



Distinct clinical phenotypes of airways disease defined by cluster analysis

M. Weatherall*, J. Travers[#], P.M. Shirtcliffe[#], S.E. Marsh[#], M.V. Williams[#],
M.R. Nowitz^{*†}, S. Aldington[#] and R. Beasley^{#,†}

ABSTRACT: Airways disease is currently classified using diagnostic labels such as asthma, chronic bronchitis and emphysema. The current definitions of these classifications may not reflect the phenotypes of airways disease in the community, which may have differing disease processes, clinical features or responses to treatment. The aim of the present study was to use cluster analysis to explore clinical phenotypes in a community population with airways disease.

A random population sample of 25–75-yr-old adults underwent detailed investigation, including a clinical questionnaire, pulmonary function tests, nitric oxide measurements, blood tests and chest computed tomography. Cluster analysis was performed on the subgroup with current respiratory symptoms or obstructive spirometric results.

Subjects with a complete dataset (n=175) were included in the cluster analysis. Five clusters were identified with the following characteristics: cluster 1: severe and markedly variable airflow obstruction with features of atopic asthma, chronic bronchitis and emphysema; cluster 2: features of emphysema alone; cluster 3: atopic asthma with eosinophilic airways inflammation; cluster 4: mild airflow obstruction without other dominant phenotypic features; and cluster 5: chronic bronchitis in nonsmokers.

Five distinct clinical phenotypes of airflow obstruction were identified. If confirmed in other populations, these findings may form the basis of a modified taxonomy for the disorders of airways obstruction.

KEYWORDS: Asthma, chronic bronchitis, chronic obstructive pulmonary disease, cluster analysis, emphysema, phenotypes

Airways disease is currently classified using potentially overlapping diagnostic labels, such as asthma, chronic bronchitis and emphysema. These disorders are characterised by defined symptom complexes, exposure to certain risk factors, variable patterns of airflow obstruction and airway hyperresponsiveness, and different types of airways inflammation with structural changes [1–5]. The definitions of these disorders have evolved over time with greater knowledge of their pathogenesis and clinical characteristics; however, this has led to greater awareness of their imprecision, with the overlap of phenotypes resulting in difficulties in differentiating the disorders from each other. For example, reversible airflow obstruction, the characteristic physiological and clinical feature of asthma, may be present in a substantial proportion of subjects with chronic bronchitis and/or emphysema [6, 7]. In addition, patients with long-standing asthma may show an

accelerated loss of lung function, resulting in irreversible airflow obstruction [8–10]. It is also recognised that different phenotypes of asthma can differ markedly in their response to treatment [11, 12]. Chronic bronchitis is defined by the presence of chronic sputum production, a symptom complex that is commonly present in asthma [13]. Although the use of gas transfer measurements and computed tomography (CT) has enhanced the ability to diagnose emphysema, the definition remains dependent upon pathological appearances [14]. Furthermore, the current definition of chronic obstructive pulmonary disease (COPD) is based on the presence of incompletely reversible airflow obstruction (post-bronchodilator forced expiratory volume in 1 s (FEV₁)/forced vital capacity (FVC) of <0.7), independent of the presence of chronic cough and sputum or the existence of emphysematous structural changes in the lungs [14].

AFFILIATIONS

*University of Otago Wellington,
[#]Medical Research Institute of New Zealand,
[†]Pacific Radiology, Wellington, New Zealand.
^{††}University of Southampton, Southampton, UK.

CORRESPONDENCE

R. Beasley
Medical Research Institute of New Zealand
P.O. Box 10055
Wellington 6143
New Zealand
E-mail: Richard.Beasley@mrinz.ac.nz

Received:

Nov 17 2008

Accepted after revision:

March 19 2009

First published online:

April 08 2009

These limitations have led to recommendations that a new taxonomy is required in order to better define the disorders of airways obstruction [3, 15–18]. Conceptually, the phenotypic overlap of the disorders of airways obstruction has been represented in Venn diagram format [1, 2, 19]. However, this has resulted in ≥ 15 phenotypes that are difficult to differentiate in terms of their pathogenesis, response to treatment and prognosis. Similar limitations exist with the recent proposal to classify asthma into multiple overlapping phenotypes [18].

Another approach is to use techniques such as cluster analysis to describe the patterns of disease based on the clinical, pathogenetic and physiological features of airways disease. This approach has previously been employed in a patient population with well-characterised asthma and COPD, in which at least four phenotypic subgroups of airways disease were identified [3]. It has also been employed in three separate patient populations with asthma, in which several clinically relevant phenotypic subgroups of asthma were identified [12].

The burden of airways disease is not, however, confined to diagnosed patient populations. Many individuals with airways disease in the community do not meet standard criteria for a diagnosis of asthma [20] or COPD [21], or meet the criteria but are undiagnosed [19]. In the present study, a survey of a random community population was undertaken, utilising detailed clinical, immunological, physiological and radiological data. Those with airways disease were identified by the presence of current respiratory symptoms or obstructive spirometric results; a prior diagnosis of asthma, chronic bronchitis or COPD was not required. Cluster analysis was then used to explore clinical phenotypes of airways disease in this community population.

