Mining Association Rules in Clinical Trial Safety Data

Li Zhu[1], Amy Xia[1], Qi Jiang[1], William Y. Go[2], Liron Walsh[1]

[1]Amgen, Inc., Thousand Oaks, CA
[2]Kite Pharma, Inc., Santa Monica, CA

Abstract

Clinical trial safety data are routinely analysed to help the drug developer gain knowledge on the safety profile of an investigational drug, and detect potential safety signals in the pre-marketing stage. Typical safety endpoints such as adverse events (AE) are usually summarized in incidence tables, tested on two-by-two contingency tables, or analysed through drug-AE model that are based on aggregated data. However, although regularly collected in clinical trial data, the information about within-subject correlations is often ignored when AEs are analysed in such aggregated fashions. Association rule is an important statistical learning technique that has been commonly used for mining commercial databases. We propose to extend the application of association rule mining to analyse clinical trial safety data. Our result shows association rule mining could provide interesting findings on associations among safety endpoints as well as their association with investigational drug or other patient characteristics in both one-to-one and many-to-one mappings. It also offers a new way of grouping adverse events that is alternative to traditional approaches such as MedDRA (medical dictionary for regulatory activities) hierarchical coding or SMQ (standard MedDRA queries) search strategies.

Keywords: Adverse event, association rule, clinical trial safety data, safety signal detection

1. Introduction

Safety data is typically collected in a standard way from all clinical trials to aid the drug developer establish a robust safety profile of each investigational drug, detect potential safety signals early on, and devise corresponding plans for risk mitigation and risk management. A comprehensive assessment of safety data is critical for protecting patient's safety as well as supporting a benefit:risk assessment in regulatory approvals. Adverse event (AE), as the most typical safety endpoint, is usually summarized in incidence tables, tested on two-by-two contingency tables, or analysed through drug-AE model that are based on aggregated data. Because the classical hypothesis testing approach applied to efficacy endpoints is problematic when applied to safety data analysis due to endpoint multiplicity and data sparseness, many advanced statistical methods were developed in the last two decades. The mainstream method for clinical trial AE signal detection has been to focus on multiplicity adjustment, where each AE is tested separately, and the resulting raw p-value for every AE are put together to go through certain adjustment to control the overall error rate. Benjamini and Hochberg (1995) [1] proposed a multiplicity adjustment approach that controls false discovery rate (FDR), which is more suitable for AE signal detection as the traditional family-wise error rate controlling methods had lower power in general to flag potential signals. Several variations of the FDR controlling multiplicity adjustment methods were developed based on Benjamini and Hochberg's work, including some multiplicity adjustment methods [2, 3] that designed to suit the hierarchical AE coding structure of MedDRA (medical dictionary for regulatory activities) [4]. In addition to multiplicity adjustment approaches, there are a few other approaches proposed for analysing AE data, such as methods based on Bayesian hierarchical mixture model [5, 6]. However, all these statistical methods are based on aggregated data. Within-subject correlation (i.e., associations among the safety endpoints at the subject-level) naturally resides in the safety data that has been routinely collected

from clinical trials. This information, however, gets ignored when data is further summarized into aggregated fashion, and therefore, it has never been utilized in common analytical methods that are developed for aggregate data.

While extensive statistical methods have been developed in the framework of traditional hypothesis testing and parametric modelling, data mining, as an alternative approach of great potential for safety analyses, has only recently been explored by researchers. Southworth and O'Connell [7] pointed out that the evaluation of drug safety data from clinical trials should better be considered as an exercise in data mining rather than hypothesis testing. They proposed classification tree ensembles based methods to identify important AEs that are potentially caused by the investigational drug. However, tree based methods only focus on associations between drug and each AE individually. Associations among AEs, clustered AEs as associated with specific drug or patient characteristics have not been studied.

