Visual Analytics for Multi-Level Triage and Investigation of Incident Reports

by

Tabassum Kakar

A Dissertation

Submitted to the Faculty of
WORCESTER POLYTECHNIC INSTITUTE

in partial fulfillment of the requirements for the Degree of
Doctor of Philosophy
in
Data Science

February 2020

Approved:

Dr. Elke A. Rundensteiner, Professor, WPI. Co-Advisor.

Dr. Lane Harrison, Assistant Professor, WPI. Co-Advisor.

Dr. Eleanor Loiacono, Professor, WPI.

Dr. Hendrik Strobelt, IBM Research, External Member.

Abstract

Regulatory agencies in domains from health-care, finance, and aviation receive a large number of incident reports related to the products or services being overseen. Analysts at these agencies must conduct triage and investigation tasks to uncover critical incidents of concern and take appropriate action to stop such incidents from happening in the future. The number and complexity of these incident reports pose challenges for the analysts who need to make critical yet timely decisions by combing through these reports. Yet computational techniques to analyze these reports alone are not sufficient for the effective triage and investigation of incident reports due to requiring human judgment and thus involvement in the process.

In this dissertation, we design visual analytics techniques based on the analysts' workflows to augment their capabilities to examine incident reports. We approach this goal through three primary research tasks: (1) Exploratory triage of incidents at an overview level by leveraging machine-learning and Natural Language Processing (NLP) generated meta-data to help in hypothesis formation about potential incidents of concern. Specifically, we design interactive visualizations guided by domain-workflows that help analysts explore, screen, and prioritize critical incidents. We evaluate these systems via user studies and case studies. (2) Confirmatory triage of individual reports associated with a screened incident to decide if it warrants further investigation. We design a domain-informed glanceable visual summary that assists analysts in getting the gist of all the information crucial for decision making. We conduct a user study with 20 domain experts to evaluate the effectiveness of our visual summary on triage performance and experience as compared to using a tabular baseline used by the FDA. (3) Investigatory analysis of incident reports to build cases supporting particular concerning incidents to take regulatory action. We design user-driven visualizations and interactions that support analysts in the evidence identification, collection, and management for alarming incidents. We evaluate these designs via case studies and user interviews with the domain experts from the FDA.

Successful results in this research may have a significant impact on the current practices of incident analysis in regulatory agencies, particularly Pharmacovigilance activities performed in the majority of western countries. Our research forms the basis for designing visual analytics systems that facilitate the seamless analysis of the huge influx of incoming critical incident reports requiring immediate attention. We discuss the implications of using visual analytics for triage in critical workflows. This lays the groundwork for further research targeting transparency and uncertainty in text-based visual designs for triage.

Acknowledgements
I express my deepest gratitude to my co-advisor Prof. Elke Rundensteiner, without her support and guidance this work would not have been possible. Thank you for all the feedback and late night edits on my manuscripts. I am also grateful to my co-advisor Prof. Lane Harrison, whose expertise and valuable insights in HCI and visual analytics have helped me in improving my work and communication skills. I have learned a lot and I will always be thankful to you for helping me discover my passion for data visualization. I am lucky to have you both as my mentors and the lessons I learned in the past couple of years from both of you helped me develop personally and professionally and are going to help me throughout my life.

Special thanks to the members of my Ph.D. committee, Prof. Eleanor Loiacono and Dr. Hendrik Strobelt from IBM Watson. I am also thankful to my research collaborator Dr. Xiao Qin for his valuable discussions to improve my research strategy. I am particularly grateful to Dr. Susmitha Wunnava, my collaborator and friend, who has been listening ears, my confidante, and provided valuable insights that made a huge impact on my research journey. Thanks to my other collaborators, former undergraduate students Cory Tapply, Derek Murphy, Oliver Spring, Daniel Yun, Andrew Schade, Brian McCarthy, Huy Quoc Tran, Brian Zylich for help with the implementation and all other previous and current DSRG members for their useful discussion and feedback.

I am grateful to the U.S. Department of State for granting me Fulbright scholarship, ORAU for granting me the ORISE fellowship to conduct research with the FDA, and WPI for additional funding. I am thankful to my FDA mentors, Suranjan De, Sanjay Sahoo, and Thang La for assisting me in understanding current reports review challenges at the FDA. I am also thankful to all drug safety staff at the FDA particularly, Jo Wyeth, Monica Munoz, Christian Cao, Oanh Dang, Miriam Chehab, Debra Ryan, Samantha Cotter, Madhuri Patel, Tingting Gao, Cathy A. Miller, Sarah Kang, Lopa Thambi, Karen Long, Charlene Flowers, Jessica Weintraub, Laurelle Cascio, Wana Manitpisitkul, and others for the discussions and user studies.

Last but not least, I am thankful to my husband, family, and friends for believing in me and supporting me both practically and emotionally during this wonderful life-changing journey.

I could not have done this without you all and for that I am always going to be indebted to you.

Publications

Task 1 - Exploratory Triage and Screening of Incident Reports

1. Tabassum Kakar, Xiao Qin, Cory Tapply, Derek Murphy, Daniel Yun, Oliver Spring, Elke A. Rundensteiner, Lane Harrison, Thang La, Sanjay K. Sahoo, Suranjan De. *MEV: Visual Analytics for Medication Error Detection*. International conference on Information Visualization Theory and Application (IVAPP), 2019.

2. Tabassum Kakar, Xiao Qin, Cory Tapply, Derek Murphy, Daniel Yun, Oliver Spring, Elke A. Rundensteiner, Lane Harrison, Thang La, Sanjay K. Sahoo, Suranjan De. *Designing a Visual Analytics System for Medication Error Screening and Detection*. International Joint Conference on Computer Vision, Imaging and Computer Graphics. Springer, Cham, 2019.

   Relationship to this dissertation. This work demonstrates the detailed design and evaluation of MEV, a visual analytics tool that summarizes incident reports, discussed in Chapter 3.

3. Tabassum Kakar, Xiao Qin, Elke A. Rundensteiner, Lane Harrison, Sanjay K. Sahoo, and Suranjan De. *DIVA: Towards Validation of Hypothesized Drug-Drug Interactions via Visual Analysis*. EuroVis, 2019.

   Relationship to this dissertation. This work discusses the design study of DIVA, a multiple-coordinated views based tool to explore machine-generated hypothesized incidents, summarized in Chapter 4.

4. Tabassum Kakar, Xiao Qin, Susmitha Wunnava, Brian McCarthy, Andrew Schade, Huy Quoc Tran, Brian Zylich, Elke A. Rundensteiner, Lane Harrison, Sanjay K. Sahoo, Suranjan De. *DEVES: Interactive Signal Analytics for Drug Safety*. Conference on Information and Knowledge Management (CIKM), 2018,

5. Xiao Qin, Tabassum Kakar, Susmitha Wunnava, Brian McCarthy, Andrew Schade, Huy Quoc Tran, Brian Zylich, Elke A. Rundensteiner, Lane Harrison, Sanjay K. Sahoo, Suranjan De. *Mediar: multi-drug adverse reactions analytics*. IEEE 34th international conference on data engineering (ICDE), pp. 1565-1568, 2018.

   Relationship to this dissertation. Short papers describing the challenges and effectiveness of domain knowledge integration for the exploration and sense-making of hypothesized incidents generated by machine learning.

Task 2 - Confirmatory Triage of Incident Reports

6. Tabassum Kakar, Xiao Qin, Elke A. Rundensteiner, Thang La, Sanjay K. Sahoo, Suranjan De, Lane Harrison. *Towards Understanding Incident Reports Analysis at the FDA*. In submission to INTERACT'21.
7. Tabassum Kakar, Xiao Qin, Elke A. Rundensteiner, Thang La, Sanjay K. Sahoo, Suranjan De, Lane Harrison. *Design and Evaluation of a Glanceable Visual Summary for Drug Incident Reports Triage*. Under review by SIGCHI Conference on Human Factors in Computing Systems (CHI), 2021.

   Relationship to this dissertation. In this work, we design and evaluate a glanceable visual summary that assists analysts in quick identification of information crucial for reports triage. This work is discussed in Chapter 5.

Task 3 - Investigatory Analysis of Incident Reports

8. Tabassum Kakar, Xiao Qin, Elke A. Rundensteiner, Thang La, Sanjay K. Sahoo, Suranjan De, Lane Harrison. *ConText: Supporting the Pursuit and Management of Evidence in Text-based Reporting Systems*. Under review by SIGCHI Conference on Human Factors in Computing Systems (CHI), 2021.

   Relationship to this dissertation. In this work, we design an analytics prototype that supports the investigatory tasks of evidence identification, collection, and management in a unified manner by close coupling of information foraging and synthesis. This work is presented in Chapter 6.

Other Publications

Other research I have undertaken during my PhD at WPI include the following publications on the topic of machine learning and natural language processing for adverse event reports and electronic health records.

9. Xiao Qin, Tabassum Kakar, Susmitha Wunnava, Elke Rundensteiner, and Lai Cao. *Maras: Signaling multi-drug adverse reactions*. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2017 (pp. 1615-1623).
10. Susmitha Wunnava, Xiao Qin, Tabassum Kakar, Vimig Socrates, Amber Wallace, and Elke Rundensteiner. *Towards Transforming FDA Adverse Event Narratives into Actionable Structured Data for Improved Pharmacovigilance*. Proceedings of the Symposium on Applied Computing. 2017.
11. Susmitha Wunnava, Xiao Qin, Tabassum Kakar, M. L. Tlachac, Xiangnan Kong, Elke A. Rundensteiner, Sanjay K. Sahoo, and Suranjan De. *Multi-layered Learning for Information Extraction from Adverse Drug Event Narratives*. In International Joint Conference on Biomedical Engineering Systems and Technologies, pp. 420-446. Springer, Cham, 2018.

12. Susmitha Wunnava, Xiao Qin, Tabassum Kakar, Elke A. Rundensteiner, and Xiangnan Kong, *One Size Does Not Fit All: An Ensemble Approach Towards Information Extraction from Adverse Drug Event Narratives*. HEALTHINF. 2018.

13. Susmitha Wunnava, Xiao Qin, Tabassum Kakar, Elke A. Rundensteiner, and Xiangnan Kong, *Bidirectional LSTM-CRF for adverse drug event tagging in electronic health records*. International Workshop on Medication and Adverse Drug Event Detection. 2018.

14. Susmitha Wunnava, Xiao Qin, Tabassum Kakar, Elke A. Rundensteiner, and Xiangnan Kong, *Deep Learning Strategies for Automatic Detection of Medication and Adverse Drug Events from Electronic Health Records*. AMIA. 2018.

Contents

1 Introduction **14**
1.1 Motivating Use-Case from the U.S. Food and Drug Administration (FDA) 14
1.2 Incident Reports Analysis – An Opportunity for Visual Analytics 15
1.3 Research Problems in Incident Reports Analysis 16
1.4 State-of-the-Art in Incident Report Analysis 17
1.5 Research Tasks for Incident Report Analysis 18
1.6 Anticipated Research Outcomes 19
1.7 Dissertation Outline 19

2 Incident Report Analysis Background **21**
2.1 Incident Reporting Systems 21
2.2 FAERS Incident Reports 21
2.3 Incident Reports Analysis Workflow 21

3 Interactive Exploratory Triage and Screening of Incident Reports **23**
3.1 Introduction 23
3.2 MEV System Framework 25
3.3 The MEV System Overview 25
3.3.1 The Treemap Panel 26
3.3.2 The Timeline Panel 27
3.3.3 The Demographics Panel 27
3.3.4 The Reports View 27
3.3.5 System Implementation 28
3.4 Evaluation 28
3.4.1 Pharmacovigilance Usage Scenario 28
3.4.2 User Study 29
3.4.3 Expert interviews & Overall Impression of MEV 31
3.5 Related Work 32
3.6 Discussion 33
3.7 Conclusion 34

4 Visual Analytics for Exploration and Investigation of Hypothesized Incidents **35**
4.1 Introduction 35
4.2 Related Work 37
4.2.1 Association Rule Visualization 37
4.2.2 Drug-Interaction Visualization and Network Diagrams 38
4.3 Task Characterization 39
4.3.1 Requirements Analysis 39

4.4 Augmented Interaction Model . . . 40
4.4.1 FAERS Reports . . . 40
4.4.2 Augmented Interaction Model (AIM) . . . 40
4.4.3 The AIM Model Instantiation . . . 42
4.5 DIVA System Overview . . . 43
4.6 Design of DIVA Interactive Views . . . 45
4.6.1 Signal Screening Overview . . . 45
4.6.2 Signal Triage View . . . 46
4.6.3 Signal Forensics View . . . 48
4.6.4 Reports View: Revealing Reports . . . 49
4.6.5 DIVA Capability for Signal Annotation . . . 50
4.6.6 Interactions on Linked Views . . . 50
4.7 DIVA Evaluation . . . 50
4.7.1 Case Studies . . . 50
4.7.2 Interviews with Domain Experts . . . 52
4.8 Discussion . . . 53
4.9 Conclusion . . . 54

5 Visual Summary for Confirmatory Triage of Incident Reports 56
5.1 Background and Related Work . . . 57
5.1.1 Visual Analysis of Text Documents . . . 57
5.1.2 Triage in Other Domains . . . 58
5.1.3 Incident Report Analytics . . . 59
5.2 Understanding Incident Report Triage . . . 59
5.3 Triage Requirements Elicitation and Characterization . . . 60
5.3.1 Triage Cues: The Critical Information Used in Triage . . . 60
5.3.2 Challenges in Reports Triage . . . 61
5.4 Design of a Glanceable Visual Summary for Reports Triage . . . 62
5.4.1 Overall SumRe Layout . . . 63
5.4.2 Design Considerations for Triage Cues Depictions . . . 64
5.5 Evaluation . . . 65
5.5.1 Overall Study Design . . . 66
5.5.2 Results . . . 69
5.6 Discussion . . . 73
5.6.1 Triage Design Shapes Analysis, Exploration, and Insights . . . 74
5.6.2 Design Implications for Triage Workflows . . . 75
5.6.3 Limitations . . . 75
5.7 Conclusion And Future Work . . . 76

6 Interactive Case Building and Management for Incidents of Concern 77
6.1 Introduction . . . 77
6.2 Background and Related Work . . . 78
6.2.1 Incident Report Analysis . . . 78
6.2.2 Visual Analytics for Document Analysis . . . 79
6.2.3 Analytics Record Keeping Tools . . . 79
6.2.4 Investigative Tools . . . 80
6.3 Understanding the IIA Workflow at FDA . . . 80
6.3.1 Background on Medical Incident Reports . . . 81
6.3.2 Current IIA Practices at the FDA . . . 81

6.3.3 Motivating Real World IIA Scenario from FDA 81
6.3.4 Challenges in Current IIA Practices 82
6.4 Data Abstraction and Design Requirements 82
6.4.1 Instance-based Incident Analysis (IIA) Data Model 82
6.4.2 IIA Tasks and Design Requirements 84
6.5 Designing ConText for IIA 85
6.5.1 ConText User Interface 85
6.6 Development Process and Insights 87
6.7 Evaluating ConText for IIA in Pharmacovigilance 88
6.7.1 Case Studies 89
6.7.2 Expert Interviews 91
6.8 Discussion 92
6.9 Conclusion 95

7 Conclusion **96**

8 Discussion and Future Work **97**
8.1 Exploratory Triage of Incident Reports 97
8.1.1 Analytics for Exploring and Monitoring a Drug's Lifetime Incidents 97
8.1.2 Designing Transparent Overviews for Incident Triage 98
8.2 Confirmatory Triage of Incident Reports 99
8.2.1 Studying the Effect of a Visual Summary on User Engagement and Reading Behavior 99
8.2.2 Studying the Effect Contextual Information in Incident Report Triage 101
8.3 Investigatory Analysis of Incident Reports 101
8.3.1 Designing and Studying Semantic Similarity Measure for Incident Investigation 101

List of Figures

1.1 *Multi-state fungal meningitis outbreak involving the New England Compounding Center (NECC).* 15

1.2 *Research tasks to detect potential critical incidents of concern. (d) is out of the scope of this thesis.* 18

3.1 The Medication Error Visual Analytics (MEV) Approach 24

3.2 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 report narratives of the screened errors. 26

3.3 Reports view for the medication errors screened through the *MEV Interface* with no personal information (Left). Example of a de-identified text narrative (Right) 26

3.4 MEV with the selected date, demographics, drugname and medication error type reveals that the majority of the errors are happening during the preparation stage and packaging and confusion in the name of the drug are causing the errors. 28

3.5 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*. 32

4.1 The user interface of DIVA, a web-based visual analytics system for exploring and verifying Drug-Drug Interactions (DDIs) proposed via machine learning methods. (a) Screening Overview – showing candidate Drug-Drug Interactions for selected score. (b) Signal Triage View – enables drug-centric analysis of interactions, including the number of severe Adverse Reactions (ADRs). (c) Signal Forensics View – a view of the interaction profile of a drug of interest, including all ADRs triggered by each signal. (d) Controls facilitate navigation between views, and direct filtering by drugs of interest. (e) Legends for colors. 37

4.2 Augmented Interaction Model (AIM) 41

4.3 Overview of the DIVA framework consisting of two main components: the AIM Constructor and the DIVA Visualizer. 43

4.4 FAERS Reports associated with interaction Lansoprazole and Digoxin. Every report has Furosemide which is used to treat kidney disorders. 44

4.5 (a): Screening View gives an overview of interactions between drugs. Each node is a drug and edge depicts an interaction between two drugs. (b): Color Legends for: (i) Interestingness score in all views, (ii) Severe adverse reactions (or "DME") count in the Triage View, (iii) Shape of links representing label status in the Screening and Forensics view. 45

4.6 The Triage View - (a) Drug of interest Omeprazole. Interacting drugs with an (b) unknown and high score signal, (c) Unknown and low score signal, (d) Known and low score signal. (e) Pin a drug. (f) Zoom in to view the Profile of Omeprazole. (g) Color of box represents total number of severe reactions present in all signals related to Omeprazole. (h) Sort. (i) Pinned drugs. 47

4.7 Signal Forensics View: a tree layout allows an analyst to view and analyze the whole Augmented Interaction Model of a drug. (a) The root node represents the drug of interest- Omeprazole. (b) The second level represents drugs interacting with Omeprazole. (c) The third level represents the reactions related to each DDI. (d) DDI leading to severe reactions - purple color. (e) shows a signal with highest score - dark color. (f) Link between Omeprazole-Simavastatin depicts the aggregated score and unknown status because some of the signals are unknown. Each path starting from root until leaf is a hypothesized signal and represents an association rule. 48

4.8 Forensics View for drug Lansoprazole. Interaction with Digoxin leading to acute kidney injury, a DME, is unknown, and is highly scored by the rule mining hence worthy of further investigation. 50

4.9 Forensics View for Byetta. Interaction with Victoza and Januvia is just a co-occurrence and not an interaction as all three of these drugs belong to same drug class that treats diabetes. 52

4.10 Forensics View for Ondansetron. Interaction with lithium leading to severe and rare ADR 'serotonin syndrome' has been labeled by the FDA recently. 53

5.1 The current tool used for incident triage at the FDA uses a standard Tabular layout. Our aim in this work is to design and evaluate an alternative tool, SumRe, which draws on visualization principles to better encode and communicate critical safety report information. 58

5.2 We find that one key part of incident report triage involves what we call *triage cues*. Analysts first seek primary cues about the patient, incident, and key events first to form an assessment of plausibility, severity, and likelihood of further investigation. Analysts then review the secondary cues that can confirm or reject their hypothesis. (X*n*) represents the attribute type (C= Categorical, N=Numeric) and it cardinality. E.g. (C*) means a high-dimensional categorical attribute with hundreds of categories. This characterization informs the visual encodings used in SumRe. 60

5.3 SumRe layout to triage incident reports. (a) Profile Panel presenting information about the patient and the reporter. (b) Incident Panel representing all the information associated with the reaction (incident). (c) Analysis Panel allows analysts to add comments and triage action (Investigate or Do Not Investigate). Icons: (i) Regular drug, (ii) Onset, (iii) Severe Reaction, (iv) Serious outcome, (v) Dechallenge and Rechallenge respectively. Missing information is represented with grey color. Arrows are added for illustration and are not part of the design. . 63
5.4 Icons for the Primary Triage Cues. Red color depicts a serious or plausible cue, blue color encodes non-serious or implausible cues, and grey represents missing information. 64
5.5 Triage task to select an action (Investigate or not) on each individual report. SumRe has significantly high precision, i.e., correct selection of reports indicative of investigation. Recall is the measure of correct selection of all the reports. 69
5.6 Overall accuracy and time for identifying if reports are complete. Although insignificant, but the Tabular layout outperforms SumRe. . . . 70
5.7 Overall accuracy and time for pattern detection questions (Q4-Q6). SumRe significantly outperforms the Tabular layout in both accuracy and speed. 70
5.8 (a). Average count of identified unrelated issues. (b). Average number of times when the reference for triage criteria was viewed 73
5.9 (a) Count of participants able to recall information along-with type of information recalled. (b) Participants' preference (count) for using the layouts for reports triage (left), along with the reason for selecting their preference (right). 74
5.10 Average ratings for participant's perceived confidence and ease on completing the tasks. 74

6.1 Instance-based Incident Analysis (IIA) workflow in the FDA with tools supporting each task. Tasks (a & d) are out of the scope of this work. . . 80
6.2 Instance-based Incident Analysis (IIA) data model. Incident reports represent the instances of a reported (suspected) problem. IIA involves building case-series for each of the suspected problems by identifying reports that can serve as evidence to confirm theses problems. The data inside each report, which can vary, is evidence to the problem in each instance. 83
6.3 The Incident Analysis View allows analyst to identify evidential reports with line-listing of reports on the left and Content Analysis Panel on the right. (a) Current weeks reports. (b) Panel to view searched reports. (c) The dashboard panel with active cases. (d) To add a report to a case. (e) Menu to correct and control the annotations in the narrative. (f) Option to search across the reports and creating notes on a report. Due to privacy concerns, a FAERS-like publicly available safety report narrative [54] is shown in the Content Panel. 85
6.4 The Recommendation Query Panel with both NLP and user generated descriptors to allow analysts steer investigations by finding relevant evidential reports. 86

6.5 Dashboard view to build and monitor cases. Dashboard for the Case 'Cisatracurium'. (a) Reports within the case. (b) Recommended reports for the case. (c) Controls to set the recommendation query. (d) Case summary. (e) Alert of getting new evidential reports for the case Exforge. 89
6.6 Participants feedback (ratings) on the importance of features. Ratings are at 5-point likert scale (1=least important, 5=most important.. 92

8.1 Mockup of a drug's lifetime incident triage. 98
8.2 Baseline table and plain textual narrative (TP) condition. 100
8.3 Mockup for the baseline table and highlighted textual narrative (TH) condition. 100
8.4 Mockup for the visual Summary and textual narrative with word-scale visuals. 100
8.5 Mockup for the visual summary with contextual information. An analyst can view the report attributes associated with a drug and compares to the reports usually received for that drug. 101

List of Tables

3.1 List of Tasks designed to evaluate the effectiveness of MEV 30
3.2 U-Test for both time and perceived ease. Tasks T8 & T9 are not supported by the *control* (current tools). 31

4.1 Questions covered in a two hour interview with a group of drug safety analysts from the FDA 52

5.1 Task Questions and Description 68

Chapter 1

Introduction

Incident reporting systems are a common element of many organizations today. Customer service complaint systems [188, 67], for example, allow customers to submit a report of an issue they encountered. A company may review and consider such reports to take action. Other examples of such systems that target critical areas exist. In the aviation industry, for example, it is mandatory for the US Federal Aviation Administration to investigate service incident reports to better identify ongoing maintenance issues with aircraft [7, 118]. Similarly, one responsibility of the US Consumer Financial Protection Bureau involves analyzing consumer complaints about unfair, deceptive, or abusive financial practices [37].

One such example is the incident reports received in Pharmacovigilance, that is, the practices involved in the monitoring of the safety of medications after they have been licensed for use [78]. These semi-structured incident reports contain information about potential adverse reactions or medication errors that happened to the patients. Drug safety analysts in Pharmacovigilance regularly screen and investigate thousands of these incoming reports to identify incidents of concern by collecting supporting evidence in a timely fashion to take regulatory action to prevent such incidents from happening in the future.

Thus, sense-making of these large number of complex reports is challenging for analysts who need to make critical yet timely decisions on a daily basis. Recently developed computational techniques [181, 98, 79] although useful, are not completely accurate. Hence need the analysts judgment in assessing the outcome of these techniques. A need for the human-in-the-loop makes the incident analysis process a suitable candidate for visual analytics, which is the *focus of this dissertation*.

Below, we consider a use case from the Pharmacovigilance agency in the U.S., the Food and Drug Administration (FDA), revealing the importance of incident report analysis and the challenges involved in finding and leveraging information from these reports for decision making.

1.1 Motivating Use-Case from the U.S. Food and Drug Administration (FDA)

In 2012, the drug safety analysts at the FDA were conducting their routine review of incident reports when they observed several reports of unexpected death and hospitalization from across the country [100, 62, 101]. This seriousness of the reports alerted the analysts to find more information about these particular incidents

(Fig. 1.2a). Initial digging revealed that the adverse reaction "fungal meningitis" being reported is in fact a rare reaction. Upon executing further queries, the analysts uncovered that patients who are taking steroid injections were reporting these incidents. But steroid injections are not known to cause fungal meningitis. The analysts thus began to investigate if it was a false alarm and something else was causing these deadly events, or if the injections were the culprits in these incidents.

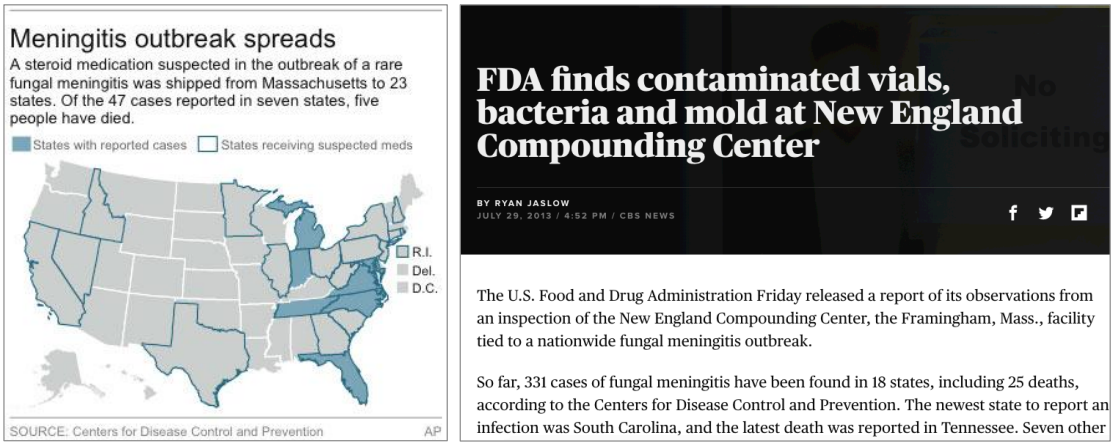


Figure 1.1: *Multi-state fungal meningitis outbreak involving the New England Compounding Center (NECC).*

To investigate this incident, the analysts had to examine individual reports associated with these meningitis incidents to find more information about these incidents and understand their root cause (Fig. 1.2b). Upon extensive search across the reports and close observation of the text narratives they found that the affected patients received the steroid injections from the New England Compounding Center.

After finding more reports pointing to NECC (Fig. 1.2c), they built a case and presented to their administration for decision making (Fig. 1.2d). Upon the inspection of this NECC facility, it was revealed that the product was contaminated due to the violation of FDA standard. Thus regulatory actions were taken that led to the indictment of the NECC [62]. This incident corresponds the well-known fungal meningitis outbreak scandal in Massachusetts that killed 76 people and hospitalized 700 nationwide [101]. Clearly, the more effectively we can support the tasks of such investigative analysis cycle, the faster we can solve potentially life threatening public health threats, such as the crisis described above.

1.2 Incident Reports Analysis – An Opportunity for Visual Analytics

From the above motivating scenario we learn that incident reports are submitted so that potentially avoidable or preventable incidents are brought to the attention of regulatory agencies. Analysts need to carefully analyze these reports in a timely manner to ensure safety issues are not missed and appropriate action is taken to mitigate any identified issues. Hence, the reports review process need analysts's judgment through each step. However, the number and complexity of these reports requiring analysts' judgment and attention to detail along with the high-stakes involved in the decision make it challenging for the analysts to efficiently perform reports review activities. A complete automation of this process of incident reports analysis with the human-in-the-loop does

not solve the problem, and thus is a viable candidate for visual analytics, an area that has not been explored in this context. In this dissertation, we design visual analytics to support the incident reports review process.

The vast majority of existing visualization techniques [111, 63, 161, 166] as well as commercialized tools such as Tableau and Microsoft Power BI provide limited support for the exploration of textual documents. On the other hand, techniques designed for text analysis allow sense making of the documents at a high level such as understanding the overall topics discussed in a corpus [111]. While incident analysis requires a more focused and thorough examination of the reports following a review workflow. Prior works have used visualizations to detect adverse reactions related to a particular drug from incident reports [30, 31, 154]. These works have focused on visualizing and understanding specific drug or safety issue, while our goal is to support the analysts' drug safety review workflow to detect any reported incidents of concern.

1.3 Research Problems in Incident Reports Analysis

Before designing visual analytics to help the incident reports review process, based on our more than two years of collaboration with the Pharmacovigilance domain analysts (Chapter 2), we identify several challenges that need to be addressed.

Challenge 1: Analyzing large collections of incidents. Incident reports are received in large numbers. In 2016 only, FDA received approximately 1.7 million incident reports [65]. Moreover, the number of incidents being reported and overseen during the analysis is large as well. For instance, every week FDA receives and investigates hundreds of reports about thousands of approved medications causing hundreds of adverse reactions or medication errors [65]. Recently develop computational methods generate even larger number of hypotheses about potential incidents [135]. Hence, it is challenging for the analysts to screen and prioritize the most critical incidents from a large pool of reported incidents to form a hypothesis about a potential concerning incident in a timely manner.

Challenge 2: Making sense of each individual report at a glance: After prioritizing a critical incident for analysis, analysts must review each and every report associated with that incident to make sure not to miss a single potential threat. Analysts must review certain information within each report to decide if further investigation is needed. Investigation at this stage means if they need to review the report narrative. Not every reported incident, however, requires further investigation, as some of these reports can be about known issues with drugs (thus no additional investigation to the one ongoing maybe needed) while others either provide confounding or incomplete information to assess the incident. Hence, it is challenging for analysts to get the gist of all the information within a report at a glance to form a hypothesis about a concerning incident given the huge influx of reports.

Challenge 3: Monitoring the investigation of multiple concerning incidents: When a hypothesis is formed about a potentially concerning incident, an in-depth investigation is started to collect evidence to help the analysts validate the hypothesis. This is because the decision of the investigation is critical involving regulatory action such as banning a drug from the market. Hence, careful examination of each report as well as collection of similar reports that serve as evidence is needed to conclude a finding to take a regulatory action. Moreover, an investigation may be open for a

long time if enough supporting evidence is not found. This thus requires the analysts to keep track of multiple ongoing investigations as well to analyze newly reported incidents which can be challenging, if done manually.

1.4 State-of-the-Art in Incident Report Analysis

Much of the existing work in Pharmacovigilance has focused on applying computational techniques on the incident reports to detect reactions or generate hypothesis about severe reactions [98, 143, 79]. Substantial amount of work is done on developing natural language processing (NLP) techniques [181, 182] to extract key information from these text reports. These techniques are useful but not completely accurate, and thus are not sufficient to completely automate the reports review workflow.

