Drug Safety
https://doi.org/10.1007/s40264-025-01548-3

CURRENT OPINION

Interplay of Spontaneous Reporting and Longitudinal Healthcare Databases for Signal Management: Position Statement from the Real-World Evidence and Big Data Special Interest Group of the International Society of Pharmacovigilance

Salvatore Crisafulli[1] · Andrew Bate[2,3] · Jeffrey Stuart Brown[4,5] · Gianmario Candore[6] · Rebecca E. Chandler[7] · Tarek A. Hammad[8] · Samantha Lane[9,10] · Judith Christina Maro[5] · G. Niklas Norén[11] · Antoine Pariente[12,13] · Mulugeta Russom[14,15] · Maribel Salas[16,17] · Andrej Segec[18] · Saad Shakir[9,10] · Andrea Spini[1] · Sengwee Toh[5] · Marco Tuccori[1] · Eugène van Puijenbroek[19,20] · Gianluca Trifirò[1] on behalf of the Real-World Evidence and Big Data Special Interest Group of the International Society of Pharmacovigilance

Accepted: 25 March 2025

Abstract
Signal management, defined as the set of activities from signal detection to recommendations for action, is conducted using different data sources and leveraging data from spontaneous reporting databases (SRDs), which represent the cornerstone of pharmacovigilance. However, the exponentially increasing generation and availability of real-world data collected in longitudinal healthcare databases (LHDs), along with the rapid evolution of artificial intelligence-based algorithms and other advanced analytical methods, offers a wide range of opportunities to complement SRDs throughout all stages of signal management, especially signal detection. Integrating information derived from SRDs and LHDs may reduce their respective limitations, thus potentially enhancing post-marketing surveillance. The aim of this position statement is to critically evaluate the complementary role of SRDs and LHDs in signal management, exploring the potential benefits and challenges in integrating information coming from these two data sources. Furthermore, we presented successful cases of the interplay between SRDs and LHDs for signal management, along with future opportunities and directions to improve such interplay.

1 Introduction

Medicines are authorized following a positive benefit/risk assessment of data from clinical development programs; however, at the time of marketing, uncertainties remain about rare or delayed adverse drug reactions (ADRs) as well as drug–drug interactions, owing to the inherent limitations of pivotal clinical trials. With the aim to address these uncertainties, pharmacovigilance activities are legally required in many countries to further characterize the safety profile of medicines in the real-world settings.

Pharmacovigilance is the discipline involved in detection, assessment, characterization, and prevention of risks related to medicines. With the aim of identifying new risks as early as possible and re-evaluating the benefit/risk balance of medicinal product use in routine care, pharmacovigilance systems are aimed at the early detection of safety signals, a concept described in detail by CIOMS group VIII [1]. These signals are then assessed, and actions for minimizing risks are taken. The process from signal detection to the recommended actions is known as signal management, which involves multiple steps and stakeholders. According to EU legislation (IR 520/2012), the signal management process includes signal detection, validation, confirmation, analysis and prioritization, assessment, and recommendations for action [2]. Other regulatory agencies follow a similar signal management pathway despite using different terminologies for each phase.

The initial step of signal detection involves highlighting the associations that are suggestive of a causal relationship between the exposure to a medicinal product and suspected adverse events. These safety signals may arise from multiple sources, mostly spontaneous reporting databases

Published online: 13 April 2025

Key Points

Signal management encompasses activities from detection to actionable recommendations, traditionally relying on spontaneous reporting databases, which are crucial to pharmacovigilance.

The exponential growth of data collected in longitudinal healthcare databases, together with advanced analytical tools, offers significant opportunities to enhance signal management by complementing evidence from spontaneous reporting databases.

Combining evidence generated from spontaneous reporting databases and longitudinal healthcare databases can address the individual limitations of these sources, potentially improving post-marketing surveillance, and providing a framework for evaluating benefits, challenges, and successful use cases in signal management.

(SRDs) as well as patient registries, clinical trials, and/or pharmacoepidemiologic studies.

Beyond the analysis of the SRDs, as part of the signal management, assessors generally collect pre-clinical, clinical, biological/pharmacological, and epidemiological data that could support the causal relationship or otherwise suggest an alternative explanation for the suspected safety signal. On the basis of the gathered information, causality assessment in case series may be conducted considering the totality of evidence, based on criteria such as the Bradford–Hill criteria that include the systematic evaluation of the strength of association, temporality, specificity, consistency, biological plausibility, dose–response relationship of the association, coherence, and analogy of the association, as well as other aspects of the evidence, such as dechallenge and rechallenge [3, 4]. In this regard, real-world data (RWD), defined as routinely collected data concerning patient health status and delivery of healthcare services [5], and more specifically longitudinal healthcare databases (LHDs), can contribute to signal management by providing extensive patient data over long time-periods and allowing the real-world assessment of a medication's benefit–risk profile.

The distinct design purposes of SRDs and LHDs highlight their focus on different areas. In addition to SRDs and LHDs, other sources of RWD include social media and person-generated health data (PGHD) via wearables and mobile devices [6]. However, such data sources are outside the focus of this paper as currently they are not consistently and routinely used in the context of drug safety signal management.

The aim of this position statement is to delve into the synergistic and complimentary role of SRDs and LHDs in the different phases of signal management, exploring the opportunities and challenges in integrating information coming from these two types of data sources. Successful cases of the interplay of SRDs and LHDs for different phases of signal management are also presented.

2 Key Features of Spontaneous Reporting and Longitudinal Healthcare Databases in the Context of Signal Management

SRDs and LHDs are two important data sources for pharmacovigilance. SRDs (e.g., the World Health Organization global database of adverse event reports for medicines and vaccines [VigiBase], the US Food and Drug Administration Adverse Event Reporting System [FAERS database], the European pharmacovigilance database [EudraVigilance], and the Japanese Adverse Drug Event Report [JADER]) have the specific scope of monitoring safety of medicinal products. They are generated by collecting reports of suspected adverse reactions of medicines and vaccines submitted by healthcare professionals, consumers, caregivers, the pharmaceutical industry, and others, and can be managed by local, national, or international organizations. In addition, there are databases created and managed by the pharmaceutical industry limited to medicinal products for which they are marketing authorization holders. Table 1 reports examples of SRDs that are frequently used for pharmacovigilance activities.

LHDs encompass electronic health records (EHRs, containing comprehensive health information including medical history, drug prescriptions, diagnostic test results, and other data from either primary or secondary care), administrative healthcare databases (containing data about billing and claims for healthcare services provided to citizens), and patient/drug registries (e.g., longitudinal data collection of specific information for a population defined by a particular disease, condition, or exposure).

EHRs and claims databases are among the most widely used LHDs for conducting real-world studies. Each data source has distinct strengths and limitations, making them suitable for different types of research. Compared with claims databases, EHRs generally provide richer clinical detail and patient-level clinical information that is often missing in claims data. EHRs also offer a more comprehensive view of patient care, capturing prescribed

medications, disease severity, and clinical outcomes such as biomarkers and treatment responses. In addition, they enable more granular tracking of adverse events and disease progression. Conversely, claims databases excel in providing standardized and complete longitudinal records of healthcare utilization, particularly for billing, reimbursement, and cost-related research. They cover large populations over extended periods, ensuring better continuity of patient records across different healthcare settings. Moreover, because claims data are structured and systematically recorded for reimbursement purposes, they tend to be more complete and less prone to missing data compared with EHRs, which may suffer from documentation variability and fragmentation across healthcare providers [7]. Table 2 reports examples of LHDs that are commonly used for pharmacoepidemiologic studies.

LHDs can contribute to large national or multinational networks that allow access to large scale harmonized data [8]. Valuable examples of major distributed data networks used for real-world evidence generation and health data analysis are represented by the Data Analysis and Real-World Interrogation Network (DARWIN EU) in Europe [9, 10], and the Sentinel System in the USA [11]. DARWIN EU is the European medicines regulators' network of RWD sources, expertise, and services. Its establishment began in 2022, following the recommendations of the EMA-HMA Big Data Task Force for a network of databases allowing timely analysis of real-world healthcare data at the scale required to support any decision taken on medicines [12, 13]. The Sentinel System is an electronic system that assesses the safety and effectiveness of medical products, such as drugs, vaccines, biologics, and medical devices funded by the US Food and Drug Administration (FDA) [11]. Other distributed data networks funded by regulatory agencies are the Canadian network for observational drug effect studies (CNODES) (https://www.cnodes.ca/) in Canada, and the Medical Information Database Network (MID-NET) developed by the Japanese Pharmaceuticals and Medical Devices Agency (PMDA, https://www.pmda.go.jp/safety/mid-net/0001.html) [14].

Both SRDs and LHDs have different strengths and limitations that can complement each other in the context of signal management (Table 3). Advantages of SRDs include the coverage of all the types of medicinal products, including over-the-counter and prescribed medicines, vaccines, herbal products, traditional medicines, and even illicit drugs, as well as all types of suspected ADRs, including less severe events, which do not require medical attention and might therefore be not recorded in LHDs, and very rare events, which may not be properly captured in LHDs unless they are part of large-scale distributed database networks. Conversely, LHDs usually collect information on either prescribed or dispensed medicines, are restricted to either primary or secondary care, and, as far as administrative databases are concerned, they may be restricted by the type of medicines and healthcare services reimbursed in a specific patient population by the national health system or insurance [15].

