**Causes of variability in latent phenotypes of childhood wheeze**

Ceyda Oksel PhD1§, Raquel Granell PhD2§, Osama Mahmud PhD2, Adnan Custovic MD PhD1§, A. John Henderson MD2§

on behalf of STELAR3 and Breathing Together4 investigators

§Equal contribution

1Section of Paediatrics, Department of Medicine, Imperial College London, UK

2Department of Population Health Sciences, Bristol Medical School, University of Bristol, UK

*Correspondence and requests for reprints:*

Professor Adnan Custovic, Section of Paediatrics, Department of Medicine, Imperial College London, UK.

Tel: +44 20 7594 3274; fax: +44 20 7594 3984;

Email: a.custovic@imperial.ac.uk

*Funding*: CO is funded through the Wellcome Trust Strategic Award 108818/15/Z

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. A comprehensive list of grants funding is available on the ALSPAC website (<http://www.bristol.ac.uk/alspac/external/documents/grant-acknowledgements.pdf>). STELAR cohorts are funded by The UK Medical Research Council (MRC) Grants G0601361 and MR/K002449/1. This research was specifically funded by the Wellcome Trust Strategic Award 108818/15/Z

*Disclosure of potential conﬂict of interest:*The authors declare that they have no conflict of interests.

Abstract word count: 273

Word count: 3286

**ABSTRACT**

**Background:** Latent class analysis (LCA) has been used extensively to identify (latent) phenotypes of childhood wheezing. However, the number and trajectory of discovered phenotypes differed substantially between studies.

**Objective:** To investigate sources of variability affecting the classification of phenotypes, identify key time points for data collection to understand wheeze heterogeneity, and ascertain the association of childhood wheeze phenotypes with asthma and lung function in adulthood.

**Methods:** We used LCA to derive wheeze phenotypes among 3167 participants in the ALSPAC cohort who had complete information on current wheeze recorded at 14 time points from birth to age 16½ years. We examined the effects of sample size, data collection age and intervals on the results, and identified important and redundant time points. We examined the associations of derived phenotypes with asthma and lung function at age 23-24 years.

**Results:** There was a strong interplay between sample size, number of data collection points, and the optimal number of phenotypes identified by LCA. A relatively large sample size (>2000) underestimated the number of phenotypes under some conditions (e.g. number of time points <11). Increasing the number of data points resulted in an increase in the optimal number of phenotypes, but an identical number of randomly selected follow-up points led to different solutions. A variable selection algorithm identified 8 informative time points (months 18, 42, 57, 81, 91, 140, 157 and 166). The proportion of asthmatics at age 23-24 years differed between phenotypes, while lung function was lower among persistent wheezers.

**Conclusions:** Sample size, frequency, and timing of data collection have a major influence on the number and type of wheeze phenotypes identified by LCA in longitudinal data.

**Clinical implications**

The number and the nature of wheeze phenotypes identified by latent class analysis are dependent on the sample size, frequency, timing and distribution of data collection time points, model dimensionality, and combinations of these factors. Key informative time points for follow up are months 18, 42, 57, 81, 91, 140, 157 and 166.

**Capsule Summary**

Understanding the causes of variability in wheeze sub-classification is key to interpreting existing studies, and informing the design of future cohort studies with sufficient power to identify true wheeze phenotypes.

**Keywords**

Childhood asthma,wheeze phenotypes, longitudinal analysis, latent class analysis. ALSPAC

***Abbreviations used***

ALSPAC : Avon Longitudinal Study of Parents and Children

PIAMA : Prevention and Incidence of Asthma and Mite Allergy

TCRS : Tucson Children's Respiratory Study

LCA : Latent class analysis

BIC : Bayesian information criterion

AIC : Akaike information criterion

LMR : Lo–Mendell–Rubin

ARI : Adjusted Rand index

GLI : Global lung function initiative

**INTRODUCTION**

Wheeze is a common symptom in the early years of life, with nearly one third of children experiencing it at least once before their third birthday.1-3 However, wheezing which starts in infancy is often transient, and only a small proportion of wheezy infants continue to wheeze in later childhood and adulthood. Conversely, the majority of patients with persistent asthma start wheezing in early childhood.2 However, at the onset of symptoms, patients with “transient wheeze” and “persistent wheeze” look very similar, and it is difficult to predict which of the early childhood wheezers will stop wheezing (and when), and which will develop persistent wheezing and asthma.

