Prediction Models for Childhood Asthma: A Systematic Review

Authors

Dilini M. Kothalawala, BSc[a,b], Latha Kadalayil, PhD[a], Veronique B.N. Weiss[a], M. Aref. Kyyaly, PhD[c,d], Hasan S. Arshad, DM[b,c,d], John W. Holloway, PhD[*a,b], Faisal I. Rezwan, PhD[*a,e]

Institutional affiliations

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

[b]NIHR Southampton Biomedical Research Centre, University Hospitals Southampton

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

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

[e]School of Water, Energy and Environment, Cranfield University, Cranfield, Bedfordshire, MK43 0AL, UK.

*Equal contribution

Correspondence to:

Name: Professor John W Holloway, PhD

Address: Duthie Building, MP808, Southampton General Hospital, SO16 6YD, United Kingdom

Tel: +44 23 8120 8758

Email: j.w.holloway@soton.ac.uk

Count: abstract - 243 words; main text - 3978 words; 6 tables; 1 figure, supplementary tables: 4

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

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

Abstract

BACKGROUND: The inability to objectively diagnose childhood asthma before age five often results in both under- 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 utilising 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 utilised machine learning approaches (n=5). Nine studies conducted validation of six regression-based models. Fifteen (out of 21) models required additional clinical tests. Overall model performance, assessed by Area Under the Receiver-Operator-Curve (AUC), ranged between 0.66-0.87. Models demonstrated moderate ability to either rule in or rule out asthma development, but not both. Where external validation was performed, models demonstrated modest generalisability (AUC range: 0.62-0.83).

CONCLUSION: Existing prediction models demonstrated moderate predictive performance, often with modest generalisability 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.

Key Words:

Childhood, Wheeze, Asthma, Risk scores, Prediction model

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

Introduction

Asthma is the most common chronic disease in children [1,2]. The clinical presentation of childhood asthma is highly heterogeneous. Whilst 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- 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 personalised 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].

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.

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].

Search strategy

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

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 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.

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 which improved upon existing models were included. Studies which 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, randomised control trials, proceedings, letters, editorials and non-English articles were excluded.

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 I)). 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 generalisability. 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.

Results

The literature search identified 4187 articles (Figure I). 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 II) machine learning approaches (n=4) (Appendix– Table EIV). The remaining four studies were external validations of previously developed models (Table V).

Regression-based models

Twenty-one regression-based prediction models were described in 20 studies (Table II). Thirteen of 21 models were novel whilst eight were modifications of existing models: six modified the Asthma Predictive Index (API) [24-29]; one updated the PIAMA risk score [30] 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].

Target population

Of the 21 models carried forward for qualitative analysis (Table II), 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].

Predictors

Thirty-eight different predictors were used among the 21 identified models, including seven variations of wheeze and two different measures for both allergic sensitisation and pulmonary

function (Table III). The number of predictors used to construct the models ranged between 3 and 10. Twenty out of 38 predictors were each included in just one of the 21 models (last column, Table III). 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] whilst 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 sensitisation 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).

Outcome

The prediction outcome in most studies (19/20) was school-age asthma, yet nine different definitions of asthma were used (Table IV). 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].

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].

Performance measures

Three studies failed to report any model performance measures detailed in Table I. Of these, the modified Asthma Predictive Index (mAPI), developed within a randomised 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 1.94-21 whilst the ability to rule out disease (LR-) ranged from 0.13-0.87.

Validation

Nine studies performed external validation: four validated the loose and/or stringent API, two validated PIAMA and PARC whilst PAPS and PARS were each validated once (Table V). 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. Whilst 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.

Critical appraisal

The overall risk of bias was deemed high for all 21 models due to: i) 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 ii) biased analysis due to an

inappropriate number of candidate predictors, inappropriate handling of missing data, failure in reporting performance measures (e.g. calibration) or failure in treating models for potential overfitting or performance optimisation as detailed in the PROBAST checklist (Table VI).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.

Machine Learning Approaches

Four studies which utilised 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 EIV).

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 utilised 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.

Choice of population

The performance of any predictive model is highly dependent on its developmental setting and may not generalise 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 population samples. 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 V).

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 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 excluded from the model [37].

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 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 [38].

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.

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 generalisability and potential clinical implementation.

