DOI: 10.1111/cgf.13674
Eurographics Conference on Visualization (EuroVis) 2019
M. Gleicher, H. Leitte, and I. Viola
(Guest Editors)

Volume 38 (2019), Number 3

DIVA: Exploration and Validation of Hypothesized Drug-Drug Interactions

T. Kakar[1,2] , X. Qin[1,2], E. A. Rundensteiner[1], L. Harrison[1], S. K. Sahoo[2], and S. De[2]

[1] Computer Science Department, Worcester Polytechnic Institute, Worcester, USA
[2] Center for Drug Evaluation and Research, U.S. Food and Drug Administration, Maryland, USA

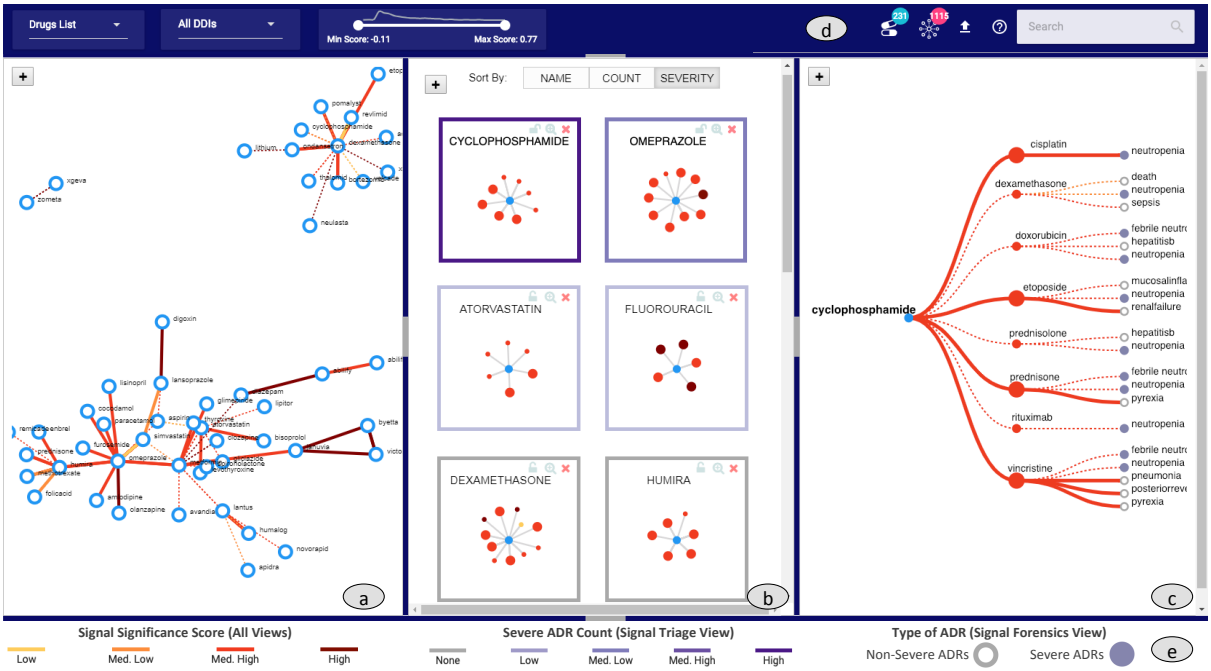

Figure 1: *The user interface of DIVA, a web-based visual analytics system for exploring and verifying Drug-Drug Interactions (DDIs) proposed via machine learning methods. (a) Screening Overview – showing candidate Drug-Drug Interactions for selected score. (b) Signal Triage View – enables drug-centric analysis of interactions, including the number of severe Adverse Reactions (ADRs). (c) Signal Forensics View – a view of the interaction profile of a drug of interest, including all ADRs triggered by each signal. (d) Controls facilitate navigation between views, and direct filtering by drugs of interest. (e) Legends for colors.*

Abstract

Adverse reactions caused by drug-drug interactions are a major public health concern. Currently, adverse reaction signals are detected through a tedious manual process in which drug safety analysts review a large number of reports collected through post-marketing drug surveillance. While computational techniques in support of this signal analysis are necessary, alone they are not sufficient. In particular, when machine learning techniques are applied to extract candidate signals from reports, the resulting set is (1) too large in size, i.e., *exponential to the number of unique drugs and reactions in reports, (2) disconnected from the underlying reports that serve as evidence and context, and (3) ultimately requires human intervention to be validated in the domain context as a true signal warranting action. In this work, we address these challenges though a visual analytics system, DIVA, designed to align with the drug safety analysis workflow by supporting the detection, screening, and verification of candidate drug interaction signals. DIVA's abstractions and encodings are informed by formative interviews with drug safety analysts. DIVA's coordinated visualizations realize a proposed novel augmented interaction data model (AIM) which links signals generated by machine learning techniques with domain-specific metadata critical for signal analysis. DIVA's alignment with the drug review process allows an analyst to interactively screen for important signals, triage signals for in-depth investigation, and validate signals by reviewing the underlying reports that serve as evidence. The evaluation of DIVA encompasses case-studies and interviews by drug analysts at the US Food and Drug Administration - both of which confirm that DIVA indeed is effective in supporting analysts in the critical task of exploring and verifying dangerous drug-drug interactions.*

1. Introduction

Adverse drug reactions (ADRs) caused by drug-drug interactions are a major cause of mortality, resulting in more than 100,000 deaths annually with a yearly cost of over $170 billion in the U.S. alone [LPC98, EG01]. Polypharmacy, the use of multiple drugs to treat medical conditions, is also rising. For example, approximately 29% of elderly patients are taking six or more drugs, which increases the chance of harmful and possibly fatal Drug-Drug Interactions (DDIs) [BMS*08].

Before approval for use, new drugs are tested for interactions with existing drugs using both clinical trials & animal studies (in vivo) and tests on cells (in vitro) methods [ZZZH09]. However, any given drug may interact with other drugs in numerous, unexpected ways. These interactions make it impossible to test all possible drug combinations before a drug is released to the market. Drug regulatory authorities therefore collect, analyze, and make regulatory decisions based on reports on adverse drug events via a process called post marketing drug surveillance, which aims to detect unanticipated adverse reactions that were not discovered during clinical trials. Early detection makes it possible for authorities to take actions that minimize patient exposure to harmful drug combinations.

The U.S. Food and Drug Administration (FDA) conducts post marketing surveillance via the FDA Adverse Event Reporting System (FAERS) [FA15]. Similar systems are also in operation internationally, including the World Health Organization [Lin08], as well as in Canada [Can16] and Britain [AAB*11]. In 2016, FAERS received approximately 1.7 million drug reaction reports [FA15]. Direct information about drug interactions is not captured in these reports. Rather, each report includes the drugs being taken by the patient along with the observed adverse reactions. Prior studies [HCF10, QKW*17] have suggested that these reports are a critical information source for discovering potential drug-drug interaction signals worthy of investigation. Such interactions may represent causal effects between a combination of drugs that result in dangerous adverse reactions.

One challenge is that the manual approaches currently used for detecting and investigating candidate signals in large sets of drug safety reports are tedious, and time consuming. Complicating the problem in practice is the reality that, due to staff limitations, a small team of roughly fifty analysts at the U.S. FDA have dedicated time for reviewing these reports. Given these restrictions, the primary workflow of the FDA primarily focuses on signal detection related to single drug adverse reactions, with drug-drug interaction findings remaining more a matter of chance, despite its significant risks.

Automated approaches to drug reaction analysis are also insufficient. Machine learning techniques proposed to mine drug reaction reports for signal hypotheses tend to generate a large number of candidate signals [SFG16, ADK*03, HCF10, ISAE16, CLH*17]. For example, n distinct drugs and m unique adverse reactions across a set of reports result in up to $\mathcal{O}(2^{n+m})$ signals in the worst case. Regardless, machine-generated signals require inspection by drug safety analysts, who must analyze and validate signals as worthy of escalation, or dismiss signals because of insufficient evidence.

In this paper, we propose to address these challenges through a visual analytics framework called Drug Drug Interactions via Visual Analysis, that supports drug safety analysts in analyzing drug-interaction signals mined from the drug surveillance reports.

