ERJ Express. Published on January 13, 2011 as doi: 10.1183/09031936.00120810

Identifying adult asthma phenotypes using a clustering approach

Valérie Siroux[1,2], Xavier Basagaña[3,4,5,6], Anne Boudier[1,2], Isabelle Pin[1,2,7], Judith Garcia-Aymerich[3,4,5,6], Aurélien Vesin[1,2], Rémy Slama[1,2], Deborah Jarvis[8], Josep M Anto[3,4,5,6], Francine Kauffmann[9,10], Jordi Sunyer[3,4,5,6]

1. Team of Environmental Epidemiology applied to Reproduction and Respiratory Health, Inserm, U823, Grenoble, France
2. Université Joseph Fourier, Grenoble, France
3. Centre for Research in Environmental Epidemiology (CREAL), Barcelona, Spain
4. Institut Municipal d'Investigació Mèdica (IMIM-Hospital del Mar), Barcelona, Spain
5. Department of Experimental Sciences and Health, Universitat Pompeu Fabra, Barcelona, Spain
6. CIBER en Epidemiologia y Salud Pública (CIBERESP), Barcelona, Spain
7. CHU Grenoble, pédiatrie, Grenoble, France
8. Respiratory Epidemiology and Public Health Group, Imperial College, London, United-Kingdom
9. Respiratory and Environmental epidemiology, CESP Centre for research in Epidemiology and Population health, U1018, Inserm, F-94807, Villejuif, France
10. Université Paris Sud 11, UMRS 1018, F-94807,Villejuif, France

Correspondance and reprint requests:

Valérie SIROUX

Centre de Recherche INSERM/UJF U823

Institut Albert Bonniot

BP 170

38042 Grenoble Cedex 9

Tel : 33-(0)4-76-54-95-56

Fax : 33-(0)4-76-54-94-13

Mail : valerie.siroux@ujf-grenoble.fr

Funding support : The EGEA study was supported in part by grants from Merck Sharp & Dohme (MSD); Hospital program of clinical research (PHRC)-Paris; National Research Agency - Health environment, health-work program; National Research Agency (ANR)-Biological collections for health program; French Agency of health safety, environment and work (AFSSET); the national scientific committee of the medico-technology support at home (AGIR à dom) and the Isere committee against respiratory diseases (COMARES).

Judith Garcia-Aymerich has a researcher contract from the Instituto de Salud Carlos III (CP05/00118), Ministry of Health, Spain.

Word count: 2920

ABSTRACT

There is a need to improve asthma characterization by integrating multiple aspects of the disease.

The aim 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 Genetics and Environment of Asthma (EGEA2, n=641) and the European Community Respiratory Health Survey (ECRHSII, n=1895) 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/mild untreated asthma, that differed by atopy status and age of asthma onset (childhood/adulthood). The phenotypes clearly discriminated populations in terms of quality-of-life and blood eosinophil and neutrophil counts.

The latent class analyses 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.

Key words: asthma heterogeneity; asthma phenotypes; latent class analysis

INTRODUCTION

Asthma is a complex disorder that includes distinct phenotypes with potentially different etiologies, natural histories and responses to treatment [1]. Distinct adult asthma phenotypes have been identified for some time but have been based on limited number of characteristics. Allergic and non allergic asthma are probably the most commonly discussed phenotypes. Other phenotypes defined by clinical or physiological categories (i.e severity, age at onset, chronic airflow obstruction), by asthma triggers (i.e. exercise, allergens, occupational allergens or irritants) or by their pathobiology (i.e. eosinophilic, neutrophilic asthma) have also been proposed [1]. It is expected that a comprehensive examination protocol of asthma patients that incorporates several domains of the disease would make possible to identify more distinct asthma phenotypes. Such widening of the asthma characterization may allow a better understanding of the etiology 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), like clustering methods, have already been applied in the respiratory epidemiology field [3-7] and have recently been underlined as being 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 in clinical, physiologic and inflammatory parameters [5]. The 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 which, compared to clinical population, are 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 by applying LCA in two large epidemiological studies conducted in adults, ECRHS (the European Community and Respiratory Health Survey) a European population-based study and EGEA (the Epidemiological Study on the Genetics of Asthma, bronchial hyperresponsiveness) a French case-control and family-based study, for being applied in etiological studies of asthma.

METHODS

Details regarding the methods are provided in the online data supplement.

