Study Protocol

Validation of the accuracy of the childhood asthma model for clinical decision support: a study protocol

Na Dong[1#], Beirong Wu[1#], Bingru Yin[1], Wei Dong[2], Xiaoqun Jin[3], Miao Wang[4], Xiuhe Xu[5], Canghong Zhi[6], Dandan Zhao[6], Min Lu[1], Haoxiang Gu[1], Rong Qiao[7]

[1]Department of Respiration, Children's Hospital of Shanghai, Shanghai, China; [2]Department of Pediatrics, Nanxiang Hospital of Jiading District, Shanghai, China; [3]Department of Pediatrics, People's Hospital of Shanghai Putuo District, Shanghai, China; [4]Department of Medical Affairs, Children's Hospital of Shanghai, Shanghai, China; [5]Department of Pediatrics, Shibei Hospital of Shanghai, Shanghai, China; [6]Joincare Pharmaceutical Group Industry Co., Ltd., Shenzhen, China; [7]Department of Gastroenterology, Children's Hospital of Shanghai, Shanghai, China
[#]These authors contributed equally to this work.
Correspondence to: Min Lu, PhD; Haoxiang Gu, MD. Department of Respiration, Children's Hospital of Shanghai, No. 24, Lane 1400, Beijing West Road, Shanghai 200040, China. Email: lumin61@aliyun.com; guhx@shchildren.com.cn. Rong Qiao, MD. Department of Gastroenterology, Children's Hospital of Shanghai, No. 355, Luding Road, Shanghai 200040, China. Email: qiaor@shchildren.com.cn.

Background: In China, the average prevalence of asthma in children aged 0–14 years increased by approximately 50% every 10 years. Hence, a specific decision support system that fits China's situation is needed for childhood asthma. This prospective, multicenter, observational study aims to assess the accuracy of the Childhood Asthma Model for Clinical Decision Support (CAMCDS) in clinical practice in four hospitals in Shanghai in China.

Methods: The study will be conducted in two phases. Phase I of the study aims to evaluate the accuracy of the CAMCDS for diagnosis, while phase II of the study aims to examine the treatment predicting accuracy of the CAMCDS model. In total, 817 children diagnosed with stable asthma and 545 suspected asthma will be enrolled. The accuracy of the CAMCDS model will be calculated using the receiver operating characteristic (ROC) curve compared with the results of pediatrician's diagnosis. Besides, the treatment patterns from CAMCDS and real-world environment for Chinese children with stable asthma will be assessed, and the factors that affect the CAMCDS implementation in routine clinical practice will be explored.

Conclusions: This will be the first study to examine the diagnostic accuracy and treatment predicting accuracy of a clinical decision support system in children with asthma in China. We hope that the CAMCDS will be help pediatricians in basic-level hospitals to improve the diagnosis and treatment strategy of asthma.

Trial Registration: Chinese Clinical Trial Registry Identifier: ChiCTR2100045283.

Keywords: Accuracy; children; asthma; clinical decision support system

Submitted Apr 20, 2021. Accepted for publication Sep 13, 2021.
doi: 10.21037/jtd-21-668
View this article at: https://dx.doi.org/10.21037/jtd-21-668

Introduction

Asthma is the most common chronic disease (1,2), and asthma exacerbations are common events in school-age children (3). Exacerbations involve huge medical costs and are associated with considerable complications. In addition, asthma exacerbations may lead to the progressive loss of lung function and greater asthma severity over time (4,5). In China, since the 1990s, the average prevalence of asthma in children aged 0–14 years increased by approximately 50% every 10 years (6-8).

Globally, educational programs and guidelines have emphasized the importance of symptom control and future risk of adverse outcomes, including the Expert Panel Report-3 of the National Asthma Education and Prevention Program (9), the Global Initiative for Asthma (GINA)

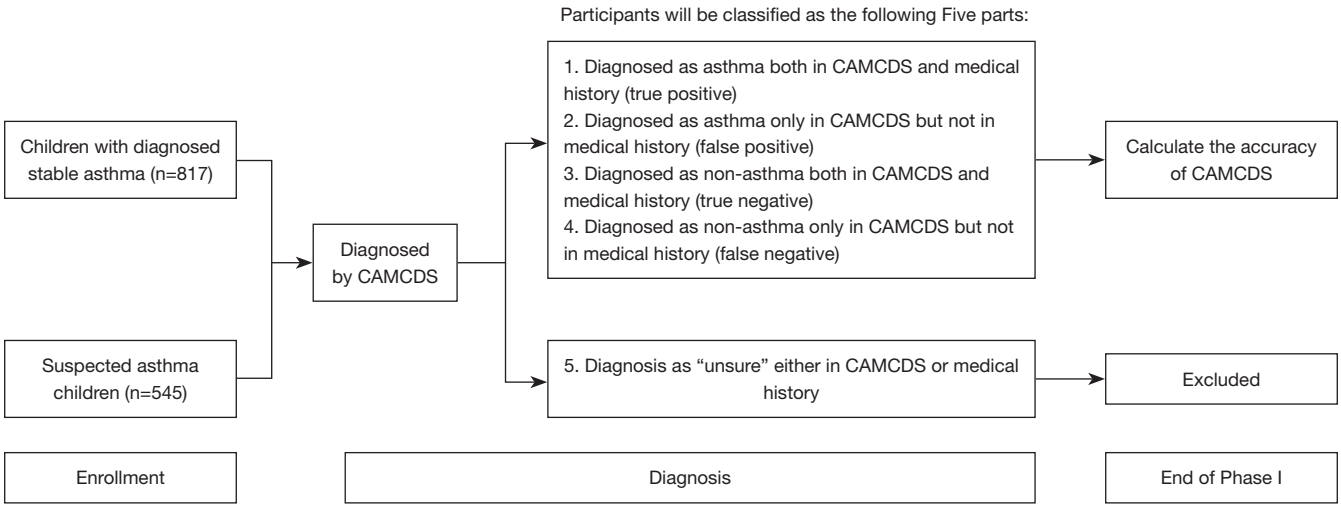

Figure 1 Study schema of Phase I. CAMCDS, Childhood Asthma Model for Clinical Decision Support.

