



Published in final edited form as:

Clin Exp Allergy. 2014 May ; 44(5): 724–735. doi:10.1111/cea.12306.

The Diversity of Young Adult Wheeze; a Cluster Analysis in a Longitudinal Birth Cohort

Ramesh J Kurukulaaratchy^{1,2}, Hongmei Zhang³, Abid Raza¹, Veeresh Patil¹, Wilfried Karmaus³, Susan Ewart⁴, and S Hasan Arshad^{1,2}

¹The David Hide Asthma and Allergy Research Centre, St Mary's Hospital, Newport, Isle of Wight, United Kingdom.

²Clinical and Experimental Medicine, Faculty of Medicine, University of Southampton, United Kingdom.

³Department of Epidemiology and Biostatistics, Arnold School of Public Health, University of South Carolina, Columbia, SC, USA.

⁴Large Animal Clinical Sciences, Michigan State University, East Lansing, Michigan, USA.

Abstract

Background—Cluster analyses have enhanced understanding of the heterogeneity of both paediatric and adult wheezing. However while adolescence represents an important transitional phase, the nature of young adult wheeze has yet to be clearly characterised.

Objectives—To use cluster analysis to define, for the first time, clinically relevant young adult wheeze clusters in a longitudinal birth cohort.

Methods—K-Means Cluster analysis was undertaken among 309 currently wheezing subjects at 18-years in the Isle of Wight Birth Cohort (N=1456). Thirteen disease characterising clustering variables at 18-years were used. Resulting clusters were then further characterised by severity indices plus potential risk factors for wheeze development throughout the 1st 18-years of life.

Corresponding Author: Professor S Hasan Arshad MP 810, South Block, Southampton General Hospital, Tremona Road, Southampton, United Kingdom Tel: +44 (0)1983 534373, Fax: +44 (0)1983 822928 sha@soton.ac.uk.

RJK contributed to study design and conduct, conceived the idea for the paper, contributed to data analysis, and wrote the first draft of the manuscript.

HZ contributed to study design, conducted data analysis, and contributed to manuscript preparation.

AR contributed to study conduct, data analysis, and manuscript preparation.

VP contributed to data analysis and manuscript preparation.

WK contributed to study design and manuscript preparation.

SE contributed to study design and manuscript preparation.

SHA contributed to study design, data analysis, manuscript preparation and acts as guarantor for the study. As corresponding author, SHA confirms that he had full access to all the data in the study and had final responsibility for the decision to submit for publication. All authors had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Competing Interest Declaration: “All authors have completed the Unified Competing Interest form at www.icmje.org/coidisclosure.pdf (available on request from the corresponding author) and declare that (1) RJK, HZ, AR, VP, WK, SE and SHA have no support from any company for the submitted work; (2) RJK, HZ, AR, VP, WK, SE and SHA have no relationship to any company that might have an interest in the submitted work in the previous 3 years; (3) their spouses, partners, or children have no specified financial relationships that may be relevant to the submitted work; and (4) RJK, HZ, AR, VP, WK, SE and SHA have no non-financial interests that may be relevant to the submitted work”.

Results—Six wheeze clusters were identified. Cluster 1 (12.3%) male-early-childhood-onset-atopic-wheeze-with-normal-lung function had male predominance, normal spirometry, low BDR (bronchodilator reversibility), intermediate BHR (bronchial hyper-responsiveness), high atopy prevalence, and more admissions. Cluster 2 (24.2%) early-childhood-onset-wheeze-with-intermediate-lung-function had no specific sex association, intermediate spirometry, BDR, BHR, more significant BTS step therapy and admissions. Cluster 3 (9.7%) female-early-childhood-onset-atopic-wheeze-with-impaired-lung-function showed female predominance, high allergic disease comorbidity, more severe BDR and BHR, greatest airflow obstruction, high smoking prevalence, higher symptom severity and admissions. Cluster 4 (19.4%) *female-undiagnosed-wheezers* had adolescent onset non-atopic wheeze, low BDR and BHR, impaired but non-obstructed spirometry, high symptom frequency and highest smoking prevalence. Cluster 5 (24.6%) female-late-childhood-onset-wheeze-with-normal-lung-function showed no specific atopy association, normal spirometry, low BDR, BHR, and symptom severity. Cluster 6 (9.7%) male-late-childhood-onset-atopic-wheeze-with-impaired-lung-function had high atopy and rhinitis prevalence, elevated BDR and BHR, moderately impaired spirometry, high symptom severity and higher BTS step therapy.

Conclusions & Clinical Relevance—Young adult wheeze is diverse and can be classified into distinct clusters. More severe clusters merit attention and are associated with childhood onset, atopy, impaired lung function and in some, smoking. Smoking associated undiagnosed-wheezers also merit recognition. Better understanding of young adult wheeze could facilitate better later adult respiratory health.

Keywords

Asthma; Cluster Analysis; Morbidity; Severity; Smoking; Wheeze

Introduction

Asthma and wheezing illnesses represent a spectrum of common conditions in both paediatric and adult populations, which place considerable burden on healthcare resources [1]. A growing awareness has emerged of the diverse nature of wheezing illnesses in the past two decades [2]. That has been enhanced by cluster analysis techniques to study the multidimensional nature of wheezing illness. Recent studies in paediatric populations have highlighted numerous wheeze clusters characterised by varying severity and associated risk factors [3-7]. More specifically, analysis of the Severe Asthma Research Program (SARP) population [3] identified a wheezing cluster characterised by early-onset atopic status, early-onset more severe childhood asthma and airflow limitation. That study also identified clusters with severe bronchial hyperresponsiveness (BHR) and aggravating comorbidities (sinusitis and gastro-oesophageal reflux) plus other clusters with later onset less atopic childhood asthma. Cluster analyses have identified childhood clusters at risk of asthma exacerbation in both the primary [6] and secondary care [4] populations. An atopic nature for many severe childhood asthma clusters has been demonstrated [3,4,7]. Adult asthma cluster analyses have also identified severe adult atopic disease groups [8-10]. However, adult studies have also shown greater diversity with more non-atopic groups [8,9], comorbidities [6], associations to obesity [6,8,9], greater female predominance [9] and more

significant airflow limitation [9,10]. An overlapping nature of adult asthma, emphysema and chronic bronchitis has also been identified by cluster analysis [10]. Most such analyses have been in cross-sectional study populations with potential for recall or information biases.

Adolescence is an important developmental phase associated with dynamic changes that lead to peak physiological status, and consequently may provide a window of vulnerability for disease expression. While the diverse nature of both paediatric and adult wheeze is increasingly well defined, understanding of the nature of adolescent/young adult wheeze remains largely uncharacterised territory. To date, cluster analysis of wheezing illnesses has not scrutinised this specific age group. Recent analysis of the Isle of Wight whole population birth cohort at 18-years showed a substantial prevalence of wheezing illness (22%) at that age, 80% of which was associated with an asthma diagnosis [11]. Wheezing subjects who did not receive an asthma diagnosis showed scanty phenotypic features of atopic asthma with low levels of atopy, bronchodilator reversibility (BDR), and bronchial hyperresponsiveness (BHR) [11]. With high levels of personal tobacco smoking in adolescence, such *undiagnosed-wheezers* may represent a distinct phenotype. Following such work, we hypothesised that young adult wheeze is likely to encompass a collection of distinct, clinically-relevant wheezing phenotypes that could be identified by adopting a cluster analysis technique using relevant clinical characteristics at 18-years. Furthermore, by undertaking such analysis in a longitudinal birth cohort we postulated that having identified young adult wheezing clusters, assessment of prospectively-recorded potential risk factors could further define the nature of those clusters. In this paper we describe a cluster analysis of wheezing illnesses in the Isle of Wight cohort at 18-years.

