

Disentangling the heterogeneity of allergic respiratory diseases by latent class analysis reveals novel phenotypes

Rita Amaral^{1,2} | Jean Bousquet^{3,4,5,6,7} | Ana M. Pereira^{1,8} | Luís M. Araújo^{8,9} |
Ana Sá-Sousa¹ | Tiago Jacinto^{1,2} | Rute Almeida¹ | Luís Delgado^{8,9} |
Joao A. Fonseca^{1,8,10}

¹CINTESIS - Center for Health Technology and Services Research, Faculty of Medicine, University of Porto, Porto, Portugal

²Department of Cardiovascular and Respiratory Sciences, Porto Health School, Porto, Portugal

³MACVIA-France, Contre les MALadies Chroniques pour un Vieillissement Actif en France European Innovation Partnership on Active and Healthy Ageing Reference Site, Montpellier, France

⁴INSERM U 1168, VIMA: Ageing and Chronic Diseases Epidemiological and Public Health Approaches, Villejuif, France

⁵UMR-S 1168, Université Versailles St-Quentin-en-Yvelines, Montigny le Bretonneux, France

⁶Euforea, Brussels, Belgium

⁷Euforea, Charité, Germany

⁸Allergy Unit, Instituto & Hospital CUF, Porto, Portugal

⁹Basic & Clinical Immunology, Pathology Department, Faculdade Medicina da, Universidade do Porto, Porto, Portugal

¹⁰MEDCIDS, Department of Community Medicine, Information, and Health Sciences, Faculty of Medicine, University of Porto, Porto, Portugal

Correspondence

Rita Amaral, CINTESIS - Center for Health Technology and Services Research, Faculty of Medicine, University of Porto, Porto, Portugal.

Email: rita.s.amaral@gmail.com

Funding information

Fundação para a Ciência e a Tecnologia, Grant/Award Number: PD/BD/113659/2015, PTDC/SAU-SAP/119192/2010

Abstract

Background: Refined phenotyping of allergic diseases may unravel novel phenotypes. Conjunctivitis as an independent disorder has never been approached.

Aim: To identify distinct classes of allergic respiratory diseases using latent class analysis (LCA) and distinguish each class using classification and regression tree (CART) analysis.

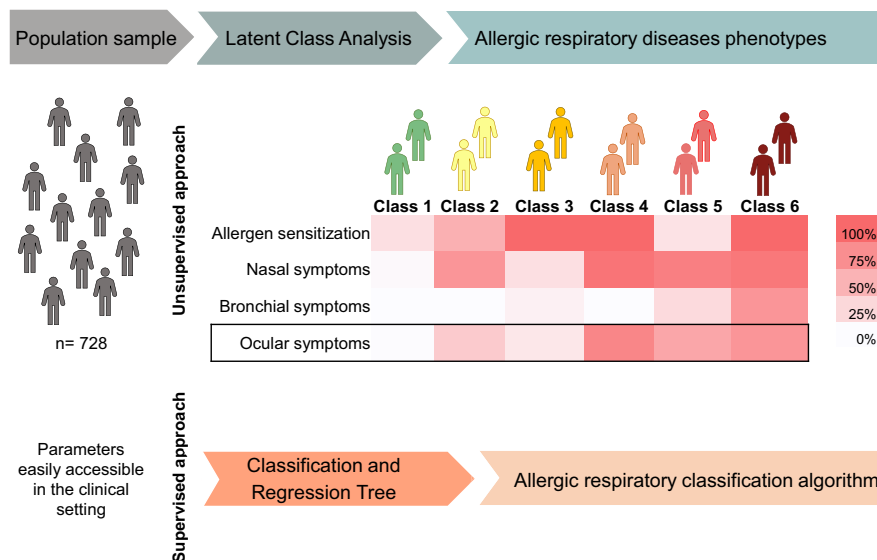
Methods: Seven hundred and twenty-eight adults from the Portuguese general population study ICAR had a structured medical interview combined with blood collection, skin prick tests, spirometry with bronchodilation, and exhaled nitric oxide. LCA was applied to 19 variables. The CART algorithm selected the most likely variables distinguishing LCA-classes.

Results: A six-class model was obtained. Class 1 (25%): nonallergic participants without bronchial or ocular symptoms. Classes 2 (22%) and 3 (11%): nasal and ocular (low levels) symptoms without nasal impairment, monosensitized (Class 2) or polysensitized (Class 3). Class 4 (13%): polysensitized participants with high levels of nasal and ocular symptoms, and nasal impairment. Classes 5 (16%) and 6 (14%): high level of nasal, bronchial and ocular symptoms with nasal impairment (non-allergic or polysensitized, respectively). Participants in classes 5 and 6 had more bronchial exacerbations and unscheduled medical visits ($P < 0.001$). Ocular symptoms were significantly higher in classes with nasal impairment, compared to those without impairment ($P < 0.001$) or no nasal symptom ($P < 0.001$). CART highlighted ocular symptoms as the most relevant variable in distinguishing LCA-classes.

Conclusion: Novel severe phenotypes of participants with co-occurrence of ocular, nasal and bronchial symptoms, and exacerbation-prone were identified. The tree algorithm showed the importance of the ocular symptoms in the expression of allergic diseases phenotypes.

KEYWORDS

airways symptoms, allergic sensitization, cluster analysis, ocular symptoms, phenotype



GRAPHICAL ABSTRACT

Novel phenotypes of allergic respiratory diseases were identified by latent class analysis, in adults from a general population sample. For the first time, ocular symptoms were used independently and ranked by classification and regression tree analysis as the most relevant variable to separate those phenotypes.

1 | INTRODUCTION

Allergic respiratory diseases (ARD) such as asthma, allergic rhinitis, and conjunctivitis, often co-exist in the same patient.¹ Allergic multimorbidity is independent of Immunoglobulin E (IgE) sensitization,² but the multimorbid polysensitized phenotype is among the most severe ones.³⁻⁵

Multimorbidity is complex and may be difficult to manage in clinical practice.⁶ Disease stratification using classical diagnostic methods may not be sufficient for a more effective and personalized patient care. Identification of “treatable traits” by data-driven approaches (unsupervised methods) may be a promising approach^{7,8} using machine learning tools.^{9,10} Previous studies used data-clustering (unsupervised) methods, such as latent class analysis (LCA), to identify separately phenotypes of asthma,^{11,12} rhinitis,^{13,14} and other allergic diseases,^{15,16} but nasal and ocular symptoms were considered together.¹⁶

The LCA seeks to identify homogeneous groups in a heterogeneous descriptive, a theoretical, and non-inferential manner without providing information on the comparative relevance of variables or on the classification decision rule.¹⁷ The classification and regression tree (CART) analysis can develop reliable clinical decision rules to be used in the development of new classification of patients into categories,^{18,19} because of their easily interpretive nature and ability to handle missing data.¹⁹ The development of a classification tree to classify participants in the ARD phenotypes, using data readily accessible in most outpatient settings, is expected to improve patient classification and management.

