



# Propensity score-adjusted three-component mixture model for drug-drug interaction data mining in FDA Adverse Event Reporting System

Xueying Wang<sup>1</sup>  | Lang Li<sup>2</sup> | Lei Wang<sup>1</sup> | Weixing Feng<sup>1</sup> | Pengyue Zhang<sup>2</sup> 

<sup>1</sup>Institute of Intelligent System and Bioinformatics, College of Automation, Harbin Engineering University, Harbin, China

<sup>2</sup>Department of Biomedical Informatics, College of Medicine, The Ohio State University, Columbus, Ohio

## Correspondence

Weixing Feng, Institute of Intelligent System and Bioinformatics, College of Automation, Harbin Engineering University, No. 145-1, Nantong Street, Nangang, Harbin 150001, Heilongjiang, China.

Email: fengweixing@hrbeu.edu.cn  
Pengyue Zhang, Department of Biomedical Informatics, College of Medicine, The Ohio State University, 250 Lincoln Tower, 1800 Cannon Drive, Columbus, OH 43210.  
Email: pengyue.zhang@osumc.edu

## Funding information

National Institutes of Health, DK102694, GM10448301-A1, R01GM117206, R01LM011945; National Natural Science Foundation of China, 61471139; Natural Science Foundation of Heilongjiang Province, F20160062; HEU Fundamental Research Funds for the Central University, 3072019CFG0401, HEUCFP201722; National Science Foundation, NSF1622526

With increasing trend of polypharmacy, drug-drug interaction (DDI)-induced adverse drug events (ADEs) are considered as a major challenge for clinical practice. As premarketing clinical trials usually have stringent inclusion/exclusion criteria, limited comedication data capture and often times small sample size have limited values in study DDIs. On the other hand, ADE reports collected by spontaneous reporting system (SRS) become an important source for DDI studies. There are two major challenges in detecting DDI signals from SRS: confounding bias and false positive rate. In this article, we propose a novel approach, propensity score-adjusted three-component mixture model (PS-3CMM). This model can simultaneously adjust for confounding bias and estimate false discovery rate for all drug-drug-ADE combinations in FDA Adverse Event Reporting System (FAERS), which is a preeminent SRS database. In simulation studies, PS-3CMM performs better in detecting true DDIs comparing to the existing approach. It is more sensitive in selecting the DDI signals that have nonpositive individual drug relative ADE risk (NPIRR). The application of PS-3CMM is illustrated in analyzing the FAERS database. Compared to the existing approaches, PS-3CMM prioritizes DDI signals differently. PS-3CMM gives high priorities to DDI signals that have NPIRR. Both simulation studies and FAERS data analysis conclude that our new PS-3CMM is a new method that is complement to the existing DDI signal detection methods.

## KEYWORDS

adverse drug event, drug-drug interaction, false discovery rate, FDA adverse event reporting system, propensity score

## 1 | INTRODUCTION

Drug-drug interaction (DDI) can change either the pharmacokinetics or pharmacodynamics (PD) drug response when two or more drugs are coadministered together.<sup>1</sup> Clinically, adverse drug events (ADEs) are the primary safety concern

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. *Statistics in Medicine* published by John Wiley & Sons, Ltd.

of DDIs. The incidence of DDI-induced ADEs is expected to increase due to the increasing trend of polypharmacy. Currently in US, 21.8% of US population receives three or more prescription drugs simultaneously, and 10.7% of the persons even have five or more.<sup>2</sup> Comparatively, these percentages were only 11.0% and 3.6% before 2000.<sup>2</sup> DDI-induced ADEs have become a major challenge to the clinical care setting, as DDIs are suspected to be related with up to 59.1% of ADEs, and they became a common cause for drugs withdraw from the market.<sup>3,4</sup> However, only limited DDIs have been detected in premarketing clinical trials; most of these trials focus on single-drug effect, and comedICATIONS due to the comorbidities are usually limited by stringent inclusion/exclusion criteria. Thus, most DDIs are discovered through analyzing the postmarketing pharmacovigilance databases.<sup>5</sup>

The spontaneous reporting system (SRS) is an important type of pharmacovigilance database for the postmarketing ADE detection.<sup>5</sup> SRS collects ADE reports from a variety of sources including patients, pharmaceutical companies, and healthcare professionals. As the SRS collects only “cases”, ADE incidence rates are usually inflated. Despite the potential bias, it is commonly acceptable to use the SRS and test whether a drug or a drug combination has an increased ADE relative to the other patients who did not have this drug or drug combination. This analysis is also called disproportionality analyzes (DPAs) for the drug-induced ADE signal detection. DPAs are originally developed to detect single-drug-ADE signals by comparing the observed report frequency for a drug-ADE combination (ie, observed frequency) to the baseline frequency under the no association assumption (ie, expected frequency).<sup>6</sup> Well-known DPAs include the Information Component (IC),<sup>7</sup> IC with Dirichlet prior,<sup>8</sup> empirical Bayes geometric mean (EBGM) also known as the multigamma Poisson shrinkage (MGPS),<sup>9</sup> and the three-component mixture model (3CMM).<sup>10</sup> On the other hand, the likelihood ratio test (LRT) and its extension zero-inflated Poisson model-based LRT (ZIP-LRT) are able to test the associations between an ADE and many drugs.<sup>11,12</sup>

The DPA method has also been used for investigating the DDI-induced ADEs by comparing the observed report frequency of a drug-drug-ADE combination to its baseline frequency assuming no interactions between two drugs. By considering the drug-drug pair as a “new drug”, the observed report frequency can be obtained from the SRS database. Therefore, the single-drug-ADE signals detection methods can be extended to identify the DDI signals by designing a new baseline frequency. For example, by extending the baseline frequency from a two-item set to the three-item set based on a so-called concept of “market basket problem”, the EBGM approach was extended to identify the suspected DDIs.<sup>13,14</sup> The IC with Dirichlet prior approach was extended for detecting DDI signals based on a proposed three-order IC function.<sup>8</sup> The three-order IC is defined as the deviation between the IC under the condition of the third event and the original IC, and the baseline frequency account for both the main effects and the two-drug interactions. Despite these extensions of DPA methods, Thakrar et al<sup>15</sup> proposed multiplicative and additive models to detect DDI signals in the FDA Adverse Event Reporting System (FAERS) database. Under the assumption of no interaction, the multiplicative model assumes that the risk associated with one drug multiplies to the background risk, and the relative risk (RR) associated with this drug is not influenced by the other drug; the additive model assumes that the risk associated with one drug is additive to the background risk, and this drug's risk is not influenced by the other drug. Norén et al<sup>16</sup> also proposed a well-known  $\Omega$  shrinkage method for the suspected DDIs detection in individual case safety reports (ICSRs) data. It determines the baseline model with an additive model and calculates a shrinkage observed-to-expected ratio of the ADE risk for the drug-drug pair. The shrinkage penalizes the signals of rare drug-drug-ADE combinations with a prespecified parameter.

Detecting DDI signals from SRS has several major challenges. SRS analysis subjects to confounding bias from thousands of comedICATIONS. In single-drug-ADE association analysis, Dumouchel et al<sup>9</sup> proposed a stratified expectation by using the Mantel-Haenszel formula.<sup>17</sup> This approach calculates the expected frequency by using the expected frequency within each strata. The stratified approach can be generalized for LRT, IC, and their aforementioned extensions as well. However, stratification is limited by the number of suspected confounders. As the increased number of strata yields reduced sample size within each strata, the power to detect DDI signals is reduced.<sup>18</sup> Besides stratification, confounding bias can be addressed by multiple regression models in which confounding variables (eg, demographic variables and comedICATIONS) are treated as covariates. For example, van Puijenbroek et al<sup>19</sup> proposed a multiple logistic regression model for detecting DDI signals. Though multiple regression approach is not necessarily limited by the sample size challenge, techniques for properly selecting confounding variables are still underdeveloped.<sup>20</sup> False positive control is another significant challenge in detecting DDI signals from hundreds of thousands drug pairs.<sup>10</sup> False discovery rate (FDR) is more practical than traditional hypothesis testing techniques (eg, *P*-value) for large-scale DDI signals detection. For instance, Bayesian FDR and local FDR were proposed to control false positive rate for detecting single-drug signals.<sup>21,22</sup> While these approaches have not been developed for DDI signals detection.

In this article, our aim is to detect suspected DDI signals with a low FDR from the FAERS, which is a prominent SRS database. Specifically, we propose a novel approach to compute the expected report frequency, in which confounding bias is adjusted by using the comedication information. Then, we model the observed and expected frequencies for the

drug-drug-ADE combinations under an empirical Bayes model frame work. By defining the null risk (eg,  $RR = 1$ ), the proposed approach estimated local FDR for each signal. The rest of the article is organized as follow. In Section 2, we review the  $\Omega$  shrinkage method and describe the proposed method. In Section 3, we conduct simulation studies to evaluate and compare the performance of the proposed approach and the  $\Omega$  shrinkage method. In Section 4, we illustrate the utilization of the proposed approach by analyzing the FAERS. Last, Section 5 concludes and discusses the proposed approach.

