Advanced Visual Analytics Interfaces for Adverse Drug Event Detection

Sebastian Mittelstädt
University of Konstanz
mittelstaedt@dbvis.de

Ming C. Hao
Hewlett Packard
ming.hao@hp.com

Umeshwar Dayal
Hewlett Packard
umeshwar.dayal@hp.com

Mei-Chun Hsu
Hewlett Packard
meichun.hsu@hp.com

Joseph Terdiman
joe.terdiman@gmail.com

Daniel A. Keim
University of Konstanz
keim@dbvis.de

ABSTRACT

Adverse reactions to drugs are a major public health care issue. Currently, the Food and Drug Administration (FDA) publishes quarterly reports that typically contain on the order of 200,000 adverse incidents. In such numerous incidents, low frequency events that are clinically highly significant often remain undetected. In this paper, we introduce a visual analytics system to solve this problem using (1) high scalable interfaces for analyzing correlations between a number of complex variables (e.g., drug and reaction); (2) enhanced statistical computations and interactive relevance filters to quickly identify significant events including those with a low frequency; and (3) a tight integration of expert knowledge for detecting and validating adverse drug events. We applied these techniques to the FDA Adverse Event Reporting System and were able to identify important adverse drug events, such as the known association of the drug Avandia with myocardial infarction and Seroquel with diabetes mellitus, as well as low frequency events such as the association of Boniva with femur fracture. In our evaluation, we found over 90% of the adverse drug events that were published in the Institute for Safe Medication Practices (ISMP) reports from 2009 to 2012. In addition, our domain expert was able to identify some previously unknown adverse drug events.

Categories and Subject Descriptors

H.5.m. [**Information Interfaces and Presentation (e.g. visual analytics)**]: Miscellaneous

General Terms

Design

©Authors 2014. This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive version was published in Source Publication, http://dx.doi.org/10.1145/2598153.2598156.

Keywords

Adverse drug event detection; visual analytics; pixel-based technique

1. INTRODUCTION

An Adverse Drug Event (ADE) is defined by the Institute of Medicine as "an injury event resulting from medical intervention related to a drug" [16]. Each year, about 98,000 drug-related deaths are reported in the USA. Even after a successful clinical trial in which a drug has been cleared for marketing to the medical community, unsuspected ADEs are occasionally detected. This is due to the fact that the trials are usually limited to short time periods and include only a small number of patients. Also, the frequency of the adverse drug events may be so low that they are not detected in the clinical trials. For example, femur fracture (see Figure 1) was not recognized as an adverse reaction to Boniva until after the drug was prescribed to thousands of patients [23].

There exist several sophisticated approaches and tools for detecting significant adverse events; however, most current tools are primarily capable of detecting events with frequencies above predefined thresholds. Thus, low frequency events are often overlooked. The decision of significance in about whether an event with a close-to-noise frequency is clinically significant cannot be done by automatic systems alone. Another type of problem is confounding by indication. For example, insulin is administered for diabetes. Myocardial infarction is a common disease for patients with diabetes and thus, detecting the event myocardial infarction for insulin is a false positive ("confounding effect"). For automatic systems, this domain knowledge has either to be manually included or automatically extracted from appropriate data. The manual approach will increase the accuracy of results. However, it will require significant effort from both, domain experts and data analysts. These issues (low-frequency, automatic compensation for confounding effects) make this domain problem a challenging research problem.

Our Contribution: We propose to use visual analytics for this domain problem. The goal of this paper is to illustrate an expert system and a means for domain experts to detect low frequent, but significant unexpected data features in large heterogeneous databases. In contrast to existing techniques, we introduce a novel solution to identify and validate important adverse drug events just above the noise level with a

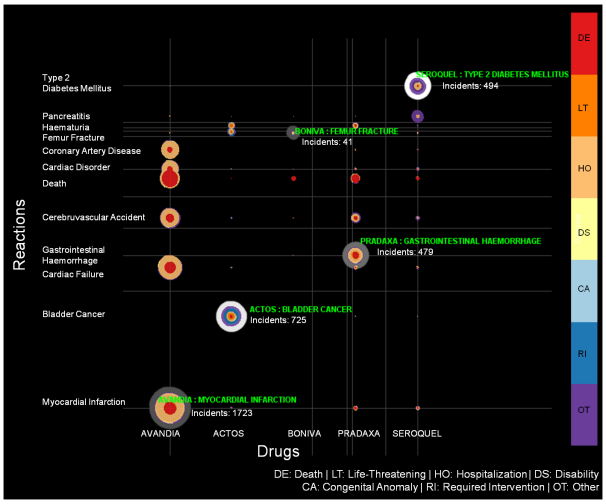


Figure 1: A non-overlapping x-y pixel interface for detecting adverse drug events (x-axis: drug name, y-axis: adverse reaction, color: outcome, white: signal strength). Each pixel represents one adverse drug-reaction event, whose color encodes the outcome in the AERS database in 2011 (3rd quarter). The visualization shows a number of known and some previously unknown adverse drug events: The known association of the drug Avandia with myocardial infarction [19] occurred 1,723 times with 280 associated deaths. The association of Actos with bladder cancer shows a high signal (white circle), which was described in the ISMP report in 2012 [14]. Boniva has a low frequency association with femur fracture (41 adverse events) and 42 associated deaths [12], an adverse drug event which has been detected by our technique. Also, it has been recently discovered that Pradaxa causes gastrointestinal hemorrhages resulting in death and hospitalization [13].

tightly integrated domain expert. We claim the following contributions: 1) A visual analytics approach to access massive volumes of events by interactive relevance filtering; 2) Detection and validation of low frequency events by enhanced statistical computations and interactive analysis; 3) Elimination of confounding effects by using discriminative heuristics.

