Received: 20 January 2020 | Revised: 27 February 2020 | Accepted: 28 February 2020

DOI: 10.1111/pai.13247

ORIGINAL ARTICLE

WILEY

Prediction models for childhood asthma: A systematic review

Dilini M. Kothalawala[1,2] | Latha Kadalayil[1] | Veronique B. N. Weiss[1] | Mohammed Aref Kyyaly[3,4] | Syed Hasan Arshad[2,3,4] | John W. Holloway[1,2] | Faisal I. Rezwan[1,5]

[1]Human Development and Health, Faculty of Medicine, University of Southampton, Southampton, UK

[2]NIHR Southampton Biomedical Research Centre, University Hospitals Southampton, Southampton, UK

[3]The David Hide Asthma and Allergy Research Centre, St. Mary's Hospital, Isle of Wight, UK

[4]Clinical and Experimental Sciences, Faculty of Medicine, University of Southampton, Southampton, UK

[5]School of Water, Energy and Environment, Cranfield University, Cranfield, UK

Correspondence
John W. Holloway, Southampton General Hospital, Duthie Building, MP808, Southampton SO16 6YD, UK.
Email: j.w.holloway@soton.ac.uk

Funding information
This work was supported by the NIHR Southampton Biomedical Research Centre and a University of Southampton Presidential Research Studentship.

Editor: Ömer Kalaycı

Abstract

Background: The inability to objectively diagnose childhood asthma before age five often results in both under-treatment and over-treatment of asthma in preschool children. Prediction tools for estimating a child's risk of developing asthma by school-age could assist physicians in early asthma care for preschool children. This review aimed to systematically identify and critically appraise studies which either developed novel or updated existing prediction models for predicting school-age asthma.

Methods: Three databases (MEDLINE, Embase and Web of Science Core Collection) were searched up to July 2019 to identify studies utilizing information from children ≤5 years of age to predict asthma in school-age children (6-13 years). Validation studies were evaluated as a secondary objective.

Results: Twenty-four studies describing the development of 26 predictive models published between 2000 and 2019 were identified. Models were either regression-based (n = 21) or utilized machine learning approaches (n = 5). Nine studies conducted validations of six regression-based models. Fifteen (out of 21) models required additional clinical tests. Overall model performance, assessed by area under the receiver operating curve (AUC), ranged between 0.66 and 0.87. Models demonstrated moderate ability to either rule in or rule out asthma development, but not both. Where external validation was performed, models demonstrated modest generalizability (AUC range: 0.62-0.83).

Conclusion: Existing prediction models demonstrated moderate predictive performance, often with modest generalizability when independently validated. Limitations of traditional methods have shown to impair predictive accuracy and resolution. Exploration of novel methods such as machine learning approaches may address these limitations for future school-age asthma prediction.

KEYWORDS

asthma, childhood, prediction model, risk scores, wheeze

Abbreviations: AUC, Area under the receiver operating curve; FeNO, Fractional exhaled nitric oxide; LASSO, Least absolute shrinkage and selection operator; LR−, Negative likelihood ratio; LR+, Positive likelihood ratio; NPV, Negative predictive value; PPV, Positive predictive value; PROBAST, Prediction model Risk Of Bias ASsessment Tool; RAST, Radio-allergosorbent test; SPT, Skin prick test.

Holloway and Rezwan equally contributed.

1 | INTRODUCTION

Asthma is the most common chronic disease in children.[1,2] The clinical presentation of childhood asthma is highly heterogeneous. While hallmark symptoms include wheeze, shortness of breath, cough and chest tightness, children may present with one or a combination of these symptoms, which may be intermittent or persistent.[3-7]

Asthma symptoms usually manifest in early life. However, in a large proportion of children, these symptoms are transient, often disappearing by school-age (6-13 years). For example, wheeze, the primary symptom observed in asthmatic children, affects half of all preschool children, of whom only one third go on to develop asthma.[8,9] In addition, a study of children enrolled onto the Tucson Children's Respiratory Study in the United States identified that 20% of school-age asthmatics were asymptomatic in early life.[8] As a result, it is difficult to predict which pre-schoolers will develop asthma later in childhood and whose symptoms will subside. Unsurprisingly, there is a window of uncertainty in clinical decision-making,[10] resulting in both under-diagnosis and over-diagnosis of probable asthmatic pre-schoolers.[11,12]

Prediction models which can distinguish true future asthmatics from a group of high-risk, symptomatic preschool children can assist physicians in providing early diagnoses and interventions. However, models which can also identify future asthmatics within a general population of pre-schoolers have the additional benefits of identifying late-onset asthmatics and stratifying individuals by asthma risk to subsequently promote asthma prevention among moderate/low-risk children. Besides being cost-effective, such strategies, as already demonstrated in other disease areas,[13-16] could promote personalized asthma care, limit unnecessary exposure to the adverse effects of asthma medications, and reduce the wastage of healthcare resources.[11,17]

To be of clinical value, the performance of any predictive tool needs to be reproducible in independent populations with comparable characteristics. Although several prediction models for childhood asthma exist, not all have been validated in independent populations. Surprisingly, none have yet been incorporated into clinical practice.[18-20]

1.1 | Objectives

This systematic review critically evaluates existing prediction models for school-age asthma development by assessing their predictive performance, statistical methodology and their potential clinical utility. Where relevant, external validation studies of these models were assessed. Finally, potential issues which might be responsible for the lack of clinical utility of existing asthma prediction models were identified and recommendations for future research priorities presented.

2 | METHODS

This systematic review (PROSPERO registration number: CRD42019146638) was conducted in accordance with the guidelines reported in the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) statement.[21]

Key Message

This study reviewed childhood asthma prediction models that have been developed and/or validated to date and identifies key methodological limitations which may account for their lack of current clinical utility. This critical evaluation informs physicians of the strengths and limitations of tools currently available to assist preschool asthma care management, and provides researchers with key recommendations for future studies developing clinically relevant childhood asthma prediction models.

2.1 | Search strategy

An electronic search of three databases, MEDLINE, Embase and Web of Science Core Collection, was performed on 26 July 2019. Free-text and MeSH terms were used to identify articles related to predictive modelling for childhood asthma (Table S1-S3).

