MS SILVIA COLICINO (Orcid ID : 0000-0002-0262-6280)

Article type : Systematic reviews and Meta-analysis

Validation of childhood asthma predictive tools: a systematic review

Silvia Colicino[1*] and Daniel Munblit[2,3,4,5*], Cosetta Minelli[1], Adnan Custovic[2], Paul Cullinan[1]

1 National Heart and Lung Institute, Imperial College London, United Kingdom
2 Department of Paediatrics, Imperial College London, United Kingdom
3 Faculty of Paediatrics, I. M. Sechenov First Moscow State Medical University, Russian Federation
4 International Inflammation (in-FLAME) network of the World Universities Network
5 Solov'ev Research and Clinical Center for Neuropsychiatry, Moscow, Russian Federation

Correspondence: s.colicino@ic.ac.uk
* These authors contributed equally to this work.

Abstract

Background: There is uncertainty about the clinical usefulness of currently available asthma predictive tools. Validation of predictive tools in different populations and clinical settings is an essential requirement for the assessment of their predictive performance, reproducibility and generalizability. We aimed to critically appraise asthma predictive tools which have been validated in external studies.

Methods: We searched MEDLINE and EMBASE (1946-2017) for all available childhood asthma prediction models and focused on externally validated predictive tools alongside the studies in which they were originally developed. We excluded non-English and non-original studies. PROSPERO registration number is CRD42016035727.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/cea.13336
This article is protected by copyright. All rights reserved.

Accepted Article

Results: From 946 screened papers, 8 were included in the review. Statistical approaches for creation of prediction tools included chisquare tests, logistic regression models and the least absolute shrinkage and selection operator. Predictive models were developed and validated in general and high-risk populations. Only three prediction tools were externally validated: the Asthma Predictive Index, the PIAMA, and the Leicester asthma prediction tool. A variety of predictors has been tested, but no studies examined the same combination. There was heterogeneity in definition of the primary outcome among development and validation studies, and no objective measurements were used for asthma diagnosis. The performance of tools varied at different ages of outcome assessment. We observed a discrepancy between the development and validation studies in the tools' predictive performance in terms of sensitivity and positive predictive values.

Conclusions: Validated asthma predictive tools, reviewed in this paper, provided poor predictive accuracy with performance variation in sensitivity and positive predictive value.

Key messages

What is the key question?

▸ Are available asthma predictive tools clinically useful in different populations and clinical settings?

What is the bottom line?

▸The number of external validation studies is limited and they show poor predictive ability

▸Estimates of the predictive performance in development and validation studies vary due to several reasons including flaws in study design and execution of the validation studies, poor data reporting, application of inappropriate statistical analysis, and heterogeneity in outcome definitions, characteristics of study populations, clinical settings and age at

This article is protected by copyright. All rights reserved.

Accepted Article

outcome assessment.

Why read on?

‣ This systematic review summarises the findings on all asthma predictive tools that have been externally validated, providing a critical appraisal of their predictive performance in validation studies in terms of sensitivity, specificity, predictive values, areas under the (ROC) curve, and generalizability in different clinical settings and populations.

‣ It provides suggestions for improvement of future research in the field, including suggestions for recruitment and screening of subjects, outcome definitions, selection of predictors, statistical modelling and clinical evaluation.

Introduction

Wheeze, cough, shortness of breath and chest tightness are cardinal symptoms of asthma. Wheeze tends to start in early life, often disappears during school age, but in some individuals persists for life (1, 2). In clinical practice, accurate tools for the early identification of children at high risk of persistent asthma would facilitate patient-physician communication based on information that is more objective. Physicians would be able to inform individuals of their risk of having the condition later in life, and to notify patients and their caregivers whether further testing and/or treatment are advisable (3, 4); the early exclusion of persistent asthma would help prevent unnecessary tests and over-treatment, and reduce costs.

This article is protected by copyright. All rights reserved.

Accepted Article

Accepted Article

Disease prediction models are clinically valuable when they accurately assess the risk of having the condition in the future (3), and have an impact on clinical decision making, patient outcomes and health costs.(5-10) In medical research, predictive tools can be used to screen and stratify patients for experimental studies, based on their risk of having the health outcome of interest.(4) Such models have clinical utility when they are able not only to distinguish patients with and without the disease with high accuracy, but also when their performance is reproducible across different populations and clinical settings.(11) In this context, it is necessary to test and validate predictive models in external studies, using different populations with comparable characteristics and similar clinical settings, as this enables the quantification of their predictive ability and an assessment of their generalizability.(6, 7, 12) The appropriate specification of the target population for external validation is crucial in the assessment of the validity and clinical utility of predictive models. The importance of validation has been demonstrated many times in different fields of medical research (13, 14), and a significant reduction in the predictive ability of a model when applied to a new dataset has been consistently reported.(14) Despite this, few studies have attempted to validate their models in different populations and settings; methods for internal, or 'apparent', validation are commonly used, but these approaches can lead to over-optimistic results.(6, 8, 12, 15)

Many prediction tools have been proposed to forecast asthma in children from the general population (16-20) and high-risk groups (21-28), but their predictive performance, clinical usefulness and reproducibility in external populations are limited, making them difficult to use in clinical practice and/or research.(29, 30) Despite lack of validation and uncertainty of predictive performance in different populations, the U.S. National Asthma Education and Prevention Program Expert Panel Report 3 (EPR-3) recommends the use of a modified

This article is protected by copyright. All rights reserved.

Asthma predictive index (API).(31) In contrast, two recent comprehensive systematic reviews (29, 30) concluded that none of the existing models can robustly predict or rule out persistent asthma. These reviews focused on development studies, without exploring the reliability and usefulness of the predictive tools in different populations and clinical settings. In the current study, we focused on validation studies together with the original development studies, and assessed their statistical methods, predictive abilities and clinical usefulness in children from both the general population and high-risk groups.

Methods

Search strategy and selection criteria

We focused only on studies which externally validated existing asthma predictive tools, alongside the studies in which they were originally developed. Our methods were based on the CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies (CHARMS) (32) and the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) statements.(33) The protocol was registered on PROSPERO (CRD42016035727).

An electronic search was performed on 11 July 2017 by a search of two databases, MEDLINE and EMBASE, using both free text and MESH terms. The search strategies are supplied as supplementary material (Figure E1 and Figure E2). In addition, further studies were traced through cross-checking of reference lists from identified relevant papers. Studies of all designs were included only when they: (i) investigated the prediction of asthma in external validation studies and in original studies that had developed the validated tools; (ii) performed outcome assessment in children between the ages of 6 and 18; (iii) tested at

This article is protected by copyright. All rights reserved.

Accepted Article

least three candidate predictors in the model development; (iv) evaluated at least one performance measure, including overall accuracy, discrimination and calibration measures; (v) were written in English. We excluded letters, editorials and conference papers. Overall, more than 900 papers were identified as potentially eligible; these were screened for eligibility based on title, abstract and full text when necessary, by two independent reviewers (SC and DM). The reason for exclusion was recorded; a flow diagram of our selection process is displayed in Figure 1.

Data extraction

Data extraction was performed in duplicate by two reviewers (SC and DM) using a form developed with the CHARMS checklist (32). The reviewers resolved any discrepancies through discussion with two other authors (CM and PC), until a consensus was reached. We extracted performance measures and calculated predictive values and likelihood ratios (LR) if they were not reported.

