MEV: Visual Analytics for Medication Error Detection

Tabassum Kakar[1,2], Xiao Qin[1,2], Cory M. Tapply[1], Oliver Spring[1], Derek Murphy[1], Daniel Yun[1],
Elke A. Rundensteiner[1], Lane Harrison[1], Thang La[2], Sanjay K. Sahoo[2] and Suranjan De[2]

[1]*Department of Computer Science, Worcester Polytechnic Institute, Worcester, Massachusetts, U.S.A.*

[2]*Center for Drug Evaluation and Research, U.S. Food and Drug Administration, Silverspring, Maryland, U.S.A.*

{tkakar, xqin, cmtapply, obspring, dmurphy2, dyun, rundenst, ltharrison}@wpi.edu,
{thang.la, sanjay.sahoo, suranjan.de}@fda.hhs.gov

Abstract: To detect harmful medication errors and inform regulatory actions, the U.S. Food & Drug Administration uses the FAERS spontaneous reporting system to collect medication error reports. Drug safety analysts, however, review the submitted report narratives one by one to pinpoint critical medication errors. Based on a formative study of the review process requirements, we design an interactive visual analytics prototype called Medication Error Visual analytics (MEV), to facilitate the medication error review process. MEV visualizes distributions of the reports over multiple data attributes such as products, types of error, etc., to guide analysts towards most concerning medication errors. MEV supports interactive filtering on key data attributes that aim to help analysts hone in on the set of evidential reports. A multi-layer treemap visualizes the count and severity of the errors conveyed in the underlying reports, while the interaction between these layers aid in the analysis of the corresponding data attributes and their relationships. The results of a user study conducted with analysts at the FDA suggests that participants are able to perform the essential screening and review tasks more quickly with MEV and perceive tasks as being easier with MEV than with their existing tool set. Post-study qualitative interviews illustrates analysts' interest in the use of visual analytics for FAERS reports analysis operations, opportunities for improving the capabilities of MEV, and new directions for analyzing critical spontaneous reports at scale.

1 INTRODUCTION

A medication error is a preventable event that may cause or lead to inappropriate medication use or patient harm while the medication is in the control of the health care professional, or patient. Every year, serious preventable medication errors occur in 3.8 million inpatient admissions and 3.3 million outpatient visits with an estimated annual cost burden of $20 billion (err, 2010). A medication error involves mistakes that are caused by wrong administration or handling of drug due to ambiguity of drug label or carton. Hence, these errors are preventable and should be detected and corrected earlier to avoid further damage.

To be able to take immediate regulatory actions towards the medical products that are prone to harmful medication errors, the U.S. Food & Drug Administration (FDA) uses the Adverse Event Reporting System, *FAERS* in short, to collect medication error reports from health care professionals, consumers, and drug manufacturers.

At the FDA, the Division of Medication Errors Prevention and Analysis (DMEPA) is responsible for ensuring the safe use of medications by minimizing use errors related to the drugname, such as drugnames that sound or look similar, labeling, packaging, or design. It is their responsibility to monitor and analyze reports about medication errors submitted via FAERS to identify concerns that can be addressed through regulatory action. These actions may include revising container labels or instructions for use, communicating safety issues to the public, and in rare cases, changing a proprietary drugname.

A safety analyst may determine that a reported incident corresponds to a more general medication error concern that may potentially warrant a label change, drug withdrawal, or other similar action. Such incident report is then evaluated based on various factors including the severity, type and the cause of the error. This evaluation tends to require the analysis of many other reports over a longer period of time. In these reports, some useful information such as demographics of the affected patients are explicitly

captured in the structured fields associated with each report, while the details of the error are discussed in-depth only in the text narrative itself. Statistics about FAERS reports can also be important. For example, to understand how severe an error is, the analysts may want to know how its severity compares to that of a similar type of error within the overall set of reports.

Currently, drug safety analysts at the FDA use tools that are supported by Structured Query Language (SQL) to retrieve reports from FAERS that refer to their assigned set of products. The information about a specific error is gathered by reading through each narrative report. They may alternatively use SQL to collect basic statistics about a collection of reports, e.g., they may compute the total number of reports for a given error type in the last two weeks or the age distribution of those affected by this error (e.g., to determine if an older population is disproportionately affected).

Such mechanisms become problematic as the volume of reports grows. First, a systematic method of exploring and categorizing reports based on their content is missing. Second, information embedded in the unstructured narrative text can only be extracted manually. This manual information extraction from text is inefficient, time consuming and cognitively demanding. Third, no comprehensive representation that conveys the overall global statistics of the suspected errors or products with respect to different subsets of FAERS reports is available. Our overarching objective is to design interactive visualization and analytics techniques to address these shortcomings.