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] utilised 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].

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, optimisation 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.

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 [60-62]. 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.

Validation

Models tend to perform best within their development population. External validation studies, which assess the true performance of models in independent populations, are essential to assess the generalisability 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 V). Whilst 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 V). Validation of all existing models within a single independent population using a single outcome definition is necessary to standardise 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.

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 I. 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 generalisability 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 optimising 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 optimise 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, whilst 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 utilising the API [29] or, more frequently, parental history, to assess asthma risk.

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.

1. Study design and data availability

Improving model generalisability across all population settings could be achieved by standardising predictor and outcome definitions across settings, and addressing issues of population bias and data availability. Whilst 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, standardised questionnaires with unambiguous descriptions which are consistent across both clinical and research settings. Where parental-reported data is used, clinical jargon should be deconstructed and/or be supported by auditory or visual aids to minimise recall bias and misclassification wherever possible.

In addition, only easily derivable and commonly available clinical predictors should be used. Whilst 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.

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 by past 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, due to population-specific influences and predictor selection methodological limitations of these studies as previously discussed, a better approach would be for future studies to utilise a robust predictor selection method e.g. 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.

3. Model development methodologies

The majority of existing studies have utilised regression-based methods and have developed a number of similar prediction models, few generalising 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 are necessary to evaluate a model's generalisability and subsequently promote its clinical application for predicting school-age asthma.

Acknowledgements

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.

Author contributions

Conceived and designed the study: DMK, FIR, JWH and SHA. Conducted systematic review screening: DMK, VBNW, MAK, LK and FIR. Performed critical appraisal: DMK and LK. Analysed findings: DMK. Drafted manuscript: DMK. Revised manuscript: All. Approved final manuscript: All.

References

1. Akdis CA, Allergy EAo, Immunology C, Agache I. *Global Atlas of Asthma.* 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. *The Lancet.* 2018;391(10118):350-400.
4. Global Initiative for Asthma G. *Global Strategy for Asthma Management and Prevention.* 2018.
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 FD WA, Taussig LM, Holberg CJ, Halonen M, Morgan WJ. Asthma and wheezing in the first six years of life. the Group Health Medical Associates. *The New England Journal of Medicine.* 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. Isabella Annesi-Maesano CS, Denis Caillaud, Fréderic de Blay, François Lavaud, Denis Charpin and Chantal Raherisson. Factors related to under-diagnosis and undertreatment of childhood asthma in metropolitan France. *Multidisciplinary Respiratory Medicine.* 2012;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. 2009.
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. *The Lancet Respiratory Medicine.* 2015;3(12):973-984.

