Finding Causal Mechanistic Drug-Drug Interactions from Observational Data

Sanjoy Dey, PhD[1], Ping Zhang, PhD[2], Mohamed Ghalwash, PhD[1], Chandramouli Maduri, MS[1], Daby Sow, PhD[1], Zach Shahn, PhD[1]
[1]IBM T. J. Watson Research Center, Yorktown Heights, NY, USA
[2]The Ohio State University, Columbus, OH, USA

Abstract

Many adverse drug reactions (ADRs) are caused by drug-drug interactions (DDIs), meaning they arise from concurrent use of multiple medications. Detecting DDIs using observational data has at least three major challenges: (1) The number of potential DDIs is astronomical; (2) Associations between drugs and ADRs may not be causal due to observed or unobserved confounding; and (3) Frequently co-prescribed drug pairs that each independently cause an ADR do not necessarily causally interact, where causal interaction means that at least some patients would only experience the ADR if they take both drugs. We address (1) through data mining algorithms pre-filtering potential interactions, and (2) and (3) by fitting causal interaction models adjusting for observed confounders and conducting sensitivity analyses for unobserved confounding. We rank candidate DDIs robust to unobserved confounding more likely to be real. Our rigorous approach produces far fewer false positives than past applications that ignored (2) and (3).

1 Introduction

Adverse drug reactions (ADRs) lead to nearly 74,000 emergency room visits and 195,000 hospitalizations each year in the United States[1,2]. Many ADRs are brought about by drug-drug interactions (DDIs), meaning that they arise as a result of concurrent use of multiple medications. Concurrent medication use is particularly common among the elderly[3]. DDIs are difficult to identify from clinical trials, which typically only randomize one treatment at a time and have sample sizes too low to reliably detect interaction effects for rare outcomes. Thus, DDI detection must depend on large scale observational data analysis.

In the United States, the Food Drug Administration (FDA) has implemented the FDA Adverse Event Reporting System (FAERS) database[4]. FAERS is a system gathering adverse event reports, medication error reports and product quality complaints resulting in adverse events submitted to the FDA by healthcare professionals (such as physicians, pharmacists, nurses, and others), consumers (such as patients, family members, lawyers, and others) and manafacturers (such as pharmaceutical companies). FAERS has been a critical source of data for pharmaco-vigilance analytics[5] mostly focusing on detecting adverse effects of individual drugs. In this work, we propose to use FAERS to establish and detect DDIs.

Identifying DDIs causing an adverse drug reaction from observational databases such as FAERS poses at least three main challenges. First, the number of potential candidate drug-drug interactions is extremely large, which is computationally challenging. Second, associations between drugs and ADRs in observational data may not be causal due to observed or unobserved confounding. For example, a candidate DDI with two interacting drugs A_1 and A_2 causing ADR Y (as shown in Figure 1) may have both observed and unobserved confounders. In FAERS, observed confounders comprise co-prescribed drugs (C_1 in Figure 1) and comorbidities captured as indications of co-prescribed drugs (C_2 in Figure 1). Unobserved confounders include demographic factors, unrecorded indicators of health status such as vital signs or lab tests, and more. Third, pairs of drugs that each independently cause an ADR and are frequently co-prescribed do not necessarily causally interact, where causal interaction means that there are at least some patients who would only experience the ADR if they take both drugs.

We address the first challenge of too many candidate DDIs by using data mining algorithms to pre-filter. Most existing studies[6] in this area stop here and ignore the remaining challenges. Many of the candidate DDIs these studies flag are not causal, since they did not adjust for confounding. We address the remaining challenges by fitting causal interaction models to adjusting for observed confounders and conducting sensitivity analyses for

possible unobserved confounding. Candidate DDIs robust to strong unobserved confounding in the sensitivity analyses are ranked as more likely to be real.

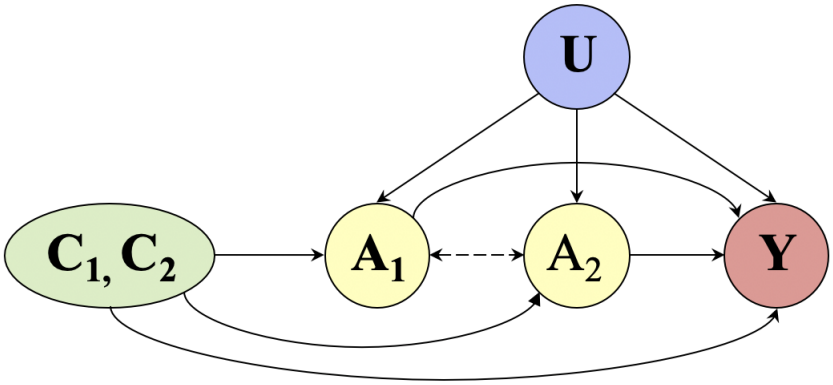

FIGURE 1: Possible confounders (C_1, C_2 and U) for a potential DDI $A_1 \& A_2 \Rightarrow Y$.

