Comparison of phenotypes of childhood wheeze and cough in 2 independent cohorts

Ben D. Spycher, PhD,[a] Michael Silverman, MD,[b] Anina M. Pescatore, MSc,[a] Caroline S. Beardsmore, PhD,[b] and Claudia E. Kuehni, MD, MSc[a] *Bern, Switzerland, and Leicester, United Kingdom*

Background: Among children with wheeze and recurrent cough there is great variation in clinical presentation and time course of the disease. We previously distinguished 5 phenotypes of wheeze and cough in early childhood by applying latent class analysis to longitudinal data from a population-based cohort (original cohort).

Objective: To validate previously identified phenotypes of childhood cough and wheeze in an independent cohort.

Methods: We included 903 children reporting wheeze or recurrent cough from an independent population-based cohort (validation cohort). As in the original cohort, we used latent class analysis to identify phenotypes on the basis of symptoms of wheeze and cough at 2 time points (preschool and school age) and objective measurements of atopy, lung function, and airway responsiveness (school age). Prognostic outcomes (wheeze, bronchodilator use, cough apart from colds) 5 years later were compared across phenotypes.

Results: When using a 5-phenotype model, the analysis distinguished 3 phenotypes of wheeze and 2 of cough as in the original cohort. Two phenotypes were closely similar in both cohorts: *Atopic persistent wheeze* (persistent multiple trigger wheeze and chronic cough, atopy and reduced lung function, poor prognosis) and *transient viral wheeze* (early-onset transient wheeze with viral triggers, favorable prognosis). The other phenotypes differed more between cohorts. These differences might be explained by differences in age at measurements.

Conclusions: Applying the same method to 2 different cohorts, we consistently identified 2 phenotypes of wheeze (atopic persistent wheeze, transient viral wheeze), suggesting that these represent distinct disease processes. Differences found in other phenotypes suggest that the age when features are assessed is critical and should be considered carefully when defining phenotypes. (J Allergy Clin Immunol 2013;■■■:■■■-■■■.)

Key words: *Wheeze, cough, asthma, children, phenotypes, latent class analysis, cluster analysis, cohort study, allergy, bronchial responsiveness*

Wheezing in early childhood varies greatly in clinical presentation and time course.[1,2] A single disease label, "asthma," has long been applied to all children with wheezing illness. But there is increasing evidence that the heterogeneity of clinical presentations and long-term outcomes is caused by the coexistence of different disease entities.[2-5] Diverse underlying pathologies may also explain the heterogeneity of clinical presentation in children with recurrent cough.[6-8] Some might suffer from a variant of asthma, but this remains controversial.[6,9,10]

Early attempts to define more homogenous subgroups, or "phenotypes," of childhood wheezing were based on expert opinion, such as the distinction into virus-induced wheeze and multiple trigger wheeze,[3,11] or the distinction by time course into transient early, persistent, and late-onset wheeze.[2] Recently, we[12,13] and others[14] proposed more objective data-driven methods for defining phenotypes of wheezing in children.[15] These and similar methods have since been used in other studies.[16-20] Using latent class analysis (LCA), Henderson et al[14] analyzed data on the presence of wheeze at different time points throughout the first 7 years of life in the Avon Longitudinal Study of Parents and Children (ALSPAC).[14] They identified 5 phenotypes of wheezing that were differently associated with measures of lung function, atopy, and bronchial responsiveness in mid-childhood. These findings were partially reproduced in the Dutch Prevention and Incidence of Asthma and Mite Allergy cohort study.[16]

In a previous study, we took a different approach and defined phenotypes on the basis of multiple clinical features assessed simultaneously, rather than solely on the basis of the presence of wheeze. This more closely resembles the differentiation of clinical presentations by the physician. By using LCA, we analyzed longitudinal data on severity and triggers of wheeze and cough, and measures of lung function, atopy, and bronchial responsiveness from the Leicestershire Respiratory Cohort Study.[13] Our approach identified 3 phenotypes of wheeze (labeled "atopic persistent wheeze," "nonatopic persistent wheeze," and "transient viral wheeze") and 2 phenotypes of cough ("persistent cough" and "transient cough"), which differed in outcomes 5 and 10 years later.[13] However, because this method is exploratory, the identified phenotypes require validation in independent populations.

From [a]the Institute of Social and Preventive Medicine, University of Bern, and [b]the Division of Child Health, Department of Infection, Immunity & Inflammation, University of Leicester and Institute for Lung Health.

The work presented in this article was funded by Asthma UK (grant no. 07/048) and the Swiss National Science Foundation (grant no. PDFMP3-123162 and 32003B-144068). B. D. Spycher is the recipient of a European Respiratory Society/Marie Curie Joint Research Fellowship (MC 1614-2010). The research leading to these results has received funding from the European Respiratory Society and the European Community's Seventh Framework Programme FP7/2007-2013—Marie Curie Actions under grant agreement RESPIRE, PCOFUND-GA-2008-229571. Data collection has been funded by UK National Asthma Campaign, University Hospitals of Leicester NHS Trust (R&D), Leicestershire & Rutland Partnership Trust, Medisearch, Trent NHS Regional Health Authority, and the UK Department of Health.

Disclosure of potential conflict of interest: B. D. Spycher has received grants from Swiss National Science Foundation, and the European Respiratory Society/Marie Curie Actions Joint Research Fellowship. C. S. Beardsmore and M. Silverman have received a grant from Asthma UK. C. E. Keuhni has received a grant from Swiss National Science Foundation. A. M. Pescatore declares no relevant conflicts of interest.

Received for publication September 7, 2012; revised August 8, 2013; accepted for publication August 9, 2013.

Corresponding author: Ben D. Spycher, PhD, Institute of Social and Preventive Medicine, University of Bern, Finkenhubelweg 11, CH-3012 Bern, Switzerland. E-mail: bspycher@ispm.unibe.ch.

0091-6749/$36.00

© 2013 American Academy of Allergy, Asthma & Immunology

http://dx.doi.org/10.1016/j.jaci.2013.08.002

Abbreviations used
ALSPAC: Avon Longitudinal Study of Parents and Children
BIC: Bayesian Information Criterion
BHR: Bronchial hyper-responsiveness
BLRT: Bootstrap likelihood ratio test
LCA: Latent class analysis

The aim of the present study was to investigate the robustness of the respiratory phenotypes we identified, by repeating our previous analysis in an independent cohort. We compared phenotype characteristics, including prognostic outcomes 5 years later, with the phenotypes identified in the original cohort. We assumed that if the identified phenotypes are a good reflection of the underlying disease processes in the population, they should be reproducible and robust against minor differences in study design.

METHODS

For ease of comparison, we present our original analysis[13] and the new analysis in the independent validation cohort in parallel. In both cohorts, we identified phenotypes by applying LCA to data on multiple disease dimensions, including symptom pattern and severity assessed at preschool and school age, and physiological measurements taken at school age. We then compared prognostic outcomes between the different phenotypes and an asymptomatic control group 5 years after the school-age survey.

Study design and study populations

The cohorts we used for identifying phenotypes were 2 independent representative samples of children born in the counties of Leicestershire and Rutland, United Kingdom: The original cohort consisted of 1422 children born between 1985 and 1989; the validation cohort consisted of 6970 children born between 1993 and 1997. The 2 cohort studies are similar in many aspects, including sampling and data collection procedures (routine data, questionnaires, laboratory measurement protocols) as described in detail elsewhere.[21] Postal questionnaires with detailed questions on respiratory symptoms were sent to parents of children in the original cohort in 1990, 1992-1994 (subsample), 1998, 2003, and 2010 and to those of the validation cohort in 1998, 1999, 2001, 2003, 2006, and 2010. All studies were approved by the local Research Ethics Committee, and informed consent was obtained from all parents and children participating in laboratory measurements.

The main difference between the cohorts is that physiological measurements, including spirometry, bronchial challenge, and skin prick testing, were taken at ages 4 to 8 years in the original cohort and at ages 8 to 13 years in the validation cohort. In addition, the original cohort consisted of white children only, while the validation cohort included children of South Asian ethnicity.

As in the original sample, we included children who responded to a survey at preschool age and at school age, and who participated in laboratory measurements (Table I; see Fig E1 in this article's Online Repository at www.jacionline.org). For the purpose of identifying phenotypes, children whose parents had reported wheeze or recurrent cough in the preschool or school age survey were included in the LCA. Children without cough or wheeze were designated asymptomatic and used as a control group for comparison of prognosis.

Symptoms included

To identify phenotypes, we included the following symptoms: previous episodes of wheezing; number of episodes in the past 12 months; shortness of breath; triggers of episodes; night cough; and triggers of cough (see Table E1 in this article's Online Repository at www.jacionline.org).[13] In both cohorts, the wording of the questions on symptoms was similar for most questions, with 2 exceptions: (1) night symptoms (original cohort: "wheezing is worse at night"; replaced in validation cohort with "sleep is disturbed by wheezing") and (2) seasonality/seasonal triggers of symptoms (original cohort: "symptoms are worst in spring or summer"; replaced in validation cohort with "wheeze is triggered by dust, grass, animals or food or drinks").