## 2 | MATERIALS AND METHODS

### 2.1 | The FAERS database and data processing

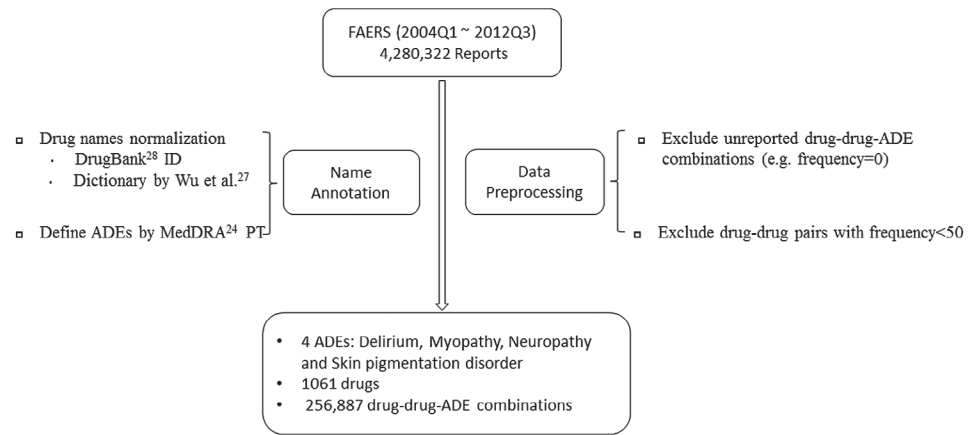
The FAERS is a prominent SRS maintained by the U.S. Food and Drug Administration. The FAERS database collects adverse event reports directly from manufacturers, pharmacists, physicians, nurses, and consumers.<sup>23</sup> Each report consists of information including: (1) patients demographics; (2) drugs information which classified as “primary suspect”, “secondary suspect”, “concomitant”, and “interacting”; (3) ADEs coded by using the Medical Dictionary for Regulatory Activities' (MedDRA) Preferred Terms (PTs)<sup>24</sup>; (4) patient outcomes; (5) report sources; and (6) indications coded by using MedDRA PT code for the reported drugs. For FAERS analysis, the ADEs can be defined by using MedDRA PTs. However, the drug names may involve abbreviations, synonyms, brand names, and spelling mistakes. Thus, normalization of the drug names is necessary for FAERS analysis. The FAERS database has two major advantages: (1) massive sample size (>18 million reports)<sup>25</sup> and (2) the implementation of MedDRA PT allows the ADEs to be directly available. Despite the advantages, the FAERS database also has some limitations. First, the true ADE incidence rates cannot be obtained, as the FAERS database only contains ADE reports. Second, there is a great uncertainty between the drugs and the ADEs on a report, as the report collects all possible drugs and ADEs for a patient.<sup>9</sup> Third, the quality of the FAERS database may be compromised by inaccurate reports and missing data (eg, age and gender). For instance, over 30% of the FAERS reports involve missing data, as reporting of demographics is voluntary.<sup>26</sup> Although the FAERS database has the aforementioned limitations, it is a valuable source for DDI study.

In this article, we selected FAERS reports from 2004Q1 to 2012Q3 as our data source. To avoid the spurious associations caused by the duplicated reports, we only left the reports with the latest primary ID and removed other reports with the same primary IDs. As there was a great uncertainty of the drugs' roles (eg, primary suspect), we used all the drugs on each report including “primary suspect,” “secondary suspect”, “concomitant”, and “interacting”. Furthermore, the drug names were normalized into generic names by using the dictionary proposed by DrugBank and Wu et al<sup>27,28</sup> Frequent drug names (report frequency > 999) that failed to be normalized were manually checked and corrected. After data processing, our FAERS dataset contained 4 280 322 reports with 1180 generic drug names and 15 445 MedDRA PT ADE names. Of these 4 280 322 reports used for data analysis, about 38% of the reports involved with missing age information and nearly 10% of the reports involved with missing gender information.

To detect potential DDI signals, we extracted the information for drug-drug combinations and ADEs from our processed FAERS dataset. The total number of drugs in our FAERS dataset was 1180. In order to limit the search space of the drug-drug combinations and maintain sufficient candidate drug-drug combinations, we set criteria to filter the drug-drug combinations by their frequencies. Specifically, all drug-drug combinations appeared in our FAERS dataset with the frequency bigger than 50 were used in our illustrative FAERS analysis. After filtering the report frequencies of drug-drug combinations (eg, >50), we identified 81 312 drug-drug combinations generated by 1061 drugs. Furthermore, we selected four primary ADEs for analysis including delirium, myopathy, neuropathy, and skin pigmentation disorder.<sup>10</sup> Our final dataset contained 256 887 drug-drug-ADE combinations which were formed from four ADEs and 1061 drugs (Figure 1).

### 2.2 | Notations and definitions

For a drug-drug-ADE combination (eg, Drug1 and Drug2), the FAERS reports can be summarized into eight frequencies (Table 1) In Table 1,  $a$ ,  $b$ ,  $c$ ,  $d$ ,  $e$ ,  $f$ ,  $g$ , and  $h$  denote the eight report frequencies for a drug-drug-ADE combination in the FAERS, according to the status of Drug1, Drug2, and the ADE. For instance,  $g$  is the number of reports that contain Drug1, Drug2, and the ADE; and  $h$  is the number of reports that contain Drug1 and Drug2 but do not contain the ADE. Furthermore, we define the observed relative reporting rate for the ADE under the four different drug exposure status as

**FIGURE 1** Overview of FAERS data processing**TABLE 1** Report frequencies and rates for a drug pair and ADE in FAERS database

Exposure status		ADE status		Relative reporting rate
Drug1	Drug2	Yes	No	
No	No	$a$	$b$	$r_{00} = a/(a+b)$
Yes	No	$c$	$d$	$r_{10} = c/(c+d)$
No	Yes	$e$	$f$	$r_{01} = e/(e+f)$
Yes	Yes	$g$	$h$	$r_{11} = g/(g+h)$

Abbreviation: ADE, adverse drug event.

$r_{00} = \frac{a}{a+b}$ ,  $r_{10} = \frac{c}{c+d}$ ,  $r_{01} = \frac{e}{e+f}$ , and  $r_{11} = \frac{g}{g+h}$  (Table 1). We define  $N = g$  as the observed report frequency,  $E$  as the expected report frequency,  $\lambda$  as the RR, and  $\hat{\lambda} = \frac{N}{E}$  as the observed RR for a drug-drug-ADE combination. For different DPAs,  $E$  are calculated differently.

### 2.3 | A review of the $\Omega$ shrinkage method

In this section, we review the well-known DPA to detect DDI signals from SRS by Norén et al<sup>16</sup> This method, also known as the  $\Omega$  shrinkage method, penalizes the signals of less frequent drug-drug-ADE combinations. Under this approach, based on the additive risk model:

$$r_{11} + r_{00} = r_{10} + r_{01}$$

The expected frequency  $E$  is calculated as:

$$E = \left[ 1 - \frac{1}{\max\left(\frac{r_{10}}{1-r_{10}}, \frac{r_{00}}{1-r_{00}}\right) + \max\left(\frac{r_{01}}{1-r_{01}}, \frac{r_{00}}{1-r_{00}}\right) - \frac{r_{00}}{1-r_{00}} + 1} \right] \times (g+h). \quad (1)$$

In Equation (1),  $(g+h)$  represents the total number of reports that contain both Drug1 and Drug2 (Table 1). The maximum functions on denominator are equivalent as  $r_{10} = \max(r_{10}, r_{00})$  and  $r_{01} = \max(r_{01}, r_{00})$ . These restrictions guarantee the ADE risk caused by single drug is higher than its background risk. And the  $\Omega$  shrinkage measure is calculated as

$$\Omega = \log_2 \frac{N + \alpha}{E + \alpha}. \quad (2)$$

In Equation (2),  $\alpha$  is prespecified (eg,  $\alpha = 0.5$ ). The  $\Omega$  shrinkage method can also be derived under the Bayesian framework by assuming  $N \sim Pois(\lambda E)$  and  $\lambda \sim \Gamma(\alpha, \alpha)$ . The posterior distribution of  $\lambda$  is also a gamma distribution:  $\lambda | N,$

$E \sim \Gamma(N + \alpha, E + \alpha)$ . To detect DDIs, the signal generation can be based on the logarithm of the lower limits of 95% credibility interval for  $\lambda$  denoted as  $\Omega_{0.25}$ . According to Noren et al<sup>16</sup>, positive  $\Omega_{0.25}$  value indicates DDI signals.

