Paper 2470-2018

Association Rule Mining of Polypharmacy Drug Utilization Patterns in Health Care Administrative Data Using SAS® Enterprise Miner™

Dingwei Dai, Chris Feudtner, The Children's Hospital of Philadelphia

ABSTRACT

Pediatric polypharmacy is prevalent in the both outpatient and inpatient setting, and is associated with increased exposure to adverse drug events (ADEs). The current definition of polypharmacy is simply a count of concurrent medication exposures for a given patient being equally to or greater than five, but this patient-level approach does not offer insights regarding the reasons why polypharmacy occurs that could be gained by a complementary drug-level approach, examining polypharmacy patterns not only across patients but also across medications. Health care administrative databases offer opportunities to examine drug-level relationships and patterns. Association rule mining (ARM) is a well-established data mining technique that has been commonly used for mining commercial transactional databases. Link Analysis (LA) is a popular social network analysis technique that is used to discover and visualize associations between different items. We transformed administrative data to a transaction data format suitable for mining rules and applied ARM and LA to analyze drug utilization and polypharmacy patterns in health care administrative database using SAS® Enterprise Miner™. Our results demonstrate how ARM can find co-utilization associations among drugs, thereby enabling the description of various polypharmacy patterns, and the subsequent detection of patient characteristics associated with these patterns. We believe that ARM and LA, along with their visualization graphics, provide valuable methods to analyze drug utilization and polypharmacy, and that this approach could be used in mining other databases such as administrative claims data and electronic medical records.

Key Words: *polypharmacy; adverse drug events; health care administrative databases; association rule mining; link analysis; data visualization; Pharmacoepidemiology*

INTRODUCTION

Polypharmacy, typically defined as exposure of a patient to 5 or more medications concurrently, is known to pose an increased hazard to patients of experiencing an adverse drug event (ADE) [1]. ADEs are major cause of increased morbidity, mortality, and health care costs: ADEs are ranked as the 4th-to-6th leading causes of death in inpatients [1,2], and in the United States, for every dollar spent on medication in 2000, more than a dollar was estimated to have been spent on direct medical costs related to drug misadventures [3]. Although often considered to be of most concern for the elderly, polypharmacy is common among pediatric patients in the United States [4-6], resulting in many patients who are treated with polypharmacy regimens, especially for children with complex chronic conditions, being exposed to multiple potential drug-drug interactions [5].

Currently the prevailing method of assessing polypharmacy is simply a count the number of generic drug concurrently used by patients [4-6]. This patient-level approach offers a limited range of insight regarding why polypharmacy occurs, since the combinations of particular drugs is lost by the simplicity of a single number for each patient. We therefore sought to examine polypharmacy from a drug-level analytic approach. Specifically, we considered how data mining could be applied, as this methodology offers great potential for drug safety analytics [7,9]. For example, Chen et al. applied association rule mining (ARM) to a linked dataset comprised of a pharmaceutical prescribing dataset and a hospital admissions dataset in order to identify groups of patients who are more likely to have an ADE to ACE inhibitors [8]. Zhu et al performed clinical trial safety data analytics using ARM [9]. Harpaz et al. mined muti-item drug ADEs associations in spontaneous reporting systems [10]. To date, data mining approaches have mostly focused on identifying associations between specific drugs and either specific ADE or specific clusters ADEs.

In this paper we describe our application of ARM and LA in health care administrative databases to analyze and visualize polypharmacy drug utilization patterns at the drug-level en mass, and then to

examine the association of different polypharmacy patterns with patient demographic and clinical characteristics.

ASSOCIATION RULES AND DRUG UTILIZATION

ARM finds interesting association or correlation relationships among a large set of data items. With massive amounts of data continuously be stored in databases and available, many industries are becoming interested in mining association rules from their databases. ARM was initially used for market basket analysis, where interested lies in finding associations between the different items that customers place in their “shopping baskets” as recorded in commercial transaction databases. The discovery of such associations can help retailers develop marketing strategies by gaining insight into which items are frequently purchased together by customers. We applied ARM in clinical administrative databases to analyze drug utilization patterns and polypharmacy, polypharmacy associated certain medical conditions, or certain patient subpopulations.

The basic concepts of AMR were established originally on transaction data. Initially AMR was implemented by Agarwal et al [11,12]. ARM has been divided into two phases of process as follows: Phase 1: Identify the sets of frequent items or itemsets or pattern within the set of transaction using user-specified support threshold. Phase 2: Generate inferences or rules from these above patterns using user-specified confidence threshold. The above two phases are generated strong association rules from dataset. The first phase is called frequent itemset construction or mining. That is extremely computational expensive than phase 2. The second phase is called association rule generation. That is, straight forward process. In a given database, denote I = {i1, i2, …, ik} as a set of k distinct items. A set of items is also called as an itemset. An association rule (AR) is a pair (A, B) of sets of items, denoted by A → B. A is the antecedent and B is the consequent of the rule A → B. The essential parameters associated to an AR are its support and confidence. The support measures the “prevalence” of A U B. It is defined as the fraction of transactions in the database which contain all items in a specific rule. This can be written as:

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

Where #{A U B} is the number transactions (itemset) which contain both A and B and N represents the total number of transactions (itemset) in the database. The confidence measures the “predictability” of the rule:

$$\text{Confidence } (A \longrightarrow B) = \frac{\text{Support } (A \longrightarrow B)}{\text{Support } (A)} = \frac{\#\{A \cup B\}}{\#\{A\}}$$

Clearly, the confidence of A → B is an estimation of probability that a record that contains the items A, chosen at random, will contain the items B. Where #{A} is the number transactions (itemset) which contain A. Another measure of the rule is lift.

$$\text{Lift } (A \longrightarrow B) = \frac{\text{Confidence } (A \longrightarrow B)}{\text{Expected Confidence } (A \longrightarrow B)} = \frac{\#\{A \cup B\} \times N}{\#\{A\} \times \#\{B\}}$$

Lift is the ratio of the confidence of a rule to the expected confidence of the rule. The expected confidence is calculated under the assumption that the left hand side of a rule is independent from the right hand side of the rule. Consequently, lift is a measure of association between left hand side and right hand side of the rule. Values that are greater than one represents positive association between A and B. Values that are equal to one represent A and B are independent. Values that are less than one represents negative association between A and B. The lift serves as a reference to interpret the importance of the AR.

In the transaction data, for the rule A → 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 administrative data, the rule A → B can be mined using different aspect of the databases to answer many important pharmacoepidemiology questions, such as all specific drug use in specific population, specific drug polypharmacy,

polypharamacy associated medical conditions, procedures and ADEs. The rules should not be interpreted as a direct causation, but as an association between two or more drugs. AR does not create rules about repeating items, such as “if item A is part of an event, then another item A is also part of the event X% of the time.” In association analysis, it does not matter whether an individual patient takes one or multiple pills of drug A in the day; only the administration of drug A in the prescription is relevant. However, identifying creditable associations between two or more specific drugs (drug-drug combination or polypharmacy) can help clinicians make decisions such as when to choose which medications to adjust, when they intend to escalate or de-escalate medications as a patient’s condition or stage changes.

LINK ANALYSIS

LA, also known as Social Network Analysis, is a mathematical and graphical analysis highlighting the linkage between items (objects), including drugs, medical conditions, organizations, people and transactions. LA is used to identify and visualize the associations (links) between different items and discover item clusters and has been used for investigation of criminal activity (fraud detection, counterterrorism, and intelligence), computer security analysis, search engine optimization, market research, medical research, and art [13]. The Link Analysis node in SAS Enterprise Miner provides a very powerful but easy to use platform for performing LA and visualize the results in a succinct manner. The Link Analysis node processes transactional data in the following steps [14,15]:

1. The node first discovers association or sequence rules computing confidence, support, expected confidence and lift using equations stated above.
2. The link analysis node then transforms the rules data into a network graph data in the form of nodes and links where the support of each item becomes the node weight and the strength of the association (confidence of the rule) becomes the link weight. The two-item sequence rules are transformed into a directed graph data and the association rules into an undirected graph data.
3. The node calculates several centrality measures and detects item clusters from the link graph.
4. Finally, it scores the transactional data. There are two score properties under the link analysis node. The node can either produce a next-best-offer list using the association /sequence rules or produce customer segmentation information for scoring using the item clusters.

There are some overlaps on association or sequence rules discovery between AR and LA. For more information, see the SAS Enterprise Miner: References Help. In this paper, we applied ARM to analyze drug utilization and polypharmacy and used LA to visualize the drug-drug combination patterns.

