

# A Comparative Study of Data Mining Algorithms used for Signal Detection in FDA AERS Database

Viswam Subeesh<sup>1,\*</sup>, Eswaran Maheswari<sup>2</sup>, Ganesan Rajalekshmi Saraswathy<sup>2</sup>, Ann Mary Swaroop<sup>3</sup>, Satya Sai Minnikanti<sup>3</sup>

<sup>1</sup>Research Scholar, Department of Pharmacy Practice, Faculty of Pharmacy, M.S Ramaiah University of Applied Sciences, Bengaluru, Karnataka, INDIA.

<sup>2</sup>Professor, Department of Pharmacy Practice, Faculty of Pharmacy, M.S Ramaiah University of Applied Sciences, Bengaluru, Karnataka, INDIA.

<sup>3</sup>Associate Professor, Department of Pharmacy Practice, Faculty of Pharmacy, M.S Ramaiah University of Applied Sciences, Bengaluru, Karnataka, INDIA.

## ABSTRACT

**Objective:** Signal detection is a technique in pharmacovigilance for the early detection of new, rare reactions (desired or undesired) of a drug. This study aims to compare and appraise the performance of data mining algorithms used in signal detection. **Method:** Most commonly used three data mining algorithms (DMAs) (Reporting Odds Ratio (ROR), Proportional Reporting Ratio (PRR) and Information Component (IC)) were selected and applied retrospectively in USFDA Adverse Event Reporting System database to detect five confirmed Drug Event Combinations. They were selected in such a way that the drug is withdrawn from the market or label change between 2006-2015. A value of  $ROR-1.96SE > 1$ ,  $PRR \geq 2$ ,  $\chi^2 > 4$  or  $IC-2SD > 0$  were considered as the positive signal. The data mining algorithms were compared for their sensitivity and early detection. **Result:** Among the three data mining algorithms, Information Component was found to have a maximum sensitivity (100%) followed by Reporting Odds Ratio (60%) and Proportional Reporting Ratio (40%). Sensitivity associated with the number of reports per drug event combination and early signal detection suggested that information component needs comparatively fewer reports to show positive signal than the other two data mining algorithms. ROR and PRR

showed comparable results. **Conclusion:** Early detection of a reaction is possible using signal detection technique. Information component was found to be sensitive method compared with other two data mining algorithms in FDA Adverse Event Reporting System database. As the number of reports of drug event combination increased, the sensitivity and comparability of data mining algorithm also increased.

**Key words:** Signal Detection, Data mining algorithms, FDA AERS Database, Disproportionality Analysis, Pharmacovigilance.

## Correspondence

**Mr. Subeesh K Viswam**, Research Scholar, Department of Pharmacy Practice, Faculty of Pharmacy, M.S Ramaiah University of Applied Sciences, Bengaluru, Karnataka, INDIA.

Phone: +91-9895843219

Email: subeeshkviswam@gmail.com

DOI: 10.5530/jyp.2018.10.97

## INTRODUCTION

The US Food and Drug Administration (FDA) has maintained the Adverse Events Reporting System (AERS) database since 1968 with the intention of continuous monitoring of drugs during the post-marketing surveillance.<sup>1</sup> AERS database is a collection of suspected Adverse Event (AEs) reports from pharmaceutical companies, consumers and health-care professionals.<sup>2</sup> To date, FDA AERS database contains more than 10 million AE reports and receive nearly half a million each year. Many approaches have been adopted into post-marketing studies, including prescription-event monitoring (PEM), spontaneous reporting, case review, observational study and drug utilization review.

Compared to clinical trials and traditional epidemiologic studies, the computer-assisted Data Mining Algorithms (DMAs) are relatively new and characterized by providing a fast and efficient way of detecting possible Adverse Drug Reactions (ADRs) signal.<sup>3</sup> Several DMAs have been well described in the literature, mainly including the Reporting Odds Ratio (ROR),<sup>4</sup> Multi-item Gamma Poisson Shrinker (MGPS),<sup>5</sup> the Proportional Reporting Ratio (PRR),<sup>6</sup> and the Information Component (IC).<sup>7</sup> By integrating such computer-aided techniques, incorporating statistical analyses and centralizing different data resources, it is not only possible to minimize the human efforts and errors but also assist the regulatory bodies and safety councils.<sup>8</sup>

A.M. Wilson<sup>9</sup> defined 'data mining' as 'the use of statistical techniques, such as disproportionality measures for database or large information sources for extracting an unknown information'. At present, three major

DMAs such as the PRR of the Netherlands, the ROR of the United Kingdom and the IC of the WHO are widely used.<sup>9-10</sup>

Speaking of the DMAs, one of the frequently discussed and yet to be resolved question is which algorithm has superior performance. The absence of gold standard,<sup>11</sup> enormous ADR reports, different coding systems, a wide range of data mining processes and structural dissimilarities of databases made the comparisons across the DMAs difficult. Limited studies have been conducted to compare the DMAs thus far. This study aims to compare and appraise the performance of signal detection techniques used in data mining.

