Drug Recommendation toward Safe Polypharmacy

Wen-Hao Chiang, and Xia Ning*

***Abstract*—Adverse drug reactions (ADRs) induced from high-order drug-drug interactions (DDIs) due to polypharmacy - simultaneous use of multiple drugs - represent a significant public health problem. Unfortunately, computational effects to facilitate future polypharmacy, particularly to assist safe multi-drug prescription, are still in their infancy. We formally formulate the to-avoid and safe drug recommendation problems when multiple drugs have been taken simultaneously. We develop a joint model with a recommendation component and an ADR label prediction component to recommend for a prescription a set of to-avoid/safe drugs that will induce/will not induce ADRs if taken together with the prescription. We also develop real drug-drug interaction datasets and corresponding evaluation protocols.**

***Index Terms*—Adverse Drug Reaction, Polypharmacy, Drug Recommendation.**

I. Introduction

Adverse drug reactions (ADRs) induced from high-order drug-drug interactions (DDIs) due to polypharmacy - simultaneous use of multiple drugs - represent a significant public health problem. ADRs refer to undesired or harmful reactions due to drug administration. One of the major causes for ADRs is DDIs that happen when the pharmacological effects of a drug are altered by the actions of other drugs, leading to unpredictable clinical consequences. The increasing popularity of polypharmacy continues to expose a significant and growing portion of the population to unknown or poorly understood DDIs and associated ADRs. The National Health and Nutrition Examination Survey [1] reports that more than 76% of the elderly Americans take two or more drugs every day. Another study [2] estimates that about 29.4% of elderly American patients take six or more drugs every day. This has made understanding high-order DDIs and their associated ADRs urgent and important.

Current research on DDIs and their associated ADRs is primarily focused on mining and detecting DDIs for knowledge discovery. In applying the knowledge in practice for preventive, proactive and person-centered healthcare, it also requires predictive power to deal with unknown DDIs, and to provide evidence-based suggestions to facilitate future drug use. As wet-lab based experimental validation scheme for DDI study still falls behind due to its low throughput and lack of scalability, but meanwhile huge amounts of electronic medical record data become increasingly available, data-driven computational methodologies appear more appealing to tackle DDI and ADR problems. Unfortunately, computational effects to facilitate future polypharmacy, particularly to assist safe multi-drug prescription, are still in their infancy.

In this paper, we present a computational method toward the goal of facilitating future safe multi-drug prescription. Here we use the term "prescription" to represent a set of drugs that have been taken together, even though there could be non-prescription drugs. Thus, we tackle the following *to-avoid drug recommendation* problem and *safe drug recommendation* problem.

Definition 1. *To-avoid Drug Recommendation: given the multiple drugs in a prescription that have been taken simultaneously, recommend a short, ranked list of drugs that should be avoided taking together with the prescription in order to avoid a particular ADR.*

Definition 2. *Safe Drug Recommendation: given the multiple drugs in a prescription that have been taken simultaneously, recommend a short, ranked list of safe drugs that, if taken together with the prescription, are not likely to induce a particular ADR.*

Please note that in this study we only consider one ADR and thus drug safety is only considered with respect to that particular ADR. To the best of our knowledge, this is the first work to formally formulate the above problems and provide a computational solution framework. These two recommendation problems are significant particularly for healthcare practice. They can enable evidence-based suggestions to help polypharmacy decision making, and induce novel hypotheses on new high-order DDIs and associated ADRs. Our contributions to solving the two new drug recommendation problems are summarized as follows:

- We formally define the drug recommendation problems.
- We developed a joint sparse linear recommendation and logistic regression model (`SLR`), with a drug recommendation component and an ADR label prediction component (Section V) to solve the problems. The recommendation component captures drug co-prescription patterns among ADR and non-ADR inducing prescriptions, respectively, and uses such patterns to recommend drugs. The ADR label prediction component learns and predicts the ADR probabilities from drugs in prescriptions. These two components are learned concurrently in `SLR` so that the recommended drugs are more likely to introduce expected ADR labels.
- We developed a protocol to mine high-order DDIs and their associated ADRs, and provided a DDI dataset to the public [1] (Section VIII).
- We developed new evaluation protocols and evaluation met-

Wen-Hao Chiang is with the Department of Computer & Information Science, Indiana University - Purdue University Indianapolis, Indianapolis, IN, 46202, USA e-mail: chiangwe@iupui.edu.

Xia Ning is with the Department of Biomedical Informatics, The Ohio State University, Columbus, OH, 43210, USA e-mail: ning.104@osu.edu.

*Corresponding author.

[1]The dataset will be available upon the acceptance of this manuscript.

rics to evaluate the performance of drug recommendation [2] (Section IX).

- We conducted comprehensive experiments to evaluate SLR, and provided case study on recommended drugs (Section X). Our experiments demonstrate strong performance of SLR.

II. Related Work

A. Computational Methods for DDI and ADR Studies

Significant research efforts have focused on detecting DDIs. These efforts can be broadly classified into four categories. Methods of the first category analyze medical literature and/or electronic medical records, and extract mentioned drug pairs [2]. Methods in the second category integrate various biochemical and molecular drug/target data to measure drug-drug similarities and score/predict pairwise DDIs. These data include phenotypic and genomic information [3], and drug side effects [4], etc. Methods of the third category leverage healthcare information on social media and online communities to detect DDIs [5]. The fourth category focuses on using numerical models to predict the dose responses to multiple drugs [6]. A recent thread is dedicated to understanding the interaction patterns among high-order DDIs, and how such patterns can relate to induced ADRs [7].

B. Recommender Systems

Top-N recommender systems, which recommend the top-N items that are most likely to be preferred by users, have been used in a variety of applications in e-commerce and social networking, etc [8]. Recommender systems have been recently applied to prioritizing healthcare information, due to a tremendous need for personalized healthcare [9]. Current applications along this line include recommending physicians to patients on specific diseases [10], and recommending drugs for patient symptoms [11], etc. However, to the best of our knowledge, very little research has been done on new prescription recommendation particularly with ADR concerns.

III. Definitions and Notations

In this paper, we use d to represent a drug, and a binary matrix $A \in \mathbb{R}^{m\times n}$ to represent prescription data. Each row of A represents a prescription, denoted as $\mathbf{a}_i$ ($i = 1, \cdots, m$), and each column of A corresponds to a drug. Thus, there are m prescriptions and n drugs in total represented by A. If a prescription $\mathbf{a}_i$ contains a drug d_j, the j-th entry in $\mathbf{a}_i$ (i.e., $a_{i,j}$) will be 1, otherwise 0. Thus, $\mathbf{a}_i$ is a binary row vector representing the existence of drugs in the i-th prescription. Note that all the prescriptions are unique in A. Drug dosage in prescriptions is not considered in this study. When no ambiguity arises, the terms "a prescription" and "a binary vector $\mathbf{a}$" are used exchangeably, and "a set of prescriptions" and "a matrix A" are also used exchangeably. In addition, a label y_i is assigned to $\mathbf{a}_i$ to indicate whether $\mathbf{a}_i$ induces a certain ADR (denoted as $y_i = 1$ or y_i^+) or not (denoted as $y_i = -1$ or y_i^-). In this paper, we only consider one ADR and thus the labels are binary. In addition, superscript $^+$ indicates information/data related to ADR induction (i.e., positive information), superscript $^-$ indicates information/data related to no ADR induction (i.e., negative information) and accent $\tilde{A}$ indicates estimated A. In this manuscript, all the vectors are by default row vectors and represented using lower-case bold letters (e.g., $\mathbf{a}$); all the matrices are represented using upper-cases letters (e.g., A).

[2]The source code and datasets are publicly available in https://github.com/chiangwe/SLR.

