

Original Paper

# Automatically Recognizing Medication and Adverse Event Information From Food and Drug Administration's Adverse Event Reporting System Narratives

Balaji Polepalli Ramesh<sup>1,2</sup>, MS Comp Sc, PhD; Steven M Belknap<sup>3</sup>, MD; Zuofeng Li<sup>1,2</sup>, MD, PhD; Nadya Frid<sup>1,2</sup>, MS; Dennis P West<sup>4</sup>, PhD; Hong Yu<sup>1,5,6</sup>, PhD

<sup>1</sup>University of Massachusetts Medical School, Worcester, MA, United States

<sup>2</sup>University of Wisconsin Milwaukee, Milwaukee, WI, United States

<sup>3</sup>Northwestern University Feinberg School of Medicine, Department of Dermatology and Department of Medicine, Division of General Internal Medicine and Geriatrics, Chicago, IL, United States

<sup>4</sup>Northwestern University Feinberg School of Medicine, Department of Dermatology, Chicago, IL, United States

<sup>5</sup>VA Central Western Massachusetts, Leeds, MA, United States

<sup>6</sup>University of Massachusetts, Department of Computer Science, Amherst, MA, United States

**Corresponding Author:**

Hong Yu, PhD

University of Massachusetts Medical School

AS6-2071

368 Plantation St

Worcester, MA, 01605

United States

Phone: 1 (508) 856 3474

Fax: 1 (508) 856 8993

Email: [Hong.Yu@umassmed.edu](mailto:Hong.Yu@umassmed.edu)

## Abstract

**Background:** The Food and Drug Administration's (FDA) Adverse Event Reporting System (FAERS) is a repository of spontaneously-reported adverse drug events (ADEs) for FDA-approved prescription drugs. FAERS reports include both structured reports and unstructured narratives. The narratives often include essential information for evaluation of the severity, causality, and description of ADEs that are not present in the structured data. The timely identification of unknown toxicities of prescription drugs is an important, unsolved problem.

**Objective:** The objective of this study was to develop an annotated corpus of FAERS narratives and biomedical named entity tagger to automatically identify ADE related information in the FAERS narratives.

**Methods:** We developed an annotation guideline and annotate medication information and adverse event related entities on 122 FAERS narratives comprising approximately 23,000 word tokens. A named entity tagger using supervised machine learning approaches was built for detecting medication information and adverse event entities using various categories of features.

**Results:** The annotated corpus had an agreement of over .9 Cohen's kappa for medication and adverse event entities. The best performing tagger achieves an overall performance of 0.73 F1 score for detection of medication, adverse event and other named entities.

**Conclusions:** In this study, we developed an annotated corpus of FAERS narratives and machine learning based models for automatically extracting medication and adverse event information from the FAERS narratives. Our study is an important step towards enriching the FAERS data for postmarketing pharmacovigilance.

(*JMIR Med Inform* 2014;2(1):e10) doi: [10.2196/medinform.3022](https://doi.org/10.2196/medinform.3022)

**KEYWORDS**

natural language processing; pharmacovigilance; adverse drug events

## Introduction

### Background

An adverse event (AE) is an injury or untoward medical occurrence to a patient or clinical investigation subject who has been administered a pharmaceutical product and the AE does not necessarily have a causal relationship with the administered treatment [1,2]. An adverse drug event (ADE) is an injury resulting from a medical intervention related to a drug, including harm caused by the drug (adverse drug reactions and overdoses), and harm from the use of the drug (including dose reductions and discontinuations of drug therapy) [3,4]. Studies have reported that ADEs account for nearly 20% of all adverse events that occur in hospitalized patients [5-7]. In the United States alone, ADEs account for more than 770,000 injuries and deaths annually [8-10], and an increased average length of stay in hospitals at a cost of between \$1.56 and \$5.60 billion annually [3,11]. Improved methods for ADE detection and analysis may identify novel drug safety signals and lead to improved methods for avoiding ADEs, with their attendant burden of morbidity, mortality, and cost. As part of a major effort to support postmarketing drug safety surveillance, the US Food and Drug Administration (FDA) receives mandatory reports on ADEs from manufacturers through the FDA Adverse Event Reporting System (FAERS). The FAERS is a database that captures information concerning adverse events and medication errors associated with FDA-approved prescription drugs. Currently, FAERS contains over four million reports of adverse events dating from 1969 to present [12]. It serves as a rich resource for pharmacovigilance—the study of drug-related injuries for the purpose of making warning or withdrawal recommendations for pharmaceutical products [4]. A typical FAERS report incorporates both structured data and unstructured free text, as

shown in Figure 1. The structured data entries incorporate each patient's personal and demographic information, a list of prescribed drugs, and the class of drug reaction (in this example, "anaphylactic reaction") (Figure 1). The Event/Problem narrative contains additional information relevant to describing the event, assessing causality, and grading severity (Figure 1). In this example, the narrative text contains the phrase that indicates causality between paclitaxel and the anaphylactic reaction while "experienced a life threatening anaphylactic reaction" shows the severity of the event, which is not coded in the structured data.

Although FAERS is an excellent resource to study drug effects, as stated in Tatonetti et al [13], the structured data does not incorporate confounding factors including concomitant medications and patient medical histories, which limits FAERS' effectiveness for pharmacovigilance. In contrast, such confounding factors are frequently described in the FAERS narratives. Making this data computationally available is critical for pharmacovigilance.

Currently, manual abstraction is required for identification of relevant data in FAERS narratives. Manual abstraction is expensive and often not practical, given the current size of the FAERS dataset, which contains millions of records. Therefore, it is important to develop computational approaches to automatically extract information from FAERS narratives. In this study, we report the development of both a corpus of FAERS narratives annotated with medication and adverse event information and a Natural Language Processing (NLP) system called AETagger that automatically extracts this information from the narratives and is adapted from existing tools. This is an important step towards enriching the existing FAERS' capacity for pharmacovigilance.

**Figure 1.** A sample AERS Report with structured data and narrative text.

FDA Adverse Event Reporting System (FAERS)

ISR Report for report # ISR 4901614-2

**Reporter Name:**

**Reporter Org:**

**Reporter Street:**

**Reporter Zip:**

**Product (s)**

TAXOL

KYTRIL

DECADRON PHOSPHATE

GLUCOPHAGE

ACIPHEX

LEVAQUIN

PYRIDIUM

**Reaction (s)**

ANAPHYLACTIC REACTION

**Disease/Surgical Procedure**

FOOD ALLERGY

DRUG HYPERSENSITIVITY

DIABETES MELLITUS

BREAST CANCER

DEPRESSION

#### **Event/ Problem Narrative**

A nurse reported that a 60- year old female consumer experienced a delayed allergic reaction while on paclitaxel therapy. The patient received her first dose of paclitaxel 280 mg on [DATE]. On [DATE], hours later, the consumer presented at the emergency room with pain in her hands and feet, feeling of throat closing and a rash (generalized). Epinephrine was administered and her symptoms resolved. She was discharged to home on the same day. The patient's medical history was a significant for diabetes, left breast cancer, depression and allergies of shell fish and sulfa. Supplemental information was received from the reporting nurse on [DATE]; The patient experienced a life threatening anaphylactic reaction after receiving one dose of paclitaxel, on [DATE] for treatment of breast cancer. The Anaphylactic reaction was characterized by symptoms of pain in her hands and feet, feeling of throat closing and rash (generalized).

#### **Related Work**

There is extensive research related to AE and ADE detection and analysis from a variety of data sources. Earlier work examined patients' paper medical records determining whether AEs and ADEs can be reliably abstracted based on the information conveyed in those records. For example, Hiatt et al (1989) [14] was among one of the early studies that defined an AE as an injury caused at least in part by medical mismanagement (negligence). They then manually abstracted ADEs from patients' paper-based clinical medical records. Similarly, other early studies (eg, [3,7,15]) defined AEs and ADEs and manually abstracted them from clinical records. These studies indicate the feasibility and value of clinical records for ADE surveillance and prevention.

When electronic medical records (EMRs) became available, computational approaches were developed to automatically identify AE and ADE information from EMRs. Studies used rule-based approaches for detecting ADEs from EMR data [16-18]. Tinoco et al [19] compared a rule-based computer surveillance system called Health Evaluation through Logical

**Reporter Phone:**

**Reporter City:**

**Reporter State:**

**Reporter Country:**

**Route**

INTRAVENOUS

INTRAVENOUS

INTRAVENOUS

**Indication (s)**

BREASTCANCER

PREMEDICATION

PREMEDICATION

Processing (HELP) [20] with manual chart reviews on 2137 patient admissions. They reported that HELP detected as many ADEs as were found by manual chart review, suggesting that NLP systems could improve ADE detection from EMR narrative data.

Many studies applied NLP to detect AEs and then inferred a causality relationship between a drug and an AE (called an ADE) using logical rules, statistical analyses, and supervised machine learning (ML) approaches. Hazlehurst et al [21] developed MediClass, a knowledge-based system that deploys a set of domain-specific logical rules to medical concepts that are automatically identified from EMR narratives (eg, progress notes) or precoded data elements (eg, medication orders). The system achieved a precision of 64% for detecting vaccine-related AEs [22]. A number of studies applied the NLP system [23-25], MedLEE [26], to detect AEs from discharge summaries and hospitalization records. For example, Wang et al [23] applied MedLEE to detect terms and mapped them to the Unified Medical Language System (UMLS) semantic types. Subsequently, they detected medication and AEs when the terms were mapped to the UMLS concepts with the semantic types

of Clinical Drug (T200) and Disease or Symptom (T047), respectively. The causality relationship between a medication and an AE was extracted from 25K discharge summaries based on a  $\chi^2$ -statistical analysis of medication and AE. Evaluation of seven drugs for known ADEs led to a recall and precision of 75% and 31% respectively. Aramaki et al [27] manually annotated 435 discharge summaries for drugs and ADEs and then applied supervised machine learning techniques to detect these named entities. They identified the causality between drugs and AEs using pattern matching and SVM techniques. They reported a recall and precision score of 0.81 and 0.87 for drug, and 0.80 and 0.86 for AE detection respectively. For inferring causality they achieved recall and precision of 0.92 and 0.41 using pattern matching, and 0.62 and 0.58 using SVM technique respectively.