Past applications in this field have produced a very large number of false positives. For example, another popular DDI detection algorithm based on FAERS data (Twosides[7]) produces thousands of DDIs with 'high confidence' of causing each outcome we considered. This is medically implausible, as DDIs are not that common. Our approach is considerably more conservative than Twosides, i.e. less likely to generate false positives, for two reasons. First, while Twosides employs ad hoc heuristics to identify associations, we explicitly target a metric that implies causal interaction. Second, while Twosides assumes that its matching procedure is sufficient to remove confounding, our sensitivity analysis allows for the likely eventuality that we have failed to fully adjust for confounding. While validation is a challenge in this area because there do not exist reliable sources of ground truth about DDIs, it is more justifiable to devote limited resources to further investigate the few candidate DDIs flagged by our approach than the many flagged by alternative less rigorous approaches with high false positive rates.

2 Related Work

Significant research efforts have focused on detecting and predicting DDIs. One group of methods leverages various biochemical and molecular drug/target data to measure drug-drug similarities and score/predict DDIs. These data include chemical structures[8,9], target information[10], compound-target docking scores[11], and drug side effects[12]. There have also been efforts at predicting DDIs by integrating molecular and pharmacological data[13,14]. Another group of methods analyze medical literature and/or electronic medical records to extract potential DDIs. Various computational techniques such as traditional regression models[15], association rule mining[6,16,17], disproportionality ratios[18,19], biclustering[20], and text mining[21] have been used to detect DDIs in a fully data-driven manner. Some of these methods[7,22] consider issues of confounding. However, none of the data-driven methods explicitly targets a *causal interaction* effect or incorporates sensitivity analysis to acknowledge that some confounding will be unobserved.

3 Methods

We divide our approach into four stages. Stage 1 entails pre-filtering to identify potential drug-drug-ADR triplets in which the two drugs may interact to cause the ADR. In stage 2, for each triplet identified in Stage 1, we identify a set of potential confounding variables that are associated with either of the drugs or the outcome. In stage 3, we estimate the causal interaction effects in the triplets identified in stage 1 adjusting for the possible confounders identified in stage 2. In stage 4, we perform sensitivity analyses to assess the robustness of the causal interaction effect estimates obtained in stage 3 to possible unobserved confounding. In the following, we refer to drug-drug-ADR triplets in which the drugs actually interact to cause the ADR as DDIs, and we refer to triplets in which the drugs may or may not interact to cause the ADR as candidate DDIs.

3.1 Data and Notation

The FAERS database comprises self-reported suspected adverse reactions. Each self-report includes the adverse outcome (suspected to be a reaction to some medication), all medications taken by the patient leading up to the adverse outcome, and the indications for which all the listed medications were prescribed. A strength of such self-reported data is that we can have high confidence that reported drugs were actually taken, in contrast to EHR data where non-adherence is a concern.

For a given candidate DDI, let A_1, A_2, and Y denote binary random variables over the space of reports indicating presence of the first drug of interest, the second drug of interest, and the outcome of interest, respectively. Let C be a random vector denoting additional possible confounding variables (i.e. presence of indications and other drugs) to be adjusted for when estimating the effect of A_1 and A_2 on Y. We will use upper case letters to refer to random variables and lower case letters to refer to their particular realizations.

If a report is for an adverse outcome other than the ADR of interest, $Y = 0$. A shortcoming of the dataset is that when $Y = 0$ we do not actually know for sure that the patient who is the subject of the report did not also experience the ADR of interest in addition to whatever outcome motivated the report. However, if the ADR of interest is rare in the population, this will not significantly bias our results.

We adopt the counterfactual framework for causal inference[23] in which it is assumed that corresponding to each possible treatment assignment (a_1, a_2) of A_1 and A_2 there exists a counterfactual outcome $Y(a_1, a_2)$ which is the outcome that would have been observed had the patient actually received treatment assignment (a_1, a_2). We say that a *mechanistic interaction* between A_1 and A_2 is present if for at least some patients $Y(1,1) = 1$ but $Y(1,0) = Y(0,1) = Y(0,0) = 0$[24], i.e. if at least some patients would only experience the outcome if they were to receive both treatments. Our goal will be to identify mechanistic interactions, since compared to other possible definitions of a DDI they are most relevant to public health.