On the other hand, online publicly available interactive tools provided by the FDA such as OpenFDA [99] and OpenFDAVigil [26] help to filter drug related adverse reactions, and compare drugs based on their safety profiles (adverse reactions) using the incident reports. These tools can be helpful to promote awareness in the public as well as medical professionals about the potential adverse reactions of a drug before using or prescribing the drugs. These systems, however, do not support the analysts in the in-depth analysis of incident reports for regulatory decision making.

Approaches that have combined data mining and visualization to detect drug related adverse reactions also exist [187]. Yildirim et al. [187] use clustering on FDA incident reports to highlight relationships between Erythromycin, an antibiotic drug, and its adverse reactions based on the patients demographics and event outcomes. Similarly, ADRVis [92] uses disproportionality analysis to visualize a candidate drug's correlations with the reported adverse reactions. GraphSAW [154] uses a network diagram to visualize known drug-drug interactions by integrating multiple data sources. These existing tools focused on a particular drug or drug safety issue, and do not address the broader scope of drug safety review activities performed by the analysts on a regular basis.

With the significant growth of textual data, recently, visual text analysis has emerged in various domains including but not limited to digital humanities, social media, and medicine [110, 103]. In the medical domain, the focus of these existing tools has been on the analysis and summarization of electronic health records (EHR) and patient summaries [138, 55]. These tools are not suitable for incident report analysis because of their different analytics goals. For instance, in EHR, majority of these tools display overviews of patient's medical history to help the medical professionals with diagnosis or treatment. While, in case of the incident reports, the analysts need to perform in-depth analysis of the reports to assess the root cause of the incidents to build a case to take a regulatory action.

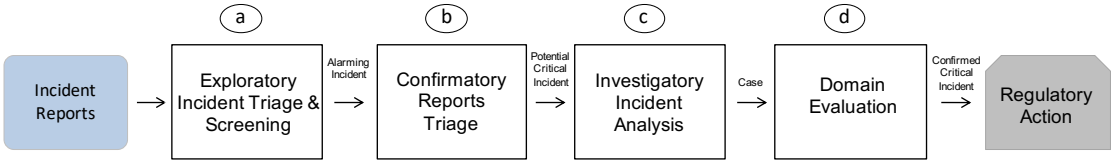


Figure 1.2: *Research tasks to detect potential critical incidents of concern. (d) is out of the scope of this thesis.*

1.5 Research Tasks for Incident Report Analysis

In this dissertation, we address the above mentioned challenges by designing interactive visual analytics approaches based on domain-workflows to support the analysts in the triage and investigation of the influx of incident reports received on a regular basis. We aim to achieve this goal through three research tasks as depicted in Figure 1.2.

Research Task 1: Exploratory Incident Triage.  Given the large collection of incident reports received every week, screening and triage of these reports for critical incidents is important. We propose two approaches to triage these reports at an overview level (Fig. 1.2a). Their results are evaluated with user studies and case studies.

a. In the first work, we design *a systematic approach that combines NLP-based meta-data with interactive visualizations* to assist analysts in prioritizing the most critical incidents by exploring the overall distribution of the core reported data elements [96]. The results of our user studies show that incident screening tasks can be performed quicker and are perceived to be easier using our proposed interactive visual analytics system compared to the existing tools used by the analysts **(Chapter 3)**.

b. In the second work, we design a *visual analytics approach (DIVA) to help analysts explore and triage hypothesized incidents* generated by computational techniques to guide analysts towards most concerning incidents. We design multiple coordinated views integrating domain knowledge and workflows to help analysts explore potentially critical incidents [95] . Expert interviews and case studies conducted using incident reports confirm the effectiveness of DIVA in exploration and triage of potential critical incidents **(Chapter 4)**.

Research Task 2: Confirmatory Incident Reports Triage.  Incidents screening and triage by exploring reports using 'overviews' guide analysts towards potentially critical incidents. Analysts, however, need to analyze certain information within each individual report associated with these triaged incidents to assess the incident and its context and decide if further investigation is needed (Fig. 1.2c). We design and evaluate *a glanceable visual summary* that provides compact view of the crucial information within each report to help in forming a hypothesis about a potential critical incident. A user study with 20 domain experts demonstrates the effectiveness of a glanceable visual summary for reports triage **(Chapter 5)**.

Research Task 3: Investigatory Incident Report Analysis.  After forming a hypothesis about an incident of concern, analysts often pivot their analyses toward finding and organizing similar incident reports to build a case for action. We design *interactive visual displays and features (ConText) to aid analysts in evidence collection and management for incidents of concern* by a close coupling of information foraging and synthesis (Fig. 1.2c). We conduct case studies as well as interviews with the Pharmacovigilance experts to evaluate the effectiveness of ConText in building and managing multiple ongoing cases **(Chapter 6)**.

1.6 Anticipated Research Outcomes

The tasks involved in the incident report analysis are substantially different than the regular exploratory document analysis [110], as incident reports require a deeper and closer look into the text stating the incidence. The tools resulting from our research tasks have the potential to aid the incident review process by providing *user-driven computation-supported design techniques*. Particularly, analysts in the Pharmacovigilance domain will be able to efficiently screen and triage reports to not only identify incidents of concern but also conduct in-depth investigations. Analysts can interactively collect evidence for any concerning incidents and automatically track ongoing investigations in a mixed-initiative manner [86]. Moreover, our research lays the foundation for designing visual analytics systems not only for Pharmacovigilance but also for regulatory agencies in other domains such as finance [37] and aviation [7], where similar analyses are performed.

The outcome of this work is useful for researchers in diverse fields from machine-learning to visualization. In the machine-learning domain, our work opens research questions such as predicting a potential new incident based on the features of the previously identified incidents, or the automatic recommendation of an action while triaging for incident reports. Our research opens opportunities for studying user's behavior while performing detailed analysis on incident reports and understanding the process of decision making for flagging an incident as 'critical'. Additionally, we aim to develop techniques that can learn from a user's interaction with the system to identify incidents of concern from a pool of incident reports. In the visual analytics field, our research uncovers potential opportunities for designing trustworthy and transparent visual overviews that would help the analysts view the uncertainty inherited from computational techniques [49]. One way to achieve this is by integrating statistics and confidence distributions regarding the accuracy of the data generated by NLP and machine-learning techniques. Such information, if communicated visually via carefully designed encodings to help analysts identify the uncertainty in the underlying data, can help improve users' trust [116].

1.7 Dissertation Outline

This dissertation thesis is organized as follows. Chapter 2 characterizes the background information about the Pharmacovigilance domain and the incident reports analysis workflow. Chapter 3 describes the design of *MEV*, an interactive treemap-based overview of incident reports to interactively explore and triage critical incidents. Chapter 4 summarizes the design of our visual analytics tool, *DIVA*, for the exploration

and triage of hypothesized incidents by providing multiple coordinated views. Chapter 5 discusses the design and evaluation of a glanceable visual summary, *SumRe*, for triage of individual incident report associated with a screened incident. Chapter 6 illustrates our work on *ConText* an interactive case building and management tool for efficiently monitoring multiple incidents of concern. Chapter 7 concludes by sketching possible avenues for future work in visual analytics for incident triage and investigation which build on the areas of study in this dissertation.

Chapter 2

Incident Report Analysis Background

This chapter provides the background information about the incident reports and their analysis.

2.1 Incident Reporting Systems

Drugs after being released to the market are monitored by drug regulatory authorities to detect unanticipated adverse reactions that were not discovered during clinical trials. This is achieved by a process called post-marking drug surveillance. In the U.S., the Food and Drug Administration (FDA) conducts post marketing surveillance via the FDA Adverse Event Reporting System (FAERS) [65]. Similar systems are also in operation internationally, including the World Health Organization [108], as well as in Canada [40] and Britain [16]. Analysts in these organizations perform in-depth analysis of hundreds of incident reports on a daily basis, with the aim of identifying critical incidents that can be rectified and in some cases even be stopped from happening in the future through regulatory action. Early detection makes it possible for authorities to minimize patient exposure to harmful adverse reactions.

2.2 FAERS Incident Reports

Reports submitted to the FDA Adverse Event Reporting System (FAERS) are semi-structured in nature. These reports contain structured information about patient demographics, drugs taken, therapies, and adverse reactions. They also contain an unstructured textual narrative that describes the adverse reactions in detail and contains richer information such as a patient's medical history and the details of the incident to help analysts decide if incident is worthy of investigation. This collection includes mandatory reports submitted by drug manufacturers and voluntary reports submitted by health care professionals and consumers. To ensure the **reproducibility** of our research, we focus on a public version of FAERS data that includes the structured information with no personal identification and is publicly available on a quarterly basis [65]. For some tasks, we have redacted the text narratives to avoid any privacy breaches.

2.3 Incident Reports Analysis Workflow

For this dissertation, through our collaboration, we worked closely with domain experts at the FDA who serve as drug safety analysts. We organized a series of semi-formal interview sessions with drug safety reviewers. A primary aim was to understand the current incident report analysis process and to identify challenges that analysts face in analyzing incidents. The goal of incident report analysis is to

identify potential safety issues related to drugs, and to escalate cases for further action if sufficient supporting evidence is found during the evaluation of potential incidents. The drug review process is composed of iterative steps as depicted in Figure 1.2. Each safety analyst receives reports related to the drugs assigned to them.

The drug analysts we interviewed screen their assigned reports for any alarming incidents. As a next step, the analysts perform in-depth analysis of the report narratives to confirm if the incident needs further investigation. If a potential critical incident is identified, then the analysts build a case by searching for similar reports. If the analysts find sufficient evidence to move forward, they proceed by evaluating patient's medical histories and relevant medical literature to find additional evidence supporting a potential critical incident. If there is sufficient evidence to confirm an incident, then analysts formulate their recommendations that can lead to regulatory action, such as changing drug labels or restricting drug usage. In severe cases, drugs are removed entirely from the market [183]. Currently, most of these analysts use SQL-based query interfaces to retrieve the reports of interest, and use Microsoft Excel to manage their ongoing investigations, and as they find a new report related to a case, they add it to the Excel file.

Chapter 3

Interactive Exploratory Triage and Screening of Incident Reports

In this chapter, we discuss the design of a visual analytics system for screening and triage of incidents caused by the medication errors at a macro-level via interactive overview of the reports. This work is completed and published in the International Conference on Information Visualization Theory and Application (IVAPP), 2019. An extended version is now in submission to the Communications in Computer and Information Science (CCIS), Springer book series 2019.

3.1 Introduction

Every year, serious preventable incidents caused due to medication errors occur in 3.8 million inpatient admissions and 3.3 million outpatient visits with an estimated annual cost burden of $20 billion [6]. 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. 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 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 actions. These actions may include revising container labels or instructions for use, communicating safety issues to the public, and in rare cases, changing a proprietary drug name.

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

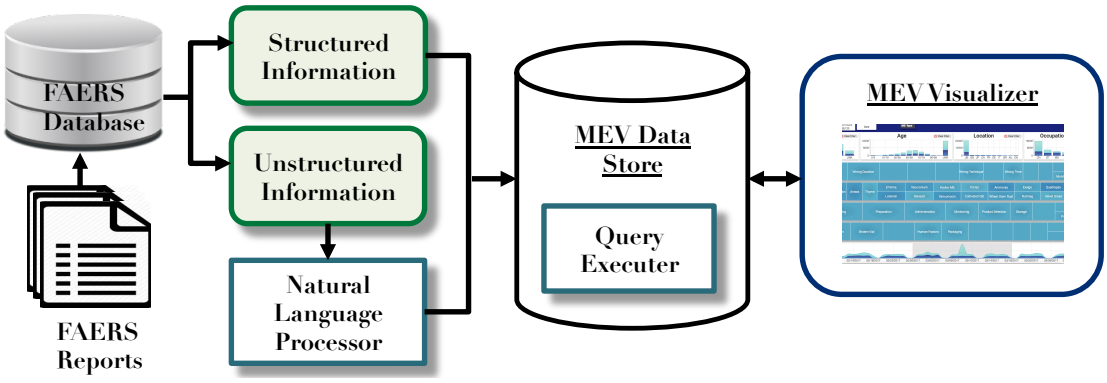


Figure 3.1: The Medication Error Visual Analytics (MEV) Approach

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 [147, 184, 13, 181]. MEV then displays this information along with other attributes associated with a given report such as drugnames on the treemap visualization (Fig. 3.2). 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 data 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.

3.2 MEV System Framework

Following the workflow of domain analysts, MEV depicted in Fig. 3.1 is designed to explore the reports efficiently. As described earlier, FAERS reports contain both structured as well unstructured text narrative explaining the reported 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 [181] to extract key information from the text narrative.

We use domain specific lexicons [124, 33] 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. 3.1) 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 [58] for smooth exploration and analysis. Currently, analysts manually summarize each narrative by adding these terminologies into the Excel spreadsheet. After standardization, on average each extracted entity contains approximately 15-20 categories.

The extracted information along with structured information about demographics is stored in the MEV *Data Store* (Fig. 3.1). 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 the user experience. The MEV assists analysts in exploring the data interactively using linked interactive visualizations described below.

3.3 The MEV System Overview

Our MEV tool consists of four main interactive displays (Fig. 3.2), the treemap view, the demographics panel, the timeline panel and the reports view.

3.3.1 The Treemap Panel

A treemap visualization (Fig. 3.2b) 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.

Figure 3.2: 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 report narratives of the screened errors.

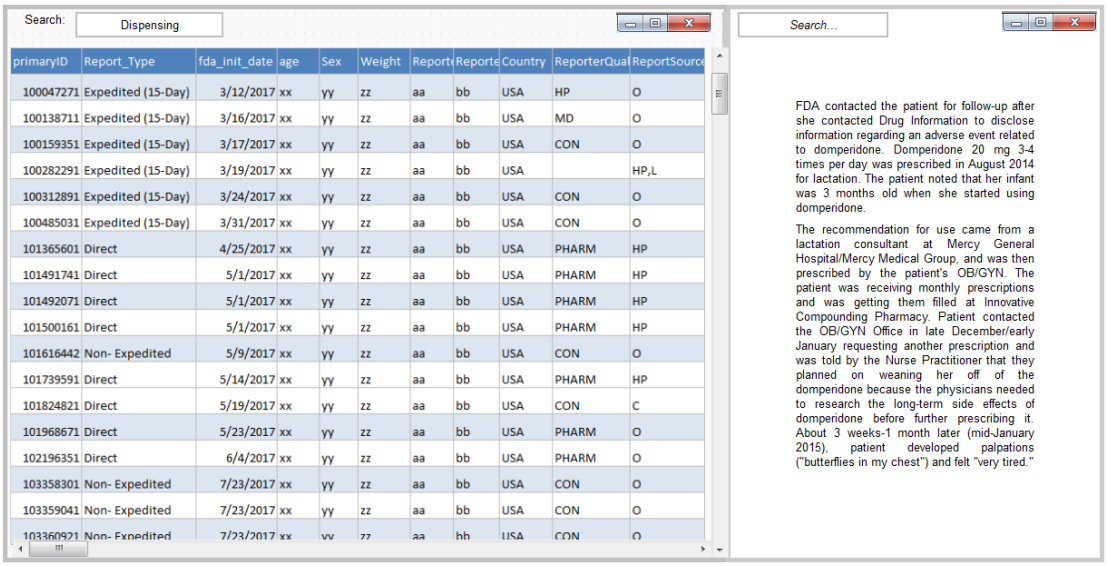


Figure 3.3: Reports view for the medication errors screened through the *MEV Interface* with no personal information (Left). Example of a de-identified text narrative (Right)

.

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 [114]. 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 [161] having similar functionality may have desirable properties.

3.3.2 The Timeline Panel

The timeline panel (Fig. 3.2c) displays the overall report distribution as well as their severity over a 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.

3.3.3 The Demographics 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. 3.2a) 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.

3.3.4 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. 3.2d) the selected reports are accessible. The reports view displays the line listing of the screened data elements (Fig. 3.3-left). Analysts can drill into the narrative of each report to further examine the report in great detail (Fig. 3.3-right).

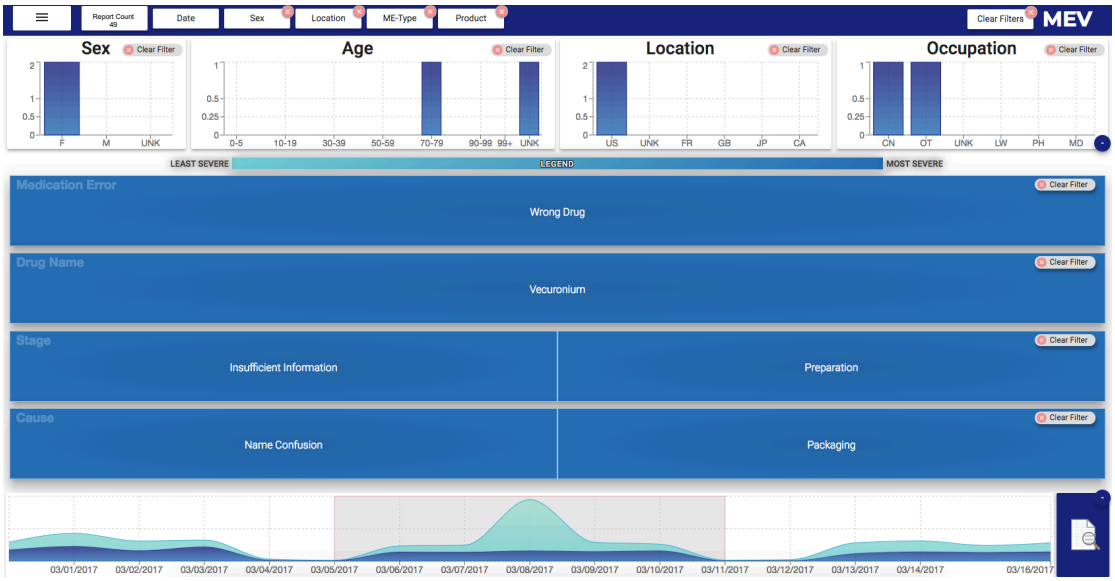


Figure 3.4: MEV with the selected date, demographics, drugname and medication error type reveals that the majority of the errors are happening during the preparation stage and packaging and confusion in the name of the drug are causing the errors.

3.3.5 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.

3.4 Evaluation

3.4.1 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. 3.2 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 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. 3.4-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. 3.4). She speculates whether these reports indeed have compelling evidence about these errors. She clicks on the reports icon (Fig. 3.2d) to read the details of each narrative in the reports view (Fig. 3.3). Hence, MEV interactively guides the analyst towards concerning errors by supporting exploration and screening of reports.

3.4.2 User Study

To have a quantitative analysis of performance of our proposed approach over the existing tool, we conducted a user study with domain experts at the FDA.

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. 3.1) 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 (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.

Table 3.1: List of 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?

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. 3.1) 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 [32] 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.

Study Results

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

Quantitative Analysis

From Table. 3.2, 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 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. 3.2). 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. 3.5 (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.

Tasks	Siginificance Test Time (α = 0.05)	Siginificance Test Easiness (α = 0.05)
T1	**(U = 29, 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

Table 3.2: U-Test for both time and perceived ease. Tasks T8 & T9 are not supported by the *control* (current tools).

Similarly, for perceived ease, Fig. 3.5 (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 [32]. MEV received an SUS score of 85 out of 100.

3.4.3 Expert interviews & 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

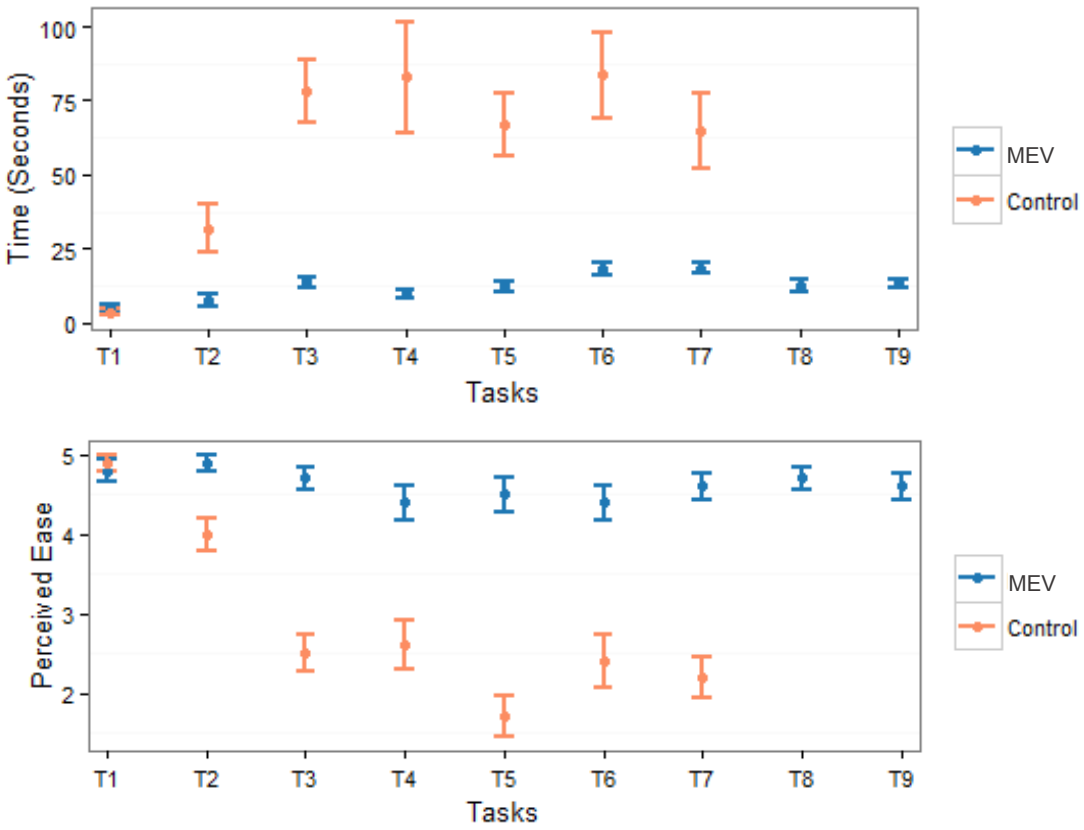


Figure 3.5: 95% Confidence Interval for performing tasks using both MEV and existing tool (*Control*). (Right): Perceived Ease Score. (Left): Time (in sec). Task 8 & 9 are not supported by *Control*.

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.

3.5 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 [159] supports interactive visualizations to explore facets of a dataset, however, it does not support discovering relationships among facets. FacetLens [107] extends FacetMap to help users observe trends and explore relationships within faceted datasets. Most of these faceted systems [107, 159] 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 [14] have been widely used in visualization systems [114, 80]. For

example, SellTrend [114], a visualization tool for displaying temporal categorical data, displays transaction failures using treemaps. NV [80] utilizes treemaps and histograms to allow security analysts and system administrators 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 [161], 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 [129] and clinical information systems [91]. Varkey et al. [173] 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 [129]. Other tools are designed as interfaces to provide a user-friendly and efficient mean of error reporting [157]. Clinical decision support systems have been proven to reduce medication errors during prescription [91]. 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.

3.6 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 [133]. 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 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 [181] 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 [88] 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 [74] would also add value by allowing analysts to share their thought-processes and findings with their team members.

3.7 Conclusion

In this chapter, 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.

Chapter 4

Visual Analytics for Exploration and Investigation of Hypothesized Incidents

This work describes the design of a visual analytics tool to aid analysts in the exploration of machine-generated hypothesized incidents. This work is completed and published at EuroVis, 2019.

4.1 Introduction

Medical incidents such as Adverse Drug Reactions (ADRs) caused by drug-drug interactions are a major cause of mortality, resulting in more than 100,000 deaths annually with a yearly cost of over $170 billion in the U.S. alone [106, 60]. Polypharmacy, the use of multiple drugs to treat medical conditions, is also rising. For example, approximately 29% of elderly patients are taking six or more drugs, which increases the chance of harmful and possibly fatal drug-drug interactions (DDIs)[38].

Before approval for use, new drugs are tested for interactions with existing drugs using both clinical trials & animal studies (in vivo) and tests on cells (in vitro) methods [189]. However, any given drug may interact with other drugs in numerous, unexpected ways. These interactions make it impossible to test all possible drug combinations before a drug is released to the market.

Incident reports submitted to FDA are a critical information source for discovering potential drug-drug interaction signals worthy of investigation. Such interactions may represent causal effects between a combination of drugs that result in dangerous adverse reactions. One challenge is that the manual approaches currently used for detecting and investigating candidate signals in large sets of drug reaction reports are tedious, time consuming, and error prone. Complicating the problem in practice is the reality that, due to staff limitations, a small team of roughly fifty analysts at the US FDA have dedicated time for reviewing these reports. Given these restrictions, the primary workflow of the FDA primarily focuses on signal detection related to single drug adverse reactions, with drug-drug interaction findings remaining more a matter of chance, despite its significant risks.

Automated approaches to drug reaction analysis are also insufficient. Machine learning techniques proposed to mine drug reaction reports for signal hypotheses tend to generate a large number of candidate signals [160, 10, 79, 87, 39]. For example, n distinct drugs and m unique adverse reactions across a set of reports result in up to $\mathcal{O}(2^{n+m})$ signals in the worst case. Regardless, machine-generated signals require

inspection by drug safety analysts, who must analyze and validate signals as worthy of escalation, or dismiss signals because of insufficient evidence.

In this work, we propose to address these challenges through a visual analytics framework called Drug Drug Interactions via Visual Analysis, (**DIVA** [1]), that supports drug safety analysts in analyzing drug-interaction signals mined from drug surveillance reports.

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

[1] A latest web-based release of DIVA is accessible at http://diva.wpi.edu:3000/ that follows the same core paradigm and design principles as DIVA

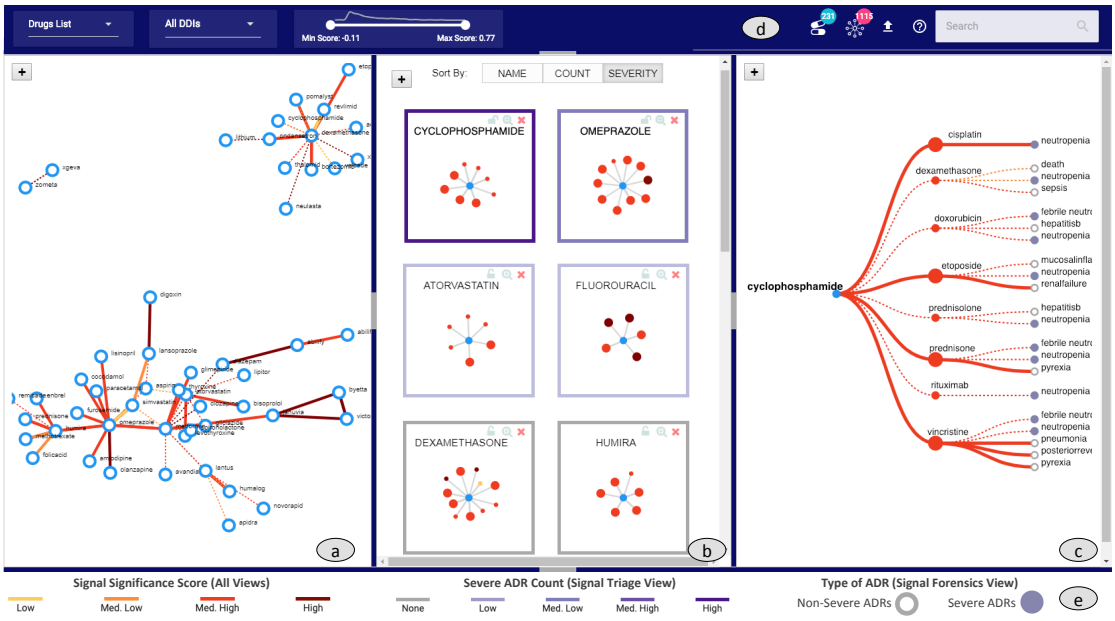


Figure 4.1: The user interface of DIVA, a web-based visual analytics system for exploring and verifying Drug-Drug Interactions (DDIs) proposed via machine learning methods. (a) Screening Overview – showing candidate Drug-Drug Interactions for selected score. (b) Signal Triage View – enables drug-centric analysis of interactions, including the number of severe Adverse Reactions (ADRs). (c) Signal Forensics View – a view of the interaction profile of a drug of interest, including all ADRs triggered by each signal. (d) Controls facilitate navigation between views, and direct filtering by drugs of interest. (e) Legends for colors.

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

4.2 Related Work

DIVA draws from prior work in visualization spanning at least two primary categories, including visualization techniques for association rules, and visualization systems focusing on the discovery and analysis of drug interactions.

4.2.1 Association Rule Visualization

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

Romero et al. [141] represent association rules with simple text in a tabular format. Grids (2D matrices) [128, 35] and 3D matrices [178] have also been proposed to visualize association rules for smaller data sets. Matrices with a fish-eye view [52, 51]

visualize association rules in more detail. InterVisAR [47] uses a two-dimensional bar chart approach to allow users to search for particular rules.

In Mosaic plots [83], individual antecedent items are shown as horizontal bars along the x-axis and the support of an association is represented by the height of the vertical column above the specified item. Existing graph based association rule visualization tools [81, 76, 168, 112] tend to focus on an overview of the generated rules, rather than the investigation and validation of a set of rules.

In parallel coordinates-based rule visualization [186, 77] each vertical line depicts a set of items and a rule is represented by lines or splines. Some initial work has combined two techniques to visualize rules [36, 151, 24]. Buono et al. [36] used both graphs and parallel coordinates to get an overview as well as a detailed view of selected rules. Sekhavat et al. [151] used matrices as overview of rules and graphs to analyze a selected subset of rules. Another approach includes a virtual arena [24] where rules are represented as spheres positioned by the steps of an arena. Similarly, glyphs have been used to represent quantitative values associated with the rules [134].

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

4.2.2 Drug-Interaction Visualization and Network Diagrams

Several recent studies have developed visualization systems for analyzing interaction between drug and proteins as well as with other drugs. Kegg [97], like other online tools [3, 2], is a search interface for known drug-drug interactions. In this work we have used such tools to extract known signals into a hypothesis-driven exploratory system for discovering and analyzing unknown signals.

Stitch [104] integrates data from various sources and uses a network visualization to represent Drug-Protein interactions. Promiscuous [175] integrates data from three different molecular databases and visualizes Drug-Target interactions and drug-related adverse reactions using node-link diagrams. Both of these tools focus on data integration and allow exploration of drug related chemicals, however, these tools are not designed to support drug safety analysis workflows. GraphSAW [154] integrates data about known drug-drug interactions from various sources. A radial network graph is used to visualize a set of adverse reactions and drug interactions. Network visualization techniques are also used to analyze vaccine related adverse events [19, 31, 29]. For example, Botsis et.al propose AENA [31], which uses a network diagram with an edge weighing algorithm to identify outliers in the U.S. Vaccine Adverse Event Reporting System.

While these tools do not support the specific analytic activity of conducting pharmacovigalance by analyzing drug reports for unknown interactions, they do form a broader landscape of tools that aid in the overall pharmacovigalance mission by providing access to known interactions between drugs and drug compounds.

4.3 Task Characterization

To design DIVA, we worked closely with domain experts at the FDA who serve as drug safety analysts. We used an interview-based iterative design process, presenting the analysts with progressively refined prototype visualizations to characterize the requirements in support of their workflow. In doing so, we arrived at the Augmented Interaction Model, a data abstraction as described by Munzner's nested design model [122], which serves as a basis for the visualizations and interactions in DIVA.

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

4.3.1 Requirements Analysis

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

Screening for Possible Drug-drug Interactions:

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

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

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

R4: ***Facilitate identification of unknown signals.*** Drug review analysts are interested in uncovering unknown, novel signals that constitute a hypothesis worth

escalation and further investigation. Therefore, unknown signals must be easily recognizable so they can remain a priority.

