Drug-Induced Liver Injury (DILI) Classification using US Food and Drug Administration (FDA)-Approved Drug Labeling and FDA Adverse Event Reporting System (FAERS) data

Qais Hatim[1], Minjun Chen[1], Eileen Navarro Almario[1], Monica Munoz[1], Allen Brinker[1], Marc Stone[1], Sonja Brajovic[1], Kendra Worthy[1], Lilliam Rosario[1], Tom Sabo[2], Emily McRae[2], Soundar Kumara[3]

1 U.S. Food and Drug Administration, 2 SAS Institute Inc., 3 Pennsylvania State University/University Park

ABSTRACT

Defining DILI positive and negative is challenging, which needs to consider the causality, incidence, and severity of the liver injury events caused by each drug. The previous approach, based on the FDA approved drug labels, partly considered these issues and classified the drugs into most-, less-, and no-DILI-concern categories. We incorporated the causality assessment information from literature with the drug label based approach and developed a new approach to classify drugs into Most-, Less-, and No-DILI-concern plus a group of drugs as ambiguous DILI, which causality were not confirmed by literature reports (Minjun Chen 2016). The FDA FAERS database provides comprehensive post-marketing surveillance data; it is therefore prudent to improve the DILI classification by integrating the post-marketing data into the drug-label based approach to further improve the accuracy of DILI classifications, which subsequently could further refine model development for better predicting DILI in humans.

INTRODUCTION

Many drugs have either been discontinued from clinical trials or withdrawn from the market after being approved because of hepatic adverse effects (Maddrey 2005) & (Senior 2007). Some of these adverse events can be serious in nature as evidenced by drug-induced liver injury (DILI) being listed as the leading cause of acute liver failure in the US (Ostapowicz G and Group. 2002). Thus, DILI has become one of the most important concerns in the drug development and approval process (Kaplowitz 2001). DILI has also been identified by the FDA Regulatory Science Initiatives as a key area of focus in a concerted effort to broaden the agency's knowledge for the better evaluation of tools and safety biomarkers (http://www.fda.gov/ScienceResearch/SpecialTopics/RegulatoryScience/ucm228131.htm). Some drugs are more likely to cause hepatotoxicity or liver injury than others, and severe DILI is of most concern. The FDA published guidelines in 2009 for assessing the potential for a drug to cause severe DILI in premarketing clinical evaluation (CDER 2009).

The toxicological community has made great efforts in developing biomarkers and methodologies to assess hepatotoxicity, including DILI beyond classical animal testing, for all chemicals. The representative methods include, but are not limited to, QSAR assessments (Rodgers 2010), in vitro assays (Obach 2008), high-content screening assays (Xu 2008) and 'omics' studies (Zidek 2007). Some of these approaches are being evaluated by large government-initiated efforts for developing alternative methodologies for toxicity assessment, such as Tox21 (Shukla 2010) and ToxCast (Benigni 2010) in the USA, and the REACH program

(Schoeters 2010) in Europe. These efforts require a list of drugs with well-annotated DILI potential to guide the methodology development and assess their performance characteristics (i.e. sensitivity and specificity) (Temple 2006).

A drug classification scheme is essential to facilitate the community-wide effort to evaluate the performance characteristics of existing DILI biomarkers and discover novel DILI biomarkers. However, there is no commonly adopted practice by which the research community can classify a drug's DILI potential in humans (Chen , Vijay, et al. 2011). *Chen, et al.* focused on using FDA-approved drug labels to develop a systematic and objective classification scheme, which is named as Rule-of-two (RO2), for categorizing the DILI potential of a drug.

The Rule-of-two (RO2) prediction model to identify propensity of DILI risk of new drugs is based on assessment of drug attributes (lipophilicity and dose >100 mg/day). The listing of drugs in this model is based on their identification as the primary drug of interest. Using the labeling of DILI in the Warning or Precautions section of the final product label, drugs with these two properties are scored as having DILI propensity. Within this framework, the RO2 has achieved adequate sensitivity and a specificity of in identifying DILI risk.

The limitation of RO2 model is that labeling is highly context specific, the relative rarity of DILI in the premarket experience and the complex phenotypes of DILI. Furthermore, drugs are often used in combination with other medications, which may have their own DILI liability and dosing can be modified from the labeled dose based on the indication or disease severity. This research aimed to enrich the RO2 model based on machine learning and data-mining modeling of premarket and post market DILI narrative reports and present the methods and findings from this initial effort. We utilized the FDA FAERS database that provides comprehensive post-marketing surveillance data in order to improve the DILI classification by integrating the post-marketing data into the drug-label based approach to further improve the accuracy of DILI classifications. This research will develop a statistical prediction model for better predicting DILI in humans.

DATA EXTRACTION/PREPROCESSING/VISUALIZATION

Three data platforms have been utilized in this research. Empirica Signal, Drug Safety Analytics Dashboards, and Rule-of-two dataset. The following section will highlight some of the data reprocessing that has been performed to prepare the data for modeling.

1. EMPIRICA SIGNAL

Empirica Signal is a tool used to monitor signals and their evolution over time. Empirica Signal with Signal Management:

1) creates a pharmacovigilance environment;
2) track and document day-to-day pharmacovigilance activities;
3) conducts periodic reviews and assessments of the latest safety information;
4) includes configurations and time-stamping capabilities to support analyses in different databases and between different points in time;

5) takes advantage of public data sources including the FAERS and Vaccine Adverse Event Reporting System (VAERS) databases, the VigiBase ADR (adverse drug reaction) database from the WHO Collaborating Center for International Drug Monitoring, as well as proprietary internal databases; and
6) provides drill-down capabilities to display case details collected in case reporting systems (Signal 2017).

Data-mining techniques are utilized for detecting safety signals of adverse events from spontaneous reports. These data mining algorithms, which are widely used for signal detection, are a complement for the traditional expert review of the reports as well as provides the capability to efficiently analyze large amounts of accumulated data. These data-mining techniques are used to explore databases of spontaneous reports for hidden associations between drugs and reported adverse events that may not be obvious during a manual case assessment (Harpaz 2013). FDA uses these techniques (commonly known as signal detection algorithms) with FAERS to monitor, prioritize, and identify new safety signals of adverse drug events that authorizing further investigation.

In this research, Empirica Signal served as the source of data retrieval based on the preferred term (PT) or standard MedDRA query (SMQ) equal to 'Drug related hepatic disorders - severe events only (SMQ) [narrow]'. 171,890 cases have been retrieved with the most data mining statistics that are widely used for signal detection. These statistics are proportional reporting ratio (PRR), empirical Bayes geometric mean (EBGM), lower 5th percentile of the posterior observed-to-expected distribution (EB05), reporting odds ratio (ROR), and reporting ratio (RR).

Prioritizing investigations might be based on scores for statistical significance, rather than for association, to avoid following up potential associations that could have arisen merely by chance. However, unnecessary focus on drugs and events that are common overall in the database can be an outcome of using a PRR or ROR p-value to rank associations. For instance, frequently reported drugs and events may have reporting ratios that are only slightly greater than 1, but have very tiny p values which will cause such unnecessary focus. Investigating these drug-event combinations can potentially eclipse larger reporting ratios for less-frequently reported drugs or events.

Employing the confidence limits with p values, both the statistical significance and the reporting ratio can contribute to a prioritization system. Therefore, ranking drug-event combinations by their lower confidence limits (for multi-item gamma Poisson shrinker (MGPS), by using EB05 rather than EBGM) reduces chance for false alarms due to chance fluctuation.
Therefore, in this research, we prioritize investigations based on both significance and association scores, rather than relying on only one score. A threshold of EBGM>2 and EB05>1 are used and therefore the number of cases reduced to only 14,436 cases from the initial retrieved 171,890 cases.

Moreover, observational analysis was performed to understand the most dominate preferred terms (PT) based on both EBGM values (Figure 1) and EB05 values (Figure 2). For example, Figure 2 shows that Alanine Aminotransferase Increased is the dominate preferred term when the

EB05 is between 1.01-1.73, while Cholestasis is the dominate one when EB05 is between 3.99-1099.8.

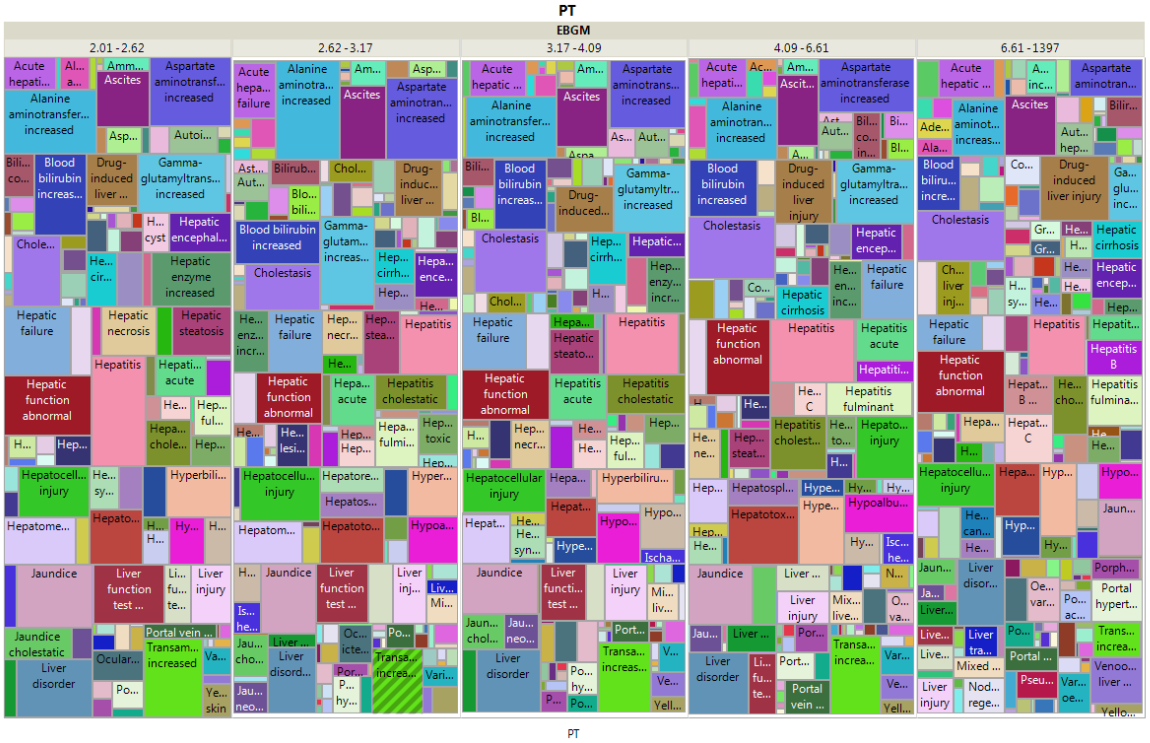


Figure 1: Preferred terms grouped by EBGM

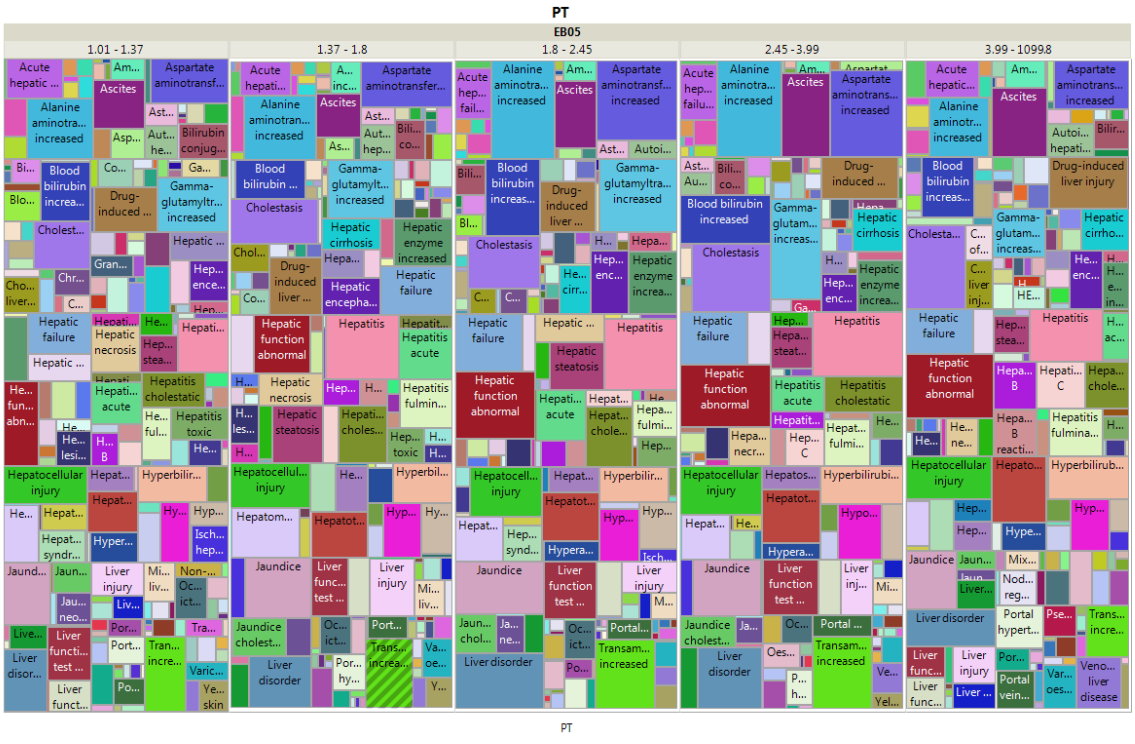


Figure 2: Preferred terms grouped by EB05

2. DRUG SAFETY ANALYTICS DASHBOARDS

The Drug Safety Analytics Dashboards provide a centralized place to monitor adverse events and perform aggregate case reviews for all report types while simultaneously offering an enhanced user experience with regards to performance, functionality, content, and aesthetics (Platform 2018) .