Understanding the heterogeneity of wheezing disorders and distinguishing wheeze phenotypes in early childhood is critical to developing interventions targeted at those who will persist with wheezing into later childhood, and to avoid overtreatment of individuals with transient wheeze.4 Over the last two decades, substantial effort has been devoted to understanding the heterogeneity of childhood wheezing illness (reviewed in5-8). In general, population-based birth cohorts are regarded as the optimal data sources for understanding temporal patterns of wheezing, and relating them to different risk factors, since the information is collected prospectively and therefore free from recall bias.9 The initial approach of hypothesis testing using data on wheezing collected at ages three and six years in the Tucson Children’s Respiratory Study described three wheezing phenotypes (transient early, late-onset and persistent).2 This finding was confirmed in several independent cohorts.3,10,11 Subsequently, the methodology to discover “wheeze phenotypes” was extended to the use of unsupervised, data-driven approaches such as the latent class analysis (LCA).1,12-16 These analyses revealed different structure within the data and suggested the existence of one,17,18 or two further intermediate phenotypes.1,15,16 It is important to emphasize that although “wheeze phenotypes” derived from different analyses tend to share the same nomenclature, phenotypes with the same assignment often differ substantially in terms of the age of onset, temporal trajectory, distributions within a population6 and associated risk factors, making comparison between studies difficult, and clinical application uncertain.6,8 For example, late‐onset wheezers were reported to start experiencing symptoms after age 3years,17 4 years,14 or5 years11 in different studies The inconsistencies between studies may be partly attributed to differences in study design, or could be due to true differences between different populations. However, this seems unlikely, as most evidence comes from broadly similar population-based studies with comparable ethnic mixes.

If we are to understand factors associated with patterns of wheezing with different long-term consequences, then “phenotypes” must be consistent and reproducible. Despite the widespread use of LCA, little is known about the external factors that influence the outcomes of LCA models in phenotype identification. We propose that sample size and the timing and frequency of data collection affect the number and type of discovered wheeze phenotypes in LCA, and that not all time points carry useful information (and therefore some might be redundant, or even cause uncertainty in the results). To provide a better understanding of the influence of input data characteristics on the identified longitudinal trajectories of wheezing, we investigated the effect of the number of data points, age at which information was collected, and sample size on the number and/or the nature of wheeze phenotypes discovered by LCA. We also sought to identify data collection points which are most informative in distinguishing wheeze phenotypes.

**METHODS**

**Study design, setting and participants**

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 age 1 year. Participants were recruited prenatally and followed prospectively. Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and Local Research Ethics Committees. The study website 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/](http://www.bris.ac.uk/alspac/researchers/data-access/data-dictionary/). Details of the study protocol can be found elsewhere.19

**Data sources and definition of outcomes**

Study mothers were sent a self-completion questionnaire about the health of their child at 14 time points from birth to age 16½ years: months 6, 18, 30, 42, 57, 69, 81, 91, 103, 128, 140, 157, 166 and 198. Current wheeze was defined as a positive answer to the question “In the last 12 months has he/she had any periods when there was wheezing or wheezing with whistling on his/her chest when he/she breathed?”.20 A total of 3167 participants had complete reports of wheeze at all 14 time points from birth to age 16½ years.

Study subjects attended research clinic at age 23-24 years in which lung function was measured using spirometry performed according to the ATS/ERS criteria.21,22 Post-bronchodilator FEV1 measurements were taken 15 minutes after administration of 400 mg of salbutamol. We expressed FEV1 as % predicted against the GLI-curves.23 Self-reported asthma ever was defined as a positive answer to the question “Have you ever had asthma?”. Self-reported current asthma was defined at age 23 years as asthma ever together with a positive answer to either “Have you had any wheezing or whistling in the past 12 months?” or “Have you taken asthma medication in the last 12 months?”.