guideline (10), and the Chinese Guideline for the diagnosis and optimal management of asthma in children (11). Although great efforts have been made toward managing children with asthma to reduce the recurrence, a considerable number of children cannot get an accurate assessment and standardized grading treatment due to the complex assessment process. Therefore, several clinical decision support systems, also referred to as the clinical prediction model, have been proposed and tested to facilitate decision-making in pediatric asthma in various developed countries (12). Moreover, the "MyAsthma portal" developed by Children's Hospital of Philadelphia also provides a model for using technology to foster shared decision-making in ambulatory care settings (13). However, a systemic review examined the effectiveness of such computerized decision support systems (CDSSs), and the results indicated difficulty in the improvements of outcomes for participants with asthma because they were rarely used and the advice was not followed in the real world (14). On the other hand, China also needs its own decision support system for childhood asthma due to the unique healthcare system and the uneven distribution of pediatric workload (15).

Therefore, this prospective, multicenter, observational study has been designed to verify the diagnostic accuracy and treatment predicting accuracy of our newly established Childhood Asthma Model for Clinical Decision Support (CAMCDS) model and provide data for the improvement and optimization of the model. We hope that the model will assist doctors in primary hospitals in improving the standardized diagnosis and treatment for children with asthma.

We present the following article in accordance with the SPIRIT reporting checklist (available at https://dx.doi.org/10.21037/jtd-21-668).

Methods

Study design

This study aims to assess the diagnostic accuracy and the treatment predicting accuracy of the CAMCDS in clinical practice in four hospitals in Shanghai in China. The study was approved by the Ethics Review Committee of Children's hospital of Shanghai (NO. 2021R009-E01).

The study will be conducted in two phases. Phase I of the study aims to evaluate the accuracy of CAMCDS for diagnosis. In phase I, children aged 4–14 years, who are diagnosed with stable asthma in routine clinical practice or have asthma-like symptoms but have not been diagnosed in the outpatient clinic will be enrolled. The study flowchart for phase I is shown in *Figure 1*. Briefly, baseline information, including demographic data, disease inducement, symptom, response to bronchodilators, use of antibiotics, peak expiratory flow (PEF) rate, concomitant medications, and diagnosis, will be collected and inputted into the CAMCDS. The accuracy of the CAMCDS model will be evaluated by using the receiver operating characteristic (ROC) curve compared with the results of pediatrician's diagnosis. The evaluation indexes, including accuracy, precision, sensitivity, specificity, false-positive rate (FPR), false-negative rate (FNR), and area under the curve (AUC), will also be

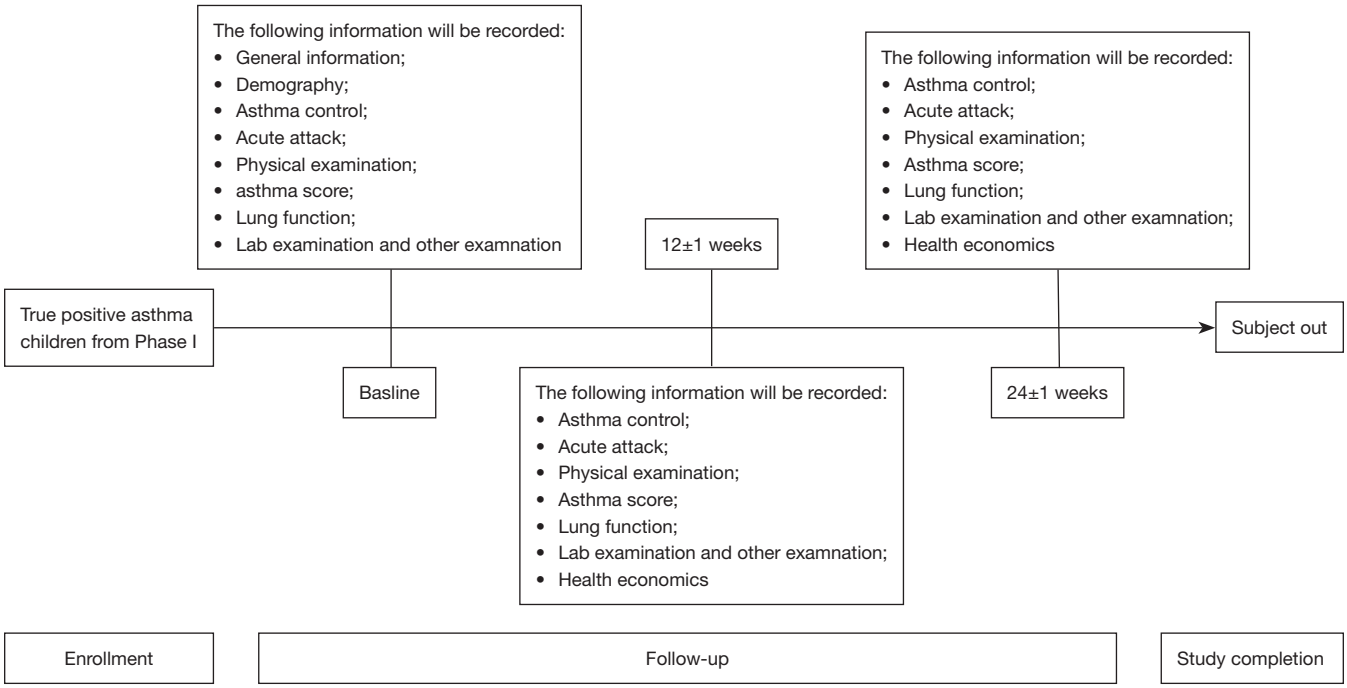

Figure 2 Study schema of Phase II.

calculated.