1.1 | Aim

The present cross-sectional study aimed to identify distinct classes of allergic respiratory diseases, in adults from a general population

sample, using latent class analysis and to explore the most relevant clinical variables that could be used to distinguish each class, using classification and regression tree analysis. Additionally, we included ocular symptoms that are usually not considered independently in allergic multimorbidity but were supported by a study using mobile technology²⁰ and suggested by MeDALL.¹⁶

2 | METHODS

2.1 | Sample and study design

This is a secondary analysis of a national and cross-sectional study conducted in the Portuguese general population, ICAR (Control and Burden of Asthma and Rhinitis) study.²¹ It comprised anthropometric measurements, lung function, inflammation and allergy tests, a structured clinical assessment and standardized questionnaires. Detailed information is provided in the Online Supplement. ICAR study enrolled 858 participants, from 3 to 89 years old, with and without self-reported asthma and/or rhinitis (Figure S1) and, for this analysis, we considered all adults (≥ 18 years old, $n = 728$) (Table S1).

Ethical approval was obtained from the Hospital Ethics Committee (Comissão de Ética do Hospital São João, E.P.E) and national data protection committee (n.12372/2011). All participants gave written informed consent. The paper follows the STROBE guidelines for reporting observational studies.²²

2.2 | Variables

Demographic characteristics, such as age, gender, body mass index (BMI), and smoking status were analyzed. Data regarding nasal, bronchial

and ocular symptoms, bronchial exacerbation and unscheduled medical visit, in the last 12 months, were collected by clinical interview. Age at onset of the bronchial symptoms was self-reported. Further details on variable definition are described in Online Supplement.

Nasal severity score was adapted from the ARIA severity score,²⁰ and it was calculated using questions regarding impact of nasal symptoms on daily activities, work, and sleep, ascribed with the score 1 if “Yes” and 0 if “No.” The nasal severity score was then categorized as “no/mild impairment” (ranging 0-2) and “severe impairment” (score = 3).

Additionally, a diagnosis of rhinitis, asthma, and other allergic diseases was established by an allergy specialist and the use of asthma/allergy medication in the last 12 months was also analyzed.

2.3 | Measurements

Allergic sensitization (AS) was assessed by skin-prick tests (SPT), following the GA2LEN recommendations.²³ The standardized allergen panel included 28 allergens (Stallergenes Greer®, France),²⁴ categorized into six groups (mites, dog and cat epitheliums, tree, grass and weed pollens mixtures, and molds). Monosensitization and polysensitization were defined, respectively, as sensitization to only one and to two or more groups of allergens. Furthermore, sensitization data of 25 participants with current antihistamine medication were considered as missing.

Total immunoglobulin E (IgE), serum eosinophilic cationic protein (S-ECP) and blood eosinophilic count (B-Eos) were obtained from blood sampling. Phadiatop® (Thermo Fisher Scientific, Uppsala, Sweden) was used as a screening test; if ≥ 0.35 kU/l the sample was considered as Phadiatop-positive, and additional determinations were performed to assess individual allergen-specific IgE (sIgE) antibody concentrations.

Predicted values of basal forced expiratory volume in one second (FEV_1) and bronchodilator response were obtained by spirometry procedure, following the ATS/ERS recommendations.^{25,26} FEV_1 were considered abnormal if less than the lower limit of normal (LLN).²⁷ Fractional exhaled Nitric Oxide (FeNO) measurements were performed using NIOX MINO® (Aerocrine, Solna, Sweden), following the ATS/ERS criteria.²⁸

2.4 | Biases

To reduce the risk of bias, several quality assurance measures were followed: research assistants performing the evaluations were blinded to the subject classification at screening; data validity was periodically verified soon after being collected and custom statistic algorithms were used to detect extreme, illogical and missing value and amendments to the protocol were done if necessary.

2.5 | Statistical analyses

Categorical variables are presented as absolute frequencies and proportions. Continuous variables were presented according to their distributions. The socio-demographic and clinical variables of the classes derived from the LCA were described and compared using chi-squared test, one-way analysis of variance and Kruskal-Wallis test with Bonferroni correction.

Mplus 6.12 (Los Angeles, CA: Muthén & Muthén) was used to conduct LCA analysis and R 3.3.3 (<https://www.r-project.org/>) to establish the classification model and build the respective decision tree, using the “rpart” and “DMwR” packages, respectively. All other analyses were performed using SPSS Statistics 25.0 (Armonk, NY: IBM Corp) and *P*-values < 0.05 were defined as statistically significant.

2.6 | Unsupervised analysis

LCA was applied to identify underlying unobserved (latent), mutually exclusive subgroups (classes) based on categorical manifest variables without the need for historical or *a priori* assumptions.²⁹ Nineteen dichotomic variables (defined as “Yes”/“No”) were chosen regarding nasal, ocular and bronchial symptoms and the 6 groups of AS.

2.7 | Supervised analysis

CART analysis was performed to obtain the classification tree algorithm, using Gini impurity index and the Cost-Complexity pruning algorithm.³⁰ The algorithm allows to select the variables most likely to identify LCA-classes, which included parameters easily accessible in most outpatient settings: patient's gender, age, and BMI; assessment of nasal, bronchial and ocular symptoms, nasal impairment; and objective diagnostic tests (number of AS groups, FEV_1 , and FeNO). The variable importance is given by Gini index (ranging 0%-100%).³¹ Additionally, we randomly divided the dataset into a training (70%) and a test set (30%) to obtain a reliable estimate of the model's predictive performance. Cohen's kappa coefficient (kappa) was used to evaluate model performance for imbalanced datasets.³²

Details regarding LCA and CART methodology and definitions of each included respiratory/ocular symptoms are provided in Online Supplement.

3 | RESULTS

3.1 | Sample characteristics

Seven hundred and twenty-eight adults (63% female) were included, mean (SD) age of 43.9 (15.2) years, 61% were non-smokers and 11% were taking inhaled corticosteroids (ICS). Demographic and clinical characteristics of the six classes are shown in Table 1.

3.2 | Latent class analysis

A six-class model was selected as the best solution for these data, with a significantly better fitting than a five-class model ($P = 0.013$), and a non-significantly different fit from a seven-class model ($P = 0.363$) (Table S2). Furthermore, entropy of the six-class model was 0.873, a good overall certainty in classification.

Figure 1 presents the probability of latent class membership for each of the six-class LCA model and Table S3 shows the stratification of the LCA-classes, according to clinical and allergic profiles.