**Statistical analysis**

We performed LCA to investigate how latent class subpopulation structure varied by the timing and frequency of observations, using maternal reports of current wheezing collected at 14 time points from birth to 16½ years of age. Starting with a latent model including 4 phenotypes, we compared models with varying sample sizes (3167, 2500, 2000, 1500, 1000 and 500), number of latent classes (4 to 6) and number of time points (14, 11, 8 and 6) based on their statistical fit, including the Akaike information criterion (AIC), Bayesian information criterion (BIC), Lo–Mendell–Rubin (LMR) and Bootstrapped likelihood ratio, model quality (model entropy) and interpretability. We identified critical data collection points for the identification of distinct phenotypes of wheeze based on stochastic evolutionary search via a genetic algorithm (see Online Repository for more details on the methodology for selection of informative data collection points). The Adjusted Rand Index (ARI) was used as a similarity measure when comparing different clustering results. Variable specific entropy values were used to show how well individual data collection points identify the latent classes. We calculated Confidence Intervals (Cis) for the difference of population proportions to compare the frequency of participants with asthma at age 23 years between different phenotypes. All analyses were performed in Stata v15, Mplus 8, and R using the packages poLCA,24 DiagrammeR and LCAvarsel.25

**RESULTS**

In line with our previous results,15,16 the best-fitting model based on 3167 children with complete reports of wheezing at all 14 time points resulted in 6 distinct wheezing phenotypes: Never/infrequent wheeze; Persistent wheeze; two early-onset transient classes (Early-onset preschool remitting and Early onset mid-childhood remitting); and two late-onset persisting classes (School age-onset and Late childhood-onset).

**Influence of sample size**

We varied the sample size from 3167 to 500, and developed 12 different models based on randomly selected sub-samples of six different sizes (N=500, 1000, 1500, 2000, 2500 and 3167), holding all else constant. Figure 1a shows the best-fitting models based on different sample sizes and the prevalence of each phenotype based on the estimated model. Four phenotypes (Never/infrequent, Persistent, Transient early and Late onset) were identified with a sample size of 500. The best fitting model based on 1000 participants with complete report of wheeze at 14 time points resulted in 4 to 5 phenotypes.

Larger sample sizes (2000 participant or more) were needed to detect smaller phenotypes (<5% frequency). LCA identified six latent wheeze phenotypes in samples of ≥2000 children with complete data. There was no evidence of improved fit for a 6-class solution based on the BIC and entropy when the sample size was increased from 2000 to 3167 (i.e. holding data frequency and interval constant, but varying only the total number). Of note, classification certainty as assessed by relative entropy values (Figure 1a) decreased (from 0.90 to 0.84) as the total sample size increased (from 500 to 3167).

**Influence of data collection frequency**

We then varied the frequency of data collection time points from 6 to the full set of 14, and developed 10 different models based on randomly selected time points while maintaining a constant sample size (N=3167). Adding more time points to the latent model increased the number of wheeze phenotypes that were identified (Figure 1b). However, in some cases, an identical number of (randomly selected) data collection points (e.g. 11 time points) led to the identification of a different optimal number of phenotypes, which shared similar wheeze patterns, depending on the intervals between time points. This suggests that, in addition to sampling frequency, timing and distribution of time points at which data are collected may influence wheeze phenotype identification, and that there might be critical data collection points which are more informative in distinguishing wheeze phenotypes.

**Combined effects of sample size and data collection frequency**

To examine how both the frequency of data collection (number of time points) and the size of the studied population affects the optimal number, trajectory, and frequency of the identified phenotypes, we varied the number of data collection points from 6 to 11 and randomly selected sub-samples of four different sizes, resulting in a total of 12 data conditions (Figure 1c).

