

A simple asthma prediction tool for preschool children with wheeze or cough

Anina M. Pescatore, MSc,^a Cristian M. Dogaru, MD, PhD,^a Lutz Duembgen, PhD, MSc,^{a,b} Michael Silverman, MD,^c Erol A. Gaillard, MD,^c Ben D. Spycher, PhD,^{a,*} and Claudia E. Kuehni, MD, MSc^{a,*} *Bern, Switzerland, and Leicester, United Kingdom*

Background: Many preschool children have wheeze or cough, but only some have asthma later. Existing prediction tools are difficult to apply in clinical practice or exhibit methodological weaknesses.

Objective: We sought to develop a simple and robust tool for predicting asthma at school age in preschool children with wheeze or cough.

Methods: From a population-based cohort in Leicestershire, United Kingdom, we included 1- to 3-year-old subjects seeing a doctor for wheeze or cough and assessed the prevalence of asthma 5 years later. We considered only noninvasive predictors that are easy to assess in primary care: demographic and perinatal data, eczema, upper and lower respiratory tract symptoms, and family history of atopy. We developed a model using logistic regression, avoided overfitting with the least absolute shrinkage and selection operator penalty, and then simplified it to a practical tool. We performed internal validation and assessed its predictive performance using the scaled Brier score and the area under the receiver operating characteristic curve.

Results: Of 1226 symptomatic children with follow-up information, 345 (28%) had asthma 5 years later. The tool consists of 10 predictors yielding a total score between 0 and 15: sex, age, wheeze without colds, wheeze frequency, activity disturbance, shortness of breath, exercise-related and aeroallergen-related wheeze/cough, eczema, and parental history of asthma/bronchitis. The scaled Brier scores for the internally validated model and tool were 0.20 and 0.16, and the areas under the receiver operating characteristic curves were 0.76 and 0.74, respectively.

Conclusion: This tool represents a simple, low-cost, and noninvasive method to predict the risk of later asthma in symptomatic preschool children, which is ready to be tested in other populations. (*J Allergy Clin Immunol* 2013;■■■:■■■-■■■.)

Key words: Asthma, wheeze, cough, children, prediction, prognosis, persistence, longitudinal, cohort study

Many preschool children present to primary care with recurrent wheeze or cough. These symptoms are a burden to families and lead to treatment with inhalers, antibiotics, or cough mixtures; hospitalizations; and considerable health care costs.¹ In this age group wheezing illness is heterogeneous and includes different phenotypes with varying prognoses.²⁻⁵ Fortunately, only some children will have persistent problems until school age. The ability to predict the persistence of wheeze up to school age would allow preventative and therapeutic efforts to be directed to those most in need⁶ and would reassure parents of children with transient problems. It would also help to select children for intervention studies, which aim to alter the course of disease.⁷

Several groups have presented tools for prediction of later asthma in preschool children,⁸⁻¹⁶ but their use for primary care is limited.¹⁷ Some tools were developed in study populations untypical for primary care. For instance, they included asymptomatic children,^{8,10,14,16} children with mild symptoms who never visited the doctor,^{13,15} or only high-risk children hospitalized for bronchiolitis.¹² Several studies excluded children with chronic cough,^{13,15} who might actually have a variant of asthma.^{4,18} Some tools included predictors, such as parental education, that are not easily generalizable to other populations.⁹ Other tools involve invasive measurements (blood tests or skin prick tests) that might not be accepted by all families in primary care.^{8,11,13,14} Finally, the methods commonly used to develop the prediction tools are prone to overfitting the data.^{9,11,13} Overfitting leads to reduced performance when tools are applied to other populations.^{19,20}

In this study we aimed to develop a simple tool to predict asthma at school age in preschool children with wheeze or chronic cough. We designed the tool for application in clinical practice, particularly primary care, by (1) studying a population of symptomatic children who had presented to the doctor for wheeze or cough, (2) defining a clinically relevant outcome, (3) considering only predictive factors easily assessed during a single consultation (a detailed symptom history but no blood or skin prick tests and no repeated observations), and (4) developing a robust model that performs well in internal validation and relevant sensitivity analyses but does not overfit the data and is therefore likely to be transferable to other populations.

METHODS

Study population

We analyzed data from a population-based childhood cohort from Leicestershire, United Kingdom, that has been described in detail elsewhere.²¹⁻²³ In brief, we recruited a representative population-based sample of 6808 children of white and South Asian ethnic origin born in 1993-1997.

From ^athe Institute of Social and Preventive Medicine (ISPM) and ^bthe Institute of Mathematical Statistics and Actuarial Science, University of Bern, and ^cthe Department of Infection, Immunity & Inflammation, University of Leicester.

*These authors contributed equally to this work.

Supported by the Swiss National Science Foundation (PDFMP3-123162 and 3200B0-122341) and Asthma UK 07/048. B.D.S. is the recipient of a European Respiratory Society/Marie Curie Joint Research Fellowship (MC 1614-2010).

Disclosure of potential conflict of interest: C. E. Kuehni has received grants from the Swiss National Science Foundation (SNF). M. Silverman has received grants from Asthma UK. B. D. Spycher has received grants from the European Respiratory Society/Marie Curie Actions Joint Research Fellowship. The rest of the authors declare that they have no relevant conflicts of interest.

Received for publication October 12, 2012; revised May 24, 2013; accepted for publication June 3, 2013.

Corresponding author: Claudia E. Kuehni, MD, MSc, Institute of Social and Preventive Medicine, Finkenhubelweg 11, CH-3012 Bern, Switzerland. E-mail: kuehni@ispm.unibe.ch.

0091-6749/\$36.00

© 2013 American Academy of Allergy, Asthma & Immunology

<http://dx.doi.org/10.1016/j.jaci.2013.06.002>

Abbreviations used

API: Asthma Predictive Index
 AUC: Area under the receiver operating characteristic curve
 LASSO: Least absolute shrinkage and selection operator

Perinatal data were collected at birth, and data on growth and development were acquired prospectively during childhood. Upper and lower respiratory morbidity, treatments and health care use, family history of atopic disease, and individual and family-related exposures were assessed by using repeated questionnaires (1998, 1999, 2001, 2003, 2006, and 2010). The study was approved by the Leicestershire Health Authority Research Ethics Committee.

Presentation at baseline (inclusion criteria)

Our analysis included all cohort children aged 1 to 3 years at baseline with parent-reported wheeze or chronic cough (cough without colds or cough at night) with 1 or more visits to the doctor for wheeze or cough during the past 12 months (Fig 1, highlighted in gray). The original questions are provided in Fig E1 in the *Methods* section in this article's Online Repository at www.jacionline.org. We included chronic cough because some children with chronic cough might have a variant of asthma or be at risk for asthma later in life.^{4,18} Information on symptoms at baseline was taken from the 1998 or 1999 questionnaire, favoring the questionnaire when children were closest to age 2.0 years.

Any asthma at school age (definition of outcome)

We defined a clinically relevant outcome as the combination of current wheeze plus use of asthma medication during the past 12 months at the age of 6 to 8 years (ie, 5 years later; see Fig E2 in this article's Online Repository for original questions). Asthma medication included short- or long-acting β_2 -agonists, inhaled corticosteroids, leukotriene receptor antagonists, or oral corticosteroids.

We used the Fisher exact test to compare characteristics of children with and without the outcome (Table 1 and see Table E1 in this article's Online Repository at www.jacionline.org), as well as to compare characteristics of children by availability of follow-up information (see Table E2 in this article's Online Repository at www.jacionline.org).

Choice of potential predictive factors

We used the following approach to compile the list of potential predictors. First, we reviewed the literature to identify relevant risk factors for the incidence or persistence of childhood asthma.^{3,24-31} From these, we only selected factors that are readily available in primary care and do not require repeated observations or additional investigations, such as blood tests or skin prick tests. The final list contained 24 potential predictors (see Table E1): demographic and perinatal data; eczema; upper and lower respiratory tract symptoms, particularly those reflecting triggers and severity of wheeze; and parental history of wheeze, asthma, bronchitis, or hay fever (see Fig E3 in this article's Online Repository for original questions). We did not include environmental or socioeconomic information because their prevalence and interpretation are likely to vary between populations, and thus their inclusion might reduce the generalizability of the tool.

Model development

We used least absolute shrinkage and selection operator (LASSO)-penalized logistic regression to develop the prediction model.^{32,33}

This approach allows us to identify important predictors and to estimate their influence on later asthma without overfitting the data. Traditional methods used for selecting predictors, such as stepwise backward or forward selection, tend to overfit the data, resulting in models that predict outcomes in the current dataset well but become unreliable in other datasets.²⁰ For our

analysis, we recoded all potential predictors with more than 2 response categories into multiple binary variables. Thus 38 binary variables derived from the 24 questions entered the variable selection process (see the *Methods* section in this article's Online Repository at www.jacionline.org for details). LASSO regression selects predictors in the order of their predictive importance. The final prediction model allows calculation of a prediction score and the probability of later asthma for each child.