METHODS

Subjects

Participants in the Wellington Respiratory Survey ($n=3,500$) were randomly selected from the electoral register, and equally distributed by sex across the five decade age groups ranging 25–75 yrs [19, 22, 23]. Subjects were sent a simple postal questionnaire seeking demographic, respiratory and smoking history data. The 2,319 subjects who returned completed questionnaires were invited to undertake an interviewer-administered questionnaire followed by pulmonary function tests, exhaled nitric oxide fraction (F_{eNO}) measurements, blood tests, CT scanning and a peak flow diary. The survey was approved by the Central Regional Ethics Committee (Ministry of Health, Wellington, New Zealand), and written informed consent was obtained from each subject.

Detailed questionnaire

All participants completed a detailed written questionnaire, compiled from a series of validated questionnaires and administered by a trained interviewer in a standardised manner, as previously described [22] (see online supplementary material).

Pulmonary function testing

Pulmonary function tests were carried out using two whole-body constant-volume plethysmographs with heated pneumotachographs and gas analysers (Masterlab 4.5 and 4.6; Erich-Jaeger,

Würzburg, Germany), following which peak flow monitoring was undertaken, as described in detail elsewhere [23] (see online supplementary material).

Blood tests

Total serum immunoglobulin (Ig) E levels were measured using the Roche Modular analyser (Roche, Basle, Switzerland). The blood eosinophil level was measured by the Sysmex XE-2100 automated complete blood cell count analyser (Sysmex, Mundelein, IL, USA).

FeNO measurements

Measurements of F_{eNO} were undertaken using an online NO monitor (NIOX[®]; Aerocrine, Solna, Sweden), as previously described [24] (see online supplementary material).

CT scanning

CT scans of the chest were performed using a single machine (GE Prospeed; GE Yokogawa Medical Systems, Tokyo, Japan), as described previously [25] (see online supplementary material).

Statistical methods

Cluster analysis is a technique that defines the distances of each subject from each other based on the combined values, the multidimensional vector, of their measured characteristics. Using a matrix of distance measurements, cluster analysis finds groups of subjects more similar to each other than to those in other groups. It can be used to describe the phenotypes of disease without the need for historical or arbitrary *a priori* assumptions about classification. Subjects were selected for inclusion in the cluster analysis if they had a pre-bronchodilator FEV_1/FVC ratio of $<70\%$ and/or had had wheeze within the last 12 months.

Cluster analysis was carried out on the data set with the subjects defined as described using the following nine variables: 1) pre-bronchodilator FEV_1/FVC ratio expressed as a percentage; 2) pre-bronchodilator FEV_1 expressed as a percentage of the predicted value; 3) post-bronchodilator change in FEV_1 , expressed as percentage from baseline; 4) functional residual capacity expressed as a percentage of the predicted value; 5) diffusing capacity of the lung for carbon monoxide/alveolar volume adjusted for haemoglobin level and expressed as a percentage of the predicted value; 6) natural logarithm of the serum IgE concentration; 7) mean F_{eNO} ; 8) sputum production, defined as a positive response to the question “Do you usually bring up sputum from your chest or have sputum in your chest that is difficult to bring up when you don’t have a cold?”; and 9) cumulative tobacco cigarette consumption (in pack-years).

The variables were chosen to provide measures of airflow obstruction, bronchodilator reversibility, hyperinflation, gas transfer, atopy, eosinophilic airway inflammation, sputum production and lifetime smoking. The variables were equally weighted and the logarithmic function of IgE was used to ensure that extreme values did not form clusters.

For the purposes of the cluster analysis, the Gower distance measure, which is meaningful in the presence of a mixture of continuous and categorical variables, was used [26].

Two different methods were used for the cluster analysis. The primary method was the agglomerative algorithm using the agnes cluster function (R statistical package). The secondary method was the divisive algorithm using the diana function (R statistical package) [26]. Agglomerative methods start with each individual contributing to the cluster analysis being treated as a single cluster and then joins individuals who are closest together on the particular distance metric into clusters. Divisive methods start with all individuals combined into one cluster and then split off individuals and then groups who are furthest apart on the distance metric. Different algorithms use slightly different techniques to efficiently work through the very large number of possible combinations of individuals and clusters at each step of the joining or splitting process. The clusters generated by these techniques were then described using their summary statistics. Tree diagrams or dendrograms were presented to show the progressive divisions or joinings of the clustering process. Cut points in the tree diagrams were chosen to avoid having clusters of less than five subjects, and to aim to have ≥ 10 subjects per cluster. In addition to the nine primary variables, other phenotypic characteristics of the clusters were also presented in summary form.

RESULTS

The 749 subjects who completed the investigative modules showed similar characteristics to the 2,319 who completed the screening questionnaire, as previously reported [22]. Of the 749 subjects, 250 met the criteria for inclusion in the cluster analysis, with 103 (41.2%) having a pre-bronchodilator FEV₁/FVC ratio of <0.7 alone, 86 (34.4%) reporting wheeze within the last 12 months and 61 (24.4%) meeting both criteria. Some subjects were missing data, particularly for FeNO measurements, due to technical problems with the equipment on the day of testing. Exclusion of these subjects with missing data resulted in the cluster analysis being carried out on the 175 subjects with a complete data set. The characteristics of the subjects are shown in table 1. The subjects with data for all variables were similar to those with missing data.

