

Endotypes identified by cluster analysis in asthmatics and non-asthmatics and their clinical characteristics at follow-up: the case-control EGEA study

Rachel Nadif ,¹ Mickael Febrissy,² Miora Valérie Andrianjafimasy ,¹ Nicole Le Moual,¹ Frederic Gormand,³ Jocelyne Just,⁴ Isabelle Pin,^{5,6} Valerie Siroux,⁵ Régis Matran,^{7,8} Orianne Dumas ,¹ Mohamed Nadif⁹

To cite: Nadif R, Febrissy M, Andrianjafimasy MV, *et al*. Endotypes identified by cluster analysis in asthmatics and non-asthmatics and their clinical characteristics at follow-up: the case-control EGEA study. *BMJ Open Resp Res* 2020;**7**:e000632. doi:10.1136/bmjresp-2020-000632

► Additional material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/bmjresp-2020-000632>).

Received 8 May 2020
Revised 5 October 2020
Accepted 28 October 2020



© Author(s) (or their employer(s)) 2020. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to
Dr Rachel Nadif;
rachel.nadif@inserm.fr

ABSTRACT

Background Identifying relevant asthma endotypes may be the first step towards improving asthma management. We aimed identifying respiratory endotypes in adults using a cluster analysis and to compare their clinical characteristics at follow-up.

Methods The analysis was performed separately among current asthmatics (CA, n=402) and never asthmatics (NA, n=666) from the first follow-up of the French EGEA study (EGEA2). Cluster analysis jointly considered 4 demographic, 22 clinical/functional (respiratory symptoms, asthma treatments, lung function) and four blood biological (allergy-related, inflammation-related and oxidative stress-related biomarkers) characteristics at EGEA2. The clinical characteristics at follow-up (EGEA3) were compared according to the endotype identified at EGEA2.

Results We identified five respiratory endotypes, three among CA and two among NA: CA1 (n=53) with active treated adult-onset asthma, poor lung function, chronic cough and phlegm and dyspnoea, high body mass index, and high blood neutrophil count and fluorescent oxidation products level; CA2 (n=219) with mild asthma and rhinitis; CA3 (n=130) with inactive/mild untreated allergic childhood-onset asthma, high frequency of current smokers and low frequency of attacks of breathlessness at rest, and high IgE level; NA1 (n=489) asymptomatic, and NA2 (n=177) with respiratory symptoms, high blood neutrophil and eosinophil counts. CA1 had poor asthma control and high leptin level, CA2 had hyper-responsiveness and high interleukin (IL)-1Ra, IL-5, IL-7, IL-8, IL-10, IL-13 and TNF- α levels, and NA2 had high leptin and C reactive protein levels. Ten years later, asthmatics in CA1 had worse clinical characteristics whereas those in CA3 had better respiratory outcomes than CA2; NA in NA2 had more respiratory symptoms and higher rate of incident asthma than those in NA1.

Conclusion These results highlight the interest to jointly consider clinical and biological characteristics in cluster analyses to identify endotypes among adults with or without asthma.

INTRODUCTION

Asthma is undoubtedly a heterogeneous disease encompassing several phenotypes that may share common underlying

Key messages

- Can a clustering approach jointly considering clinical and biological data identify distinct respiratory endotypes among adult asthmatics and never asthmatics? Did their clinical characteristics differ at follow-up according to their endotypes at baseline?
- Using a clustering approach for mixed data integrating clinical and biological characteristics, this paper identified distinct respiratory endotypes among asthmatics and never-asthmatics in adults (≥ 16 years old) showing different clinical characteristics at follow-up.
- The present study highlights for the first time the interest to jointly consider clinical and biological characteristics in cluster analyses to identify distinct respiratory endotypes among adults with or without asthma.

mechanisms. In 2018, the *Lancet* Commission recommended deconstructing asthma into component parts before planning treatment, focusing in particular on ‘treatable’ traits.¹ Identifying treatable traits with specific clinical or molecular characteristics that could be targeted with treatment will help to understand which asthma subtype a patient has and how it should be treated.²

Asthma endotypes are usually defined as asthma subtypes characterised by a distinct functional or pathobiological mechanism.³ Identifying relevant asthma endotypes may be the first step towards improving its management. Over the last decade, studies mainly focused on patients with severe asthma who do not fully respond to currently available medications.^{3–5} Several clustering approaches have been used to identify asthma endotypes; however, there is still an unmet need to identify and characterise distinct asthma endotypes beyond severe asthma and beyond type2 (T2) asthma.^{6,7}



Asthma is a chronic inflammatory and oxidative-stress related disease.⁸ Blood eosinophilia and neutrophilia are recognised features of asthma related to specific phenotypes, able to orchestrate T2 and non-T2 immune responses.^{9 10} Previously, we highlighted the interest of using blood neutrophil and eosinophil counts to identify inflammatory phenotypes in adults with asthma in large-scale epidemiological studies.^{11 12} Recently, among several biomarkers related to oxidative stress, high level of fluorescent oxidation products (FLOPs)—a global biomarker of oxidation processes¹³ was associated with asthma characteristics and worse respiratory health among adult asthmatics.^{14 15} Up to now, continuous biological characteristics have been rarely selected simultaneously with clinical or functional data in cluster analyses, and almost always transformed into binary data.^{16 17}

As with other chronic diseases, asthma is not a dichotomous disease, and must be figured as a continuum in view of the current knowledges on its pathophysiology and natural history. Identifying specific respiratory endotypes among never asthmatics (NA) may add values to prevent the development of the disease. To date, no cluster analysis has identified specific respiratory endotypes among non-severe asthmatics and NA in an epidemiological study, and none has incorporated blood eosinophil and neutrophil counts and plasma FLOPs level jointly with demographic, clinical and functional asthma characteristics.

In the framework of the French longitudinal epidemiological study on the genetic and environment of asthma (EGEA), we previously identified four phenotypes among asthmatics by latent class analyses using only demographic and clinical/functional asthma characteristics.¹⁸ In the present paper, we hypothesised that distinct respiratory endotypes exist in both asthmatics and NA which differ on their long-term evolution. We first performed a cluster analysis by jointly considering clinical and biological data. Then to evaluate which endotypes might need a better disease management, we compared their clinical characteristics at follow-up according to the endotype identified at baseline.

METHODS

Study design and participants

The French EGEA study is a longitudinal study with an initial group of asthma cases and their first-degree relatives, and controls followed-up for over 20 years (first survey: EGEA1, 1991–1995, <https://egeanet.vjf.inserm.fr/>). The protocol and descriptive characteristics have been described previously.^{19 20} Briefly, 2047 participants from five cities were enrolled at EGEA1. Between 2003 and 2007, they were contacted for the second survey (EGEA2). As a follow-up study of EGEA2, the third survey (EGEA3) was conducted between 2011 and 2013 using a self-questionnaire (see the online supplemental materials for more details).

We first identified endotypes using data from EGEA2 including adult participants (≥ 16 years) without asthma (NA) or with current asthma (CA, [figure 1](#)). The analysis was carried out on 1068 participants after an imputation step. Then, we compared their clinical characteristics at EGEA3 respective to the endotype identified at EGEA2. This analysis was performed in 917 participants (86%) followed up at EGEA3.