Phase II of the study aims to examine the treatment predicting accuracy of the CAMCDS model. In phase II, children diagnosed as true positive asthma in phase I will be followed up 12±1 and 24±1 weeks after baseline. The study flowchart for phase II is shown in *Figure 2*. Information such as asthma symptom score, use of antibiotics, PEF rate, and concomitant medications of all children will be collected during follow-up. Treatment predicting accuracy of the CAMCDS model will be assessed by calculating the matching degree between clinical prescription and CAMCDS prescription.

Study participants

All participants from each site who fulfill the following inclusion and exclusion criteria will be consecutively invited to attend the study.

Participants eligible for this study must meet all of the following criteria:

(I) Age ≤14 and ≥4 years.

(II) Related complaints of asthma-like symptoms such as recurrent wheezing, coughing, shortness of breath, and chest distress within 3 months.

(III) Children diagnosed with stable asthma: if children meet one of the following criteria according to the medical history within 6 months, they will be classified as stable asthma.

(i) After anti-inflammatory treatment (such as inhaled glucocorticoids and/or anti-leukotriene drugs) for 4–8 weeks, the increase of forced expiratory volume in one second (FEV_1) ≥12% (11).

(ii) Variation rate of the PEF (continuous monitoring for 2 weeks) ≥13%.

Suspected asthma: participants not diagnosed with asthma before enrollment (13).

(IV) One of the caregivers signs the informed consent form; also for children, if with decision-making ability (aged >10 years) (16).

Participants meeting any of the following criteria must not be enrolled in the study:

(I) Children or caregivers of children having problems in expression, understanding, reading, or writing;

(II) Any other participants not suitable to participate in the project judged by investigators of this study, such as participants who are not able to cooperate with the pulmonary function test.

All participants fulfilling the inclusion/exclusion criteria, including those who do not attend the study should be documented in the screening log. The reasons provided by

(suspected) participants who are not willing to be enrolled in the study should also be recorded in the screening log for further enrollment bias analysis.

CAMCDS model

The CAMCDS is established based on authoritative guidelines, pediatricians' experience, and mature model construction technology. The Bronchial Asthma in Children Guideline for Its Diagnosis and Treatment (2016) (11) and the Global Strategy for Asthma Management And Prevention (GINA, 2019) (17) guidelines will be adopted. Pediatricians' experiences are based on the correctness of knowledge guaranteed by respiratory experts in Children's Hospital of Shanghai. The CAMCDS model consists of two portions: diagnosis model and clinical pathway.

The diagnosis model is based on the eXtreme Gradient Boosting (XGBoost) algorithms (18), which is driven by the data that contain participants' complaints, physical signs, past medical history, family history, examination information and diagnosis results. The feature engineering based on data is necessary in the XGBoost algorithms. Three kinds of feature engineering exist in the diagnosis model. The first is to standardize the continuous data so as to standardize input data into the same dimension system so that the convergence of the model can be ensured. The second is feature selection, which aims to find out the features that may contribute most to the model and screen out similar features. The main methods include variance analysis, correlation analysis and mutual information method. The third is feature dimensionality reduction, which aims to compress and reconstruct multi-dimensional features and reduce feature redundancy. During the training process of the XGBoost algorithms, the strategy to find the best parameters is k-fold cross validation (19), which divides the total data set into training set, verification set and test set. The details of diagnosis model are shown in *Figure 3*.

For the clinical pathway, knowledge modeling is mainly used, which can be divided into three stages: knowledge management, knowledge representation, and knowledge reasoning.

Knowledge management refers to the collection and management of relevant data. To treat children's asthma, we referred to authoritative guidelines at home and abroad (11,17). We extracted key recommendations from the guidelines as a framework for treatment. To ensure the authority and effectiveness of treatment recommendations, we invited experts from top level hospitals to provide and refine the recommendations according to their professional opinions. When the guidelines were inconsistent with clinical practice, we adjusted the plan according to local clinical practice.

Knowledge representation is a method to describe knowledge or a set of conventions of knowledge. A data structure that can be accepted by a computer to describe knowledge is the basis of knowledge recognition in the area of artificial intelligence. Firstly, we described the reasoning process of treatment methods with the formation of a decision tree, which similar to the decision-making thinking process of professional doctors. Then, we constructed production rules, to translate the graphical decision tree into language that computer can recognize. Finally, we used inference engine to carry out knowledge reasoning. An inference engine is a module used to complete inference functions in the application system. The inference engine generally consists of three parts: scheduler, actuator, and consistency coordinator. After all the above steps, five types of treatments are considered for asthma control: short-acting beta-2 agonists (SABAs); short-acting muscarinic antagonists (SAMA); inhaled corticosteroids (ICS); oral corticosteroids (OCS) and oxygen therapy. The details of Clinical pathway are shown in *Figure 4*.