Physiological measurements

We included measurements of prebronchodilator lung function, bronchial responsiveness, and atopy. We used sex- and height-standardized z scores[22] of $FEV_{0.5}$ in the original cohort and similarly standardized z scores[23] of FEV_1 in the validation cohort. We used $FEV_{0.5}$ instead of FEV_1 in the original cohort because the children were young at the time of measurement and therefore some may have exhaled all their vital capacity within less than 1 second.[24] The use of $FEV_{0.5}$ in young children is also recognized as a suitable outcome measure during challenge testing.[25] We measured bronchial responsiveness as provocative concentration of methacholine causing a 20% decrease in transcutaneous oxygen tension[26] in the original cohort, and causing a 20% decrease in FEV_1 (PC_{20}) in the validation cohort. Previous studies have shown that provocative concentration of methacholine causing a 20% decrease in transcutaneous oxygen tension is a safe and reliable outcome measure for assessing responsiveness to methacholine in young children[27] and changes are closely correlated with respiratory resistance in this age group.[28] These concentrations were log-transformed for analysis.[29] Response to skin prick testing was assessed for cat hair, dog danders, *Dermatophagoides pteronyssinus*, and mixed grass pollen. Children who reacted to 1 or more of these allergens were designated atopic (see the Online Repository at www.jacionline.org for details).

Outcomes 5 years later in preadolescence

We compared the following prognostic outcomes between phenotypes: any episodes of wheeze in the past 12 months, 4 or more episodes in the past 12 months, bronchodilator use, and cough apart from colds. These outcomes were assessed at a survey (referred to as preadolescence survey) taken about 5 years after the school-age survey used for phenotype identification (see Fig E1).

Statistical analysis

To identify phenotypes, we used LCA, a statistical method that allows identifying groups of subjects with similar characteristics within a heterogeneous population. The method assumes that all associations between the included variables (in this case, symptoms and physiological measurements) are entirely due to the existence of distinct subpopulations called latent classes. Within the latent classes, which we interpret as phenotypes, all variables are assumed to be independent.[30,31] For consistency with our previous study,[13] we fitted the models by using an adapted version of Multimix, a Fortran program.[32] The model allowed for both categorical (symptoms, atopy) and continuous (FEV_1 z scores and bronchial challenge)[32] variables. Conditional questions (ie, questions relevant only to children with wheeze, such as frequency and triggers of episodes) were treated appropriately,[12] and missing values were treated as missing at random.[33] We chose to use a model with 5 phenotypes to compare with our original study, as this was the number of phenotypes identified in that study. However, we also fitted models with up to 7 phenotypes and calculated the Bayesian information criterion (BIC) and bootstrap likelihood ratio tests (BLRTs),[31] as in the original study, to assess the optimal number of phenotypes in the validation sample. For each child, we computed posterior probabilities of belonging to the phenotypes identified by the fitted models. Children were allocated to the phenotypes for which they had the highest probability.[31] Finally, we examined associations between phenotype groups and later prognostic outcomes by using logistic regression models. For this, we created 5 data sets in which the children were assigned to the phenotypes by using random sampling from the posterior membership probabilities (multiple imputation). The regression models were fitted in each of these data sets and results combined by using Rubin's rules.[34] These analyses were done by using Stata 12.1 software (StataCorp, College Station, Tex).

TABLE I. Characteristics of the symptomatic study populations used to define phenotypes in the original and validation cohort (excluding control group)

	Original sample (N = 319)	Validation sample (N = 903)
Demographic characteristics		
Years of birth (range)	1985-1989	1993-1997
Females	160/319 (50.2%)	428/903 (47.4%)
Ethnicity		
White	319/319 (100.0%)	697/903 (77.2%)
South Asian	0/319 (0.0%)	206/903 (22.8%)
Preschool questionnaire		
Age (y), median (range)	3.3 (0.3-5.4)	2.6 (1.0-5.0)
Wheeze ever	145/319 (45.5%)	488/903 (54.0%)
Cough apart from colds	178/319 (55.8%)	432/892 (48.4%)
School-age questionnaire		
Age (y), median (range)	6.3 (4.1-8.8)	6.6 (5.0-8.9)
Wheeze ever	159/319 (49.8%)	442/903 (48.9%)
Cough apart from colds	126/313 (40.3%)	525/898 (58.5%)
Laboratory measurements		
Age (y), median (range)	6.3 (4.1-8.8)	12.4 (8.6-14.1)
Skin prick test positivity (≥1 positive)	53/198 (26.8)	387/903 (42.9)

Data are presented as number of children with characteristic/total number of children with available information (%) if not otherwise stated.

RESULTS

A total of 903 children from the validation cohort reported wheeze or recurrent cough in the preschool and/or school-age survey and were included in the analysis for identifying phenotypes (see Fig E1) compared with 319 children in the original cohort. Response rates in the validation cohort to the preschool and school-age surveys were similar to those of the original cohort, in the order of 80% and 60%, respectively (see Fig E1). Median age of the children at the preschool and school-age surveys was comparable between cohorts, but there was a difference of about 6 years in median age at laboratory measurements between the 2 cohorts (Table I). Five years after the school-age survey, when the children were in the preadolescent stage, information on prognostic outcomes was available on 283 and 800 symptomatic children in the original and validation cohort, respectively, and on 159 and 369 controls (asymptomatic at preschool and school-age surveys), respectively (see Fig E1).

Identification of phenotypes

In the validation sample, the BIC preferred a model with 5 phenotypes and the BLRT a model with 7 or more phenotypes, while in the original sample, the BIC preferred a model with 2 phenotypes and the BLRT a model with 5. Because the original study was exploratory, we presented the more liberal model with 5 phenotypes.[13] For ease of comparability with that study, we focus on results of the 5-phenotype model which, in the validation sample, was also supported statistically, albeit by a different criterion. However, we report the full results of both the 5- and 7-phenotype models of the validation sample in the Online Repository (see Tables E3 and E4 at www.jacionline.org). We did not fit models with more than 7 phenotypes because of increasing computation time and problems of convergence.

Fig 1 shows how the children were reassigned to phenotypes as the number of phenotypes in the model increased. In the original sample, most phenotypes showed stability; that is, children grouped together in one model tended to be grouped together again in other models. In the validation sample, there was more regrouping, particularly among children assigned to 2A, 2B, and 2D in the 5-phenotype model.

We present the phenotypes of the original sample in the same order (1A, …, 1E) and with the same labels used in the original study.[13] Although the ordering of phenotypes in the validation sample (2A, …, 2E) can be chosen arbitrarily, we gave the same alphabetic character to phenotypes in both samples that had similar distinguishing characteristics regarding dominant symptom pattern (cough, wheeze) and symptom persistence.

Comparison of phenotype characteristics

The main characteristics of phenotypes are summarized in Table II and presented in more detail in Table III (objective features) and Fig 2 (reported symptoms). Cough apart from colds was the predominant symptom of the 2 most prevalent phenotypes (A and B) in both samples (Fig 2). In the original sample, phenotypes 1A and 1B represented a persistent and transient cough phenotype, respectively. In the validation sample, this distinction, though also present, was less pronounced: in 2A, the probability of cough apart from colds increased from 0.39 to 1.00 between preschool and school age and decreased in 2B from 0.86 to 0.34. Furthermore, 2B showed a tendency for transient wheeze (preschool probability of current wheeze 0.36) and bronchial responsiveness that was not seen in 1B (Fig 2, Table III).

Wheeze was the predominant symptom of the 3 remaining phenotypes (C, D, and E) in both samples. Both 1C (atopic persistent wheeze) and 2C were characterized by persistent wheeze (preschool and school-age probability of current wheeze for 1C. 0.75 and 0.68; for 2C. 0.70 and 0.76) with multiple triggers, chronic cough, reduced FEV_1, bronchial hyperresponsiveness (BHR), and atopy (Fig 2, Table III). Current wheeze was also likely in 1D (nonatopic persistent wheeze) and 2D, but with opposing trends from preschool to school age (1D, 0.79 and 0.46; 2D, 0.36 and 0.90). The D phenotypes were less likely than the C phenotypes to have frequent wheeze and cough apart from colds (Fig 2). While 1D was characterized by a low probability of atopy and somewhat increased bronchial responsiveness compared with asymptomatic controls, 2D was characterized by both atopy and BHR (Table III). The last wheeze phenotype represented a transient phenotype in both samples: children were likely to report wheeze ever, particularly at preschool age (1E, 0.91; 2E, 0.84), but less likely to report current wheeze at the preschool survey (1E, 0.28; 2E, 0.31), and unlikely to report current wheeze at the school-age survey (1E, 0.05; 2E, 0.00). Detailed model results are presented in Tables E2 and E3 in this article's Online Repository at www.jacionline.org.

