

Endotyping early childhood asthma by quantitative symptom assessment

Hans Bisgaard, MD, DMSci, Christian Bressen Pipper, MSc, PhD, and Klaus Bønnelykke, MD, PhD *Copenhagen, Denmark*

Background: Asthmatic symptoms in young children reflect a heterogeneous group of diseases. Symptoms remain the primary end-point in both research and clinical management, but there is a need for standardized symptom assessment.

Objective: We sought to explore endotyping of early childhood asthma by prospective daily diary recordings of globally assessed symptoms during the first 6 years of life.

Methods: Globally assessed troublesome lung symptoms were recorded in daily diaries during the first 6 years of life in the Copenhagen Prospective Studies on Asthma in Childhood birth cohort born of mothers with asthma. Symptom recordings adjusted for missing values were used to categorize children based on the temporal symptom pattern. We propose an alternative approach of quantitating symptom frequency and longitudinal assessment of age-at-onset to segment children.

Results: Six years of daily diary recordings were available in 307 children (75% of the birth cohort). We confirmed the archetypal temporal categories of transient early, persistent, and late-onset troublesome lung symptoms based on 3-year periods, finding no benefit from a finer temporal categorization of 2- or 1-year periods. Restricting categorization to symptoms during the summer improved specificity at the expense of sensitivity. Our alternative approach quantitating symptom frequency and age-at-onset exhibited a more powerful association with *ORMDL3*, whereas the study power was lost by restricting to doctor-verified wheeze.

Conclusions: We propose a novel method for endotyping of early childhood asthma based on the frequency and age-of-onset of globally assessed troublesome lung symptoms analyzed longitudinally. This method showed the closest association with genetic variants, hence underlying molecular mechanisms and endotypes. (J Allergy Clin Immunol 2011;■■■:■■■-■■■.)

Key words: Asthma, young children, patterns, endotypes, *ORMDL3*, symptoms

Asthma and other wheezing disorders are among the most common causes of hospitalization and other health care use in young children representing a significant burden to the child, the family, the health care system, and society.^{1,2}

These very common symptoms in the first years of life represent a heterogeneity of diseases with little to differentiate the clinical presentations, which is a cause of inadequate treatment.^{3,4} Asthma and other wheezy disorders are likely to embody several endotypes (endophenotypes) associated with distinct clinical features, divergent underlying molecular causes, and different treatment responses.⁵ Segmentation of patients into such endotypes is important for establishing effective preventive measures and determining appropriate individualized treatments.⁶

Objective assessments are rarely accessible in young children, and diagnosis and monitoring is reliant on second-hand symptom description with inaccurate terminology. It has been documented that parental reporting of the specific term wheeze carries a large risk of misclassification and lacks cross-cultural validity and that the term has low sensitivity for identifying disease.⁷⁻¹²

Therefore we used a global description of asthma associated symptoms, hereafter called troublesome lung symptoms, instead of attempting assessment of any specific terminology.^{13,14} The global symptoms were recorded prospectively in day-to-day diary cards from birth to age 6 years in the Copenhagen Prospective Studies on Asthma in Childhood (COPSAC) birth cohort study to provide a stringent quantitative framework for assessing symptoms during the first 6 years of life. Doctor-verified wheeze on auscultation during acute episodes was compared with such global symptom assessment.

The temporal symptom pattern during preschool age has been used to segment children into early transient, late-onset, and persistent asthma symptom categories in early childhood with differential associations to risk factors and intermediate atopic traits.^{15,16} We explored in detail the completeness of such temporal symptom patterns, conceiving a novel method of extracting individual measures of symptom load unbiased from missing registrations.

We propose an alternative approach using longitudinal analysis of symptom frequency and age-at-onset to characterize early

From the Copenhagen Prospective Studies on Asthma in Childhood; Health Sciences, University of Copenhagen; and the Danish Pediatric Asthma Center, Copenhagen University Hospital, Gentofte, Copenhagen.

The Copenhagen Prospective Studies on Asthma in Childhood (COPSAC) is funded by private and public research funds, all of which are listed on www.copsac.com. The Lundbeck Foundation, the Pharmacy Foundation of 1991, the Augustinus Foundation, the Danish Medical Research Council, and the Danish Pediatric Asthma Centre provide core support for COPSAC. The funding agencies did not have any role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Disclosure of potential conflict of interest: H. Bisgaard is a lecturer for AstraZeneca; is a lecturer and consultant for Merck; is a consultant for Chiesi; and has provided legal consultation/expert witness testimony for the European Medicines Agency: guidelines on pediatric studies for documenting asthma drugs. The rest of the authors have declared that they have no conflict of interest.

Received for publication August 25, 2010; revised February 1, 2011; accepted for publication February 2, 2011.

Reprint requests: Hans Bisgaard, MD, DMSci, Copenhagen Prospective Studies on Asthma in Childhood; Health Sciences, University of Copenhagen; and the Danish Pediatric Asthma Center, Copenhagen University Hospital, Gentofte, Ledreborg Alle 34, 2820 Gentofte, Copenhagen, Denmark. E-mail: Bisgaard@copsac.com.

0091-6749/\$36.00

© 2011 American Academy of Allergy, Asthma & Immunology
doi:10.1016/j.jaci.2011.02.007

Abbreviations used

COPSAC: Copenhagen Prospective Studies on Asthma in Childhood
 HR: Hazard ratio
 IQR: Interquartile range
 IR: Intensity ratio
 OR: Odds ratio
 PR: Prevalence ratio

childhood asthma. We hypothesize that such an approach provides a more precise measure for identifying endotypes.

These day-to-day symptom recordings allowed us accurate insight into early symptom patterns and allowed comparison of different methods of segmenting patients with troublesome lung symptoms based on their association with asthma at age 6 years and with the established genetic risk variant *ORMDL3*.¹⁷

METHODS**Subjects**

COPSAC is a longitudinal clinical study of a birth cohort of 411 infants born to mothers with a history of asthma. Mothers were interviewed during pregnancy, and the newborns were enrolled at the age of 1 month; key exclusion criteria were severe congenital abnormality, gestational age of less than 36 weeks, and lung symptoms before enrollment. The cohort's enrollment was previously described in detail.^{13,14,18}

Troublesome lung symptom day

Troublesome lung symptoms were recorded by the parents in daily diaries from 1 month until 6 years of age. Troublesome lung symptoms were explained to the parents as wheeze or whistling sounds, breathlessness, or recurrent troublesome cough severely affecting the well-being of the infant. Emphasis was given to symptoms from the lower airways that were not just audible but actually affecting the sleep, activity, or well-being of the child. Daily symptoms were recorded as composite dichotomized scores (yes/no) on each day; ie, a global assessment was recorded. The complexity of symptoms was detailed in a book that was given to the parents and presented on the COPSAC Web site (http://www.copsac.com/userfiles/books/Asthma_in_young/index.html). The doctors at the research unit reviewed symptom definitions and the diary entries with the parents at every 6-month clinical session for 6 years.