The cluster analysis using the agnes methodology identified five distinct clusters based on the nine variables included in the model (tables 2–4). Cluster 1 grouped subjects with features suggesting an overlap syndrome with atopic asthma, chronic bronchitis and emphysema. This cluster exhibited severe airflow obstruction and the greatest degree of bronchodilator reversibility and peak flow variability. Additional characteristics included concomitant eczema and rhinitis, heavy cigarette smoking and markedly reduced quality of life.

Cluster 2 grouped subjects with features of emphysema but without features of concomitant asthma or chronic bronchitis. Similarly to cluster 1, these subjects showed severe airflow obstruction and were heavy cigarette smokers, but, in contrast, exhibited minimal bronchodilator reversibility and were nonatopic.

Cluster 3 grouped subjects with asthma who were highly atopic and with concomitant eczema and rhinitis and markedly raised FeNO, indicative of eosinophilic airways inflammation. This cluster did not have the features of chronic bronchitis or emphysema.

Cluster 4 grouped subjects with mild airflow obstruction but without features suggesting emphysema or chronic bronchitis, a dominant atopic predisposition or smoking history.

Cluster 5 grouped subjects with airflow obstruction and chronic sputum production, but without concomitant emphysema or a major smoking history. Other features of this group included concomitant eczema and rhinitis and a modest increase in FeNO and blood eosinophil level.

The cluster analysis using the diana methodology identified four disease clusters based on the nine variables included in the model (tables 5–7). Clusters 1, 2 and 4 were very similar to clusters 1, 2 and 5 using the agnes methodology. Cluster 3 represented a combination of agnes clusters 3 and 4 without differentiating on the basis of FeNO or IgE levels.

The dendrograms for the cluster analyses using the agnes and diana methods are shown in figures 1 and 2, respectively.

DISCUSSION

The present cluster analysis suggests that the syndrome of airways obstruction can be classified according to five distinct phenotypes. This provides a naturalistic classification that, if confirmed in other studies, could provide a modified taxonomy for the disorders of airways obstruction.

There are a number of methodological issues relevant to the interpretation of the present findings. The first was the selection of patients from a random population sample, thereby ensuring that the results were broadly generalisable. Subjects who had lung function test results or current respiratory symptoms indicating airflow obstruction were then included in the cluster analysis, thereby ensuring that all forms of airflow obstruction were included. This approach differed from that used previously, in which subjects with a diagnosis of asthma and/or COPD were included and patients

TABLE 1 Characteristics of subjects

| | Complete data set | Incomplete data set [#] | |
|--|-------------------|----------------------------------|--------------|
| | | Subjects n | Data |
| Subjects n | 175 | 75 | |
| Age yrs | 57.4 ± 13.5 | 75 | 56.4 ± 11.8 |
| Pre-BD FEV₁/FVC % | 67.0 ± 9.8 | 75 | 67.0 ± 10.3 |
| Pre-BD FEV₁ % pred | 80.1 ± 18.4 | 75 | 79.7 ± 21.5 |
| Reversibility % | 8.7 ± 8.8 | 75 | 8.4 ± 8.3 |
| DL_{co}/VA[†] % pred | 95.9 ± 15.4 | 75 | 93.4 ± 16.4 |
| FRC % pred | 107.6 ± 23.7 | 65 | 104.2 ± 20.3 |
| Log IgE g·L⁻¹ | 4.3 ± 1.5 | 72 | 4.1 ± 1.7 |
| FeNO ppb | 33.5 ± 26.0 | 5 | 29.8 ± 36.5 |
| Smoking history pack-yrs | 10.8 ± 18.4 | 75 | 10.5 ± 14.8 |

Data are presented as mean ± sd, unless otherwise indicated. BD: bronchodilator; FEV₁: forced expiratory volume in 1 s; FVC: forced vital capacity; % pred: % predicted; DL_{co}: diffusing capacity of the lung for carbon monoxide; VA: alveolar volume; FRC: functional residual capacity; Ig: immunoglobulin; FeNO: exhaled nitric oxide fraction. [#]: data missing for at least one variable; [†]: adjusted for haemoglobin concentration.

TABLE 2 Characteristics of clusters of subjects with full data set[#] defined by the agnes method

| | Cluster [†] | | | | |
|--|----------------------|------------|------------|------------|------------|
| | 1 | 2 | 3 | 4 | 5 |
| Subjects n | 15 | 14 | 30 | 78 | 38 |
| Pre-BD FEV₁/FVC % | 54.7±10.4 | 56.4±9.2 | 65.5±9.1 | 70.1±7.0 | 70.5±8.9 |
| Pre-BD FEV₁ % pred | 58.0±15.5 | 60.2±18.4 | 81.0±19.1 | 87.0±14.1 | 81.1±15.3 |
| Reversibility % | 17.1±15.2 | 6.7±6.7 | 11.7±10.5 | 6.1±5.0 | 8.9±8.6 |
| DL_{CO}/VA[‡] % pred | 75.4±10.7 | 76.5±13.2 | 100.3±10.6 | 100.4±13.3 | 98.5±13.2 |
| FRC % pred | 132.7±28.1 | 120.6±34.4 | 108.2±18.4 | 103.2±20.7 | 101.6±19.2 |
| Log IgE g·L⁻¹ | 5.3±1.4 | 3.1±1.0 | 5.3±1.2 | 4.1±1.4 | 4.0±1.7 |
| FeNO ppb | 33.1±35.6 | 23.7±11.4 | 69.2±26.1 | 22.6±10.5 | 31.6±23.9 |
| Sputum production n (%) | 15 (100) | 0 (0) | 0 (0) | 0 (0) | 38 (100) |
| Smoking history pack-yrs | 36.3±22.9 | 38.1±31.5 | 2.6±7.1 | 5.8±10.6 | 7.4±10.4 |