POPULATION

The ECRHS study is an European population-based study on adults with a 8-year follow-up (ECRHSI (1991-1993), n=18356; ECRHSII (1999-2002), n=10933) [9,10]. The EGEA is a French case-control and family-based study with protocols and questionnaires similar to ECRHS (EGEA1 (1991-1995) n=2047; EGEA2 (2003-2007), n= 1601) [11-13] (see methods, Figure E1 and Figure E2 in the online data supplement).

The present cross-sectional analysis was conducted on 1895 subjects with asthma ever at ECRHSII (positive answer to "Have you ever had asthma?") and on 641 adults with asthma ever at EGEA2 (positive answer to "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

Latent Class Analysis (LCA), a latent variable model that serves to cluster subjects into classes was used to identify distinct asthma phenotypes [14]. This approach allows identifying 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 data supplement). 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), on phenotypic characteristics (asthma symptoms over the past 12 months, age of asthma onset, asthma exacerbation, allergic characteristics, lung function and airway hyperresponsiveness) and on asthma treatment 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 (see Table E1 in the online data supplement) was reduced using an exploratory factor analysis, a multivariate approach which allowed identifying variables that represented similar dimensions(see methods, Table E2 and E3 in the online data supplement). 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 past 12 months, the type of asthma treatment, eczema, rhinitis, atopy (skin prick tests or specific IgE), total IgE and FEV1. Airway hyperresponsiveness (AHR) (PD20≤1mg methacholine) was not included in the factor analysis because it was missing for all subjects with low lung function at baseline (FEV1 < 70% for ECRHSII and <80% for EGEA2 precluded individuals from undergoing bronchial challenges).

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 he had the highest membership probability [6].

To validate the identified phenotypes, we assessed their discriminative properties according to Health-Related Quality of Life (HRQL) assessed with the total Asthma Quality of Life Questionnaire (AQLQ) score [15]. We hypothesized that HRQL differences observed between the phenotypes identified by LCA were of stronger magnitudes than HRQL differences observed between phenotypes identified on single variable included in the classification (atopy, age of asthma onset and asthma treatment) or on a composite score such as asthma control assessed following GINA guidelines [13,16,17]. To allow for the comparison of the HRQL differences observed across the variables, effect-sizes were computed by the ratio of the mean difference between the 2 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 women compared with the individuals in EGEA2. The prevalence of asthma symptoms over the past 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 often 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 in the ECRHSII data. The mean highest posterior probability was high (83%), indicating that participants were assigned to classes with a fairly high probability. The phenotype A (36.1%), labelled "Active treated allergic childhood-onset asthma" is characterized by individuals with atopic asthma with an active disease (asthma symptoms and asthma treatment) at the time of examination (Table 2). Compared to the other 3 groups, individuals belonging to this group had more often AHR. Phenotype B (19.2%), labelled « Active treated adult-onset asthma » was characterized by older subjects with adult-onset asthma (compared to the 3 other groups), they were mostly women, with an active disease at the time of examination; many of them had an asthma symptom score of 3 or more and reported an asthma attack in the last 12 months. Compared to the other 3 groups, the probability of chronic cough or phlegm was the highest in this group. Phenotype C (28.9%) and D (15.9%) were both characterized 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 non-allergic asthma". Compared to the 3 other groups, allergic-related variables (rhinitis, atopy, IgE<100IU/ml) and AHR were the lowest in phenotype D.

Similarly, in the EGEA2 population, the best-fitting model had four latent classes (labelled E to H). The mean highest posterior probability was similar to the one observed in ECRHS (88 %). Phenotype E (34.6%), labelled "Active treated allergic childhood-onset asthma" was composed of young individuals with childhood-onset asthma and atopy and an active disease at the time of examination (table 3). Phenotype F (15.0%), labelled "Active treated adult-onset asthma" was characterized by older subjects with adult-onset asthma and an active disease at the time of examination (92% had > 1 symptoms and 68% used daily asthma treatment). Compared to the three other classes, the individuals belonging to this group more often reported asthma exacerbation and chronic cough or phlegm and had more often a FEV1 < 80% predicted. 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%) regarded "Inactive/mild untreated allergic childhood-onset asthma", and phenotype H (25.6%) regarded "Inactive/mild untreated adult onset asthma". Compared to the three other phenotypes, allergic-related variables (atopy, IgE<100IU/ml) and AHR were the lowest in phenotype H.