Episodes of troublesome lung symptoms

Episode was defined from the diary cards as 3 consecutive days with troublesome lung symptoms, which was used as the base unit.

Recurrent episodes of troublesome lung symptoms

Recurrent episodes of troublesome lung symptoms were defined from the diaries as 5 episodes within 6 months or 4 weeks of consecutive symptoms.

Episodes with doctor-verified wheeze

During the first 3 years of the child's life, the families were instructed to attend the clinical research unit at every 3-day episode (as defined above) for a clinical evaluation as part of a randomized controlled trial.^{13,19} This included auscultation of the lungs by a medical doctor from the research unit recording wheeze, prolonged expiration, and/or rhonchi.

Current asthma by age 6 years

Asthma was diagnosed according to the international guidelines, as previously detailed,^{13,14} from a doctor's interviews of the parents at the

clinical research unit based on a predefined algorithm including all of the following: (1) recurrent episodes of troublesome lung symptoms (as defined above); (2) symptoms typical of asthma (recently proposed to be termed multi-trigger wheeze²⁰; eg, exercise-induced symptoms, prolonged nocturnal cough, recurrent cough outside the common cold, and symptoms causing waking at night); (3) intermittent rescue use of inhaled β_2 -agonist; and (4) response to a 3-month course of inhaled corticosteroids of 400 $\mu\text{g}/\text{d}$ and relapse when stopping treatment. Relapse was treated with a defined course of 6 months of 400 $\mu\text{g}/\text{d}$ inhaled corticosteroids and subsequent relapses by further 12-month courses only to be prolonged after documented relapse.

Current asthma by age 6 years was defined as a child receiving inhaled corticosteroid treatment for asthma for at least part of the sixth year of life on the basis of the described algorithm.

Genetic variants

The *ORMDL3* single nucleotide polymorphism rs7216389 was determined in the cohort, as previously reported.²¹ rs7216389 is a C/T polymorphism in which the T allele has consistently been associated with childhood asthma.

Baseline characteristics

Atopic heredity was assessed from interviews by the research doctors. Information on sociodemographics, prenatal exposures, gestational age, mode of delivery, Apgar scores, and parental age at birth of the child was collected by means of interview in the research clinic at enrollment. Information on breast-feeding, day care, presence of cats and dogs in the household, and presence of older siblings in the first year of life was collected by interview at 1, 6, and 12 months. Date of birth was grouped into 4 seasons: winter (December-February), spring (March-May), summer (June-August), and fall (September-November). Maternal education was accessed as the highest level of finished education when the child was 1 year of age.

Children's height and length were measured at the research clinic at 2 and 4 weeks and 6 and 12 months. Z scores were calculated from the World Health Organization standards, and birth weight z scores, birth length z scores, and birth body mass index z score were estimated from linear regression of z scores from 0 to 1 year.

Nicotine levels in hair at 1 year of age²² were analyzed as previously described.

Statistical analyses

Adjusting for missing data. We calculated an individual index of troublesome lung symptom load based on the available diary recordings as the ratio between the individual average and the cohort population average for a given age (days) by using a novel algorithm explained below. The index is standardized for age rather than date because the cohort covers a 3-year age span per calendar year (recruited 1998-2001) and because of the strong age dependency of symptoms (Fig 1).

Table E1 (available in this article's Online Repository at www.jacionline.org) exemplifies 2 subjects who both recorded the same total number of symptom days and were both missing half the diary data, but the average population symptom burden differs through the missing periods, with fewer symptoms in the population during the period recorded for subject 2 and more symptoms in the population during the period recorded for subject 1. Therefore subject 2 is more abnormal (ill) relative to the population, which could be calculated as shown in the table. Note that this adjustment also classifies both subjects as more ill than an average subject in the population that would correspond to a symptom load equal to 1. The index classifies each child relative to the concurrent average population symptom load at a given age, whereby 1 equals the average symptom burden in the population at similar age and values greater than or less than 1 equals a symptom burden of greater than or less than the population average at a similar age (ie, high or low symptom load). This index is unaffected by missing values for a given child.

Categories of troublesome lung symptoms. We calculated the individual symptom load index for each of the age intervals studied

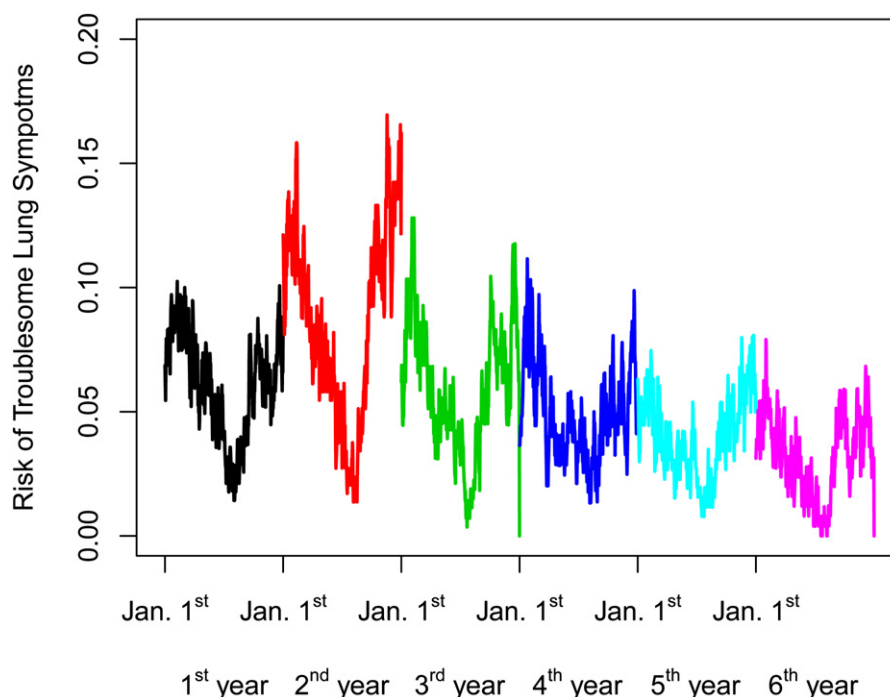


FIG 1. Day-to-day population frequency of troublesome lung symptoms normalized for calendar day.

(3-year, 2-year, and 1-year intervals, respectively). Categorization based on 2-year intervals (0-2 years, 2-4 years, and 4-6 years) results in 8 possible categories of temporal asthmatic pattern (see Table E2 in this article's Online Repository at www.jacionline.org). The expected frequencies in these 8 categories were calculated as the product of the observed frequency for each age group assuming independence between these age groups. These expected frequencies were compared with the observed frequencies (see Table E2). Also, we calculated an alternative categorization based solely on symptoms recorded during the summer (June, July, and August).