We utilized the Drug Safety Analytics Dashboards (MERCADO) to retrieve FAERS data regarding hepatic failure from November 1997 till March 2018. However, some of the reports were received prior to November 1997 therefore no narrative text had been entered in FAERS data. Since drug-induced liver injury (DILI) is increasingly being recognized as a cause of clinically significant acute and chronic liver disease (Fontana 2010), we customized our event using Standard MedDRA Query (SMQ) in order to select drug related hepatic disorders-severe events only with narrow scope searches. Using such a custom search enabled groupings of terms from one or more MedDRA System Organ Classes (SOCs) related to defined medical condition or area of interest as well as including terms may relate to signs, symptoms, diagnoses, syndromes, physical findings, laboratory and other physiologic test data, etc., related to medical condition or area of interest in this case DILI. Moreover, the narrow scope- specificity is to retieve all cases highly likely to be condition of interest while the broad scope-sensitivity is to retrieve all possible cases. Data was downloaded in seven time intervals since MERCADO allows only around 40,000 cases to be retrieved in one search. Some obsevational analysis was exploited at each time interval in order to understand the characteristics of the data at each time interval. For instance, the following analysis, not all inclusive, was performed for a time interval from January 01, 2014 thru December 31, 2016 (Tables 1,2,3,4,5) .

Patient Sex	Total Cases	% of Cases
Female	19,567	45.7%
Male	18,250	42.6%
Not Reported	4,990	11.7%
Unknown	12	0.0%
Total	**42,819**	**100.0%**

Table 1: Case Count by Patient Sex

Reported Outcomes	Total Cases
Death	9,324
Hospitalized	20,190
Life Threatening	3,027
Disabled	965
Congenital Anomaly	58
Required Intervention	115
Other Outcome	27,090
Total (Distinct Cases)	**42,819**

Table 2: Case Count by Reported Outcomes

		Report Type			Seriousness		Reported Outcomes						
Age Group	Total Cases	Direct	Expedited	Non-Expedited	Non Serious	Serious	DE	HO	LT	DS	CA	RI	OT
<1 year	134	3	125	6	1	133	50	58	21	4	5	0	69
1 - <3 years	174	16	153	5	3	171	67	84	15	2	2	0	76
3 - <7 years	225	23	198	4	7	218	50	110	31	0	0	1	119
7 - <17 years	770	49	685	36	16	754	131	397	71	14	1	3	469
17 - <65 years	19,456	901	17,385	1,170	753	18,703	3,849	10,106	1,530	482	8	70	12,100
>=65 years	10,444	363	9,559	522	320	10,124	2,811	5,819	934	242	2	36	6,139
NOT REPORTED	11,616	112	10,157	1,347	1,057	10,559	2,366	3,616	425	221	40	5	8,118
Total	**42,819**	**1,467**	**38,262**	**3,090**	**2,157**	**40,662**	**9,324**	**20,190**	**3,027**	**965**	**58**	**115**	**27,090**

Table 3: Displays total case count by age group, report type, seriousness and outcome.

		Report Type			Seriousness		Reported Outcomes						
Country	Total Cases	Direct	Expedited	Non-Expedited	Non Serious	Serious	DE	HO	LT	DS	CA	RI	OT
Foreign	27,364	34	26,912	418	187	27,177	6,420	13,362	2,207	525	31	26	17,840
USA	15,429	1,412	11,347	2,670	1,964	13,465	2,897	6,815	814	440	27	88	9,241
Not Reported	26	21	3	2	6	20	7	13	6	0	0	1	9
Total	**42,819**	**1,467**	**38,262**	**3,090**	**2,157**	**40,662**	**9,324**	**20,190**	**3,027**	**965**	**58**	**115**	**27,090**

Table 4: Displays total case count by country, report type, seriousness and outcome.

	Report Type		
Initial FDA Received Year	Direct	Expedited	Non-Expedited
2016	445	14,392	998
2015	547	12,950	1,303
2014	475	10,920	789

Table 5: Displays Case Count by Initial FDA Received Year or Event Year

After a high level understanding for our data corpus which is around 304,000 cases, the data was prepared for both the unsupervised and supervised learning. In unsupervised learning, for instance, it is important to reject variables which are unnecessary or irrelevant to the stated objective(s), in our case is a binary objective serious vs. non-serious event. For example, the basis variables used in the unsupervised learning, clustering algorithms, should be meaningful to the analysis objective; low correlation between input variables; intervals variables as categorical variables have a propensity to take over a cluster information; and low kurtosis and skewness to reduce the possibility of producing small outlier clusters for DILI cases. Likely basis variables

include case demographic, products information, patient history, report type, and reporter information. Since, almost our data variables are class variables and text; we utilized text mining to transfer these variables to interval ones using some techniques such as text clustering, text rule builder, and text profile. More than 241 variables were available for modeling but these were reduced to a smaller set, 121 variables, which had the possible to be analytically beneficial.

Since the data is dominated by cases with serious outcome value of Yes (Y=1), building any model with such dominate outcome will be biased towards predicting serious adverse event mostly. To compensate for the rare proportion of No (No=0) in the raw data, over-sampling of the data was done to produce a more balanced data set as well as to the patterns that appear in the data will be traceable in the sample. Over-sampling rare classes often leads to more accurate predictions. To illustrate the data over-sampling, we will use the FAERS data collected till December 31, 2000. Figure 3 shows that 88% of the target level was Yes (Y=1) while only 12% was No (N=0).

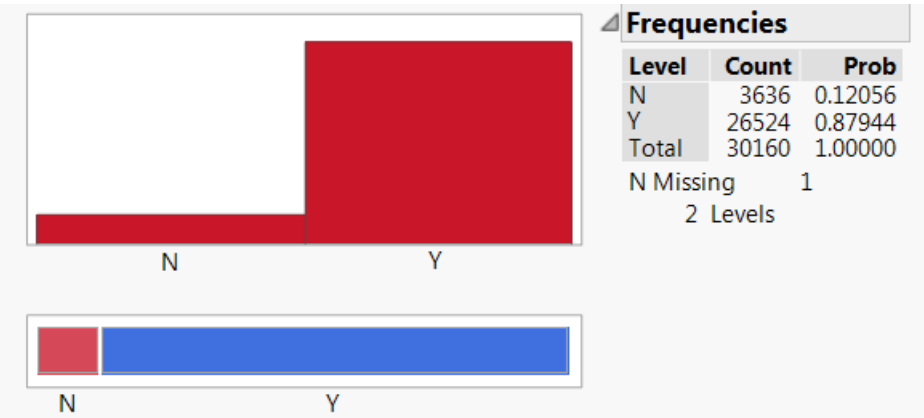


Figure 3: Summary Statistic for Serious Outcome

To account for working with rare events, oversampling technique was employed adjusting the frequency for the oversampling in order to create a frequency variable with sampling weights. The final No (N=0) proportion was increased to 34%.

Before building any predictive models and in order to get the correct decision consequences, we specify the inverse priors based on the original proportion of rare events (12%) to correctly adjust model predictions regardless of what the proportions in the training set are. If no adjusted prior probabilities are used, the estimated posterior probability for the No event class will be over-estimated. SAS Enterprise Miner uses profit matrix with elements equal to the inverse of the prior distribution for each outcome instead of a traditional profit matrix. The reason for such modification (i.e. inverse prior distribution) is to get accurate specification from model-based decision since it is difficult, if not impossible, task to tune and assess predictive models based on the profit or loss consequence of model-based decision (SAS Course Note 2016).

$$Let\ \pi_i = prior\ distribution\ for\ serious\ outcome\ at\ level\ i$$

Therefore, the inverse prior profit matrix for serious outcome, binary target, will be

		Decision	
		1	0
Serious Outcome	1	$\frac{1}{\pi_1}$	0
	0	0	$\frac{1}{\pi_0}$

Using the above modification, cases predicted more likely than average to have serious outcome with level=1 (primary outcome) will receive primary decision (decision=1). This instinctively related to the fact that the inverse prior profit matrix for binary response variable allocates decision=1 to each case with a posterior probability more than prior distribution probability for serious outcome at level 1(i.e. π_1).

3. RULE-of-TWO (RO2) DATASET

A drug list that was developed by (Chen, Suzuki, et al. 2016) is utilized in this research. This dataset has the following criteria: 1) has an FDA-approved label; 2) for human use only; 3) contains a single active molecule in the dosage form; 4) administered through oral or parenteral route; 5) approved for five years and 6) commercially available and affordable for future study. A total of 1036 FDA-approved unique drugs with a single active molecule for human use were collected from the DailyMed database. By using the verification process for drug induced liver injury (DILI) annotation *Chen et al.* 2016, 1036 FDA- approved drugs were classified into 192 vMost-DILI concern, 278 vLess-DILI concern, and 312 vNo-DILI concern drugs, all of which were verified by the evidenced causality, and leaving out 254 drugs as 'Ambiguous DILI concern' drugs (Figure 4).

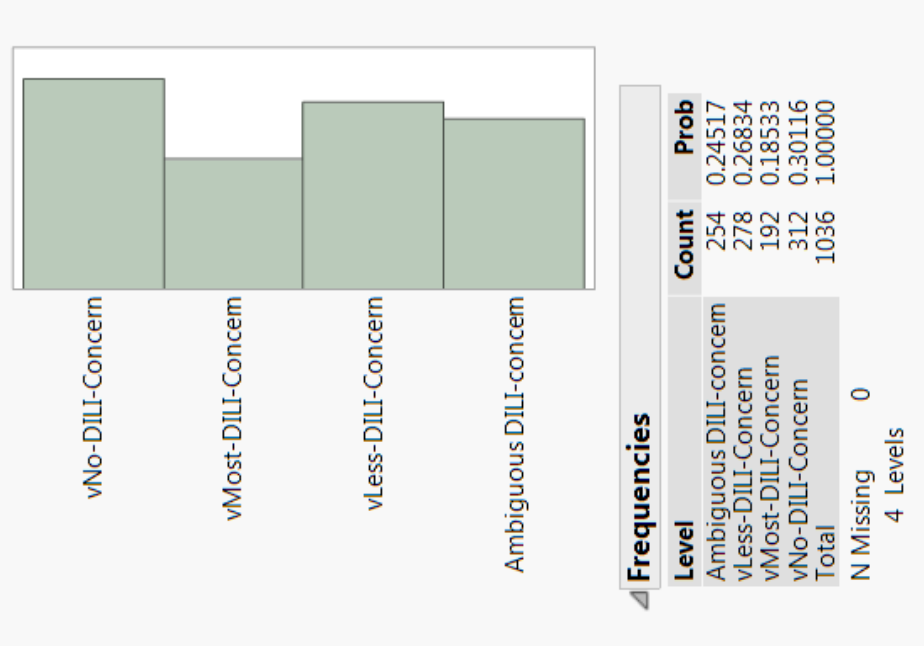


Figure 4: A summary for drug-induced lever injury (DILI) annotation in RO2 dataset

Moreover, this dataset has defined the rule-of-two test for these 1036 drugs. The rule-of-two (RO2) (Chen, Borlak, et al. 2013) is defined by the combing two factors which are the daily dose and lipophilicity in assigning DILI positive and DILI negative. Based on these two factors, high risk for hepatotoxicity (odds ratio [OR], 14.05; $P < 0.001$) is observed for drugs given at dosages $\geq$ 100 mg/day and octanol-water partition coefficient (logP) ≥ 3.

