DIVA: Towards Validation of Hypothesized **D**rug-Drug **I**nteractions via **V**isual **A**nalysis

Tabassum Kakar, Xiao Qin, Elke A. Rundensteiner, Lane Harrison, Sanjay K. Sahoo, Suranjan De

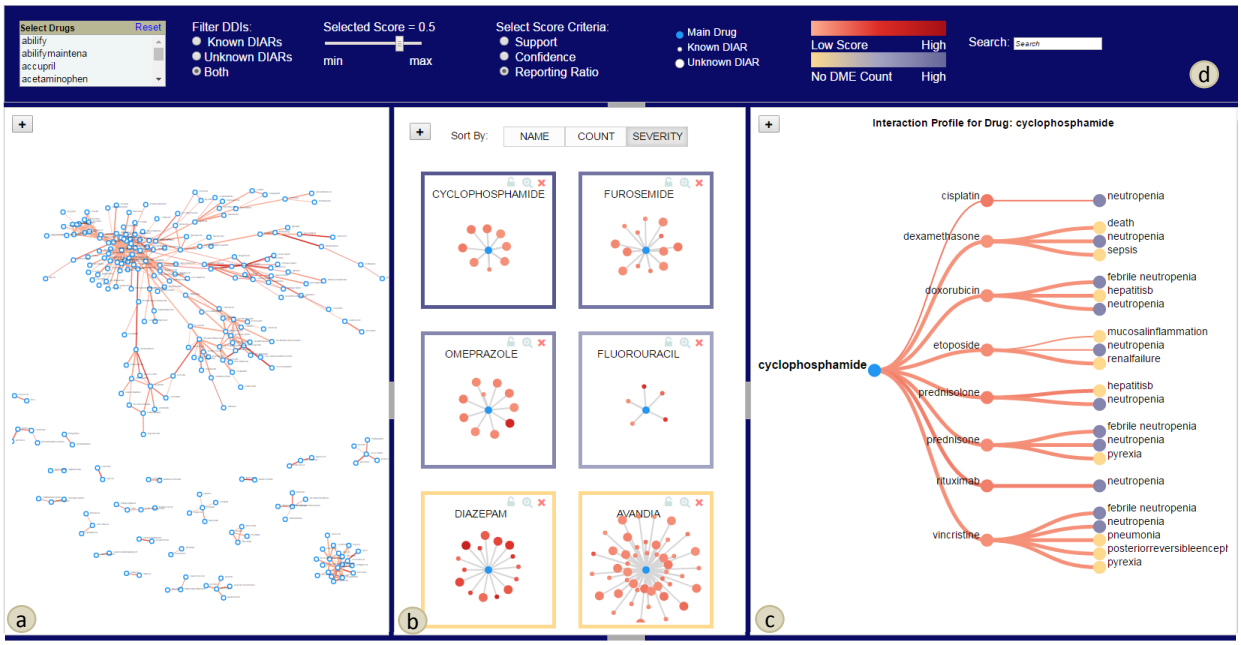

Fig. 1. The user interface of DIVA, a web-based visual analytics system for exploring and verifying Drug-Drug Interactions (DDIs) predicted by machine learning methods. (a) Overview- showing all inferred Drug-Drug Interactions (b) Small Multiples Graph View- enables drug-by-drug focused analysis of interactions, including the number of severe Adverse Reactions (ADRs). (c) Profile View- a view of the interaction profile of a drug of interest, including all ADRs triggered by each interacting drug. (d) Controls facilitate navigation between views, and direct filtering by drugs of interest.

Abstract— Severe adverse drug reactions (ADRs) caused by drug-drug interactions (DDIs) are a major public health concern. Currently, drug-drug interaction related adverse reaction (DIAR) signals are detected either manually or by applying data mining techniques on ADR reports that are submitted by drug manufacturers, health care professionals, and consumers via post drug marketing surveillance systems. Any given ADR report might inform a drug safety analyst's hypothesis of a DIAR signal that requires validation. However, drug safety review is a lengthy and tedious process, consisting of four main stages, i.e., signal screening, detection, evaluation and documentation, requiring significant evidence collection and validation by analysts. In this work, we propose a visual analytics tool, **DIVA**, that aligns with drug safety analysts' workflow to enable them to systematically generate and verify hypothesized DIAR signals. The design and feature set of **DIVA** is informed by conducting formative interviews with drug safety analysts. **DIVA** enables drug safety analysts to discover unknown DIARs inferred from ADR reports, through a hierarchy of visualizations that support analysts in prioritizing DIARs for further analysis. **DIVA**'s four layered view abstractions provide links to the underlying ADR reports and is designed to integrate with the actual drug review process, enabling analysts to determine whether a potential DIAR warrants action within their organization. We contribute an evaluation consisting of case-studies and drug analysts interviews, which validate the effectiveness of DIVA in supporting analysts in the discovery and validation of new critical DIAR signals.

Index Terms—Pharmacovigilance, Visual analytics, Hypothesis validation, Drug safety review, Node-link diagrams, Small multiples.

1 Introduction

Adverse drug reactions are a major cause of mortality, resulting in more than 100,000 deaths annually [34] with a yearly cost of over $170 billion [23] in the U.S. alone. Simultaneously, polypharmacy, the use of multiple drugs to treat medical conditions, is rising each day. According to one study, approximately 29.4% of elderly patients are taking six or more drugs [15]. Severe drug-drug interactions (DDIs) are avoidable, however, if drug safety analysts detect them early enough to take action with minimum patient exposure.

- *Tabassum Kakar is with Worcester Polytechnic Institute (WPI). E-mail: tkakar@wpi.edu.*
- *Xiao Qin is with WPI. E-mail: xqin@wpi.edu.*
- *Elke Rundensteiner is with WPI. E-mail:rundenst@wpi.edu.*
- *Lane Harrison is with WPI. E-mail:ltharrison@wpi.edu.*
- *Sanjay K. Sahoo MS. MBA. is with U.S. Food and Drug Administration (FDA). E-mail: Sanjay.Sahoo@fda.hhs.gov.*
- *Suranjan De MS. MBA. is with the FDA. E-mail: Suranjan.De@fda.hhs.gov.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

Before approval, new drugs are tested for interactions with existing drugs using both in vivo and in vitro methods [53], where in vivo refers to experimentation using living organisms such as animal studies and clinical trials, and in vitro refers to the technique of performing a given procedure in a controlled environment outside of a living organism, such as using a test tube to examine a cell.

However, any given drug may interact with other drugs in a number of ways. Further, other factors such as genetics and the environment may also influence interaction. This makes it impractical to completely test drugs before they get released to the public. It is necessary therefore for drug regulatory authorities to collect, analyze and interpret information relevant to patient safety via post marketing surveillance also known as Pharmacovigilance.

The U.S. Food and Drug Administration (FDA) conducts post marketing surveillance via the FDA Adverse Event Reporting System (FAERS). Similar systems operate internationally by WHO [35] and other western countries such as Canada [17] or Britain [10]. Post marketing surveillance is essential for detecting unanticipated interactions that occur between drugs used concurrently by the general population.

Every year, FAERS receives approximately one million ADR reports [24]. FAERS reports include mandatory reports submitted by drug manufacturers and voluntary reports submitted by health care professionals and consumers. These reports contain semi-structured information about patients' demographics, drugs being taken, ADRs being observed, therapies and outcomes. These reports are rich source of information for discovering potential DDI related ADR (DIAR) signals defined as possible causal relations between a set of ADRs and a drug combination. Due to the large number of reports and limited staff to analyze them, manual detection of potential DIAR signals is not only extremely tedious and error prone but also rather inefficient.

One approach to address this problem is to utilize computational methods such as probabilistic soft logic [45], bayesian methods [8] and association rule mining (ARM) [16, 27, 31] to detect potential DIARs. Such approaches are a step beyond manual analysis, as they support large scale analysis as well as the detection of novel and unexpected signals during post marketing surveillance.