Approvals were obtained from the relevant Ethics Committees and Institutional Review Board Committees: INSERM, RBM ‘Recherche BioMédicale’ RBM 91-005 and RBM 01-11; CNIL ‘Commission Nationale de l’Informatique et des Libertés’ no 109427 (04/1990), no 900198 (10/2000) and no 1769319 (2014); Institutional Review Board Committees (no 01-07-07, 04-05-03, 04-11-13 and 04-11-18); DGS ‘Direction Générale de la Santé’ no 2002/0106 and no 910048. All participants signed a written informed consent.

Patient and public involvement

Patients or the public were not involved in the design, or conduct, or reporting, or dissemination plans of the present research.

Definitions of ever asthma and CA

At EGEA2, the participants with ever asthma were those who answered positively to at least one of the two following questions: ‘Have you ever had attacks of breathlessness at rest with wheezing?’ or ‘Have you ever had asthma attacks?’, or were recruited as asthmatic cases at EGEA1. NA were those who answered negatively to the two questions above, and were not recruited as asthmatic cases at EGEA1. They were recruited as family members, spouses or control, could report respiratory symptoms but did not fulfil the proposed strict criteria to define asthma.¹⁹ Among participants with ever asthma, CA was defined by the report of respiratory symptoms (wheeze, nocturnal chest tightness, or attacks of breathlessness following strenuous activity, at rest or at night time) or asthma attacks or use of inhaled and/or oral medicines because of breathing problems in the past 12 months.¹⁴

Participants with ever asthma but without CA were excluded from the analyses. In order to facilitate reading, participants with CA are called ‘asthmatics’ and those without asthma called ‘never-asthmatics’.

Demographic, clinical and biological characteristics included in the cluster analysis

Twenty-three characteristics common to NA and asthmatics, and five supplementary specific characteristics only for asthmatics were selected and included in the cluster analysis. These characteristics were selected to reflect as comprehensively as possible the demographic, clinical and biological characteristics of the participants. We selected the native variables instead of their combinations (eg, ‘chronic cough’ and ‘chronic phlegm’ instead

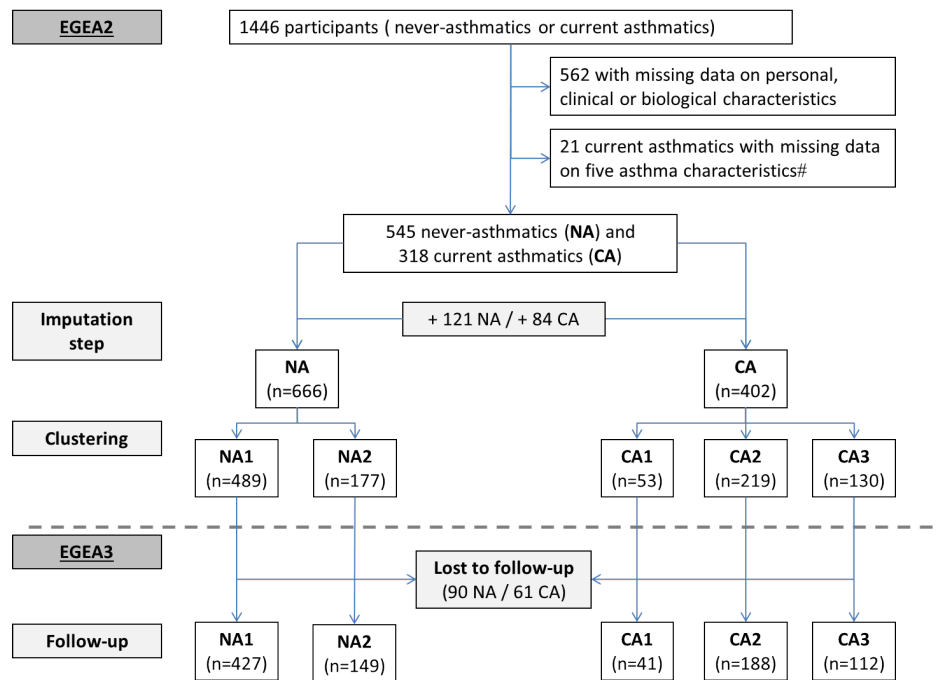


Figure 1 Flow chart of the participants included in the cross-sectional (EGEA2) and in the longitudinal analyses (EGEA3). EGEA, epidemiological study on the genetics and environment of asthma. #: the five asthma characteristics that are missing data among the 21 current asthmatics are: (1) age of asthma onset (continuous), (2) asthma attacks in the last 12 months, (3) hospital or (4) emergency admissions in the last 12 months and (5) use of oral medicines because of breathing problems in the last 12 months.

of ‘chronic cough or phlegm’, or each of the five respiratory symptoms that compose the symptom score instead of the symptom score itself). We excluded variables that were missing for a large number of participants (eg, certain biomarkers).

None of the variables were redundant. As all variables were considered to be of equal interest, no prioritisation in the data was performed.

Herein, the list of selected characteristics used to perform the cluster analyses:

1. Demographic characteristics: age (continuous), sex, current smoking status (non-smokers, ex-smokers or current smokers) and body mass index (BMI, continuous).
2. Clinical/functional characteristics: respiratory symptoms in the last 12 months (shortness of breath and wheezing, attacks of breathlessness following strenuous activity, at rest or at night time, nocturnal chest tightness and cough, chronic cough and phlegm), dyspnoea (severity grade 3, Medical Research Council scale), skin prick test positivity for at least one of 12 aeroallergens (indoor: cat, *Dermatophagoides pteronyssinus*, *Blattella germanica*, outdoor: olive, birch, *Parietaria judaica*, timothy grass, *Cupressus* and ragweed pollen, and moulds: *Aspergillus*, *Cladosporium herbarum*, *Alternaria tenuis*), current rhinitis, ever eczema and use of inhaled medicines because of breathing problems in the past 12 months, forced expiratory volume in 1 s (FEV_1 , continuous) and forced vital capacity (FVC, continuous) measured with spirometry.

For asthmatics, the five supplementary asthma specific characteristics were: (1) age of asthma onset (continuous), (2) asthma attacks in the last 12 months, (3) hospital or (4) emergency admissions for asthma in the last 12 months and (5) use of oral medicines because of breathing problems in the last 12 months.

3. Biological characteristics: blood neutrophil and eosinophil counts, total serum IgE level, and plasma FLOPs level, all expressed as continuous.

In summary, 23 characteristics were included in the cluster analysis for 666 NA and 28 characteristics for 402 asthmatics, respectively (online supplemental table E1). Detailed definitions of all demographic, clinical, functional and biological characteristics at EGEA2 and at follow-up (EGEA3) are provided in the online supplemental materials.

Clinical and biological characteristics not included in the cluster analysis

Clinical and biological characteristics not included in the cluster analyses were compared between endotypes after the identification of the latter. These characteristics are combination of variables included in the cluster analysis: the asthma symptom score and the asthma control, categorisation of a continuous variables: $FEV_1 < 80\%$, or characteristics available only in subsamples: airway hyperresponsiveness, 8-isoprostanes in exhaled breath condensate, serum interleukin (IL) among which the T2 IL-5, IL-6, IL-10, IL-13 cytokines and high-sensitivity C reactive



protein (hs-CRP). Details on their definitions and availabilities are provided in the online supplemental materials.

Cluster analysis and statistical methods

To deal with both continuous and categorical data, we relied on factor analysis of mixed data (FAMD) which combines principal components analysis for continuous variables and a multiple correspondence analysis for qualitative variables.²¹ FAMD converts all variables into continuous components that are not correlated.

