



Identification and validation of asthma phenotypes in Chinese population using cluster analysis

Lei Wang, MD^{*}; Rui Liang, MD^{*}; Ting Zhou, MD[†]; Jing Zheng, MD^{*}; Bing Miao Liang, MD[‡]; Hong Ping Zhang, MD^{*}; Feng Ming Luo, MD[‡]; Peter G. Gibson, MBBS[§]; Gang Wang, MD, PhD^{*}

^{*} Pneumology Group, Department of Integrated Traditional Chinese and Western Medicine, State Key Laboratory of Biotherapy of China, West China Hospital, Sichuan University, Chengdu, People's Republic of China

[†] Health Management Center, West China Hospital, Sichuan University, Chengdu, People's Republic of China

[‡] Department of Respiratory and Critical Care Medicine, West Chia Hospital, Sichuan University, Chengdu, People's Republic of China

[§] Department of Respiratory and Sleep Medicine, John Hunter Hospital, Hunter Medical Research Institute, University of Newcastle, Priority Research Centre for Healthy Lungs, New Lambton Heights, New South Wales, Australia

ARTICLE INFO

Article history:

Received for publication March 31, 2017.

Received in revised form July 9, 2017.

Accepted for publication July 12, 2017.

ABSTRACT

Background: Asthma is a heterogeneous airway disease, so it is crucial to clearly identify clinical phenotypes to achieve better asthma management.

Objective: To identify and prospectively validate asthma clusters in a Chinese population.

Methods: Two hundred eighty-four patients were consecutively recruited and 18 sociodemographic and clinical variables were collected. Hierarchical cluster analysis was performed by the Ward method followed by *k*-means cluster analysis. Then, a prospective 12-month cohort study was used to validate the identified clusters.

Results: Five clusters were successfully identified. Clusters 1 (*n* = 71) and 3 (*n* = 81) were mild asthma phenotypes with slight airway obstruction and low exacerbation risk, but with a sex differential. Cluster 2 (*n* = 65) described an “allergic” phenotype, cluster 4 (*n* = 33) featured a “fixed airflow limitation” phenotype with smoking, and cluster 5 (*n* = 34) was a “low socioeconomic status” phenotype. Patients in clusters 2, 4, and 5 had distinctly lower socioeconomic status and more psychological symptoms. Cluster 2 had a significantly increased risk of exacerbations (risk ratio [RR] 1.13, 95% confidence interval [CI] 1.03–1.25), unplanned visits for asthma (RR 1.98, 95% CI 1.07–3.66), and emergency visits for asthma (RR 7.17, 95% CI 1.26–40.80). Cluster 4 had an increased risk of unplanned visits (RR 2.22, 95% CI 1.02–4.81), and cluster 5 had increased emergency visits (RR 12.72, 95% CI 1.95–69.78). Kaplan-Meier analysis confirmed that cluster grouping was predictive of time to the first asthma exacerbation, unplanned visit, emergency visit, and hospital admission (*P* < .0001 for all comparisons).

Conclusion: We identified 3 clinical clusters as “allergic asthma,” “fixed airflow limitation,” and “low socioeconomic status” phenotypes that are at high risk of severe asthma exacerbations and that have management implications for clinical practice in developing countries.

© 2017 American College of Allergy, Asthma & Immunology. Published by Elsevier Inc. All rights reserved.

Reprints: Gang Wang, MD, PhD, Pneumology Group, Department of Integrated Traditional Chinese and Western Medicine, West China Hospital, Sichuan University, Chengdu 610041, People's Republic of China; E-mail: wucms-respiration@hotmail.com.

Drs Lei Wang and Rui Liang contributed equally to this article.

Disclosures: Authors have nothing to disclose.

Funding Sources: This study was supported by the National Natural Science Foundation of China (81370122, 81570023, and 81670023), the Science and Technology Foundation of Sichuan Province (2015SZ0121), the Science and Technology Foundation of Chengdu (2014-HM01-00294-SF), the Youth Science Funding of Sichuan University (2011SCU04B17), and the Program for New Century Excellent Talents in University (NCET-12-0380).

<http://dx.doi.org/10.1016/j.anai.2017.07.016>

1081-1206/© 2017 American College of Allergy, Asthma & Immunology. Published by Elsevier Inc. All rights reserved.

Introduction

Asthma is a common, chronic respiratory disease affecting 1% to 18% of the population in different countries.¹ Uncontrolled and severe asthma results in not only poor health outcomes but also increased economic burden.² Effective clinical care is complicated by heterogeneity in clinical manifestations, physiologic responses, and pathologic and molecular abnormalities that are associated with uncontrolled and severe asthma. The identification of asthma phenotypes might improve our understanding of the disease pathogenesis and could facilitate the use of targeted therapies.

Recently, the use of statistical approaches has been of great benefit to identify phenotypes using an unbiased approach. One

approach is cluster analysis, which categorizes data objects based on their features and their relationships.³ This method explores how objects within one group are similar or related to one another. These groupings or clusters are described as a “phenotype” when clinical relevance is demonstrated. Several recently published studies have described an asthma phenotype with early onset of diseases and atopy and a phenotype showing more severe disease with late onset and persistent symptoms regardless of therapy. Most of these cluster analyses are cross-sectional analyses of a group of patients as they look today without an analysis of how they will look in the future. Furthermore, most participants are derived from clinical trials rather than a real-world population, and studies based on cluster analysis involving the Chinese population have not been published to date.

A phenotype is defined as the “observable properties of an organism that are produced by the interactions of the genotype and the environment.”⁴ Thus, the phenotype of an individual with asthma is identified by clinical presentation rather than a single

blood test result. Currently, the goal of asthma treatment is to achieve optimal asthma control including symptom control and future risk of asthma exacerbations.¹ In this study, we hypothesized there would be several clusters based on clinical expression in a real-world Chinese population with asthma. We used a 2-step cluster analysis to identify clusters and then a longitudinal cohort study to assess whether the clusters determined at baseline would predict asthma exacerbations and to determine which clusters were likelier than others to experience more asthma exacerbations during 1 year of follow-up.

Methods

Design Overview

This study was a post hoc analysis of our previously published cohort study⁵ (ChiCTR-CH-10000819), which consisted of 2 parts (Fig 1). Study 1 was a cross-sectional survey to consecutively collect the demographic and clinical features of patients with asthma to

