

Original Paper

Medication Use for Childhood Pneumonia at a Children's Hospital in Shanghai, China: Analysis of Pattern Mining Algorithms

Chunlei Tang^{1*}, PhD; Huajun Sun^{2*}, PhD; Yun Xiong^{3*}, PhD; Jiahong Yang⁴, MPH; Christopher Vitale⁵, PharmD; Lu Ruan³, MSc; Angela Ai¹, BA; Guangjun Yu², PhD; Jing Ma⁶, MD, PhD; David Bates¹, MD, MSc

¹Division of General Internal Medicine and Primary Care, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States

²Children's Hospital of Shanghai, Shanghai Jiaotong University School of Medicine, Shanghai, China

³Shanghai Key Laboratory of Data Science, School of Computer Science, Fudan University, Shanghai, China

⁴Shanghai Shengkang Hospital Development Center, Shanghai, China

⁵Clinical Informatics for the Integrated Health Model Initiative, American Medical Association, Chicago, IA, United States

⁶Department of Population Medicine, Harvard Medical School, Boston, MA, United States

*these authors contributed equally

Corresponding Author:

Guangjun Yu, PhD

Children's Hospital of Shanghai

Shanghai Jiaotong University School of Medicine

No 24, Lane 1400 West Beijing Road,

Shanghai, 200040

China

Phone: 86 18917762998

Email: gjyu@shchildren.com.cn

Abstract

Background: Pattern mining utilizes multiple algorithms to explore objective and sometimes unexpected patterns in real-world data. This technique could be applied to electronic medical record data mining; however, it first requires a careful clinical assessment and validation.

Objective: The aim of this study was to examine the use of pattern mining techniques on a large clinical dataset to detect treatment and medication use patterns for childhood pneumonia.

Methods: We applied 3 pattern mining algorithms to 680,138 medication administration records from 30,512 childhood inpatients with diagnosis of pneumonia during a 6-year period at a children's hospital in China. Patients' ages ranged from 0 to 17 years, where 37.53% (11,453/30,512) were 0 to 3 months old, 86.55% (26,408/30,512) were under 5 years, 60.37% (18,419/30,512) were male, and 60.10% (18,338/30,512) had a hospital stay of 9 to 15 days. We used the FP-Growth, PrefixSpan, and USpan pattern mining algorithms. The first 2 are more traditional methods of pattern mining and mine a complete set of frequent medication use patterns. PrefixSpan also incorporates an administration sequence. The newer USpan method considers *medication utility*, defined by the dose, frequency, and timing of use of the 652 individual medications in the dataset. Together, these 3 methods identified the top 10 patterns from 6 age groups, forming a total of 180 distinct medication combinations. These medications encompassed the top 40 (73.66%, 500,982/680,138) most frequently used medications. These patterns were then evaluated by subject matter experts to summarize 5 medication use and 2 treatment patterns.

Results: We identified 5 medication use patterns: (1) antiasthmatics and expectorants and corticosteroids, (2) antibiotics and (antiasthmatics or expectorants or corticosteroids), (3) third-generation cephalosporin antibiotics with (or followed by) traditional antibiotics, (4) antibiotics and (medications for enteritis or skin diseases), and (5) (antiasthmatics or expectorants or corticosteroids) and (medications for enteritis or skin diseases). We also identified 2 frequent treatment patterns: (1) 42.89% (291,701/680,138) of specific medication administration records were of intravenous therapy with antibiotics, diluents, and nutritional supplements and (2) 11.53% (78,390/680,138) were of various combinations of inhalation of antiasthmatics, expectorants, or corticosteroids. Fleiss kappa for the subject experts' evaluation was 0.693, indicating moderate agreement.

Conclusions: Utilizing a pattern mining approach, we summarized 5 medication use patterns and 2 treatment patterns. These warrant further investigation.

KEYWORDS

drug therapy; combination; computer-assisted; pattern recognition; data mining; precision medicine; childhood pneumonia; hospital

Introduction

Childhood pneumonia remains the single largest cause of death in young children worldwide [1,2]. According to a recent World Health Organization (WHO) report, an estimated 922,000 children under the age of 5 passed away because of pneumonia in 2015 alone, accounting for 16% of all deaths in this age group [2,3]. In China, it is an especially grave concern, driven in part by environmental pollution [3] and abuse of antibiotics [4]. For virtually all high-mortality settings, the WHO's case management plan uses algorithms as the basis for pneumonia management [5]. Globally, there are many clinical recommendations for treating pediatric pneumonia, but some guidelines are outdated and can be vague [6]. Variation in treatment regimens within the same clinic could indicate poor clinical practice leading to increased recurrence rates that are more likely to have complications [7,8]. Multiple medication therapies are common practice and may provide more effective treatment with a lower concentration of individual components lessening the risk of side effects and toxicity [9,10]. What is missing, however, is the knowledge derived directly from real-world clinical practice.

Electronic medical record (EMR) data represent a rich source of information that includes 2 types of clinical knowledge [8,11,12]. First, it reflects the medical knowledge— and specialty-based clinical practice of a group of doctors within a certain time period. This encompasses medical solutions originating from various clinical standards developed by physicians as a professional group. Second, it represents the clinical experience of an individual practicing clinician and their prescribing style, as well as external factors such as pricing and regulation. Meaningful use of these data can yield information that can guide clinical practice. Challenges of using EMR data, however, have included the large data volume and the substantial variations between different EMR systems.

Data mining techniques have great potential for use in exploring EMR data. Pattern mining is an important subfield of data mining. Its goal is to find all possible salient and persistent patterns in a dataset [13,14], including, but not limited to, direct or indirect associations, trends, periodic patterns, sequential rules, and high-utility patterns of real-life events. Pattern mining is an overloaded term that is used to refer to different technologies in different domains. These techniques have been used in economics, for example, to analyze the stock market and supermarket sales, but have rarely been used in medical research and, specifically, aiming to a variable length of hospital stay is the first time. In one context, it means statistical properties for handling continuous attributes, and in another context, it refers to an ordinal relation, usually based on temporal or spatial precedence that exists among events occurring in the data. Furthermore, the tools and evaluation methods used in each context are different. We aimed to consider the sequential

information that may be valuable for identifying recurring features of a dynamic system or predicting future occurrences of certain events. In addition, most of the previous research on discovering medication-using patterns involves text-based methods and is limited to one or two medications [15,16]. In this study, we explored the utility of a pattern mining approach to the entire record of inpatient EMR-based medications administered for childhood pneumonia in a large children's hospital in China across several years.

Methods

Study Population

The Children's Hospital of Shanghai (CHS), one of the top comprehensive children's hospitals in China, admits approximately 5000 inpatients annually from different parts of the country and internationally. Formerly known as Underprivileged Children's Hospital, it is one of the oldest children's hospitals in Asia.