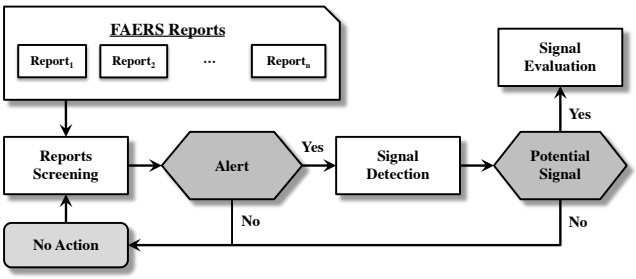

Figure 2: *The current Drug Review Process, composed of Signal Screening, Detection, and Evaluation phases.*

To design a visual paradigm that aids the review process, we first study the full life cycle of how drug interaction signals are screened, studied and eventually used as evidence for recommending regulatory action. By interviewing drug safety analysts and observing their review routines, we construct a data abstraction, the Augmented Interaction Model (AIM), that captures the core data concepts and their relationships critical for drug analysts to explore and validate candidate interaction signals. Through these interviews, we also extract key requirements essential to drug analysts' review process, which guide the design of DIVA's visual displays and underlying operations to allow analysts to explore and validate mined drug interaction signals. Following an iterative design process and evaluations, DIVA's resulting visualizations (Fig. 1) include a network visualization that shows a summary of interactions in focus reports, a small-multiples node-link view that supports drug-centric inspection of signals, and a profile view that enables in-depth investigation of a signal, including the underlying reports that serve as evidence.

The primary contributions of this work include:

- A characterization of the drug safety review process via formative interviews, which serves as a basis for extracting requirements for drug interaction signal detection and verification.
- A domain knowledge driven data abstraction, the Augmented Interaction Model (AIM), that integrates the diverse critical information composed of machine-generated signals, relevant domain knowledge extracted from external sources, and evidential reports into one unified model for sensemaking.
- Interactive visual displays, each built on top of the AIM model, that enable drug analysts to explore relevant slices of the AIM model from multiple perspectives in support of the specific subtasks in their review process.
- An evaluation consisting of case studies with domain experts, which demonstrate the utility of visual analytics approaches for exploring, analyzing, and validating drug interaction signals, particularly when juxtaposed with current state-of-the-art practices.

The results of evaluations with U.S. FDA analysts suggest that DIVA's visual analytics for drug interaction signal screening fills a tangible need in the early detection of severe interactions. DIVA allows analysts to move between levels of abstraction – building trust in the results of the computational techniques augmented with human interaction for better decision making. By aligning with their workflow, DIVA aims to support analysts in identifying dangerous drug interaction signals from an overwhelming set of candidates.

2. Background

DIVA draws from prior work in visualization spanning at least two primary categories, including visualization techniques for associa-

tion rules, and visualization systems focusing on the discovery and analysis of drug interactions.

2.1. Association Rule Visualization

Drug-drug interactions are often associated with a consistent set of adverse reactions. As such, these relationships have been modeled using association rule techniques in prior work, particularly in data mining contexts. DIVA therefore draws on prior work in visualization that centers on association rule visualization.

Romero et al. [RLRV11] represent association rules with simple text in a tabular format. Grids (2D matrices) [OONL02, BKK97] and 3D matrices [WWT99] have also been proposed to visualize association rules for smaller data sets. Matrices with a fish-eye view [CRC07, CHYN07] visualize association rules in more detail. InterVisAR [CSW16] uses a two-dimensional bar chart approach to allow users to search for particular rules.

In Mosaic plots [HSW00], individual antecedent items are shown as horizontal bars along the x-axis and the support of an association is represented by the height of the vertical column above the specified item. Existing graph based association rule visualization tools [HHHW98, HCHB11, TD05, LS16] tend to focus on an overview of the generated rules, rather than the investigation and validation of a set of rules.

In parallel coordinates-based rule visualization [Yan05, HC00] each vertical line depicts a set of items and a rule is represented by lines or splines. Some initial work has combined two techniques to visualize rules [BC05, SH13, BGB03]. Buono et al. [BC05] used both graphs and parallel coordinates to get an overview as well as a detailed view of selected rules. Sekhavat et al. [SH13] used matrices as overview of rules and graphs to analyze a selected subset of rules. Another approach includes a virtual arena [BGB03] where rules are represented as spheres positioned by the steps of an arena. Similarly, glyphs have been used to represent quantitative values associated with the rules [QKW*18].

These prior approaches support rule analysis with the primary goal of visualizing the structure of these machine-generated rules, such as common consequents and antecedents. While the design space covered by DIVA shares some of these goals, other key analytics tasks differ. For example, the work context in which DIVA was designed requires support for in-depth analysis of the content of these rules, including features such as severity and relations to other drug and adverse reaction pairs.

2.2. Drug-Interaction Visualization and Network Diagrams

Several recent studies have developed visualization systems for analyzing interaction between drug and proteins as well as with other drugs. Kegg [KGF*10], like other online tools [Drub, DDI], is a search interface for known drug-drug interactions. In this work we have used such tools to extract known signals into a hypothesis-driven exploratory system for discovering and analyzing unknown signals.

Stitch [KvMC*08] integrates data from various sources and uses a network visualization to represent Drug-Protein interactions. Promiscuous [VEMD*11] integrates data from three different molecular databases and visualizes Drug-Target interactions and drug-related adverse reactions using node-link diagrams. Both of these tools focus on data integration and allow exploration of drug related chemicals, however, these tools are not designed to support drug safety analysis workflows. GraphSAW [SHK*15] integrates data about known drug-drug interactions from various sources. A radial network graph is used to visualize a set of adverse reactions and drug interactions. Network visualization techniques are also used to analyze vaccine related adverse events [BB11a, BSG*14, BB11b]. For example, Botsis et.al propose AENA [BSG*14], which uses a network diagram with an edge weighing algorithm to identify outliers in the U.S. Vaccine Adverse Event Reporting System. While these tools do not support the specific analytic activity of conducting pharmacovigalance by analyzing drug reports for unknown interactions, they do form a broader landscape of tools that aid in the overall pharmacovigalance mission by providing access to known interactions between drugs and drug compounds.

3. Task Characterization

To design DIVA, we worked closely with domain experts at the FDA who serve as drug safety analysts. We used an interview-based iterative design process, presenting the analysts with progressively refined prototype visualizations to characterize the requirements in support of their workflow. In doing so, we arrived at the Augmented Interaction Model which serves as a basis for the visualizations and interactions in DIVA.

3.1. Interviews with Domain Experts

We organized a series of semi-formal interviews with drug safety reviewers. A primary aim was to understand the current drug review process and to identify challenges that reviewers face in analyzing drug-drug interactions. From these interviews, we learned that certain information was critical to their workflow. We also observed how they transformed certain data throughout the analysis. This then informed our creation of a data abstraction which we call the Augmented Interaction Model (AIM). The AIM captures the critical information analysts need to identify possible adverse reactions caused by drug interactions (signals, for short) by integrating domain specific meta-information with mined drug-drug interactions.

To design and refine the specific visualizations DIVA uses, we showed the analysts sketches of design alternatives, such as glyphs and variations of network diagrams. This activity helped us gather additional design requirements. In subsequent interviews, we presented analysts with a working prototype of DIVA to evaluate their perceptions of the degree to which DIVA meets their needs, and to receive further feedback on the visual and interaction design. In the final session, a larger group of analysts used DIVA to explore FAERS data. This activity led to additional insights on the utility of DIVA, a visual analytics tool in supporting the drug review process.

3.2. The Drug Review Process

The goal of drug safety analysis is to identify potential safety issues related to drug-drug interactions, and to escalate cases for further action if sufficient supporting evidence is found during the evaluation of potential signals. The drug review process is composed of iterative steps as depicted in Fig. 2. Each safety analyst receives reports related to the drugs assigned to them.

The drug analysts we interviewed screen their assigned reports for red flags such as a severe adverse reaction. The primary mechanism the analysts use for retrieving these reports are pre-computed database queries. As a next step, the analysts explore whether a candidate signal needs to be escalated for further review and action by searching for similar reports and reading their associated text narratives in detail. If the analyst finds sufficient evidence to move

forward, they proceed by evaluating patient's medical histories to find additional evidence supporting a potential signal.

If there is sufficient evidence, then analysts formulate their recommendations along-with the supporting material for escalation. Post-escalation action can led to regulatory action, such as changing drug labels or restricting drug usage. In severe cases, drugs are removed entirely from the market [WS05]. Drug review analysis has many challenging sub-problems.

Given the complexity of inferring and investigating drug-drug interactions, analysts often focus on single drugs and their adverse effects. DDIs are sometimes investigated only incidentally, if a hypothesis is formed during routine report analysis.These factors, combined with known limitations of purely computational approaches, motivate the need for visual analytics systems that improve the drug review process. Hence, we set out to design displays and interactions to realize a visual analytics drug review workflow to explore and validate machine-generated hypothesized signals interactively.

3.3. Requirements Analysis

Throughout our interviews with the drug review analysts, we established and incrementally refined a set of requirements to guide the design of DIVA. While these requirements were iteratively updated throughout the course of the project, the following list represents the final version of the requirements used to inform the development and evaluation of DIVA.

Screening for Possible Drug-drug Interactions:

R1: ***Provide an overview of all signals.*** Given the large number of drugs and ADRs, approximately tens of thousands in three months of data, the possible relationships between drugs and ADRs (signals) extracted from this data is large. Analysts expressed a need for an overview of potential candidates drug interactions to gain a quick preview of their tasks. Such an overview should help an analyst screen for low-importance DDIs, and narrow down the search space to focus on those that are both likely and severe.

