

The Characteristics of Doctors, Patients and Question Records in Online Medical Communities

Hao Liu^{1#}, Shi-Qi Wang^{2#}, Xiao-Ni Hou¹, Hong-Mei Duan^{1*}, Kun-ling Shen³

¹School of Nursing, Beijing University of Chinese Medicine, Beijing 10029, China, ²Department of Information Management, Peking University, China, ³National Clinical Medical Research Center for Respiratory Diseases, Beijing Children's Hospital, The Capital Medical University.

***Corresponding Author:** Hong-mei Duan, School of Nursing, Beijing University of Chinese Medicine, Beijing 10029, China. # - Contributes Equally to this Work.

Abstract

Objective: This article starts to explore the characteristics of users, doctors and the question & answer records related to childhood asthma, by analyzing the data from three online medical communities. It then develops a topic model to identify the topics in the question-answer records which can help to find specific needs of users.

Method: We developed a topic model to identify the topics in the question-answer records which can help to find specific needs of users. After data cleaning and word segmentation, we applied a text mining technique, Latent Dirichlet Allocation (LDA) topic modeling, to analyze the themes of the question records and the characteristics of patients and doctors.

Results: Finally, 2173 question records were divided into eight themes which included knowledge requirements, emotional needs and nursing needs. Network data were found to correspond with childhood asthma epidemiology. Fewer doctors from children's speciality hospitals were engaged with the online medical communities than other types of facilities.

Conclusion: The data in online medical communities can reflect the needs of users, which can help medical staff to improve medical services.

Keywords: online medical communities, Latent Dirichlet Allocation, thematic analysis, childhood, asthma, health information needs.

INTRODUCTION

In recent years, online medical communities established a platform for health information exchange between doctors and patients. By 2017, the number of Internet medical users had reached 253 million, and the usage rate had reached 32.7% in China. Online medical communities have developed "reservation registration", "patients forum", "medical questions and answers (Medical Q&A)", "health information" and other functions. "Medical Q&A" is the main way for users (patients or patients' caregivers) to communicate with health professionals. The records from these questions, comments and answers provide various data sources for scholars to study the health behaviors of specific populations. Colineau¹ found that from the user's point of view, most patients were seeking information about their medical conditions and to interact with peers online. Goh² found that

the emergence of online health communities has an important impact on improving the difference between urban and rural health care. Buis³ found that patients experiencing low-survival-rate cancers may have a greater desire for informational support online than patients experiencing high-survival-rate cancers by comparing the posts in eight online communities. The research methods of the online medical communities have evolved from survey interviews, to manual coding to text mining.^{4,5} The LDA (Latent Dirichlet Allocation) topic model is widely used in text feature mining. It is a dynamic topic model based on probability and statistical methods, which can be used to identify potential topics from the corpus.⁶

Most scholars focus on adult diseases. There are fewer studies on pediatric diseases and the caregivers of child patients. Childhood asthma is a heterogeneous disease which has characteristics include chronic

The Characteristics of Doctors, Patients and Question Records in Online Medical Communities

airway inflammation and high airway response. In China, the prevalence of asthma was 3.02% in children under 14 years old and has been increasing at a rate of 50% per decade.^{7,8} The long disease duration, relapse, and lung function damage caused by improper control of childhood asthma is an important factor in the incidence of adult asthma and chronic obstructive pulmonary disease. Therefore, children with asthma and their caregivers (in China the main caregiver is usually the mother) need long-term disease management with the help of medical staff to control the disease and to improve the quality of life. Online medical communities provide a mechanism for caregivers to obtain health information and to express their health needs and doubts about asthma. We can use the LDA method to extract the topics of concern

to caregivers and the health information needs of caregivers from the question records in the online medical communities, which can help to improve the quality of medical services and the disease management ability of asthmatic children and their caregivers.

METHODS

Data Collection

From the Chinese website traffic analysis website “Stationhouse”, we selected the top three online medical communities with the highest number of views, and crawled their medical questions and answers module as the data source. The data collection date was June 1, 2019. All websites and the number of data records collected are provided in Table 1.

