signal detection afford a clearer path to insights from spontaneous reporting systems.”

-Joyce de Langen, PharmD, Senior Product Solution Manager, Elsevier

**Powerful Search Strategies**

Monitoring information for pharmacovigilance can be a daunting task, especially if relying on manual searches with results stored and managed in spreadsheets or standard reference management systems. The risk of incomplete information assessments can be high under these circumstances. Heterogeneity of language and data types further complicates finding and tracking relevant references, as multiple synonyms and related search terms must be considered to achieve full coverage of a topic. While harmonization efforts, such as MedDRA coding of adverse events reports, have already boosted discoverability of relevant information, insights from any source are only as good as the telescope used to find them. An information framework that incorporates authoritative dictionaries and thesauri and supports mapping search criteria from one source to another greatly facilitates monitoring and integrating diverse data. The outcome is a streamlined search workflow with improved accuracy and comprehensiveness of results.

**Always Know What You Are Learning**

With the hypothesis-driven approach, signal detection involves tracking results from multiple detection strategies at any time point. A comprehensive overview of implemented strategies and the hypotheses they tackle helps ensure that the sum of all strategies affords detailed but extensive coverage of the adverse events landscape surrounding a drug of interest. Signal detection strategies should be coordinated and complementary. They should eliminate potentially confounding factors or other ambiguities to highlight causal associations between medicinal products and adverse events. Furthermore, detection strategies evolve over time as new and more information feeds into the pipeline. A time course of that evolution can be informative in constructing new detection strategies. The right information framework for the hypothesis-driven approach maintains an overview of developed and implemented signal detection strategies. A profile of each detection strategy tracks what information was used in constructing its underlying hypothesis, how it fits with other detection strategies, and how it has been adapted with successive iterations. The goal is that stakeholders in the pharmacovigilance process know at any time what they are learning from probing spontaneous reports data.

**Informed and Current**

Information from a broad scope of disciplines and sources is at the core of creating new and adapting existing detection strategies, but processing that information is challenging. Automated alerts about the availability of new references or data that could play a role in assessing the risks and benefits of a drug are a first step to facilitate finding and capturing important input. However, triage of raw data can be streamlined by an information framework that conducts additional processing, such as text-mining literature to highlight information relevant to adverse events, or excerpting data about the properties of known compounds. Expert data curation built into an information framework generates a more detailed understanding of potential connections between a medicinal product and adverse events that can be examined, explored and expanded through hypothesis-based detection strategies.

**The Right Questions from Start to Finish**

Ultimately, the goal of each hypothesis-driven signal detection strategy is to deliver potential signals of adverse events that can be validated along a specific investigative path. The signals that emerge from spontaneous reports based on these strategies meet very specific definitions, which strengthens validation by providing an evaluative context. The questions asked to validate signals are guided by the hypotheses underlying the detection strategies. At the same time, outcomes of each investigative path during validation can and should inform new signal detection strategies. One outcome can lead to a new perspective that should be explored in spontaneous reports data. Another investigative path may highlight unanswered questions that can be tested in spontaneous reports data with another hypothesis-driven detection strategy. An information framework that spans the pharmacovigilance pipeline allows integrating feedback loops to support this exchange where detection informs validation and vice versa.
More Data, but a Clearer Picture

There is no question that pharmacovigilance efforts have the responsibility to maximize insights from the growing body of information relevant to the performance of a medicinal product in the market. Leveraging this information effectively safeguards the utility of medicines, the safety of patients, and the contribution of pharmaceutical companies to shaping a healthcare system that meets today and tomorrow’s demands.

At first instance, managing the sheer volumes of data poses an arduous undertaking. In spontaneous reports databases, the mass of information is additionally complicated by heterogeneity of data that clutters the landscape with biases, misleading connections, and noisy background. With the right information solution, however, it is possible to disentangle some of that complexity. de Langen explains: “A platform that supports generating and tracking hypotheses derived from integrating multiple information sources can guide the detection and subsequent validation of potential signals. This guidance can streamline pharmacovigilance workflows and create transparency to better align the efforts of all parties involved in guaranteeing patient safety. Complementing a broad, sweeping collection of possible signals, information frameworks that support hypothesis-driven signal detection afford a clearer path to insights from spontaneous reporting systems.”

REFERENCES

1. MedDRA®, the Medical Dictionary for Regulatory Activities terminology is the international medical terminology developed under the auspices of the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH). www.meddra.org.
PharmaPendium informs critical drug development decisions on safety and efficacy, risk assessments and mitigation and study designs with fully searchable FDA and EMA drug approval documents and FAERS data, a drug-drug interaction risk calculator and comparative safety, pharmacokinetic, efficacy, and metabolizing enzyme and transporter data.

For more information about PharmaPendium, visit elsevier.com/solutions/pharmapendium-clinical-data

Elsevier offices

ASIA AND AUSTRALIA
Tel: +65 6349 0222

JAPAN
Tel: +81 3 5561 5034

KOREA AND TAIWAN
Tel: +82 2 6714 3000

EUROPE, MIDDLE EAST AND AFRICA
Tel: +31 20 485 3767

NORTH AMERICA, CENTRAL AMERICA AND CANADA
Tel: +1 888 615 4500

SOUTH AMERICA
Tel: +55 21 3970 9300