
Early version, also known as pre-print

Link to published version (if available): 10.1513/AnnalsATS.201811-837OC

Link to publication record in Explore Bristol Research

PDF-document

*University of Bristol - Explore Bristol Research*

*General rights*

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: http://www.bristol.ac.uk/pure/user-guides/explore-bristol-research/ebr-terms/
Distinguishing wheezing phenotypes from infancy to adolescence: a pooled analysis of five birth cohorts

Ceyda Oksel¹, Raquel Granell², Sadia Haider¹, Sara Fontanella¹, Angela Simpson³, Steve Turner⁴, Graham Devereux⁵, Syed Hasan Arshad⁷,⁸, Clare S Murray³, Graham Roberts⁶,⁷,⁸, John W. Holloway⁶,⁷, Paul Cullinan⁹, John Henderson², Adnan Custovic¹⁰ on behalf of STELAR¹⁰ and Breathing Together¹¹ investigators

¹Equal contribution

¹Department of Paediatrics, Imperial College London, UK

²MRC Integrative Epidemiology Unit, Population Health Sciences, Bristol Medical School, University of Bristol, UK

³Division of Infection, Immunity and Respiratory Medicine, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre, UK

⁴Child Health, University of Aberdeen, Aberdeen, UK

⁵Liverpool School of Tropical Medicine

⁶Human Development and Health, Faculty of Medicine, University of Southampton, Southampton, UK

⁷David Hide Asthma and Allergy Research Centre, Isle of Wight, UK

⁸Respiratory Biomedical Research Unit, University Hospitals Southampton NHS Foundation Trust, Southampton, UK

⁹National Heart and Lung Institute, Imperial College, London, UK

Correspondence to: Professor John Henderson, Department of Population Health Sciences, Bristol Medical School, University of Bristol, UK; Email: A.J.Henderson@bristol.ac.uk

Contributions: C.O., A.C., R.G and J.H conceived and planned the study, and wrote the manuscript. C.O and R.G analysed the data. S.H., S.F., A.S., S.T., G.D., S.H.A., C.S.M, G.R, J.W.H. and P.C. contributed to the interpretation of the results. All authors provided critical feedback and helped shape the research, analysis and manuscript.

Funding: This analysis is funded through the Wellcome Trust Strategic Award 108818/15/Z. STELAR consortium is funded by the UK Medical Research Council (MRC) Grants G0601361 and MR/K002449/1. The UK Medical Research Council and the Wellcome Trust (Grant ref: 102215/2/13/2) and the University of Bristol provide core support for ALSPAC.

Short running head: Distinguishing wheezing phenotypes: a pooled analysis of five birth cohorts

Descriptor: 1.17

Abstract word count: 250

Word count: 35022
At a Glance Commentary:

Scientific knowledge on the subject: Clustering of wheezing phenotypes has provided insights into the heterogeneity of childhood wheeze but variation between studies means uncertainties remain about the number and characteristics of discrete phenotypes.

What this study adds: This is the first study to jointly model data on wheezing from birth to adolescence in multiple population-based birth cohorts and investigate the nature of within-class heterogeneity by looking at individual patterns of wheeze within different thresholds of class assignment. By pooling the cohorts and restricting our analyses to participants with complete data, we obtained a better representation of the latent structure, while maintaining the power to detect less prevalent phenotypes and making meaningful analyses of their associations with risk factors and outcomes. Obstructive pattern of lung function, with significantly diminished FEV$_1$ and FEV$_1$/FVC ratio, was a feature of all wheeze phenotypes. This suggests that the notion that episodic wheeze in early life has a benign long-term prognosis may not be true for a proportion of transient wheezers. This is particularly important in the context of findings that low FEV$_1$ at its physiological plateau in early adult age is important in the genesis of chronic obstructive pulmonary disease, early cardio-vascular mortality, and is a marker of premature death from all causes.
ABSTRACT

Rationale: Pooling data from multiple cohorts and extending the time-frame across childhood should minimize study-specific effects, enabling better characterization of the childhood wheezing.

Objective: To analyze wheezing patterns from early childhood to adolescence using combined data from five birth cohorts.

Methods: We used latent class analysis to derive wheeze phenotypes among 7719 participants from five birth cohorts with complete report of wheeze at five time-periods. We tested the association of derived phenotypes with late asthma outcomes and lung function, and investigated the uncertainty in phenotype assignment.

