ARTICLE

Three-Component Mixture Model-Based Adverse Drug Event Signal Detection for the Adverse Event Reporting System

Pengyue Zhang[1], Meng Li[2,3], Chien-Wei Chiang[1], Lei Wang[1,2], Yang Xiang[1], Lijun Cheng[1], Weixing Feng[2], Titus K. Schleyer[4], Sara K. Quinney[5], Heng-Yi Wu[1], Donglin Zeng[6] and Lang Li[1]*

The US Food and Drug Administration (FDA) Adverse Event Reporting System (FAERS) is an important source for detecting adverse drug event (ADE) signals. In this article, we propose a three-component mixture model (3CMM) for FAERS signal detection. In 3CMM, a drug-ADE pair is assumed to have either a zero relative risk (RR), or a background RR (mean RR = 1), or an increased RR (mean RR >1). By clearly defining the second component (mean RR = 1) as the null distribution, 3CMM estimates local false discovery rates (FDRs) for ADE signals under the empirical Bayes framework. Compared with existing approaches, the local FDR's top signals have noninferior or better sensitivities to detect true signals in both FAERS analysis and simulation studies. Additionally, we identify that the top signals of different approaches have different patterns, and they are complementary to each other.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?
☑ The FAERS' risk structure (i.e., background drug-ADE risks generated by comedications and true signals) has not yet been adequately modeled. The FDR of the drug ADE signal detection has not been investigated.
WHAT QUESTION DID THIS STUDY ADDRESS?
☑ The FAERS' risk structure is characterized by the proposed 3CMM. The 3CMM initially estimates local FDR for each drug-ADE pair with respect to the background ADE risk.

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?
☑ The local FDR adds precision in detecting the true drug-ADE signals. Different signal detection methods have different strengths in ranking drug-ADE signals.
HOW MIGHT THIS CHANGE DRUG DISCOVERY, DEVELOPMENT, AND/OR THERAPEUTICS?
☑ The 3CMM can generate drug-ADE signals with a desired false-positive rate. Signals generated by different methods are complementary to each other. These signals are valuable for pharmacological research.

In the United States, adverse drug events (ADEs) account for >3.5 million physician office visits[1] and ∼125,000 hospital admissions each year.[2] About 53% of elders have their hospital stays complicated by ADEs.[3] Many of these ADEs cannot be detected in premarketing clinical trials. Regulatory agencies maintain spontaneous reporting systems (SRSs), which collect reports including patients' medication and ADE information. One of the well-known SRS databases is the US Food and Drug Administration's (FDA's) Adverse Event Reporting System (FAERS).[4] In the past decades, a significant amount of novel ADE knowledge was revealed by SRS analyses.[5] For the aberrations in this article, their full names can be found in **Supplementary Table S1**.

Disproportionality analysis (DPA) is a major method for SRS analysis.[6] For a drug-ADE pair, DPAs compare its reported frequency to expected frequency (expectation) under the assumption of no association between drug and ADE. The ratio of the observed report frequency over its expectation (i.e., relative risk (RR)) or other similarly constructed statistics are used to assess drug-ADE associations. Notable frequentist DPAs includes proportional reporting ratio (PRR),[7] reporting odds ratio (ROR),[8] and likelihood ratio test (LRT).[9] Besides frequentist approaches, the empirical Bayesian approach includes the well-known Empirical Bayesian Geometric Mean (EBGM),[10] and the Bayesian approach includes information component (IC).[11] Details of these DPAs will be reviewed in another section below. Briefly, these approaches do not require sophisticated modeling techniques, are efficient for computation, and are capable of examining different ADEs simultaneously or the whole SRS database at one time.[12] Frequentist DPAs utilize P values to detect signals, whereas the Bayesian and empirical Bayesian DPAs are based on

[1]Department of Biomedical Informatics, College of Medicine, the Ohio State University, Columbus, Ohio, USA; [2]Biomedical Engineering Institute, College of Automation, Harbin Engineering University, Harbin, Heilongjiang, China; [3]CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Shanghai, China; [4]Department of Medicine, Indiana University, Indianapolis, Indiana, USA; [5]Department of Obstetrics and Gynecology, Indiana University, Indianapolis, Indiana, USA; [6]Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA. *Correspondence: L Li (Lang.Li@osumc.edu)
Pengyue Zhang and Meng Li contributed equally to this work.

Model-Based Adverse Drug Event Signal Detection
Zhang *et al*.

500

posterior probabilities.[13] According to a recent comparison performance analysis on the known drug-ADEs pairs[14] using the FAERS data, PRR, ROR, and EBGM have decent performances. Their areas under the receiver operating characteristic (ROC; area under the curve (AUC)) range from 0.71–0.75.[15]

Due to the nature of SRS, the drug-ADE pairs detected by DPAs are often mixed with false positives.[16] For instance, an SRS report includes every drug and ADE related to the patient, whereas the true causal relations between these drugs and ADEs are uncertain. To be specific, there is an average of four drugs per FAERS report, as well as an average of four ADEs. If each ADE is caused by one drug only, the amount of observed false drug-ADE pairs is significantly greater than the amount of true drug-ADE pairs. Besides the uncertainties between reported drugs and ADEs, false positives are generated by incorrectly reported drug and ADE names as well. For instance, about 2,000 drugs are approved by the FDA; although the amount of drug names identified in our FAERS analysis is >300,000. Hence, drug and ADE names must be normalized cautiously to minimize false drug-ADE associations. These false-positive drug-ADE pairs can be considered as background noises and their properties have not been investigated. Even though the DPA signals are contaminated with false positives, they were shown to have high enough specificity for further investigation.[17] Each DPA has its unique strength in ranking the top drug-ADE associations. However, the differences among DPA signal ranking are not well studied yet. As a result, there is a great deal of confusion in selecting top signals for further investigation.

