Original Research   **Open Access**

A Generalized Likelihood Ratio Test for Estimating Drug Adverse Reactions

Yeqian Liu

*Correspondence: yeqian.liu@mtsu.edu

CrossMark
← Click for updates

Department of Mathematical Sciences, Middle Tennessee State University, Murfreesboro, TN, USA.

Abstract

Likelihood ratio test is widely used for detecting adverse reactions (ARs) of single drug in biomedical studies. However, it is difficult to detect adverse reactions of multiple drugs simultaneously. Corresponding to this, we consider a generalized likelihood ratio test procedure to detect adverse reactions simultaneously for all the drugs of the same class. An extensive simulation study is performed to evaluate the proposed test procedural as well as its power and sensitivity. The simulation study suggest that the proposed test procedure seems to work well for practical situations and an illustrative example is also provided.

Keywords: Likelihood ratio test, Adverse reaction, Type I error, Sensitivity

Introduction

Reporting of drug or medical device related adverse reactions (ARs) is usually voluntary. One of the major postmarket safety surveillance databases is the World Health Organization's (the "WHO") global pharmacovigilance database, which contains reports of suspected ADRs, so called Individual Case Safety Reports (ICSRs), collected by national drug authorities in over 110 countries and span over more than 100 000 different medicinal products. Clinical reviewers evaluate adverse reactions reports to look for new safety concerns that might be related to a marketed product, or for a manufacturer's compliance to reporting regulations.

There are several statistical methods available for adverse reactions detection in pharmacovigilance databases. These methods include the reporting odds ratio (ROR, [7]), proportional reporting ratio (PRR, [3]), multi-gamma Poisson shrinker (MGPS, [2,8]), Bayesian confidence propagation neural network (BCPNN, [1]), Bayesian method based on a new Information Component (IC, [6]), simplified Bayes (sB, [5]), among others. A likelihood ratio test-based method ([4]), that assumes the number of adverse reactions follows a Poisson distribution, was developed to identify signals of adverse events for a particular drug or to identify drugs for a particular adverse event. In post-marketing surveillance, the signals of adverse reactions

within a drug class, or drug signals for a group of adverse reaction may be of interest to medical reviewers. In this article, we develop a generalized likelihood ratio test to identify adverse reactions that have high reporting rates compared to other adverse reactions associated with all the drugs of the same class or with similar treatment indications. The drug class refers to a set of drugs which have similar chemical structure such as the antibiotics drug class containing Penicillins, Tetracyclines, Cephalosporins, Quinolones, Lincomycins, Macrolides, Sulfonamides, Glycopeptides, Aminoglycosides and Carbapenems, etc. A group of adverse reactions refers to a set of preferred terms such as hepatic failure, alanine aminotransferase abnormal, ascites, blood bilirubin abnormal, cholestatic liver injury, hepatic atrophy, hepatomegaly, Reye's syndrome, and so on that are all related to hepatocellular injury.

This article is organized as followed. In section 2, a brief overview of World Health Organization's global pharmacovigilance database is provided. In Section 3, we give a brief review of the likelihood ratio test procedure for adverse reactions detection for a single drug, then we propose a generalized likelihood ratio test procedure, namely GLRT, to detect multiple ARs in a drug class. The performance of GLRT is evaluated using simulated datasets in Section 4. In Section 5, both the LRT and the GLRT are applied to the 2000-2005 and 2005-2010 data from WHO's

global database. Section 6 contains some discussion and concluding remarks.

2 World Health Organization's Pharmacovigilance Database

The WHO's global pharmacovigilance database consists of the individual reports with demographical information, route of administration, drug/biological information, medical history, treatment indication, therapy start dates, and end dates. For adverse reaction detection, Medical Dictionary for terminology of preferred terms is often used to identify the adverse events, such as Death, Stroke, Myocardial infarction, and so on. There are also verbatim drug names in the file for drug/biologic information. In studying the drug-AE association, the generic name of the drug is used, which refers to the unique chemical makeup of a drug.