Main Results: We identified five phenotypes: Never/Infrequent wheeze (52.1%), Early-onset pre-school remitting (23.9%), Early-onset mid-childhood remitting (9%), Persistent (7.9%) and Late-onset wheeze (7.1%). Compared to the Never/infrequent wheeze, all phenotypes had higher odds of asthma and lower FEV₁ and FEV₁/FVC in adolescence. The association with asthma was strongest for Persistent wheeze (adjusted odds ratio 56.54, 95%CI 43.75–73.06). We observed considerable within-class heterogeneity at individual level, with 913 (12%) children having low membership probability (<0.60) of any phenotype. Class membership certainty was highest in Persistent and Never/infrequent, and lowest in Late-onset wheeze (with 51% of participants having membership probabilities<0.80). Individual wheezing patterns were particularly heterogeneous in Late-onset wheeze, while many children assigned to Early-onset pre-school remitting class reported wheezing at later time points.
Conclusions: All wheeze phenotypes had significantly diminished lung function in school-age, suggesting that the notion that early-life episodic wheeze has a benign prognosis may not be true for a proportion of transient wheezers. We observed considerable within-phenotype heterogeneity in individual wheezing patterns.

KEYWORDS: wheezing phenotypes, childhood, adolescence, latent class
INTRODUCTION

Wheezing is one of the most common symptoms in infants and young children. Childhood wheezing is highly heterogeneous(1-4), and distinguishing subtypes of wheezing and their underlying mechanisms is a prerequisite to develop interventions to target children whose wheezing will persist, while avoiding overtreatment of transient wheeze.(5)

Over the last two decades, substantial effort has been devoted to understanding the heterogeneity of childhood wheezing.(5-9) Population-based birth cohorts, with prospectively collected information free from recall bias, are the optimal data sources for understanding temporal patterns of symptoms.(6) In a pioneering study using clinical insights about the age of onset, progression and remission of wheezing, Martinez et al described three mutually-exclusive wheezing phenotypes (transient early, late-onset and persistent).(1) These findings were confirmed in several independent cohorts.(4, 10, 11) Subsequently, the approach to discovering longitudinal wheezing phenotypes was extended to data-driven approaches such as the latent class analysis (LCA), which suggested the existence of one or two(12) further intermediate phenotypes. However, although phenotypes derived using LCA in different studies are usually designated with the same names share the same name, they often differ in temporal trajectories, distributions within a population and associated risk factors.(9) This may be in part a consequence of the study characteristics, including sample size and the timing and frequency of data collection.(8) A further reason may arise from the heterogeneity between children within each of the derived phenotypes, which has seldom been investigated. For example, in most LCA models, there will be individuals who cannot be classified with a high degree of certainty. Latent class model estimate class membership probabilities (each individual’s probability of
membership of each class), which are traditionally used to assign individuals to the latent class with the highest posterior probability. (13-15) Such maximum-probability phenotype assignment is equivalent to fixing individuals’ membership probabilities of their highest phenotype to 1 and all others to 0, and may introduce within-class heterogeneity. Therefore, models that are weighted for the probability of each subject belonging to each phenotype should be used to compensate for uncertainty in phenotype assignment. (3, 16)

We propose that pooling data from multiple birth cohorts and extending the time frame from early childhood to adolescence should minimize study-specific effects, thereby enabling us to better understand the true variation in the natural history of wheezing, as well as the individual differences within apparently homogeneous phenotypes. We therefore proceeded to analyze wheezing patterns from early childhood to adolescence/young adulthood using combined data from five birth cohorts in the UK STELAR consortium (6), and their associations with early-life risk factors, asthma-related outcomes and lung function measures in late childhood. We then investigated the extent of uncertainty in phenotype assignment, and individual variations in wheezing within and across phenotypes.

We acknowledge that the wheeze classes discovered using data-driven approaches are not observed but are latent by nature, and ideally should not be referred to as “phenotypes” (i.e. observable characteristics); however, as the term “phenotype” has been used in this context for over a decade, we will maintain this nomenclature in this manuscript.
METHODS

Study design, setting and participants

The STELAR consortium(6) brings together five UK population–based birth cohorts: Avon Longitudinal Study of Parents and Children (ALSPAC)(17), Ashford(18) and Isle of Wight (IOW)(4, 19) cohorts, Manchester Asthma and Allergy Study (MAAS)(20) and the Aberdeen Study of Eczema and Asthma to Observe the Effects of Nutrition (SEATON)(21). The cohorts are described in detail in the online supplement. All studies were approved by research ethics committees. Informed consent was obtained from parents, and study subjects gave their assent/consent when applicable. Data were imported into a web–based knowledge management platform, Asthma eLab (www.asthmaelab.org). (6)

Data sources and definition of outcomes

Validated questionnaires were completed on multiple occasions from infancy to adolescence (14 time points in ALSPAC over 16.5 years, 7 in ASHFORD over 14 years, 6 in MAAS over 16 years, 6 in SEATON over 14 years, and 5 in IOW over 18 years). The cohort–specific time points and sample size included in this analysis are shown in Table E1. We used data on current wheeze collected at proximate time-periods across cohorts: infancy (½–1 year); early childhood (2–3 years); pre–school/early school age (4–5 years); middle childhood (8–10 years); and adolescence (14–18 years). Current wheeze was defined as a positive response to the question “Has your child had wheezing or whistling in the chest in the last 12 months?”. Definitions of other variables are provided in the online supplement (Table E2).
Skin prick testing was carried out at age 4 years in IOW, age 5 years in SEATON, ASHFORD and MAAS, and 7.5 years in ALSPAC. Allergic sensitization was defined as a wheal diameter of 3mm greater than the negative control to one or more common allergens.