To design Medication Error Visual analytics (MEV) tool, we first characterize the current practices in medication error detection and prevention through formative interviews with drug safety analysts at the FDA. This leads us to gain an understanding of the analysts' pain points and limitations of current tools. We then utilize these insights to guide the design of MEV. The result is MEV– a visual analytics approach that aims to support the exploration and analysis of medication error reports. MEV first extracts key information about the reported incident from the respective text narrative using recently developed biomedical natural language processing techniques (Savova et al., 2010; Xu et al., 2010; Aronson and Lang, 2010; Wunnava et al., 2017). MEV then displays this information along with other attributes associated with a given report such as drugnames on the treemap visualization (Fig. 2). MEV provides several visual interactions aim to help safety analysts sift through these reports to uncover pertinent information about suspected medication errors.

MEV defines criticality scores for different types of medication errors based on the severity of the error and the count of reports reflecting that same error. This information is encoded in visual features of the visualization, such as the shape and size of the treemap components making the severe reports more quickly discernible as compared to less severe ones. A timeline view allows analysts to see the overall distribution of the reports over a period of time. Demographic displays enable visual analytics based on the structured information from FAERS reports such as age, gender and occupation. These interactive visualizations are intended to allow analysts to see faceted distributions of the patient characteristics for selected drugs or errors. Analysts can interactively choose particular attributes and analyze the resulting reports.

A user study with 10 drug safety analysts at the FDA , who were not involved in the design process of MEV, suggests that performing several common review related exploration tasks with MEV is faster and easier than their existing tool. Further, qualitative interviews show participants' enthusiasm regarding the use of visual analytics for medication error detection and highlight opportunities for future improvements.

2 RELATED WORK

We study existing techniques that align with our data type and goals. The key data elements extracted using NLP such as type and cause of an error are categorical, called facets. Facets have been widely used as interactive filters for searching and browsing data. FacetMap (Smith et al., 2006) supports interactive visualizations to explore facets of a dataset, however, it does not support discovering relationships among facets. FacetLens (Lee et al., 2009) extends FacetMap to help users observe trends and explore relationships within faceted datasets. Most of these faceted systems (Lee et al., 2009; Smith et al., 2006) divide their interfaces between a main viewing area and a secondary facet area which allows to browse only one data item at a time. For medication error screening, however, it is crucial to see the effect of selection of one item on others, so that data points representing concerning errors can be identified quickly.

Treemaps (Asahi et al., 2003) have been widely used in visualization systems (Liu et al., 2009; Harrison et al., 2012). For example, SellTrend (Liu et al., 2009), a visualization tool

for displaying temporal categorical data, displays transaction failures using treemaps. NV (Harrison et al., 2012) utilizes treemaps and histograms to allow security analysts to discover, analyze, and manage vulnerabilities on their networks. However, these tools do not have support for extracting name entities from textual data, neither do they visualize temporal patterns and demographics within the data. JigSaw (Stasko et al., 2008), on the other hand, is a powerful tool for investigating text data by visualizing name entities and their relationships to reveal hidden plots in criminal reports. However, there is a need to support temporal data analysis for reports screening and review.

In the medical domain there has been work on designing systems to avoid medication errors from arising in the first place, such as medication-reconciliation tools (Ozturk et al., 2014) and clinical information systems (Jia et al., 2016). Varkey et al. (Varkey et al., 2007) study the effect of interventions on decreasing medication errors related to the administration of drugs. A patient's one year long prescription history is visualized using timeline charts to be used by clinicians and the emergency room staff (Ozturk et al., 2014). Other tools are designed as interfaces to provide a user-friendly mean of error reporting (Singh et al., 2008). Clinical decision support systems have been proven to reduce medication errors during prescription (Jia et al., 2016). However, these tools are designed with the goal of reducing medication errors from happening in the first place during the prescription or the administration of the drugs.

Our work instead starts after the medication errors have already occurred and have been reported to the concerned authorities such as the FDA. For example, if two drugs have look-alike carton labels for different dosages and FDA receives error reports about these dosages being prescribed interchangeably. Then FDA drug safety analysts after careful examination of such reports can recommend to change the product carton label so that different products or dosages can be differentiated easily. This prevents such errors from happening in the future. To the best of our knowledge, no visual analytics tool exists that can be used to help analysts explore medication error reports.

3 REQUIREMENT ANALYSIS

Before designing a system for medication error analysis, we conducted formative interviews with the FDA drug safety analysts to understand their data, current workflow, exiting tools for reports review and their limitations and challenges.

3.1 Interviews with Domain Experts

We organized a series of semi-formal interview sessions with five drug safety analysts at the Division of Medication Error Prevention and Analysis (DMEPA) at the FDA. Our primary objective was to understand the current report review process and to identify the challenges these analysts face in analyzing medication error reports. From these interviews, we learned that certain information was critical to their workflow. We also observed the limitations of current tools.

To develop and refine the design of the specific visualizations MEV uses, we showed these analysts sketches of design alternatives, such as parallel coordinates and variations of node-link diagrams. This activity helped us gather additional design requirements, such as readability of the visualization. In subsequent interviews, we presented these analysts with a working prototype of MEV to evaluate their perceptions of the degree to which MEV meets their needs, and to receive further feedback on the visual and interaction design. In the final session, a larger group of analysts (ten), who were not involved in the design process of MEV, participated in the user study to evaluate MEV and provide additional insights on the utility of MEV.