The WHO's global pharmacovigilance database includes reports since 1980, however researchers and reviewers are more interested in data from recent years. In this article we focus on cases reported to WHO between 2000 and 2010 for more than 6500 drugs and 14,000 Adverse reactions. For any particular adverse event, the investigators consider all suspect and concomitant drugs.

3 Inference Procedure

3.1 Test Procedure for Adverse Reactions Detection of A single Drug

After summarizing the data files, the WHO pharmacovigilance data can be presented in a tabular form with, say, adverse reactions (ARs) as the row variable and drugs as the column variable (as in **Table 1**), with n_{ij} as the cell count for ith AR and jth drug, n_i as the sum of counts for ith AR (ith row total) and n_{ij} as the sum of counts for jth drug (jth column total).

We collapse the $I \times J$ table into multiple 3×3 tables. For a fixed jth drug, we have I such tables (**Table 2**), each associated with an AR ($i=1,...,I$). We assume that $n_{ij} \sim$ Poisson($n_i \times p_{ij}$), where p_{ij} is the reporting rate of jth drug for ith AE; and $n_j - n_{ij} \sim$ Poisson($n_j - n_i)q_{ij}$) where q_{ij} is the reporting rate of jth drug for other AEs combined excluding ith AR. We also assume n_{ij} and $n_j - n_{ij}$ are independent. Since drug j is fixed, unless stated otherwise, we suppressed the notational dependence of p_{ij} and q_{ij} on jth drug. We define the null hypothesis,

H_{0j}: $p_{ij}=q_{ij}=p_0$ for all ARs in drug j(ie.H_{0j}:$\cap_i p_{ij}=q_{ij}=p_0$);

and the alternative hypothesis (one sided) as

H_{aj}: $p_{ij}>q_{ij}$ for at least one AR in drug j(ie.H_{aj}:$\cup_i p_{ij}>q_{ij}$).

Under the null hypothesis, the maximum likelihood estimates (MLEs) for $p_0 = \dfrac{n_{.j}}{n_{..}}$ and the expected number of cases for ith AR and jth drug is $E_{ij} = n_{i.} \times \dfrac{n_{.j}}{n_{..}}$. Under the two-sided alterna-

Table 1. Data Structure.

	$Drug_1$	$Drug_2$	. . .	. . .	$Drug_J$	$n_{i.}$
AR_1	n_{11}	n_{12}	. . .	. . .	n_{1J}	$n_{1.}$
AR_2	n_{21}	n_{22}	. . .	. . .	n_{2J}	$n_{2.}$
. . .	. . .	. . .	. . .	. . .	. . .	. . .
. . .	. . .	. . .	. . .	. . .	. . .	. . .
AR_I	n_{I1}	n_{I2}	. . .	. . .	n_{IJ}	$n_{I.}$
$n_{.j}$	$n_{.1}$	$n_{.2}$	. . .	. . .	$n_{.J}$	$n_{..}$

Table 2. 3×3 tables for jth drug.

	jth Drug	Other Drugs	Marginal total
ith AR	n_{ij}	$n_{i.} - n_{ij}$	$n_{i.}$
other ARs	$n_{.j} - n_{ij}$	$(n_{..} - n_{i.}) - (n_{.j} - n_{ij})$	$n_{..} - n_{i.}$
Marginal Total	$n_{.j}$	$n_{..} - n_{.j}$	$n_{..}$

tive hypothesis ($p_{ij} \neq q_{ij}$), the maximum likelihood estimates (MLEs) for p_{ij} and q_{ij} are $\hat{p}_{ij} = \dfrac{n_{ij}}{n_{i.}} and \hat{q}_{ij} = \dfrac{n_{.j} - n_{ij}}{n_{..} - n_{i.}}$.

