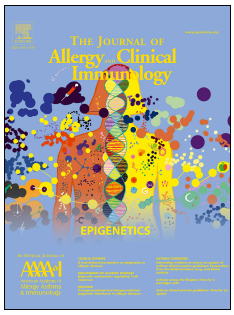
Accepted Manuscript



Causes of variability in latent phenotypes of childhood wheeze

Ceyda Oksel, PhD, Raquel Granell, PhD, Osama Mahmoud, PhD, Adnan Custovic, MD PhD, A. John Henderson, MD, on behalf of STELAR, Breathing Together investigators

PII: S0091-6749(18)31723-8

DOI: https://doi.org/10.1016/j.jaci.2018.10.059

Reference: YMAI 13755

To appear in: *Journal of Allergy and Clinical Immunology*

Received Date: 8 June 2018

Revised Date: 29 August 2018

Accepted Date: 12 October 2018

Please cite this article as: Oksel C, Granell R, Mahmoud O, Custovic A, Henderson AJ, on behalf of STELAR, Breathing Together investigators, Causes of variability in latent phenotypes of childhood wheeze, *Journal of Allergy and Clinical Immunology* (2019), doi: https://doi.org/10.1016/j.jaci.2018.10.059.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Causes of variability in latent phenotypes of childhood wheeze

Ceyda Oksel PhD[1§], Raquel Granell PhD[2§], Osama Mahmoud PhD[2], Adnan Custovic MD PhD[1§], A. John Henderson MD[2§]

on behalf of STELAR[3] and Breathing Together[4] investigators

[§]Equal contribution

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

[2]Department 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, Norfolk Place, London W2 1PG, 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 analysis was 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: 248

Word count: 3537

ACCEPTED MANUSCRIPT

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 time points. We examined the associations of derived phenotypes with asthma and lung function at age 23-24 years.

Results: 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.

Key Messages

- 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.
- Not all data collection points carry useful information in distinguishing wheeze phenotypes.

Capsule Summary

We determined dependence of phenotype discovery on frequency, timing and distribution of data collection, identifying eight informative, time-specific follow-up points from infancy to adolescence.

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] Although the symptoms of most infants with wheeze seem to remit by the time the child reaches school age[4], infantile wheeze may also persist into later childhood and adulthood after a period of remission.[5, 6] 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.[7] Over the last two decades, substantial effort has been devoted to understanding the heterogeneity of childhood wheezing illness (reviewed in[7-10]). 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.[11] The initial approach of hypothesis testing using data on wheezing collected at ages three and six years in the Tucson Children's Respiratory Study (TCRS) described three wheezing phenotypes (transient early, late-onset and persistent).[2] This finding was confirmed in several independent cohorts.[3, 12, 13] 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, 14-18] These analyses revealed

98 different structure within the data and suggested the existence of one,[19, 20] or two further
99 intermediate phenotypes.[1, 17, 18] It is important to emphasize that although "wheeze
100 phenotypes" derived from different analyses tend to share the same nomenclature,
101 phenotypes with the same assignment often differ substantially in terms of the age of onset,
102 temporal trajectory, distributions within a population[8] and associated risk factors, making
103 comparison between studies difficult, and clinical application uncertain.[8, 10] For example, late-
104 onset wheezers were reported to start experiencing symptoms after the age of three,[19] four,[16]
105 or five years[13] in different studies. The inconsistencies between studies may be partly
106 attributed to differences in study design or could be due to true differences between different
107 populations. However, this seems unlikely, as most evidence comes from broadly similar
108 population-based studies with comparable ethnic mixes.

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

119 by LCA. We also sought to identify data collection points which are most informative in
120 distinguishing wheeze phenotypes.

ACCEPTED MANUSCRIPT

METHODS

Study design, setting and participants

The Avon Longitudinal Study of Parents and Children (ALSPAC) is a population-based birth cohort established in 1991 in Avon, UK. It recruited 14,701 children born between 1st April 1991 and 31st December 1992. Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and Local Research Ethics Committees. Details of the study protocol can be found elsewhere.[21] 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/.

Data sources and definition of outcomes

Participating 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?".[22]

Study subjects attended research clinic at age 23-24 years in which lung function was measured using spirometry.[23, 24] Post-bronchodilator FEV_1 was ascertained 15 minutes after administration of 400 mcg of salbutamol. We expressed FEV_1 as % predicted against the GLI-curves.[25] 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

Children with complete reports of wheezing at all 14 time points from birth to age 16½ years (n=3167) were included in the analysis to obtain better representation of the latent structure. We performed LCA to investigate how latent class subpopulation structure varied by the timing and frequency of observations. 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. The best fitting model in each run was selected based on the lowest BIC. We then repeated our analyses among 12,290 participants with at least 2 questionnaire responses. 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. Differences in lung function were tested using one-way ANOVA

162 and Tukey's HSD (honestly significant difference) test. All analyses were performed in Stata
163 v15, Mplus 8, and R using the packages poLCA,[26] DiagrammeR and LCAvarsel.[27]

ACCEPTED MANUSCRIPT

164 **RESULTS**

165 A total of 3167 participants had complete reports of wheeze at all 14 time points. In line with
166 our previous results,[17, 18] the best-fitting model resulted in six distinct wheezing phenotypes:
167 Never/infrequent wheeze; Persistent wheeze; two early-onset transient classes (Early-onset
168 preschool remitting and Early onset mid-childhood remitting); and two late-onset persisting
169 classes (School age-onset and Late childhood-onset).

170 **Influence of sample size**

171 We varied the sample size from 3167 to 500, and developed 11 different models based on
172 randomly selected sub-samples of six different sizes (N=500, 1000, 1500, 2000, 2500 and 3167),
173 holding all else constant. Figure 1a shows the best-fitting models based on different sample
174 sizes and the prevalence of each phenotype based on the estimated model. Four phenotypes
175 (Never/infrequent, Persistent, Transient early and Late onset) were identified with a sample
176 size of 500. The best fitting model based on 1000 participants resulted in four to five
177 phenotypes.

178 Larger sample sizes (2000 participant or more) were needed to detect smaller phenotypes (<5%
179 frequency). LCA identified six latent wheeze phenotypes in samples of ≥2000 children with
180 complete data (Figure 1a), and in samples of ≥5000 children with incomplete data (Figure E2).

181 **Influence of data collection frequency**

182 We then varied the frequency of data collection time points from 6 to 14, and developed 10
183 different models based on randomly selected time points while maintaining a constant sample
184 size (N=3167). Adding more time points to the latent model increased the number of wheeze

185 phenotypes that were identified (Figure 1b). However, in some cases, an identical number of
186 (randomly selected) data collection points (e.g. 11 time points) resulted in different optimal
187 number of phenotypes, depending on the intervals between time points. This suggests that, in
188 addition to sampling frequency, timing and distribution of time points at which data are
189 collected may influence wheeze phenotype identification, and that there might be critical data
190 collection points which are more informative in distinguishing wheeze phenotypes.

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

192 To examine how both the frequency of data collection (number of time points) and the size of
193 the studied population affects the optimal number, trajectory, and frequency of the identified
194 phenotypes, we varied the number of data collection points from 6 to 11 and randomly
195 selected sub-samples of four different sizes, resulting in a total of 12 data conditions (Figure
196 1c).

197 Models with small sample sizes (N<2500) did not identify low-frequency phenotypes (<5%),
198 regardless of the frequency of data sampling. However, there was a clear link between sample
199 size, number of data points and the optimal number of wheeze phenotypes. The model with
200 sample size of ≥2500 identified six phenotypes when the number of data collection points
201 included in the analysis was relatively high. However, models with decreasing number of data
202 points were unable to detect six phenotypes, and models with the same sample size did not
203 identify small phenotypes (<5% frequency) under certain conditions (e.g. number of time points
204 <11).

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

206 Figure E1 shows the correlations (phi coefficients) between wheeze reports at different time
207 points. Time points close to each other were moderately correlated (e.g. month 157 and 166;
208 month 81 and 91 etc.), suggesting that some of the adjacent time points convey similar
209 information. In order to discard the non-informative data collection points, we performed
210 stochastic evolutionary search via a genetic algorithm, which retained 8 informative time points
211 (months 18, 42, 57, 81, 91, 140, 157, and 166), while 6 were dropped as uninformative (months
212 6, 30, 69,103, 128 and 198). Comparing the clustering of the models using eight time points to
213 the clustering from the model using the full dataset showed a satisfactory level of agreement,
214 with Rand and Adjusted Rand indices of 82 and 64%, respectively (Table 1).

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

216 To understand how the trajectories and estimated phenotypes changed over a sequence of
217 increasing number of classes, and how children move from one class to another in models with
218 an increasing number of classes, we developed three LCA models with four, five and six classes.
219 Persistent and never/infrequent wheeze classes had similar patterns in all three models, with a
220 slight decrease in estimated prevalence from four to six-class solution (Figure 2 panel A). With
221 the addition of a fifth latent class, Transient-early wheeze divided into two remitting classes
222 (Pre-school and Mid-childhood resolution, Figure 2 panel B), while the Late-onset wheeze
223 remained almost identical. The addition of a sixth class resulted in the division of the Late-
224 onset wheeze into two similar-sized sub-groups (School-age and Late childhood onset, Figure 2
225 panel C). We then assigned participants to the most likely phenotype based on the maximum
226 membership probability, and calculated transition probabilities reflecting the proportion of

227 participants moving from one phenotype to another when the number of phenotypes increased
228 from four up to six. Figure 3 shows whether members of distinct phenotypes remained in the
229 same phenotype or shift into another one (either existing or newly formed) with increasing
230 number of phenotypes. The figure also demonstrates where the intermediate phenotypes arise
231 from, and which phenotypes become separated or remain undivided with increasing number of
232 latent classes. The results based on analysis of participants with incomplete reports of wheeze
233 (12,290 participants with at least 2 responses to questionnaires about wheezing) did not
234 materially differ from those obtained among children with complete data set and are presented
235 in this article's Online Repository (Figures E2-E6).

