Received: 30 September 2021 | Revised: 20 December 2021 | Accepted: 17 January 2022

DOI: 10.1002/psp4.12765

REVIEW

Advancing drug safety science by integrating molecular knowledge with post-marketing adverse event reports

Theodoros G. Soldatos[1] | **Sarah Kim**[2] | **Stephan Schmidt**[2] | **Lawrence J. Lesko**[2] | **David B. Jackson**[1]

[1]Molecular Health GmbH, Heidelberg, Germany

[2]Department of Pharmaceutics, Center for Pharmacometrics and Systems Pharmacology, University of Florida, Orlando, Florida, USA

Correspondence
David B. Jackson, Molecular Health GmbH, Kurfürsten-Anlage 21, 69115 Heidelberg, Germany.
Email: david.jackson@molecularhealth.com

Funding information
No funding was received for this work.

Abstract

Promising drug development efforts may frequently fail due to unintended adverse reactions. Several methods have been developed to analyze such data, aiming to improve pharmacovigilance and drug safety. In this work, we provide a brief review of key directions to quantitatively analyzing adverse events and explore the potential of augmenting these methods using additional molecular data descriptors. We argue that molecular expansion of adverse event data may provide a path to improving the insights gained through more traditional pharmacovigilance approaches. Examples include the ability to assess statistical relevance with respect to underlying biomolecular mechanisms, the ability to generate plausible causative hypotheses and/or confirmation where possible, the ability to computationally study potential clinical trial designs and/or results, as well as the further provision of advanced features incorporated in innovative methods, such as machine learning. In summary, molecular data expansion provides an elegant way to extend mechanistic modeling, systems pharmacology, and patient-centered approaches for the assessment of drug safety. We anticipate that such advances in real-world data informatics and outcome analytics will help to better inform public health, via the improved ability to prospectively understand and predict various types of drug-induced molecular perturbations and adverse events.

INTRODUCTION

Lack of efficacy, unpredicted toxicities, and unexpected adverse events (AEs) may compromise otherwise potentially successful drug development projects. To compensate for the associated costs and to protect relevant investments, several efforts have been initiated toward the development of approaches that aim to detect or predict such toxicities as early as possible. Innovative safety characterization efforts are therefore fundamentally required throughout the whole drug development process (see Figure 1).

Although randomized controlled clinical trial results and their systematic reviews remain the gold standard for evidence-based medicine,[1] recent years have seen the development of novel analytics in pharmacovigilance studies through postmarketing data mining. One key source of such data is information about AEs, collected regularly in spontaneous reporting repositories and monitored by

(inter-)national regulatory authorities worldwide. Major instances of such safety data collections include:

- The Sentinel Initiative of the US Food and Drug Administration (FDA),[2,3]
- The FDA Adverse Event Reporting System (FAERS; formerly AERS), maintained by the FDA,[4]
- The Vaccine Adverse Event Reporting System (VAERS), maintained by the FDA, co-managed by the Centers for Disease Control and Prevention,[5]
- The European database of suspected adverse drug reaction reports (EudraVigilance), maintained by the European Medicines Agency (EMA),[6] and
- VigiBase – the global database of individual case safety reports (ICSRs), maintained by the World Health Organization's (WHO) Uppsala Monitoring Center (UMC).[7]

These resources contain real-world data independent from randomized controlled trial settings and each contains information for millions of AE cases, emphasizing the distinctive informational contribution they deliver to drug safety and pharmacovigilance analytics.

Despite the value and plurality of this data, no universal proposal exists to jointly justify hypotheses of causal associations derived from the statistical inference and/or assessment of signals based on such data alone.[1] This is particularly important when considering the broader complexity of the regulatory drug safety environment. Critical assessment of current practices in drug approval and pharmacovigilance is a multifaceted challenge requiring constant monitoring and coordinating of a multitude of stakeholders and real-time content, including technical, scientific, regulatory, political, economic, and data management (sharing/access) aspects. To this end, frequent updates of relevant practices, standards, and guidelines, are published by regulatory experts, as well as other scientific contributions, often in collaboration with academic and/or industry partners.[8,9]

Yet, one key shift toward a more evidence-driven framework for drug evaluation has been combining AE information with additional data layers, such as chemical or molecular descriptors.[10–13] This paradigm accommodates a new theoretical standard for approaching pharmacology and postmarketing safety signals.[1] Ultimately, it provides an information-driven approach containing the evidence necessary to characterize postmarketing or clinical trial observations with a plausible casual base, relying on a hierarchy of additional data layers (including patient behavior, genomic, molecular, and chemical dimensions).

In this work, we reflect on some of the innovative potential presented by the molecular expansion of AE data. Whereas discussing the broader biomolecular context at large, we often refer to prime examples that are more specific. First, we briefly summarize the analytical landscape pertaining to the study of AEs. We then seek to identify key data augmentation opportunities (see Figure 1) and focus on the novel insights that an ensemble of molecular information may provide. We delve into these perspectives from a systems pharmacology point of view and also highlight examples with respect to advanced technologies, such as machine learning (ML) and big data handling. Moreover, we investigate existing software or service options provided by industry or academia that may help understand implications underlying disease and therapeutics based on the computational analysis of real-time AE data with clinical and molecular knowledge. Finally, we discuss the impact of these developments on the progress of future approaches, especially with respect to the assessment of the safety of therapeutic interventions.

BACKGROUND

In this section, we briefly outline the key dimensions pertaining the analysis of AEs.

Quantitative approaches and challenges

One of the most commonly used approaches to analyze AE data relies on disproportionality measures, which are metrics aiming to quantify the relative congruence of pairwise entity relations as based on their co-occurrence in subsets of AE cases.[10,14,15] These methods rely somewhat on the premise that as the number of AE cases for an observation accumulates over time, there comes a point where the signal may strongly suggest causation—without, however, demonstrating it. Advantages of these quantitative estimates of association include that they are computationally efficient, easy-to-understand, and frequently implement additional statistical testing techniques, such as χ^2 and Fisher's exact tests.[14,16]

Next to disproportionality analysis, Bayesian correlation models may also be used to assess the degree of congruence, with various publicly or commercially available software programs generating both types of scores by default.[10,14,15,17] Other approaches that have been used include time-series forecasting, item-set, and association rule mining, as well as more sophisticated ML.[10,18–20] Which algorithm is ultimately more favorable depends on the task at hand,[18,21,22] as the resources and types of questions examined may be highly variable.[10,14,15,23–25] These may include AE signal detection, AE identification, and AE prediction, as well as challenges like drug (off-)target identification, drug interaction analysis, drug