Data collection

The data collection activities are also divided into two phases. For phase I, the baseline information and diagnosis results will be collected. For phase II, the non-asthma children in phase I will be considered as "complete the study", whereas children diagnosed with true positive asthma will be considered "on study", until the withdrawal of consent, loss to follow-up, or study termination/closure. The more detailed variables that will be collected for the study are shown in Table S1 and Appendix 1.

As this is an observational study, treatment and care procedures will be at the clinician's discretion following routine care practice and not dictated by the protocol (*Figures 3*,*4* and https://cdn.amegroups.cn/static/public/jtd-21-668-1.docx). Protocol-defined data will be collected on sites (baseline and the second visit) or by telephone (the first visit). All of the data will be collected by trained study nurses via case report forms and will be manually entered into the CDSS. For aggregation, every site will get an analysis script to run on their data that produces an aggregated output that will be collected at Children's Hospital of Shanghai for the final analysis. M.M. Na Dong (Department

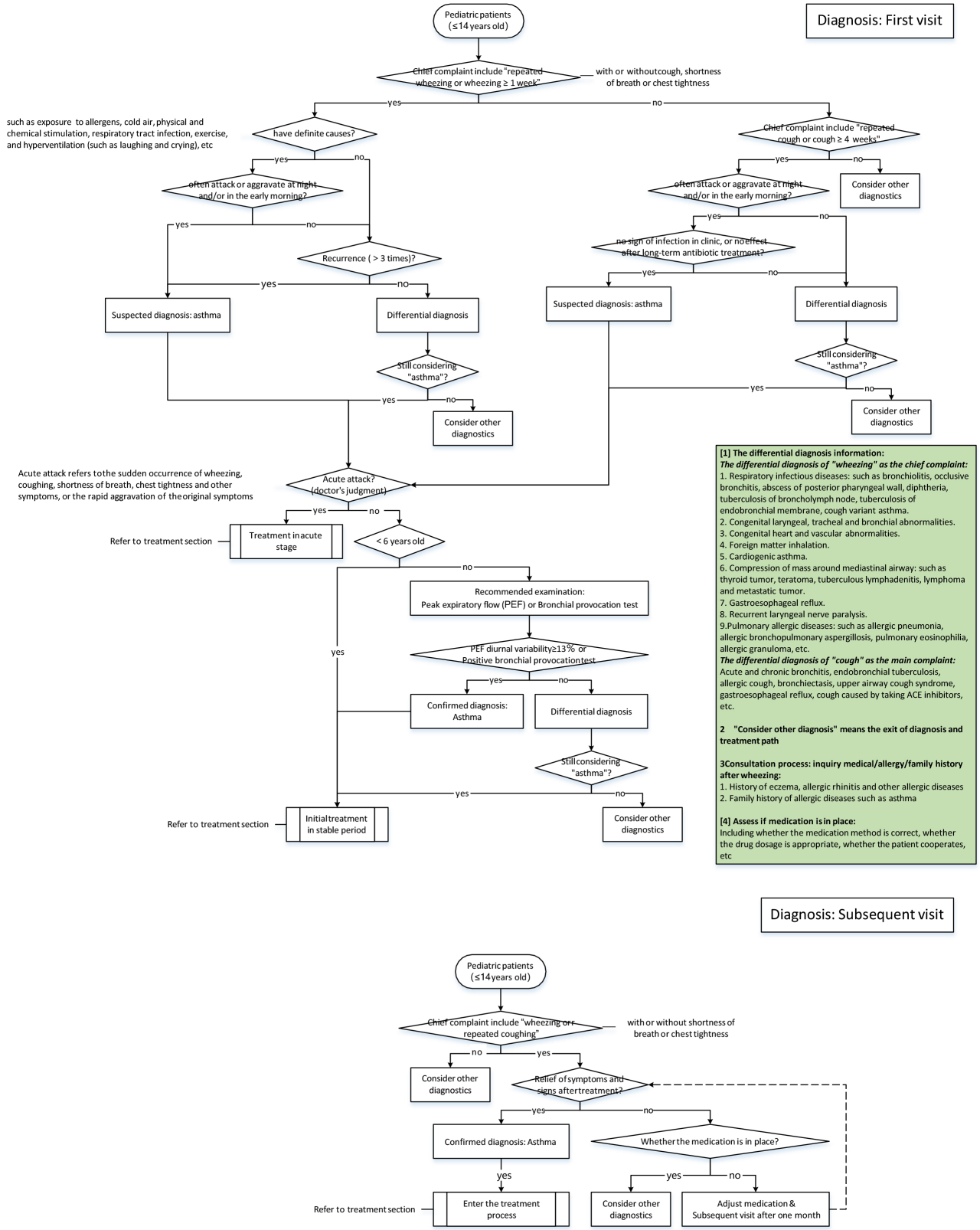

Figure 3 Diagnosis model of CAMCDS. CAMCDS, Childhood Asthma Model for Clinical Decision Support.