R2: ***Allow analysts to segment and prioritize signals.*** We found that drug safety analysts review reports based on a set of roughly hundred drugs assigned to them. This implies a need to segment signals based on the assigned drugs. As each drug may interact with hundreds of other drugs, possibly outside the analysts' assigned list, functions for the prioritization of signals are required.

R3: ***Integrate previously known signals.*** The mining and data integration process generates both signals that are known (that is, previously discovered and already documented by the community) and unknown/unverified interactions. Analysts need ready access to such prior domain knowledge to determine if a candidate signal is indeed unknown and thus worth of exploration. Without that, huge overhead may be wasted by looking up external resources, duplicating work or worse yet, taking guesses based on their recollections.

R4: ***Facilitate identification of unknown signals.*** Drug review analysts are interested in uncovering unknown, novel signals that constitute a hypothesis worth escalation and further investigation. Therefore, unknown signals must be easily recognizable so they can remain a priority.

R5: ***Facilitate identification of severe adverse reactions.*** Drug interactions leading to severe adverse reactions (ADRs) such as heart attacks, kidney failure, or death (as opposed to non-severe ADRs such as headaches or nausea) must be given greater attention. Severe ADRs must thus be easily identifiable.

R6: ***Ready access to evidence supporting signals.*** Domain experts have indicated that it is essential to have direct access to the actual reports, because these reports form the key evidence for a suspected signal candidate. Views must be designed to provide rapid access to the reports.

4. The Augmented Signal Model

4.1. FAERS Reports

Reports submitted to the FDA Adverse Event Reporting System (FAERS) contain structured information about patient demographics, drugs taken, therapies, and adverse reactions. They also contain an unstructured textual narrative that describes the adverse reactions in detail and contains richer information such as a patient's medical history. This collection includes mandatory reports submitted by drug manufacturers and voluntary reports submitted by health care professionals and consumers. To ensure the **reproducibility** of our research, we focus on a public version of FAERS data that includes the structured information with no personal identification and is available on a quarterly basis [FA15]. The core data elements, such as drugs and ADRs within each report processed by our machine learning module, are available in this structured FAERS. While the inclusion of the actual narratives can be easily provided by DIVA internally for the FDA analysts, however, for privacy reasons they cannot be published publicly.

4.2. Augmented Interaction Model (AIM)

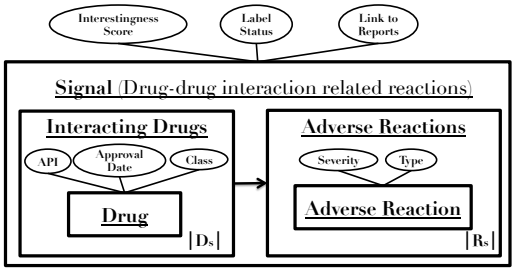

Figure 3: *Augmented Interaction Model (AIM) describes the data entities, their properties and relationships which consist of a drug-drug interaction related reaction signal. Each signal encodes a casual relationship from a set of interacting drugs $\mathcal{D}_s$ to a set of triggered adverse reactions $\mathcal{R}_s$.*

After rounds of interviews and initial design alterations, we developed a data abstraction [Mun09] that reflects a unit of exploration from the FDA analysts perspectives. Drug interaction related signals information consists of data from various external sources, generated automatically as well as manually. We capture all this information in the form of entities, attributes and their relationships into an Augmented Interaction Model (AIM).

The AIM provides all the information essential for an analyst to be able to explore and analyze signals, i.e., screen the important ones and validate them. In this section, we define the core entities that form the AIM abstraction and explain how each component in AIM is instantiated by DIVA's visualizations and interactions.

Definition 1 Drug Entity. A drug entity *DE* refers to a single drug product d from a list of approved drugs $\mathcal{D}$. Each drug d is associated with a set of attributes $\mathcal{A}_d = \{a_1, \cdots, a_n\}$ that describe d.

Below are the attributes most useful in the review process.

a_1 **API (Active Pharmaceutical Ingredient):** The central ingredient that produces the effects of the drug.

a_2 **Approval Date:** The date when the drug was approved by a regulatory agency.

a_3 **Class:** A specific drug category that the drug belongs to.

These attributes are independent from the reported adverse events and the generated signals. A signal is composed of at least two distinct drugs, i.e., the interacting drugs.

An adverse drug reaction (ADR) is an unwanted reaction possibly triggered by the administration of a medication.

Definition 2 Adverse Reaction Entity. An adverse reaction entity *AE* refers to a single reaction r from a reaction vocabulary $\mathcal{R}$. Each reaction r has a set of attributes $\mathcal{A}_r = \{a_1, \cdots, a_m\}$ that describes r. The attributes relevant to the review process include:

a_1 **Severity:** The severity of an ADR is a binary attribute that indicates if the ADR is serious, determined and maintained by FDA.

a_2 **Type:** A specific ADR class as defined by a medical dictionary.

A signal describes the interacting drugs and the resulting reactions which are the outcome of the interaction.

Definition 3 Signal. A drug-drug interaction related adverse reaction signal s models a causal relationship between a set of interacting drugs $\mathcal{D}_s$ and a set of triggered reactions $\mathcal{R}_s$, denoted as $s = \mathcal{D}_s \rightarrow \mathcal{R}_s$ where $\mathcal{D}_s \subseteq \mathcal{D}$ and $\mathcal{R}_s \subseteq \mathcal{R}$. Each signal s is associated with a set of attributes $\mathcal{A}_s = \{a_1, \cdots, a_n\}$ that explains it.

Signals are generated from a set of FAERS reports using computational methods. Attributes related to the signal critical for the review process include:

a_1 **Interestingness Score:** A numeric variable that quantifies how significant a signal is with respect to a given set of reports. The significance reflects how likely this signal is true and worth of further investigation. This score is calculated by the machine learning techniques.

a_2 **Label Status:** The label status is a binary variable indicating whether or not this signal is already known to the FDA, or it is currently unknown.

a_2 **Links to Reports:** Links to all evidential reports from which the signal is derived.

The AIM model represents the above mentioned entities namely, drugs and reactions, the signals composed of these entities and the domain knowledge that augments these signals as depicted in Fig. 3.

Definition 4 The AIM Model. Given a set of reports $\mathcal{T}$, an Augmented Interaction Model (AIM) $\mathcal{M}^{\mathcal{T}}$ can then be captured by a set of signals $\mathcal{S}^{\mathcal{T}} = \{s_1, \cdots, s_n\}$ derived from $\mathcal{T}$. The attributes of the drug entity *DE*, adverse reaction entity *AE* and the signal s are populated based upon $\mathcal{T}$ and other domain knowledge such as Drugs.com [Drub].

4.3. The AIM Model Instantiation

The AIM model captures rich information about drugs, ADRs and possible signals extracted from a given set of reports. Next, we discuss how this model is instantiated.

Instantiation of Entities and their Attributes. The FDA maintains a list of approved medical products, including drugs currently in the market [druc]. Each drug is documented with detailed information such its active ingredients, approval date, and drug class. In this study, we extract these attributes from FDA resources [druc] and construct a drug entity repository for use in DIVA.

For an adverse reaction (ADR) entity, we use the *Preferred Terms* from the MedDRA Hierarchy [med] to form an adverse reaction vocabulary. To specify the **severity** of these reactions (**R5**), we leverage the list of Designated Medical Events (DMEs) also known as severe ADRs maintained internally by FDA for review purposes. A severe ADR such as heart failure or liver injury is more alarming than nausea or headache. Thus it must be prioritized over less severe concerns to avoid further patient exposure.

Instantiation of Signals and their Attributes. Drug-drug interaction related adverse reaction signals are the core components of the AIM model. They correspond to severe candidate DDIs extracted from a set of reports. Prior studies [HCF10, QKW*17] have suggested signal generation by modeling associations between drugs and reactions using their co-occurrence in the surveillance database. That is, *frequent pattern mining* methods have been applied to extract signals.

In this work, we adopt *MARAS* [QKW*17] technology to mine potential signals as sketched below. MARAS adopts association rule learning to identify relationships among objects that occur together in a database. In the surveillance database (FAERS), each record can be modeled as a combination of a reported drug set and the reported observed ADR set. The rules that model the relationship between a drug set and an ADR set are signals that need exploration and validation.

A brief formulation of MARAS follows. Let each report t_i be represented as a set of distinct drugs ($\mathcal{D}$) and a set of distinct adverse reactions ($\mathcal{R}$). The generated rules modeling drug interactions are in the form of:

$$Rule = \mathcal{D} \rightarrow \mathcal{R}. \quad (1)$$

Many measures [Sah10] exist to evaluate the significance of a rule. Two common measures are *support* and *confidence*.

