

# Distinguishing Wheezing Phenotypes from Infancy to Adolescence

## A Pooled Analysis of Five Birth Cohorts

Ceyda Oksel<sup>1</sup>, Raquel Granell<sup>2</sup>, Sadia Haider<sup>1</sup>, Sara Fontanella<sup>1</sup>, Angela Simpson<sup>3</sup>, Steve Turner<sup>4</sup>, Graham Devereux<sup>5</sup>, Syed Hasan Arshad<sup>6,7</sup>, Clare S. Murray<sup>3</sup>, Graham Roberts<sup>6,7,8</sup>, John W. Holloway<sup>6,8</sup>, Paul Cullinan<sup>9</sup>, John Henderson<sup>2\*</sup>, and Adnan Custovic<sup>1\*</sup>; on behalf of STELAR<sup>‡</sup> and Breathing Together Investigators<sup>§</sup>

<sup>1</sup>Department of Paediatrics, and <sup>9</sup>National Heart and Lung Institute, Imperial College London, London, United Kingdom; <sup>2</sup>MRC Integrative Epidemiology Unit, Department of Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom; <sup>3</sup>Division of Infection, Immunity and Respiratory Medicine, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, United Kingdom; <sup>4</sup>Child Health, University of Aberdeen, Aberdeen, United Kingdom; <sup>5</sup>Liverpool School of Tropical Medicine, Liverpool, United Kingdom; <sup>6</sup>David Hide Asthma and Allergy Research Centre, Isle of Wight, United Kingdom; <sup>7</sup>Respiratory Biomedical Research Unit, University Hospitals Southampton NHS Foundation Trust, Southampton, United Kingdom; and <sup>8</sup>Human Development and Health, Faculty of Medicine, University of Southampton, Southampton, United Kingdom

ORCID ID: 0000-0001-8393-5060 (S.T.).

### Abstract

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

**Objectives:** 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 7,719 participants from five birth cohorts with complete report of wheeze at five time periods. We tested the associations of derived phenotypes with late asthma outcomes and lung function, and investigated the uncertainty in phenotype assignment.

**Results:** We identified five phenotypes: never/infrequent wheeze (52.1%), early onset preschool remitting (23.9%), early onset midchildhood remitting (9%), persistent (7.9%), and late-onset wheeze (7.1%). Compared with the never/infrequent wheeze, all phenotypes had higher odds of asthma and lower forced expiratory volume in 1 second and forced expiratory volume in 1 second/forced

vital capacity in adolescence. The association with asthma was strongest for persistent wheeze (adjusted odds ratio, 56.54; 95% confidence interval, 43.75–73.06). We observed considerable within-class heterogeneity at the 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, whereas many children assigned to early onset preschool remitting class reported wheezing at later time points.

**Conclusions:** All wheeze phenotypes had significantly diminished lung function in school-age children, 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

(Received in original form November 29, 2018; accepted in final form March 19, 2019)

Ⓒ This article is open access and distributed under the terms of the Creative Commons Attribution Non-Commercial No Derivatives License 4.0 (<http://creativecommons.org/licenses/by-nc-nd/4.0/>). For commercial usage and reprints, please contact Diane Gern ([dgern@thoracic.org](mailto:dgern@thoracic.org)).

\*These authors contributed equally.

‡A complete list of members may be found before the beginning of the REFERENCES.

§A complete list of members may be found before the beginning of the REFERENCES.

**Author Contributions:** C.O., R.G., J.H., and A.C. conceived and planned the study and wrote the manuscript. C.O. and R.G. analyzed 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.

Correspondence and requests for reprints should be addressed to Professor John Henderson, M.D., Department of Population Health Sciences, Bristol Medical School, University of Bristol, 1-5 Whiteladies Road, Bristol BS8 1NU, UK. E-mail: [A.J.Henderson@bristol.ac.uk](mailto:A.J.Henderson@bristol.ac.uk)

This article has an online supplement, which is accessible from this issue's table of contents at [www.atsjournals.org](http://www.atsjournals.org).

Ann Am Thorac Soc Vol 16, No 7, pp 868–876, Jul 2019

Copyright © 2019 by the American Thoracic Society

DOI: 10.1513/AnnalsATS.201811-837OC

Internet address: [www.atsjournals.org](http://www.atsjournals.org)

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 persists, 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). In a pioneering study using clinical insights about the age of onset, progression, and remission of wheezing, Martinez and coworkers (1) described three mutually exclusive wheezing phenotypes (transient early, late-onset, and persistent). 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 name, they often differ in temporal trajectories, distributions within a population, and associated risk factors (9). This is 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 phenotype, which has seldom been investigated. For example, in most LCA models, there are individuals who cannot be classified with a high degree of certainty. Latent class model estimates 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, and 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 U.K. 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, because the term “phenotype” has been used in this context for more than a decade, we maintain this nomenclature in this manuscript.