Discriminative properties of the identified subgroups with regard of 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 (figure 3). In ECRHSII, the difference of HRQL 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 of HRQL score observed for the other asthma classifications (effect sizes < 1.3) (Figure 1A). In EGEA2, similarly, the strongest difference in the total AQLQ score between all asthma phenotypes was observed for two asthma phenotypes identified using the LCA (phenotypes F and G, effect size 2.0). In comparison, the effect size comparing controlled and uncontrolled asthma was 1.7 (Figure 1 B). In both samples, the phenotype "Active treated adult-onset asthma" (B and F) was associated with the poorest HRQL.

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. 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 latent class analysis of two large epidemiological studies allowed to reveal 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/mil 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 HRQL more efficiently than simple clinical asthma classification. Blood eosinophil and neutrophil counts were significantly associated with these phenotypes.

One strength of the present study lies on the use of two well characterized 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 studies, allowing 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 subject with asthma ever, and not only current asthma as in clinical settings, may bring complementary insight in the understanding of persistent *versus* remittent asthma. The analysis, conducted in the two asthma populations independently, allowed assessing to which degree the phenotypes obtained differed between these two epidemiological studies relying on standardized protocols but different designs (a European community-based study and a French case-control ad 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 considering 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 to address whether the identified phenotypes are relevant from clinical and etiological 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 article show overlap with clusters described by Haldar et al. and Moore et al, which relied on different study designs and different a priori list of selected variables [4,5]. 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 asthma was split in two groups according to atopy in ECRHSII and age at onset in EGEA2. Interestingly, compared to the three other phenotypes, Phenotypes B in ECRHSII (which consisted mainly of more women with late onset disease and no atopy) and F in EGEA2 (which consisted of subject mainly with late onset disease no atopy and with airflow limitation) showed similar characteristics to phenotype 5 in the paper from Moore et al. Moreover, neutrophils were highest in phenotype 5 in the study by Moore et al [5], and phenotype F in EGEA2.

Our findings suggest that treatment is an important feature to consider when identifying sub-group 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 heath care utilization and social criteria are associated with the use of asthma treatment, this later 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 sub-groups of subjects with a differential severity, as suggested by the GINA 2002 guidelines, has firstly been suggested by epidemiological results in populations [20]. Furthermore, a genome-wide linkage analysis on asthma quantitative sore conducted in the EGEA study showed that scoring asthma severity based on clinical items and asthma treatment increased power to detect linkage as compared to clinical items only [21].

One of the earliest approach to identify asthma phenotypes was the differentiation between allergic and non-allergic asthma [22-24]. Allergic-related variables played a critical role in the classification in both studies, but in a greater extent in ECRHSII where two allergic phenotypes (A and C) and non allergic 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 to 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. also showed in their population of adults with more severe asthma the importance of the age of onset in phenotyping asthma [5]. 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. and Haldar et al. also identified groups of subjects composed of older subjects [4,5].

The distinct asthma phenotypes defined by the LCA exhibited strong differences in HRQL, even stronger than when using other existing asthma classifications. Also, it is reassuring that our results using an exploratory method are in line with observations from clinical practice, with phenotype B and F, labelled “active treated adult-onset asthma”, being associated with the poorest HRQL and the lowest FEV1.These results show the strong discriminative properties of our classification with regard to HRQL, the self-perception of the patient regarding its 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 for more than a decade, the genetic basis of asthma remains largely unknown [28]. Recent GWAS 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, as those defined by a multivariate approach like ours, appear as a promising approach in the direction.

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 allow to identify novel risk factors, genetic as well as environmental, and to improve the understanding of the disease.

ACKNOWLEDGMENTS

We thank the EGEA cooperative group:

Coordination: F Kauffmann; F Demenais (genetics); I Pin (clinical aspects).

Respiratory epidemiology: Inserm U 700, Paris M Korobaeff (Egea1), F Neukirch (Egea1); Inserm U 707, Paris : I Annesi-Maesano ; Inserm CESP/U 1018, Villejuif : F Kauffmann, N Le Moual, R Nadif, MP Oryszczyn ; Inserm U 823, Grenoble : V Siroux

Genetics : Inserm U 393, Paris : J Feingold ; Inserm U 946, Paris : E Bouzigon , F Demenais, MH Dizier ; CNG, Evry : I Gut , M Lathrop.