We performed spirometry and recorded forced expiratory volume in the first second (FEV₁) and forced vital capacity (FVC). To reduce variability, we restricted our analyses to lung function measured at comparable ages (8½ years in ALSPAC, 10 years in SEATON and IOW, and 11 years in MAAS).

**Statistical analysis**

A detailed description of statistical analysis is presented in the online supplement. Briefly, we used LCA to identify longitudinal trajectories of wheeze among 7719 children with complete data on wheezing at five time periods. Cohort ID was included as an additional covariate by transforming the 5-category variable into a set of four dummy variables. The optimal number of classes was selected based on the Bayesian information criterion (BIC) and interpretability. We repeated analyses among 15,941 children by retaining children with at least two observations, and examined the stability of allocation of 7719 children with complete data to their most probable class when using a larger but incomplete data set. We also assessed the stability of the optimal number of classes excluding the largest cohort, ALSPAC. To understand how children are reassigned to different phenotypes when the number of phenotypes increases, we compared LCA models constrained to 4 classes to models with 5 classes.

To validate the derived wheezing phenotypes, we then examined the relationships of the derived phenotypes with asthma and asthma medication use at the latest follow up, and lung
function (Z-scores for adjusted FEV$_1$, FVC and FEV$_1$/FVC) using logistic regression models weighted for the probability of each individual belonging to each phenotype. Models were adjusted for potential confounders, including allergic sensitization (mid-childhood), maternal history of asthma/allergy, maternal smoking and low birth weight. We used weighted multinomial logistic regression models to ascertain early-life risk factors associated with each phenotype; results are reported as relative risk ratios (RRR) with 95% confidence intervals (CI).

We then investigated individual variations in wheezing within and across phenotypes. We defined categories of class-membership certainty based on class membership probabilities (high≥0.80, medium=0.60-0.80, low<0.60) to elicit the extent of within-class homogeneity in different certainty thresholds. Analyses were carried out using Mplus 8, R (http://www.r-project.org/) and Stata 14 (StataCorp, College Station, Tex).
RESULTS

Characteristics of the study population

Of 7719 children with complete data on wheezing, 50.3% were male, 14.1% had current asthma and 13.4% reported using asthma medication at the last available follow up. Mean (SD) birth weight in the combined dataset was 3.44 kg (0.53), with 314 (4.1%) children having a low birth weight (<2.5kg); mean maternal age at delivery was 29.1 years (SD 4.64), and 1198 (16%) mothers reported smoking during pregnancy (Table 1).

Wheeze phenotypes from infancy to adolescence and transitions in phenotype membership

A five-class solution was selected as the optimal model based on statistical fit (Table E3, see online supplement for more details on model selection). Trajectories of wheeze for each class are shown in Figure 1. Based on the onset and duration of wheeze, the classes were labeled as: (1) Never/Infrequent wheeze (52.1%); (2) Early–onset pre–school remitting wheeze (23.9%), with 50% to 60% prevalence of wheeze during infancy, decreasing to 20% around early–childhood and to less than 10% afterwards; (3) Early–onset middle–childhood remitting wheeze (9%), with early onset and peak prevalence (~75%) in pre/early–school age, decreasing to 23% in mid–childhood, and diminishing further by adolescence; (4) Persistent wheeze (7.9%) with 53% wheeze prevalence during infancy, and an increasing prevalence thereafter to a high prevalence of ~80%; (5) Late–onset wheeze (7.1%) with a low prevalence of 27% until pre/early–school age, increasing rapidly to a peak prevalence of 74% in adolescence. Characteristics of children across different phenotypes are presented in Table 1. LCA of 4 cohorts with complete data (excluding the largest
cohort, ALSPAC) provided very similar results, as did the analysis among 15,941 children with at least two observations (Table E3, Figure E1).

Figure 2a shows the changes in the allocation of 7719 individuals with complete data from the model using only children with data on wheezing at all five time periods to their most probable class when using a larger data set including 15,941 individuals with at least two observations. The optimal model solution in the model using 15,941 children remained 5 classes (Table E3, Figure E1), and was very similar to that derived from a complete data set.