We first performed an imputation step. The {missMDA} package was used to impute missing data on smoking in NA, rhinitis in asthmatics, and ever eczema, ICS use, allergic sensitisation and total IgE level in both groups.²² Then, we used the mixture approach²³ to identify clusters separately among CA and NA. Specifically, we relied on Gaussian mixture models for their flexibility, and the availability of variety of covariance structures that can be obtained by means of an Eigen decomposition. We used the {mclust} package which provides a comprehensive strategy for clustering and density estimation, and an integrated approach to finite mixture models with functions that combine model-based hierarchical clustering, expectation-maximisation for mixture estimation and several tools for model selection.²⁴ The selections of the number of mixture components (clusters) and of the covariance parameterisation for each cluster have been addressed by the Bayes information criterion.^{25 26} As pointed out by Chang,²⁷ the structure in clusters can appear in any dimension, even in those with smallest eigenvalues; we therefore included all the components obtained by FAMD in the cluster analysis. Therefore, a discriminant analysis to specify which variables had the best discriminatory value is not required. Each participant was assigned to the group whose probability of belonging was the highest. All cluster analyses were performed using R statistical software (V.3.5).

To test the robustness of our results, we also used the same mixture approach to identify the endotypes among participants without missing data, respectively, 318 CA and 545 NA (figure 1 and online supplemental materials).

Standard statistical tests including χ^2 exact test, variance analyses and Scheffe's test were performed at EGEA2 to compare the demographic, clinical and biological characteristics across endotypes among NA and asthmatics separately. Multiple regression models considering age (continuous), sex, smoking status (never smokers, ex-smokers or current smokers) at EGEA2 as potential confounding factors were used to compare the clinical characteristics at EGEA3 according to the endotype identified at EGEA2. Models with further adjustment for BMI were also performed as sensitivity analysis. NA1 was the reference group among NA and CA2 the reference group among the asthmatics. Due to the familial aggregation of the data, multivariate analyses were conducted using generalised estimated equations to take into account

dependence between observations. All statistical analyses were performed using SAS software (V.9.4; SAS Institute).

RESULTS

At EGEA2, 1068 participants were included in the analysis (figure 1). They were more often women, reported less chronic cough and phlegm, more eczema, and had lower IgE level and eosinophil count than the 378 participants not included in the analyses (online supplemental table E2). The two groups did not differ for age, BMI, other respiratory symptoms, lung function, allergic sensitisation, blood neutrophil counts and FLOPS level.

Table 1 describes the characteristics of the 402 asthmatics (CA) and 666 NA. CA were 39 years old on average, half of them were men, 24% were current smokers, reported often shortness of breath and wheezing in the last 12 months, nocturnal symptoms, use of inhaled corticosteroids in the last 12 months, and 83% of them had allergic sensitisation. NA were 47 years old on average, less than half of them were men, 21% were current smokers, 27% reported attacks of breathlessness following strenuous activity and nocturnal cough in the last 12 months, and 36% had allergic sensitisation.

Identification of distinct endotypes among NA and CA at EGEA2

The cluster analysis identified three distinct endotypes among CA (table 2, online supplemental figure E1): CA1 included 53 asthmatics predominantly characterised by adult-onset asthma with poor lung function, use of asthma treatments, cough and phlegm, asthma exacerbations, high neutrophil count and high FLOPs level; CA2 included 219 asthmatics predominantly with rhinitis and low IgE level; CA3 included 130 asthmatics predominantly young men, with childhood-onset asthma, allergic sensitisation and high IgE level. In particular, we observed gradual decreases in age, age of asthma onset, and neutrophil count, and in frequencies of shortness of breath and wheezing, asthma attacks, attacks of breathlessness at rest, nocturnal symptoms, treatments, chronic cough and phlegm, dyspnoea, and gradual increases in current smokers and allergic sensitisation frequencies from CA1 to CA3. More specifically, the CA1 cluster had statistically significantly higher BMI, lower FEV₁ and FVC % predicted, and higher FLOPs level and frequencies of chronic cough, phlegm and dyspnoea than CA2 and CA3, whereas the CA3 cluster had statistically significantly higher IgE level and higher frequencies of current smokers and allergic sensitisation, and lower frequency of attacks of breathlessness at rest than CA1 and CA2. Regarding the clinical characteristics not included in the cluster analysis, CA1 had a higher asthma symptom score and a poorer asthma control than CA2 and CA3, and CA2 had high airway hyperresponsiveness (online supplemental table E3). Differences were also observed between the three clusters for hs-CRP, leptin, IL-1Ra, IL-5, IL-7, IL-8, IL-10, IL-13 and TNF- α levels (figure 2).

Table 1 Description of the characteristics included in the cluster analysis at EGEA2 in current asthmatics and in never asthmatics

	Current asthmatics (CA, n=402)	Never asthmatics(NA, n=666)
Age, year	39.3 (16.5)	46.7 (15.8)
Sex, women	199 (49.5%)	371 (55.7%)
Smoking habits		
Non-smokers	201 (50.0%)	331 (49.7%)
Ex-smokers	104 (25.9%)	195 (29.3%)
Smokers	97 (24.1%)	140 (21.0%)
Body mass index, kg/m ²	24.3 (4.4)	24.6 (4.0)
Age of asthma onset	15.1 (14.9)	/
FEV ₁ % predicted	94.4 (18.4)	107 (16.4)
FVC % predicted	106 (16.4)	112 (17.6)
Shortness of breath and wheezing (last 12 months)	242 (60.2%)	38 (5.70%)
Asthma attacks (last 12 months)	142 (45.3%)	/
Attacks of breathlessness at rest, (last 12 months)	115 (28.6%)	18 (2.70%)
Attacks of breathlessness following strenuous activity (last 12 months)	255 (63.4%)	180 (27.0%)
Nocturnal symptoms (last 12 months)		
Cough	164 (40.8%)	177 (26.6%)
Chest tightness	220 (54.7%)	10 (1.50%)
Shortness of breath	95 (23.6%)	83 (12.5%)
Hospital admissions, (last 12 months)	4 (1.0%)	/
Emergency admissions (last 12 months)	9 (2.2%)	/
Inhaled corticosteroids (last 12 months)	200 (49.8%)	20 (3.00%)
Oral corticosteroids (last 12 months)	62 (15.4%)	/
Chronic cough	51 (12.7%)	30 (4.95%)
Chronic phlegm	45 (11.2%)	28 (4.20%)
Dyspnoea grade 3	74 (18.4%)	40 (6.00%)
Skin prick test positivity	333 (82.8%)	237 (35.6%)
Rhinitis (last 12 months)	243 (60.4%)	134 (20.1%)
Eczema ever	205 (51.0%)	180 (27.0%)
Total IgE, IU/mL, GM (Q1–Q3)	151 (65.4–385)	43.6 (16.4–115)
White blood cell counts		
Eosinophils/mm ³	262 (202)	162 (118)
Neutrophils/mm ³	4087 (1515)	3971 (1334)
Fluorescent oxidation products, RFU/mL GM (Q1–Q3)	92.2 (79.1–103)	94.7 (81.7–108)

Data are means (SD) or n (%) unless otherwise stated.

Dyspnoea grade 3 was defined according to the Medical Research Council scale. Skin prick test positivity was defined by a mean weal diameter ≥ 3 mm than the negative control for at least one of 12 aeroallergens.

