DR. JOACHIM HEINRICH (Orcid ID : 0000-0002-9620-1629)
DR. VALÉRIE SIROUX (Orcid ID : 0000-0001-7329-7237)

Article type : Original Article: Epidemiology and Genetics

Data-driven adult asthma phenotypes based on clinical characteristics are associated with asthma outcomes twenty years later

Anne Boudier[1], Sébastien Chanoine[1,2,3], Simone Accordini [4], Josep M. Anto[5], Xavier Basagaña[5,6,7], Jean Bousquet[8], Pascal Demoly[9], Judith Garcia-Aymerich[5,6,7], Frederic Gormand[10], Joachim Heinrich[11], Christer Janson[12], Nino Künzli[13], Régis Matran[14], Christophe Pison[15], Chantal Raherison[16], Jordi Sunyer[5,6,7], Raphaelle Varraso[8], Deborah Jarvis[17], Bénédicte Leynaert[18], Isabelle Pin[19], Valérie Siroux[1]

1. Inserm, Univ. Grenoble Alpes, CNRS, IAB, Team of environmental epidemiology applied to reproduction and respiratory health, 38000 Grenoble, France
2. Faculté de Pharmacie, Université Grenoble Alpes, Grenoble, France
3. Pôle Pharmacie, CHU Grenoble Alpes, Grenoble, France
4. Unit of Epidemiology and Medical Statistics, Department of Diagnostics and Public Health, University of Verona, Verona, Italy,
5. ISGlobal, Centre for Research in Environmental Epidemiology (CREAL), Barcelona, Spain
6. Universitat Pompeu Fabra (UPF), Barcelona, Spain
7. CIBER Epidemiología y Salud Pública (CIBERESP), Barcelona, Spain
8. INSERM, U1168: Aging and chronic diseases. Epidemiological and public health approaches, Villejuif, France

9. CHU Montpellier, Pneumology department, Montpellier, France
10. CHU de Lyon, Pneumology Department, Lyon, France
11. Institute and Outpatient Clinic for Occupational, Social and Environmental Medicine, university Hospital of Ludwig Maximilians University, Munich, Comprehensive Pneumology Centre Munich, German Centre for Lung Research. Ziemssenstrasse 1, 80336 Muenchen
12. Department of Medical Sciences: Respiratory, Allergy and Sleep Research, Uppsala University, Uppsala, Sweden
13. Swiss Tropical and Public Health Institute, University of Basel, Basel, Switzerland
14. CHU, Univ Lille, Lille, France
15. Université Grenoble Alpes, Clinique Universitaire de Pneumologie, Pôle Thorax et Vaisseaux, CHU de Grenoble, INSERM U1055, Grenoble, France.
16. Univ. Bordeaux, Inserm, Bordeaux Population Health Research Center, team EPICENE, UMR 1219, F-33000 Bordeaux, France
17. National Heart and Lung Institute, Imperial College, London, United-Kingdom
18. Inserm, Unit 1152, Team of Epidemiology, University Paris-Diderot, Paris, France
19. CHU Grenoble, Peadiatric department, Grenoble, France

- Risk for poor asthma outcomes varied between cluster-based asthma phenotypes defined 20 years earlier
- The clinical prognosis of the cluster-based asthma phenotypes was stronger as compared to classical phenotypes
- There was a tracking of asthma activity and lung function over the life course of each asthma cluster

Correspondence:

Valérie Siroux

Institute for Advanced Biosciences, Inserm U1209 /CNRS UMR5309, Univ Grenoble-Alpes,

Team of Environmental Epidemiology applied to reproduction and respiratory health

Site Santé, Allée des Alpes

38706 La Tronche cedex

Tel: +33 4 76 54 95 56

Mail: valerie.siroux@univ-grenoble-alpes.fr

Short title: twenty-year evolution of asthma phenotypes**.**

Manuscript Abstract

Funding information: The present analysis was funded by National PHRC 2012, the scientific committee "AGIR pour les Maladies Chroniques".

EGEA data collection were funded in part by Hospital program of clinical research (PHRC)-Paris, PHRC-Grenoble, national PHRC 2012, Scientific committee "AGIR pour les Maladies Chroniques", Merck Sharp & Dohme (MSD) and the GA2LEN project (Global Allergy and Asthma European Network).

The ECRHS data collection has had multiple funding sources within each centre (listed in the online data supplement). The coordination of ECRHS3 was funded by the Medical Research Council (Grant Number 92091).

ABSTRACT

Background: Research based on cluster analyses led to the identification of particular phenotypes confirming phenotypic heterogeneity of asthma. The long-term clinical course of asthma phenotypes defined by clustering analysis remains unknown, although it is a key aspect to underpin their clinical relevance. We aimed to estimate risk of poor asthma events between asthma clusters identified 20 years earlier.

Methods: The study relied on two cohorts of adults with asthma with 20-year follow-up, ECRHS (European Community Respiratory Health Survey) and EGEA (Epidemiological study on Genetics and Environment of Asthma). Regression models were used to compare asthma characteristics (current asthma, asthma exacerbations, asthma control, quality of life and FEV_1) at follow-up and the course of FEV_1 between seven cluster-based asthma phenotypes identified 20 years earlier.

Results: The analysis included 1325 adults with ever asthma. For each asthma characteristic assessed at follow-up, the risk for adverse outcomes differed significantly between the seven asthma clusters identified at baseline. As compared with the mildest asthma phenotype, ORs (95%CI) for asthma exacerbations varied from 0.9 (0.4 to 2.0) to 4.0 (2.0 to 7.8) and the regression estimates (95%CI) for FEV_1 % predicted varied from 0.6 (-3.5 to 4.6) to -9.9 (-14.2 to -5.5) between clusters. Change of FEV_1 over time did not differ significantly across clusters.

Conclusion: Our findings show that the long-term risk for poor asthma outcomes differed between comprehensive adult asthma phenotypes identified 20-years earlier, and suggest a strong tracking of asthma activity and impaired lung function over time.

Key words: asthma, clustering, follow-up, lung function, phenotypes.