Methods

An unselected whole population birth cohort (n=1456) was established on the Isle of Wight (UK) in 1989 to study the natural history of asthma. Participants were assessed at 1-, 2-, 4-, 10- and 18-years. Methodology for the first decade of follow-up has been published previously [11-14].

The Local Research Ethics Committee (06/Q1701/34) approved follow-up at 18-years. Participants gave informed consent and provided information on respiratory, nasal and dermatological symptoms. Study-specific plus International Study of Asthma and Allergies in Childhood (ISAAC) [15] questionnaires were used. Questions about specific exposures are provided in the online supplement.

Participation at age 18 was in person, by telephone or by post. Participants attending in person also performed spirometry, fractional exhaled nitric oxide (FeNO) measurement, methacholine challenge test and skin prick test (SPT). Blood was also taken for total IgE measurement. Identical methodology, published previously [14], was used for spirometry and challenge testing at 10- and 18-years. FeNO (Nioxmino,®Aerocrine AB, Solna, Sweden) and SPT to common allergens (ALK-Abello, Horsholm, Denmark) were performed as reported previously [16]. Relevant methodology at 10 and 18-years is summarised in the online supplement.

Definitions

Asthma at 10 and 18-years was defined as affirmative responses to the questions “Have you ever had asthma?” and either of “Have you had wheezing in the last 12 months?” or “Have you had asthma treatment in the last 12 months?” Rhinitis at 10 and 18 was defined as affirmative responses to “Have you ever had a problem with sneezing, runny or blocked nose in the absence of cold or flu?” plus “Have you had symptoms in the last 12 months?” Eczema at 10 and 18 was defined by affirmative responses to “Have you ever been diagnosed with eczema?” plus “Have you had an itchy rash in the past 12 months?” Diagnostic definitions used at early life follow-ups are given in the online supplement. Atopy was defined by positive SPT (mean wheal diameter 3mm negative control) to at least one allergen.

Bronchodilator reversibility was defined as percentage change in FEV₁ after inhaling 600 micrograms salbutamol. At both 10 and 18-years, a continuous dose-response slope (DRS) measure of BHR was estimated by least-square regression of percentage change in FEV₁ upon cumulative methacholine dose for each child. The DRS was transformed as Log₁₀(DRS+10), to satisfy normality and homoscedasticity, with higher values inferring greater BHR. We used a continuous dose-response measure of BHR since not all subjects who undergo bronchial provocation testing will demonstrate a 20% fall in FEV₁ that enables calculation of a PC20 to indicate BHR. Reliance on PC20 would have meant that a proportion of subjects would not provide meaningful data on BHR.

Statistical methods

All statistical analyses were performed using the SAS statistical package version 9.2 (SAS Institute, Cary, NC, USA).

Cluster analysis was performed on the population that either had wheezing at age 18 or who had taken asthma medication in the past 12 months but may not have experienced wheezing (n=309). Cluster variables were selected as being defining clinical characteristics at 18-years. These included both questionnaire-derived data and variables from objective testing. In total, 13 variables were selected for the cluster analyses: atopic status at 18, rhinitis at 18, eczema at 18, age that wheeze appeared, total IgE (log 10), BDR, BHR DRS, mean FeNO (log10), FEV₁, FVC (Forced Vital Capacity), FEV₁/FVC, FEF₂₅₋₇₅ (Forced Expiratory Flow 25-75%), and British Thoracic Society (BTS) Step of asthma management [17]. As in linear regressions, clustering methods are sensitive to the scale of the variables. To put every continuous variable on a common scale, each variable was standardized by subtracting its mean and then divided by its standard deviation [18]. To remove gender and height effect on FEV₁, FVC, and FEF₂₅₋₇₅, we regressed these variables on gender and height; the residuals with gender and height effect excluded were used in the analyses. We used this approach in preference to using % predicted lung function values. With % predicted values, people with the same gender and height would register the same value of % predicted estimated from existing formulas based on reference populations. Those reference ranges may not be applicable to our population and would not necessarily reflect the measure of lung function for each individual participant in our cohort. Due to the skewness of Total IgE and FeNO measures in the original scale, those variables were log₁₀-transformed to improve normality

and homoscedasticity. We calculated the correlations among the continuous variables (Table E1), between continuous and binary variables (Table E2), and between binary variables (Table E3) which are provided in the online supplement. Most correlations were smaller than 0.51, a moderate correlation as defined by Cohen. The method of K-means (PROC FASTCLUS in SAS) was used to perform the cluster analysis. To determine the number of clusters, we used two criteria: the cubic clustering criterion (CCC) [19] and the pseudo F statistic [20]. The CCC criterion is a measure of the deviation of the clusters from the distribution expected if data points were drawn from a uniform distribution. The pseudo F statistic is intended to capture the homogeneity of the clusters. It is a ratio of the mean sum of squares between groups to the mean sum of squares within group. Larger values of CCC and the pseudo F indicate a better cluster solution. In our analysis, we considered different numbers of clusters, and then for each case obtained the values of CCC and the pseudo F. In general, the patterns of CCC and the pseudo F were quadratic with respect to the number of clusters. The final choice of the number of clusters was determined as a balance between large CCC and the pseudo F and simple clusters (Table E4; online supplement).

To evaluate the cluster profiles, for each clustering variable, across the clusters, ANOVA was used for continuous variables and Chi-Square tests for binary variables. Pair-wised t-tests with Bonferonni multiple testing correction were applied to continuous variables, and pairwised proportion tests were applied to binary variables with the overall significance level set at $\alpha=0.05$. Having defined distinct clusters via cluster analysis, we sought to further characterize those clusters by assessing morbidity variables including symptom frequency and severity plus healthcare utilization; A&E (Accident & Emergency) department attendance and hospital admission. To compare each morbidity variable between different clusters, for categorical variables, chi-square tests were implemented and for continuous variables Kruskal-Wallis tests were used due to the violation of the normality assumption in the residuals. Potential associations with risk factors recorded prospectively during the lifetime of the cohort were then assessed. These included male sex, family history (parent or sibling) of asthma, cord blood IgE, low (<2.5kg) birth weight, maternal smoking in pregnancy, exclusive breast feeding in the first 3 months of life, recurrent chest infections in infancy, past or current personal smoking, paracetamol use at 18-years, and Body Mass Index (BMI) at 18-years. The same methods as those for morbidity variables were used to compare risk factors between different clusters. The significance level was set at 0.05. The Bonferroni method was used to correct for multiple testing.

Results

High (90%; n= 1313) cohort follow-up was achieved at 18-years. Of the overall cohort, 16.8% (n=210) were seen at 10 but not 18-years and 6.4% (n=80) were seen at 18-years but not at 10. Of subjects seen for a Centre Visit at 18-years (n=864), 90.5% (n=762) were also seen at 10-years. Previously published data [16] (see Supplementary Table E5) demonstrated that participants attending the centre for a “full visit” (n=864) at 18-years did not differ significantly from the overall cohort participation. At 18-years, whole population prevalence of diagnosed asthma was 17.9% (234/1306), rhinitis was 35.8% (468/1309), eczema was 12.3% (161/1306) and atopy was 41.3% (352/853).

