A Data Mining Approach for the Identification of Adverse Drug Events (ADE) resulting from Drug-Drug Interactions (DDI) to improve pharmacovigilance

Thesis submitted to

COCHIN UNIVERSITY OF SCIENCE AND TECHNOLOGY

In partial fulfillment of the requirements for the degree of

Doctor of Philosophy

By

Sindhu M.S.
Reg: No. 2768

Under the guidance of
Dr. B. Kannan

DEPARTMENT OF COMPUTER APPLICATIONS
COCHIN UNIVERSITY OF SCIENCE AND TECHNOLOGY
COCHIN 682 022
Kerala

July 2013

Dedicated to my Beloved Parents .....

DECLARATION

*I hereby declare that the present work entitled "**A Data Mining Approach for the Identification of Adverse Drug Events (ADE) resulting from Drug-Drug Interactions (DDI) to improve pharmacovigilance**" is based on the original work done by me under the guidance of Dr. B. Kannan, Associate Professor, Department of Computer Applications, Cochin University of Science and Technology, Kochi-22 and the work does not form part of any dissertation submitted for the award of any degree, diploma, associateship or any other title or recognition from any University or Institution.*

Kochi -22 ***Sindhu M.S***

Date: 23/07/2013

Dr. B. Kannan (Supervising Guide)
Associate Professor
Dept of Computer Applications
Cochin University of Science and Technology

Certificate

*Certified that the work presented in this thesis entitled "**A Data Mining Approach for the Identification of Adverse Drug Events (ADE) resulting from Drug-Drug Interactions (DDI) to improve pharmacovigilance**" is based on the bona fide research work done by Sindhu M.S. under my guidance in the Department of Computer Applications, Cochin University of Science and Technology, Kochi -22 and has not been included in any other thesis submitted previously for the award of any degree.*

Kochi- 22

Date: 23/07/2013 **Dr.B. Kannan**

Dr. B. Kannan (Supervising Guide)
Associate Professor
Dept of Computer Applications
Cochin University of Science and Technology

Certificate

This is to certify that all the relevant corrections and modifications suggested by the audience during the Pre-synopsis seminar and recommended by the Doctoral Committee of the candidate have been incorporated in the thesis.

Kochi- 22

Date: 23/07/2013

Dr. B.Kannan

Acknowledgements

First, I would like to thank my guide and mentor, Dr. B. Kannan, Associate Professor, Department of Computer Applications, Cochin University of Science and Technology for his invaluable directions and tutelage throughout my research. His patience and motivation enabled me to a great extent to complete this thesis.

I now express my sincere thanks to Dr. K.V. Pramod, Associate Professor, Cochin University of Science and Technology for sparing his valuable time for discussions and encouraging the persistence in me.

I would like to thank Mr. Richard W. Meyer, Dean and Director, Georgia Institute of Technology, Atlanta, United States, for providing me the permission to access resources necessary for my research, without which the completion of this thesis would not have been possible.

I am also thankful to other faculty members, office staff, and technical staff of the Department of Computer Applications who helped me during various stages of my research.

I express my sincere gratitude to my parents for their unconditional love and support in all my endeavors. I learned to aim high and zest to achieve from them. To them and only to them, I dedicate my thesis.

A special hug to my beloved daughter Prathibha for the compromise she made for me while I was involved in my research. And now I express my gratitude to my husband for his love, support and persistent encouragement without which I could not have completed this thesis.

Finally I express my worship and gratitude to Lord Almighty whose light, guidance and immeasurable power guided me throughout my life.

Sindhu M.S.

ABSTRACT

Adverse drug events are unexpected and unwanted side effects caused by drugs when administered together to treat the same disease or different health conditions. When multiple drugs interact, they alter the desired level of activity and induce reactions that are unpredictable based on the side effects of the drugs observed when administered individually. These drug-drug interactions (DDIs) are often fatal; they are associated with increased hospital admission rates, treatment failures and avoidable medical complications. DDIs are so common even in paediatric hospitals. In addition, the elderly are more likely to be prescribed multiple drugs, increasing their susceptibility to DDIs. All in all, DDIs cause unnecessary patient exposure and are responsible for the withdrawal of drugs from the market, considerably increasing research overheads. Both the Food and Drug Administration (FDA) and the World Health Organization (WHO) have identified DDIs as an important safety concern and maintain databases to gather information regarding side effects caused by various medications. These databases contain drugs and their side effects, but their size makes it difficult for healthcare professionals to find timely information to diminish the patients' risk. It then stands to reason that a more efficient and effective method should be developed to determine potential DDIs.

In the current study, epidemiology study is done by means of literature survey in groups identified to be at higher potential for DDIs as well as in other cases to explore patterns of DDIs and the factors affecting them. The structure of the FDA Adverse Event Reporting System (FAERS) database is studied and analyzed in detail to identify issues and challenges in data mining the drug-drug interactions. The necessary pre-processing algorithms are developed based on the analysis and the Apriori algorithm is modified to suit the process. Finally, the modules are integrated into a tool to

identify DDIs. The results are compared using standard drug interaction database for validation. 31% of the associations obtained were identified to be new and the match with existing interactions was 69%. This match clearly indicates the validity of the methodology and its applicability to similar databases. Formulation of the results using the generic names expanded the relevance of the results to a global scale. The global applicability helps the health care professionals worldwide to observe caution during various stages of drug administration thus considerably enhancing pharmacovigilance.

Contents

List of Tables

List of Figures

Chapter 1
INTRODUCTION

Contents

A drug-drug interaction (DDI) is a change in the effect of a drug when administered with another drug or group of drugs. This interaction may be synergistic, increasing the effect of the drug, or antagonistic, decreasing the effects of any of the drugs or even causing an adverse effect that is not usual with either of the drugs. Adverse Drug Reaction (ADR) is the unexpected or unwanted effect resulting from the use of a therapeutic drug or the use of two or more drugs. The occurrence of an ADR is called an Adverse Event (AE).

1.1 Motivation

Drug interactions are common and cause increased hospital admission rates, treatment failure, avoidable medical complications, and even deaths [1]. The occurrence of drug interactions is reported to be a permanent cause of risk even in paediatric hospitals [2]. The incidence of potential drug-drug interactions is close to 40% in patients taking 5 drugs and 80% in patients taking 7 or more medications [3]. Drug adverse reactions cause up to 5%

hospital admissions, 28% emergency room visits, and 5% hospital deaths [4], [5], [6].

The elderly are more likely to be prescribed multiple drugs, and are therefore at a higher risk of DDIs [7]. This has been verified by several studies. Due to the number of drugs needed to treat comorbidities in the elderly and the strong correlation between increasing age and the number of drugs prescribed, the potential for drug-drug interactions is found to be higher in the elderly compared to other demographics [8], [9], [10], [11]. According to Leidholm, elderly patients experience DDIs at a higher frequency than their younger counterparts [12]. Additionally, almost 5% of hospitalizations among the elderly are due to DDIs [13]. This is of immense importance because senior citizens aged 65 and above account for approximately 15% of the western population and receive almost one-third of all drug prescriptions [14]. Clearly, improvement in drug safety will be invaluable in terms of morbidity and mortality.

In addition, advancement in drug safety through the identification of DDIs will have positive economic consequences [15]. Almost 10% of general hospital admissions are caused by DDIs and are associated with longer ICU stay [16], [17]. DDIs cause 1, 95,000 hospital admissions yearly in the U.S. [18]. All in all, ADRs impose a huge financial burden with an estimated cost of $16,000 per hospitalization [13], [19], [20]. Some adverse events are life threatening and cause withdrawal of popular medications like Siruvudine, Astemizole etc. from the market, leading to futile research overheads [21].

The Food and Drug Administration (FDA), the U.S agency responsible for regulating the safety of medications, medicinal supplements,

and food items, reckons drug-drug interaction as a critical factor in the benefit-risk assessment of a drug during development and regulatory review, and has created an online database, the FDA Adverse Event Reporting System (FAERS), which contains worldwide adverse event data [22]. The number of adverse event reports received and processed by the FDA increases marginally every year as indicated by Figure 1-1 [23].

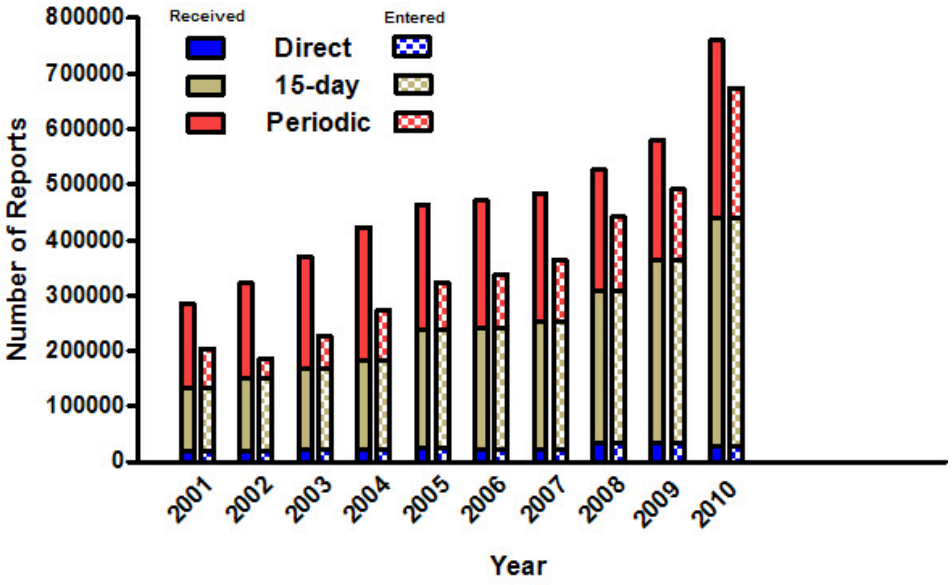


Figure 1-1 Yearly Adverse Event Reports received by FDA

Figure 1-1 illustrates the number of reports received (solid bars) and entered (checkered bars) into FAERS by type of report since the year 2001 until the end of 2010.

Figure 1-2 shows that polypharmacy has been increasing among patients who take prescription drugs, multiplying the gravity of the situation [24].

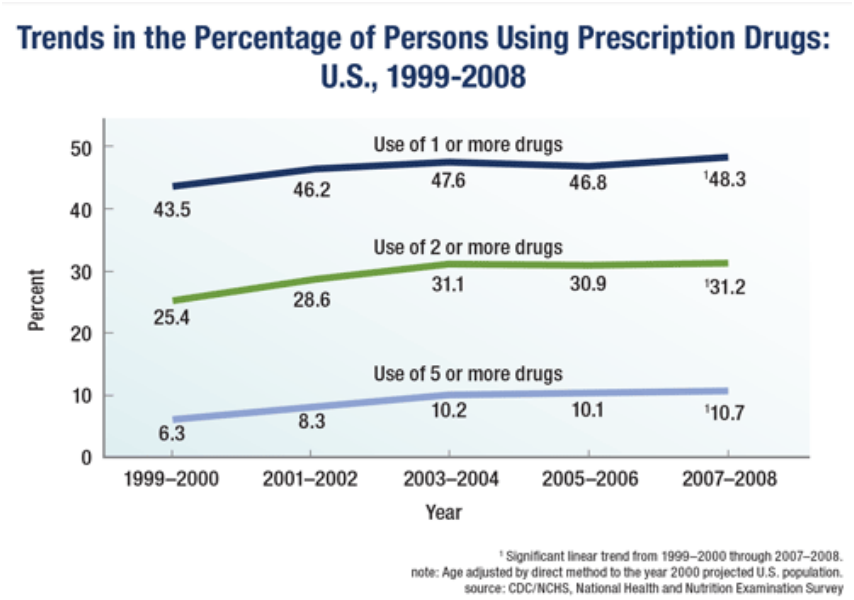


Figure 1-2 Trends in Prescription drug usage

As more drugs are introduced to the market, the importance of excavating DDIs increases.

1.2 Problem Statement

Drug-drug interactions (DDIs) are very common and cause hospitalizations, emergency visits, and even deaths [1]. DDIs cause 1,95,000 hospital admissions every year in the U.S. with an estimated cost of $16,000 per hospitalization, posing huge treatment costs for the elderly as their comorbidities are treated with concomitant drugs, leading to higher chance of DDIs [7], [13], [18], [19], [20]. Research overhead is also accrued, DDIs being the major cause of withdrawal of marketed drugs [21].

Present Adverse Event (AE) databases contain AEs of only individual drugs and drug interaction databases contain data from small cohort studies or interaction studies of a small group of drugs administered for the same

symptom. Present methods analyze single drug AEs from databases or conduct text search of biomedical literature, but both fail to find novel potential DDIs. The set of textual patterns proposed for text mining were found to be inadequate to identify many interactions [25]. The cross matching of possibly every drug that has an AE with every other drug must be done to minimize patient exposure. The FDA Adverse Event Reporting System (FAERS) database is a repository of mandatory AEs collected from manufacturers and voluntary reports from health care professionals and consumers worldwide, and can be excavated to find potential DDIs. Association rule mining methodology, which is an established method suitable for finding associations between multiple items in large databases, is used to mine new drug-drug interactions with an appropriate implementation of the Apriori algorithm. Data pre-processing and data mining algorithms are developed for the current problem. The modules developed are integrated into a tool for DDI identification. The rules thus formulated are compared by a pharmacological expert with standard database for validation.

1.3 Types of Adverse Drug Reactions

Adverse Drug Reaction often refers to the side effects caused by a drug and can be categorized based on the mechanism by which they are caused [26]. Their knowledge and understanding is necessary for practitioners to monitor drug therapy and ADR detection.

1.3.1 Type A ADRs

These are caused by the normal pharmacological effect and effects of a substance. Usually, these are identified in the manufacturing phase. Adequate

labelling and other information can be used to make these evident as being predictable and dose related [27].

Example 1: Respiratory depression with opioids or bleeding with Warfarin.

Type A reactions also include those effects that are indirectly related to the desired pharmacological effect of the concerned drug.

Example 2: Dry mouth associated with tricyclic antidepressants.

1.3.2 Type B ADRs

These are relatively uncommon, unpredictable and non-dose related. These are responses not expected from the known actions of a drug and can only be discovered for the first time after the drug becomes available for general use. Idiosyncratic and immunological categories like these are detected during the pharmacovigilance phase of the drug's shelf life using Adverse Event reporting databases called Spontaneous Reporting System (SRS) like FAERS or Yellow Card [28].

Example: Anaphylaxis with penicillin or skin rashes with antibiotics.

1.3.3 Type C ADRs

Cumulative toxic effect of a drug when used over time leads to this type of ADRs. In other words, these are continuous reactions that persist for relatively long time intervals. In this case, AEs increase gradually.

Example: Osteonecrosis of the jaw with bisphosphonates.

The SRS contains Type C ADR data as it collects data for the life time of the drug.

1.3.4 Type D ADRs

These rare ADRs become evident only sometime after treatment has been completed. These reactions are delayed, making them difficult to detect.

Example: Leucopoenia caused by the medication Lomustine which can take up to six weeks to occur after the dosage has been discontinued.

1.3.5 Type E ADRs

Abrupt discontinuation of a drug which has been used long term leads to Type E ADRs. These ADRs are called 'end of use' reactions, caused by sudden withdrawal of a medicine.

Example: Insomnia, anxiety and perceptual disturbances due to withdrawal of Benzodiazepines.

In conclusion, Type A are remedied premarketing and Type D are difficult to detect. Type E can be easily detected as discontinuation and consequences are evident both to the patient and physician. Type B and Type C are difficult to detect and need to be analyzed with the current database resources available. Type B becomes evident in inpatient and outpatient settings and type C is usually detected in outpatient settings, due to their nature. The data collected in SRS contains signals of both Type B and Type C drug reactions.

1.4 The Concept of Drug-Drug Interactions

Drugs interact with each other for a variety of reasons. Based on the underlying mechanism of interactions, the DDIs can be divided into 4 types as shown in Figure 1-3 [29].

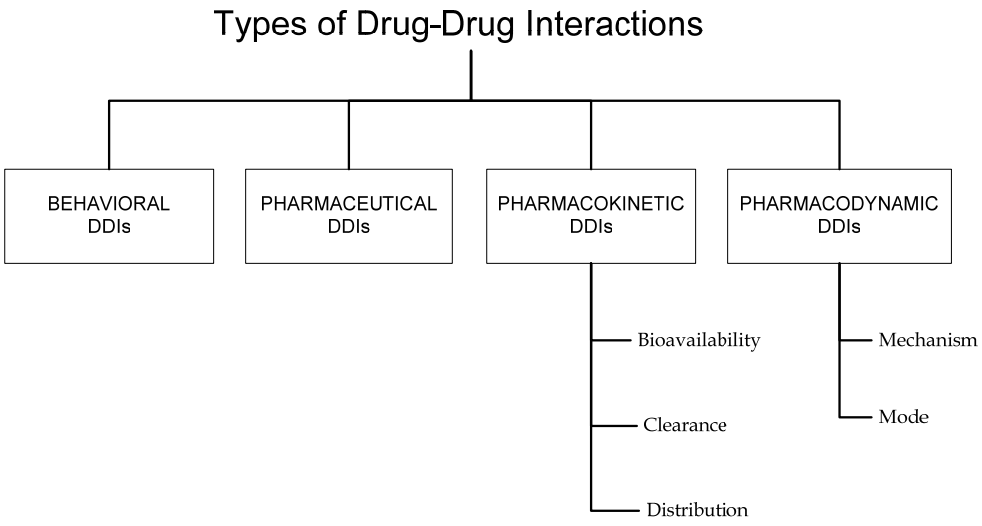


Figure 1-3 Types of Drug-Drug Interactions

1.4.1 Behavioral DDIs

Behavioral drug-drug interactions are initiated when a drug changes the patient's behaviour towards compliance with another drug. For example, a patient taking an anti-depressant may show better compliance with the drug as symptoms improve [30].

1.4.2 Pharmaceutical DDIs

Pharmaceutical drug-drug interactions usually occur in lab. These interactions occur when the formulation of one drug is altered by another drug prior to administration. The medium used to prepare syrup as in the case of Sulfanilamide and diethelene glycol or precipitating sodium thiopentone and vecuronium in an intravenous giving set can be taken as examples [31].

1.4.3 Pharmacokinetic DDIs

Pharmacokinetic drug-drug interaction occurs when systemic concentration of one drug is changed by another drug, i.e. alteration of the duration and strength of a drug at the action site. This can be further classified as being due to (a) bioavailability caused by absorption variation, (b) clearance due to changed metabolism or excretion of active drug, or (c) distribution caused by change of cell membrane transport to the site of action. For example, in the concomitant administration with Sulbactam, Ampicillin provides better antibiotic coverage than when it is administered alone [32].

1.4.4 Pharmacodynamic DDIs

Pharmacodynamic drug-drug interactions occur when the interacting drugs have either additive effect (the overall effect is multiplied) or negative and opposing effects (the overall effect is diminished or cancelled out). The pharmacodynamic drug-drug interactions can be due to (a) mechanism (molecular signal) or (b) mode (physiological effect). The positive effect can be exemplified by the use of anti-hypertensive or the use of two different antibiotics to combat drug resistant bacteria. The negative effect is exemplified by the opposing effect of a reduced bronchodilator beta with a non-selective beta blocker [33].

Behavioural DDIs do not cause any serious patient outcomes. Due to increased awareness of pharmacovigilance, pharmaceutical DDIs are now remedied in premarketing stage itself and have minimal patient exposure. The remaining categories namely pharmacodynamic and pharmacokinetic DDIs need to be analysed and monitored.

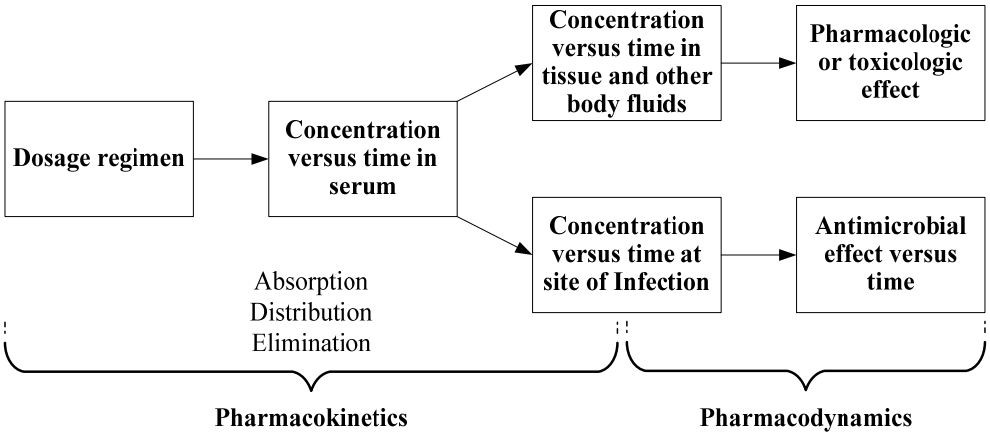


Figure 1-4 Pharmacokinetics and Pharmacodynamics

The stages in which pharmacokinetics and pharmacodynamics affects (Figure 1-4) are different and therefore the methodology of studies of the drugs involving the two also differ.

1.5 Identification and Importance of DDIs in Pharmacovigilance

The World Health Organization (WHO) defines pharmacovigilance as the pharmacological science relating to the detection, assessment, understanding, and prevention of adverse effects, particularly long term and short term adverse effects in medicine [34].

1.5.1 Objectives of Pharmacovigilance

The objectives of pharmacovigilance, defined by WHO are summarized below [35]:

- Improvement of patient care and safety in relation to the use of medicines and all medical and paramedical interventions.

- Improvement of public health and safety in relation to the use of medicines.
- Detection of problems related to the use of medicines and communication of the findings in a timely manner.
- Contribution to the assessment of benefit, harm, and effectiveness and risk of medicines, leading to the prevention of harm and maximization of benefit.
- Encouragement of safe, rational and more effective (including cost-effective) use of medicines.
- Promotion of the understanding, education, and clinical training in pharmacovigilance and its effective communication to the public.

1.5.2 Pharmacovigilance - Activities and Partnership

The activities in pharmacovigilance fall within three realms (Figure 1-5). They are regulatory, industry, and academia [36].

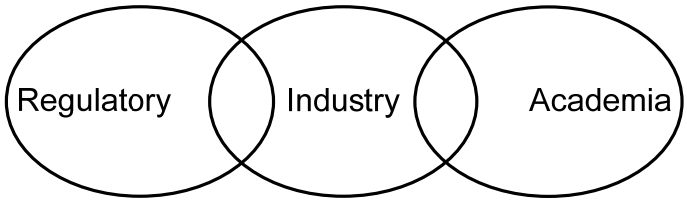


Figure 1-5 The three realms of pharmacovigilance

a. Regulatory Authorities

Pharmacovigilance is defined by regulatory authorities present in national and, increasingly, in the international level. These authorities play a crucial role in pharmacovigilance and regulatory pharmacovigilance is defined

as “the process of evaluating and improving the safety of medicines” by Waller, Caulson and Wood [37]. They demarcated the responsibilities of national governments in drug safety monitoring and these responsibilities were accepted by many nations in consequence to the following incident:

Sulfanilamide (Prontosil) used for treating streptococcus infection since 1932, was introduced in a syrup form with diethelene glycol as solvent, launched in 1937 without adequate safety evaluation. This new drug caused the death of 105 patients (34 children and 71 adults) in the U.S. and diethylene glycol was identified as the cause [31]. Consequently, in 1938, the American Congress approved the Food, Drug and Cosmetic Act mandating manufacturers to show scientific evidences before sales. Most nations have regulatory authorities for drug safety monitoring initiated by national governments. In India, an ADR monitoring system consisting of 12 regional centres was proposed in 1986. There are numerous drug brands in India. As per current statistics, there are more than 6000 licensed manufacturers and over 60,000 branded formulations. India is an emerging clinical trial hub and the fourth largest manufacturer of drugs in the world. Previously, data collected in western countries was being used in drug safety. But since the time lag between the launch of a drug and its availability in India has been considerably reduced and more and more new drugs are being manufactured in India, the need for better drug safety measures became imminent. The twelve centres mentioned previously collect information and are instrumental in enforcing the standards in India.

b. Pharmaceutical Industry

Drug manufacturers play a key role in pharmacovigilance. The sector interests in epidemiology, causality assessment, and risk management. Many countries require reports of drug adverse events along with other details to be submitted to the authorities before product marketing.

c. Academia: the Science of Pharmacovigilance

Pharmacovigilance is a scientific discipline which studies and monitors drug safety [38]. Individual and population level safety of clinical drugs is studied based on usage and harmful effects. The knowledge thus generated is applied to clinical practice and promotes safer drug usage. Universities and other academic settings have developed, used, and evaluated several statistical methodologies for pharmacovigilance [39], [40].

1.5.3 Methods of Achieving Pharmacovigilance

Pharmacovigilance uses the following methods to achieve its objectives [36].

a. Pre-marketing Studies

Pre-marketing studies are done via different stages of clinical trials, but, these have been found inadequate as it is not possible to detect rare ADRs owing to limited patient participation. The short duration of clinical trials makes it inefficient in detecting ADRs with long latency. Also, the characteristic of the population in which the trials are conducted may not correspond to the characteristic of the population where it is later used, making the extrapolation of the results difficult. This is increasingly applicable to the elderly, women, and minorities. To conclude, careful post-marketing studies

are necessary for the study of rare ADRs, ADRs of long latency and ADRs focused in specific populations.

b. Post-marketing Studies

The Post marketing studies can be either descriptive or analytical. The descriptive studies formulate hypothesis to describe occurrence of events by collecting and studying toxicity and efficacy data. Analytical studies test hypothesis to find the seriousness of the AEs based on whether the connection between drugs and AEs are causal. Descriptive studies form important part of post-marketing surveillance as they are capable of generating hypotheses which can start the analytical studies. Spontaneous Reporting and Intensive Monitoring are the two different types of descriptive studies which are being discussed in the subsequent sections.

c. Spontaneous Reporting System (SRS)

This is a major method of collecting post-marketing data about drug safety. The Spontaneous Reporting Systems primarily function as the early detectors of new, rare and serious ADR 'signals'. Through the SRS, physicians, pharmacists, and patients are able to report suspected ADRs to pharmacovigilance centres, when new ADR 'signals' are detected. The pharmacovigilance centre collects and analyses these reports and informs the stakeholders regarding the potential risk of adverse events. Pharmaceutical industry uses the Spontaneous Reporting Systems for the collection of information of its drugs. The SRS enables the monitoring of all the drugs in the market throughout the life cycle at a low cost. The most important criticism about the SRS is that the authorities do not have control over the voluntary reporting and cannot enforce voluntary reporting which leads to

underreporting. Enthusiastic consumer reporting and the subsequent problem of duplicate reports is another factor the SRS needs to take care of. The FAERS database used in the present problem, UK Yellow Card System and the WHO Uppsala Monitoring (VigiBase) are all examples of Spontaneous Reporting Systems.

d. Intensive Monitoring

Intensive monitoring systems (Intensive Medicines Monitoring Programme of New Zealand and Prescription Event Monitoring of UK) use the prescription data to identify users of certain drugs. The prescriber of the drug is asked to report the AEs occurring during the drug usage while monitoring. The data thus collected is used to analyse new signals like prescription event and pro-active safety surveillance. The methods differ from SRS as it is based on a small observational cohort and only monitors selected drugs during a certain period of time. But the real world data without over-reporting and underreporting as in the case of SRS is collected by the methodology. Intensive monitoring is event based and is capable of identifying signals for events that are not mere suspects. But the proportion of reports that go without reporting to doctors is unknown.

e. Database Studies

In the UK, all patient care is coordinated by the general practitioner (GP), and the data from this source provides the complete details of a patient, the patients' illness and treatment. The GPs, being members of the General Practice Research Database, collect data from about three million patients, representing the general UK population in terms of sex, age, and geographic distribution. The collected data include demographic information, routine care,

hospitalization and emergency care records, date, location of the event etc. Prescription data include date of prescription, formulation strength, quantity and dosing instructions, indications for new prescriptions, and withdrawal events of a drug treatment. But all the data represent only 5% of the UK population.

f. Quantitative Signal Detection

Signal detection is defined as the search of information on a causal relationship between a drug and its AE, of which the reporting is unknown or not documented previously. This is an active area of clinical research informatics where data analysis methods involving signal detection activities are being applied. Signal detection was previously done based on case by case report analysis but has paved way for data mining techniques in the recent years. The methodology of the techniques differs, but all the algorithms express the extent to which the number of observed cases differs from the number of expected cases. The Proportional Reporting Ratio, PRR, is an analytic approach currently being used to compare proportion of reports for a specific ADR reported for a drug with the proportion for that ADR in all other drugs.

Considering the general nature of the data available and its nature of data collection over the life cycle of a product, the use of FAERS in the current study can be justified.

1.5.4 Pharmacovigilance Synergistics

Shuka et al. describes the parties in effective pharmacovigilance as in Figure 1-6 [41]. Efficient pharmacovigilance can only be achieved by an

effective two way communication. The consumers or patients and relatives or public will have to report the Adverse Events. Physicians and health care professionals are in the middle, reporting the events and have the responsibility for implementing the decisions as per the policy maker (regulator) regulations. Press or media have the responsibility to direct the attention of the intelligentsia towards the problem, so that remedial research can be started at the earliest. The reporting also makes physicians and consumers aware of the situation. Policy makers are usually national governments who receive, analyse and enforce regulations through information collected by means of databases like Spontaneous Reporting Systems.