## METHODOLOGY

Most commonly used three DMAs (ROR, PRR and IC) were selected based on a literature survey. DMAs were applied retrospectively (Table 1 and 2) in US FDA AERS database to detect five confirmed Drug Event Combinations (DEC) (Table 3). The DEC was selected in regard to the withdrawal of the drug from the market between 2006-2015 or the change in labelling criteria or black box warning of the drug during the time period of 2006-2015. The time period is important because the data available in USFDA AERS database for signal detection is from 2006.

## Data source

AE reports from the FDA AERS database were used for the study. It is a surveillance program used for detecting serious AEs that have not been identified during premarketing analysis.<sup>12</sup>

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as the author is credited and the new creations are licensed under the identical terms.

**Table 1: The 2 × 2 table for the calculation of the signal.**

|                 | Drug of Interest | Other Drugs |
|-----------------|------------------|-------------|
| ADR of interest | A                | B           |
| Other ADR       | C                | D           |

A: The number of reports containing both suspected drug and suspected ADRs  
 B: The number of reports containing drug of choice but with other ADRs  
 C: The number of reports containing the event of interest but with other medications  
 D: The number of reports concerning other medications and other ADRs

**Table 2: Formula for the computation of signal.**

| Serial No. | Measures | Computation   | Threshold   |
|------------|----------|---|---|
| 1          | ROR      | $ROR = (A/B)/(C/D)$<br>$S.E = \sqrt{\frac{1}{A} + \frac{1}{B} + \frac{1}{C} + \frac{1}{D}}$                       | $ROR - 1.96SE > 1$                                      |
| 2          | PRR      | $PRR = (A(A+C))/(B(B+D))$<br>$SE = \sqrt{\left(\frac{1}{A} - \frac{1}{A+C} + \frac{1}{B} + \frac{1}{B+D}\right)}$ | $PRR \geq 2$<br>$\chi^2 > 4$<br>$\geq 3$ cases reported |
| 3          | IC       | $IC = \log_2 \frac{p(x,y)}{p(x)p(y)}$   | $IC - 2SD > 0$  |

$p(x)$  = Probability of a suspected drug being reported in a case report;  $p(y)$  = Probability of a suspected event being reported in a case report;  $p(x, y)$  = Probability that suspected drug and event simultaneously being reported in a case report

**Table 3: Drug Event Combinations.**

| Sl. No. | Drug          | Event   | Present status       |
|---------|---------------|---|----------------------|
| 1       | Propoxyphene  | Serious toxicity to the heart                 | Withdraw from market |
| 2       | Sibutramine   | Cardio and cerebrovascular outcomes and death | Withdraw from market |
| 3       | Rosiglitazone | Serious cardiac toxicity                      | Withdraw from market |
| 4       | Sitagliptin   | Renal failure                                 | Change in label      |
| 5       | Canagliflozin | Urinary Tract Infection                       | Change in label      |

Source: FDA annual report

## Study Procedure

FDA AERS database downloaded from USFDA official website. The database converted into excel from text format for the ease of analysis and computation. Primary suspects and secondary suspect case IDs of the drug of interest were noted from DRUG file. AE pertaining to those case IDs were filtered and listed. Different parameters of the DMA equation (Table 1) were computed and finally applied to the equation (Table 2).

## Statistical analysis

The threshold was predefined as PRR of  $\geq 2.0$  with a Chi-squared test of  $\geq 4.0$ , at least three reports ( $n \geq 3$ ) of that preferred term (PT), IC with  $IC - 2SD > 0$  and ROR with  $ROR - 1.96SE > 1$ . The confidence interval (95%) were considered to be statistically significant.<sup>13</sup>

## RESULTS

ROR, PRR and IC were applied on USFDA AERS database to detect five confirmed DEC and to compare the DMAs.

### Propoxyphene

Propoxyphene was approved in 1957 and was withdrawn from the market in 2010 owing to serious cardiac toxicity. A total of 366 DEC were reported from 2005Q1 to 2010Q4 in USFDA AERS database (Table 4). ROR, 1.43 (95% CI, **1.26**-1.52) and PRR, 2.8 (95% CI, **2.24**- 3.32) showed positive signal from 2009Q1 and IC, 0.62 (95% CI, **0.06** -1.02) showed positive signal from 2007Q1 for the selected DEC. PTs used for data mining were “cardiac arrest” and “cardio-respiratory arrest”. Positive signals were highlighted with bold letters.