2. RELATED WORK

The ability to detect ADEs in patient records is essential in healthcare. Numerous detection techniques of ADEs have been developed over the past years, such as Bayesian classification, decision trees, category association, and data visualization. Some of these techniques are closely related to our work. They can be classified into two categories: data mining and visual analytics.

Data Mining Methods: Algorithms for finding drug-event associations typically rely on first detecting frequent item sets in the data. Association rule-finding algorithms have been heavily discussed in the database literature in the past 15 years. Since 1998, the FDA [21] has been exploring automated Bayesian data mining methods using the Multi-Item Gamma Poisson Shrinker (MGPS) program that computes signal scores for any combination of drugs and reactions that are significantly more frequent than their previous pair-wise associations. In 2003, a review of software packages for data mining was published by Haughton et al. [10], which showed that SAS Enterprise Miner has the best coverage of statistical methods.

In France, Chazard et al. [7] began to use decision trees with associations between ADEs and potential ADE to identify adverse drug events. In 2011, Chazard et al. [6] used a common data model including diagnoses, drug administrations, laboratory results, and text records to generate association rules to automatically detect adverse drug events. Chazard's decision tree generates useful rules for physicians to follow. In a study of 115,447 complete past hospital stays extracted from 6 French, Danish and Bulgarian hospitals, 236 ADE-detection rules were generated and validated by a committee of experts.

Visual Analytics: In 2003, Atherton [1] used event charts to summarize individual patient data and to display clinically significant changes over time. Different from traditional histograms, bar charts and plots, Chazard et al. [8] began to use treemaps [20] to represent medical data.

The Export Explore [2] is a tool designed by Idea Advertising to validate the clinical cases and the assessment of the decision rules to detect adverse drug events. It uses tables and charts to display diagnoses, drugs, and lab results. In 2011, Marcilly et al. [17] designed novel adverse drug-event scorecards to provide hospitals with summary information about adverse drug events on the web. The data used in the Scorecards is routinely updated and report ADEs detected by data mining processes. This work is an ongoing project in the Department of Public Health in France.

Wongsuphasawat and Shneiderman [25] introduced Similan, an interactive tool that facilitates similarity searching and visualization for temporal categorical data. Their plot panel and generalized scatter plot [15] provide an overall distribution of search results to help users quickly find similar records from temporal categorical data. They designed a new LifeFlow system [24] to support an interactive visualization for retrieving event sequences based on similarity for hospital physicians to observe interesting events.

In 2012, Miselbauer et al. employed the Smart Super Views [18], which use fuzzy logic rule visualization techniques to enable clinicians to analyze different spatial regions in medical images based on their domain knowledge. Miselbauer uses a relevance driven user interface to arrange the generated views, which is similar to our technique to determine relevant drug-reaction pairs from huge amounts of FDA events. The key difference is that Miselbauer selects relevant medical images based on image recognition and we select important relevant drug-reaction pairs according to their significance.

While existing data mining methods and visualizations have produced useful results, they are limited by their ability to detect signals that are well above the background noise level and may suffer under false positive associations that occur by chance or are due to confounding. Our method allows a detection of ADE, which are much closer to but above the noise level and reduces confounding effects.

3. EVENT DETECTION ALGORITHM & VISUAL INTERFACES

In order to create a visual analytics system for the needs of medical investigators, we incorporated domain experts

iteratively into the research and development process. The two domain experts were not affiliated with the drug industry, but familiar with the clinical use of drugs. They are also familiar with ADE detection as part of post-marketing drug surveillance, the FDA's efforts to monitor ADEs and traditional epidemiological studies of ADEs.

The authors and domain experts were located in different countries and therefore, video conferences were the basis for collaboration. Weekly one hour meetings were set up for two months. In the first phase of our study, we discovered the available data source and the domain problem of detecting low frequency but relevant ADE. We researched the current literature and state-of-the-art tools of the domain and further discussed with the experts where they see potential gaps and improvements. On this basis, we started to implement fast alternative solutions with state-of-the-art visual analysis tools to demonstrate the effectiveness of "non-standard" visualization techniques on the domain problem. These solutions helped in the discussions to discover concrete tasks and requirements and also to narrow down the solution space. Paper mock-ups were designed to meet these requirements and discussed with the experts. The finalized designs were implemented in an iterative process, in which the domain experts were weekly updated with new features and findings. Presentations, findings, and prototypes were carefully prepared, such that the experts could analyze and validate findings during the meetings. In this way, we felt that the domain experts were convinced of the tool and gave concrete feedback. We also discovered tasks that were not clear in beginning. The feedback from the domain experts were carefully analyzed to enhance the system. For the final evaluation, they were invited to test the system.