Results

Target Population

Studies that aimed to identify children who will have asthma in later life could be grouped into two main categories: (a) studies of the general population (with two studies recruiting participants at birth (16, 19) and four approaching generally healthy children recruited through healthcare practices, schools or local registries (17, 18, 34, 35)); and (b) studies in children at "high-risk" of persistent asthma (with two studies recruiting infants born to

This article is protected by copyright. All rights reserved.

Accepted Article

allergic parents (24, 26), eight studies approaching symptomatic children recruited at healthcare practices or schools (21-23, 25, 34, 36-38) and a single study looking at children attending hospital with asthma symptoms).(27)

In total, only three prediction tools were externally validated and therefore considered in this review: (1) the Asthma Predictive Index (API) (16), which was validated in two studies (35, 36); (2) the Prevention and Incidence of Asthma and Mite Allergy (PIAMA) tool (23), which was validated in one study (34), with a further study assessing the predictive capabilities of both API and PIAMA tool (37); (3) a simple asthma prediction tool developed by Pescatore *et al* (19), which was validated in one study.(38) We reviewed these five validation studies (34-38) and the three related development studies.(16, 19, 23) The API was the only tool developed in a general population, while most of its validation studies were undertaken in high-risk children.

Table 1 shows the main characteristics of these studies, which differed widely in their sample sizes, from a minimum of 130 participants (37) to more than 3,000 children.(35) Study populations included cohorts from the UK (19, 35), central Europe and Scandinavia (23, 34, 36, 38), USA (16) and South America.(37) The prevalence of asthma at outcome assessment varied from 5.8% (34) to 53.6%.(37)

Predictors

In total, 16 different candidate predictors were used in the models' development. All models included a history of self-reported or doctor-diagnosed eczema, the presence of wheeze aside from colds, and a measure of the frequency of early wheezing. Other variables included gender (19, 23), age (19), allergic rhinitis and self-reported or doctor-diagnosed

This article is protected by copyright. All rights reserved.

Accepted Article

parental asthma.(16, 19) The full list of predictors used in the development models is available in Table 2. While a wide variety of predictors has been tested, no studies examined the same combination. Aside from blood eosinophilia (16), other potential objective measurements (e.g. total and specific IgE levels, skin prick tests etc.) were not considered. Environmental exposures such as crowding, indoor pollution and pets in the house, potentially associated with asthma development (39), were not included in the models assessed.

Outcome definition

There was, among the development studies, heterogeneity in the way researchers defined the primary outcome (Table E1). The most commonly used criterion was wheezing in the previous 12 months. To define asthma for the API, Castro-Rodriguez and co-authors used a combination of doctor-diagnosed asthma and at least one wheezing episode in the last 12 months, or three or more episodes of wheeze regardless of a doctor's diagnosis.(16) In the PIAMA study, Caudri *et al.* defined childhood asthma at age of outcome assessment based on any of the following: at least one wheezing episode in the last 12 months, doctor diagnosis of asthma or asthma medication between the ages of seven and eight years.(23) The risk score developed by Pescatore *et al.* defined asthma as a combination of wheezing and asthma medication in the last 12 months.(19) No objective measurements (e.g. lung function tests) were used to assess asthma in any of the development studies.

While some validation studies (37, 38) defined outcomes using the same criteria as in the development study, others modified them. In the validation of the API, one study increased the required number of wheezing episodes in the preceding 12 months from three to four

This article is protected by copyright. All rights reserved.

Accepted Article

(35), while another (36) used a different combination of criteria (Table E1). Hafkamp-de Groen *et al.* (34) did not have data on asthma medication at the age of outcome assessment, and validated the PIAMA index using the definition of asthma from Leonardi *et al*.(35)

The age at outcome assessment varied across the development studies, with the API providing asthma prediction from six and up to 13 years (16), while other tools examined the outcome only up to the age of eight. The validation studies assessed asthma at an age which was one or two years different from the age in the development studies.

Statistical modelling and predictive performance

Statistical approaches used to develop the predictive models in the development studies included the chi-square test to compare proportions and logistic regression analysis, as well as more advanced penalised regression methods. Details of the statistical methods are reported in Table E2.

The reporting of performance measures in the development and validation studies varied. The discriminative ability measured by the area under the ROC curve (AUC) was reported in five studies (two development (19, 23) and three validation studies (34, 35, 38)) and ranged between 0.63 and 0.83 (Table E3). In general, predictive models with high sensitivity had lower specificity, and discrimination measures were better in studies in high-risk groups with a higher prevalence of outcome, compared to those in the general population. Likelihood ratios (LR) are additional measures of diagnostic accuracy and indicate the probability of a subject having, or not, the condition given the results of a predictive model. Overall, LRs were not reported in development studies; only two validation studies

This article is protected by copyright. All rights reserved.

Accepted Article

calculated these performance measures without providing their confidence intervals. We calculated point estimates of LRs to attempt a comparison among prediction models; however, the lack of confidence intervals may lead to misinterpretation of results and predictive performance.

The models with higher positive LR, at the same time, resulted in a higher negative LR (Figure 2). Positive LR varied between 1.07 and 7.43, and negative LR between 0.26 and 0.88. The highest LR were reported in the 'stringent API development study.(16)

The performance of the tools varied at different ages of outcome assessment (Figure 3-A1, Figure 3-A2 and Table E3). The 'stringent' and 'loose' versions of the API used in validation studies, showed higher sensitivity compared to the original study. The maximum values were reported in the studies carried out in a high-risk population, but had similar or lower specificity in the external validation studies at all age groups. All externally validated studies of the API reported much wider confidence intervals for the performance measurements compared to the development study (Table E3).

The PIAMA index has been externally validated in two studies, with only one providing most performance measures. The original PIAMA study (23) showed higher sensitivity and positive LR, similar specificity and much lower positive predictive value at the age of 7–8 years when compared with the study reported by Rodriguez-Martinez *et al*., which assessed children at 5–6 years (37) (Figure 3-B and Table E3).

A predictive score developed by Pescatore *et a*l.(19) was validated in a single study (38), which showed, at the age of eight, higher sensitivity, positive LR and AUC with a similar specificity and lower positive predictive value compared to the development study which assessed childhood asthma between the age of 6 to 8 years (Figure 3-B and Table E3).

This article is protected by copyright. All rights reserved.

Accepted Article

Risk of bias

The risk of bias (Table E4) was assessed using the CHARMS checklist (32), using previously reported criteria.(30) The risk of participant selection bias was low in all the studies but one.(37) The risk of bias for predictors and outcome assessment was low in all development and validation studies. Loss to follow-up resulted in a moderate attrition for seven studies out of eight; the risk of attrition bias was high in Rodriguez-Martinez and co-authors' study as loss to follow-up was higher than 20% and differences on key characteristics between participants and those who were lost-to-follow-up were not fully described. Half of the studies provided relevant aspects of the analysis resulting in a low risk of analysis bias (19, 23, 35, 38), while the other four had a moderate risk of such bias.(16, 34, 36, 37)

Discussion

Among all available asthma predictive tools, only API (16), PIAMA (23) and the Leicester tool (19) have been externally validated. These validated predictive models included in this systematic review showed limited accuracy in forecasting persistent wheeze and asthma at school age, and thus limited clinical usefulness in practice, but there is still room for other prediction models to be validated in the future. Overall, we observed a discrepancy between the original and validated models in predictive performance, especially when the tool was developed in a general population and validated in a high-risk group. The validation studies tended to report higher sensitivities, lower specificities and lower predictive values compared to the original studies. We may hypothesise that when predictive models are applied in external studies, they have a greater ability to identify children who will have persistent asthma compared to those established in development studies, but a propensity

This article is protected by copyright. All rights reserved.

