

Adverse Event Profiles for Multi-drug Combinations

Srinivasan Iyer, Kushal Tayal, Siddhi Soman, Jeff Ullman
CS 341 Report

1. Introduction

The practice of polypharmacy has become widespread today and it is well known that this leads to an increased risk of adverse events. It is infeasible for drug manufacturers to test exhaustively for adverse events caused by interaction of drugs with other co-prescribed medications owing to the sheer number of combinations. Owing to this, most interactions are realized after the drug has been in use for some time and it is vital that such reactions be identified and acted upon quickly.

At present, drug interaction effects are primarily available in the form of voluntary reports through Spontaneous Reporting Systems(SRS) such as FDA's AERS and these reports are subject to various biases[1]. Electronic Health Records(EHRs) provide more detailed information than traditional SRSs and their use and availability is growing very fast. They are also not subject to reporting and several other kinds of bias. The use of EHRs for drug safety is still in its infancy and most research is focused on finding adverse reactions of single drugs. We show that it is possible to mine drug interaction information using EHRs and tractably construct an exhaustive database, listing possible interactions between various drug pairs.

Multi-drug profiling presents several new problems which are non-existent in the single drug profiling case. Firstly, data is much more scarce for patients receiving multiple drugs within a short timespan and consequently developing an adverse event. We have solved this problem by an exhaustive search through all possible drug combinations of a given size, thereby surfacing any interactions which have significant statistical support. It is also much more difficult to establish causality in the multi-drug case, as there are many more cases to argue about. We exploit the timestamps associated with drugs and events and build our model accordingly, and this helps us establish causality with a higher certainty.

We demonstrate two methods with differing computational complexities for identifying drug interactions. The first one is a variant of the Apriori algorithm proposed by Brin et. al.[2] to reduce the search space, however, maintaining the interesting results, and this method easily scales to three and four drug combinations as well. This method produces results for only those drug pairs which are not individually related to each other, or the adverse event in question. This method, though requiring less computational power, fails to capture interactions in which one of the drugs increases/decreases the effectiveness of a medication to prevent the adverse event and most interactions are of this type.

The introduction of the Map Reduce paradigm[3] and its easy availability via Amazon Web Services [4] has significantly raised the bar of data volume that can be processed in a reasonable amount of time. Using a map reduce implementation on 20 machines, we were able to generate the complete adverse event profile for an event for all drug pairs within 11 hours of wall clock time. We ran our analysis on two adverse events i.e. Hypoglycemia and Myocardial Infarction and generated a table of 2919 and 307 risky drug combinations respectively for these events. We also validated our results using the Micromedex database [5] and were able to achieve an accuracy of 60%. Owing to the lack of a gold standard for true negatives, we were unable to determine our false discovery rate, but hope to establish this by testing random pairs of drugs for interactions.

2. Datasets

Our results are based on data from the Stanford Clinical Data Warehouse (STRIDE). This dataset comprises approximately 10M EHRs corresponding to ~ 1M patients, who have visited the Stanford Hospital between 1994 and 2011. The EHRs were originally in an unstructured free text format and were

converted into a structured binary matrix using the annotation pipeline built by LePendou et. al.[6]. We reproduce the steps here for completeness(Fig 2.1).