Never/infrequent and Persistent phenotype assignments were very stable, but there were considerable transitions between other classes (e.g. 39% of children assigned to the Early-onset middle-childhood remitting wheeze using complete data were assigned to Persistent wheeze in the model with incomplete data).

To understand how class allocation changes in models with increasing number of phenotypes, we compared LCA models with 4 and 5 phenotypes using complete data (Figure 2b). Never/infrequent and Late-onset phenotypes were similar in both models. With the addition of a fifth phenotype, Transient–early wheeze divided into two remitting classes (Pre−school and Mid−childhood resolution), while a third of children from the Persistent wheeze in a 4−class solution transitioned, mainly to Early−onset mid−childhood remitting wheeze.

**Wheeze phenotypes, family history, early−life factors, environmental exposures and atopy**

Results of univariate analyses among participants with complete and incomplete data are shown in Tables E4 and E5. Table E6 shows results of multivariable logistic regression models weighted for the probability of each individual belonging to each phenotype, using the Never/infrequent wheeze as the reference. Males had a higher risk of developing the three Early−onset
phenotypes, but not Late−onset wheezing. Maternal asthma was and smoking were associated with all four phenotypes, with the strongest association being that of maternal asthma with Persistent wheezing (RRR 3.12, 95%CI 2.51−3.89, p<0.0001). Compared to never/infrequent wheeze, all wheezing phenotypes except Early−onset pre−school remitting wheeze, showed significant associations with maternal smoking during pregnancy. There was little evidence of association between wheeze phenotypes and pet ownership or paternal smoking. Paternal asthma was significantly associated with persistent wheeze in both genders, but it was a predictor of late−onset wheeze in males only (Table E4).

The strongest association with allergic sensitization (ages 4 to 7.5 years) was observed for Persistent wheeze (RRR 5.12, 95%CI 4.04−6.47, p<0.0001). Two early−onset remitting classes differed substantially in their association with sensitization: there was little evidence of association between sensitization and Pre−school remitting wheeze, but clear evidence of an association between sensitization and Mid−childhood remitting wheeze (RRR 1.61, 95%CI 1.24−2.11, p<0.0001).

Asthma and lung function

Associations of wheeze phenotypes with current asthma and asthma medication use at the last follow−up are shown in Table 2. Compared to Never/Infrequent wheeze, all four wheeze phenotypes were associated with higher asthma prevalence. This association was weakest for Early−onset pre−school remitting wheeze (RRR 2.28, 95%CI 1.76, 2.96, p<0.0001) and strongest for Persistent wheeze (RRR 56.54, 95%CI 43.75, 73.06, p<0.0001). We obtained similar results in the analysis of 8223 children with incomplete reports of wheezing (excluding 7719 children with
complete reports at five time-points, Table E7). Characteristics of children with complete report of wheeze and those with missing data are shown in Table E8.

Associations of wheeze phenotypes with lung function in middle childhood (8.5 to 11 years) is shown in Table 3. FEV$_1$ and FEV$_1$/FVC ratio were significantly lower in all wheeze phenotypes compared to Never/infrequent class. Association was particularly strong for Persistent wheeze (Z scores $-0.60$, 95%CI $-0.70$, $-0.50$, p<0.0001). There was no evidence of differences in FVC between the classes.

**Individual variation within and across wheezing phenotypes**

Figure 3 shows the presence/absence of wheeze at five time periods amongst each individual participant across latent classes, stratified by cohort and class membership certainty. More than 80% of individuals in Never/infrequent and Persistent wheeze classes were assigned with a high degree of certainty (membership probability >0.80), while class membership certainty was the lowest for Late-onset wheeze, with 51% of its members having posterior probabilities <0.80. Sorting individuals in the vertical space and stratifying them by class assignment probability revealed fairly-consistent between-individual patterns when assignment probability was ≥0.80, but heterogeneity increased as assignment became less certain (<0.80). Individual wheezing patterns were particularly heterogeneous in Late-onset wheeze, whereby participants reported wheezing only once, up to three times persistently, or intermittently. Of 113 children who were assigned to the Early-onset pre-school remitting class, but reported current asthma at the last follow up, 78 (69%) reported wheezing at later time points.
Over a tenth of individuals (913/7719, 12%) had low posterior probabilities (<0.60) of membership to any class (Figure 3). Figure E2 shows a heat map of each individual’s posterior class membership probabilities, lying between 0 (low) and 1 (high). Inspection of the figure reveals that there are individuals whose patterns do not fit well within the assigned phenotype (for example, some have a ~0.40 probability of belonging to Never/infrequent wheeze and ~0.40 probability of belonging to Early-onset pre-school remitting wheeze).
DISCUSSION