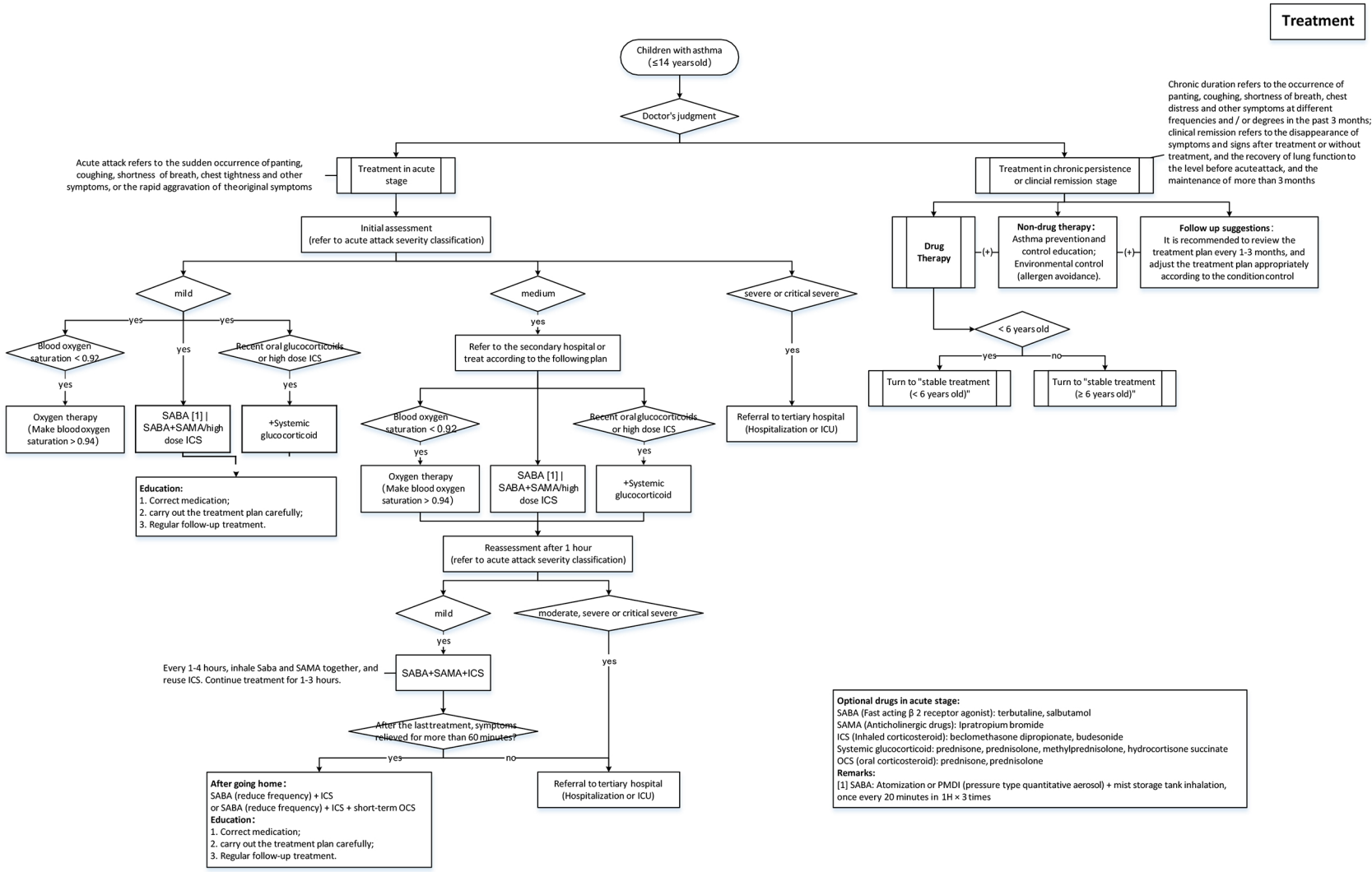

Figure 4 Clinical pathway of CAMCDS. CAMCDS, Childhood Asthma Model for Clinical Decision Support.

of Respiration, Children's Hospital of Shanghai) will be responsible for data collecting and gathering. While M.M. Beirong Wu (Department of Respiration, Children's Hospital of Shanghai) will be responsible for data analysis and interpretation.

Statistical analysis

Sample size

We estimated the sample size based on the following assumptions:

(I) In the accuracy evaluation of the auxiliary diagnosis, it is assumed that the sensitivity of CDSS diagnosis is 95%, the tolerance error is 2%, the specificity is 85%, and the tolerance error is 5%. In the case of 0.05 significance level and 95% confidence level, considering that 20% of the data are invalid or lost to follow-up, the required sample size of asthma children is 817.

(II) Considering that children with diagnosed asthma accounting for 60% of children with asthma like symptoms (20), the total screened sample size comes to 1,362.

Data presentation methods

All variables will be analyzed descriptively and be presented as number of children, mean, standard deviation, median, minimum and maximum for continuous variables, and frequency and percentages for categorical variables. The calculation will be based on non-missing data unless otherwise specified. Missing data will not be imputed but will be reported. All statistical tests will be at the 0.05 significant level, two-sided, and a 95% CI will be constructed wherever applicable. Statistical analyses will be conducted by Statistical Analysis System (SAS®, version 9.3).

Categorical data (including binary data) will be summarized by presenting the rate and 95% CI according to Pearson-Clopper. Logistic regression adjusting confounders will also be considered. Continuous data will be summarized using mean, standard deviation, coefficient of variation, geometric mean, median, and minimum and maximum. The linear regression model adjusting confounders will also be considered.

Data sets

In phase I, all recruited participants who completely fulfill the inclusion/exclusion criteria will be included in the primary analysis population. In phase II, participants who received at least one follow-up will be included in the primary analysis population. Other analysis populations may be defined based on more restrictive criteria, such as the fulfillment of eligibility criteria or a minimum duration of the observation period.

Full analysis set (FAS) is defined as all children from the database who fulfill the inclusion/exclusion criteria will be included in the primary analysis population. FAS will be used for all analyses.