Computational approaches to post marketing surveillance also introduce new challenges and suffer from significant limitations. One challenge is the complexity and sheer cardinality of hypothesized DIAR signals generated by computational techniques. For example, for a dataset of n distinct items, association rule mining generates a set of rules in the order O(2^n), where each rule represents a possible DIAR signal. Therefore, a set of thousands of drugs and ADRs will generate millions of hypothesized DIAR signals, not all of which are truly a cause for action. This poses a major challenge for the drug review analysts, who are ultimately responsible for investigating all possible signals. A second severe shortcoming of state-of-the-art is that computational techniques dissociate candidate interactions from the actual FAERS reports supporting their hypothesis. Based on our interviews with FDA analysts, the low-level analysis of raw reports is the backbone of the drug review process for the validation of potential interactions requiring regulatory action. The limitations of computation can be addressed to some degree, namely pruning and ranking strategies [27, 31] that reduce the number of generated DIAR signals and highlighting the most meaningful ones. However, these strategies still leave thousands of rules for examination, and the priority set by pruning and ranking algorithms bring interpretability challenges to analysts who must ensure adequate coverage of the FAERS reports.

Another possible direction to tackle these challenges is to draw from prior work on visually representing association rules [11, 20, 42]. However, these displays are primarily designed to support a **global** overview of rules. This is in contrast to drug review analysts' goals, providing means of sense-making of the items within the rules, i.e., drugs and ADRs in a domain context to help in decision making of whether a particular rule (DIAR signal) really is a potential safety signal and worth further investigation.

What is needed is a visual analytics approach- one that integrates the analysis process of drug analysts with computational techniques and visual interfaces. In this paper we therefore propose a system - Drug Drug Interactions via Visual Analysis (**DIVA**), that presents a new perspective on integrated computational and visualization approach for analyzing DIARs. On the one hand, **DIVA** adopts computational methods to detect DIAR signals from raw FAERS reports, on other, provides multiple visualizations with varying levels of detail. In addition, basic provenance functions designed to aid drug analysts by the discovery, analysis, and vetting of harmful DIARs.

The design of **DIVA** was informed via studying the work-flow of FDA drug safety analysts. At the *macroscopic* level, the interactions between all drugs are revealed via a network visualization using information obtained via association rule mining. We introduce a novel middle layer, a small multiples node-link representation, which enables drug review analysts to parse first-order drug interaction information, and to focus on a set of drugs that are assigned to the analysts, reflecting the real-world situation at the FDA. At the *microscopic* level, each candidate interaction can be viewed in context of the raw reports associated with it for further validation. When an analyst has vetted a particular DIAR, they can report it as an actionable finding or suggest to continue monitoring it using future reports.

DIVA's visual analytics approach to the drug review screening fills a tangible need in the early detection of severe DIARs. It also presents a perspective on the interpretability of large-scale data mining techniques that allow analysts to move between levels of abstraction building trust in the results of computational approaches. **DIVA** can be considered as a personalized analytics system for domain experts as it aims to help analysts with specific expertise more effectively identify drugs of interest from a potentially overwhelming set of choices, guiding them towards useful information that is not only interesting but also unknown.

The primary contributions of this work are:

- A comprehensive study of the drug safety review process and requirements, forming a basis for visualization centric interactive DIAR detection and verification systems.
- Our interactive visual analytics strategy allows drug safety analysts to systematically explore the space of DIAR signals generated by association rule mining techniques.
- The **DIVA** system, which realizes the proposed requirements and functions to aid drug safety analyst in different stages of the review process such as signal screening, detection and verification.
- Case studies with real world datasets that demonstrate the utility of visual analytic tools like **DIVA** for finding and validating severe unknown DIARs.

Below table summarizes the terminologies used throughout the paper.

Terms	Meaning
DDI	Drug-Drug Interaction
HDDI	Hypothesized DDIs generated by data mining
ADRs	Adverse Drug Reactions
Signal	A possible causal relation between a drug and ADRs
DIAR Signal	ADRs that might have been caused by a DDI
DME	Designated Medical Event - a severe ADR

2 TASK CHARACTERIZATION

In this section, we describe the procedure used for extracting user requirements, and present our design rationale.

2.1 Working with Domain Experts

To design DIVA, we worked closely with two domain experts involved in the drug safety review process. We organized a series of formal interview sessions with these experts. In the first two interviews, we had in-depth conversations with them to understand current drug review process and identify their challenges in analyzing DDIs. We also showed them sketches of our initial design to gather a list of concrete design requirements. In the third interview, we presented a working prototype of our system to these experts to confirm whether it meets the design requirements. In the last session, a larger group of domain experts explored real FAERS data using our refined and fully working

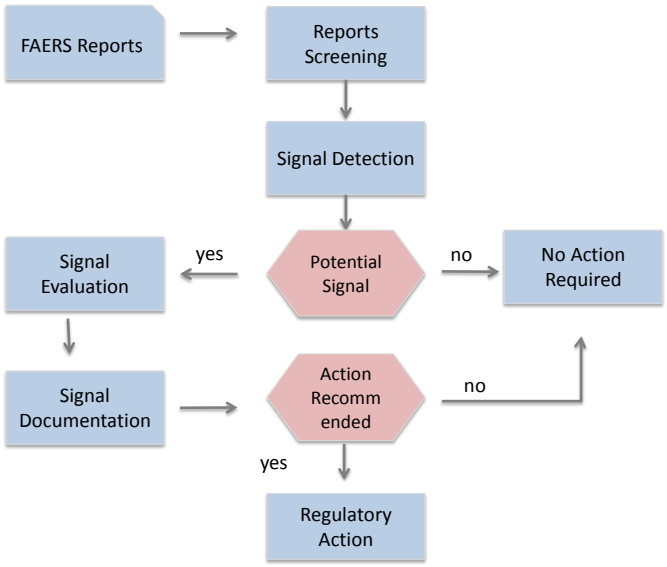

Fig. 2. Current Drug Review Process: Drug safety analysts first screen the huge-pile of reports (1 million a year) for any red flags such as severe ADR or severe outcome such as death. Second, they explore the signal by reviewing the narratives and other similar cases. If it is considered a potential signal then further investigation is done by evaluating the data from clinical studies and literature. After enough supporting evidence is found then recommended action is taken.

system which helped in identifying the strengths and weaknesses of our system. Moreover, the use cases discussed later in section 6.1 are derived based on this last session.

2.2 Current Drug Review Process

At a high-level, the goal of drug safety analysts is to identify potential novel safety signals related to drug-drug interactions, and take action if enough supporting evidence is found during the evaluation of a potential signal. However, this process is complex and composed of many steps as shown in Figure 2. For signal detection related to single drugs, FAERS captures structured information that analysts can use to screen for any red flags such as a severe ADR or serious outcomes of a given report such as death or hospitalization using pre-computed database queries. As a next step these analysts explore whether a screened signal really is a potential safety signal by collecting information from other similar reports. If they find evidence supporting further investigation then they evaluate the finding by examining the actual reports, their text narratives, data from clinical trials and other relevant medical literature to find evidence supporting a potential signal. If enough evidence is found the review is documented and recommendations are formulated. Appropriate regulatory action is taken based on the recommendations, such as changing the drug label or restricting the drug usage or in severe cases even banning drug from the market. In the case of DDIs, FAERs does not directly capture structured information about drug interactions. Thus the only way to screen for DDIs is reading the narratives. This however, can be very time-consuming and cumbersome given the huge size of the data received every year. Moreover, these domain experts have other regulatory tasks to perform besides the analysis of huge reports for detection of safety signals. This makes the need of a visual analytical system that can make this whole process efficient, even crucial. Hence, our proposed technique can present a set of hypothesis about DIARs that can be further explored and validated interactively in an efficient manner.

2.3 Functional Requirements

Based on our discussions with the domain experts, we have formulated a set of requirements to guide the design of our system. The overall project took ten months. The design requirements were formulated iteratively throughout the course of the project.