Clustering Outcome

Cluster analysis was conducted on a population of 309 subjects who either had wheezing at age 18 or who had taken asthma medication in the previous 12 months but may not have experienced wheezing. Of the 309 cluster population, 68.9% (n=199) were seen at both 10 and 18-years. Among the cluster population, diagnosed asthma was present in 76.0% (n=234). Of subjects with diagnosed asthma with available information, 98.2% (n=221) had ever received asthma treatment. Statistical summary of standardized variables used for cluster analysis is in the online supplement (Table E6). Cluster analysis using the 13 selected variables identified 6 discrete wheezing clusters. The proportion of these clusters within the wheeze population varied at 18-years; cluster 1 (12.3%), cluster 2 (24.2%), cluster 3 (9.7%), cluster 4 (19.4%), cluster 5 (24.6%) and cluster 6 (9.7%).

Cluster Variable Characteristics for Wheezing Clusters

Each cluster variable showed a significant difference between clusters as indicated by the comparisons between clusters using ANOVA for continuous variables or Chi-Squared tests for binary variables (Table 1). Separation of clusters by cluster variables is outlined in Table 1 (last column) and resulting cluster variable based characterization is summarised below. Age of onset was defined as “early childhood” (<6-years), “late childhood” (6-12-years) and adolescent (>12-years). Distribution of these “age of onset categories” across the 6 wheeze clusters is shown in Figure 1. Prevalence of clusters across BTS steps of therapy is shown in Table 2. No subjects were treated at BTS step 5.

Cluster 1; Early-childhood-onset-atopic-wheeze-with-normal-lung-function (12.3%) had early childhood onset disease. At 18-years they had normal spirometry, low BDR, intermediate BHR and high atopy prevalence. Half of these subjects required asthma drug treatment at 18-years, most of which was at BTS steps 1 and 2. However, they also constituted 19.0% of subjects on BTS steps 3 or 4 therapy at 18.

Cluster 2; Early-childhood-onset-wheeze-with-intermediate-lung-function (24.2%) had early childhood onset disease. At 18-years they had intermediate spirometry, BDR, and BHR. There was no specific atopy association. The majority (57.3%) were taking asthma therapy at 18-years, mostly at BTS steps 1 and 2. However, they accounted for 38.1% of subjects on BTS steps 3 or 4 therapies at 18.

Cluster 3; Early-childhood-onset-atopic-wheeze-with-impaired-lung-function (9.7%) had youngest childhood onset disease. They showed high allergic disease comorbidity, atopy prevalence, more severe BDR and BHR, plus greatest airflow obstruction at 18-years. This group had the highest prevalence of asthma therapy at 18-years (70.0%), but most of this was at BTS steps 1 and 2 and this cluster constituted only 9.5% of those on BTS steps 3 or 4.

Cluster 4; Undiagnosed-wheezers (19.4%) had adolescent onset non-atopic wheeze, low allergic comorbidity, low BDR and BHR plus impaired but non-obstructed spirometry at 18. Just under half (46.7%) were on asthma therapy at 18-years, all of which was at BTS steps 1 or 2.

Cluster 5; Late-childhood-onset-wheeze-with-normal-lung-function (24.6%) had late childhood onset disease. They showed no specific atopy association, low allergic comorbidity, normal spirometry, plus low BDR and BHR at 18. This group had the lowest prevalence of asthma therapy at 18-years (34.2%), most of which was at BTS steps 1 or 2. This group accounted for 4.8% of BTS step 3 or 4 therapy at 18.

Cluster 6; Late-childhood-onset-atopic-wheeze-with-impaired-lung-function (9.7%) had late childhood onset disease. They had high atopy and rhinitis prevalence with high FeNO, elevated BDR and BHR, plus moderately impaired spirometry at 18. Asthma therapy was prescribed for 60.0% of this cluster at 18-years, mostly at BTS steps 1 or 2. However, this group also constituted 28.6% of subjects needing BTS steps 3 or 4 therapy at 18-years.

Morbidity Measures for 18-year Clusters

Measures of symptom severity are illustrated in Table 3. Clusters 1 (early-childhood-onset-atopic-wheeze-with-normal-lung-function) and 5 (late-childhood-onset-wheeze-with-normal-lung-function) showed lowest expression of exercise-induced, speech-limiting or sleep-disturbing symptoms. Cluster 3 (early-childhood-onset-atopic-wheeze-with-impaired-lung-function) had high reporting of exercise-induced and sleep-disturbing symptoms, cluster 6 (late-childhood-onset-atopic-wheeze-with-impaired-lung-function) had high prevalence of sleep-disturbing symptoms and cluster 4 (*undiagnosed-wheezers*) had high prevalence of sleep-disturbing and speech-limiting symptoms. Wheezing symptom frequency at 18-years for the clusters is shown in Table 3. This was lowest for cluster 5 (late-childhood-onset-wheeze-with-normal-lung-function). Cluster 6 (late-childhood-onset-atopic-wheeze-with-impaired-lung-function) showed highest prevalence of moderate symptom frequency, while cluster 4 (*undiagnosed-wheezers*) showed highest prevalence of high symptom frequency. Healthcare attendance for wheezing is shown in Table 3. Cluster 4 (*undiagnosed-wheeze*) showed lowest A&E/ hospital admission levels. Cluster 6 (late-childhood-onset-atopicwheeze-with-impaired-lung-function) showed high A&E attendances but low admission rates. Clusters 1(early-childhood-onset-atopic-wheeze-with-normal-lung-function), 2 (*early-childhood-onset-wheeze-with-intermediate-lung-function*) and 3 (early-childhood-onset-atopic-wheeze-with-impaired-lung-function) showed highest A&E and hospital admissions.

Risk Factor Associations for Wheeze Clusters

Ten potential factors were assessed for association with the 6 wheezing clusters (Table 4). Male sex and personal history of smoking differed significantly between the clusters. Male sex was highest in cluster 6 (late-childhood-onset-atopic-wheeze-with-impaired-lung-function; 73.3%) and lowest in cluster 3 (early-childhood-onset-atopic-wheeze-with-impaired-lung-function; 26.7%). Cluster 1 (early-childhood-onset-atopic-wheeze-with-normal-lung-function) also showed male predominance. Other clusters showing female predominance were clusters 4 (*undiagnosed-wheeze*) and 5 (late-childhood-onset-wheeze-with-normal-lung-function). Personal history of smoking was highest in cluster 4 (*undiagnosed-wheezers*; 74.6%) and lowest in cluster 5(late-childhood-onset-wheeze-with-normal-lung-function; 49.3%). Maternal smoking in pregnancy, cord IgE at birth, family history of asthma, low birth weight, exclusive breastfeeding, recurrent chest infections in

infancy, paracetamol use at 18-years and BMI at 18-years did not differ significantly between clusters (Table 4).