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

Verifying Hypothesized Drug-drug Interactions:

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

R7: ***Incorporate analysts feedback.*** Analysts may find a signal that is marked as significant by the mining algorithm, yet it turns out to be uninteresting in practice. Hence, the analyst should be able to annotate their findings to avoid repeated work.

4.4 Augmented Interaction Model

4.4.1 FAERS Reports

Reports submitted to the FDA Adverse Event Reporting System (FAERS) contain structured information about patient demographics, drugs taken, therapies, and adverse reactions. They also contain an unstructured textual narrative that describes the adverse reactions in detail and contains richer information such as a patient's medical history. This collection includes mandatory reports submitted by drug manufacturers and voluntary reports submitted by health care professionals and consumers. To ensure the **reproducibility** of our research, we focus on a public version of FAERS data that includes the structured information with no personal identification and is available on a quarterly basis [65]. The core data elements, such as drugs and ADRs within each report processed by our machine learning module, are available in this structured FAERS. While the inclusion of the actual narratives can be easily provided by DIVA to the FDA analysts, however, for privacy reasons they cannot be published publicly and thus are not included in our manuscript or publicly available prototype [2].

4.4.2 Augmented Interaction Model (AIM)

After rounds of interviews and initial design alterations, we developed a data abstraction [122] that reflects a unit of exploration from the FDA analysts perspectives. Drug interaction related signals information consists of data from various external sources, generated automatically as well as manually. We capture all this information in the form of entities, attributes and their relationships into an Augmented Interaction Model (AIM). The AIM provides all the information essential for an analyst to be able to explore and analyze signals, i.e., screen the important ones and validate them. In this section, we define the core entities that form the AIM abstraction and explain how

[2] http://diva.wpi.edu:3000/

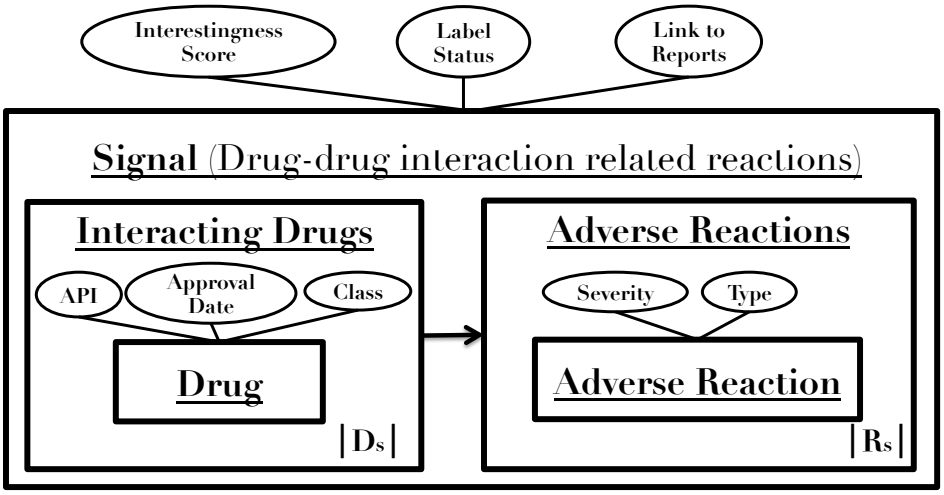


Figure 4.2: Augmented Interaction Model (AIM)

each component in AIM is instantiated by DIVA's visualizations and corresponding interactions.

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

The attributes determined as most useful in the drug review process include:

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

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

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

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

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

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

The attributes relevant to the review process include:

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

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

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

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

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

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

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

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

The AIM model represents the above mentioned entities namely, drugs and reactions, the signals composed of these entities and the domain knowledge that augments these signals as depicted in Fig. 4.2.

Definition 4. ***The AIM Model.*** *Given a set of reports $\mathcal{T}$, an Augmented Interaction Model (AIM) $\mathcal{M}^{\mathcal{T}}$ can then be captured by a set of signals $\mathcal{S}^{\mathcal{T}} = \{s_1, \cdots, s_n\}$ derived from $\mathcal{T}$. The attributes of the drug entity DE, adverse reaction entity AE and the signal s are populated based upon $\mathcal{T}$ and other domain knowledge such as Drugs.com [3].*

4.4.3 The AIM Model Instantiation

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

Instantiation of Entities and their Attributes. The FDA maintains a list of approved medical products, including drugs currently in the market [4]. Each drug is documented with detailed information such its active ingredients, approval date, and drug class. In this study, we extract these attributes from FDA resources [4] and construct a drug entity repository for use in DIVA.

For an adverse reaction (ADR) entity, we use the *Preferred Terms* from the MedDRA Hierarchy [5] to form an adverse reaction vocabulary. To specify the **severity** of these reactions (**R5**), we leverage the list of Designated Medical Events (DMEs) also known as severe ADRs maintained internally by FDA for review purposes. A severe ADR such as heart failure or liver injury is more alarming than nausea or headache. Thus it must be prioritized over less severe concerns to avoid further patient exposure.

Instantiation of Signals and their Attributes. Drug-drug interaction related adverse reaction signals are the core components of the AIM model. They correspond to severe candidate DDIs extracted from a set of reports. FAERS reports do not capture direct information about drug interactions. However, each report captures the drugs being taken by the patient along with the observed ADRs. Prior studies [79, 135] have suggested signal generation by modeling associations between drugs and reactions using their co-occurrence in the surveillance database. That is, *frequent pattern mining* methods have been applied to extract signals.

In this work, we adopt *MARAS* [135] technology to mine potential signals as sketched below. MARAS adopts association rule learning to identify relationships among objects that occur together in a database. In the surveillance database (FAERS), each record can be modeled as a combination of a reported drug set and the reported observed ADR set. The rules that model the relationship between a drug set and an ADR set are signals that need exploration and validation.

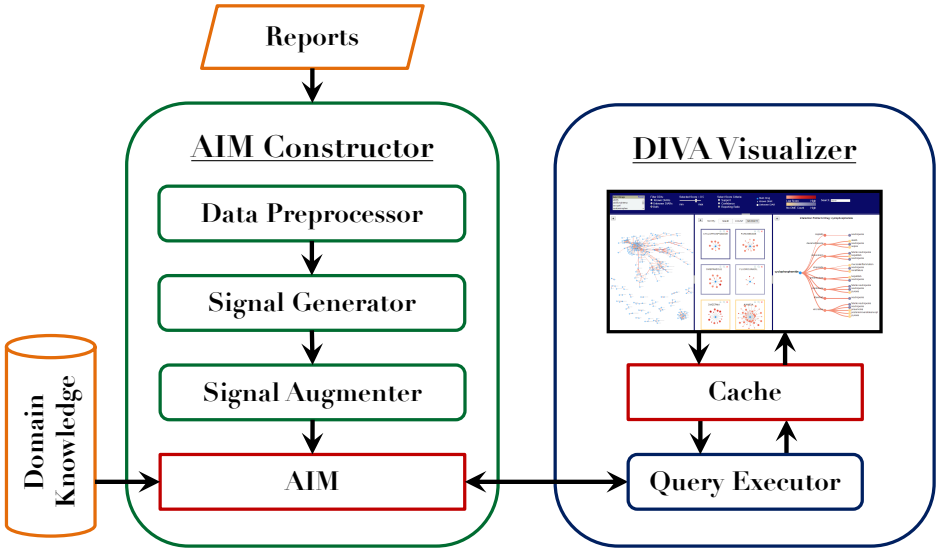


Figure 4.3: Overview of the DIVA framework consisting of two main components: the AIM Constructor and the DIVA Visualizer.

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

$$Rule = \mathcal{D} \rightarrow \mathcal{R}, \tag{4.1}$$

MARAS addresses issues related to avoiding misleading rules in this context through a *contrast* measure (See [135]). This evaluates how likely the ADRs are caused by a drug interaction. The intuition behind this measure is that if ADRs are triggered by the drug-drug interaction, then they should have less chance of being triggered by any of the individual drugs in the signal. The *contrast* measure is used as an **interestingness score** to help prioritize signals in DIVA. Signals with higher scores have more chances of being true signals warranting action, therefore need to be prioritized by safety analysts. Other proposed measures [142] can also be plugged in to the system. The linkage (Case-Ids in FAERS reports) to raw reports used to extract the signals is maintained during the mining process so that analysts can access these reports for signal validation (**R6**).

Computer-generated signals may be both already known (labeled) signals as well as unknown (unlabeled) signals. These generated signals can help drug safety analysts form hypotheses where they identify novel and severe signals worthy of further investigation. Therefore, such unknown signals must be distinguishable from the already known ones as such by their *Label Status*. Information about the status of a signal is not available in FAERS, neither does the signal extraction method provide this information. Hence, to assist analysts, we incorporate such information into our AIM model (**R3**) by extracting it from external sources [3].

4.5 DIVA System Overview

The DIVA framework depicted in Fig. 4.3 consists of two major components, namely the **AIM Constructor** that generates AIMs from reports and the **DIVA Visualizer** that supports interactive visual analytics of AIMs from multiple perspectives.

The **AIM Constructor** has three modules: The *Data Preprocessor*, *Signal Generator* and *Signal Augmenter*. The *Data Preprocessor* transforms the original FAERS reports

into the format required for the signal generation algorithm. During the preparation, duplicate reports are removed and drug names are cleaned due to different variants of same drugs and spelling mistakes.

The *Signal Generator* module adapts MARAS [135], a drug interaction signal extraction and scoring technique. Other machine learning techniques [160, 10, 79, 87, 39] are also candidates for producing signals that serve as input to DIVA. These generated signals model the association between drugs and ADRs, depicted in Fig. 4.2, along with its interestingness score (*contrast*) [135]. The *Signal Augmenter* populates the rejoining attributes of the signals to produce AIMs. Domain knowledge such as the label status and severity of ADRs is obtained from external data resources.

The **DIVA Visualizer** module, consisting of multiple coordinated views (Fig. 4.1), provides multiple perspectives into the Augmented Interaction Model. The views aim to align with the work-flow of the analysts so that they can explore the major components of the AIM in an iterative manner. The underlying queries and data exchanges between the AIM and the analysts are supported by the *Query Executor*. A cache is used to optimize the query execution time.

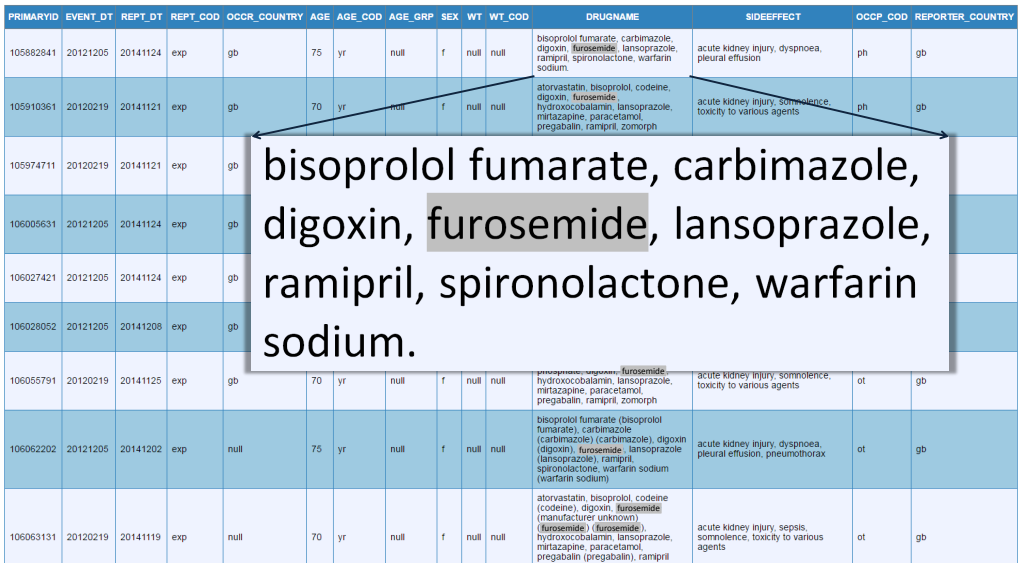


Figure 4.4: FAERS Reports associated with interaction Lansoprazole and Digoxin. Every report has Furosemide which is used to treat kidney disorders.

The *Screening* view gives an overview of all hypothesized drug-drug interactions supporting an analyst in screening unknown and high scored signals (Fig. 4.1-a). The *Triage* view, composed of small multiples, shows all the drug interactions associated with a particular drug or set of drugs. It helps analysts prioritize a drug for review based on the aggregated interestingness of its interactions (Fig. 4.1-b). The *Forensics* view includes adverse reactions related to each drug-drug interaction for further exploration (Fig. 4.1-c). To further investigate a drug interaction, at the lowest level, the *Reports* view visualizes the line-listings and text narratives of reports associated with a selected drug interaction (Fig. 4.4). We developed the visual interface of DIVA following the aforementioned design rationale (Section 4.3). All views are coordinated via brushing and linking, supporting hypothesis generation, exploration and validation in the context of drug interactions.

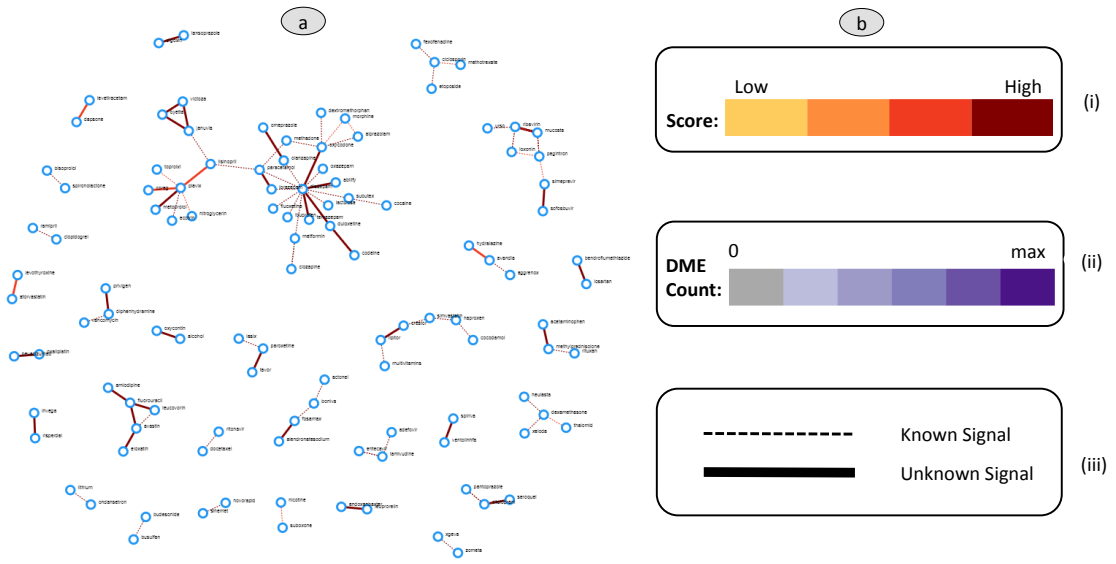


Figure 4.5: (a): Screening View gives an overview of interactions between drugs. Each node is a drug and edge depicts an interaction between two drugs. (b): Color Legends for: (i) Interestingness score in all views, (ii) Severe adverse reactions (or "DME") count in the Triage View, (iii) Shape of links representing label status in the Screening and Forensics view.

4.6 Design of DIVA Interactive Views

DIVA, a Web-based system, has been designed to fulfill the requirements elicited in Section 4.3. DIVA is composed of multiple coordinated interactive views that based on the tasks provide different perspectives into the AIM data model because of its richer content (Figure 4.1). The visual encodings reflect aspects of the AIM data model including information such as the drugs and reactions that compose a signal, an interestingness score, label status, and severity of adverse reactions (DMEs).

4.6.1 Signal Screening Overview

The Screening view provides an overview into drug-drug interactions (**R1**). This view allows analysts to see the entire space of the machine generated drug-drug interactions through a node-link diagram (Figure 4.5). Here nodes represent the drugs, while edges depict an interaction between a pair of drugs. The shape and size of the edge encodes whether an interaction is known (dotted and thin) or unknown (solid and thick). The color of an edge is mapped to the strength of the interaction as determined by the mining technique (the interestingness score) derived from the support and confidence (see Section 4.4.3).

A pair of drugs can contribute to multiple signals, each of which can have a different score. To avoid confusion and repetition of data in the Screening view, each drug is represented only once. This way an analyst can instantly examine the degree of possible interactions between drugs. When multiple signals are caused by the same drug pair, an aggregated score encodes the interestingness score represented by edges. In such cases, we use the maximum score of all generated signals as an aggregated score to represent the drug-drug interaction. A maximum score is used as an aggregated score so the analyst can know quickly that at least one of the signals related to a

particular drug pair is interesting as opposed to using an average of scores that might hide a highly scored signal by averaging it. Similarly, even if one of the multiple signals related to a drug pair is unknown, the edge is marked unknown (solid and thicker) to grasp the analyst's attention. This helps avoid missing the detection of novel signals. While it is possible to augment these views with more nuanced information, for example through glyphs or more complex color schemes, We instead emphasize visual cues based on drug analyst's reported work-flows.

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

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

- Table View: The simplest method to show rules or data with relationships is a tabular format [141], where each attribute can be a column and each row corresponds to a signal. TableLens [137] which is designed to allow users to detect patterns, correlations, and outliers in the data set using tables is suitable for presenting numerical data. However, most of our data is categorical. Also thousands of drugs and ADRs with many to many relationships form the signals. Thus a tabular format may be cognitively demanding and tedious for exploration. Moreover, a tabular format does not provide enough visual dimensions to encode AIM models as units (**R1, R4, R5**).

- Adjacency Matrix: An alternate design to visually encode drug interactions would be an adjacency matrix where row and column dimensions map to the drugs and each cell depicts a DDI [35, 178]. Adjacency matrices have two shortcomings. First, drugs might interact with other drugs as nearly disjoint sets, that is, each drug on the x-axis might have a different set of interacting drugs on y-axis. This would render the matrix mostly empty. Second, the AIM model requires several visual dimensions to encode information about signal interestingness and label status. With matrices, color alone did not appear to be sufficient to encode both (**R1, R4, R5**).

4.6.2 Signal Triage View

To align with the work-flow of the analysts, we aim to help them prioritize which drugs and interactions to analyze first from a pool of drugs assigned to them (**R2**). For this, we designed a Triage view using the small multiples technique with a small node-link diagram using the force-directed layout [170, 12, 139, 61].

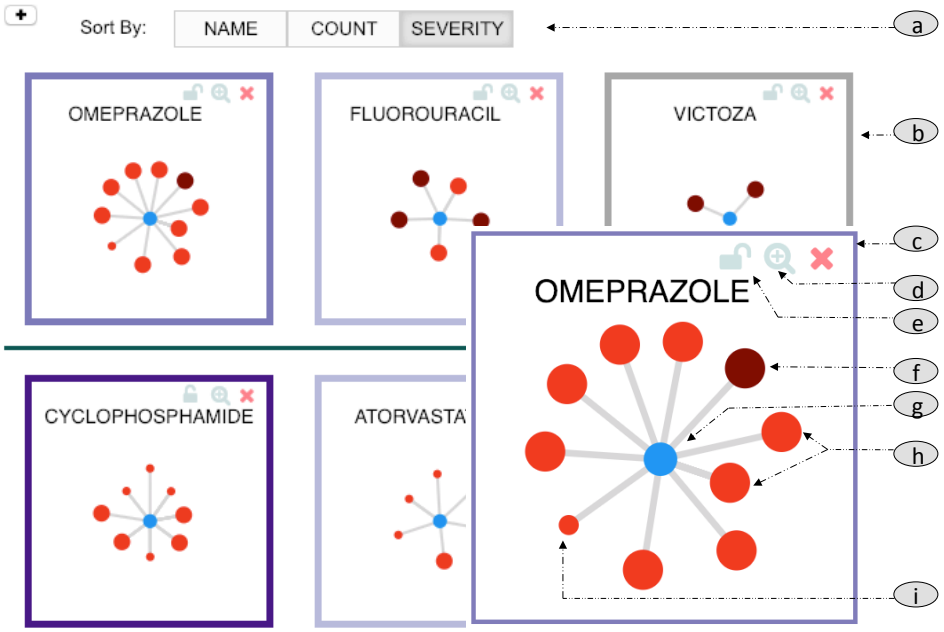


Figure 4.6: The Triage View - (a) Drug of interest Omeprazole. Interacting drugs with an (b) unknown and high score signal, (c) Unknown and low score signal, (d) Known and low score signal. (e) Pin a drug. (f) Zoom in to view the Profile of Omeprazole. (g) Color of box represents total number of severe reactions present in all signals related to Omeprazole. (h) Sort. (i) Pinned drugs.

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

The center node represents the drug of interest and nodes surrounding it depict all other drugs interacting with the drug of interest (Figure 4.6). At a glance, the analyst can get an overview of each of her assigned drugs. She can pick the drug with a comparatively larger number of DMEs, i.e., severe ADRs (**R4**). Or she can focus on the most interesting drug-drug interactions without any overwhelming details about the signal such as the related adverse reactions.

The option to pin a drug is provided to facilitate the analyst in prioritizing a drug to further explore it (**R5**). Pinning helps an analyst maintain context, so that they can resume their work where they left off in case they do not finish the review of a particular drug in one go. Similar to the *Screening view* (Section 4.6.1), if a drug-pair has many signals, the maximum interestingness score is mapped to the color of the nodes to facilitate attention.

The drug name represented by a node along with other information about the interaction such as the number of reports supporting the interaction, their ADRs, their scores etc. is revealed via a tool-tip upon hovering over a node. The Triage view and the Screening view represent information about AIMs differently and support different tasks. The Triage view helps an analyst prioritize drugs to be looked at and analyzed first (**R2**), while the Screening view gives them high level information on all interactions related to their assigned drugs to empower them to screen out the unimportant ones (**R1**).

Design Alternatives: Other design alternatives are possible for the Triage view such as a compact barchart or treemaps where the size of a chunk (in treemap) or

the height of a bar (in barchart) represents the interestingness score and color depicts the label status of a DDI. A graph view has two desirable properties. First, a Triage view based on node-link diagrams is consistent with other views [136]. Second, with a central drug, graphs can display a larger number of interacting drugs visually separable in a small space. This makes the comparison of different interactions easy as compared to a compact treemap or barchart.

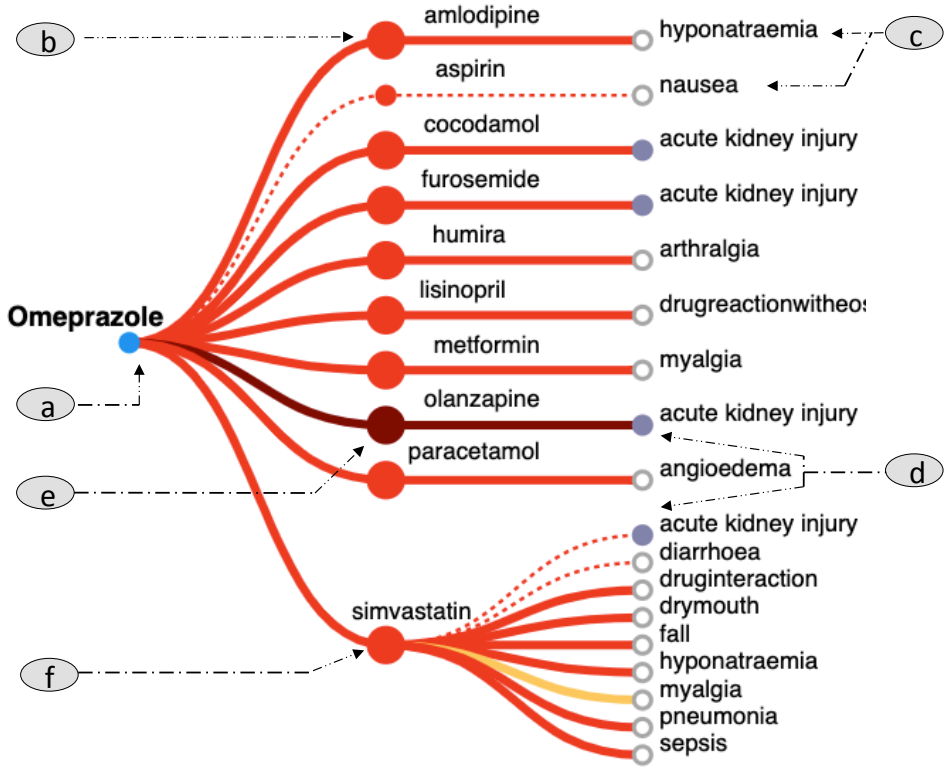


Figure 4.7: Signal Forensics View: a tree layout allows an analyst to view and analyze the whole Augmented Interaction Model of a drug. (a) The root node represents the drug of interest- Omeprazole. (b) The second level represents drugs interacting with Omeprazole. (c) The third level represents the reactions related to each DDI. (d) DDI leading to severe reactions - purple color. (e) shows a signal with highest score - dark color. (f) Link between Omeprazole-Simavastatin depicts the aggregated score and unknown status because some of the signals are unknown. Each path starting from root until leaf is a hypothesized signal and represents an association rule.

4.6.3 Signal Forensics View

Once analysts screen and prioritize the drug of interest by interacting with the Screening and Triage views, they can explore the signals further using this view by analyzing the related reactions (**R4, R5**). The Forensics view (Figure 4.7) depicts a full Augmented Interaction Model (AIM) for one particular drug of focus. It not only displays but also differentiates among the scored interactions of a drug and their corresponding reactions along with their severity. Given the views are linked, an analyst can interactively access the Forensics view of the corresponding drug through the zoom button or by clicking the central node in the Triage view. There is a wide range of potential visual designs to represent drug related reactions, both contain categorical

attributes. DIVA uses a modified version of tree diagram, which was found to be among the most appropriate alternative designs compared to adjacency matrices or parallel coordinates. While a tree encoding is more commonly adopted to visualize hierarchical data, they are a good fit to visualize AIMs due to the data structure and analyst's workflow patterns. A three-level tree layout represents the core attributes of AIM, i.e., the two interacting drugs and their respective reactions for every signal. Also, a drug interaction may be associated with multiple sets of reactions leading to different signals each with different attributes, i.e., interestingness scores and label status (**R2, R4**). Hence each signal can be visually encoded with a tree layout.

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

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

Node duplication can be used to reduce edge crossing [185]. Therefore, we keep the duplicate nodes in the third layer that represents similar ADRs in some cases. Moreover, to see the common ADRs of all DDIs of a drug, all similar ADRs within the view are highlighted upon hover. In addition, an analyst can annotate a signal as being a potential signal or a co-occurrence after analysis (**R7**).

Design Alternatives: One possible design alternative to display signals is parallel coordinates plots (PCP). PCPs have been used in the past to visualize association rules [77, 186]. A PCP with three axis could be used to visualize drug interaction signals, two for the interacting drugs and one for the associated reactions. However, this design choice has been found to not be appropriate for two reasons. One, PCPs work well only for a small number of items along the axis. Two, PCPS are good for global patterns but local patterns are difficult to see because of high clutter. For signals, it is difficult to relate each reaction set with its corresponding drug interaction because of the extensive edge crossings and overlaps. Hence, PCPs would fail to capture the detailed view of an AIM or individual signals (**R3, R4, R5**), which is indeed a core purpose of this view.

4.6.4 Reports View: Revealing Reports

As per (**R6**), analysts need to access the underlying reports as evidence when validating a signal. The analysts also requested to see the text narratives related to each case, as the narratives have richer information then the structured meta-data, a patient's medical history and the details of the adverse event potentially helpful in the evaluation of a particular signal. We thus design the Report View (Figure 4.4). This view provides a line listing of the cases (reports) related to a particular selected drug or drug interaction. The Report View is linked with all three views namely the Screening view, Triage view, and the Forensics view, to give the analyst direct access to the relevant reports supporting a signal. Similarly, selecting a case in the report view will highlight the corresponding drugs and ADRs in all views. The narrative section provides options to search for keywords in the text. Specific narratives are not shown in this work due to

data confidentiality as they contain sensitive information related to patients.

4.6.5 DIVA Capability for Signal Annotation

Based on the analysis, using the Forensics view an analyst can annotate a particular signal (**R7**) to belong to one of these two categories: (1) Continue Monitoring: if they need to further investigate it by monitoring it, (2) Co-occurrence: if drugs in the signal are appearing by chance and there is no logical reason for an interaction. This annotation is recorded by the system. It can subsequently be used by the *Signal Augmenter* to remove such drug interactions from the results of the mining process to facilitate future analysis.

4.6.6 Interactions on Linked Views

DIVA is designed to provide rapid exploration capabilities, at least compared to traditional workflows. All views are interlinked with each other, that is, all views are updated automatically as the selection is changed in any one view. Moreover, to give the analyst control, each view can be updated via the selection menus (Figure 4.1-d). For example, the sorting feature in the view allows an analyst to sort drugs either alphabetically, by the number of interactions they have, or the total number of DMEs (severe ADRs) present within the signals. Each view can be maximized and viewed independently. An option to select an alternate scoring criteria is also provided (Figure 4.1-d) for advanced users.

4.7 DIVA Evaluation

We evaluated the effectiveness and improvement opportunities in DIVA by conducting in-depth case studies and semi-structured interviews with domain experts who are drug safety analysts at the FDA. These experts also helped us in the iterative design of DIVA. After introducing the system to them, we observe them exploring the data in a think-aloud manner, and noting their feedback. During the interview, experts used the DIVA system on data from Quarter 4, 2014 (Oct-Dec, 2014) of FAERS. In total, MARAS [135] generated 1265 distinct ranked signals from this data.

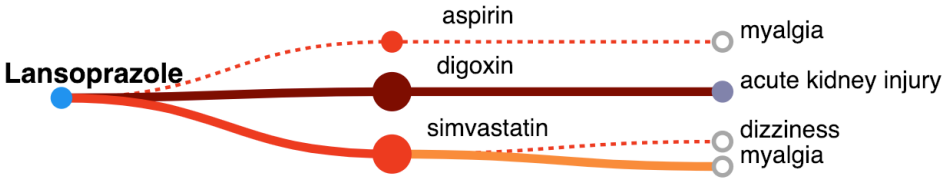


Figure 4.8: Forensics View for drug Lansoprazole. Interaction with Digoxin leading to acute kidney injury, a DME, is unknown, and is highly scored by the rule mining hence worthy of further investigation.

4.7.1 Case Studies

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

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

After looking at the assigned drugs from the Triage view sorted by the severity of reactions, he chooses "Lansoprazole" (Figure 4.8) to explore first (**R2**). He explains "First, it seems to have more DMEs compared to others. Second, it has a highly scored unknown interaction". He also mentions "it might be quicker to start the analysis with it as there are only three interactions to analyze". To view the adverse reactions (ADR), he clicks on the zoom button using the Forensics view (Figure 4.8). He then comments "Lansoprozole, a protein pump inhibitor usually interacts with Digoxin but the resulting ADR acute kidney injury, which is a DME, is not labeled yet" (**R4, R5**).

The analyst is now curious to see if this interaction leading to acute kidney injury indeed is a safety signal. He clicks on the interaction represented by link between Digoxin and Lansoprozole to view the reports and get the details of the cases that were used to extract the signal (**R6**). He starts to explore the relevant reports (Figure 4.4). He then comments "I see almost all of these reports also have another drug "Furosemide", which is used to treat kidney disorders". He further adds "the patients who were taking Lansoprozole were also taking Furosemide. That means, they might already be having a kidney disorder and Furosemide was prescribed to them for treatment. However, because of the DME we should keep an eye on it". He annotates the signal as "Continue Monitoring" (**R7**) and pins the drug for further investigation using the pin button on the Triage view.