The underlying populations covered by individual SRDs tend to be larger than those covered by LHDs in the same country, as most countries have nationwide spontaneous reporting systems, whereas LHDs are often restricted to subsets of the country's population, either based on geographical regions or healthcare providers [16, 17]. However, there are notable exceptions such as the Scandinavian

Table 1 Examples of spontaneous reporting databases used for pharmacovigilance activities

Database name	Description
VigiBase	The global database of the WHO, which contains over 38 million ICSRs (unique case reports) from more than 150 countries (including the majority of reports of Eudravigilance, FAERS and VAERS)
EudraVigilance	The European database of the EMA, which contains over 25 million ICSRs from the European Economic Area and other countries
FDA Adverse Events Reporting System (FAERS) and the Vaccine Adverse Events Reporting System (VAERS)	FAERS is the US database of the FDA, which contains over 29.6 million drug-focused ICSRs (including follow-up reports) from the USA and other countries, corresponding to 20.1 million unique cases. The vaccine-focused version, VAERS, contains information about more than 2.6 million reports of adverse events following immunization since 1990
Japanese Adverse Drug Event Report (JADER)	JADER is the database of the Pharmaceuticals and Medical Devices Agency (PMDA) in Japan, established in 2024. It contains more than 1 million ICSRs, mainly from Japan.
UK Yellow Card scheme	The Yellow Card Scheme was created in 1964 and is run by the MHRA. Its main purpose is to serve as a system for the early detection of unexpected ADRs. It contains more than 1 million ICSRs

ADR adverse drug reaction, *EHRs* electronic health records, *EMA* European Medicines Agency, *FDA* Food and Drug Administration, *ICSRs* individual case reports, *MHRA* Medicines and Healthcare products Regulatory Agency, *UK* United Kingdom, *WHO* World Health Organization

Table 2 Examples of longitudinal healthcare databases used for pharmacoepidemiologic studies

Database name	Coverage	Data content
Electronic health records		
Clinical Practice Research Datalink (CPRD)	UK	Primary care data, including demographics, diagnoses, symptoms, laboratory test results, prescriptions, vaccinations, lifestyle factors, and referrals
The Health Improvement Network (THIN)	UK	General practice records with patient demographics, prescriptions, clinical diagnoses, referrals, and consultations
NHS Scotland Databases	Scotland	Comprehensive healthcare data including primary and secondary care records, prescriptions, laboratory tests, hospital admissions, and mortality
The Information System for Research in Primary Care (SIDIAP)	Spain	Primary care records covering demographics, prescriptions, clinical diagnoses, lifestyle factors, and healthcare utilization
All of Us Research Program	USA	A diverse, population-based cohort collecting EHRs, genetic data, biospecimens, survey responses, wearable device data, and social determinants of health
Cerner Real-World Data	USA	A large-scale de-identified EHR database covering hospital visits, diagnoses, medications, laboratory results, and clinical notes
TriNetX	Global coverage	A real-time, cloud-based network providing de-identified patient-level EHR data from healthcare organizations, including diagnoses, procedures, medications, and laboratory results
Claims databases		
Système National des Données de Santé (SNDS)	France	Nationwide healthcare claims data including prescriptions, hospitalizations, medical procedures, chronic disease registries, and socioeconomic status
Medicare	USA	Data for individuals aged ≥ 65 years and disabled persons, covering inpatient and outpatient services, pharmacy claims, physician visits, and long-term care
Medicaid	USA	It covers low-income populations, including inpatient and outpatient claims, prescription drug use, and physician visits
Merative MarketScan	USA	Proprietary claims data from employers and insurers, including inpatient and outpatient services, medical procedures, prescriptions, and lab results
National Health Insurance Research Database (NHIRD)	Taiwan	Nationwide health claims data covering demographics, hospitalizations, diagnoses, prescriptions, laboratory tests, and medical procedures
Korean National Health Insurance Service (NHIS) Database	South Korea	Nationwide healthcare claims data including prescriptions, hospitalizations, medical procedures, and mortality records
Electronic health records and claims databases		
Optum Clinformatics Data Mart	USA	Includes de-identified claims data covering medical claims, pharmacy claims, laboratory results, patient demographics, and provider details
Health Maintenance Organization Databases (e.g., Kaiser Permanente, HealthPartners)	USA	Data from managed care organizations containing comprehensive medical claims, prescriptions, patient demographics, diagnoses, healthcare utilization, and laboratory test results
PHARMO Database Network	The Netherlands	Links community pharmacy, hospital discharge, primary care, and mortality data, including medication use, clinical outcomes, and healthcare utilization
Danish National Registers	Denmark	Population-based registries containing prescription drug use, hospital admissions, primary care visits, and mortality data
Finnish National Health Registers	Finland	Nationwide registries linking prescription, hospital discharge, and mortality data for pharmacoepidemiologic research

Table 2 (continued)

Database name	Coverage	Data content
Patients' registries		
Swedish Cancer Register	Sweden	Detailed information on all diagnosed cancer cases in Sweden, including tumor type, treatment, and survival outcomes
UK Biobank	UK	Genetic, lifestyle, clinical, and health data for over 500,000 participants, used for a wide range of health-related research
Danish National Patient Registry (NPR)	Denmark	Hospital diagnoses, procedures, and treatments for all residents of Denmark

EHR electronic health records, *NHS* National Health Service, *USA* United States of America, *UK* United Kingdom

population-based registers [18, 19] and the French national healthcare database [20].

Advantages of LHDs include not relying on the reporter's judgment that the event experienced by the patient may have been drug-induced, thus reducing the risk of missing certain events due to underreporting. In addition, unlike spontaneous reports, which include information only for patients that experience a suspected ADR, LHDs also collect information on patients taking the medicine without experiencing ADRs, as well as those who experience the same adverse events without taking a medicine. This allows for the calculation of population-level association measures (e.g., relative risks) of clinical events for patients taking a medicine, as compared with those not taking the medicine or taking another comparator drug, can be calculated, providing measures of clinical impact [21, 22].

Data available in SRDs is near real-time and can typically be analyzed within a few days or weeks. Conversely, data from LHDs are usually available with a larger time lag (depending on data processing and type of use) and the analytical process may require several weeks or months (including ethical approval and more complex study design).

SRDs sometimes contain more detailed clinical information about the case of interest (e.g., the course and burden of the event), and these are often provided in an unstructured narrative format. This makes this source particularly suitable for causality assessment [21]. When available, clinical data in LHDs are generally recorded in free text fields, and they capture detailed and unstructured patient information that is often missing from structured fields. Such data could be helpful for a better characterization of drug exposure, a more comprehensive identification of the outcomes, and analysis adjustment for potential confounders [23]. However, in LHDs clinical information is usually not collected for research purposes and for this reason may be generally less detailed. Consequently, evaluating causality associations using these data requires suitable analytical methods.

A primary reason for the use of spontaneous reporting is signal detection, based on rapid detection of suspected unknown ADRs of newly marketed products, as well as those used off-label or used incorrectly (e.g., abuse, misuse, and overdose), occurring anywhere in the world. In addition to disproportionality analysis using both frequentist [i.e., proportional reporting ratios (PRR), reporting odds ratios (ROR), and relative RORs] and Bayesian methods [i.e., information components (IC) and multi-item gamma-Poisson-shrinker (MGPS) algorithms] [24, 25], signal detection methods in SRDs include the use of the time-to-onset algorithm (focused on the difference in the time-to-onset of an adverse event) [26] and methods for the detection of drug–drug interaction signals (e.g., Ω shrinkage measure, additive and multiplicative models) [25]. This makes it a tried and tested early warning system that enables rapid identification of safety concerns. However, once safety concerns become more broadly known, there may be stimulated reporting of those safety issues making continued quantitative analysis of the spontaneous reporting data fraught with difficulty [27]. This stimulated reporting, along with well-known limitations of spontaneous reports, means that formal hypothesis testing cannot be done in such data [28–30]. In contrast, LHDs generally do not suffer from these weaknesses, at least not to the same extent. Therefore, LHDs have a long, well-established ability to be used for pharmacoepidemiologic studies [31], aimed at investigating a potential safety signal to confirm or refute the causal association of the drug and the adverse event. On the contrary, signal detection using LHDs is challenging, given the lack of universal data capture, the time needed to have data available for analysis, the complexity of data collection, lack of consistency in coding adverse events, as well as the lack of clinical suspicion of the link between adverse events and drugs or vaccines [32]. For this reason, after two decades from the first pilot studies and publicly funded projects that explored the potential of LHDs for detection of safety signals without any prior hypothesis, currently, LHDs are not routinely used in pharmacovigilance for signal detection.

Table 3 Main differences between spontaneous reporting databases and longitudinal healthcare databases

	Spontaneous reporting databases		Longitudinal healthcare databases	
	Strengths	Limitations	Strengths	Limitations
Medicines coverage	Broad coverage, including over-the-counter medicines, prescribed medicines, vaccines, herbal products, traditional medicines, and illicit drugs	–	–	Limited coverage, often restricted to prescribed (or dispensed) medicines into the specific healthcare settings (and which are reimbursed by national health systems or health insurance)
Population coverage	Large population coverage, generally nationwide or multi-country.	–	–	Restricted population coverage, often based on geographical regions or healthcare providers
Reporting method	–	Mainly relies on voluntary reporting and reporters' judgment, leading to potential underreporting Higher potential for stimulated reporting and duplication	Not reliant on voluntary reporting, hence it is not affected by underreporting	Potential for outcome and exposure misclassifications Data on drug use can often prevent appropriate capture of the actual drug exposure and/or treatment duration
Information availability	Near real-time data availability, enabling rapid signal detection and analysis	Subject to duplicated reports and inconsistent reporting delays	–	Data available with some months lag and requiring several weeks/months for analysis
Clinical information detail	Contain detailed clinical information in unstructured narrative format, suitable for causality assessment	Requires advanced techniques to extract insights from unstructured data	When available, clinical notes are helpful to provide a more comprehensive understanding of patient health	Clinical information is usually not collected for research purposes, less detailed, requiring suitable methods for causality estimation taking into account potential confounding
Risk measures	–	Only possible to evaluate disproportionality reporting of certain suspected ADR for specific medications, but it is not possible to estimate absolute and relative risks owing to absence of the denominator (i.e., number of users) and underreporting	Useful for calculating association measures at population level, providing measures of clinical relevance or impact	Requires advanced statistical adjustments to account for confounding factors

ADR adverse drug reaction

3 Integration of Information and Interplay of SRDs and LHDs

In this section, the main approaches for the integration of information and interplay of SRDs and LHDs are described following the signal management phases (Table 4), along with case studies and examples.