Association rule (AR) mining is one of the most important statistical leaning techniques, which was first introduced by Agrawal *et al.* in early 1990s [8, 9], and has been studied extensively in the data mining literature. It was initially used for Market Basket Analysis, where interest lies in finding the associations among items purchased by grocery store customers as recorded in transaction databases. As it became a popular and well researched method for discovering relations between variables in large databases, AR mining has also been applied to other areas such as telecommunication networks, web usage mining, intrusion detection, and inventory control. We propose to apply AR mining in clinical trial safety data to detect meaningful patterns and relationships among AEs, AEs associated with certain treatment groups, or certain patient subpopulations. The AR mining works on the subject-level data and best utilizes the within-subject correlation, and it also synthesize information across all the safety endpoints (such as all the AEs been analysed) instead of treating pairwise relations between each endpoint with drug as independent to all other endpoints. Therefore, it could provide unique insight into interesting patterns in clinical trial safety data.

2. Association Rule Mining for Clinical Trial Safety Data

The basic concepts of association rule mining were established originally on transaction data [8]. In a given database, denote $I = \{i_1, i_2, ..., i_m\}$ as a set of m distinct attributes, or *items*. A set of items is also termed as an *itemset*. A *transaction* T is an itemset such that $T \subseteq I$. An *association rule*, or a *rule*, is an implication in the form of $A \longrightarrow B$, where $A \subset I$, $B \subset I$ are two itemsets, and $A \cap B = \emptyset$. A and B are referred as *antecedent* and *consequent*, respectively, or the *rule body* and *rule head*. A transaction T *contains* the set of items A if $A \subset T$. When ARs are mined in a transaction database, two basic measures are essential: support and confidence. The *support* measures the "prevalence" of $A \cup B$:

$$\text{Support}(A \longrightarrow B) = \frac{\#\{\, T\colon (A \cup B) \subset T \,\}}{N},$$

where the numerator is the number of transactions that contain both itemsets A and B, and the denominator, N, is the total number of transactions in the database. Notation $\#\{\cdot\}$ represent "the number of". The *confidence* measures the "predictability" of the rule:

$$\text{Confidence}(A \longrightarrow B) = \frac{\text{Support}(A \longrightarrow B)}{\text{Support}(A)} = \frac{\#\{\, T\colon (A \cup B) \subset T \,\}}{\#\{\, T\colon A \subset T \,\}}.$$

Considering all possible combinations of itemsets, theoretically there are going to be almost infinitely many rules even in a database of small to moderate size. For example in the case of m =20, namely, there are only 20 distinct rules considered in the database and any transaction can only include a subset from

these 20 items in I, the possible rules in total will be well above 3 billion rules. Therefore, we usually set the acceptable thresholds on support and confidence, so that we can eliminate numerous unimportant rules and only identify the important ones. To achieve more focused statistical learning, the common practice is to only mine rules with single-item rule head, in which case we can leverage the knowledge about the attribute(s)/item(s) in rule body A to identify the corresponding associated attribute/item in the rule head B. Many fast algorithms, such as Apriori [9], Eclat [10], and FP-growth [11], have been developed to mine the rules of interest. These fast algorithms nowadays have been implemented in many statistical software packages.

In addition to support and confidence, several other measures for an AR have been proposed, among which the *lift* is most commonly considered [12, 13]. The lift of a rule is defined as

$$\text{Lift}(A \longrightarrow B) = \frac{\#\{\, T: (A \cup B) \subset T \,\} \times N}{\#\{\, T: A \subset T \,\}\, \#\{\, T: B \subset T \,\}}.$$

It can be viewed as the ratio of the confidence and *expected confidence* of the rule, with the latter is defined as follows, and is the expected value that confidence would take if A and B were in fact independent.

$$\text{Expected Confidence}(A \longrightarrow B) = \frac{\#\{\, T: A \subset T \,\}\, \#\{\, T: B \subset T \,\}}{\#\{\, T: A \subset T \,\}} = \#\{\, T: B \subset T \,\}.$$

The lift measures the extent to which A and B are not independent, and is often served as a reference to interpret the importance of the rule. It is hard to define a minimum acceptable lift value, but rules with lift value below 1 are deemed not important. For rule body A and rule head B, generally in the case that one is not always concurred with the other, lift gives extra weight to rules if either A or B has considerably low incidence.