METHODS

DATA SOURCES

This study used the Pediatric Health Information System (PHIS) database, a national health care administrative database containing resource utilization data for pediatric inpatient, emergency department, ambulatory surgery, and observation unit patient encounters from 49 not-for-profit, tertiary care pediatric hospitals in the US, representing most major U.S. metropolitan areas and approximately 85% of freestanding pediatric acute care hospital admissions in the United States [16]. These participating hospitals are located in 32 US states and the District of Columbia. All these hospitals are affiliated with the Children’s Hospital Association (CHA, Kansas City, KS), a national business alliance of children’s hospitals. The PHIS database includes patient demographics, diagnosis, and procedures as well as detailed pharmacy information, including ICD-9-CM (International Classification of Diseases, Ninth Revision, Clinical Modification) codes, CPT (Current Procedural Terminology) codes, and CTC (Clinical Transaction Classification) codes for each diagnosis, procedure, generic drug entity dispensed, and clinical services for each day of hospital stay of each patient (Figure 1). Each patient in PHIS is assigned a unique identifier allowing records to be linked among different datasets and longitudinally linked. Details about the PHIS database have been reported previously [4,5]. A joint effort between participating hospitals, a data manager (Truven Health Care Analytics, IBM), and the CHA ensures maintenance of data quality and reliability. PHIS is a relational database and all data in PHIS are deidentified. Our study included all children <18 years of age, who admitted to ED, and were discharged from participating hospitals between 1 January 2014, and 31 December 2014.

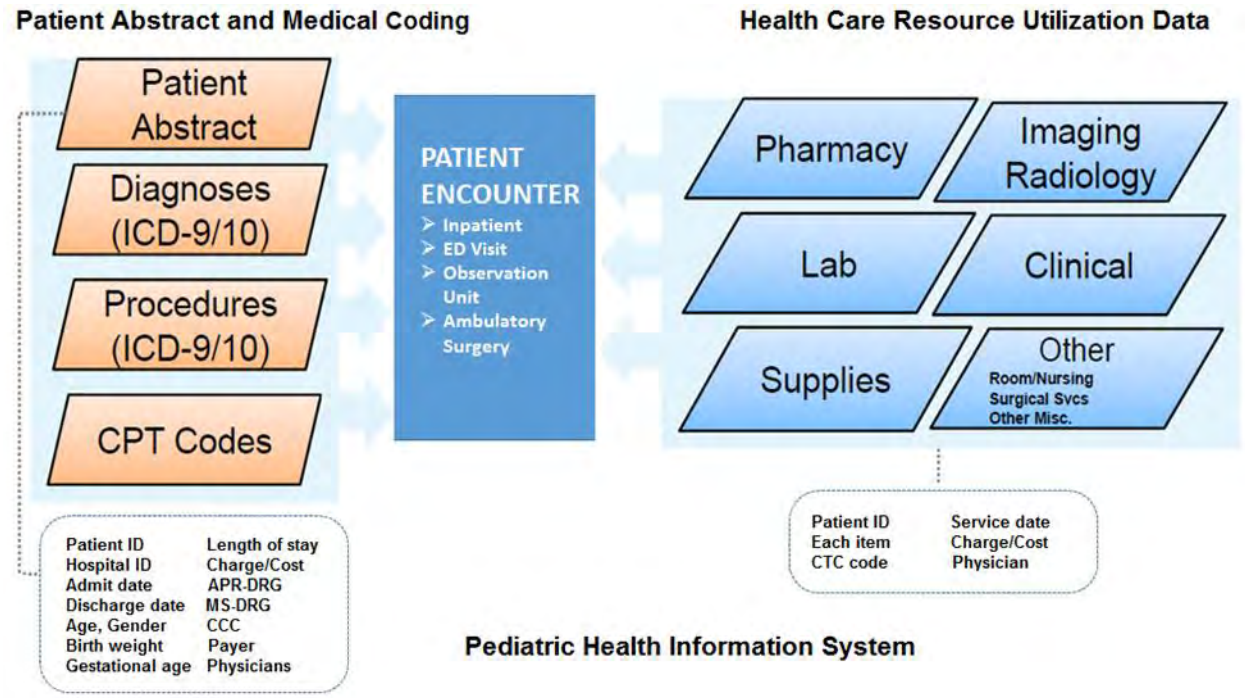


Figure 1. PHIS Databases

DATA PREPARATION

In this project, we used only two datasets in the databases. One is Patient Abstract, which contains demographic characteristics, up to 41 diagnosis codes, and up to 25 procedure codes; the other one is Pharmacy which contains detail daily drug use including service date, drug name, dose, route of administration, and charges. We first cleaned the data, corrected data errors, removed some duplicates, and made service day variable consistent cross all hospitals. For these two data, we selected only variables we needed. Then we mapped the administrative datasets to transactional data formats. The process consisted of three steps. (1). drug names were mapped to their corresponding generic names to reduce drug naming redundancy and reduce computing time. We implemented this by using previously developed PHIS drug dictionary [4]. There are 28 variables in the pharmacy data, we kept only Patient_ID, Service_Day, Generic_Name (Table 1). The variable Service_Day identifies the sequence in which the drugs were used. In this example, all the drugs were used at the same day. We do not know the order in which the drugs were administered by nurses. When the sequence is taken into account, association analysis is known as sequence analysis. Sequence analysis is not demonstrated in this paper. (2). For the purpose of this study, we selected only Patient ID, age, and up to 41 diagnoses from the Patient Abstract data (Table 2). Patient records data were then transformed into transactional format (Table 3). (3). Create the work dataset by combining all generated data (Table 1 and Table 3). All data management were conducted by using SAS version 9.4 (SAS Institute Inc., Cary, NC).

Table 1: Pharmacy Data

Patient ID	Service Day	Generic Name
P00000001	1	Acetaminophen
P00000001	1	Albuterol sulfate
P00000001	1	Amoxicillin trihydrate

P00000001	1	Ceftriaxone sodium
P00000001	1	Dexamethasone
P00000001	1	Prednisolone
P00000002	1	Acetaminophen
P00000002	1	Heparin sodium
P00000002	1	Lidocaine HCl
P00000002	1	Morphine sulfate
P00000003	1	Albuterol sulfate
P00000003	1	Ibuprofen
…		
P00010000	1	Ibuprofen

Table 2: Patient Diagnosis (Administrative data format)

Patient ID	Count_DX	DX1	DX2	DX3	…	DX41
P00000001	3	Asthma	Otitis Media	Vomiting		
P00000002	3	Headache	Abdominal pain	Acute lymphoid leukemia		
P00000003	2	Acute bronchitis	Unspecified viral infection			
…	..					
P00010000	1	Pharyngitis				

Table 3: Patient Diagnosis (Transactional data format)

Patient ID	DX
P00000001	Asthma
P00000001	Otitis Media
P00000001	Vomiting
P00000002	Headache
P00000002	Abdominal pain
P00000002	Acute lymphoid leukemia
P00000003	Acute bronchitis
P00000003	Unspecified viral infection
…	
P00010000	Pharyngitis

Table 4: Patient Diagnosis and Drug Use (work data)

Patient ID	Rx DX
P00000001	Acetaminophen
P00000001	Albuterol sulfate
P00000001	Amoxicillin trihydrate
P00000001	Ceftriaxone sodium
P00000001	Dexamethasone
P00000001	Prednisolone
P00000001	Acetaminophen
P00000001	Asthma
P00000001	Otitis Media
P00000001	Vomiting
P00000002	Acetaminophen
P00000002	Heparin sodium
P00000002	Lidocaine HCl
P00000002	Morphine sulfate
P00000002	Acetaminophen
P00000002	Abdominal pain
P00000002	Acute lymphoid leukemia
P00000003	Albuterol sulfate
P00000003	Ibuprofen
P00000003	Acute bronchitis
P00000003	Unspecified viral infection
…	
P00010000	Ibuprofen
P00010000	Pharyngitis