3.1 Data Mining for Signal Detection Using LHDs

Signal detection in pharmacovigilance traditionally relies on spontaneous reporting system databases. However, since more than 15 years ago, a number of international initiatives, such as the European project "Exploring and Understanding Adverse Drug Reactions" (EU-ADR) [33], the Observational Medical Outcomes Partnership (OMOP) [34], and the Pharmacoepidemiological Research on Outcomes of Therapeutics by a European Consortium (PROTECT) [35], explored if and when data mining of large, distributed LHDs networks may help in identifying safety signals [32, 35–40], including in pediatric populations [41]. Likewise, in more recent years, regulatory agencies have begun to explore the feasibility and values of using LHDs in the context of signal detection. For example, the Sentinel System has employed a statistical data mining method called TreeScan to identify potential adverse neonatal and infant outcomes after maternal medication exposures using Merative MarketScan [42], and potential safety issues associated with the use of semaglutide for type 2 diabetes mellitus [43]. TreeScan is particularly suited for signal detection in healthcare databases because it leverages the hierarchical structure of commonly used coding systems (e.g., ICD-10-CM) and simultaneously scans for safety signals at various levels of these coding systems [44]. In Europe, findings from the European project, EU-ADR, previously showed the potential value of LHDs for signal detection, especially for the assessment of adverse events that are not so easily attributable to drugs, such as the ones that are multifactorial and have a high background incidence in the general population (e.g., acute myocardial infarction, hip fracture, and upper gastrointestinal bleeding) [45, 46]. LHDs also support prospective safety surveillance as post-marketing data of a medication's use accrued over time. These surveillance activities can scan for many outcomes or target specific outcomes of interest. For example, Sentinel has conducted prospective surveillance of saxagliptin use on the risk of acute myocardial infarction [47] and rivaroxaban use on the risks of ischemic stroke, intracranial hemorrhage, and gastrointestinal bleeding [48]. Unlike some of the traditional methods used in SRDs, methods used for signal detection in LHDs could adjust for confounding and other biases using approaches such as propensity score matching [47–49].

Nevertheless, data mining of LHDs may lead to the identification of additional spurious signals, thus substantially increasing the number of potential safety signals to be validated, and ultimately, the burden of the signal management process on the regulatory agency [50]. As such, as mentioned earlier, this approach has not yet been routinely implemented into the current pharmacovigilance activities owing to the high possibility and negative impact of spurious signals, which are likely to be captured mining large amount of EHRs.

Exploration into the use of LHDs for the detection of types of common harms from medicines, such as signals relating to medication errors and inappropriate use of medicines, which are challenging for SRDs, should be prioritized. In this regard, the development and assessment of specific quality of care indicators (e.g., process indicators, patient outcomes, and healthcare resources utilization) using claims databases and/or primary care databases is one of the most suitable approaches to conduct real-world studies on these topics.

3.2 Observed-Over-Expected Analysis of Adverse Events

In SRDs, adverse events may be reported on the basis of an assumed causal relationship with medicine by the reporter, but it cannot be ruled out that these adverse events might have occurred by coincidence after medicines use. Often, more formal epidemiological studies are needed to confirm the safety signals. In these circumstances, well-conducted observed-over-expected (O/E) analysis can be used to strengthen or refute the initial safety signal [4]. O/E analysis is commonly used to assess the imbalance between the actual number of adverse events being reported within a well-defined population (e.g., cohort of drug users) and the anticipated number of adverse events in that population based on estimated background incidence rates using different data sources, such as LHDs [51–53]. In this way, the observed frequency of an adverse event in a SRD can be compared with the expected background frequency as derived from LHDs, providing insights into the epidemiology of the adverse events of interest. O/E analysis is frequently applied in post-marketing surveillance of vaccines, because when known proportions of a population receive vaccinations, expected values of adverse events (especially serious adverse events) in the overall vaccinated population (and in its age- and sex-strata) may be calculated.

Regarding the observed number of adverse events, as mentioned above, SRD data often underestimate adverse events owing to underreporting and reporting bias, making

Table 4 Main types of integration and interplay between spontaneous reporting databases and longitudinal healthcare databases

Type of interplay	Main characteristics
Data mining for signal detection using longitudinal healthcare databases	Data mining allows scanning for a large number of outcomes or target specific outcomes of interest Helpful for detecting adverse events with high background incidence or not strongly attributable to drugs Supports prospective safety surveillance and adjusts for confounding factors using methods such as propensity score matching Data mining in SRDs can increase the risk of identifying spurious signals, leading to an increased number of potential safety signals to be validated. LHDs can improve specificity by validating signals using population-based study designs As compared with SRDs, LHDs provide structured, population-based data with lower reporting bias, enabling adjustment for confounders Not routinely implemented in pharmacovigilance activities
Observed-over-expected analysis	Used to explore the potential imbalance between the actual number of events occurring within a well-defined population and the anticipated number of events in that population, based on estimated background incidence rates. It is commonly applied in vaccine pharmacovigilance to compare observed frequency with expected frequency derived from other real world data sources SRDs analysis can provide the observed number of adverse events reported following exposure to a medical product or vaccine LHDs analysis can provide background incidence rates of adverse events in the general population or among specific subgroups, offering real-world background rates for observed-over-expected rates It helps in determining potential causal relationships between reported adverse events and exposure to medical products/vaccines It requires careful consideration of biases, underreporting, and (regarding the expected number of events) variations in data sources
Analyses of patterns in spontaneous reporting databases to inform the conduct of pharmacoepidemiologic studies	SRDs can be analyzed to identify potential risk factors, confounders and effect modifiers of a certain signal which may inform the protocol of a formal pharmacoepidemiology study using LHDs for signal assessment Time-to-onset analysis in SRS aids in understanding the temporal relationship between drug exposure and the onset of adverse drug reaction, and to inform risk windows for future pharmacoepidemiologic studies exploring that association using LHDs SRDs may help to more precisely define the outcome to be investigated in a formal pharmacoepidemiology study using LHDs, based on actually reported adverse events (e.g., specific type of infections reported for a specific class of immunosuppressant drugs)

LHDs longitudinal healthcare databases, *SRDs* spontaneous reporting databases

it difficult to determine precise numbers in a well-defined population. However, in cases where underreporting or other forms of biases are likely to be low or absent (e.g., in the case of a specific clinical entity that is very likely to be captured by the SRD) SRD can provide more accurate data. An example is the occurrence of thrombosis with thrombocytopenia syndrome (TTS) associated with Vaxzevria, one of the initially marketed coronavirus disease 2019 (COVID-19) vaccines, for which in some countries data of all suspected adverse events occurring within the underlying population were shared and reported on a national level [54]. Another example is the occurrence of conditions that require treatment in a limited number of specialized centers, allowing for a correct registration and combination of the observed number of adverse events [55]. Severe cutaneous reactions such as toxic epidermal necrolysis, which require treatment in dedicated burn centers because of the extent of skin loss, support this concept [56].

To calculate the expected number of adverse events in the population, similar considerations apply. The first requirement is that for the event under study, people will contact a healthcare professional, and the event is coded in a reliable and consistent way in the employed data source. The background incidence rate, used to calculate the expected number, is the number of adverse events divided by the amount of person-time in the population [55]. However, the types of data available for the population under study may vary. An international study aimed to generate background

incidence rates of adverse events of special interest (AESI) in the context of the COVID-19 pandemic, based on ten data sources from seven European countries, showed differences between the various data sources due to heterogeneity of data collection, coding systems, and healthcare practices [57, 58]. In addition, calculation of the expected number of events should be based on the risk period for which an increased risk for the occurrence of the event can reasonably be expected. Furthermore, differences in background incidence rates by sex, age category, or additional risk factors may require stratified analyses [55].

Several requirements should be fulfilled for a successful comparison of observed versus expected events. The definitions used for calculating the observed and expected frequencies should align as much as possible. As an example, in vaccine safety, the Brighton Collaboration case definitions may be used to define the observed number of events in SRDs as well as the observed number of adverse events in LHDs that do not necessarily use the same diagnostic criteria [59]. Different health care settings may differ in the way diagnoses are made, their level of certainty, and the severity of adverse events. A final point of attention is that different coding systems may have been used to code the observed cases (e.g., the Medical Dictionary for Regulatory Activities [MedDRA] terminology) and the cases used to calculate the expected number of events (e.g., the International Classification of Diseases, ninth [ICD-9] and tenth [ICD-10] revisions, the Systematized Nomenclature of Medicine Clinical Terms [SNOMED CT], and the International Classification of Primary Care [ICPC]).

O/E analysis is an example of an approach in which data from various sources are utilized to gain a deeper understanding of the relationship between exposure to a medical product, either drug or vaccine, and a reported adverse event. Whenever the number of observed events is much higher than the number of expected events in a specific exposure cohort, this would point towards a causal relationship between the drug/vaccine exposure and the adverse event. The availability of large-scale distributed database networks in several geographic areas globally offer the opportunity to obtain more rapidly background data on adverse events, thus potentially enabling a more frequent and expeditious application of this approach [60].