The maximum likelihoods under both the null and the two-sided alternative hypotheses are obtained by replacing the parameters with their MLEs in the likelihood functions, leading to the likelihood ratio, for ith AR and jth drug as,

$$LR_{ij} = \frac{L_a(\hat{p}_{ij}, \hat{q}_{ij})}{L_0(\hat{p}_0)} = \frac{(\frac{n_{ij}}{n_{i.}})^{n_{ij}} (\frac{n_{.j} - n_{ij}}{n_{..} - n_{i.}})^{(n_{.j} - n_{ij})}}{(\frac{n_{.j}}{n_{..}})^{n_{.j}}} = (\frac{n_{ij}}{E_{ij}})^{n_{ij}} (\frac{n_{.j} - n_{ij}}{n_{.j} - E_{ij}})^{n_{.j} - n_{ij}} \quad (1)$$

where $E_{ij} = \dfrac{n_{i.} \times n_{.j}}{n_{..}}$ is the expected number of counts for ith

AR and jth drug. The likelihood ratio test statistic for testing H_{0j}: $p_i = q_i$ for all ARs in drug j vs. H_{aj}: $p_i > q_i$ for at least one AR, is the maximum likelihood ratio $MLR_j = max_i (LR_{ij})$, where the maximum is over ARs $i=1,...,I$. We calculate

$$LR_{ij} = (\frac{n_{ij}}{E_{ij}})^{n_{ij}} (\frac{n_{.j} - n_{ij}}{n_{.j} - E_{ij}})^{n_{.j} - n_{ij}} I(\hat{p}_{ij} > \hat{q}_{ij})$$

and define

$MLR_i = max(LR_{ij})$ as the test statistic. For computational convenience, we may sometimes work with the log-likelihood ratio $log(LR_{ij})$ which is the monotone function of LR_{ij}.

The distribution of MLR under H_0 is not analytically tractable and is obtained using Monte Carlo simulation as defined below. First, the number of cases for each AR, for a given drug j, are simulated under H_0. Under H_0, since $n_{1j},...,n_{Ij}$, given the margin totals $n_{1.},...,n_{I.}$ are independent Poisson($n_i \cdot p_0$), $i=1,...,I$, the joint distribution of ($n_{1j},...,n_{Ij}$) conditioning on $n_{.j}$ and ($n_{1.},...,n_{I.}$) is

$$(n_{1j}, n_{2j}, ..., n_{Ij}) | n_{.j}; n_{1.},...,n_{I.} \sim multinomial(n_{.j}, (\frac{n_{1.}}{n_{..}},...,\frac{n_{I.}}{n_{..}})).$$

A total of 4999 datasets under H_0 are simulated from the multinomial distribution, and 500 MLRs are calculated (using 999 from the simulated datasets and one from the observed dataset). The null hypothesis is rejected at the α=0.05 level if the value of MLR from the observed dataset is greater than the 95th percentile of the 500 MLR values (threshold, T_a). The corresponding p-value is then 1-R/500, where R is the rank of the observed MLR among all the 500 MLR values. If the p-value of the observed MLR is less than $α$ (say, 0.05), then the AR associated with this MLR is the strongest signal among all ARs for the jth drug under consideration. Having found the strongest signal, we can then move to the second largest LR_{ij}, and so on, and declare them as signals if their LR_{ij} are greater than Tα or the corresponding p-values are less than α.

The likelihood ratio test is shown, analytically and through extensive simulation study, to control type-I error and false discovery rate (FDR) while retaining good power and sensitivity [**4,5**]. In the next section, we generalize the likelihood ratio test procedure to detect all AR signals in a drug class. The methods to detect drug signals for a set of prespecified ARs can be performed in a similar fashion.

3.2 Test Procedure for Adverse Reaction Detection of Multiple Drugs

In order to develop a test statistic that can identify adverse reaction signals in a drug class including multiple drugs, we assume that a drug class has k different drugs (usually k is a small number), and we assume that for kth drug the number of reports for ith AR and all other ARs (excluding ith AR) still remains a Poission distribution:

$$n_{ik} \sim Poisson(n_{i.} p_{ik})$$

$$n_{.k} - n_{ik} \sim Poisson((n_{..} - n_{i.})q_{ik})$$

where p_{ik} is the reporting rate of kth drug for ith AR, and q_{ik} is the reporting rate of kth drug for the other AEs. The null and alternative hypotheses for detecting AE signal in drug k are, H_{0k}: $p_{ik} = q_{ik} = p_{0k}$ for all AEs in drug k versus H_{ak}: $p_{ik} > q_{ik}$ for at least one AE.