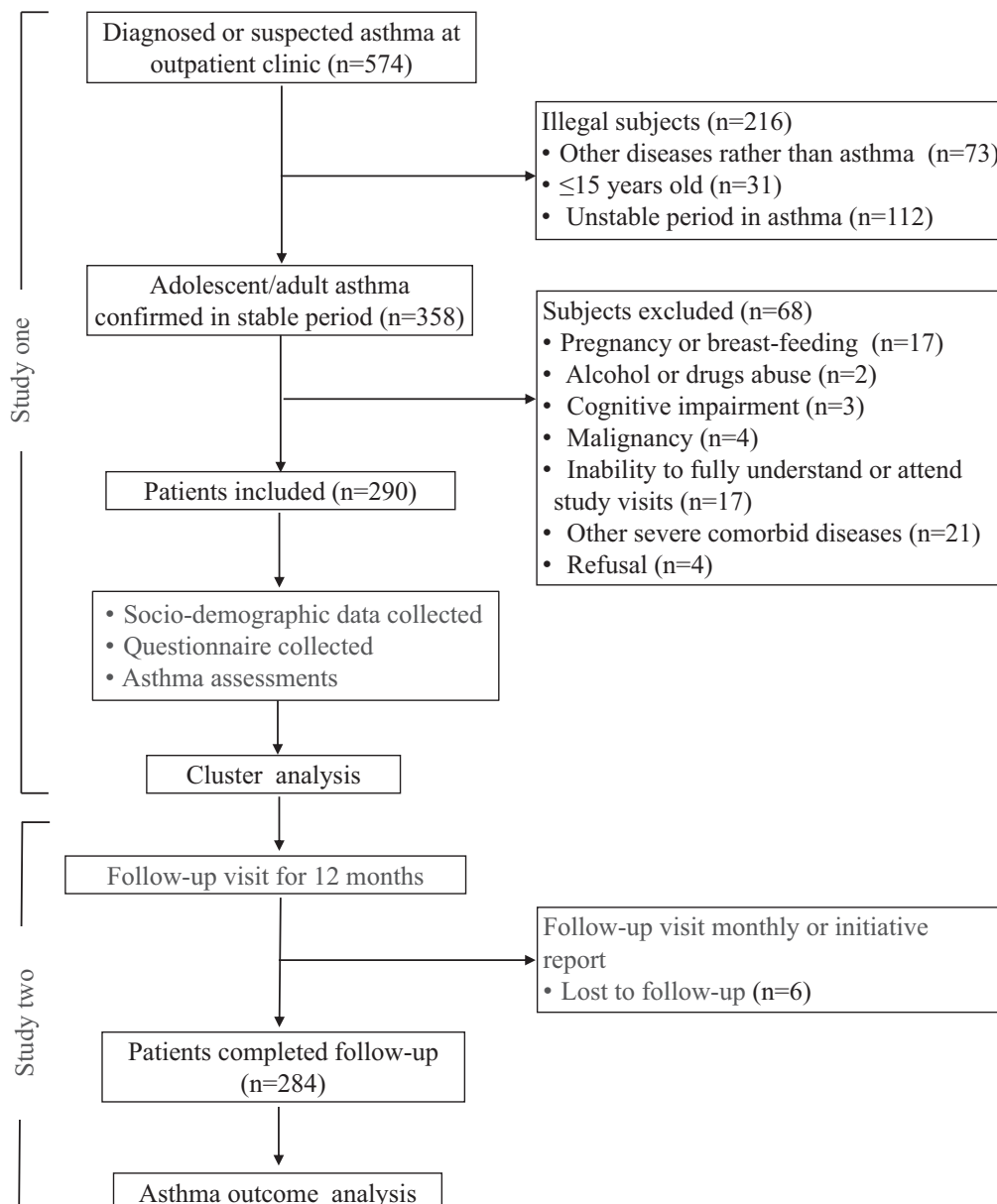


Figure 1. Flowchart of study recruitment and the number of individuals at baseline in study 1 for cluster analysis and in the observational longitudinal research in study 2 for validation.

identify asthma clusters using a 2-step cluster analysis. Study 2 was a prospective noninterventional cohort study with a 12-month follow-up period that was designed to validate these clusters in relation to future risk of asthma exacerbations in a real-world setting of patients with asthma from the previous cross-sectional survey. Treatments received by patients during the study were determined by their treating physicians' standard practices and discretion. The institutional review board at West China Hospital, Sichuan University (Chengdu, China) reviewed and approved this study (ChiECRCT-2010004).

Subjects

Subjects with suspected asthma and at least 15 years old were consecutively recruited at the asthma clinic of the West China Hospital from September 2007 to October 2009. Asthma was diagnosed according to American Thoracic Society (ATS) and Global Initiative for Asthma (GINA) guidelines based on current (within the previous 12 months) episodic symptoms, physician's diagnosis, and confirmed evidence of airway hyperresponsiveness or at least 12% improvement in forced expiratory volume in 1 second (FEV₁) after bronchodilator. To be eligible, all patients were required to be clinically stable (no lower respiratory tract infection, recent asthma exacerbation, or change in maintenance therapy in the previous 4 weeks).⁶ Subjects were excluded if they were pregnant or breastfeeding, had cognitive impairment, had current solid organ malignancy, or could not fully understand the questionnaires and/or attend study visits. All participants gave written informed consents.

Data Collection and Assessments

Data at baseline were collected during a period of stable asthma and included demographics, medications, asthma history, the Asthma Control Test (ACT)^{7–9} questionnaire, and the Asthma Quality of Life Questionnaire (AQLQ),^{10,11} which were translated into and validated in Chinese. Anxiety and depression symptoms were assessed using the 14-item Hospital Anxiety and Depression Scale,¹² which reflects how subjects have felt during the past week and has been validated as screening tool for anxiety and depressive disorders in hospitalized and primary care patients with chronic diseases including asthma. The scores were totaled to produce an anxiety score and a depression score (range 0–21), and a score of at least 8 was used as a cutoff point for identifying anxiety and depression symptoms, with 100% sensitivity and 80% specificity for anxiety and 100% sensitivity and 90% specificity for depression.¹³ Spirometry (SensorMedics 6200, Yorba Linda, California) was performed according to ATS and European Respiratory Society (ERS) standards.¹⁴ Predicted FEV₁ and force vital capacity (FVC) were calculated using data from the Chinese population.¹⁵ Atopy was assessed by skin prick testing with allergen extracts for house dust mites (*Dermatophagoides pteronyssinus*, *Dermatophagoides farinae*), mold (*Alternaria tenuis*, *Aspergillus* species), dog hair, cat hair, pollen (ragweed, birch, London plane), and cockroach, in addition to positive (histamine) and negative (saline) controls.¹⁶

To assess the prognostic value (asthma exacerbations) of these asthma phenotypes based on cluster analysis for predicting asthma exacerbation risk, all included participants entered a cohort study with follow-up visits at the outpatient or asthma clinic or by phone once a month during the following 12 months. During the following 12-month period, based on the ATS and ERS statement of asthma control and exacerbations,¹⁷ asthma exacerbation was identified as the primary outcome, and the secondary outcomes included unplanned visits, emergency visits, hospital admissions, and intensive care unit (ICU) admissions. All research assistants in this study were trained to administer the questionnaires in the same manner.

Cluster and Statistical Analyses

This study collected 18 distinct variables, which are readily assessed in clinical practice. The variables were sex, age, body mass index (BMI), atopy, history of smoking, socioeconomic status (SES; including educational level, health insurance, monthly income status, and employment status), asthma duration, age at asthma onset, asthma severity, FEV₁ (liters), FEV₁/FVC ratio, FEV₁ percentage predicted (FEV₁%predicted), ACT score, AQLQ score, and psychological symptoms. Variables that were clinically redundant were reduced using multicollinearity analysis^{18,19} and reflected certain physiologic parameters of asthma. Subsequently, variables for cluster modeling were selected based on their contribution to characterizations of the asthma phenotype. As a result, 5 continuous variables (age, BMI, ACT, age at asthma onset, and FEV₁% predicted) and 5 categorical variables (sex, psychological symptoms, monthly income status as SES, smoking, and atopy) were selected for cluster analysis.

