Why is FDA Interested in Natural Language Processing (NLP) of Clinical Texts?

Applications to Pharmacovigilance and Pharmacoepidemiology

Robert Ball, MD, MPH, ScM

Deputy Director

Office of Surveillance and Epidemiology

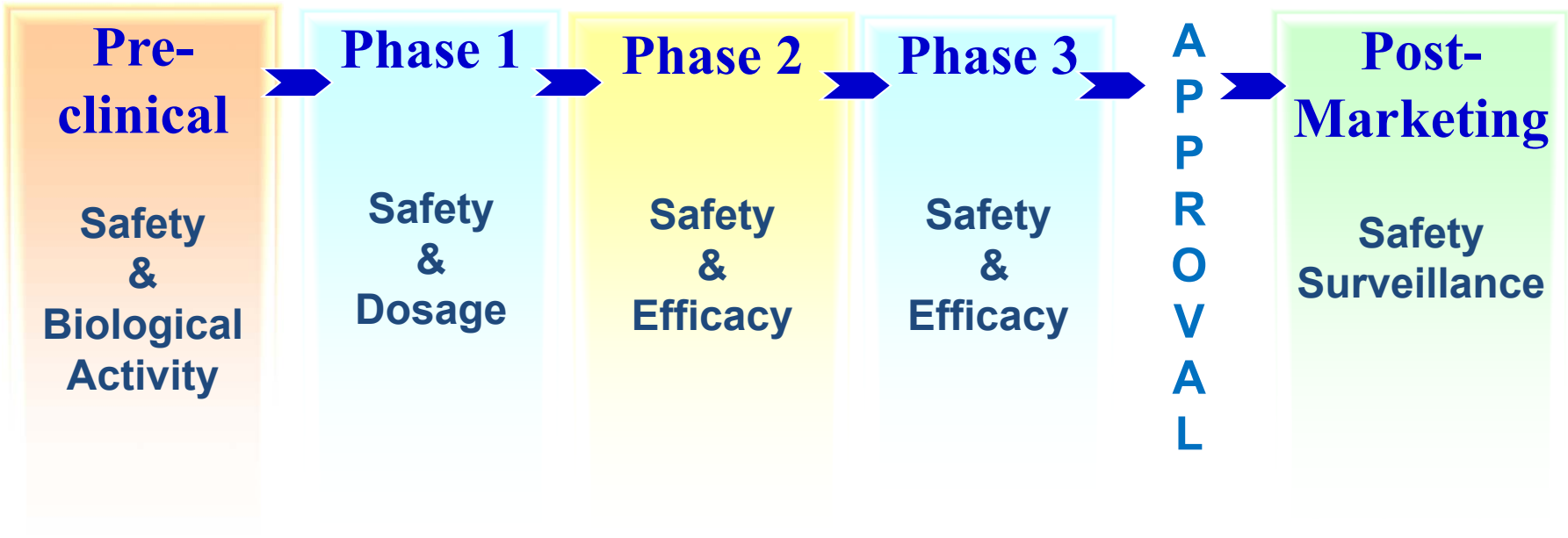
Center of Drug Evaluation and Research

June 15, 2017

FDA's Interest in NLP of Clinical Texts

- Improve the efficiency and scientific validity of FDA analyses
  - e.g. application to narratives of individual case safety reports for pharmacovigilance (including IND safety reports, post-marketing safety reports, literature reports)
- Be prepared for regulatory submissions that might use NLP
  - e.g. automated population of case report forms (CRFs) in clinical trials
- Understand how this technology can support the Evidence Generation and the Learning Healthcare System visions
  - e.g. Identification of key information from clinical narratives in EHRs for case classification in the Sentinel System