We extracted 680,138 inpatient medication administration records from 30,512 childhood patients with an initial diagnosis of pneumonia, from January 1, 2010, to December 31, 2015, at the CHS, China. The EMR dataset contains 18,419 males and 12,093 females, among which 60.10% (18,338/30,512) had a hospital stay of 9 to 15 days. If rehospitalized, the patients had 1 record for each hospitalization. Raw data were deidentified. We pulled age and sex, initial diagnosis, and admission and discharge dates, as well as routes and doses of administered medications. According to Chinese legislation, ethical approval from the regional ethical review board is not needed for this type of study of deidentified EMR data. The data used in this study were anonymous. Although ethical approval from the regional ethical review board is not needed for this type of study of deidentified EMRs according to Chinese legislation, we still applied for and received approval from the Institutional Review Board of the CHS.

Data Preparation

We first merged similar routes of administration for the same medication, and then converted commonly used medications to their shortened reference names as shown in [Multimedia Appendix 1](#). We then removed diluents (or carriers) from the dataset after our expert panel determined that these were not intended to have a therapeutic effect. The final dataset contains 652 individual medications (including traditional Chinese patent medicine [17]), among which the top 40 medications appeared in 73.66% (500,982/680,138) of all the medication administration records ([Multimedia Appendix 1](#)).

Selected Pattern Mining Algorithms: FP-Growth, PrefixSpan, and USpan

Our initial goal was to generate sets of frequently appearing items (ie, frequent item sets) to better understand *medication*

use patterns. However, doing so would involve excluding too many temporal data points. An example of this temporal information is the time at which an item (eg, medication) was prescribed by a physician. Thus, while the classic frequency-based framework often leads to many patterns being identified, most of them are not informative enough for further clinical investigation. Recent efforts have been made to incorporate utility into the pattern selection framework, so that high-utility patterns, regardless of frequency, are mined. To obtain more reliable and relevant results, we first developed a unified platform to draw a parallel comparison among all medication use patterns produced by 3 pattern mining algorithms (FP-Growth, PrefixSpan, and USpan). We then asked the panel of subject matter experts to review these results based on their clinical expertise and to make recommendations for meaningful grouping. We also compared these results with what experts found using summary statistics in our previous work [18].

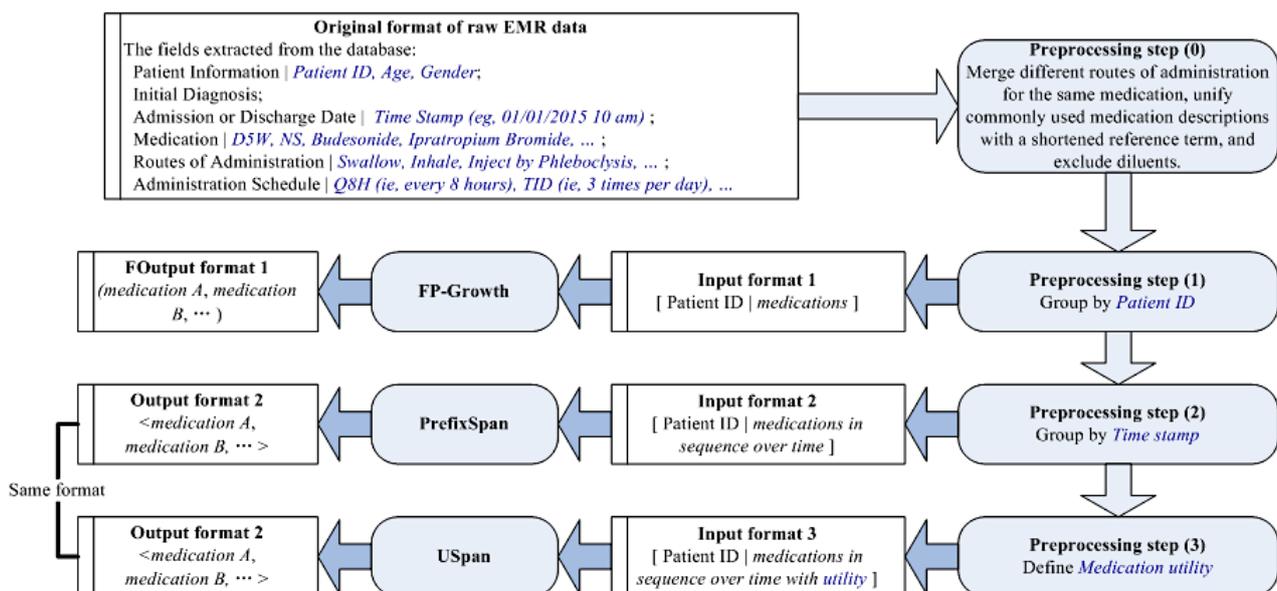
The advantage of progressively using 3 algorithms instead of applying only 1 is twofold: (1) the 3 methods allow for more comprehensive data mining that includes frequency, timing, and utility of dosing of medication administration and (2) it can reduce approach-specific limitations. FP-Growth and PrefixSpan are 2 classic pattern mining algorithms, which were commonly used in business. The earliest and most classic example of their use was examining the cosale of diapers and beer [19]. FP-Growth allows mining of a complete set of frequent patterns by pattern fragment growth without using candidate generations [20]. PrefixSpan offers ordered growth and reduced projected databases over a particular interval [21]. However, the 2 algorithms can only find patterns that appear with high frequency. For example, if there is a medication use pattern of

(a, b) where one of the medications (ie, a or b) was administered with low frequency, the 2 algorithms would not be able to discover this combination. It is possible that some interesting patterns (eg, a medication prescribed at a low frequency but at a high dose) may be filtered out of the results because of the medication’s low frequency of appearance. To address this problem, we used the more advanced USpan algorithm (first proposed by Yin et al [22]) and proposed a new definition called *medication utility*. The term *utility* originated from economics and considers quantities, profits, and time orders of items simultaneously [23]. The method is unique in its business consideration in dollar value for customers in financial markets. We define *medication utility* using both the dosage and frequency of medication administration over time to obtain more detailed evidence of medication use. USpan reflects a recent effort to incorporate utility into the pattern selection framework, so that high-utility (frequent or infrequent) patterns are mined which address some of the concerns involved in exploratory factor analysis, such as dollar value associated with each pattern. To achieve this, we implemented the USpan algorithm (see [Multimedia Appendix 2](#) with its explanation).

Application of Pattern Mining Algorithms to the Electronic Medical Record Data

We summarized our approach in [Figure 1](#), including (1) the original format of the raw EMR data, (2) a diagram outlining the 3-step data processing method, and (3) the steps taken by a particular algorithm, and its input and output formats. The thresholds used in the algorithms are as follows: the minimum support of both FP-Growth and PrefixSpan is 0.15, and the minimum utility of USpan is 30,000. We only utilized the top 10 results from our algorithmic outputs.