The following design requirements are identified for *HDDIs screening*:

- **R1:** ***Provide an overview of all HDDIs.*** Given the huge number of HDDIs generated by the computational methods, a compact overview of HDDIs is required to get the gist of the data. This can help an analyst screen the non-important DDIs and narrow down the search space to focus on interesting ones.
- **R2:** ***Facilitate the detection of interesting HDDIs.*** Because of the huge size of the HDDIs, the interesting HDDIs should be highlighted by the system so that the analyst could further investigate them quickly, without losing focus by wasting time in exploring the non-important HDDIs.
- **R3:** ***Enable analysts to change importance criteria for HDDIs interactively.*** The importance criteria is defined by the data mining technique that rank the most strong drug-drug interactions. For example, reports count supporting a HDDI or strength of causality of the DIAR signal. As various ranking measures have been proposed in recent studies, analysts should be able to change this criteria interactively.
- **R4:** ***Facilitate detection of severe ADRs.*** Analysts are more interested in DDIs leading to a severe ADR such as heart attack or kidney failure even sudden death rather than headache or nausea. Hence, these severe ADRs should be highlighted to be more easily detectable.
- **R5:** ***Facilitate prioritization of drugs.*** Analysts review reports based on a set of drugs assigned to them (on average 25). Hence they need a way to prioritize particular drugs in the system.

For *HDDI verification* we identify the following design requirements:

- **R6:** ***Linkage to reports.*** (a). Domain experts have suggested that showing a line-listing of reports associated with a DDI is essential to evaluate findings. (b) Moreover, the option to view the text narrative of the reports should be provided to be able to further investigate the hypothesis, as the text contains richer information about the patients, such as their medical history etc, that can help an analyst in the validation process.
- **R7:** ***DDI Annotation.*** Analysts may find a DDI that is marked interesting based on a given interestingness criteria, uninteresting in practice. Hence, the analyst should be able to annotate such finding. The system should record such feedback and incorporate it into the data for future analysis.

Following requirements are important for designing the whole system:

- **R8:** ***Support smooth and interactive exploration of FAERS data.*** As the number of FAERS reports increases day by day, the system should be scalable to support interactive exploratory analysis of a large dataset.
- **R9:** ***Use familiar visual metaphors and respect user's mental model about drugs and ADRs.*** Since only few of our target users have experience with advanced visual analytics systems, it is important to keep the visual designs simple, intuitive and easy to understand. Therefore, we considered these aspects during making the design choices.

3 Data Abstraction

3.1 DIAR Signal Extraction

FAERS reports contain information about patient demographics, drugs taken by patients and the observed ADRs. These reports do not capture direct information about DDIs. However, many studies such as [28] have emphasized the importance of this data providing an opportunity to uncover DDIs using data mining algorithms (DMAs), as these reports contain populations that are not well represented in clinical trials. However, further analysis of the results generated by these DMAs by humans is required [46]. These works include Bayesian learning [8] and Probabilistic soft logic [45] to extract DDIs.

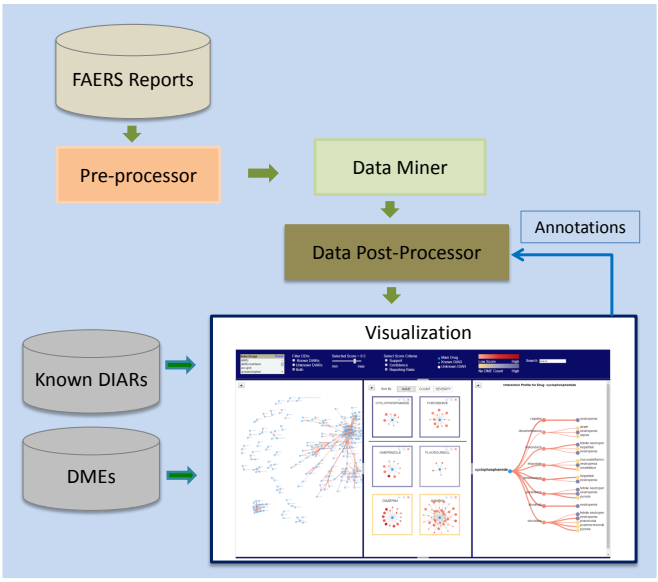

Fig. 3. Overview of DIVA framework consisting of a data preprocessor, data miner, and data visualization module. Annotations from visualization are used to remove certain DIARs that are validated by analysts as duplication or co-occurrence.

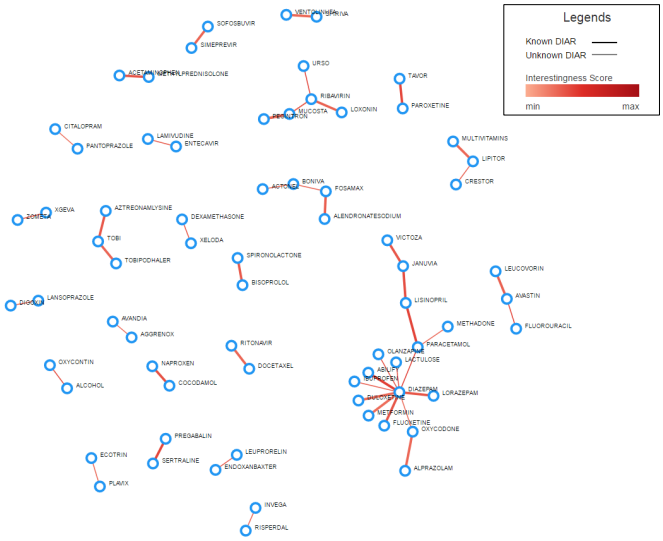

Fig. 4. DDI-Overview: To get a gist of the number and type of interaction between drugs. Each node is a drug and edge is an interaction between two drugs. Edges width is mapped to whether a DIAR is known or unknown. Color of edges depicts the interestingness score.

Association rule mining [6], a well established DMA designed for discovering interesting relationships among variables in the data, making it a natural fit to extract DIAR signals [16, 27, 31].

3.2 Association Rule Mining (ARM)

In order to understand ARM, a brief introduction is given below: Let $\mathcal{I} = \{i_1, i_2, ..., i_n\}$ represent a set of **items** in a collection of database transactions $\mathcal{T} = \{t_1, t_2, ..., t_m\}$ such that each transaction in $\mathcal{T}$ contains a subset of the items in $\mathcal{I}$.

Definition 1 *An **association rule** is an expression of the form* $\mathcal{R} \equiv \mathcal{X} \Rightarrow \mathcal{Y}$*, where* $\mathcal{X}$ *and* $\mathcal{Y}$ *are itemsets and* $\mathcal{X} \subseteq \mathcal{I}$, $\mathcal{Y} \subseteq \mathcal{I} \setminus \mathcal{X}$.

Here the left hand side of the arrow, i.e., X, denotes the antecedent and the right side, i.e. Y, denotes the consequent. Each association rule is referred to as an itemset. An itemset contains k items in a k-itemset. In the case of signaling DDI related ADRs, X denotes a set of drugs while Y denotes a set of ADRs. For example, the rule Warfarin, Aspirin $\rightarrow$ bleeding is a 3-itemset rule that implies that taking Warfarin and aspirin together might lead to bleeding.

Though ARM is very good at finding hidden relationships among data items, the major drawback of ARM is the huge amount of generated rules. The total number of rules generated from a dataset with n distinct items is in the order $O(2^n)$. To overcome this many studies have suggested ranking strategies based on measures to select interesting associations among items.

Few of the existing measures to select interesting rules are given below while research continues to explore new ones.

Measures. Many measures [41] have been proposed to evaluate the interestingness of associations. The most common ones *support*, *confidence* and *lift* (RR) for an association $\mathcal{R}$ are defined as follows:

$$support(\mathcal{R}) = P(\mathcal{X} \cup \mathcal{Y}) = |\mathcal{X} \cup \mathcal{Y}|. \quad (1)$$

$$confidence(\mathcal{R}) = P(\mathcal{Y}|\mathcal{X}) = \frac{|\mathcal{X} \cup \mathcal{Y}|}{|\mathcal{X}|}. \quad (2)$$

$$lift(\mathcal{R}) = \frac{P(\mathcal{Y}|\mathcal{X})}{P(\mathcal{Y})} = \frac{P(\mathcal{X}|\mathcal{Y})}{P(\mathcal{X})} = \frac{|\mathcal{X} \cup \mathcal{Y}| \times m}{|\mathcal{X}| \times |\mathcal{Y}|}. \quad (3)$$

Support (Formula 1) describes the proportion of the transactions that *contain* all items in the association. *Confidence* (Formula 2) describes the probability of finding the *consequent* $\mathcal{Y}$ of the association under the condition that these transactions also *contain* the *antecedent* $\mathcal{X}$. It is a maximum likelihood estimate of the conditional probability $P(\mathcal{Y}|\mathcal{X})$. *Lift* (Formula 3) measures how many times more often $\mathcal{X}$ and $\mathcal{Y}$ occur together than expected if they are statistically independent. Lift is also known as relative reporting ratio (RR), which attempts to quantify the degree of unexpectedness of a drug-AE association [9]. Both the FDA and WHO use an adjusted version of RR as a basis for monitoring safety signals in their systems.