An observational analysis has been performed to understand this data set. In Figure 5 below, % of total (Daily Dose in mg/day) vs. LogP have been grouped based on their values in label section. For instance, a compound name Aplaviroc is the dominate compound in the discontinued label section with LogP=2.58 and % of Total (Daily Dose in mg/day) = 49.08% for total daily dose at this label section. While, in the box warning label section, Divalproex Sodium is the dominate compound name with LogP=3.55 and % of Total (Daily Dose in mg/day) = 21.54% for total daily dose at the box warning section

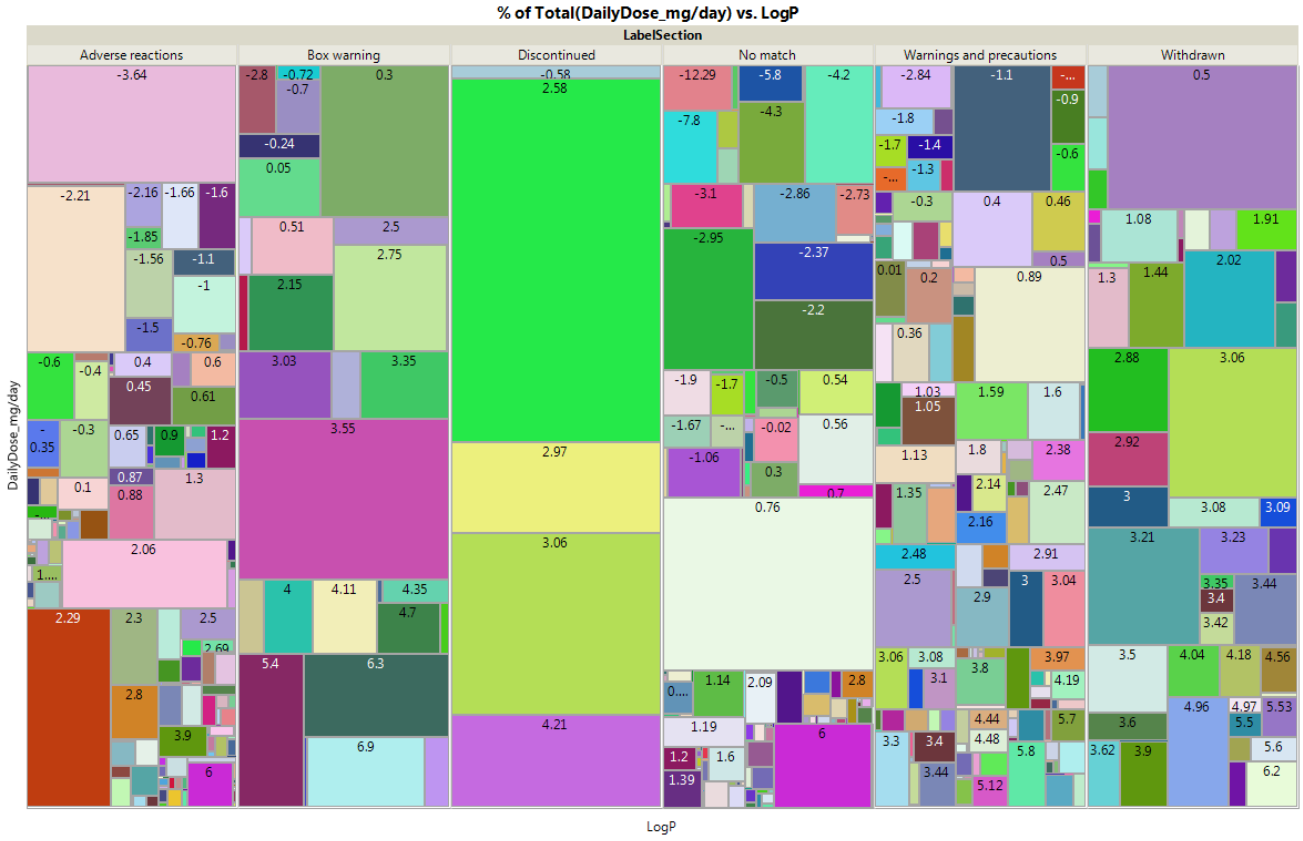


Figure 5: % of total (Daily Dose in mg/day) vs. LogP grouped by the label section

METHODOLOGY

The following steps summarize the research methodology (Figure 6) that this research developed:

1. Association analysis model is developed on the data retrieved from Empirica Signal. The objective is to determine which preferred terms go together. Such information can be useful for investigating associative relationships in DILI preferred terms.
2. Domain experts from FDA have been consulted to assign topics names for the three different association analysis scenarios outcomes regarding the DILI preferred terms.
3. After building the association model, the three data sets (i.e., Empirica Signal, Drug Safety Analytics Dashboards, Role-of-two) have been aggregated by performing queries using both SAS Proc SQL as well as JMP query builder.
4. The combined data set contains both structured and unstructured data. Therefore, text analysis model is built for the unstructured data. Moreover, supervised and unsupervised models are also constructed to examine the structured data as well as the outcome from the text mining model.

Figure 6: Methodology

Due to the space limitation, some of the developed models in the above steps will be illustrated in the following section.

ANALYTICS APPLICATIONS and RESULTS

1. ASSOCIATION ANALYSIS

Association analysis is a popular technique that is used to identify and visualize relationships (association) between different objects. For this research, the following question could be nontrivial to be answered manually: What linkage of DILI preferred terms (events) can be observed from post-market data? Such relationships can be addressed using association analysis by defining association rules and calculating the support for the combination of the preferred terms(PTs). The relationship between two preferred term sets is defined by an association rule. An association rule consists of a condition item set (PTs) and a consequent item set (PTs). Antecedents are the individual items in the condition item set. Association analysis identifies association rules, which predict that a consequent item set will be in an event, given that the condition item set is already in the event. Some association rules are stronger, and therefore more useful, than others. Support, confidence, and lift are the three performance measures describe the

strength of an association rule. Designate the condition item set by *X* and the consequent item set by *Y*. An association rule with condition set *X* and consequent set *Y* is denoted as $X \Rightarrow Y$.

Support is the proportion of events in which an item set (PTs) appears. A high value for support indicates that the item set occurs frequently.

$$Support\ (X \cup Y) = \frac{Number\ of\ Events\ Containing\ both\ Items\ X\ and\ Y}{Total\ Number\ of\ Events}$$

Measuring the strength of implication of an association rule (predictive power) is performed by calculating the *Confidence* which is the proportion of events that contain the consequent item set (PTs), given that the condition item set (PTs) is in the transaction.

$$Confidence\ (X \Longrightarrow Y) = \frac{Number\ of\ Events\ that\ Contain\ both\ Items\ X\ and\ Y}{Number\ of\ Events\ that\ Contain\ Item\ X}$$

Finally, *lift* measures how much the consequent item set depends on the presence of the condition item set. *Lift* is the ratio of an association rule's confidence to its expected confidence, with the assumption that the condition and consequent item sets appear in events independently.

$$Lift = \frac{\left(\frac{Number\ of\ Events\ that\ Contain\ both\ Items\ X\ and\ Y}{Number\ of\ Events\ that\ Contain\ Item\ X}\right)}{\left(\frac{Number\ of\ Events\ that\ Contain\ Item\ Y}{Total\ Number\ of\ Events}\right)}$$

Three scenarios are developed for the subset data from Empirica Signal (i.e., the 14,436 cases after utilizing both significance and association scores in prioritizing DILI cases). For each scenario, association model is built based on different settings for minimum support, minimum confidence, minimum lift, maximum antecedents, and maximum rule size. Resetting these values allow us to cover more association rules as well as understand the optimal setting that provides more informative rules.

For instance, setting a minimum support=0.1, minimum confidence=0.4, minimum lift=3.2, maximum antecedents=10, and maximum rule size=250, the rules table (Table 7) illustrates the generated rules with their associated expected confidence, confidence, support, and lift values.

Expected Confidence(%)	Confidence(%)	Support(%)	Lift	Transaction Count	Rule ▲
1.89	62.50	1.44	32.99	25.00	Hepatotoxicity & Aspartate aminotransferase abnormal ==> Transaminases increased & Hyperbilirubinaemia & Alanine aminotransferase abnormal
2.64	88.89	1.38	33.66	24.00	Oesophageal varices haemorrhage & Cholestasis & Ascites ==> Varices oesophageal & Portal vein thrombosis
2.64	88.89	1.38	33.66	24.00	Oesophageal varices haemorrhage & Cholestasis & Blood bilirubin increased ==> Varices oesophageal & Portal vein thrombosis
2.01	75.00	1.38	37.33	24.00	Oesophageal varices haemorrhage & Cholestasis ==> Varices oesophageal & Hyperbilirubinaemia & Hepatotoxicity
2.24	75.00	1.38	33.50	24.00	Oesophageal varices haemorrhage & Cholestasis ==> Varices oesophageal & Portal hypertension & Hepatotoxicity
2.41	88.89	1.38	36.87	24.00	Oesophageal varices haemorrhage & Hepatitis C & Blood bilirubin increased ==> Portal vein thrombosis & Hepatic cirrhosis
2.41	85.71	1.38	35.55	24.00	Oesophageal varices haemorrhage & Hepatitis C & Hepatic failure ==> Portal vein thrombosis & Hepatic cirrhosis
1.78	68.57	1.38	38.53	24.00	Oesophageal varices haemorrhage & Hepatitis C ==> Portal vein thrombosis & Hepatic cirrhosis & Blood bilirubin increased
1.89	68.57	1.38	36.20	24.00	Oesophageal varices haemorrhage & Hepatitis C ==> Portal vein thrombosis & Hepatic failure & Hepatic cirrhosis
1.95	71.43	1.44	36.60	25.00	Oesophageal varices haemorrhage & Hepatitis C ==> Portal vein thrombosis & Hyperbilirubinaemia & Hepatic cirrhosis
1.95	68.57	1.38	35.13	24.00	Oesophageal varices haemorrhage & Hepatitis C ==> Varices oesophageal & Portal vein thrombosis & Hepatic cirrhosis
2.18	71.43	1.44	32.74	25.00	Oesophageal varices haemorrhage & Hepatitis C ==> Varices oesophageal & Portal vein thrombosis & Hyperbilirubinaemia
3.04	100.00	1.49	32.87	26.00	Oesophageal varices haemorrhage & Hepatotoxicity & Blood bilirubin increased ==> Varices oesophageal & Hyperbilirubinaemia
2.58	96.00	1.38	37.16	24.00	Oesophageal varices haemorrhage & Hepatotoxicity & Hepatic cirrhosis ==> Varices oesophageal & Transaminases increased
3.04	100.00	1.44	32.87	25.00	Oesophageal varices haemorrhage & Hepatotoxicity & Hepatic failure ==> Varices oesophageal & Hyperbilirubinaemia
2.07	77.42	1.38	37.46	24.00	Oesophageal varices haemorrhage & Hepatotoxicity ==> Varices oesophageal & Hyperbilirubinaemia & Cholestasis
2.35	77.42	1.38	32.89	24.00	Oesophageal varices haemorrhage & Hepatotoxicity ==> Varices oesophageal & Portal hypertension & Cholestasis
2.53	87.10	1.55	34.48	27.00	Oesophageal varices haemorrhage & Hepatotoxicity ==> Varices oesophageal & Portal hypertension & Hyperbilirubinaemia
2.18	77.42	1.38	35.49	24.00	Oesophageal varices haemorrhage & Hepatotoxicity ==> Varices oesophageal & Portal vein thrombosis & Hyperbilirubinaemia
2.35	77.42	1.38	32.89	24.00	Oesophageal varices haemorrhage & Hepatotoxicity ==> Varices oesophageal & Portal vein thrombosis & Portal hypertension
2.30	77.42	1.38	33.72	24.00	Oesophageal varices haemorrhage & Hepatotoxicity ==> Varices oesophageal & Transaminases increased & Hepatic cirrhosis
1.89	77.42	1.38	40.87	24.00	Oesophageal varices haemorrhage & Hepatotoxicity ==> Varices oesophageal & Transaminases increased & Hyperbilirubinaemia
2.18	77.42	1.38	35.49	24.00	Oesophageal varices haemorrhage & Hepatotoxicity ==> Varices oesophageal & Transaminases increased & Portal hypertension
2.58	92.31	1.38	35.73	24.00	Oesophageal varices haemorrhage & Hyperbilirubinaemia & Cholestasis ==> Varices oesophageal & Hepatotoxicity
2.12	73.53	1.44	34.62	25.00	Oesophageal varices haemorrhage & Hyperbilirubinaemia & Hepatic cirrhosis ==> Portal vein thrombosis & Hepatitis C
2.12	76.47	1.49	36.00	26.00	Oesophageal varices haemorrhage & Hyperbilirubinaemia & Hepatic cirrhosis ==> Varices oesophageal & Hepatitis C
2.12	72.73	1.38	34.24	24.00	Oesophageal varices haemorrhage & Hyperbilirubinaemia & Hepatic encephalopathy ==> Varices oesophageal & Hepatitis C
2.41	83.33	1.44	34.56	25.00	Oesophageal varices haemorrhage & Hyperbilirubinaemia & Hepatitis C ==> Portal vein thrombosis & Hepatic cirrhosis
2.58	85.71	1.38	33.18	24.00	Oesophageal varices haemorrhage & Hyperbilirubinaemia & Hepatotoxicity ==> Varices oesophageal & Transaminases increased
1.78	61.36	1.55	34.48	27.00	Oesophageal varices haemorrhage & Hyperbilirubinaemia ==> Portal vein thrombosis & Hepatic cirrhosis & Blood bilirubin increased
1.66	56.82	1.44	34.13	25.00	Oesophageal varices haemorrhage & Hyperbilirubinaemia ==> Portal vein thrombosis & Hepatitis C & Hepatic cirrhosis
1.61	56.82	1.44	35.35	25.00	Oesophageal varices haemorrhage & Hyperbilirubinaemia ==> Varices oesophageal & Hepatitis C & Blood bilirubin increased
1.66	54.55	1.38	32.76	24.00	Oesophageal varices haemorrhage & Hyperbilirubinaemia ==> Varices oesophageal & Portal hypertensive gastropathy & Portal hypertension
1.89	63.64	1.61	33.59	28.00	Oesophageal varices haemorrhage & Hyperbilirubinaemia ==> Varices oesophageal & Portal vein thrombosis & Hepatic encephalopathy
2.07	68.18	1.72	32.99	30.00	Oesophageal varices haemorrhage & Hyperbilirubinaemia ==> Varices oesophageal & Portal vein thrombosis & Hepatic failure
1.66	56.82	1.44	34.13	25.00	Oesophageal varices haemorrhage & Hyperbilirubinaemia ==> Varices oesophageal & Portal vein thrombosis & Hepatitis C
1.55	54.55	1.38	35.19	24.00	Oesophageal varices haemorrhage & Hyperbilirubinaemia ==> Varices oesophageal & Portal vein thrombosis & Hepatotoxicity
1.78	64.86	1.38	36.45	24.00	Oesophageal varices haemorrhage & Jaundice ==> Portal vein thrombosis & Hepatic cirrhosis & Blood bilirubin increased
1.95	64.86	1.38	33.23	24.00	Oesophageal varices haemorrhage & Jaundice ==> Portal vein thrombosis & Hyperbilirubinaemia & Hepatic cirrhosis
1.95	64.86	1.38	33.23	24.00	Oesophageal varices haemorrhage & Jaundice ==> Varices oesophageal & Portal vein thrombosis & Hepatic cirrhosis
1.78	63.16	1.38	35.49	24.00	Portal hypertension & Hepatitis C ==> Portal vein thrombosis & Hepatic cirrhosis & Blood bilirubin increased
1.89	65.79	1.44	34.73	25.00	Portal hypertension & Hepatitis C ==> Portal vein thrombosis & Hepatic failure & Hepatic cirrhosis
1.95	68.42	1.49	35.06	26.00	Portal hypertension & Hepatitis C ==> Portal vein thrombosis & Hyperbilirubinaemia & Hepatic cirrhosis
1.95	71.05	1.55	36.40	27.00	Portal hypertension & Hepatitis C ==> Varices oesophageal & Portal vein thrombosis & Hepatic cirrhosis
2.64	88.89	1.84	33.66	32.00	Portal hypertension & Oesophageal varices haemorrhage & Blood bilirubin increased ==> Varices oesophageal & Portal vein thrombosis
2.58	85.71	1.38	33.18	24.00	Portal hypertension & Oesophageal varices haemorrhage & Cholestasis ==> Varices oesophageal & Hepatotoxicity
2.64	89.29	1.44	33.81	25.00	Portal hypertension & Oesophageal varices haemorrhage & Hepatomegaly ==> Varices oesophageal & Portal vein thrombosis
2.58	85.71	1.38	33.18	24.00	Portal hypertension & Oesophageal varices haemorrhage & Hepatotoxicity ==> Varices oesophageal & Transaminases increased
2.12	70.59	1.38	33.23	24.00	Portal hypertension & Oesophageal varices haemorrhage & Hyperbilirubinaemia ==> Varices oesophageal & Hepatitis C
1.89	70.59	1.38	37.26	24.00	Portal hypertension & Oesophageal varices haemorrhage & Hyperbilirubinaemia ==> Varices oesophageal & Portal hypertensive gastropathy
2.64	91.18	1.78	34.53	31.00	Portal hypertension & Oesophageal varices haemorrhage & Hyperbilirubinaemia ==> Varices oesophageal & Portal vein thrombosis
1.55	51.06	1.38	32.95	24.00	Portal hypertension & Oesophageal varices haemorrhage ==> Portal vein thrombosis & Portal hypertensive gastropathy
1.55	51.06	1.38	32.95	24.00	Portal hypertension & Oesophageal varices haemorrhage ==> Varices oesophageal & Portal hypertensive gastropathy & Ascites
1.49	51.06	1.38	34.21	24.00	Portal hypertension & Oesophageal varices haemorrhage ==> Varices oesophageal & Portal hypertensive gastropathy & Hyperbilirubinaemia