Figure 1. Application of algorithms to the dataset (medication A is a placeholder for a drug name and a indicates the frequency of appearance of medication A). EMR: electronic medical record; D5W: Dextrose 5% in Water; NS: normal saline; ID: identification.



Step 1: Group by Patient Identification

As FP-Growth disregards the quantity of items and time information, we first simplified the raw EMR data to a list of all the administered medications and their corresponding patients. FP-Growth outputs an unordered pattern represented by several medications enclosed in parentheses, for example, (*medication A*, *medication B*, and *medication C*). For example, consider 2 patients, one who used *medication A* on January 1 and *medication B* on January 2, and another who used *medication B* on January 10 and *medication A* on January 21. In this scenario, the output of FP-Growth would be (*medication A* and *medication B*), meaning that *medication A* and *medication B* were administered during the same hospitalization.

Step 2: Group by Time Stamp

To make a suitable input for PrefixSpan that includes temporality, we took the results from Step 1 and determined the sequence of medication administration over time. PrefixSpan outputs a temporal order enclosed in a pair of angled brackets. For example, the combination $\langle (\textit{medication A}, \textit{medication B}), \textit{medication C}, \dots \rangle$ means that there exists a temporal relation between (*medication A*, *medication B*) and *medication C*. Put another way, *medication A* and *medication B* were administered at the same time, and both were administered before *medication C*. For example, if a patient uses *medication A* for the first day, *medication B* for the first day, *medication D* for the fifth day, *medication A* for the eighth day, and *medication B* for the ninth day, while FP-Growth would output the grouping (*medication A*, *medication B*, *medication D*) without considering the repeat medications, PrefixSpan's output would be of the form $\langle (\textit{medication A}, \textit{medication B}), \textit{medication D}, \textit{medication A}, \textit{medication B} \rangle$.

Step 3: Define Medication Utility

We defined *medication utility* as follows. Let $I = \{i_1, i_2, \dots, i_n\}$ be the universal set of distinct medications. Each item $i_k \in I$ ($1 \leq k \leq n$) is associated with a utility value, denoted as $P(i_k)$, which shows the utility of dosing of each specific medication to treat a given disease. A medication usage is represented as an ordered pair (i_k, q_k) , where $i_k \in I$ is a medication and q_k is a positive number representing the quantity of I , which shows how much of this medication is taken (ie, dose). We thus define *medication utility* as $U(i_k, q_k) = P(i_k) \times q_k$ representing a single medication's dosage during one patient's hospitalization.

Since USpan takes both the time sequence and medication utility of each item into account, every administration record that occurred at the same time will be enclosed in a pair of parentheses, such as (*medication A*, a), in which *medication A* is a single drug and a is the frequency of *medication A*'s appearance. Each medication is attached to its utility, and each medication-utility pair is separated by time, indicated by a comma. USpan results are in the same output format as PrefixSpan.

Experts' Reviewing of the Results

To determine which machine-identified patterns most accurately reflect real-world clinical practice [24], we invited a panel of

subject matter experts including 3 physicians (MD), 3 pharmacists, and 2 researchers (1 DS and 1 MPH). The panelists were invited to review, adjusting process inputs and outputs if needed, and offer opinions on the clinical validity of the patterns obtained from the 3 algorithms. We then conducted a thorough literature review (by putting the drug names seen in a specific pattern into Google's search engine to find related papers) to further analyze the combinations that our panel considered clinically interesting. Our experts then conducted a final review of the results of the literature search to further validate the medication use patterns. We calculated Fleiss kappa to measure interrater reliability for the experts shown in [Multimedia Appendix 3](#).

Results

Study Population

The majority (86.55%, 26,408/30,512) of the study population ([Table 1](#)) was under 5 years of age. Among the under-5 population, 66.55% (20,305/30,512) were under 2 years of age and 37.53% (11,453/30,512) were newborns. There were more male (60.37%, 18,419/30,512) than female patients.

Simple Exploration of Frequency of Use of the Raw Electronic Medical Record Data

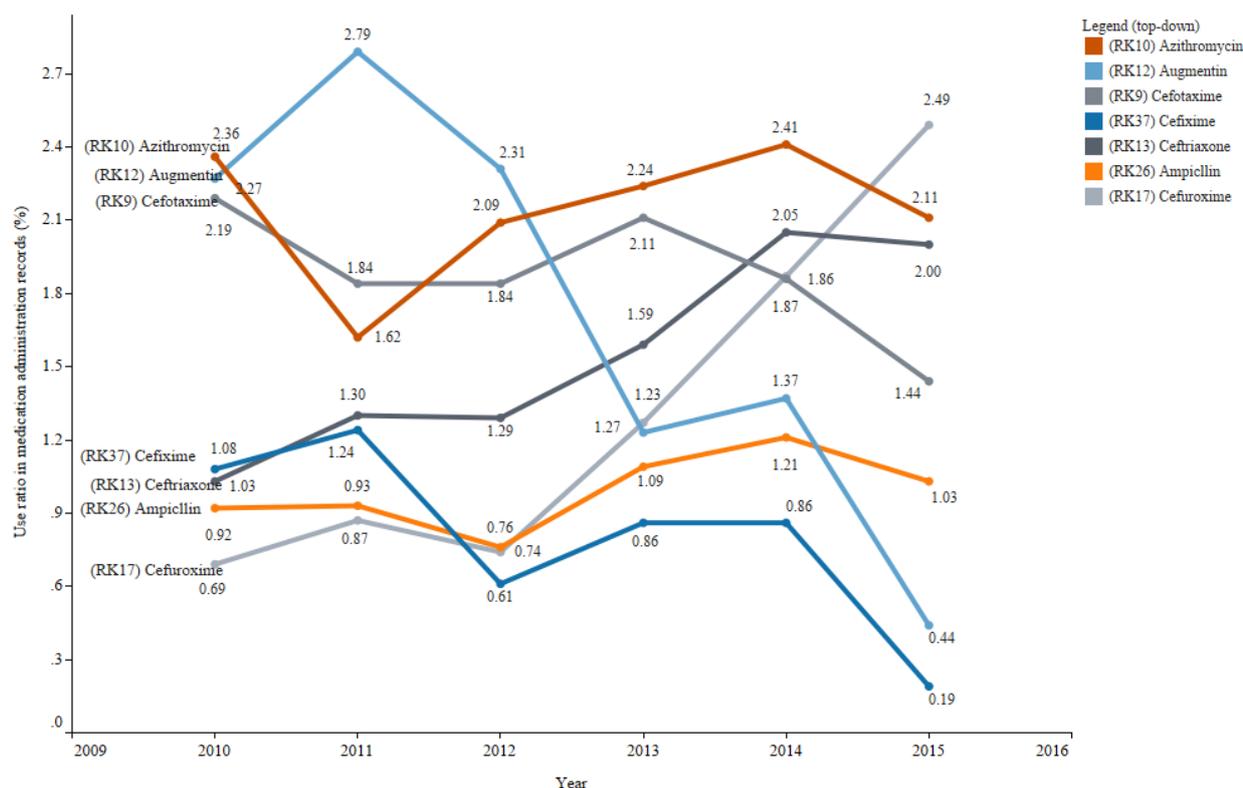
To better understand the raw EMR data, we first examined the top 40 most frequently prescribed medications across age groups and found that 2 diluents (D5W, Dextrose 5% in Water and NS, Normal Saline) were ranked the first- and second-most frequent ([Multimedia Appendix 4](#)). After removing all diluents from the dataset, some nutritional supplements, such as fat emulsion and vitamins, were ranked within top 10 medications for each of the 6 age groups ([Multimedia Appendix 5](#)). Besides diluents and nutritional supplements, antibiotics were also among the top 10 most frequently prescribed in each age group. These medications were present in 42.89% (291,701/680,138) of all specific medication administration records and they were all administered through intravenous (IV) therapy. Thus, *treatment pattern 1* is the combination of antibiotics, diluents, and nutritional supplements via IV therapy.