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

237 Of 3797 participants who attended age 23-24 years follow-up, 1492 had complete reports of
238 wheezing (14 points), of whom 240 (16%) reported current asthma; 1345 had valid lung
239 function. The proportion of subjects with current asthma was highest in the Persistent wheeze
240 (99.7%), Table 2. In the two early-onset transient phenotypes, the proportion of asthmatics
241 was significantly higher in Mid-childhood remitting (60.4%) compared to the Pre-school
242 remitting (6.4%) (Mean difference 0.54, 95%CI 0.40-0.68, p<0.0001). In the two late-onset
243 phenotypes, the proportion of asthmatics was significantly higher in School-age onset (88.4%)
244 compared to Late-childhood onset (68.1%) (Mean difference 0.20, 95%CI 0.05-0.36, p<0.02).
245 Pre- and post-bronchodilator lung function differed significantly across phenotypes (p=0.005
246 and p=0.04 respectively, ANOVA), and was significantly lower in Persistent wheeze and Early-
247 onset pre-school remitting wheeze compared to Never/infrequent wheeze, with little evidence
248 of differences between other phenotypes (Tables 3, E1-3). The Preschool-onset remitting

249 phenotype mostly overlapped with no asthma (94%), but the pre- and post-bronchodilator lung

250 function at age 24 was significantly lower in this class compared to Never/infrequent wheeze.

ACCEPTED MANUSCRIPT

251 **DISCUSSION**

252 *Key results*

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

Limitations

There are several limitations to our findings. Despite latent models' usefulness in disentangling disease complexity, one unresolved issue in the application of LCA is that there is not one commonly accepted statistical indicator for deciding on the number of subgroups in a study population. The limitation of this study 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 best-available approximation of the 'true classification'. In the absence of clear statistical requirements for identifying clinically important groups of small size, validation of the phenotypes with late asthma outcomes provides the only clues about their clinical relevance. However, we acknowledge that in our study information on asthma and lung function measures at age 23-24 years was available for ~45% of participants used to derive wheeze phenotypes.

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.[28]

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)[22] and lung function[29] may allow better distinction of the underlying pathophysiological mechanisms.

295 The key advantage of our study is the large sample size with complete data on wheezing
296 collected frequently and prospectively. Another advantage is that participants were followed
297 from birth to late adolescence, covering a longer period compared to most prior studies.[1, 13, 18,
298 19, 30]

299 Finally, it is worth noting that subtypes discovered using data-driven methods are not observed,
300 but are latent by nature, and ideally should not be referred to as "phenotypes" (i.e. observable
301 characteristics). However, as the term "phenotype" has been used in this context for over a
302 decade, we have maintained this nomenclature.

303 *Interpretation*

304 A number of previous studies (including our own) embarked on identifying wheeze phenotypes
305 from birth to mid-school age (summarized in Table E4). However, the inconsistency of findings
306 has led to a debate on the validity and clinical value of phenotyping studies,[10, 31, 32] hampering
307 the discovery of pathophysiological endotypes and translation into clinically actionable insights.
308 The four phenotypes of persistent, never, transient early and late-onset wheeze have been long
309 postulated in descriptive,[2] and data-driven studies.[33] We found that when the sample size is
310 relatively small, a particular wheeze phenotype that exists in the population may be
311 undetectable. Therefore, relatively smaller sample size in some studies might have contributed
312 to the inability to detect intermediate wheeze phenotypes with a relatively low prevalence.
313 Using more time points allowed the identification of less common phenotypes (<5% frequency)
314 by increasing possible response patterns. When the data collection was frequent (>11 time
315 points), a sample size of ~2500 was found to be sufficiently large to distinguish six phenotypes.
316 However, even a larger sample size of 3167 might be insufficient to detect uncommon

317 phenotypes (<5% frequency) under certain conditions (e.g. data collection points <11). Our
318 findings suggest that increasing data collection frequency may help compensate for a modest
319 sample size in phenotype identification. In line with this finding, Depner *et al*.[30] identified an
320 intermediate phenotype in the PASTURE cohort that existed during the first six years of life,
321 using a similar sample size but more data collection points than those used in the TCRS.[2]
322 However, the selection of follow up points needs a careful thought. Our analyses have shown
323 that although adding more time points to the latent model increased the number of identified
324 phenotypes with distinguishable interpretations, in some cases the same number of randomly
325 selected data collection points resulted in a different optimal solution. This suggests that the
326 timing and distribution of follow-ups is important, and that there might be critical data
327 collection points which are more informative than others. A variable selection method which
328 we applied to the data identified 6 time points which were not carrying additional useful
329 information (months 6, 30, 69, 103, 128 and 198).

330 The proportion of asthmatics was highest in the Persistent wheeze (98.5%), and subjects in this
331 phenotype had diminished pre and post-bronchodilator lung function (at the time of maximally
332 attained physiological lung function plateau[29]) compared to all other phenotypes. The
333 proportion of asthmatics differed between intermediate phenotypes (15.1% and 75.3% in two
334 transient early phenotypes, Pre-school remitting and Mid-childhood remitting respectively;
335 91.3% and 70.0% in two late-onset phenotypes, Late childhood and School-age onset). These
336 findings suggest that all phenotypes are distinct and that this may be a true classification.
337 However, we acknowledge that the observed associations may also be a proxy of severity.

338 The preschool-onset remitting phenotype mostly overlapped with no asthma (94%) but the pre
339 and post-bronchodilator lung function at age 24 was significantly lower in this class compared
340 to Never/infrequent wheeze. Although this may be seen as a contradiction, we would stress
341 that diminished lung function does not equate to asthma.[29] There is evidence that early
342 transient wheeze is associated with low lung function[34-37]; as lungs/airways grow the symptoms
343 regress but lung function impairments may persist. In TCRS, the lowest infant lung function test
344 values were associated with low lung function at 22 years[38]; so early wheeze that remits may be
345 a marker of low lung function in early life that persists to adulthood, but without the
346 development of airway inflammation or asthma.

347 In conclusion, our findings add to the understanding of childhood wheeze phenotypes by
348 extending the knowledge on potential causes of variability in classification of wheezing. Sample
349 size, frequency, and timing of data collection have a major influence on the number and type of
350 phenotypes identified by data-driven techniques. Our results, which include information on the
351 most informative follow-up points, are important to interpret (or reanalyze) existing studies
352 and to inform better design of future cohorts. However, we wish to note that these data
353 collection points should not be regarded as absolute; rather, they should be treated as relative
354 values with respect to our population, and considerations for investigators when designing
355 future studies.

356

357 **Acknowledgments:** Asthma UK Grant ref. 06/005 funded the collection of the wheezing and
358 asthma data in ALSPAC. We are extremely grateful to all the families who took part in this
359 study, the midwives for their help in recruiting them, and the whole ALSPAC team, which

360 includes interviewers, computer and laboratory technicians, clerical workers, research
361 scientists, volunteers, managers, receptionists, and nurses.