IV. Background

A. Sparse Linear Method for top-N Recommendation

Sparse Linear Method (SLIM) [12] is an efficient and state-of-the-art algorithm for top-N recommendation that was initially designed for e-commerce applications. In the drug recommendation problem, given a drug prescription $\mathbf{a}_i$, SLIM models the score of how likely an additional drug d_j should be co-prescribed with $\mathbf{a}_i$ as a sparse linear aggregation of the drugs in $\mathbf{a}_i$, that is,

$$\tilde{a}_{ij} = \mathbf{a}_i \mathbf{w}_j^\intercal, \tag{1}$$

where $\tilde{a}_{ij}$ is the estimated score of d_j in $\mathbf{a}_i$ (can have values rather than 0 and 1), and $\mathbf{w}_j^\intercal$ is a sparse column vector of aggregation coefficients. Note that $a_{ij} = 0$, that is, d_j is not included $\mathbf{a}_i$ originally. Drugs with high scores calculated as above will be recommended to the prescription. Thus, the scores are referred to as recommendation scores, and a prescription composed of $\mathbf{a}_i$ and a recommended drug d_j is referred to as a new prescription with respect to $\mathbf{a}_i$, denoted as $\mathbf{a}_i \cup \{d_j\}$. The intuition of using SLIM for drug recommendation will be discussed later in Section V-A.

To learn $W = [\mathbf{w}_1^\intercal, \mathbf{w}_2^\intercal, \cdots, \mathbf{w}_n^\intercal]$, SLIM solves the following optimization problem,

$$\begin{aligned} \min_{W} \quad & \texttt{SLIM}(A; W, \alpha, \lambda) = \frac{1}{2}\|A - AW\|_F^2 + \frac{\alpha}{2}\|W\|_F^2 + \lambda\|W\|_{\ell_1} \\ \text{subject to} \quad & W \geq 0, \text{diag}(W) = 0, \end{aligned} \tag{2}$$

where $\|W\|_{\ell_1} = \sum_{i=1}^n \sum_{i=1}^n |w_{ij}|$ is the entry-wise ℓ_1-norm of W, and $\|\cdot\|_F$ is the matrix Frobenius norm. The SLIM formulation can also be considered as a way to reconstruct A by itself through learning the patterns (i.e., W) in A. Therefore, W converts a binary A into its estimation $\tilde{A}$ of floating values, which could recover unseen non-zero entries in A.

B. Logistic Regression for Label Prediction

We can formulate the problem of predicting whether a prescription of multiple drugs induces a particular ADR as a binary classification problem, and solve the problem using logistic regression (LogR). In LogR, the probability of a prescription $\mathbf{a}_i$ inducing the ADR is modeled as follows,

$$p(y_i|\mathbf{a}_i; \mathbf{x}, c) = (1 + \exp(-y_i(\mathbf{a}_i\mathbf{x}^\intercal + c)))^{-1}, \tag{3}$$

where $\mathbf{x}^\intercal$ and c are the parameters. To learn the parameters, LogR solves the following optimization problem,

$$\begin{aligned} \min_{\mathbf{x},c} \quad & \texttt{LogR}(\mathbf{y}|A; \mathbf{x}, c, \beta, \gamma) \\ & = \sum\nolimits_{i=1}^m \log\{1 + \exp[-y_i(\mathbf{a}_i\mathbf{x}^\intercal + c)]\} + \frac{\beta}{2}\|\mathbf{x}\|_2^2 + \gamma\|\mathbf{x}\|_1, \end{aligned} \tag{4}$$

where $\mathbf{y}$=$[y_1; y_2; \cdots; y_m]$, $\|\mathbf{x}\|_1$=$\sum_{i=1}^n |x_i|$, and $\|\mathbf{x}\|_2^2$=$\sum_{i=1}^n {x_i}^2$.

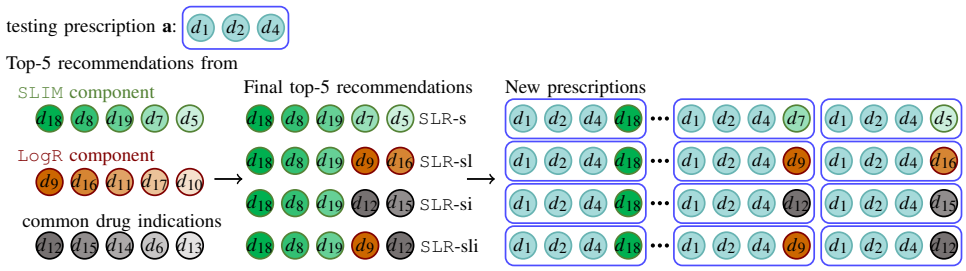

Fig. 1: SLR recommendation methods

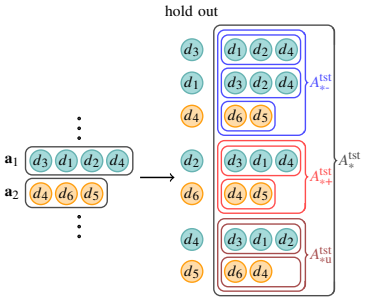

Fig. 2: Testing set generation

V. Joint SLIM and LogR Model: SLR

We propose a new model to conduct the task of to-avoid and safe drug recommendation to existing prescriptions. The model recommends a set of additional to-avoid drugs to an existing prescription such that each of the recommended drugs should not be taken together with the prescription in order to avoid ADRs. The model also recommends safe drugs that can be taken together with the prescription. This novel model consists of a drug recommendation component and an ADR prediction component. The recommendation component is instantiated from SLIM (Section IV-A), and the ADR prediction component uses logistic regression LogR (Section IV-B). This model is referred to as joint sparse linear recommendation and logistic regression model and denoted as SLR.

SLR learns the SLIM and LogR components jointly through solving the following optimization problem:

$$\begin{aligned} \min_{W^+, W^-, \mathbf{x}, c} \quad & \texttt{SLIM}(A^+; W^+, \alpha, \lambda) + \texttt{SLIM}(A^-; W^-, \alpha, \lambda) + \\ & \omega\{\texttt{LogR}(\mathbf{y}^+ | \tilde{A}^+; \mathbf{x}, c, \beta, \gamma) + \texttt{LogR}(\mathbf{y}^- | \tilde{A}^-; \mathbf{x}, c, \beta, \gamma)\} \\ \text{subject to} \quad & \tilde{A}^+ = A^+ W^+, \tilde{A}^- = A^- W^-, \\ & W^+ \geq 0, W^- \geq 0, \text{diag}(W^+) = 0, \text{diag}(W^-) = 0. \end{aligned} \tag{5}$$

In SLR, A^+/A^- is a set of training prescriptions known to induce ADRs/not to induce ADRs, respectively, and $\tilde{A}^+/\tilde{A}^-$ is the respective estimation from a SLIM model (Equation 1), that is, $\tilde{A}^+ = A^+W^+$ and $\tilde{A}^- = A^-W^-$, where W^+ and W^- are the SLIM model parameters. The optimization algorithm for problem 5 is presented in Section I in the supplementary materials.

A. Learning Co-Prescription Patterns

SLR learns drug co-prescription patterns using its SLIM component (i.e., $\tilde{\mathbf{a}} = \mathbf{a}W$, or explicitly, $\tilde{a}_{i,j} = \sum_{k,k\neq j} a_{i,k} W_{k,j}$), that is, whether a drug should be included in a prescription is modeled as a linear function of other drugs that are included in the prescription. This modeling scheme is motivated by the existence of strong co-prescription drug pairs as we observe from real data (Section X-E). Meanwhile, linearity is a simplistic relation to model the co-prescription patterns, in which larger coefficients represent stronger co-prescription relations. Note that the zero-diagonal constraint on the coefficient matrix W in SLIM (i.e., $\text{diag}(W^+) = 0$, $\text{diag}(W^-) = 0$ in Equation 5) excludes the possibility that a drug is co-prescribed with itself. In addition, the non-negativity constraint on W (i.e., $W^+ \geq 0$, $W^- \geq 0$ in Equation 5) ensures that the learned relation is on drug co-appearance. The regularization term $\|W\|_{\ell_1}$ induces sparsity in W because not all the drugs tend to be co-prescribed.

In SLR, the patterns in ADR inducing and non-ADR inducing prescriptions are modeled using two SLIM models (i.e., $\tilde{A}^+ = A^+W^+$ and $\tilde{A}^- = A^-W^-$) because it is expected these two patterns are different, and thus the prospectively learned W^+ and W^- will present different patterns. The pattern difference is indicated during the data pre-processing (Section VIII-A, Table I). The W coefficients in the SLIM component are learned by minimizing the difference between A and $\tilde{A}$ estimated from SLIM. Note that A is a binary matrix but $\tilde{A}$ can have floating values rather than 0 and 1. Once the linear aggregation parameter W is learned, it can be used to recommend additional drugs that are likely to be co-prescribed with an existing prescription.