Besides signal ranking, another purpose of a DPA is to differentiate true signals from false-positive signals.[18] In order to select true signals, the SRS nature structure must be considered. We assume the drug-ADE pairs belong to three different groups (mean RR = 0, = 1, or >1). First, we assume that many RRs are equal to 0, as we identify most (90% in our analysis and 70% in DuMouchel[10]) drug-ADE pairs' reported frequencies are 0. From a practical view, an example is that tablets/capsules have no risk to the ADE injection site pain. Second, for the drug-ADE pairs reported at least once, we assume many of them are false positives. As we mentioned above, their report frequencies are generated from either incorrectly reported drug/ADE names or comedications. The observed report frequencies of these drug-ADE pairs will be closely distributed around their expectations. Hence, their RRs are distributed around one. In other words, the distribution of the false-positive frequency can be characterized by the expected reported frequencies of these drug-ADE pairs. In the following sections, we refer this group as the background RR distribution. Third, the remaining drug-ADE pairs with positive reported frequencies have greater RRs than those belonging to the background RR group. Hence, we propose a three-component mixture model (3CMM) for drug-ADE signal detection. By using the background RR distribution as null hypothesis (i.e., H_0 : RR=1), the local false discovery rate (FDR) can be used to identify drug-ADE pairs with increased RRs. Moreover, the properties of top-ranked drug-ADE pairs and the signal detection performances by different DPAs will be investigated.

METHODS
Definitions and notations

Subscript i indicates the ith drug ($1 \leq i \leq I$) and subscript j indicates the jth ADE ($1 \leq j \leq J$). The report frequency N_{ij} is the count of reports containing a drug-ADE pair. Further, $N_{i+}=\sum_j N_{ij}$ is the marginal summation of drug i, and $N_{+j}=\sum_j N_{ij}$ is the marginal summation of ADE j, and $N_{++}=\sum_i \sum_j N_{ij}$ is the summation over all drugs and ADEs. For DPA analysis, $E_{ij}=\frac{N_{+j}}{N_{++}}\times N_{i+}$ is the expectation of N_{ij}. In other words, E_{ij} is the expected frequency under no drug-ADE association. Moreover, let R_{i+} and R_{+j} be the number of reports containing drug i and ADE j, respectively, and R_{++} be the total number of reports.

Review of DPAs

ROR and PRR. $PRR_{ij}=\frac{N_{ij}/R_{i+}}{\left(R_{+j}-N_{ij}\right)/\left(R_{++}-R_{i+}\right)}$, which is the ratio of observed reporting rates. Similarly, $ROR_{ij}=\frac{N_{ij}/\left(R_{i+}-N_{ij}\right)}{\left(R_{+j}-N_{ij}\right)/\left(R_{++}-R_{+j}-R_{i+}+N_{ij}\right)}$.

The variances of PRR and ROR can be calculated by the delta method. Signal detection can be based on the lower bounds of their 95% confidence intervals, which are known as ROR_025 and PRR_025.

LRT. Under the LRT approach, let $N_{ij}^{-}=N_{+j}-N_{ij}$, and $E_{ij}^{-}=\frac{N_{+j}}{N_{++}}\times(N_{++}-N_{i+})$ is the expectation of N_{ij}^{-}. Both N_{ij} and N_{ij}^{-} are assumed to follow Poisson distributions such that $N_{ij} \sim \text{Poisson}(N_{i+}\times p_{ij})$ and $N_{ij}^{-} \sim \text{Poisson}[(N_{++}-N_{i+})\times p_{ij}^{-}]$. The log-likelihood ratio (llr) $\text{llr}_{ij}=N_{ij}\log\left(N_{ij}/E_{ij}\right)+N_{ij}^{-}\log\left[N_{ij}^{-}/E_{ij}^{-}\right]$ is used to test the null hypothesis $H_0 : p_{ij}=p_{ij}^{-}$. Further, for an ADE, $MLR_j=\max_i(\text{llr}_{ij})$ is used for testing $H_0 : p_{ij}=p_{ij}^{-}$ ($1 \leq i \leq I$) vs $H_1=\neg H_0$. Under the null, MLR_j doesn't have a closed form distribution. As a consequence, the authors simulated the null distribution to calculate P values.[9]

IC. This approach assumes that R_{i+} follows a binomial distribution $R_{i+} \sim \text{Bin}(R_{++}, p_{i+})$; and p_{i+} is assumed to have a uniform prior such that $p_{i+} \sim \text{Uniform}(0, 1)$.[11] Similarly, $R_{+j} \sim \text{Bin}(R_{++}, p_{+j})$ and $p_{+j} \sim \text{Uniform}(0, 1)$. N_{ij} is assumed to follow a binomial distribution with a beta prior such that $N_{ij} \sim \text{Bin}(R_{++}, p_{ij})$ and $p_{ij} \sim \text{Beta}\left[1, \left(p_{i+}\times p_{+j}\right)^{-1}\right]$.[11] Dirichelet distribution can be used as the prior distributions as well.[19] Defined as $IC_{ij}=\log_2\left(\frac{p_{ij}}{p_{i+}p_{+j}}\right)$, IC is a measurement of disproportionality. Its posterior expectation was Eq. 1,

$$E\left(IC_{ij}\right)=\log_2\left[\frac{(R_{ij}+1)(R_{++}+2)^2}{(R_{++}+2)^2+R_{++}(R_{i+}+1)(R_{+j}+1)}\right]. \quad (1)$$

and its variance, $V(IC_{ij})$, can be obtained via the delta method. The lower bound of 95% confidence interval (IC_025) is used for detecting signals.

EBGM. This approach assumes $N_{ij} \sim \text{Poisson}(\mu_{ij})$.[10] The relative risk (RR) is $\lambda_{ij}=\frac{\mu_{ij}}{E_{ij}}$ and the observed RR is $\hat{\lambda}_{ij}=\frac{N_{ij}}{E_{ij}}$. The RR is assumed to follow a mixture distribution in which the first component has mean <1 and the second component has mean >1:

Model-Based Adverse Drug Event Signal Detection
Zhang *et al*.

501

$$\lambda_{ij} \sim P\Gamma\left(\lambda_{ij}; \alpha_1, \beta_1\right) + (1-P)\Gamma\left(\lambda_{ij}; \alpha_2, \beta_2\right). \qquad (2)$$

where in Eq. 2, $\Gamma(\lambda; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}\exp(-\beta\lambda)$. For the parameters in Eq. 2, their maximum likelihood estimators can be estimated from the observed likelihood of N_{ij}s. The EBGM is an estimate of drug-ADE association.

$$EBGM_{ij} = 2^{\left[E\left(\log\lambda_{ij}|N_{ij}, E_{ij}\right)/\log 2\right]}. \qquad (3)$$

The 5th percentile of λ_{ij}'s posterior distribution, EB_05, is used for signal detection.

