Environmental data do not improve a clinical asthma prediction tool for children

Anina M. Pescatore, MSc[1], Ben D. Spycher, PhD[1], Maja Jurca, MD[1], Erol A. Gaillard, MD, PhD[2], Claudia E. Kuehni, MD, MSc[1]

[1]Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland

[2]Division of Child Health, Department of Infection, Immunity and Inflammation, University of Leicester, Leicester, UK

Correspondence

Prof. Claudia Kuehni

Institute of Social and Preventive Medicine

Finkenhubelweg 11, CH-3012 Bern, Switzerland

Tel.: +41 (0)31 631 35 07

Fax: +41 (0)31 631 35 20

E-mail: kuehni@ispm.unibe.ch

Word count

1045/1000

Number of figures and tables

1 table

Funding

This study was funded by Swiss National Science Foundation (PDFMP3-123162, 3200B0-122341, 32003B_144068, PDFMP3 137033) and Asthma UK 07/048. BD Spycher was supported by a Swiss National Science Foundation fellowship (PZ00P3_147987/1).

28

29 **Conflict of interest**

30 None

31

32 **Capsule summary**

33 Available asthma prediction tools perform only moderately well. We expanded the Childhood Asthma
34 Risk Assessment tool (CARAT) to include environmental and socioeconomic information, and found
35 that its performance was not improved. [31/35]

36

37 **Key words**

38 Wheeze, cough, children, prediction, persistence, longitudinal, cohort study, socioeconomic factors

39

40

To the Editor:

Many preschool children present to their doctor with respiratory symptoms, but not all of them develop asthma. Prediction tools can help distinguish children with a high risk of developing asthma from children whose risk is low. A good prediction tool selects children who need therapeutic intervention, and reassures parents whose children have transient problems.

Five tools to predict school-age asthma in symptomatic preschool children are currently available. All predict only moderately well (Youden index ≤0.43 or area under receiver operating characteristics curve [AUC] ≤0.74).[1-5] They include the asthma predictive index[4], the PIAMA prediction tool[3] and the Childhood Asthma Risk Assessment Tool (CARAT)[2] , which we developed previously. The latter consists of 10 clinical predictors, including current respiratory symptoms. To develop the CARAT, we deliberately included only predictors that can be easily assessed clinically. Thus, we considered neither physiological measurements, nor environmental or socioeconomic factors. These factors might not be generalizable to other populations.

Nonetheless, environmental and socioeconomic factors have been associated with respiratory symptoms in children, such as second hand tobacco smoke or house pets.[6, 7] Some asthma prediction tools for children and young adults do include such exposures, namely maternal smoking or parental education.[3, 5] In this study, we test the addition of environmental exposures and socioeconomic factors to see if they improve the predictive performance of the CARAT.

Our study population was the same that we used to develop the CARAT. We used questionnaire data from a population-based cohort from Leicestershire, United Kingdom, described in detail elsewhere.[8] We included children aged 1-3 years at baseline (in 1998) with parent-reported wheeze or chronic cough (cough without colds, or cough at night), who visited their doctor for wheeze or cough at least once during the past 12 months. The outcome "any asthma" was assessed five years later, at the age of 6-8 years. "Any asthma" was defined as current wheeze plus use of asthma medication within the past 12 months. For each child, we calculated the CARAT risk score for developing asthma (range of score: 0-15).[2] We then investigated if the following environmental and socioeconomic factors, assessed at baseline, improved the accuracy of the score's prediction: nursery care, number of