In addition to EMRs, studies have explored other data sources for ADE information, including biomedical literature [28,29], social media and the Internet [30-32]. Shetty and Dalal [33] mined ADEs from PubMed citations. They first built a document classifier to identify relevant documents that incorporate ADE relationships using Medical Subject Headings (MeSH) terms. For example, if an article is assigned “chemically induced” or “adverse effects,” then the article is likely to incorporate an ADE. They then identified ADE signals using disproportionality analysis in which the rate at which a particular AE of interest co-occurs with a given drug is compared to the rate an AE occurs without the drug in the collection. Their evaluation on a predefined set of 38 drugs and 55 AEs showed that their literature-based approach could uncover 54% of ADEs prior to FDA warnings.

There is a rich store of literature for ADE detection on Spontaneous Reporting Systems (SRS) such as the FAERS reports and WHO Vigibase [34]. Studies have explored several statistical data mining and machine learning techniques on SRS for the detection of ADE signals [13,35-60]. However, all aforementioned approaches for ADE detection from FAERS are based on its structured data. In this study, we report the development and evaluation of supervised machine learning approaches for automatically detecting medication information and adverse events from the FAERS narratives. We speculate that such information can be a useful addition to the FAERS structured data for ADE detection.

## Methods

### Annotation Data and Procedure

Through our collaboration at Northwestern University [61], we obtained a collection of 150 de-identified FAERS narratives; a sample is shown in Figure 1. The data collection originally came as a scanned PDF image file. With Institutional Review Board (IRB) approval from Northwestern University and University of Wisconsin Milwaukee, we manually transcribed the PDF file into a computer-readable text file.

We randomly selected a set of 28 narratives for developing the annotation guideline (Multimedia Appendix 1). Our annotation guideline was based on the i2b2 challenges in NLP for Clinical Data Medication Extraction [62,63]. A balanced interdisciplinary

team consisting of a linguist (NF), a physician (SB), two informaticians (BPR and HY) and a physician informatician (ZFL) developed the annotation guideline through an iterative process. At the end of reviewing 28 narratives, we obtained a guideline that all the members of the team agreed upon.

Following the final annotation guideline, two annotators (ZFL, designated as AnnPhy, and NF, designated as AnnLing), both of whom were the primary annotators for the i2b2 medication event detection challenge [63] in which we participated, independently annotated the remaining 122 AERS narratives. The different backgrounds of the annotators aids in building a corpus that is both linguistically driven and clinically correct. A physician (SB) served as a tiebreaker and resolved annotation disagreements. This collection of 122 narratives is comprised of approximately 23,000 word tokens and the average number of words per narrative is 190.2 (SD 130.3).

The annotation was carried out using Knowtator [64], a plugin for Protégé [65]. The Knowtator interface allows users to define entities that need to be annotated and configure the relationships between them. The 122 annotated narratives were used as both training and testing data for machine learning approaches described below. The annotated data was grouped into four collections each containing 122 narratives: *AnnPhy* and *AnnLing* –data annotated by annotators AnnPhy (ZFL) and AnnLing (NF), respectively; *Comb* –a joint set of annotations agreed upon by both AnnPhy and AnnLing, and *Tie* –a joint set of AnnPhy and AnnLing annotations where disagreements were resolved by the tiebreaker SB. We also report Cohen’s kappa, a well-known statistic used to assess the degree of Inter-Annotator Agreement (IAA) between annotators [66]. We use these four sets of data to capture all named entities and build robust supervised machine learning classifiers to identify them.

### Supervised Machine Learning

#### Machine Learning Techniques

Three supervised machine learning approaches were explored for automatically identifying medication information and adverse events: Naïve Bayes (NB), Support Vector Machines (SVMs) and Conditional Random Fields (CRFs) [67]. We built NB and SVM classifiers using Weka [68] and the CRF model was built using the A Biomedical Named Entity Recognizer (ABNER) toolkit [69]. NB is a simple model that assumes all attributes of the examples which are independent of each other given the context of the class. SVMs are a well-known statistical machine learning algorithm and have shown very good performance in many classification tasks [70,71]. CRFs have shown success in named entity recognition in the biomedical domain [69,72].

#### Learning Features

We explored a variety of features such as syntactic features, semantic features based on the external knowledge resource (UMLS), morphological and contextual features, presence of negation, hedging and discourse connectives as a feature in addition to ABNER default features which include bag of words and orthogonal features. We describe each of these in detail below.

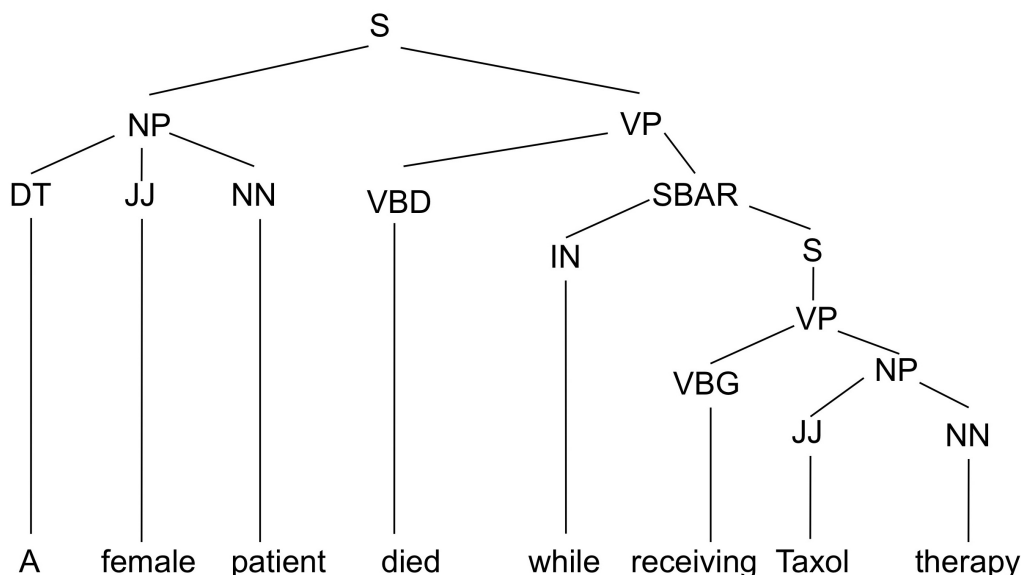
The syntactic features include the part-of-speech (POS), the phrasal class of each token, and the POS of the token immediately to the left of the token under consideration. The syntactic features were extracted from the constituency parse tree generated by the Charniak-Johnson parser [73] trained in the biomedical domain. This parser was determined to have the best performance when tested on the GENIA corpus [74]. Figure 2 shows a sample constituency parse tree. In this example, the POS features determiner (DT), adjective (JJ), noun (NN) are the POS of tokens “A”, “female”, and “patient” respectively. Further, the phrasal class for all the three tokens is noun phrase (NP). The left sibling POS value of “A” is NONE assuming it is the start of the sentence. The left sibling POS of “female” and “patient” tokens are DT and JJ respectively.

We applied the UMLS Metamap [75,76] to extract semantic features, which are concepts and semantic types represented in

the UMLS Metathesaurus. The morphological features were obtained by considering various characteristics of the word. We took attributes of the word, such as whether it was a digit, was capitalized, its alphanumeric order (ie, if the token started with letters and was followed by numerals or vice versa), and the presence of punctuation such as commas and hyphens. These features were extracted using a simple pattern-matching technique. The first (prefix) and last (suffix) three and four characters of the token were added as affix features.

We added as features, negation and hedging cues with their scope that were detected automatically by the systems described in the literature [77,78]. We also added presence of discourse connectives that were automatically detected by the discourse parser [79].

**Figure 2.** The sample constituency parse tree. S: simple declarative clause, NP: noun phrase, VP: verb phrase, DT: determiner, JJ: adjective, NN: noun, VBD: verb, past tense, SBAR: subordinate clause, IN: preposition or subordinating conjunction, VBG: verb, gerund or present participle.



## Systems

### Overview

We developed several taggers to evaluate the complexity of the task for identifying medication information and adverse events and the impact of features.

### Systems to Evaluate Task Complexity

In this experiment, we built two baseline systems to compare the performance of ML algorithms. The first system *BaseDict* is a simple dictionary-matching system. A lexicon of medications and AEs is compiled from the UMLS Metathesaurus using the semantic types as defined by Wang et al [23], where terms having the semantic types Clinical Drug (T200) and Disease or Symptom (T047) were considered as drug and adverse event respectively. The baseline system *BaseDict*, tags all instances of the lexicon that match within the text. The second system, *MetaMapTagger*, is a UMLS Metamap [75] based system that tags phrases as AEs or medications using UMLS semantic types similar to *BaseDict*.

The baseline systems were compared with taggers built using bag of words as the default feature –*NBTagger*, a NB-based tagger, *SVMTagger*, a SVM-based tagger, and *SimpleTagger*, a CRF-based tagger built using ABNER default features. We then evaluate the taggers by adding all the features defined in the Learning Features section, which we call *NBTagger<sup>+</sup>*, *SVMTagger<sup>+</sup>* and *CombinedTagger* for NB, SVM- and CRF-based taggers respectively.

### Systems to Evaluate Impact of Features

We evaluate the impact of various features on the performance of tagger. We used the ML technique found to have the best performance in our previous experiment. In addition to the default features trained as *SimpleTagger*, we individually added syntactic features (*SyntacticTagger*), semantic features (*SemanticTagger*), morphological features (*MorphologicalTagger*), affix features (*AffixTagger*), negation and hedging features (*NegHedgeTagger*), discourse connective features (*ConnectiveTagger*), and a tagger incorporating all the features (*CombinedTagger*) which were trained to identify the named entities.