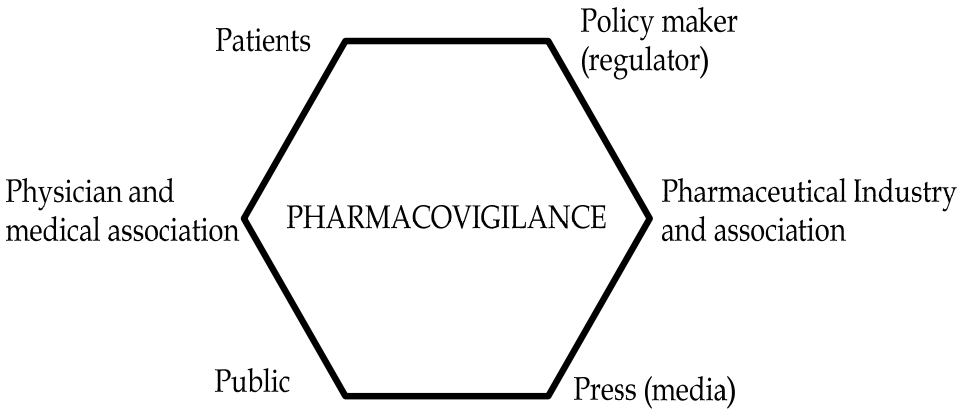


Figure 1-6 Partners of Pharmacovigilance Synergistics

The communication and co-ordination between every partner from regulator to consumer is necessary to achieve effective pharmacovigilance.

1.5.5 Pharmacovigilance: Efforts and Challenges

National pharmacovigilance centres function as an international network coordinated by WHO Program for International Drug Monitoring.

Significant achievement has been made in improving the activities, support, and recognition of individual national pharmacovigilance centres. The programme plays a key role in communication and acts as a training centre and clearing house for information on the safety of medicines. The centre, located in Uppsala, Sweden, manages the international database of adverse event reports received from national centres [42]. The darker regions in Figure 1-7 show the members of the WHO centre for International Drug Monitoring.

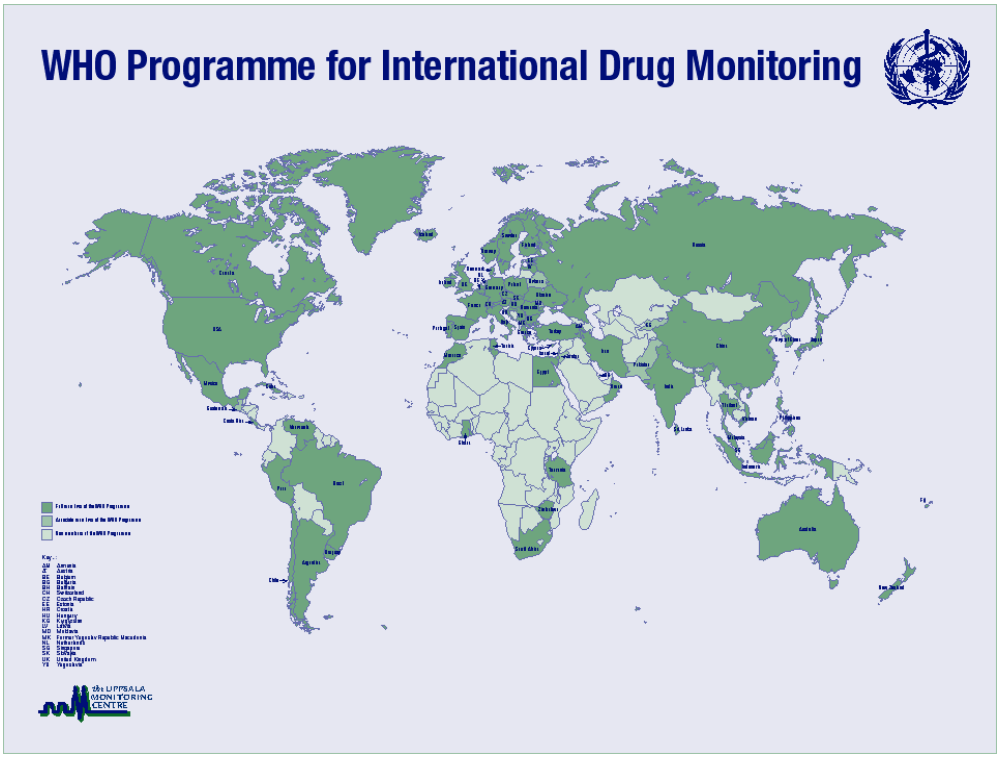


Figure 1-7 The WHO members of International Drug Monitoring

Figure 1-8 illustrates the important realms of challenges in pharmacovigilance, identified as the huge volume of data to be analyzed,

unavailability of accurate medical information, obsolete data mining methodologies, and regulatory reporting.

Figure 1-8 Current Pharmacovigilance Challenges

1.6 Objectives of the Present Study

The objectives of the present study can be summarized as follows:

1. Conduct epidemiology survey to study ADRs and identify factors affecting DDIs by means of literature published in the subject area.
2. Find a suitable method considering the large data volume and the type of result (association rules) to be obtained; select a suitable implementation method.
3. Establish the feasibility of the methodology and its implementation.

4. Increase the efficiency of the data mining techniques selected to suit the problem under consideration i.e., large volume of data and finding multiple drug Adverse Event associations.
5. Find novel DDIs by analysing the AEs so that physicians and pharmacologists can exercise caution during prescription and administration of medications, thus improving pharmacovigilance.
6. Create an integrated DDI mining tool.
7. Validate the results by cross matching with known ADEs using the standard database with the help of pharmacology expert.

1.7 Scope of the Thesis

The applicability and scope of the thesis is discussed in this section:

1. The Adverse Event database is one of the core components for DDI identification. The FAERS database used for the present study, though maintained by FDA, contains worldwide data received from manufacturers, healthcare professionals and consumers. Figure 1-9 demonstrates the number of domestic (U.S.) and foreign (rest of the world) reports in FAERS database since the year 2001 until the end of 2010. The data of the year 2012 released by the FDA contains reports from 168 countries including 1709 reports from India.
2. Pre-processing algorithms are developed to remove duplicate and irrelevant data.

3. Algorithm is developed for finding drug-drug interactions by comparing and analysing the Adverse Event data to be used by physicians and pharmacologists.
4. The quality of the results is improved using proportionality ratios as per WHO and FDA recommendations.
5. The result is generated using the generic names to enable the applicability worldwide.
6. The different modules are integrated into a DDI mining tool.

Figure 1-9 shows that almost 50% of the reports received by the FDA are from outside the U.S.

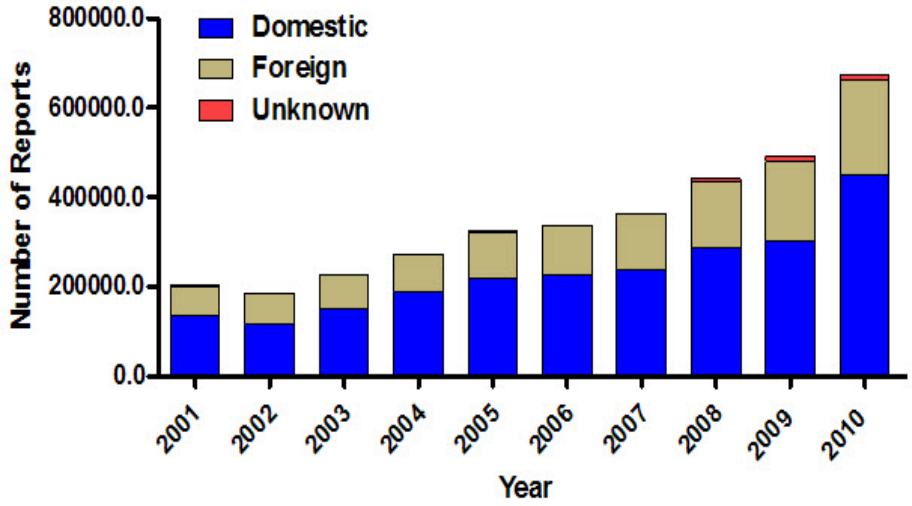


Figure 1-9 Domestic and foreign reports received by FAERS

1.8 Method of Mining the Data: the Association Rule Mining

Association rule mining is a data mining methodology invented to combat the problem of market basket analysis. It was first introduced by

Agarwal et al. in 1993 and is a well-researched and important data mining methodology [43], [44]. The method is capable of finding correlations, frequent patterns, or associations of interest from very large transactional and data repositories. Association Rule Mining is widely used in telecommunication networks, market and risk management, inventory control etc.

The data mining methodology finds answers to questions like the items bought together often, or the items purchased in sequence to one another. For example, bread and butter are bought together and anti-viral software is bought after purchasing a computer. The analysis is used by business managers for store layout, discounting product using coupons, and cross marketing.

The method finds association rules satisfying a predefined support and confidence from a given database. The problem is divided into two sub problems. The first step finds itemsets whose occurrence exceeds a predefined threshold in the database. The second step generates association rules from these large itemsets with minimal confidence constraints. Support and confidence are two measures of rule interestingness. These reflect the usefulness and certainty of discovered rules respectively. Support measures how often items occur together as a percentage of total transactions and confidence measures how much a particular item is dependent on another.

Example: Consider the following association rule:

Computer →antivirus-software [support=2%, confidence=60%]

The left hand side of the equation is called the antecedent and right hand side is called the consequent. A support of 2% means that 2% of all the transactions under analysis indicate that computer and antivirus-software are

purchased together. Confidence of 60% indicates that 60% of customers purchased the two items together.

The present problem tries to find association between drugs and their Adverse Effects the same way like group of items being purchased together.

In the current problem, the Apriori algorithm is used for association rule mining. The Apriori algorithm is a proven technique to mine association rules and is often used together [45]-[53].

1.9 Contemporary Works

Currently, ADE databases contain thousands of drugs and adverse effects which make the statistical analysis difficult. The traditional method of manual case reviews by medical professionals became impossible as the database size grew larger and more complex. Moreover, most data mining algorithms explore only bivariate associations i.e., pairs involving one drug and one adverse event.

Example:

Lisinopril → cough.

But the current problem demands excavation of multiple associations like:

Aspirin + Warfarin → bleeding

Or

Azithromycin+ Fleroxacin → sleep disorder + abdominal pain

Most of the published works in biomedical and pharmacological research is based on specific drugs, diseases, or demographic groups. The text mining algorithms to find DDIs fail due to lack of generality and inadequacy of textual patterns used for mining the associations. Therefore, a new and more general technique needs to be developed.

1.10 Organization of the Thesis

The following chapters are organized as follows. In the next chapter, the epidemiology of DDIs is investigated in detail in an attempt to enable a better understanding of the patterns of DDIs and find the causative factors. Propriety of the methodology and its implementation are also discussed. Towards the end of the chapter, the need of identifying DDIs for pharmacovigilance, current research, and the improvements required are explored. In Chapter 3, details of the database, statistics of the data and the necessity of pre-processing are discussed. The chapter also elucidates the pre-processing performed before the process of data mining. Chapter 4 depicts the implementation of the association rule mining methodology using modified Apriori, and integrating the components into a DDI mining tool. Result Analysis is also done in this chapter. Lastly, Chapter 5 discusses the contributions of the thesis and future studies.

Chapter 2

LITERATURE REVIEW

Contents

2.1 Epidemiology of Drug-Drug Interactions

Epidemiology is a branch of medical science that deals with the patterns, distribution, and control of a disease in a defined population. In the current case studies, emphasis is given to oncology, HIV treatment, and drug-drug interactions in the elderly. ADRs are found to be higher with combination therapy than in monotherapy which indicates potential DDIs [54]. Rate of DDIs from other studies is also explored to understand rates and factors affecting interactions.

2.1.1 Case Studies

Epidemiology survey is conducted to find patterns of DDIs. The group is selected which were identified as cohorts where, comorbidities, or presence of two or more medical conditions in addition to an initial diagnosis, are treated.

a. Antiretroviral/HIV

HIV related morbidity and mortality have been considerably reduced by the triple combination antiretroviral therapy [55]. Antiretroviral therapy (ART) is widely used in HIV infection treatment and in majority of these cases,

polypharmacy is required for treatment as the patients will have comorbidities. A DDI study was conducted at New York City HIV Specialty Clinics with 550 patients who were prescribed an average of 6 medications [56]. Of these, 14% showed at least one DDI. Another study in a Swiss HIV cohort involving 1497 patients reported that 68% of patients had at least one DDI [57]. This involved mainly central nervous system drugs (49%), cardiovascular drugs (45%), and methadone (19%). Clinically significant DDIs of 41.2% were reported among 153 randomly selected patients in a 2006 antiretroviral therapy study [58]. A Kenyan study among 334 patients receiving ART reported a 33.5% DDI rate [59]. It is reported that patients aged 42 and above are prone to more DDIs [60]. Other factors posing high risk of DDIs by this study were presence of more than three comorbidities, administration of more than five non-ART agents, and treatment with more than four ART medications. Some other studies have stressed the need of therapeutic drug monitoring in ART [61]. Another study reiterated the fact that DDIs increased with age and taking ART along with other medications [62]. A study at the University of Liverpool has identified polypharmacy and age as key factors increasing DDIs [63]. The analysis indicates that the potential of DDIs increases with age and greatly depends on the number of medications.

b. Oncology

Psychotropics are often prescribed with other drugs to cancer patients due to pre-existing psychological disorders and treatment related anxieties [64]. The patients are also prescribed supportive medications along with cytotoxic drugs [65]. For example, patients in oncology clinics are usually treated for nausea and distress along with other symptoms [66]. The number of

drugs prescribed is higher in these cases, and so is the chance of DDIs. Studies in the field of cancer related nausea and distress report a 44.8% rate of DDI when considering 440 medicinal combinations. 250 potential drug interactions were identified among 115 patients, of which 31% showed at least one DDI, the most common interaction being Warfarin and Phenytoin [67]. A South Indian study among 75 patients indicated 213 drug interactions of which 6.1% were DDIs among anti-cancer drugs and 6.5% were between anti-cancer drugs and medications for other co-morbidities [68]. A study in metastatic breast cancer (MBC) found the patients to be under risk for drug interactions due to the heavily pre-treated nature of the disease and narrow therapeutic window [69]. Two oncology patients treated for DDIs died among the hospitalized patients for treatment of DDIs [70]. Analysis of the studies indicate that cancer patients have a higher chance of being prescribed a large number of drugs irrespective of age, and the higher the number of drugs, the higher the number of potential DDIs, which can be fatal.

c. Elderly Population

Patients in upper age groups will have comorbidities and are another group which demands DDI study. Studies indicate that hospital admissions in elderly patients have resulted from known DDIs, many of which can be avoided if identified [11]. Research has also shown that DDIs due to ADRs in elderly patients is highly significant (Table 2-1) [71], [72].

Table 2-1 Distribution of patients exposed to potential Drug-Drug Interactions, attendance in primary care [71]

Age	Number (%) of patients exposed to drug – drug interactions(n=521)			
	Major	Moderate	Minor	Total
45-59	37 (37.0)	134 (36.8)	19 (34.6)	193 (36.6)
60-69	35 (35.0)	118 (32.0)	22 (40.0)	176 (33.4)
70-79	18 (18.0)	79 (21.6)	10 (18.2)	108 (20.5)
80-94	10 (10.0)	35 (9.6)	4 (7.2)	50 (9.5)
All	100 (19.2)	366 (70.2)	55 (10.6)	521 (100.0)

A study among Brazilian elderly indicated that 26.5% of the elderly population included in the study was prescribed medications involving DDIs affecting 64.4% of the women and 50.75% of the men [73]. Patients of age 70 and older showed more DDI episodes as more medications were prescribed [74]. DDIs among cardiology patients also increased with age (Table 2-2) [75].

Table 2-2 DDI rate in cardiology based on age

Age (Years)	Patients: n (%)
$\leq$ 30	25 (6.25)
31-45	64 (16)
46-64	115 (28.75)
$\geq$ 65	196 (49)

Studies have shown that DDI increases with age even in outpatients (Table 2-3) [76].

Table 2-3 Annualized ambulatory visits involving clinically important DDIs

Age	Visit rate
< 25	0.44
25-44	0.59
45-64	6.07
65-74	38.45
> 74	70.12

An Iranian General Hospital study indicated increase in DDIs due to age and number of prescribed drugs [77]. Age related changes and polypharmacy have been identified as one of the six reasons for DDIs among the elderly (Figure 2-1) [78].

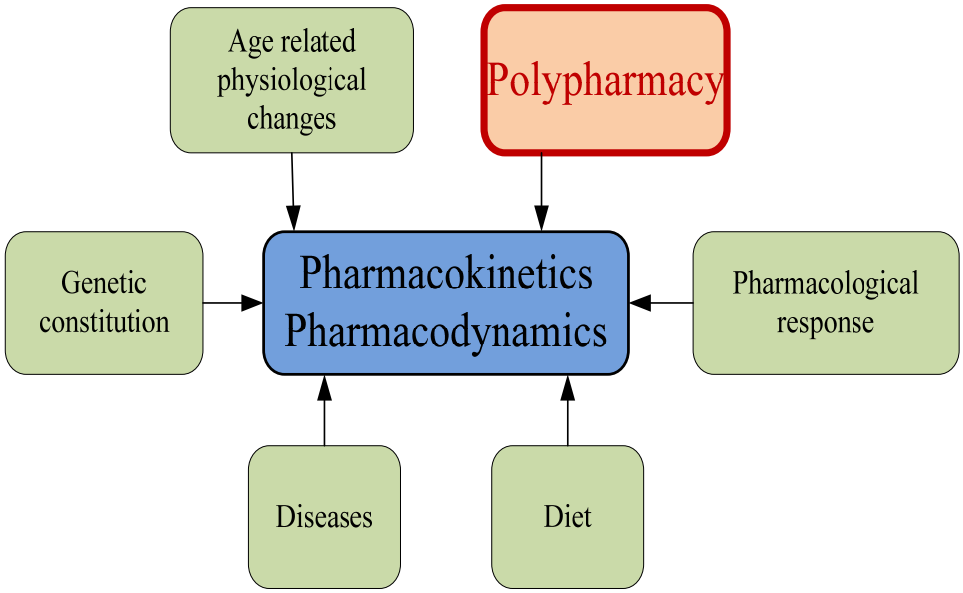


Figure 2-1 Factors affecting drug response and predisposing to DDIs in the elderly

There is little that science can do to age-related physiological changes, genetic constitution, pharmacological responses and diseases in terms of drug interactions. Diet can be controlled to avoid interactions. It is the polypharmacy factor or multiple drug usage that drug-drug interactions studies have to take care of.

Johnell did an extensive study of the effects of polypharmacy usage in the elderly. The study indicates a direct correlation between the number of drugs used and clinically relevant Type C and Type D DDIs (Figure 2-2, Figure 2-3) [79].

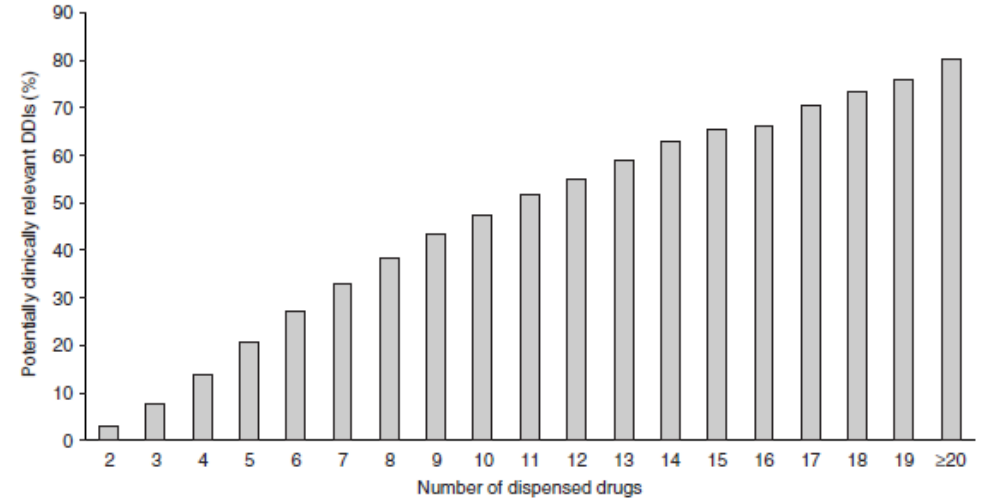


Figure 2-2 Number of drugs and clinically relevant Drug-Drug Interactions

Prevalence of potentially clinically relevant Type C DDIs as a function of number of dispensed drugs among 6,30,743 people aged 75 years and above from the Swedish Prescribed Drug Register (Figure 2-2) [79].

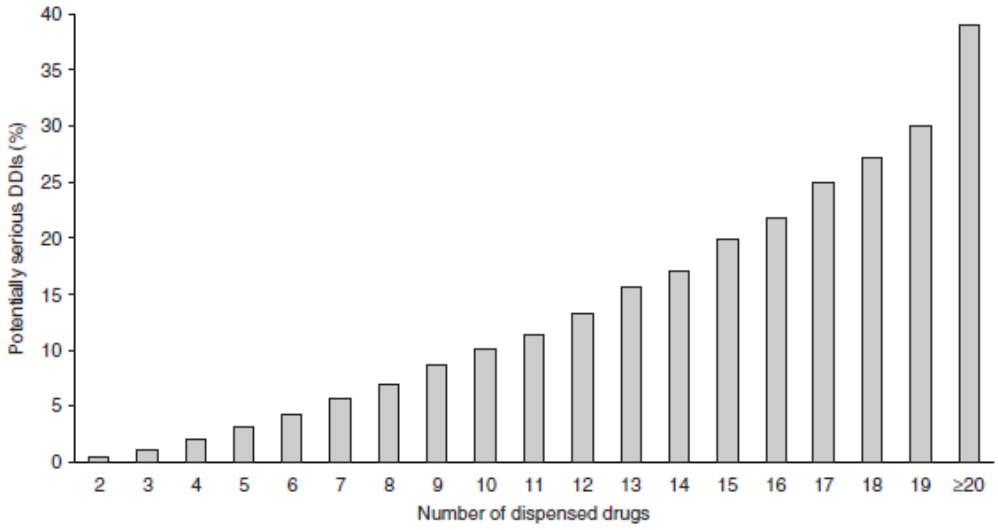


Figure 2-3 Serious Drug Interactions depending on number of drugs

Prevalence of potentially serious Type D DDIs as a function of number of dispensed drugs among 6,30,743 people aged 75 years and above from the Swedish Prescribed Drug Register (Figure 2-3) [79].

d. Statistics From Other Case Studies

The rate of drug interactions from other case studies (Table 2-4) clearly indicates a directly proportional relationship between DDI rate and cohort size. That is, larger the size of the population studied, the higher the number and types of drugs involved and the more likely it is for drugs to interact.

Table 2-4 Drug Interaction statistics from different case studies

No	Type	% of DDI	Number	Year	Reference
1	Antimicrobial drugs	39.5	226	2011	[80]
2	Cardiothoracic ICU	56.25	400	2010	[81]
3	Primary care	66.3	23733	2012	[74]
4	General	56	142295	2007	[82]
5	Inpatient	80	188020	2004	[83]

2.2 Contemporary Research

Microdosing has been used as a method to predict potential DDIs using volunteers [84]. The *in vivo* method is implemented by administering sub-pharmacological new chemical entity (NCE) micro doses to human volunteers. The subsequent pharmacokinetics is extrapolated for higher doses [85], [86]. This involves a risk element since reactions even in small doses can cause hazards. Regulatory authorities demand very less safety toxicology data before further studies prelude to market launch [87]. In addition, pharmacovigilance demands minimum patient exposure before product marketing.

A number of data mining algorithms are used to identify ADRs [88], [89], [90], [91]. These algorithms use SRS populations to interpret ADR rates.

The FDA drug labels have also been mined to find ADRs [92]. All these methods are useful in finding ADRs of individual drugs, but not DDIs. Also, these methods cannot predict new ADRs due to drug interactions. These applications analyse adverse reactions involving bivariate relations, i.e. a particular drug and its adverse event, but fail to excavate interactions between multiple drugs.

Pharmacokinetic computer programs are used to mine potential DDIs in treating infectious diseases [93]. Myopathy related DDIs are predicted and assessed using electronic medical records [94]. But all these methods are specific to a group of drugs or medical conditions and lack generality.

Whenever a drug interaction is reported, pharmaceutical and biomedical research is conducted on the issue and articles are published. These articles are identified as sources for finding DDIs. Text mining methods have also been implemented to identify DDIs from pharmacological documents and biomedical texts [95], [96], [97]. However these generally contain interactions of specific group of drugs or disease conditions. All these methods can explore existing, but not novel DDIs. Besides, the set of textual patterns proposed by pharmacologists, which forms the basis of textual searches, was found to be inadequate to identify many interactions [25]. As found in the survey of the epidemiology of DDIs in the previous section, studies concentrate on smaller cohorts, a particular disease, a number of interacting drugs administered concomitantly for a disease, or a particular age group. Biomedical and pharmacological documents also retain this property. It was also found that the number of DDIs increases with the number of medications administered together [74]. Moreover, the data in the previous section indicates higher DDI

rate for larger populations and hence requires a more general study (Table 2-4). All these shortcomings are to be overcome by a new method capable of identifying novel DDIs by cross matching a drug with every other drug.

2.3 Selection of the Methodology and Its Implementation

Association Rule Mining is reckoned as an efficient method for identifying relationships among variables in large databases. It was introduced by Agarwal et al. in 1994 and is based on the requirement to analyse large amounts of supermarket basket data, which inspired the invention [98]. The method has been used for numerous applications ever since. Association Rule mining is used for customer targeting , e-commerce, quality improvement of production process, Network Event Analysis, and analysis of spatial data sets [47]–[52]. Negative association rules can also be mined using the methodology to find items absent in transactions and hence can be used for efficient inventory management [53]. Association rules in large biomedical/biological databases can be mined using this method. HIV-Human Protein Interactions, gene ontology, and gene expression databases have been mined using this method in order to explore association rules [99], [100], [101], [102]. Association Rule Mining was applied to find ADEs in HIV drugs, hospital inpatient data and even in hospital infection control [45], [46], [103], [104]. Medical records of Chinese medicinal formulae have been searched for association rules for the prevention and treatment of breast cancer [105]. All these are very large databases that demand very high computational capability.

Frequent itemsets have been mined using the Apriori algorithm [106]. In the first step, the algorithm generates the frequent itemsets. In the subsequent step, the database is scanned to find support count of the

corresponding itemsets. Apriori algorithm has been identified as an efficient method for mining association rules and can also be used for multilevel association rule mining in large databases [107], [108]. Almost all the applications mentioned in the previous paragraph use the Apriori algorithm.

Systematic and random samplings are used for Association Rule Mining, using the Apriori and FP-growth algorithms, and analysis indicates that different sampling methods can be efficiently used and behave similarly in terms of accuracy [109].

The AIS, DHCP, and Partition algorithms have been compared to the Apriori algorithm and Apriori was determined to be the best among the four [110]. The number of association rules generated by the Apriori algorithm was larger for all confidence and support levels and the margin improved considerably as the number of transactions increased (Figure 2-4, Table 2-5).

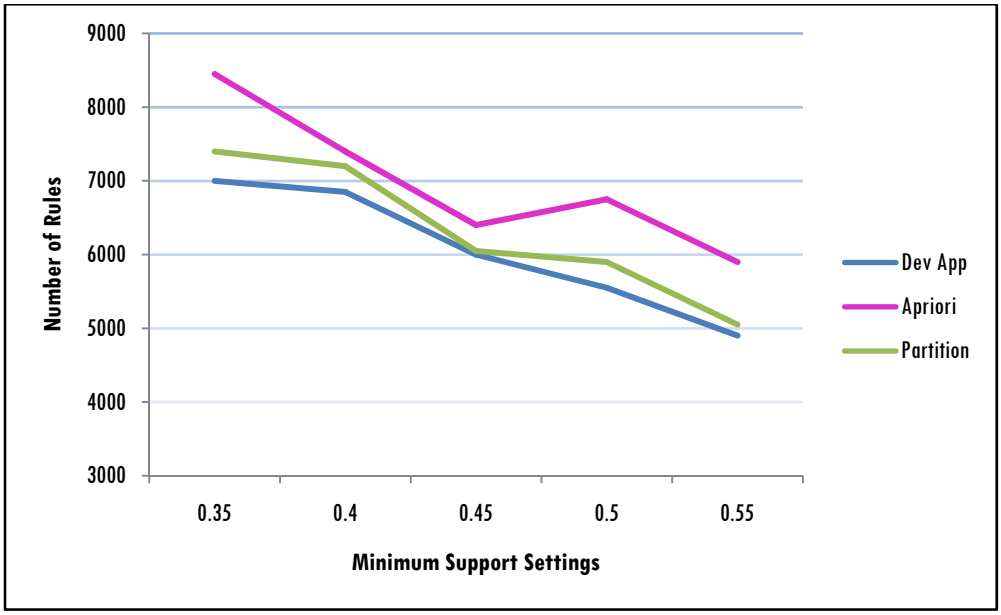


Figure 2-4 Apriori Performance Comparison

Table 2-5 Performance Comparison of the Apriori

	Data set 1	Data set 2	Data set 3
No. of transactions	120,000	400,000	750,000
No. of items	420	557	682
Max items/ transaction	18	26	38
Min items/ transaction	5	7	11
Support %	0.35%, 0.40 %, 0.45%, 0.50%, 0.55%	0.35%, 0.40 %, 0.45%, 0.50%, 0.55%	0.35%, 0.40 %, 0.45%, 0.50%, 0.55%
Confidence %	0.40%, 0.47 %, 0.55 % 0.60%, 0.65%	0.40%, 0.47 %, 0.55 % 0.60%, 0.65%	0.40%, 0.47 %, 0.55 % 0.60%, 0.65%
Avg # rules / FARMA	6262	6823	8118
Avg # rules / Apriori	7005	7597	9011
Avg # rules /Partition	6618	7107	8391

2.4 Pharmacovigilance

So far, the discussion has been regarding how the DDIs can be found. In the current section, the possible applicability of the DDIs identified using existing systems is discussed.