Opportunity for NLP of Clinical Text in the Lifecycle of FDA-regulated Products



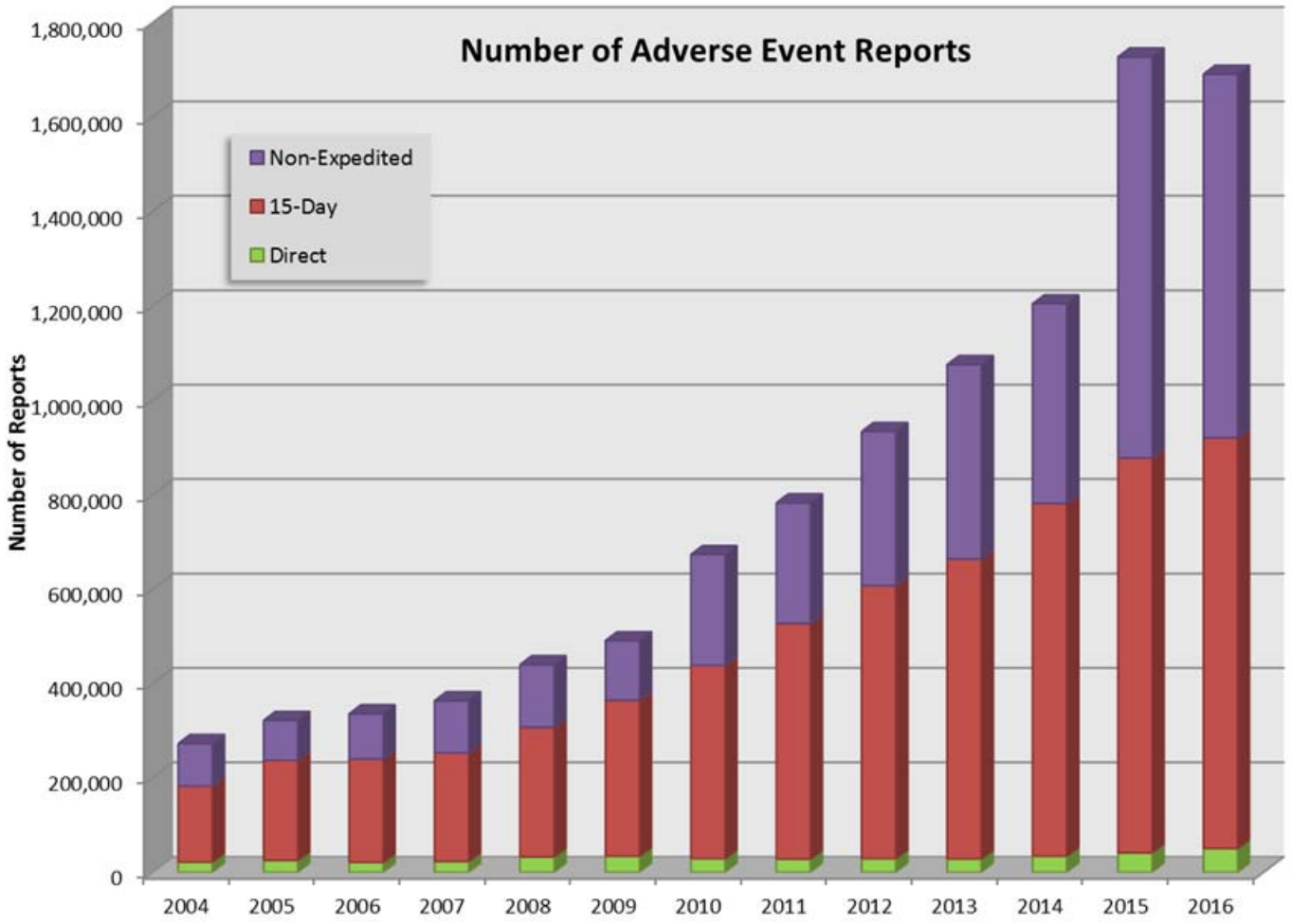
Selected Contexts in which Clinical Texts Arise

- IND safety reports
- NDA/BLA submissions
  - Trials
    - Patient identification/enrollment
    - Data collection (e.g. auto populate CRF's)
- Labels
  - Is an adverse event "labeled"?
- Post-market Individual Case Safety Reports of Adverse Events
  - Key information often in clinical narratives
- Pharmacoepidemiological studies
  - Sentinel System
- Social Media/Internet query logs

Some Application Areas for NLP of Clinical Texts

- Adverse Events (IND Safety Reports and Post-market ICSR's)
  - Exposure, outcome, temporal association, alternative explanations (e.g. other exposures, past medical history)
- Outcomes
  - Code validation, complex data (e.g. laboratory, imaging, pathology, electrophysiology)
  - Case Definitions (aka computational phenotypes)
- Cohort Selection
  - Clinical trials
  - Epidemiological studies
- Confounder extraction (e.g. smoking history, BMI)
- Identifying drug use, abuse, misuse

FDA Adverse Event Reporting System (FAERS)



Introduction to Spontaneous Reporting Systems (SRS)

Reporter

1. State	2. County where administered	3. Date of birth	4. Patient age	5. Sex
MD	USA	mm dd yy		M

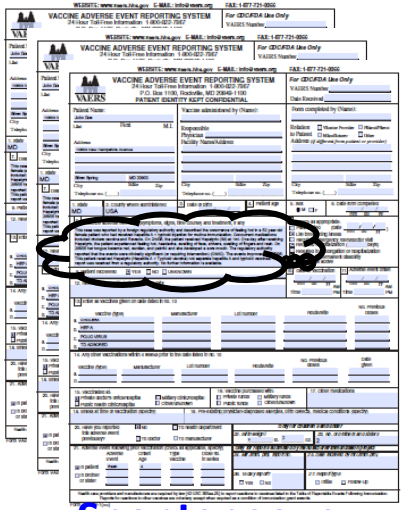
7. Describe adverse events(s) (symptoms, signs, time course) and treatment, if any

This case was reported by a foreign regulatory authority and described the occurrence of feeling hot in a 52 year old female patient who had received hepatitis A + typhoid injection for routine immunization. Concurrent medications included cholera vaccine and Revaxis. On 2/4/05, the patient received Hepatyrix (IM) at 1ml. One day after receiving Hepatyrix, the patient experienced feeling hot, headache, swelling of face, shivers, swelling of fingers and rash. On 2/6/05 her tongue became red, swollen, and painful and she developed a sore mouth. The regulatory authority reported that the events were clinically significant (or requiring intervention) (OMIC). The events improved on 2/7/05. This patient received Hepatyrix (Hepatitis A + Typhoid vaccine) not seperate hepatitis A and typhoid vaccines. This report was received from a regulatory authority. No further information is available.

8. Chec
Patie
Life th
Requ
Requ
Resul
Resul
None

9. Patient recovered ■ YES ☐ NO ☐ UNKNOWN

10. Date



Spontaneous Reports

(MedDRA) Codes Assignment

Spontaneous Reporting System

Case Review

Safety Use Cases & NLP

A. NARRATIVE of SAFETY REPORT

Patient received Smallpox vaccination on **4/21/2006** in left deltoid. **12 days after** vaccination he developed increased left arm pain and pleuritic substernal chest pain. **5/11/06** transferred to hospital with chest pain, right arm pain. Final dx of acute myopericarditis, serum reaction, allergic reaction, anemia, abnormal reaction to vaccine. Medical records from previous hospitalization obtained on **5/14/06** showed PMHx of Stevens-Johnson syndrome; family hx reveals patient's father had myocardial infarction.