3.2 Pre-filtering

We prefiltered candidate DDIs according to the following criteria:

- Rare outcome: $P(Y = 1) \leq .01$
- Unadjusted statistical interaction: We only considered candidate DDIs with unadjusted statistical interaction $\frac{P(Y=1|A_1=A_2=1)}{P(Y=1|A_1=A_2=0)} - \frac{P(Y=1|A_1=1,A_2=0)}{P(Y=1|A_1=A_2=0)} - \frac{P(Y=1|A_1=0,A_2=1)}{P(Y=1|A_1=A_2=0)} + 1 \geq 2$.
- Sufficient power: We required power ≥ 0.8 to test whether the unadjusted statistical interaction was greater than 2 at a .01 level assuming that the true unadjusted statistical interaction was equal to the larger of the observed unadjusted statistical interaction or 3.

The rare outcome assumption is necessary to be able to approximate relevant quantities using a case-control design[25]. The other two assumptions are simply to restrict candidate DDIs to those both more likely to be real and have sufficient support in the data to estimate their interaction effects. The unadjusted statistical interaction between drugs in almost all true DDIs will be greater than 2 for reasons that will become clear below. And effect estimates for candidate DDIs with higher power are less likely to be pure noise arising from sampling variability. In addition, for a given ADR, we also curated the set of drugs that are prescribed for the treatment of the same ADR as indication using the SIDER[26] database. We removed these drugs from the potential candidate DDIs to cause the ADR under consideration.

3.3 Identification of Possible Confounders

In a careful study of an individual candidate DDI, confounder selection would be based on expert knowledge of the causes of the outcome and drugs of interest[27]. In this work, we are designing an automated algorithm to assess large numbers of candidate DDIs, and an automated confounder selection approach cannot require expert knowledge about individual candidate DDIs. For each drug and outcome included in each candidate DDI, we

identify the variables (i.e. medications and indications) that are most strongly associated with the it (e.g., $C1$ and $C2$ of Figure 1). In particular, we find up to 50 drugs and indications that are associated with the ADR under consideration, and another 50 indications that are related to each drug of the DDI. Thus, we generated a list of up to 150 variables comprising the potential confounding variables for each candidate DDI and then we adjust for these confounders while estimating causal interaction effects as described in the following subsection. We use mutual information to identify candidate confounders of each DDI.

3.4 Causal Interaction Estimation

In this step, we want to find the mechanistic interaction between two drugs after adjusting for all the observerd confounders obtained from the previous step. In particular, we compute the Relative Excess Risk due to Interaction (RERI) of a particular DDI, which is defined as

$$RERI \equiv RR_{11} - RR_{10} - RR_{01} + 1, \tag{1}$$

where $RR_{a_0 a_1}$ denotes $P(Y(a_0,a_1)=1)/P(Y(a_0,a_0)=1)$. The RERI is an important quantity because an RERI greater than 2 implies a mechanistic interaction[24]. We cannot directly estimate the RERI as limitations of the FAERS dataset force us to adopt a case-control design. However, for rare outcomes,

$$RERI \approx RERI_{OR} \equiv OR_{11} - OR_{10} - OR_{01} + 1, \tag{2}$$

where $OR_{a_0 a_1}$ denotes $\frac{P(Y(a_0,a_1)=1)/P(Y(a_0,a_1)=0)}{P(Y(0,0)=1)/P(Y(0,0)=0)}$. Note that $OR_{a_0 a_1}$ and therefore $RERI_{OR}$ are counterfactual quantities. If we specify a logistic regression model

$$logit(P(Y=1|A_1=a_1,A_2=a_2,C=c)) = \gamma_0 + \gamma_1 a_1 + \gamma_2 a_2 + \gamma_3 a_1 a_2 + \gamma_4^{'} c \tag{3}$$

and assume that C contains all common causes of A_1 and Y and all common causes of A_2 and Y, then

$$RERI_{OR} = e^{\gamma_1+\gamma_2+\gamma_3} - e^{\gamma_1} - e^{\gamma_2} + 1. \tag{4}$$

Hence we can estimate $RERI_{OR}$ by plugging in the parameter estimates of our logistic regression (3) into (4).

3.5 Sensitivity Analysis

It is a very strong assumption that C is sufficient to adjust for all confounding between both treatments and the outcome. Causal sensitivity analyses[24,28,29] posit that there exists an unobserved confounder U not included in C as in Figure 1. For a range of parameter values determining the prevalence of U at different levels of the treatments and the strength of U's association with the outcome, we can compute the $RERI_{OR}$ estimate we would have obtained had we been able to adjust for U assuming the causal graph in Figure 1. If U-adjusted $RERI_{OR}$ estimates are large for a wide range of sensitivity parameters governing the impact of hypothetical unobserved confounder U, this is an indication that these estimates are robust to unobserved confounding. (The level of robustness to confounding is determined by a complex function of the dependencies between treatments, observed confounders, and the outcome, i.e. robustness is not simply equivalent to a large estimated effect.)