Analysis of accuracy

The accuracy of diagnosis and treatment prediction of CDSS model will both be based on the comparison between ROC curve and the results of pediatrician's diagnosis and prescriptions. The definition of asthma diagnosis and treatment were from the recommendations derived by a panel of advisors including pediatricians from China after consideration of international guidelines (10,11). Each clinical diagnosis and treatment will be determined by an adjudication panel comprising three consultant paediatric clinicians (median 10 years of specialist practice). Two members will review each subject independently, with a third member acting as tie-breaker in the event of non-agreement. The panel will arrive at diagnoses and prescriptions after assessment of all available clinical data. There will be three outcomes: "YES", "NO" or "UNSURE". The outcome of "UNSURE" indicates that the case is not entirely met due to lack of information and these cases will be excluded from the endpoint (21). After setting up the real diagnosis cases and definition of evaluation indexes including accuracy, precision, sensitivity, specificity, the FPR, and AUC, the CDSS model can be considered better if the accuracy, precision, sensitivity, specificity, and AUC are higher and FPR is lower. The accuracy of the recommended diagnosis refers to its consistency with the diagnosis made by pediatricians. If the CDSS recommended diagnosis and prescriptions are consistent with clinicians' decisions, the record will be flagged as positive. If not, the record will be flagged as negative. The measurements of the analysis are shown in *Table 1*.

The definition and calculation formula of relevant indexes are as follows: (I) sensitivity is the proportion of diseased participants correctly identified: Se = a/(a + b); (II) specificity is the proportion of healthy participants correctly identified: Sp = d/(c + d); (III) Youden's J statistic. J = Se + Sp − 1; (IV) PPV is the probability that participants with a positive screening test truly have the disease, PPV = a/(a + c); (V) NPV is the probability that participants with a negative screening test truly

Table 1 Classification of diagnosis and prescription results of CAMCDS *vs.* pediatricians' responses

Pediatricians	Result based on the index of CAMCDS		
	Positive/Yes	Negative/No	Total
Positive/yes	a (TP)	b (FN)	a + b
Negative/no	c (FP)	d (TN)	c + d
Total	a + c	b + d	N

CAMCDS, Childhood Asthma Model for Clinical Decision Support; TP, true positive; FN, false negative; FP, false positive; TN, true negative.

don't have the disease, NPV =d/(b + d); (VI) False positive rate (FPR) (α) = type I error = 1 – Sp = c/(c + d); (VII) False negative rate (FNR) (β) = type II error = 1 – Se = b/(a + b); (VIII) False discover rate (FDR) = c/(a + c); (7) Accuracy = (a + d)/(a + b + c + d); (IX) Positive likelihood ratio = Se/(1 – Sp); (X) Negative likelihood ratio = (1 – Se)/Sp.

Discussion

The CAMCDS is the first developed model developed for clinical decision support in Chinese children with asthma. The present study is a real-world study that aims to test the accuracy of the CAMCDS. The multi-site design of this study can helped ensure the recruitment of children with asthma from a representative regional population.

In this study, we choose XGBoost algorithms to model asthma diagnoses and treatment because it is an advanced implementation of a gradient-boosting decision-tree algorithm and has been used in a few studies to predict asthma hospital visits (22-25). Although there is no study comparing the prediction accuracy of XGBoost algorithm and gradient boosting decision tree algorithm in predicting the diagnosis and treatment of asthma, a retrospective analysis compared the differences of accuracy in predicting hospitalization risk among different machine learning algorithms. The area under curves for each model were: logistic regression 0.82 (95% CI: 0.81–0.82), random forests 0.82 (95% CI: 0.81–0.83), and gradient boosting machines 0.85 (95% CL 0.84–0.86), which showed that gradient boosting machines model was the most successful at predicting need for hospitalization at the time of triage in pediatric children presenting with asthma exacerbation (23). XGBoost is advantageous because of its high speed and well performance, making it dominant in applied machine learning for structured data. XGBoost can also offers regularized gradient boosting and feature importance scores using a trained predictive model for feature selection (8,22).

We have designed several methods to minimize the potential bias. Firstly, we will apply a data cleaning process when the system is launched in each hospital. Data cleaning refers to the correction or amelioration of data problems, including missing values, incorrect or out-of-range values, or responses that are logically inconsistent with other responses in the database. While all registries strive for "clean data", in reality, this is a relative term. In this analytical study, the data validation rules will be applied before analysis. Missing values will be traced to the source from children' medical records if applicable. A data management report includes processes of logical checks for out-of-range values and explains how missing values and values that are logically inconsistent. Secondly, we have prepared the solutions for missing data. For baseline information, simple methods such as the complete patient analysis method will be used. Multiple imputation methods will be applied to compensate for missing data. Finally, we have designed the current multi-centre study with opinions and insights from experienced clinical experts. This will help to narrow the gaps between guidelines and real clinical practice.

However, several uncontrolled limitations may exist in the database. First of all, all of the participants are from the sites in Shanghai region and hence the results may not represent the whole of China. Then, loss to follow-up or attrition of children may threaten the generalizability. Moreover, the bias in the process of design and operation may interfere data clean.

Acknowledgments

Funding: This study has been supported by the major new drug development program of the 13th Five-year Plan (No. 2017ZX09201002-002); and the key area R&D program of Guangdong province (No. 2019B020204001).