Clinical centers : Grenoble : I Pin, C Pison; Lyon : D Ecochard (Egea1), F Gormand, Y Pacheco ; Marseille : D Charpin (Egea1), D Vervloet ; Montpellier : J Bousquet ; Paris Cochin : A Lockhart (Egea1), R Matran (now in Lille) ; Paris Necker : E Paty, P Scheinmann ; Paris-Trousseau : A Grimfeld, J Just.

Data and quality management : Inserm ex-U155 (Egea1) : J Hochez ; Inserm CESP/U 1018, Villejuif : N Le Moual, Inserm ex-U780 : C Ravault ; Inserm ex-U794 : N Chateigner ; Grenoble : J Ferran

The authors thank all those who participated to the setting of the EGEA 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 subjects were examined.

We thank the Principal Investigators and Senior Scientific Team in ECRHS

Belgium: South Antwerp & Antwerp City (P Vermeire, J Weyler, M Van Sprundel, V Nelen). **Estonia:** Tartu (R Jogi, A Soon). **France:** Paris (F Neukirch, B Leynaert, R Liard, M Zureik), Grenoble (I Pin, J Ferran-Quentin). **Germany:** Erfurt (J Heinrich, M Wjst, C Frye, I Meyer). **Iceland:** Reykjavik (T Gislason, E Bjornsson, D Gislason, T Blondal, KB Jorundsdottir). **Italy:** Turin (M Bugiani, P Piccioni, E Caria, A Carosso, E Migliore, G Castiglioni), Verona (R de Marco, G Verlato, E Zanolin, S Accordini, A Poli, V Lo Cascio, M Ferrari), Pavia (A Marinoni, S Villani, M Ponzio, F Frigerio, M Comelli, M Grassi, I Cerveri, A Corsico). **Netherlands:** Groningen & Geleen (J Schouten, M Kerkhof). **Norway:** Bergen (A Gulsvik, E Omenaas, C Svanes, B Laerum). **Spain:** Barcelona (JM Antó, J Sunyer, M Kogevinas, JP Zock, X Basagana, A Jaen, F Burgos), Huelva (J Maldonado, A Pereira, JL Sanchez), Albacete (J Martinez-Moratalla Rovira, E Almar), Galdakao (N Muniozguren, I Urritia), Oviedo (F Payo). **Sweden:** Uppsala (C Janson, G Boman, D Norback, M Gunnbjornsdottir), Goteborg (K Toren, L Lillienberg, AC Olin, B Balder, A Pfeifer- Nilsson, R Sundberg), Umea (E Norrman, M Soderberg, K Franklin, B Lundback, B Forsberg, L Nystrom). **Switzerland:** Basel (N Künzli, N. Probst-Hensch). **United Kingdom:** Norwich (D Jarvis, B Harrison), Ipswich (D Jarvis, R Hall, D Seaton) Centres taking part at their own expense**: France:** Bordeaux (A Taytard, C Raherison), Montpellier (J Bousquet, P Demoly). **Germany:** Hamburg (K Richter). **USA:** Portland (M Osborne, S Buist, W Vollmer, L Johnson).

The authors are indebted to all the individuals who participated to the ECRHS and EGEA studies without whom that study would not have been possible.

REFERENCES

1. Wenzel SE. Asthma: defining of the persistent adult phenotypes. *Lancet* 2006; 368: 804-13.

2. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM. Finding the missing heritability of complex diseases. *Nature* 2009; 461: 747-53.

3. Henderson J, Granell R, Heron J, Sherriff A, Simpson A, Woodcock A, Strachan DP, Shaheen SO, Sterne JA. Associations of wheezing phenotypes in the first 6 years of life with atopy, lung function and airway responsiveness in mid-childhood. *Thorax* 2008; 63: 974-80.

4. Haldar P, Pavord ID, Shaw DE, Berry MA, Thomas M, Brightling CE, Wardlaw AJ, Green RH. Cluster analysis and clinical asthma phenotypes. *Am J Respir Crit Care Med* 2008; 178: 218-24.

5. Moore WC, Meyers DA, Wenzel SE, Teague WG, Li H, Li X, D'Agostino JR, Castro M, Curran-Everett D, Fitzpatrick AM, Gaston B, Jarjour NN, Sorkness R, Calhoun WJ, Chung KF, Comhair SA, Dweik RA, Israel E, Peters SP, Busse WW, Erzurum SC, Bleecker ER. Identification of Asthma Phenotypes using Cluster Analysis in the Severe Asthma Research Program. *Am J Respir Crit Care Med* 2010; 181: 315-23.