Abbreviations

EGEA: Epidemiological study of the Genetics and Environment of Asthma, bronchial hyper-responsiveness and atopy

ECRHS: European Community Respiratory Health Survey

LTA: Latent Transition Analysis

BHR: Bronchial Hyper Responsiveness

LF: Lung Function

ACT: Asthma Control Test

AQLQ: Asthma Quality of life Questionnaire

ICS: Inhaled Cortico-Steroids

FEV_1: Forced Expiratory Volume in 1 second

FVC: Forced Vital Capacity

INTRODUCTION

Asthma is a common respiratory disease characterized by phenotypic heterogeneity.(1) Asthma encompasses different disease subtypes, (2) each one possibly resulting from specific causes and characterized by different natural histories. Taking into account the heterogeneity of the disease is critical both in asthma management and research.(1-3) Improving the phenotypic characterization may help identify unshared risk factors across phenotypes, as illustrated by genetic studies that identified gene variants associated with early-onset asthma and asthma with severe exacerbations.(4-6) From a clinical point of view, distinguishing different asthma phenotypes is needed to improve asthma treatment, one example being the development of biotherapeutic agents that target asthma with high IgE, Interleukin 5

mediated inflammation, or Th2 endotype.(7, 8) Yet, more research is needed to improve our ability to disentangle asthma heterogeneity, allowing better prediction of the subsequent risk for poor asthma events and to better tailor asthma management strategies.

In addition to studies based on clinical hypotheses, data-based approaches relying on numerous asthma-related characteristics have been used to unravel disease heterogeneity.(3, 9) Applying clustering analysis in various populations of adults and children with asthma, different asthma phenotypes have been identified.(10-14) These cluster-based phenotypes have shown some similarities across different studies, as well as consistencies with the clinical phenotypes.(15) In particular, most of the studies identified the age of asthma onset and the level of current disease activity as key features in the discrimination of the phenotypes.(2) However, the clinical relevance of these cluster-based asthma phenotypes has not been rigorously investigated. In particular, few studies investigated the outcomes associated with cluster grouping, and most relied on short to medium follow-up (1 to 3 years).(16-19) A long-term prospective study was conducted on 299 child participants in the Childhood Asthma Management Program study (CAMP) and showed that differences in clinical characteristics observed between children assigned to different clusters persisted into young adulthood.(20) Further studies are needed to explore the long-term prognostic value of these phenotypes in a large population of adults with asthma.

In a previous study, we showed a strong consistency in the structure of adult asthma phenotypes identified twice ten years apart.(12) In addition, by modeling the individuals' phenotypic transition across these asthma phenotypes, we showed that the probabilities of remaining in the same phenotype at each time point varied between 54 to 88% across phenotypes. In the present study, we took advantage of a new follow-up of this population, 20 years after recruitment, to further investigate these asthma phenotypes previously identified by clustering approaches, and specifically to investigate the long-term prognosis of these

phenotypes.(12) We hypothesized that the comprehensive cluster-based asthma phenotypes are strongly associated with asthma outcomes assessed 20-years later.

METHODS

Study design

We analyzed longitudinal data from two large epidemiological cohorts, the European Community Respiratory Health Survey (ECRHS) and the Epidemiological Study on the Genetics and Environment of Asthma (EGEA).(21, 22) ECRHS is an international population-based study focused on respiratory health over 18,000 participants were originally recruited (ECRHS I) from 29 centers in 14 countries in 1991–1993 using population-based registers, with an additional sampling of asthmatics. Two follow-ups have since taken place: ECRHS II in 1999–2003 and ECRHS III in 2010–2014. EGEA (https://egeanet.vjf.inserm.fr) is a French cohort including a group patients with asthma and their first-degree relatives and a group of control subjects. In total, 2047 adults and children were recruited from 1991 to 1995 (EGEA1). A first follow-up of the EGEA population was conducted from 2003 to 2007 (EGEA2; 1845 subjects), and a second follow-up was conducted from 2011 to 2013 (EGEA3; 1558 subjects). In both studies, all surveys included a detailed and standardized respiratory questionnaire and lung function testing, measures of bronchial responsiveness, skin prick tests, and total IgE measurement (except for EGEA3 consisting of a self-completed questionnaire). The EGEA and ECRHS studies were approved by the appropriate ethics committees.

Cluster-based asthma phenotypes at baseline (EGEA1/ECRHSI)

The baseline asthma phenotypes considered in this study were obtained using the Latent Transition Analysis (LTA), a cluster-based model developed for longitudinal data,(23) and have previously been published.(12, 23) Shortly, LTA was performed in adults with ever asthma using two time points and using nine variables covering personal and phenotypic characteristics: respiratory symptoms, asthma treatment, allergic characteristics, lung function, and Bronchial Hyper Responsiveness (BHR) (see Table S1 in the online data supplement). The models identified seven phenotypes at EGEA1/ECRHSI (Figure S1 and table S1): phenotype B was composed of subjects with no or few symptoms, no treatment and few allergic sensitization; phenotype A showed similar characteristics to phenotype B but almost all subjects had allergic sensitization; phenotypes E, F and G were composed of subjects with moderate symptoms: almost all subjects belonging to phenotype E were allergic and had bronchial hyper-reactivity, subjects belonging to phenotype F were predominantly allergic and had a normal lung function and subjects belonging to phenotype G were predominantly non allergic and had no asthma treatment; phenotype C and D were composed of subjects with high symptoms and asthma treatment, but differ in the allergic sensitization.

Asthma outcomes defined at the 20-year follow-up (EGEA3/ECRHSIII)

The follow-up asthma outcomes considered in this study were: current asthma (defined by asthma attacks or asthma treatment use in the past 12 months), asthma exacerbations (defined by the report of oral corticosteroid use for breathing difficulties or hospitalization in the past 12 months), asthma control (defined by the asthma control test (ACT)),(24) quality of life (defined by the total Asthma Quality of Life Questionnaire (AQLQ) score)(25) and spirometry parameters (using the FEV_1 and FEV_1/FVC in ECRHS) (26) (more details in online data supplement).