Table 1. Basic Information of the Online Medical Communities and Data Collection (Children’s Asthma Theme)

Website Name	Website Address	The Number of Items Related to Children’s Asthma
Answer all questions	www.120ask.com	658
39Health	www.39.net	795
Ask the doctors	www.xywy.com	720

Data Collection and Pre-Process

A crawler software program is used to collect information on the childhood asthma questions and answers records from three websites. The data includes the name of the questioner, the month, the content of the questions, and the name, unit and job title of responders. After the data text was exported, we performed data preprocessing, which included deleting blank fields, numbers, English letters, special characters, and short meaningless texts. In the end, 2173 standardized data packets were obtained.

Question Text Process

The “Jieba” module in Python was applied to segment the question text and “Sogou” medical lexicon was

imported as the word dictionary. Conventional stop words and single Chinese characters which usually cannot express semantic features were removed. Only the words which have more than one Chinese character were kept, then lower frequency words were deleted. These processed words were used to create a corpus and a document-word matrix. Word frequency statistics on the processed words, and some high-frequency words are shown in Table 2. Some high-frequency words, such as symptoms and trade names of drugs, were used to merge the vocabulary with the same meaning and expression rules.⁷ The mapping rules for partial synonym merging are shown in Table 3.

Table 2. Some High Frequency Words.

Words	Frequency
asthma	2723
treatment	899
cough	455
hospital	455
medical examination	379

Table 3. Some Mapping Rules.

Before Mapping	After Mapping
Pediatric asthma	childhood asthma
Singulair	Montelukast
The noise of lung	wheezing
chest feel oppressed	chest distress
can’t breathe	dyspnea

The Question Records Analysis

The question records were analyzed by using the LDA topic model, and the asthmatic children and medical staff characteristics were manipulated using statistical methods for analysis. The LDA topic model was proposed by Blei, et al, in 2003. It can uncover unobserved information based on observable data.⁵ The model is shown in Figure 1; where the symbol w means observable; the symbol φ is the distribution of the subject, and $p(w/z)$ obeys the Dirichlet β

distribution; θ is the subject distribution of the document, and $p(z/d)$ obeys the Dirichlet distribution α . The main task of the solution is to estimate the value of the implicit variable subject distribution θ and the word distribution φ . In the study, Gibbs sampling method is used to train the model, and the subject of each word is randomly given. When other variables remain unchanged, a new topic of each word is generated according to the transition probability sampling. After repeated iterations, the convergence result is the topic distribution and word distribution.

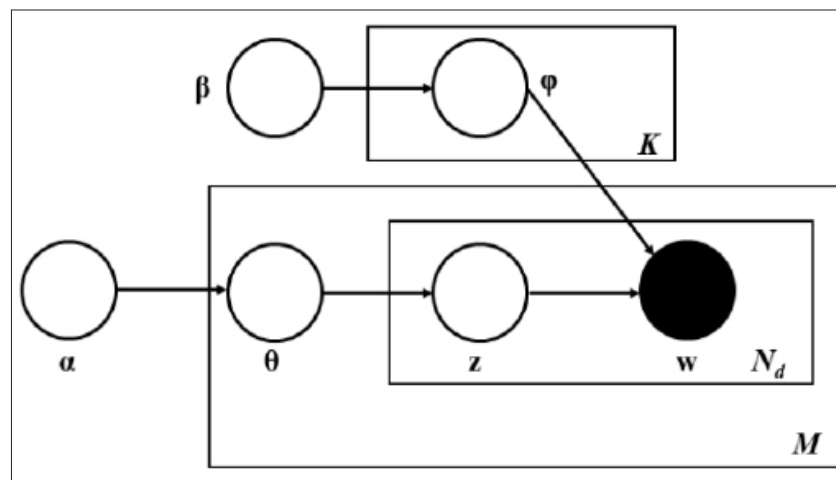


Fig 1. LDA Model

Applying the machine learning module “sklearn” in Python to process the LDA topic modeling, 500 high frequency words with frequency higher than 10 were extracted as feature keywords to generate a feature word vector space matrix. After the model was established, the number of iterations was set to 500, and the number of topics, K , was set to 5, 8, 10, and 20 separately to perform the modeling. The keyword prior Dirichlet distribution parameters η and α were selected as default values, and the solution algorithm used ‘batch’.