3.2 FAERS Data Description

We briefly describe the data reviewed by FDA analysts based on our initial discussions with the domain experts. FDA maintains an Adverse Event Reporting System (FAERS) (FDA, 1995) as a part of its post-marketing drug surveillance program for medications and therapeutic biologic products. Reports submitted to FAERS include mandatory reports submitted by drug manufacturers and voluntary reports submitted by health care professionals and consumers. These reports are semi-structured in nature, that is, they contain structured information about patient demographics, drugs taken, therapies, and adverse reactions or medication errors. They also contain an unstructured textual narrative that describes the incidents associated with medication errors or adverse reactions in detail and contains richer information such as the details of the incident to help analysts decide if the incident is worthy of investigation. Majority of the key information used in the analysis is categorical, with drugs having the highest number of categories (50-100) per analyst.

Safety analysts review reports based on the classes of products assigned to them. On average a safety analyst can have 50-100 medical products and reviews 200 reports on average on a weekly basis. However, for detail analysis of a product thousands of reports are retrieved for several months. These numbers vary from team to team based on the assigned products. For example, a new approved product might be causing more medication errors as compared to an old product that has been in market for a long time and people are familiar with its proper usage.
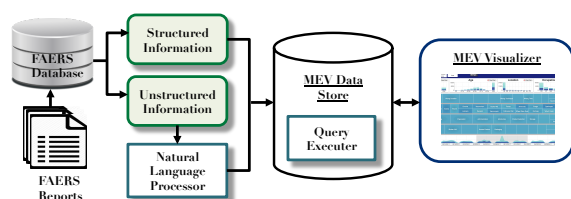
Figure 1: The MEV framework.

3.3 Domain Experts Workflow

At the FDA, drug safety analysts review FAERS reports daily for potential severe medication errors. The analysts receive reports related to the drugs assigned to them. The first task is to screen the reports for potential signals based on severe outcomes such as hospitalization or death, age groups, products and error types. Then the report narrative of the prioritized errors are analyzed. Once any alarming error is found, then other sources are analyzed to investigate a particular error.

3.4 Limitations of Current Tools – MEV Design Rationale

Currently, existing tools at the FDA are mostly SQL based. That is, they allow finding reports related to an age group or a product. However, our interviews revealed that the tool lacks the ability to interactively guide the user towards most critical errors and ultimately help them in forming hypothesis about a potential dangerous medication error. That is, by using the current tools it is not trivial to analyze data distributions and relationships among core attributes. Particularly, it is cumbersome to ask questions such as *which errors are associated with a particular product*, or *which age group is more prevalent in the reports associated with a particular error?* Moreover, questions such as *what stage a particular error is happening at?* or *what are most reported root causes of error?* are not supported by the tool at all.

The reason is that first, the key information such as stages and causes of error must be learned from the text narrative, as this information is not stored in the structured fields associated with a given report. Second, these existing tools are only designed to help analysts filter a set of reports using the structured information, which can be analyzed further using Microsoft Excel spreadsheets.

Therefore, an automatic way is needed to first extract this information from text narratives and then to allow analysts to interactively query this information along with other structured information to make the report review process efficient. Our MEV system is designed to assist analysts in this signal screening phase by supporting the analysis of data distributions to help in hypothesis formation about critical errors.

3.5 The MEV Framework

Following the workflow of domain analysts, MEV depicted in Fig. 1 is designed to explore the reports efficiently. As described earlier, FAERS reports contain both structured as well unstructured text narrative explaining the event in detail. In case of medication errors, the core information related to the type or cause of the errors is not captured in the structured parts of the report. Instead, it tends to be mostly mentioned within the text narrative. To support analysts in finding important information concerning medication errors quickly, we use rule-based name-entity recognition techniques (Wunnava et al., 2017) to extract key information from the text narrative.

We use domain specific lexicons (NCC-MERP, 1995; Brown et al., 1999) to extract key data attributes. These attributes include types of medication errors (e.g., taking a wrong drug or dosage), the root causes of the errors (e.g., name confusion and container label confusion), and the stage in which error has occurred (e.g., dispensing and administration). The *Natural Language Processor* (Fig. 1) after preprocessing the text, such as stemming and tokenizing, extracts these core data elements. This extracted information is then standardized by mapping it to NCC-MERP terms using *edit distance* based string matching (Du, 2005) for smooth exploration and analysis. Currently, analysts manually summarize each narrative by adding these terminologies into the Excel spreadsheet. After standardization, on average each of the extracted entity contain approximately 15-20 categories.

The extracted information along with structured information about demographics is stored in the

Figure 2: The user interface of MEV (a) The demographics panel. (b) The treemap panel. (c) The timeline panel. (d) Reports icon to access the reports view to analyze the associated report narratives.