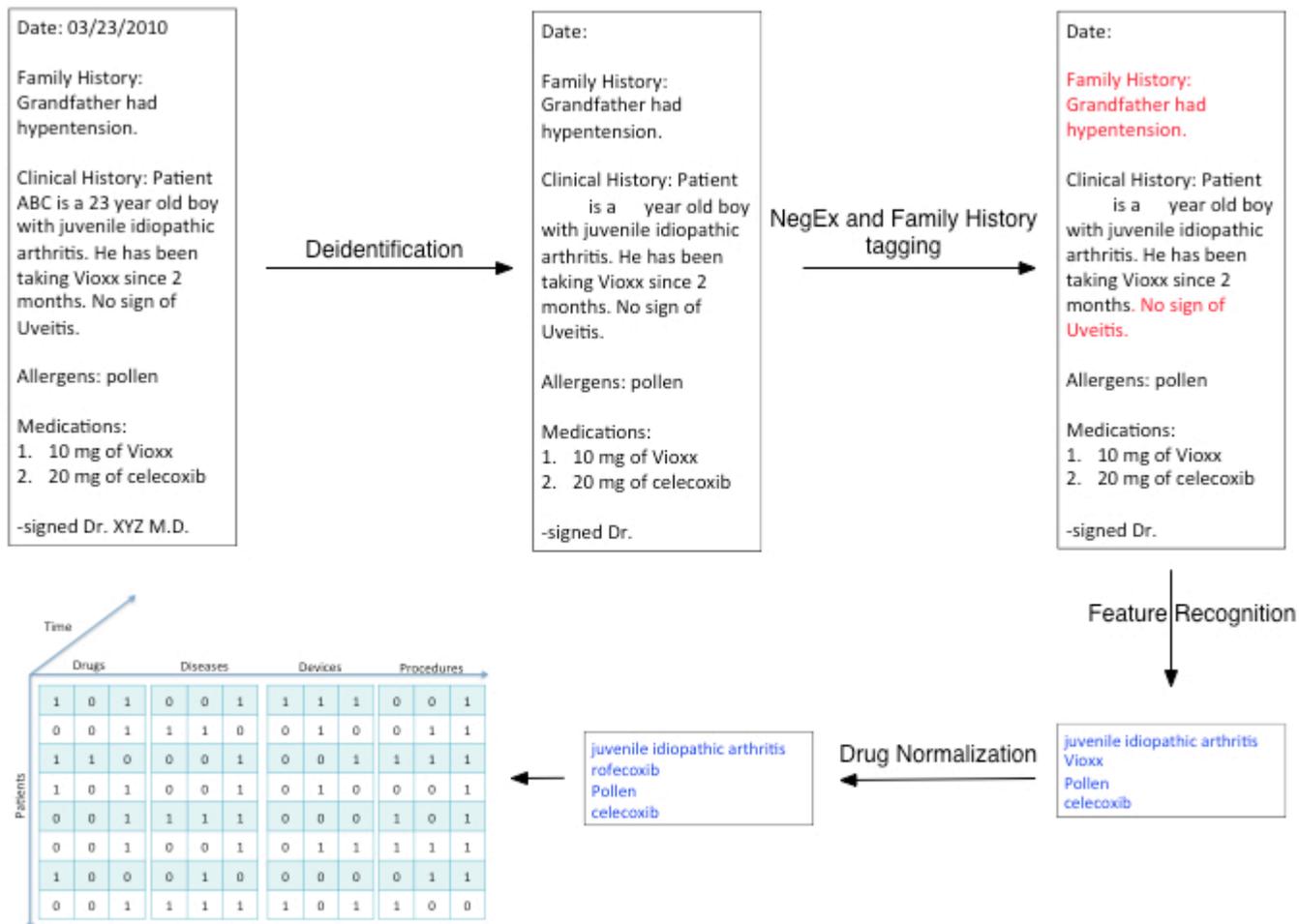


Fig 2.1 Pipeline for the annotation of unstructured EHR data, adopted from LePendou et. al. [6].

Following patient deidentification, the EHR texts are annotated with term ids, using a term extraction annotator. The annotator uses a dictionary comprising ~3M terms built using strings from the NCBO Bioportal Library. Following annotation, portions of the EHR which pertain to the patient's family history are then tagged and removed. Terms which negated are also removed using an adaptation of the NegEx algorithm[7]. The remaining terms are then translated to concepts as defined by the Unified Medical Language System(UMLS)[8] set of ontologies. The entire corpus then consisted of 27891 drug concepts, 49006 disease concepts, 2931 device concepts and 18152 procedure concepts. Normalization of drug concepts into ingredient concepts was done using the RxNORM ontology from UMLS, and this reduced the number of drug concepts to 2892. The resulting data can be represented as a binary matrix B , where $B_{ijk} = 1$, means that patient i had concept J at time instant k .

3. Methods

Our aim in this report is to mine EHRs for associations between combinations of drugs and adverse events i.e. associations of the form

$$\{drug_1, drug_2, drug_3, \dots\} \rightarrow Adverse\ Event$$

We do this by constructing contingency tables to learn the joint probability distribution of the variables. We demonstrate two methods of varying computational complexity to perform the aforesaid task. We shall refer to these as the Apriori method and the Map Reduce method.

3.1 Apriori Method

We adopted the method described in the paper “Beyond Market Baskets: Generalizing Association Rules to Correlations” [2] to find such correlations. Although our aim was to find associations, we generalized it to correlations. The paper discusses measuring the significance of association rules via support and chi-squared test for correlation. It uses two support parameters. One is the standard support as used in data mining. The second being the support fraction parameter. This parameter ensures that a certain fraction of the counts in the contingency table are above the support threshold.

The algorithm start with finding correlations between pairs of items and labelling them as significant or nonsignificant based on the support and chi-square values. In the second phase, it removes pairs that were significant and using the nonsignificant pairs, it forms 3-tuples of items and labels them as significant or nonsignificant. In the third phase it again considers 3-tuples that are not significant and forms 4-tuples. The algorithm progresses in similar fashion. so on. The algorithm makes use of the downward closure property of support and upward closure property of correlation (chi-squared test).