Due to the high frequency of antibiotic use, we then used statistics on individual medications to examine the frequency of use of the 7 commonly administered antibiotic monotherapies over the study period (2010 to 2015) and found a dramatically increased use of ceftriaxone (a third-generation cephalosporin antibiotic) and cefuroxime (a second-generation cephalosporin) over this period of time ([Figure 2](#)). Meanwhile, the use of a more traditional first-generation antibiotic—augmentin—had dramatically declined. In addition, we found that the administration of certain types of antibiotics varied with patients' age ([Figure 3](#)). For instance, cefotaxime (a third-generation cephalosporin) was more commonly used in newborns under 3 months of age, but azithromycin was rarely administered in that age group. This observation might reflect the results of some published studies which reported an increased risk of cardiovascular death associated with the use of azithromycin, specifically, in infants [25,26].

Table 1. Study population demographics.

Characteristics	Year					
	2010	2011	2012	2013	2014	2015
Total population (N)	4216	4062	3935	5080	6198	7021
Sex, n (%)						
Male	2602 (61.72)	2447 (60.24)	2411 (61.27)	3030 (59.65)	3718 (59.99)	4211 (59.98)
Female	1614 (38.28)	1615 (39.76)	1524 (38.73)	2050 (40.35)	2480 (40.01)	2810 (40.02)
Age groups^a, n (%)						
0-3 months	1678 (39.80)	1617 (39.81)	1609 (40.89)	2075 (40.85)	2297 (37.06)	2177 (31.01)
3-6 months	356 (8.44)	413 (10.17)	360 (9.15)	351 (6.91)	410 (6.62)	549 (7.82)
6-12 months	441 (10.46)	491 (12.09)	386 (9.81)	482 (9.49)	580 (9.36)	716 (10.20)
1-2 years	426 (10.10)	421 (10.36)	421 (10.70)	504 (9.92)	723 (11.67)	822 (11.71)
2-5 years	781 (18.52)	725 (17.85)	711 (18.07)	902 (17.76)	1279 (20.64)	1705 (24.28)
>5 years	534 (12.67)	395 (9.72)	448 (11.39)	766 (15.08)	909 (14.67)	1052 (14.98)

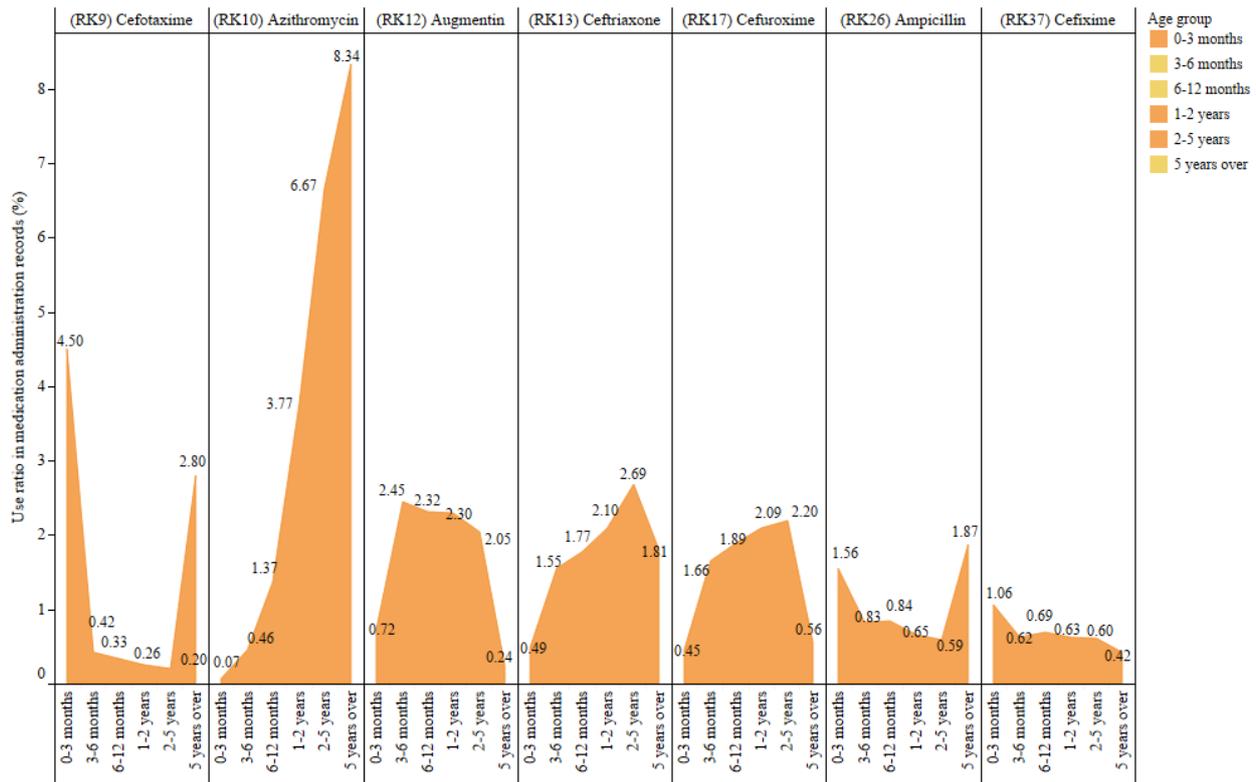
^aAge indicates the age at admission and was calculated as the difference between *Admission Date* and *Birthday*.

Figure 2. The historic proportion of antibiotics administered according to the calendar year.