ACCEPTED MANUSCRIPT

362 **REFERENCES**

363 1. Henderson J, Granell R, Heron J, Sherriff A, Simpson A, Woodcock AA, et al. Associations of
364 wheezing phenotypes in the first six years of life with atopy, lung function and airway
365 responsiveness in mid childhood. Thorax 2008; 63:974-80.
366 2. Martinez FD, Wright AL, Taussig LM, Holberg CJ, Halonen M, Morgan WJ. Asthma and wheezing
367 in the first six years of life. New England Journal of Medicine 1995; 332:133-8.
368 3. Lowe LA, Simpson A, Woodcock A, Morris J, Murray CS, Custovic A, et al. Wheeze phenotypes
369 and lung function in preschool children. Am J Respir Crit Care Med 2005; 171:231-7.
370 4. Savenije OE, Granell R, Caudri D, Koppelman GH, Smit HA, Wijga A, et al. Comparison of
371 childhood wheezing phenotypes in 2 birth cohorts: ALSPAC and PIAMA. J Allergy Clin Immunol
372 2011; 127:1505-12. e14.
373 5. Arshad SH, Holloway JW, Karmaus W, Zhang H, Ewart S, Mansfield L, et al. Cohort Profile: The
374 Isle Of Wight Whole Population Birth Cohort (IOWBC). International journal of epidemiology
375 2018.
376 6. Pike K, Rose-Zerilli M, Osvald EC, Inskip H, Godfrey KM, Crozier S, et al. The relationship between
377 infant lung function and the risk of wheeze in the preschool years. Pediatric pulmonology 2011;
378 46:75-82.
379 7. Deliu M, Belgrave D, Sperrin M, Buchan I, Custovic A. Asthma phenotypes in childhood. Expert
380 Rev Clin Immunol 2017; 13:705-13.
381 8. Howard R, Rattray M, Prosperi M, Custovic A. Distinguishing Asthma Phenotypes Using Machine
382 Learning Approaches. Curr Allergy Asthma Rep 2015; 15:38.
383 9. Just J, Bourgoin-Heck M, Amat F. Clinical phenotypes in asthma during childhood. Clin Exp
384 Allergy 2017; 47:848-55.
385 10. Belgrave D, Simpson A, Custovic A. Challenges in interpreting wheeze phenotypes: the clinical
386 implications of statistical learning techniques. Am J Respir Crit Care Med 2014; 189:121-3.
387 11. Custovic A, Ainsworth J, Arshad H, Bishop C, Buchan I, Cullinan P, et al. The Study Team for Early
388 Life Asthma Research (STELAR) consortium 'Asthma e-lab': team science bringing data, methods
389 and investigators together. Thorax 2015; 70:799-801.
390 12. Cano-Garcinuño A, Mora-Gandarillas I. Wheezing phenotypes in young children: an historical
391 cohort study. Primary Care Respiratory Journal 2014; 23:60-6.
392 13. Kurukulaaratchy R, Fenn M, Waterhouse L, Matthews S, Holgate S, Arshad S. Characterization of
393 wheezing phenotypes in the first 10 years of life. Clinical & Experimental Allergy 2003; 33:573-8.
394 14. Lazic N, Roberts G, Custovic A, Belgrave D, Bishop C, Winn J, et al. Multiple atopy phenotypes
395 and their associations with asthma: similar findings from two birth cohorts. Allergy 2013;
396 68:764-70.
397 15. Spycher BD, Silverman M, Brooke AM, Minder CE, Kuehni CE. Distinguishing phenotypes of
398 childhood wheeze and cough using latent class analysis. European Respiratory Journal 2008;
399 31:974-81.
400 16. Lodge CJ, Zaloumis S, Lowe AJ, Gurrin LC, Matheson MC, Axelrad C, et al. Early-life risk factors
401 for childhood wheeze phenotypes in a high-risk birth cohort. The Journal of pediatrics 2014;
402 164:289-94. e2.
403 17. Granell R, Henderson AJ, Sterne JA. Associations of wheezing phenotypes with late asthma
404 outcomes in the Avon Longitudinal Study of Parents and Children: A population-based birth
405 cohort. Journal of Allergy and Clinical Immunology 2016; 138:1060-70. e11.
406 18. Savenije OE, Granell R, Caudri D, Koppelman GH, Smit HA, Wijga A, et al. Comparison of
407 childhood wheezing phenotypes in 2 birth cohorts: ALSPAC and PIAMA. Journal of Allergy and
408 Clinical Immunology 2011; 127:1505-12. e14.

19. Belgrave DC, Simpson A, Semic-Jusufagic A, Murray CS, Buchan I, Pickles A, et al. Joint modeling of parentally reported and physician-confirmed wheeze identifies children with persistent troublesome wheezing. Journal of Allergy and Clinical Immunology 2013; 132:575-83. e12.
20. Valk R, Caudri D, Savenije O, Koppelman GH, Smit HA, Wijga AH, et al. Childhood wheezing phenotypes and FeNO in atopic children at age 8. Clinical & Experimental Allergy 2012; 42:1329-36.
21. Boyd A, Golding J, Macleod J, Lawlor DA, Fraser A, Henderson 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; 42:111-27.
22. Belgrave DC, Granell R, Simpson A, Guiver J, Bishop C, Buchan I, et al. Developmental profiles of eczema, wheeze, and rhinitis: two population-based birth cohort studies. PLoS Med 2014; 11:e1001748.
23. Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, Coates A, et al. Standardisation of spirometry. European respiratory journal 2005; 26:319-38.
24. Beydon N, Davis SD, Lombardi E, Allen JL, Arets HG, Aurora P, et al. An official American Thoracic Society/European Respiratory Society statement: pulmonary function testing in preschool children. Am J Respir Crit Care Med 2007; 175:1304-45.
25. Quanjer P, Stanojevic S, Stocks J, Hall G, Prasad K, Cole T, et al. Changes in the FEV1/FVC ratio during childhood and adolescence: an intercontinental study. European Respiratory Journal 2010; 36:1391-9.
26. Linzer DA, Lewis JB. poLCA: An R package for polytomous variable latent class analysis. Journal of Statistical Software 2011; 42:1-29.
27. 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; 11:2080-110.
28. Lowe L, Murray CS, Martin L, Deas J, Cashin E, Poletti G, et al. Reported versus confirmed wheeze and lung function in early life. Arch Dis Child 2004; 89:540-3.
29. Belgrave DCM, Granell R, Turner SW, Curtin JA, Buchan IE, Le Souef PN, et al. Lung function trajectories from pre-school age to adulthood and their associations with early life factors: a retrospective analysis of three population-based birth cohort studies. Lancet Respir Med 2018.
30. Depner M, Fuchs O, Genuneit J, Karvonen AM, Hyvärinen A, Kaulek V, et al. Clinical and epidemiologic phenotypes of childhood asthma. American journal of respiratory and critical care medicine 2014; 189:129-38.
31. Belgrave D, Henderson J, Simpson A, Buchan I, Bishop C, Custovic AJJoA, et al. Disaggregating asthma: big investigation versus big data. 2017; 139:400-7.
32. Brand PL, Schultz AJERJ. To track or not to track: wheeze phenotypes in preschool children. 2018; 51:1800042.
33. Chen Q, Just AC, Miller RL, Perzanowski MS, Goldstein IF, Perera FP, et al. Using latent class growth analysis to identify childhood wheeze phenotypes in an urban birth cohort. Annals of Allergy, Asthma & Immunology 2012; 108:311-5. e1.
34. Grad R, Morgan WJJJoA, Immunology C. Long-term outcomes of early-onset wheeze and asthma. 2012; 130:299-307.
35. Morgan WJ, Stern DA, Sherrill DL, Guerra S, Holberg CJ, Guilbert TW, et al. Outcome of asthma and wheezing in the first 6 years of life: follow-up through adolescence. 2005; 172:1253-8.
36. Sears MRJJoA, Immunology C. Predicting asthma outcomes. 2015; 136:829-36.
37. Belgrave DC, Buchan I, Bishop C, Lowe L, Simpson A, Custovic AJAjor, et al. Trajectories of lung function during childhood. 2014; 189:1101-9.
38. Stern DA, Morgan WJ, Wright AL, Guerra S, Martinez FDJTL. Poor airway function in early infancy and lung function by age 22 years: a non-selective longitudinal cohort study. 2007; 370:758-64.

[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

ACCEPTED MANUSCRIPT

LEGEND FOR FIGURES

Figure 1. The optimal number, shape and prevalence of wheeze phenotypes identified by using latent class analysis:

a) 11 latent models based on randomly selected sub-samples of six different sample sizes (N=500, 1000, 1500, 2000, 2500 and 3167) while maintaining a constant number of follow-up points (14 time points)

b) 10 latent models based on randomly selected time points (6, 8, 11 and 14 TPs) while maintaining a constant sample size (N=3167)

c) 12 latent models based on randomly selected sub-samples of four different sample sizes (N=500, 1500, 2500 and 3167) and different number of time points (6, 8 and 11 TPs)

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.

ACCEPTED MANUSCRIPT

Table 1: Clustering summary of the LCA model fitted to the data sub-set (8 time points identified via a genetic algorithm search) and its comparison the model fitted to the full dataset (14 data collection points): based on 3167 participants with complete information on current wheeze recorded at 14 time points

Model Characteristics		**Variable Selection (stochastic search)**	
No. of classes	6	Selected time points (months)	Univariate entropy
BIC	15508	18	0.502
		42	0.581
Entropy	0.87	57	0.590
		81	0.578
Rand Index	0.82	91	0.588
Adjusted Rand Index	0.64	140	0.549
Jaccard Index	0.70	157	0.576
		166	0.582

ACCEPTED MANUSCRIPT

Table 2: Proportion of asthmatics at age 23-24 in each phenotype

Wheezing Phenotypes 0-16½ years		Self-reported asthma ever		Current asthma at age 23		Asthma medication use at age 23	
		No. of asthmatics/total	Percent*	No. of asthmatics/total	Percent*	No. of med. users/total	Percent*
Never-infrequent		105/1111	9.4	50/985	5.1	33/985	3.3
Pre-school remitting	Transient early	54/355	15.1	19/295	6.4	9/295	3.2
Mid-childhood remitting	Transient early	72/95	75.3	30/49	60.4	14/49	29.5
School-age onset	Late-onset	56/61	91.3	38/43	88.4	25/43	58.3
Late-childhood onset	Late-onset	58/82	70.0	38/55	68.1	25/55	45.3
Persistent wheeze		81/82	98.5	65/65	99.7	53/65	82.1

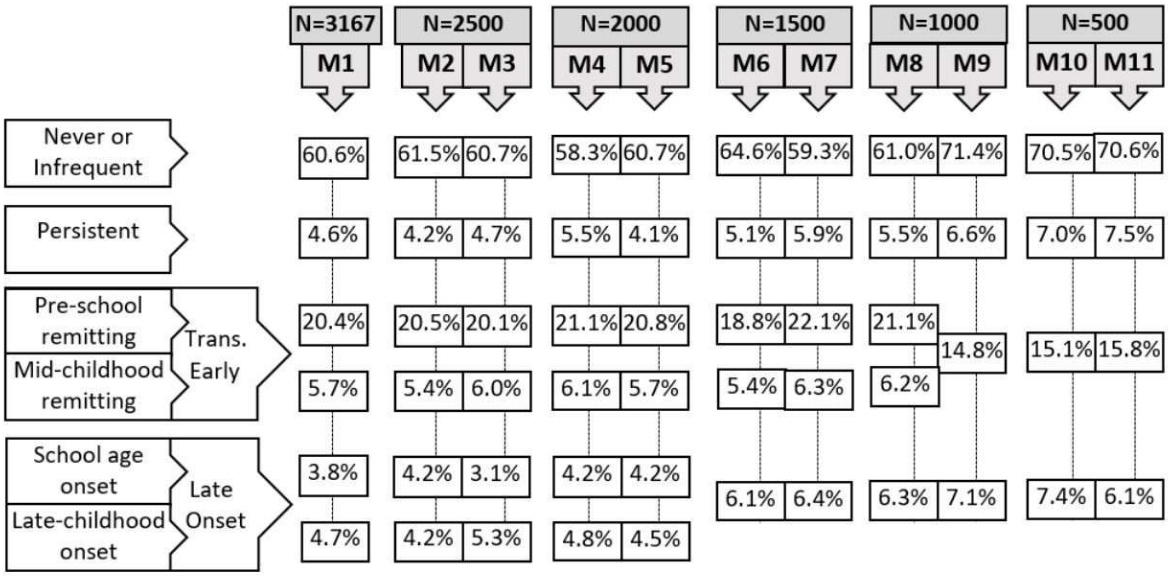
*The percentage is estimated from weighted cross tabulations.