3.1.1 Algorithm

```
INPUT: D = drugs U {EVENT},  $\alpha$ , p, s, records
Phase 1:
for d1,d2 in D
    if O(d1)>s && O(d2)>s
        t = cont-table(d1,d2, records)
        if #(cells in t > s) > p%
            if pval(t)> $\alpha$ 
                NOTSIG2 += {d1,d2}
            else
                SIG2 += {d1,d2}

Phase 2:
for d1,d2 in NOTSIG2 && d3 in D
    if O(d3)>s
        if (d1,d3) in NOTSIG2 && (d2,d3) in NOTSIG2
            t = cont-table(d1,d2,d3, records)
            if #(cells in t > s) > p%
                if pval(t)> $\alpha$ 
                    NOTSIG3 += {d1,d2,d3}
                else
                    SIG3 += {d1,d2,d3}
```

We implemented the above pseudo code using map-reduce. We had 2 map-reduce phases. The first phase would build pairwise drug correlations. The second phase used the non-significant drug pairs from previous pair and checked correlation of the pair with the adverse event. This phase also produced 3-tuples of drug correlations for consequent phases. However, we did not go beyond drug pair correlations with adverse event since the number of results were very few after this phase.

After analysing the results from this method, we came across a few confounding factors that were leading to a lot of false positives. Firstly, we had to take into consideration the time lapse between the drug(s) taken and the adverse event. Secondly, the patient demographics also hampered the results. The patients of different age, sex affect results significantly. Apart from these confounding factors, we also concluded that all drug pairs are important. Removal of significant pairs was resulting in most of the medications being taken for the adverse event to be removed from consideration. However, most

interesting drug interactions occur when one drug nullifies the effect of a medication for the adverse event in question.

3.2 Map Reduce Method

In this method, we corrected the shortcomings from the previous approach. We made sure the drugs being considered are taken less than a year before the adverse event takes place. We also calculated the fraction of patients getting the adverse event based on their age and gender and fit a curve to best match the distribution. We then used this curve to adjust the counts in the contingency table, thus normalizing for age and gender. We also do not throw away any non-significant drugs as in the previous method.

We moved to using the Fisher test instead of the chi-squared test. Fisher test works better for low counts and also helps in identifying the causal relationship. As we have a 3x3 contingency table, we split them into three 2x2 contingency tables and measured the odds ratio and p-value for these 3 tables. First table measures counts for both drugs v/s the event as against not taking both drugs v/s event. The second table measures counts for one drug and not the other v/s the event as against not taking both drugs v/s the event. The third table measures vice versa. Thus we can check if taking both drugs increases the odds of having the adverse event as against each drug considered individually.

3.2.1 Algorithm

```
INPUT:
dataFile = {[d1, data1], [d2, data2], [d3, data3], ...},
eventData, drugList

Map(d1, data1)
  for each d in drugList
    output({d, d1}, {d1: data1})
Reduce({d1, d2}, {d1: data1, d2: data2})
  t = cont-table(data1, data2, eventData)
  output(compute-stats(t))
```

3.2.2 Optimization

We implemented this using a single phase of map-reduce. Our initial plan used one reducer for each pair of drugs. Given that we had about 3000 drugs, and assuming each drug pair would have a data of 1MB, it resulted in intermediate data of ~4.5T. We could not run this job fully and decided to terminate the job and optimize the code.

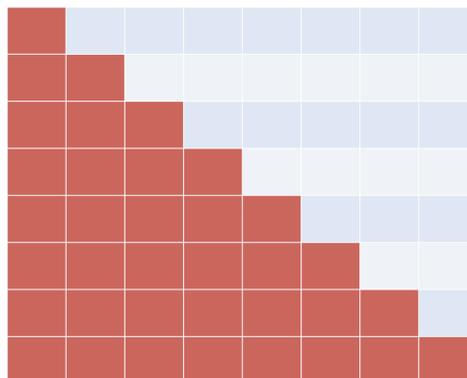


Fig 3.1 Map Reduce Optimization

We partitioned the drugs into 20 groups, each group now having about 150 drugs. Each reducer now received a set of two such groups. Thus a total of 200 reducers were needed and each reducer received data of size 300MB, reducing the intermediate data written to 60GB. However, each reducer now has to

process 150x150 pairs. Given powerful machines, this turned out to be an excellent trade off reducing our run time by a huge factor.