## Machine Learning Evaluation Metrics

All the AE taggers trained were evaluated using ten-fold cross-validation. We reported recall, precision, and F1 score. Recall is the ratio of the number of entities of a certain class correctly identified by the system and the number of entities of that class in the gold standard. Precision is the ratio of the number of entities of a certain class correctly identified by the system and the number of entities of that class predicted by the system. F1 score is the harmonic mean of precision and recall.

## Results

### Corpus Characteristics and Annotation Agreement

Table 1 shows the definitions of adverse event and medication-related named entities, the number of annotated instances, and Cohen's kappa value. The annotation agreement

is calculated based on two criteria: *strict* in which the two annotations have an exact match, and *unstrict* in which there exists an overlap of at least one word between the two annotations. We measured the agreement using *unstrict* criteria to estimate the agreement between annotators when entity boundary is ignored. The table also shows the number of instances annotated in all four data sets.

As shown in Table 1, *adverse event (AE)* was the most frequently annotated entity followed by *medication* entity. *Duration* had the least number of annotated instances and lowest kappa value (.34) for *strict* criteria. *Indication* had the second highest kappa value for *unstrict* criteria (.93) after *medication* (.95), since most of the *indication* entities were followed by explicit and unambiguous patterns such as “for the treatment of”, “diagnosed with”, “due to”, “enrolled in breast cancer study”, and so on.

**Table 1.** Named entity definition, number of annotated instances, and inter-annotator agreement measured by Cohen's kappa for both strict and unstrict criterion.

Named entity	Definition	Number of instances annotated				kappa (strict)	kappa (unstrict)
		AnnPhy	AnnLing	Comb	Tie		
Medication	Name of the drug they administered to patient including drug class name or medications referred to with	1231	1278	1152	1286	.92	.95
Dosage	Amount of a single medication used in each administration	143	315	137	205	.59	.82
Route	Method for administering the medication	115	244	107	132	.59	.64
Frequency	How often each dose of the medication should be taken	25	56	21	42	.58	.74
Duration	How long the medication is to be administered	34	153	24	51	.34	.87
Indication	Medical conditions for which the medication is given	175	148	126	175	.76	.93
Adverse event (AE)	Harm directly caused including the pronouns referring to it by the drug at normal doses and during normal use	1689	2083	1646	1842	.83	.93
Other signs, symptoms and diseases (OSSD)	Other symptoms associated with the disease	234	140	90	147	.50	.71
Treatment	Treatment the patient received for the disease	77	216	62	153	.39	.77
Total		3723	4633	3365	4033		

### Results of Supervised Learning

Table 2 reports recall, precision, and F1 score of the AETaggers for identifying the AE and other medication-related named entities on each of the four data sets as described in Annotation and data procedure section.

The baseline system *BaseDict* that matches only *AE* and *medication* achieved an F1 score of 0.45, 0.41, 0.46, and 0.42 on the *AnnPhy*, *AnnLing*, *Comb*, and *Tie* datasets respectively. The *MetamapTagger* also had similar performance. Among the taggers using bag of words as features, the CRF-based *SimpleTagger* had the best performance. The addition of features improved the performance of the ML classifiers. The *CombinedTagger* achieved best performance with F1 scores of 0.69, 0.74, and 0.73 on the *AnnPhy*, *AnnLing*, and *Comb* datasets

respectively. The *SVMTagger*<sup>+</sup> had the best performance with a 0.66 F1 score on the *Tie* dataset. The difference in performance between *CombinedTagger* and *SVMTagger*<sup>+</sup> taggers was statistically significant only on *AnnLing* dataset (*t* test, *P*=.003). The ML-based taggers clearly outperform the baseline method. The CRF-based tagger had the best overall performance and was therefore chosen as the system to be adopted for subsequent experiments measuring impact of features.

We trained the CRF-based AETaggers using different features as described in the Learning Features section. The results show that the *CombinedTagger* achieved the highest performance on all datasets. Our results also show that the *AnnLing* dataset has the highest performance while *Tie* performs the lowest. *Comb* outperforms both *Tie* and *AnnPhy*.

Since the *Comb* dataset's performance (0.73 F1 score) is close to the highest (0.74 F1 score) and contains annotations agreed upon by both annotators, we further report feature analyses using the *Comb* dataset. Table 3 shows how different learning features affect AETagger's performance. The results show that adding a single feature added little to the overall performance, although the performance of different entities varied. Affix features improved *route* and *duration* but decreased *AE*,

*medication*, and *dosage*. Connective features increased the performance of *dosage*, *route*, and *indication*; however, the performance of *medication* decreased. Other features (morphological, negation, hedge, semantic, and syntactic) showed similar patterns. On the other hand, when all features were added, the overall performance increased to 0.73 F1 score (default 0.71), although the increase was not statistically significant ( $t$  test,  $P=.08$ ).

**Table 2.** The precision, recall, and F1 score of Taggers on each of the four annotated data sets ( $t$  test,  $P<.01$ ).

Machine learning	AnnPhy			AnnLing			Combined			Tie		
	Mean (SD)			Mean (SD)			Mean (SD)			Mean (SD)		
	F1	Preci- sion	Recall	F1	Preci- sion	Recall	F1	Preci- sion	Recall	F1	Preci- sion	Recall
<b>Task complexity</b>												
BaseDict	0.45 (0.10)	0.86 (0.08)	0.31 (0.09)	0.41 (0.09)	0.91 (0.07)	0.27 (0.08)	0.46 (0.12)	0.82 (0.06)	0.32 (0.11)	0.42 (0.10)	0.86 (0.13)	0.28 (0.08)
MetaMapTagger	0.41 (0.17)	0.41 (0.16)	0.42 (0.18)	0.41 (0.10)	0.47 (0.20)	0.37 (0.15)	0.42 (0.18)	0.41 (0.17)	0.43 (0.19)	0.40 (0.16)	0.46 (0.19)	0.36 (0.14)
NBTagger	0.22 (0.08)	0.39 (0.17)	0.15 (0.05)	0.23 (0.08)	0.45 (0.14)	0.16 (0.06)	0.24 (0.08)	0.40 (0.17)	0.17 (0.05)	0.20 (0.06)	0.47 (0.19)	0.13 (0.04)
SVMTagger	0.55 (0.05)	0.77 (0.10)	0.44 (0.04)	0.55 (0.05)	0.78 (0.07)	0.43 (0.05)	0.58 (0.04)	0.78 (0.10)	0.46 (0.04)	0.59 (0.04)	0.80 (0.05)	0.46 (0.05)
SimpleTagger	0.67 (0.09)	0.77 (0.09)	0.60 (0.09)	0.72 (0.08)	0.81 (0.06)	0.66 (0.10)	0.71 (0.08)	0.81 (0.09)	0.63 (0.08)	0.63 (0.09)	0.69 (0.08)	0.55 (0.10)
NBTagger <sup>+</sup>	0.45 (0.09)	0.38 (0.10)	0.56 (0.06)	0.44 (0.06)	0.39 (0.07)	0.50 (0.06)	0.46 (0.09)	0.37 (0.11)	0.60 (0.04)	0.43 (0.07)	0.38 (0.08)	0.51 (0.07)
SVMTagger <sup>+</sup>	0.66 (0.07)	0.78 (0.10)	0.58 (0.06)	0.67 (0.07)	0.78 (0.07)	0.59 (0.07)	0.70 (0.06)	0.80 (0.11)	0.63 (0.05)	0.66 (0.07)	0.78 (0.06)	0.57 (0.08)
CombinedTagger	0.69 (0.09)	0.77 (0.10)	0.62 (0.09)	0.74 (0.08)*	0.81 (0.07)	0.68 (0.09)	0.73 (0.08)	0.81 (0.10)	0.66 (0.07)	0.65 (0.08)	0.71 (0.08)	0.60 (0.09)
<b>Impact of features</b>												
SimpleTagger	0.67 (0.09)	0.77 (0.09)	0.60 (0.09)	0.72 (0.08)	0.81 (0.06)	0.66 (0.10)	0.71 (0.08)	0.81 (0.09)	0.63 (0.08)	0.63 (0.09)	0.69 (0.08)	0.55 (0.10)
AffixTagger	0.67 (0.09)	0.78 (0.09)	0.60 (0.09)	0.73 (0.09)	0.81 (0.06)	0.66 (0.10)	0.70 (0.08)	0.81 (0.09)	0.63 (0.08)	0.61 (0.09)	0.70 (0.08)	0.52 (0.10)
ConnectiveTagger	0.67 (0.09)	0.77 (0.09)	0.60 (0.09)	0.73 (0.08)	0.81 (0.06)	0.66 (0.10)	0.71 (0.08)	0.81 (0.09)	0.63 (0.08)	0.63 (0.09)	0.70 (0.07)	0.57 (0.10)
MorphologicalTagger	0.68 (0.09)	0.77 (0.08)	0.60 (0.10)	0.73 (0.08)	0.81 (0.06)	0.66 (0.09)	0.71 (0.08)	0.80 (0.09)	0.63 (0.08)	0.64 (0.08)	0.71 (0.07)	0.59 (0.09)
NegHedgeTagger	0.66 (0.09)	0.77 (0.09)	0.59 (0.10)	0.72 (0.08)	0.81 (0.06)	0.65 (0.10)	0.71 (0.08)	0.81 (0.09)	0.63 (0.08)	0.61 (0.09)	0.69 (0.08)	0.54 (0.10)
SemanticTagger	0.68 (0.09)	0.77 (0.10)	0.61 (0.09)	0.70 (0.09)	0.78 (0.07)	0.64 (0.10)	0.72 (0.09)	0.80 (0.11)	0.65 (0.08)	0.63 (0.09)	0.69 (0.10)	0.58 (0.09)
SyntacticTagger	0.68 (0.09)	0.78 (0.09)	0.61 (0.10)	0.72 (0.08)	0.80 (0.06)	0.65 (0.09)	0.71 (0.08)	0.80 (0.09)	0.64 (0.08)	0.63 (0.08)	0.70 (0.08)	0.58 (0.09)
CombinedTagger	0.69 (0.09)	0.77 (0.10)	0.62 (0.09)	0.74 (0.08)	0.81 (0.07)	0.68 (0.09)	0.73 (0.08)	0.81 (0.10)	0.66 (0.07)	0.65 (0.08)	0.71 (0.08)	0.60 (0.09)