## Methods

### Study Design, Setting, and Participants

The STELAR consortium (6) brings together five U.K. population-based birth cohorts: Avon Longitudinal Study of Parents and Children (ALSPAC) (17), Ashford (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](http://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 yr, 7 in ASHFORD over 14 yr, 6 in MAAS over 16 yr, 6 in SEATON over 14 yr, and 5 in IOW over 18 yr). The cohort-specific time points and sample size included in this analysis are shown in Table E1 in the online supplement. We used data on current wheeze collected at proximate time periods across cohorts: infancy (0.5–1 yr), early childhood (2–3 yr), preschool/early school age (4–5 yr), middle childhood (8–10 yr), and adolescence (14–18 yr). 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 (*see* Table E2).

Skin prick testing was performed 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 3 mm greater than the negative control to one or more common allergens.

We performed spirometry and recorded forced expiratory volume in one second (FEV<sub>1</sub>) and forced vital capacity (FVC). To reduce variability, we restricted our analyses to lung function measured at comparable ages (8.5 yr in ALSPAC, 10 yr in SEATON and IOW, and 11 yr 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 7,719 children with complete data on wheezing at five time periods. Cohort ID was included as an additional covariate by transforming the five-category variable into a set of four dummy variables. The optimal number of classes was selected based on the Bayesian information criterion and interpretability. We repeated analyses among 15,941 children with at least two observations, and examined the stability of allocation of 7,719 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 four classes to models with five classes.

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<sub>1</sub>, FVC, and FEV<sub>1</sub>/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 (midchildhood), 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  $\geq 0.80$ , medium = 0.60–0.80, low  $< 0.60$ ) to elicit the extent of within-class homogeneity in different certainty thresholds. Analyses were performed using Mplus 8, R (<http://www.r-project.org/>) and Stata 14 (StataCorp).

## Results

### Characteristics of the Study

#### Population

Of 7,719 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 (standard deviation) birth weight in the combined dataset was 3.44 kg (0.53), with 314 (4.1%) children having a low birth weight ( $\leq 2.5$  kg); mean maternal age at delivery was 29.1 years (standard deviation, 4.64), and 1,198 (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 (*see*

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 preschool remitting wheeze (23.9%), with 50–60% prevalence of wheeze during infancy, decreasing to 20% around early childhood and to less than 10% afterward; 3) early onset midchildhood remitting wheeze (9%), with early onset and peak prevalence ( $\sim 75\%$ ) in preschool/early school age, decreasing to 23% in midchildhood, 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 approximately 80%; and 5) late-onset wheeze (7.1%) with a low prevalence of 27% until preschool/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 four cohorts with complete data (excluding ALSPAC) provided very similar results.

Figure 2A shows the changes in the allocation of 7,719 individuals with complete data from the model using only children with data on wheezing at all five time periods to their most probable class in the model using 15,941 individuals with at least two observations. The optimal solution in the model using 15,941 children remained five classes (*see* 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 midchildhood 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 four and five 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 (preschool and midchildhood resolution), whereas a third of children from the persistent wheeze in a four-class solution transitioned, mainly to early onset midchildhood 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 associated with all four phenotypes, with the strongest association with persistent wheezing (RRR, 3.12; 95% CI, 2.51–3.89;  $P < 0.0001$ ). Compared with never/infrequent wheeze, all wheezing phenotypes except early onset preschool 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 was a predictor of late-onset wheeze in males only (*see* Table E4).

The strongest association with allergic sensitization (ages 4–7.5 yr) 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 preschool remitting wheeze, but clear evidence of an association between sensitization and midchildhood 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 with never/infrequent wheeze, all four wheeze phenotypes were associated with higher asthma prevalence. This association was weakest for early onset preschool 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 8,223 children with incomplete reports of wheezing (excluding 7,719 children with complete reports at five time points) (*see* Table E7). Characteristics of children with complete report of wheeze and those with missing data are shown in Table E8.