B. Predicting ADR Labels

SLR uses its LogR component (i.e., $p(y_i|\tilde{\mathbf{a}}_i; \mathbf{x}, c) = (1 + \exp(-y_i(\tilde{\mathbf{a}}_i \mathbf{x}^\intercal + c)))^{-1}$ to predict whether a prescription will induce ADRs or not. LogR produces a probability of ADR induction from a linear combination of drugs in a prescription (Equation 3). Note that the estimated prescriptions $\tilde{A}$ from SLIM rather than A is used to train LogR. This connects SLIM and LogR to further enforce that the SLIM component learns co-prescription patterns that better correlate with their ADR induction labels (i.e., $\mathbf{y}$ in Equation 5). Meanwhile, the use of $\tilde{A}$ also generalizes the ability of LogR to predict for new prescriptions, as $\tilde{A}$ will have new drugs compared to A (as discussed in Section IV-A).

VI. SLR Drug Recommendation

SLR recommends drugs for each existing prescription $\mathbf{a}$ following a five-step procedure, resulting in four different recommendation methods. We will discuss these steps and the corresponding recommendation methods in detail later in this section. Note that in SLR, each recommended drug is considered individually together with the prescription $\mathbf{a}$. It is also possible in real practice that multiple drugs all together should be avoided as they, together with the prescription, may induce ADRs. However, this problem is significantly

more difficult as the combinatorial synergy of high-order drug combinations is highly non-trivial to estimate. This problem will be tackled in our future research.

A. Step 1: Recommendation from SLIM component

Given a prescription $\mathbf{a}$, the recommendation scores of all drugs with respect to $\mathbf{a}$ are calculated as $\tilde{\mathbf{a}}^+ = \mathbf{a}W^+$ and $\tilde{\mathbf{a}}^- = \mathbf{a}W^-$, that is, $\tilde{\mathbf{a}}^+/\tilde{\mathbf{a}}^-$, the scores of being potential to-avoid/safe drugs, are calculated from W^+/W^-. Based on $\tilde{\mathbf{a}}^+$ and $\tilde{\mathbf{a}}^-$, the top-N scored drugs that are not in prescription $\mathbf{a}$ will be selected as to-avoid and safe drug candidates for $\mathbf{a}$, respectively. This set of N drugs is denoted as $\{d_{r_i}\}_{\texttt{SLIM}}$ ($i = 1 \cdots N$). The recommendation process in this step is identical to that in the original SLIM method [12]. However, the SLIM component is learned from SLR (Equation 5). This recommendation method, together with the SLR learning method, is referred to as SLR-s.

B. Step 2: Removing low-frequency recommendations

The drugs in $\{d_{r_i}\}_{\texttt{SLIM}}$ (i.e., the recommendation from the SLIM component in SLR) are removed if their co-prescription frequency with $\mathbf{a}$ is lower than a threshold η. Specifically, for each drug d_{r_i} and each drug $d \in \mathbf{a}$, we count how many times both of them are prescribed together within a prescription and sum up the counts over all the drugs in $\mathbf{a}$ as the co-prescription frequency for drug d_{r_i}. The intuition underlying the low-frequency drug removal is to explicitly follow the co-prescription patterns that have been often observed in standard of care. The reduced drug recommendation list is denoted as $\{d_{r_i}\}_{\texttt{SLIM}_\eta}$.

C. Step 3: Recommendation from LogR component

For a prescription $\mathbf{a}$, each of all possible drugs that are not in $\mathbf{a}$ is first combined with $\mathbf{a}$ to form a new prescription. Then the LogR component is applied on the new prescription to calculate its probability of inducing ADRs (Equation 3). The drugs that lead to the top-N highest/lowest predicted probabilities are recommended as the to-avoid/safe drugs. Note that the LogR component is learned from SLR. The generated recommendation list is denoted as $\{d_{r_j}\}_{\texttt{LogR}}$.

D. Step 4: Recommendation based on drug indications

In addition to recommendations from SLIM and LogR components, we recommend another list of drugs based on drug indications. The intuition is that drugs that have similar indications tend to be prescribed together. Table S1 in the supplementary materials presents some examples of co-prescribed drugs with similar indications. For each drug $d \notin \mathbf{a}$ and each drug $d_i \in \mathbf{a}$, we first calculate the number of common indications between d and d_i, and sum up all such numbers over d_i as the number of common indications for d with $\mathbf{a}$. The drugs which have the top-N most common indications with $\mathbf{a}$ form a recommendation list, denoted as $\{d_{r_k}\}_{\text{ind}}$. The drug indication information is extracted from the Side Effect Resource (SIDER) [13].

E. Step 5: Combining recommendations

To recommend top-N to-avoid/safe drugs, we combine drugs from the list generated by different components, leading to the following four methods.

- SLR-s method: use the recommendation list $\{d_{r_i}\}_{\texttt{SLIM}}$ generated from Step 1 as the final recommendation list;
- SLR-sl method: use the top-N drugs from the concatenation of $\{d_{r_i}\}_{\texttt{SLIM}_\eta}$ and $\{d_{r_j}\}_{\texttt{LogR}}$ as the final recommendation list.
- SLR-si method: use the top-N drugs from the concatenation of $\{d_{r_i}\}_{\texttt{SLIM}_\eta}$ and $\{d_{r_k}\}_{\text{ind}}$ as the final recommendation list.
- SLR-sli method: use the top-N drugs from the concatenation of $\{d_{r_i}\}_{\texttt{SLIM}_\eta}$, the top $\{d_{r_j}\}_{\texttt{LogR}}$ and the top $\{d_{r_j}\}_{\text{ind}}$ as the final recommendation list.

Figure 1 demonstrates the four different methods.

VII. Comparison Methods

A. Random Method (Rand)

In the first comparison method, we fully randomly recommend drugs for each prescription as to-avoid drugs and safe drugs, respectively. This method is referred to as random method and denoted as Rand. Rand serves as a baseline in which no learning is involved.

B. Logistic Regression Only (LogR)

In this comparison method, only a LogR model is learned to predict ADR labels. Prescriptions that induce ADRs are used as positive training data, and prescriptions that do not induce ADRs are used as negative training data. No SLIM models are learned in this method. The recommendation in this baseline method is the same as the procedure to generate a LogR recommendation list in SLR (Section VI-A). This method is referred to as logistic regression only and denoted as LogR.

C. Sparse Linear Method Only (SLIM)

In this method, a $\texttt{SLIM}^+$ model is learned on ADR inducing prescriptions, and a separate $\texttt{SLIM}^-$ model is learned on non-ADR inducing prescriptions. No LogR is learned. For a prescription, the top drugs recommended from $\texttt{SLIM}^+$ are considered as to-avoid drugs, and the top drugs recommended from $\texttt{SLIM}^-$ are considered as safe drugs. This method is referred to as SLIM only and denoted as SLIM.

D. Separate Models (S+LR)

In this method, a SLIM and a LogR model are learned separately, and used together for recommendation. Specifically, a $\texttt{SLIM}^+$ and a $\texttt{SLIM}^-$ as in SLIM are learned first. Meanwhile, a LogR model as in LogR is also learned independently. The top recommendations are then generated by combining recommendations from $\texttt{SLIM}^+/\texttt{SLIM}^-$ and LogR model in an identical way as to that of SLR-sl. The difference is that SLIM and LogR are learned independently, not jointly as in SLR. This method is referred to as separate SLIM-LogR model and denoted as S+LR.

VIII. Materials

A. Training Data Generation

We use the prescription dataset A_{FAERS} that we generated from FDA Adverse Event Reporting System (FAERS) [3] in our previous research [14] and consider myopathy as the particular ADR of interest in this work. Detailed description of dataset generation protocol is available in section 5.1 of [14]. Table I summarizes the dataset. Specifically, A_{FAERS} consists of four subsets of prescriptions: 1). A_0^+ has all the prescriptions that are reported both with myopathy (case events) among some patients and without myopathy (control events) among other patients, and the Odds Ratio (OR) between these two cases is above 1, indicating that more possibly these prescriptions will induce myopathy; 2). A_0^- has all the prescriptions that are also reported both with and without myopathy as those in A_0^+, but the OR is below 1, indicating that more possibly these prescriptions will not induce myopathy; 3). A_+^+ consists of prescriptions that only have reports of myopathy (i.e., only in case events); and 4). A_-^- consists of prescriptions that only have reports of non-myopathy (i.e., only in control events). The set of prescriptions in case events is denoted as A^+ (i.e., $A^+ = A_+^+ \cup A_0^+$) and is used as the positive set in A_{FAERS}. The set of prescriptions in control events is denoted as A^- (i.e., $A^- = A_-^- \cup A_0^-$) and is used as the negative set in A_{FAERS}.