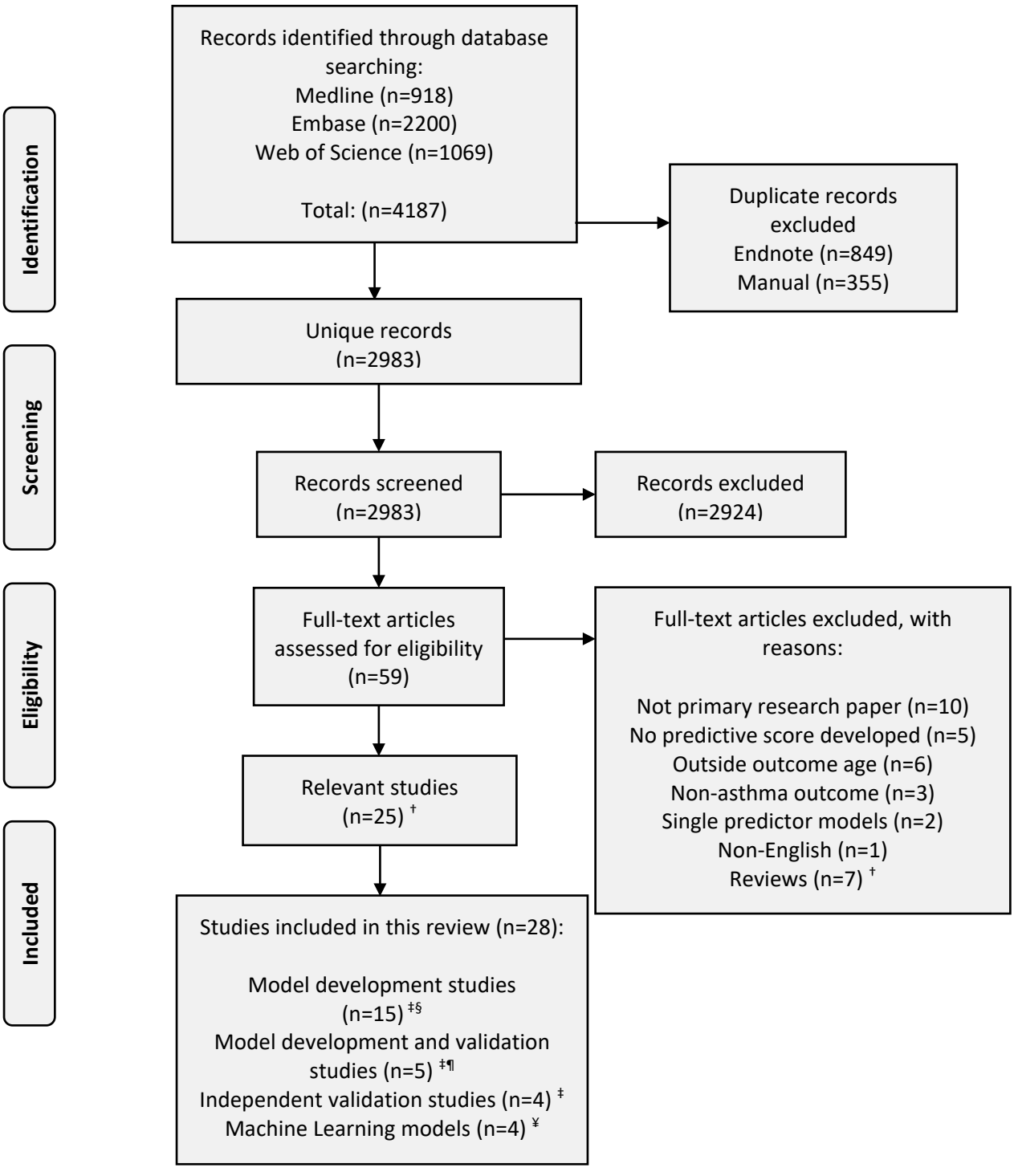
19. Luo G, Nkoy FL, Stone BL, Schmick D, Johnson MD. A systematic review of predictive models for asthma development in children. *BMC Medical Informatics and Decision Making.* 2015;15(1).
20. Colicino S, Munblit D, Minelli C, Custovic A, Cullinan P. Validation of childhood asthma predictive tools: A 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. In: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 JA HC, Wright AL, Martinez FD. A Clinical Index to Define Risk of Asthma in Young Children with Recurrent Wheezing. *American Journal of Respiratory and Critical Care Medicine.* 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. *American Journal of Respiratory and Critical Care Medicine.* 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. *The Journal of Allergy and Clinical Immunology: In Practice.* 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. *Journal of Allergy and Clinical Immunology.* 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.
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 SB, de Jong C, Halbeisen F, Ramette A, Gaillard EA, Granell R, Henderson AJ, Kuehni CE. A simple 10-iten PARC tool to predict childhood asthma - an external validation.pdf. *The Journal of Allergy and Clinical Immunology: In Practice.* 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. Lodrup 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. *Journal of Allergy and Clinical Immunology.* 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. *British Journal of General Practice: the journal of the Royal College of General Practitioners* 2005;55(511):125-131.
41. Kurukulaaratchy RJ, Matthews S, Holgate ST, Arshad SH. Predicting persistent disease among children who wheeze during early life. *European Respiratory Journal.* 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. *Journal of Allergy and Clinical Immunology.* 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. *Journal of 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. *Journal of Allergy and Clinical Immunology.* 2014;133(1):111-118.e113.
45. van der Mark LB vWK, 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. *Primary Care Respiratory Journal.* 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. *Pediatric Pulmonology.* 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. *Clinical & Experimental 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. *Computational and Mathematical Methods in Medicine.* 2013;2013:1-6.
51. Chatzimichail E, Paraskakis E, Rigas A. Predicting Asthma Outcome Using Partial Least Square Regression and Artificial Neural Networks. *Advances in Artificial Intelligence.* 2013;2013:1-7.
52. Aligne CA, Auinger P, Byrd RS, M. W. Risk Factors for Pediatric Asthma Contributions of Poverty, Race, and Urban Residence. *American Journal of Respiratory and Critical Care Medicine.* 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. *Human Reproduction.* 2013;28(1):77-86.
54. Ranganathan P, Pramesh CS, Aggarwal R. Common pitfalls in statistical analysis: Logistic regression. *Perspectives in clinical research.* 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. *Egyptian Informatics Journal.* 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. *Archivos Argentinos de Pediatria.* 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.* 1 ed: Springer-Verlag New York; 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.

