UC Riverside

Title

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

Permalink

https://escholarship.org/uc/item/90h5n575

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

More than 400,000 preventable adverse drug reactions occur every year[1], each costing over $3500 with an increased hospital stay of over 3 days[2], so their discovery and prevention would address one of the leading causes of death in the US. We are also witnessing a rise in polypharmacy, which is the use of multiple concomitant drugs to treat medical conditions, with many people taking 3 or more drugs. In fact, one study estimates that 29.4% of elderly patients[3] are on 6 or more drugs. Drug interactions that lead to adverse reactions are a potentially avoidable[4] consequence of this practice, accounting for more than 30% of all drug reactions[5]; and their early detection is vital[6].

New drugs are usually tested for interactions with existing drugs before market approval using *in-vivo* and *in-vitro* methods[7]. However, owing to the sheer number of ways by which drugs can interact[8], it is infeasible and expensive to test for every kind of interaction. Also, many drug interactions manifest after a certain period of exposure and it takes several exposures for rare drug interactions to occur[9]. Therefore, post marketing surveillance is necessary to detect unanticipated interactions that occur when the drug is in use in the general population. The US Food and Drug Administration (FDA) enables such surveillance by the active monitoring of spontaneous reporting systems (SRS) such as Adverse Event Reporting System (AERS) and similarly, the World Health Organization's VigiBase. There have been several studies[10-13] that have successfully inferred drug interactions from these sources, overcoming problems of reporting biases[14] and duplicate reporting[15].

Electronic health records (EHRs) complement these existing SRSs, providing a source of observational data without the same biases. Initiatives like the Observational Medical Outcomes Partnership (OMOP) in the US and the Exploring and Understanding Adverse Drug Reactions (EU-ADR) project in Europe are focusing on building EHR based surveillance systems. These projects mainly utilize the structured diagnosis and prescription data of the EHRs for identifying single drug adverse reactions. Most efforts aimed at finding drug interactions use reported sources for signal detection and use EHRs as a means of validation. For instance, Tatonetti et. al.[10] found 171 new drug interactions from AERS and used the EHRs at Stanford to validate them. Another study by Duke et. al.[16] mined MEDLINE abstracts for hypothesis generation and validated the signals on EHRs.

However, in addition to structured data, EHRs contain rich information in the unstructured notes and reports taken by doctors, nurses and other practitioners. By ignoring the unstructured text, we could be missing a substantial portion of adverse events[17]. Many studies[18] have shown that coded information like ICD-9 are inadequate to accurately build patient cohorts and there is a considerable advantage[19] in using the unstructured clinical text of EHRs. We argue that such an advantage would also extend to drug safety signal detection. Indeed, there is already some work[20,21] demonstrating the discovery of the adverse event profiles for single drugs using unstructured notes.

Therefore, given increasing adoption and access to medical records for research, we expect efforts to shift more toward directly mining EHRs for signal generation with an increased attention on the use of unstructured data[22]. In this paper, we apply data mining methods on the textual portion of EHRs to *learn* signals of drug-drug interactions. To our knowledge, this is the first study of its kind.

Methods

Preparation of Gold Standard

In order to ensure reliability, we limit our study to 1,164 drug ingredients common to three sources: the UMLS RxNORM ontology (4,993 ingredients), DrugBank[23] (6,711 drugs) and the Anatomical Therapeutic Chemical

Classification (ATC) ontology (4,406 drugs). We use known interactions from DrugBank and the Medi-Span® Drug Therapy Monitoring System™ (Wolters Kluwer Health, Indianapolis, IN) as positive interactions in our gold standard. Each of these databases provides a textual monograph describing the interaction. The short monographs follow a basic template, so we use regular expressions to extract and manually validate *drug-drug-event* relations that indicate known interactions. Our set of positive interactions consists of 591 different drug ingredients and 14 distinct events. For estimating false discovery, we simulate negative examples by generating random drug-drug-event tuples, and removing any known interactions according to DrugBank, Medi-Span or Drugs.com[24]. We also remove pairs for which the adverse event is an indication (from Medi-Span, DrugBank, Drugs.com, UMLS and SIDER[25]) for either drug individually.