MEV *Data Store* (Fig. 1). The MEV *Query Executor* handles processes requests on the data store specified through online MEV visual interface. Results from frequent interactions are cached to improve user experience. The MEV assists analysts in exploring the data interactively using linked interactive visualizations described below.

4 MEV INTERFACE OVERVIEW

Our MEV tool consists of four main interactive displays (Fig. 2), the treemap view, the demographics panel, the timeline panel and the reports view.

4.1 The Treemap Panel

A treemap visualization (Fig. 2b) displays the distribution of each of the multi-value categorical attributes extracted from structured data as well as unstructured text. These attributes include drugname, the root cause of the error, the stage where the error has occurred, and the error type. Each of these attributes have multiple values. In each treemap, each rectangle represents a data value within an attribute, e.g., for the product treemap, each rectangle represents a drugname. The size of each rectangle is mapped to the count of reports related to that specific data value, while the color depicts the count of severe outcomes which is a structured data field.

This treemap design allows analysts to

interactively filter even large number of items, such as a large number of drugnames can be visualized in a compact way (Liu et al., 2009). The analyst can select one or multiple data values on each treemap and the system will immediately show what other data attributes correspond to a selected value. This direct manipulation of data allows the analysts to narrow down their search based on the items distributions that need the most attention, which may be achieved through multiple tidy steps using their current tools.

Although, treemaps are often used to visualize hierarchical data, here we leverage the capability of displaying categorical data as well as showing many values though space filling techniques. Another advantage of treemaps is their ability to effectively make use of both size and color for encoding additional properties about each categorical choice. While alternate multi-dimensional visualization techniques, such as parallel coordinates or scatter plot matrices are possible, for scalability and avoiding visual clutter, treemaps are used to guide analysts in the screening of their assigned reports. Treemaps are one possible design, but other design choices including bar-charts or lists (Stasko et al., 2008) having similar functionality may have desirable properties.

4.2 The Timeline Panel

The timeline panel (Fig. 2c) displays the overall report distribution as well as their severity over a

Figure 3: Reports view with no personal information (Left). Example of a de-identified text narrative (Right).

period of time using a temporal area chart. This allows us to detect a spike in the severity associated in the incidence of certain products. Interactive brushing and selection through zooming is provided to allow the safety analysts to drill into a particular date range and explore the associated reports. Once a date range is selected, other displays are updated to reflect only data from the selected date range.

4.3 The Demographic Panel

The demographics of patients also play an important role in the analysis of the reports. For instance, for a particular drug there might be many more severe outcomes in a particular age group than in the other groups. The graphs in the demographics panel (Fig. 2a) assist the analysts in selecting reports related to a particular demographic attribute, such as, location, gender or age group. Drug safety analysts can not only prioritize reports based on these attributes to hone in on respective reported medication errors, but they can also upon selecting any data value immediately view the distribution of reports for each demographic view through linked displays.

4.4 The Reports View

After safety analysts select a particular product or medication error of interest, they can view the respective reports and investigate them further to find if the reports indeed are indicative of errors with serious consequences for patient health warranting

regulatory action. For this, by clicking on the reports icon (Fig. 2d) the selected reports are accessible. The reports view displays the line listing of the screened data elements (Fig. 3-left). Analysts can drill into the narrative of each report to further examine the report in great detail (Fig. 3-right).

4.5 System Implementation

MEV is a web based tool developed using React and JavaScript for front-end and PostgreSQL for the back-end database. The tool also leverages a cache (Redis) for efficient data retrieval and to improve user experience. The extracted data elements are stored in the database along-with other structured information.