Model performance

We assessed our prediction model in terms of overall performance, discrimination, and calibration. To assess overall performance, we calculated the scaled Brier score,²⁰ a measure of the discrepancy between the predicted probability and the actual outcome. A scaled Brier score with a value of zero means that the model does not predict later asthma in a subject better than if it had been informed only by the average prevalence of asthma at school age; the maximal value of 1 indicates perfect prediction. To determine the discriminative ability of the model (ie, its ability to distinguish between children with and without later asthma), we plotted the receiver operating characteristic curve and calculated the area under the receiver operating characteristic curve (AUC), which is also known as the c-statistic.^{20,34} The AUC can take on values from 0 to 1, with 1 being a perfectly discriminating model. Discrimination is considered not better than chance if the AUC is 0.5, moderate if the AUC is 0.6 to 0.8, and good if the AUC is greater than 0.8.³⁴ The calibration of the model (how well the predicted probabilities agree with the prevalence of the outcome in subgroups of children) was tested by using the Hosmer-Lemeshow goodness-of-fit-test^{20,35} and visualized with a calibration plot.²⁰ A Hosmer-Lemeshow goodness-of-fit-test result of less than 0.05 indicates that the predicted probabilities and the actual outcome agree poorly. In the calibration plot a perfect calibration curve would lie exactly on the diagonal line.

Internal validity

A prediction model can be validated internally to provide a more accurate estimate of model performance in other populations. As an internal validation of our model, we used the leave-one-out cross-validation method^{20,34} assessing overall performance (Brier), discrimination (AUC), and calibration (see the *Methods* section in this article's Online Repository for further explanations).

Sensitivity analyses

To test the robustness of the model developed in our original study population (P0), we performed sensitivity analyses using modified inclusion criteria at baseline or modified definitions of the outcome, resulting in slight changes of the study populations (P1 to P4, described in more detail in Tables E3 and E4 in this article's Online Repository at www.jacionline.org).

We first applied our existing prediction model to these modified populations and calculated the scaled Brier score and AUC (sensitivity analysis I). Second, we developed new models within the slightly modified study populations P1 to P4 and assessed their performance (sensitivity analysis II).

Clinical prediction tool

To simplify our model to a practical tool, we considered 3 different approaches: (1) multiplying regression coefficients by factors 10, 5, and 3 and rounding them to the nearest integer²⁰; (2) setting the penalty of the LASSO-penalized logistic regression so that only a few important predictors (5 or 3) were retained; and (3) considering a model with frequency of wheeze as the only predictor.¹⁹ All these approaches aimed to reduce the number of variables while maintaining a comparable predictive performance.

RESULTS**Study population**

At the baseline survey, 5878 of 6808 children were aged 1 to 3 years. Fig 1 shows how many of the 1- to 3-year-old children reported episodes of wheeze, cough without colds, or cough at night

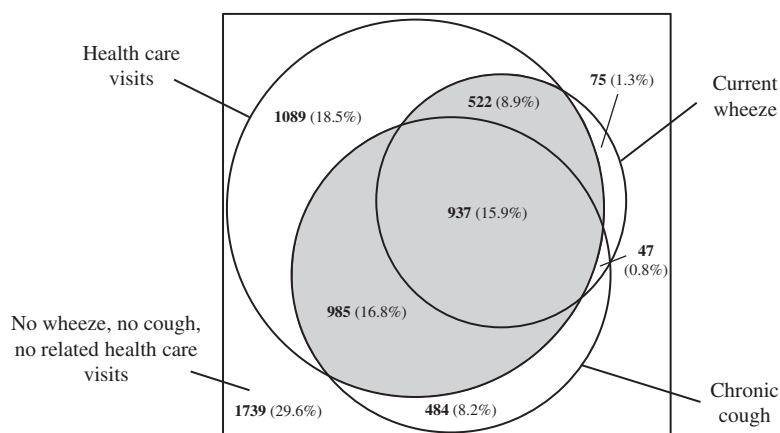


FIG 1. Wheeze, cough, and health care visits in 1- to 3-year-old-children. The illustration shows a proportional Venn diagram for children aged 1 to 3 years, showing frequency of health care visits caused by wheeze or cough, current wheeze, and chronic cough (cough without colds or cough at night). The gray area represents our study population.

in the past 12 months and in addition reported visits to a doctor ($n = 2444$), making them eligible for the study. For 1226 children, we had information on any asthma 5 years later. Their characteristics are shown in Table I for the variables selected by the main model and in Table E1 for all potential predictors considered. At baseline, 336 (27.4%) children were aged 1 year, 702 (57.3%) children were aged 2 years, and 188 (15.3%) children were aged 3 years. The mean prediction interval from baseline to outcome was 4.5 ± 0.5 ($\pm$ SD) years. At school age, 345 (28.1%) had any asthma.

Table E2 compares eligible children with and without follow-up information. The groups were comparable in many aspects (chronic cough, upper respiratory infections, eczema, and parental history), but those with follow-up information were more likely to be of white ethnicity and less likely to have wheeze at baseline.

Main prediction model

Of the 38 binary predictors that entered variable selection, the LASSO-penalized logistic regression retained 22 (Table II). The 5 most important predictors were, in order of importance, shortness of breath, frequent wheeze, wheeze triggered by exercise, disturbance by wheeze, and wheeze/cough with cough, activity disturbance. In addition, the model included aeroallergen-related wheeze/cough, male sex, age, birth weight, gestational age, eczema, upper respiratory tract symptoms, and parental history of wheeze, asthma, bronchitis, or hay fever.

In the original study population, the overall performance of the main model measured by using the scaled Brier score was 0.23, and its discriminative ability (AUC) was 0.78. In internal validation these measures were comparable (0.20 and 0.76, respectively). The calibration plot (Fig 2) shows good agreement between the predicted probabilities of later asthma and the observed frequencies in internal validation. The same was indicated by using the Hosmer-Lemeshow test ($P = .6$).

Sensitivity analyses

Sensitivity analysis I. The main model was robust to changes in baseline criteria (see P1 and P2 in Table E3). When

the outcome definition was changed to wheeze plus a doctor's diagnosis of asthma (P3) or to moderately severe asthma (≥ 4 attacks plus inhaled corticosteroids; P4), the AUC improved to 0.80 and 0.87, respectively (P3 and P4 in Table E3).

Sensitivity analysis II. The performance of new models developed in these alternative study populations was comparable with that of the main model for P1 to P3 and slightly improved for P4 (see Table E4). The selected predictors and estimated coefficients in the newly developed models (see Table E5 in this article's Online Repository at www.jacionline.org) were comparable with those of the main model. Severity-related predictors (wheeze without colds, frequent attacks, shortness of breath, and activity disturbance) gained comparatively more weight when predicting moderately severe asthma (P4).

Clinical prediction tool

We then simplified the model by using the 3 planned approaches. Our preferred simplification includes 10 variables (13 binary predictors), each of which contributes with one of 3 values (1, 2, or 3) to the prediction score (Fig 3; an online version of the prediction tool is available at www.leicestercohorts.org).

This tool was derived from the original model by multiplying all regression coefficients by 3 and rounding them to the nearest integer, dropping variables with coefficients rounded to zero.²⁰ It had almost the same discriminative ability (AUC = 0.775) as the main model (AUC = 0.782, Fig 4). Other approaches to simplification retained more predictors (making the tool complicated with little benefit) or had reduced discriminative ability (see Table E6 in this article's Online Repository at www.jacionline.org), particularly the model with frequency of wheeze only.

In internal validation the prediction tool showed only a minor decrease in performance compared with the main model: the scaled Brier score was 0.16, and the AUC was 0.74.