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

When analyzing this first drug, "Lansoprazole" he does not find any threatening signals to further evaluate via reading the case narratives or by examining data from clinical trials. He then moves on to study his next assigned drug, as his job is to screen through all signals related to the drugs assigned to him. This time he screens "Cyclophosphamide" by examining the Triage view. He explains "though there are not many high scored unknown signals but the higher count of DMEs cannot be ignored". He then explores it further using the Forensics view (Figure 4.1-c). "I see all the interactions have the DME *neutropenia* listed as an ADR, which is a labeled ADR for Cyclophosphamide itself. They all have a similar color (low score)". He hovers over the edges and explains "The report count is same for all signals, they must have been extracted from same set of reports."

Then the analyst selects his next drug from the Triage View "Byetta" and explains "I noticed a highly scored and an unknown interaction with Victoza" (Figure 4.9). He points out "Both of these drugs are anti-diabetic and are used to control blood sugar level". He explains further "this cannot be an interaction but it must be a mere co-occurrence. The reason can be that the patient might have changed therapies during treatment and hence these two drugs were reported together". The analyst annotates the interaction as "co-occurrence" (**R7**) by right clicking on the edge. The analyst gives

#	Aim	Question
1	Visual Design	Is it easy/hard to read the Ego-centric view? Why?
2	Visual Design	Is it easy/hard to detect unknown interesting signals? Why?
3	Visual Design	Is it easy/hard to read the 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 signals? How?
6	General	Which part of the visual interface can be further improved in your opinion? How?

Table 4.1: Questions covered in a two hour interview with a group of drug safety analysts from the FDA

similar remarks for the interacting drug Januvia which is also an anti-diabetic drug from the same class.

As a next drug to examine the analyst selects the "Ondansetron" drug in the Triage view as it has a highly score interaction. Upon zooming in and viewing its details in the Forensics view (Figure 4.10), he observes that the ADR associated with the highly scored DDI is a DME "Serotonin Syndrome". He explains, "Ondansetron is used to treat vomiting and nausea caused by chemotherapy or radiations". He further comments "FDA added a warning in Ondansetron's label based on reports to avoid concomitant (at the same time) use of Lithium that might develop serotonin syndrome". Pointing to this known interaction, he adds "the fact that I can find the information about this signal being already labeled visually is very convenient and saves my time by not having to search for it in a separate tool" (**R3**).

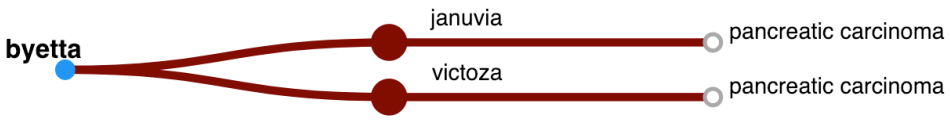


Figure 4.9: Forensics View for Byetta. Interaction with Victoza and Januvia is just a co-occurrence and not an interaction as all three of these drugs belong to same drug class that treats diabetes.

4.7.2 Interviews with Domain Experts

We interviewed multiple target domain experts to assess the effectiveness of DIVA and validate our design choices. Before presenting the prototype to a larger audience, we first invited two domain experts for the pilot study. The goal of the pilot study was to identify potential usability issues and to gather initial feedback on the workflow of the system. The two participants explored the system on their own after we had introduced the visual designs to them. One of the participants said "The Forensics view is very intuitive and easy to read, having the focused drug at first. Then we can see how this drug is interacting with other drugs and then we see the ADRs for each interaction. Following each path is easier to understand and the DMEs being highlighted make it very easy to grasp an interesting signal." They had a few suggestions too. At first, we had separate windows for the Forensics view. However, they suggested to keep

everything within one window and instead give the user control to choose a view to maximize or minimize. We added this capability to our system (Figure 4.1). They also suggested to make the report view available on demand only, i.e., whenever a user wants to see the relevant reports. Other minor suggestions included, highlighting the focus drug, and keeping the report view as simple as possible.

After the pilot study, we interviewed a larger group of 10 drug safety analysts to gain a more detailed assessment of the individual components of the system. These analysts were familiar with basic visualizations such as bar charts and pie charts. Our participants tried out the system themselves. These semi-structured interviews were guided by the questions provided in Table 4.1. We noted their feedback during the interview. Overall, the feedback was positive. Limitations in the current system were also collected.

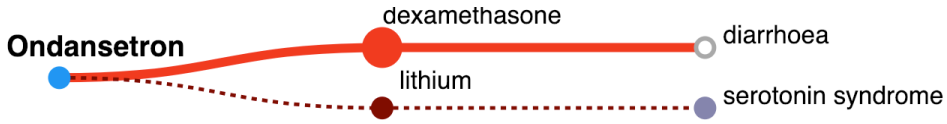


Figure 4.10: Forensics View for Ondansetron. Interaction with lithium leading to severe and rare ADR 'serotonin syndrome' has been labeled by the FDA recently.

The analysts' comments are summarized below:

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

4.8 Discussion

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

More broadly, the *scalability of visual analytics workflows* was one challenge encountered in the design and evaluation of DIVA. DIVA currently works with one year of data consisting of 1178 drugs (nodes) and 2763 candidate signals (edges). In theory, we find that DIVA can easily scale to data from several years. Any resulting clutter caused by an increase in the drugs can readily be addressed by existing scalable visualization techniques, such as edge bundling [84] for the node-link

diagram. However, it is less straightforward to scale the interplay between the drug analysts and the underlying data processing algorithms and corresponding workflow in DIVA. While there are a host of design models and activities for general visualization design, *i.e.*, Munzner's Nested Model [122], exactly how these activities and models map to visual analytics settings remains unclear, a gap which may be addressed in future work.

At a more practical level, our interviews with drug analysts revealed a need for the ability to incorporate additional domain knowledge. Information about known signals is available in drug package inserts (labels) and via online sources. In DIVA, data from one of these online sources (Drugs.com) is used. This dataset is known to be incomplete [154], and this gap was noticed by analysts during evaluation. To have a more complete list of known signals, integration of additional data sources such as DrugBank [1] is needed. More broadly, the dynamic nature of knowledge in the drug analysis space points to a future need for ways to incorporate dynamic heterogenous data, rather than just a static one-time data gathering phase as is current practice. Dynamic data integration, however, will raise new challenges for investigative analysis, such as when new information is learned that may inform prior (*i.e.* closed) investigations.

Beyond drug analysis, although DIVA is designed for a specific domain application, some components may be adapted to other domains that focus on low-level investigative analysis. For instance, similar reporting systems exist in other domains, such as the aviation industry, where the Federal Aviation Administration manages Service Difficulty Reporting system [7, 118] that collects reports about any malfunctions or defects in the planes. The idea of generating hypotheses about a faulty airplane using machine learning from a huge set of reports and providing means to explore and validate these hypotheses with interactive visualizations and domain expertise can be used in these application areas as well.

4.9 Conclusion

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

In the future, we plan to integrate additional knowledge sources into the mining process to provide more accurate information for the review process. We also intend to integrate analysts' feedback and interaction as annotation of signals in the mining process, so that signal generation can be improved. More broadly, we plan to explore visual analytics approaches for drug analysts' style of investigative analysis, which relies heavily on evidence collection from raw reports. Finally, to address the fact that drug interactions may impact sub-populations differently, we will incorporate

demographics in the visual analytics pipeline from signal generation to the visual representation.

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

Chapter 5

Visual Summary for Confirmatory Triage of Incident Reports

In this work, we focus on *incident reports triage* at the U.S. Food and Drug Administration (FDA)— the analysts' first encounter with hundreds of reports in their caseloads— where they review information within each report to determine if the report needs further investigation. Analysts must dedicate substantial time and attention to minimize the risk that life-threatening safety issues are missed amidst the influx of reports received on a daily basis.

Challenges in incident report triage are further complicated by the complexity of the possible underlying drug safety problems. For example, differences in patients' demographics, medical histories, and other drugs concurrently taken can make an outsize impact on the criticality of a given report, even among reports that focus on the same drugs and adverse reactions. Moreover, the criteria for report triage is not explicit, as it may vary from one incident to another. For instance, for some drugs, analysts may prioritize reports where a patient is hospitalized or deceased, while in the reports about drugs that are used to treat chronic diseases such as cancer, the death of the patient may not be as alarming to warrant escalation of the incident. Due to these complexities, reports triage currently requires a human to be in the loop to decide upon a report's relevance to a particular safety issue. Given that complete automation is not a viable approach at this time, reports triage emerges as a candidate for visualization, which seeks to encode data in ways that enable people to more rapidly and accurately process data.

The promise of visualization approaches for improving reports triage at the FDA is underscored by examining current practices. Currently, at the FDA, analysts use a traditional SQL-style table layout, which we will hereafter refer to as the Tabular layout, to summarize key information in the report text narratives, (see Figure 5.1). With the Tabular layout, however, analyzing key triage cues at a glance is difficult, and often requires scrolling, hindering efficient comprehensive analysis. After analyzing FDA analysts' behavior and process with this tool, we observe that one possible means for improving the reports triage process is to design an information-rich, compact visual summary of the same information. If individual incident reports can be made "glanceable" for triage analysis, a subsequent evaluation may reveal strengths and further opportunities to improve reports triage workflows.

In this work, we introduce SumRe, a glanceable visual summary for incident reports, designed to facilitate analysts reports triage workflows. We summarize

in-person interviews and a series of follow-up discussions with six (6) drug-safety analysts at the FDA to characterize the key data elements required for report triage, which we call *triage cues*, and develop a set of requirements. We then describe the design process of SumRe, which incorporates principles from information visualization [123] and glyph-based visual design [117], shaped by domain-specific workflow considerations to provide a compact overview of individual incident reports. Finally, we report the results of a controlled user-study with twenty (20) domain experts comparing SumRe to the Tabular baseline, which suggest that SumRe can aid in accurate identification of patterns and reports indicative of investigation. Moreover, results suggest that SumRe leads to better recall, qualitatively different insights, and participants reported an overall pleasant and enjoyable experience.

We summarize the main contributions of this work as follows:

- We contribute a task characterization of incident report triage following observations of 6 domain experts at work using think-aloud protocols and a series of followup discussions. These activities form the basis of a set of design requirements for a compact yet expressive design to support accurate and efficient *reports triage*.
- We design and develop a *glanceable visual summary*, SumRe, drawing on principles from information visualization and glyph-based visual design. SumRe facilitates tasks from domain-specific workflows using visual encodings such as spatial alignment, color, and word-scale icons with the goal of reducing analysis effort by supporting *glanceability* of a report.
- We develop an empirical study composed of triage-centered tasks to compare SumRe and the Tabular baseline. Results from 20 domain experts suggest SumRe facilitates accurate detection of important reports, as well as patterns across reports, with comparable performance across other key metrics. Overall, the promise of visualization for incident report triage is validated, and new opportunities for further improvements emerge and are discussed.

5.1 Background and Related Work

SumRe relates to previous work in three ways, including visualization for text documents and corpora; visualization for triage in domains such as cyber security; and efforts in improving drug-safety issue management. We briefly cover prior work in these areas, focusing on prior work which we draw from in the design, development, and evaluation of SumRe.

5.1.1 Visual Analysis of Text Documents

The vast majority of text visualization techniques designed for the exploration of a document corpus display metadata about the corpus, such as results of document clustering [44], topics [111], and name-entities [70]. Feature Lens [56] allows the visual exploration of frequent text patterns in text collections. TextTile [63] allows users to explore a set of documents by providing interaction operations and views. These approaches provide an aggregated summary of a set of documents using multiple linked views for exploratory analysis, our focus instead is to provide a glancable summary of each individual report for triage.

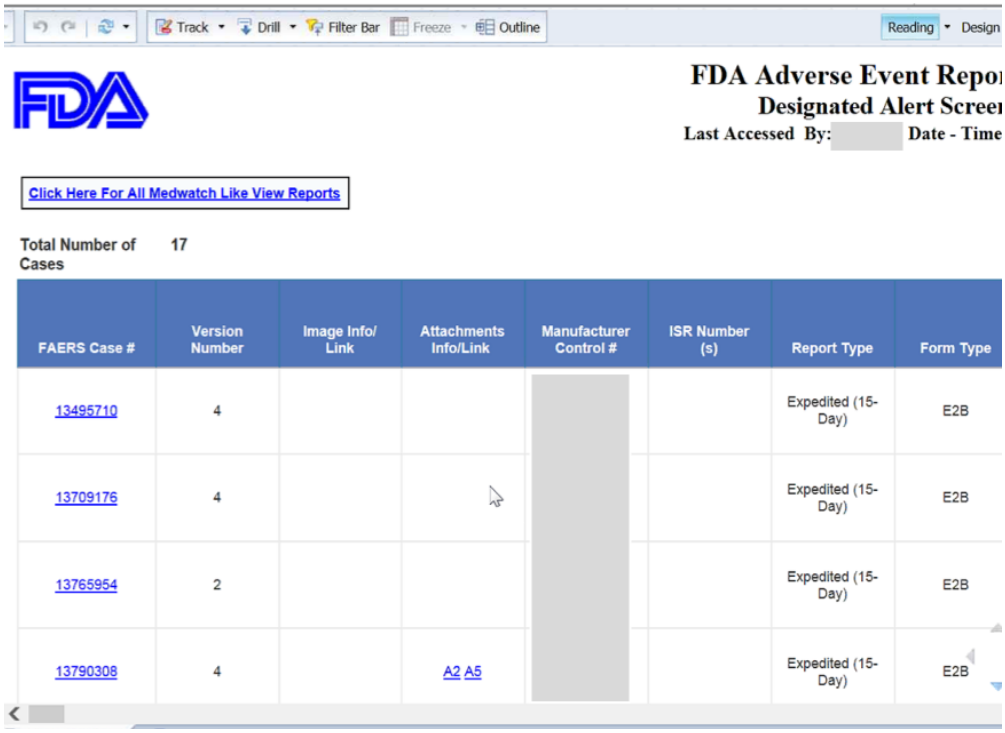


Figure 5.1: The current tool used for incident triage at the FDA uses a standard Tabular layout. Our aim in this work is to design and evaluate an alternative tool, SumRe, which draws on visualization principles to better encode and communicate critical safety report information.

Visualization techniques for a single document also exist. Word-clouds, also known as tag-clouds, have been widely used as an exploratory tool to provide a high-level overview of a single or multiple text documents [174, 105]. Docuburst [50] uses a space-filling approach to visualize document content by depicting relevant terms along with their semantic relationships. Word Trees [176], a graphical version of 'keyword-in-context', visualizes sentences sharing the same beginning in the form of a tree. Phrase Nets [172] uses a graph-based visualization to display relationships among words. Similarly, work on visual summaries of individual text documents also exists. For instance, Liu et al [109] visually summarize the emotions in a text document. In the literary domain, long texts such as books and novels have been visually summarized at multiple levels of abstraction to facilitate navigation within the document [102]. For example, Document Cards [166] provide a compact summary of a publication consisting tf-idf-based keywords and images from the publication to support the interactive exploration of the document.

These approaches are useful for exploratory analysis, that is, when users, not familiar with the data, aim to get an overview of the document. Incident reports, on the other hand, are problem-centric documents, where every report discusses an instance of a drug safety problem. In the later, analysts seek specific information to make sense of the severity of the report to rapidly assess if further in-depth investigation is warranted.

5.1.2 Triage in Other Domains

Document triage – a well-studied area in information retrieval and digital libraries – is the process of selecting a web page or document from the search results for further reading [115]. This is somewhat similar to incident reports triage. However, the

difference is that, in document triage, users stop searching for documents if they find what they are looking for. While, in case of incident reports, analysts have to triage *all reports* to maintain due diligence even if they find a report that is indicative of investigation. Second, in document triage users seek documents relevant to a known topic or search query, while in case of incident reports analysts seek reports that are indicative of unknown drug safety issues that have not been discovered yet. Third, for incident reports the criteria for report selection may change from incident to incident, while in document triage the criteria of a document selection depends upon a known topic or search query. This makes reports triage more complex and a viable candidate for visual analytics.

Existing research in document triage has focused on designing thumbnails [15] or image-based [41] web-page previews to help users select relevant documents. Other efforts have studied user behavior during the triage of research articles [18] such as skimming through the headings and titles. Trist [93], an information triage tool, provides an overview of large document corpora to help in document comparison and trend analysis. Approaches on designing visualizations to triage emails also exist [125]. TagSNet-Email [75] uses a network visualization to aid the triage of evidence during a forensics investigation. CueT [11], a network alarm triage tool, uses machine learning to prioritize network alarms to help operators quickly identify and fix them. While we share CueT's idea of helping analysts in identifying important reports, CueT is not designed nor evaluated for incident report triage.

5.1.3 Incident Report Analytics

Much of the existing work in Pharmacovigilance has focused on applying computational techniques such as natural language processing (NLP) on incident reports to extract name-entites from unstructured text [181, 98, 143]. Publicly available online tools such as OpenFDA [99] and OpenFDAVigil [26] help the general public explore incident reports and learn basic statistics about a certain drug or reaction. Other approaches have visualized relationships between drugs and their reported reactions [187, 92, 95]. These tools help in exploring the reports at an aggregate level; but they are not designed to visually summarize an individual report.

5.2 Understanding Incident Report Triage

Incident reports triage is the process where analysts review each individual report to assess the incident, that is, the association between the drug and reaction, and decide whether a report needs further investigation or not. More specifically, analysts review certain structured information that we call **triage cues** in each report and formulate a hypothesis as to whether the report is indicative of a potential issue and should be investigated. Investigation beyond this stage means that the analyst would need to read the text narrative in depth to investigate the details of the incident. While steps beyond triage are important, they are sufficiently complex to require attention beyond the scope of this work. The complexity of the post-triage process, combined with the fact that FDA analysts spend a substantial portion of their time on triage itself, leads us to constrain the scope of this work to focus on triage with its own distinct set of needs and challenges.

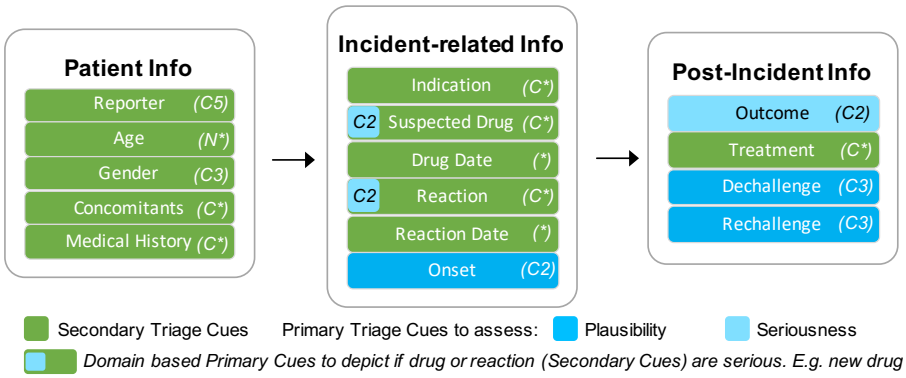


Figure 5.2: We find that one key part of incident report triage involves what we call *triage cues*. Analysts first seek primary cues about the patient, incident, and key events first to form an assessment of plausibility, severity, and likelihood of further investigation. Analysts then review the secondary cues that can confirm or reject their hypothesis. (X*n*) represents the attribute type (C= Categorical, N=Numeric) and it cardinality. E.g. (C*) means a high-dimensional categorical attribute with hundreds of categories. This characterization informs the visual encodings used in SumRe.

5.3 Triage Requirements Elicitation and Characterization

To better understand what requirements were necessary to aid analysts in the identification of reports indicative of investigation, we conducted one-hour in-person semi-structured interviews with six (6) drug safety analysts at the FDA. In these sessions, we observed the analysts performing their regular triage tasks, and asked follow-up questions to characterize their thought processes. In these efforts, our goal was to understand what parts of the report summary are attended to during triage, and how these parts are analyzed to render an assessment. We recorded and transcribed these sessions to facilitate the synthesis of requirements. As an intermediary goal, following these sessions we had additional remote discussions with these analysts to clarify and crystallize a model of their process and key requirements. The details of the minor requirements are reflected in the design process as discussed in Section 5.4.

5.3.1 Triage Cues: The Critical Information Used in Triage

One major requirement that emerged from interviews with FDA analysts is the relative weighting of report summary information, which we characterize as *triage cues*. In particular, analysts seek information about the patient, such as their demographics and history, the incident, such as the drug and reaction, and the events after the incident, such as, interventions taken to mitigate the incident. Figure 5.2 depicts common triage cues (data attributes) that are assessed during triage. The majority of these triage cues are multi-faceted, that is, they have many facets or categories, such as different drugs or reactions. For instance, Aspirin and Tylenol are two categories of the drug cue, while nausea and headache are categories of the reaction cue.

Primary Triage Cues for Hypothesis Formation

Here we describe how these triage cues are assessed during triage. Some triage cues help analysts to determine whether the report should be investigated or not. We call these cues **Primary Triage Cues** as analysts assess them first to know if the

incident is serious and/or plausible, depending upon the quantity and quality of useful information present in the report.

Primary Triage Cues for Seriousness. To assess seriousness, analysts look at the outcome of the report, where "serious" typically means a negative outcome, such as the patient dying or being hospitalized (Fig. 5.2). Analysts also prioritize a report if the drug is a new molecular entity, i.e., a new drug. This is because analysts closely monitor new drugs as they have not been in the market for a long time and chances of them causing undiscovered adverse reactions are high compared to those in long-term use. Analysts also look for the severity of the reported adverse reaction. For instance, a renal failure or seizure is a severe reaction and thus worth investigating as compared to a headache or nausea.

Primary Triage Cues for Plausibility. Analysts also form a hypothesis about the importance of a report by evaluating the plausibility of the incident by reviewing the 'Onset', 'Dechallenge', and 'Rechallenge' triage cues (Fig. 5.2). Onset or Time-to-Onset is the duration between the date when the drug was taken and the date when the reaction was observed. Onset helps analysts assess the possibility of the reaction being associated with the drug. Dechallenge and Rechallenge, on the other hand, are clinical actions taken by medical professionals to assess if the reaction is associated with the drug. Dechallenge implies that the reaction disappeared after stopping the drug, and rechallenge that the reaction recurred after restarting the drug.

Secondary Triage Cues to Support or Reject the Hypothesis

After forming a hypothesis about whether the association between the drug and reaction is plausible or if the report is serious and hence important by reviewing the Primary Triage Cues, analysts seek additional information to further support their initial hypothesis. We call these cues **Secondary Triage Cues**. For instance, the analyst may want to know if the drug is known to be causing the reaction, or patient's medical history may be the reason for the incident (reaction). Similarly, if a patient is taking multiple drugs simultaneously, those could be the cause of the reaction. These secondary cues support or reject the analyst's hypothesis about whether the report needs investigation or not.

Missing Information. Information including key triage cues can be missing in incident reports due to poor reporting – a well-known world-wide problem plaguing Pharmacovigilance [22]. Missing information may complicate the triage process because based on the type of missing triage cues, the assessment of a report can become challenging. However, if a report has many triage cues missing and there is less information to assess an incident, then analysts are more likely to quickly form a hypothesis that the report does not need investigation.

5.3.2 Challenges in Reports Triage

Based on our interviews (Sec. 5.3), we identified the following key challenges in the above triage process.

- **Comprehensive analysis of triage cues.** Analysts review multiple triage cues (Fig. 5.2) collectively to decide if a report needs further investigation. That is, while the primary cues help analysts in forming their preliminary hypothesis about the report's importance, the secondary triage cues then guide them in supporting or rejecting that hypothesis.

- **Identification of heterogeneous triage cues.** Analysts have to collectively analyze multiple triage cues which have different precedence in decision making. Currently, with the Tabular layout, all cues are displayed in the same textual manner, making it difficult to visually discriminate between primary and secondary cues.
- **Knowledge Externalization.** From our interviews we observed that during triage, analysts create pieces of evidence about each report being either indicative of investigation or not. Currently, analysts typically rely on their memory to store this knowledge as the system does not allow them to annotate a report to capture their analyses. This creates burden on the analysts who need to keep track of which reports have been triaged. This is particularly challenging during task switches which cause interruptions in the workflow.

To summarize, a reported incident does not mean that the reaction is actually associated with the suspected drug [66]. Instead, many factors could be associated with the reaction and need to be assessed. Moreover, lack of definite triage criteria along with missing information to assess the triage cues add to the challenges of sensemaking of these reports. These characteristics of incident report triage lay the groundwork for alternative designs and evaluation criteria. Henceforth, we aim to design a visual summary to make an individual incident report glanceable by highlighting the relevant critical information to effectively support this triage task.

Based on the above requirement analysis, we define as set of guidelines for the development of the design components of SumRe based on which the remainder of this work is based.

DG1 **Providing a compact view to facilitate comprehensive analysis.** As analysts need to glance over all the triage cues to make a decision, our goal is to provide a compact view of this expansive and informative data to make glancing feasible.

DG2 **Differentiating among diverse triage cues.** Analysts prioritize primary cues for forming a hypothesis about the report, and only thereafter tend to focus on secondary cues to seek supporting evidence for their hypothesis. Triage cues also include patient and incident related information. The design thus needs to support an ease in differenting between these classes of these triage cues.

DG3 **Facilitating capture of triage artifacts.** Our goal is to allow analysts capture their triage related comments and actions; with the aim to support them in keeping track of their analyses as well as facilitate information recall at a later stage.

5.4 Design of a Glanceable Visual Summary for Reports Triage

Following our collaboration with FDA analysts and initial task characterization activities, we developed an alternative summary method, SumRe. SumRe seeks to address the identified challenges in incident report triage, while drawing on visualization principles to effectively align with the identified primary/secondary cue workflow we observed from analysts. The final design as depicted in Fig. 5.3 is a result of multiple iterations and discussions with the domain experts. After initial prototype designs, we refined the aforementioned requirements by obtaining further details on how analysts process triage cues via discussions with the FDA experts. One outcome of this activity was the use of icons to summarize information. Further iterations explored alternatives related to the order and visual encodings of the triage cues.

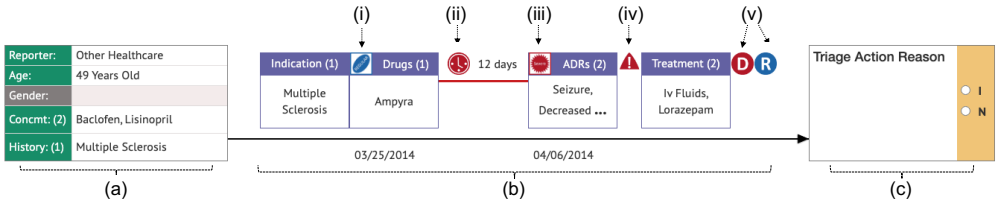


Figure 5.3: SumRe layout to triage incident reports. (a) Profile Panel presenting information about the patient and the reporter. (b) Incident Panel representing all the information associated with the reaction (incident). (c) Analysis Panel allows analysts to add comments and triage action (Investigate or Do Not Investigate). Icons: (i) Regular drug, (ii) Onset, (iii) Severe Reaction, (iv) Serious outcome, (v) Dechallenge and Rechallenge respectively. Missing information is represented with grey color. Arrows are added for illustration and are not part of the design.

5.4.1 Overall SumRe Layout

Analysts currently read reports from left to right. Therefore, in our design the information is also structured from left to the right. All triage cues are categorical with the exception for the age, drug and reaction dates. We used a primary visual channel—spatial position— to represent the sequence of the cues to follow the natural flow of the events. For instance, the summary can be read starting from the patient and reporter, to the details of the incident, to the events after the incident.

One example outcome of the design process comes from the placement of the indication component, which is the disease for which drug is taken or prescribed. During our discussions with the domain experts, they suggested to place indication next to the drug as it would help them know right away why the person took the drug. According to one expert, "Indication is important for our analysis, as most of the times one drug can be used to treat multiple symptoms. Such as, Propranolol is used to lower blood pressure but it is also prescribed to prevent migraines. Knowing the drug's association with the reaction for a certain indication is helpful". Below we discuss further components of the design.

Profile Panel. The Profile Panel (Fig .5.3a) displays the triage cues related to the patient. Basically it answers the questions "who did the incident happen to, and who reported it". These cues represent supplementary information and are not associated directly with the incident. Following expert feedback, we added the counts for the high-dimensional cues to allow analysts to assess the information on the fly. We were told that having many underlying conditions (history) and taking multiple drugs at a time hinders the analysis. According to one expert, "We are looking for confounders. e.g., If we have a report where patient has taken 10 drugs and has a reaction, that report is not gonna help us in assessing the incident as compared to a report where a patient has taken only one or maybe two drugs".

Incident Panel. The Incident Panel (Fig .5.3b) summarizes the information associated with the reaction from the time the drug was taken until the patient recovered from the reaction. The arrow on the timeline depicts the sequence of the events. The information on timeline reads like a story. That is, for a certain disease, the patient took the drug on a date and after [onset] days the patient experienced the reaction with a serious outcome [hospitalization/death] and the patient was then treated with [Treatment]. In some cases, the drug was dechallenged and rechallenged to assess its association with the reaction.

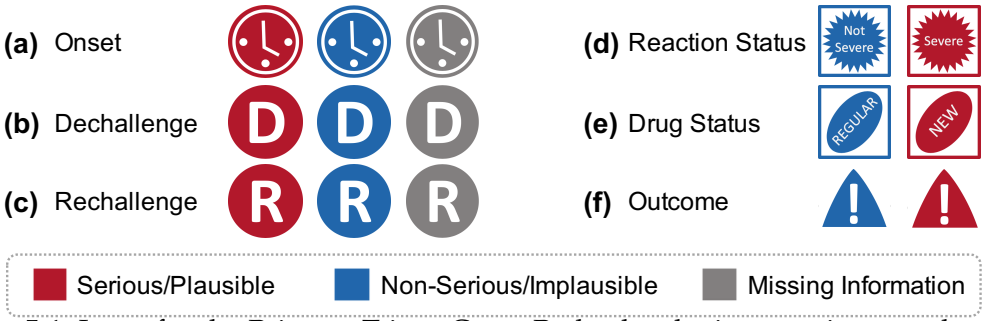


Figure 5.4: Icons for the Primary Triage Cues. Red color depicts a serious or plausible cue, blue color encodes non-serious or implausible cues, and grey represents missing information.

Analysis Panel. The Analysis Panel (Fig .5.3c) allows analysts to add their comments as well as triage action to a report summary to help them in distinguishing between triaged and untriaged reports as well as help them in reviewing their assessment about the report at a later time (**DG3**). The triage actions at this stage are if they will investigate (I) a report further or not (N). The Analysis Panel is placed at the end of the summary following the workflow of analysts, that is, they read a report from left-to-right and end by adding notes to the end of the report when writing a review.

5.4.2 Design Considerations for Triage Cues Depictions

We design icons for the primary cues using Borgo et al.'s [27] guidelines, used text, space, and color to represent rest of the triage cues (**DG2**). Redundant encodings are used to help analysts easily differentiate among various triage cues [28] and to facilitate multiple analysis stratgies. For instance, patient-related cues have the most left spatial position and a green color (Fig. 5.3).

Overall Visual Encoding for Icons. The primary triage cues contain 2-3 categories, at most. For instance, the primary cue *outcome* has two categories, *serious* and *non-serious*. Therefore, the visual channel with the strongest 'popout' effect to differentiate between these categories is color after position, as position has been used for the overall layout of SumRe. We use color hues to differentiate between categories of a primary triage cue. The colors are chosen based on the semantic meaning of the category. For instance, red represents information that is serious or plausible and needs attention, grey encodes missing information and blue represents non-serious or implausible. For consistency, similar color encoding is used across all icons. Similarly, the shape channel [117] is selected to differentiate among different primary triage cues. The details on selection of each shape is discussed below.

