AHRQ Grant Final Progress Report

Title of Project:

Encoding and Processing Patient Allergy Information in EHRs

Principal Investigator:
Li Zhou, MD, PhD
Associate Professor
Brigham and Women's Hospital, Harvard Medical School

Team Members:

Co-Investigators	David Bates, Diane Seger, Foster Goss, Sarah Slight, Paige Wickner, Kimberly Blumenthal, Max Topaz, Aleena Banerji
Data Coordinator/Manager	Frank Chang
Postdoc Research Fellows	Maxim Topaz, Adrian Wong, Chunlei Tang
Research Assistants	Jason Lau, Joseph Plasek, Kenneth Lai, Neil Dhopeshwarkar, Feiran Zhong, Sean Bethard, Suzanne Blackley, Victor Lei, Kevin Lai, Raymond Doan, Warren Acker, Leigh Kowalski, Carlos Ortega
Domain Expert Panel	Kin Wah Fung, John Kilbourne, Shelly Spiro, Robert C. McClure, George Robinson

Organization: Brigham and Women's Hospital

Inclusive Dates of Project: 09/30/2013-09/30/2018

AHRQ Project Officer: Steve Bernstein

Grant Award Number: 5R01HS022728

Acknowledgement of Agency Support: This project was supported by grant number 5R01HS022728 from the Agency for Healthcare Research and Quality.

Structured Abstract (≤ 250 words)

Purpose: The goal of this study was to explore an important and understudied area in the Health Information Technology and patient safety domains, focusing particularly on allergy terminology standards, natural language processing, and allergy reconciliation.

Scope: The project was conducted at a large integrated healthcare system located in the Boston, Massachusetts, area and using longitudinal electronic health records.

Methods: We mapped structured and free-text allergy entries to standard terminologies and developed qualitative metrics to assess mapping accuracy. We also compared differences between these standard terminologies and the institutional terminology used at Partners. We developed an Allergy Information Extraction and Encoding module to identify allergy information from free-text clinical documents using natural language processing and machine learning methods.

Results: We created a comprehensive knowledge base and an information model for representing allergy information. We created lexicons for food, drug, environmental, and contrast media allergens. The NLP module achieved an F-measure of 96.7% for the detection of allergens. We developed a standalone web application to display allergy information based on output from the NLP module. The application can batch process many notes, display extracted allergies, and compare the extracted allergies with those in the patient's structured allergy list. We integrated our allergy reconciliation NLP Module with Epic NoteReader in the development environment.

Key Words: allergy, electronic medical records, patient safety, health information technology, natural language processing, clinical informatics

Purpose

The goal of this study was to explore an important and understudied area in the Health Information Technology (HIT and patient safety domains, focusing particularly on allergy terminology standards, natural language processing (NLP, and allergy reconciliation. During the course of this project, we developed innovative solutions to address major challenges associated with allergy documentation through the completion of the following specific aims.

Specific Aim 1: Build a comprehensive knowledge base for allergy information
Specific Aim 2: Develop and evaluate an NLP system for processing allergy information
Specific Aim 3: Use NLP output to facilitate allergy reconciliation
Specific Aim 4: Distribute our methods and tool

Scope

Background and Context

Allergies are linked to a number of chronic and serious illnesses (e.g., asthma, sinusitis, hay fever, atopic eczema),and, for some, allergic reactions can be severe or even fatal. Obtaining allergy information is a critical step toward safe prescribing, preventing adverse drug events (ADEs), and reducing the cost of care. Proper documentation and exchange of patients' allergy information within electronic health record (EHR) systems is vital to the appropriate management of patient care, safety, and education.

Clinicians routinely elicit allergy information during the medical interview; however, allergies are often poorly documented in the EHR. Currently, many EHR systems contain an allergy module that allows entry of allergies and adverse reactions in a structured/coded fashion (e.g., from a drop-down menu or pick list). Nevertheless, the entry of free-text allergies continues to be prevalent. When the terminology used by the EHR does not contain the precise allergy that clinicians are looking for, the allergy is often entered as uncoded free-text. The resulting information is neither interoperable across clinical information systems nor easily reusable for other applications (e.g., clinical decision support).

The automatic processing of free-text allergy information is inherently complex. Not only does it involve many of the challenges associated with NLP in general (e.g., lexical variation, spelling errors, ambiguity, implicit information, and so on) but it also is subject to additional challenges unique to the clinical domain. An individual allergy record often contains multiple data elements, including allergy type, allergen name, reaction, episode, severity, and more. Each entry must be processed to allow representation of a complete allergy record with all clinically relevant details, including the semantic relationships among the different elements. Furthermore, automatically encoding these data elements requires integrating multiple standard terminologies.

Settings and Participants

The project was conducted at a large, integrated healthcare system, Partners HealthCare System, located in the Boston, Massachusetts, area. During this study, Partners transitioned from LMR, a homegrown EHR system, to Epic EHR.

Methods

Data Sources and Collection

Partners HealthCare System is an integrated healthcare system in the Boston, Massachusetts, area. Founded by Brigham and Women's Hospital and Massachusetts General Hospital, it also includes multiple community hospitals, a physician network, and other health-related entities. Partners Enterprisewide Allergy Repository (PEAR) is a longitudinal allergy database containing more than 3 million active allergies for more than 2 million patients and covering the timespan between 1990 and 2015.

