Chronic obstructive pulmonary disease (COPD) is a complex, multidimensional and heterogeneous disease with a large number of subtypes and multifactorial background. It may be caused by different pathophysiologic mechanisms (sometimes referred to as endotypes) but may share similar observed characteristics (phenotypes). These phenotypes divide all patients into several groups with common features which helps patients to receive effective care and achieve better clinical results.

**Objective**

The main purpose of this study is, through the development of statistical methods, to identify clinical phenotypes among adults suffering from COPD. This problem of grouping objects may be solved by cluster analysis. Clustering was applied to understand, manage and better predict future risks and optimize treatment selection based on the new groupings of patients. Furthermore, in this study, missing data and dimension-reduction, which are present in any large dataset of observational data, were handled.

**Cluster Analysis**

In this application, 178 patients were described by 85 multiple and huge sets of variables that structured into seven groups. This study was conducted by the Pneumology Department of the University Hospital of Liège. At the first step, the missing values were imputed by multiple factor analysis (MFA). After single imputation, MFA was applied for reducing the complexity of high-dimensional data. After this, hierarchical clustering was performed using Ward’s criterion on the selected principal components. In this step, the optimal number of clusters was selected based on the hierarchical tree, tool within-cluster sum of square and silhouette method. In the final step, K-means was performed to improve the initial partition obtained from hierarchical clustering. All statistical analyses were performed using R software.

**Statistical Analysis**

Three different phenotypes were defined in COPD. All variables were compared between these clusters by using Kruskal-Wallis and Chi-squared tests for quantitative and qualitative variables, respectively.

**Conclusion**

In the past years, classification methods in COPD have been applied based on ignoring missing values with limited or selected number of variables which have missed more complex phenotypes. In this study, two issues are solved. Then, with advanced statistical methods, patients are divided into three distinct clusters. These clinically meaningful clusters of patients with common characteristics can be used to predict outcomes of patients with COPD, to aid in the development of personalized therapy.