Designing Onset's Icon. As onset is the duration between the date when the reaction was observed and the date on which the drug was taken, we use a time symbol to represent it. Moreover, onset represents a connection between the drug and reaction so we represent the connection using a semantic encoding of link [117]. Due to poor quality of reporting, sometimes the onset is missing in the report or it may even be implausible, such as depicting that the reaction happened before the drug. Although only domain experts can verify the actual plausibility of the onset as it varies from drug to drug, we consider an onset only to be plausible if it is at least a positive value depicting that the reaction happened after taking the drug. According to one domain expert "For some drugs a certain reaction may appear the same day, while for others it may take months, so the onset really depends on the drug and reaction". Therefore,

we only represent if onset is reported or is missing and leave further assessment to the human experts.

Designing Dechallenge & Rechallenge's Icons. Dechallenge means the reaction's disappearance after stopping the drug (a positive fact), while rechallenge means the reaction's recurrence after restarting the drug (a negative fact). Our first designs for these elements included variations of metaphoric icons to represent 'stop' and 'restart' signs to represent dechallenge and rechallenge, respectively. Other design alternatives included using the alphabets 'D' and 'R' with and without background. All these designs were shown to the experts and multiple rounds of discussions led to the final design as depicted in Fig. 5.4 (b & c). The circles were added to the alphabet to improve their visibility in the presence of other triage cues in SumRe (Fig. 5.3)

Designing Icons for Drug and Reaction Status. Two of the primary triage cues correspond to domain knowledge about the reaction and drug – both critical for triage. For drug, the status represents if the drug is new or old (regular), and for reaction, status means if it is severe or non-severe (Fig. 5.4d,e). At the FDA, a drug is considered new for three years after its approval and being in the market. We considered various design options for the drug-status cue including pictures of medicine and different shapes of pills. For the reaction-status, metaphoric representations, such as, variation of 'star shapes' to depict issue or negative effect were preferred by experts as compared to the alphabetic representation (A for ADR).

After many design iterations, we selected the final design as depicted in Fig. 5.4(d,e) with the surrounding rectangle used to enhance their popout effect when displayed along-side the secondary cues (Fig. 5.3).

Designing Outcome's Icon. The outcome of a reaction has two values: serious and non-serious. We designed multiple variations of metaphoric symbols for 'danger' and 'alarming' signs using a variety of shapes (circular, triangular) as well as using alphabetical representations (S and NS). Based on the feedback by the experts, the triangle symbol (Fig. 5.4f) is selected as compared to a circle to avoid confusion with other circular icons, i.e., icons for onset, dechallenge and rechallenge.

Representing Missing Information. As missing information is key in report quality assessment, we use a grey color to represent missing information in SumRe for both primary and secondary cues. In informal pilot studies, we observed that participants had difficulty perceiving the amount of grey presented and were unconsciously biased towards the missing secondary cues as compared to primary cues, due to the small size of icons. We conducted followup pilots using different shades of grey for both primary and secondary cues. The final shade that best perceived is presented in Fig. 5.3.

5.5 Evaluation

SumRe is designed to support effective triage of incident reports by providing a glanceable summary of each individual report. To evaluate the design of SumRe, we have conducted a controlled user study to assess its usefulness in the reports triage in comparison with the current techniques used for processing these reports at the FDA.

5.5.1 Overall Study Design

Participants.

We recruited 20 medical experts (5 males, 10 females, 5 unspecified) using the following pre-screen criteria from Prolific [131]. We used Prolific's filter criteria to recruit participants who identify as Doctors, Nurses, Pharmacists, and Emergency medical workers. Recruited participants were largely familiar with drugs and their related adverse reactions, particularly judging from their free-response answers. Based on completion times (around 55 minutes) in pilot experiments, each participant was paid $10.00 in order to exceed US Minimum Wage. All participants viewed an IRB-approved consent form.

Study Design

The experiment followed a within-subjects design where each participant performed the tasks using both conditions (layouts) to minimize the effects of participant's expertise and domain knowledge. Both conditions and datasets were balanced across participants.

Baseline Condition.

Analysts at the FDA currently use a Tabular layout (Fig. 5.1) to skim the structured information of a report to make a decision on whether to dive into reading the actual report and more deeply investigate the narrative. We use a Tabular Layout similar to (Fig. 5.1) as a baseline to reflect this current triage workflow. The order of the data elements in the Tabular layout follows the same order used in current tool at the FDA for triage (Fig. 5.1). We added two columns appended to the end of the Table, one for adding comments and other for adding the proposed triage action. However, in practice, the current Tabular layout at the FDA (Fig. 5.1) does not capture any annotations and it is read only. For consistency and sake of a fair comparison, we present similar information in both the Tabular layout and SumRe.

Data Set

For our datasets we used FDA Adverse Event Reports from 2014-2019 [65]. As the study was within-subjects, we curated two data sets, each containing a total of eight reports. For both datasets, six of the reports were about one drug, 'Ampyra' and 'Harvoni' respectively, while two additional reports were about other drugs. This was done to reflect the domain workflow, where analysts may come across important reports not related to the drugs they are responsible for, yet their identification is crucial. Thus, we added two additional reports to evaluate the analysts' capability of "serendipitous" discovery following a similar task used in network security analysis [169].

To reflect the real ratio of reports with incomplete to complete information received by the FDA, half of the reports were chosen with missing information, while the other half had complete information. We consider a report with more than 80% of triage cues as complete [22], and others as incomplete. We selected one report as indicative of investigation in each of these complete and incomplete subsets (3 reports each), verified

by the domain experts. In practice, the ratio of complete to incomplete and important to non-important is small, but for study purposes we kept it balanced.

We modified the reports in several ways to keep the two sets as similar to each other as possible to avoid bias. This included simplifying very large medical histories, and removing personal patient information from the histories. For the complete reports in both datasets, we manually extracted information about medical history, treatment, dechallenge and rechallenge which is not present in the structured format.

Procedure of the Study

After completing the consent form, participants were presented with two video tutorials; one for demonstrating the study task and other for the layouts, followed by a guided tour of both layouts. Participants were allowed time to practice with each of the layouts. They were given two mock tasks with multiple-choice questions to ensure their understanding of the layouts and tasks. After accurate completion of the mock task with as many attempts as possible, participants started the study with the Triage task followed by a set of pattern detection and exploration questions. After completing the Triage task twice each with one of the two layouts, participants were asked to fill out a demographic questionnaire and qualitative survey to provide feedback on both layouts as well as their preferred layout. Thereafter, participants performed the Recall task. A help reference for the concepts used in the study such as factors that make a report "indicative" or "not indicative" of investigation was provided in each layout.

Study Tasks

We designed the study tasks to be reflective of the triage tasks performed by the analysts at the FDA. That is, analysts review sets of reports, one by one, examining the reported information to assess if it requires further investigation.

Triage Task. The goal of the triage task was to assess the participants' performance in discerning the gist of a given set of reports with both layouts. Participants were asked to put themselves into the role of a drug safety analyst who needs to analyze each report and decide whether the report demands further investigation by taking the respective triage action.

Once the triage action was captured for all reports, participants were asked questions (Table. 5.1) to assess their performance in accurately assessing reports. For the pattern detection questions,

we ask questions for participants to search and interpret information from reports following the tasks for overviews for text analysis [85].

Recall Task. Recall is important in reports triage as analysts review reports regularly and come across safety issues that they may have seen earlier. This could further trigger them to raise alarm and investigate more deeply if they were to encounter the same issue a second time. Therefore, we include a recall task to assess the effectiveness of SumRe. For the recall task, we follow the tasks designed by Bateman et al. [20]. Participants were not told about this task during the instructions in the beginning of the study to prevent intentional learning. After completing the triage task for both layouts, participants filled the demographic and qualitative survey to clear their visual and linguistic memory before starting the recall task. For each layout, we presented equally blurred summaries of two reports to the participant, including some that were indicative and others they were non-indicative of investigation from

No	Tasks	Description
Q1	Triage	Select triage action by identifying reports that are or are not indicative of investigation.
Q2	Overview	Identify reports that are the most worthy of investigation.
Q3	Overview	Identify reports with the most complete information.
Q4	Pattern Detection	What happened to the patient after the serious outcome in the report with [X] search criteria?
Q5	Pattern Detection	What common pattern do you observe about [X] cues in reports with [Y] search criteria?
Q6	Pattern Detection	Contrast report [X] and report [Y].
Q7	Open-ended Exploration	Explore the reports freely and report on your findings. Is there anything surprising or interesting?

Table 5.1: Task Questions and Description

the triage task. The order of the layouts and datasets followed the same order from the triage task to ensure similar duration between recall and the triage task. For each layout, participants were asked to recall and report as much information about the two reports as possible.

Measurements

We collect a set of both *qualitative and quantitative measures* throughout the study. For each trial, we capture the *start* and *end time*, this allows us to evaluate the average time spent on tasks and training for each condition. During the task, we measure time spent on triaging the reports and answering the questions, submitting the answers, participants' confidence in the submitted answer, and the perceived difficulty of the task, all on a 7-point Likert scale. Through free-response questions, we also collect qualitative answers and feedback for each task which help us assess participants expressiveness and reasoning about the report information. We also collect demographics and free-response feedback on the overall summary and tasks design. Following [144], we also asked participants to select their preferred layout for future tasks and provide a reason for their preference.

When calculating correctness, we use non-binary rules that map to a 0–1 scale, corresponding to key "parts" of an answer. For example, we give 0.5 points for an answer that contains only the severity of the reaction, if the task asked for identifying severity and count. Additional details on the scoring method for each task in the supplementary material.

Pilots, Analysis, and Experiment Planning

We conducted several pilot studies to evaluate system usability, data collection, tasks, measures and clarity of our procedure. Due to the limitations of null hypothesis significance testing, we base our analysis on best practices for fair statistical communication in HCI [57] by reporting confidence intervals and effect sizes. We compute 95% bootstrapped confidence intervals [53] and effect sizes using Cohen's d to indicate a standardized difference between two means. For each task, we display the accuracy and time results in the form of CI, along with p-values from non-parametric

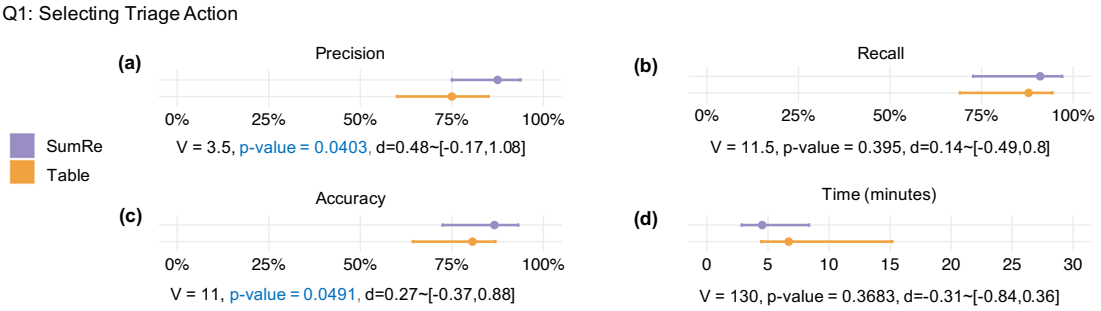


Figure 5.5: Triage task to select an action (Investigate or not) on each individual report. SumRe has significantly high precision, i.e., correct selection of reports indicative of investigation. Recall is the measure of correct selection of all the reports.

Wilcox Signed-Rank Test (given the sample size of ¡=20) in our figures to supplement the analysis.

Hypotheses

We developed a set of hypotheses to assess how the two summary layouts would compare for different types of tasks. We present the hypotheses below and later use them to frame and discuss our results.

- **H1 Efficient Reports Triage.** Analysts should be able to triage reports quickly and accurately with SumRe due to the compact layout and the visual encodings designed to highlight and differentiate between different triage cues.
- **H2 Report Quality Assessment.** Analysts should be able to identify reports with more missing information accurately with SumRe due to the visual encodings.
- **H3 Accurate Pattern Detection.** Analysts should perform better in identifying common patterns among parts of reports (patient and incident) with SumRe due to the spatial alignment of patient and incident related information in SumRe.
- **H4 Insight Generation.** When freely exploring the reports between the SumRe and Tabular layouts, the different structures and emphasis of these summary designs may lead to different types of insights.
- **H5 Serendipitous Discovery.** Analysts will be able to identify unrelated safety issues with SumRe due to the visual encodings to highlight primary triage cues.
- **H6 Better Information Recall.** Analysts would be able to recall more items with SumRe due to the memorable nature of glyphs and visual cues.
- **H7 Triage User Experience.** Overall participants will be report a positive experience when completing the tasks with SumRe.

5.5.2 Results

We report on the results of the study conducted with 20 participants (in Prolific, an additional 14 started but returned the study before completion). We group the results based on our hypotheses. While recall expresses the ability to find all relevant instances in a dataset, precision expresses the proportion of the data points our model says was relevant actually were relevant.

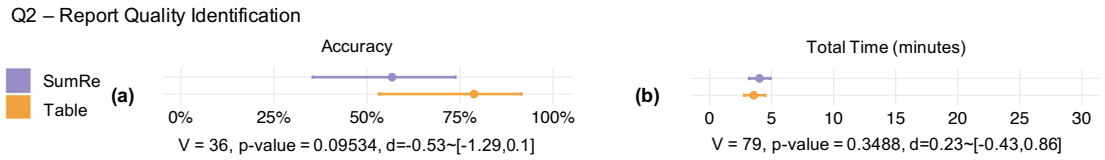


Figure 5.6: Overall accuracy and time for identifying if reports are complete. Although insignificant, but the Tabular layout outperforms SumRe.

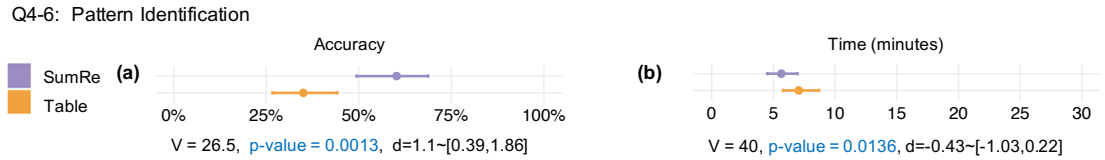


Figure 5.7: Overall accuracy and time for pattern detection questions (Q4-Q6). SumRe significantly outperforms the Tabular layout in both accuracy and speed.

Reports Triage.

Shown in Figure 5.5.c, participants had a higher accuracy in selecting the triage action (Q1) with SumRe ($M = 0.87\ [0.72, 0.93], d = 0.27\ [0.11, 1.33]$) as compared to the Tabular layout ($M = 0.81\ [0.64, 0.87]$). Qualitative analysis of the triage action indicates that participants were able to correctly identify the reports indicative of investigation, as verified with the precision measure (Fig. 5.5.a), SumRe ($M = 0.88\ CI[0.74, 0.94]$), the Tabular layout ($M = 0.75\ CI[0.59, 0.86], d = 0.48\ [-0.17, 1.08]$). This is crucial for triage to not miss important reports and a visual summary is likely to help in correct identification of such reports. Although, for the reports not indicative of investigation, the spread in mistakes was similar with both layouts.

There was little difference in participants' time spent on taking the triage action SumRe ($M = 4.5\ minutes\ [2.8, 8.3], d = -0.31\ [-0.84, 0.36]$) compared to the Tabular layout ($M = 6.7\ minutes\ [4.5, 15.1]$).

One interesting observation we made is that the reference (help) for triage criteria provided to the participants to help them assess the report's importance based on some basic guidelines (primary triage cues, quality of information, etc.) was accessed much more frequently under baseline layout ($M = 1.9[0.9, 2.9]$) as compared to SumRe ($M = 0.8\ [0.3, 1.3], d = -0.6\ [-1.18, 0.03]$), as depicted in Figure 5.8b. This means that with the Tabular layout, participants tend to need to refresh their understanding of a concept on average at least once more than SumRe.

This partially supports (H1). While triage with visual layout is not more efficient in terms of time, it is comparatively accurate, particularly in identifying reports that are indicative of investigation.

Identification of Report Quality.

For this task, participants spent relatively more time on questions with SumRe ($M = 4.02\ [3.2, 4.9]$) as compared to the Tabular layout ($M = 3.56\ [2.79, 4.48]$). Participants also had comparatively better accuracy in identifying the report quality with the Tabular layout ($M = 0.79\ [0.53, 0.91]$) than SumRe ($M = 0.57\ [0.35, 0.74]$). Both of these differences are highly insignificant (Fig. 5.6). Qualitative analysis of selecting the wrong reports as complete indicates that in SumRe participants were not able to

accurately detect the missing information for the patient related triage cues and the dechallenge and rechallenge icons. This might be due to the variable shapes and sizes of these cues making it difficult to assess which information is missing or present, as compared to the Tabular layout, where all triage cues look similar and missing information is represented as an empty cell. We note that this completely fails our hypothesis (H2).

For Q3 to select reports most worthy of investigation, there were also insignificant differences in the accuracy with the Tabular layout ($M = 0.85\ [0.6, 0.95], d = 0.08\ [-0.55, 0.68]$) and SumRe ($M = 0.88\ [0.62, 0.95]$). This could be because, participants had better understanding of the reports after taking the triage action and assessing report's quality (Q1 & Q2).

Pattern Identification.

The overall accuracy and time for all three tasks (Q4-Q6) are shown in Figure 3. There was a significant difference in overall accuracy between SumRe ($M = 0.58\ [0.48, 0.68], d = 1.1\ [0.39, 1.86]$) and the Tabular layout ($M = 0.34\ [0.25, 0.43]$). Participants also took less time with SumRe ($M = 5.6\ [4.5, 6.95], d = -0.43\ [-1.03, 0.22]$) to identify these patterns as compared to the Tabular layout ($M = 7.05\ [5.8, 8.7]$). This supports our hypothesis (H3) that visual cues help in quickly noticing commonality and important information across reports.

For the comparison task (Q6), there was no significant difference between the two layouts, that is SumRe ($0.53[0.44, 0.68], p - value = 0.3261, d = 0.29\ [-0.41, 0.86]$), and the Tabular layout ($M = 0.47\ [0.38, 0.55]$). Qualitative analysis showed that participants were able to easily compare data across reports with the Tabular layout due to the visual proximity and simplicity. We also noticed that with SumRe, participants missed the comparison between the patient information, while with the Tabular layout, information in the columns towards the end of the table such as dechallenge, rechallenge were missed. This is because, most of the participants did not scroll through the table to compare the reports. In the future, interactions such as grouping, sorting and highlighting could be added to improve this with SumRe.

Insight Generation.

Q7 instructed participants to freely explore the data and report on any insights they derived from their exploration. To analyze these responses, we performed qualitative coding of the responses following the guidelines provided in [146]. We consider one fact or observation about the data as an insight and do not count general comments toward insights, such as, 'hard to read information in this layout'.

For SumRe we received a total of 24 insights of which 21 were distinct, while for the Tabular layout, 23 insights of which 15 were distinct. We consider insights duplicate if two insights are discussing the same facts, for instance, "I was shocked at the lack of useful information in some", and "There is a lot of missing information". We categorized insights into a set of codes that were derived by an initial open coding of the data.

Two types of insights were markedly more common in the SumRe layout: **Report-level**, and **Unexpected**. We categorize an insight as **Unexpected** if it has not been part of the answers in the previous tasks and is an observation about the data, such as, *"Two of the patients had mental problems and they may have taken the drug*

incorrectly", and *"I'm sure that there are other things to mention but the one that immediately popped up was in R5, Cardiac Arrest is described as non severe ADR? That's odd! "*. In the first insight, the participant is reasoning about patient's capability of administering the drug correctly, while in the second, the reaction status of cardiac arrest is being questioned. FDA does not consider cardiac arrest a severe reaction due to its high background rate making the assessment of its association with the drug difficult.

Report-level insights were observations about multiple triage cues in a single or multiple reports. For instance, *"The reports submitted by pharmacists are surprisingly undetailed, they lack further information about the patient and nature of the ADRs which would be helpful in the triage process. Being a pharmacist myself I thought that the reports submitted by pharmacists would at least contain more information about the patient's concomitant medications"*, and *"It was interesting that so many of the very long adverse reactions reported had an implausible onset"*. These insights are considering multiple attributes, such as patient information and reporter in the first example, and reactions and onsets in the second.

For the Tabular layout, the most common insights were categorized as **Attribute-Level** and **Guided**. Attribute-level insights are those that are focused on a single triage cue or attribute. For instance, *"5 reporters were other healthcare"*, and *"There appears to be more men than women taking part"*. The first example is focused on the frequency of a certain reporter type (Other Healthcare), and the second is about the gender.

On the other hand, **Guided** insights were those that have been observed during completing the previous tasks. Such as, *"Also the information in some reports was not much and it did not help me and that was a big problem for me"*, and *"The most plausible adverse reactions to Harvoni were seizure, hypotension, and acute renal failure so pretty serious"*. The first insight is pointing towards the quality of information, one of the questions from the Overview task. The second insight indicates the severity of the reactions which was part of the triage criteria. Although, both layouts had these categories, their frequencies varied with each layout. Here we reported the more prominent ones.

We also saw twice as many insights about missing information in the Tabular layout as compared to SumRe. This also confirms that participant's observe missing information better with the Tabular layout. Another reason could be that in the Tabular layout, missing information is the only visual cue, making it easily detectable.

Overall, we were impressed by the extent and the quality of the insights, the engagement of the participants, and the ability of both layouts to reveal insights of various types, albeit with different frequency. A possible explanation for having more attribute-level insights with the Tabular layout could be that column data can be viewed and compared easily in the Tabular layout due to visual proximity. On the other hand, SumRe is designed to make all the report data accessible, resulting in more Report-level insights with SumRe. Consequently, we consider hypothesis H4 to be supported.

Serendipitous Discovery.

Shown in the Figure 5.8a, participants tend to identify more unrelated safety issues with SumRe ($M = 0.7$ $[0.35, 0.95], d = 0.61$ $[-0.13, 1.25]$) as compared to the Tabular layout ($M = 0.3$ $[0.05, 0.57]$). Although the difference is slightly insignificant, but with more data the results can be clearer. Serendipitous discovery is important in reports triage to ensure safety issues are not missed, even if it is unrelated to the drugs

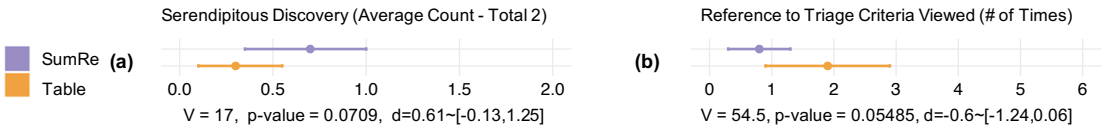


Figure 5.8: (a). Average count of identified unrelated issues. (b). Average number of times when the reference for triage criteria was viewed

analysts are monitoring currently. Overall, with the Tabular layout, 5 participants were able to identify unexpected issues as opposed to SumRe with 11 participants. Qualitative analysis indicates that participants were able to easily identify the primary triage cues despite of missing information in these reports. On the other hand, with the Tabular layout primary cues are not prominent and hence participants mostly focused on the missing information in these reports considering them as non-indicative of investigation. This supports our hypothesis (H5) that visual layout will lead to more discoveries.

Information Retention (Recall).

For the recall task, 4 out of 20 participants were able to recall information with the Tabular layout, resulting in a total of 6 triage cues, including 1 primary and 5 secondary cues. With SumRe, 12 participants were able to recall a total of 77 triage cues consisting of 70 primary and 7 secondary cues (Fig. 5.9a). Responses were scored based on specific answers. Because of the blurred images, the report text was not readable. However, with SumRe, due to the redundant encodings, that is, spatial alignment as well as use of other visual channels, the triage cues particularly icons for primary cues could be interpreted, resulting in a high count of information retention. The high recall due to redundant encodings aligns the findings from [28].

Preference and Qualitative Feedback

For participant's preference we followed the task from [145] and after the completion of Triage task, asked participants to select their preferred layout if they were to perform another triage task. 90% of the participants preferred to use SumRe layout for Triage tasks, considering it faster, easier, appealing, engaging and enjoyable (Fig. 5.9.b). This is also verified by significant differences in participant's overall perceived ease and confidence ratings for all the tasks (Fig. 5.10). Some of comments by participants in the favor of Visual layout included *"The visual layout was more appealing to use and I enjoyed the work more, it felt less like monotonous work and more like a pattern recognition game almost. It was more user friendly"*, *"The visual layout made it significantly easier and faster to perform the tasks. Moreover, it was less straining for my eyes compared to the table"*. For the Tabular layout, participants' remarks included *"For me Table was easy to use and see the ADRs"*.

5.6 Discussion

Overall, our results show that a visual summary significantly outperforms the Tabular layout in detecting patterns across reports and identifying reports that are indicative of investigation, which is crucial for not missing critical issues. The results of our

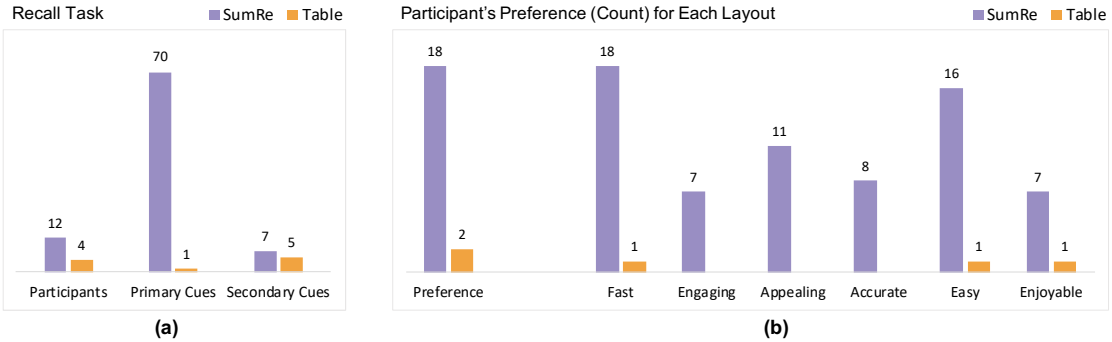


Figure 5.9: (a) Count of participants able to recall information along-with type of information recalled. (b) Participants' preference (count) for using the layouts for reports triage (left), along with the reason for selecting their preference (right).

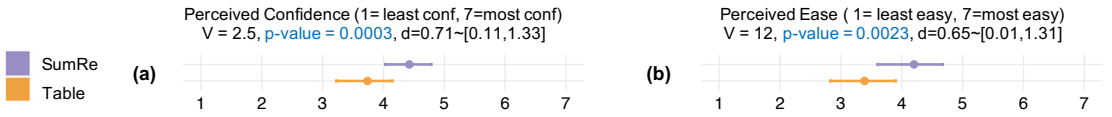


Figure 5.10: Average ratings for participant's perceived confidence and ease on completing the tasks.

experiment generally support our hypothesis that a visual summary can improve triage experience. Our goal now is to discuss the implications of these findings more broadly and make recommendations for the design and evaluation of visual summaries for triage.

5.6.1 Triage Design Shapes Analysis, Exploration, and Insights

In our study, participants were more engaged and expressive with the visual layout as compared to the Tabular layout when we look at the qualitative feedback. For instance, participants left a total of 23 voluntary comments (avg length 37.5 characters) with the Tabular layout, as compared to 36 with visual (avg length 47.2 characters). Our analysis showed that the order in which layouts were randomly displayed, did not have an effect on their commenting. We observed similar behavior in our pilots as well. The question remains whether this effect is true or is because of the size of the comment box which is broader in SumRe as compared to the Tabular layout.

We also observed that participants are able to correct their triage actions later in the study due to the visual feedback, that is, color for Investigate and Do not Investigate actions (Fig. 4.1), while we did not see this behavior with the Tabular layout. This was an expected outcome of the design of our Analyses panel (Section. 5.4), which we did not evaluate exclusively. This could be useful in the real-world scenario to re-assess one's analyses and correct their mistakes if needed.

We also noticed that the insights provided by participants while using SumRe were deeper, that is, reasoning about the data beyond the provided facts. This also supports previous findings on visualization resulting in insights [42]. It would be interesting to investigate what actually caused this behavior. Could it be that with scrolling the Tabular layout left and right, the participants felt drained and could not think beyond the provided facts?

5.6.2 Design Implications for Triage Workflows

Our visual design for SumRe is an initial step towards introducing visualization for the lowest-level analysis of textual data in critical workflows. One of the challenges we faced during the design of SumRe was visualizing the missing information, that is, there was a trade-off between selecting visual encodings that were consistent but less prominent vs. inconsistent but with larger pop-out effect. For instance, using a consistent color for missing information across the layout was hiding the Dechallenge and Rechallenge icons that were comparatively smaller than the boxes in the Incident and Patient Panel. On the other hand, using different shades of grey based on the size of the box displaying the triage cue was creating inconsistency in the visual encodings hence creating confusion.

We observed this during our studies as well. Participants were able to recognize missing information accurately with the Tabular layout as compared to Visual, although the difference was insignificant. Similarly, with the Tabular layout many of the participants' insights were about missing information. One possible solution could be using a visual indicator (count or a bar) for the amount of missing information, or adding interactions to highlight all the missing information. While designing for uncertainty is widely studied in visualization community, exploring design options in triage contexts involving textual information can be an interesting direction to follow.

Other suggestions involve the incident-panel with the timeline which currently displays cumulative information about the drugs and onset, to keep it as similar to the Tabular layout as possible. However, more sophisticated designs to display multiple onsets for multiple drugs while keeping the design compact are possible. For instance, one approach could be adding interactions and overlays with further detailed information on top of SumRe. This would need further investigation to weigh it's benefits.

5.6.3 Limitations

One main challenge was to design the tasks for the report-level triage of incident reports due to lack of precedent to draw from. Our data was mostly textual, however, there are few in-depth evaluations of the proposed approaches in the text visualization community [8]. Moreover, existing work with quantitative studies do not fit our single-report analysis. Due to lack of work on designing and evaluating visualization for document-level triage, we pulled tasks from many areas including cybersecurity [169], digital libraries [18, 115], overviews [85], and visualization [20, 145] and aligning them with the workflow of analysts to keep it practical.

Otherwise, our study of evaluating a visual summary for triage was within a limited scope, suggesting a need for future studies, in at least two ways: (1) participants' background, and (2) interactions. First, the participants' for our study were medical experts and hence closer to domain experts, however, they were not domain experts and had to be made aware of drug safety practices and triage tasks. Second, we did not add any interactions to the designs and compared the static layouts. Adding interactions and comparing layouts would be an interesting direction to investigate in future.

Another consideration is generalizability. We designed SumRe based on the Pharmacovigilance workflows at the FDA, however, Pharmacovigilance is a large domain world-wide [90] collecting incident reports with similar quality issues and data

elements [22], and our designed solutions can be a baseline to improve their systems.

5.7 Conclusion And Future Work

We present a design study to visualize this data on a report-level for efficient triage of individual reports. Our current design for the glanceable visual summary displays the tabular information in a compact visual fashion using word-size icons and visual encodings. We also present a quantitative study that compares and evaluates the impact of such a visual summary on participants' experience and performance when performing triage tasks. The results of our study suggests that a visual summary can be helpful in accurate identification of both important reports and patterns across reports, while concurrently providing a more engaging experience to the analysts.

In the future, we plan on testing variants of the Tabular layout with additional visualizations that may align with aspects of the broader Pharmacogivilance workflow. For example, we plan to include population statistics across weekly reports as part of the triage cues, such as medical history and drugs involved. Additionally, future work must examine how to best combine an effective summary with a similarly effective underlying text narrative interface, a concrete step towards better incident report triage management.

Chapter 6

Interactive Case Building and Management for Incidents of Concern

This chapter describes our design study of ConText, a investigation tool for the incident reports that support analysts in interactively building and monitoring cases by identifying and collecting supporting evidence. We address the challenges we observed in the analysis of these reports by first defining an Incident Investigative Analysis (IIA) model. We then design an interactive visualization system to realize the IIA model. We will evaluate the system with domain experts via case-studies and interviews.