6. Spycher BD, Silverman M, Brooke AM, Minder CE, Kuehni CE. Distinguishing phenotypes of childhood wheeze and cough using latent class analysis. *Eur Respir J* 2008; 31: 974-81.

7. Pillai SG, Tang Y, van den OE, Klotsman M, Barnes K, Carlsen K, Gerritsen J, Lenney W, Silverman M, Sly P, Sundy J, Tsanakas J, von BA, Whyte M, Ortega HG, Anderson WH, Helms PJ. Factor analysis in the Genetics of Asthma International Network family study identifies five major quantitative asthma phenotypes. *Clin Exp Allergy* 2008; 38: 421-9.

8. Fahy JV. Identifying clinical phenotypes of asthma: steps in the right direction. *Am J Respir Crit Care Med* 2010; 181: 296-7.

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

10. The European Community Respiratory Health Survey II. *Eur Respir J* 2002; 20: 1071-9.

11. Kauffmann F, Dizier MH, Pin I, Paty E, Gormand F, Vervloet D, Bousquet J, Neukirch F, Annesi I, Oryszczyn MP, Lathrop M, Demenais F, Lockhart A, Feingold J. 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, Bresson JL, Cracowski JL, Ferran J, Gormand F, Just J, Le Moual N, Morange S, Nadif R, Oryszczyn MP, Pison C, Scheinmann P, Varraso R, Vervloet D, Pin I, Kauffmann F. Phenotypic determinants of uncontrolled asthma. *J Allergy Clin Immunol* 2009; 124: 681-7.

14. Hagenaars J, McCutcheon A. Applied latent class analysis.Cambridge, UK: Cambridge University Press; 2002.

15. Juniper EF, Guyatt GH, Epstein RS, Ferrie PJ, Jaeschke R, Hiller TK. 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. Bethesda (MD): National Heart, Lung, and Blood Institute, National Institutes of Health; 1995(updated 2008). NIH publication no. 95-3659. URL:http://www.ginasthma.org.

17. Cazzoletti L, Marcon A, Janson C, Corsico A, Jarvis D, Pin I, Accordini S, Almar E, Bugiani M, Carolei A, Cerveri I, Duran-Tauleria E, Gislason D, Gulsvik A, Jogi R, Marinoni A, Martinez-Moratalla J, Vermeire P, de Marco R. Asthma control in Europe: A real-world evaluation based on an international population-based study. *J Allergy Clin Immunol* 2007; 120: 1360-7.

18. Cohen J. In: Lawrence Earlbaum Associates, editor. Statistical power analysis for the behavioral sciences (2nd ed.).Hillsdale, NJ: 1988.

19. Nadif R, Siroux V, Oryszczyn MP, Ravault C, Pison C, Pin I, Kauffmann F. Heterogeneity of asthma according to blood inflammatory patterns. *Thorax* 2009; 64: 374-80.

20. Liard R, Leynaert B, Zureik M, Beguin FX, Neukirch F. Using Global Initiative for Asthma guidelines to assess asthma severity in populations. *Eur Respir J* 2000; 16: 615-20.

21. Bouzigon E, Siroux V, Dizier MH, Lemainque A, Pison C, Lathrop M, Kauffmann F, Demenais F, Pin I. Scores of asthma and asthma severity reveal new regions of linkage in EGEA study families. *Eur Respir J* 2007; 30: 253-9.

22. Rackeman FM. A working classification of asthma. *Am J Med* 1947; **33**: 601-6.

23. Stein RT, Holberg CJ, Morgan WJ, Wright AL, Lombardi E, Taussig L, Martinez FD. Peak flow variability, methacholine responsiveness and atopy as markers for detecting different wheezing phenotypes in childhood. *Thorax* 1997; 52: 946-52.

24. Romanet-Manent S, Charpin D, Magnan A, Lanteaume A, Vervloet D, and the EGEA study group. Allergic vs nonallergic asthma: what makes the difference? *Allergy* 2002; 57: 607-13.

25. Miranda C, Busacker A, Balzar S, Trudeau J, Wenzel SE. Distinguishing severe asthma phenotypes: role of age at onset and eosinophilic inflammation. *J Allergy Clin Immunol* 2004; 113: 101-8.