TABLE 1 Demographics and clinical characteristics of the 6 LCA-derived classes

	Total (n = 728)	Class 1 (n = 182; 25%)	Class 2 (n = 157; 22%)	Class 3 (n = 77 ; 11%)	Class 4 (n = 102; 14%)	Class 5 (n = 114; 16%)	Class 6 (n = 96; 13%)	P-value
Female, n (%)	461 (63)	105 (58)	89 (57)	45 (58)	65 (64)	95 (83)	62 (65)	<0.001*
Age, mean (SD)	43.9 (15.2)	46.9 (16.5)	45.5 (16.4)	39.9 (14.8)	40.3 (12.9)	45.2 (13.4)	41.3 (13.2)	<0.001**
BMI, mean (SD)	25.9 (4.7)	26.6 (4.7)	25.1 (4.0)	25.8 (4.3)	25.2 (4.6)	25.8 (5.0)	27.1 (5.3)	0.003**
Age of bronchial symptoms onset, median (P ₂₅ -P ₇₅)	8.0 (3.0-20.0)	6.0 (1.0-13.0)	3.0 (2.0-18.0)	5.0 (3.0-7.5)	7.0 (2.0-12.0)	15.0 (3.0-25.0)	12.0 (4.0-30.0)	0.06***
Smoking status, n (%)								
Non-smoker	441 (61)	108 (59)	95 (60)	47 (61)	67 (66)	75 (66)	49 (51)	0.01*
Smoker	154 (21)	47 (26)	33 (21)	20 (26)	17 (17)	22 (19)	15 (16)	
Ex-smoker	133 (18)	27 (15)	29 (18)	10 (13)	18 (18)	17 (15)	32 (33)	
Packs-year, mean (SD)	5.9 (12.8)	8.3 (15.6)	6.5 (14.2)	5.0 (10.6)	3.6 (8.4)	4.2 (8.4)	5.7 (14.0)	0.03**
Current medication, n (%)								
ICS	81 (11)	3 (2)	6 (4)	14 (18)	5 (5)	18 (16)	35 (36)	<0.001*
Number of AS groups, median (P ₂₅ -P ₇₅)	2.0 (0-6.0)	0 (0-0.7)	1.0 (0-2.0)	4.0 (3.0-5.0)	4.0 (3.0-5.0)	0 (0-0)	4.0 (3.0-6.0)	<0.001***
Level of sensitization, n (%)								
Monosensitization ^a	100 (14)	37 (21)	37 (24)	0 (0)	0 (0)	22 (19)	4 (4)	<0.001*
Polysensitization ^b	332 (46)	8 (4)	50 (32)	77 (100)	101 (100)	4 (3)	92 (96)	
Group of AS, n (%)								
Mites	336 (46)	31 (17)	55 (35)	73 (95)	77 (76)	19 (17)	81 (84)	<0.001*
Cat and dog epithelium	220 (30)	2 (1)	18 (11)	64 (83)	60 (59)	7 (6)	69 (72)	<0.001*
Molds	135 (19)	0 (0)	11 (7)	33 (43)	40 (40)	1 (1)	50 (52)	<0.001*
Pollens ^c	335 (46)	16 (9)	55 (35)	70 (91)	101 (100)	3 (3)	90 (94)	<0.001*
Lung function								
FEV ₁ % predicted, mean (SD)	97.4 (15.5)	99.4 (16.1)	99.6 (13.3)	95.3 (13.9)	98.9 (13.9)	97.2 (15.6)	90.1 (18.4)	<0.001**
FEV ₁ <LLN, n (%)	69 (9)	16 (9)	8 (5)	6 (8)	9 (9)	8 (7)	22 (23)	<0.001*
Positive BD, n (%)	55 (8)	11 (6)	5 (3)	7 (9)	4 (4)	12 (11)	16 (17)	0.001*

AS, allergen sensitizations; BD, Bronchodilatation; BMI, body mass index; FEV₁, Forced Expiratory Volume in the first second; ICS, inhaled corticosteroids; LLN, lower limit of normal; P₂₅-P₇₅, 25th percentile-75th percentile.

^aSensitization to 1 group of allergens.

^bSensitization to 2 or more groups of allergens.

^cIncluding tree, grass and weed sensitizations.

*Chi-squared test.

**One-way analysis of variance (ANOVA).

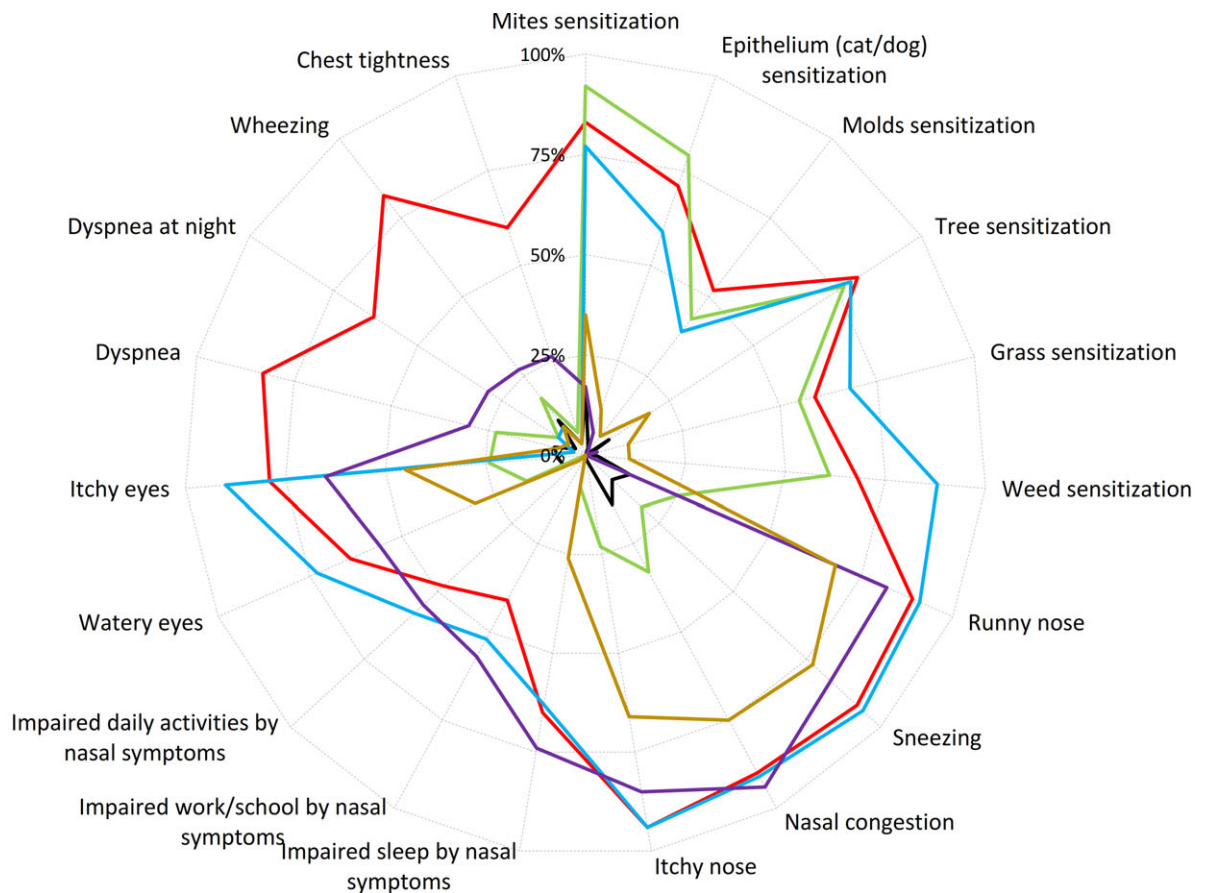
***Kruskal-Wallis test.