### Sibutramine

Sibutramine was approved in 1997 as an oral anorexiant. Later in 2010, it was withdrawn from the market due to severe cardio and cerebrovascular accidents and associated death. A total of 4182 DEC were reported from 2005Q1 to 2010Q4 in USFDA AERS database (Table 4). Only IC, 0.28 (95% CI, **0.05** – 0.54) showed a positive signal for sibutramine from 2009Q1 whereas ROR and PRR failed to show a positive signal for the given DEC in the specified duration. PT used for data mining were “cerebrovascular disease” and “stroke”. Positive signals were highlighted with bold letters.

### Rosiglitazone

Rosiglitazone was approved in 1999 for diabetes mellitus and was withdrawn from the market because of cardiac toxicity and associated death in 2010. A total of 11839 DEC were reported from 2005Q1 to 2010Q4 in USFDA AERS database (Table 4). ROR, 1.4 (95% CI, **1**-1.8) showed positive signal from 2009Q3, PRR, 2.7 (95% CI, **2.1**-3.6) from 2009Q4 and IC, 0.24 (95% CI, **0.02** - 0.5) showed positive signal from 2008Q3 for the given DEC in the specified period. PT used for data mining were “Myocardial Infarction” (MI). Positive signals were highlighted with bold letters.

### Sitagliptin

Sitagliptin was approved in 2006 for diabetes mellitus. Recently (2012), the labelling criteria of sitagliptin had changed to include a black box warning of renal failure associated with sitagliptin. A total of 854 DEC were reported from 2005Q1 to 2015Q2 in USFDA AERS database (Table 5). Only IC, 0.1 (95% CI, **0** – 0.23) showed positive signal (from 2011Q2) whereas ROR and PRR failed to show positive signal. PT used for data mining were “renal failure acute”, “creatinine renal clearance decreased” and “blood creatinine increases”. Positive signals were highlighted with bold letters.

### Canagliflozin

Canagliflozin is approved in 2013 for the treatment of Diabetes Mellitus and in 2015 it was subjected to a change in labelling criteria to incorporate safety warning of Urinary Tract Infection (UTI). A total of 566 DEC were reported from 2013Q2 to 2015Q4 in USFDA AERS database (Table 5). IC, 0.78 (95% CI, **0.53**- 0.98) and ROR, 1.42 (95% CI, 1.11-1.94) showed a positive signal at 2015Q3 and 2015Q4 respectively. PT used for data mining were “urinary tract infection”. Positive signals were highlighted with bold letters.

### Sensitivity of DMAs

The sensitivity of three DMAs were assessed (Table 6) in accordance with its potential to identify the signals prior to the withdrawal date. Out of

**Table 4: DMA of propoxyphene and reporting of cardiac arrest, Sibutramine and reporting of cerebrovascular disease and Rosiglitazone and reporting of MI.**