Associations with prognostic outcomes 5 years later (preadolescence)

In both samples, phenotype C showed the greatest risk for later wheeze. The odds ratio (95% CI) for current wheeze in preadolescence, comparing 1C and 2C with the control group, was 20 (9-45) and 19 (11-32), respectively. For the outcome frequent wheeze, the odds ratio was 30 (8-106) and 20 (9-48), respectively; for bronchodilator use, 31 (14-72) and 19

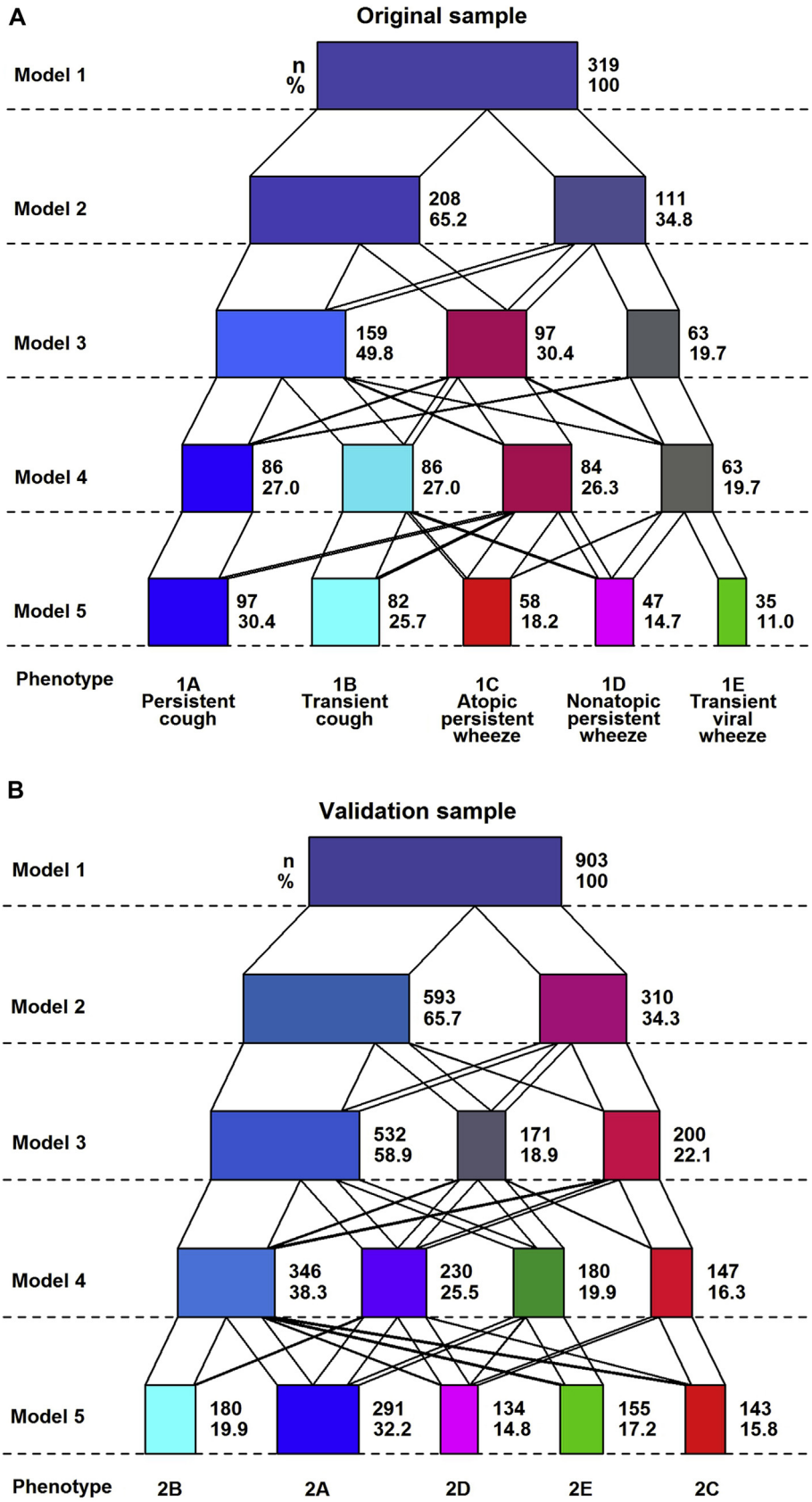

FIG 1. Stepwise identification of phenotypes in the 2 cohorts. Box width is proportional to the number of children allocated to a phenotype. Final phenotypes in the original sample are labeled as in the original publication.[13] The figure for the original sample is adapted with permission from Spycher et al.[13]

TABLE II. Main characteristics of phenotypes identified

Original cohort*	Validation cohort
1A (persistent cough): Cough apart from colds and night cough at preschool and school age; reduced FEV_1; increased bronchial responsiveness	**2A:** Cough apart from colds and night cough common at school age but less common at preschool age
1B (transient cough): Cough apart from colds and night cough at preschool age but rarely at school age	**2B:** Cough apart from colds and night cough common at preschool age but less common at school age; early wheeze with remission before school age in some; increased bronchial responsiveness
1C (atopic persistent wheeze): Wheeze, cough apart from colds and night cough at preschool and school age; episodes of wheeze frequent, also apart from colds, and often with SOB; atopy likely; reduced FEV_1; increased bronchial responsiveness	**2C:** Wheeze, cough apart from colds and night cough at preschool and school age; episodes of wheeze frequent, often with atopic triggers or exercise and often with SOB; atopy likely; reduced FEV_1; increased bronchial responsiveness
1D (nonatopic persistent wheeze): Wheeze at preschool and school age; episodes less common than in 1C also apart from colds, and often with SOB; atopy unlikely; increased bronchial responsiveness	**2D:** Wheeze common at school age but less common at preschool age; episodes less common than in 1C often with atopic triggers or exercise and often with SOB; atopy likely; reduced FEV_1; increased bronchial responsiveness
1E (transient viral wheeze): Early wheeze with remission before school age; episodes only with colds in most; increased bronchial responsiveness	**2E:** Early wheeze with remission before school age; episodes only with colds; reduced FEV_1

SOB, Shortness of breath.
*Labels in parentheses as assigned in previous publication.[13]

TABLE III. Characteristics of identified phenotypes compared with an asymptomatic control group according to objective features included in the LCA

	Phenotype					
The original cohort	**1A: Persistent cough**	**1B: Transient cough**	**1C: Atopic persistent wheeze**	**1D: Nonatopic persistent wheeze**	**1E: Transient viral wheeze**	**Control**
Children assigned (n)	97	82	58	47	35	169
Female	0.56	0.52	0.51	0.51	0.27	0.49
>3 y old at preschool survey	0.61	0.46	0.70	0.39	0.64	0.49
≥1 positive skin prick test result	0.19	0.16	0.70	0.09	0.22	0.11
$FEV_{0.5}$ (*z* score[22])	−1.59 ± 1.19	−1.18 ± 1.02	−1.80 ± 1.19	−1.47 ± 0.75	−1.09 ± 0.98	−1.33 ± 1.21
Bronchial challenge PC_{20}-$P_{tc,O2}$ (g/L)	2.42 (1.4-4.1)	2.75 (1.5-5.2)	1.26 (0.6-2.6)	2.32 (1.4-3.9)	2.48 (1.4-4.3)	3.82 (2.7-6.2)
	Phenotype					
The validation cohort	**2A**	**2B**	**2C**	**2D**	**2E**	**Control**
Children assigned (n)	291	180	143	134	155	435
Female	0.53	0.45	0.40	0.44	0.50	0.46
>3 y old at preschool survey	0.20	0.31	0.37	0.30	0.20	0.39
≥1 positive skin prick test result	0.34	0.35	0.64	0.67	0.28	0.31
FEV_1 (*z* score[23])	−0.12 ± 1.01	−0.24 ± 1.00	−0.53 ± 1.06	−0.38 ± 0.90	−0.31 ± 0.97	−0.09 ± 0.97
Bronchial challenge PC_{20} (g/L)	14.43 (6.0-34.6)	9.83 (4.1-23.8)	7.11 (2.4-20.7)	4.70 (2.1-10.4)	16.52 (6.4-42.5)	14.82 (5.4-33.5)

Data are presented as estimated probabilities, mean ± SD, or geometric mean (interquartile range) based on the latent class model for phenotypes and on sample values for controls. The table for the original cohort is adapted with permission from Spycher et al.[13]
PC_{20}-$P_{tc,O2}$, Provocative concentration of methacholine causing a 20% decrease in transcutaneous oxygen tension.

(11-32), respectively; and for cough apart from colds, 7 (3-14) and 9 (5-14), respectively (Table IV, Fig 3). Lower but still elevated risks were observed for phenotypes E in both samples, although CIs tended to include 1. For phenotypes D, risks of later outcomes were intermediate between C and E in both samples, but were more similar to C in the validation sample. In the original sample, the persistent cough phenotype A1 showed an elevated risk compared with controls for all outcomes with lower confidence limits above 1, while the transient cough phenotype B1 had risks comparable to controls. In the validation sample, the inverse situation was observed for the outcomes current wheeze and bronchodilator use, with elevated risks for phenotype B2 but not for A2 (Table IV, Fig 3).

DISCUSSION

Summary of findings

By applying a similar 5-phenotype model to an independent data set, we identified 2 phenotypes of cough and 3 phenotypes of wheeze as in our previously published study. Two of these phenotypes were closely similar in both samples: (1) *Atopic persistent wheeze* characterized by frequent multiple trigger wheeze and chronic cough at preschool and school age, atopy and reduced lung function at school age, and a poor prognosis; and (2) *Transient viral wheeze* characterized by early wheeze with viral triggers, remission by school age, and a favorable prognosis. The 2 cough phenotypes and the intermediate wheeze phenotype differed more between the cohorts.