Table 7: Rules Table

The first rule is [***Hepatotoxicity & Aspartate aminotransferase abnormal ==> Transaminases increased & Hyperbilirubinaemia & Alanine aminotransferase abnormal***] which indicates that with a confidence of 62.5 % of the events where the preferred terms *Hepatotoxicity & Aspartate aminotransferase abnormal* appear in DILI cases, the preferred terms *Transaminases increased & Hyperbilirubinaemia & Alanine aminotransferase abnormal* will also appears. The value of Lift is 32.99, indicating that there is a likely dependency since a lift ratio greater than 1 indicates that the consequent item ***"Transaminases increased & Hyperbilirubinaemia & Alanine aminotransferase abnormal"*** set has an affinity for the condition item set "***Hepatotoxicity & Aspartate aminotransferase abnormal"***. Therefore, the consequent item set occurs more often with the condition item set than one would expect by chance alone.

Even though the rules generated above might be sufficient for understanding the degree of association among the DILI preferred terms, additional analysis was performed so that similar PTs are grouped together using a matrix reducing methodology (i.e., Singular value decomposition (SVD)). Singular value decomposition (SVD) reducing the PTs matrix, which is denoted as *transaction item matrix* in association analysis modeling, to a manageable number of dimensions. The transaction listing (Figure 8) will be the entries of the transaction item matrix for which each row corresponds to a transaction ID and each column corresponds to an item

(PT). The entries of the matrix are zeros and ones. If an item (PT) occurs in a transaction, the corresponding row and column entry is one. Otherwise, the row and column entry is zero.

Transaction Listing

Transaction ID	Item Set
3	Acute hepatic failure, Alanine aminotransferase abnormal, Alanine aminotransferase increased, Ascites, Aspartate aminotransferas...
5	Acute hepatic failure, Aspartate aminotransferase increased, Drug-induced liver injury, Hepatic enzyme increased, Hepatic necrosi...
6	Acute hepatitis C, Alanine aminotransferase abnormal, Alanine aminotransferase increased, Aspartate aminotransferase increased,...
7	Ascites, Blood bilirubin increased, Gamma-glutamyltransferase increased, Hepatic enzyme increased, Hepatitis, Hepatomegaly, He...
9	Alanine aminotransferase abnormal, Biliary cirrhosis, Hepatic enzyme increased, Hepatitis B, Hepatitis B reactivation, Liver function...
10	Hepatic haemorrhage
11	Alanine aminotransferase increased, Aspartate aminotransferase increased, Gamma-glutamyltransferase increased, Hepatic enzy...
15	Ammonia increased, Blood bilirubin increased, Gamma-glutamyltransferase increased, Hepatitis, Hepatitis fulminant, Hepatitis tox...
16	Alanine aminotransferase increased, Aspartate aminotransferase increased, Blood bilirubin increased, Cholestasis, Gamma-glutam...
19	Cholestasis, Hepatitis, Hepatocellular injury
20	Hepatitis
21	Hepatic failure
22	Cholestasis, Hepatitis cholestatic, Hepatocellular injury, Hepatomegaly, Ischaemic hepatitis, Jaundice cholestatic, Oesophageal vari...
23	Aspartate aminotransferase increased
24	Cholestasis, Hepatitis cholestatic
25	Cholestasis, Hepatitis
27	Hepatitis, Hepatitis fulminant, Jaundice neonatal
28	Acute fatty liver of pregnancy, Acute hepatic failure, Acute on chronic liver failure, Alanine aminotransferase, Alanine aminotransfe...
41	Acute hepatic failure, Liver injury
49	Cholestasis, Gamma-glutamyltransferase increased, Hepatitis, Hepatitis acute, Hepatitis cholestatic, Hepatocellular injury, Jaundice...
51	Acute hepatic failure, Hepatic failure, Hepatotoxicity
54	Hepatitis
55	Aspartate aminotransferase increased, Blood bilirubin increased, Gamma-glutamyltransferase increased, Hepatitis, Hepatocellular ...
66	Alanine aminotransferase increased, Aspartate aminotransferase increased, Bilirubin conjugated increased, Blood bilirubin increas...
72	Acute hepatic failure, Hepatic failure
77	Acute hepatic failure, Hepatic failure
78	Acute hepatic failure, Hepatic failure, Hepatic function abnormal, Hepatic necrosis, Hepatic steatosis, Hepatitis, Hepatocellular inju...
80	Acute hepatic failure, Alanine aminotransferase increased, Ammonia increased, Aspartate aminotransferase increased, Blood biliru...
83	Acute hepatic failure, Drug-induced liver injury, Hepatic encephalopathy
88	Acute hepatic failure, Alanine aminotransferase increased, Aspartate aminotransferase increased, Hepatic encephalopathy, Hepati...
91	Transaminases increased
92	Acute hepatic failure
94	Acute hepatic failure, Alanine aminotransferase increased, Aspartate aminotransferase increased, Blood bilirubin increased, Drug-i...
95	Cholestasis, Gamma-glutamyltransferase increased, Hepatic encephalopathy, Hepatic necrosis, Hepatic steatosis, Hepatitis, Hepati...
97	Acute hepatic failure, Alanine aminotransferase increased, Blood bilirubin increased, Gamma-glutamyltransferase abnormal, Hepa...
98	Acute hepatic failure, Alanine aminotransferase increased, Ammonia increased, Aspartate aminotransferase increased, Bilirubin co...
107	Acute hepatic failure, Ammonia increased, Hepatic encephalopathy, Hepatic failure, Hepatic necrosis, Ischaemic hepatitis, Jaundic...
113	Acute hepatic failure, Ammonia increased, Hepatic congestion, Hepatic encephalopathy, Hepatic failure, Hepatic necrosis, Ischae...
119	Acute hepatic failure
123	Cholestasis, Drug-induced liver injury, Gamma-glutamyltransferase increased, Hepatic encephalopathy, Hepatic function abnorma...
124	Acute hepatic failure, Hepatic cirrhosis, Hepatic enzyme increased, Hepatic necrosis, Liver injury, Transaminases abnormal, Transa...
126	Ammonia increased, Hepatic failure, Hyperammonaemia

Figure 8: Transaction listing

Then, rotating the SVD by performing a varimax rotated singular value decomposition of the transaction item matrix to produce groups of similar transactions called topics. The grouped PTs are then presented to domain experts from FDA (6 medical officers) to assign informative topic name for each group based on the experts' judgments. Experts independently provided their assigned topic names and the experts' outcomes are aggregated and majority consistent in topic naming are employed to assign name(s) for the generated topics (Table 7,8,9,10,11).

Item	Topic Name
Bacillary angiomatosis	**Various hepatic disorders, particularly vascular, Hepatic Infection/vascular, Hepatic vascular disorders, complications of liver transplantation, nonspecific clinical finding, infectious hepatitis, liver injury clinical finding**
Hepatic cyst infection	
Hepatic artery stenosis	
Perihepatic abscess	
Hepatic artery aneurysm	
Portal vein stenosis	
Splenorenal shunt	
Hepatitis infectious mononucleosis	
Hepatic vein stenosis	
Portal vein occlusion	
Portal vein phlebitis	
Chronic graft versus host disease in liver	
Hepatic artery occlusion	

Table 7: Topic 1

Bile output abnormal	**Various liver Injury and associated lab/exam findings, Hepatic Infection/Injury/Toxicity, Liver injury, infectious hepatitis, liver injury clinical finding, nonspecific lab finding, liver injury lab finding**
Hepatobiliary infection	
Hepatic artery thrombosis	
Hepatitis toxic	
Cholestasis	
Gamma-glutamyltransferase increased	
Hepatocellular injury	
Transaminases increased	
Hepatitis acute	
Hepatitis	
Hepatitis cholestatic	
Hepatotoxicity	
Cholestatic liver injury	
Jaundice	
Liver function test abnormal	
Drug-induced liver injury	
Hepatomegaly	
Alanine aminotransferase increased	
Hepatic enzyme increased	
Aspartate aminotransferase increased	
Liver injury	

Table 8: Topic 2

Hepatic candidiasis	**Hepatic fungal/viral infections, lab findings, misc, Hepatic Infection/Vascular, Liver infection in immunocompromised host , Liver infection , vascular sequelae of malignancy or infiltrative disease, vascular sequelae of immunocompromise, vascular complication, infectious hepatitis, nonspecific clinical finding, liver injury lab finding, liver injury clinical finding**
Hepatosplenic candidiasis	
Adenoviral hepatitis	
Retrograde portal vein flow	
Hepatitis D	
Hepatic infection fungal	
Hepatitis E	
Blood bilirubin abnormal	
Acute hepatitis B	
Budd-Chiari syndrome	
Venoocclusive liver disease	
Aspartate aminotransferase abnormal	
Transaminases abnormal	
Hepatitis B reactivation	
Hepatic vein occlusion	
Herpes simplex hepatitis	

Table 9: Topic 3

Hepatitis chronic persistent	**Chronic failure and associated lab/physical manifestations, misc, Chronic liver disease, Chronic liver disease, Chronic liver disease with cirrhosis, infectious hepatitis, liver injury clinical finding, nonspecific lab finding, nonspecific clinical finding,**
Perihepatic discomfort	
Oedema due to hepatic disease	
Ammonia decreased	
Hepatic hydrothorax	
Acute on chronic liver failure	
Hepatic amoebiasis	
Ultrasound liver abnormal	
Chronic hepatic failure	
Alanine aminotransferase decreased	
Ammonia abnormal	
Aspartate aminotransferase decreased	
Child-Pugh-Turcotte score increased	
Varices oesophageal	
Oesophageal varices haemorrhage	
Hepatitis chronic active	

Table 10: Topic 4

Periportal sinus dilatation	**Varices and other hepatic complications, misc, Hepatic Vascular Disorders, Hepatic vascular disorders, portal hypertension, liver injury clinical finding, nonspecific clinical finding, nonspecific lab finding, liver injury histological finding.**
Intrahepatic portal hepatic venous fistula	
Intestinal varices	
Stomal varices	
Anorectal varices haemorrhage	
Pseudocirrhosis	
Gastric varices haemorrhage	
Portal hypertension	
Portal shunt	
Gastric varices	
Portal vein thrombosis	
Hepatic lesion	
Oesophageal varices haemorrhage	
Blood bilirubin abnormal	

Table 11: Topic 5

The above five topics are generated based on examining the SVD plot (Figure 9) in which approximately five different groups can be detected. However, analysts might assign different number based on their judgments and domain expert.