4. Results

We have two sets of results, corresponding to the two techniques that we implemented. We will call these Apriori results and Map Reduce results.

4.1 Apriori Results

Drug 1	Drug 2	Significant?	Chi-square	p-val
hydrocodone	prednisolone	SIG	4.11616	0.0424753
ribavirin	formalin	NOTSIG	3.74414	0.0529928
ampicillin	diazepam	SIG	3.87785	0.0489272
bupivacaine	valproic acid	NOTSIG	3.83524	0.0501857
ofloxacin	docusate	SIG	3.98806	0.0458239
thalidomide	formalin	SIG	4.15149	0.0415978
zinc sulfate	formalin	SIG	4.12099	0.0423543
ampicillin	cortisone	NOTSIG	3.68865	0.0547839
cefaclor	formalin	NOTSIG	3.76764	0.0522533
cocaine	warfarin	NOTSIG	3.76801	0.0522419
cefazolin	rituximab	SIG	4.12413	0.0422758
hydrocodone	stearyl alcohol	NOTSIG	3.64182	0.0563455
betamethasone	warfarin	SIG	4.07735	0.0434617
acetaminophen	terbutaline	SIG	3.99439	0.0456521
heparin	lamotrigine	SIG	3.9002	0.0482803
riboflavin	loratadine	NOTSIG	3.54786	0.0596222
ibuprofen	esmolol	NOTSIG	3.74223	0.0530535
myocardial infarction	tamoxifen	SIG	3.89466	0.0484398
hydrocodone	isotretinoin	NOTSIG	3.63818	0.0564685

Significant?	Event	Drug 1	Drug 2	Chi square
SIG	myocardial infarction	montelukast	rituximab	123.854
SIG	myocardial infarction	doxorubicin	montelukast	110.924
SIG	myocardial infarction	carbamazepine	cyclophosphamide	66.9601
SIG	myocardial infarction	cyclophosphamide	mometasone	65.1719
SIG	myocardial infarction	megestrol	mometasone	59.8776
SIG	myocardial infarction	mometasone	rituximab	59.2914
SIG	myocardial infarction	filgrastim	mometasone	51.9082
SIG	myocardial infarction	interferon alfa-2a	mometasone	51.0973
SIG	myocardial infarction	interferon beta-1a	mometasone	51.0973
SIG	myocardial infarction	doxorubicin	mometasone	47.0905
SIG	myocardial infarction	mometasone	recombinant interferon	45.8728

Fig 4.1: Phase 1 and Phase 2 of the Apriori type algorithm respectively

Using $s = 50$ and $p = 0.25$ (see Algorithm 3.1.1), the two phases of our algorithm completed within 2 hours with Heart Attack as the event. The results of the two phases are summarized in Fig 4.1. Out of 3.5 million pairs with sufficient support, ~737000 pairs were found to be significant and were excluded from Phase 2. However, in Phase 2, only 11 pairs were found, which were significant together with Heart Attack. This hinted at the fact that there would be extremely few results for 3 drugs and more and therefore, we did not take this analysis further. Moreover, the results mostly comprised cancer related drugs, and it could be argued that the presence of cancer was a confounder in our analysis.

4.2 Map Reduce Results

We evaluated our Map Reduce algorithm on two events - Hypoglycemia and Heart Attack.

4.2.1 Hypoglycemia

4.2.1.1 True Positives - Micromedex

After formulating the algorithms for mining EHRs for associations between pairs of drugs and adverse events, we required a way to validate our approach. We utilized the Micromedex database for this purpose. Micromedex is a database containing information about a number of drugs. More importantly it contains information about known drug interactions - i.e. drugs which are known to interact and lead to adverse events. We extracted drug interaction information from the Micromedex database such that the pairs of drugs are known to interact and lead to Hypoglycemia. This formulated our set of true positives. We then proceeded to evaluate our approach on this set of drug pairs known to increase the risk of Hypoglycemia.

The set of true positive had 89 pairs. Since our data consists of doctor's notes, we do not have enough number of patients corresponding to all these drug combinations. For the data that we do have, we applied a number of filters providing us variations in the number of pairs retained along with the number of pairs we were able to identify correctly.

1. Filter : $a > 0$ - Where a is the number of patients who have taken both the drugs and have had the event:

29 found positive, 13 not found.

2. Filter : $a > 10$

12 found positive, 6 not found.

3. Filter : p_{vald1} , p_{vald2} and $p_{vald1d2} < 0.05$

7 found positive, 5 not found.

Rest of the pairs did not have enough data for us to infer anything.

