

2016

# Using free-text searching and ICD-10 codes to identify prospective patients with drug-induced liver injury

Cassandra Coffman

Follow this and additional works at: <http://commons.emich.edu/theses>



Part of the [Medicine and Health Sciences Commons](#)

---

## Recommended Citation

Coffman, Cassandra, "Using free-text searching and ICD-10 codes to identify prospective patients with drug-induced liver injury" (2016). *Master's Theses and Doctoral Dissertations*. 803.  
<http://commons.emich.edu/theses/803>

This Open Access Thesis is brought to you for free and open access by the Master's Theses, and Doctoral Dissertations, and Graduate Capstone Projects at DigitalCommons@EMU. It has been accepted for inclusion in Master's Theses and Doctoral Dissertations by an authorized administrator of DigitalCommons@EMU. For more information, please contact [lib-ir@emich.edu](mailto:lib-ir@emich.edu).

Using Free-Text Searching and ICD-10 Codes to Identify Prospective Patients with Drug-  
Induced Liver Injury

by

Cassandra Coffman

Thesis

Submitted to the College of Health and Human Services

Eastern Michigan University

in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

in

Clinical Research Administration

Thesis Committee:

Irwin G. Martin Ph.D, Chair

Stephen A. Sonstein, Ph.D.

December 14, 2015

Ypsilanti, Michigan

### **Acknowledgments**

This thesis would not have been possible without the help and guidance of Dr. Irwin Martin, for always being available for help and guidance when needed and for his support throughout the entire program.

I would also like to thank Dr. Stephen Sonstein for his support of my studies throughout the program and the development that he continues to do for clinical research on a broader scale.

I would like to thank Dr. Robert Fontana for giving me the opportunity to develop my career in clinical research as I complete my master's degree and for helping me develop a research topic that will hopefully be of benefit to our research together.

Finally, I would like to thank my coworkers at Sidetrack Bar and Grill for covering some of my shifts to allow me to work on this research and providing hospitality so that I would not have to drive one hour home after 24 hours of work. I am forever grateful.

### **Abstract**

Drug-induced liver injury (DILI) is a rare cause of liver disease that is difficult to diagnose due to the lack of a confirmatory test. The Drug-Induced Liver Injury Network (DILIN) was developed to enroll patients with DILI to understand the clinical outcomes of the disease. The present study aims to develop a tool to identify patients with DILI for enrollment into the DILIN study. The use of ICD-10 codes and liver injury terms were used to search the electronic medical records (EMRs) of patients at the University of Michigan. It was hypothesized that free-text searching using key words would be superior to the use of ICD-10 codes. Two million patient encounters within a six month window were searched using ICD-10 codes related to toxic liver disease. A total of 489 patients were identified and after manual review 32 cases were confirmed. Using the medical record numbers from the patients identified in the initial search, six liver injury terms were used to search free-text in the EMRs. Twelve cases of DILI were identified using “drug-induced liver injury” and none were found using “hepatotoxicity.” The results of this study showed that it is feasible to identify prospective DILI cases using ICD-codes and free-text searching of liver injury terms.

**Table of Contents**

Acknowledgments.....	ii
List of Tables .....	vi
List of Figures .....	vii
Abstract.....	iii
Chapter 1: Introduction.....	1
1.1 Introduction .....	1
1.2 Study Purpose.....	2
1.3 Thesis Statement.....	3
1.3.1 Specific Aims.....	3
1.3.2 Hypothesis.....	3
Chapter 2: Background .....	4
2.1 Background.....	4
2.2 Literature Review .....	5
2.3 The DILIN Prospective Study.....	8
2.4 Electronic Medical Record .....	9
Chapter 3: Methods.....	10
3.1 Medical Record Search Techniques Using Data Direct and EMERSE .....	10
3.2 Medical Record Review .....	12
Chapter 4: Results.....	15

4.1 Initial Search and Manual Chart Review.....	15
4.2 Search Refinement.....	17
4.3 EMERSE: Review of Searches 1 and 2.....	18
4.4 Characteristics of Medical Records and DILI Cases.....	21
Chapter 5: Discussion .....	24
5.1 Utilization of ICD-10 Codes vs Free-Text Searching .....	24
5.2 Possible Reasons for Low Number of Cases.....	28
5.3 Study Limitations and Recommendations for Future Research .....	30
Chapter 6: Conclusion.....	32
References.....	33
Appendix A: DILI Inclusion and Exclusion Criteria.....	36
Appendix B: Result of ICD-10 Codes from Data Direct Search.....	37

**List of Tables**

<u>Table</u>	<u>Page</u>
1. Search Sensitivities Using “Drug-Induced Liver Injury” as a Key Term.....	18
2. Search 1 Relative Sensitivities Using 6 Liver Injury Terms.....	20
3. Search 2 Relative Sensitivities Using 6 Liver Injury Terms.....	21
4. Disqualified DILI Cases From Enrollment into the DILIN Study.....	22
5. Common ICD-10 Codes for Confirmed DILI Cases.....	23

**List of Figures**

<u>Figure</u>	<u>Page</u>
1. Word segmentation of keyword selections.....	12
2. Prevalence of key terms used in patient’s EMRs.....	16
3. Output search results using free-text searching.....	27



## **Chapter 1: Introduction**

### **1.1 Introduction**

Drug-induced liver injury (DILI) is an adverse reaction that is largely unexpected and caused by the administration of a drug or herbal dietary supplement (HDS) (Overby et al., 2013). In the United States, DILI is the leading cause of acute liver failure (ALF) and the most common adverse drug reaction (ADR) for medications. About 14 in 100,000 treated patients per year will develop DILI (Bell & Chalasani, 2009; Bjornsson, Bergmann, Kvaran, & Sigurdur, 2013). Because there is no objective laboratory test to diagnose DILI, this rate may be even higher. Thus, the Drug-Induced Liver Injury Network study (DILIN) was developed in 2003 by the National Institute of Diabetes and Digestive Kidney Diseases (NIDDK) to better understand the natural history, clinical, environmental, immunological, and genetic risk factors of DILI (Fontana et al., 2009).

Diagnosing DILI is challenging. Many patients are on multiple medications and herbal dietary supplements. Furthermore, these products are taken at and for different periods of time. Much medical history, laboratory tests, and exclusion of competing causes are necessary to provide a more descriptive picture. Because no single lab test exists to confirm a DILI diagnosis, causality is assessed by experts, based on the above criteria (Fontana et al., 2013).

With the growing numbers of patients taking medications and herbal dietary supplements, there is an increasing chance of developing an ADR of DILI (Bell & Chalasani, 2009). The need to enroll DILI cases into registry studies is of high

importance so that the etiology and mechanisms of DILI can be better understood and we can develop ways to treat, diagnose, and prevent DILI.

Traditional methods of screening patients for DILI rely largely on physician referrals to liver specialists to confirm a DILI diagnosis. With advancements in technology, data mining in electronic medical records (EMRs) serves as a useful tool to help identify patients with DILI. Currently, there are no effective methods to screen for patients with DILI. As more drugs have been approved each year, the number of patients with drug-induced liver injury has continued to increase (Hayashi & Chalasani, 2015).

One way to search medical records for potential patients is to use the International Classification of Diseases (ICD) coding system. The ICD codes are commonly used to code the diagnoses of patients and billing of physician services. It is also used to monitor the prevalence of diseases. (International Classification of Diseases (ICD-10-CM/PCS) Transition- Background, 2015).

## **1.2 Study Purpose**

The aim of this study is to develop an efficient, effective and specific screening method using free-text searching and ICD-10 codes to identify patients with drug-induced liver injury prospectively. By developing a searching strategy to identify patients with DILI, it is hoped that increased enrollment into prospective registries will provide a better understanding of the etiologies of DILI.

