Associations of Four sensitization patterns revealed by Latent Class Analysis with Clinical symptoms: A multicenter study of China

Xiangqing Hou,a,# Wenting Luo,b,# Liting Wu,b,# Yuemin Chen,b,# Guoping Li,c,# Rongfang Zhang,d,# Hong Zhang,e,# Jing Wu,f,# Yun Sun,g,# Lina Xu,h,# Peiru Xu,i,# Yongmei Yu,j,# Dongming Huang,k,# Chuangli Hao,h,** and Baoqing Sunb*

aFaculty of Health Sciences, University of Macau 999078, Macau, China
bDepartment of Allergy and Clinical Immunology, State Key Laboratory of Respiratory Disease, National Clinical Research Center of Respiratory Disease, Guangzhou Institute of Respiratory Health, First Affiliated Hospital of Guangzhou Medical University, Guangzhou Medical University, Guangzhou 510120, Guangdong, China
cDepartment of Pulmonary and Critical Care Medicine, Laboratory of Allergy and Precision Medicine, Chengdu Institute of Respiratory Health, the Third People’s Hospital of Chengdu, Branch of National Clinical, 610000, Chengdu, China
dDepartment of Pediatrics, Gansu Provincial Maternity and Child Care Hospital 730400, Gansu, China
eDepartment of Pulmonary and Critical Care Medicine, Laboratory of Allergy and Precision Medicine, Inner Mongolia Cancer Hospital 021000, Inner Mongolia, China
fDepartment of Pediatrics, Yinchuan Maternal and Child Health Care Hospital 750001, Yinchuan, China
gDepartment of Pediatrics, Children’s Hospital, Soochow University 215025, Suzhou, China
hDepartment of Respirology, Children’s Hospital, Soochow University, Suzhou 215025, Jiangsu, China
iDepartment of Pediatrics, First Affiliated Hospital of Xinjiang Medical University 830000, China
jDepartment of Pediatrics, First Affiliated Hospital of Kunming Medical University 650000, Kunming, China
kDepartment of Pediatrics, Bo’ai Hospital of Zhongshan City 511462, Zhongshan, China

Summary

Background Because of the significant regional differences in the distribution of allergens, the relationship between anaphylaxis and allergic sensitization is complex in China. Using this large-scale epidemiologic survey, we explore the potential patterns of sensitization to common allergens in mainland China and investigate their relationship with various clinical symptoms.

Method The participants were recruited from 13 medical centers in mainland China from October 2019 to June 2021. Skin prick test (SPT) results that cover 18 common allergens were utilized to diagnose atopic sensitization. The demographic characteristics and clinical information were collected through questionnaires during routine medical follow-up. Latent class analysis (LCA) was conducted to determine the optimal sensitization patterns. The logistic regression was used to assess the associations of different sensitization patterns with allergy symptoms.

Findings A total of 1089 patients who had a positive SPT to at least one of 18 allergens were included for formal analysis. An optimal LCA model with 4 classes was obtained in this study, and the corresponding labels were as follows: Class1, house dust mite sensitization; Class2, low pollen sensitization; Class3, middle pollen sensitization; Class4, high pollen sensitization. The prevalence of different classes varied widely in geographical distribution, which was characterized by Class1 being very common in south and east as well as Class2 in north and west of China. Compared with patients in Class1, those in middle and high pollen sensitization clusters had the higher odds ratios (ORs) of allergic rhinitis and allergic conjunctivitis when controlling for other confounders. However, there was no significant difference between low pollen sensitization and house dust mite sensitization groups in the risks for various clinical performances except dermatitis. Additionally, the adjusted ORs (95% confidence interval) of allergic conjunctivitis and dermatitis for participants in pollen sensitization clusters (Class2, 3 and 4) were 1.56 (1.18, 2.06) and 1.43 (1.09, 1.88) respectively compared with those in Class1.

Abbreviations: CRD, component-resolved diagnosis; LCA, latent class analysis; IgE, Immunoglobulin E; sIgE, specific IgE; SPT, skin prick test; AIC, Akaike Information Criterion; CAIC, the “consistent AIC”; BIC, Bayesian Information Criterion; SSA-BIC, the adjusted BIC using Rissanen’s sample size adjustment; AS, asthma; AR, allergic rhinitis; 95% CI, 95% confidence interval
*aCorresponding author: Baoqiong Sun, First Affiliated Hospital of Guangzhou Medical University, 151 Yanjiangxi Road, Guangzhou 510120, Guangdong, China. Tel: +86 20 8306 2865; Fax: +86 20 8306 2729.
**Chuangli Hao, Department of Respirology, Children’s Hospital, Soochow University, Suzhou 215025, Jiangsu, China. Tel: +86-51280693588; Fax: +86-512-80696309.
E-mail addresses: hcl_md@163.com (C. Hao), sunbaoqing@vip.163.com (B. Sun).
# These authors should be considered co-first authors.

www.thelancet.com Vol 46 Month April, 2022
Interpretation In this study, we identified four sensitization clusters with specific risks of various clinical symptoms using common allergens by adopting LCA. Our findings may contribute to improved diagnosis and potential immunotherapy approaches to allergy in mainland China.

Funding This study was supported by the National Natural Science Foundation of China (81802076 and 81877736), the Guangzhou Science and Technology Foundation (202102010327), the Foundation of SKLRD (MS-2019-06 and Z-2022-09), and the Foundation of GYYY (ZH201904) and ZNSA-2020012.

Copyright © 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

Keywords: sensitization patterns; latent class analysis; skin prick test; clinical symptoms

Research in context

Evidence before this study

The associations of anaphylaxis with allergic sensitization patterns are very complex in China resulting from the significant regional differences in the distribution of allergens. We searched PubMed using the Mesh terms (“sensitization pattern” or “atopic sensitization”), (“asthma” or “allergic rhinitis” or “dermatitis” or “allergic conjunctivitis”), “China”, and (“Latent Class Analysis” or “LCA”) for articles published up to January 1, 2022 and found no multi-center epidemiologic study of the association between potential sensitization patterns to common allergens revealed by LCA and clinical symptoms. Although several studies have confirmed the causal associations of allergen sensitization with asthma, allergic rhinitis, conjunctivitis or dermatitis in China, most of them just focused on specific allergens on the risk of clinical symptoms and neglected the combined effects of allergens. Therefore, it’s necessary to identify sensitization clusters of common allergens and further to investigate their highly complex relationship with allergy symptoms among the Chinese population.