4.2.1.2 Validation via the Facts and Comparisons database

Instead of proceeding to test our approach on various combinations of all drugs, we evaluated our approach on different combinations of these suspect drugs. There were overall 57 drugs in the set of true positives that we obtained for Hypoglycemia. We eliminated the pairs which were already included in the set obtained from Micromedex. We applied the different filters on the 1547 pairs that were newly generated.

1. Filter : $a > 0$ - Where a is the number of patients who have taken both the drugs and have had the event:

325 new pairs.

2. Filter : $a > 10$

103 new pairs.

3. Filter : p_{vald1} , p_{vald2} and $p_{vald1d2} < 0.05$

30 new pairs.

There is another database similar to the Micromedex database called the Facts and Comparisons database. This database also contains drug interaction information i.e. drugs which are known to interact and increase the risk of the event, Hypoglycemia, in our case.

From the new pairs generated from the list of suspect drugs, we found a number of drug pairs which were not present in the Micromedex database but were verified by the Facts and Comparisons database. A few of them are as follows:

1. Glyburide and Rosiglitazone –

$P[\text{Hypo} | \text{both}] = 13.5\%$ - $[\text{OR} = 10.36, P: 5.8 \text{ e-}11]$

$P[\text{Hypo} | \text{Glyburide only}] = 6.6\%$ - $[\text{OR} = 4.54, P: 2.2 \text{ e-}25]$

$P[\text{Hypo} | \text{Rosiglitazone only}] = 5.6\%$ - $[\text{OR} = 4.02, P: 1.7 \text{ e-}09]$

Glyburide, Rosiglitazone - Used as diet and exercise medicines and in the treatment of type 2 diabetes.

2. Trimethoprim and Pioglitazone–

$P[\text{Hypo} | \text{both}] = 10.9\%$ - $[\text{OR} = 8.87, P: 4.4 \text{ e-}05]$

$P[\text{Hypo} | \text{Trimethoprim only}] = 1.7\%$ - $[\text{OR} = 7.7, P: 2.7 \text{ e-}4]$

$P[\text{Hypo} | \text{Pioglitazone only}] = 4.2\%$ - $[\text{OR} = 3.09, P: 5.6 \text{ e-}10]$

Trimethoprim - Used for treating urinary infections, in combination with other drugs to treat pneumonia.

Pioglitazone - Used as a diet and exercise medicine and in the treatment of type 2 diabetes.

3. Glipizide and Clarithromycin–

$P[\text{Hypo} | \text{both}] = 12.14\%$ - $[\text{OR} = 9.19, P: 0.023]$

$P[\text{Hypo} | \text{Glipizide only}] = 7.26\%$ - $[\text{OR} = 5.1, P: 1.0 \text{ e-}32]$

P[Hypo|Clarithromycin only] = 0.64% - [OR= 0.42, P: 8.3 e-6]
 Glipizide - Used as a diet and exercise medicine and in the treatment of type 2 diabetes.
 Clarithromycin - Used for treating bacterial infections, pneumonia, bronchitis etc.

4.2.1.3 New Pairs

We also discovered a number of pairs which were not present in either the Micromedex or the Facts and Comparisons database as follows:

1. Pioglitazone and Rosiglitazone–

P[Hypo | both] = 11.6% - [OR= 8.59, P: 8.64 e-06]
 P[Hypo|Pioglitazone only] = 4.18% - [OR= 2.76, P:2.05e-08]
 P[Hypo|Rosiglitazone only] = 6.04% - [OR= 4.17, P: 3.30 e-11]

Pioglitazone, Rosiglitazone - Used as diet and exercise medicines and in the treatment of type 2 diabetes.

4.2.1.4 Evaluation on all drug pairs

After validating our approach as above, we proceeded to evaluate it on all pairs of drugs. We have ~2892 drugs and correspondingly we tested 2,023,066 drug pairs. Among the pairs tested, 2919 pairs were found to be statistically significant. Among these we decided to filter out pairs such that, none of the individual drugs are related to Hypoglycemia, but the drug pair seems to increase the risk of Hypoglycemia. Following is a snapshot of the database with these 17 drug pairs filtered out (Fig 4.2).