Bayesian false discovery rate. This approach calculates the posterior probabilities with respect to a predefined null hypothesis.[13,20] The estimated posterior probabilities were named the Bayesian False Discovery Rate (BFDR), which refers to the Bayesian false discovery rate. For instance, BFDR is initially derived from the EBGM model,[13] in which the posterior probability of λ_{ij} to be greater than a predefined cutoff point is used for signal selection.

$$BFDR_{ij}(\lambda_{thres}) = P\left(\lambda_{ij} > \lambda_{thres}|N_{ij}, E_{ij}\right). \qquad (4)$$

3CMM and the local FDR. The 3CMM assumes the RRs are distributed either at 0, with mean = 1, or with mean > 1 (Eq. 5):

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

$$\alpha_2=\beta_2, \; \alpha_3 > \beta_3, \; P_1+P_2+P_3=1, \; \text{and} \qquad (5)$$

$$\Gamma(\lambda; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}\exp(-\beta\lambda).$$

The first component in the above model is an identity distribution. It characterizes the drug-ADE pairs with their RRs equal to 0. Under the assumption of $\alpha_2=\beta_2$, the second component in Eq. 5 describes drug-ADE pairs having background RRs. The second component will be defined as the "null distribution" for establishing the local FDR statistic.[21] The third component has its mean >1. It represents the increased RRs generated from true drug-ADE associations. For those drug-ADE pairs that have background or increased RRs, we assume the distribution of their observed report frequencies N_{ij}s to be $Poisson(\lambda_{ij}\times E_{ij})$. If the drug-ADE pairs' RRs are equal to 0, we assume their observed report frequencies to have an identity distribution such that $I(N_{ij}=0)$. The distribution of N_{ij} under 3CMM is:

$$P(N_{ij}) = P_1 I\left(N_{ij}=0\right) + \sum_{l=2}^{3}\left[P_l F\left(N_{ij}; \alpha_l, \beta_l, E_{ij}\right)\right];$$

$$\alpha_2=\beta_2, \; \alpha_3 > \beta_3 \; \text{and} \; P_1+P_2+P_3=1. \qquad (6)$$

In Eq. 6, $F\left(N_{ij}; \alpha_l, \beta_l, E_{ij}\right) = \frac{\Gamma\left(N_{ij}+\alpha_l\right)\times E_{ij}^{N_{ij}}\times\beta_l^{\alpha_l}}{\Gamma(\alpha_l)\times N_{ij}!\times\left[E_{ij}+\beta_l\right]^{N_{ij}+\alpha_l}}$ is the negative binomial distribution. The log-likelihood function of Eq. 6 is:

$$ll\left(\boldsymbol{N_{ij}}; \boldsymbol{E_{ij}}, \{\alpha_2=\beta_2, \alpha_3, \beta_3, P_1, P_2\}\right) = \sum_i\sum_j\log P(N_{ij});$$

$$P_3=1-P_1-P_2. \qquad (7)$$

For pharmacovigilance study, drug-ADE pairs with positive report frequencies are of more interest. Their conditional distribution is:

$$P\left(N_{ij}=k|N_{ij}>0\right) = \frac{F\left(k; \alpha_2, \beta_2, E_{ij}\right) + \frac{P_3}{P_2}\times F\left(k; \alpha_3, \beta_3, E_{ij}\right)}{P\left(N_{ij}>0; \alpha_2, \beta_2, E_{ij}\right) + \frac{P_3}{P_2}\times P\left(N_{ij}>0; \alpha_3, \beta_3, E_{ij}\right)}. \qquad (8)$$

In Eq. 8, $P\left(N_{ij}>0; \alpha_l, \beta_l, E_{ij}\right)$ is the probability of $N_{ij}>0$, based on $F\left(N_{ij}; \alpha_l, \beta_l, E_{ij}\right)$. Their report frequencies can be modeled by the conditional log-likelihood function in Eq. 9:

$$ll\left(\boldsymbol{N_{ij}}; \boldsymbol{E_{ij}}, \{\alpha_2=\beta_2, \alpha_3, \beta_3, r\}\right) = \sum_i\sum_j\log\left[P\left(N_{ij}=k|N_{ij}>0\right)\right] \qquad (9)$$

Instead of estimating P_2 and P_3 separately, the conditional model in Eq. 9 considers their ratio as a single parameter. Thus, Eq. 9 has four parameters $\{\alpha_2=\beta_2, \alpha_3, \beta_3, r=P_3/P_2\}$. The local FDR statistic is:

$$local\; FDR(N_{ij}) = \frac{P_2\times F\left(N_{ij}; \alpha_2, \beta_2, E_{ij}\right)}{P_1\times I\left(N_{ij}=0\right) + \sum_{l=2}^{3}\left[F\left(N_{ij}; \alpha_l, \beta_l, E_{ij}\right)\times P_l\right]}. \qquad (10)$$

It is equivalent to the posterior probability of a drug-ADE pair to have a null RR. If $N_{ij}>0$, the local FDR can be simplified as:

$$local\; FDR\left(N_{ij}=k|N_{ij}>0\right) = \frac{F\left(k; \alpha_2, \beta_2, E_{ij}\right)}{F\left(k; \alpha_2, \beta_2, E_{ij}\right) + r\times F\left(k; \alpha_3, \beta_3, E_{ij}\right)}. \qquad (11)$$

Like the EBGM model, under the 3CMM, posterior expectations of the RRs for the drug-ADE pairs with positive report frequencies can be estimated simultaneously. The 3CMM-based EBGM (3C_EBGM) is:

$$3C_{EBGM_{ij}} = 2^{\frac{E\left(\log\lambda_{ij}|N_{ij}, E_{ij}\right)}{\log 2}}. \qquad (12)$$

In Eq. 12, $E(\log\lambda_{ij}|N_{ij}, E_{ij}) = local\; FDR(N_{ij})\times\left[\Psi(\alpha_2+N_{ij}) - \log(\beta_2+E_{ij})\right] + \left[1-local\; FDR(N_{ij})\right]\times\left[\Psi(\alpha_3+N_{ij}) - \log(\beta_3+E_{ij})\right]$, where $\Psi(x)$ is the digamma function.

