Identifying adult asthma phenotypes using a clustering approach

V. Siroux*,#, X. Basagaña¶,+,§,ƒ, A. Boudier*,#, I. Pin*,#,**, J. Garcia-Aymerich¶,+,§,ƒ,
A. Vesin#,##, R. Slama*,#, D. Jarvis¶¶, J.M. Anto¶,+,§,ƒ,
F. Kauffmann++,§§ and J. Sunyer¶,+,§,ƒ

ABSTRACT: There is a need to improve asthma characterisation by integrating multiple aspects of the disease. The aim of the present study was to identify distinct asthma phenotypes by applying latent class analysis (LCA), a model-based clustering method, to two large epidemiological studies.

Adults with asthma who participated in the follow-up of the Epidemiological Study on the Genetics and Environment of Asthma (EGEA2) (n=641) and the European Community Respiratory Health Survey (ECRHSII) (n=1,895) were included. 19 variables covering personal characteristics, asthma symptoms, exacerbations and treatment, age of asthma onset, allergic characteristics, lung function and airway hyperresponsiveness were considered in the LCA.

Four asthma phenotypes were distinguished by the LCA in each sample. Two phenotypes were similar in EGEA2 and ECRHSII: active treated allergic childhood-onset asthma and active treated adult-onset asthma. The other two phenotypes were composed of subjects with inactive or mild untreated asthma, who differed by atopy status and age of asthma onset (childhood or adulthood). The phenotypes clearly discriminated populations in terms of quality of life, and blood eosinophil and neutrophil counts.

The LCAs revealed four distinct asthma phenotypes in each sample. Considering these more homogeneous phenotypes in future studies may lead to a better identification of risk factors for asthma.

KEYWORDS: Asthma heterogeneity, asthma phenotypes, latent class analysis

Asthma is a complex disorder that includes distinct phenotypes, potentially with different aetiologies, natural histories and responses to treatment [1]. Distinct adult asthma phenotypes have been identified for some time but have been based on a limited number of characteristics. Allergic and nonallergic asthma are probably the most commonly discussed phenotypes. Other phenotypes defined by clinical or physiological categories (*i.e.* severity, age at onset and chronic airflow obstruction), asthma triggers (*i.e.* exercise, allergens, occupational allergens or irritants) or their pathobiology (*i.e.* eosinophilic or neutrophilic asthma) have also been proposed [1]. It is expected that a comprehensive examination protocol of asthma patients incorporating several domains of the disease would make it possible to identify more distinct asthma phenotypes. Such widening of the asthma characterisation may allow a better understanding of the aetiology of asthma, by increasing the

power to detect environmental and genetic risk factors [2].

For such a purpose, multivariate statistical methods centred on the subjects (and not on the variables, as in regression analysis), such as clustering methods, have already been applied in the respiratory epidemiology field [3–7] and have recently been described as "steps in the right direction" [8]. This approach, applied to populations of adult asthma patients, identified asthma phenotypes that exhibited differences in clinical response to treatment [4], and clinical, physiological and inflammatory parameters [5]. Latent class analysis (LCA), a clustering model-based method, has been applied in two populations of children from the general population and identified several wheezing phenotypes [3, 6]. These approaches have never been applied in adults with asthma from population-based studies that, compared with the clinical population, are

---

This article has supplementary material available from www.erj.ersjournals.com

expected to cover a larger range of asthma phenotypes, by including patients with current and remittent asthma.

The aim of the present study was to identify distinct asthma phenotypes for use in aetiological studies, by applying LCA in two large epidemiological studies conducted in adults: the European Community and Respiratory Health Survey (ECRHS), a European population-based study; and the Epidemiological Study on the Genetics and Environment of Asthma (EGEA), a French case–control and family-based study.

METHODS
Details regarding the methods are provided in the online supplementary data.

Population
The ECRHS is a European population-based study on adults with an 8-yr follow-up (ECRHSI: 1991–1993, n=18,356; ECRHSII: 1999–2002, n=10,933) [9, 10]. The EGEA is a French case–control and family-based study with protocols and questionnaires similar to ECRHS (EGEA1: 1991–1995, n=2,047; EGEA2: 2003–2007, n=1,601) [11–13] (supplementary methods, and figs E1 and E2 in the online supplementary data).