Drug 1	Drug 2	P[both]	P[first]	P[second]
tobramycin	fluticasone	6.142857298	0.503416033	0.68292683
methotrexate	povidone-iodine	4.530313239	0.73236404	0.678604003
methotrexate	iodine	4.187327996	0.694600027	0.788215548
sodium fluoride	iodine	2.603143454	0.673709996	0.687475502
sodium fluoride	povidone-iodine	2.568751946	0.744706718	0.589228189
polyethylene glycols	fluticasone	2.524038404	1.077150926	0.611310638
dexamethasone	polyethylene glycols	2.436406624	0.580326002	1.083518378
beclomethasone	hydrocodone	2.257114835	0.483870972	0.657759886
dexamethasone	sodium fluoride	1.949634403	0.52393428	0.74081393
nitrogen	hydrocodone	1.731504314	0.521896966	0.548347598
iodine	oxymetazoline	1.602364704	0.569866691	0.708445814
diphenhydramine	acetaminophen	1.521938574	0.457922928	0.595073169
methamphetamine	amoxicillin	1.349481009	0.54941061	0.685010012
guaifenesin	hydrocodone	1.339491922	0.574529218	0.590580096
fluticasone	hydrocodone	1.185944397	0.541024841	0.578221679
amoxicillin	hydrocodone	1.048466843	0.684538949	0.521258544
acetaminophen	triclosan	1.042102184	0.277416734	0.390553358

Drug 1	Drug 2	Rank	P[Both]	P[First]	P[Second]
aspirin	probenecid	63.57904515	6.666667014	0.165289256	4.229206592
ofloxacin	aspirin	49.1219208	4.964539036	4.223433137	0.118800125
dobutamine	raloxifene	29.21938343	18.51851791	1.217263006	9.641774744
carvedilol	hydroxyzine	25.20994387	14.88095224	1.044896059	8.40652436
dobutamine	meperidine	22.73236847	16.82243049	1.303006057	9.554024786
aspirin	nortriptyline	20.29101468	3.671970591	0.157480314	4.219571501
meperidine	clopidogrel	18.43052537	14.79289979	9.280373901	1.279391441
cyproterone	aspirin	17.52920219	3.850509599	4.198403656	0.201460579
tetracycline	clopidogrel	17.18538431	14.28571492	9.437453747	1.25831645
quinapril	aspirin	15.96673869	6.917164475	4.172873124	0.718132872
famotidine	milrinone	15.32639247	13.82113844	6.765067577	1.842364483
hydroxyzine	aspirin	15.07524906	4.618117213	4.149386659	0.340942782
levofloxacin	milrinone	15.02779808	15.7407403	6.707317382	2.458137088
aspirin	beclomethasone	14.3368143	4.766269401	0.37834004	4.188149795
fluconazole	milrinone	14.05498264	11.92660555	7.609988004	1.329901349
furosemide	hydroxyzine	13.96032215	5.996472761	0.516214408	4.989605024
ciprofloxacin	milrinone	12.76495628	12.42236048	6.797385961	1.778471842
albuterol	isosorbide dinitrate	12.55899313	14.71571922	10.45112759	1.649852097

(a) (b)
 Fig 4.2: Risky drug combinations for (a) Hypoglycemia (b) Heart Attack

Following are some of the pairs from the list above:

1. Tobramycin - Used as an antibiotic to treat bacterial infections.
 Fluticasone - Used in nasal sprays.
2. Diphenhydramine - Used in the treatment of redness, itchy eyes, running nose etc due to allergies.
 Acetaminophen - Used as a painkiller.

3. Hydrocodone - Used in combination with other drugs as a painkiller and to relieve cough
Guaifenesin - Used to relieve chest congestion
4. Hydrocodone - Used in combination with other drugs as a painkiller and to relieve cough
Amoxicillin - Used to treat bacterial infections, bronchitis, pneumonia etc.

We are currently in the process of talking to medical experts about investigating these suspect drug pairs for an increased risk of Hypoglycemia.

4.2.2 Heart Attack

Our algorithm surfaced 307 risky drug combinations which cause an increased risk of heart attack(Fig 4.2). The curves from section 3 were used for age and gender normalization. There are several pairs, in which one drug is taken to prevent the occurrence of a Heart Attack, and the other drug nullifies its effect. For instance, 6.7% of patients taking Milrinone alone suffer a heart attack. However, among the patients also taking Levofloxacin at the same time, 15.7% suffer a heart attack and this represents a significant increase. The reason for this pair being interesting is that Levofloxacin is an antibiotic and is not related to the cardiovascular system per se. We were unable to validate these results owing to the lack of a gold standard.

4.3 Graphical representation

The appendix shows a pictorial depiction of the top 300 results for Hypoglycemia and Heart Attack. The nodes are the drugs and there is an edge between the nodes if the pair was among the pairs of suspects (interacting drug pairs). The nodes are arranged in a circular layout to easily make of their degree. We observe that most nodes with a high degree are taken to prevent the adverse event and are convicted when taken along with some other non-related drug. It is exactly these type of pairs, which we hoped to find, using our Map reduce implementation.