Define $\lambda \equiv \frac{E[Y|A_1=a_1,A_2=a_2,C=c,U=1]}{E[Y|A_1=a_1,A_2=a_2,C=c,U=1]}$ assumed constant over a_1,a_2,c. And let $\rho_{a_1 a_2}$ denote $P(U=1|A_1=a_1,A_2=a_2,C=c)$ assumed constant over c. Then the $U-$adjusted estimate of $RERI_{OR}$ is

$$\hat{RERI}_{OR}(\lambda,\rho) = \frac{e^{\hat{\gamma}_1+\hat{\gamma}_2+\hat{\gamma}_3}}{\frac{1+(\lambda-1)\rho_{11}}{1+(\lambda-1)\rho_{00}}} - \frac{e^{\hat{\gamma}_1}}{\frac{1+(\lambda-1)\rho_{10}}{1+(\lambda-1)\rho_{00}}} - \frac{e^{\hat{\gamma}_2}}{\frac{1+(\lambda-1)\rho_{01}}{1+(\lambda-1)\rho_{00}}} + 1. \tag{5}$$

For each candidate DDI, we compute $\hat{RERI}_{OR}(\lambda,\rho)$ over a grid of λ and ρ values spanning $[1,4]$ by increments of .5 for λ and $[0,1]$ by increments of .25 for $\rho_{11},\rho_{10},\rho_{01}$, and ρ_{00}. We refer to the resulting grid of U-adjusted $RERI_{OR}$ estimates as the 'sensitivity grid' for a candidate DDI. We also estimate by bootstrap a covariance matrix $\hat{\Sigma}$ for $\hat{RERI}_{OR}(\lambda,\rho)$ estimates over all candidate DDIs and values of λ and ρ.

3.6 Ranking Candidate DDIs Based on Sensitivity Analysis Output

We consider multiple schemes for ranking candidate DDIs based on their sensitivity grids and the confidence intervals for the entries of those grids.

- Confounding naive approach: Rank candidate DDIs by their RERI estimates
- Conservative confounding naive approach: Rank candidate DDIs by the lower bound of the 0.95 confidence interval of the RERI estimate.
- Confounding aware approach: Rank candidate DDIs by the p^{th} (e.g. 0.25) quantile of the sensitivity grid.
- Conservative confounding aware approach: Rank candidate DDIs by the lower bound of the 0.95 confidence interval for the p^{th} (e.g. 0.25) quantile of the sensitivity grid.

4 Experiments

We applied our approach to detect DDIs causing Acute Renal Failure (ARF), Acute Myocardial Infarction (AMI), and Acute Liver Injury (ALI) in the FAERS database. We used the OMOP definition of UMLS concept ids to map the FAERS data to these three ADRs. Our dataset comprised 6,293,657 total reports collected from Q4 of 2012 to Q4 of 2014. 13,138 were regarding instances of ARF, 25,045 were regarding instances of AMI, and 12,911 were regarding instances of ALI. The pre-filtering process reduced these numbers to just 69 candidate DDIs for ARF, 60 for AMI, and 32 for ALI. For each outcome, we report all four scores described in the previous subsection for top candidate DDIs.

TABLE 1

(a) Top 15 DDIs of ARF ranked by the .25 quantile of sensitivity grid

Drug1	Drug2	RERI
Lisinopril	Sodium Phosphate, Dibasic	192.7
Heparin	Mannitol	10.7
Potassium Chloride	dabigatran etexilate	4.6
Lisinopril	dabigatran etexilate	3.1
carvedilol	dabigatran etexilate	2.8
Allopurinol	dabigatran etexilate	2.4
Diltiazem	dabigatran etexilate	2.2
dabigatran etexilate	Spironolactone	2.2
Albuterol	dabigatran etexilate	1.9
Acetaminophen	dabigatran etexilate	1.6
Digoxin	dabigatran etexilate	1.6
Aspirin	dabigatran etexilate	1.6
Simvastatin	dabigatran etexilate	1.6
Metformin	Indapamide	1.5
Bisoprolol	Bumetanide	1.5

(b) Top 15 DDIs of ARF ranked by the .25 quantile of sensitivity grid

Drug1	Drug2	Sensitivity adjusted RERI
Lisinopril	Sodium Phosphate, Dibasic	140.4
Heparin	Mannitol	8.8
Potassium Chloride	dabigatran etexilate	3.9
Lisinopril	dabigatran etexilate	2.7
carvedilol	dabigatran etexilate	2.4
Allopurinol	dabigatran etexilate	2.2
Diltiazem	dabigatran etexilate	2
dabigatran etexilate	Spironolactone	1.9
Albuterol	dabigatran etexilate	1.7
Acetaminophen	dabigatran etexilate	1.5
Digoxin	dabigatran etexilate	1.5
Aspirin	dabigatran etexilate	1.5
Simvastatin	dabigatran etexilate	1.5
Metformin	Indapamide	1.4
Heparin	Sodium Bicarbonate	1.4