The null and alternative hypotheses for detecting AE signals among this drug class with drugs are,

$$H_0: \cap_k \cap_i p_{ik} = q_{ik} = p_{0k}$$

where $i=1,\dots,I$ and $k=1,\dots,K$, versus

$$H_a: \cup_k \cup_i p_{ik} > q_{ik}$$

Another way to write the null and alternative hypotheses are $H_0: \cap_k H_{0k}$ versus $H_a: \cap_k H_{ak}$. Under the null hypothesis H_0, the MLE for p_{0k} is $\hat{p}_{0k} = \dfrac{n_{.k}}{n_{..}}$ and the expected number of counts for ith AE and th drug is $E_{ik} = n_{i.} \times \dfrac{n_{.k}}{n_{..}}$.   Under the two-

sided alternative hypothesis ($\cup_k \cup_i p_{ik} \neq q_{ik}$), the MLEs for p_{ik} and q_{ik} are $\hat{p}_{ik} = \dfrac{n_{ik}}{n_{i.}}$ and $\hat{q}_{ik} = \dfrac{n_{.k} - n_{ik}}{n_{..} - n_{i.}}$.

The maximum likelihood under both the null and the two-sided alternative hypotheses are obtained by replacing the parameters with their MLEs in the likelihood functions, leading to the likelihood ratio for ith AR in kth drug as:

$$LR_{ik} = \frac{L_a(\hat{p}_i, \hat{q}_i)}{L_0(\hat{p}_0)} = \frac{(\frac{n_{ik}}{n_{i.}})^{n_{ik}} (\frac{n_{.k} - n_{ik}}{n_{..} - n_{i.}})^{(n_{.k} - n_{ik})}}{(\frac{n_{.k}}{n_{..}})^{n_{.k}}} = (\frac{n_{ik}}{E_{ik}})^{n_{ik}} (\frac{n_{.k} - n_{ik}}{n_{.k} - E_{ik}})^{n_{.k} - n_{ik}} \quad (2)$$

The likelihood ratio test statistic for testing H_{0k}: $p_{ik} = q_{ik} = p_{0k}$, versus, H_{ak}: $p_{ik} > q_{ik}$ for at least one AR, is the maximum likelihood ratio $MLR_k = max_i (LR_{ik})$, $i=1,\dots,I$. The test statistic for testing $H_0: \cap_k \cap_i p_{ik} = q_{ik} = p_{0k}$ versus $H_a: \cup_k \cup_i p_{ik} > q_{ik}$ is

$$MLR = max_k(MLR_k) = max_k(max_i(LR_{ik} I(p_{ik} > q_{ik}))),$$

where $i=1,\dots,I$ and $k=1,\dots,K$.

Because the distribution of MLR under H_0 is not analytically tractable, we still use a Monte Carlo simulation to obtain its distribution. For each drug k in the drug class under H_0 we generate 500 datasets using

$$(n_{1k}, n_{2k}, \dots, n_{Ik}) | n_{.k}; n_{i.}, \dots, n_{I.} \sim multinomial(n_{.k}, (\frac{n_{1.}}{n_{..}}, \dots, \frac{n_{I.}}{n_{..}})) \quad (3)$$

and compute 500 values of MLR including the one from the real data, for $k=1,\dots,K$. This results into *500×K* MLR values. The nulll hypothesis is rejected at *α=0.05* level if the value of MLR from the observed dataset is greater than the (*1-α*)th percentile of the *500×K* MLR values *Ta*. After AR associated with the largest LR_{ik} is identified as signal $LR_{ik} > Ta$, we move to the AR with the second largest value of LR_{ik}, determine if it is a signal and so on. This way, the generalized likelihood ratio test procedure controls Type-I error. It also controls the false discovery rate (FDR) with *FDR ≤ α*.