## 2.4 | Propensity score-adjusted three-component mixture model

Our proposed statistics for DDI signal detection is named as PS-adjusted three-component mixture model (PS-3CMM). We will first introduce the calculation for the expected frequency  $E$  and subsequently the empirical Bayes mixture model for adjusted-FDR estimation.

### 2.4.1 | Logistic regression-based expected frequency for drug-drug-ADE combinations

If there are no confounders, logistic regression model (3) can be used to estimate the individual drug and DDI effect.

$$\text{Logit}[P(\text{ADE} = 1)] = \beta_0 + \beta_1 D1 + \beta_2 D2 + \beta_3 D1D2, \quad (3)$$

where in Equation (3),  $D1$  and  $D2$  are the binary drug exposure status,  $P(\text{ADE} = 1)$  is the risk of ADE, and  $\beta_0, \beta_1, \beta_2, \beta_3$  are the coefficients.

The estimation for the background odds of the ADE in the absence of both  $D1$  and  $D2$  is:

$$\text{Odds}_{00} = \frac{r_{00}}{1 - r_{00}} = \exp(\hat{\beta}_0). \quad (4)$$

Similarly, the estimations for the odds of the ADE in the absence of one of  $D1$  and  $D2$  are

$$\text{Odds}_{10} = \frac{r_{10}}{1 - r_{10}} = \exp(\hat{\beta}_0 + \hat{\beta}_1), \quad \text{and} \quad (5)$$

$$\text{Odds}_{01} = \frac{r_{01}}{1 - r_{01}} = \exp(\hat{\beta}_0 + \hat{\beta}_2). \quad (6)$$

The estimation for the odds of the ADE in present of both  $D1$  and  $D2$  under the condition of no interaction is

$$\text{Odds}_{11} = \frac{r_{11}}{1 - r_{11}} = \exp(\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2). \quad (7)$$

From Equations (4) to (7), we get the relation for the reporting rate (8)

$$\frac{r_{11}}{1 - r_{11}} \times \frac{r_{00}}{1 - r_{00}} = \frac{r_{10}}{1 - r_{10}} \times \frac{r_{01}}{1 - r_{01}}. \quad (8)$$

Therefore, the baseline model can be seen as a multiplicative odds model.

From Equations (7) and (8), the expected report frequency of taking both drugs under a non-DDI assumption can be calculated as:

$$E = \left[ 1 - \frac{1}{\frac{\frac{r_{10}}{1-r_{10}} \times \frac{r_{01}}{1-r_{01}}}{\frac{r_{00}}{1-r_{00}}} + 1} \right] \times (g + h) = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2)}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2)} \times (g + h). \quad (9)$$

### 2.4.2 | PS-adjusted logistic regression-based expected frequency for drug-drug-ADE combinations

Because the large number of comedication confounders cannot be straightforwardly controlled in the SRS dataset, we propose a PS-adjusted expectation to address confounding bias. Assuming that the unobserved confounders can be

characterized by the comedications in our dataset, the logistic regression (10) and the principal components (PCs) derived from comedications are used to compute the probability (ie, propensity score, PS) of taking the drug.

$$\text{Logit}[P(\text{Drug} = 1)] = \alpha_0 + \sum_{i=1}^K \alpha_i \times PC_i. \quad (10)$$

In Equation (10),  $K$  denote the number of PCs that we used to calculate the PS. PCs are derived from the comedications and  $K$  are selected by a prespecified threshold for the fraction of the total explained variance. Then, for each drug-drug-ADE combination, we used the logistic regression model (11) to estimate the individual drug and DDI effect with adjustment of two PSs.

$$\text{Logit}[P(\text{ADE} = 1)] = \beta_0 + \beta_1 D1 + \beta_2 D2 + \beta_3 D1D2 + \beta_4 PS1 + \beta_5 PS2. \quad (11)$$

In Equation (11),  $D1$  and  $D2$  are the binary indicators of two drugs and  $PS1$  and  $PS2$  are their PSs, respectively. The expected report frequency is calculated as:

$$E = \sum_{i=1}^{g+h} \frac{\text{Exp}(\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_4 PS1_i + \hat{\beta}_5 PS2_i)}{1 + \text{Exp}(\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_4 PS1_i + \hat{\beta}_5 PS2_i)}. \quad (12)$$

In Equation (12), for report  $i$ ,  $\frac{\text{Exp}(\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_4 PS1_i + \hat{\beta}_5 PS2_i)}{1 + \text{Exp}(\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_4 PS1_i + \hat{\beta}_5 PS2_i)}$  is the estimation of the PS-adjusted probability for the ADE occurrence with only the individual drug effects. In another word,  $E$  is the expected report frequency of taking both drugs under a non-DDI assumption while adjusted for potential confounders.

### 2.4.3 | FDR estimation for drug-drug-ADE combinations

To model the observed and expected report frequencies (ie,  $Ns$  and  $Es$ ), we adopt the empirical Bayes mixture model framework. As nearly 83% of report frequencies for drug-drug-ADE combinations in our dataset is equal to 0, we choose the three-component empirical Bayes mixture model proposed by Zhang et al<sup>10</sup> to model the RR for the drug-drug-ADE combinations. Moreover, the 3CMM framework also provides a much needed false positive control for the signals. We assume the drug-drug-ADE combinations belong to three groups. They are (i) zero DDI risk: drug-drug-ADE combinations have 0 frequency (eg, neither the drug combination nor the individual drugs causes the ADE); (ii) background noise: the report frequencies for drug-drug-ADE combinations are close to their expected frequencies due to the additive effects from two drugs; and (iii) DDI signals: drug-drug-ADE combinations have much higher RRs than the background additive effect.

We assume the unobserved RR ( $\lambda$ ) follows the 3CMM (13).

$$\lambda \sim P_1 \{I(\lambda = 0)\} + \sum_{l=2}^3 [P_l \Gamma(\lambda; \alpha_l, \beta_l)];$$

$$\alpha_2 = \beta_2, \alpha_3 > \beta_3, \text{ and } P_1 + P_2 + P_3 = 1. \quad (13)$$

In Equation (13),  $I(\lambda = 0)$  is an identity distribution that characterizes the drug-drug-ADE combinations with  $RR = 0$  (eg, no ADE risk).  $\Gamma(\lambda; \alpha_l, \beta_l)$  is the gamma distribution such that  $\Gamma(\lambda; \alpha_l, \beta_l) = \frac{\beta_l^{\alpha_l}}{\Gamma(\alpha_l)} \lambda^{\alpha_l-1} \exp(-\beta_l \lambda)$ . There are two gamma distributions with different restrictions in the 3CMM. One gamma distribution ( $l = 2$ ) restricts  $\alpha_2 = \beta_2$ , and has its mean equal to 1. Therefore, this gamma distribution characterizes the drug-drug-ADE combinations to have a background ADE risks (ie, mean  $RR = 1$ ). The other gamma distribution ( $l = 3$ ) restricts  $\alpha_3 > \beta_3$  and has its mean greater than 1. Therefore, this gamma distribution characterizes drug-drug-ADE combinations with increased ADE risks (ie, mean  $RR > 1$ ), which represents signals of DDI. Furthermore, from Equation (13), we assume  $N \sim \text{Pois}(\lambda \times E)$ . Thus, the observed distribution of  $N$  is



$$P(N) = P_1 I(N = 0) + \sum_{l=2}^3 [P_l F(N; \alpha_l, \beta_l, E)], \quad (14)$$

$$\alpha_2 = \beta_2, \alpha_3 > \beta_3, \text{ and } P_1 + P_2 + P_3 = 1.$$

In Equation (14),  $F(N; \alpha_l, \beta_l, E) = \frac{\Gamma(N+\alpha_l) \times E^N \times \beta_l^{\alpha_l}}{\Gamma(\alpha_l) \times N! \times [E+\beta_l]^{N+\alpha_l}}$  is the negative binomial distribution function.