RUN THE ASSOCIATION NODE AND THE LINK ANALYSIS NODE

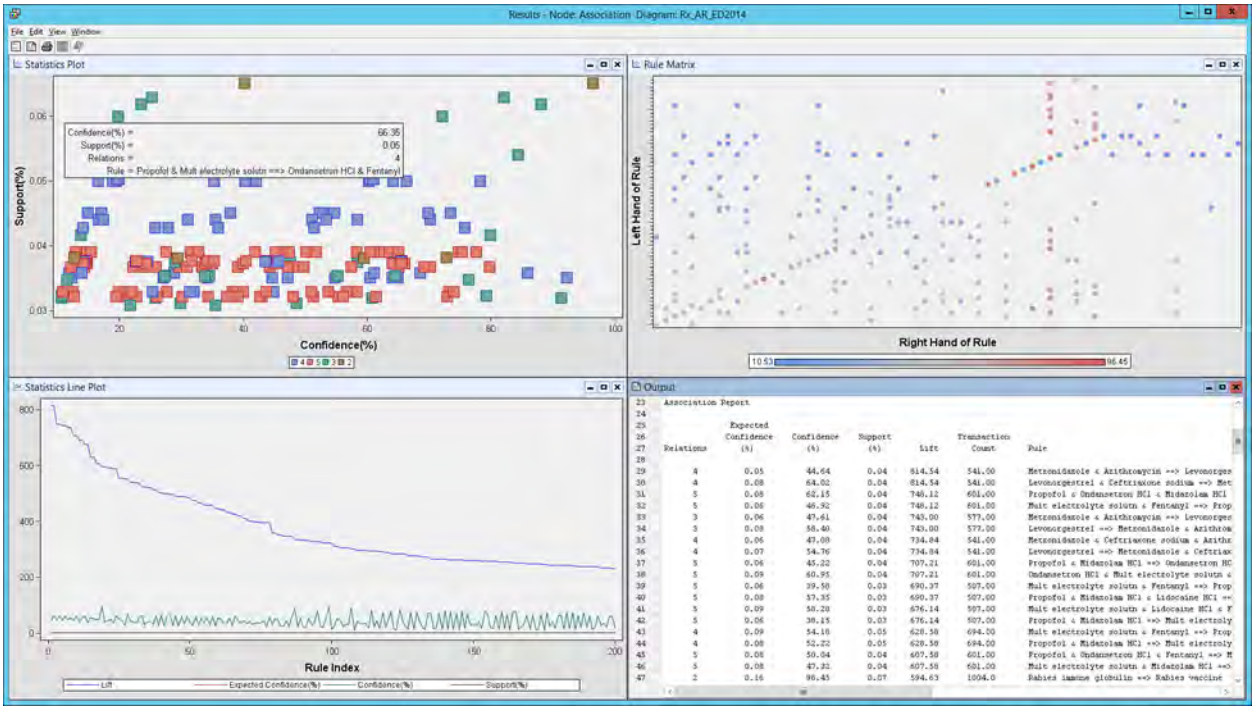
Observe the Association properties subgroup of the Association node (SAS® Enterprise Miner™14.2, SAS Institute Inc., Cary, NC). These properties determine how large each association can be and how association rules are formed. Set the value of the Maximum Items property to 5. This indicates that only associations among up to five drugs are generated. In many drug safety reports, some specific drug led to clinical side effects or one or more ADEs. ADEs can involve single or multiple drugs and describe single or multiple adverse drug interactions. Our previous studies that ≥3 drug combinations are very common [4,5]. A greater number of drugs used concurrently may have a potentially multiplicative increase in the number of ADEs, especially for specific drug classes such as sedation [17] or QT-prolonging medications [18].In the absence of a gold standard (the set of all true multi-drug combinations and the frequency of the combinations are unknown) which would have enabled us to calibrate or determine the

most appropriate thresholds in a quantitative manner, we resorted to a data driven and heuristic approach guided by clinical expert knowledge. The support threshold was set high enough value to highlight the more frequent drug-drug combination patterns. Setting the threshold to a lower value resulted in a much larger set of associations. We set the threshold as support ≥ 0.01% in this example. As we discussed in previous section, The Confidence for the rule is the conditional probability that drug B is in the prescription given that drug A is present. Rules with highest confidence and support percentages are recommended to discover high correlation items. Our objective is to find more specific drug-drug combinations which may pose a risk for ADEs, so we set Confidence threshold as 1% in this example.

We set the thresholds for both support and confidence same as running the Association Node. The remaining properties are set to the default. Once the settings are completed, connect the input data node with the Association node and the Link Analysis node respectively and run the project.

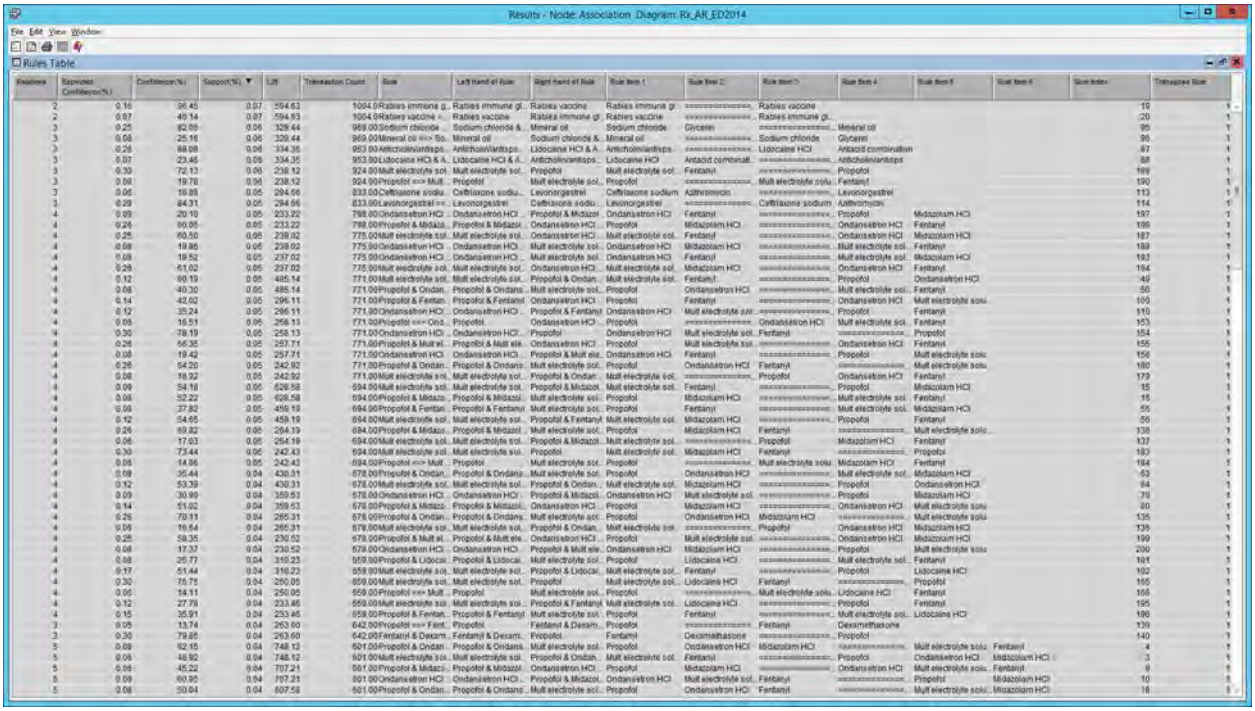
RESULTS

The Results window of the Association node shows the association statistics (Display 1). When one click any square (color coded based on the number of drug-drug combination, brown means 2 drugs combination, green 3, blue 4, red 5), the window displays the rule and the two association parameters: support and confidence.



Display 1. The Association Statistics

In the Results window of the Association node, select View ➡ Rules ➡ Rule Table from the main menu, it will display the Rules Table (Display 2). The Rules Table displays information about each rule that was created. This includes the rule, support, confidence, expected confidence, lift, number of occurrence, and the items in the rule. For this project, the rule is the drug-drug combination (polypharmacy). The number of occurrence is the number of patients, who were exposed to the drug-drug combination. The support can be used as proportion or prevalence of the drug combination exposed in the population. The Rule Table can be used to screen the drug-drug combinations which have a risk for drug-drug interactions and lead to ADEs. We listed top 5 most common two to five drugs combination exposed among pediatric patients in ED visits in the U.S.A. 2014 (Table 5).