To use more frequent and more confident prescriptions for model training, we further generated a prescription dataset from A_{FAERS}. The detailed generation protocol is presented in section 5.2 in [14]. This dataset is denoted as A_* and summarized in Table I. Note that A_* is the set of labeled prescriptions that are used for model training.

TABLE I: Data Statistics [14]

dataset	stats	A^-		A^+	
		A_-^-	A_0^-	A_0^+	A_+^+
A_{FAERS}	#{**a**}	621,449	1,264	8,986	27,387
	#{d}	1,209	417	881	1,201
	avgOrd	6.100	2.351	3.588	7.096
	avgFrq	1.761	225.317	13.730	1.402
	avgOR	-	0.546	16.343	-
A_*	#{**a**}	2,200	1,264	2,464	1,000
	#{d}	562	417	692	679
	avgOrd	2.678	2.351	3.809	7.615
	avgFrq	42.082	225.317	20.565	5.520
	avgOR	-	0.546	31.998	-

In this table, "#{**a**}" and "#{d}" represent the number of prescriptions and the number of involved drugs, respectively. In each set of prescriptions, "avgOrd"/"avgFrq"/"avgOR" is the average number of drugs/average frequency/average OR of all prescriptions.

IX. Evaluation Protocols and Metrics

A. Five-Fold Cross Validation

The performance of different methods is evaluated via five-fold cross validation. The dataset A_* is randomly split into five folds of equal size. Four folds are used for model training and the rest fold is used for testing. The corresponding training set is denoted as A_*^{trn}. In the fold for testing, each prescription of order k ($k = 2, 3, \cdots$) is first duplicated into k copies. Then each of the k drugs is excluded from each of the k duplicates, respectively, to generate a testing prescription of order k-1. We further unify these testing prescriptions and exclude those testing prescriptions that appear in A_*^{trn}. The remaining testing prescriptions constitute the testing set, denoted as A_*^{tst}. Figure 2 demonstrates the testing set generation process. The drug recommendation will be conducted and evaluated on A_*^{tst} prescriptions for five times, with one fold for testing each time. The final result is the average of the five experiments.

[3]https://www.fda.gov/drugs/informationondrugs/ucm135151.htm

B. Knowledge Pool Generation

To evaluate the performance of various methods, we first define a knowledge pool of labeled prescriptions, denoted as A_{pool}, which the recommended new prescriptions can be searched from and evaluated against. This knowledge pool serves as the entire knowledge space which is unknown during training. For a training set A_*^{trn}, we consider $A_{\text{FAERS}} \setminus A_*^{\text{trn}}$ as the knowledge pool A_{pool}. A_*^{trn} is excluded from A_{pool} because we can always search testing prescriptions in A_*^{trn}, where prescriptions have known labels at the time of training, in order to identify corresponding to-avoid prescriptions in A_*^{trn}, and therefore no recommendations will be needed. By excluding A_*^{trn} from A_{pool}, we care about recommending particularly to-avoid/safe drugs to prescriptions such that the new prescriptions have not been observed before. We believe this application is of more interest and significance in real practice.

C. Evaluation Metrics

Each of the methods will generate N recommended to-avoid/safe drugs for each testing prescription. Given a knowledge pool A_{pool}, the performance of these methods is evaluated using *Normalized Recall on Positives/Negatives*, denoted as $\overline{\texttt{rec}_\text{P}}/\overline{\texttt{rec}_\text{N}}$, defined as follows,

$$\texttt{rec}_\text{P} = \#\text{TP}/\#\text{P}, \quad \overline{\texttt{rec}_\text{P}} = \texttt{rec}_\text{P}/\max(\texttt{rec}_\text{P}), \tag{6}$$

$$\texttt{rec}_\text{N} = \#\text{TN}/\#\text{N}, \quad \overline{\texttt{rec}_\text{N}} = \texttt{rec}_\text{N}/\max(\texttt{rec}_\text{N}), \tag{7}$$

where #TP/#TN is the number of recommended to-avoid/safe prescriptions for A_*^{tst} that have positive/negative labels in A_{pool}, #P/#N is the number of all possible ground-truth to-avoid/safe prescriptions for A_*^{tst} that have positive/negative labels in A_{pool}, and $\max(\texttt{rec}_\text{P})/\max(\texttt{rec}_\text{N})$ is the maximum possible $\texttt{rec}_\text{P}/\texttt{rec}_\text{N}$ value. In particular, $\max(\texttt{rec}_\text{P})/\max(\texttt{rec}_\text{N})$ is achieved when all the top-N recommendations lead to true-positive/true-negative (to-avoid/safe) prescriptions. Note that $\max(\texttt{rec}_\text{P})/\max(\texttt{rec}_\text{N})$ can be smaller than 1. This is because for each testing prescription, the number of recommended prescriptions is limited to N and can be smaller than the number of all its possible ground-truth to-avoid/safe prescriptions in A_{pool}. Thus, $\overline{\texttt{rec}_\text{P}}$ and $\overline{\texttt{rec}_\text{N}}$ measure how much each method can achieve to its best. Higher $\overline{\texttt{rec}_\text{P}}$ and $\overline{\texttt{rec}_\text{N}}$ values indicate better performance. In particular, higher $\overline{\texttt{rec}_\text{P}}$ indicates stronger performance on recommending to-avoid drugs, and higher $\overline{\texttt{rec}_\text{N}}$ indicates stronger performance on recommending safe drugs. The maximum possible values of $\texttt{rec}_\text{P}$ and $\texttt{rec}_\text{N}$ (i.e., $\max(\texttt{rec}_\text{P})$ and $\max(\texttt{rec}_\text{N})$) are presented in Table S2 in the supplimentary materials.

TABLE II: Overall Performance Comparison on A_*^{tst}

mdl	best $\overline{\text{rec}_\text{P}}$			best $\overline{\text{rec}_\text{N}}$			best HM_rec		
	$\overline{\text{rec}_\text{P}}$	$\overline{\text{rec}_\text{N}}$	HM_rec	$\overline{\text{rec}_\text{P}}$	$\overline{\text{rec}_\text{N}}$	HM_rec	$\overline{\text{rec}_\text{P}}$	$\overline{\text{rec}_\text{N}}$	HM_rec
Rand	0.0020	0.0003	0.0006	0.0020	0.0003	0.0006	0.0020	0.0003	0.0006
LogR	0.1621	0.2320	0.1908	0.1614	0.2397	0.1927	0.1614	0.2397	0.1927
SLIM	0.2564	0.3739	0.3041	0.2072	0.3747	0.2668	0.2564	0.3739	0.3041
S+LR	**0.2662**	0.3604	0.3061	0.2574	0.3777	0.3061	0.2649	0.3666	0.3074
SLR-s	0.2576	0.3739	0.3049	0.2468	0.3823	0.2999	0.2574	0.3755	0.3053
SLR-sl	0.2633	0.3381	0.2959	0.2608	0.3910	0.3127	0.2608	0.3910	0.3127
SLR-si	0.2583	0.3943	0.3121	0.2472	**0.3978**	0.3048	0.2583	0.3943	0.3121
SLR-sli	0.2654	0.3836	0.3136	0.2474	**0.3978**	0.3050	0.2654	0.3836	**0.3136**

The column "mdl" corresponds to models. The best overall performance is **bold**.

The third metric is the harmonic mean of $\overline{\text{rec}_\text{P}}$ and $\overline{\text{rec}_\text{N}}$, that is,

$$\text{HM}_\text{rec} = 2 \times \overline{\text{rec}_\text{P}} \times \overline{\text{rec}_\text{N}} / (\overline{\text{rec}_\text{P}} + \overline{\text{rec}_\text{N}}), \tag{8}$$

and is denoted as HM_rec. Higher HM_rec values indicate better performance. Note that we did not use conventional ranking-based metrics such as average precision at k because we believe in clinical practice of multi-drug prescription, people care whether all the to-avoid drugs can be correctly recommended (i.e., a notion of recall) much more than how the to-avoid drugs are ranked.