MARAS addresses issues related to avoiding misleading rules in this context through a *contrast* measure (See [QKW*17]). This evaluates how likely the ADRs are caused by a drug interaction. The intuition behind this measure is that if ADRs are triggered by the drug-drug interaction, then they should have less chance of being triggered by any of the individual drugs in the signal. The *contrast* measure is used as an **interestingness score** to help prioritize signals in DIVA. Signals with higher scores have more chances of being true signals warranting action, therefore need to be prioritized by safety analysts. Other proposed measures [Sah10] can also be plugged in to the system. The linkage (Case-Ids in FAERS reports) to raw reports used to extract the signals is maintained during the mining process so that analysts can access these reports for signal validation (**R6**).

Computer-generated signals may be both already known (labeled) signals as well as unknown (unlabeled) signals. These generated signals can help drug safety analysts form hypotheses where they identify novel and severe signals worthy of further investigation. Therefore, such unknown signals must be distinguishable from the already known ones as such by their *Label Status*. Information about the status of a signal is not available in FAERS, neither does the signal extraction method provide this information. Hence, to assist analysts, we incorporate such information into our AIM model (**R3**) by extracting it from external sources [Drub].

5. DIVA System Overview

The DIVA framework depicted in Fig. 4 consists of two major components, namely the **AIM Constructor** that generates AIMs

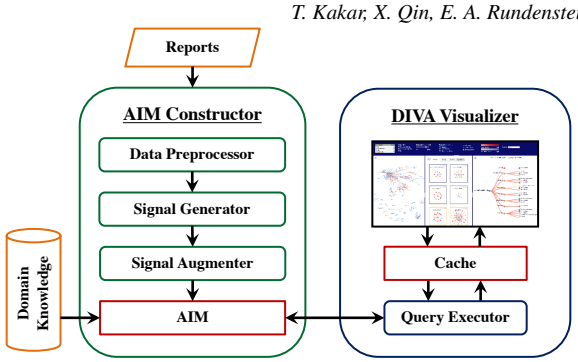

Figure 4: *Overview of the DIVA framework consisting of two main components: the AIM Constructor and the DIVA Visualizer.*

from reports and the **DIVA Visualizer** that supports interactive visual analytics of AIMs from multiple perspectives.

The **AIM Constructor** has three modules: The *Data Preprocessor*, *Signal Generator* and *Signal Augmenter*. The *Data Preprocessor* transforms the original FAERS reports into the format required for the signal generation algorithm. During the preparation, duplicate reports are removed and drug names are cleaned due to different variants of same drugs and spelling mistakes.

The *Signal Generator* module adapts MARAS [QKW*17], a drug interaction signal extraction and scoring technique. Other machine learning techniques [SFG16, ADK*03, HCF10, ISAE16, CLH*17] are also candidates for producing signals that serve as input to DIVA. These generated signals model the association between drugs and ADRs, depicted in Fig. 3, along with its interestingness score (*contrast*) [QKW*17]. The *Signal Augmenter* populates the rejoining attributes of the signals to produce AIMs. Domain knowledge such as the label status and severity of ADRs is obtained from external data resources.

The **DIVA Visualizer** module, consisting of multiple coordinated views (Fig. 1), provides multiple perspectives into the Augmented Interaction Model. The views aim to align with the work-flow of the analysts so that they can explore the major components of the AIM in an iterative manner. The underlying queries and data exchanges between the AIM and the analysts are supported by the *Query Executor*. A cache is used to optimize the query execution time.

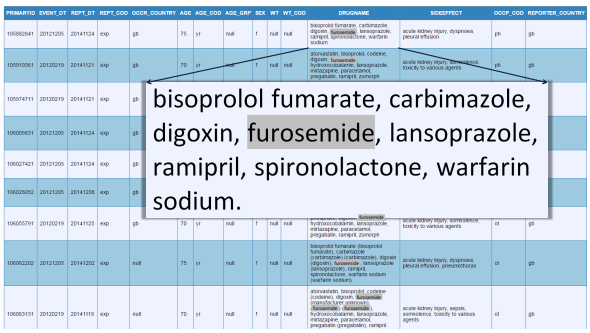

Figure 5: *FAERS Reports associated with interaction Lansoprazole and Digoxin. Every report has Furosemide which is used to treat kidney disorders.*

The *Screening* view gives an overview of all hypothesized drug-drug interactions supporting an analyst in screening unknown and high scored signals (Fig. 1-a). The *Triage* view, composed of small multiples, shows all the drug interactions associated with a particular drug or set of drugs. It helps analysts prioritize a drug for review based on the aggregated interestingness of its interactions (Fig. 1-b). The *Forensics* view includes adverse reactions related to each drug-drug interaction for further exploration (Fig. 1-c). To further investigate a drug interaction, at the lowest level, the *Reports* view visualizes the line-listings and text narratives of reports associated with a selected drug interaction (Fig. 5). We developed the visual interface of DIVA following the aforementioned design rationale (Section 3). All views are coordinated via brushing and linking, supporting hypothesis generation, exploration and validation in the context of drug interactions.

6. Design of DIVA Interactive Views

DIVA, a Web-based system, has been designed to fulfill the requirements elicited in Section 3. DIVA is composed of multiple coordinated interactive views, that based on the tasks, provide different perspectives into the AIM data model because of its richer content. The visual encodings reflect aspects of the AIM data model including information such as the drugs and reactions that compose a signal, an interestingness score, label status, and severity of adverse reactions (DMEs).

6.1. Signal Screening Overview

The Screening view provides an overview into drug-drug interactions (**R1**). This view allows analysts to see the entire space of the machine generated drug-drug interactions through a node-link diagram (Fig. 6). Here nodes represent the drugs, while edges depict an interaction between a pair of drugs. The shape and size of the edge encodes whether an interaction is known (dotted and thin) or unknown (solid and thick). The color of an edge is mapped to the strength of the interaction as determined by the mining technique (the interestingness score) derived from the support and confidence (see Section 4.3). All colors are selected based on their encoded data types using ColorBrewer [HB03]. The shape channel is used for encoding the binary signal status attribute, while color is used to depict score which is divided into four bins based on the distribution of the scores.

A pair of drugs can contribute to multiple signals, each of which can have a different score. To avoid confusion and repetition of data in the Screening view, each drug is represented only once. This way an analyst can instantly examine the degree of possible interactions between drugs. When multiple signals are caused by the same drug pair, an aggregated score encodes the interestingness score represented by edges. In such cases, we use the maximum score of all generated signals as an aggregated score to represent the drug-drug interaction. A maximum score is used as an aggregated score so the analyst can know quickly that at least one of the signals related to a particular drug pair is interesting as opposed to using an average of scores that might hide a highly scored signal by averaging it. Similarly, even if one of the multiple signals related to a drug pair is unknown, the edge is marked unknown (solid and thicker) to grasp the analyst's attention. This helps avoid missing the detection of novel signals. While it is possible to augment these views with more nuanced information, for example through glyphs or more complex color schemes, we instead emphasize visual cues based on drug analyst's reported work-flows.

The length of the edge or the position of the node is a by-product

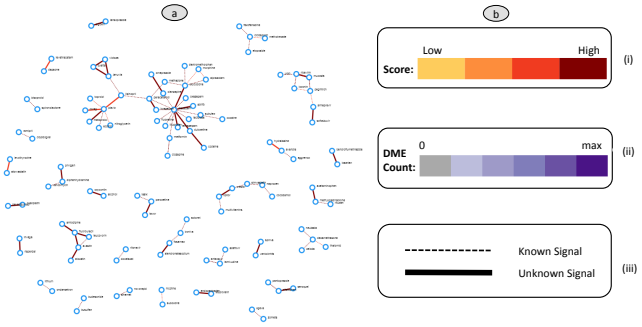

Figure 6: *(a): Screening View gives an overview of interactions between drugs. Each node is a drug and edge depicts an interaction between two drugs. (b): Color Legends for: (i) Interestingness score in all views, (ii) Severe adverse reactions (or "DME") count in the Triage View, (iii) Shape of links representing label status in the Screening and Forensics view.*

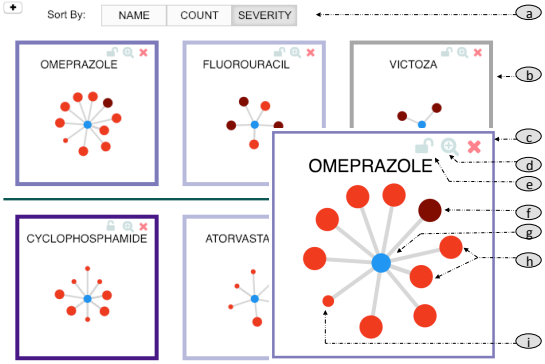

Figure 7: *The Triage View - (a) Sort. (b) Pinned drugs (c) Color of box represents total number of severe reactions present in all signals related to Omeprazole. (d) Zoom in to view the Profile of Omeprazole. (e) Pin a drug. Interacting drugs with an (f) unknown and high score signal, (h) unknown and low score signal, (i) known and low score signal.*

of the force-directed layout to assure efficient use of space. This view invites high-level comparisons between DDIs to help an expert in the screening of non-important DDIs (**R2**). Analysts may be more interested in a DDI leading to a particular ADR that is not discovered yet via clinical trials or post-marketing surveillance, as their goal is the detection of novel signals with minimum patient exposure. Moreover, an overview can enable a team leader to track where in the space of possible DDIs their analysts should invest their time at.