Using Pattern Mining to Explore the Electronic Medical Record Dataset

We utilized the 3 algorithms to produce the top 10 medication use patterns from 6 age groups spanning from 0 to 17 years, resulting in 180 distinct medication combinations (Multimedia Appendix 1). On the basis of these results, our expert panel summarized 5 clinically interesting medication use patterns as shown in Table 2. Checkmarks indicate that the pattern appeared in that age group. *Pattern 1* is antiasthmatics (albuterol) and expectorants (ipratropium bromide) and corticosteroids

(budesonide). *Pattern 2* is antibiotics and (antiasthmatics or expectorants or corticosteroids). *Pattern 3* is the only use of third-generation antibiotics or the use of these medications followed by traditional antibiotics. *Pattern 4* is antibiotics and medications for enteritis (probiotics: bifid triple viable, smectite, and clostridium butyricum) or medication for skin diseases (antiseptics: zinc oxide and drapolene). Finally, *Pattern 5* is (antiasthmatics or expectorants or corticosteroids) and (medications for enteritis or skin diseases). Detailed descriptions of the patterns for 3 major groups are provided in the following sections.

Figure 3. The historic proportion of antibiotics administered according to patient age group.

Inhaled Medications

Medication use patterns 1 and 2 were both revealed by USpan, because it takes into account a high medication utility for administered medications ([Multimedia Appendix 5](#)). For pattern 1, we found that all 3 medications reflect nonsequential medication administrations, indicating that they were administered concurrently instead of the more traditional sequential application of each medication. Our expert panel confirmed that this medication pattern could shorten medication administration durations, for example by combining and administering multiple inhaled medications simultaneously via nebulization. This approach was also confirmed by the chemical stability of such a mixture [27,28]. Further investigation of the route of administration of pattern 1 medication use indicated that inhalation therapy accounted for 11.53% (78390/680138) of the medication administration records, which we considered as *Treatment pattern 2*. Medication pattern 2 was these inhaled medications plus antibiotics indicating the management pattern of difficult-to-treat infections in patients with pneumonia.

Antibiotics

FP-Growth and PrefixSpan, the 2 classical algorithms, showed a similar pattern 3 ([Table 2](#)), that is, the only use of third-generation antibiotics or the use of these medications with (or followed by) traditional antibiotics. FP-Growth resulted in

2 combinations. One is the combination of 2 third-generation cephalosporins (cefotaxime and cefixime). The other is the combination of a third-generation cephalosporin (ceftriaxone) with a traditional first-generation antibiotic (azithromycin). Interestingly, PrefixSpan showed not only the same 2 medication combinations but also similar timings: the average time gap between the 2 medications was 7 and 5 days, respectively.

Medications for Enteritis or Skin Diseases

Patterns 4 and 5 showed a correlation between pneumonia-specific medications and medications for 2 other conditions: enteritis and skin diseases. For example, pattern 4 demonstrated a third-generation antibiotic (cefotaxime) and a medication for skin rashes (zinc oxide) were followed by a probiotic (bifid triple viable) that appears only in the *0 to 3 months* age group. Pattern 5 showed that the 3 inhalation medications (albuterol, ipratropium bromide, and budesonide) were followed by probiotics without using antibiotics that mainly appeared in the *3 to 6 months* age group. Due to the fact that PrefixSpan only determines the order of different medications without considering a time window, we further examined the time interval between medications for pneumonia following medication for enteritis or skin diseases ([Multimedia Appendix 4](#)). The findings revealed that the average time intervals were similar (4 vs 6 days) in the abovementioned 2 examples for patterns 4 and 5.

Table 2. A select list of clinically interesting results produced by 3 pattern mining algorithms.

#	Medication use patterns	FP-Growth	PrefixSpan	USpan	Age groups					
					0-3 m	3-6 m	6-12 m	1-2 y	2-5 y	>5 y
1	anti-asthmatics AND expectorants AND corticosteroids	(Albuterol, Ipratropium Bromide, Budesonide)	(Albuterol, Ipratropium Bromide, Budesonide)	(Albuterol, Ipratropium Bromide, Budesonide), (Pholcodine)	✓ ^a	✓	✓	✓	✓	✓
2	antibiotics AND (anti-asthmatics OR expectorants OR corticosteroid)	— ^b	—	(Albuterol, Ipratropium Bromide, Budesonide), (Azithromycin)			✓	✓	✓	✓
		—	—	(Albuterol, Ipratropium Bromide, Augmentin, Budesonide), (Azithromycin)						✓
3	third-generation cephalosporin antibiotics with (or followed by) traditional antibiotics	(Cefotaxime, Cefixime)	(Cefotaxime), (Cefixime)	—	✓					
		(Ceftriaxone, Azithromycin)	(Ceftriaxone), (Azithromycin)	—						✓
4	anti-biotics AND (medications for enteritis OR skin diseases)	(Bifid Triple Viable, Cefotaxime, Zinc Oxide)	(Cefotaxime, Zinc Oxide), (Bifid Triple Viable)	—	✓					
		(Bifid Triple Viable, Cefotaxime, Drapolene)	(Bifid Triple Viable, Cefotaxime)	—	✓					
		(Cefotaxime, Zinc Oxide, Drapolene)	(Cefotaxime, Zinc Oxide, Drapolene)	—	✓					
		(Smecitite, Bifid Triple Viable, Cefotaxime)	—	—	✓					
5	(antiasthmatics OR expectorants OR corticosteroids) AND (medications for enteritis OR skin diseases)	(Bifid Triple Viable, Albuterol, Ipratropium Bromide, Budesonide)	(Albuterol, Ipratropium Bromide, Budesonide), (Bifid Triple Viable)	—		✓				

^a✓ indicates that the pattern appeared within the specified age groups.

^bNo result from the specific algorithm.

Discussion

Principal Findings

There is a considerable demand for appropriate and proven approaches to expanding the use of medication analytics. By simply checking the frequency of use of medications in the raw EMR data, we found that IV administration is the most common administration route (treatment pattern 1) for treating inpatient childhood pneumonia in this hospital. This finding is in line with concerns about heavy use of IV therapy in China reported by the WHO [29]. Even then, statistical analysis is limited in its ability to discover relationships among medications, and it typically only produces a list of monotherapies with frequency rankings [30]. Indeed, one could argue that from the perspective of treating a given disease, a medication administered with higher frequency implies higher efficacy. However, our results illustrate that this is not always the case.

By utilizing 3 pattern mining algorithms (FP-Growth, PrefixSpan, and USpan), which are commonly used for business applications, we not only found 5 medication use patterns but also identified an additional treatment pattern (treatment pattern 2), that is, medication administration via inhalation route.