For Aim 1, we used clinical allergy entries stored in PEAR between 2002 and 2013 and free-text clinical documents from Partners hospitals. We used three different document types: discharge summaries (30%), outpatient visit notes (40%), and emergency room notes (30%), each obtained from their respective settings (inpatient, outpatient, and emergency department). Free-text data were retrieved from the Partners clinical research data repository.

For Aim 2, because allergy information is not contained in all clinical notes, we used a stratified convenience sample designed to contain richer allergy information. We excluded notes that typically do not contain allergy information (e.g., procedure notes) by examining the subject heading. We then developed an annotation schema and guidelines detailing specifications for annotating allergens, adverse reactions, severity, instances of negation, and other information that were based upon which annotation was conducted by domain experts, including physicians and pharmacists.

For Aim 3, free-text (i.e., clinical notes) and structured (i.e., structured allergy lists) data were obtained from Partners' ambulatory EHR system for a random sample of 200 patients.

Study Design and Measures

Aim 1

We mapped both structured and free-text allergy entries in PEAR to Systematized Nomenclature of Medicine - Clinical Terms (SNOMED-CT) and RxNorm and developed qualitative metrics to assess the level of mapping accuracy, including exact, partial, incorrect, and no match. We also compared differences between these standard terminologies and the institutional terminology used at Partners at both the lexical and the semantic levels. We compared four common adverse sensitivity information models---Health Level 7 Allergy and Intolerance Domain Analysis Model, HL7-DAM; the Fast Healthcare Interoperability Resources, FHIR; the Consolidated Continuity of Care Document, C-CDA; and OpenEHR---and evaluated their coverage on a corpus of inpatient and outpatient notes.

We assembled a Domain Expert Panel to provide advice on the study at multiple stages, particularly on the development, validation, and distribution of the allergy knowledge base and terminology subsets. By collaborating with national experts in HIT, terminology standards, and medication safety, we were able to ensure that our findings were generalizable and readily adoptable by other healthcare institutions and EHR vendors, furthering our goal of promoting standards adoption, data interoperability, and meaningful use of HIT.

We analyzed 2,471,004 adverse reactions from 2.7 million patients' allergy data stored in PEAR. Using the NLP tool developed as part of Aim 2 (see below, we processed both structured and free-text reaction entries and mapped them to SNOMED-CT. We calculated the frequencies of reaction concepts, including rare, severe, and hypersensitivity reactions. We then compared PEAR concepts to a Federal Health Information Modeling and Standards value set and University of Nebraska Medical Center data and created an integrated value set.

As part of our goal of constructing a comprehensive knowledge base of allergy information, we also conducted several epidemiologic studies in which we described the prevalence of drug, food, environmental, and other allergies along with other adverse reactions, such as hypersensitivities and intolerances. We also conducted epidemiologic studies of a number of specific allergens (e.g., opioids, nonsteroidal anti-inflammatory drugs) and reactions (e.g., anaphylaxis, severe cutaneous adverse reactions).

Aim 2

We developed an Allergy Information Extraction and Encoding module and integrated it with our previously developed NLP tool, MTERMS (Medical Text Extraction, Reasoning and Mapping System). The module consists of a lexicon, an information model, a named entity recognizer, and a semantic tagger applying both rule-based and machine learning approaches, and it can be used to identify and encode food, drug, and environmental allergies and allergic reactions. The module included updates to our lexicon using standard terminologies and novel disambiguation algorithms. We developed an annotation schema and annotated 400 emergency department notes that served as a gold standard in order to evaluate the module's accuracy. We successfully implemented our allergy module, as well as preliminary versions of MTERMS' medication and problem modules, with Epic's NoteReader, a module within Epic that provides a user interface for reconciling allergy, medication, and problem information.

Lexicon lookup and regular expressions are the most commonly used NLP techniques for named entity recognition (NER). Although allergy information appearing in the allergy section of a clinical note is relatively simple to identify, identifying this information in other sections (e.g., history of present illness, hospital course) is far more difficult, in part because, in these sections, a named entity can have multiple meanings. For example, a mention of a drug name may be in reference to that drug as an allergen, but it could also be referring to a medication used to treat a clinical condition, a medication that a patient refused to take, or something else. Similarly, a mention of a food may be in reference to a food to which the patient is allergic, but it could also be mentioned in the context of dietary or nutritional advice. Therefore, in order to resolve such ambiguity and assign the terms to the appropriate semantic classes, additional rule-based and machine learning methods are needed.

We first tried an approach based on hand-crafted rules that considered contextual information (e.g., section header, the existence of trigger words such as "allergic to"). For example, if a drug name was mentioned in the allergy section, we could infer that the drug belongs to the "allergen" semantic class. We also tried an assortment of machine learning-based approaches, including a Naive Bayes classifier, a Maximum Entropy classifier, and a Support Vector Machine (SVM).