X. Experimental Results

A. Overall Comparison

We present the overall performance comparison of all the methods in Table II. Note that Table II presents the performance of different methods in terms of their best $\overline{\text{rec}_\text{P}}$, best $\overline{\text{rec}_\text{N}}$ and HM_rec, respectively, and lists the corresponding values on the other two evaluation metrics when each respective best performance is achieved. Optimal parameters are not presented in the tables due to space limit. Overall, recommendation-based methods (i.e., SLIM, S+LR and SLR-based methods) substantially outperform LogR and Rand. LogR learns and predicts the labels of prescriptions, but it does not learn or leverage the patterns among the drugs in prescriptions, even though LogR learns the labels substantially better than random (Rand). In terms of best $\overline{\text{rec}_\text{P}}$, SLIM, S+LR, SLR-s, SLR-sl, SLR-si and SLR-sli are very comparable (all around 0.26), with S+LR ($\overline{\text{rec}_\text{P}}$ = 0.2662) slightly better than the others. In terms of best $\overline{\text{rec}_\text{N}}$, SLR-sl ($\overline{\text{rec}_\text{N}}$ = 0.3910), SLR-si and SLR-sli ($\overline{\text{rec}_\text{N}}$ = 0.3978) are comparable. SLR-si and SLR-sli are 5.3% better than S+LR and 6.2% better than SLIM, where both of S+LR and SLIM learn a SLIM component independently. In terms of best HM_rec, SLR-sl (HM_rec =0.3127) and SLR-sli (HM_rec = 0.3136) are also comparable. SLR-sli is 2.0% better than S+LR and 3.1% better than SLIM. Overall, SLR-sl and SLR-sli are very comparable, and both are slightly better than SLR-s and SLR-si. SLR-based methods (i.e., SLR-s, SLR-sl, SLR-si and SLR-sli) outperform SLIM, a strong baseline for top-N recommendation but learned without consideration of recommendation labels. SLR-based methods also outperform S+LR, which considers recommendation labels but learns labels independently of recommendations. These results demonstrate the strong potential of SLR methods in recommending to-avoid/safe drugs for existing prescriptions. Detailed parameter study is presented in Section III-B and Section III-C in the supplementary materials.

B. Comparison on Different Testing Subsets

Among all the testing prescriptions in A_*^{tst}, some prescriptions can be found in A_{FAERS} along with known ADR labels, while the others have unknown ADR labels. We separate A_*^{tst} into the following subsets according to their labels in A_{FAERS}:

- A_{*+}^{tst}: the set of prescriptions in A_*^{tst} that have positive labels in A_{FAERS};
- A_{*-}^{tst}: the set of prescriptions in A_*^{tst} that have negative labels in A_{FAERS};
- $A_{*\text{u}}^{\text{tst}}$: the set of prescriptions in A_*^{tst} that do not have labels in A_{FAERS}.

Thus, $A_{*+}^{\text{tst}} \subset A_{\text{FAERS}}$, $A_{*-}^{\text{tst}} \subset A_{\text{FAERS}}$, $A_{*\text{u}}^{\text{tst}} \not\subset A_{\text{FAERS}}$, $A_{*+}^{\text{tst}} \cap A_{*-}^{\text{tst}} = \emptyset$ and $A_{*+}^{\text{tst}} \cup A_{*-}^{\text{tst}} \cup A_{*\text{u}}^{\text{tst}} = A_*^{\text{tst}}$. Note that if a prescription with an unknown label contains only one drug and can be found in $\mathcal{D}_{\text{Myo}}$ (i.e., the set of drugs which induce myopathy on their own), we still categorize it as in A_{*+}^{tst}. Different subsets here correspond to different concerns of interest. For A_{*+}^{tst}, we are more concerned with safe drug recommendations (i.e., $\overline{\text{rec}_\text{N}}$ in evaluation) so as not to further increase the ADR risks of prescriptions. For A_{*-}^{tst}, we focus more on to-avoid drug recommendation (i.e., $\overline{\text{rec}_\text{P}}$ in evaluation) so as not to introduce ADRs in prospective prescriptions. For $A_{*\text{u}}^{\text{tst}}$, we are interested in both to-avoid and safe drug recommendations. Table III presents the statistics of different sets of testing prescriptions. In particular, the column $\mathcal{D}_{\text{Myo}}(\%)$ in Table III presents the average percentage of the $\mathcal{D}_{\text{Myo}}$ drugs in a prescription. In specific, we first calculate the percentage of $\mathcal{D}_{\text{Myo}}$ drugs in each prescription and take the average. A_{*+}^{tst}, in which all the prescriptions are positive, contains the most $\mathcal{D}_{\text{Myo}}$ drugs on average.

TABLE III: Statistics on Different Testing Subsets

Subset	#{**a**}	avgOrd	$\mathcal{D}_{\text{Myo}}$ (%)
A_{*+}^{tst}	2,182	2.9	25.51
A_{*-}^{tst}	3,057	3.0	19.22
$A_{*\text{u}}^{\text{tst}}$	12,065	10.4	16.14
A_*^{tst}	17,304	7.8	17.87

" #{**a**}" represents the total number of unique prescriptions; "avgOrd" is the average order of prescriptions; and "$\mathcal{D}_{\text{Myo}}$ (%)" is the average percentage of $\mathcal{D}_{\text{Myo}}$ drugs in prescriptions.

1) Performance Comparison on A_{*+}^{tst}: Table IV presents the performance of different methods on A_{*+}^{tst}. In terms of best $\overline{\text{rec}_\text{P}}$, SLR-based methods and S+LR have similar performance (around 0.2528) and slightly better than SLIM. In terms of best $\overline{\text{rec}_\text{N}}$, SLR-based methods and SLIM have similar performance (around 0.41), with SLR-si and SLR-sli slightly better than the others. In terms of best HM_rec, still the SLR-based methods are similar to SLIM. These results show that for A_{*+}^{tst}, that is, the set of prescriptions that induce ADRs, joint learning SLIM and LogR does not introduce additional benefits on top of SLIM, although all these methods are significantly better than Rand and LogR. However, comparing the best $\overline{\text{rec}_\text{N}}$ values in Table IV and those in Table II, SLR-based methods have better performance (best $\overline{\text{rec}_\text{N}}$ =0.4179) on A_{*+}^{tst} than on A_*^{tst} (best $\overline{\text{rec}_\text{N}}$ =0.3978). This indicates that for ADR-inducing prescriptions, our methods are able to recommend safe drugs that taken together with the ADR-

inducing prescriptions would actually lead to no/less ADR effects.

TABLE IV: Performance Comparison on A^{tst}_{*+}

mdl	best $\overline{\text{rec}_\text{P}}$			best $\overline{\text{rec}_\text{N}}$			best HM_rec		
	$\overline{\text{rec}_\text{P}}$	$\overline{\text{rec}_\text{N}}$	HM_rec	$\overline{\text{rec}_\text{P}}$	$\overline{\text{rec}_\text{N}}$	HM_rec	$\overline{\text{rec}_\text{P}}$	$\overline{\text{rec}_\text{N}}$	HM_rec
Rand	0.0010	0.0003	0.0004	0.0010	0.0003	0.0004	0.0010	0.0003	0.0004
LogR	0.1964	0.1796	0.1875	0.1964	0.1796	0.1875	0.1964	0.1796	0.1875
SLIM	0.2519	0.4110	0.3123	0.2519	0.4110	0.3123	0.2519	0.4110	0.3123
S+LR	0.2528	0.4046	0.3111	0.2528	0.4046	0.3111	0.2528	0.4046	0.3111
SLR-s	**0.2530**	0.4111	0.3131	0.2486	0.4138	0.3104	0.2530	0.4111	0.3131
SLR-sl	0.2529	0.4119	0.3133	0.2493	0.4113	0.3104	0.2529	0.4119	0.3133
SLR-si	0.2528	0.4150	0.3141	0.2482	**0.4179**	0.3112	0.2528	0.4150	**0.3141**
SLR-sli	**0.2530**	0.4131	0.3137	0.2482	**0.4179**	0.3112	0.2530	0.4131	0.3137

The column "mdl" corresponds to models. The best overall performance is **bold**.