Data are presented as mean±SD, unless otherwise indicated. BD: bronchodilator; FEV₁: forced expiratory volume in 1 s; FVC: forced vital capacity; % pred: % predicted; DL_{CO}: diffusing capacity of the lung for carbon monoxide; VA: alveolar volume; FRC: functional residual capacity; Ig: immunoglobulin; FeNO: exhaled nitric oxide fraction. #: n=175; †: listed in order of pre-BD FEV₁/FVC; ‡: adjusted for haemoglobin concentration.

who did not meet these strict diagnostic criteria were excluded [3]. It has previously been shown that the inclusion of only those subjects who meet the strict diagnostic criteria for asthma and/or COPD results in the majority of patients with asthma or COPD in the community being excluded from such studies [20, 21]. The cluster analysis was not undertaken in the whole population sample as any signal related to airflow obstruction

would have been overwhelmed by the noise of the far larger nonobstructed group.

Another issue was the selection of the variables included in the cluster analysis. The analysis was limited to nine variables, chosen to reflect putative mechanisms and clinical features of the different phenotypes, and to have the best chance of

TABLE 3 Additional characteristics of subjects included in the five clusters defined by the agnes method

| | Cluster | | | | |
|--|-----------|-----------|-----------|-----------|-----------|
| | 1 | 2 | 3 | 4 | 5 |
| Subjects n | 15 | 14 | 30 | 78 | 38 |
| Male | 11 (73.3) | 12 (85.7) | 19 (63.3) | 46 (59.0) | 21 (55.3) |
| Macroscopic emphysema[#] | 5 (33.3) | 3 (21.4) | 1 (3.3) | 1 (1.3) | 1 (2.6) |
| PEF variability[‡] | 6 (40.0) | 4 (28.6) | 4 (13.3) | 7 (9.0) | 8 (21.1) |
| Eczema | 7 (46.7) | 5 (35.7) | 15 (50.0) | 29 (37.2) | 19 (50.0) |
| Rhinitis | 11 (73.3) | 3 (21.4) | 26 (86.7) | 46 (59.0) | 27 (71.1) |
| Hospitalised[†] ever | 4 (26.7) | 2 (14.3) | 8 (26.7) | 10 (12.8) | 9 (23.7) |
| Hospitalised[†] in last year | 1 (6.7) | 0 (0) | 0 (0) | 1 (1.3) | 4 (10.5) |
| Inhaled β₂-agonist | 8 (53.3) | 4 (28.6) | 13 (43.3) | 22 (28.2) | 13 (34.2) |
| Inhaled corticosteroids | 4 (26.7) | 3 (21.4) | 10 (30.3) | 18 (23.1) | 16 (42.1) |
| Inhaled antimuscarinic | 0 (0) | 1 (17.1) | 0 (0) | 0 (0) | 1 (2.6) |
| Any inhaler | 8 (53.3) | 5 (35.7) | 14 (46.7) | 29 (32.1) | 17 (44.7) |
| Age yrs | 64.7±7.6 | 65.5±8.6 | 52.4±14.6 | 56.1±14.5 | 58.2±11.6 |
| Eosinophils 10⁹ cells·L⁻¹ | 0.27±0.17 | 0.21±0.19 | 0.22±0.09 | 0.18±0.14 | 0.27±0.16 |
| MMEF % pred | 28.6±12.9 | 34.1±13.9 | 58.9±25.1 | 69.6±22.9 | 66.9±24.5 |
| sGaw % pred | 46.6±28.9 | 52.3±24.0 | 61.6±22.9 | 77.0±26.2 | 73.3±29.2 |
| Maximum PEF variability[§] | 22.5±15.0 | 16.8±14.4 | 12.2±8.4 | 10.9±7.0 | 13.9±16.4 |
| SGRQ score | 28.0±18.2 | 21.9±16.8 | 12.6±11.9 | 9.9±11.1 | 23.6±18.3 |

Data are presented as mean±SD or n (%), unless otherwise indicated. PEF: peak expiratory flow; MMEF: maximal mid-expiratory flow; % pred: % predicted; sGaw: specific airway conductance; SGRQ: St George's Respiratory Questionnaire. #: radiological (computed tomography) evidence of definite emphysematous changes; ‡: variability of >20% during 1 week of testing; †: due to respiratory disorder; §: amplitude of mean.