Added value of this study

To our knowledge, we are the first to identify four sensitization clusters with obvious separation and divided grass pollen sensitization into three degrees (low, middle, high) using 18 common allergens in mainland China by adopting LCA. We described that the prevalence of four sensitization patterns varied widely in geographical distribution, which was characterized by house dust mite sensitization (HDMs) groups being very common in south and east of China, while pollen sensitization clusters were concentrated in north or west of China. We found that compared with participants in HDMs groups, those in middle and high pollen sensitization clusters were associated with higher risks of allergic rhinitis and allergic conjunctivitis when controlling for other confounders. Besides, we revealed no significant difference in the risks of various clinical symptoms between low pollen sensitization and HDMs groups except dermatitis.

Implications of all the available evidence

This study provides new insights into the epidemiology distribution of potential sensitization patterns and gives important clues for the prevention and immunotherapy of allergy in mainland China.

Introduction

There is an abundance of evidence to support the associations of allergic sensitization with several clinical symptoms, for instance, asthma or rhinitis. However, the complicated distribution of dominating allergens in different regions and sensitization to multiple allergens concurrently enable the patients to present complex anaphylaxis, which poses a huge hinder in clinical diagnosis and therapy of allergy. In traditional clinical practices, clinicians commonly use a skin prick test (SPT) or specific Immunoglobulin E (sIgE) measurement results in combination with the corresponding exposure history of allergens for clinical diagnosis of allergy. However, it’s not uncommon that the co-sensitization of allergens exists among allergic patients. Therefore, the diagnosis precision of specific allergens might be confused by complicated sensitization patterns. Recently, advanced molecular-based diagnosis technique for allergy has been developed rapidly, for instance, component-resolved diagnostics (CRD) has facilitated the clarification of the biological mechanism of allergic diseases. Given conventional analysis for co-sensitization characteristics using the Venn diagram couldn’t disclose the complexity of sensitization patterns, many researchers have begun to utilize some more robust statistic methods for exploring the heterogeneity of potential sensitization patterns.

Latent class analysis (LCA) is a typical unsupervised machine learning method regarding categorical latent variable measured with categorical items. Individuals can be divided into subgroups or latent classes based on unobservable constructs using multiple correlated categorical variables. The latent classes are mutually
exclusive and characterized by having a high probability distribution to a set of specific observed categorical variables. Although several latent variable models have been successfully formulated to identify underlying patterns of sensitization to common allergens or component-specific IgE responses, most of them just focused on school-age children or specific allergic symptoms in many developed countries. To our best knowledge, few studies have been conducted in China to comprehensively investigate the associations between the patterns of sensitization to common allergens revealed by LCA and different clinical symptoms.

The sensitization patterns of allergy among Chinese are highly complicated, which results from significant regional differences in the distribution of allergens and herd susceptibility. Previous studies have reported that house dust mite is common in the southern and eastern areas of China, while pollen allergies are prevailing in the western and northern China. In this study, we recruited all qualified patients with allergies from 13 medical centers that are distributed to four regions of China, including south, north, west and east areas. A total of 18 common allergens which covered house dust mite, German cockroach, aspergillus fumigatus, penicillium and 14 grass pollen allergens were utilized to perform SPT.

Methods

Study design

This study was a large-scale multi-center epidemiology survey that was conducted in China from October 2019 to June 2021. The participants were obtained from Allergy/Pediatrics/Respiratory department from 13 medical centers that cover 11 provinces of mainland China, moreover, those research sites were divided into four regions including north (4 sites), south (3 sites), east (1 site), west (5 sites) areas in the final analysis. Patients attending outpatient clinics at 13 medical centers and diagnosed with allergic disorders were invited to participate in this survey. With the written consent forms, skin prick tests (SPT) were performed among patients in the above centers only after completing the questionnaires which were specifically designed for patients in this study. The including criteria were as follows: (1) participants with allergic disorders, including rhinitis, conjunctivitis, asthma and/or dermatitis; (2) participants underwent SPT with 18 common allergens and without missing value of test results; (3) participants who had a positive SPT to at least one of 18 allergens; (4) participants who completed questionnaire by themself or their guardian during routine medical visits. The participants with autoimmune diseases, parasitic infections or cancer were excluded from our study. Additionally, subjects with serious skin damage or recent medication were also excluded due to these may interfere with the accuracy of SPT results.

At last, a total number of 1089 patients (670 were men and 419 were women) aged 0-74 years were entered in formal statistics analysis. Data on demographics and clinical characteristics of participants were extracted from the standardized questionnaire by the trained physicians or research nurses, including gender (man or woman), age, smoking exposure (Does anyone in your family smoke?), family history of allergy (Does anyone in your family have allergies?), mode of delivery (normal delivery or caesararean section), habitual residence (rural or urban), ethnic group (Han or non-Han Chinese). Additionally, skin prick tests (Inmunotek, Inmunotek SL, Spain) were performed by two trained nurses for all the patients among all the centers. The wheal and flare results were obtained at 15 minutes and the orthogonal diameter greater and equal to 3mm was recognized as positive SPT. All processes were strictly following the standardized operation procedures determined before this study. Each completed questionnaire and skin test report was verified by two well-trained nurses and the results were double-checked by the principal investigator.

The research plan has been approved by the ethical committee of the First Affiliated Hospital of Guangzhou Medical University (GYFY-2018-93) and all patient’s informed consent forms were obtained before completing the SPT.