Annotation of Electronic Health Records

We use the Stanford Translational Research Integrated Database Environment (STRIDE) dataset comprising 9,078,736 textual notes corresponding to 1,044,979 patients. There are 565,898 patients (53% female) in STRIDE, with at least one drug or event from our selection mentioned in their records. 857 out of the 1,164 drugs actually appeared at least once in our dataset.
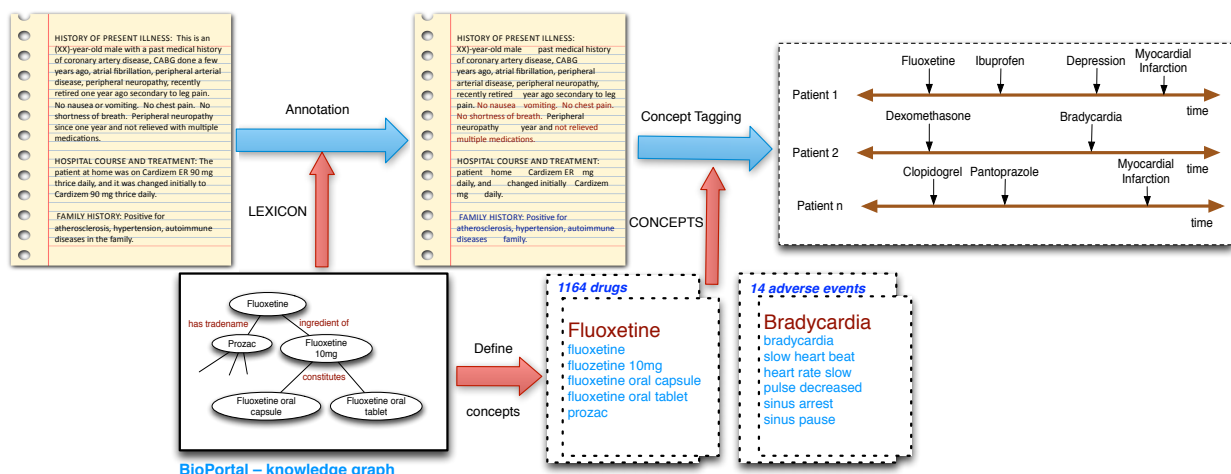
Figure 1. The Annotator Workflow – The annotator program uses a lexicon derived from biomedical ontologies and transforms clinical notes into a set of patient timelines with markers for the drug and disease mentions.

As described in our previous work[26], we define drug and event concepts as sets of terms derived from biomedical ontologies. For drugs, we include trade names and other forms of the drug from the RxNORM ontology. To improve our precision, we remove terms that occur in common English usage[27], followed by manual curation. The average size of the set of terms used to identify a drug is 8.3 terms. We then use a fast text annotator to tag clinical notes with these concepts and order them by the note's timestamp, thus forming a set of patient timelines. The tool also takes into account negation and family history contextual cues to reduce false attribution of concepts. We focus our study on 14 adverse events based on existing literature (a list published by Trifiro et al[28]), their presence in our gold standard, their prevalence in STRIDE and our ability to successfully detect their presence from EHRs.

Identification of DDI signals

In traditional drug safety surveillance with single drugs, odds ratios (OR) are computed using a 2-by-2 contingency table to detect signals [Figure 2]. For our method, the *exposed group* represents patients who have taken both drugs and the *comparison group* represents patients who have taken at most one drug. Based on the ordering of the drug and event mentions, we classify each patient into one of the cells of the contingency table [Figure 3]. Any drugs that appear after the first occurrence of the event are ignored.

We use the ratio of the odds of getting the adverse event in the exposed group and in the comparison groups as a measure of the strength of the interaction.

Figure 2. 2x2 contingency table for a drug-drug-event association

$$OR = \frac{a/b}{c/d}$$

We calculate 95% confidence intervals for the odds ratio and signal an interaction if the lower bound of the confidence interval is greater than a threshold, which is picked using ROC curves depending on the desired sensitivity and specificity. Similar to approaches used in SRSs[29] to reduce the effect of confounding, we use propensity score matching (PSM) to match at most 10 control patients for every case patient, and then generate an adjusted OR. In addition to matching on average age at the time of exposure, gender, and race, we also match on note count, drug count and disease count, which serve as a proxy for the overall health of the patient[22].

Patient Timeline	Cell in 2x2 table	Reason
	d	No drug taken and event not encountered
D	d	Single drug taken and no event encountered
E	c	Event encountered without taking either drug
D E	c	Single drug precedes adverse event
D1 D2	b	Drugs possibly interact, but adverse event not encountered
D1 D2 E	a	Drugs possibly interact, followed by adverse event

Figure 3. Assignment of patients to various cells in the 2x2 contingency table. The portion of the timeline after the first occurrence of the event is ignored. D=drug, E=event.