Drug-ADE signal ranking

For data analyses and simulations, local FDR is used to rank top drug-ADE signals generated from our 3CMM model. The PRR, IC, and EBGM are used to rank top drug-ADE signals generated from their models. For LRT, it is computationally infeasible to estimate extreme P values (i.e., $<1^{-10}$) through simulation studies. Thus, llr is used to rank signals. They are consistent with their P value-based rankings.[9] We calculated the BFDR under the EBGM model

and set $\lambda_{thres} = 2$ according to Ahmed *et al.*[13] The BFDR is used to rank top drug-ADE signals.

Control of confounding bias

In this section, we define a propensity score (PS) adjusted expectation to control confounding bias. Both the regular expectation (defined in the Definitions and Notations section) and the adjusted expectation will be utilized for signal detection. To calculate the adjusted expectation, principle components (PCs) are derived from the drug matrix, in which each column corresponds to a drug and each row is a report. For our analysis, we use the first 100 PCs and the logistic regression model in Eq. 13 to calculate the PS.

$$\text{logit}[P(Drug=1)] = \beta_0 + \sum_{i=1}^{100} \beta_i \times PC_i \qquad (13)$$

For a drug-ADE pair, we fit another logistic regression model in Eq. 14, in which response variable is the ADE status and covariates include the drug status and its PS.

$$\text{logit}[P(ADE=1)] = \beta_0 + \beta_1 \times Drug + \beta_2 \times PS \qquad (14)$$

From Eq. 14, for K reports, the PS adjusted expectation is defined as:

$$E = \sum_{k=1}^{K} [P(ADE=1|Drug=0 \& PS_k) \times 1(D_k=1)] \qquad (15)$$

In other words, Eq. 15 is the production of the drug frequency and the average of PS adjusted drug absent ADE risks.

Parameter estimation

We utilized the particle swarm optimization (PSO) to estimate the maximum likelihood estimators.[22] The practices are five-dimensional and four-dimensional vectors with respect to the log-likelihood (Eq. 7) and conditional log-likelihood (Eq. 9). Let $\boldsymbol{X}$ be the particle's position, $\boldsymbol{V}$ be the particle's velocity, subscript t ($t=1, \ldots, T$) indicates the tth particle, and superscript s ($s=1, \ldots, S$) indicates the sth step. In each step, the PSO identifies the local and global best such that $\boldsymbol{L}_t = \arg\max_{\boldsymbol{X}_t^s \ s=1,\ldots,\ S} \ ll(\boldsymbol{N}_{ij}; \boldsymbol{E}_{ij}, \boldsymbol{X}_t^s)$ and $\boldsymbol{G} = \arg\max_{\boldsymbol{L}_t \ t=1,\ldots,\ T} \ ll(\boldsymbol{N}_{ij}; \boldsymbol{E}_{ij}, \boldsymbol{L}_t)$. Then, the particles' positions ($\boldsymbol{X}$) and velocities ($\boldsymbol{V}$) are updated by:

$$\boldsymbol{V}_t^s = w^s \boldsymbol{V}_t^{s-1} + U(0,1)C_1\left(\boldsymbol{L}_t^{s-1} - \boldsymbol{X}_t^{s-1}\right) +$$
$$U(0,1)C_2\left(\boldsymbol{G}^{s-1} - \boldsymbol{X}_t^{s-1}\right) \text{ and } \boldsymbol{X}_t^s = \boldsymbol{X}_t^{s-1} + \boldsymbol{V}_t^s. \qquad (16)$$

In Eq. 16, the weight w^s is set to be $w^s = (w_{\max} - w_{\min}) \times \frac{(iter_{\max} - iter)}{iter_{\max}} + w_{\min}$, and $U(0,1)$ is a number generated uniformly from 0 to 1.[23] The PSO is carried out by setting the particles with random starting points, and iterating until they converge.

FAERS data processing

The FAERS reports are stored quarterly for each year. The reports from the first quarter of 2004 to the third quarter of 2012 were selected. The primary ID numbers of the reports

Table 1 Risk profiles for full FAERS data and the four ADE data

Group	Full data		Four ADE data with regular expectation		Four ADE data with adjusted expectation	
	Mean RR [SD]	%	Mean RR [SD]	%	Mean RR [SD]	%
Background risk	1 [0.73]	81	1 [0.66]	90	1 [0.75]	85
Increased risk	4 [11.5]	19	5.10 [8.17]	10	4.96 [6.52]	15

ADE, adverse drug events; FAERS, US Food and Drug Administration Adverse Event Reporting System; RR, relative risk.

were used to filter duplications. Moreover, drugs recorded to treat indicated ADEs were removed in order to reduce the indication bias. Our initial database contains 4,070,770 reports, 15,445 unique MedDRA[24] preferred terms, and 356,734 distinct drug names. Drug Bank IDs[25] were adopted to normalize the drug names. Manual corrections were made for the frequent drug names (>999 reports) that cannot be mapped with Drug Bank ID. These corrections included correction for incorrect formations or drug names with additional information. For instance, "simvastatin tablets 20 mg" was manually revised to "simvastatin." The final data had 1,735 distinct drug names.

We selected reports containing 92 MedDRA[24] preferred terms that belong to myopathy, neuropathy, delirium, and skin pigmentation disorder for simulations and analyses as well. This dataset is named as four ADE data in the following sections. The ADE names and their frequencies are given in **Supplementary Table S2**. Due to computational burden, adjusted expectations were only calculated for the four ADE data. The adjusted expectations will be used to calculate EBGM, BFDR, and local FDR.

Drug-ADE signal validation and evaluation

The side effect resource (SIDER)[26] is a database established from drug labels. As it includes the labeled drug-ADE associations, we adopted it to validate the top signals of different DPAs. In the validation process, we utilized the drug name mapping tool developed by Wu *et al.*[27] to normalize drug names between SIDER and FAERS.

The Observational Medical Outcomes Partnership (OMOP) gold standard[14] was designed to establish a reference set for pharmacovigilance study. It contains 399 drug-ADE pairs that were made up of 181 drugs and 4 ADEs (acute myocardial infarction, acute renal failure, acute liver injury, and gastrointestinal bleeding). These 399 drug-ADE pairs are classed as 165 true positives and 234 true negatives. The performances of signal detection for local FDR and other DPAs will be evaluated by the OMOP gold standard. The signal detection performances are evaluated by areas under the ROC curve (AUCs).

