



International Journal of Emerging Technologies in Computational and Applied Sciences (IJETCAS)

www.iasir.net

Study of Adverse Drug effects Using Knowledge Discovery Methods

Vasudha Satalkar¹, Mansee Rathore² & D. S. Joshi

Dept. of Biotechnology, MGM University of Health Sciences and MGM College of Engineering and Technology, Kamothe, Navi Mumbai. 410209 India

Abstract: This is pilot study to find associations, trends and patterns resulting in useful knowledge about drugs and their side effects or adverse drug reactions reported in scientific literature. Method of manual addition of data about drugs, side effects/adverse drug reactions, disease reported in scientific literature has been used to create tables as data collection technique. The current analysis has been done on approximately 2000 records. Several point estimates like the Reporting Odds Ratio (ROR), Proportional ADR Reporting Ratio (PRR) and Yule's Q, have been used, in combination with additional estimators of the precision of point estimates such as the Chi-square test, or the lower limits of the 95% confidence intervals of the point estimates. Around fourteen potentially significant associations (signals) have been detected.

Keywords— Adverse Drug Events, Signal detection, Associations, Data mining

I. INTRODUCTION

Clinical trials involve several thousand patients at most; less common side effects and ADRs are often unknown at the time a drug enters the market. Even very severe ADRs such as liver damage are often undetected because study populations are small. Post-marketing pharmacovigilance uses tools such as data mining and investigation of case reports to identify the relationships between drugs and ADRs.

"Close monitoring of adverse drug reactions (ADRs) is necessary due to the limited size of pre-marketing trials. A timely detection of either new ADRs or an increase of the frequency of ADRs which are already known to be associated with the drugs involved is necessary" to prevent harm to patients [1].

"New chemical entities (NCEs) admitted to the market between 1975 and 1999, it was found that out of 548 NCEs, 45(8.2%) acquired black box warnings and 16(2.9%) were withdrawn from the market", which is a huge loss to pharmaceutical industry.[2]

"Signals originating from statistical methods measuring disproportionality of reporting of drug-event pairs are referred to as "Signals of Disproportionate Reporting (SDR)". The specific disproportionality measure implemented in the EudraVigilance Data Analysis System is the Proportional Reporting Ratio (PRR) [3].

A statistical approach called disproportionality analysis (DPA) is used to identify events that occur at a greater frequency than would be expected by chance. The DPA calculation is derived from a two-by-two table shown below

Table 1: Disproportionality Analysis

	Drug X	All other drugs
Event of interest	A	C
All other events	B	D

Measure	Definition	Utilization
Reporting Odds Ratio (ROR)	$[A/C] / [B/D]$	Netherlands Pharmacovigilance Foundation
Yules Q ratio	$[AD-BC] / [AD+BC]$	
Proportional reporting ratio (PRR)	$[A/(A+C)] / [C/(C+D)]$	UK Yellow Card

Wilson, A. M., Thabane, L., & Holbrook, A. has quoted in their paper "To our knowledge, measures of disproportionality are the only techniques which have been used to identify ADEs" [4] Detection of bradycardia associated with propranolol hydrochloride has been illustrated by Gavali et al [5]. Many others have used illustrated this data mining technique in their papers. [6-11]

Such statistical analysis is not sufficient by itself, as there are always biases in the data collection and analysis. Therefore, researchers verify all signals identified by such data mining. And while such data mining can be an important tool in the of post-market product surveillance, it is intended only to enhance current pharmacovigilance techniques, not replace them.

II. MATERIALS AND METHODS

All the work has been done using computer and necessary software for data collection and analysis.

Data collection

About 183 journal issues were read physically. Most papers read fall between years 2000 to 2010. Terminology used by the authors reporting adverse drug reactions associated with the drug was used to collect data about adverse drug reactions. A spreadsheet was created identifying drug, side effect/ adverse reaction, journal citation; disease the drug is intended to cure using Table 2 to feed information manually.

Inclusion/Exclusion Criteria:

For Drugs: Single drugs having valid chemical structures listed in SIDER Side Effects database or Pub-Chem were included. Biologics were excluded

For ADR's: Drug – Drug interactions were excluded.

Table 2: Data Collection Format

S.No.	Drug Name	Adverse Effect	Journal	Disease name
1	Aripirazole	Severe muscle rigidity and severe influenza	Clin Alert. Sept 30, 2008 v46 No.18	Schizophrenia
2	Bupropion	Severe Constipation	Clin Alert. Sept 30, 2008 v46 18 p1(16).	Depression
3	Clozapine	Severe muscle rigidity and severe influenza	Clin Alert. Sept 30, 2008 v46 18 p1(16).	Schizophrenia
4	Disulfiram	Punding	Clin Alert. Sept 30, 2008 v46 18 p1(16).	Alcoholism

Data cleaning

- Searched for duplicate records,
- Missing values,
- Spelling mistakes,
- Same ADR with different spellings in different journals.