### **1.3 Thesis Statement**

The current study will aim to develop a simple searching method using free-text searching of EMRs and ICD-10 codes to identify patients with DILI. An attempt will be made to answer the following questions:

#### **1.3.1. Specific Aims.**

1. Which terms used for a free-text searching algorithm yield the highest results in identifying the number of probable DILI cases?
2. Does the use of free-text searching methods yield higher positive predictive or negative predictive values when compared with ICD-10 codes?
3. Are free-text searching methods more or less specific and sensitive when compared with ICD-codes?
4. Will the development of a free-text searching algorithm increase the number of subjects enrolled into the Drug-Induced Liver Injury study?

#### **1.3.2. Hypothesis.**

Based on current literature, it is expected that free-text searching methods will be superior in identifying a higher number of probable DILI cases. It is unclear as to which terms will yield higher results. ICD-10 codes should be more sensitive, specific, and yield lower positive predictive values (PPV) when compared to free-text searching methods. It is expected that over time, the number of subjects enrolled in to the DILIN study will increase; however, this would be dependent on an effective algorithm and would need to be examined over time to determine its reliability. Therefore, the results of this study may be limited to aims 1–3.

## Chapter 2: Background

### 2.1 Background

There is an increasing need for retrieving clinical data for research purposes. The data can be used for measuring health care quality, monitoring indicators of adverse events, screening for enrollment into clinical trials (Baldwin, 2008). Most of the data, however, lies in the narrative reports of the electronic medical records (EMR), and difficulty lies in its retrieval. A study by all of the departments of family medicine at Swedish Universities demonstrated that it was feasible to extract and store data from patients' medical records (Mansson, Nilsson, Bjorkelund, & Strender, 2004).

Extracting clinical data can be done through the use of natural language processing (NLP). Natural language processing analyzes linguistics found in the narrative text of EMRs. Using computer software, "word stemming," "segmentation," and "normalization" can transform the natural text into smaller phrases where the prefixes, suffixes, and punctuations are removed to reduce variation and then map them to a structured coding system where they can be analyzed (Travers & Haas, 2003). There are many challenges with NLP systems. When phrases are used with the same meaning but typed in the notes differently, they would get coded differently. For example, "drug-induced liver injury" would get coded differently than "liver injury due to drugs" even though they have the same meaning. In addition, natural language processing systems are quite expensive, and free-text searching methods may be just as effective (Baldwin, 2008).

The narrative text can also be coded according to International Classification of Diseases (ICD) codes as a way of coding narrative text. According to the Centers for Disease Control and Prevention, the ICD-9 index has over 14,025 diagnosis codes and 3,824 procedure codes, yet there is no code for DILI (International Classification of Diseases (ICD-10-CM/PCS) Transition- Background, 2015). Jinjuvadia, Kwan, and Fontana previously demonstrated that even using ICD-9 codes of “acute liver injury” and “drug poisoning” and cross linking them with the names of specific medications known to cause DILI was unsuccessful in identifying patients with idiosyncratic DILI (Jinjuvadia, Kwan, & Fontana, 2007). In October of 2015, the new ICD-10 index was released. There are 69,823 diagnosis codes and 71,924 procedure codes (International Classification of Diseases (ICD-10-CM/PCS) Transition- Background, 2015). Several of these codes such as “toxic liver disease,” which includes “drug-induced idiosyncratic (unpredictable) liver disease” and “hepatic failure (acute) (chronic) due to drugs,” are specific to idiosyncratic DILI (International Classification of Diseases (ICD-10-CM/PCS) Transition- Background, 2015). To date, there have been no studies examining the use of ICD-10 codes to identify patients with idiosyncratic DILI. Using the more specific ICD-10 coding index has the potential to be useful in screening for prospective patients with DILI and is worth studying.

## **2.2 Literature Review**

Recent studies have demonstrated that NLP can be a useful tool in data extraction from narrative text. In a literature review by Warrer et al. 2015, it was demonstrated that data mining in EMRs using free-text searching and NLP methods were capable of identifying ADRs from medication use. The review demonstrated that more advanced

methods like NLP identified more ADRs than manual chart reviews. However, it was also shown that more ADRs were missed with NLP when compared with free-text searching. This study was limited by the number of studies (seven) included in the review and the differing study designs. (Warrer, Hansen, Juhl-Jensen, & Aagaard, 2011).

In a review of the studies included in the Warrer et al. (2011) literature review, Honigman et al. (2001) conducted a study evaluating four different search methods to identify ADRs. They looked at ICD-9 codes, allergy codes, free-text searching, and computerized event monitoring systems. Of the four methods, free-text searching of notes found in the EMR had the highest number of ADRs identified with an overall sensitivity of 90.6%, but the positive predictive value (PPV) was relatively low at 7.2%. The combined PPV for all four search methods was slightly higher at 7.5%. The negative predictive value (NPV) fared much better at 99.2% (Honigman et al., 2001).

Another study included in the Warrer et al. (2011) literature review examined strategies for identifying ADRs in the elderly. Field et al. (2004) found that when comparing free-text searching to provider notes from an internal ADR reporting system, manual reviews of discharge summaries, administrative incident reports, and emergency department notes, and computer-generated signals, free-text searching was superior in identifying ADRs. Sensitivity was lower in this study, yet the PPV was much higher at 39% and 12% respectively (Field et al., 2004).

In a study conducted by Heidemann, Law, and Fontana (2015), 101 DILI cases were identified using a complex natural language processing algorithm to identify retrospective patients with DILI that were attributed to eight specific drugs. A total of 2,564 potential cases were reviewed upon the use of 14 liver injury terms when cross

linked with the names of eight specific medications (Heidemann, Law, & Fontana, 2015). These terms included *drug-induced liver toxicity*, *drug-induced liver injury*, *DILI*, *drug-induced hepatitis*, *liver injury*, *drug-induced liver disease*, *hepatotoxicity*, *liver damage*, *liver toxicity*, *drug-induced hepatotoxicity*, *drug-induced liver damage*, *drug-hepatotoxicity*, and *adverse liver reaction* (Heidemann, Law, & Fontana, 2015). When all 14 terms were used, a total of 2,564 potential DILI cases were found; after review of all of these, only 101 of them were true DILI. No true cases were found using the terms *drug-induced hepatotoxicity*, *drug-induced liver damage*, *drug-hepatotoxicity*, or *adverse liver reaction*, while *liver disease* yielded the most results; with 2,268 potential cases, only 57 were true DILI. The positive predictive value was only 4% when all search terms were used. When the four high yielding liver injury terms were used (*DILI*, *drug-induced liver injury*, *drug-induced liver toxicity* and *drug-induced hepatitis*) the positive predictive value increased to 64% (Heidemann, Law, & Fontana, 2015).

A study by Warrer et al. (2015) looked at clinical notes of 207 patients with type 2 diabetes mellitus who were on either one of three classes of glucose-lowering medications. The objectives of the study were to see if there was potential using clinic notes in the electronic medical records of patients and if there was enough information to attribute causality of the drugs to any adverse drug reactions (ADRs). Results showed that 163 ADRs were found. These corresponded to 27 terms grouped by organ system. They include, but aren't limited to *increased heart rate*, *stomach ulcers*, *abdominal pain*, *gastro esophageal reflux*, *angioedema* and *skin reactions*. Regarding causality assessment, 14% of the ADRs were definite, 60% were probable, and 26% were possible. Fifteen of the 163 ADRs were unlabeled. The data from this study suggest that the use of

clinic notes in EMRs may serve to be a useful tool in detecting new ADRs and assessing causality (Warrer et al., 2015).