3.3 Analyses of Reported ADR Patterns in SRDs to Inform Epidemiological Studies in LHDs

Patient characteristics such as age, sex, as well as underlying medical conditions (including pregnancy status) and concomitant medications of patients experiencing suspected ADRs are generally available in SRDs, but such information may also be inaccurate owing to the voluntary nature of spontaneous reporting [61–63]. In addition, regarding pregnancy, the adverse event report structure is insufficient as it has not been developed to capture information on the mother–child adverse events relationship.

Stratified descriptive and disproportionality analyses might identify patient subgroups at increased risk of experiencing specific ADRs [64, 65]. Therefore, information from SRDs can allow the identification of some potential confounders and effect modifiers, as well as the hypothesis generation related to drug–drug interactions that should be considered when planning the design of a pharmacoepidemiology study using LHDs and aimed at validating a safety signal. However, as for any statistical analysis of SRDs, such patterns should be interpreted as hypothetical and need to be critically reviewed by pharmacovigilance experts before being included in subsequent epidemiological analyses. Accordingly, it was demonstrated in FAERS that subgroup analyses performed better for age and sex, while for covariates not well-captured in FAERS, such as underlying condition and pregnancy, additional data sources should be considered [66]. It should also be acknowledged that the covariates available in SRDs have some statistical limitations (e.g., the sparse-data bias) that should be taken into account when adjusting the analyses, as they could affect sensitivity and/or specificity of the disproportionality analysis. Furthermore, adjustments in disproportionality analyses not only pose a theoretical risk of sparse-data biases, potentially causing even large databases to suffer from a lack of statistical power, but also introduce biases when covariates have missing data that are not missing at random [67, 68].

Moreover, another piece of information available in SRDs that can be leveraged for the design of a pharmacoepidemiology validation study is the time from the beginning of the drug exposure to the occurrence of the suspected ADR of interest (i.e., time-to-onset) [69]. Information concerning time-to-onset is crucial for understanding the temporal relationship between drug administration and the occurrence of adverse events. In addition, it helps in distinguishing between adverse events that are likely to be related to drug exposure and those which are not [70]. As for pharmacoepidemiologic validation studies, this information can be used, along with available published scientific literature, to define appropriate time windows in which the event of interest should be investigated in relation to the timing of the treatment exposure (i.e., risk period). For example, time-to-onset information can be particularly relevant in "case-only studies", where the investigator needs to select a specific time window during which the events of interest are likely to occur. If the event of interest occurs after the time window chosen by the investigator, then it is less likely that the event is related to treatment exposure. However, it should be noted that time-to-onset reporting in SRDs may not be fully accurate, and

a considerable amount of case reports have missing values. As such, information on time-to-onset should be cautiously interpreted [71, 72].

Therefore, preliminary analyses of SRDs can help researchers better planning pharmacoepidemiologic studies for confirming/refuting safety signals using LHDs by retrieving important information that is hard to derive from scientific literature, such as potential risk factors for specific adverse event–drug combinations, time-to-onset to define risk windows, and additional clinical details (e.g., the type of infection for a drug that potentially increases the risk of infection is more likely to be reported in SRDs in association with a medicine that is known to increase overall the risk of infection) for a more precise outcome definition.

3.4 Case Studies and Examples

Historically, signal detection of approved medicines has been conducted using spontaneous reports, and pharmacoepidemiology studies subsequently conducted to evaluate specific safety signals as part of signal management activities [73]. An example of this paradigm working in practice is dabigatran-induced bleeding, which was observed in the spontaneous reporting data. Although the existence of reports per se was not surprising given the frequency of bleeds in the treated population, the volume of reporting seemed unexpectedly high, thus leading to a safety signal. FDA's Sentinel (miniSentinel at the time) was used to examine the potential association and did not find any significantly increased rate of bleeding, when using warfarin as a comparator [74–76].

The signal concerning the association between fenfluramine/phentermine and valvular heart disease was identified with case reports [77] and subsequent pharmacoepidemiology studies confirmed and strengthened the safety signal [78, 79].

Other notable examples include the concerns surrounding cutaneous small vessel vasculitis (CSVV) following direct oral anticoagulant (DOAC) exposure, which first emerged from the US FAERS database with a median time-to-onset of 11 days. A parallel study was begun in the US Sentinel System with the aim of providing further contextual information in an LHD with more complete coverage of the exposed population [80]. To characterize any differential risk among DOACs and warfarin, an active new user cohort, propensity score matched study was performed. No elevated risk was found in six pairwise comparisons that included rivaroxaban, dabigatran, apixaban, and warfarin. CSVV ranged from 3.3 to 5.6 per 10,000 in the cohort of new users of the drug who were affected by atrial fibrillation [81].

While the CSVV outcome was being assessed, a similar study of DOACs was being conducted in the US Sentinel System for severe uterine bleeding. The study enrolled more than 1 million new users of DOACs and found an increased risk of surgical intervention to correct severe uterine bleeding with rivaroxaban as compared with other DOACs or warfarin, with statistically significant hazard ratios that ranged from 1.19 to 1.34 depending on the comparator. These findings contributed to a class-wide labelling change for DOACs to raise awareness of the risks, highlighting the utility of combining complementary evidence from SRDs and LHDs.

Finally, a signal assessment of the drug–event combination of nintedanib and ischemic colitis demonstrated the integrated use of spontaneous and observational data in a review of the Bradford–Hill causality criteria [82]. Review of the case reports from SRD provided evidence for the criteria of specificity, consistency, and analogy. Calculations of incidence rates of colitis in new users of nintedanib across multiple populations in observational health databases provided evidence of the strength of the association and further evidence of consistency. The signal assessment was supplemented with characterization of real-world users and an exploration of potential risk factors using observational health data.

4 Future Directions and Opportunities

In this section, initiatives that can improve the interplay between SRDs and LHDs and the overall efficiency of the signal management process are highlighted.

4.1 Networks of LHDs

As mentioned, one of the limitations of even the largest LHDs is their inability to cover an entire population or even multiple populations and different healthcare settings, which can limit their utility in the investigation of rare adverse events. In these contexts, solutions could be found in multinational studies that have nowadays transformed into multi-database studies (e.g., Vaxzevria and TTS) [83]. The adoption of common data models and of federated analyses allowed us to make these analyses more efficient and to access more data sources. In addition to the previously mentioned US Sentinel System, other similar large-scale initiatives will make these approaches more common.

DARWIN EU uses the OMOP common data model (CDM), based on which standardized analytics are built as packages in R [84]. Studies are conducted as federated analyses and typically involve several data partners using the same analytical code. Code is executed by data partners locally and there is no data transfer; only aggregated results

are provided back to the DARWIN EU Coordination center who operate the network and perform the studies [85]. Generic approvals of different types of analyses can speed up institutional review board approval and further reduce time needed to deliver study results.

4.2 New Methods for Signal Detection

Current research developments, in terms of methods, mainly concern signal detection and reside in the evaluation of the performances of machine learning techniques for this detection. In particular, several studies showed that the application of natural language processing (NLP) and machine learning techniques to EHR may help extract relevant clinical information from unstructured data and improve the detection of potential safety signals. However, the performance yielded by these models varied significantly across models and tasks, thus underlining the need for ongoing research and validation on diverse clinical applications. Furthermore, no study has demonstrated a concrete advantage of these machine learning approaches over traditional data mining methods in administrative databases so far, and no widely accepted guidelines currently exist for reporting and critically assessing the use of this technique in the analysis of EHR notes [86]. Developments are thus still needed in this field to identify methods that will ultimately contribute more efficiently to signal detection.

4.3 Additional Data Sources

Hospital data warehouses have been less explored so far, as it was complicated by their mixed content of structured and unstructured data and their use of various languages. It is indeed very recently that progress has been made in NLP, which has widely extended their use for drug safety monitoring [87, 88]. The short-term potentialities of implementing and further developing signal management techniques through database interplay might essentially reside in the more important integration of hospital data warehouse use [89]. Especially their mining for the identification of supplementary cases of potential ADRs at the early stage of a signal investigation could be extremely valuable [90]. Instead of waiting passively for further cases with extensive clinical characterization to be reported, contributing hospitals could be asked to actively search for cases in their databases and thus potentially reduce the time needed for signal confirmation.

Beyond signal detection, the integration of machine learning, and generative artificial intelligence in particular, into case review and other routine pharmacovigilance activities holds considerable potential to save time and resources by significantly reducing the workload of pharmacovigilance professionals [91–93]. More specifically, generative artificial intelligence, such as large language models (LLMs) can help minimize manual effort by automating repetitive tasks and data entry, improving accuracy and efficiency, and accelerating processing to ensure timely regulatory submissions [94]. This applies not only to SRDs, but also to LHDs. In this context, ambient artificial intelligence scribes, which are machine learning-based tools that can automate parts of the clinical documentation process, hold considerable potential to alleviate documentation burden. Ultimately, this could improve both the quality and availability of clinical data in LHDs, particularly in EHRs [95].

4.4 Embedding Case Reporting Within LHDs

An area of opportunity requiring further research is the potential value of embedding technological solutions for individual case reporting of suspected ADRs directly within and based on existing information contained in LHDs. This may facilitate the creation of spontaneous reports by (semi-) automatically populating reports with information available in the healthcare database, such as the patient's age, concomitant and past medicines, medical history, etc. This would not only reduce the effort for reporters and ensure more complete capture of basic data elements, but may free up reporters to spend time and effort on providing higher-grade information that is often missing in pharmacovigilance assessment, such as reflections by the reporter around possible alternative causes (including factors that they have ruled out or considered implausible), or around novel aspects of a known ADR that the reporter wants to highlight to the pharmacovigilance system. Ideally, the causality assessment and additional information thus captured in developing the spontaneous report would be retained in the healthcare database, increasing its value for pharmacovigilance and pharmacoepidemiologic follow-up. Linder et al. reported a successful and user-acceptable implementation of such an integration, with automatic generation of ADR reports from EHRs [96]. This led to reporting of ADRs among physicians who had not reported any in the previous 12 months. Such automation and integration may result in improvements with under-reporting of ADRs, which has been reported to varying degrees ranging between 19–42% [97] and 99% [96, 98].