The present cross-sectional analysis was conducted on 1,895 subjects who had ever had asthma at ECRHSII (answered ''yes'' to the question ''Have you ever had asthma?'') and on 641 adults who had ever had asthma at EGEA2 (answered ''yes'' to the question ''Have you ever had attacks of breathlessness at rest with wheezing?'' or ''Have you ever had asthma attacks?'', or being recruited as an asthma case in chest clinics).

Analysis strategy
LCA, a latent variable model that serves to cluster subjects into classes, was used to identify distinct asthma phenotypes [14]. This approach allows identification of a set of latent classes of individuals who are similar to each other according to the variables used in the analysis (see Methods in the online supplementary data). As our objective was to identify homogeneous asthma phenotypes to better assess risk factors for asthma, we decided to focus on personal characteristics (age and sex), phenotypic characteristics (asthma symptoms over the previous 12 months, age of asthma onset, asthma exacerbation, allergic characteristics, lung function and airway hyperresponsiveness (AHR)) and asthma treatment. Asthma treatment was included as it has a direct impact on the clinical features of the disease, may partly reflect the activity of asthma and has already been used in a previous study with a similar purpose [5]. To comply with the conditional independence assumption of LCA (*i.e.* the assumption that, within each latent class, all input variables are statistically independent of each other), the original list of 18 variables (table E1 in the online supplementary data) was reduced using an exploratory factor analysis, a multivariate approach that allowed identifying variables that represented similar dimensions (supplementary methods, and tables E2 and E3 in the online supplementary data). The 18 variables were thus reduced to 14 independent variables: age, sex, age of asthma onset, woken up by attack of coughing, asthma symptom score, chronic cough or phlegm, asthma attacks, and asthma exacerbation in the previous 12 months, the type of asthma treatment, eczema, rhinitis,

atopy (skin prick tests or specific immunoglobulin (Ig)E), total IgE, and forced expiratory volume in 1 s (FEV_1). AHR (provocative dose causing a 20% fall in $FEV_1 \leqslant 1$ mg methacholine) was not included in the factor analysis because it was missing for all subjects with low lung function at baseline ($FEV_1 <70\%$ predicted for ECRHSII and <80% pred for EGEA2 precluded individuals from undergoing bronchial challenges).

In order to determine the number of latent classes, models with different numbers of latent classes were compared using the Bayesian information criterion (BIC) and the model with the lowest BIC was selected. Each subject was assigned to the latent class for which they had the highest membership probability [6].

In order to validate the identified phenotypes, we assessed their discriminative properties according to health-related quality of life (HRQoL), which was assessed using the total Asthma Quality of Life Questionnaire (AQLQ) score [15]. We hypothesised that HRQoL differences observed between the phenotypes identified by LCA were of stronger magnitudes than HRQoL differences observed between phenotypes identified on a single variable included in the classification (atopy, age of asthma onset and asthma treatment) or a composite score, such as asthma control assessed following Global Initiative for Asthma (GINA) guidelines [13, 16, 17]. In order to allow for the comparison of the HRQoL differences observed across the variables, effect sizes were computed as the ratio of the mean difference between the two groups divided by the pooled standard deviation, as proposed by COHEN [18].

In the EGEA2 study, a further dimension of validity was studied by comparing two inflammatory markers, blood eosinophil and neutrophil counts, between the phenotypes identified by the LCA.

RESULTS
Description of populations
The populations studied are described in table 1. The individuals in ECRHSII were older and more often female compared with the individuals in EGEA2. The prevalence of asthma symptoms over the previous 12 months was comparable between the two studies after adjustment for age and sex, except for shortness of breath following activity (less often reported in ECRHSII than in EGEA2) and nocturnal shortness of breath (more frequently reported in ECRHSII than in EGEA2) (table 1). Because of the different study designs in the two studies, individuals in ECRHSII had less early-onset asthma, allergic characteristics and severe exacerbations.

Latent class analysis
Using the BIC criteria, a model with four latent classes was selected as the best model for the ECRHSII data. The mean highest posterior probability was high (83%), indicating that participants were assigned to classes with a fairly high probability. Phenotype A (36.1%), ''active treated allergic childhood-onset asthma'', is characterised by individuals with atopic asthma and active disease (asthma symptoms and asthma treatment) at the time of examination (table 2). Compared with the other three groups, individuals belonging