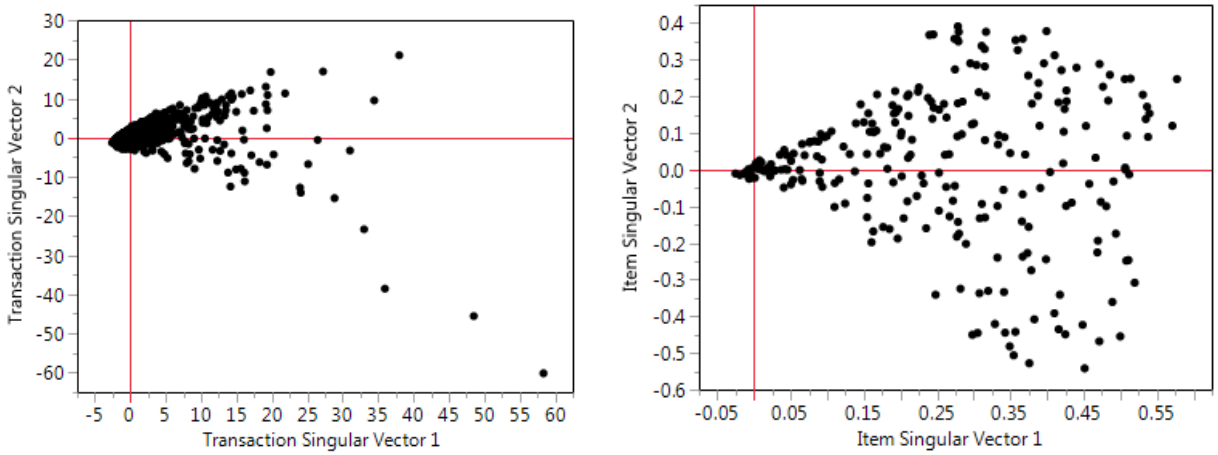


Figure 9: Singular Value Decomposition Plot

2. DATA SETS AGGREGATION

The data that have been utilized in this research has two different domains (i.e., pre-marketing and post-marketing). RO2 dataset was based on information gathered from drug labeling as well as incorporating information about whether the drugs were verified for their causality of DILI in humans, using publicly available resources. While, data cathered from Empirica Signal, Drug Safety Analytics Dashboards are based on FAERS data which is post-marketing data. Therefore, we build several customized SQL to match the RO2 compound names (1036 unique drugs) with 182474 DILI cases from FAERS that have more unique drugs than the RO2 dataset.
For instance, the primary suspect drug list in FAERS has 4520 unique drugs while one of the concomitant drug list has 5257 unique drugs. Therefore, tables joining, concatenating, and updating are performed using JMP Custom SQL. For illustration, the following SQL is to compare the RO2 compound name with the FAERS primary suspect drug list and the concomitant drug list which can be up to 10 drugs for one case.

```
New SQL Query(
	Version( 130 ),
	Connection( "JMP" ),
	JMP Tables(
		["All cases for DILI" =>
		"M:\Eileen_Qais Project_Pre_Post Market\Narrative for DILI data\All Cases for DILI till
Nov 21_2017\All cases for DILI.jmp",
		"RO2 in both Primary_P2_P3_P4_P5_P6_P7_P8_P9_P10" =>
		"M:\Eileen_Qais Project_Pre_Post Market\Narrative for DILI data\Data Comparison between
RO2 and FAERs Data\RO2 in both Primary_P2_P3_P4_P5_P6_P7_P8_P9_P10.jmp"
		]
	),
	QueryName( "SQLQuery7" ),
	CustomSQL(
		"SELECT t1.\!"RO2 matching either the Primary, P2, P3, P4, P5, P6, P7, P8, P9, P10 or
their Active ingredient\!", t1.\!"N Rows of RO2 matching either the Primary Suspect or Active
ingredient\!", t1.\!"N Rows of RO2 matching either the P2 or Active ingredient\!", t1.\!"N Rows of RO2
matching either the P3 or Active ingredient\!",
	t1.\!"N Rows of RO2 matching either the P4 or Active ingredient\!", t1.\!"N Rows of RO2 matching
either the P5 or Active ingredient\!", t1.\!"N Rows of RO2 matching either the P6 or Active ingredient\!",
t1.\!"N Rows of RO2 matching either the P7 or Active ingredient\!",
	t1.\!"N Rows of RO2 matching either the P8 or Active ingredient\!", t1.\!"N Rows of RO2 matching
either the P9 or Active ingredient\!", t1.\!"N Rows of RO2 matching either the P10 or Active
ingredient\!", t2.\!"FAERS Case #\!",
	t2.\!"Version #\!", t2.\!"Image Info/Link\!", t2.\!"Attachments Info/Link\!", t2.\!"Manufacturer
Control #\!",
	t2.\!"ISR #(s)\!", t2.\!"Report Type\!", t2.\!"Form Type\!", t2.\!"Initial FDA Received Date\!",
	t2.\!"Latest FDA Received Date\!", t2.\!"Latest MFR Received Date\!", t2.\!"Data Entry Completion
Date\!", t2.\!"Patient ID\!",
	t2.\!"Age in Years\!", t2.DOB, t2.Sex, t2.\!"Weight (kg)\!",
	t2.\!"Medical History /Medical History Comments\!", t2.\!"Sender Organization\!", t2.\!"Reporter
Organization\!", t2.\!"Reporter Last Name\!",
	t2.\!"Reporter First Name\!", t2.\!"Reporter City\!", t2.\!"Reporter State\!", t2.\!"Country
Derived\!",
	t2.\!"Reporter Qualifications\!", t2.\!"Health Professional\!", t2.\!"Report Source\!",
t2.Narrative,
	t2.\!"Case Event Date\!", t2.\!"All LLTs\!", t2.\!"All PTs\!", t2.\!"All HLTs\!",
	t2.\!"All HLGTs\!", t2.\!"All SOCs\!", t2.\!"Medication Errors Narrow SMQ (PTs)\!",
t2.\!"Medication Errors Narrow SMQ (LLTs)\!",
	t2.\!"Medication Errors Broad SMQ (PTs)\!", t2.\!"PT Term Event 1\!", t2.\!"Start Date Event 1\!",
t2.\!"PT Term Event 2\!",
	t2.\!"Start Date Event 2\!", t2.\!"PT Term Event 3\!", t2.\!"Start Date Event 3\!", t2.\!"PT Term
Event 4\!",
	t2.\!"Start Date Event 4\!", t2.\!"PT Term Event 5\!", t2.\!"Start Date Event 5\!", t2.\!"PT Term
Event 6\!",
	t2.\!"PT Term Event 7\!", t2.\!"PT Term Event 8\!", t2.\!"PT Term Event 9\!", t2.\!"PT Term Event
10\!",
	t2.\!"PT Term Event 11\!", t2.\!"PT Term Event 12\!", t2.\!"Serious Outcome?\!", t2.\!"All
Outcomes\!",
	t2.\!"All Suspect Product Names\!", t2.\!"ALL Suspect Product Active Ingredients\!", t2.\!"All
Suspect Active Ingredients\!", t2.\!"ALL Suspect Verbatim Products\!",
	t2.\!"All Concomitants\!", t2.\!"Product 1 Product Name\!", t2.\!"Product 1 Product Active
Ingredient\!", t2.\!"Product 1 Reported Verbatim\!",
	t2.\!"Product 1 Role\!", t2.\!"Product 1 Reason for Use\!", t2.\!"Product 1 Strength\!",
t2.\!"Product 1 Strength (Unit)\!",
	t2.\!"Product 1 Dose (Amount)\!", t2.\!"Product 1 Dose (Unit)\!", t2.\!"Product 1 Dosage Text\!",
t2.\!"Product 1 Dosage Form\!",
	t2.\!"Product 1 Route\!", t2.\!"Product 1 Frequency\!", t2.\!"Product 1 Dechallenge\!",
t2.\!"Product 1 Rechallenge\!",
	t2.\!"Product 1 Start Date\!", t2.\!"Product 1 Stop Date\!", t2.\!"Product 1 Therapy Duration
(Days)\!", t2.\!"Product 1 Therapy Duration (Verbatim)\!",
	t2.\!"Product 1 Time To Onset (Days)\!", t2.\!"Product 1 Manufacturer Name\!", t2.\!"Product 1
Application Type\!", t2.\!"Product 1 Application #\!",
	t2.\!"Product 1 NDC #\!", t2.\!"Product 1 Lot #\!", t2.\!"Product 2 Product Name\!", t2.\!"Product
2 Product Active Ingredient\!",
```

```
        t2.\!"Product 2 Reported Verbatim\!", t2.\!"Product 2 Role\!", t2.\!"Product 2 Reason for Use\!",
t2.\!"Product 2 Strength\!",
        t2.\!"Product 2 Strength (Unit)\!", t2.\!"Product 2 Dose (Amount)\!", t2.\!"Product 2 Dose
(Unit)\!", t2.\!"Product 2 Dosage Text\!",
        t2.\!"Product 2 Dosage Form\!", t2.\!"Product 2 Route\!", t2.\!"Product 2 Frequency\!",
t2.\!"Product 2 Dechallenge\!",
        t2.\!"Product 2 Rechallenge\!", t2.\!"Product 2 Start Date\!", t2.\!"Product 2 Stop Date\!",
t2.\!"Product 2 Therapy Duration (Days)\!",
        t2.\!"Product 2 Therapy Duration (Verbatim)\!", t2.\!"Product 2 Time To Onset (Days)\!",
t2.\!"Product 2 Manufacturer Name\!", t2.\!"Product 2 Application Type\!",
        t2.\!"Product 2 Application #\!", t2.\!"Product 2 Lot #\!", t2.\!"Product 3 Product Name\!",
t2.\!"Product 3 Product Active Ingredient\!",
        t2.\!"Product 3 Reported Verbatim\!", t2.\!"Product 3 Role\!", t2.\!"Product 3 Reason for Use\!",
t2.\!"Product 3 Strength\!",
        t2.\!"Product 3 Strength (Unit)\!", t2.\!"Product 3 Dose (Amount)\!", t2.\!"Product 3 Dose
(Unit)\!", t2.\!"Product 3 Dosage Text\!",
        t2.\!"Product 3 Dosage Form\!", t2.\!"Product 3 Route\!", t2.\!"Product 3 Frequency\!",
t2.\!"Product 3 Dechallenge\!",
        t2.\!"Product 3 Rechallenge\!", t2.\!"Product 3 Start Date\!", t2.\!"Product 3 Stop Date\!",
t2.\!"Product 3 Therapy Duration (Days)\!",
        t2.\!"Product 3 Therapy Duration (Verbatim)\!", t2.\!"Product 3 Time To Onset (Days)\!",
t2.\!"Product 3 Manufacturer Name\!", t2.\!"Product 3 Application Type\!",
        t2.\!"Product 3 Application #\!", t2.\!"Product 3 Lot #\!", t2.\!"Product 4 Product Name\!",
t2.\!"Product 4 Product Active Ingredient\!",
        t2.\!"Product 4 Reported Verbatim\!", t2.\!"Product 4 Role\!", t2.\!"Product 4 Reason for Use\!",
t2.\!"Product 4 Strength\!",
        t2.\!"Product 4 Strength (Unit)\!", t2.\!"Product 4 Dose (Amount)\!", t2.\!"Product 4 Dose
(Unit)\!", t2.\!"Product 4 Dosage Text\!",
        t2.\!"Product 4 Dosage Form\!", t2.\!"Product 4 Route\!", t2.\!"Product 4 Dechallenge\!",
t2.\!"Product 4 Rechallenge\!",
        t2.\!"Product 4 Start Date\!", t2.\!"Product 4 Stop Date\!", t2.\!"Product 4 Therapy Duration
(Days)\!", t2.\!"Product 4 Therapy Duration (Verbatim)\!",
        t2.\!"Product 4 Time To Onset (Days)\!", t2.\!"Product 4 Manufacturer Name\!", t2.\!"Product 4
Application Type\!", t2.\!"Product 4 Application #\!",
        t2.\!"Product 4 Lot #\!", t2.\!"Product 5 Product Name\!", t2.\!"Product 5 Product Active
Ingredient\!", t2.\!"Product 5 Reported Verbatim\!",
        t2.\!"Product 5 Role\!", t2.\!"Product 5 Reason for Use\!", t2.\!"Product 5 Strength\!",
t2.\!"Product 5 Strength (Unit)\!",
        t2.\!"Product 5 Dose (Amount)\!", t2.\!"Product 5 Dose (Unit)\!", t2.\!"Product 5 Dosage Text\!",
t2.\!"Product 5 Dosage Form\!",
        t2.\!"Product 5 Route\!", t2.\!"Product 5 Dechallenge\!", t2.\!"Product 5 Rechallenge\!",
t2.\!"Product 5 Start Date\!",
        t2.\!"Product 5 Stop Date\!", t2.\!"Product 5 Therapy Duration (Days)\!", t2.\!"Product 5 Therapy
Duration (Verbatim)\!", t2.\!"Product 5 Time To Onset (Days)\!",
        t2.\!"Product 5 Manufacturer Name\!", t2.\!"Product 5 Application Type\!", t2.\!"Product 5
Application #\!", t2.\!"Product 5 Lot #\!",
        t2.\!"Product 6 Product Name\!", t2.\!"Product 6 Product Active Ingredient\!", t2.\!"Product 7
Product Name\!", t2.\!"Product 7 Product Active Ingredient\!",
        t2.\!"Product 8 Product Name\!", t2.\!"Product 8 Product Active Ingredient\!", t2.\!"Product 9
Product Name\!", t2.\!"Product 9 Product Active Ingredient\!",
        t2.\!"Product 10 Product Name\!", t2.\!"Product 10 Product Active Ingredient\!",
t2.\!"Race/Ethnicity\!", t2.\!"Product 4 Frequency\!",
        t2.\!"Product 5 Frequency\!", t2.\!"Product 1 Combination Product\!", t2.\!"Product 2 Combination
Product\!", t2.\!"Product 2 NDC #\!",
        t2.\!"Product 3 Combination Product\!", t2.\!"Product 3 NDC #\!", t2.\!"Product 4 NDC #\!",
t2.\!"Product 1 Compounded Product\!",
        t2.\!"Product 4 Combination Product\!", t2.\!"Product 5 Combination Product\!", t2.\!"Product 2
Compounded Product\!"
FROM \!"RO2 in both Primary_P2_P3_P4_P5_P6_P7_P8_P9_P10\!"  t1
        LEFT OUTER JOIN \!"All cases for DILI\!" t2
                ON  (  ( t1.\!"RO2 matching either the Primary, P2, P3, P4, P5, P6, P7, P8, P9, P10 or
their Active ingredient\!" = t2.\!"Product 1 Product Name\!" ) OR  ( t1.\!"RO2 matching either the
Primary, P2, P3, P4, P5, P6, P7, P8, P9, P10 or their Active ingredient\!" = t2.\!"Product 1 Product
Active Ingredient\!" )  AND  ( t1.\!"RO2 matching either the Primary, P2, P3, P4, P5, P6, P7, P8, P9, P10
or their Active ingredient\!" = t2.\!"Product 2 Product Name\!" )  OR  ( t1.\!"RO2 matching either the
Primary, P2, P3, P4, P5, P6, P7, P8, P9, P10 or their Active ingredient\!" = t2.\!"Product 2 Product
Active Ingredient\!" )  AND  ( t1.\!"RO2 matching either the Primary, P2, P3, P4, P5, P6, P7, P8, P9, P10
```

```
or their Active ingredient\!" = t2.\!"Product 3 Product Name\!" )  OR  ( t1.\!"RO2 matching either the
Primary, P2, P3, P4, P5, P6, P7, P8, P9, P10 or their Active ingredient\!" = t2.\!"Product 3 Product
Active Ingredient\!" )  AND  ( t1.\!"RO2 matching either the Primary, P2, P3, P4, P5, P6, P7, P8, P9, P10
or their Active ingredient\!" = t2.\!"Product 4 Product Name\!" )  OR  ( t1.\!"RO2 matching either the
Primary, P2, P3, P4, P5, P6, P7, P8, P9, P10 or their Active ingredient\!" = t2.\!"Product 4 Product
Active Ingredient\!" )  AND  ( t1.\!"RO2 matching either the Primary, P2, P3, P4, P5, P6, P7, P8, P9, P10
or their Active ingredient\!" = t2.\!"Product 5 Product Name\!" )  OR  ( t1.\!"RO2 matching either the
Primary, P2, P3, P4, P5, P6, P7, P8, P9, P10 or their Active ingredient\!" = t2.\!"Product 5 Product
Active Ingredient\!" )  AND  ( t1.\!"RO2 matching either the Primary, P2, P3, P4, P5, P6, P7, P8, P9, P10
or their Active ingredient\!" = t2.\!"Product 6 Product Name\!" )  OR  ( t1.\!"RO2 matching either the
Primary, P2, P3, P4, P5, P6, P7, P8, P9, P10 or their Active ingredient\!" = t2.\!"Product 6 Product
Active Ingredient\!" )  AND  ( t1.\!"RO2 matching either the Primary, P2, P3, P4, P5, P6, P7, P8, P9, P10
or their Active ingredient\!" = t2.\!"Product 7 Product Name\!" )  OR  ( t1.\!"RO2 matching either the
Primary, P2, P3, P4, P5, P6, P7, P8, P9, P10 or their Active ingredient\!" = t2.\!"Product 7 Product
Active Ingredient\!" )  AND  ( t1.\!"RO2 matching either the Primary, P2, P3, P4, P5, P6, P7, P8, P9, P10
or their Active ingredient\!" = t2.\!"Product 8 Product Name\!" )  OR  ( t1.\!"RO2 matching either the
Primary, P2, P3, P4, P5, P6, P7, P8, P9, P10 or their Active ingredient\!" = t2.\!"Product 8 Product
Active Ingredient\!" )  AND  ( t1.\!"RO2 matching either the Primary, P2, P3, P4, P5, P6, P7, P8, P9, P10
or their Active ingredient\!" = t2.\!"Product 9 Product Name\!" )  OR  ( t1.\!"RO2 matching either the
Primary, P2, P3, P4, P5, P6, P7, P8, P9, P10 or their Active ingredient\!" = t2.\!"Product 9 Product
Active Ingredient\!" )  AND  ( t1.\!"RO2 matching either the Primary, P2, P3, P4, P5, P6, P7, P8, P9, P10
or their Active ingredient\!" = t2.\!"Product 10 Product Name\!" )  OR  ( t1.\!"RO2 matching either the
Primary, P2, P3, P4, P5, P6, P7, P8, P9, P10 or their Active ingredient\!" = t2.\!"Product 10 Product
Active Ingredient\!" )  )  ;")) << Run;
```