26. Anto JM, Sunyer J, Basagana X, Garcia-Esteban R, Cerveri I, de Marco R., Heinrich J, Janson C, Jarvis D, Kogevinas M, Kuenzli K, Leynaert B, Svanes C, Wjst M, Gislason T, Burney P. Risk factors of new-onset asthma in adults: a population-based international cohort study. *Allergy* 2010; 65: 1021-30.

27. Bush A, Menzies-Gow A. Phenotypic differences between pediatric and adult asthma. *Proc Am Thorac Soc* 2009; 6: 712-9.

28. Postma DS, Koppelman GH. Genetics of asthma: where are we and where do we go? *Proc Am Thorac Soc* 2009; 6: 283-7.

29. Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, Heath S, von Mutius E, Farrall M, Lathrop M, Cookson WO. A large-scale, consortium-based genomewide association study of asthma. *N Engl J Med* 2010; 363: 1211-21.

FIGURE LEGENDS

FIGURE 1: Differences on the total AQLQ between the asthma phenotypes identified by latent class analysis and other a priori asthma sub-groups (atopy, age of asthma onset, asthma treatment, asthma control defined following the GINA 2006 guidelines). Results observed in ECRHSII (A) and EGEA2 (B) are presented.

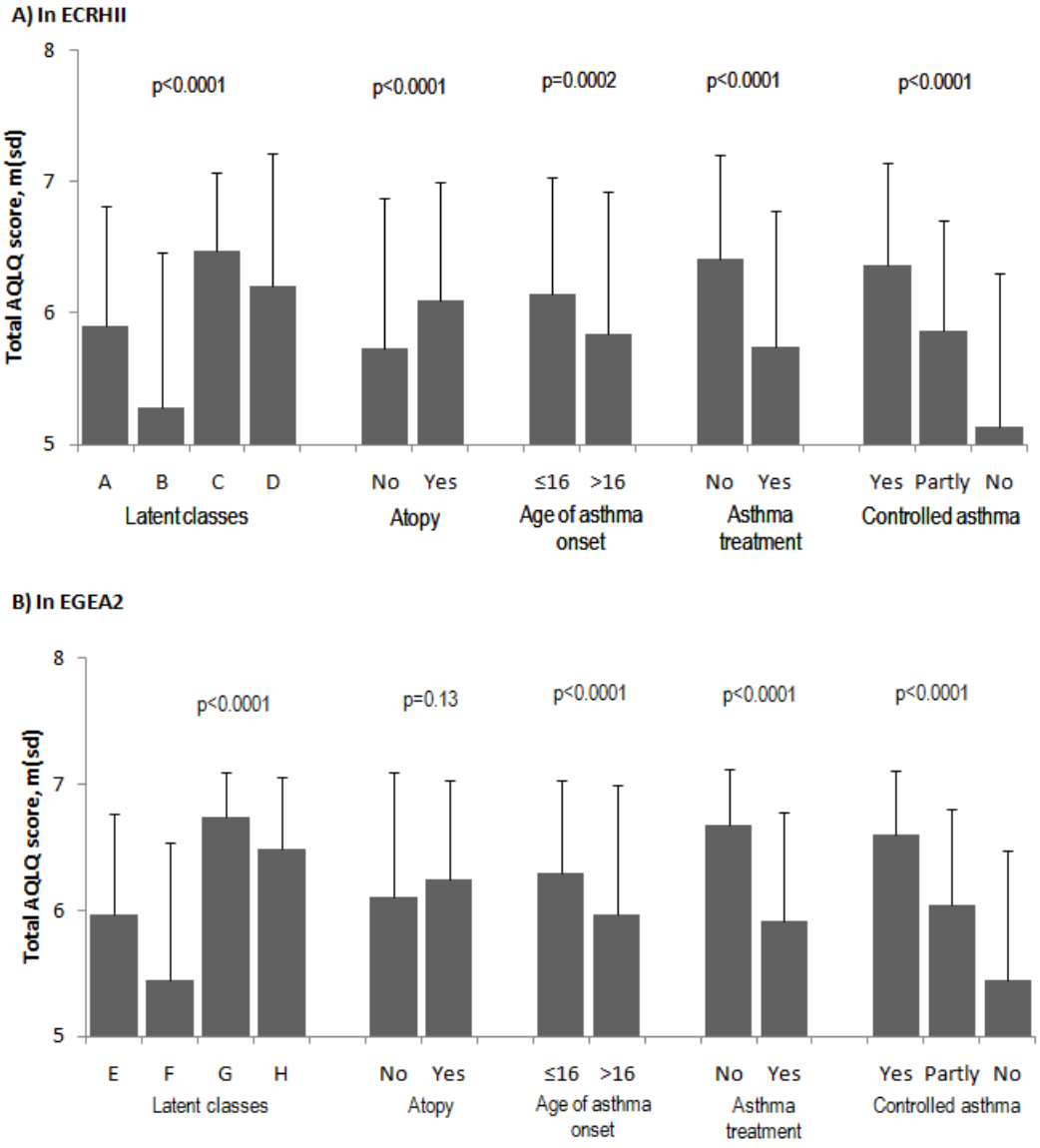

TABLE 1. Description of the individuals with asthma in the ECRHSII and EGEA2 studies

	ECRHS II n=1895	**EGEA2** n=641	**P value adjusted on age and sex***
Age: <40, n (%)	760 (40.1)	359 (56.0)	**<0.0001**
≥ 40, n (%)	1135 (59.9)	282 (44.0)	
Sex: Men, n (%)	780 (41.2)	337 (52.6)	**<0.0001**
Women, n (%)	1115 (58.8)	304 (47.4)	
Age of asthma onset: ≤4 years, n (%)	289 (15.8)	189 (31.4)	**<0.0001**
]4-16] years, n (%)	558 (30.5)	210 (34.9)	
>16 years, n (%)	982 (53.7)	203 (33.7)	
Wheezing with breathlessness, 12 months			0.16
No, n (%)	871 (46.4)	312 (48.9)	
Yes, n (%)	1006 (53.6)	326 (51.1)	
Woken up with feeling of tightness, 12 months			0.25