2) Performance Comparison on A^{tst}_{*-}: Table V presents the performance of different methods on A^{tst}_{*-}. The SLR-based methods and SLIM have a similar performance in terms of their best $\overline{\text{rec}_\text{P}}$, $\overline{\text{rec}_\text{N}}$ and HM_rec. However, comparing the results in Table V with the results in Table IV and Table II, SLR-based methods can achieve even better $\overline{\text{rec}_\text{P}}$ values ($\overline{\text{rec}_\text{P}}$ =0.2847) on A^{tst}_{*-} than those on A^{tst}_{*+} ($\overline{\text{rec}_\text{P}}$ =0.2530 and A^{tst}_{*} ($\overline{\text{rec}_\text{P}}$ =0.2662). As for A^{tst}_{*-}, we care more on recommending to-avoid drugs. The better $\overline{\text{rec}_\text{P}}$ values on A^{tst}_{*-} indicate the capability of SLR-based methods in recommending to-avoid drugs that taken together with the non ADR-inducing prescriptions would not induce ADRs. In addition, SLR-based methods also have better performance on A^{tst}_{*-} in terms of $\overline{\text{rec}_\text{N}}$ and HM_rec, compared to that on A^{tst}_{*+} and A^{tst}_{*}.

TABLE V: Performance Comparison on A^{tst}_{*-}

mdl	best $\overline{\text{rec}_\text{P}}$			best $\overline{\text{rec}_\text{N}}$			best HM_rec		
	$\overline{\text{rec}_\text{P}}$	$\overline{\text{rec}_\text{N}}$	HM_rec	$\overline{\text{rec}_\text{P}}$	$\overline{\text{rec}_\text{N}}$	HM_rec	$\overline{\text{rec}_\text{P}}$	$\overline{\text{rec}_\text{N}}$	HM_rec
Rand	0.0000	0.0004	0.0000	0.0000	0.0004	0.0000	0.0000	0.0004	0.0000
LogR	0.1423	0.1409	0.1414	0.1410	0.1461	0.1431	0.1412	0.1460	0.1432
SLIM	0.2804	0.4499	0.3454	0.2792	0.4508	0.3448	0.2804	0.4499	0.3454
S+LR	0.2808	0.3740	0.3208	0.2790	0.3745	0.3197	0.2808	0.3740	0.3208
SLR-s	0.2844	0.4525	0.3492	0.2844	**0.4525**	0.3492	0.2844	0.4525	0.3492
SLR-sl	**0.2847**	0.4525	0.3494	0.2847	**0.4525**	0.3494	0.2847	0.4525	**0.3494**
SLR-si	0.2842	0.4407	0.3455	0.2842	0.4407	0.3455	0.2842	0.4407	0.3455
SLR-sli	**0.2847**	0.4407	0.3459	0.2847	0.4407	0.3459	0.2847	0.4407	0.3459

The column "mdl" corresponds to models. The best overall performance is **bold**.

3) Performance Comparison on A^{tst}_{*u}: Table VI presents the performance of different methods on $A^{\text{tst}}_{*\text{u}}$. Different from the performance on A^{tst}_{*+} and A^{tst}_{*-}, on $A^{\text{tst}}_{*\text{u}}$, S+LR has the best performance in terms of $\overline{\text{rec}_\text{P}}$ ($\overline{\text{rec}_\text{P}}$ =0.2745) compared to SLR-sli, the best performing SLR-based methods ($\overline{\text{rec}_\text{P}}$ =0.2700). SLIM is worse than SLR-based methods in terms of $\overline{\text{rec}_\text{P}}$. However, in terms of best $\overline{\text{rec}_\text{N}}$, on $A^{\text{tst}}_{*\text{u}}$, SLR-sli and SLR-sl are significantly better than other methods, followed by SLR-si and SLR-s, while SLIM and LogR do not perform comparably. A similar trend holds in terms of best HM_rec. The difference of the performance trends on $A^{\text{tst}}_{*\text{u}}$ compared to A^{tst}_{*+} and A^{tst}_{*-} may be due to the different characteristics of $A^{\text{tst}}_{*\text{u}}$ (Table III). For example, the larger size of drug combinations may allow LogR alone to better predict prescription labels, particularly the positive labels. Overall, on $A^{\text{tst}}_{*\text{u}}$, SLR-based methods outperform SLIM and LogR on average. This indicates the capability of SLR-based methods in recommending safe drugs when the testing prescriptions do not have an ADR labels.

TABLE VI: Performance Comparison on $A^{\text{tst}}_{*\text{u}}$

mdl	best $\overline{\text{rec}_\text{P}}$			best $\overline{\text{rec}_\text{N}}$			best HM_rec		
	$\overline{\text{rec}_\text{P}}$	$\overline{\text{rec}_\text{N}}$	HM_rec	$\overline{\text{rec}_\text{P}}$	$\overline{\text{rec}_\text{N}}$	HM_rec	$\overline{\text{rec}_\text{P}}$	$\overline{\text{rec}_\text{N}}$	HM_rec
Rand	0.0031	0.0026	0.0026	0.0031	0.0026	0.0026	0.0031	0.0026	0.0026
LogR	0.1470	0.2504	0.1852	0.1491	0.2468	0.1858	0.1491	0.2468	0.1858
SLIM	0.2512	0.3053	0.2754	0.2512	0.3053	0.2754	0.2512	0.3053	0.2754
S+LR	**0.2745**	0.2824	0.2782	0.2714	0.2689	0.2700	0.2714	0.2689	0.2700
SLR-s	0.2562	0.2251	0.2395	0.2471	0.3275	0.2817	0.2471	0.3275	0.2817
SLR-sl	0.2685	0.3263	0.2933	0.2615	0.3736	0.3071	0.2658	0.3661	0.3075
SLR-si	0.2566	0.3073	0.2795	0.2461	0.3527	0.2898	0.2476	0.3532	0.2909
SLR-sli	0.2700	0.3447	0.3025	0.2499	**0.4025**	0.3082	0.2541	0.4005	**0.3107**

The column "mdl" corresponds to models. The best overall performance is **bold**.

C. Recommendation Method Comparison

Table VII presents the contribution from various recommendation components of SLR-sli (i.e., the SLIM component, the LogR component and the indication component as in Figure 1). In Table VII, the average number of true positive prescriptions that are recommended by SLR-sli for A^{tst}_{*+}, A^{tst}_{*-} and $A^{\text{tst}}_{*\text{u}}$ in terms of best HM_rec is 503.0, 249.0, and 901.8, respectively, and the corresponding average number of true negative prescriptions are 1204.0, 1234.2, and 829.8, respectively. For A^{tst}_{*+} and A^{tst}_{*-}, the recommended true positives are fewer than the true negatives. This may be due to the fact that, as in Table I, the knowledge pool A_{FAERS} has more negative prescriptions, which leads to a higher probability of a recommended prescription being true negative in A_{FAERS}. However, for $A^{\text{tst}}_{*\text{u}}$, there are more recommended true positives than true negatives. The $A^{\text{tst}}_{*\text{u}}$ testing prescriptions have higher average order than the A^{tst}_{*+} and A^{tst}_{*-} testing prescriptions as in Table III. Such higher order leads to higher risks of those $A^{\text{tst}}_{*\text{u}}$ testing prescriptions on their own being true positives (as in Table I, true positives tend to have higher order than true negatives). The fact that SLR-sli recommends more true positives than true negatives for $A^{\text{tst}}_{*\text{u}}$ demonstrates its strong capability of identifying to-avoid drugs particularly when the existing prescriptions already have high risks of ADRs.