ACCEPTED MANUSCRIPT

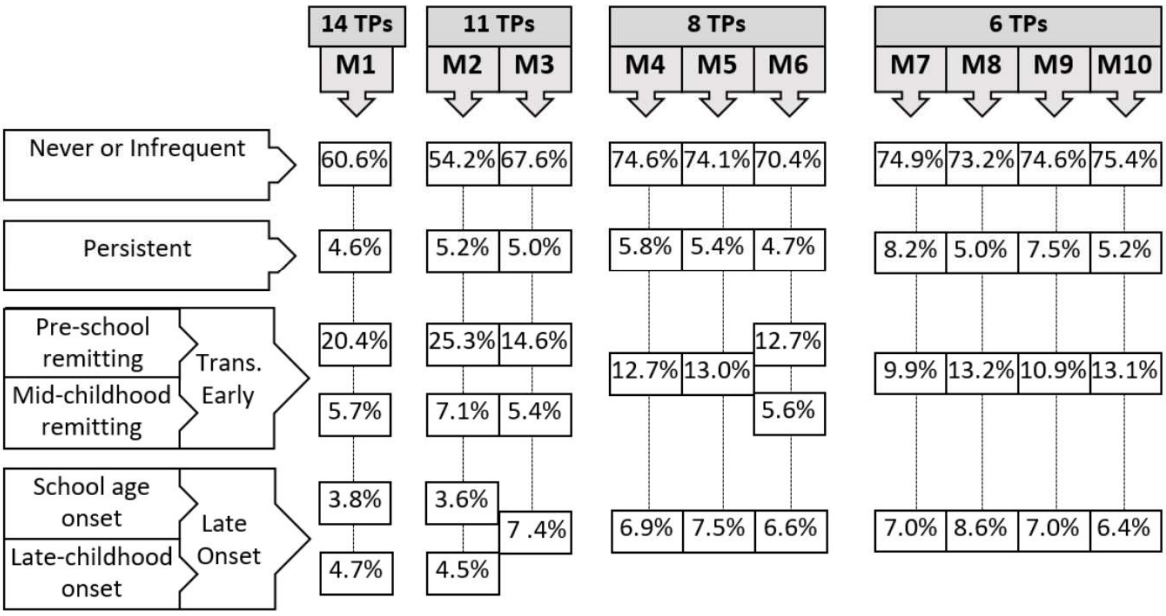
ACCEPTED MANUSCRIPT

Table 3: Lung function at age 24 years by wheeze phenotype (restricted to 1343 participants with FEV_1 % predicted data and 1351 with FEV_1/FVC data)

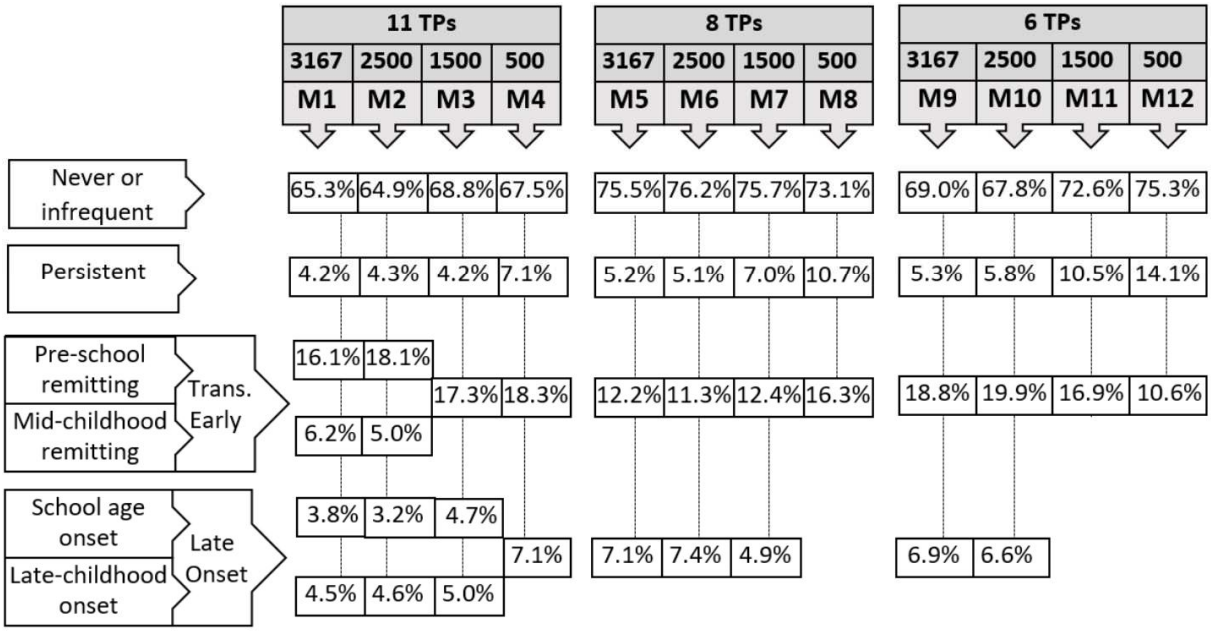
Wheezing Phenotypes 0-16½ years		Baseline lung function at 24y				Post-bronchodilator lung function at 24y			
		FEV_1 % predicted		FEV_1/FVC		FEV_1 % predicted		FEV_1/FVC	
		No.	Mean (SD)	No.	Mean (SD)	No	Mean (SD)	No.	Mean (SD)
Never-infrequent		1004	95.0 (11.7)	1009	0.84 (0.06)	830	97.9 (11.7)	834	0.86 (0.06)
Pre-school remitting	Transient early	329	93.4 (11.4)	330	0.82 (0.07)	274	96.8 (10.9)	275	0.85 (0.06)
Mid-childhood remitting	Transient early	89	93.5 (11.4)	91	0.82 (0.06)	71	97.5 (11.8)	73	0.84 (0.05)
School-age onset	Late-onset	61	95.4 (11.2)	61	0.81 (0.08)	47	100.8 (10.8)	47	0.86 (0.06)
Late-childhood onset	Late-onset	79	94.0 (12.1)	80	0.82 (0.07)	62	98.7 (10.8)	63	0.85 (0.05)
Persistent wheeze		80	91.6 (12.4)	80	0.79 (0.09)	59	96.5 (11.1)	59	0.83 (0.07)