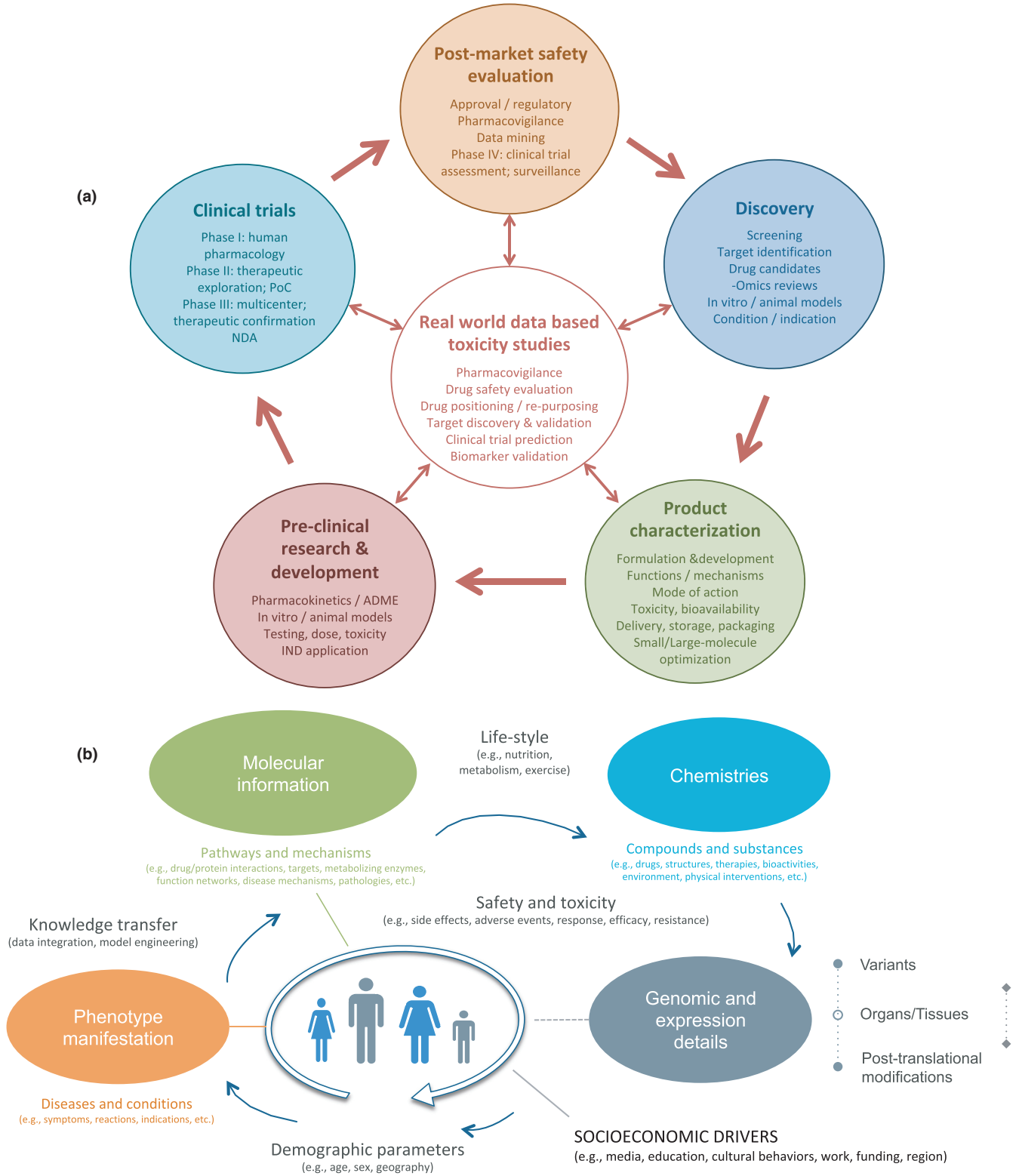

FIGURE 1 Innovative postmarket data mining and novel methods to analyze and detect toxicities are key to (pre-)clinical modeling. (a) The standard drug development process is aligned to characterization steps spanning (animal) model screening, target identification and validation, product formulation and development, preclinical model optimization, preclinical and clinical testing, approval and postmarket monitoring. (b) Advanced pharmacovigilance studies may benefit from data augmentation opportunities that consider multiple layers of information. ADME, absorption, distribution, metabolism, and excretion; IND, investigational new drug; NDA, new drug application; PoC, proof of concept

indication prediction, drug repositioning, drug (class) interaction discovery, or characterization of pathological mechanisms. Many of these approaches have become increasingly quantitative and have to deal with a number of limitations (see Table 1). A primary challenge has been data standardization and efforts to address this issue have used knowledge engineering strategies that help provide AE resources in an easily usable form.[12,26–28]

Toward evidence expansion

Although such efforts have helped accelerate drug safety characterization by enabling the systematic analysis of AE case studies (e.g., refs. 29–41), research frequently highlights the advantages that additional scientific data or advanced feedback from experts may provide. To this end, some have used Natural Language Processing (NLP) to mine information directly from text (such as clinical narratives, electronic health or medical records, health insurance, or billing data),[10,15,17] whereas others have incorporated evidence coming from additional layers of reference knowledge in their data mining efforts.[10,12,15,17,23]

These developments have led to an evidence amalgamation paradigm, where safety questions are addressed in novel ways. Indeed, many apply well-established pharmacometrics and mechanistic modeling methods, whereas others incorporate analysis of biological networks, side effect information coming from drug labels, exploitation of molecular information, and other data-driven strategies to analyze AEs and drug interactions.[11,12,42–44] Example data augmentation layers may include:

- Drug-centric information (such as therapeutic class, chemical, side effect, and target or other drug interaction descriptions),
- Molecular knowledge (such as target- and metabolizing-partners, binding affinities, protein interactions, expression, pathways, pathogenic mechanisms, and genomic mutations) and/or
- Ontologies/hierarchies describing phenotypic manifestation relationships.

Importantly, many of the above implications also apply to pharmacovigilance/safety analysis of other therapeutic modalities, such as vaccines, where challenges are largely similar.[45,46]

The data augmentation route

Incorporation of reference data in drug safety analytics aims largely to provide additional support to understanding the (biological) plausibility of identified (AE) signals.[10,15,17,23] Reference information sources may often contain organized data (e.g., regarding chemistry and [patho-]physiology) coming from multiple resources, such as DrugBank,[47] PubChem,[48] Off-sides or Two-sides,[42] SIDER,[49] STRING,[50]

TABLE 1 Sample challenges and limitations hindering the analysis and management of adverse event data

AE content	Knowledge engineering	Data mining
Lack of specific severity grading for described conditions (indications, reactions).	Scalability (and increasing data sizes)	Multi-pharmacy (co-medications)
Disambiguation of symptoms (e.g., disease vs. reactions)	Lack of universal benchmark data	Comorbidities
Polypharmacy and drug interference	Visualization	Statistical normalization / control background
Detail regarding patient/event history (+)	Management of dictionaries and ontologies	Semantics and hierarchies
Missing, incorrect, or vague information	Data standardization	Pattern identification / synergistic effects
Handling of duplicates and/or multiple reports per case (e.g., follow-ups)	Data safety, ownership and transparency	Signal does not provide proof of causation
Reporting bias (over-/under-reporting)	Disambiguation and synchronization of reports between systems	Unbalanced data sets
Data entry and coding	Different and changing reporting requirements between systems	Definition of (risk) populations / cohorts
Difficulty in verifying AE occurrence	Inconsistent database structure	Observation of signals over time
Not all AEs are reported	Biomedical plausibility	Biomedical plausibility

Note: (+)Detailed patient/event history may include treatment duration, cycles, timing, dosage or previous therapies and co-morbidities, de-/re-challenge information, demographics, disease stage, laboratory, and clinical parameters.

Abbreviation: AE, adverse event.

UniProt,[51] Reactome,[52] KEGG,[53] BindingDB,[54] Open Targets,[55] ATC,[56] MedDRA,[57] and so on.

Other used resources may be only partially structured or not at all. Examples of such data sets include biomedical literature, patent contents, social media and health web-forums, clinical trials, as well as real-world outcome data from electronic health or medical records, federal/national repositories, patient event narratives, or collections of hospital records.[10,15,17,23,58] Evidence derived from such data-rich sources provides complementary strategies to approaching key safety tasks, like confirming or rejecting AE knowledge, discovering or evaluating AE-related hypotheses, and ranking or prioritizing AEs.

Notwithstanding, data extraction from semi-structured or unstructured content may be complicated, for example, from the use of nonstandardized drug names,[59] and may require the application of advanced text-mining techniques (such as NLP). Moreover, data augmentation practices may have to cope with multiple demanding knowledge engineering challenges (see Table 1).

Finally, novel resources used in retrospective AE analytics may also include genomics data, such as patient-specific or genomewide association study (GWAS) data, thus emphasizing the potential for pharmacogenetics and pharmacogenomics capabilities in pharmacovigilance studies.[10,15,43,60]

MOLECULAR EXPANSION

As already mentioned, analysis of AEs alone suffers mainly from shortage of insight regarding potential molecular etiologies underlying observed signals. Yet, recent pharmacology, chemo-, and bio-informatics advances have allowed information about biological systems to provide mechanism-based context in drug safety assessment and prediction studies.[10,61] These systems are often represented in the form of biological networks describing the relationships between interacting components, such as genes, proteins or drugs—an approach commonly utilized in biology and medicine to provide larger mechanistic contexts as well as emergent associations.[10,62] For example, this approach has been used for the systematic characterization of proteins affected by drugs[63] (e.g., using drug similarity[13] or protein interaction network analysis[64]), an important step in exploring the drug mode of action hypotheses. Indeed, the systems approach has proven useful in pharmacology too, often complementing available experimental data and knowledge about very specific interactions, or lack thereof. Combined with integrative data augmentation, systems pharmacology approaches have been applied to achieve key goals, such as predicting AEs or understanding underlying mechanisms.[10,15,65]

Integrative molecular interrogation of adverse events

Molecular Analysis of Side Effects (MASE)[12] is a prototypic example of these integrative systems approaches that allows enrichment of existing AE information with drug target knowledge and additional levels of evidence.[10,15] Perhaps its most characteristic elements are that it provides a standardized approach to AE exploration and a strategy to characterize target-phenotype associations directly from their occurrence in real-world data, as opposed to using indirect drug-centric properties or observations (e.g., refs. 26, 49). This type of analysis has been termed "Target Adverse Event" (TAE) analysis to distinguish it from the current drug-focused standards of pharmacovigilance. Importantly, it permits phenotypes to be analyzed and compared at the level of any molecular perturbation or specific clinical and molecular feature (including drugs, drug-classes, targets, transporters, or enzymes) and enables the user to efficiently generate hypotheses about cause-effect relationships, or lack thereof.[12]

Its ability to systematically facilitate retrospective computational AE analytics (e.g., refs. 29, 31, 32, 34–36) comes together with other project-specific data standardization and augmentation efforts (e.g., refs. 30, 33, 37–39, 66). However, the distinct ability to provide a biomolecular rationale to detected safety signals is critical to pharmacovigilance and regulatory scientists use MASE in this process (e.g., at the FDA[12,17,67,68]).