All articles underwent a two-stage duplicate removal: first electronically using EndNote X8.2[22] followed by a manual removal of remaining duplicates. Two independent reviewers conducted a title and abstract screening to assess the relevance of the remaining articles. Discrepancies were resolved through discussion among the reviewers. A full-text screening and additional screening of citations in selected papers and reviews of prediction models for childhood asthma were conducted. Identified studies underwent data extraction and qualitative analysis.

2.2 | Study selection

Articles were included if they met the following criteria: the study detailed the development of a novel prediction model or updated a pre-existing model; the target population was children aged ≤5 years; the main prediction outcome was future childhood asthma or wheeze persistence at school-age (6-13 years old); and at least two risk predictors were used to construct the model. Models developed in both general and high-risk populations were considered. Validation studies that improved upon existing models were included. Studies that externally validated existing models in populations unrelated to that in which they were developed were also included.

Articles were excluded if a final prediction score was not developed or studies failed to report any performance measures for model evaluation. Conference papers, randomized control trials, proceedings, letters, editorials and non-English articles were excluded.

2.3 | Data extraction

Information on study design, candidate predictors, statistical methodology for model development and prediction outcome were collected from model derivation studies.

Model performance was evaluated using prediction measures of: discrimination, sensitivity, specificity, positive and negative predictive values (PPV and NPV, respectively) and positive and negative likelihood ratios (LR+ and LR−, respectively; Table 1). Where absent, likelihood ratios were calculated using reported sensitivity and specificity. Where applicable, performance measures were collected from both derivation and validation studies in order to assess model generalizability. The Prediction model Risk Of Bias ASsessment Tool (PROBAST) checklist[23] was used to critically appraise the risk of bias and applicability of each article.

3 | RESULTS

The literature search identified 4187 articles (Figure 1). Following the removal of 1204 duplicate articles, 2983 articles underwent title and abstract screening. The screening process identified 59 articles for full-text review. Of these, 25 studies were deemed relevant. An additional citation screening of relevant articles and the seven identified review papers on childhood asthma prediction tools identified a further three studies. These 28 studies were classified into two categories based on the methods used for developing the predictive models: regression-based (n = 20; Table S4) and machine learning approaches (n = 4; Table S5). The remaining four studies were external validations of previously developed models (Table 3).

3.1 | Regression-based models

Twenty-one regression-based prediction models were described in 20 studies (Table S4). Thirteen of the 21 models were novel while eight were modifications of existing models: six modified the Asthma Predictive Index (API)[24-29]; one updated the PIAMA risk score30 and one adapted the Obstructive Airway Disease (OAD) risk score.[31] Additionally, nine studies externally validated six prediction models, detailed within either developmental (n = 5) or independent validation studies (n = 4).[32-35]

3.1.1 | Target population

Of the 21 models carried forward for qualitative analysis (Table S4), six were developed in the general population[24,31,36-38] and 15 within high-risk populations, the latter restricting inclusion to children with a parental history of allergy/asthma (four models)[25,28-29,39] or asthma-like symptoms (11 models,[30,40] with nine specifically targeting children experiencing wheeze[26-27,41-47]). Only one model was derived based on predictors initially associated with childhood asthma within a low-income, Puerto Rican population.[37]

3.1.2 | Predictors

Thirty-eight different predictors were used among the 21 identified models, including seven variations of wheeze and two different measures for both allergic sensitization and pulmonary function (Table S6). The number of predictors used to construct the models ranged between 3 and 10. Twenty out of 38 predictors were each included in

TABLE 1 Definitions of the main measures used to evaluate prediction model performance

Performance measures	Definition
Calibration	How well the model's predictions compare to the observed outcomes (goodness of fit)
Discrimination	How well the model distinguishes between those with and without the disease, measured by the area under the receiver operating curve (AUC)
Sensitivity	The proportion of individuals with the disease who are correctly predicted to have the disease
Specificity	The proportion of individuals without the disease who are correctly predicted as disease-free
Positive predictive value (PPV)	The proportion of individuals with a positive disease prediction who truly have the disease
Negative predictive value (NPV)	The proportion of individuals with a negative disease prediction who are truly disease-free
Positive likelihood ratio (LR+)	The ratio of true positive predictions against false positive predictions which indicates a model's ability to rule in disease
Negative likelihood ratio (LR−)	The ratio of false negative predictions against true negative predictions which indicates a model's ability to rule out disease