Accepted Article

to produce prediction errors forecasting the condition in children whose symptoms will not endure. Since predictive values depend on the prevalence of the condition, tools developed in high-risk populations have higher positive predictive values compared to general population studies.

Likelihood ratios are useful in clinical practice as they allow physicians to calculate the probability of disease for specific patients by directly relating pre-test and post-test probabilities. Calculation of the post-test probability and comparison between pre- and post-test probabilities can also be assessed using the Fagan's nomogram which is a graphical tool where a straight line drawn from a patient's pre-test probability of condition (left axis) through the LR of the test (middle axis) is intersected with the post-test probability of disease (right axis). In general, the higher the pre-test probability or prevalence, the higher is the post-test probability. The European Academy of Allergy and Clinical Immunology (EAACI) suggests using a LR interpretation of IgE sensitization tests (40), and a similar approach may be used for asthma predictive tools.

Leonardi et al. (35), Devulapalli et al. (36) and Rodriguez-Martinez et al. (37) focused on external validation of the API tools but in different geographical regions, target populations and clinical settings. The performance of the 'stringent' version of the API at all age group varies significantly between validation studies, which makes meaningful comparison difficult. This probably reflects the initial population selection, as the 'stringent' version includes only children with early frequent wheeze. In contrast, the 'loose' version of the API seems to be more reproducible, showing better ability in asthma prediction but performing slightly worse in ruling out disease in the external validation studies in all age groups. The positive predictive values of both versions of the API vary among studies, highlighting the

This article is protected by copyright. All rights reserved.

Accepted Article

importance of population definitions based on early wheezing symptoms as well as the dependency between this performance measure and prevalence of the disease.

The variation in the predictive performance between development and validation studies can be explained by artefactual and clinical mechanisms.(41) Artefactual mechanisms refer to performance variability arising from flaws in the design and execution of the validation studies; these include misspecification of the target population in terms of inclusion criteria for the enrolment of participants, poor data reporting, application of inappropriate statistical analysis and lack of a standardisation in asthma definition.(42) The use of self- or proxy-reported questionnaires to establish asthma diagnosis is an important drawback of the included predictive models. It may lead to over-diagnosis of asthma, as a single episode of wheeze per annum reported by parents without any supportive clinical information may not accurately describe asthma. Recent data show that asthma is often over-diagnosed even in clinical settings (43), with parental reported medical history of cough and wheeze being the main reasons behind inaccurate diagnosis. This shows that self- or proxy-reported information has limited reliability as part of asthma definition, highlighting the importance of objective measurements and potentially explaining the difference in asthma prevalence in the studies assessed. However, no objective measurements were used to characterize asthma in both development and validation studies.

Possible clinical mechanisms include “case-mix variation” (44-46), which is a special type of the “spectrum effect”; the latter describes the variation in sensitivity, specificity and accuracy among different populations and subgroups (47-49), while case-mix variation refers to differences between development and validation studies in clinical settings, subject characteristics, age of outcome assessment, outcome prevalence and/or incidence.(44) In

This article is protected by copyright. All rights reserved.

Accepted Article

particular, the larger the difference between the characteristics of the original and external validation populations, the higher the discrepancy in predictive performance and the poorer the generalizability of findings.

Sensitivity, specificity and likelihood ratios are generally used as benchmarks for comparison of model performance as they are thought to be independent of the prevalence of the outcome (50-52), while predictive values vary with the prevalence of the disease. However, several studies have shown that sensitivity and specificity, together with likelihood ratios, may also vary with the prevalence of the disease.(42, 53-55) Disease prevalence is often predetermined by inclusion criteria, target population specifications and outcome definitions; for example, when individuals are selected based on their symptoms and risk profile (a 'high-risk' population), the prevalence is higher than in the general population (e.g. the Tucson Children's Respiratory Study is a general population and, as expected, the prevalence of asthma is lower compared to the high-risk validation studies established in Norway and Colombia). Although it is important to assess the predictive performance of a model in a new sample of individuals, from different locations and clinical settings, the external population should be homogeneous with that of the development study in order to reduce variation in performance, facilitate comparison among between models and establish the generalizability of the results. Among three models reviewed Pescatore prediction tool was the only one validated using exact parameters of original instrument. Authors of some API validation studies made some changes to API criteria, which potentially could influence instrument predictive abilities.

This article is protected by copyright. All rights reserved.

Accepted Article

Differences in age of outcome assessment can also lead to case-mix variation and change in predictive performance. Prediction is usually challenging, and predictive performance may be limited when the outcome is assessed later in the future using information collected very early in life. The majority of existing studies on childhood asthma prediction aimed to develop and validate asthma predictive tools between 6 and 10 years of age using predictors collected at 3-5 years of life, making prediction potentially relatively simple and reliable. The validation studies assessed asthma at a similar age to their development counterparts leading to comparable results, but it is notable that no studies have aimed to develop and externally validate asthma predictive models beyond the age of 13 years.

At the present, asthma is considered an 'umbrella term' (56), bringing together a selection of different conditions that share common clinical features such as cough and wheeze, shortness of breath, and bronchial obstruction.(57) While the concept of distinct wheezing and asthma endotypes has been proposed, existing guidelines based on clinical symptoms and predictive models are still aimed at a single disease.(58) This may partially explain the inaccuracy of asthma prediction models; future work should consider endotypes to a larger extent in the development of predictive models. Another important issue, potentially affecting predictive performance, is individual genetic predisposition to disease development. This important factor was not taken into account in the development and validation studies included in this systematic review.

With more research coming from Genome-wide association studies (GWAS) genetic variants that predispose individuals to asthma have been identified (59); for example, a meta-analysis of GWAS reported seven asthma genetic risk loci (HLA-DQ, IL33, ORMDL3/GSDMB, IL1RL1/IL18R1, IL2RB, SMAD3, and TSLP).(60, 61) Considering these in the development of

This article is protected by copyright. All rights reserved.

Accepted Article

future predictive models may lead to improved predictive performance.

To provide better assessment of the predictive ability of existing asthma prediction models, future research should aim at collecting information using well standardised questionnaires and clinical tests to allow for better harmonisation of predictor and outcome definitions. Collaborative international, multicentre efforts would offer the opportunity to increase sample size, and to develop predictive models using information from subsets of studies and then validate them in the remaining ones. More studies run on general and high-risk populations in parallel will allow for meaningful evaluation of model performance. Clinical tests including lung function tests, SPTs, allergen-specific IgE, Fractional exhaled Nitric Oxide (FeNO) and bronchodilator response might also be used to increase the capability of a predictive tool to correctly identify children at high-risk of asthma in later childhood and young adulthood. Moreover, external study populations should be carefully selected, with characteristics similar to those of the study populations where the predictive models were developed in order to have low heterogeneity within subpopulations.

In conclusion, validated tools for predicting childhood asthma provide poor predictive accuracy, with some performance variation in sensitivity and positive predictive value. Those available asthma predictive tools that have not been externally validated should be evaluated to assure their reliability. Larger cohorts are needed to include more predictors into the models and to use more advanced statistical methods. More work on standardisation of predictors and outcome assessment needs to be done.

Contributors
Silvia Colicino and Dr Daniel Munblit contributed equally to this work. These two authors searched MEDLINE and EMBASE for all available childhood asthma prediction models; they screened the papers and extracted a number of information from eligible studies. Silvia Colicino and Daniel Munblit contributed to creation of

This article is protected by copyright. All rights reserved.

Accepted Article

tables, graphs and interpretation of results; they also participated in drafting the article.