3.1 Problem Characterization and Abstraction

The FDA Adverse Event Reporting System (AERS) data source publishes numerous adverse drug event reports in each quarter. These reports contain demographic information about the patient (age, gender etc.), the drug that was used during the therapy (e.g., Avandia) and the patient reaction (e.g., myocardial infarction). The drug-reaction pair is considered an adverse drug event. The indication (why the drug has been administered) and the outcome (such as hospitalization or death) are also recorded. In each quarter hundreds of thousands of reports are published. The large multivariate data set is a valuable source of information and the domain applies sophisticated automatic methods to detect adverse drug events. However, our experts describe that these methods often overlook low frequency events.

On the basis of the FDA data, our domain experts were able to define concrete **analysis tasks**: 1) Explore relevant adverse events of drugs and signals for "unexpected" adverse reactions. 2) Explore gender and age distribution for ADE. 3) Summarize demographic information. 4) Manual validation of signals.

To abstract the domain problem, we see that the experts are faced with a large multidimensional data set, where automatic methods overlook low frequent correlations. Further, correlations may lack causality or are false positives. Therefore, the solution requires an expert validation of results. We identified the **requirements** of the solution as follows: 1) Interactive filtering of relevant information must be supported by automatic methods. 2) Low frequency events must be detected by automatic methods. 3) Visual interfaces and analysis techniques must highlight relevant candidates and compensate for the bias of inconsistencies and false positives (confounding effects). 4) Enable manual validation of events by providing the full record of events in the database.

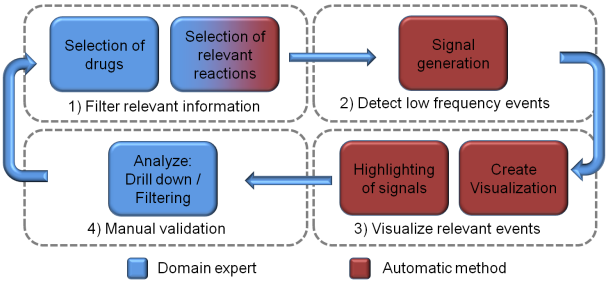


Figure 2: Visual Analytics process for detecting low frequency adverse drug events.

3.2 Concept and Idea

Based on the given tasks, requirements, and the assumption that automatic methods are outperformed by human experts, we decided to integrate the analyst as soon as possible in the analysis process. The complexity of the task has to be reduced into several steps (as illustrated in Figure 2): First, the entry point for the analysis is to be found and relevant information is extracted from the database. Our system enables exploratory (without hypothesis) and hypothesis driven approaches where an initial set of target drugs is selected. Relevant drug-reaction events are carefully separated from the noise level in order to reduce complexity. In the second step, statistical measures and heuristics are applied to identify significant adverse drug events. All drug-reaction events are then visualized in an x-y plane interface, where significant events are highlighted. The pixel-based interface enables the user to filter and drill down interactively in order to validate events or to form new hypothesis and thus, the user may start the process again.

3.3 Filtering of Relevant Events

In order to exclude confounding effects and to identify significant adverse events, the selected drugs should ideally be administered for the same indication. This can be done by querying for the most frequent co-occurring indications for drugs and vice versa from the AERS data base. The results are represented to the user for further refinement. Thus, drugs can be selected by: 1) **Drug class**: defined by domain experts; 2) **indication**: drugs that are administered for the same indication are determined automatically (querying for the drugs that co-occur frequently with the selected indication); 3) **specific drug**: the indication of the drug is determined automatically by querying for the most co-occurring indication (and then proceed as in 2).

For example, in Figure 3 the user is interested in the adverse events of the drug Avandia. In Figure 3 (B), the system suggests the most frequent co-occurring indications with Avandia, which is diabetes. In Figure 3 (C), the system queries the database for the most co-occurring drugs for diabetes in order to find the drugs for the same treatment to exclude confounding effects. The user selects a meaningful subset or searches for other drugs, since aspirin and lisinopril are not used to treat diabetes. The selection is used for

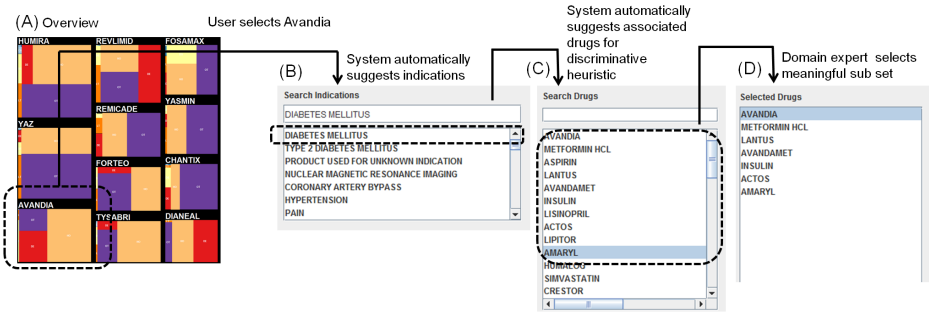


Figure 3: Semi-automatic approach to select relevant drugs for the discriminative signal generator.

signal generation and visualization to detect adverse drug events (see Figure 5).