RESULTS

FAERS risk profile

The conditional 3CMM (Eq. 9) had been applied to both full FAERS data and the four ADE data. The estimated risk structures for the full data, four ADE data with regular expectation, and four ADE data with adjusted expectation are shown in **Table 1**. Generally, among all drug-ADE pairs
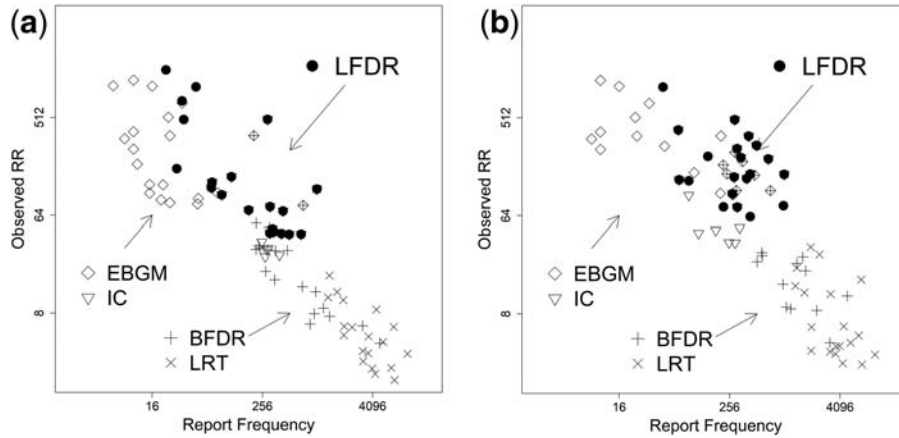
Model-Based Adverse Drug Event Signal Detection
Zhang *et al*.

503

Figure 1 (**a**) The report frequencies and the observed relative risks (RRs) for the top-20 ranked signals by different methods with regular expectation. (**b**) The report frequencies and the observed RRs for top-20 ranked signals by different methods with adjusted expectation. BFDR, Bayesian False Discovery Rate; EBGM, Empirical Bayesian Geometric Mean; IC, information component; LFDR, local false discovery rate; LRT, likelihood ratio test.

with non-zero risk, 19% of them have an average of RR of 4. We also fitted the EBGM model (Eq. 2) to the full FAERS data. Under EBGM model, the drug-ADE pairs have two RR mean estimates: 0.76 or 3.67. The EBGM model's second component (mean = 3.67) is similar to the third component of 3CMM (mean = 4.0). On the other hand, the first
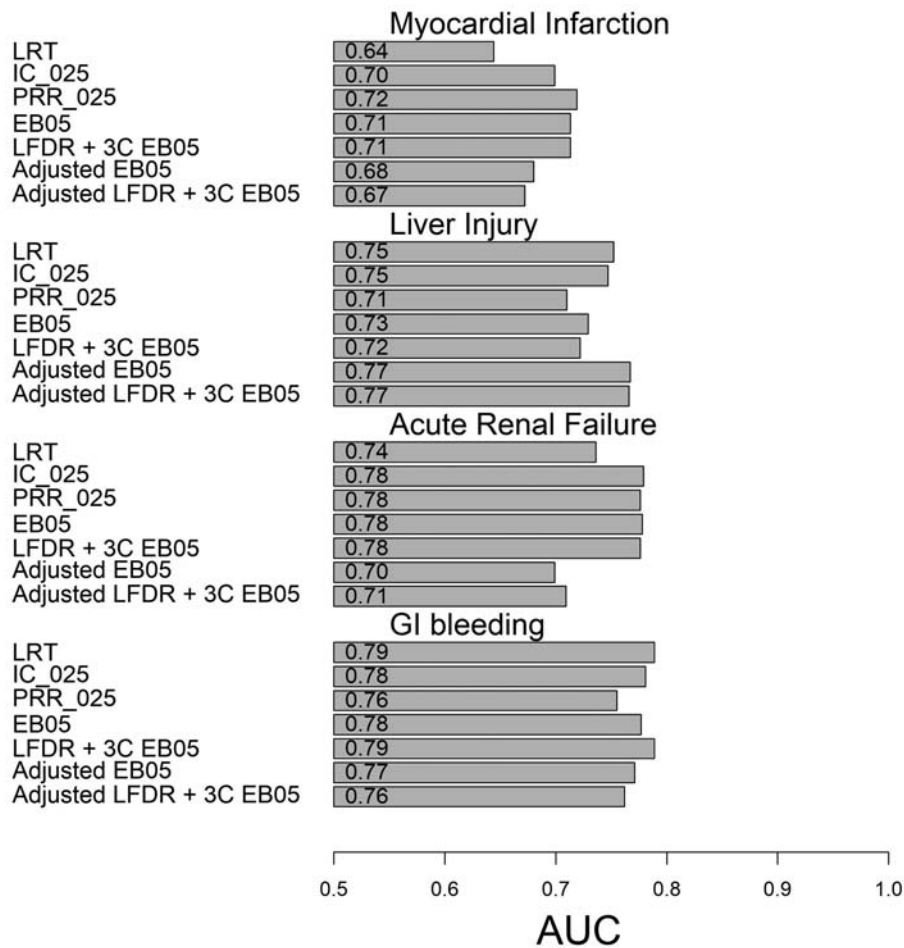

Figure 2 Signal detection algorithm performances (area under the curve (AUC)) classified by event. IC, information component; LFDR, local false discovery rate; LRT, likelihood ratio test; PRR, proportional reporting ratio.
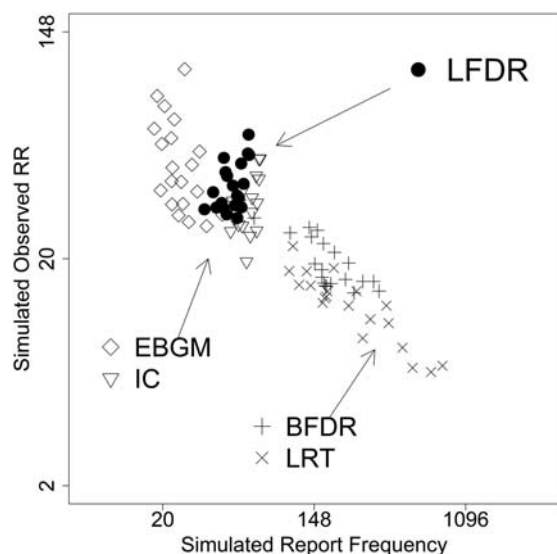
504

Figure 3 The average simulated report frequencies and the simulated observed relative risks (RRs) for top-20 ranked signals by different methods. BFDR, Bayesian False Discovery Rate; EBGM, Empirical Bayesian Geometric Mean; IC, information component; LFDR, local false discovery rate; LRT, likelihood ratio test.

component of the EBGM model (mean = 0.76) is a mixture of 3CMM's components one and two. By using regular expectation, 10% drug-ADE pairs in the four ADE data have an average of RR of 5.10. Adjusted expectation yielded the risk structure such that 15% drug-ADE pairs have an average of RR of 4.96. More information about model fitting is presented in the **Supplementary Figures S1-S3**.