**Table 3.** The F1 score of different named entities with different features on *Comb* dataset.

Feature group	AE mean (SD)	Medication mean (SD)	Dosage mean (SD)	Frequency mean (SD)	Route mean (SD)	Duration mean (SD)	Indication mean (SD)	OSSD mean (SD)	Treatment mean (SD)	Overall mean (SD)
Default	0.70 (0.10)	0.82 (0.10)	0.59 (0.35)	0.57 (0.46)	0.36 (0.33)	0.20 (0.42)	0.57 (0.12)	0.44 (0.45)	0.60 (0.52)	0.71 (0.08)
Affix	0.69 (0.11)	0.81 (0.12)	0.58 (0.37)	0.59 (0.45)	0.55 (0.37)	0.40 (0.52)	0.57 (0.09)	0.51 (0.44)	0.60 (0.52)	0.70 (0.08)
Connective	0.70 (0.10)	0.81 (0.10)	0.69 (0.31)	0.57 (0.46)	0.44 (0.36)	0.20 (0.42)	0.60 (0.15)	0.44 (0.45)	0.60 (0.52)	0.71 (0.08)
Morphological	0.70 (0.10)	0.82 (0.10)	0.57 (0.35)	0.59 (0.45)	0.32 (0.32)	0.20 (0.42)	0.62 (0.12)	0.47 (0.43)	0.60 (0.52)	0.71 (0.08)
NegHedge	0.69 (0.10)	0.82 (0.10)	0.56 (0.36)	0.59 (0.45)	0.36 (0.33)	0.20 (0.42)	0.59 (0.11)	0.50 (0.43)	0.60 (0.52)	0.71 (0.08)
Semantic	0.71 (0.11)	0.82 (0.11)	0.56 (0.35)	0.65 (0.40)	0.34 (0.33)	0.30 (0.48)	0.64 (0.13)	0.43 (0.39)	0.60 (0.52)	0.72 (0.09)
Syntactic	0.70 (0.10)	0.81 (0.11)	0.61 (0.35)	0.59 (0.45)	0.32 (0.31)	0.34 (0.47)	0.58 (0.11)	0.44 (0.45)	0.60 (0.52)	0.71 (0.08)
All	0.72 (0.10)	0.83 (0.11)	0.61 (0.37)	0.59 (0.44)	0.32 (0.31)	0.34 (0.47)	0.65 (0.11)	0.55 (0.39)	0.60 (0.52)	0.73 (0.08)

## Annotation Disagreements

### Overview

We manually analyzed the annotation disagreements and found they can be organized into three main categories: (1) boundary inconsistencies –disagreement due to assignment of inconsistent boundaries to entities; (2) missed named entity annotations –disagreement where one of the annotators annotated an entity and the other annotator missed it; (3) inconsistent named entity annotations –disagreement due to inconsistent categorization of entities.

There were a total of 2955 disagreed token instances, of which 1591 (53.84%) were related to *AE* and *medication* named entities.

### Textbox 1. Examples.

*Example 1:* She received approximately less than two minutes of therapy with intravenous Taxol (paclitaxel), 280 mg in a *three hour* [duration] *infusion* [route] for phase IIID ovarian cancer, when the symptoms occurred.

*Example 2:* The patient then became *lightheaded* [AE], *collapsed* [AE], and was *unconscious* [AE].

*Example 3:* Investigator considers that *haematologic toxicity* [AE] of methotrexate could be increased by interaction with apranax (naproxene) and sintrom (acenocoumarol).

*Example 4:* On [words marked], the patient died, presumed to be a result of *cardiogenic shock* [AE]. Prior to death, the patient was noted for having an increase in troponin T level, and found to be more unresponsive.

*Example 5:* Moderate anaphylactoid symptom appeared after administration of docetaxel and recovered later. After the end of administration, convulsion appeared. Anti-convulsion agent could not be administered due to *allergy* [AE].

*Example 6:* ...days after the last Vinorelbine intake patient was hospitalized due to *NCI/CTC grade 4 neutropenia* [AE] without fever [OSSD]...

### Missed Named Entity Annotations

Missed named entity annotation was the major cause for disagreement. Among 2955 disagreed token instances, 2355 or approximately 79.69% belong to this category. Table 4 shows instances of *medication* that were annotated by one annotator and missed by other. Examples 1-5 (Table 4) were annotated by AnnPhy but missed by AnnLing; examples 6-10 (Table 4) were annotated by AnnLing but missed by AnnPhy.

AnnLing explained that “blood transfusion”, “fluids”, and “red packed cells” shown in examples 1, 2, and 5 were not *medication*, but referred to a kind of treatment or medical procedure. In example 3, AnnLing missed annotating “normal saline” as *medication*. In example 4, “oxygen” was not annotated because AnnLing felt it did not represent *medication*. Annotators

### Boundary Inconsistencies

We found that inconsistencies related to boundary accounted for nearly 13.94% (412/2955) of disagreement. In all the examples in the article, the named entity instance is shown in italics and the named entity type is shown within the “[ ]”.

In Textbox 1 Example 1, AnnLing annotated “three hour” as *duration* and “infusion” as *route*, AnnPhy annotated “three hour infusion” as *duration* only. This inconsistency exemplifies differences between the linguist and the physician. While the linguist can separate the linguistic differences between different named entities, we found that physicians (both ZFL and SB) frequently overlook the differences, which leads to inconsistent annotations.

did not reach any consensus on annotating “oxygen” as *medication* or not. The differences here exemplify the strength of the physician as a domain-expert who may interpret the semantics of EMR notes more accurately than the linguist.

In examples 7, 8, and 10, in Table 4, AnnPhy did not annotate “treatment”, “Re-exposure”, and “chemotherapy” as these entities were anaphoric references; AnnLing, being a linguist, annotated these anaphoric references as *medication*. In example 6 (Table 4), AnnLing annotated “drug” as *medication* but AnnPhy did not annotate the entity because the text did not refer to any *medication*. Later, AnnLing agreed that where there is mention of entities but they do not refer to specific entities, such as “drug” in example 6, they should not be annotated. Example 9 in Table 4 was a special case where “concomitant drug” refers



to the role or function of the drug, “Solupred”, rather than referring to a drug. AnnPhy did not annotate such instances. These examples demonstrated that annotating medical texts is

a complex and cumbersome task. Further refinement of guidelines in such instances may improve the consistency of annotations.

**Table 4.** Disagreement in medication annotation (medication text is italicized).

Annotation	Medication annotation
Annotated by AnnPhy but not annotated by AnnLing	<ol style="list-style-type: none"> <li>1. Given multiple <i>blood transfusions</i> (hemoglobin: 4.8).</li> <li>2. Pressors continued with <i>fluids</i>.</li> <li>3. He was admitted to the hospital and hydrated with <i>normal saline</i>.</li> <li>4. The event was treated with steroids and <i>oxygen</i>.</li> <li>5. Pancytopenia, treated with G-CSF, erythropoetin, and <i>red packed cells</i>.</li> </ol>
Annotated by AnnLing but not annotated by AnnPhy	<ol style="list-style-type: none"> <li>6. Causality assessment <i>drug</i> relationship is unable to determine for Taxol.</li> <li>7. The 4th previous courses of <i>treatment</i> were well tolerated.</li> <li>8. During the first infusion of paclitaxel, the patient experienced a decrease in blood pressure and was unconscious for a short while. <i>Re-exposure</i> elicited the same symptoms.</li> <li>9. The <i>concomitant drug</i> prescribed was oral Solupred instead of Solumedrol.</li> <li>10. A female patient possibly received non-therapeutic dosages of intravenous Taxol (paclitaxel), Paraplatin (carboplatin), and/or Platinol (cisplatin) for the treatment of ovarian cancer and subsequently expired. It was reported that the pharmacist possibly diluted the <i>chemotherapy</i> improperly.</li> </ol>

### Inconsistent Named Entity Category Annotations

We have annotated a total of nine categories of named entities, as shown in Table 1. The third type of inconsistency was caused by inconsistent named entity assignments. Among 2955 disagreed token instances, 188 (6.36%) belong to inconsistent named entity categorization. We manually examined few instances and Examples 2-6 in Textbox 1 show the annotated sentences where inconsistency occurred. Example 2 in Textbox 1 is an example where both annotators agreed on the AE annotation.

Example 3 in Textbox 1, however, shows an instance where AnnPhy and the tiebreaker agreed on “haematologic toxicity” as an AE whereas AnnLing did not initially annotate the entity. This inconsistency suggests that domain knowledge is required for annotation. After discussion with two other annotators, AnnLing agreed that “haematologic toxicity” should be annotated an AE.

Example 4 in Textbox 1 shows an instance where AnnLing and the tiebreaker agreed on “cardiogenic shock” as an AE but AnnPhy annotated it as OSSD. AnnPhy argued that “cardiogenic shock” caused “death” and therefore “death” should be an AE and “cardiogenic shock” is the reason for death and therefore was annotated as OSSD. This example shows the complexity of clinical cause.

In Textbox 1 Example 5, the tiebreaker annotated “allergy” as an AE, whereas AnnPhy annotated it as OSSD and AnnLing did not annotate it as an AE because it refers to the patient’s history of “allergy” and does not represent a current instance of AE. We will need to refine our annotation guideline to add current or past status in addition to the named entity annotation.

Example 6 in Textbox 1 shows an instance of boundary inconsistency. AnnPhy and AnnLing both annotated “NCI/CTC grade 4 neutropenia without fever” as an AE whereas the tiebreaker annotated “NCI/CTC grade 4 neutropenia” as an AE and “fever” as OSSD. This is a case in which annotators interpret clinical texts differently. Such an inconsistency is

difficult to address due to the nature of ambiguity in clinical texts.