In addition, we provide an exploratory approach, which is based on an overview of the whole dataset and allows the user to search and filter for patterns and anomalies in order to generate hypotheses for further investigation. Figure 4 (A) visualizes all reports in 2011 in a scatter plot where interesting age and temporal patterns become visible, and (B) shows the report frequency and patient outcome for the 100 most frequently reported drugs in a treemap overview, where Avandia shows a high association to severe patient outcomes that indicates the need for further investigation. The user can interactively select interesting drugs in the overview interfaces and initiate the visual analytics process as shown in Figure 3.

Separating relevant reactions from irrelevant is a critical step in our approach, because it may also remove relevant low frequency events. Our algorithm queries the database for all drug (d) - reaction (r) co-occurrences and measures their frequencies $f(d, r)$. Then it calculates the relevance $R(d, r) = f(d, r)/n(d)$ with the frequency $f(d, r)$ divided by the number of records $n(d)$ containing drug d.

If the relevance is below a user defined threshold t, the drug-reaction pair will be removed. A threshold of $t = 0.01$ removes all adverse events that occur in less than 1% of reports associated with a drug and thus, are considered as noise. The user can select a different sensitivity interactively. All relevant reactions of the currently selected drugs form the set of reactions that are further analyzed and visualized.

3.4 Signal Generation

Many related automatic systems are based on methods that scan the database and generate signals [21]. A signal is an event (drug-reaction pair), whose frequency in a patient population is higher than expected. There are two disadvantages to this strategy: First, low frequency events may be lower than some thresholds and never generate a signal. Second, confounding effects may lead to highlighting candidates that are expected because of the intended medical use of the drug. These events are usually of high frequency and thus, may mask unknown events with a lower frequency.

In order to avoid these issues, we follow a higher than expected approach by significance ordering, which highlights the n most significant events for each drug (these candidates are called signals from now on) and informs the domain expert of potential (even low frequency) events.

A well established general statistic for determining the significance of categorical data is the chi-square statistic. It will generate a signal if the co-occurrence of X and Y is higher or lower than expected. Since it also creates signals for co-occurrence of events less frequent than expected, chi-square is not suitable for our task. Odds ratio "have become widely used in medical reports. They are primarily used for estimating the relationship between two binary variables" [4]. We are using the contingency table of our drug-reaction pairs to apply our significance statistics, which is based on odds ratio (1). The intuition behind the formula is: if reaction Y co-occurs with drug X more frequently than with other drugs, then a and d will be high and b and c will be low. Note, that in the formula we add one to b and c to avoid a division by zero. We also increased the influence of a, since it represents the co-occurrence of the drug-reaction pair. The reaction can now be ordered according to their odds. The drug-reaction pairs with the highest odds are highlighted in the visual interface in order to steer the analyst's attention to these potentially valid and unknown ADEs for validation.

$$odds(x, y) = \frac{a^2 \cdot d}{(b+1) \cdot (c+1)} \tag{1}$$

	Reaction Y	**Other Reactions**
Drug X	a	b
Other Drugs	c	d

Table 1: Contingency table (e.g., a := co-occurrence of X and Y).

The "other" (in Table 1) could be defined as all drugs in the database (global). Another way is to only include a certain group of drugs (local) by indications or drug classes. For example, if a member of the "hypoglycemic" drug class is compared to the entire database, it is known that these drugs are significantly associated with heart problems (e.g., myocardial infarction) compared to other drugs. Thus, detecting myocardial infarction is a confounding effect. However, if only the drugs in a class are considered as "others", these confounding effects will be compensated. Applying the significance measure (1) in this way reveals, for example, that myocardial infarction is only significant for Avandia (a hypoglycemic drug). This ADE was identified by the FDA in 2007. In summary, our significance measure works the best if drugs are grouped together with those are supposed to have similar reactions. Thus, our adapted odds ratio can be interpreted as a discriminative heuristic. Note, that we interpret the heuristic as an updateable module, which may be replaced by other means of significance ordering.