ADEs cause hospitalization, cost billions of dollars annually and is often fatal to patients [111], [112]. Pharmacovigilance intends to minimize Adverse Drug Events through minimum patient exposure before product marketing. Studies show that pharmacovigilance plays a crucial role in combating against DDIs due to the ever increasing range and potency of drugs [113]. DDIs pose additional and serious danger and must be identified and prevented. Drug Interaction Management demands tools to improve guideline quality [114]. A number of software tools have been developed to

cross match drugs to avoid DDIs during various stages of administration, using the currently known drug interactions. The Prescribing Optimization Method (POM) for polypharmacy prescription to the elderly, computerized pharmacy system for DDI checking, and High Speed drug interaction search system for clinical use, are all examples for DDI checking systems [115], [116], [117]. These tools can reduce the risk of ADRs to some extent [3]. Studies have also been conducted to compare standard tools for known DDI interaction checking by Perkins, Murphy et al. Five programs scored perfect sensitivity scores: DrugIx, ePocrates Rx, ePocrates Rx Pro, Lexi-Interact, and the Tarascon pocket Pharmacopoeia. Of these, the Mosbylx programs scored the highest in specificity (1.0). A comparison of the various tools is given in Table 2-6 [118].

Table 2-6 Comparison of various drug interactions checking software

Results by Program								
Program	**TP**	**FP**	**TN**	**FN**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
DrugIx (Handbook of Adverse Drug Reactions)	16	5	16	0	1.0	0.76	0.76	1.0
ePocrates Rx	16	2	19	0	1.0	0.9	0.89	1.0
ePocrates Rx Pro	16	2	19	0	1.0	0.9	0.89	1.0
iFacts (Facts and Comparisons Drug interactions)	14	1	20	2	0.88	0.96	0.93	0.91
Lexi-Interact (Lexi-Comp)	16	10	11	0	1.0	0.52	0.62	1.0
mobileMICROMEDEX	15	6	15	1	0.94	0.71	0.71	0.94
MosbyIx (Mosby's Drug Consult)	13	0	21	3	0.81	1.0	1.0	0.88
Tarascon pocket Pharmacopoeia (deluxe ed.)	16	10	11	0	1.0	0.52	0.62	1.0
Mean (SD)					0.96 (0.07)	0.78 (0.19)	0.80 (0.15)	0.97 (0.06)
Range					0.81-1	0.52-1	0.62-1	0.88-1
FN = false negative; FP = false positive; NPV = negative predictive value; PPV = positive predictive value; TN = true negative; TP = true positive.								

Adherence to drug label recommendations for certain drug interactions is stressed by some studies using data analysis of drug dispensing [119]. Research has found that data mining can be used to discover DDIs and unexpected ADRs to enable better drug safety [120]. Better pharmacovigilance can be achieved by using data mining tools for monitoring prescriptions or inpatient administration and adequate drug labelling, but these can only be made possible by the identification of potential novel DDIs.

2.5 Summary and Conclusion

Epidemiological study of drug-drug interactions indicates that they cause huge financial burden. DDIs are a common problem regardless of age and gender. The chances of DDIs increase with age and number of medications taken concomitantly. Since the elderly experience higher number of comorbidities, they take more medications and have more drug interactions. This again reiterates the increase in drug interactions due to multiple drug usage.

Percentage of DDIs increased when larger populations were involved in various studies, indicating that larger cohorts present more medicinal combinations, increasing the chance of DDIs, i.e., cross matching of drugs is to be made at a larger scale and more general cohort. Therefore, the data resource like FAERS which contain data collected worldwide and of a general nature can be used for the mining process.

As described in section 2.2, the current data mining methods like bivariate drug–AE identification and text mining are inadequate due to their specificity and inability to find new DDIs.

Based on the size of the database under consideration and the information to be mined (the association rules between drugs and their AEs from large data bases), Association Rule Mining can be used to find association rules using a suitable implementation of the Apriori algorithm.

Different existing drug safety tools can be more efficiently used only if novel DDIs are identified and this can substantially improve pharmacovigilance.

Chapter 3

THE DATA BASE AND DATA PRE-PROCESSING

Contents

3.1 Introduction

Data Mining is an essential step in the knowledge discovery process and involves iterative sequence of steps namely data cleaning, data integration, data selection, data transformation, data mining, pattern evaluation, and knowledge presentation. The pre-stage of data mining is the data pre-processing, which involves data cleaning, data reduction, data integration, and data transformation. According to Fayyad, data pre-processing is essential to ensure that useful knowledge is derived from the data. Blind data mining methods are dangerous, easily leading to the discovery of meaningless and invalid patterns [121].

In data cleaning, the data is cleaned by filling the missing values, smoothing the noisy data, identifying and removing the outliers, and resolving inconsistencies. Data reduction is the process of obtaining a reduced representation of the data set.

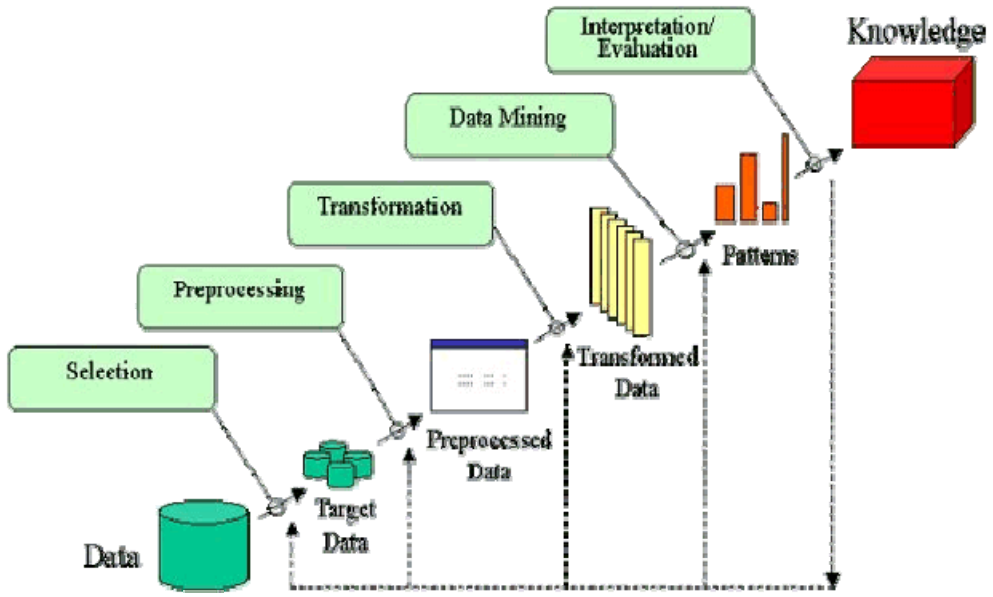


Figure 3-1 Overview of the Steps that compose the KDD (Knowledge Discovery in Databases) Process [121]

Though smaller in volume, the results produced are analytically the same. It involves dimensionality reduction and numerocity reduction. Data integration is the merging of data from multiple data stores. Redundancy and inconsistency avoidance follows the data integration. Accuracy and speed of data mining is improved by the process integration. Using data transformation the data is converted into a form appropriate for data mining. The stages of the KDD process are represented in Figure 3-1 [121].

3.2 Data Resources

The American Food and Drug Administration Adverse Event Reporting System (FAERS) is a repository which contains ADR and medication error reports submitted to the FDA. The database is designed to support the FDA's post-marketing safety surveillance program. As per FDA regulations, the

manufacturers are required to submit suspected ADRs before product marketing. The manufacturers are also mandated to report ADRs reported by consumers. FDA also receives voluntary reports from healthcare professionals and consumers all over the world. The different forms for data entry vary in their formats, for each category. A sample form for mandatory reporting is shown in Appendix A. Following a manual review, these reports are entered into the FAERS database. The adverse events are described at the MedDRA (Medical Dictionary for Regulatory Activities) Preferred Term (PT) level. Details regarding MedDRA are given in Appendix B. Drug information is in R_xNorm context and includes R_xNorm code, method of administration, dosage, and brand information of each drug. Description of R_xNorm as applicable to the current study is given in Appendix C. The database is reported to be a good resource for finding potential DDIs and the existing algorithms for mining bivariate relations are incapable of exploring higher order ADE associations [122], [123]. Patient outcomes, therapy dates, reporting sources, and MedDRA coded indications for reported drugs are also entered into the database.

The FAERS is an online database and contains over five million reports from 1969 till present. As the size and complexity of the database became unmanageable by the traditional method of manual case reviews, data mining algorithms were designed to explore relations of drugs with adverse events.

The report of the first three quarters from 2012 shows representation from 168 countries, including 1709 from India. Figure 3-2 indicates countries which have contributed more than 1000 reports.

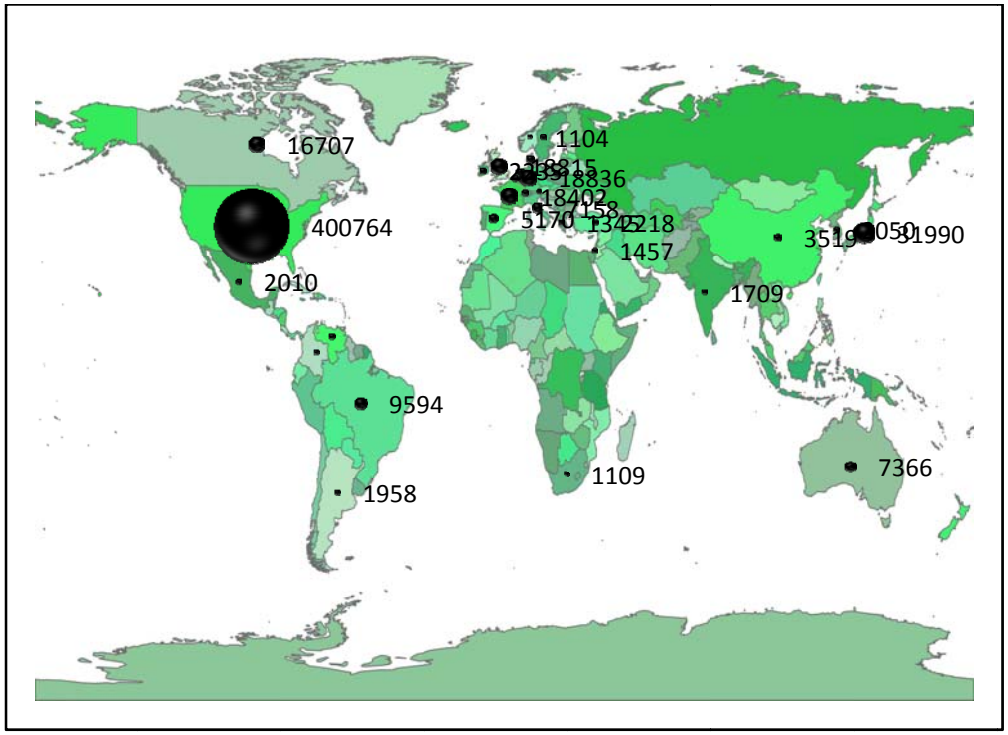


Figure 3-2 Number of reports received by FDA (Country wise, 2012)

Figure 1-1 and Figure 1-9 show statistics of the FAERS database regarding total reports received and reports from inside and outside U.S.

The analysis based on the type of reporter indicates that consumers, though not mandatory reporters, are also keen in the process of reporting.

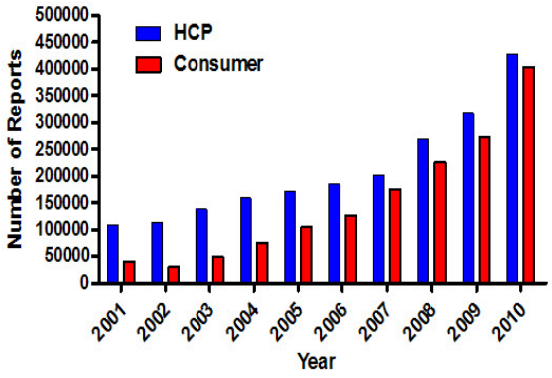


Figure 3-3 Reports received by FDA based on source

Figure 3-3 illustrates the number of reports in FAERS by type of reporter (Healthcare Professional [HCP] or consumer) since the year 2001 until the end of 2010.

There are seven categories for analysing Adverse Event outcomes in the FAERS database. The percentage represented by each outcome in the 2012 reports is represented in Figure 3-4.

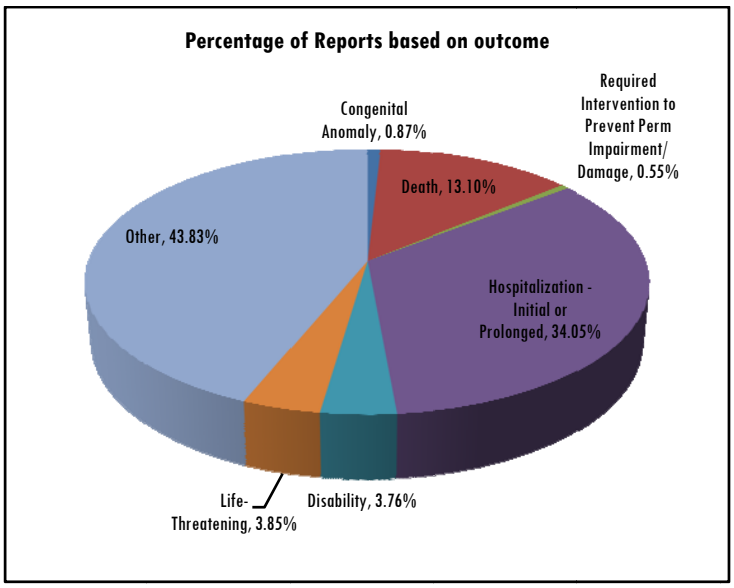


Figure 3-4 Number of reports (%) received by FDA based on Patient Outcome

The percentage of reports based on various age groups is represented in Figure 3-5. The analysis shows that age group 40 to 60 contributes to the highest % of Adverse Event reports.

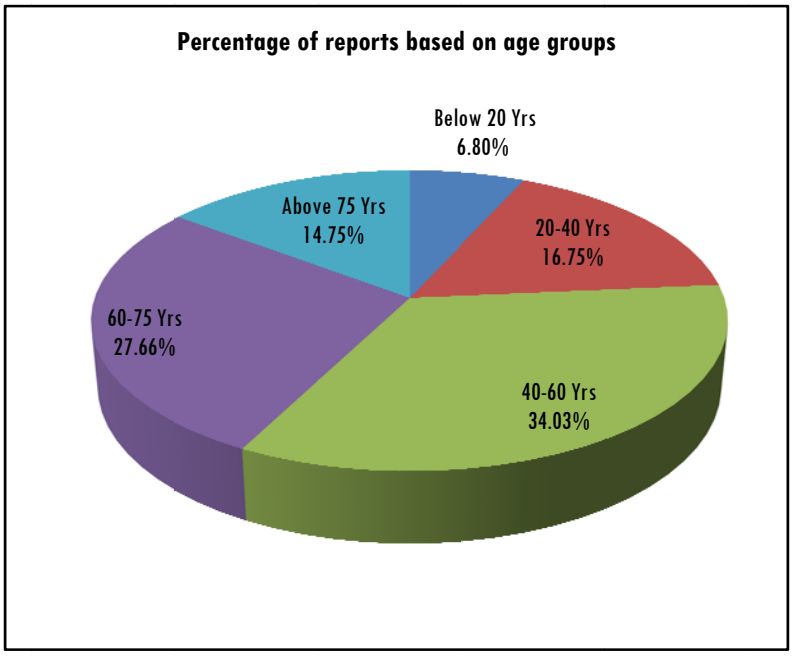


Figure 3-5 Number of reports received by FDA based on age group

3.3 FAERS Database Schema

A detailed schema of the FAERS database is given below. A few sample records of each file are shown in Appendix D.

3.3.1 Demographic File

The file contains patient demographic and administrative information. There will be only a single record for each event report.

Table 3-1 Demographic File

NAME	DESCRIPTION
ISR	Number that uniquely identifies an FAERS Report. Primary link field between data files.
CASE	Number for identifying an FAERS case. A case consists of one or more reports (ISRs). If correctly linked, a follow up report will have the same CASE number as the initial report.
I_F_COD	Code for initial or follow up status of report, as reported by manufacturer. CODE MEANING_TEXT ---- ---------- I Initial F Followup
FOLL_SEQ	The sequence number of a follow-up report, as reported by manufacturer. For initial reports, there will be no data for this field.
IMAGE	Identifier for an FAERS report image. Character field consisting of the ISR number, a dash, and a check digit or letter. (ex: 3123456-X)
EVENT_DT	Date adverse event occurred or began. (YYYYMMDD format)
MFR_DT	Date manufacturer first received initial (or follow-up) information. (YYYYMMDD format)
FDA_DT	Date FDA received report. (YYYYMMDD format)
REPT_COD	Code for the type of report submitted. CODE MEANING_TEXT ---- ---------- EXP Expedited (15-Day) PER Periodic DIR Direct
MFR_NUM	Manufacturer's unique report identifier.
MFR_SNDR	Verbatim name of manufacturer sending report.
AGE	Numeric value of patient's age at event.
AGE_COD	Unit abbreviation for patient's age. CODE MEANING_TEXT ---- ---------- DEC DECADE YR YEAR MON MONTH WK WEEK DY DAY HR HOUR

GNDR_COD	Code for patient's sex. CODE MEANING_TEXT ---- ---------- UNK Unknown M Male F Female NS Not Specified
E_SUB	Whether (Y/N) this report was submitted under the electronic submissions procedure for manufacturers.
WT	Numeric value of patient's weight.
WT_COD	Unit abbreviation for patient's weight. CODE MEANING_TEXT ---- ---------- KG Kilograms LBS Pounds GMS Grams
REPT_DT	Date report was sent. (YYYYMMDD format)
OCCP_COD	Abbreviation for initial reporter's type of occupation. CODE MEANING_TEXT ---- ---------- MD Physician PH Pharmacist OT Other health-professional LW Lawyer CN Consumer
DEATH_DT	Date patient died. (YYYYMMDD format) This field remains but is no longer populated with data due to privacy concerns
TO_MFR	Whether or not (Y/N) voluntary reporter also notified manufacturer (blank for manufacturer reports).
CONFID	Whether or not (Y/N) voluntary reporter stated that his identity should not be disclosed to the product manufacturer (blank for manufacturer reports).
REPORTER_COUNTRY	Reporters are asked to give us their addresses. This is usually the country the event occurred in.

3.3.2 Drug File

The file contains drug/biologic information for as many medications as were reported for the event. There can be more than one for each event.

Table 3-2 Drug File

NAME	DESCRIPTION
ISR	Number that uniquely identifies an FAERS report. Primary link field between data files.
DRUG_SEQ	Drug sequence number for identifying a drug for an ISR. To link to the RUGyyQq.TXT data file, both the ISR number (primary key) and the DRUG_SEQ number (secondary key) are needed.(For an explanation of the DRUG_SEQ number, including an example.
ROLE_COD	Code for drug's reported role in event. CODE MEANING_TEXT ---- ---------- PS Primary Suspect Drug SS Secondary Suspect Drug C Concomitant I Interacting
DRUGNAME	Name of medicinal product. If a "Valid Trade Name" is populated for this ISR, then DRUGNAME = Valid Trade Name; if not, then DRUGNAME = "Verbatim" name, exactly as entered on the report. For the great majority of reports, there is a "Valid Trade Name."
VAL_VBM	Code for source of DRUGNAME. CODE MEANING_TEXT ---- ---------- 1 Validated trade name used 2 Verbatim name used
ROUTE	The route of drug administration.
DOSE_VBM	Verbatim text for dose, frequency, and route, exactly as entered on report.
DECHAL	Dechallenge code, indicating if reaction abated when drug therapy was stopped. CODE MEANING_TEXT ---- ---------- Y Positive dechallenge N Negative dechallenge U Unknown D Does not apply
RECHAL	Rechallenge code, indicating if reaction recurred when drug therapy was restarted. CODE MEANING_TEXT ---- ---------- Y Positive rechallenge N Negative rechallenge U Unknown D Does not apply
LOT_NUM	Lot number of the drug.
EXP_DT	Expiration date of the drug. (YYYYMMDD format)
NDA_NUM	Verbatim NDA number, exactly as reported.

3.3.3 Reaction File

The file contains all (MedDRA) "Medical dictionary for Regulatory Activities") terms coded for the adverse event. There can be more than one adverse event.

Table 3-3 Reaction File

NAME	DESCRIPTION
ISR	Number that uniquely identifies an AERS report. Primary link field between data files.
PT	"Preferred Term" level medical terminology describing the event, using the Medical Dictionary for Regulatory Activities (MedDRA). The order of the terms for a given event does not imply priority. In other words, the first term listed is not necessarily considered more significant than the last one listed.

3.3.4 Outcome File

This file contains patient outcome for each of the events. The events may be zero or more depending on the outcome of the patient caused by the adverse effect.

Table 3-4 Outcome File

NAME	DESCRIPTION
ISR	Number that uniquely identifies an AERS report. Primary link field between data files.
OUTC_COD	Code for a patient outcome. CODE MEANING_TEXT ---- ---------- DE Death LT Life-Threatening HO Hospitalization - Initial or Prolonged DS Disability CA Congenital Anomaly RI Required Intervention to Prevent Permanent Impairment/Damage OT Other

3.3.5 Report Source File

The file contains report source for the event. The source of the report may be foreign or domestic, and may be from health care professionals or social media, or even consumers.

Table 3-5 Report Source File.

Name	Description
ISR	Number that uniquely identifies an AERS report. Primary link field between data files.
RPSR_COD	Code for an initial source of the report. CODE MEANING_TEXT ---- ---------- FGN Foreign SDY Study LIT Literature CSM Consumer HP Health Professional UF User Facility CR Company Representative DT Distributor OTH Other

3.3.6 Therapy File

The file contains drug therapy start dates and end dates for the reported drugs. There can be zero or more drugs for each event.

Table 3-6 Therapy File

NAME	DESCRIPTION
ISR	Number that uniquely identifies an FAERS report. Primary link field between data files.
DRUG_SEQ	Drug sequence number for identifying a drug for an ISR. To link to the RUGyyQq.TXT data file, both the ISR number (primary key) and the DRUG_SEQ number (secondary key) are needed
START_DT	A date therapy was started (or re-started) for this drug. (YYYYMMDD)
END_DT	A date therapy was stopped for this drug. (YYYYMMDD)
DUR	Numeric value of the duration (length) of therapy
DUR_COD	Unit abbreviation for duration of therapy CODE MEANING TEXT ---- ---------- YR Years MON Months WK Weeks DAY Days HR Hours MIN Minutes SEC Seconds

3.3.7 Indications File

This file contains all MedDRA terms coded for the indications for use (diagnoses) for the reported drugs. These can be zero or more per drug per event.

Table 3-7 Indications File

NAME	DESCRIPTION
ISR	Number that uniquely identifies an FAERS report. Primary link field between data files.
DRUG_SEQ	Drug sequence number for identifying a drug for an ISR. To link to the RUGyyQq.TXT data file, both the ISR number (primary key) and the DRUG_SEQ number (secondary key) are needed
INDI_PT	"Preferred Term" level medical terminology describing the Indication for use, using the Medical Dictionary for Regulatory Activities (MedDRA).

The FAERS website is periodically uploaded with the received reports. Currently, the first 3 quarter reports of 2012 are available online. In the present problem, these reports were used in the mining process. The latest reports contain information about new drugs and therefore can be sources of potential novel DDIs. Moreover, many reports from the earlier dates would have become obsolete due to the withdrawal of drugs due to various reasons.

3.3.8 Data Element Contents and Maximum Lengths

The following table provides the maximum length and type of each data element.

Table 3-8 Data Element Details

Data Element	Data Content	Max Length
ISR	N(numeric)	10
CASE	N	10
I_F_CODE	A(alphabetic)	1
FOLL_SEQ	AN(alphanumeric)	2
IMAGE	AN	12
EVENT_DT	N (or D,date)	8
MFR-DT	N (or D)	8
FDA_DT	N (or D)	8
REPT_CODE	A	3
MFR_NUM	AN	100
MFR_SNDR	AN	60
AGE	N	7(including 2 decimal places)
AGE_COD	A	3
GNDR_COD	A	3
E-SUB	A	1
WT	N	11(including 5 decimal places)
WT_COD	A	3
REPT_DT	N(or D)	8
OCCP_COD	A	2
DEATH_DT	N (or D)	8
TO_MFR	A	1
CONFID	A	1
Reporter-Country	A	50
DRUG-SEQ	N	10
ROLE_COD	A	2
DRUGNAME	AN	70
VAL_VBM	N	1
ROUTE	A	28
DOSE_VBM	AN	100
DECHAL	A	1
RECHAL	A	1
LOT_NUM	AN	35
EXP_DT	N(or D)	8
NDA_NUM	AN	7
PT	AN	100
OUTC_COD	A	2
RPSR_COD	A	3
START_DT	N(or D)	8
END_DT	N(or D)	8
DUR	N	5
DUR_COD	A	3
INDI_PT	AN	100

3.4 Data Pre-processing

The FAERS database required to be pre-processed to suit the problem under consideration, for reducing the unnecessary complexity and also due to duplicity.

3.4.1 Data Pre-Processing Techniques

Figure 3-6 illustrates the data pre-processing done on the FAERS database. The data pre-processing module which performs three tasks, namely, elimination of single drug-AEs, duplicate report elimination, and report selection based on outcome is implemented in Python and runs on Intel® Core™2 Duo processor at 3.0 GHz.

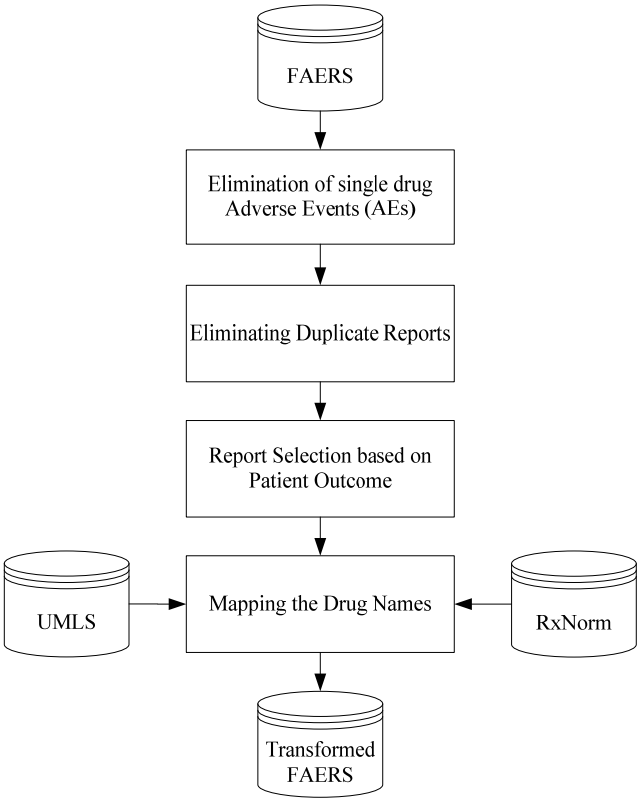


Figure 3-6 Data Pre-processing

3.4.2 Elimination of Single Drug Adverse Events

The FAERS database contains numerous reports which show adverse events caused by single drug usage. These are not sources of drug-drug interactions and are not relevant in the present study. Therefore, these reports from the DRUG file are searched and removed based on the count of the drug sequence number (DRUG_SEQ). Now the corresponding report from the REACTION file is also deleted being irrelevant. The process is shown in Figure 3-7. (An ISR number uniquely identifies a FAERS report.)

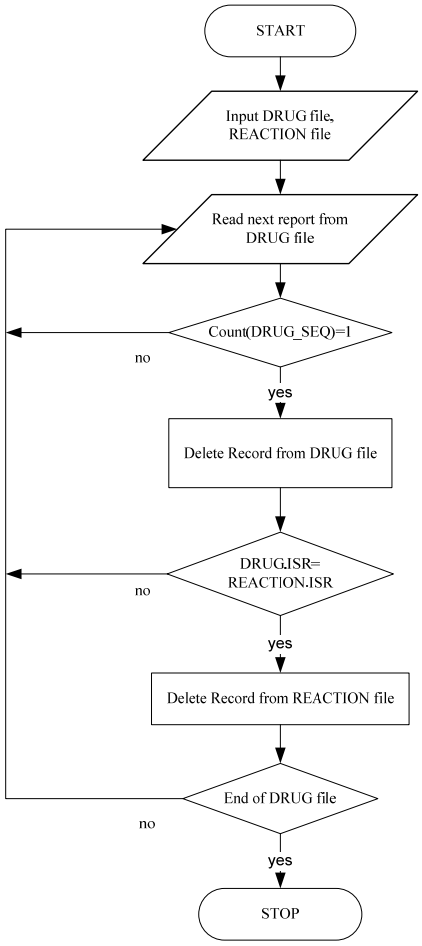


Figure 3-7 Eliminating single drug Adverse Events

Pseudocode

1. *Input DRUG file and REACTION file.*

2. *Read next report from DRUG file.*

3. If DRUG_SEQ count=1
Delete record from DRUG file
Else
step 2.

4. IF DRUG_ISR=REACTION_ISR
Delete record from REACTION file
Else
step2.

5. If not end of DRUG file
Step 2
Else
STOP.

3.4.3 Elimination of Duplicate Reports

The greatest challenge in the FAERS database is its duplicity [124]. The voluntary reports received are often duplicates reporting the same event. Some duplicate reports are present due to error in health care professionals reporting the same event. Manufacturers also receive reports from consumers or relatives regarding the adverse events. These will also create duplicity. The FAERS statistics in Figure 3-3 demonstrates the consumer participation in voluntary reporting.

The database does not contain patient name or date of birth. Death date field (DEATH_DT) is no longer populated due to privacy reasons as per the explanation given on the FAERS schema. Therefore, elimination of duplicate reports required comparison of several data fields from multiple files. The date of Adverse Event occurrence (EVENT_DT), age (AGE), gender (GNDR_COD) and weight (WT) from DEMOGRAPHIC file and outcome

code (OUTC_COD) from OUTCOME file are compared to find duplicate entries and the corresponding reports from the DRUG file and REACTION file are deleted. The process is shown in Figure 3-8.