Figure 10 illustrates the matching RO2 compound names with FAERS product names or their active ingredients and the number of cases for such matching. It is obvious that some drugs dominate the DILI cases such as ADALIMUMAB, SORAFENIB, ETANERCEPT and INTERFERON BETA-1A.

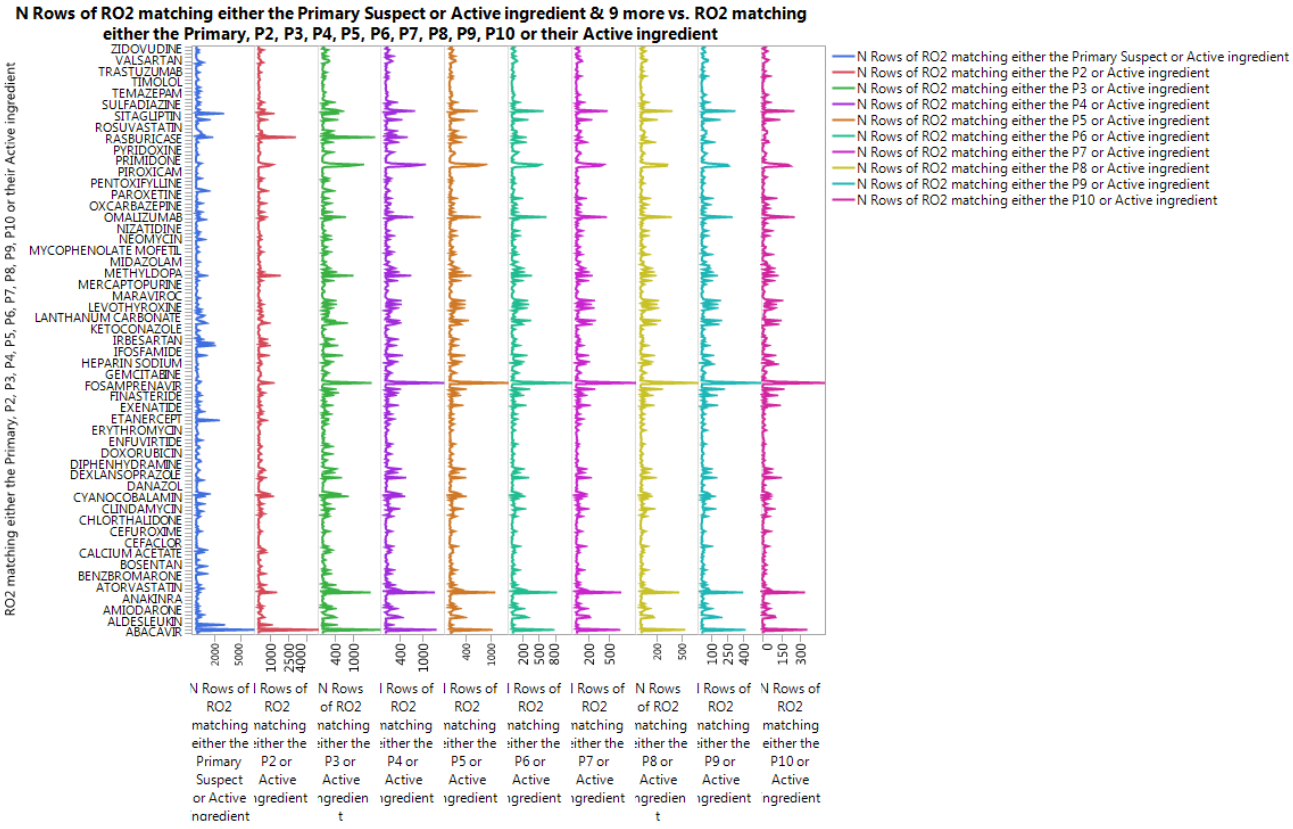


Figure 10: Number of cases that RO2 list matching FAERS data for DILI.

Out of the 1,036 RO2 unique compound names, only 472 compound names or their active ingredients are reported in FAERS. Therefore, the aggregated data has 8,288 cases that will be as input data for text analysis model, supervised and unsupervised models in the next section.

3. PREDICTIVE ANALYSIS

The data set contains 122 variables with Topic1 through Topic5 as target variables. In this paper, we used Topic 2 as a target variable and all the other topics are rejected. However, we already built a model for each topic but for space constraint we will demonstrate the predictive model where Topic 2 is a target variable.

1. TEXT MINING_ TEXT FILTER and CONCEPT LINKS

Text Mining starts with text parsing which identifies unique terms in the text variable and identifies parts of speech, entities, synonyms and punctuation (Rajman and Besancon 1997). The terms identified from text parsing are used to create a term-by-document matrix with terms as rows and documents as variables. A typical text mining problem has more terms than documents resulting in a sparse rectangular terms-by-document matrix. Stop lists help in reducing the number of rows in the matrix by dropping some of the terms (SAS Enterprise Miner 2018). Stop list is a dictionary of terms that are ignored in the analysis. A standard stop list removes words such as "a, about, again, and, after, etc." However, a custom stop lists can be designed by analyst for getting more informed text mining results. Based on the preliminary analysis for our aggregated data as well as by communicating with experts at the FDA, we created a costumed stop list that includes terms appearing in fewer than 5 FAERS cases as well as terms with highest frequencies (i.e. drop term with frequency more than 6000). These terms are deemed as to not add any value to the analysis. Examples of such terms in the custom stop lists are patient, drug, liver, FDA, and so on. We also created a custom synonym data set using the terms extracted from the four data sets. For instance, terms hepatic, hepaticopsida, leafy liverwort, and liver failure are considered as synonyms for this research.

Even after using customized stop lists, in a corpus of several thousands of documents, the term-by-document matrix can contain hundreds and thousands of terms. It becomes computationally very difficult to analyze a matrix with high dimensional sparse data. Singular Value Decomposition (SVD) creates orthogonal columns that characterize the terms data set in fewer dimensions than the document by term matrix. Therefore, SVD can be used to reduce the dimensionality by transforming the matrix into a lower dimensional and more compact form. A high number of SVD dimensions usually summarizes the data better but requires a lot of computing resources. In addition, the higher the number, the higher the risk of fitting to noise (Sanders and DeVault 2004). However, a careful decision needs to be made on how many SVD high dimensions to use. A high number for SVDs can give better results, but high computing

resources are required. It is recommended to try low, medium, and high different values for number of dimensions and compare the results. In this paper, we selected 25 SVD dimensions. Each term identified in text parsing is given a weight based on different criteria. The log frequency weighting (local weights) is selected to assign weights to term/document matrix to control the effect of high-frequency terms in a document. Moreover, mutual information is selected for term weights (global weights) to help in identifying significant terms in separating cases from other cases in the corpus by distinguishing terms that occur in only few documents, but occur many times in those few documents. Text filter is used to reduce the total number of parsed terms or documents that are analyzed. Therefore, we eliminated unnecessary information so that only the most valuable and relevant information is considered. Experimental analysis and expert inputs have been applied to remove unwanted terms and to keep only documents that discuss a liver injury. This help us in reducing data set to smaller one rather than using the original collection that contain hundreds of thousands of documents and hundreds of thousands of distinct terms.

Zipf's Law identifies important terms for purposes such as describing concepts and topics. The number of meanings of a word is inversely proportional to its rank (Konchady 2006). Figure 11 exhibits the exponential decay for the Zipf's Law which is typical for the English language and indicates that our data does not deviate from this law.