"Classical" asthma phenotypes at baseline (EGEA1/ECRHSI)

We ascertained whether the comprehensive cluster-based asthma phenotypes were more strongly associated with long term asthma characteristics than asthma phenotypes more classically used in epidemiological studies: current *vs.* past asthma (defined by positive answer to the following question "have you had an attack of asthma in the past 12 months" or asthma treatment use in the past 12 months), atopic *vs.* non-atopic asthma (defined by ≥ 1 positive skin-prick tests to any of 9 aero-allergens for ECRHSI and 11 aero-allergens for EGEA1) and childhood *vs.* adult-onset asthma (<16 years, ≥16 years).

Statistical analyses

To study the clinical prognosis of asthma phenotypes identified by clustering method at baseline, we fitted logistic or linear regression models with the asthma characteristics estimated at the 20-year follow-up as dependent variables and the cluster-based asthma phenotypes at baseline as predictors. Unadjusted models and models adjusted on age, sex, smoking status and occupational status were conducted. The LTA model provides for each subject a probability to belong to each class (the membership probability); in our analysis we assigned each subject to the latent class for which he had the highest membership probability, and we weighted the regression model on this highest membership probability to account for the uncertainty in the latent class assignment. To investigate the added prognostic value of these cluster-based asthma phenotypes to the asthma phenotypes more classically used in epidemiological studies, we compared the magnitude of the Odds Ratios (ORs) for poor asthma outcomes from cluster-based asthma phenotypes and from classical asthma phenotypes. Moreover risks for current asthma and asthma exacerbation at follow-up associated with the cluster-based asthma phenotypes were estimated in a model further

adjusted on current asthma or asthma exacerbation at baseline to assess the statistical independent effect of the cluster-based phenotypes.

The main analysis was conducted in the pooled data (EGEA and ECRHS). To address the robustness of the results within each cohort, the same models as above were applied separately in EGEA and ECRHS participants. A sensitivity analysis was also conducted in men and women separately and we addressed whether regular use of ICS during the follow-up modified the associations.

In ECRHS, the 20-year decline of lung function was compared between the cluster-based asthma phenotypes. For each asthma phenotype, a linear mixed-effects model for repeated measurements (ECRHSI, II, and III) was fitted with FEV_1 (in ml or FEV_1 divided by the square of the height in meters) as the dependent variable, time (0 for baseline or the number of years between the follow-up and baseline for each follow-up) as the predictor, and a random intercept to account for repeated data. Sex, age, height, occupational status, and smoking were added in the model as covariates. To statistically compare the FEV_1 decline between the asthma phenotypes, the linear mixed-effect model was conducted with the addition of the cluster-based asthma phenotypes and its interaction with time among the predictors. Similar analysis was conducted for FEV_1/FVC.

To account for multiple hypotheses testing, the Bonferroni method was applied to the main analysis, which consisted in estimating the association between seven cluster-based asthma phenotypes and six asthma outcomes at follow-up (threshold p-value=0.0012). All analyses were performed with the SAS 9.4 statistical software.

RESULTS

Population description at baseline

The population comprised 1325 adults with asthma (1026 from ECRHS and 299 from EGEA) with available data for cluster-based asthma phenotypes at EGEA1/ECRHSI and followed-up at EGEA3/ECRHSIII (mean follow-up was 19.7±1.1 years) (see Figure S2 in the online data supplement). Among participants with asthma, those followed-up were older, less often smokers and more often managers, and reported slightly more often allergic rhinitis but did not strongly differ from the non-participants with regard to the clinical characteristics at baseline (Table S2). Mean age of the included population at baseline was 34.9 years and 43% were men (Table 1). Half of the population had never smoked and one third had an occupational status of manager. More than half of the subjects had asthma onset after 16 years. At baseline, 60% of subject had current asthma in the past 12 months. Regarding allergic characteristics, more than half of the subjects had eczema (51.7%), allergic rhinitis (64.8%) and atopy (67.8%).

In our study population, the frequencies of the cluster-based asthma phenotypes at baseline were: 14.6% and 18.2% for asthma phenotypes composed of few symptoms (B and A respectively), 10.2%, 14.5% and 10.4% for asthma phenotypes composed of moderate symptoms (G, F, and E respectively), and 9% and 23% for asthma phenotypes composed of high symptoms (C and D respectively) (Table 1 and Figure S1). The cluster-based asthma phenotypes are associated with the classical phenotypes at baseline (Table S1).

Risk for poor asthma outcomes after 20 years of follow-up, according to cluster-based asthma phenotypes at baseline

Asthma outcomes observed at follow-up significantly differed between the cluster-based asthma phenotypes identified 20 years earlier (Table 2). More than 80% of subjects belonging to phenotype C ("non-allergic, high symptoms, treatment") or D ("allergic, high symptoms, treatment, BHR"), the most symptomatic phenotypes at baseline, had current asthma at follow-up compared with only about one third for phenotypes B ("non-allergic, few symptoms, no treatment") and A ("allergic, few symptoms, no treatment"). In addition, subjects belonging to phenotypes C and D at baseline were more likely to have asthma exacerbations, poor asthma control, worse quality of life and a low FEV_1 at follow-up than subjects belonging to the other clusters.

The associations between cluster-based asthma phenotypes at baseline and asthma characteristics at follow-up remained in the adjusted analysis (Figure 1; Table S3). For most of the asthma characteristics, higher risks of poor outcomes were observed for phenotypes C ("non-allergic, high symptoms, treatment"), D ("allergic, high symptoms, treatment, BHR") and G ("non-allergic, moderate symptom, no treatment") than for phenotype B ("non-allergic, few symptoms, no treatment"). For example, OR point estimates for increased prevalence of current asthma were 1.1, 2.2, 2.9, 4.1, 8.5 and 12.9 for phenotypes A, G, F, E, C and D as compared with phenotype B, respectively. Models further adjusted on current asthma or asthma exacerbations at baseline led to similar conclusions (Table S4).