Properties of the top-ranked signals

The comparison of top-ranked drug-ADE signals among DPAs was performed on the four ADE datasets. We observed that the top 20 signals generated from different DPAs had different report frequencies and observed RRs (**Figure 1a,b**). For both regular and adjusted expectations, local FDR top signals have moderate report frequencies but a greater magnitude on the observed RRs. We discovered that EBGM and IC yielded similar top-20 ranked drug-ADE pairs. Their report frequencies are small. On the contrary, the report frequencies of BFDR's and LRT's top signals are significantly larger.

Top-20 signals were further investigated by SIDER.[26] For the frequentist methods, 10 of the top 20-ranked signals identified by LRT can be validated in SIDER, and 1 by PRR. The Bayesian method, IC, has 5 drug-ADE pairs. For the empirical Bayesian methods with adjusted expectation, EMGM, BFDR, and local FDR have 4, 4, and 5 overlapped drug-ADE, respectively. Alternatively, with regular expectation, EBGM, BFDR, and local FDR have 3, 6, and 5 overlapped pairs. Interestingly, the top-20 ranked signals by different methods are complementary to each other, such that nearly all methods' top-20 signals can identify unique SIDER documented drug association(s). The top 20 ranked signals by each method are in **Supplementary Table S3**.

DPA performance evaluation by OMOP gold standard

Using OMOP gold standard,[14] performances of EBGM, PRR, and ROR were compared by Ryan *et al.*,[14] in which EBGM was shown to have the best performance (i.e., AUC). In this study, we extend the comparisons to local FDR, LRT, and IC. Additionally, the performances of EBGM and local FDR under regular and adjusted expectations will be evaluated. As 3CMM generates both 3C_EB05 and local FDR, their combination can be used for performance evaluation. We used a weighted combination of 3C_EB05 and local FDR (local FDR+3C_EB05), where the weights are derived from logistic regressions. As we mentioned in the Drug-ADE signal ranking section, llr is used to evaluate the performance of LRT. In our analysis, BFDR is not examined because BFDR is equivalent to EB05 (i.e., BFDR($\lambda_{Thres}=2$) < 0.05 is equivalent with EB05 > 2). Their performances are shown in **Figure 2**. For frequentist approaches, LRT has better AUCs than PRR in three out of four ADEs. The IC_025, the only Bayesian approach, has comparable AUCs to the other DPAs. By using regular expectation, local FDR+3C_EB05 is noninferior to or better in three of four ADEs. Alternatively, by using adjusted expectation, local FDR+3C_EB05 has best performance in liver injury. The AUCs of local FDR+3C_EB05 under equal weights were examined as well. Under regular expectation, the AUCs are 0.71 for myocardial infraction, 0.72 for liver injury, 0.78 for acute renal failure, and 0.76 for gastrointestinal bleeding. Under adjusted expectation, the AUCs are 0.67 for myocardial infraction, 0.77 for liver injury, 0.71 for acute renal failure, and 0.76 for gastrointestinal bleeding. For the ROC curves, please visit **Supplementary Figure S4**.

Simulation study

In order to maintain the drugs' correlation structure, we chose a random subset of FAERS containing 40,000 reports. We further selected 100 drugs randomly for our simulation study. In the simulation studies, 20 ADEs were simulated by assuming 5 causal drugs per ADE. For multiple causal drugs on a report, we assumed the risks to be multiplicative. For each simulation, ADE status for each report was simulated first. Then, we summarized the reports into report frequencies of 2,000 drug-ADE pairs. The causal drug-ADE pairs represented true signals, and the background signals were generated from the FAERS' comedication structure.

We conducted 1,000 simulations. First, the top 50 and 20 drug-ADE pairs ranked by different DPAs have at least 90% to be casual drug-ADE pairs except for PRR. Second, the average observed RRs and report frequencies for different DPAs' top-20 signals were examined. Results show a consistent pattern as we observed from our FAERS analysis (**Figure 3**). The report frequencies of the EBGM and IC top signals are small; the report frequencies of the LRT and BFDR top signals are large; and the report frequencies of the local FDR top signals are in the middle. For the observed RRs, a reverse trend was observed.

In addition, simulation studies were conducted to examine the true positive rate (TPR) and the local FDR consistency. We examined the TPR for the top-20, 50, 100, and 200 ranked signals. Results show that EBGM, IC, and local FDR rankings have nearly 100% TPRs (**Supplementary Figure**

Model-Based Adverse Drug Event Signal Detection
Zhang *et al*.

505

S5). From another simulation, we observed that the local FDR estimators are consistent, if the drug-ADE pairs are following the 3CMM independently (**Supplementary Figure S6a**). Although, if the independent assumption does not hold, then the local FDR estimators are underestimated (**Supplementary Figure S6b**). More detailed data and information are presented in **Supplementary Figure S2**.

CONCLUSIONS

This article presents a novel model (3CMM) for detecting drug-ADE associations. Under 3CMM, local FDR is defined to be the posterior probability of a drug-ADE pair to have an increased RR with respect to the null. On one hand, local FDR's top signals show reasonable power to detect true signals from the FAERS database. In addition, local FDR have comparable or improved performances in OMOP analysis as well. On the other hand, simulation studies show local FDR has noninferior or better abilities to select casual drug-ADE pairs. Thus, local FDR is a decent statistic for signal ranking/detection. Additionally, local FDR is more statistically meaningful regarding FDR, compared to traditional DPA statistics.