Aim 3

The goal of this aim was to test whether the NLP module developed in Aim 2 could be used to improve the accuracy and completeness of allergy lists using data extracted from electronic clinical texts and, if so, whether it would help improve medication management and clinical decision support. To assess this, we collected a random sample of patient notes written in 2013 and 2014 from the Partners clinical research data repository. We then used our NLP module to process these notes to identify patients who have had any allergy information documented in their notes. Of the positively identified notes, we randomly selected a subsample of 200 patients for evaluation by domain experts to validate the output of the NLP module.

The next step was to obtain the structured allergy lists for these patients and compare them to the allergies identified in free-text by the NLP module. We assessed how many allergy records in the structured allergy lists are inaccurate or missing information based on the data found in the clinical notes via NLP. We also reviewed the patients' medication lists to check if any medication for the allergy was ordered before or after that allergen had been documented in the notes and to assess how many drug-allergy alerts had been triggered using the additional allergy information found via NLP.

Results

Aim 1

We created a comprehensive knowledge base and an information model for representing allergy information. We created lexicons for food, drug, environmental, and contrast media allergens, and we compiled a comprehensive lexicon of allergic reaction terms composed of 92% of all reactions stored in PEAR, sorted by frequency. We also developed a new food allergen grouping schema that improved upon previous coding systems.

Aim 2

MTERMS achieved an overall F-measure of 96.7% for the detection of allergen names and no known allergies. However, identifying allergy information outside the allergy section remains difficult. On a separate set of 70 ED notes that were likely to have allergy information but no allergy section, the F-measure remained low, at 57.1%. Therefore, we obtained a set of 217 ED notes (including the 70 mentioned above) that contain the words "allergy" or "allergic" but do not have an allergy section. We used MTERMS to extract all possible allergens from these notes —all instances of drugs, foods, environmental allergens, etc. Of the 1,259 substances found, 211 were allergens. We trained several classifiers (including a Naive Bayes classifier, a Maximum Entropy classifier, and a Support Vector Machine [SVM]) on this data. Features included a "bag of words," additional word features, the type of substance (food, environmental, blood, or other), and whether an allergy indicator or reaction was nearby. We evaluated the performance of each model using 10-fold cross-validation. Our performing model was an SVM, which achieved an F-measure of 74.0%.

Aim 3

We developed a standalone web application to display allergy information based on output from the Allergy Information Extraction and Encoding NLP module. The application can batch process a large number of notes, display allergies extracted from notes, and compare the extracted allergies with those listed in the patient's structured allergy list.

In the following, we describe some of our study's major findings. We provide a list of publications from this study at the end of the report.

NLP, Machine Learning and Value Sets

1. Extraction and Encoding of Allergy Information from Clinical Notes
   We conducted a study to explore the feasibility of extracting and encoding allergy information from emergency department clinical notes using MTERMS, our NLP system. We developed an annotation schema and annotated 400 ED notes that served as a gold standard for comparison to MTERMS output. MTERMS achieved an F-measure of 87.6% for the detection of allergen names and no known allergies, 90% for identifying true reactions in each allergy statement in which true allergens were also identified, and 69% for linking reactions to their allergen. These preliminary results demonstrate the feasibility using NLP to extract and encode allergy information from clinical notes.
2. Comparison and Evaluation of Existing Standards for Adverse Sensitivity Information Models
   Despite the existence of several standards for adverse sensitivity (e.g., allergy and intolerance) information models, many clinicians record information as free-text. To understand how the data interoperability and similarities between current adversity sensitivity information models, we designed a study that aimed to 1) identify and compare the existing common adverse sensitivity information models and 2) evaluate the coverage of the adverse sensitivity information models for representing allergy information on a subset of inpatient and outpatient adverse sensitivity clinical notes. We compared four common adverse sensitivity information models---Health Level 7 Allergy and Intolerance Domain Analysis Model, HL7-DAM; the Fast Healthcare Interoperability Resources, FHIR; the Consolidated Continuity of Care Document, C-CDA; and OpenEHR---and evaluated their coverage on a corpus of inpatient and outpatient notes (n = 120). We found that allergy specialists' notes had the highest frequency of adverse sensitivity attributes per note, whereas emergency department notes had the fewest attributes. Overall, the models had many similarities in the central attributes, which covered between 75% and 95% of adverse sensitivity information contained within the notes. However, representations of some attributes (especially the value sets) were not well aligned between the models, which is likely to present an obstacle for achieving data interoperability. Also, adverse sensitivity exceptions were not well represented among the information models. Although we found that common adverse sensitivity models cover a significant portion of relevant information in the clinical notes, our results highlight areas that need to be reconciled between the standards for data interoperability; this is important as adverse sensitivity information is a critical component of any EHR system.
3. Food Adverse Sensitivity Concepts in the EHR
   We used MTERMS to examine, encode, and group foods that cause any adverse sensitivity in a large allergy repository using natural language processing and standard terminologies. Using the Medical Text Extraction, Reasoning, and Mapping System (MTERMS), we processed both structured and free-text entries stored in an enterprise-wide allergy repository (Partners' Enterprisewide Allergy Repository), normalized diverse food allergen terms into concepts, and encoded these concepts using the Systematized Nomenclature of Medicine – Clinical Terms (SNOMED-CT) and Unique