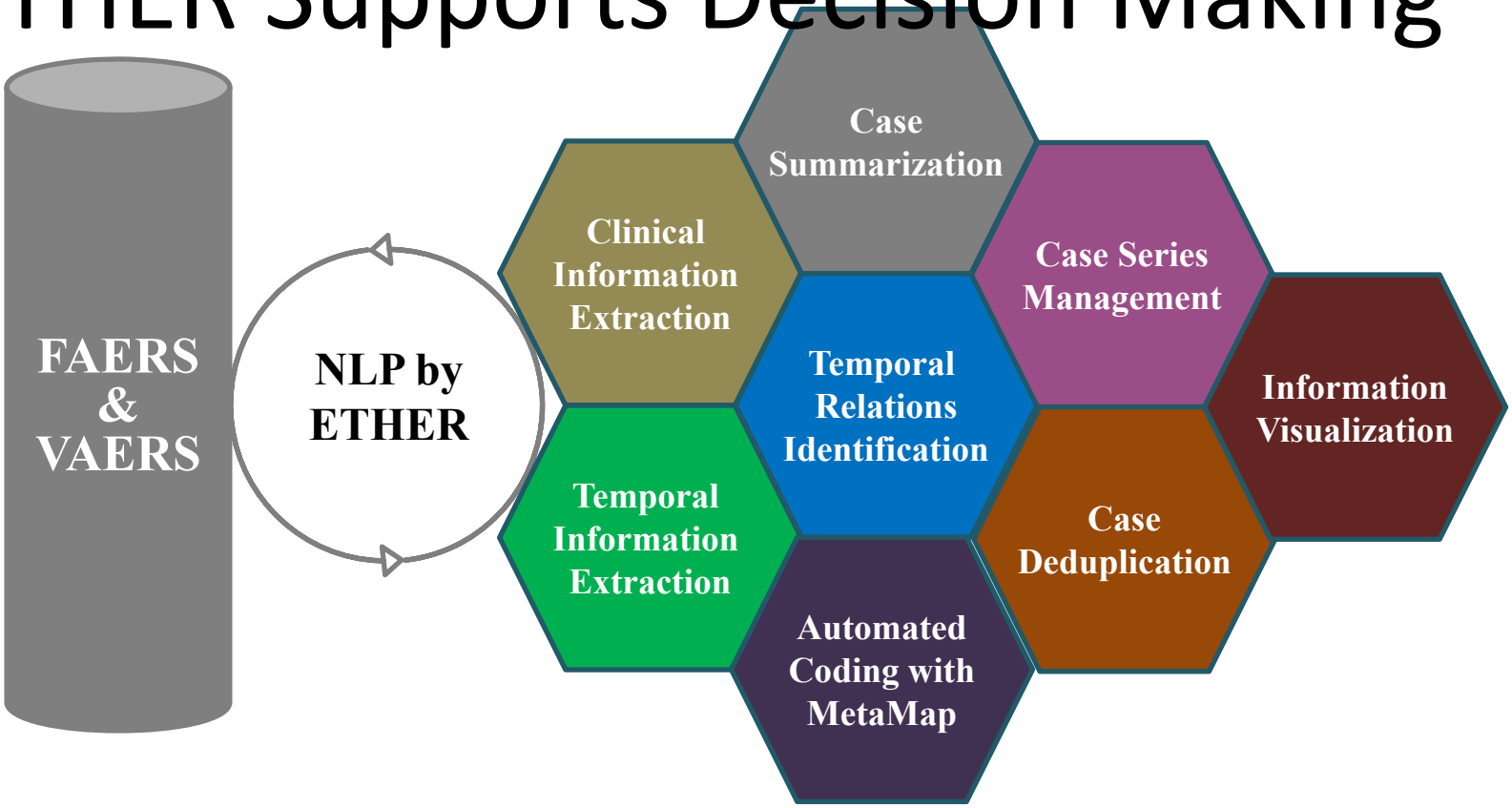
B. CLINICAL & TEMPORAL INFORMATION EXTRACTION with AUTOMATED CODING

CLINICAL INFORMATION (TYPE)	CLINICAL INFORMATION (TEXT)	DATE STAMP	MedDRA PT
VACCINE	Smallpox	**4/21/2006**	
SECONDARY DIAGNOSES	increased left arm pain	**5/3/2006**	Pain in extremity
	pleuritic substernal chest pain	**5/3/2006**	Chest pain
SYMPTOMS	chest pain	**5/11/06**	Chest pain
	right arm pain	**5/11/06**	Pain in extremity
PRIMARY DIAGNOSES	acute myopericarditis	**5/11/06**	Myocarditis
	serum reaction	**5/11/06**	Serum sickness-like reaction
	allergic reaction	**5/11/06**	Hypersensitivity
	anemia	**5/11/06**	Anaemia
	abnormal reaction to vaccine	**5/11/06**	Immunisation reaction
MEDICAL HISTORY	stevens johnson	**N/A**	Stevens-Johnson syndrome
FAMILY HISTORY	myocardial infarction	**N/A**	Myocardial infarction

C. REPORT SUMMARIZATION

A 32.0 years old female with medical history of stevens johnson and family history of myocardial infarction was treated with smallpox vaccination on 2006-04-21. The patient was diagnosed with increased left arm pain and pleuritic substernal chest pain on 2006-05-03; and acute myopericarditis, serum reaction, allergic reaction, anemia, abnormal reaction on 2006-05-11.