Footnote

Reporting Checklist: The authors have completed the SPIRIT reporting checklist. Available at https://dx.doi.org/10.21037/jtd-21-668

Peer Review File: Available at https://dx.doi.org/10.21037/jtd-21-668

Conflicts of Interest: All authors have completed the ICMJE

uniform disclosure form (available at https://dx.doi.org/10.21037/jtd-21-668). Canghong Zhi and Dandan Zhao report funding from major new drug development program of the 13th Five-year Plan (No.2017ZX09201002-002) and key area R&D program of Guangdong province (No.2019B020204001) for this study. The other authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. This study was approved by the Ethics Review Committee of Children's hospital of Shanghai (NO. 2021R009-E01).

References

1. Akinbami OJ. The State of Childhood Asthma: United States, 1980-2005. US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics, 2006.
2. Akinbami LJ, Moorman JE, Garbe PL, et al. Status of childhood asthma in the United States, 1980-2007. Pediatrics 2009;123 Suppl 3:S131-45.
3. Akinbami LJ, Moorman JE, Bailey C, et al. Trends in asthma prevalence, health care use, and mortality in the United States, 2001-2010. NCHS Data Brief 2012;(94):1-8.
4. O'Byrne PM, Pedersen S, Lamm CJ, et al. Severe exacerbations and decline in lung function in asthma. Am J Respir Crit Care Med 2009;179:19-24.
5. O'Brian AL, Lemanske RF Jr, Evans MD, et al. Recurrent severe exacerbations in early life and reduced lung function at school age. J Allergy Clin Immunol 2012;129:1162-4.
6. National Pediatric Asthma Prevention and Control Group. A nationwide surrey on the prevalence of asthma among 0–14 year old population in China (1988~1990). Chinese Journal of Tuberculosis and Respiratory Diseases 1993;16:64-8.
7. Chen YZ; National Cooperation Group On Childhood Asthma. A nationwide survey in China on prevalence of asthma in urban children. Zhonghua Er Ke Za Zhi 2003;41:123-7.
8. Chinese Center for Disease Control and Prevention. Third nationwide survey of childhood asthma in urban areas of China. Zhonghua Er Ke Za Zhi 2013;51:729-35.
9. National Asthma Education and Prevention Program, Third Expert Panel on the Diagnosis and Management of Asthma. Expert Panel Report 3: Guidelines for the Diagnosis and Management of Asthma. Bethesda (MD): National Heart, Lung, and Blood Institute (US); 2007 Aug. Available online: https://www.ncbi.nlm.nih.gov/books/NBK7232/
10. Global Initiative for Asthma. Global Strategy for Asthma a Management and Prevention, 2017. Available online: www.ginasthma.org
11. Subspecialty Group of Respiratory Diseases, Society of Pediatrics, Chinese Medical Association; Editorial Board, Chinese Journal of Pediatrics. Guideline for the diagnosis and optimal management of asthma in children(2016). Zhonghua Er Ke Za Zhi 2016;54:167-81.
12. Fiks AG, Mayne S, Karavite DJ, et al. A shared e-decision support portal for pediatric asthma. J Ambul Care Manage 2014;37:120-6.
13. Toll DB, Janssen KJ, Vergouwe Y, et al. Validation, updating and impact of clinical prediction rules: a review. J Clin Epidemiol 2008;61:1085-94.
14. Daines L, McLean S, Buelo A, et al. Clinical prediction models to support the diagnosis of asthma in primary care: a systematic review protocol. NPJ Prim Care Respir Med 2018;28:15.
15. Zhang Y, Huang L, Zhou X, et al. Characteristics and Workload of Pediatricians in China. Pediatrics 2019;144:e20183532.
16. National Medical Products Administration. Technical Guidelines for Population Drug Clinical Trialsin Pediatrics. Available online: https://www.nmpa.gov.cn/index.html
17. Global Initiative for Asthma. Global Strategy for Asthma a Management and Prevention, 2019. Available online: www.ginasthma.org
18. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. 2016.
19. Wiens TS, Dale B, Boyce MS, et al. Three way k-fold cross-validation of resource selection functions. Ecol Modell 2008;212:244-55.

20. de Jong CCM, Pedersen ESL, Mozun R, et al. Diagnosis of asthma in children: the contribution of a detailed history and test results. Eur Respir J
21. Porter P, Abeyratne U, Swarnkar V, et al. A prospective multicentre study testing the diagnostic accuracy of an automated cough sound centred analytic system for the identification of common respiratory disorders in children. Respir Res 2019;20:81.
22. Nam SM, Peterson TA, Seo KY, et al. Discovery of Depression-Associated Factors From a Nationwide Population-Based Survey: Epidemiological Study Using Machine Learning and Network Analysis. J Med Internet Res 2021;23:e27344.
23. Patel SJ, Chamberlain DB, Chamberlain JM. A Machine Learning Approach to Predicting Need for Hospitalization for Pediatric Asthma Exacerbation at the Time of Emergency Department Triage. Acad Emerg Med 2018;25:1463-70.
24. Tong Y, Messinger AI, Luo G. Testing the Generalizability of an Automated Method for Explaining Machine Learning Predictions on Asthma Patients' Asthma Hospital Visits to an Academic Healthcare System. IEEE Access 2020;8:195971-9.
25. Brownlee J. XGBoost With Python: Gradient Boosted Trees with XGBoost and scikit-learn. Machine Learning Mastery; 2018.