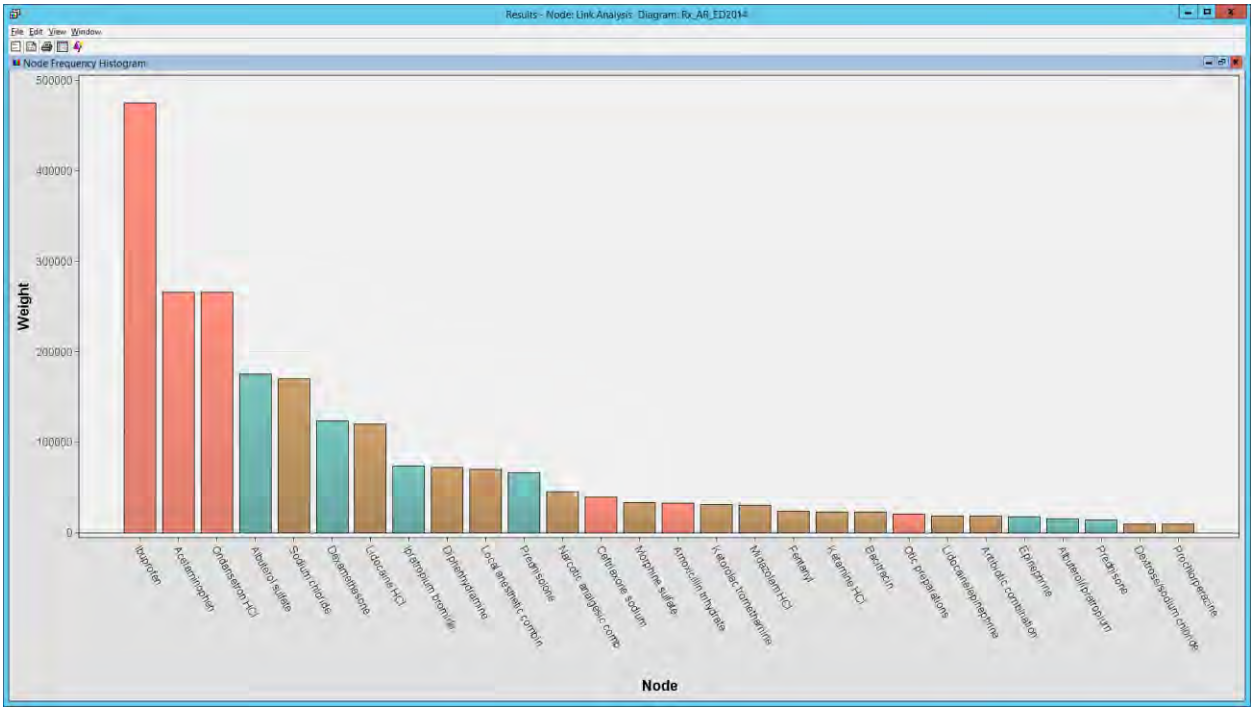
Display 2. Table of association rules from ARM

Table 5. Top 5 most common two to five drug-drug combinations exposed among Sick Children in ED Visits in the US Children's Hospitals, 2014

Drug-drug Combinations	Number of Patients Exposed	%, Exposed
Two drug combinations (D2)		
Albuterol sulfate & Ipratropium bromide	72,956	2.37
Albuterol sulfate & Prednisolone	43,769	1.42
Albuterol sulfate & Dexamethasone	42,171	1.37
Ibuprofen + Ondansetron HCl	40,070	1.30
Acetaminophen + Ibuprofen	39,488	1.28
Three drug combinations (D3)		
Albuterol sulfate & Ipratropium bromide & Prednisolone	26,805	0.87
Albuterol sulfate & Dexamethasone & Ipratropium bromide	26,175	0.85
Lidocaine HCl & Ondansetron HCl & Sodium chloride	11,376	0.37
Albuterol sulfate & Ibuprofen & Ipratropium bromide	9,292	0.27
Acetaminophen & Ibuprofen & Ondansetron HCl	6,959	0.23
Four drug combinations (D4)		
Diphenhydramine & Ketorolac tromethamine & Prochlorperazine & Sodium chloride	3,326	0.11
Albuterol sulfate & Ipratropium bromide & Prednisolone &	3,017	0.10

Sodium chloride		
Albuterol sulfate & Ibuprofen & Ipratropium bromide & Prednisolone	2,901	0.10
Albuterol sulfate & Ipratropium bromide & Ibuprofen & Dexamethasone	2,587	0.09
Albuterol sulfate & Acetaminophen & Dexamethasone & Ipratropium bromide	2,341	0.08
Five drug combinations (D5)		
Diphenhydramine & Ketorolac tromethamine & Lidocaine HCl & Prochlorperazine & Sodium chloride	1,433	0.05
Fentanyl & Midazolam HCl & Mult electrolyte solutn & Ondansetron HCl & Propofol	601	0.03
Fentanyl & Lidocaine HCl & Midazolam HCl & Ondansetron HCl & Propofol	580	0.02
Fentanyl & Lidocaine HCl & Midazolam HCl & Mult electrolyte solutn & Ondansetron HCl	574	0.02
Fentanyl & Lidocaine HCl & Mult electrolyte solutn & Ondansetron HCl & Propofol	566	0.02

In the Results window the Link Analysis node, select View ➡Plots ➡ Node Frequency Histogram (by item cluster) from the main menu, it will display the Node Frequency Histogram (Display 3), which shows frequency of specific drug utilization. The weight represents number of patients who took the particular drug. One can see that ibuprofen is the most common drug used among pediatric patients in ED visits.



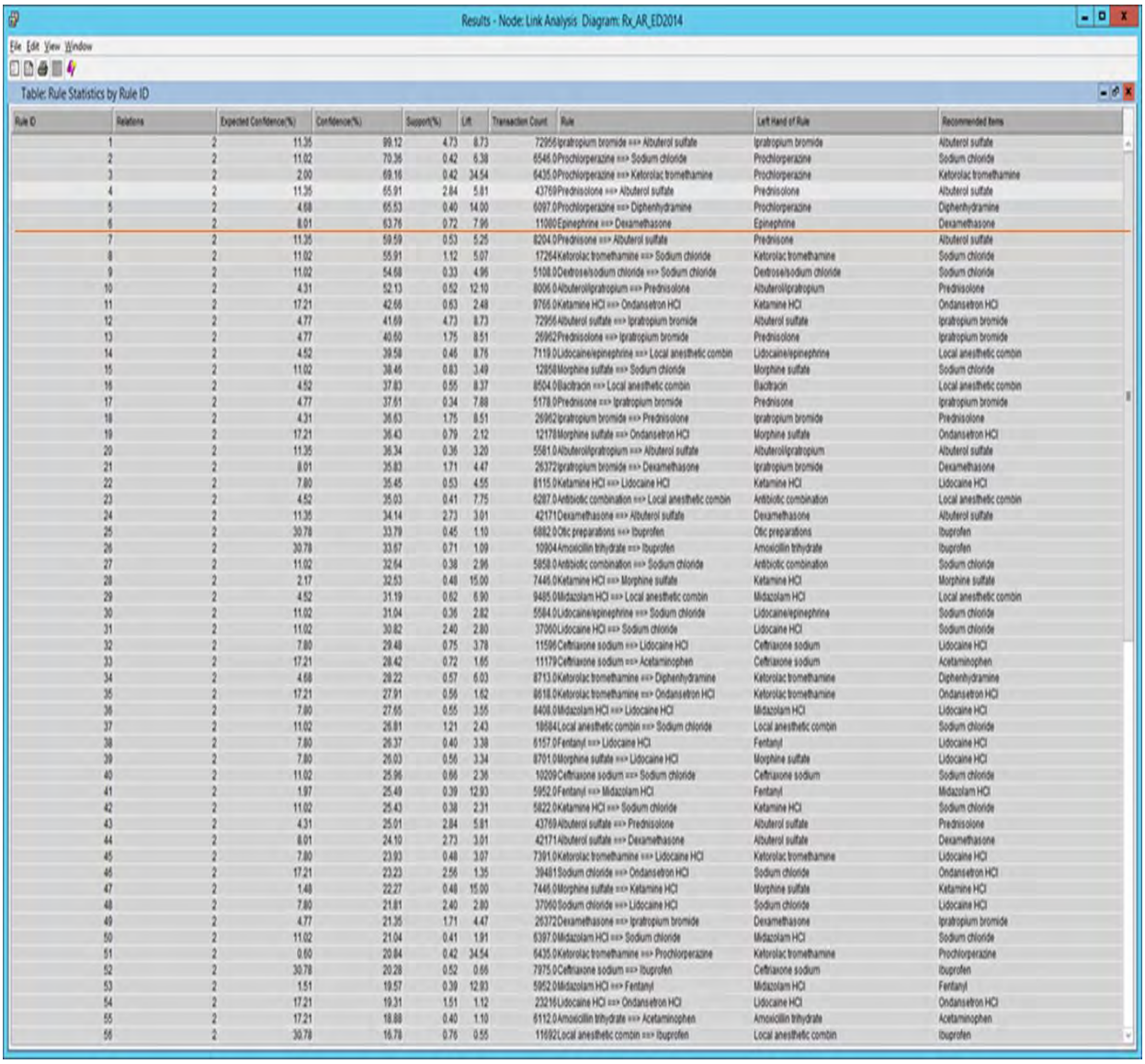
Display 3: Node Frequency Histogram

The prevalence of drug use can be calculated by dividing the number of patients who had used the certain drug by total number of patients in the cohort. The top 12 most commonly used drugs listed in Table 6.