Models with small sample sizes (N<2500) did not identify low-frequency phenotypes (<5%), regardless of the frequency of data sampling. However, there was a clear link between sample size, number of data points and the optimal number of wheeze phenotypes identified by the model. The model with sample size of ≥2500 identified 6 phenotypes when the number of data collection points included in the analysis was relatively high. However, models with decreasing number of data points were unable to detect 6 phenotypes, and models with the same sample size did not identify small phenotypes (<5% frequency) under certain conditions (e.g. number of time points <11).

**Selection of the most informative data collection points**

Figure E1 shows the correlation coefficients between wheeze reports at different time points. Time points close to each other were moderately correlated (e.g. month 157 and 166; month 81 and 91 etc.), suggesting that some of the adjacent time points convey similar information. In order to discard the non-informative and redundant data collection points, we performed stochastic evolutionary search via a genetic algorithm, which retained 8 informative time points (months 18, 42, 57, 81, 91, 140, 157, and 166), while 6 were dropped as uninformative (months 6, 30, 69,103, 128 and 198). Comparing the clustering of the models using eight time points to the clustering from the model using the full dataset showed a satisfactory level of agreement, with Rand and Adjusted Rand indices of 82 and 64%, respectively (Table 1).

**Latent transition probabilities with increasing number of classes**

To understand how the trajectories and estimated phenotypes changed over a sequence of increasing number of classes, and how children move from one class to another in models with an increasing number of classes, we developed three LCA models with 4 to 6 classes. Persistent and never/infrequent wheeze classes had similar patterns in all three models, with a slight decrease in estimated prevalence from four to six-class solution (Figure 2 panel A). With the addition of a fifth latent class, Transient-early wheeze divided into two remitting classes (Pre-school and Mid-childhood resolution, Figure 2 panel B), while the late-onset class remained almost identical. The addition of a sixth class resulted in the division of the Late-onset wheeze into two similar-sized sub-groups (School-age and Late childhood onset, Figure 2 panel C). We then assigned participants to the most likely phenotype based on the maximum membership probability, and calculated transition probabilities reflecting the proportion of participants moving from one phenotype to another when the number of phenotypes increased from 4 up to 6. Figure 3 shows whether members of distinct phenotypes remained in the same phenotype or shift into another one (either existing or newly formed) with increasing number of phenotypes. The figure also demonstrates where the intermediate phenotypes arise from, and which phenotypes become separated or remain undivided with increasing number of latent classes.

**Asthma and lung function in adulthood in different wheeze phenotypes**

A total of 3797 participants attended the follow-up at age 23-24 years, with 3038 having valid lung function data. The proportion of subjects reporting asthma at age 23-24 years in each phenotype is shown in Table 2 (restricted to participants with complete reports of wheeze). As expected, the Never/infrequent wheeze had the smallest proportion of participants with self-reported current asthma (5.1%), whereas Persistent wheeze had the highest proportion (99.7%). In the two early-onset transient phenotypes, the proportion of asthmatics was significantly higher in Mid-childhood remitting (60.4%) compared to the Pre-school remitting (6.4%) (Mean difference 0.54, 95%CI 0.40-0.68, p<0.0001). In the two late-onset phenotypes, the proportion of asthmatics was significantly higher in School-age onset (88.4%) compared to Late-childhood onset (68.1%) (Mean difference 0.20, 95%CI 0.05-0.36, p<0.02).

Both pre- and post-bronchodilator lung function was significantly lower in the Persistent wheeze, with little evidence of differences between other wheeze phenotypes (Table 3).