69 older siblings, heating or cooking with gas, pet ownership (cat, dog, other furry pets, bird), mother
70 smoking during pregnancy, exposure to environmental tobacco smoke (mother or other persons in the
71 household smoking), duration of breastfeeding, ethnicity (white vs. South Asian), crowding, single
72 parenthood, parental education, Townsend deprivation index[9], living in an urban area, and self-
73 reported traffic density at home address.
74 As when we developed the CARAT, we used least absolute shrinkage and selection operator
75 (LASSO)-penalized logistic regression to identify important predictors without over-fitting the data.[10]
76 The penalty for the regression coefficients is set using the penalization parameter λ. For large values
77 of λ, no predictors enter the model. With decreasing λ, more predictors enter the model, in order of
78 their added predictive value. For our main model, we set λ to a value that maximized the AUC of
79 resulting predictions in 10-fold cross validation.
80 We varied λ, to explore the order in which predictors entered the final model. We also used
81 conventional logistic regression without penalization to estimate univariable associations of each
82 potential predictor with later asthma, to see how estimates changed when adjusted for the CARAT
83 score, and when adjusted for all potential predictors. Methodological details are in the online
84 repository.
85
86 We had baseline data from 6808 children, of whom 2444 reported respiratory symptoms and a visit
87 to the doctor due to their symptoms. We had outcome data at age 6-8 years for 1226/2444 children
88 (50%), of whom 28% (345/1226) had asthma.
89 Those with school-age asthma differed little from those without in respect to environmental
90 exposures and socioeconomic factors, with few exceptions (maternal smoking, cooking fuel). (Table
91 I)
92 Of 31 potential predictors that entered variable selection, the CARAT score was the only variable that
93 remained in the final model (maximal AUC of 10-fold cross validations = 0.783; λ=0.083).

When we reduced λ of our main model by 70%, absence of nursery care entered the model as a second predictor (AUC=0.780). When λ was lowered by 79%, maternal smoking (AUC=0.780) and absence of crowding entered as well (AUC=0.781).

In the regression models without penalization, few potential predictors showed an association with asthma ($p<0.05$) (Table S1). In multivariable logistic regressions, adjusted for the score and all potential predictors, only absence of nursery care was associated with later asthma.

This study found no evidence that addition of environmental and socioeconomic data improves the predictive performance of the CARAT.

The CARAT already contains rather detailed information on respiratory symptoms. Environmental stimuli (air pollution, allergens, infections) do directly affect prevalence of these symptoms in toddlers. It is conceivable that they don't have an additional effect on persistence of symptoms (i.e. prediction of later asthma). One might even hypothesize that toddlers who wheeze and cough a lot because of increased exposure to infectious agents (nursery care, crowding) will tend to have a better prognosis compared to peers who have these symptoms in the absence of exposure. This might explain the trend towards a poorer prognosis in children who were not in nursery care and did not live in crowded households, which was seen in some of the models with reduced penalization.

Our findings contrast with those reported by Balemans et al, who found that maternal smoking while children were toddlers predicted asthma in young adults in their cohort.[5] Balemans only included a few symptoms as potential predictors and used stepwise logistic regression to derive the final model, which might explain why maternal smoking was a better predictor in their model than in ours. In our cohort, maternal smoking was one of the first predictors joining the CARAT score when we lowered the penalization, but it did not improve the predictive performance of CARAT.

The strength of our study lies in its large sample size and clinically relevant population. We used an objective approach for variable selection that minimized over-fitting the data. A limitation common to other tools is that symptoms and exposures are parent-reported. This reflects the situation in clinical

practice, where many decisions are based on medical history taken from parents. Future research should evaluate if such tools can be improved by including results from clinical tests.

In summary, the asthma risk assessment tool CARAT, which uses detailed clinical data, performs moderately well. Adding information on environmental and socioeconomic exposures did not improve the CARAT's predictive performance.

Anina M. Pescatore, MSc[1],

Ben D. Spycher, PhD[1],

Maja Jurca,MD[1],

Erol A. Gaillard, MD, PhD[2],

Claudia E. Kuehni ,MD, MSc[1]

[1]Institute of Social and Preventive Medicine, University Bern, Bern, Switzerland

[2]Division of Child Health, Department of Infection, Immunity and Inflammation, University of Leicester, Leicester, UK

Acknowledgements

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

References

1. van der Mark LB, van Wonderen KE, Mohrs J, van Aalderen WM, Ter Riet G, Bindels PJ. Predicting asthma in preschool children at high risk presenting in primary care: development of a clinical asthma prediction score. Prim Care Respir J 2014.
2. Pescatore AM, Dogaru CM, Duembgen L, Silverman M, Gaillard EA, Spycher BD, Kuehni CE. A simple asthma prediction tool for preschool children with wheeze or cough. J Allergy Clin Immunol 2014; 133:111-8 e13.
3. Caudri D, Wijga A, CM AS, Hoekstra M, Postma DS, Koppelman GH, Brunekreef B, Smit HA, de Jongste JC. 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.
4. 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.
5. 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.
6. Vork KL, Broadwin RL, Blaisdell RJ. Developing asthma in childhood from exposure to secondhand tobacco smoke: insights from a meta-regression. Environ Health Perspect 2007; 115:1394-400.
7. Svanes C, Heinrich J, Jarvis D, Chinn S, Omenaas E, Gulsvik A, Kunzli N, Burney P. Pet-keeping in childhood and adult asthma and hay fever: European community respiratory health survey. J Allergy Clin Immunol 2003; 112:289-300.
8. 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.
9. Townsend P, Philimore P, Beattie A. Health and deprivation: inequality and the North. Bristol: Croom Helm; 1988.
10. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw 2010; 33:1-22.