TABLE 1	Description of the individuals with asthma in the European Community and Respiratory Health Survey (ECRHS)II and Epidemiological Study on the Genetics and Environment of Asthma (EGEA)2 studies		
		ECRHS II	EGEA2	p-value[#]
Subjects n		1895	641	
Age yrs				<0.0001
<40		760 (40.1)	359 (56.0)	
⩾40		1135 (59.9)	282 (44.0)	
Sex				<0.0001
Males		780 (41.2)	337 (52.6)	
Females		1115 (58.8)	304 (47.4)	
Age at asthma onset yrs				<0.0001
⩽4		289 (15.8)	189 (31.4)	
4–16		558 (30.5)	210 (34.9)	
>16		982 (53.7)	203 (33.7)	
Wheezing with breathlessness[¶]				0.16
No		871 (46.4)	312 (48.9)	
Yes		1006 (53.6)	326 (51.1)	
Woken up with feeling of tightness[¶]				0.25
No		1072 (56.6)	344 (53.9)	
Yes		822 (43.4)	294 (46.1)	
Attack of shortness of breath at rest[¶]				0.69
No		1397 (73.8)	478 (75.2)	
Yes		496 (26.2)	158 (24.8)	
Shortness of breath during activity[¶]				0.0008
No		961 (50.8)	280 (44.0)	
Yes		931 (49.2)	357 (56.0)	
Woken up with shortness of breath[¶]				0.02
No		1409 (74.5)	506 (79.7)	
Yes		482 (25.5)	129 (20.3)	
Woken up by attack of coughing[¶]				0.07
No		1046 (55.3)	395 (62.0)	
Yes		846 (44.7)	242 (38.0)	
Chronic cough or phlegm				<0.0001
No		1463 (77.5)	542 (85.3)	
Yes		426 (22.5)	93 (14.6)	
Asthma symptom score[¶]				0.29
0 symptoms		452 (24.2)	135 (21.5)	
1–2 symptoms		750 (40.2)	276 (43.9)	
⩾3 symptoms		665 (35.6)	218 (34.6)	
Asthma attack[¶]				0.0005
No		1061 (55.5)	406 (64.0)	
Yes		818 (43.5)	228 (36.0)	
Exacerbation[¶]				0.001
No		1492 (89.8)	534 (85.2)	
Yes		170 (10.2)	93 (14.8)	
Asthma treatment[+]				<0.0001
No asthma treatment		691 (44.0)	254 (42.2)	
Other than daily ICS		611 (38.9)	203 (33.7)	
Daily ICS		268 (17.1)	145 (24.1)	
Eczema				0.0007
No		807 (42.8)	326 (51.3)	
Yes		1077 (57.2)	309 (48.7)	
Rhinitis				0.002
No		659 (34.9)	171 (27.2)	
Yes		1230 (65.1)	457 (72.8)	

TABLE 1	Continued		
	ECRHS II	EGEA2	p-value[#]
Atopy			<0.0001
No	534 (36.3)	110 (20.3)	
Yes	939 (63.7)	431 (79.7)	
IgE IU·mL^{-1}			<0.0001
<100	802 (54.4)	224 (38.6)	
⩾100	673 (45.6)	357 (61.4)	
FEV1 % pred			0.05
⩾80%	1353 (87.0)	491 (85.0)	
<80%	202 (13.0)	87 (15.0)	
AHR[§]			0.55
No	521 (51.5)	156 (52.3)	
Yes	490 (48.5)	142 (47.4)	

Data are presented as n (%), unless otherwise stated. ICS: inhaled corticosteroids; Ig: immunoglobulin; FEV1: forced expiratory volume in 1 s; % pred: % predicted; AHR: airway hyperresponsiveness. [#]: adjusted for age and sex, except for age and sex, for which unadjusted p-value are given; [¶]: in the previous 12 months; [+]: in the previous 3 months; [§]: provocative dose causing a 20% fall in FEV1 ⩽1 mg methacholine.

to this group had AHR more often. Phenotype B (19.2%), ''active treated adult-onset asthma'', was characterised by older subjects with adult-onset asthma (compared with the three other groups); they were mostly females with active disease at the time of examination, and many of them had an asthma symptom score of three or more and reported an asthma attack in the previous 12 months. Compared with the other three groups, the probability of chronic cough or phlegm was the highest in this group. Phenotypes C (28.9%) and D (15.8%) were both characterised by individuals with no or few asthma symptoms and no asthma treatment at the time of examination, these two groups differed in atopy status: phenotype C was labelled ''inactive/mild untreated allergic asthma'' and phenotype D was labelled ''inactive/mild untreated nonallergic asthma''. Compared with the other three groups, allergy-related variables (rhinitis, atopy and IgE <100 IU·mL^{-1}) and AHR were lowest in phenotype D.