Another valuable approach for such integration has been described by Powell et al., who reported the development of a digital bidirectional clinical communication channel between healthcare professionals and pharmacovigilance experts for exchanging adverse event information. This channel was successfully implemented in an EHR platform, leading to a 96% reduction in time needed to collect complete information on adverse events, as compared with traditional methods [99].

However, evidence is still needed that when such initiatives scale, they not only lead to increased volumes of reports but also that this translates to increased receipt of reports with actionable data [100].

In the longer term, advances in technology and further integration of information from summary of product characteristics, databases of known ADR associations, etc., together with further artificial intelligence (AI) methods may lead to improved ADR recognition and management via automated flags to physicians at the point of prescribing or at the point of consulting for possible ADRs, lending itself to prevention and risk minimization activities. For example, pre-specified tests can be built in EHRs upfront depending on applicable risk minimization measures for medicines that require baseline testing, continuous monitoring, or in those with additional risk factors, comorbidities, and comedication. As novel AI methods will likely be trained on existing known "true positive" associations, every effort should be made to allow for the unpredictable and bizarre reactions to be detected, and such methods to be validated [101]. Reflections on the use of such technologies relating to medicines across their lifecycle have been published in Europe [101].

5 Data Integration and Linkage

In the previous sections, we discussed how analyses from both SRDs and LHDs are used in the signal management process, and how their different strengths and limitations determine their suitability for the different phases of this process. We also highlighted case examples where these data sources complement each other. An extreme example of this complementarity is when the two types of data sources would be eventually integrated and linked with each other.

Successful integration necessitates continuous collaboration among stakeholders, emphasizing shared standards and transparent methodologies [102]. Quality control procedures should be in place to validate the accuracy of linkages and minimize errors. Several crucial aspects influence the success of integrating data from SRDs and LHDs. Addressing these aspects lays the groundwork for robust signal management across diverse data sources:

5.1 Standardization and Harmonization of Data Formats and Terminologies

Achieving consistency in data formats and terminologies is essential for effective integration. Standardization involves the development and implementation of consistent rules and conventions for representing data. In the context of SRDs and LHDs, the use of common terminologies seeks to ensure that data are formatted in a consistent manner across different sources. This includes adopting common data elements, codes, and terminologies as much as possible. For instance, using standardized medical coding systems, such as MedDRA, SNOMED (https://www.snomed.org/) or the Logical Observation Identifiers Names and Codes (LOINC) (https://loinc.org/), as well as drug reference terminologies such as the Anatomical Therapeutic Chemical Classification System (https://atcddd.fhi.no/atc_ddd_index/), the WHODrug [103], and RxNorm (https://www.nlm.nih.gov/research/umls/rxnorm/index.html), ensures that medical concepts are represented uniformly, facilitating seamless integration. Harmonization goes a step further by reconciling differences between similar concepts represented differently in various databases. It would involve creating mappings or crosswalks to establish semantic equivalence, allowing for meaningful linkage and analyses.

5.2 CDMs and Availability of Variables

The existence of CDMs helps to establish a unified structure for diverse datasets. In the context of SDRs, the International Council for Harmonization (ICH) E2B guideline effectively represents the standard CDM for SRDs [104]. In general, CDM provides a standardized way of structuring and organizing data, enabling interoperability across different systems. In the integration of spontaneous reporting and LHDs, adopting a sort of shared CDM may promote consistency. The availability of key variables needed for integration is vital in ensuring that essential data elements required for a specific analysis or signal detection are present in both datasets. It is understandable that spontaneous reports do not have complete information, but this is a reason to leverage healthcare databases to better understand the clinical context of these spontaneous reports. This requires collaborative efforts to define essential variables necessary for meaningful integration and encourage their inclusion in both spontaneous reporting and LHDs moving forward.

5.3 Ensuring Protection of Sensitive Data

The integration of SRDs and LHDs for signal management, however, would bring forth significant privacy and data protection challenges, particularly concerning identifiable data. Combining datasets from these sources requires careful consideration of patient privacy, as the data often contains sensitive health information. Privacy regulations, such as General Data Protection Regulations (GDPR) and the Health Insurance Portability and Accountability Act (HIPAA) of 1996, impose strict guidelines on the use and sharing of patient data. Complementary and combined use of SRDs and LHDs requires meticulous adherence to

regulatory standards. In addition, the integration process should align with existing pharmacovigilance regulations to ensure that safety signals and adverse events are appropriately identified and reported [105, 106]. Ensuring compliance with these regulations, while still allowing for meaningful data integration, is a delicate balance. Difficulty in matching patient identifiers across disparate datasets is a common challenge and data linkage from pharmacovigilance databases is generally not possible. If any data linkage should be considered and implemented in the future, appropriate safeguards must be in place to de-identify and protect patient identities, and researchers must navigate ethical considerations associated with data linkage. Employing advanced matching algorithms and encryption techniques can help address these challenges, striking a balance between the need for linkage and preserving patient privacy.

Addressing issues of patient consent is a critical ethical consideration in the integration of different healthcare data including both LHDs and SRDs. In SRDs, where data is often collected without direct patient consent, ethical challenges arise when linking this data with LHDs. Developing strategies to navigate these challenges, such as anonymizing or de-identifying data is necessary. Some of the National Patient-Centered Clinical Research Network (PCORnet) Clinical Research Centers have successfully implemented an anonymous linkage tool using encrypted hashed identifiers to identify overlapping patients across a large urban region [107]. Ethical considerations in linking sensitive healthcare data involve respecting patient privacy and autonomy. It necessitates clear communication with patients about the potential uses of their data and providing mechanisms for opting out if they choose to do so.

5.4 Organizational and Operational Barriers to Achieving Effective Data Integration

While the integration of SRDs and LHDs offers significant advantages in signal management, several organizational and operational barriers must be addressed to realize this vision. The technical and regulatory challenges outlined earlier are critical, but additional factors related to stakeholder coordination, sustainability, and real-world implementation also play a crucial role in determining the feasibility of such integration.

One key challenge is the fragmentation of responsibilities among stakeholders in pharmacovigilance and healthcare data management. While regulatory agencies set safety monitoring standards, pharmaceutical companies, academic researchers, and healthcare providers generate and analyze data using their own methodologies. The absence of a centralized coordinating body to oversee data integration efforts leads to variability in implementation, inconsistent engagement levels, and delays in adopting best practices. A collaborative governance model that aligns incentives and responsibilities across these groups is necessary to maintain consistency and ensure sustained commitment to integration efforts.

Another barrier is the misalignment of priorities between pharmacovigilance and clinical research. The lack of dedicated analytical frameworks tailored to linking SRD and LHD data means that even when data are successfully integrated, they may not be optimized for detecting actionable safety signals in a regulatory context. Establishing fit-for-purpose analytical tools and refining research questions to align with pharmacovigilance objectives will be essential to maximize the value of integrated data.

Financial and resource constraints also present significant obstacles. Large-scale data integration requires long-term investment in infrastructure, personnel, and ongoing data curation efforts. However, funding mechanisms for such initiatives are often fragmented, with short-term grants or project-based investments rather than sustained financial support. Without dedicated funding sources and industry-wide commitment, integration efforts may stall before realizing their full impact. Developing public–private partnerships or embedding integration initiatives within broader healthcare digital transformation efforts may help secure the necessary financial and technical resources.

Finally, cultural and organizational inertia within institutions can slow progress. Many organizations rely on entrenched data-sharing policies, risk-averse regulatory interpretations, and legacy IT systems that resist modernization. Overcoming these barriers requires change management strategies, including stakeholder education, incentives for data sharing, and pilot programs that demonstrate the feasibility and benefits of integration. A phased approach to implementation, where smaller-scale integration efforts yield tangible benefits before full-scale adoption, may help organizations transition more effectively.

Addressing these barriers requires more than just technological solutions—it demands strategic coordination, sustainable funding, and a cultural shift toward open and collaborative data sharing done in a timely manner. Without these foundational changes, the full potential of integrated pharmacovigilance data will remain unrealized.

6 Conclusions

Over the years, SRDs, primarily for signal detection, and LHDs, especially for confirming/refuting signals once detected, have been consolidated as the cornerstones for the post-marketing evaluation of drug safety. The interplay

between these data sources holds a significant promise for optimization of signal management in pharmacovigilance. The increasing generation of RWD and the development of advanced AI-based analytical methods have enhanced the potential for utilizing this data to assess the risk/benefit of drugs and vaccines and ensure patient safety. In particular, the complementarity of SRDs and LHDs could be helpful for a wide range of pharmacovigilance activities, especially signal detection and assessment. However, challenges hampering the integrated and/or combined use of these data sources, including data linkage, data structure and harmonization, as well as data protection, need to be addressed to fully realize the potential of such a synergistic approach.

Although several initiatives concerning the integration of information coming from SRDs and LHDs have been conducted and described in the literature, the potential of RWD for signal management (especially concerning signal detection) has still to be explored. Hence, continued collaboration by pharmacovigilance and pharmacoepidemiology researchers and advancements in analytical methodologies are essential to optimize the interplay of SRDs and LHDs and to promote the use of RWD for all the phases of signal management.

Declarations

Funding Open access funding provided by Università degli Studi di Verona within the CRUI-CARE Agreement.