Inspired by the plethora of utilities enabled by the systematic molecular interrogation of AEs, we highlight the next few important perspectives derived from this context (see Figure 2):

- ***Systems pharmacology***: In a systems pharmacology view of drug action, drugs interact with multiple primary and secondary targets and pathways, existing within a complex network and mediate therapeutic response via intended "by design" interactions ("on-targets") or via unintended ones ("off-targets").[10] Accordingly, the systems pharmacology approach will combine pharmacokinetic, biochemical, and systems biology information into a unifying model that comprises of parameters interlinked by connecting the underlying (patho-)physiology and drug mode of action knowledge examined each time.[69] The MASE strategy is therefore an essential paradigm toward enabling the study of AEs via the prism of systems pharmacology, consistent with the position that drugs interact with each other and with multiple targets, as opposed to the simpler "one-drug, one-target" approach. Moreover, it can facilitate a "bidirectional" approach to systems pharmacology in which the "reverse" course to identify mechanistic differences, to explain

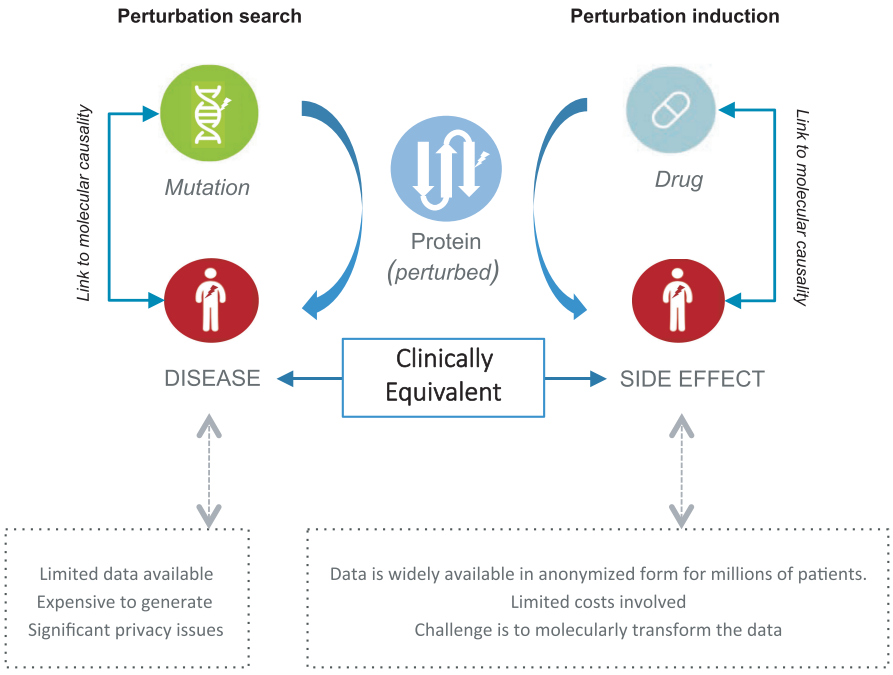

FIGURE 2 Safety reports contain clinical outcomes data for millions of real-world patient treatments with details about drug-induced phenotypes (i.e., side effects) that can be linked to chemical and molecular information about drugs and their targets. Similar to model-based approaches, perturbation studies are key to identifying the molecular underpinnings of human disease, via the process of comparing individuals and/or cohort-populations for differences at the molecular-level. Therefore, linking the molecular functions targeted by therapeutics to resultant side-effect phenotypes—and this for millions of patients—should enable the computational dissection of the molecular mechanisms underlying these effects. This approach for the molecular analysis of side effects (MASE) enables production of datasets valuable to the molecular characterization of human disease, directly from real-world treatment outcomes/safety data. This strategy provides a standardized approach in a variety of contexts; prototype results can be accessed via Molecular Health's Effect[86] platform. Figure adapted from Soldatos et al[12]

AE manifestations, or to understand therapeutic effects is driven by insights derived from the signals and scores pertaining to annotated molecular components, such as TAEs (see Figure 2). Finally, in addition to identifying and explaining idiosyncratic AEs, innovative tools may in the future include advanced models and additional parameters that will help reliably address more complex challenges, such as genetic, lifestyle, and comorbidity-induced AE susceptibility.

- ***Hypothesis generation/validation***: One important feature underlying molecular integration approaches is the resolution provided in terms of different target, reaction, or drug classes. This aspect influences the degree to which target-specific phenotype associations can be systematically screened and, in turn, the accuracy of hypothesis generation (or confirmation) strategies that can be associated with such phenotypic profiling (see Figure 2). Although statistical characterization derived directly from AEs may be more definitive,[12] the MASE approach enables a new strategy based on the induction of target-specific perturbation using drug-treatments (i.e., TAEs). Two examples toward the prospective prediction of side effects in this way rely on comparative drug-safety analysis[12]—one approach argues that comparison of targets and their agonism/antagonism modulation-states may reveal in many cases dissimilar phenotypic consequences, whereas the other argues that drug side effect dissimilarity could be attributed to differences in their molecular landscape (i.e., perturbation of targets specific to each drug may affect different molecular systems). Importantly, such perturbation studies are key to understanding not only side effects but also human disease, because side effects can for all intents and purposes be seen as drug-induced disease states.
- ***Clinical trials***: An overarching conduit to this direction is that the large-scale phenotypic profiling of molecular targets is derived directly from clinical data of human patients.[12] One important aspect of this feature is that it provides a promising approach in developing quickly testable molecular hypotheses regarding the potential clinical utility of a drug (see Figure 2). One MASE analysis showed, for example, that inhibition of beta-adrenergic receptors might improve mortality of patients with skin cancer,[1270]—in this instance, the AE study was conducted by measuring the disposition of death as outcome in two treatment groups of a specific cohort. In all, this can be seen as an example that represents a virtual trial, enabled by the side-by-side

investigation of clinical features associated with any two (or more) entities, as derived from the molecular profiling of individual patient prescriptions. Providing additional support in clinical trial model design or study result prediction, molecular profiling of phenotypes permits the systematic analysis and comparison of TAE profiles within predefined cohorts of patients.

- ***Machine learning***: In a recent study, TAE profiles were used to predict AEs at the time of drug approval.[71] In this approach, TAE profiles generated by MH EFFECT (a technology encompassing the MASE methodology), were used to predict postmarket label changes by incorporating molecular features into ML. Integration of such domain-specific knowledge may provide further features into advanced ML models studying AEs by encoding chemo- and bio-descriptors about the physical, chemical, and biological characteristics of the involved components of interest (e.g., drugs).[72] Analyzing large and complex datasets and the ability to discover novel and hidden but precious knowledge in data are key advantages to using ML techniques and may thus be applied to a wide range of predictive safety settings.[15,23,72,73] Notably, in the context of the coronavirus disease 2019 (COVID-19) pandemic, the FDA issued a landmark emergency use authorization to an enhanced artificial intelligence (AI)-powered tool to predict AEs.[74–79]
- ***Patient centered focus***: Rather than using molecular interrogation of AEs to assess biological plausibility of detected signals, the biomolecular systems level representation can be useful also in the analysis of individual AE case studies, helping to predict or explain a patient's (observed) response.[11] Such personalized AE prediction and therapeutic safety assessment tools may permit deeper understanding of AE circumstances, especially with respect to drug interaction and polypharmacy considerations.[11]

Selected case studies that demonstrate examples of approaches addressing the above perspectives are listed in Table 2. The listed summaries in Table 2 discuss how each study benefited from the integrative methodology and what data were used to augment AE reports in each case. Specifically, the list provides information that reflects on each study's broader goals and results, and on the novel insight potential facilitated by the respective type of approach used for the examination of AEs: analyses that were performed per study, novel conclusions that were derived with respect to the AE focus category highlighted every time, as well as other implications deemed important to emphasize in the context of handling or organizing AEs.