6.1 Introduction

Analysts investigate if the incident indeed warrants an action after finding a report that helps in forming a hypothesis about a potential incident. Analysts analyze and combine information about a potential issue from several such reports to elevate reports that serve as relevant evidence in support of taking a corrective action. We refer to these activities analysts perform in such critical text-based reporting systems as *Instance-based Incident Analysis* (IIA). A common goal for IIA is to analyze and then organize incoming reports such that the relevant incident reports serving as evidence for a particular issue are captured within one coherent collection [94]. This IIA collection can then be used as evidence for regulatory decision-making on products or services.

Investigative activities such as IIA require analysts' judgement throughout the process to make decisions on the importance and relevance of each report to an investigation, hence making IIA a suitable candidate for interactive analytics approaches. Existing investigative analysis tools such as JigSaw [162, 71] focus on finding relationships among documents or entity mentions within and across documents to uncover one larger terrorist plot. As opposed to JigSaw, in IIA instances of independent incident reports about a suspected problem are analyzed to identify and collect evidence to confirm the problem. In short, JigSaw focuses on hypothesis formation, while our focus is on finding evidence to confirm a hypothesis about a potential problem.

Similarly, general analytics history and annotation approaches [156, 46] have been developed to allow analysts to record and manage insights during exploratory analysis. These approaches however focus on how to best record the insights

interactively, while our goal instead is to take a step further and utilize such captured insights to assist analysts in steering investigations by collecting and managing evidence.

In Pharmacovigilance, analysts monitor batches of medical incident reports submitted by patients to identify adverse drug reactions or medication errors associated with drugs [78]. Robust detection of adverse reactions and medication errors represents a life-critical task, as adverse reactions are one of the leading causes of death worldwide [106]. Given this criticality, drug safety analysts must perform due diligence on hundreds of incident reports on a daily basis, with the aim of identifying reports as evidence to a suspected drug safety problem. When these analysts find an incident of concern with a particular drug or medical treatment, they tend to pivot their analyses toward finding and organizing similar incidents to *build a case for action* [94]. Moreover, a case needs continuous management as instances of evidence are received over time. Cases take longer to conclude as enough evidence is needed to confirm a potential problem. Thus, analysts continuously monitor multiple cases and add evidence to them as incident reports are received every week.

In this chapter, we present a design study involving a prototype tool, ConText, designed in collaboration with Pharmacovigilance analysts that aims to support analysts in Instance-based Incident Analysis (IIA). ConText assists an analyst to interactively examine individual text narratives aided by domain-informed NLP-generated interest points *to identify evidential information within a report.* ConText realizes a systematic approach for assisting analysts in finding instances of evidence to *build and strengthen a case* supporting the occurrence of a critical incident, and *interactively managing multiple ongoing cases* by leveraging analysts' knowledge and findings. We evaluate ConText via case studies and interviews to better understand the needs of IIA in an operational context.

Contributions of this work include:

- Characterization of the Instance-based Incident Analysis (IIA) workflow and ongoing challenges through in-depth analysis of IIA practices via a series of interviews and follow-up discussions with domain experts at the FDA.
- ConText, an iteratively designed IIA-focused prototype that includes a composition of visualization and domain-informed computational elements designed to aid analysts in evidence collection and management by providing a unified workflow.
- Recommendations following insights gained during the development and evaluation of ConText that opens opportunities in human-computer interaction and visualization for designing future systems supporting IIA.

6.2 Background and Related Work

6.2.1 Incident Report Analysis

Much of the existing work in Pharmacovigilance has focused on applying computational techniques on incident reports to detect potential drugs causing reactions [98, 143]. Extensive work developing natural language processing (NLP) techniques [181] to extract key information from these text reports also exists. Computational techniques alone, however, are not sufficient due to the analysts'

judgement being core to successful IIA. Hence, this makes IIA a viable candidate for interactive analytics.

On the other hand, publicly available online tools such as OpenFDA [99] and OpenFDAVigil [26] allow the general public explore incident reports and get basic statistics about the reports. Other approaches have visualized relationships and correlations between drugs and their reported reactions [187, 92, 94]. These works do neither chararcterize nor support the broader scope of IIA activities performed by the analysts. Apart from this, we are not aware of studies to understand IIA tasks or approaches that use interactive analytics to support such tasks performed in other domains, such as, aviation [7] and financial services [37].

6.2.2 Visual Analytics for Document Analysis

The majority of the existing visualization techniques for text analysis center around designing summaries of a set of documents using topic analysis, word-clouds or meta information from text [110, 148, 71, 96]. These overviews may be potentially applied to screen a particular problem represented by a set of reports. Our goal instead is to support the identification and collection of evidence for a potential problem, screened using one of these overview techniques [96, 71].

Extensive research work has been done to design visual tools for 'close reading' of literary texts, such as books and poems [89]. These tools help in understanding the structure and content of the literary documents. Literary texts are different than incident reports in the sense that the later corresponds to *problem-centric* documents to be investigated with the sole purpose to assess the reported incidents.

In the medical domain, visual analytics tools target the analysis and summarization of electronic health records (EHR) and clinical notes [132, 138, 153, 179, 167]. The majority of these tools display overviews of patient's medical history consisting of multiple events to help the medical professionals with diagnosis or treatment. Conversely, in IIA, analysts identify an issue that is happening to multiple patients and build a case by collecting evidence about similar incidents to take a regulatory action.

6.2.3 Analytics Record Keeping Tools

Extensive research has been conducted to record, manage and annotate findings and insights during exploration [45, 43, 156, 73] which resembles our *case management task.*

Harvest [155] provides a mixed-initiative based approach to automatically recommend notes and concepts from past analyses that are relevant to the current analysis. Aruvi [156] supports the analytical reasoning process by providing support for recording analysis artifacts such as findings and hypotheses. Sandbox [180] is a gesture-based editor for collecting and managing insights and discoveries on a visualization. Click2annotate [43] provides automatically computed editable templates for interpreting outliers and clusters. Jigsaw's successor [113] supports insight capturing through a tablet view using text-diagramming, that allows an investigator to sketch their findings on a timeline.

These systems target the design of interactive techniques to allow analysts frame relevant information together and clarify connections between data points. Our goal instead is to use analysts' captured insights to support evidence collection for ongoing investigations to assist the IIA workflow.

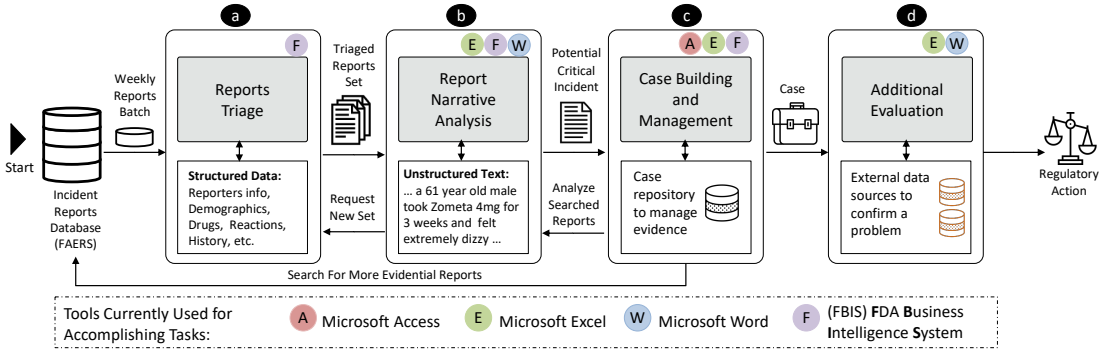


Figure 6.1: Instance-based Incident Analysis (IIA) workflow in the FDA with tools supporting each task. Tasks (a & d) are out of the scope of this work.

6.2.4 Investigative Tools

Investigative tools such as Jigsaw [162, 71] focus on finding relationships among data entities within and across documents, constructing visualizations around these relationships. While like Jigsaw, we extract key information from the text narratives, one difference is that exploration in JigSaw ends with reading the individual documents [162], while ConText's workflow starts with the analysis of individual report and mainly focuses on collecting independent instances of evidences for similar problem. Moreover, ConText supports interactive tracking of multiple simultaneous cases required for IIA as depicted in Fig. 6.2. For IIA, Jigsaw can be used to screen a certain problem, while ConText supports the tasks of evidence identification, collection, and management for the screened problem.

On the other hand, bug tracking tools aim to resolve bugs related to software or issues related to products or services [152, 17]. The majority of such tools [152] aim to resolve software bugs by assigning the problem tickets to respective agents and track the status of their resolution. In contrast, our goal is to support the intermediate steps, i.e., the evidence identification and collection to be able to initiate an action to be taken towards the resolution of a problem. Similarly, in cybersecurity, tools have been designed to help operators quickly identify and address network intrusions and incidents [11]. Like cybersecurity, our goal of finding critical issues remains, but the data, challenges, and workflows of security and drug incident response domains vary— requiring thorough design, development, and evaluation of incident management and response tools for Pharmacovigilance.

6.3 Understanding the IIA Workflow at FDA

This work is part of our over two and a half years collaboration with the FDA to design visual analytics to improve their drug safety review process. For this project, we conducted a series of semi-structured in-person interviews and follow-up discussions with five (5) domain experts who were drug safety analysts, to observe and understand their IIA workflow, tools and challenges. In our preliminary interviews, we observed the drug safety analysts perform their routine review tasks in a think-aloud manner. We recorded these interviews and transcribed them to get concrete design requirements for ConText.

After these preliminary interviews we had biweekly remote meetings and email discussions with three of these experts to get more feedback initially on the requirements and eventually on the design of ConText during the later stages of the

project. Overall the project took about eight months from requirement analysis to development. We formulated and refined the design requirements throughout this time.

6.3.1 Background on Medical Incident Reports

Based on our interviews we learned that the U.S. Food and Drug Administration (FDA) regularly receives medical incident reports about medication errors and adverse reactions through their adverse event reporting system know as FAERS [65]. These reports are submitted by consumers, health-care professionals and drug manufacturers. Each report has structured information such as demographics of the patients and therapy related information and an unstructured narrative describing the details of the adverse event suspected to be caused by the drug.

6.3.2 Current IIA Practices at the FDA

Based on our extensive analysis and discussions with the domain experts, we express our understanding of the current IIA workflow at the FDA as depicted in Fig. 6.1. Incident reports about suspected drug-safety problems are received on a daily basis yet tend to be investigated every week in a batch-wise manner to find potential reactions and errors caused by medical products. The goal of the analysts is to find if any of these reports is discussing a real potential problem and indeed is worthy of further investigation and ultimately warranting regulatory action.

Teams of drug safety analysts review these reports based on the medical products assigned to them to detect potential critical incidents. The analysts triage reports related to a suspected drug-safety problem based on structured information associated with each report in the FAERS, such as outcome and severity of the incidents (Fig. 6.1a). If the analysts find a report indicative of a problem, they examine the text narratives of the triaged report to analyze the details of the incident and identify if the narrative has enough information and can serve as evidence to the reported problem (Fig. 6.1b).

Once an incident is identified as evidence to a suspected problem, other reports that could potentially further corroborate the incident, are collected to build a case to be presented to the management for further evaluation (Fig. 6.1c). Once sufficient evidence is compiled that can confirm the problem then regulatory actions are taken (Fig. 6.1d). These actions include adding warnings to the drug label or in worst case removing the product from the market [78]. In this work, we only focus on narrative analysis to identify evidential reports and case building and management to collect and monitor evidence (Fig. 6.1b & c).

Next we consider a real example revealing the importance of IIA and the challenges involved in finding and leveraging information from these reports for decision making.

6.3.3 Motivating Real World IIA Scenario from FDA

In 2012, the FDA drug safety analysts were conducting their routine review of incident reports to find drug safety problems. During the analysis of one of the report's narrative (Fig. 6.1b), the analyst observed that a patient using steroid injection was hospitalized due to a rare adverse reaction "fungal meningitis". This was an unexpected adverse event causing serious damage. The analyst thus decided to open an investigation about this incident and searched through the database to see if other similar incidents had been reported (Fig. 6.1c).

After exhaustively searching and collecting other evidential reports, they built a case and found that these patients received steroid injections from the New England Compounding Center (Fig. 6.1d). The investigation was concluded with an order of inspection of the facility. Later, it was revealed that the product was contaminated due to the violation of safety standards. Thus regulatory actions were taken [62]. This incident corresponds to the well-known fungal meningitis outbreak scandal in Massachusetts that killed 64 people and hospitalized 700 nationwide [62].

Clearly, the more effectively we can support the tasks of such investigative process (Fig. 6.1b & c), the faster we can solve potentially life threatening health issues, such as the crisis described above.

6.3.4 Challenges in Current IIA Practices

Analysts currently perform IIA tasks manually using a variety of tools to analyze the narratives (Fig. 6.1b) and collect further evidence to build a case (Fig. 6.1c) if the report is indicative of investigation. For instance, analysts use FAERS Business Intelligence System (FBIS) [23] to compose SQL-based search queries to retrieve relevant reports from the FAERS database using the structured information. Analysts use Microsoft (MS) Excel and MS Access to keep track of their ongoing investigations by manually recording reports evidence to these investigations. Similarly, MS Word is used to search through the narratives text when investigating a particular incident.

These current practices utilise tools and techniques that require a manual trawl through the reports, which is time-consuming, laborious and error prone. Moreover, in IIA, evidence is collected over time as reports are reviewed every week, manually keeping track of the open investigations and adding evidence as they are received, solely relies on the analysts' memory. This becomes even challenging for the investigations that are open for a longer period of time, i.e., months even years. While, analysts are the drivers of the IIA process – their judgment and perspective is crucial to make decision on a report's importance and opening an investigation – a unified tool is needed that leverages computational and interactive features to help them efficiently achieve their goals of evidence identification, collection, and management.

6.4 Data Abstraction and Design Requirements

In this section, we characterize the data and tasks involved in IIA based on our understanding of the FDA review workflow, to help us formulate the design goals for ConText.

6.4.1 Instance-based Incident Analysis (IIA) Data Model

First, we introduce the data model of IIA by defining the basic components (Fig. 6.2).
Suspect. A suspect is an entity that is the potential cause of the incident. In case of FDA, drugs are the suspects.
Issue. Issue is an entity that the incident is about. For instance, reactions or medication errors are the issues at FDA.
Suspected Problem. An *unconfirmed* drug-safety problem presenting a possible association between the suspect and issue, or, in this case a drug and a reaction or medication error.

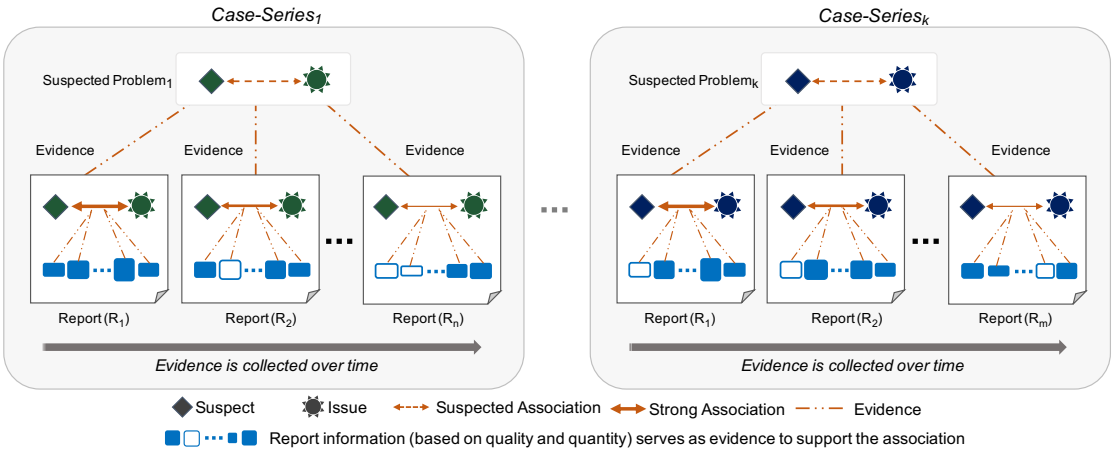


Figure 6.2: Instance-based Incident Analysis (IIA) data model. Incident reports represent the instances of a reported (suspected) problem. IIA involves building case-series for each of the suspected problems by identifying reports that can serve as evidence to confirm theses problems. The data inside each report, which can vary, is evidence to the problem in each instance.

Incident Report. Incident reports are about suspected problems pointing toward an association between a drug and reaction. A report, however, does not necessarily mean that there is an association and therefore analysts assess each report to ensure that critical issues are not missed. Incident reports are intended for regulatory agencies to bring their attention toward a problem so that they can take an appropriate action to stop or mitigate it.

Evidence. An incident report is evidence to a suspected problem, if the information provided in the report can serve as evidence to support the association between the suspect and issue. That is, information necessary to assess a suspected problem is present in the report. Evidence can be stronger or weaker depending upon the quality of report's information.

Case-Series. A case-series represents a set of incident reports that each can serve as evidence for the suspected problem. A case-series is created when the first evidence of a suspected problem is identified and eventually with more evidence the suspected problem can be confirmed, leading to regulatory action to prevent future incidents from happening.

Instance-based Incident Analysis. Instance-based Incident Analysis (IIA) involves building case-series for the suspected problems by identifying instances of incident reports that are evidence to each of these problems. Investigations that involve regulatory actions are expensive in terms of resources and consequences, therefore, before starting an investigation, extensive evidence is collected to build a case. Our goal is to support this evidence collection and case-building process.

These definitions are based on the Pharmacovigilance, however, they can be adapted for other domains. For instance, for criminal investigation, a person (suspect) can be associated with a particular crime (issue) such as sexual harassment or corruption, and multiple instances (reports) of similar issue are collected to start an investigation against the person. In case of FAA, multiple instances of a particular maintenance issue related to an aircraft (suspect) are identified to taken an action. This is converse to the kind of investigation Jigsaw [162] supports which involves uncovering a hidden major plot by identifying connections between entities within and across documents. In case of IIA, documents are only associated with the suspected

problem, and independent documents (instances) are analyzed to identify if they can serve as evidence to the problem (Fig. 6.2) and further evidence is sought to build a case.

6.4.2 IIA Tasks and Design Requirements

Below we describe the core tasks derived from our interviews and study of the IIA work-flow for Pharmacovigilance.

T1 Identify evidential information within a report. Analysts review every report to make sure not to miss any potential problem that is critical and unknown to the organization. The goal is to identify if any of the reports can serve as evidence to the problem. Analysts seek specific information within a report to assess if the report has evidential information. This key information includes the demographics of the patient, the drugs taken by the patient and the observed reactions. Other factors such as medical history and symptoms etc. are also assessed.

T2 Collect evidence and build case-series. Once a suspected problem is escalated, substantial evidence (reports) is needed to support that a drug might be causing the reaction. When the first report serving evidence to a problem, is discovered, reports are searched over a longer period of time to see if further evidence exists. Analysts formulate queries and refine them continuously to get all the relevant reports that represent additional evidence.

T3 Track multiple case-series. Due to a high number of potential suspected problems (thousands of drugs and reactions), a large number of case-series may be opened simultaneously for ongoing investigations of multiple problems, as investigations for drug-safety problems generally take longer to be concluded. In addition, due to *recurrence*, new case-series are formed as new reports are received and analyzed every week. Thus, keeping track of all active case-series and collecting relevant evidence for them while keeping an eye on the new batch of reports to not miss an alarming problem is extremely challenging.

As opposed to doing these tasks manually and with no guidance, ConText combines them into a single workflow and provides computational and interactive features to augment the analysis. To formally justify the design components and features discussed throughout the rest of this chapter, we first define a set of guidelines that ConText should adhere to:

DG1 Highlight evidential information within report. The key information within a text narrative that can be used as evidence to confirm the suspected problem should be highlighted to make it easily identifiable.

DG2 Support instance-based evidence search. Analysts extensively formulate and refine queries based on keywords from an identified evidential report to find other similar instances of reports that can serve as evidence to the suspected problem. The system should compose such queries automatically and allow analysts to interactively curate them.

DG3 Facilitate automatic tracking of multiple cases. The system should keep track of the open cases (investigations) by automatically identifying instances of reports evidence to these cases as they are received every week and notify the analysts to validate such findings.

ConText contains three visual components to fulfill these requirements. For **DG1**, the Incident Analysis View is designed to help analysts identify evidential information within a report. In consideration of **DG2** and **DG3**, the Case Management Dashboard is designed to help in evidence collection to build and monitor cases, respectively.

6.5 Designing ConText for IIA

6.5.1 ConText User Interface

ConText is designed to achieve the IIA tasks to identify, collect and monitor evidence for ongoing investigations in a unified manner. There are two main components of the ConText prototype, the *Incident Analysis View* (Fig. 6.3) and the *Case Management Dashboard View* (Fig. 6.5).

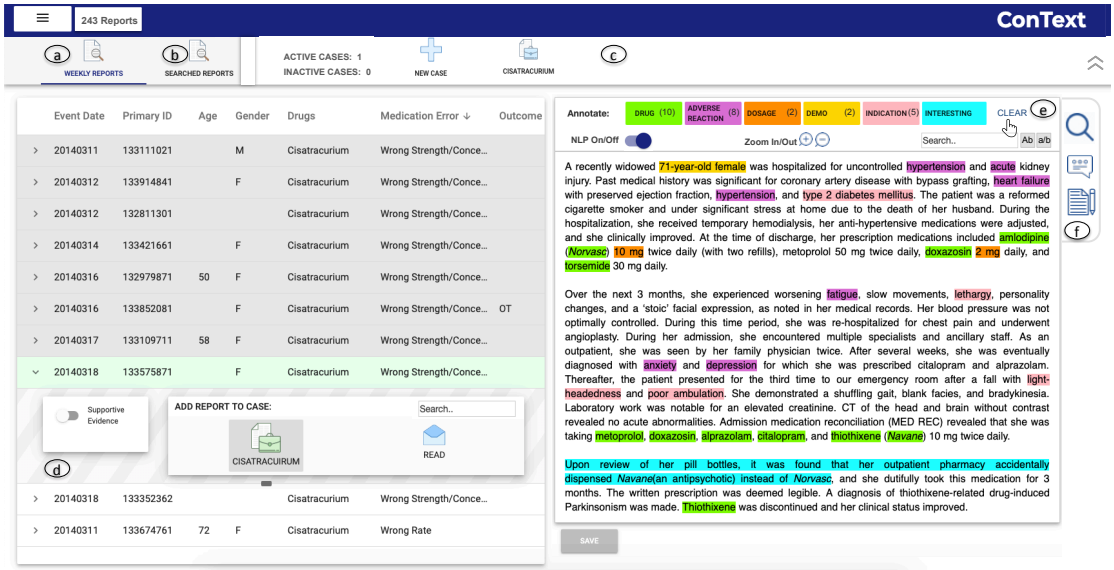


Figure 6.3: The Incident Analysis View allows analyst to identify evidential reports with line-listing of reports on the left and Content Analysis Panel on the right. (a) Current weeks reports. (b) Panel to view searched reports. (c) The dashboard panel with active cases. (d) To add a report to a case. (e) Menu to correct and control the annotations in the narrative. (f) Option to search across the reports and creating notes on a report. Due to privacy concerns, a FAERS-like publicly available safety report narrative [54] is shown in the Content Panel.

Incident Analysis View.

The incident analysis view provides the line-listing of the reports (Fig. 6.3-Left) along with the Content analysis panel (Fig. 6.3-Right) for narrative analysis. An analyst can select a report from the line-listing and analyze its text narrative to interactively identify evidential information using the features below.

Domain-informed Access Points. Analysts need to identify the evidential information in the text narratives that can help them assess if the report indeed necessitates the opening of an investigation. ConText aids analysts to locate the important relevant information (access points) within the text quickly via natural language processing (NLP). These domain-informed access points include the data elements requested by the analysts to be extracted such as drugs, reactions and demographics. Though various techniques to extract name entities from biomedical text exist [9, 181], we leverage the MEFA framework, that uses a combination of rule-based and machine-learning-based name-entity recognition techniques to extract the key information from the FAERS reports narratives [181].

Fig. 6.3 depicts a text narrative with highlighted access points extracted using NLP to guide the analyst. We provide an option to interactively correct or update any inaccurate extracted information using the annotation menu (Fig. 6.3f). Further, these access points can be toggled on or off based on the analyst's preference. Options to search within the narrative are also provided (Fig. 6.3e).

User Driven Annotations. Besides the access points, ConText also provides the analyst with the ability to add free-form comments linked with an evidential report or mark interesting keywords or phrases while reading and analyzing a document. Such direct annotations from the analyst are very useful. First, analysts can remember their insights and findings when the document is reviewed again at a later time. Second, these annotations can be helpful in capturing the information that is important to the analysts for assessing an incident, and thus can be leveraged to facilitate the next steps of investigation as discussed below.

Case Management Dashboard

To provide an interactive approach to build and monitor case-series, we have designed the incident case management dashboard (Fig. 6.5) that provides the features described below.

Case Creation. At any time during the analysis if an analyst considers a report important enough to open an investigation, ConText allows her to interactively create a case ((Fig. 6.3c) and add the report into the case ((Fig. 6.3d). The analyst can also select if a report is primary (strong evidential information) or supportive evidence (weak information) to the case (Fig. 6.3d).

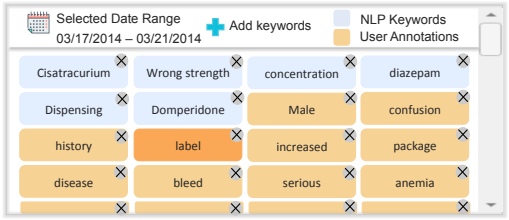


Figure 6.4: The Recommendation Query Panel with both NLP and user generated descriptors to allow analysts steer investigations by finding relevant evidential reports.

Case Building. When a case is created, ConText recommends other reports with similar evidential information to assist the analyst in building a case. For this we leverage the keywords marked within the report, either using NLP or explicitly annotated by the analyst as they convey important information about the reports — therefore are conduits towards constructing descriptors of a case. These case descriptors are used to automatically recommend reports that are most similar to the reports within a a case. A similar approach is taken by Cheng et al. [48] to recommend relevant web pages based on user's notes. More advanced document recommendation techniques using word embeddings [21]

can also be plugged into ConText.

An analyst can interactively update the domain-informed "recommendation query" by modifying the descriptors as well as operation (Fig. 6.4). By default, the full case descriptor is used in the recommendation query to retrieve documents using the 'OR' operation, which can be changed to an 'AND' operation by double clicking (Fig. 6.4- the descriptor *label* uses 'AND' operation). We leverage the well-known *Okapi BM25* [140] information retrieval model with the *inverted index* built upon incident reports to implement such recommendation (using Apache Lucene).

By default, we use reports from the past two weeks for the search and score calculation, but this can be interactively adjusted. The reports that match the query descriptors are ranked based on the proportion of the matched descriptors from the query represented as glyph (Fig. 6.5b). These ranked reports are then recommended to the analysts for review (Fig. 6.5b) with the matched descriptors highlighted in the recommended reports narratives.

Case Summary. ConText provides an overview of the reports collected within a case-series by displaying a summary of the key information including the phrases marked by the analyst during the analysis of the narratives (Fig. 6.5d). For the analysts annotations, we display each keyword as a bar with its length mapped to the frequency of the keyword. The count of reports associated with a keyword is also displayed to help analysts in composing their query.

An analyst can also view the distribution of the reports marked as strong and supportive evidence, along with the comments added to reports within a case. This helps the analyst to quickly review the important details of their case especially when a case is open for a longer time.

Case Monitoring. An analyst might be working with many active cases, with some new and others older ones. Therefore, ConText provides an alerting feature to notify analysts about the arrival of new evidential reports that are relevant to any existing active case-series. Every week new reports are added to the existing unread reports, thus creating a new batch. When new reports are received that match an existing case-series in terms of the specified recommendation query, then analysts are alerted on their dashboard (Fig. 6.5e). By default the alert operation executes every week using the case descriptors from the case summary to recommend the relevant reports. When desired, the analyst can customize the alert criteria by modifying the query (Fig. 6.4).

At any time during the investigation. the state of the analysis is saved and can be retrieved upon login. Once a decision has been reached and a case is marked as inactive, it would no longer be visible in currently active cases view.

6.6 Development Process and Insights

As previously mentioned, ConText is part of our over two years of collaboration with the FDA. In this section, we discuss some of the main steps of the design and development process during this period. Our goal is to highlight some of the challenges we faced during this time – this will be the basis for some of the lessons we present in Section 6.8.

Though our design study does not explicitly follow the nine stages suggested by Sedlmair et al. [149], we did naturally progress through the main stages of the process. Due to our ongoing collaboration, we had observed that evidence collection and management were one of the core and most frequently performed tasks. Thus

our first stage started with the *core phase* to discover and characterize the problem [149], that is, requirement analysis as discussed in Section 6.3. After the preliminary interviews, we had biweekly remote meetings with three of these experts for follow-up discussions on refining the requirements and design goals.

During the *design phase*, we iteratively refined our designed features, visualizations and interface based on the feedback from the experts. As design alternatives, we presented sketches of tag-clouds with tf-idf and topic clusters using Latent Dirichlet Allocation [25] for report analysis. The experts, however, found these approaches suitable for someone unfamiliar with these reports to get general insights. Analysts instead emphasized on keeping the raw reports intact as they seek certain information within these reports for decision making. For the recommended documents our initial designs included 2D representations (tiles) of the most similar documents with score mapped to visual cues, but analysts wanted to access the actual report along with the score right away without additional mouse clicks (Fig. 6.5b).

The line-listing in the Incident Analysis Panel was designed similar to the current tools that FDA uses for skimming through the reports (Fig. 6.3). In our initial design, the text narrative was displayed in a new tab, but experts recommended to have it side-wise along with the line-listing (Fig. 6.3). Regarding features, for the narrative view, in our initial designs, we only included domain-informed access points generated by NLP. Conversations with the experts led to including user-driven annotations to allow analysts capture their insights both at report and case-level, which we later became a core component for the evidence recommendation.

During the *implement phase*, after ConText was developed and tested in lab for usability, we conducted a pilot test with these three experts, which highlighted a few usability issues as well as suggestions on improving ConText features. For instance, experts suggested to let analysts choose which keywords they want to use for search in the Recommendation Query, and allowing them to set a customized alert for case monitoring as the criteria for evidence might vary from incident to incident. Other suggestions included using visual cues to differentiate between NLP-based keywords and user annotations for efficient query editing. After including these features in the final version of ConText presented in Section 6.5.1, we evaluated the system by conducing case-studies and interviews with ten analysts that were not involved in the design process and discuss it in detail in Section 6.7.

Finally, the insights we gained during the design and development of ConText in Section 6.8 corresponds to the *write* and *reflect* stages of writing design studies [149].

6.7 Evaluating ConText for IIA in Pharmacovigilance

We evaluated the effectiveness of ConText for IIA by conducting case studies followed by semi-structured interviews with ten domain experts (˜20% of the workforce) who are drug safety analysts at the FDA. These experts were not involved in the design process of ConText and were familiar with basic visualizations such as bar charts.

These evaluation sessions were each 1.25 hour long. During the first 15 minutes, we demonstrated the ConText prototype to these experts. After getting familiar to ConText (˜10 minutes), we asked these experts to explore the system to perform IIA tasks by analyzing their reports set in a think-aloud manner for about 30 minutes, while the follow-up semi-structured interviews and discussions took approximately twenty minutes. We recorded their feedback during the interview. The FAERS data from 2014

was used during the evaluation.

6.7.1 Case Studies

To illustrate the workflow of ConText and how it supports the Instance-based Incident Analysis (IIA) tasks, we present the following case studies derived directly from observing two of the analysts using ConText for reviewing reports for medication errors and adverse reactions. Due to space limitation, we present two case-studies here, but other analysts were able to derive similar findings. One of the analyst volunteered for follow-ups regarding the evidence alert feature of the system, as illustrated in the following case study. The name of the analyst in this case study is anonymized.