The association between the diagnosis of current asthma at 6 years of age, the *ORMDL3* gene variants, and the categories of troublesome lung symptoms was evaluated by means of logistic regression. Odds ratio (OR) estimates are reported together with the 95% Wald CIs calculated on a log scale and back transformed. Also reported is the Wald test *P* value for $\log(\text{OR})$ equal to zero.

Quantitative assessments of troublesome lung symptoms. The risk of development of troublesome lung symptoms from the *ORMDL3* gene variant was assessed by age-at-onset analyses with Kaplan-Meier curves and Cox regression. Hazard ratio (HR) estimates are reported and accompanied by 95% Wald CI calculated on a log scale and back transformed. Also reported is the Wald test *P* value for $\log(\text{HR})$ equal to zero.

The number of events (symptom days, number of episodes, and number of episodes with doctor-verified wheeze) during a half-year period of age were analyzed by means of Poisson regression with log link and log number of registration days as offset and adjusting for age in half-year intervals expressed as intensity ratios (IRs).

Point prevalence's of recurrent episodes of troublesome lung symptoms were analyzed by using a log linear model adjusting for age in half-year intervals.

Within-child correlation is accounted for by independent working generalized estimating equations. *P* values correspond to robust Wald tests. Robust 95% Wald CIs for the prevalence ratios (PRs) are calculated on a log scale and back transformed.

All *P* values were evaluated at a 5% significance level. Analyses were performed with SAS version 9.1.3 software (SAS Institute, Inc, Cary, NC) and the open source statistical programming environment R version 2.7.1.

RESULTS

Daily diary recordings until age 6 years were available in 307 (157 girls) of the COPSAC birth cohort (75% of the birth cohort). The overall registration frequency over the 6-year period was 80.2%. The overall symptom frequency was 3.5% disregarding missing registrations and 4.3% among actual registration days.

Table I describes the baseline characteristics of the COPSAC birth cohort and compares the study group with the group lost for this follow-up. Subjects in the study group were more likely to have a father with asthma or allergic rhinitis and to come from families with higher incomes, had parents with a higher mean age, and were more likely to have a dog in the household. There were no significant differences with respect to other risk factors, such as male sex, *ORMDL3* genotype, or tobacco smoke exposure.

Days with troublesome lung symptoms observed during the first 6 years of life

Fig 1 shows the average symptom load during the first 6 years of life in the birth cohort normalized for calendar day. We see a monotonous seasonal variation of low symptom levels during the summer months, a rapid increase in August-September, and a high level during winter, tapering off slowly toward the following summer. The winter symptom burden peaked during the first 2 winters and decreased thereafter. Symptoms during the summer showed less age dependency.

Sensitivity analysis of categorical temporal patterns based on 1-, 2-, or 3-year periods

Categorization based on 3-year periods (0-3 years and 3-6 years of age) yielded 17% persistent troublesome lung symptoms

TABLE I. Baseline description in the full cohort and stratified according to dropout status

	Full cohort	Diary, 0-6 y	Lost to follow-up	
No. of patients	411	307	104	—
Reasons for dropout (n)				
Death			2	
Disabling disease			3	
Social (removed from parents)			1	
Emotional distress to study procedures			7	
Emigration			11	
Maternal disease			1	
Parental lack of time			6	
No specified reason			73	
Genetics				
Male sex (%)	49	49	51	NS*
White race (%)	97	97	95	NS*
<i>ORMDL3</i>				
CT (%)	48	48	50	NS*
TT (%)	29	28	30	
Atopic heredity				
Mother with asthma (%)	100	100	100	—
Mother with allergic rhinitis (%)	76	77	73	NS*
Mother with eczema (%)	48	49	45	NS*
Father with asthma (%)	17	20	7	.003*
Father with allergic rhinitis (%)	33	37	20	.002*
Father with eczema (%)	13	14	11	NS*
Sociodemographics				
Household income, mean (SD), 1,000 DKK	494 (207)	511 (205)	430 (202)	.002†
Mother's education				
Low (%)	60	59	61	NS*
Medium (%)	27	27	24	
High (%)	14	13	15	
Mother's occupation				
Student (%)	11	10	14	NS*
Unemployed (%)	10	8	17	
Nonprofessional (%)	34	34	32	
Professional (%)	46	48	36	
Urban living (%)	93	93	92	NS*
Father's age (y), mean (SD)	32.0 (5.2)	32.5 (5.1)	30.6 (5.2)	.001†
Mother's age (y), mean (SD)	30.0 (4.5)	30.4 (4.3)	29.0 (5.1)	.007†
Birth				
Birth weight (z score), mean (SD)	0.06 (1.12)	0.09 (1.17)	−0.04 (0.96)	NS†
Birth length (z score), mean (SD)	0.11 (1.15)	0.13 (1.20)	0.06 (0.97)	NS†
Birth BMI (z score), mean (SD)	0.00 (1.02)	0.03 (1.05)	−0.10 (0.94)	NS†
Gestational age (wk), mean (SD)	39.9 (1.6)	39.9 (1.6)	39.8 (81.4)	NS†
Cesarean section (%)	21	22	16	NS*
Apgar score >9 (%)	97	96	98	NS*
Season of birth				
Winter (%)	23	23	24	NS*
Spring (%)	21	20	25	
Summer (%)	27	29	22	
Fall (%)	29	29	29	
Prenatal exposures				
Mother smoking in 3rd trimester (%)	15	14	19	NS*
Antibiotic use in 3rd trimester (%)	14	15	12	NS
Postnatal exposures				
Solely breast-fed > 4 wk (%)	82	83	76	NS*
Day care, age at start, median (IQR)	346 (246-433)	343 (246-429)	359 (246-463)	NS‡
Cat in household 1st y (%)	14	13	14	NS*
Dog in household 1st y (%)	13	15	4	.02*
Older siblings				
0 (%)	61	60	65	NS*
1 (%)	28	30	20	
Nicotine in hair at 1 y, ng/mg, median (IQR)	0.74 (0.3-2.5)	0.70 (0.3-2.3)	0.95 (0.4-4.5)	NS‡

BMI, Body mass index; DKK, Danish kroner; NS, not significant ($P > .05$).* χ^2 Test.†Unpaired t test.

‡Wilcoxon rank sum test.