The maximum score a child can attain by using the prediction tool is 15, corresponding to a 95% probability of having any asthma 5 years later (Fig 3). Sensitivity and specificity of the tool are 0.72 and 0.71 for a score of 5 and 0.22 and 0.98 for a score of 10 (additional performance measures are reported in Table E7 in

TABLE I. Characteristics of the study population (n = 1226) at baseline by development of asthma 5 years later*

	5 y later: Asthma (n = 345)		5 y later: No asthma (n = 881)		P value†
	No.	Percent	No.	Percent	
Demographic and perinatal data					
Male sex	224	64.9	454	51.5	<.001
Age (y)					
1	85	24.6	251	28.5	.388
2	204	59.1	498	56.5	
3	56	16.2	132	15.0	
Gestational age <37 wk	35	10.1	49	5.6	.006
Birth weight <2500 g	41	11.9	68	7.7	.025
Wheeze-related symptoms‡					
Current wheeze	272	78.8	425	48.2	<.001
Wheeze without colds	127	36.8	95	10.8	<.001
Frequency of attacks					
0	81	23.5	476	54.0	<.001
1-3	111	32.2	281	31.9	
4-12	117	33.9	106	12.0	
>12	36	10.4	18	2.0	
Activity disturbance					
No	141	40.9	649	73.7	<.001
Little	129	37.4	185	21.0	
Moderate	57	16.5	39	4.4	
A lot	18	5.2	8	0.9	
Shortness of breath					
Never	129	37.4	668	75.8	<.001
Sometimes	166	48.1	190	21.6	
Always	50	14.5	23	2.6	
Exercise-related wheeze/cough§	196	56.8	286	32.5	<.001
Aeroallergen-related wheeze/ cough	52	15.1	37	4.2	<.001
Other symptoms‡					
Cough without colds	233	67.5	536	60.8	.030
Duration of colds (wk)					
<1	75	21.7	203	23.0	.194
1-2	198	57.4	533	60.5	
>2	72	20.9	145	16.5	
Nasal symptoms	186	53.9	350	39.7	<.001
Eczema (ever)	190	55.1	343	38.9	<.001
Parental history					
Wheeze, asthma, or bronchitis					
None	142	41.2	499	56.6	<.001
Father	68	19.7	136	15.4	
Mother	85	24.6	182	20.7	
Both	50	14.5	64	7.3	
Hay fever					
None	152	44.1	474	53.8	.001
Father	56	16.2	144	16.3	
Mother	93	27.0	203	23.0	
Both	44	12.8	60	6.8	

*This table includes all predictors that were selected for the main model.

†Fisher exact test.

‡During the last 12 months.

§Wheeze or cough with running, playing, laughing, or crying.

this article's Online Repository at www.jacionline.org). In our study sample 840 (69%) children were at low risk (score ≤5), 288 (23%) were at medium risk (score ≥6 and ≤9), and 98 (8%) were at high risk (score ≥10) of any asthma 5 years later. The percentage of children with any asthma at school age was 16%, 48%, and 79% in the low-, medium-, and high-risk groups, respectively.

TABLE II. Important factors for prediction of asthma at school age in symptomatic preschool children (selected by using penalized logistic regression)

	OR§	RC: Main model	Simplified RC*: Tool	Order of inclusion
Demographic and perinatal data				
Male sex	1.48	0.394	1	9
Age >1 y	1.19	0.171	1	16
Gestational age <37 wk	1.11	0.108		18
Birth weight <2500 g	1.17	0.154		17
Wheeze-related symptoms†				
Current wheeze	1.18	0.163		13
Wheeze without colds	1.40	0.337	1	3
Frequency of attacks >3	1.65	0.500	2	2
Activity disturbance				
Any	1.28	0.243	1	4
Moderate or a lot	1.16	0.144		7
A lot	1.63	0.491	1	13
Shortness of breath				
Sometimes or always	1.98	0.684	2	1
Always	1.56	0.442	1	6
Exercise-related wheeze/cough‡	1.26	0.233	1	5
Aeroallergen-related wheeze/cough	1.22	0.198	1	10
Other symptoms†				
Cough without colds	1.09	0.086		18
Duration of cold ≥1 wk	0.97	−0.031		22
Nasal symptoms	1.17	0.157		12
Eczema (ever)	1.52	0.420	1	7
Parental history				
Wheeze, asthma, or bronchitis				
Mother or father	1.23	0.203	1	10
Both parents	1.26	0.235	1	13
Hay fever				
Mother or father	1.03	0.025		21
Both parents	1.12	0.110		18
No. of binary predictors	22	22	13	22
No. of variables	17	17	10	17

OR, Odds ratio; RC, regression coefficient.

*Regression coefficient of the main model multiplied by 3 and rounded to the nearest integer (simplification approach in which the number of variables was substantially reduced without a relevant decrease in predictive performance).

†During the last 12 months.

‡Wheeze or cough with running, playing, laughing, or crying.

§CIs for the odds ratios are not provided because odds ratio estimates result from penalized logistic regression, which is primarily a method for variable selection rather than statistical inference. Estimates are deliberately biased toward the null, with the benefit of reducing their variance and improving overall prediction. CIs are misleading in this context.

DISCUSSION

Summary of findings

We have developed a new tool for predicting asthma at school age in preschool children who see a doctor for wheeze or cough. Our tool includes 10 predictors representing wheeze severity and triggers, male sex, age, eczema, and parental respiratory history. It showed good internal validity and is distinguished by ease of use in primary care and epidemiologic studies.

Comparison with previous prediction models

Several prediction models have been proposed for estimating the risk of persistent asthma in preschool children.⁸⁻¹⁶

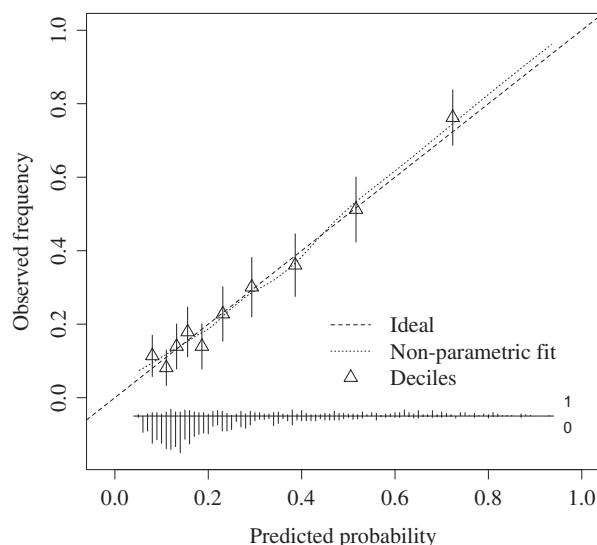


FIG 2. Calibration plot of the main model (assessed in leave-one-out cross-validation). Children are grouped into deciles of their predicted probability. The average predicted probability for later asthma among children within each decile is plotted against the actual observed frequency (prevalence) of asthma in that group. As a visual aid, a smoothing technique (locally weighted polynomial regression) was applied to these data. The *straight line* represents perfect calibration.

Table III^{8,9,13,36} summarizes inclusion criteria, outcome, methods used to derive the tool, predictors, and performance for 3 tools that used a similar prediction interval as ours and had a sample size of greater than 300. In short, Castro-Rodriguez et al⁸ (Tucson Children's Respiratory Study) used data from 2- to 3-year-olds with and without respiratory symptoms to develop 2 prediction tools for asthma at school age (loose and stringent Asthma Predictive Index [API], Table III). Kurukulaaratchy et al¹³ (Isle of Wight birth cohort) proposed a score for persistence of early wheeze up to age 10 years. Caudri et al⁹ (Prevention and Incidence of Asthma and Mite Allergy birth cohort) developed a clinical risk score for 0- to 4-year-olds with wheeze or cough to predict asthma at age 7 to 8 years.

The performance of these tools was comparable or slightly less than ours (Table III), with a Youden index³⁶ (sensitivity + specificity – 1) varying from 0.32⁸ to 0.38¹³ (calculated based on the maximal sum of sensitivity and specificity reported in the respective studies) compared with 0.43 in our study. The Youden index ranges between 0 and 1. Values close to 1 indicate large predictive effectiveness, and values close to 0 indicate limited effectiveness.

The method used to derive the APIs is difficult to replicate,⁸ whereas methods used for the other tools (logistic regression with stepwise variable selection)^{9,13} tend to overfit the data; that is, the models might be overly influenced by the random variation in the data used to develop them. This limits the application of the models to other populations.

Only Caudri et al⁹ performed an internal validation of their prediction model and reported an AUC (0.72) similar to the one we obtained (0.74). They included 8 predictors with exact regression coefficients, whereas our model includes 10 predictors with simplified regression coefficients that facilitate calculation of individual risks in a clinical setting. The Prevention and Incidence of Asthma and Mite Allergy risk

Asthma Prediction Tool

1. What is the child's sex? Female ☐ 0
Male ☐ 1
2. How old is the child? (in years) 1 ☐ 0
2 ☐ 1
3 ☐ 1
3. In the last 12 months, has the child had wheezing or whistling in the chest even without having a cold or flu? No ☐ 0
Yes ☐ 1
4. How many attacks of wheeze has the child had during the last 12 months? 0-3 ☐ 0
>3 ☐ 2
5. In the last 12 months, how much did wheezing interfere with the child's daily activities? No ☐ 0
A little ☐ 1
A lot ☐ 2
6. Do these wheezing attacks cause him/her to be short of breath? Never ☐ 0
Sometimes ☐ 2
Always ☐ 3
7. In the last 12 months, did exercise (playing, running) or laughing, crying or excitement cause wheezing or coughing in the child? No ☐ 0
Yes ☐ 1
8. In the last 12 months, did contact with dust, grass, pets or other animals cause wheezing or coughing in the child? No ☐ 0
Yes ☐ 1
9. Has the child ever had eczema? No ☐ 0
Yes ☐ 1
10. Have the child's parents ever suffered from wheezing, asthma or bronchitis? None ☐ 0
Mother ☐ 1
Father ☐ 1