ETHER Supports Decision Making

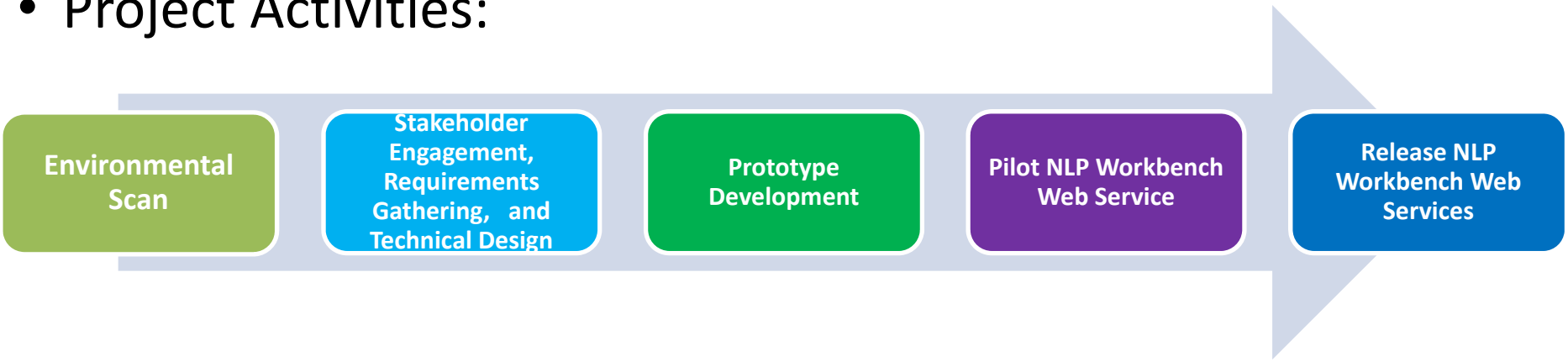


ETHER: **E**vent-based **T**ext-**M**ining of **H**ealth **E**lectronic **R**ecords
FAERS: **F**DA **A**dverse **E**vent **R**eporting **S**ystem
VAERS: **V**accine **A**dverse **E**vent **R**eporting **S**ystem
NLP: **N**atural **L**anguage **P**rocessing

Botsis et al, JBI 2016

Patient-Centered Outcomes Research Trust Fund (PCORTF) Collaborative Project – CDC and FDA

- Two Year Project (2016-2018)
- Project Goals:
  - Develop a Natural Language Processing (NLP) Workbench that utilizes Web Services for analyzing unstructured clinical information
  - Pilots for use in cancer registries and safety surveillance domains
  - Workbench will be used to develop NLP applications for other clinical domains
- Project Activities:



NLP and Machine Learning in FAERS

Journal of the American Medical Informatics Association, 0(0), 2017, 1–9
doi: 10.1093/jamia/ocx022
Research and Applications

Research and Applications

Development of an automated assessment tool for MedWatch reports in the FDA adverse event reporting system

Lichy Han,[1] Robert Ball,[2] Carol A Pamer,[2] Russ B Altman,[3,4] and Scott Proestel[2]

[1]Biomedical Informatics Training Program, Stanford University, Stanford, CA, USA, [2]Office of Surveillance and Epidemiology, Center for Drug Evaluation and Research, Food and Drug Administration, Silver Spring, MD, USA, [3]Department of Genetics, Stanford University and [4]Department of Bioengineering, Stanford University

Corresponding Author: Scott Proestel, Division of Epidemiology, Office of Biostatistics and Epidemiology, FDA Center for Biologics Evaluation and Research, White Oak Building 71, Room 1260, 10903 New Hampshire Avenue, Silver Spring, MD 20993, USA. Phone: (240) 402-0396. E-mail: Scott.Proestel@fda.hhs.gov

Received 1 September 2016; Revised 30 January 2017; Accepted 24 February 2017

ABSTRACT

Objective: As the US Food and Drug Administration (FDA) receives over a million adverse event reports associated with medication use every year, a system is needed to aid FDA safety evaluators in identifying reports most likely to demonstrate causal relationships to the suspect medications. We combined text mining with machine learning to construct and evaluate such a system to identify medication-related adverse event reports.

Methods: FDA safety evaluators assessed 326 reports for medication-related causality. We engineered features from these reports and constructed random forest, L1 regularized logistic regression, and support vector machine models. We evaluated model accuracy and further assessed utility by generating report rankings that represented a prioritized report review process.

Results: Our random forest model showed the best performance in report ranking and accuracy, with an area under the receiver operating characteristic curve of 0.66. The generated report ordering assigns reports with a higher probability of medication-related causality a higher rank and is significantly correlated to a perfect report ordering, with a Kendall's tau of 0.24 ($P = .002$).

Conclusion: Our models produced prioritized report orderings that enable FDA safety evaluators to focus on reports that are more likely to contain valuable medication-related adverse event information. Applying our models to all FDA adverse event reports has the potential to streamline the manual review process and greatly reduce reviewer workload.

Key words: drug-related side effects and adverse reactions, supervised machine learning

BACKGROUND AND SIGNIFICANCE

The US Food and Drug Administration (FDA) receives more than 4000 medication safety reports every day, and the number of reports received each year has been increasing exponentially over the last decade. These reports are stored in a database known as the FDA Adverse Event Reporting System (FAERS), which has collected over 11 million reports since its inception in 1969.[1] In the United States, reporting these adverse events, medication errors, and product quality issues by health care professionals and consumers via the MedWatch program is voluntary, but it is mandatory for drug manufacturers.[2] The FDA uses these reports to detect safety issues that may not have been identified during pre-market clinical trials used as the basis for medication approval. Among the reasons for not detecting safety issues during pre-market evaluation are that the reporting systems, including the FDA Adverse Event Reporting System. Over the last decade, the number of adverse event reports has increased exponentially, resulting in a substantial workload for reviewers. Delays in detecting drug adverse events can have costly and detrimental effects on public health, and thus a system to identify reports most likely to contain information demonstrating causal drug events would be highly beneficial. Researchers have investigated such approaches using the US Vaccine Adverse Event Reporting System, in which extracted text features[16,17] were used with multiple classification algorithms to create an effective report classification model.[18–20]

The success of text classification in the Vaccine Adverse Event Reporting System and previous computational discoveries of new medication-related adverse events in FAERS[23–27] have generated significant interest in developing a classification system for FAERS. To accomplish this, we built models to classify and rank adverse event reports based on the likelihood of medication-related causality. In addition, we showed the potential utility of our models to assist manual adjudicators by shifting reports with a higher probability of medication-related causality to a higher priority in rank order.