3.3 Hypothesized DIAR Signals

Each rule generated by association rule mining has a set of interacting drugs and a set of ADRs forming the hypothesis that the ADRs might have been triggered by the drug-drug interaction. For sake of simplicity and practical implications we only considered two drug-interactions. We call each rule an inferred or hypothesized DIAR signal. Some of these DIAR signals might be mere co-occurrences of drugs in the ADR reports, while others might be a potential signal worth an investigation.

4 DIVA System Overview

The architecture of DIVA system consists of four major components as shown in Figure 3: a data pre-processing module, a data mining module, a data post-processing module and a visualization module.

4.1 Framework

Data Pre-processor: To ensure the **reproducibility** of this experiment, we used the public version of FAERS data which is available online on a quarterly basis. The data has information about patient demographics, drugs taken, therapies, ADRs, outcomes etc. The data pre-processor module integrates these separate data files and removes any duplicate cases. Due to different variants of same drug and spelling mistakes, the drug names are cleaned for analysis purposes using drug names lexicons and regular expressions.

Data Miner: The data miner module uses association rule mining to extract and score the DIAR signals using the pre-processed data. **DIVA** adopts the *Apriori* algorithm [7] with a minimum support threshold of 10, to align with the fact that the FDA investigates a case even if there are few reports indicating a safety signal based on the quality of the reports. This module can be replaced by any other data mining techniques that are capable of extracting and scoring DIAR signals.

Data Post-processor: The data generated by the data mining module is transformed to a tabular format to be visualized. Each report has a unique id, a combination of case id and case version that uniquely identifies the reports. During the mining process, this information is

kept and is used to retrieve the relevant reports for each HDDI. The post-processor also maps the ADRs and drug names back to their original terms used in the reports. Additionally, the resulting data is manipulated to adjust for two way drug interactions. For instance, the interaction Aspirin-Avandia leading to hypertension is available as an interaction for both Aspirin and Avandia. After verification, if an interaction is evaluated to be a co-occurrence or duplication, the post-processor uses that feedback to remove such rules during future mining processes.

Visualization Module: The visualization module allows the user to explore and validate DIAR signals using multiple coordinated views (Figure 1). At the macroscopic level, the DDI-Overview gives a glimpse of all hypothesized drug drug interactions (Figure 1-a). The Galaxy view, composed of small multiples, helps an analyst get an overview of interactions related to a particular drug or set of drugs and helps them screen and prioritize a drug for review based on the aggregated interestingness of the interactions within a galaxy (Figure 1-b). The Profile view includes ADRs related to each DDI to further explore a particular drug-drug interaction (Figure 1-c). In order to further investigate a drug interaction, at a lowest level we provide a reports view that visualizes the line-listing and text narrative of raw reports associated with a selected drug or DDI (Figure 9). We develop the visual interface of DIVA by following the aforementioned design rationales (Section 2). All views are coordinated via interactive linking, allowing for exploration and validation of DDIs.

4.2 External Data Sources

Besides FAERS reports, the data sources below are used.

Known DIAR Knowledge: The results generated by ARM can correspond to both known and unknown DIAR signals. By known, we mean a DIAR signal that is already known by the literature or post marketing surveillance, where as unknown means a novel DIAR signal that has not yet been noticed in clinical trials or in the literature. While analysts want to keep the knowledge of known DIAR signals handy, they are more interested to explore and validate novel DIAR signals. Information about known DDIs is available via many online resources such as DrugBank [2], Drugs.com [3] etc. To make the screening efficient, we extracted the known DIAR knowledge using Drugs.com and incorporated it into our system. This empowers the analyst to easily pin-point an interesting novel DIAR signal and explore and validate it.

Designated Medical Events (DMEs:) FDA maintains an internal list of designated medical events (DMEs) that are adverse events that are considered rare, serious, and associated with a high drug-attributable risk. Such events constitute an alarm with as few as one to three reports. Examples include hepatic failure, aplastic anemia, etc. This list is created to be used by drug safety analysts to focus and prioritize safety signal detection activities.

5 DIVA Design

We developed DIVA to fulfill the aforementioned requirements (Section 2). In this section, we describe the visual design of DIVA. DIVA is composed of four coordinated interactive views that assist the exploration of the drug-drug interactions by presenting all DDIs, drug specific interactions, drug-specific DIAR signals and reports. When designing DIVA we follow the design mantra of overview first, filter and details on demand [43], and iterative check-ins with domain experts as per Munzner's Nested Design Model [36] as a guiding framework for interaction and exploration.

5.1 Overview: Getting the Gist

The overview enables analysis of the drug-drug interactions at a macroscopic level (**R1**). This allows analysts to see the entire space of HDDIs. We use a node-link diagram view to characterize drug interactions as depicted in Figure 4. The familiar node-link diagram allows us visually encode relationships among drugs and ADRs (**R9**). Here nodes represent the drugs, while edges depict an interaction between two drugs. The color of the edge encodes whether an interaction is known (orange) or unknown (maroon). The size of the edges is mapped to the strength of the interaction as determined by the mining technique (an interestingness score) based on support and confidence, discussed in

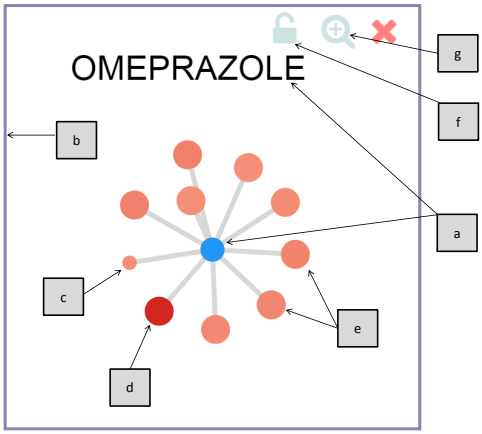

Fig. 5. Galaxy View for drug Omeprazole, each node is a drug : (a) Drug of interest Omeprazole is shown in center. (b) Color of box represents total number of DMEs present in all interactions of Omeprazole. (c) Interacting drug with a known and low score DIAR signal. (d) Interacting drug with at least one unknown and high score DIAR signal. (e) Interacting drugs with an unknown and low score DIAR signal. (f) Option to pin this galaxy. (g) Option to zoom in to view the profile of Omeprazaole.

Section 3.2). For example, if Aspirin has an interaction with Avandia, then Aspirin is connected both to Avandia (directly) and all other drugs that are interacting with Avandia (indirectly) and so on.

The goal of this view is to invite high-level comparisons between DDIs to help an expert in 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, as their goal is the detection of a novel DIAR signals with minimum patient exposure. Moreover, an overview can enable a team leader to track where in the space of possible DDIs their analysts are investigating.

5.2 Galaxy View

We designed the galaxy view using the small multiples technique [48] to align with the work-flow as analysts are typically assigned a set of drugs to review. Small multiples, either in the form of images or matrices, are used widely in information visualization [48] to study dynamic graphs, animations and trends [9, 39]. However, to the best of our knowledge, they have not been used to visualize association rules or entity specific information.