Final Summary Cluster Characterisation

A final summary of the clusters incorporating cluster characteristics, morbidity measures and risk factors is shown in Figure 2. Clusters 1 (*male-early-childhood-onset-atopic-wheeze-with-normal-lung-function*), 2 (*early-childhood-onset-wheeze-with-intermediate-lung-function*), 3 (*female-early-childhood-onset-atopic-wheeze-with-impaired-lung-function*), and 6 (*male-late-childhood-onset-atopic-wheeze-with-impaired-lung-function*) all showed high prevalence of diagnosed asthma at 18-years (Figure 2). Cluster 4 (*female-undiagnosed-wheeze*) showed lowest prevalence of diagnosed asthma.

Discussion

Cluster analysis demonstrated a heterogeneous nature to wheezing illness in young adulthood. Using 13 parameters representing core disease characteristics, our study identified 6 wheeze clusters at 18-years with differing associations to age of wheezing onset, atopy, allergic comorbidity, lung function, BDR, BHR and level of asthma therapy. Further characterisation of those clusters also indicated differing associations with healthcare utilisation, symptom severity and frequency. Potential risk factors were identified including sex, and smoking habit. Two higher severity clusters were identified; female-early-childhood-onset-atopic-wheeze-with-impaired-lung-function and male-late-childhood-onset-atopic-wheeze-with-impaired-lung-function. These both showed high prevalence of atopy and allergic comorbidity, plus impaired lung function, high BDR and high BHR at 18-years. High smoking prevalence was seen in female-early-childhood-onset-atopic-wheeze-with-impaired-lung-function. In addition 20% of wheezing subjects (*female-undiagnosed-wheezers*) showed high smoking status but low associations with atopy, BHR, BDR or airflow obstruction, suggesting a nature distinct from asthma.

Cluster analysis potentially offers an assessment of the diversity of wheezing illness that is less biased by preconceived assumptions. It has successfully described the nature of wheezing in adults [8-10,21] and children [3-7,22-23]. To the best of our knowledge this is the first study using this technique to study wheezing during the transition to early adulthood in a longitudinal whole population birth cohort. The use of longitudinally-acquired data enables reliable assessment of potential risk factors for these wheezing clusters, particularly in a study with high long-term follow-up such as ours. Another strength is that rather than the high severity cohorts studied by other groups [3-5, 8-9] our study was derived from a whole population cohort. That increases likelihood of detecting clusters reflecting the true nature of wheeze, covering all grades of severity, rather than those relevant only to severe disease clinics. Major adult cluster analyses have excluded smokers [8-9]. A further strength of using an unselected birth cohort was the inclusion of smokers; (28% of our 18-year population currently smoked), permitting assessment of smoking in wheezing clusters. One potential criticism of our study might be the choice of potentially correlated cluster variables such as lung function, BHR and BDR. Since FEV₁/FVC is not linearly associated with FEV₁ and FVC individually we did not expect strong correlations; that was confirmed by

evaluation (correlation between FEV₁/FVC and FEV₁ is 0.45 and between FEV₁/FVC and FVC is -0.32; see online supplement). Potentially higher correlations occurred between FEV₁ and FVC (correlation=0.68) and between FEV₁ and FEF₂₅₋₇₅ (correlation=0.75), so when performing cluster analyses, subjects with similar levels of FEV₁ and FVC were more likely to be in the same cluster. Similar expectation is applied to FEV₁ and FEF₂₅₋₇₅. However, we do not expect this influenced the overall cluster pattern compared to the pattern from cluster analyses by taking out FEV₁ (since this is the variable most highly correlated with FVC and FEF₂₅₋₇₅) because we included many other variables that are not highly correlated and will influence the cluster patterns. If all the variables were highly correlated, then we would not get clear cluster patterns, but fortunately this is not the case in our data. While we cannot exclude potential biases as a result of collinearity, our tests of correlation did not identify very strong correlations between these variables, suggesting that the extent of such effects were limited. Another possible criticism of our study is the wider applicability of our findings, given a unique island environment. Nevertheless, our island population is genetically similar to mainland England (data on file) and previous findings [12-14] have been similar to other international cohorts. Our present findings require replication in other birth cohorts and among larger samples.

Wheeze cluster studies have shown variable associations to sex. Paediatric and adult clusters have shown female [5,8,9], male [5,8] or no sex [3,6,9] predominance. We identified 3 female predominant and 2 male predominant clusters with both female and male predominant higher severity clusters (3 and 6, respectively). Prior reports have associated disease severity with early-onset atopic clusters [3,4,7-9]. Female-early-childhood-onset-atopic-wheeze-with-impaired-lung-function were consistent with that notion (mean age of onset 4.75-years). However, emphasising the diversity of young adult wheeze clusters, male-late-childhood-onset-atopic-wheeze-with-impaired-lung-function demonstrated significant disease severity but a much later age of onset (8.12-years). That is contrary to expectations that impaired lung function in early adulthood might be the preserve of clusters with earlier disease onset. The identification of a severe female early-onset cluster and a severe male later-onset cluster is also contrary to expectations that more severe early-onset disease might be male predominant and later onset severe disease more female predominant [24]. It does though highlight the presence of impaired young adult lung function in more severe wheeze clusters and invites speculation as to what the later adult outcome of such states will be.

In common with adult cluster analyses on samples drawn from general [8,10] or primary care [8] populations, our analysis also identified a milder disease cluster (female-late-childhood-onset-wheeze-with-normal-lung-function). No single cluster demonstrated higher wheezing morbidity across all parameters. Three clusters (male-early-childhood-onset-atopic-wheeze-with-normal-lung-function, early-childhood-onset-wheeze-with-intermediate-lung-function and female-early-childhood-onset-atopic-wheeze-with-impaired-lung-function) accounted for most wheezing-related acute healthcare utilisation. These three clusters also had highest diagnosed asthma prevalence at 18-years. Clusters showed varying discordance between chronic symptom expression and exacerbation risk. Female-early-childhood-onset-atopicwheeze-with-impaired-lung-function showed higher symptom severity and hospital admissions. However, male-early-childhood-onset-atopic-wheeze-with-normal-lung-function showed low symptom severity but high admission rates while

male-late-childhood-onset-atopic-wheeze-with-impaired-lung-function showed high symptom severity and frequency but lower admissions. This extends previous findings of discordance between wheezing symptoms and exacerbation risk in paediatric [3] and adult [6] cluster analyses. The paediatric SARP wheeze clusters [3] demonstrated equally high exacerbation risks regardless of background severity. Our findings may reflect the varying discordance between symptoms and underlying pathophysiology shown in adult clusters. Adult clusters with disproportionately high symptoms for degree of airflow obstruction [9] or airway inflammation [8] plus those with high airway inflammation but low symptom expression have been identified [8]. It is plausible that treatment may have differentially influenced these findings in some clusters. For instance the higher treatment levels of male-late-childhood-onset-atopic-wheeze-with-impaired-lung-function may have reduced severe exacerbations but impacted less on chronic symptoms in that cluster.