Consistent patterns of associations were observed in separate analysis in EGEA and ECRHS (Table S5), in men and women (Table S6) and when adjusting on cohorts (data not shown). In addition, separate analysis by use of ICS during the follow-up showed associations with total AQLQ score and FEV_1 level of similar magnitude in each sub-group (Table S7, p value for the interaction=0.56 and 0.61 respectively).

Risk for poor asthma outcomes after 20 years of follow-up, according to "classical" asthma phenotypes at baseline

Considering the classical asthma phenotypes at baseline, subjects with current asthma at baseline had a significant higher risk for all asthma characteristics investigated at follow-up, except for quality of life (Table 3). No marked differences were observed for phenotypes defined by allergic sensitization status and age at asthma-onset (Table 3). As compared with associations relying on classical asthma phenotypes, associations with cluster-based asthma phenotypes were usually of similar or larger magnitude and systematically exhibited the strongest risk. For example, ORs [95% CI] observed for current asthma varied from 0.9 (0.8 to 1.3) to 3.3 (2.6 to 4.2) between classical asthma phenotypes at baseline (Table 3), while these ORs increased to 12.9 (7.8 to 21.2) for phenotype D when considering cluster-based asthma phenotypes at baseline. The 95% CI observed for phenotype D do not overlap with any of those estimated for classical phenotypes (Table S3).

Long-term changes in lung function between cluster-based asthma phenotypes

Cross-sectional analyses for the three ECRHS surveys (ECRHSI, II, and III) showed that mean FEV_1 and FEV_1/FVC %predicted significantly differed between the cluster-based asthma phenotypes, with similar differences between phenotypes at each survey (p values were <0.0001 for FEV_1 and FEV_1/FVC at each survey) (Figure 2). FEV_1 %predicted values were consistently higher in cluster A ("allergic, few symptoms, no treatment") and lower in cluster E ("allergic, moderate symptoms, BHR") and D ("allergic, higher symptoms, treatment, BHR"). Mean level of FEV_1 at baseline varied from 3090 ml for cluster D ("allergic, higher symptoms, treatment, BHR") to 3738 ml for cluster A ("allergic, few symptoms, no treatment") (Table 4). Mean FEV_1 change in our population was -33.7 (SE=0.7) ml/year. Changes in FEV_1 did not differ significantly between the cluster-based

asthma phenotypes, except for subjects belonging to cluster D ("allergic, higher symptoms, treatment, BHR") who had a slower FEV_1 decline than those in cluster B ("non-allergic, few symptoms, no treatment") ($p<0.0001$) (Table 4). Similarly, changes in FEV_1/FVC did not differ significantly between the cluster-based asthma phenotypes, except for cluster D ("allergic, higher symptoms, treatment, BHR") (Figure 2; Table S8). Subjects with excessive FEV_1 decline between ECRHSI and ECRHSIII (≥50ml/yr, n=138) were equally distributed across the cluster-based asthma phenotype at baseline ($p=0.17$, data not shown).

DISCUSSION

To our knowledge, this is the first study that shows the long-term prognosis of adult asthma phenotypes, defined by an hypothesis-free approach combining clinical and functional characteristics. We observed that adults with asthma classified in cluster-based phenotypes exhibit statistically significant differences in several clinical and functional characteristics 20 years later; in a general way, phenotypes characterized by the highest level of symptoms or use of asthma treatment at baseline showed the highest risk for adverse outcomes. Our study suggests a stronger clinical prognosis of the cluster-based asthma phenotypes as compared to classical phenotypes. These results underline the importance of considering unsupervised asthma phenotypes, that are able to account for multiple asthma characteristics, to understand the long-term evolution of asthma patients.

Our study with long-term follow-up adds to the existing literature supporting the clinical significance of adult asthma clusters over shorter periods. A Korean study with a 1-year follow-up showed that change in FEV_1 and use of oral corticosteroids significantly differed across the four groups of patients previously identified using a clustering method.(17) Five distinct phenotypes identified by cluster analysis in primary care showed differences in asthma outcomes after 12-month follow-up, the "exacerbators" phenotypes exhibiting the

highest risk for adverse outcomes, despite using the most medication.(16) In our study, the clinical relevance of these phenotypes is strengthened by analysis showing that 1) these phenotypes differ for asthma characteristics at follow-up after adjusting for these similar characteristics at baseline, and 2) for each adverse outcome, the strongest risk was systematically observed among the cluster-based asthma phenotypes when compared to classical asthma phenotypes. Our findings also extend previous results observed in children to adult populations. Indeed, a study relying on children with asthma included in the CAMP cohort identified five phenotypes and showed significant differences in lung function level and use of ICS two years later.(27) In addition, a more recent study in the CAMP population showed that differences in clinical characteristics observed between clusters persisted into young adulthood.(28) On the contrary, some recent studies based on moderate to severe asthma questioned the relevance of these cluster methods and the stability over time of the identified phenotypes.(18, 19, 29) Our study is the only one with a population based cohort of people with asthma and yet showing that our population based and our clinical cohort were homogeneous regarding their classification in unsupervised phenotypes. Differences in the study design as well as in the mix of patients recruited in the reported clinical studies may explain the lack of consistency. By including subjects with ever asthma, -both with remittent or current asthma and with a wide range of asthma severity, our study led to the identification of broader range of phenotypes, which are expected to be more stable over-time as compared with severe asthma sub-types. Nevertheless, we cannot totally dismiss that part of the subjects belonging to phenotype A and B wrongly reported ever asthma, which might have led to overestimate the prognostic value of the clusters.

Our findings showing that subjects belonging to the clusters characterized by low lung function at baseline still had lower lung function 20 years later, support a strong tracking of lung function over the life course, as recently underlined by two studies.(31, 32) Lange and

colleagues showed that a substantial proportion of individuals who developed COPD had a low FEV_1 level in early adulthood and that COPD group did not necessarily follow a trajectory of rapid decline in FEV_1.(32) A further recent study showed that low lung function levels in early childhood predicted cardiopulmonary mortality.(31) These studies highlight the major role of the maximally attained lung function in early adulthood in the future respiratory health events.