Ingredient Identifiers (UNII) terminologies. Concept coverage also was assessed for these two terminologies. We further categorized allergen concepts into groups and calculated the frequencies of these concepts by group. Finally, we conducted an external validation of MTERMS's performance when identifying food allergen terms, using a randomized sample from a different institution. We identified 158,552 food allergen records (2140 unique terms) in the Partners repository, corresponding to 672 food allergen concepts. High-frequency groups included shellfish (19.3%), fruits or vegetables (18.4%), dairy (9.0%), peanuts (8.5%), tree nuts (8.5%), eggs (6.0%), grains (5.1%), and additives (4.7%). Ambiguous, generic concepts such as "nuts" and "seafood" accounted for 8.8% of the records. SNOMED-CT covered more concepts than UNII in terms of exact (81.7% vs 68.0%) and partial (14.3% vs 9.7%) matches. Adverse sensitivities to food are diverse, and existing standard terminologies have gaps in their coverage of the breadth of allergy concepts. From this, we concluded that new strategies are needed to represent and standardize food adverse sensitivity concepts, to improve documentation in EHRs.

4. Reaction Value Set

   As part of this study, we developed comprehensive value set for documenting and encoding adverse reactions in the allergy module of an electronic health record. To do this, we analyzed 2,471,004 adverse reactions stored in Partners Healthcare's Enterprise-wide Allergy Repository (PEAR) of 2.7 million patients. Using MTERMS, we processed both structured and free-text reaction entries and mapped them to Systematized Nomenclature of Medicine-Clinical Terms. We calculated the frequencies of reaction concepts, including rare, severe, and hypersensitivity reactions. We compared PEAR concepts to a Federal Health Information Modeling and Standards value set and University of Nebraska Medical Center data, and then we created an integrated value set. We identified 787 reaction concepts in PEAR. Frequently reported reactions included rash (14.0%), hives (8.2%), gastrointestinal irritation (5.5%), itching (3.2%), and anaphylaxis (2.5%). We identified an additional 320 concepts from Federal Health Information Modeling and Standards and the University of Nebraska Medical Center to resolve gaps due to missing and partial matches when comparing these external resources to PEAR. This yielded 1106 concepts in our final integrated value set. The presence of rare, severe, and hypersensitivity reactions was limited in both external datasets. Hypersensitivity reactions represented roughly 20% of the reactions within our data. We developed a value set for encoding adverse reactions using a large dataset from one health system, enriched by reactions from two large external resources. This integrated value set includes clinically important severe and hypersensitivity reactions. This work contributes a value set, harmonized with existing data, to improve the consistency and accuracy of reaction documentation in electronic health records, providing the necessary building blocks for more intelligent clinical decision support for allergies and adverse reactions.

5. Viability of Mutual Information Clustering for Discovering Food Cross-Reactions

   Mutual information clustering is an agglomerative hierarchical clustering method that has been used to group random variables or sets thereof. Some researchers have found that the normalization method used can lead to oddly sized clusters that do not line up with expected results. We introduce a new normalization parameter to control the size of the clusters and apply it to food allergy data from a large allergy repository from an electronic health record, treating the distributions of food allergies in our population as random variables. Our method was able to identify previously known food cross-reaction

groups (with an adjusted Rand index of 0.971, outperforming alternative clustering algorithms) in addition to proposing possible new groups. Our results demonstrate the viability of mutual information clustering as an approach for discovering possible food cross-reactions.

6. Narrative Review of NLP Fundamental and Application Medication Safety

   We also conducted a narrative review that illustrated the fundamentals of NLP and discussed NLP's application to medication safety in four data sources: electronic health records, internet-based data, published literature, and reporting systems. We found that the safety of medication use has been a priority in the United States since the late 1930s. Recently, it has gained prominence due to the increasing amount of data suggesting that a large amount of patient harm is preventable and can be mitigated with effective risk strategies that have not been sufficiently adopted. Adverse events from medications are part of clinical practice, but the ability to identify a patient's risk and to minimize that risk must be a priority. The ability to identify adverse events has been a challenge due to limitations of available data sources, which are often free-text. The use of natural language processing (NLP) may help address these limitations. NLP is the artificial intelligence domain of computer science that uses computers to manipulate unstructured data (i.e., narrative text or speech data) in the context of a specific task. The NLP application to medication safety in electronic health records, internet-based data, published literature, and reporting systems is a growing area is the use of computer algorithms to help automatically detect associations between medications and adverse effects. The main benefit of NLP is in the time savings associated with automation of various medication safety tasks, such as the medication reconciliation process facilitated by computers, as well as the potential for near-real-time identification of adverse events for postmarketing surveillance, such as those posted on social media that would otherwise go unanalyzed. NLP is limited by a lack of data sharing between healthcare organizations due to insufficient interoperability capabilities, inhibiting large-scale adverse event monitoring across populations. We anticipate that future work in this area will focus on the integration of data sources from different domains to improve the ability to identify potential adverse events more quickly and to improve clinical decision support with regard to a patient's estimated risk for specific adverse events at the time of medication prescription or review.