5 EVALUATION

5.1 Pharmacovigilance Usage Scenario

To understand how MEV can help with reports screening, we now discuss a use case of Alex, a drug safety analyst, reviewing reports related to her assigned products using MEV. From the timeline panel, she sees an overall weekly distribution of the number of new reports received over this last month. At a glance, she can see in Fig. 2 that no reports have been submitted over the weekend, while new reports have been received during weekdays. She explains that the FDA does not populate any reports into the database during the weekend. She notices a
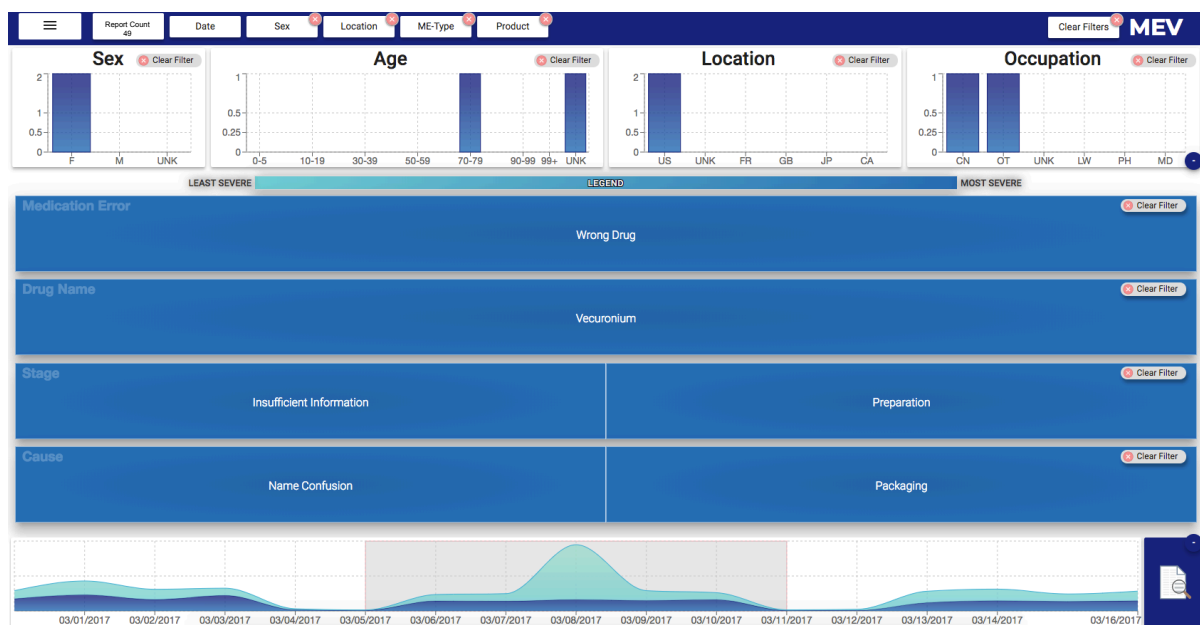
Figure 4: MEV with selection of date range, demographics, drugname and medication error type.

spike in the number of reports between "3/5/2017 - 3/11/2017" of which 39% of reports are severe and 61% are non-severe reports (Fig. 4-bottom). She decides to investigate reports by selecting this week using the brush tool on the timeline panel. She observes that the number of reports for this one week are 28,123. The demographics and treemap charts both are updated for the selected date. She notices there are more female patients than male and the age group is mostly between "30-80" years old. That is expected, as her assigned products are mostly for elderly women. From the demographics, she selects females with a location in the U.S. to see the reported drugs and errors. This reduced the target set to 11,174 reports.

On the treemap, she now notices that the medication error "wrong-technique" has most of the count with severe outcomes. She questions which products are administered with this "wrong-technique" error. Alex thus selects wrong-technique in the first treemap by clicking on the rectangle labeled 'wrong-technique'. This reduces the reports count to 2,786 reports. She observes the reported drug Lotensin has the highest number of severe reports. She selects Lotensin from drugnames on the second treemap. Now she wants to know what causes this "wrong-technique" error in Lotensin products and at which stage these errors arise. Looking at 3rd and 4th treemaps corresponding to the cause and stage of errors respectively, she notices that most have causes such as "name confusion" and "packaging". She adds "It seems an

error in preparation of the drug". She also observes that a total of 49 reports remain that she needs to analyze in detail (Fig. 4). She speculates whether these reports indeed have compelling evidence about these errors. She clicks on the reports icon (Fig. 2d) to read the details of each narrative in the reports view (Fig. 3). Hence, MEV interactively guides the analyst towards concerning errors by supporting exploration and screening of reports.

5.2 User Study

5.2.1 Study Design

We invited eleven drug safety evaluators (10 females, 1 male) at the Division of Medication Error and Prevention Analysis (DMEPA) at the FDA for a one hour in-person study session. One of the participants withdrew participation. These participants were within the age range of 30-50 years with the majority having experience with basic visualizations. These participants were pharmacists, conducting regular report reviews to identify any medication error that would need regulatory action.

Assessment Measures. We specified a set of nine tasks (Table. 1) commonly performed during the report review process to evaluate the usefulness of MEV. These tasks were derived from the initial interviews conducted with the users to understand the review workflow. These tasks varied from a one-step task of finding a particular attribute value

Table 1: List of 9 Tasks designed to evaluate the effectiveness of MEV.

Task #	Description
T1	How many total reports have been reported during a time period?
T2	Which medication error is reported the most for a time period?
T3	Which drug has most severe outcomes for a selected medication error?
T4	Which gender and age have most severe outcomes?
T5	Which age group is most prevalent in reports related to a selected product?
T6	What are the two most frequent medication errors reported with a select product, age group, and gender?
T7	Given the report distribution of a drug for female patients with a specified age group, what are the critical medication errors that need to be analyzed?
T8	What are the two most frequent root causes of error for a selected drug and medication error?
T9	What are the two most common reported stages of errors for a drug and a medication error?

(T1-T2) to two-step tasks of finding reports associated with analysis of two attributes (T3-T5). We included multi-step tasks of finding interesting reports to be prioritized based on the distribution of multiple data attributes (T6-T9). These composite tasks involved filtering based on examination of relationships among data attributes. We considered two metrics, one, time to successfully complete each task and two, how easy the participants rated each task. Data loading in their existing tool takes longer time, so the task completion time was recorded **after** FAERS data for one week (from 2017) was loaded in both tools. The perceived ease from each task was recorded on a 5-point Likert scale (5 extremely easy and 1 extremely difficult). We reported the time taken by each participant to successfully accomplish each task. Participants were asked to perform the same set of tasks with their existing tool (*Control*) as well as MEV to compare both tools.