Average posterior probabilities were at least 89% for all classes, indicating a low chance of misclassification.

Two classes were characterized by non-allergic participants (>70% with negative SPT): classes 1 (n = 182; 25%) and 5 (n = 114; 16%). Class 1 had very low probability of having respiratory or ocular symptoms. Class 5 had a very high probability of having nasal, bronchial, and ocular symptoms with nasal severe impairment (nasal severity score ≥ 3).

Three classes were predominantly allergic (100% sensitization): classes 3 (n = 77; 11%), 4 (n = 96; 13%), and 6 (n = 102; 14%). Class 3 had a high probability of nasal and ocular symptoms without severe nasal impairment (score ≤ 2). Classes 4 and 6 predominantly had high nasal and ocular symptoms with nasal impairment, differing by the absence (Class 4) or presence of bronchial symptoms (Class 6).

Class 2 (n = 157; 22%) was characterized by participants with a very high probability of having nasal symptoms without severe nasal



Class 1: n = 182 (25%); **Class 2:** n = 157 (22%); **Class 3:** n = 77 (11%); **Class 4:** n = 102 (14%); **Class 5:** n = 114 (16%); **Class 6:** n = 96 (13%)

FIGURE 1 Proportions for the latent classes based on the estimated posterior probabilities [Colour figure can be viewed at wileyonlinelibrary.com]

impairment, with a moderately increased probability of ocular symptoms, and 55% of them were allergic.

3.3 | Latent class characteristics

There were significant differences among the six LCA-classes in all variables described in Table 1, except in age of bronchial symptoms onset ($P = 0.06$). A female predominance across all classes was observed, particularly in Class 5 (83%).

Most participants in classes 3, 4, and 6 were polysensitized, with the more frequent AS groups being: mites, pollens and cat/dog epithelia (Table 1 and Table S3). Moreover, half of the participants in Class 2 ($n = 87$; 55%) were sensitized, particularly to mites ($n = 55$; 35%) and classes 1 and 5 were mainly non-allergic.

Regarding the presence of ocular symptoms (Figure 2), the proportion of participants with ocular symptoms is significantly higher in those with severe nasal impairment ($\text{score} \geq 3$), compared to those without nasal impairment ($P < 0.001$). Similarly, the rate of participants with ocular symptoms having severe nasal impairment was significantly higher when compared to those with mild nasal impairment ($\text{score} \leq 2$) ($P < 0.001$) or without nasal symptoms

($P < 0.001$). Also, among participants without nasal symptoms, the proportion of ocular symptoms was very low (12%).

Classes 1 and 6 represented two extreme phenotypes: Class 1 was the mildest phenotype, whereas Class 6 was the most severe, including participants with the lowest and the highest values of B-Eos, FeNO and total IgE and proportions of urgent medical care, respectively (Table 2 and Figure 3). Class 6 participants had significantly higher proportions of current use of ICS, abnormal lung function and positive bronchodilation as compared to other classes (Table 1). After Class 6, the non-allergic Class 5 had the second highest proportions of participants with current use of ICS (16%), positive bronchodilation (11%), and bronchial exacerbations, comparing to other classes.

An "Intermediate" phenotype was found. Class 3 had the lowest mean age, and when compared to other classes, had significantly higher proportions of participants with sensitization to indoor allergens, with 18% of them having current ICS medication (Table 1). Participants in Class 3 had a significantly higher proportion of bronchial exacerbations in the past 12 months, compared to classes 1 and 2 ($P < 0.001$) (Figure 3).

Serum total IgE levels and Phadiatop[®] sIgE were highest in sensitized groups (classes 3, 4, and 6) (Table 2). B-Eos values were

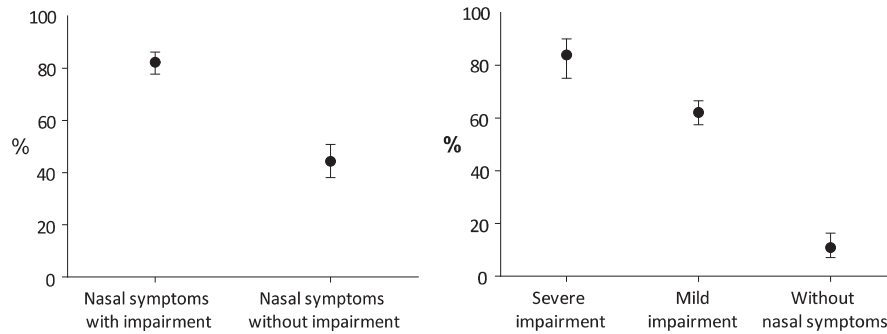


FIGURE 2 Proportions of participants with ocular symptoms, stratified by (A) the presence of nasal-related impairment and (B) ARIA severity score. The presence of nasal-related impairment defined as at least one of the following questions: impairment of daily activities by nasal symptoms, impairment of school/work by nasal symptoms and/or disturbed sleep due to nasal symptoms

significantly higher in classes 2, 4, and 6 comparing with Class 1. FeNO values were highest in classes 4 and 6, by comparison to classes 1, 3, and 5 ($P < 0.03$) (Figure S2).

Furthermore, medical diagnosis of rhinitis was common in classes 2, 4, 5, and 6 (Table S4), while asthma diagnosis was more frequent in Class 6 (75%), and less, but also prevalent, in classes 3 and 5 (35% and 31%, respectively). A high proportion of conjunctivitis was diagnosed in participants belonging to classes 4, 5, and 6. Class 1 participants had the lowest proportion of diseases medical-diagnosed, except for the proportion of other respiratory diseases (20%) (Table S4).

3.4 | CART analysis

Nine terminal nodes were formed in the classification tree (Figure 4), with a kappa(95% CI) = 0.75(0.72-0.79). Classification tree showed that ocular symptoms were the variable with the highest relative contribution to the model (37%), followed by number of AS groups (21%), having impairment of school/work by nasal symptoms (15%), presence of bronchial symptoms (13%), having impairment of daily activities (12%) and sleep disturbance due to nasal symptoms (3%).