587 **Figures**



588

589 Figure I. PRISMA flow diagram of study search strategy
590 †Citation screening of articles identified three additional studies
591 ‡Included in the final qualitative analysis
592 §One study transformed a diagnostic model into a prediction model upon external validation
593 (considered a developmental study in this review).
594 ¶Validated the developmental study model (n=2) or an existing model (n=3).
595 ¥Excluded from the main qualitative analysis

Tables

Table I. 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.

599 Table II. Summary of existing childhood asthma prediction models

	Risk Score	Year	Target population¤, age	Prediction population, age	Study size, prevalence (n, %)¥	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	LR+	LR-	Discrimination
Validated Studies												
API	Loose Asthma Predictive Index (API)* [24]	2000	General population, ≤3	6, 8, 11, 13	986 (57.1)	41.60	84.70	59.10	73.20	2.72¥	0.69¥	-
	Stringent Asthma Predictive Index (API)* [24]	2000	General population, ≤3	6, 8, 11, 13	1002 (57.1)	15.70	97.40	76.60	68.30	6.04¥	0.87¥	-
PIAMA	Prevalence and Incidence of Asthma and Mite Allergy (PIAMA) [§^☖] [42]	2009	High-risk¤, 0-4	7-8	2171 (11.1)	19.00	97.00	42.00	91.00	6.33¥	0.84¥	0.72
PAPS	Persistent Asthma Predictive Score (PAPS)[^] [43]	2011	High-risk¤, <2	6	200 (47.5)	42.40	89.60	66.70	75.90	4.06	0.64	0.66
PARC	Predicting Asthma Risk in Children (PARC) Tool[^§☖] [44]	2014	High-risk¤, 1-3	6-8	1226 (28.1)	72.00	71.00	49.00	86.00	2.47	0.40	0.74
PARS	Paediatric Asthma Risk Score (PARS)[^☖] [39]	2018	High-risk$^{\mu}$, ≤3	7	589 (16.1)	68.00	77.00	37.00	93.00	3.02	0.41	0.80
Exploratory Studies												
API	Modified Asthma Predictive Index (mAPI)*[ᵠ] [28,29]	2004	High risk$^{\mu}$, 2-3	4-6	259 (28.2)	17.00	99.00	-	-	21.00	0.84	-
	Singer et al. risk score (API + FeNO)[§] [26]	2013	High risk¤, ≤ 4	6-10	166 (41.0)	75.00	62.30	58.00	78.20	1.99¥	0.40¥	-
	Modified Asthma Predictive Index (m^2API)* [28]	2014	High risk$^{\mu}$, 1-3	6, 8, 11	259 (28.2)	30.00	98.00	-	-	16.00	0.71	-