Dr Cosetta Minelli, Prof Adnan Custovic and Prof Paul Cullinan participated in revising the manuscript critically for important intellectual content and gave their final approval of the version to be submitted.

Declaration of interests
We declare no competing interests.

Role of the funding source
The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

References

1. Masoli M, Fabian D, Holt S, Beasley R, Global Initiative for Asthma P. The global burden of asthma: executive summary of the GINA Dissemination Committee report. Allergy. 2004;59(5):469-78.
2. Global strategy for asthma management and prevention. www.ginasthma.org (updated 2018).
3. Lee YH, Bang H, Kim DJ. How to Establish Clinical Prediction Models. Endocrinol Metab (Seoul). 2016;31(1):38-44.
4. Steyerberg EW. Clinical prediction models: a practical approach to development, validation, and updating. Springer Science & Business Media. 2008 Dec 16.
5. Riley RD, Hayden JA, Steyerberg EW, Moons KG, Abrams K, Kyzas PA, et al. Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. PLoS Med. 2013;10(2):e1001380.
6. Steyerberg EW, Moons KG, van der Windt DA, Hayden JA, Perel P, Schroter S, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. PLoS Med. 2013;10(2):e1001381.
7. Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. BMJ. 2009;338:b606.
8. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. BMJ. 2009;338:b605.
9. Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: Developing a prognostic model. BMJ. 2009;338:b604.
10. Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? BMJ. 2009;338:b375.
11. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ. 2015;350:g7594.
12. Moons KG, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: II. External validation, model updating, and impact assessment. Heart. 2012;98(9):691-8.
13. Zhang R, Borisenko O, Telegina I, Hargreaves J, Ahmed AR, Sanchez Santos R, et al. Systematic review of risk prediction models for diabetes after bariatric surgery. Br J Surg. 2016;103(11):1420-7.
14. Ivanescu AE, Li P, George B, Brown AW, Keith SW, Raju D, et al. The importance of prediction model validation and assessment in obesity and nutrition research. Int J Obes (Lond). 2016;40(6):887-94.
15. Steyerberg EW, Harrell FE, Jr., Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. J Clin Epidemiol. 2001;54(8):774-81.

This article is protected by copyright. All rights reserved.

Accepted Article

16. Castro-Rodriguez JA, Holberg CJ, Wright AL, Martinez FD. A clinical index to define risk of asthma in young children with recurrent wheezing. Am J Respir Crit Care Med. 2000;162(4 Pt 1):1403-6.
17. van der Werff SD, Junco Diaz R, Reyneveld R, Heymans MW, Ponce Campos M, Gorbea Bonet M, et al. Prediction of asthma by common risk factors: a follow-up study in Cuban schoolchildren. J Investig Allergol Clin Immunol. 2013;23(6):415-20.
18. Cano-Garcinuno A, Mora-Gandarillas I, Group SS. Wheezing phenotypes in young children: an historical cohort study. Prim. 2014;23(1):60-6.
19. Pescatore AM, Dogaru CM, Duembgen L, Silverman M, Gaillard EA, Spycher BD, et al. A simple asthma prediction tool for preschool children with wheeze or cough. J Allergy Clin Immunol. 2014;133(1):111-8.e1-13.
20. Ro AD, Simpson MR, Storro O, Johnsen R, Videm V, Oien T. The predictive value of allergen skin prick tests and IgE tests at pre-school age: The PACT study. Pediatric Allergy and Immunology. 2014;25(7):691-8.
21. 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. 2014;23(1):52-9.
22. Eysink PED, ter Riet G, Aalberse RC, van Aalderen WMC, Roos CM, van der Zee JS, et al. Accuracy of specific IgE in the prediction of asthma: Development of a scoring formula for general practice. British Journal of General Practice. 2005;55(511):125-31.
23. Caudri D, Wijga A, CM AS, Hoekstra M, Postma DS, Koppelman GH, et al. Predicting the long-term prognosis of children with symptoms suggestive of asthma at preschool age. J Allergy Clin Immunol. 2009;124(5):903-10.e1-7.
24. Chang TS, Lemanske RF, Guilbert TW, Gern JE, Coen MH, Evans MD, et al. Evaluation of the modified asthma predictive index in high-risk preschool children. Journal of Allergy and Clinical Immunology: In Practice. 2013;1(2):152-6.
25. 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;2013:240182.
26. Amin P, Levin L, Epstein T, Ryan P, LeMasters G, Khurana Hershey G, et al. Optimum Predictors of Childhood Asthma: Persistent Wheeze or the Asthma Predictive Index? Journal of Allergy and Clinical Immunology: In Practice. 2014;2(6):709-15.e2.
27. 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):729-36.
28. Klaassen EM, van de Kant KD, Jobsis Q, van Schayck OC, Smolinska A, Dallinga JW, 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):201-7.
29. 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.
30. Smit HA, Pinart M, Anto JM, Keil T, Bousquet J, Carlsen KH, et al. Childhood asthma prediction models: a systematic review. Lancet Respir Med. 2015;3(12):973-84.
31. National asthma education and prevention program - Expert panel report 3 (EPR-3): Guidelines for the diagnosis and management of asthma - Summary report 2007. J Allergy Clin Immun. 2007;120(5):S94-S138.
32. Moons KGM, de Groot JAH, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies: The CHARMS Checklist. Plos Medicine. 2014;11(10).
33. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med. 2009;6(7):e1000100.

This article is protected by copyright. All rights reserved.

34. Hafkamp-de Groen E, Lingsma HF, Caudri D, Levie D, Wijga A, Koppelman GH, 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):1303-10.
35. Leonardi NA, Spycher BD, Strippoli MPF, Frey U, Silverman M, Kuehni CE. Validation of the Asthma Predictive Index and comparison with simpler clinical prediction rules. J Allergy Clin Immun. 2011;127(6):1466-72.e6.
36. Devulapalli CS, Carlsen KC, Haland G, Munthe-Kaas MC, Pettersen M, Mowinckel P, et al. Severity of obstructive airways disease by age 2 years predicts asthma at 10 years of age. Thorax. 2008;63(1):8-13.
37. 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. Pediatric Pulmonology. 2011;46(12):1175-81.
38. Grabenhenrich LB, Reich A, Fischer F, Zepp F, Forster J, Schuster A, et al. The novel 10-item asthma prediction,tool: external validation in the german,MAS Birth cohort. PLoS ONE. 2014;9(12).
39. Rodriguez-Martinez CE, Sossa-Briceno MP, Castro-Rodriguez JA. Factors predicting persistence of early wheezing through childhood and adolescence: a systematic review of the literature. J Asthma Allergy. 2017;10:83-98.
40. Roberts G, Ollert M, Aalberse R, Austin M, Custovic A, DunnGalvin A, et al. A new framework for the interpretation of IgE sensitization tests. Allergy. 2016;71(11):1540-51.
41. Leeflang MMG, Rutjes AWS, Reitsma JB, Hooft L, Bossuyt PMM. Variation of a test's sensitivity and specificity with disease prevalence. Can Med Assoc J. 2013;185(11):E537-E44.
42. Leeflang MM, Bossuyt PM, Irwig L. Diagnostic test accuracy may vary with prevalence: implications for evidence-based diagnosis. J Clin Epidemiol. 2009;62(1):5-12.
43. Looijmans-van den Akker I, van Luijn K, Verheij T. Overdiagnosis of asthma in children in primary care: a retrospective analysis. Br J Gen Pract. 2016;66(644):e152-7.
44. Riley RD, Ensor J, Snell KI, Debray TP, Altman DG, Moons KG, et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. BMJ. 2016;353:i3140.
45. Debray TP, Riley RD, Rovers MM, Reitsma JB, Moons KG, Cochrane IPDM-aMg. Individual participant data (IPD) meta-analyses of diagnostic and prognostic modeling studies: guidance on their use. PLoS Med. 2015;12(10):e1001886.
46. Debray TP, Moons KG, Ahmed I, Koffijberg H, Riley RD. A framework for developing, implementing, and evaluating clinical prediction models in an individual participant data meta-analysis. Stat Med. 2013;32(18):3158-80.
47. Mulherin SA, Miller WC. Spectrum bias or spectrum effect? Subgroup variation in diagnostic test evaluation. Ann Intern Med. 2002;137(7):598-602.
48. Willis BH. Spectrum bias--why clinicians need to be cautious when applying diagnostic test studies. Fam Pract. 2008;25(5):390-6.
49. Usher-Smith JA, Sharp SJ, Griffin SJ. The spectrum effect in tests for risk prediction, screening, and diagnosis. BMJ. 2016;353:i3139.
50. Linden A. Measuring diagnostic and predictive accuracy in disease management: an introduction to receiver operating characteristic (ROC) analysis. J Eval Clin Pract. 2006;12(2):132-9.
51. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. JAMA. 1994;271(9):703-7.
52. Guyatt G, Sackett DL, Haynes RB. Evaluating diagnostic tests. Clinical epidemiology: how to do clinical practice research Lippincott williams & wilkins. 2012 Mar 29.
53. Mulherin SA, Miller WC. Spectrum bias or spectrum effect? Subgroup variation in diagnostic test evaluation. Annals of Internal Medicine. 2002;137(7):598-602.
54. Feinstein AR. Misguided efforts and future challenges for research on "diagnostic tests". J Epidemiol Community Health. 2002;56(5):330-2.