TABLE 4 Patterns of variables of subjects with full data set[#] defined by the agnes method

| | Cluster | | | | |
|--|---------|----|-----|-----|-----|
| | 1 | 2 | 3 | 4 | 5 |
| Pre-BD FEV ₁ /FVC % | - | - | +/- | +/- | +/- |
| Pre-BD FEV ₁ % pred | -- | -- | +/- | +/- | +/- |
| Reversibility % | ++ | -- | + | -- | + |
| DL _{co} /VA [*] % pred | -- | -- | +/- | +/- | +/- |
| FRC % pred | ++ | + | +/- | +/- | +/- |
| Log IgE g·L ⁻¹ | ++ | -- | ++ | +/- | +/- |
| FeNO ppb | +/- | -- | ++ | -- | +/- |
| Sputum production % | 100 | 0 | 0 | 0 | 100 |
| Smoking history pack-yrs | ++ | ++ | -- | -- | -- |

BD: bronchodilator; FEV₁: forced expiratory volume in 1 s; FVC: forced vital capacity; % pred: % predicted; DL_{co}: diffusing capacity of the lung for carbon monoxide; VA: alveolar volume; FRC: functional residual capacity; Ig: immunoglobulin; FeNO: exhaled nitric oxide fraction. --: ≥20% below mean; -: ≥10% and <20% below mean; +/-: within 10% of mean; +: ≥10% and <20% above mean; ++: ≥20% above mean. #: n=175; *: adjusted for haemoglobin concentration.

discriminating between underlying clusters, ensuring that the variables were not too close to measuring the same characteristics. Some, but not all, of the variables chosen have previously been shown to contribute to separate dimensions in the description of asthma and COPD [3, 27, 28]. Variables, such as sex and age, which represented characteristics of the participants rather than expression of disease phenotypes were not used. Smoking history was an exception because of the common clinical practice of allocating smokers with wheeze or airflow limitation to a diagnosis of COPD.

There was no formal means of assessing the optimal number of clusters with the algorithms used in the present analysis. The approach used was to move the cut points in the tree diagrams to give ≥10 subjects per cluster. Examination of the tree plots suggested that this was a reasonable approach.

The cluster analysis using the divisive diana algorithm identified four distinct groups. These were very similar to the five distinct groups identified with the agglomerative agnes algorithm, with the exception that one cluster represented a combination of agnes clusters 3 and 4, without differentiating subjects on the basis of FeNO or IgE levels. Owing to evidence that measures of FeNO are an important phenotypic feature that reflects the degree of airways eosinophilic inflammation and predicts response to inhaled corticosteroid (ICS) therapy [29, 30], we consider that cluster analysis using the agnes algorithm is of greater clinical relevance.

An important consideration, which is inherent to any study of this kind, is that disease-modifying treatment has the potential to modify clinical variables. This relates primarily to ICS therapy, which was prescribed in 21–42% of subjects in each of the groups. It is not possible to overcome this confounding through study design or analysis, since to include only subjects

TABLE 5 Characteristics of clusters of subjects with full data set[#] defined by the diana method

| | Cluster [†] | | | |
|--|----------------------|------------|------------|------------|
| | 1 | 2 | 3 | 4 |
| Subjects n | 11 | 10 | 112 | 42 |
| Pre-BD FEV ₁ /FVC % | 49.8±6.8 | 56.6±10.9 | 68.3±8.1 | 70.3±8.6 |
| Pre-BD FEV ₁ % pred | 52.3±13.5 | 57.2±20.3 | 84.7±15.9 | 80.4±14.8 |
| Reversibility % | 21.0±16.2 | 8.3±7.0 | 7.5±7.3 | 8.7±8.2 |
| DL _{co} /VA [*] % pred | 74.5±12.4 | 74.6±15.4 | 99.7±12.9 | 96.5±14.0 |
| FRC % pred | 145.2±20.2 | 125.7±39.5 | 104.7±19.9 | 101.3±18.6 |
| Log IgE g·L ⁻¹ | 5.6±1.4 | 3.1±1.1 | 4.4±1.4 | 4.1±1.7 |
| FeNO ppb | 37.3±40.2 | 21.0±10.9 | 35.3±26.1 | 30.6±23.3 |
| Sputum production n (%) | 11 (100) | 0 (0) | 0 (0) | 42 (100) |
| Smoking history pack-yrs | 34.8±26.4 | 50.5±28.7 | 5.0±9.7 | 10.6±14.1 |

Data are presented as mean±SD, unless otherwise indicated. BD: bronchodilator; FEV₁: forced expiratory volume in 1 s; FVC: forced vital capacity; % pred: % predicted; DL_{co}: diffusing capacity of the lung for carbon monoxide; VA: alveolar volume; FRC: functional residual capacity; Ig: immunoglobulin; FeNO: exhaled nitric oxide fraction. #: n=175; †: listed in order of pre-BD FEV₁/FVC; ‡: adjusted for haemoglobin concentration.

not prescribed ICS therapy would result in the exclusion of those with the most significant clinical disease, and thereby not be generalisable to a random population with airflow obstruction. However, the observation that the use of ICSs was broadly similar between the groups suggests that it may not have markedly influenced the clusters observed.