	Risk Score	Year	Target population¤, age	Prediction population, age	Study size, prevalence (n, %)¥	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	LR+	LR-	Discrimination
	University of Cincinnati (ucAPI) [25]	2014	High risk[µ], 3	7	589 (17.5)	44.00	94.10	60.30	89.30	7.50	0.60	-
	Klaassen et al. (API + biomarkers)[§☖] [27]	2015	High risk[¤], 2-4	6	198 (38.4)	88.00	90.00	90.00	89.00	8.80[¥]	0.13[¥]	0.86
IOW	Recurrent Wheeze Score (Isle of Wight, IOW) [larr] [41]	2003	High risk[¤], 4	10	1034 (12.1)	52.50	84.60	68.40	73.70	3.41[¥]	0.56[¥]	-
RAST	Radio-Allergosorbent Testing (RAST)[§ larr] [40]	2005	High risk[†], 1-4	6	123 (26.8)	-	-	-	-	-	-	0.87
OAD	Obstructive Airway Disease (OAD) score for asthma[☖ larr] [36]	2007	General population[θ], 2	10	449 (21.6)	55.60	86.40	52.90	87.60	4.09[¥]	0.51[¥]	0.78
	Combined IgE antibodies and OAD (OAD + IgE) [31]	2010	General population[θ], 2	10	371 (50.0)	-	-	-	-	-	-	-
PIAMA	Updated PIAMA [^] [30]	2013	High-risk[‡], 0-4	6-7	5048 (5.5)	63.80	73.90	12.40	97.20	2.44	0.49	0.75
Lødrup Carlsen et al.	Lødrup Carlsen et al.[larr] [38]	2010	General population, at birth	10	607 (11.0)	64.00	67.00	19.00	94.00	1.94[¥]	0.54[¥]	0.72
CAPS	Clinical Asthma Prediction Score (CAPS) [^§☖ larr] [45]	2013	High-risk[¤], 1-5	6	438 (42.7)	-	-	74.30	78.40	-	-	0.73
Boersma et al.	Boersma et al.[larr] [46]	2017	High risk[¤], 1-3	6	116 (62.9)	-	-	-	-	-	-	0.79
Szentpetery et al.	Szentpetery et al. [¬ larr] [37]	2017	General population, 1-4	8	2339 (5.0)	-	-	-	-	-	-	-

	Risk Score	Year	Target population¤, age	Prediction population, age	Study size, prevalence (n, %)¥	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	LR+	LR-	Discrimination
MAAS-APT	Manchester Asthma and Allergy Study Asthma Prediction Tool (MAAS-APT)^§⌂ Ɫ [47]	2018	High-risk[¤], 3	8-11	336 (34.8)	47.00	93.00	75.00	78.00	6.30	0.60	0.79

PPV: positive predictive value; NPV: negative predictive value; LR+: positive likelihood ratio; LR-: negative likelihood ratio

Shaded: modified prediction models

[†] Prevalence = proportion of cases with the outcome included in the selected study sample– for the IOW model which details a stratified outcome of wheeze, prevalence of persistent wheeze is reported; prevalence for the loose and stringent API refers to the reported active asthma in at least one survey within the prediction window.

[‡]The loose and stringent API, mAPI and m^2API were all evaluated at 6, 8, 11 and 13 (loose and stringent API only) years. Study details and performance measures are given for asthma prediction in at least one survey within the prediction window for the loose and stringent API and at age 6 for the mAPI/m^2API.

[§] Internal validation during model development was performed using bootstrapping (API+biomarkers, API+FeNO, MAAS-APT, RAST, PIAMA and CAPS) or leave-one-out cross validation (PAPS). Where applicable, the internal validation performance measures are reported. Unbootstrapped discrimination was reported for the API+biomarkers (AUC=0.95), RAST (AUC=0.87), and PIAMA risk score (AUC=0.74).

[¶] Performance measures extracted from Chang et al.'s study

[^] Models provided performance measures over a range of thresholds. Performance measures are reported at the threshold recommended in their developmental study.

[º] The study initially developed a diagnostic model targeting and predicting childhood asthma at age 6. For external validation in the BAMSE cohort, this model was transformed into a prediction model targeting children between ages 1-4 to predict asthma at age 8. The latter model was considered as a developmental study in this review and study details are reported for the BAMSE population in which the prediction model was evaluated.

[¥] Where unreported, likelihood ratios were calculated based on reported sensitivity and specificity as: LR+ = sensitivity/ 1- specificity, LR- = 1- sensitivity/ specificity.

[⌂] Model calibration was evaluated in the study

High-risk study cohort specified by parental history of allergy/asthma ([µ]), presence of asthma-like symptoms (ɬ), asthma-like symptoms specifically wheeze ([¤])

[θ] Nested case-control study within a general population birth cohort of children age 2 with obstructive airway disease.

