# Mining Adverse Drug Reactions from Electronic Health Records

Henry Z. Lo, Wei Ding
Department of Computer Science
University of Massachusetts Boston
Boston, Massachusetts 02125–3393
Email: {henryzlo, ding}@cs.umb.edu

Zohreh Nazeri
MITRE Corporation
McLean, Virginia 22102
Email: nazeri@mitre.org

***Abstract*—Over 2 million serious side effects, including 100,000 deaths, occur due to adverse drug reactions (ADR) every year in the US. Though various NGOs monitor ADRs through self reporting systems, earlier detection can be achieved using patient electronic health record (EHR) data available at many medical facilities. This paper presents an algorithm which allow existing ADR detection methods, which were developed for spontaneous reporting systems, to be applied directly to the longitudinal EHR data, as well as a new ADR detection method specifically for this type of data. Preliminary results show that the new method outperforms existing methods on EHR datasets. Future work on the method will extend it to detecting potential cause-effect relationships between events in other types of longitudinal data, handling multiple cause and effect items, and automatically selecting surveillance windows.**

## I. Introduction

In the U.S., adverse drug reactions (ADR) send 700,000 patients to emergency rooms and hospitalize 120,000 people yearly [3]. Organizations such as the FDA and the WHO have developed spontaneous reporting systems (SRS) to monitor ADRs in the general population. Yet, these systems suffer from underreporting and insufficient information [1], and death rates due to ADRs continue to rise [11].

To catch ADRs as soon as possible, it is necessary to monitor ADRs in healthcare institutions, rather than rely on spontaneous reports. Hospitals often already contain databases of patient electronic health records (EHR), which contain longitudinal patient-level information about drugs administered and conditions experienced. Yet despite the potential benefit of mining ADRs from EHR data, methods have only been applied to SRS data [2], [4], [5], [9], [12]. All these methods require counting the co-occurrence of drugs and conditions, something that is easily done in spontaneous reporting data, but not directly applicable to EHR data.

Co-occurrence counts are necessary in calculating proportionality ratios, which form the basis of all ADR mining methods. In SRS data, if a drug and condition are both present in the same report, then a co-occurrence is noted. However, in EHR, linkages between drugs and conditions must be established using only timestamp data. This paper presents a method to handle this problem.

The second contribution in this paper is a new technique for mining ADR methods. This method, which is based on comparing the rates of condition occurrence, do not suffer from the weaknesses of disproportionality methods used in SRS data when applied in the EHR context. This method specifically applies to events occuring over time as in EHR datasets, but also generalizes to any problem related to extracting temporal associations from time-stamped data.

| | Condition | No condition |
|---|---|---|
| Drug | $a$ | $b$ |
| No drug | $c$ | $d$ |

TABLE I. Contingency table. $a$ refers to the number of subjects who have taken the drug and experienced the condition, and so on.

These contributions leverage the advances in mining SRS data to EHR data, and also provide a new method specifically for this task. Together, this forms a basis of comparison for future ADR mining work from EHRs. The long-term goal of this project is to further improve EHR data mining techniques, in the hopes that adoption will catch ADRs earlier and more reliably and limit the potential damage they cause.

## II. ADR Detection Methods

Most accepted methods use disproportionality measures, which quantify how often a drug and condition co-occur compared to some baseline. These measures are ratios or combinations of the four count values present in a contingency table, as shown in Table I. The variable names $a$,$b$,$c$, and $d$ used in that table will be used to define each of the disproportionality measures.

All disproportionality methods have only been applied to self-reporting data. The Poisson method proposed later in this paper is especially designed for longitudinal data, by considering rates of occurrence, using time as a denominator.

### A. Proportional reporting ratio

The proportional reporting ratio (PRR), as defined in Equation 1, has been used in the UK Medicines Control Agency (MCA) and is widely adopted. [5]. A larger PRR signifies a potential effect of a drug on the condition.

$$PRR = \frac{a/(a+b)}{c/(c+d)} \tag{1}$$

| | |
|---|---|
| $x_1$ | Occurrence count of given condition in the presence of a given drug. |
| $x_2$ | Occurrence count of given condition in the absence of a given drug. |
| $n_1$ | Number of individuals who have taken the given drug. |
| $n_2$ | Number of individuals who have not taken the given drug. |
| $\lambda_1$ | $x_1/n_1$, the condition's rate of occurrence with drug exposure. |
| $\lambda_2$ | $x_2/n_2$, the condition's rate of occurrence without drug exposure. |

TABLE II. NOTATION FOR POISSON FORMULAS.