We found no marked differences for change in lung function between the asthma phenotypes, except for a lower FEV_1 decline observed in the most "severe" phenotype at baseline (labeled "allergic, higher symptoms, treatment, BHR"). Several explanations may account for the lower lung function decline in this asthma phenotype; first, patients from this group are symptomatic, therefore might have received higher corticosteroids as compared to patients with milder asthma phenotypes, thus preventing a further decline in lung function. This hypothesis is supported by the observation that 19% of ECRHS participants belonging to phenotype D reported to have used ICS each year during the follow-up period, whereas this number was below 7% for the milder asthma phenotypes. Second, these allergic patients might have higher clinical benefits of corticosteroids as compared with phenotype C ("non-allergic, high symptoms, treatment") characterized by non-allergic inflammation.(33, 34) An interesting point is that the phenotype showing a trend for a faster decline in FEV_1 is the "mildest" asthma phenotype ("non-allergic, few symptoms, no treatment"), which suggest that some asymptomatic and well controlled patients might benefit from early treatment.

Our study has a number of strengths, including the use of two large epidemiological studies, with long follow-up (20 years) and with good and large phenotypic characterizations. Standardized protocols and questionnaires were used in both studies. The similarity of the results in the two cohorts, which have been set-up using different study design (ECRHS is a population-based study and EGEA is a cohort including a group of cases recruited in chest

clinics), indicates that the study design did not influence our findings and adds to the generalizability of our findings. Importantly, each asthma phenotypes comprises more than 100 subjects.

One limitation of our study is the lack of biomarkers in the characterization of the clusters to account for the underlying molecular mechanisms. However, most of the variables considered in the clustering are easily available in primary care, where most of the patients with asthma are followed. Smoking, a factor previously shown to drive some of the clinical asthma clusters,(35) was not included in the cluster analysis. However, in a complex picture, where many other lifestyle, environmental and genetic factors are expected to contribute to the clinical asthma heterogeneity, accounting for a single determinant is arguable and it might lead to bias in future analyses aimed at identifying risk factors for these asthma phenotypes. Another limitation of our study is the loss of follow up, with followed-up participants showing differences in sociodemographic characteristics (age, smoking, socio-economic position), as usually observed in many cohorts. However, a 20-year follow-up rate of 54% in population-based studies can be considered as acceptable and we showed that the followed-up participants did not strongly differ from the non-participants in regard to the main clinical characteristics, and in particular across the cluster-based asthma phenotypes, limiting the risk for selection bias in our study. The applied clustering method relied on characteristics measured at a single point in time, therefore does not account for the short-term fluctuations of the parameters that might contain information for asthma phenotyping, as recently shown for the lung function parameters.(36) The analysis relies on multiple testing, possibly leading to inflation of the false discovery rate. However, most of the associations observed in the main analysis remained significant when correcting for multiple testing with the conservative Bonferroni correction method, supporting the robustness of our findings. Finally, asthma treatments (ICS) used in the past 12 months at baseline were used in the clustering methods

to identify the asthma phenotypes, but the level of and adherence to asthma treatment during the follow-up were not considered in the association studies. Nevertheless, the similarity of the estimates observed for FEV_1 level in regular and no/irregular ICS users during the follow-up suggests that ICS did not modify the association observed. Further studies are needed to address if the tracking effect observed both for the disease activity and lung function means that treatment is not changing the natural history of the disease.

In conclusion, we confirmed the clinical relevance of asthma clusters by showing the long-term prognosis of our previously identified cluster-based adult asthma phenotypes. Our results lend further support to the strong tracking of asthma activity and lung function over the life course for each asthma cluster.

AUTHOR CONTRIBUTIONS: V.S. and I.P. defined the research question. A.B., S.C. and V.S. designed analytical and statistical methods. A.B. analyzed the data. A.B., V.S. and I.P. interpreted the results, and wrote the paper. DJ coordinated the multicenter data collection as part of the ECHRS including design of the protocols for questionnaire and clinical assessment. All authors have contributed to the data collection and to the manuscript. All authors approved the manuscript.

CONFLICT OF INTEREST: DJ reports grants from MRC, grants from EU, grants from Asthma UK, during the conduct of the study. IP reports other from Astra Zeneca, other from Novartis, other from GSK, outside the submitted work. PD reports personal fees from ALK, personal fees from Stallergènes Greer, personal fees from MEDA-MYLAN, personal fees from Chiesi, personal fees from ThermoFisher Scientific, personal fees from Menarini, personal fees from AstraZeneca, personal fees from ASIT Biotech, outside the submitted

work. CP reports grants, personal fees and non-financial support from GSK France, grants, personal fees and non-financial support from Novartis France, grants, personal fees and non-financial support from Actélion France, grants, personal fees and non-financial support from BIF France, outside the submitted work. JB reports personal fees and other from Chiesi, Cipla, Hikma, Menarini, Mundipharma, Mylan, Novartis, Sanofi-Aventis, Takeda, Teva, Uriach, other from Kyomed, from null, outside the submitted work. SC reports personal fees from AstraZeneca, non-financial support from Boehringer Ingelheim, non-financial support from Actelion Pharmaceuticals, non-financial support from MSD, outside the submitted work. VS reports personal fees from TEVA, AstraZeneca and Novartis-France, outside the submitted work. All other authors declare no conflicts of interests.

ACKNOWLEDGMENTS: The authors thank all those who participated to the setting of the EGEA and ECRHS study and on the various aspects of the examinations. The authors are grateful to the three CIC-Inserm centres of Necker, Grenoble and Marseille who were in charge of EGEA participants examination. They are indebted to all the individuals who participated without whom that study would not have been possible.

We would like to thank EGEA cooperative group and steering committee for the ECRHS (listed in the online data supplement).

SUPPORTING INFORMATION: Additional supporting information may be found online in the Supporting Information section at the end of the article.