By jointly modelling data on parent/self-reported wheezing from birth to adolescence in five population-based birth cohorts, we identified five distinct phenotypes. In addition to never/infrequent wheezing, the latent structure of wheezing disorders from infancy to adolescence comprised two persisting (Early-onset and Late-onset) and two remitting (Pre-school remitting and Mid-childhood remitting) phenotypes. The model assigned 33% of children to remitting (24% remitting in Pre-school, and 9% in Mid-childhood), and 15% to persisting phenotypes (7.9% with onset in early life, and 7.1% with onset in later childhood). By pooling the cohorts and restricting our analyses to participants with complete data, we obtained a better representation of the latent structure, while maintaining the power to detect less prevalent phenotypes and making meaningful analyses of their associations with risk factors and outcomes. This allowed us to determine that an obstructive pattern of lung function, with significantly diminished FEV₁ and FEV₁/FVC-ratio, was a feature of all four wheeze phenotypes. Differing associations between wheeze phenotypes and asthma in adolescence confirmed the validity of derived phenotypes.

To our knowledge, this is the first study to investigate the nature of within-phenotype heterogeneity by looking at individual patterns of wheeze within different thresholds of class assignment. In the context of wheeze phenotyping, data driven techniques such as LCA have been used to discover subgroups within the study population which were presumed to be homogeneous. However, we have demonstrated that this may not always be the case. A consistent finding across all five cohorts was that a substantial number of children were classified imprecisely and did not follow wheeze patterns suggested by the label ascribed to the class when
the individual’s likelihood of belonging to the assigned class is low (was <0.80). Moreover, the precision with which children were assigned to a “label” varied across different classes. The greatest level of uncertainty in assignment was for the Late-onset class. This within-class heterogeneity may, in part, be responsible for a lack of consistent associations of phenotypes with risk factors reported in previous studies.(22)

Limitations and strengths

One limitation of this multi-cohort study is the heterogeneity between cohorts that were integrated retrospectively (e.g. in recruitment criteria, data collection time and intervals, question types and wording, and tools for collecting data). We observed some level of variation in the proportion of children allocated to each phenotype in different cohorts. For example, ASHFORD had the lowest prevalence of Late-onset wheeze and highest prevalence of Early-onset remitting wheeze, while IOW had the lowest prevalence of Early-onset pre-school remitting wheeze and the highest prevalence for Late-onset wheeze. We addressed this limitation by including ‘cohort site’ as a covariate for determining phenotypes.

The integration of cohorts is a difficult task requiring a significant investment in a knowledge management platform for managing the data, metadata and collaborative analyses, mapping the study variables across the cohorts and data harmonization, as well as protocol standardization to ensure data quality and remove any inconsistencies from various data sources. STELAR cohorts have harmonized data collection tools and procedures, and have created a secure web-based research environment to support consistent recording, description and sharing of data, statistical methods and findings across the cohorts(6), which facilitated current analyses.
One of the advantages of our multi-cohort approach is that individual studies that might not provide conclusive evidence to make inference about the general population due to cohort-specific effects and biases can contribute to revealing a more accurate picture when integrated together. The integration of five cohorts and their pooled analysis enhanced the credibility and generalizability of the phenotyping results to the UK population. A further advantage is to minimize the study-specific biases (including cohort-specific effects, attrition effects, different participant-recruitment strategies, and geographical factors) affecting the certainty of allocation of individuals to each latent class, while maximizing the benefits of individual cohort studies (e.g. potentially important risk factors and outcomes are captured in some, but not all cohorts). Another strength of pooling cohort data is that a multi-cohort design allowed us to analyze a large sample with complete data on wheeze from birth to adolescence, thus increasing statistical power to detect less prevalent phenotypes.

**Interpretation of findings**

Several birth cohorts have investigated wheeze phenotypes in childhood.(1-4, 10-12, 16, 23-26) The number of derived phenotypes varied by study, with four common classes identified in all cohorts: Never/infrequent wheeze, Transient–early, Late–onset and Persistent wheeze. In our joint analysis of five cohorts, a four-class solution identified the same classes, while the optimum solution included a fifth phenotype, Early–onset mid–childhood remitting wheeze, which was previously observed in ALSPAC(3, 12, 16) but not in other independent cohort.(3) School–age onset persisting wheeze that was previously identified as a sixth phenotype in ALSPAC(12, 16) was not detected in our joint analysis. One possible explanation could be that we have focused on common ages that are approximately shared across all cohorts to ensure
compatibility, resulting in the exclusion of potentially informative time points for identifying additional phenotypes(8). An alternative explanation is that some phenotypes may arise as an artefact of the analysis (e.g. due to the increasing data collection frequency).