Investigating Medication Errors

Amy, a drug safety analyst responsible for medication error detection starts exploration of the reports using the Incident Analysis View (Fig. 6.3a). She has received 243 reports for this week that she needs to analyze one by one. From the line-listing, she observes that the drug 'Cisatracurium' is reported to have a serious outcome, thus she filters the reports for the drug 'Cisatracurium' and opens the narrative of the first report to read it (Fig. 6.3). The first thing she notices are the highlighted drug names and medication errors. She immediately reads the sentence that has the mention of the drugs. It says "Pharmacist called stating that they may be experiencing potency issues with Cisatracurium and Padilaxel, but they are not sure..".

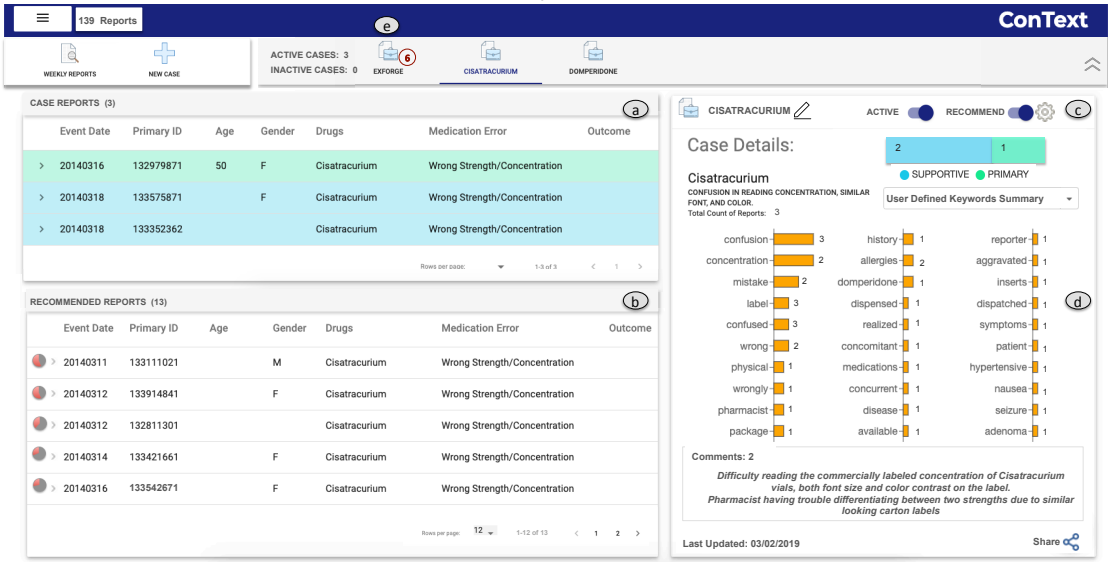


Figure 6.5: Dashboard view to build and monitor cases. Dashboard for the Case 'Cisatracurium'. (a) Reports within the case. (b) Recommended reports for the case. (c) Controls to set the recommendation query. (d) Case summary. (e) Alert of getting new evidential reports for the case Exforge.

Skimming through the rest of the narrative, she comments "this narrative does not give much information about the incident, so i will not read it further". She marks the report as 'read'. It turns grey in the reports panel (Fig. 6.3-Left) to allow quick recognition of its 'read' state. She next opens the second narrative. While reading through the narrative, she finds the term 'internal bleeding' and highlights it using the interesting marking on the annotation menu (Fig. 6.3e). She marks the narrative as read, and adds the comment "Insufficient information".

She repeats the same process for the next five narratives and keeps highlighting the interesting words in each narrative, when she scans over them during her analysis. She also adds comments where needed. She states, "this is our regular routine, most of the reports are false alarms and usually don't get investigated further".

For the next report's narrative, she reads the sentence "several ICU nurses were having difficulty in reading the commercially labeled concentration of Cisatracurium vials" with 'Cisatracurium' highlighted. She says, "Cisatracurium is a muscle paralyzing agent used on patients before surgery, if the concentration is mistaken then it can have critical health consequences". She is now suspicious and is reading the full narrative from start to end. As she is reading it, she is highlighting other phrases of interest to her. She seems interested in this narrative and explains, "this narrative can be describing a potential medication error with the drug Cisatracurium having similar labels for different concentrations" (**T1**). She adds "Difficulty reading the commercially labeled concentration of Cisatracurium vials, both font size and color contrast on the label" as a summary of the narrative by clicking on the comment button on the Content Analysis view (Fig. 6.3f). She clicks on the 'Create New Case' icon (Fig. 6.3c) and names it 'Cisatracurium'. She adds the description 'Confusion in reading concentration, similar font and color' to the case. By clicking the case name, she includes the report in this case as strong evidence (Fig. 6.3d). She can see the details of this newly created case in the dashboard (Fig. 6.5).

She now investigates if there are other evidential reports in the database with similar characteristics. She clicks on the "recommendation setting" button (Fig. 6.5c) and uses the default option of selecting all keywords in the query (Fig. 6.4). In a few seconds, she sees 13 reports in the recommended reports panel (Fig. 6.5b) that match the search criteria which she needs to investigate further one by one (**T2**).

She repeats the same process the next two weeks. At that time, she has identified two new reports for this case. She also has created two more cases 'Domperidone' and 'Exforge' through her findings during these past two weeks. She again is curious about the reports inside the 'Cisatracurium' case. By clicking on her 'Cisatracurium' case under her list of active cases (Fig. 6.5), she reads the description and remembers that this case was about some confusion in reading different concentrations of Cisatracurium. She also sees on the summary panel, that one of the three reports is a strong evidence while two are supportive.

Now, she is curious about the most important keywords in this case that she has been marking during her analysis. Using the drop-down menu (Fig. 6.5d), she selects "User-defined keywords" and sees a distribution of the most frequently words marked by her. She observes that the most frequent terms in all these cases are "confusion" and "concentration". Reading through the comments, she says "I remember, these reports were talking about confusion in reading the concentration information". She now goes back to her routine report analysis.

Case Monitoring. One week later, when Amy logs into ConText, looking at her dashboard, she notices a notification of six newly received reports on her 'Exforge' case (Fig. 6.5e). She opens the case and sees the list of recommended reports (**T3**). Upon reading their narratives one by one, she comments "Although these reports discuss issues with Exforge, none of these reports however have sufficient information, so I will just mark them as read". Amy then proceeds to this week's routine analysis. At any time during her analysis, she can go back to investigate a particular case.

In this way, ConText allows Amy to efficiently perform her IIA tasks by assisting her

in identifying evidential information within the reports, guiding her towards further evidence to build and monitor multiple investigations.

Investigating Adverse Reactions

Next we describe the case study reflecting the detection of adverse reactions conducted by one of the drug safety analysts. The analyst is to explore ConText for the 121 reports he received and analyze if there is any potential problem that needs investigation.

While skimming through the reports in the Incident Analysis View (Fig. (Fig. 6.3), he notices one report with a serious outcome for the drug 'Optiray' (given before CT scans) being associated with the reaction "throat tightness". He considers this report important and opens the narrative to read the details and observes that it's reported by a radiologic technologist. He adds "if reporter is a medical professional, it adds to the reliability of the report, because there are certain reactions that only medical professionals can recognize better, for instance, Stephen Johnson syndrome". He further reads the sentence with highlighted reaction, "After the injection of Optiray, the patient described that she felt pressure on her throat and was treated with 50 mg IV Benadryl (diphenhydramine) and two 4 mg prednisone tablets". Reading through rest of the narrative that "the patient recovered after the given fluids" while highlighting the important keywords such as patient's history, the analyst considers this report important and creates a case with the name 'Optiray'.

Now, the analyst wants to know if there are other reports about Opitray with similar complains, that is, he wants to find more evidence. He opens the Recommendation Query Panel (Fig. 6.4) and changes the dates to past four weeks and includes the keyword 'allergic' to the keywords. In a few seconds, six reports are displayed. He skims through the narrative of each of these reports in a similar fashion focusing on the highlighted entities. These reports were pointing to allergic reactions such as rash, and red and puffy eyes. In one of the sentences with highlighted reactions he reads "After the administration of Optiray 350, the patient developed an itchy throat and sneezed over 15 times. She was treated with 25 mg IV diphenhydramine (Benadryl)". He adds these reports to the case, and the case-summary is populated with all the highlighted keywords. He elaborates that "the fact all the patients in these reports developed the reaction within few minutes of administration of Optiray and were recovered after a doze of Benadryl states that Optiray might be associated with these allergic reactions and this should be further investigated". This way ConText facilitates the identification and collection of evidence for IIA.

6.7.2 Expert Interviews

As part of the evaluation, after the aforementioned case-studies, we conducted semi-structured interviews with these 10 domain experts to get further feedback on the strengths and weaknesses of the system. These interviews were guided by the following questions.

Q1. Which features of the ConText prototype do they believe could make a difference in their daily IIA practices?

Q2. Why are these features important to them?

Q3. Express opinions on the overall positive and negative aspects of ConText.

Overall, the feedback was encouraging. Fig. 6.6 depicts the features that were considered the most important for the workflow by the majority of our participants. From Fig. 6.6 we see that 'analysts annotations' currently supported by current tools (Excel) is comparatively rated lower than the novel features that are not supported by the current workflow.

Justifications for the lower ratings (3-score) provided by the majority of participants included concerns about the accuracy of the computational techniques, that is, recommending relevant evidence for new as well as ongoing investigations and identifying key information within the narratives using NLP. One participant mentioned "The highlights in the text (pointing towards NLP-generated access points) were not always correct, an adverse reaction was highlighted as an indication. It would be great if these mistakes are not that often". A few participants mentioned that if the system consistently guides them towards the true evidential reports, then they would trust these features and ConText would be beneficial for their tasks. Other features that were selected by fewer participants included 'marking a report strong or supportive evidence' and 'searching based on exact and fuzzy match'.

Participant's lower rating is due to lack of trust on the computational techniques such as NLP or recommended evidence because they are used to manually analyzing actual report text. While user trust is a big challenge in information visualization [119], one way to address this challenge beside developing more accurate algorithms is to design transparent visualizations to inform user about the uncertainties in the data [119] – a possible direction to investigate in the future.

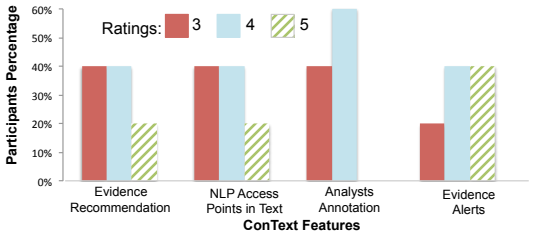


Figure 6.6: Participants feedback (ratings) on the importance of features. Ratings are at 5-point likert scale (1=least important, 5=most important..

Regarding the highly rated features (Q2), analysts comments are summarized next. P3: "Sometimes these reports are very lengthy and looking for a drug or error can be tedious. This (domain-informed access points) can save us a lot of time" (**DG1**). P6: "Currently, we have to check our list of monitored drugs and match it with current reports to see if we have got a new report, or rely on our memory. Having an alert system that notifies us about new possibly relevant reports can be very helpful" (**DG3**). P10: "Searching for reports that are similar in content is very hectic currently, the recommend functionality can make our lives easier" (**DG2**). In short, analysts had positive feedback and were excited to have a ConText-like unified tool to assist them in their IIA tasks.

Limitations in the current prototype were also recorded. In particular, few participants suggested more transparent visual query functions, such as interactive ways of building logical and/or queries. Others suggested to use visual cues to distinguish their commented 'read' reports from the uncommented ones to easily identify them.

6.8 Discussion

Design studies are a growing part of visualization, however, due to being subjective and specific to a domain, their value and contribution is often difficult to assess.

A recent work [120] discusses six criteria for rigor in visualization design study research that can guide researchers in communicating and assessing knowledge claims. According to these criteria, a design study should be *informed, reflexive, resonant, plausible, abundant,* and *transparent*. Our design study of ConText aligns with these criteria in the following five ways.

Our over two years of close collaboration with domain experts at the FDA provided **abundant** knowledge about the IIA problem, its challenges and opportunities for interactive features and displays. Our design of ConText is **informed** by the literature review as well as the needs of the analysts and described as the design goals (Sec. 6.4.1). We presented the abstraction of data and tasks (Sec. 6.4) to communicate our implicit understanding of IIA at the FDA. This helps to promote **transparency** of our particular knowledge which is later embedded in the interactive views and features provided by ConText. The design of ConText prototype makes our study **plausible** as the views and features realize our goals of assisting analysts with the IIA tasks.

Our overall approach is **resonant** as it can help in knowledge transerferability by setting a foundational step to adapt interactive analytics in other domains having IIA workflows. Although, ConText is designed for the FDA, Pharmacovigilance is a large domain world-wide [90] having similar analytics goals and challenges [22]. Particularly, the data abstraction along with the features to collect and maintain evidence can be adapted by other domains such as the FAA.

Contributing to these criteria of rigor [120], we further highlight the insights we learned through the design and evaluation of ConText. These insights were extracted through the analysis of the discussions and interviews conducted with the domain analysts. This would help us achieve *resonance* by expanding and transferring our knowledge of analysts needs and desires which can *inform* the designs of future systems and techniques for assisting analysts in IIA.

Actual text reports are the backbone of IIA. During the initial interviews, we observed that analysts continuously emphasized the importance of reading the actual text narratives to make a decision if a reported incident is worthy of opening an investigation. The importance of reading the actual text document is highlighted previously in other investigative domains as well [72]. Our initial designs included 2D document representations for the recommended documents. Analysts wanted to see the document first hand without having to click through the 2D tile (abstraction). Existing work to aid actual documents with visualizations, without disrupting the raw text, is limited to typographic features such as highlighting [164]. Other text analytics approaches provide summaries and abstractions of text documents, such as clustering, or topic analysis [89]. If future designs consider adding a layer of such abstraction to these text reports, careful design and thorough evaluation is needed to examine if the abstraction improves or worsens the efficiency of the IIA workflow, because analysts have to read the narrative in any case.

Goal-oriented analysis outweighs general insight seeking. During the initial design phase of the project, we discussed sketches of designs consisting word clouds and topic analysis [25] – common visual document analysis approaches – to help analysts examine the incident reports and get insights about these reports. The insights from these techniques were general (unique or similar keywords/topics) and were not helpful in making a decision whether a certain report is indicative of a potential problem. The analysts instead emphasized on the importance of certain information within these reports for decision making. This was confirmed during evaluation as

well when analysts would first read the sentences with highlighted NLP-generated access points, paying attention to specific information such as drug and treatment. While insight generation is one of the main metrics for evaluating visual analytics [126], it's important that such insights are aligned with the analysts goals, which we have witnessed during the development of ConText. Therefore, for more focused tasks such as IIA, analysis goals should be prioritized over general insight seeking during design.

Basic accessible features are crucial for analysts to achieve their analytics goals. During our observational interviews we repeatedly noticed analysts struggling with their current tools due to the lack of elementary features crucial for the analysis. For instance, analysts were not able to see the searched keywords highlighted in the narratives, or search with multiple keywords and operations simultaneously, or zoom the text (in case of bad eyesight), or align a messy narrative to reduce the cognitive load. The lack of these features exhausted the analysts, thus compelling them to use alternative accessible tools that can help them achieve their desired goals (MS Word). Therefore, along with focusing on advance sophisticated features, such simpler operations should not be ignored when designing new systems.

Smooth transition between concurrent investigative tasks is important. While analysts were using ConText during evaluation, we observed that they often switched between analysing their weekly batch and collecting evidence for a particular problem they found interesting while reviewing a narrative. One analyst, in particular, spotted a report indicative of an adverse reaction from the reports recommended for a potential medication error and started investigating it by creating a new case. In short, analysts do not review reports in a linear fashion, and ends up going in multiple directions based on their findings, which ConText supports. During evaluation, analysts also mentioned that sometimes they lag behind their weekly reports analysis due to an active and time-intensive investigation of an ongoing safety problem. Therefore, it is crucial for future designs to keep track of the state of multiple analyses at all times by providing occasional reminders about their incomplete tasks, so that they can easily pick up where they left off an analysis.

Sharing analysis threads across teams needs thorough consideration. During the evaluation, we observed that analysts were excited about adding and sharing their insights with team members using ConText. However, they were concerned about the content of comments being shared. Further discussions with analysts revealed that they share insights with varied styles and levels of details with different team members based on the needs of an ongoing investigation. For instance, a team member may be well aware of the ongoing investigation, hence, analysts are comfortable sharing their informal notes depicting minor details to illustrate the issue. On the other hand, the higher management being unaware of the issue may require more formal and detailed analysis of the problem under investigation. Henceforth, future designs can benefit by providing support for sharing insights at multiple levels of granularity.

Providing support for tracking analysts' operations leading to insights. When analysts were performing IIA tasks during our evaluation, several analysts pointed out the need for the system to remember their queries used during searching for evidence. Particularly, upon the confirmation of a report being evidence to a suspected problem, the analysts want to use the same search query possibly with minor modifications, for future searches to collect evidence for a certain case. Therefore, future designs for IIA systems should incorporate analysts' interactions with the system, i.e., provide support

for analytic provenance, to allow analysts quickly access their reasoning process that lead to the insights in the first place.

6.9 Conclusion

In this chapter, we contribute a design study for an analytics prototype ConText to support the Instance-based Incident Analysis (IIA). Our design of interactive operations and features is based on an in-depth analysis of the Pharmacovigilance workflows at the US FDA. ConText is designed as a unified system to support the identification of an incident of concern, finding evidence to build and strengthen a case supporting the incident, and interactively managing multiple ongoing cases over a large weekly batch of semi-structured text reports. We discuss our experience of designing and evaluating ConText, and share the insights we gained during this process to benefit future IIA tools and techniques for this and similar real-world problems concerning public safety.

In the future we plan to empirically study the long-term usability of ConText in performing IIA tasks. This would give us further insights into the adoption of an interactive analytics tool such as ConText in life-critical workflows. Other research directions include designing and evaluating interactive trustworthy visual displays and features for analyzing textual data [119].

Chapter 7

Conclusion

In the first part of this dissertation we study the Pharmacovigilance domain workflows and challenges (Chapter 2) via user interviews and contextual inquiries at the FDA. In Chapter 3 we present Medication Error Visualization (MEV), an interactive visual analytics system that provides overview of the regularly received incident reports to triage the most critical incidents. MEV uses a treemap-based visualization to display reports distribution along with crucial reports parameters extracted via natural language processing to help analysts prioritize incidents for further analysis. In Chapter 4 we present DIVA, an interactive visual analytics system that provides multiple coordinated views to help analysts in the smooth exploration and triage of hypothesized drug-drug interaction incidents generated by machine learning techniques. DIVA's node-link diagrams based views follow the report screening and triage tasks of domain workflows and provide an effective way to make sense of machine generated output.

In Chapter 5 we discuss the need and challenges in the triage and analysis of individual reports associated with a screened incident. We present SumRe, a glanceable visual summary, that provides compact view of the crucial information within each report to help in forming a hypothesis about a potential critical incident. A user study with 20 domain experts demonstrates the effectiveness of a glanceable visual summary for reports triage.

Chapter 6 presents, ConText, an interactive tool to support the in-depth investigation of an incident of concern by helping analysts in efficient identification of reports that could serve as evidence to build a case for action. ConText aids analysts in evidence collection and management for incidents of concern by a close coupling of information foraging and synthesis.

Chapter 8

Discussion and Future Work

Building on the ideas developed in this dissertation for interactive visual analytics to explore, triage, and investigate incident reports, we now outline the following interesting future avenues for each task for further study.

8.1 Exploratory Triage of Incident Reports

Working on the first research task (Chapter 3 & 4) helped us identify the following research avenues to explore in the future.

8.1.1 Analytics for Exploring and Monitoring a Drug's Lifetime Incidents

The adverse reactions or medication errors of a drug are regularly monitored from the time a drug is approved [82]. During its lifespan, by regulation a drug is reviewed multiple times and with each review actions such as updating the drug label are recommended. Apart from these major reviews, drug labels can be updated if a safety issue is confirmed during the routine analysis of incident reports.

Currently, our overview analysis (Chapter 3 & 4) helps with reports triage related to a particular drug or incident. However, these design fail to provide a complete picture of a drug's lifetime incidents, such as when a certain incident was first reported, how it evolved over time, and if the recommended actions had any impact on them. Answering these questions can help the regulatory agencies monitor the impact of their decisions as well.

Figure 8.1 depicts a mocked visualization of a drug from the time it is approved. The (zoomed) timeline shows major events (e.g., reviews) that have happened in the past five years. Frequency of reported medication errors along with report outcome (severe or non-severe) during these major events are shown along the blue axis. For instance, the medication error ME1 which was first reported after fifteen days of approval increased in frequency over time. An analyst can drill up and down the timeline to have a high-level as well as detailed analysis based on months, weeks, or days. Moreover, similar to [179] the visualization can be aligned by any of the major events. For instance, if an analyst has just reviewed the drug a few months ago, she can align the timeline based on the last review to analyze the reports trends since then.

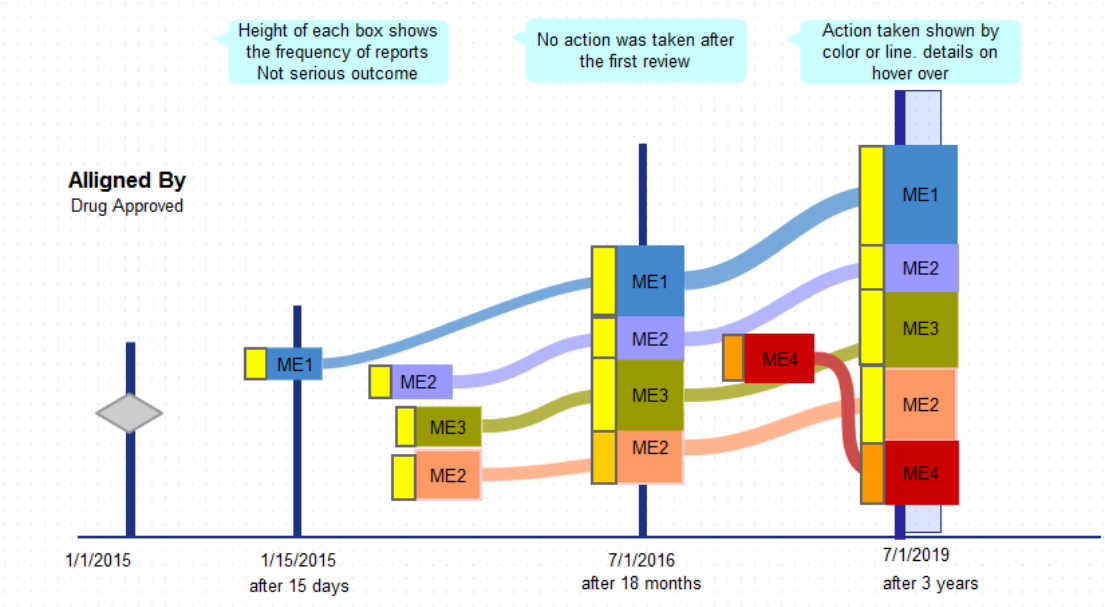


Figure 8.1: Mockup of a drug's lifetime incident triage.

8.1.2 Designing Transparent Overviews for Incident Triage

During our evaluation of the MEV and ConText tool, we observed analysts questioning the uncertainty of quality of information extracted from text using NLP. Uncertainty in any real world data is unavoidable and exists in many domains including medicine, aviation, and weather forecast to name a few [158]. Effectively communicating uncertainty is necessary for establishing scientific transparency and to help make informed decisions. One research direction for incident reports could be to characterize uncertainty in the Pharmacovigilance and identifying opportunities to model and visually communicate uncertainty for more informed decisions.

Like many domains [158], drug incident reports contain uncertainty at multiple levels. These include errors during data entry and extraction (Measurement), missing or unreported information (Completeness), uncertainty in inferring if a reaction is caused by the drug or other medical condition and drugs are the caveats (Inference), credibility of the reporter and report source, and disagreement among drug safety analysts on contradicting evaluation of a signal.

The uncertainty due to inference and the disagreement among analysts is more subjective and can be reduced if rest of the uncertainties are communicated effectively [158]. To address these uncertainties, we explore recently developed technology and research in other domains to communicate uncertainty for incident reports.

Inference and Measurement Uncertainty. There has been recent work to model uncertainty using Bayesian and probabilistic approaches [130]. We can leverage these modeling techniques to visualize uncertainty raised due to computational techniques, such as, natural language processing to extract key data elements from the narrative and signal detection algorithm to identify potential signals. As both of these processes inherently provide uncertain information, providing a probability distribution of the range of accuracy for a certain data element instead of only providing the output (data) can improve transparency and user's trust in the system.

Completeness Uncertainty. Out of all uncertainties, accurately detecting missing

information is one of the challenges we faced during the evaluation of SumRe. There are many ways to visualize uncertainty [130]. For instance using color, transparency, alignment, and icons or symbols [130]. In SumRe, we used a grey color to represent missing information, with the underlying data being visually encoded with variable sizes and shapes. This made it difficult to detect all the missing information accurately. A future direction is to empirically study and investigate different visual encodings to represent missing information in incident report overviews and summaries. This can be achieved by following the study design from [116] which evaluates the intuitiveness of various visual semiotics for uncertainty visualization.

8.2 Confirmatory Triage of Incident Reports

During the design and evaluation of the visual report summary (Chapter 5 we discovered that the following research avenues can be explored in the future.

8.2.1 Studying the Effect of a Visual Summary on User Engagement and Reading Behavior

Another direction is to empirically study the effect of a glanceable visual summary on reading the narrative for further investigation. This is a step further to the reports triage discussed in Section 5 in the reports review workflow.

Conditions. To study the effect of SumRe on reading behavior, we intend to compare against the baseline tabular layout (Fig. 8.2) with plain text narrative currently used at the FDA. In the literature, many text analytics systems highlight key information in the text [64, 161]. We also include this condition to compare against the state-of-the-art techniques (Fig. 8.3) following the guidelines for typographic encodings [165]. For the visual summary condition, we leverage word-scale visualization from the summary (SumRe) to highlight key information in the text (Fig. 8.4) following the guidelines from Goffin et al. [69]. We also consider these three layouts with and without interactions. The interactions considered in this case include hovering over the data in the summary will highlight the data in the narrative.

Tasks and Metrics. For this study, tasks from narrative visualization domain can be leveraged, where a visualization is used to tell a story generally for media and journalism purposes [150]. These tasks include information extraction and comprehension. Participants will be provided with a summary and narrative one at a time and asked questions about searching particular information from the summary or the narrative or both. A similar task is also performed by Goffin et al. [69] where the effect of different placement of word-scale visualizations on reading behavior is studied. In this study the effect of a particular summary and narrative design on comprehension, reading behavior, and experience can be quantified. This will include capturing the time spent on each report as well as answering the questions, the accuracy of the answers, and user interactions with the summary and narrative.

Reporter	Age	Gender	History	ADRs	AdrDate	Outcome	Drugs	Concmt	Indication	Dechallenge
Other Healthcare	49 Years Old		Multiple Sclerosis	Seizure, Decreased Interest	04/06/2014	Serious	Ampyra	Baclofen, Lisinopril	Multiple Sclerosis	Positive

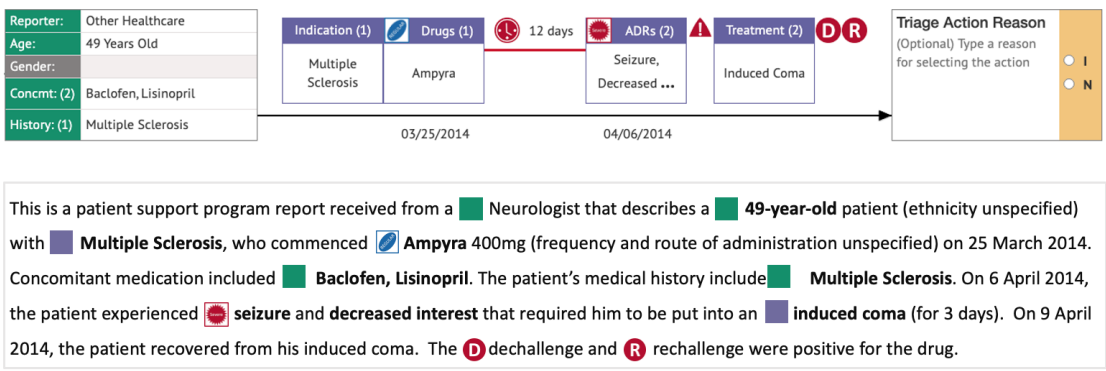
This is a patient support program report received from a neurologist that describes a 49-year-old patient (ethnicity unspecified) with Multiple Sclerosis, who commenced Ampyra 400mg (frequency and route of administration unspecified) on 25 March 2014. Concomitant medication included Bclofen, Lisinopril. The patient's medical history was not reported. On 6 April 2014, the patient experienced seizure and decreased interest that required him to be put into an induced coma (for 3 days). On 9 April 2014, the patient recovered from his induced coma. The dechallenge and rechallenge were positive for the drug.

Figure 8.2: Baseline table and plain textual narrative (TP) condition.

Reporter	Age	Gender	History	ADRs	AdrDate	Outcome	Drugs	Concmt	Indication	Dechallenge
Other Healthcare	49 Years Old		Multiple Sclerosis	Seizure, Decreased Interest	04/06/2014	Serious	Ampyra	Baclofen, Lisinopril	Multiple Sclerosis	Positive

This is a patient support program report received from a neurologist that describes a 49-year-old patient (ethnicity unspecified) with Multiple Sclerosis, who commenced Ampyra 400mg (frequency and route of administration unspecified) on 25 March 2014. Concomitant medication included Bclofen, Lisinopril. The patient's medical history includes Multiple Sclerosis. On 6 April 2014, the patient experienced seizure and decreased interest that required him to be put into an induced coma (for 3 days). On 9 April 2014, the patient recovered from his induced coma. The dechallenge and rechallenge were positive for the drug.

Figure 8.3: Mockup for the baseline table and highlighted textual narrative (TH) condition.



This is a patient support program report received from a Neurologist that describes a **49-year-old** patient (ethnicity unspecified) with **Multiple Sclerosis**, who commenced **Ampyra** 400mg (frequency and route of administration unspecified) on 25 March 2014. Concomitant medication included **Baclofen, Lisinopril**. The patient's medical history include **Multiple Sclerosis**. On 6 April 2014, the patient experienced **seizure** and **decreased interest** that required him to be put into an **induced coma** (for 3 days). On 9 April 2014, the patient recovered from his induced coma. The D dechallenge and R rechallenge were positive for the drug.

Figure 8.4: Mockup for the visual Summary and textual narrative with word-scale visuals.

8.2.2 Studying the Effect Contextual Information in Incident Report Triage

Our designed visual summary for the confirmatory triage, SumRe, contains information about a report instance and does not include the overall context of reports submitted to the FDA. Such information can be helpful to understand the usual patterns of an incident when analyzing an individual incident report to make more informed triage decisions. One future research direction is to provide a fluid triage experience considering other reports and evaluate if it can tell a better story by conducting a user study.

Figure 8.5 depicts a mockup to include the context of reports Statistics using scented widgets [177] that helps in comparing a certain week's reports to usual trend of such reports. Visual interactions can help drill down into an attribute's effect on the rest of the population, for instance, hovering over a drug can display how the ADR affects the population. Statistics such as likelihoods, number of occurrences, and conditional probabilities can be leveraged to provide context.