**DISCUSSION**

*Key results*

Our results suggest that the number and the nature of wheeze phenotypes from infancy to adolescence identified by LCA are dependent on several factors including sample size, frequency, timing and distribution of data collection time points, model dimensionality, as well as the combination of these factors. Transition analysis revealed that subjects assigned to Never or Persistent wheeze tend to stay in these phenotypes, whilst most of the switching goes on in the intermediate classes. Given the strong interplay between the birth cohort design (including the number of participants, data collection frequency and distribution) and the optimal number of phenotypes identified by means of developmental trajectory modelling, care should be taken when interpreting wheeze phenotypes emerging from small studies with few data collection points. When the sample size is small, a wheeze phenotype that exists in the population may be unidentifiable, whereas excessive data collection may result in the identification of trivial or clinically-irrelevant phenotypes. In general, increasing data collection frequency helps detect more complex structure and larger number of phenotypes by capturing less-frequently observed subgroups. However, it also increases the risk of violating the fundamental assumption of LCA modelling where indicator variables (e.g. presence/absence of wheeze at subsequent ages) are independent of each other. When frequent data collection and large sample sizes are not obtainable, collecting data at critical time points may help counterbalance the effects of sub-optimal conditions (e.g. smaller sample size and infrequent data collection). In our study, time points which proved most informative in distinguishing wheeze phenotypes were months 18, 42, 57, 81, 91, 140, 157 and 166.

*Limitations*

There are several limitations to our findings. The key one is that we do not know how many true phenotypes there are, and we assumed that the classification obtained on the largest sample and using all time points corresponded to the "true" classification. We were able to look at the associations of wheeze phenotypes with asthma and lung function in early adulthood (age 23-24 years). The proportion of asthmatics differed between intermediate phenotypes (15.1% and 75.3% in two transient early phenotypes, Pre-school remitting and Mid-childhood remitting respectively; 91.3% and 70.0% in two late-onset phenotypes, Late childhood and School-age onset). The proportion of asthmatics was highest in the Persistent wheeze (98.5%), and subjects in this phenotype had diminished pre and post-bronchodilator lung function (at the time of maximally attained physiological lung function plateau26) compared to all other phenotypes. These findings suggest that all phenotypes are distinct and that this may be a true classification. However, we acknowledge that the observed associations may also be a proxy of severity.

Another limitation is that we could only vary conditions using the sampling framework that was available to us, which was fixed by the study design, so this analysis has limited direct application to other studies that have used different sampling frames. We also acknowledge that the definition of current wheeze which we used in our models is based on parental reporting using validated questionnaires (as in most other epidemiological studies) and that this may lead to overestimation of the true prevalence.27

As most previous studies, we used information on current wheeze for our modelling. It is possible that a more holistic examination of other features (e.g. frequency and severity of wheeze) and/or other symptoms (e.g. cough, atopic dermatitis and rhinitis)20 and lung function26 may allow better distinction of the underlying pathophysiological mechanisms.

The key advantage of our study is the large sample size with complete data on wheezing collected frequently and prospectively. Another advantage is that participants were followed from birth to late adolescence, covering a longer period compared to many prior studies.1,11,16,17,28

Finally, it is worth noting that subtypes discovered using data-driven methods 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 have maintained this nomenclature.

*Interpretation*

A number of previous studies (including our own) embarked on identifying wheeze phenotypes from birth to mid-school age based on the temporal pattern of wheezing (summarized in Table E1). However, the inconsistency of ﬁndings has led to a debate on the validity and clinical value of phenotyping studies, hampering the translation of this knowledge into clinically actionable insights and pathophysiological endotypes. The four phenotypes of persistent, never, transient early and late-onset wheeze have been long postulated in descriptive,2 and data-driven studies.29 We found that when the sample size is relatively small, a particular wheeze phenotype that exists in the population may be undetectable. Therefore, relatively smaller sample size in some studies might have contributed to the inability to detect intermediate wheeze phenotypes with a relatively low prevalence. Using more time points allowed the identification of less common phenotypes (<5% frequency) by increasing possible response patterns. When the data collection was frequent (>11 time points), a sample size of ~2500 was found to be sufficiently large to distinguish 6 wheeze phenotypes. However, even a larger sample size of 3167 might be insufficient to detect uncommon wheeze phenotypes (<5% frequency) under certain conditions (e.g. data collection points <11). Our findings suggest that increasing data collection frequency may help compensate for a modest sample size in phenotype identification. In line with this finding, Depner *et al*.28 identified an intermediate phenotype in the Protection against Allergy—Study in Rural Environments (PASTURE) cohort that existed during the first six years of life, using a similar sample size but more data collection points than those used in the Tucson Children's Respiratory Study.2 However, the selection of follow up points needs a careful thought. Our analyses have shown that although adding more time points to the latent model increased the number of identified phenotypes with distinguishable interpretations, in some cases the same number of randomly selected data collection points resulted in a different optimal solution. This suggests that the timing and distribution of follow-ups is important, and that there might be critical data collection points which are more informative than others. A variable selection method which we applied to the data identified 6 time points which were not carrying additional useful information (months 6, 30, 69, 103, 128 and 198).