Table 5. Top dozen most common drug used among Sick Children in ED Visits in the US Children's Hospitals, 2014

Generic Drug Name	Number of Patients Exposed	%, Exposed
Ibuprofen	474,589	15.43
Acetaminophen	265,448	8.63
Ondansetron HCl	265,355	8.62
Albuterol sulfate	174,998	5.69
Sodium chloride	169,930	5.52
Dexamethasone	123,520	4.02
Lidocaine HCl	120,244	3.91
Ipratropium bromide	73,605	2.39
Diphenhydramine	72,182	2.34
Prednisolone	66,408	2.16
Ceftriaxone sodium	39330	1.28
Morphine sulfate	33429	1.09

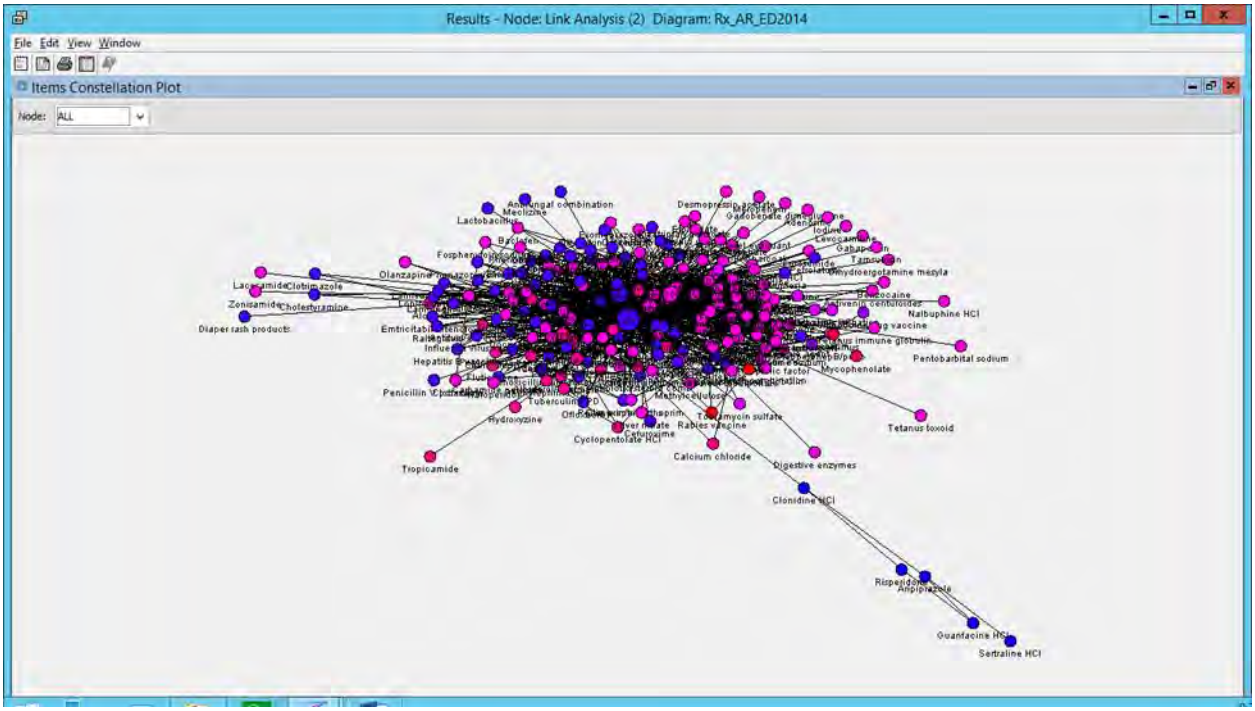
In the results window of the Link Analysis node, users can find view a table of association rules. Below is a screenshot of part of the rule table (Display 4). We will take rule 6 as example to illustrate how to read this table. The expected confidence of rule 6 is 8.01%, indicating that epinephrine will be exposed to 8.01% of the patients who have ED visits and a support of 0.72% indicates that epinephrine and dexamethasone are used together by 0.72% of patients during the ED visit. When a patient already took epinephrine, he/she has a 63.76% of chance to also take dexamethasone as shown by the confidence, which is 7.69 times the chance of taking dexamethasone within the population, shown by the lift value of 7.69. This information can be useful to understanding drug utilization patterns. LA can generate only two item (drug) association rules. This table is similar to the Rules Table (Relations = 2) in the results window of the Association Node (Display 2), which can generate two or more item association rules.