It can be argued that most recent clinical advancements have emerged from analysis of patient-derived molecular data as opposed to model organism-based observations.[12] Integrating AE-data with molecular knowledge at the patient level expedites the identification of safety problems via a prospective approach,[11,12] but may in addition provide a rational data stream to collecting a broad array of critical system-level insight required for the realization of more comprehensive, evidence-based decision support for personalized and precision medicine.[62,80] Altogether, prospective initiatives toward feed-forward mechanisms integrating patient-specific data on gene, protein, and other interactions with clinical knowledge of disease and pharmacology could provide an extended context to network medicine toward understanding AEs and treating illnesses, toward identifying novel disease pathways, and predicting patient drug responses.[62]

Integrative software platforms

Given its importance, we explored the degree to which software solutions exist that adopt the molecular data integrative approach for AEs. Although recent advances have highlighted a number of quantitative models or approaches to understand and predict various types of drug-induced AEs, a key challenge is the large amount of data involved—especially when it comes to incorporating translational, chemical, and molecular dimensions in an integrated software platform. In addition, the variety of applications that this approach is useful for makes it difficult to make a single defined point of access for one only service/analysis scenario. It is, therefore, not straightforward to directly compare AE analytic platforms and Table 3 provides an overview of some parameters that may be useful to assess the suitability of such complex software tools and solutions.

A comparison between AE technologies from the public and private domains shows high variability and diverse use-case scenarios—to summarize main aspects and key software features, we selected representatives from each sector, which we split into three application classes:

- **Group A**: Regulatory portals, enabling visual exploration of original AE data—these include AE browsers, such as the FAERS Public Dashboard,[81] or the VAERS[82] and EudraVigilance[6] gateways.
- **Group B**: Public data projects, tools, and software packages for mining or analyzing AEs—these include the OpenVigil suit,[83,84] the ezFAERS,[27] AERS*Mine*,[28,85] and AEOLUS[26] repositories.
- **Group C**: (Semi-)commercial solutions offered by companies, such as Molecular Health's (MH) Effect,[86] Evidex from Advera Health,[87] OFF-X from Clarivate's BioInfoGate,[88] FDAble,[89] or Elsevier's PharmaPendium.[90]

TABLE 2 Selected studies that demonstrate examples or discuss ways to gain potentially new insights from the interrogation of AEs at additional data layers

Description of case study (+) (Summary of example with respect to key AE extension perspectives)	
Title citation; year	*Src* activation by *β*-adrenoreceptors is a key switch for tumor metastasis[70]; 2013
Focus category	*Virtual perturbation studies*
Study's domain	Oncology
Purpose/emphasis	To examine the potential clinical impact of model (e.g., experimental, laboratory) findings.
Method/approach	Retrospective computational mining; Statistical analysis Bayesian inference
Molecular data	Indirect; use of ATC classification hierarchy to automatically identify beta-blockers/AEs.
Hypothesis/goal	Molecular hypothesis validation using outcome data from human AE observations.
Implications	*Main points*: Construction of virtual AE cohorts with specific patient characteristics; Clinical setting based on drug MoA parameters; AE based evidence for hypothesis confirmation. *Result/conclusion*: Beta-blocker usage associates with reduced cancer-related mortality.
Title citation; year	Association Between Serotonin Syndrome and Second-Generation Antipsychotics via Pharmacological Target-Adverse Event Analysis [32]; 2018
Focus category	*Systems pharmacology*
Study's domain	Serotonin syndrome (SS)
Purpose/emphasis	Insight into the molecular mechanisms of SS
Method/approach	Systems pharmacology; Data mining Disproportionality (PRR)
Molecular data	Drug interactors (T), Pathways (P)
Hypothesis/goal	Molecular characterization of AEs can help determine relationship between SS and SGAs.
Implications	*Main points*: Comparison of targets annotated in AE groups; Hypotheses generated based on signals of targets perturbed in the AEs of each defined cohort (TAE); Use of information about drug-target pharmacological action (induction/inhibition) to define cohorts; AE grouping based on drug categorization (SGAs, SSRIs); Inquiry for confounding or synergistic effects. *Result/conclusion*: Identification of *5-HT1A* agonism and *5-HT2A* antagonism as potential mechanism of SGA-associated SS.
Title citation; year	Adverse Event Circumstances and the Case of Drug Interactions[11]; 2019
Focus category	*Patient centered focus*
Study's domain	Circumstances of AEs
Purpose/emphasis	Preventable AEs, DDIs, personalized therapeutic optimization
Method/approach	Systems level analysis; Graph representation Network co–occurrence
Molecular data	Drug interactors (T), Pathways (P), DDIs (D)
Hypothesis/goal	Molecular level interrogation of therapeutic setting can help optimize use of co-medications.
Implications	*Main points*: Individual patient-specific clinical case models; Detection of avoidable DDIs; Determination of molecular level (TAE) interactions (i.e., via targets, enzymes, or pathway load). *Result/conclusion*: Many AEs may be explicable by avoidable scenarios and/or attributed to previously known factors.
Title citation; year	Target Adverse Event Profiles for Predictive Safety in the Postmarket Setting[71]; 2021
Focus category	*Machine learning*
Study's domain	Prediction of unlabeled adverse effects at the time of drug approval
Purpose/emphasis	Prediction of post-market label changes using molecular features
Method/approach	Ensemble of classification methods; Data mining Disproportionality (PRR)
Molecular data	Drug interactors (T)
Hypothesis/goal	Parameters from the molecular layer may provide important information to predictive safety.

(Continues)

TABLE 2 (Continued)

Description of case study (+) (Summary of example with respect to key AE extension perspectives)	
Implications	*Main points*: Comparator drugs determined via shared target information; Normalization of reaction/effect terms into designated medical events; Direct target-to-reaction/effect profiling based on AE co-occurrence (TAE); Generated profiles used as input features to prediction models. *Result/conclusion*: Machine-learning features derived from target level profiles are an important information component to predictive performance. Extensions to this modeling approach include features from chemical structures of lead proteins or compounds, binding assay results, gene expression profiles, comparator selection based on (sub-) structure similarity and/or shared pathway signaling, etc.

Note: (+)Case studies sorted by year of publication (ascending order); all studies utilized the US Food and Drug Adverse Event Reporting System (FAERS) data, organized based on the Molecular Analysis of Side Effects (MASE) extension principle. (T) Information about drug interactors refers to targets, metabolizing enzymes, carriers, and transporters (and may include potential pharmacokinetic and pharmacodynamics aspects). (P) Should a drug interact with components that belong to a certain (functional and molecular) pathway network, then that pathway may be transitively annotated via its (drug, or drug interacting) members. (TAE) Indicates type of target adverse event (TAE) analysis implementation. (D) Between medication products reported for the same patient case.

Abbreviations: AE, adverse event; ATC, anatomic therapeutic chemical; DDI, drug-drug interaction; MoA, mode of action; PRR, Proportional Reporting Ratio; SGA, second generation antipsychotic; SS, serotonin syndrome; SSRI, selective serotonin reuptake inhibitor.

From the above groups, we exclude platforms, such as Oracle's Argus[91] or Emperica[92] that are tailored to create a wide set of automatically generated analytical outputs and reusable tables and graphs. In general, most of groups A and C solutions utilize a dashboard approach to visualizing data into informative tables and/or matrices, whereas group B projects tend to develop their own graphical user interfaces (GUIs). Table 4 summarizes key aspects that may affect typical GUI or web-service provision characteristics when it comes to displaying or organizing AE information.

Although group A usually relies on raw content without extensive statistical support, openFDA[93] provides indexing and formatting services to FDA data and some public front end examples provide enhanced access to FAERS data,[94] including disproportionality scores and graphic options. On the contrary, groups B and C tend to apply data transformation processes to strengthen the integrity of the available information, as well as the capability to provide ontological aggregations and high-dimensional cohort-based analyses.

However, challenges in available compute capacities may often limit AE mining resources. For example, group B options may come with less intuitive and slower GUIs, and may apply techniques to reduce the analytical load (see also Table 4). Finally, group C incorporates additional data sets that enable extended AE analytics and visualization. These solutions may involve big data and cloud-native technologies, may come as "software as a service," may provide integration options to external infrastructures, or may be customizable to customer projects.

Importantly, group C enables the molecular analytics paradigm, each from a different standpoint:

- MH Effect[86] is a computational technology for the molecular analysis of real-world data,
- Evidex[87] is a platform for accessing, analyzing, and tracking drug safety issues,
- OFF-X[88] is a portal for translational drug safety intelligence, and
- PharmaPendium[90] provides comprehensive data for drug safety and risk assessment.