4 Applications

In the following, we present the results from applying the likelihood ratio test procedure discussed in Section 3 to the "Monoamine oxidase inhibitors" (MAOIs). The MAOIs are used to treat several conditions. They include, but are not limited to: depression, generalized anxiety disorder, agitation, obsessive compulsive disorders (OCD), manic-depressive disorders, childhood enuresis (bedwetting), major depressive disorder, diabetic peripheral neuropathic pain, neuropathic pain, social anxiety disorder, posttraumatic stress disorder (PTSD) etc. The drug class includes Nardil (phenelzine), Parnate (tranylcypromine), Marplan (isocarboxazid), Emsam (selegiline), etc. We select four MAOIs labeled as MAOI1, MAOI2, MAOI3, MAOI4 and MAOI5 (not in any specific order to mask their names) using the WHO 2000--2005 and 2005-2010 data set.

The purpose of this analysis is to identify the ARs signals (with high disportionality rates) associated with MAOIs drug class. We apply the likelihood ratio test (LRT) and generalized likelihood ratio test (GLRT) for detecting Adverse Reactions.

The results of MAOIs drug class using both the LRT and GLRT are listed in **Table 3**. By using the likelihood ratio test procedure to each of the four drugs in the drug class, there are 66, 37, 74, 45 ARs detected for the four MAOIs drugs; while using the generalized likelihood ratio test, there are 61, 32, 68, 39 ARs detected, respectively. Across the four drugs, the GLRT detects less ARs than the LRT. By cross-checking the ARs in the four MAOIs drugs, there are 23 common ARs detected within this drug class. The top ARs are listed in **Tables 4** and **5**, and postural hypotension, high blood pressure, fainting, abnormal heart rhythm, dizziness, headache, drowsiness are the most strong ARs for this MAOIs class.

Table 3. Number of signals detected by the LRT and GLRT in Monoamine oxidase inhibitors (MAOIs) drug class.

Drug	2000-2005		2005-2010	
	LRT	GLRT	LRT	GLRT
MAOI1	66	61	95	88
MAOI2	37	32	52	45
MAOI3	74	68	90	89
MAOI4	45	39	73	68

5 A Simulation Study
5.1 Data Simulation

We then study the performance of the generalized likelihood ratio test (GLRT) using simulated datasets. We simulate datasets based on the four drugs in the monoamine oxidase inhibitors drug class in WHO's global pharmacovigilance database.

Under the null hypothesis, the data are simulated from multinomial distribution (3). Under the alternative hypothesis, data are generated as follow,

$$n_{1k},\ldots,n_{Ik} \mid n_{.k};n_{1.},\ldots,n_{I.} \sim multinomial(n_{.k},(rr_{1k} \times r_{0k} \times \frac{n_{1.}}{n_{..}}),\ldots,(rr_{Ik} \times r_{0k} \times \frac{n_{I.}}{n_{..}}))$$

(4)

where $k=1,\ldots,5$, and $rr_{1k},\ldots,rr_{Ik}$ are the relative reporting rates for $AE_1,\ldots,AE_I$ in k drugs with constraints $0 \le rr_{ik} * r_{0k} * \frac{n_{i.}}{n_{..}}$ and .

$$\sum_{i=1}^{I} rr_{ik} * rr_{0k} * \frac{n_{i.}}{n_{..}} = 1$$

Relative reporting rates rr_{ik} are specified as follows: rr_{ik} are assigned a value; higher than 1 for ARs selected as signals and 1 for all other ARs not selected as signals. r_{0k} can be regarded as baseline risks for drug k, and r_{0k} can be different from one drug to another.

Table 4. Total common reactions found for the five drugs in MAOIs class in 2000-2005 data set.