Drug-induced liver injury is the most common ADR that leads to an arrest in the development of new medications and withdrawal of existing drugs from the market (Bell & Chalasani, 2009). With the increase in the number of DILI cases arising from drugs and herbal and dietary supplements (HDS) and no effective screening method to identify a large number of patients, the current study will aim to develop a simple searching method using free-text searching of EMRs and ICD-10 codes to identify patients with DILI.

### **2.3 The DILIN Prospective Study**

The DILIN network was established by the NIDDK in 2003 to help physicians understand the etiologies, outcomes, and risk factors of patients diagnosed with DILI. There are currently six sites enrolling patients into the DILIN study, the University of Michigan being one of those six. The prospective arm of the study enrolls patients who have had a liver injury due to a drug or HDS within six months of onset and follows these subjects prospectively for either 6 or 24 months. Inclusion criteria are dependent upon lab values and the suspected agent attributed to their liver injury. Patients who have competing causes of liver injury such as Hepatitis A, B, C, cholangitis or alcoholic hepatitis, or who have suffered a liver injury due to acetaminophen toxicity are excluded from this study. A complete list of inclusion and exclusion study criteria can be found in Appendix A.



Traditionally, potential subjects are referred by physicians at the University of Michigan or outside hospitals. Other methods include data-mining in EMRs for eligible subjects. A HIPPA waiver was obtained by the University of Michigan from the Institutional Review Board (IRB) to be able to search medical records of patients. All subjects who participate in the study are required to sign written, informed consent that has been approved by the IRB.

## **2.4 Electronic Medical Record**

The University of Michigan Health System is host to 2.1 million patient visits and 47,000 hospital stays in 1000 hospital beds each year (Facts and Figures, 2015). All encounters both inpatient and outpatient is captured in the University of Michigan Health System's EMR. In 2012, the University of Michigan Health System purchased an EMR known as MiChart from Epic Systems Corporation. MiChart has the capabilities of running customized reports based on specific search criteria such as problem lists, emergency visit encounters, and ICD codes. Other institutionally supported tools than can be used to generate reports using EMRs include Data Direct and The Electronic Medical Record Search Engine (EMERSE). These enable one to generate aggregate counts and statistics and search dictations from the EMRs for specific text and language used to answer a specific question (Self-Serve Data Tools, 2015). Both of these tools will be used for the purpose of this study.

## Chapter 3: Methods

### 3.1 Medical Record Search Techniques Using Data Direct and EMERSE

Data Direct has the capabilities of creating specific queries based on specific filters with defining criteria such as demographics, encounters, medications administered and ordered, labs, diagnoses, ICD codes, procedures, and problem lists. Queries were created based upon the inclusion and exclusion criteria from the DILIN study and the use of liver injury specific ICD-10 codes. As part of the inclusion criteria for the DILIN prospective study, individuals must have had their liver injury from a drug or HDS within 6 months of the date of enrollment into the study. Meaning, a person who had a liver injury greater than 6 months of the search date used would be excluded from the prospective study.

In the initial data direct search, a query was generated that searched the problem list and text of all inpatient encounters from June 23, 2015 and December 23, 2015. The search was limited to living patients who were greater than two years old. Key terms specific to DILI were searched when selecting the ICD-10 codes. For example, when searching for *DILI*, the terms *drug-induced liver injury*, *drug-induced liver disease*, *drug-induced injury of the liver*, and *drug induced liver disease* were the terms that resulted. These were all coded in the same manner (<K71.9> [573.3] [E980.5]) or (K75.9). They were used in combination with ICD-10 diagnosis codes, specific to DILI that were inclusive of ICD-9 codes. These codes were *toxic liver disease unspecified* (K71.9), *inflammatory liver disease unspecified* (K75.9), *poisoning by unspecified drug or medicinal substance— undetermined whether accidentally or purposely inflicted—* (E980.5) and *hepatitis unspecified* (573.3).

A second technique using EMERSE was used to create a second query. This technique used the list of specifically generated medical record numbers (MRNs) from the data direct search and cross-linked terms specific to the diagnosis of DILI and filtered sources in the patient's medical record that matched these terms and highlighted the specific terms that were searched. After examining the discharge summaries and dictations from the last 20 patients enrolled into the DILIN study, the chart reviews of the generated MRNs from the data direct search, and the terms used in Heidemann, Law and Fontana (2015), a list of four terms were selected that would likely indicate a diagnosis of DILI. These terms are as follows: *drug-induced liver injury*, *drug-induced hepatotoxicity*, *drug-induced hepatitis*, and *drug-induced liver toxicity*. Two additional terms were added after manual review revealed many potential cases. These terms were *liver abnormalities* and *DRESS* (drug rash or reaction with eosinophilia and systemic symptoms). The EMERSE search engine then integrated documents from multiple sources within MiChart, such as dictations written by physicians, physician assistants, nurse practitioners, residents, and any documents scanned from outside institutions, lab results, imaging results, and procedures. Of note, the EMERSE search engine accounts for acronyms, abbreviations, spelling mistakes, and word variations to account for error. For example, if *drug-induced liver injury* is inputted into the search criteria, *drug induced liver injury*, *DILI*, and *liver injury* are screened. Figure 1 shows an example of this word segmentation.

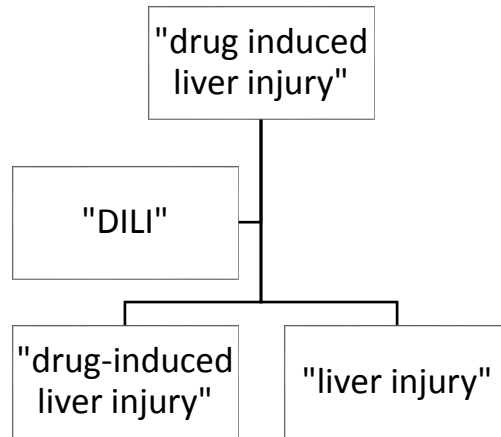


Figure 1 Word segmentation of keyword selections.

### 3.2 Medical Record Review

The output data from the data direct search yielded the subject's medical record number (MRN) that was manually entered into MiChart to review the patients' labs, problem list, and dictations. Each individual MRN derived from the data direct output was examined to determine whether or not this was a potential DILI case. The problem summary was first examined to see if there was mentioning of DILI or some variation of DILI such as *transaminitis*, *elevated LFTs*, *hepatotoxicity*, or any other description related to the possible diagnosis of DILI. It was also noted if key words that had any relation to the words *drug-induced liver injury*, or relationship with the liver were mentioned such as *drug-induced pancytopenia*, or *traumatic brain injury*. When there were no indications or references of DILI, a quick search was initiated within the patient's medical record to search for terms that would likely indicate a DILI diagnosis. Because MiChart is not sensitive, the terms were searched for individually. The terms that were used were *drug*, *induced*, *liver*, and *injury*. If this did not yield any pertinent results, the patients' labs were checked to see if there were any elevations in liver

function tests (LFTs). If there were elevations in alkaline phosphatase, total bilirubin, alanine transaminase, or aspartate transaminase, however slight, it was documented.

Using the same MRNs, from the initial Data Direct Search, the MRNs were then inputted into EMERSE, which pulled up and highlighted any dictation or problem noted in the patients' snapshot of the selected search terms mentioned in section 3.2. These notes from each patient's medical record were then examined further by manual chart review to confirm a DILI diagnosis based on the inclusion and exclusion criteria for the DILIN prospective study and reviewed by the physician investigator (PI) Robert J. Fontana. After careful review of each EMR, each case was determined if it was "probable," "possible," or non-DILI case. If the subject had met the inclusion and exclusion criteria, the patients' healthcare provider was contacted by the PI and the study coordinator then contacted the patient for potential enrollment into the DILIN study.