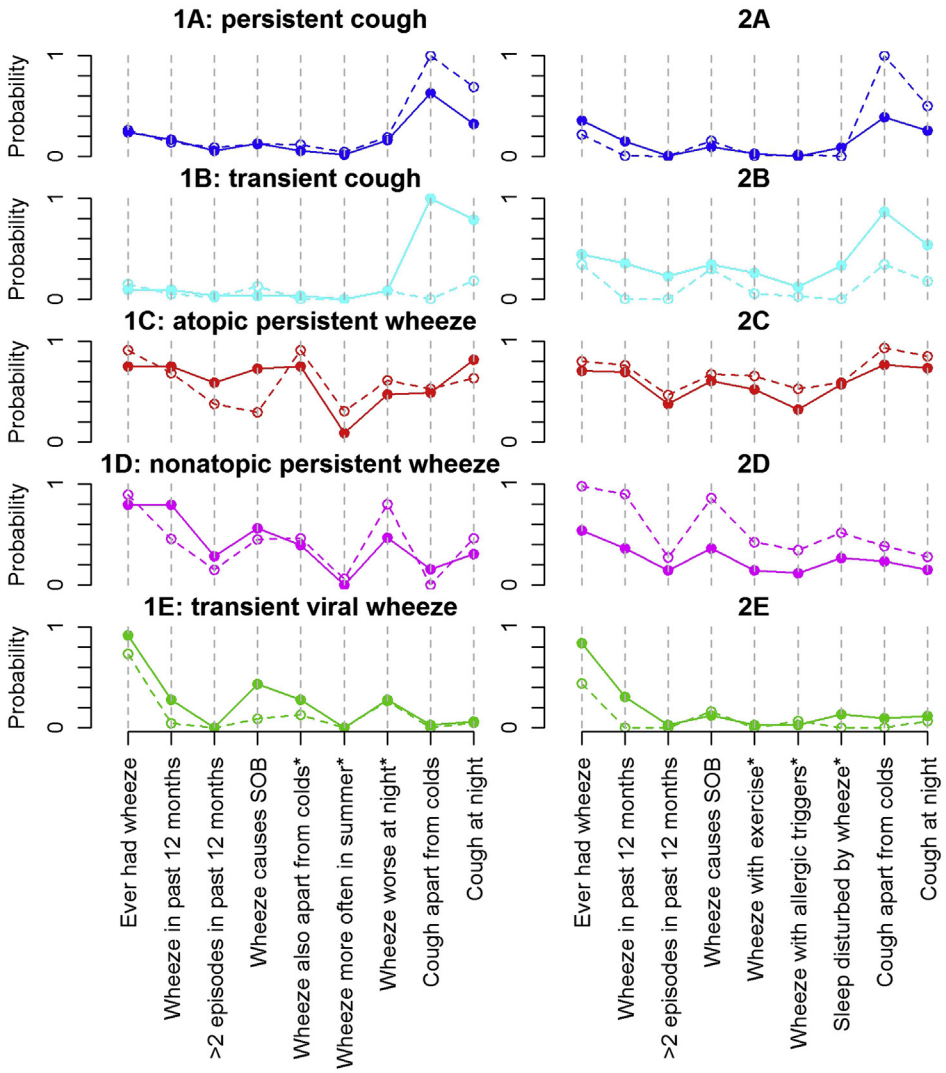

FIG 2. Symptom profiles of identified phenotypes at preschool *(solid lines, filled circles)* and school age *(dotted lines, empty circles)* according to parent-reported symptoms included in the latent class model. Data are probabilities for reporting symptoms given phenotype membership as estimated by the model. *SOB*, Shortness of breath.

TABLE IV. Prevalence of prognostic outcomes in preadolescence (5 y later) of identified phenotypes and an asymptomatic control group

	Phenotype											
The original cohort	**1A: Persistent cough**		**1B: Transient cough**		**1C: Atopic persistent wheeze**		**1D: Nonatopic persistent wheeze**		**1E: Transient viral wheeze**		**Control**	
Current wheeze	27	17-37	8	1-14	71	57-84	34	18-50	25	9-40	11	6-16
≥4 Wheeze episodes	11	4-18	1	0-4	35	22-48	14	3-26	8	0-18	2	0-4
Bronchodilater use	31	21-42	4	0-10	78	67-90	46	29-62	27	12-42	11	6-16
Cough apart from colds	44	33-55	18	9-28	47	32-61	22	8-37	19	5-33	12	7-17
	Phenotype											
The validation cohort	**2A**		**2B**		**2C**		**2D**		**2E**		**Control**	
Current wheeze	12	7-16	16	10-21	61	51-71	53	44-62	14	8-20	8	5-11
≥4 Wheeze episodes	4	2-7	3	1-6	28	20-37	23	15-31	5	1-8	2	1-3
Bronchodilater use	10	6-15	14	8-20	59	50-68	60	50-69	17	10-23	8	5-10
Cough apart from colds	59	52-66	49	41-57	78	70-87	67	58-76	43	33-52	29	24-33

Data are prevalence in % (first column) and 95% CIs (second column). These data are shown as odds ratios for the outcomes comparing children assigned to the identified phenotypes with controls in Fig 3. All the outcomes refer to the past 12 months.

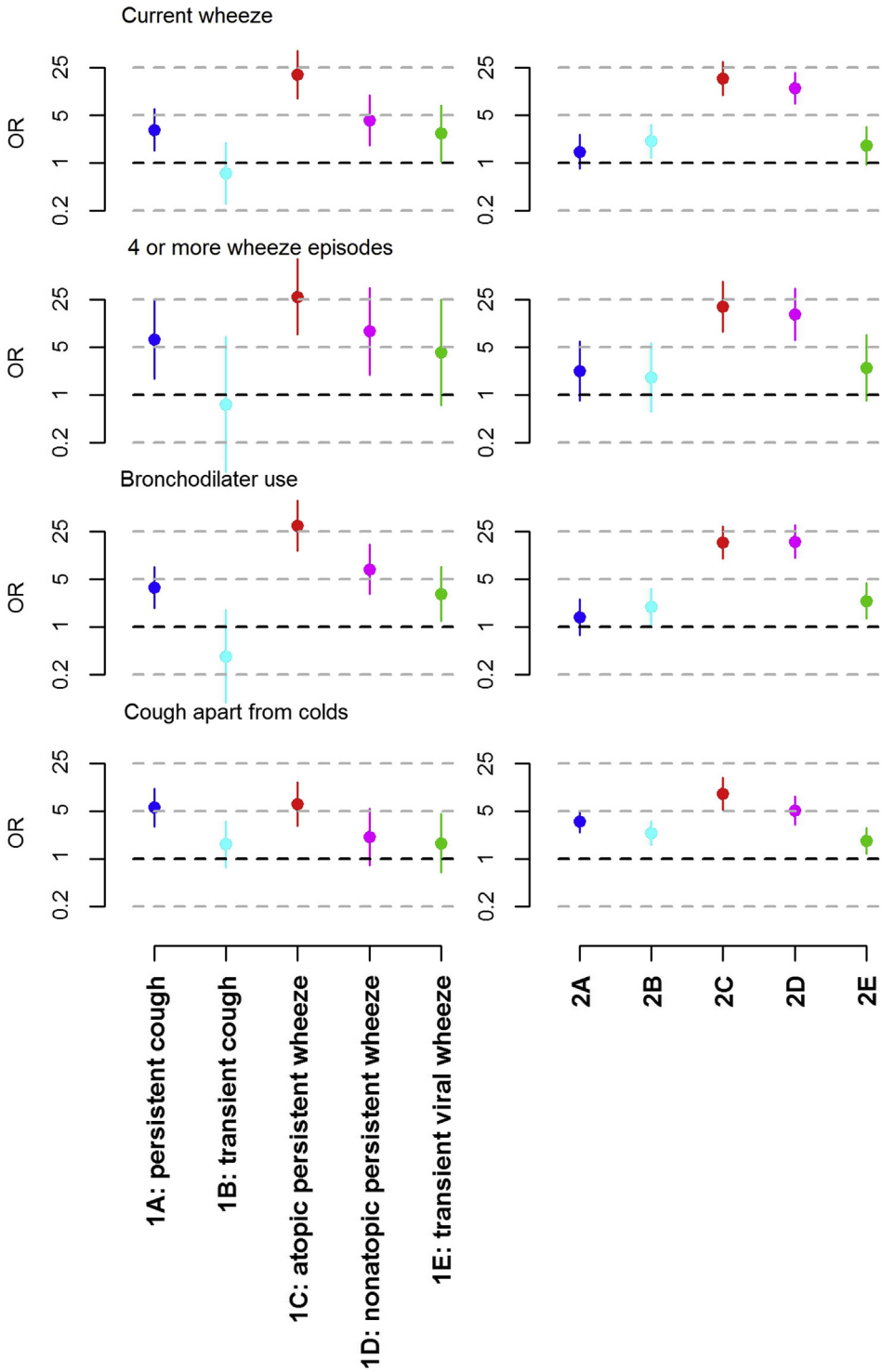

FIG 3. Odds ratios *(ORs)* and 95% CIs *(error bars)* for prognostic outcomes in preadolescence (5 years later) comparing children assigned to the identified phenotypes with an asymptomatic control group. Outcomes refer to the past 12 months.

Comparison with other clustering approaches

Other studies have used a clustering approach to identify phenotypes of respiratory disease in children.[14,16-20] These studies have investigated either repeated measurements of the single disease dimension, current wheeze,[14,16] or single, cross-sectional measurements of multiple disease dimensions, for example, different symptoms.[17-19] To our knowledge, the present study is the only study that used repeated measurements of multiple asthma-related dimensions to identify phenotypes.