Results

Prevalence of known interactions

DrugBank contains 10,906 distinct drug interactions, of which 7,400 corresponded to our list of drugs. Similarly, of 40,475 drug interactions from Medi-Span, 15,621 corresponded to our list of drugs. Together, these formed a set of 19,356 interactions, which resulted in 6,845 distinct drug-drug-event tuples constituting the positive examples in our gold standard. We calculated ORs, using STRIDE data, for 560 true examples with sufficient support, and also for an equal number (560) of random negative examples [see Methods]. We calculated the prevalence of each event among patients on drug combinations representing true interactions from the 2x2 contingency table [see Table 1]. This gold standard of 1,120 interactions, together with the prevalence information for the 560

Table 1. The drug combinations with the highest event prevalence (Prev. = a/(a+b)) in STRIDE, for each event.

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

Table 2. Performance at selected values of the threshold

Threshold	TP	TN	FP	FN	Sensitivity(%)	Specificity(%)	PPV(%)
1.5	177	450	16	289	37.98	96.56	91.71
1.3	205	442	24	261	43.99	94.85	89.52
1.0	244	423	43	222	52.36	90.77	85.02

true interactions forms a unique resource that can be used to test the performance of other methods for mining DDIs and to prioritize known interactions [see Discussion].

Evaluation

Using an odds ratio [see Methods] as a measure of interaction, and taking the lower bound of the 95% confidence interval to account for variance, we obtained a sensitivity of 37.86% at a specificity of 86.79% and a positive predictive value (PPV) of 74.13%. We found that the performance of the method was below average for Acute Renal Failure, Nephrotoxicity, Hypokalemia and Hyperglycemia [see Discussion]. On removing these events, and following adjustment by PSM, our specificity improved to 96.56% with a PPV of 91.71% (threshold=1.5)[see Table 2]. Figure 4 shows a Receiver Operator Characteristic (ROC) curve showing all possible values of sensitivity and specificity that can be achieved by varying the cutoff threshold. Overall performance suggests that the unstructured text contains relevant data for drug interaction detection, achieving 81.5% area under the ROC curve for signaling known DDIs.

Discussion

The importance of the early identification of DDIs is paramount and most of the existing methods use SRS databases to learn interactions, and use coded information present in EHRs for validation and prioritization. We have shown that it is possible to detect signals of DDIs directly from the unstructured text of EHRs using text mining methods and odds ratios. Since our methods are easy to implement and run rapidly, we could deploy them as an active monitoring tool for detecting unknown interactions for new and existing drugs, thus serving as an important step forward for Phase IV surveillance of drugs and meaningful use of EHRs.

Unlike SRS systems, EHRs have good longitudinal coverage of patient history, a larger number of measured covariates and are less affected by reporting and publicity biases and thus, can provide a more accurate measure of the prevalence of a particular drug interaction in the real world. Augmenting existing drug interaction databases with this prevalence information[30] would in itself be a major step forward in prioritizing interaction alerts in Computerized Physician Order Entry (CPOE) systems; where at present 49% to 96% of all alerts are overridden[31] owing to alert fatigue. Furthermore, this prevalence information could help choose between drugs used in combination therapy. For example, a study found that use of Tacrolimus with statin therapy is safer than use of Cyclosporine A and reduces the risk of rhabdomyolysis[32].

We tested our method on a gold standard comprising known drug interactions as positive examples, and randomly generated interactions as negative examples. Most interaction studies[33,34] use similar techniques to build a gold standard and this aids in comparison of results. Indeed, some events did not perform well in this setting, and this could be due in part to the gold

Figure 4. Receiver Operator Characteristic (ROC) curve showing sensitivity and specificity levels that can be achieved by varying the threshold.

standard or to the methods themselves. The databases from which the known interactions are derived are not exhaustive and this may introduce certain biases. In future work, we propose to make a distinction between interactions where the event is unrelated to the drugs, and interactions in which either of the drugs is already known to be associated with the event. It is possible that these simple methods work for some kinds of interactions and more sophisticated methods are required for others.