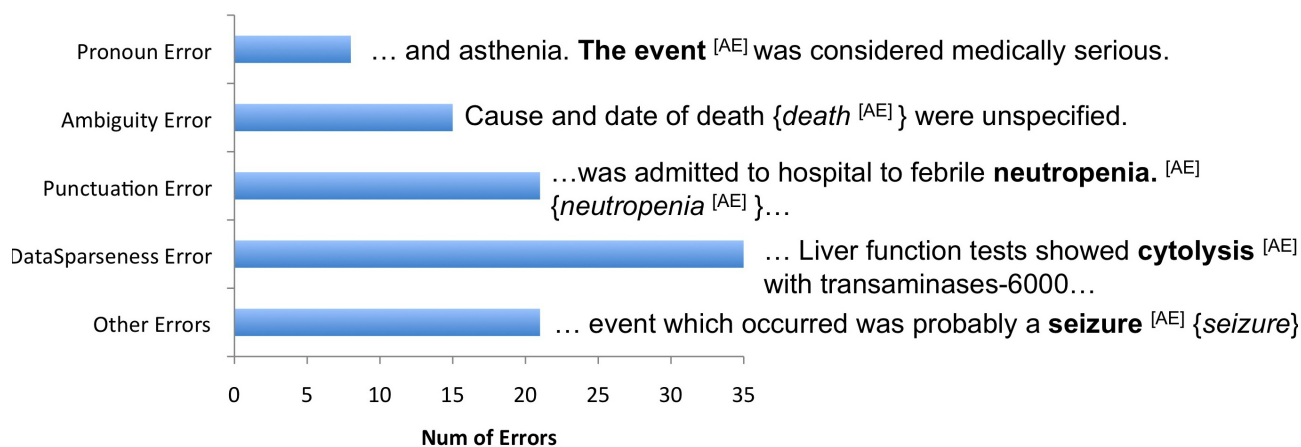
### Error Analyses

For error analyses, we focused on *CombinedTagger* because it yielded the highest performance (as shown in Table 2) and the *Comb* dataset because it contained annotations agreed on by both annotators. We randomly selected 100 named entities predicted wrongly by *CombinedTagger* and manually analyzed them. As shown in Figure 3, we group the named entities into a total of five types of errors and give an illustrative example for each. In all examples, annotated named entities are shown in bold, the tagger output in {*italicized*} and the named entity type is shown within “[ ]”. The leading type of error was data sparseness (35%). Data sparseness is a common problem and the major cause of poor performance. For instance, the gold standard consisted of a number of singleton instances (instances that appear only once) like “cytolysis”, “sodium chloride solution 0.9% 100ml”, and “neoplasm of unspecified nature of respiratory system” that created sparseness in the data.

The second cause of error was inconsistent inclusion of punctuation (21%). The gold standard had inconsistency in inclusion of punctuation (eg, a period [.] in “neutropenia.”) as a part of a named entity. This boundary inconsistency reduced the overall performance. Figure 3 shows an instance where the gold standard included a period as part of named entity “neutropenia.” but the tagger failed to include it (“neutropenia”). This was followed by an error caused by ambiguous named entities (15%). The instances in the gold standard that were assigned to multiple named entity categories resulted in ambiguous entities. For example, “death” was annotated as either *AE* or *OSSD*. This could have confused the ML algorithm and yielded a lower performance. In Figure 3, the instance “death” was not annotated as *AE* in the *Comb* dataset due to disagreements between annotators, but the tagger identified it as an *AE*. The missed pronoun annotations such as “the event” contributed to 8% of the errors. The final category was other type of errors (21%), for which the exact cause of error could

not be determined. In Figure 3, “seizure” was annotated as an *AE* but the tagger failed to identify it. The exact cause for miscategorization could not be determined.

**Figure 3.** Error categories, their frequency, and an illustrative example of error category on 100 randomly sampled instances. The annotated entities are shown in bold, the annotated named entity type is shown within “[ ]” and tagger output is {italicized}. AE: adverse events.



### Annotation Inconsistencies

As predicted, annotation inconsistency played an important role on AETaggers' performance as our Pearson correlation results (coefficient of 0.73) show that the IAA value (Cohens' kappa) is positively correlated with the machine learning performance of named entity recognition. This is not surprising because inconsistent annotations confuse the machine learning systems.

Our manual analysis of inconsistency revealed that nearly 20% of errors were due to inconsistent inclusion of punctuation in annotation. When we removed the inconsistency in punctuation, the F1 score of *CombinedTagger* increased from 0.73 to 0.79, which was statistically significant (*t* test,  $P=.001$ ). Although the missed pronoun annotations of *AE* and *medication* can be fixed readily, they also contributed to the lower performance of the tagger.

### Data Sparseness

Data sparseness is a common problem and the major cause of poor performance. The performance of AETagger was positively

correlated with the size of the annotated data for each named entity (a Pearson correlation coefficient of 0.64). In the cases of *frequency*, *duration*, *OSSD*, and *treatment* entities, data was very sparse (Table 1) and taggers showed low performance on these named entities. In addition to low performance, data sparseness also contributed to a higher standard deviation (Table 3). When the training data incorporate instances of a named entity but the testing data do not, the precision decreases. When the training data misses instances of a named entity but the testing data do not, then recall suffers.

### Learning Features

To further understand the contribution of learning features on the performance of AETagger, we first trained the tagger with all the features and used it as a baseline system (*CombinedTagger*). We then removed each feature category one at a time. Table 5 shows the performance of taggers with each feature category removed. Consistent with Table 3, the results show that each feature contributed to the performance differently.

**Table 5.** The precision, recall, and F1 score of Taggers with feature categories removed one at a time on each of the four annotated data sets.

Tagger	AnnPhy			AnnLing			Combined			Tie		
	Mean (SD)			Mean (SD)			Mean (SD)			Mean (SD)		
	F1	Preci- sion	Recall	F1	Preci- sion	Recall	F1	Preci- sion	Recall	F1	Preci- sion	Recall
All features	0.67 (0.09)	0.77 (0.10)	0.62 (0.09)	0.74 (0.08)	0.81 (0.07)	0.68 (0.09)	0.73 (0.08)	0.81 (0.10)	0.66 (0.07)	0.65 (0.08)	0.71 (0.08)	0.60 (0.09)
No affix features	0.68 (0.09)	0.76 (0.10)	0.62 (0.09)	0.71 (0.10)	0.78 (0.07)	0.65 (0.11)	0.71 (0.09)	0.79 (0.11)	0.64 (0.09)	0.64 (0.08)	0.70 (0.08)	0.60 (0.08)
No connective features	0.69 (0.09)	0.77 (0.10)	0.62 (0.09)	0.74 (0.08)	0.81 (0.06)	0.69 (0.09)	0.73 (0.08)	0.81 (0.10)	0.66 (0.07)	0.65 (0.08)	0.71 (0.08)	0.60 (0.09)
No morphological features	0.69 (0.09)	0.78 (0.10)	0.62 (0.09)	0.73 (0.08)	0.81 (0.06)	0.66 (0.09)	0.73 (0.08)	0.82 (0.10)	0.66 (0.07)	0.65 (0.08)	0.72 (0.08)	0.60 (0.08)
No negation and hedge features	0.68 (0.09)	0.77 (0.10)	0.62 (0.09)	0.74 (0.08)	0.81 (0.07)	0.68 (0.09)	0.72 (0.08)	0.81 (0.10)	0.65 (0.07)	0.64 (0.09)	0.71 (0.09)	0.59 (0.09)
No semantic features	0.67 (0.08)	0.77 (0.08)	0.60 (0.09)	0.74 (0.08)	0.82 (0.05)	0.68 (0.10)	0.71 (0.08)	0.80 (0.09)	0.64 (0.08)	0.64 (0.08)	0.71 (0.07)	0.59 (0.08)
No syntactical features	0.68 (0.09)	0.77 (0.10)	0.61 (0.09)	0.73 (0.08)	0.80 (0.07)	0.68 (0.09)	0.71 (0.09)	0.80 (0.11)	0.64 (0.08)	0.64 (0.08)	0.70 (0.09)	0.58 (0.09)

## Discussion

### Principal Findings

Our results show that medication and adverse events can be reliably annotated (Cohen's kappa value of .64-.95 IAA as shown in Table 1) in the FAERS narratives. Many named entities (eg, *indication*) that had shown low annotation agreements in the i2b2 challenge [63] had good annotation agreements in our dataset. The improvements were attributed to improved annotation guidelines and the quality and domain specificities of the FAERS narratives.

With a good IAA, we still found room to further improve the annotation guideline. For example, our error analyses (Figure 3) show that inconsistencies were introduced by annotation boundary; therefore it can be further refined. Although *medication* had the highest IAA (.95), our analysis (Table 4) found that the inconsistency in *medication* was introduced by whether instances like “fluids” could be considered as medication or not. In the future, we may separate *medication* into two classes: *strict medication* and *unstrict medication*. The names and mentions of all drugs appearing in the United States Pharmacopeia will belong to *strict medication*; any substances or chemicals—including oxygen, fluids, drinks, and others—given to patients during the treatment will be classified as *unstrict medication*. Refining the guideline to annotate previous and potential AEs like “allergy” (Example 5) may further reduce the inconsistency.

We explored various ML methods and compared them with a baseline string matching and Metamap-based systems to assess the complexity of the task. The CRF-based tagger had the best performance. Further analyses of the CRF tagger found that data sparseness affected the taggers' performance (Figure 3). For example, the standard deviation of *treatment* is high because we found that the testing data did not incorporate *treatment*

instances. Similar behavior was also observed for other sparse entities (Table 3).

Using the best performing ML technique, we explored a variety of features (Table 2 and Table 3). The features had a mixed effect on the performance of the taggers and the combination of all the features improved overall performance slightly. This suggests the robustness of the default features for CRFs. Since most of the features were extracted automatically (eg, negation, hedge cues, and discourse connectives were extracted using the taggers [77,78] and parser [79] we developed), the accuracy of the extracted features played an important role in overall performance of the tagger. To avoid the noise introduced by automatic feature extraction, one may explore the features manually annotated such as POS in the PennTree Bank [80]. This is, however, expensive. An alternative is to further improve the performance of the BioNLP systems for feature extraction.