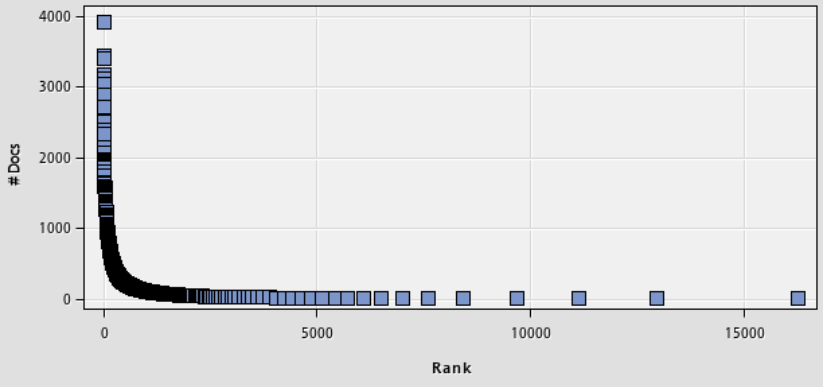


Figure 11: Zipf Plot

The number of documents by frequency plot (Figure 12) exhibits a monotonic behavior while frequency counts that deviate substantially from an approximate linear relationship are suspicious and usually indicate data quality problem. Therefore, the data preprocessing was beneficial in preparation for modeling to obtain useful information.

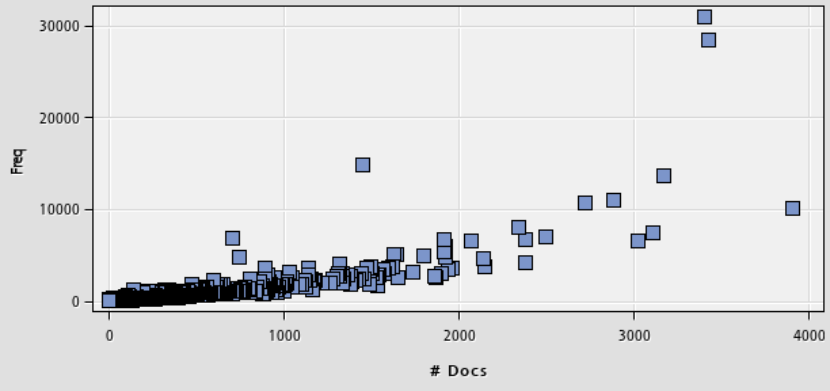


Figure 12: Number of Documents by Frequency

Domain knowledge was utilized to suggest expected frequency distribution for the Role by Freq table. Based on the guidelines from FDA Medical Officers as well as other health care experts, verbs, nouns, adjectives, noun group, and miscellaneous proper nouns should be the expected frequency distribution in the FAERS data (Figure 13).

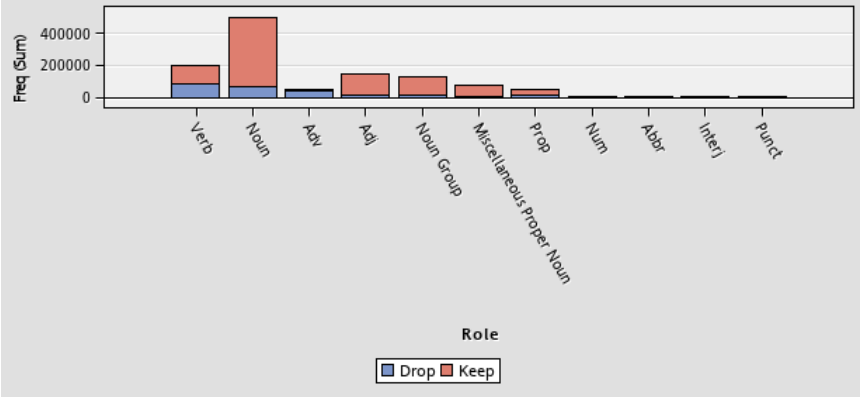


Figure 13: Role by Frequency Distribution

To understand the association between words identified in the corpus, concept linking is an interactive view that illustrates for a given pairs of terms their strength of association with one another which is computed using the binomial distribution (Cerrito 2006). Concept linking is graphical representation where the width of the line between the centered term and a concept link represents how closely the terms are associated. A thicker line indicates a closer association. As an example, Figure 14 shows below the concept linking for the noun group "Hepatobiliary Enzyme".

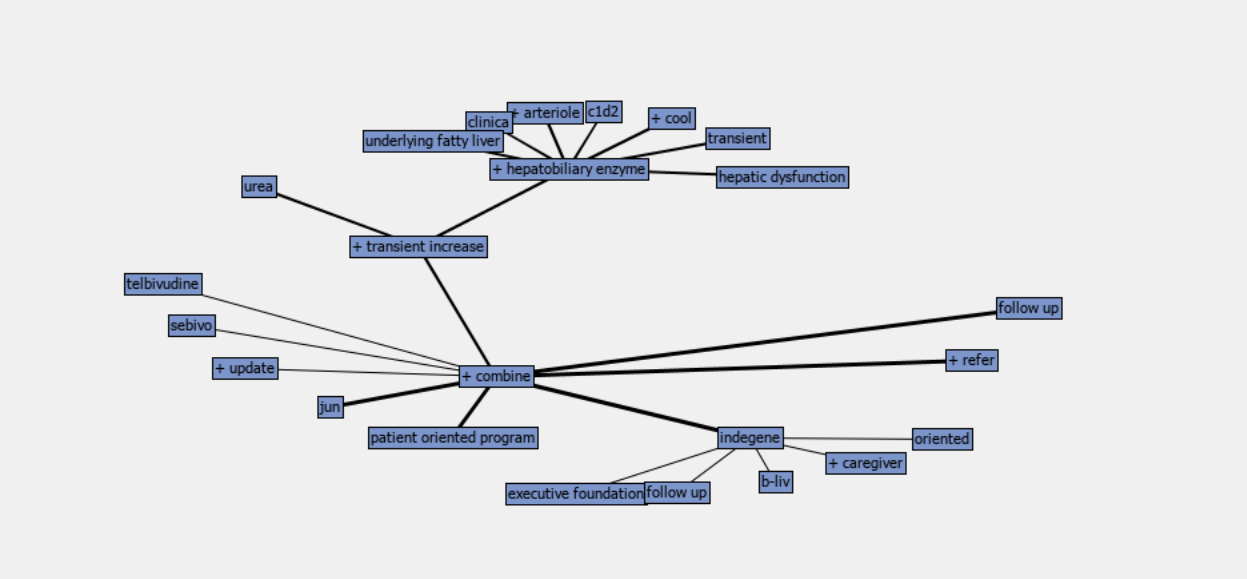


Figure 14: Hepatobiliary Enzyme Concept Linking

The concept Hepatobiliary Enzyme is mainly associated with terms such as transient increase, hepatic dysfunction, underlying fatty liver, arteriole, etc. which indicates that FAERS cases with Hepatobiliary Enzyme mentioned in the case narrative might be serious adverse events or need careful investigation for identifying the causality of death.

2. DECISION TREE

The goal of data mining is to create a good predictive model, which provides us with knowledge and the ability to identify key attributes of business processes that target opportunities (for example, target customers, control risks, or identify fraud). Decision tree models represent one of the most popular types of predictive modeling. Decision trees partition large amounts of data into smaller segments by applying a series of rules. These rules split the data into pieces until no further splits can occur on those pieces. The goal of these rules is to create subgroups of cases that have a lower diversity than the overall sample of population. The purpose of partitioning the data is to isolate concentrations of cases with identical target values. Decision trees are visually represented as upside-down trees with the root at the top and branches emanating from the root. Branches terminate with the final splits (or leaves) of the tree.

In this paper, decision tree is developed to perform the three essential tasks that predictive models performed which are predict new cases, select useful inputs, and optimize complexity. Each of these essential tasks applies to a general principle as shown in (Table 12) below. Decision trees, like other modeling methods, address each of the modeling essentials. Cases are scored using prediction rules. A split-search algorithm facilitates input selection. Model complexity is addressed by pruning.

Predictive Modeling Task	General Principle	Decision Trees
Predict new cases	Decide, rank, or estimate	Prediction Rules
Select useful inputs	Eradicate redundancies and irrelevancies	Split Search
Optimize complexity	Tune models with validation data	Pruning

Table 12: Decision Tree Essential Tasks

Moreover, we utilized the three methods for constructing decision tree models which are interactive method or by hand, the automatic method, and the autonomous method. Many parameters settings for building decision tree have been adjusted in this work. These parameters can be divided in to five groups 1) the number of splits to create at each partitioning opportunity, 2) the metric used to compare different splits, 3) the rules used to stop the autonomous tree growing process, 4) the method used to prune the tree model, and 5) the method used to treat missing values.

Decision tree models are constructed using a recursive algorithm that attempts to partition the input space into regions with mostly primary outcome cases and regions with mostly secondary outcome cases. Model predictions are based on the percentage of primary outcome cases found in each partition. The models can easily accommodate missing values and therefore do not require imputed data. Decision tree models make few assumptions regarding the nature of the association between input and target, making them extremely flexible predictive modeling tools.

To utilize unstructured data in building the decision tree, a text cluster is built prior to the decision tree. The aim of a text cluster is to create clusters that will help with identifying the desired value of the target variable (serious outcome). FAERS cases are assigned to mutually exclusive clusters so each document can belong to only one cluster which is described by a set of terms (Figure 13). This is achieved by deriving a numeric representation for each document. Producing the numeric representation for each cluster is implemented through Singular Value Decomposition (SVD) to organize terms and documents into a common semantic space based upon term co-occurrence. When cases are parsed, a frequency matrix is generated. Depending on the application, the user can define the number of dimensions. For text segmentation, a recommended number of dimensions' ranges from 2 to 50, but for prediction and classification higher values from 30 to 200 are used (Berry and Kogan 2010). Therefore, we selected the number of cluster to be 25. Figure 13 shows part of these 25 cluster.

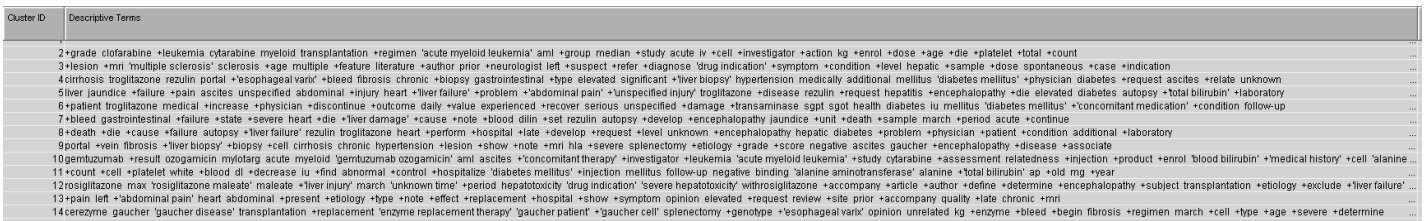
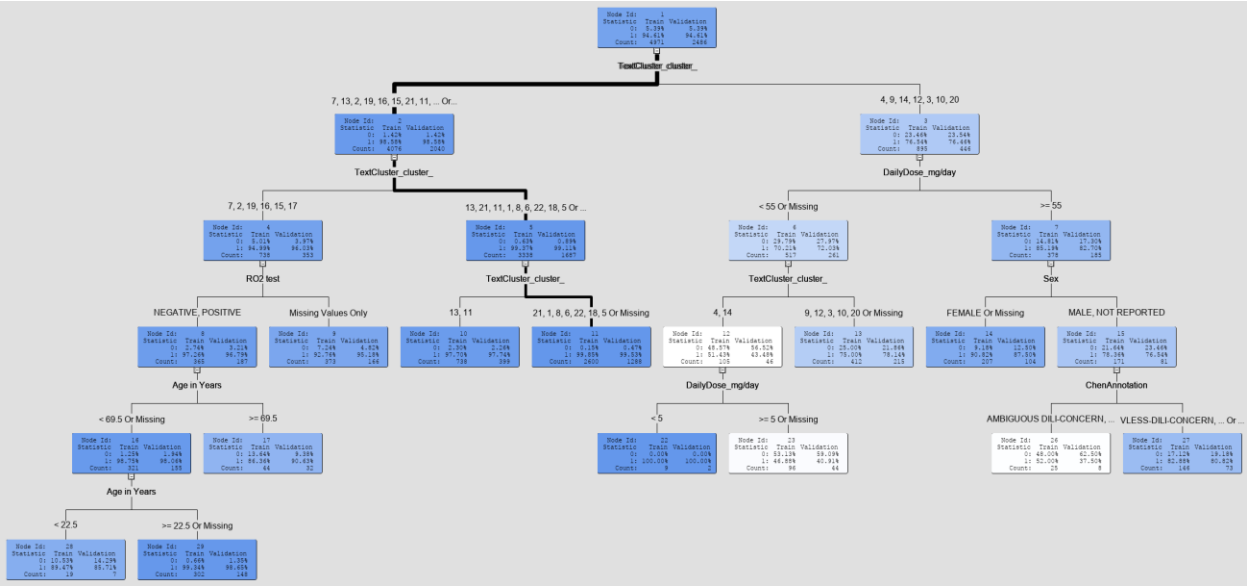


Figure 15: Cluster Description

The output from the cluster analysis is the input to the decision tree modeling. Two decision tree models have been developed. In the first one, the numeric values for the 25 SVDs have been assigned rejected role so that only the nominal values of cluster numbers (TextCluster_cluster_)

will input the decision tree modeling with other FAERS input variables. While on the second model, the SVDs assigned new role as input to the decision tree with other FAERS variables and cluster number variable has been rejected. Figure 16 demonstrates the tree construction for the first model as well as the variables that were important in growing this decision tree.