4.1 Results

Of the 69 candidate DDIs for ARF remaining after pre-filtering, after confounding adjustment only 8 had estimated RERIs ≥ 2 (Table 1(a)). The 0.25 quantile of the sensitivity grid was ≥ 2 for 6 of those 8 (Table 1(b)), and

TABLE 2: Top 15 DDIs for ARF ranked by the lower .95 CI of .25 quantile of sensitivity grid

Drug 1	Drug 2	RERI	Lower .95 CI RERI	Sensitivity adjusted RERI	Sensitivity adjusted RERI (Lower .95 CI)
Potassium Chloride	dabigatran etexilate	4.6	0.9	3.9	1.4
Diltiazem	dabigatran etexilate	2.2	0.6	2.0	1.0
Digoxin	dabigatran etexilate	1.6	0.6	1.5	0.9
Allopurinol	dabigatran etexilate	1.9	0.8	2.2	0.9
Oxycodone	Metoprolol	1.5	0.7	1.3	0.8
Acetaminophen	lansoprazole	1.6	0.7	1.4	0.8
dabigatran etexilate	Omeprazole	1.4	0.6	1.0	0.8
levothyroxine	dabigatran etexilate	1.3	0.7	1.3	0.7
Simvastatin	Citalopram	1.6	0.6	1.5	0.7
Ondansetron	Hydromorphone	1.4	0.5	1.4	0.6
Aspirin	Insulin Glargine	1.6	0.8	1.0	0.6
Potassium Chloride	Esomeprazole	1.5	0.5	0.8	0.6
Warfarin	Bisoprolol	1.4	0.7	1.1	0.5
Ciprofloxacin	Furosemide	1.2	0.6	1.0	0.5
Fentanyl	Furosemide	1.0	0.4	0.9	0.5

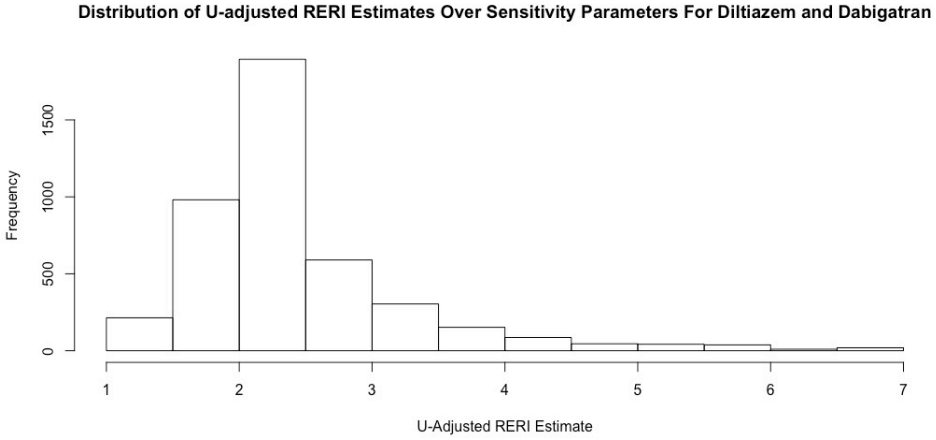

FIGURE 2: Distribution of U-adjusted RERI (of Diltiazen and Dabigatran on ARF) Estimates Over Sensitivity Parameters

the 0.1 quantile of the sensitivity grid was ≥ 2 for 5 of the 8. (For illustration, Figure 2 depicts the distribution of interaction effect estimates on ARF over the grid of sensitivity parameters for dabigatran and diltiazem. Note that under some sensitivity parameters, the RERI would actually be larger, i.e. if the confounding lessened the effect estimate.) For only 1 of the 8 was the 0.95 confidence interval lower bound of the RERI point estimate ≥ 2. Table 2 shows all four scores mentioned in Section 3.6 (i.e., the RERI point estimate, the lower 95% CI of RERI estimate, the .25 quantile of the sensitivity grid, and the lower 95% CI of the .25 quantile of the sensitivity grid) for the top 15 candidate DDIs ranked by the conservative confounding aware approach, i.e, the lower bound of the 95% CI for the 0.25 quantile of the sensitivity grid. No candidate DDI has the lower bound of the 0.25 quantile of its sensitivity grid ≥ 2 (Table 2). Hence, there was not a single candidate DDI that we would confidently declare to mechanistically cause ARF.