Nevertheless, little emphasis has been paid to integrating molecular data in a single platform and most of these options come as a constellation of confederated data and management solutions. In this sense, only MH Effect ultimately implements a pure MASE implementation, whereas, among the discussed group C options, only MH Effect, OFF-X, and Oracle solutions are routinely used by (or engage in research/material transfer agreement with) regulatory authorities.[17]

DISCUSSION

Augmented AE analytics provides an innovative new approach to drug safety risk assessment. This work focused on perspectives enabled by the integrative paradigm whereby molecular data expansion can be used to enhance the insight gained through AE analysis. Although recent, these integrative advances accommodate the development of improved systems' strategies to support contemporary pharmacology and personalized medicine.[95] In our view, the molecular data integration paradigm markedly enhances AE data analytics by providing insights that help understand drug-target (protein) interactions, both desirable and undesirable, within the context of a biological system. Finally, we argue that these developments bear significant implications for drug development, regulatory review, and clinical practice.

Aspects that may challenge the performance of "traditional" AE analytic approaches include both qualitative and quantitative issues. Examples of qualitative

TABLE 3 Selected features for assessing and comparing AE data analysis software and platforms

Category	Feature	Description (value, parameters, examples)
Audience	Availability	Is the technology a commercial product or an academic/open-source solution? What is the licensing status?
	Target groups	May refer to (a combination) of regulatory, interdisciplinary health/life-experts, medical, clinical, "pharma" or other industry stakeholders, including clinical pharmacologists, medicinal chemists, pharmacovigilance professionals, preclinical toxicologists, translational researchers, epidemiologists, health policy makers, or even the public.
Core capabilities	Analytical power	May refer to signal detection capabilities, such as: • The mathematical methods to define the degree of statistical association between entities (e.g., drug and side effect), including disproportionality/frequentist or Bayesian methods, identification of confounders, and so on. • Or, the use of sophisticated analytical technologies such as knowledge graphs, ML(+), and BD(+).
	Content diversity	Key integration parameters include: • Data breadth: does it integrate drug/chemistry (e.g. structure, synonyms, descriptions) or molecular (e.g., targets, metabolizing enzymes, transporters, signaling/metabolic pathway models, DDIs(+)) knowledge, preclinical toxicity, label information, evidence from literature, etc. • Datasets (e.g., public or proprietary data, RWD(+), AEs(+), CTs(+), and so on) • Data depth (e.g., dictionaries, ontologies, hierarchies, resolution in GUI(+))
	Key functionalities	Types of analyses may address multiple use cases, such as: • Drug safety signal detection and prediction • Comparative/competitive safety analysis • (In-)validation / assessment of emergent signals • Design of rational drug combinations • Molecular analysis of emergent safety signals • Genotype to phenotype analysis • Biological plausibility analysis • Drug repositioning • Prediction of (target) AE(+) profiles and DDIs(+) • Rational design of combination therapies
Systems pharmacology	Systems level molecular analysis of real-world drug safety data	Main abilities include: • Integrates AEs(+)/RWD(+) with molecular knowledge • Analyzes clinical effects of drug targets (e.g., safety profile) • Analyzes potential clinical effects of novel drug targets (e.g., using neighborhood analysis) • Analyzes the clinical effects of biological pathways • Examines patient level drug-mode of action models • Analyzes DDIs(+) at the level of biological systems • Compares clinical safety profiles for independent drug targets or pathways • Constructs and compares independent cohorts of patients based on user-defined clinical and molecular parameters
Technology	Easy to understand and user-friendliness	May include aspects such as intuitiveness of GUI(+), search/browse utilities, speed and responsiveness, graphical data visualization functionalities.
	Information access	May refer to aspects regarding: • Web accessibility, report generation and/or data download capabilities • Format (e.g., DB(+), cloud-native, BD(+), SaaS(+)) • Can it be integrated or customized and in what scope (scalable or consultancy/project-based approach)?
	Analytical features	May refer to safety signal management, provision of advanced search capabilities, access to patient level case reports, cohort building functionalities, depth of (GUI(+)) exploration/data resolution (e.g., seriousness, demographics, outcomes, hierarchies, synonyms, etc.), number of statistic/metric scores, embedded analytical scenarios, advanced/customizable graphical visualization options (charts, dashboards, etc.), molecular mechanism/drug-MoA(+) insights.

(Continues)

TABLE 3 (Continued)

Category	Feature	Description (value, parameters, examples)
	Quality	May refer to quality management and control procedures, such as proprietary data integration and transformation processes, benchmarking, or compliance with industry certified quality assurance standards.
	Validation	Main points of endorsement include: • Use by regulators • Industry acceptance (customer projects and market usage and advocacy) • Patent protection • Peer reviewed studies and recognition

Abbreviations: AE, adverse event; BD, big data; CT, clinical trial; DB, database; DDI, drug-drug interaction; GUI, graphical user interface; ML, machine learning; MoA, mode of action; RWD, real-world data; SaaS, software-as-a-service.

TABLE 4 Aspects that may reflect on the structure of the discussed web-based services or GUIs(+) with respect to the organized exploration of complicated AE information

GUI(+) or service parameter	Aspect description (examples / discussion)
Search forms	Identifying sets of AEs with specific characteristics is one main GUI or service functionality. Some of these options may be simpler or more complicated. Commonly, searches rely on text queries matching drug or symptom names. Often, advanced search forms are used to include more structured queries utilizing logical operators and complex hierarchies. Finally, some allow for customizable result views and/or cohort definition options, using different content characteristics as filters.
Data dissemination	Main dimensions when exploring AE content (in turn, display pages, or views) include information regarding reported outcomes, seriousness, timelines, demographics (e.g., age, sex, or geographical/regional origin), and/or symptoms (indications or reactions) the resolution and detail with which such condition layers are described and/or presented is variable among the different tools.
Content resolution	In some cases, aggregation of results may rely on pre-calculated views rather than on on-demand computations. Access to individual AE case-level descriptions may be available too.
Data irregularities	May refer to unprocessed data or irregular representation structures, including free-text descriptions (e.g., generic vs. product drug names, medicinal products vs. active substances).
Data delivery	Generated data may be shared via views organized in web-reports, and sometimes may be exportable into a commonly downloadable form (such as ASCII, Excel, JSON, CSV, or otherwise formatted files).
Analytic resources (mining and storage)	Addressing availability of resources (or rather, the lack thereof) is key—whether it is about handling (limited) computational capacity or (large) computational requirements. Lead ways to handle the topic may include: (a) narrowed breadth: some web services (or GUIs) apply techniques to reduce the analytical load (e.g., force the user to specify the exact multi-parameterized combination of records to be analyzed each time, or restrict the computation to a limited number of top only records); (b) resort to benefits that rely on third party solutions (e.g., OpenVigil FDA[83] that runs over openFDA's services[93]) comes with a limited number of embedded comparison scenarios, whereas the OpenVigil 2[110,111] GUI comes with none, as yet. In comparison, FAERS data (enabled also by openFDA[93]) shared by the Open Targets[55] resource—a platform maintained by a larger consortium—come with both select filters and statistics, over records mapped via own computational pipelines.[112]
Visualization provision	Above factors (including data and resource management) may directly affect visualization decisions, and on the degree that data delivery is organized in a dynamic fashion (e.g., in terms of charts, tables, or graphs and whether these are configurable by the user). In comparison, ezFAERS[27] and AEOLUS[26] are available mainly as static downloadable options.

Abbreviations: AE, adverse event; FAERS, US Food and Drug Adverse Event Reporting System; FDA, US Food and Drug Administration; GUI, graphical user interface.

limitations include possible misstatement of indications, cases attributed to additionally co-administered medications, difficulty for risk assessment of over-the-counter drugs and supplements, food-drug interactions, uncertainty that a given AE is causally related to the reported drug product, or due to other risk factors, such as patient characteristics (e.g., age, smoking status, or history of coronary artery disease).[31,32,34,35,96]

Whereas Table 1 lists many parameters that can be largely limiting, it also provides only a generalization and

some aspects may not apply to all methods, tools, and/or data sets. For example, the various repositories may handle reports differently, with FAERS allowing multiple reports per AE, whereas VigiBase allows only one (e.g., latest active data per case, with most up-to-date valid information). VigiBase also flags possible duplicates, and highlights cases coming from FAERS. Other aspects include implications, the handling of which cannot possibly address all potential scenarios uniformly (e.g., detailed information about treatment setting [such as the number of cycles or whether first- or second-line therapy]) and, therefore, a certain modeling compromise may be required.

Assessment of the scientific validity of the study that generated the AE data is also important (e.g., risks of bias, imprecise measurements, and so on). For example, there are different considerations to account for when AE information comes from spontaneous reporting repositories, or from the results of a carefully controlled and monitored phase III randomized control trial. Technically, patient behavior can play an important role here as well, such as doubling up on missing doses or misuse of modified release formulations, which can contribute to AEs and may be hard to control or capture.