PRRs greater than two, along with a count of over three, and a $\chi^2$ value over four signify a potential ADR [5]. The $\chi^2$ test is performed on the contingency table.

### B. Reporting odds ratio

The reporting odds ratio (ROR), shown in Equation 2, currently being used by the Netherlands Pharmacovigilance Centre to detect ADRs, is also widely deployed [7].

$$ROR = \frac{a/b}{c/d} \tag{2}$$

The drug-condition pair is signaled if the lower limit of the two-sided 95% confidence interval exceeds 1 [7], [10]. This confidence interval can be calculated using Equation 3.

$$e^{\ln(ROR) \pm 1.96\sqrt{(\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d})}} \tag{3}$$

## III. CONTRIBUTIONS

This section tackles two problems associated with extracting ADRs: first, a method to detect ADRs based on comparing rates of condition occurrence; and second, an efficient algorithm for extracting contingency counts ($a, b, c, d$ in Table I) to use disproportionality methods on EHR data.

### A. Poisson method

The proposed method compares rates of condition occurrence, as with disproportionality methods, but uses time as a denominator. This avoids using condition non-occurrence, which is required in disproportionality methods cannot truly be counted in longitudinal data.

Comparing (within some time window) the rate of condition occurrence in drug takers $\lambda_1$ and the rate in non-drug takers $\lambda_2$, we may be able to determine whether or not a drug has had a significant effect on the condition. Assuming these rates occur according to the Poisson distribution, we can detect an ADR a by detecting a significant difference between the two. This can be done using either the popular C-test [8] or the newer and more powerful E-test [6]. As discovering ADRs is more important than being accurate, we set the $\alpha$ value of both tests at $0.1$, and use two-sided tests.

For the following sections, notation is given in Table II.

*1) C-test:* The C-test compares the two condition occurrence rates, $\lambda_1$ and $\lambda_2$, using the fact that the joint distribution of two counts $x_1, x_2$ follow the binomial distribution when conditioning on $\rho = \frac{\lambda_1}{\lambda_1+\lambda_2}$ and $\mu = x_1 + x_2$. The binomial function is then used to calculate the $p$-value for the two-sided hypothesis test $\lambda_1 = \lambda_2$. Derivation is shown in the original paper [8].

$$p(x_1, x_2|\rho, \mu) = \frac{\mu^m e^{-\mu}}{m!}\binom{m}{x_1}\rho^{x_1}(1-\rho)^{m-x_1} \tag{4}$$

For our two sided $C$-test, with the null hypothesis $\lambda_1 = \lambda_2$, the $p$-value is given by the formula:

$$2 * \min(P(X_1 \geq x_1|m, \rho), P(X_1 \leq x_1|m, \rho)) \tag{5}$$

where

$$P(X_1 \leq x_1|m, \rho) = \sum_{i=x_1}^{m} p(i|m, \rho)$$

and we reject the hypothesis that the two rates are equal (i.e. the rates of condition occurrence differs significantly) when Equation 5 yields a value less than $\alpha = 0.1$.

*2) E-test:* The E-test uses the standardized difference as a pivot statistic. If the null hypothesis is that the two observed rates $x_1/n_1$ and $x_2/n_2$ are equal, then the standardized difference

$$T_{x_1,x_2} = \frac{x_1/n_1 - x_2/n_2}{\sqrt{\hat{V}_k}} \tag{6}$$

should be 0. In this formula, $\hat{V}_k$ is estimated variance, given by

$$\hat{V}_k = \frac{x_1/n_1}{n_1} + \frac{x_2/n_2}{n_2}$$

The $p$-value can be calculated using the following formula, which is derived in the original paper [6]:

$$\sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty}\frac{(n_1\hat{\lambda}_2)^{k_1}e^{-n_1\hat{\lambda}_2}}{k_1!}\frac{(n_2\hat{\lambda}_2)^{k_2}e^{-n_2\hat{\lambda}_2}}{k_2!}[I(T_{k_1,k_2} = T_{x_1,x_2})] \tag{7}$$

where $I$ is the indicator function.

### B. Count extraction algorithm

We separate the calculation of disproportionality measures or pivot statistics from extracting the counts needed for these methods. This saves processing time, as extracting counts is costly for large datasets, but calculating measures, though potentially difficult, do not depend directly on the size of the dataset.

The count extraction algorithm described in Algorithm 1 utilizes the schema in Table IV. It returns patient drug exposure information in the form of $P_d$, from which $a + b$ can be extracted; $N_{+,}$, which counts condition occurrences for patients exposed to a given drug ($a$); and $N_{-,}$, which counts condition occurrences for unexposed patients ($c$). Given that we have the total number of people, all contingency counts in Table I can be inferred.