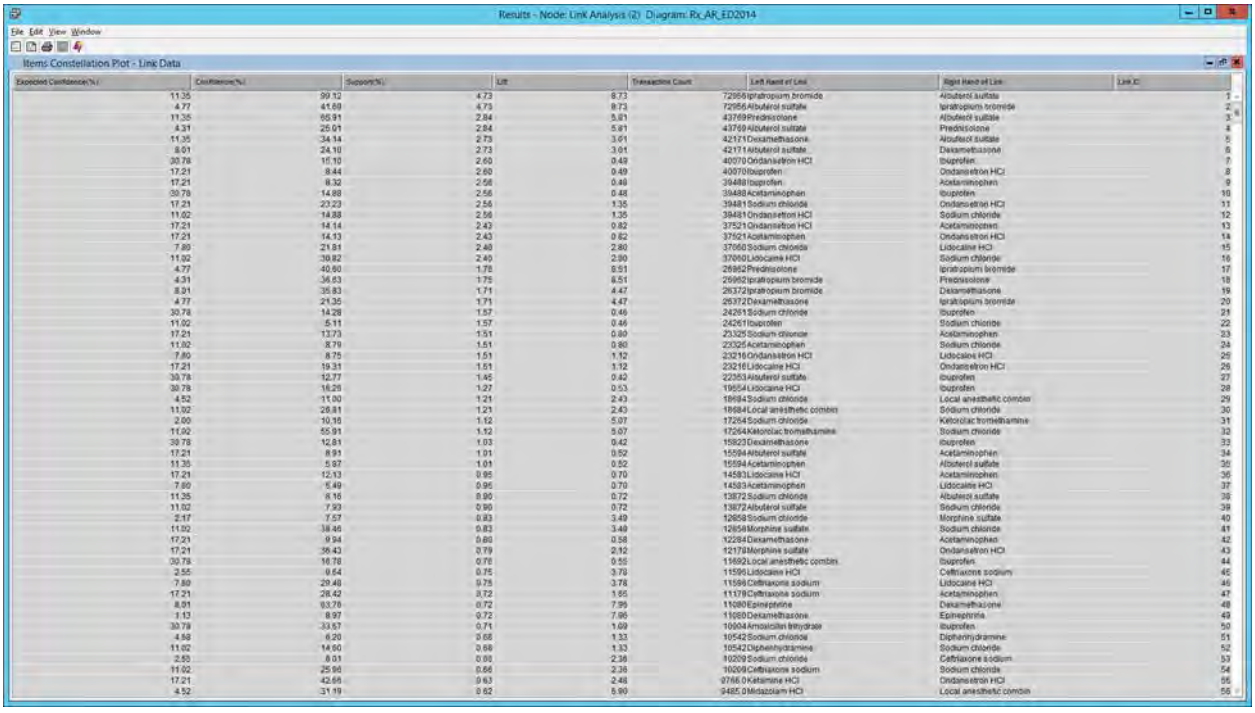
Rule ID	Relations	Expected Confidence(%)	Confidence(%)	Support(%)	Lift	Transaction Count	Rule	Left Hand of Rule	Recommended Items
1	2	11.35	99.12	4.73	8.73	72956	Ipratropium bromide ==> Albuterol sulfate	Ipratropium bromide	Albuterol sulfate
2	2	11.02	70.36	0.42	6.38	6546.0	Prochlorperazine ==> Sodium chloride	Prochlorperazine	Sodium chloride
3	2	2.00	69.16	0.42	34.54	6435.0	Prochlorperazine ==> Ketorolac tromethamine	Prochlorperazine	Ketorolac tromethamine
4	2	11.35	65.91	2.84	5.81	43769	Prednisolone ==> Albuterol sulfate	Prednisolone	Albuterol sulfate
5	2	4.68	65.53	0.40	14.00	6097.0	Prochlorperazine ==> Diphenhydramine	Prochlorperazine	Diphenhydramine
6	2	8.01	63.76	0.72	7.96	11080	Epinephrine ==> Dexamethasone	Epinephrine	Dexamethasone
7	2	11.35	59.59	0.53	5.25	8204.0	Prednisone ==> Albuterol sulfate	Prednisone	Albuterol sulfate
8	2	11.02	55.91	1.12	5.07	17264	Ketorolac tromethamine ==> Sodium chloride	Ketorolac tromethamine	Sodium chloride
9	2	11.02	54.68	0.33	4.96	5108.0	Dextrose/sodium chloride ==> Sodium chloride	Dextrose/sodium chloride	Sodium chloride
10	2	4.31	52.13	0.52	12.10	8006.0	Albuterol/ipratropium ==> Prednisolone	Albuterol/ipratropium	Prednisolone
11	2	17.21	42.66	0.63	2.48	9766.0	Ketamine HCl ==> Ondansetron HCl	Ketamine HCl	Ondansetron HCl
12	2	4.77	41.69	4.73	8.73	72956	Albuterol sulfate ==> Ipratropium bromide	Albuterol sulfate	Ipratropium bromide
13	2	4.77	40.60	1.75	8.51	26962	Prednisolone ==> Ipratropium bromide	Prednisolone	Ipratropium bromide
14	2	4.52	39.58	0.46	8.76	7119.0	Lidocaine/epinephrine ==> Local anesthetic combin	Lidocaine/epinephrine	Local anesthetic combin
15	2	11.02	38.46	0.83	3.49	12858	Morphine sulfate ==> Sodium chloride	Morphine sulfate	Sodium chloride
16	2	4.52	37.83	0.55	8.37	8504.0	Bacitracin ==> Local anesthetic combin	Bacitracin	Local anesthetic combin
17	2	4.77	37.61	0.34	7.88	5178.0	Prednisone ==> Ipratropium bromide	Prednisone	Ipratropium bromide
18	2	4.31	36.63	1.75	8.51	26962	Ipratropium bromide ==> Prednisolone	Ipratropium bromide	Prednisolone
19	2	17.21	36.43	0.79	2.12	12178	Morphine sulfate ==> Ondansetron HCl	Morphine sulfate	Ondansetron HCl
20	2	11.35	36.34	0.36	3.20	5581.0	Albuterol/ipratropium ==> Albuterol sulfate	Albuterol/ipratropium	Albuterol sulfate
21	2	8.01	35.83	1.71	4.47	26372	Ipratropium bromide ==> Dexamethasone	Ipratropium bromide	Dexamethasone
22	2	7.80	35.45	0.53	4.55	8115.0	Ketamine HCl ==> Lidocaine HCl	Ketamine HCl	Lidocaine HCl
23	2	4.52	35.03	0.41	7.75	6287.0	Antibiotic combination ==> Local anesthetic combin	Antibiotic combination	Local anesthetic combin
24	2	11.35	34.14	2.73	3.01	42171	Dexamethasone ==> Albuterol sulfate	Dexamethasone	Albuterol sulfate
25	2	30.78	33.79	0.45	1.10	6882.0	Otic preparations ==> Ibuprofen	Otic preparations	Ibuprofen
26	2	30.78	33.67	0.71	1.09	10904	Amoxicillin trihydrate ==> Ibuprofen	Amoxicillin trihydrate	Ibuprofen
27	2	11.02	32.64	0.38	2.96	5858.0	Antibiotic combination ==> Sodium chloride	Antibiotic combination	Sodium chloride
28	2	2.17	32.53	0.48	15.00	7446.0	Ketamine HCl ==> Morphine sulfate	Ketamine HCl	Morphine sulfate
29	2	4.52	31.19	0.62	6.90	9485.0	Midazolam HCl ==> Local anesthetic combin	Midazolam HCl	Local anesthetic combin
30	2	11.02	31.04	0.36	2.82	5584.0	Lidocaine/epinephrine ==> Sodium chloride	Lidocaine/epinephrine	Sodium chloride
31	2	11.02	30.82	2.40	2.80	37060	Lidocaine HCl ==> Sodium chloride	Lidocaine HCl	Sodium chloride
32	2	7.80	29.48	0.75	3.78	11596	Ceftriaxone sodium ==> Lidocaine HCl	Ceftriaxone sodium	Lidocaine HCl
33	2	17.21	28.42	0.72	1.65	11179	Ceftriaxone sodium ==> Acetaminophen	Ceftriaxone sodium	Acetaminophen
34	2	4.68	28.22	0.57	6.03	8713.0	Ketorolac tromethamine ==> Diphenhydramine	Ketorolac tromethamine	Diphenhydramine
35	2	17.21	27.91	0.56	1.62	8618.0	Ketorolac tromethamine ==> Ondansetron HCl	Ketorolac tromethamine	Ondansetron HCl
36	2	7.80	27.65	0.55	3.55	8408.0	Midazolam HCl ==> Lidocaine HCl	Midazolam HCl	Lidocaine HCl
37	2	11.02	26.81	1.21	2.43	18684	Local anesthetic combin ==> Sodium chloride	Local anesthetic combin	Sodium chloride
38	2	7.80	26.37	0.40	3.38	6157.0	Fentanyl ==> Lidocaine HCl	Fentanyl	Lidocaine HCl
39	2	7.80	26.03	0.56	3.34	8701.0	Morphine sulfate ==> Lidocaine HCl	Morphine sulfate	Lidocaine HCl
40	2	11.02	25.96	0.66	2.36	10209	Ceftriaxone sodium ==> Sodium chloride	Ceftriaxone sodium	Sodium chloride
41	2	1.97	25.49	0.39	12.93	5952.0	Fentanyl ==> Midazolam HCl	Fentanyl	Midazolam HCl
42	2	11.02	25.43	0.38	2.31	5822.0	Ketamine HCl ==> Sodium chloride	Ketamine HCl	Sodium chloride
43	2	4.31	25.01	2.84	5.81	43769	Albuterol sulfate ==> Prednisolone	Albuterol sulfate	Prednisolone
44	2	8.01	24.10	2.73	3.01	42171	Albuterol sulfate ==> Dexamethasone	Albuterol sulfate	Dexamethasone
45	2	7.80	23.93	0.48	3.07	7391.0	Ketorolac tromethamine ==> Lidocaine HCl	Ketorolac tromethamine	Lidocaine HCl
46	2	17.21	23.23	2.56	1.35	39481	Sodium chloride ==> Ondansetron HCl	Sodium chloride	Ondansetron HCl
47	2	1.48	22.27	0.48	15.00	7446.0	Morphine sulfate ==> Ketamine HCl	Morphine sulfate	Ketamine HCl
48	2	7.80	21.81	2.40	2.80	37060	Sodium chloride ==> Lidocaine HCl	Sodium chloride	Lidocaine HCl
49	2	4.77	21.35	1.71	4.47	26372	Dexamethasone ==> Ipratropium bromide	Dexamethasone	Ipratropium bromide
50	2	11.02	21.04	0.41	1.91	6397.0	Midazolam HCl ==> Sodium chloride	Midazolam HCl	Sodium chloride
51	2	0.60	20.84	0.42	34.54	6435.0	Ketorolac tromethamine ==> Prochlorperazine	Ketorolac tromethamine	Prochlorperazine
52	2	30.78	20.28	0.52	0.66	7975.0	Ceftriaxone sodium ==> Ibuprofen	Ceftriaxone sodium	Ibuprofen
53	2	1.51	19.57	0.39	12.93	5952.0	Midazolam HCl ==> Fentanyl	Midazolam HCl	Fentanyl
54	2	17.21	19.31	1.51	1.12	23216	Lidocaine HCl ==> Ondansetron HCl	Lidocaine HCl	Ondansetron HCl
55	2	17.21	18.88	0.40	1.10	6112.0	Amoxicillin trihydrate ==> Acetaminophen	Amoxicillin trihydrate	Acetaminophen
56	2	30.78	16.78	0.76	0.55	11692	Local anesthetic combin ==> Ibuprofen	Local anesthetic combin	Ibuprofen

Display 4. Table of association rules from LA

In the Results window of the Link Analysis node, select View ➡ Plots ➡ Items Constellation Plot from the main menu, it will display the link graph (Display 5), which represents associations between all drugs and visualizes drug-drug combinations. The plot provides better understanding for clinicians who want to know prescription patterns of certain medical condition. The rules data yield nodes and links data in the plot, each drug is represented by a circle called node and the association is represented by a line called edge or link. The larger the node is, the more frequently the drug had been used. The thicker the link is, the more frequently the two drugs had been used concurrently. In this example, Ibuprofen was the most commonly used drug. We found that there were 1507 two drug combinations. The top 5 most common drug combinations are listed in Table 6. One can see the detail drug combinations in the link data (Display 6).