The observed RRs and report frequencies for different DPAs' top signals were examined. An interesting finding is that DPAs have different patterns of their top signals. The report frequencies of the EBGM and IC top-20 signals range from 10 to 50, which are among the smallest. Although their top signals' observed RRs are the highest (between 20 and 200). The LRT and BFDR top-20 signals show a contrary pattern. Their report frequencies are at least 200 and up to a few thousand. Although their observed RRs are between 2 and 15. For local FDR, both its top-20 signals' report frequencies (20 to 400) and observed RRs (16 to 200) are moderate. To summarize, the IC and EBGM top signals are rare drug-ADE pairs with large RRs; LRT's and BFDR's top signals are common drug-ADE pairs with low RRs; and local FDR's top signals are drug-ADE pairs with moderate frequencies and RRs. Comparing the top-signals, nearly all methods can identify unique SIDER[26] documented drug-ADE association(s). These are important evidences that the signal detection methods are complementary to each other. Using the OMOP golden standard drug-ADE pairs, we show that combining both 3CMM statistics, 3C_EB05 and local FDR, they have comparable or better AUC performance in selecting the true drug-ADE signals. In detecting liver injury-related drugs, PC adjusted 3C_EB05+local FDR has the best AUC. For real application, equal weights can be used to combine local FDR+3C_EB05. Further, none of the method has uniformly better performance than other methods. This is more evidence that different DPAs are complementary.

The 3C_EBGM and 3C_EB05 generated by the 3CMM are consistent with DuMouchel's EBGM[10] and EB05 (**Supplementary Figure S7**). In the four ADE data analysis, DuMouchel's EBGM and our 3C_EBGM have the same top-20 ranked drug-ADE pairs. Through 3CMM, local FDR is naturally defined with respect to the null RR distribution.

As described in the 3CMM and the local FDR section, 3CMM is derived based on the nature structure of the ADE risks. The background risk (null) is properly defined only under 3CMM. Alternatively, under a two-component mixture model in which one component represents background risk (mean $RR = 1$) and the other component characterizes increased risk (mean $RR > 1$), the null distribution is misspecified. As a consequence, the local FDR will be improper. The rationale of 3CMM is also supported by the fitted EBGM model, in which the first component have mean $RR = 0.76$. The first component of the EBGM model represents a mixture of the background risk (mean $RR = 1$) and the zero-risk component. Under 3CMM, the zero-risk component is not identifiable from the data. Although, this issue can be solved by the conditional inference approach, in which the background and increased risks can be estimated. Compared with DuMouchel's model, 3CMM initially estimated the background risk of the FAERS database. Another contribution is the added local FDR statistics, which measures the false-positive drug-ADE signals. Simulation results show that the model-based local FDR are consistent with empirical false discovery rates. Thus, the uncertainty is not a major challenge for using the local FDR estimates. The local FDR statistics can be used to prioritize drug-ADE signals alone. Alternatively, it can be combined with other methods. For instance, local FDR can be used to evaluate the FDR for the top-signals generated by different methods as well.

In this analysis, PS was used to control confounding variables. Particularly, 100 PCs were used to estimate the PSs. A plot of number of PCs vs. percentage of variation explained is given in the (**Supplementary Figure S8**), and 100 PCs explain 57.36% of the total variation. For our analysis, incorporating 100 PCs costs about 10 GB of computational memory (sample size about 4.07 million) to fit the propensity score model. Both computational resources and statistical knowledge (i.e., percentage of variation explained) are two factors to determine the number of PCs for the propensity score model. In our analysis, we use PS as a covariate. Additionally, PS can be also used to match observations or as weight. Our choice is a computational-driven approach, as PS-matching is computationally expensive and using PS as weight may yield instable model estimation. Moreover, confounding variables, such as demographic variables and clinical variables, can be controlled by either multiple regression or propensity score analysis,[28,29] which are multivariate extension of PRR or POR methods. However, integrate confounding variables into the other DPA methods are not straightforward, and need significant further methodology development. For instance, vast FAERS reports are missing age and gender information. Further, we identify that the adjusted expectation is not performed uniformly better than regular expectation in the OMOP analysis. We identify techniques, such as traditional multiple regression and propensity score analysis, are underpowered on handling highly correlated drugs. Such correlations are generated by the increasing trend of co-prescriptions and polypharmacy.[30] As a consequence, in such a situation, multivariate analyses may yield reduced power to detect signals compared with univariate analyses.

Last, the proposed local FDR derived from 3CMM is fundamentally different from the BFDR.[13] Under 3CMM, if a drug-ADE pair has a positive report frequency, its RR would either follow a gamma distribution with mean = 1, or a gamma distribution with mean >1. These two distributions represent the background (null) and increased (alternative) RR distributions. The 3CMM shares same model frame work with Efron's local FDR model, in which null and alternative distributions are clearly specified. Additionally, the local FDR from 3CMM follows the theory of Storey.[31] The BFDR is derived from the EBGM model which does not have null and alternative distributions. The BFDR is defined to be the posterior probability of RR to be greater than a predefined threshold. Thus, the BFDR null hypothesis does not characterize the null or false positive distribution. On the contrary, 3CMM characterizes the SRS nature risk structure; models the RR by specifying the null and alternative distributions; and estimates the local FDR purely from the drug-ADE pairs' report frequencies.

Conflict of Interest. The authors declared no competing interests for this work. As an Associate Editor for *CPT: Pharmacometrics & Systems Pharmacology*, Lang Li was not involved in the review or decision process for this paper.

Author Contributions. P.Z., S.K.Q., and L.L. wrote the manuscript. L.L. designed the research. P.Z. and M.L. performed the research. P.Z., M.L., and C.C. analyzed the data. H.W., L.W., Y.X., L.C., T.K.S., W.F., and D.Z. contributed new reagents/analytical tools.