For the first phase of this study, we chose to focus on reports with assessments of *Certain* to *Unassessable*, as they constituted

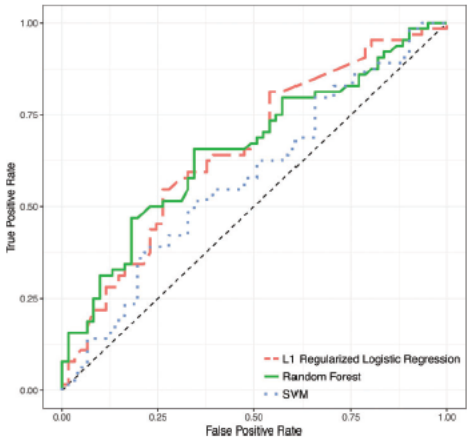


Figure 2. ROC curves for all classification models.

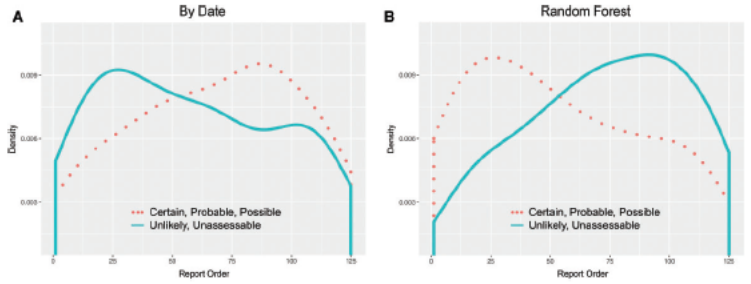


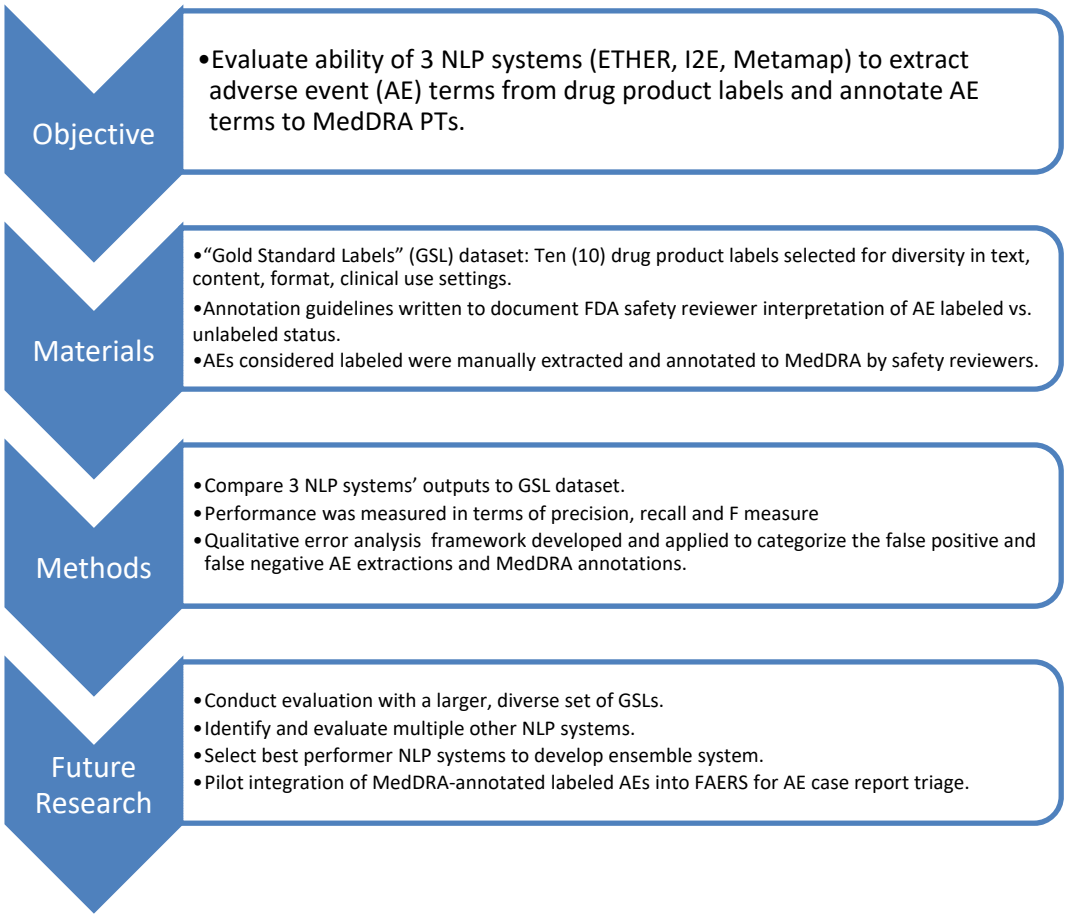
Figure 3. Comparison of report orderings in the held-out test set by (A) date and (B) random forest with assessments of *Certain*, *Probable*, or *Possible* vs assessments of *Unlikely* or *Unassessable*.