After several experiments, we found that the larger the K value is set, the more overlapping feature words will be in the subject, and it was difficult to generalize the independent and low-relevance topics. After substituting the smaller K value (5~10) into the LDA topic model, it was found that when the K value increases from 5 to 8, the distribution probability of the feature words in each topic is gradually increased. However, when the category was small, according to the theme, the correlation between the topics identified by the feature words is relatively large and the degree of discrimination is not strong. When the K value is 9 to

10, the distribution probability of the feature words in each topic decreases. By comparing the experimental results, it was found that when the subject number K value is 8, the probability distribution of the feature words in the subject-character word probability distribution matrix is relatively high, and the related topics were less relevant. So, we set the number of topics to 8 and summarized the corresponding theme features based on professional knowledge.

RESULTS

Gender and Age Distribution of Asthmatic Children

The asthmatic children data is collected from information provided by caregivers when they registered their accounts in the online medical communities database. From the description in the tables below, it can be seen that there are more asthmatic boys than girls in the online medical community. In terms of age, the highest proportion is preschool children, followed by school-age children, and then infants and toddlers.

Table 4. Gender Distribution of Patients.

Gender	Count	Percentage
Male	1316	60.56%
Female	801	36.86%
Missing Information	56	2.58%
Total	2173	100%

Table 5. Age Distribution of Patients.

Age	Count	Percentage
0~3	460	21.17%
4~7	955	43.95%
8~14	735	33.82%
14~	23	1.06%
Total	2173	100%

According to three epidemiological surveys of childhood asthma in China⁸, the prevalence of asthmatic boys was higher than girls as 1.49:1, 1.66:1, and 1.57:1. Preschool children are the highest in age distribution, and the prevalence of school-age children is close to that of infants. The results above show that the network data reflect the epidemiological characteristics of childhood asthma to some degree.

Characteristics of the Question Month

After counting the number of questions in different months, it was found that April and May were the peaks in the first half year. In the second half year, the number of questions increased significantly after October. The growth trend of questions matched with the characteristics of the disease, as shown in Figure

2. Childhood asthma is closely related to allergy factors, and frequent pollen in spring and autumn is an important cause of asthma exacerbation.⁹ Winter respiratory infection is also an important cause of asthma attacks. Therefore, in the above two times, the number of questions shows an increasing trend. The number of questions drops from December to January. It may be the Chinese children are busy on study in the end of the term, so parents prefer to consult the doctors on internet. We also observe that the visiting rate in the asthma department always increases in the winter vacation after December. This fact is consistent with our team’s previous experience of a dramatic increase in children with asthma during the winter vacation.

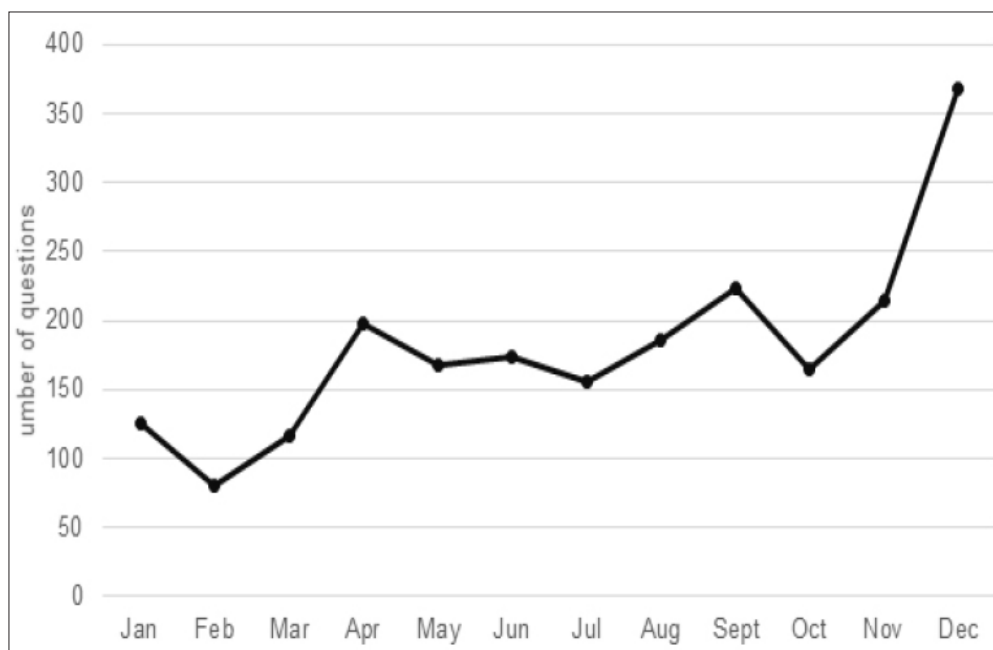


Fig 2. Monthly Trend of Questions

Characteristics of Practices and Titles of Medical Staff

Medical staff from many different disciplines, including doctors, nurses, technicians, and psychologists, are involved in the online medical

communities. Doctors and nurses were the main participants, among which doctors have the highest active response rates. Those doctors with junior and intermediate job titles are more active than those with senior titles, as shown in Figure-3.