Conflict of interest G.T. has served, over the last 3 years, on advisory boards/seminars funded by Sanofi, MSD, Eli Lilly, Sobi, Celgene, Daiichi Sankyo, Novo Nordisk, Gilead, and Amgen on topics not related to content of this paper; he is also a scientific coordinator of the academic spin-off "INSPIRE srl," which has received funding from several pharmaceutical companies (Kiowa Kirin, Shionogi, Shire, Novo Nordisk, and Daiichi Sankyo) for conducting observational studies and additional consultancy services on topics not related to content of this paper. In addition, he is currently a consultant for Viatris in a legal case concerning an adverse reaction to sertraline. G.C. is a full-time employee of Bayer AG. A.B. is an employee of GSK and holds both stock and stock options with GSK. J.S.B. is an employee of TriNetX, LLC. T.A.H. is an employee at Takeda Pharmaceutical, United States. When the project started, M.S. was an employee at Daiichi Sankyo Inc. but currently, she is an employee at Bayer Inc. and receives stocks from the organization. However, the organization did not sponsor this manuscript, and the views expressed in this document do not represent the organization. A.Se. is an employee of the European Medicines Agency. A.B., G.N.N., S.S., M.T., E.v.P., and G.T. are Editorial Board members of Drug Safety, but were not involved in the selection of peer reviewers for the manuscript nor in any of the subsequent editorial decisions. The other authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the article apart from those disclosed.

Ethics approval Not applicable. This study does not involve any human or animal subjects, or any sensitive information.

Consent to participate Not applicable, as this study does not involve any human participants.

Consent for publication Not applicable, as this study does not involve any individual data or images that would require consent for publication.

Availability of data and material Data sharing does not apply to this article as no datasets were generated or analyzed during the current study.

Code availability Not applicable, as no custom code or software application was developed or used in this study.

Author contributions S.C.: writing—original draft, writing—review and editing, and visualization. A.B.: writing—review and editing. J.S.B.: writing—review and editing. G.C.: writing—review and editing. R.E.C.: writing—review and editing. T.A.H.: writing—review and editing. S.L.: writing—review and editing. J.C.M.: writing—review and editing. G.N.N.: writing—review and editing. A.P.: writing—review and editing. M.R.: writing—review and editing. M.S.: writing—review and editing. A.Se.: writing—review and editing. S.S.: writing—review and editing. A.Sp.: writing—review and editing. S.T.: writing—review and editing. M.T.: writing—review and editing. E.v.P.: writing—review and editing. G.T.: writing—original draft, writing—review and editing, supervision, and conceptualization. All authors read and approved the final version.

References

1. Council for International Organizations of Medical Sciences. Practical aspects of signal detection in pharmacovigilance. Report of CIOMS Working Group VIII. 2010. https://cioms.ch/wp-content/uploads/2018/03/WG8-Signal-Detection.pdf.
2. Heads of Medicines Agencies (HMA), European Medicines Agency (EMA). Guideline on good pharmacovigilance practices (GVP). Module IX—signal management (Rev 1). 2017. https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-good-pharmacovigilance-practices-gvp-module-ix-signal-management-rev-1_en.pdf.
3. Shakir SAW, Layton D. Causal association in pharmacovigilance and pharmacoepidemiology: thoughts on the application of the Austin Bradford–Hill criteria. Drug Saf. 2002;25:467–71.
4. Hammad TA, Afsar S, McAvoy LB, Le Louet H. Aspects to consider in causality assessment of safety signals: broadening the thought process. Front Drug Saf Regul. 2023. https://www.frontiersin.org/journals/drug-safety-and-regulation/articles/10.3389/fdsfr.2023.1193413/full.

5. Cave A, Kurz X, Arlett P. Real-world data for regulatory decision making: challenges and possible solutions for Europe. Clin Pharmacol Ther. 2019;106:36–9.
6. Dreyer NA, Blackburn SCF. Power to the people: why person-generated health data are important for pharmacoepidemiology. Am J Epidemiol. 2024;193:1215–8.
7. Anand P, Zhang Y, Merola D, Jin Y, Wang SV, Lii J, et al. Comparison of EHR data-completeness in patients with different types of medical insurance coverage in the United States. Clin Pharmacol Ther. 2023;114:1116–25.
8. European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP). ENCePP Guide on methodological standards in pharmacoepidemiology. Chapter 8: approaches to data collection. 2023. https://encepp.europa.eu/encepp-toolkit/methodological-guide/chapter-8-approaches-data-collection_en.
9. Darwin EU. The Darwin EU® Data Network 2024. https://darwin-eu.org/index.php/data/data-network.
10. European Medicines Agency (EMA). Data Analysis and Real World Interrogation Network (DARWIN EU). 2021. https://www.snds.gouv.fr/SNDS/Accueil.
11. Food and Drug Administration. FDA's sentinel initiative. 2024. https://www.fda.gov/safety/fdas-sentinel-initiative.
12. European Medicines Agency (EMA). Initiation of DARWIN EU® Coordination Centre advances integration of real-world evidence into assessment of medicines in the EU. 2022. https://www.ema.europa.eu/en/news/initiation-darwin-eur-coordination-centre-advances-integration-real-world-evidence-assessment-medicines-eu.
13. Heads of Medicines Agencies (HMA), European Medicines Agency (EMA). HMA-EMA joint big data taskforce phase II report: "Evolving Data-Driven Regulation." 2019.
14. Yamaguchi M, Inomata S, Harada S, Matsuzaki Y, Kawaguchi M, Ujibe M, et al. Establishment of the MID-NET® medical information database network as a reliable and valuable database for drug safety assessments in Japan. Pharmacoepidemiol Drug Saf. 2019;28:1395–404.
15. Faillie J-L, Montastruc F, Montastruc J-L, Pariente A. Pharmacoepidemiology and its input to pharmacovigilance. Therapie. 2016;71:211–6.
16. Trifirò G, Isgrò V, Ingrasciotta Y, Ientile V, L'Abbate L, Foti SS, et al. Large-scale postmarketing surveillance of biological drugs for immune-mediated inflammatory diseases through an Italian distributed multi-database healthcare network: the VALORE Project. BioDrugs. 2021;35:749–64.
17. Spini A, Pellegrini G, Ingrasciotta Y, L'Abbate L, Bellitto C, Carollo M, et al. Switching patterns of biological drugs in patients with psoriasis and psoriatic arthritis: insight from the VALORE database network. Expert Opin Biol Ther. 2024;24:399–409.
18. Laugesen K, Ludvigsson JF, Schmidt M, Gissler M, Valdimarsdottir UA, Lunde A, et al. Nordic health registry-based research: a review of health care systems and key registries. Clin Epidemiol. 2021;13:533–54.
19. Ludvigsson JF, Almqvist C, Bonamy A-KE, Ljung R, Michaëlsson K, Neovius M, et al. Registers of the Swedish total population and their use in medical research. Eur J Epidemiol. 2016;31:125–36.
20. Système National des Données de Santé. The French National Healthcare Data System. 2023. https://www.snds.gouv.fr/SNDS/Accueil.
21. European Medicines Agency (EMA). Screening for adverse reactions in EudraVigilance. 2016. https://www.ema.europa.eu/system/files/documents/other/wc500218606_en.pdf.
22. Strandell J, Caster O, Bate A, Norén N, Edwards IR. Reporting patterns indicative of adverse drug interactions: a systematic evaluation in VigiBase. Drug Saf. 2011;34:253–66.
23. Walker AM, Zhou X, Ananthakrishnan AN, Weiss LS, Shen R, Sobel RE, et al. Computer-assisted expert case definition in electronic health records. Int J Med Inform. 2016;86:62–70.
24. Cutroneo PM, Sartori D, Tuccori M, Crisafulli S, Battini V, Carnovale C, et al. Conducting and interpreting disproportionality analyses derived from spontaneous reporting systems. Front Drug Saf Regul. 2024. https://www.frontiersin.org/journals/drug-safety-and-regulation/articles/10.3389/fdsfr.2023.1323057/full.
25. Noguchi Y, Yoshimura T. Detection algorithms for simple two-group comparisons using spontaneous reporting systems. Drug Saf. 2024;47:535–43.
26. Van Holle L, Tavares Da Silva F, Bauchau V. Signal detection based on time-to-onset: extending a new method from spontaneous reports to observational studies. Pharmacoepidemiol Drug Saf. 2014;23:849–58.
27. Gordillo-Marañón M, Szmigiel A, Yalmanová V, Caplanusi I, Genov G, Olsen DB, et al. COVID-19 vaccines and heavy menstrual bleeding: the impact of media attention on reporting to EudraVigilance. Drug Saf. 2024;47(8):783–98.
28. Bate A, Evans SJW. Quantitative signal detection using spontaneous ADR reporting. Pharmacoepidemiol Drug Saf. 2009;18:427–36.
29. Eland IA, Belton KJ, van Grootheest AC, Meiners AP, Rawlins MD, Stricker BH. Attitudinal survey of voluntary reporting of adverse drug reactions. Br J Clin Pharmacol. 1999;48:623–7.
30. Thomas SH, Rawlins MD. Spontaneous reporting of fatal adverse drug reactions. Int J Risk Saf Med. 1992;3:179–81.
31. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. J Clin Epidemiol. 2005;58:323–37.
32. Bate A, Hornbuckle K, Juhaeri J, Motsko SP, Reynolds RF. Hypothesis-free signal detection in healthcare databases: finding its value for pharmacovigilance. Ther Adv Drug Saf. 2019;10:2042098619864744.
33. Coloma PM, Trifirò G, Schuemie MJ, Gini R, Herings R, Hippisley-Cox J, et al. Electronic healthcare databases for active drug safety surveillance: is there enough leverage? Pharmacoepidemiol Drug Saf. 2012;21:611–21.
34. Reich C, Ostropolets A, Ryan P, Rijnbeek P, Schuemie M, Davydov A, et al. OHDSI Standardized Vocabularies-a large-scale centralized reference ontology for international data harmonization. J Am Med Inform Assoc. 2024;31:583–90.
35. Wisniewski AFZ, Bate A, Bousquet C, Brueckner A, Candore G, Juhlin K, et al. Good signal detection practices: evidence from IMI PROTECT. Drug Saf. 2016;39:469–90.
36. Coloma PM, Schuemie MJ, Trifirò G, Gini R, Herings R, Hippisley-Cox J, et al. Combining electronic healthcare databases in Europe to allow for large-scale drug safety monitoring: the EU-ADR Project. Pharmacoepidemiol Drug Saf. 2011;20:1–11.
37. Norén GN, Hopstadius J, Bate A, Star K, Edwards IR. Temporal pattern discovery in longitudinal electronic patient records. Data Min Knowl Discov. 2010;20:361–87.
38. Trifirò G, Coloma PM, Rijnbeek PR, Romio S, Mosseveld B, Weibel D, et al. Combining multiple healthcare databases for postmarketing drug and vaccine safety surveillance: why and how? J Intern Med. 2014;275:551–61.
39. Whalen E, Hauben M, Bate A. Time series disturbance detection for hypothesis-free signal detection in longitudinal observational databases. Drug Saf. 2018;41:565–77.
40. Zhou X, Douglas IJ, Shen R, Bate A. Signal detection for recently approved products: adapting and evaluating self-controlled case series method using a US claims and UK electronic medical records database. Drug Saf. 2018;41:523–36.
41. Ferrajolo C, Coloma PM, Verhamme KMC, Schuemie MJ, de Bie S, Gini R, et al. Signal detection of potentially drug-induced