Continuous and categorical variables were included in the analysis using a 2-step cluster analysis. In the first step, hierarchical cluster analysis using the Ward method generated a cluster feature tree. We used the log-likelihood method to measure the distance between clusters for an estimation of the number of likely clusters among the studied subjects. This estimate was prespecified in a *k*-means cluster analysis that was used as the principal clustering technique. Selected variables to construct the cluster model were standardized using *z* scores for continuous variables and 0 or 1 for categorical variables. Then, the statistical system classified the leaf node of the dendrogram using the agglomerative method and then identified each cluster by the Bayesian information criterion. We also used the noise handling method to avoid missing data caused by a branch of the cluster feature tree that was too large, if present, in step 1.^{18,20,21} In recognition that clustering is dependent on user-defined inputs, we used an iterative approach over a range of clusters (3 clusters up to a maximum of 6 clusters) to define the optimal number of clusters to characterize the data.

All continuous variables were expressed as mean $\pm$ SD, and differences among clusters were analyzed with analysis of variance or with nonparametric statistics if continuous variables were not normally distributed. Categorical data were compared using the χ^2 test or Fisher test. The estimation of future risk across clusters used survival curves, and results are presented as the time to first asthma end-point events. In addition, we used logistic regression models to evaluate the prospective impact of these asthma phenotypes based on cluster analysis on future risk of asthma exacerbations during longitudinal follow-up, followed by calculation of the adjusted relative risk (RR) with 95% confidence intervals (CIs). Cluster and statistical analysis was performed with SPSS 21.0 for IBM Professional (SPSS, Inc, Chicago, Illinois) and STATA 13.0 (Stata Corp LP, College Station, Texas). A 2-sided *P* value less than .05 was considered statistically significant.

Results

Subject Characteristics

The initial dataset included 574 patients who were diagnosed with asthma and suspected asthma at an asthma clinic. After screening using the inclusion and exclusion criteria, 290 eligible subjects were included in the cross-sectional study. After they completed the questionnaire survey and clinical assessment, they entered the 12-month follow-up, and 284 subjects completed the 12-month cohort study (Fig 1). The demographics and clinical characteristics of included subjects are listed in Table 1. Sixty-two percent of participants were women, had a mean age of 39.1 years (SD 12.1), and a BMI of 22.2 kg/m² (SD 3.7), and 25% were current or ex-smokers. Most (55.6%) had atopy, with asthma

Table 1
Sociodemographic and Clinical Characteristics of Subjects With Stable Asthma

Variables	Total cohort (N = 284)
Age (y), mean $\pm$ SD	39.1 $\pm$ 12.1
Women/men	176/108
Atopy, n (%)	156 (55.6)
BMI (kg/m ²), mean $\pm$ SD	22.2 $\pm$ 3.7
Smoking, n (%)	71 (25.0)
Age of asthma onset (y), mean $\pm$ SD	32.9 $\pm$ 13.8
Asthma duration (y), mean $\pm$ SD	6.2 $\pm$ 9.2
Spirometry, mean $\pm$ SD	
FEV ₁ (L)	2.60 $\pm$ 0.84
FEV ₁ (% predicted)	85.48 $\pm$ 21.74
FEV ₁ /FVC (%)	79.79 $\pm$ 15.64
ACT score, mean $\pm$ SD	17.6 $\pm$ 3.1
AQLQ score, mean $\pm$ SD	3.7 $\pm$ 1.2

Abbreviations: ACT, Asthma Control Test; AQLQ, Asthma Quality of Life Questionnaire; BMI, body mass index; FEV₁, forced expiratory volume in 1 second; FVC, forced vital capacity.

duration of 6.2 years (SD 9.1), and mild impairment of lung function, with 85.5% (SD 21.7%) of FEV₁%predicted. Most patients had poorly controlled asthma with ACT scores of 17.6 (SD 3.1) and poor AQLQ scores of 3.7 (SD 1.2), and 70.3% of subjects had psychological states presenting as anxiety and/or depression symptoms. Twenty-eight of these subjects (9.9%) met the GINA criteria for severe asthma.¹

Cluster Profile

A total of 284 consecutive patients with stable asthma entered the 2-step cluster approach outlined in the Methods section. The study showed 5 distinct phenotypic clusters. The silhouette coefficient of clustering quality was 0.2 (range –1 to 1, with the higher score indicating better clustering quality), indicating a satisfactory and better performance than using 3, 4, or 6 clusters. The relative importance of variable predictors in clustering indicated psychological symptoms, sex, smoking, atopy, age, ACT score, monthly income status, FEV₁%predicted, and age at asthma onset were crucial, but obesity was not crucial (Fig 2A). As a result, these 5 clusters and their phenotypic characteristics differed significantly by age, sex, atopy, smoking, onset at asthma onset, FEV₁%predicted, ACT score, monthly income status, psychological symptoms, occupation, and education status but by not by BMI and health insurance (Tables 2 and 3, Fig 2B). Clusters 1 and 3 had a sex differential but had less airway obstruction, which would be “mild asthma phenotypes.” Cluster 2 had a high prevalence of atopy, signifying an “allergic” phenotype. Cluster 4 featured a “fixed airflow limitation” phenotype with smoking, and cluster 5 was a “low SES” phenotype. It was evident that patients in clusters 2, 4, and 5 had a distinctly low SES and more psychological symptoms than clusters 1 and 3.

Cluster 1

Twenty-five percent of subjects (n = 71) were grouped into cluster 1, which was composed entirely of men (100%) and more than half of which were current or ex-smokers (50.7%), had anxiety or depression symptoms (42.3%), and most had a higher SES in monthly income status and employment status. Patients in the cluster 1 had well-controlled asthma with the highest ACT score and the best lung function. This group was composed of the youngest subjects and their age at asthma onset was generally young.

Cluster 2

Cluster 2 consisted of 65 subjects, most of whom were women. All had atopy, were nonsmokers, had anxiety and depression symptoms, and a larger proportion (18.5%) had severe

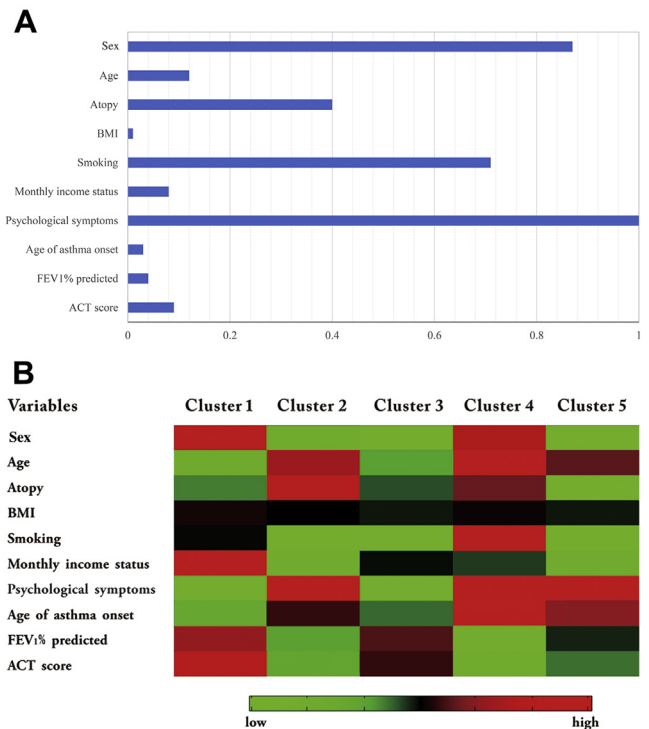


Figure 2. Relative importance of 10 selected variable predictors in (A) clustering and (B) heat map of phenotypic characteristics in 5 clusters that showed statistical differences in normalized variables. Low (green) to high (red) expression of presented variables. Abbreviations: ACT, Asthma Control Test; BMI, body mass index; FEV₁, forced expiratory volume in 1 second.

asthma. In addition, low-income earners accounted for a larger proportion (72.3%).