To this end, trustworthy real-world data summaries are warranted not only to generate reliable real-world evidence but also to provide quality estimates for confident decision making. In support of evidence-based medicine, the GRADE approach provides one such framework to the systematic assessment of the certainty in a body of evidence leading to clinical practice recommendations.[97]

Coordinated strategies are, therefore, vital for overcoming regular obstacles that impede the widespread implementation of systematic approaches to the analysis of AEs (e.g., ref. 98–101). Such strategies may include collaborations between multiple healthcare stakeholders, including policy, regulatory, scientific, technological and commercial entities, and the general public. They may also involve the multilayered integration of sources of evidence in drug safety studies, including a range of parameters spanning from demographic, socioeconomic, and behavioral dimensions to lifestyle, molecular, genomic, and expression details (see also Figure 1).

Interdisciplinary sources of data are important to the inference of casual mechanisms, because the actual reasons underlying the occurrence of an AE ("real causes") cannot always be proven and, usually, accompanying evidence may be limited by the scope of observations.[31,32,34–36] Data integration thus helps to extend the application and interpretation of causal inference approaches that may be used to evaluate hypothesized cause-effect relationships derived from observational evidence (e.g., Bradford-Hill criteria[102,103]). Biological plausibility analysis plays a pivotal role in this direction, which may be further complicated from the fact that a given AE etiology may be attributed to multiple factors (e.g., treatment, interaction with comedications, inherent disease, diet, smoking, etc.), rather than a single "prominent cause."

In this work, we discussed in more detail the aspects pertaining to the expansion of AE data sets with molecular data. In several instances, we used the term "molecular" in a somewhat vague fashion to denote the broader context that can be incorporated (e.g., chemical or protein substructures, binding assays, protein interaction networks, gene expression profiles, and so on), whereas in others we focused on more specific dimensions (such as information about drugs and their interactions with targets, metabolizing enzymes or with other drugs, about biomolecular pathways, and hierarchical therapeutic classes). Table 2 summarizes examples of four such studies that incorporated different levels of molecular detail, each highlighting a different perspective category enabled.

Indeed, the impact of integrating molecular level evidence on AE analytics has been successful in advancing the pharmacology systems approach, and enabling better identification of risk factors.[12,95] Routine signal assessment in this manner may further help improve regulatory efficiency, providing not only speedy capture, management, or prioritization of potential safety issues but also opportunities for real-time interventions, including the timely identification of populations at risk, of product contamination patterns, or of areas requiring public health education and assistance.[11,17]

Overall, the systematic interrogation of AEs at the molecular level can have multiple applications, including both the identification of causal mechanisms, or separating causal explanations of AEs from noncausal or alternative explanations. As indicated by the MASE paradigm, TAE profiles may, for example, help predict AEs, perform virtual "in-patient" perturbation experiments, encode advanced ML models, analyze individual patient cases, and in generating or validating hypothesis regarding pathological disease mechanisms[11,12,32,71] (see also Table 2). Importantly, such inferences based on molecular targets can quickly be explored with laboratory-based analysis or further human studies.

In the future, we expect that AE-based studies will use more biomolecular and real-world data sets, emphasizing the importance of refined knowledge engineering in drug safety and model-informed drug development.[104,105] In addition, we look forward to more MASE comprehensive platforms that will enable intuitive integrative analytics. Tables 3 and 4 list some important features to help assess such AE data analysis software. This is of primary importance also from a regulatory point of view, as it becomes challenging to "handle" an increasing abundance of tailored digital tools (or methods).[106] Furthermore, we believe that some promising approaches to advance

molecular AE mining will be achieved from the greater use of annotation benchmarks regarding biomedical concepts and mechanisms.

In addition, we expect that molecular analysis of real-world data and outcome analytics will generate important opportunities, forwarding more predictive (as opposed to reactive) responses and informed guidance to clinical trials and personalized medicine. The potential of AI in biomedicine, health care, and quality of life is vast in this perspective[107,108] and modern advances in big data and ML are especially important to the feasibility of incorporating molecular information in these applications. In comparison to other approaches, these efforts will be able to benefit from the combined advantages of:

- Outcome data that provide an augmented source of real-world scenarios regarding drug use and combinations, and
- ML techniques, capable of coping with larger volumes of data, as well as of capturing multilayered features from complex data.

Specifically, AE data contain phenotypes and conditions not studied in randomized controlled trials, and also include information for many more patients.[12] Importantly, safety reports and disease models are linked directly via human information (see also Figure 2). As the need for more human-specific disease models increases and reliance on model organisms decreases, molecular analysis of real-world data may therefore provide a seamless new path to approaching drug safety and informing drug development.[12,109]

In summary, molecular data expansion provides an elegant way to extend mechanistic modeling, systems pharmacology, and patient-centered approaches for the assessment of the safety of therapeutic interventions. We anticipate that such advances in real-world data informatics and outcome analytics will help to better inform public health, via the improved ability to understand and predict various types of drug-induced molecular perturbations and AEs.

ACKNOWLEDGEMENTS

The authors would like to thank the reviewers for constructive comments and feedback, and for recommendations and suggestions improving our manuscript, as well as colleagues in the University of Florida and Molecular Health GmbH for their support.

CONFLICT OF INTEREST

D.B.J. and T.G.S. have patents associated with the analysis of adverse event data, are employees of Molecular Health GmbH, and holders of stock options in the same company.

ORCID

Theodoros G. Soldatos https://orcid.org/0000-0001-6051-6605
Sarah Kim https://orcid.org/0000-0002-9179-0735

REFERENCES

1. Landes J, Osimani B, Poellinger R. Epistemology of causal inference in pharmacology. *Euro J Phil Sci*. 2018;8:3-49.
2. Homepage | Sentinel Initiative. https://www.sentinelinitiative.org/. Accessed September 28, 2021.
3. 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;99:265-268.
4. Research Questions and Answers on FDA's Adverse Event Reporting System (FAERS). FDA. https://www.fda.gov/drugs/surveillance/questions-and-answers-fdas-adverse-event-reporting-system-faers. Accessed September 23, 2021.
5. VAERS - About Us. https://vaers.hhs.gov/about.html. Accessed September 23, 2021.
6. European database of suspected adverse drug reaction reports. https://www.adrreports.eu/en/index.html. Accessed September 23, 2021.
7. UMC | VigiBase. https://www.who-umc.org/vigibase/vigibase/. Accessed September 23, 2021.
8. Scientific Publications by FDA Staff. https://www.accessdata.fda.gov/scripts/publications/. Accessed September 23, 2021.
9. Scientific publications by EMA staff European Medicines Agency. https://www.ema.europa.eu/en/news-events/publications/scientific-publications. Accessed September 23, 2021.
10. Boland MR, Jacunski A, Lorberbaum T, Romano JD, Moskovitch R, Tatonetti NP. Systems biology approaches for identifying adverse drug reactions and elucidating their underlying biological mechanisms. *Wiley Interdiscip Rev Syst Biol Med*. 2016;8:104-122.
11. Soldatos TG, Jackson DB. Adverse event circumstances and the case of drug interactions. *Healthcare*. 2019;7(1):45.
12. Soldatos TG, Taglang G, Jackson DB. In silico profiling of clinical phenotypes for human targets using adverse event data. *High-Throughput*. 2018;7:37.
13. Campillos M, Kuhn M, Gavin A-C, Jensen LJ, Bork P. Drug target identification using side-effect similarity. *Science*. 2008;321:263-266.
14. van Puijenbroek EP, Bate A, Leufkens HGM, Lindquist M, Orre R, Egberts ACG. A comparison of measures of disproportionality for signal detection in spontaneous reporting systems for adverse drug reactions. *Pharmacoepidemiol Drug Saf*. 2002;11:3-10.
15. Ho T-B, Le L, Thai DT, Taewijit S. Data-driven approach to detect and predict adverse drug reactions. *Curr Pharm Des*. 2016;22:3498-3526.
16. Fisher RA. On the interpretation of $\chi 2$ from contingency tables, and the calculation of P. *J Roy Stat Soc*. 1922;85:87-94.
17. Duggirala HJ, Tonning JM, Smith E, et al. Use of data mining at the Food and Drug Administration. *J Am Med Inform Assoc*. 2016;23:428-434.
18. Pham M, Cheng F, Ramachandran K. A comparison study of algorithms to detect drug-adverse event associations: frequentist, Bayesian, and machine-learning approaches. *Drug Saf*. 2019;42:743-750.
19. Noguchi Y, Ueno A, Otsubo M, et al. A new search method using association rule mining for drug-drug interaction based on spontaneous report system. *Front Pharmacol*. 2018;9.