Based on the results of the initial search, a second search using exclusion criteria was created to decrease the amount of false positive patients. These two searches were then compared to see if there were any significant differences in sensitivities, and positive and negative predictive values between using only codes for DILI versus excluding competing causes of liver injury. Competing causes of liver injury include hepatocellular carcinomas, specified hepatitis such as A, B, C, E, autoimmune hepatitis (AIH) and HIV, poisoning by unspecified drugs, and livers replaced by transplants. The same process of manually reviewing each patient's EMR was applied, and the same key terms were used in EMERSE search for Search 2.

Terms that were similar were then grouped together to be able to identify the codes and terms that were giving positive results. For example, Hepatitis A, B, C, E,

HIV, alcoholic hepatitis, and AIH would all fall under the category *hepatitis specified*.

Other groupings included any problem summary with the word *drug, induced, liver* and *injury* excluding *drug-induced liver injury*, which was categorized as DILI. If a patient's problem list contained multiple references to the liver or terminology used, then all problems and diagnoses were included. If no terminology was found, but the patient had elevated liver enzymes of any sort, they were categorized into *elevated LFTs*.

## Chapter 4: Results

### 4.1 Initial Search and Manual Chart Review

A total of 489 patients were found in the initial search when the terms *drug-induced liver injury*, *drug-induced liver disease*, *drug-induced injury of the liver*, and *drug induced liver disease* were used to search for potential patients. These were all coded in the same manner (<K71.9> [573.3] [E980.5]) or (K75.9). After manual review of each patient, it was found that *any* of these words or codes were mentioned in the patient's problem list, dictation notes, or any other document in the patient's medical record. It was also found that any elevation of LFTs yielded positive results. A list of exact terms that were noted in the problem lists of the medical record and the corresponding ICD-10 code can be found in Appendix B.

A manual chart review was conducted for all 489 patients. It took an estimated time of 500 total minutes to review each medical record (1 minute per case). Because there were 57 different ICD-10 codes that resulted from the search, similar categories were grouped together to see where the results were coming from. Figure 2 shows the prevalence of each term grouped into categories.

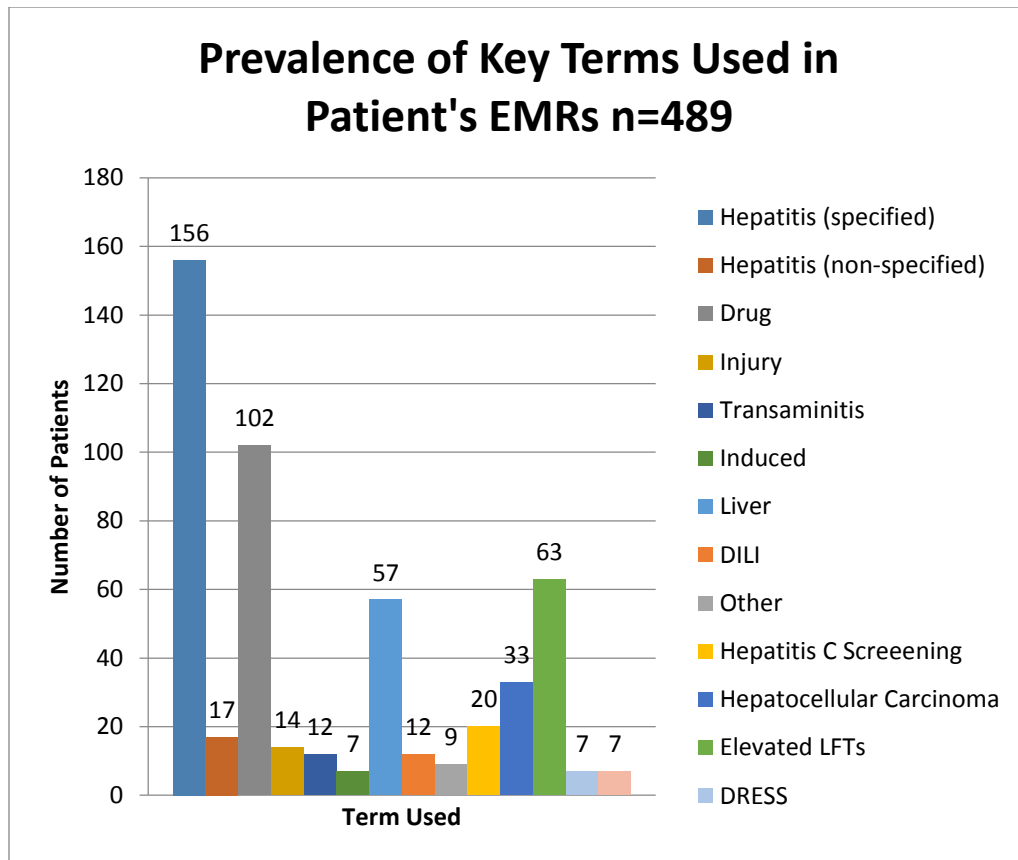


Figure 2 Prevalence of key terms used in patient's EMRs.

The data in Figure 2 indicate that the greatest number of patients had a specified hepatitis listed in their summary of problems, followed by mentioning of the words *drug* and *induced*. *Drug-induced liver injury* (DILI) was only accounted for in the problem list 12 times. When a manual chart review of all 489 patients was conducted, 32 patients were found to have a confirmed DILI diagnosis. Of the total 32 DILI cases that were found manually, when "DILI" was mentioned in the patient's problem list it had only 61.5% sensitivity, missing a total of 20 DILI cases.



## 4.2 Search Refinement

After review of the first search method, it was concluded that there were a greater number of terms that were included in the results of the first algorithm in proportion to the number true positive DILI cases. There were 57 different ICD-10 codes and only 11 that corresponded with the 32 true positive cases, and only 4 different codes that pertained to 11 cases that were eligible for enrollment into the DILIN study. Therefore, the second search was created in hopes to reduce the number of patients to review and increase the positive predictive value (PPV). Unlike the first search where no terms or ICD codes were excluded, the second search included a list of codes pertaining to specified hepatitis such as A, B, C, E, AIH, HIV and alcoholic, hepatocellular carcinomas, poisoning by unspecified drugs, and livers replaced by transplant that were excluded from the results. The results of the second search yielded 345 patients. A manual chart review was conducted in a similar fashion as the Search 1. This took approximately 360 minutes to review each patient. After manual review, 23 cases of DILI were confirmed. Once again, when “DILI” was searched in the problem list of the medical record, only 12 cases of 23 were confirmed. Table 1 shows the sensitivities between each search when DILI was listed in the problem list and used as the only method of review. We can see that 20 cases were missed in Search 1 and 11 in Search 2, indicating that this method was not sensitive. When McNemar’s chi-squared analysis was calculated, a p-value of 0.10 resulted, indicating that there is no statistical difference in the search sensitivities. Additionally, the PPV for each Search 1 and Search 2 were 6.5% and 6.6%, respectively. Time to review each case (1 minute per case) was also the same.