Variable Importance

Variable Name	Label	Number of Splitting Rules	Importance	Validation Importance	Ratio of Validation to Training Importance
TextCluster_cluster_		4	1.0000	1.0000	1.0000
DailyDose_mg_day	DailyDose_mg/day	2	0.4159	0.2864	0.6887
ChenAnnotation	ChenAnnotation	1	0.2208	0.2298	1.0405
Sex	Sex	1	0.1866	0.1523	0.8163
Age_in_Years	Age in Years	2	0.1357	0.0916	0.6752
RO2_test	RO2 test	1	0.0946	0.0000	0.0000

Figure 16: Decision Tree where the role of cluster numbers set as input role

The classification chart for assigning Topic 2 for this model (Figure 17) shows that 76.56 % of the Topic 2 with level =1 (i.e., it is under topic 2) was classified correct (Topic 2=1) while 19% of the Topic 2 with level=0 (i.e., it is not under topic 2) was misclassified as (Topic 2=1).

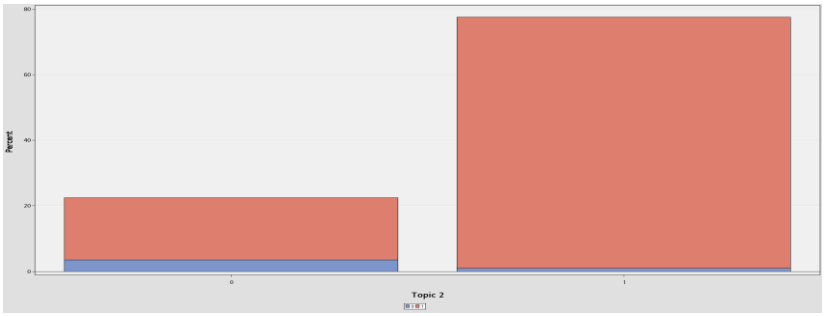
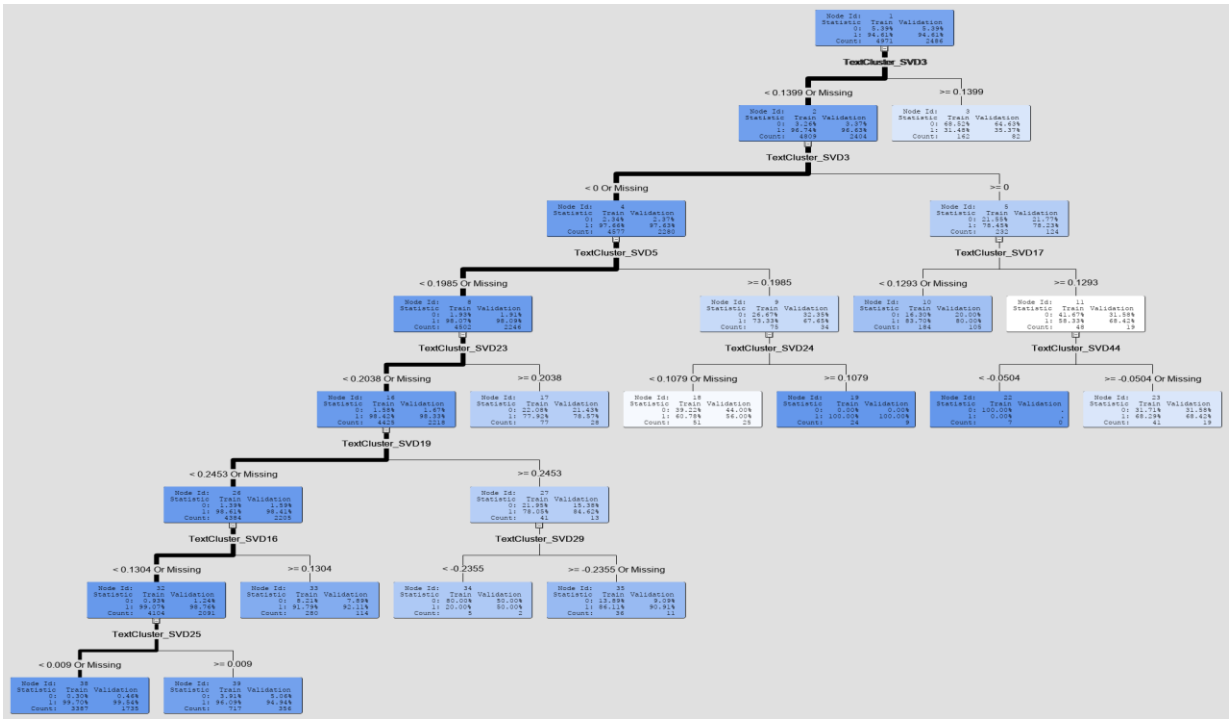


Figure 17: Classification Chart: Topic 2? where the role of cluster numbers set as input role

On the other hand, when SVDs assigned a role of inputs in the metadata while rejecting the cluster number, Figure 18 shows more variables have been contributed to construct this tree.



Variable Importance

Variable Name	Label	Number of Splitting Rules	Importance	Validation Importance	Ratio of Validation to Training Importance
TextCluster_SVD3		2	1.0000	1.0000	1.0000
TextCluster_SVD5		1	0.2455	0.2964	1.2074
TextCluster_SVD23		1	0.2061	0.1778	0.8629
TextCluster_SVD44		1	0.1930	0.0753	0.3904
TextCluster_SVD24		1	0.1831	0.1979	1.0809
TextCluster_SVD17		1	0.1808	0.0000	0.0000
TextCluster_SVD29		1	0.1601	0.0627	0.3917
TextCluster_SVD19		1	0.1514	0.0733	0.4842
TextCluster_SVD16		1	0.1364	0.1168	0.8563
TextCluster_SVD25		1	0.1015	0.1319	1.3000

Figure 18: Decision Tree where the role of SVDs set as input role

The classification chart for the second model (Figure 19) does differ from the first model. Only 16% of the cases were misclassified as Topic 2 with level=1 while 78% of the cases correctly classified as Topic 2 with level =1. Therefore, overall misclassification rate on validation data set is 71% for the first model while the overall misclassification rate on the validation data set for the second mode 59.23%. Therefore, using SVDs as input variable improve the model by reducing the misclassification rate about 12%.

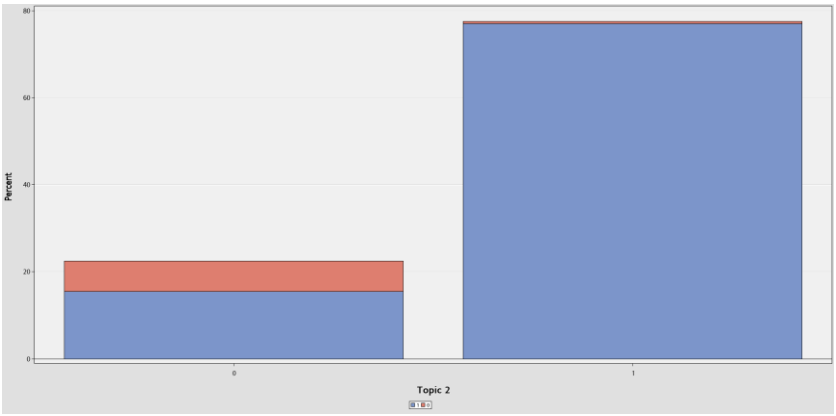


Figure 19: Classification Chart: Topic 2? where the role of SVDs set as input role

CONCLUSION and FUTURE WORK

In this research, post-marketing and pre-marketing DILI databases were combined and utilized in building an analytical model with text mining to advance the DILI facts. Both structured and unstructured data were utilized to increase our predictive power and provide an informative analysis. Our work illustrates a proof of concept of modeling two different data domains (i.e., post- and pre- marketing database) and the feasibility of utilizing the unstructured data in such modeling. This work is in progress and more improvement will be adapted for refining the analysis and utilizing more powerful techniques.

Bibliography

Benigni, R. et al. 2010. "Exploring in vitro/in vivo correlation: lessons learned from analyzing phase I results of the US EPA's ToxCast Project." *J. Environ. Sci. Health C: Environ. Carcinog. Ecotoxicol. Rev.* 28: 272–286.

Berry, M., and J. Kogan. 2010. *Text Mining-Application and Theory.* John Wiley & Sons, Ltd. .

CDER. 2009. *Drug-induced Liver Injury: Premarketing Clinical Evaluation.* https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM174090.pdf.

Cerrito, B.P. 2006. *Introduction to Data Mining using SAS Enterprise Miner.* SAS Publishing .

Chen , Minjun, Vikrant Vijay, Qiang Shi, Zhichao Liu, Hong Fang, and Weida Tong. 2011. "FDA-approved drug labeling for the study of drug-induced liver injury." *Drug Discovery Today* 16: 697-703.

Chen, Minjun, Ayako Suzuki, Shraddha Thakkar, Ke Yu, Chuchu Hu, and Weida Tong. 2016. "DILIrank: the Largest Reference Drug List Ranked by the Risk for Developing Drug-Induced Liver Injury in Humans." *Drug Discovery Today* 648-653.

Chen, Minjun, Jurgen Borlak, and Weida Tong. 2013. "High Lipophilicity and High Daily Dose of Oral Medications Are Associated With Significant Risk for Drug-Induced Liver Injury." *Hepatology* 58 (1): 388-396.

Fontana, R., Seeff, L., Andrade, R., Bjornsson, E., Day, C., Serrano, J., & Hoofnagle, J. 2010. "Standardization of Nomenclature and Causality Assessment in Drug-Induced Liver Injury: Summary of a Clinical Research Workshop." *Hepatology* 52 (2): 730-742.

Harpaz , R et al. . 2013. "Performance of Pharmacovigilance Signal-Detection Algorithms for the FDA Adverse Event Reporting System." *Clinical Pharmacology & Therapeutics* 93: 539-546.

Kaplowitz, N. 2001. "Drug-induced liver disorders: implications for drug development and regulation." *Drug Safety* 24: 483-490.

Konchady, Manu. 2006. *Text Mining Application Programming.* Boston: Charles River Media.

Maddrey, W.C. 2005. "Drug-induced hepatotoxicity." *J. Clin. Gastroenterol* 39 (Suppl.2): 83-89.

Minjun Chen, Ayako Suzuki, Shraddha Thakkar, Ke Yu, Chuchu Hu and Weida Tong. 2016. "DILIrank: the largest reference drug list ranked by the risk for developing drug-induced liver injury in humans." *Drug Discovery Today* 21: 648-653.

Obach, R.S. et al. 2008. "Can in vitro metabolism-dependent covalent binding data in liver microsomes distinguish hepatotoxic from nonhepatotoxic drugs? An analysis of 18 drugs with consideration of intrinsic clearance and daily dose." *Chem. Res. Toxicol.* 21: 1814-1822.

Ostapowicz G, Fontana RJ, Schiødt FV, Larson A, Davern TJ, Han SH, McCashland TM, Shakil AO, Hay JE, Hynan L, Crippin JS, Blei AT, Samuel G, Reisch J, Lee WM, and U.S. Acute Liver Failure Study Group. 2002. "Results of a prospective study of acute liver failure at 17 tertiary care centers in the United States." *Ann. Intern. Med.* 137: 947-954.

Platform, CDER Informatics. 2018. Accessed 2017. http://inside.fda.gov:9003/downloads/cder/officeofsurveillanceandepidemiology/ucm577005.pdf.

Rajman, M., and R. Besancon. 1997. *Text Mining: Natural Language Techniques and Text Mining Applications.* Lausanne, Switzerland: Chapman & Hall.

Rodgers, A.D. et al. 2010. "Modeling liver-related adverse effects of drugs using k nearest neighbor quantitative structure–activity relationship method." *Chem. Res. Toxicol.* 23: 724-732.

Sanders, Annette, and Craig DeVault. 2004. "Using SAS at SAS: The Mining of SAS Technical Support." *SUGI 29 Analytics.* Cary, NC.

SAS Course Note, E. 2016. "Advanced Predictive Modeling Using SAS Enterprise Miner." Cary: SAS Institute Inc.

SAS Enterprise Miner. 2018. "Intriduction to Text Miner." Cary, NC.: SAS Institue Inc.

Schoeters, G. 2010. "The REACH perspective: toward a new concept of toxicity testing." *J. Toxicol. Environ. Health B: Crit. Rev.* 13: 232-241.

Senior, J.R. 2007. "Drug hepatotoxicity from a regulatory perspective." *Clin. Liver Dis.* 11: 507-524.

Shukla, S.J. et al. 2010. "The future of toxicity testing: a focus on in vitro methods using a quantitative high-throughput screening platform." *Drug Discov. Today* 15: 997-1007.

Signal, Empirica. 2017. August . Accessed 2017. http://inside.fda.gov:9003/CDER/OfficeofTranslationalSciences/CDERDataMiningGroup/ucm352563.htm.

Temple, R. 2006. "Hy's law: predicting serious hepatotoxicity." *Pharmacoepidemiol* 15: 241-243.

Xu, J.J. et al. 2008. "Cellular imaging predictions of clinical drug-induced liver injury." *Toxicol. Sci.* 105: 97–105.

Zidek, N. et al. 2007. "Acute hepatotoxicity: a predictive model based on focused illumina microarrays." *Toxicol. Sci.* 2007: 289–302.