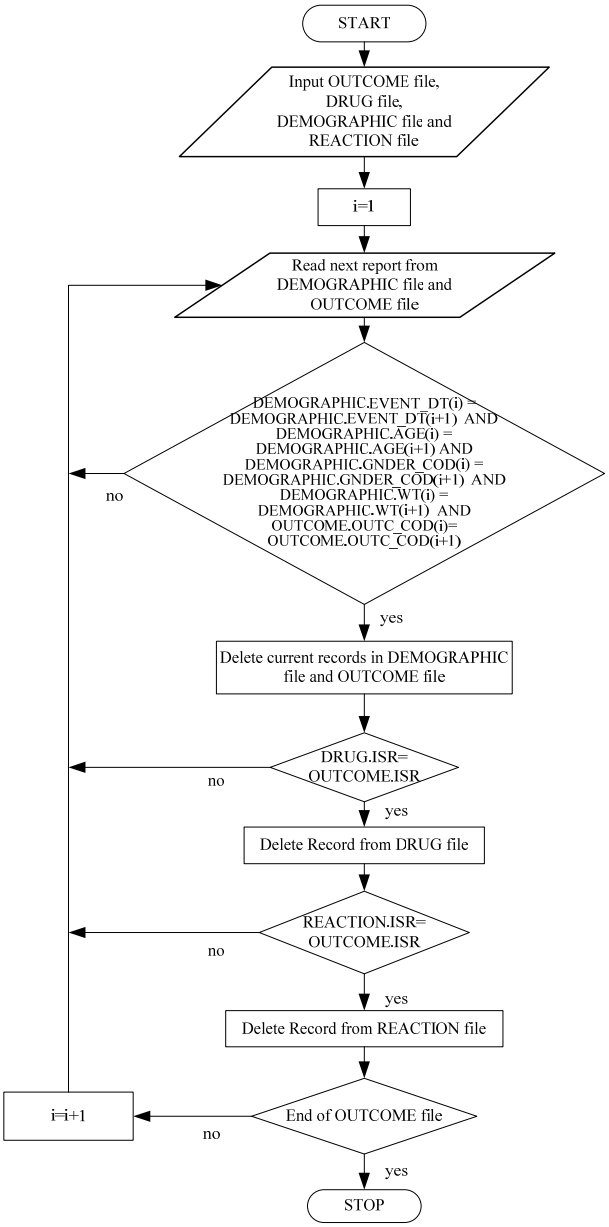


Figure 3-8 Eliminating duplicate reports

Pseudocode

1. Input DRUG file ,OUTCOME FILE ,REACTION file AND DEMOGRAPHIC file.

2.Initialize OUTCOME file pointer.

3.Read next reports from DEMOGRAPHIC file and OUTCOME file.

4. IF

DEMOGRAPHIC.EVENT_DT(i)=DEMOGRAPHIC.EVENT_DT(i+1) AND DEMOGRAPHIC.AGE(i)= DEMOGRAPHIC.AGE(i+1)

AND

DEMOGRAPHIC.GNDER_COD(i)= DEMOGRAPHIC.GNDER_COD(i+1) AND DEMOGRAPHIC.WT(i)= DEMOGRAPHIC.WT(i+1)

AND

OUTCOME.OUTCOME_COD(i)=DEMOGRAPHIC.OUTCOME_COD(i+1)

delete current records from OUTCOME file and DEMOGRAPHIC file.

ELSE
Step 3.

5. IF DRUG.ISR=OUTCOME.ISR
delete record from DRUG file
ELSE
Step 3.

6.IF REACTION.ISR=OUTCOME.ISR
Delete record from REACTION file
ELSE
Step 3.

7. IF not end of OUTCOME file
Increment OUTCOME file pointer;
Step 3
ELSE
STOP.

3.4.4 Report Selection Based on Patient Outcome

The OUTCOME file of the FAERS database contains codes 'DE' for death, 'LT' for life-threatening,' HO' for hospitalization, 'DA' for disability, 'CA' for Congenial Anomaly, 'OT' for Other categories and 'RI' for required intervention. Reports containing serious patient outcomes like death, life-threatening, hospitalization, or disability are to be selected as these can be the indicators of potential DDIs. The OUTCOME file is searched for the non-serious patient outcomes and the non-specified other category and the corresponding reports from DRUG and REACTION file are deleted. The process is illustrated in Figure 3-9.

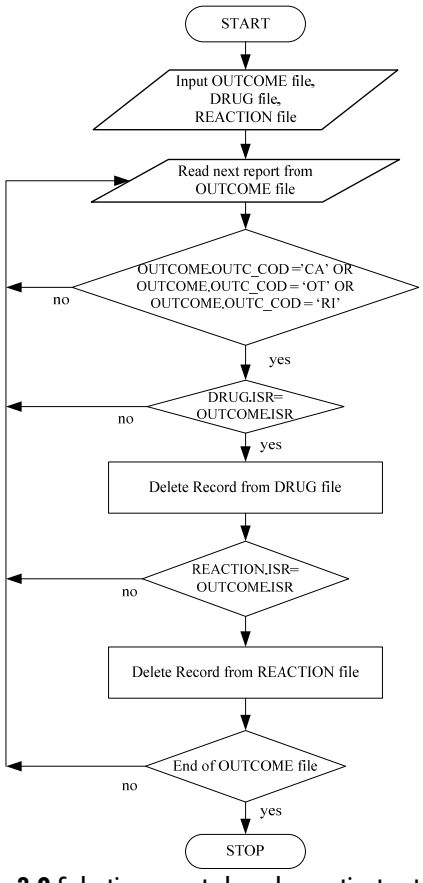


Figure 3-9 Selecting reports based on patient outcome

Pseudocode

1. Input OUTCOME file DRUG file and REACTION file.

2.Read next report from OUTCOME file.

3. If

OUTCOME.OUTC_COD = 'CA' OR
OUTCOME.OUTC_COD = 'OT' OR
OUTCOME.OUTC_COD = 'RI'

If DRUG.ISR=OUTCOME.ISR

Delete record from DRUG file.

ELSE
Step 2.

4.If REACTION.ISR=OUTCOME.ISR
Delete record from REACTION file

ELSE
Step 2.

5. If not end of OUTCOME file
Step 2
ELSE

STOP.

3.4.5 Method of Mapping the Drug Names

The drug information in the database is in the R_xNorm context. R_xNorm is a standardized nomenclature for clinical drugs and is produced by NLM, the National Library of Medicine.

In the R_xNorm context, clinical drug is a pharmaceutical product given to patients with a therapeutic or diagnostic intent. Therefore, in R_xNorm, the name of a clinical drug i.e. brand name, combines its ingredients, strength, and form. In other words, depending on weight and method of administration, the same drug has different codes.

For example, consider the 2012 first quarter report of the FAERS database. The drug LIPITOR is given different codes based on dosage.

Table 3-9 Different drug codes for the same medicine

Code	Dosage
1018426025	10 mg tablet
1018422348	20 mg tablet
1017686798	40 mg tablet
1018425841	80 mg tablet

If a medicine is given as injection or syrup, it will have multiple codes depending on dosage. In the mining process, this means that there will be multiple codes for any given drug, increasing the complexity of the algorithm and creating irrelevant combinations, i.e. comparing the same drug.

In order to combat this problem, every drug obtained from a report is assigned a UMLS (Unified Medical Language System) drug code using the standard tool MedLEE (Medical Language Extraction and Encoding System). The goal of MedLEE is to extract, structure, and encode clinical information in textual patient reports so that the data can be used by subsequent automated processes. Finally, the UMLS codes are mapped to the generic using R_xNorm. Therefore, all the codes mentioned in the example corresponding to the brand

name LIPITOR can be mapped to the corresponding UMLS-CODE C0286651 for the generic ATROVASTATIN.

3.5 Conclusion

The elimination of duplicate reports and reports containing single drug-adverse event pairs reduced the search space of the data mining algorithm and removal of irrelevant data. Report selection based on serious patient outcomes helped to further reduce search space and increase 'signals' of DDIs. Also, drug name mapping further reduced the search space and made the results globally applicable.

Chapter 4

DEVELOPING THE TOOL FOR DDI IDENTIFICATION

Contents

4.1 Introduction

Data Mining is an analytic process designed to explore data in search of consistent patterns and/or systematic relationship between variables, and validating the findings by applying the detected patterns to new subset of data [125]. Association rule mining is a well-researched method for discovering relationships between variables in large data sets. These powerful exploratory techniques have a wide range of applications in many areas of research [126]. These techniques are efficient in uncovering hidden patterns in large data sets. The aim of the technique described here is to detect relationships or associations between drugs and their adverse events.

4.2 Apriori Algorithm

Apriori is currently one of the best known algorithms to mine association rules. In this section, the use of Apriori algorithm to mine drugs and their adverse events is explained. Apriori employs a bottom-up, breadth-

first search that enumerates every single frequent pattern in a database. It starts by finding all frequent patterns of size 1, which is then used to find all frequent patterns of size 2 etc.

Figure 4.1illustrates frequent itemset selection. In the current context, this is a collection of frequently prescribed drugs and frequent AEs.

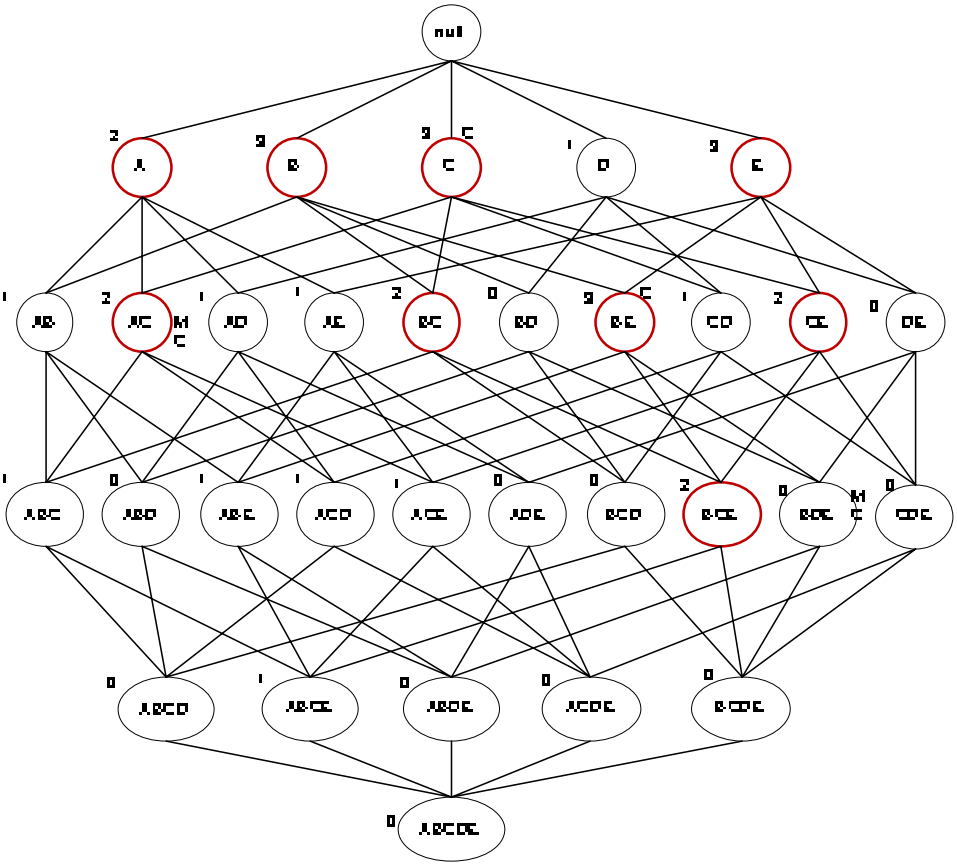


Figure 4-1 Frequent itemset generation

Apriori uses the downward closure property of pattern support, all subsets of a frequent pattern must themselves be frequent to prune the search space. Thus, only frequent patterns of size *k* are used to generate patterns of

size $k+1$.In the DDI mining context this means if a drug or adverse event is uncommon, then the associations formulating from this will also be uncommon and can be eliminated from consideration. In this step, the uncommon drugs and AEs are eliminated. This property can significantly reduce the search space as illustrated Figure in 4.2.

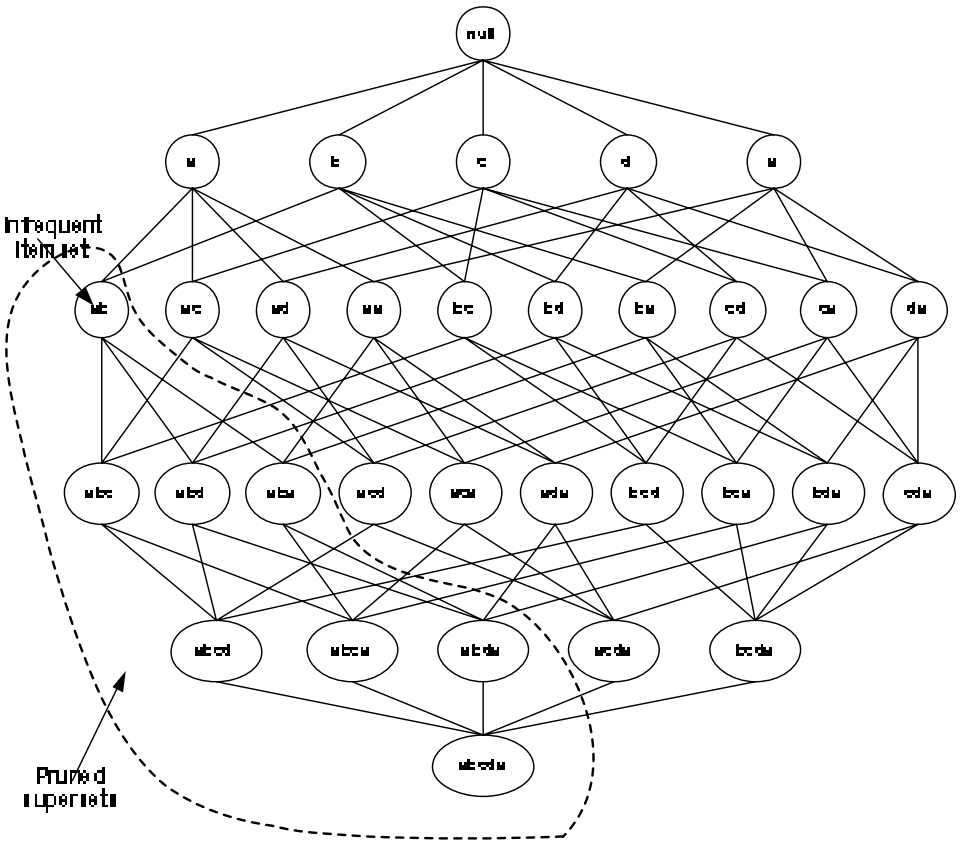


Figure 4-2 Infrequent itemsets

Association rules are formulated from the frequent itemsets generated. In DDI mining, the rules are formulated using the frequently occurring drugs and AEs.

Figure 4-3 demonstrates identification of Association Rules.

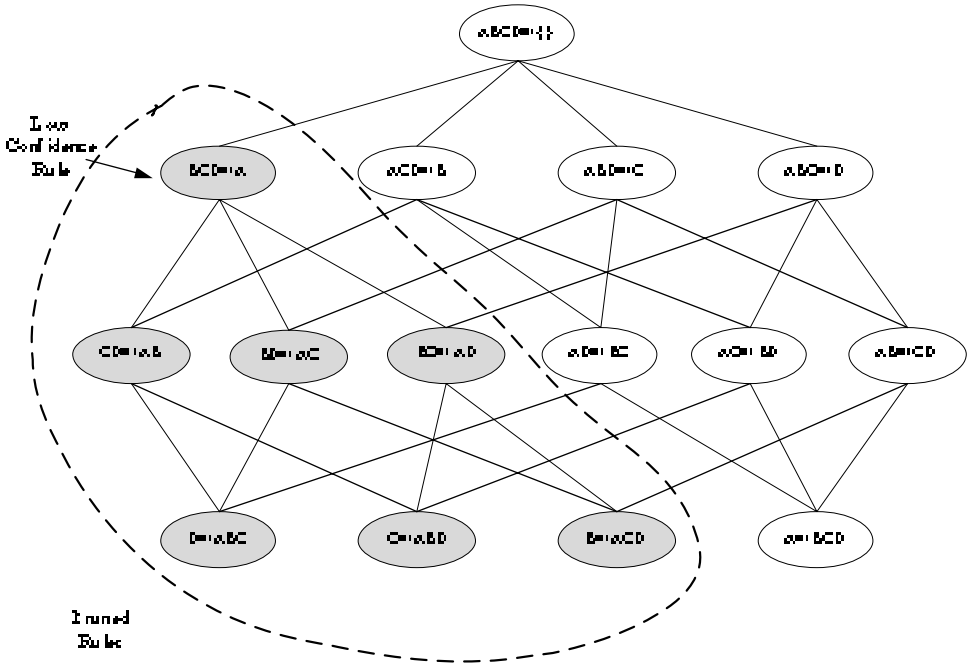


Figure 4-3 Association Rule generation

The Apriori Algorithm is given below.

- A **frequent itemset** is an itemset whose support is greater than some user-specified minimum support (denoted L_k, where k is the size of the itemset)
- A **candidate itemset** is a potentially frequent itemset (denoted C_k, where k is the size of the itemset)

Algorithm

Input: DRUGS and REACTION set.

Output: Frequent Group of Drugs and AEs.

Pass 1

1. Generate the candidate itemsets in C_1
2. Save the frequent itemsets in L_1

Pass *k*

1. Generate the candidate itemsets in C_k from the frequent itemsets in L_{k-1}
    1. Join $L_{k-1}p$ with $L_{k-1}q$, as follows:
       insert into C_k
       select $p.\text{item}_1, p.\text{item}_2, \ldots, p.\text{item}_{k-1}, q.\text{item}_{k-1}$
       from $L_{k-1}p, L_{k-1}q$
       where $p.\text{item}_1 = q.\text{item}_1, \ldots p.\text{item}_{k-2} = q.\text{item}_{k-2}, p.\text{item}_{k-1} < q.\text{item}_{k-1}$
    2. Generate all (k-1)-subsets from the candidate itemsets in C_k
    3. Prune all candidate itemsets from C_k where some (k-1)-subset of the candidate itemset is not in the frequent itemset L_{k-1}
2. Scan the transaction database to determine the support for each candidate itemset in C_k
3. Save the frequent itemsets in L_k

The Apriori uses an iterative approach called level-wise search where n drugs/adverse events are used to explore *(n+1)*itemsets. First the set of frequent items i.e., drugs and AEs, say DAE1, is found by scanning the database to accumulate the count for each drug/ AE set 1 and collecting the drugs/AEs that satisfy minimum support. The resulting set is denoted by L_1.

Next, L_1 is used to find L_2, the set of frequent itemset 2, which is used to find L_3 and so on until no more frequent n drugs/AEs can be found. Finding each L_k requires one full scan of the database. Apriori property is that all non-empty subsets of frequent sets must also be frequent. By definition, if a drug/AE set S do not satisfy the minimum support threshold, *minisup*, then S is not frequent, i.e. P(S) <*minisup*. If the Drug A or adverse event A is added to the set S, then resulting association cannot occur more frequently than S. Therefore, (SUA) is not frequent either; i.e. P(SUA) <*minisup*. This property belongs to a special category of properties called anti-monotonicity in the sense that, if a set cannot pass a test, all its supersets will fail the same test as well [127]. It is called anti-monotone, as the property is monotonic in the context of failing a test. Here a two-step process is considered, join and prune. The sample training decision tables are shown in tables 4-1,4-2,4-3 and 4-4.

Training Decision Tables

Table 4-1 Training Decision Table- DAE1

Drug/AE	Support Count
Drug 1	3
Drug 2	6
AE 1	7
AE 2	5
Drug 3	4
Drug 4	2
Drug 5	9
Drug 6	6
AE 3	7
Drug 7	7
Drug 8	10
AE 4	7

Table 4-2 Training Decision Table- L1

Drug/AE	Support Count
Drug 2	6
AE 1	**7**
AE 2	5
Drug 5	9
Drug 6	6
AE 3	**7**
Drug 7	**7**
Drug 8	10
AE 4	**7**

Table 4-3 Training Decision Table- L2

Drug/AE	Support Count
Drug 2, AE1	6
Drug 2, Drug5	5
Drug5, Drug 6	6
AE4, Drug7	5
Drug 7, Drug 8	7
Drug 8, AE4	6

Table 4-4 Training Decision Table- L3

Drug/AE	Support Count
Drug 8, Drug 7, AE 4	5

In the first iteration of the algorithm, each drug and AE is a member of the set of candidate 1 itemset DAE1. The algorithm simply scans all the reports in order to count the number of occurrences of each drug and AE. Suppose that minimum support count required is 5, i.e. *minsup*=5. The set of frequent 1 itemset L1 can be determined. It consists of the candidate 1 itemsets satisfying minimum support to discover the set of frequent 2 itemset L2. The algorithm uses the join step L1 JOIN L2 to generate the candidate of set 2 itemsets. Next, it finds the third frequent itemsets using the join step L2 JOIN L3 and determines the L3 sets. Then the algorithm uses L3 JOIN L4 to generate candidate sets of fourth frequent itemsets and thus DAE 4 is null set.

Imposing the constraint that the antecedent can have only drug names and consequent can have only AEs derives the following association rules.

Sample Associations Rules

The following association rules can be formulated from the example.

Drug 2→AE1

Drug 8, Drug7→AE 4

Drug 8→AE4

Drug7→AE4

4.3 Algorithm Performance Evaluation

The algorithm performance was tested on an Intel® Core™2 Duo processor at 3.0 GHz and the performance graphs are shown in Figures 4-4, 4-5, 4-6 and 4-7.

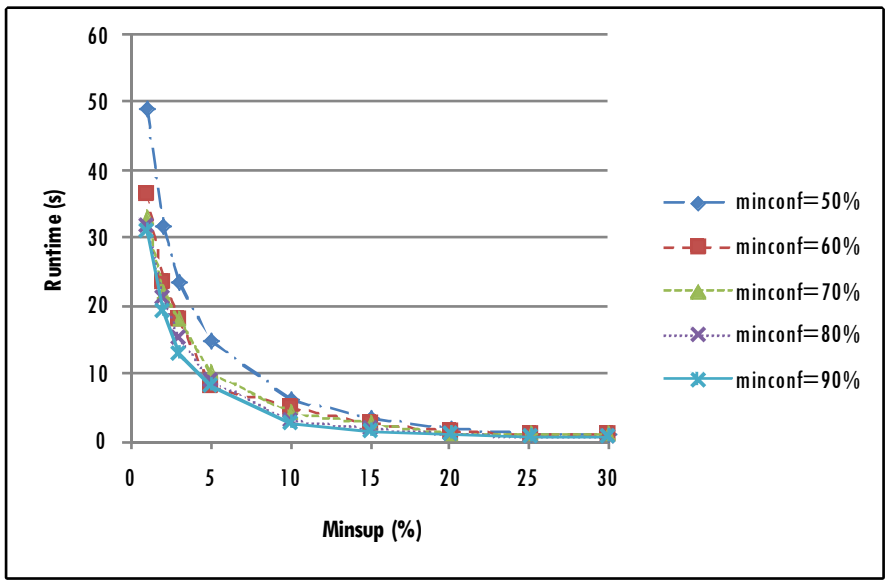


Figure 4-4 Runtime for different *minisup* and *miniconf* values

Change of minimum confidence does not affect the runtimemuch, but the runtime exponentially increased as minimum support became smaller.

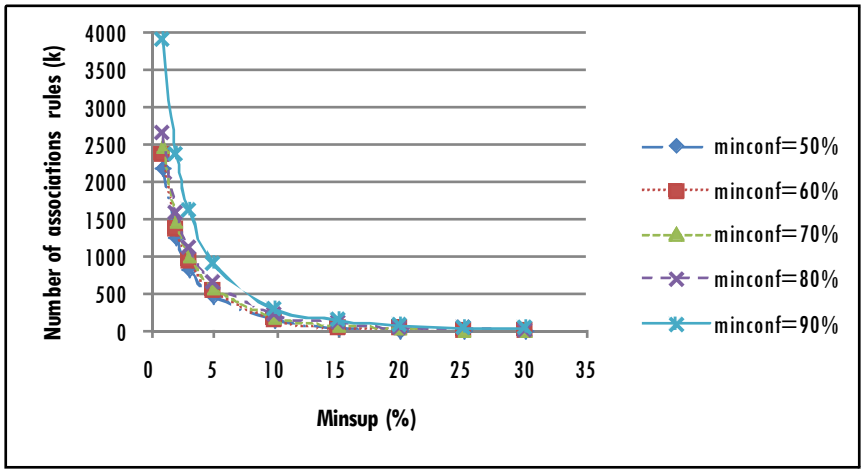


Figure 4-5 Number of associations for various minisup and miniconf values

Minimum confidence level has less influence on the number of associations generated but the associations generated for low minimum support settings exponentially increases the number of rules generated due to the increase in the itemsets for low minimum support settings.

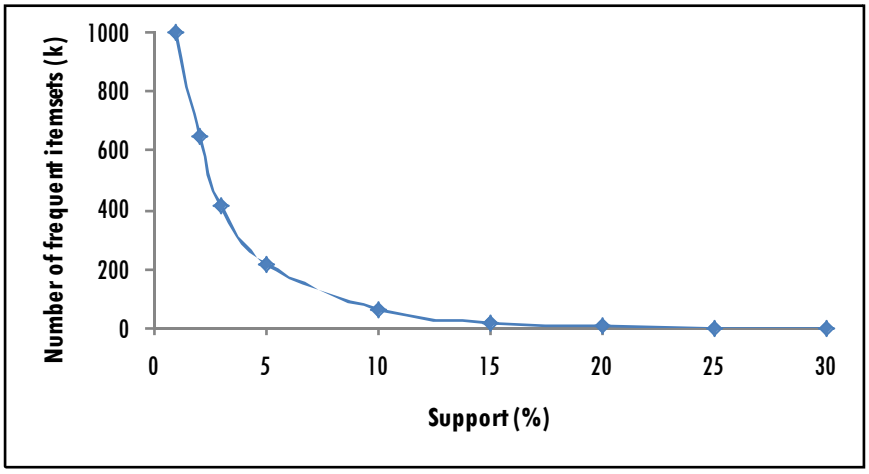


Figure 4-6 Number of frequent itemsets for different minimum support values

Number of frequent itemsets exponentially increases as support becomes smaller. The above results indicate that *minisup* or the anti-monotonicity property of itemsets is very effective to prune non-frequent itemsets.

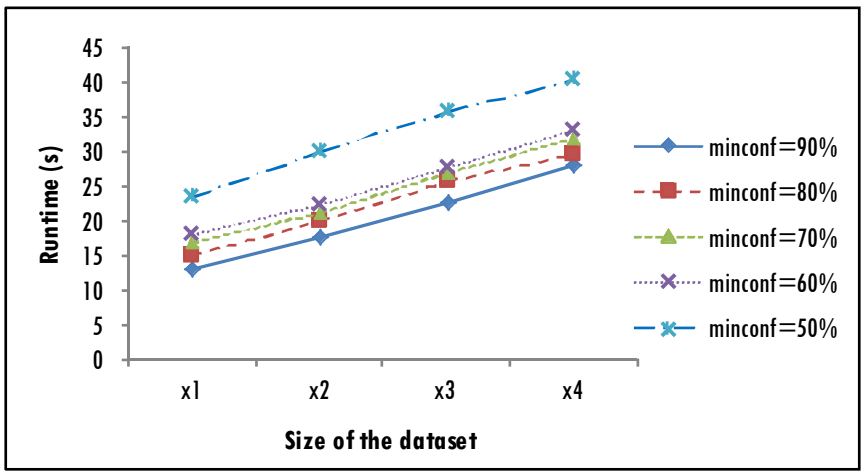


Figure 4-7 Runtime for different itemsets for *minisup* 5

Runtime linearly increases as the number of transactions becomes larger. Therefore under certain distribution of items, *minsup* influences runtime than both *miniconf* and number of transactions in Apriori.

4.4 Mining the FAERS Database

The DRUG file and REACTION file are merged for the ease of data mining. The combined DRUGREACTION file now contains unique ISR numbers corresponding to each Adverse Event, the drug codes for each event, and the adverse effects in MedDRA coded Terminology.

4.4.1 System Flowchart

The various stages in the process of data mining are depicted in Figure 4-8.

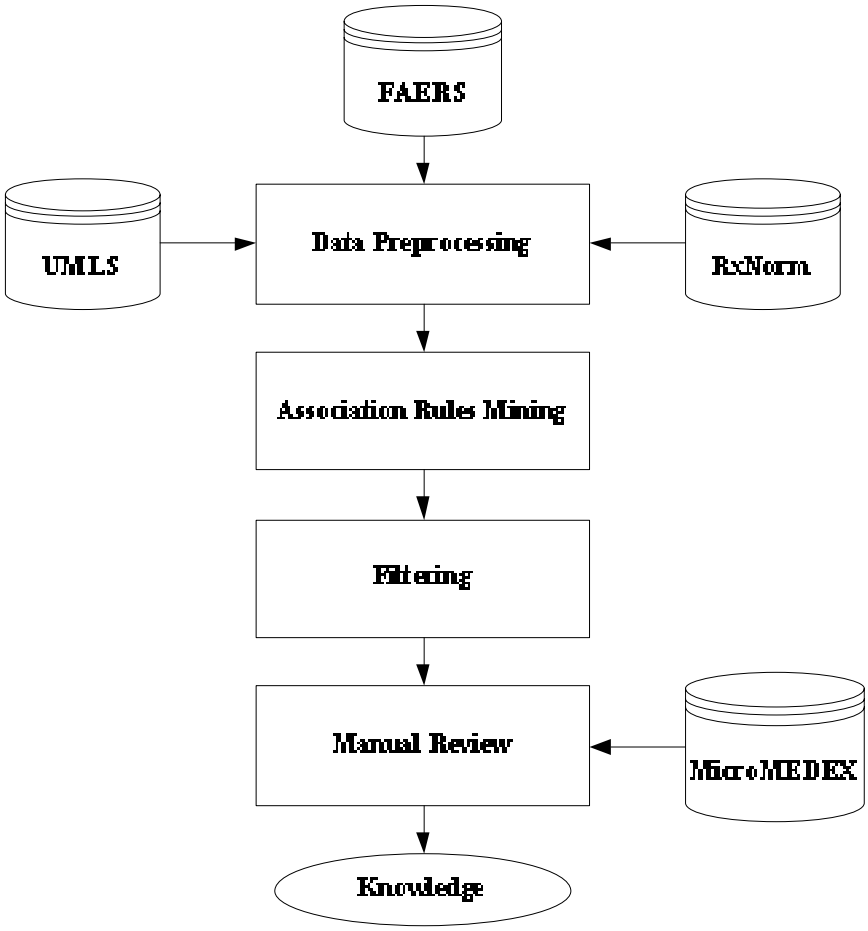


Figure 4-8 The data Mining Process

4.4.2 Process of Data Mining

The search space of concomitantly administered medications is extremely large and discovering the associations between these drugs is computationally difficult considering the size of the large database. If 100 unique drugs and adverse events are to be considered, the number of DDI associations for 2 drugs and 2 AEs that need to be explored becomes $100^4 = 10^8$.