Similarly, in the EGEA2 population, the best-fitting model had four latent classes (phenotypes E–H). The mean highest posterior probability was similar to the one observed in ECRHSII (88%). Phenotype E (34.6%), ''active treated allergic childhood-onset asthma'', was composed of young individuals with childhood-onset asthma and atopy, and active disease at the time of examination (table 3). Phenotype F (15.0%), ''active treated adult-onset asthma'', was characterised by older subjects with adult-onset asthma and active disease at the time of examination (92% had more than one symptom and 68% used daily asthma treatment). Compared with the other three classes, the individuals belonging to this group reported asthma exacerbation and chronic cough or phlegm more often, and more often had an FEV1 <80% pred. Phenotypes G and H were both composed of subjects with no or few asthma symptoms and asthma treatment, but differed mainly in age, age of asthma onset and allergic phenotypes. Phenotype G (24.8%) was labelled ''inactive/mild untreated allergic

| TABLE 2 | Characteristics of the European Community and Respiratory Health Survey (ECRHS)II population and probability of individuals presenting the characteristics, by inclusion in each of the four phenotypes identified by the latent class analysis |

	All subjects	Phenotype A	Phenotype B	Phenotype C	Phenotype D
Subjects n (%)	1895 (100.0)	685 (36.1)	363 (19.2)	548 (28.9)	299 (15.8)
Age ⩾40 yrs	0.60	0.50	0.76	0.55	0.73
Male sex	0.41	0.49	0.25	0.46	0.33
Age at asthma onset yrs					
⩽4	0.16	0.20	0.04	0.16	0.20
4–16	0.30	0.40	0.12	0.38	0.17
>16	0.54	0.40	0.84	0.46	0.63
Woken by coughing[#]	0.45	0.45	0.79	0.26	0.35
Asthma symptom score[#]					
0 symptom	0.24	0.03	0.02	0.52	0.50
1–2 symptoms	0.40	0.38	0.30	0.45	0.48
⩾3 symptoms	0.36	0.59	0.67	0.03	0.02
Chronic cough or phlegm	0.22	0.24	0.46	0.09	0.17
Asthma attack[#]	0.43	0.75	0.73	0.06	0.04
Exacerbation[#]	0.10	0.17	0.22	0.00	0.01
Asthma treatment[¶]					
No treatment	0.44	0.08	0.20	0.75	0.79
Other than daily ICS	0.39	0.64	0.49	0.19	0.18
Daily ICS	0.17	0.28	0.31	0.06	0.03
Eczema	0.57	0.63	0.44	0.59	0.58
Rhinitis	0.65	0.78	0.54	0.75	0.31
Atopy	0.64	0.96	0.05	0.98	0.00
IgE ⩾100 IU·mL^{-1}	0.46	0.74	0.15	0.50	0.11
FEV$_1$ <80% pred	0.13	0.19	0.18	0.05	0.08
AHR[+]	0.48	0.76	0.44	0.36	0.20

Data are presented as the probability of individuals presenting the characteristics, unless otherwise stated. ICS: inhaled corticocosteroid; Ig: immunoglobulin; FEV$_1$: forced expiratory volume in 1 s; % pred: % predicted; AHR: airway hyperresponsiveness. [#]: in the previous 12 months; [¶]: in the previous 3 months; [+]: provocative dose causing a 20% fall in FEV$_1$ ⩽1 mg methacholine.

childhood-onset asthma'' and phenotype H (25.6%) was labelled ''inactive/mild untreated adult onset asthma''. Compared to the three other phenotypes, allergy-related variables (atopy and IgE <100 IU·mL^{-1}) and AHR were the lowest in phenotype H.

Discriminative properties of the identified subgroups with regard to quality of life, and blood eosinophil and neutrophil counts

In both studies, strong associations were found between the four phenotypes and the total AQLQ score (fig. 1). In ECRHSII, the difference in HRQoL score between the two most contrasted asthma phenotypes (B and C) identified by the LCA corresponded to an effect size of 1.4. This effect size was larger than any of the differences in HRQoL score observed for the other asthma classifications (effect sizes <1.3) (fig. 1a). Similarly, in EGEA2, the strongest difference in the total AQLQ score between all asthma phenotypes was observed for two asthma phenotypes identified using the LCA (F and G; effect size 2.0). In comparison, the effect size comparing controlled and uncontrolled asthma was 1.7 (fig. 1b). In both samples, the phenotype ''active treated adult-onset asthma'' (B and F) was associated with the poorest HRQoL.