The Poisson method does not use the values in the contingency tables, but rather uses the count of occurrences of a condition given that a user has taken or not taken a drug. This can be calculated from $N_{+,}$ and $N_{-,}$, respectively.

In the algorithm, $D$ is the set of drugs; $P$, the set of persons; $X_C$ is the `CONDITION_OCCURRENCE` table; $X_D$

**Result**: $N_{+.}, N_{-.}, P_d$

1 $P_d \leftarrow$ map ;
2 **foreach** *Row* $x \in X_D$ **do**
3 | $P_d[x.person, x.drug] \leftarrow x.drug\_start\_date$ ;
4 **end**
5 $N_{+.} \leftarrow$ map ;
6 $N_{-.} \leftarrow |P|$ map ;
7 **foreach** *Drug* $d \in D$ **do**
8 | **foreach** *Row* $x \in X_C$ **do**
9 | | $s \leftarrow P_d[x.person, d]$ ;
10 | | **if** $s \neq NA$ **then**
11 | | | **if** $s \leq x.condition\_start\_date \leq s + t$ **then**
12 | | | | $N_{+.}[x.person, d, x.condition] + +$ ;
13 | | | **end**
14 | | **else**
15 | | | $s \leftarrow$ random time ;
16 | | | **if** $s \leq x.condition\_start\_date \leq s + t$ **then**
17 | | | | $N_{-.}[x.person, d, x.condition] + +$ ;
18 | | | **end**
19 | | **end**
20 | **end**
21 **end**

**Algorithm 1:** Count extraction program.

is the DRUG_EXPOSURE table; $t$ is the time window; and $s$ is a start time of a drug for a person.

An analysis of the runtime of Algorithm 1:

- Lines 1-4 populate $P_d$, which holds the date of first occurrence of a drug for a patient, or $NA$ if the patient never takes the drug. Runtime depends solely on $|X_D|$.
- Lines 5-6 create $N_{+.}$ and $N_{-.}$, sparse matrices which count condition occurrences.
- In lines 7-21, how many runs the for loop goes on for depends on $|D||X_C|$.
- In line 16, $s$ should be after the start of the subject's observation, but with at least $t$ before the end of the observation period.
- In total, the algorithm is $O(|X_D| + |D||X_C|)$.

### C. Extracting disproportionality ratios

Calculating the values in the contingency tables is necessary to derive ROR and PRR for each drug and condition. To get these values, convert every count value in $N_{-.}$ and $N_{+.}$ above two to one, then sum both matrices along the "person" dimension. Then for each drug and condition:

- $a$ is the respective cell in the summed $N_{+.}$ matrix.
- $c$ is the respective cell in the summed $N_{-.}$ matrix.
- $b = m - a$, where $m$ is the number of people who took $d$, according to $P_d$.
- $d = n - (a + b + c)$, where $n$ is the total number of people.

The Poisson method does not use contingency values, but rather uses the actual count values in $N_{-.}$ and $N_{+.}$, without converting nonzero counts to one.

| Dataset | Drugs | Conditions | Persons |
|---|---|---|---|
| 1 | 12 | 10 | 10000 |
| 2 | 25 | 20 | 10000 |
| 3 | 50 | 40 | 10000 |
| 4 | 12 | 10 | 10000 |
| 5 | 25 | 20 | 10000 |
| 6 | 50 | 40 | 10000 |
| 7 | 12 | 10 | 10000 |
| 8 | 25 | 20 | 10000 |
| 9 | 50 | 40 | 10000 |

TABLE III. PARAMETERS USED FOR DATASETS GENERATED.

| Table | Columns |
|---|---|
| CONDITION_OCCURRENCE | PERSON_ID<br>CONDITION_ID<br>CONDITION_START_DATE |
| DRUG_EXPOSURE | PERSON_ID<br>DRUG_ID<br>DRUG_START_DATE<br>DRUG_END_DATE |
| OBSERVATION_PERIOD | PERSON_ID<br>OBSERVATION_START_TIME<br>OBSERVATION_END_TIME |
| DRUG_OUTCOMES | DRUG_ID<br>CONDITION_ID<br>DRUG_EFFECT_SIZE<br>DRUG_EFFECT_CATEGORY |

TABLE IV. A TABLE REPRESENTING THE USEFUL TABLES AND ATTRIBUTES OF THE OSIM OUTPUT SCHEMA.

## IV. EXPERIMENTS

### A. Data

EHR data was generated using the OMOP OSIM1 software. OSIM uses probability distributions of variables such as demographic information, personal characteristics, and drug and condition characteristics extracted from a real EHR database, and preserves these distributions for generated datasets, while allowing experimenters to scale and modify parameters at will.