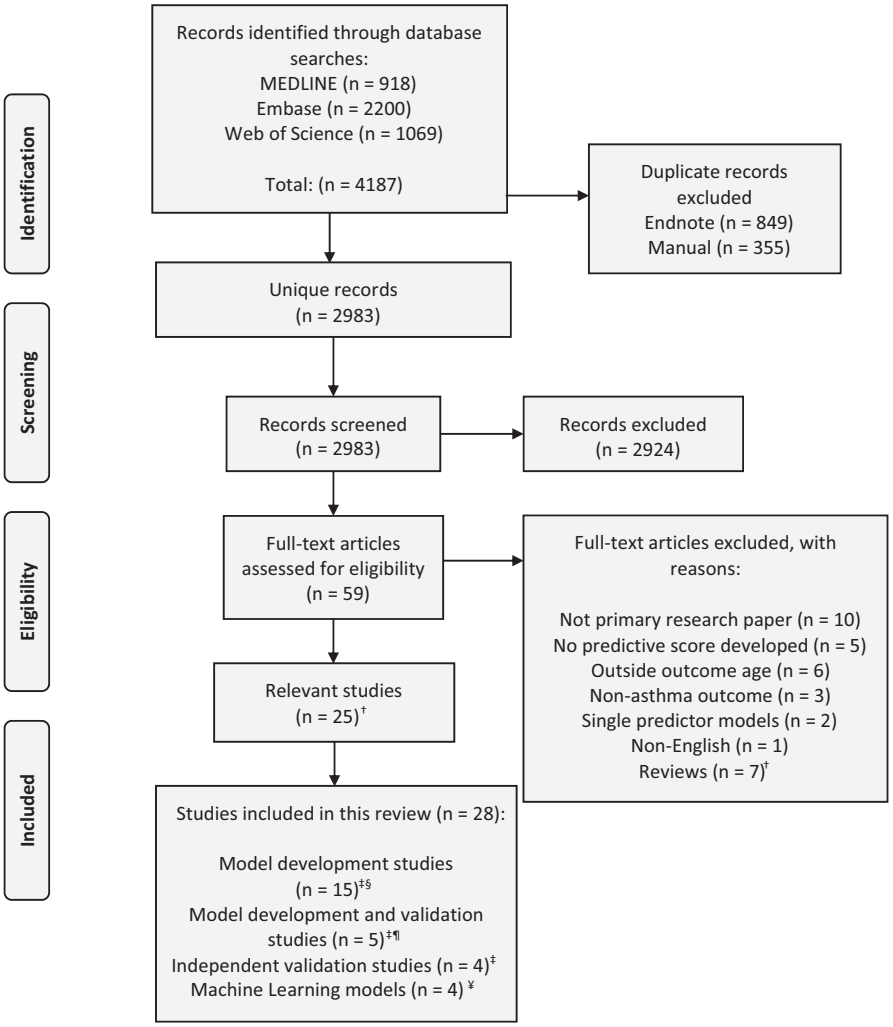

FIGURE 1 PRISMA flow diagram of study search strategy. †Citation screening of articles identified three additional studies. ‡Included in the final qualitative analysis. §One study transformed a diagnostic model into a prediction model upon external validation (considered a developmental study in this review). ¶Validated the developmental study model (n = 2) or an existing model (n = 3). ¥Excluded from the main qualitative analysis

just one of the 21 models (last column, Table S6). For example, familial pollen allergy was a predictor in RAST alone, while race was only included in PARS. A history of parental asthma and personal eczema were the most frequently used predictors of childhood asthma, each incorporated into 14 models. Three studies used data only available in early life (≤2 years)[31,36,43] while another only used predictor data collected at birth.[38] Predictor information was mainly collected from parent-reported questionnaires or standard clinical assessments. Sixteen models required data from additional clinical tests such as blood or skin prick tests (SPT) to assess allergic sensitization status (14 models); measures of pulmonary function (two models); biomarkers of volatile organic compounds in exhaled breath condensate (one model); and gene expression in peripheral blood (one model).

3.1.3 | Outcome

The prediction outcome in most studies (19/20) was school-age asthma, yet nine different definitions of asthma were used (Table 2). Seventeen studies included asthma-like symptoms, twelve included a doctor diagnosis, and nine incorporated objective pulmonary tests as components in their asthma definition. One study used persistent wheeze determined through the frequency of wheezing episodes as the prediction outcome.[41] The most common definition (in 5/20 studies) specified a combination of asthma-like symptoms, use of asthma medications and/or objective respiratory tests. All studies identified a child's asthma status by evaluating the outcome criteria within the last 12 months except one which evaluated the asthma criteria across two consecutive years.[42]

3.1.4 | Model construction

The API and its modifications are clinical indices requiring a combination of major and minor criteria to be met. The other prediction models are weighted scoring systems based on derivations of each predictor's regression coefficients, with the exception of two unweighted scoring systems.[37,41]

3.1.5 | Performance measures

Three studies failed to report any model performance measures detailed in Table 1. Of these, the modified Asthma Predictive Index

TABLE 2 Nine main classes of asthma definitions used among asthma prediction model developmental studies

Asthma outcome definitions	Number of studies	Study reference
1. Doctor diagnosis only	1	[26]
2. Symptoms only	1	[41]
3. Doctor diagnosis and symptoms	4	[24,30,37,47]
4. Doctor diagnosis and medication	2[a]	[28]
5. Symptoms and medication	2	[44,46]
6. Doctor diagnosis, symptoms and medication	1	[42]
7. Symptoms, medication and lung function tests[b]	5	[27d,e], [25d,e], [40e], [45d,e], [38c]
8. Doctor diagnosis, symptoms and lung function tests[b]	1	[39c,e]
9. Doctor diagnosis, symptoms, medication and lung function tests[b]	3	[36c], [31c], [43d]

[a]The asthma outcome for the mAPI was extracted from the m[2]API study which evaluated the model's performance.

[b]Lung function tests comprised of one or a combination of exercise tests ([c]), spirometry assessing reversibility to bronchodilators ([d]) and bronchial hyper-responsiveness to methacholine or histamine ([e]).

TABLE 3 Model performance of externally validated asthma prediction models

	Author	Population geography	Risk group	Variation in predictors	Variation in outcome	Study size (prevalence, %)	Study asthma prevalence (%)
Loose API	Castro-Rodriguez et al[24]	USA	General population			986	57.1
	Rodriguez-Martinez et al[32]	Colombia	High risk	-	-	93	22.5
	Leonardi et al[35]	UK	General population	✓	-	1731	11.5
						1291	10.5
	Devulapalli et al[36]	Norway	General population	✓	✓	459	21.1
Stringent API	Castro-Rodríguez et al[24]	USA	General population			1002	57.1
	Rodriguez-Martinez et al[32]	Colombia	High risk	-	-	93	22.5
	Leonardi et al35	UK	General population	✓	-	1683	11.5
						1257	10.5
	Caudri et al[42]	Netherlands	High risk	✓	✓	1177	11.7
	Devulapalli et al[36]	Norway	General population	✓	✓	459	21.1
PIAMA	Caudri et al[42]	Netherlands	High risk			2171	11.1
	Hafkamp-de Groen et al[30]	Netherlands	High risk	✓	✓	2877	6.0
	Rodriguez-Martinez et al[32]	Colombia	High risk	-	✓	123	53.6
PARC	Pescatore et al[44]	UK	High risk			1226	28.1
	Grabenhenrich et al[33]	Germany	High risk	✓	-	140	20.0
	Pedersen et al[34]	UK	High risk	✓	-	2690	14.0
PAPS	Vial Dupuy et al[43]	France	High risk			200	47.5
	Vial Dupuy et al[43]	France	High risk	-	-	227	18.9
PARS	Biagini Myers et al[39]	USA	High risk			589	16.1
	Biagini Myers et al[39]	UK	General population	✓	✓	1098	-

(Continues)

(mAPI), developed within a randomized clinical trial protocol, did not evaluate the model's performance.[29] Performance measures for the mAPI were extracted from Chang et al's study[28] which evaluated and compared the mAPI to another modified API (m^2API).[28] The other two studies only reported single performance measures of population attributable risk[37] and Nagelkerke R^2.[31]

Discriminative ability was reported for 12 models and ranged between 0.66 and 0.87. Sixteen models reported sensitivity (range: 15.7%-88%) and specificity (range: 62.3%-99%). PPV and NPV were reported for 15 models, ranging between 12.4%-90% and 68.3%-97.2%, respectively. Likelihood ratios were reported for eight models and were derived for an additional eight models using reported sensitivity and specificity. The ability to rule in disease (LR+) ranged from 1.94 to 21 while the ability to rule out disease (LR−) ranged from 0.13 to 0.87.