Total Score = SUM = _____

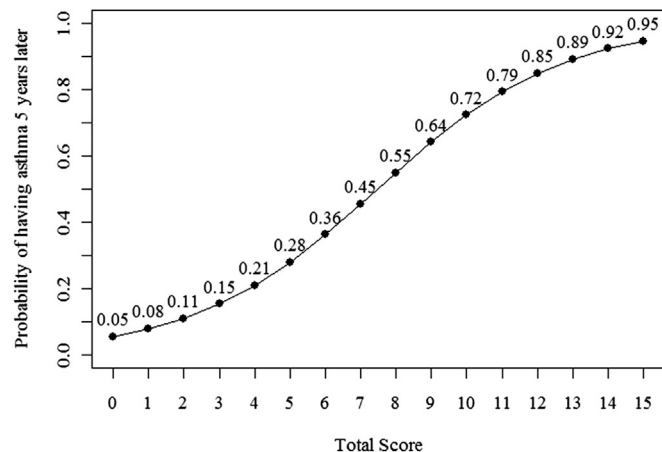


FIG 3. Asthma prediction tool. For any 1- to 3-year-old child seeking health care because of wheeze or cough, the applicable predictors are summed to a total score in the upper part of the figure. The estimated probability of having asthma 5 years later is given below for different total scores.

score and the API have been tested in a small external population.^{19,37}

In comparison with our tool, previous asthma prediction rules included at most 2 descriptors of wheeze (out of frequency, duration, or wheeze without colds).^{8-10,14} In addition, they relied on blood or skin prick tests,^{8,11-13,15} which are more time-consuming and costly than a detailed symptom history. Socioeconomic position is a proxy measure for a variety of exposures and health care access and might have a variable effect in different populations.⁹

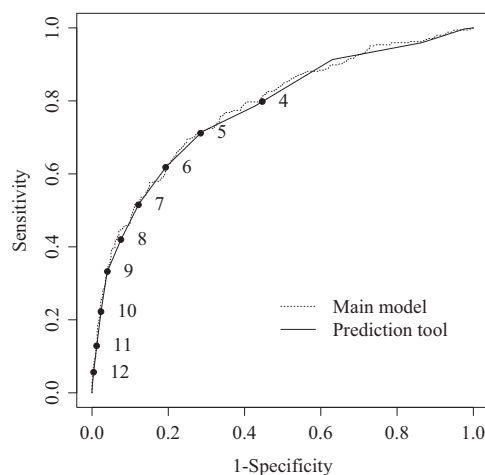


FIG 4. Receiver operating characteristic curves for the main asthma prediction model and the prediction tool. The dots represent the sensitivity and specificity for different cutoff values of the prediction tool.

Strengths and limitations

The main strengths of our tool are the objective approach used for its development and its clinical applicability. We used a population-based sample of an adequate size to develop the model. We included only children with health care visits for wheeze or cough, ensuring that the sample represents a clinically relevant population. We defined a clinically relevant outcome measure (wheeze needing treatment). When defining a more severe outcome (moderately severe asthma, which was defined as ≥ 4 attacks per year and inhaled corticosteroid treatment), the tool performed even better. All predictors are obtained routinely when taking a respiratory history for a child presenting with chronic cough or wheeze, and predictors are easy to assess, even during a short primary care consultation or in a questionnaire survey. We used a method that minimizes overfitting and is less affected by sampling variability compared with stepwise variable selection procedures,³⁸ and we performed an internal validation. Finally, our model predicts a range of probabilities rather than predicting only a low or high risk, as with the API.⁸

Like other studies,^{8,9,11,13} ours relies on parent-reported questionnaire data. However, it uses standardized questions, mostly from the International Study of Asthma and Allergies in Childhood,³⁹ and reflects, to some extent, the clinical situation in which parents report respiratory symptoms. The applied questionnaire showed good repeatability.⁴⁰ We did not use objective measurements to define our outcome. However, for a subsample of our study population ($n = 451$), we assessed bronchodilator response in a later survey conducted in 2006 (see Table E8 in this article's Online Repository at www.jacionline.org). By using the same outcome definitions (any asthma and moderately severe asthma), mean percentage change in FEV₁ was significantly higher in children with any asthma compared with that seen in those without (5.5% [95% CI, 3.6% to 7.3%] vs 2.6% [95% CI, 2.0% to 3.2], $P < .001$). For maximal expiratory flow at 50% of vital capacity, the mean percentage change was 16.7% (95% CI, 12.8% to 20.5%) and 10.7% (95% CI, 8.8% to 12.5%), respectively ($P = .003$). This is less than the cutoffs recommended for

clinical situations.⁴¹ However, our measurements came not from hospital-based children referred when they were unwell but from community-based children who were usually asymptomatic when measured. Our results are in line with data from Galant et al,⁴² in which bronchodilator responses for FEV₁ were 7.3% (95% CI, 4.2% to 10.4%) in patients with mild persistent asthma and 7.6% (95% CI, 5.8% to 9.5%) in patients with mild intermittent asthma compared with 2.2% (95% CI, 0.2% to 4.3%) in nonasthmatic subjects. Children with and without follow-up information were comparable (see Table E2), although we cannot exclude that selection bias has affected the composition of the final model. Finally, we interpreted missing values in potential predictor variables as an absence of the respective risk factor, which might also have affected the results. However, the number of missing values did not exceed 5.8% in any of the potential predictor variables.

Meaning of the study

Our model was robust, and results changed little with modifications of the inclusion criteria and outcomes. In fact, the performance improved (AUC = 0.89 vs 0.78) when we predicted moderately severe asthma rather than any asthma. After internal validation, the AUCs of the main model and tool were similar to those before validation, suggesting that there was little overfitting.

Our tool used only information on symptoms that can be gathered in a simple patient's history. Despite that, it had a similar or better predictive performance than previous tools, including more complex measurements.^{8,11,13-15} This suggests that a detailed description of presented symptoms might predict later asthma equally as well as more invasive methods, including measurement of blood eosinophilia or skin prick tests.^{8,11,13-15} Seven of 10 predictors (including the 5 strongest) describe the following symptoms: frequency of attacks, activity disturbance, shortness of breath, triggers (wheeze apart from colds, exercise, and aeroallergens), and eczema. This is consistent with the old knowledge that frequent wheeze strongly predicts asthma persistence^{10,43} and with our previous report, showing that the frequency of wheeze predicted asthma nearly as well as the complicated API rule.¹⁹ In our tool, adding more symptoms (in addition to wheeze frequency) improved performance (AUC after internal validation = 0.74 for the tool vs 0.57 for wheeze frequency only, see Table E6). This shows that more detailed assessment of symptoms in preschool children improves prediction of later asthma.

Future research

External validation in independent samples is necessary to further evaluate the predictive performance of the proposed tool and assess its generalizability to other populations.³⁴ Therefore we encourage the application and validation of this tool in ongoing epidemiologic studies and clinical care (particularly primary care). Some earlier prediction models^{8,9,13} performed similarly in external populations, but their performance remained modest.^{15,19,37}

Compared with other prediction rules, our tool includes a detailed description of symptom severity and pattern. This raises the possibility that further refinement in the description of

TABLE III. Comparison of 4 asthma prediction tools for preschool children

	Leicester (present study [Leicestershire Respiratory Cohort Studies])	Tucson ⁸ (API* [Tucson Children's Respiratory Study])	IoWBC ¹³ (Isle of Wight Birth Cohort)	PIAMA ⁹ (Prevention and Incidence of Asthma and Mite Allergy)
No. (included in analysis)	1226	776	336	2054
Inclusion criteria				
Age (y)	1-3	2-3	4	1-4
Symptoms	Health care visit because of respiratory problems plus ≥ 1 of the following symptoms in the past 12 mo: wheeze, cough without colds, cough at night	Entire cohort (including a majority of children without symptoms)	Wheeze at ages 1, 2, and 4 y	Wheeze or cough at night without colds (or both) in the past 12 mo
Outcome definition				
Age (y)	6-8	8	10	7-8
Prediction interval (y)	4-5	5	6	3-7
Criteria	Wheeze plus asthma medication (past 12 mo)	Doctor's diagnosis of asthma plus current wheeze or >3 wheeze episodes (past 12 mo)	Current wheeze	At ages 7 and 8 y: current wheeze or prescription of inhaled corticosteroids or doctor's diagnosis of asthma (past 12 mo)
Outcome prevalence	28.1%	13.7%	37.2%	11.7%
Predictor variables included in tool	Male sex, age >1 y, wheeze without colds, frequent wheeze, activity disturbance, shortness of breath, exercise-related wheeze/cough, [†] aeroallergen-related wheeze/cough, eczema, parental asthma or wheeze bronchitis	Wheeze, frequent wheeze, [‡] wheeze without colds, eczema, parental asthma, blood eosinophilia, allergic rhinitis	Family history of asthma, recurrent chest infections (at 2 y), skin prick test response positivity (at 4 y), nasal symptoms (at 1 y)	Male sex, postterm delivery, wheeze/dyspnea without colds, frequent wheeze, eczema, respiratory infections, inhalation medication (parents), parental education
Method used to derive tool	Penalized logistic regression	Combination of predictors was chosen that yielded the highest PPV and specificity	Stepwise backward logistic regression	Stepwise backward logistic regression
Performance measures§	Score cutoff ≥ 5	Loose API	Score cutoff ≥ 3	Score cutoff ≥ 20
Youden index ³⁶	0.43	0.32	0.38	0.36
Sensitivity (%)	72	51	53	60
Specificity (%)	71	81	85	76
PPV (%)	49	29	68	23
NPV (%)	86	91	74	94

NPV, Negative predictive value; PPV, positive predictive value.