Epidemiological Studies

Using MTERMS, we were able to describe the epidemiology of drugs allergies, food allergies, and allergic reactions using longitudinal electronic health record data. The following studies were published and distributed.

1. Prevalence of Drug Allergies

   The prevalence of drug allergies documented in electronic health records (EHRs) of large patient populations is understudied. We aimed to describe the prevalence of common drug allergies and patient characteristics documented in EHRs of a large healthcare network over the past two decades. Drug allergy data were obtained from EHRs of patients who visited two large tertiary care hospitals in Boston from 1990 to

2013. The prevalence of each drug and drug class was calculated and compared by sex and race/ethnicity. The number of allergies per patient was calculated and the frequency of patients having 1, 2, 3…, or 10+ drug allergies was reported. We also conducted a trend analysis by comparing the proportion of each allergy to the total number of drug allergies over time. Among 1,766,328 patients, 35.5% of patients had at least one reported drug allergy, with an average of 1.95 drug allergies per patient. The most commonly reported drug allergies in this population were to penicillins (12.8%), sulfonamide antibiotics (7.4%), opiates (6.8%), and nonsteroidal anti-inflammatory drugs (NSAIDs) (3.5%). The relative proportion of allergies to angiotensin-converting enzyme (ACE) inhibitors and HMG CoA reductase inhibitors (statins) have more than doubled since early 2000s. Drug allergies were most prevalent among women and White patients---except for NSAIDs, ACE inhibitors, and thiazide diuretics, which were more prevalent in Black patients. Women and White patients may be more likely to experience a reaction from common medications. An increase in reported allergies to ACE inhibitors and statins is noteworthy.

2. Prevalence of Food Allergies

   Food allergy prevalence is reported to be increasing, but epidemiological data using patients' electronic health records (EHRs) remain sparse. We sought to determine the prevalence of food allergy and intolerance documented in the EHR allergy module. Using allergy data from a large healthcare organization's EHR between 2000 and 2013, we determined the prevalence of food allergy and intolerance by sex, racial/ethnic group, and allergen group. We examined the prevalence of reactions that were potentially IgE mediated and anaphylactic. Data were validated using radioallergosorbent test and ImmunoCAP results, when available, for patients with reported peanut allergy. Among 2.7 million patients, we identified 97,482 patients (3.6%) with one or more food allergies or intolerances (mean, 1.4 ± 0.1). The prevalence of food allergy and intolerance was higher in women (4.2% vs 2.9%; $P < .001$) and Asians (4.3% vs 3.6%; $P < .001$). The most common food allergen groups were shellfish (0.9%), fruit or vegetable (0.7%), dairy (0.5%), and peanut (0.5%). Of the 103,659 identified reactions to foods, 48.1% were potentially IgE mediated (affecting 50.8% of food allergy or intolerance patients), and 15.9% were anaphylactic. About 20% of patients with reported peanut allergy had a radioallergosorbent test/ImmunoCAP performed, of which 57.3% had an IgE level of grade 3 or higher. Our findings are consistent with previously validated methods for studying food allergy, suggesting that the EHR's allergy module has the potential to be used for clinical and epidemiological research. The spectrum of severity observed with food allergy highlights the critical need for more allergy evaluations.

3. Multiple Drug Intolerance and Multiple Drug Allergy Syndrome Epidemiology

   The epidemiology of multiple drug intolerance syndrome (MDIS) and multiple drug allergy syndrome (MDAS) is poorly characterized. We used electronic health record (EHR) data to describe prevalences of MDIS and MDAS and to examine associations with anxiety and depression. Patients with ≥3 outpatient encounters at Partners HealthCare System from 2008 to 2015 were included. Patients with MDIS had intolerances to ≥3 drug classes, and patients with MDAS had hypersensitivities to ≥2 drug classes. Psychiatric conditions and comorbidities were defined from the EHR and used in multivariable logistic regression models to assess the relation between anxiety/depression and MDIS/MDAS. Of 746,888 patients, 47,634 (6.4%) had MDIS and

8615 (1.2%) had MDAS; 3171 (0.4%) had both. Anxiety (adjusted odds ratio [aOR] 1.72 [1.65, 1.80]), depression (aOR 1.46 [1.41, 1.52]), and both anxiety and depression (aOR 1.97 [1.86, 2.08]) were associated with increased odds of MDIS. Depression was associated with increased odds of MDAS (aOR 1.41 [1.28, 1.56]), but there were no clear associations with anxiety (aOR 1.13 [0.99, 1.30]) or both depression and anxiety (aOR 1.13 [0.92, 1.38]). Although 6% of patients had MDIS, only 1% had MDAS. MDIS was associated with both anxiety and depression; patients with both anxiety and depression had an almost twofold increased odds of MDIS. MDAS was associated with a 40% increased odds of depression, but there was no significant association with anxiety. Psychological assessments may be useful in the evaluation and treatment of patients with MDIS and MDAS; physiologic causes for MDAS warrant further investigation.