The idea of transaction data can be generalized as follows. In the rule $A \longrightarrow B$, the itemsets A and B can be both regarded as certain events, and the rule describes how the event B is associated with the event A, i.e., how the occurrence of event A could imply the occurrence of event B. In the setting of clinical trial safety data, the rule $A \longrightarrow B$ can be mined using different aspect of the data to answer many important questions including, but not limited to, the followings.

1) "Does the drug cause certain AE or cluster of AEs?" For example, after mining AR in patients who received either investigational drug Drug X or the control drug, there is a resulting rule {cardiac arrhythmia}→{Drug X}, cardiac arrhythmia is then identified as a potential safety signal for Drug X. To explore this type of question, we can mine the ARs in the entire study population by treating the drug group for each subject as the rule head B and the AEs experienced by the subject as the rule body A. The identified ARs will summarize which AE(s) tend to occur in each drug group, where the interest often lies on the AE(s) that tend to occur in the patients exposed to the drug under investigation or drug of interest. This is particularly useful for the purpose of safety signal detection, where prior knowledge on whether the drug causes the AE(s) is not available.

2) "Does the event of interest tend to co-occur with any other AEs in patients treated with the investigational drug?" For example, suppose ocular toxicity is a known adverse reaction for Drug X, we could run AR mining in the Drug X group and control group separately, and investigate the clustering behavior of ocular toxicity with other AEs. If {skin infection, rash, eczema}→{ocular toxicity} is a resulting rule from AR mining in the Drug X group but there is no rule indication any AE associated with ocular toxicity in control group (or there is no case of ocular toxicity in

control at all), this could provide data supporting that ocular toxicity, skin infection, rash, and eczema, as a series of AEs, are likely to occur with exposure to Drug X. To explore this type of questions, we can mine the ARs in each subset of patients as divided by the drug group, and treat the adverse event of special interest as the rule head *B* and any subset of all other reported AEs as the rule body *A*. Using AR mining in this way could help with clustering of AEs and provide a new way to group the AEs as a series of potential risks for an investigational drug. Notice in the example here, ocular toxicity belongs to the SOC (system organ class) of eye disorders according to the MedDRA hierarchical coding, while skin infection, rash, and eczema all belong to another SOC – skin and subcutaneous tissue disorders. Therefore, this application of AR mining provide a new way of grouping AEs that is different from traditional approaches such as gouping AEs in preferred terms according to the system organ class per MedDRA hierarchical coding or SMQ (standardised MedDRA queries) [14] search strategies.

3) "Are there any risk factor for certain event(s) that we were not aware of?" Once certain pattern is identified by AR mining in approach 1) or 2) as described above, clinical knowledge should first be applied to interpret the result. If a resulting AR cannot be explained by clinical and biological rationale (e.g., the identified AEs that are associated with the drug, or the identified AEs that are associated with an event of special interest in certain population), we may need to investigate other aspects of data to get a comprehensive profile, such as examining data from medical or surgical history, concomitant medication, genetic characteristic, etc. of the subset of patients who contributed to the resulting rule for some unexplainable pattern by direct clinical interpretation. If some underlying similarity that is unique in the patient subset can be found, then such similarity could constitute a composite risk factor. The knowledge of this risk factor could help identify subjects at increased risk and may allow further selection for personalized treatment.

4) "Are there any demographic characteristics that are associated with certain AE reporting pattern?" For example, in a global oncology phase III clinical trial, unlike in all the other participating countries, in the county South Africa, drug of G-CSF (Granulocyte-colony stimulating factor) is not commonly given to chemotherapy-treated patients unless the patient really developed some noteworthy neutropenia. Then an AR {serious neutropenia}→{South Africa} would bring reviewer's attention on this specific regional-specific medical practice. Also, sometimes some specific nomenclature convention in one particular linguistic region or nationality may also be associated with a higher reporting rate of certain event. Such patterns could be detected by leveraging the AR mining as well. In this type of applications, AR mining could be performed in the full population or subpopulation of interest, and the rules to be detected are of the form {AE1, AE2, etc.} →{characteristic Z}, where the status of a demographical characteristics is put as the rule head *B* and the AE(s) reported is put as the rule body *A*.