FEV₁, forced expiratory volume in 1 s; FVC, forced vital capacity; GM, geometric mean; (Q1–Q3), first and third quartiles; RFU, relative fluorescence intensity.

In particular, gradual decreases in leptin and in hs-CRP levels were observed from CA1 to CA3, and CA3 had lower hs-CRP level than CA1 and CA2 ($p < 0.05$, Scheffé's test). Furthermore, the levels of IL-1Ra, IL-5, IL-7, IL-8, IL-10, IL-13 and TNF- α were significantly lower in

CA3 as compared with CA2 whereas no differences were observed between CA1 and CA2.

Some overlaps exist between the CA endotypes and the four phenotypes identified by latent class analysis at EGEA2¹⁸ (online supplemental table E4): CA1 with

Table 2 Description of the characteristics included in the cluster analysis (EGEA2) according to each current asthmatics (CA) endotype

	CA 1 (n=53)	CA 2 (n=219)	CA 3 (n=130)	P value
Age, year	53.3 (15.9)	40.8 (15.7)	31.1 (13.1)	<0.0001
Sex, women	29 (54.7%)	121 (55.2%)	49 (37.7%)	0.005
Smoking habits				<0.0001
Non-smokers	26 (49.1%)	99 (45.2%)	76 (58.5%)	
Ex-smokers	21 (39.6%)	70 (32.0%)	13 (10.0%)	
Smokers	6 (11.3%)	50 (22.8%)	41 (31.5%)	
Body mass index kg/m ²	27.0 (5.28)	24.6 (4.24)	22.8 (3.70)	<0.0001
Age of asthma onset	26.5 (17.6)	15.4 (14.8)	9.96 (10.7)	<0.0001
FEV ₁ % predicted	76 (21)	97 (18)	98 (15)	<0.0001
FVC % predicted	95.3 (19.9)	110 (16)	106 (13)	<0.0001
Shortness of breath and wheezing (last 12 months)	44 (83.0%)	141 (64.4%)	57 (43.9%)	<0.0001
Asthma attacks (last 12 months)	40 (75.5%)	112 (51.1%)	30 (23.1%)	<0.0001
Attacks of breathlessness at rest (last 12 months)	26 (49.1%)	72 (32.9%)	17 (13.1%)	<0.0001
Attacks of breathlessness following strenuous activity (last 12 months)	38 (71.7%)	142 (64.8%)	75 (57.7%)	0.17
Nocturnal symptoms (last 12 months)				
Cough	32 (60.4%)	107 (48.9%)	25 (19.2%)	<0.0001
Chest tightness	39 (73.6%)	146 (66.7%)	35 (26.9%)	<0.0001
Shortness of breath	24 (45.3%)	63 (28.8%)	8 (6.2%)	<0.0001
Hospital admissions (last 12 months)	3 (0.57%)	1 (0.46%)	0 (0%)	0.005
Emergency admissions (last 12 months)	7 (13.2%)	2 (0.91%)	0 (0%)	<0.0001
Inhaled corticosteroids (last 12 months)	43 (81.1%)	116 (53.0%)	41 (31.5%)	<0.0001
Oral corticosteroids (last 12 months)	20 (37.7%)	37 (16.9%)	5 (3.9%)	<0.0001
Chronic cough	22 (41.5%)	25 (11.4%)	4 (3.08%)	<0.0001
Chronic phlegm	16 (30.2%)	24 (10.9%)	5 (3.85%)	<0.0001
Dyspnoea grade 3	29 (54.7%)	33 (15.1%)	12 (9.2%)	<0.0001
Skin prick test positivity	37 (69.8%)	174 (79.5%)	122 (93.8%)	<0.0001
Rhinitis (last 12 months)	29 (54.7%)	147 (67.1%)	67 (51.5%)	0.01
Eczema (ever)	20 (37.7%)	110 (50.2%)	75 (57.7%)	0.047
Total IgE, IU/ml, GM (Q1 – Q3)	136 (63.8–372)	132 (48.0–314)	198 (95.2–531)	0.03
White blood cell counts				
Eosinophils/mm ³	293 (314)	258 (179)	257 (183)	0.5
Neutrophils/mm ³	4968 (1646)	4183 (1583)	3567 (1095)	<0.0001
Fluorescent oxidation products, RFU/mL, GM (Q1 – Q3)	105 (93.3–117)	91.9 (79.2–105)	87.7 (76.4–98.4)	<0.0001

Data are means (SD) or n (%) unless otherwise stated.

Dyspnoea grade 3 was defined according to the Medical Research Council scale.

Skin Prick Test positivity was defined by a mean weal diameter ≥ 3 mm than the negative control for at least one of 12 aeroallergens.

FEV₁, forced expiratory volume in 1 s; FVC, forced vital capacity; RFU, relative fluorescence intensity; GM, geometric mean; (Q1–Q3), first and third quartiles.

‘active treated adult-onset asthma’ and ‘active treated allergic childhood-onset asthma’, CA2 with ‘active treated allergic childhood-onset asthma’ and ‘inactive/mild untreated adult-onset asthma’ and CA3 with ‘inactive/mild untreated allergic childhood asthma’ and ‘active treated allergic childhood asthma’.

The cluster analysis applied on the 666 NA identified two endotypes (table 3, online supplemental figure E2): NA1 with 489 NA predominantly asymptomatic and NA2 characterised by 177 NA predominantly current smokers, with high BMI, reporting respiratory symptoms including shortness of breath and wheezing, nocturnal symptoms,

