**Title: Clinical Features and Later Prognosis of Replicable Early-Life Wheeze Clusters from 2 Birth Cohorts 12-years Apart**

**Authors:**

Suzanne Y Ngo, MD 1,2

Carina Venter, PhD, RD 1

William C Anderson III, MD 1

Kaci Picket, MS 1

Hongmei Zhang, PhD, MS 3

S Hasan Arshad, DM 4,5,6

Ramesh J Kurukulaaratchy, DM 4,5,6

**Affiliations:**

1Section of Allergy and Immunology, Department of Pediatrics, University of Colorado School of Medicine, Children’s Hospital Colorado, Aurora, Colo.

2Division of Allergy and Clinical Immunology, Department of Medicine, University of Colorado School of Medicine, Aurora, CO.

3Division of Epidemiology, Biostatistics, and Environmental Health, School of Public Health, University of Memphis, Memphis, TN.

4School of Clinical and Experimental Sciences, Faculty of Medicine, University of Southampton, UK.

5National Institute for Health Research (NIHR) Southampton Biomedical Research Centre at University Hospital Southampton NHS Foundation Trust, Southampton, UK.

6The David Hide Asthma & Allergy Centre, St Mary’s Hospital, Newport, Isle of Wight, UK.

**Running Title**: Replicable Early-Life Wheeze Clusters

**Corresponding Author:**

Ramesh J Kurukulaaratchy, DM

Clinical Experimental Sciences, Mailpoint 810, F-Level, South Academic Block, Southampton General Hospital, Tremona Road, Southampton, SO16 6YD.

+44 238120 5232

R.J.Kurukulaaratchy@soton.ac.uk

**Word Count Abstract:** 246

**Word Count Manuscript Body:** 2835

**Conflicts of Interest**

Dr. Venter reports grants from Reckitt Benckiser, Food Allergy Research and Education, National Peanut Board; personal fees from Reckitt Benckiser, Nestle Nutrition Institute, Danone, Abbott Nutrition, Else Nutrition, Sifter and Before Brands.

Dr. Anderson has served as an advisory board member for Regeneron, Sanofi, and Genentech. He has grant support from the Colorado Medicaid Supplemental Funding Program and the COPIC Medical Foundation.

Dr. Zhang’s work was supported by R01AI121226.

All other authors have no conflicts of interest to report.

**Financial Support:**

The Isle of Wight Birth Cohort assessments have been supported by the National Institutes of Health USA (Grant no. R01 HL082925 and R01 AI121226), Asthma UK (Grant no. 364) and the David Hide Asthma and Allergy Research Trust. The Food Allergy and Intolerance Research (FAIR) cohort was funded by the Food Standards Agency UK and the National Institute of Health Research (NIHR) UK.

**Abstract**

**Background**

Clustering techniques can define the heterogeneity of asthma and wheezing. Defining early-life wheezing clusters and associated asthma risk could potentially inform patient management strategies. Clustering models that yield replicable cluster groups will have greater validity and clinical utility. This study sought to identifyearly-life wheezing clusters that are translatable into clinical practice and assess their stability over time in two whole-population birth cohorts established a decade apart from the same geographical location.

**Methods**

Non-parametric K-means cluster analysis was performed separately on 2 birth cohorts from the Isle of Wight, UK; the Isle of Wight Birth Cohort (IOWBC) and Food Allergy and Intolerance Research Cohort (FAIR), using clinically defining variables in wheezing subjects in the 1st 3-4 years. Associations of resulting clusters with potential early-life risk factors and 10-year asthma outcomes were further assessed.

**Results**

Five clusters were identified in both cohorts: 1) infantile-onset-transient-non-atopic-wheeze, 2) infantile-onset-persistent-non-atopic-wheeze, 3) infantile-onset-atopic-wheeze, 4) early-childhood-onset-non-atopic-wheeze and 5) early-childhood-onset-atopic-wheeze. Two atopic wheezing clusters (3 and 5) were associated with greatest early-life wheeze frequency, highest wheeze persistence and asthma prevalence at 10-years. Cluster 1 was commonest but had lowest early-life wheeze frequency and asthma prevalence at 10-years. Cluster 2, characterized by limited atopy but recurrent infantile respiratory infections and ongoing early-life wheezing had high 10-year asthma prevalence only in IOWBC.

**Conclusions**

Early-life wheeze comprises several disease clusters (2 more severe and 3 mild-moderate) with differing relationships to later childhood asthma, which can be replicated over time supporting their potential validity and clinical utility.

**Key words:** asthma, early-life wheezing, wheezing phenotypes, atopy

**Abbreviations**

IOWBC Isle of Wight Birth Cohort

FAIR Food Allergy and Intolerance Research birth cohort

SPT skin prick test

CCC cubic clustering criterion

ANOVA analyses of variance

BMI body mass index

NAUW Non-Atopic Uncontrolled Wheeze

**Introduction**

There is growing understanding of unique phenotypes of early-life wheezing.1 Gaining insight into what clinical features predispose a child who wheezes to develop asthma may help in earlier identification, risk modification and treatment to alter clinical course. Prior longitudinal studies evaluating wheezing patterns in young children identified differences in characteristics and outcomes based on the timing of wheezing onset and duration.2–8 Further evaluation has identified additional factors, including personal and family history of atopy, lung function, ethnicity, sex and other environmental exposures that modulate asthma risk.3,5,9 Better understanding of clinically relevant classifications for both asthma and wheezing have been advanced through unbiased techniques of cluster and latent class analyses.10–13 Establishing cluster groupings with similar features and outcomes in different populations would increase validity of a clustering model. Similarly, stability among clustering groups over different time periods would indicate longevity of a particular model.

In this study, we employed a cluster analysis using clinically relevant and readily available variables to classify early-life wheezing clusters that reflect real-world clinical practice. Our study utilized two prospectively followed birth cohorts from the Isle of Wight (United Kingdom; UK) that commenced 12-years apart to evaluate the reproducibility of the cluster modelling over time. Following cluster identification, associated clinical characteristics and 10-year outcomes were assessed to gain insight on possible clinical and prognostic implications of each cluster.

**Methods**

Two whole population birth cohorts established on the Isle of Wight 12 years apart were evaluated in this study. The Isle of Wight birth cohort (IOWBC) (n=1456) was established in 1989 and the Food Allergy and Intolerance Research birth cohort (FAIR) (n=969) was established in 2001-2002 as whole population birth cohorts to study the natural history of asthma and allergic disease. Details of methodology and follow-up studies for the IOWBC and FAIR studies have been published previously.7,14–20 Ethical approval was obtained for the IOWBC from the Isle of Wight NHS Ethics committee (No 05/89; dated 08/22/1988) and for the FAIR cohort from the NRES South Central - Southampton B Research Ethics Committee (REF 10/H0504/11).