**Table 1.** Per cohort and combined characteristics of children with complete reports of wheezing

		Most Likely Wheezing Phenotypes					
		Never/Infrequent Wheeze	Early Onset Preschool Remitting	Early Onset Midchildhood Remitting	Persistent	Late Onset	
SEATON, <i>n</i> (%)		499	305 (61)	108 (22)	16 (3)	31 (6)	39 (8)
ALSPAC, <i>n</i> (%)		5,149	3,119 (61)	1,058 (20)	476 (9)	253 (5)	243 (5)
MAAS, <i>n</i> (%)		667	421 (63)	96 (14)	34 (5)	84 (13)	32 (5)
Ashford, <i>n</i> (%)		492	233 (47)	176 (36)	28 (6)	50 (10)	5 (1)
IOW, <i>n</i> (%)		912	548 (60)	46 (5)	37 (4)	66 (7)	215 (24)
5-Cohort data, <i>n</i> (%)		7,719	4,626 (60)	1,484 (19)	591 (8)	484 (6)	534 (7)
Sex	Male, <i>n</i> (%)	3,881 (50)	2,162 (47)	830 (56)	343 (58)	288 (60)	258 (48)
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 (0.52)
Low birth weight ( $\leq 2.5$ )	Yes, <i>n</i> (%)	314 (4.1)	161 (3)	63 (4)	36 (6)	31 (6)	23 (4)
Maternal age	Mean (SD)	29.1 (4.64)	29.2 (4.57)	28.9 (4.58)	28.8 (4.73)	28.7 (4.78)	28.8 (5.07)
Advanced maternal age ( $\geq 35$ )	Yes, <i>n</i> (%)	947 (12)	572 (13)	177 (12)	72 (12)	53 (11)	73 (14)
Maternal smoking	Yes, <i>n</i> (%)	6,491 (16)	647 (14)	249 (17)	117 (20)	88 (18)	97 (18)
Paternal smoking	Yes, <i>n</i> (%)	1,702 (27)	979 (26)	306 (25)	144 (31)	132 (32)	141 (29)
Pet ownership	Yes, <i>n</i> (%)	4,149 (54)	2,449 (54)	818 (56)	334 (58)	264 (55)	284 (54)
Sensitization (midchildhood)	Yes, <i>n</i> (%)	1,042 (17)	461 (13)	134 (12)	106 (23)	181 (47)	160 (39)
Current asthma (last follow-up)	Yes, <i>n</i> (%)	1,000 (14)	121 (3)	113 (8)	86 (15)	352 (75)	328 (67)
Asthma ever	Yes, <i>n</i> (%)	1,737 (26)	452 (11)	291 (24)	274 (59)	374 (91)	346 (72)
Asthma medication (last follow-up)	Yes, <i>n</i> (%)	949 (13)	120 (3)	113 (8)	87 (15)	345 (73)	284 (58)
Asthma medication ever	Yes, <i>n</i> (%)	2,125 (42)	591 (14)	294 (20)	374 (64)	451 (93)	415 (89)

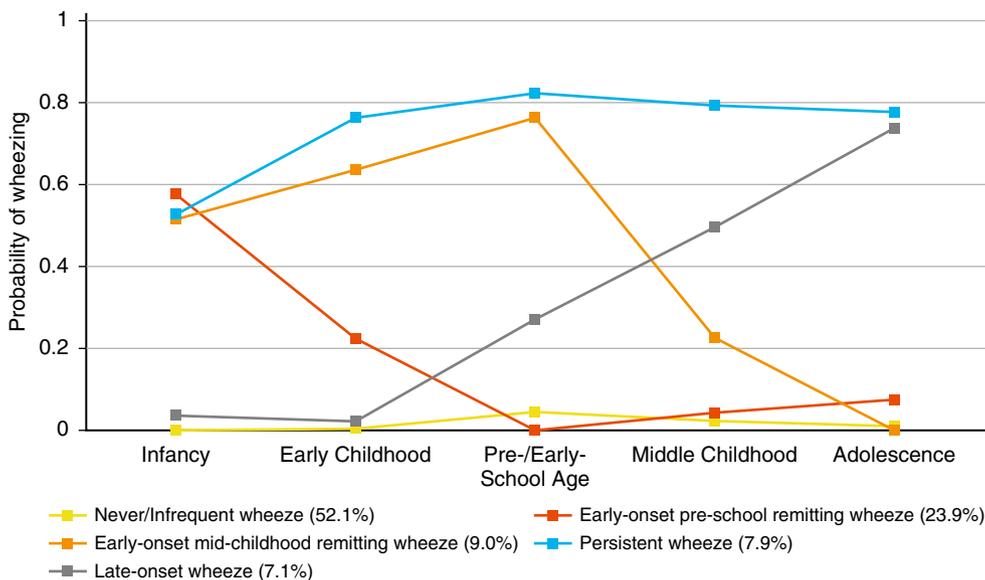
Definition of abbreviations: ALSPAC = Avon Longitudinal Study of Parents and Children; IOW = Isle of Wight; MAAS = Manchester Asthma and Allergy Study; SD = standard deviation; SEATON = Aberdeen Study of Eczema and Asthma to Observe the Effects of Nutrition study.