This article is protected by copyright. All rights reserved.

Accepted Article

55. Brenner HE, Gefeller OL. Variation of sensitivity, specificity, likelihood ratios and predictive values with disease prevalence. Statistics in medicine. 1997 May 15;16(9):981-91.
56. Marks GB. Identifying asthma in population studies: from single entity to a multi-component approach. Eur Respir J. 2005;26(1):3-5.
57. Belgrave D, Simpson A, Custovic A. Challenges in interpreting wheeze phenotypes: the clinical implications of statistical learning techniques. Am J Respir Crit Care Med. 2014;189(2):121-3.
58. Deliu M, Belgrave D, Sperrin M, Buchan I, Custovic A. Asthma phenotypes in childhood. Expert Rev Clin Immunol. 2017;13(7):705-13.
59. Belsky DW, Sears MR, Hancox RJ, Harrington H, Houts R, Moffitt TE, et al. Polygenic risk and the development and course of asthma: an analysis of data from a four-decade longitudinal study. Lancet Respir Med. 2013;1(6):453-61.
60. Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, Heath S, et al. A large-scale, consortium-based genomewide association study of asthma. N Engl J Med. 2010;363(13):1211-21.
61. Torgerson DG, Ampleford EJ, Chiu GY, Gauderman WJ, Gignoux CR, Graves PE, et al. Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. Nat Genet. 2011;43(9):887-92.

Figure 1: Flow diagram of manuscript selection

Figure 2: Likelihood ratios among predictive tools

Figure 3: Performance measures among predictive tools

Accepted Article

This article is protected by copyright. All rights reserved.

Accepted Article

Table 1: Characteristics of studies included in the systematic review by target population

Source of study	Target population	ID	Country	Study name (recruitment year)	Inclusion criteria	Sample size (% of those eligible)	Prediction rule	Age at outcome assessment (years)	Asthma prevalence at outcome n (%)
Developmental studies	General population	Castro-Rodriguez 2000(16)	USA	The Tucson Children's Respiratory Study (TCRS) (1980-1984)	Healthy new-born babies recruited in primary healthcare practices	1,246 (78)	Stringent API and Loose API	6	112 (11)
								8	113 (14)
								11	150 (16)
								13	122 (17.5)
								6-13	250 (35)
	High-risk population	Caudri 2009(23)	The Netherlands	The Prevention and Incidence of Asthma and Mite Allergy (PIAMA) (1996-1997)	Children with asthma-like symptoms at age 1-4 years	2,171 (55)	PIAMA	7-8	362 (16.7)
		Pescatore 2014(19)	UK	Leicestershire Respiratory Cohort Studies (1993-1997)	Children aged 1 to 3 years with parent-reported wheeze or chronic cough	2,444 (36)	Leicester tool	6-8	345 (28.1)

This article is protected by copyright. All rights reserved.

Accepted Article

Table 1: Characteristics of studies included in the systematic review by target population (*Continued*)

Source of study	Target population	ID	Country	Study name (recruitment year)	Inclusion criteria	Sample size (% of those eligible)	Prediction rule	Age at outcome assessment (years)	Asthma prevalence at outcome n (%)
Validation studies	General population	Leonardi 2011(35)	UK	Leicester Respiratory Cohort (1998)	Children, between 0-4 yrs, registered in local child health database	3392 (79)	Stringent API and Loose API	7	220 (11.5)
								10	147 (10.5)
	High-risk population	Devulapalli 2008(36)	Norway	Environment and Childhood Asthma study (NR)	Bronchial obstruction by age 2	612 (NR)	Stringent API and Loose API	10	239 (52.07)
		Rodriguez-Martinez 2011(37)	Colombia	NR (2006-2007)	Parental reported wheezing	130 (NR)	Stringent API and Loose API	5-6	33 (35.5)
							PIAMA		66 (53.6)
		Hafkamp-de Groen 2013(34)	The Netherlands	Generation R (2002-2006)	≥1 positive response to wheeze at 1 yr	2877 (73)	PIAMA	6	168 (5.8)
		Grabenhenrich 2014(38)	Germany	The German Multicentre Allergy Study (MAS-90) (1990)	Children with wheezing or cough in in the previous 12 months at the age of 3yrs	140 (17)	Leicester tool	8	28 (20)

Abbreviations – NR: not reported

This article is protected by copyright. All rights reserved.

Table 2: List of candidate predictors explored among development studies included in the systematic review

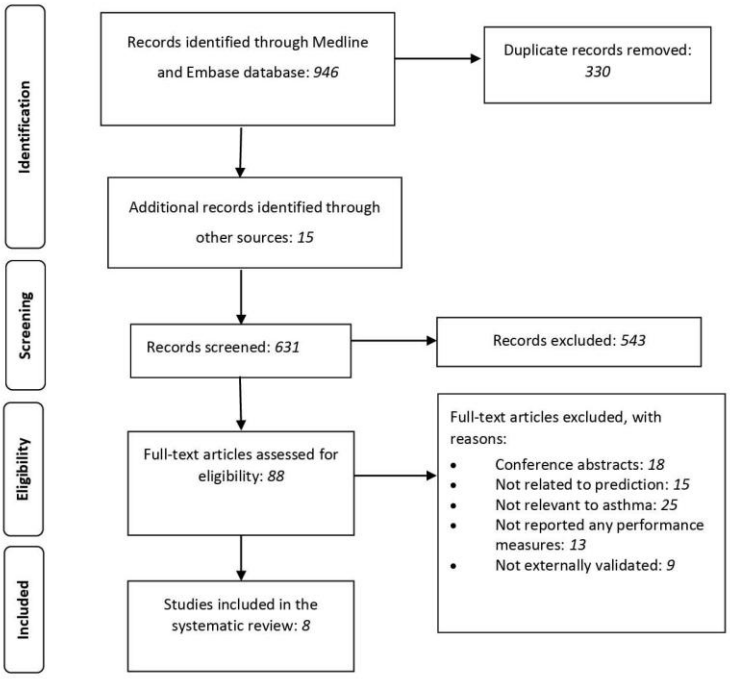
Candidate predictors	Castro-Rodriguez 2000(16) (API)	Caudri 2009(23) (PIAMA)	Pescatore 2014(19) (Leicester tool)
Gender		✓	✓
Age			✓
Post-term delivery		✓	
Parental education		✓	
Parental inhalation medication		✓	
Wheezing frequency	✓*	✓	✓
Wheezing apart from colds	✓	✓	✓
Infections		✓	
Eczema	✓	✓	✓
Parental asthma	✓		✓
Allergic rhinitis	✓		
Eosinophilia (≥4%)	✓		
Activity disturbance			✓
Shortness of breath			✓
Exercise-related wheeze/cough			✓
Aeroallergen-related wheeze/cough			✓