169

Table I. Characteristics of the study population (children seeing a doctor for wheeze or cough at age 1-3years, by asthma outcome at age 6-8 ; N=1226)

		5 yrs later: Asthma (n=345)		5 yrs later: No asthma (n=881)		
		n	(%)	n	(%)	p-value‖
Demographic factors						
Male		224	(64.9)	454	(51.5)	<0.001
Age (years)	1	85	(24.6)	251	(28.5)	0.388
	2	204	(59.1)	498	(56.5)	
	3	56	(16.2)	132	(15.0)	
Ethnicity	White	267	(77.4)	643	(73.0)	0.127
	South Asian	78	(22.6)	238	(27.0)	
Current wheeze and total asthma prediction score						
Current wheeze		272	(78.8)	425	(48.2)	<0.001
CARAT score (mean [SD])*		6.7	(3.2)	3.7	(2.3)	<0.001
Environmental exposures						
Nursery care		164	(47.5)	451	(51.2)	0.254
Older siblings	0	106	(30.7)	281	(31.9)	0.548
	1 or 2	202	(58.6)	523	(59.4)	
	>2	37	(10.7)	77	(8.7)	
Heating	central heating only	245	(71.0)	638	(72.4)	0.621
	gas, coal, other	100	(29.0)	243	(27.6)	
Cooking fuel	electrical stove only	102	(29.6)	197	(22.4)	0.010
	gas, other	243	(70.4)	684	(77.6)	
Pet ownership	cat	63	(18.3)	161	(18.3)	1.000
	dog	66	(19.1)	153	(17.4)	0.507
	other furry pet	43	(12.5)	78	(8.9)	0.070
	bird	13	(3.8)	38	(4.3)	0.752
Mother smoking during pregnancy		53	(15.4)	121	(13.7)	0.467
Mother smoking (number of cigarettes /day)	1 to 10	39	(11.3)	90	(10.2)	0.045
	>10	40	(11.6)	65	(7.4)	
Other person smoking in household (number of cigarettes /day)	1 to 10	38	(11.0)	131	(14.9)	0.201
	>10	37	(10.7)	87	(9.9)	
Breastfed (months)	<1	39	(11.3)	89	(10.1)	0.320
	1 to 3	54	(15.7)	164	(18.6)	
	4 to 6	40	(11.6)	120	(13.6)	
	> 6	57	(16.5)	162	(18.4)	
Self-reported traffic density (at home)	low	142	(41.2)	343	(38.9)	0.384
	moderate	176	(51.0)	447	(50.7)	
	high	27	(7.8)	91	(10.3)	
Socioeconomic factors						
Crowding (persons/room)	≤1	277	(80.3)	676	(76.7)	0.353
	1.1 to 1.5	53	(15.4)	152	(17.3)	
	>1.5	15	(4.3)	53	(6.0)	

Single parents		42	(12.2)	84	(9.5)	0.175
Higher parental education†		200	(58.0)	506	(57.4)	0.898
Townsend deprivation index‡	more affluent	71	(20.6)	172	(19.5)	0.695
	affluent	73	(21.2)	168	(19.1)	
	average	74	(21.4)	187	(21.2)	
	deprived	59	(17.1)	181	(20.5)	
	more deprived	68	(19.7)	173	(19.6)	
Living in an urban area§		178	(51.6)	461	(52.3)	0.849

*Range: 0 to 15 points, 0 represents low risk for having asthma 5 years later, 15 high risk[1]

†Age at the end of education is >16 years

‡The categories cover the following Townsend Deprivation Index intervals: [-5.522, -2.981], [-2.886, -1.264], [-1.250, 0.908], [0.909, 4.403], [4.418, 11.072]

§Living in Leicester post code areas LE1 to LE5

‖Fisher's exact test

170