TABLE VII: Recommendation Statistics from SLR-sli

subset	Metric	#TP				#TN			
		total	SLIM	LogR	Ind	total	SLIM	LogR	Ind
A^{tst}_{*+}	$\overline{\text{rec}_\text{P}}$	503.0	503.0	0.0	0.0	1204.0	1126.8	0.0	77.2
	$\overline{\text{rec}_\text{N}}$	493.4	488.4	3.4	1.6	1218.0	1138.2	0.0	79.8
	HM_rec	503.0	503.0	0.0	0.0	1204.0	1126.8	0.0	77.2
A^{tst}_{*-}	$\overline{\text{rec}_\text{P}}$	249.0	248.6	0.4	0.0	1234.2	1179.4	0.0	54.8
	$\overline{\text{rec}_\text{N}}$	249.0	248.6	0.4	0.0	1234.2	1179.4	0.0	54.8
	HM_rec	249.0	248.6	0.4	0.0	1234.2	1179.4	0.0	54.8
$A^{\text{tst}}_{*\text{u}}$	$\overline{\text{rec}_\text{P}}$	958.0	828.8	88.0	41.2	714.4	180.6	0.0	533.8
	$\overline{\text{rec}_\text{N}}$	886.8	674.8	128.2	83.8	833.8	93.2	478.0	262.6
	HM_rec	901.8	757.2	95.2	49.4	829.8	145.6	452.6	231.6

Column "Metric" presents to the evaluation metric. Column "#TP"/"#TN" presents TP/TN prescriptions. Column "total" presents the average total number of TP/TN prescriptions. Column "SLIM", "LogR" and "Ind" present the average TP/TN prescriptions from SLIM, LogR and drug indication component, respectively.

For A^{tst}_{*+} and A^{tst}_{*-}, almost all the true positive and true negative recommendations are from the SLIM component of SLR-sli. However, for $A^{\text{tst}}_{*\text{u}}$, the recommendations are contributed from all of the three components (e.g., 757.2, 95.2, and 49.4 true positives from SLIM, LogR and indication component, respectively). This indicates that for high-order prescriptions, to-avoid/safe drug recommendation may require the collective consideration of various factors including the

drug co-prescription patterns and drug indications, etc. This also demonstrates the capability of SLR-sli in leveraging the recommendation lists strategically from various recommendation components in order to generate reliable to-avoid/safe drug recommendations.

D. ADR-Relevant Drug Identification

We analyze how SLR-sli can identify ADR-relevant single drugs using $\mathbf{x}$ in the LogR component. The $\mathbf{x}$ values represent how strong the association between the corresponding drug and the ADR label is. The larger positive/negative value of x_i indicates higher possibility for drug d_i in inducing/not inducing ADR. The top-5 largest positive and negative weights in the $\mathbf{x}$ from SLR-sli and the corresponding drugs are presented in Table VIII. Note the largest positive/negative weights correspond to drugs that are most predictive of ADR/non-ADR induction, and the SLR-sli corresponds to the best HM_{rec} performance on $A^{\text{tst}}_{*\text{u}}$ as in Table VII, where LogR contributes most. In Table VIII, it turns out only two $\mathbf{x}$ values are positive, corresponding to drug atorvastatin and simvastatin. These two drugs have been reported in SIDER as to induce myopathy, and 28.72% of the ADR inducing prescriptions have at least one of these two drugs. This indicates that the LogR component in SLR-sli is able to identify single drugs that are most ADR relavant. Meanwhile, the top-5 largest negative weights correspond to warfarin, ibuprofen, metformin, quetiapine, and valproic acid. In A_*, the frequency of these 5 drugs in A^+ is 119, 110, 176, 46, and 54, respectively, and much higher than the average frequency 14.05 of all the drugs; the frequency of these 5 drugs in A^- is 124, 86, 148, 102, and 95, respectively, and also much higher than the average frequency 7.32 of all the drugs. This indicates that the LogR component in SLR-sli is also able to identify single drugs that are most ADR safe.

TABLE VIII: Most Indicating Drugs from SLR-sli

Top-5		Bottom-5	
x	d	x	d
7.944	**atorvastatin**	-10.783	warfarin
6.430	**simvastatin**	-9.492	ibuprofen
-	-	-3.592	metformin
-	-	-1.761	quetiapine
-	-	-0.673	valproic acid

"x" is the value of the entry corresponding to the drug in $\mathbf{x}$; "d" corresponds to the drug; and "Top-5" and "Bottom-5" correspond to the top-5 largest and smallest values in $\mathbf{x}$, respectively. Drugs that are reported in SIDER to induce myopathy on their own are **bold**.

E. Co-Prescription Patterns

We analyze the co-prescription patterns using W^+ and $W^-$from SLR-sli of the best HM_{rec} performance. W^+/W^- captures the co-prescription patterns among drugs that often appear in ADR inducing/no-ADR inducing prescriptions, respectively. The values in W^+/W^- can be considered as measurement of such patterns: higher/lower values correspond to stronger/weaker co-prescription patterns. The top-10 largest values in W^+ and W^- from the optimal SLR-sli model and the corresponding co-prescribed drug pairs on A^{tst}_* are presented in Table IX.

In Table IX, the learned W^+ has a density 10.89%, and its top-5 largest values correspond to the following drug pairs: (ethinyl estradiol, etonogestrel), (hydrochlorothiazide, triamterene), (sulfamethoxazole, trimethoprim), (acetaminophen, hydrocodone), and (fluticasone propionate, salmeterol). These drug pairs are frequently prescribed together in A^+. The number of prescriptions in A^+ that contain these drug pairs is 38, 32, 31, 133, and 44, respectively ("freq" in Table IX), compared to the average number of A^+ prescirptions of all possible drug pairs, which is 2.48. This demonstrates that SLR-sli is able to capture the co-prescription patterns among perscriptions. In addition, many of the co-prescribed drugs share similar medical purposes. For example, ethinyl estradiol and etonogestrel are often used for birth control, and acetaminophen and hydrocodone are often prescribed together to relieve moderate to severe pains and treat fever.

In Table IX, the learned W^- has a density 18.20%, and its top-5 largest values correspond to the following drug pairs: (ethinyl estradiol, norelgestromin), (emtricitabine, tenofovir), (acetaminophen, hydrocodone), (fluticasone propionate, salmeterol), and (lopinavir, ritonavir). These drug pairs are also frequently co-prescribed in A^-. The number of A^- prescriptions that have these drugs pairs is 23, 35, 36, 35 and 36, respectively, compared to 2.07, the average number of A^- prescriptions that have all possible drug pairs. These drugs pairs also share similar medical purposes. For example, emtricitabine and tenofovir, and lopinavir and ritonavir are commonly used in HIV cocktail therapy, and fluticasone propionate and salmeterol are prescribed together for preventing difficulty in breathing.

It is also noted that in Table IX, the learned W^+ and W^- values are not necessarily in order with the drug pair frequencies. This indicates that the SLR-sli is able to discover patterns related to ADRs that are not simply represented by prescription frequencies. In addition, the two lists of top-10 drug pairs corresponding to W^+ and W^- values have 3 common pairs (e.g., (acetaminophen, hydrocodone), (fluticasone propionate, salmeterol) and (emtricitabine, tenofovir)) and 7 different pairs, respectively. This may illustrate the common practices in both ADR-inducing and non-ADR-inducing drug perscribing, but meanwhile the different distributions of ADR-inducing and non-ADR inducting prescriptions. Co-prescription patterns corresponding to the best HM_{rec} on A^{tst}_{*+}, A^{tst}_{*-} and $A^{\text{tst}}_{*\text{u}}$ are presented in Table S6, S7, and S8, respectively, in the supplementary materials.

F. Case Study

Table S9 in the supplementary materials presents some examples of testing prescriptions and their recommended to-avoid drugs from SLR-sli such that the corresponding new prescriptions (i.e., testing prescriptions and recommended drugs together) are ADR-inducing. When the testing prescriptions already induce ADRs (i.e., in A^{tst}_{*+}; the first row block in Table S9), SLR-sli is able to identify additional single drugs such that when the recommended drug is taken together with the testing prescription, the OR of the new prescription increases over that of the testing prescription. For example, the

TABLE IX: Co-Prescription Patterns from SLR-sli on A_*^{tst}

W^+				W^-			
w	d_1	d_2	freq	w	d_1	d_2	freq
0.7021	ethinyl estradiol	etonogestrel	38	0.6275	ethinyl estradiol	norelgestromin	23
0.5903	hydrochlorothiazide	triamterene	32	0.5932	**emtricitabine**	**tenofovir**	35
0.5874	sulfamethoxazole	trimethoprim	31	0.5916	acetaminophen	hydrocodone	36
0.5719	acetaminophen	hydrocodone	133	0.5523	fluticasone propionate	salmeterol	35
0.5302	fluticasone propionate	salmeterol	44	0.5189	lopinavir	**ritonavir**	36
0.5034	amoxicillin	clavulanate	22	0.4668	budesonide	formoterol	17
0.4967	carbidopa	levodopa	17	0.4496	fluorouracil	leucovorin	18
0.4595	acetaminophen	codeine	27	0.4459	conjugated estrogens	medroxyprogesterone	17
0.4492	**emtricitabine**	**tenofovir**	22	0.4104	buprenorphine	naloxone	9
0.4268	pamidronate	zoledronate	42	0.3837	amphetamine	dextroamphetamine	8