acute liver injury in children using a multi-country healthcare database network. Drug Saf. 2014;37:99–108.
42. Suarez EA, Nguyen M, Zhang D, Zhao Y, Stojanovic D, Munoz M, et al. Novel methods for pregnancy drug safety surveillance in the FDA Sentinel System. Pharmacoepidemiol Drug Saf. 2023;32:126–36.
43. Sentinel. Outcome monitoring following Ozempic use in patients with type 2 diabetes: a signal identification analysis. 2022. https://www.sentinelinitiative.org/studies/drugs/individual-drug-analyses/outcome-monitoring-following-ozempic-use-patients-type-2.
44. Kulldorff M, Dashevsky I, Avery TR, Chan AK, Davis RL, Graham D, et al. Drug safety data mining with a tree-based scan statistic. Pharmacoepidemiol Drug Saf. 2013;22:517–23.
45. Pacurariu AC, Straus SM, Trifirò G, Schuemie MJ, Gini R, Herings R, et al. Useful interplay between spontaneous ADR reports and electronic healthcare records in signal detection. Drug Saf. 2015;38:1201–10.
46. Patadia VK, Schuemie MJ, Coloma PM, Herings R, van der Lei J, Sturkenboom M, et al. Can electronic health records databases complement spontaneous reporting system databases? A historical-reconstruction of the association of rofecoxib and acute myocardial infarction. Front Pharmacol. 2018;9:594.
47. Toh S, Reichman ME, Graham DJ, Hampp C, Zhang R, Butler MG, et al. Prospective postmarketing surveillance of acute myocardial infarction in new users of saxagliptin: a population-based study. Diabetes Care. 2018;41:39–48.
48. Chrischilles EA, Gagne JJ, Fireman B, Nelson J, Toh S, Shoaibi A, et al. Prospective surveillance pilot of rivaroxaban safety within the US Food and Drug Administration Sentinel System. Pharmacoepidemiol Drug Saf. 2018;27:263–71.
49. Wang SV, Maro JC, Gagne JJ, Patorno E, Kattinakere S, Stojanovic D, et al. A general propensity score for signal identification using tree-based scan statistics. Am J Epidemiol. 2021;190:1424–33.
50. Hauben M, Reich L, Van Puijenbroek EP, Gerrits CM, Patadia VK. Data mining in pharmacovigilance: lessons from phantom ships. Eur J Clin Pharmacol. 2006;62:967–70.
51. Gordillo-Marañón M, Candore G, Hedenmalm K, Browne K, Flynn R, Piccolo L, et al. Lessons learned on observed-to-expected analysis using spontaneous reports during mass vaccination. Drug Saf. 2024;47:607–15.
52. Luteijn JM, Morris JK, Garne E, Given J, de Jong-van den Berg L, Addor M-C, et al. EUROmediCAT signal detection: a systematic method for identifying potential teratogenic medication. Br J Clin Pharmacol. 2016;82:1110–22.
53. van der Boom MDX, van Eekeren R, van Hunsel FPAM. Observed-over-expected analysis as additional method for pharmacovigilance signal detection in large-scaled spontaneous adverse event reporting. Pharmacoepidemiol Drug Saf. 2023;32:783–94.
54. Greinacher A, Thiele T, Warkentin TE, Weisser K, Kyrle PA, Eichinger S. Thrombotic thrombocytopenia after ChAdOx1 nCov-19 vaccination. N Engl J Med. 2021;384:2092–101.
55. Mahaux O, Bauchau V, Van Holle L. Pharmacoepidemiological considerations in observed-to-expected analyses for vaccines. Pharmacoepidemiol Drug Saf. 2016;25:215–22.
56. Noe MH, Micheletti RG. Diagnosis and management of Stevens–Johnson syndrome/toxic epidermal necrolysis. Clin Dermatol. 2020;38:607–12.
57. Li X, Ostropolets A, Makadia R, Shoaibi A, Rao G, Sena AG, et al. Characterising the background incidence rates of adverse events of special interest for COVID-19 vaccines in eight countries: multinational network cohort study. BMJ. 2021;373: n1435.
58. Willame C, Dodd C, Durán CE, Elbers R, Gini R, Bartolini C, et al. Background rates of 41 adverse events of special interest for COVID-19 vaccines in 10 European healthcare databases—an ACCESS cohort study. Vaccine. 2023;41:251–62.
59. Brighton collaboration. Case definitions. 2024. https://brightoncollaboration.org/case-definitions/.
60. Durand J, Dogné J-M, Cohet C, Browne K, Gordillo-Marañón M, Piccolo L, et al. Safety monitoring of COVID-19 vaccines: perspective from the European Medicines Agency. Clin Pharmacol Ther. 2023;113:1223–34.
61. Food and Drug Administration. FAERS quarterly data files documentation. 2015. https://www.fda.gov/drugs/questions-and-answers-fdas-adverse-event-reporting-system-faers/faers-quarterly-data-files-documentation.
62. Goldman SA. Limitations and strengths of spontaneous reports data. Clin Ther. 1998;20(Suppl C):C40-44.
63. Lindquist M. VigiBase, the WHO Global ICSR database system: basic facts. Ther Innov Regul Sci. 2008;42:409–19.
64. Petronijevic M, Ilic K. Associations of gender and age with the reporting of drug-induced hepatic failure: data from the VigiBase™. J Clin Pharmacol. 2013;53:435–43.
65. Sandberg L, Taavola H, Aoki Y, Chandler R, Norén GN. Risk factor considerations in statistical signal detection: using subgroup disproportionality to uncover risk groups for adverse drug reactions in VigiBase. Drug Saf. 2020;43:999–1009.
66. Mahaux O, Powell G, Haguinet F, Sobczak P, Saini N, Barry A, et al. Identifying safety subgroups at risk: assessing the agreement between statistical alerting and patient subgroup risk. Drug Saf. 2023;46:601–14.
67. Hopstadius J, Norén GN, Bate A, Edwards IR. Impact of stratification on adverse drug reaction surveillance. Drug Saf. 2008;31:1035–48.
68. Noguchi Y, Tachi T, Yoshimura T. Caveats of covariate adjustment in disproportionality analysis for best practices. Drug Saf. 2025;48:1–5.
69. Leroy F, Dauxois J-Y, Théophile H, Haramburu F, Tubert-Bitter P. Estimating time-to-onset of adverse drug reactions from spontaneous reporting databases. BMC Med Res Methodol. 2014;14:17.
70. Agbabiaka TB, Savović J, Ernst E. Methods for causality assessment of adverse drug reactions. Drug-Safety. 2008;31:21–37.
71. Scholl JHG, van Puijenbroek EP. The value of time-to-onset in statistical signal detection of adverse drug reactions: a comparison with disproportionality analysis in spontaneous reports from the Netherlands. Pharmacoepidemiol Drug Saf. 2016;25:1361–7.
72. Scholl JHG, van Hunsel FPAM, Hak E, van Puijenbroek EP. Time to onset in statistical signal detection revisited: a follow-up study in long-term onset adverse drug reactions. Pharmacoepidemiol Drug Saf. 2019;28:1283–9.
73. Edwards R, Faich G, Tilson H, International Society of Pharmacovigilance. Points to consider: the roles of surveillance and epidemiology in advancing drug safety. Pharmacoepidemiol Drug Saf. 2005;14:665–7.
74. 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;99:265–8.
75. Mini-Sentinel Medical Product Assessment. A protocol for assessment of Dabigatran. 2015. https://www.sentinelinitiative.org/sites/default/files/Drugs/Assessments/Mini-Sentinel_Protocol-for-Assessment-of-Dabigatran_0.pdf.
76. Southworth MR, Reichman ME, Unger EF. Dabigatran and postmarketing reports of bleeding. N Engl J Med. 2013;368:1272–4.
77. Connolly HM, Crary JL, McGoon MD, Hensrud DD, Edwards BS, Edwards WD, et al. Valvular heart disease associated with fenfluramine-phentermine. N Engl J Med. 1997;337:581–8.
78. Jick H, Vasilakis C, Weinrauch LA, Meier CR, Jick SS, Derby LE. A population-based study of appetite-suppressant drugs