20. Pinheiro LC, Candore G, Zaccaria C, Slattery J, Arlett P. An algorithm to detect unexpected increases in frequency of reports of adverse events in EudraVigilance. *Pharmacoepidemiol Drug Saf*. 2018;27:38-45.
21. Seabroke S, Candore G, Juhlin K, et al. Performance of stratified and subgrouped disproportionality analyses in spontaneous databases. *Drug Saf*. 2016;39:355-364.
22. Candore G, Juhlin K, Manlik K, et al. Comparison of statistical signal detection methods within and across spontaneous reporting databases. *Drug Saf*. 2015;38:577-587.
23. Vilar S, Friedman C, Hripcsak G. Detection of drug-drug interactions through data mining studies using clinical sources, scientific literature and social media. *Brief Bioinform*. 2018;19:863-877.
24. Phillips R, Sauzet O, Cornelius V. Statistical methods for the analysis of adverse event data in randomised controlled trials: a scoping review and taxonomy. *BMC Med Res Methodol*. 2020;20:288.
25. Guo JJ, Pandey S, Doyle J, et al. A review of quantitative risk-benefit methodologies for assessing drug safety and efficacy-report of the ISPOR risk-benefit management working group. *Value Health*. 2010;13:657-666.
26. Banda JM, Evans L, Vanguri RS, et al. A curated and standardized adverse drug event resource to accelerate drug safety research. *Scientific Data*. 2016;3:160026.
27. Xu D, Caylor ML & Niu J et al. ezFAERS: Encoding FDA adverse event reporting system with drug identifiers for improved data quality and query performance (web access: http://dxulab.pharmacy.isu.edu/curry/faers.php)
28. Sarangdhar M, Tabar S, Schmidt C, et al. Data mining differential clinical outcomes associated with drug regimens using adverse event reporting data. *Nat Biotechnol*. 2016;34:697-700.
29. Moore PW, Burkhart KK, Jackson D. Drugs highly associated with infusion reactions reported using two different data-mining methodologies. *J Blood Disorders Transfusion*. 2014;5:1-6.
30. Siafis S, Papazisis G. Detecting a potential safety signal of antidepressants and type 2 diabetes: a pharmacovigilance-pharmacodynamic study. *Br J Clin Pharmacol*. 2018;84:2405-2414.
31. Soldatos TG, Dimitrakopoulou-Strauss A, Larribere L, Hassel JC, Sachpekidis C. Retrospective side effect profiling of the metastatic melanoma combination therapy ipilimumab-nivolumab using adverse event data. *Diagnostics (Basel)*. 2018;8(4):76.
32. Racz R, Soldatos TG, Jackson D, Burkhart K. Association between serotonin syndrome and second-generation antipsychotics via pharmacological target-adverse event analysis. *Clin Transl Sci*. 2018;11(3):322-329.
33. Culbertson VL, Rahman SE, Bosen GC, et al. Implications of off-target serotoninergic drug activity: an analysis of serotonin syndrome reports using a systematic bioinformatics approach. *Pharmacotherapy*. 2018;38:888-898.
34. Soldatos TG, Iakovou I, Sachpekidis C. Retrospective toxicological profiling of radium-223 dichloride for the treatment of bone metastases in prostate cancer using adverse event data. *Medicina (Kaunas)*. 2019;55(5):149.
35. Sachpekidis C, Jackson DB, Soldatos TG. Radioimmunotherapy in non-Hodgkin's lymphoma: retrospective adverse event profiling of Zevalin and Bexxar. *Pharmaceuticals (Basel)*. 2019;12:141.
36. Schaefer A, Sachpekidis C, Diella F, et al. Public adverse event data insights into the safety of pembrolizumab in melanoma patients. *Cancers (Basel)*. 2020;12:E1008.
37. Papazisis G, Spachos D, Siafis S, et al. Assessment of the safety signal for the abuse potential of pregabalin and gabapentin using the FAERS database and big data search analytics. *Front Psychiatry*. 2021;12. https://doi.org/10.3389/fpsyt.2021.640264
38. Stamatellos V-P, Siafis S, Papazisis G. Disease-modifying agents for multiple sclerosis and the risk for reporting cancer: A disproportionality analysis using the US Food and Drug Administration Adverse Event Reporting System database. *Br J Clin Pharmacol*. 2021;87(12):4769-4779.
39. Spachos D, Siafis S, Bamidis P, Kouvelas D, Papazisis G. Combining big data search analytics and the FDA Adverse Event Reporting System database to detect a potential safety signal of mirtazapine abuse. *Health Informatics J*. 2020;26:2265-2279.
40. Khamisy-Farah R, Damiani G, Kong JD, Wu J-H, Bragazzi NL. Safety profile of Dupilumab during pregnancy: a data mining and disproportionality analysis of over 37,000 reports from the WHO individual case safety reporting database (VigiBase™). *Eur Rev Med Pharmacol Sci*. 2021;25:5448-5451.
41. Mahroum N, Damiani G, Watad A, et al. Higher rates of COVID-19 but less severe infections reported for patients on Dupilumab: a Big Data analysis of the World Health Organization VigiBase. *Eur Rev Med Pharmacol Sci*. 2021;25:5865-5870.
42. Tatonetti NP, Ye PP, Daneshjou R, Altman RB. Data-driven prediction of drug effects and interactions. *Sci Transl Med*. 2012;4:125ra31.
43. Berger SI, Iyengar R. Role of systems pharmacology in understanding drug adverse events. *Wiley Interdiscip Rev Syst Biol Med*. 2011;3:129-135.
44. Wilson JL, Racz R, Liu T, et al. PathFX provides mechanistic insights into drug efficacy and safety for regulatory review and therapeutic development. *PLoS Comput Biol*. 2018;14:e1006614.
45. Soldatou V, Soldatos A, Soldatos T. Examining socioeconomic and computational aspects of vaccine pharmacovigilance. *Biomed Res Int*. 2019;2019:6576483.
46. Bragazzi NL, Gianfredi V, Villarini M, et al. Vaccines Meet Big Data: State-of-the-Art and Future Prospects. From the Classical 3Is ('Isolate-Inactivate-Inject') Vaccinology 1.0 to Vaccinology 3.0, Vaccinomics, and Beyond: A Historical Overview. *Front Public Health*. 2018;6.
47. Wishart DS, Feunang YD, Guo AC, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res*. 2018;46:D1074-D1082.
48. Kim S, Chen J, Cheng T, et al. PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Res*. 2021;49:D1388-D1395.
49. Kuhn M, Letunic I, Jensen LJ, Bork P. The SIDER database of drugs and side effects. *Nucleic Acids Res*. 2016;44:D1075-D1079.
50. Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019;47:D607-D613.
51. UniProt Consortium. UniProt. the universal protein knowledgebase in 2021. *Nucleic Acids Res*. 2021;49:D480-D489.
52. Jassal B, Matthews L, Viteri G, et al. The reactome pathway knowledgebase. *Nucleic Acids Res*. 2020;48:D498-D503.

53. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res*. 2000;28:27-30.
54. Gilson MK, Liu T, Baitaluk M, Nicola G, Hwang L, Chong J. BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Res*. 2016;44:D1045-D1053.
55. Ochoa D, Hercules A, Carmona M, et al. Open Targets Platform: supporting systematic drug-target identification and prioritisation. *Nucleic Acids Res*. 2021;49:D1302-D1310.
56. WHOCC - Structure and principles. https://www.whocc.no/atc/structure_and_principles/. Accessed March 21, 2017.
57. MedDRA, 2 April 2018. http://meddra.org/.
58. Xu R, Wang Q. Large-scale combining signals from both biomedical literature and the FDA Adverse Event Reporting System (FAERS) to improve post-marketing drug safety signal detection. *BMC Bioinformatics*. 2014;15:17.
59. Soldatos TG, Perdigão N, Brown NP, Sabir KS, O'Donoghue SI. How to learn about gene function: text-mining or ontologies? *Methods*. 2015;74:3-15.
60. Berger SI, Iyengar R. Network analyses in systems pharmacology. *Bioinformatics*. 2009;25:2466-2472.
61. Abernethy DR, Woodcock J, Lesko LJ. Pharmacological mechanism-based drug safety assessment and prediction. *Clin Pharmacol Ther*. 2011;89:793-797.
62. Jacunski A, Tatonetti NP. Connecting the dots: applications of network medicine in pharmacology and disease. *Clin Pharmacol Ther*. 2013;94:659-669.
63. Kuhn M, Campillos M, González P, Jensen LJ, Bork P. Large-scale prediction of drug-target relationships. *FEBS Lett*. 2008;582:1283-1290.
64. Kuhn M, Al Banchaabouchi M, Campillos M, et al. Systematic identification of proteins that elicit drug side effects. *Mol Syst Biol*. 2013;9:663.
65. Trame MN, Biliouris K, Lesko LJ, Mettetal JT. Systems pharmacology to predict drug safety in drug development. *Eur J Pharm Sci*. 2016;94:93-95.
66. Andronis C, Silva JP, Lekka E, et al. Molecular basis of mood and cognitive adverse events elucidated via a combination of pharmacovigilance data mining and functional enrichment analysis. *Arch Toxicol*. 2020;94:2829-2845.
67. Lesko LJ, Fang L, Schmidt S, Trame MN. An integrated bioinformatics and quantitative modeling approach to investigate potential claims of oral generic drug product bioinequivalence: introduction. *J Clin Pharmacol*. 2019;59:1245-1248.
68. Ventola CL. Big data and pharmacovigilance: data mining for adverse drug events and interactions. *P T*. 2018;43:340-351.
69. Zhang X-Y, Trame MN, Lesko LJ, Schmidt S. Sobol sensitivity analysis: a tool to guide the development and evaluation of systems pharmacology models. *CPT Pharmacometrics Syst Pharmacol*. 2015;4:69-79.
70. Armaiz-Pena GN, Allen JK, Cruz A, et al. Src activation by β-adrenoreceptors is a key switch for tumour metastasis. *Nat Commun*. 2013;4:1403.
71. Schotland P, Racz R, Jackson DB, et al. Target adverse event profiles for predictive safety in the postmarket setting. *Clin Pharmacol Ther*. 2021;109:1232-1243.
72. Nguyen DA, Nguyen CH, Mamitsuka H. A survey on adverse drug reaction studies: data, tasks and machine learning methods. *Brief Bioinform*. 2021;22:164-177.
73. Lee CY, Chen Y-PP. Prediction of drug adverse events using deep learning in pharmaceutical discovery. *Brief Bioinform*. 2021;22:1884-1901.
74. Center for Devices and Radiological Health Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices FDA. https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices. Accessed September 23, 2021.
75. Our New AI-based Solution, Approved for ICU Care. Clew. https://clewmed.com/clew-medical-receives-fda-clearance-for-ai-based-predictive-analytics-technology-to-support-adult-icu-patient-assessment/. Accessed September 23, 2021.
76. Center for Devices and Radiological Health January 2021 510(K) Clearances. *FDA*. https://www.fda.gov/medical-devices/510k-clearances/january-2021-510k-clearances. Accessed September 23, 2021.
77. Authorization of Emergency Use of Certain Medical Devices During COVID-19; Availability. Federal Register, 2020. https://www.federalregister.gov/documents/2020/11/20/2020-25603/authorization-of-emergency-use-of-certain-medical-devices-during-covid-19-availability. Accessed September 23, 2021.
78. 510(k) Premarket Notification AccessData.FDA at https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfPMN/pmn.cfm?ID=K200717. Accessed September 29, 2021.
79. Emergency Use Authorization Letter FDA. FDA (media) at https://www.fda.gov/media/138369/download. Accessed September 29, 2021.
80. Soldatos TG, Kaduthanam S, Jackson DB. Precision oncology-the quest for evidence. *J Pers Med*. 2019;9:E43.
81. FDA Adverse Events Reporting System (FAERS) Public Dashboard. https://fis.fda.gov/sense/app/d10be6bb-494e-4cd2-82e4-0135608ddc13/sheet/7a47a261-d58b-4203-a8aa-6d3021737452/state/analysis. Accessed September 23, 2021.
82. The Vaccine Adverse Event Reporting System (VAERS). https://wonder.cdc.gov/vaers.html. Accessed September 23, 2021.
83. Böhm R, von Hehn L, Herdegen T, et al. OpenVigil FDA - inspection of U.S. American Adverse Drug Events Pharmacovigilance Data and Novel Clinical Applications. *PLoS One*. 2016;11:e0157753.
84. OpenVigil Pharmacovigilance Search Engines. http://openvigil.sourceforge.net/. Accessed September 23, 2021.
85. AERSMine: Visualize Differential Therapeutic Outcomes. https://research.cchmc.org/aers/home. Accessed September 23, 2021.
86. Effect - Molecular Health. https://www.molecularhealth.com/en/applications-effect/. Accessed September 23, 2021.
87. Advera Health. Analytics Pharmacovigilance Signal Detection and Management | Advera Health. https://www.adverahealth.com/. Accessed September 23, 2021.
88. OFF X. https://targetsafety.info/. Accessed September 23, 2021.
89. FDAble: Search FDA MedWatch FAERS, AERS & MAUDE Databases. http://www.fdable.com/. Accessed September 23, 2021.
90. Elsevier Preclinical & clinical data – PharmaPendium | Elsevier. Elsevier.com at https://www.elsevier.com/en-gb/solutions/pharmapendium-clinical-data. Accessed September 23, 2021.
91. Argus - Safety Case Management. https://www.oracle.com/industries/life-sciences/argus-safety-case-management.html. Accessed September 23, 2021.

92. Empirica - Safety Signal Management. https://www.oracle.com/industries/life-sciences/empirica-safety-signal-management.html. Accessed September 23, 2021.
93. openFDA. https://open.fda.gov/. Accessed September 23, 2021.
94. openFDA - shiny app reporting ratios example. https://openfda.shinyapps.io/RR_D/. Accessed September 23, 2021.
95. Lesko LJ, Zheng S, Schmidt S. Systems approaches in risk assessment. *Clin Pharmacol Ther*. 2013;93:413-424.
96. Stephenson WP, Hauben M. Data mining for signals in spontaneous reporting databases: proceed with caution. *Pharmacoepidemiol Drug Saf*. 2007;16:359-365.
97. Atkins D, Best D, Briss PA, et al. Grading quality of evidence and strength of recommendations. *BMJ*. 2004;328:1490.
98. Gini R, Sturkenboom MCJ, Sultana J, et al. Different strategies to execute multi-database studies for medicines surveillance in real-world setting: a reflection on the European model. *Clin Pharmacol Ther*. 2020;108:228-235.
99. Postigo R, Brosch S, Slattery J, et al. EudraVigilance medicines safety database: publicly accessible data for research and public health protection. *Drug Safety*. 2018;41:665–675. https://doi.org/10.1007/s40264-018-0647-1
100. Banovac M, Candore G, Slattery J, et al. Patient reporting in the EU: analysis of EudraVigilance data. *Drug Saf*. 2017;40:629-645.
101. Wisniewski AFZ, Bate A, Bousquet C, et al. Good signal detection practices: evidence from IMI PROTECT. *Drug Saf*. 2016;39:469-490.
102. Hill AB. The environment and disease: association or causation? *Proc R Soc Med*. 1965;58:295-300.
103. Fedak KM, Bernal A, Capshaw ZA, Gross S. Applying the Bradford Hill criteria in the 21st century: how data integration has changed causal inference in molecular epidemiology. *Emerg Themes Epidemiol*. 2015;12:14.
104. Natsiavas P, Malousi A, Bousquet C, Jaulent M-C, Koutkias V Computational advances in drug safety: systematic and mapping review of knowledge engineering based approaches. *Front Pharmacol*. 2019;10. https://doi.org/10.3389/fphar.2019.00415
105. Lesko LJ. Perspective on model-informed drug development. *CPT Pharmacometrics Syst Pharmacol*. 2021;10:1127-1129.
106. Di Giovanni R, Cochrane A, Lewis DJ. Adverse events in the digital age: finding the sharpest tool in the box. *Ther Innov Regul Sci*. 2022;56:23-37.
107. Lee DH, Yoon SN Application of artificial intelligence-based technologies in the healthcare industry: opportunities and challenges. *Int J Environ Res Public Health*. 2021;18:271. https://doi.org/10.3390/ijerph18010271
108. Goecks J, Jalili V, Heiser LM, Gray JW. How machine learning will transform biomedicine. *Cell*. 2020;181:92-101.
109. Pound P. Are animal models needed to discover, develop and test pharmaceutical drugs for humans in the 21st century?. *Animals*. 2020;10:2455. https://doi.org/10.3390/ani10122455
110. OpenVigil 2. http://h2876314.stratoserver.net:8080/OV2/search/. Accessed September 23, 2021.
111. Böhm R, Bulin C, Waetzig V, Cascorbi I, Klein H-J, Herdegen T. Pharmacovigilance-based drug repurposing: The search for inverse signals via OpenVigil identifies putative drugs against viral respiratory infections. *Br J Clin Pharmacol*. 2021;87:4421-4431.
112. Open Targets | Pharmacovigilance. Accessed September 23, 2021.

How to cite this article: Soldatos TG, Kim S, Schmidt S, Lesko LJ, Jackson DB. Advancing drug safety science by integrating molecular knowledge with post-marketing adverse event reports. *CPT Pharmacometrics Syst Pharmacol*. 2022;11:540-555. doi:10.1002/psp4.12765