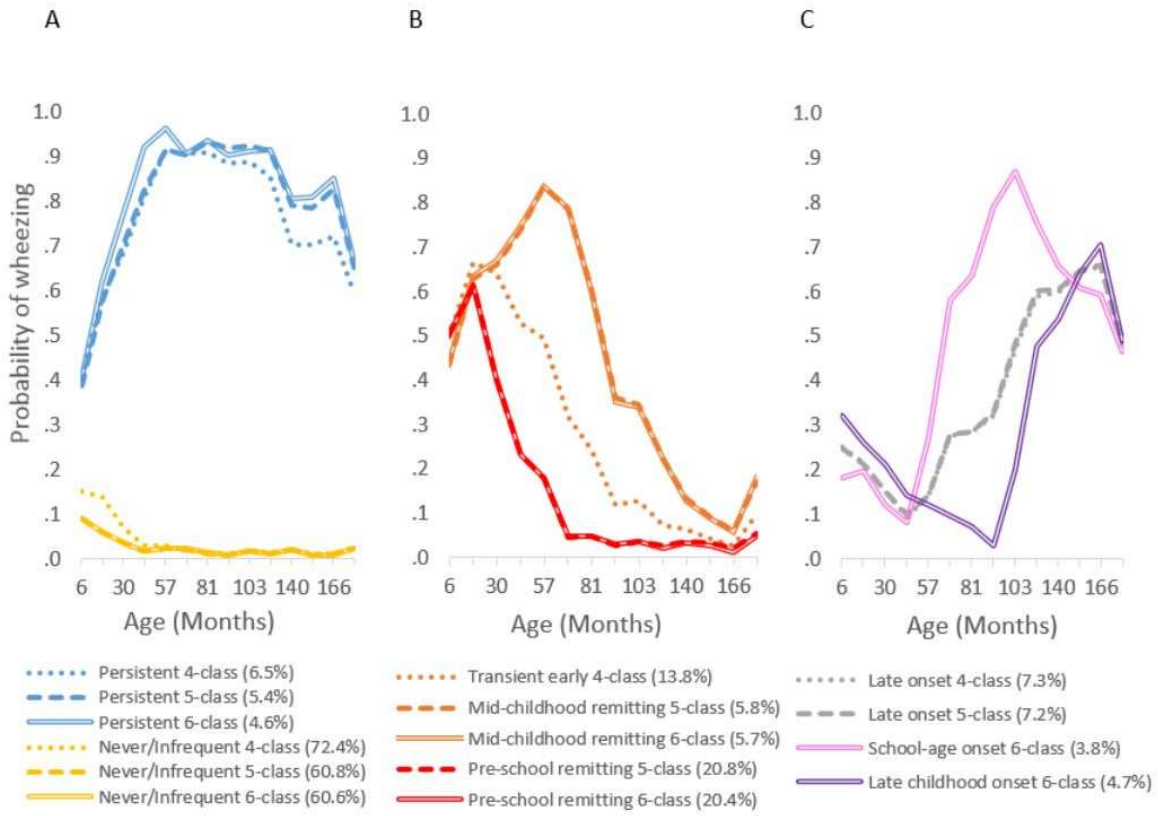
ACCEPTED MAN



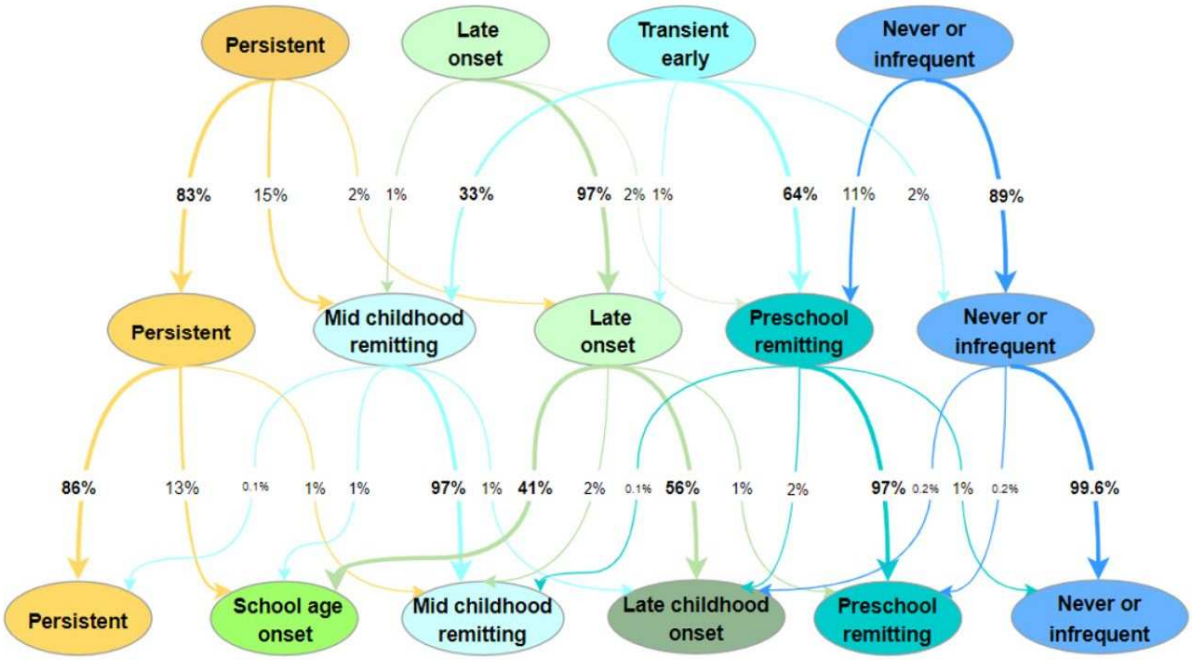
ACCEPTED MANUSCRIPT



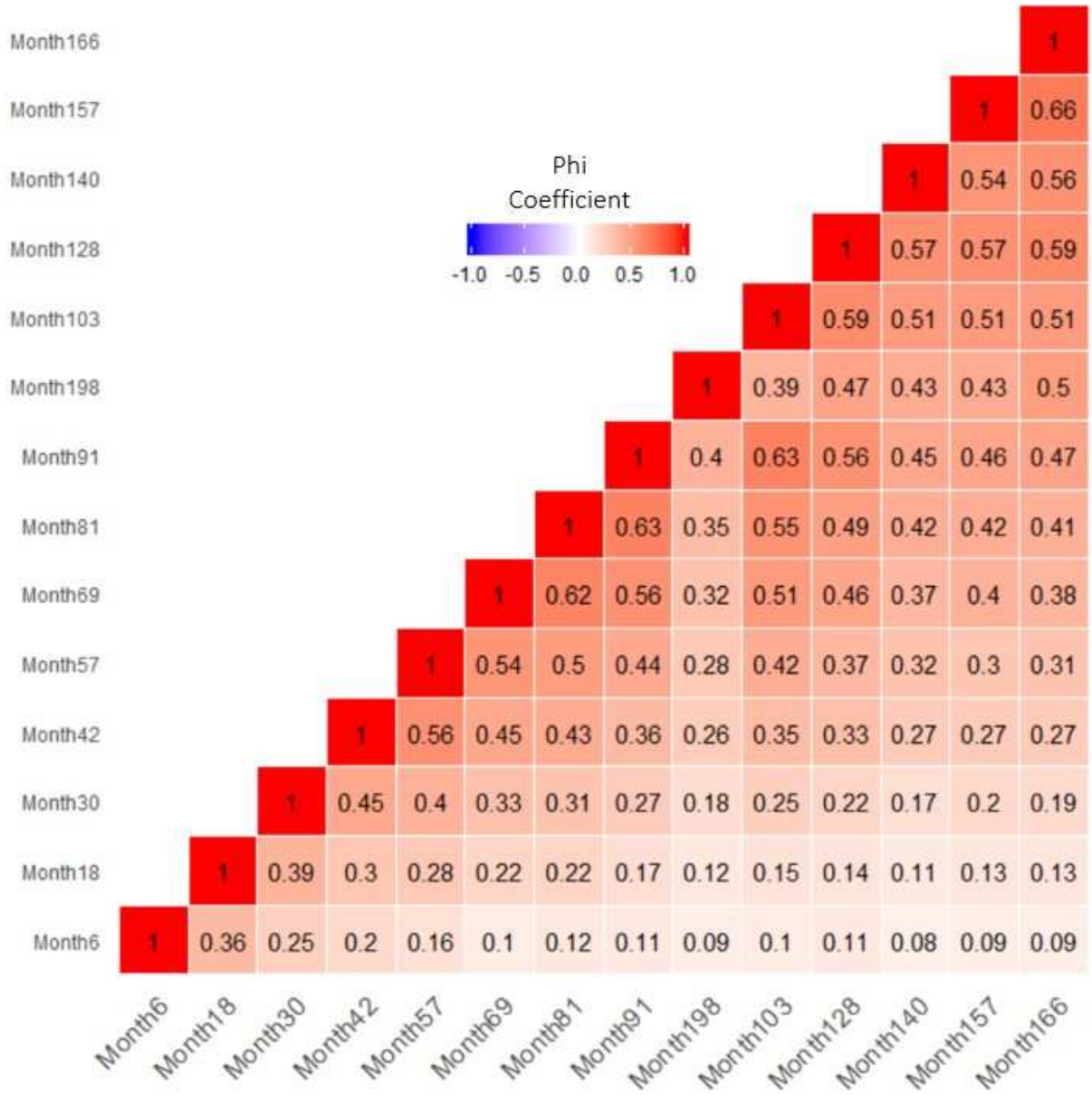
ACCEPTED MANUSCRIPT

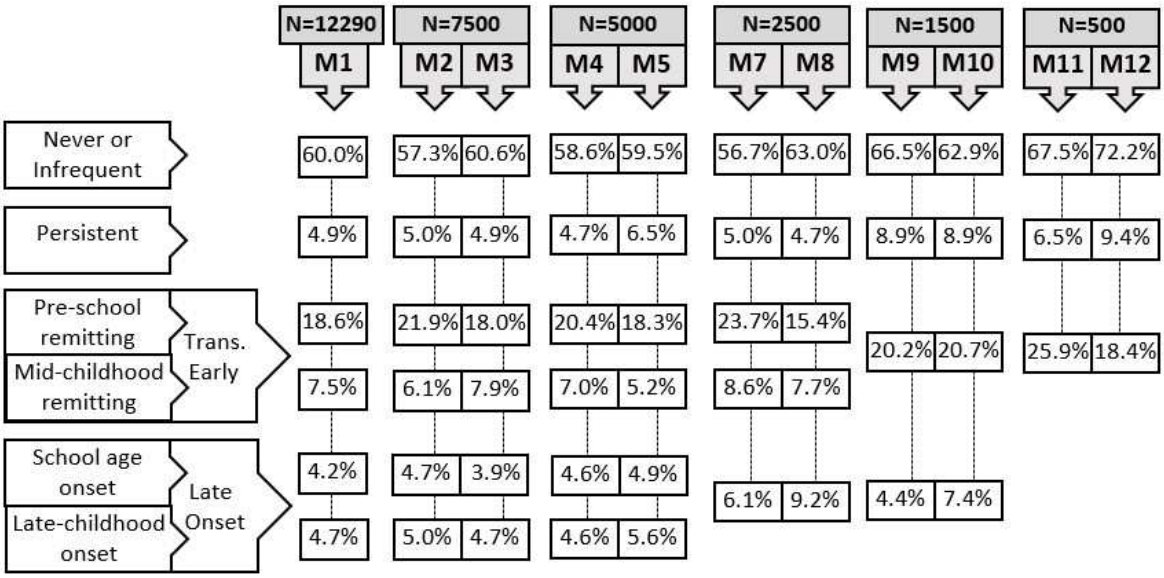

A
B
C
Probability of wheezing
1.0
.9
.8
.7
.6
.5
.4
.3
.2
.1
.0
6 30 57 81 103 140 166
Age (Months)
Persistent 4-class (6.5%)
Persistent 5-class (5.4%)
Persistent 6-class (4.6%)
Never/Infrequent 4-class (72.4%)
Never/Infrequent 5-class (60.8%)
Never/Infrequent 6-class (60.6%)
Transient early 4-class (13.8%)
Mid-childhood remitting 5-class (5.8%)
Mid-childhood remitting 6-class (5.7%)
Pre-school remitting 5-class (20.8%)
Pre-school remitting 6-class (20.4%)
Late onset 4-class (7.3%)
Late onset 5-class (7.2%)
School-age onset 6-class (3.8%)
Late childhood onset 6-class (4.7%)