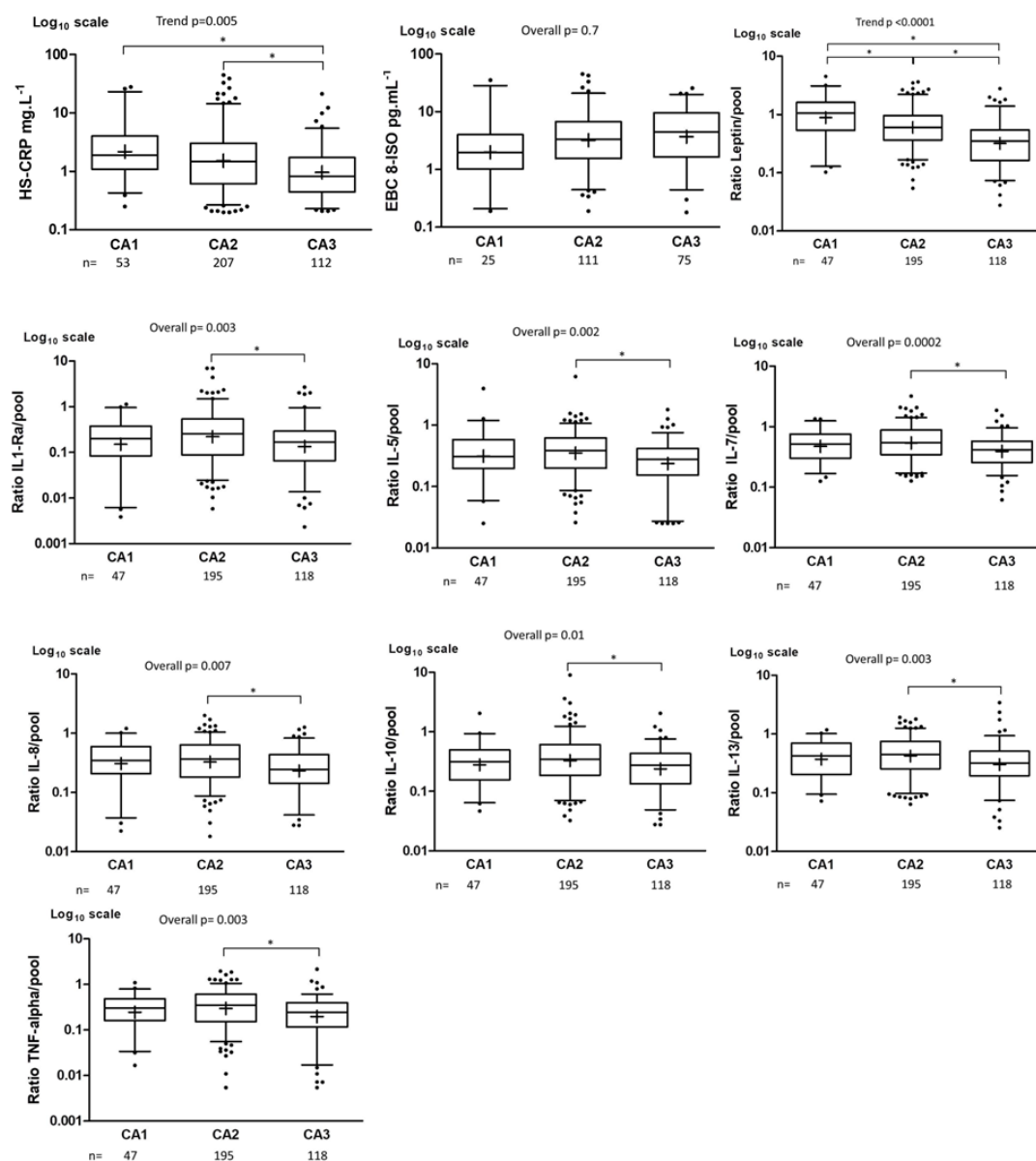


Figure 2 Box plots of hs-CRP, 8-isoprostanes, leptin and seven cytokines plotted in CA1, CA2 and CA3 endotypes among asthmatics. The plots show the median (bar), the first and third quartiles (box), the 1st and 99th percentiles (whiskers) and the outliers (.) for each endotype. P values are adjusted for age, gender and medication use (* $p < 0.05$, Scheffe's test). CA, current asthma.

chronic cough and phlegm and dyspnoea, use of inhaled corticosteroids in the last 12 months and high neutrophil and eosinophil counts.

Regarding the clinical characteristics not included in the cluster analysis, NA2 had a high asthma symptom score, poor lung function, high airway hyper-responsiveness (online supplemental table E5), and high leptin and hs-CRP levels (online supplemental figure E3) as compared with NA1.

The distributions of the clusters according to the centres are provided in the supplementary materials (online supplemental tables E6 and E7).

Sensibility analyses performed among participants without missing data by using the same mixture model

approach showed similar results (online supplemental tables E8 and E9).

Comparison of the clinical characteristics at EGEA3 according to the endotype identified at EGEA2

Sixty-one asthmatics and 90 NA (14%) were lost to follow-up at EGEA3. In comparison to those not followed up, the 341 followed-up asthmatics were more often women, reported more often shortness of breath and wheezing and rhinitis, and had better lung function and lower blood neutrophil count (online supplemental table E10). In comparison to those not followed up, the 576 followed-up NA were older, more often women,

Table 3 Description of the characteristics included in the cluster analysis at EGEA2 according to each never asthmatics (NA) endotype

	NA1 (n=489)	NA2 (n=177)	P value
Age, year	46.3 (15.5)	47.8 (16.6)	<0.0001
Sex, women	262 (53.6%)	109 (61.6%)	0.07
Smoking habits			0.005
Non-smokers	249 (50.9%)	82 (46.3%)	
Ex-smokers	152 (31.1%)	43 (24.3%)	
Smokers	88 (18.0%)	52 (29.4%)	
Body mass index, kg/m ²	24.3 (3.82)	25.4 (4.39)	0.002
FEV ₁ % predicted	108 (16.0)	105 (17.4)	0.04
FVC % predicted	113 (17.5)	111 (17.7)	0.13
Shortness of breath and wheezing (last 12 months)	2 (0.41%)	36 (20.3%)	<0.0001
Attacks of breathlessness at rest (last 12 months)	3 (0.61%)	15 (8.47%)	<0.0001
Attacks of breathlessness following strenuous activity, (last 12 months)	104 (21.3%)	76 (42.9%)	<0.0001
Nocturnal symptoms (last 12 months)			
Cough	103 (21.14%)	74 (41.8%)	<0.0001
Chest tightness	2 (0.41%)	8 (4.52%)	0.0006
Shortness of breath	51 (10.4%)	32 (18.1%)	0.008
Inhaled corticosteroids (last 12 months)	1 (0.20%)	19 (10.7%)	<0.0001
Chronic cough	1 (0.20%)	32 (18.1%)	<0.0001
Chronic phlegm	0 (0%)	28 (15.8%)	<0.0001
Dyspnoea grade 3	4 (0.82%)	36 (20.3%)	<0.0001
Skin prick test positivity	175 (35.8%)	62 (35.0%)	0.9
Rhinitis (last 12 months)	84 (17.2%)	50 (28.2%)	0.002
Eczema (ever)	133 (27.2%)	47 (26.6%)	0.9
Total IgE, IU/ml, GM (Q1–Q3)	42.0 (16.4–107)	52.6 (16.4–135)	0.07
White blood cell counts			
Eosinophils/mm ³	156 (108)	180 (139)	0.04
Neutrophils/mm ³	3907 (1269)	4150 (1488)	0.02
Fluorescent oxidation products, RFU/mL, GM (Q1–Q3)	94.0 (81.4–108)	96.6 (82.1–109)	0.16

Data are means (SD) or n (%) unless otherwise stated.

Dyspnoea grade 3 was defined according to the Medical Research Council scale.

Skin Prick Test positivity was defined by a mean weal diameter ≥ 3 mm than the negative control for at least one of 12 aeroallergens.

FEV₁, forced expiratory volume in 1 s; FVC, forced vital capacity; GM, geometric mean; (Q1–Q3), first and third quartiles; RFU, relative fluorescence intensity.

and reported less often be awakened by cough (online supplemental table E11).

Characteristics of the participants at EGEA3 according to their endotype at EGEA2 are presented in online supplemental tables E12 and E13).

The comparison of the clinical characteristics at EGEA3 according to the endotype identified at EGEA2 showed that asthmatics in CA1 had higher asthma symptom score and poorer asthma control, reported more often nocturnal symptoms and dyspnoea as compared with CA2 (table 4). Conversely, CA3 had lower asthma symptom score, better asthma control and reported less often asthma attacks, exacerbations, nocturnal symptoms and

dyspnoea at follow-up as compared with CA2. Similarly, NA2 reported more often respiratory symptoms and had higher rate of incident asthma than NA1 (table 5).