Evaluation of Natural Language Processing (NLP) Systems to Annotate Drug Product Labeling with MedDRA Terminology: A Pilot Study

Problem: FDA receives nearly 2 million FAERS adverse event (AE) case reports annually. FDA safety reviewers manually review many cases to identify serious, unlabeled AEs. Case triage using MedDRA, the FAERS coding dictionary, could increase AE case review efficiency. However MedDRA is not used to describe AEs in drug product labels. Manual annotation of AEs with MedDRA in all drug labels would be resource-intensive.

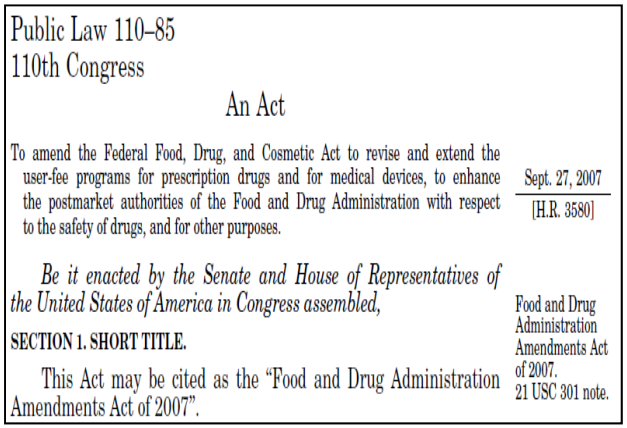
Solution: Develop NLP systems ensemble to automate annotation of drug labels with MedDRA. Identify potential NLP systems for an ensemble by evaluating operating characteristics when used to extract AE terms from drug product labels and annotate AE terms with MedDRA.

Approach:



2007 FDA Amendments Act (FDAAA)

- Post Marketing Requirements
- Safety Labeling Changes
- Risk Evaluation and Mitigation Strategies (REMS)
- Required Safety Reviews ("915" and "921")
- **Active post-market Risk Identification and Analysis (ARIA) system**
  - **FDA Sentinel Initiative**



Public Law 110–85
110th Congress

An Act

To amend the Federal Food, Drug, and Cosmetic Act to revise and extend the user-fee programs for prescription drugs and for medical devices, to enhance the postmarket authorities of the Food and Drug Administration with respect to the safety of drugs, and for other purposes.

Sept. 27, 2007
[H.R. 3580]

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,

SECTION 1. SHORT TITLE.

This Act may be cited as the "Food and Drug Administration Amendments Act of 2007".

Food and Drug Administration Amendments Act of 2007.
21 USC 301 note.

Defining ARIA

ARIA uses a subset of Sentinel System's full capabilities to fulfill the FDAAA mandate to conduct active safety surveillance

* Pre-defined, parameterized, and re-usable to enable faster safety surveillance in Sentinel (in contrast to protocol based assessments with customized programming)

† Electronic claims data, without manual medical record review

What is Sufficiency?

- Adequate data
  - Drug/biologic of interest and comparator
  - Confounders and covariates
  - Health outcome of interest
- Appropriate methods
- To answer the question of interest
  - assess a known serious risk related to the use of the drug/biologic
  - assess signals of serious risk related to the use of the drug/biologic
  - identify an unexpected serious risk when available data indicate the potential for a serious risk
- To lead to a satisfactory level of precision

When are automated queries insufficient?

- Inadequate data
  - Health outcome of interest
    - Claims alone not always adequate (e.g. anaphylaxis)
  - Creating cohorts – variety of reasons (e.g. cancer and missing data)
  - Confounders and other data of interest – not always available in common data model (e.g. BMI, smoking history, race, provider specialty, zip code)

Can NLP help?

Journal of the American Medical Informatics Association Advance Access published February 5, 2016

Extracting information from the text of electronic medical records to improve case detection: a systematic review

RECEIVED 12 May 2015
REVISED 13 October 2015
ACCEPTED 26 October 2015

Elizabeth Ford,[1] John A Carroll,[2] Helen E Smith,[1] Donia Scott,[2] and Jackie A Cassell[1]

ABSTRACT

Background Electronic medical records (EMRs) are revolutionizing health-related research. One key issue for study quality is the accurate identification of patients with the condition of interest. Information in EMRs can be entered as structured codes or unstructured free text. The majority of research studies have used only coded parts of EMRs for case-detection, which may bias findings, miss cases, and reduce study quality. This review examines whether incorporating information from text into case-detection algorithms can improve research quality.

Methods A systematic search returned 9659 papers, 67 of which reported on the extraction of information from free text of EMRs with the stated purpose of detecting cases of a named clinical condition. Methods for extracting information from text and the technical accuracy of case-detection algorithms were reviewed.

Results Studies mainly used US hospital-based EMRs, and extracted information from text for 41 conditions using keyword searches, rule-based algorithms, and machine learning methods. There was no clear difference in case-detection algorithm accuracy between rule-based and machine learning methods of extraction. Inclusion of information from text resulted in a significant improvement in algorithm sensitivity and area under the receiver operating characteristic in comparison to codes alone (median sensitivity 78% (codes + text) vs 62% (codes), $P = .03$; median area under the receiver operating characteristic 95% (codes + text) vs 88% (codes), $P = .025$).

Conclusions Text in EMRs is accessible, especially with open source information extraction algorithms, and significantly improves case detection when combined with codes. More harmonization of reporting within EMR studies is needed, particularly standardized reporting of algorithm accuracy metrics like positive predictive value (precision) and sensitivity (recall).