On the right side of the tree (corresponding to participants sensitized to ≥ 3 AS groups), the presence of bronchial symptoms, distinguished Class 6 from all the others. When bronchial symptoms were absent, the presence/absence of impairment in daily activities and sleep disturbance by nasal symptoms, differentiated classes 3 and 4. Using a training ($n = 509$) and a test set ($n = 219$), the obtained CART algorithm was identical (Figure S3) and similar kappa was also obtained (0.73[0.67-0.80]).

4 | DISCUSSION

We identified from a general population sample using an unsupervised method six distinct classes of ARD in adults having clinically relevant differences in disease profiles. This study identified classes not determined by a clinical organ-centered diagnosis, but rather clinical characteristics and allergy profiles focusing on allergic

multimorbidities. There were two non-allergic ARD phenotypes (classes 1 and 5), three allergic phenotypes (classes 3, 4, and 6), and one with half of the participants being allergic (Class 2). Moreover, we identified novel severe ARD phenotypes, with participants having concomitant ocular, nasal and bronchial symptoms (classes 5 and 6) who are prone to exacerbations. CART analysis ranked ocular symptoms as the most relevant variable to separate LCA-classes, among respiratory symptoms, nasal-related impairment and diagnostic tests readily accessible in most outpatient settings.

4.1 | Strengths and limitations of the study

Limitations of this study should be recognized. The most important is that this work lacks external validity, and therefore generalizability. A potential source of bias for unsupervised analysis is how individuals were included. In our study, this risk is very low because not only we included data in LCA that was obtained in the clinical assessment rather than data from the initial screening (self-reported), but also, more importantly, pre-screening did not comprise data regarding ocular symptoms. Although we included participants from all regions of mainland Portugal, the sensitization profile to some allergens vary in Europe and results might differ depending on geographic areas. The choice of clinical variables included in the unsupervised model may condition the obtained LCA-classes. However, their inclusion was justified not only by their clinical relevance, being currently used and validated questions comprehending the main dimensions of the ARDs,^{1,33,34} but also compensating the fact that a high number of variables potentially increases the risk of some domains allocate too much weight in the modelling process.¹⁷ The inclusion of diagnostic tests was based on parameters being easily accessible at a visit to a specialist in clinical practice and that could prove useful in disease stratification and management. A sensitivity analysis was conducted using multiple imputation of incomplete data³⁵ and the obtained LCA-classes were similar when imputing all missing values (data not shown). Finally, the cross-sectional design of this study does not allow us to address latent classes stability over time.

TABLE 2 Distribution and comparison of FeNO and laboratory tests, according to each obtained LCA-classes

	Class 1 (n = 182; 25%)	Class 2 (n = 157; 22%)	Class 3 (n = 77; 11%)	Class 4 (n = 102; 14%)	Class 5 (n = 114; 16%)	Class 6 (n = 96; 13%)	Significant P-values*
FeNO, geom mean (CI 95%)	17.9 (16.0-19.9)	18.1 (16.2-20.0)	24.9 (18.9-30.9)	26.5 (22.0-31.0)	19.3 (16.1-22.4)	35.1 (28.7-41.5)	C1/C4, C1/C6, C3/C4, C3/C6, C4/C5, C5/C6
S-ECP, geom mean (CI 95%)	13.2 (11.2-15.2)	14.5 (11.9-17.1)	13.2 (10.6-15.9)	15.1 (11.7-18.5)	12.0 (9.8-14.3)	16.8 (13.7-20.0)	N.S.
B-Eos, geom mean (CI 95%)	158.7 (133.6-183.8)	177.2 (152.3-202.0)	200.9 (169.0-232.7)	205.3 (174.7-236.0)	189.3 (161.8-216.8)	237.2 (197.4-277.1)	C1/C2, C1/C4, C1/C6
Phadiatop® sIgE, median (P ₂₅ -P ₇₅)	0 (0-0)	0 (0-1.6)	3.6 (0-16.5)	7.4 (1.9-23.2)	0 (0-0)	13.2 (0.6-31.3)	C1/C3, C1/C4, C1/C6, C2/C3, C2/C4, C2/C6, C3/C5, C4/C5, C5/C6
Total IgE, geom mean (CI 95%)	85.5 (52.4-118.7)	76.9 (54.7-99.2)	202.8 (138.5-267.0)	184.8 (111.0-258.6)	52.2 (36.2-68.2)	227.6 (170.0-285.3)	C1/C3, C1/C4, C1/C6, C2/C3, C2/C4, C2/C6, C3/C5, C4/C5, C5/C6

B-Eos, blood eosinophilic count; C, class; CI, confidence interval; FeNO, fractional exhaled nitric oxide; N.S., non-significant; S-ECP, serum eosinophilic cationic protein; sIgE, specific IgE.
*Significant if P-values<0.05, using Kruskal-Wallis test with Bonferroni correction.

The unsupervised analysis based on airways symptoms and allergy profile data remains a powerful approach toward ARD phenotyping.^{10,16,36} In our study, we not only identified data-driven phenotypes but also, we provided information on the relevance of the variables that better distinguish the LCA-classes, using the supervised method CART analysis. While some previous studies used only unsupervised-clustering methods to identify phenotypes of ARD,¹¹⁻¹⁴ the present study extended this approach into providing information on the importance of the variables that best distinguish between the obtained classes, using CART analysis. To our knowledge, this is the first time CART analysis is studied in a population-based study of ARD. This analysis has various advantages over other methods, including multi-variable logistic regression: it is a nonparametric method; results are summarized in a tree, much simpler to interpret and more practical in a clinical setting; and measures the variable relevance in the model (relative impact of the predictors on the output).³¹ Also, our tree algorithm based on parameters already used in clinical practice performed well using a training and a test set, suggesting a high potential to be further applied.

A major strength of the study is that we used ocular symptoms independently for the first time in LCA and they were found to be extremely informative.

4.2 | Interpretation of the results

4.2.1 | Profiles differed by their association with IgE allergy

There were two non-allergic ARD phenotypes (classes 1 and 5), three allergic phenotypes (classes 3, 4, and 6) and one with 50% participants being allergic (Class 2). Class 1 was the mildest phenotype, without symptoms or inflammation. Class 5 comprised predominantly females, with low FeNO and B-Eos, and a high proportion of urgent health care use, suggesting that bronchial symptoms can be linked with non-allergic rhinitis³⁷ or rhinosinusitis. Moreover, Class 5 participants have some characteristics similar to the non-allergic rhinitis phenotype obtained by cluster analysis in other studies.^{13,14}

4.2.2 | Multimorbidity is a necessary component of ARD

In our study, the unsupervised analysis did not identify clusters of participants having asthma/rhinitis/conjunctivitis only, supporting the concept of “one airway one disease”³⁸ and considering the view that rhinitis, conjunctivitis, and asthma are different manifestations of the same disease. Particularly, there was no asthma cluster without rhinitis. This accords with epidemiologic studies such as ECRHS where asthma alone represents <10% of the asthmatic population. Thus, it seems that in clinical practice, multimorbidity should always be investigated in ARD patients. This fact also holds for a precision medicine approach based on treatable traits, rather than diagnostic labels, in the clinical management of the ARDs.^{18,39} One surprising issue is the prevalence of conjunctivitis in four classes. However, the diagnosis was made by a physician.

