Distinguishing Wheezing Phenotypes from Infancy to Adolescence: A Pooled Analysis of Five Birth Cohorts

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

[1]Department of Paediatrics, Imperial College London, UK; [2]MRC Integrative Epidemiology Unit, Population Health Sciences, Bristol Medical School, University of Bristol, UK; [3]Division 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; [4]Child Health, University of Aberdeen, Aberdeen, UK; [5]Liverpool School of Tropical Medicine; [6]Human Development and Health, Faculty of Medicine, University of Southampton, Southampton, UK; [7]David Hide Asthma and Allergy Research Centre, Isle of Wight, UK; [8]Respiratory Biomedical Research Unit, University Hospitals Southampton NHS Foundation Trust, Southampton, UK; [9]National Heart and Lung Institute, Imperial College, London, UK
[§]Equal contribution
[*]See list of STELAR investigators before the References
‡See list of Breathing Together investigators before the References

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

Author 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

Keywords: wheezing phenotypes, childhood, adolescence, latent class

Descriptor: 1.17

Word Count: 3522

Online Data Supplement: This article has an online data supplement, which is accessible from this issue's table of content online at www.atsjournals.org

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.

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 FEV_1 and FEV_1/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.

Abstract Word Count: 250

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) 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 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 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 (FEV_1) 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 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.

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 ALSPAC) provided very similar results.

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 in the model using 15,941 individuals with at least two observations. The optimal 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 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 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 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 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.

Individual Variation within and across Wheezing Phenotypes

Figure 3 shows the presence/absence of wheeze at five time periods amongst each individual across latent classes, stratified by 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), 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 FEV_1 and FEV_1/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 <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)

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.

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 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.

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, 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, 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 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 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_1 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 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 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.

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 understand 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.

STELAR investigators:

Dr John Curtin

Silvia Colicino

Professor Ashley Woodcock

‡*Breathing Together investigators:*

Prof Andrew Bush

Prof Sejal Saglani

Prof Clare M Lloyd

Dr Benjamin Marsland

Prof Jonathan Grigg

Prof Jurgen Schwarze

Prof Mike Shields

Dr Peter Ghazal

Dr Multan Power

References

1. Martinez FD, Wright AL, Taussig LM, Holberg CJ, Halonen M, Morgan WJ. Asthma and wheezing in the first six years of life. *New England Journal of Medicine* 1995; 332: 133-138.
2. Belgrave DC, Simpson A, Semic-Jusufagic A, Murray CS, Buchan I, Pickles A, Custovic A. Joint modeling of parentally reported and physician-confirmed wheeze identifies children with persistent troublesome wheezing. *Journal of Allergy and Clinical Immunology* 2013; 132: 575-583. e512.
3. Savenije OE, Granell R, Caudri D, Koppelman GH, Smit HA, Wijga A, de Jongste JC, Brunekreef B, Sterne JA, Postma DS. Comparison of childhood wheezing phenotypes in 2 birth cohorts: ALSPAC and PIAMA. *J Allergy Clin Immunol* 2011; 127: 1505-1512. e1514.
4. Kurukulaaratchy R, Fenn M, Waterhouse L, Matthews S, Holgate S, Arshad S. Characterization of wheezing phenotypes in the first 10 years of life. *Clinical & Experimental Allergy* 2003; 33: 573-578.
5. Deliu M, Belgrave D, Sperrin M, Buchan I, Custovic A. Asthma phenotypes in childhood. *Expert Rev Clin Immunol* 2017; 13: 705-713.
6. Howard R, Rattray M, Prosperi M, Custovic A. Distinguishing Asthma Phenotypes Using Machine Learning Approaches. *Current allergy and asthma reports* 2015; 15: 38.
7. Saglani S, Custovic A. Childhood Asthma: Advances Using Machine Learning and Mechanistic Studies. *Am J Respir Crit Care Med* 2018.
8. Oksel C, Granell R, Mahmoud O, Custovic A, Henderson AJ, Stelar, Breathing Together i. Causes of variability in latent phenotypes of childhood wheeze. *J Allergy Clin Immunol* 2018.
9. 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. Wheezing phenotypes in young children: an historical cohort study. *Primary Care Respiratory Journal* 2014; 23: 60-66.
11. Lowe LA, Simpson A, Woodcock A, Morris J, Murray CS, Custovic A. Wheeze phenotypes and lung function in preschool children. *American Journal of Respiratory and Critical Care Medicine* 2005; 171: 231-237.
12. Henderson J, Granell R, Heron J, Sherriff A, Simpson A, Woodcock A, Strachan DP, Shaheen SO, Sterne JA. Associations of wheezing phenotypes in the first 6 years of life with atopy, lung function and airway responsiveness in mid-childhood. *Thorax* 2008; 63: 974-980.
13. Depner M, Fuchs O, Genuneit J, Karvonen AM, Hyvärinen A, Kaulek V, Roduit C, Weber J, Schaub B, Lauener R. 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, Slama R, Jarvis D, Anto JM, Kauffmann F. Identifying adult asthma phenotypes using a clustering approach. *European Respiratory Journal* 2011: erj01208-02010.

15. Spycher BD, Silverman M, Pescatore AM, Beardsmore CS, Kuehni CE. Comparison of phenotypes of childhood wheeze and cough in 2 independent cohorts. *Journal of allergy and clinical immunology* 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. *Journal of Allergy and Clinical Immunology* 2016; 138: 1060-1070. e1011.

17. Golding J, Pembrey M, Jones R. ALSPAC-the Avon longitudinal study of parents and children. I. Study methodology. *Paediatric and perinatal epidemiology* 2001; 15: 74-87.

18. Atkinson W, Harris J, Mills P, Moffat S, White C, Lynch O, Jones M, Cullinan P, Taylor AN. Domestic aeroallergen exposures among infants in an English town. *European Respiratory Journal* 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. The national asthma campaign Manchester asthma and allergy study. *Pediatric Allergy and Immunology* 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. *American journal of respiratory and critical care medicine* 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, Taussig LM, Wright AL, Martinez FD. 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 ½ years. *International journal of epidemiology* 2001; 30: 1473-1484.

26. Valk R, Caudri D, Savenije O, Koppelman GH, Smit HA, Wijga AH, Postma DS, Kerkhof M, Brunekreef B, Jongste J. Childhood wheezing phenotypes and FeNO in atopic children at age 8. *Clinical & Experimental Allergy* 2012; 42: 1329-1336.

27. Lange P, Celli B, Agusti A, Boje Jensen G, Divo M, Faner R, Guerra S, Marott JL, Martinez FD, Martinez-Camblor P, Meek P, Owen CA, Petersen H, Pinto-Plata V, Schnohr P, Sood A, Soriano JB, Tesfaigzi Y, Vestbo J. 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, Carr TF, Guerra S, Morgan WJ, Wright AL, Martinez FD. 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, Bowatte G, Gurrin L, Johns DP, Thompson BR, Hamilton GS, Frith PA, James AL, Thomas PS, Jarvis D, Svanes C, Russell M, Morrison SC, Feather I, Allen KJ, Wood-Baker R, Hopper J, Giles GG, Abramson MJ, Walters EH, Matheson MC, Dharmage SC. 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 Souef PN, Simpson A, Henderson AJ, Custovic A. 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.