Cluster 3

Cluster 3 was the largest group (n = 81, 28% of subjects). It consisted of slightly younger patients, all of whom were women, most were nonsmokers, and most had slight airway obstruction. Nearly 50% had an atopic history and psychological symptoms.

Cluster 4

Cluster 4 was the smallest group (n = 33, 12% of subjects). In contrast to cluster 2, all subjects in this group were smokers and men were predominant (75.8%). In this cluster, almost all subjects (97%) had anxiety and depression and they had the worst ACT scores and evident airway obstruction in FEV₁%predicted and FEV₁/FVC%.

Cluster 5

Thirty-four subjects were grouped into cluster 5. This group was composed of women without atopy who were nonsmokers. Low monthly income status accounted for a larger proportion (73.5%) in this cluster. All had anxiety and depression symptoms, the worst ACT scores, and slight airflow obstruction.

Prognostic Value of Identified Asthma Phenotypes Based on Cluster Analysis

During the 12-month follow-up period, the number of asthma exacerbations as the primary outcome and the secondary outcomes such as unplanned visits, emergency visits, and hospital admissions differed statistically among the 5 phenotypes based on cluster analysis (Table 4). Rank conversion of all data in asthma outcomes was conducted because of non-normal distribution, and then 1-way analysis of variance was used. It further indicated that the

Table 2
Sociodemographic Characteristics of 5 Phenotypes Based on Cluster Analysis

Variables	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	P value
Patients, n (%)	71 (25.0)	65 (22.9)	81 (28.5)	33 (11.6)	34 (12.0)	
Age (y), mean $\pm$ SD	34.5 $\pm$ 10.5	42.4 $\pm$ 11.7 ^a	36.5 $\pm$ 11.9 ^b	47.1 $\pm$ 11.3	40.7 $\pm$ 11.9 ^{a,b}	<.0001
Women/men	0/71	53/12	81/0	8/25	34/0	<.0001
Atopy, %	43.7	100.0	48.1	69.7	0.0	<.0001
Body mass index (kg/m ²), mean $\pm$ SD	22.5 $\pm$ 4.2	22.3 $\pm$ 4.1	22.0 $\pm$ 3.4	23.1 $\pm$ 2.4	21.3 $\pm$ 3.3	.346
Smoking, %	50.7	0.0	2.4	100.0	0.0	<.0001
Monthly income status, n (%)						<.0001
Low	23 (32.4)	47 (72.3)	43 (53.1)	22 (66.7)	25 (73.5)	
Moderate	35 (49.3)	18 (27.7)	33 (40.7)	7 (21.2)	8 (3.5)	
High	13 (18.3)	0 (0.0)	5 (6.2)	4 (12.1)	1 (2.9)	
Education levels, n (%)						<.0001
Primary school or less	4 (5.6)	20 (30.8)	9 (11.1)	8 (24.2)	7 (20.6)	
High school	27 (38.0)	37 (56.9)	47 (58.0)	19 (57.6)	22 (64.7)	
at least university degree	40 (56.3)	8 (12.3)	25 (30.9)	6 (18.2)	5 (14.7)	
Employment status, n (%)						<.0001
Unemployed	9 (12.7)	23 (35.4)	17 (21.0)	12 (36.4)	10 (29.4)	
Retired/student/home duties	2 (2.8)	13 (16.0)	10 (12.3)	7 (21.2)	4 (11.8)	
Part time	29 (40.8)	7 (10.8)	16 (19.8)	5 (15.2)	7 (20.6)	
Full time	31 (43.7)	22 (33.8)	38 (46.9)	9 (27.3)	13 (38.2)	
Health insurance, n (%)	41 (57.7)	24 (36.9)	36 (44.4)	16 (48.5)	13 (38.2)	.138

^aP < .05 vs cluster 1.

^bP < .05 vs cluster 2.

number of outcomes such as asthma exacerbation, unplanned visits, emergency visits, and hospital admissions in clusters 2, 4, and 5 was significantly larger than those in cluster 2 or 3 ($P < .0001$ for all comparisons; Table 4), which suggested that clusters 2, 4, and 5 would be at high risk of asthma exacerbations. However, we did not find any difference in primary and secondary exacerbations among clusters 3, 4, and 5. Furthermore, survival analyses in time to first asthma outcome such as exacerbation, unplanned visits, emergency visits, hospital admission, and ICU admission were undertaken among these prespecified clusters. Kaplan-Meier survival curves demonstrated statistical differences in asthma exacerbation, unplanned visits, emergency visits, and hospital admissions ($P < .0001$ for all comparisons; Fig 3) except ICU admissions ($P = .359$). Based on these data, we analyzed the future risks of asthma exacerbations, unplanned visits, emergency visits, hospital admissions, and ICU admissions using logistic regression models (Fig 4). The models indicated that, if cluster 1 was taken as the reference, then cluster 2 had a statistically increased risk of asthma exacerbations (RR 1.13, 95% CI 1.03–1.25), unplanned visits (RR 1.98, 95% CI 1.07–3.66), and emergency visits (RR 7.17, 95%

CI 1.26–40.80), cluster 4 was at higher risk of unplanned visits (RR 2.22, 95% CI 1.02–4.81), and cluster 5 had a higher risk for emergency visits (RR 12.72, 95% CI 1.95–69.78).