There were differing opinions within our expert panel as to what pattern 4 (antibiotics and medications for enteritis or skin diseases) indicates. One possible explanation, offered by 4 US experts, is that because of the immaturity of infant organs, damage to one organ or organ system could cause reactions in other organs [31]. Furthermore, from a pharmaceutical perspective, broad-spectrum antibiotics (eg, cefotaxime) cause diarrhea by irradiating normal gastrointestinal (GI) flora [32], and physicians can treat this type of diarrhea in the pediatric population with bifid triple viable, a probiotic used to re-establish normal GI flora. Additionally, diarrhea can cause diaper rash that can be treated with zinc oxide. Chinese experts believed that these findings support an underlying internal

relationship between the lungs and large intestine (or the lungs and skin), which is consistent with established theories of traditional Chinese medicine (TCM) [33]. TCM believes that the human body is an organic whole, that there is a relationship between the lung and large intestine, and that the lungs govern the skin and hair [34]. Despite this rationale, our expert panel generally agreed that pattern 5 was unexpected (ie, the use of medications for enteritis or skin diseases following inhalation treatment, without concurrent use of antibiotics). The overall Fleiss kappa for these 7 patterns was 0.693 (Table 2) indicating moderate agreement [35]. Our findings neither contradict nor confirm these hypothesized relationships as our data target treatment protocols which cannot, on their own, speak to the relationship between organ systems. However, according to a recent study on innate lymphoid cells, lung inflammation might originate in the gut [36], and there exists a link between intestinal microbiota and lung diseases (eg, asthma) [37].

Among the strengths of this study is its novel application of the 3 pattern mining algorithms to medical data. We determined 5 medication use patterns that can succinctly reflect prescriber style among complex real-world hospital EMR use. This prescriber style reflects decision factors beyond efficacy (eg, availability and cost of a medication, frequency of administration in a busy setting, and hospital formulary). Although patients with an initial EMR diagnosis of *pneumonia* could include all varieties of pneumonia to which various treatment protocols may apply, most of our results correlated with our literature searches regarding the medications' use. For example, a combination of 2 antibiotics produced by FP-Growth (cefuroxime and azithromycin) was evaluated by Vergis et al [38], who found no increased risk of mortality associated with prescribing the 2 antibiotics continuously or simultaneously. Another example is by Rubio et al [39], which evaluated the sequential combination of cefotaxime and cefixime, finding that prescribing them within 2 to 3 days of each other may result in shorter hospital stays, a pattern that was also found by PrefixSpan. Results that have not been mentioned in the current literature (eg, pattern 3) warrant further investigation.

One limitation of this study is that we focused on one source of data from a single pediatric hospital in Shanghai, China. Although other pediatric hospitals in China might not have unique medical practices, this Shanghai-based pediatric hospital has been the leading medical institution with a strong link to the international pediatric community; it contains up-to-date technology, and its clinicians undergo continuous medical training. Our findings provide firsthand insight into understanding Chinese pediatricians' experiences in the treatment of pneumonia. Although our results may not be generalizable across all demographics, especially in remote areas of China, it is likely that most medication use patterns are universal in the more populated areas of China.

Our work is the first step toward better synthesis of current practice and establishing more realistic treatment protocols for childhood pneumonia. Moving forward, we plan to expand this study within and across childhood pneumonia groups from multiple health care organizations. We also envision that a knowledge base of medication use patterns will serve as an informative guide to researchers and clinicians. Our discovered treatment patterns have the potential for inclusion in treatment protocols. We hope to further integrate clinical and nonclinical information into our algorithms to help determine cost and efficacy of treatments, as well as readmission, mortality, and morbidity rates.

Conclusions

We used a pattern mining approach to automatically acquire knowledge of prior medication treatment combinations for childhood pneumonia. An expert panel summarized 5 medication use patterns. These, together with 2 identified treatment patterns that also targeted skin disease and enteritis, may warrant further investigation. Additionally, our findings suggest the following starting points for further discussion: (1) a comparison of IV therapy before and after the publication of China's new deal on the rational use of medicines, the 2013 *Principle of Rational Use of Medicines* [40], (2) validation and comparison of efficacy of various medication use patterns, and (3) a potential relationship between the lungs and skin.

Acknowledgments

The authors would like to thank the nonauthors in our expert panel: Datian Che, MD, Diane L. Seger, RPh, and Changzheng Yuan, DS, for their time and advice. Also, we would like to thank Siyuan Cheng, MSc, Hai Cao, MSc, Suzanne V. Blackley, MA, and Joseph M Plasek, MS, for helping us improve USpan and revise the manuscript. The work was partially funded by the National Natural Science Foundation of China Project #71473164, 71874110, U1636207 and 91546105, the Shanghai Science and Technology Development Fund #16JC1400801, 17511105502, 17511101702, and the Suzhou Science and Technology Bureau Technology Demonstration Project (SS2017 12, SS201812).

Authors' Contributions

All authors provided substantial contribution to the conception and design of this work, its data analysis and interpretation, and helped draft and revise the manuscript. All the authors are accountable for the integrity of this work. CT, HS, LR, YX, and GY conceived and designed the experiments. CT, HS, VC, LR, YX, and GY analyzed the data. CT, LR, and YX performed the experiments. CT, VC, LR, AA, YX, GY, JM, and DWB contributed reagents, materials, or analysis tools. CT, JY, VC, AA, LR, YX, GY, JM, and DWB wrote and revised the paper. CT, SH, and YX contributed equally to this work. JM and DB are joint senior authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1

The implementation of the USpan algorithm.

[\[PDF File \(Adobe PDF File\), 66KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

Related commonly administered medications with their shortened reference names.

[\[PDF File \(Adobe PDF File\), 98KB-Multimedia Appendix 2\]](#)

Multimedia Appendix 3

Values for Fleiss kappa (N=7, n=8, k=2) as substantial agreement.

[\[PDF File \(Adobe PDF File\), 99KB-Multimedia Appendix 3\]](#)

Multimedia Appendix 4

Appearance of “medications for enteritis or skin diseases” in pneumonia treatment course within patients’ medication administration records.

[\[PDF File \(Adobe PDF File\), 66KB-Multimedia Appendix 4\]](#)

Multimedia Appendix 5

The top 10 results distinguished by age groups.