When using 15,941 children with at least two observations of wheeze, individual allocation to Persistent and Never/infrequent phenotypes were consistent with the analysis using 7719 complete cases, but there was a considerable transition in other classes (e.g. almost 40% of children in Early-onset middle-childhood remitting class from complete data switched to Persistent wheeze). This finding raises questions about the confidence with which allocation to early-onset remitting classes can be applied in cohorts with considerable missing data. The analyses to understand how children’s class allocations changed from one phenotype to another in models with increasing number of classes have shown that when moving from a four- to five-class solution, Transient early wheeze was divided into two remitting classes. Pre-school remitting phenotype included children who were assigned to Transient early and Never/infrequent wheeze, while Mid-childhood remitting phenotype was formed by children assigned to Transient early and Persistent wheeze in a four-class model. These newly-formed remitting phenotypes showed different associations with risk factors, lung function and later asthma and lung function, suggesting that they may be distinct entities, or that phenotype allocation also reflects wheeze severity.

We have extended previous observations of the association between Persistent and Late-onset wheezing with atopy and allergic sensitization(1), by demonstrating that two early-onset remitting classes differed in their association with sensitization: there was little evidence of association between sensitization and pre-school remitting wheeze, but sensitization was strongly
associated with mid-childhood remitting wheeze. This suggests that some transient wheezers may have longer wheeze duration in association with the development of sensitization. Most previous studies have shown that lung function is impaired in some, but not all wheeze phenotypes. (11, 23, 24) In our study, all wheeze phenotypes were associated with decrements of diminished lung function compared to Never/infrequent wheeze, with the greatest impairment in Persistent wheeze. This is particularly important in the context of findings that low FEV<sub>1</sub> at its physiological plateau in early adult age is important in the genesis of chronic obstructive pulmonary disease (COPD) (27), and early cardio-vascular mortality (28). Several recent studies have shown that low childhood lung function trajectory is associated with future COPD risk (29-31), and our findings of impaired lung function in school age across all childhood wheezing phenotypes suggest that the notion that episodic wheeze in early life has a benign long-term prognosis may not be true for a proportion of transient wheezers.

It is often stated that temporal analyses may be crucial for distinguishing between different endotypes of asthma (6), and that derived latent classes (“phenotypes”) may be used to improve discovery process in genetic and environmental studies. (5) However, to date, there has been a lack of consistency in associations between wheeze phenotypes and genetic and environmental factors, and a scarcity of information on temporal characteristics of individual profiles of symptoms within the longitudinal latent classes. Our findings suggest that one of the reasons for the inconsistencies may be because associations are diluted due to within-phenotype heterogeneity. We observed considerable within-class heterogeneity, with 12% of children having low membership probability of any phenotype. The heterogeneity in between-individual wheezing patterns, within-phenotypes which increased as the assignment became less certain.
Individual wheezing patterns were particularly heterogeneous in the Late−onset wheeze. It is also of note that a proportion of children in the Early−onset pre−school remitting class reported wheezing at later time points. This can explain the increased risk of asthma in transient wheezers, as ~70% of individuals with later−life asthma within this class reported wheezing beyond age 10.

We suggest In conclusion, our results suggest that early−life episodic wheeze may not have a benign prognosis for a proportion of transient wheezers. Up to 12% of children have low membership probability of any wheezing phenotype, and that the mere presence or absence of a single symptom wheeze may be insufficient to derive homogenous disease subtypes phenotypes for probing causality. It is likely that more holistic indicators of wheeze development (including frequency and severity) are needed to describe individual trajectories and derive more homogenous phenotypes, and that we may need to move beyond LCA to facilitate better understand the true variation in the natural history and causality of childhood wheezing illness.

Acknowledgements: We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole MAAS, IOW, ASHFORD, SEATON and ALSPAC teams, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, and nurses.
References


10**STELAR investigators**

Dr John Curtin
Silvia Colicino
Professor Ashley Woodcock

11**Breathing Together investigators**

Prof Andrew Bush
Prof Sejal Saglani
Prof Clare M Lloyd
Prof Jonathan Grigg
Prof Jurgen Schwarze
Prof Mike Shields
Dr Peter Ghazal
Dr Multan Power
LEGEND FOR FIGURES

**Figure 1.** Estimated prevalence of wheeze for each of the five wheezing phenotypes identified by latent class analysis in 7719 children (infancy: age ½–1, early childhood: age 2–3, pre–school age / early–school age: age 4–5, middle childhood: age 8–10, adolescence: age 14–18)

*Class proportions shown in the figure legend are computed based on estimated posterior probabilities.*

**Figure 2.**

A) Transition of most likely membership class between latent models constructed with complete (n=7719) and incomplete (n=15942) data.