Keywords: electronic health records, review, text mining, data quality, case detection

INTRODUCTION

Information recorded in electronic medical records (EMRs), clinical reports, and summaries has the possibility of revolutionizing health-related research. EMR data can be used for disease registries, epidemiological studies, drug safety surveillance, clinical trials, and healthcare audits.

Information recording in EMRs

In most EMRs there is the possibility for the clinician both to code their findings in a structured format and also to enter information in narrative free text. There are various nomenclatures for structuring or coding information; the most widely used are International Classification of Diseases version 10,[1] Systematized Nomenclature of Medicine – Clinical Terms,[2] and the International Classification of Primary Care.[3] Within multi-modal EMRs there are also laboratory, pathology, and radiology reports, admission and discharge summaries, and chief complaints fields, which are in unstructured or semi-structured text. The balance of recording by the clinician, between codes and narrative text, is likely to vary by institution, EMR system, department, disease type, and component of the record.

Why do EMRs contain free text instead of being completely structured?

Clinicians experience a tension between choosing to code information and expressing it in text.[4] Among the main motivators for clinicians to code rather than use text is the increased ease of search, access, and retrieval.[5,6] A coded record allows the clinician to readily demonstrate that appropriate care has been provided, accurate diagnoses are made, and targets met.[7] This is especially important for billing after episodes of care, or for incentive based systems such as the National Health Service (NHS) Quality and Outcomes Framework in UK primary care.[8]

Coded data can be analyzed and summarized easily and on a large scale, whereas free text cannot. In contrast to structured data, narrative text is highly variable,[9] but is more engaging, captures the patient's narrative, can be told from different perspectives, and allows expression of feelings.[10] It is a better reminder for the clinician of the human encounter.[7]

Additionally, clinicians have given a number of reasons why they find coding onerous; the choices available in coded data may be too limiting, and may not allow for the expression of nuances.[11] The process of finding and entering codes on the computer represents an additional cognitive load,[9] and may take longer than summarizing the consultation in text.[6] Free text may be chosen when no code precisely describes clinical findings, or when there is a need to give supporting evidence for a diagnosis or suspicion.[12] Clinicians use free text as a pragmatic solution to recording vague diagnoses or strange collections of symptoms, when diagnoses need qualification, and for psychosocial problems.[7] Text can summarize processes of deduction, and modal language can be used to convey a range of possible outcomes.

Correspondence to Elizabeth Ford, Division of Primary Care and Public Health, Brighton and Sussex Medical School, Mayfield House, Village Way, Falmer, Brighton, BN1 9QH, UK; e.m.ford@bsms.ac.uk; Tel: (+44) 01273 641974. For numbered affiliations see end of article.

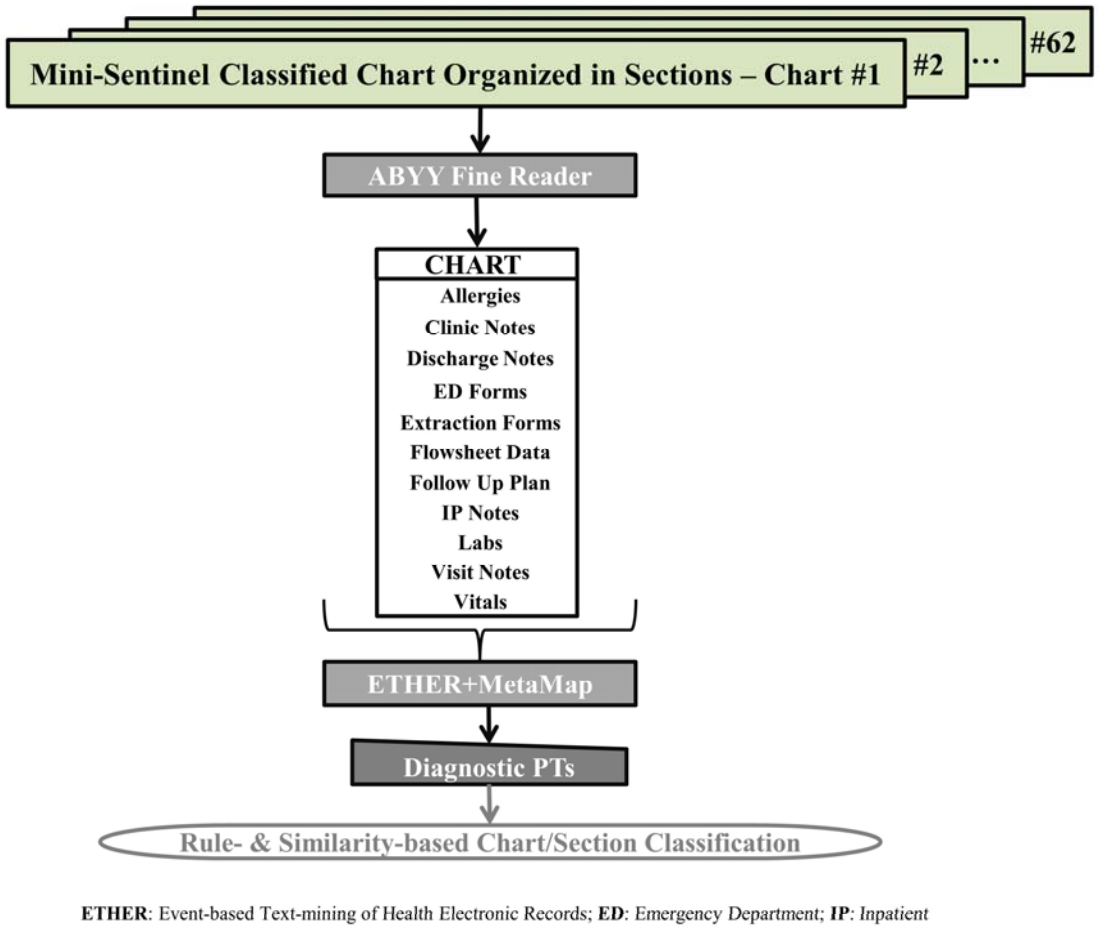
Ford E et al. J Am Med Inform Assoc 23 (5), 1007-1015. 2016.