Design Alternatives: Based on the requirements discussed in Section 2, we explored a large design space of visual encodings. Several candidate views were iteratively eliminated based on the analysis of the elicited requirements through the periodic interviews with analysts.

- Table View: The simplest method to show rules or data with relationships is a tabular format [RLRV11], where each attribute can be a column and each row corresponds to a signal. TableLens [RC94] which is designed to allow users to detect patterns, correlations, and outliers in the data set using tables is suitable for presenting numerical data. However, most of our data is categorical. Also thousands of drugs and ADRs with many to many relationships form the signals. Thus a tabular format may be cognitively demanding and tedious for exploration. Moreover, a tabular format does not provide enough visual dimensions to encode AIM models as units (**R1, R4, R5**).
- Adjacency Matrix: An alternate design to visually encode drug interactions would be an adjacency matrix where row and column dimensions map to the drugs and each cell depicts a DDI [BKK97, WWT99]. Adjacency matrices have two shortcomings. First, drugs might interact with other drugs as nearly disjoint sets, that is, each drug on the x-axis might have a different set of interacting drugs on y-axis. This would render the matrix mostly empty. Second, with hundreds of drugs, the matrix size grows which affects readability.

6.2. Signal Triage View

To align with the work-flow of the analysts, we aim to help them prioritize which drugs and interactions to analyze first from a pool of drugs assigned to them (**R2**). For this, we designed a Triage view using the small multiples technique with a small node-link diagram using the force-directed layout [Tuf91, APP11, RFF*08, FHQ11].

Each small multiple represents a drug and all its associated drug interactions by nodes. The outer box represents the count of severe adverse reactions (also known as designated medical events (DMEs)) present within each small multiple with a continuous color scale, where gray color is used to encode the absence of severe ADRs.

The center node represents the drug of interest and nodes surrounding it depict all other drugs interacting with the drug of interest (Fig. 7). At a glance, the analyst can get an overview of each of her assigned drugs. She can pick the drug with a comparatively larger number of DMEs, i.e., severe ADRs (**R4**). Or she can focus on the most interesting drug-drug interactions without any overwhelming details about the signal such as the related adverse reactions.

The option to pin a drug is provided to facilitate the analyst in prioritizing a drug to further explore it (**R5**). Pinning helps the analysts maintain context, so that they can resume their work where they left off in case they do not finish the review of a particular drug in one go. Similar to the *Screening view* (Section 6.1), if a drug-pair has many signals, the maximum interestingness score is mapped to the color of the nodes to facilitate attention.

We worked with analysts through alternate variations of encodings and visual channels, arriving at final views based on analysts' consensus on the views they perceived as being best for their tasks and workflow. Links are associated with several visual channels in the Screening view, as each node is associated with multiple drugs and this stage of the workflow requires analysts to reason about specific links between drugs. In contrast, links are only used as an association visual channel in the Triage view, as each node represents one drug and analysts reported difficulty in reading these views when links were encoded. Other visual channels were not found to be helpful at this stage. For example, the size channel for nodes is not currently used in the Screening and Triage views in part because the relatively small size may hinder expressive quantitative encodings of size, and additional categorical representations were seen as unnecessary in this first prototype system. At the same time, these choices are meant to serve as a baseline for future evaluations, where analysts may find that additional encodings are useful to enable more rapid or robust exploration of the data.

The drug name represented by a node along with other information about the interaction such as the number of reports supporting the interaction, their ADRs, their scores etc. is revealed via a tool-tip upon hovering over a node. The Triage view and the Screening view

represent information about AIMs differently and support different tasks. The Triage view helps the analysts prioritize drugs to be looked at and analyzed first (**R2**), while the Screening view gives them high level information on all interactions to empower them to screen out the unimportant ones (**R1**).

Design Alternatives: Other design alternatives are possible for the Triage view such as a compact barchart where the height of a bar represents the interestingness score and color depicts the label status of a DDI. A graph view has two desirable properties. First, a Triage view based on node-link diagrams is consistent with other views [QH16]. Second, with a central drug, graphs can display a larger number of interacting drugs visually separable in a small space. This makes the comparison of different interactions easy as compared to a compact barchart.

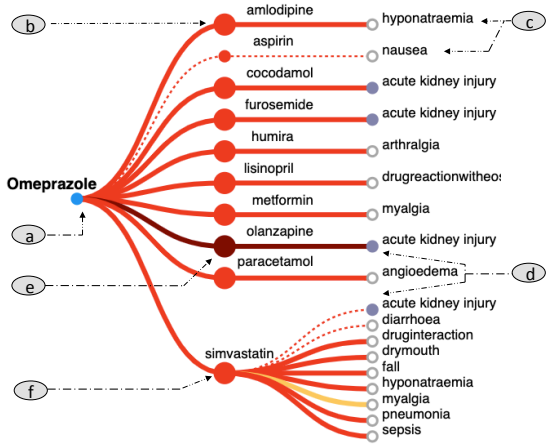

Figure 8: *Signal Forensics View: a tree layout allows an analyst to view and analyze the whole Augmented Interaction Model of a drug. (a) The root node represents the drug of interest – Omeprazole. (b) The second level represents drugs interacting with Omeprazole. (c) The third level represents the reactions related to each DDI – grey color represents non-severe reactions. (d) Severe reactions. (e) A signal with highest score - dark color. (f) Link between Omeprazole-Simavastatin depicts the aggregated score and unknown status because some of the signals are unknown.*

6.3. Signal Forensics View

Once analysts screen and prioritize the drug of interest by interacting with the Screening and Triage views, they can explore the signals further using this view by analyzing the related reactions (**R4, R5**). The Forensics view (Fig. 8) depicts a full Augmented Interaction Model (AIM) for one particular drug of focus. It not only displays but also differentiates among the scored interactions of a drug and their corresponding reactions along with their severity. There is a wide range of potential visual designs to represent drug related reactions, both containing categorical attributes. DIVA uses a modified version of tree diagram, which was found to be among the most appropriate alternative designs compared to adjacency matrices or parallel coordinates. While a tree encoding is more commonly adopted to visualize hierarchical data, they are a good fit to visualize AIMs due to the data structure and analyst's workflow patterns. A three-level tree layout represents the core attributes of AIM, i.e., the two interacting drugs and their respective reactions for every signal. Also, a drug interaction may be associated with multiple sets of reactions leading to different signals each with different attributes, i.e., interestingness scores and label status (**R2, R4**). Hence each signal can be visually encoded with a tree layout.

In the Forensics view the root node represents the drug of interest. The nodes in second level represent all the drugs that interact with the drug of interest. The nodes in third level represents the reactions related to each DDI. A horizontal tree layout is used instead of a vertical one so that reactions and labels can be easily read.

One drug pair can lead to different reactions forming multiple signals. Thus, the link between two drugs depicts the aggregate score and label status of the associated signals. For the sake of consistency among the other views, links and nodes that represent interacting drugs are encoded with an aggregated score as well as a label status. Both the link width and shape are mapped to the label status to make them visually differentiable. That is, a thick solid line means an unknown signal, while a thin dotted line represents a known signal.

We keep the repetition of ADRs for different DDIs for two reasons. One, to avoid clutter and edge crossing [yHBF08]. Two, the commonality of ADRs across many interactions is not significant in the analysis, because an ADR unknown of one DDI may be known for the others. Moreover, to see the common ADRs of all DDIs of a drug, all similar ADRs within the view are highlighted upon hover.

Design Alternatives: One possible design alternative to display signals is parallel coordinates plots (PCP). PCPs have been used in the past to visualize association rules [HC00, Yan05]. A PCP with three axis could be used to visualize drug interaction signals, two for the interacting drugs and one for the associated reactions. However, this design choice has been found to not be appropriate for two reasons. One, PCPs work well only for a small number of items along the axis. Two, PCPS are good for global patterns but local patterns are difficult to see because of high clutter. For signals, it is difficult to relate each reaction set with its corresponding drug interaction because of the extensive edge crossings and overlaps. Hence, PCPs would fail to capture the detailed view of an AIM (**R3, R4, R5**), which is indeed a core purpose of this view.