In both cohorts, informed consent from the subject’s parents and information on first degree family history of allergy, parental smoking, socioeconomical status, household pets and birth weight were obtained. Subjects were evaluated at predetermined intervals. Cohort subjects were evaluated at birth and ages 1, 2, 4 and 10-years for the IOWBC and at birth and ages 1, 2, 3 and 10-years for FAIR. Feeding and weaning practices, immunization history, and exposure to relevant environmental factors (e.g., domestic pets and tobacco smoke) were recorded at the 1-year evaluation. At 1, 2 and either 3 or 4-years of age, parents also completed a series of questionnaires to provide information on clinical asthma and allergy status. These questionnaires recorded symptoms of atopy, including features for eczema, rhinitis, and food allergy. Presence of wheezing within the last 12 months, history of recurrent chest infections, presence of nocturnal cough or sleep disturbance, diagnosis of asthma and treatment for asthma were also assessed. Specific definitions/ questions used in the 2 cohorts are given in supplementary Table E1. Notable differences include different wheezing ranges used for the ordinal scale of wheezing frequency and requirement for 3 separate episodes of wheezing and/or cough in addition to physician diagnosis of asthma in the IOWBC to meet criteria for diagnosis of asthma at 4-years. Current asthma at 10-years was defined as “physician diagnosed asthma ever” plus “current wheeze.” Skin prick testing (SPT) was performed in all consenting subjects at 3- (FAIR) or 4-years (IOWBC) and again at 10-years to a standard aeroallergen panel.17,19

*Statistical methods*

All statistical analyses were performed using the SAS statistical package version 9.4 (SAS Institute, Cary, NC, USA). The two cohorts were analyzed separately. A non-parametric K-means clustering method (PROC FASTCLUS in SAS) was applied to detect the clusters of subjects. This approach is selected due to its simplicity and robustness to different shapes of data. The cluster analysis was performed in each cohort on subjects with wheezing at any point within the past 12 months at ages 1, 2, 3 or 4-years. Subjects with no recorded wheezing in the first 3- (FAIR study) to 4-years (IOWBC study) of life were excluded. Cluster variables were selected that defined clinical characteristics in the first 3- to 4-years. These included both questionnaire-derived data and variables from objective testing. Six clinical disease defining variables derived from data obtained in the first 3- (FAIR) to 4-years (IOWBC) were selected for the cluster analyses: age of wheeze onset, wheeze frequency at age 3 or 4-years, atopy (positive SPT) at 3- or 4-years, eczema by 3- or 4-years, food allergy at 3- or 4-years, and rhinitis at 3- or 4-years. We standardized continuous variables to reduce clustering bias due to different scales or units. The final choice for the number of clusters was determined by an overall evaluation of CCC21 and pseudo F,22 combined withan R2 statistic measuring between cluster variation (Appendix A, Table E3).

To evaluate variation of each clustering variable across the clusters, analyses of variance (ANOVA) was used for continuous variables and Chi-square tests for categorical variables. To test differences between clusters, pairwise t-tests were applied to continuous variables, and forcategorical variables pairwise proportion tests were applied. Associations with potential risk factors recorded during the first 4-years were then assessed. These included male sex, family history (parent or sibling) of asthma, lower socio-economic status at birth, mode of delivery, maternal smoking in pregnancy, exclusively breastfed, recurrent chest infections in infancy (age 1- or 2-years), birth weight, Body Mass Index (BMI) z-score at 2-years.In these assessments, Chi-square tests and ANOVA were applied for categorical and continuous variables, respectively.

**Results**

Among cohort subjects, 367 (23.9%) had previously recorded wheeze during the first 4-years of life in the IOWBC and 485 (50%) during the first 3-years of life in FAIR. These subjects with a history of wheeze formed the respective populations for cluster analysis.

Variables used in the cluster analysis and additional potential risk factors between the IOWBC and FAIR studies showed significant differences (Table 1). In the respective clustered populations, estimated mean onset of wheeze was later in the IOWBC compared to FAIR (1.5-years vs. 0.8-years, p<0.001). Frequency of wheezing at 3- or 4-years was higher in the IOWBC (1.2 vs. 0.5 on ordinal scale, p<0.001), however scales used for analysis differed. In the IOWBC, there was also significantly higher prevalence of atopic sensitization (26% vs. 13%, p<0.001) and food allergy (17% vs. 7%, p<0.001), but lower prevalence of eczema (29% vs. 65%, p<0.001) and rhinitis (11% vs. 36%, p<0.001). Diagnosis of asthma was also significantly higher in the IOWBC at both preschool age (49% vs. 17%, p<0.001) and at 10-years of age (29% vs. 13%, p<0.001).

In separate cluster analyses of the IOWBC and FAIR early-childhood wheezers, five clusters were identified: 1) infantile-onset-transient-non-atopic-wheeze, 2) infantile-onset-persistent-non-atopic-wheeze, 3) infantile-onset-atopic-wheeze, 4) early-childhood-onset-non-atopic-wheeze and 5) early-childhood-onset-atopic-wheeze. These 5 clusters from the two separate cohorts were paired based on similar characteristics among clustering variables as outlined in Table 2. Age of wheeze onset separated into three early-onset (1, 2, 3) and two late-onset (4, 5) clusters. The overall proportional split of early-onset and late-onset wheeze was similar in the 2 cohorts, though there were differences in prevalence of individual clusters. Atopy and allergic diseases (eczema, food allergy and rhinitis) segregated into lower (1, 2, 4) and higher (3, 5) prevalence clusters. At age 3-4, there was significant difference among clusters for prevalence of current wheezing (p < 0.001), diagnosis of asthma (p < 0.001) and treatment for asthma (p < 0.001) with consistently lowest prevalence in cluster 1 (indicating an early, transient wheeze pattern) but highest for clusters 3 and 5 in both cohorts (consistent with greater disease severity). Furthermore, cluster 2 demonstrated moderate to high current wheezing, asthma diagnosis and treatment needs in both cohorts but these trends were more pronounced in IOWBC.

Associations of the early-childhood wheeze clusters with potential early-childhood risk factors were also evaluated. In the IOWBC, no significant differences were observed between clusters for birthweight, BMI z-score at 2-year, male sex, maternal smoking during pregnancy, family history of asthma, lower socioeconomic status at birth, vaginal delivery, exclusive breastfeeding for the first 3 months of life and pet exposure during infancy (Supplemental Table E2). However, there were significant differences between clusters in the FAIR cohort (Table 3). Cluster 3 had the highest while cluster 5 had the lowest BMI z-scores (0.9 vs. -0.3, p = 0.017), prevalence of family history of atopy (76% vs. 35%, p = 0.0014). There was also significant difference among clusters based on families with an annual household income less than £50K with similar trend with higher prevalence in cluster 3 compared to cluster 5 (49% vs. 20%, p > 0.001). There was also a similar non-significant trend for maternal smoking in pregnancy in the FAIR cohort data (p = 0.064) such that there was higher prevalence in cluster 3 (31%) compared to cluster 5 (16%).

Asthma and wheezing outcomes at 10-years were evaluated for each cluster (Figure 1). Across both cohorts, there were significant differences for both 10-year wheezing frequency (p < 0.001) and diagnosis of asthma (p < 0.001) among clusters with prevalence lowest in cluster 1 and highest among clusters 3 and 5, which were associated with higher prevalence of atopy and greater wheeze frequency in earlier childhood. Discrepancies in asthma outcomes at 10-years of age were noted between cohorts for cluster 2, with subjects from the IOWBC continuing to exhibit higher asthma prevalence (43% vs. 25%, p = 0.020) and treatment need (47% vs 29%, p = 0.028) compared to FAIR. . Additionally, in the FAIR cohort, while cluster 2 showed moderate 10-year asthma treatment needs, their asthma and wheeze prevalence was comparably low to that of subjects with persistent non-atopic wheezing seen in cluster 4. In the IOWBC, cluster 2 had the highest rates of recurrent respiratory infections at ages 1-2 (54%, p < 0.001) plus highest 4-year wheezing frequency among clusters with infantile-onset wheezing (1-3) (100%, p < 0.001). This pattern was not observed in FAIR, where 3-year wheeze frequency was low in cluster 2 (34%).