TABLE II. Categorical asthma symptom categories in young children based on temporal patterns of 2- or 3-year periods

	With or without troublesome lung symptoms all year, age 0-3 and 3-6 y	With or without troublesome lung symptoms all year, age 0-2, 2-4, and 4-6 y	With or without troublesome lung symptoms during summer, age 0-3 and 3-6 y
Asymptomatic	54%	51%	62%
Persistent	17%	12%	12%
Late onset	11%	14%	9%
Transient	18%	23%	17%
Sensitivity of persistent and late-onset categories for current asthma by age 6 y (95% CI)	0.97 (0.84-1.00)	0.95 (0.81-0.99)	0.77 (0.61-0.88)
Specificity of persistent and late-onset category for current asthma by age 6 y (95% CI)	0.83 (0.78-0.87)	0.85 (0.80-0.89)	0.88 (0.83-0.91)

TABLE III. Comparison of principles of categorizing wheezy symptom patterns in young children (3-year intervals) for genetic risk analyses (*ORMDL3* gene variant rs7216389, risk genotype TT)

	Transient early	Persistent	Late onset
<i>ORMDL3</i>			
All year (CI)	1.86 (0.97-3.58), <i>P</i> = .06	1.65 (0.83-3.25), <i>P</i> = .15	0.43 (0.14-1.31), <i>P</i> = .14
Summer symptoms (CI)	2.04 (1.06-3.95), <i>P</i> = .03	1.90 (0.88-4.10), <i>P</i> = .10	0.77 (0.27-2.16), <i>P</i> = .62

The asymptomatic category is used as the reference category.

(high-high symptom load), 18% transient early troublesome lung symptoms (high-low symptom load), 11% late-onset troublesome lung symptoms (low-high symptom load), and 54% asymptomatic children (low-low symptom load). Thirty-eight of 39 children with asthma by age 6 years were categorized as persistent or late-onset troublesome lung symptoms (sensitivity for current asthma by age 6 years, 0.97; [95% CI, 0.84-1.00]), whereas 46 of 266 nonasthmatic subjects were categorized as persistent or late-onset (specificity for current asthma by age 6 years, 0.83; CI, 0.78-0.87; Table II).

Refining the age intervals to 2-year periods resulted in 8 categorical patterns. Expected and observed symptom frequencies for these patterns are presented in Table E2. Symptom patterns between age groups were significantly interdependent ($P < .0001$, χ^2 test). Patterns of low-high-low (3%) and high-low-high (1%) were more rare than expected (14% and 6%, respectively; $P < .0001$, 1-sided). By ignoring these 2 subsets, we could confirm a general pattern similar to the classical 4 categories. Persistent troublesome lung symptoms (high-high-high) was more commonly observed than expected and constituted 12% of our birth cohort. Likewise, healthy children (low-low-low) were more common than expected and classified 51% of the children. Transient early (high-low-low and high-high-low) and late-onset (low-low-high and low-high-high) troublesome lung symptoms classified 23% and 14% of children, which was less common than expected. There was a high level of agreement between this 2-year classification and the 3-year classification (simple κ coefficient = 0.84 [0.79-0.89]). The sensitivity of this categorization of persistent and late onset in relation to 6-year current asthma was 0.95 (95% CI, 0.81-0.99), and the specificity was 0.85 (95% CI, 0.80-0.89; Table II). This was very similar to the 3-year categorization. We therefore chose the 3-year categorization for future analyses because it is always possible to classify a child using this method.

When further reducing the intervals to 1-year symptom periods, 96 children could not be classified in the 4 categories described above, and 40 children changed category. The 96

children who could not be classified exhibited what graphically seemed a random pattern from which no new archetype was apparent. This inverse relation between finer age grouping and the number of children classified into archetypes is mainly induced by instabilities in the index of symptom load as age grouping narrows. The sensitivity of this categorization of persistent and late onset in relation to 6-year current asthma was 0.97 (CI, 0.80-1.00), and the specificity was 0.90 (CI, 0.85-0.94). However, this was at the expense of a large group of children who could not be meaningfully categorized, and it was therefore not considered a useful categorization.

Classification based on summer symptoms in 3-year intervals yielded 12% with persistent troublesome lungs symptoms during summer 9% with late-onset summer symptoms 17% with transient early summer symptoms and 62% asymptomatic children. The use of summer symptoms for categorizing children into the groups of persistent and late-onset troublesome lung symptoms increased the specificity to 0.88 (95% CI, 0.83-0.91) for a diagnosis of current asthma by age 6 years, whereas the sensitivity decreased to 0.77 (95% CI, 0.61-0.88).

Capturing genetic risk by means of categorical temporal patterns

ORMDL3 showed a nonsignificant trend as a risk factor for transient early troublesome lung symptoms (OR, 1.86; 95% CI, 0.97-3.58; P = .06), whereas there was no significant association between this variant and persistent troublesome lung symptoms (OR, 1.65; 95% CI, 0.83-3.25; P = .15) or late-onset troublesome lung symptoms (OR, 0.43; 95% CI, 0.14-1.31; P = .14) by using the asymptomatic category as a reference (Table III).

Categories defined solely from symptoms during summer showed *ORMDL3* to be a significant risk factor for transient early troublesome lung symptoms (OR, 2.04; 95% CI, 1.06-3.95; P = .03), whereas there was no significant association between this variant and persistent (OR, 1.90; 95% CI, 0.88-4.10; P = .10) or late-onset troublesome lung symptoms (OR, 0.77; 95% CI, 0.27-2.16; P = .62) (Table III).

TABLE IV. Capturing genetic risk by using age-at-onset and quantitative assessment of troublesome lung symptoms

Symptom definition	Age-at-onset			Symptom burden 0-6 y		
	HR	CI	P value	PR/IR	CI	P value
Troublesome lung symptom-days	1.25	0.99–1.58	.06	1.41	1.08–1.83	.01
Troublesome lung symptom episodes	1.37	1.08–1.73	.01	1.42	1.07–1.90	.02
Recurrent troublesome lung symptom episodes	1.61	1.03–2.53	.04	2.42	1.36–4.33	.003
Episodes with wheeze at auscultation	1.27	0.90–1.80	.17	1.26	0.81–1.95	.31