Input parameters for the 9 generated datasets are shown in Table III. In addition to these, condition and drug prevalence were also increased (set to the 1-10% category), as the default settings resulted in extremely few condition / drug reactions. All other parameters remained at their default values.

The relevant tables generated from the OSIM process, and their schemas, are shown in Table IV. CONDITION_OCCURRENCE and DRUG_EXPOSURE consist of patient data; DRUG_OUTCOMES contains known ADRs, and serves as the ground truth.

### B. Experimental Design

Experiments compared ROR, PRR, and the proposed C-test and E-test methods. All 9 datasets were used.

The primary evaluation metric is recall, which measures the rate of false negatives, and is commonly used metric in ADR detection [10], [12]. The secondary metric is precision, which measures the rate of false positives.

The length of the surveillance period, within which a drug-condition pair may be considered to have co-occurred, was selected manually. 180 and 365-day time windows were used in these experiments.

| Dataset | Precision | | | | Recall | | | |
|---|---|---|---|---|---|---|---|---|
| | ROR | PRR | C-test | E-test | ROR | PRR | C-test | E-test |
| 1 | **0.033** | 0.022 | **0.033** | 0.029 | **1.000** | **1.000** | **1.000** | **1.000** |
| 2 | 0.038 | 0.034 | 0.039 | **0.042** | 0.733 | **0.867** | 0.733 | **0.867** |
| 3 | **0.024** | 0.023 | 0.023 | 0.023 | 0.737 | 0.842 | 0.737 | **0.868** |
| 4 | **0.033** | 0.022 | **0.033** | 0.029 | **1.000** | **1.000** | **1.000** | **1.000** |
| 5 | 0.038 | 0.035 | 0.039 | **0.042** | 0.733 | **0.867** | 0.733 | **0.867** |
| 6 | **0.024** | 0.023 | 0.023 | 0.023 | 0.737 | 0.842 | 0.737 | **0.868** |
| 7 | **0.033** | 0.022 | **0.033** | 0.029 | **1.000** | **1.000** | **1.000** | **1.000** |
| 8 | 0.039 | 0.034 | 0.039 | **0.042** | 0.733 | **0.867** | 0.733 | **0.867** |
| 9 | **0.024** | 0.023 | 0.023 | 0.023 | 0.737 | 0.842 | 0.737 | **0.868** |
| mean | **0.032** | 0.026 | **0.032** | 0.031 | 0.823 | 0.903 | 0.823 | **0.912** |

TABLE V. RECALL AND PRECISION RESULTS FOR EACH DATASET, USING A 180-DAY TIME WINDOW. BEST RESULTS FOR EACH DATASET ARE PRESENTED IN BOLD.

| Dataset | Precision | | | | Recall | | | |
|---|---|---|---|---|---|---|---|---|
| | ROR | PRR | C-test | E-test | ROR | PRR | C-test | E-test |
| 1 | 0.028 | 0.022 | **0.031** | 0.028 | **1.000** | **1.000** | **1.000** | **1.000** |
| 2 | **0.044** | 0.039 | 0.041 | 0.040 | **0.867** | **0.867** | **0.867** | **0.867** |
| 3 | **0.126** | 0.046 | 0.021 | 0.021 | 0.632 | 0.711 | 0.737 | **0.816** |
| 4 | 0.028 | 0.022 | **0.032** | 0.029 | **1.000** | **1.000** | **1.000** | **1.000** |
| 5 | **0.043** | 0.039 | 0.041 | 0.040 | **0.867** | **0.867** | **0.867** | **0.867** |
| 6 | **0.122** | 0.046 | 0.021 | 0.022 | 0.632 | 0.711 | 0.737 | **0.842** |
| 7 | **0.033** | 0.022 | **0.033** | 0.029 | **1.000** | **1.000** | **1.000** | **1.000** |
| 8 | 0.038 | 0.034 | 0.039 | **0.042** | 0.733 | **0.867** | 0.733 | **0.867** |
| 9 | **0.024** | 0.023 | 0.023 | 0.023 | 0.737 | 0.842 | 0.737 | **0.868** |
| mean | **0.054** | 0.032 | 0.031 | 0.030 | 0.830 | 0.874 | 0.853 | **0.903** |

TABLE VI. RECALL AND PRECISION RESULTS FOR EACH DATASET, USING A 365-DAY TIME WINDOW. BEST RESULTS FOR EACH DATASET ARE PRESENTED IN BOLD.

### C. Results

Results are shown in Tables V and VI.