Key features of each cluster, associated risk factors and 10-year outcomes are summarized in Figure 2.

**Discussion**

To our knowledge, this is the first study to characterize early-life wheezing clusters in 2 longitudinal birth cohorts enrolled within the same geographical location a decade apart. We defined early-life wheeze clusters, their associations with later childhood asthma outcomes and the stability of these clusters over time, thus providing confidence in their clinical utility. Five distinct wheezing clusters were identified in both. These clusters were distinguished by clinically relevant and easily obtainable variables including age of wheeze onset, wheeze frequency, atopic status, allergic comorbidity, early-life chest infections, plus prevalence of current wheeze, asthma, and asthma treatment at 3-4-years. We demonstrated associations among these clusters with later childhood asthma outcomes and the stability of these clusters over time, thus providing confidence in their clinical utility.

Two higher severity atopic clusters were identified; one with onset in infancy (cluster 3;infantile-onset-persistent-atopic-wheeze) and the other in early childhood (cluster 5; early-childhood-onset-persistent-atopic-wheeze). These 2 clusters showed stronger association with asthma in later childhood. Milder non-atopic wheezing clusters (1; infantile-onset-transient-non-atopic-wheeze and 4; early-childhood-onset-transient-non-atopic-wheeze) were also demonstrated that did not show strong associations with later childhood asthma. An infantile-onset-non-atopic-wheezecluster (2) showed moderate-high wheeze/asthma features in early-life and was associated with later childhood asthma and treatment need.

Similar patterns have been observed in other studies on early-life wheezing. Early-onset, transient wheezing, which resolves by 3-4-years of age as characterized by cluster 1, is the commonest group of early-life wheezers and have the lowest risk for developing asthma.2,3,11 This cluster had lowest association to allergic disease among all clusters. Therefore, infantile-onset wheezers with a brief preschool wheezing history and no history of atopy are unlikely to require additional asthma-type management.

In contrast, persistent and later onset wheezing can result in higher rates of asthma later in childhood, and this risk can be further increased by a history of atopy.1 By differentiating clusters by both timing of onset and presence of atopy, this study’s cluster analysis demonstrated that history of atopy, regardless of when onset of wheezing occurred, was a strong predictor of higher prevalence of asthma at 10-years of age. Both cluster 3 and cluster 5 resulted in similar higher persistence of asthma at age 10, while cluster 4 had lower persistent asthma at age 10. These findings further support the inclusion of allergen sensitization and history of eczema as criteria for higher risk of asthma persistence in several childhood prediction models.23

While there was strong agreement in 10-year outcomes in clusters 1, 3, 4 and 5, there were discrepant findings among the two cohorts for cluster 2 (infantile-onset-persistent-non-atopic-wheeze) with higher rates of persistent asthma among subjects in the earlier established IOWBC. This discrepancy may be associated with an increased history of recurrent respiratory infections among cluster 2 patients observed in the IOWBC. Data on the history of recurrent respiratory infections was not obtained in the FAIR cohort to corroborate this possible association. Other studies have observed increased risk for asthma with higher frequency of early-life wheezing. Cluster analysis of the Trousseau Asthma Program cohort identified the Non-Atopic Uncontrolled Wheeze (NAUW) cluster, characterized by increased wheezing frequency that persisted beyond early childhood in the absence of atopy and had poorer response to inhaled therapy.11 Given possible absence of increased respiratory infections in the later FAIR cohort, this may reflect better understanding and mitigation of risk of early life respiratory infections to reduce incidence of asthma. Further studies to understand this trend over time are needed.

Within the FAIR cohort, prevalence of family history of asthma, increased BMI z-score, lower household income and maternal smoking during pregnancy was highest in cluster 3 and lowest in cluster 5. These two clusters have high levels of atopy and primarily differed by timing of wheezing onset, so these factors may play a role in earlier onset of wheezing in atopic children. Studies have shown that family history of atopy,24 low socioeconomic level,25 infant BMI26 and maternal smoking during pregnancy27 do increase the risk for wheezing and asthma, but there has not been a clear association with earlier onset of wheezing. Maternal smoking during pregnancy also increases the risk for early-life viral infections, including bronchiolitis, which may result in earlier presentation of wheezing in this subset of patients.28 There were no statistically significant differences in associated early-life risk factors among clusters in the IOWBC, including to any of those identified in the FAIR cohort. Prior analysis of the IOWBC did demonstrate increased risk for persistent wheezing with family history of atopy, parental smoking and lower social class at birth, but these trends were not observed among studied clusters.29 The significant difference in environmental factors on wheezing and asthma outcomes in the FAIR cohort may indicate they are playing a larger role on disease in recent years. Some of these factors, such as low socioeconomic level, infant BMI and parental smoking, can reflect social disadvantages that are concerningly causing more pronounced effects over time. This could highlight an increasing need to identify and address such disparities with public health measures.

A major strength of this study is the ability to compare the wheezing patterns of children from two different time periods. By studying two cohorts from the same geographical region, many confounding variables were minimized, though significant differences in several characteristics were noted. The prevalence of wheezing between 3-4 years of age was significantly higher in FAIR (50% of cohort) compared to IOWBC (29% of cohort). Recognition of wheezing may have increased over time causing higher reported wheezing rates in the more recent FAIR cohort. Potentially aligned to that concept, prevalence of milder wheeze (represented by cluster 1) accounted for a higher proportion of early-childhood wheezing in FAIR than IOWBC.

Additionally, among these subjects with early-life wheezing, 49% met the definition for having asthma in the IOWBC compared to only 17% in FAIR at 3-4 years of age. The definition of asthma at this early age in IOWBC was based on “physician diagnosis plus recurrent wheeze” while in FAIR it was “parentally reported.” Such different criteria may partly explain these differences as may the fact that the FAIR early-childhood assessment was one year earlier than IOWBC. Furthermore, over time in the UK, it has become widely accepted that making a secure asthma diagnosis before the age of 6 is tenuous and may have led to lower early-childhood asthma labelling in FAIR. Nevertheless, there also remained a higher prevalence of asthma among subjects in the IOWBC at 10-years of age. The FAIR cohort was established in 2001-2002, so this pattern may also reflect the reported decrease in asthma prevalence observed after 2006, though this trend would not account for the entire magnitude of difference between both cohorts.30 Other early-life factors, such as allergen sensitization, breastfeeding rates and socioeconomic level, may also contribute to these differences in asthma prevalence.