Associations of wheeze phenotypes with lung function in middle childhood (8.5–11 yr) are shown in Table 3. FEV<sub>1</sub> and FEV<sub>1</sub>/FVC were significantly lower in all wheeze phenotypes compared with

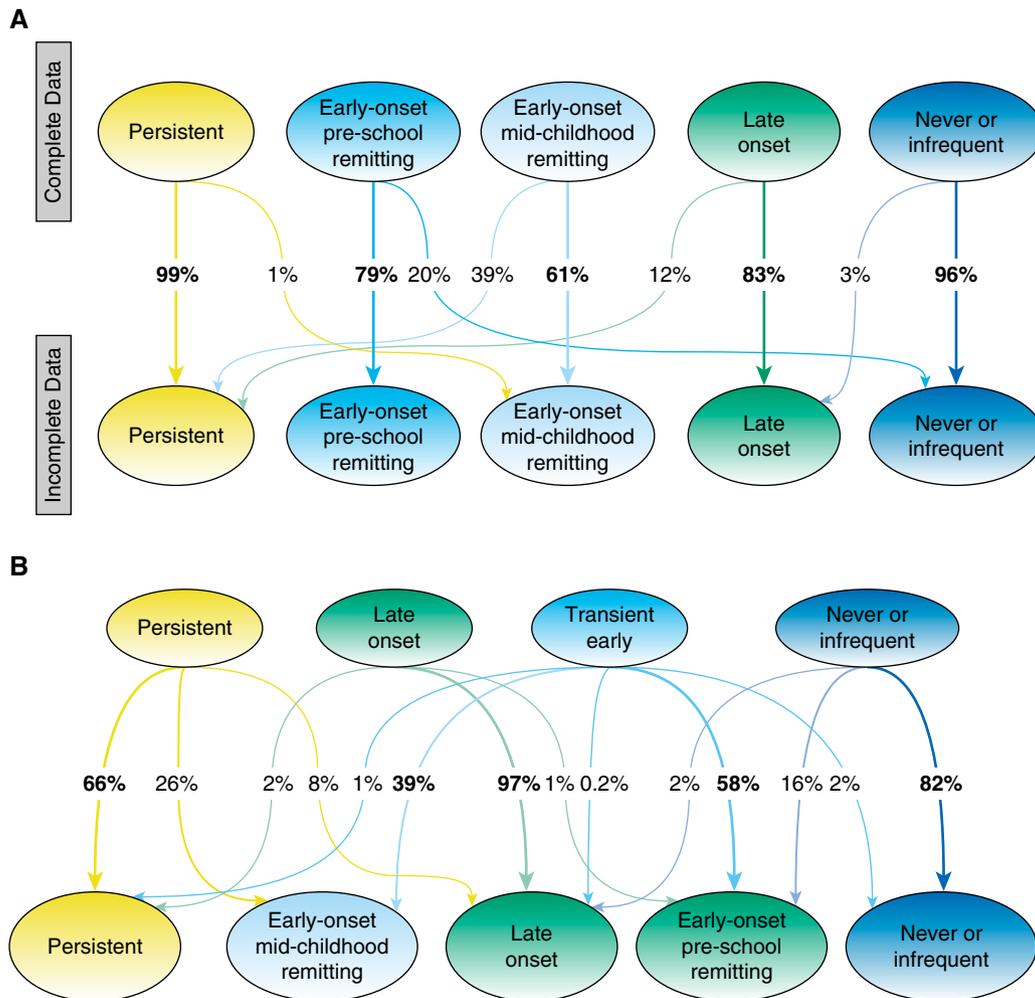
never/infrequent class. Association was particularly strong for persistent wheeze (*z* scores  $-0.60$ ; 95% CI,  $-0.70$  to  $-0.50$ ;  $P < 0.0001$ ). There was no evidence of differences in FVC.