Amongst the five distinct phenotypes identified with the agglomerative agnes algorithm, there were two groups with features characteristic of emphysema secondary to tobacco smoking. These two groups exhibited the most severe degree of airflow obstruction and hyperinflation, a marked reduction in gas transfer, a higher prevalence of macroscopic emphysema on CT scanning and similarly heavy exposure to tobacco smoke. However, these two groups were markedly different in terms of other characteristics. One group (cluster 2) had the classical characteristics of emphysema, with no chronic sputum production, minimal bronchodilator reversibility and low levels of serum IgE and FeNO. In contrast, the other group (cluster 1) showed the greatest degree of bronchodilator reversibility and peak flow variability of all five clusters, was strongly atopic in terms of both raised IgE levels and the presence of eczema and rhinitis, and exhibited chronic sputum production. As a result, this group could be considered to represent an overlap disorder with features characteristic of emphysema, atopic asthma and chronic bronchitis.

The other group with chronic sputum production (cluster 5) more closely represented simple chronic bronchitis, without features of emphysema and not associated with heavy tobacco smoking. These subjects commonly exhibited rhinitis and eczema, suggesting a possible role of atopy in the disease pathogenesis.

TABLE 6 Additional characteristics of subjects included in the four clusters defined by the diana method

| | Cluster | | | |
|--|-----------|-----------|-----------|-----------|
| | 1 | 2 | 3 | 4 |
| Subjects n | 11 | 10 | 112 | 42 |
| Male | 8 (72.7) | 9 (90.0) | 68 (60.7) | 24 (57.1) |
| Macroscopic emphysema[#] | 4 (36.4) | 2 (20.0) | 3 (2.7) | 2 (4.8) |
| PEF variability[†] | 5 (45.5) | 4 (40.0) | 11 (9.8) | 9 (21.4) |
| Eczema | 6 (54.6) | 2 (20.0) | 47 (42.0) | 20 (47.6) |
| Rhinitis | 8 (72.7) | 2 (20.0) | 73 (65.2) | 30 (71.4) |
| Hospitalised⁺ ever | 4 (36.4) | 2 (20.0) | 18 (16.1) | 9 (21.4) |
| Hospitalised⁺ in last year | 1 (9.1) | 0 (0) | 1 (0.9) | 4 (9.5) |
| Inhaled β₂-agonist | 7 (63.6) | 3 (30.0) | 36 (32.1) | 14 (33.3) |
| Inhaled corticosteroids | 3 (27.3) | 2 (20.0) | 29 (25.9) | 17 (40.5) |
| Inhaled antimuscarinic | 0 (0) | 1 (10.0) | 0 (0) | 1 (2.4) |
| Any inhaler | 7 (63.6) | 4 (40.0) | 40 (35.7) | 18 (42.9) |
| Age yrs | 65.9±6.7 | 64.0±9.4 | 55.5±14.5 | 58.5±11.4 |
| Eosinophils 10⁹ cells·L⁻¹ | 0.24±0.15 | 0.22±0.21 | 0.19±0.13 | 0.28±0.16 |
| MMEF % pred | 24.0±8.8 | 30.9±13.7 | 65.8±24.0 | 64.4±24.8 |
| sGaw % pred | 35.7±25.0 | 44.9±18.3 | 72.7±26.1 | 73.6±28.0 |
| Maximum PEF variability[‡] | 24.7±15.7 | 18.4±17.0 | 11.4±7.3 | 13.9±15.9 |
| SGRQ score | 30.9±19.1 | 23.7±13.8 | 10.9±11.9 | 23.3±17.8 |

Data are presented as mean±SD or n (%), unless otherwise indicated. PEF: peak expiratory flow; MMEF: maximal mid-expiratory flow; % pred: % predicted; sGaw: specific airway conductance; SGRQ: St George's Respiratory Questionnaire. [#]: radiological (computed tomography) evidence of definite emphysematous changes; [†]: variability of >20% during 1 week of testing; [‡]: due to respiratory disorder; [§]: amplitude of mean.

The other two groups (clusters 3 and 4) could primarily be differentiated on the basis of their FeNO. One group (cluster 3) could be considered to show allergic asthma with eosinophilic airways inflammation, with a markedly raised FeNO (mean 69.2 ppb), high IgE levels and significant bronchodilator reversibility. The other group (cluster 4) showed mild airflow

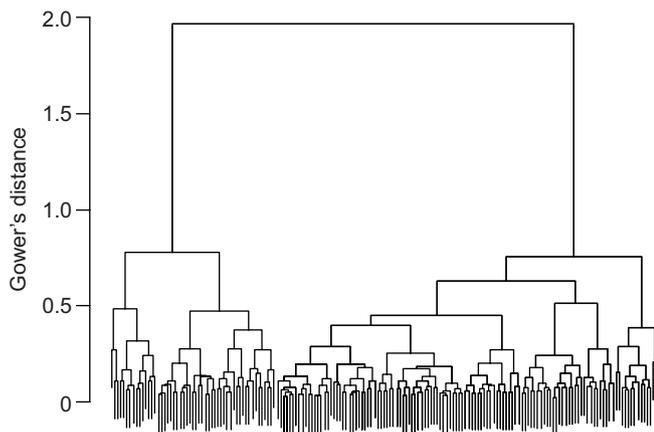


FIGURE 1. Tree plot for the agnes method. Gower's distance refers to the distance at which clusters are joined. The distance separating the individual branches is unitless.