Paediatric cluster studies have identified association of atopic status with higher severity childhood wheezing states [3,4]. Atopy has been associated with persistent patterns of childhood wheeze [22-23] while multiple early allergic sensitisation may be a significant risk factor for childhood asthma, impaired spirometry, high BHR and elevated asthma admission risk [7]. Adult studies too have identified atopic clusters with impaired lung function [9,10], greater BHR [8,9] or BDR [10], higher medication needs [9] and greater acute healthcare utilisation [9]. Two of our young adult wheezing clusters mirrored these findings (female-early-childhood-onset-atopic-wheeze-with-impaired-lung-function and male-late-childhood-onset-atopic-wheeze-with-impaired-lung-function) in association with highly atopic status. However, another highly atopic cluster (male-early-childhood-onset-atopic-wheeze-with-normal-lung-function) showed intermediate severity wheeze. This heterogeneity of atopic wheeze adds to previous demonstration of milder atopic wheeze clusters in younger [4] and older [9, 21] populations. Adult cluster studies have identified more low-atopic wheeze clusters than paediatric ones, though a severe low-atopy cluster has been shown in later childhood [4]. Low-atopic adult wheeze clusters with high [8,9] and low [21] morbidity have been characterised. We found that 19% of young adult wheeze fell within an adolescent-onset low-atopic wheeze cluster. This was predominantly female, consistent with prior description of later adult low-atopy wheeze clusters [8,9]. However, those adult female low-atopy clusters differed from ours, showing association with obesity, high medication use and acute healthcare utilisation. Our population, by comparison, may have been too young to show those specific features.

We labelled our main low-atopy wheeze cluster *undiagnosed-wheezers* as they had lowest proportion of diagnosed asthma. At 18, with low proportions of BHR or BDR, normal FeNO, and no airflow obstruction they showed scant phenotypic features of asthma though they reported wheezing and had mildly impaired lung function. They also showed lowest family history of asthma suggesting low genetic predisposition to asthma. Irritant-induced bronchitis may be one explanation for later-onset adolescent wheeze without BHR or BDR. We recently characterised a similar undiagnosed-wheeze phenotype in this population by simply assessing 18-year olds who wheezed without an asthma diagnosis [11]. That approach may be open to predetermined bias, but our present cluster analysis using an unbiased approach suggests that this phenotype is a valid entity. That recognition of this

state is poor in clinical practice is suggested by the fact that they showed higher symptom frequency and severity than many clusters but received little pharmacotherapy judged by BTS step of therapy and had lowest levels of asthma diagnosis. Recognition in the literature is also poor but evidence for early abnormal airway pathophysiology in adolescent smokers is emerging [25]. Early adult lowatopy wheeze associated with smoking has been previously reported in other studies [26-28] though generally assumed to represent undiagnosed asthma. Our study suggests that this group is distinct from mainstream asthma.

Few wheeze cluster studies have incorporated impact of personal tobacco smoking. In addition to *female-undiagnosed-wheezers*, our study identified high smoking prevalence in female-early-childhood-onset-atopic-wheeze-with-impaired-lung-function who showed many hallmarks of more severe asthma. Recent work demonstrated negative impact of adolescent smoking on symptoms and lung function in adolescent asthmatics, though more pronounced in non-atopic subjects [29]. Adolescent smoking has high familial association [30]. In such clusters, there may be inherent susceptibility to harmful effects of both passive and later active tobacco smoking in relation to their airways disease. In this regard early [31] and late [32] passive smoke exposure may interact with 17q12-21 gene variants in childhood asthma risk. Adolescence is a period of dynamic change that should result in peak physiological status. Young adult asthmatics show poorer lung function [33] while smoking asthmatics may show worse lung function [29]. Poor young adult lung function is a risk factor for impaired lung function in middle age [34]. Smoking and asthma have additive effects on lung function decline in the lead up to middle age,[33] which may predispose to high morbidity overlapping smoking related clusters [10]. Smoking young adult wheezers may be at increased risk of adult COPD and smoking cessation efforts in such individuals should be paramount.

In conclusion, this cluster analysis in a longitudinal birth cohort provides clinically relevant insight into the diversity of young adult wheezing. High morbidity clusters with impaired lung function merit early recognition and those who smoke should be a focus for behavioural intervention; so too should other “less asthmatic” young adult smoking wheezers. An improved understanding of the pathophysiological characteristics and future outcome of these clusters could facilitate better clinical management of young adult wheeze. These clusters should be subject to further assessment to both investigate mechanisms predisposing to their development and their relevance to later respiratory health.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The 18-year follow-up of this study was funded by the National Institutes of Health USA (Grant5 R01 HL082925) and National Institute of Allergy and Infectious Diseases (award number R01 AI091905). The authors gratefully acknowledge the cooperation of the children and parents who have participated in this study. We also thank Brian Yuen, Professor Graham Roberts, Sharon Matthews, Roger Twiselton, Paula Williams, Monica Fenn, Linda Terry, Stephen Potter and Rosemary Lisseter for their considerable assistance with many aspects of the 18-year follow-up of this study. Finally we would like to highlight the role of the late Dr David Hide in starting this study.

Funding: The National Institutes of Health USA (Grant5 R01 HL082925), National Institute of Allergy and Infectious Diseases (award number R01 AI091905), and National Institute of Allergy and Infectious Diseases (award number R21 AI099367). Contributor Statement

References

1. Braman SS. The global burden of asthma. *Chest*. 2006; 130(1 Suppl):4S–12S. [PubMed: 16840363]
2. Borish L, Culp JA. Asthma: a syndrome composed of heterogeneous diseases. *Ann Allergy Asthma Immunol*. Jul; 2008 101(1):1–8. [PubMed: 18681077]
3. Fitzpatrick AM, Teague WG, Meyers DA, Peters SP, Li X, Li H, Wenzel SE, Aujla S, Castro M, Bacharier LB, Gaston BM, Bleecker ER, Moore WC, National Institutes of Health/National Heart, Lung, and Blood Institute Severe Asthma Research Program. 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*. Feb; 2011 127(2):382–389. e1–13. [PubMed: 21195471]
4. Just J, Gouvêas-Echraghi R, Rouve S, Wanin S, Moreau D, Annesi Maesano I. Two novel severe asthma phenotypes identified during childhood using a clustering approach. *Eur Respir J*. 2012; 40:55–60. [PubMed: 22267763]
5. Benton AS, Wang Z, Lerner J, Foerster M, Teach SJ, Freishtat RJ. Overcoming heterogeneity in pediatric asthma: tobacco smoke and asthma characteristics within phenotypic clusters in an African American cohort. *J Asthma*. Sep; 2010 47(7):728–34. [PubMed: 20684733]
6. Ortega H, Miller DP, Li H. Characterization of asthma exacerbations in primary care using cluster analysis. *J Asthma*. Mar; 2012 49(2):158–69. [PubMed: 22300042]
7. Simpson A, Tan VY, Winn J, Svensén M, Bishop CM, Heckerman DE, Buchan I, Custovic A. Beyond atopy: multiple patterns of sensitization in relation to asthma in a birth cohort study. *Am J Respir Crit Care Med*. Jun 1; 2010 181(11):1200–6. [PubMed: 20167852]
8. Haldar P, Pavord ID, Shaw DE, Berry MA, Thomas M, Brightling CE, Wardlaw AJ, Green RH. Cluster analysis and clinical asthma phenotypes. *Am J Respir Crit Care Med*. Aug 1; 2008 178(3):218–24. [PubMed: 18480428]
9. Moore WC, Meyers DA, Wenzel SE, Teague WG, Li H, Li X, D'Agostino R Jr, Castro M, Curran-Everett D, Fitzpatrick AM, Gaston B, Jarjour NN, Sorkness R, Calhoun WJ, Chung KF, Comhair SA, Dweik RA, Israel E, Peters SP, Busse WW, Erzurum SC, Bleecker ER, National Heart, Lung, and Blood Institute's Severe Asthma Research Program. Identification of asthma phenotypes using cluster analysis in the Severe Asthma Research Program. *Am J Respir Crit Care Med*. Feb 15; 2010 181(4):315–23. [PubMed: 19892860]
10. Weatherall M, Travers J, Shirtcliffe PM, Marsh SE, Williams MV, Nowitz MR, Aldington S, Beasley R. Distinct clinical phenotypes of airways disease defined by cluster analysis. *Eur Respir J*. Oct; 2009 34(4):812–8. [PubMed: 19357143]
11. Raza A, Kurukulaaratchy RJ, Grundy JD, Clayton CB, Mitchell FA, Roberts G, Ewart S, Sadeghnejad A, Arshad SH. What does adolescent undiagnosed-wheeze represent? Findings from the Isle of Wight Cohort. *Eur Respir J*. Sep; 2012 40(3):580–8. [PubMed: 22267759]
12. Arshad SH, Hide DW. Effect of environmental factors on the development of allergic disorders in infancy. *J Allergy Clin Immunol*. Aug; 1992 90(2):235–41. [PubMed: 1500628]
13. Tariq SM, Matthews SM, Hakim EA, et al. The prevalence of and risk factors for atopy in early childhood: a whole population birth cohort study. *J Allergy Clin Immunol*. 1998; 101:587–93. [PubMed: 9600493]
14. Kurukulaaratchy RJ, Fenn M, Twiselton R, Matthews S, Arshad SH. The prevalence of asthma and wheezing illnesses amongst 10-year-old schoolchildren. *Respir Med*. Mar; 2002 96(3):163–9. [PubMed: 11905550]
15. Asher MI, Keil U, Anderson HR, et al. International Study of Asthma and Allergies in Childhood (ISAAC): rationale and methods. *Eur Respir J*. Mar; 1995 8(3):483–91.
16. Scott M, Raza A, Karmaus W, et al. Influence of atopy and asthma on exhaled nitric oxide in an unselected birth cohort. *Thorax*. 2010; 65:258–62. [PubMed: 20335297]
17. British Thoracic Society/ Scottish Intercollegiate Guidelines Network. British Guideline on the Management of Asthma. A national clinical guideline. Jan. 2012 Revised