Discussion

Identifying asthma phenotypes can facilitate an increased understanding of its pathogenesis and facilitate a more personalized approach to management. To our knowledge, this is the first study to identify asthma phenotypes based on cluster analysis in a real-world setting based on clinical expression in a Chinese population and to further validate the clinical relevance of these clusters using a prospective longitudinal cohort study. We used cluster analysis to show that clinical phenotypes can be defined in patients with asthma based on clinically accessible measurements including sex, age, BMI, monthly income status, atopy, smoking, age at asthma onset, FEV₁%predicted, and asthma control level at baseline. We identified 3 clinical clusters that correspond to recognizable phenotypes of “allergic asthma,” “fixed airflow limitation,” and “low SES.” Each of these phenotypes was at increased future risk of

Table 3
Clinical Characteristics of 5 Phenotypes Based on Cluster Analysis

Variables	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	P value
Patients, n (%)	71 (25.0)	65 (22.9)	81 (28.5)	33 (11.6)	34 (12.0)	
Age at asthma onset (y), mean $\pm$ SD	30.1 $\pm$ 12.8	33.7 $\pm$ 15.2	31.9 $\pm$ 11.7	38.5 $\pm$ 15.1	34.7 $\pm$ 15.4 ^b	.049
Spirometry, mean $\pm$ SD						
FEV ₁ (% predicted)	90.84 $\pm$ 19.77	81.24 $\pm$ 22.27	88.62 $\pm$ 16.63 ^b	74.99 $\pm$ 29.16	85.11 $\pm$ 23.11 ^b	<.0001
FEV ₁ /FVC (%)	84.83 $\pm$ 13.98	76.20 $\pm$ 14.66	83.69 $\pm$ 13.15	68.15 $\pm$ 19.83	78.09 $\pm$ 14.50	.0120
ACT score	19.3 $\pm$ 2.9	16.5 $\pm$ 2.8	18.0 $\pm$ 2.7	15.7 $\pm$ 3.4	16.9 $\pm$ 3.0 ^{a,b}	<.0001
AQLQ score	4.5 $\pm$ 1.2	3.0 $\pm$ 1.2	4.0 $\pm$ 1.0	3.0 $\pm$ 1.0	3.4 $\pm$ 1.0	<.0001
Severity in GINA patterns, n (%)						<.0001
Intermittent	24 (33.8)	11 (16.9)	28 (34.6)	2 (6.1)	5 (14.7)	
Mild	38 (53.5)	30 (46.2)	34 (42.0)	14 (42.4)	21 (61.8)	
Moderate	8 (11.3)	12 (18.5)	12 (14.8)	11 (33.3)	6 (17.6)	
Severe	1 (1.4)	12 (18.5)	7 (8.6)	6 (18.2)	2 (5.9)	
Psychological status, n (%)						<.0001
Neither	41 (57.8)	0	43 (53.1)	1 (3.0)	0	
Either	25 (35.2)	0	38 (46.9)	0 (0.0)	0	
Both	5 (7.0)	65 (100.0)	0 (0.0)	32 (97.0)	34 (100.0)	

Abbreviations: ACT, Asthma Control Test; AQLQ, Asthma Quality of Life Questionnaire; FEV₁, forced expiratory volume in 1 second; FVC, forced vital capacity; GINA, Global Initiative for Asthma.

^aP < .05 vs cluster 1.

^bP < .05 vs cluster 2.

Table 4

Difference in Asthma Exacerbations Within 12-Month Follow-up in 5 Phenotypes Based on Cluster Analysis

Variables	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	P value
Patients, n (%)	71 (25.0)	65 (22.9)	81 (28.5)	33 (11.6)	34 (12.0)	
Exacerbations						<.0001
Median (Q1, Q3)	1 (0, 2)	6 (5, 10)	1 (0, 3)	6 (3, 8)	5 (0, 8.25)	
Mean $\pm$ SD	2.0 $\pm$ 3.8	11.6 $\pm$ 28.6 ^a	4.5 $\pm$ 13.6 ^b	6.9 $\pm$ 7.9 ^a	4.9 $\pm$ 4.4 ^a	
Unplanned visits						<.0001
Median (Q1, Q3)	0 (0, 0)	2 (0, 3)	0 (0, 0)	2 (0, 3)	1.5 (0, 3)	
Mean $\pm$ SD	0.2 $\pm$ 0.5	2.3 $\pm$ 2.6 ^a	0.3 $\pm$ 0.6 ^b	1.7 $\pm$ 1.6 ^a	2.2 $\pm$ 2.6 ^a	
Emergency visits						<.0001
Median (Q1, Q3)	0 (0, 0)	1 (0, 1)	0 (0, 0)	1 (0, 1)	0.5 (0, 1)	
Mean $\pm$ SD	0.03 $\pm$ 0.2	0.9 $\pm$ 1.2 ^a	0.04 $\pm$ 0.2 ^b	0.6 $\pm$ 0.7 ^a	0.9 $\pm$ 1.2 ^a	
Hospital admission						<.0001
Median (Q1, Q3)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)	
Mean $\pm$ SD	0.0 $\pm$ 0.0	0.3 $\pm$ 0.8 ^a	0.01 $\pm$ 0.1 ^b	0.2 $\pm$ 0.5	0.3 $\pm$ 0.7 ^a	

Abbreviations: Q1, quartile 1; Q3, quartile 3.

^aP < .05 vs cluster 1.^bP < .05 vs cluster 2.

severe asthma exacerbations, which has important implications for clinical practice.

Cluster analysis is a collection of methods for defining groups of individuals based on clinical characteristics, so that the individuals are grouped based on their differences (or similarities) into clusters.²² The groupings are constructed such that the degree of association is strong between members of the same cluster and weak between members of different clusters. To design a cluster analysis, we needed to consider at least 3 major issues.²³ The first relates to the selection of individuals. We recruited consecutive patients from a tertiary hospital, and their diagnosis of asthma was confirmed by evidence of airway hyperresponsiveness or bronchodilator reversibility. They represented different severities of asthma, which was different from other studies.^{24–28} For example, in the study by Boudier et al,²⁴ participants with asthma were identified by self-report based on answers to questionnaires, and there would be inherent issues with misdiagnosis of diseases, especially in cohorts that include smokers who could have other etiologies for respiratory symptoms. The studies by Ning et al²⁶ and Weatherall et al²⁵

were not restricted to asthma and included disparate individuals with respiratory symptoms or obstructive spirometric results, which are relevant to obstructive airway disease generally rather than specifically. In 2 other studies,^{27,28} although the cluster analysis algorithm discriminated the clusters at entry, these clusters did not behave differently concerning asthma control-related outcomes. This could be because they included individuals with severe or refractory asthma who were too similar and within a relatively homogenous group.

The second key issue is the selection of variables; if the variables represent epiphenomena, then clusters might present these rather than underlying pathogenic features. Because our aim was to define clinical asthma phenotypes based on clinical characteristics, all selected variables in our study had social demographic or clinical relevance. Sex, age at asthma onset, atopy, psychological symptoms, and BMI are clearly inherent characteristics. Although smoking exposure is an environmental factor, there is evidence to suggest that tobacco smoking can have epigenetic effects in promoting disease development.^{29,30} Asthma phenotypes identified in