ACCEPTED

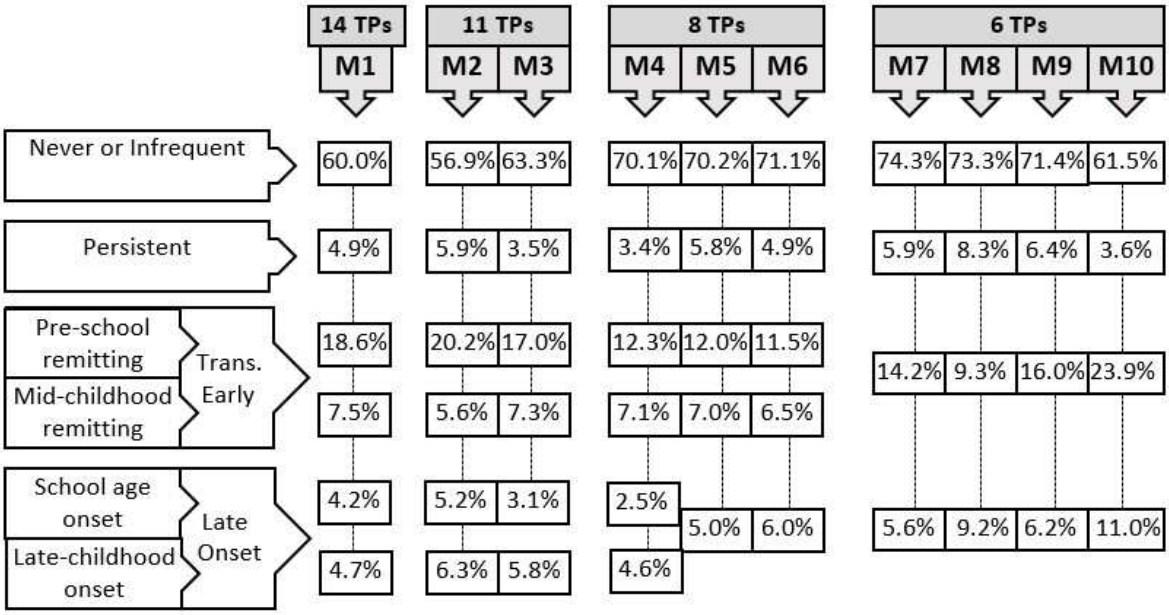


ACCEPTED MANUSCRIPT

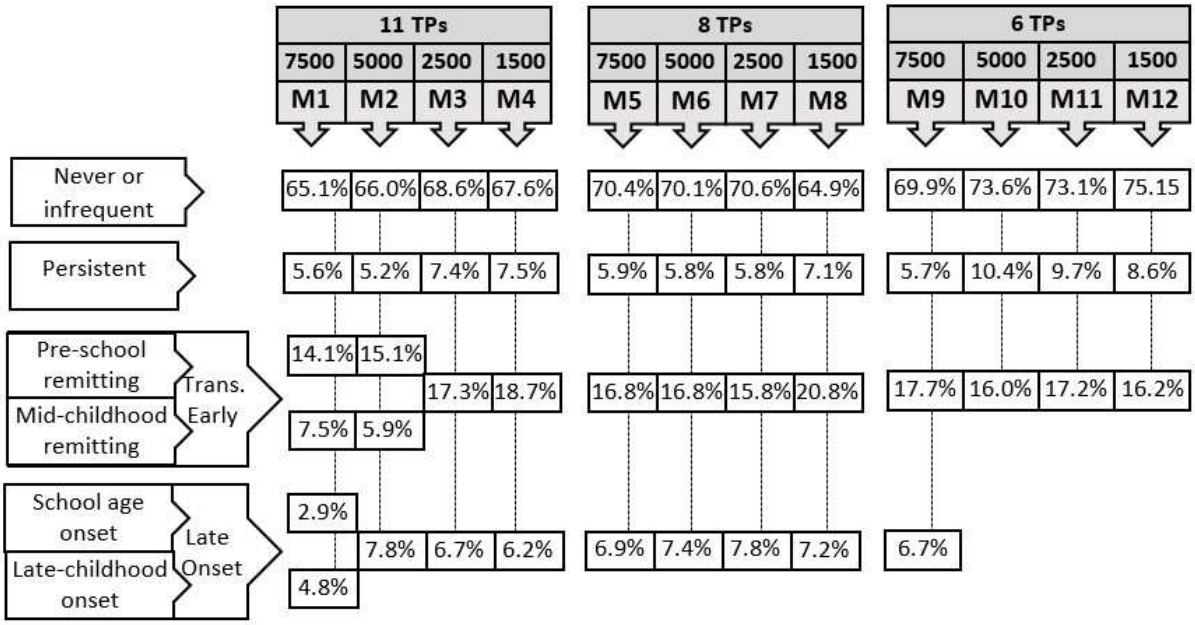

Phi
Coefficient
-1.0 -0.5 0.0 0.5 1.0
Month166 1
Month157 1 0.66
Month140 1 0.54 0.56
Month128 1 0.57 0.57 0.59
Month103 1 0.59 0.51 0.51 0.51
Month198 1 0.39 0.47 0.43 0.43 0.5
Month91 1 0.4 0.63 0.56 0.45 0.46 0.47
Month81 1 0.63 0.35 0.55 0.49 0.42 0.42 0.41
Month69 1 0.62 0.56 0.32 0.51 0.46 0.37 0.4 0.38
Month57 1 0.54 0.5 0.44 0.28 0.42 0.37 0.32 0.3 0.31
Month42 1 0.56 0.45 0.43 0.36 0.26 0.35 0.33 0.27 0.27 0.27
Month30 1 0.45 0.4 0.33 0.31 0.27 0.18 0.25 0.22 0.17 0.2 0.19
Month18 1 0.39 0.3 0.28 0.22 0.22 0.17 0.12 0.15 0.14 0.11 0.13 0.13
Month6 1 0.36 0.25 0.2 0.16 0.1 0.12 0.11 0.09 0.1 0.11 0.08 0.09 0.09
Month6 Month18 Month30 Month42 Month57 Month69 Month81 Month91 Month198 Month103 Month128 Month140 Month157 Month166