Limitations in the study include reliance on subjective reports for many variables assessed, including wheezing frequency, other atopic symptoms, and nature of early-life chest infections. Asthma outcomes evaluated were also restricted to subjective reports of ongoing wheezing, asthma diagnosis and use of asthma therapies. Definitions utilized between the two cohorts were not equivalent and may have accounted for some of the differences in characteristics observed. Further studies to evaluate other objective outcomes, including lung function and response to specific therapies, could better delineate cluster outcomes and provide guidance for management approaches for each cluster. Additionally, a geographically limited population limits wider applicability of these findings due to the homogenous patient population residing on an island. Expansion of this cluster analysis model to other populations can further establish its clinical relevance.

The findings of this cluster analysis on two similar populations adds to the growing understanding of the heterogenicity of early-life wheezing. Clustering patterns observed in this study indicate that early-life wheezing associated with atopy or other allergic comorbidities, regardless of wheezing age of onset, has the highest rates of asthma at age 10. This highlights the importance for evaluation of allergic comorbid conditions to identify and treat these patients earlier. In children who do not have atopy, early-onset of wheezing that persists beyond 3-4 years of age also have a higher risk of carrying a diagnosis of asthma at 10 years of age. Differences in clusters between cohorts may also indicate longitudinal trends in factors that may affect onset of wheezing and asthma. Specifically, risk increased by higher frequency of early-life respiratory infections may be less pronounced over time with better recognition of their role as a risk factor. Conversely, there is increasing importance of environmental factors including low socioeconomic level, infant BMI and maternal tobacco smoking that require addressing. Further study of these childhood wheeze clusters should focus on whether earlier detection and management and mitigation of risk factors has potential to improve longer-term health outcomes.

**Impact Statement**

Replicable early-life wheezing clusters based on clinically relevant parameters, in two birth cohorts established a decade apart, identify distinct patterns of childhood wheezing associated with differing morbidity and persistence risk. This study highlights higher risk childhood wheeze phenotypes that merit close attention in clinical practice.

**References**

1. Kwong CG, Bacharier LB. Phenotypes of wheezing and asthma in preschool children. *Curr Opin Allergy Clin Immunol*. 2019;19(2):148-153. doi:10.1097/ACI.0000000000000516

2. Martinez FD, Wright AL, Taussig LM, Holberg CJ, Halonen M, Morgan WJ. Asthma and Wheezing in the First Six Years of Life. *N Engl J Med*. 1995;332(3):133-138. doi:10.1056/nejm199501193320301

3. Henderson J, Granell R, Heron J, et al. Associations of wheezing phenotypes in the first 6 years of life with atopy, lung function and airway responsiveness in mid-childhood. *Thorax*. 2008;63(11):974-980. doi:10.1136/thx.2007.093187

4. Stern DA, Morgan WJ, Halonen M, Wright AL, Martinez FD. Wheezing and bronchial hyper-responsiveness in early childhood as predictors of newly diagnosed asthma in early adulthood: a longitudinal birth-cohort study. *Lancet*. 2008;372(9643):1058-1064. doi:10.1016/S0140-6736(08)61447-6

5. Midodzi WK, Rowe BH, Majaesic CM, Saunders LD, Senthilselvan A. Predictors for wheezing phenotypes in the first decade of life. *Respirology*. 2008;13(4):537-545. doi:10.1111/j.1440-1843.2008.01284.x

6. Granell R, Henderson AJ, Sterne JA. Associations of wheezing phenotypes with late asthma outcomes in the Avon Longitudinal Study of Parents and Children: A population-based birth cohort. *J Allergy Clin Immunol*. 2016;138(4):1060-1070.e11. doi:10.1016/j.jaci.2016.01.046

7. Kurukulaaratchy RJ, Fenn MH, Waterhouse LM, Matthews SM, Holgate ST, Arshad SH. Characterization of wheezing phenotypes in the first 10 years of life. *Clin Exp Allergy*. 2003;33(5):573-578. doi:10.1046/j.1365-2222.2003.01657.x

8. Savenije OE, Granell R, Caudri D, et al. Comparison of childhood wheezing phenotypes in 2 birth cohorts: ALSPAC and PIAMA. *J Allergy Clin Immunol*. 2011;127(6). doi:10.1016/j.jaci.2011.02.002

9. Tse SM, Rifas-Shiman SL, Coull BA, Litonjua AA, Oken E, Gold DR. Sex-specific risk factors for childhood wheeze and longitudinal phenotypes of wheeze. *J Allergy Clin Immunol*. 2016;138(6):1561-1568.e6. doi:10.1016/j.jaci.2016.04.005

10. Just J, Gouvis-Echraghi R, Couderc R, Guillemot-Lambert N, Saint-Pierre P. Novel severe wheezy young children phenotypes: Boys atopic multiple-trigger and girls nonatopic uncontrolled wheeze. *J Allergy Clin Immunol*. 2012;130(1):103-110.e8. doi:10.1016/j.jaci.2012.02.041

11. Just J, Saint-Pierre P, Gouvis-Echraghi R, et al. Wheeze phenotypes in young children have different courses during the preschool period. *Ann Allergy, Asthma Immunol*. 2013;111(4):256-261.e1. doi:10.1016/j.anai.2013.07.002

12. Fitzpatrick AM, Bacharier LB, Guilbert TW, et al. Phenotypes of Recurrent Wheezing in Preschool Children: Identification by Latent Class Analysis and Utility in Prediction of Future Exacerbation. *J Allergy Clin Immunol Pract*. 2019;7(3):915-924.e7. doi:10.1016/j.jaip.2018.09.016

13. Moore WC, Meyers DA, Wenzel SE, et al. Identification of asthma phenotypes using cluster analysis in the severe asthma research program. *Am J Respir Crit Care Med*. 2010;181(4):315-323. doi:10.1164/rccm.200906-0896OC

14. Arshad SH, Hide DW. Effect of environmental factors on the development of allergic disorders in infancy. *J Allergy Clin Immunol*. 1992;90(2):235-241. doi:10.1016/0091-6749(92)90077-F

15. Arshad SH, Stevens M, Hide DW. The effect of genetic and environmental factors on the prevalence of allergic disorders at the age of two years. *Clin Exp Allergy*. 1993;23(6):504-511. doi:10.1111/j.1365-2222.1993.tb03238.x

16. Tariq SM, Matthews SM, Hakim EA, Stevens M, Arshad SH, Hide DW. The prevalence of and risk factors for atopy in early childhood: A whole population birth cohort study. *J Allergy Clin Immunol*. 1998;101(5):587-593. doi:10.1016/S0091-6749(98)70164-2

17. Kurukulaaratchy RJ, Fenn M, Twiselton R, Matthews S, Arshad SH. The prevalence of asthma and wheezing illnesses amongst 10-year-old schoolchildren. *Respir Med*. 2002;96(3):163-169. doi:10.1053/rmed.2001.1236

18. Arshad SH, Holloway JW, Karmaus W, et al. Cohort profile: The isle of wight whole population birth cohort (ioWBC). *Int J Epidemiol*. 2018;47(4):1043-1044I. doi:10.1093/ije/dyy023

19. Pereira B, Venter C, Grundy J, Clayton CB, Arshad SH, Dean T. Prevalence of sensitization to food allergens, reported adverse reaction to foods, food avoidance, and food hypersensitivity among teenagers. *J Allergy Clin Immunol*. 2005;116(4):884-892. doi:10.1016/j.jaci.2005.05.047

20. Venter C, Patil V, Grundy J, et al. Prevalence and cumulative incidence of food hyper-sensitivity in the first 10 years of life. *Pediatr Allergy Immunol*. 2016;27(5):452-458. doi:10.1111/pai.12564

21. Sarle WS. *SAS Technical Report A-108, Cubic Clustering Criterion*. Cary, NC; 1983. https://support.sas.com/documentation/onlinedoc/v82/techreport\_a108.pdf.