3.1.6 | Validation

Nine studies performed external validation: four validated the loose and/or stringent API, two validated PIAMA and PARC while PAPS and PARS were each validated once (Table 3). Upon validation, most models demonstrated a trade-off between improvements in sensitivity at the expense of specificity, resulting in increased false positive predictions and a decline in PPV and LR+ estimates compared to their derivation models. While the PARS model showed comparable performance upon validation, only the PARC model demonstrated superior performance, with improvement in LR+ (2.47 vs 2.63) and AUC (0.74 vs 0.83) compared to the derivation model.

3.1.7 | Critical appraisal

The overall risk of bias was deemed high for all 21 models due to: (a) predictor and outcome bias (21 and 17 models, respectively), predominantly due to the subjective interpretation of their definitions, particularly those based on parent-reported information; and (b) biased analysis due to an inappropriate number of candidate predictors, inappropriate handling of missing data, failure in reporting performance measures (eg calibration) or failure in treating models for potential overfitting or performance optimization as detailed in

TABLE 3 (Continued)

Target age	Prediction age	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	LR+	LR−	Discrimination
≤3	6-13	41.6	84.7	59.1	73.2	2.72[a]	0.69[a]	-
1-3	5-6	71.4	33.3	23.8	80	1.07	0.86	-
2-3	7	57	80	26	94	2.85[a]	0.54[a]	0.68
2-3	10	57	81	25	94	3.00[a]	0.53[a]	0.69
3	10	59.8	79	43.9	87.7	2.85[a]	0.51[a]	-
≤3	6-13	15.7	97.4	76.6	68.3	6.04[a]	0.87[a]	-
1-3	5-6	42.9	79.2	37.5	82.6	2.06	0.72	-
2-3	7	37	93	40	93	5.29[a]	0.68[a]	0.65
2-3	10	32	94	35	92	5.33[a]	0.72[a]	0.63
0-4	7-8	20	92	25	90	2.50[a]	0.87[a]	0.62
3	10	56.7	83	47.8	87.4	3.34[a]	0.52[a]	-
0.4	7-8	19	97	42	91	6.33[a]	0.84[a]	0.74
1-4	6	-	-	-	-	-	-	0.74
1-3	5-6	54.5	78.9	75.0	60	2.59	0.58	-
1-3	6-8	72	71	49	86	2.47	0.40	0.74
3	8	82	69	40	94	2.63	0.26	0.83
1.5-3.5	7.5	69	76	32	94	2.87	0.41	0.77
<2	6	42.4	89.6	66.7	75.9	4.06	0.64	0.66
<2	13	62.8	67.4	31	88.6	1.93[a]	0.55[a]	0.65
≤3	7	68	77	37	93	3.02	0.41	0.80
2	10	67	79	36	93	3.25	0.41	0.79

Note: Shaded rows: prediction models as reported in the developmental studies; unshaded rows: external validation studies. ✓Used altered definitions in the external validation study compared to the original developmental study: predictors = exclusions or surrogate variables used; outcome = variation in components used to determine asthma.

Abbreviations: LR−, negative likelihood ratio; LR+, positive likelihood ratio; NPV, negative predictive value; PPV, positive predictive value.

[a]Likelihood ratios calculated based on reported sensitivity and specificity as: LR+ = sensitivity/(1 − specificity), LR− = (1 − sensitivity)/specificity.

TABLE 4 Critical appraisal of each study's risk of bias and applicability using the PROBAST checklist

	Loose API[24]	Stringent API[24]	mAPI[29]	m²API28	API + FeNO[26]	API + biomarkers[27]	ucAPI[25]	IOW[41]	RAST[40]	OAD[36]
Risk of bias										
Participants	L	L	H	H	H	H	H	H	H	L
Predictors	H	H	H	H	H	H	H	H	H	H
Outcome	H	H	H	H	H	L	H	H	H	L
Analysis	H	H	H	H	H	H	H	H	H	H
Overall risk	H	H	H	H	H	H	H	H	H	H
Concern regarding applicability										
Participants	L	L	H	H	H	H	H	H	H	L
Predictors	L	L	L	L	L	H	L	L	L	L
Outcome	L	L	L	L	L	L	L	L	L	L
Overall risk	L	L	H	H	H	H	H	H	H	L

(Continues)

the PROBAST checklist (Table 4). The 15 studies which used high-risk developmental populations presented with low risk of bias (assuming their intended use in settings similar to their developmental study) but high concern regarding applicability to a general population.

3.2 | Machine learning approaches

Four studies which utilized machine learning approaches to develop five prediction models for childhood asthma within a paediatric hospital population of diagnosed asthma patients were identified.[48-51] These studies presented with ambiguity in their study design with regard to unclear predictor definitions, time points of predictor measurements and population characteristics. Additionally, due to limitations of using an asthma diagnosis as a predictor, the small study size for machine learning applications, and signs of overfitting in the reported results, these studies were excluded from the main qualitative analysis. However, they are included in this review to highlight novel methodologies currently being explored for childhood asthma prediction (Table S5).

4 | DISCUSSION

This review identified 26 prediction models for predicting childhood asthma at school-age but none have been widely implemented into standard clinical practice. Only the API is mentioned in asthma management guidelines[4] and has been utilized with caution (upon modification), in the recruitment of participants into clinical trials.[29] Against this background, a critical evaluation of these studies aimed to identify potential problems surrounding the lack of applicability of these models. The key issues centred on: the choice of population for model derivation and/or validation, predictor and outcome definitions, methodologies employed for predictor selection, methods of data collection, study power and the interpretability of models.