Definition of atopic sensitization

SPT results were utilized to determine atopic sensitization in this study. If the patients were positive to at least one of 18 common allergens (dermatophagoides pteronyssinus, blattella germanica, aspergillus fumigatus, penicillium notatum, ambrosia elatior, artemisia vulgaris, chenopodium album, betula verrucose, phleum pratense, brassica napus, triticum aestivum, ulmus campestris, salix fragilis, cynodon dactylon, populus alba, mediterranean cypress, platanus hispanica, phragmites communis), they were diagnosed with atopic sensitization. Histamine (10 mg/ml) and diluent were utilized as the positive and negative control, respectively. Allergen extracts and control solutions were put on the volar side of the forearm for SPT. After 15 minutes, the average value of the longest diameter and the length of the perpendicular line through its middle would be determined through the wheal reaction. When subtraction of the negative control, a positive skin reaction was defined as a wheal size ≥ 3 mm.
Definition of clinical outcomes
The clinical outcomes were obtained from a well-trained clinician based on clinical examination and questionnaire which was specifically designed for participants with potential allergic diseases. ‘Allergic rhinitis’ was defined as a positive answer to the question, ‘Have you had nasal itching, nasal congestion or spraying in the last 12 months?’; ‘Allergic conjunctivitis’ was defined as a positive answer to the question, ‘Have you had gritty sensation, redness, tearing, itching in your eyes in the last 12 months?’; ‘Asthma’ was defined as a positive answer to the question, ‘Have you had the symptoms of dyspnea, wheezing, chest tightness at night or shortness of breath or cough in the last 12 months?; ‘Dermatitis’ was defined as a positive answer to the question ‘Have you had urticaria or other symptoms of skin allergy?’.

Statistical analysis
The categories variables in this study were represented with frequency and proportional. A Chi-square test (χ²) was utilized to compare the group’s different clinical characteristics and allergens sensitization. Latent class analysis (LCA) was utilized to identify the potential sensitization patterns. The participants can be assigned to the class in which they had the highest posterior probability based on the maximum-probability assignment rule. An optimal latent class model with 4 classes was selected using 18 common allergens since the model had the lowest Bayesian information criterion (BIC) value and good clinical interpretability. To assess the performance and stability of our model, we additionally reported the fit statistics including Akaike information criterion (AIC), the “consistent AIC” (CAIC) and the adjusted BIC using Rissanen’s sample size adjustment (SSA-BIC) as well as the entropy of the models. Satellite imagery was applied to describe the prevalence of sensitization patterns with clinical symptoms using the SAS macro. Moreover, multivariate models was established to adjust for some potential confounding factors with a p-value less than 0.05 in the univariate analysis, including age, smoking exposure, habitual residence, ethnic group and region. We further established multivariate model2 to exclude the effects of region since the adjusted ORs of asthma in different sensitization patterns had extremely changed from model2 to model1 when controlling for region. Given geographical distribution was also a potentially important confounder factor, thus a subgroup analysis stratified by geographical regions was performed in this study to explore the association between sensitization patterns and the risk of asthma. Additionally, the multilevel model was established to investigate the association between sensitization patterns and asthma by utilizing PROC GLIMMIX in SAS 9.4. The 13 medical centers were recognized as a level-2 unit in this multilevel model.

Role of the funding source
The funders had no roles in study design, data collection, data analysis, interpretation of the data, as well as in the writing of the report and in the decision to submit the paper for publication. Baoqing Sun and Chuangli Hao had full access to the data in the study and had final responsibility for the decision to submit for publication.

Results
The characteristics of study populations
A total of 1089 participants met the inclusion criteria and were enrolled into formal statistics analysis. The median age (1st quartile, 3rd quartile) of the patients was 9(6, 15) years and 24.0% of the participants were adults (≥18 years). The male to female ratio was approximately 1.5:1. Most of the patients (74.9%) were urban dwellers. About ninety percent (89.2%) of patients were diagnosed with allergic rhinitis, and nearly sixty percent (58.2%) of participants declared to have allergic conjunctivitis. Additionally, the reported positive rates of asthma and dermatitis were 54.4% and 49.9% respectively.

The included participants were derived from 13 medical centers, which are distributed in south China (n=327), east China (n=170), north China (n=267) and west China (n=325), separately.

Characteristics of sensitization Patterns
In this study, the sensitization profile with 4 patterns was revealed by the LCA model according to their specific characteristics of sensitization as well as the lowest BIC criteria (Figure 1, Supplemental Table 1). Besides, the clinical interpretation had also been carefully considered since the four-class LCA model has an obvious separation and divided grass pollen sensitization into three degrees, that is, low, middle and high according to the extent of sensitization. Class1 is titled ‘house dust...
mite sensitization' due to it has an obvious high sensitized rate of house dust mite and low rates of other allergens. Class2 which is named 'low pollen sensitization' is characterized by lower levels of pollen allergens. Class3 is labeled as 'middle pollen sensitization' as this group overall has a higher prevalence of pollen sensitization than Class2. And Class4, a visible 'high pollen sensitization' cluster, which depicts a higher probability of sensitization to most pollens. The detailed sensitization rates of specific allergens among four classes can be seen in Table 1.

Demographics and clinical characteristics among different sensitization patterns were compared in Table 2. The results reveal that the Class4 group tends to have high proportions of school-age children (7-17, years) and the lowest percentage of women than other clusters. The proportion of the Han Chinese population in Class1 is larger than 95%, which is different from the other three groups (P<0.001). Besides, compared with patients in the house dust mite sensitization pattern (Class1), the proportions of urban residents are higher in pollen sensitization patients (Class2, 3 and 4). From the perspective of living regions, many patients with pollen sensitization can be seen in the north and west of China, while participants in class1 with house dust mite sensitization are mainly localized in the south and east area (Supplemental Figure 1).

The geographical distribution of various sensitization patterns

Figure 2 shows the prevalence of various sensitization patterns based on the geographical distribution of mainland China. The house dust mite sensitization pattern (Class1) is mainly distributed in the south and east of China. On the contrary, the low pollen sensitization clusters (Class2) are very common in the north and west regions of China. Additionally, it can be observed that the prevalence of the middle pollen sensitization clusters (Class3) is comparatively high in northern China. Our model identifies limited patients belonging to the high pollen sensitization group (Class4), and most of them can be found in the north of China.