**Individual Variation within and across Wheezing Phenotypes**

Figure 3 shows the presence/absence of wheeze at five time periods among each individual across latent classes, stratified by



**Figure 1.** Estimated prevalence of wheeze for each of the five wheezing phenotypes identified by latent class analysis in 7,719 children (infancy, age 0.5–1; early childhood, age 2–3; preschool 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 ( $n = 7,719$ ) and incomplete ( $n = 15,942$ ) 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), whereas the values along the arrow represent the percentage 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.

cohort and class membership certainty. More than 80% of children in never/infrequent and persistent wheeze classes were assigned with a high degree of certainty (membership probability  $>0.80$ ), whereas class membership certainty was the lowest for late-onset wheeze, with 51% of its members having posterior probabilities less than 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 greater than or equal to 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 preschool remitting class, but reported current asthma at the last follow-up, 78 (69%) reported wheezing at later time points.

Over a tenth of individuals (913/7,719; 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 (e.g., some have  $\sim 0.40$  probability of belonging to never/infrequent

wheeze and  $\sim 0.40$  probability of belonging to early onset preschool remitting wheeze).

### Discussion

By jointly modeling 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 (preschool remitting and midchildhood remitting) phenotypes. The model assigned 33% of children to remitting (24% remitting in preschool, and 9% in

**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

	Adjusted Odds Ratio (95% CI)*				
	Current <sup>†</sup> Asthma	Asthma Ever	Current <sup>†</sup> Asthma Medication	Asthma Medication Ever	Eczema Ever
Never/infrequent	Reference	Reference	Reference	Reference	Reference
Early onset preschool remitting	2.34	2.05	2.35	1.36	1.24
95% CI	1.73 to 3.17	1.70 to 2.46	1.74 to 3.18	1.14 to 1.62	1.07 to 1.42
P value	<0.0001	<0.0001	<0.0001	0.001	0.003
Early onset midchildhood remitting	3.65	6.41	3.53	5.97	1.31
95% CI	2.55 to 5.21	5.11 to 8.03	2.47 to 5.05	4.85 to 7.35	1.06 to 1.61
P value	<0.0001	<0.0001	<0.0001	<0.0001	0.010
Persistent	48.31	37.95	42.45	38.67	3.22
95% CI	35.87 to 65.07	27.78 to 51.84	31.57 to 57.09	28.27 to 52.90	2.55 to 4.08
P value	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Late onset	36.39	12.00	24.51	17.75	2.03
95% CI	26.89 to 49.25	9.41 to 15.31	18.07 to 33.23	13.59 to 23.17	1.62 to 2.53
P value	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001

Definition of abbreviations: ALSPAC = Avon Longitudinal Study of Parents and Children; CI = confidence interval; IOW = Isle of Wight; MAAS = Manchester Asthma and Allergy Study; SEATON = Aberdeen Study of Eczema and Asthma to Observe the Effects of Nutrition Study.

\*Adjusted for sensitization (midchildhood), sex, maternal history of asthma or allergy, maternal smoking, and low birth weight.

<sup>†</sup>Available at the latest follow-up (e.g., 18 yr in IOW, 16 yr in MAAS, 15 yr in SEATON, 15 yr in Ashford, and 14 yr in ALSPAC).

midchildhood), 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 FEV<sub>1</sub> and FEV<sub>1</sub>/FVC, 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 was less than 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).