623 Table III. Predictors included in the final asthma prediction models

		Loose API [24]	Stringent API [24]	mAPI* [29]	m²API* [28]	API + FeNO [26]	API + biomarkers [27]	ucAPI [25]	IOW [41]	RAST [40]	OAD [36]	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]	Total†
Subject characteristics																							
	Sex												X	X	X		X			X			5
	Age									X							X	X					3
	Race																					X	1
	Gestation time												X	X									2
Clinical Symptoms																							
Wheeze*	Any									X													1
	Early wheeze	X				X																X	3
	Frequent wheeze			X	X			X					X	X			X						6
	Early frequent wheeze		X																				1
	Exercise-induced																X				X		2
	Aeroallergen-induced																X						1
	Wheeze without cold	X	X	X	X	X	X	X					X	X			X	X				X	12
	Eczema	X	X	X	X	X	X	X					X	X		X	X		X		X	X	14
	Allergic rhinitis	X	X			X	X	X												X			6
	Shortness of breath																X				X		2
	Nasal symptoms								X														1
	Nocturnal symptoms																	X					1
	Cough on exertion																				X		1

		Loose API [24]	Stringent API [24]	mAPI* [29]	m²API* [28]	API + FeNO [26]	API + biomarkers [27]	ucAPI [25]	IOW [41]	RAST [40]	OAD [36]	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]	Total†
	Recurrent chest infection								X														1
	Respiratory tract infection												X										1
	Hospital admission for respiratory symptoms										X	X							X				3
	Number of bronchial obstruction episodes										X	X											2
	Duration of bronchial obstruction										X	X											2
	Disturbances to activity																X						1
	Obesity																			X			1
Familial characteristics																							
	Parental asthma	X	X	X	X	X	X	X	X					X		X	X	X		X		X	14
	Familial allergy to pollen									X													1
	Parental inhaled medication												X										1
	Alcohol intake during pregnancy														X								1
Environmental exposures																							
	Parental education												X	X									2
	Ease of acquiring a babysitter														X								1
	Family network														X								1

		Loose API [24]	Stringent API [24]	mAPI* [29]	m²API* [28]	API + FeNO [26]	API + biomarkers [27]	ucAPI [25]	IOW [41]	RAST [40]	OAD [36]	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]	Total†
Clinical tests																							
Atopy/ allergic sensitisation	Specific IgE (RAST^ or other assay)			X	X		X			X		X				X		X	X				8
	Skin prick test (SPT)							X	X												X	X	4
Pulmonary Function	Fraction of exhaled nitric oxide (FeNO)					X																	1
	Lung function (V_E§)														X								1
Other	Blood eosinophilia	X	X	X	X																		4
	Exhaled volatile organic compound						X																1
	Gene expression						X																1
Total ‡		6	6	6	6	6	7	6	4	4	3	4	8	7	5	3	10	5	3	6	5	6	

† Total number of models that use each predictor

‡ Total number of predictors included in each model

§ V_E = minute ventilation

628 Table IV. Nine main classes of asthma definitions used amongst asthma prediction model
629 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[⌂]	[28]
5. Symptoms and medication	2	[44,46]
6. Doctor diagnosis, symptoms, medication	1	[42]
7. Symptoms, medication, lung function tests[†]	5	[27 §¶, 25 §¶, 40 ¶, 45 §¶, 38‡]
8. Doctor diagnosis, symptoms, lung function tests[†]	1	[39‡ ¶]
9. Doctor diagnosis, symptoms, medication, lung function tests[†]	3	[36‡, 31‡, 43 §]

630 [†]Lung function tests comprised of one or a combination of exercise tests ([‡]), spirometry assessing
631 reversibility to bronchodilators ([§]) and bronchial hyper-responsiveness to methacholine or histamine
632 ([¶]).
633 [⌂]The asthma outcome for the mAPI was extracted from the m^{2}API study which evaluated the
634 model's performance.