| Time Period | Propoxyphene and reporting of cardiac arrest |            |        | Sibutramine and reporting of cerebrovascular disease |            |        | Rosiglitazone and reporting of MI |            |        |
|-------------|--|------------|--------|--|------------|--------|-----------------------------------|------------|--------|
|             | ROR-1.96SE                                   | PRR-1.96SE | IC-2SD | ROR-1.96SE   | PRR-1.96SE | IC-2SD | ROR-1.96SE                        | PRR-1.96SE | IC-2SD |
| 2005Q1      | <b>-1.72</b>                                 | -0.69      | -2.83  |  |            | -6.64  | -2.16                             | -1.15      | -1.86  |
| 2005Q2      | <b>-2.03</b>                                 | -1.03      | -2.86  | -4.89  | -4.89      | -3.83  | -2.69                             | -1.6       | -1.81  |
| 2005Q3      | <b>-0.87</b>                                 | 0.13       | -1.63  | -4.38  | -4.38      | -2.7   | -2.95                             | -1.84      | -1.39  |
| 2005Q4      | <b>0.42</b>                                  | 1.41       | -0.27  | -3.81  | -3.81      | -1.91  | -2.87                             | -1.75      | -1.13  |
| 2006Q1      | <b>0.4</b>                                   | 1.39       | -0.26  | -3.74  | -3.74      | -1.71  | -2.97                             | -1.83      | -1.02  |
| 2006Q2      | <b>0.42</b>                                  | 1.42       | -0.18  | -4.01  | -4.01      | -1.65  | -2.79                             | -1.67      | -0.87  |
| 2006Q3      | <b>0.41</b>                                  | 1.4        | -0.18  | -3.97  | -3.97      | -1.52  | -2.78                             | -1.78      | -0.83  |
| 2006Q4      | <b>0.53</b>                                  | 1.53       | -0.05  | -3.97  | -3.97      | -1.52  | -2.64                             | -1.53      | -0.76  |
| 2007Q1      | <b>0.66</b>                                  | 1.65       | 0.06   | -3.97  | -3.97      | -1.52  | -2.53                             | -1.53      | -0.73  |
| 2007Q2      | <b>0.63</b>                                  | 1.62       | 0.04   | -3.33  | -3.14      | -1.09  | -1.71                             | -0.66      | -0.43  |
| 2007Q3      | <b>0.65</b>                                  | 1.64       | 0.05   | -2.92  | -2.79      | -0.87  | -1.25                             | -0.22      | -0.27  |
| 2007Q4      | <b>0.64</b>                                  | 1.63       | 0.05   | -2.53  | -2.53      | -0.74  | -0.99                             | 0.06       | -0.19  |
| 2008Q1      | <b>0.91</b>                                  | 1.89       | 0.35   | -2.42  | -2.33      | -0.68  | -0.76                             | 0.28       | -0.13  |
| 2008Q2      | <b>0.94</b>                                  | 1.92       | 0.38   | -2.15  | -2.08      | -0.59  | -0.39                             | 0.63       | -0.01  |
| 2008Q3      | <b>0.94</b>                                  | 1.93       | 0.39   | -2.08  | -2.08      | -0.58  | -0.29                             | 0.72       | 0.02   |
| 2008Q4      | <b>0.97</b>                                  | 1.96       | 0.42   | -0.32  | -0.31      | -0.02  | 0.18                              | 1.17       | 0.21   |
| 2009Q1      | <b>1.26</b>                                  | 2.23       | 0.78   | -0.18  | -0.17      | 0.05   | 0.64                              | 1.6        | 0.48   |
| 2009Q2      | <b>1.25</b>                                  | 2.22       | 0.77   | -0.12  | -0.11      | 0.07   | 0.73                              | 1.69       | 0.55   |
| 2009Q3      | <b>1.24</b>                                  | 2.22       | 0.77   | 0.1  | 0.1        | 0.18   | 1                                 | 1.94       | 0.78   |
| 2009Q4      | <b>1.26</b>                                  | 2.24       | 0.79   | 0.31   | 0.3        | 0.28   | 1.17                              | 2.1        | 0.96   |
| 2010Q1      | <b>1.4</b>                                   | 2.37       | 1      | 0.4  | 0.39       | 0.34   | 1.27                              | 2.19       | 1.07   |
| 2010Q2      | <b>1.38</b>                                  | 2.35       | 0.97   | 0.58   | 0.56       | 0.45   | 1.4                               | 2.31       | 1.24   |
| 2010Q3      | <b>1.4</b>                                   | 2.37       | 1      | 0.77   | 0.75       | 0.6    | 1.59                              | 2.48       | 1.54   |
| 2010Q4      | <b>1.43</b>                                  | 2.4        | 1.04   | 0.86   | 0.83       | 0.66   | 1.69                              | 2.57       | 1.7    |

ROR=Reporting Odds Ratio; PRR= Proportional Reporting Ratio; IC= Information Component; SD=Standard Deviation; SE= Standard Error. Bold Letters: Positive Signal

five, IC showed maximum sensitivity (100%) followed by ROR (60%) and PRR (40%).

#### *The sensitivity of DMA based on Early Detection*

DMAs were assessed for their ability for early detection of a signal (Table 7). Index Date of Withdrawn/Label change (IDW/L) were identified from literature or official websites of USFDA. The quadrant from which DEC started showing positive signals were allocated as Index Date of Detection (IDD). Early detection was quantified by IDD subtracted from IDW/L. IC was found to be the most sensitive, as it detects positive signal well before other DMAs and there is no remarkable difference in sensitivity of early detection between ROR and PRR.

#### *Sensitivity associated with the number of reports per DEC*

DMAs were assessed for their sensitivity with respect to the number of reports required to show positive signal (Figure 1). Sensitivity associated with the number of reports per DEC suggested that IC need compara-

tively fewer reports to show positive signal than ROR and PRR. Parallel to the above result, ROR and PRR are almost identical in their sensitivity.