III. DATABASE CREATION AND QUERYING

The spreadsheets were then imported into MySQL using SQLyog Community Edition – MySQL GUI v 8.01 © 2001-2009 provided by Webyog Softworks Pvt. Ltd. MySQL is a open source relational database management system. It comes with tools that assist you in organizing your data and simplifying information retrieval and modification. Various queries were fired in MySQL to find information about data characteristics. Frequency of all drugs that showed up in our database was calculated individually. A total of 350 distinct drugs were found. Similarly, frequency of all ADR's that showed up in our database was found out. A total of 891 distinct ADR's were found. Values of A, B, C and D illustrated in Table 1 were also computed and subjected to further calculations to find associations between a drugs and adverse event using standards established in literature enumerated below.

Association measures and signal generation criteria

As per literature minimum criteria for a signal is 3 or more cases, Proportional reporting ratio (PRR) at least 2. A similar signal generation criterion for Reporting Odds Ratio (ROR) is $ROR > 1$. Yule's Q ratio is based on the odds ratio and a symmetric measure taking on values between -1 and +1. 1 (one) implies perfect negative or positive association, 0 (zero) no association. A potential signal may warrant further investigation for a $PRR > 2$, where 'A' is > 3 and Chi square > 4 .

“As per standards, number of case reports should be more than or equal to 3 (Three) and the signal detected by PRR and ROR more than or equal to 3 (Three) shall be considered as a potent indicator for producing ADR in large population⁵.”

Data Analysis

Data collected by manual addition of journal citations from our database, that indicated $A > 3$ or equal to 3, was subjected to calculations shown in Table 1 above and illustrated in papers [3, 6, 7 & 9] and tabulated as shown

below in Table 6. However, most drug / Adverse Effect combination from the results in Table 6 were not reported before marketing by the manufacturers and are therefore post-marketing signals.

IV. RESULTS AND DISCUSSION

The cost of adverse drug reactions to patients is well illustrated by Philip J. Schneider in his paper “Measuring the Cost of Adverse Drug Events and Medication Errors” *Pharma Times*, Vol. 45 - No. 5 - May 2013 discussed below.

The frequency of harm resulting from the use of medicines is common and the associated costs are significant. In the Harvard Medical Practice Study, it was found that 3.7% of patients are harmed while they are in the hospital. The most common cause of harm is drug complications, accounting for 19% of all adverse medical events.

Table 3: Clinical outcomes of harm resulting from the use of medicines

Clinical outcome	Example of cost-related outcomes
Additional laboratory tests	Serum drug concentrations, electrolyte concentrations, serum glucose concentrations, complete blood count, platelet count
Non-invasive procedures	Angiography, CT scan, ECG, Doppler ultrasonography, EEG, echocardiography fluoroscopy, MRI scan, roentgenography, pulmonary function studies
Additional treatment	Medications, aerosolizations, transfusions,, oxygen
Invasive monitoring or procedures	Coronary angiography, myocardial perfusion studies, pericardiocentesis, PTCA, surgery, dialysis
Increased length of stay in the hospital	Daily room rate
Transfer to intensive care	ICU bed, insertion of arterial line, intubation, insertion of Swan-Ganz catheter, medical emergency

Frequency distribution resulting from our study:

Tables 4 & 5 below shows most frequent adverse effects and drugs reported in our pilot study.

Adverse Drug Reactions and Drugs Reported in Literature Reviewed: Stevens Johnson Syndrome, Rash and Nausea are most reported adverse drug reactions in literature resources read by us manually and from our analysis.

Table 4: Top Ten Adverse Drug Reactions and Drugs reported

TOP TEN ADR'S		TOP TEN DRUGS	
Frequency	Adverse Drug Reaction	Frequency	Drug
47	Stevens johnson syndrome	56	Methotrexate
39	Rash	46	Varenicline
35	Nausea	44	Hydroxychloroquine
32	Death	44	Sitagliptin
27	Toxic epidermal necrolysis	35	Olanzapine
22	Diarrhea	33	Abacavir
22	Vomiting	31	Leflunomide
21	Thrombocytopenia	30	Chlorthalidone
21	Pancreatitis	28	Reserpine
18	Headache	28	Carbamazepine
18	Fatigue	28	Atomoxetine

Association detected:

Following associations have been detected from our studies by performing statistical calculations illustrated in Table 1 and tabulated below in Table 6.