REFERENCES

1. Pavord ID, Beasley R, Agusti A, Anderson GP, Bel E, Brusselle G, et al. After asthma: redefining airways diseases. *The Lancet* 2017;**391**:350-400.
2. Wenzel SE. Asthma phenotypes: the evolution from clinical to molecular approaches. *Nature Medicine* 2012;**18**:716-725.
3. Siroux V, Garcia-Aymerich J. The investigation of asthma phenotypes. *Curr Opin Allergy Clin Immunol* 2011;**11**:393-399.
4. Bouzigon E, Corda E, Aschard H, Dizier MH, Boland A, Bousquet J, et al. Effect of 17q21 variants and smoking exposure in early-onset asthma. *N Engl J Med* 2008;**359**:1985-1994.
5. Bonnelykke K, Sleiman P, Nielsen K, Kreiner-Moller E, Mercader JM, Belgrave D, et al. A genome-wide association study identifies CDHR3 as a susceptibility locus for early childhood asthma with severe exacerbations. *Nat Genet* 2014;**46**:51-55.
6. Dizier MH, Margaritte-Jeannin P, Madore AM, Moffatt M, Brossard M, Lavielle N, et al. The nuclear factor I/A (NFIA) gene is associated with the asthma plus rhinitis phenotype. *J Allergy Clin Immunol* 2014;**134**:576-582 e571.
7. Borish L. The immunology of asthma: Asthma phenotypes and their implications for personalized treatment. *Ann Allergy Asthma Immunol* 2016;**117**:108-114.
8. Krug N, Hohlfeld JM, Kirsten AM, Kornmann O, Beeh KM, Kappeler D, et al. Allergen-induced asthmatic responses modified by a GATA3-specific DNAzyme. *N Engl J Med* 2015;**372**:1987-1995.
9. Deliu M, Sperrin M, Belgrave D, Custovic A. Identification of Asthma Subtypes Using Clustering Methodologies. *Pulm Ther* 2016;**2**:19-41.
10. Moore WC, Meyers DA, Wenzel SE, Teague WG, Li H, Li X, 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.
11. Siroux V, Basagana X, Boudier A, Pin I, Garcia-Aymerich J, Vesin A, et al. Identifying adult asthma phenotypes using a clustering approach. *Eur Respir J* 2011;**38**:310-317.

12. Boudier A, Curjuric I, Basagana X, Hazgui H, Anto JM, Bousquet J, et al. Ten-year follow-up of cluster-based asthma phenotypes in adults. A pooled analysis of three cohorts. *American journal of respiratory and critical care medicine* 2013;**188**:550-560.

13. Fitzpatrick AM, Teague WG, Meyers DA, Peters SP, Li X, Li H, et al. Heterogeneity of severe asthma in childhood: Confirmation by cluster analysis of children in the National Institutes of Health/National Heart, Lung, and Blood Institute Severe Asthma Research Program. *J Allergy Clin Immunol* 2011;**127**:382-389.

14. Ilmarinen P, Tuomisto LE, Niemela O, Tommola M, Haanpaa J, Kankaanranta H. Cluster Analysis on Longitudinal Data of Patients with Adult-Onset Asthma. *J Allergy Clin Immunol Pract* 2017;**5**:967-978 e963.

15. Prosperi MC, Sahiner UM, Belgrave D, Sackesen C, Buchan IE, Simpson A, et al. Challenges in identifying asthma subgroups using unsupervised statistical learning techniques. *Am J Respir Crit Care Med* 2013;**188**:1303-1312.

16. Khusial RJ, Sont JK, Loijmans RJB, Snoeck-Stroband JB, Assendelft PJJ, Schermer TRJ, et al. Longitudinal outcomes of different asthma phenotypes in primary care, an observational study. *NPJ Prim Care Respir Med* 2017;**27**:55.

17. Park SY, Baek S, Kim S, Yoon SY, Kwon HS, Chang YS, et al. Clinical significance of asthma clusters by longitudinal analysis in Korean asthma cohort. *PLoS One* 2013;**8**:e83540.

18. Bourdin A, Molinari N, Vachier I, Varrin M, Marin G, Gamez AS, et al. Prognostic value of cluster analysis of severe asthma phenotypes. *J Allergy Clin Immunol* 2014;**134**:1043-1050.

19. Newby C, Heaney LG, Menzies-Gow A, Niven RM, Mansur A, Bucknall C, et al. Statistical cluster analysis of the British Thoracic Society Severe refractory Asthma Registry: clinical outcomes and phenotype stability. *PLoS One* 2014;**9**:e102987.

20. Howrylak JA, Fuhlbrigge AL, Strunk RC, Zeiger RS, Weiss ST, Raby BA, et al. Classification of childhood asthma phenotypes and long-term clinical responses to inhaled anti-inflammatory medications. *J Allergy Clin Immunol* 2014;**133**:1289-1300, 1300 e1281-1212.

21. Burney PG, Luczynska C, Chinn S, Jarvis D. The European Community Respiratory Health Survey. *Eur Respir J* 1994;**7**:954-960.

22. Kauffmann F, Dizier MH, Pin I, Paty E, Gormand F, Vervloet D, 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.

23. Lanza ST, Bray BC. Transitions in Drug Use among High-Risk Women: An Application of Latent Class and Latent Transition Analysis. *Advances and applications in statistical sciences* 2010;**3**:203-235.

24. Nathan RA, Sorkness CA, Kosinski M, Schatz M, Li JT, Marcus P, et al. Development of the asthma control test: a survey for assessing asthma control. *J Allergy Clin Immunol* 2004;**113**:59-65.

25. 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.

26. Standardization of Spirometry, 1994 Update. American Thoracic Society. *Am J Respir Crit Care Med* 1995;**152**:1107-1136.

27. Haldar P, Pavord ID, Shaw DE, Berry MA, Thomas M, Brightling CE, et al. Cluster analysis and clinical asthma phenotypes. *Am J Respir Crit Care Med* 2008;**178**:218-224.

28. Howrylak JA, Moll M, Weiss ST, Raby BA, Wu W, Xing EP. Gene expression profiling of asthma phenotypes demonstrates molecular signatures of atopy and asthma control. *J Allergy Clin Immunol* 2016;**137**:1390-1397 e1396.

29. Zaihra T, Walsh CJ, Ahmed S, Fugere C, Hamid QA, Olivenstein R, et al. Phenotyping of difficult asthma using longitudinal physiological and biomarker measurements reveals significant differences in stability between clusters. *BMC Pulm Med* 2016;**16**:74.