Figure Legends:

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

			Most likely Wheezing Phenotypes				
			Never/ Infrequent wheeze	Early−onset pre−school remitting	Early−onset mid−childhood remitting	Persistent	Late onset
SEATON, N (%)		499	305 (61%)	108 (22%)	16 (3%)	31 (6%)	39 (8%)
ALSPAC, N (%)		5149	3119 (61%)	1058 (20%)	476 (9%)	253 (5%)	243 (5%)
MAAS, N (%)		667	421 (63%)	96 (14%)	34 (5%)	84 (13%)	32 (5%)
ASHFORD, N (%)		492	233 (47%)	176 (36%)	28 (6%)	50 (10%)	5 (1%)
IOW, N (%)		912	548 (60%)	46 (5%)	37 (4%)	66 (7%)	215 (24%)
5−COHORT DATA, N (%)		7719	4626 (60%)	1484 (19%)	591 (8%)	484 (6%)	534 (7%)
Gender	Male, N (%)	3881 (50%)	2162 (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 (≤2.5)	Yes, N (%)	314 (4.1%)	161 (3%)	63 (4%)	36 (6%)	31 (6%)	23 (4%)
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 (5.07)
Advanced MA (≥35)	Yes, N (%)	947 (12%)	572 (13%)	177 (12%)	72 (12%)	53 (11%)	73 (14%)
Maternal smoking	Yes, N (%)	6491 (16%)	647 (14%)	249 (17%)	117 (20%)	88 (18%)	97 (18%)
Paternal smoking	Yes, N (%)	1702 (27%)	979 (26%)	306 (25%)	144 (31%)	132 (32%)	141 (29%)
Pet ownership	Yes, N (%)	4149 (54%)	2449 (54%)	818 (56%)	334 (58%)	264 (55%)	284 (54%)
Sensitization (mid−childhood)	Yes, N (%)	1042 (17%)	461 (13%)	134 (12%)	106 (23%)	181 (47%)	160 (39%)
Current asthma (last follow−up)	Yes, N (%)	1000 (14%)	121 (3%)	113 (8%)	86 (15%)	352 (75%)	328 (67%)
Asthma ever	Yes, N (%)	1737 (26%)	452 (11%)	291 (24%)	274 (59%)	374 (91%)	346 (72%)
Asthma med. (last follow−up)	Yes, N (%)	949 (13%)	120 (3%)	113 (8%)	87 (15%)	345 (73%)	284 (58%)
Asthma med. ever	Yes, N (%)	2125 (42%)	591 (14%)	294 (20%)	374 (64%)	451 (93%)	415 (89%)

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** asthma	Asthma ever	Current** asthma med.	Asthma med. ever	Eczema ever
Never/Infrequent	Reference	Reference	Reference	Reference	Reference
Early−onset pre−school remitting	2.34	2.05	2.35	1.36	1.24
(95%CI)	(1.73, 3.17)	(1.70, 2.46)	(1.74, 3.18)	(1.14, 1.62)	(1.07, 1.42)
p value	*<.0001*	*<.0001*	*<.0001*	*0.001*	*0.003*
Early−onset mid−childhood remitting	3.65	6.41	3.53	5.97	1.31
(95%CI)	(2.55, 5.21)	(5.11, 8.03)	(2.47, 5.05)	(4.85, 7.35)	(1.06, 1.61)
p value	*<.0001*	*<.0001*	*<.0001*	*<.0001*	*0.010*
Persistent	48.31	37.95	42.45	38.67	3.22
(95%CI)	(35.87, 65.07)	(27.78, 51.84)	(31.57, 57.09)	(28.27, 52.90)	(2.55, 4.08)
p value	*<.0001*	*<.0001*	*<.0001*	*<.0001*	*<.0001*
Late onset	36.39	12.00	24.51	17.75	2.03
(95%CI)	(26.89, 49.25)	(9.41, 15.31)	(18.07, 33.23)	(13.59, 23.17)	(1.62, 2.53)
p value	*<.0001*	*<.0001*	*<.0001*	*<.0001*	*<.0001*

*****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.

	Mean Difference (95%CI)* ***P value***		
	Z scores for FEV_1**	Z scores for FVC**	Z scores for FEV_1/FVC**
Never/ Infrequent	reference	reference	Reference
Early−onset pre−school remitting	−0.11 (−0.17, −0.04) 0.002	−0.05 (−0.12, 0.02) 0.144	−0.09 (−0.16, −0.02) 0.001
Early−onset mid−childhood remitting	−0.19 (−0.28, −0.09) <.0001	−0.03 (−0.13, 0.06) 0.531	−0.26 (−0.35, −0.16) <.0001
Persistent	−0.34 (−0.44, −0.24) <.0001	0.03 (−0.07, 0.13) 0.577	−0.60 (−0.70, −0.50) <.0001
Late onset	−0.22 (−0.32, −0.12) <.0001	−0.04 (−0.14, 0.06) 0.406	−0.30 (−0.40, −0.20) <.0001

* Adjusted for maternal history of asthma or allergy, maternal smoking and low birth weight
** Sex−, age−, and height−adjusted SD units

Figure 1

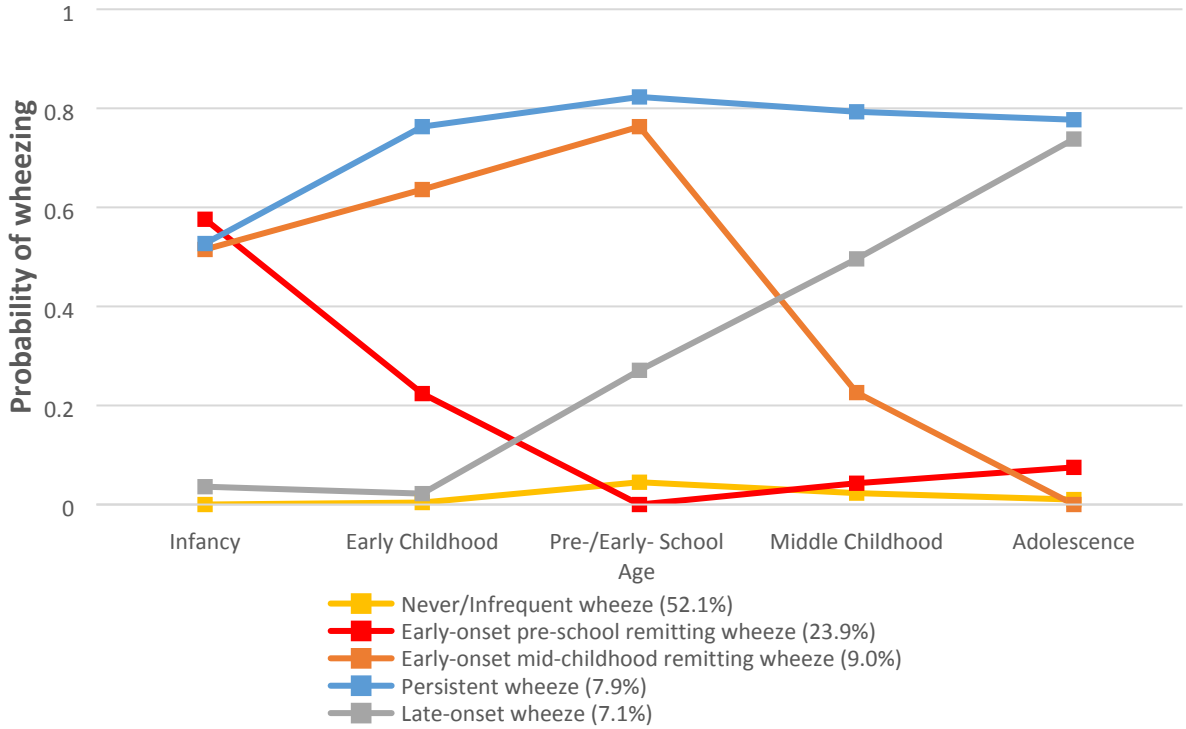
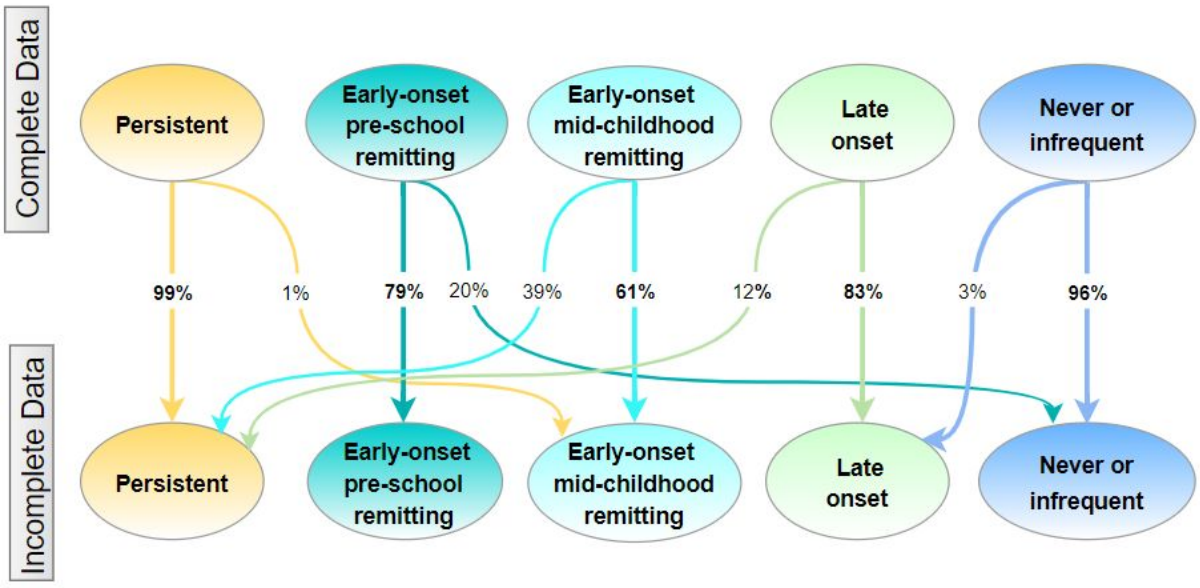

