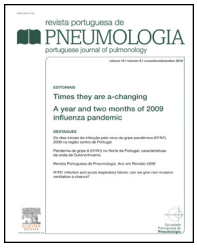
Rev Port Pneumol. 2015;21(6):299–306

revista portuguesa de PNEUMOLOGIA
portuguese journal of pulmonology
www.revportpneumol.org



ORIGINAL ARTICLE

Cluster analysis in phenotyping a Portuguese population

C.C. Loureiro[a,b,*], P. Sa-Couto[c], A. Todo-Bom[d,b], J. Bousquet[e,f]

[a] *Pneumology Unit, Hospitais da Universidade de Coimbra, Centro Hospitalar e Universitário de Coimbra, Coimbra, Portugal*
[b] *Centre of Pneumology, Faculty of Medicine, University of Coimbra, Portugal*
[c] *Center for Research and Development in Mathematics and Applications, Department of Mathematics, University of Aveiro, Aveiro, Portugal*
[d] *Immunoallergology Unit, Hospitais da Universidade de Coimbra, Centro Hospitalar e Universitário de Coimbra, Coimbra, Portugal*
[e] *University Hospital Arnaud de Villeneuve, Montpellier, France*
[f] *Inserm, CESP Centre for Research in Epidemiology and Population Health, U1018, Respiratory and Environmental Epidemiology Team, Villejuif, France*

Received 28 May 2015; accepted 17 July 2015
Available online 3 September 2015

KEYWORDS
Asthma;
Phenotypes;
Cluster analysis

Abstract
Background: Unbiased cluster analysis using clinical parameters has identified asthma phenotypes. Adding inflammatory biomarkers to this analysis provided a better insight into the disease mechanisms. This approach has not yet been applied to asthmatic Portuguese patients.
Aim: To identify phenotypes of asthma using cluster analysis in a Portuguese asthmatic population treated in secondary medical care.
Methods: Consecutive patients with asthma were recruited from the outpatient clinic. Patients were optimally treated according to GINA guidelines and enrolled in the study. Procedures were performed according to a standard evaluation of asthma. Phenotypes were identified by cluster analysis using Ward's clustering method.
Results: Of the 72 patients enrolled, 57 had full data and were included for cluster analysis. Distribution was set in 5 clusters described as follows: cluster (C) 1, early onset mild allergic asthma; C2, moderate allergic asthma, with long evolution, female prevalence and mixed inflammation; C3, allergic brittle asthma in young females with early disease onset and no evidence of inflammation; C4, severe asthma in obese females with late disease onset, highly symptomatic despite low Th2 inflammation; C5, severe asthma with chronic airflow obstruction, late disease onset and eosinophilic inflammation.

Abbreviations: ACT, asthma control test; α-1 AT, α-1 antitrypsin; ALQ, asthma life quality; BMI, body mass index; CARAT, control asthma and allergic rhinitis test; COPD, chronic obstructive pulmonary disease; GERD, gastroesophageal reflux disease; HDAC2, histone deacetylase 2; ICS, inhaled corticosteroids; OCS, oral corticosteroids; OSAS, obstructive sleep apnoea syndrome; SOA, severity of asthma score; SPT, skin prick-tests.

* Corresponding author.
E-mail address: cchloureiro@gmail.com (C.C. Loureiro).

http://dx.doi.org/10.1016/j.rppnen.2015.07.006

Conclusions: In our study population, the identified clusters were mainly coincident with other larger-scale cluster analysis. Variables such as age at disease onset, obesity, lung function, FeNO (Th2 biomarker) and disease severity were important for cluster distinction.