* - for stringent API only

Accepted Article

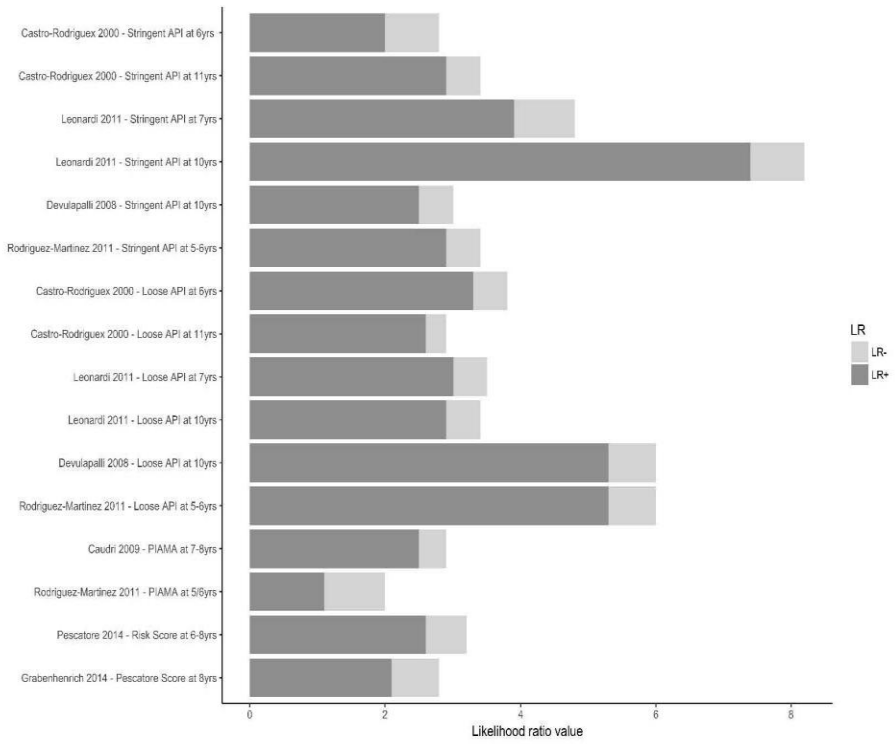
This article is protected by copyright. All rights reserved.

Accepted Article



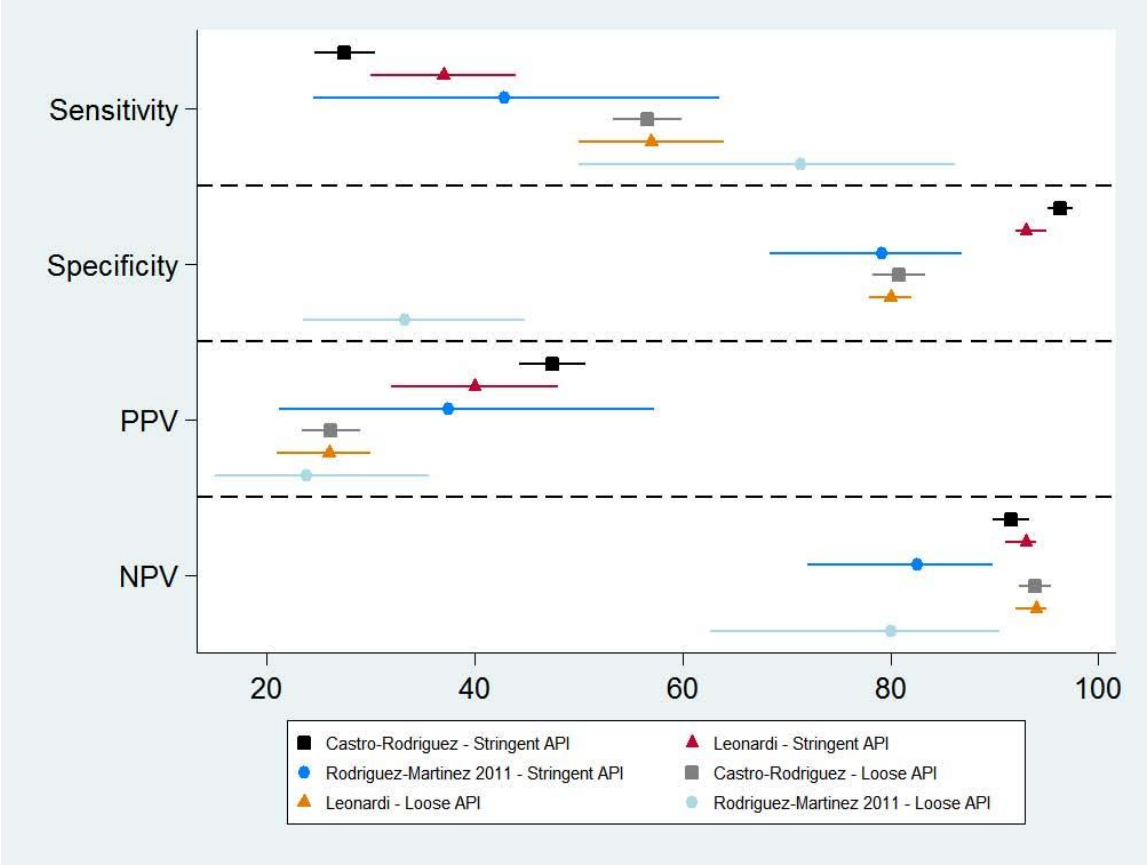
This article is protected by copyright. All rights reserved.