For the adverse DDIs detection, our concern is drug-drug-ADE combinations with positive report frequencies ( $N > 0$ ). Therefore, to reduce the excessive amount of drug-drug-ADE combination with  $N = 0$ , the conditional distribution (15) can be used for inference, which will not affect local FDR estimation accord to Zhang et al<sup>10</sup>

$$P(N = k | N > 0) = \frac{F(N = k; \alpha_2, \beta_2, E) + \frac{P_3}{P_2} \times F(N = k; \alpha_3, \beta_3, E)}{P(N > 0; \alpha_2, \beta_2, N) + \frac{P_3}{P_2} \times P(N > 0; \alpha_3, \beta_3, E)}. \quad (15)$$

In Equation (15),  $P(N > 0; \alpha_2, \beta_2, N)$  is the probability for  $N > 0$ ,  $F(N = k; \alpha_2, \beta_2, E)$  is the distribution function for  $P(N > 0; \alpha_2, \beta_2, N)$ . The conditional log-likelihood function for Equation (15) is:

$$ll \left( N; E, \left\{ \alpha_2 = \beta_2, \alpha_3, \beta_3, \frac{P_3}{P_2} \right\} \right) = \sum \log[P(N = k | N > 0)]. \quad (16)$$

By considering  $\frac{P_3}{P_2}$  as one parameter, the four parameters  $\left\{ \alpha_2 = \beta_2, \alpha_3, \beta_3, \frac{P_3}{P_2} \right\}$  in Equations (16) can be obtained by maximizing the observed likelihood (eg, maximum likelihood estimate). In this study, the MLEs are obtained by using the R function `nlminb`, which is able to conduct both unconstrained and box-constrained optimization by using Newton-like methods.<sup>29</sup>

As the mean of the first gamma component in Equation (16) equals 1, it will be served as the null distribution to compute the local FDR accord to Efron et al<sup>22</sup> The local FDR for a drug-drug-ADE combination is defined as:

$$\text{Local FDR}(N = k | k > 0) = \frac{F(N = k; \alpha_2, \beta_2, E)}{F(N = k; \alpha_2, \beta_2, E) + \frac{P_3}{P_2} \times F(N = k; \alpha_3, \beta_3, E)}. \quad (17)$$

In another word, local FDR represents the posterior probability that a drug-drug-ADE combination has the background RR (null). As we use PSs to adjust confounding bias, we name the statistic in Equation (17) as adjusted-FDR for screening adverse DDIs.

## 2.5 | Proportional reporting ratio for DDI

Proportional reporting ratio (PRR) is a well-known method for pharmacovigilance proposed by Evan et al.<sup>30</sup> PRR is developed for detecting signals of single-drug-induced ADE from a SRS database. Specifically, PRR calculates the ratio of ADE reporting rates with and without a drug. Based on the report frequencies of Drug1 and Drug2 (Table 1),  $PRR_{D1}$  and  $PRR_{D2}$  for Drug1 and Drug2 are defined as:

$$\begin{aligned} PRR_{D1} &= \frac{(c + g)/(c + d + g + h)}{(a + e)/(a + b + e + f)}, \\ PRR_{D2} &= \frac{(e + g)/(e + f + g + h)}{(a + c)/(a + b + c + d)}. \end{aligned} \quad (18)$$

For the Drug1-Drug2-ADE combination, we define its PRR by considering the Drug1-Drug2 pair as a “new drug”, based on the drug pair’s ADE frequency (Table 1), the  $PRR_{D1D2}$  for the Drug1-Drug2-ADE combination is:

$$PRR_{D1D2} = \frac{g/(g + h)}{(a + c + e)/(a + b + c + d + e + f)}. \quad (19)$$

The lower bound of the 95% confidence interval (CI) for the PRR is defined as:

$$PRR_{0.25} = e^{\ln PRR - 1.96SD}, \quad (20)$$

where SD denotes the SD, for Drug1,  $SD_{D1} = \sqrt{\frac{1}{c+g} - \frac{1}{c+d+g+h} + \frac{1}{a+e} - \frac{1}{a+b+e+f}}$ ; similarly, for Drug2  $SD_{D2} = \sqrt{\frac{1}{e+g} - \frac{1}{e+f+g+h} + \frac{1}{a+c} - \frac{1}{a+b+c+d}}$ , for Drug1-Drug2 combination  $SD_{D1D2} = \sqrt{\frac{1}{g} - \frac{1}{g+h} + \frac{1}{a+c+e} - \frac{1}{a+b+c+d+e+f}}$ .<sup>31</sup> Let  $PRR_{0.25 D1}$ ,  $PRR_{0.25 D2}$ , and  $PRR_{0.25 D1D2}$  to be the values of  $PRR_{0.25}$  for Drug1, Drug2, and Drug1-Drug2 combination, respectively.

Similar to the approach proposed by Almenoff et al<sup>32</sup>, a drug pair with  $PRR_{0.25 D1D2} > \max(PRR_{0.25 D1}, PRR_{0.25 D2})$  is considered as a signal of DDI.

Comparing to adjusted-FDR and  $\Omega_{0.25}$ , PRR and  $PRR_{0.25}$  are measurements without shrinkage adjustment. Therefore, in this study, they will be served as a reference for the  $\Omega$  shrinkage and our proposed approach.

### 3 | SIMULATION STUDY

We conducted extensive simulation studies to evaluate the performance of our proposed method and compare it with the  $\Omega$  shrinkage method. In each method, the area under the receiver-operating characteristic (ROC) curve (AUC) was calculated by using the simulated true positive DDIs and negative controls. Let PS be the PS,  $D$  be the binary drug exposure status, and  $Y$  be the binary ADE status. Specifically, the data were simulated as:

1. Simulate drug exposure status by  $D \sim \text{Bin}(1, \text{PS})$ , where  $\text{PS} \sim \text{Beta}(\alpha, \beta)$ .
2. Simulate ADE status for a given drug pair by:

$$\text{Logit}[P(Y = 1)] = \beta_0 + \beta_1 D1 + \beta_2 D2 + \beta_3 D1 * D2 + \beta_4 \text{PS1} + \beta_5 \text{PS2}. \quad (21)$$

In order to evaluate and compare the performance of our proposed method to the  $\Omega$  shrinkage method, we first define two types of DDIs. One type is DDI with nonpositive individual drug relative ADE risk (DDI-NPIRR) (eg,  $r_{01} - r_{00} \leq 0$  or  $r_{10} - r_{00} \leq 0$ ) and the other is DDI with positive individual drug relative ADE risk (DDI-PIRR) (eg, both  $r_{01} - r_{00} > 0$  and  $r_{10} - r_{00} > 0$ ).

We simulated data of the following three situations:

- (a) Data contain both DDI-NPIRR ( $\beta_1 > 0$ ,  $\beta_2 < 0$ , and  $\beta_3 > 0$ ) and DDI-PIRR ( $\beta_1 > 0$ ,  $\beta_2 > 0$ , and  $\beta_3 > 0$ ), and without confounding bias ( $\beta_4 = \beta_5 = 0$ ).
- (b) Data only contain DDI-NPIRR, and without confounding bias.
- (c) Data contains both DDI-NPIRR and DDI-PIRR, and with confounding bias ( $\beta_4 \neq 0$  and  $\beta_5 \neq 0$ ).

For all the three situations, the total number of reports was chosen to be 100 000, and PS was generated from beta distribution such that  $\text{PS} \sim \text{Beta}(1, 6)$ . In each simulation, 10 000 positive DDI signals and 10 000 negative controls were simulated from the uniformly distributed parameters shown in Table 2.

The ROC curves and AUC values are shown in Figure 2. In the first simulation, Figure 2A, we display the AUCs in the absence of confounding bias [ $\beta_4 = \beta_5 = 0$ ] in Equation (21) for both DDI-NPIRR and DDI-PIRR. Although both proposed PS-3CMM and the  $\Omega$  shrinkage methods have good performance, PS-3CMM performs better than  $\Omega$  shrinkage,  $\text{AUC} = 0.94$  vs  $\text{AUC} = 0.88$ . In the second simulation with only DDI-NPIRR, PS-3CMM has even better DDI detection performance than the  $\Omega$  shrinkage method,  $\text{AUC} 0.93$  vs  $0.78$ , respectively. In the third simulation, Figure 2C, when confounding bias is present, [ $\beta_4 \neq 0$  and  $\beta_5 \neq 0$ ] in Equation (21), the PS-3CMM's DDI detection performance did not reduce ( $\text{AUC} = 0.94$ ) much from the simulation where there was no confounding. On the other hand, the  $\Omega$  shrinkage method has obvious decreased performance from 0.88 to 0.84 when confounding bias was presented.