1
0.8
0.6
0.4
0.2
0
Probability of wheezing
Infancy
Early Childhood
Pre-/Early- School Age
Middle Childhood
Adolescence
Never/Infrequent wheeze (52.1%)
Early-onset pre-school remitting wheeze (23.9%)
Early-onset mid-childhood remitting wheeze (9.0%)
Persistent wheeze (7.9%)
Late-onset wheeze (7.1%)

Figure 2

A


Complete Data
Persistent
Early-onset pre-school remitting
Early-onset mid-childhood remitting
Late onset
Never or infrequent
99%
1%
79%
20%
39%
61%
12%
83%
3%
96%
Incomplete Data
Persistent
Early-onset pre-school remitting
Early-onset mid-childhood remitting
Late onset
Never or infrequent

B

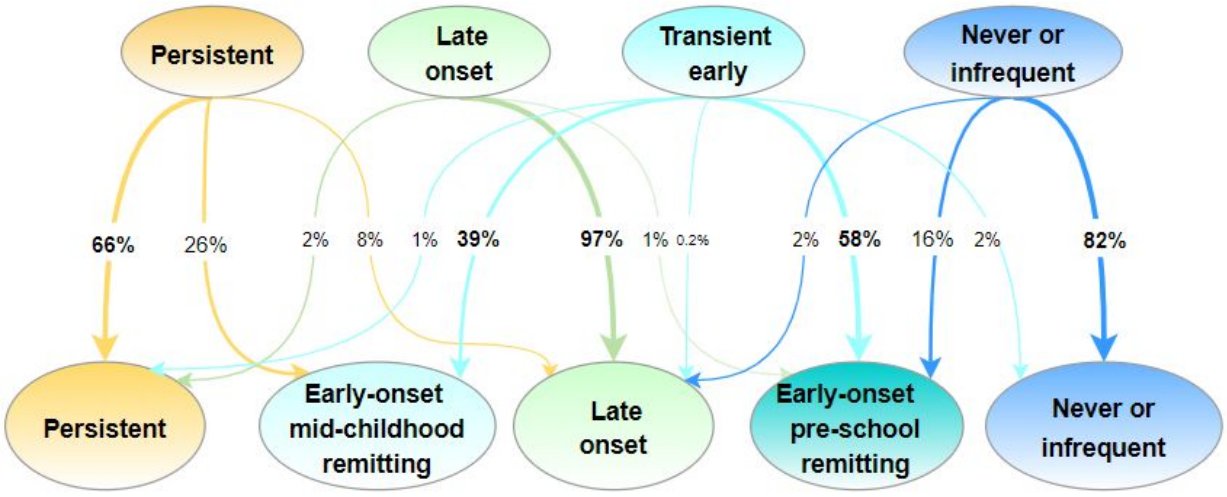
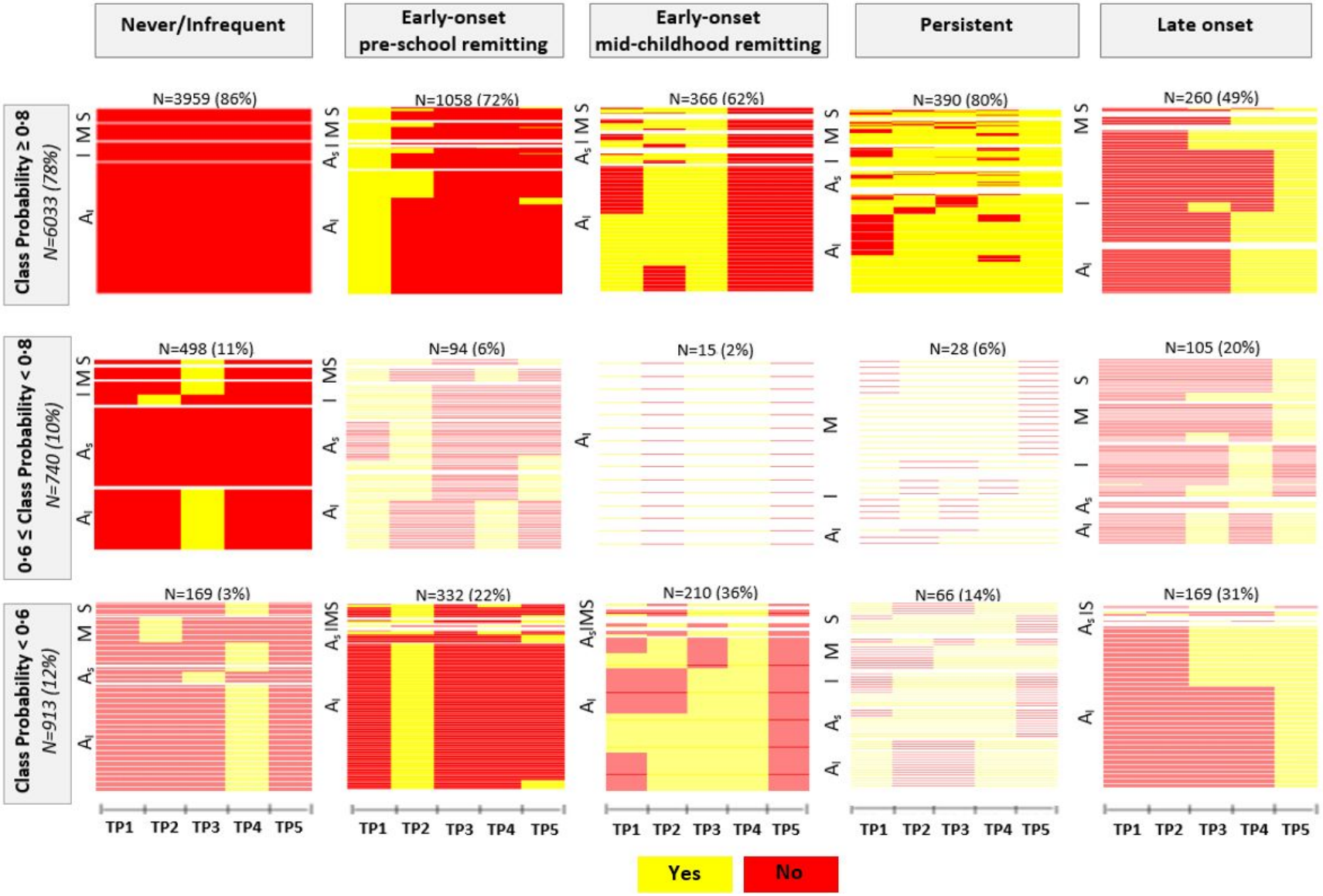