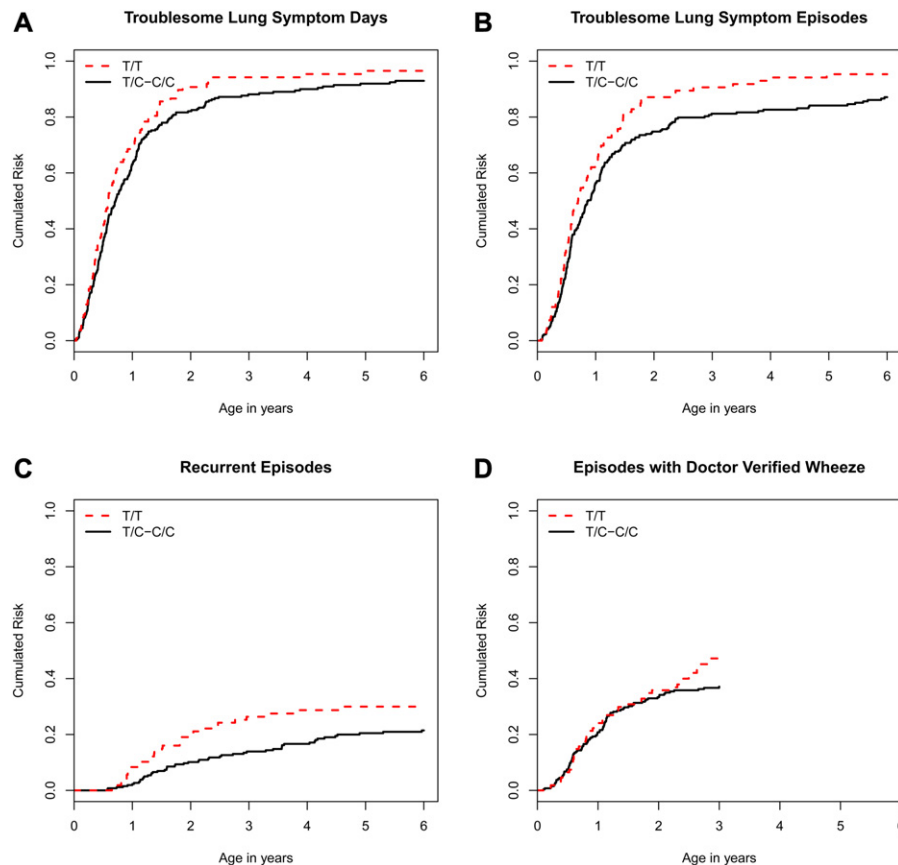


FIG 2. Symptom development is illustrated as time to event by using the Kaplan-Meier curves (inverted) and estimated hazard rates of troublesome lung symptoms in early life from the *ORMDL3* risk variant, exploring different quantitative assessments of symptom days. **A**, Troublesome lung symptom days recorded in daily diary cards (HR, 1.25; 95% CI, 0.99–1.58; $P = .06$). **B**, Episodes of 3 consecutive days with troublesome lung symptoms recorded in daily diary cards (HR, 1.37; 95% CI, 1.08–1.73; $P = .01$). **C**, Recurrent episodes of troublesome lung symptoms. Algorithm based on 5 recurrent episodes in 6 months (HR, 1.61; 95% CI, 1.03–2.53; $P = .04$). **D**, Doctor-verified wheeze during acute episodes (HR, 1.273; 95% CI, 0.90–1.80; $P = .17$).

Capturing genetic risk by quantitative and age-at-onset assessments of troublesome lung symptoms

Age-at-onset (Fig 2, Table IV). *ORMDL3* was a significant risk factor for the development of episodes of troublesome lung symptoms (HR, 1.37; 95% CI, 1.08–1.73; $P = .01$) and recurrent episodes of troublesome lung symptoms (HR, 1.61; 95% CI, 1.03–2.53; $P = .04$), as previously reported,²¹ and with a borderline trend for symptom days (HR, 1.25; 95% CI, 0.99–1.58; $P = .06$). In contrast, there was no increased risk from *ORMDL3* on episodes with doctor-verified wheeze (HR, 1.27; 95% CI, 0.90–1.80; $P = .17$) (Table IV).

IR and PR (Fig 3, Table IV). *ORMDL3* was a significant risk factor for accumulated symptom days (IR, 1.41; CI, 1.08–1.83; $P = .01$), episodes (IR, 1.42; CI, 1.07–1.90; $P = .02$), and recurrent

episodes of troublesome lung symptoms (PR, 2.42; CI, 1.36–4.33; $P = .003$) but not for episodes with doctor-verified wheeze (IR, 1.26; CI, 0.81–1.95; $P = .31$). The 4 different symptom assessments are compared in Fig 3 as the absolute 6-month rates for the *ORMDL3* T/T versus T/C or C/C variants.

DISCUSSION

Key findings

Day-to-day recordings of globally assessed troublesome lung symptoms during the first 6 years of life documented a general pattern of low symptom levels during the summer months, rapidly increasing levels in August–September to high levels during the winter, and tapering off slowly toward the summer (northern hemisphere). The winter symptom burden peaked during the first