Accepted Article



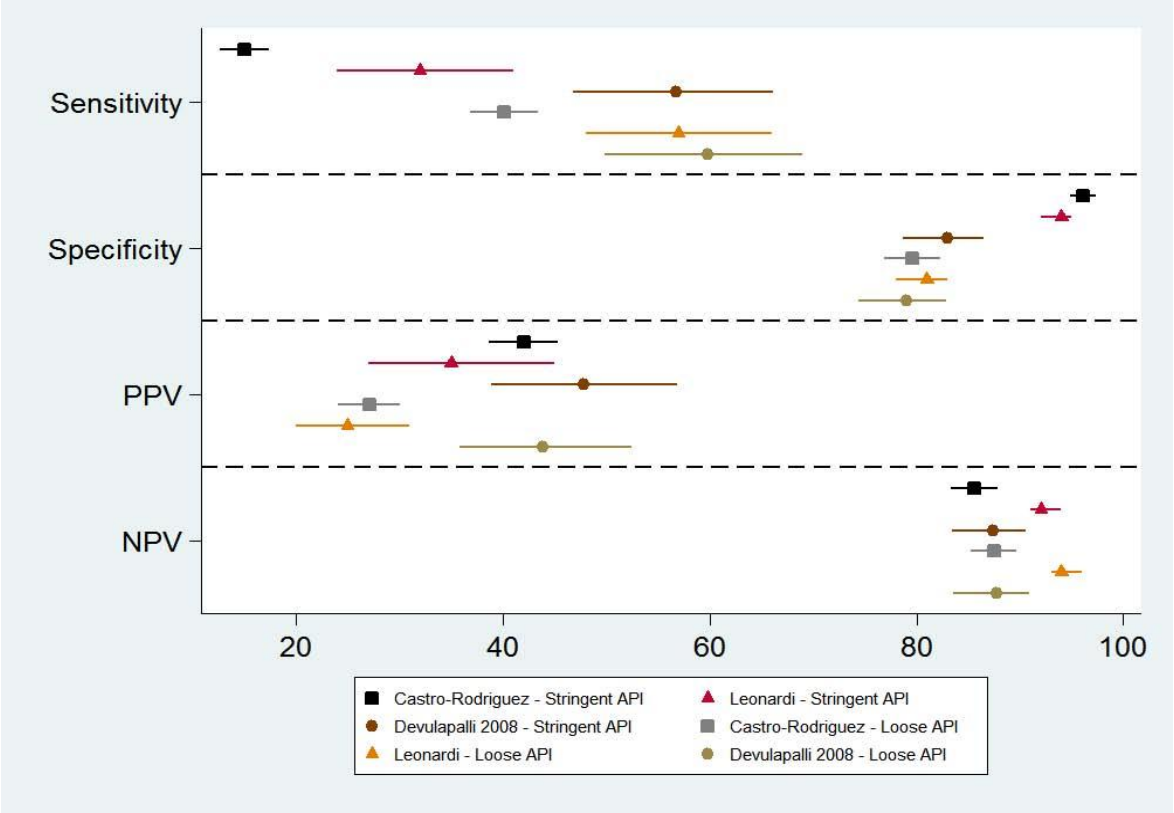
This article is protected by copyright. All rights reserved.

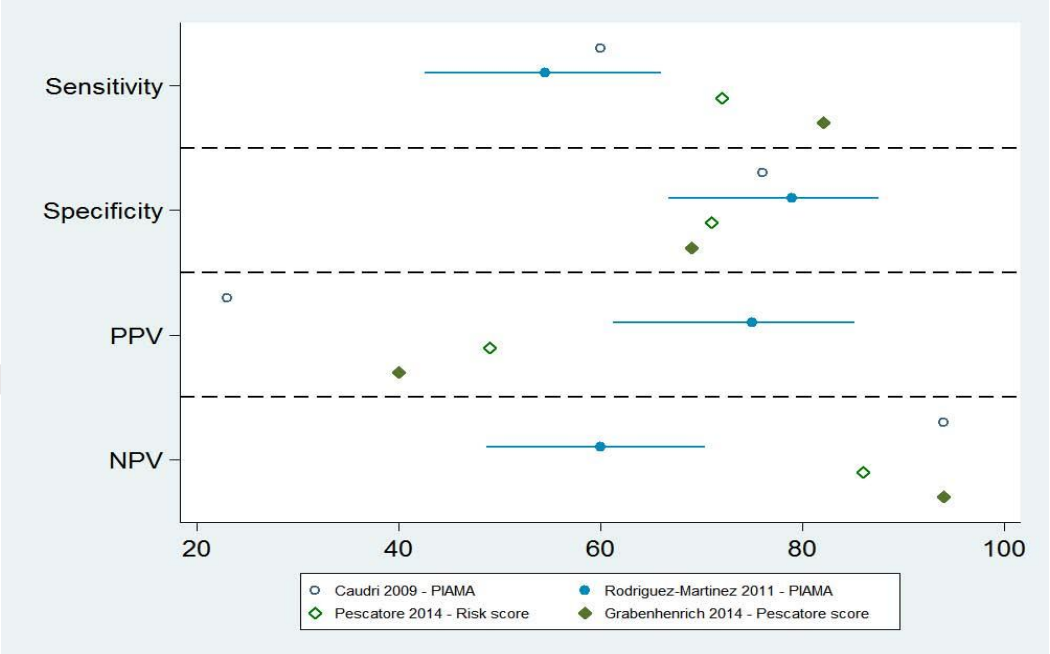
Accepted Article



This article is protected by copyright. All rights reserved.

Accepted Article





This article is protected by copyright. All rights reserved.

Validation of childhood asthma predictive tools: a systematic review

Silvia Colicino[1*] and Daniel Munblit[2,3,4*], Cosetta Minelli[1], Adnan Custovic[2], Paul Cullinan[1]

1 National Heart and Lung Institute, Imperial College London, United Kingdom
2 Department of Paediatrics, Imperial College London, United Kingdom
3 Faculty of Paediatrics, I. M. Sechenov First Moscow State Medical University, Russia Federation
4 International Inflammation (in-FLAME) network of the World Universities Network

Correspondence: s.colicino@ic.ac.uk
***** These authors contributed equally to this work.

Web extra Material

Figure E1: Search on 14-7-17 - Ovid MEDLINE(R) 1946 to July Week 2 2015

#	Searches	**Results**
1	exp asthma/	118835
2	(asthma* or wheez*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]	167279
3	1 or 2	167279
4	("predictive score*" or "prediction score*" or "predictive index*" or "prediction index*" or "predictive model*" or "prediction model*" or "predictive ability" or "prediction ability" or "predictive performance" or "prediction performance").mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]	35058
5	3 and 4	317
6	("asthma score*" or "asthma risk score*" or "asthma index*" or "asthma prediction*" or "prediction of asthma" or "asthma predictive" or "wheez* score*" or "wheez* index*" or "prediction of wheez*").mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]	477
7	5 or 6	702
8	**limit 7 to humans and "all child (0 to 18 years)"**	437

Figure E2: Search on 14-7-17 - Ovid Embase Classic+Embase 1947 to 2017 July 11

#	Searches	Results
1	exp asthma/	358605
2	(asthma* or wheez*).mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]	448144
3	exp wheezing/	34122
4	1 or 2 or 3	455184
5	("predictive score*" or "prediction score*" or "predictive index*" or "prediction index*" or "predictive model*" or "prediction model*" or "predictive ability" or "prediction ability" or "predictive performance" or "prediction performance").mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]	82746
6	4 and 5	900
7	("asthma score*" or "asthma risk score*" or "asthma index*" or "asthma prediction*" or "prediction of asthma" or "asthma predictive" or "wheez* score*" or "wheez* index*" or "prediction of wheez*").mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]	1192
8	6 or 7	1829
9	limit 8 to human	1451
10	**limit 9 to (infant <to one year> or preschool child <1 to 6 years> or school child <7 to 12 years> or adolescent <13 to 17 years>)**	809

Table E1: Outcome definition across studies *(✔= part of the asthma definition; -= is not included in the asthma definition)*