Throughout the study, we found that additional features may be further included. For example, we observed that *OSSD* most often appeared in the patient's medical history. We therefore added a feature representing patient history and found that the performance of the *CombinedTagger* on *OSSD* increased 1.2% (results not reported in the Result section), although the increase was not statistically significant ( $t$  test,  $P=.25$ ).

### Limitations

Our study has limitations. The AETaggers were trained on the FAERS corpus we constructed. Like any other NLP system, the performance of the tagger on other types of EMRs can vary based on the structure and content of the narrative text. On the other hand, since our selection of the FAERS corpus was through a random process, we speculate that the data is representative. Although the taggers performed well, the training and evaluation was based on a relatively small training dataset. In the future, we would increase the size of the training data and explore other semi-supervised machine learning approaches to further improve the performance.

## Conclusions

In this study, we developed an annotation guideline for medication and adverse event information from the FAERS narratives, our annotation of 122 FAERS narratives (a total of approximately 23,000 tokens) showed a reliable inter-rater annotation agreement (an overall kappa of .82). We then developed machine learning models for automatically extracting medication and adverse event information from the FAERS narratives. We explored utilizing different learning features in

the machine learning models. The results showed that features such as syntactic, semantic, morphological, and affix improved the performance and the best performing system had an overall F1 score of 0.73. In the future, we would like to refine further the annotation guideline, explore additional features and increase the annotation size to improve system performance. We will also explore approaches for normalizing the entities by mapping them to standard terminologies like MedDRA and identify the causal relation between a medication and an adverse event.

## Acknowledgments

We would like to acknowledge Dr Dennis W Raisch for providing the FAERS data from FDA and Ricky J Sethi for providing valuable comments on the manuscript. The research reported in this publication was supported in part by the National Institutes of Health (NIH) the National Institute of General Medical Sciences under award number 5R01GM095476 and the National Center for Advancing Translational Sciences under award number UL1TR000161. The content is solely the responsibility of the authors and does not necessarily represent the official views of NIH.

## Conflicts of Interest

None declared.

## Authors' Contributions

Conceived and designed the experiments: BPR HY. Performed Annotation: ZF NF SB. Performed the experiments: BPR. Obtained the data: DW. Analyzed the data: BPR HY. Wrote the paper: BPR HY.

## Multimedia Appendix 1

Annotation Guideline.

[[PDF File \(Adobe PDF File\), 183KB-Multimedia Appendix 1](#)]

## References

1. Kohn LT, Corrigan JM, Donaldson MS. To err is human: building a safer health system. In: To err is human: building a safer health system. Washington, D.C: National Academy Press; 2000.
2. E 2 A Clinical Safety Data Management: Definitions and Standards for Expedited Reporting - WC500002749. pdf URL: [http://www.ema.europa.eu/docs/en\\_GB/document\\_library/Scientific\\_guideline/2009/09/WC500002749.pdf](http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002749.pdf) [accessed 2013-10-01] [WebCite Cache ID 6K2qiistb]
3. Bates DW, Cullen DJ, Laird N, Petersen LA, Small SD, Servi D, et al. Incidence of adverse drug events and potential adverse drug events. Implications for prevention. ADE Prevention Study Group. JAMA 1995 Jul 5;274(1):29-34. [Medline: [7791255](#)]
4. Nebeker JR, Barach P, Samore MH. Clarifying adverse drug events: a clinician's guide to terminology, documentation, and reporting. Ann Intern Med 2004 May 18;140(10):795-801. [Medline: [15148066](#)]
5. Forster AJ, Halil RB, Tierney MG. Pharmacist surveillance of adverse drug events. Am J Health Syst Pharm 2004 Jul 15;61(14):1466-1472. [Medline: [15332694](#)]
6. Thomas EJ, Studdert DM, Burstin HR, Orav EJ, Zeena T, Williams EJ, et al. Incidence and types of adverse events and negligent care in Utah and Colorado. Med Care 2000 Mar;38(3):261-271. [Medline: [10718351](#)]
7. Leape LL, Brennan TA, Laird N, Lawthers AG, Localio AR, Barnes BA, et al. The nature of adverse events in hospitalized patients. Results of the Harvard Medical Practice Study II. N Engl J Med 1991 Feb 7;324(6):377-384. [doi: [10.1056/NEJM199102073240605](#)] [Medline: [1824793](#)]
8. Classen DC, Pestotnik SL, Evans RS, Lloyd JF, Burke JP. Adverse drug events in hospitalized patients. Excess length of stay, extra costs, and attributable mortality. JAMA 1997;277(4):301-306. [Medline: [9002492](#)]
9. Cullen DJ, Sweitzer BJ, Bates DW, Burdick E, Edmondson A, Leape LL. Preventable adverse drug events in hospitalized patients: a comparative study of intensive care and general care units. Crit Care Med 1997 Aug;25(8):1289-1297. [Medline: [9267940](#)]
10. Cullen DJ, Bates DW, Small SD, Cooper JB, Nemeskal AR, Leape LL. The incident reporting system does not detect adverse drug events: a problem for quality improvement. Jt Comm J Qual Improv 1995 Oct;21(10):541-548. [Medline: [8556111](#)]
11. Bates DW, Spell N, Cullen DJ, Burdick E, Laird N, Petersen LA, et al. The costs of adverse drug events in hospitalized patients. Adverse Drug Events Prevention Study Group. JAMA 1997;277(4):307-311. [Medline: [9002493](#)]