FIGURE 3 Proportions of participants with bronchial exacerbations and unscheduled medical visits, stratified by LCA-classes

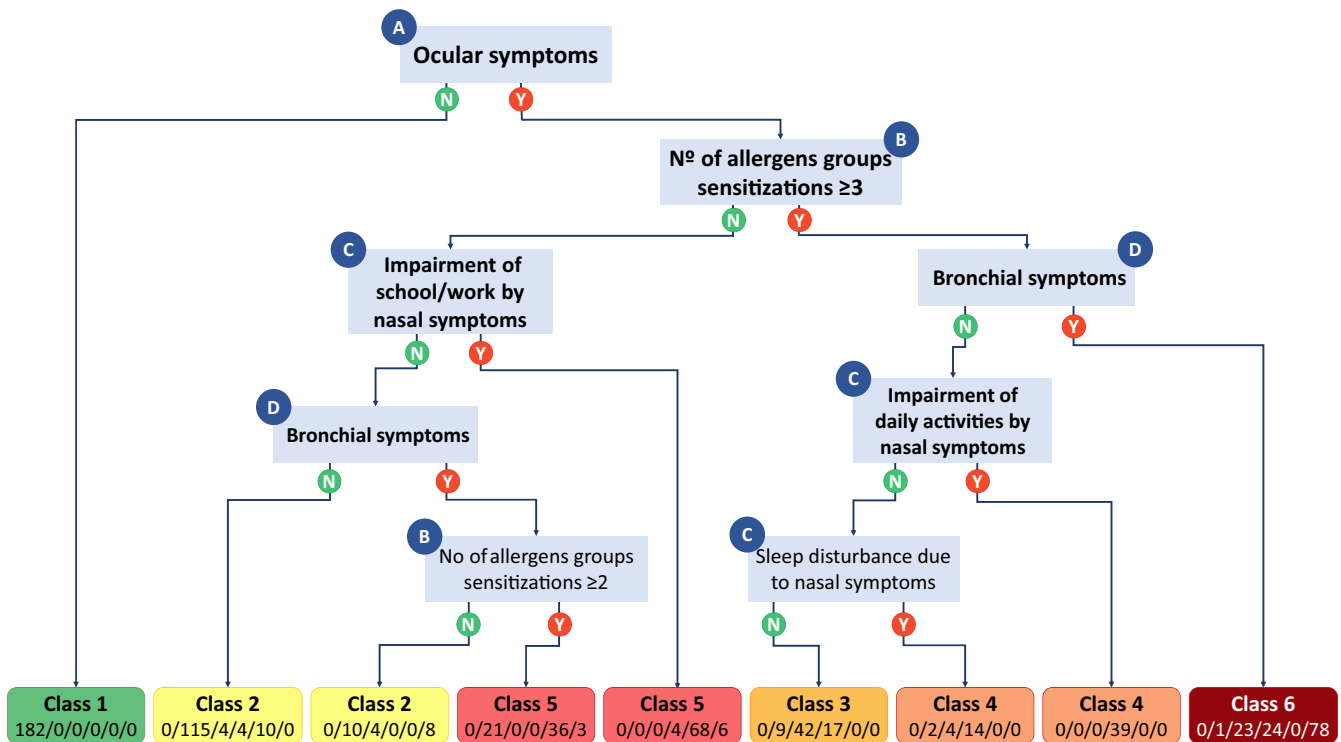
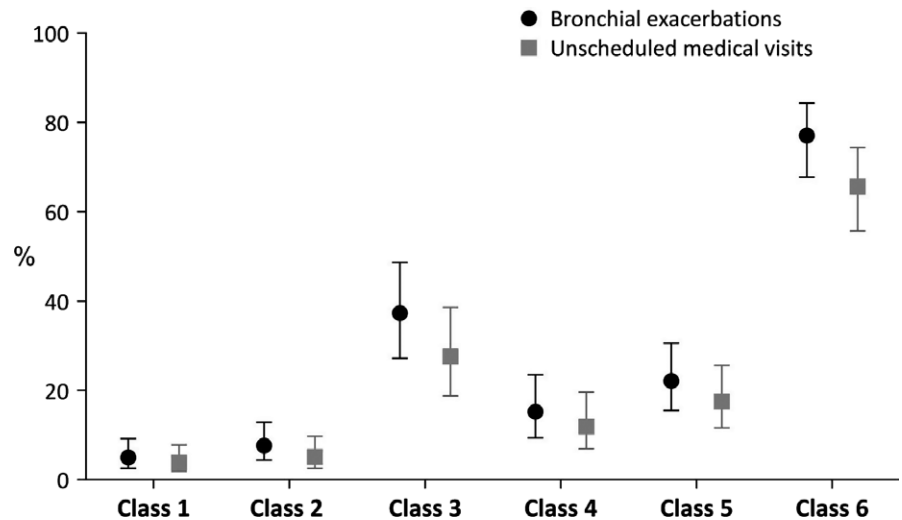


FIGURE 4 Classification tree algorithm generated by CART analysis using easily accessible parameters available in the clinical setting. (A) At least 1 ocular symptom: Watery eyes and/or Itchy eyes. (B) Mites, Cat/dog epitheliums, Tree, Grass, Weed, Molds. (C) Impairment due to nasal symptoms: Sneezing, Rhinorrhea, Nasal pruritus, Nasal congestion; (D) At least one bronchial symptom: Dyspnea, Dyspnea at night, Wheezing and/or Chest tightness. N: No; Y: Yes [Colour figure can be viewed at wileyonlinelibrary.com]

The obtained LCA-classes found in our sample, derived from the general population, were very similar to previously published clinical phenotypes in non-allergic and allergic rhinoconjunctivitis patients.^{40,41} However, we not only reinforced the different patterns of multimorbidity in participants with rhinitis, but also, we identified a specific class of participants polysensitized predominantly to indoor allergens (Class 3), with mild nose and ocular symptoms but presenting bronchial exacerbations, suggesting that different phenotypes of ARD can be obtained when analyzing a more comprehensive sample (population-based) and not derived from a secondary healthcare setting.