In practice, the application of AR mining can be generalized to other safety endpoints such as laboratory outcomes and vital signs. In terms of the clinical trial database, in addition to data from one single clinical trial of adequate size (usually a phase III trial), data pooled from multiple clinical trials, such as integrated data supporting clinical summary of safety for regulatory submission, could also be used to perform the AR mining.

3. Case Study

For purposes of illustrating the application of AR mining, results from the SOLVD study are briefly discussed here. SOLVD (studies of left ventricular dysfunction) [15, 16] was an extensive clinical study program including two large randomized clinical trials (treatment studies to evaluate the effects of enalapril on mortality, morbidity and quality of life of patients with left ventricular dysfunction) with

several sub-studies, and a large registry study (designed to describe the clinical course of an unselected group of patients). The data utilized for the purposes of this cases study is from one treatment study of SOLVD conducted in late 1980s to early 1990s. The objective of this treatment study is to determine if intervention with enalapril would reduce mortality and hospitalizations in patients with low ejection fraction. This is a multi-center, randomized, double-blind and controlled clinical trial including two treatment arms: 1285 subjects on enalapril arm, and 1284 subjects on placebo arm.

The AR mining was applied to the data from this trial. The threshold for support is set as ≥0.1% in the AR mining here. As discussed in section 2, the threshold on support defines the desired prevalence of the event in the rule. In general, the threshold on support is set at a higher level if all AEs are collected uniformly in the same fashion. However in this study, only 10 AEs ("altered taste", "angioneurotic edema", "blurred vision", "cough", "dizziness/fainting", "fatigue", "forgetfulness", "nausea", "skin rash", and "symptomatic hypotension") were pre-specified and were collected in a systematic way at each clinic visit; all other AEs were only collected sporadically as "other" AEs, supplied with the free text of the event. As a result, out of the over 23000 AE records in the data, less than 5% were those AEs reported as "other" events that were not the 10 pre-specified AEs. To accommodate this specific characteristic of the AE data from this trial, we adjusted the support to ≥0.1% when applying the AR mining, in order for the "other" AEs to also get included in this analysis.

All the AE reported terms in this dataset were first coded into preferred terms according to MedDRA version 17.0. To explore the potential drug related AEs as described in 1) of section 2, all rules are detected by setting the threshold as support >= 0.1% and confidence >= 60%. Among all the detected rules, orthopnoea is an AE that associated with the investigational drug enalapril. The rules with high confidence that involves the event orthopnoea are listed as follows. The support, confidence, and lift for each rule are listed. The numerator (number of subjects with the event on the left-hand side and they were also on enalapril) and denominator (number of subjects with the event on the left-hand side) of the confidence are also presented in the parenthesis following the value of confidence.

Rule	**Support(%)**	**Confidence**	**Lift**
{orthopnoea} → {enalapril}	0.3	100% (7/7)	1.95
{orthopnoea, fatigue} → {enalapril}	0.3	100% (7/7)	1.95
{orthopnoea , fatigue, dizziness, syncope} → {enalapril}	0.2	100% (5/5)	1.95
{orthopnoea, fatigue, memory impairment} → {enalapril}	0.1	100% (3/3)	1.95
{orthopnoea, fatigue, vision blurred, cough} → {enalapril}	0.1	100% (3/3)	1.95
{orthopnoea, fatigue, dizziness, syncope, nausea} → {enalapril}	0.1	100% (3/3)	1.95
{orthopnoea, fatigue, dizziness, syncope, cough} → {enalapril}	0.1	100% (3/3)	1.95