B) Assignment of children into wheeze phenotypes over a sequence of latent class analysis models with four and five classes based on most likely membership class

Ellipse nodes show class membership (most likely phenotype) whilst the values along the arrow represent the % of children moving from one class to another. This figure reflects whether the members of distinct wheeze phenotypes will remain in the same group or shift into another

**Figure 3.** Presence/Absence of wheeze at five time periods (TP) across latent classes stratified by cohort (Al: ALSPAC; As: ASHFORD; I: IOW; M: MAAS; S: SEATON) and class membership probabilities that have been derived by including ‘cohort site’ as a covariate.
Table 1  Per–cohort and combined characteristics of children with complete reports of wheezing

<table>
<thead>
<tr>
<th>Most likely Wheezing Phenotypes</th>
<th>Never/Infrequent wheeze</th>
<th>Early–onset pre-school remitting</th>
<th>Early–onset mid–childhood remitting</th>
<th>Persistent</th>
<th>Late onset</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEATON, N (%)</td>
<td>499</td>
<td>(61%)</td>
<td>(22%)</td>
<td>(3%)</td>
<td>(6%)</td>
</tr>
<tr>
<td>ALSPAC, N (%)</td>
<td>5149</td>
<td>(61%)</td>
<td>(20%)</td>
<td>(9%)</td>
<td>(5%)</td>
</tr>
<tr>
<td>MAAS, N (%)</td>
<td>667</td>
<td>(63%)</td>
<td>(14%)</td>
<td>(5%)</td>
<td>(13%)</td>
</tr>
<tr>
<td>ASHFORD, N (%)</td>
<td>492</td>
<td>(47%)</td>
<td>(36%)</td>
<td>(6%)</td>
<td>(10%)</td>
</tr>
<tr>
<td>IOW, N (%)</td>
<td>912</td>
<td>(60%)</td>
<td>46</td>
<td>37</td>
<td>66</td>
</tr>
<tr>
<td>5–COHORT DATA, N (%)</td>
<td>7719</td>
<td>(60%)</td>
<td>(19%)</td>
<td>(8%)</td>
<td>(6%)</td>
</tr>
</tbody>
</table>

| Gender Male, N (%)            | 3881                    | (50%)                            | 2162                              | 830        | 343       | 288 | 258 |
| Birth weight (kg)             | Mean (SD)               | 3.44 (0.53)                       | 3.44 (0.51)                       | 3.46 (0.54)| 3.41 (0.58) | 3.48 (0.57) | 3.43 |
| Low birth weight (≤2.5) Yes, N (%) | 314 (4.1%) | 161 (3%)                        | 63 (4%)                           | 36 (6%)    | 31 (6%)   | 23  |
| Maternal age (MA)             | Mean (SD)               | 29.1 (4.64)                      | 29.2 (4.57)                       | 28.9 (4.58)| 28.8 (4.73)| 28.7 (4.78)| 28.8 |
| Advanced MA (≥35) Yes, N (%)  | 947 (12%)               | 572 (13%)                        | 249 (12%)                         | 117 (12%)  | 72 (12%)  | 53  | 73  |
| Paternal smoking Yes, N (%)   | 1702 (27%)              | 979 (26%)                        | 306 (25%)                         | 144 (31%)  | 132 (32%) | 141 |
| Pet ownership Yes, N (%)      | 4149 (54%)              | 2449 (54%)                       | 818 (56%)                         | 334 (58%)  | 264 (55%) | 284 |
| Sensitization (mid–childhood) Yes, N (%) | 1042 (17%) | 461 (13%)                     | 134 (12%)                         | 106 (23%)  | 181 (47%) | 160 |
| Current asthma (last follow–up) Yes, N (%) | 1000 (14%) | 121 (3%)                    | 113 (8%)                          | 86 (15%)   | 352 (75%) | 328 |
| Asthma ever (last follow–up) Yes, N (%) | 1737 (26%) | 452 (11%)                     | 291 (24%)                         | 274 (59%)  | 374 (91%) | 346 |
| Asthma med. ever Yes, N (%)   | 949 (13%)               | 120 (3%)                         | 113 (8%)                          | 87 (15%)   | 345 (73%) | 284 |
| Asthma med. (last follow–up) Yes, N (%) | 2125 (42%) | 591 (14%)                     | 294 (20%)                         | 374 (64%)  | 451 (93%) | 415 |
Table 2 Adjusted associations of wheezing phenotypes with late–outcomes including current asthma, asthma ever, current use of asthma medication, asthma medication ever and eczema ever using weighted membership probabilities.