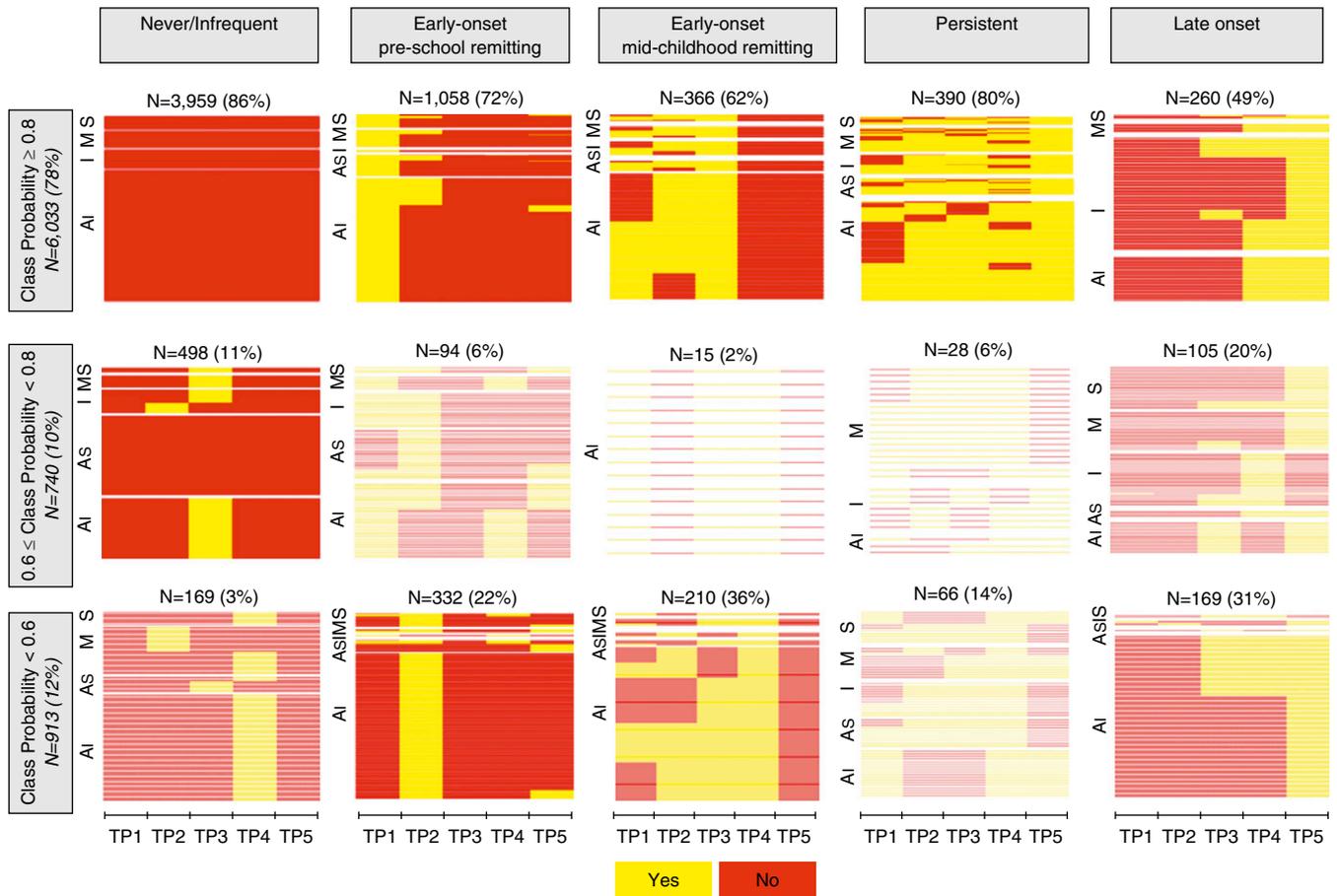
**Table 3.** Associations of wheezing phenotypes with lung function measures adjusted for sex, height, and gender using weighted membership probabilities

	Mean Difference (95% CI)* P value		
	z Scores for FEV <sub>1</sub> <sup>†</sup>	z Scores for FVC <sup>†</sup>	z Scores for FEV <sub>1</sub> /FVC <sup>†</sup>
Never/infrequent	Reference	Reference	Reference
Early onset preschool remitting	-0.11 (-0.17 to -0.04)	-0.05 (-0.12 to 0.02)	-0.09 (-0.16 to -0.02)
	0.002	0.144	0.001
Early onset midchildhood remitting	-0.19 (-0.28 to -0.09)	-0.03 (-0.13 to 0.06)	-0.26 (-0.35 to -0.16)
	<0.0001	0.531	<0.0001
Persistent	-0.34 (-0.44 to -0.24)	0.03 (-0.07 to 0.13)	-0.60 (-0.70 to -0.50)
	<0.0001	0.577	<0.0001
Late onset	-0.22 (-0.32 to -0.12)	-0.04 (-0.14 to 0.06)	-0.30 (-0.40 to -0.20)
	<0.0001	0.406	<0.0001

Definition of abbreviations: CI = confidence interval; FEV<sub>1</sub> = forced expiratory volume in 1 second; FVC = forced vital capacity.

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

<sup>†</sup>Sex-, age-, and height-adjusted standard deviation units.



**Figure 3.** Presence/absence of wheeze at five time periods across latent classes stratified by cohort and class membership probabilities that have been derived by including cohort site as a covariate. AI = Avon Longitudinal Study of Parents and Children; AS = Ashford; I = Isle of Wight; M = Manchester Asthma and Allergy Study; S = Aberdeen Study of Eczema and Asthma to Observe the Effects of Nutrition Study; TP = time period.

One limitation of this multicohort 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, whereas IOW had the lowest prevalence of early onset preschool remitting wheeze and the highest prevalence for late-onset wheeze. We addressed this limitation by including “cohort site” as a covariate for determining phenotypes.