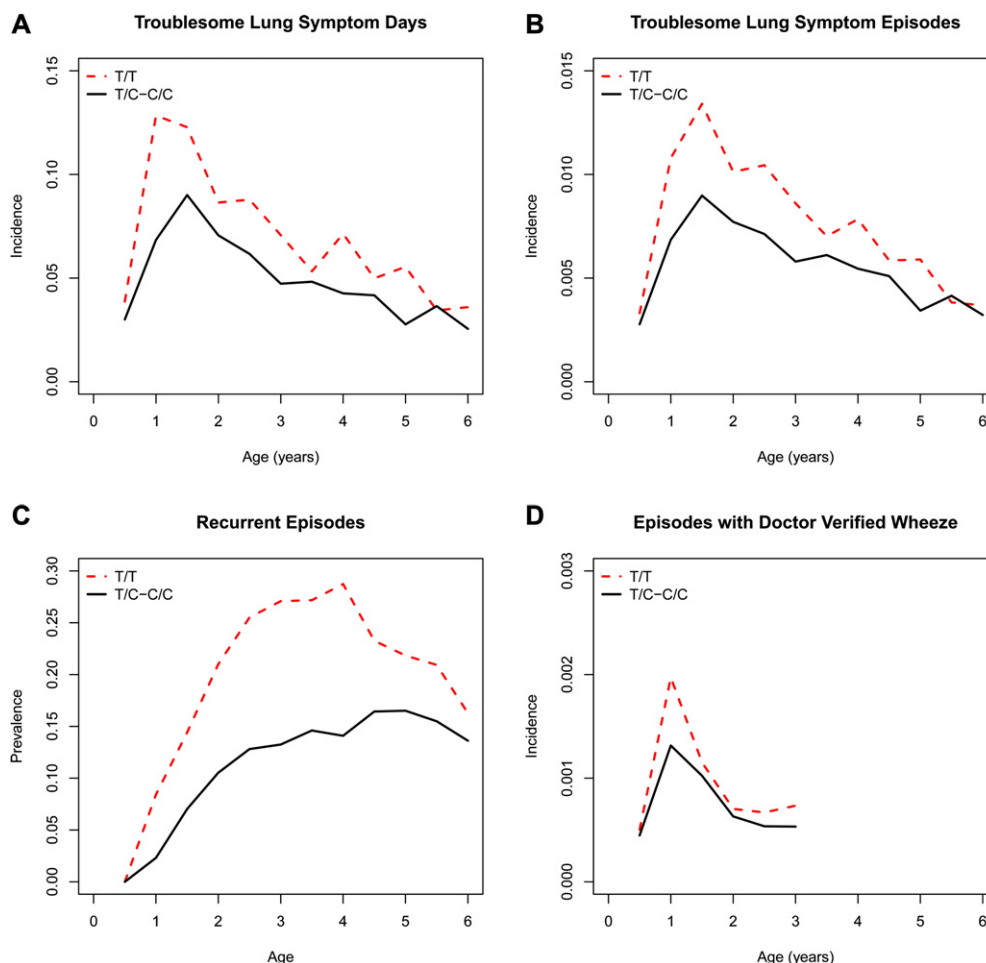


FIG 3. Intensity and prevalence rates of troublesome lung symptoms in early life from the *ORMDL3* risk variant, exploring different quantitative assessments of troublesome lung symptoms days. This suggests that the algorithm-based “recurrent episodes of troublesome lung symptoms” is the more powerful end point assessing asthma-related disease in early life. **A**, Troublesome lung symptom days recorded in daily diary cards (IR, 1.410; CI, 1.084-1.834; $P = .01$). **B**, Troublesome lung symptoms episodes of 3 consecutive days with symptoms recorded in daily diary cards (IR, 1.420; CI, 1.065-1.895; $P = .02$). **C**, Recurrent episodes of troublesome lung symptoms: algorithm based on 5 recurrent episodes in 6 months (PR, 2.423; CI, 1.356-4.330; $P = .003$). **D**, Doctor-verified wheeze during acute episodes (IR, 1.255; CI, 0.810-1.946; $P = .3092$).

2 winter seasons with a general decrease thereafter, whereas symptoms during the summer exhibited a more stable pattern.

The diary data confirmed the archetypal temporal patterns of transient early, persistent, and late-onset troublesome lung defined from the 3-year supervised classification, showing a higher frequency of these categories than expected by chance. Finer age grouping in 2- or 1-year periods did not provide more useful or meaningful categorization of children.

Quantitative assessment of daily symptoms provided a versatile and powerful tool for the description and estimation of the burden of troublesome lung symptoms, allowing calculation of risk ratios from the symptom burden over a given age span and HRs from age-at-onset information. In particular, a prospective, algorithm-based definition of recurrent episodes of troublesome lung symptoms provided the most powerful assessment of asthma burden in early life by using a known genetic asthma variant as the gold standard and was superior to the qualitative assessment of doctor-verified wheeze, as well as any temporal segmentation. The genetic risk variant reflects a molecular

mechanism; that is, our proposed quantitative global assessment of troublesome lung symptoms exhibited the best association with such an endotype.

Strengths of the study

The key strength of this study is that troublesome lung symptoms were monitored prospectively from daily diary cards during the first 6 years of life. This reduces the risk of recall bias, likely skewing information when a symptom history is obtained retrospectively.

Compliance was acceptable with complete 6-year data in 75% of the birth cohort and 80% overall registration.

The symptom definitions were preplanned in the study protocol.^{13,14,18,19} We defined an episode as 3 consecutive days of troublesome lung symptoms, which was used as the base unit. Recurrent troublesome lung symptoms and asthma were prospectively diagnosed based on an algorithm including 5 such episodes within a 6-month period. This protocol was used from 1 month of

age including a randomized controlled trial of antiasthma agents during the first 3 years of life¹³ and longitudinal observational studies of both environmental exposure effects¹⁴ and genetic effects.²¹

It is a strength that the data emphasized symptom frequency and age at onset by estimating a quantitative index instead of relying on parental interpretations of past history.

It is a strength that we applied a global assessment of troublesome lung symptoms instead of attempting a qualitative assessment of particular symptoms by using specific terms, such as wheeze, with a high risk of misclassification. The validity of the global symptom assessments was supported by 6-month reviews by the research doctor together with the parents and supported by comprehensive written material detailing the typical symptoms associated with asthma. Furthermore, the validity was strengthened from the mothers' own experiences of asthma.

The specific term wheeze carries a large risk of misclassification because of its inaccurate use by clinicians and caregivers.⁷⁻¹² Wheeze as a term has little cross-cultural validity. Limiting asthma symptoms to those reporting wheeze builds on a long tradition but one with no studies documenting the specificity or sensitivity of this term. Randomized controlled trials have shown that control of asthma symptoms is more closely associated with composite scores of associated lung symptoms than with wheeze or any other particular symptom.²³⁻²⁶ A recent study showed that cough was correlated with lung function and atopy at preschool age similarly to and independent of wheeze.²⁷ Studies in older children confirm that wheeze is not audible by using a stethoscope until a late stage of airway obstruction.²⁸ In the current study objective assessment of wheeze by the doctors at the research unit could not separate children at risk from their known genetic variant (Figs 2 and 3), confirming the limited value of wheeze as a symptom. This evidence supports our preplanned strategy to emphasize a global description of asthma associated symptoms that were severely troubling the child rather than using any specific terminology.

Furthermore, it is a strength of the analysis that we have developed an algorithm accounting for missing data. Long-term recordings of daily symptoms inevitably contain missing observations. Often missing data are ignored¹⁵ or the analysis is restricted to complete cases, seriously reducing the study base¹⁶ while ignoring the bias this incurs. Others have imputed missing values from neighboring data from the same subject, which can also cause bias because data are rarely missing at random. Intuitively, subjects who are symptomatic during periods when the average population is without symptoms (eg, the summer months) might be considered more abnormal or ill than subjects who are symptomatic during periods when the average population is also symptomatic (eg, the winter months). For the same reason, missing data weigh differently depending on the average population symptom burden during the missing period.

It is a strength of this study that we used a well-established genetic asthma risk variant as a gold standard. Otherwise, validation of diagnostic tools for asthma in early childhood is difficult because of the lack of consensus on asthma diagnosis at preschool age and the large variation in diagnostic traditions between physicians and countries. However, the genetic variation at the 17q21 locus (*ORMDL3*) is the strongest and best replicated genetic risk variant for early childhood asthma and has been replicated in multiple populations of both white and nonwhite origin.^{17,21,29} We believe that such a genetic marker is the most valid reference in the attempt to

establish the most sensitive measure of asthma symptoms. The genetic asthma variant reflects an underlying molecular mechanism and hence defines an asthma endotype.

Limitations of the study

It is a limitation of the study that we emphasized classical statistical methods to explore the categorical temporal disease patterns.¹⁵ Unsupervised approaches, such as principal component analysis or clustering techniques, can be used to uncover novel phenotypes, as previously reported.^{16,27}

It is a limitation to the external validity that the mothers all have a history of asthma. As such, the classification is internal to our population, and the prevalence of the categories cannot be generalized. However, the proposed principle of assessing troublesome lung symptoms symptoms in early life is unaffected by this.