As studies have shown that small multiples display [49] help users to quickly explore data, we designed a middle layer (Galaxy View) using this technique. Each small multiple (galaxy) represents a drug and all its associated HDDIs. The analyst can get an overview of the assigned drugs and at a glance they can pick the drug with comparatively larger number of DMEs (**R4**) without viewing any details of the DIAR signal such as ADRs. The aggregated DME count is mapped to the color of the outer box of the galaxy. The center node represents the drug of interest and nodes surrounding it depict all other drugs interacting with the drug of interest shown in Figure 5. The option to pin a Galaxy is provided to facilitate the analyst in prioritizing and locking a drug to further explore and analyze (**R5**). It also helps an analyst to maintain context, so that they can resume their work and pin on where they left off next time in case they do not finish the review of a particular drug in one go. The color of nodes is mapped to the aggregated interaction score discussed below:

Let $\mathcal{R} \equiv \{\mathcal{R}1, R2, ..., Rn\}$ represent n DIAR signals (association rules) with two drugs D1 and D2 but different set of ADRs. Let $\mathcal{S} \equiv \{\mathcal{S}1, S2, ..., Sn\}$ be the interestingness scores for each rule respectively. The aggregated score for the interaction of D1 and D2 is given as:

$$Score_{max}(\mathcal{S}) = max\{\mathcal{S}1, S2, ..., Sn\}. \quad (4)$$

An interaction between two drugs can lead to different DIAR signals, each of which can have a different score. Because each interaction is

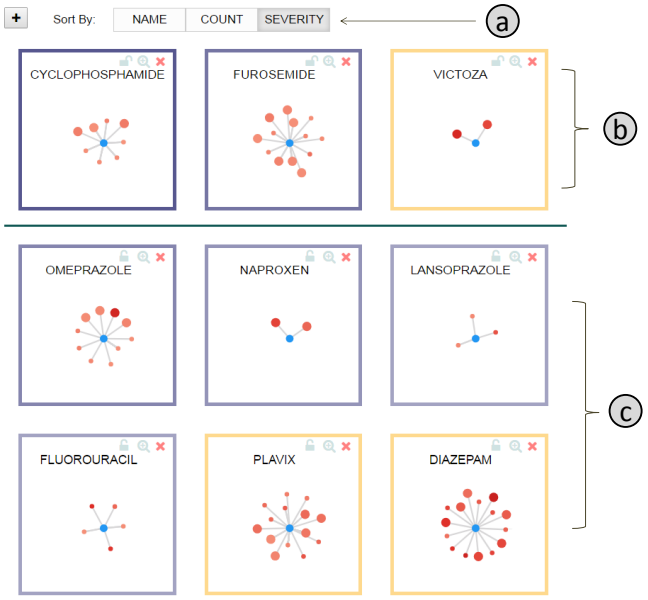

Fig. 6. The Galaxy View specific to selected set of drugs sorted by number of DME (severe ADRs) counts for the drug of interest within each galaxy. (a) Option to sort galaxies. (b) Drugs pinned by an analyst based on their interestingness for further analysis. (c) Selected drugs.

only represented once, we use the maximum score given in Formula 4 of all DIAR signals as an aggregated score to represent a DDI in the Galaxy view. We chose the maximum score as an aggregated score mapped to the color of the node even if one DIAR signal for a DDI is scored higher, so the analyst can point it quickly as opposed to using an average of scores that might cause a high score DIAR signal to be missed. Similarly, even if one of the DIAR signals is unknown, the node is marked unknown (larger size) in the galaxy to grasp analyst's attention and avoid missing out detection of any novel signal.

Additional 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 any node. It is important to understand that both Galaxy and Overview are needed because they represent information about DDIs differently. The Galaxy view helps an analyst prioritize drugs to look at first while the DDI-Overview gives them high level information on number of interactions and interestingness.

5.3 Profile View

Although the Galaxy view provides an overview of a selected drug, but in order to see the ADRs related to each DDI for further analysis of the DIAR signal we designed the Drug Profile View (Figure7). An analyst can zoom into any galaxy to view the profile of the corresponding drug. Though there is a wide range of visual representation techniques that could potentially be used for categorical data, we chose a modified version of tree diagram to be the most appropriate among the alternatives such as adjacency matrices or parallel coordinates, even though tree is more commonly adopted to visualize hierarchical data. The first goal was to visually separate all three attributes, i.e., the two drugs and the ADRs for every DIAR signal. Second, one should be able to easily find the ADRs associated with each DDI. Third, it is usual for a drug interaction to be associated with multiple sets of ADRs leading to different DIAR signals.

The Profile view is a tree with three levels. The first level or root node is mapped to the drug of interest. The second level represents all the drugs interacting with the drug of interest. The third level represents the ADRs related to each DDI. We chose the horizontal layout instead of vertical so that ADRs can be easily read. Note that, as the data represents association rules, there can be many rules representing the same DDI each with a distinct ADR. Hence each DIAR signal might have a different score. Further, as there can be multiple DIAR signals presented by the same DDI, the link between two drugs depicts the aggregated score and status (whether known or unknown) of the associated DIAR signals, as discussed in the Galaxy view. Similar color encodings are used in all views for sake of consistency. For easy readability, only DIAR signals related to the selected DDI are highlighted upon mouse-over.

5.4 Reports View: Revealing Original/Raw Reports

To enable analysts to examine the raw data at the lowest level (**R6**), we design the Report View (Figure 9). It provides a line listing of the cases (reports) related to a particular selected drug or drug interaction. Similarly, selecting a case in the report view, highlights the corresponding drugs and ADRs in the Overview, Galaxy and Profile view. The analysts also requested to see the text narratives related to each case when needed, as the narratives have richer information about a patient's medical history and the adverse event that can help an analyst in the evaluation of a particular DIAR signal besides the demographics and adverse event data present in the line listing. It has also been suggested to keep the report view simple and intuitive. The Report View is linked with all three views namely Overview, Galaxy and Profile View, to give the analyst immediate access to the reports. Options to search for keywords is provided for the narrative section. The narrative is not shown in this paper due to data confidentiality as it contains sensitive information related to the patient.

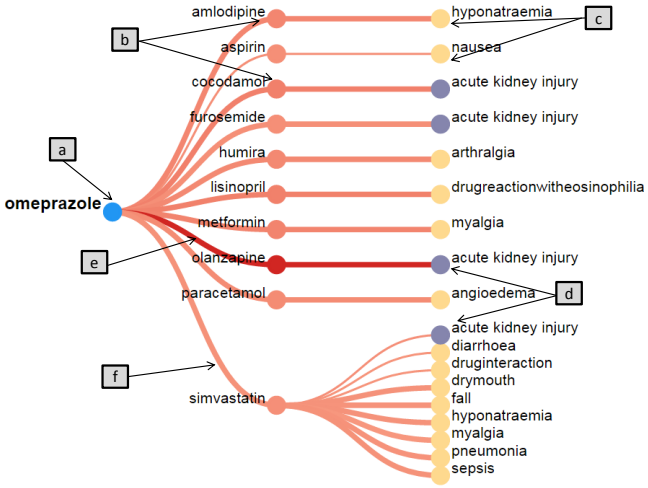

Fig. 7. Profile View: a tree layout specific to a drug of interest- Omeprazole. (a) The root node represents the drug of interest. (b) The second level represents drugs interacting with the root node. (c) The third level represents the ADRs related to each DDI. (d) DDI leading to severe ADRs - purple color. (e) shows a DIAR signal with highest score - dark color. (f) Link between Omeprazole-Simavastatin depicts the aggregated score and unknown status (wide edge) as some of the DIAR signals are unknown. Each path starting from root till leaf is a hypothesized DIAR signal and represents an association rule.

5.5 DDI Annotation

Based on the analysis, using the Profile view an analyst can annotate a particular DDI (**R7**) as one of these three categories: (1) Continue Monitoring: if they need to further investigate it by monitoring it, or (2) Duplication: if a DDI is a therapeutic duplication, i.e. drugs taken together for the same treatment, or (3) Co-occurrence: if drugs in the DIAR signal are appearing by chance and there is no logical reason for an interaction. This annotation is recorded by the system and is used during the post-processing phase to remove such DDIs from the results of the mining process for future analysis.

5.6 Interactions and Linked Views

DIVA is designed with goal of providing robust exploration capabilities (**R8**). All views are interlinked with each other, that is, all views are updated automatically as the selection is changed in any one view. Moreover, to give the analyst control, each view can be updated via the selection menu (Figure 1-d). The sorting feature in the Galaxy View helps an analyst to sort drugs alphabetically, by the number of interactions they have, total number of DMEs present within the DIAR signals. Each view can be maximized and viewed independently. Also, an option to plug in an alternate data mining scoring criteria (**R3**) is provided (Figure 1-d).