30. Moore WC. The natural history of asthma phenotypes identified by cluster analysis. Looking for chutes and ladders. *Am J Respir Crit Care Med* 2013;**188**:521-522.

31. Vasquez MM, Zhou M, Hu C, Martinez FD, Guerra S. Low Lung Function in Young Adult Life Is Associated with Early Mortality. *Am J Respir Crit Care Med* 2017;**195**:1399-1401.

32. Lange P, Celli B, Agusti A, Boje Jensen G, Divo M, Faner R, et al. Lung-Function Trajectories Leading to Chronic Obstructive Pulmonary Disease. *N Engl J Med* 2015;**373**:111-122.

33. Fajt ML, Wenzel SE. Asthma phenotypes and the use of biologic medications in asthma and allergic disease: the next steps toward personalized care. *J Allergy Clin Immunol* 2015;**135**:299-310; quiz 311.

34. Fahy JV. Type 2 inflammation in asthma--present in most, absent in many. *Nat Rev Immunol* 2015;**15**:57-65.

35. Lefaudeux D, De Meulder B, Loza MJ, Peffer N, Rowe A, Baribaud F, et al. U-BIOPRED clinical adult asthma clusters linked to a subset of sputum omics. *J Allergy Clin Immunol* 2017;**139**:1797-1807.

36. Delgado-Eckert E, Fuchs O, Kumar N, Pekkanen J, Dalphin JC, Riedler J, et al. Functional phenotypes determined by fluctuation-based clustering of lung function measurements in healthy and asthmatic cohort participants. *Thorax* 2018;**73**:107-115.

Table 1. Description of the population at baseline

	ALL (n=1325)		EGEA (n=299)		ECRHS (n=1026)	
	n	**statistic**	**n**	**statistic**	**n**	**statistic**
Personnel characteristics						
Age, mean ± sd	1325	34.9 ± 9.0	299	38.0 ± 13.2	1026	34.0 ± 7.1
Sex, % men	568	42.9	137	45.8	431	42.0
BMI, mean ± sd	1164	23.9 ± 4.2	296	23.0 ± 3.3	868	24.2 ± 4.4
Smoking status						
Never smokers, %	662	50.1	161	54.4	501	48.8
Ex-smokers, %	316	23.9	81	27.4	235	22.9
Smokers, %	344	26.0	54	18.2	290	28.3
Occupational status						
No profession, %	283	21.5	54	18.5	229	22.3
Managers, %	430	32.6	100	34.2	330	32.2
Technicians, %	305	23.1	104	35.6	201	19.6
Skilled manual, %	300	22.8	34	11.6	266	25.9
Asthma characteristics						
Age of asthma onset						
≤4 years, %	209	16.4	49	17.6	160	16.0
]4-16] years, %	376	29.4	84	30.2	292	29.2
>16 years, %	692	54.2	145	52.2	547	54.8
Asthma attacks 12 months, %	588	44.7	167	57.2	421	41.2
Current asthma 12 months, %	781	59.9	214	75.9	567	55.5
Asthma exacerbations 12 months, %	180	14.8	108	38.8	72	7.7
FEV_1%predict, mean ± sd	1253	92.9 ± 15.9	295	89.6 ± 18.2	958	93.9 ± 15.0
FEV_1 in ml, mean ± sd	1253	3397.9±858.1	443	3183.5±871.5	958	3463.9±843.5
Inhaled corticosteroids 12m, %	311	25.2	134	45.4	177	18.8
Allergic characteristics						
Total IgE, ≥100 IU/ml, %	583	49.7	183	64.4	400	44.9
Eczema, %	684	51.7	132	44.4	552	53.8
Allergic rhinitis, %	857	64.8	195	65.7	662	64.5
Atopy, %	812	67.8	207	76.6	605	66.6
Cluster-based asthma phenotypes, %						
B: Non-allergic, few symptoms, no treatment	193	14.6	27	9.0	166	16.2
A: Allergic, few symptoms, no treatment	241	18.2	57	19.1	184	17.9
G: Non-allergic, moderate symptoms, no treatment	135	10.2	11	3.7	124	12.1
F: Allergic, moderate symptoms, normal LF	192	14.5	30	10.0	162	15.8
E: Allergic, moderate symptoms, BHR	138	10.4	22	7.4	116	11.3
C: Non-allergic, high symptoms, treatment	119	9.0	41	13.7	78	7.6
D: Allergic, high symptoms, treatment, BHR	307	23.2	111	37.1	196	19.1

LF: lung function

BHR: bronchial hyper-responsiveness

Table 2. Associations between cluster-based asthma phenotypes at baseline and different asthma characteristics assessed at follow-up