635 Table V. Model performance of externally validated asthma prediction models

	Author	Population geography	Variation in predictors	Variation in outcome	Study size (prevalence, %)	Study asthma prevalence (%)	Target age	Prediction age	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	LR+	LR-	Discrimination
Loose API	Castro-Rodriguez et al. [24]	USA			986	57.1	<3	6-13	41.6	84.7	59.1	73.2	2.72†	0.69†	-
	Rodriguez-Martinez et al. [32]	Colombia	-	-	93	22.5	1-3	5-6	71.4	33.3	23.8	80	1.07	0.86	-
	Leonardi et al. [35]	UK	✓	-	1731	11.5	2-3	7	57	80	26	94	2.85†	0.54†	0.68
					1291	10.5	2-3	10	57	81	25	94	3.00†	0.53†	0.69
	Devulapalli et al. [36]	Norway	✓	✓	459	21.1	3	10	59.8	79	43.9	87.7	2.85†	0.51†	-
Stringent API	Castro-Rodriguez et al. [24]	USA			1002	57.1	<3	6-13	15.7	97.4	76.6	68.3	6.04†	0.87†	-
	Rodriguez-Martinez et al. [32]	Colombia	-	-	93	22.5	1-3	5-6	42.9	79.2	37.5	82.6	2.06	0.72	-
	Leonardi et al. [35]	UK	✓	-	1683	11.5	2-3	7	37	93	40	93	5.29†	0.68†	0.65
					1257	10.5	2-3	10	32	94	35	92	5.33†	0.72†	0.63
	Caudri et al.	Netherlands	✓	✓	1177	11.7	0-4	7-8	20	92	25	90	2.50†	0.87†	0.62
	Devulapalli et al. [36]	Norway	✓	✓	459	21.1	3	10	56.7	83	47.8	87.4	3.34†	0.52†	-
PIAMA	Caudri et al. [42]	Netherlands			2171	11.1	0.4	7-8	19	97	42	91	6.33†	0.84†	0.74
	Hafkamp-de Groen et al. [30]	Netherlands	✓	✓	2877	6.0	1-4	6	-	-	-	-	-	-	0.74
	Rodriguez-Martinez et al. [32]	Colombia	-	✓	123	53.6	1-3	5-6	54.5	78.9	75.0	60	2.59	0.58	-
PARC	Pescatore et al. [44]	UK			1226	28.1	1-3	6-8	72	71	49	86	2.47	0.40	0.74
	Grabenhenrich et al. [33]	Germany	✓	-	140	20.0	3	8	82	69	40	94	2.63	0.26	0.83
	Pedersen et al. [34]	UK	✓	-	2690	14.0	1.5-3.5	7.5	69	76	32	94	2.87	0.41	0.77

	Author	Population geography	Variation in predictors	Variation in outcome	Study size (prevalence, %)	Study asthma prevalence (%)	Target age	Prediction age	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	LR+	LR-	Discrimination
PAPS	Dupuy et al. [43]	France			200	47.5	<2	6	42.4	89.6	66.7	75.9	4.06	0.64	0.66
	Dupuy et al. [43]	France	-	-	227	18.9	<2	13	62.8	67.4	31	88.6	1.93†	0.55†	0.65
PARS	Biagini Myers et al. [39]	USA			589	16.1	≤3	7	68	77	37	93	3.02	0.41	0.80
	Biagini Myers et al. [39]	UK	✓	✓	1098	-	2	10	67	79	36	93	3.25	0.41	0.79

PPV: positive predictive value; NPV: negative predictive value; LR+: positive likelihood ratio; LR-: negative likelihood ratio

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

† Likelihood ratios calculated based on reported sensitivity and specificity as: LR+ = sensitivity/ 1- specificity, LR- = 1- sensitivity/ specificity

Table VI. Critical appraisal of each study's risk of bias and applicability using the PROBAST checklist

	Loose API [24]	Stringent API [24]	mAPI [29]	m²API [28]	API + FeNO [26]	API + biomarkers [27]	ucAPI [25]	IOW [41]	RAST [40]	OAD [36]	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]
Risk of Bias																					
Participants	L	L	H	H	H	H	H	H	H	L	L	H	H	L	H	H	H	H	L	H	H
Predictors	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H
Outcome	H	H	H	H	H	L	H	H	H	L	L	H	H	H	U	H	H	H	H	H	H
Analysis	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H
Overall risk	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**H**
Concern regarding Applicability																					
Participants	L	L	H	H	H	H	H	H	H	L	L	H	H	L	H	H	H	H	L	H	H
Predictors	L	L	L	L	L	H	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
Outcome	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
Overall risk	**L**	**L**	**H**	**H**	**H**	**H**	**H**	**H**	**H**	**L**	**L**	**H**	**H**	**L**	**H**	**H**	**H**	**H**	**L**	**H**	**H**

Risk of bias and applicability were assessed as: H = High risk, L=Low risk, U= Unclear risk using the criteria as 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, unclear if ≥1 domain was considered unclear and all other domains were low-risk.