4. Drug-Induced Anaphylaxis

   Although drugs represent a common cause of anaphylaxis, few large studies of drug-induced anaphylaxis have been performed. To describe the epidemiology and validity of reported drug-induced anaphylaxis in the electronic health records (EHRs) of a large United States healthcare system. Using EHR drug allergy data from 1995 to 2013, we determined the population prevalence of anaphylaxis, including anaphylaxis prevalence over time, and the most commonly implicated drugs/drug classes reported to cause anaphylaxis. Patient risk factors for drug-induced anaphylaxis were assessed using a logistic regression model. Serum tryptase and allergist visits were used to assess the validity and follow-up of EHR-reported anaphylaxis. Among 1,756,481 patients, 19,836 (1.1%) reported drug-induced anaphylaxis; penicillins (45.9 per 10,000), sulfonamide antibiotics (15.1 per 10,000), and nonsteroidal anti-inflammatory drugs (NSAIDs) (13.0 per 10,000) were most commonly implicated. Patients who were White (odds ratio [OR] 2.38, 95% CI 2.27-2.49), were women (OR 2.20, 95% CI 2.13-2.28), had systemic mastocytosis (OR 4.60, 95% CI 2.66-7.94), had Sjögren's syndrome (OR 1.94, 95% CI 1.47-2.56), and had asthma (OR 1.50, 95% CI 1.43-1.59) had an increased odds of drug-induced anaphylaxis. Serum tryptase was performed in 135 (<1%) anaphylaxis cases, and 1587 patients (8.0%) saw an allergist for follow-up. EHR-reported anaphylaxis occurred in approximately 1% of patients, most commonly from penicillins, sulfonamide antibiotics, and NSAIDs. Women, White patients, and patients with mastocytosis, Sjögren's syndrome, or asthma had increased odds of reporting drug-induced anaphylaxis. The low observed frequency of tryptase testing and specialist evaluation emphasize the importance of educating providers on anaphylaxis management.

5. Hypersensitivity Reactions

   Hypersensitivity reactions (HSRs) are immunologic responses to drugs. Identification of HSRs documented in the electronic health record (EHR) is important for patient safety. We examined HSR epidemiology using longitudinal EHR data from a large United States healthcare system. Patient demographic information and drug allergy data were obtained from the Partners Enterprisewide Allergy Repository for two large tertiary care hospitals from 2000 to 2013. Drug-induced HSRs were categorized into immediate and delayed HSRs based on typical phenotypes. Causative drugs and drug groups were assessed. The prevalence of HSRs was determined, and sex and racial differences were analyzed. Among 2.7 million patients, 377,474 (13.8%) reported drug-induced HSRs, of whom 70.3% were women and 77.5% were White. In total, 580,456 HSRs were reported, of which 53.1% were immediate reaction phenotypes. Common immediate HSRs included

hives (48.8%), itching (15.0%), and angioedema (14.1%). Delayed HSR phenotypes (46.9%) were largely rash (99.0%). Penicillins were associated with the most immediate (33.0%) and delayed (39.0%) HSRs. Although most HSRs were more prevalent in women and White patients, notable differences were identified for certain rare HSRs including acute interstitial nephritis, which appeared more commonly in men (0.02% vs 0.01%, P < .001). Asian patients had more fixed drug eruptions (0.007% vs 0.002%, P = .021) and severe cutaneous adverse reactions (0.05% vs 0.04%, P < .001). Drug HSRs were reported in 13.8% of patients. Almost one half of reported immediate HSR phenotypes were hives, and almost all reported delayed HSR phenotypes were rash. HSRs largely affected women and White patients, but differences were identified for specific rare HSRs.

6. Contrast Allergen Documentation in the EHR

   We systematically identified medical imaging drugs and class terms in an integrated EHR allergy repository for patients seen at a large healthcare system between 2000 and 2013. Structured and free-text contrast allergy records were normalized and categorized by inciting agent and nature of adverse reaction. Allergen records were evaluated by their level of specificity. Reaction records were evaluated by whether the reaction was known or unknown and by whether known reactions would be categorized as allergic-like or physiologic. Among 2.7 million patients, we identified 36,144 patients (1.3%) with at least one of 40,669 contrast allergy records associated with 49,000 reactions. Contrast allergens were more likely than other allergens to be entered as free-text (15.2% vs. 6.3%; OR 2.69, 95% CI 2.61-2.76). There were 1305 unique contrast allergen records, which we grouped into 141 concepts. Most contrast allergen records were ambiguous contrast concepts (69.1%) rather than imaging modality-specific class terms (19.4%) or specific contrast agents (11.5%). Contrast reactions were occasionally entered as free-text (24.8%), which together with structured entries were grouped into 183 concepts. A known reaction was documented in 71.8% of cases; however, 12.2% were not allergic-like reactions. Contrast allergy records in EHRs are diverse and commonly low quality. Continued EHR enhancements and training are needed to support contrast allergy documentation to facilitate improved patient care and medical research.

Drug-Allergy Interaction Alerts

Additionally, we also examined drug-allergy interaction alerts, and our results are described below.