In conclusion, our findings add to the understanding of childhood wheeze phenotypes by extending the knowledge on potential causes of variability in classification of wheezing. Sample size, frequency, and timing of data collection have a major influence on the number and type of wheeze phenotypes identified by data-driven techniques. This information, which includes information on the most informative follow-up points) is critical to interpret (or reanalyze) existing studies and inform the better design of future cohorts.

**Acknowledgments:** Asthma UK Grant ref. 06/005 funded the collection of the wheezing and asthma data in ALSPAC. 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 ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, and nurses.

**REFERENCES**

1. Henderson J, Granell R, Heron J, et al. Associations of wheezing phenotypes in the first six years of life with atopy, lung function and airway responsiveness in mid childhood. *Thorax* 2008; **63**(11): 974-80.

2. 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**(3): 133-8.

3. Lowe LA, Simpson A, Woodcock A, et al. Wheeze phenotypes and lung function in preschool children. *Am J Respir Crit Care Med* 2005; **171**(3): 231-7.

4. Belgrave D, Henderson J, Simpson A, Buchan I, Bishop C, Custovic A. Disaggregating asthma: Big investigation versus big data. *The Journal of allergy and clinical immunology* 2017; **139**(2): 400-7.

5. Deliu M, Belgrave D, Sperrin M, Buchan I, Custovic A. Asthma phenotypes in childhood. *Expert Rev Clin Immunol* 2017; **13**(7): 705-13.

6. Howard R, Rattray M, Prosperi M, Custovic A. Distinguishing Asthma Phenotypes Using Machine Learning Approaches. *Current allergy and asthma reports* 2015; **15**(7): 38.

7. Just J, Bourgoin-Heck M, Amat F. Clinical phenotypes in asthma during childhood. *Clin Exp Allergy* 2017; **47**(7): 848-55.

8. Belgrave D, Simpson A, Custovic A. Challenges in interpreting wheeze phenotypes: the clinical implications of statistical learning techniques. *Am J Respir Crit Care Med* 2014; **189**(2): 121-3.

9. Custovic A, Ainsworth J, Arshad H, et al. The Study Team for Early Life Asthma Research (STELAR) consortium 'Asthma e-lab': team science bringing data, methods and investigators together. *Thorax* 2015; **70**(8): 799-801.

10. Cano-Garcinuño A, Mora-Gandarillas I. Wheezing phenotypes in young children: an historical cohort study. *Primary Care Respiratory Journal* 2014; **23**: 60-6.

11. 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**(5): 573-8.

12. Lazic N, Roberts G, Custovic A, et al. Multiple atopy phenotypes and their associations with asthma: similar findings from two birth cohorts. *Allergy* 2013; **68**(6): 764-70.

13. Spycher BD, Silverman M, Brooke AM, Minder CE, Kuehni CE. Distinguishing phenotypes of childhood wheeze and cough using latent class analysis. *European Respiratory Journal* 2008; **31**(5): 974-81.

14. Lodge CJ, Zaloumis S, Lowe AJ, et al. Early-life risk factors for childhood wheeze phenotypes in a high-risk birth cohort. *The Journal of pediatrics* 2014; **164**(2): 289-94. e2.

15. 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**(4): 1060-70. e11.

16. Savenije OE, Granell R, Caudri D, et al. Comparison of childhood wheezing phenotypes in 2 birth cohorts: ALSPAC and PIAMA. *Journal of Allergy and Clinical Immunology* 2011; **127**(6): 1505-12. e14.