In the EGEA2 study, blood eosinophil and neutrophil counts were strongly associated with the phenotypes (p<0.0001) (table 4). Eosinophil count was highest in phenotype E (active treated allergic childhood-onset asthma) and lowest in phenotype H (inactive/mild untreated adult onset asthma). Neutrophil count was highest in phenotype F (active treated adult-onset asthma) and lowest in phenotype G (inactive/mild untreated allergic childhood-onset asthma).

DISCUSSION

Our LCA of two large epidemiological studies revealed four distinct asthma phenotypes. Two of these phenotypes were similar in both populations (''active treated allergic childhood-onset asthma'' and ''active treated adult-onset asthma'') and corresponded to phenotypes encountered in clinical practice. The other two phenotypes were composed of subjects with inactive or mild untreated asthma, which differed between each other by atopy and age of asthma onset. Interestingly, the asthma phenotypes identified by the LCA significantly discriminated levels of HRQoL more efficiently than simple clinical asthma classification. Blood eosinophil and neutrophil counts were significantly associated with these phenotypes.

TABLE 3 Characteristics of the Epidemiological Study on the Genetics and Environment of Asthma (EGEA)2 population and probability of individuals presenting the characteristics, by inclusion in each of the four phenotypes identified by the latent class analysis

	All subjects	Phenotype E	Phenotype F	Phenotype G	Phenotype H
Subjects n (%)	641 (100.0)	222 (34.6)	96 (15.0)	159 (24.8)	164 (25.6)
Age $\geqslant$40 yrs	0.44	0.24	0.96	0.07	0.76
Male sex	0.53	0.47	0.54	0.68	0.44
Age at asthma onset yrs					
$\leqslant$4	0.31	0.50	0.00	0.43	0.13
4–16	0.35	0.43	0.05	0.50	0.27
>16	0.34	0.07	0.95	0.07	0.60
Woken by coughing[#]	0.38	0.50	0.53	0.18	0.33
Asthma symptom score[#]					
0 symptom	0.21	0.00	0.08	0.48	0.33
1–2 symptoms	0.44	0.31	0.42	0.52	0.55
$\geqslant$3 symptoms	0.35	0.69	0.50	0.00	0.12
Chronic cough or phlegm	0.15	0.16	0.31	0.04	0.14
Asthma attack[#]	0.36	0.68	0.61	0.08	0.05
Exacerbation[#]	0.15	0.26	0.37	0.00	0.00
Asthma treatment[¶]					
No treatment	0.42	0.05	0.00	0.86	0.75
Other than daily ICS	0.34	0.68	0.32	0.09	0.12
Daily ICS	0.24	0.27	0.68	0.05	0.13
Eczema	0.49	0.56	0.30	0.58	0.40
Rhinitis	0.73	0.87	0.56	0.72	0.64
Atopy	0.80	0.98	0.59	1.00	0.47
IgE $\geqslant$100 IU·mL^{-1}	0.61	0.79	0.54	0.75	0.29
FEV$_1$ <80% pred	0.15	0.15	0.51	0.01	0.07
AHR[+]	0.48	0.63	0.42	0.48	0.24

Data are presented as the probability of individuals presenting the characteristics, unless otherwise stated. ICS: inhaled cortiocosteroid; Ig: immunoglobulin; FEV$_1$: forced expiratory volume in 1 s; % pred: % predicted; AHR: airway hyperresponsiveness. [#]: in the previous 12 months; [¶]: in the previous 3 months; [+]: provocative dose causing a 20% fall in FEV$_1$ $\leqslant$1 mg methacholine.

One strength of the present study lies on the use of two well characterised and large populations of individuals with asthma. To our knowledge, this is the first time that a clustering approach aimed at identifying asthma phenotypes has been applied in adults with asthma (ever) recruited in a population-based study, allowing us to cover a large range of asthma phenotypes. The purpose of trying to single out homogeneous asthma phenotypes in epidemiological settings is to increase the power to identify risk factors associated with asthma and, therefore, to better understand diseases mechanisms. In this context, including subjects who had ever had asthma and not only current asthma, as in clinical settings, may bring complementary insights to the understanding of persistent *versus* remittent asthma. The analysis, conducted in the two asthma populations independently, allowed assessment of to what degree the phenotypes obtained differed between these two epidemiological studies relying on standardised protocols but different designs (a European community-based study, and a French case–control and family-based study). It is remarkable that similar results were observed in both populations. The lack of availability of biomarkers in ECRHSII did not allow to include markers of inflammation in the cluster analysis. Inflammation markers have previously been identified as major phenotyping criteria [4, 19].