Supporting Information

Table EI. Search strategy used for Embase database search

Embase (1947 to 25th July 2019)
1. exp asthma/
2. asthma*.mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word]
3. wheezing/
4. wheez*.mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word]
5. 1 or 2 or 3 or 4
6. exp child/
7. (child or children or childhood or paediatric* or pediatric* or infant* or school-age or preschool or pre-school or early life or toddler*).mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word]
8. 6 or 7
9. "prediction and forecasting"/ or prediction/ or computer prediction/
10. scoring system/
11. ((forecast* or predict* or risk*) adj3 (score* or model* or system* or formula* or value* or index* or tool*)).mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word]
12. exp machine learning/
13. exp artificial intelligence/
14. intelligent system*.mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word]
15. 9 or 10 or 11 or 12 or 13 or 14
16. onset age/
17. (develop* or onset or outcome).mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word]
18. 16 or 17
19. 5 and 18
20. 8 and 19
21. 15 and 20

Table EII. Search strategy used for Medline database search

Medline Search Strategy (1946 to 25th July 2019)
1. exp Asthma/
2. asthma*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
3. wheez*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
4. 1 or 2 or 3
5. exp Child/
6. (child or children or childhood or paediatric* or pediatric* or infant* or school-age or preschool or pre-school or early life or toddler*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
7. exp Infant/
8. 5 or 6 or 7
9. ((forecast* or predict* or risk*) adj3 (score* or model* or system* or formula* or algorithm* or value* or index* or tool*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
10. exp Artificial Intelligence/
11. exp Machine Learning/
12. exp algorithms/
13. intelligent system*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
14. 9 or 10 or 11 or 12 or 13
15. "age of onset"/
16. (develop* or onset or outcome).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
17. 15 or 16
18. 4 and 17
19. 8 and 18
20. 14 and 19

3 Table EIII. Search strategy used for Web of Science database search

Web of Science Search Strategy
#1 TOPIC: (asthma*) OR TOPIC: (wheez*) DocType=All document types; Language=All languages;
#2 TOPIC: ((child OR children OR childhood OR paediatric* OR pediatric* OR infant* OR school-age OR preschool OR pre-school OR "early life" OR toddler*)) DocType=All document types; Language=All languages;
#3 TOPIC: (((forecast* OR predict* OR risk*) NEAR/3 (score* OR model* OR system* OR formula* OR algorithm OR value* OR index* or tool*))) DocType=All document types; Language=All languages;
#4 TOPIC: ("machine learning" OR "artificial intelligence" OR algorithm* OR "intelligent system*") DocType=All document types; Language=All languages;
#5 TOPIC: (develop* OR onset OR outcome*) DocType=All document types; Language=All languages;
#6 #4 OR #3 DocType=All document types; Language=All languages;
#7 #6 AND #2 AND #1 DocType=All document types; Language=All languages;
#8 #5 AND #1 DocType=All document types; Language=All languages;
#9 #8 AND #6 AND #2 DocType=All document types; Language=All languages;

4

Table EIV. Summary of identified studies using machine learning approaches to developed prediction models for childhood asthma onset

Study	Method of predictor selection	Number of predictors	Model Algorithm	Study size	Sensitivity	Specificity	PPV	NNV	LR+	LR-	AUC	External validation
[48]	Correlation analysis	10	MLP	112	1.00	1.00	-	-	NA†	0.00†	-	No
[49]	Genetic algorithm	4	ANN	112	-	-	-	-	-	-	-	No
[50]	Principal Component Analysis	18	Least square Support Vector Machine	112	0.95	0.96	-	-	21.64†	0.05†	-	No
[51]	Partial least square regression	9	MLP	112	0.96	1.00	-	-	NA†	0.04†	-	No
			Probabilistic Neural Network		1.00	0.80	-	-	5.00†	0.00†	-	No

† Likelihood ratios calculated based on reported sensitivity and specificity as: LR+ = sensitivity/ 1- specificity, LR- = 1- sensitivity/ specificity, NA= calculation error

- Not reported