*To have a prediction interval comparable with the one in our tool, we focused here on the API for prediction at 8 years.

[†]Wheeze or cough with running, playing, laughing, or crying.

[‡]This variable is only part of the stringent API but not of the loose API.

§Reported for cutoff where the sum of sensitivity and the specificity was maximal. It is possible that a higher sum of sensitivity and specificity exists at a cutoff point that was not reported in the respective studies.

the preschool wheeze phenotype might improve the precision of prediction of later asthma. Additional gains might be made by using detailed assessment of age-related changes, physiologic measurements (lung function, bronchial hyperresponsiveness, exhaled nitric oxide, and atopy), and environmental, socioeconomic, and genetic risk factors.¹⁷ However, all this could compromise the tool's simplicity.

Conclusions

This tool represents a simple, low-cost, and noninvasive method to predict the risk of later asthma in symptomatic preschool children and is ready to be tested in other populations.

We thank all the children and parents of Leicestershire respiratory cohorts for participating in the study and Kali Tal for her editorial assistance.

Clinical implications: The proposed asthma prediction tool is simple and uses information that is noninvasive and easy to assess. This makes it an ideal instrument for use in clinical practice and research.

REFERENCES

1. Stevens CA, Turner D, Kuehni CE, Couriel JM, Silverman M. The economic impact of preschool asthma and wheeze. *Eur Respir J* 2003;21:1000-6.
2. 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.
3. 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.

4. 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.
5. Spycher BD, Silverman M, Kuehni CE. Phenotypes of childhood asthma: are they real? *Clin Exp Allergy* 2010;40:1130-41.
6. Caudri D, Wijga AH, Smit HA, Koppelman GH, Kerkhof M, Hoekstra MO, et al. Asthma symptoms and medication in the PIAMA birth cohort: evidence for under and overtreatment. *Pediatr Allergy Immunol* 2011;22:652-9.
7. Robertson CF, Price D, Henry R, Mellis C, Glasgow N, Fitzgerald D, et al. Short-course montelukast for intermittent asthma in children: a randomized controlled trial. *Am J Respir Crit Care Med* 2007;175:323-9.
8. Castro-Rodriguez JA, Holberg CJ, Wright AL, Martinez FD. A clinical index to define risk of asthma in young children with recurrent wheezing. *Am J Respir Crit Care Med* 2000;162:1403-6.
9. Caudri D, Wijga A, Schipper CM, Hoekstra M, Postma DS, Koppelman GH, et al. Predicting the long-term prognosis of children with symptoms suggestive of asthma at preschool age. *J Allergy Clin Immunol* 2009;124:903-10, e1-7.
10. Devulapalli CS, Carlsen KC, Haland G, Munthe-Kaas MC, Pettersen M, Mowinckel P, et al. Severity of obstructive airways disease by age 2 years predicts asthma at 10 years of age. *Thorax* 2008;63:8-13.
11. Eysink PE, ter Riet G, Aalberse RC, van Aalderen WM, Roos CM, van der Zee JS, et al. Accuracy of specific IgE in the prediction of asthma: development of a scoring formula for general practice. *Br J Gen Pract* 2005;55:125-31.
12. Korppi M. Asthma predictive factors in infants with bronchiolitis: asthma risk at 13-20 years of age. *Eur Respir J* 2010;36:221-2.
13. Kurukulaaratchy RJ, Matthews S, Holgate ST, Arshad SH. Predicting persistent disease among children who wheeze during early life. *Eur Respir J* 2003;22:767-71.
14. Lodrup Carlsen KC, Soderstrom L, Mowinckel P, Haland G, Pettersen M, Munthe Kaas MC, et al. Asthma prediction in school children: the value of combined IgE-antibodies and obstructive airways disease severity score. *Allergy* 2010;65:1134-40.
15. Matricardi PM, Illi S, Keil T, Wagner P, Wahn U, Lau S. Predicting persistence of wheezing: one algorithm does not fit all. *Eur Respir J* 2010;35:701-3.
16. Balemans WA, van der Ent CK, Schilder AG, Sanders EA, Zielhuis GA, Rovers MM. Prediction of asthma in young adults using childhood characteristics: development of a prediction rule. *J Clin Epidemiol* 2006;59:1207-12.
17. Savenije OE, Kerkhof M, Koppelman GH, Postma DS. Predicting who will have asthma at school age among preschool children. *J Allergy Clin Immunol* 2012;130:325-31.
18. Corrao WM, Braman SS, Irwin RS. Chronic cough as the sole presenting manifestation of bronchial asthma. *N Engl J Med* 1979;300:633-7.
19. Leonardi NA, Spycher BD, Strippoli MP, Frey U, Silverman M, Kuehni CE. Validation of the Asthma Predictive Index and comparison with simpler clinical prediction rules. *J Allergy Clin Immunol* 2011;127:1466-72.e6.
20. Steyerberg E. Clinical prediction models: a practical approach to development, validation, and updating. Berlin: Springer-Verlag; 2009.
21. Kuehni CE, Brooke AM, Strippoli MP, Spycher BD, Davis A, Silverman M. Cohort profile: the Leicester respiratory cohorts. *Int J Epidemiol* 2007;36:977-85.
22. Dogaru CM, Strippoli MP, Spycher BD, Frey U, Beardsmore CS, Silverman M, et al. Breastfeeding and lung function at school age: does maternal asthma modify the effect? *Am J Respir Crit Care Med* 2012;185:874-80.
23. Spycher BD, Silverman M, Egger M, Zwahlen M, Kuehni CE. Routine vaccination against pertussis and the risk of childhood asthma: a population-based cohort study. *Pediatrics* 2009;123:944-50.
24. Bisgaard H, Bonnelykke K. Long-term studies of the natural history of asthma in childhood. *J Allergy Clin Immunol* 2010;126:187-99.
25. Frank PI, Morris JA, Hazell ML, Linehan MF, Frank TL. Long term prognosis in preschool children with wheeze: longitudinal postal questionnaire study 1993-2004. *BMJ* 2008;336:1423-6.
26. Lowe L, Custovic A, Woodcock A. Childhood asthma. *Curr Allergy Asthma Rep* 2004;4:159-65.
27. Matricardi PM, Illi S, Gruber C, Keil T, Nickel R, Wahn U, et al. Wheezing in childhood: incidence, longitudinal patterns and factors predicting persistence. *Eur Respir J* 2008;32:585-92.
28. Midodzi WK, Rowe BH, Majaesic CM, Saunders LD, Senthilselvan A. Predictors for wheezing phenotypes in the first decade of life. *Respirology* 2008;13:537-45.
29. Piippo-Savolainen E, Korppi M. Wheezy babies—wheezy adults? Review on long-term outcome until adulthood after early childhood wheezing. *Acta Paediatr* 2008;97:5-11.
30. Sherriff A, Peters TJ, Henderson J, Strachan D. Risk factor associations with wheezing patterns in children followed longitudinally from birth to 3(1/2) years. *Int J Epidemiol* 2001;30:1473-84.
31. Van Bever HP, Desager KN, Hagendorens M. Critical evaluation of prognostic factors in childhood asthma. *Pediatr Allergy Immunol* 2002;13:77-83.
32. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* 2010;33:1-22.
33. Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc Ser B* 1996;58:267-88.
34. Harrell F. Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. Berlin: Springer; 2001.
35. Hosmer DW, Hosmer T, Le Cessie S, Lemeshow S. A comparison of goodness-of-fit tests for the logistic regression model. *Stat Med* 1997;16:965-80.
36. Youden WJ. Index for rating diagnostic tests. *Cancer* 1950;3:32-5.
37. Rodriguez-Martinez CE, Sossa-Briceno MP, Castro-Rodriguez JA. Discriminative properties of two predictive indices for asthma diagnosis in a sample of pre-schoolers with recurrent wheezing. *Pediatr Pulmonol* 2011;46:1175-81.
38. Austin PC, Tu JV. Automated variable selection methods for logistic regression produced unstable models for predicting acute myocardial infarction mortality. *J Clin Epidemiol* 2004;57:1138-46.
39. Asher MI, Keil U, Anderson HR, Beasley R, Crane J, Martinez F, et al. International Study of Asthma and Allergies in Childhood (ISAAC): rationale and methods. *Eur Respir J* 1995;8:483-91.
40. Strippoli MP, Silverman M, Michel G, Kuehni CE. A parent-completed respiratory questionnaire for 1-year-old children: repeatability. *Arch Dis Child* 2007;92:861-5.
41. Pellegrino R, Viegi G, Brusasco V, Crapo RO, Burgos F, Casaburi R, et al. Interpretative strategies for lung function tests. *Eur Respir J* 2005;26:948-68.
42. Galant SP, Morpew T, Amaro S, Liao O. Value of the bronchodilator response in assessing controller naive asthmatic children. *J Pediatr* 2007;151:457-62.e1.
43. Phelan PD, Robertson CF, Olinsky A. The Melbourne Asthma Study: 1964-1999. *J Allergy Clin Immunol* 2002;109:189-94.