### 4 | FAERS ANALYSIS

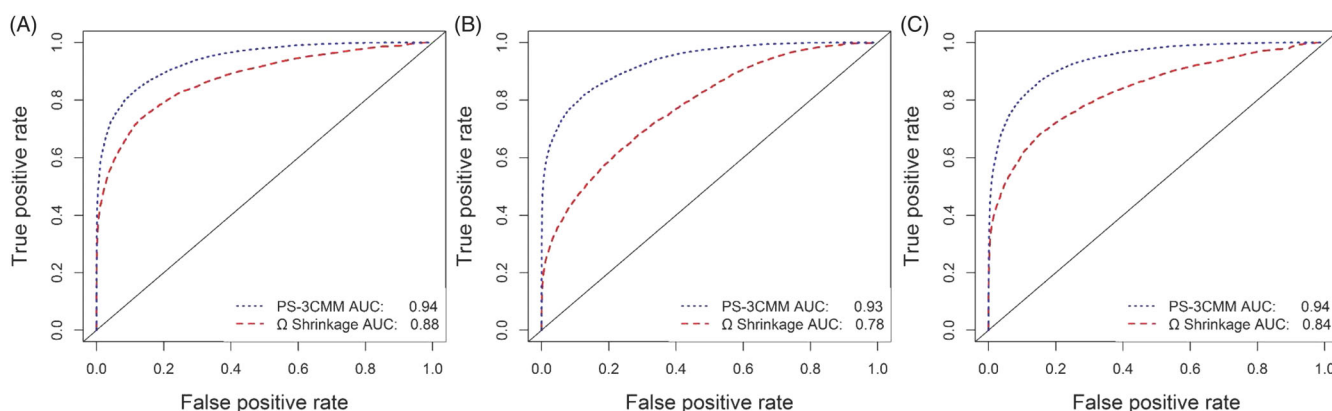
We demonstrate the utilization of the proposed approach in analyzing the FAERS database including 256 887 drug-drug-ADE combinations. In our analysis, 160 PCs, [ $K = 160$ ] in Equation (10), were used to calculate the PS and



**TABLE 2** Parameters setting for the simulation study

	Situation (a)		Situation (b)		Situation (c)	
	DDI signal	Negative control	DDI signal	Negative control	DDI signal	Negative control
$\beta_0$	[-6,-5]	[-6,-5]	[-6,-5]	[-6,-5]	[-6,-5]	[-6,-5]
$\beta_1$	[-0.8,0.8]	[-0.8,0.8]	[0.1,2]	[-0.8,0.8]	[-0.8,0.8]	[-0.8,0.8]
$\beta_2$	[-0.8,0.8]	[-0.8,0.8]	[-2,0.1]	[-0.8,0.8]	[-0.8,0.8]	[-0.8,0.8]
$\beta_3$	[0.1,1]	[-1,-0.1]	[0.1,1]	[-1,-0.1]	[0.1,1]	[-1,-0.1]
$\beta_4$	0	0	0	0	[-2,1]	[-1,1]
$\beta_5$	0	0	0	0	[-2,1]	[-1,1]

Abbreviation: DDI, drug-drug interaction.



**FIGURE 2** A, Receiver-operating characteristic (ROC) curves of propensity score-adjusted three-component mixture model (PS-3CMM) and  $\Omega$  shrinkage without confounding bias for all drug-drug interactions (DDIs); B, ROC curves of PS-3CMM and  $\Omega$  shrinkage without confounding bias for DDI-nonpositive individual drug relative ADE risk; and C, ROC curves of PS-3CMM and  $\Omega$  shrinkage with confounding bias for all DDIs. [Color figure can be viewed at wileyonlinelibrary.com]

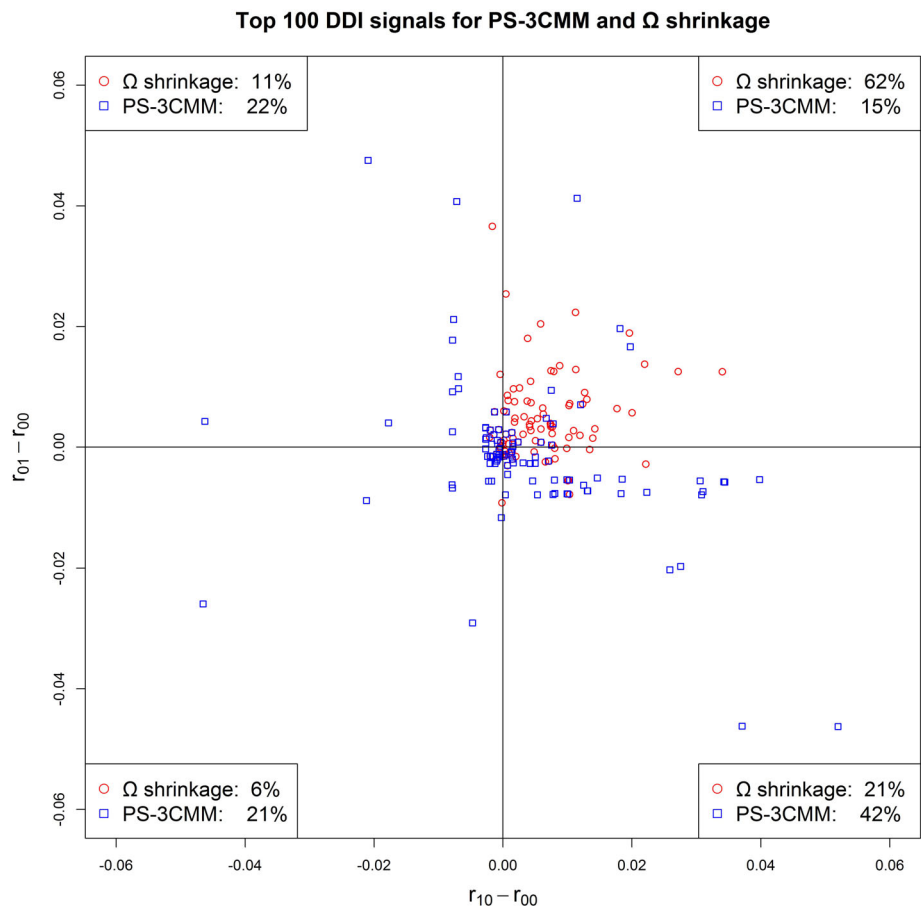
expected report frequency, in which the 160 PCs explained 70% of total variation of all comedICATIONS. The estimated parameters for the conditional likelihood in Equation (16) are  $\hat{\alpha}_2 = \hat{\beta}_2 = 2.19$ ,  $\hat{\alpha}_3 = 0.22$ ,  $\hat{\beta}_3 = 0.04$ , and  $\hat{P}_3 = 0.08$ . These estimates mean that the component 2 (background noise group) had mean RR = 1 with SD = 0.68 and the component 3 (increased risk group) had mean RR = 5.33 with SD = 11.33. For each drug-drug-ADE combination, the adjusted-FDR and  $\Omega_{0.25}$  were calculated for signal detection and comparison. We also calculate the value of PRR<sub>0.25</sub> for each drug-drug-ADE combination as a reference for the DDI signals detected by PS-3CMM and  $\Omega$  shrinkage methods.

### 4.1 | Performance comparisons between PS-3CMM and $\Omega$ shrinkage

If both drugs cause an ADE, we define it as the DDI-PIRR, which yields additive or synergistic risk for the ADE. In terms of reporting rate, DDI-PIRR has the pattern  $r_{11} > r_{10}, r_{01} > r_{00}$  (Table 1). Alternatively, if only one drug has ADE risk and the other one does not, we define it as DDI-NPIRR, which increases the ADE risk when two drugs are coprescribed. In terms of reporting rate, DDI-NPIRR has the pattern  $r_{11} > r_{10} > r_{00} \geq r_{01}$  or  $r_{11} > r_{01} > r_{00} \geq r_{10}$ .

We compared the top-100 DDI signals between two methods in Figure 3: PS-3CMM and  $\Omega$  shrinkage. The x-axis is the difference of the marginal reporting rate of drug 1 and the reporting rate of no drug (eg,  $r_{10} - r_{00}$ ); and the y-axis is the difference of the marginal reporting rates of drug 2 and the reporting rate of no drug (eg,  $r_{01} - r_{00}$ ). Clearly, top-100 DDI signals of PS-3CMM and  $\Omega$  shrinkage have different patterns. Two lists only have 25 overlapped DDIs (Table 5). Among the  $\Omega$  shrinkage's top-100 signals, 62% of them have  $r_{10}, r_{01} > r_{00}$ . On the other hand, only 15% of the PS-3CMM's top-100 signals are DDI-PIRR. For DDI-NPIRR ( $r_{00} \geq r_{10}$  or  $r_{00} \geq r_{01}$ ), the percentages are 32% for  $\Omega$  shrinkage and 64% for PS-3CMM.

**FIGURE 3** Top-100 drug-drug interaction signals for propensity score-adjusted three-component mixture model and  $\Omega$  shrinkage methods. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



## 4.2 | Evaluation of top-100 DDI signals detected by adjusted-FDR and $\Omega_{0.25}$

The drug side-effect database (SIDER) and DrugBank database were used to evaluate the top-100 adjusted-FDR DDI signals (Table S1) and the top-100  $\Omega_{0.25}$  DDI signals (Table S2). SIDER collects drug-ADE association data from drug labels.<sup>33</sup> DrugBank documented DDIs from a variety of sources.<sup>28</sup> Our evaluation was conducted manually through searching the drug-drug-ADE combinations in SIDER and DrugBank. The results for the adjusted-FDR DDI signals and the  $\Omega_{0.25}$  DDI signals are summarized in Tables 3 and 4, respectively. As Table 3 shows, 49 out of the top-100 DDI signals have supporting evidence from DrugBank. Among these 49 DDI signals, seven pairs have both drugs labeled with the ADE; 17 pairs have only one drug labeled with the ADE; and 25 combinations have neither drugs labeled with the ADE. For the remaining 51 DDI signals that are not documented in DrugBank, four have both drugs labeled with the ADE; 20 have only one drug labeled with the ADE; and 27 have neither drug labeled with the ADE. In Table 4, 56 out of the top-100 DDI signals have supporting evidence from DrugBank. Among these 56 DDI signals, eight pairs have both drugs labeled

**TABLE 3** Evaluation of the top-100 drug-drug-ADE combinations ranked by adjusted-FDR

Number of drugs have ADE on its label <sup>a</sup>	DDI evidence in DrugBank <sup>b</sup>		Total
	Yes	No	
2	7	4	11
1	17	20	37
0	25	27	52
<b>Total</b>	<b>49</b>	<b>51</b>	<b>100</b>

Abbreviations: ADE, adverse drug event; DDI, drug-drug interaction; FDR, false discovery rate.