5. Conclusions

We tried two methods to mine multi drug associations from EMRs. Firstly, we used an Apriori style method to find three drug combinations which lead to some adverse event. We required every subset of these drugs to be unrelated to the event and exploited the fact that correlation is upward closed. However, after the second Phase, very few drugs remained thus proving this method to be too restrictive.

We then focussed our attention on pairs of drugs which lead to an adverse event, allowing for individual drugs to be related to the event. We looked for an increase in risk of the event when both drugs were taken together, as compared to each drug taken individually. Using this method we were able to find 2919 risky drugs pairs causing Hypoglycemia(validated with Micromedex) and 307 risk pairs causing Heart Attack. We corrected for age and gender and generated adjusted odds ratios for the effects of either and both the drugs.

Judging from the fact that these pairs of drugs have survived all the strict rules what we have applied throughout the analysis, it is very likely that many of these results are true. However, we also expect several false positives owing to confounding by unmodeled variables. One possible issue would be that the two drugs may be taken as medications for certain conditions, which are responsible for an increased risk of the adverse event. In this case, we would wrongly attribute the cause to the drugs. Although we have made good use of the timing information to establish causality, it would also be necessary to adjust for co-morbidities, a task which we hope to undertake in future work.

6. Acknowledgements

We would like to thank Jeff Ullman(Dept of CS, Stanford University) for his mentorship and ideas, which helped us at every stage of this work. We are also very thankful to Nigam Shah (Dept. of Bioinformatics, Stanford University) and his group for giving us access to the dataset and for providing insightful suggestions from time to time. Much thanks to Tyler Cole for helping us make sense of the results.

7. References

1. Schneeweiss, S. and J. Avorn, A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol*, 2005. 58(4): p. 323-337.
2. S. Brin, R.Motwani, and C. Silverstein. Beyond Market Baskets: Generalizing Association Rules to Correlations. In *The Proceedings of SIGMOD*, pages 265-276, AZ, USA, 1997.
3. Jeffrey Dean, Sanjay Ghemawat: *MapReduce: a flexible data processing tool. Commun. ACM* 53(1): 72-77 (2010)
4. Amazon Web Services: <http://aws.amazon.com/>
5. MICROMEDEX Thomson Reuters, 3 Times Square, New York, NY 10036, USA.
6. Paea LePendu, Srinivasan Iyer, Cedrick Fairon and Nigam Shah, Annotation Analysis for Testing Drug Safety Signals using Unstructured Clinical Notes in *Journal of BioMedical Semantics*, 2011.
7. Chapman, W.W., et al., A simple algorithm for identifying negated findings and diseases in discharge summaries. *Journal of Biomedical Informatics*, 2001. 34(5): p. 301-10.
8. The Unified Medical Language System(UMLS): <http://www.nlm.nih.gov/research/umls/>