5.7 Design Alternatives

Based on the requirements discussed in Section 2, we explored a larger design space of visual encodings suitable for this data and tasks. These candidate views were eliminated based on analysis of the requirements collected, and through periodic interviews with drug analysts.

5.7.1 Table View

The simplest method to show association rules or data with relationships is tabular format [40], where each attribute can be a column and each row corresponds to a DIAR signal. TableLens [38] that is designed to allow users to detect patterns, correlations, and outliers in the data set using tables works better when presenting numerical data. However, as most of our data is categorical and there are thousands of drugs and ADRs having many to many relationships, a tabular format will be a cognitively demanding and tedious task. Therefore, a more effective way to communicate these complex relationships is to encode them in a graphical representation using variables such as shape, color, position and size etc.

5.7.2 Adjacency Matrix

An alternate design to encode the DDIs visually would be to use an adjacency matrix as it has been used to visualize association rules [13,51] where rows and columns map to the drugs and each cell depicts a DDI. This technique has two shortcomings, sparsity and scalability. As the number of drugs increases the size of the matrix grows as well. Second, it is not necessary that two drugs interact with a similar set of other drugs. This causes a lot of empty cells in the matrix making it difficult to read. Moreover, if the aggregated score of a DDI is mapped to the color of a cell, then mapping the known/unknown information is not possible. If the size of the cell is used to encode the status of the DDI (known/unknown), then it will be a heat map which will jeopardize the readability of DDIs.

5.7.3 Parallel Coordinates

Another design alternative is parallel coordinates (PC) also suitable to show relationships among attributes and have been used in past to visualize association rules [26, 52]. A PC with three axis can be used to visualize our DIAR signals, two for the interacting drugs and one for the associated ADRs. However, using PC is not suitable for two reasons: one, they work well only for a small number of items along the axis and two, it is difficult to relate each ADR set with its corresponding drug interaction because of the extensive edge crossings and overlaps.

6 Evaluation

6.1 Case Studies

We evaluated the effectiveness and usefulness of DIVA by conducting in-depth case studies, use-cases and semi-structured interviews with domain experts who are the drug safety analysts. These experts also helped us in the iterative design of DIVA. After introducing the system, we observed them exploring the tool in a think-aloud manner and noted their feedback. During the interview, experts used the DIVA system on data from Quarter 4, 2014 (Oct-Dec, 2014). There were an ~100,000 ADR reports and ~300,000 rules generated with a minimum support threshold of 10. After ranking and pruning the rules that did not had two drugs as antecedents and ADRs as consequents, approximately 3500 DIAR signals left.

From the overview, an analyst saw that Aspirin and Avandia which were manufactured a long time ago (~1999) had many interactions compared to Budesonide and Victoza (2010-2013) that were manufactured recently. The analyst said, "It makes sense, as these are old drugs, many reports are submitted for these drugs that leads to the possibility of many HDDIs".

6.1.1 Prioritizing Drug Review

To start with, one analyst wanted to determine with which drug to start investigation from a set of drugs assigned to him. After selecting the assigned drugs from the Galaxy view, he chose Omeprazole (Figure 5) to explore first. He explained "First, it seemed to have more DMEs compared to others. Second, it had a highly scored interactions)".

Fig. 8. Profile View for the selected drug Lansoprazole. Interaction with Digoxin leading to acute kidney injury, a DME, is unknown, and is highly scored by AR mining hence worth further investigation

6.1.2 Detecting Unknown DDI

Several patterns were identified by analysts when they used the Galaxy view for screening the interesting drugs. Out of the assigned drugs, one analyst chose Lansoprozole from the galaxy view to further investigate. He said "The first thing that I noticed is that this drug has some DME listed with it and also some unknown interactions are scored higher". After zooming in to view the ADRs using the Profile view (Figure 8), he commented "Lansoprozole, a protein pump inhibitor usually interacts with Digoxin but the resulting ADR acute kidney injury which is a DME is not labeled yet". The analyst started to explore the relevant reports (Figure 9). He commented "I see almost all of these reports also have Furosemide, which is used to treat kidney disorders, but because of the DME we should keep an eye on it" and pinned the drug for further investigation. "On the other hand, interaction between Simavastatin and Lansoprozole leading to myalgia (muscle pain) and dizziness is known, and it is very obvious from the visuals".

6.1.3 Detecting Highly Scored DDIs

One analyst screened "Cyclophosphamide" from her set of selected drugs and wanted to explore it further using the Profile View (Figure 1-c). "I see all the interactions have the DME "neutropenia" listed as an ADR, which is a labeled ADR for Cyclophosphamide itself, and they all having a low score. This makes sense".

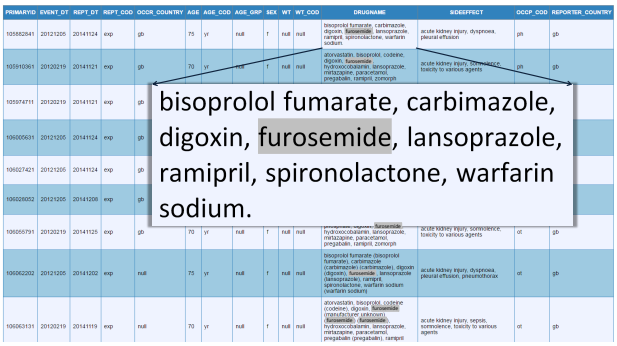

Fig. 9. FAERS Reports associated with interaction Lansoprazole and Digoxin. Almost every report has the drug Furosemide which is used to treat kidney disorders.

6.1.4 Severe ADRs Detection

One analyst who selected the Ondansetron drug in the galaxy view, noticed a high score interaction. Upon zooming in and viewing the Profile (Figure 10), he observed that the ADR associated with the high scored DDI is a DME "Serotonin Syndrome". He explained, "Ondansetron is used to treat vomiting and nausea caused by chemotherapy or radiations". He further commented "FDA added a warning in Ondansetron's label based on reports to avoid concomitant use of Lithium that might develop serotonin syndrome".

6.1.5 Recording Feedback

One analyst who was responsible for a class of drugs called "Incretin mimetics" or drugs that control blood sugar level, selected "Byetta" and noticed a high scored and unknown interaction with Victoza (Figure 11). The analyst pointed out "Both of these drugs are anti-diabetic and are used to control blood sugar level". She explained further "this is not an interaction but 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 annotated the interaction as therapeutic duplication. The analyst gave similar remarks for the interacting drug Januvia which is also an anti-diabetic drug from same class and annotated it.

6.2 Interviews with Domain Experts

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

After the pilot study, we interviewed a larger group of 10 drug safety analysts to have a detailed assessment of the individual components of the system. These analysts were familiar with basic visual techniques 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 down their feedback during the interview. Overall, the feedback was positive. Though we noticed a few limitations in the current system. such as, participants had some difficulty reading the overview because of edge crossings. The analysts' comments are summarized below:

All domain experts agreed that the Galaxy and Profile view were intuitive, easy to read and informative. For the overall system they commented "This is a very useful system, the Galaxy view really helps to prioritize a drug for review and steps for further investigations are smooth using the Profile 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 drug interactions is very effective in narrowing down our investigation", "the galaxy 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 of 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 work flow"

Although, we are unable to conduct a controlled user study due to the lack of base-line systems but we plan to perform long term usability studies and record users' experience using the system.

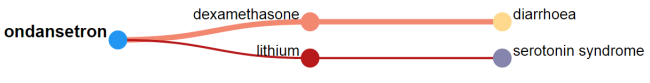

Fig. 10. Profile View for Ondansetron. Interaction with lithium leading to severe and rare ADR serotonin syndrome is labeled by FDA recently

Fig. 11. Profile View for Byetta. Interaction with Victoza and Januvia is just a co-occurrence and not an interaction as all three of these drugs i.e., Victoza, Januvia and Byetta belong to same drug class that treats diabetes.

7 Discussion

The results of our case-studies and user interviews indicate that DIVA is effective and desirable to aid FDA analysts in identifying and verifying potential DIAR signals, compared to their current screening process which involves manual analysis of multiple individual ADR reports and manual investigation by retrieving similar reports.