Persistent
Late onset
Transient early
Never or infrequent
66%
26%
2%
8%
1%
39%
97%
1%
0.2%
2%
58%
16%
2%
82%
Persistent
Early-onset mid-childhood remitting
Late onset
Early-onset pre-school remitting
Never or infrequent

Figure 3


Never/Infrequent
Early-onset pre-school remitting
Early-onset mid-childhood remitting
Persistent
Late onset
Class Probability ≥ 0·8 N=6033 (78%)
0·6 ≤ Class Probability < 0·8 N=740 (10%)
Class Probability < 0·6 N=913 (12%)
N=3959 (86%)
N=1058 (72%)
N=366 (62%)
N=390 (80%)
N=260 (49%)
N=498 (11%)
N=94 (6%)
N=15 (2%)
N=28 (6%)
N=105 (20%)
N=169 (3%)
N=332 (22%)
N=210 (36%)
N=66 (14%)
N=169 (31%)
TP1 TP2 TP3 TP4 TP5
Yes
No

ONLINE DATA SUPPLEMENT FOR:

Distinguishing wheezing phenotypes from infancy to adolescence: a pooled analysis of five birth cohorts

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

Description of cohorts

The Isle of Wight (IOW) is a birth cohort study established in 1989 in the UK. It included 1 456 children born between 1st January 1989 and 28th February 1990. Participants were recruited prenatally and followed up prospectively. The Skin Prick Test (SPT) was performed on 980, 1036 and 853 participants at 4, 10 and 18 years of age to check allergic reactions to common allergens. At 10 and 18 years, spirometry and methacholine challenge tests were performed to diagnose lung problems.

The Avon Longitudinal Study of Parents and Children (ALSPAC) is a birth cohort study established in 1991 in Avon, UK. It included 14,701 children born between 1st April 1991 and 31st December 1992 who were alive at 1 year of age. Participants were recruited prenatally and followed up prospectively. Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and Local Research Ethics Committees. The study Web site contains details of all the data that are available through a fully searchable data dictionary at www.bris.ac.uk/alspac/researchers/data-access/data-dictionary/. SPT was carried out at 7−8 years of age. Spirometry test was conducted at $8^{1/2}$ and 15 years while the methacholine challenge test was performed at $8^{1/2}$ years to assess airway responsiveness.

The Ashford study is a birth cohort study established in 1991 in Ashford, UK. It included 642 children born between 1992 and 1993. Participants were recruited prenatally and followed up for 14 years. SPT was performed on 551, 548 and 420 children at 5, 8 and 14 years of ages. At 5 and 8 years, a spirometry test was performed to assess lung function. At the 14 year follow−up, the Exhaled Nitric Oxide (ENO) test was carried out to measure the concentration of nitric oxide in exhaled breath.

The Manchester Asthma and Allergy Study (MAAS) is a birth cohort study established in 1995 in Manchester, UK. It included 1186 children born between February 1996 and April 1998. Participants were recruited prenatally and followed up for 16 years. Atopic sensitization was ascertained by skin prick testing and measurement of sIgE at each clinical follow−up to a mix of common inhalant and food allergens. At age 3−16 years, specific airway resistance (sRaw) was

assessed using whole−body plethysmography. Additionally, the spirometry test was conducted at 5, 8, 11 and 16 years.

The Study of Eczema and Asthma to Observe the influence of Nutrition (SEATON) is a birth cohort study established in 1997 in Aberdeen, UK. It included 1 924 children born between April 1998 and December 1999. Participants were recruited prenatally and followed up by self−completion questionnaire at ½, 1, 2, 5, 10 and 15 years of age. SPT to common allergens (egg, cat dander, grass pollen and house dust mite) was performed at 5, 10 and 15years. At 10 and 15 years of age, study children were invited to attend for a clinical investigation that included spirometry and measurement of exhaled nitric oxide.

Definitions of variables (demographic, exposures and outcomes)

Postal questionnaires were used in ALSPAC and SEATON, while interviewer−administered questionnaires were employed in other cohorts.

Parental history of asthma, eczema and hay fever was defined based on the responses given to the question “have you (and/or your partner) ever had asthma/eczema/hay fever”. Maternal and paternal smoking were defined based on the response given to the question “do you (or does your partner) smoke”, administered during pregnancy. Low birth weight was defined as birth weight less than 2500 g based on NHS birth records.

Information on asthma was obtained from the responses given to the question “Has your child ever suffered from asthma” combined with current asthma questions such as “Has your child had asthma during the past 12 months” at some point during adolescence. Based on the responses, children were divided into two groups: children who had asthma in past (responded “yes” to at least one asthma question) and children who never had asthma (responded “no” to all asthma questions available).

Information about the use of asthma medication during adolescence was obtained from parent’s reports of whether their child had used any medication and/or received any treatment for asthma in the past 12 months. Based on the responses, children were divided into two groups: children who received asthma treatment (responded “yes” to at least one asthma medication

question) and children who never received asthma treatment (responded “no” to all asthma medication questions available).

Information on parent−reported eczema during adolescence obtained from the responses given to the question “has your child ever had eczema” combined with current eczema questions. Accordingly, participants were divided into two groups: children who had eczema in past (responded “yes” to at least one eczema question) and children who never had eczema (responded “no” to both eczema questions).

Early−life risk factors were divided into four groups according to timing of exposure; maternal and child characteristics (gender, maternal smoking during pregnancy and maternal history of asthma), perinatal (low birth weight adjusted for gestational age), environmental (pet ownership, smoke exposure after birth) and allergic sensitization (defined based on positive skin prick test to cat, house dust mite or grass) variables.

Data analysis

Latent Class Analysis

We used LCA to identify longitudinal trajectories of wheeze based on pooled analysis among children with complete data on wheezing at five time periods that were approximately shared across all cohorts. To control for cohort−specific variation, Cohort ID was included in the LCA model as an additional predictor by transforming the 5−category variable into a set of four dummy variables and including them as covariates. The largest cohort, ALSPAC, was treated as the non−coded category to which all other cohorts were compared. The expectation maximization algorithm was used to estimate relevant parameters, with 100,000 iterations and 500 replications.

Model Selection

To assess model fit, we used (1) the Bayesian information criterion (BIC), (2) the Akaike information criterion (AIC), (3) Lo−Mendel−Rubin likelihood ratio test (LMR), (4) Bootstrapped likelihood ratio and, (4) quality of classification certainty (model entropy). The BIC is an index

used in Bayesian statistics to choose among a set of competing models; the model with the lowest BIC is preferred. Using the lowest BIC as a selection criterion, the best fitting model was chosen as the five-class solution with a nominal covariate (BIC:31340).

Stability and heterogeneity of class allocation

We then repeated latent class analyses among 15,941 children with at least two observations and examined how participants with complete data (n=7719) transitioned across classes when 15,941 were included in the analyses. To understand how children, move between phenotypes when the number of phenotypes increases, we compared two LCA models over a sequence of 4 to 5 classes. We used a horizontal line plot for sorting individuals assigned to each phenotype in the vertical space and stratified them by class membership probability and cohort type to reveal individual differences in wheezing trajectories within and across phenotypes. Three class−membership certainty thresholds were defined based on individuals' posterior probabilities of membership in each latent class (≥0.8, 0.8−0.6, <0.6) to elicit the extent of within−class homogeneity in different certainty thresholds.