Our work has several limitations. Since our workflow relies on text annotation methods to recognize concepts, its performance depends on the accuracy of concept recognition in EHRs. In this study, we use frequency based methods[35] and manual curation to remove ambiguous terms corresponding to concepts, at the risk of reducing our sensitivity. However, we have found that some concepts are hard to detect in EHRs even after such filtering. Another limitation of the current method is that it fails to identify DDIs that are dependent on the dosage of the drugs in question. Resolving these issues requires more advanced natural language processing methods[36], which proportionally increases the computation time required. Also, we do not put a restriction on the time window in which drugs interact and drug exposures spaced far away in time may cause several false associations. Lastly, our current methods may suffer from confounding by other variables, which we aim to resolve in future work by adjusting for confounding by co-morbidities and co-prescriptions.

Conclusion

In this paper, we demonstrate the feasibility of using the textual portion of EHRs for learning signals of DDIs and to estimate the prevalence of existing interactions. In the process, we created a gold standard of DDIs that may be useful for detailed characterization of future methods. Owing to the use of simple and fast text mining methods, we are able to learn DDI signals on a database containing millions of text notes without the use of extensive computing infrastructure. To the best of our knowledge, this is the first study to use the textual notes from a clinical data warehouse to generate hypotheses about drug interactions and examine the prevalence of DDIs.