22. Caliñski T, Harabasz J. A Dendrite Method Foe Cluster Analysis. *Commun Stat*. 1974;3(1):1-27. doi:10.1080/03610927408827101

23. Smit HA, Pinart M, Antó JM, et al. Childhood asthma prediction models: A systematic review. *Lancet Respir Med*. 2015;3(12):973-984. doi:10.1016/S2213-2600(15)00428-2

24. Bjerg A, Hedman L, Perzanowski MS, Platts-Mills T, Lundbäck B, Rönmark E. Family history of asthma and atopy: In-depth analyses of the impact on asthma and wheeze in 7- to 8-year-old children. *Pediatrics*. 2007;120(4):741-748. doi:10.1542/peds.2006-3742

25. Galobardes B, Granell R, Sterne J, et al. Childhood Wheezing, Asthma, Allergy, Atopy, and Lung Function: Different Socioeconomic Patterns for Different Phenotypes. *Am J Epidemiol*. 2015;182(9):763-774. doi:10.1093/aje/kwv045

26. Ali GB, Bui DS, Lodge CJ, et al. Infant body mass index trajectories and asthma and lung function. *J Allergy Clin Immunol*. 2021;148(3):763-770. doi:10.1016/j.jaci.2021.02.020

27. Farber HJ, Batsell RR, Silveira EA, Calhoun RT, Giardino AP. The impact of tobacco smoke exposure on childhood asthma in a medicaid managed care plan. *Chest*. 2016;149(3):721-728. doi:10.1378/chest.15-1378

28. Carroll KN, Gebretsadik T, Griffin MR, et al. Maternal asthma and maternal smoking are associated with increased risk of bronchiolitis during infancy. *Pediatrics*. 2007;119(6):1104-1112. doi:10.1542/peds.2006-2837

29. Kurukulaaratchy RJ, Matthews S, Arshad SH. Does Environment Mediate Earlier Onset of the Persistent Childhood Asthma Phenotype? *Pediatrics*. 2004;113(2):345-350. doi:10.1542/peds.113.2.345

30. Bloom CI, Saglani S, Feary J, Jarvis D, Quint JK. Changing prevalence of current asthma and inhaled corticosteroid treatment in the UK: Population-based cohort 2006–2016. *Eur Respir J*. 2019;53(4). doi:10.1183/13993003.02130-2018

**Table 1: Comparison of Patient Demographics and Wheezing Characteristics**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Entire Cohort | | | | Population for Cluster Analysis | | | |
|  | **IOWBC (N=1536)** | **FAIR (N=969)** | **Total (N=2505)** | **p value** | **IOWBC (N=367)** | **FAIR**  **(N=485)** | **Total (N=909)** | **p value** |
| Estimated age of Wheeze Onset (years) |  |  |  | **<0.0001** |  |  |  | **<0.001** |
| Mean (sd) | 1.6 (1.2) | 0.8 (0.8) | 1.2 (1.1) |  | 1.5 (1.2) | 0.8 (0.8) | 1.1 (1.0) |  |
| N-Not Collected | 1018 | 450 | 1468 |  | 20 | 0 | 20 |  |
| Continuous/ Ordinal Wheeze Frequency† |  |  |  | **<0.001** |  |  |  | **<0.001** |
| Mean (sd) | 0.9 (1.2) | 0.3 (0.6) | 0.5 (0.9) |  | 1.3 (1.3) | 0.5 (0.8) | 0.8 (1.1) |  |
| N-Not Collected | 1018 | 79 | 1097 |  | 20 | 0 | 20 |  |
| Atopy at 3 or 4-years |  |  |  | **<0.001** |  |  |  | **<0.001** |
| Yes | 193 (20%) | 70 (10%) | 263(16%) |  | 81 (26%) | 46 (13%) | 127 (19%) |  |
| No | 787 (80%) | 660 (90%) | 1387(84%) |  | 235 (74%) | 317 (87%) | 552 (81%) |  |
| N-Not Collected | 556 | 299 | 855 |  | 51 | 122 | 230 |  |
| Eczema in the 1st 3-4-years |  |  |  | **<0.001** |  |  |  | **<0.001** |
| Yes | 263 (18%) | 579 (60%) | 842 (35%) |  | 105 (29%) | 314 (65%) | 419 (49%) |  |
| No | 1185 (82%) | 380 (40%) | 1565 (65%) |  | 261 (71%) | 170 (35%) | 431 (51%) |  |
| N-Not Collected | 88 | 10 | 98 |  | 1 | 1 | 2 |  |
| Food allergy in the first 3-4-years |  |  |  | 0.014 |  |  |  | **<0.001** |
| Yes | 127 (9%) | 58 (6%) | 185 (8%) |  | 61 (17%) | 34 (7%) | 95 (11%) |  |
| No | 1292 (91%) | 894 (94%) | 2186 (92%) |  | 305 (83%) | 450 (93%) | 755 (89%) |  |
| N-Not Collected | 117 | 17 | 134 |  | 1 | 1 | 20 |  |
| Rhinitis at 3-4-years |  |  |  | **<0.001** |  |  |  | **<0.001** |
| Yes | 65 (5%) | 279 (31%) | 344 (16%) |  | 39 (11%) | 176 (36%) | 215 (25%) |  |
| No | 1149 (95%) | 611 (69%) | 1760 (84%) |  | 327 (89%) | 308 (64%) | 635 (75%) |  |
| N-Not Collected | 322 | 79 | 401 |  | 1 | 1 | 2 |  |
| Asthma Diagnosis at 3 or 4-year |  |  |  | **<0.001** |  |  |  | **<0.001** |
| Yes | 181 (15%) | 83 (9%) | 302 (14%) |  | 179 (49%) | 83 (17%) | 262 (31%) |  |
| No | 1033 (85%) | 807 (91%) | 1799 (86%) |  | 186 (51%) | 401 (83%) | 587 (69%) |  |
| N-Not Collected | 322 | 79 | 401 |  | 2 | 1 | 3 |  |
| Asthma Diagnosis at 10-year |  |  |  | 0.62 |  |  |  | **<0.001** |
| Yes | 201 (15%) | 101 (14%) | 302 (14%) |  | 114 (32%) | 60 (13%) | 174 (21%) |  |
| No | 1167 (85%) | 632 (86%) | 1799 (86%) |  | 238 (68%) | 411 (87%) | 649 (79%) |  |
| N-Not Collected | 168 | 236 | 404 |  | 15 | 14 | 29 |  |
| Gender |  |  |  | 0.91 |  |  |  | 0.26 |
| Female | 594 (49%) | 462 (48%) | 1056 (49%) |  | 154 (42%) | 223 (46%) | 377 (44%) |  |
| Male | 624 (51%) | 492 (52%) | 1116 (51%) |  | 213 (58%) | 261 (54%) | 474 (56%) |  |
| N-Not Collected | 318 | 15 | 333 |  | 0 | 1 | 1 |  |
| Lowest socioeconomic level |  |  |  | **<0.001** |  |  |  | **<0.001** |
| Yes | 73 (10%) | 340 (35%) | 413 (24%) |  | 32 (14%) | 179 (37%) | 211 (29%) |  |
| No | 650 (90%) | 623 (65%) | 1273 (76%) |  | 205 (86%) | 302 (63%) | 507 (71%) |  |
| N-Not Collected | 813 | 6 | 819 |  | 130 | 4 | 134 |  |
| Family history of asthma |  |  |  | **0.021** |  |  |  | 0.16 |
| Yes | 595 (47%) | 411 (42%) | 1006 (45%) |  | 186 (58%) | 255 (53%) | 441 (55%) |  |
| No | 660 (53%) | 558 (58%) | 1218 (55%) |  | 135 (42%) | 230 (47%) | 365 (45%) |  |
| N-Not Collected | 281 | 0 | 281 |  | 46 | 0 | 64 |  |
| Smoking During Pregnancy |  |  |  | 0.52 |  |  |  | 1 |
| Yes | 384 (25%) | 225 (24%) | 609 (25%) |  | 102 (28%) | 131 (28%) | 233 (29%) |  |
| No | 1137 (75%) | 712 (76%) | 1849 (75%) |  | 264 (72%) | 342 (72%) | 568 (71%) |  |
| N-Not Collected | 15 | 32 | 47 |  | 1 | 12 | 13 |  |
| Exclusively Breastfed > 3 mo |  |  |  | **<0.001** |  |  |  | **0.0032** |
| Yes | 419 (33%) | 191 (20%) | 800 (36%) |  | 81 (25%) | 80 (17%) | 269 (33%) |  |
| No | 865 (67%) | 759 (80%) | 1434 (64%) |  | 243 (75%) | 403 (83%) | 538 (67%) |  |
| N-Not Collected | 252 | 19 | 271 |  | 43 | 2 | 45 |  |