Validation of the phenotypes

We tested the validity of derived phenotypes by examining their relationships parent-reported asthma, asthma medication use and lung function (height−, age− and sex−adjusted Z−scores for 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 maternal history of asthma/allergy, maternal smoking and low birth weight. Early−life risk factors were assessed for each phenotype using (weighted) multinomial logistic regression models; results are reported as relative risk ratios (RRR) with 95% confidence intervals (CI). We repeated all analyses in children with available data at two or more-time periods (n=15,941). Analyses were carried out using Mplus 8,[1] R (http://www.r−project.org/) and Stata 14 (StataCorp, College Station, Tex).

RESULTS

Table E1. The time period and size of data included in the analyses

Birth Cohort:	IOW	MAAS	SEATON	ASHFORD	ALSPAC
Year of birth	1989	1995	1997	1992	1991
Questionnaire	Interviewer −administered	Interviewer −administered	Postal	Interviewer −administered	Postal
Data collection age (years)	1, 2, 4, 10, 18	1, 3, 5, 8, 16	1, 2, 5, 10, 15	1, 2, 5, 8, 14	½, $2^{1/2}$, $4^{3/4}$, $8^{1/2}$, 14
No. of children with complete data on wheezing at five selected time points	912	667	499	492	5149
No. of children with $\geq$*2 observations on wheezing at five selected time points*	1460	1150	1535	620	11176
SPT test Age (y)	1*, 2*, 4, 10, 18	1*, 3, 5, 8, 11, 16	5, 10, 15	5, 8, 14	5*, $7^{1/2}$
Lung function Measurement Age (y)	10, 18	3, 5, 8, 11, 16	5, 10, 15	5, 8, 14	8, 15

**A subset of participants only*

Table E2. Definition of variables

Variable	Definition
	Cohort: IOW
Wheezing	Presence of wheeze since previous review (year 1, 2, 4, 10 and 18)
Mother−asthma	Do you or have you suffered from asthma or wheezing (recruitment)
Mother−eczema	Do you or have you suffered from eczema (recruitment)
Mother−hay fever	Do you or have you suffered from hay fever or allergic rhinitis (recruitment)
Father−asthma	Does your partner or has your partner suffered from asthma or wheezing (recruitment)
Father−eczema	Does your partner or has your partner suffered from eczema (recruitment)
Father−hay fever	Does your partner or has your partner suffered from hay fever or allergic rhinitis (recruitment)
Mother smoking	Do you smoke in the house? (recruitment)
Father smoking	Does your partner smoke in the house? (recruitment)
Pet	Do you have any pets in the house? (recruitment)
Asthma	Child ever had asthma (10 year and 18 year Questionnaires) combined
Eczema	Child ever had eczema (18 year Questionnaire) combined with eczema questions being asked at year 1, 2, 4, 10 and 18
Asthma Medication	Child ever had asthma treatment (18 year Questionnaire) combined with asthma treatment questions being asked at year 1, 2, 4, 10 and 18
	Cohort: ASHFORD
Wheezing	Which one best describes your child's wheeze in past 12 months? ‘Yes’ (B:1−6, C:7+), ‘No’ (A:0) (year 1, 2, 5, 8 and 14)
Mother−asthma	Do you have or have you ever been told you to have asthma? (recruitment)
Mother−eczema	Do you have or have you ever been told you to have eczema? (recruitment)
Mother−hay fever	Do you have or have you ever been told you to have hay fever? (recruitment)
Father−asthma	The father has had or has ever been told to have asthma? (recruitment)
Father−eczema	The father has had or has ever been told to have eczema? (recruitment)
Father−hay fever	The father has had or has ever been told to have hay fever? (recruitment)
Mother smoking	Do you smoke cigarettes? (recruitment)
Father smoking	Does your partner smoke cigarettes? (recruitment)
Pet	Do you have any pets in the home? (recruitment)
Asthma	Has she/he ever suffered from asthma? (year 1, 2, 3, 14) combined with In the past twelve months has your daughter suffered from eczema? (year 4, 5 and 8)
Eczema	Has she/he ever suffered from eczema? (year 1, 2, 3, 4, 5, 8 and 14 combined)
Asthma Medication	Over the last twelve months has your daughter taken any of the following treatments for asthma? (year 5, 8, 14)
	Cohort: ALSPAC
Wheezing	Two questions combined: Occurrence of wheezing and wheezing with whistling on the chest (month 6, 30, 57, 103, 166)
Mother−asthma	Have you ever had asthma? (recruitment)
Mother−eczema	Have you ever had eczema? (recruitment)
Mother−hay fever	Have you ever had hay fever? (recruitment)
Father−asthma	Partner ever had asthma? (recruitment)
Father−eczema	Partner ever had eczema? (recruitment)

Father−hay fever	Partner ever had hay fever? (recruitment)
Mother smoking	Mother smoked when expecting (recruitment)
Father smoking	Partners daily nicotine intake (recruitment)
Pet	Do you have any pets (recruitment)
Asthma	Asthma ever (month 198)
Eczema	Eczema ever (month 198)
Asthma Medication	Asthma medication at months 77, 91, 103, 128, 140, 166 and 198 combined
Cohort: MAAS	
Wheezing	Has your child had wheezing or whistling in the chest in the last 6/12 months (year 1, 3, 5, 8 and 16)
Mother−asthma	Has a doctor ever told you that you had asthma? (recruitment)
Mother−eczema	Has a doctor ever told you that you had eczema? (recruitment)
Mother−hay fever	Has a doctor ever told you that you had eczema? (recruitment)
Father−asthma	Has a doctor ever told you that you had asthma? (recruitment)
Father−eczema	Has a doctor ever told you that you had eczema? (recruitment)
Father−hay fever	Has a doctor ever told you that you had eczema? (recruitment)
Mother smoking	Do you smoke− mother (recruitment)
Father smoking	Do you smoke− father (recruitment)
Pet	Do you own a pet? (recruitment)
Asthma	Has your child ever suffered from asthma (year 8, 11 and 16 combined)
Eczema	Has your child ever suffered from eczema (year 8 and 11) combined with current eczema (year 16)
Asthma Medication	Asthma medication (year 1, 3, 5, 8, 11 and 16 combined)
Cohort: SEATON	
Wheezing	Has your child had wheezing in the chest in the last 12 months (year 1, 2, 5, 10 and 15)
Mother−asthma	Do you suffer from asthma? (recruitment)
Mother−eczema	Do you suffer from eczema? (recruitment)
Mother−hay fever	Do you suffer from hay fever? (recruitment)
Father−asthma	Does your baby's father suffer from asthma? (recruitment)
Father−eczema	Does your baby's father suffer from eczema? (recruitment)
Father−hay fever	Does your baby's father suffer from hay fever? (recruitment)
Mother smoking	Which of the following best describes your smoking status? (recruitment)
Father smoking	Who smokes in home (6 months)
Pet	Dog, cat and pet in the home combined (6 months)
Asthma	"Has your child ever suffered from asthma?" (year 10) combined with "Have you ever suffered from asthma?" (year 15)
Eczema	"Has your child ever suffered from eczema?" (year 10) combined with "Have you ever suffered from eczema?" (year 15)
Asthma Medication	Has your child (year 5 and 10) or Have you been (year 15) prescribed medicines/inhalers for asthma in the last 12 months

Table E3. Model fit statistics

Model selection in LCA is an area of active research; in accordance with recommendations[2], the best fitting model in each run was selected based on the lowest BIC.