4.1 | Choice of population

The performance of any predictive model is highly dependent on its developmental setting and may not generalize well in alternative risk populations. Fifteen of the twenty-one regression-based models were developed in high-risk populations. High-risk populations, which have a higher asthma prevalence compared to the general population, are commonly used for model development in the hope of increasing the power for predictor selection and the detection of true asthmatics. However, such models may overestimate asthma risk within the general population. At present, only PARS has assessed this and was able to show comparable predictive performance in high-risk and general populations. In contrast, the loose and stringent API, developed in a general population, demonstrated a substantial improvement in sensitivity, although at the cost of increasing false positive predictions, when validated in high-risk populations (Table 3).

4.2 | Population-specific predictors

Most models were developed in European/predominantly Caucasian cohorts. Exposures specific to less developed countries, such as poverty and pollution, are typically not considered as important predictors of asthma in these models due to inadequate representation of such populations in the study cohorts.[52] For example, Szentpetery et al initially developed a diagnostic model, identifying

TABLE 4 (Continued)

OAD + IgE[31]	PIAMA[42]	Updated PIAMA[30]	Lødrup Carlsen et al[38]	PAPS[43]	PARC[44]	CAPS[45]	Boersma et al[46]	Szentpetery et al[37]	MAAS-APT[47]	PARS[39]
L	H	H	L	H	H	H	H	L	H	H
H	H	H	H	H	H	H	H	H	H	H
L	H	H	H	U	H	H	H	H	H	H
H	H	H	H	H	H	H	H	H	H	H
H	H	H	H	H	H	H	H	H	H	H
L	H	H	L	H	H	H	H	L	H	H
L	L	L	L	L	L	L	L	L	L	L
L	L	L	L	L	L	L	L	L	L	L
L	H	H	L	H	H	H	H	L	H	H

Note: Risk of bias and applicability were assessed as: H = High risk, L = Low risk, U = Unclear risk using the criteria outlined in the PROBAST checklist.[23] For each domain, the risk of bias or concern of applicability is considered: high—if ≥1 signalling question in the PROBAST critical appraisal criteria were answered "no" or "probably no"; low—if the answer to the signalling questions were all "yes"; unclear—if relevant information was missing to answer the signalling question and none of the signalling questions were answered "no". The overall risk of bias and applicability were deemed low if all domains were evaluated as low risk, high risk if ≥1 domain was considered high-risk, and unclear if ≥1 domain was considered unclear and all other domains were low-risk.

gun violence and an unhealthy diet as predictors of childhood asthma in a Puerto Rican population. However, when validated as a prediction model in a Swedish cohort, data for these two predictors were unavailable, potentially due to low concern for these risk factors in this population, and were excluded from the model.[37]

4.3 | Prediction window

Due to the transient nature of asthma-like symptoms in early life, the evaluation of clinical predictors from 4-5 years of age is more predictive of school-age asthma.[8] However, for prediction models developed with the intention of preventing asthma development rather than targeting children for early therapeutic intervention, predictions made at 4-5 years may already be too late. Four models used predictor data available before age 2[31,36,38] but only one was externally validated.[43] Lødrup Carlsen et al's model[38] only used predictor data collected at birth; however, the need to perform neonatal lung function tests (rarely conducted outside of a research setting) greatly impairs its potential clinical applicability.

4.4 | Data collection

Most studies collected predictor information through parent-completed questionnaires, a method prone to recall bias and misclassifications. A recent study identified that one third of parents change their answer after watching a recording of wheeze.[53] Such under/overestimations of parent-reported predictors can result in poor model performance compared to models using data collected from physicians, healthcare records or objective measurements.

4.5 | Predictor availability

Thirty-eight different predictors indicative of well-documented asthma risk factors were used across the 21 regression-based models. This variation reflects the inherent heterogeneity of childhood asthma across different populations and variability in predictor availability between studies. Sixteen models required additional clinical tests, most commonly blood and skin prick tests (SPT) to determine a child's atopic status. These tests were the main amendment in four of the seven modified prediction models. Four other studies demonstrated that the addition of IgE as a predictor in their models improved predictive power compared to their models without IgE.[31,40,45,46] One modification of the API included biomarkers of volatile organic compounds in exhaled breath condensate and gene expression[27]; despite ranking second in terms of AUC (AUC = 0.86, unboot-strapped AUC = 0.95), the use of this model is unlikely to be feasible outside of a specialist/research setting. Models developed with predictors which are not readily available, or which require the use of additional healthcare resources, can be limited in their generalizability and potential clinical implementation.

4.6 | Predictor selection

Methodology for the selection of predictors varied between the 20 regression-based studies. Models used either a priori knowledge,[24,28,29] univariate analysis,[24] multivariate regression analysis[26,30] or a combination of univariate and multivariate regression.[25,41-42,46,47] Despite the latter two-stage combination

approach being an established method used across biomedical research, this method can introduce significant bias to the feature selection process due to inconsistencies between univariate and subsequent multivariate analyses.[54,55] To address this, some studies adopted a stepwise backward or forward selection multivariate regression approach,[27,37-40,45] and the PARC model[56] utilized LASSO (least absolute shrinkage and selection operator).[57] However, none of these studies address the issue of multicollinearity between candidate predictors which can introduce noise and subsequently reduce model performance. Among the four machine learning studies identified, supervised and unsupervised machine learning algorithms were used for feature selection.[48-51] Indeed, machine learning algorithms, particularly those such as random forest, recursive feature elimination and genetic algorithms, are more robust in handling the relatedness between predictors and may promote better predictor selection compared to regression-based methods.[57,58]

4.7 | Outcome

Nine asthma outcome definitions were used across the 20 regression-based studies. This may have led to an artificial variation in the prevalence of asthma across studies influencing the construction, optimization and subsequent performance of predictive models. Childhood asthma is often considered an umbrella term describing a syndrome of different respiratory symptoms.[3] As a result, models developed to predict childhood asthma are predicting a subjective entity. A consensus on an objective definition acceptable to the clinical and research community is essential.