Association of sensitization patterns with clinical outcomes

Figure 3 depicts the prevalence of various clinical symptoms among 4 classes. Compared with the participants with house dust mite sensitization, the risk of allergic rhinitis is significantly increased among those in the middle (OR = 2.08, 95% CI: 1.04–4.17) and high pollen sensitization classes (OR = 5.07, 95% CI: 1.17–21.93) when controlling for other potential confounders (Table 3). Similarly, significantly increased adjusted ORs of allergic conjunctivitis can be observed in Class3 (OR=2.17) and Class4 (OR=2.03) than Class1. The risks of dermatitis in pollen sensitization clusters (Class2, 3 and 4) are larger than those in the house dust mite sensitization groups though the adjusted OR values couldn’t achieve significance in Class4. However, there is no significant difference between the house dust mite group and low pollen sensitization cluster for the risks of allergic rhinitis, allergic conjunctivitis and asthma. In this study, we further combined Class2, Class3 and Class4 as a pollen sensitization cluster for comparison with Class1 on the risks of various clinical symptoms. The results reveal that compared with house dust mite
sensitized participants, individuals with pollen sensitization have the higher ORs of allergic conjunctivitis (OR=1.36, 95% CI: 1.18–1.56) and dermatitis (OR=1.43, 95% CI: 1.09–1.88) after adjusting for age, smoking exposure, habitual residence, ethnic group and region. Moreover, the results (Table 3) suggest that ORs calculated by the BCH method are similar to the corresponding results reported by the logistic regression model, which further confirm that our findings are stable and reliable.

Additionally, given geographical distribution is probably an important confounder factor to the effects of various sensitization patterns on clinical symptoms. We further excluded the confounding effects of the region for establishing model2. Table 3 represents the robustness of our findings depending on two multivariable logistics regression models, except the adjusted OR value of asthma in stratum2 has extremely changed from 1.41 in model2 to 0.83 in model1 when controlling for region. Therefore, a subgroup analysis stratified by geographical regions was further performed to explore the association between sensitization patterns and the risk of asthma. The results suggest that there are no significant associations between sensitization patterns and asthma in different regions (Supplemental Table 2). In this study, it could be better to treat the medical centers as a level-2 unit in the multilevel model. The result (Supplemental Table 3) also suggests that there is no significant association between sensitization patterns and asthma in the adjusted multilevel model which is consistent with the results of the subgroup analysis.

**Discussion**

It has been well known that allergic diseases are complex disorders with several clinical symptoms due to different potential pathogenicity mechanisms caused by complicated allergens. The identification of underlying patterns of sensitization to common allergens is facilitated to evaluate the risks of various clinical performance as well as the severity of disease. In this study, we included 18 common allergens which are distributed in four regions (south, west, north and east) of mainland China to identify underlying sensitization patterns, and further assess their associations with typical allergic symptoms. Finally, a four-class sensitization pattern was selected by applying LCA according to fit statistics and clinical interpretability. The four latent classes were labeled as ‘house dust mite sensitization’, ‘low pollen sensitization’, ‘middle pollen sensitization’ and ‘high pollen sensitization’ based on their specific characteristics of sensitization to allergens. Our finding suggests that patients in the ‘middle and high pollen sensitization’ classes show higher odds ratios (ORs) of allergic rhinitis and conjunctivitis than those in the ‘house dust mite sensitization’ class. However, there is no significant difference between Class1 and Class2 in the risks of various clinical symptoms except dermatitis. Additionally, compared with those in the house dust mite sensitization patterns, participants with pollen sensitization clusters are associated with higher ORs of allergic conjunctivitis and dermatitis when other confounders are under control.

**Table 1**: The positive rate of common 18 allergens among four sensitization patterns

| Variables                          | Class1 N=424(38.9) | Class2 N=394(36.2) | Class3 N=194(17.8) | Class4 N=77(7.1) | P value |
|------------------------------------|--------------------|--------------------|--------------------|-----------------|--------|
| Dermatophagoides pteronyssinus     | 398(93.9)          | 98(24.9)           | 61(31.4)           | 21(27.3)        | <0.0001|
| Blattella germanica                | 95(22.4)           | 46(11.7)           | 36(18.6)           | 18(23.4)        | <0.0001|
| Aspergillus fumigatus             | 1(0.2)             | 35(8.9)            | 12(6.2)            | 10(13.0)        | <0.0001|
| Penicillium notatum               | 3(0.7)             | 45(11.4)           | 21(10.8)           | 8(10.4)         | <0.0001|
| Ambrosia elatior                  | 0(0.0)             | 122(31.0)          | 117(60.0)          | 70(90.9)        | <0.0001|
| Artemisia vulgaris                | 18(4.2)            | 209(53.0)          | 148(76.3)          | 77(100.0)       | <0.0001|
| Chenopodium album                 | 0(0.0)             | 113(28.7)          | 117(60.3)          | 71(92.2)        | <0.0001|
| Betula verrucosa                  | 9(2.1)             | 54(13.7)           | 91(46.9)           | 68(88.3)        | <0.0001|
| Phleum pratense                   | 3(0.7)             | 43(10.9)           | 41(21.1)           | 62(80.5)        | <0.0001|
| Brassica napus                    | 2(0.5)             | 31(7.9)            | 91(46.9)           | 63(81.8)        | <0.0001|
| Triticum aestivum                 | 4(0.9)             | 24(6.1)            | 41(21.1)           | 60(77.9)        | <0.0001|
| Ulmus campesiris                  | 8(1.9)             | 81(20.6)           | 142(73.2)          | 72(93.5)        | <0.0001|
| Salix fragilis                    | 3(0.7)             | 38(9.6)            | 86(44.3)           | 69(89.6)        | <0.0001|
| Cynodon dactylon                  | 6(1.4)             | 61(15.5)           | 108(55.7)          | 77(100.0)       | <0.0001|
| Populus alba                      | 3(0.7)             | 64(16.2)           | 123(63.4)          | 70(90.9)        | <0.0001|
| Mediterranean cypress             | 12(2.8)            | 34(8.6)            | 37(19.1)           | 31(40.3)        | <0.0001|
| Platanus hispanicus              | 7(1.7)             | 48(12.2)           | 98(50.5)           | 68(88.3)        | <0.0001|
| Phragmites communis               | 3(0.7)             | 32(8.1)            | 50(25.8)           | 52(67.5)        | <0.0001|