DISCUSSION

This study aimed at identifying distinct adult respiratory endotypes based on demographic, clinical, functional and blood biological characteristics by using a cluster analysis for mixed data. We identified five endotypes: three among asthmatics and two among NA with distinct clinical, functional and biological characteristics. Comparison of the clinical characteristics at follow-up

Table 4 Comparison of the clinical characteristics at EGEA3 respective to the current asthma (CA) endotypes identified at EGEA2

	Number (No.)	CA1 versus CA2		CA3 versus CA2	
		No. events	OR (95% CI)	No. events	OR (95% CI)
Asthma attacks (last 12 months)	314	20/134	1.08 (0.54 to 2.16)	27/134	0.35 (0.19 to 0.62)
Exacerbations (last 12 months)	258	13/65	2.19 (0.84 to 5.70)	8/65	0.28 (0.12 to 0.65)
Asthma symptom score, last 12 months (0–5) (score more than 2 vs less or equal to 2)	313	16/100	2.23 (1.02 to 4.91)	15/100	0.20 (0.10 to 0.39)
Nocturnal symptoms (last 12 months)					
Cough	275	22/116	3.48 (1.40 to 8.66)	32/116	0.79 (0.46 to 1.37)
Chest tightness	278	19/114	2.65 (1.09 to 6.48)	28/114	0.41 (0.23 to 0.73)
Shortness of breath	281	13/46	4.16 (1.64 to 10.5)	9/46	0.40 (0.16 to 1.00)
Asthma control test (last 4 weeks) (score more or equal to 20 vs less than 20)	282	17/58	3.43 (1.54 to 7.61)	9/58	0.42 (0.50 to 0.97)
Dyspnoea grade 3	328	19/72	3.52 (1.64 to 7.57)	12/72	0.46 (0.22 to 0.97)

Data are expressed as adjusted OR with a 95% CI.

Endotype CA2 was the reference (OR=1).

OR (95% CI) were adjusted on age, sex and current smoking status at EGEA2, taking into account familial dependence of the participants.

*†‡§Results did change after further adjustment on BMI for asthma symptom score OR=2.18 (0.98 to 4.82), shortness of breath OR=0.43 (0.17 to 1.08), asthma control test OR=0.45 (0.19 to 1.02), and dyspnoea OR=0.51 (0.24 to 1.09).

Dyspnoea grade 3 was defined according to the Medical Research Council scale.

according to the endotype identified at baseline showed different clinical outcomes.

The main strength of our study was the simultaneous inclusion of various asthma characteristics and of neutrophil, eosinophil and FLOPs, that are key biological markers related both to asthma and to inflammatory or oxidative stress pathways. We did not transform any continuous biological data into binary data: no consensus exists on the best cut-off to choose for white blood cell

counts and FLOPs level, and a priori cut-offs may lead to loss information and bias the results. Most of the asthmatics were recruited in chest clinics as asthma cases, with careful procedures set up to include true asthmatics, and others were recruited as first-degree relatives of asthmatic cases. This design leads to the recruitment of participants with wide range of asthma severity/control. No evident follow-up bias related to asthma status and asthma-related phenotypes was shown, and among

Table 5 Comparison of the clinical characteristics at EGEA3 respective to the never asthmatics (NA) endotypes identified at EGEA2

	Number (No.)	NA2 vs NA1	
		No. events	OR (95% CI)
New onset-asthma	574	14/31	2.34 (1.09 to 5.02)
Wheezing, last 12 months	572	19/48	1.69 (0.95 to 3.02)
Attacks of breathlessness at rest, (last 12 months)	569	10/26	1.91 (0.87 to 4.24)
Attacks of breathlessness following strenuous activity, (last 12 months)	549	48/134	1.66 (1.07 to 2.57)
Nocturnal symptoms (last 12 months), cough	485	33/103	1.44 (0.90 to 2.30)
Asthma symptom score, last 12 months (0–5) (score more than 2 vs less or equal to 2)	535	9/17	2.68 (1.10 to 6.53)
Dyspnoea grade 3	563	25/51	2.78 (1.46 to 5.09)
Rhinitis, last 12 months	565	69/130	2.09 (1.42 to 3.09)

Data are expressed as adjusted OR with a 95% CI.

Endotype NA1 was the reference (OR=1).

OR (95% CI) were adjusted on age, sex, and current smoking status at EGEA2, taking into account familial dependence of the participants.

Result did change after further adjustment on body mass index for asthma symptom score OR=2.13 (0.86 to 5.29).

Dyspnoea grade 3 was defined according to the Medical Research Council scale.



asthmatics, those included in the present analyses were representative of the EGEA adult cases and their first degree relatives with asthma at inclusion. Five centres participated in the recruitment of the participants, but given the standardised protocols, it is unlikely that there was a recruitment bias in the formation of the clusters. We acknowledge that our results are not generalisable to the general population, and that the lack of replication may be seen as a weakness. However, we did not identify any epidemiological study with similar phenotypic and biological characterisation of the participants to replicate our analyses. We performed the cluster analyses before the imputation step to test the robustness of the results. We also acknowledge that the available data did not allow us to study the stability of clusters over time. We compared the endotypes identified among CA in the present study with four asthma phenotypes previously revealed by latent class analyses using part of the same data.¹⁸ Interestingly, neutrophil count, that was not initially included in the latent class analyses, which is highest in CA1 and lowest in CA3, was significantly higher among asthmatics in 'active treated adult-onset asthma' phenotype that shows the highest but not complete overlap with CA1, and was significantly lower among asthmatics in 'inactive/mild untreated allergic childhood asthma' phenotype that shows the highest but not complete overlap with CA3. These results suggest that clusters sharing similar phenotypic characteristics may have different underlying mechanisms, and highlight the interest to jointly integrate clinical and biological characteristics in cluster analysis to refine the identification.

Among asthmatics, we identified three endotypes with contrasted clinical and biological characteristics. The CA1 endotype was characterised by severe treated uncontrolled adult-onset asthma, high BMI, high neutrophil count and high FOPs, hs-CRP and leptin levels, and showed the worst clinical characteristics 10 years later. The high leptin level in CA1 can be discussed in relation with the high BMI, and the high number of blood neutrophils and FOPs level observed in this cluster. Leptin is an adipocyte-derived proinflammatory protein, and the high leptin level may be partly explained by the hypothesis that adiposity affects asthma activity through increased leptin level. Animal studies support the biological plausibility of this hypothesis: leptin administration to wild type mice resulted in increased airway inflammation,²⁸ and leptin is also known to induce production of proinflammatory cytokines, and reactive oxygen species²⁹ which are likely involved in the pathophysiology process of asthma. The CA3 endotype was characterised by well-controlled allergic childhood-onset asthma, low neutrophil count, low leptin, hs-CRP and cytokines levels, and the best clinical characteristics 10 years later. These results were consistent with previous associations observed among asthmatics at EGEA2 between high neutrophil counts or high FOPs levels and worse respiratory health.^{12 14 15} Comparisons with results from the literature are very limited: beyond studies studying severe