Drug	Common signals
MAOI1	postural hypotension, fainting, abnormal heart rhythm, dizziness, headache, drowsiness, insomnia, constipation, nausea, diarrhea, sexual dysfunction, weight gain or weight loss, seizures, rash, blurred vision, hepatitis
MAOI2	high blood pressure, fainting, abnormal heart rhythm, dizziness, headache, drowsiness, insomnia, anxiety, nausea, diarrhea, sexual dysfunction, weight gain or weight loss, blurred vision, hepatitis
MAOI3	postural hypotension, dizziness, headache, drowsiness, insomnia, anxiety, constipation, nausea, diarrhea, sexual dysfunction, weight gain or weight loss, seizures, rash, blurred vision, hepatitis
MAOI4	fainting, headache, drowsiness, postural hypotension, high blood pressure, insomnia, anxiety, constipation, nausea, diarrhea, sexual dysfunction, weight gain or weight loss, hepatitis

Table 5. Total common reactions found for the five drugs in MAOIs class in 2005-2010 data set.

Drug	Common signals
MAOI1	postural hypotension, dizziness, fainting, abnormal heart rhythm, headache, drowsiness, insomnia, constipation, nausea, diarrhea, sexual dysfunction, weight gain or weight loss, seizures, rash, blurred vision, hepatitis
MAOI2	abnormal heart rhythm, dizziness, high blood pressure, fainting, headache, drowsiness, insomnia, anxiety, nausea, diarrhea, sexual dysfunction, weight gain or weight loss, blurred vision, hepatitis
MAOI3	headache, dizziness, postural hypotension, drowsiness, insomnia, anxiety, diarrhea, sexual dysfunction, constipation, nausea, weight gain or weight loss, seizures, rash, blurred vision, hepatitis
MAOI4	fainting, drowsiness, postural hypotension, headache, high blood pressure, insomnia, nausea, anxiety, constipation, diarrhea, sexual dysfunction, weight gain or weight loss, hepatitis

We evaluate how the relative reporting rate (*rr*), the sample size ($n_{.k}$) and the number of signals affect the performance of the GLRT through the following four scenarios:

• Scenario 1: one signal is randomly assigned to one drug, and the remaining of other four drugs are free of signals. Without loss of generality, we assign one signal to the drug with the column total as 12000.

• Scenario 2: we randomly assign 30 common signals in each drug over the drug class with homogeneous relative reporting rate.

• Senario 3: we randomly assign 30 signals in each drug using homogeneous relative reporting rates (*rr*) across the drug class, but signals are not necessarily common between drugs.

• Senario 4: we take a similar process as Scenario 3, randomly select 30 signals for each drug independently, but we use inhomogeneous *rr*. A rate of *2×rr* is assigned to those AR signals for which n_i (the total number of reports for the ARs) fall between 35,000 and 40,000, a rate of *3×rr* to those AR signals for which n_i fall between 20,000 and 25,000, a rate of *4×rr* to those AR signals for which fall between 15,000 and 20,000, and a rate of *5×rr* to those AR signals for which n_i fall between 6,000 and 12,000. *rr* is assigned to 1 for those ARs that are not selected as signals.

In each simulation, we generate 1,000 datasets.

5.2 Performance Characteristics Evaluation

The performance of the proposed methods is evaluated by using Power, sensitivity (ST) and false discovery rate (FDR). First, power is defined as:

$$Power = \frac{\text{Number of times reject } H_0}{L},$$

where *L=1,000* is the total number of simulations. H_0 will be rejected when at least one AR in any one drug (in the drug class) is signal.

The sensitivity of a test is the proportion of positive results that are correctly identified. In our case, sensitivity is defined as:

$$ST = \frac{1}{L}\sum_I \frac{\text{Number of reactions correctly detected in ith simulated data}}{\text{Total number of true reactions in the ith simulated data}}$$

The definition of FDR can be illustrated by a *2×2* table as in **Table 6**, where V is the number of falsely deteced signals, S is the number of correctly detected signals. FDR is defined as

$$E(\frac{V}{V+S})$$, the expected proportion of rejected null

hypotheses which are erroneously rejected. It's estimated by,

$$FDR = \frac{1}{L}\sum_I \frac{\text{Number of reactions falsely detected in ith simulated data(V)}}{\text{Total number of reactions detected in the ith simulated data(V+S)}}.$$

Table 6. Illustration of false discovery rate (FDR).