Table 1

*Search Sensitivities Using “Drug-Induced Liver Injury” as a Key Term*

	Number of True DILI Cases	Number of Cases with DILI Problem List	Cases Missed	Search Sensitivity%
Search 1	32	12	20	61.5
Search 2	23	12	11	67.6

### 4.3 EMERSE: Review of Searches 1 and 2

After review of Heidemann, Law, and Fontana’s 2015 study on text-searching, the four highest yielding liver terms were selected to be included in the EMERSE search criteria. They were as follows: *drug-induced liver toxicity*, *drug-induced liver injury*, *drug-induced hepatitis*, and *drug induced hepatotoxicity*. Additionally, based on the confirmed DILI cases from the manual review of the data direct search, *DRESS* and *liver abnormalities* were selected to be part of the search as those terms yielded eight additional positive DILI cases. Each of these six terms were inputted into the EMERSE search engine. The medical record numbers for all 489 patients in Search 1 were copied into the patient list. A search was then run to see which patients contained any of the six terms. Table 2 shows the relative sensitivities and positive and negative predictive values (NPV) for each specific term relative to the number of potential patients. Of the 489 patients, 32 confirmed cases of DILI were found when the charts were manually reviewed; no cases were missed. When *drug-induced liver toxicity* was searched, only one patient had this listed in their medical record. Upon manual review, it was then confirmed that the patient had *drug-induced liver toxicity* listed in either a dictation or under the patient’s problem list. However, based on the number of true DILI cases that were found via manual chart review, 31 cases of DILI were missed. Although specificity,

PPV and NPV were 100%, the sensitivity was low at only 3.1%. Similarly when *drug-induced hepatotoxicity* was searched, zero cases were found. *Drug-induced liver injury* yielded the highest sensitivity at 43.8% and a PPV of 73.6%, followed by *drug-induced hepatitis*, and *liver abnormalities* with sensitivities of 15.6 and 25%, and PPVs of 45.4% and 44.4%, respectively. Though *DRESS* had the highest number of potential patients with 68 mentionings, it had one of the lowest PPVs with 14.7% and a sensitivity of 31.2%. However, it did fare well with finding true DILI cases. A total of 10 cases were found when *DRESS* was searched coming in closely behind *drug-induced liver injury* with 14 total cases.

Table 2

*Search 1 Relative Sensitivities Using 6 Liver Injury Terms*

<b>EMERSE Term</b>	<b>Number of Potential Patients (using specific term)</b>	<b>Number of True Patients</b>	<b>Number of Cases Missed</b>	<b>Search Sensitivity</b>	<b>Search Specificity</b>	<b>PPV%</b>	<b>NPV%</b>	<b>Time To Review (min)</b>
<b>All Terms</b>	489	32	0	100	100	6.5	100	500
<b>Drug-Induced Liver Toxicity</b>	1	1	31	3.1	100	100	100	1
<b>Drug-Induced Liver Injury</b>	19	14	18	43.8	100	73.6	100	20
<b>DRESS</b>	68	10	22	31.2	100	14.7	100	60
<b>Drug-Induced Hepatitis</b>	11	5	27	15.6	100	45.4	100	10
<b>Liver Abnormalities</b>	18	8	24	25	100	44.4	100	20
<b>Drug-Induced Hepatotoxicity</b>	0	0	32	0	100	0	100	0

After the initial EMERSE search was run, a second search was run using the same parameters in order to compare the relative sensitivities and PPVs. Results of search 2 were similar to Search 1. *Drug-induced liver toxicity* had the highest PPV at 100% but only found one confirmed DILI case and had a search sensitivity of 4.3%. *Drug-induced liver injury* identified the most DILI cases (12), while also having a high PPV of 85.7% and a sensitivity of 52.1%. *Liver abnormalities* successfully identified eight confirmed DILI cases and had a PPV of 72.7% and a search sensitivity of 34.7%. Similar to the

results of Search 1, *hepatotoxicity* did not yield any confirmed DILI cases. Every term in both Search 1 and Search 2 had NPVs of 100%. Table 3 shows the results of relative sensitivities and PPVs for each search term.

Table 3

*Search 2 relative sensitivities using 6 liver injury terms*

<b>EMERSE Term</b>	<b>Number of Potential Patients (using specific term)</b>	<b>Number of True Patients</b>	<b>Number of Cases Missed</b>	<b>Search Sensitivity</b>	<b>Search Specificity</b>	<b>PPV%</b>	<b>NPV%</b>	<b>Time To Review (Min)</b>
<b>All Terms</b>	345	23	0	100	100	6.6	100	360
<b>Drug-Induced Liver Toxicity</b>	1	1	22	4.3	100	100	100	1
<b>Drug-Induced Liver Injury</b>	14	12	11	52.1	100	85.7	100	15
<b>DRESS</b>	34	3	20	13	100	8.8	100	30
<b>Drug-Induced Hepatitis</b>	7	5	18	21.7	100	71.4	100	10
<b>Liver Abnormalities</b>	11	8	15	34.7	100	72.7	100	10
<b>Drug-Induced Hepatotoxicity</b>	0	0	23	0	100	0	100	0

#### 4.4 Characteristics of Medical Records and DILI Cases

The 32 confirmed DILI cases were reviewed for potential enrollment into the DILIN study. After careful review by the study coordinator and PI, it was determined that only 11 of the 32 cases were eligible for enrollment. Of the 21 disqualified cases, four of them were previously enrolled into the study, while four cases had liver injury onset

greater than 6 months. Two patients were deceased, two patients refused consent, and two patients received a liver transplant. The remaining seven cases had other competing causes of liver injury that would thereby exclude them from participation and enrollment. A complete list of competing causes can be found in Table 4.

Table 4

*Disqualified DILI cases from enrollment into the DILIN study*

<b>Reason for Disqualification</b>	<b>Number of Patients</b>
<b>Not Within 6 Month Time Frame</b>	4
<b>Already Enrolled</b>	4
<b>Cholangitis</b>	1
<b>Dead</b>	2
<b>Mononucleosis</b>	1
<b>Alcoholic Hepatitis</b>	1
<b>Hepatitis C</b>	2
<b>Hepatitis B</b>	2
<b>Received Liver Transplant</b>	2
<b>Refused Consent</b>	2

The ICD-10 codes of the remaining 11 eligible patients were also examined. The majority of ICD-10 codes used to code these cases were found to have been predominately from *toxic liver disease* (K71.9) and *DRESS Syndrome* (L27.0). Other synonyms for toxic liver disease that can be included were *drug induced liver disease*, *drug-induced disorder of the liver*, and *liver disease drug induced*. Synonyms for DRESS Syndrome included *acniform drug eruption*, *dermatitis due to drug and/or medicine taken internally*, *drug rash* and *eruption due to drug* (International Classification of Diseases (ICD-10-CM/PCS) Transition- Background, 2015). A complete list of ICD-10 codes for the 11 eligible patients can be found in Table 5.

Table 5

*Common ICD-10 Codes for Confirmed DILI Cases*

<b>ICD-10 Code</b>	<b>Diagnosis</b>	<b>Number of Patients</b>
<b>R74.0</b>	Transaminitis	1
<b>K71.9</b>	Toxic liver disease (DILI)	4
<b>L27.0</b>	DRESS Syndrome	4
<b>K75.9</b>	Hepatitis	1
<b>No code: (found from labs and dictations)</b>	DILI	1

## Chapter 5: Discussion

### 5.1 Utilization of ICD-10 Codes vs Free-Text Searching

Given that DILI is difficult to diagnose, and that no objective, diagnostic test exists to successfully determine if a patient has DILI, identifying cases for enrollment is challenging. Over the recent years, a variety of methods have been used to search for potential retrospective DILI cases using ICD-9 codes and natural language processing of EMRs. To date, no studies have been done using ICD-10 codes to search for bona fide prospective DILI cases. In the current study, a unique search method was developed using two existing software programs at the University of Michigan to screen for prospective DILI cases.