The LCA, a model-based clustering approach, has been chosen because it is well designed to treat categorical variables included in the analysis; it handles missing data and, therefore, allows consideration of the whole sample in the analysis. The application of this method in children with asthma led to the identification of different wheezing phenotypes [3, 6]. Although very interesting, the findings provided by exploratory analyses have to be interpreted in the context of future work in order to address whether the identified phenotypes are relevant from clinical and aetiological perspectives.

Replication of the results in other datasets is important when using these exploratory approaches; however, such replication is difficult, as the phenotypes identified are dependent on the populations under study (clinical or population based) and on the set of selected variables. Nevertheless, it is noteworthy that phenotypes identified in the present study show overlap with clusters described by HALDAR *et al.* [4] and MOORE *et al.* [5], which relied on different study designs and different *a priori* lists of selected variables. All three studies identified a phenotype composed of subjects with early-onset atopic asthma. As previously identified in a primary care dataset [4], we also identified groups of benign (mild) asthma; mild
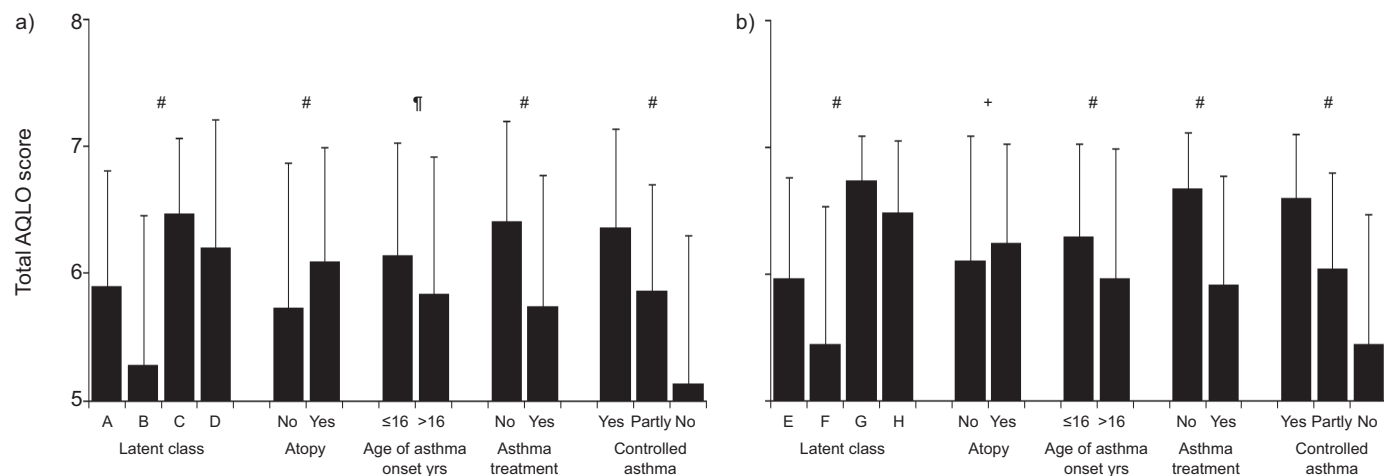
FIGURE 1. Differences on the total Asthma Quality of Life Questionnaire (AQLQ) score between the asthma phenotypes identified by latent class analysis and other *a priori* asthma subgroups (atopy, age of asthma onset, asthma treatment, asthma control defined following the Global Initiative for Asthma 2006 guidelines [17]). Results observed in the a) European Community and Respiratory Health Survey II and b) Epidemiological Study on the Genetics and Environment of Asthma 2 are presented. Bars represent means and whiskers represent standard deviations. #: $p < 0.0001$; ¶: $p = 0.0002$; +: $p = 0.13$.

asthma was split in two groups according to atopy in ECRHSII and age at onset in EGEA2. Interestingly, compared with the other three phenotypes, phenotypes B in ECRHSII (which consisted mainly of females with late-onset disease and no atopy) and F in EGEA2 (which consisted mainly of subjects with late-onset disease no atopy and with airflow limitation) showed similar characteristics to phenotype 5 in the study by MOORE *et al.* [5]. Moreover, neutrophils were highest in phenotype 5 in the study by MOORE *et al.* [5] and phenotype F in EGEA2 in the present study.