From the resulting rules, it demonstrates that the AE orthopnoea is strongly associated with the treatment of enalapril. Out of the 7 subjects who had orthopnoea, all of them were on the enalapril arm and none of them were on the placebo arm. The lift is also higher than 1, indicating it is highly unlikely that the event orthopnoea is independent of treatment. Together with orthopnoea, the resulting rules about drug associated AEs above also includes some other AEs, such as fatigue, dizziness, syncope, etc. The driving factor for all these rules about drug associated AEs is the single event orthopnoea as the very first rule indicates ({orthopnoea} → {enalapril} with confidence of 100%) The other rules show that the orthopnea

event is sometimes accompanied by some other AEs such as fatigue, but without orthopnea, those other AEs do not associate with enalapril since no association rule supporting was detected with the threshold support >= 0.1% and confidence >= 60%. For example, although {orthopnoea, fatigue} → {enalapril} is a strong rule with confidence being 100%, out of all the 1743 subjects with fatigue in the data, only 862 were on enalapril, namely, the rule {fatigue} → {enalapril} only has confidence as 49%.

Another treatment associated event also detected in this AR mining exercise is palpitations. The corresponding rules detected are listed below.

Rule	Support(%)	Confidence	Lift
{palpitations} → {Enalapril}	0.8	60% (18/30)	1.2
{palpitations, nausea} → {Enalapril}	0.4	90% (9/10)	1.8
{palpitations, nausea, fatigue, dizziness, syncope} → {Enalapril}	0.4	89% (8/9)	1.7
{palpitations, memory impairment} → {Enalapril}	0.4	80% (8/10)	1.6
{palpitations, memory impairment, fatigue, dizziness, syncope} → {Enalapril}	0.3	78% (7/9)	1.5
{palpitations, nausea, memory impairment} → {Enalapril}	0.3	100% (7/7)	1.9
{palpitations, nausea, memory impairment, fatigue, dizziness, syncope} → {Enalapril}	0.3	100% (6/6)	1.9

Unlike the event orthopnoea, palpitations by itself is associated with the treatment to some extent, but is associated with treatment more strongly through a cluster of a few other AEs. The confidence for the rule {palpitations} → {Enalapril} is 60%, while the confidence for {palpitations, nausea} → {Enalapril} is 90%, and that for {palpitations, nausea, memory impairment} → {Enalapril} is 100%. In other words, only 60% subjects with palpitations are on enalapril, but for subjects with palpitations, if they also had nausea and memory impairment, they are 100% on enalapril.

To further investigate the clustering behaviour of palpitations with other AEs, AR mining was later performed in each treatment group of the trial to detect rules of the form {AEs} → {palpitations}. In the enalapril group (18 subjects with palpitations), there are >20 rules of {AEs} → {palpitations} with confidence >= 40%. Two examples of detected rules are listed below. This result indicates that the event palpitations tends to co-occur with certain other AEs in patients on enalapril.

Rule	Support(%)	Confidence	Lift
{fatigue, memory impairment, nausea, vision blurred, dysgeusia, diarrhoea} → {palpitations}	0.2	50% (2/4)	30.9
{fatigue, memory impairment, nausea, vision blurred, diarrhoea} → {palpitations}	0.2	40% (2/5)	24.7

On the other hand, when AR mining is performed in the placebo group (12 subjects with palpitations), there is no rule of the form {AEs} → {palpitations} with the minimum confidence set as low as 3%. The event palpitations occurred more randomly in the placebo group, and no pattern is observed. Clearly there

is some difference in the clustering behaviour of palpitations with other AEs in the two treatment groups. The clustered events in the enalapril group, such as palpitations together with nausea and memory impairment, may be a manifestation of some underlying similarity among the subjects who had the cluster of the events. Careful clinical review is desired to interpret such interesting patterns. Other data domains such as demography and medical history, etc., should be investigated to aid the review.