METHODS

Development of the main prediction model

We used the R package *glmnet* to fit the penalized logistic regression. The parameter α was set to 1 so that only a LASSO-type penalty was included. This tends to retain only the most influential predictors. The parameter λ , which determines the magnitude of the penalty, was set to a value that maximized the area under the receiver operating characteristic curve of resulting predictions in 10-fold cross-validation.^{E1} All potential predictors with more than 2 response categories were coded as binary variables. If the original categories were ordered, these dichotomous variables represented all possible cutoff points separating lower from higher categories. For instance, the frequency of wheezing episodes in the past 12 months (0, 1-3, 4-12, and >12) was coded into 3 binary variables indicating greater than 0, greater than 3, and greater than 12 episodes, respectively. This procedure resulted in 38 binary variables entering variable selection.

CIs for the odds ratios are not provided because odds ratio estimates result from penalized logistic regression, which is primarily a method for variable selection rather than statistical inference. Estimates are deliberately biased toward the null, with the benefit of reducing their variance and improving overall prediction. CIs are misleading in this context.

Data were prepared with STATA 11.0 software (StataCorp, College Station, Tex) and analyzed with R version 2.12.2. We used the R package *ROCR* to assess discrimination and the functions *hosmerlem* and *val.prob.ci* to assess calibration.^{E2}

Clinical prediction tool

To simplify our model to a practical tool, we considered 3 different approaches: (1) multiplying regression coefficients by factors 10, 5, and 3 and rounding them to the nearest integer^{E3}; (2) setting the penalty of the LASSO-penalized logistic regression so that only a few important predictors (5 or 3) were retained; and (3) considering a model with frequency of wheeze as the only predictor.^{E4} All these approaches aimed to reduce the number of variables while maintaining a comparable predictive performance.

In Table E7 the performances of these tools are compared with the main model in the sample (the sample used for model development) and by means

of internal validation (see below). In a final step, we recalibrated the probabilities for later asthma of the preferred tool by rerunning a logistic regression of the outcome on simplified scores.

Internal validation

To assess the reliability of our result of model performance within our study sample (ie, to test its repeatability within our development sample), we tested our model in leave-one-out cross-validation. The first step in this technique is to omit the first of total n observations and to use the remaining $n-1$ observations from the entire study sample to develop a new model. By using this new model, the probability for later asthma is estimated for the one observation left out before. In total, this procedure is repeated n times, each time omitting an observation that has not previously been left out. In the end, the internal validity of the model is tested based on these estimated probabilities.

Because the purpose was to test the main model's predictive performance and not how the method performs (including variable selection), we chose leave-one-out cross-validation as an internal validation technique that aims to fit models that are very similar to the main model. Other approaches, such as bootstrapping, would result in fitting models that are less similar to the main model and thus would have tested the repeatability of the method (variable selection approach and estimation of regression coefficients) rather than having validated the main model itself.

REFERENCES

- E1. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* 2010;33:1-22.
- E2. Homepage for clinical prediction models. Available at: http://survey.erasmusmc.nl/wiki/mgz-clinpred/doku.php?id=rcode_and_data:chapter15. Accessed November 9, 2011.
- E3. Steyerberg E. *Clinical prediction models: a practical approach to development, validation, and updating*. Berlin: Springer-Verlag; 2009.
- E4. Leonardi NA, Spycher BD, Strippoli MP, Frey U, Silverman M, Kuehni CE. Validation of the Asthma Predictive Index and comparison with simpler clinical prediction rules. *J Allergy Clin Immunol* 2011;127:1466-72.e6.

1. Has your child had **wheezing or whistling in the chest** in the last 12 months? yes ☐ no ☐
2. Does your child usually have a **cough apart from colds**? yes ☐ no ☐
3. In the last 12 months, has your child had a **dry cough at night**, apart from a cough associated with a cold or a chest infection? yes ☐ no ☐
4. How often did your child see the GP for coughing or wheezing during the last 12 months?
 never ☐ once ☐ 2 - 3 times ☐ 4 - 6 times ☐ 7 or more times ☐
5. In the last 12 months, has wheezing or asthma resulted in your child:
 - being referred to a consultant in hospital yes ☐ no ☐
 - being admitted to hospital yes ☐ no ☐
 - attending the casualty (A and E) department yes ☐ no ☐
 - attending (or calling) the GP in an emergency yes ☐ no ☐

FIG E1. Original questions used to define inclusion criteria at baseline.

6. Has your child had wheezing or whistling in the chest in the last 12 months? yes ☐ no ☐

7. Did your child take any of the following during the last 12 months?

<u>a blue inhaler</u> (Salbutamol, Ventolin, Bricanyl or other)	yes	☐	no	☐	don't know	☐
<u>a brown or orange inhaler</u> (Pulmicort, Flixotide, Becotide, Beclovent or other)	yes	☐	no	☐	don't know	☐
<u>Serevent or Oxis</u> (a green or green-white inhaler)	yes	☐	no	☐	don't know	☐
<u>Seretide or Symbicort</u> (a violet or red-white inhaler)	yes	☐	no	☐	don't know	☐

FIG E2. Original questions used to assess outcome at follow-up.

8. Has your child had **wheezing or whistling in the chest** in the last 12 months? yes ☐ no ☐

9. In the last 12 months, has your child had wheezing or whistling in the chest even **without** having a cold or flu? yes ☐ no ☐

10. **How many attacks of wheezing** has your child had during the last 12 months?
None ☐ 1 to 3 ☐ 4 to 12 ☐ more than 12 ☐

11. In the last 12 months, how much did **wheezing interfere with your child's daily activities**? not at all ☐ a little ☐ a moderate amount ☐ a lot ☐

12. Do these attacks cause him/her to be **short of breath**?
yes, always ☐ yes, occasionally ☐ no, never ☐

13. In the last 12 months, how often, on average, has your child's **sleep been disturbed due to wheezing**?
Never woken with wheezing ☐ less than one night per week ☐ one or more nights per week ☐

14. In the last 12 months did the following things cause **wheezing** in your child?
 • exercise (playing or running) yes ☐ no ☐ don't know ☐
 • laughing, crying or excitement yes ☐ no ☐ don't know ☐
 • contact with pets or other animals yes ☐ no ☐ don't know ☐
 • food or drinks yes ☐ no ☐ don't know ☐

15. Does your child usually have a **cough apart from colds**? yes ☐ no ☐

16. In the last 12 months, has your child had a **dry cough at night**, apart from a cough associated with a cold or a chest infection? yes ☐ no ☐

17. In the last 12 months, how many times has your child had a **cold or flu**?
never ☐ 1 - 3 times ☐ 4 - 6 times ☐ 7 - 10 times ☐ more than 10 times ☐

18. **How long does a cold** usually last in your child?
less than 1 week ☐ 1 to 2 weeks ☐ 2 to 4 weeks ☐ more than 4 weeks ☐

19. In the past 12 months, has your child had **ear infections**?
no, never ☐ yes, once ☐ yes, more than once ☐

20. In the past 12 months, has your child had a problem with **sneezing, or a runny, or blocked nose** when he/she **did NOT** have a cold or the flu? yes ☐ no ☐

21. Over the past 12 months, has your child **snored** at night? yes ☐ no ☐

22. In the past 12 months, has your child had **eczema**? yes ☐ no ☐

23. Has the **child's father** ever suffered from any of the following conditions?
 • **wheezing**? yes ☐ no ☐ don't know ☐
 • **asthma**? yes ☐ no ☐ don't know ☐
 • **bronchitis**? yes ☐ no ☐ don't know ☐
 • **hayfever**? yes ☐ no ☐ don't know ☐

24. Has the **child's mother** ever suffered from any of the following conditions?
 • **wheezing**? yes ☐ no ☐ don't know ☐
 • **asthma**? yes ☐ no ☐ don't know ☐
 • **bronchitis**? yes ☐ no ☐ don't know ☐
 • **hayfever**? yes ☐ no ☐ don't know ☐

FIG E3. Original questions used as potential predictive factors.