## DISCUSSION

In this study, the most commonly used three DMAs are ROR, PRR and IC which are applied retrospectively in USFDA AERS database to detect five known and confirmed ADRs associated with the drug withdrawal or change in labelling criteria. Further comparisons across the selected DMAs were executed to identify the sensitivity by means of early detection and number of reports.

Hitherto, there is no clearly defined method to compare the DMAs with respect to their sensitivity or performance. The major drawback in comparing the DMAs is the lack of golden standards.<sup>14</sup> In E.P. van Puijenbroek, A. Bate<sup>11</sup> study, they compared the DMAs like PRR, Yule's Q and Chi-square with IC, which was considered as a golden standard by them. Another study conducted by K. Kubota<sup>15</sup> in Japanese spontaneous reports, compared five DMAs. According to Kubota *et al.*, the number of

**Table 5: Sitagliptin and reporting of renal failure and Canagliflozin and reporting of UTI.**

| Time Period | Sitagliptin and reporting of renal failure |            |        | Canagliflozin and reporting of UTI |            |        |
|-------------|--|------------|--------|------------------------------------|------------|--------|
|             | ROR-1.96SE                                 | PRR-1.96SE | IC-2SD | ROR-1.96SE                         | PRR-1.96SE | IC-2SD |
| 2006Q4      | -0.13                                      | 0.89       | -2.84  | -                                  | -          | -      |
| 2007Q1      | 0.09                                       | 1.1        | -1.22  | -                                  | -          | -      |
| 2007Q2      | -0.37                                      | 0.64       | -0.74  | -                                  | -          | -      |
| 2007Q3      | -0.3                                       | 0.7        | -0.49  | -                                  | -          | -      |
| 2007Q4      | -0.2                                       | 0.8        | -0.34  | -                                  | -          | -      |
| 2008Q1      | -0.18                                      | 0.82       | -0.28  | -                                  | -          | -      |
| 2008Q2      | -0.13                                      | 0.87       | -0.27  | -                                  | -          | -      |
| 2008Q3      | -0.22                                      | 0.78       | -0.26  | -                                  | -          | -      |
| 2008Q4      | -0.26                                      | 0.74       | -0.23  | -                                  | -          | -      |
| 2009Q1      | -0.25                                      | 0.75       | -0.2   | -                                  | -          | -      |
| 2009Q2      | -0.25                                      | 0.76       | -0.17  | -                                  | -          | -      |
| 2009Q3      | -0.26                                      | 0.75       | -0.16  | -                                  | -          | -      |
| 2009Q4      | -0.22                                      | 0.79       | -0.13  | -                                  | -          | -      |
| 2010Q1      | -0.17                                      | 0.83       | -0.1   | -                                  | -          | -      |
| 2010Q2      | -0.17                                      | 0.83       | -0.08  | -                                  | -          | -      |
| 2010Q3      | -0.15                                      | 0.86       | -0.07  | -                                  | -          | -      |
| 2010Q4      | -0.14                                      | 0.86       | -0.06  | -                                  | -          | -      |
| 2011Q1      | -0.09                                      | 0.91       | -0.03  | -                                  | -          | -      |
| 2011Q2      | -0.05                                      | 0.96       | 0      | -                                  | -          | -      |
| 2011Q3      | 0.01                                       | 1.01       | 0.03   | -                                  | -          | -      |
| 2011Q4      | 0  | 0.99       | 0.03   | -                                  | -          | -      |
| 2012Q1      | 0  | 1          | 0.04   | -                                  | -          | -      |
| 2012Q2      | -0.01                                      | 0.99       | 0.04   | -                                  | -          | -      |
| 2012Q3      | -0.01                                      | 1          | 0.05   | -                                  | -          | -      |
| 2012Q4      | 0.01                                       | 1.01       | 0.06   | -                                  | -          | -      |
| 2013Q1      | -0.01                                      | 0.99       | 0.05   | -                                  | -          | -      |
| 2013Q2      | -0.01                                      | 0.98       | 0.06   | -1.39                              | -0.36      | -4.4   |
| 2013Q3      | -0.03                                      | 0.97       | 0.06   | 0.08                               | 3.63       | -1.81  |
| 2013Q4      | -0.05                                      | 0.95       | 0.06   | 0.18                               | 4.14       | -1.04  |
| 2014Q1      | -0.05                                      | 0.95       | 0.06   | 0.2                                | 3.09       | -0.87  |
| 2014Q2      | -0.1                                       | 0.9        | 0.05   | -0.21                              | 0.8        | -0.84  |
| 2014Q3      | -0.16                                      | 0.85       | 0.04   | -0.18                              | 0.82       | -0.62  |
| 2014Q4      | -0.18                                      | 0.83       | 0.04   | -0.06                              | 0.96       | -0.49  |
| 2015Q1      | -0.17                                      | 0.83       | 0.04   | 0.41                               | 1.39       | -0.17  |
| 2015Q2      | -0.22                                      | 0.78       | 0.03   | -0.55                              | 0.48       | -0.15  |
| 2015Q3      | -  | -          | -      | 0.9                                | 1.83       | 0.53   |
| 2015Q4      | -  | -          | -      | 1.11                               | 1.97       | 0.76   |