[\[PDF File \(Adobe PDF File\), 55KB-Multimedia Appendix 5\]](#)

References

1. World Health Organization. 2018 May 24. The top 10 causes of death URL:<https://www.who.int/en/news-room/fact-sheets/detail/the-top-10-causes-of-death> [accessed 2019-01-21] [[WebCite Cache ID 75ZumyoMe](#)]
2. World Health Organization. 2016 Nov 6. Pneumonia URL:<http://www.who.int/mediacentre/factsheets/fs331/en> [[WebCite Cache ID 73rKrnS8q](#)]
3. Guan X, Silk BJ, Li W, Fleischauer AT, Xing X, Jiang X, et al. Pneumonia incidence and mortality in Mainland China: systematic review of Chinese and English literature, 1985-2008. PLoS One 2010 Jul 23;5(7):e11721 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0011721](https://doi.org/10.1371/journal.pone.0011721)] [Medline: [20668535](#)]
4. Currie J, Lin W, Meng J. Addressing antibiotic abuse in China: an experimental audit study. J Dev Econ 2014 Sep 1;110:39-51 [[FREE Full text](#)] [doi: [10.1016/j.jdeveco.2014.05.006](https://doi.org/10.1016/j.jdeveco.2014.05.006)] [Medline: [26949284](#)]
5. Mulholland K. Problems with the WHO guidelines for management of childhood pneumonia. Lancet Glob Health 2018 Dec;6(1):e8-e9 [[FREE Full text](#)] [doi: [10.1016/S2214-109X\(17\)30468-0](https://doi.org/10.1016/S2214-109X(17)30468-0)] [Medline: [29241619](#)]
6. Pritchard JR, Lauffenburger DA, Hemann MT. Understanding resistance to combination chemotherapy. Drug Resist Updat 2012 Oct;15(5-6):249-257 [[FREE Full text](#)] [doi: [10.1016/j.drug.2012.10.003](https://doi.org/10.1016/j.drug.2012.10.003)] [Medline: [23164555](#)]
7. Ostapchuk M, Roberts DM, Haddy R. Community-acquired pneumonia in infants and children. Am Fam Physician 2004 Sep 1;70(5):899-908 [[FREE Full text](#)] [Medline: [15368729](#)]
8. Cordoba G, Siersma V, Lopez-Valcarcel B, Bjerrum L, Llor C, Aabenhus R, et al. Prescribing style and variation in antibiotic prescriptions for sore throat: cross-sectional study across six countries. BMC Fam Pract 2015 Jan 29;16:7 [[FREE Full text](#)] [doi: [10.1186/s12875-015-0224-y](https://doi.org/10.1186/s12875-015-0224-y)] [Medline: [25630870](#)]
9. Wu M, Sirota M, Butte AJ, Chen B. Characteristics of drug combination therapy in oncology by analyzing clinical trial data on ClinicalTrials.gov. Pac Symp Biocomput 2015:68-79 [[FREE Full text](#)] [Medline: [25592569](#)]
10. Chou TC. Drug combination studies and their synergy quantification using the Chou-Talalay method. Cancer Res 2010 Jan 15;70(2):440-446 [[FREE Full text](#)] [doi: [10.1158/0008-5472.CAN-09-1947](https://doi.org/10.1158/0008-5472.CAN-09-1947)] [Medline: [20068163](#)]
11. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. J Am Med Inform Assoc 2013 Jan 1;20(1):144-151 [[FREE Full text](#)] [doi: [10.1136/amiajnl-2011-000681](https://doi.org/10.1136/amiajnl-2011-000681)] [Medline: [22733976](#)]
12. Miotto R, Li L, Kidd BA, Dudley JT. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. Sci Rep 2016 Dec 17;6:26094 [[FREE Full text](#)] [doi: [10.1038/srep26094](https://doi.org/10.1038/srep26094)] [Medline: [27185194](#)]
13. Aggarwal CC, Han J, editors. Frequent Pattern Mining. New York City: Springer; 2014:471.