Introduction

Asthma is heterogeneous and cluster approaches have been developed to better define phenotypes.[1–3] Although variable selection and analysis methodology differ, some studies have reached comparable results. Identification of each corresponding endotype requires the definition of their clinical characteristics, biomarkers, lung physiology, genetic aspects, disease history and therapeutic response.[4] According to Wenzel, the first step to better understand asthma is to define phenotypes by cluster analysis.[5]

The objective of this observational study was to cluster an asthmatic Portuguese population, treated in secondary medical care, combining clinical, inflammatory, lung function and severity parameters.

Methods

Study design

Consecutive patients with asthma were recruited from the outpatient clinic, using written informative material. They were included if aged between 18 and 79 years and excluded in the presence of any of the following: cystic fibrosis, interstitial lung disease, auto-immune disease, neoplastic disease, untreated cardiac failure.

Asthma was defined on the basis of a relevant symptom history, plus one or more of the following: history of airway reversibility to salbutamol according to GINA guidelines,[6] positive test for airway hyperresponsiveness using methacholine.[7]

Asthma control was measured under optimized treatment, with validated asthma questionnaires (asthma control test (ACT), control asthma and allergic rhinitis test (CARAT), asthma life quality (ALQ), and severity of asthma score (SOA)).

Severe exacerbations were defined as events requiring urgent action to prevent a serious outcome and at least one of the following: use of systemic corticosteroids or an increase from a stable maintenance dose, for at least 3 days, and/or hospitalization or emergency room visit due to asthma requiring systemic corticosteroid treatment.[8]

The risk of future adverse events was evaluated considering loss of control, number of exacerbations in the previous year, number of systemic corticosteroids in the previous year, accelerated decline in lung function, and side-effects of treatment. A low basal FEV_1 (% predicted value) and low reversibility were also taken into account.

Lung function was assessed with spirometry and plethysmography according to ATS criteria;[9,10] skin prick tests (SPT) were performed using standardized allergens and/or specific serum IgE, FeNO measurement was done using a CLD 88 SP (EcoMedics®) analyzer before any forced expiratory manoeuvers[11] and sputum induction and collection was performed using hypertonic saline 4.5% if stable asthma, delivered via an ultrasonic nebulizer.[12] Induced sputum was collected by a trained nurse, stored on ice and processed within two hours after expectoration. Sputum processing and immunophenotypical analysis of sputum cells was performed according to laboratory procedures.

If the need to exclude other diagnoses occurred, specific procedures were scheduled (Table E1).

Data processing

Variables were selected according to previous studies[1,2] and included demographic data; comorbidities; evaluation of disease control, quality of life and risk assessment; lung function and blood biomarkers (Table 1). Patients with missing data were excluded. Summary statistics were reported as mean and standard deviation values for continuous variables and as percentages and counts for categorical variables. Geometric mean was reported for total serum IgE.

Clinically redundant variables (or with correlation values above 0.9 in module) were reduced. Binary questionnaire data and data with a spectrum of responses were transformed into ''composite variables'' (Table 1) to capture multiple questions in a ranked ordinal scale.[2,13,14]

A cluster analysis of 22 variables (Table 1) was applied to identify groups of patients with the same characteristics. Ward's minimum-variance hierarchical clustering method was performed using an agglomerative approach, and the linkage measure was the squared Euclidean distance with standardization in *z* scores, according to Moore et al.[1] By visual inspection of dendrogram, no single member, small clusters or observations with large distances from all other observations were observed. Therefore, no formal method for outliers' detection was used. Analysis of variance (ANOVA or Kruskal–Wallis) or contingency table tests (Person chi-square or Fisher) were used to compare differences between clusters. All statistical analyses were performed using SPSS® Software, version 20.0 (SPSS, Inc., Chicago, IL), and *p*-values under 0.05 were considered significant.

Results

Of the 72 patients enrolled, 57 had completed data for the considered variables (79.2% of total sample).

The optimum number of clusters was estimated by visual inspection of the dendrogram (Fig. 1) and by representation of the difference between consecutive clusters (Fig. E1). A second cluster method was used (two-step cluster approach) to ensure that Ward's cluster solutions were not