No, n (%)	1072 (56.6)	344 (53.9)	
Yes, n (%)	822 (43.4)	294 (46.1)	
Attack of shortness of breath at rest, 12 months			0.69
No, n (%)	1397 (73.8)	478 (75.2)	
Yes, n (%)	496 (26.2)	158 (24.8)	
Shortness of breath during activity, 12 months			**0.0008**
No, n (%)	961 (50.8)	280 (44.0)	
Yes, n (%)	931 (49.2)	357 (56.0)	
Woken up shortness of breath, 12 months			**0.02**
No, n (%)	1409 (74.5)	506 (79.7)	
Yes, n (%)	482 (25.5)	129 (20.3)	
Woken up by attack of coughing, 12 months			0.07
No, n (%)	1046 (55.3)	395 (62.0)	
Yes, n (%)	846 (44.7)	242 (38.0)	
Chronic cough or phlegm: No, n (%)	1463 (77.5)	542 (85.3)	**<0.0001**
Yes, n (%)	426 (22.5)	93 (14.6)	
Asthma symptom score 12 months			0.29
0 symptom, n (%)	452 (24.2)	135 (21.5)	
1 or 2 symptoms, n (%)	750 (40.2)	276 (43.9)	
≥3 symptoms, n (%)	665 (35.6)	218 (34.6)	
Asthma attack 12 months: No, n (%)	1061 (55.5)	406 (64.0)	**0.0005**
Yes, n (%)	818 (43.5)	228 (36.0)	
Exacerbation 12 months: No, n (%)	1492 (89.8)	534 (85.2)	**0.001**
Yes, n (%)	170 (10.2)	93 (14.8)	
Asthma treatment 3 months:			**<0.0001**
No asthma treatment, n (%)	691 (44.0)	254 (42.2)	
Other than daily ICS, n (%)	611 (38.9)	203 (33.7)	
Daily ICS, n (%)	268 (17.1)	145 (24.1)	
Eczema: No, n (%)	807 (42.8)	326 (51.3)	**0.0007**
Yes, n (%)	1077 (57.2)	309 (48.7)	
Rhinitis: No, n (%)	659 (34.9)	171 (27.2)	**0.002**
Yes, n (%)	1230 (65.1)	457 (72.8)	
Atopy: No, n (%)	534 (36.3)	110 (20.3)	**<0.0001**
Yes, n (%)	939 (63.7)	431 (79.7)	
IgE : <100 IU/ml, n (%)	802 (54.4)	224 (38.6)	**<0.0001**
≥100 IU/ml, n (%)	673 (45.6)	357 (61.4)	
FEV1: ≥80% predicted, n (%)	1353 (87.0)	491 (85.0)	0.05
<80% predicted, n (%)	202 (13.0)	87 (15.0)	
AHR, PD20≤1mg: No, n(%)	521 (51.5)	156 (52.3)	0.55
Yes, n(%)	490 (48.5)	142 (47.4)	