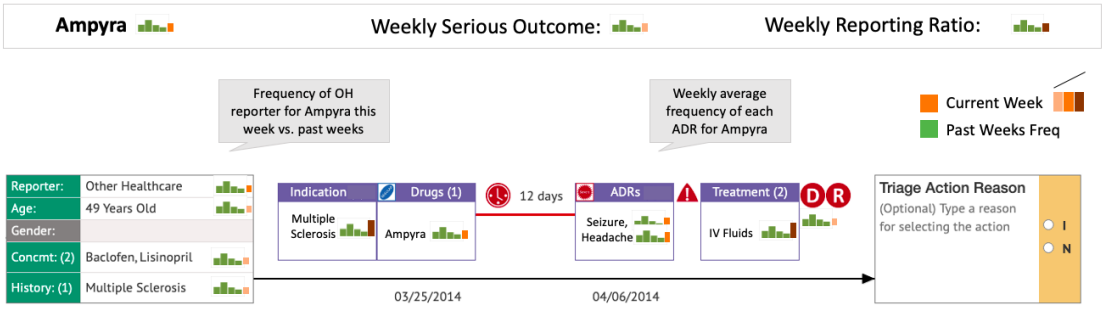


Figure 8.5: Mockup for the visual summary with contextual information. An analyst can view the report attributes associated with a drug and compares to the reports usually received for that drug.

8.3 Investigatory Analysis of Incident Reports

Working on ConText (Chapter 6 to help analysts collect evidential reports similar to a triaged report, we realized that the following research direction can be further investigated in the future.

8.3.1 Designing and Studying Semantic Similarity Measure for Incident Investigation

During the investigation of a particular drug-related-reaction also called *signal*, analysts manually screen through thousands of incident reports to select the ones that can be used as evidence for the signal. At the core of this process, analysts are comparing reports with certain search criteria and selecting the most similar ones. Currently, after applying the search filters, analysts review these reports one-by-one to select the most relevant subset of reports. This also happens when analysts find a report interesting, they seek other similar reports. The challenges in the above task include:

Multi-Faceted Attributes Comparison. The report summaries used during screening have a variety of diverse multi-faceted attributes, where a change in single

facet can make the report different. For instance, a change in the medical history that can be confounding to the reaction, may make the report useless for the investigation. Therefore, highlighting the differences in the reports is crucial.

In-explicit Similarity Criteria. The criteria for similarity is highly dependent upon the signal under investigation. For instance, in certain cases reports about drugs under a similar therapeutic class might be considered similar, while investigating drug-drug interactions, drugs within same class might be considered different. Therefore, analysts should have the control to define the similarity criteria for each incident under investigation.

As discussed in Chapter 5, analysts seek specific information in the form of report summaries and we designed a *glanceable* visual summary to display such information as shown in the Figure 5.3. However, due to richer visual encodings in SumRe, it is challenging to simply highlight similar information, and hence more sophisticated and interactive techniques are needed to visually map the differences among reports.

Our designed summary, SumRe, for multiple reports resemble the concept of small-multiples, where the visualization is similar but the data varies across the charts [171]. The major difference is that in small-multiples the data changes across a few attributes. While, in case of summaries, data changes across many attributes, and even across facets within an attribute. One research direction is to design interactions and features to help analysts efficiently and effectively spot differences among reports in relation to a reference report. The goal is to allow analysts not only find the syntactic differences but also semantic differences in the reports. For instance, two reports about patients with age 60 and 61 are not considered different based on age. Similarly, medical histories containing the terms cardiac arrest, myocardial infarction, and heart attack might not be considered different when evaluating a reaction related to heart disease. We intend to leverage domain informed definitions along with the best practices from information visualization to help analysts compare individual reports.

Following are a few directions to identify differences in the reports.

- **Interactive Attribute Simplification-based Similarity Definition.** One approach is to empower analysts to define similarity for the signal under investigation. For instance, defining an age group 'senior' which includes any patients with age fifty or higher, or treating a list of drugs within a therapeutic drug class, such as Tylenol and Paracetamol. Similarly, allowing analysts to interactively remove items from comparison, such as reporter type or gender depending upon the need. A similar approach is taken by Monroe et al. [121] to interactively simplify event sequences for summarizing data. Such approach can help simplify the complex differences in these reports.

- **Dynamic Relative Delta Visualization.** To visually highlight the differences among report summaries, techniques for visual comparison can be leveraged [68]. One approach would be to compare the summaries based on the spatial position assigned to specific data, for instance, change in patient's information, or post-incident information. Out of the three major techniques for comparing visualizations including juxtaposition, superimposition, and explicit encodings, superimposition might not be effective for report comparison as it is prone to visual clutter. A hybrid approach by combining the other two techniques, that is, juxtaposition and explicit encoding can be adapted for our multi-faceted summaries. For instance, explicit encoding can be used to only highlight the

differences between juxtaposed reports by performing a subtraction operation. This will need to mute the attributes that are similar, and only highlight the differences among summaries.

- **Human-in-the-Loop-based Similarity Score.** As the analyst reviews the reports, she can discard or select a report as evidence. A weighted similarity score will be updated with each selection and rejection, and analyst can penalize certain information in the reports that can be confounding to the causality of reaction. The machine-learning algorithm can update the weights in real-time highlighting the highly scored reports.

Evaluation Existing work to evaluate visual comparison techniques are limited to numeric data such as maxima and delta [127]. In case of incident report summaries, the majority of the data is categorical, hence textual. Thus designing tasks to evaluate the report comparison techniques would include domain scenarios as well as sub-tasks of finding visual differences from the information visualization literature [127]. One direction can be to study and compare the effect of different similarity functions on triage. For instance, studying triage experience and performance using a sorted list of reports based on a distance-based similarity measure such as cosine similarity and [163], a learned distance function [34], and a randomly ordered list of reports. A step forward in this direction would be having users evaluate the similarity based on visual encodings of the report summaries. A recent work on semantic interactions where Endert et al. [59] update a similarity function based on analysts's interactions with the documents can be adapted for incident reports as well.

Bibliography

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

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

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

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

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

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

[7] F. A. Administration. SDR. http://av-info.faa.gov/sdrx/, 2010. Accessed: 2018-3-3.

[8] M. Alharbi and R. S. Laramee. Sos textvis: An extended survey of surveys on text visualization. *Computers*, 8(1):17, 2019.

[9] M. Allahyari, S. Pouriyeh, M. Assefi, S. Safaei, E. D. Trippe, J. B. Gutierrez, and K. Kochut. A brief survey of text mining: Classification, clustering and extraction techniques. *arXiv preprint arXiv:1707.02919*, 2017.

[10] 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.

[11] S. Amershi, B. Lee, A. Kapoor, R. Mahajan, and B. Christian. Cuet: human-guided fast and accurate network alarm triage. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 157–166, 2011.

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

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

[14] T. Asahi, D. Turo, and B. Shneiderman. Visual decision-making: using treemaps for the analytic hierarchy process. In *The Craft of Information Visualization*, pages 235–236. Elsevier, 2003.

[15] A. Aula, R. M. Khan, Z. Guan, P. Fontes, and P. Hong. A comparison of visual and textual page previews in judging the helpfulness of web pages. In *Proceedings of the 19th International Conference on World Wide Web*, pages 51–60, 2010.

[16] 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'. *Health Technology Assessment*, 15(20):1–234, 2011.

[17] Y. Avnon and S. L. Boggan. Fit and finish using a bug tracking system: challenges and recommendations. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems*, pages 4717–4720. ACM, 2010.

[18] R. Badi, S. Bae, J. M. Moore, K. Meintanis, A. Zacchi, H. Hsieh, F. Shipman, and C. C. Marshall. Recognizing user interest and document value from reading and organizing activities in document triage. In *Proceedings of the 11th International Conference on Intelligent User Interfaces*, pages 218–225, 2006.

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

[20] S. Bateman, R. L. Mandryk, C. Gutwin, A. Genest, D. McDine, and C. Brooks. Useful junk? the effects of visual embellishment on comprehension and memorability of charts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2573–2582, 2010.

[21] M. Berger, K. McDonough, and L. M. Seversky. cite2vec: Citation-driven document exploration via word embeddings. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):691–700, 2016.

[22] T. Bergvall, G. N. Norén, and M. Lindquist. vigigrade: a tool to identify well-documented individual case reports and highlight systematic data quality issues. *Drug safety*, 37(1):65–77, 2014.

[23] BIFACT. *FAERS Business Intelligence System (FBIS)*. `http://www.bifact.com/faers-bifact.html` [Accessed: 2019-06-19].

[24] J. Blanchard, F. Guillet, and H. Briand. Exploratory visualization for association rule rummaging. In *KDD-03 Workshop on Multimedia Data Mining (MDM-03)*, 2003.

[25] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3(Jan):993–1022, 2003.

[26] R. Böhm, L. von Hehn, T. Herdegen, H.-J. Klein, O. Bruhn, H. Petri, and J. Höcker. Openvigil fda–inspection of us american adverse drug events pharmacovigilance data and novel clinical applications. *PloS one*, 11(6):e0157753, 2016.

[27] R. Borgo, J. Kehrer, D. H. Chung, E. Maguire, R. S. Laramee, H. Hauser, M. Ward, and M. Chen. Glyph-based visualization: Foundations, design guidelines, techniques and applications. In *Eurographics (STARs)*, pages 39–63, 2013.

[28] M. A. Borkin, Z. Bylinskii, N. W. Kim, C. M. Bainbridge, C. S. Yeh, D. Borkin, H. Pfister, and A. Oliva. Beyond memorability: Visualization recognition and recall. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):519–528, 2015.

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

[30] T. Botsis, C. Jankosky, D. Arya, K. Kreimeyer, M. Foster, A. Pandey, W. Wang, G. Zhang, R. Forshee, R. Goud, et al. Decision support environment for medical product safety surveillance. *Journal of biomedical informatics*, 64:354–362, 2016.

[31] T. Botsis, J. Scott, R. Goud, P. Toman, A. Sutherland, and R. Ball. Novel algorithms for improved pattern recognition using the us fda adverse event network analyzer. *Studies in health technology and informatics*, 205:1178–1182, 2014.

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

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

[34] E. T. Brown, J. Liu, C. E. Brodley, and R. Chang. Dis-function: Learning distance functions interactively. In *2012 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 83–92. IEEE, 2012.

[35] C. Brunk, J. Kelly, and R. Kohavi. Mineset: An integrated system for data mining. In *Knowledge Discovery and Data mining*, pages 135–138, 1997.

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

[37] C. F. P. Bureau. CFPB. `https://www.consumerfinance.gov/`, 2011. Accessed: 2018-4-3.

[38] 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.

[39] 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.

[40] 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.

[41] R. Capra, J. Arguello, and F. Scholer. Augmenting web search surrogates with images. In *Proceedings of the 22nd ACM International Conference on Information and Knowledge Management*, pages 399–408, 2013.

[42] R. Chang, C. Ziemkiewicz, T. M. Green, and W. Ribarsky. Defining insight for visual analytics. *IEEE Computer Graphics and Applications*, 29(2):14–17, 2009.

[43] Y. Chen, S. Barlowe, and J. Yang. Click2annotate: Automated insight externalization with rich semantics. In *Visual Analytics Science and Technology (VAST), 2010 IEEE Symposium on*, pages 155–162. IEEE, 2010.

[44] Y. Chen, L. Wang, M. Dong, and J. Hua. Exemplar-based visualization of large document corpus (infovis2009-1115). *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1161–1168, 2009.

[45] Y. Chen, J. Yang, S. Barlowe, and D. H. Jeong. Touch2annotate: generating better annotations with less human effort on multi-touch interfaces. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems*, pages 3703–3708. ACM, 2010.

[46] Y. Chen, J. Yang, and W. Ribarsky. Toward effective insight management in visual analytics systems. In *Visualization Symposium, 2009. PacificVis' 09. IEEE Pacific*, pages 49–56. IEEE, 2009.

[47] 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*, pages 175–184. ACM, 2016.

[48] W.-H. Cheng and D. Gotz. Context-based page unit recommendation for web-based sensemaking tasks. In *Proceedings of the 14th International Conference on Intelligent User Interfaces (IUI)*, pages 107–116. ACM, 2009.

[49] J. Chuang, D. Ramage, C. Manning, and J. Heer. Interpretation and trust: Designing model-driven visualizations for text analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 443–452. ACM, 2012.

[50] C. Collins, S. Carpendale, and G. Penn. Docuburst: Visualizing document content using language structure. In *Computer graphics forum*, volume 28, pages 1039–1046. Wiley Online Library, 2009.

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

[52] 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*, pages 258–267. Springer, 2007.

[53] G. Cumming. *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge, 2013.

[54] B. A. Da Silva and M. Krishnamurthy. The alarming reality of medication error: a patient case and review of pennsylvania and national data. *Journal of Community Hospital Internal Medicine Perspectives*, 6(4):31758, 2016.

[55] F. Dabek, E. Jimenez, and J. J. Caban. A timeline-based framework for aggregating and summarizing electronic health records. In *2017 IEEE Workshop on Visual Analytics in Healthcare (VAHC)*, pages 55–61. IEEE, 2017.

[56] A. Don, E. Zheleva, M. Gregory, S. Tarkan, L. Auvil, T. Clement, B. Shneiderman, and C. Plaisant. Discovering interesting usage patterns in text collections: integrating text mining with visualization. In *Proceedings of the Sixteenth ACM Conference on Information and Knowledge Management*, pages 213–222, 2007.

[57] P. Dragicevic. Fair statistical communication in hci. In *Modern statistical methods for HCI*, pages 291–330. Springer, 2016.

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

[59] A. Endert, P. Fiaux, and C. North. Semantic interaction for visual text analytics. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 473–482, 2012.

[60] 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.

[61] M. Farrugia, N. Hurley, and A. Quigley. Exploring temporal ego networks using small multiples and tree-ring layouts. *Proceedings of Advances in Computer-Human Interactions*, 2011:23–28, 2011.

[62] FDA. New england compounding center pharmacist sentenced for role in nationwide fungal meningitis outbreak. `https://www.fda.gov/ICECI/CriminalInvestigations/ucm594800.htm`, 2018. Accessed: 2018-02-10.

[63] C. Felix, A. V. Pandey, and E. Bertini. Texttile: an interactive visualization tool for seamless exploratory analysis of structured data and unstructured text. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):161–170, 2016.

[64] C. Felix, A. V. Pandey, and E. Bertini. Texttile: an interactive visualization tool for seamless exploratory analysis of structured data and unstructured text. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):161–170, 2017.

[65] U. Food and D. Administration. FDA Adverse Event Reporting System (FAERS) data statistics. `https://www.fda.gov/drugs/surveillance/questions-and-answers-fdas-adverse-event-reporting-system-faers`, 2015. Accessed: 2017-01-10.

[66] U. Food and D. Administration. FDA Adverse Event Reporting System (FAERS) public dashboard. `https://www.fda.gov/drugs/questions-and-answers-fdas-adverse-event-reporting-system-faers/fda-adverse-event-reporting-system-faers-public-dashboard`, 2019. Accessed: 2020-08-20.

[67] A. P. Fundin and B. L. Bergman. Exploring the customer feedback process. *Measuring Business Excellence*, 7(2):55–65, 2003.

[68] M. Gleicher, D. Albers, R. Walker, I. Jusufi, C. D. Hansen, and J. C. Roberts. Visual comparison for information visualization. *Information Visualization*, 10(4):289–309, 2011.

[69] P. Goffin, W. Willett, A. Bezerianos, and P. Isenberg. Exploring the effect of word-scale visualizations on reading behavior. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, pages 1827–1832, 2015.

[70] C. Görg, Z. Liu, J. Kihm, J. Choo, H. Park, and J. Stasko. Combining computational analyses and interactive visualization for document exploration and sensemaking in jigsaw. *IEEE Transactions on Visualization and Computer Graphics*, 19(10):1646–1663, 2012.

[71] C. Görg, Z. Liu, J. Kihm, J. Choo, H. Park, and J. Stasko. Combining computational analyses and interactive visualization for document exploration and sensemaking in jigsaw. *IEEE Transactions on Visualization and Computer Graphics*, 19(10):1646–1663, 2013.

[72] C. Görg, Z. Liu, and J. Stasko. Reflections on the evolution of the jigsaw visual analytics system. *Information Visualization*, 13(4):336–345, 2014.

[73] D. Gotz, M. X. Zhou, and V. Aggarwal. Interactive visual synthesis of analytic knowledge. In *Visual Analytics Science And Technology, 2006 IEEE Symposium On*, pages 51–58. IEEE, 2006.

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

[75] J. Haggerty, S. Haggerty, and M. Taylor. Forensic triage of email network narratives through visualisation. *Information Management and Computer Security*, 2014.

[76] 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(6):2021–2025, 2011.

[77] 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*, pages 244–253. ACM, 2000.

[78] L. Härmark and A. Van Grootheest. Pharmacovigilance: methods, recent developments and future perspectives. *European Journal of Clinical Pharmacology*, 64(8):743–752, 2008.

[79] 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.

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

[81] 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.

[82] K. B. Hoffman, M. Dimbil, N. P. Tatonetti, and R. F. Kyle. A pharmacovigilance signaling system based on fda regulatory action and post-marketing adverse event reports. *Drug safety*, 39(6):561–575, 2016.

[83] 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*, pages 227–235. ACM, 2000.

[84] D. Holten. Hierarchical edge bundles: Visualization of adjacency relations in hierarchical data. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):741–748, 2006.

[85] K. Hornbæk and M. Hertzum. The notion of overview in information visualization. *International Journal of Human-Computer Studies*, 69(7-8):509–525, 2011.

[86] E. Horvitz. Principles of mixed-initiative user interfaces. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 159–166. ACM, 1999.

[87] H. Ibrahim, A. Saad, A. Abdo, and A. S. Eldin. Mining association patterns of drug-interactions using post marketing fda's spontaneous reporting data. *Journal of Biomedical Informatics*, 60:294–308, 2016.

[88] A. N. Jagannatha and H. Yu. 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, 2016.

[89] S. Jänicke, G. Franzini, M. F. Cheema, and G. Scheuermann. On close and distant reading in digital humanities: A survey and future challenges. In *Eurographics Conference on Visualization (EuroVis)-STARs. The Eurographics Association*, 2015.

[90] G. Jeetu and G. Anusha. Pharmacovigilance: a worldwide master key for drug safety monitoring. *Journal of Young Pharmacists*, 2(3):315–320, 2010.

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

[92] W. Jian-Xiang, Z. Yun-Xia, S. Jun, X. Hou-Ming, L. Ming, and S. Yue-Hong. Adrvis: an information visualization platform for adverse drug reactions. *International Journal of u-and e-Service, Science and Technology*, 8(10):139–150, 2015.

[93] D. Jonker, W. Wright, D. Schroh, P. Proulx, B. Cort, et al. Information triage with trist. In *2005 International Conference on Intelligence Analysis*, pages 2–4. Citeseer, 2005.

[94] T. Kakar, X. Qin, E. Rundensteiner, L. Harrison, S. Sahoo, and S. De. Diva: Exploration and validation of hypothesized drug-drug interactions. In *Computer Graphics Forum*, volume 38, pages 95–106. Wiley Online Library, 2019.

[95] T. Kakar, X. Qin, E. A. Rundensteiner, L. Harrison, S. K. Sahoo, and S. De. Diva: Towards validation of hypothesized drug-drug interactions via visual analysis. In *Eurographics*, 2019.

[96] T. Kakar, X. Q. Qin, C. M. Tapply, O. Spring, D. Murphy, D. Yun, E. A. Rundensteiner, L. Harrison, T. La, S. K. Sahoo, and S. De. Mev: Visual analytics for medication error detection. In *International Conference on Information Visualization Theory and Applications (IVAPP), Prague, 2019*. SciTePress, 2019.

[97] 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.

[98] S. Karimi, C. Wang, A. Metke-Jimenez, R. Gaire, and C. Paris. Text and data mining techniques in adverse drug reaction detection. *ACM Computing Surveys (CSUR)*, 47(4):56, 2015.

[99] T. A. Kass-Hout, Z. Xu, M. Mohebbi, H. Nelsen, A. Baker, J. Levine, E. Johanson, and R. A. Bright. Openfda: an innovative platform providing access to a wealth of fda's publicly available data. *Journal of the American Medical Informatics Association*, 23(3):596–600, 2015.

[100] E. S. Kastango. Lessons learned from compounding tragedies. *The Canadian Journal of Hospital Pharmacy*, 66(3):152, 2013.

[101] T. W. Kay Lazar. 14 executives, staff charged in tainted drug case. `https://www.bostonglobe.com/metro/2014/12/17/feds-arrest-tied-necc-closed-pharmacy-blamed-causing\\-deaths-with-tainted-drugs/24QytISVdaqXDNsSUOsNYJ/story.html`, 2014. Accessed: 2017-01-10.

[102] S. Koch, M. John, M. Wörner, A. Müller, and T. Ertl. Varifocalreader—in-depth visual analysis of large text documents. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1723–1732, 2014.

[103] K. Kucher and A. Kerren. Text visualization techniques: Taxonomy, visual survey, and community insights. In *Visualization Symposium (PacificVis), 2015 IEEE Pacific*, pages 117–121. IEEE, 2015.

[104] 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.

[105] B. Y. Kuo, T. Hentrich, B. M. Good, and M. D. Wilkinson. Tag clouds for summarizing web search results. In *Proceedings of the 16th international conference on World Wide Web*, pages 1203–1204. ACM, 2007.

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

[107] B. Lee, G. Smith, G. G. Robertson, et al. 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, 2009.

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

[109] H. Liu, T. Selker, and H. Lieberman. Visualizing the affective structure of a text document. In *CHI'03 extended abstracts on Human factors in computing systems*, pages 740–741. ACM, 2003.

[110] S. Liu, X. Wang, C. Collins, W. Dou, F. Ouyang, M. El-Assady, L. Jiang, and D. Keim. Bridging text visualization and mining: A task-driven survey. *IEEE Transactions on Visualization and Computer Graphics*, 2018.

[111] S. Liu, M. X. Zhou, S. Pan, Y. Song, W. Qian, W. Cai, and X. Lian. Tiara: Interactive, topic-based visual text summarization and analysis. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 3(2):25, 2012.

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

[113] Z. Liu, C. Görg, J. Kihm, H. Lee, J. Choo, H. Park, and J. Stasko. Data ingestion and evidence marshalling in jigsaw vast 2010 mini challenge 1 award: Good support for data ingest. In *Visual Analytics Science and Technology (VAST), 2010 IEEE Symposium on*, pages 271–272. IEEE, 2010.

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

[115] F. Loizides. *Understanding and conceptualising the document triage process through information seekers' visual and navigational attention*. PhD thesis, City University London, 2012.

[116] A. M. MacEachren, R. E. Roth, J. O'Brien, B. Li, D. Swingley, and M. Gahegan. Visual semiotics & uncertainty visualization: An empirical study. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2496–2505, 2012.

[117] E. J. Maguire. *Systematising glyph design for visualization*. PhD thesis, Oxford University, UK, 2014.

[118] K. B. Marais and M. R. Robichaud. Analysis of trends in aviation maintenance risk: An empirical approach. *Reliability Engineering & System Safety*, 106:104–118, 2012.

[119] E. Mayr, N. Hynek, S. Salisu, and F. Windhager. Trust in information visualization. In *EuroVis Workshop on Trustworthy Visualization (TrustVis)*, volume 20191187, 2019.

[120] M. Meyer and J. Dykes. Criteria for rigor in visualization design study. *IEEE Transactions on Visualization and Computer Graphics*, 2019.

[121] M. Monroe, R. Lan, H. Lee, C. Plaisant, and B. Shneiderman. Temporal event sequence simplification. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2227–2236, 2013.

[122] T. Munzner. A nested model for visualization design and validation. *IEEE Transactions on Visualization and Computer Graphics*, 15(6), 2009.

[123] T. Munzner. *Visualization analysis and design*. CRC press, 2014.

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

[125] C. Neustaedter, A. B. Brush, and M. A. Smith. Beyond" from" and" received" exploring the dynamics of email triage. In *CHI'05 extended abstracts on Human factors in Computing Systems*, pages 1977–1980, 2005.

[126] C. North. Toward measuring visualization insight. *IEEE Computer Graphics and Applications*, 26(3):6–9, 2006.

[127] B. Ondov, N. Jardine, N. Elmqvist, and S. Franconeri. Face to face: Evaluating visual comparison. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):861–871, 2018.

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

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

[130] L. Padilla, M. Kay, and J. Hullman. Uncertainty visualization.

[131] S. Palan and C. Schitter. Prolific. ac—a subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, 17:22–27, 2018.

[132] H. Park and J. Choi. V-model: a new innovative model to chronologically visualize narrative clinical texts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 453–462. ACM, 2012.

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

[134] X. Qin, T. Kakar, S. Wunnava, B. MacCarthy, A. Schade, H. Q. Tran, B. Zylich, E. Rundensteiner, L. Harrison, S. Sahoo, et al. Mediar: Multi-drug adverse reactions analytics. In *2018 IEEE 34th International Conference on Data Engineering (ICDE)*, pages 1565–1568. IEEE, 2018.

[135] X. Qin, T. Kakar, S. Wunnava, E. A. Rundensteiner, and L. Cao. MARAS: Signaling multi-drug adverse reactions. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1615–1623. ACM, 2017.

[136] Z. Qu and J. Hullman. Evaluating visualization sets: Trade-offs between local effectiveness and global consistency. In *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization*, pages 44–52. ACM, 2016.

[137] R. Rao and S. K. Card. Table lens: Merging graphical and symbolic representations in an interactive focus+ context visualization for tabular info. In *SIGCHI Conference on Human Factors in Computing Systems*, pages 318–322. ACM, 1994.

[138] A. Rind, T. D. Wang, W. Aigner, S. Miksch, K. Wongsuphasawat, C. Plaisant, B. Shneiderman, et al. Interactive information visualization to explore and query electronic health records. *Foundations and Trends® in Human–Computer Interaction*, 5(3):207–298, 2013.

[139] 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.

[140] S. Robertson, H. Zaragoza, et al. The probabilistic relevance framework: Bm25 and beyond. *Foundations and Trends® in Information Retrieval*, 3(4):333–389, 2009.

[141] 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.

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

[143] T. Sakaeda, A. Tamon, K. Kadoyama, and Y. Okuno. Data mining of the public version of the fda adverse event reporting system. *International journal of medical sciences*, 10(7):796, 2013.

[144] B. Saket, A. Endert, and J. Stasko. Beyond usability and performance: A review of user experience-focused evaluations in visualization. In *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization*, pages 133–142, 2016.

[145] B. Saket, C. Scheidegger, and S. Kobourov. Comparing node-link and node-link-group visualizations from an enjoyment perspective. In *Computer Graphics Forum*, volume 35, pages 41–50. Wiley Online Library, 2016.

[146] P. Saraiya, C. North, and K. Duca. An evaluation of microarray visualization tools for biological insight. In *IEEE Symposium on Information Visualization*, pages 1–8. IEEE, 2004.

[147] G. K. Savova, J. J. Masanz, P. V. Ogren, et al. 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, 2010.

[148] J. Schrammel, M. Leitner, and M. Tscheligi. Semantically structured tag clouds: an empirical evaluation of clustered presentation approaches. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2037–2040. ACM, 2009.

[149] M. Sedlmair, M. Meyer, and T. Munzner. Design study methodology: Reflections from the trenches and the stacks. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2431–2440, 2012.

[150] E. Segel and J. Heer. Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1139–1148, 2010.

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

[152] N. Serrano and I. Ciordia. Bugzilla, itracker, and other bug trackers. *IEEE Software*, 22(2):11–13, 2005.

[153] B. Shneiderman, C. Plaisant, and B. W. Hesse. Improving healthcare with interactive visualization. *Computer*, 46(5):58–66, 2013.

[154] 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.

[155] Y. B. Shrinivasan, D. Gotzy, and J. Lu. Connecting the dots in visual analysis. In *Visual Analytics Science and Technology, 2009. VAST 2009. IEEE Symposium on*, pages 123–130. IEEE, 2009.

[156] Y. B. Shrinivasan and J. J. van Wijk. Supporting the analytical reasoning process in information visualization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1237–1246. ACM, 2008.

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

[158] M. Skeels, B. Lee, G. Smith, and G. G. Robertson. Revealing uncertainty for information visualization. *Information Visualization*, 9(1):70–81, 2010.

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

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

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

[162] J. Stasko, C. Görg, and Z. Liu. Sensemaking across text documents: human-centered, visual exploration with jigsaw. In *Sensemaking Workshop@ CHI 2008*, 2008.

[163] A. Strehl, J. Ghosh, and R. Mooney. Impact of similarity measures on web-page clustering. In *Workshop on artificial intelligence for web search (AAAI 2000)*, volume 58, page 64, 2000.

[164] H. Strobelt, D. Oelke, B. C. Kwon, T. Schreck, and H. Pfister. Guidelines for effective usage of text highlighting techniques. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):489–498, 2015.

[165] H. Strobelt, D. Oelke, B. C. Kwon, T. Schreck, and H. Pfister. Guidelines for effective usage of text highlighting techniques. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):489–498, 2016.

[166] H. Strobelt, D. Oelke, C. Rohrdantz, A. Stoffel, D. A. Keim, and O. Deussen. Document cards: A top trumps visualization for documents. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1145–1152, 2009.

[167] N. Sultanum, D. Singh, M. Brudno, and F. Chevalier. Doccurate: A curation-based approach for clinical text visualization. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):142–151, 2018.

[168] K. Techapichetvanich and A. Datta. Visar: A new technique for visualizing mined association rules. In *International Conference on Advanced Data Mining and Applications*, pages 88–95. Springer, 2005.

[169] R. S. Thompson, E. M. Rantanen, W. Yurcik, and B. P. Bailey. Command line or pretty lines?: comparing textual and visual interfaces for intrusion detection. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, page 1205. ACM, 2007.

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

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

[172] F. Van Ham, M. Wattenberg, and F. B. Viégas. Mapping text with phrase nets. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1169–1176, 2009.

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

[174] F. B. Viégas and M. Wattenberg. Timelines tag clouds and the case for vernacular visualization. *interactions*, 15(4):49–52, 2008.

[175] 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.

[176] M. Wattenberg and F. B. Viégas. The word tree, an interactive visual concordance. *IEEE Transactions on Visualization & Computer Graphics*, (6):1221–1228, 2008.

[177] W. Willett, J. Heer, and M. Agrawala. Scented widgets: Improving navigation cues with embedded visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1129–1136, 2007.

[178] 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*, pages 120–123. IEEE, 1999.

[179] K. Wongsuphasawat, J. A. Guerra Gómez, C. Plaisant, T. D. Wang, M. Taieb-Maimon, and B. Shneiderman. Lifeflow: visualizing an overview of event sequences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1747–1756. ACM, 2011.

[180] W. Wright, D. Schroh, P. Proulx, A. Skaburskis, and B. Cort. The sandbox for analysis: concepts and methods. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 801–810. ACM, 2006.

[181] S. Wunnava, X. Qin, T. Kakar, et al. 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, 2017.

[182] S. Wunnava, X. Qin, T. Kakar, X. Kong, E. A. Rundensteiner, S. K. Sahoo, and S. De. One size does not fit all: An ensemble approach towards information extraction from adverse drug event narratives. pages 176–188, 2018.

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

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

[185] N. y Henr, A. Bezerianos, and J.-D. Fekete. Improving the readability of clustered social networks using node duplication. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1317–1324, 2008.

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

[187] P. Yildirim, M. Bloice, and A. Holzinger. Knowledge discovery and visualization of clusters for erythromycin related adverse events in the fda drug adverse event reporting system. In *Interactive Knowledge Discovery and Data Mining in Biomedical Informatics*, pages 101–116. Springer, 2014.

[188] M. Zairi. Managing customer dissatisfaction through effective complaints management systems. *The TQM magazine*, 12(5):331–337, 2000.

[189] L. Zhang, Y. D. Zhang, P. Zhao, and S.-M. Huang. Predicting drug–drug interactions: an FDA perspective. *The American Association of Pharmaceutical Scientists journal*, 11(2):300–306, 2009.