- Varenicline and aggressiveness, suicide attempt, suicide ideation, depression
- Warfarin and bleeding
- Carbamazepine and Stevens Johnson Syndrome
- Heparin and thrombocytopenia
- Carbamazepine and Toxic Epidermal Necrolysis
- Leflunomide and Peripheral Neuropathy
- Simvastatin and Rhabdomyolysis
- Clozapine and Agranulocytosis
- Phenytoin and Stevens Johnson Syndrome
- Didanosine and Pancreatitis

- Abacavir and Hypersensitivity

Table 6: Associations between drug and ADR detected

Drug	Adverse effects	Reporting Odds Ratio (ROR)	Yules Q Ratio	Upper 95% CIFOR PRR	Lower 95% CIFOR PRR	Proportional Reporting Ratio PRR	Chi Square X2	Disease
Carbamazepine	Stevens Johnson syndrome	7.18	0.65	2.44	1.02	5.64	22.04	Epilepsy
Varenicline	Depression	14.07	0.77	3.5	1.53	12.36	35.67	Anti psychotic
Heparin	Thrombocytopenia	18.13	0.83	3.53	1.73	13.84	44.39	Anti coagulant
Carbamazepine	Toxic epidermal necrolysis	8.52	0.71	2.87	1.08	7.18	20.99	Epilepsy
Leflunomide	Peripheral neuropathy	23.01	0.85	4.49	1.59	20.88	31.32	Rheumatoid arthritis
Simvastatin	Rhabdomyolysis	58.53	0.93	5.1	2.47	44.15	78.00	Hyperlipidemia
Clozapine	Agranulocytosis	32.96	0.9	4.33	2.04	24.25	50.8	Antipsychotic
Varenicline	Aggressiveness	14.72	0.78	4.09	1.16	13.83	19.35	Smoking cessation
Varenicline	Suicide attempt	19.63	0.81	4.49	1.34	18.41	23.43	Smoking cessation
Varenicline	Suicidal ideation	9.81	0.73	3.58	0.87	9.24	13.97	Smoking cessation
Phenytoin	Stevens Johnson syndrome	5.43	0.6	2.54	0.46	4.48	7.48	Epilepsy
Didanosine	Pancreatitis	24.47	0.87	3.84	1.78	16.65	37.45	Infection HIV
Warfarin	Bleeding	238.91	0.96	7.44	3.04	187.93	137.26	Embolicism
Abacavir	Hypersensitivity	12.3	0.77	3.35	1.38	10.64	29.65	Infection HIV

Association Validation:

Using SIDER Resource: These associations were checked in SIDER side effects database created by European Molecular Biology Laboratory (EMBL) to confirm if they were already reported by manufacturers of the drugs. “SIDER is organized around the side-effects observed with marketed drugs, as determined primarily by product inserts.”¹¹ Most reports are either post marketing or without any frequency information.

Table 7: Comparison with SIDER side effects resource

Drug	Adverse Effect	Sider database	Report
Carbamazepine	Stevens johnson syndrome	Reported	No frequency information
Varenicline	Depression	Reported	Post marketing
Heparin	Thrombocytopenia	Reported	Post marketing
Carbamazepine	Toxic epidermal necrolysis	Reported	Rare
Leflunomide	Peripheral neuropathy	Reported	Post marketing
Simvastatin	Rhabdomyolysis	Reported	Post marketing
Clozapine	Agranulocytosis	Reported	Boxed warning
Varenicline	Aggressiveness	Reported	Post marketing
Varenicline	Suicide attempt	Reported	Post marketing
Varenicline	Suicidal ideation	Reported	Post marketing
Phenytoin	Stevens johnson syndrome	Reported	No frequency information
Didanosine	Pancreatitis	Reported	Post marketing
Warfarin	Bleeding	Reported	Boxed warning
Abacavir	Hypersensitivity	Reported	No frequency information

Association Validation:

Using PubMed database: These associations were also checked for number of articles in PubMed central to affirm their existence.