The first search was created using Data Direct, a software program that was designed to determine the feasibility of patients to recruit for research studies. Because the University of Michigan's electronic medical record has over two million patients, searching individually is burdensome for research coordinators. In addition, traditional recruitment strategies involve the physician investigator, co-investigators to be aware of the DILIN study. If they see any patients in clinic or while rounding that qualify for the study, the study coordinators are notified. Little opportunity is left for the research coordinator to actively recruit patients. Data Direct is a self-serve tool that allows the coordinator to input specific criteria and parameters into the search engine to populate a list of MRNs of individuals who meet the search criteria (Self-Serve Data Tools, 2015).

While NLP can serve as a useful tool, generalizability is often prevented due to misspellings and synonyms of the same word. In addition, it also involves complex algorithms and can be very expensive (Hersh, Campbell, & Malveau, 1997). The use of



ICD-10 codes provides for a more specific and generalizable way to search specifically for diagnosed DILI cases. With the new ICD-10 coding system, there are now specific codes for DILI. Codes such as *toxic liver disease (K71.9)* include *drug-induced idiosyncratic (unpredictable) liver disease* and *hepatic failure (acute) (chronic) due to drugs* and are specific to idiosyncratic DILI (International Classification of Diseases (ICD-10-CM/PCS) Transition- Background, 2015). Searching specifically for this code, as opposed to text, eliminates idioms, misspellings, and other syntax errors.

Over a 6-month window, an initial search was created (Search 1). There were 489 potential DILI cases that were identified. Of these 489 cases, only 32 were true DILI cases after manual review. After careful review of the ICD-10 codes, however, only 12 cases were coded using K71.9, resulting in a PPV of 2.5% when all 489 cases were considered. When all codes were used for qualifying DILI cases, the PPV value increased to 6.5%, suggesting that while the code for DILI is specific, not all cases were being coded correctly or including the K71.9 code.

A second observation was that most of the cases that resulted from the initial search were non-DILI cases. Manual review of these cases determined that the search was picking up on terms that related to the words *drug, induced, liver and injury* as shown in Figure 2. For example, patients with any type of hepatitis who were greater than two years old that had a hospital visit in the last six months were included in the results of the 489 potential patients. This amounted to 156 cases, all of which were excluded. The high number of non-DILI cases suggests that the search was too sensitive in efforts to identify DILI cases.

Search 1 resulted in a high number of potential DILI cases that took approximately 500 minutes of review time. Because of the amount of time that it took to manually review, a second search (Search 2) was created using the same parameters with excluding diagnoses. After review of Search 1, ICD-10 codes for all forms of hepatitis, (with the exception of non-specific hepatitis), liver transplant, and hepatocellular carcinoma were excluded in effort to create a less sensitive test and cut down on review time. A total of 345 potential patients resulted in Search 2, with 23 total DILI cases. The PPV improved slightly from 6.5% to 6.6% and 12 cases were identified using K71.9. Time to review decreased from 500 minutes to 360 minutes. Similar to Search 1, Search 2 also showed to be too sensitive and including exclusion parameters made no difference identifying bona fide DILI cases. More DILI cases were also missed using Search 2 suggesting that actual DILI cases also contain ICD-10 codes that were on the exclusion list. For example, a subject who was suspected to have DILI but also had acute hepatitis B would have been excluded in the search if hepatitis B (B19.10) was coded and toxic liver disease (K71.9) was not (International Classification of Diseases (ICD-10-CM/PCS) Transition- Background, 2015).

A similar concept was applied using a second search method. EMERSE is an electronic medical record search engine that allows one to search through a patients chart for specific words or phrases. Unlike traditional NLP where grammatical errors and synonyms must be broken down and accounted for, EMERSE automatically accounts for these, making it easier to search and identify eligible patient. Additionally, each inputted word is highlighted and the specific lab, image or dictation is pulled up directly. Figure 3 shows an example of the search output. When each individual mosaic is selected, it

navigates directly to the dictation or section the the EMR containing the highlighted term.

This makes it convenient for researchers to be brought directly to the source of the information instead of having to manually screen through their entire chart.

Patients List 2 (345 Patients)  
 Dates 06/15/2015 to 12/15/2015  
 Terms "drug-induced liver toxicity" "drug induced liver injury" DRESS "drug-induced hepatitis" "liver abnormalities" "drug-induced hepatotoxicity"  
 Searching...  
 Gradient Mosaic Searching patient 247 of 345 06/15/2015 to 12/15/2015 Relevance Low High

MRN	Patient Name	MiChart	Careweb	Radiology	Pathology	Other
34537537						
33129403						
35105372						
35229231						
34664742						
36207607						
36698068						
22874676		■				

Figure 3

*Output search results using free-text searching*

In similar fashion, two searches were run using EMERSE. Search 1 contained the list of 489 MRNs and Search 2 contained 345 MRNs. The same keywords were selected with both searches. When comparing the two searches, for all terms, the PPV was higher even though the total amount of identified cases was less. Having less false positives will increase the PPV and cut down on review time. In Tables 2 and Table 3, the number of potential patients corresponding with each term were significantly reduced, making it much easier to identify DILI cases.

On a term-by-term basis, for both searches 1 and 2, *drug-induced liver injury* had the highest PPV at 73.6% and 85.7%, while *drug-induced hepatotoxicity* had the lowest at 0%, respectively. This indicates that *drug-induced liver injury* is a better predictive term in identifying bona fide DILI cases, while *drug-induced hepatotoxicity* does not. In relation to the potential amount of patients identified with each term, “drug-induced liver injury” also was the most sensitive, and it correctly identified patients with DILI better than *drug-induced liver toxicity*, *DRESS*, *drug-induced hepatitis*, *liver abnormalities*, and *drug-induced hepatotoxicity*.

When comparing ICD-10 searching methods in Data Direct with free-text searching in EMERSE, free-text searching was shown to be superior to the use of ICD-10 codes. In Searches 1 and 2 using ICD-10 codes, the PPVs were 6.5% and 6.6%. In Searches 1 and 2 using free-text methods, all PPVs were greater than 6.6% showing the superiority of free-text searching over the use of ICD-10 codes. Of note, because some patients contained more than one of the selected terms, there is no way to determine the relative PPV for all of the terms combined versus all codes used. However, based on each individual result, we can see that the PPV was higher in every circumstance. Negative predictive values were equal across the board and well as the specificity.

## **5.2 Possible Reasons for Low Numbers of Cases**

There are 69,823 diagnosis codes and 71,924 procedure codes listed in the ICD-10 index (International Classification of Diseases (ICD-10-CM/PCS) Transition-Background, 2015). Because it is up to the physicians, nurses, and medical assistants to update the patients problem list and code each diagnosis, there is a lot of room for variability and inconsistencies in coding. For example, a patient who was seen in the

hospital and found to have DILI would also have elevated liver enzymes by the nature of the disease. This can be specifically coded as DILI, but it would not be incorrect to code it as elevated LFTs, hepatitis unspecified, or transaminitis. This makes it difficult to search using only ICD-10 codes for DILI. Additional codes for hepatitis or transaminitis can be used; however this will likely increase the number of potential patients to screen and increase the amount of time spent manually reviewing charts.

Differential diagnoses also cloud the free-text search. Most of the time when a patient arrives, they are not diagnosed immediately. There are instead multiple possibilities to consider, second opinions, consultations, and follow-ups. All of which are dictated in the patient's medical chart. The following information from a patient's medical record can be examined to demonstrate:

A 52 year old menopausal woman with history of polycystic kidney and liver disease who presents with a 2-week history of progressive nausea, abdominal pain, pruritis and found to have grossly elevated LFTs. Differential diagnosis includes drug-induced liver injury in the setting of recent herbal supplements with black cohosh as a substance notorious for hepatotoxicity, autoimmune hepatitis given positive antinuclear antibodies (ANA+) and family history, as well as viral hepatitis. No acute portal clot on right upper quadrant (RUQ) ultrasound. No known cardiac history, so less suspicious for congestive hepatopathy picture.