K−Class*	BIC	Entropy	Lo−Mendell−Rubin	Parametric Bootstrapped LRT
			pvalue	pvalue
			(k−1 vs k classes)	(k−1 vs. k classes)
LCA model without covariates in 7719 children				
3	31610	0.8	0.0000	0.0000
4	31456	0.8	0.0000	0.0000
5	31506	0.8	0.4638	0.6667
6	31559	0.7	0.9693	1.0000
LCA model with a nominal covariate (cohort ID) in 7719 children				
3	31597	0.8	0.0000	0.0000
4	31379	0.7	0.0000	0.0000
5	31340	0.7	0.0000	0.0000
6	31367	0.7	0.0000	0.0000
LCA model with a nominal covariate (cohort ID) in 15,941 children (with incomplete data)				
3	56658	0.7	0.0000	0.0000
4	56416	0.6	0.0000	0.0000
5	56337	0.6	0.0000	0.0000
6	56346	0.7	0.0000	0.0000
LCA model with a nominal covariate (cohort ID) in 2570 children (without ALSPAC)				
3	11096	0.8	0.0028	0.0000
4	10975	0.8	0.0000	0.0000
5	11015	0.7	0.1474	0.0000
6	11058	0.7	0.0672	0.0000

*Equality constraints on some of the parameters were imposed to ensure model identifiability.

Table E4. Associations of wheezing phenotypes with early−life risk factors and skin test responses: results from univariable multinomial logistic regression using children with complete wheeze data (reference class: never/infrequent wheeze)

	Unadjusted Relative Risk Ratio (95%CI)			
	Early−onset pre−school remitting	Early−onset mid−childhood remitting	Persistent	Late onset
Gender (Male)	1.28 (1.17, 1.39)	1.50 (1.34, 1.67)	1.61 (1.44, 1.78)	1.15 (0.97, 1.33)
P value	*<.0001*	*<.0001*	*<.0001*	*0.130*
Low birth Weight	1.12 (0.83, 1.40)	1.47 (1.10, 1.85)	1.94 (1.58, 2.30)	1.09 (0.62, 1.55)
P value	*0.451*	*0.044*	*0.0003*	*0.728*
Mother Smoking	1.14 (0.99, 1.30)	1.50 (1.25, 1.67)	1.36 (1.14, 1.58)	1.31 (1.07, 1.54)
P value	*0.090*	*0.0004*	*0.008*	*0.027*
Pet at home	1.042 (0.93, 1.15)	1.15 (0.99, 1.32)	1.092 (0.90, 1.26)	1.012 (0.83, 1.19)
P value	*0.474*	*0.088*	*0.315*	*0.897*
Mother− asthma	1.50 (1.33, 1.68)	2.03 (1.80 2.26)	3.09 (2.87, 3.31)	1.77 (1.51, 2.03)
P value	*<.0001*	*<.0001*	*<.0001*	*<.0001*
Mother− eczema	1.17 (1.04, 1.31)	1.46 (1.27, 1.65)	1.84 (1.66, 2.03)	1.09 (0.87, 1.31)
P value	*0.022*	*0.0001*	*<.0001*	*0.422*
Mother− hay fever	1.10 (0.97, 1.22)	1.35 (1.17, 1.52)	1.79 (1.61, 1.97)	1.14 (0.95, 1.34)
P value	*0.136*	*0.001*	*<.0001*	*0.182*
Father smoking	0.94 (0.80, 1.08)	1.23 (1.03, 1.42)	1.31 (1.11, 1.51)	1.20 (0.99, 1.41)
P value	*0.367*	*0.043*	*0.010*	*0.088*
Father− asthma	1.23 (1.04, 1.41)	1.65 (1.40, 1.90)	2.62 (2.39, 2.85)	1.51 (1.24, 1.78)
P value	*0.030*	*0.000*	*<.0001*	*0.003*
Father−asthma male only	1.19 (0.91, 1.55)	1.64 (1.16, 2.33)	2.63 (1.92, 3.61)	1.97 (1.36, 2.84)
P value	*0.214*	*0.005*	*<.0001*	*<.0001*
Father−asthma female only	1.29 (1.00, 1.67)	1.71 (1.19, 2.45)	2.69 (1.91,3.80)	1.13 (0.75, 1.70)
P value	*0.048*	*0.004*	*<.0001*	*0.550*
Father− eczema	1.06 (0.88, 1.24)	1.14 (0.88, 1.40)	1.43 (1.18, 1.68)	1.23 (0.96, 1.49)
P value	*0.505*	*0.336*	*0.005*	*0.135*
Father− hay fever	1.07 (0.94, 1.21)	1.19 (0.99, 1.39)	1.56 (1.37, 1.76)	0.94 (0.72, 1.16)
P value	*0.317*	*0.094*	*<.0001*	*0.599*
SPT− cat	0.92 (0.56, 1.27)	2.05 (1.66, 2.43)	8.22 (7.93, 8.51)	4.41 (4.07, 4.75)
P value	*0.631*	*0.0003*	*<.0001*	*<.0001*
SPT− HDM	1.03 (0.77, 1.29)	2.04 (1.74, 2.34)	6.87 (6.63, 7.12)	4.59 (4.32, 4.86)
P value	*0.817*	*<.0001*	*<.0001*	*<.0001*
SPT− grass	0.91 (0.67, 1.15)	1.39 (1.08, 1.69)	3.79 (3.54, 4.03)	2.28 (1.99, 2.56)
P value	*0.450*	*0.035*	*<.0001*	*<.0001*

Table E5. Associations of wheezing phenotypes with early−life risk factors and skin test responses: results from univariable multinomial logistic regression using children with missing wheeze data only (n=8223, children with complete wheeze data are not included)

	Unadjusted Relative Risk Ratio (95%CI)			
	Early−onset pre−school remitting	Early−onset mid−childhood remitting	Persistent	Late onset
Gender (Male)	1.27 (1.15, 1.38)	1.38 (1.23, 1.53)	1.49 (1.34, 1.63)	1.09 (0.92, 1.25)
P value	*0.0001*	*<.0001*	*<.0001*	*0.329*
Low birth Weight	1.17 (0.91, 1.43)	1.52 (1.22, 1.81)	1.53 (1.23, 1.83)	1.05 (0.67, 1.43)
P value	*0.241*	*0.007*	*0.005*	*0.812*
Mother Smoking	1.2 (1.06, 1.34)	1.37 (1.19, 1.54)	1.37 (1.19, 1.55)	1.16 (0.96, 1.36)
P value	*0.012*	*0.001*	*0.001*	*0.142*
Pet at home	0.99 (0.87, 1.11)	1.07 (0.92, 1.22)	0.97 (0.82, 1.12)	0.91 (0.74, 1.07)
P value	*0.884*	*0.375*	*0.650*	*0.246*
Mother− asthma	1.36 (1.18, 1.54)	1.70 (1.48, 1.91)	2.53 (2.33, 2.72)	1.47 (1.22, 1.71)
P value	*0.001*	*<.0001*	*<.0001*	*0.003*
Mother− eczema	1.11 (0.96, 1.26)	1.32 (1.14, 1.51)	1.55 (1.38, 1.73)	1.04 (0.82, 1.26)
P value	*0.189*	*0.003*	*<.0001*	*0.711*
Mother− hay fever	1.1 (0.97, 1.23)	1.44 (1.27, 1.60)	1.69 (1.54, 1.85)	1.05 (0.86, 1.23)
P value	*0.160*	*<.0001*	*<.0001*	*0.637*
Father− asthma	1.15 (0.94, 1.36)	1.37 (1.10, 1.63)	1.73 (1.49, 1.97)	1.18 (0.89, 1.46)
P value	*0.195*	*0.021*	*0.000*	*0.264*
Father− eczema	1.02 (0.80, 1.25)	1.15 (0.87, 1.44)	1.27 (1.01, 1.54)	0.95 (0.63, 1.26)
P value	*0.846*	*0.324*	*0.076*	*0.723*
Father− hay fever	0.99 (0.84, 1.15)	1.10 (0.90, 1.30)	1.20 (1.01, 1.40)	0.89 (0.67, 1.11)
P value	*0.934*	*0.360*	*0.059*	*0.309*
SPT− cat	1.28 (0.86, 1.71)	1.57 (1.08, 2.06)	4.87 (4.51, 5.24)	2.61 (2.13, 3.09)
P value	*0.252*	*0.072*	*<.0001*	*0.0001*
SPT− HDM	0.73 (0.35, 1.10)	3.71 (3.14, 4.27)	4.48 (4.09, 4.86)	1.63 (1.17, 2.08)
P value	*0.096*	*<.0001*	*<.0001*	*0.036*
SPT− grass	0.92 (0.59, 1.26)	1.05 (0.65, 1.45)	2.66 (2.36, 2.97)	2.06 (1.69, 2.44)
P value	*0.649*	*0.799*	*<.0001*	*0.0002*