One of the advantages of our multicohort approach is that individual studies that might not provide conclusive evidence to make inference about the general population because of 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 U.K. population. A further advantage is to minimize the study-specific biases (including cohort-specific effects, attrition effects, different recruitment strategies, and geographic 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 multicohort 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.

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, a four-class solution identified the same classes, whereas the optimum solution included a fifth phenotype, early onset midchildhood remitting wheeze, which was previously observed in ALSPAC (3, 12, 16) but not in another 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 artifact of the analysis (e.g., caused by 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 was consistent with the analysis using 7,719 complete cases, but there was a considerable transition in other classes (e.g., almost 40% of children in early onset midchildhood 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. Preschool remitting phenotype included children who were assigned to transient early and never/infrequent wheeze, whereas midchildhood 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, 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 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 preschool remitting wheeze, but sensitization was strongly associated with midchildhood 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 diminished lung function compared with 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 physiologic plateau in early adult age is important in the genesis of chronic obstructive pulmonary disease (27) and early cardiovascular mortality (28). Several recent studies have shown that low childhood lung function trajectory is associated with future chronic obstructive pulmonary disease 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 in wheezing patterns, 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 preschool remitting class reported wheezing at later time points. This can explain the increased risk of asthma in transient wheezers, because approximately 70% of individuals with later-life asthma within this class reported wheezing beyond age 10.

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 the mere presence or absence of a wheeze may be insufficient to derive homogenous phenotypes for probing causality. It is likely that more holistic indicators of wheeze (including frequency and severity) are needed to describe individual trajectories and derive more homogenous phenotypes to facilitate better understanding of the natural history and causality of childhood wheezing illness. ■

**Author disclosures** are available with the text of this article at [www.atsjournals.org](http://www.atsjournals.org).

**STELAR investigators:** Dr. John Curtin, Silvia Colicino, and Professor Ashley Woodcock.

**Breathing Together investigators:** Professor Andrew Bush, Professor Sejal Saglani, Professor Clare M. Lloyd, Dr. Benjamin Marsland, Professor Jonathan Grigg, Professor Jurgen Schwarze, Professor Mike Shields, Dr. Peter Ghazal, and Dr. Multan Power.

**Acknowledgment:** The authors thank all the families who took part in this study; the midwives for their help in recruiting them; and the MAAS, IOW, Ashford, SEATON, and ALSPAC teams, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, and nurses.

## References

- Martinez FD, Wright AL, Taussig LM, Holberg CJ, Halonen M, Morgan WJ; The Group Health Medical Associates. Asthma and wheezing in the first six years of life. *N Engl J Med* 1995;332:133–138.
- Belgrave DC, Simpson A, Semic-Jusufagic A, Murray CS, Buchan I, Pickles A, et al. Joint modeling of parentally reported and physician-confirmed wheeze identifies children with persistent troublesome wheezing. *J Allergy Clin Immunol* 2013;132:575–583.
- Savenije OE, Granell R, Caudri D, Koppelman GH, Smit HA, Wijga A, et al. Comparison of childhood wheezing phenotypes in 2 birth cohorts: ALSPAC and PIAMA. *J Allergy Clin Immunol* 2011;127:1505–1512.
- 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;33:573–578.
- Deliu M, Belgrave D, Sperrin M, Buchan I, Custovic A. Asthma phenotypes in childhood. *Expert Rev Clin Immunol* 2017;13:705–713.
- Howard R, Rattray M, Prosperi M, Custovic A. Distinguishing asthma phenotypes using machine learning approaches. *Curr Allergy Asthma Rep* 2015;15:38.
- Saglani S, Custovic A. Childhood asthma: advances using machine learning and mechanistic studies. *Am J Respir Crit Care Med* 2019;199:414–422.
- Oksel C, Granell R, Mahmoud O, Custovic A, Henderson AJ; STELAR; Breathing Together investigators. Causes of variability in latent phenotypes of childhood wheeze. *J Allergy Clin Immunol* 2019;143:1783–1790.
- Oksel C, Haider S, Fontanella S, Frainay C, Custovic A. Classification of pediatric asthma: from phenotype discovery to clinical practice. *Front Pediatr* 2018;6:258.