18. Gelman A. Scaling regression inputs by dividing by two standard deviations. *Statistics in Medicine*. 2008; Vol. 27:2865–2873. [PubMed: 17960576]
19. Sarle, WS. SAS Technical Report A-108. SAS Institute Inc.; Cary, NC: 1983. Cubic Clustering Criterion.
20. Calinski T, Harabasz J. A dendrite method for cluster analysis. *Communications in Statistics*. 1974; 3:1–27.
21. Siroux V, Basagaña X, Boudier A, Pin I, Garcia-Aymerich J, Vesin A, Slama R, Jarvis D, Anto JM, Kauffmann F, Sunyer J. Identifying adult asthma phenotypes using a clustering approach. *Eur Respir J*. Aug; 2011 38(2):310–7. Epub 2011 Jan 13. [PubMed: 21233270]
22. Spycher BD, Silverman M, Brooke AM, Minder CE, Kuehni CE. Distinguishing phenotypes of childhood wheeze and cough using latent class analysis. *Eur Respir J*. May; 2008 31(5):974–81. Epub 2008 Jan 23. [PubMed: 18216047]
23. Henderson J, Granell R, Heron J, Sherriff A, Simpson A, Woodcock A, Strachan DP, Shaheen SO, Sterne JA. Associations of wheezing phenotypes in the first 6 years of life with atopy, lung function and airway responsiveness in mid-childhood. *Thorax*. Nov; 2008 63(11):974–80. Epub 2008 Aug 4. [PubMed: 18678704]
24. Almqvist C, Worm M, Leynaert B, working group of GA2LEN WP 2.5 Gender. Impact of gender on asthma in childhood and adolescence: a GA2LEN review. *Allergy*. Jan; 2008 63(1):47–57. Epub 2007 Sep 5. [PubMed: 17822448]
25. Van Miert E, Sardella A, Bernard A. Biomarkers of early respiratory effects in smoking adolescents. *Eur Respir J*. Dec; 2011 38(6):1287–93. Epub 2011 May 12. [PubMed: 21565920]
26. Butland BK, Strachan DP. Asthma onset and relapse in adult life: the British 1958 birth cohort study. *Ann Allergy Asthma Immunol*. 2007; 98:337–43. [PubMed: 17458429]
27. Court CS, Cook DG, Strachan DP. Comparative epidemiology of atopic and non-atopic wheeze and diagnosed asthma in a national sample of English adults. *Thorax*. 2002; 57:951–57. [PubMed: 12403877]
28. Genuneit J, Weinmayr G, Radon K, et al. Smoking and the incidence of asthma during adolescence: results of a large cohort study in Germany. *Thorax*. 2006; 61:572–8. [PubMed: 16537668]
29. Yoo S, Kim HB, Lee SY, Kim BS, Kim JH, Yu J, Kim BJ, Lee DH, Seong MW, Hong SJ. Effect of active smoking on asthma symptoms, pulmonary function, and BHR in adolescents. *Pediatr Pulmonol*. Oct; 2009 44(10):954–61. [PubMed: 19728392]
30. Brook JS, Saar NS, Zhang C, Brook DW. Familial and non-familial smoking: effects on smoking and nicotine dependence. *Drug Alcohol Depend*. Apr 1; 2009 101(1-2):62–8. Epub 2008 Dec 19. [PubMed: 19101100]
31. Bouzigon E, Corda E, Aschard H, Dizier MH, Boland A, Bousquet J, Chateigner N, Gormand F, Just J, Le Moual N, Scheinmann P, Siroux V, Vervloet D, Zelenika D, Pin I, Kauffmann F, Lathrop M, Demenais F. Effect of 17q21 variants and smoking exposure in early-onset asthma. *N Engl J Med*. Nov 6; 2008 359(19):1985–94. Epub 2008 Oct 15. [PubMed: 18923164]
32. Flory JH, Sleiman PM, Christie JD, Annaiah K, Bradfield J, Kim CE, Glessner J, Imielinski M, Li H, Frackelton EC, Cuiping H, Otieno G, Thomas K, Smith R, Glaberson W, Garris M, Chiavacci R, Allen J, Spergel J, Grundmeier R, Grunstein M, Magnusson M, Grant SF, Bønnelykke K, Bisgaard H, Hakonarson H. 17q12-21 variants interact with smoke exposure as a risk factor for pediatric asthma but are equally associated with early-onset versus late-onset asthma in North Americans of European ancestry. *J Allergy Clin Immunol*. Sep; 2009 124(3):605–7. [PubMed: 19660801]
33. James AL, Palmer LJ, Kicic E, Maxwell PS, Lagan SE, Ryan GF, Musk AW. Decline in lung function in the Busselton Health Study: the effects of asthma and cigarette smoking. *Am J Respir Crit Care Med*. Jan 15; 2005 171(2):109–14. Epub 2004 Oct 14. [PubMed: 15486340]
34. Kalhan R, Arynchyn A, Colangelo LA, Dransfield MT, Gerald LB, Smith LJ. Lung function in young adults predicts airflow obstruction 20 years later. *Am J Med*. May; 2010 123(5):468, e1–7. [PubMed: 20399325]