4.8 | Study power

Upon critical appraisal, at least eight studies were identified as lacking sufficient power to develop stable prediction models; these studies had a ratio of candidate predictors to total number of cases lower than recommended (at least 20 cases per candidate predictor) to achieve sufficient power.[30,38,41-42,44-45,47] Underpowered studies risk important predictors not being selected (under-fitting—Type II error), the incorrect selection of predictors (overfitting—Type I error) as well as the misrepresentation of the associated directionality between predictors and the outcome.[59]

Compared to traditional regression methods, machine learning approaches possess superior power and resolution for pattern recognition. By allowing a larger number of candidate predictors to be considered and being more robust to the relatedness between predictors, there is potential to identify novel predictors and exclude redundant predictors which may have been previously overlooked by traditional predictor selection approaches.[60-6260-62] Despite the potential benefits offered by machine learning methods, the four machine learning studies reported to date remain underpowered.[48-51] Further studies are necessary to determine whether machine learning approaches can develop better performing asthma prediction models over regression-based methods.

4.9 | Validation

Models tend to perform best within their developmental population. External validation studies, which assess the true performance of models in independent populations, are essential to assess the generalizability of a model. However, only six of the 21 identified regression-based models were externally validated. None of the five machine learning models were externally validated (Table 3). While the PARS and PARC models demonstrated comparable performance when validated, the other models demonstrated poorer predictive performance, particularly in terms of PPV and likelihood ratios. This may be due to inconsistencies between the derivation and validation study designs, mainly with regard to the predictor/outcome definitions and the exclusion or use of surrogate variables for unavailable predictor information (Table 3). Validation of all existing models within a single independent population using a single outcome definition is necessary to standardize inconsistencies in study design and population effect to facilitate a comparative analysis between models. However, this remains difficult in practice due to the need for a reference population of sufficient size with data available for all 38 predictors.

4.10 | Interpretability

At present, a quantitative evaluation of the performance of existing models is difficult as not all studies report the standard performance measures listed in Table 1. Discrimination (AUC) is often used to compare the overall performance between models, with a discriminative threshold of 0.80 considered to identify a very good predictive model.[63] Three developmental models reached this threshold but only one, PARS, was externally validated. The good generalizability of PARS (AUC = 0.79) has facilitated its transformation into an online interactive tool and mobile app for use by both physicians and parents.[39]

However, using discrimination alone to compare model performance is inappropriate as models with similar AUC can show large variations in sensitivity and specificity. There is a clear trade-off between optimizing both of these performance measures, with no one model able to achieve both high sensitivity and specificity. Therefore, clear aims of whether a model intends to optimize towards higher sensitivity or specificity for the future application of prevention or asthma symptom management, respectively, would benefit the evaluation of a model's predictive power and viability.[63]

Finally, the API and its modifications provide a dichotomous outcome of asthma risk, while the remaining regression-based models present asthma risk across a range of potential scores, often stratifying individuals into groups of low, medium or high risk. However, physicians are already able to make similar predictions

upon clinical assessment which may explain the lack of clinical uptake of existing models. The exploration of novel approaches, such as machine learning, for the development of prediction models with greater probabilistic resolution of an individual's asthma risk is warranted.

Yet, existing prediction models are not redundant—the use of well-performing, externally validated models should be considered for use in clinical trials to support the stratification of participants for inclusion or treatment allocation. These models are likely to offer superior predictions compared to trials currently utilizing the API[29] or, more frequently, parental history, to assess asthma risk.

5 | CONCLUSIONS AND FUTURE RECOMMENDATIONS

Based on the findings of this review, a number of key considerations are needed for the development of future prediction models.

5.1 | Study design and data availability

Improving model generalizability across all population settings could be achieved by standardizing predictor and outcome definitions across settings, and addressing issues of population bias and data availability. While the perfect solution would be to establish a single, general population, prospective cohort of sufficient size for model development with an independent reference population for validation, this is unrealistic.

Instead, studies should specify and closely match the developmental population of the model for its future application. Data should be collected using objective measurements and high-quality, standardized questionnaires with unambiguous descriptions which are consistent across both clinical and research settings. Where parental-reported data are used, clinical jargon should be deconstructed and/or be supported by auditory or visual aids to minimize recall bias and misclassification wherever possible.

In addition, only easily derivable and commonly available clinical predictors should be used. While biomarkers can have high predictive power, their predictive benefit needs to be measured against the cost of test availability across different healthcare settings, patient/physician time and demand on healthcare resources. Yet, the exploration and identification of novel biomarkers, particularly in early life, may encourage the transition from asthma management to prevention.

5.2 | Isolating predictors for model development

Due to the heterogeneity of childhood asthma, a number of candidate predictors have been associated with childhood asthma. One approach to identifying predictors for model development is to isolate a subset of the most frequently used predictors from previous studies. For example, parental asthma, eczema, wheeze without cold, specific IgE, frequent wheeze, allergic rhinitis and sex have been used in at least a quarter of existing models. However, as previously discussed, population-specific influences and predictor selection methodological limitations exist in these studies. A better approach would be for future studies to utilize a robust predictor selection method (such as recursive feature elimination), which is sufficiently powered and able to address the multicollinearity between predictors, in order to distinguish strong predictors from redundant variables within their specific population.

5.3 | Model development methodologies

The majority of existing studies have utilized regression-based methods and have developed a number of similar prediction models, few generalizing well in independent populations, and none widely implemented into clinical practice. Alternative methods such as machine learning approaches have advantages over these statistical methods as already discussed, particularly with regard to addressing frequently overlooked concerns of predictor relatedness, distinguishing between predictive and redundant predictors, and improving the resolution of predictions. Such methods have not been adequately implemented; hence, future studies using robust study designs are needed to assess their potential benefits for childhood asthma prediction.

Finally, it is crucial for any developed model to undergo external validation within a population similar to its future application. Non-validated models are not clinically useful and are largely limited as exploratory studies. Reporting of all standard performance measures for both development and validation is necessary to evaluate a model's generalizability and subsequently promote its clinical application for predicting school-age asthma.

ACKNOWLEDGMENTS

The authors gratefully acknowledge Ms Paula Sands, research engagement librarian at University of Southampton, for her guidance and assistance in conducting the systematic review database searches.

CONFLICT OF INTEREST

The authors have no funding relationships or conflicts of interest related to this article to disclose.