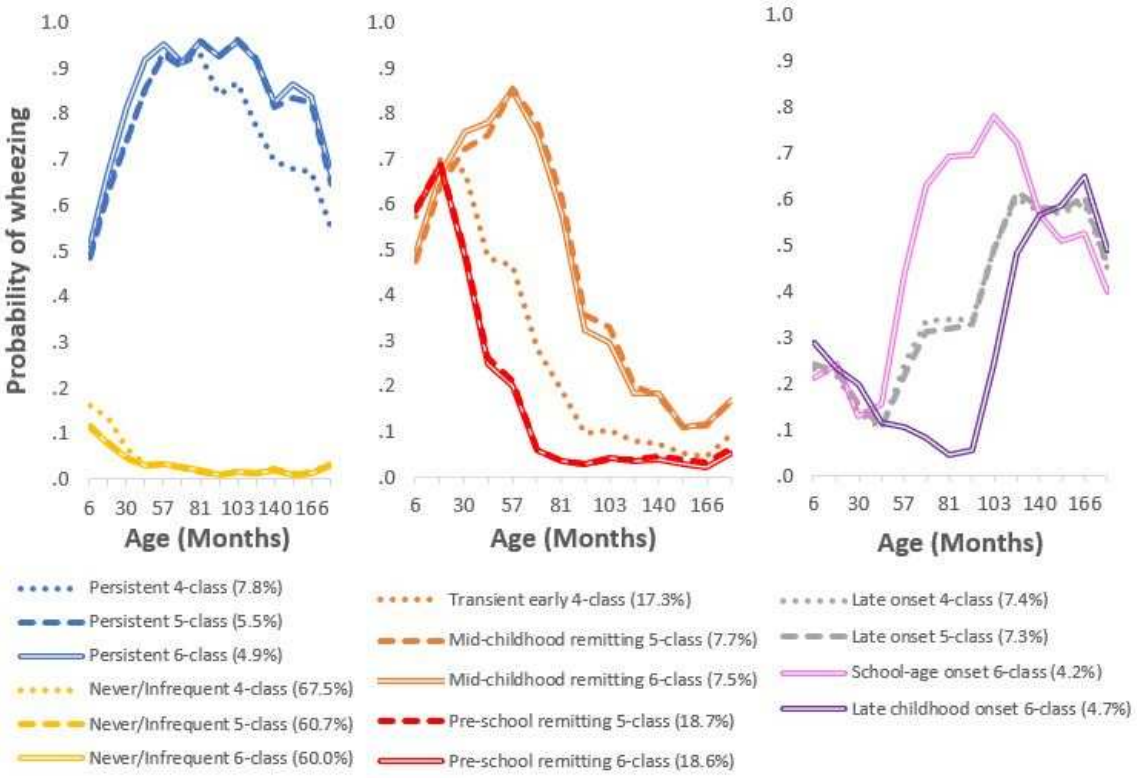
ACCEPTED MANUSCRIPT



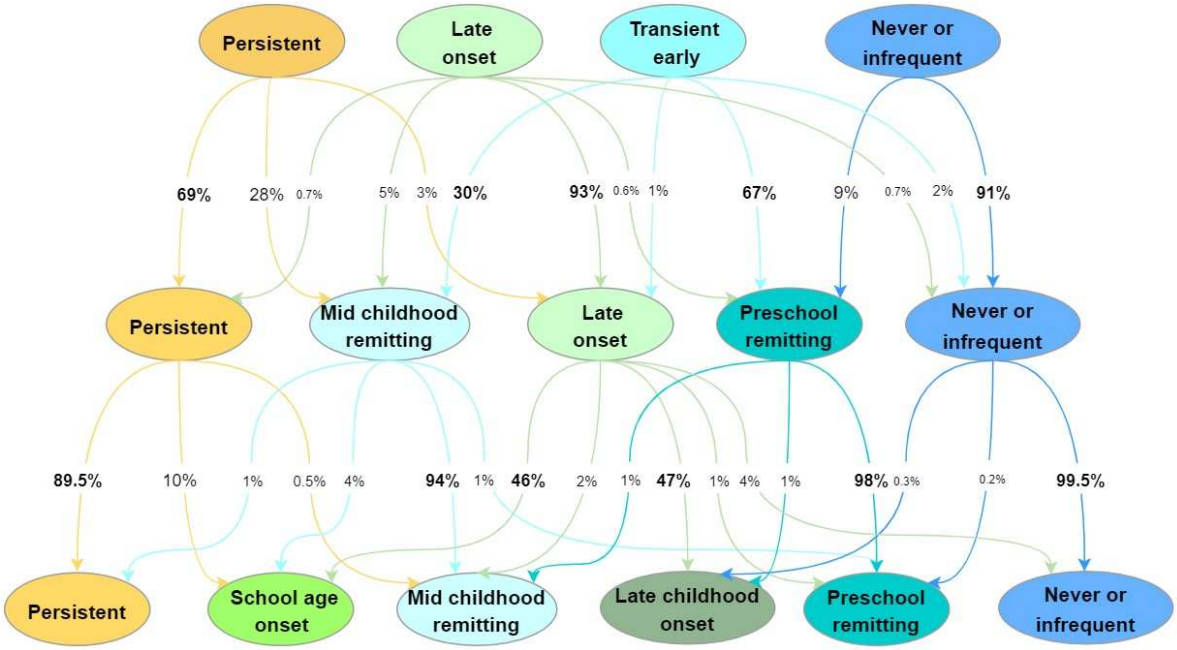
ACCEPTED MANUSCRIPT



ACCEPTED MANUSCRIPT



ACCEPTED MANUSCRIPT



ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Causes of variability in latent phenotypes of childhood wheeze

Ceyda Oksel PhD[1§], Raquel Granell PhD[2§], Osama Mahmoud PhD[2], Adnan Custovic MD PhD[1§], A. John Henderson MD[2§]

on behalf of STELAR[3] and Breathing Together[4] investigators

[§]Equal contribution

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

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

ONLINE DATA SUPPLEMENT

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

METHODS

Study design: Unselected birth cohort study.

Setting: ALSPAC is based on the former administrative County of Avon, United Kingdom covering a population of approximately 0.9 million.

Screening and recruitment: ALSPAC initially recruited 14,541 pregnant women resident in Avon, UK with expected dates of delivery between April 1, 1991 and December 31, 1992. This initial number of pregnancies, known as core sample, included the mothers enrolled in the ALSPAC study and had either returned at least one questionnaire or attended a 'Children in Focus' research clinic by 19th July 1999. These initial pregnancies had a total of 14,676 fetuses, resulting in 14,062 live births and 13,988 children who were alive at age 1 year. When the oldest children were approximately seven years of age, an attempt was made to bolster the initial sample with eligible cases who had failed to join the study originally. As a result, there are extra data available when considering variables collected from the age of seven years onwards. The number of new pregnancies, not in the core sample, known as phases II and III enrolments, is 706 (452 and 254 recruited during Phases II and III respectively), resulting in an additional 713 children being enrolled. The phases of enrolment are described in more detail in the cohort profile paper.[1,2] Therefore, the total sample size for analyses using any data collected after the age of seven years is therefore 15,247 pregnancies, resulting in 15,458 fetuses with 14,775 live births and 14,701 alive children at 1 year of age.

Spirometry: Performed according to American Thoracic Society/European Respiratory Society guidelines,[3,4] using a Vitalograph pneumotachograph system with animated incentive software (Spirotrac, Vitaograph, UK) in a dedicated research clinic by trained technicians. Calibration checks were performed with a standard 3L calibration syringe according to the manufacturer's instructions at the start of each half-day clinic session. Subjects were seated with a nose clip in place and were asked to inhale to total lung capacity (TLC), then instructed to perform a forced expiration, through a mouthpiece, to residual volume (RV). The test was repeated at intervals of 30 seconds until 3 technically acceptable traces were obtained from a maximum of eight

attempts. Forced expiratory volume in one second (FEV_1) and Forced vital capacity (FVC) were recorded and the data expressed as FEV_1 % predicted and FEV_1/FVC ratio.

Definition of outcomes

Current wheeze: 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?"

Current asthma: Self-reported current asthma at 23 years based on asthma ever at 22+ together with current wheezing and/or current treatment: "Have you had any wheezing in the past 12 months?" and/or "Have you taken asthma medication in the last 12 months?".

Asthma ever: Positive answer to the question "Have you ever had asthma?" at age 22+.

Study data were collected and managed using REDCap electronic data capture tools hosted at ALSPAC facilities.[5] Please note that the study website contains details of all the data that is available through a fully searchable data dictionary and variable search tool (http://www.bris.ac.uk/alspac/researchers/data-access/data-dictionary/).

Statistical analysis

Measures of fit in LCA

An optimal model is defined as the free model that best fits the data. To assess model fit, we used (1) the Bayesian information criterion (BIC), a function of the likelihood that rewards parsimony and (2) entropy, an assessment of model classification based on the posterior class membership probabilities. 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. Entropy is a measure of classification certainty that that ranges from 0 to 1, with values near 1 indicating a clear delineation of classes and values near 0 indicating low certainty in classification.

Selection of informative data collection points

A genetic algorithm was employed to search for the optimal set of clustering variables (e.g. time points) to distinguish wheeze subgroups, using the LCAvarsel R package.[6] During the search, multiple sets of clustering variables were considered at the same time; then, for each set, a latent class analysis model was estimated on the clustering variables and a regression/independence model was estimated on the non-clustering ones. Different sets were generated by various genetic operators and the fittest individuals were selected. The fitness function was defined as the BIC of the joint distribution of both clustering and non-clustering variables, where clustering variables were modeled via a latent class analysis model and non-clustering variables were modeled via multinomial logistic regression. Variable specific entropy contribution of each time point was used to assess how well individual time points identified the latent classes. These univariate entropies varying between 0 and 1 were directly comparable to each other, with large values indicating the clear separation of classes. The Rand Index (RI) and Adjusted Rand Index (ARI) was used as a similarity measure when comparing different clustering results. More specifically, ARI was used to measure the level of agreement between two partitions, the model fitted to the data subset and the full dataset. A larger RI and ARI means a higher agreement between two partitions.

ACCEPTED MANUSCRIPT

REFERENCES

1. Boyd A, Golding J, Macleod J, Lawlor DA, Fraser A, Henderson 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; 42:111-27.
2. Fraser A, Macdonald-Wallis C, Tilling K, Boyd A, Golding J, Davey Smith G, et al. Cohort profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. 2012; 42:97-110.
3. Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, Coates A, et al. Standardisation of spirometry. European respiratory journal 2005; 26:319-38.
4. Beydon N, Davis SD, Lombardi E, Allen JL, Arets HG, Aurora P, et al. An official American Thoracic Society/European Respiratory Society statement: pulmonary function testing in preschool children. Am J Respir Crit Care Med 2007; 175:1304-45.
5. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JGJJBI. A metadata-driven methodology and workflow process for providing translational research informatics support. 2009; 42:377-81.
6. 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; 11:2080-110.
7. Chen Q, Just AC, Miller RL, Perzanowski MS, Goldstein IF, Perera FP, et al. Using latent class growth analysis to identify childhood wheeze phenotypes in an urban birth cohort. Annals of Allergy, Asthma & Immunology 2012; 108:311-5. e1.
8. 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; 332:133-8.
9. Depner M, Fuchs O, Genuneit J, Karvonen AM, Hyvärinen A, Kaulek V, et al. Clinical and epidemiologic phenotypes of childhood asthma. American journal of respiratory and critical care medicine 2014; 189:129-38.
10. Belgrave DC, Simpson A, Semic-Jusufagic A, Murray CS, Buchan I, Pickles A, et al. Joint modeling of parentally reported and physician-confirmed wheeze identifies children with persistent troublesome wheezing. Journal of Allergy and Clinical Immunology 2013; 132:575-83. e12.
11. Savenije OE, Granell R, Caudri D, Koppelman GH, Smit HA, Wijga A, et al. Comparison of childhood wheezing phenotypes in 2 birth cohorts: ALSPAC and PIAMA. J Allergy Clin Immunol 2011; 127:1505-12. e14.
12. Henderson J, Granell R, Heron J, Sherriff A, Simpson A, Woodcock AA, 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; 63:974-80.