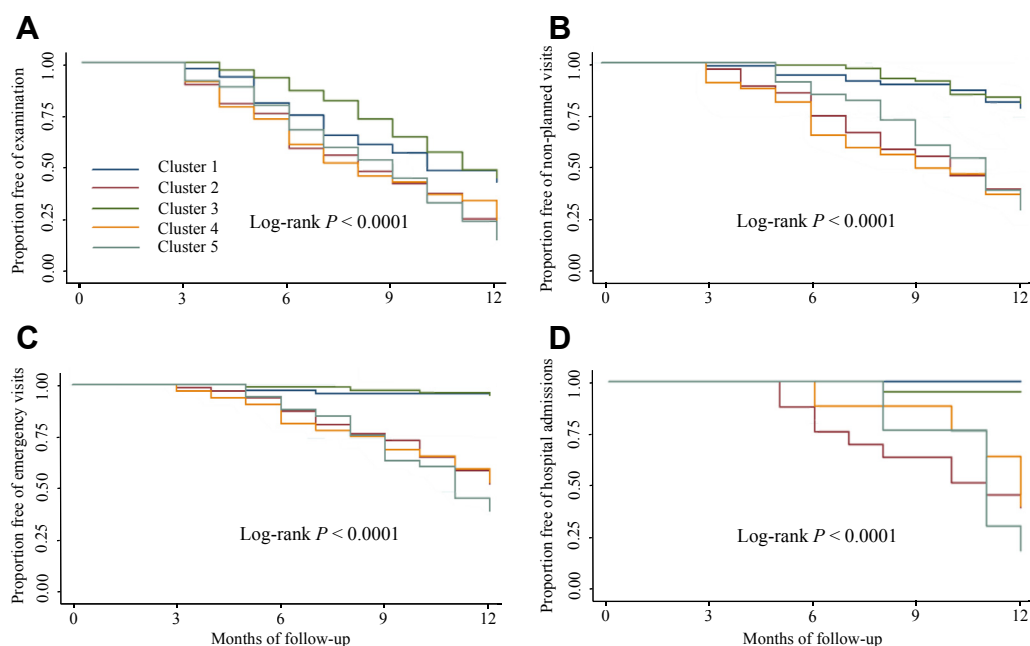


Figure 3. Kaplan-Meier plot of time to first asthma exacerbations in the 5 phenotypes based on cluster analysis during 12-month follow-up period indicating clusters 2, 4, and 5 had significantly shorter time to (A) asthma exacerbation, (B) unplanned visit, (C) emergency department visit, and (D) hospital admission.

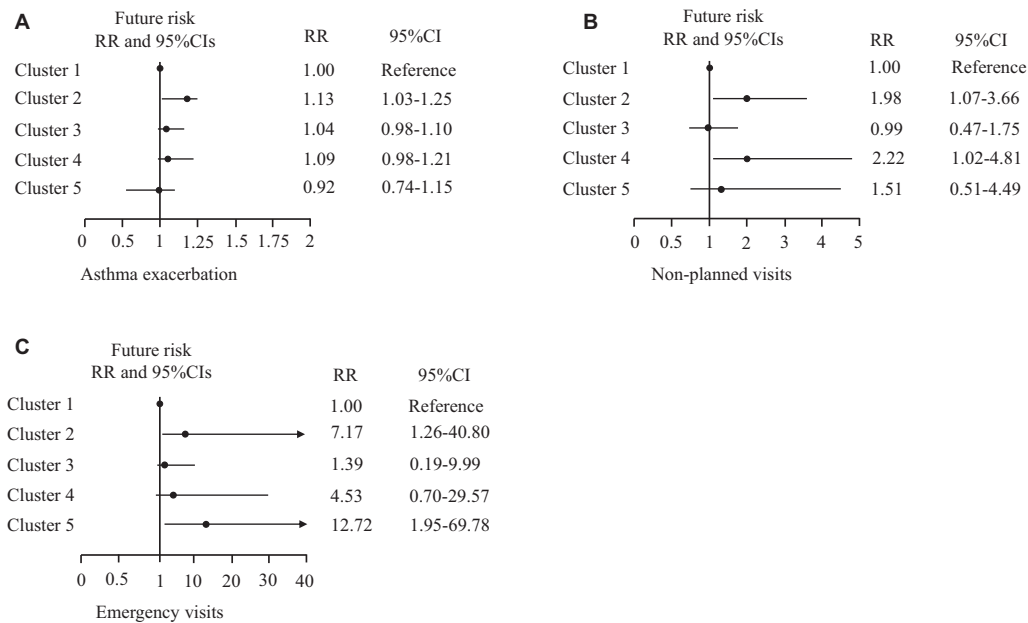


Figure 4. Relation of the 5 phenotypes based on cluster analysis to risk of asthma exacerbations (A), non-planned visits (B), and emergency visits (C) during 12-month follow-up period, with cluster 1 as the reference. CI, confidence interval; RR, relative ratio.

previous studies resulted from developed rather than developing countries; monthly income status as a selected variable is included in our cluster analysis because SES is a strong predictor of poor asthma control.³¹ Obviously, treatments (eg, inhaled corticosteroids affect those related to variable airflow obstruction) would modify the values of variables such as ACT score and FEV₁%predicted, which can lead to cluster instability or transitions.^{24,32} Although the choice of clustering parameters was broad, we cannot exclude the possibility that other variables might be of greater significance in developing clinically meaningful phenotypes.

Third, a consideration is the number of variables to choose to enter into a cluster analysis. In previously published studies, the number of variables varied from 5 to at least 25.^{21,33–35} Weatherall et al²³ suggested approximately 10 variables would be a useful number because it is difficult for too many variables to describe the clusters in a meaningful way. As a result, our study demonstrated a satisfying performance in cluster identification and had significantly differential characteristics except for BMI across clusters. It is likely explained by geography and ethnicity,^{32,36} which is supported by a study by Sakagami et al.³⁶ Another plausible issue is that the prevalence of overweight or obesity is relatively low in China.^{21,37–39} Although the global prevalence of overweight and obesity in developing countries such as China is increasing,^{40,41} obesity was not an important contributor to these clusters, which was supported by analysis of the relative variable importance in clustering shown in Figure 2A.

The long-term goals of asthma management are to achieve good symptom control and to minimize future risk of exacerbations.⁴² Although cross-sectional cluster analysis identified different phenotypes, they need to be validated for clinical relevance in a longitudinal study. In this study, we identified 3 clusters at increased risk of asthma exacerbations in a Chinese population. It is evident that the patients in clusters 2, 4, and 5 rather than clusters 1 and 3 had distinctly lower SES. Further, clusters 2, 4, and 5 had more psychological symptoms and were at high risk of severe or refractory asthma.⁴³ Studies from Korea⁴¹ and Japan⁴⁰ also reported clusters with increased exacerbation risk. Our clusters 1 and 3 are similar to cluster D and our cluster 4 is similar to cluster B in the study by Kim et al.²¹

In this study, we identified for the first time a “low SES” phenotype in cluster 5 with no health insurance in two thirds of subjects, which increased exacerbation risk. This is not surprising but not demonstrated by previously published cluster analyses from developed countries. The relation between asthma and SES has been confirmed previously. Recent studies by Uphoff et al⁴⁴ and Thakur et al⁴⁵ indicated that SES played an important role in predicting asthma but also correlated with worse asthma control and poor prognosis of asthma,^{31,46} which could be explained by a lower SES exacerbating the perceived effect on asthma. Intriguingly, some recently published studies indicated that low SES is related to low-grade systemic inflammation such as C-reactive protein and interleukin-6,^{47–49} which would modify the clinical characteristics or worsen symptoms of asthma.⁵⁰ Cluster 4 is the “fixed airflow limitation” phenotype identified in this study that is similar to the “persistent airflow obstruction” phenotype⁵¹ or the “accelerated lung function decline”³⁶ phenotype, which could be explained by a larger proportion of patients with severe asthma, smokers and patients with more psychological symptoms. Our recent studies^{52–54} indicated that psychological symptoms predisposed to greater perceived dyspnea intensity and was associated with significantly increased risk of asthma exacerbations presenting as hospitalizations, unscheduled doctor visits, and emergency department visits because of asthma. This could be explained by a weaker bronchodilator response and altered pattern of inflammation that could be refractory to standard therapies that would constitute a possibly distinct “neuropsychological” phenotype.^{5,52,54–56}

The strength of this study is represented by the prospective longitudinal design, recruitment of participants from a real-world setting in which interventions are instigated according to clinical guidelines, confirmed evidence of bronchial reversibility for asthma diagnosis, rigorous statistical methodology, and the ability to control for potentially important confounders. This study also has some potential limitations. First, because the cluster analysis was not prespecified, potential cluster variables could be identified only post hoc based on availability in our prospectively registered cohort study.⁵ Second, no biologic markers other than pulmonary function and skin testing were included, because our aim was to define clinical phenotypes based on clinical characteristics.