More broadly, the design of this system and our interactions with drug analysts highlight several ongoing threads of research in the visualization and visual analytics communities. In particular, efforts in this domain will be directly impacted by the continued advances in semi-structured text analysis and visualization, as well as advances in frameworks for ensuring the results of machine-learning and data mining techniques are understandable by end-users.

Though the current narrative view provides features for viewing text, there is room for improvement. There is an opportunity for visualization to support structured analysis of multiple documents. Several visualization systems exist focusing on text data [4, 21, 22]. These generally focus on overviews at a high level, rather than low-level tasks such as evidence collection and narrative comparison. An exception is the Overview system [12], which explicitly focuses on low- and mid-level analysis. Advances in this area, particularly those that continue to blend computational approaches with analytical workflows, will positively impact the domain of Pharmacovigilance.

Other advances will stem from work in interpretable machine-learning and data mining. Currently, the mining algorithm in this system is ran before importing into the visualization. However, as reports are generated month-by-month, tighter integration of the algorithm and visual analytics system could further reduce the turnaround between reports and action by drug regulation agencies.

At a more practical level, our interviews with drug analysts revealed requests for testing data that is closer to what they use in practice. Information about known DIARs is available in drug package inserts (labels) and via online sources. In the current prototype, data from one of these online sources (Drugs.com) is used. This information is not complete, however, and which was noticed by the analysts. To have a more concrete list of known DIARs, integration of various data sources such as DrugBank [2], is needed.

Another request from our interviews with drug analysts is an even tighter integration with the specifics of the review process. During verification, analysts often delve into medical literature as part of evidence collection. One analyst suggested integrating literature sources such as PubMed [5], DailyMed [1] for supporting further investigation of a potential DIAR signal. This raises opportunities for further supporting and investigating analytic provenance in the context of drug review.

Although many visualizations techniques presented in this system are tailored to the specific domain of drug safety analysis, some of the components may be adapted for other domains. For instance, small multiple node link diagrams can be applied to a wide range of domain areas that use association rule mining. We are not aware of prior

#	Aim	Question
1	Visual Design	Is it easy/hard to read the Galaxy View? Why?
2	Visual Design	Is it easy/hard to detect unknown interesting DDIs? Why?
3	Visual Design	Is it easy/hard to read the Profile 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 HDDIs? How?
6	General	Which part of the visual interface can be further improved in your opinion? How?

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

work exploring small multiple node link diagrams as a means for understanding rules or relationships among data.

8 Related Work

Although to the best of our knowledge, a single system does not exist that provides visualization, exploration and validation of DDIs to help the review process, but our work closely relates to solutions proposed in the following areas.

8.1 Association Rule Visualization

Visualization of association rules (ARs) is a well studied topic and various visual techniques have been explored to visualize, generalize and filter ARs. Starting with representing ARs with simple text in tabular format [40], however this work lacks an overview of the relationships to find interesting patterns. On the other hand, grids, 2D and 3D matrices [13, 37, 51] have been used to visualize association rules for small data sets. 2D Grids [13, 37] representation shows one-to-one relationships where each x and y-axis represents antecedents and consequents respectively. On the other hand, the 3D matrix view [51] shows item-to-rule relationships. Rows represent the items and columns represent the rules. The third dimension is exploited to visualize the values of support and confidence of each rule. Matrices with fish-eye view [19, 20] have also been used to visualize association rules with details. Matrix-based visualization techniques typically use numerical itemset IDs on the x- and y-axis due to the limited axis space, specially for large item sets. Rules with numerical IDs are not human-readable and require a mapping process to translate the IDs to real item names. This delays the interpretablity of the rules. Moreover, the limitation of these approaches is a lack of scalability for larger item sets.

InterVisAR [18], an interactive visualization system for association rules is proposed to search for specific rules from a huge set of rules using a two-dimentional bar chart like approach, however, the system is designed for rules with fixed length consequents. Mosaic plots [30] have also been used to visualize ARs. In this technique, 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. Rules represented by this approach are difficult to read and does not scale very well. Parallel Coordinates are used by [26, 52] to visualize rules. Each vertical line shows set of items and a rule is represented by lines or splines. However, parallel coordinates work well for smaller sets of items. Graph views are another technique that is widely used to visualize association rules [25, 29, 47]

Attempts have also been made to combine two techniques to visualize rules. [14] have used both graphs and parallel coordinates to get an overview as well as detail view of selected rules. [42] have used graphs and matrices to visualize association rules. Other attempts include a virtual arena [11] where ARs are represented as spheres positioned by the steps of an arena. The radius of the spheres is proportional to the support. The spheres are set on a cone whose height represents the confidence. This tool allows the users to select a rule (sphere) and to analyze in more details the rules related to the items of the selected rule. This approach is attractive but does not appear suitable for many rules.

Each of these techniques presents advantages and drawbacks. It is necessary to take them into account for the initial choice of our design. The effectiveness of these approaches is dependent on the input data size and type. These representations are understandable for small quantities of data but become complex when these quantities increase. Notice that our DDI-specific association rules are both huge and different with two huge distinct itemsets (drugs and ADRs) than the traditional rules with one itemset, hence design modifications are needed to accommodate data. Lastly, our goal is not just visual representation of association rules, but allowing an analyst to explore and validate drug interactions by combining their knowledge with raw data.

8.2 Drug-Interaction Visualization

In recent years, studies have explored techniques to visualize drug-interactions with proteins (targets) as well as other drugs in bio-medical domain. STITCH [33] represents Drug-Protein interactions using a network view, however it does not support ADRs or DDIs. On the other hand, PROMISCOUS [50] integrates data from three different molecular databases and visualizes Drug-Target interactions and drug related ADRs. However, it does not support DDIs. KEGG MEDICUS [32] like other free online tools is a drug interaction checker, where one drug is searched against another drug to see if interaction exists and the ADRs related to the interaction. GraphSAW [44] integrates data about known DDIs from various sources. A radial graph is used to visualize a set of ADRs and DDIs, however neither it can scale up to visualize a large number of DDIs and ADRs nor it supports exploration and validation of inferred DDIs. Moreover, none of these systems are developed to support pharmacovigilance using the ADR reports.

9 Conclusion and Future work

In this paper, we contribute a design study for a visual analytical tool, DIVA, that allows analysts efficiently discover novel drug-drug interaction (DDIs) signals from a pool of hypothesized DDIs generated by data mining techniques. DIVA, designed based on interviews and requirements garnered through collaboration with drug safety analysts, uses a set of views that enable analysts to visualize, explore and verify HDDIs. DIVA provides an overview of the drug interaction space, a novel 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 a visual analytics such as DIVA for Pharmacovigilance.

In the future, we plan to optimize the mining algorithm to scale for processing data across several years. We also plan to integrate additional known DIAR knowledge sources to provide more concrete and accurate information for the review process. 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 rule-mining to the interface.

Drug interaction remains a serious public health issue. However, the use of computation and visual analytics approaches show promise in improving the analytics that leads to regulatory action.

Acknowledgments

We are grateful to Dr. Marni Hall who was Sr. Program Director of Regulatory Science with OSE at FDA for introducing us to Pharmacovigilance in general and the ADR problem in particular. We are also thankful to all domain experts who are involved in the case studies. This work is supported in part by Oak Ridge Institute for Science and Education (ORISE) managed for the U.S. Department of Energy (DOE) by Oak Ridge Associated Universities (ORAU).

REFERENCES

[1] DailyMed. https://dailymed.nlm.nih.gov/dailymed/. Accessed: 2017-03-15.

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

[3] Drugs.com. https://www.drugs.com/. Accessed: 2016-11-12.

[4] IN-SPIRE. http://in-spire.pnnl.gov/. Accessed: 2017-01-06.

[5] PubMed. https://www.ncbi.nlm.nih.gov/pubmed/. Accessed: 2017-03-15.

[6] R. Agrawal, T. Imieliński, and A. Swami. Mining association rules between sets of items in large databases. In *ACM sigmod record*, vol. 22, pp. 207–216. ACM, 1993.

[7] R. Agrawal, R. Srikant, et al. Fast algorithms for mining association rules. In *Proc. 20th int. conf. very large data bases, VLDB*, vol. 1215, pp. 487–499, 1994.