TABLE 7 Patterns of variables of subjects with full data set[#] defined by the diana method

| | Cluster | | | |
|--|---------|-----|-----|-----|
| | 1 | 2 | 3 | 4 |
| Pre-BD FEV₁/FVC % | -- | - | +/- | +/- |
| Pre-BD FEV₁ % pred | -- | -- | +/- | +/- |
| Reversibility % | ++ | +/- | - | +/- |
| DL_{CO}/VA[†] % pred | -- | -- | +/- | +/- |
| FRC % pred | ++ | + | +/- | +/- |
| Log IgE g·L⁻¹ | ++ | -- | +/- | +/- |
| FeNO ppb | + | -- | +/- | +/- |
| Sputum production % | 100 | 0 | 0 | 100 |
| Smoking history pack-yrs | ++ | ++ | -- | +/- |

BD: bronchodilator; FEV₁: forced expiratory volume in 1 s; FVC: forced vital capacity; % pred: % predicted; DL_{CO}: diffusing capacity of the lung for carbon monoxide; VA: alveolar volume; FRC: functional residual capacity; Ig: immunoglobulin; FeNO: exhaled nitric oxide fraction. --: ≥20% below mean; -: ≥10% and <20% below mean; +/-: within 10% of mean; +: ≥10% and <20% above mean; ++: ≥20% above mean. [#]: n=175; [†]: adjusted for haemoglobin concentration.

obstruction with minimal bronchodilator reversibility or peak flow variability, no features to suggest a chronic bronchitis or emphysema phenotype, and a low FeNO. It is possible that these latter subjects represent a group with mild disease and intermittent airflow obstruction.

These findings complement those from the previous cluster analyses of patients with diagnosed asthma or COPD [3]. In particular, they confirm the presence of an overlap atopic asthma/emphysema/chronic bronchitis group with severe airflow obstruction and marked bronchodilator reversibility. This cluster exhibited the most severe disease based on the degree of airflow obstruction, requirement for hospital admission, prescribed treatment and quality of life, and, as a result, represents a particular priority for research into its pathogenesis

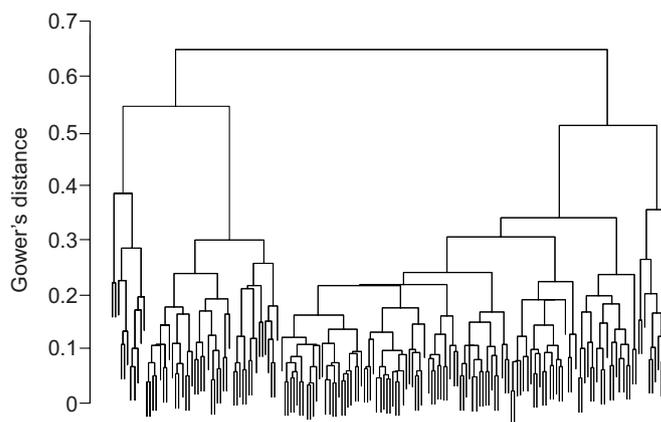


FIGURE 2. Tree plot for the diana method. Gower's distance refers to the distance at which clusters are separated. The distance separating the individual branches is unitless.

and treatment. These findings also support previous findings that there may be more than one asthma group, without emphysema or chronic bronchitis, defined by the presence or absence of atopy and eosinophilic airways inflammation [3, 12]. This finding is of clinical relevance in view of the preferentially greater response to ICS therapy in patients with marked eosinophilic airways inflammation [11, 12, 29, 30].

The present findings are also relevant to the generalisability of the results from randomised controlled trials of asthma and COPD [20, 21]. Subjects in the overlap atopic asthma/emphysema/chronic bronchitis group with severe airflow obstruction would have been excluded from the major randomised controlled trials of asthma due to their smoking history, and from trials of COPD due to their marked bronchodilator reversibility. As a result, there is an inadequate evidence base for the treatment of this group of subjects with the most severe form of airways obstruction.

Application of the current definitions of asthma, chronic bronchitis, emphysema and COPD results in substantial overlap, with ≥ 15 potential phenotypes. This potentially represents a major source of confusion due to the difficulties in distinguishing between these phenotypes in terms of their aetiology, clinical features and responses to treatment. We have demonstrated that there is evidence to consider the syndrome of airflow obstruction in terms of the five distinct phenotypes defined by the major variables included in the present cluster analysis. If these findings are confirmed in other populations, these five phenotypes may form the basis of a modified taxonomy for the disorders of airways obstruction. The priority would then be to determine whether the phenotypes vary in their response to different pharmacological treatments. This knowledge could lead to treatment specifically targeted at defined phenotypic groups, rather than asthma or COPD in general, which represents the current management approach.