	Cluster-based asthma phenotypes at baseline							
Asthma characteristics assessed at follow-up	Phenotype B *"non-allergic, few symptoms, no treatment"*	Phenotype A *"allergic, few symptoms, no treatment"*	Phenotype G *"non-allergic, moderate symptoms, no treatment"*	Phenotype F *"allergic, moderate symptoms, normal LF"*	Phenotype E *"allergic, moderate symptoms, BHR"*	Phenotype C *"non-allergic, high symptoms, treatment"*	Phenotype D *"allergic, high symptoms, treatment, BHR"*	Global p value
Current asthma 12m, n (%)	64 (35.0)	82 (37.8)	64 (53.3)	107 (60.1)	86 (66.1)	96 (82.8)	260 (86.1)	<0.0001
OR [95% CI]	1	1.1 [0.7;1.7]	2.2 [1.3;3.7] [a]	3.0 [1.8;4.8] [a]	4.2 [2.4;7.2] [a]	8.9 [4.9;16.4] [a]	13.0 [7.9;21.2] [a]	
Asthma exacerbations 12m, n (%)	21 (12.1)	28 (13.4)	24 (21.0)	17 (11.0)	19 (15.7)	36 (34.3)	59 (21.8)	<0.0001
OR [95% CI]	1	1.1 [0.5;2.1]	2.0 [1.0;4.1] [a]	0.8 [0.4;1.8]	1.4 [0.6;2.9]	4.0 [2.1;7.7] [a]	2.3 [1.3;4.1] [a]	
Asthma Control Test ≤19	7 (4.4)	6 (3.2)	21 (21.0)	16 (11.2)	8 (7.8)	23 (28.0)	65 (26.3)	<0.0001
OR [95% CI]	1	0.6 [0.2;2.2]	6.0 [2.3;15.8] [a]	2.8 [1.0;7.7] [a]	2.2 [0.7;6.8]	8.6 [3.3;22.5] [a]	8.1 [3.4;19.4] [a]	
AQLQ total score among current asthma, mean±sd	6.4±0.6	6.4±0.6	5.6±1.1	6.2±0.7	6.3±0.7	5.8±1.1	5.9±0.9	<0.0001
β [95% CI]	0	0.03 [-0.4;0.5]	-0.8 [-1.2;-0.3] [a]	-0.1 [-0.6;0.3]	-0.1 [-0.5;0.4]	-0.7 [-1.1;-0.3] [a]	-0.5 [-0.9;-0.1] [a]	
FEV_1 %pred at ECRHSIII, mean±sd	90.5±16.1	95.0±13.8	91.6±16.6	92.6±12.4	83.0±15.9	84.3±19.7	83.2±18.8	<0.0001
β [95% CI]	0	5.0 [1.2;8.7] [a]	0.5 [-3.6;4.5]	2.5 [-1.4;6.5]	-8.8 [-13.2;-4.4] [a]	-6.6 [-11.2;-1.9] [a]	-7.9 [-11.5;-4.3] [a]	

[a] p value < 0.05

Table 3. Associations between classical asthma phenotypes identified at baseline and different asthma characteristics assessed at follow-up

	Classical asthma phenotypes at baseline					
	Current asthma		Atopy		Age of asthma onset	
Asthma characteristics assessed at follow-up	**No (n=523)**	**Yes (n=781)**	**No (n=385)**	**Yes (n=812)**	**≤16 years (n=585)**	**>16 years (n=692)**
Current asthma, 12m						
Unadjusted-OR	1	3.4 [2.6;4.3] [b]	1	1.7 [1.3;2.2] [b]	1	1.0 [0.8;1.3]
Adjusted-OR[a] [95% CI]	1	3.3 [2.6;4.2] [b]	1	1.8 [1.4;2.3] [b]	1	0.9 [0.8;1.3]
Asthma exacerbations, 12m						
Unadjusted-OR	1	1.6 [1.2;2.3] [b]	1	0.8 [0.6;1.2]	1	0.9 [0.7;1.2]
Adjusted-OR[a] [95% CI]	1	1.7 [1.2;2.4] [b]	1	1.0 [0.7;1.4]	1	0.7 [0.5;1.0]
Asthma Control Test ≤19						
Unadjusted-OR	1	3.0 [2.0;4.6] [b]	1	0.9 [0.6;1.3]	1	1.3 [0.9;1.9]
Adjusted-OR[a] [95% CI]	1	3.2 [2.1;4.9] [b]	1	1.0 [0.7;1.5]	1	1.2 [0.8;1.7]
AQLQ total score among current asthma						
Unadjusted- β	0	-0.1 [-0.3;0.07]	0	0.3 [0.1;0.5] [b]	0	-0.03 [-0.3;-0.01][b]
Adjusted-β[a] [95% CI]	0	-0.1 [-0.3;0.03]	0	0.2 [0.01;0.4] [b]	0	-0.03 [-0.2;0.1]

[a]Models were adjusted on age, sex, smoking status and occupational status at baseline

[b]p value < 0.05

Table 4. Level (in ml) and change (in ml/year) in FEV_1 between cluster-based asthma phenotypes assessed at ECRHSI

	[a] mean level of FEV_1 (ml) at ECRHSI		[ab] mean change in FEV_1 (ml/yr)	[c] Regression estimates for height-standardized FEV_1 decline (/yr)	
	n	**Mean (se)**	**Mean (se)**	**Beta (se)**	**P value**
B _ Non-allergic, few symptoms, no treatment	142	3626.5 (45.4)	-38.0 (1.8)	Ref	--
A _ Allergic, few symptoms, no treatment	150	3738.2 (44.4)	-34.2 (1.6)	1.3 (0.9)	0.13
G _ Non-allergic, moderate symptoms, no treatment	109	3591.8 (52.8)	-33.8 (1.8)	0.9 (1.0)	0.35
F _ Allergic, moderate symptoms, normal LF	140	3629.8 (47.8)	-33.3 (1.7)	1.7 (0.9)	0.06
E _ Allergic, moderate symptoms, BHR	91	3247.4 (62.0)	-35.2 (2.6)	0.8 (1.0)	0.44
C _ Non-allergic, high symptoms, treatment	64	3437.8 (65.3)	-37.3 (3.2)	-0.5 (1.1)	0.62
D _ Allergic, high symptoms, treatment, BHR	168	3090.2 (41.4)	-27.7 (2.1)	3.8 (0.9)	<0.0001

[a]Analyses were adjusted on age, sex, smoking status, occupational status and height at ECRHSI.

[b]Number of subjects was 864 and number of measurements was 2471.

[c]The height-standardized values for FEV_1 correspond to the FEV_1 divided by the square of the height in meters. Number of subject was 958 and number of measurements was 2471. Analyses were adjusted on age, sex, smoking status, and occupational status at ECRHSI.

FIGURE LEGENDS

Figure 1. Adjusted-association between cluster-based asthma phenotypes at baseline and different asthma characteristics (Upper: current asthma, exacerbations and asthma control, Lower: asthma quality of life questionnaire) assessed at follow-up (adjusted on age, sex, smoking status, and occupational status at baseline)

Figure 2. Twenty-year trajectories of lung function parameters (Upper: FEV_1 %pred, Lower: FEV_1/FVC %pred) according to cluster-based asthma phenotypes assessed at ECRHSI (adjusted for age, sex, smoking status, and occupational status at baseline)