Table E6. Associations of wheezing phenotypes with early−life risk factors and skin test responses: results from weighted multinomial logistic regression using children with complete data on risk factors (reference class: never/infrequent wheeze)

	(Adjusted) Relative Risk Ratio (95%CI)			
	Early−onset pre−school remitting	Early−onset mid−childhood remitting	Persistent	Late onset
Maternal and child characteristics (adjusted by each other)				
Gender (Male)	1.28 (1.14, 1.43)	1.50 (1.27, 1.76)	1.61 (1.35, 1.91)	1.15 (0.96, 1.38)
P value	*<.0001*	*<.0001*	*<.0001*	*0.126*
Maternal smoking	1.15 (0.98, 1.34)	1.47 (1.19, 1.82)	1.38 (1.10, 1.74)	1.32 (1.04, 1.67)
P value	*0.083*	*<.0001*	*0.005*	*0.024*
Maternal history of asthma	1.51 (1.27, 1.80)	2.05 (1.63, 2.59)	3.12 (2.51, 3.89)	1.79 (1.38, 2.32)
P value	*<.0001*	*<.0001*	*<.0001*	*<.0001*
Perinatal characteristics adjusted by gestational age, maternal and child characteristics				
Low birth weight	1.09 (0.75, 1.61)	1.07 (0.62, 1.84)	1.57 (0.95, 2.61)	1.39 (0.78, 2.48)
P value	*0.640*	*0.805*	*0.080*	*0.266*
Environmental characteristics adjusted by maternal, child, perinatal and env. characteristics				
Cat ownership	0.90 (0.78, 1.04)	0.94 (0.77, 1.16)	0.87 (0.69, 1.09)	1.12 (0.91, 1.37)
P value	*0.151*	*0.595*	*0.222*	*0.301*
Dog ownership	1.03 (0.87, 1.21)	1.25 (1.01, 1.57)	1.02 (0.80, 1.31)	1.12 (0.89, 1.42)
P value	*0.748*	*0.045*	*0.859*	*0.332*
Father smoking	0.94 (0.81, 1.08)	1.17 (0.96, 1.43)	1.29 (1.05, 1.59)	1.17 (0.94, 1.44)
P value	*0.396*	*0.130*	*0.170*	*0.159*
Atopic characteristics adjusted by maternal, child, perinatal and env. characteristics				
Sensitization at 4−7.5y	0.91 (0.73, 1.12)	1.61 (1.24, 2.11)	5.12 (4.04, 6.47)	3.44 (2.69, 4.39)
P value	*0.374*	*<.0001*	*<.0001*	*<.0001*
Sens. to cat at 4−7.5y	0.90 (0.60, 1.37)	2.10 (1.35, 3.26)	7.00 (4.96, 9.86)	3.97 (2.73, 5.77)
P value	*0.636*	*0.001*	*<.0001*	*<.0001*
Sens. to hdm at 4−7.5y	1.00 (0.74, 1.36)	1.83 (1.28, 2.60)	6.23 (4.68, 8.30)	4.05 (3.00, 5.46)
P value	*0.987*	*0.001*	*<.0001*	*<.0001*
Sens. to grass at 4−7.5y	0.87 (0.65, 1.16)	1.24 (0.86, 1.79)	3.57 (2.67, 4.78)	2.18 (1.59, 2.99)
P value	*0.343*	*0.242*	*<.0001*	*<.0001*

Table E7. Adjusted associations of wheezing phenotypes with late−outcomes including asthma ever, asthma treatment ever and eczema ever: results from logistic regression using children with missing wheeze data (children with complete wheeze data are not included)

	Adjusted Odds Ratio (95%Cl)*				
	Current** asthma	Asthma ever	Current** asthma med.	Asthma med. ever	Eczema ever
Never/ Infrequent	Reference	reference	Reference	reference	Reference
Early−onset pre−school remitting	2.19	2.05	2.33	1.27	1.17
(95%Cl)	(1.31, 3.65)	(1.55, 2.71)	(1.32, 4.11)	(1.02, 1.59)	(0.94, 1.45)
P value	*0.003*	*<.0001*	*0.003*	*0.032*	*0.172*
Early−onset mid−childhood remitting	1.33	4.22	1.69	3.40	1.42
(95%Cl)	(0.65, 2.73)	(3.04, 5.87)	(0.80, 3.58)	(2.69, 4.29)	(1.06, 1.91)
P value	*0.433*	*<.0001*	*0.169*	*<.0001*	*0.018*
Persistent	23.02	21.96	23.41	18.21	2.47
(95%Cl)	(14.64, 36.19)	(15.43, 31.27)	(14.40, 38.07)	(13.93, 23.81)	(1.87, 3.25)
P value	*<.0001*	*<.0001*	*<.0001*	*<.0001*	*<.0001*
Late onset	22.25	5.73	17.75	6.58	1.23
(95%Cl)	(14.27, 34.69)	(4.27, 7.69)	(10.91, 28.86)	(5.08, 8.53)	(0.95, 1.60)
P value	*<.0001*	*<.0001*	*<.0001*	*<.0001*	*0.115*

**Adjusted for 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 E8. Characteristics of children with complete report of wheeze at five time−periods vs. children with missing data

	Children with complete wheezing data	**Children with 2-4 wheezing data points**
Male gender	50.3% (3881/7719)	53.0% (4348/8208)
Low birth weight	4.1% (314/7604)	5.6% (439/7842)
Maternal smoking	15.6% (1198/7689)	28.9% (1914/6621)
Maternal asthma	11.9% (915/7666)	13.0% (990/7631)
Maternal eczema	22.1% (1695/7664)	19.5% (1487/7626)
Cohort- ALSPAC	46.1% (5149/11176)	53.9% (6027/11176)
Cohort- MAAS	58.0% (667/1150)	42.0% (483/1150)
Cohort- SEATON	32.5% (499/1535)	67.5% (1036/1535)
Cohort- IOW	62.5% (912/1460)	37.5% (548/1460)
Cohort- ASHFORD	79.3% (492/620)	20.7% (128/620)

LEGEND FOR FIGURES

Figure E1. Comparison of LCA models with 5 cohorts (7719 children), 4 cohorts (2570 children, excluding ALSPAC) and incomplete data (15,941 children) on wheezing

Figure E2. Heatmap showing posterior probabilities (0 to 1) of LCA phenotypes

Values close to 0 and 1 represent individuals whose patterns fit well within the assigned class with a high probability (shown in green), and with a very low probability of belonging to other classes (shown in red). Participants who have low posterior probabilities (≤0·5) for all classes indicates uncertainty in class assignment (shown in yellow/orange).