* Except for age and sex for which unadjusted p value are given

TABLE 2. Characteristics of the ECRHSII population and probability of individuals presenting the characteristics given membership in each of the 4 phenotypes identified by the latent class analysis

	Frequency of each variable in the whole sample	Phenotype A Active treated allergic childhood-onset asthma	Phenotype B Active treated adult-onset asthma	Phenotype C Inactive/mild untreated allergic asthma	Phenotype D Inactive/mild untreated non allergic asthma
Subjects, % (n)	100.0 (1895)	36.1 (685)	19.2 (363)	28.9 (548)	15.8 (299)
Age ≥40 years	0.60	0.50	0.76	0.55	0.73
Sex, men	0.41	0.49	0.25	0.46	0.33
Age of asthma onset					
≤4 years	0.16	0.20	0.04	0.16	0.20
]4-16] years	0.30	0.40	0.12	0.38	0.17
>16 years	0.54	0.40	0.84	0.46	0.63
Woken by coughing 12 months	0.45	0.45	0.79	0.26	0.35
Asthma symptom score 12 months					
0 symptom	0.24	0.03	0.02	0.52	0.50
1 or 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 12 months	0.43	0.75	0.73	0.06	0.04
Exacerbation 12 months	0.10	0.17	0.22	0.00	0.01
Asthma treatment 3 months					
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	0.46	0.74	0.15	0.50	0.11
FEV1 <80% predicted	0.13	0.19	0.18	0.05	0.08
AHR, PD20≤ 1mg	0.48	0.76	0.44	0.36	0.20

TABLE 3. Characteristics of the EGEA2 population and probability of individuals presenting the characteristics given membership in each of the 4 phenotypes identified by the latent class analysis

	Frequency of each variable in the whole sample	Phenotype E Active treated allergic childhood-onset asthma	Phenotype F Active treated adult-onset asthma	Phenotype G Inactive/mild untreated allergic childhood-onset asthma	Phenotype H Inactive/mild untreated adult-onset asthma
Subjects, % (n)	100.0 (641)	34.6 (222)	15.0 (96)	24.8 (159)	25.6 (164)
Age ≥40 years	0.44	0.24	0.96	0.07	0.76
Sex, men	0.53	0.47	0.54	0.68	0.44
Age of asthma onset					
≤4 years	0.31	0.50	0.00	0.43	0.13
]4-16] years	0.35	0.43	0.05	0.50	0.27
>16 years	0.34	0.07	0.95	0.07	0.60
Woken by coughing 12 months	0.38	0.50	0.53	0.18	0.33
Asthma symptoms score 12 months					
0 symptom	0.21	0.00	0.08	0.48	0.33
1 or 2 symptoms	0.44	0.31	0.42	0.52	0.55
≥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 12 months	0.36	0.68	0.61	0.08	0.05
Exacerbation 12 months	0.15	0.26	0.37	0.00	0.00
Asthma treatment 3 months					
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 ≥100 IU/ml	0.61	0.79	0.54	0.75	0.29
FEV1 <80% predicted	0.15	0.15	0.51	0.01	0.07
AHR, PD20≤ 1mg	0.48	0.63	0.42	0.48	0.24

TABLE 4: Blood eosinophil and neutrophil counts in EGEA2 population according to the phenotypes identified by the latent class analysis

	Phenotype E (n=222) Active treated allergic childhood-onset asthma	Phenotype F (n=96) Active treated adult-onset asthma	Phenotype G (n=159) Inactive/mild untreated allergic childhood-onset asthma	Phenotype H (n=164) Inactive/mild untreated adult-onset asthma	p value
Neutrophil count, GM* [95% CI]	3866 [1897-7878]	4358 [2177-8724]	3524 [2018-6152]	3775 [2056-6933]	<0.0001
Eosinophil count, GM* [95% CI]	243 [72-817]	212 [47-956]	190 [57.0-636]	174 [50.2-600]	<0.0001

*GM: Geometric Mean