Henderson et al, using data from the ALSPAC, applied LCA to repeated assessments of current wheeze from birth to age 7 years. They identified 6 phenotypes (trajectories) of wheeze, which

differed by age of onset and persistence.[14] In a validation study applying the same method in an independent cohort, 5 phenotypes were identified.[16] Some of these were closely similar to those of the original study, while others differed somewhat. Wheeze phenotypes that were similar between the studies also showed consistent associations with lung function, BHR, and atopy.

Subsequently, 3 French studies investigated respiratory phenotypes based on multiple variables in children, using different clustering methods[17-19]: A recent LCA of data on respiratory symptoms (wheeze, night cough, rhinitis) and atopy in children aged 18 months from the Pollution and Asthma Risk Infant Study identified a mild phenotype, a nonatopic severe phenotype, and an atopic severe phenotype.[19] In 2 separate studies of patients with asthma from the Trousseau Asthma Program in Paris, a broad range of clinical and physiological parameters was analyzed in children aged younger than 3 years[17] and 6 to 12 years,[18] respectively, showing both similarities and differences in identified phenotypes between the age groups. By applying LCA to data on respiratory symptoms in children aged 8 to 12 years, a recent cross-sectional study from Spain identified 3 phenotypes with wheeze and 3 phenotypes with cough as the predominant symptom.[20]

Identified phenotypes in context of the literature

The atopic persistent phenotype (1C, 2C) was strongly consistent across the 2 cohorts, and was comparable to phenotypes identified by using other approaches. In the ALSPAC study, 2 phenotypes characterized by early- and intermediate-onset wheeze, respectively, and persistence up to age 7 years were associated with skin prick test positivity, BHR, and reduced lung function.[14] Similar findings were obtained in the Tucson cohort, where children with early-onset persisting wheeze had poor prognosis.[2,35,36] Genetic associations with the well-replicated childhood asthma locus ORMDL3/GSDMB on chromosome 17q21 were strong for early-onset persistent wheeze but absent for early transient wheeze, suggesting distinct genetic origins.[5,37] This locus does not, however, appear to be associated with atopy.[5,37,38]

The other consistently reproduced phenotype was that of transient wheeze (1E and 2E), with mild, virus-triggered symptoms that subsided during preschool or early school years. This is consistent with findings from Tucson, where children with early transient wheeze were less likely than persistent wheezers to have frequent episodes of wheeze or wheeze apart from colds in infancy.[2] In addition, one of the phenotypes emerging in the Trousseau Asthma Program cluster analysis for age younger than 3 years was characterized by mild episodic viral wheeze. However, it should be noted that despite mild symptoms in early life, the children with transient wheeze in our study remained more likely than controls to have current wheeze and use bronchodilators in preadolescence.

In both our cohorts, a third wheeze phenotype was identified (1D and 2D), which was intermediate between the atopic persistent and the viral wheeze phenotype regarding prognosis in preadolescence. This phenotype was not entirely consistent between the original cohort and the validation cohort. Most notably, it included mainly nonatopic children in the original cohort, but a large proportion of atopic children in the validation cohort. The measurements (skin prick tests) had been taken at a later age in the validation cohort, and this is likely to have contributed to the higher prevalence of atopy in the validation sample (43%) compared with that in the original sample (27%) (Table I). Possibly, phenotype 2D of the validation cohort is associated with late-onset atopy—and late-onset wheeze as suggested by the higher prevalence of current wheeze at school age than at preschool age—and was not identified as such in the original sample because of the earlier measurement of atopy. Interestingly, the Trousseau Asthma Program studies revealed a cluster of nonatopic uncontrolled wheeze in the younger age group (<3 years), whereas all 3 clusters in the older age group (6-12 years) showed atopic features, and atopy lost importance as a distinguishing feature.[17,18] Similarly, the nonatopic severe wheeze phenotype identified in the recent analysis from the Pollution and Asthma Risk Infant Study was based on symptoms and measurements assessed at a young age (18 months).[19]

In both cohorts, LCA distinguished 2 phenotypes of recurrent cough. In our original study (based on the original cohort),[13] we suggested that the persistent cough phenotype (1A) could correspond to what has been labeled cough variant asthma, because these children showed increased bronchial responsiveness and an increased risk of later wheezing and bronchodilator use (Fig 3). The 2 cough phenotypes were not replicated consistently in the validation cohort. Again, 2 phenotypes were distinguished, but they were somewhat different from those defined in the original cohort. It should be noted, however, that the LCA was based on a large number of symptoms relating to wheeze, while only 2 questions relating to recurrent cough (cough apart from colds and night cough) were included. This might not have been enough to represent differences in clinical picture within children who cough. Again, age differences at the time of measurement might have contributed to the difference in phenotypes.

Further research is needed to assess the robustness of the posited entities cough variant asthma and nonatopic asthma.

Strengths and limitations

Our study used a repeatable and objective method to identify phenotypes. By including a wide range of symptoms assessed at 2 time points, we identified phenotypes that reflect both temporal patterns (transience, persistence) and symptom picture at a given time point (triggers, severity). Because our study samples were population-based, identified phenotypes cover the entire severity spectrum.

In accordance with our original study, we did not include children without symptoms of cough or wheeze (designated as asymptomatic) in the sample used for phenotype definition. Because these symptoms are also included as variables in the LCA, the central assumption of LCA, that variables are independent within classes, does not strictly hold. This selection may therefore have biased our results. To investigate the extent of bias, we repeated the LCA by using a sample that included the asymptomatic children (data not shown). The best model according to the BIC was again a model with 5 phenotypes, 4 of which each closely corresponded to one of the phenotypes 2A to 2D. The last phenotype included almost all asymptomatic children and children allocated to 2E in the present analyses (corresponding to transient viral wheeze in our original study). These findings are in line with previous studies in which children with few or mild symptoms formed a separate phenotype.[14,16]

We therefore judge that the bias in the present analyses was minimal.

There was a considerable difference between the 2 cohorts in age when the measurements were taken. For most children of the validation cohort, laboratory measurements and the information on prognostic outcomes were, in fact, collected at around the same time. The importance of included phenotypic characteristics such as atopy varies with age, and the age when these were assessed is therefore likely to have affected the results of our study. Other differences between the 2 cohorts, such as the different assessment of night symptoms and seasonality, might have also affected our results. The inclusion of children of South Asian ethnicity in the validation cohort might also have influenced results, but only if we assume that phenotypes differ between whites and South Asians. Our study included only few symptoms relating to cough. Inclusion of additional clinical features (upper and lower respiratory tract symptoms, cough challenges, or other measurements) might allow one to more reliably identify cough phenotypes.

Our study shows that phenotypes of wheeze can be identified consistently from observed data by using more objective statistical tools, rather than relying on expert opinion. Two phenotypes—atopic persistent wheeze and transient viral wheeze—emerged consistently in both cohorts and are likely to reflect distinct disease processes. The age when symptoms are assessed and measurements carried out may have a major influence on results and must be considered carefully whenever defining phenotypes and comparing them across studies.

We are grateful to all the children and their parents for participating in the study. We thank Erol Gaillard for critically reviewing an earlier version of the article and Kali Tal for her editorial assistance. The early phase of data collection in the original cohort was carried out by Drs Hamish Simpson, David Luyt, and Adrian Brooke.

Key messages

- **Using LCA, we identified 2 phenotypes of cough and 3 phenotypes of wheeze in 2 independent cohort studies.**
- **Two phenotypes—atopic persistent wheeze and transient viral wheeze—emerged consistently in both cohorts and might represent distinct disease entities.**
- **The age when symptoms are assessed and measurements taken may have a major influence on results and should be considered carefully when defining and comparing phenotypes.**