asthma and T2 asthma, only two studies have identified asthma endotypes by performing clustering methods based on clinical characteristics, eosinophils and neutrophils. None of them have included biomarkers related to oxidative stress. The first study identified eight endotypes among 198 patients with mild-to-severe asthma and 21 controls by using topological data analysis and Bayesian network analysis on 103 clinical, physiological and inflammatory parameters including neutrophils and eosinophils in induced sputum.³⁰ Interestingly, the neutrophilic cluster (n=9) was characterised by poor lung function, a result coherent with the characteristics of CA1. The second study identified 6 endotypes among 100 participants with mild/moderate to severe asthma by using the same method on 21 characteristics among which blood eosinophils and neutrophils expressed as continuous data.³¹ In contrast to our study, the authors reported that neutrophil count did not play a significant role in the characterisation of the clusters. Low IL-5, IL-10 and IL-13 levels were one of the characteristics of the CA3 endotype, but no differences in the cytokines levels were found between CA2 and CA1. Even if cytokine measurements were not available for all the participants of each endotype, our results suggest that the key T2 cytokines would not be helpful to distinguish between these endotypes. The largest studies (100 to 726 participants) that have been conducted to identify asthma phenotypes/endotypes using unsupervised computational modelling of many clinical features and biomarkers are the Severe Asthma Research Program (SARP) in the USA, the Leicester study conducted in the UK, the unbiased Biomarkers in Prediction of respiratory Disease Outcome (U-BIOPRED) study and the transcriptomic endotypes of asthma (TEA) study.³² Despite significant differences in these cohorts such as disease features, computational approaches, and the number of clusters identified, four clusters have been consistently reported across studies, and two of which are also found in our study. The first one is characterised by late-onset asthma, low lung function, no allergic status, sometimes obesity and non-T2 inflammation named 'neutrophilic cluster' (SARP cluster 5, U-BIOPRED cluster 4, UK cluster 2 or TEA cluster 2) which are the main characteristics of CA1. The second one is characterised by early-onset asthma, preserved lung function, and allergic status (SARP cluster 1 and 2, U-BIOPRED cluster 1, UK cluster 3 or TEA cluster 3) which are the main characteristics of CA3. To date, studies often focused on induced sputum, and on eosinophils arguing that they are related to T2 inflammation, but this is not always synonymous with such endotype.³ Our results support the T2 versus non-T2 asthma mechanistic paradigm, and suggest that neutrophils should also be integrated in future cluster analyses to identify distinct endotypes.

By applying the same cluster analysis among NA, we identified two endotypes: NA2 was characterised by worse respiratory health, high blood neutrophil and eosinophil counts, high leptin and hs-CRP levels, worse asthma

symptoms and higher rate of ‘new-onset asthma’ 10 years later as compared with ‘asymptomatic’ NA1. Even if NA2 endotype included participants who did not fulfil our asthma definition, they shared some of the respiratory and biological characteristics of asthma endotypes, especially those of CA1. We acknowledge that we have no way to discern if these participants could have undiagnosed asthma, chronic bronchitis and/or with ‘asthma COPD overlap syndrome’. These preliminary results highlight the interest to consider the continuum of the disease. Identifying respiratory endotypes among ‘respiratory healthy’ adults should be further investigated, and the impact of treating the symptomatic adults to prevent the development of respiratory diseases should be studied.

Identifying relevant asthma endotypes is a challenging issue. There are numerous clustering methods that have been used to identify asthma endotypes, each with its strengths and weaknesses,^{16 17} and their results are influenced by choice of variables, their encoding/categorisation and transformation, and choice of statistical method.³³ The mixture models are able to combine continuous and qualitative data in a unified framework. Our strategy for selecting biological markers was linked to the idea that relying on a single biomarker to identify asthma endotypes is not realistic. An ‘omics’ approach in first line may not be realistic too if we consider that ‘omics’ may be viewed as several sets of variables that could be grouped together based on biological processes or pathways. The choice of eosinophils and neutrophils was based on their bottom role to orchestrate immune responses, and the choice of FLOPs as they reflect a mixture of oxidation products from DNA, proteins and lipids¹³ and were associated with asthma activity.¹⁴ They therefore could be used as first step in a more complex strategy to identify endotypes.

In summary, by adding key biomarkers to an extensive characterisation of respiratory health, we identified distinct respiratory endotypes both among asthmatics and NA. In the future, identifying asthma endotypes could follow a two-step strategy: first identifying asthma endotypes by using clinical, physiological and key biomarkers that orchestrate or reflect the main ‘pathways’ in asthma, found to be robust across studies and clustering methods, and second by using high throughput ‘omics’ platforms to identify biomarkers in systemic and lung compartments related to these endotypes. This strategy could be helpful to open the way to study and develop new biologics that will improve outcomes with non-eosinophilic or T2-low asthma.

Author affiliations

¹Université Paris-Saclay, UVSQ, Univ. Paris-Sud, INSERM, Equipe d’Epidémiologie Respiratoire Intégrative, CESP, 94807 Villejuif, France

²Université de Paris, LIPADE, Paris, France

³CHU de Lyon, Pneumology Department, Lyon, France

⁴Service d’Allergologie, APHP, Hôpital Trousseau, Sorbonne Université, Paris, France

⁵Univ. Grenoble Alpes, INSERM, CNRS, Team of Environmental Epidemiology Applied to Reproduction and Respiratory Health, IAB, 38000 Grenoble, France

⁶CHU de Grenoble-Alpes, Pédiatrie, Grenoble, France

⁷Université de Lille Nord de France, Lille, France

⁸CHU de Lille, Laboratoire de Biochimie et Biologie Moléculaire, Pôle de Biologie Pathologie Génétique, Lille, France

⁹Université de Paris, CNRS, Centre Borelli, 75005 Paris, France

Acknowledgements The authors thank all those who participated to the setting of the study and on the various aspects of the examinations involved: interviewers, technicians for lung function testing and skin prick tests, blood sampling, IgE determinations, coders, those involved in quality control, data and sample management and all those who supervised the study in all centers. The authors are grateful to the three CIC-Inserm of Necker, Grenoble and Marseille who supported the study and in which participants were examined, and to the biobanks in Lille (CIC-Inserm), and at Annemasse (Etablissement français du sang) where biological samples are stored. They are indebted to all the participating individuals without whom the study would not have been possible. The authors are also grateful to Samir Nadif for his help in formatting all the tables.

EGEA cooperative group members as follows: Coordination: V Siroux (epidemiology, PI since 2013); F Demenais (genetics); I Pin (clinical aspects); R Nadif (biology); F Kauffmann (PI 1992-2012). Respiratory epidemiology: Inserm ex-U 700, Paris: M Korobaëff (Egea1), F Neukirch (Egea1); Inserm ex-U 707, Paris: I Annesi-Maesano (Egea1-2); Inserm U 1018, Villejuif: O Dumas, F Kauffmann, N Le Moual, R Nadif, MP Oryszczyn (Egea1-2), R Varraso; Inserm U 1209 Grenoble: J Lepeule, V Siroux. Genetics: Inserm ex-U 393, Paris: J Feingold; Inserm UMR 1124, Paris: E Bouzigon, MH Dizier, F Demenais; CNG, Evry: I Gut (now CNAG, Barcelona, Spain), M Lathrop (now Univ McGill, Montreal, Canada). Clinical centers: Grenoble: I Pin, C Pison; Lyon: D Ecochard (Egea1), F Gormand, Y Pacheco; Marseille: D Charpin (Egea1), D Vervloet (Egea1-2); Montpellier: P Bousquet; Paris Cochin: A Lockhart (Egea1), R Matran (now in Lille); Paris Necker: E Paty (Egea1-2), P Scheinmann (Egea1-2); Paris-Trousseau: A Grimfeld (Egea1-2), J Just. Data management and quality: Inserm ex-U155, Paris: J Hochez (Egea1); Inserm U 1018, Villejuif: N Le Moual, L Orsi; Inserm ex-U780, Villejuif: C Ravault (Egea1-2); Inserm ex-U794, Evry: N Chateigner (Egea1-2); Inserm UMR 1124, Paris: H Mohamdi; Inserm U1209, Grenoble: A Boudier, J Quentin (Egea1-2).