6.4. Reports View: Revealing Reports

As per (**R6**), analysts need to access the underlying reports as evidence when validating a signal. The analysts also requested to see the text narratives related to each case, as the narratives have richer information then the structured meta-data, a patient's medical history and the details of the adverse event potentially helpful in the evaluation of a signal. We thus design the Report View (Fig. 5). This view provides a line listing of the reports related to a particular selected drug or drug interaction. The Report View is linked with all three views namely the Screening view, Triage view, and the Forensics view, to give the analyst direct access to the relevant reports supporting a signal. Similarly, selecting a report highlights the corresponding drugs and ADRs in all views. The narrative section provides options to search for keywords in the text. Specific narratives are not shown in this paper due to data confidentiality as they contain sensitive information related to patients.

6.5. Visual Interactions on Linked Views

DIVA is designed to provide rapid exploration capabilities, at least compared to traditional workflows. All views are interlinked with each other, that is, all views are updated automatically as the selection of a drug is changed in any view. For instance, any drug can be made a focus by clicking on any node at any point in the Screening and Triage view. Hovering over any node highlights the

corresponding drugs in other views and also provides a tooltip with additional information. Moreover, the drug of focus in Forensics view is highlighted in the other views for context. Additionally, to give the analyst control, each view can be updated via the selection menus (Fig. 1-d). For example, the sorting feature in the Triage view allows the analysts to sort drugs either alphabetically, by the number of interactions they have, or the total number of severe reactions present within the signals. The drugs in the Screening and Triage view are by default loaded based on the analyst-assigned set of drugs. However, more drugs can be added through the drug list (Fig. 1-d). Each view can be maximized and viewed independently.

7. Evaluation

We evaluated the effectiveness and improvement opportunities in DIVA by conducting in-depth case studies and semi-structured interviews with domain experts who are drug safety analysts at the FDA. These experts also helped us in the iterative design of DIVA. After introducing the system to them, we observe them exploring the data in a think-aloud manner, and noting their feedback. During the interview, experts used the DIVA system on data from Quarter 4, 2014 (Oct-Dec, 2014) of FAERS. In total, MARAS [QKW*17] generated 1115 distinct ranked signals from this data.

7.1. DIVA Evaluation Using Case Studies

Next we describe the case study reflecting the exploration, and vetting of unknown drug interaction-related adverse reaction signals conducted by one of the drug safety analysts. The analyst is to explore the signals related to the drugs assigned to him and analyze if there is any potential new signal that needs further investigation.

Figure 9: *Forensics View for drug Lansoprazole. Interaction with Digoxin leading to acute kidney injury, a DME, is unknown, and is highly scored by the rule mining hence worthy of further investigation.*

The analyst first selects his set of assigned drugs from the drug selection menu. He begins exploring with all views updated for his set of assigned drugs. He first examines the Screening view in the DIVA system and says "At a glance I can see that I have a few 'dark red' unknown signals to investigate today" (**R1**). He adds these drugs with higher scores to the Triage view by clicking on them. Next he determines which drug to start to investigate among his selected set of drugs.

After looking at the assigned drugs in the Triage view sorted by the severity of reactions, he chooses "Lansoprazole" (Fig. 9) to explore first (**R2**). He explains "First, it seems to have more DMEs compared to others. Second, it has a highly scored unknown interaction". He also mentions "it might be quicker to start the analysis with it as there are only three interactions to analyze". To view the reactions, he opens the Forensics view (Fig. 9) by clicking on the node. He then comments "Lansoprozole, a protein pump inhibitor usually interacts with Digoxin but the resulting ADR acute kidney injury, which is a DME, is not labeled yet" (**R4, R5**).

The analyst is now curious to see if this interaction leading to acute kidney injury indeed is a safety signal. He clicks on the interaction represented by link between Digoxin and Lansoprozole to view the reports and get the details of the cases that were used to extract the signal (**R6**). He starts to explore the relevant reports (Fig. 5). He then comments "I see almost all of these reports also have another drug "Furosemide", which is used to treat kidney disorders". He further adds "the patients who were taking Lansoprozole were also taking Furosemide. That means, they might already be having a kidney disorder and Furosemide was prescribed to them for treatment. However, because of the DME we should keep an eye on it". He pins the drug for further investigation using the pin button on the Triage view.

Coming back to the Forensics view, he then analyzes the other interacting drugs, i.e., Simavastatin and Aspirin. He comments, "We are aware of the interaction with Aspirin but it is not very severe, and from its light color it seems that it is not ranked high by the mining process either. Also, interaction with Simvastatin which is used to treat high cholesterol and triglyceride level leading dizziness is labeled too. So I will not analyze them further" (**R3**).

When analyzing this first drug, "Lansoprazole" he does not find any threatening signals to further evaluate via reading the case narratives or by examining data from clinical trials. He then moves on to study his next assigned drug, as his job is to screen through all signals related to the drugs assigned to him. This time he screens "Cyclophosphamide" by examining the Triage view. He explains "though there are not many high scored unknown signals but the higher count of DMEs cannot be ignored". He then explores it further using the Forensics view (Fig. 1-c). "I see all the interactions have the DME *neutropenia* listed as an ADR, which is a labeled ADR for Cyclophosphamide itself. They all have a similar color (low score)". He hovers over the edges and explains "The report count is same for all signals, they must have been extracted from same set of reports."

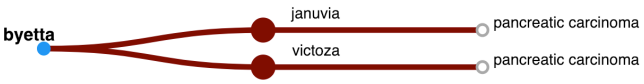

Figure 10: *Forensics View for Byetta. Interaction with Victoza and Januvia is just a co-occurrence and not an interaction as all three of these drugs belong to same drug class that treats diabetes.*

Then the analyst selects his next drug from the Triage View "Byetta" and explains "I noticed a highly scored unknown interaction with Victoza" (Fig. 10). He points out "Both of these drugs are anti-diabetic and are used to control blood sugar level". He explains further "this cannot be an interaction but it must be a mere co-occurrence. The reason can be that the patient might have changed therapies during treatment and hence these two drugs were reported together". The analyst gives similar remarks for the interacting drug Januvia which is also an anti-diabetic drug from the same class.

To examine next, the analyst selects the "Ondansetron" drug in the Triage view as it has a high score signal. Upon viewing its details in the Forensics view (Fig. 11), he observes that the ADR associated with this signal is "Serotonin Syndrome", a severe ADR. He explains, "Ondansetron is used to treat vomiting and nausea caused by chemotherapy or radiations". He further comments "FDA added a warning in Ondansetron's label based on reports to avoid concomitant (at the same time) use of Lithium that might develop

serotonin syndrome". Pointing to this known interaction, he adds "the fact that I can find the information about this signal being already labeled visually is very convenient and saves my time by not having to search for it in a separate tool" (**R3**).

7.2. Interviews with Domain Experts

We interviewed multiple target domain experts to assess the effectiveness of DIVA and validate our design choices. Before presenting the prototype to a larger audience, we first invited two domain experts for the pilot study. The goal of the pilot study was to identify potential usability issues and to gather initial feedback on the workflow of the system. The two participants explored the system on their own after we had introduced the visual designs to them. One of the participants said "The Forensics view is very intuitive and easy to read, having the focused drug at first. Then we can see how this drug is interacting with other drugs and then we see the ADRs for each interaction. Following each path is easier to understand and the DMEs being highlighted make it very easy to grasp an interesting signal." They had a few suggestions too. At first, we had separate windows for the Forensics view. However, they suggested to keep everything within one window and instead give the user control to choose a view to maximize or minimize. We added this capability to our system (Fig. 1). They also suggested to make the report view available on demand only, i.e., whenever a user wants to see the relevant reports. Other minor suggestions included, highlighting the focus drug, and keeping the report view as simple as possible.

After the pilot study, we interviewed a larger group of 10 drug safety analysts to gain a more detailed assessment of the individual components of the system. These analysts were familiar with basic visualizations such as bar charts and pie charts. Our participants tried out the system themselves. These semi-structured interviews were guided by the questions provided in Table 1. We noted their feedback during the interview. Overall, the feedback was positive. Limitations in the current system were also collected.

The analysts' comments are summarized below:

All domain experts agreed that the Triage and Forensics view were intuitive, easy to read and informative. For the overall system, they commented "This is a very useful system, the Triage view helps us to prioritize a drug for review and steps for further investigations are smooth using the Forensics and Report view", "It is easy to differentiate DMEs from non-DMEs through the highlights, as compared to reading the list or trusting one's memory", "having the ability to highlight interesting and highly scored signals is very effective in narrowing down our investigation", "the Triage view even helps in comparing two drugs, by their number of interactions or interesting interaction or the DMEs present for each drug", "this can really help us in screening potential signals faster and then finding similar case reports, without searching for them manually", "this has not been done before, it is very useful and aligns with our workflow".