TABLE E1. Characteristics of the study population (n = 1226) at baseline by development of asthma 5 years later (all potential predictors considered in the analysis)

Question no.		Total study population (n = 1226)		5 y later: Asthma (n = 345)		5 y later: No asthma (n = 881)		P value†
		No.	Percent	No.	Percent	No.	Percent	
	Demographic and perinatal data							
	Male sex	678	55.3	224	64.9	454	51.5	<.001
	Age (y)							
	1	336	27.4	85	24.6	251	28.5	.388
	2	702	57.3	204	59.1	498	56.5	
	3	188	15.3	56	16.2	132	15.0	
	Gestational age <37 wk	84	6.9	35	10.1	49	5.6	.006
	Birth weight <2500 g	109	8.9	41	11.9	68	7.7	.025
	South Asian ethnicity (vs white)	316	25.8	78	22.6	238	27.0	.127
	Wheeze-related symptoms‡							
8	Current wheeze	697	56.9	272	78.8	425	48.2	<.001
9	Wheeze without colds	222	18.1	127	36.8	95	10.8	<.001
10	Frequency of attacks							
	0	557	45.4	81	23.5	476	54.0	<.001
	1-3	392	32.0	111	32.2	281	31.9	
	4-12	223	18.2	117	33.9	106	12.0	
	>12	54	4.4	36	10.4	18	2.0	
11	Activity disturbance							
	No	790	64.4	141	40.9	649	73.7	<.001
	Little	314	25.6	129	37.4	185	21.0	
	Moderate	96	7.8	57	16.5	39	4.4	
	A lot	26	2.1	18	5.2	8	0.9	
12	Shortness of breath							
	Never	797	65.0	129	37.4	668	75.8	<.001
	Sometimes	356	29.0	166	48.1	190	21.6	
	Always	73	6.0	50	14.5	23	2.6	
13	Sleep disturbance							
	Never	790	64.4	148	42.9	642	72.9	<.001
	<1	269	21.9	122	35.4	147	16.7	
	≥1	167	13.6	75	21.7	92	10.4	
14	Exercise-related wheeze/cough§	482	39.3	196	56.8	286	32.5	<.001
14	Aeroallergen-related wheeze/cough	89	7.3	52	15.1	37	4.2	<.001
14	Food-related wheeze/cough	186	15.2	54	15.7	132	15.0	.791
	Other symptoms‡							
15	Cough without colds	769	62.7	233	67.5	536	60.8	.030
16	Cough at night	631	51.5	190	55.1	441	50.1	.127
17	Frequency of colds							
	<4	447	36.5	101	29.3	346	39.3	.001
	4-6	461	37.6	134	38.8	327	37.1	
	>6	318	25.9	110	31.9	208	23.6	
18	Duration of colds (wk)							
	<1	278	22.7	75	21.7	203	23.0	.194
	1-2	731	59.6	198	57.4	533	60.5	
	>2	217	17.7	72	20.9	145	16.5	
19	Ear infection(s)							
	0	599	48.9	151	43.8	448	50.9	.020
	1	351	28.6	99	28.7	252	28.6	
	>1	276	22.5	95	27.5	181	20.5	
20	Nasal symptoms	536	43.7	186	53.9	350	39.7	<.001
21	Snoring	880	71.8	267	77.4	613	69.6	.006
22	Eczema (ever)	533	43.5	190	55.1	343	38.9	<.001
	Parental history							
23/24	Wheeze, asthma, or bronchitis							
	None	641	52.3	142	41.2	499	56.6	<.001
	Father	204	16.6	68	19.7	136	15.4	
	Mother	267	21.8	85	24.6	182	20.7	
	Both	114	9.3	50	14.5	64	7.3	

(Continued)

TABLE E1. (Continued)

Question no.		Total study population (n = 1226)		5 y later: Asthma (n = 345)		5 y later: No asthma (n = 881)		P value†
		No.	Percent	No.	Percent	No.	Percent	
23/24	Hay fever							
	None	626	51.1	152	44.1	474	53.8	.001
	Father	200	16.3	56	16.2	144	16.3	
	Mother	296	24.1	93	27.0	203	23.0	
	Both	104	8.5	44	12.8	60	6.8	

*See Fig E3 for the original questions used in questionnaires.

†Fisher exact test.

‡During the last 12 months.

§Wheeze or cough with running, playing, laughing, or crying.

TABLE E2. Characteristics of children at baseline by availability of follow-up information (n = 2444)

	Follow-up information available (n = 1226)		Follow-up information not available (n = 1218)		<i>P</i> value*
	No.	Percent	No.	Percent	
Demographic and perinatal data					
Male sex	678	55.3	633	52.0	.105
Gestational age <37 wk	84	6.9	86	7.1	.874
Birth weight <2500 g	109	8.9	86	7.1	.101
South Asian ethnicity (vs white)	316	25.8	386	31.7	.001
Wheeze-related symptoms†					
Current wheeze	697	56.9	762	62.6	.004
Wheeze without colds	222	18.1	272	22.3	.010
Frequency of attacks					
0	557	45.4	482	39.6	.012
1-3	392	32.0	419	34.4	
4-12	223	18.2	269	22.1	
>12	54	4.4	48	3.9	
Activity disturbance					
No	790	64.4	725	59.5	.044
Little	314	25.6	371	30.5	
Moderate	96	7.8	91	7.5	
A lot	26	2.1	31	2.5	
Shortness of breath					
Never	797	65.0	749	61.5	.193
Sometimes	356	29.0	387	31.8	
Always	73	6.0	82	6.7	
Sleep disturbance					
Never	790	64.4	728	59.8	.059
<1	269	21.9	304	25.0	
≥1	167	13.6	186	15.3	
Exercise-related wheeze/cough‡	482	39.3	531	43.6	.033
Aeroallergen-related wheeze/cough	89	7.3	104	8.5	.261
Food-related wheeze/cough	186	15.2	196	16.1	.540
Other symptoms†					
Cough without colds	769	62.7	798	65.5	.152
Cough at night	631	51.5	612	50.2	.571
Frequency of colds					
<4	447	36.5	420	34.5	.498
4-6	461	37.6	484	39.7	
>6	318	25.9	314	25.8	
Duration of colds (wk)					
<1	278	22.7	268	22.0	.897
1-2	731	59.6	737	60.5	
>2	217	17.7	213	17.5	
Ear infection(s)					
0	599	48.9	613	50.3	.481
1	351	28.6	322	26.4	
>1	276	22.5	283	23.2	
Nasal symptoms	536	43.7	569	46.7	.143
Snoring	880	71.8	877	72.0	.928
Eczema (ever)	533	43.5	548	45.0	.464
Parental history					
Wheeze, asthma, or bronchitis					
None	641	52.3	647	53.1	.581
Father	204	16.6	178	14.6	
Mother	267	21.8	276	22.7	
Both	114	9.3	117	9.6	
Hay fever					
None	626	51.1	646	53.0	.702
Father	200	16.3	199	16.3	
Mother	296	24.1	271	22.2	
Both	104	8.5	102	8.4	

*Fisher exact test.

†During the last 12 months.

‡Wheeze or cough with running, playing, laughing, or crying.

TABLE E3. Sensitivity analysis I: Testing performance of the main asthma prediction model in alternative study populations

Study population	Baseline criteria: 1- to 3-year-olds			Outcome definition: 5 y later			No. (total)	No. (outcome)	Percent	Brier (scaled)	AUC
	Health care visit and any wheeze or chronic cough	Health care visit and any wheeze	Any wheeze	Any wheeze and asthma medication	Any wheeze and ever doctor-diagnosed asthma	>4 Episodes of wheeze and inhaled corticosteroids					
P0 (used for main model)	✓			✓			1226	345	28.1	0.23	0.78
P1			✓	✓			769	285	37.1	0.21	0.77
P2		✓		✓			697	272	39.0	0.22	0.77
P3	✓				✓		1239	331	26.7	0.25	0.80
P4	✓					✓	1053	71	6.7	−0.51*	0.87

Baseline and outcome criteria refer to the past 12 months, if not otherwise stated.

*The negative scaled Brier score is due to the large difference in the prevalence of the outcome in P0 and P4. A simple recalibration without changing the score would lead to a scaled Brier score of 0.24.