STATEMENT OF INTEREST

Statements of interest for R. Beasley and the study itself can be found at www.erj.ersjournals.com/misc/statements.dtl

REFERENCES

- Soriano JB, Davis KJ, Coleman B, *et al.* The proportional Venn diagram of obstructive lung disease: two approximations from the United States and the United Kingdom. *Chest* 2003; 124: 474–481.
- Viegi G, Matteelli G, Angino A, *et al.* The proportional Venn diagram of obstructive lung disease in the Italian general population. *Chest* 2004; 126: 1093–1101.
- Wardlaw AJ, Silverman M, Siva R, *et al.* Multi-dimensional phenotyping: towards a new taxonomy for airway disease. *Clin Exp Allergy* 2005; 35: 1254–1262.
- Hargreave FE, Parameswaran K. Asthma, COPD and bronchitis are just components of airway disease. *Eur Respir J* 2006; 28: 264–267.
- Celli BR, Roger S. Mitchell Lecture. Chronic obstructive pulmonary disease phenotypes and their clinical relevance. *Proc Am Thorac Soc* 2006; 3: 461–465.
- Calverley PM, Burge PS, Spencer S, *et al.* Bronchodilator reversibility testing in chronic obstructive pulmonary disease. *Thorax* 2003; 58: 659–664.
- Anthonisen NR, Wright EC. Bronchodilator response in chronic obstructive pulmonary disease. *Am Rev Respir Dis* 1986; 133: 814–819.
- Silva GE, Sherrill DL, Guerra S, *et al.* Asthma as a risk factor for COPD in a longitudinal study. *Chest* 2004; 126: 59–65.
- Vonk JM, Jongepier H, Panhuysen CI, *et al.* Risk factors associated with the presence of irreversible airflow limitation and reduced transfer coefficient in patients with asthma after 26 years of follow up. *Thorax* 2003; 58: 322–327.
- Peat JK, Woolcock AJ, Cullen K. Rate of decline of lung function in subjects with asthma. *Eur J Respir Dis* 1987; 70: 171–179.
- Berry M, Morgan A, Shaw DE, *et al.* Pathological features and inhaled corticosteroid response of eosinophilic and non-eosinophilic asthma. *Thorax* 2007; 62: 1043–1049.
- Haldar P, Pavord ID, Shaw DE, *et al.* Cluster analysis and clinical asthma phenotypes. *Am J Respir Crit Care Med* 2008; 178: 218–224.
- Mannino DM, Gagnon RC, Petty TL, *et al.* Obstructive lung disease and low lung function in adults in the United States. Data from the National Health and Nutrition Examination Survey, 1988–1994. *Arch Intern Med* 2000; 160: 1683–1689.
- Global Initiative for Chronic Obstructive Lung Disease. Global Strategy for the Diagnosis, Management and Prevention of COPD (Updated 2008). Medical Communications Resources, Inc., 2008. www.goldcopd.org/Guidelineitem.asp?l1=2&l2=1&intId=2003 Date last updated: January 2009. Date last accessed: July 2, 2009.
- Asthma: still more questions than answers. *Lancet* 2008; 372: 1009.
- Anderson GP. Endotyping asthma: new insights into key pathogenic mechanisms in a complex, heterogeneous disease. *Lancet* 2008; 372: 1107–1119.
- A plea to abandon asthma as a disease concept. *Lancet* 2006; 368: 705.
- Wenzel SE. Asthma: defining of the persistent adult phenotypes. *Lancet* 2006; 368: 804–813.
- Marsh SE, Travers J, Weatherall M, *et al.* Proportional classifications of COPD phenotypes. *Thorax* 2008; 63: 761–767.
- Travers J, Marsh S, Williams M, *et al.* External validity of randomised controlled trials in asthma: to whom do the results of the trials apply? *Thorax* 2007; 62: 219–223.
- Travers J, Marsh S, Caldwell B, *et al.* External validity of randomized controlled trials in COPD. *Respir Med* 2007; 101: 1313–1320.
- Shirtcliffe P, Weatherall M, Marsh S, *et al.* COPD prevalence in a random population survey: a matter of definition. *Eur Respir J* 2007; 30: 232–239.
- Marsh S, Aldington S, Williams M, *et al.* Complete reference ranges for pulmonary function tests from a single New Zealand population. *N Z Med J* 2006; 119: U2281.
- Travers J, Marsh S, Aldington S, *et al.* Reference ranges for exhaled nitric oxide derived from a random community survey of adults. *Am J Respir Crit Care Med* 2007; 176: 238–242.
- Marsh S, Aldington S, Williams MV, *et al.* Utility of lung density measurements in the diagnosis of emphysema. *Respir Med* 2007; 101: 1512–1520.
- Everitt B. *An R and S-PLUS® Companion to Multivariate Analysis*. London, Springer-Verlag, 2005.
- Rosi E, Ronchi MC, Grazzini M, *et al.* Sputum analysis, bronchial hyperresponsiveness, and airway function in asthma: results of a factor analysis. *J Allergy Clin Immunol* 1999; 103: 232–237.
- Lapperre TS, Snoeck-Stroband JB, Gosman MM, *et al.* Dissociation of lung function and airway inflammation in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2004; 170: 499–504.
- Szeffler SJ, Phillips BR, Martinez FD, *et al.* Characterization of within-subject responses to fluticasone and montelukast in childhood asthma. *J Allergy Clin Immunol* 2005; 115: 233–242.
- Smith AD, Cowan JO, Brassett KP, *et al.* Exhaled nitric oxide: a predictor of steroid response. *Am J Respir Crit Care Med* 2005; 172: 453–459.