[8] J. S. Almenoff, W. DuMouchel, L. A. Kindman, X. Yang, and D. Fram. 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):517–521, 2003.

[9] D. Archambault, H. Purchase, and B. Pinaud. Animation, small multiples, and the effect of mental map preservation in dynamic graphs. *IEEE Transactions on Visualization and Computer Graphics*, 17(4):539–552, 2011.

[10] A. J. Avery, C. Anderson, C. Bond, H. Fortnum, A. Gifford, P. C. Hannaford, L. Hazell, J. Krska, A. Lee, D. J. Mclernon, et al. Evaluation of patient reporting of adverse drug reactions to the uk'yellow card scheme': literature review, descriptive and qualitative analyses, and questionnaire surveys. *Health Technology Assessment*, 15(20):1–234, 2011.

[11] J. Blanchard, F. Guillet, and H. Briand. Exploratory visualization for association rule rummaging. In *KDD-03 workshop on multimedia data mining (MDM-03)*, 2003.

[12] M. Brehmer, S. Ingram, J. Stray, and T. Munzner. Overview: The design, adoption, and analysis of a visual document mining tool for investigative journalists. *IEEE transactions on visualization and computer graphics*, 20(12):2271–2280, 2014.

[13] C. Brunk, J. Kelly, and R. Kohavi. Mineset: An integrated system for data mining. In *KDD*, pp. 135–138, 1997.

[14] P. Buono and M. F. Costabile. Visualizing association rules in a framework for visual data mining. In *From Integrated Publication and Information Systems to Information and Knowledge Environments*, pp. 221–231. Springer, 2005.

[15] R. L. Bushardt, E. B. Massey, T. W. Simpson, J. C. Ariail, and K. N. Simpson. Polypharmacy: misleading, but manageable. *Clinical interventions in aging*, 3(2):383, 2008.

[16] R. Cai, M. Liu, Y. Hu, B. L. Melton, M. E. Matheny, H. Xu, L. Duan, and L. R. Waitman. Identification of adverse drug-drug interactions through causal association rule discovery from spontaneous adverse event reports. *Artificial Intelligence in Medicine*, 2017.

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

[18] C.-W. Cheng, Y. Sha, and M. D. Wang. Intervisar: An interactive visualization for association rule search. In *Proceedings of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, pp. 175–184. ACM, 2016.

[19] O. Couturier, T. Hamrouni, S. B. Yahia, and E. M. Nguifo. A scalable association rule visualization towards displaying large amounts of knowledge. In *IV*, pp. 657–663, 2007.

[20] O. Couturier, J. Rouillard, and V. Chevrin. An interactive approach to display large sets of association rules. In *Symposium on Human Interface and the Management of Information*, pp. 258–267. Springer, 2007.

[21] W. Dou, X. Wang, D. Skau, W. Ribarsky, and M. X. Zhou. Leadline: Interactive visual analysis of text data through event identification and exploration. In *Visual Analytics Science and Technology (VAST), 2012 IEEE Conference on*, pp. 93–102. IEEE, 2012.

[22] W. Dou, L. Yu, X. Wang, Z. Ma, and W. Ribarsky. Hierarchicaltopics: Visually exploring large text collections using topic hierarchies. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2002–2011, 2013.

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

[24] U. Food and D. Administration. FDA Adverse Event Reporting System (FAERS) data statistics. www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/ucm070434.htm, 2015. Accessed: 2017-01-10.

[25] M. Hahsler, S. Chelluboina, K. Hornik, and C. Buchta. The arules r-package ecosystem: analyzing interesting patterns from large transaction data sets. *Journal of Machine Learning Research*, 12(Jun):2021–2025, 2011.

[26] J. Han and N. Cercone. Ruleviz: a model for visualizing knowledge discovery process. In *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 244–253. ACM, 2000.

[27] R. Harpaz, H. S. Chase, and C. Friedman. Mining multi-item drug adverse effect associations in spontaneous reporting systems. *BMC bioinformatics*, 11(9):S7, 2010.

[28] M. Hauben, D. Madigan, C. M. Gerrits, L. Walsh, and E. P. Van Puijenbroek. The role of data mining in pharmacovigilance. *Expert opinion on drug safety*, 4(5):929–948, 2005.

[29] E. Hetzler, W. M. Harris, S. Havre, and P. Whitney. Visualizing the full spectrum of document relationships. *Advances in Knowledge Organization*, 6:167–174, 1998.

[30] H. Hofmann, A. P. Siebes, and A. F. Wilhelm. Visualizing association rules with interactive mosaic plots. In *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 227–235. ACM, 2000.

[31] H. Ibrahim, A. Saad, A. Abdo, and A. S. Eldin. Mining association patterns of drug-interactions using post marketing fdas spontaneous reporting data. *Journal of biomedical informatics*, 60:294–308, 2016.

[32] M. Kanehisa, S. Goto, M. Furumichi, M. Tanabe, and M. Hirakawa. Kegg for representation and analysis of molecular networks involving diseases and drugs. *Nucleic acids research*, 38(suppl 1):D355–D360, 2010.

[33] M. Kuhn, C. von Mering, M. Campillos, L. J. Jensen, and P. Bork. Stitch: interaction networks of chemicals and proteins. *Nucleic acids research*, 36(suppl 1):D684–D688, 2008.

[34] J. Lazarou, B. H. Pomeranz, and P. N. Corey. Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. *Jama*, 279(15):1200–1205, 1998.

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

[36] T. Munzner. A nested model for visualization design and validation. *IEEE transactions on visualization and computer graphics*, 15(6), 2009.

[37] H.-H. ONG, K.-L. Ong, W.-K. Ng, and E. P. LIM. Crystalclear: Active visualization of association rules. 2002.

[38] R. Rao and S. K. Card. The table lens: merging graphical and symbolic representations in an interactive focus+ context visualization for tabular information. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 318–322. ACM, 1994.

[39] G. Robertson, R. Fernandez, D. Fisher, B. Lee, and J. Stasko. Effectiveness of animation in trend visualization. *IEEE Transactions on Visualization and Computer Graphics*, 14(6), 2008.

[40] C. Romero, J. M. Luna, J. R. Romero, and S. Ventura. Rm-tool: A framework for discovering and evaluating association rules. *Advances in Engineering Software*, 42(8):566–576, 2011.

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

[42] Y. A. Sekhavat and O. Hoeber. Visualizing association rules using linked matrix, graph, and detail views. 2013.

[43] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Visual Languages, 1996. Proceedings., IEEE Symposium on*, pp. 336–343. IEEE, 1996.

[44] A. Shoshi, T. Hoppe, B. Kormeier, V. Ogultarhan, and R. Hofestädt. 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):15, 2015.

[45] D. Sridhar, S. Fakhraei, and L. Getoor. A probabilistic approach for collective similarity-based drug–drug interaction prediction. *Bioinformatics*, 32(20):3175–3182, 2016.

[46] W. P. Stephenson and M. Hauben. Data mining for signals in spontaneous reporting databases: proceed with caution. *Pharmacoepidemiology and drug safety*, 16(4):359–365, 2007.

[47] K. Techapichetvanich and A. Datta. Visar: A new technique for visualizing mined association rules. In *International Conference on Advanced Data*

Mining and Applications, pp. 88–95. Springer, 2005.

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

[49] S. van den Elzen and J. J. van Wijk. Small multiples, large singles: A new approach for visual data exploration. In *Computer Graphics Forum*, vol. 32, pp. 191–200. Wiley Online Library, 2013.

[50] J. Von Eichborn, M. S. Murgueitio, M. Dunkel, S. Koerner, P. E. Bourne, and R. Preissner. Promiscuous: a database for network-based drug-repositioning. *Nucleic acids research*, 39(suppl 1):D1060–D1066, 2011.

[51] P. C. Wong, P. Whitney, and J. Thomas. Visualizing association rules for text mining. In *Information Visualization, 1999.(Info Vis' 99) Proceedings. 1999 IEEE Symposium on*, pp. 120–123. IEEE, 1999.

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

[53] L. Zhang, Y. D. Zhang, P. Zhao, and S.-M. Huang. Predicting drug–drug interactions: an fda perspective. *The AAPS journal*, 11(2):300–306, 2009.