TABLE E4. Sensitivity analysis II: testing performance of newly developed asthma prediction models based on alternative study populations

populations													
Study population	Baseline criteria: 1- to 3-year-olds			Outcome definition: 5 y later									
	Health care visit and any wheeze or chronic cough	Health care visit and any wheeze	Any wheeze	Any wheeze and asthma medication	Any wheeze and ever doctor-diagnosed asthma	>4 Episodes of wheeze and inhaled corticosteroids	No. of binary predictors in the model	No. (total)	No. (outcome)	Percent	Brier (scaled)	AUC	
P0 (used for main model)	✓				✓		22	1226	345	28.1	0.23	0.78	
P1			✓	✓	✓		25	769	285	37.1	0.22	0.77	
P2		✓		✓			23	697	272	39.0	0.23	0.78	
P3	✓					✓	26	1239	331	26.7	0.26	0.81	
P4	✓						20	1053	71	6.7	0.28	0.89	

Baseline and outcome criteria refer to the past 12 months, if not otherwise stated.

TABLE E5. Selected predictors in sensitivity analysis II and corresponding odds ratios

	Main model,* OR	New models (alternative populations), OR			
		P1†	P2‡	P3§	P4
Demographic and perinatal data					
Male sex	1.48	1.43	1.49	1.68	1.00
Age (y)					
≥2	1.19	1.53	1.51	1.28	1.00
3	1.00	1.00	1.01	1.06	0.95
Gestational age <37 wk	1.11	1.13	1.00	1.16	1.00
Birth weight <2500 g	1.17	1.18	1.28	1.34	1.00
South Asian ethnicity (vs white)	1.00	1.00	1.00	1.00	0.53
Wheeze-related symptoms¶					
Current wheeze	1.18	1.00	1.00	1.59	1.46
Wheeze without colds	1.40	1.55	1.45	1.42	2.11
Frequency of attacks					
≥1	1.00	1.00	1.00	1.05	1.00
>3	1.65	1.53	1.60	1.37	1.16
>12	1.00	1.00	1.00	1.00	2.10
Activity disturbance					
Any	1.28	1.30	1.25	1.28	1.49
Moderate or a lot	1.16	1.31	1.17	1.14	1.00
A lot	1.63	1.94	1.87	1.81	2.18
Shortness of breath					
Sometimes or always	1.98	1.90	1.91	1.84	2.06
Always	1.56	1.40	1.41	2.10	2.70
Sleep disturbance					
≥1/wk	1.00	1.00	1.00	1.10	1.00
>1/wk	1.00	1.00	1.00	1.00	1.20
Exercise-related wheeze/cough#	1.26	1.09	1.15	1.40	1.27
Aeroallergen-related wheeze/cough	1.22	1.05	1.04	1.33	1.00
Food-related wheeze/cough	1.00	1.03	1.02	0.97	1.00
Other symptoms¶					
Cough without colds	1.09	1.10	1.07	1.16	1.37
Cough at night	1.00	1.12	1.13	1.06	1.00
Frequency of colds					
>3	1.00	1.00	1.00	1.00	1.06
>6	1.00	0.97	1.00	1.00	1.00
Duration of colds (wk)					
≥1	0.97	0.89	0.90	0.80	1.00
>2	1.00	1.00	1.00	1.00	1.00
Ear infection(s)					
≥1	1.00	1.13	1.00	1.00	1.00
>1	1.00	1.00	1.00	1.00	1.00
Nasal symptoms	1.17	1.14	1.13	1.18	1.14
Snoring	1.00	1.00	1.00	1.00	1.00
Eczema (ever)	1.52	1.42	1.50	1.39	1.62
Parental history					
Wheeze or bronchitis					
Mother or father	1.23	1.14	1.06	1.45	1.07
Mother or both	1.00	1.00	1.00	1.00	1.00
Both parents	1.26	1.57	1.36	1.39	2.02
Hay fever					
Mother or father	1.03	1.00	1.00	1.00	1.09
Mother or both	1.00	1.05	1.01	1.00	1.00
Both parents	1.12	1.28	1.37	1.41	1.34

Baseline and outcome criteria refer to the past 12 months, if not otherwise stated.

OR, Odds ratio.

*Inclusion criteria: 1- to 3-year-olds with health care visit plus either wheeze or cough without colds or cough at night; outcome: wheeze plus asthma medication at age 6 to 8 years.

†Inclusion criterion: 1- to 3-year-olds with wheeze; outcome: same as in main model.

‡Inclusion criteria: 1- to 3-year-olds with health care visit plus wheeze; outcome: same as in main model.

§Inclusion criteria: same as in main model; outcome: current wheeze plus doctor's diagnosis of asthma (ever) at age 6 to 8 years.

||Inclusion criteria: same as in main model; outcome: >4 episodes of wheeze and using inhaled corticosteroids.

¶During the last 12 months.

#Wheeze or cough with running, playing, laughing, or crying.

TABLE E6. Predictive performance of simplified versions of the main asthma prediction model

Simplification approach	No. of binary predictors in the model	Brier score (scaled)		AUC	
		Before validation§	After validation	Before validation§	After validation
Main model					
No simplification	22	0.23	0.20	0.78	0.76
Rounded model*					
Factor 10	20	0.23	0.19	0.78	0.75
Factor 5	19	0.23	0.21	0.78	0.77
Factor 3#	13	0.22	0.16	0.78	0.74
Reduced model					
First 5 predictors only†	5	0.14	0.13	0.75	0.64
First 3 predictors only‡	3	0.12	0.11	0.73	0.60
Frequent wheeze only¶	3	0.13	0.12	0.70	0.57

*Using simplified regression coefficients of the model (regression coefficients of main model multiplied by 10, 5, or 3, respectively, and rounded to the next integer).

†Shortness of breath caused by wheeze, frequent wheeze episodes (>3), wheeze without colds, and activity disturbance caused by wheeze (exercise-related wheeze/cough).

‡Shortness of breath caused by wheeze, frequent wheeze episodes (>3), and wheeze without colds.

§Before internal validation: assessment using same sample as used to develop the model.

||After internal validation: assessment using leave-one-out cross-validation.

¶A 4-level variable coded as 3 binary dummy variables; analysis using logistic regression without penalization.

#Preferred model.

TABLE E7. Performance measures of the prediction tool for different cutoff values (calculated in the sample used to develop the tool without cross-validation)

Score cutoff	Sensitivity	Specificity	PPV	NPV	LR+	LR–
0	>0.99	<0.01	0.28	NA	1.00	*
1	>0.99	0.02	0.29	0.95	1.02	0.12
2	0.96	0.14	0.30	0.89	1.11	0.30
3	0.91	0.37	0.36	0.92	1.45	0.23
4	0.79	0.57	0.42	0.87	1.84	0.37
5	0.72	0.71	0.49	0.86	2.47	0.40
6	0.62	0.80	0.55	0.84	3.18	0.47
7	0.52	0.88	0.62	0.82	4.19	0.55
8	0.42	0.92	0.68	0.80	5.53	0.63
9	0.33	0.96	0.77	0.79	8.32	0.70
10	0.22	0.98	0.79	0.76	9.36	0.80
11	0.13	0.99	0.80	0.74	10.45	0.88
12	0.06	>0.99	0.83	0.73	12.77	0.95
13	0.02	>0.99	0.89	0.72	20.43	0.98
14	0.01	>0.99	>0.99	0.72	*	0.99
15	<0.01	>0.99	NA	0.72	*	>0.99

Sensitivity, specificity, positive predictive value, and negative predictive value are restricted to values between 0 and 1.

LR +, Positive likelihood ratio; LR –, negative likelihood ratio; NPV, negative predictive value; PPV, positive predictive value.

*Great uncertainty of estimate because of sensitivity and specificity close to 0 or 1.

TABLE E8. Comparison of percentage change in FEV₁ and maximal expiratory flow at 50% of vital capacity after bronchodilator administration by questionnaire-based outcome definitions assessed at the same time

Outcome criteria	Any asthma (current wheeze and asthma medication)		Moderately severe asthma (>4 episodes of wheeze in past 12 mo and inhaled corticosteroids)	
	Yes	No	Yes	No
Fulfilling outcome criteria				
No.(FEV1)	111	340	30	389
Mean % change in FEV ₁ after bronchodilator administration	5.46 (95% CI, 3.58-7.34)	2.59 (95% CI, 1.96-3.21)	9.10 (95% CI, 3.74-14.45)	2.76 (95% CI, 2.15-3.38)
No.(MEF50)	109	334	29	382
Mean % change in MEF ₅₀ after bronchodilator administration	16.66 (95% CI, 12.80-20.53)	10.65 (95% CI, 8.75-12.54)	18.60 (95% CI, 9.75-27.46)	11.21 (95% CI, 9.39-13.03)

t Tests: Any asthma: $P_{FEV1} \leq .001$ and $P_{MEF50} = .003$; moderately severe asthma: $P_{FEV1} < .001$ and $P_{MEF50} = 0.039$.

No.(FEV1), Number of children with data on forced expiratory volume in one second; No.(MEF50), Number of children with data on maximal expiratory flow at 50% of vital flow capacity.