AUTHOR CONTRIBUTIONS

DMK, FIR, JWH and SHA: Conceptualization and design. DMK, VBNW, MAK, LK and FIR: Systematic review screening. DMK and LK: Critical appraisal. DMK: Finding analysis. DMK: Writing—original draft. All authors: Manuscript revision and final approval.

ORCID

Dilini M. Kothalawala https://orcid.org/0000-0002-5804-0457
Latha Kadalayil https://orcid.org/0000-0002-3757-5487

Veronique B. N. Weiss https://orcid.org/0000-0003-4116-950X
Mohammed Aref Kyyaly https://orcid.org/0000-0002-1684-9207
Syed Hasan Arshad https://orcid.org/0000-0001-5988-235X
John W. Holloway https://orcid.org/0000-0001-9998-0464
Faisal I. Rezwan https://orcid.org/0000-0001-9921-222X

REFERENCES

1. Akdis CA, Agache I. *Global Atlas of Asthma*. Zürich: European Academy of Allergy and Clinical Immunology; 2013.
2. Licari A, Castagnoli R, Brambilla I, et al. Asthma endotyping and biomarkers in childhood asthma. *Pediatr Allergy Immunol Pulmonol*. 2018;31(2):44-55.
3. Pavord ID, Beasley R, Agusti A, et al. After asthma: redefining airways diseases. *Lancet*. 2018;391(10118):350-400.
4. Global Initiative for Asthma (GINA). Global strategy for asthma management and prevention. 2018. Available from: https://www.ginasthma.org
5. Carr TF, Bleecker E. Asthma heterogeneity and severity. *World Allergy Organ J*. 2016;9(1):41.
6. Bacharier LB, Boner A, Carlsen KH, et al. Diagnosis and treatment of asthma in childhood: a PRACTALL consensus report. *Allergy*. 2008;63(1):5-34.
7. Wenzel SE. Asthma phenotypes: the evolution from clinical to molecular approaches. *Nat Med*. 2012;18(5):716-725.
8. Martinez FDWA, 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(3):133-138.
9. 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(2):325-331.
10. Bacharier LB, Guilbert TW. Diagnosis and management of early asthma in preschool-aged children. *J Allergy Clin Immunol*. 2012;130(2):287-296; quiz 297-288.
11. Annesi-Maesano I, Sterlin C, Caillaud D, et al. Factors related to under-diagnosis and undertreatment of childhood asthma in metropolitan France. *Multidiscip Respir Med*. 2012;7(24). https://doi.org/10.1186/2049-6958-7-24
12. Bush A, Fleming L, Saglani S. Severe asthma in children. *Respirology*. 2017;22(5):886-897.
13. Noble D, Mathur R, Dent T, Meads C, Greenhalgh T. Risk models and scores for type 2 diabetes: systematic review. *BMJ*. 2011;343:d7163.
14. Vogenberg FR. Predictive and prognostic models: implications for healthcare decision-making in a modern recession. *Am Health Drug Benefits*. 2009;2:218-222.
15. Oh SM, Stefani KM, Kim HC. Development and application of chronic disease risk prediction models. *Yonsei Med J*. 2014;55(4):853-860.
16. Lloyd-Jones DM. Cardiovascular risk prediction: basic concepts, current status, and future directions. *Circulation*. 2010;121(15):1768-1777.
17. Caudri D. *Asthma and wheezing in childhood: perinatal risk factors and early detection* [Ph.D. thesis], Erasmus University Rotterdam; 2010.
18. Smit HA, Pinart M, Antó JM, et al. Childhood asthma prediction models: a systematic review. *Lancet Respir Med*. 2015;3(12):973-984.
19. Luo G, Nkoy FL, Stone BL, Schmick D, Johnson MD. A systematic review of predictive models for asthma development in children. *BMC Med Inform Decis Mak*. 2015;15(99). https://doi.org/10.1186/s12911-015-0224-9
20. Colicino S, Munblit D, Minelli C, Custovic A, Cullinan P. Validation of childhood asthma predictive tools: a systematic review. *Clin Exp Allergy*. 2019;49(4):410-418.
21. Moher D, Shamseer L, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev*. 2015;4:1.
22. Reuters T. EndNote™ X8. 2017.
23. Moons KGM, Wolff RF, Riley RD, et al. PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. *Ann Intern Med*. 2019;170(1):W1-W33.
24. Castro-Rodríguez JAHC, 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(4):1403-1406.
25. Amin P, Levin L, Epstein T, et al. Optimum predictors of childhood asthma: persistent wheeze or the Asthma Predictive Index? *J Allergy Clin Immunol Pract*. 2014;2(6):709-715.
26. Singer F, Luchsinger I, Inci D, et al. Exhaled nitric oxide in symptomatic children at preschool age predicts later asthma. *Allergy*. 2013;68(4):531-538.
27. Klaassen EMM, van de Kant KDG, Jöbsis Q, et al. Exhaled biomarkers and gene expression at preschool age improve asthma prediction at 6 years of age. *Am J Respir Crit Care Med*. 2015;191(2):201-207.
28. Chang TS, Lemanske RF, Guilbert TW, et al. Evaluation of the modified asthma predictive index in high-risk preschool children. *J Allergy Clin Immunol Pract*. 2013;1(2):152-156.
29. Guilbert TW, Morgan WJ, Zeiger RS, et al. Atopic characteristics of children with recurrent wheezing at high risk for the development of childhood asthma. *J Allergy Clin Immunol*. 2004;114(6):1282-1287.
30. Hafkamp-de Groen E, Lingsma HF, Caudri D, et al. Predicting asthma in preschool children with asthma-like symptoms: validating and updating the PIAMA risk score. *J Allergy Clin Immunol*. 2013;132(6):1303-1310.e1306.
31. Lødrup Carlsen KC, Söderström L, Mowinckel P, et al. Asthma prediction in school children; the value of combined IgE-antibodies and obstructive airways disease severity score. *Allergy*. 2010;65:1134-1140.
32. Rodriguez-Martinez CE, Sossa-Briceno MP, Castro-Rodriguez JA. Discriminative properties of two predictive indices for asthma diagnosis in a sample of preschoolers with recurrent wheezing. *Pediatr Pulmonol*. 2011;46(12):1175-1181.
33. Grabenhenrich LB, Reich A, Fischer F, et al. The novel 10-item asthma prediction tool: external validation in the German MAS birth cohort. *PLoS ONE*. 2014;9(12):e115852.
34. Pedersen ESL, Spycher BD, de Jong CCM, et al. A simple 10-iten PARC tool to predict childhood asthma - an external validation. *J Allergy Clin Immunol Pract*. 2019;7(3):943-953.e944.
35. 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(6):1466-1472.e1466.
36. Devulapalli CS, Carlsen KCL, Haland G, et al. Severity of obstructive airways disease by age 2 years predicts asthma at 10 years of age. *Thorax*. 2008;63(1):8-13.
37. Szentpetery SS, Gruzieva O, Forno E, et al. Combined effects of multiple risk factors on asthma in school-aged children. *Respir Med*. 2017;133:16-21.
38. Lødrup Carlsen KC, Mowinckel P, Granum B, Carlsen KH. Can childhood asthma be predicted at birth? *Clin Exp Allergy*. 2010;40(12):1767-1775.
39. Biagini Myers JM, Schauberger E, He H, et al. A Pediatric Asthma Risk Score to better predict asthma development in young children. *J Allergy Clin Immunol*. 2018;143(5):1803-1810.e1802.
40. Eysink PE, ter Riet G, Aalberse RC, 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(511):125-131.