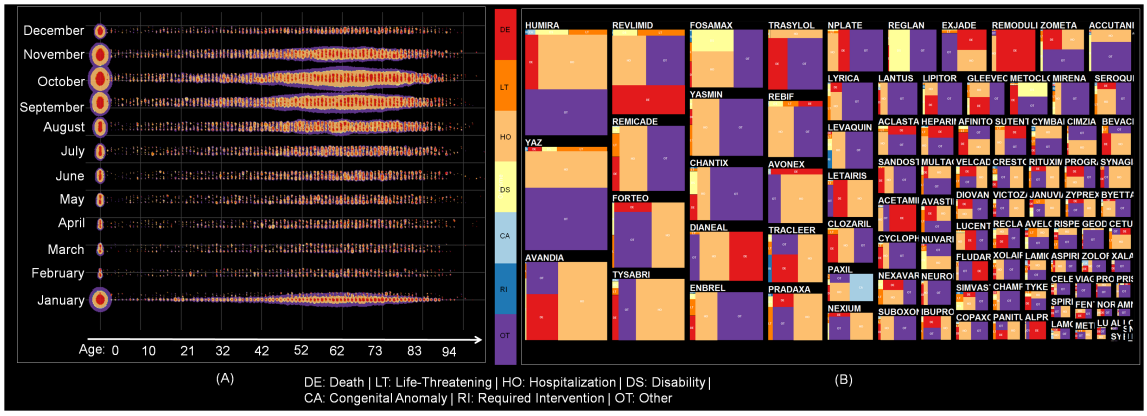


Figure 4: Overview of the whole data set of 2011 4th quarter. (A) Shows an interactive scatter plot (x-axis: age; y-axis: event month; color: outcome) Many reports are associated with middle aged patients in January and from August to November, which might be due to the cold season. The user is enabled to select interesting regions in the scatter plot. The selected drugs are then visualized in a treemap overview as in (B) and can be used as entry points for the analysis process. (B) Treemap overview of the 100 most reported drugs in this quarter or in the selection (size: frequency; color: outcome). Humira is mentioned in many reports but Pradaxa, Dianeal, and Avandia show higher frequency with death associations.

4. VISUAL REPRESENTATION & DESIGN DECISIONS

In order to meet the requirements, the system must provide the drugs under investigation, the co-occurring reactions, the frequency of the drug-reaction pair, and also the severity of each event. However, drug-reaction pairs are not the only interesting variables in the data set for experts to identify and validate adverse drug events. For example, timestamps and demographic information are also important variables to consider in identifying target populations in which ADEs occur. The system is intended for non-visual-analytics experts and thus, there is a need for intuitive and easy to use visual interfaces that are flexible enough to visualize different data types while maintaining the same look and feel.

In an exploratory approach, there are lots of possible and relevant drug-reaction candidates for further investigation, and it may happen, that the smallest of thousand drug-reaction pairs are the most interesting ones. Due to these requirements, there is a clear need for highly scalable visualizations techniques that also allow the perception of small emergent data properties.

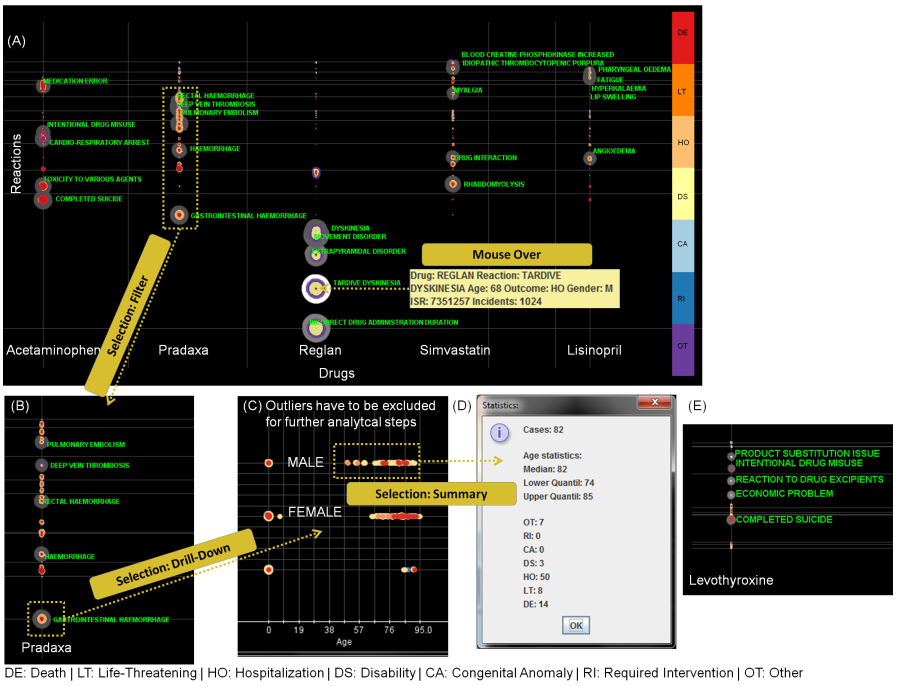
4.1 Visual Analysis and Validation of Events

Scatter plots are flexible enough to represent all combinations of variables and data types (categorical, ordinal, and continuous data), and to show data in different views and maintain the format of representation. However, both x- and y- axes can be categorical and there are multiple drug-reaction data points with the same x-y coordinates in a patient population. The pixels will highly overlap and an estimation of the number of events will not be possible. To avoid this, Keim et. al. [15] introduced repositioning of pixels to generate a high scalable and overlap-free representation. If there is more than one event with the same x-y coordinates, the system will use a circular pixel placement algorithm. It sorts the overlapping events according to their third dimension (e.g., outcome mapped to color) and places them around the center (x,y) coordinate in concentric rings. The area of each ring is proportional to the number of events with the particular outcome. For example, in Figure 1 Avandia has 1723 co-incidents with myocardial infarction that would overlap without pixel replacements. The remaining empty space in the categorical version is used to place labels at significant adverse events.

This technique allows visualizing large amounts of data. For example, more than 80,000 reports can be visualized on one display (Figure 4 (A)) and each individual report can be selected and queried on demand (Figure 5(A)). Patterns and anomalies can be easily recognized, such as the temporal and age patterns in Figure 4 (A). Drugs and reactions can be grouped by similarity in the categorical plot (see below). Thus, the domain experts can detect and eliminate the effects of inconsistencies and confounding effects. Another advantage is that this technique offers the ability to encode a third dimension with color, such as the severity of the event.