Although subjective symptoms are difficult to assess for accuracy in airway inflammation of asthma, the clinical phenotypes based on cluster analysis in this study might provide potential insight into pathobiologic mechanisms that are related to future risk of exacerbations. Third, detailed information on asthma treatments and their adherence was not available; therefore, we cannot provide information on the evolution of the treatment requirements and how treatment modified these clinical phenotypes or led to cluster instability.

Fourth, these asthma phenotypes based on cluster analysis resulted from data objects from a tertiary hospital, which need to be further validated in a primary-care population dataset.³⁸ Fifth, although 5 clusters were distinctly identified and would provide potential implications for clinical practice in developing countries, it was inevitable that some subjects with asthma on the “edge” would be incorrectly assigned to the identified clusters because of asthma heterogeneity, instability of clinical phenotypes, and few selected variables involved in the pathophysiology of airway inflammation in this study.^{24,32,57,58}

In conclusion, our study in a Chinese population with asthma conducted in a real-world setting identified several recognizable clinical phenotypes such as “allergic asthma,” “fixed airflow limitation,” and “low SES” phenotypes that were associated with high risk of future asthma exacerbations. This study adds new evidence to support the interest in cluster-derived asthma phenotypes for a better understanding of this complex disease and for asthma management in developing countries. Further, it will provide health policy makers in developing countries with evidence to improve health insurance because asthma is not usually covered by health care insurance in outpatient clinics. The concept of the phenotype has been suggested to be the prelude to that of the endotype, which would force us to further explore these phenotypes and their underlying pathobiology with identifiable biomarkers in the ongoing Australasian Severe Asthma Network involving Australia, New Zealand, and China.⁵⁹