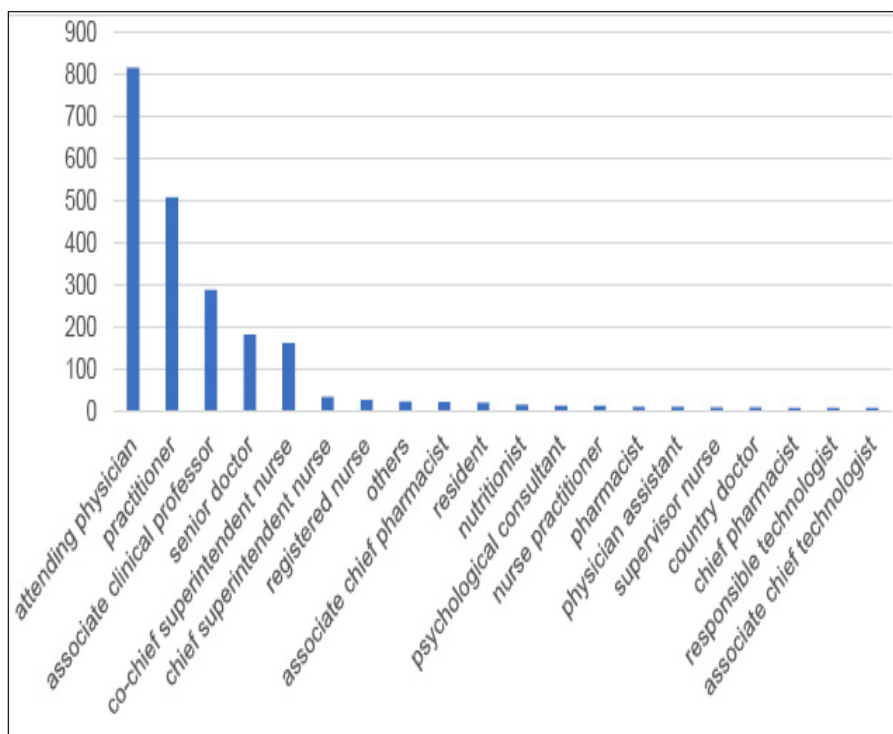


Fig 3. Response Characteristics by Medical Staff Practices.

Hospital Characteristics of Medical Staff

The hospital affiliations of the medical staff range from urban health clinics to 3A (Class Three/Grade A) hospitals, with comprehensive hospitals as the mainstay, and fewer children’s specialty hospitals. By the end of 2018, there were 228 children’s specialty hospitals in China, but the number of pediatricians per 1,000 population was only 0.63. The network data also reflected a shortage of pediatric medical resources in

China and the low participation of pediatricians in children’s specialty hospitals.

The top five active-response hospitals are Feicheng Mining Central Hospital, China-Japan Friendship Hospital, Beiliu People’s Hospital, Lianyungang Municipal Oriental Hospital, and Second Affiliated Hospital of Suzhou University. The basic information of the top 20 hospitals in response to activity is shown in Table 6.

Table 6. Basic Information of Hospitals with High Activity

Analysis of theme characteristics of the question records

Location	Number	Type of Hospital	Number	Grade of Hospital	Number
First tier city	6	General hospital	19	Class Three/Grade A hospital	18
Second tier city	5	Specialized hospital	1	Class Three/Grade B hospital	1
Other city	9			Class Two/Grade A hospital	1

The LDA model divides 2173 question records into eight major themes, which reflects, to a certain extent, the caregiver’s health information

needs and concerns for childhood asthma. The theme and top ten keywords are shown in Table 7.