DDI network can be constructed using PAJEK, a program for large-network analysis and visualization. The constructed DDI network is represented in Figure 4-9 and Figure 4-10[128]. Figure 4-9 consists of 966 drugs represented by dots and 3351 known interactions represented by lines. The dots represent interacting drugs and the lines indicate the interactions that were detected for a given pair of drugs. Top 40 interacting drugs are indicated by red dots.

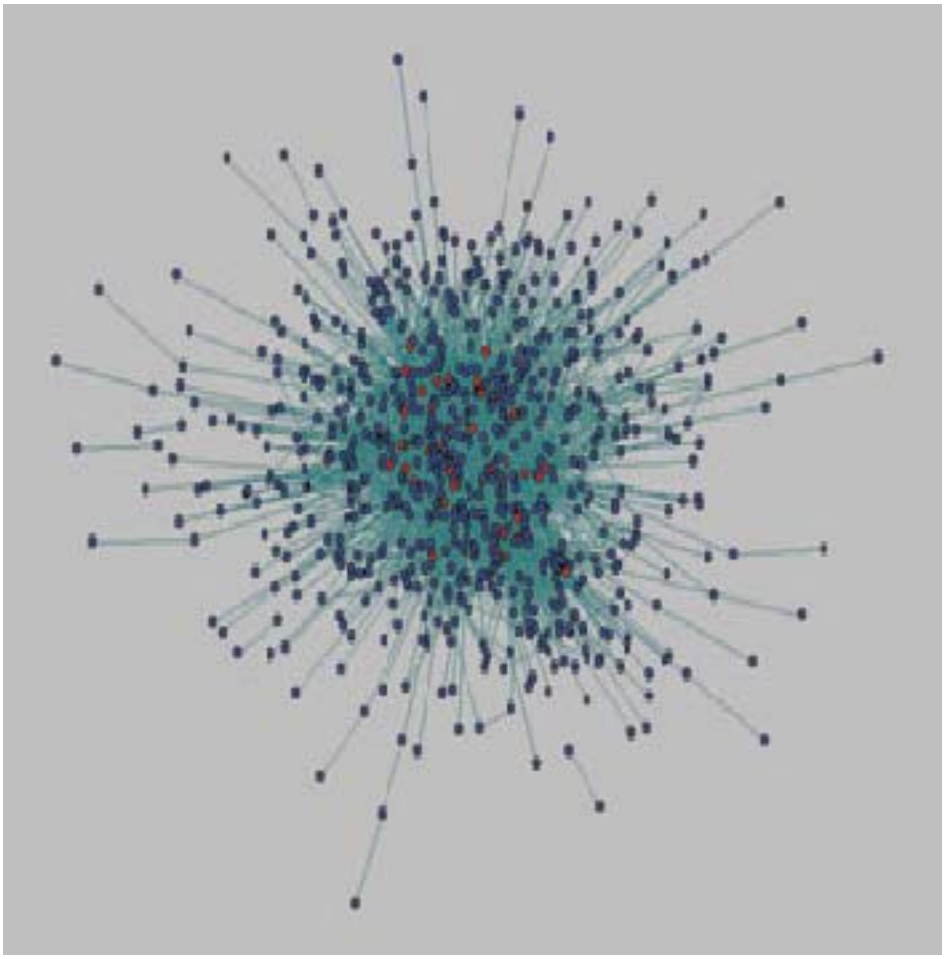


Figure 4-9 Visualization of drug-drug interaction network

The following figure clearly shows interacting drug groups.

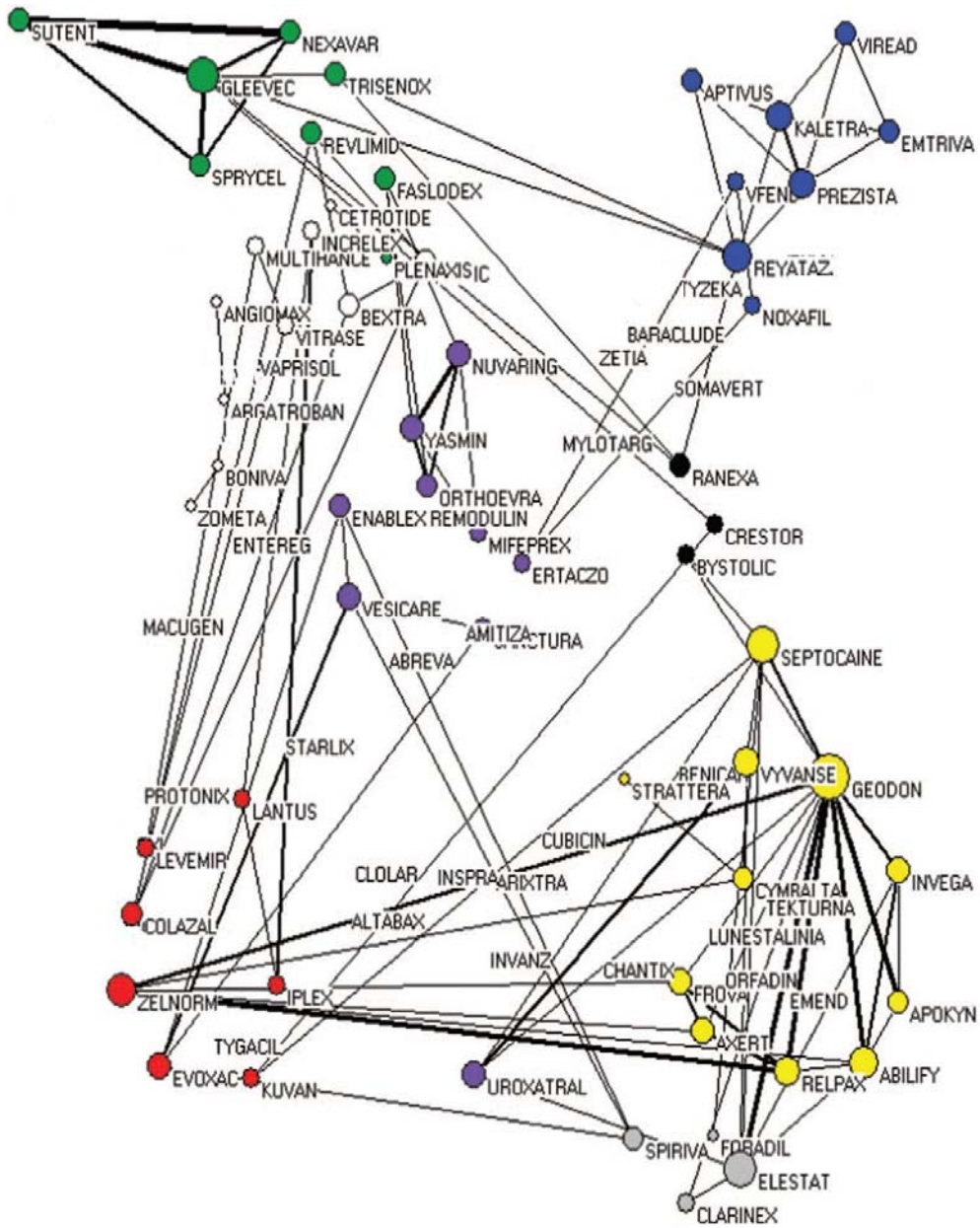


Figure 4-10 Visualization of DDIs

Complexity of matching every possible drug interaction can be imagined effectively from the visualization. Moreover, other associated statistics also need to be prepared.

The process of data mining is performed in the following three steps:

Step 1

Reduce the complexity of the algorithm. To this purpose, the drug names are mapped to their corresponding generic names. This will also help in reducing redundant searches and will strengthen signals.

Step 2

A set of candidate drug-drug ADE associations was generated using the Apriori algorithm.

Step 3

These associations are filtered to remove false associations to identify potential DDIs as shown in Figure 4-13.

4.4.3 Mining Association Rules

The FAERS database is huge and the rules to be explored corresponding to two or more drugs is computationally intractable. To improve the efficiency of the algorithm and reduce search space, additional criteria to increase support and confidence of the association rules and the rules with high priority, such as rules containing certain number of items or set of items, are enforced in finding DDIs. The general Apriori algorithm which is optimized for the above criteria is implemented.

The Apriori algorithm uses the downward closure property of frequency to prune the search space of association rules. In the process of DDI mining this means that if some combination of drugs and AEs are infrequent, then the larger set of combinations which gets formulated on the infrequent one will also be infrequent and can be eliminated from consideration. For example, if the combination of drugs Warfarin and Aspirin are infrequent, then the association rule that can be constructed over these, such as Warfarin+ Aspirin→bleeding, will also be infrequent and can be eliminated.

The Apriori first explores itemsets for a minimum support, and then formulates association rules based on a certain confidence from these generated itemsets. The item generation process poses more challenge as it is based on these that all the possible rules are formulated. Hence, in order to tackle the problem with the large quantity of data in the FAERS database, the algorithm demands enhancement.

Some of the rules generated by the algorithm may contain only drugs or AEs. These are not agreeable to the ADE association definition. Therefore, the constraint is set so that only itemsets with a set of drugs on the left-hand side and at least an AE on the right hand side are considered as potential association rules. This method decreases the search space of possible multi-item associations. Thus the rules to be generated will also be reduced. Indexing based on the drug names and Adverse Events is used to decrease search of the entire database.

The above mentioned optimization techniques such as elimination of non-agreeable associations and indexing by drug names and AEs increased the computational gain to several thousand folds (Figure 4-11, Figure 4-12).

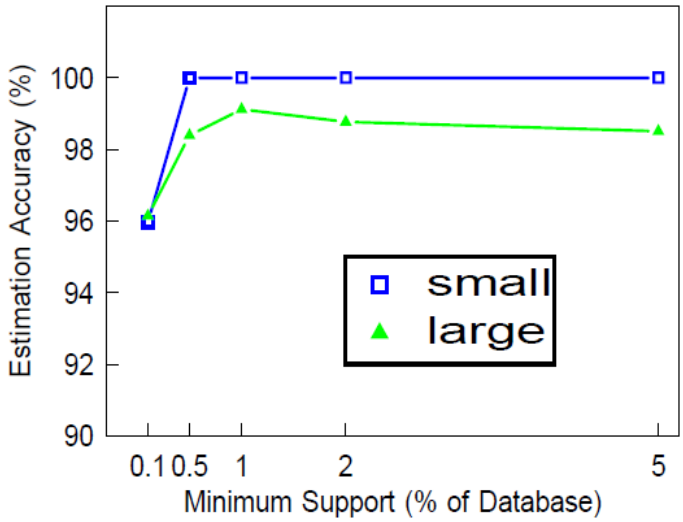


Figure 4-11 Apriori performance enhancement

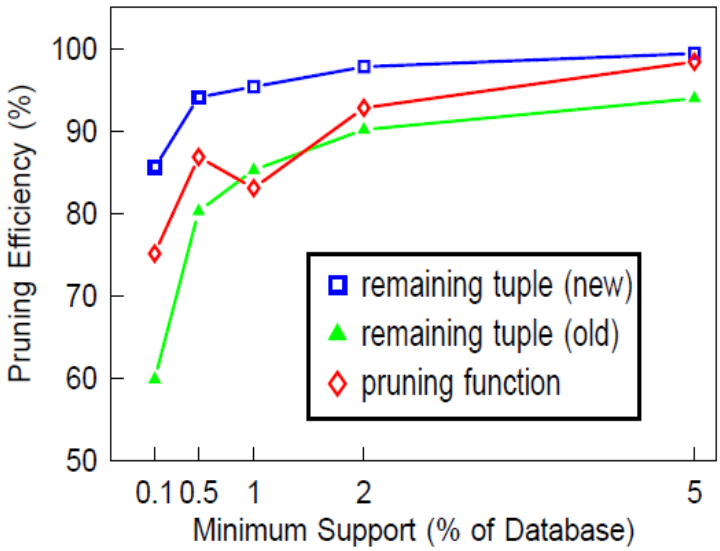


Figure 4-12 Apriori performance enhancement

4.4.4 Increasing the Association Strength: Rule Filtering Using the Proportional Reporting Ratios (PRR)

Frequent Adverse Events like headache or nausea usually generate large confidence values in spite of the drugs associated with them. Also, infrequent adverse events will produce small confidence levels even though they are strongly associated to some drugs. This has been verified by the present study as well as other studies [45], [46], [103]. Besides, mining the DDIs encompasses associations between multiple items and these will naturally generate rules of less support and confidence. It is found that, in the current case, the confidence criteria do not generate rules of interest. The FDA and WHO use PRR to monitor safety signals in their databases [129], [130]. The data mining algorithms using SRS databases use RRR (Relative Reporting Ratios) to quantify drug adverse event unexpectedness [131]. Further analysis of potential ADEs is done using a pre-defined disproportionality threshold.

In the current study, instead of confidence, PRR is used to enhance association strength and rule interestingness i.e., PRR is used instead of confidence for rule filtering. The ratio is a means which can be employed to summarize the extent of occurrence of an adverse event for a person taking a particular drug compared to the occurrence of the persons' taking some other drug(s). PRR is the amount of deviation of the joint probability of the drugs and AEs from statistical independence. Large value indicates that the drug-AE association is due to a reason and not by chance, i.e., due to a strong relation between drug and the AE. A PRR value greater than 1 indicates that the adverse event occurs in a person taking a particular drug compared to one taking other drugs. The PRRs are similar to the proportional mortality ratios in epidemiology which are based on the knowledge that the proportional frequency of adverse events reported to the UK Yellow Card system (The UK Adverse Event

Reporting System, similar to FAERS in US) is comparatively constant in spite of the significant increase in total reports under consideration [122].

Consider the following contingency table for PRR computation.

Table 4-5 is a 2x2 contingency table for a drug (X)-adverse event (Y) combination, in Spontaneous Reported data [131].

Table 4-5 Computation of PRR

	Adverse event (Y)	Not adverse event (Y)	Total
Using drug X	a	b	(a+b)
All other adverse events	c	d	(c+d)
Total	(a+c)	(b+d)	(a+b+c+d)

The table is constructed based on the number of reports of the combinations of interest. Drugs are reported as suspected or concomitant medications. The count of reports of a combination of interest can be based on all the reports for a drug, or only those where it is suspected as causal. When seeking interactions, all drugs are to be considered.

Here, 'a' is the observed number of reports of drug X and AE Y, and 'b' the number of reports without AE Y. The expected count is $\frac{(a+b)(a+c)}{a+b+c+d}$, assuming no association between X and Y.

Measures of disproportionality, is used to determine unexpectedness relative to the background of the rest of the database [131], [132].

$$\text{PRR} = \frac{a}{a+b} \Big/ \frac{c}{c+d}$$

Previous quantitative analysis studies involving the Spontaneous Reporting databases have used the proportional reporting ratios, and the characteristics of their performance post marketing surveillance and regulatory databases have been evaluated [132], [133].

With a PRR value greater than 2 for all adverse events occurring with frequency greater than 2, in the UK Yellow Card database, it was found that around 60% of the signals were of known adverse events. The PRR threshold used in the present study is set to 2 based on similar studies [39], [129], [132]. Large PRR value indicates that occurrence of a drug-AE combination is not by chance, but due to a reason behind the association. The support threshold to validate the adverse drug event association rules is set to 50. This is to accommodate the size of the database as well as to justify more frequent patterns. The threshold resulted in less variation in content instead of larger set of associations generated by smaller values. Support value of fifty balanced between the associations generated, size, drugs, and Adverse Events appearing in the associations.

4.4.5 Modification of the Apriori Algorithm

To suit the problem of finding associations between drugs and their adverse effects, PRR is computed for each drug in the ISR based on other drugs of the particular ISR. The Apriori is implemented based on PRR instead of confidence to find the associations. As stated in Section 4.4.4., PRR is used instead of confidence to avoid higher confidence generating frequent AEs like headache or nausea from being part of the rules since these AEs are usually not associated with drugs. Also, infrequent AEs generating lesser confidence found to be strongly associated with drugs will be part of the rules using PRR [45], [46],[103]. Finally, the rule base is populated with the associations containing at least two drug codes and an adverse event. The figure on the next page shows enhancements made on the Apriori algorithm.

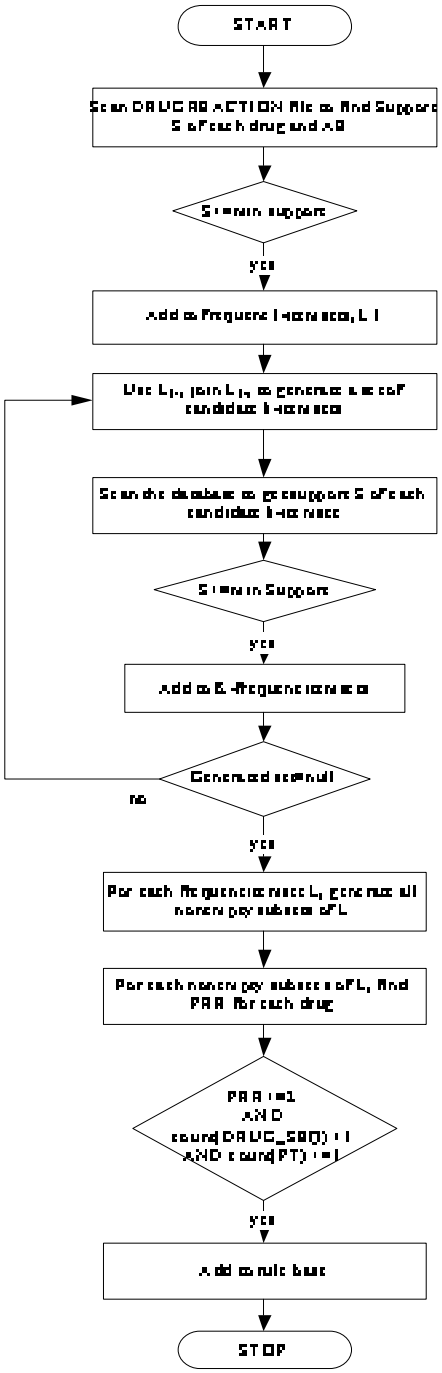


Figure 4-13 Modifying of the Apriori Algorithm

4.5 Integration of the Tool for DDI Identification

The different modules are integrated to form a tool for DDI identification. The major components include the FAERS database,data pre-processing module, the standard tool MedLEE, and the data mining module.

4.5.1 Medical Language Extraction and Encoding System MedLEE

The Architecture of MedLEE is shown in Figure 4-14.

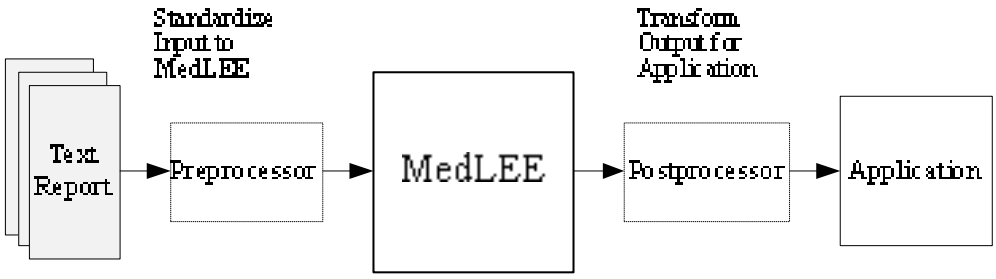


Figure 4-14 Architecture of MedLEE

MedLEE has six knowledge based components namely Abbreviation, WSD Rules, Lexicon, Grammar, Mapping, and Coding Table (Figure 4-15) [134]. The programming components are Pre-processor, Parser, Error Recovery, Phrase Regularization, and Encoder. The Pre-processor is capable of dividing the given text into segments as sections, paragraphs, sentences, and words. It also performs lexical lookup for the identification and classification of words and phrases. Lexicon determines the canonical forms of these outputs. The component also takes care of abbreviations using abbreviation table and performs Word Sense Disambiguation (WSD) based on context rules. Using grammar, the parser determines the initial structure form of a sentence. Syntactic and semantic rules specifying the language component structure and

their interpretation are included in the grammar. Remaining elements are value modifiers. Frame based structure is in the form of a list, the first element indicating type of information and the second corresponding to the value. The error recovery part discovers a parse on failure of the initial effort. Possible recovery is made by skipping words and segmenting texts. The component named phrase regularization composes multi-word phrase for sentences containing non-contiguous phrase. This process also separates each word of the phrase. The component uses a compositional table with the phrase and its corresponding compositional structure which is used to compose non-contiguous terms to form the one same as for the sentence with contiguous phrases. The compositional table is generated using MedLEE and gets created only once while the Lexicon for MedLEE is being compiled. The phrase regularization component in addition, uses domain knowledge to add information to the output that is implicitly in the domain. The Encoder uses an encoding table to add codes, like the UMLS code. Encoding is the process of mapping the primary finding to a coded form. Creation of the coding table is a tedious manual process as many of the entries are to be entered manually. Every table entry contains a target term generated by MedLEE and the corresponding UMLS term and code.

New MedLEE uses an improved method in which an automated process is employed to create a structured coding table of clinical information. Four different steps, namely term selection, term preparation, MedLEE parsing, and table generation are employed in the process. The different term variants are taken care of in the process as exemplified below.

Example: Synonyms associated with myocardial infarction.

C0027051^myocardial infarction|myocardial infarction

C0027051^myocardial infarction|heart attack

C0027051^myocardial infarction|myocardial infarction syndrome

C0027051^myocardial infarction|myocardial necrosis

C0027051^myocardial infarction|attack coronary

C0027051^myocardial infarction|necrosis myocardium

C0027051^myocardial infarction|myocardium necrosis syndrome

C0027051^myocardial infarction|cardiopathy necrotic

C0027051^myocardial infarction|infarction of heart

C0027051^myocardial infarction|infarction,myocardial

C0027051^myocardial infarction|infarction myocardial

In the process of drug name mapping, all the different codes will be mapped to a single code corresponding to the generic using MedLEE, i.e., all the codes 1018426025, 1018422348, 1017686798, 1018425841 in the example of section 3.4.5 are mapped to UMLS-CODE C0286651 for the generic ATROVASTATIN.

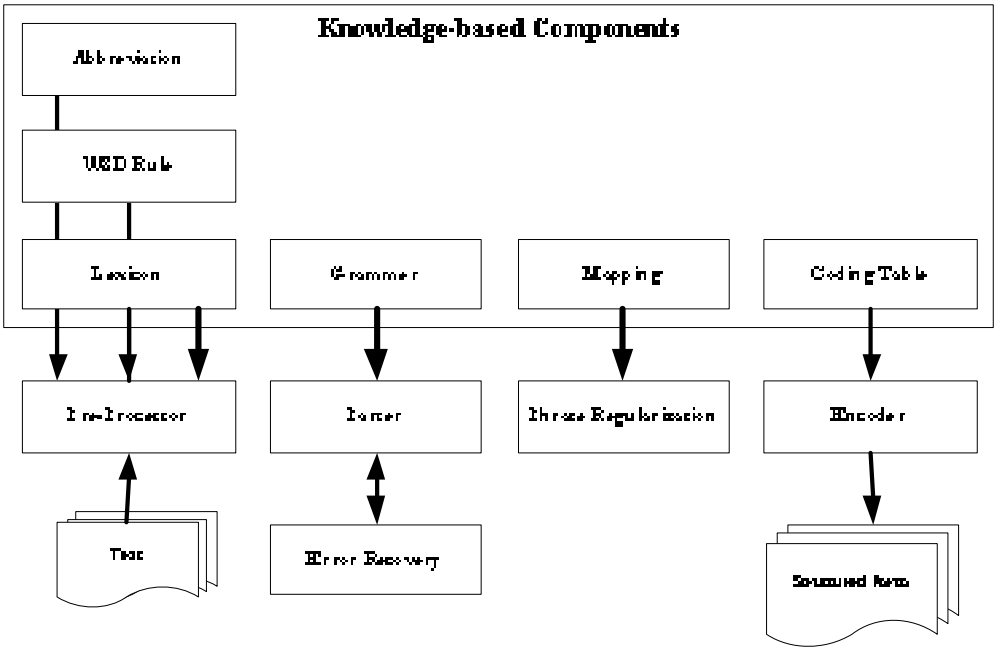


Figure 4-15 Components of MedLEE

The table 4-6 shows how MedLEE parses the sentence ***"New maculopapular rash on trunk."***

Table 4-6 MedLEE Parsing Illustration

Pre-processor	[new,maculopapular,rash,on,trunk,'.']
Parser	[problem,rash,[status,new],[descriptor,maculopapular],[bodyloc,trunk]]
Encoder	[problem,rash,[status,new],[descriptor,maculopapular],[bodyloc,trunk,[code,C0460005^trunk]],[code,C0241488^trunk maculopapular rash]
XML Translator	<problem v = "rash"> <status v = "new"></status> <descriptor v = "maculopapular"></descriptor> <bodyloc v = "trunk"> <code v = "C046005^trunk"</code> </bodyloc> <code v = "C0241488^trunk maculopapular rash" </code> </problem>

Some other Applications of MedLEE are briefly discussed in Appendix E.

Sample screenshots of MedLEE are shown in Figure 4-16 and 4-17.

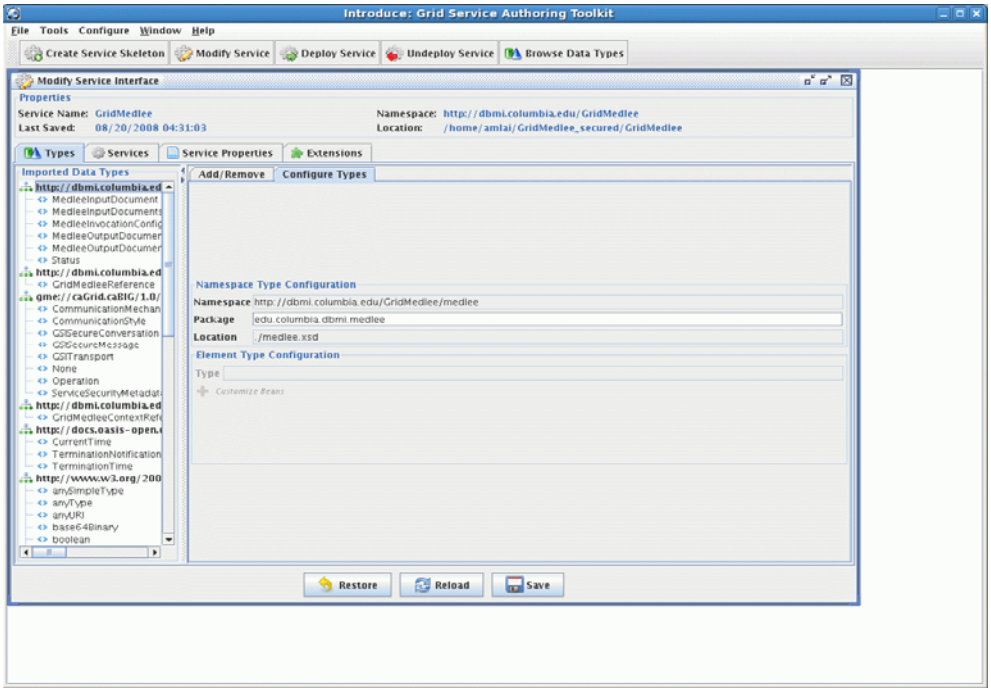


Figure 4-16 MedLEE: Sample Screen 1

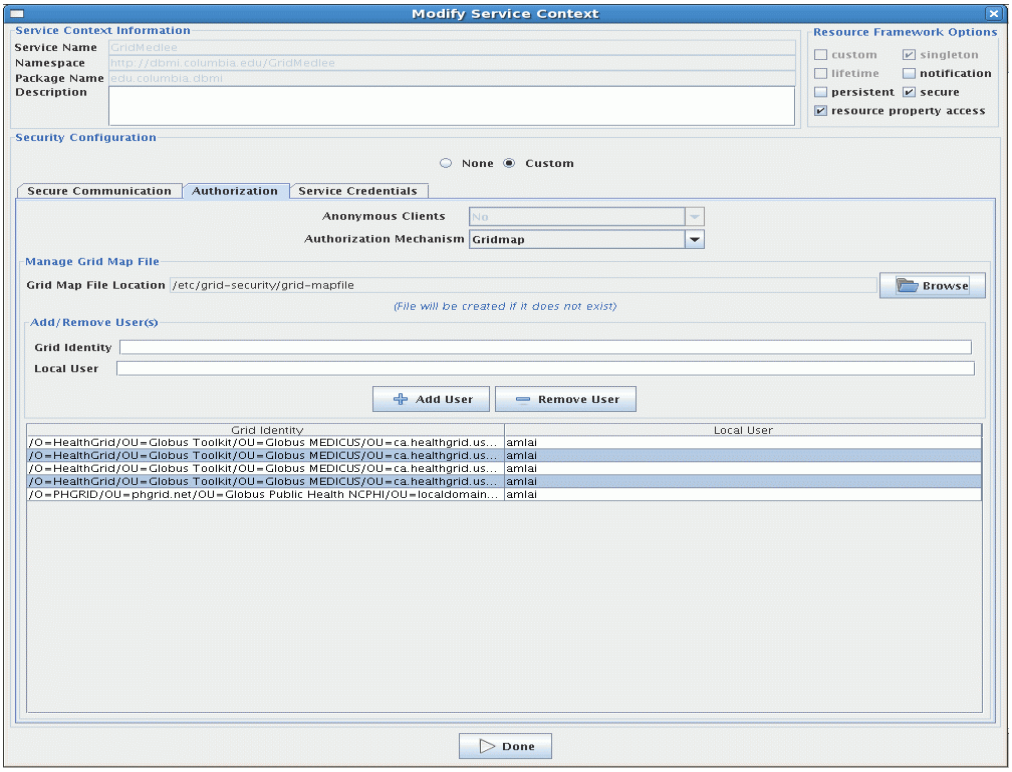


Figure 4-17 MedLEE: Sample Screen 2

4.5.2 An Integrated Tool for DDI Mining

The major components of the tool are the FAERS database, data pre-processing module, standard tool MedLEE and the data mining module (Figure 4-18). It also uses databases like UMLS and R_xNorm. The data mining module which performs the association rule mining, implemented in Python, needed 3.5 hrs.to run with Intel® Core™2 Duo processor at 3.0 GHz.

