UC Riverside

Title

Learning signals of adverse drug-drug interactions from the unstructured text of electronic health records.

Permalink

https://escholarship.org/uc/item/7cd017bx

Authors

Iyer, Srinivasan V
Lependu, Paea
Harpaz, Rave
et al.

Publication Date

2013

Peer reviewed

Learning Signals of Adverse Drug-Drug Interactions from the Unstructured Text of Electronic Health Records

Srinivasan V Iyer, Paea LePendu, Rave Harpaz, Anna Bauer-Mehren, Nigam H Shah
Stanford University, Stanford, CA

Abstract

Drug-drug interactions (DDI) account for 30% of all adverse drug reactions, which are the fourth leading cause of death in the US. Current methods for post marketing surveillance primarily use spontaneous reporting systems for learning DDI signals and validate their signals using the structured portions of Electronic Health Records (EHRs). We demonstrate a fast, annotation-based approach, which uses standard odds ratios for identifying signals of DDIs from the textual portion of EHRs directly and which, to our knowledge, is the first effort of its kind. We developed a gold standard of 1,120 DDIs spanning 14 adverse events and 1,164 drugs. Our evaluations on this gold standard using millions of clinical notes from the Stanford Hospital confirm that identifying DDI signals from clinical text is feasible (AUROC=81.5%). We conclude that the text in EHRs contain valuable information for learning DDI signals and has enormous utility in drug surveillance and clinical decision support.

Introduction

Electronic Health Record (EHR) systems are becoming highly prevalent today and have the potential[1] to add to the success of spontaneous reporting systems for post marketing surveillance of drugs, especially to identify risky drug interactions. Although most EHR based drug safety research relies on coded information, studies have established several advantages to using the clinical text. In this work, we apply data mining methods on the textual portion of EHRs to *learn* drug-drug interaction (DDI) signals, which, to our knowledge is the first study of its kind.

Methods

We use an ontology-based text processing workflow to tag a corpus of 9,078,736 time-stamped textual notes from the Stanford Hospital (STRIDE) with drug and disease concepts (1,164 drugs and 14 events) and create a concept timeline for each patient. We check for a *drug-drug-event* interaction signal, by comparing the odds of the event among patients on both drugs to that in patients on at most one of the drugs. We adjust the odds ratio for confounding by age, gender, race, drug count, disease count and note count. Finally, we build a gold standard of 560 known interactions and an equal number of random unknown interactions and use it to validate our method.

Results

Our method achieves an area under the Receiver Operator Characteristic curve of 81.5% on the gold standard, with a sensitivity of 37.98% at a specificity of 96.56% and a positive predictive value of 91.71%. We also calculate the prevalence of the event among patients on drug combinations representing true interactions (Table 1), and argue that such prevalence information can be a unique resource to prioritize known interactions.

Table 1. The drug combinations with the highest event prevalence in STRIDE, for each event. (Prev = #(Exposed and affected)/#(Exposed))

Adverse Event (#Patients)	Drug1	Drug2	a	b	Prev. (%)
Parkinsonian Symptoms (3541)	levodopa	lorazepam	176	235	42.82
Cardiac Arrhythmias (88555)	potassium chloride	lisinopril	1091	1615	40.32
Neutropenia (14322)	paclitaxel	trastuzumab	140	567	19.8
Bradycardia (22906)	amiodarone	metoprolol	796	3671	17.82
Hypoglycemia (11150)	glipizide	lisinopril	367	2160	14.52
Acute Renal Failure (32197)	hydrochlorothiazide	ibuprofen	884	8375	9.55
Hyperkalemia (4973)	potassium chloride	spironolactone	349	3471	9.14
Hyperglycemia (19189)	prednisone	salmeterol	379	4612	7.59
Nephrotoxicity (1460)	fluconazole	tacrolimus	85	1208	6.57
Pancytopenia (8718)	mercaptopurine	azathioprine	15	278	5.12
Hypokalemia (8405)	prednisone	salmeterol	222	4982	4.27
Serotonin Syndrome (674)	tramadol	duloxetine	57	1301	4.2
QT prolongation (1260)	amiodarone	ciprofloxacin	46	2487	1.82
Rhabdomyolysis (1378)	ciprofloxacin	simvastatin	50	5184	0.96

Discussion

Our work shows the feasibility of the use of EHR text to generate hypotheses about DDIs and to examine their real world prevalence, which could serve as an important step forward for Phase IV drug surveillance and meaningful use of EHRs. Having longitudinal coverage and being minimally affected by reporting and publicity bias, makes the prevalence information generated from EHRs extremely useful for Computerized Physician Order Entry (CPOE) systems to tackle the problem of alert fatigue; and in choosing combination therapy from among many drug combinations. Confounding, accuracy of concept recognition, detection of the drug interaction period and other limitations of this method will be addressed in future work.

References

1. Schuemie MJ, Coloma PM, Straatman H, et al. Using Electronic Health Care Records for Drug Safety Signal Detection: A Comparative Evaluation of Statistical Methods. *Medical care.* 2012.