**The p-values are for the tests comparing the means (for continuous variables) or for comparing the distribution patterns of a categorical variable between the IOWBC and FAIR cohorts.**

*†Wheeze frequency was defined by a categorical scale; For IOWBC: 0 = 0 episodes in past 12 months, 1 = <3 episodes in past 12 months, 2 = 3 episodes in past 12 months, 3 = >3 episodes in past 12 months; For FAIR:0 = 0 episodes in past 12 months, 1 = <3 episodes in past 12 months, 2 = 4-12 episodes in past 12 months, 3 = >12 episodes in past 12 months*

**Table 2: Participant Characteristics by Early-Life Wheezing Cluster**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | |  | | Cluster 1 | | Cluster 2 | | | Cluster 3 | | Cluster 4 | | Cluster 5 | p-value |
| n (%) | | IOWBC | | 143  (39%) | | 77  (21%) | | | 29  (7.9%) | | 56  (15%) | | 62  (16.9%) |  |
|  | | FAIR | | 241  (49.7%) | | 69  (14.2%) | | | 37  (7.6%) | | 112  (23%) | | 26  (5.3%) |  |
| Variables Utilized in Cluster Analysis | | | | | | | | | | | | | | |
| Continuous Variables; Mean (SD) | | | | | | | | | | | | | | |
| Age of Wheeze Onset (yrs) | | IOWBC | | ***0.57*** (0.42) | | 0.73 (0.47) | | | 1.72 (0.73) | | **2.99** (0.62) | | 2.78 (0.67) | <0.001 |
|  | | FAIR | | 0.4 (0.2) | | ***0.3*** (0.2) | | | ***0.3*** (0.2) | | 1.9 (0.5) | | **2.2** (0.5) | <0.001 |
| Wheeze Frequency | | IOWBC‡ | | 0.13 (0.33) | | **2.68** *(0.47)* | | | 1.21 (0.62) | | ***0.30*** (0.46) | | 2.63 (0.49) | <0.001 |
|  | | FAIR§ | | ***1.0*** (0.0) | | ***1.0*** (0.0) | | | **2.2** (0.4) | | ***1.0*** (0.0) | | **2.2** (0.4) | <0.001 |
| Categorical Variables; % (n/N0) | | | | | | | | | | | | | | |
| Atopy | | IOWBC at 4-year | | ***9.3%***  (11/118) | | 31.9%  (23/72) | | | **68.0%**  *(17/25)* | | 17.4% (8/46) | | 40.0%  (22/55) | <0.001 |
|  | | FAIR at 3-year | | ***7%*** (12/172) | | 16%  (9/58) | | | 21%  (6/28) | | 13%  (11/84) | | **38%**  (8/21) | 0.0006 |
| Eczema | | IOWBC in first 4-years | | 22.4%  (32/143) | | ***31.2%***  (24/77) | | | **82.8%**  *(24/29)* | | 14.6% (8/55) | | 27.4%  (17/62) | <0.001 |
|  | | FAIR in first 3-years | | 63%  (151/241) | | ***61%***  (42/69) | | | **89%**  (33/37) | | 63%  (71/112) | | 68%  (17/25) | 0.0287 |
| Food Allergy | | IOWBC in first 4-years | | 11.9%  (17/143) | | 14.3%  (11/77) | | | **65.5%**  (19/29) | | ***10.9%***  (6/55) | | 12.9%  *(8/62)* | <0.001 |
|  | | FAIR in first 3-years | | ***4%***  (10/240) | | 12%  (8/69) | | | 16%  (6/37) | | ***4%***  (5/112) | | **19%**  (5/26) | 0.0019 |
| Rhinitis | | IOWBC in first 4-years | | ***2.1%***  (3/143) | | 13.0%  (10/77) | | | **48.3%**  (14/29) | | 5.5%  (3/55) | | 14.5%  *(9/62)* | <0.001 |
|  | | FAIR in first 3-years | | ***29%***  (69/240) | | 45%  (31/69) | | | **59%**  (22/37) | | 37%  (41/112) | | 50%  (13/26) | <0.001 |
| Additional Clinical Features, % (n/N­0)† | | | | | | | | | | | | | | |
| Nocturnal Symptoms | | IOWBC nocturnal symptom frequency *≥* 3 at 4-years | | 82.4%  (28/34) | | **95.9%**  (70/73) | | | 78.3% *(18/23)* | | ***69.2%*** (27/39) | | 89.7%  (52/58) | 0.0013 |
|  | FAIR sleep disturbance in 12 months at 3-years | | ***0%***  (1/241) | | 30%  (21/69) | | 41%  (15/37) | 14%  (16/112) | | **46%**  (12/26) | | <0.001 | | |
| Recurrent Chest Infections | IOWBC at 1-2 years | | 50.8%  (67/132) | | **53.9%**  (35/65) | | 32.0%  *(8/25)* | ***17.0%***  (8/47) | | 36.5%  (19/52) | | <0.001 | | |
| Wheeze in past 12 months | IOWBC at 4-year | | ***30.1%***  (43/143) | | **100%**  (77/77) | | 89.7%  (26/29) | 94.5%  (53/56) | | **100%**  *(62/62)* | | <0.001 | | |
|  | FAIR at 3-year | | **2%**  (4/241) | | 99%  (68/69) | | **100%**  (37/37) | 49%  (55/112) | | **100%**  (26/26) | | <0.001 | | |
| Asthma Diagnosis | IOWBC at 4-year | | ***13.4%***  (19/142) | | 89.6%  (69/77) | | 58.6%  (17/29) | 29.1%  (16/55) | | **93.6%**  *(58/62)* | | <0.001 | | |
|  | FAIR at 3-year | | ***3%***  (7/241) | | 34%  (23/68) | | **62%**  (23/37) | 17%  (19/112) | | 42%  (11/26) | | <0.001 | | |
| Asthma Treatment | IOWBC at 4-year | | ***17.5%***  (25/143) | | **77.9%**  (60/77) | | 48.3%  (14/29) | 51.8%  (29/56) | | 72.6%  *(45/62)* | | <0.001 | | |
|  | FAIR at 3-year | | ***9%***  (22/237) | | 70%  (46/66) | | **80%**  (28/35) | 38%  (42/111) | | 76%  (19/25) | | <0.001 | | |