ROR=Reporting Odds Ratio; PRR= Proportional Reporting Ratio; IC= Information Component; SD=Standard Deviation; SE= Standard Error. Bold Letters: Positive Signal

**Table 6: The DEC detected by the three DMAs.**

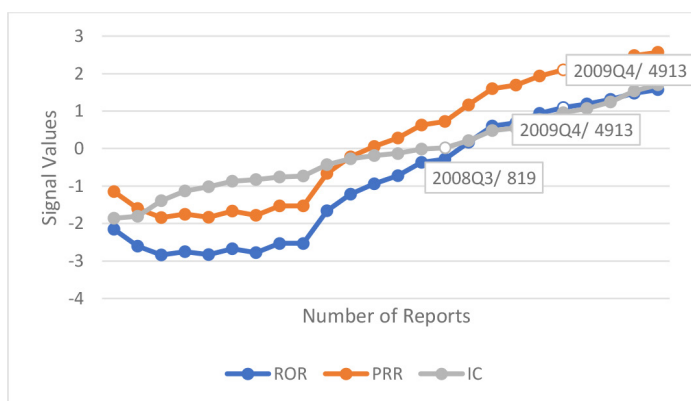
| Sl. no | Drug Event Combination                    | ROR | PRR | IC   |
|--------|---|-----|-----|------|
| 1      | Propoxyphene and cardiac arrest           | √   | √   | √    |
| 2      | Sibutramine and cerebrovascular disorders | X   | X   | √    |
| 3      | Rosiglitazone and MI                      | √   | √   | √    |
| 4      | Sitagliptin and renal failure             | X   | X   | √    |
| 5      | Canagliflozin and UTI                     | √   | X   | √    |
|        | Sensitivity                               | 60% | 40% | 100% |

√ means that the DMA could identify the signals prior to the withdrawal date; X means that algorithms did not identify the signals prior to the withdrawal date.

**Table 7: Sensitivity of each DMA in terms of early detection.**

| Sl. No. | Drug          | Event                     | IDW    | IDD    |        |                  | IC     | Earlier Quarters | Status change date | Present status       |
|---------|---------------|---------------------------|--------|--------|--------|------------------|--------|------------------|--------------------|----------------------|
|         |               |                           |        | ROR    | PRR    | Earlier Quarters |        |                  |                    |                      |
| 1       | Propoxyphene  | Cardiac Arrest            | 2010Q4 | 2009Q1 | 2009Q1 | 3                | 2007Q1 | 14               | 2010               | Withdraw from market |
| 2       | Sibutramine   | Cerebrovascular disorders | 2010Q4 | —      | —      | —                | 2009Q1 | 7                | 2010               | Withdraw from market |
| 3       | Rosiglitazone | MI                        | 2010Q4 | 2009Q3 | 2009Q4 | 4                | 2008Q3 | 9                | 2010               | Withdraw from market |
| 4       | Sitagliptin   | Renal failure             | 2012Q4 | —      | —      | —                | 2011Q2 | 4                | 2012               | Change in label      |
| 5       | Canagliflozin | Urinary Tract Infection   | 2016Q2 | 2015Q4 | 2015Q4 | 1                | 2015Q3 | 2                | 2016               | Change in label      |

Abbreviations: ROR = Reporting Odds Ratio; PRR = Proportional Reporting Ratio; IC = Information Component; IDW = Index Date of Withdrawn; IDD = Index Date of Detection.



**Figure 1:** Sensitivity associated with the number of reports per DEC. Abbreviations: ROR = Reporting Odds Ratio; PRR = Proportional Reporting Ratio; IC = Information Component.

DECs identified as signals were considered as the measure of sensitivity. Compared to the above two studies,<sup>11,15</sup> the findings of the present study seemed to be more reliable because the reference standard considered is more robust. Comparison of DMAs is possible in many ways, but the main aspect which directly influences the DMA values is the number of reports. Thus, according to this study, a cumulative number of reports should be considered as the measure of sensitivity. Early detection of ADRs is the main advantage of DMAs,<sup>16-17</sup> therefore it can be considered as a measure of sensitivity.