Our results were qualitatively different from those obtained by Twosides. Twosides ranked each candidate DDI on a scale from 1 to 5 of confidence that the DDI is real. 3,716 of the 17,945 candidate DDIs potentially causing

TABLE 3: Top 15 DDIs for AMI ranked by the lower .95 CI of .25 quantile of sensitivity grid

Drug 1	Drug 2	RERI	Lower .95 CI RERI	Sensitivity adjusted RERI	Sensitivity adjusted RERI (Lower .95 CI)
bisoprolol	nilotinib	11.1	2.8	9.0	2.2
ivabradine	furosemide	1.6	1.4	1.5	1.3
ivabradine	omeprazole	2.2	1.3	1.9	1.2
insulin glargine	testosterone	1.5	1.2	1.4	1.1
insulin, isophane	furosemide	1.5	1.2	1.4	1.1
bisoprolol	sitagliptin	1.4	1.1	1.3	1.1
sitagliptin	lansoprazole	1.2	1.1	1.1	1.1
sitagliptin	esomeprazole	1.2	1.1	1.1	1.1
allopurinol	darbepoetin alfa	1.4	1.1	1.2	1.0
glimepiride	lansoprazole	1.3	1.1	1.2	1.0
enalapril	ezetimibe	1.8	1.1	1.6	1.1
darbepoetin alfa	lansoprazole	1.7	1.1	1.5	1.0
furosemide	aluminum hydroxide	1.1	1.0	1.1	1.0
esomeprazole	torsemide	1.1	1.0	1.1	1.0
colchicine	pantoprazole	1.1	1.0	1.1	1.0

ARF considered by Twosides were given a score of 5. Since strong DDIs are fairly rare, we believe that this is an implausibly large proportion of drug pairs to mechanistically interact and cause ARF. We would argue our method provided more measured evidence of interaction effects.

Table 3 contains the the top 15 candidate DDIs for Acute Myocardial Infarction (AMI) ranked by the lower 95% CI of the 0.25% estimate of the sensitive grid. Only one out of the 60 initial candidate DDIs (bisoprolol and nilotinib) had a score ≥ 2 by this conservative measure. By contrast, out of 28,683 candidate DDIs for AMI considered by TwoSides, 5,119 were reported to have the highest confidence score of 5 and 19,328 a confidence score ≥ 3. Again, since true causal DDIs are rare, such a large number of DDIs indicates a preponderance of false positive results.

ALI results followed a similar pattern (Table 4), where only one candidate DDI (efavirenz and rifampin) out of the 32 candidate DDIs had lower .95 CI of senstivity adjusted RERI ≥ 2. In contrast, out of 8,479 total candidate DDIs for ALI considered by TwoSides, 1,180 had confidence score 5 and 5,469 had confidence ≥ 3. Once again, this is an implausibly large number.

5 Discussion

One contribution of this work is to emphasize the importance of causality in DDI detection. Prior DDI detection work has not explicitly targeted drug pairs that mechanistically interact to cause adverse outcomes. Yet mechanistic interaction effects (in which at least some instances of the outcome would only occur if both drugs are taken) are most relevant to public health. Hence, we focus on finding drug-drug-outcome triplets with high RERI, since high RERI implies mechanistic interaction.

Another contribution of this work is to point out the importance of sensitivity analysis to DDI detection. It is very unlikely that observed variables are sufficient to adjust for confounding of any given interaction effect. Especially in a limited dataset like FAERS, common causes of each drug and the outcome will almost certainly be omitted. In the absence of sensitivity analysis, this will lead to false positives and overconfidence that DDIs are truly causal. Sensitivity analysis allows us to incorporate the assumption that we are not actually fully adjusting for confounding into our estimates of interaction effects, hence reducing overconfidence and false positive rates.

Even our sensitivity analysis does not account for all possible sources of bias from analyses of FAERS data. For example, selection on reporting of the adverse event may induce collider bias[30]. We may also induce bias by adjusting for certain covariates if the assumed causal structure of Figure 1 does not hold and so-called M-structures[31]

TABLE 4: Top 15 DDIs for ALI ranked by the lower .95 CI of .25 quantile of sensitivity grid