Finally, it is a limitation to the generalizability of our study that this is a first proof of concept. Validation of this approach is needed in other cohorts.

Interpretation

The 6-year daily recordings demonstrate for the first time the complete natural history of troublesome lung symptoms in early childhood. The dramatic acceleration of symptoms in August-September mimics what has been reported for asthma-related hospitalization in schoolchildren.³⁰ This has been interpreted as a reflection of the accelerating spread of infections when children return to institutions after summer vacation.

The temporal patterns of troublesome lung symptoms confirmed the archetypes of transient early, late-onset, and persistent symptoms.¹⁵ These temporal categories have often been used to categorize asthmatic disorders in young children and have been useful in research into the origins of asthma. Temporal patterns might reflect different underlying endotypes, as demonstrated by differential associations with risk factors.^{15,16} However, these categories were supervised because they were defined from the presence of symptoms at 2 time points (the first 3 years of life and the sixth year of life). Any disease will fall into one of 4 categories because these are the only possible outcomes from a dichotomized end point (yes/no) at 2 time points; that is, the categories basically describe children never having symptoms, growing out of symptoms, growing into symptoms, or continuing to have symptoms. Still, our sensitivity analyses refining resolution of symptom periods into finer age groupings confirmed this simple temporal classification. Another birth cohort analyzed complete cases (approximately half of the cohort) by using an approach that was similar but based on 7 different time points during the first 6 years of life and a cluster (latent class) analysis approach proposing 6 different phenotypes.¹⁶ Future studies applying different approaches in the same group of children are needed to determine to which extent the identified categories are identical, overlapping, or complementary.

The relevance of temporal categories for describing troublesome lung symptom burden might be improved by emphasizing symptoms during the summer. Considering the significant seasonality of symptoms (Fig 1), it is plausible that recurrent episodes during the summer provide different information than recurrent episodes during the winter. Troublesome lung symptoms during the summer probably occur less often because of infection and might therefore provide a better measure of asthma endotypes. Accordingly, this approach reduced the sensitivity but improved the specificity for current asthma by age 6 years and suggested a stronger association with known genetic risk variants for early asthma.

We propose that quantitating symptom frequency and longitudinal analyses of age-at-onset improve the statistical power over that of a cross-sectional analysis. The prospective nature of the end point definition allows age-at-onset analyses of symptom development and longitudinal assessment of symptom frequency and disease prevalence. Complex human diseases have variable ages of onset, and the subject's age of onset carries extra information about the cause of the disease. Heritability penetrates most strongly early in life, whereas environmental effects play increasing roles later in life.³¹ This assessment of symptom frequency and age at onset provided a better risk assessment of the genetic asthma variant *ORMDL3* than segmenting children based on temporal categories (transient, persistent, and late onset). Genetic variants reflect a molecular mechanism and define asthma endotypes.³² This suggests that quantitative longitudinal assessment of troublesome lung symptoms is a better tool for identifying asthma endotypes in early life than temporal symptom patterns.

Our quantitative assessment of any troublesome lung symptoms also showed improved statistical power for detection of the underlying genetic susceptibility compared with doctor verified assessment of wheeze. Our approach might therefore also be a better measure of treatment response in clinical practice and clinical trials of antiasthma medication.

In conclusion, asthmatic symptoms in young children represent a heterogeneous group of diseases. The options for objective assessments in young children are limited, and symptom history remains the main end-point in clinical practice, as well as research. We propose a novel phenotyping tool based on quantitative global assessment of troublesome lung symptoms: emphasizing the recurrence of episodes of troublesome lung symptoms rather than the presence of any specific lung symptom to avoid symptom misclassification. Our proposed algorithm is based on a predefined number of days and episodes of troublesome lung symptoms. This quantitative approach to disease categorization instead of qualitative characteristics or temporal patterns showed a stronger relation to an underlying molecular mechanism exemplified by the *ORMDL3* genetic variant and therefore a closer relation to underlying endotypes. This approach might be helpful in the refinement of outcomes for research into the origins and treatment of early asthma, clinical practice, and clinical trials in young children.

We thank the children and parents participating in the COPSAC cohorts, as well as the COPSAC study team.

Key messages

- Temporal patterns of early transient, persistent, and late-onset troublesome lung symptoms from 3-year periods were confirmed, with no benefit from a finer temporal categorization.
- We propose a novel method for endotyping of early childhood asthma by the frequency and age-of-onset of globally assessed troublesome lung symptoms analyzed longitudinally.
- This novel approach showed stronger association with a validated genetic risk variant than traditional temporal patterns, whereas the study power was lost by restricting to doctor-verified wheeze.