Contributors RN, OD and MN designed and conducted the research. NLM, FG, JJ, IP, RM and VS provided essential materials. RN, MF, MVA and MN analysed data and performed statistical analyses. RN, OD and MN wrote the manuscript and had primary responsibility for final content. All authors read, edited and approved the final manuscript.

Funding The French Agency for Food, Environmental and Occupational Health & Safety (ANSES PNR-EST 2017), the Région Hauts de France, National Hospital program of clinical research (PHRC-National 2012, EvAdA), ANR-CES-2009, the Fonds AGIR pour les maladies chroniques.

Competing interests IP reports personal fees from AGIRadom, and other fees from ASTRA ZENECA and NOVARTIS outside the submitted work. JJ reports personal fees from ALK and Thermofischer, and grants and personal fees from Novartis and Astra Zeneca outside the submitted work.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement No data are available. Due to third party restrictions, EGEA data are not publicly available. See the following URL for more information: <https://egeanet.vjf.inserm.fr/index.php/en/contacts-en>. Interested researchers should contact Egea.cohorte@inserm.fr with further questions regarding data access.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Rachel Nadif <http://orcid.org/0000-0003-4938-9339>

Miora Valérie Andrianjafimasy <http://orcid.org/0000-0002-5791-0772>
 Orianne Dumas <http://orcid.org/0000-0001-8423-2826>

REFERENCES

- 1 Pavord ID, Beasley R, Agusti A, *et al.* After asthma: redefining airways diseases. *Lancet* 2018;391:350–400.
- 2 Chung KF, Adcock IM. Precision medicine for the discovery of treatable mechanisms in severe asthma. *Allergy* 2019;74:1649–59.
- 3 Bush A. Pathophysiological mechanisms of asthma. *Front Pediatr* 2019;7:68.
- 4 McGregor MC, Krings JG, Nair P, *et al.* Role of biologics in asthma. *Am J Respir Crit Care Med* 2019;199:433–45.
- 5 Chupp GL, Kaur R, Mainardi A. New therapies for emerging Endotypes of asthma. *Annu Rev Med* 2020;71:289–302.
- 6 Diamant Z, Vijverberg S, Alving K, *et al.* Toward clinically applicable biomarkers for asthma: an EAACI position paper. *Allergy* 2019;74:1835–51.
- 7 Carr TF, Zeki AA, Kraft M. Eosinophilic and Noneosinophilic asthma. *Am J Respir Crit Care Med* 2018;197:22–37.
- 8 Mishra V, Banga J, Silveyra P. Oxidative stress and cellular pathways of asthma and inflammation: therapeutic strategies and pharmacological targets. *Pharmacol Ther* 2018;181:169–82.
- 9 Wenzel SE. Asthma phenotypes: the evolution from clinical to molecular approaches. *Nat Med* 2012;18:716–25.
- 10 Lambrecht BN, Hammad H. The immunology of asthma. *Nat Immunol* 2015;16:45–56.
- 11 Nadif R, Siroux V, Oryszczyn M-P, *et al.* Heterogeneity of asthma according to blood inflammatory patterns. *Thorax* 2009;64:374–80.
- 12 Nadif R, Siroux V, Boudier A, *et al.* Blood granulocyte patterns as predictors of asthma phenotypes in adults from the EGEEA study. *Eur Respir J* 2016;48:1040–51.
- 13 Wu T, Willett WC, Rifai N, *et al.* Plasma fluorescent oxidation products as potential markers of oxidative stress for epidemiologic studies. *Am J Epidemiol* 2007;166:552–60.
- 14 Andrianjafimasy M, Zerimech F, Akiki Z, *et al.* Oxidative stress biomarkers and asthma characteristics in adults of the EGEEA study. *Eur Respir J* 2017;50. doi:10.1183/13993003.01193-2017. [Epub ahead of print: 28 Dec 2017].
- 15 Akiki Z, Andrianjafimasy M, Zerimech F, *et al.* High level of fluorescent oxidation products and worsening of asthma control over time. *Respir Res* 2019;20:203.
- 16 Howard R, Rattray M, Prosperi M, *et al.* Distinguishing asthma phenotypes using machine learning approaches. *Curr Allergy Asthma Rep* 2015;15:38.
- 17 Deliu M, Sperrin M, Belgrave D, *et al.* Identification of asthma subtypes using clustering methodologies. *Pulm Ther* 2016;2:19–41.
- 18 Siroux V, Basagaña X, Boudier A, *et al.* Identifying adult asthma phenotypes using a clustering approach. *Eur Respir J* 2011;38:310–7.
- 19 Kauffmann F, Dizier MH, Pin I, *et al.* Epidemiological study of the genetics and environment of asthma, bronchial hyperresponsiveness, and atopy: phenotype issues. *Am J Respir Crit Care Med* 1997;156:S123–9.
- 20 Kauffmann F, Dizier MH, Annesi-Maesano I, *et al.* EGEEA (Epidemiological study on the Genetics and Environment of Asthma, bronchial hyperresponsiveness and atopy)-- descriptive characteristics. *Clin Exp Allergy* 1999;29 Suppl 4:17–21.
- 21 Pagès J. *Multiple factor analysis by example using R: Chapman and Hall/CRC*, 2014.
- 22 Audigier V, Husson F, Josse J. A principal component method to impute missing values for mixed data. *Adv Data Anal Classif* 2016;10:5–26.
- 23 McLachlan G, Peel D. *Wiley series in probability and statistics. Finite Mixture Models*, 2000: 420–7.
- 24 Scrucca L, Fop M, Murphy TB, *et al.* mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *R J* 2016;8:289–317.
- 25 Schwarz G. Estimating the dimension of a model. *The Annals of Statistics* 1978;6:461–4.
- 26 Fraley C, Raftery AE. Model-Based clustering, discriminant analysis, and density estimation. *J Am Stat Assoc* 2002;97:611–31.
- 27 Chang W-C. On using principal components before separating a mixture of two multivariate normal distributions. *Appl Stat* 1983;32:267–75.
- 28 Sood A, Shore SA, Sood A. Adiponectin, leptin, and resistin in asthma: basic mechanisms through population studies. *J Allergy* 2013;2013:1–15.
- 29 Tilg H, Moschen AR. Adipocytokines: mediators linking adipose tissue, inflammation and immunity. *Nat Rev Immunol* 2006;6:772–83.
- 30 Hinks TSC, Brown T, Lau LCK, *et al.* Multidimensional endotyping in patients with severe asthma reveals inflammatory heterogeneity in matrix metalloproteinases and chitinase 3-like protein 1. *J Allergy Clin Immunol* 2016;138:61–75.
- 31 Agache I, Strasser DS, Pierlot GM, *et al.* Monitoring inflammatory heterogeneity with multiple biomarkers for multidimensional endotyping of asthma. *J Allergy Clin Immunol* 2018;141:442–5.
- 32 Kaur R, Chupp G. Phenotypes and endotypes of adult asthma: moving toward precision medicine. *J Allergy Clin Immunol* 2019;144:1–12.
- 33 Prosperi MCF, Sahiner UM, Belgrave D, *et al.* Challenges in identifying asthma subgroups using unsupervised statistical learning techniques. *Am J Respir Crit Care Med* 2013;188:1303–12.