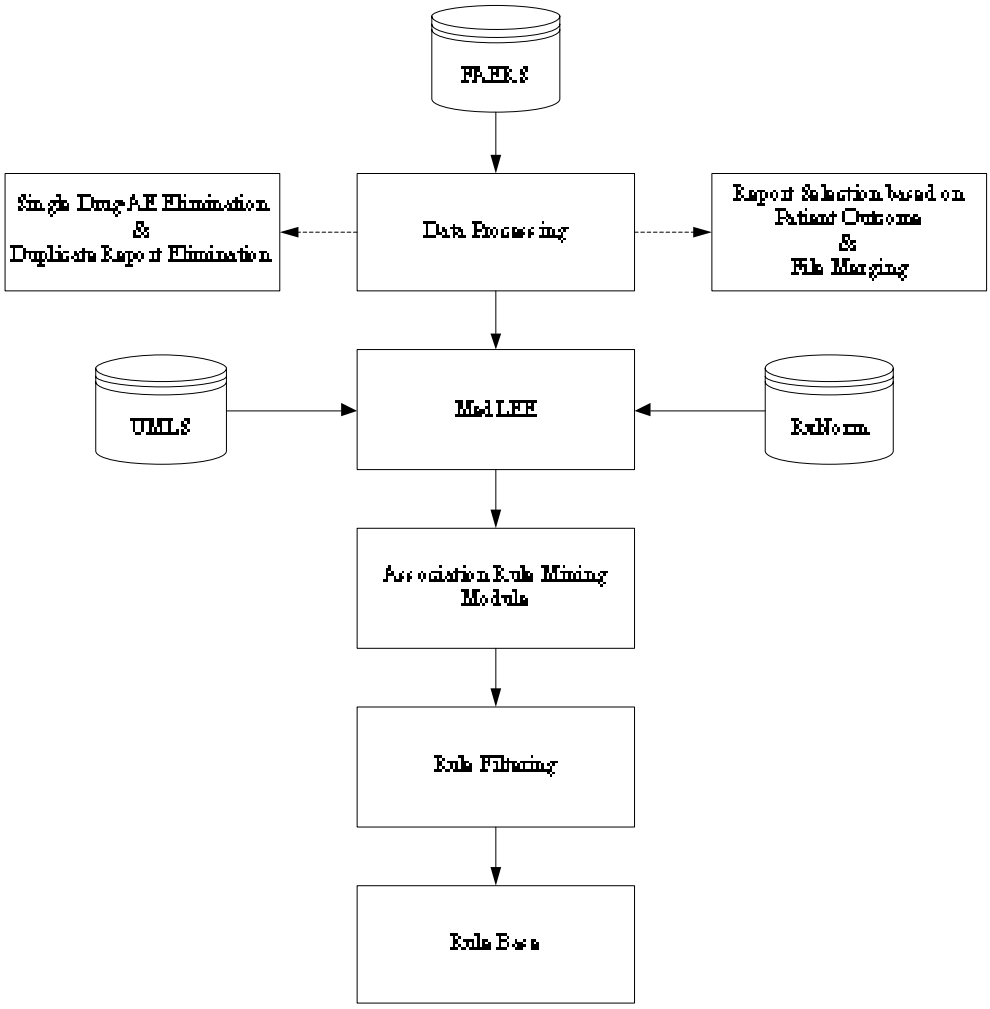


Figure 4-18 Architecture of DDI mining tool

The data pre-processing and data mining modules are integrated into a DDI mining tool using MATLAB.

4.6 Sample Results

Sample screen shots of the tool including the association rules generated are shown in Figures 4-19, 4-20, 4-21, 4-22 and 4-23.

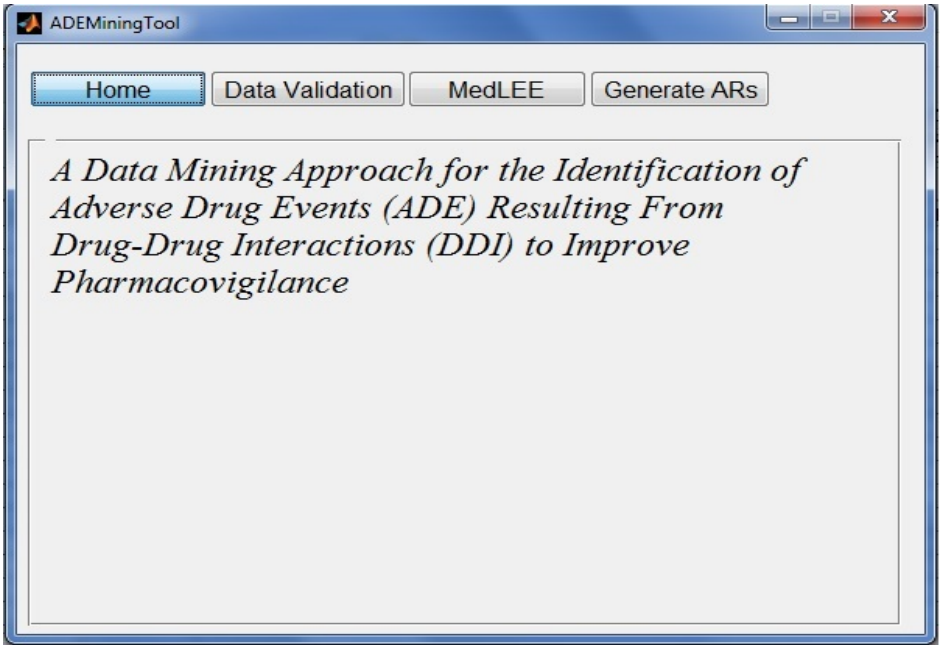


Figure 4-19 ADE Mining Tool- Home page

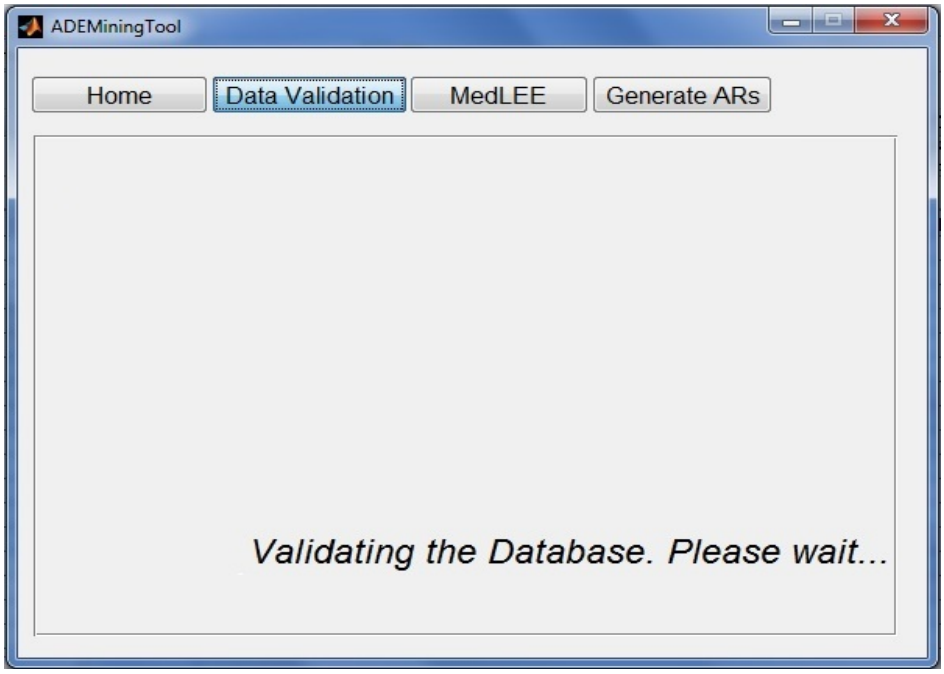


Figure 4-20 ADE Mining Tool- Data Validation page

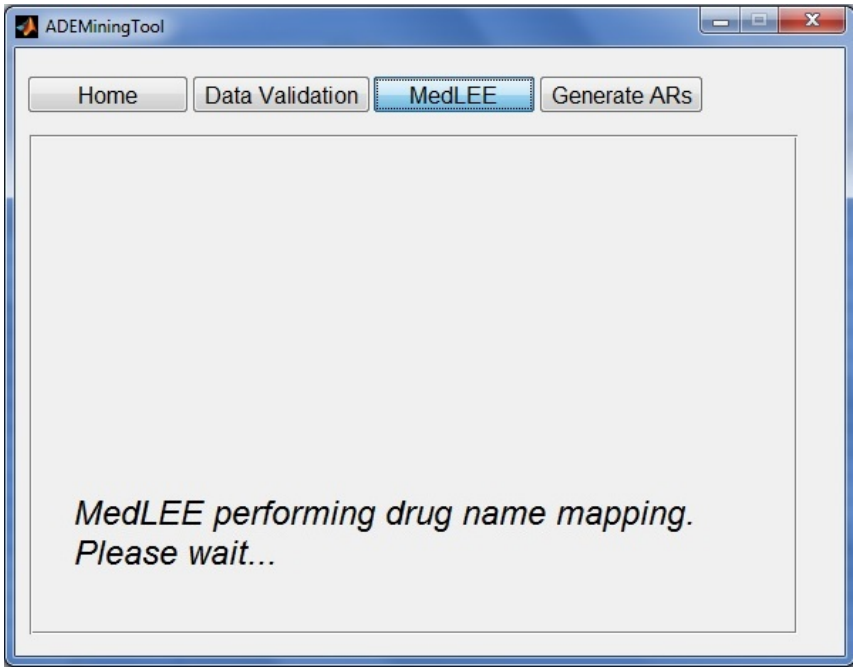
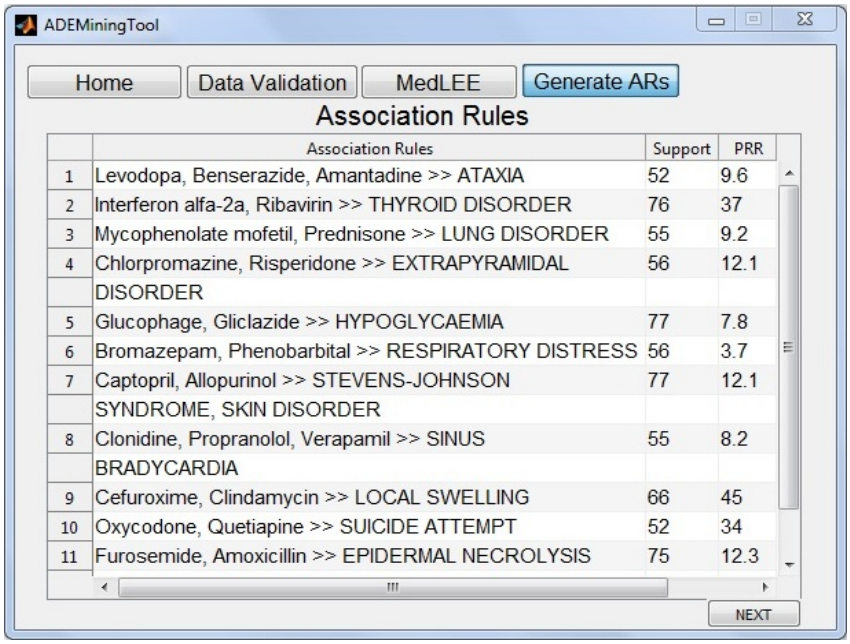


Figure 4-21 ADE Mining Tool- drug name mapping page



	Association Rules	Support	PRR
1	Levodopa, Benserazide, Amantadine >> ATAXIA	52	9.6
2	Interferon alfa-2a, Ribavirin >> THYROID DISORDER	76	37
3	Mycophenolate mofetil, Prednisone >> LUNG DISORDER	55	9.2
4	Chlorpromazine, Risperidone >> EXTRAPYRAMIDAL	56	12.1
	DISORDER		
5	Glucophage, Gliclazide >> HYPOGLYCAEMIA	77	7.8
6	Bromazepam, Phenobarbital >> RESPIRATORY DISTRESS	56	3.7
7	Captopril, Allopurinol >> STEVENS-JOHNSON	77	12.1
	SYNDROME, SKIN DISORDER		
8	Clonidine, Propranolol, Verapamil >> SINUS	55	8.2
	BRADYCARDIA		
9	Cefuroxime, Clindamycin >> LOCAL SWELLING	66	45
10	Oxycodone, Quetiapine >> SUICIDE ATTEMPT	52	34
11	Furosemide, Amoxicillin >> EPIDERMAL NECROLYSIS	75	12.3

Figure 4-22 ADE Mining Tool- Generate ARs page 1/2

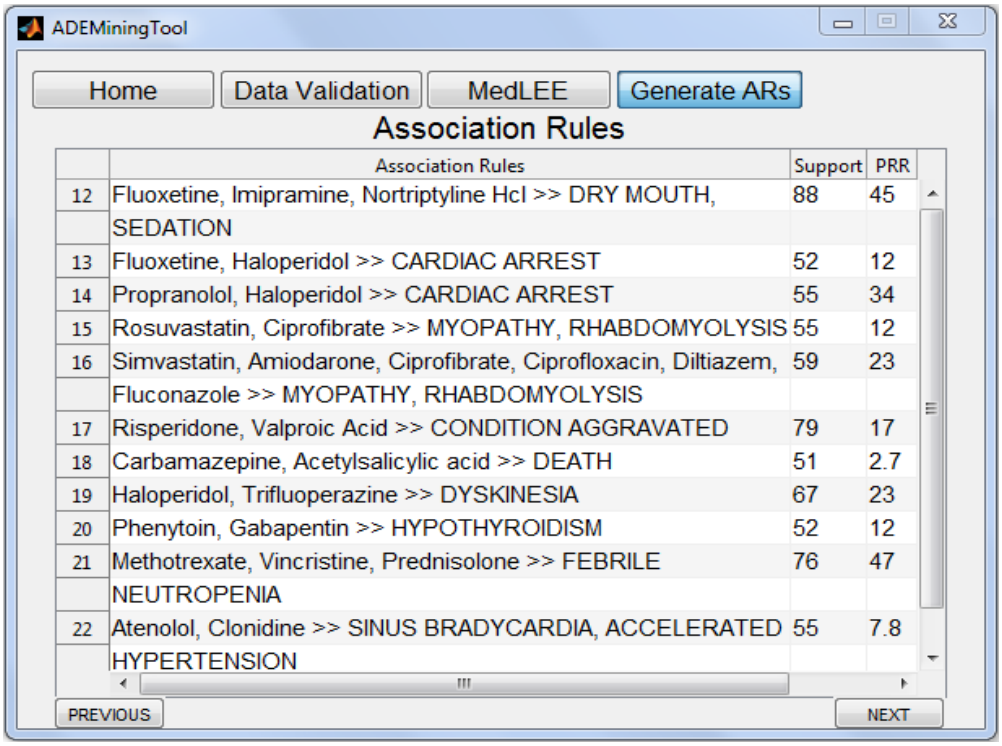


Figure 4-23 ADE Mining Tool- Generate ARs page 2/2

4.7 Method for Evaluating the Results

Micromedex is recommended as a reliable reference for the result evaluations by clinical subject matter experts [135]. This healthcare information system is developed by Truven Health Analytics. It is an online database that includes referenced information about drugs, toxicology, diseases, acute care, and alternative medicine for healthcare professionals to make informed clinical diagnosis and treatment decisions. The database is updated regularly.

Micromedex uses different types of online systems as described below:

CareNotes System

CareNotes is a patient education product with information about aspects of patient care, medical conditions, treatment, medications, and health, in up to 15 languages.

RED BOOK Online (Access through MICROMEDEX 2.0)

This component provides access to drug pricing and descriptive information for more than 2,00,000 active and deactivated FDA-approved prescription and over-the-counter (OTC) medications, nutraceuticals, bulk chemicals as well as some medical devices and supplies.

Formulary Advisor

This is an online formulary management tool intended to manage and update a hospital's formulary and communicate the most current formulary information throughout the facility.

PDR Electronic Library

The PDR Electronic Library makes access available to FDA approved drug information including drug interactions, side effects, recommended dosages, contraindications, etc.

The service of a clinical subject matter expert was employed in evaluating the association rules generated using the Micromedex.

4.8 Result Statistics and Analysis

Result statistics is prepared based on the database and the association rules generated. Finally, the association rules generated are categorized and analyzed manually for clinical and pharmacological interpretations.

4.8.1 Statistics

The statistics based on the database and generated association rules are summarized in Table 4-7, illustrated in Figure 4-24 and Figure 4-25.

Table 4-7 Database and Result Statistics

Total Number of reports	593679
Reports involving more than one drug	379184
Total distinct drug names	60189
Total generic drug codes after drug name mapping	17196
Total MedDRA coded AEs	11598
Average number of drugs per association	3.2
Average number of AEs per association	3.5
Total items per association	6.7
Total number of associations generated	1224

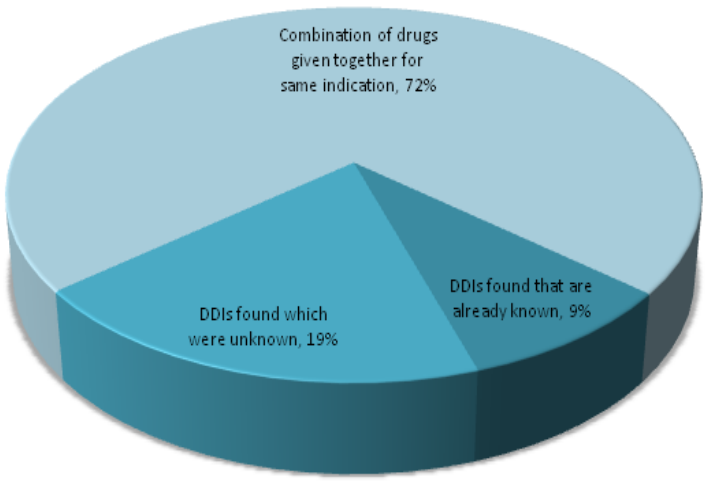


Figure 4-24 Drugs (in %) involved in the Associations

The statistics indicates that newly introduced drugs are more involved in creating DDIs. Most of the associations were composed of drugs that are usually given together like antibiotics or drugs given for the same disease like cancer or diabetes.

Association Statistics Based on formulated Rules

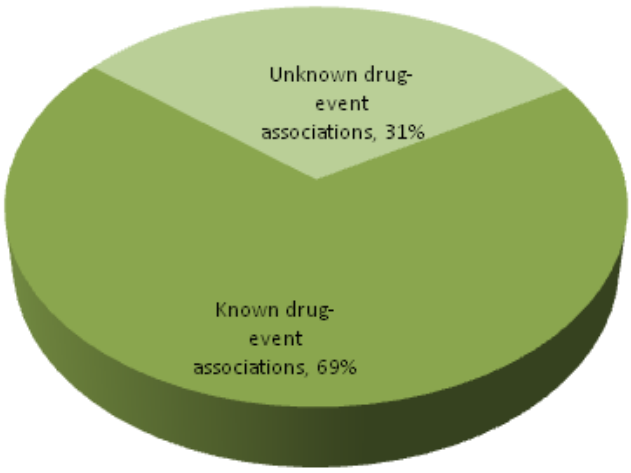


Figure 4-25 Percentages of Drug-AE Associations, Known and Unknown

The Adverse Events available in the Micromedex database were referred to categorize the generated associations into known and unknown. If an Adverse Reaction associated with an event is not caused by any of the individual drugs in the concerned association, it is considered to be a clear evidence of DDI.

1224 associations were generated, of which 844 (69%) were known and 380 (31%) were unknown. 69% known drug event associations substantiate the validity of the result and represents the research till date, *in vitro*, *in vivo* and *in silico*. The success and advantages of the methodology, is that unlike the usual

pharmacological and small cohort study methods done previously, it can be easily be repeated on an updated database to get new results.

4.8.2 Result Analysis and Significance

The generated associations are analyzed to study and validate the applicability of the results. Some of the drug and Adverse Effect combinations in the association rules generated are shown in Table 4-8.

Table 4-8 Sample DDI Associations by classification

NO	Classification	DDIs	Support	PRR
1	Known	Methotrexate,Prednisolone,Rituximab,Daunorubicin→FEBRILE NEUTROPENIA	65	52
2	Unknown	Methotrexate,Vincristine, Prednisolone→FEBRILE NEUTROPENIA	76	47
3	Known	Atenolol, Metformin→NAUSEA	55	8.1
4	Unknown	Amitryptiline, Dextropropoxyphene→ MEMORY IMPAIRMENT,FATIGUE,INSOMNIA,NAUSEA, ABNORMAL DREAMS	56	2467
5	Unknown	Esomeprazole,Pravastatin→DYSPNOEA	62	14
6	Unknown	Lisinopril,Pravastatin→ DYSPNOEA	57	4.5

Patterns of drugs and adverse effects combinations in the association rules generated can be illustrated using the above table of representative sample.

The first and thirdassociations in Table 4-8 describe DDIs that are known and caused by medications that are often prescribed together.

Atenolol and Metformin are known for their drug-drug interactions, and since both drugs are associated with nausea, nausea can be predicted in patients administered both the drugs.

The three drugs in the second association, Methotrexate, Vincristine, and Prednisolone are used to treat certain types of cancer, and febrile neutropenia is a known Adverse Drug Effect consequent to the treatment. Prednisolone is administered with other drugs in association 1 for the treatment of different conditions like inducing immune suppression, but is not known to cause the reported ADE. Based on previous ADE analysis, this demonstrates the potential of the method in finding important DDIs that are not recognized.

The extremely high PRR value of 2467 assigned to association 4 was further investigated. The corresponding reports from the FAERS database were examined and it was found that each of the concerned patients was taking 19 or more drugs. In spite of the numerous resulting combinations, the Amitryptiline and Dextropropoxyphene combination was more commonly associated than any other combinations. The indication is that the potential DDI enhances the effect of both on the brain; i.e. both the drugs influence the central nervous system (as per the AE indication in the association generated) and some of the AEs are due to the combination. This kind of analysis helps to direct research in a direction identifying drugs influencing a particular part they influence i.e. drug interaction of the pharmacokinetic type.

Pharmacology expert identified associations 5 and 6 in Table 4-8 are drugs known to be given together and known to interact, but the ADEs reported in the current study are clinically unknown. Lisinopril and Pravastatin are prescribed together for various medical conditions like diabetes, but the reported adverse drug event dyspnoea is not a known reaction of any of the drugs. Different antecedents Esomeprazole and Pravastatin cause the same adverse event, but the event is not known to be caused by any of the two drugs

when administered individually. This is a clear indication that the drugs have interacted to cause the event. When an AE is new to a group of drugs and is not reported of individual drugs, it indicates a DDI.

The results indicate that of the associations, 72% consists of drugs usually given together like antibiotics or given together for same condition as in cancer treatment (Figure 4-24). 69% of the associations generated were of known DDIs; each of the drugs was known to cause each of the AEs (Figure 4-25). 31% of the associations were unknown DDIs. The 69% represents all the previous research to identify DDIs. This again demonstrates the validity of the current method as the result matches with the previous findings. The newly proposed DDIs are considerably useful for pharmacologists and physicians who can exercise caution while prescribing drugs thus improving pharmacovigilance.

4.9 Summary and Conclusion

All the four challenges of pharmacovigilance are addressed effectively through present research. The Association Rule Mining method of finding the DDI associations with the efficiency enhancements addresses the problems of the accuracy of information retrieval and outdating of data mining tools. The centralized regulatory repository of the Adverse Events is mined and the problem of large volume data is addressed with an efficient pruning method without compromising accuracy. Positive and negative DDIs are mined alike and this can enhance therapeutic decision making in combination drug therapy, as in the case of combination of antibiotics like Clavulanic acid and Amoxicillin to combat a disease resistant bacteria or antihypertensive, to

control hypertension. Regulatory reporting has a legal aspect, but if the adverse events are reported accurately and the resultant drug-drug interactions mined through the program are reported to health care professionals, it will enhance pharmacovigilance significantly.

The program can be run on an updated database to derive new DDIs. As the DDIs are mined based on the generic names, the results are valid and applicable regardless of where the drugs are manufactured; i.e. the result can be applied to every drug marketed worldwide.

Chapter 5
CONCLUSION

Contents

An integrated tool is developed to mine drug-drug interactions using the Adverse Event database. The methodology and results obtained are analysed based on the contributions in the following sections. Possible future enhancements are also discussed.

5.1 Contributions

5.1.1 Technical Contributions

1. The method of Association Rule Mining is successfully implemented in finding DDIs using the Adverse Event database. Thus, the method is proven to be capable of mining association rules in the FAERS database and similar Spontaneous Reporting Databases. This is a considerable improvement over the existing methods as the method is especially easy to understand and the availability of an established method of implementation, the Apriori algorithm, enhances the ease of usage.
2. Unlike other methods to find DDIs like pharmacological, small cohort and other *in vitro* and *in vivo* techniques, the application can be

repeated on an updated database to find new drug-AE associations. Repeatability of the application is its best advantage.

3. The FAERS database found to be a source of signals of Adverse Events is now proven to be a source of DDIs also. The current work establishes the FAERS database as a source of DDIs.
4. The methodology of mapping the drug names implemented in the current method can be used to improve current applications and can be used in future research to improve the scope and applicability of the data mining algorithms in pharmacoinformatics.
5. The hashing and pruning techniques used in the current solution establish the previous methods of improving the Apriori algorithm.
6. The four challenges of pharmacovigilance, namely analysis of medical information, inefficiency in reporting to regulatory authorities, unmanageability of the volume of data, and inefficiency of the obsolete applications, are efficiently addressed through the present application.
7. The method implemented is established to be capable of handling large sized database and hence addresses the pharmacovigilance challenge of data volume.
8. The increase in the efficiency of the algorithm and the mapping of drug names overcome the inefficiency of existing data and text mining methods and increase the applicability of the program, considerably improving pharmacovigilance worldwide. As the rules are formulated on generic names, the result can be used by physicians and pharmacologists worldwide.

9. Excavating the drug-drug interaction associations is a contribution in the analysis of adverse drug events and improves pharmacovigilance by enhancing drug safety considerably.
10. The current work establishes the significance of timely reporting of drug interactions to the authorities. Efficient reporting of DDIs to the authorities will lead to updates in the current drug interaction databases like Micromedex which is a widely accepted resource for DDI exploration, and this will also improve pharmacovigilance significantly.

5.1.2 Social Contributions

1. DDIs are present irrespective of age, sex, and geography. Potential repercussions of drug interactions include patient injury up to disability and death, disease progression, lost wages, increase in healthcare costs, and research overheads. The identified DDIs help to reduce the above considerably.
2. The method outputs the results in the form of association rules between drugs and their adverse events. The drugs are associated based on generics rather than brand names. This increases the applicability of the results extensively. The result is applicable to any drug marketed in any country and the obvious social contribution achieved by this is substantial.
3. The increase in healthcare costs poses a financial burden in most of the families and federal governments alike depending on policies in different countries. This burden can be reduced by applying results of the current study.

4. The increase in research overhead is a consequence of withdrawal of drugs from market due to DDIs. If a generic is identified to be interacting with another drug in the current problem, the research overhead can be considerably reduced by pre-market surveillance.
5. Epidemiology studies indicate that the elderly are more affected by DDIs as they are administered more prescription medications due to comorbidities. Quality of life can be improved with DDI identification and subsequent enforcement of standards. Rather than living longer with interacting drugs, DDIs provide a better quality of life along with longevity.

5.2 Future Works

Lowering the support can produce rarer associations but the number of comparisons to be made increases the complexity enormously. Algorithmic improvement is to be made to combat this problem.

Since the drug names were mapped irrespective of dosage and form, the method reduced the complexity and increased the applicability. But, a study based on severity of adverse events pertaining to dosage levels is possible by algorithmic improvements.

Selecting report with very large and small PRR values, analysis can be made regarding other factors affecting DDIs.

By analyzing the results, drug discovery can be enhanced through the categorization of drugs involved in the generated DDI associations into pharmacodynamic and pharmacokinetic types.

LIST OF PUBLICATIONS

1. Sindhu M.S. and Dr.B. Kannan. “Bayesian Networks: Data Modeling in Bioinformatics and cybercrime detection”. *Proceedings of the UGC Sponsored National Seminar on Cyber Criminology*, March 2011, BPC College, Piravom. ISBN 978-81-009720-4-8.
2. Sindhu M.S. and Kannan B., “Investigating the Factors affecting drug-drug interactions,” *International Journal of Pharma and Biosciences*. (Accepted for Publication)
3. Sindhu M.S. and Kannan.B., ”Detecting Signals of Drug-Drug Interactions using Association Rule Mining Methodology.”, *International Journal of computer Science and Information Technologies. (*Accepted for Publication.)

Appendix A

FAERS MANDATORY REPORTING FORM

Form Approved: OMB No. 09 10-029 1, Expires 12/31/11
See OMB statement on reverse.