Table 7. Eight Themes Derived from the Question Text

Number	Theme	Key Words	Number of Questions
1	Treatment of childhood asthma	children, cough, wheeze, severity, how to deal with, dyspnea, will appear, recently, condition, discover	415
2	Asthma treatment and medical needs	radical cure, treat, want to know, children, how, attack, what, children's hospital, childhood asthma, help	395
3	Worry about illness	relatively, family, worry, extremely, want to know, suffer from asthma, in the heart, don't know, now, how	306
4	Clinical manifestations of asthma	gasp, symptoms, which, sneeze, chest distress, feeling, breathe, childhood asthma, cough, sound	273
5	Pathogeny and prognosis of asthma	childhood asthma, can or not, cure, consult, inheritance, don't know, how, what is, be caused by, treatment	209
6	The medical examination of asthma	repeatedly, examination, my son, medical laboratory, a few times, around, effect, doctor, asthmatic attack, hospital	202
7	The nursing and Chinese traditional treatment of childhood asthma	children, pay attention to, what is, relatively, what's the matter, don't know, influenza, can cure, Chinese traditional medicine, effect	196
8	Allergen prevention long-term safety of asthma medicine	relapse, how to prevent, how to cure asthma, control allergen, side-effect, long term, fluticasone propionate, cough, chest distress, allergen	177

DISCUSSION

Main Theme Analysis and Lessons from the Themes of the Question Text

The first and fourth themes are related to childhood asthma symptoms and treatment. In the records, "What should I do if my child has severe asthma?", "The baby always coughs, like he cannot breathe", show caregivers' concerns about asthma symptoms and their apprehension about the special medications when the symptoms occur. Asthma can be triggered by stimuli such as exercise, contact with allergens, weather changes, laughter or car exhaust, and viral infections (respiratory syncytial virus, rhinovirus, etc.) are also main predisposing factors.¹⁰ A multi-perspective qualitative study shows that some caregivers expressed difficulty in identifying triggers and symptoms of asthma and lacked awareness regarding the appropriate management of the disease.¹¹ Asthma management requires caregivers to identify and

properly process the symptoms of asthma attacks, and to ensure that the child follows the treatment regimen, participates in further consultations, and avoids asthma-causing factors.¹² At present, the medical faculty focuses on drug therapies and dosage details, and most caregivers do not know how to identify and control asthma symptoms.¹³ Therefore, clinical work should focus on the health education of symptom recognition and treatment methods. We should also notice that the majority of caregivers in developing countries have a poor educational background. Thus clinical work should provide a form of educational campaigns that are easily understood.¹⁴

The second and seventh themes are related to the treatment of asthma and daily care management. In the records, statements like, "The major hospital says that you have asthma, how to treat it?", "What are the best ways for traditional Chinese medical practice to treat asthma?", "What do we need to pay

attention to during daily life, etc.”, show caregivers uncertainty about asthma treatment and care. The lack of communication between doctors and patients is an important reason for caregivers’ lack of understanding of treatment options. Therefore, medical staff should provide a good communication support system for asthma families to alleviate their doubts during the visit and treatment.¹² In addition to routine treatment counseling, traditional Chinese medicine options have also received attention. Hao¹⁵ and others compared China’s online medical question and answer communities with the United States’ and found that caregivers’ concerns about children’s health are more reflected in the comments related to the practice of traditional Chinese medicine. Most often, Chinese parents prefer to take their children to visit traditional Chinese medicine specialists to avoid the potential side effects of biomedicine.

The third theme is caregivers’ concerns about the condition of their children. In the records we see comments such as “The child has been suffering from asthma for a long time, and I am worried about that.”, “The child is very sick after exercise, I am very worried.”, etc. These comments are seen as expressions of caregivers’ concerns about the condition of their child. In the process of asthma management, caregivers, especially mothers, often are uncertain about the disease and they fear asthma attacks. When they find out that they are unable to control their children’s symptoms, they will be sad, and may show symptoms of depression and anxiety.¹⁶ There is an interplay between childhood asthma and parents’ mental health. A mother’s symptoms of depression can exacerbate her child’s asthma symptoms, and the progression of a child’s asthma can adversely affect a mother’s mental health. Ameliorating maternal depressive symptoms may result in better asthma outcomes for children.¹² Health information support and psychological intervention from medical staff can help caregivers understand their own dilemmas and respond to psychological problems, so they can take care of their children better.¹⁷