REFERENCES

1. Silverman M. Out of the mouths of babes and sucklings: lessons from early childhood asthma. Thorax 1993;48:1200-4.
2. Martinez FD, Wright AL, Taussig LM, Holberg CJ, Halonen M, Morgan WJ. Asthma and wheezing in the first six years of life. N Engl J Med 1995;332:133-8.
3. Silverman M, Grigg J, Mc Kean M. Virus-induced wheeze in young children—a separate disease? In: Johnston S, Papadopoulos N, editors. Respiratory infections in allergy and asthma. New York: Marcel Dekker; 2002. p. 427-71.
4. Edwards CA, Osman LM, Godden DJ, Douglas JG. Wheezy bronchitis in childhood: a distinct clinical entity with lifelong significance? Chest 2003;124:18-24.
5. Spycher BD, Henderson J, Granell R, Evans DM, Smith GD, Timpson NJ, et al. Genome-wide prediction of childhood asthma and related phenotypes in a longitudinal birth cohort. J Allergy Clin Immunol 2012;130:503-9.e7.
6. van Asperen PP. Cough and asthma. Paediatr Respir Rev 2006;7:26-30.
7. Brodlie M, Graham C, McKean MC. Childhood cough. BMJ 2012;344:e1177.
8. Goldsobel AB, Chipps BE. Cough in the pediatric population. J Pediatr 2010;156:352-8.
9. Turcotte SE, Lougheed MD. Cough in asthma. Curr Opin Pharmacol 2011;11:231-7.
10. Lougheed MD, Turcotte SE, Fisher T. Cough variant asthma: lessons learned from deep inspirations. Lung 2012;190:17-22.
11. Wassall HJ, Devenny AM, Daud Khan S, Ninan TK, Russell G. A comparison of virus-associated and multi-trigger wheeze in school children. J Asthma 2005;42:737-44.
12. Spycher BD, Minder CE, Kuehni CE. Multivariate modelling of responses to conditional items: new possibilities for latent class analysis. Stat Med 2009;28:1927-39.
13. Spycher BD, Silverman M, Brooke AM, Minder CE, Kuehni CE. Distinguishing phenotypes of childhood wheeze and cough using latent class analysis. Eur Respir J 2008;31:974-81.
14. Henderson J, Granell R, Heron J, Sherriff A, Simpson A, Woodcock A, et al. Associations of wheezing phenotypes in the first 6 years of life with atopy, lung function and airway responsiveness in mid-childhood. Thorax 2008;63:974-80.
15. Spycher BD, Silverman M, Kuehni CE. Phenotypes of childhood asthma: are they real? Clin Exp Allergy 2010;40:1130-41.
16. 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.
17. Just J, Gouvis-Echraghi R, Couderc R, Guillemot-Lambert N, Saint-Pierre P. Novel severe wheezy young children phenotypes: boys atopic multiple-trigger and girls nonatopic uncontrolled wheeze. J Allergy Clin Immunol 2012;130:103-10.
18. Just J, Gouvis-Echraghi R, Rouve S, Wanin S, Moreau D, Annesi-Maesano I. Two novel, severe asthma phenotypes identified during childhood using a clustering approach. Eur Respir J 2012;40:55-60.
19. Herr M, Just J, Nikasinovic L, Foucault C, Le Marec AM, Giordanella JP, et al. Risk factors and characteristics of respiratory and allergic phenotypes in early childhood. J Allergy Clin Immunol 2012;130:389-96.e4.
20. Weinmayr G, Keller F, Kleiner A, du Prel JB, Garcia-Marcos L, Batlles-Garrido J, et al. Asthma phenotypes identified by latent class analysis in the ISAAC phase II Spain study. Clin Exp Allergy 2013;43:223-32.
21. Kuehni CE, Brooke AM, Strippoli M-PF, Spycher BD, Davis A, Silverman M. Cohort profile: the Leicester respiratory cohorts. Int J Epidemiol 2007;36:977-85.
22. Nystad W, Samuelsen SO, Nafstad P, Edvardsen E, Stensrud T, Jaakkola JJ. Feasibility of measuring lung function in preschool children. Thorax 2002;57:1021-7.
23. Rosenthal M, Bain SH, Cramer D, Helms P, Denison D, Bush A, et al. Lung function in white children aged 4 to 19 years: I—spirometry. Thorax 1993;48:794-802.
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. Vilozni D, Livnat G, Dabbah H, Elias N, Hakim F, Bentur L. The potential use of spirometry during methacholine challenge test in young children with respiratory symptoms. Pediatr Pulmonol 2009;44:720-7.
26. Wilson NM, Phagoo SB, Silverman M. Use of transcutaneous oxygen tension, arterial oxygen saturation, and respiratory resistance to assess the response to inhaled methacholine in asthmatic children and normal adults. Thorax 1991;46:433-7.
27. Wilson NM, Bridge P, Phagoo SB, Silverman M. The measurement of methacholine responsiveness in 5 year old children: three methods compared. Eur Respir J 1995;8:364-70.
28. Wang J, Mochizuki H, Muramatsu R, Arakawa H, Tokuyama K, Morikawa A. Evaluation of bronchial hyperresponsiveness by monitoring of transcutaneous oxygen tension and arterial oxygen saturation during methacholine challenge in asthmatic children. J Asthma 2006;43:145-9.
29. Chinn S. Methodology of bronchial responsiveness. Thorax 1998;53:984-8.
30. Lazarsfeld PF, Henry NW. Latent structure analysis. Boston: Houghton Mifflin; 1968.
31. McLachlan G, Peel D. Finite mixture models. New York: John Wiley & Sons; 2000.

32. Hunt L, Jorgensen M. Mixture model clustering using the MULTIMIX program. Aust N Z J Stat 1999;41:153-71.
33. Hunt L, Jorgensen M. Mixture model clustering for mixed data with missing information. Comput Stat Data Anal 2003;41:429-40.
34. Little RJA, Rubin DB. Statistical analysis with missing data. Chichester, NY: John Wiley & Sons; 2002.
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. Am J Respir Crit Care Med 2005;172:1253-8.
36. Stern DA, Morgan WJ, Halonen M, Wright AL, Martinez FD. Wheezing and bronchial hyper-responsiveness in early childhood as predictors of newly diagnosed asthma in early adulthood: a longitudinal birth-cohort study. Lancet 2008;372:1058-64.
37. Granell R, Henderson AJ, Timpson N, St Pourcain B, Kemp JP, Ring SM, et al. Examination of the relationship between variation at 17q21 and childhood wheeze phenotypes. J Allergy Clin Immunol 2013;131:685-94.
38. Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, Heath S, et al. A large-scale, consortium-based genomewide association study of asthma. N Engl J Med 2010;363:1211-21.

Comparison of phenotypes of childhood wheeze and cough in two independent cohorts

Authors

Ben D Spycher (PhD)[1]

Michael Silverman (MD)[2]

Anina M Pescatore (MSc)[1]

Caroline S Beardsmore (PhD)[2]

Claudia E Kuehni (MD, MSc)[1]

[1]Institute of Social and Preventive Medicine (ISPM), University of Bern, Bern, Switzerland;

[2]The Leicester Children's Asthma Centre, Division of Child Health, Department of Infection, Immunity & Inflammation, University of Leicester, Leicester, UK

Address for correspondence

Ben D Spycher, Institute of Social and Preventive Medicine (ISPM), University of Bern, Finkenhubelweg 11, CH-3012, Bern, Switzerland

Phone: +41 31 631 56 40, Fax: 41 31 631 35 20, e-mail: bspycher@ispm.unibe.ch

Online Repository Tables E1-E4

Pulmonary function

Spirometry was performed according to the guidelines current at the time of testing,[E1, E2] with the child seated and wearing a noseclip. An electronic spirometer was used for both cohorts (Pneumoscreen II/1 Erich Jaeger, Germany and Pneumotrac, Vitalograph, Ireland for Cohorts 1 and 2 respectively). Short-acting ß-agonist bronchodilators were avoided for at least 12 hours before baseline testing, after which children were given salbutamol (4 puffs of 100 mg each via spacer (Ventolin, Glaxo Smith Kline, UK)) using standard techniques.[E3] Spirometry was repeated 15 minutes after bronchodilator administration.

We reported the best of three repeatable measurements of forced expiratory volume in 0.5 seconds ($FEV_{0.5}$) or forced expiratory volume in one second (FEV_1) for Cohorts 1 and 2 respectively. We reported $FEV_{0.5}$ in preference to FEV_1 in the original cohort because many of the children in this younger age group reached their forced vital capacity within one second. Sex- and height-standardized z-scores of $FEV_{0.5}$ and FEV_1 were used for analysis as appropriate.[E4,E5]

Bronchial Challenge Testing

Children who were wheezing on examination were excluded from the bronchial challenge test. In the original cohort, the change in transcutaneous oxygen tension (tc-PO_2) using a transcutaneous oxygen tension probe (Transoxode, Drager, Hemel Hempstead, UK) was employed as an effort-independent, indirect measure of the response to inhaled methacholine.[E6] After a 20-min stabilization period, baseline measurements of tc-PO_2 were recorded. Following inhalation of phosphate-buffered saline, doubling doses of methacholine were administered via a face mask with the nose clipped and the subject seated according to the tidal breathing method.[E7] The protocol for children from the validation cohort was identical, except that the outcome measure was FEV_1 assessed by spirometry instead of tc-PO_2. The starting concentration of methacholine was either 0.25 or 2.0 g/L depending on whether or not the child had wheezed in the last year, respectively.[E8] Children who reported

current cough but no wheeze started at a concentration of 1.0g/L. Doses were then doubled until the maximum dose of 32.0g/L was reached or a response (> 20% fall in tc-pO_2 from pre-saline baseline or a reduction in FEV_1 of 20%) was obtained. All subjects from the original cohort except 2 responded by the final concentration. Children responding to methacholine were given 2.5 mg salbutamol via a nebulizer. Children from the validation cohort were given salbutamol via MDI and spacer. Tc-PO_2 or FEV_1 were measured 15 min afterwards to ensure return to baseline. For latent class analysis, the values of PC_{20} and $PC_{20}P_{tc,O2}$ were logarithmically transformed to improve approximation by the normal distribution,[E9] the assumed distribution within latent classes.

Assessment of atopic status

Both cohorts underwent skin prick testing against common allergens. Any anti-histamine therapy was suspended 48 hours before the test, and systemic steroids in excess of 10mg Prednisolone/day were suspended 24 hours before testing. Children were tested against cat, dog, *Dermatophagoides pteronyssinus*, and mixed grass pollens (Bencard, Brentford, Middlesex, UK (original cohort), Alk Abello 2, UK (validation cohort)). The positive control was histamine 1mg/ml; negative control was the carrier solution. The reaction was assessed 5 to 15 min after skin prick testing according to the method of Pepys.[E10] Responses were deemed positive if a wheal of diameter 2mm or more was observed and was larger than any response to the negative control. If one or more positive reactions were observed, the child was designated as atopic.