Cite this article as: Dong N, Wu B, Yin B, Dong W, Jin X, Wang M, Xu X, Zhi C, Zhao D, Lu M, Gu H, Qiao R. Validation of the accuracy of the childhood asthma model for clinical decision support: a study protocol. J Thorac Dis 2021;13(10):6052-6061. doi: 10.21037/jtd-21-668

Appendix 1 Data collection elements

The following data will be collected at baseline for all participants:

1. Demographic data: date of birth, gender, ethnicity, medical insurance type, and kindred relationship of caregiver; education status of caregiver, family economic status, occupation of caregiver and residency.
2. Disease inducement.
3. asthma-like symptom: if there any wheezing, shortness of breath, cough, chest tightness/chest pain, nasal obstruction, nose itch, runny nose, dyspnea, attack or aggravation at night and/or early morning within the four weeks before enrollment and their frequency.
4. Response to bronchodilators: such as if asthma can be relieved by aerosol inhalation controller or reliever drugs.
5. Family history of asthma.
6. History of asthma.
7. Other history of illness: history of respiratory infection, sinusitis, obstructive sleep apnea, gastroesophageal reflux.
8. History of lung surgery.
9. History of allergy: eczema, allergic rhinitis, etc.
10. Acute attack information: such as frequency of hospitalization or emergency room admission due to acute asthma-like symptom within previous 12 weeks.
11. Physical examination and signs: height, weight, pulse rate, respiratory rate, body temperature, cyanosis, wet rales, three concave signs, wheezing sound, anxious/fidgety, shortness of breath.
12. Spirometry examination.
13. Protocol specified laboratory tests.
14. Protocol specified imaging examination.
15. Diagnosis.
16. Asthma-related drugs in the previous 12 weeks.
17. Other concomitant medication.

The following data will be collected for participants diagnosed as stable asthma:

1. Asthma control: defined according to *Bronchial Asthma in Children Guideline for Its Diagnosis and Treatment (2016)*.
2. Asthma symptom score: children aged 4–11 years old will be assessed by the C-ACT score, while the older children (12–14 years of age) will be assessed by ACT score.

The following data will be collected for 1st follow-up (12±1 weeks):

1. Asthma-like symptom: if there any wheezing, shortness of breath, cough, chest tightness/chest pain, nasal obstruction, nose itch, runny nose, dyspnea, attack or aggravation at night and/or early morning within the four weeks before enrollment and their frequency.
2. Asthma control: stage of asthma (acute attack stage/chronic duration stage/clinical remission stage), control level classification [defined according to *Bronchial Asthma in Children Guideline for Its Diagnosis and Treatment (2016)*].
3. Acute attack information: such as frequency of hospitalization or emergency room admission due to acute asthma-like symptom within previous 12 weeks.
4. Asthma-related drugs in the previous 12 weeks.*
5. Asthma symptom score: children aged 4–11 years old will be assessed by the C-ACT score, while the older children (12–14 years of age) will be assessed by ACT score.
6. Medical cost for the diagnosis and treatment of asthma.
7. Study discontinuation status.

https://dx.doi.org/10.21037/jtd-21-668

The following data will be collected for 2nd follow-up (24±1 weeks):

1. Asthma-like symptom: if there any wheezing, shortness of breath, cough, chest tightness/chest pain, nasal obstruction, nose itch, runny nose, dyspnea, attack or aggravation at night and/or early morning within the four weeks before enrollment and their frequency.
2. Asthma control: stage of asthma (acute attack stage/chronic duration stage/clinical remission stage), control level classification [defined according to *Bronchial Asthma in Children Guideline for Its Diagnosis and Treatment (2016)*].
3. Acute attack information: such as frequency of hospitalization or emergency room admission due to acute asthma-like symptom within previous 12 weeks.
4. Physical examination and signs: height, weight, pulse rate, respiratory rate, body temperature, cyanosis, wet rales, three concave sign, wheezing sound, anxious/fidgety, shortness of breath.
5. Spirometry examination.
6. Protocol specified laboratory tests.
7. Protocol specified imaging examination.
8. asthma-related drugs in the previous 12 weeks.*
9. Asthma symptom score: children aged 4–11 years old will be assessed by the C-ACT score, while the older children (12–14 years of age) will be assessed by ACT score.
10. Medical cost for the diagnosis and treatment of asthma.
11. Study discontinuation status.

*, treatment change include but not limit within the following situations: drug dosing change, drug administration schedule change, change drug (drug brand change is not included), add new treatment.

In addition to the data elements mentioned above, any AE/SAE reported at any time will be recorded.

https://dx.doi.org/10.21037/jtd-21-668

Table S1 Data collection plan

Data collection	Baseline[a]	1st Visit	2nd Visit
Window of visits (week ± week)	0±1	12±1	24±1
Informed consent	X		
Demographics	X		
Disease inducement	X		
Asthma-like symptom	X	X	X
Acute attack information	X	X	X
Physical examination and signs	X		X
Asthma-related drugs in the previous 12 weeks	X	X	X
Response to bronchodilators	X		
Family history of asthma	X		
History of asthma	X		
History of lung surgery	X		
History of allergy	X		
Other past history of illness	X		
Spirometry examination	X		X
Protocol specified laboratory tests	X		X
Protocol specified imaging examination	X		X
Diagnosis	X		
Concomitant medication	X	X	X
Asthma control[b]	X	X	X
Asthma symptom score[b]	X	X	X
AEs and SAEs		X	X
Cost of diagnosis and treatment of asthma (if applicable)		X	X
Study discontinuation status		X	X

[a], to help assess the rates of participation, sites should maintain a log of the number of eligible participants who decline to participate in the study. [b], the information of participants with diagnosed stable asthma in Phase I need to be collected. AE, adverse event; SAE, severe adverse events.

https://dx.doi.org/10.21037/jtd-21-668