CONCLUSION

Everywhere in the world, Asthma control is an arduous challenge in children. According to the research, over 2 million children in the US have uncontrolled asthma, and mild asthma attacks occur at least weekly in 11% of children. Serious attacks occur at least annually in

35%.^{18, 19} Lack of knowledge is an important reason for the lack of control in asthmatic children. In Brazil, 75 percent of parents believe that inhaling drugs can lead to addiction, and two-thirds believe that outdoor smoking has no effect on children.²⁰ In South Korea, school-aged children did not have enough knowledge to sufficiently manage their asthma.²¹ The online medical communities provides a platform for users to acquire asthma knowledge, such as understanding about diagnosis and treatment of the disease, to learn how to perform daily management needs, and to gain access to health information. For patients with chronic diseases, the mutual-assisted communication mode of a variety of users can play an important and positive role in disease control.¹¹

Using “childhood asthma” as an entry, we analyzed the characteristics of patients and medical staff, and found that the network data reflects the epidemiological characteristics of childhood asthma in China. Different kinds of medical institutions participate in online communities Q&A, and medical staff from multiple practice disciplines participate, though mainly doctors with intermediate job titles. The eight themes that LDA uncovered show different aspects of the needs and doubts of childhood asthma, including disease knowledge, medical needs, emotional needs, and daily care, etc. In developing countries, nurses are the primary care providers who are responsible for health education. However, health education is not always emphasized since nurses often lack the time to investigate the patient’s personalized needs due to their heavy workload. According to our results, health staff can provide more practical medical service based on patient’s real needs rather than simply from the individual experience of the practitioner.

The results also suggest that medical staff should strengthen communications with caregivers, especially focusing on the daily management of the disease. Staff also should pay attention to any psychological changes in patients and caregivers. For the builders of the online medical communities, the categories should be refined according to the characteristics of different diseases, and the corresponding modules should be set up according to the needs of users to increase the convenience and accuracy of obtaining health information.¹² In addition to all of the findings above, the health department should vigorously cultivate pediatric professionals to alleviate the current shortage of pediatric medical resources in

China. Finally, pediatricians in specialty hospitals should be encouraged to participate in the Q&A of the online medical communities.

LIMITATIONS AND RECOMMENDATIONS

This study only focuses on the problems and concerns seen in asthmatic children and their caregivers as reflected by the online medical communities in China. In the future, we may conduct more in-depth research to compare domestic and international online communities. Also, other health problems need to be a focus for further study. In addition, the deep mining and analysis of text data, such as named entity recognition and sentiment analysis, are also worthy topics for future research.

ACKNOWLEDGEMENT

We thank to Jim Stout, Indiana University (ret) for editing help.

FUNDING

The Research is Supported by the Key Laboratory of Evidence Based Medicine and Knowledge Translation of Gansu Province (No.GSEBMKT-2020YJ04).