**For continuous variables, the Analysis of Variance (ANOVA) was used to compare means across the clusters. For categorical variables, the Chi-square tests were applied to compare the proportions across the clusters. Significantly different clusters are highlighted in bold (highest)/ bold italics (lowest), and are identified based on pair-wise t-tests (continuous variables) or proportion tests (categorical variables). Definitions of each variable are detailed in supplemental table E1.**

†For each variable, the number of subjects in total, denoted by N0, can be different from the size of each cluster (denoted by N).

*‡ Wheeze frequency was defined by a categorical scale; 0 = 0 episodes in past 12 months, 1 = <3 episodes in past 12 months, 2 = 3 episodes in past 12 months, 3 = >3 episodes in past 12 months*

*§ Wheeze frequency was defined by a categorical scale; 0 = 0 episodes in past 12 months, 1 = <3 episodes in past 12 months, 2 = 4-12 episodes in past 12 months, 3 = >12 episodes in past 12 months*

**Table 3: Association of Early-Life Factors with Early-Life Wheeze Clusters in FAIR Cohort**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Overall | Total Cluster Cohort | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Cluster 5 | p-value |
| Continuous Variables; Mean (SD) | | | | | | | | |
| Birthweight (kg) | 3.4 (0.6) | 3.4 (0.6) | 3.4 (0.6) | 3.4 (0.6) | 3.3 (0.5) | 3.3 (0.5) | 3.3 (0.7) | 0.60 |
| BMI z-score at 2 years | 0.2 (1.1) | 0.3 (1.1) | 0.1 (1.0) | 0.8 (1.2) | **0.9 (2.0)** | 0.4 (0.8) | ***-0.3 (0.9)*** | **0.017** |
| Categorical Variables; % (n/N0)† | | | | | | | | |
| Male | 52%  (492/954) | 54%  (261/484) | 57%  (138/241) | 54%  (37/68) | 57%  (21/37) | 44%  (49/112) | 62%  (16/26) | 0.17 |
| Maternal Smoking in Pregnancy | 24%  (225/937) | 28%  (131/485) | 27%  (64/237) | 41%  (27/66) | 31%  (11/36) | 23%  (25/109) | 16%  (4/25) | 0.064 |
| Family History of Asthma | 42%  (411/969) | 53%  (255/485) | 48%  (116/241) | 65%  (45/69) | **76%**  **(28/37)** | 51%  (57/112) | **35%**  **(9/26)** | **0.0014** |
| Annual household income less than 50K | 35%  (340/963) | 37%  (179/481) | 37%  (88/238) | 51%  (35/69) | 49%  (18/37) | 29%  (33/112) | 20%  (5/25) | **> 0.001** |
| Vaginal Delivery | 78%  (755/966) | 74%  (359/484) | 74%  (178/240) | 77%  (53/69) | 70%  (26/37) | 75%  (84/112) | 69%  (18/26) | 0.92 |
| Exclusively Breast-fed > 3 months | 40%  (381/950) | 39%  (188/483) | 39%  (93/241) | 32%  (22/69) | 32%  (12/37) | 45%  (50/111) | 44%  (11 /25) | 0.39 |
| Pet Exposure | 58%  (563/969) | 60%  (293/485) | 59%  (141/241) | 70%  (48/69) | 68%  (25/37) | 59%  (66/112) | 50%  (13/26) | 0.30 |

**The p-values are for the tests comparing the means (for continuous variables) across the 5 clusters or for comparing the distribution patterns of a categorical variable in each cluster. For continuous variables ANOVA was used and, for categorical variables, chi-square tests were implemented. #Significantly differing clusters are highlighted in bold (highest) and bold italics (lowest) (significance = p<0.05), and are identified based on pair-wise proportion tests (categorical variables).**

†For each variable, the number of subjects in total, denoted by N0, can be different from the size of each cluster (denoted by N).

**Figure Legends**

**Figure 1. Longitudinal Asthma Characteristics of the Early Childhood Wheeze Clusters.**

a) Percent of patients by cluster with asthma diagnosis at 10-year. Asthma diagnosis is defined as a history of physician diagnosed asthma plus wheezing in the past 12 months at 10 years of age. b) Percent of patients by cluster with current wheezing at 10 years of age. Current wheezing is defined as presence of any wheezing in the past 12 months at 10 years of age.

**Figure 2. Summary of Early-Life Wheezing Cluster Characteristics and 10-year Asthma Outcome**

Proportion of patients within each of the clusters among a) combined clusters, b) IOWBC and c) FAIR. Darker shaded area indicates percentage of patients within each cluster who met criteria for a diagnosis of asthma at age 10-years.

**a)**

**b)**

**Figure 1. Longitudinal Asthma Characteristics of the Early Childhood Wheeze Clusters.**

a) Percent of patients by cluster with asthma diagnosis at 10-year. Asthma diagnosis is defined as a history of physician diagnosed asthma plus wheezing in the past 12 months at 10 years of age. b) Percent of patients by cluster with current wheezing at 10 years of age. Current wheezing is defined as presence of any wheezing in the past 12 months at 10 years of age.

a) Combined Cohorts Chart, sunburst chart

Description automatically generated

Darker shaded region represents asthma prevalence at 10-years within each cluster

b) IOWBC

Chart, sunburst chart

Description automatically generated

c) FAIR

Chart, sunburst chart

Description automatically generated

**Figure 2. Summary of Early-Life Wheezing Cluster Characteristics and 10-year Asthma Outcome**

Proportion of patients within each of the clusters among a) combined clusters, b) IOWBC and c) FAIR. Darker shaded area indicates percentage of patients within each cluster who met criteria for a diagnosis of asthma at age 10-years.