Drug 1	Drug 2	RERI	Lower .95 CI RERI	Sensitivity adjusted RERI	Sensitivity adjusted RERI (Lower .95 CI)
efavirenz	rifampin	279.0	110.2	223.4	88.3
lamivudine	efavirenz	1.4	1.3	1.3	1.2
lamivudine	tenofovir disoproxil	1.4	1.2	1.3	1.2
temozolomide	levetiracetam	1.3	1.1	1.2	1.1
tenofovir disoproxil	efavirenz	1.2	1.1	1.2	1.0
amoxicillin	clindamycin	1.1	1.1	1.1	1.0
heparin	esomeprazole	1.1	1.0	1.0	1.0
tramadol	ramipril	1.1	1.0	1.0	1.0
tramadol	enoxaparin	1.1	1.0	1.0	1.0
acyclovir	tacrolimus	1.0	1.0	1.0	1.0
acetaminophen	cefazolin	1.1	1.0	1.0	1.0
enoxaparin	esomeprazole	1.0	1.0	1.0	1.0
ramipril	lansoprazole	1.0	1.0	1.0	1.0
atenolol	lansoprazole	1.0	1.0	1.0	1.0
aspirin	ceftriaxone	1.0	1.0	1.0	1.0

are present or certain covariates are actually influenced by the treatments.

Still, because of our conservative and rigorous approach to causal interaction and confounding, we believe that the candidate DDIs that we surfaced are worthy of further investigation. Each of the candidates that we surfaced with >2 lower confidence bounds on sensitivity adjusted RERI estimates have plausible biomedical mechanisms. In our candidate AMI DDI, one drug (bisoprolol) is known to increase concentration of the other (nilotinib)[32]. In our candidate DDI for ALI, one drug (efavirenz) is known to cause ALI on its own and higher doses are required when the other drug (rifampin) is being used concurrently[33]. When alternative screening approaches detect thousands of potential DDIs, it is difficult to justify devoting resources to following up on any one in particular.

The methods we apply in this paper are well known in epidemiology but for some reason have not taken hold in the data mining community that has been most active in DDI detection. We demonstrate that careful consideration of causality and bias can fit neatly into a data mining framework and be applied in automated screening tasks at a large scale.

References

1. Eugen Lounkine, Michael J Keiser, Steven Whitebread, Dmitri Mikhailov, Jacques Hamon, Jeremy L Jenkins, Paul Lavan, Eckhard Weber, Allison K Doak, Serge Côté, et al. Large-scale prediction and testing of drug activity on side-effect targets. *Nature*, 486(7403):361–367, 2012.

2. Frank R Ernst and Amy J Grizzle. Drug-related morbidity and mortality: updating the cost-of-illness model. *Journal of the American Pharmaceutical Association (1996)*, 41(2):192–199, 2001.

3. United States National Center for Health Statistics. Health. Number of prescription drugs used in the past 30 days by percentage of the usa population (age-adjusted estimates): Prescription drug use in the past 30 days, by sex, age, race, and hispanic origin: United states, selected years 1988–1994 through 2007–2010. *Trends Pharmacol Sci*, 2012.

4. United States National Center for Health Statistics. Health. The fda adverse event reporting system (faers). *FDA*, 2015.

5. Rave Harpaz, William DuMouchel, Paea LePendu, Anna Bauer-Mehren, Patrick Ryan, and Nigam H Shah. Performance of pharmacovigilance signal-detection algorithms for the fda adverse event reporting system. *Clinical Pharmacology & Therapeutics*, 93(6):539–546, 2013.

6. Rave Harpaz, Herbert S Chase, and Carol Friedman. Mining multi-item drug adverse effect associations in spontaneous reporting systems. In *BMC bioinformatics*, volume 11, page S7. BioMed Central, 2010.

7. Nicholas Tatonetti, Ye Patrick, Roxana Daneshjou, and Russ Altman. Data-driven prediction of drug effects and interactions. *Science translational medicine*, 4(125), 2012.

8. Sanjoy Dey, Heng Luo, Achille Fokoue, Jianying Hu, and Ping Zhang. Predicting adverse drug reactions through interpretable deep learning framework. *BMC bioinformatics*, 19(21):476, 2018.

9. Santiago Vilar, Rave Harpaz, Eugenio Uriarte, Lourdes Santana, Raul Rabadan, and Carol Friedman. Drug—drug interaction through molecular structure similarity analysis. *Journal of the American Medical Informatics Association*, 19(6):1066–1074, 2012.

10. Jialiang Huang, Chaoqun Niu, Christopher Green, Lun Yang, Hongkang Mei, and Jing-Dong Han. Systematic prediction of pharmacodynamic drug-drug interactions through protein-protein-interaction network. *PLoS Comput Biol*, 9(3):e1002998, 2013.

11. Heng Luo, Ping Zhang, Hui Huang, Jialiang Huang, Emily Kao, Leming Shi, Lin He, and Lun Yang. Ddi-cpi, a server that predicts drug–drug interactions through implementing the chemical–protein interactome. *Nucleic Acids Research*, 42(W1).

12. Ping Zhang, Fei Wang, Jianying Hu, and Robert Sorrentino. Label propagation prediction of drug-drug interactions based on clinical side effects. *Scientific Reports*, 5:12339, 2015.