- Varenicline and aggressiveness (31 articles in PubMed central on Dec 10, 2012)
- Varenicline and suicide attempt (10 articles in PubMed central on Dec 10, 2012)
- Varenicline and suicide ideation (3 articles in PubMed central on Dec 10, 2012)
- Warfarin with bleeding (5052 articles in PubMed central on Dec 10, 2012)
- Carbamazepine and Stevens Johnson Syndrome (258 articles in PubMed central on Dec 10, 2012)

- Varenicline and depression (372 articles in PubMed central on Dec 10, 2012)
- Heparin and thrombocytopenia (3279 articles in PubMed central on Dec 10, 2012)
- Carbamazepine and Toxic Epidermal Necrolysis (179 articles in PubMed central on Dec 10, 2012)
- Leflunomide and Peripheral Neuropathy (41 articles in PubMed central on Dec 10, 2012)
- Simvastatin and Rhabdomyolysis (501 articles in PubMed central on Dec 10, 2012)
- Clozapine and Agranulocytosis (389 articles in PubMed central on Dec 10, 2012)
- Phenytoin and Stevens Johnson Syndrome (218 articles in PubMed central on Dec 10, 2012)
- Abacavir and Hypersensitivity (369 articles in PubMed central on Dec 10, 2012)
- Didanosine and Pancreatitis (195 articles in PubMed central on Dec 10, 2012).

There is scientific consensus that signals of disproportionate reporting identified with statistical methods must be considered with a medical judgment. The ADRs which are considered to warrant further evaluation should always be systematically medically assessed.

For a causal relationship to be established between a drug and an ADR, the strength of the clinical diagnosis must be assessed by studying the individual reports or by the use of controlled studies. Observational, interventionist, or mechanistic studies can then be used to test hypotheses of causality, and attempt to quantify the risk.

“In the current regulatory environment, where efficacious drugs are brought to market as soon as possible, post-marketing drug surveillance (PMS) has become increasingly important in order to characterize cost-effectiveness and harm in real practice [11].

V. ACKNOWLEDGEMENTS

Author wishes to thank Ms. Anagha Komawar and Mr. S. Suresh Faculty at International Institute of Information Technology, Pune for their continuous support and guidance.

REFERENCES

- [1] Egberts, A. C. G., Meyboom, R. H. B., & van Puijenbroek, E. P. (2002). Use of measures of disproportionality in pharmacovigilance: three Dutch examples. *Drug safety: an international journal of medical toxicology and drug experience*, 25(6).
- [2] Bender, A., Scheiber, J., Glick, M., Davies, J. W., Azzaoui, K., Hamon, J., Urban, L., et al. (2007). Analysis of pharmacology data and the prediction of adverse drug reactions and off-target effects from chemical structure. *ChemMedChem*, 2(6), 861-73.
- [3] European Medicines Agency. (2008). Guideline on the use of statistical signal detection methods in Eduravigilance data analysis system. *Risk Management*, (June), 1-22.
- [4] Bate, A., Lindquist, M., & Edwards, I. R. (2008). The application of knowledge discovery in databases to post-marketing drug safety : example of the WHO database. *Fundamental & Clinical Pharmacology*, 22, 127-140.
- [5] Gaval, D. K., Kulkarni, K. S., Kumar, A., & Chakraborty, B. S. (2009). Therapeutic class-specific signal detection of bradycardia associated with propranolol hydrochloride. *Indian journal of pharmacology*, 41(4), 162-6.
- [6] Hastie et. al. (2001). *The elements of statistical learning – Data mining, Inference and Prediction*. New York: Springer Series in Statistics.
- [7] Hauben, M., Patadia, V., Gerrits, C., Walsh, L., & Reich, L. (2005). Data mining in pharmacovigilance: the need for a balanced perspective. *Drug safety : an international journal of medical toxicology and drug experience*, 28(10), 835-42.
- [8] Roux, E., Thiessard, F., Fourier, A., Bégaud, B., & Tubert-Bitter, P. (2005). Evaluation of statistical association measures for the automatic signal generation in pharmacovigilance. *IEEE transactions on information technology in biomedicine : a publication of the IEEE Engineering in Medicine and Biology Society*, 9(4), 518-27.
- [9] van Puijenbroek, E. P., van Grootheest, K., Diemont, W. L., Leufkens, H. G., & Egberts, a C. (2001). Determinants of signal selection in a spontaneous reporting system for adverse drug reactions. *British journal of clinical pharmacology*, 52(5), 579-86.
- [10] Wilson, A. M., Thabane, L., & Holbrook, A. (2003). Application of data mining techniques in pharmacovigilance. *British Journal of Clinical Pharmacology*, 57(2), 127-134.
- [11] Kuhn M, Campillos M, Letunic I, Jensen LJ, Bork P. (2010) A side effect resource to capture phenotypic effects of drugs. *Mol Syst Biol*;6:343. Epub 2010 Jan 19.