1. Drug-Allergy Alerts for Opioids

   This study examined trends in drug-allergy interaction (DAI) alert overrides for opioid medications - the most commonly triggered alerts in the computerized provider order entry (CPOE). We conducted an observational analysis of the DAI opioid alerts triggered over the past decade (2004-2013, n=342,338) in two large academic hospitals in Boston (United States). We found an increasing rate of DAI alert overrides, culminating in 89.7% in 2013. Allergic reactions included a high proportion (38.2%) of non-immune-mediated opioid reactions (e.g., gastrointestinal upset). The DAI alert override rate was high for immune-mediated (88.6%) and life-threatening reactions (87.8%). Exact allergy-medication matches were overridden less frequently (about 70%) compared with non-exact

matches within allergy groups (over 90%). About one third of the alert override reasons pointed to irrelevant alerts (i.e., "Patient has tolerated the medication before"), and 44.9% were unknown. Those findings warrant further investigation into providers' reasons for high override rates. User interfaces should evolve to enable less interruptive and more accurate alerts to decrease alert fatigue.

2. Drug Allergy Alerts and Provider Alert Fatigue

   There have been growing concerns about the impact of drug allergy alerts on patient safety and provider alert fatigue. The authors aimed to explore the common drug allergy alerts over the past 10 years and the reasons why providers tend to override these alerts. This was a retrospective observational cross-sectional study (2004-2013). Drug allergy alert data (n = 611,192) were collected from two large academic hospitals in Boston, MA (USA). Overall, the authors found an increase in the rate of drug allergy alert overrides, from 83.3% in 2004 to 87.6% in 2013 (P < .001). Alarmingly, alerts for immune-mediated and life-threatening reactions with definite allergen and prescribed medication matches were overridden 72.8% and 74.1% of the time, respectively. However, providers were less likely to override these alerts compared with possible (cross-sensitivity) or probable (allergen group) matches (P < .001). The most common drug allergy alerts were triggered by allergies to narcotics (48%) and other analgesics (6%), antibiotics (10%), and statins (2%). Only slightly more than one third of the reactions (34.2%) were potentially immune mediated. Finally, more than half of the overrides' reasons pointed to irrelevant alerts (i.e., patient has tolerated the medication before, 50.9%), and providers were significantly more likely to override repeated alerts (89.7%) rather than first-time alerts (77.4%, P < .001). These findings underline the urgent need for more efforts to provide more accurate and relevant drug allergy alerts to help reduce alert override rates and improve alert fatigue.

Other Findings

1. Comparing ADRs in EHR Data with Social Media Data

   Large databases of clinician-reported (e.g., allergy repositories) and patient-reported (e.g., social media) adverse drug reactions (ADRs) exist; however, whether patients and clinicians report the same concerns is not clear. Our objective was to compare electronic health record data and social media data to better understand differences and similarities between clinician-reported ADRs and patients' concerns regarding aspirin and atorvastatin. This pilot study explored a large repository of electronic health record data and social media data for clinician-reported ADRs and patients concerns for two common medications: aspirin (n = 31,817 ADRs accessible in clinical data; n = 19,186 potential ADRs accessible in social media data) and atorvastatin (n = 15,047 ADRs accessible in clinical data; n = 23,408 potential ADRs accessible in social media data). We found that the most frequently reported ADRs matched the most frequent patients' concerns. However, several less frequently reported reactions were more prevalent on social media (i.e., aspirin-induced hypoglycemia was discussed only on social media). Overall, we found a relatively strong positive and statistically significant correlation between the frequency ranking of reactions and patients' concerns for atorvastatin (Pearson's r = 0.61, p < 0.001) but not for aspirin (Pearson's r = 0.1, p = 0.69). Future studies should develop

additional natural language methods for a more detailed data analysis (i.e., identifying causality and temporal aspects in the social media data).

Publications and Products

1. Goss FR, et al. An evaluation of a natural language processing tool for identifying and encoding allergy information in emergency department clinical notes. AMIA 2014 Annual Symposium; 2014 Nov 15-19, Washington, DC.
2. Blumenthal KG, Wickner PG, Lau JJ, et al. Stevens-Johnson syndrome and toxic epidermal necrolysis: a cross-sectional analysis of patients in an integrated allergy repository of a large health care system. J Allergy Clin Immunol Pract. 2015 Mar-Apr;3(2):277-280.e1. PMID: 25609329.
3. Lai KH, Topaz M, Goss FR, et al. Automated misspelling detection and correction in clinical free-text records. J Biomed Inform. 2015 Jun;55:188-95. PMID: 25917057
4. Zhou L, Baughman AW, Lei VJ, et al. Identifying Patients with Depression Using Free-text Clinical Documents. Stud Health Technol Inform. 2015;216:629-33. PMID: 26262127.
5. Topaz M, et al. High override rate for opioid drug-allergy interaction alerts: current trends and recommendations for future. 15th World Congress on Health and Biomedical Informatics (MEDINFO 2015); 2015 Aug 19-23, São Paulo, Brazil.
6. Zhou L, et al. Rising drug allergy alert overrides in a computerized provider order entry system: a decade of experience. AMIA 2015 Annual Symposium; 2015 Nov 14-18, San Francisco, CA.
7. Topaz M, Seger DL, Goss F, et al. Standard information models for representing adverse sensitivity information in clinical documents. Methods Inf Med. 2016;55(2):151-7. doi: 10.3414/ME15-01-0081. PMID: 26905461
8. Topaz M, Lai K, Dhopeshwarkar N, Seger DL, et al. Clinicians' reports in electronic health records versus patients' concerns in social media: a pilot study of adverse drug reactions of aspirin and atorvastatin. Drug Saf. 2016 Mar;39(3):241-50. PMID: 26715498.
9. Blumenthal KG, et al. Reported incidence of hypersensitivity reactions to non-steroidal anti-inflammatory drugs in the electronic health record. AAAAI 2016 Annual Meeting; 2016 Mar 4-7, Los Angeles, CA.
10. Banerji A, et al. Epidemiology and incidence of ACE inhibitor angioedema utilizing a large electronic health record. AAAAI 2016 Annual Meeting; 2016 Mar 4-7, Los Angeles, CA.
11. Plasek JM, Goss FR, Lai KH, et al. Food entries in a large allergy data repository. J Am Med Inform Assoc. 2016 Apr;23(e1):e79-87. PMID: 26384406.
12. Topaz M, Seger DL, Slight SP, et al. Rising drug allergy alert overrides in electronic health records: an observational retrospective study of a decade of experience. J Am Med Inform Assoc. 2016 May;23(3):601-8. PMID: 26578227.
13. Zhou L, Dhopeshwarkar N, Blumenthal KG, et al. Drug allergies documented in electronic health records of a large healthcare system. Allergy. 2016 Sep;71(9):1305-13. PMID: 26970431.
14. Topaz M, et al. Expert recommendations on redesigning drug allergy alerts in electronic health record systems. AMIA 2016 Annual Symposium; 2016 Nov 12-16, Chicago, IL.