Note: Class1, house dust mite sensitization; Class2, low pollen sensitization; Class3, middle pollen sensitization; Class4, high pollen sensitization.
Given the complicated cross-reactions of allergic sensitization, studies to explore the associations between various allergens and clinical symptoms using machine learning methods, such as network analysis or latent class analysis, have been emerging as a hot topic. However, most of them were performed in European countries, and the evidence from the Asia area was scarce. China has the largest population around the world and wide geographical distribution in Asia, which results in significant individual heterogeneity of the associations between sensitization panels and clinical performance. Though several studies have confirmed the causal associations of atopic sensitization with asthma, rhinitis or dermatitis in China, most of them just focused on specific allergens on the risk of clinical symptoms and neglected the combined effects of allergens. Therefore, it’s necessary to identify sensitization clusters of common allergens and further to investigate their complex relationship with allergy symptoms among the Chinese population using such machine learning methodologies. To our knowledge, this study was the first to explore the underlying patterns of sensitization to common allergens based on large-scale multicenter epidemiology data in China by using LCA. Additionally, this study of ours comes from samples in 13 medical centers, including representative 12 tertiary hospitals and 1 secondary hospital in their local areas according to the classification of Chinese hospitals. They are responsible for providing medical care to multiple regions and serve as medical hubs for conducting medical research. Therefore, the participants in this study have good representativeness of samples and can give implications for other areas with similar panels of allergen sensitization patterns as China at the national-specific level.

### Table 2: Demographic and clinical characterization of participants among four sensitization patterns

| Variables                  | Class1 N=424(38.9) | Class2 N=394(36.2) | Class3 N=194(17.8) | Class4 N=77(7.1) | P value |
|---------------------------|---------------------|---------------------|---------------------|------------------|---------|
| Age, years                |                     |                     |                     |                  |         |
| 0–6                       | 142(33.5)           | 128(32.5)           | 63(32.5)            | 19(24.7)         | 0.001   |
| 7–17                      | 198(46.7)           | 151(38.3)           | 80(41.2)            | 47(61.0)         |         |
| ≥18                       | 84(19.8)            | 115(29.2)           | 51(26.3)            | 11(14.3)         |         |
| Gender                    |                     |                     |                     |                  | 0.294   |
| Man                       | 255(60.1)           | 239(60.7)           | 121(62.4)           | 55(71.4)         |         |
| Woman                     | 169(39.9)           | 155(39.3)           | 73(37.6)            | 22(28.6)         |         |
| Smoking exposure          |                     |                     |                     |                  | 0.029   |
| Yes                       | 214(50.5)           | 233(59.1)           | 93(47.9)            | 41(53.2)         |         |
| No                        | 210(49.5)           | 161(40.9)           | 101(52.1)           | 36(46.8)         |         |
| Family history of allergy |                     |                     |                     |                  | 0.279   |
| Yes                       | 254(59.9)           | 212(53.8)           | 115(59.3)           | 47(61.0)         |         |
| No                        | 170(40.1)           | 182(46.2)           | 79(40.7)            | 30(39.0)         |         |
| Habitual residence        |                     |                     |                     |                  | 0.005   |
| Urban                     | 295(69.6)           | 301(76.4)           | 156(80.4)           | 64(83.1)         |         |
| Rural                     | 129(30.4)           | 93(23.6)            | 38(19.6)            | 13(16.9)         |         |
| Ethnic group              |                     |                     |                     |                  | <0.001  |
| Han Chinese               | 406(95.8)           | 335(85.0)           | 170(87.6)           | 63(81.8)         |         |
| Non-Han Chinese           | 18(4.2)             | 59(15.0)            | 24(12.4)            | 14(18.2)         |         |
| Region                    |                     |                     |                     |                  | <0.001  |
| South                     | 222(52.4)           | 71(18.0)            | 23(11.9)            | 11(14.3)         |         |
| North                     | 44(10.4)            | 139(35.3)           | 67(34.5)            | 17(22.1)         |         |
| East                      | 143(33.7)           | 20(5.1)             | 7(3.6)              | 0(0.0)           |         |
| West                      | 15(3.5)             | 164(41.6)           | 97(50.0)            | 49(63.6)         |         |
| Season                    |                     |                     |                     |                  | 0.255   |
| Spring                    | 73(17.2)            | 72(18.3)            | 40(20.6)            | 12(15.6)         |         |
| Summer                    | 155(36.6)           | 138(35.0)           | 69(35.6)            | 23(29.9)         |         |
| Autumn                    | 110(25.9)           | 114(28.9)           | 61(31.4)            | 30(39.0)         |         |
| Winter                    | 86(20.3)            | 70(17.8)            | 24(12.4)            | 12(15.6)         |         |
| Mode of delivery          |                     |                     |                     |                  | 0.536   |
| Normal delivery           | 269(63.4)           | 265(67.3)           | 130(67.0)           | 47(61.0)         |         |
| Caesarean section         | 155(36.6)           | 129(32.7)           | 64(33.0)            | 30(39.0)         |         |

Note: Spring (3, 4 and 5 months); Summer (6, 7 and 8 months); Autumn (9, 10 and 11 months); Winter (1, 2 and 12 months). Class1, house dust mite sensitization; Class2, low pollen sensitization; Class3, middle pollen sensitization; Class4, high pollen sensitization.
In the end, an optimal 4-class model which included three pollen sensitization patterns and one house dust mite allergic sensitization was identified by LCA. This model has not only the lowest BIC and CAIC (Supplementary Table 1) but also an excellent clinical interpretation. Although the LCA model with 5 classes has a lower AIC and SSA-BIC than the 4-class model, the latter one has higher entropy than the former one (0.81 VS 0.78). Additionally, according to Occam’s Law of Razor, the superiority model for achieving optimal results is a model with fewer variables, thus, the 4-class model can be more straightforward in clinical practice. Consistent with others, this study indicates that age is a significant influence factor of various clusters, and the adults have the lowest proportion in all clusters. The possible explanation is that children lack robust autoantibodies, which result in increased risks of sensitization to allergens compared to adults. Besides, it can be also observed that school-aged children tend to be in Class4 (high pollen sensitization) than pre-school children. As we all know, compared with pre-school children, school-aged children commonly have much time to take outdoor activities, thus, increasing the chance of exposure to outdoor pollen allergens. Additionally, a similar multi-center study performed in China suggested that the patterns of sensitization to outdoor allergens in the environment varied widely in different regions because China spans mid-temperate, warm-temperate, subtropical and tropical zones. In this study, we also highlighted varied widely geographical distribution among four classes. Overall, the most common sensitization pattern in the south and east of China is house dust mite sensitization, while the prevalence of low pollen sensitization is extremely higher in the west and north of China. In particular, the proportions of middle and high pollen sensitization patterns are comparatively higher in north China than in other regions. Therefore, our findings provide robust evidence to guide allergen detection planning and medical resource allocation in different regions. In this study, the seasons of SPT were determined by the prick date. Our findings suggest no significant seasonal variability (spring, summer, autumn and winter) among four sensitization patterns (P=0.255). Besides, a study performed in Europe suggests that seasonal variation in specific IgE antibodies of grass-pollen allergen was not associated with clinical symptoms. Therefore, we do not adjust for the seasons of SPT in the multivariate models. Moreover,
significant differences among the four clusters can be observed in several exposure factors, including smoking exposure, habitual residence and ethnic group in this study.