13. Assaf Gottlieb, Gideon Stein, Yoram Oron, Eytan Ruppin, and Roded Sharanb. Indi: a computational framework for inferring drug interactions and their associated recommendations. *Mol Syst Biol*, 8:592, 2012.

14. Feixiong Cheng and Zhongming Zhao. Machine learning-based prediction of drug–drug interactions by integrating drug phenotypic, therapeutic, chemical, and genomic properties. *Journal of the American Medical Informatics Association*, 21(e2):e278–e286, 2014.

15. X Han, SK Quinney, Z Wang, P Zhang, J Duke, Z Desta, JS Elmendorf, DA Flockhart, and L Li. Identification and mechanistic investigation of drug-drug interactions associated with myopathy: A translational approach. *Clinical Pharmacology Therapeutics*, 98(3):321–327, 2015.

16. Ruichu Cai, Mei Liu, Yong Hu, Brittany L Melton, Michael E Matheny, Hua Xu, Lian Duan, and Lemuel R Waitman. Identification of adverse drug-drug interactions through causal association rule discovery from spontaneous adverse event reports. *Artificial intelligence in medicine*, 76:7–15, 2017.

17. Xia Ning, Chi Zhang, Kai Wang, Zhongming Zhao, and Ewy Mathé. The international conference on intelligent biology and medicine 2019: computational methods for drug interactions. *BMC Medical Informatics and Decision Making*, 20(2):1–2, 2020.

18. Niklas Noren, Rolf Sundberg, Andrew Bate, and Ralph Edwards. A statistical methodology for drug–drug interaction surveillance. *Stat Med*, 27(16):3057–3070, 2008.

19. Srinivasan Iyer, Rave Harpaz, Paea LePendu, Anna Bauer-Mehren, and Nigam Shah. Mining clinical text for signals of adverse drug-drug interactions. *J Am Med Inform Assoc*, 21(2):353–62, 2014.

20. R Harpaz, H Perez, HS Chase, R Rabadan, G Hripcsak, and C Friedman. Biclustering of adverse drug events in the fda's spontaneous reporting system. *Clin Pharmacol Ther*, 89(2):243–250, 2011.

21. Su Yan, Xiaoqian Jiang, and Ying Chen. Text mining driven drug-drug interaction detection. In *IEEE Int Conf Bioinformatics Biomed Proceedings*, volume 2013, pages 349–355. IEEE, 2013.

22. Ruichu Cai, Mei Liu, Yong Hu, Brittany Melton, Michael Matheny, Hua Xu, Lian Duan, and Lemuel Waitman. Identification of adverse drug-drug interactions through causal association rule discovery from spontaneous adverse event reports. *Artificial intelligence in medicine*, 76, 2017.

23. Donald Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688, 1974.

24. Tyler VanderWeele and Mirjam Knol. A tutorial on interaction. *Epidemiological Methods*, 3(1):33–72, 2014.

25. Tyler VanderWeele. *Explanation in causal inference: methods for mediation and interaction*. Oxford University Press, 2015.

26. Michael Kuhn, Ivica Letunic, Lars Juhl Jensen, and Peer Bork. The sider database of drugs and side effects. *Nucleic acids research*, 44(D1):D1075–D1079, 2016.

27. Tyler VanderWeele. Principles of confounder selection. *European Journal of Epidemiology*, 34, 2019.

28. Paul Rosenbaum and Donald Rubin. Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society: Series B (Methodological)*, 45(2), 1983.

29. Peng Ding and Tyler VanderWeele. Sensitivity analysis without assumptions. *Epidemiology*, 27(3), 2016.

30. Miguel Hernan, Sonia Hernandez-Diaz, and James Robins. A structural approach to selection bias. *Epidemiology*, pages 615–625, 2004.

31. Wei Liu, Alan Brookhart, Sebastian Schneeweiss, Xiaojuan Mi, and Soko Setoguchi. Implications of m bias in epidemiologic studies: a simulation study. *American journal of epidemiology*, 176(10):938–948, 2012.

32. A Haouala, N Widmer, M Montemurro, T Buclin, and L Decosterd. Drug interactions with the tyrosine kinase inhibitors imatinib, dasatinib, and nilotinib. *Blood*, 117(8):75–87, 2011.

33. LF Lopez-Cortes, R Ruiz-Valderas, P Viciana, A Alarcon-Gonzalez, J Gomez-Mateos, E Leon-Jimenez, M Sarasanacenta, Y Lopez-Pua, and J Pachon. Pharmacokinetic interactions between efavirenz and rifampicin in hiv-infected patients with tuberculosis. *Clinical Pharmacokinetics*, 41(9):681–690, 2002.