ACCEPTED MANUSCRIPT

LEGEND FOR FIGURES

Figure E1. Heatmap showing the phi coefficient of pairwise comparison between data collection points

Figure E2. The optimal number, shape and prevalence of wheeze phenotypes with different (12,290, 7500, 5000, 2500, 1500, 500) sample sizes based on children with at least 2 observations of wheeze (the optimal was chosen based on the lowest BIC)

Figure E3. The optimal number, shape and prevalence of wheeze phenotypes with different number of data collection points (6, 8, 11 and 14) based on children with at least 2 observations of wheeze.

Figure E4. The optimal number, shape and prevalence of wheeze phenotypes with combined effects of sample size and data collection frequency based on children with at least 2 observations of wheeze.

Figure E5. Estimated prevalence of wheeze for each wheezing phenotype in four-, five- and six-latent class solutions identified by LCA based on children 12,290 children with at least 2 observations of wheeze

Figure E6. 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 (12,290 children with at least 2 observations of wheeze). 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.

ACCEPTED MANUSCRIPT

Table E1: Differences of lung function (FEV_1 % predicted) at 24 years between wheezing phenotypes: ANOVA and Tukey HSD test of pairwise comparisons.

ANOVA

Variation source	df		Sum square		Mean square		F value		Pr(>F)	
	Pre-BD FEV1 % pred.	Post-BD FEV1 % pred.	Pre-BD FEV1 % pred.	Post-BD FEV1 % pred.	Pre-BD FEV1 % pred.	Post-BD FEV1 % pred.	Pre-BD FEV1 % pred.	Post-BD FEV1 % pred.	Pre-BD FEV1 % pred.	Post-BD FEV1 % pred.
Phenotypes	5	5	2243	1463	448.6	292.6	3.313	2.252	0.005561	0.04709
Residual	1651	1351	223565	175521	135.4	129.9				

Tukey's HSD (Honestly Significant Difference) Test

Pairwise comparison of wheeze phenotypes	Mean Differences		Significant (p adjusted<0.05)?		95 % CI of differences (lower, upper)	
	Pre-BD FEV_1 % pred.	Post-BD FEV_1% pred	Pre-BD FEV_1% pred.	Post-BD FEV_1 % pred.	Pre-BD FEV_1 % pred.	Post-BD FEV_1 % pred.
Mid-childhood remitting AND Late-childhood onset	-0.66	-1.61	No	No	(-5.82, 4.49)	(-7.27, 4.05)
Never/infrequent AND Late-childhood onset	0.69	-1.34	No	No	(-3.11, 4.50)	(-5.52, 2.84)
Pre-school remitting AND Late-childhood onset	-1.59	-3.11	No	No	(-5.78, 2.60)	(-7.69, 1.47)
Persistent AND Late-childhood onset	-3.26	-3.41	No	No	(-8.44, 1.93)	(-9.26, 2.44)
School-age onset AND Late-childhood onset	1.51	1.70	No	No	(-4.11, 7.12)	(-4.45, 7.86)
Never/infrequent AND Mid-childhood remitting	1.36	0.27	No	No	(-2.40, 5.12)	(-3.85, 4.39)
Pre-school remitting AND mid-childhood remitting	-0.93	-1.50	No	No	(-5.08, 3.22)	(-6.03, 3.03)
Persistent AND Mid-childhood remitting	-2.59	-1.80	No	No	(-7.75, 2.56)	(-7.61, 4.00)
School-age onset AND Mid-childhood remitting	2.17	3.31	No	No	(-3.42, 7.76)	(-2.80, 9.43)
Pre-school remitting AND Never/infrequent	-2.29	-1.77	**Yes**	No	(-4.55, -0.02)	(-4.20, 0.66)
Persistent AND Never/infrequent	-3.95	-2.07	**Yes**	No	(-7.76, -0.15)	(-6.45, 2.30)
School-age onset AND Never/infrequent	0.81	3.04	No	No	(-3.56, 5.18)	(-1.73, 7.82)
Persistent AND Pre-school remitting	-1.66	-0.30	No	No	(-5.86, 2.52)	(-5.06, 4.46)
School-age onset AND Pre-school remitting	3.10	4.81	No	No	(-1.61, 7.81)	(-0.32, 9.94)
School-age onset AND Persistent	4.77	5.12	No	No	(-0.85, 10.38)	(-1.17, 11.4)

ACCEPTED MANUSCRIPT

Table E2: Differences of lung function (FEV_1/FVC ratio) at 24 years between wheezing phenotypes: ANOVA and Tukey HSD test of pairwise comparisons.

ANOVA

Variation source	df		Sum square		Mean square		F value		Pr(>F)	
	Pre-BD FEV_1/FVC ratio	Post-BD FEV_1/FVC ratio	Pre-BD FEV_1/FVC ratio	Post-BD FEV_1/FVC ratio	Pre-BD FEV_1/FVC ratio	Post-BD FEV_1/FVC ratio	Pre-BD FEV_1/FVC ratio	Post-BD FEV_1/FVC ratio	Pre-BD FEV_1/FVC ratio	Post-BD FEV_1/FVC ratio
Phenotypes	5	5	0.3269	0.1527	0.06539	0.03054	14.51	8.33	6.2E-14	9.0E-08
Residual	1600	1358	7.482	4.977	0.00451	0.00366				

Tukey's HSD (Honestly Significant Difference) Test

Pairwise comparison of wheeze phenotypes	Mean Differences		Significant (p adjusted<0.05)?		95 % CI of differences (lower, upper)	
	Pre-BD FEV_1/FVC ratio	Post-BD FEV_1/FVC ratio	Pre-BD FEV_1/FVC ratio	Post-BD FEV_1/FVC ratio	Pre-BD FEV_1/FVC ratio	Post-BD FEV_1/FVC ratio
Mid-childhood remitting AND Late-childhood	0.01	-0.01	No	No	(-0.02, 0.04)	(-0.04, 0.02)
Never/infrequent AND Late-childhood onset	0.02	0.01	No	No	(0.00, 0.04)	(-0.02, 0.03)
Pre-school remitting AND Late-childhood onset	0.00	-0.01	No	No	(-0.03, 0.02)	(-0.04, 0.01)
Persistent AND Late-childhood onset	-0.03	-0.03	No	No	(-0.06, 0.00)	(-0.06, 0.00)
School-age onset AND Late-childhood onset	0.00	0.00	No	No	(-0.04, 0.03)	(-0.03, 0.03)
Never/infrequent AND Mid-childhood remitting	0.01	0.02	No	No	(-0.01, 0.04)	(0.00, 0.04)
Pre-school remitting AND mid-childhood	-0.01	0.00	No	No	(-0.04, 0.01)	(-0.02, 0.02)
Persistent AND Mid-childhood remitting	-0.03	-0.02	**Yes**	No	(-0.06, 0.00)	(-0.05, 0.01)
School-age onset AND Mid-childhood remitting	-0.01	0.01	No	No	(-0.04, 0.02)	(-0.02, 0.04)
Pre-school remitting AND Never/infrequent	-0.03	-0.02	**Yes**	**Yes**	(-0.04, -0.01)	(-0.03, -0.01)
Persistent AND Never/infrequent	-0.05	-0.04	**Yes**	**Yes**	(-0.07, -0.03)	(-0.06, -0.01)
School-age onset AND Never/infrequent	-0.03	-0.01	**Yes**	No	(-0.05, 0.00)	(-0.03, 0.02)
Persistent AND Pre-school remitting	-0.02	-0.02	No	No	(-0.05, 0.00)	(-0.04, 0.01)
School-age onset AND Pre-school remitting	0.00	0.01	No	No	(-0.03, 0.03)	(-0.01, 0.04)
School-age onset AND Persistent	0.02	0.03	No	No	(-0.01, 0.05)	(0.00, 0.06)

ACCEPTED MANUSCRIPT

Table E3: The FEV_1 reversibility at age 24 years by phenotype (restricted to 1364 participants with FEV_1 reversibility data)

Wheezing Phenotypes 0-16½ years		FEV_1 reversibility at 24y		P-value*	(optional) % positive FEV_1 reversibility
		No.	Mean (SD)		
Never-infrequent		898	2.87 (6.2)	Baseline	80.0
Pre-school remitting	Transient early	224	4.02 (5.4)	0.104	81.3
Mid-childhood remitting	Transient early	69	4.03 (5.3)	0.627	84.1
School-age onset	Late-onset	49	4.94 (4.9)	0.171	89.8
Late-childhood onset	Late-onset	65	4.22 (5.4)	0.490	83.1
Persistent wheeze		59	6.29 (5.9)	0.0003	91.5

*** Tukey test.**

ACCEPTED MANUSCRIPT

Table E4: Wheeze phenotypes from birth up to 9 years of age identified based on temporal pattern

Cohort	Sample Size	Number of time points	Years covered	Number of phenotypes
CCCEH[7]	689	15	9	4
TUSCON[8]	826	2	6	4
PASTURE[9]	953	6	6	5
MAAS[10]	1184	8	8	5
PIAMA[11]	2810	8	8	5
ALSPAC[11]	5760	8	8	6
ALSPAC[12]	6265	7	7	6

ACCEPTED MANUSCRIPT