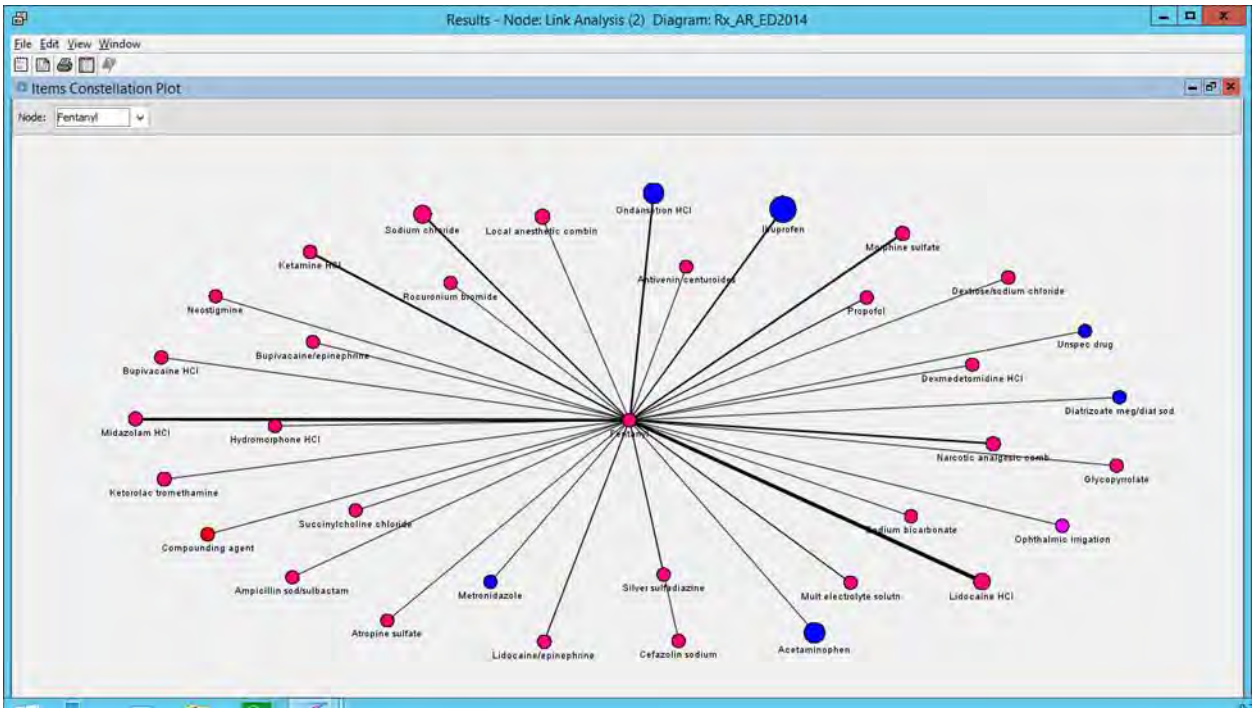
Display 5. Items Constellation Plot.



Display 6. Link Data Screenshot

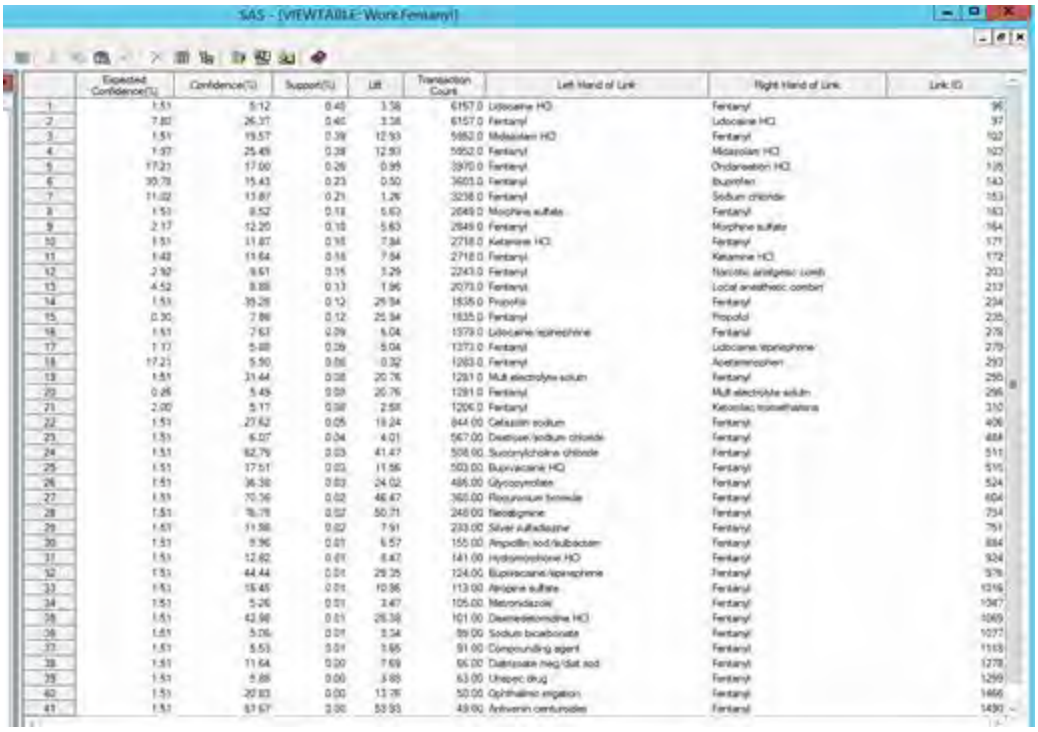
If one is interested in particular node (for example, fentanyl in this case), select the node from the Node list. Any drug of interest can be selected. Once selected, all nodes connected to the selected node you

are displayed, as shown in Display 7. The plot shows fentanyl concurrently used with 34 other drugs respectively.



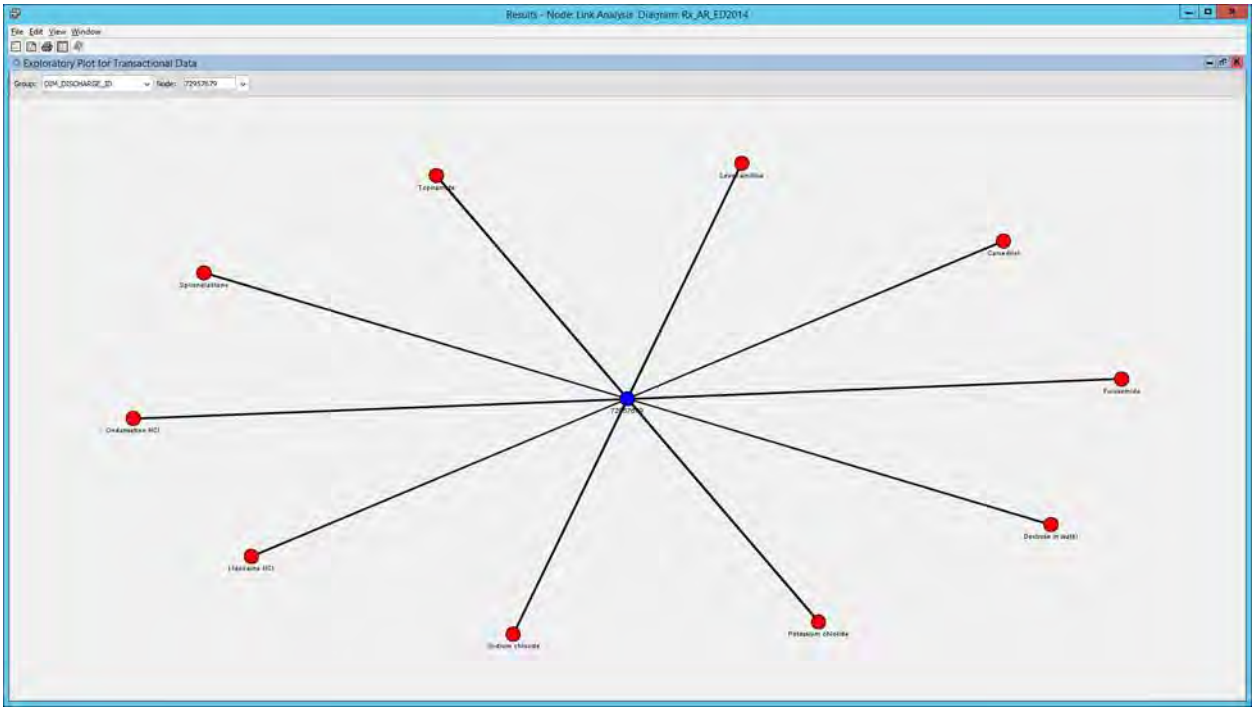
Display 7. Drug Fentanyl Concurrently Used With

We subset the link data with fentanyl involved (Display 8) and identified that three of the drug combinations (Fentanyl + Midazolam HCl, Fentanyl + Morphine sulfate, fentanyl + Hydromorphone HCl) may cause adverse drug-drug interactions and ADEs (additive respiratory depression) using DRUG-REAX® system (Thomson Micromedex®, Truven Health Analytics Inc., Greenwood Village, Co, USA) [5].