Figure E1 Comparison of LCA models with 5 cohorts (7719 children), 4 cohorts (2570 children, excluding ALSPAC) and incomplete data (15,941 children) on wheezing

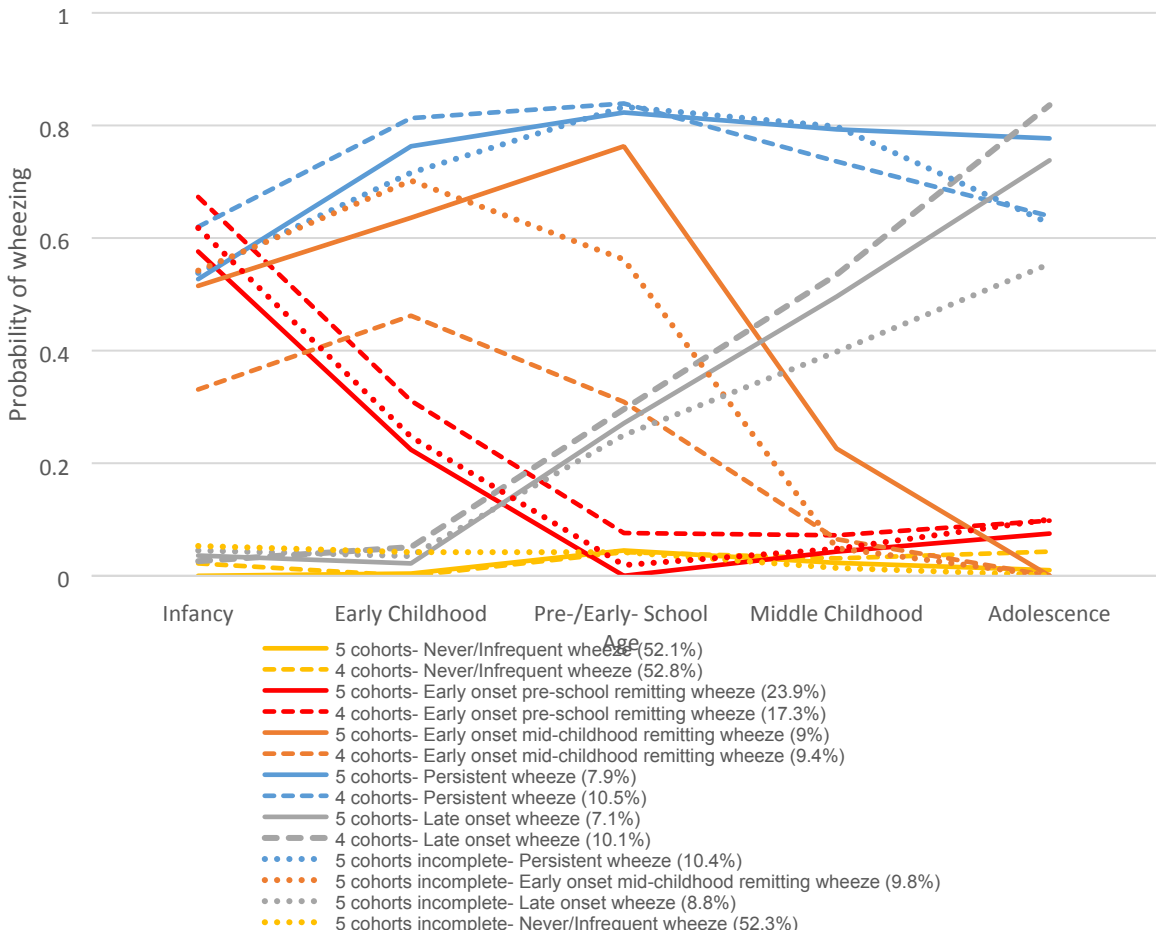
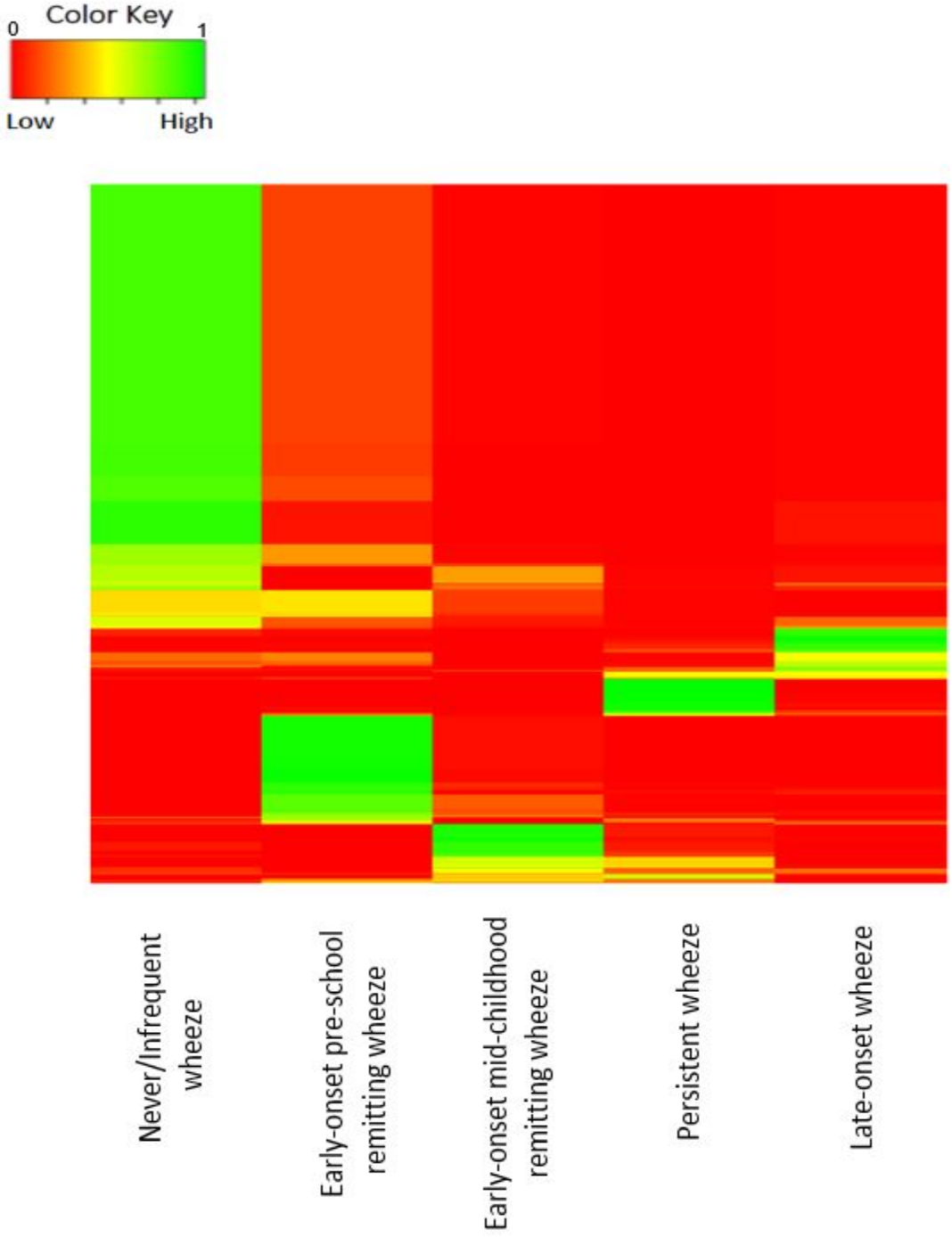


Figure E2. Heatmap showing posterior probabilities (0 to 1) of LCA phenotypes

Values close to 0 and 1 represent individuals whose patterns fit well within the assigned class with a high probability (shown in green), and with a very low probability of belonging to other classes (shown in red). Participants who have low posterior probabilities (≤0·5) for all classes indicates uncertainty in class assignment (shown in yellow/orange).



REFERENCES

1. Muthén LK, Muthén BO. Mplus. Statistical analysis with latent variables. Version 2007; 3.
2. Nylund KL, Asparouhov T, Muthén BO. Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. Structural equation modeling 2007; 14:535-69.