The E-test method recognized the most true ADRs in every experimental condition, as evidenced by its high recall. On average, the PRR test had the second highest recall, followed by C-test and then ROR.

In the 180-day time window condition, the E-test precision was comparable to other methods, only negligibly lower than the C-test and ROR methods. In the 365-day condition, ROR had the highest precision by far, but also the lowest recall.

## V. DISCUSSION

### A. Future Experiments

Though results show the E-test to be superior in every experiment ran, more sophisticated methods such as MGPS and BCPNN were not extended to EHR data and tested against. This is because as Bayesian methods, they are sensitive to initial conditions and have many implementation details. In the future, the methods described here must be tested against the more state-of-the-art Bayesian methods.

### B. Future Applications

The Poisson method described here is applicable to any task involving detecting abnormal temporal co-occurrences, for example, finding potential causes and effects. To demonstrate the utility and flexibility of this technique, future experiments should be run in different domains and on different problems.

The framework described here only detects single-cause single-effect associations, but needs to be extended to handling multiple causes and effects. This presents considerable difficulty, as co-occurrence will then need to be defined for sets of drugs and conditions larger than two.

One possible method to make this easier would be to automatically identify a time window $t$. If the likelihood of a cause-effect association decreases as the gap between the two increases, then perhaps this definition could be extended to model the time gap between multiple causes and effects as well.

### C. Conclusion

This paper explores the usage of EHR data for ADR detection, tackling two issues: one, the establishment of a counting method, which is necessary to apply existing ADR mining methods to EHRs, and second, a native method for mining ADRs from longitudinal data, which outperforms two popular and currently used methods. Discussions show the benefits of working with EHR data, and it is hoped that future work will benefit actual ADR detection in the future.

## REFERENCES


[1] L. Aagaard, J. Strandell, L. Melskens, P. S. Petersen, and E. Holme Hansen. Global patterns of adverse drug reactions over a decade: analyses of spontaneous reports to VigiBase. *Drug Saf*, 35(12):1171–1182, Dec 2012.

[2] Andrew Bate. *The use of Bayesian confidence propagation neural network in pharmacovigilance*. PhD thesis, Ume University, Pharmacology and Clinical Neuroscience, 2003.

[3] Budnitz DS, Pollock DA, Weidenbach KN, Mendelsohn AB, Schroeder TJ, and Annest JL. National surveillance of emergency department visits for outpatient adverse drug events. *JAMA*, 296(15):1858–1866, 2006.

[4] William Dumouchel. Bayesian data mining in large frequency tables, with an application to the fda spontaneous reporting system. *The American Statistician*, 53(3):177–190, 1999.

[5] Stephen J. Evans, Patrick C. Waller, and S. Davis. Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiology and drug safety*, 10(6):483–486, 2001.

[6] K. Krishnamoorthy and Jessica Thomson. A more powerful test for comparing two poisson means. *Journal of Statistical Planning and Inference*, 119(1):23 – 35, 2004.

[7] Nicholas Moore, Carmen Kreft-Jais, Franoise Haramburu, Catherine Noblet, Michel Andrejak, Michel Ollagnier, and Bernard Bgaud. Reports of hypoglycaemia associated with the use of ACE inhibitors and other drugs: a case/non-case study in the French pharmacovigilance system database. *Br J Clin Pharmacol*, 44(5):513–518, Nov 1997.

[8] J. Przyborowski and H. Wilenski. Homogeneity of results in testing samples from poisson series: With an application to testing clover seed for dodder. *Biometrika*, 31(3-4):313–323, 1940.

[9] Kenneth J. Rothman, Stephan Lanes, and Susan T. Sacks. The reporting odds ratio and its advantages over the proportional reporting ratio. *Pharmacoepidemiology and Drug Safety*, 13(8):519–523, 2004.

[10] Toshiyuki Sakaeda, Akiko Tamon, Kaori Kadoyama, and Yasushi Okuno. Data mining of the public version of the FDA adverse event reporting system. *Int J Med Sci*, 10(7):796–803, 2013.

[11] G. Shepherd, P. Mohorn, K. Yacoub, and D. W. May. Adverse drug reaction deaths reported in United States vital statistics, 1999-2006. *Ann Pharmacother*, 46(2):169–175, Feb 2012.

[12] Eugne. P. van Puijenbroek, Andrew Bate, Hubert G. Leufkens, Marie Lindquist, Roland Orre, and Antoine C. Egberts. A comparison of measures of disproportionality for signal detection in spontaneous reporting systems for adverse drug reactions. *Pharmacoepidemiol Drug Saf*, 11(1):3–10, 2002.