U.S. Department of Health and Human Services
Food and Drug Administration

MEDWATCH

FORM FDA 3500A (6/10)

For use by user-facilities, importers, distributors and manufacturers for MANDATORY reporting

Page 1 of ____

Mfr Report #
UF/Importer Report #
FDA Use Only

A. PATIENT INFORMATION

1. Patient Identifier	2. Age at Time of Event: or Date of Birth:	3. Sex	4. Weight
In confidence		☐ Female ☐ Male	____ lbs or ____ kgs

B. ADVERSE EVENT OR PRODUCT PROBLEM

1. ☐ **Adverse Event** and/or ☐ **Product Problem** *(e.g., defects/malfunctions)*

2. **Outcomes Attributed to Adverse Event** *(Check all that apply)*

☐ Death: _____ *(mm/dd/yyyy)*
☐ Life-threatening
☐ Hospitalization - initial or prolonged
☐ Required Intervention to Prevent Permanent Impairment/Damage (Devices)
☐ Disability or Permanent Damage
☐ Congenital Anomaly/Birth Defect
☐ Other Serious (Important Medical Events)

3. Date of Event *(mm/dd/yyyy)*	4. Date of This Report *(mm/dd/yyyy)*

5. **Describe Event or Problem**

(Continue on page 3)

6. **Relevant Tests/Laboratory Data, Including Dates**

(Continue on page 3)

7. **Other Relevant History, Including Preexisting Medical Conditions** *(e.g., allergies, race, pregnancy, smoking and alcohol use, hepatic/renal dysfunction, etc.)*

(Continue on page 3)

Submission of a report does not constitute an admission that medical personnel, user facility, importer, distributor, manufacturer or product caused or contributed to the event.

C. SUSPECT PRODUCT(S)

1. **Name** *(Give labeled strength & mfr/labeler)*
#1
#2

2. Dose, Frequency & Route Used	3. Therapy Dates *(If unknown, give duration)* from/to (or best estimate)
#1	#1
#2	#2

4. Diagnosis for Use *(Indication)*	5. Event Abated After Use Stopped or Dose Reduced?
#1	#1 ☐ Yes ☐ No ☐ Doesn't Apply
#2	#2 ☐ Yes ☐ No ☐ Doesn't Apply

6. Lot #	7. Exp. Date	8. Event Reappeared After Reintroduction?
#1	#1	#1 ☐ Yes ☐ No ☐ Doesn't Apply
#2	#2	#2 ☐ Yes ☐ No ☐ Doesn't Apply

9. **NDC# or Unique ID**

10. **Concomitant Medical Products and Therapy Dates** *(Exclude treatment of event)*

(Continue on page 3)

D. SUSPECT MEDICAL DEVICE

1. **Brand Name**
2. **Common Device Name**
3. **Manufacturer Name, City and State**

4. Model #	Lot #	5. Operator of Device
Catalog #	Expiration Date *(mm/dd/yyyy)*	☐ Health Professional ☐ Lay User/Patient ☐ Other:
Serial #	Other #	

6. If Implanted, Give Date *(mm/dd/yyyy)*	7. If Explanted, Give Date *(mm/dd/yyyy)*

8. **Is this a Single-use Device that was Reprocessed and Reused on a Patient?** ☐ Yes ☐ No

9. **If Yes to Item No. 8, Enter Name and Address of Reprocessor**

10. **Device Available for Evaluation?** *(Do not send to FDA)*
☐ Yes ☐ No ☐ Returned to Manufacturer on: _____ *(mm/dd/yyyy)*

11. **Concomitant Medical Products and Therapy Dates** *(Exclude treatment of event)*

(Continue on page 3)

E. INITIAL REPORTER

1. Name and Address	Phone #

2. Health Professional?	3. Occupation	4. Initial Reporter Also Sent Report to FDA
☐ Yes ☐ No		☐ Yes ☐ No ☐ Unk.

MEDWATCH

FORM FDA 3500A (6/10) *(continued)* Page 2 of ____

FDA USE ONLY

F. FOR USE BY USER FACILITY/IMPORTER *(Devices Only)*

1. **Check One** ☐ User Facility ☐ Importer
2. **UF/Importer Report Number**
3. **User Facility or Importer Name/Address**
4. **Contact Person**
5. **Phone Number**
6. **Date User Facility or Importer Became Aware of Event** *(mm/dd/yyyy)*
7. **Type of Report** ☐ Initial ☐ Follow-up # _____
8. **Date of This Report** *(mm/dd/yyyy)*
9. **Approximate Age of Device**
10. **Event Problem Codes** *(Refer to coding manual)*
Patient Code ____ - ____ - ____
Device Code ____ - ____ - ____
11. **Report Sent to FDA?** ☐ Yes _____ *(mm/dd/yyyy)* ☐ No
12. **Location Where Event Occurred** ☐ Hospital ☐ Home ☐ Nursing Home ☐ Outpatient Treatment Facility ☐ Outpatient Diagnostic Facility ☐ Ambulatory Surgical Facility ☐ Other: _____ *(Specify)*
13. **Report Sent to Manufacturer?** ☐ Yes _____ *(mm/dd/yyyy)* ☐ No
14. **Manufacturer Name/Address**

G. ALL MANUFACTURERS

1. **Contact Office - Name/Address** ***(and Manufacturing Site for Devices)***
2. **Phone Number**
3. **Report Source** *(Check all that apply)* ☐ Foreign ☐ Study ☐ Literature ☐ Consumer ☐ Health Professional ☐ User Facility ☐ Company Representative ☐ Distributor ☐ Other: _____ _____ _____ _____
4. **Date Received by Manufacturer** *(mm/dd/yyyy)*
5. (A)NDA # _____ IND # _____ STN # _____ PMA/510(k) # _____ Combination Product ☐ Yes Pre-1938 ☐ Yes OTC Product ☐ Yes
6. **If IND, Give Protocol #**
7. **Type of Report** *(Check all that apply)* ☐ 5-day ☐ 30-day ☐ 7-day ☐ Periodic ☐ 10-day ☐ Initial ☐ 15-day ☐ Follow-up # ____
8. **Adverse Event Term(s)**
9. **Manufacturer Report Number**

H. DEVICE MANUFACTURERS ONLY

1. **Type of Reportable Event** ☐ Death ☐ Serious Injury ☐ Malfunction ☐ Other: _____
2. **If Follow-up, What Type?** ☐ Correction ☐ Additional Information ☐ Response to FDA Request ☐ Device Evaluation
3. **Device Evaluated by Manufacturer?** ☐ Not Returned to Manufacturer ☐ Yes ☐ Evaluation Summary Attached ☐ No *(Attach page to explain why not)* or provide code: _____
4. **Device Manufacture Date** *(mm/yyyy)*
5. **Labeled for Single Use?** ☐ Yes ☐ No
6. **Evaluation Codes** *(Refer to coding manual)*
Method ____ - ____ - ____ - ____
Results ____ - ____ - ____ - ____
Conclusions ____ - ____ - ____ - ____
7. **If Remedial Action Initiated, Check Type** ☐ Recall ☐ Repair ☐ Replace ☐ Relabeling ☐ Notification ☐ Inspection ☐ Patient Monitoring ☐ Modification/Adjustment ☐ Other: _____
8. **Usage of Device** ☐ Initial Use of Device ☐ Reuse ☐ Unknown
9. **If action reported to FDA under 21 USC 360i(f), list correction/removal reporting number:**
10. ☐ **Additional Manufacturer Narrative** and / or 11. ☐ **Corrected Data**

The public reporting burden for this collection of information has been estimated to average 66 minutes per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to:

Department of Health and Human Services
Food and Drug Administration
Office of Chief Information Officer
1350 Piccard Drive, Room 400
Rockville, MD 20850

Please DO NOT RETURN this form to this address.

OMB Statement:
"An agency may not conduct or sponsor, and a person is not required to respond to, a collection of information unless it displays a currently valid OMB control number."

(CONTINUATION PAGE)
For use by user-facilities,
importers, distributors, and manufacturers
for MANDATORY reporting

MEDWATCH

FORM FDA 3500A (6/10) *(continued)*

Page 3 of____

B.5. **Describe Event or Problem** *(continued)*
B.6. **Relevant Tests/Laboratory Data, Including Dates** *(continued)*
B.7. **Other Relevant History, Including Preexisting Medical Conditions** *(e.g., allergies, race, pregnancy, smoking and alcohol use, hepatic/renal dysfunction, etc.) (continued)*
Concomitant Medical Products and Therapy Dates *(Exclude treatment of event) (For continuation of C.10 and/or D.11; please distinguish)*
Other Remarks

Appendix B

MedDRA, THE MEDICAL DICTIONARY FOR REGULATORY ACTIVITIES

Website:http://www.meddramsso.com/

MedDRA is a medical terminology used to classify adverse event information associated with the use of biopharmaceuticals and other medical products like medical devices and vaccines. Coding these data to a standard set of MedDRA terms enables the health authorities and the pharmaceutical industry to easily exchange and analyse data related to the safe use of medical products.

MedDRA, developed by the International Conference on Harmonisation (ICH), is continuously enhanced to meet the emerging requirements of its consumers, including regulators and industry worldwide. ICH has created a governance structure to nurture and protect the integrity of MedDRA. The ICH MedDRA Management Board, appointed by the ICH Steering Committee, has overall responsibility for the direction of MedDRA, and controls and coordinates all the activities of the MedDRA Maintenance and Support Services Organization, MSSO.

The MSSO - Maintenance and Support Services Organization The Organisation serves as the repository, maintainer, and distributor of MedDRA as well as the source for the most recent information regarding MedDRA and its application within the l industry and regulators. MedDRA subscribers submit proposed changes to the terminology. The MSSO includes an

international group of physicians who review all proposed subscriber changes and provide a timely response directly to the subscriber.

MedDRA terms

The MedDRA dictionary is organized by System Organ Class (SOC), divided into High-Level Group Terms (HLGT), High-Level Terms (HLT), Preferred Terms (PT) and finally into Lower-Level Terms (LLT). The MedDRA dictionary also includes Standardized MedDRA Queries (SMQs). SMQs are groupings of terms that relate to a defined medical condition or area of interest.

Appendix C

RxNorm Technical Documentation

Website: http://www.nlm.nih.gov/research/umls/rxnorm/overview.html

RxNorm is composed of two things:

1. A normalized naming system for branded and generic drugs and
2. A tool for supporting semantic interoperation linking pharmacy knowledge base systems and drug terminologies.

Purpose of RxNorm

RxNorm aims to allow computer systems to communicate drug-related information efficiently and unambiguously. Hospitals, pharmacies, and other organizations use computer systems for recording and processing drug information. As these systems use many different sets of drug names, it is difficult for one system to communicate with another. To combat this problem, RxNorm provides normalized names and unique identifiers for medicines and drugs.

Scope of RxNorm

RxNorm contains the names of prescription and many over-the-counter drugs available in the United States. RxNorm includes generic and branded clinical Drugs and drug packs:

- Clinical drugs - pharmaceutical products given to (or taken by) a patient with therapeutic or diagnostic intent.

- Drug packs - packs that contain multiple drugs, or drugs designed to be administered in a specified sequence

The creation of RxTerms follows the following principles:

1. Unambiguity – the use of the interface terminology should lead to the unambiguous identification of the clinical drug being prescribed. In practice, this means it should be able to produce the RXCUI of either the SCD or SBD from what the user enters
2. Naturalness – the names it uses and the way they are organized should conform to the usual prescribers' vocabulary and prescription writing habits
3. Efficiency – the user should be able to identify the drugs they are prescribing quickly with a small number of mouse clicks and keystrokes
4. Coverage – there should be adequate coverage of drugs used in the prescribing environment. At present, we have focused on currently available drugs in the U.S.

Preservation of Content and Meaning from Source Vocabularies

R_xNorm reflects and preserves the meanings, concept names, and relationships from its source vocabularies in the same way that the Metathesaurus preserves content and meaning.

R_xNorm contains source vocabularies produced by different copyright holders.

Source Vocabularies

The following source vocabularies are included in RxNorm:

SAB	Source Name
GS	Gold Standard Drug Database
MDDB	Medi-Span Master Drug Data Base
MMSL	Multum MediSource Lexicon
MMX	Micromedex RED BOOK
MSH	Medical Subject Headings (MeSH)
MTHFDA	FDA National Drug Code Directory(NDC)
MTHSPL	FDA Structured Product Labels
NDDF	FDB MedKnowledge (formerly NDDF Plus)
NDFRT	Veterans Health Administration National Drug File - Reference Terminology
SNOMEDCT	SNOMED Clinical Terms (drug information)
VANDF	Veterans Health Administration National Drug File

Ambiguity in RxNorm Source Data

The variety of source providers to the RxNorm dataset, along with their individual schedules for updating and modifying their datasets, has created the possibility of atom strings coming from source providers that look similar, but have different meanings. It is common for these ambiguous strings to be placed in separate concepts within RxNorm in order to maintain the separate level of meaning for each string.

The ambiguity is a result of the inability to resolve the correct meaning for a given string from a particular source. For example, attributes of an atom string may conflict with the string of the atom. One attribute may refer to an "Oral Tablet", while another may refer to "Extended Release Tablet", while the string that names the atom might disagree with these attributes. This could result in the string in question being placed in a concept separate from similar looking string(s) whose meaning is more clearly apparent in the name.

Another case of ambiguity could involve different quantities of a particular drug. In the case of an inhaler, a source may provide a string for a particular brand of inhaler, without specifying the number of actuations in the container. Another source may provide this information for a similar drug string, and research at NLM will confirm the different sizes of containers for the particular brand, thus making the first string ambiguous in meaning.

As drug source providers have refined and updated their datasets over time, new information has been provided about existing drug strings, making the old information obsolete and ambiguous. R_xNorm maintains this old data as R_xNorm (SAB=RXNORM) atoms of term type 'OCD' in RXNCONSOOCD.RRF. Attribute and semantic type information for these data are kept in RXNSATOCD.RRF and RXNSTYOCD.RRF, respectively.

NLM strives to review these ambiguous strings on a regular basis in both the R_xNorm and UMLS data sets. In the UMLS, relationships are created between ambiguous concepts and a unique name is created for each string with multiple meanings.

R_xNorm normal forms (e.g. SCD, SBD, GPCK, BPCK) are created when the meaning of the source string is clear at the appropriate level of abstraction. In this sense, SAB=RXNORM normal form atoms are never ambiguous.

Drug Dose Forms and the Modified Format for Semantic Clinical Drug (SCD) and Semantic Branded (SBD) Normal Forms

To accurately represent factors of time, measures, or strength for certain dose forms within R_xNorm, an additional format for SCD and SBD forms has

been utilized along with a more succinct set of dose forms for certain categories of forms. The factor of time, measure, or strength can now be represented by a Quantity Factor appended to the beginning of the SCD or SBD name. This quantity factor consists of a number followed by a unit measure. The number might represent strength, a unit of time, or a quantity of dosages depending upon the dosage form in use at the time. This quantity factor is stored as an attribute to the RxNorm SAB atom as the value of the ATN='RXN_QUANTITY' in RXNSAT.RRF. Some dose forms have been retired and replaced by more generic dose form names, allowing for greater flexibility in representing drug names. Other dose forms have been added.

The following dose forms have been replaced by the more general dose form shown to the right of the name. The Quantity number is variable according to the clinical drug being represented.

Old Dose Form	New Dose Form	Quantity number	Quantity Unit	Quantity Factor Required
12 hour Extended Release Capsule	Extended Release Capsule	12	HR	No
12 hour Extended Release Tablet	Extended Release Tablet	12	HR	No
24 hour Extended Release Capsule	Extended Release Capsule	24	HR	No
24 hour Extended Release Tablet	Extended Release Tablet	24	HR	No
16 Hour Transdermal Patch	Transdermal Patch	16	HR	No
24 Hour Transdermal Patch	Transdermal Patch	24	HR	No
72 Hour Transdermal Patch	Transdermal Patch	72	HR	No
Biweekly Transdermal Patch	Transdermal Patch	84	HR	No
Weekly Transdermal Patch	Transdermal Patch	168	HR	No

Example RxNorm forms before the data transformation: Bupropion 300 MG 24 hour Extended Release Tablet [Wellbutrin]

Clonidine 0.00833 MG/HR Weekly Transdermal Patch

And after the data transformation:

24 HR Bupropion 300 MG Extended Release Tablet [Wellbutrin]

168 HR Clonidine 0.00833 MG/HR Transdermal Patch

Example forms are shown below:

3 ML Insulin, Aspart, Human 100 UNT/ML Prefilled Syringe

2 ML hyaluronate 10 MG/ML Prefilled Syringe [Hyalgan]

Examples of RXNSAT.RRF entries for an RXCUI (ATN="RXN_QUANTITY")

104420|||2701594|AUI|104420|||RXN_QUANTITY|RXNORM|24 HR|N||

104873|||2701596|AUI|104873|||RXN_QUANTITY|RXNORM|72 HR|N||

Syrup and Elixir Dose Forms

In order to reduce the discrepancies related to NDC codes on R_xNorm forms, beginning with the March 2008 release, R_xNorm has chosen to use the Oral Solution dose form in lieu of the Syrup and Elixir dose forms. Existing R_xNorm normal forms (SCD, SBD, SCDF, SBDF) using Syrup and Elixir were transformed into R_xNorm normal forms using the Oral Solution dose form. In some cases, this process resulted in new R_xNorm forms being created, in others, an R_xNorm form already existed and was reused. The R_xNorm normal forms that contained Syrup and Elixir dose forms have been archived.

Extended Release Products

Specified (with quantity factor) and unspecified (without quantity factor) versions of extended release products are connected by the

"has_quantified_form" and "quantified_form_of" reciprocal relationship attributes (RELA) in the RXNREL.RRF file. Specified forms are connected to a single unspecified form, while an unspecified form may be connected to multiple specified forms (8 HR, 12 HR, etc.). These relationships exist from SCD to SCD and from SBD to SBD only.

Examples:

RXCUI (1)	TTY (1)	STR (1)	RELA	RXCUI (2)	TTY (2)	STR (2)
1099759	SCD	24 HR Valproic Acid 250 MG Extended Release Capsule	has_quantified_form	1099760	SCD	Valproic Acid 250 MG Extended Release Capsule
1014577	SBD	12 HR cetirizine hydrochloride 5 MG / Pseudoephedrine Hydrochloride 120 MG Extended Release Tablet [Cetiri-D]	has_quantified_form	1014578	SBD	cetirizine hydrochloride 5 MG / Pseudoephedrine Hydrochloride 120 MG Extended Release Tablet [Cetiri-D]
849487	SCD	Naproxen sodium 825 MG Extended Release Tablet	quantified_form_of	849484	SCD	24 HR Naproxen sodium 825 MG Extended Release Tablet
1098612	SBD	lamotrigine 300 MG Extended Release Enteric Coated Tablet [Lamictal]	quantified_form_of	1098610	SBD	24 HR lamotrigine 300 MG Extended Release Enteric Coated Tablet [Lamictal]

FAERS Sample Records

DEMOGRAPHIC file

ISR	8226072	7919673
CASE	8479397	8188761
I_F_COD	—	F
FOLL_SEQ	5	1
IMAGE	8226072-4	7919673-1
EVENT_DT	20110123	20110816
MFR_DT	20120305	20111031
FDA_DT	20120319	20111115
REPT_COD	EXP	EXP
MFR_NUM	214715	2011-3891
MFR_SNDR	LEO PHARMA A/S	TERCICA INC
AGE	41	16
AGE_COD	YR	YR
GNDR_COD	F	M
E_SUB	N	N
WT	38.3	33
WT_COD	KG	KG
REPT_DT	20120313	20111108
OCCP_COD	OT	MD
DEATH_DT		
TO_MFR		
CONFID		
REPORTER_COUNTRY	INDIA	UNITED STATES

DRUG file

ISR	DRUG_SEQ	ROLE_COD	DRUGNAME	VAL_VBM	ROUTE	DOSE_VBM	DECHAL	RECHAL	LOT_NUM	EXP_DT	NDA_NUM
8024444	1018459863	PS	VICTRELIS	1	ORAL	800 MG; Q8H; PO	D	D	1HCE005	20130331	202258
8024677	1018451717	PS	BUPIVACAINE HCL	1	TRANSPLACENTAL	TRANSPLACENTAL	D	D	04528 EV	20130301	71810
8028041	1018462833	PS	VIVITROL	1	INTRAMUSCULAR	(INTRAMUSCULAR)	D	D	2011-005P	20131001	21897

INDICATIONS file

ISR	DRUG_SEQ	INDI_PT
7805388	1017738582	PRODUCT USED FOR UNKNOWN INDICATION

OUTCOME File

ISR	OUTC_COD
7805388	OT
7805388	HO
7810716	HO

REACTION file

ISR	PT
7805388	COLON CANCER
7805388	MELAENA
7810716	CONFUSIONAL STATE

REPORT SOURCE File

ISR	RPSR_COD
7810716	HP
7811037	CSM
7811037	FGN

THERAPY file

ISR	DRUG_SEQ	START_DT	END_DT	DUR	DUR_COD
7805388	1017738577	20100316	20100423	3	DAY
7805388	1017738578	20100315	20100315	5	DAY

Appendix E

MedLEE Overview

Overview of MedLEE

Extracts, structures, and encodes clinical information in narrative patient reports

Comprehensive coverage

Can be used for diverse clinical applications

Development started in 1991

Used at Columbia University Medical Centre since 1995

Numerous independent evaluations

Applications using MedLEE

Biosurveillance, Syndromic surveillance, Adverse Drug Event detection, Decision Support, Clinical Research, Clinical Trials, Quality Assurance, Automated Encoding, Patient Management, Data mining – finding trends and associations, Linking patient record to the literature, Summarization.

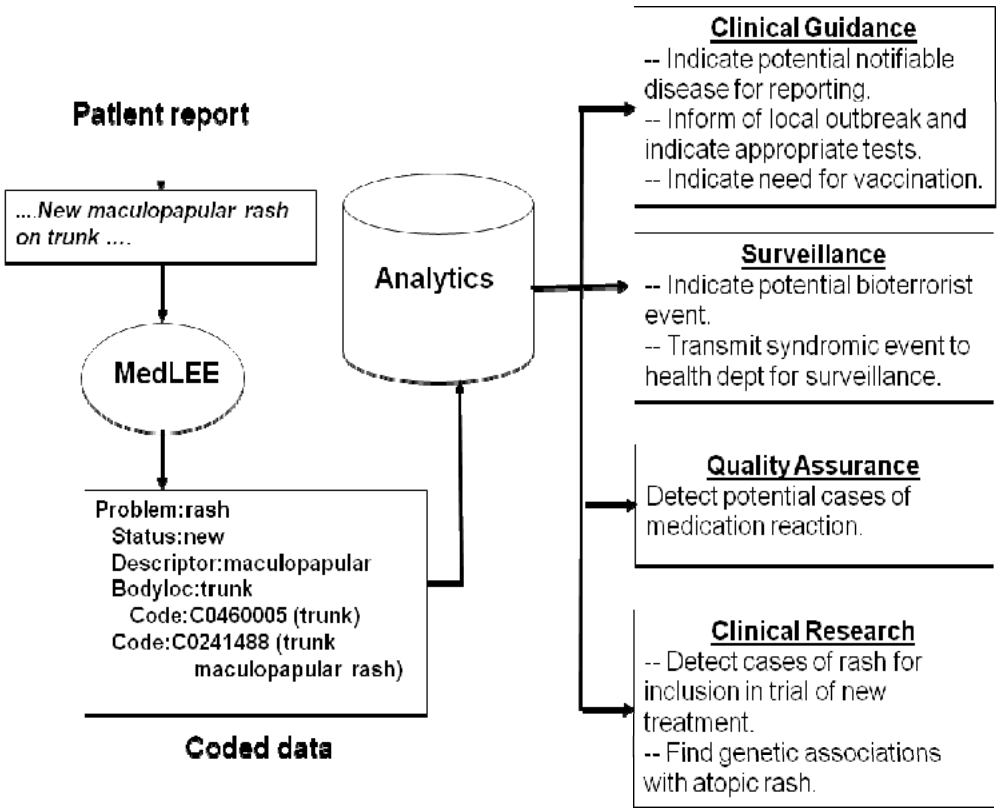


Figure: Overview of MedLEE.

REFERENCES

1. ANATI-OTONG, D., Adverse Drug Reactions Associated with Antipsychotics, Antidepressants and Mood Stabilizers. *Nursing Clinics of North America* (2003) Volume 38, p161-176
2. QORRAJ-BYTYQI, HASIME., HOXHA, REXHEP., KRASNIQI, SHAIP., BAHTIRI, ELTON., and KRANSIQI, VALON., The incidence and clinical relevance of drug interactions in pediatrics. *Journal of Pharmacology and Pharmacotherapeutics* (2012) Volume 3, Issue 4, p304-07
3. GRATTAGLIANO, IGNAZIO., PORTINCASA, PIERO., D'AMBROSIO, GAETANO., PALMIERI, VINCENZO O., and PALASCIANO, GIUSEPPE., Avoiding drug interactions: Here's help. *The Journal of Family Practice* (2010) Volume 59, Number 6: p322-29
4. PIRMOHAMED, M., BRECKENRIDGE, AM., KITTERINGHAM, NR., and PARK, BK, Adverse Drug Reactions, *British Medical Journal,* (1998), Volume 316: p1295-98.
5. PATEL, P., and Zed, PJ,, Drug-related visits to the emergency department: How big is the problem?, *Pharmacotherapy* (2002), volume 22: p915-23.
6. JUNTTI-PATINENL, NEUVONEN PJ., Drug-related deaths in a University central hospital. *European Journal of Pharmacology*.2002, volume 58: p479-82.

7. SITAR, DANIEL S., Aging Issues in Drug Disposition and Efficacy. Proc. West. *Pharmacology Soc* (2007) Volume 50: p16-20.

8. JANCHAWEE, B., OWATRANPORN, T., MAHATTHANATRAKUL, W., and CHONGSUVIVATWONG, V., Clinical drug interactions in outpatients of a university hospital in Thailand. *Journal of Clinical Pharmacy and Therapeutics* (2005) Volume 30, p583-90.

9. BJORKMAN, IK., FASTBOM, J., SCHMIDT, IK., and BERNSTEN, CB., Drg-drug interactions in the elderly, *Annals of Pharmacotherapy* (2002): Volume 36: p1675-81.

10. CRUCIOL-SOUZA, JM., and THOMSON, JC., Prevalence of potential Drug-drug interactions and associated factors in a Brazilian teaching hospital, *Journal of Pharmacology and pharmaceutical science*, (2006), Volume 9; p427-33.

11. JUURLINK, DAVID N., MAMDANI, MUHAMMAD., KOPP, ALEXANDER., LAUPACIS, REAS., and REDELMEIER, DONALD A., Drug-Drug Interactions Among Elderly Patients Hospitalized for Drug Toxicity. *Journal of the American Medical Association* (2003) Volume 289, Issue 13: p1652-58.

12. MERLO, J., LIEDHOLM, H., LINDBLAD, U., BJÖRCK-LINNÉ, A., FÄLT, J., LINDBERG, G., ET AL. Prescriptions with potential drug interactions dispensed at Swedish pharmacies in January 1999: cross sectional study. *BMJ* (2001) Volume 323: p427-28.

13. BECKER, M L., KALLEWAARD, M., CASPERS, P W., VISSER, L E., LEUFKENS, HG., and STRICKER, B H., Hospitalizations and

emergency department visits due to drug-drug interactions: A Literature Review. *Pharmacoepidemiology Drug Safety* (2007) Volume 16: p641-51.

14. PAILLE, F., Over-consumption of Drug by Elderly Patients, *Therapie* (2004), Volume 59; p215-22.
15. MANNHEIMER, B., ULFVARSON, J., EKLOF, S., BERGQVIST, M., ANDERSEN-KARLSSON, E., et al, Drug-related problems and pharmacotherapeutic advisory intervention at a medicine clinic, *European journal of clinical Pharmacology*, (2006), Volume 62: p1075-81.
16. MARCHETTI, S., MAZZANTI, R., BEIJNEN, J H., and SCHELLENS, J H M., Concise Review: Clinical Relevance of drug-drug and herb-drug Interactions mediated by the ABC Transporter ABCB1 (MDR1, P-glycoprotein). *The Oncologist* (2007) Volume 12: p927-41.
17. MOURA, CRISTIANO., PRADO, NÍLIA., and ACURCIO, FRANCISCO., Potential Drug-Drug Interactions Associated with Prolonged Stays in the Intensive Care Unit, A Retrospective Cohort Study. *Clinical Drug Investigation* (2011) Volume 31, Issue 5: p309-16
18. National Centre for Health Statistics (2012*) Health, United States. 2011: With Special Feature on Socioeconomic Status and Health*, National Centre for Health Statistics.
19. JHA, A.K., KUPERMAN, G J., RITTENBERG, E., TEICH, J M., and BATES, D W., Identifying hospital admissions due to adverse

drug events using a computer-based monitor. *Pharmacoepidemiology Drug Safety* (2001) Volume 10: p113-21.

20. SHAD, M U., MARSH, C., and PRESKORN, S H., The economic consequences of a drug-drug interaction. *Journal of Clinical Psychopharmacology* (2001), Volume 21: p119-20.
21. HUANG, S.M., LESCO,L.J. Drug-drug, drug-dietary supplement and drug-citrus fruit and other food interactions: what have we learned?, *Journal of Clinical Pharmacology*, (2004), volume 44, issue 6: p559-69.
22. ZHANG, LEI., ZHANG, YUANCHAO., ZHAO, PING., and HUANG, SHIEW-MEI., Predicting Drug–Drug Interactions: An FDA Perspective. *The AAPS Journal,* (2009), Volume 11, Number 2: p300-6.
23. FDA Adverse event reporting system. http://www.fda.gov/drugs/ guidancecomplianceregulatoryinformation/surveillance/adversedrug effects/default.htm
24. Centers for Disease Control/ National Center for Health Statistics. *National Hospital Ambulatory Medical Care Survey* 2009 (2010).
25. SEGURA-BEDMAR, ISABEL., MARTÍNEZ, PALOMA., and PABLO-SÁNCHEZ, CÉSAR DE., Extracting drug-drug interactions from biomedical texts. *BMC Bioinformatics,* (2010), Volume 11, Supplement 5 , P9: p1-2.
26. MIDDLETON, R., (2006), Adverse drug reactions and toxicology. In Gerbino PP(Ed) *Remington: The Science and Practice of Pharmacy*. Twenty first edition. Lippincot, Williams and Wilkins, Philadelphia PA.