Our findings suggest that treatment is an important feature toconsider when identifying subgroups of subjects in asthma populations in developed countries. This observation did not seem dependent on geographical differences in clinical practices, given the international and multicentre nature of ECRHS. Although factors related to healthcare utilisation and social criteria are associated with the use of asthma treatment, the latter is highly associated with the activity and severity of the disease. The approach that consists of combining clinical features with the level of asthma treatment to distinguish subgroups of subjects with a differential severity, as suggested by the GINA 2002 guidelines, has been suggested by epidemiological results in populations [20]. Furthermore, a genome-wide linkage analysis on asthma quantitative score conducted in the

EGEA study showed that scoring asthma severity based on clinical items and asthma treatment increased power to detect linkage, compared with clinical items only [21].

One of the earliest approaches to identifying asthma phenotypes was the differentiation between allergic and nonallergic asthma [22–24]. Allergy-related variables played a critical role in the classification in both studies, but to a greater extent in ECRHSII, where two allergic phenotypes (A and C) and two nonallergic phenotypes (B and D) were clearly identified. The less critical role of allergy in the classification in EGEA2 may be explained by the higher prevalence of atopy in EGEA2 compared with ECRHSII, probably resulting from different study designs, with inclusion of cases from chest clinics and children, a population more prone to allergic asthma in EGEA2. Early-onset and adult-onset asthma are well-established asthma phenotypes [25]. A recent study conducted in ECRHS also provided epidemiological evidence for distinguishing adult-onset from early-onset asthma [26]. MOORE *et al.* [5] also showed, in their population of adults with more severe asthma, the importance of the age of onset in phenotyping asthma. Asthma phenotypes show age-related variations [27] and, accordingly, the age at examination was identified in the EGEA2 study as a major phenotyping criterion. MOORE *et al.* [5] and HALDAR *et al.* [4] also identified groups of subjects composed of older subjects.

TABLE 4 Blood eosinophil and neutrophil counts in the Epidemiological Study on Genetics and Environment of Asthma (EGEA)2 population according to the phenotypes identified by the latent class analysis

	Phenotype E	Phenotype F	Phenotype G	Phenotype H	p-value
Subjects n	222	96	159	164	
Neutrophils cells per mL	3866 (1897–7878)	4358 (2177–8724)	3524 (2018–6152)	3775 (2056–6933)	<0.0001
Eosinophils cells per mL	243 (72–817)	212 (47–956)	190 (57.0–636)	174 (50.2–600)	<0.0001

Data are presented as geometric means (95% CI), unless otherwise stated.

The distinct asthma phenotypes defined by the LCA exhibited strong differences in HRQoL, even stronger than when using other existing asthma classifications. Also, it is reassuring that our results using an exploratory method are consistent with observations from clinical practice, with phenotypes B and F (both ''active treated adult-onset asthma'') being associated with the poorest HRQoL and the lowest FEV1.These results show the strong discriminative properties of our classification with regard to HRQoL, the patient's own perception of their health status. It is also noteworthy that the phenotypes showed significant differences in eosinophil and neutrophil counts, two objective measurements of the inflammatory component of the disease [19].

Despite all the research efforts in asthma genetics over the last decade, the genetic basis of asthma remains largely unknown [28]. Recent genome-wide association studies have confirmed the genetic heterogeneity of asthma according to age at onset [29]. To better benefit from the existing genomic data, there is a need to reduce phenotypic heterogeneity by the improvement of the phenotype definition [2]. Genetic studies relying on more homogeneous phenotypes, such as those defined by a multivariate approach like ours, appear to be a promising approach to this problem.

In summary, the current analyses provide further evidence for asthma heterogeneity in adults in the general population and support the use of multivariate statistical techniques that allow a more integrated classification of asthma. Considering these more homogeneous phenotypes in future studies could lead to identification of novel risk factors, genetic as well as environmental, and to improve the understanding of the disease.

STATEMENT OF INTEREST
A statement of interest for the study itself can be found at www.erj. ersjournals.com/site/misc/statements.xhtml

REFERENCES
1 Wenzel SE. Asthma: defining of the persistent adult phenotypes. *Lancet* 2006; 368: 804–813.
2 Manolio TA, Collins FS, Cox NJ, *et al.* Finding the missing heritability of complex diseases. *Nature* 2009; 461: 747–753.
3 Henderson J, Granell R, Heron J, *et al.* Associations of wheezing phenotypes in the first 6 years of life with atopy, lung function and airway responsiveness in mid-childhood. *Thorax* 2008; 63: 974–980.
4 Haldar P, Pavord ID, Shaw DE, *et al.* Cluster analysis and clinical asthma phenotypes. *Am J Respir Crit Care Med* 2008; 178: 218–224.
5 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.