Table E1: Variables included in latent class analyses to define phenotypes in cohorts 1 and 2

Original cohort	Validation cohort
Demographic variables	
- Sex	- Sex
- Age at preschool survey (categories: 0-2 yrs, 3-5 yrs)	- Age at preschool survey (categories: 1-2 yrs, 3-4 yrs)
Physiological measurements	
- Skin-prick tests: cat, dog, house dust mite, grasses (all negative, ≥1 positive)	- Skin-prick tests: cat, dog, house dust mite, grasses (all negative, ≥1 positive)
- $FEV_{0.5}$ (sex and height-standardized z-scores[4])	- FEV_1 (sex and height-standardized z-scores[5])
- Bronchial responsiveness $PC_{20}P_{tc}O_2$ (log transformed[9])	- Bronchial responsiveness PC_{20} (log transformed[9])
Symptoms assessed at preschool and school-age	
- Ever had attacks of wheezing (never, at preschool-age, at preschool- and school-age, at school-age)	- Ever had whistling or wheezing in the chest (never, preschool, preschool and school-age, school-age)
- Number of wheezing attacks during the past 12 months (0, 1-2, ≥ 3)*	- Number of wheezing attacks during the past 12 months (0, 1-2, ≥ 3)*†
- Wheezing attacks cause shortness of breath (no, yes)*	- Wheezing attacks cause shortness of breath (no, yes)*
- Wheeze with or without colds	- Wheeze during or soon after colds (no, yes)*

(only with colds, occasionally apart from colds [due to running, playing, eating, drinking])*	- Wheeze caused by running, playing, laughing, crying or excitement (no, yes)*
- Time of year with more frequent wheezing attacks (indifferent, winter, summer)*	- Wheeze caused by dust, grass, pets or other animals, eating or drinking (no, yes)*
- Time of day when wheezing is worse (indifferent, night, day)*	- Sleep disturbed due to wheezing (no, yes)*
- Cough (no cough, only with colds, also apart from colds)	- Cough (no cough, only with colds, also apart from colds)
- Wakened by cough at night without chest infection (no, yes)	- Dry cough at night without chest infection (no, yes)

* Asked only to those reporting wheeze ever in respective survey

† In a subsample of 548 children different categories were used, with the number of attacks of wheezing being divided into 0, 1-3, and >3.

Table E2: Fitted parameter values for 5-class latent class model in the original cohort

Phenotype		A1	B1	C1	D1	E1
Prevalence in sample		0.31	0.25	0.18	0.15	0.11
Sex	Female	0.56	0.52	0.51	0.51	0.27
	Male	0.44	0.48	0.49	0.49	0.73
Age at preschool survey	0 to 2 yrs	0.39	0.54	0.30	0.61	0.36
	3 to 5 yrs	0.61	0.46	0.70	0.39	0.64
Wheeze ever	Never	0.66	0.82	0.00	0.00	0.00
	P yes, S no	0.08	0.02	0.09	0.10	0.27
	P yes, S yes	0.16	0.07	0.66	0.69	0.65
	P no, S yes	0.10	0.08	0.25	0.21	0.09
P, frequency of attacks†	0	0.07	0.00	0.00	0.00	0.63
	1 to 2	0.11	0.06	0.16	0.51	0.28
	>2	0.06	0.04	0.59	0.28	0.00
P, attacks with shortness of breath†	No	0.11	0.06	0.02	0.23	0.48
	Yes	0.13	0.04	0.73	0.56	0.43
P, triggers of wheeze†	Only colds	0.18	0.06	0.00	0.40	0.64
	Colds and other	0.06	0.04	0.75	0.39	0.28
P, season with most frequent attacks†	Indifferent	0.18	0.09	0.43	0.41	0.83
	Winter	0.04	0.00	0.22	0.39	0.08
	Summer	0.02	0.00	0.09	0.00	0.00
P, time of day with worse attacks†	Indifferent	0.08	0.00	0.23	0.23	0.64
	Night	0.16	0.08	0.47	0.46	0.28
	Day	0.00	0.01	0.04	0.10	0.00
P, wakened by cough at night	No	0.68	0.21	0.18	0.69	0.94
	Yes	0.32	0.79	0.82	0.31	0.06
P, cough	No cough	0.09	0.00	0.11	0.12	0.32
	Only with colds	0.29	0.00	0.40	0.72	0.65

	Also without colds	0.63	1.00	0.49	0.16	0.03
S, frequency of attacks†	0	0.12	0.10	0.23	0.44	0.69
	1 to 2	0.05	0.02	0.31	0.31	0.05
	≥ 3	0.09	0.02	0.38	0.15	0.00
S, attacks with shortness of breath†	No	0.14	0.02	0.61	0.45	0.64
	Yes	0.12	0.13	0.30	0.45	0.09
S, triggers of wheeze†	Only colds	0.14	0.15	0.00	0.43	0.60
	Colds and other	0.12	0.00	0.91	0.46	0.13
S, season with most frequent attacks†	Indifferent	0.08	0.15	0.29	0.08	0.73
	Winter	0.13	0.00	0.32	0.75	0.00
	Summer	0.04	0.00	0.31	0.06	0.00
S, time of day with worse attacks†	Indifferent	0.07	0.04	0.18	0.00	0.46
	Night	0.19	0.09	0.61	0.80	0.27
	Day	0.00	0.02	0.12	0.09	0.00
S, wakened by cough at night	No	0.31	0.81	0.37	0.54	0.95
	Yes	0.69	0.19	0.63	0.46	0.05
S, cough	No cough	0.00	0.15	0.09	0.17	0.33
	Only with colds	0.00	0.85	0.38	0.83	0.67
	Also without colds	1.00	0.00	0.53	0.00	0.00
Skin prick tests	All negative	0.81	0.84	0.30	0.91	0.78
	≥ 1 positive	0.19	0.16	0.70	0.09	0.22
$FEV_{0.5}$ (z-scores)	Mean	-1.59	-1.18	-1.80	-1.47	-1.09
	Variance	1.41	1.05	1.41	0.57	0.96
log PC_{20}-$P_{tc,O2}$	Mean	0.88	1.01	0.23	0.84	0.91
	Variance	0.60	0.90	1.18	0.58	0.65

Data presented as proportion (prevalence), probability (symptoms) mean and variance (continuous measures). P: preschool survey; S: school-age survey; † Conditional probability given a positive report of wheeze ever in that survey.[E11]

Table E3: Fitted parameter values for 5-class latent class model in the validation cohort

Phenotype		A2	B2	C2	D2	E2
Prevalence in sample		0.30	0.22	0.16	0.15	0.17
Sex	Female	0.53	0.45	0.40	0.44	0.50
	Male	0.47	0.55	0.60	0.56	0.50
Age at preschool survey	0 to 2 yrs	0.80	0.69	0.63	0.70	0.80
	3 to 5 yrs	0.20	0.31	0.37	0.30	0.20
Wheeze ever	Never	0.58	0.52	0.13	0.00	0.00
	P yes, S no	0.20	0.13	0.07	0.02	0.56
	P yes, S yes	0.16	0.31	0.63	0.52	0.28
	P no, S yes	0.06	0.03	0.16	0.46	0.16
P, frequency of attacks†	0	0.58	0.20	0.02	0.33	0.63
	1-2	0.39	0.29	0.45	0.40	0.33
	≥ 3	0.02	0.51	0.54	0.27	0.03
P, attacks with shortness of breath†	No	0.74	0.23	0.14	0.33	0.85
	Yes	0.26	0.77	0.86	0.67	0.15
P, wheeze with colds†	No	0.26	0.06	0.03	0.20	0.35
	Yes	0.74	0.94	0.97	0.80	0.65
P, weeze triggered by exercise†	No	0.92	0.42	0.26	0.73	0.97
	Yes	0.08	0.58	0.74	0.27	0.03
P, wheeze with allergic triggers†	No	1.00	0.72	0.54	0.78	0.97
	Yes	0.00	0.28	0.46	0.22	0.03
P, sleep disturbed by wheezing†	No	0.76	0.25	0.19	0.50	0.84
	Yes	0.24	0.75	0.81	0.50	0.16
P, dry cough at night	No	0.75	0.46	0.27	0.85	0.88
	Yes	0.25	0.54	0.73	0.15	0.12
P, cough	No cough	0.21	0.01	0.00	0.16	0.27
	Only with colds	0.40	0.13	0.24	0.60	0.63
	Also without colds	0.39	0.86	0.76	0.24	0.10