<table>
<thead>
<tr>
<th></th>
<th>Current** asthma</th>
<th>Asthma ever</th>
<th>Current** asthma med.</th>
<th>Asthma med. ever</th>
<th>Eczema ever</th>
</tr>
</thead>
<tbody>
<tr>
<td>Never/Infrequent</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Early–onset pre–school remitting</td>
<td>2.342.28</td>
<td>2.051.84</td>
<td>2.352.18</td>
<td>1.361.34</td>
<td>1.241.19</td>
</tr>
<tr>
<td>(95%CI)</td>
<td>(1.73, 3.17)</td>
<td>(1.56, 2.17)</td>
<td>(1.74, 3.18)</td>
<td>(1.14, 1.62)</td>
<td>(1.07, 1.42)</td>
</tr>
<tr>
<td>p value</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
<td>0.001 &lt;.0001</td>
<td>5.00E–03 &lt;.0010</td>
</tr>
<tr>
<td>Early–onset mid–childhood remitting</td>
<td>3.65 3.74</td>
<td>6.416.84</td>
<td>3.53 3.59</td>
<td>6.295.97</td>
<td>1.311.41</td>
</tr>
<tr>
<td>(95%CI)</td>
<td>(2.55, 5.21)</td>
<td>(5.11, 8.03)</td>
<td>(2.47, 5.05)</td>
<td>(4.85, 7.35)</td>
<td>(1.06, 1.61)</td>
</tr>
<tr>
<td>p value</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Persistent</td>
<td>48.31 56.54</td>
<td>37.9545.37</td>
<td>42.45 49.81</td>
<td>38.67 48.01</td>
<td>3.223.97</td>
</tr>
<tr>
<td>(95%CI)</td>
<td>(35.87, 65.07)</td>
<td>(27.78, 51.84)</td>
<td>(31.57, 57.09)</td>
<td>(28.27, 52.90)</td>
<td>(2.55, 4.08)</td>
</tr>
<tr>
<td>p value</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Late onset</td>
<td>36.3947.62</td>
<td>13.2412.00</td>
<td>24.5126.83</td>
<td>17.7519.93</td>
<td>2.192.03</td>
</tr>
<tr>
<td>(95%CI)</td>
<td>(28.89, 49.25)</td>
<td>(10.68, 16.41)</td>
<td>(18.07, 33.23)</td>
<td>(13.59, 23.17)</td>
<td>(1.62, 2.53)</td>
</tr>
<tr>
<td>p value</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

*Adjusted for sensitization (mid-childhood), sex, maternal history of asthma or allergy, maternal smoking and low birth weight

**Available at the latest follow–up (e.g. 18 years in IOW, 16 years in MAAS, 15 years in SEATON, 15 years in ASHFORD and 14 years in ALSPAC.)
Table 3 Associations of wheezing phenotypes with lung function measures adjusted for sex, height and gender using weighted membership probabilities.

<table>
<thead>
<tr>
<th>Wheezing Phenotype</th>
<th>Mean Difference (95%CI)*</th>
<th>P value</th>
<th>Z scores for FEV$_1$**</th>
<th>Z scores for FVC**</th>
<th>Z scores for FEV$_1$/FVC**</th>
</tr>
</thead>
<tbody>
<tr>
<td>Never/ Infrequent</td>
<td>reference</td>
<td>reference</td>
<td>reference</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Early–onset pre–school remitting</td>
<td>-0.11 (-0.17, -0.04)</td>
<td>$2.00E-03$</td>
<td>0.002</td>
<td>-0.05 (-0.12, 0.02)</td>
<td>0.144</td>
</tr>
<tr>
<td>Early–onset mid–childhood remitting</td>
<td>-0.19 (-0.28, -0.09)</td>
<td>&lt;.0001</td>
<td>0.531</td>
<td>-0.03 (-0.13, 0.06)</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Persistent</td>
<td>-0.34 (-0.44, -0.24)</td>
<td></td>
<td></td>
<td>0.03 (-0.07, 0.13)</td>
<td>-0.60 (-0.70, -0.50)</td>
</tr>
<tr>
<td>Late onset</td>
<td>-0.22 (-0.32, -0.12)</td>
<td></td>
<td></td>
<td>-0.04 (-0.14, 0.06)</td>
<td>-0.30 (-0.40, -0.20)</td>
</tr>
</tbody>
</table>

* Adjusted for maternal history of asthma or allergy, maternal smoking and low birth weight

** Sex-, age-, and height–adjusted SD units

Figure 1
Figure 2

- Never/Infrequent wheeze (52.1%)
- Early-onset pre-school remitting wheeze (23.9%)
- Early-onset mid-childhood remitting wheeze (9.0%)
- Persistent wheeze (7.9%)
- Late-onset wheeze (7.1%)