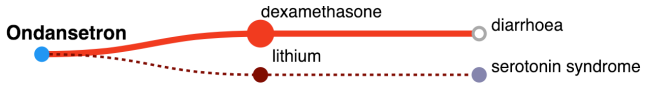

Figure 11: *Forensics View for Ondansetron. Interaction with lithium leading to severe and rare adverse reaction (ADR) 'serotonin syndrome' has been labeled by the FDA recently.*

Table 1: *Questions covered in a two hour interview with a group of drug safety analysts from the FDA.*

#	Aim	Question
1	Visual Design	Is it easy/hard to read the Triage view? Why?
2	Visual Design	Is it easy/hard to detect unknown interesting signals? Why?
3	Visual Design	Is it easy/hard to read the Forensics view? Why?
4	General	Do you think the views are intuitive and align with the work-flow? How?
5	General	Do you think the system is useful in screening and investigation of signals? How?
6	General	Which part of the visual interface can be further improved in your opinion? How?

8. Discussion

The results of these evaluation activities, which include case-studies and interviews with drug analysts, suggest that DIVA is effective at aiding analysts in identifying and verifying potential drug interaction signals. Several comments from drug analysts compared DIVA favorably to their current screening process, which primarily involves manual analysis of individual reports and manual approaches for retrieving similar reports. The multiple coordinated views in DIVA appear to align with analysts' preferred workflow of interactively escalating and investigating signals from a pool of possible candidates, in this case generated from mining techniques.

More broadly, the *scalability of visual analytics workflows* was one challenge encountered in the design and evaluation of DIVA. DIVA currently works with one year of data consisting of 1178 drugs (nodes) and 2763 candidate signals (edges). However, it is less straightforward to scale the interplay between the drug analysts and the underlying data processing algorithms and corresponding workflow in DIVA. Solutions to overcome the scalability issues in node-link diagrams such as NodeTrix [HFM07] should be explored in future. However, these interventions should also be coupled with evaluations to ensure that the domain analysts can reliably interpret and act on these hybrid displays. While there are a host of design models and activities for general visualization design, *i.e.* Munzner's Nested Model [Mun09], exactly how these activities and models map to visual analytics settings remains unclear, a gap which may be addressed in future work.

At a more practical level, our interviews with drug analysts revealed a need for the ability to incorporate additional domain knowledge. Information about known signals is available in drug package inserts (labels) and via online sources. In DIVA, data from one of these online sources (Drugs.com) is used. This dataset is known to be incomplete [SHK*15], and this gap was noticed by analysts during evaluation. To have a more complete list of known signals, integration of additional data sources such as DrugBank [Drua] is needed. More broadly, the dynamic nature of knowledge in the drug analysis space points to a future need for ways to incorporate dynamic heterogenous data, rather than static as is current practice. Dynamic data integration, however, will raise new challenges for investigative analysis, such as when new information is learned that may inform prior (*i.e.* closed) investigations.

Beyond drug analysis, although DIVA is designed for a specific domain application, some components may be adapted to other domains that focus on low-level investigative analysis. For instance, there exist similar reporting systems in other domains, such as the aviation industry, where the Federal Aviation Administration

manages Service Difficulty Reporting system [SDR, MR12] that collects reports about any malfunctions or defects in the planes. The idea of generating hypotheses about a faulty airplane using machine learning from a huge set of reports and providing means to explore and validate these hypotheses with interactive visualizations and domain expertise can be used in these application areas as well.

9. Conclusion And Future Work

In this paper, we contribute a design study for a visual analytics tool, DIVA, that supports analysts in discovering novel drug-drug interaction signals from a pool of hypothesized signals generated by machine learning techniques. DIVA, designed through interviews and requirements garnered through collaboration with drug safety analysts, uses a data abstraction augmented by a set of attributes and their relationships important for the review process. DIVA then uses a set of views that provides different perspectives of this abstraction to enable analysts to explore and verify the mined signals. DIVA provides an overview of the drug interaction space, a middle layer view consisting of small multiples node-link diagrams to show coarse-level signals, and a detail view to support validation. The results of our case-studies and interviews with drug safety analysts illustrate the effectiveness of visual analytics approaches such as DIVA for supporting Pharmacovigilance workflows.

In the future, we plan to integrate additional knowledge sources into the mining process to provide more accurate information for the review process. We also intend to integrate analysts' feedback and interaction as annotation of signals in the mining process, so that signal generation can be improved. More broadly, we plan to explore visual analytics approaches for drug analysts' style of investigative analysis, which relies heavily on evidence collection from raw reports. Finally, to address the fact that drug interactions may impact sub-populations differently, we will incorporate demographics in the visual analytics pipeline from signal generation to the visual representation. Drug interaction remains a serious public health issue. However, the use of computation and in particular visual analytics approaches show promise in improving the analytics that lead to regulatory action.

Acknowledgement

The authors are grateful for partial support from Oak Ridge Institute for Science and Education (ORISE) managed for the U.S. Department of Energy (DOE) by Oak Ridge Associated Universities (ORAU), NSF IIS-1815866, CRI-1305258 and REU CNS-1560229.

References

[AAB*11] AVERY A. J., ANDERSON C., BOND C., FORTNUM H., GIFFORD A., HANNAFORD P. C., HAZELL L., KRSKA J., LEE A., MCLERNON D. J., ET AL.: Evaluation of patient reporting of adverse drug reactions to the uk 'yellow card scheme'. *Health Technology Assessment 15*, 20 (2011), 1–234. 2

[ADK*03] ALMENOFF J. S., DUMOUCHEL W., KINDMAN L. A., YANG X., FRAM D.: Disproportionality analysis using empirical bayes data mining: a tool for the evaluation of drug interactions in the post-marketing setting. *Pharmacoepidemiology and Drug Safety 12*, 6 (2003), 517–521. 2, 6

[APP11] ARCHAMBAULT D., PURCHASE H., PINAUD B.: Animation, small multiples, and effect of mental map preservation in dynamic graphs. *IEEE Trans. on Visualization and Computer Graphics 17*, 4 (2011), 539–552. 7

[BB11a] BALL R., BOTSIS T.: Can network analysis improve pattern recognition among adverse events following immunization reported to vaers? *Clinical Pharmacology & Therapeutics 90*, 2 (2011), 271–278. 3

[BB11b] BOTSIS T., BALL R.: Network analysis of possible anaphylaxis cases reported to the us vaccine adverse event reporting system after h1n1 influenza vaccine. *Studies in health technology and informatics 169* (2011), 564–568. 3

[BC05] BUONO P., COSTABILE M. F.: Visualizing association rules in a framework for visual data mining. In *From Integrated Publication and Info. Systems to Info. and Knowledge Environments*. Springer, 2005, pp. 221–231. 3

[BGB03] BLANCHARD J., GUILLET F., BRIAND H.: Exploratory visualization for association rule rummaging. In *KDD-03 Workshop on Multimedia Data Mining (MDM-03)* (2003). 3

[BKK97] BRUNK C., KELLY J., KOHAVI R.: Mineset: An integrated system for data mining. In *Knowledge Discovery and Data mining* (1997), pp. 135–138. 3, 7

[BMS*08] BUSHARDT R. L., MASSEY E. B., SIMPSON T. W., ARIAIL J. C., SIMPSON K. N.: Polypharmacy: misleading, but manageable. *Clinical Interventions in Aging 3*, 2 (2008), 383. 2

[BSG*14] BOTSIS T., SCOTT J., GOUD R., TOMAN P., SUTHERLAND A., BALL R.: Novel algorithms for improved pattern recognition using the us fda adverse event network analyzer. *Studies in health technology and informatics 205* (2014), 1178–1182. 3

[Can16] CANADA H.: Canada Vigilance Adverse Reaction Online Database. http://www.hc-sc.gc.ca/dhp-mps/medeff/databasdon/index-eng.php, 2016. Accessed: 2016-10-09. 2

[CHYN07] COUTURIER O., HAMROUNI T., YAHIA S. B., NGUIFO E. M.: A scalable association rule visualization towards displaying large amounts of knowledge. In *Information Visualization* (2007), pp. 657–663. 3

[CLH*17] CAI R., LIU M., HU Y., MELTON B. L., MATHENY M. E., XU H., DUAN L., WAITMAN L. R.: Identification of adverse drug-drug interactions through causal association rule discovery from spontaneous adverse event reports. *Artificial Intelligence in Medicine* (2017). 2, 6

[CRC07] COUTURIER O., ROUILLARD J., CHEVRIN V.: An interactive approach to display large sets of association rules. In *Symposium on Human Interface and the Management of Information* (2007), Springer, pp. 258–267. 3

[CSW16] CHENG C.-W., SHA Y., WANG M. D.: Intervisar: An interactive visualization for association rule search. In *Proceedings of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics* (2016), ACM, pp. 175–184. 3

[DDI] Druginteractionchecker. https://reference.medscape.com/drug-interactionchecker. Accessed: 2019-02-24. 3