S, frequency of attacks†	0	0.97	0.99	0.04	0.08	1.00
	1-2	0.03	0.01	0.37	0.64	0.00
	>2	0.00	0.00	0.58	0.28	0.00
S, attacks with shortness of breath†	No	0.30	0.12	0.16	0.12	0.63
	Yes	0.70	0.88	0.84	0.88	0.37
S, wheeze with colds†	No	0.99	0.91	0.10	0.08	0.95
	Yes	0.01	0.09	0.90	0.92	0.05
S, weeze triggered by exercise†	No	0.97	0.83	0.18	0.57	1.00
	Yes	0.03	0.17	0.82	0.43	0.00
S, wheeze with allergic triggers†	No	0.91	0.92	0.34	0.65	0.85
	Yes	0.09	0.08	0.66	0.35	0.15
S, sleep disturbed by wheezing†	No	1.00	1.00	0.26	0.47	1.00
	Yes	0.00	0.00	0.74	0.53	0.00
S, dry cough at night	No	0.50	0.82	0.15	0.72	0.93
	Yes	0.50	0.18	0.85	0.28	0.07
S, cough	No cough	0.00	0.22	0.00	0.11	0.31
	Only with colds	0.00	0.43	0.07	0.50	0.69
	Also without colds	1.00	0.34	0.93	0.38	0.00
Skin prick tests	All negative	0.66	0.65	0.36	0.33	0.72
	≥ 1 positive	0.34	0.35	0.64	0.67	0.28
FEV_1 (z-scores)	Mean	-0.12	-0.24	-0.53	-0.38	-0.31
	Variance	1.02	1.00	1.13	0.82	0.93
log PC20	Mean	2.67	2.29	1.96	1.55	2.80
	Variance	1.68	1.72	2.50	1.40	1.96

Data presented as proportion (prevalence), probability (symptoms) mean and variance (continuous measures). P: preschool survey; S: school-age survey; † Conditional probability given a positive report of wheeze ever in that survey.[E11]

Table E4: Fitted parameter values for 7-class latent class model in the validation cohort

Phenotype		Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7
Prevalence in sample		0.22	0.20	0.16	0.14	0.14	0.09	0.06
Sex	Female	0.52	0.46	0.44	0.36	0.50	0.50	0.58
	Male	0.48	0.54	0.56	0.64	0.50	0.50	0.42
Age at preschool survey	0 to 2 yrs	0.84	0.69	0.65	0.64	0.82	0.66	0.80
	3 to 5 yrs	0.16	0.31	0.35	0.36	0.18	0.34	0.20
Wheeze ever	Never	0.67	0.64	0.25	0.00	0.00	0.00	0.00
	P yes, S no	0.16	0.06	0.07	0.04	0.50	0.43	0.45
	P yes, S yes	0.13	0.30	0.56	0.54	0.30	0.41	0.24
	P no, S yes	0.05	0.00	0.12	0.42	0.20	0.16	0.31
P, frequency of attacks†	0	0.40	0.11	0.00	0.36	0.77	0.70	0.04
	1-2	0.54	0.26	0.42	0.37	0.23	0.30	0.66
	≥ 3	0.06	0.63	0.58	0.27	0.00	0.00	0.30
P, attacks with shortness of breath†	No	0.69	0.19	0.15	0.27	0.94	0.61	0.61
	Yes	0.31	0.81	0.85	0.73	0.06	0.39	0.39
P, wheeze with colds†	No	0.00	0.00	0.03	0.22	0.46	0.59	0.00
	Yes	1.00	1.00	0.97	0.78	0.54	0.41	1.00
P, weeze triggered by exercise†	No	0.83	0.44	0.22	0.66	1.00	0.81	0.83
	Yes	0.17	0.56	0.78	0.34	0.00	0.19	0.17
P, wheeze with allergic triggers†	No	1.00	0.70	0.53	0.70	0.96	0.95	0.95
	Yes	0.00	0.30	0.47	0.30	0.04	0.05	0.05
P, sleep disturbed by wheezing†	No	0.74	0.08	0.15	0.50	0.93	0.76	0.67
	Yes	0.26	0.92	0.85	0.50	0.07	0.24	0.33
P, dry cough at night	No	0.93	0.50	0.17	0.85	0.95	0.21	0.81
	Yes	0.07	0.50	0.83	0.15	0.05	0.79	0.19
P, cough	No cough	0.28	0.01	0.00	0.11	0.31	0.03	0.18
	Only with colds	0.48	0.14	0.19	0.61	0.64	0.15	0.57

	Also without colds	0.23	0.86	0.81	0.28	0.05	0.83	0.25
S, frequency of attacks†	0	1.00	1.00	0.04	0.09	1.00	0.97	0.06
	1-2	0.00	0.00	0.39	0.53	0.00	0.03	0.85
	>2	0.00	0.00	0.57	0.38	0.00	0.00	0.09
S, attacks with shortness of breath†	No	0.00	0.09	0.14	0.17	0.64	0.36	0.00
	Yes	1.00	0.91	0.86	0.83	0.36	0.64	1.00
S, wheeze with colds†	No	0.94	0.92	0.09	0.13	0.97	0.94	0.00
	Yes	0.06	0.08	0.91	0.87	0.03	0.06	1.00
S, weeze triggered by exercise†	No	1.00	0.83	0.19	0.39	1.00	0.95	0.91
	Yes	0.00	0.17	0.81	0.61	0.00	0.05	0.09
S, wheeze with allergic triggers†	No	1.00	0.94	0.39	0.44	0.84	0.86	1.00
	Yes	0.00	0.06	0.61	0.56	0.16	0.14	0.00
S, sleep disturbed by wheezing†	No	1.00	1.00	0.27	0.39	1.00	1.00	0.65
	Yes	0.00	0.00	0.73	0.61	0.00	0.00	0.35
S, dry cough at night	No	0.56	0.87	0.15	0.58	0.94	0.42	1.00
	Yes	0.44	0.13	0.85	0.42	0.06	0.58	0.00
S, cough	No cough	0.00	0.24	0.00	0.09	0.34	0.01	0.18
	Only with colds	0.00	0.41	0.07	0.35	0.64	0.23	0.82
	Also without colds	1.00	0.34	0.93	0.56	0.02	0.75	0.00
Skin prick tests	All negative	0.63	0.67	0.44	0.20	0.69	0.74	0.73
	≥ 1 positive	0.37	0.33	0.56	0.80	0.31	0.26	0.27
FEV_1 (z-scores)	Mean	-0.12	-0.27	-0.36	-0.61	-0.34	-0.14	-0.09
	Variance	0.94	0.91	1.02	0.96	0.68	1.45	1.27
log PC20	Mean	2.62	2.31	2.16	1.24	2.73	2.59	2.81
	Variance	1.35	1.67	2.86	1.24	2.42	1.04	0.77

Data presented as proportion (prevalence), probability (symptoms) mean and variance (continuous measures). P: preschool survey; S: school-age survey; † Conditional probability given a positive report of wheeze ever in that survey.[E11]

References

E1. Standardization of Spirometry, 1994 Update. American Thoracic Society. Am J Respir Crit Care Med 1995; 152:1107-36.

E2. Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, Coates A, et al. Standardisation of spirometry. Eur Respir J 2005; 26:319-38.

E3. 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.

E4. Nystad W, Samuelsen SO, Nafstad P, Edvardsen E, Stensrud T, Jaakkola JJ. Feasibility of measuring lung function in preschool children. Thorax 2002; 57:1021-7.

E5. Rosenthal M, Bain SH, Cramer D, Helms P, Denison D, Bush A, et al. Lung function in white children aged 4 to 19 years: I--Spirometry. Thorax 1993; 48:794-802.

E6. Wilson NM, Phagoo SB, Silverman M. Use of transcutaneous oxygen tension, arterial oxygen saturation, and respiratory resistance to assess the response to inhaled methacholine in asthmatic children and normal adults. Thorax 1991; 46:433-7.

E7. Cockcroft DW, Killian DN, Mellon JJ, Hargreave FE. Bronchial reactivity to inhaled histamine: a method and clinical survey. Clin Allergy 1977; 7:235-43.

E8. Hargreave FE, Ryan G, Thomson NC, O'Byrne PM, Latimer K, Juniper EF, et al. Bronchial responsiveness to histamine or methacholine in asthma: measurement and clinical significance. J Allergy Clin Immunol 1981; 68:347-55.

E9. Chinn S. Methodology of bronchial responsiveness. Thorax 1998; 53:984-8.

E10. Pepys J. Skin tests in diagnosis. In: Gell PH, Coombs RRA, Lach PJ, editors. Clinical Aspects of Immunology. 3 ed. Oxford: Blackwell Scientific Publications; 1975. p. 55-80.

E11. Spycher BD, Minder CE, Kuehni CE. Multivariate modelling of responses to conditional items: New possibilities for latent class analysis. Stat Med 2009; 28:1927-39.

Figure E1: Flow diagram of study subjects in the original cohort and the validation cohort.

Children reporting wheeze or cough in the preschool and/or school-age survey were included in the sample used to identify phenotypes (■). Identified phenotypes were compared with each other and an asymptomatic control group with respect to outcomes at 5 year follow-up (□). In the original cohort pre-school data came from the 1990 survey (when children were aged 0-5 years; response rate 86%) and school-age data came from the1992-94 survey (when children were aged 4-8 years; response rate 61%). In the validation cohort, pre-school data came for each child from the survey when their age at response was closest to 3.0 years, but between 1 and 4 years (either 1998 or 1999 survey; response rates 80% and 78% respectively), and school-age data came from the survey when their age was closest to 7.0 years, but between 5 and 8 years (either survey 2001 or 2003, response rates 59% and 52% respectively). The pre-adolescent survey was the 1998 survey for the original cohort and, for the validation cohort, the 2006 or 2010 survey, respectively, depending on whether the school-age survey 2001 or 2003 was used.