1. Bourgeois, F.T., Shannon, M.W., Valim, C. & Mandl, K.D. Adverse drug events in the outpatient setting: an 11-year national analysis. *Pharmacoepidemiol. Drug Saf.* **19**, 901–910 (2010).
2. Budnitz, D.S., Pollock, D.A., Weidenbach, K.N., Mendelsohn, A.B., Schroeder, T.J. & Annest, J.L. National surveillance of emergency department visits for outpatient adverse drug events. *JAMA* **296**, 1858–1866 (2006).
3. Lucado, J., Paez, K. & Elixhauser, A. Medication-related adverse outcomes in U.S. hospitals and emergency departments, 2008. <http://www.hcup-us.ahrq.gov/reports/statbriefs/sb109.pdf> (2008).
4. US Food and Drug Administration (FDA) Adverse Event Reporting System (FAERS). <https://www.fda.gov/Drugs/InformationOnDrugs/ucm135151.htm> (2017). Accessed 1 November 2017.
5. Szarfman, A., Tonning, J.M. & Doraiswamy, P.M. Pharmacovigilance in the 21st century: new systematic tools for an old problem. *Pharmacotherapy* **24**, 1099–1104 (2004).
6. Montastruc, J.L., Sommet, A., Bagheri, H. & Lapeyre-Mestre, M. Benefits and strengths of the disproportionality analysis for identification of adverse drug reactions in a pharmacovigilance database. *Br. J. Clin. Pharmacol.* **72**, 905–908 (2011).
7. Evans, S.J., Waller, P.C. & Davis, S. Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiol. Drug Saf.* **10**, 483–486 (2001).
8. van Puijenbroek, E.P., Bate, A., Leufkens, H.G., Lindquist, M., Orre, R. & Egberts, A.C. A comparison of measures of disproportionality for signal detection in spontaneous reporting systems for adverse drug reactions. *Pharmacoepidem. Drug Saf.* **11**, 3–10 (2002).
9. Huang, L., Zalkikar, J.N. & Tiwari, R.C. A likelihood ratio test based method for signal detection with application to FDA's Drug Safety Data. *J. Am. Stat. Assoc.* **106**, 1230–1241 (2011).
10. DuMouchel, W. Bayesian data mining in large frequency tables, with an application to the FDA Spontaneous Reporting System. *Am. Stat.* **53**, 177–190 (1999).
11. Bate, A. *et al.* A Bayesian neural network method for adverse drug reaction signal generation. *Eur. J. Clin. Pharmacol.* **54**, 315–321 (1998).
12. Harpaz, R., DuMouchel, W., Shah, N.H., Madigan, D., Ryan, P. & Friedman, C. Novel data-mining methodologies for adverse drug event discovery and analysis. *Clin. Pharmacol. Ther.* **91**, 1010–1021 (2012).
13. Ahmed, I. *et al.* Bayesian pharmacovigilance signal detection methods revisited in a multiple comparison setting. *Stat. Med.* **28**, 1774–1792 (2009).
14. Ryan, P.B., Schuemie, M.J., Welebob, E., Duke, J., Valentine, S. & Hartzema, A.G. Defining a reference set to support methodological research in drug safety. *Drug Saf.* **36** Suppl 1, S33–S47 (2013).
15. Harpaz, R., DuMouchel, W., LePendu, P., Bauer-Mehren, A., Ryan, P. & Shah, N.H. Performance of pharmacovigilance signal-detection algorithms for the FDA Adverse Event Reporting System. *Clin. Pharmacol. Ther.* **93**, 539–546 (2013).
16. Ali, A.K. Pharmacovigilance analysis of adverse event reports for aliskiren hemifumarate, a first-in-class direct renin inhibitor. *Ther. Clin. Risk. Manag.* **7**, 337–344 (2011).
17. Szarfman, A., Machado, S.G. & O'Neill, R.T. Use of screening algorithms and computer systems to efficiently signal higher-than-expected combinations of drugs and events in the US FDA's spontaneous reports database. *Drug Saf.* **25**, 381–392 (2002).
18. Wang, L., Jiang, G., Li, D. & Liu, H. Standardizing adverse drug event reporting data. *J. Biomed. Semantics* **5**, 36 (2014).
19. Norén, G.N., Bate, A., Orre, R. & Edwards, I.R. Extending the methods used to screen the WHO drug safety database towards analysis of complex associations and improved accuracy for rare events. *Stat. Med.* **25**, 3740–3757 (2006).
20. Ahmed, I., Dalmasso, C., Haramburu, F., Thiessard, F., Broët, P. & Tubert-Bitter, P. False discovery rate estimation for frequentist pharmacovigilance signal detection methods. *Biometrics* **66**, 301–309 (2010).
21. Efron, B., Tibshirani, R., Storey, J.D. & Tusher, V.G. Empirical Bayes analysis of a microarray experiment. *J. Am. Stat. Assoc.* **96**, 1151–1160 (2001).
22. Kennedy, J. & Eberhart, R. Particle swarm optimization. 1995 IEEE International Conference on Neural Networks Proceedings, November 27-December 1, 1995. Perth, WA, Australia.
23. Chaturvedi, K.T., Pandit, M. & Srivastava, L. Self-organizing hierarchical particle swarm optimization for nonconvex economic dispatch. *IEEE Trans. Power Syst.* **23**, 1079–1087 (2008).
24. MedDRA. <https://www.meddra.org> (2017). Accessed 1 November 2017.
25. Drug Bank. <https://www.drugbank.ca> (2017). Accessed 1 November 2017.
26. SIDER: Side Effect Resource. <http://sideeffects.embl.de> (2015). Accessed 1 November 2017.
27. Wu, H.Y. *et al.* An integrated pharmacokinetics ontology and corpus for text mining. *BMC Bioinformatics* **14**, 35 (2013).
28. Tatonetti, N.P., Ye, P.P., Daneshjou, R. & Altman, R.B. Data-driven prediction of drug effects and interactions. *Sci. Transl. Med.* **14**, 125ra31 (2012).
29. Tatonetti, N.P. *et al.* Detecting drug interactions from adverse-event reports: interaction between paroxetine and pravastatin increases blood glucose levels. *Clin. Pharmacol. Ther.* **90**, 133–142 (2011).
30. Prybys, K.M. Melville, K.A. & Hanna, J.R. Polypharmacy in the elderly: clinical challenges in emergency practice - part 1 overview, etiology, and drug interactions. *Emerg. Med. Rep.* **23**, 145–153 (2002).
31. Storey, J. A direct approach to false discovery rates. *J. R. Stat. Soc. Series B. Stat. Methodol.* **64**, 479–498 (2002).

Supplementary information accompanies this paper on the *CPT: Pharmacometrics & Systems Pharmacology* website (http://psp-journal.com)