6 Spycher BD, Silverman M, Brooke AM, *et al.* Distinguishing phenotypes of childhood wheeze and cough using latent class analysis. *Eur Respir J* 2008; 31: 974–981.

7 Pillai SG, Tang Y, van den OE, *et al.* Factor analysis in the Genetics of Asthma International Network family study identifies five major quantitative asthma phenotypes. *Clin Exp Allergy* 2008; 38: 421–429.

8 Fahy JV. Identifying clinical phenotypes of asthma: steps in the right direction. *Am J Respir Crit Care Med* 2010; 181: 296–297.

9 Burney P. The changing prevalence of asthma? *Thorax* 2002; 57: Suppl. 2, II36–II39.

10 The European Community Respiratory Health Survey II Steering Committee. The European Community Respiratory Health Survey II. *Eur Respir J* 2002; 20: 1071–1079.

11 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–S129.

12 Kauffmann F, Dizier MH. EGEA (Epidemiological study on the Genetics and Environment of Asthma, bronchial hyperresponsiveness and atopy) – design issues. EGEA Co-operative Group. *Clin Exp Allergy* 1995; 25: Suppl. 2, 19–22.

13 Siroux V, Boudier A, Bousquet J, *et al.* Phenotypic determinants of uncontrolled asthma. *J Allergy Clin Immunol* 2009; 124: 681–687.

14 Hagenaars J, McCutcheon A. Applied Latent Class Analysis. Cambridge, Cambridge University Press, 2002.

15 Juniper EF, Guyatt GH, Epstein RS, *et al.* Evaluation of impairment of health related quality of life in asthma: development of a questionnaire for use in clinical trials. *Thorax* 1992; 47: 76–83.

16 Global Initiative for Asthma. Global Strategy for Asthma Management and Prevention. NIH publication no. 95-3659. Bethesda, National Institutes of Health, 1995.

17 Cazzoletti L, Marcon A, Janson C, *et al.* Asthma control in Europe: a real-world evaluation based on an international population-based study. *J Allergy Clin Immunol* 2007; 120: 1360–1367.

18 Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd Edn. Hillsdale, Laurence Earlbaum Associates, 1988.

19 Nadif R, Siroux V, Oryszczyn MP, *et al.* Heterogeneity of asthma according to blood inflammatory patterns. *Thorax* 2009; 64: 374–380.

20 Liard R, Leynaert B, Zureik M, *et al.* Using Global Initiative for Asthma guidelines to assess asthma severity in populations. *Eur Respir J* 2000; 16: 615–620.

21 Bouzigon E, Siroux V, Dizier MH, *et al.* Scores of asthma and asthma severity reveal new regions of linkage in EGEA study families. *Eur Respir J* 2007; 30: 253–259.

22 Rackeman FM. A working classification of asthma. *Am J Med* 1947; 33: 601–606.

23 Stein RT, Holberg CJ, Morgan WJ, *et al.* Peak flow variability, methacholine responsiveness and atopy as markers for detecting different wheezing phenotypes in childhood. *Thorax* 1997; 52: 946–952.

24 Romanet-Manent S, Charpin D, Magnan A, *et al.* Allergic *vs* non-allergic asthma: what makes the difference? *Allergy* 2002; 57: 607–613.

25 Miranda C, Busacker A, Balzar S, *et al.* Distinguishing severe asthma phenotypes: role of age at onset and eosinophilic inflammation. *J Allergy Clin Immunol* 2004; 113: 101–108.

26 Anto JM, Sunyer J, Basagana X, *et al.* Risk factors of new-onset asthma in adults: a population-based international cohort study. *Allergy* 2010; 65: 1021–1030.

27 Bush A, Menzies-Gow A. Phenotypic differences between pediatric and adult asthma. *Proc Am Thorac Soc* 2009; 6: 712–719.

28 Postma DS, Koppelman GH. Genetics of asthma: where are we and where do we go? *Proc Am Thorac Soc* 2009; 6: 283–287.

29 Moffatt MF, Gut IG, Demenais F, *et al.* A large-scale, consortium-based genomewide association study of asthma. *N Engl J Med* 2010; 363: 1211–1221.