Study Procedure. To get detailed feedback from the participants and observe them closely interacting with the system, the study was conducted via a one hour in-person interview session. Upon successful completion of the demonstration and training session (20 minutes), the participants were asked to perform the set of prescribed tasks (Table. 1) using the FDA adverse event reporting (FAERS) data from 2017 using both the MEV tool as well as their existing tool. At the end of each session, participants were provided with a post-study questionnaire, which was not timed. The first section of the questionnaire contained questions related to the demographics of the participant such as age and gender. The second part had questions about the usability (Brooke et al., 1996) of MEV on a 5-point Likert scale (5 strongly agree & 1 strongly disagree). Finally, an open-ended questionnaire was offered to solicit qualitative feedback about MEV.

Analysis. For some tasks, the time and perceived ease score collected from the study were not normally distributed. Hence, to find out whether performing the prescribed tasks is quicker and easier with MEV than the existing tools, we performed the non-parametric Mann-Whitney U Test (Wilcoxon Rank Sum Test) to compare conditions. We also report the 95% confidence intervals for both time as well as perceived ease score for all tasks.

5.2.2 Study Results

We now analyze the participants' performance on the tasks and their response about the overall system usability.

Quantitative Analysis: From Table. 2, we see that for majority of the tasks, there are significant differences between the recorded time and perceived ease score for completing them using our proposed system and their existing tool *control* with the exception of T1. T1 was a one-step task involving finding the total number of reports for a given duration of time. One possible explanation for this difference is that participants were used to their current tool and knew exactly where they will find this information. On the other hand, being new to MEV tool they took little longer (M=5.11 seconds [3.47, 8.76]) as compared to their current tool (M=3.62 [1.80, 5.44]). This task was also scored easier under *control* condition than using MEV. Neither time nor perceived ease were significantly different for T1. T2 involved finding the most reported medication errors for a selected time period. There was significant difference between the performance under *control* condition (M = 7.54 seconds [3.57, 11.52]) and using MEV (M =31.84 [15.78, 47.91]). In addition to time, participants also found it easy to perform the task using MEV (M = 4.9 [4.70, 5.10]) than under control

Table 2: U-Test for both time and perceived ease.

Tasks	Siginificance Test Time (α = 0.05)	Siginificance Test Easiness (α = 0.05)
T1	(U = 29.5,p = 0.13104)	(U = 45, p = 0.72786)
T2	(U = 14, p = 0.00736)	(U = 14, p = 0.00736)
T3	(U = 0, p = 0.00018)	(U = 1.5, p = 0.00028)
T4	(U = 0, p = 0.00018)	(U = 7.5, p = 0.00152)
T5	(U = 4, p = 0.00058)	(U = 1 , p = 0.00024)
T6	(U = 4, p = 0.00058)	(U = 7, p = 0.00132)
T7	(U = 11, p = 0.00362)	(U = 0, p = 0.00018)
T8	Not Applicable	Not Applicable
T9	Not Applicable	Not Applicable

condition (M=4.0 [3.59, 4.41]).

For the multi-step tasks (T3-T7), that involved retrieving data based on analyzing distribution and severity across multiple attributes, both time and perceived ease have significant differences (Table. 2). Tasks T8 and T9 involved composite filtering to retrieve the root causes and stages of errors related to severe outcomes. As these data entities were extracted using NLP and their current tools do not provide them, the comparison was not possible. Additionally, from Fig. 5 (Left), we see that participant's performance is relatively consistent/stable for all tasks, that is, all participants were able to quickly perform the tasks using MEV. On the other hand, participants had highly varied performance for tasks (T2-T7) using the existing tool.

Similarly, for perceived ease, Fig. 5 (Right) depicts that participants perceived it easier to perform tasks (T2-T9) using MEV than the existing tool. T5 was rated the most difficult to perform under *control* condition, as it involved analysis of distribution of age for a selected product. Exploring the distribution of data attributes with their existing tool is tedious as it requires filtering for each attribute value individually and then analyzing the outcome.

Lastly, we aggregated the responses from all participants on the system usability (SUS) questionnaire (Brooke et al., 1996). MEV received an SUS score of 85 out of 100.

Qualitative Analysis & Overall Impression of MEV: The focus of qualitative questionnaires was on the participants' subjective impression of the tool and their experience using it. Our analysis of comments on the questionnaire suggests that the participants' experiences with the tool differed depending on their prior experience with similar interactive visualizations. For instance, some participants found the timeline visualization difficult to interact with, while others liked it.