Interactive Analysis and Drill Down: The interactive analysis and drill down functionalities are illustrated and explained in Figure 5. (A) Shows the overview of the events for drugs in the ISMP report of the 1st quarter of 2011 (see also Table 2). Since all records are visualized as pixels on the display, the user can demand details about every single event by mouse over (here: within 1024 reports for Reglan co-occurring with tardive dyskinesia; the selected patient at the age of 68 was sent to hospital). Pradaxa shows a high frequency of associated patient deaths and signals for hemorrhage. The user is able to select and zoom into this point of interest (B). The user drills down to gastrointestinal hemorrhage to visualize detailed demographic information about the patients in (C) in order to identify potential target groups (x-axis: age, y-axis: gender). The user is able to select again a meaningful subset, (D) revealing that 82 incidents with 14 deaths are associated to male patients in the 74-85 year range. (E) shows the low frequency signals for levothyroxine, which can be detected in addition to the dominant drugs in (A).

Ordering and clustering of drugs and reactions: Drugs and reactions have to be ordered to create effective



DE: Death | LT: Life-Threatening | HO: Hospitalization | DS: Disability | CA: Congenital Anomaly | RI: Required Intervention | OT: Other

Figure 5: Interactive exploration and validation of events.

adverse drug event visualizations. This can be done by generating a feature vector for each drug that consists of drug-reaction pair frequencies. These vectors and their dimensions can be ordered by similarity based on a hierarchical single-linkage clustering. In this way, similar drugs and reactions are grouped together. In case name inconsistencies result in multiple instances of the same drug, it is likely that those instances will have similar appearances in the visualization. Also drug classes as well as subclasses become visible and can be further analyzed or combined.

4.2 Data Overviews and Visual Comparison of Ratios

In the design process, we discovered that the overview functionality of the scatter plot is not enough to meet the requirements of our domain experts. For the exploratory approach the overview has to visualize the number of reports about certain drugs and the ratio of the patient outcomes at one glance such that interesting ratios can be perceived. The scatter plot will suffer under huge over-plotting with large categorical data and is not suitable for this task. Therefore, we proposed standard techniques such as bar charts, pie charts and glyph representations. We, however, excluded pie charts and glyphs due to the lack of accuracy and intuitiveness. Bar charts are very accurate, however, with six categorical outcomes and hundreds of drugs this results in an overloaded display. Further, comparing the number of reports over different drugs requires cognitive aggregation of all outcome categories. Stacked bar charts and treemaps are well suited for this task. Both techniques scale well and allow the comparison of ratios. Since treemaps have already found their way into the medical community [8], we propose squarified treemaps as a space filling technique for the overview that allows accurate ratio comparison [5]. We use size to visualize the frequency of events. The hierarchy encodes drugs on the first level and corresponding outcomes on the second level (see Figure 4). This enables the analyst to compare efficiently the number of reports for the drugs and also their associated outcomes. The main drawback of the technique is that absolute values cannot be encoded, which we solve with interaction (mouse over) and tooltips.

4.3 Visual Encoding

After the categorical drug-reaction information, the "outcome" is the most important dimension to reveal the severity of an adverse drug event, which scales from common medical investigation to the death of patients. The data type of this variable can be interpreted as ordinal rather than categorical since the outcomes can be ordered according to the severity of the event. We see different constraints for an appropriate color mapping. 1) The data is ordered from low to high severity of ADE, which is usually coded by an uniform scale or multi-hue scale with linear increasing lightness or saturation [3]. This would highlight saturated or bright ADE in the visualization and attract the analyst's attention towards more severe events. 2) However, steering of attention can be harmful in our particular task, since the system must treat each ADE equal in the visualization, because the impact of real significant but unexpected ADE can only be estimated by the expert and not by the system. Therefore, saturation and lightness should not be used for the encoding.

Categorical color scales have the disadvantage that the natural ordering of hues varies widely from subject to subject. Semantics and the use of metaphors, however, can ensure correct ordering. Our choice is the temperature metaphor and thus, we vary from cold colors (violet, blue, cyan) to warm colors (yellow, orange, red). This also concerns the western culture that associates severe events with red. Studies have shown that humans are more attracted to warm colors than they are to cold colors, which suppresses cold color that

Drug	Medical Use	Approved ADE	Top 5 Signals				
PRADAXA	Atrial Fibrillation	Hemorrhage, Gastrointestinal Hemorrhage	Gastrointestinal Hemorrhage	Pulmonary embolism	Hemorrhage	Rectal Hemorrhage	Deep vein thrombosis
REGLAN	Nausea	Tardive Dyskinesia	Tardive Dyskinesia	Extrapyramidal Disorder	Incorrect Duration	Movement Disorder	Dyskinesia
ACETAMINOPHEN-HYDROCODONE	Pain	Overdose / Suicide	Completed Suicide	Intentional Drug Misuse	Medication Error	Cardio Respiratory Arrest	Toxicity to various Agents
LEVOTHYROXINE	Hormone replacement	Overdose / Suicide	Completed Suicide	Intentional Drug Misuse	Product Substitution Error	Reaction To Drug Excipients	Economic Problem
SIMVASTATIN	Lipid-lowering	Muscle Damage	Rhabdomyolysis	Drug Interaction	Blood creatine increased	Myalgia	Idiopathic thrombocyto-penic purpura
LISINOPRIL	Hypertension	Hypersensitivity	Angioedema	Fatigue	Hyperkalae-mia	Lip Swelling	Pharyngeal Edema