<sup>a</sup>Drug label information was collected from SIDER.

<sup>b</sup>DDI evidence was collected from DrugBank.

Number of drugs have ADE on its label <sup>a</sup>	DDI evidence in DrugBank <sup>b</sup>		
	Yes	No	Total
2	8	1	9
1	25	18	43
0	23	25	48
Total	56	44	100

Abbreviations: ADE, adverse drug event; DDI, drug-drug interaction.

<sup>a</sup>Drug label information was collected from SIDER.

<sup>b</sup>DDI evidence was collected from DrugBank.

**TABLE 4** Evaluation of the top-100 drug-drug-ADE combinations ranked by  $\Omega_{0.25}$

**TABLE 5** Overlap top-100 drug-drug-ADE combinations

Drug1	Drug2	ADE	$-\log_{10}$ (adjusted-FDR) [Rank]	$\Omega_{0.25}$ [Rank]	PRR <sub>0.25</sub> D1	PRR <sub>0.25</sub> D2	PRR <sub>0.25</sub> D1D2
<u>BUPROPION</u> <sup>a</sup>	<u>METHAMPHETAMINE</u>	Neuropathy	35.0380 [9]	3.2731 [84]	2.2115	3.9370	36.1433
BEVACIZUMAB	VALPROIC ACID <sup>a</sup>	Delirium	23.1426 [18]	3.8676 [9]	0.6918	3.2050	66.0302
PEGFILGRASTIM	PHENYTOIN <sup>a</sup>	Delirium	22.7221 [19]	4.9747 [1]	1.9341	2.0910	162.7558
BEVACIZUMAB	PHENYTOIN <sup>a</sup>	Delirium	22.3023 [22]	4.1906 [4]	0.6918	2.0910	49.2345
PEGFILGRASTIM	VALPROIC ACID <sup>a</sup>	Delirium	18.6823 [26]	4.2639 [3]	1.9341	3.2050	106.5403
<u>CARBOPLATIN</u> <sup>a</sup>	<u>VARENICLINE</u>	Neuropathy	18.4633 [27]	3.6936 [17]	2.2121	0.3180	52.1493
<u>NALOXONE</u>	<u>BUDESONIDE</u>	Neuropathy	17.4483 [28]	3.8090 [13]	0.8758	0.7835	32.8251
PEGFILGRASTIM	BEVACIZUMAB	Delirium	16.6822 [31]	3.4410 [51]	1.9341	0.6918	14.1947
VALPROIC ACID <sup>a</sup>	CARBOPLATIN	Delirium	15.4286 [34]	4.0513 [6]	3.2050	1.5340	89.2514
<u>WARFARIN</u>	<u>TETRACYCLINE</u>	Delirium	14.7344 [38]	3.2537 [93]	1.4521	2.3461	34.1982
METAXALONE	TERIPARATIDE	Delirium	14.0579 [40]	3.2475 [98]	1.2557	0.3933	24.9570
<u>PHENYTOIN</u>	<u>CARBOPLATIN</u>	Delirium	13.3545 [46]	4.3066 [2]	2.0910	1.5340	76.0028
<u>CARVEDILOL</u>	<u>NALOXONE</u>	Neuropathy	13.0354 [49]	3.2530 [94]	1.8283	0.8758	28.1420
METAXALONE	ROSUVASTATIN	Delirium	12.0456 [55]	3.3886 [60]	1.2557	0.5564	24.7431
<u>ROFECOXIB</u> <sup>a</sup>	<u>LAMOTRIGINE</u> <sup>a</sup>	Neuropathy	11.8226 [59]	3.5168 [35]	1.9221	1.0212	31.8124
<u>BUPIVACAINE</u>	<u>PREDNISON</u> <sup>a</sup>	Delirium	11.2454 [67]	3.4850 [42]	2.0987	0.9646	54.0550
<u>TETRACYCLINE</u>	<u>ESCITALOPRAM</u> <sup>a</sup>	Delirium	10.9939 [68]	3.3108 [73]	2.3461	1.9282	50.7768
<u>DULOXETINE</u>	<u>ROFECOXIB</u> <sup>a</sup>	Neuropathy	10.9655 [69]	3.5546 [30]	2.6122	1.9221	45.9861
ROSIGLITAZONE	BIMATOPROST	Neuropathy	10.8976 [71]	3.4377 [52]	1.0349	0.7282	25.5408
<u>GEMCITABINE</u>	<u>DULOXETINE</u>	Delirium	10.6277 [74]	3.3538 [65]	0.8139	1.9728	45.1643
<u>VALDECOXIB</u> <sup>a</sup>	<u>TELMISARTAN</u> <sup>a</sup>	Neuropathy	10.6144 [76]	3.4979 [38]	1.1240	0.8844	30.3932
<u>GEMCITABINE</u>	<u>ROFECOXIB</u>	Delirium	10.4108 [79]	3.4159 [56]	0.8139	0.8227	29.1456
BUPIVACAINE	ZOLEDRONATE	Delirium	10.3610 [80]	3.4567 [45]	2.0987	1.1792	58.8019
FILGRASTIM	PHENYTOIN <sup>a</sup>	Delirium	10.2910 [81]	3.8094 [12]	1.9407	2.0910	50.6368
<u>ROFECOXIB</u>	<u>GRANISETRON</u>	Delirium	10.2715 [82]	3.5553 [29]	0.8227	1.4602	42.8081

Note: The underlined drug names represent DDI evidence in DrugBank database.

Abbreviations: ADE, adverse drug event; DDI, drug-drug interaction; FDR, false discovery rate.

<sup>a</sup>Drugs labeled the ADE in SIDER database.

with the ADE; 25 pairs have only one drug labeled with the ADE; and 23 combinations have neither drugs labeled with the ADE. For the remaining 44 DDI signals that are not documented in DrugBank, one have both drugs labeled with the ADE; 18 have only one drug labeled with the ADE; and 25 have neither drug labeled with the ADE. By using DrugBank as a gold standard to evaluate the top-100 signals, the proposed method has a positive-predicted values (PPVs) of 49% and the  $\Omega$  shrinkage method has a PPV of 56%.

Using  $PRR_{0.25}$  as a reference, for the top-100 adjusted-FDR DDI signals (Table S1), 91 DDI signals have increased  $PRR_{0.25}$  (ie,  $PRR_{0.25 D1D2} > \max(PRR_{0.25 D1}, PRR_{0.25 D2})$ ), nine DDI signals have decrease  $PRR_{0.25}$  (ie,  $PRR_{0.25 D1D2} < \max(PRR_{0.25 D1}, PRR_{0.25 D2})$ ). Among these nine DDI signals, four have supporting evidence from DrugBank and two have both drugs labeled with the ADE. In the top-100  $\Omega_{0.25}$  DDI signals (Table S2), all of the 100 DDI signals have increased  $PRR_{0.25}$ . Among the top 100 ranking DDI signals detected by adjusted-FDR and  $\Omega_{0.25}$ , there are 25 common DDI signals (Table 5). Among these 25 DDI signals, 14 drug pairs have supporting evidence from DrugBank; only two of the drug pairs have both drugs labeled with ADE and 11 of drug pairs have only one drug labeled with ADE. Moreover, all these 25 common DDI signals have increased  $PRR_{0.25}$ .

Except the overlap 25 DDI signals, both our proposed method and the  $\Omega$  shrinkage method can detect some DDI signals that cannot be detected by each other, therefore, the PS-3CMM is an important complementary to the  $\Omega$  shrinkage method.

## 5 | CONCLUSION AND DISCUSSION

In this article, we propose a new approach, PS-3CMM to detect signals of DDI from the FAERS database. This method controls false positive rate and adjusts for confounding bias due to comedICATIONS. Specifically, this approach first uses the comedICATIONS to derive the PSs. These PSs are then utilized to calculate adjusted expected frequencies for drug-drug-ADE combinations under the no DDI assumption. The empirical Bayes mixture model is used to characterize the RR between the DDI-induced ADE frequency and the expected ADE frequency of drug-drug combination under no DDI effect assumption. This 3CMM estimates adjusted-FDR for each drug-drug-ADE combination.