12. FDA Adverse Events Reporting System (FAERS) > FDA Adverse Event Reporting System (FAERS) Statistics. Available. Online URL: <http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/ucm070093.htm> [accessed 2013-10-01] [WebCite Cache ID 6K2rK4ifG]
13. Tatonetti NP, Ye PP, Daneshjou R, Altman RB. Data-driven prediction of drug effects and interactions. *Sci Transl Med* 2012 Mar 14;4(125):125ra31 [FREE Full text] [doi: [10.1126/scitranslmed.3003377](https://doi.org/10.1126/scitranslmed.3003377)] [Medline: [22422992](https://pubmed.ncbi.nlm.nih.gov/22422992/)]
14. Hiatt HH, Barnes BA, Brennan TA, Laird NM, Lawthers AG, Leape LL, et al. A study of medical injury and medical malpractice. *N Engl J Med* 1989 Aug 17;321(7):480-484. [doi: [10.1056/NEJM198908173210725](https://doi.org/10.1056/NEJM198908173210725)] [Medline: [2761585](https://pubmed.ncbi.nlm.nih.gov/2761585/)]
15. Brennan TA, Leape LL, Laird NM, Hebert L, Localio AR, Lawthers AG, et al. Incidence of adverse events and negligence in hospitalized patients. Results of the Harvard Medical Practice Study I. *N Engl J Med* 1991 Feb 7;324(6):370-376. [doi: [10.1056/NEJM199102073240604](https://doi.org/10.1056/NEJM199102073240604)] [Medline: [1987460](https://pubmed.ncbi.nlm.nih.gov/1987460/)]
16. Jha AK, Kuperman GJ, Teich JM, Leape L, Shea B, Rittenberg E, et al. Identifying adverse drug events: development of a computer-based monitor and comparison with chart review and stimulated voluntary report. *J Am Med Inform Assoc* 1998 Jun;5(3):305-314 [FREE Full text] [Medline: [9609500](https://pubmed.ncbi.nlm.nih.gov/9609500/)]
17. Murff HJ, Forster AJ, Peterson JF, Fiskio JM, Heiman HL, Bates DW. Electronically screening discharge summaries for adverse medical events. *J Am Med Inform Assoc* 2003 Aug;10(4):339-350 [FREE Full text] [doi: [10.1197/jamia.M1201](https://doi.org/10.1197/jamia.M1201)] [Medline: [12668691](https://pubmed.ncbi.nlm.nih.gov/12668691/)]
18. Hwang SH, Lee S, Koo HK, Kim Y. Evaluation of a computer-based adverse-drug-event monitor. *Am J Health Syst Pharm* 2008 Dec 1;65(23):2265-2272. [doi: [10.2146/ajhp080122](https://doi.org/10.2146/ajhp080122)] [Medline: [19020194](https://pubmed.ncbi.nlm.nih.gov/19020194/)]
19. Tinoco A, Evans RS, Staes CJ, Lloyd JF, Rothschild JM, Haug PJ. Comparison of computerized surveillance and manual chart review for adverse events. *J Am Med Inform Assoc* 2011 Aug;18(4):491-497 [FREE Full text] [doi: [10.1136/amiajnl-2011-000187](https://doi.org/10.1136/amiajnl-2011-000187)] [Medline: [21672911](https://pubmed.ncbi.nlm.nih.gov/21672911/)]
20. Gardner RM, Pryor TA, Warner HR. The HELP hospital information system: update 1998. *Int J Med Inform* 1999 Jun;54(3):169-182. [Medline: [10405877](https://pubmed.ncbi.nlm.nih.gov/10405877/)]
21. Hazlehurst B, Frost HR, Sittig DF, Stevens VJ. MediClass: A system for detecting and classifying encounter-based clinical events in any electronic medical record. *J Am Med Inform Assoc* 2005 Oct;12(5):517-529 [FREE Full text] [doi: [10.1197/jamia.M1771](https://doi.org/10.1197/jamia.M1771)] [Medline: [15905485](https://pubmed.ncbi.nlm.nih.gov/15905485/)]
22. Hazlehurst B, Naleway A, Mullooly J. Detecting possible vaccine adverse events in clinical notes of the electronic medical record. *Vaccine* 2009 Mar 23;27(14):2077-2083. [doi: [10.1016/j.vaccine.2009.01.105](https://doi.org/10.1016/j.vaccine.2009.01.105)] [Medline: [19428833](https://pubmed.ncbi.nlm.nih.gov/19428833/)]
23. Wang X, Hripcsak G, Markatou M, Friedman C. Active computerized pharmacovigilance using natural language processing, statistics, and electronic health records: a feasibility study. *J Am Med Inform Assoc* 2009 Jun;16(3):328-337 [FREE Full text] [doi: [10.1197/jamia.M3028](https://doi.org/10.1197/jamia.M3028)] [Medline: [19261932](https://pubmed.ncbi.nlm.nih.gov/19261932/)]
24. Haerian K, Varn D, Vaidya S, Ena L, Chase HS, Friedman C. Detection of pharmacovigilance-related adverse events using electronic health records and automated methods. *Clin Pharmacol Ther* 2012 Aug;92(2):228-234 [FREE Full text] [doi: [10.1038/clpt.2012.54](https://doi.org/10.1038/clpt.2012.54)] [Medline: [22713699](https://pubmed.ncbi.nlm.nih.gov/22713699/)]
25. Melton GB, Hripcsak G. Automated detection of adverse events using natural language processing of discharge summaries. *J Am Med Inform Assoc* 2005 Aug;12(4):448-457 [FREE Full text] [doi: [10.1197/jamia.M1794](https://doi.org/10.1197/jamia.M1794)] [Medline: [15802475](https://pubmed.ncbi.nlm.nih.gov/15802475/)]
26. Friedman C, Alderson PO, Austin JH, Cimino JJ, Johnson SB. A general natural-language text processor for clinical radiology. *J Am Med Inform Assoc* 1994 Apr;1(2):161-174 [FREE Full text] [Medline: [7719797](https://pubmed.ncbi.nlm.nih.gov/7719797/)]
27. Aramaki E, Miura Y, Tonoike M, Ohkuma T, Masuichi H, Waki K, et al. Extraction of adverse drug effects from clinical records. *Stud Health Technol Inform* 2010;160(Pt 1):739-743. [Medline: [20841784](https://pubmed.ncbi.nlm.nih.gov/20841784/)]
28. Yang C, Srinivasan P, Polgreen PM. Automatic adverse drug events detection using letters to the editor. *AMIA Annu Symp Proc* 2012;2012:1030-1039 [FREE Full text] [Medline: [23304379](https://pubmed.ncbi.nlm.nih.gov/23304379/)]
29. Gurulingappa H, Mateen-Rajput A, Toldo L. Extraction of potential adverse drug events from medical case reports. *J Biomed Semantics* 2012;3(1):15 [FREE Full text] [doi: [10.1186/2041-1480-3-15](https://doi.org/10.1186/2041-1480-3-15)] [Medline: [23256479](https://pubmed.ncbi.nlm.nih.gov/23256479/)]
30. Chee BW, Berlin R, Schatz B. Predicting adverse drug events from personal health messages. *AMIA Annu Symp Proc* 2011;2011:217-226 [FREE Full text] [Medline: [22195073](https://pubmed.ncbi.nlm.nih.gov/22195073/)]
31. Nikfarjam A, Gonzalez GH. Pattern mining for extraction of mentions of Adverse Drug Reactions from user comments. *AMIA Annu Symp Proc* 2011;2011:1019-1026 [FREE Full text] [Medline: [22195162](https://pubmed.ncbi.nlm.nih.gov/22195162/)]
32. White RW, Tatonetti NP, Shah NH, Altman RB, Horvitz E. Web-scale pharmacovigilance: listening to signals from the crowd. *J Am Med Inform Assoc* 2013 May 1;20(3):404-408 [FREE Full text] [doi: [10.1136/amiajnl-2012-001482](https://doi.org/10.1136/amiajnl-2012-001482)] [Medline: [23467469](https://pubmed.ncbi.nlm.nih.gov/23467469/)]
33. Shetty KD, Dalal SR. Using information mining of the medical literature to improve drug safety. *J Am Med Inform Assoc* 2011 Oct;18(5):668-674 [FREE Full text] [doi: [10.1136/amiajnl-2011-000096](https://doi.org/10.1136/amiajnl-2011-000096)] [Medline: [21546507](https://pubmed.ncbi.nlm.nih.gov/21546507/)]
34. Lindquist M. VigiBase, the WHO Global ICSR Database System: Basic Facts. *Therapeutic Innovation & Regulatory Science* September 2008 Sep;42(5):409-419. [doi: [10.1177/009286150804200501](https://doi.org/10.1177/009286150804200501)]
35. Hauben M, Madigan D, Gerrits CM, Walsh L, Van Puijenbroek EP. The role of data mining in pharmacovigilance. *Expert Opin Drug Saf* 2005 Sep;4(5):929-948. [doi: [10.1517/14740338.4.5.929](https://doi.org/10.1517/14740338.4.5.929)] [Medline: [16111454](https://pubmed.ncbi.nlm.nih.gov/16111454/)]
36. Hauben M, Bate A. Decision support methods for the detection of adverse events in post-marketing data. *Drug Discov Today* 2009 Apr;14(7-8):343-357. [doi: [10.1016/j.drudis.2008.12.012](https://doi.org/10.1016/j.drudis.2008.12.012)] [Medline: [19187799](https://pubmed.ncbi.nlm.nih.gov/19187799/)]

37. Harpaz R, DuMouchel W, Shah NH, Madigan D, Ryan P, Friedman C. Novel data-mining methodologies for adverse drug event discovery and analysis. *Clin Pharmacol Ther* 2012 Jun;91(6):1010-1021 [[FREE Full text](#)] [doi: [10.1038/clpt.2012.50](https://doi.org/10.1038/clpt.2012.50)] [Medline: [22549283](#)]
38. Bate A, Evans SJ. Quantitative signal detection using spontaneous ADR reporting. *Pharmacoepidemiol Drug Saf* 2009 Jun;18(6):427-436. [doi: [10.1002/pds.1742](https://doi.org/10.1002/pds.1742)] [Medline: [19358225](#)]
39. Almenoff JS, Pattishall EN, Gibbs TG, DuMouchel W, Evans SJ, Yuen N. Novel statistical tools for monitoring the safety of marketed drugs. *Clin Pharmacol Ther* 2007 Aug;82(2):157-166. [doi: [10.1038/sj.clpt.6100258](https://doi.org/10.1038/sj.clpt.6100258)] [Medline: [17538548](#)]
40. Evans SJ, Waller PC, Davis S. Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiol Drug Saf* 2001 Nov;10(6):483-486. [doi: [10.1002/pds.677](https://doi.org/10.1002/pds.677)] [Medline: [11828828](#)]
41. van Puijjenbroek EP, Bate A, Leufkens HG, Lindquist M, Orre R, Egberts AC. A comparison of measures of disproportionality for signal detection in spontaneous reporting systems for adverse drug reactions. *Pharmacoepidemiol Drug Saf* 2002 Feb;11(1):3-10. [doi: [10.1002/pds.668](https://doi.org/10.1002/pds.668)] [Medline: [11998548](#)]
42. Norén GN, Sundberg R, Bate A, Edwards IR. A statistical methodology for drug-drug interaction surveillance. *Stat Med* 2008 Jul 20;27(16):3057-3070. [doi: [10.1002/sim.3247](https://doi.org/10.1002/sim.3247)] [Medline: [18344185](#)]
43. Harpaz R, Vilar S, Dumouchel W, Salmasian H, Haerian K, Shah NH, et al. Combining signals from spontaneous reports and electronic health records for detection of adverse drug reactions. *J Am Med Inform Assoc* 2013 May 1;20(3):413-419 [[FREE Full text](#)] [doi: [10.1136/amiainjnl-2012-000930](https://doi.org/10.1136/amiainjnl-2012-000930)] [Medline: [23118093](#)]
44. Sakaeda T, Kadoyama K, Okuno Y. Adverse event profiles of platinum agents: data mining of the public version of the FDA adverse event reporting system, AERS, and reproducibility of clinical observations. *Int J Med Sci* 2011;8(6):487-491 [[FREE Full text](#)] [Medline: [21897761](#)]
45. Vilar S, Harpaz R, Chase HS, Costanzi S, Rabadan R, Friedman C. Facilitating adverse drug event detection in pharmacovigilance databases using molecular structure similarity: application to rhabdomyolysis. *J Am Med Inform Assoc* 2011 Dec;18 Suppl 1:i73-i80 [[FREE Full text](#)] [doi: [10.1136/amiainjnl-2011-000417](https://doi.org/10.1136/amiainjnl-2011-000417)] [Medline: [21946238](#)]
46. Norén GN, Bate A, Orre R, Edwards IR. Extending the methods used to screen the WHO drug safety database towards analysis of complex associations and improved accuracy for rare events. *Stat Med* 2006 Nov 15;25(21):3740-3757. [doi: [10.1002/sim.2473](https://doi.org/10.1002/sim.2473)] [Medline: [16381072](#)]
47. Hochberg AM, Hauben M, Pearson RK, O'Hara DJ, Reisinger SJ, Goldsmith DI, et al. An evaluation of three signal-detection algorithms using a highly inclusive reference event database. *Drug Saf* 2009;32(6):509-525. [Medline: [19459718](#)]
48. Bate A, Lindquist M, Edwards IR, Orre R. A data mining approach for signal detection and analysis. *Drug Saf* 2002;25(6):393-397. [Medline: [12071775](#)]
49. Bate A. Bayesian confidence propagation neural network. *Drug Saf* 2007;30(7):623-625. [Medline: [17604417](#)]
50. DuMouchel W, Pregibon D. Empirical bayes screening for multi-item associations. In: *KDD-2001*. New York, NY, USA: Association for Computing Machinery; 2001 Presented at: Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 26-29, 2001; San Francisco, CA, USA.
51. Bate A, Lindquist M, Edwards IR, Olsson S, Orre R, Lansner A, et al. A Bayesian neural network method for adverse drug reaction signal generation. *Eur J Clin Pharmacol* 1998 Jun;54(4):315-321. [Medline: [9696956](#)]
52. Genkin A, Lewis DD, Madigan D. Large-Scale Bayesian Logistic Regression for Text Categorization. *Technometrics* 2007 Aug;49(3):291-304. [doi: [10.1198/004017007000000245](https://doi.org/10.1198/004017007000000245)]
53. Caster O, Norén GN, Madigan D, Bate A. Large-scale regression-based pattern discovery: The example of screening the WHO global drug safety database. *Stat Anal Data Min* 2010;3(4):197-208. [doi: [10.1002/sam.10078](https://doi.org/10.1002/sam.10078)]
54. Solomon R, Dumouchel W. Contrast media and nephropathy: findings from systematic analysis and Food and Drug Administration reports of adverse effects. *Invest Radiol* 2006 Aug;41(8):651-660. [doi: [10.1097/01.rli.0000229742.54589.7b](https://doi.org/10.1097/01.rli.0000229742.54589.7b)] [Medline: [16829749](#)]
55. Rouane-Hacene M, Toussaint Y, Valtchev P. Mining Safety Signals in Spontaneous Reports Database Using Concept Analysis. In: *AIME Proceedings / Lecture Notes in Artificial Intelligence*. Berlin, Heidelberg: Springer; 2009 Presented at: 12th Conference on Artificial Intelligence in Medicine in Europe (AIME); July 18-22, 2009; Verona, Italy p. 285-294.
56. Harpaz R, Chase HS, Friedman C. Mining multi-item drug adverse effect associations in spontaneous reporting systems. *BMC Bioinformatics* 2010;11 Suppl 9:S7 [[FREE Full text](#)] [doi: [10.1186/1471-2105-11-S9-S7](https://doi.org/10.1186/1471-2105-11-S9-S7)] [Medline: [21044365](#)]
57. Harpaz R, Perez H, Chase HS, Rabadan R, Hripcsak G, Friedman C. Biclustering of adverse drug events in the FDA's spontaneous reporting system. *Clin Pharmacol Ther* 2011 Feb;89(2):243-250 [[FREE Full text](#)] [doi: [10.1038/clpt.2010.285](https://doi.org/10.1038/clpt.2010.285)] [Medline: [21191383](#)]
58. Ball R, Botsis T. Can network analysis improve pattern recognition among adverse events following immunization reported to VAERS? *Clin Pharmacol Ther* 2011 Aug;90(2):271-278. [doi: [10.1038/clpt.2011.119](https://doi.org/10.1038/clpt.2011.119)] [Medline: [21677640](#)]
59. Harpaz R, DuMouchel W, LePendu P, Bauer-Mehren A, Ryan P, Shah NH. Performance of pharmacovigilance signal-detection algorithms for the FDA adverse event reporting system. *Clin Pharmacol Ther* 2013 Jun;93(6):539-546. [doi: [10.1038/clpt.2013.24](https://doi.org/10.1038/clpt.2013.24)] [Medline: [23571771](#)]
60. Yom-Tov E, Gabrilovich E. Postmarket drug surveillance without trial costs: discovery of adverse drug reactions through large-scale analysis of web search queries. *J Med Internet Res* 2013;15(6):e124 [[FREE Full text](#)] [doi: [10.2196/jmir.2614](https://doi.org/10.2196/jmir.2614)] [Medline: [23778053](#)]