15. Topaz M, Goss F, Blumenthal K, Lai K, et al. Towards improved drug allergy alerts: multidisciplinary expert recommendations. Int J Med Inform. 2017 Jan;97:353-355. PMID: 27729200.
16. Blumenthal KG, Lai KH, Huang M, et al. Adverse and Hypersensitivity Reactions to Prescription Nonsteroidal Anti-Inflammatory Agents in a Large Health Care System. J Allergy Clin Immunol Pract. 2017 Jan 18. pii: S2213-2198(16)30669-9. PMID: 28110055.
17. Blumenthal KG, Acker WW, Li Y, Holtzman NS, Zhou L. Allergy entry and deletion in the electronic health record. Ann Allergy Asthma Immunol. 2017 Mar;118(3):380-381. doi: 10.1016/j.anai.2016.12.020. PMID: 28126435.
18. Banerji A, Blumenthal KG, Lai KH, et al. Epidemiology of ACE Inhibitor Angioedema Utilizing a Large Electronic Health Record. J Allergy Clin Immunol Pract. 2017 Apr 1. pii: S2213-2198(17)30163-0. PMID: 28377081.
19. Acker WW, Plasek JM, Blumenthal KG, et al. Prevalence of Food Allergies and Intolerances Documented in Electronic Health Records. J Allergy Clin Immunol. 2017 May 23. pii: S0091-6749(17)30672-3. PMID: 28577971.
20. Lai K, Blackley S, Zhou L. Using Mutual Information Clustering to Discover Food Allergen Cross-Reactivity. 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). 2017 Nov 13-6, Kansas City, MO.
21. Goss FR, Lai KH, Topaz M, et al. A value set for documenting adverse reactions in electronic health records. J Am Med Inform Assoc. 2018 Jun 1;25(6):661-669. PMID: 29253169.
22. Blumenthal KG, Li Y, Acker WW, et al. Multiple Drug Intolerance Syndrome and Multiple Drug Allergy Syndrome: Epidemiology and Associations with Anxiety and Depression. Allergy. 2018 Oct;73(10):2012-2023. PMID: 29574787.
23. Wong A, Plasek JM, Montecalvo SP, et al. Natural language processing and its implications for the future of medication safety: a narrative review of recent advances and challenges. Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy. 2018 Aug;38(8):822-841. PMID: 29884988.
24. Dhopeshwarkar N, Sheikh A, Doan R, et al. Drug-Induced Anaphylaxis Documented in Electronic Health Records. J Allergy Clin Immunol Pract. 2018;7(1):103-111. PMID: 29969686.
25. Saff RR, Li Y, Santhanakrishnan N, et al. Identification of Inpatient Allergic Drug Reactions Using ICD-9-CM Codes. J Allergy Clin Immunol Pract. 2018 Jul 31. pii: S2213-2198(18)30457-4. PMID: 30075337.
26. Wolfson AR, Zhou L, Li Y, et al. Drug Rash Eosinophilia and Systemic Symptoms (DRESS) Syndrome Identified in the Electronic Health Record Allergy Module. J Allergy Clin Immunol Pract.2018 Aug 31. pii: S2213-2198(18)30548-8. PMID: 30176295.
27. Wong A, Seger D, Goss F, et al. Drug Hypersensitivity Reactions Documented in Electronic Health Records within a Large Health System. J Allergy Clin Immunol Pract. 2018 Dec 1. pii: S2213-2198(18)30751-7. PMID: 30513361.
28. Deng F, Li MD, Wong A, et al. Quality of documentation of contrast agent allergies in electronic health records. Journal of the American College of Radiology. 2019. (accepted)