REFERENCES

1. Kocevar VS, Bisgaard H, Jonsson L, Valovirta E, Kristensen F, Yin DD, et al. Variations in pediatric asthma hospitalization rates and costs between and within Nordic countries. *Chest* 2004;125:1680-4.
2. Weiss KB, Sullivan SD, Lyttle CS. Trends in the cost of illness for asthma in the United States, 1985-1994. *J Allergy Clin Immunol* 2000;106:493-9.
3. Warren P. Asthma as a disease concept. *Lancet* 2006;368:1416.
4. Bush A. Practice imperfect—treatment for wheezing in preschoolers. *N Engl J Med* 2009;360:409-10.
5. Anderson GP. Endotyping asthma: new insights into key pathogenic mechanisms in a complex, heterogeneous disease. *Lancet* 2008;372:1107-19.
6. Wenzel SE. Asthma: defining of the persistent adult phenotypes. *Lancet* 2006;368:804-13.
7. Cane RS, Ranganathan SC, McKenzie SA. What do parents of wheezy children understand by “wheeze”? *Arch Dis Child* 2000;82:327-32.
8. Cane RS, McKenzie SA. Parents' interpretations of children's respiratory symptoms on video. *Arch Dis Child* 2001;84:31-4.
9. Crane J, Mallol J, Beasley R, Stewart A, Asher MI. Agreement between written and video questions for comparing asthma symptoms in ISAAC. *Eur Respir J* 2003;21:455-61.
10. Van Sickle D. Perceptions of asthma among physicians: an exploratory study with the ISAAC video. *Eur Respir J* 2005;26:829-34.
11. Michel G, Silverman M, Strippoli MP, Zwahlen M, Brooke AM, Grigg J, et al. Parental understanding of wheeze and its impact on asthma prevalence estimates. *Eur Respir J* 2006;28:1124-30.
12. Mohango AD, de Koning HJ, Hafkamp-de GE, van der Wouden JC, Jaddoe VW, Moll HA, et al. A comparison of parent-reported wheezing or shortness of breath among infants as assessed by questionnaire and physician-interview: the Generation R study. *Pediatr Pulmonol* 2010;45:500-7.
13. Bisgaard H, Hermansen MN, Loland L, Halkjaer LB, Buchvald F. Intermittent inhaled corticosteroids in infants with episodic wheezing. *N Engl J Med* 2006;354:1998-2005.
14. Bisgaard H, Hermansen MN, Buchvald F, Loland L, Halkjaer LB, Bonnelykke K, et al. Childhood asthma after bacterial colonization of the airway in neonates. *N Engl J Med* 2007;357:1487-95.
15. Martinez FD, Wright AL, Taussig LM, Holberg CJ, Halonen M, Morgan WJ. Asthma and wheezing in the first six years of life. The Group Health Medical Associates. *N Engl J Med* 1995;332:133-8.
16. 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.
17. Moffatt MF, Kabesch M, Liang L, Dixon AL, Strachan D, Heath S, et al. Genetic variants regulating *ORMDL3* expression contribute to the risk of childhood asthma. *Nature* 2007;448:470-3.
18. Bisgaard H. The Copenhagen Prospective Study on Asthma in Childhood (COPSAC): design, rationale, and baseline data from a longitudinal birth cohort study. *Ann Allergy Asthma Immunol* 2004;93:381-9.
19. Bisgaard H, Hermansen MN, Bonnelykke K, Stokholm J, Baty F, Skytt NL, et al. Association of bacteria and viruses with wheezy episodes in young children: prospective birth cohort study. *BMJ* 2010;341:c4978.
20. Brand PL, Baraldi E, Bisgaard H, Boner AL, Castro-Rodriguez JA, Custovic A, et al. Definition, assessment and treatment of wheezing disorders in preschool children: an evidence-based approach. *Eur Respir J* 2008;32:1096-110.
21. Bisgaard H, Bonnelykke K, Sleiman PM, Brasholt M, Chawes B, Kreiner-Moller E, et al. Chromosome 17q21 gene variants are associated with asthma and exacerbations but not atopy in early childhood. *Am J Respir Crit Care Med* 2009;179:179-85.
22. Sorensen M, Bisgaard H, Stage M, Loft S. Biomarkers of exposure to environmental tobacco smoke in infants. *Biomarkers* 2007;12:38-46.
23. Knorr B, Franchi LM, Bisgaard H, Vermeulen JH, LeSouef P, Santanello N, et al. Montelukast, a leukotriene receptor antagonist, for the treatment of persistent asthma in children aged 2 to 5 years. *Pediatrics* 2001;108:e48.
24. Bacharier LB, Phillips BR, Zeiger RS, Szefer SJ, Martinez FD, Lemanske RF Jr, et al. Episodic use of an inhaled corticosteroid or leukotriene receptor antagonist in preschool children with moderate-to-severe intermittent wheezing. *J Allergy Clin Immunol* 2008;122:1127-35.
25. Bisgaard H, Munck SL, Nielsen JP, Petersen W, Ohlsson SV. Inhaled budesonide for treatment of recurrent wheezing in early childhood. *Lancet* 1990;336:649-51.
26. Bisgaard H, Gillies J, Groenewald M, Maden C. The effect of inhaled fluticasone propionate in the treatment of young asthmatic children: a dose comparison study. *Am J Respir Crit Care Med* 1999;160:126-31.

27. Smith JA, Drake R, Simpson A, Woodcock A, Pickles A, Custovic A. Dimensions of respiratory symptoms in preschool children: population-based birth cohort study. *Am J Respir Crit Care Med* 2008;177:1358-63.
28. Meslier N, Charbonneau G, Racineux JL. Wheezes. *Eur Respir J* 1995;8:1942-8.
29. Galanter J, Choudhry S, Eng C, Nazario S, Rodriguez-Santana JR, Casal J, et al. ORMDL3 gene is associated with asthma in three ethnically diverse populations. *Am J Respir Crit Care Med* 2008;177:1194-200.
30. Sears MR, Johnston NW. Understanding the September asthma epidemic. *J Allergy Clin Immunol* 2007;120:526-9.
31. Thomsen SF, Duffy DL, Kyvik KO, Backer V. Genetic influence on the age at onset of asthma: a twin study. *J Allergy Clin Immunol* 2010;126:626-30.
32. Bisgaard H, Bonnelykke K. Long-term studies of the natural history of asthma in childhood. *J Allergy Clin Immunol* 2010;126:187-97.

TABLE E1. Adjusting diary symptom data for missing observations

Day	Population	Subject 1			Subject 2		
	Average symptom load	Diary record	Symptom	Population adjustment	Diary record	Symptom	Population adjustment
1	0.05	0	0	0×0.05	1	0	1×0.05
2	0.10	0	0	0×0.10	1	0	1×0.10
3	0.15	0	0	0×0.15	1	1	1×0.15
4	0.20	0	0	0×0.20	1	1	1×0.20
5	0.25	0	0	0×0.25	1	1	1×0.25
6	0.30	1	0	1×0.30	0	0	0×0.30
7	0.30	1	1	1×0.30	0	0	0×0.30
8	0.30	1	1	1×0.30	0	0	0×0.30
9	0.30	1	1	1×0.30	0	0	0×0.30
10	0.30	1	0	1×0.30	0	0	0×0.30
Total			3	1.5		3	0.75
Adjusted symptom load			$3/1.5 = 2$			$3/0.75 = 4$	

This example shows subjects 1 and 2, who recorded similar total numbers of symptom days and are missing half the diary data. However, the average population symptom burden differs through the missing periods, with fewer symptoms in the population during the period symptoms were recorded for subject 2 and more symptoms in the population during the period symptoms were recorded for subject 1. Hence subject 2 could be considered more abnormal (ill) relative to the population, which could be calculated as shown in the table. Note that this adjustment also classifies both subjects as more ill than an average subject in the population, which would correspond to a symptom load equal to 1.

TABLE E2. Observed pattern frequencies of dichotomized asthmatic symptom loads during the 0- to 2-, 2- to 4-, and 4- to 6-year intervals and expected frequencies if the asthmatic symptom loads were independent between these intervals

Pattern	Expected frequency	Observed frequency (n)	Revised categories	Frequency of revised categories
Low, low, low	0.34	0.49 (150)	Asymptomatic	51%
High, high, high	0.03	0.12 (36)	Persistent	12%
Low, low, high	0.12	0.08 (24)	Late onset	14%
Low, high, high	0.05	0.05 (16)	Late onset	
High, low, low	0.18	0.13 (39)	Transient early	23%
High, high, low	0.08	0.09 (28)	Transient early	
Low, high, low	0.14	0.03 (10)	Transient early	
High, low, high	0.06	0.01 (4)	Transient early	