References

- [1] Global Initiative for Asthma. *Global strategy for asthma management and prevention*. www.ginasthma.org. Accessed December 25, 2016.
- [2] Slejko JF, Ghushchyan VH, Sucher B, et al. Asthma control in the United States, 2008–2010: indicators of poor asthma control. *J Allergy Clin Immunol*. 2014;133:1579–1587.
- [3] Tan PN, Steinbach M, Kumar V. *Introduction to Data Mining*. 1st ed. Edinburgh, UK: Person Education Limited; 2014.
- [4] *Merriam-Webster's Collegiate Dictionary*. 11th ed. Springfield, MA: Merriam-Webster; 2008.
- [5] Wang G, Zhou T, Wang L, et al. Relationship between current psychological symptoms and future risk of asthma outcomes: a 12-month prospective cohort study. *J Asthma*. 2011;48:1041–1050.
- [6] Wang G, Baines KJ, Fu JJ, et al. Sputum mast cell subtypes relate to eosinophilia and corticosteroid response in asthma. *Eur Respir J*. 2016;47:1123–1133.
- [7] Schatz M, Zeiger RS, Vollmer WM, et al. Determinants of future long-term asthma control. *J Allergy Clin Immunol*. 2006;118:1048–1053.
- [8] Jia CE, Zhang HP, Lv Y, et al. The Asthma Control Test and Asthma Control Questionnaire for assessing asthma control: systematic review and meta-analysis. *J Allergy Clin Immunol*. 2013;131:695–703.
- [9] Zhou X, Ding FJ, Yin K. Validity of asthma control test for asthma control assessment in Chinese primary care settings. *Chest*. 2009;135:904–910.
- [10] Juniper EF, Buist AS, Cox FM, et al. Validation of a standardized version of the Asthma Quality of Life Questionnaire. *Chest*. 1999;115:1265–1270.
- [11] Xu KF, Luo XC, Chen Y, et al. The use of Juniper's asthma quality of life questionnaire in Chinese asthmatics. *Zhonghua Nei Ke Za Zhi*. 2003;42:760–763.
- [12] Zigmond AS, Snaith RP. The Hospital Anxiety and Depression Scale. *Acta Psychiatr Scand*. 1983;67:361–370.
- [13] Bjelland I, Dahl AA, Haug TT, et al. The validity of the Hospital Anxiety and Depression Scale. An updated literature review. *J Psychosom Res*. 2002;52:69–77.
- [14] Miller MR, Hankinson J, Brusasco V, et al. Standardisation of spirometry. *Eur Respir J*. 2005;26:319–338.
- [15] Zhen JP. *Textbook of Pulmonary Function Testing*. 1st ed. Guangzhou, Guangdong, China: Guangdong Science and Technology Publishing House; 2007.
- [16] Zheng J, Zhang X, Zhang L, et al. Interactive effects between obesity and atopy on inflammation: a pilot study for asthma phenotypic overlap. *Ann Allergy Asthma Immunol*. 2016;117:716.
- [17] Reddel HK, Taylor DR, Bateman ED, et al. An official American Thoracic Society/European Respiratory Society statement: asthma control and exacerbations: standardizing endpoints for clinical asthma trials and clinical practice. *Am J Respir Crit Care Med*. 2009;180:59–99.
- [18] SPSS Inc. *Two-step cluster analysis. Technical report*. Published 2004. http://support.spss.com/tech/stat/Algorithms/12.0/twostep_cluster.pdf. Accessed August 27, 2016.
- [19] Moore WC, Meyers DA, Wenzel SE, et al. Identification of asthma phenotypes using cluster analysis in the Severe Asthma Research Program. *Am J Respir Crit Care Med*. 2009;181:315–323.
- [20] Johann B, Knut W, Melanie V. *SPSS Two-Step Cluster: A First Evaluation*. Nuremberg, Germany: Sozialwissenschaftliches Institut Nürnberg; Lehrstuhl für Soziologie: Arbeits- und Diskussionspapiere; Lehrstuhl für Soziologie's Press; 2004.
- [21] Kim TB, Jang AS, Kwon HS, et al. Identification of asthma clusters in two independent Korean adult asthma cohorts. *Eur Respir J*. 2013;41:1308–1314.
- [22] McLachlan GJ. Cluster analysis and related techniques in medical research. *Stat Methods Med Res*. 1992;1:27–48.
- [23] Weatherall M, Shirlcliffe P, Travers J, et al. Use of cluster analysis to define COPD phenotypes. *Eur Respir J*. 2010;36:472–474.
- [24] Boudier A, Curjuric I, Basagaña X, et al. Ten-year follow-up of cluster-based asthma phenotypes in adults. A pooled analysis of three cohorts. *Am J Respir Crit Care Med*. 2013;188:550–560.
- [25] Weatherall M, Travers J, Shirlcliffe PM, et al. Distinct clinical phenotypes of airways disease defined by cluster analysis. *Eur Respir J*. 2009;34:812–818.
- [26] Ning P, Guo YF, Sun TY, et al. Study of the clinical phenotype of symptomatic chronic airways disease by hierarchical cluster analysis and two-step cluster analyses. *Zhonghua Nei Ke Za Zhi*. 2016;55:679–683.
- [27] Bourdin A, Molinari N, Vachier I, et al. Prognostic value of cluster analysis of severe asthma phenotypes. *J Allergy Clin Immunol*. 2014;134:1043–1050.
- [28] Schatz M, Hsu JW, Zeiger RS, et al. Phenotypes determined by cluster analysis in severe or difficult-to-treat asthma. *J Allergy Clin Immunol*. 2014;133:1549–1556.
- [29] Zhang HP, Wang L, Fu JJ, et al. Association between histone hyperacetylation status in memory T lymphocytes and allergen-induced eosinophilic airway inflammation. *Respirology*. 2016;21:850–857.
- [30] Tommola M, Ilmarinen P, Tuomisto LE, et al. The effect of smoking on lung function: a clinical study of adult-onset asthma. *Eur Respir J*. 2016;48:1298–1306.
- [31] Thakur N, Martin M, Castellanos E, et al. Socioeconomic status and asthma control in African American youth in SAGE II. *J Asthma*. 2014;51:720–728.
- [32] Newby C, Heaney LG, Menziesgow A, et al. Statistical cluster analysis of the British Thoracic Society Severe Refractory Asthma Registry: clinical outcomes and phenotype stability. *PLoS One*. 2014;9:e102987.
- [33] Seys SF, Scheers H, Van den Brande P, et al. Cluster analysis of sputum cytokine-high profiles reveals diversity in T(h)2-high asthma patients. *Respir Res*. 2017;18:39.
- [34] Loza MJ, Djukanovic R, Chung KF, et al. Validated and longitudinally stable asthma phenotypes based on cluster analysis of the ADEPT study. *Respir Res*. 2016;17:165.
- [35] Cabral AL, Sousa AW, Mendes FA, et al. Phenotypes of asthma in low-income children and adolescents: cluster analysis. *J Bras Pneumol*. 2017;43:44–50.
- [36] Sakagami T, Hasegawa T, Koya T, et al. Cluster analysis identifies characteristic phenotypes of asthma with accelerated lung function decline. *J Asthma*. 2014;51:113–118.
- [37] Sutherland ER, Goleva E, King TS, et al. Cluster analysis of obesity and asthma phenotypes. *PLoS One*. 2012;7:e36631.
- [38] Haldar P, Pavord ID, Shaw DE, et al. Cluster analysis and clinical asthma phenotypes. *Am J Respir Crit Care Med*. 2008;178:218–224.
- [39] Moore WC, Hastie AT, Li X, et al. Sputum neutrophil counts are associated with more severe asthma phenotypes using cluster analysis. *J Allergy Clin Immunol*. 2014;133:1557–1563.
- [40] Ng M, Fleming T, Robinson M, et al. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet*. 2014;384:766–781.
- [41] World Health Organization. http://www.who.int/gho/ncd/risk_factors/overweight/en/. Accessed February 10, 2017.
- [42] Reddel HK, Bateman ED, Becker A, et al. A summary of the new GINA strategy: a roadmap to asthma control. *Eur Respir J*. 2015;46:622–639.
- [43] Hahn DL. Comment on: international ERS/ATS guidelines on definition, evaluation and treatment of severe asthma. *Eur Respir J*. 2013;43:343–373.
- [44] Uphoff E, Cabieses B, Pinart M, et al. A systematic review of socioeconomic position in relation to asthma and allergic diseases. *Eur Respir J*. 2015;46:364–374.
- [45] Thakur N, Oh SS, Nguyen EA, et al. Socioeconomic status and childhood asthma in urban minority youths. The GALA II and SAGE II studies. *Am J Respir Crit Care Med*. 2013;188:1202–1209.
- [46] Chen W, Marra CA, Lynd LD, et al. The natural history of severe asthma and influences of early risk factors: a population-based cohort study. *Thorax*. 2016;71:267–275.
- [47] Stringhini S, Batty GD, Bovet P, et al. Association of lifecourse socioeconomic status with chronic inflammation and type 2 diabetes risk: the Whitehall II Prospective Cohort Study. *PLoS Med*. 2013;10:e1001479.

- [48] Stepanikova I, Bateman LB, Oates GR. Systemic inflammation in midlife: race, socioeconomic status, and perceived discrimination. *Am J Prev Med.* 2017;52: S63–S76.
- [49] Bonaccio M, Castelnuovo AD, Pounis G, et al. Relative contribution of health-related behaviours and chronic diseases to the socioeconomic patterning of low-grade inflammation. *Int J Public Health.* 2017;62:551–562.
- [50] Fu JJ, McDonald VM, Baines KJ, et al. Airway IL-1 β and systemic inflammation as predictors of future exacerbation risk in asthma and COPD. *Chest.* 2015; 148:618–629.
- [51] Konstantellou E, Papaioannou AI, Loukides S, et al. Persistent airflow obstruction in patients with asthma: characteristics of a distinct clinical phenotype. *Respir Med.* 2015;109:1404.
- [52] Zhang L, Zhang X, Zheng J, et al. Co-morbid psychological dysfunction is associated with a higher risk of asthma exacerbations: a systematic review and meta-analysis. *J Thorac Dis.* 2016;8:1257–1268.
- [53] Li HL, He XL, Liang BM, et al. Anxiety but not depression symptoms are associated with greater perceived dyspnea in asthma during bronchoconstriction. *Allergy Asthma Proc.* 2015;36:447–457.
- [54] Zhu M, Liang Z, Wang T, et al. Th1/Th2/Th17 cells imbalance in patients with asthma with and without psychological symptoms. *Allergy Asthma Proc.* 2016; 37:148–156.
- [55] Han YY, Forno E, Marsland AL, et al. Depression, asthma and bronchodilator response in a nationwide study of U.S. adults. *J Allergy Clin Immunol Pract.* 2015;4:68–73.e1.
- [56] Brehm JM, Ramratnam SK, Tse SM, et al. Stress and bronchodilator response in children with asthma. *Am J Respir Crit Care Med.* 2015;192: 47–56.
- [57] Moore WC, Fitzpatrick AM, Li X, et al. Clinical heterogeneity in the severe asthma research program. *Ann Am Thorac Soc.* 2013;10(suppl): S118–S124.
- [58] Moore WC. The natural history of asthma phenotypes identified by cluster analysis. Looking for chutes and ladders. *Am J Respir Crit Care Med.* 2013;188: 521–522.
- [59] Wang G, Wang F, Gibson PG, et al. Severe and uncontrolled asthma in China: a cross-sectional survey from the Australasian Severe Asthma Network. *J Thorac Dis.* 2017;9:1333–1344.