	H_0 is true	H_a
Positive result	V	S
Negative result	U	T

All power, ST and FDR have values between 0 and 1. As we shall see in next section, GLRT high sensitivity, low FDR, and to control Type-I error α. which indicates its superiority over the conventional likehood ratio test.

5.3 Simulation Results

The simulation results shown in **Table 7** include power, sensitivity, and false discovery rate for the different scenarios described in Section 5.1.

Table 7. Simulation results for the GLRT.

marginal count	$n_{.j}$	rr	Power	Sensitivity	FDR
Scenario 1: Select one signal in one drug with marginal total as 28,216 or 4,362					
28216	500	3	0.073	0.13	0.0565
28216	1500	3	0.315	0.278	0.0442
28216	3000	3	0.83	0.753	0.044
28216	5000	3	0.992	0.983	0.0368
28216	8000	3	1	1	0.0257
28216	12000	1	0.061	NA	0.053
28216	12000	2	0.758	0.751	0.043
28216	12000	3	1	1	0.0351
28216	12000	5	1	1	0.0326
28216	12000	7	1	1	0.0316
4362	12000	1	0.077	NA	0.077
4362	12000	2	0.084	0.005	0.090
4362	12000	3	0.152	0.079	0.073
4362	12000	4	0.50	0.364	0.0569
4362	12000	5	1	1	0.0393
4362	12000	7	1	1	0.0397
Scenario 2: Randomly select 30 common signals with homogenerous rr					
NA	AS	1	0.062	NA	0.064
NA	AS	2	0.744	0.01207	0.0381
NA	AS	3	1	0.1347	0.0057
NA	AS	5	1	0.4794	0.0014
NA	AS	7	1	0.7315	0.0011
NA	AS	9	1	0.8520	0.0009
Scenario 3: Randomly select 30 signals (not necessarily common) with homogenerous rr					
NA	AS	1	0.085	NA	0.081
NA	AS	2	0.80	0.0163	0.0293
NA	AS	3	1	0.1426	0.0050
NA	AS	5	1	0.4450	0.0018
NA	AS	7	1	0.6829	0.0013
NA	AS	9	1	0.8018	0.0010
Scenario 4: Randomly select 30 signals with inhomogenerous rr					
NA	AS	1	0.056	NA	0
NA	AS	3	1	0.93912	0.0009
NA	AS	5	1	0.98467	0.006
NA	AS	7	1	0.99887	0.018

AS: actual sample size; NA: not applicable.

In Scenario 1, one signal was assigned to the ARs with the relative large or moderate marginal counts (28,216 and 4362). With fixed *rr=3*, n_i=28216 and sample size n_j=500, the power is 0.073, ST is 0.13 and FDR is 0.0565. As the sample size n_j increases, the power and ST increase to 1, and FDR decreases from 0.06 to 0.03. When the sample size of AR is fixed at n_i=28216, with the increase of *rr* from 1 to 7, the power increases from 0.06 to 0.75, and then to 1. The same increasing trend is also observed for ST. FDR decreases from 0.05 to 0.03, a value much lower than the level of significance. The effect of sample size is also evaluated when n_i is fixed at 4362. The trends remain similar for the Power, ST and FDR, though the change in trends is relatively slower.

In Scenario 2 where 30 common signals are assigned to all the five drugs in the drug class, when *rr=1*, the power and FDR are both 0.06. As *rr* increases, the power increases to 1, the ST increases from 0.01 to 0.85, and the FDR decreases from 0.064 to 0.0009. Because multiple signals are assigned randomly, we use actual sample size (AS) for n_j. Similar trends of power, sensitivity and FDR are also found for Scenarios 3 and 4.

