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 2014;133:111-8.)

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 need[6] 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 [a]the Institute of Social and Preventive Medicine (ISPM) and [b]the Institute of Mathematical Statistics and Actuarial Science, University of Bern, and [c]the 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.

Available online July 24, 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

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 I 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 if 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 integer[20]; (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