Table 2: Evaluation of the signal generation (AERS database: 1st quarter of 2011). Left: the drug, its medical use and its major adverse drug effect that has been discussed in the ISMP report [13]. Right: The top 5 signals that our system generates without user interaction and domain knowledge (green events are approved by ISMP reports).

are spatially close to warm colors [22]. Hence, we must decrease the effect by ordering the records in the circular layouting algorithm by decreasing severity from inner to outer radius. Thus, the inner (severe) records are more perceptually striking, but do not cover larger parts of the visualization, which reduces the bias for the analyst.

Due to the limitations of human color perception [11], it is most effective if only few qualitative or diverging colors are used to represent a categorical or ordinal variable [9]. We adapted our colors from a categorical scale of Brewer et al. [9], because in terms of attention and separation, they fulfill the second constraint. We selected a set of three cold colors in the hue range from 180° to 270° and four warm colors from 0° to 60°.

4.4 Highlighting

To show the statistical significance of each adverse drug event, we use white highlighting circles. This has two reasons: First, this increases the size and thus, the visual importance of the highlighted events in the visualization, which enhances the visibility of low frequency events. Second, we are able to encode the level of significance of the adverse event with different shades of gray. We are aware of the potential bias that the outer ring can have on the other categorical "outcome" colors. However, since we are only using seven colors that can be clearly separated, we consider the risk as acceptable. For further highlighting, we place labels next to the significant events.

5. APPLICATION & EVALUATION OF RESULTS

In order to demonstrate the effectiveness of our system, we evaluated our automatic signal generation on the AERS database of the years 2009-2012. This allowed us to validate the generated signals and findings since the ISMP (Institute for Safe Medication Practices) reports of known adverse drug events were available. Further, we illustrate a use case with one dependent domain expert.

Table 2 and Figure 5 show the drugs that are mentioned in the ISMP report for the first quarter of 2011. The table reveals their medical use and validated major ADEs. We examined each drug in the corresponding quarter's database by using it as starting point for our system as shown in Figure 3. Our system was configured to automatically determine the indication from the database and thus, determined ten drugs for the discriminative heuristic in order to eliminate confounding effects. We excluded events that occur less than 1% for each drug by a relevance threshold $t = 0.01$ (see Section 3.3). We configured the system to generate signals for the five most significant reactions of each drug, such as hemorrhage for Pradaxa or tardive dyskinesia for Reglan (see Figure 5 (A)). Even the low frequency suicide events were detected for acetominophen and levothyroxine, for which only 313 reports existed in the whole database (with only six suicidal events). Expressing the complete results with precision-recall measures, our system detected all the major adverse drug events in the ISMP report of 2012 with a precision of 40%. We found that this performance (recall $> 90\%$ and precision $> 40\%$) holds for all the reports from 2009 to 2012. Since a domain expert is filtering the signals during the analysis process, a high recall is far more important than a high precision.

For the use case, one domain expert was given the task to detect and validate "unknown" adverse drug events in the AERS data. In the 3rd quarter of 2011 the FDA reported 198,777 adverse drug events from over 28,000 drugs that caused over 9,000 different reactions. The domain expert was involved in the design process and was already familiar with the interfaces and the visual analytics process. However, the ISMP reports mentioned below were not published when the expert performed the task and therefore, the expert was considered independent of the reports' insights.

The expert found several unexpected adverse drug events summarized in Figure 1. For example, the expert found that Actos, which is used to treat diabetes, has a significant association with bladder cancer (reported in 2012 by the ISMP [14]). Also the expert identified a low frequency association between Cymbalta and serious withdrawal symptoms. This finding was later described in the ISMP report of the 1st quarter of 2012 [14]. The expert reported that our system

and analysis strategy is very useful to detect low frequency adverse drug events and that there is a significant potential for visual analytics interfaces in post-marketing surveillance.

6. CONCLUSION

In this paper, we presented a visual analytics approach to access a massive volume of events by enhanced statistical computations and advanced interfaces to incorporate expert knowledge for validating the individual event by interactive relevance filtering. We demonstrated that experts are able to detect relevant low frequency events from massive volumes of adverse drug events, such as bladder cancer for Actos, hemorrhage for Pradaxa, and femur fracture for Boniva. Our next step is to continue our visual analytics work on drug - drug interactions and drug usage over time. These tasks pose new challenges to the analysis and will likely need different measures and additional visual interfaces.

7. ACKNOWLEDGMENTS

The authors would like to thank Nicholas Tatonetti for his feedback and advise on the system.