8. Appendix

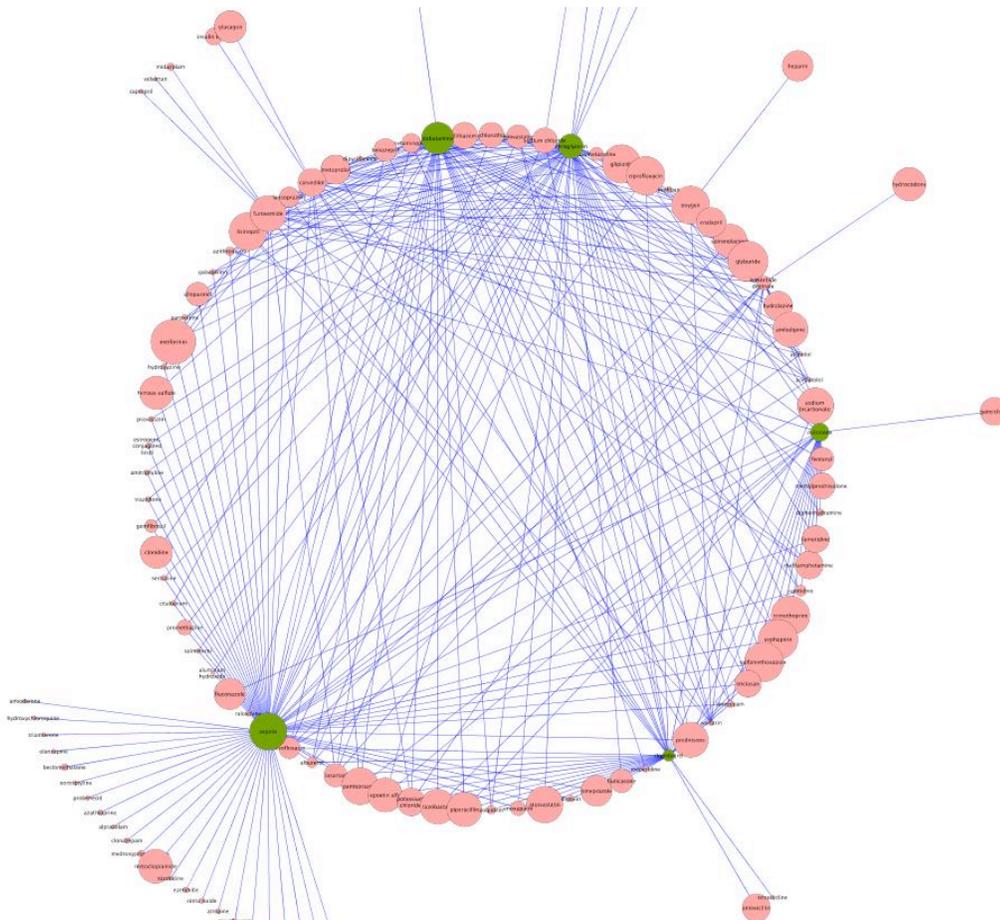


Fig: Graphical Representation of risky drug combinations for Heart Attack

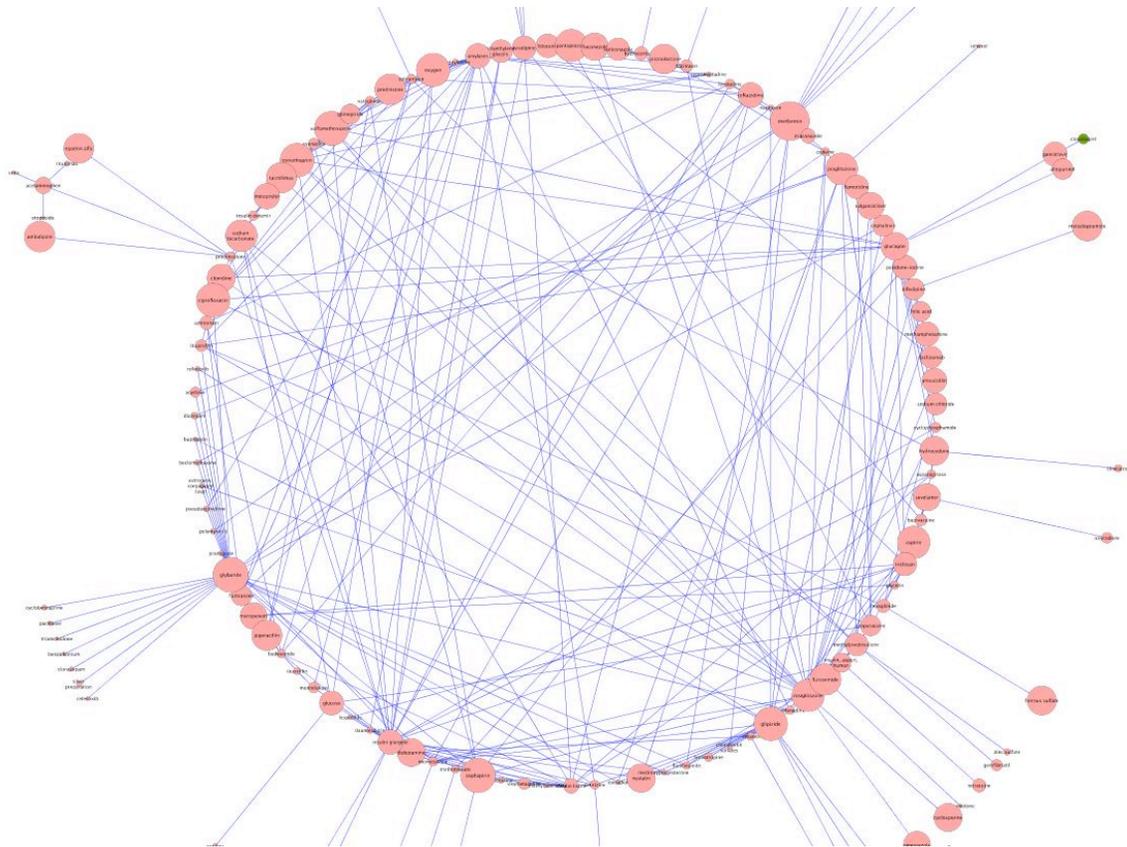


Fig: Graphical Representation of top 300 risky drug combinations for Hypoglycemia