References

1. Bates DW. Preventing medication errors: a summary. *Am. J. Health. Syst. Pharm.* 2007;64(14 Supplement 9):S3-S9.
2. Hug BL, Keohane C, Seger DL, Yoon C, Bates DW. The costs of adverse drug events in community hospitals. *Jt. Comm. J. Qual. Patient Saf.* 2012;38(3):120-126.
3. Bushardt RL, Massey EB, Simpson TW, Ariail JC, Simpson KN. Polypharmacy: misleading, but manageable. *Clin. Interv. Aging.* 2008;3(2):383-389.
4. Strandell J, Bate A, Lindquist M, Edwards IR. Drug-drug interactions - a preventable patient safety issue? *Br. J. Clin. Pharmacol.* 2008;65(1):144-146.
5. Pirohamed M. *Drug Interactions of Clinical Importance*. London: Chapman and Hall; 1998.
6. Huang SM, Temple R, Throckmorton DC, Lesko LJ. Drug interaction studies: study design, data analysis, and implications for dosing and labeling. *Clin. Pharmacol. Ther.* 2007;81(2):298-304.
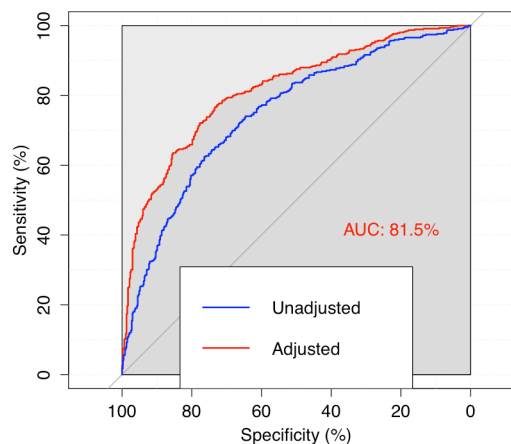
7. Zhang L, Zhang YD, Zhao P, Huang SM. Predicting drug-drug interactions: an FDA perspective. *The AAPS journal.* 2009;11(2):300-306.
8. Kroner BA. Common Drug Pathways and Interactions. *Diabetes Spectrum.* 2002;15(4):249-255.
9. Triaridis S, Tsiropoulos G, Rachovitsas D, Psillas G, Vital V. Spontaneous haematoma of the pharynx due to a rare drug interaction. *Hippokratia.* 2009;13(3):175-177.
10. Tatonetti NP, Fernald GH, Altman RB. A novel signal detection algorithm for identifying hidden drug-drug interactions in adverse event reports. *J. Am. Med. Inform. Assoc.* 2012;19(1):79-85.
11. van Puijenbroek EP, Egberts AC, Heerdink ER, Leufkens HG. Detecting drug-drug interactions using a database for spontaneous adverse drug reactions: an example with diuretics and non-steroidal anti-inflammatory drugs. *Eur. J. Clin. Pharmacol.* 2000;56(9-10):733-738.
12. Thakrar BT, Grundschober SB, Doessegger L. Detecting signals of drug-drug interactions in a spontaneous reports database. *Br. J. Clin. Pharmacol.* 2007;64(4):489-495.
13. Harpaz R, Chase HS, Friedman C. Mining multi-item drug adverse effect associations in spontaneous reporting systems. *BMC bioinformatics.* 2010;11 Suppl 9:S7.
14. McAdams M, Staffa J, Dal Pan G. Estimating the extent of reporting to FDA: a case study of statin-associated rhabdomyolysis. *Pharmacoepidemiol Drug Saf.* 2008;17(3):229-239.
15. Noren GN, Orre R, Bate A, Edwards IR. Duplicate detection in adverse drug reaction surveillance. *Data Min. Knowl. Discov.* 2007;14(3):305-328.
16. Duke JD, Han X, Wang Z, et al. Literature Based Drug Interaction Prediction with Clinical Assessment Using Electronic Medical Records: Novel Myopathy Associated Drug Interactions. *PLoS Comput Biol.* 2012;8(8):e1002614.
17. Classen DC, Resar R, Griffin F, et al. 'Global trigger tool' shows that adverse events in hospitals may be ten times greater than previously measured. *Health Aff (Millwood).* 2011;30(4):581-589.
18. Birman-Deych E, Waterman AD, Yan Y, Nilasena DS, Radford MJ, Gage BF. Accuracy of ICD-9-CM codes for identifying cardiovascular and stroke risk factors. *Medical care.* 2005;43(5):480-485.
19. Xu H, Fu Z, Shah A, et al. Extracting and integrating data from entire electronic health records for detecting colorectal cancer cases. *AMIA. Annu. Symp. Proc.* 2011;2011:1564-1572.
20. Lependu P, Iyer SV, Fairon C, Shah NH. Annotation Analysis for Testing Drug Safety Signals using Unstructured Clinical Notes. *Journal of biomedical semantics.* 2012;3 Suppl 1:S5.
21. 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;92(2):228-234.
22. 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.
23. Wishart DS, Knox C, Guo AC, et al. DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic acids research.* 2008;36(Database issue):D901-906.
24. Drugs.com. http://www.drugs.com/. Accessed May, 2011.
25. Kuhn M, Campillos M, Letunic I, Jensen LJ, Bork P. A side effect resource to capture phenotypic effects of drugs. *Mol. Syst. Biol.* 2010;6:343.
26. Liu Y, Lependu P, Iyer S, Shah NH. Using temporal patterns in medical records to discern adverse drug events from indications. *AMIA Summits on Translational Science proceedings.* 2012;2012:47-56.
27. MOBY Project. 2000; http://icon.shef.ac.uk/Moby/.
28. Trifiro G, Pariente A, Coloma PM, et al. Data mining on electronic health record databases for signal detection in pharmacovigilance: which events to monitor? *Pharmacoepidemiol Drug Saf.* 2009;18(12):1176-1184.
29. Tatonetti NP, Ye PP, Daneshjou R, Altman RB. Data-driven prediction of drug effects and interactions. *Science translational medicine.* 2012;4(125):125ra131.
30. Strasberg HR. Medi-Span® Drug-Drug Interactions Study. *Wolters Kluwer Health.* 2012.
31. van der Sijs H, Aarts J, Vulto A, Berg M. Overriding of drug safety alerts in computerized physician order entry. *J. Am. Med. Inform. Assoc.* 2006;13(2):138-147.
32. Lemahieu WP, Hermann M, Asberg A, et al. Combined therapy with atorvastatin and calcineurin inhibitors: no interactions with tacrolimus. *Am. J. Transplant.* 2005;5(9):2236-2243.
33. Gottlieb A, Stein GY, Oron Y, Ruppin E, Sharan R. INDI: a computational framework for inferring drug interactions and their associated recommendations. *Mol. Syst. Biol.* 2012;8:592.
34. Vilar S, Harpaz R, Uriarte E, Santana L, Rabadan R, Friedman C. Drug-drug interaction through molecular structure similarity analysis. *J. Am. Med. Inform. Assoc.* 2012.
35. Xu R, Musen MA, Shah NH. A Comprehensive Analysis of Five Million UMLS Metathesaurus Terms Using Eighteen Million MEDLINE Citations. *AMIA. Annu. Symp. Proc.* 2010;2010:907-911.
36. D'Avolio LW, Nguyen TM, Goryachev S, Fiore LD. Automated concept-level information extraction to reduce the need for custom software and rules development. *J. Am. Med. Inform. Assoc.* 2011;18(5):607-613.