Key Points

"**Text in EMRs** is accessible, especially with open source information extraction algorithms, and **significantly improves case detection when combined with codes**. More harmonization of reporting within EMR studies is needed, particularly standardized reporting of algorithm accuracy metrics like positive predictive value (precision) and sensitivity (recall)."

Authors also noted small sample that directly compared codes to narratives and **variability in performance.**

Health Outcome of Interest: Anaphylaxis



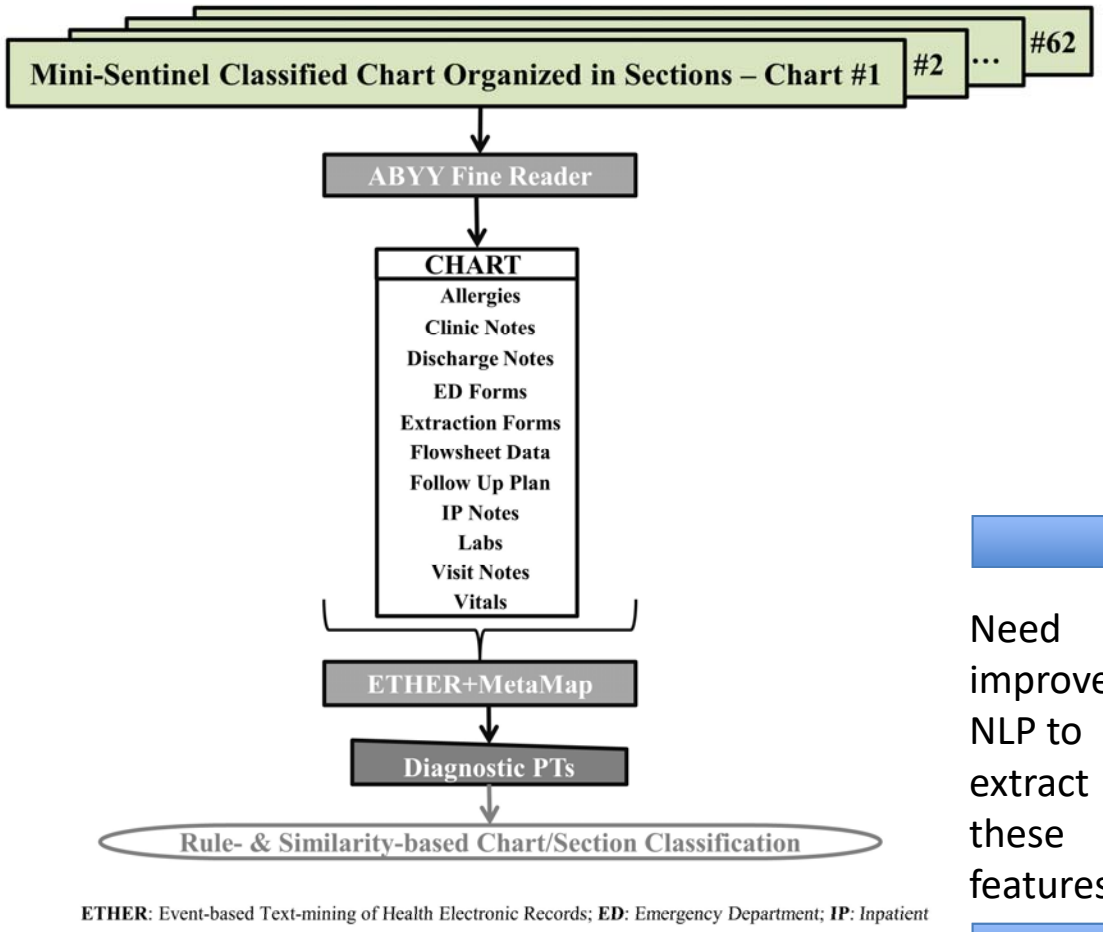
ETHER: Event-based Text-mining of Health Electronic Records; **ED**: Emergency Department; **IP**: Inpatient

KEY POINTS

- The previously developed natural language processing, rule- and similarity-based classification approaches demonstrated almost equal performance (F-measure: 0.753 vs. 0.729, recall 100% vs 100%, precision 60.3% vs 57.4%).
- These algorithms might improve recall but had **similar precision (PPV) to claims only algorithms from MS**.

Ball et al, Evaluating automated approaches to anaphylaxis case classification using unstructured data from the FDA Sentinel System, submitted

Health Outcome of Interest: Anaphylaxis



Need improved NLP to extract these features

KEY POINTS

- Reasons for misclassification included: the **inability** of the algorithms **to make the same clinical judgments as human experts** about the timing, severity, or presence of alternative explanations; the identification of terms consistent with anaphylaxis but present in conditions other than anaphylaxis.

Ball et al, Evaluating automated approaches to anaphylaxis case classification using unstructured data from the FDA Sentinel System, submitted

Summary

- Opportunity for NLP of clinical texts to help FDA improve efficiency and scientific validity of safety and effectiveness evaluation across the product lifecycle
  - Improvements in NLP and addition of machine learning likely important next steps
- Can also contribute to development of improved evidence Generation System and Learning Healthcare System

Acknowledgements

- *Taxiarchis Botsis*
- *Carol Pamer*
- *Mitra Rocca*
- *Michael Nguyen*

Thank you