14. Fournier-Viger P. The Data Mining Blog. 2013 Oct 13. An introduction to frequent pattern mining URL:<http://data-mining.philippe-fournier-viger.com/introduction-frequent-pattern-mining> [accessed 2018-11-12] [WebCite Cache ID 73rL4sloy]
15. Wright AP, Wright AT, McCoy AB, Sittig DF. The use of sequential pattern mining to predict next prescribed medications. *J Biomed Inform* 2015 Feb;53:73-80 [FREE Full text] [doi: [10.1016/j.jbi.2014.09.003](https://doi.org/10.1016/j.jbi.2014.09.003)] [Medline: [25236952](https://pubmed.ncbi.nlm.nih.gov/25236952/)]
16. Last M, Carel R, Barak D. Utilization of data-mining techniques for evaluation of patterns of asthma drugs use by ambulatory patients in a large health maintenance organization. 2007 Presented at: Proceedings of Seventh IEEE International Conference on Data Mining Workshops (ICDMW). IEEE; October 28-31, 2007; Omaha, Nebraska p. 169-174. [doi: [10.1109/ICDMW.2007.50](https://doi.org/10.1109/ICDMW.2007.50)]
17. Wikipedia. 2005. Chinese patent medicine URL:https://en.wikipedia.org/wiki/Chinese_patent_medicine [accessed 2018-11-12] [WebCite Cache ID 73rMgVxTQ]
18. Sun H, Gu Z, Gao C, Yu G. Descriptive analysis of medication mode in hospitalized children with pneumonia by electronic medical records massive data. *Pharm Care Res* 2014 Aug 30;14(4):264-267 [FREE Full text] [doi: [10.5428/pcar20140408](https://doi.org/10.5428/pcar20140408)]
19. Craig T. Pattern builders. 2011 Mar 2. Tales of Beers and Diapers URL:<https://blog.patternbuilders.com/2011/03/02/tales-of-beers-and-diapers> [accessed 2018-11-12] [WebCite Cache ID 73rLN2PrL]
20. Han J, Pei J, Yin Y. Mining frequent patterns without candidate generation. 2000 Presented at: SIGMOD '00 Proceedings of the 2000 ACM SIGMOD international conference on Management of data; May 15-18, 2000; Dallas, Texas, USA p. 1-12. [doi: [10.1145/342009.335372](https://doi.org/10.1145/342009.335372)]
21. Pei J, Han J, Mortazavi-Asl B, Wang J, Pinto H, Chen Q, et al. Mining sequential patterns by pattern-growth: the prefixspan approach. *IEEE Trans Knowl Data Eng* 2004 Nov;16(11):1424-1440. [doi: [10.1109/TKDE.2004.77](https://doi.org/10.1109/TKDE.2004.77)]
22. Yin J, Zheng Z, Cao L. USpan: an efficient algorithm for mining high utility sequential patterns. 2012 Presented at: KDD '12 Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining; August 12-16, 2012; Beijing, China p. 660-668. [doi: [10.1145/2339530.2339636](https://doi.org/10.1145/2339530.2339636)]
23. Chan R, Yang Q, Shen Y. Mining high utility itemsets. 2003 Presented at: ICDM '03 Proceedings of the Third IEEE International Conference on Data Mining; November 19-22, 2003; Melbourne, Florida p. 19-26. [doi: [10.1109/ICDM.2003.1250893](https://doi.org/10.1109/ICDM.2003.1250893)]
24. Armstrong JS. Findings from evidence-based forecasting: methods for reducing forecast error. *Int J Forecast* 2006 Apr;22(3):583-598. [doi: [10.1016/j.ijforecast.2006.04.006](https://doi.org/10.1016/j.ijforecast.2006.04.006)]
25. Knirsch CA, Chandra R. Azithromycin and the risk of cardiovascular death. *N Engl J Med* 2012 Dec 23;367(8):772-773. [doi: [10.1056/NEJMc1207269](https://doi.org/10.1056/NEJMc1207269)] [Medline: [22913693](https://pubmed.ncbi.nlm.nih.gov/22913693/)]
26. Tilelli JA, Smith KM, Pettignano R. Life-threatening bradyarrhythmia after massive azithromycin overdose. *Pharmacotherapy* 2006 Jan;26(1):147-150. [Medline: [16506357](https://pubmed.ncbi.nlm.nih.gov/16506357/)]
27. McKenzie JE, Cruz-Rivera M. Compatibility of budesonide inhalation suspension with four nebulizing solutions. *Ann Pharmacother* 2004 Jun;38(6):967-972. [doi: [10.1345/aph.1D335](https://doi.org/10.1345/aph.1D335)] [Medline: [15084687](https://pubmed.ncbi.nlm.nih.gov/15084687/)]
28. Melani AS. Effects on aerosol performance of mixing of either budesonide or beclomethasone dipropionate with albuterol and ipratropium bromide. *Respir Care* 2011 Mar;56(3):319-326 [FREE Full text] [doi: [10.4187/respcare.00955](https://doi.org/10.4187/respcare.00955)] [Medline: [21255509](https://pubmed.ncbi.nlm.nih.gov/21255509/)]
29. Kan J, Zhu X, Wang T, Lu R, Spencer PS. Chinese patient demand for intravenous therapy: a preliminary survey. *Lancet* 2015 Oct 1;386:S61. [doi: [10.1016/S0140-6736\(15\)00642-X](https://doi.org/10.1016/S0140-6736(15)00642-X)]
30. Kuhn M, Yates P, Hyde C. In: Zhang L, Kuhn M, Peers I, Altan S, editors. Nonclinical statistics for pharmaceutical and biotechnology industries. New York: Springer; Jan 22, 2016:698.
31. Marx V. Tissue engineering: organs from the lab. *Nature* 2015 Jun 18;522(7556):373-377. [doi: [10.1038/522373a](https://doi.org/10.1038/522373a)] [Medline: [26085275](https://pubmed.ncbi.nlm.nih.gov/26085275/)]
32. Ling Z, Liu X, Cheng Y, Luo Y, Yuan L, Li L, et al. Clostridium butyricum combined with Bifidobacterium infantis probiotic mixture restores fecal microbiota and attenuates systemic inflammation in mice with antibiotic-associated diarrhea. *Biomed Res Int* 2015;2015:582048 [FREE Full text] [doi: [10.1155/2015/582048](https://doi.org/10.1155/2015/582048)] [Medline: [25802855](https://pubmed.ncbi.nlm.nih.gov/25802855/)]
33. Liu P, Wang P, Tian D, Liu J, Chen G, Liu S. Study on traditional Chinese medicine theory of lung being connected with large intestine. *J Tradit Chin Med* 2012 Sep;32(3):482-487 [FREE Full text] [Medline: [23297577](https://pubmed.ncbi.nlm.nih.gov/23297577/)]
34. Yan CL, Zhu YL. The Treatment of External Diseases With Acupuncture and Moxibustion. Boulder, CO: Blue Poppy Pr; Aug 1999:253.
35. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)* 2012;22(3):276-282 [FREE Full text] [Medline: [23092060](https://pubmed.ncbi.nlm.nih.gov/23092060/)]
36. Mjösberg J, Rao A. Lung inflammation originating in the gut. *Science* 2018 Dec 5;359(6371):36-37. [doi: [10.1126/science.aar4301](https://doi.org/10.1126/science.aar4301)] [Medline: [29302003](https://pubmed.ncbi.nlm.nih.gov/29302003/)]
37. Huang Y, Mao K, Chen X, Sun M, Kawabe T, Li W, et al. S1P-dependent interorgan trafficking of group 2 innate lymphoid cells supports host defense. *Science* 2018 Dec 5;359(6371):114-119. [doi: [10.1126/science.aam5809](https://doi.org/10.1126/science.aam5809)] [Medline: [29302015](https://pubmed.ncbi.nlm.nih.gov/29302015/)]
38. Vergis EN, Indorf A, File TM, Phillips J, Bates J, Tan J, et al. Azithromycin vs cefuroxime plus erythromycin for empirical treatment of community-acquired pneumonia in hospitalized patients: a prospective, randomized, multicenter trial. *Arch Intern Med* 2000 May 8;160(9):1294-1300. [Medline: [10809032](https://pubmed.ncbi.nlm.nih.gov/10809032/)]

39. Rubio FG, Cunha CA, Lundgren FL, Lima MP, Teixeira PJ, Oliveira JC, et al. Intravenous azithromycin plus ceftriaxone followed by oral azithromycin for the treatment of inpatients with community-acquired pneumonia: an open-label, non-comparative multicenter trial. *Braz J Infect Dis* 2008 Jun;12(3):202-209 [[FREE Full text](#)] [Medline: [18833404](#)]
40. Xiao Y, Zhang J, Zheng B, Zhao L, Li S, Li L. Changes in Chinese policies to promote the rational use of antibiotics. *PLoS Med* 2013 Nov;10(11):e1001556 [[FREE Full text](#)] [doi: [10.1371/journal.pmed.1001556](#)] [Medline: [24260030](#)]

Abbreviations

CHS: Children's Hospital of Shanghai

EMR: electronic medical record

GI: gastrointestinal

IV: intravenous

TCM: traditional Chinese medicine

WHO: World Health Organization

Edited by G Eysenbach; submitted 22.10.18; peer-reviewed by Y Chen, MJ Kang, M Wan; comments to author 07.11.18; revised version received 11.11.18; accepted 20.11.18; published 22.03.19

Please cite as:

Tang C, Sun H, Xiong Y, Yang J, Vitale C, Ruan L, Ai A, Yu G, Ma J, Bates D

Medication Use for Childhood Pneumonia at a Children's Hospital in Shanghai, China: Analysis of Pattern Mining Algorithms

JMIR Med Inform 2019;7(1):e12577

URL: <http://medinform.jmir.org/2019/1/e12577/>

doi: [10.2196/12577](https://doi.org/10.2196/12577)

PMID: [30900998](https://pubmed.ncbi.nlm.nih.gov/30900998/)

©Chunlei Tang, Huajun Sun, Yun Xiong, Jiahong Yang, Christopher Vitale, Lu Ruan, Angela Ai, Guangjun Yu, Jing Ma, David Bates. Originally published in *JMIR Medical Informatics* (<http://medinform.jmir.org>), 22.03.2019. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Medical Informatics*, is properly cited. The complete bibliographic information, a link to the original publication on <http://medinform.jmir.org/>, as well as this copyright and license information must be included.