4. Discussion

Association rule mining is a powerful tool to analyse clinical trial safety data, and could provide additional insight into the data and detect important patterns that cannot be identified by standard descriptive summary or traditional statistical methods based on aggregate data. As discussed in this paper, AR mining could be applied to clinical trial safety data with great flexibility to answer unique questions related to drug safety. Performing an AR mining analysis is relatively straightforward because the fast algorithms have now been implemented in many statistical software packages. However, the results of AR mining analysis depend directly on the thresholds assigned to the support and confidence. As we have seen, the choice of threshold may be influenced by several factors such as the actual patient population, type of database, and purpose of safety review. It is currently determined in a case by case fashion, and there is no general guidance on the choice of threshold for AR mining applied in clinical trial safety data. Various thresholds may be utilized to potentially find clinical meaningful associations. AR mining has great potential in its application to analyse clinical trial safety data, but similar to all other data mining methods, AR analysis should be applied to large datasets (such as one large scale phase III trial, or integrated safety data from multiple phase II/III trials) and it has limited value in small datasets. In addition, we need to be mindful that data mining serves as an initial step in safety evaluation, and a combination of statistical inference and medical judgment is required as part of the signal detection and evaluation process.

Acknowledgments

It is recognized that Drs William Y. Go and Liron Walsh contributed equally to this manuscript. The authors would like to thank Dr Steven Snapinn for his careful review and comments.

References

[1] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society B 1995; 57:289–300.

[2] Mehrotra DV, Heyse JH. Use of the false discovery rate for evaluating clinical safety data. Statistical Methods in Medical Research 2004; 13:227–238

[3] Mehrotra D and Adewale A, Flagging clinical adverse experiences: reducing false discoveries without materially compromising power for detecting true signals Statist. Med. 2012, 31 1918–1930

[4] MedDRA (Medical Dictionary for Regulatory Activities) by MSSO (Maintenance and Support Services Organization) http://www.meddramsso.com/

[5] Berry S and Berry D, Accounting for Multiplicities in Assessing Drug Safety: A Three-Level Hierarchical Mixture Model, Biometrics 60, 418–426

[6] Xia A, Ma H, and Carlin B, Bayesian Hierarchical Modeling for Detecting Safety Signals in Clinical Trials Journal of Biopharmaceutical Statistics, 21: 1006–1029, 2011

[7] Southworth, H. and O'Connell, M. Data mining and statistically guided clinical review of adverse event data in clinical trials. Journal of Biopharmaceutical Statistics, 2009; 19: 803-817.

[8] Agrawal, R., Imielinski, T., Swami, A. Database mining: A performance perspective. *IEEE* Transactions on Knowledge and Data Engineering, 1993; 5: 914-925.

[9] Agrawal, R., Srikant, R. Fast algorithms for mining association rules in large databases. Proceedings of the 20th International Conference on Very Large Data Bases, VLDB, Santiago, Chile, 1994; 487-499.

[10] Zaki, M. Scalable algorithms for association mining. IEEE Transactions on Knowledge and Data Engineering, 2000; 12: 372-390.

[11] Han, J., Pei, J., Yin, Y., Mao, R. Mining frequent patterns without candidate generation. Data Mining and Knowledge Discovery, 2004 8: 53-87.

[12] Brin, S., Motwani, R., Ullman, J., Tsur, S. Dynamic itemset counting and implication rules for market basket data. Proceedings ACM SIGMOD International Conference on Management of Data, pages, SIGMOD, Tucson, Arizona, USA, 1997; 255-264.

[13] Hastie, T., Tibshirani, R., Friedman, J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer-Verlag 2001.

[14] SMQ (Standardised MedDRA Queries) by MedDRA http://www.meddra.org/standardised-meddra-queries/

[15] The SOLVD Investigators. Studies of left ventricular dysfunction (SOLVD): rationale, design and methods: two trials that evaluate the effect of enalapril in patients with reduced ejection fraction. American Journal of Cardiology. 1990; 66:315-322.

[16] The SOLVD Investigators. Effect of enalapril on mortality and the development of heart failure in asymptomatic patients with reduced left ventricular ejection fractions. New England Journal of Medicine. 992; 327:685-691.