Several studies have compared the various clinical performance between house dust mites (HDMs) and grass pollen sensitization populations. For instance, a study from Spain showed no obvious difference between HDMs and grass pollen sensitized children with AR in terms of diseases severity and clinical features. In comparison with previous studies, the major novel finding in this study is that we further divided grass pollen sensitization into 3 degrees according to their sensitization characteristics. We proved that the ‘middle pollen sensitization’ cluster was associated with higher odds ratios of AR and allergic conjunctivitis than the ‘house dust mite sensitization’ though no similar results can be observed in the low pollen sensitization groups when controlling for other confounders. Our findings could be better for the reasonable management of grass pollen sensitized AR patients using limited medical resources, particularly in the west and north of mainland China, which have a large-scale pollen sensitization population. Moreover, we also combined Class2, 3 and 4 into pollen sensitization clusters for comparison with Class1 on the risks of various clinical symptoms. The results suggest that compared with Class1, patients in pollen sensitization clusters are associated with increased risks of dermatitis and allergic conjunctivitis. However, the associations between sensitization patterns and the risks of asthma changed a lot from multivariate model1 to model2 when removing the region, which are probably resulting from multicollinearity. In this research, we further assess the association between sensitization patterns and asthma in different regions. Our finding reveals that when controlling for other confounders, there is no significant difference between pollen sensitization and house dust mite sensitization on the risk of asthma in various regions, although the corresponding ORs are large than 1 in North and less than 1 in other regions.

![Figure 3. The prevalence of clinical symptoms among four sensitization patterns. Shown are prevalence rates expressed as a percentage (%). Class1, house dust mite sensitization; Class2, low pollen sensitization; Class3, middle pollen sensitization; Class4, high pollen sensitization](image-url)
| Outcomes             | N   | Cases (%) | Unadjusted model | Unadjusted BCH model | Adjusted\(\alpha\) model 1 | Adjusted\(\alpha\) model 2 |
|----------------------|-----|-----------|------------------|----------------------|-----------------------------|-----------------------------|
|                      |     |           | OR (95%CI)       | P                    | OR (95%CI)                  | P                          |
| Allergic rhinitis    |     |           |                  |                      |                             |                             |
| Stratum 1            |     |           |                  |                      |                             |                             |
| Class 1              | 424 | 378(89.20)| 1.00(1.00,1.00)  | Ref.                 | 1.00(1.00,1.00)             | Ref.                        |
| Class 2              | 394 | 336(85.30)| 0.71(0.47,1.07)  | 0.098                | 0.63(0.39,1.02)             | 0.060                       |
| Class 3              | 194 | 182(93.80)| 1.85(0.95,3.57)  | 0.069                | 2.20(0.89,5.43)             | 0.089                       |
| Class 4              | 77  | 75(97.40) | 4.56(1.08,19.21) | 0.038                | 4.80(0.87,26.61)            | 0.072                       |
| Overall test         |     |           |                  | <0.001               | 0.005                       | <0.001                      |
| Stratum 2            |     |           |                  |                      |                             |                             |
| Class 1              | 424 | 378(89.20)| 1.00(1.00,1.00)  | Ref.                 | 1.00(1.00,1.00)             | Ref.                        |
| Class 2              | 394 | 336(85.30)| 0.71(0.47,1.07)  | 0.098                | 0.63(0.39,1.02)             | 0.060                       |
| Class 3              | 194 | 182(93.80)| 1.85(0.95,3.57)  | 0.069                | 2.20(0.89,5.43)             | 0.089                       |
| Class 4              | 77  | 75(97.40) | 4.56(1.08,19.21) | 0.038                | 4.80(0.87,26.61)            | 0.072                       |
| Allergic conjunctivitis |     |           |                  |                      |                             |                             |
| Stratum 1            |     |           |                  |                      |                             |                             |
| Class 1              | 424 | 219(51.70)| 1.00(1.00,1.00)  | Ref.                 | 1.00(1.00,1.00)             | Ref.                        |
| Class 2              | 394 | 229(58.10)| 1.30(0.99,1.71)  | 0.063                | 1.28(0.92,1.77)             | 0.138                       |
| Class 3              | 194 | 134(69.10)| 2.09(1.46,2.99)  | <0.001               | 2.30(1.50,3.51)             | <0.001                      |
| Class 4              | 77  | 52(67.50) | 1.95(1.17,3.25)  | 0.011                | 1.96(1.12,3.42)             | 0.019                       |
| Overall test         |     |           |                  | <0.001               | <0.001                      | <0.001                      |
| Stratum 2            |     |           |                  |                      |                             |                             |
| Class 1              | 424 | 219(51.70)| 1.00(1.00,1.00)  | Ref.                 | 1.00(1.00,1.00)             | Ref.                        |
| Class 2+3+4          | 665 | 593(89.20)| 1.00(0.68,1.48)  | 0.099                | 1.00(0.65,1.54)             | 0.099                       |
| Asthma               |     |           |                  |                      |                             |                             |
| Stratum 1            |     |           |                  |                      |                             |                             |
| Class 1              | 424 | 213(50.20)| 1.00(1.00,1.00)  | Ref.                 | 1.00(1.00,1.00)             | Ref.                        |
| Class 2              | 394 | 228(58.10)| 1.36(0.93,1.97)  | 0.029                | 1.45(1.04,2.01)             | 0.027                       |
| Class 3              | 194 | 102(52.60)| 1.10(0.78,1.54)  | 0.589                | 1.07(0.72,1.58)             | 0.731                       |
| Class 4              | 77  | 49(62.40) | 1.73(1.05,2.86)  | 0.032                | 1.86(1.07,3.25)             | 0.028                       |
| Overall test         |     |           |                  | <0.001               | <0.001                      | <0.001                      |
| Stratum 2            |     |           |                  |                      |                             |                             |
| Class 1              | 424 | 213(50.20)| 1.00(1.00,1.00)  | Ref.                 | 1.00(1.00,1.00)             | Ref.                        |
| Class 2+3+4          | 665 | 415(62.40)| 1.55(1.21,1.99)  | <0.001               | 1.62(1.24,2.13)             | <0.001                      |
| Dermatitis           |     |           |                  |                      |                             |                             |
| Stratum 1            |     |           |                  |                      |                             |                             |
| Class 1              | 424 | 189(46.60)| 1.00(1.00,1.00)  | Ref.                 | 1.00(1.00,1.00)             | Ref.                        |
| Class 2              | 394 | 205(52.00)| 1.35(0.92,1.78)  | 0.033                | 1.37(0.99,1.90)             | 0.058                       |
| Class 3              | 194 | 110(56.70)| 1.63(1.16,2.29)  | 0.005                | 1.73(1.16,2.57)             | 0.007                       |