- 10 Cano-Garcinuño A, Mora-Gandarillas I; SLAM Study Group. Wheezing phenotypes in young children: an historical cohort study. *Prim Care Respir J* 2014;23:60–66.
- 11 Lowe LA, Simpson A, Woodcock A, Morris J, Murray CS, Custovic A; NAC Manchester Asthma and Allergy Study Group. Wheeze phenotypes and lung function in preschool children. *Am J Respir Crit Care Med* 2005;171:231–237.
- 12 Henderson J, Granell R, Heron J, Sherriff A, Simpson A, Woodcock A, 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;63:974–980.
- 13 Depner M, Fuchs O, Genuneit J, Karvonen AM, Hyvärinen A, Kaulek V, et al.; PASTURE Study Group. Clinical and epidemiologic phenotypes of childhood asthma. *Am J Respir Crit Care Med* 2014;189:129–138.
- 14 Siroux V, Basagaña X, Boudier A, Pin I, Garcia-Aymerich J, Vesin A, et al. Identifying adult asthma phenotypes using a clustering approach. *Eur Respir J* 2011;38:310–317.
- 15 Spycher BD, Silverman M, Pescatore AM, Beardsmore CS, Kuehni CE. Comparison of phenotypes of childhood wheeze and cough in 2 independent cohorts. *J Allergy Clin Immunol* 2013;132:1058–1067.
- 16 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;138:1060–1070.
- 17 Golding J, Pembrey M, Jones R; ALSPAC Study Team. ALSPAC—the Avon longitudinal study of parents and children. I. Study methodology. *Paediatr Perinat Epidemiol* 2001;15:74–87.
- 18 Atkinson W, Harris J, Mills P, Moffat S, White C, Lynch O, et al. Domestic aeroallergen exposures among infants in an English town. *Eur Respir J* 1999;13:583–589.
- 19 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;96:163–169.
- 20 Custovic A, Simpson BM, Murray CS, Lowe L, Woodcock A; NAC Manchester Asthma and Allergy Study Group. The national asthma campaign Manchester asthma and allergy study. *Pediatr Allergy Immunol* 2002;13:32–37.
- 21 Martindale S, McNeill G, Devereux G, Campbell D, Russell G, Seaton A. Antioxidant intake in pregnancy in relation to wheeze and eczema in the first two years of life. *Am J Respir Crit Care Med* 2005;171:121–128.
- 22 Belgrave DC, Custovic A, Simpson A. Characterizing wheeze phenotypes to identify endotypes of childhood asthma, and the implications for future management. *Expert Rev Clin Immunol* 2013;9:921–936.
- 23 Duijts L, Granell R, Sterne JA, Henderson AJ. Childhood wheezing phenotypes influence asthma, lung function and exhaled nitric oxide fraction in adolescence. *Eur Respir J* 2016;47:510–519.
- 24 Morgan WJ, Stern DA, Sherrill DL, Guerra S, Holberg CJ, Guilbert TW, et al. Outcome of asthma and wheezing in the first 6 years of life: follow-up through adolescence. *Am J Respir Crit Care Med* 2005;172:1253–1258.
- 25 Sherriff A, Peters TJ, Henderson J, Strachan D. Risk factor associations with wheezing patterns in children followed longitudinally from birth to 3 1/2 years. *Int J Epidemiol* 2001;30:1473–1484.
- 26 van der Valk RJ, Caudri D, Savenije O, Koppelman GH, Smit HA, Wijga AH, et al. Childhood wheezing phenotypes and FeNO in atopic children at age 8. *Clin Exp Allergy* 2012;42:1329–1336.
- 27 Lange P, Celli B, Agustí A, Boje Jensen G, Divo M, Faner R, et al. Lung-function trajectories leading to chronic obstructive pulmonary disease. *N Engl J Med* 2015;373:111–122.
- 28 Vasquez MM, Zhou M, Hu C, Martinez FD, Guerra S. Low lung function in young adult life is associated with early mortality. *Am J Respir Crit Care Med* 2017;195:1399–1401.
- 29 Berry CE, Billheimer D, Jenkins IC, Lu ZJ, Stern DA, Gerald LB, et al. A distinct low lung function trajectory from childhood to the fourth decade of life. *Am J Respir Crit Care Med* 2016;194:607–612.
- 30 Bui DS, Lodge CJ, Burgess JA, Lowe AJ, Perret J, Bui MQ, et al. Childhood predictors of lung function trajectories and future COPD risk: a prospective cohort study from the first to the sixth decade of life. *Lancet Respir Med* 2018;6:535–544.
- 31 Belgrave DCM, Granell R, Turner SW, Curtin JA, Buchan IE, Le Souëf PN, et al. Lung function trajectories from pre-school age to adulthood and their associations with early life factors: a retrospective analysis of three population-based birth cohort studies. *Lancet Respir Med* 2018;6:526–534.