This study indicates that IC is more sensitive in terms of early detection as well as the number of reports. However, among the DMAs, the sensitivity difference is statistically not significant. Moreover, the sensitivity of DMAs may vary if a different database is used hence, we cannot conclude that IC is the most sensitive DMA. Nevertheless, for a given set of data and DMAs, IC showed more sensitivity than other DMAs. It has been observed that the sensitivity is proportional to the number of reports. According to V.G. Koutkias and M.-C. Jaulent,<sup>18</sup> the number of reports is an important factor for signal strength.

According to the literature review, it was observed that no study has been attempted to compare DMA to identify the sensitive method among the existing algorithms. We comply with P. Waller<sup>19</sup> study, as the selection of DMAs should be done on the basis of their specificity, sensitivity and predictive value in addition to a factor observed from our study. Early detection of ADR plays a vital role which may reduce the casualties and provide sufficient time for a regulatory decision.

It is not surprising that IDD is earlier than IDW because once the positive signal is identified, FDA requires adequate time to evaluate the situation, assess the risk-benefit profile of the drug and for a regulatory decision. However, the use of DMAs could trigger the initiation of this process earlier by recognizing signals in advance. Few studies have compared the traditional method of ADR detection with computer-based signal detection techniques. A study conducted by A.W. I<sup>20</sup> concluded that DMAs detected safety signals well before the conventional ways. According to D.J. Graham,<sup>21</sup> there were 88,000–140,000 excess of cardiac disorders associated with rofecoxib. It could have been reduced if the safety signal was detected earlier.

We rely on the progression between the early detection of an ADR signal to the final decision that drug withdrawal or labelling change could more likely be acquired due to the earlier detection of ADR signals as a result of the applications of DMAs. Consequently, the time when the FDA makes a decision will correspondingly occur earlier.

### Limitations of the study

There are some concerns regarding FDA AERS database. Under-reporting is the main concern with regards to any spontaneous reporting system (SRS) database. SRS will not reflect the actual picture of the scenario. Thus, more often the situation was underestimated. Secondly, the reporting may get biased when there is a change in labelling criteria or any special updates regarding an ADR of the drug. As a result, over-reporting of that particular ADR will occur which will decrease the signal strength of other ADRs of the same drug (change in Ni value will affect the signal strength).

The selection of brand names for data mining is another limitation which we had come across. FDA AERS database is a collection of ADR reports around the world but the main contributor is United States (US). The brand names used for data mining in this study mainly focused on the US, European countries and India. Thus, the chances of missing data cannot be ruled out.

## CONCLUSION

The aim of the study was to compare and appraise the performance of signal detection techniques used in data mining. It is the first study attempted to address the importance of early detection of ADR and identification of the sensitive method. Even though there is no statistically significant difference among three DMAs, IC was found to be sensitive method compared with other two DMAs in FDA AERS database. The sensitivity and comparability of DMA is proportionate to the number of reports of DEC.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## ABBREVIATIONS

**FDA:** Food and Drug Administration; **AERS:** Adverse Events Reporting System; **AEs:** Adverse Event; **PEM:** Prescription-Event Monitoring; **DMAs:** Data Mining Algorithms; **ADRs:** Adverse Drug Reactions; **ROR** : Reporting Odds Ratio; **MGPS:** Multi-item Gamma Poisson Shrinker; **PRR:** Proportional Reporting Ratio; **IC:** Information Component; **DEC:** Drug Event Combinations; **IDW/L:** Index Date of Withdrawn/Label change.