27. PIRMOHAMMED, M., and PARK, BK., Adverse drug reactions: back to the future, *British Journal of Clinical Pharmacology*. (2003), Volume 55, issue 5: p486-92.

28. Yellowcard https://yellowcard.mhra.gov.uk/

29. SNYDER, BEN D., POLASEK, THOMAS M., and DOOGUE, MATTHEW P., Drug interactions: principles and practice. *Australian Prescriber* (2012) Volume 35, Number 3: p85-88.

30. KANTON, W., CANTRELL, C., SOKOL, MC., CHIAO, E., and GDVIN, JM., Impact of antidepressant drug adherence on comorbid medications use and resource utilization. *Archives of Internal Medicine*, (2005), volume 165: p2497-503.

31. KUMAR, A., Past, Present and Future of Pharmacovigilance in India. *Systematic Reviews in Pharmacy* (2011) Volume 2, Issue 1: p55-58.

32. KELTNER, NORMAN L., and MOORE, RANDY L., Psychiatric Drug–Drug Interactions: A Review. *Perspectives in Psychiatric Care* (2010), Volume 46, Number 3: p244-51.

33. FALLOWFIELD, JM., and MARLOW, HF., Propranolol is contraindicated in asthma [letter]. *BMJ* (1996), Volume 313: p1486.

34. WHO website http://www.who.int/en/

35. WHO: The Importance of Pharmacovigilance. The WHO collaborating centre for international drug monitoring (2002) http://www.who-umc.org.

36. AC (Kees) van GROOTHEESt and RACHEL L. RICHESSON. Chapter 19,367-387. Clinical Research Informatics, Health Informatics, DOI 10.1007/978-1-84882-448-5-19. Copyright Springer-Verlag London Limited.

37. WALLER, PC., CAULSON, RA., and WOOD, SM., Regulatory pharmacovigilance in the United Kingdom: current principles and practice, *Pharmacoepidemiology and Drug Safety*, (1996), Volume 5, Issue 6: pp 363–75.

38. MOORE, N., The role of the clinical pharmacologist in the management of adverse drug reactions. *Drug Safety* (2001), Volume 24: p1–7.

39. HAUBEN, M., MADIGAN, D., GERRITS, CM., WALSH, L., VAN, and PUIJENBROEK EP., The role of data mining in pharmacovigilance. *Expert Opinion on Drug Safety* (2005), Volume 4, Issue 5: p929-48.

40. EGBERTS, AC., MEYBOOM, RH., and VAN, PUIJENBROEK EP., Use of measures of disproportionality in pharmacovigilance: three Dutch examples. *Drug Safety,* (2002), Volume 25: p453–58.

41. SHUKA, SS., BINA, GIDWANI., PANDEY, R., RAO, SP., SINGH, V., and AMBER, VYAS., Importance of pharmacovigilance in Indian pharmaceutical industry, *Asian Journal of Research in Pharmaceutical Sciences* (2012), Volume 2, Issue 1: pages 4-8.

42. The Uppsala Monitoring Centre , http://www.who-umc.org/

43. AGARWAL, BHARAT BHUSHAN., and TAYAL, SUMIT PRAKASH., Association Rule Mining in Large Database. *Data Mining and Warehousing*, Laxmi Publications. (2009) Chapter 7.

44. AGRAWAL, R., IMIELINSKI, T., and SWAMI, A., Mining association rules between sets of items in large databases. *Proceedings of the 1993 ACM SIGMOD international conference on Management of data* (1993): p207-16.

45. BROSSETTE, S E., SPRAGUE, AP., HARDIN, JM., WAITES, K B., JONES, W T., and MOSER, S A., Association rules and data mining in hospital infection control and public health surveillance, *Journal of the American Medical Informatics Association* (1998), Volume 5, Number 4: p373-81.

46. MA, L., TSUI, F., HOGAN, W., WAGNER, M., and MA, H., A framework for infection control surveillance using association rules. *AMIA Annual Symposium Proceedings Archive* (2003): p410-14.

47. SURENDIRAN, R., RAJAN, K P., and KUMAR, SATISH M., Study on the Customer Targeting Using Association Rule Mining. *International Journal on Computer Science and Engineering* (2010) Volume 2, Number 7: p2483-85.

48. PRASAD, PRAMOD., and MALIK, LATESH., Using Association Rule Mining for Extracting Product Sales Patterns in Retail Store Transactions. *International Journal on Computer Science and Engineering* (2011) Volume 3, Number 5: p2177-82.

49. VENKATESWARI, S., and SURESH, R.M., Association Rule Mining In Ecommerce: A Survey. *International Journal of Engineering Science and Technology* (2011) Volume 3, Number 4: p3086-89.

50. BERNARD, KAMSU-FOGUEM., FABIEN, RIGAL., and FELIX, MAUGET., Mining association Rules for Quality Improvement of the production Process, *Expert Systems with Applications*, (2013), Volume 40: p1034-45.

51. KIM, WANKYUNG., and SOH, WOOYOUNG., Applying Association Rule of the Data Mining Method for the Network Event Analysis. Computation in Modern Science and Engineering, *Proceedings on the International Conference on Computational Methods in Science and Engineering* (2007) CP 963, Volume 2, Part A: p567-70.

52. DING, WEI., EICK, CHRISTOPH F., YUAN, XIAOJING., WANG, JING., and NICOT, JEAN-PHILIPPE., A framework for regional association rule mining and scoping in spatial datasets. *Geoinformatica* (2011) Volume 15: p1–28

53. SUMALATHA, R., and RAMASUBBAREDDY, B., Mining Positive and Negative Association Rules. *International Journal on Computer Science and Engineering* (2010), Volume 2, Number 9: p2916-20.

54. ALAM, NIYAZ., KUMAR , RAJEEV., and BHARDWAJ, ALOK., Pharmacovigilance for adverse drug reactions in patterns on Non-Steroidal Anti-Inflammatory Drugs and Hepatic Dysfunction in Aceclofenac Therapy in South Delhi Hospital. *International Journal*

of Pharmacy & Pharmaceutical Sciences (2012), Supplement 5, Volume 4: p128-31.

55. PALLELA, F.J., DELANEY, K.M., MOORAN, A.C., LOVELESS, M.O., FUHRER, J., SATTEN, G.A., ASCHMAN, D.J., and HOLMBERG, S.D., Declining Morbidity and Mortality among Patients with Advanced Human Immunodeficiency Virus Infection, *The New England Journal of Medicine*, (1998), Volume 338: p853-60.

56. SHAH, SANJIV., McGOWN, J., OPULSKY, B., LIEBLEIN, A., and SAPERSTEIN, A., Identification of Drug Interactions Involving ART in New York City HIV Speciality Clinics, In: Abstracts of the Fourteenth Conference on Retroviruses and Opportunistic Infections, Los Angeles, CA, (2007), Abstract 573: p277.

57. MARZOLINI, CATIA., ELZI, LUIGIA., GIBBONS, SARA., WEBER, RAINER., FUX, CHRISTOPH., FURRER, HANSJAKOB., CHAVE, JEAN-PHILIPPE., CAVASSINI, MATTHIAS., BERNASCONI, ENOS., CALMY, ALEXANDRA., VERNAZZA, PIETRO., KHOO, SAYE., LEDERGERBER, BRUNO., BACK DAVID., and BATTEGAY, MANUEL., Prevalence of comedications and effect of Potential Drug Drug Interactions in the SWISS HIV Cohort Study. *Antiviral Therapy* (2010) Volume 15: pp413-23.

58. MILLER, C D., EL-KHOLI, R., FARAGON, J J., and LODISE, T P., Prevalence and risk factors for clinically significant drug interactions with antiretroviral therapy. *Pharmacotherapy* (2007), Volume 27: p1379–86.

59. KIGEN, GABRIEL., KIMAIYO, SYLVESTER., NYANDIKO, WINSTONE., FARAGHER, BRIAN., SANG, EDWIN., JAKAIT, BEATRICE., OWEN, REW., BACK, DAVID., GIBBONS, SARA., SEDEN, KAY., and KHO, SAYEAND H., Prevalence of Potential Drug-Drug Interactions Involving Antiretroviral Drugs in a Large Kenyan Cohort. *PLoS Computational Biology* (2011) Volume 6, Issue 2: p1-7.

60. PATEL, N., ABDELSAYED, S., VEVE, M., and MILLER, C.D. Predictors of clinically significant drug-drug interactions among patients treated with nonnucleoside reverse transcriptase inhibitor-, protease inhibitor-, and raltegravir-based antiretroviral regimens. *The Annals of Pharmacotherapy* (2011) Volume 45, Issue 3: p317–24.

61. RATHBUN, CHRIS R., and LIEDTKE, MICHELLE D., Antiretroviral Drug Interactions: Overview of Interactions Involving New and Investigational Agents and the Role of Therapeutic Drug Monitoring for Management. *Pharmaceutics* (2011) Volume 3: p745-81.

62. MARZOLINI, CATIA., BACK, DAVID., WEBER, RAINER., FURRER, HANSJAKOB., CAVASSINI, MATTHIAS., CALMY, ALEXANDRA., VERNAZZA, PIETRO., BERNASCONI, ENOS., KHOO, SAYE., BATTEGAY, MANUEL., and ELZI, LUIGIA., Ageing with HIV: medication use and risk for potential drug–drug interactions. *Journal of Antimicrobial Chemotherapy,* (2011), Volume 66 Issue 9: p2107-11.

63. University of Liverpool website http://www.liv.ac.uk/.

64. YAP, K.Y.L., TAY, W.L., CHUI, W.K., and CHAN, A., Clinically relevant drug interactions between anticancer drugs and psychotropic agents. *European Journal of Cancer Care* (2011), Volume 20: p6–32.

65. BLOWER, PETER., WIT, RONALD DE., GOODIN, SUSAN., and AAPRO, MATTI., Drug–drug interactions in oncology: Why are they important and can they be minimized? *Critical Reviews in Oncology/ Hematology* (2005), Volume 55, Issue 2: p117-42.

66. SAYLOR, MATTHEW S., and SMETANA, RONALD F., Potential for drug–drug interactions in treating cancer-related nausea and distress, *Journal of Oncology Pharmacy Practice* (2010) Volume 17, Issue 4: p403-8.

67. RIECHELMANN, RACHEL P., ZIMMERMANN, CAMILLA., CHIN, SHERAY N., WANG, LISA., O'CARROLL, AOIFE., ZARINEHBAF, SANAZ., and KRZYZANOWSKA, MONIKA K., Potential Drug Interactions in Cancer Patients Receiving Supportive Care Exclusively. *Journal of Pain and Symptom Management* (2008), Volume 35, Number 5: p535-43.

68. KANNAN, G., ANITHA., VANITHA, RANI N., THENNARASU, P., ALOSH, J., VASANTHA, J., MARTIN, J., and UMA, MRC., A study of drug–drug interactions in cancer patients of a south Indian tertiary care teaching hospital. *Journal of Postgraduate Medicine* (2011) Volume 57, Issue 3: p206-10.

69. TYLER, TIMOTHY., Drug interactions in metastatic breast cancer. *Journal of oncology pharmacy practice* (2011) Volume 17, Number 3: p236-245.

70. MIRANDA, VANESSA., FEDE, ANGELO., NOBUO, MELISSA., AYRES, VERONICA., GIGLIO, AURO., MIRANDA, MICHELE., and RIECHELMANN, RACHEL P., Adverse Drug Reactions and Drug Interactions as Causes of Hospital Admission in Oncology. *Journal of Pain and Symptom Management* (2011) Volume 42, Number 3: p342-53.

71. TEIXEIRA, JORGE JUAREZ VIEIRA., CROZATTI, MARCIA TEREZINHA LONARDONI., SANTOS, CARLOS APARECIDO DOS., and ROMANO-LIEBER, NICOLINA SILVANA., Potential Drug-Drug Interactions in Prescriptions to Patients over 45 Years of Age in Primary Care, Southern Brazil. *PLoS Computational Biology* (2012), Volume 7, Issue 10: p1-6.

72. OBRELI NETO, PR., NOBILI, A., DE LYRA, DP., PILGER, D., GUIDONI, CM., DE OLIVEIRA, BALDONI A., CRUCIOL-SOUZA, JM., DE CARVALHO, FREITAS AL., TETTAMANTI, M., GAETI, WP., and NAKAMURA, CUMAN RK., Incidence and Predictors of Adverse Drug Reactions caused by Drug-Drug Interactions in Elderly Outpatients: A Prospective Cohort Study. *Journal of Pharmacy & Pharmaceutical Sciences* (2012), Volume 15, Issue 2: p332-43.

73. SECOLI, SILVIA-REGINA., FIGUERAS, ALBERT., LIMA, MARIA LUCIA LEBRÄO FERNÂO DIAS DE., and SANTOS, J

LICIO FERREIRA., Risk of Potential Drug-Drug Interactions among Brazilian Elderly A Population-Based, Cross-Sectional Study. *Drugs & Aging* (2010) Volume 27, Issue 9: p759-70.

74. DHABALI, A.A.H., AWANG, R., and ZYOUD, S.H., Clinically important drug–drug interactions in primary care. *Journal of Clinical Pharmacy and Therapeutics* (2012), Volume 37: p426–30.

75. ISMAIL, MOHAMMAD., IQBAL, ZAFAR., KHATTAK, MUHAMMAD BHAL., KHAN, MUHAMMAD IMRAN., JAVAID, ARSHAD., and KHAN, TAHIR MEHMOOD., Potential drug-drug interactions in cardiology ward of a teaching hospital. *HealthMed* (2012) Volume 6, Issue 5: p1618-24.

76. APARASU, RAJENDE R., BAER, REBECCA., and APARASU, ANURADHA., Clinically important potential drug-drug interactions in outpatient settings. *Research in Social and Administrative Pharmacy* (2007) Volume 3: p426-37.

77. SEPEHRI, G., KHAZAELLI, P., DAHOOIE, F ARABPOUR., SEPEHRI, E., and DEHGHANI, M R., Prevalence of Potential Drug Interactions in an Iranian General Hospital. *Indian Journal of Pharmaceutical Sciences.* (2012), Volume 74, Issue 1: p72-75.

78. SPINA, EDOARDO., and SCORDO, MARIA GABRIELLA., Clinically Significant Drug Interactions with Antidepressants in the Elderly. Drugs and Aging (2002), Volume 19, Issue 4: p299-320.

79. JOHNELL, KRISTINA., and KLARIN, INGA., The Relationship between Number of Drugs and Potential Drug-Drug Interactions in

the Elderly: A Study of Over 600 000 Elderly Patients from the Swedish Prescribed Drug Register. *Drug Safety* (2007), Volume 30, Issue 10: p911-18.

80. KUMAR, B MAHENDRA., KUMARASWAMY, M., and MAHADEVAMMA, L., Incidence And Pattern Of Potential Drug Interactions Of Antimicrobial Agents In The Department Of Medicine In A Tertiary Care Teaching Hospital: A Prospective Study. *Asian Journal of Pharmaceutical and Clinical Research* (2011), Volume 4, Suppl 2: p31-36.
81. SMITHHURGER, PAMELA L., KANE-GILL, SANDRA L., and SEYHERT, AMY L., Drug-Drug Interactions in Cardiac and Cardiothoracic Intensive Care Units An Analysis of Patients in an Academic Medical Centre in the US. *Drug Safety* (2010), Volume 33, Issue 10: p879-88.
82. CHAN, AGNES L.F., LEE, HAW YU., H,O CHI-HOU., CHAIN, THAU-MING., and LIN, SHUN JIN., Cost Evaluation of Adverse Drug Reactions in Hospitalized Patients in Taiwan: A Prospective, Descriptive, Observational Study. *Current Therapeutic Research* (2008) Volume 69, Issue 2: p118-29.
83. PIRMOHAMED, MUNIR., JAMES, SALLY., MEAKIN, SHAUN., GREEN, CHRIS., SCOTT, ANDREW K., WALLEY, THOMAS J., FARRAR, KEITH., PARK, B KEVIN., and BRECKENRIDGE, ALASDAIR M., Adverse drug reactions as cause of admission to hospital: prospective analysis of 18 820 patients. *BMJ* (2004) Volume 329: p15-19

84. CROFT, MARIE., BRENDAN, KEELY., MORRIS, IAN., TANN, LAN., and LAPP, GRAHAM., Predicting Drug Candidate Victims of Drug-Drug Interactions, using Microdosing. *Clinical pharmacokinetics* (2012) Volume 51, Issue 4: p237-46.

85. NAGILLA, R., and WARD, KW., A comprehensive analysis of the role of correction factors in the allometricpredictivity of clearance from rat, dog, and monkey to humans. *Journal of Pharmaceutical Sciences* (2004) Volume 93, Issue 10: p2522-34.

86. INGS, R M., Microdosing: A valuable tool for accelerating drug development and the role of bioanalytical methods in meeting the challenge. *Bioanalysis* (2009) Volume 1, Issue 17: p1293-305.

87. ICH Topic M3 note for guidance on non-clinical safety pharmacology studies for human pharmaceutical CPMP. *International Conference on Harmonisation*. (2009). http://www.emea.europa.eu/docs/en_GB/ documentJibrary/Scientinc_guideline/2009/09/WC500002720.pdf

88. HAUBEN, MANFRED., PATADIA, VAISHALI., GERRITS, CHARLES., WALSH, LOUISA., and REICH, LESTER., Data Mining in Pharmacovigilance, The Need for a Balanced Perspective. *Drug Safety* (2005) Volume 28, Issue 10: p835-42.

89. WILSON, AM., THABANE, L., and HOLBROOK, A., Application of data mining techniques in pharmacovigilance. *British Journal of Clinical Pharmacology* (2004) Volume 57, Issue 2: p127-34.

90. ALMENOFF, J., TONNING, JM., GOULD, AL., SZARFMAN, A., HAUBEN, M., OUELLET- HELLSTROM, R., BALL, R., HORNBUCKLE, K., WALSH, L., and YEE, C., Perspectives on the use of data mining in pharmacovigilance. *Drug Safety* (2005) Volume 28, Issue 11: p981-1007.

91. LIU, MEI., MATHENY, MICHAEL E., HU, YONG., and XU, HUA., Data Mining Methodologies for Pharmacovigilance. *ACM SIGKDD Explorations Newsletter* (2012) Volume 14, Issue 1: p35-42.

92. BISGIN, HALIL., LIU, ZHICHAO., FANG, HONG., XU, XIAOWEI., and TONG, WEIDA., Mining FDA drug labels using an unsupervised learning technique - topic modeling. *BMC Bioinformatics* (2011) Volume 12, Suppl 10:s 11: p1-8.

93. SIMBARASHE, PETER ZVADA., TANGAI, EDWIN CHAGWEDERA., ROSEMARY, CHIGWANDA., and COLLEN, MUTOWEMBWA MASIMIREMBWA, An Experimental Pharmacokinetic Computer Program to Predict Potential Drug-drug Interactions. *The open Drug Metabolism Journal* (2009), Volume 3: p8-16.

94. DUKE, JON D., HAN, XU., WANG, ZHIPING., SUBHADARSHINI, ABHINITA., KARNIK, SHREYAS D., LI, XIAOCHUN., HALL, STEPHEN D., JIN, YAN., CALLAGHAN, J THOMAS., OVERHAGE, MARCUS J., FLOCKHART, DAVID A., STROTHER, R MATTHEW., QUINNEY, SARA K., and LI, LANG., Literature Based Drug Interaction Prediction with Clinical Assessment Using Electronic Medical Records: Novel Myopathy

Associated Drug Interactions. *PLoS Computational Biology* (2012) Volume 8, Issue 8 :p1-13.

95. TARI, L., ANWAR, S., LIANG, S., CAI, J., and BARAL, C., Discovering drug-drug interactions: a text-mining and reasoning approach based on properties of drug metabolism. *Bioinformatics* (2010), Volume 26, Issue 18: p547-53.

96. ZHANG, Y., LIN, H., YANG, Z., WANG, J. and LI, Y., A single kernel-based approach to extract drug-drug interactions from biomedical literature. *PLoS Computational Biology* (2012) Volume 7, Issue 11: p1-11.

97. SEGURA-BEDMAR, ISABEL., MARTÍNEZ, PALOMA., and PABLO-SÁNCHEZ, CÉSAR DE., A linguistic rule-based approach to extract drug-drug interactions from pharmacological documents. *BMC Bioinformatics* (2011), Volume 12, Suppl 2:S1: p1-11.

98. AGRAWAL, RAKESH., and SRIKANT, RAMAKRISHNAN., Fast Algorithms for Mining Association Rules. *Proceedings of the 20th VLDB Conference, Santiago, Chile* (1994).

99. MUKHOPADHYAY, ANIRBAN., MAULIK, UJJWAL., and BANDYOPADHYAY, SANGHAMITRA., A Novel Biclustering Approach to Association Rule Mining for Predicting HIV-1–Human Protein Interactions. *PLoS Computational Biology* (2012), Volume 7, Issue 4: p1-13.

100. MANDA, PRASHANTI., OZKAN, SEVAL., WANG, HUI., MCCARTHY, FIONA., and BRIDGES, SUSAN M., Cross-Ontology

Multi-level Association Rule Mining in the Gene Ontology. *PLoS Computational Biology* (2012), Volume 7, Issue 10: p1-9.

101. CREIGHTON, CHAD., and HANASH, SAMIR., Mining gene expression databases for association rules. *Bioinformatics* (2003) Volume 19, Number 1: p79–86.
102. MARTINEZ, RICARDO., PASQUIER, NICOLAS., and PASQUIER, CLAUDE., GenMiner: mining non-redundant association rules from integrated gene expression data and annotations. *Bioinformatics* (2008), Volume 24, Number 22: p2643–44.
103. ROUANE, H., TOUSSAINT, Y., and VALTCHEV, P. Mining signals in spontaneous reports database using concept analysis. *Artificial Intelligence in Medicine Lecture Notes in Computer Science* (2009), LNAI 5651: p285-94.
104. CHEN, J., HE, H., and LI, J., Representing association classification rules mined from health data. *Knowledge-Based Intelligent Information and Engineering Systems Lecture Notes in Computer Science* (2005), Volume 3683: p1225-31.
105. HE, YANHUA., ZHENG, XIAO., SIT, CINDY., LOO, WINGS TY., WANG, ZHIYU., XIE, TING., JIA, BO., YE, QIAOBO., TSUI, KAMCHUEN., CHOW, LOUIS WC., and CHEN, JIANPING., Using association rules mining to explore pattern of Chinese medicinal formulae (prescription) in treating and preventing breast cancer recurrence and metastasis. *Journal of Translational Medicine* (2012), Volume 10, Suppl 1:S 12: p1-8.

106. STANISIC, PREDRAG., and TOMOVIC, SAVO., An Efficient Procedure for Mining Statistically Significant Frequent Itemset. *Publications de L'INSTITUTE MATHEMATIQUE* (2010) Volume 87, Issue 101: p109-19.

107. ZHAO, QIANKUN., and BHOWMICK, SOURAV S., Association Rule Mining: A Survey. *Technical Report, CAIS, Nanyang Technological University, Singapore* (2003) Number 2003116: p1-20.

108. GAUTAM, PRATIMA., and PARDASANI, K. R., Algorithm for Efficient Multilevel Association Rule Mining. *International Journal on Computer Science and Engineering* (2010), Volume 2, Number 5: p1700-4.

109. UMARANI, V., and PUNITHAVALLI, M., An Empirical Analysis over the Four Different Methods of Progressive Sampling-Based Association Rule Mining. *European Journal of Scientific Research* (2011) Volume 66, Number 4: p620-30.

110. AL-ZAWAIDAH, FARAH HANNA., JBARA, YOSEF HASAN., and ABU-ZANONA, MARWAN AL-ABED., An Improved Algorithm for Mining Association Rules in Large Databases. *World of Computer Science and Information Technology Journal* (2011) Volume 1, Number 7: p31116.

111. BATES, DW., SPELL, N., CULLEN, DJ., BURDICK, E., LAIRD, N., PETERSEN, LA., SMALL, SD., SWEITZER, BJ., and LEAPE, LL., The costs of adverse drug events in hospitalized patients.

Journal of the American Medical Association (1997), Volume 277: p307-311.

112. CLASSEN, DC., PESTOTNIK, SL., EVANS, RS., LLOYD, JF., and BURKE, JP., Adverse drug events in hospitalized patients. Excess length of stay, extra costs, and attributable mortality. *Journal of the American Medical Association* (1997) Volume 277, Issue 4: p301-6.

113. JEETU, G., and ANUSHA., Pharmacovigilance a worldwide master key for drug safety monitoring. *Journal of Young Pharmacists* (2010) Volume 2, Issue 3: p315–20.

114. FLOOR-SCHREUDERING, ANNEMIEKE., SMET, PETER A.G.M. DE., AMINI, HENK BUURMA SONIA., and MARCEL, L., Clarity and Applicability of Drug-Drug Interaction Management Guidelines; A Systematic Appraisal by General Practitioners and Community Pharmacists in the Netherlands. *Drug Safety* (2011), Volume 34, Issue 8: p683-90.

115. MAANEN, A CLARA DRENTH-VAN., MARUM, ROB J VAN., KNOL, WILMA., LINDEN, CAROLIEN M J VAN DER., and JANSEN, PAUL A F., Prescribing Optimization Methods for Improving Prescribing in Elderly Patients Receiving Polypharmacy. *Drugs & Aging* (2009), Volume 26, Issue 8: p687-701.

116. MALONE, DANIEL C., ABARCA, JACOB., HANSTEN, PHILIP D., GRIZZLE, AMY J., ARMSTRONG, EDWARD P., BERGEN, ROBIN C VAN., DUNCAN-EDGAR, BABETTE S., SOLOMON, STEVEN L., and LIPTON, RICHARD B., Identification of Serious

Drug–Drug Interactions: Results of the Partnership to Prevent Drug–Drug Interactions. *Journal of the American Pharmacists Association* (2004), Volume 44, Number 2: p142-51.

117. TAKADA MASAHIRO, INADA HIROSHI,NAKAZAWA KAZUO, TANI SHOKO,IWATA MICHIAKI,SUGIMOTO YOSHIHISA,NAGATA SATORU.,A High-Speed Drug Interaction Search System for Ease of Use in the Clinical Environment. *Journal of Medical Systems* (2012), Volume 36: p3533-41.

118. PERKINS, NATALIE A., MURPHY, JOHN E., MALONE, DANIEL C.,ARMSTRONG, EDWARD P., Performance of Drug-Drug Interaction Software for Personal Assistants., *The Annals of Pharmacotherapy*. (2006), Volume 40, Issue 5: p850-5.

119. MANNHEIMER, BUSTER., WETTERMARK, BJÖRN., LUNDBERG, MICHAEL., PETTERSSON, HANS., BAHR, CHRISTER VON., and ELIASSON, ERIK., Nationwide drug-dispensing data reveal important differences in adherence to drug label recommendations on CYP2D6- dependent drug interactions. *British Journal of Clinical Pharmacology* (2010), Volume 69, Issue 4: p411-17.

120. SAVAGE, NEIL., Digging for drug facts. *Communications of the ACM* (2012) Volume 55, Number 10, p11-13.

121. USAMA, FAYYAD.,GREGORY, PIATETSY SHAPIRO., and PADHRAIC,SMYTH., From Data Mining to Knowledge Discovery in Databases, *American Association for Artificial Intelligence*, (1996): p 37-54.

122. HAUBEN, M., and ZHOU, X., Quantitative methods in pharmacovigilance: focus on signal detection. *Drug Safety* (2003), Volume 26, Issue 3: p159-86.

123. STEPHENSON, W., and HAUBEN, M., Data mining for signals in spontaneous reporting databases: proceed with caution. *Pharmacoepidemiology & Drug Safety* (2007) Volume 16, Issue 4: p359-65.

124. HAUBEN, M., REICH, L., DEMICCO, J., and KIM, K., Extreme Duplication' in the US FDA Adverse Events Reporting System Database. *Drug Safety* (2007) Volume 30, Issue 6: p551-54.

125. FRAWLEY, PIAATETSKY., Shaping Knowledge Discovery in Databases: An Overview, *The AAAI/ MIT press, Menlo Park,* (1996).

126. KARGUPTA, H., JOSHI, A., SIVAKUMAR, K., and YESHA, Y., Data Mining Next Generation Challenges and Future Directions, (2011). AAAI Press/ The MIT Press.

127. HAN, JIAWEI., KAMBER, MICHELINE., and PEI, JIAN., Data Mining Concepts and Techniques, Third Edition, Morgan Kauffman-Elsevir Publishers, ISBN :978-93-80931-91-3, (2011)

128. HU, T.M., and HAYTON, W.L., Architecture of the drug–drug interaction network. *Journal of Clinical Pharmacy and Therapeutics* (2011) Volume 36: p135–43.

129. 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 Safety* (2002) Volume 25, Issue 6: p381-92.

130. 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. *European journal of clinical pharmacology* (1998) Volume 54: p315-21.
131. BATE, A., and EVANS, SJ., Quantitative signal detection using spontaneous ADR reporting. *Pharmacoepidemiology& Drug Safety* (2009) Volume 18, Issue 6: p427-36.
132. EVANS, SJ., WALLER, PC., and DAVIS, S., Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiology& Drug Safety* (2001) Volume 10, Issue 6: p483-86.
133. MEYBOOM, RHB., EGBERTS, ACG., EDWARDS, IR., ET AL. Principles of signal detection in pharmacovigilance. *Drug Safety* (1997) Volume 16, Issue 6: p355-65.
134. FRIEDMAN, C., SHAGIN, AL., LUSSIER, Y., and HRIPCSAK, G., Automated encoding of clinical documents based on natural language processing, *Journal of the American Medical Informatics Association* (2004), Volume 11, Number.5: p392-402.
135. Micromedex: website http://www.micromedex.com/.