Besides the effect of relative reporting rate (*rr*), the effect of number of selected true signals on the performance of GLRT is also studied. In Scenario 2, if the number of signals are changed to 10 and 20, similar trends are observed for the power, sensitivity and FDR, as in **Table 8**. As *rr* increases, both the power and sensitivity increase, and the FDR decreases.

Table 8. The effect of the number of true signals on the performance of the GLRT.

marginal count	n_j	rr	Power	Sensitivity	FDR
Scenario 2.1: Randomly select 10 common signals with homogenerous rr					
NA	AS	1	0.071	NA	0.068
NA	AS	2	0.246	0.076	0.059
NA	AS	3	0.996	0.14081	0.021
NA	AS	5	1	0.45008	0.006
NA	AS	7	1	0.69561	0.0030
NA	AS	9	1	0.80329	0.00265
Scenario 2.2: Randomly select 20 common signals with homogenerous rr					
NA	AS	1	0.059	NA	0.058
NA	AS	2	0.535	0.01639	0.005
NA	AS	3	1	0.1601	0.0082
NA	AS	5	1	0.5631	0.0029
NA	AS	7	1	0.72354	0.00145
NA	AS	9	1	0.85018	0.00137
Scenario 2.3: Randomly select 30 common signals with homogenerous rr					
NA	AS	1	0.063	NA	0.063
NA	AS	2	0.837	0.01416	0.0449
NA	AS	3	1	0.1629	0.0052
NA	AS	5	1	0.5048	0.0012
NA	AS	7	1	0.7393	0.0009
NA	AS	9	1	0.9461	0.0008

AS: actual sample size; NA: not applicable.

If rr is fixed, as the number of selected signals increases, the power increases but the FDR decreases.

6 Discussion and Concluding Remarks

In this paper we generalized the likelihood ratio test procedure to detect adverse event for a class of drugs and applied it to the World Health Organization's global pharmacovigilance database. The proposed methods can also be used to detect drug adverse reactions in a group of pre-specified adverse reactions by renaming the row and column variables. One of the advantages of the generalized likelihood ratio test presented here is that the methods can be used to find multiple adverse reactions with both the Type-I error and false discovery rates controlled while retaining good power and sensitivity. We note that the GLRT tends to detect less adverse reactions than the LRT method. This is to be expected, since the threshold in the GLRT of the drug class is greater than or equal to those from each individual drug using the LRT, thus it is more conservative.

The generalized likelihood ratio test procedure provides a useful tool to identify potential adverse reactions in pharmacovigilance database. However, the final discovery of the true adverse reactions should also be based on a thorough review of all available medical records.

Competing interests
The author declares that he has no competing interests.

Publication history

References

1. Bate A, Lindquist M, Edwards IR, Olsson S, Orre R, Lansner A and De Freitas RM. **A Bayesian neural network method for adverse drug reaction signal generation**. *Eur J Clin Pharmacol*. 1998; **54**:315-21. | Article | PubMed

2. DuMouchel W. **Bayesian Data Mining in Large Frequency Tables, With an Application to the FDA Spontaneous Reporting System**. *The American Statistician*. 1999; **53**:177-190. | Article

3. Evans SJ, Waller PC and Davis S. **Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports**. *Pharmacoepidemiol Drug Saf*. 2001; **10**:483-6. | Article | PubMed

4. Baah EM. (2014), Analysis of data on spontaneous reports of adverse events associated with drugs. PhD thesis, University of Glasgow, UK.

5. Bate A and Evans SJW. (2009), Quantitative signal detection using spontaneous ADR reporting. Pharmacoepidemiol Drug Saf, 18, 427–436.

6. Noren GN, Bate A, Orre R and 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**:3740-57. | Article | PubMed

7. Rothman KJ, Lanes S and Sacks ST. **The reporting odds ratio and its advantages over the proportional reporting ratio**. *Pharmacoepidemiol Drug Saf*. 2004; **13**:519-23. | Article | PubMed

8. Szarfman A, Machado SG and O'Neill RT. **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*. 2002; **25**:381-92. | Article | PubMed