## REFERENCES

- Harpaz R, DuMouchel W, LePendou P, Bauer MA, Ryan P, Shah NH. Performance of pharmacovigilance signal-detection algorithms for the fda adverse event reporting system. *Clin Pharmacol Ther.* 2013;93(6):539-46. 10.1038/clpt.2013.24
- Bie SD, Ferrajolo C, Straus SM, Verhamme KM, Bonhoeffer J, Wong IC, et al. Pediatric drug safety surveillance in fda-aers: A description of adverse events from grip project. *PLoS One.* 2015;10(6):e0130399. 10.1371/journal.pone.0130399
- Rossi AC, Knapp DE, Anello C, et al. Discovery of adverse drug reactions: A comparison of selected phase iv studies with spontaneous reporting methods. *JAMA.* 1983;249(16):2226-8. 10.1001/jama.1983.03330400072029
- Rothman KJ, Lanes S, Sacks ST. The reporting odds ratio and its advantages over the proportional reporting ratio. *Pharmacoepidemiology and Drug Saf.* 2004;13(8):519-23. 10.1002/pds.1001
- Napoli AA, Wood JJ, Coumbis JJ, Soitkar AM, Seekins DW, Tilson HH. No evident association between efavirenz use and suicidality was identified from a disproportionality analysis using the faers database. *J Int AIDS Soc.* 2014;17(1):9214. 10.7448/ias.17.1.19214
- Evans SJ, Waller PC, Davis S. Use of proportional reporting ratios (prrs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiol Drug Saf.* 2001;10(6):483-6. 10.1002/pds.677
- Bate A, Evans SJ. Quantitative signal detection using spontaneous adr reporting. *Pharmacoepidemiol Drug Saf.* 2009;18(6):427-36. 10.1002/pds.1742
- Bate A, Lindquist M, Edwards IR, Orre R. A data mining approach for signal detection and analysis. *Drug Saf.* 2002;25(6):393-7.
- Wilson AM, Thabane L, Holbrook A. Application of data mining techniques in

- pharmacovigilance. *Br J Clin Pharmacol.* 2004;57(2):127-34. 10.1046/j.1365-2125.2003.01968.x
10. Szarfman A, Machado SG, 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(6):381-92.
  11. Puijenbroek EPV, Bate A, Leufkens HG, Lindquist M, Orre R, Egberts AC. A comparison of measures of disproportionality for signal detection in spontaneous reporting systems for adverse drug reactions. *Pharmacoepidemiol Drug Saf.* 2002;11(1):3-10. 10.1002/pds.668
  12. Shamloo B, Chhabra P, Freedman A, Potosky A, Malin J, Smith S. Novel adverse events of bevacizumab in the us fda adverse event reporting system database. *Drug Saf.* 2012;35(6):507-18. 10.2165/11597600-000000000-00000
  13. Deshpande G, Gogolak V, Smith S. Data mining in drug safety. *Pharmaceut Med.* 2010;24(1):37-43. 10.1007/BF03256796
  14. Lindquist M, Stahl M, Bate A, Edwards IR, Meyboom RH. A retrospective evaluation of a data mining approach to aid finding new adverse drug reaction signals in the who international database. *Drug Saf.* 2000;23(6):533-42.
  15. Kubota K, Koide D, Hirai T. Comparison of data mining methodologies using japanese spontaneous reports. *Pharmacoepidemiol Drug Saf.* 2004;13(6):387-94. 10.1002/pds.964
  16. Chakraborty BS. Pharmacovigilance: A data mining approach to signal detection. *Indian J Pharmacol.* 2015; 47(3):241-2. 10.4103/0253-7613.157102
  17. Hauben M, Reich L. Potential utility of data-mining algorithms for early detection of potentially fatal/disabling adverse drug reactions: A retrospective evaluation. *J Clin Pharmacol.* 2005;45(4):378-84. 10.1177/0091270004273936
  18. Koutkias VG, Jaulent MC. Computational approaches for pharmacovigilance signal detection: Toward integrated and semantically-enriched frameworks. *Drug Saf.* 2015;38(3):219-32. 10.1007/s40264-015-0278-8
  19. Waller P, Puijenbroek EV, Egberts A, Evans S. The reporting odds ratio versus the proportional reporting ratio: 'Deuce'. *Pharmacoepid and drug saf.* 2004; 13(8):525-6.
  20. I AW, Pratt NL, Kalisch LM, Roughead EE. Comparing time to adverse drug reaction signals in a spontaneous reporting database and a claims database: A case study of rofecoxib-induced myocardial infarction and rosiglitazone-induced heart failure signals in australia. *Drug Saf.* 2014;37(1):53-64. 10.1007/s40264-013-0124-9
  21. Graham DJ, Campen D, Hui R, Spence M, Cheetham C, Levy G, *et al.* Risk of acute myocardial infarction and sudden cardiac death in patients treated with cyclo-oxygenase 2 selective and non-selective non-steroidal anti-inflammatory drugs: Nested case-control study. *Lancet.* 2005;365(9458):475-81. 10.1016/S0140-6736(05)17864-7

**Article History:** Submission Date : 19-05-2018; Revised Date : 30-06-2018; Acceptance Date : 06-08-2018.

**Cite this article:** Subeesh V, Maheswari E, Saraswathy GR, Swaroop AM, Minnikanti SS. A Comparative Study of Data Mining Algorithms used for Signal Detection in FDA AERS Database. *J Young Pharm.* 2018;10(4):444-9.