41. Kurukulaaratchy RJ, Matthews S, Holgate ST, Arshad SH. Predicting persistent disease among children who wheeze during early life. *Eur Respir J*. 2003;22(5):767-771.
42. Caudri D, Wijga A, A. Schipper CM, et al. Predicting the long-term prognosis of children with symptoms suggestive of asthma at preschool age. *J Allergy Clin Immunol*. 2009;124(5):903-910.e907.
43. Vial Dupuy A, Amat F, Pereira B, Labbe A, Just J. A simple tool to identify infants at high risk of mild to severe childhood asthma: the persistent asthma predictive score. *J Asthma*. 2011;48(10):1015-1021.
44. Pescatore AM, Dogaru CM, Duembgen L, et al. A simple asthma prediction tool for preschool children with wheeze or cough. *J Allergy Clin Immunol*. 2014;133(1):111-118.e113.
45. van der Mark LB, van Wonderen KE, Mohrs J, van Aalderen WMC, ter Riet G, Bindels PJE. Predicting asthma in preschool children at high risk presenting in primary care: development of a clinical asthma prediction score. *Prim Care Respir J*. 2014;23(1):52-59.
46. Boersma NA, Meijneke RWH, Kelder JC, van der Ent CK, Balemans WAF. Sensitization predicts asthma development among wheezing toddlers in secondary healthcare. *Pediatr Pulmonol*. 2017;52(6):729-736.
47. Wang R, Simpson A, Custovic A, Foden P, Belgrave D, Murray CS. Individual risk assessment tool for school-age asthma prediction in UK birth cohort. *Clin Exp Allergy*. 2019;49(3):292-298.
48. Chatzimichail EA, Rigas AG, Paraskakis EN. An artificial intelligence technique for the prediction of persistent asthma in children. *Proceedings of the 10th IEEE International Conference on Information Technology and Applications in Biomedicine*; 2010.
49. Chatzimichail E, Paraskakis E, Rigas A. An evolutionary two-objective genetic algorithm for asthma prediction. *2013 UKSim 15th International Conference on Computer Modelling and Simulation*; 2013.
50. Chatzimichail E, Paraskakis E, Sitzimi M, Rigas A. An intelligent system approach for asthma prediction in symptomatic preschool children. *Comput Math Methods Med*. 2013;2013:1-6.
51. Chatzimichail E, Paraskakis E, Rigas A. Predicting asthma outcome using partial least square regression and artificial neural networks. *Adv Artif Intell*. 2013;2013:1-7.
52. Aligne CA, Auinger P, Byrd RS, Weitzman M. Risk factors for pediatric asthma contributions of poverty, race, and urban residence. *Am J Respir Crit Care Med*. 2000;162:873-877.
53. Hartwig IRV, Pincus MK, Diemert A, Hecher K, Arck PC. Sex-specific effect of first-trimester maternal progesterone on birthweight. *Hum Reprod*. 2013;28(1):77-86.
54. Ranganathan P, Pramesh CS, Aggarwal R. Common pitfalls in statistical analysis: logistic regression. *Perspect Clin Res*. 2017;8(3):148-151.
55. Wang H, Peng J, Wang B, et al. Inconsistency between univariate and multiple logistic regressions. *Shanghai Arch Psychiatry*. 2017;29(2):124-128.
56. Pescatore AM, Spycher BD, Jurca M, Gaillard EA, Kuehni CE. Environmental and socioeconomic data do not improve the Predicting Asthma Risk in Children (PARC) tool. *J Allergy Clin Immunol*. 2015;135(5):1395-1397.e1391-1393.
57. Cai J, Luo J, Wang S, Yang S. Feature selection in machine learning: a new perspective. *Neurocomputing*. 2018;300:70-79.
58. Jain D, Singh V. Feature selection and classification systems for chronic disease prediction: a review. *Egypt Inform J*. 2018;19(3):179-189.
59. Valero-Moreno S, Perez-Marin M, Montoya-Castilla I, Castillo-Corullon S, Ramirez-Aguilar A, Escribano-Montaner A. Emotional distress in family caregivers of adolescents with bronchial asthma: analysis of its predictors. *Arch Argent Pediatr*. 2018;116(2):e234-e240.
60. Singal AG, Mukherjee A, Elmunzer BJ, et al. Machine learning algorithms outperform conventional regression models in predicting development of hepatocellular carcinoma. *Am J Gastroenterol*. 2013;108(11):1723-1730.
61. Waljee AK, Higgins PD, Singal AG. A primer on predictive models. *Clin Transl Gastroenterol*. 2014;5:e44.
62. James G, Witten D, Hastie T, Tibshiran R. *An Introduction to Statistical Learning*, 1st edn. New York: Springer-Verlag; 2013.
63. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*. 2010;21(1):128-138.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Kothalawala DM, Kadalayil L, Weiss VBN, et al. Prediction models for childhood asthma: A systematic review. *Pediatr Allergy Immunol*. 2020;31:616–627. https://doi.org/10.1111/pai.13247