In this real-life case study, multiple forms of liver disease are being considered as we can see from this dictation. In the search algorithm that was created in the current study, *drug-induced liver injury* and *hepatotoxicity* would be a positive hit and would be highlighted in the results as a potential DILI case. Now if the diagnosis did not end up

being DILI and was instead AIH, then this would end up being a false positive, contributing to a lower amount of true DILI cases.

The third and final reason for low numbers of DILI cases is that DILI is a rare disease affecting roughly 14 per 100,000 people (Bell & Chalasani, 2009). There are about 2 million outpatient visits and 97,000 inpatient stays per year . Because patients usually have more than one outpatient visits and data have shown that the majority of DILI cases are hospitalized, it is assumed that the number of patients seen at the Univeristy of Michigan on a yearly basis is somewhere between 97,000 and 2 million. Therefore, the 32 identified potential DILI cases are actually comparable to the prevelence of 14 in 100,000. It is recommended that in the future, patient charts are updated to accurately reflect past and present problems, and only specific DILI terms are used to search for potential patients. Even if some cases are missed, several hours of review time will be saved.

### **5.3 Study Limitations and Recommendations for Future Research**

Several limitations exist with using both ICD-10 codes and free-text searching to screen for patients with DILI. Limitations of this study included that the free-text search was dependent upon the ICD-10 search. In other words, to get the list of MRNs to input into EMERSE to screen, they must have first been obtained from the Data Direct search. If the ICD-10 codes are not being picked up by the first search, then there is less potential for positive hits in the free-text search. Secondly, this study was also limited to the University of Michigan Hospitals and Health Systems. While other insitutions may use EPIC for EMRs, both Data Direct and EMERSE are unique to the University of

Michigan. Therefore, this study may be limited to study coordinators at the University of Michigan.

Future research is needed to first determine if similar tools exist at other institutions, particularly at the enrolling DILIN sites. If such tools exist, a similar study would need to be conducted to determine if similar results are found and to confidently say this is a useful method of recruitment. We will also need to evaluate its long term success and probability of identifying DILI cases. Because enrollment is ongoing, we will need to observe the number of cases enrolled into the the DILIN study from this search over an extended timeframe. We can then objectively compare current enrollment rates to prospective rates from the development of this search. Using this method going forward, there should be less cases to review since all cases within the 6 month time frame were already reviewed. When using a shorter window, the amount of cases to review will be more manageable.

## Chapter 6: Conclusion

The current study developed a simple searching method using free-text searching of EMRs and ICD-10 codes to identify patients with DILI. In total, 32 cases of DILI were identified using ICD-10 codes, 11 of which were eligible for enrollment into the prospective DILIN study. The code for *toxic liver disease*, (K71.9) had the best PPV in identifying bona fide DILI cases with ICD-10 searching, while *drug-induced liver injury* had the highest PPV when using free-text searching methods. Refinements made to the initial screen, by excluding codes pertaining to competing causes of liver injury, did not improve the number of cases identified. It did reduce the total review time, but was still proportional in terms of time to review per patient. Refinement of the search will be made, and enrollment rates will be examined going forward to determine the usefulness of this tool.



## References

- Baldwin, K. B. (2008). Evaluating Healthcare Quality Using Natural Language Processing. *Journal for Healthcare Quality*, 30(4), 24-29.
- Bell, L. N., & Chalasani, N. (2009, November). Epidemiology of Idiosyncratic Drug-Induced Liver Injury. *Seminars in Liver Disease*, 29(4), 337-347.
- Bjornsson, E. S., Bergmann, O. M., Kvaran, R. B., & Sigurdur, O. (2013, June). Incidence, Presentation and Outcomes in Patients with Drug-Induced Liver Injury in the General Population of Iceland. *Gastroenterology*, 144(7), 1419-1425.
- Facts and Figures*. (2015). Retrieved December 13, 2015, from University of Michigan Health System: [www.uofmhealth.org](http://www.uofmhealth.org)
- Field, T. S., Gurwitz, J. H., Harrold, L. R., Rothschild, J. M., Debellis, K., Seger, A. C., et al. (2004). Strategies for Detecting Adverse Drug Events Among Older Persons in the Ambulatory Setting. *Journal of the American Medical Association*, 11(6), 492-498.
- Fontana, R. J., Seeff, L. B., Andrade, R. J., Bjornsson, E., Day, C. P., Serrano, J., et al. (2013, April 4). Standardization of Nomenclature and Causality Assessment in Drug-Induced Liver Injury: Summary of Clinical Research Workshop. *Hepatology*, 52(2), 730-742.
- Fontana, R. J., Watkins, B. P., Bonkovsky, L. H., Chalasani, N., Davern, T., Serrano, J., et al. (2009). Drug-Induced Liver Injury Network (DILIN) Prospective Study Rationale, Design and Conduct. *Drug Safety*, 32(1), 55-68.

Hayashi, P. H., & Chalasani, N. P. (2015, June 24). Increasing Impacy of Drug-Induced Liver Injury. *Clinical Liver Disease*, 5(6), 136-138.

Heidemann, L., Law, J., & Fontana, J. R. (2015). A Text Searching Tool to Identify Patients with Idiosyncratic Drug-Induced Liver Injury. *Digestive Diseases and Sciences*, 1-11.

Hersh, W. R., Campbell, E. M., & Malveau, S. E. (1997). Assessing teh Feasibility of Large-Scale Natural Language Processing in a Corpus of Ordinary Medical Records: A Lexical Anyalysis. *Proceedings of the AMIA Annual Fall Symposium*, 580-584.

Honigman, B., Lee, J., Rothschild, J., Light, P., Pulling, R. M., Yu, T., et al. (2001). Using Computerized Data to Identify Adverse Drug Events in Outpatients. *Journal of the American Medical Informatics Association*, 8(3), 254-266.

*International Classification of Diseases (ICD-10-CM/PCS) Transition- Background*. (2015, October 1). Retrieved November 10, 2015, from Centers for Disease Control and Prevention: [cdc.gov](http://cdc.gov)

Jinjuvadia, K., Kwan, W., & Fontana, R. J. (2007, November 1). Searching for a Needle in a Haystack: Use of ICD-9-CM Codes in Drug-Induced Liver Injury. *American Journal of Gastroenterology*, 102, 2437-2443.

Mansson, J., Nilsson, G., Bjorkelund, C., & Strender, L.-E. (2004, March). Collection and Retrieval of Structured Clinical Data From Electronic Patient Records in General Practice. *Scandnavian Journal of Primary Health Care*, 22(1), 6-10.

Overby, C. L., Pathak, J., Gottesman, O., Haerian, K., Perotte, A., Murphy, S., et al.

(2013, December). A collaborative approach to developing an electronic health record phenotyping algorithm for drug-induced liver injury. *Journal of the American Medical Informatics Association*, 20, 243-252.

*Self-Serve Data Tools*. (2015). Retrieved December 13, 2015, from University of Michigan Medical School: [medicine.umich.edu](http://medicine.umich.edu)

The DILIN Research Group. (2015, July 1). Drug-Induced Liver Injury Network (DILIN): A Multi-Center, Longitudinal Study of Drug-and CAM-Induced Liver Injury. *Protocol*. Durham, North Carolina, United States: DILIN Data Coordinating Center.