Overall, the majority of participants agreed with the general premise of the tool, and found its goal of analyzing drug-related medication errors with severe outcomes and promoting individuals' ability to explore data to be promising and potentially useful. According to the study participant P10: "Well, I think this tool makes it very easy to see what the reports are describing without going into much detail". 6 out of 10 participants explicitly mentioned the usefulness of integrating name-entities into the visualization and the intuitiveness of the tool itself. Participant P2 mentioned: "Though the text-extraction is not perfect but it gives us a big sense of what kind of errors are being reported". Participant P5 said: "It takes sometime to get used to the tool, then it is very easy and intuitive to use".

Constructive feedback for potential improvements of the design of the tool were also solicited using an open response option. For instance, four participants suggested that an individual search option on each treemap for looking up a particular drug or error would be useful to achieve the presented tasks.

6 DISCUSSION

The aim of this work centers on developing visualization-enabled systems that support domain experts in pharmacovigilance. Our results indicate that users can in fact perform review tasks in pharmacovigilance data by analyzing the distribution of various data attributes using the provided views, and conduct investigative tasks from within MEV. More broadly, additional challenges and opportunities in the space of human-in-the-loop systems for medical professionals have been uncovered through interaction with drug analysts.

One key issue in modern systems is scale. As the goal of MEV is to be used by each drug safety analyst for reports screening of their assigned set of products on a weekly basis that constitutes a count of thousands of reports. We tested MEV with data from one year (2017) which constitutes over 1.82 million reports, where it takes several seconds to load data and transform it for the initial overview. Other challenges of scale relate to the visualizations themselves. If the analyst were to steer to a view with hundreds or more drugs, the treemap may display only tiny rectangles, a source of visual clutter (Peng et al., 2004). One solution to this clutter problem is to display a subset of drugs on the treemap along with a search option to access a desired drugname. Adding a layer of drug classes on the treemap can
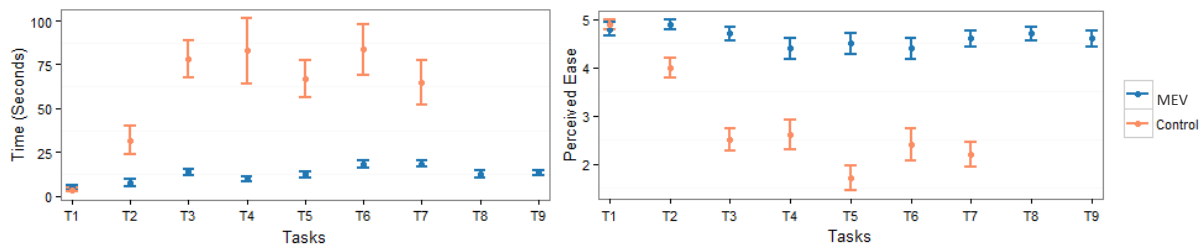
Figure 5: 95% Confidence Interval for performing tasks using both MEV and existing tool (*Control*). (Right): Perceived Ease Score. (Left): Time (in sec). Task 8 & 9 are not supported by *Control*.

be another alternative to address the scaling issue. Analyst can select a drug class and the drugs under that class can be visualized on the treemap. We could also incorporate domain practices into the system. For example, the maximum number of distinct products in the reports for each user does not exceed 100, so clutter is not a problem for typical use cases of MEV.

During our qualitative interviews while majority of the analysts acknowledged MEV's usefulness in reports screening, few analysts mentioned that they would prefer to read each and every report narrative rather than using MEV for screening, if the number of reports is few, i.e., ten or twenty. For such users, a feature of highlighting the key information within the report narratives can be added. During our user study, we also noticed that the extracted information were incorrect, when users fetched the reports to analyze the narratives. We leveraged the MEFA (Wunnava et al., 2017) name-entity extractor in this work for extracting information such as the stage and cause of the error. More advanced extraction techniques using deep learning (Jagannatha and Yu, 2016) could be plugged into MEV to improve the entity extraction accuracy. However, name-entity extraction in the medical domain itself is known to be a challenging problem and research efforts towards more accurate techniques continue.

Our user study has a number of limitations. First, participants are familiar with their existing tool; this familiarity allowed participants to complete some complicated tasks in a short time using their existing tool. Also, for a few participants some tasks were deemed as not relevant. For instance, participants who usually investigate one particular drug found it irrelevant to look for reports related to multiple drugs based on severity of reports. Study participants, while a small number, are real drug safety analysts who would be ultimately users in every day analysis. Long term studies with these analysts would help to further assess MEV in their task flow.

There are a few possible directions to work on in future. First, we plan to integrate interactive support for report text analysis into MEV to support the full workflow of the analysts. Second, direct access to external sources such as PubMed and DailyMed from within MEV so that analysts can confirm or reject a hypothesis about a possible medication error formed using the treemap by investigating these sources would simplify the analysis. Third, visual provenance (Groth and Streefkerk, 2006) would also add value by allowing analysts to share their thought-processes and findings with their team members.