REFERENCES

- [1] Colineau N, Paris C. Talking about your health to strangers: understanding the use of online social networks by patients[J]. *New Rev Hypermedia M*.2010;16(1-2):141-160. doi:10.1080/13614568.2010.496131.
- [2] Goh JM, Gao GD, Agarwal R. The creation of social value: can an online health communities reduce rural-urban health disparities? *Mis Quarterly*. 2016;40(1): 247-263. doi:10.25300/misq/2016/40.1.11.
- [3] Buis L R, Whitten P. Comparison of social support content within online communities for high- and low-survival-rate cancers. *Comput Inform Nurs*. 2011;29(8):461-467. doi:10.1097/ncn.0b013e318214093b.
- [4] Macias W1, Lewis LS, Smith TL. Health-related message boards/chat rooms on the Web: discussion content and implications for pharmaceutical sponsorships[J]. *J Health Commun*.2005;10(3):209-23. doi: 10.1080/10810730590934235.
- [5] Ku Y, Chiu C, Zhang Y, et al. Text mining self-disclosing health information for public health service. *J Assoc Inf Sci Tech*. 2014;65(5):928-947. doi:10.1002/asi.23025.
- [6] Blei DM, Ng AY, Jordan M I. Latent Dirichlet allocation[J]. *J Mach Learn Res Archive*, 2003;3: 993-1022. doi:10.3724/SP.J.1004.2014.02262.
- [7] Guo X, Li Z, Ling W, et al. Epidemiology of childhood asthma in mainland China (1988-2014): a meta-analysis. *Allergy Asthma Proc*. 2018;39:15-29. doi:10.2500/aap.2018.39.4131.
- [8] Wu JZ, Zhong TL, Zhu Y. Effects of particulate matter (PM) on childhood asthma exacerbation and control in Xiamen, China. *BMC PEDIATRICS*. 2019;19(1):194. doi:10.1186/s12887-019-1530-7.
- [9] Shrestha SK, Katelaris C, Dharmage SC, et al. High ambient levels of grass, weed and other pollen are associated with asthma admissions in children and adolescents: A large 5-year case-crossover study. *Clin Exp Allergy*. 2018;48(11): 1421-1428. doi:10.1111/cea.13225.
- [10] Caudri D, Wijga A, A Schipper CM, et al. Predicting the long-term prognosis of children with symptoms suggestive of asthma at preschool age. *J Allergy & Clin Immun*. 2009; 124(5): 903-910. doi:10.1016/j.jaci.2009.06.045.
- [11] Aidan S, Russell J, John H. Children's, parents' and health professionals' views on the management of childhood asthma: a qualitative study. *Prim Care Respir Med*.2017;27(53):1-6. doi:10.1038/s41533-017-0053-7.
- [12] Otsuki M, Eakin MN, Arceneaux LL. Prospective Relationship between Maternal Depressive Symptoms and Asthma Morbidity among Inner-City African American Children. *J Pediatr Psychol*. 2010;35(7): 758-767. doi:10.1093/jpepsy/jsp091
- [13] Searle A, Jago R, Henderson J, et al. Children's, parents' and health professionals' views on the management of childhood asthma: a qualitative study. *Prim Care Respir Med*.2017;27(1): 53. doi:10.1038/s41533-017-0053-7.
- [14] Sombans S, Lohana P. Finding a Better Solution for Asthma Control in Children. *Cureus*. 2018;10(9): e3344. doi:10.7759/cureus.3344.
- [15] Hao HJ, Zhang KP. The voice of Chinese health consumers: a text mining approach to web-based physician reviews. *J Med Internet Res*. 2016.18(5): e108. doi:10.2196/jmir.4430.

The Characteristics of Doctors, Patients and Question Records in Online Medical Communities

- [16] Chen SH, Huang JL, Yeh KW, et al. The Stress of Caring for Children With Asthma: A Qualitative Study of Primary Caregiver. *J Nurs Res*. 2015;23(4): 298-307. doi:10.1097/jnr.0000000000000083.
- [17] Chong YY, Leung D, Mak YW. When Control Exacerbates Distress: A Qualitative Study Exploring the Experiences of Hong Kong Chinese Parents in Caring for a Child with Asthma. *Int J Environ Res Public Health*. 2018;15(7): 1372. doi:10.3390/ijerph15071372.
- [18] DePriest K, Butz A, Gross D. Investigating the relationships among neighborhood factors and asthma control in African American children: A study protocol. *Res Nurs Health*. 2018;41(5):428-439. doi: 10.1002/nur.21901.
- [19] Carroll WD, Wildhaber J, Brand P L. Parent misperception of control in childhood/adolescent asthma: the Room to Breathe survey. *Eur Respir J*. 2012;39(1): 90-96. doi:10.1183/09031936.00048911.
- [20] Urrutia-Pereira M, Mocellin LP, de Oliveira RB, et al. Knowledge on asthma, food allergies, and anaphylaxis: Assessment of elementary school teachers, parents/caregivers of asthmatic children, and university students in Uruguaiana, in the state of Rio Grande do Sul, Brazil. *Allergol Immunopathol*. 2018;46(5): 421-430. doi:10.1016/j.aller.2017.09.018.
- [21] Kim, Seon Su, Insoo Kwon. Knowledge and Practice in Self-Management on Asthma of School-Aged Children with Asthma. *Child Health Nursing Research*. 2016;22(2):87-96. doi:10.4094/chnr.2016.22.2.87.

Citation: Hao Liu, Shi-Qi Wang, Xiao-Ni Hou, Hong-Mei Duan. *The Characteristics of Doctors, Patients and Question Records in Online Medical Communities. Archives of Pediatrics and Neonatology*. 2020; 3(1): 06-14.

Copyright: © 2020 Hao Liu, Shi-Qi Wang, Xiao-Ni Hou, Hong-Mei Duan. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.