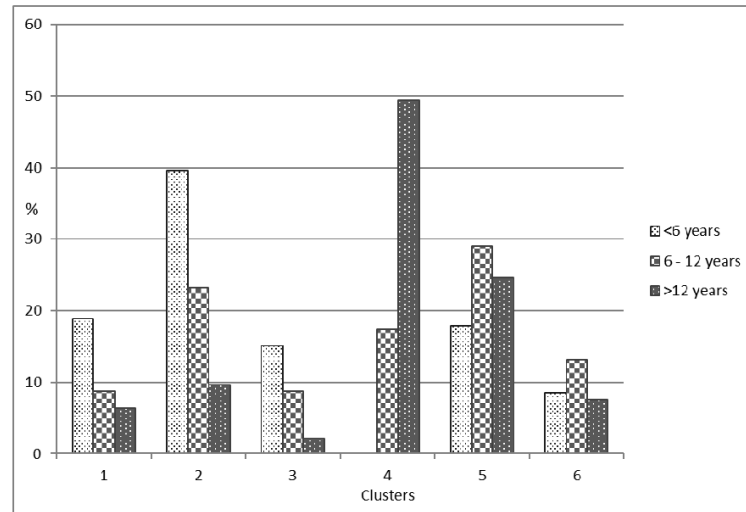


Figure 1. Distribution of Age of Onset Categories across Wheeze Clusters
Shows proportion of age of onset categories across the 6 wheeze clusters.

<6 years: early-childhood-onset

6-12 years: late-childhood-onset

>12 years: adolescent-onset

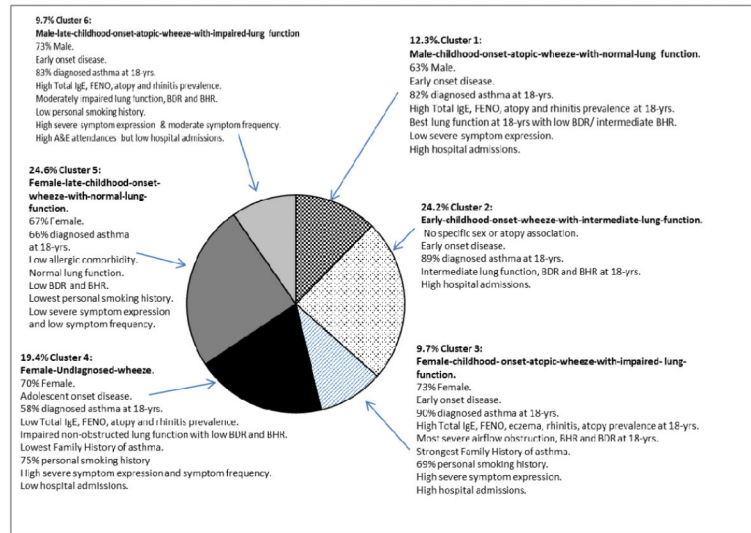


Figure 2. Wheezing Cluster Summary Characteristics

Pie-chart showing proportions and characteristic summary features of the 18-year wheeze cluster groups.

Table 1

Summary Characterisation of Cluster Variables for 18-year Wheeze Clusters

Variables	Cluster samples (n=309)	Age 18 cohort (n=864)	Cluster						Separation of the clusters #
			1 (n=38)	2 (n=75)	3 (n=30)	4 (n=60)	5 (n=76)	6 (n=30)	
Mean (SD)									
FEV1 *	-0.15 (0.48)	0 (0.46)	0.43 (0.26)	-0.27 (0.26)	-0.62 (0.54)	-0.32 (0.19)	0.13 (0.29)	-0.83 (0.33)	{{1},{2,4},{3,6},{5}}
FVC †	-0.02 (0.55)	0 (0.51)	0.58 (0.42)	0.20 (0.44)	-0.06 (0.50)	-0.40 (0.35)	-0.06 (0.39)	-0.76 (0.46)	{{1},{2,3,5},{4},{6}}
FEV1/FVC ‡	0.84 (0.08)	0.87 (0.07)	0.86 (0.034)	0.78 (0.047)	0.74 (0.10)	0.89 (0.04)	0.92 (0.04)	0.83 (0.07)	{{1,6},{1,4},{2,3},{4,5}}
FEF 25-75 §	-0.39 (1.00)	0 (0.98)	0.23 (0.45)	-1.13 (0.42)	-1.48 (0.97)	-0.34 (0.38)	0.56 (0.70)	-1.32 (0.67)	{{1,5},{2,3,6},{4}}
BDR ¶	7.49 (7.11)	5.00 (5.78)	5.85 (3.03)	9.15 (4.65)	22.68 (8.81)	4.41 (5.31)	3.11 (3.13)	8.06 (3.93)	{{1,4,5},{1,2,6},{1,4,6},{3}}
BHR DRS ¶¶	1.30 (0.42)	1.13 (0.30)	1.29 (0.26)	1.26 (0.29)	2.37 (0.33)	1.16 (0.20)	1.07 (0.20)	1.37 (0.29)	{{1,2,4,5},{1,2,3,4,6}}
FeNO **	45.36 (47.79)	27.88 (31.92)	76.24 (63.38)	41.47 (46.72)	65.35 (33.59)	14.13 (6.16)	24.17 (25.74)	93.08 (52.17)	{{1,3,6},{1,2,3},{2,5,4}}
Log ₁₀ Total IgE ††	183.19 (311.84)	174.50 (343.88)	335.68 (402.67)	145.22 (248.47)	140.68 (132.42)	41.78 (51.56)	92.96 (135.42)	797.38 (558.52)	{{1,2,3,5},{1,6},{2,3,4,5}}
BTS Step of Therapy (0-5) †††	1.80 (0.72)	1.79 (0.72)	1.89 (0.88)	2.00 (0.69)	1.76 (0.62)	1.29 (0.46)	1.65 (0.56)	2.28 (0.75)	{{1,2,3,6},{3,5,4},{1,2,3,5}}
Age wheeze 1 st appeared §§ (Years)	8.00 (5.58)	7.34 (5.29)	5.6 (4.59)	4.83 (4.14)	4.75 (4.19)	14.02 (2.77)	8.83 (5.56)	8.12 (5.01)	{{1,2,3},{1,6},{4,5,6}}
Prevalence (%)									
Rhinitis at 18	62.99	35.75	71.8	61.8	96.6	46.7	50.6	86.2	{{1,2,4,5},{1,2,6},{3,6}}
Eczema at 18	17.32	12.32	15.4	13.3	57.1	16.7	12.0	6.9	{{1,2,4,5,6},3}
¶¶ Atopy at 18	63.13	41.27	83.3	57.6	100	33.3	50.8	85.7	{{1,2,6},{1,3,6},{2,4,5}}

Age 18 cohort refers to the population at 18-years who attended for a full visit and therefore had full available objective data.

* FEV₁ = Forced expiratory volume in first second in litres (L) with standard error (S.E.) at 18-years. It is gender and height adjusted.

† FVC = Forced vital capacity in litres (L) with standard error (S.E.) at 18-years. It is gender and height adjusted.

‡ FEV₁/FVC = ratio of FEV₁ to FVC at 18-years.