4.2.3 | Ocular symptoms are essential to identify clusters of ARD patients

The inclusion of ocular symptoms, independently from other symptoms was not explored in cluster analyses except by MeDALL,¹⁶ but, in this study, only rhinoconjunctivitis was considered. We observed that the proportion of ocular symptoms was significantly higher in participants with severe ARD (Classes 3 and 4 [severe nasal disease], 5 and 6 [severe nasal and bronchial disease prone to exacerbations]) and were less expressed in participants with milder ARD (classes 2 and 3). Finally, it

was found that ocular symptoms are associated with the severity of nasal symptoms suggesting that rhinitis and rhinoconjunctivitis represent two distinct phenotypes. This accords with a paper showing that the number of allergens recognized in the two phenotypes differ (low number in rhinitis, significantly higher number for rhinoconjunctivitis, Siroux et al, submitted).

The polysensitized multimorbid phenotype (Class 6) proposed by MeDALL^{4,5} was confirmed and found to be associated with rhinitis and asthma severity. Moreover, this study also identified conjunctivitis as another component of the multimorbid ARD phenotype. This class has some characteristics similar to those found in other unsupervised clustering studies, which has labeled as “late-onset, inflammation predominant”^{11,12,42}.

4.2.4 | CART analysis ranked ocular symptoms as the most relevant variable to separate LCA-classes

The tree algorithm not only reinforced the importance of including the presence of ocular symptoms in the expression of ARD phenotypes, among other easily accessible parameters used in a routine clinical practice, but also it may help to classify patients in clinical settings and to distinguish between lower or higher degrees of airways allergic multimorbidity.

4.2.5 | This study has an important impact in epidemiology

The concept of the importance of ocular symptoms in allergic multimorbidity was raised by a big data study using an App in uncharacterized users.⁴³ This hypothesis-generating study needed to be confirmed by a classical epidemiologic study and the present study is the first to show the relevance of big data in the future of epidemiology for respiratory and allergic diseases.

4.2.6 | The characterization of airway inflammation in ARD is also essential⁴⁴

In our data, we observed that B-Eos values were lower in participants of class 1, separated those with and without nasal and ocular symptoms, and lower FeNO values in non-allergic participants and without nasal-related impairment, differentiating phenotypes with different allergy profiles and symptoms severity, supporting the view that the combination of these two biomarkers should be used for personalized treatment.^{45,46}

4.3 | Generalizability

This study performed in Portugal needs to be confirmed in other countries as regional variations exist in allergens. However, a study on multimorbidity and allergen sensitizations showed that similar data were observed in France and Sweden suggesting common biological mechanisms.⁴⁷ After validation of these phenotypes in other populations, further work should address biologic associations, such as total IgE, sIgE, B-Eos, S-ECP, and mechanisms within these

phenotypes. Also, studies on the differentiation of allergic vs non-allergic conjunctivitis, and regarding symptoms' deterioration among non-smokers vs smoker/former-smokers and among different sensitizations, should be done, especially with prospective designs.

5 | CONCLUSION

The findings of our study challenge the conventional disease classification of a “classical” clinical diagnosis organ-based approach, to a combination of unsupervised and supervised analysis bringing novel insights to phenotyping ARD while helping confirm and integrate phenotypes previously reported. Six latent classes of ARD were identified in adults from a general population sample, with clinically relevant differences in disease profiles. Novel severe phenotypes of participants with co-occurrence of ocular, nasal and bronchial symptoms and exacerbation-prone, were revealed. Furthermore, the tree algorithm not only reinforced the importance of ocular symptoms in the expression of ARD phenotypes but also may help to distinguish between lower and higher degree of allergic airway multimorbidity.

ACKNOWLEDGMENTS

First author is supported by a PhD Grant (PD/BD/113659/2015), financed by the European Social Fund and national funds of MCTES (Ministério da Ciência, Tecnologia e Ensino Superior) through FCT (Fundação para a Ciência e Tecnologia, I.P) PhD Programme ref PD/0003/2013 - Doctoral Programme in Clinical and Health Services Research (PDICSS). The data collection was part of a funded FCT project (PTDC/SAU-SAP/ 119192/2010). The funding source had no role in the study design, the data collection, analysis and interpretation, or the writing of the manuscript.

CONFLICTS OF INTEREST

Dr. Amaral has nothing to disclose; Dr. Bousquet reports personal fees and other from Chiesi, Cipla, Hikma, Menarini, Mundipharma, Mylan, Novartis, Sanofi-Aventis, Takeda, Teva, Uriach, other from KYomed-INNOV, outside the submitted work; Dr. Pereira has nothing to disclose; Dr. Araújo has nothing to disclose; Dr. Sá-Sousa has nothing to disclose; Dr. Jacinto has nothing to disclose; Dr. Almeida has nothing to disclose; Dr. Delgado has nothing to disclose; Dr. Fonseca has nothing to disclose.

AUTHOR CONTRIBUTIONS

RA, JB, LD, and JAF contributed to the conception and design of the work, analysis and interpretation of data, drafting and revising the manuscript. RA, AMP, LA, ASS, RAI, and TJ contributed to data collection, analysis, and interpretation, as well as revising the manuscript. All authors read and approved the final version of the manuscript.

