Original Article

Cluster Analysis Identifies 3 Phenotypes within Allergic Asthma

María Paz Sendín-Hernández, MD, PhD[a], Carmelo Ávila-Zarza, PhD[b,c], Catalina Sanz, PhD[c,d], Asunción García-Sánchez, PhD[c,e], Elena Marcos-Vadillo, PhD[c], Francisco J. Muñoz-Bellido, MD, PhD[c,f], Elena Laffond, MD, PhD[c,f], Christian Domingo, MD, PhD[g,h], María Isidoro-García, MD, PhD[a,c,i], and Ignacio Dávila, MD, PhD[c,e,f] *Salamanca and Barcelona, Spain*

What is already known about this topic? Unbiased approaches have shown that patients with asthma can be clustered in different phenotypes, some of them allergic. No study has phenotyped asthmatic allergic patients.

What does this article add to our knowledge? We propose for the first time new allergic asthma phenotypes based on a family history of allergy, age of onset, and severity of asthma.

How does this study impact current management guidelines? Careful classification of patients can contribute to understanding mechanisms in allergic asthma, explaining discrepancies in gene association studies, and improving the clinical approach of these patients.

BACKGROUND: Asthma is a heterogeneous chronic disease with different clinical expressions and responses to treatment. In recent years, several unbiased approaches based on clinical, physiological, and molecular features have described several phenotypes of asthma. Some phenotypes are allergic, but little is known about whether these phenotypes can be further subdivided.

OBJECTIVE: We aimed to phenotype patients with allergic asthma using an unbiased approach based on multivariate classification techniques (unsupervised hierarchical cluster analysis).

METHODS: From a total of 54 variables of 225 patients with well-characterized allergic asthma diagnosed following American Thoracic Society (ATS) recommendation, positive skin prick test to aeroallergens, and concordant symptoms, we finally selected 19 variables by multiple correspondence analyses. Then a cluster analysis was performed.

RESULTS: Three groups were identified. Cluster 1 was constituted by patients with intermittent or mild persistent asthma, without family antecedents of atopy, asthma, or rhinitis. This group showed the lowest total IgE levels. Cluster 2 was constituted by patients with mild asthma with a family history of atopy, asthma, or rhinitis. Total IgE levels were intermediate. Cluster 3 included patients with moderate or severe persistent asthma that needed treatment with corticosteroids and long-acting β-agonists. This group showed the highest total IgE levels.

CONCLUSIONS: We identified 3 phenotypes of allergic asthma in our population. Furthermore, we described 2 phenotypes of mild atopic asthma mainly differentiated by a family history of allergy. © 2017 American Academy of Allergy, Asthma & Immunology (J Allergy Clin Immunol Pract 2017;■:■-■)

[a]Department of Clinical Biochemistry, University Hospital of Salamanca, Salamanca, Spain
[b]Statistical Department, University of Salamanca, Salamanca, Spain
[c]Institute for Biomedical Research, IBSAL, Salamanca, Spain
[d]Department of Microbiology and Genetics, University of Salamanca, Salamanca, Spain
[e]Department of Biomedical and Diagnostic Sciences, University of Salamanca, Salamanca, Spain
[f]Department of Allergy, University Hospital of Salamanca, Salamanca, Spain
[g]Pulmonary Service, Corporació Sanitària Parc Taulí, Sabadell, Barcelona, Spain
[h]Department of Medicine, Universitat Autònoma de Barcelona (UAB), Barcelona, Spain
[i]Department of Medicine, University of Salamanca, Salamanca, Spain

This work was supported by the Instituto de Salud Carlos III (ISCIII) (grant PI13/00564 and RETICS ARADyAL RD16/0006/0019) with European Regional Development Fund (ERDF) confounding.

Conflicts of interest: C. Sanz, A. García-Sánchez, E. Marcos-Vadillo, and M. Isidoro-García have received research support from ISCIII. E. Laffond is on the advisory boards for AstraZeneca, ALK Abello, and Allergy Therapeutics. C. Domingo is on the boards for, has received consultancy fees from, and has provided expert testimony from Novartis, Boehringer-Ingelheim, GlaxoSmithKline, and Astra-Zeneca; has received lecture fees from Novartis, Boehringer-Ingelheim, Esteve, Almirall, Chiesi, Menarini, GSK, Takeda, Pfizer, Ferrer, Astra-Zeneca, and Stallergenes. I. Dávila has received research and travel support from ISCIII; has received consultancy fees from AstraZeneca, Sanofi, and Novartis; has received research support from ThermoFisher and Merck; has received lecture fees from AstraZeneca, GlaxoSmithKline, ALK, Stallergenes, and Chiesi. The rest of the authors declare that they have no relevant conflicts of interest.

Received for publication May 10, 2017; revised September 11, 2017; accepted for publication October 5, 2017.

Available online ■■

Corresponding author: María Isidoro-García, MD, PhD, Department of Clinical Biochemistry, University Hospital of Salamanca, Paseo San Vicente, 58-182, Salamanca 37007, Spain. (E-mail: misidoro@usal.es); Or: Catalina Sanz, PhD, Department of Microbiology and Genetics, University of Salamanca, Edificio Departamental, lab 302, Campus Miguel Unamuno, Salamanca 37007, Spain. (E-mail: catsof@usal.es).