§ FEF_{25-75%} = Forced expiratory flow 25-75% in litres per second (L/s) with standard error (S.E.) at 18-years. It is gender and height adjusted.

¶ BDR refers to %FEV₁ Bronchodilator reversibility: to 600 micrograms inhaled salbutamol at 18-years.

¶¶ BHR DRS refers to a continuous dose-response (DRS) measure of bronchial hyperresponsiveness (BHR) expressed as Log₁₀ (DRS+10). Higher values infer greater BHR at 18-years.

** FeNO refers to Log₁₀ Fractional exhaled Nitric Oxide measured in parts per billion (ppb) at 18-years at 18-years.

†† Log₁₀ Total IgE at 18-years.

‡‡ BTS (British Thoracic Society) Step of Therapy (0-4); where level of treatment increases with step.

§§ Age wheeze 1st appeared (in years; determined from prospectively collected data throughout the 18-years of the cohort).

// Atopy refers to positive skin prick test (>3mm wheal diameter) to at least allergen in standard test panel at 18-years.

ANOVA for continuous variables indicates that there is a significant difference in the statistics across the six clusters (all p-values<0.0001). Chi-squared tests on the binary variables give the same conclusion. The information provided in this column is to display the groupings. Pair-wised t-tests were applied to continuous variables, and pair-wised proportion tests were applied to binary variables both with Bonferonni multiple testing corrections. Numbers included within a pair of curly brackets indicate the statistics within those clusters are not significantly different from each other after adjusting for multiple testing. Numbers in different pairs of curly brackets indicate that there is a significant difference between those grouped clusters.

Table 2

The Distribution of BTS Treatment Stage across each Wheeze Cluster.

Cluster	BTS Treatment stages (%)			
	1	2	3	4
1 (n=38)	36.84	42.11	15.79	5.26
2 (n=75)	20.93	60.47	16.28	2.33
3 (n=30)	33.33	57.14	9.52	0
4 (n=60)	71.43	28.57	0	0
5 (n=76)	38.46	57.69	3.85	0
6 (n=30)	11.11	55.56	27.78	5.56

BTS (British Thoracic Society) Step of Therapy (0-4); where level of treatment increases with step.

BTS Step 1; short-acting inhaled beta-agonist as and when needed.

BTS Step 2; regular inhaled corticosteroid.

BTS Step 3; regular inhaled corticosteroid + inhaled long acting beta-agonist.

BTS Step 4; regular inhaled corticosteroid + multiple add-on therapy.

Table 3

Morbidity Measures for the Wheeze Clusters at 18-years

	Cluster 1 (n=38) % (count)	Cluster 2 (n=75) % (count)	Cluster 3 (n=30) % (count)	Cluster 4 (n=60) % (count)	Cluster 5 (n=76) % (count)	Cluster 6 (n=30) % (count)	p-value
Wheezing attacks in 12 months							0.06
None	16.2 (6)	16.4 (12)	6.9 (2)	8.5 (5)	26.7 (20)	10.0 (3)	
1 to 3	56.8 (21)	57.5 (42)	55.2 (16)	52.5 (31)	48.0 (36)	50.0 (15)	
4 to 12	18.9 (7)	17.81 (13)	20.7 (6)	18.6 (11)	22.7 (17)	30.0 (9)	
>12	8.1 (3)	8.2 (6)	17.24 (5)	20.3 (12)	2.7 (2)	10.0 (3)	
Sleep disturbance	10.8 (4)	13.9 (10)	21.4 (6)	22.8 (13)	8.6 (6)	24.0 (6)	0.17
Limited speech	13.5 (5)	18.3 (13)	10.7 (3)	25.9 (15)	15.7 (11)	3.9 (1)	0.17
Limited Exercise	68.4 (26)	75.7 (56)	86.2 (25)	69.5 (41)	67.1 (51)	73.3 (22)	0.44
Mean No. of overnight Hospital admissions	0.14	0.13	0.14	0.02	0.04	0.03	0.30
Mean No. of A&E visits	0.22	0.26	0.25	0.03	0.09	0.19	0.17

Sleep Disturbance; Sleep disturbed by wheeze in the past year.

Speech Limitation; Speech limited by wheezing in the past year.

Exercise Limitation; Exercise limited by wheezing in the past year.

Healthcare utilisation refers to between the ages of 10-18-years;

Overnight Hospital Admissions; Overnight hospital admissions for wheezing.

A&E/Ward visits; Attendance with wheezing at Accident & Emergency or day ward without need for admission.

the p-values are for the tests comparing the distribution patterns of a categorical variable in each cluster or comparing medians (continuous variables) across the 6 clusters. For categorical variables, chi-square tests were implemented and for continuous variables Kruskal-Wallis tests were used due to the violation of the normality assumption in the residuals.

Table 4

Risk Factor Association with the Wheeze Clusters at 18-years

	Cluster 1 (n=38) % (count)	Cluster 2 (n=75) % (count)	Cluster 3 (n=30) % (count)	Cluster 4 (n=60) % (count)	Cluster 5 (n=76) % (count)	Cluster 6 (n=30) % (count)	p-value
Male Gender	63.2 (24)	53.3 (40)	26.7 (8)	30.0 (18)	32.9 (25)	73.3 (22)	<0.0001
Family history of asthma	57.9 (22)	66.7 (50)	80.0 (24)	51.7 (31)	64.5 (49)	60.0 (18)	0.15
Low birth weight	2.7 (1)	10.7 (8)	0	3.4 (2)	1.4 (1)	3.5 (1)	0.06
Exclusively breastfed	31.0 (9)	19.7 (12)	25.0 (7)	32.6 (15)	30.7 (19)	32.0 (8)	0.66
Chest infection in Infancy	14.8 (4)	29.8 (17)	18.5 (5)	16.3 (7)	19.4 (13)	20.8 (5)	0.54
Personal smoking	51.4 (19)	55.1 (38)	69.0 (20)	74.6 (44)	49.3 (37)	51.7 (15)	0.04
Maternal smoking in pregnancy	36.8 (14)	42.7 (32)	33.3 (10)	33.3 (20)	38.2 (29)	26.7 (8)	0.71
Mean Cord IgE	-0.62	-0.77	-0.69	-0.82	-0.76	-0.78	0.55
Mean Paracetamol Use	2.32	1.92	2.11	2.79	2.45	2.81	0.61
Mean BMI	24.80	24.86	24.88	23.26	23.49	23.13	0.62

Family History of asthma = Family (parent or sibling) history of asthma.

Low Birth weight = Low Birth Weight (<2.5kg).

Exclusively Breastfed = Exclusively breastfed for > 3 months.

Chest infection in infancy = History of recurrent chest infections at 1 or 2-years.

Personal Smoking = Personal (past or current) history of cigarette smoking.

Maternal Smoking = Maternal smoking history during pregnancy.

Cord IgE = Cord blood IgE at birth.

Paracetamol Use = Paracetamol use at 18-years.

BMI = Body Mass Index (Weight [Kg]/ Height² [M]) at 18-years.

the p-values are for the tests comparing the distribution patterns of a categorical variable in each cluster or comparing medians (continuous variables) across the 6 clusters. For categorical variables, chi-square tests were implemented and for continuous variables Kruskal-Wallis tests were used due to the violation of the normality assumption in the residuals.