Table 3 (Continued)
Some limitations exist in our research. The most important one may be that our study only utilized whole allergen extracts instead of recombinant allergens to evaluate atopic sensitization, therefore, partly covered the cross-reactivity of similar molecular proteins in different allergens. However, we are indeed to consider this issue before our research. Initially, some recombinant allergen sources have no complete/representative panel of allergens yet available. Then, compared with recombinant allergens, the allergen extracts are easy to prepare and inexpensive as well as have been widely used in China for clinical diagnosis of allergy.15,17 Furthermore, to better disclose the underlying complicated sensitization patterns of common allergens depending on low-cost test data, we utilized a more robust statistic method, the latent class analysis (LCA), to explore the heterogeneity of potential sensitization patterns.3 Besides, since the clinical information of participants was collected through face-to-face routine medical investigation, some recall bias might exist in our study. Additionally, the cross-sectional study design could not certify the causal relationship between exposure and outcomes, as well as calculate the risk ratios (RRs) of exposure.

In conclusion, we applied LCA to disclose four mutually exclusive sensitization patterns among patients with atopic diseases using 18 common allergens in mainland China. Our findings confirm that compared with house dust mite sensitization, patients in pollens sensitization clusters have higher odds ratios of allergic conjunctivitis and dermatitis, although no similar results can be observed in allergic rhinitis and asthma. Additionally, patients in middle pollen sensitization are associated with high risks of allergic rhinitis and conjunctivitis than low pollen sensitization class but have no significant difference with high pollen sensitization populations. The present study could be beneficial for the precision diagnosis and clinical management of allergic diseases.

Declaration of interests
All authors declare no potential conflicts of interest.

Data sharing statement
The datasets used and/or analyzed during the current study are available from the corresponding author (e-mail: sunbaoqing@vip.163.com) on reasonable request.

Authors’ contributions
Xiangqing Hou, Wenting Luo, Liting Wu, Yuemin Chen, Guoping Li, Rongfang Zhang, Hong Zhang, Jing Wu, Yun Sun, Lina Xu, Peiru Xu, Yongmei Yu, Dongming Huang, Chuangli Hao, Baoqing Sun designed the study. Yuemin Chen, Guoping Li, Rongfang Zhang,
Hong Zhang, Jing Wu, Yun Sun, Lina Xu, Peiru Xu, Yongmei Yu, Dongming Huang completed the data management and data cleaning as well as interpretation. Xiangqiong Hou, Wenting Luo, Liting Wu performed the statistical analysis. Xiangqiong Hou, Wenting Luo, Liting Wu, Yuemin Chen, Guoping Li, Rongfang Zhang, Hong Zhang, Jing Wu, Yun Sun, Lina Xu, Peiru Xu, Yongmei Yu, Dongming Huang drafted the manuscript. Chiangli Hao and Baoping Sun edited the manuscript. All authors contributed to the critical revision of the manuscript and approved the final version.

Acknowledgments
We would like to thank everyone involved in the collection, detection, and data management of the samples. Thanks to the Biobank for Respiratory Disease in the National Clinical Research Center for Respiratory Disease (BRD-NCRCRD, Guangzhou, Southern China).

Funding
This study was supported by the National Natural Science Foundation of China (81802076 and 81871736), the Guangzhou Science and Technology Foundation (2021020327), the Foundation of SKL RD (MS-2019-06 and Z-2022-09), and the Foundation of GYY (ZH201904 and ZNSA-2020012).

Supplementary materials
Supplementary material associated with this article can be found in the online version at doi:10.1016/j.eclinm.2022.101349.