In this table, "w" is the value of the entry corresponding to the drug pairs in W^+/W^-; "d_1" and "d_2" are the two drugs in the drug pair, and "freq" is the frequency of the corresponding drug pairs in training data A^+ and A^-. Drugs that are reported in SIDER to induce myopathy on their own are **bold**.

testing prescription (fusidic acid, simvastatin) in A_{*+}^{tst} is ADR inducing with OR =39.809 (p-value=1.281×10^{-14}). SLR-sli recommends ramipril such that (fusidic acid, simvastatin, ramipril) also induces ADR with an increased OR =159.2182 (p-value=1.8109×10^{-9}). Such examples also indicate that the increased OR is due to the additional recommended drugs, and the potential directional interactions between the recommended drugs and the testing prescriptions. This conclusion may hint on the research on directional drug-drug interactions [15].

When the testing prescriptions do not induce ADRs (i.e., in A_{*-}^{tst}; the second row block in Table S9), SLR-sli is able to recommend to-avoid drugs that will introduce ADRs to the corresponding new prescriptions. In particular, there are examples in which the recommended to-avoid drugs themselves are not ADR-inducing. For example, the testing prescription (dicyclomine, gabapentin, lansoprazole, lorazepam, pamidronate, zoledronate, zolpidem) is ADR negative and the recommended to-avoid drug "oxycodone" is itself ADR negative as well. However, the corresponding new prescription is ADR inducing (OR =34.116, p-value=7.007×10^{-4}). Another example is for the testing prescription (acetylsalicylic acid, amlodipine, atorvastatin, isosorbide mononitrate, lisinopril, olmesartan, omeprazole, quinine). Although atorvastatin is ADR-positive, this prescription its own is ADR-negative. However, SLR-sli recommends another ADR-negative drug "atenolol" that results in an ADR-positive prescription (acetylsalicylic acid, amlodipine, atorvastatin, isosorbide mononitrate, lisinopril, olmesartan, omeprazole, quinine, atenolol) (OR =13.646, p-value=3.566×10^{-3}). This demonstrates that the recommendations from SLR-sli are not trivial and do capture the interaction patterns among multiple drugs in a prescription. The same conclusion can also be derived from the examples for testing prescriptions in A_{*u}^{tst} (i.e., the third row block in Table S9). For example, for the testing prescription (acetylsalicylic acid, amlodipine, esomeprazole, hydrochlorothiazide, potassium chloride, sorafenib, triamterene) in A_{*u}^{tst}, SLR-sli recommends "furosemide" which, taken together with the testing prescription, will induce ADR, although furosemide itself is ADR-negative. More examples on safe drug recommendation are available in Table S10 in the supplementary materials.

XI. Discussions and Conclusions

In this paper, we formally formulated the to-avoid drug recommendation problem and the safe drug recommendation problem. We developed a joint sparse linear recommendation and logistic regression model (SLR) to tackle the recommendation problems, and developed real datasets and new evaluation protocols to evaluate the model. To the best of our knowledge, this is the first time such problems are formally formulated and corresponding computational solutions are proposed. Our experiments demonstrate strong performance of SLR compared to other methods. In particular, SLR is able to 1). recommend safe drugs that taken together with the ADR-inducing prescriptions would lead to no/less ADR effects; 2). to recommend to-avoid drugs that taken together with the non ADR-inducing prescriptions would not induce ADRs; 3). to recommend safe drugs when the testing prescriptions do not have an ADR labels; and 4). to leverage the recommendation lists strategically from various recommendation components in order to generate reliable to-avoid/safe drug recommendations.

In SLR, we use simplistic linear models. Non-linear models (for both recommendation and label prediction) together may be able to better discover co-prescription patterns and predict labels jointly. We will investigate non-linear models in the future work. Also it is common that DDIs induce multiple ADRs simultaneously. This problem is substantially more difficult but deserve future investigation. In addition, SLR is learned on a population level without personalization to each individual patient, because the ADR labels are generated over a population. With patient data available, we may extend the methods to personalized drug recommendation.

Acknowledgment

This material is based upon work supported by the National Science Foundation under Grant Number IIS-1855501 and IIS-1827472.

References

[1] Division of Health and Nutrition Examination Surveys, "National health and nutrition examination survey," 2013, http://www.cdc.gov/nchs/nhanes.htm.

[2] S. V. Iyer, R. Harpaz, P. LePendu, A. Bauer-Mehren, and N. H. Shah, "Mining clinical text for signals of adverse drug-drug interactions," *Journal of the American Medical Informatics Association*, pp. 353–362, 2014.

[3] F. Cheng and Z. 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*, vol. 21, no. e2, pp. e278–e286, 2014.

[4] N. P. Tatonetti, J. Denny, S. Murphy, G. Fernald, G. Krishnan, V. Castro, P. Yue, P. Tsau, I. Kohane, D. Roden *et al.*, "Detecting drug interactions from adverse-event reports: interaction between paroxetine and pravastatin increases blood glucose levels," *Clinical Pharmacology & Therapeutics*, vol. 90, no. 1, pp. 133–142, 2011.

[5] H. Yang and C. C. Yang, "Harnessing social media for drug-drug interactions detection," in *Healthcare Informatics (ICHI), 2013 IEEE International Conference on*. IEEE, 2013, pp. 22–29.

[6] C.-W. Chiang, P. Zhang, X. Wang, L. Wang, S. Zhang, X. Ning, L. Shen, S. K. Quinney, and L. Li, "Translational high-dimensional drug interaction discovery and validation using health record databases and pharmacokinetics models," *Clinical Pharmacology & Therapeutics*, vol. 103, no. 2, pp. 287–295, 2018.

[7] X. Ning, T. Schleyer, L. Shen, and L. Li, "Pattern discovery from directional high-order drug-drug interaction relations," in *Healthcare Informatics (ICHI), 2017 IEEE International Conference on*. IEEE, 2017, pp. 154–162.

[8] F. Ricci, L. Rokach, and B. Shapira, "Introduction to recommender systems handbook," in *Recommender systems handbook*. Springer, 2011, pp. 1–35.

[9] M. Wiesner and D. Pfeifer, "Health recommender systems: Concepts, requirements, technical basics and challenges," *International Journal of Environmental Research and Public Health*, vol. 11, no. 3, pp. 2580–2607, 2014. [Online]. Available: http://www.mdpi.com/1660-4601/11/3/2580

[10] L. Guo, B. Jin, C. Yao, H. Yang, D. Huang, and F. Wang, "Which doctor to trust: A recommender system for identifying the right doctors," *Journal of medical Internet research*, vol. 18, no. 7, 2016.

[11] Q. Zhang, G. Zhang, J. Lu, and D. Wu, "A framework of hybrid recommender system for personalized clinical prescription," in *2015 10th International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*, Nov 2015, pp. 189–195.

[12] X. Ning and G. Karypis, "Slim: Sparse linear methods for top-n recommender systems," in *Data Mining (ICDM), 2011 IEEE 11th International Conference on*. IEEE, 2011, pp. 497–506.

[13] M. Kuhn, I. Letunic, L. J. Jensen, and P. Bork, "The sider database of drugs and side effects," *Nucleic acids research*, vol. 44, no. D1, pp. D1075–D1079, 2015. [Online]. Available: http://sideeffects.embl.de/

[14] W.-H. Chiang, L. Shen, L. Li, and X. Ning, "Drug-drug interaction prediction based on co-medication patterns and graph matching," *International Journal of Computational Biology and Drug Design*, 2018, in press.

[15] W.-H. Chiang, T. Schleyer, L. Shen, L. Li, and X. Ning, "Pattern discovery from high-order drug-drug interaction relations," *Journal of Healthcare Informatics Research*, vol. 2, no. 3, pp. 272–304, Sep 2018. [Online]. Available: https://doi.org/10.1007/s41666-018-0020-2