Travers, D. A., & Haas, S. W. (2003, August 26). Using Nurses' Natural Language Processing Entries to Build a Concept-Oriented Terminology for Patients' Chief Complaints in the Emergency Department. *Journal of Biomedical Informatics*, 36, 260-270.

Warrer, P., Hansen, E. H., Juhl-Jensen, L., & Aagaard, L. (2011). Using Text-Mining Techniques in Electronic Patient Records to Identify ADRs From Medicine Use. *British Journal of Clinical Pharmacology*, 73(5), 674-684.

Warrer, P., Jensen, P., Aagaard, L., Jensen, L., Brunak, S., Krag, M., et al. (2015). Identification of Possible Adverse Drug Reactions in Clinical Notes: The Case of Glucose-Lowering Medicines. *Journal of Research and Pharmacy Practice*, 4(2), 64-72.

## Appendix A: DILI Inclusion and Exclusion Criteria

**Liver Disease DILI Case Definition**Inclusion Criteria:

- Age > 2 years at enrollment into study.
- Evidence of liver injury that is known or suspected to be related to consumption of drug or CAM product in 6-month period prior to enrollment.
- Known chronic hepatitis B or C infection defined by detectable HBsAg or HCV RNA for at least 6 months prior to DILI onset
- Written informed consent from patient or patient's legal guardian.
- Documented clinically important DILI, defined as any of the following:
  - ALT or AST > 5 x ULN or ALP > 2 x ULN confirmed on at least 2 consecutive blood draws in patients with previously normal values.
  - If baseline ALT, AST, or ALP elevated, ALT or AST > 5 x BL or ALP > 2 x BL on at least 2 consecutive blood draws. Baseline is average of 2 measurements performed during 12-month period prior to starting implicated medication.
  - Any elevation of ALT, ALP, or AST, associated with (a) increased total bilirubin [ $\geq 2.5$  mg/dL], in absence of prior diagnosis of liver disease, Gilbert's Syndrome, or evidence of hemolysis or (b) coagulopathy with INR > 1.5 in absence of Coumadin therapy or vitamin K deficiency.

Exclusion Criteria:

Patients with any of the following will not be eligible:

- Competing cause of acute liver injury such as hepatic ischemia that is felt by investigator to be primary reason for observed liver injury and supported by laboratory tests, serologies, liver biopsy, or radiology.
- Known, pre-existing autoimmune hepatitis, primary biliary cirrhosis, primary sclerosing cholangitis, or other chronic biliary tract disease that may confound ability to make a diagnosis of DILI.
- Acetaminophen hepatotoxicity.
- Liver/allogeneic bone marrow transplant prior to development of drug- or CAM-induced liver injury (The DILIN Research Group, 2015).

## Appendix B: Result of ICD-10 Codes from Data Direct Search

<b>ICD 10 code</b>	<b>Terminology</b>	<b>ICD 10 code</b>	<b>Terminology</b>
<b>B15.9</b>	Acute hepatitis Acute type A viral hepatitis, Hepatitis A, Viral Hepatitis A	K75.9	Hepatitis, inflammatory disease of live
<b>B19.10</b>	Cirrhosis due to hepatitis B, Cirrhosis of liver due to Hepatitis B, Cirrhosis, Hepatitis B, Type B Viral Hepatitis	K76.0	Nonalcoholic fatty liver disease, fatty liver, fatty liver disease
<b>B19.2</b>	Hepatitis C, Viral Hepatitis C	K76.6	Portal hypertension
<b>B27.9</b>	EBV, infectious mononucleosis, Epstein barr virus disease	K76.89	Lesion of liver, liver nodule
<b>C22.0</b>	Liver cell carcinoma, hepatocellular carcinoma, hematoma	K76.9	Liver disease, disease of liver
<b>C78.7</b>	Secondary malignant neoplasm of liver and intrahepatic bile duct	K83.0	Cholangitis
<b>D18.09</b>	Hemangioma liver	K85.3	Drug induced acute pancreatitis
<b>D61.811</b>	Other drug-induced pancytopenia	L27.0	Generalized skin eruption due to drugs and medicaments taken internally, drug rash, due to drug rash due to drugs and medicaments, eruption due to drug
<b>D69.59</b>	Drug induced thrombocytopenia	M10.232	Drug-induced gout
<b>D70.2</b>	Drug induced neutropenia	N17.9	Injury kidney, no traumatic, acute, acute renal failure
<b>D72.1</b>	Eosinophil count raised, Eosinophilia, allergic, DRESS syndrome	R16.0	Liver mass
<b>D70.2</b>	Drug induced neutropenia, Neutropenia, drug induced	R73.9	Steroid induced hyperglycemia
<b>E06.4</b>	Drug-induced thyroiditis	R74.0	Elevated ALT, elevated alanine aminotransferase (ALT), elevated transaminase, transaminitis, increased transaminase levels

ICD 10 code	Terminology	ICD 10 code	Terminology
<b>E09.65</b>	Drug or chemical induced diabetes mellitus with hyperglycemia	S24.109	Injury of thoracic spinal cord
<b>E80.6</b>	Hyperbilirubinemia	S43.42	Rotator cuff injury
<b>F19.10</b>	Other psychoactive substance abuse, uncomplicated	T38.0X5A	Hyperglycemia due to steroid induced diabetes mellitus, steroid-induced osteopenia, corticosteroids adverse reaction
<b>G21.11</b>	neuroleptic induced due to drugs	T39.1X2	Intentional acetaminophen overdose, poisoning by 4-aminophenol derivatives intentional self-harm
<b>G21.19</b>	drug induced Parkinsonism	T50.2X5A	Adverse effect of carbonic-anhydrase inhibitors, benzothiadiazides and other diuretics, initial encounter
<b>G25.1</b>	Drug-induced tremor	T50.901A	Poisoning by unspecified drugs, medicaments and biological substances, accidental (unintentional), initial encounter
<b>G25.61</b>	Drug induced tics	T50.902A	Poisoning by unspecified drugs, medicaments and biological substances, intentional self-harm, initial encounter
<b>G62.0</b>	Drug-induced polyneuropathy	T50.905A	Adverse effect of unspecified drugs, medicaments and biological substances, initial encounter
<b>I95.2</b>	Hypotension due to drugs	T88.7	Unspecified adverse effect of drug or medicament
<b>K25.9</b>	gastric ulcer due to drugs	T88.7XXA	Unspecified adverse effect of drug or medicament, initial encounter
<b>K70.10</b>	Acute alcoholic hepatitis, Acute alcoholic liver disease, Alcoholic hepatitis, Alcoholic hepatitis, acute, Alcoholic hepatitis, chronic, Chronic alcoholic hepatitis	V42.7	Liver replaced by transplant
<b>K70.31</b>	Alcoholic cirrhosis w ascites, Ascites due to alcoholic	Z13.850	Traumatic brain injury

<b>ICD 10 code</b>	<b>Terminology</b>	<b>ICD 10 code</b>	<b>Terminology</b>
<b>K71.0</b>	Toxic liver disease with cholestasis, Drug-induced cholestatic hepatitis	Z71.7	HIV
<b>K71.9</b>	Toxic liver disease, unspecified, Drug induced liver disease, Drug-induced disorder of liver, Liver disease, drug induced · Toxic liver disease · Degeneration, degenerative liver (diffuse) K76.89 toxic (acute)	Z72.89	Drug seeking behavior
<b>K72.0</b>	Acute and sub-acute hepatic failure, Acute hepatic failure, Acute necrosis of liver	Z91.81	At risk for fall injury
<b>K75.4</b>	Autoimmune hepatitis		