2213-2198

https://doi.org/10.1016/j.jaip.2017.10.006

Key words: *Allergic asthma; Cluster analysis; Phenotypes; Inheritance*

Asthma is a chronic inflammatory disease affecting more than 300 millions of people worldwide and causing approximately 250,000 deaths annually.[1] It is considered a heterogeneous disease with different levels of severity, age of onset, type of inflammation, response to treatments, and triggers. This heterogeneity has been recognized in guidelines, such as Global Initiative for Asthma (GINA) (www.ginasthma.org) and Spanish Guideline on the Management of Asthma (GEMA) 4.0.[2] Accordingly, the interest on phenotyping asthma has almost exponentially increased during the last decade.[3] A phenotype can be considered the external manifestation of an individual's underlying genetics on interaction with environment. Asthma was first phenotyped by Rackemann,[4,5] who classified the disease as allergic or extrinsic and nonallergic or intrinsic. Initially, phenotypic classifications of asthma were performed using biased approaches, that is, based on the observation of common characteristics of patients to describe different types of asthma. Thus, these classifications were based on clinical or physiological features, triggers, or the type of inflammation.[6] In recent years, unbiased approaches, that is, based mainly on statistical methods, such as clusters or latent class analyses have been introduced. Although some bias is introduced during the process of variable selection,[7] several important studies have been published,[8-10] followed by others[11,12]; recently, molecular phenotypes have emerged.[13] In 2017 GINA guidelines[14] (www.ginasthma.org), the following phenotypes are considered: allergic asthma, nonallergic asthma, late onset asthma, asthma with fixed airway limitation, and asthma with obesity. Although not all the studies describe the same phenotypes, the allergic phenotype seems to be present in all studies. Nevertheless, the allergic phenotype is a broad phenotype with different degrees of severity.[8] In addition, atopy is present across different phenotypes. In the present study, we aimed to further phenotype allergic asthma.

METHODS

A database was elaborated from data obtained from the clinical history of patients evaluated in real life conditions in the Department of Allergy of the University Hospital of Salamanca (Spain) from 2003 to 2014. The review of clinical histories was developed from 2014 to 2016. After reviewing 700 histories, 225 patients fulfilling the inclusion criteria were selected. Data were obtained from the clinical history and electronic records. All data introduced in the database were anonymized, thus avoiding the possibility of patient identification. The study was approved by the Ethics Committee of Clinical Investigation of the hospital (2014/11/131A). All patients included in our registry were informed that their data would be anonymously used for scientific purposes.

Subjects

All patients had been evaluated by an allergist and had signed an informed written consent to be studied in the outpatient clinic. The inclusion criteria were: (1) age over 14 years; (2) asthma diagnosed following GEMA 2009,[15] including either a positive bronchodilator test or a positive methacholine test; (3) at least 1 positive skin prick test from a locally adapted battery of common aeroallergens performed following European Academy of Allergy and Clinical Immunology recommendations[16] or specific serum IgE; and (4) symptoms and signs of asthma concordant with allergen exposure. The exclusion criteria were: (1) negative skin prick tests; (2) other pulmonary diseases; and (3) patients who did not accept the tests routinely performed at our patient clinic.

Globally, 54 variables were recorded from the clinical history (see Table E1 in this article's Online Repository at www.jaci-inpractice.org). To be properly managed, variables were selected before performing the cluster analysis. Thus, those variables lacking a significant number of individuals (with missing data), being redundant or irrelevant for the purposes of this study, or with low contributions in the solution of multiple correspondence analysis,[17] were rejected (Fig 1). The finally selected variables, on the basis of their considered contribution to the asthma phenotype, were age, sex, body mass index (BMI), family history of atopy (defined as a first-degree antecedent with atopic dermatitis, asthma, or allergic rhinitis), family history of asthma, family history of rhinoconjunctivitis, age of onset of asthma, age of onset of rhinitis, atopic dermatitis, allergic rhinoconjunctivitis, smoking status, lung function, fractional exhaled nitric oxide (FENO), eosinophil peripheral blood count, serum total IgE levels, severity of asthma according to GEMA 2009,[15] use of inhaled steroids, use of long-acting bronchodilators, and use of oral corticosteroids.

Statistical analysis

To perform the analysis of the collected multidimensional information 2 multivariate techniques were performed. First, multiple correspondence analysis,[17] a nonlinear reduction technique for the reduction of dimensionality, was used in 2 ways: on the one hand, to statistically analyze all the variables initially collected in this study and measured in different scales (sociocultural and demographic variables, clinical and functional characteristics of asthma, family history, etc.), to determine among them the most informative variables; and, on the other hand, to obtain a low-dimensional solution (latent numeric variables) on which to perform additional analysis. Secondly, on the numerical solution previously obtained we carried out a cluster analysis (exploratory data analysis), a multivariate technique used to obtain classification systems. The purpose was to sort patients into groups or clusters, so the degree of association/similarity among members of the same cluster was stronger than the degree of association/similarity among members of different clusters. This statistical procedure allows discovering structures in the data that *a priori* are not usually obvious and thus contributing to the formal definition of a phenotype for a set of given objects. All statistical analyses were performed using the SPSS version 22 (IBM SPSS Statistics for Windows, Version 22.0; IBM Corp., Armonk, NY).

RESULTS

The final analysis included 225 subjects who had complete data for the 19 selected variables. The demographics of the entire cohort are presented in Table I. Briefly, 213 patients had both asthma and rhinitis and 12 had only asthma. In the global sample (Table I), there was a predominance of females (57.3%); median age was 39 years and mean BMI was 24.49 kg/m^2. Mean IgE was 362.07 kU/L; mean of peripheral eosinophils was 309.78/μL; and mean fractional nitric oxide was 48.82 ppb.

Multidimensional analysis

From the final data matrix (225 patients by 19 variables), a k-means cluster analysis was performed to identify relatively homogeneous groups of patients from the information provided by the solution of the multiple correspondence analysis,[17] performed