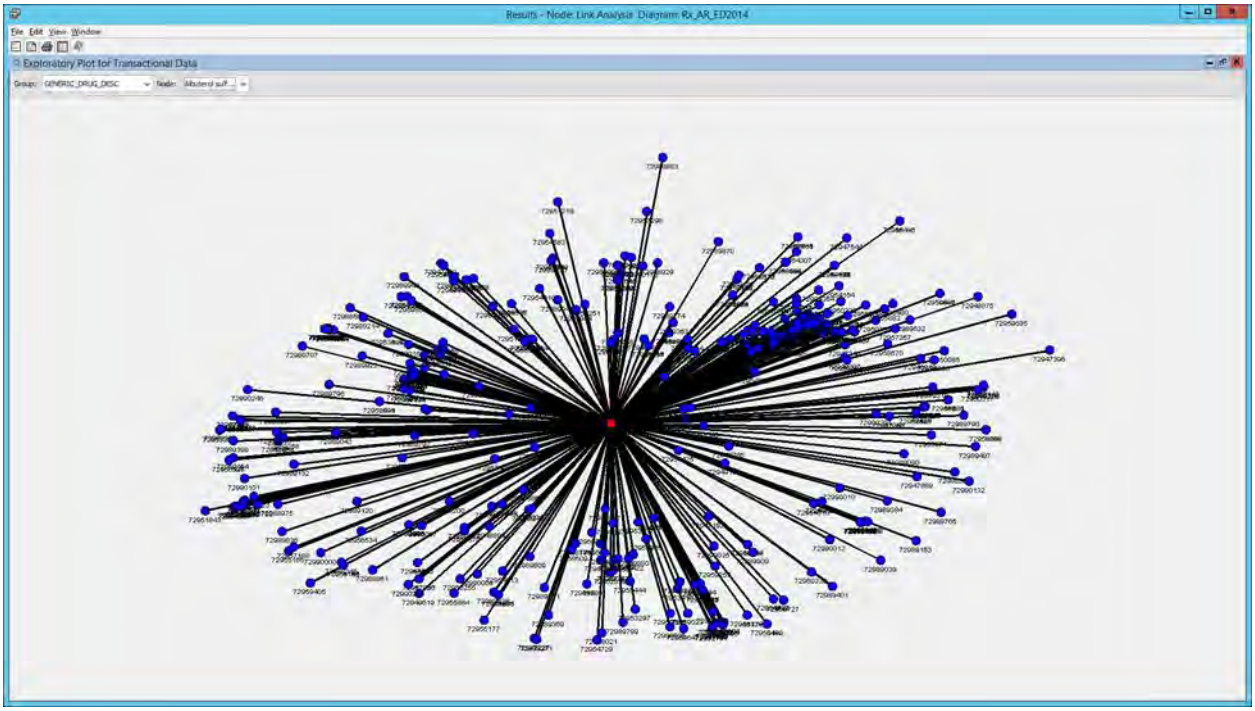
Display 8. Link Data Fentanyl

In the Results window the Link Analysis node, select View ➡ Plots ➡ Exploratory Plot from the main menu, it will display the Exploratory Plot (Display 9, Display 10) for Transactional Data to gain insight about each individual patient's drug utilization, or find each individual drug's users. For example, when one selects DIM_DISCHARGE_ID from the Group list and ID 72957679 from the Node list, one can see the various drugs that the patient took. In this example, the patient had 10 specific drugs during the ED visit (Display 9).



Display 9. Exploratory Plot for Individual Patient Drug Utilization

When one selects GENERIC_GRUG_DESC from the Group list and Albuterol Sulfate from the Node list, one sees all the individual patients who were exposed to the drug during the ED visit (Display10).



Display 10. Exploratory Plot for Individual Drug's Patient List

CONCLUSION

ARM is a powerful tool to analyze health care administrative data, and could provide additional insight into the data and detect important patterns that cannot be identified by traditional statistical methods. LA provides various graphs to visualize data. This paper shows how to use Association Node and Link Analysis Node in SAS Enterprise Miner to analyze and visualize health care administrative data. PHIS data is modified as an example of transactional data. We have shown how AMR and LA can be used to detect drug utilization and polypharmacy (drug-drug combinations) patterns and specific generic polypharmacy among sick children in emergency department visits. This application of ARM techniques has been used to analyze specific medical condition, procedure, or laboratory data associated polypharmacy in inpatient setting too (data not shown in this paper). We believe the same approach could be used in mining other databases such as administrative claims data and electronic medical records.

REFERENCES

1. Kunac DL, Kennedy J, Austin N, Reith D. Incidence, preventability, and impact of Adverse Drug Events (ADEs) and potential ADEs in hospitalized children in New Zealand: a prospective observational cohort study. Paediatr Drugs. 2009;11(2):153-160
2. Lazarou J, Pomeranz BH, Corey PN. Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. JAMA. 1998;279(15):1200–05.
3. Feudtner C, Dai D, Hexem KR, Luan X, Metjian TA.. Prevalence of polypharmacy exposure among hospitalized children in the United States. Arch Pediatr Adolesc Med. 2012;166(1):6–16.
4. Dai D, Feinstein JA, Morrison W, Zuppa AF, Feudtner C. Epidemiology of Polypharmacy and Potential Drug-Drug Interactions Among Pediatric Patients in ICUs of U.S. Children's Hospitals. Pediatr Crit Care Med. 2016;17(5):e218-28.
5. Feinstein JA, Feudtner C, Valuck RJ, Kempe A. The depth, duration, and degree of outpatient pediatric polypharmacy in Colorado fee-for-service Medicaid patients. Pharmacoepidemiol Drug Saf. 2015;24(10):1049-57.
6. Malone DC, Abarca J, Hansten PD, et al. Identification of serious drug-drug interactions: results of the partnership to prevent drug-drug interactions. Am J Pharm Assoc. 2004;44(2):142–51.
7. Harpaz R, DuMouchel W, Shah NH, Madigan D, Ryan P & Friedman C. Novel data-mining methodologies for adverse drug event discovery and analysis. Clin Pharmacol Ther 2012;91(6):1010-21.
8. Chen J, He H, Li J: Representing association classification rules mined from health data. Berlin/Heidelberg: Knowledge-Based Intelligent Information and Engineering Systems 2005;1225–31.
9. Harpaz R, Chase HS, Fridman C. Mining multi-item drug adverse effect associations in spontaneous reporting systems. BMC Bioinformatics 2010;11(Suppl 9):S7
10. Zhu L, Xia A, Jiang Q, Go WY, Walsh L. Mining association rules in clinical trial safety data. JSM 2014 – Biopharmaccutical section.
11. Agarwal R., Imielinski T., Swami AN. 1993. “Mining association rules between sets of items in large databases.” Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, Washington, D.C., United States, May 26-28, pp.207-216.
12. Agarwal R, Mannila H., Srikant R, Toivonen H, Verkamo A. 1996. “Fast discovery of association rules.” In *Advances in Knowledge Discovery and Data Mining*, eds. Fayyad,U., Piatetsky Shapiro G, Smyth P, and Uthurusamy R. MIT Press 1996, pp.307-328.
13. Aggarwal CC. *Social Network Data Analytics*, 2011 ed. New York NY: Springer 2011.
14. Liu Y, Lee T, Zhang R, and Dean J. Link Analysis Using SAS® Enterprise Miner™. https://support.sas.com/rnd/app/data-mining/enterprise-miner/papers/2014/linkAnalysis2014.pdf.

15. Agbenyegah D, Zhang C. Discover the golden paths, unique sequences and marvelous associations out of your big data using Link Analysis in SAS® Enterprise Miner™. https://www.mwsug.org/proceedings/2016/AA/MWSUG-2016-AA04.pdf.
16. Children's Hospital Association. http://www.chca.com/, Accessed December 6, 2017.
17. Neutel I, Skurtveit S, Berg C. Polypharmacy of potentially addictive medication in the older persons – quantifying usage. Pharmacoepidemiol Drug Saf. 2012; 21(2):199–206.
18. Tay KY, Ewald MB, Bourgeois FT. Use of QT-prolonging medications in US emergency departments, 1995–2009. Pharmacoepidemiol Drug Saf. 2013; 23(1):9–17.

ACKNOWLEDGMENTS

The authors thanks Arlen Enoch for helping to install and configure SAS® Enterprise Miner™14.2.

CONTACT INFORMATION

Comments and questions are valued and encouraged. Contact the lead author at:

Dingwei Dai, Ph.D.
Children’s Hospital of Philadelphia
Roberts Center for Pediatric Research
2716 South Street, 11th Floor, 11233
Philadelphia, PA 19146-2305
Email: daid@email.chop.edu