**Appendix A: Clustering Methods**

Cubic clustering criterion (CCC)and pseudo F statistic (where larger values indicate better cluster resolution) were used to determine number of clusters. Both statistics are available in the output of PROC FASTCLUS. The CCC criterion is a measure of cluster deviation from the distribution expected if data points were drawn from a uniform distribution. The pseudo F statistic captures cluster homogeneity and is a ratio of the mean sum of squares between groups to the mean sum of squares within group. Larger CCC and pseudo F indicate a better cluster resolution. In our analysis, we considered different numbers of clusters, and then for each case obtained values for CCC and pseudo F. In general, the patterns of CCC and pseudo F were quadratic with respect to the number of clusters. Pseudo F reached the largest value at 4 clusters while CCC at 5 clusters. With 5 clusters, the value of pseudo F is still high. Taking into account both pseudo F and CCC, we decided to use 5 clusters to describe the underlying clusters (Table E3).

**Appendix B: Supplemental Data**

**Table E1: Comparison of Cohort Measures**

|  |  |  |
| --- | --- | --- |
| Variable | IOWBC | FAIR |
| Presence of wheezing in first 3-4 years | Any recorded wheeze within past 12 months at 1- or 2-years or at 4-years. | Any wheezing before 1-year in the absence of infection or any wheezing within the last 12 months at 2- or 3-years |
| Wheeze Frequency at 3- or 4- years | Wheeze frequency over the past 12 months at 4-year. Frequency was defined by a categorical scale; 0 = 0 episodes in past 12 months, 1 = <3 episodes in past 12 months, 2 = 3 episodes in past 12 months, 3 = >3 episodes in the past 12 months | Wheeze frequency over the past 12 months at 3-year. Frequency was defined by a categorical scale; 0 = 0 episodes in past 12 months, 1 = <3 episodes in past 12 months, 2 = 4-12 episodes in past 12 months, 3 = >12 episodes in past 12 months |
| Atopy at 3- or 4-years | Positive skin prick testing at 4-year to dust mite, cat, dog, *Alternaria alternata*, *Cladosporium herbarium*, grass pollen mix, or tree pollen mix | Positive skin prick testing at 3-year to house dust mite, cat, or grass pollen |
| Eczema in first 3-4 years | Chronic or chronically relapsing itchy rash with characteristic morphology and distribution at 1,2 or 4-years. | Parental report a diagnosis of eczema up to 3-years. |
| Food allergy in first 3-4 years | History of vomiting, diarrhea, colic or rash within 4 hours of ingestion of a particular food on at least 2 occasions recorded at 1,2, or 4-years. | Positive food challenge (open or double blinded) and/or clinical history supported by food sensitization by 3-years |
| Rhinitis in first 3-4 years | Recurrent nasal discharge or blockage with attacks of sneezing and itchy eyes at 4-years. | Symptoms of sneezing, nasal discharge or congestion within the last 12 months at 3-year. |
| Asthma diagnosis at 3- or 4-years | Combination of physician diagnosed asthma plus 3 or more separate episodes of wheezing and/or cough at 4-years. | Parental report of a diagnosis of asthma up to 3 years of age. |
| Asthma treatment at 3- or 4-years | Any medications prescribed for asthma at 4 years of age. | Any treatment given for wheezing or asthma up to 3 years of age |
| Asthma diagnosis at 10-years | A history of physician diagnosed asthma plus wheezing in the past 12 months at 10 years of age | Parental report of physician diagnosed asthma plus wheezing in the past 12 months at 10 years of age |
| Asthma treatment at 10-years | Any medication prescribed for asthma at 10-years of age | Any treatment ever given for wheezing or asthma by 10-years of age |
| Wheezing at 10-years | Wheezing in the past 12-months at 10-years | Wheezing in the past 12-months at 10-years |
| Lower social economic level | Social Class IIIM-V of the Registrar General’s classification | Annual household income less than 50K |
| Family History of Asthma | Either parental or sibling history of asthma. | Either parental or sibling history of asthma. |
| Maternal Smoking in Pregnancy | Any history of maternal smoking during pregnancy | Any history of maternal smoking during pregnancy |
| Breastfeeding | Exclusively breastfed for >3 months | Exclusively breastfed for >3 months |

**Table E2:** **Association of Early-Life Factors with Early-Life Wheeze Clusters in IOWBC**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Overall | Total Cluster Cohort | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Cluster 5 | p-value |
| Continuous Variables; Mean (SD); N0† | | | | | | | | | |
| Birthweight (kg) | | 3.39 (0.54); 1511 | 3.39 (0.57); 364 | 3.39 (0.50);  140 | 3.35 (0.58); | 3.46 (0.67); | 3.46 (0.65);  56 | 3.35 (0.57);  62 | 0.72 |
| BMI z-score at 2-year | | 0.19 (1.21); 393 | 0.26 (1.37); 106 | 0.267 (1.20); 50 | 0.066 (1.58); 23 | 0.857 (1.01); 9 | 0.0430 (1.99); 12 | 0.401 (1.19); 12 | 0.64 |
| Categorical Variables; % (n/N0)† | | | | | | | | | |
| Male | | 48.8%  (594/1218) | 42.0%  (154/367) | 38.5%  (55/143) | 48.1%  (37/77) | 37.9%  (11/29) | 44.6%  (25/56) | 41.9%  (26/62) | 0.69 |
| Maternal Smoking in Pregnancy | | 25.25%  (384/1521) | 27.9%  (102/366) | 31.7%  (45/142) | 31.2%  (24/77) | 20.7%  (6/29) | 19.6%  (11/56) | 25.8%  (16/62) | 0.38 |
| Family History of Asthmaa | | 47.4%  (595/1255) | 58.0%  (186/321) | 51.6%  (63/122) | 67.2%  (45/67) | 59.3%  (16/27) | 54.2%  (26/48) | 63.2%  (36/57) | 0.26 |
| Social Class IIIM-V of the Registrar General’s classification | | 54.6%  (486/890) | 59.9%  (142/237) | 64.2%  (61/95) | 67.4%  (33/49) | 42.1%  (8/19) | 53.5%  (23/43) | 54.8%  (17/31) | 0.25 |
| Vaginal Delivery | | 83.2%  (1030) | 87.5%  (252/309) | 82.2%  (97/118) | 82.5%  (52/63) | 82.4%  (23/28) | 73.1%  (38/52) | 87.4%  (42/48) | 0.45 |
| Exclusively Breast-fed > 3 months | | 32.6%  (419/1284) | 25.0%  (81/324) | 26.4%  (34/129) | 19.1%  (13/68) | 29.6%  (8/27) | 22.2%  (10/45) | 29.1%  (16/55) | 0.67 |
| Pet Exposure | | 52.4%  (720/1375) | 53.2%  (185/348) | 52.5%  (73/139) | 57.5%  (42/73) | 44.4%  (12/27) | 66.0%  (33/50) | 42.4%  (25/59) | 0.11 |

**The p-values are for the tests comparing the means (for continuous variables) across the 5 clusters or for comparing the distribution patterns of a categorical variable in each cluster. For continuous variables ANOVA was used and, for categorical variables, chi-square tests were implemented.**

†For each variable, the number of subjects in total, denoted by N0, can be different from the size of each cluster (denoted by N)

**Table E3. Statistics assessing the quality of clusters.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Number of clusters | | | | |
| Statistics | 2 | 3 | 4 | 5 | 6 |
| Pseudo F | 103.59 | 159.34 | 210.75 | 185.33 | 155.32 |
| CCC | -5.41 | -3.71 | 5.54 | 6.58 | 5.27 |