and the risk of cardiac-valve regurgitation. N Engl J Med. 1998;339:719–24.
79. Khan MA, Herzog CA, St Peter JV, Hartley GG, Madlon-Kay R, Dick CD, et al. The prevalence of cardiac valvular insufficiency assessed by transthoracic echocardiography in obese patients treated with appetite-suppressant drugs. N Engl J Med. 1998;339:713–8.
80. Mohamoud M, Horgan C, Eworuke E, Dee E, Bohn J, Shapira O, et al. Complementary use of US FDA's Adverse Event Reporting System and Sentinel System to characterize direct oral anticoagulants-associated cutaneous small vessel vasculitis. Pharmacotherapy. 2020;40:1099–107.
81. Ajao A, Cosgrove A, Eworuke E, Mohamoud M, Zhang R, Shapira O, et al. A cohort study to assess risk of cutaneous small vessel vasculitis among users of different oral anticoagulants. Pharmacoepidemiol Drug Saf. 2022;31:1164–73.
82. Chandler RE. Nintedanib and ischemic colitis: signal assessment with the integrated use of two types of real-world evidence, spontaneous reports of suspected adverse drug reactions, and observational data from large health-care databases. Pharmacoepidemiol Drug Saf. 2020;29:951–7.
83. Li X, Burn E, Duarte-Salles T, Yin C, Reich C, Delmestri A, et al. Comparative risk of thrombosis with thrombocytopenia syndrome or thromboembolic events associated with different COVID-19 vaccines: international network cohort study from five European countries and the US. BMJ. 2022:e071594.
84. Darwin EU. Standardised analytics. 2023. https://darwin-eu.org/index.php/methods/standardised-analytics.
85. Darwin EU. Coordination Centre. 2023. https://darwin-eu.org/index.php/about/coordination-centre.
86. Golder S, Xu D, O'Connor K, Wang Y, Batra M, Hernandez GG. Leveraging natural language processing and machine learning methods for adverse drug event detection in electronic health/medical records: a scoping review. Drug Saf. 2025;48(4):321–37.
87. Marella WM, Sparnon E, Finley E. Screening electronic health record-related patient safety reports using machine learning. J Patient Saf. 2017;13:31–6.
88. Evans HP, Anastasiou A, Edwards A, Hibbert P, Makeham M, Luz S, et al. Automated classification of primary care patient safety incident report content and severity using supervised machine learning (ML) approaches. Health Inform J. 2020;26:3123–39.
89. Koutkias VG, Jaulent M-C. Computational approaches for pharmacovigilance signal detection: toward integrated and semantically-enriched frameworks. Drug Saf. 2015;38:219–32.
90. Abedian Kalkhoran H, Zwaveling J, van Hunsel F, Kant A. An innovative method to strengthen evidence for potential drug safety signals using Electronic Health Records. J Med Syst. 2024;48:51.
91. Painter JL, Chalamalasetti VR, Kassekert R, Bate A. Automating pharmacovigilance evidence generation: using large language models to produce context-aware structured query language. JAMIA Open. 2025;8:ooaf003.
92. Gosselt HR, Bazelmans EA, Lieber T, van Hunsel FPAM, Härmark L. Development of a multivariate prediction model to identify individual case safety reports which require clinical review. Pharmacoepidemiol Drug Saf. 2022;31:1300–7.
93. Martin GL, Jouganous J, Savidan R, Bellec A, Goehrs C, Benkebil M, et al. Validation of artificial intelligence to support the automatic coding of patient adverse drug reaction reports, using nationwide pharmacovigilance data. Drug Saf. 2022;45:535–48.
94. Pariente A, Micallef J, Lahouegue A, Molimard M, Auffret M, Chouchana L, et al. What place for intelligent automation and artificial intelligence to preserve and strengthen vigilance expertise in the face of increasing declarations? Therapie. 2023;78:131–43.
95. Tierney AA, Gayre G, Hoberman B, Mattern B, Ballesca M, Kipnis P, et al. Ambient artificial intelligence scribes to alleviate the burden of clinical documentation. NEJM Catal. 2024;5:CAT.23.0404.
96. Linder JA, Haas JS, Iyer A, Labuzetta MA, Ibara M, Celeste M, et al. Secondary use of electronic health record data: spontaneous triggered adverse drug event reporting. Pharmacoepidemiol Drug Saf. 2010;19:1211–5.
97. Januskiene J, Segec A, Slattery J, Genov G, Plueschke K, Kurz X, et al. What are the patients' and health care professionals' understanding and behaviors towards adverse drug reaction reporting and additional monitoring? Pharmacoepidemiol Drug Saf. 2021;30:334–41.
98. Hazell L, Shakir SAW. Under-reporting of adverse drug reactions: a systematic review. Drug Saf. 2006;29:385–96.
99. Powell G, Kara V, Naranjo D, Kulkarni M, Best-Sule K, Coster T, et al. Testing the feasibility of a digital point of care solution for the trusted near real-time bidirectional exchange of novel and informative adverse event information. Ther Innov Regul Sci. 2025;59:124–34.
100. Brajovic S, Piazza-Hepp T, Swartz L, Dal Pan G. Quality assessment of spontaneous triggered adverse event reports received by the Food and Drug Administration. Pharmacoepidemiol Drug Saf. 2012;21:565–70 (**discussion 571–572**).
101. European Medicines Agency (EMA). Reflection paper on the use of artificial intelligence (AI) in the medicinal product lifecycle. 2023. https://www.ema.europa.eu/en/documents/scientific-guideline/draft-reflection-paper-use-artificial-intelligence-ai-medicinal-product-lifecycle_en.pdf.
102. Sentinel. Exploration of potential for Sentinel and PCORnet data linkage. 2015. https://www.sentinelinitiative.org/about/how-sentinel-gets-its-data/exploration-potential-sentinel-and-pcornet-data-linkage.
103. Lagerlund O, Strese S, Fladvad M, Lindquist M. WHODrug: a global, validated and updated dictionary for medicinal information. Ther Innov Regul Sci. 2020;54:1116–22.
104. European Medicines Agency (EMA). ICH E2B (R3) electronic transmission of individual case safety reports (ICSRs)—data elements and message specification—implementation guide—scientific guideline. https://www.ema.europa.eu/en/ich-e2b-r3-electronic-transmission-individual-case-safety-reports-icsrs-data-elements-message-specification-implementation-guide-scientific-guideline.
105. European Medicines Agency (EMA). Legal framework: pharmacovigilance. 2015. https://www.ema.europa.eu/en/human-regulatory-overview/pharmacovigilance-overview/legal-framework-pharmacovigilance.
106. European Medicines Agency (EMA). Implementation of the pharmacovigilance legislation. 2023. https://www.ema.europa.eu/en/human-regulatory-overview/pharmacovigilance-overview/legal-framework-pharmacovigilance/implementation-pharmacovigilance-legislation.
107. Kho AN, Cashy JP, Jackson KL, Pah AR, Goel S, Boehnke J, et al. Design and implementation of a privacy preserving electronic health record linkage tool in Chicago. J Am Med Inform Assoc. 2015;22:1072–80.

Authors and Affiliations

Salvatore Crisafulli[1] · Andrew Bate[2,3] · Jeffrey Stuart Brown[4,5] · Gianmario Candore[6] · Rebecca E. Chandler[7] · Tarek A. Hammad[8] · Samantha Lane[9,10] · Judith Christina Maro[5] · G. Niklas Norén[11] · Antoine Pariente[12,13] · Mulugeta Russom[14,15] · Maribel Salas[16,17] · Andrej Segec[18] · Saad Shakir[9,10] · Andrea Spini[1] · Sengwee Toh[5] · Marco Tuccori[1] · Eugène van Puijenbroek[19,20] · Gianluca Trifirò[1] on behalf of the Real-World Evidence and Big Data Special Interest Group of the International Society of Pharmacovigilance

✉ Gianluca Trifirò
gianluca.trifiro@univr.it

1 Department of Diagnostics and Public Health, University of Verona, P.le L.A. Scuro 10, 37124 Verona, Italy

2 Global Safety, GSK, Brentford, UK

3 Department of Non-Communicable Epidemiology, London School of Hygiene and Tropical Medicine, London, UK

4 TriNetX, Cambridge, MA, USA

5 Department of Population Medicine, Harvard Medical School, Boston, MA, USA

6 Medical Affairs and Pharmacovigilance, Bayer AG, Berlin, Germany

7 Coalition for Epidemic Preparedness Innovations, Oslo, Norway

8 Takeda Development Center Americas, Inc., Cambridge, MA, USA

9 Drug Safety Research Unit, Southampton, UK

10 University of Portsmouth, Portsmouth, UK

11 Uppsala Monitoring Centre, Uppsala, Sweden

12 Université de Bordeaux, INSERM, BPH, Team AHeaD, U1219, 33000 Bordeaux, France

13 Service de Pharmacologie Médicale, CHU de Bordeaux, INSERM, U1219, 33000 Bordeaux, France

14 National Medicines and Food Administration, Ministry of Health, Asmara, Eritrea

15 Department of Medical Informatics, Erasmus Medical Center, Rotterdam, The Netherlands

16 Bayer Pharmaceuticals Inc., Whippany, NJ, USA

17 University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA

18 Data Analytics and Methods Task Force, European Medicines Agency, Amsterdam, The Netherlands

19 Netherlands Pharmacovigilance Centre Lareb, 's-Hertogenbosch, The Netherlands

20 PharmacoTherapy, Epidemiology and Economics, University of Groningen, Groningen Research Institute of Pharmacy, Groningen, The Netherlands