8. REFERENCES

[1] P. J. Atherton, B. Jasperson, A. Nibbe, K. A. Clement-Brown, C. Allmer, P. Novotny, C. Erlichman, and J. A. Sloan. What happened to all the patients? event charts for summarizing individual patient data and displaying clinically significant changes in quality of life data. *Drug information journal*, 37(1):11–21, 2003.

[2] A. Băceanu, I. Atasiei, E. Chazard, N. Leroy, et al. The expert explorer: a tool for hospital data visualization and adverse drug event rules validation. *Studies in health technology and informatics*, 148:85–94, 2009.

[3] L. D. Bergman, B. E. Rogowitz, and L. A. Treinish. A rule-based tool for assisting colormap selection. In *Proceedings of the 6th conference on Visualization'95*, pages 118–125. IEEE Computer Society, 1995.

[4] J. M. Bland and D. G. Altman. Statistics notes: the odds ratio. *BMJ: British Medical Journal*, 320(7247):1468, 2000.

[5] M. Bruls, K. Huizing, and J. J. Van Wijk. Squarified treemaps. In *Data Visualization 2000*, pages 33–42. Springer, 2000.

[6] E. Chazard, G. Ficheur, S. Bernonville, M. Luyckx, and R. Beuscart. Data mining to generate adverse drug events detection rules. *IEEE Transactions on Information Technology in Biomedicine*, 15(6):823–830, 2011.

[7] E. Chazard, C. Preda, B. Merlin, G. Ficheur, R. Beuscart, et al. Data-mining-based detection of adverse drug events. *European Federation for Medical Informatics*, 150:552–556, 2009.

[8] E. Chazard, P. Puech, M. Gregoire, and R. Beuscart. Using treemaps to represent medical data. *Studies in Health Technology and Informatics*, 124:522–527, 2006.

[9] M. Harrower and C. Brewer. Colorbrewer. org: an online tool for selecting colour schemes for maps. *The Cartographic Journal*, 40(1):27–37, 2003.

[10] D. Haughton, J. Deichmann, A. Eshghi, S. Sayek, N. Teebagy, and H. Topi. A review of software packages for data mining. *The American Statistician*, 57(4):290–309, 2003.

[11] C. G. Healey. Choosing effective colours for data visualization. In *Proceedings of Visualization'96.*, pages 263–270, 1996.

[12] ISMP QuarterWatch. Signals for varenicline, levofloxacin and fentanyl. http://www.ismp.org/quarterwatch/pdfs/2010Q2.pdf (visited on 10/06/2013), 2011.

[13] ISMP QuarterWatch. Signals for dabigatran and metoclopramide. http://www.ismp.org/quarterwatch/pdfs/2011Q1.pdf (visited on 10/06/2013), 2012.

[14] ISMP QuarterWatch. Why reports of serious adverse drug events continue to grow. http://www.ismp.org/quarterwatch/pdfs/2012Q1.pdf (visited on 10/06/2013), 2012.

[15] D. A. Keim, M. C. Hao, U. Dayal, H. Janetzko, and P. Bak. Generalized scatter plots. *Information Visualization*, 9(4):301–311, 2010.

[16] L. T. Kohn, J. M. Corrigan, M. S. Donaldson, et al. *To err is human: building a safer health system*, volume 627. National Academies Press, 2000.

[17] R. Marcilly, E. Chazard, M.-C. Beuscart-Zéphir, W. Hackl, A. Băceanu, A. Kushniruk, and E. M. Borycki. Design of adverse drug events-scorecards. *ITCH*, 164:377–381, 2011.

[18] G. Mistelbauer, A. Kochl, H. Bouzari, S. Bruckner, R. Schernthaner, M. Sramek, I. Baclija, and M. E. Groller. Smart super views - a knowledge-assisted interface for medical visualization. In *IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 163–172. IEEE, 2012.

[19] S. E. Nissen and K. Wolski. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *New England Journal of Medicine*, 356(24):2457–2471, 2007.

[20] B. Shneiderman. Tree visualization with tree-maps: 2-d space-filling approach. *ACM Transactions on graphics (TOG)*, 11(1):92–99, 1992.

[21] A. Szarfman, S. G. Machado, and R. T. O'Neill. Use of screening algorithms and computer systems to efficiently signal higher-than-expected combinations of drugs and events in the us fda's spontaneous reports database. *Drug Safety*, 25(6):381–392, 2002.

[22] L. Wang, J. Giesen, K. T. McDonnell, P. Zolliker, and K. Mueller. Color design for illustrative visualization. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1739–1754, 2008.

[23] D. Wilson. F.d.a. issues warning on bone drugs. http://prescriptions.blogs.nytimes.com/2010/10/13/f-d-a-issues-warning-on-bone-drugs/ (visited on 06/06/13), 2010.

[24] K. Wongsuphasawat, C. Plaisant, M. Taieb-Maimon, and B. Shneiderman. Querying event sequences by exact match or similarity search: Design and empirical evaluation. *Interacting with computers*, 24(2):55–68, 2012.

[25] K. Wongsuphasawat and B. Shneiderman. Finding comparable temporal categorical records: A similarity measure with an interactive visualization. In *IEEE Symposium on Visual Analytics Science and Technology*, pages 27–34. IEEE, 2009.