7 CONCLUSION

In this paper, we introduce MEV – a prototype tool for visual analytics of medication errors from spontaneous reporting databases. MEV assists analysts in exploring and screening spontaneous reports via an interactive treemap, interactive bar charts showing demographics and a timeline visualization. Analysts can pinpoint severe reports visually and compare data distributions across many weeks of data. Results from a task-based user study with 10 drug safety analysts at the FDA suggest that performing review tasks using MEV is both efficient and perceived easier than their current tool. Study results also suggest that analysts find MEV intuitive and easy to interact with and that it would likely align with the existing workflow of medication error reports analysis. Lastly, qualitative interviews suggested opportunities for improvements in the current design.

REFERENCES

(2010). *A $21 Billion Opportunity National Priorities Partnership convened by the National Quality Forum file.* http://www.nehi.net/bendthecurve/sup/documents/ Medication_Errors_Brief.pdf [Accessed: 2018-01-07].

Aronson, A. R. and Lang, ç.-M. (2010). An overview of metamap: historical perspective and recent advances. *Journal of the American Medical Informatics Association*, 17(3):229–236.

Asahi, T., Turo, D., and Shneiderman, B. (2003). Visual decision-making: using treemaps for the analytic

hierarchy process. In *The Craft of Information Visualization*, pages 235–236. Elsevier.

Brooke, J. et al. (1996). Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7.

Brown, E. G., Wood, L., and Wood, S. (1999). The medical dictionary for regulatory activities (meddra). *Drug safety*, 20(2):109–117.

Du, M. (2005). Approximate name matching. *NADA, Numerisk Analys och Datalogi, KTH, Kungliga Tekniska Högskolan. Stockholm: un*, pages 3–15.

FDA (1995). *FDA Adverse Event Reporting*. http://www.fda.gov/Drugs/GuidanceComplianceRegulatory Information/Surveillance/AdverseDrugEffects/ucm070434.htm [Accessed: 2018-01-10].

Groth, D. P. and Streefkerk, K. (2006). Provenance and annotation for visual exploration systems. *IEEE Transactions on visualization and Computer Graphics*, 12(6):1500–1510.

Harrison, L., Spahn, R., Iannacone, M., Downing, E., and Goodall, J. R. (2012). Nv: Nessus vulnerability visualization for the web. In *Proceedings of the ninth International Symposium on Visualization for Cyber Security*, pages 25–32. ACM.

Jagannatha, A. N. and Yu, H. (2016). Bidirectional rnn for medical event detection in electronic health records. In *Proceedings of the conference. Association for Computational Linguistics. North American Chapter. Meeting*, volume 2016, page 473.

Jia, P., Zhang, L., Chen, J., Zhao, P., and Zhang, M. (2016). The effects of clinical decision support systems on medication safety: an overview. *Public Library of Science One*, 11(12):e0167683.

Lee, B., Smith, G., Robertson, G. G., et al. (2009). Facetlens: exposing trends and relationships to support sensemaking within faceted datasets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1293–1302. ACM.

Liu, Z., Stasko, J., and Sullivan, T. (2009). Selltrend: Inter-attribute visual analysis of temporal transaction data. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1025–1032.

NCC-MERP (1995). *National Coordinating Council for Medication Error Reporting and Prevention*. http://www.nccmerp.org/ [Accessed: 2018-02-28].

Ozturk, S., Kayaalp, M., and McDonald, C. J. (2014). Visualization of patient prescription history data in emergency care. In *AMIA Annual Symposium Proceedings*, volume 2014, page 963. American Medical Informatics Association.

Peng, W., Ward, M. O., and Rundensteiner, E. A. (2004). Clutter reduction in multi-dimensional data visualization using dimension reordering. In *Information Visualization, 2004. INFOVIS 2004. IEEE Symposium on*, pages 89–96. IEEE.

Savova, G. K., Masanz, J. J., Ogren, P. V., et al. (2010). Mayo clinical text analysis and knowledge extraction system (ctakes): architecture, component evaluation and applications. *Journal of the American Medical Informatics Association*, 17(5):507–513.

Singh, R., Pace, W., Singh, A., et al. (2008). A visual computer interface concept for making error reporting useful at care point.

Smith, G., Czerwinski, M., Meyers, B. R., et al. (2006). Facetmap: A scalable search and browse visualization. *IEEE Transactions on Visualization and Computer Graphics*, 12(5).

Stasko, J., Görg, C., and Liu, Z. (2008). Jigsaw: supporting investigative analysis through interactive visualization. *Information Visualization*, 7(2):118–132.

Varkey, P., Cunningham, J., and Bisping, S. (2007). Improving medication reconciliation in the outpatient setting. *Joint Commission Journal on Quality and Patient Safety*, 33(5):286–292.

Wunnava, S., Qin, X., Kakar, T., et al. (2017). Towards transforming fda adverse event narratives into actionable structured data for improved pharmacovigilance. In *Proceedings of the Symposium on Applied Computing*, pages 777–782. ACM.

Xu, H., Stenner, S. P., Doan, S., et al. (2010). Medex: a medication information extraction system for clinical narratives. *Journal of the American Medical Informatics Association*, 17(1):19–24.