REFERENCES

1. Bousquet J, van CP, Khaltaev N. Allergic rhinitis and its impact on asthma. *J Allergy Clin Immunol*. 2001;108:S147-S334.
2. Pinart M, Benet M, Annesi-Maesano I, et al. Comorbidity of eczema, rhinitis, and asthma in IgE-sensitized and non-IgE-sensitized children in MeDALL: a population-based cohort study. *Lancet Respir Med*. 2014;2:131-140.
3. Johansson SG, Bieber T, Dahl R, et al. Revised nomenclature for allergy for global use: report of the Nomenclature Review Committee of the World Allergy Organization, October 2003. *J Allergy Clin Immunol*. 2004;113:832-836.
4. Anto JM, Bousquet J, Akdis M, et al. Mechanisms of the Development of Allergy (MeDALL): introducing novel concepts in allergy phenotypes. *J Allergy Clin Immunol*. 2017;139:388-399.
5. Bousquet J, Anto JM, Wickman M, et al. Are allergic multimorbidities and IgE polysensitization associated with the persistence or re-occurrence of foetal type 2 signalling? The MeDALL hypothesis *Allergy*. 2015;70:1062-1078.
6. Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. *Lancet*. 2012;380(9836):37-43.
7. Hodson R. Precision medicine. *Nature*. 2016;537(7619):S49.
8. Agusti A, Bel E, Thomas M, et al. Treatable traits: toward precision medicine of chronic airway diseases. *Eur Respir J*. 2016;47:410-419.
9. Bousquet J, Anto JM, Sterk PJ, et al. Systems medicine and integrated care to combat chronic noncommunicable diseases. *Genome Med*. 2011;3:43.
10. Bousquet J, Anto JM, Akdis M, et al. Paving the way of systems biology and precision medicine in allergic diseases: the MeDALL success story. *Allergy*. 2016;71:1513-1525.
11. Siroux V, Basagana X, Boudier A, et al. Identifying adult asthma phenotypes using a clustering approach. *Eur Respir J*. 2011;38:310-317.
12. Haldar P, Pavord ID, Shaw DE, et al. Cluster analysis and clinical asthma phenotypes. *Am J Respir Crit Care Med*. 2008;178:218-224.
13. Burte E, Bousquet J, Varraso R, et al. Characterization of rhinitis according to the asthma status in adults using an unsupervised approach in the EGEA study. *PLoS One*. 2015;10:e0136191.
14. Kurukulaaratchy RJ, Zhang H, Patil V, et al. Identifying the heterogeneity of young adult rhinitis through cluster analysis in the Isle of Wight birth cohort. *J Allergy Clin Immunol*. 2015;135:143-150.
15. Lee E, Lee SH, Kwon JW, et al. Atopic dermatitis phenotype with early onset and high serum IL-13 is linked to the new development of bronchial hyperresponsiveness in school children. *Allergy*. 2016;71:692-700.
16. Garcia-Aymerich J, Benet M, Saeys Y, et al. Phenotyping asthma, rhinitis and eczema in MeDALL population-based birth cohorts: an allergic comorbidity cluster. *Allergy*. 2015;70:973-984.
17. Collins LM, Lanza ST. The latent class model. In: Collins LM, Lanza ST, eds. *Latent Class and Latent Transition Analysis*. Hoboken, NJ: Wiley, 2010;23-47.
18. 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*. 2010;181:315-323.
19. Marshall RJ. The use of classification and regression trees in clinical epidemiology. *J Clin Epidemiol*. 2001;54:603-609.
20. Bousquet J, Arnavielhe S, Bedbrook A, et al. The Allergic Rhinitis and its Impact on Asthma (ARIA) score of allergic rhinitis using mobile technology correlates with quality of life: the MASK study. *Allergy*. 2018;73:505-510.
21. Sa-Sousa A, Pereira A, Almeida R, et al. Adult asthma scores - development and validation of multivariable scores to identify asthma in surveys. *J Allergy Clin Immunol Pract*. 2018; pii: S2213-2198(18)30428-8. <https://doi.org/10.1016/j.jaip.2018.06.024>. [Epub ahead of print].
22. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol*. 2008;61:344-349.
23. Bousquet J, Heinzerling L, Bachert C, et al. Practical guide to skin prick tests in allergy to aeroallergens. *Allergy*. 2012;67:18-24.
24. Heinzerling L, Frew AJ, Bindslev-Jensen C, et al. Standard skin prick testing and sensitization to inhalant allergens across Europe - a survey from the GA2LEN network. *Allergy*. 2005;60:1287-1300.
25. Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general U.S. population. *Am J Respir Crit Care Med*. 1999;159:179-187.
26. Miller MR. Standardisation of spirometry. *Eur Respir J*. 2005;26:319-338.
27. Quanjer PH, Cole TJ, Hall GL, Culver BH. Multi-ethnic reference values for spirometry for the 3-95 years age range: the global lung function 2012 equations. *Eur Respir J*. 2013;40:1324-1343.
28. American Thoracic S, European Respiratory S. ATS/ERS recommendations for standardized procedures for the online and offline measurement of exhaled lower respiratory nitric oxide and nasal nitric oxide, 2005. *Am J Respir Crit Care Med*. 2005;171:912-930.
29. Wang J, Wang X. Mixture modeling. In: Balding DJ, ed. *Structural Equation Modeling: Applications using Mplus*. New York: Wiley; 2012: pp 289-390.
30. Witten IH, Frank E, Hall MS. *Data Mining: Practical Machine Learning Tools and Techniques*. Burlington, VT: Elsevier; 2005.
31. Sauve M, Tuleau-Malot C. Variable selection through CART. *ESAIM Probab Stat*. 2014;18:770-798.
32. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159-174.
33. Fonseca JA, Nogueira-Silva L, Morais-Almeida M, et al. Control of Allergic Rhinitis and Asthma Test (CARAT) can be used to assess individual patients over time. *Clin Transl Allergy*. 2012;2:16.
34. Nguyen JM, Holbrook JT, Wei CY, Gerald LB, Teague WG, Wise RA. Validation and psychometric properties of the Asthma Control Questionnaire among children. *J Allergy Clin Immunol*. 2014;133:91-97.
35. Rubin DB. Multiple imputation. In: Rubin DB, ed. *Multiple Imputation for Nonresponse in Surveys*. Hoboken, NJ: John Wiley & Sons Inc; 1987:1-23.
36. Bousquet J, Anto J, Auffray C, et al. MeDALL (Mechanisms of the Development of ALLergy): an integrated approach from phenotypes to systems medicine. *Allergy*. 2011;66:596-604.
37. Shaaban R, Zureik M, Soussan D, et al. Rhinitis and onset of asthma: a longitudinal population-based study. *Lancet*. 2008;372(9643):1049-1057.
38. Grossman J. One airway, one disease. *Chest*. 1997;111(2 Suppl):11S-16S.
39. Bousquet J, Vignola AM, Demoly P. Links between rhinitis and asthma. *Allergy*. 2003;58:691-706.
40. Di Lorenzo G, Pacor ML, Amodio E, et al. Differences and similarities between allergic and nonallergic rhinitis in a large sample of adult patients with rhinitis symptoms. *Int Arch Allergy Immunol*. 2011;155:263-270.
41. Molgaard E, Thomsen SF, Lund T, Pedersen L, Nolte H, Backer V. Differences between allergic and nonallergic rhinitis in a large sample of adolescents and adults. *Allergy*. 2007;62:1033-1037.
42. de Vries R, Dagelet YWF, Spoor P, et al. Clinical and inflammatory phenotyping by breathomics in chronic airway diseases irrespective of the diagnostic label. *Eur Respir J*. 2018;51:1701817.
43. Bousquet J, Devillier P, Anto JM, et al. Daily allergic multimorbidity in rhinitis using mobile technology: a novel concept of the MASK study. *Allergy*. 2018;73:1622-1631.

44. Jeffery PK, Haahntela T. Allergic rhinitis and asthma: inflammation in a one-airway condition. *BMC Pulm Med.* 2006;6(Suppl 1):S5.
45. Malinovschi A, Fonseca JA, Jacinto T, Alving K, Janson C. Exhaled nitric oxide levels and blood eosinophil counts independently associate with wheeze and asthma events in National Health and Nutrition Examination Survey subjects. *J Allergy Clin Immunol.* 2013;132:821-827.
46. Malinovschi A, Janson C, Borres M, Alving K. Simultaneously increased fraction of exhaled nitric oxide levels and blood eosinophil counts relate to increased asthma morbidity. *J Allergy Clin Immunol.* 2016;138:1301-1308.
47. Siroux V, Ballardini N, Soler M, et al. The asthma-rhinitis multimorbidity is associated with IgE polysensitization in adolescents and adults. *Allergy.* 2018;73:1447-1458.