[Drua] Drugbank. https://www.drugbank.ca/. Accessed: 2017-01-232016-11-12. 10

[Drub] Drugs.com. https://www.drugs.com/drug_interactions.html. Accessed: 2016-11-12. 3, 5

[druc] drugs@FDA. https://www.accessdata.fda.gov/scripts/cder/daf/index.cfm/. Accessed: 2017-10-3. 5

[EG01] ERNST F. R., GRIZZLE A. J.: Drug-related morbidity and mortality: updating the cost-of-illness model. *Journal of the American Pharmaceutical Association (1996) 41*, 2 (2001), 192–199. 2

[FA15] FOOD U., ADMINISTRATION D.: FDA Adverse Event Reporting System (FAERS) data statistics. www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/ucm070434.htm, 2015. Accessed: 2017-01-10. 2, 4

[FHQ11] FARRUGIA M., HURLEY N., QUIGLEY A.: Exploring temporal ego networks using small multiples and tree-ring layouts. *Proceedings of Advances in Computer-Human Interactions 2011* (2011), 23–28. 7

[HB03] HARROWER M., BREWER C. A.: Colorbrewer. org: an online tool for selecting colour schemes for maps. *The Cartographic Journal 40*, 1 (2003), 27–37. 6

[HC00] HAN J., CERCONE N.: Ruleviz: a model for visualizing knowledge discovery process. In *Proceedings of the sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2000), ACM, pp. 244–253. 3, 8

[HCF10] HARPAZ R., CHASE H. S., FRIEDMAN C.: Mining multi-item drug adverse effect associations in spontaneous reporting systems. *BMC Bioinformatics 11*, 9 (2010), S7. 2, 5, 6

[HCHB11] HAHSLER M., CHELLUBOINA S., HORNIK K., BUCHTA C.: The arules r-package ecosystem: analyzing interesting patterns from large transaction data sets. *Journal of Machine Learning Research 12*, 6 (2011), 2021–2025. 3

[HFM07] HENRY N., FEKETE J.-D., MCGUFFIN M. J.: Nodetrix: a hybrid visualization of social networks. *IEEE transactions on visualization and computer graphics 13*, 6 (2007), 1302–1309. 10

[HHHW98] HETZLER E., HARRIS W. M., HAVRE S., WHITNEY P.: Visualizing the full spectrum of document relationships. *Advances in Knowledge Organization 6* (1998), 167–174. 3

[HSW00] HOFMANN H., SIEBES A. P., WILHELM A. F.: Visualizing association rules with interactive mosaic plots. In *Proceedings of the sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2000), ACM, pp. 227–235. 3

[ISAE16] IBRAHIM H., SAAD A., ABDO A., ELDIN A. S.: Mining association patterns of drug-interactions using post marketing fdaâĂŹs spontaneous reporting data. *Journal of Biomedical Informatics 60* (2016), 294–308. 2, 6

[KGF*10] KANEHISA M., GOTO S., FURUMICHI M., TANABE M., HIRAKAWA M.: Kegg for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Research 38*, suppl 1 (2010), D355–D360. 3

[KvMC*08] KUHN M., VON MERING C., CAMPILLOS M., JENSEN L. J., BORK P.: Stitch: interaction networks of chemicals and proteins. *Nucleic Acids Research 36*, suppl 1 (2008), D684–D688. 3

[Lin08] LINDQUIST M.: Vigibase, the who global icsr database system: basic facts. *Drug Information Journal 42*, 5 (2008), 409–419. 2

[LPC98] LAZAROU J., POMERANZ B. H., COREY P. N.: Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. *Journal of the American Medical Informatics Association 279*, 15 (1998), 1200–1205. 2

[LS16] LIU X., SHEN H.-W.: Association analysis for visual exploration of multivariate scientific data sets. *IEEE Transactions on Visualization and Computer Graphics 22*, 1 (2016), 955–964. 3

[med] MedDRA. https://www.meddra.org/. Accessed: 2016-12-29. 5

[MR12] MARAIS K. B., ROBICHAUD M. R.: Analysis of trends in aviation maintenance risk: An empirical approach. *Reliability Engineering & System Safety 106* (2012), 104–118. 11

[Mun09] MUNZNER T.: A nested model for visualization design and validation. *IEEE Transactions on Visualization and Computer Graphics 15*, 6 (2009). 4, 10

[OONL02] ONG H.-H., ONG K.-L., NG W.-K., LIM E. P.: Crystalclear: Active visualization of association rules. 3

[QH16] QU Z., HULLMAN J.: Evaluating visualization sets: Trade-offs between local effectiveness and global consistency. In *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization* (2016), ACM, pp. 44–52. 8

[QKW*17] QIN X., KAKAR T., WUNNAVA S., RUNDENSTEINER E., CAO L.: MARAS: Signaling multi-drug adverse reactions. In *Knowledge Discovery and Data mining (KDD)* (2017), ACM. 2, 5, 6, 9

[QKW*18] QIN X., KAKAR T., WUNNAVA S., MCCARTHY B., SCHADE A., TRAN H. Q., ZYLICH B., RUNDENSTEINER E. A., HARRISON L., SAHOO S. K., DE S.: MeDIAR: Multi-drug adverse reaction analytics. In *IEEE International Conference on Data Engineering (ICDE) demo paper* (2018). 3

[RC94] RAO R., CARD S. K.: Table lens: Merging graphical and symbolic representations in an interactive focus+ context visualization for tabular info. In *SIGCHI Conference on Human Factors in Computing Systems* (1994), ACM, pp. 318–322. 7

[RFF*08] ROBERTSON G., FERNANDEZ R., FISHER D., LEE B., STASKO J.: Effectiveness of animation in trend visualization. *IEEE Transactions on Visualization and Computer Graphics 14*, 6 (2008). 7

[RLRV11] ROMERO C., LUNA J. M., ROMERO J. R., VENTURA S.: Rm-tool: A framework for discovering and evaluating association rules. *Advances in Engineering Software 42*, 8 (2011), 566–576. 3, 7

[Sah10] SAHAR S.: Interestingness measures - on determining what is interesting. In *Data Mining and Knowledge Discovery Handbook, 2nd ed.* 2010, pp. 603–612. 5

[SDR] SDR. http://av-info.faa.gov/sdrx/. Accessed: 2017-3-3. 11

[SFG16] SRIDHAR D., FAKHRAEI S., GETOOR L.: A probabilistic approach for collective similarity-based drug–drug interaction prediction. *Bioinformatics 32*, 20 (2016), 3175–3182. 2, 6

[SH13] SEKHAVAT Y. A., HOEBER O.: Visualizing association rules using linked matrix, graph, and detail views. 3

[SHK*15] SHOSHI A., HOPPE T., KORMEIER B., OGULTARHAN V., HOFESTÄDT R.: Graphsaw: A web-based system for graphical analysis of drug interactions and side effects using pharmaceutical and molecular data. *BMC Medical Informatics and Decision Making 15*, 1 (2015), 15. 3, 10

[TD05] TECHAPICHETVANICH K., DATTA A.: Visar: A new technique for visualizing mined association rules. In *International Conference on Advanced Data Mining and Applications* (2005), Springer, pp. 88–95. 3

[Tuf91] TUFTE E. R.: Envisioning information. *Optometry & Vision Science 68*, 4 (1991), 322–324. 7

[VEMD*11] VON EICHBORN J., MURGUEITIO M. S., DUNKEL M., KOERNER S., BOURNE P. E., PREISSNER R.: Promiscuous: a database for network-based drug-repositioning. *Nucleic Acids Research 39*, suppl 1 (2011), D1060–D1066. 3

[WS05] WYSOWSKI D. K., SWARTZ L.: Adverse drug event surveillance and drug withdrawals in the united states, 1969-2002: the importance of reporting suspected reactions. *Archives of internal medicine 165*, 12 (2005), 1363–1369. 4

[WWT99] WONG P. C., WHITNEY P., THOMAS J.: Visualizing association rules for text mining. In *Information Visualization, 1999.(Info Vis' 99) Proceedings. 1999 IEEE Symposium on* (1999), IEEE, pp. 120–123. 3, 7

[Yan05] YANG L.: Pruning and visualizing generalized association rules in parallel coordinates. *IEEE Transactions on Knowledge and Data Engineering 17*, 1 (2005), 60–70. 3, 8

[yHBF08] Y HENR N., BEZERIANOS A., FEKETE J.-D.: Improving the readability of clustered social networks using node duplication. *IEEE Transactions on Visualization and Computer Graphics 14*, 6 (2008), 1317–1324. 8

[ZZZH09] ZHANG L., ZHANG Y. D., ZHAO P., HUANG S.-M.: Predicting drug–drug interactions: an FDA perspective. *The American Association of Pharmaceutical Scientists journal 11*, 2 (2009), 300–306. 2