61. RADAR (Research on Adverse Drug events And Reports). Available. Online URL: <http://www.feinberg.northwestern.edu/news/2013/07/radar.html> [accessed 2013-10-01] [WebCite Cache ID 6K2rG1Z3L]
62. Uzuner O, Solti I, Cadag E. Extracting medication information from clinical text. J Am Med Inform Assoc 2010 Oct;17(5):514-518 [FREE Full text] [doi: [10.1136/jamia.2010.003947](https://doi.org/10.1136/jamia.2010.003947)] [Medline: [20819854](https://pubmed.ncbi.nlm.nih.gov/20819854/)]
63. Li Z, Liu F, Antieau L, Cao Y, Yu H. Lancet: a high precision medication event extraction system for clinical text. J Am Med Inform Assoc 2010 Oct;17(5):563-567 [FREE Full text] [doi: [10.1136/jamia.2010.004077](https://doi.org/10.1136/jamia.2010.004077)] [Medline: [20819865](https://pubmed.ncbi.nlm.nih.gov/20819865/)]
64. Knowtator. URL: <http://knowtator.sourceforge.net/> [accessed 2014-05-06] [WebCite Cache ID 6PNDuSsKi]
65. Protégé. URL: <http://protege.stanford.edu/> [accessed 2014-05-06] [WebCite Cache ID 6PNEHutVo]
66. Cohen J. A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement 1960;20(1). [doi: [10.1177/001316446002000104](https://doi.org/10.1177/001316446002000104)]
67. John L, Andrew M, Fernando C N P. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. 2001 Presented at: 18th International Conference on Machine Learning (ICML); 2001; Williams College, Williamstown, MA, USA p. 282-289.
68. Weka 3 - Data Mining with Open Source Machine Learning Software in Java. Available. Online URL: <http://www.cs.waikato.ac.nz/ml/weka/> [accessed 2013-10-01] [WebCite Cache ID 6K2rPjB8r]
69. Settles B. ABNER: an open source tool for automatically tagging genes, proteins and other entity names in text. Bioinformatics 2005 Jul 15;21(14):3191-3192 [FREE Full text] [doi: [10.1093/bioinformatics/bti475](https://doi.org/10.1093/bioinformatics/bti475)] [Medline: [15860559](https://pubmed.ncbi.nlm.nih.gov/15860559/)]
70. Simon T, Koller D. Support vector machine active learning with applications to text classification. The Journal of Machine Learning Research 2002;2:45-66. [doi: [10.1162/153244302760185243](https://doi.org/10.1162/153244302760185243)]
71. Thorsten J. Text categorization with support vector machines: Learning with many relevant features. In: Machine learning: ECML-98: Lecture Notes in Computer Science Volume 1398, 1998, pp 137-142. Berlin: Springer; 1998 Presented at: 10th European Conference on Machine Learning (ECML); April 1998; Chemnitz, Germany.
72. Leaman R, Gonzalez G. BANNER: an executable survey of advances in biomedical named entity recognition. Pac Symp Biocomput 2008:652-663 [FREE Full text] [Medline: [18229723](https://pubmed.ncbi.nlm.nih.gov/18229723/)]
73. Charniak E, Johnson M. Coarse-to-fine n-best parsing and MaxEnt discriminative reranking. Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics - ACL '05 2005:173-180. [doi: [10.3115/1219840.1219862](https://doi.org/10.3115/1219840.1219862)]
74. McClosky D. Ph. Brown University: . D. thesis, , , 2010 Any domain parsing: Automatic domain adaptation for natural language parsing URL: <http://cs.brown.edu/research/pubs/theses/phd/2010/mcclosky.pdf> [accessed 2013-10-01] [WebCite Cache ID 6K3C5sx17]
75. Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. Proc AMIA Symp 2001:17-21 [FREE Full text] [Medline: [11825149](https://pubmed.ncbi.nlm.nih.gov/11825149/)]
76. MetaMap - A Tool For Recognizing UMLS Concepts in Text. URL: <http://metamap.nlm.nih.gov/> [accessed 2014-05-06] [WebCite Cache ID 6PNEdc6dD]
77. Agarwal S, Yu H. Biomedical negation scope detection with conditional random fields. J Am Med Inform Assoc 2010;17(6):696-701 [FREE Full text] [doi: [10.1136/jamia.2010.003228](https://doi.org/10.1136/jamia.2010.003228)] [Medline: [20962133](https://pubmed.ncbi.nlm.nih.gov/20962133/)]
78. Agarwal S, Yu H. Detecting hedge cues and their scope in biomedical text with conditional random fields. J Biomed Inform 2010 Dec;43(6):953-961 [FREE Full text] [doi: [10.1016/j.jbi.2010.08.003](https://doi.org/10.1016/j.jbi.2010.08.003)] [Medline: [20709188](https://pubmed.ncbi.nlm.nih.gov/20709188/)]
79. Ramesh BP, Prasad R, Miller T, Harrington B, Yu H. Automatic discourse connective detection in biomedical text. J Am Med Inform Assoc 2012 Oct;19(5):800-808 [FREE Full text] [doi: [10.1136/amiajnl-2011-000775](https://doi.org/10.1136/amiajnl-2011-000775)] [Medline: [22744958](https://pubmed.ncbi.nlm.nih.gov/22744958/)]
80. Marcus MP, Marcinkiewicz MA, Santorini B. Building a large annotated corpus of English: The Penn Treebank. Computational linguistics 1993;19(2).

## Abbreviations

- ABNER:** A Biomedical Named Entity Recognizer
- ADE:** adverse drug event
- AE:** adverse event
- CRF:** Conditional Random Fields
- EMR:** Electronic Medical Records
- FAERS:** FDA Adverse Event Reporting System
- FDA:** Food and Drug Administration
- IAA:** inter-annotator agreement
- MedDRA:** Medical Dictionary for Regulatory Activities
- ML:** machine learning
- NB:** Naïve Bayes
- NLP:** natural language processing
- OSSD:** Other Signs, Symptoms and Diseases
- POS:** part of speech
- SVM:** Support Vector Machines

*Edited by G Eysenbach; submitted 17.10.13; peer-reviewed by T Wadhwa, D Maclean, WC Su; comments to author 23.11.13; revised version received 10.12.13; accepted 10.12.13; published 27.06.14*

*Please cite as:*

*Polepalli Ramesh B, Belknap SM, Li Z, Frid N, West DP, Yu H*

*Automatically Recognizing Medication and Adverse Event Information From Food and Drug Administration's Adverse Event Reporting System Narratives*

*JMIR Med Inform 2014;2(1):e10*

URL: <http://medinform.jmir.org/2014/1/e10/>

doi: [10.2196/medinform.3022](https://doi.org/10.2196/medinform.3022)

PMID: [25600332](https://pubmed.ncbi.nlm.nih.gov/25600332/)

©Balaji Polepalli Ramesh, Steven M Belknap, Zuofeng Li, Nadya Frid, Dennis P West, Hong Yu. Originally published in JMIR Medical Informatics (<http://medinform.jmir.org>), 27.06.2014. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on <http://medinform.jmir.org/>, as well as this copyright and license information must be included.