Reference
1. Fontanella S, Frazinay C, Murray CS, et al. Machine learning to identify pairwise interactions between specific IgE antibodies and their association with asthma: A cross-sectional analysis within a population-based birth cohort. PLoS Med. 2018;15(11):e1002651.
2. Yavuz ST, Okiel Karakus C, Custovic A, et al. Four subtypes of childhood allergic rhinitis identified by latent class analysis. Pediatr Allergy Immunol. 2021;32(8):1691-1699.
3. Luo W, Wang DD, Zhang T, et al. Prevalence patterns of allergen sensitization by region, gender, age, and season among patients with allergic symptoms in Mainland China: A Four-year Multicenter Study. Allergy. 2021;76(2):589–593.
4. MA C, L C, TB C, et al. Multiple-allergen and single-allergen immunotherapy strategies in polysensitized patients: looking at the published evidence. J Allergy Clin Immunol. 2012;129(4):929–934.
5. Hu H, Hou X, Luo W, et al. The Molecule Sensitized Pattern of Atopic Dermatitis Patients Who Co-Sensitized to Shrimp, Cockroaches, Crab and House Dust Mites. J Asthma Allergy. 2021;14:399–407.
6. Gao Z, Fu WY, Sun Y, et al. Asthma and pollen allergy in China: Component-resolved diagnosis reveals allergic asthma patients have significant multiple allergen sensitization. Allergy. 2019;74(2):284–293.
7. Melozi G, Passalacqua G, Canonica GW, et al. Component-resolved diagnosis in pediatric allergic rhinoconjunctivitis and asthma. Curr Opin Allergy Clin Immunol. 2015;15(4):446–451.
8. Proper SP, Azouz NP, Mersha TB. Achieving Precision Medicine in Allergic Disease: Progress and Challenges. Front Immunol. 2021;12:720746.
9. KL. N, T.A, B. M Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. Struct Equ Model. 2007;14(3):455–469.
10. Custovic A, Sonntag HJ, Buchan IE, et al. Evolution pathways of IgE responses to grass and mite allergen throughout childhood. J Allergy Clin Immunol. 2015;136(6):1643–1648.
11. Simpson A, Lazic N, Belgrave DC, et al. Patterns of IgE responses to multiple allergen components and clinical symptoms at age 11 years. J Allergy Clin Immunol. 2015;136(6):1224–1231.
12. Forster F, Ege MJ, Gerlach J, et al. Trajectories of asthma and allergy symptoms from childhood to adulthood. Allergy. 2021;76(2):589–593.
13. Luo W, Hu H, Tang W, et al. Allergen sensitization pattern of allergic adults and children in southern China: a survey based on real life data. Allergy Asthma Clin Immunol. 2019;15:42.
14. Zeng G, Luo W, Wu Z, et al. A cross-sectional observational study on allergen-specific IgE positivity in a southeast coastal versus a southwest inland region of China. Sci Rep. 2017;7(1):9393.
15. Ma T, Chen Y, Fang Y, et al. Prevalence and risk factors of allergic rhinitis and asthma in the southern edge of the plateau grassland region of northern China: A cross-sectional study. World Allergy Organ J. 2021;14(7):100357.
16. Wu L, Luo W, Hu H, et al. A Multicenter Study Assaying Risk Factors and Aeroallergens Sensitization Characteristics in Children with Self-Reported Allergic Rhinitis in China. J Asthma Allergy. 2021;14:1453–1462.
17. Li J, Sun B, Huang Y, et al. A multicentre study assessing the prevalence of sensitizations in patients with asthma and/or rhinitis in China. Allergy. 2009;64(7):1083–1092.
18. Li J, Huang Y, Lin X, et al. Influence of degree of specific allergic sensitivity on severity of rhinitis and asthma in Chinese allergic patients. Respir Res. 2011;12:95.
19. Nagin DS. Group-based modeling of development. Cambridge, MA: Harvard University Press; 2005.
20. Bozdogan H. Model-selection and Akaike’s information criterion (AIC). The general theory and its analytical extension. Psychometrika. 1987;52:343–370.
21. Schol S. Application of Model-Selection Criteria to Some Problems in Multivariate-Analysis. Psychometrika. 1987;52(3):333–343.
22. Elliott MR, Zhao ZC, Mukherjee B, et al. Methods to Account for Uncertainty in Latent Class Assignments When Using Latent Classes as Predictors in Regression Models, with Application to Acculturation Strategy Measures. Epidemiology. 2020;31(1):104–104.
23. Dziak JJ, Bray BC, Wagner AT. LCA_Distal_BCH SAS macro users’ guide (Version 1.1). University Park, PA: The Methodology Center, Penn State; 2017.
24. Chen H, Li J, Cheng L, et al. China Consensus Document on Allergy Diagnostics. Allergy Asthma Immune Res. 2021;13(2):177–205.
25. Grunwell JR, Gillespie S, Morris CR, et al. Latent Class Analysis of School-Age Children at Risk for Asthma Exacerbation. J Aller Ci Imm-Pract. 2020;8(7):2275+.
26. Frischmeyer-Guerrerio PA, Rasooly M, Gu W, et al. IgE testing can predict food allergy status in patients with moderate to severe atopic dermatitis. Ann Allergy Asthma Immunol. 2019;122(4):393–400.e2.
27. Kleine-Tebbe J, Mailander C. Patterns of allergen sensitization in patients with severe asthma in Germany. J Asthma Allergy Immunol Pract. 2020;9(2):744–46.e3.
28. Ogulur I, Pat Y, Ardicli O, et al. Advances and highlights in biomarkers of allergic diseases. Allergy. 2021.
29. Fortino V, Wegsll L, Werner P, et al. Machine-learning-driven biomarker discovery for the discrimination between allergic and irritant contact dermatitis. Proc Natl Acad Sci U S A. 2020;117(52):34747–34758.
30. Pernes A, Schulten V, Glesner J, et al. IgE and T Cell Reactivity to a Comprehensive Panel of Cockroach Allergens in Relation to Disease. Front Immunol. 2020;11:217000.
31. Zahradkin C, Chandra P, Tufaila A, et al. Sensitization to Common Allergens Among Children with Asthma and Allergic Rhinitis in Qatar. J Asthma Allergy. 2021;14:287–292.
32. Lanza ST, Collins LM, Lemmon DR, et al. PROC LCA: A SAS Procedure for Latent Class Analysis. Struct Equ Modeling. 2007;14(4):671–694.
33. Van Den Berg HA, Occam’s razor: from Ockham’s via moderna to modern data science. Sci Prog. 2018;101(1):261–272.
Hou X, Huang H, Hu H, et al. Short-term exposure to ambient air pollution and hospital visits for IgE-mediated allergy: A time-stratified case-crossover study in southern China from 2012 to 2019. *EClinicalMedicine*. 2021;37:100949.

Larbi A, Fulop T, Pawelec G. Immune Receptor Signaling, Aging and Autoimmunity. *Adv Exp Med Biol*. 2008;640:312–324.

MA S, J M, JG G, et al. Seasonal variation in specific IgE antibodies of grass-pollen hypersensitive patients depends on the steady state IgE concentration and is not related to clinical symptoms. *J Allergy Clin Immunol*. 1989;83:486–494.

Zicari AM, Indinnimeo L, De Castro G, et al. A survey on features of allergic rhinitis in children. *Curr Med Res Opin*. 2013;29(5):415–420.