17. Belgrave DC, Simpson A, Semic-Jusufagic A, et al. Joint modeling of parentally reported and physician-confirmed wheeze identifies children with persistent troublesome wheezing. *Journal of Allergy and Clinical Immunology* 2013; **132**(3): 575-83. e12.

18. Valk R, Caudri D, Savenije O, et al. Childhood wheezing phenotypes and FeNO in atopic children at age 8. *Clinical & Experimental Allergy* 2012; **42**(9): 1329-36.

19. Boyd A, Golding J, Macleod J, et al. Cohort profile: the ‘children of the 90s’—the index offspring of the Avon Longitudinal Study of Parents and Children. *International journal of epidemiology* 2013; **42**(1): 111-27.

20. Belgrave DC, Granell R, Simpson A, et al. Developmental profiles of eczema, wheeze, and rhinitis: two population-based birth cohort studies. *PLoS Med* 2014; **11**(10): e1001748.

21. Miller MR, Hankinson J, Brusasco V, et al. Standardisation of spirometry. *European respiratory journal* 2005; **26**(2): 319-38.

22. Beydon N, Davis SD, Lombardi E, et al. An official American Thoracic Society/European Respiratory Society statement: pulmonary function testing in preschool children. *American journal of respiratory and critical care medicine* 2007; **175**(12): 1304-45.

23. Quanjer P, Stanojevic S, Stocks J, et al. Changes in the FEV1/FVC ratio during childhood and adolescence: an intercontinental study. *European Respiratory Journal* 2010; **36**(6): 1391-9.

24. Linzer DA, Lewis JB. poLCA: An R package for polytomous variable latent class analysis. *Journal of Statistical Software* 2011; **42**(10): 1-29.

25. Fop M, Smart KM, Murphy TB. Variable selection for latent class analysis with application to low back pain diagnosis. *The Annals of Applied Statistics* 2017; **11**(4): 2080-110.

26. Belgrave DCM, Granell R, Turner SW, 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.

27. Lowe L, Murray CS, Martin L, et al. Reported versus confirmed wheeze and lung function in early life. *Arch Dis Child* 2004; **89**(6): 540-3.

28. Depner M, Fuchs O, Genuneit J, et al. Clinical and epidemiologic phenotypes of childhood asthma. *American journal of respiratory and critical care medicine* 2014; **189**(2): 129-38.

29. Chen Q, Just AC, Miller RL, et al. Using latent class growth analysis to identify childhood wheeze phenotypes in an urban birth cohort. *Annals of Allergy, Asthma & Immunology* 2012; **108**(5): 311-5. e1.

**3*STELAR investigators***

Prof Syed Hasan Arshad

Silvia Colicino

Prof Paul Cullinan

Dr John Curtin

Prof Graham Devereux

Prof John Holloway

Dr Clare S Murray

Prof Graham Roberts

Prof Angela Simpson

Prof Steve Turner

**4*Breathing Together investigators***

Prof Andrew Bush

Dr Peter Ghazal

Prof Jonathan Grigg

Prof Clare M Lloyd

Dr Benjamin Marsland

Dr Ultan Power

Prof Sejal Saglani

Prof Jurgen Schwarze

Prof Mike Shields

**LEGEND FOR FIGURES**

**Figure 1.** The optimal number, shape and prevalence of wheeze phenotypes with:

**a)** Different sample sizes (N=500, 1000, 1500, 2000, 2500 and 3167)

**b)** Different number of data collection points (6, 8, 11 and 14)

**c)** Combined effects of sample size and data collection frequency

**Figure 2.** Estimated prevalence of wheeze for each wheezing phenotype in four-, five- and six- latent class solutions identified by LCA

**Figure 3.** Assignment of children into distinct wheeze phenotypes over a sequence of LC model with four, five and six classes based on most likely class membership (3167 children cohort with complete reports of wheezing at 14 time points). Ellipse nodes show class membership (most likely phenotype) whilst the values along the arrow represent the % of children moving from one class to another in models with an increasing number of classes.