Source of study	Target population	ID	Prediction rule	Age at outcome assessment (years)	Asthma symptoms in the last 12 months**	Current wheeze*	Doctor's diagnosis of asthma (DDA)	Current asthma medications	Other
Development	General population	Castro-Rodriguez 2000[15]	API	6, 8, 11, 13	-	✔	✔	-	≥3 episodes of wheezing in the last 12 months
	High-risk population	Caudri 2009[22]	PIAMA	7 - 8	-	✔	✔	✔	
		Pescatore 2014[18]	Leicester tool	6 - 8	-	✔	-	✔	
Validation	General population	Leonardi 2011[33]	API	7 and 10	-	✔	✔	-	≥4 episodes of wheezing in the last 12 months
	High-risk population	Devulapalli 2008[34]	API	10	✔	✔	✔	✔	Positive exercise test
		Rodriguez-Martinez 2011[35]	API and/or PIAMA	5 - 6	-	✔	✔	✔	≥3 episodes of wheezing in the last 12 months
		Hafkamp-de Groen 2013[32]	PIAMA	6	-	✔	✔	-	≥4 episodes of wheezing in the last 12 months
		Grabenhenrich 2014[36]	Leicester tool	8	-	✔	-	✔	

* Current wheeze is defined as wheezing or whistling in the chest in the last 12 months

**Asthma symptoms in the last 12 months is defined as wheezing apart from colds or cough or chest tightness

Abbreviations – DDA: doctor's diagnosis of asthma

Castro-Rodriguez 2000[15]: current wheeze (at least on episode of wheeze) and DDA or ≥3 episodes of wheezing in the last 12 months regardless of DDA***; Caudri 2009[22]:*** current wheeze (at least on episode of wheeze) or DDA or current asthma medications; ***Pescatore 2014[18]:*** current wheeze and current asthma medications; ***Leonardi 2011[33]:*** current wheeze and DDA or ≥4 episodes of wheezing in the last 12 months regardless of DDA***; Devulapalli 2008[34]:*** at least two of the following three criteria: dyspnoea, chest tightness and/or wheezing from 0 to 3 years and/or at 4-10 years of age or DDA or use of asthma medication at 0–3 years and/or 4– 10 years of age and one of the following criteria: dyspnoea, chest tightness and/or wheezing in the last 12 months or use of asthma medication in the last 12 months or positive exercise test; ***Rodriguez-Martinez 2011[35]:*** current wheeze and DDA or ≥3 episodes of wheezing in the last 12 months regardless of DDA (for API) / current wheeze or DDA or current asthma medications (for PIAMA); ***Hafkamp-de Groen 2013[32]:*** DDA and wheezing in the past 12 months or 4 or more episodes of wheezing in the last 12 months; ***Grabenhenrich 2014[36]:*** current wheeze and current asthma medications

Table E2: Statistical methods used among studies

ID	Model development	Prediction rule	Missing data	Model evaluation	External validation study
No modelling					
Castro Rodriguez 2000[15]	Existing knowledge to select significant predictors Chi-square test to compare proportions	Stringent API and Loose API	Handling of missing data: complete case Proportion of missing: 20.9% for Loose API and 19.6% for Stringent API	None	*Devulapalli 2007*[34]*;* *Rodriguez-Martinez 2011*[35]*;* *Leonardi 2011*[33]
Logistic regression					
Caudri 2009[22]	Existing knowledge, univariate analysis (p-value<0.05) and stepwise method based on LRT (p-value<0.01) Multivariable logistic model	PIAMA	Handling of missing data: complete case analysis and imputation using MICE Proportion of missing: predictors 15%; outcomes 12%. In total each predictor has less than 2%	Internally validated by bootstrap sampling method (200 bootstrap samples)	*Rodriguez-Martinez 2011*[35]*;* *Hafkamp-de Groen 2013*[32]
Advanced methods					
Pescatore 2014[18]	Existing knowledge and only those that were available in primary care and did not require repeated observations were selected LASSO regression	Leicester tool	Handling of missing data: NR Proportion of missing: Table E2[18]	Internally validated by leave-one-out cross-validation technique	*Grabenhenrich 2014*[36]

Table E3: Performance measures among studies and target populations

Source of study	Target population	ID	Prediction rule	Age at outcome assessment (years)	% Sensitivity (95% CI)	% Specificity (95% CI)	% PPV (95% CI)	% NPV (95% CI)	LR+	LR-	Performance measurements
Development	General population	Castro-Rodriguez J. A. 2000[15]	Stringent API	6	28 (25-31)	96 (95-98)	48 (44-51)	92 (90-93)	7.4*	0.8*	NR
				11	15 (13-17)	96 (95-97)	42 (39-45)	86 (83-88)	3.9*	0.9*	
			Loose API	6	57 (53-60)	81 (78-83)	26 (23-29)	94 (92-95)	2.9*	0.5*	
				11	40 (37-43)	80 (77-82)	27 (24-30)	88 (85-90)	2.0*	0.8*	
	High-risk population	Caudri 2009[22]	PIAMA	7-8	60 (NR)	76 (NR)	23 (NR)	94 (NR)	2.5*	0.5*	AUC=0.74
		Pescatore 2014[18]	Leicester tool	6-8	72 (NR)	71 (NR)	49 (NR)	86 (NR)	2.5	0.4	Brier = 0.16; AUC = 0.74
Validation	General population	Leonardi 2011[33]	Stringent API	7	37 (30-44)	93 (92-95)	40 (32-48)	93 (91-94)	5.3*	0.7*	Brier=16; AUC=0.65
				10	32 (24-41)	94 (92-95)	35 (27-45)	92 (91-94)	5.3*	0.7*	Brier=11; AUC=0.63
			Loose API	7	57 (50-64)	80 (78-82)	26 (21-30)	94 (92-95)	2.9*	0.5*	Brier=12; AUC=0.68
				10	57 (48-66)	81 (78-83)	25 (20-31)	94 (93-96)	3.0*	0.5*	Brier=10; AUC=0.69
	High-risk population	Devulapalli 2008[34]	Stringent API	10	57 (47-66)	83 (79-87)	48 (39-57)	87 (83-91)	3.3*	0.5*	NR
			Loose API		60 (50-69)	79 (74-83)	44 (36-53)	88 (84-91)	2.9*	0.5*	NR
		Rodriguez-Martinez 2011[35]	Stringent API	5-6	43 (25-64)	79.2 (68-87)	38 (21-57)	83 (72-90)	2.1	0.7	NR
			Loose API		71 (50-86)	33.3 (24-45)	24 (15-36)	80 (63-91)	1.1	0.9	NR
			PIAMA		55 (43-66)	79 (67-88)	75 (61-85)	60 (49-70)	2.6	0.6	NR
		Hafkamp-de Groen 2013[32]	PIAMA	6	NR	NR	NR	NR	NR	NR	AUC=0.74
		Grabenhenrich 2014[36]	Leicester tool	8	82 (NR)	69 (NR)	40 (NR)	94 (NR)	2.6	0.3	Brier= 0.22; AUC=0.83

*Derived measures to compared performance among studies and predictive tools

Abbreviations – PPV: positive predictive value; NPV: negative predictive value; LR+: positive likelihood ratio; LR-: negative likelihood ratio; AUC: area under the ROC curve; 95% CI: 95% confidence interval; NR: not reported

Table E4: Critical appraisal of the eight selected studies based on the CHARMS checklist

ID	Participant selection	Outcome	Predictors	Sample flow/attrition	Analysis
Castro-Rodriguez 2000[15]	L	L	L	M	M
Caudri 2009[22]	L	L	L	M	L
Pescatore 2014[18]	L	L	L	M	L
Leonardi 2011[33]	L	L	L	M	L
Devulapalli 2008[34]	L	L	L	M	M
Rodriguez-Martinez 2011[35]	M	L	L	H	M
Hafkamp-de Groen 2013[32]	L	L	L	M	M
Grabenhenrich 2014[36]	L	L	L	M	L

Abbreviations – L: low risk of bias; M: medium risk of bias; H: high risk of bias