The PSs are derived from PC analyses of comedication variables. The number of PCs can be determined by the percentage of total variation that explained by the selected PCs. In our illustrative FAERS analysis, the selected PCs capture 70% of variations of comedICATIONS. Therefore, we believe these comedication PCs contain sufficient variations among the comedICATIONS, and they shall capture a great deal of comorbidity variation among the patients. Our model also includes PS as continuous variable. It avoids the potential sparsity caused by the stratification of the data used in the EBGM and IC.<sup>18</sup> For instance, in our FAERS database the drug pairs' frequency has a median of 180. For  $P$  binary variables (eg, age, gender, and comedICATIONS), the number of strata is  $2^P$ . The sparsity appears even for moderate  $P$  (eg, stratify by four age groups and gender yield  $P = 8$ ). Unlike the  $\Omega$  shrinkage method in which the penalize parameter is prespecified, the proposed approach is data-driven in which the parameters in the prior distribution are estimated from data under the empirical Bayes framework. Furthermore, the adjusted-FDR provides much needed false positive control for high throughput DDI signals detection. Our simulation studies show the PS-3CMM has better performance comparing to the  $\Omega$  shrinkage method especially in detecting DDI-NPIRR. In addition, the simulation studies also demonstrate that our PS-3CMM has the advantage of adjusting for confounding bias, which is a common problem in SRS.

The application of PS-3CMM is demonstrated through the FAERS database. In FAERS analysis, we notice that the PS-3CMM and the  $\Omega$  shrinkage prioritized signals differently. PS-3CMM is more prone to identify DDI-NPIRR signals. DDI changes the expected pharmacologic or clinical response for an individual drug when two drugs are coadministered.<sup>1</sup> The incidence of ADE can be significantly increased by: (1) pharmacokinetic (PK) DDI, in which a drug's plasma concentration is increased by coprescribed drug and exceeded the maximum tolerated dose<sup>34</sup> and (2) PD DDI in which additive or synergistic effects on ADE risk occurs. For DDI-NPIRR signals, a drug's ADE risk is significantly increased by another drug which has either similar or lower ADE risk compared with background. In another word, DDI-NPIRR signals are more likely to have a mechanism similar to PK DDI. For instance, the acetaminophen-methamphetamine-neuropathy combination and acetaminophen-neostigmine-neuropathy combination are highly prioritized by PS-3CMM but not by the  $\Omega$  shrinkage method. They are ranked in 11th and 17th by PS-3CMM, and 915th and 8126th by the  $\Omega$  shrinkage method (Table S1). In both combinations, the neuropathy risks of methamphetamine and neostigmine are lower than background risks, which imply both of these DDI signals are DDI-NPIRR signals. Moreover, both drugs are known to have PK drug interaction with acetaminophen according to DrugBank.<sup>35</sup> For instance, the description of drug interactions for

acetaminophen-methamphetamine combination is “Acetaminophen may decrease the excretion rate of Metamphetamine which could result in a higher serum level”. Thus, these DDI-NPIRR signals are likely to have a mechanism of PK DDI. Alternatively, DDI-PIRR signals are more likely to have a mechanism similar to PD DDI. For DDI-PIRR signals, both individual drugs have higher than background risks, and the drug combinations have much increased risks. For instance, the cetirizine-levodopa-neuropathy combination and the ibuprofen-tinzaparin-neuropathy combination are highly prioritized by  $\Omega$  shrinkage but not by the PS-3CMM. They are ranked in 32th and 34th by the  $\Omega$  shrinkage method and 5716th and 19226th by PS-3CMM (Table S2). Both of these two DDI signals are DDI-PIRR signals. These two-drug combinations are known to have increased ADE risk according to DrugBank with no clear evidence of PK drug interaction.<sup>35</sup> In conclusion, DDI-NPIRR signals are more likely to have a mechanism similar to PK DDI and DDI-PIRR signals are more likely to have a mechanism of PD DDI. The difference of the DDI signals pattern is mainly caused by the different baseline risk assumptions, and the  $\Omega$  shrinkage method also restricts the ADE risk caused by single drug should be higher than its background risk, but the PS-3CMM do not has such restriction. Among the top-100 signals detected by PS-3CMM, 49 drug pairs have documented DDI evidences from DrugBank and 24 drug pairs have at least one drug with the ADE in its drug label. At the meantime, 75% of top 100 PS-3CMM DDI signals are not ranked top 100 by the  $\Omega$  shrinkage method. These results suggest that PS-3CMM is a different and complementary approach to the  $\Omega$  shrinkage approach, as well as the PRR approach.

As we described earlier in Section 2.4.1, instead of using PS-adjusted expected DDI-ADE frequency under the no DDI effect assumption, unadjusted expected DDI-ADE frequency can be computed in Equation (9). Although PS-adjusted analysis is clearly supreme to the unadjusted analysis in FAERS data analysis, there are other situations that unadjusted analysis is probably better. For example, if the pharmaco-epidemiological study is designed from a longitudinal cohort, such as the electronic health record (EHR) dataset, the confounding variables, such as demographics, comedications, or comorbidity can be matched and adjusted through nested case-control study design.<sup>36,37</sup> In this case, in analyzing DDI-ADE associations, the unadjusted analysis is a proper choice. Because it does not need to adjust for confounding variables, it shall have better power comparing to the adjusted analysis. Additionally, in real application, either the full model (Equation 13) or the conditional model (Equation 15) can be used for signal detection. However, the conditional model is more efficient when the primary interest is to detect drug-drug-ADE combinations with increased risk.

In this article, we illustrate the utilization of our proposed method by using the FAERS database which is a preeminent SRSs. Here, we would like to point that the proposed method can be utilized to other SRS databases. Nowadays, the importance of pharmacovigilance through SRS is internationally recognized. For instance, all European Union countries were obliged to establish patient/consumer reporting within their SRSs since 2012.<sup>38</sup> China and Japan have well-established SRS as well.<sup>39,40</sup> Additionally, the WHO's SRS (VigiBase) contains over 20 million reports which are coded by MedDRA and WHO Drug Dictionaries.<sup>41</sup> A detailed review of SRS database can be found in Huang et al<sup>42</sup> These SRS databases share similar informatics structure (eg, MedDRA), which facilitates the utilization of the proposed method. For instance, the  $\Omega$  Shrinkage method has been applied to both VigiBase and the China's SRS database,<sup>16,43-45</sup> in which the frequencies in Table 1 are able to be obtained. Thus, PS-3CMM can be straightforwardly implemented to the other SRS databases, which share similar informatics structure with FAERS. As for illustration, we only choose four ADEs in our FAERS analysis. However, the framework of PS-3CMM has no limitation with respect to the number of ADEs. We would also like to point out that the study can be initiated by either ADE(s) or drug combination(s). The current study is initiated by ADEs, as our motivation is to detect novel drug interaction signals. Alternatively, a study may be initiated with drug combinations with evidence of DDI. Such a study is more powerful to detect ADE signals and it can be used as a validation for pharmacology study.

One limitation of the proposed approach is to handle highly correlated drugs. It is a major challenge for all of the current DDI detection research. These highly correlated drugs (eg, a drug always coadministered with another) are due to increasing practice of polypharmacy. Another limitation is the proposed approach does not incorporate the drugs' duration and dosage information which are important factors for DDI, as those informations are not well captured by most SRS databases. Alternatively, EHRs databases contain detailed temporal information with respect to patients' pharmacy prescriptions, as well as the dosages. The associations between drugs and ADEs in EHR can be investigated under elegant epidemiology designs. For instance, nested case-control design with a 30-day drug exposure can be used to investigate acute ADEs, as well as the case-crossover design.<sup>46</sup> Hennessy et al provides detailed guidance for investigating DDI from EHR databases.<sup>47</sup> We would like to point out that our proposed method has the potential to analyze EHR data with proper epidemiology designs. For instance, instantaneous exposure before the ADE (ie, duration information) can be captured by the drug exposure window, and the analysis can be stratified or adjusted by dosage information. We believe the detection



of DDI signals from EHR databases can be an important direction to further expand our model in order to characterize drugs' duration and dosage information.

## ACKNOWLEDGMENTS

This work has been supported by several NIH grants, DK102694, GM10448301-A1, R01GM117206, and R01LM011945; and NSF grant, NSF1622526. It has also been supported by National Natural Science Foundation of China (61471139), Natural Science Foundation of Heilongjiang Province (F2016006), HEU Fundamental Research Funds for the Central University (3072019CFG0401, HEUCFP201722).

## DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

## ORCID

Xueying Wang  <https://orcid.org/0000-0002-6350-7998>

Pengyue Zhang  <https://orcid.org/0000-0002-3592-9042>

## REFERENCES

1. Crowther NR, Holbrook AM, Kenwright R, Kenwright M. Drug interactions among commonly used medications. Chart simplifies data from critical literature review. *Can Fam Physician*. 1997;43:1972-1976, 1979-1981.
2. Percha B, Altman RB. Informatics confronts drug-drug interactions. *Trends Pharmacol Sci*. 2013;34(3):178-184.
3. Davies EC, Green CF, Taylor S, Williamson PR, Mottram DR, Pirmohamed M. Adverse drug reactions in hospital in-patients: a prospective analysis of 3695 patient-episodes. *Plos One*. 2009;4(2):e4439.
4. Huang S-M, Strong JM, Zhang L, et al. New era in drug interaction evaluation: US Food and Drug Administration update on CYP enzymes, transporters, and the guidance process. *J Clin Pharmacol*. 2008;48(6):662-670.
5. Ventola CL. Big data and pharmacovigilance: data mining for adverse drug events and interactions. *P T*. 2018;43(6):340-351.
6. Chiang CW, Zhang P, Wang X, et al. Translational high-dimensional drug interaction discovery and validation using health record databases and pharmacokinetics models. *Clin Pharmacol Ther*. 2018;103(2):287-295.
7. Bate A, Lindquist M, Edwards IR, et al. A Bayesian neural network method for adverse drug reaction signal generation. *Eur J Clin Pharmacol*. 1998;54(4):315-321.
8. Noren GN, Bate A, Orre R, Edwards IR. Extending the methods used to screen the WHO drug safety database towards analysis of complex associations and improved accuracy for rare events. *Stat Med*. 2006;25(21):3740-3757.
9. DuMouchel W. Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. *Am Stat*. 1999;53(3):177-190.
10. Zhang P, Li M, Chiang CW, et al. Three-component mixture model-based adverse drug event signal detection for the adverse event reporting system. *CPT Pharmacomet Syst Pharmacol*. 2018;7(8):499-506.
11. Huang L, Zalkikar J, Tiwari RC. A likelihood ratio test based method for signal detection with application to FDA's drug safety data. *J Am Stat Assoc*. 2011;106(496):1230-1241.
12. Huang L, Zheng D, Zalkikar J, Tiwari R. Zero-inflated Poisson model based likelihood ratio test for drug safety signal detection. *Stat Methods Med Res*. 2017;26(1):471-488.
13. DuMouchel W, Pregibon D. Empirical Bayes screening for multi-item associations. Paper presented at: Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2001; San Francisco, CA.
14. Almenoff JS, DuMouchel W, Kindman LA, Yang X, Fram D. Disproportionality analysis using empirical Bayes data mining: a tool for the evaluation of drug interactions in the post-marketing setting. *Pharmacoepidemiol Drug Saf*. 2003;12(6):517-521.
15. Thakrar BT, Grundschober SB, Doessegger L. Detecting signals of drug-drug interactions in a spontaneous reports database. *Br J Clin Pharmacol*. 2007;64(4):489-495.
16. Noren GN, Sundberg R, Bate A, Edwards IR. A statistical methodology for drug-drug interaction surveillance. *Stat Med*. 2008;27(16):3057-3070.
17. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst*. 1959;22(4):719-748.
18. Hopstadius J, Noren GN, Bate A, Edwards IR. Impact of stratification on adverse drug reaction surveillance. *Drug Saf*. 2008;31(11):1035-1048.
19. van Puijenbroek EP, Egberts AC, Heerdink ER, Leufkens HG. Detecting drug-drug interactions using a database for spontaneous adverse drug reactions: an example with diuretics and non-steroidal anti-inflammatory drugs. *Eur J Clin Pharmacol*. 2000;56(9-10):733-738.
20. Goldstein BA, Navar AM, Pencina MJ, Ioannidis JP. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. *J Am Med Inform Assoc*. 2017;24(1):198-208.
21. Ahmed I, Haramburu F, Fourrier-Reglat A, et al. Bayesian pharmacovigilance signal detection methods revisited in a multiple comparison setting. *Stat Med*. 2009;28(13):1774-1792.
22. Efron B, Tibshirani R. Empirical Bayes methods and false discovery rates for microarrays. *Genet Epidemiol*. 2002;23(1):70-86.



23. FDA Adverse Event Reporting System (FAERS). <https://www.fda.gov/drugs/surveillance/fda-adverse-event-reporting-system-faers>.
24. MEDDRA. Medical Dictionary for Regulatory Activities; 2012.
25. Hazell L, Shakir SA. Under-reporting of adverse drug reactions: a systematic review. *Drug Saf*. 2006;29(5):385-396.
26. Poluzzi E, Raschi E, Piccinni C, De Ponti F. Data mining techniques in pharmacovigilance: analysis of the publicly accessible FDA Adverse Event Reporting System (AERS). Data mining applications in engineering and medicine. <https://www.intechopen.com/books/data-mining-applications-in-engineering-and-medicine/data-mining-techniques-in-pharmacovigilance-analysis-of-the-publicly-accessible-fda-adverse-event-re>. Published August 29, 2012.
27. Wu HY, Karnik S, Subhadarshini A, et al. An integrated pharmacokinetics ontology and corpus for text mining. *BMC Bioinform*. 2013;14:35.
28. Wishart DS, Feunang YD, Guo AC, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res*. 2018;46(D1):D1074-D1082.
29. Nash JC, Varadhan R. Unifying optimization algorithms to aid software system users: optimx for R. *J Stat Softw*. 2011;43(9):1-14.
30. Evans SJ, Waller PC, Davis S. Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiol Drug Saf*. 2001;10(6):483-486.
31. Suling M, Pigeot I. Signal detection and monitoring based on longitudinal healthcare data. *Pharmaceutics*. 2012;4(4):607-640.
32. Almenoff JS, Pattishall EN, Gibbs TG, DuMouchel W, Evans SJ, Yuen N. Novel statistical tools for monitoring the safety of marketed drugs. *Clin Pharmacol Ther*. 2007;82(2):157-166.
33. SIDER. Side effect resource. <http://sideeffects.embl.de/>.
34. Palleria C, Di Paolo A, Giofre C, et al. Pharmacokinetic drug-drug interaction and their implication in clinical management. *J Res Med Sci*. 2013;18(7):601-610.
35. DrugBank. <https://www.drugbank.ca/>.
36. Zhang P, Du L, Wang L, et al. A mixture dose-response model for identifying high-dimensional drug interaction effects on myopathy using electronic medical record databases. *CPT Pharmacomet Syst Pharmacol*. 2015;4(8):474-480.
37. Wang X, Zhang P, Chiang CW, et al. Mixture drug-count response model for the high-dimensional drug combinatory effect on myopathy. *Stat Med*. 2018;37(4):673-686.
38. Vilhelmsson A. Consumer narratives in ADR reporting: an important aspect of public health? Experiences from reports to a Swedish Consumer Organization. *Front Public Health*. 2015;3:211.
39. Li X, Li H, Deng J, et al. Active pharmacovigilance in China: recent development and future perspectives. *Eur J Clin Pharmacol*. 2018;74(7):863-871.
40. Han J, Kamber M, Pei J. *Data Mining: Concepts and Techniques*, Waltham, MA: Morgan Kaufmann; 2012.
41. VigiBase. <https://www.who-umc.org/vigibase/vigibase/know-more-about-vigibase/>.
42. Huang YL, Moon J, Segal JB. A comparison of active adverse event surveillance systems worldwide. *Drug Saf*. 2014;37(8):581-596.
43. Strandell J, Caster O, Bate A, Noren N, Edwards IR. Reporting patterns indicative of adverse drug interactions: a systematic evaluation in VigiBase. *Drug Saf*. 2011;34(3):253-266.
44. Strandell J, Caster O, Hopstadius J, Edwards IR, Noren GN. The development and evaluation of triage algorithms for early discovery of adverse drug interactions. *Drug Saf*. 2013;36(5):371-388.
45. Qian Y, Ye X, Du W, et al. A computerized system for detecting signals due to drug-drug interactions in spontaneous reporting systems. *Br J Clin Pharmacol*. 2010;69(1):67-73.
46. Solez K, Racusen LC, Walker WG. Morphologic classification of renal disease and the realities of clinical nephrology: do we need a new approach? *APMIS Suppl*. 1988;4:11-16.
47. Hennessy S, Leonard CE, Gagne JJ, et al. Pharmacoepidemiologic methods for studying the health effects of drug-drug interactions. *Clin Pharmacol Ther*. 2016;99(1):92-100.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Wang X, Li L, Wang L, Feng W, Zhang P. Propensity score-adjusted three-component mixture model for drug-drug interaction data mining in FDA Adverse Event Reporting System. *Statistics in Medicine*. 2020;39:996-1010. <https://doi.org/10.1002/sim.8457>