Open Access Full Text Article

REVIEW

Prediction models for the development of COPD: a systematic review

Melanie C Matheson,[1,2,*] Gayan Bowatte,[1,3,*] Jennifer L Perret,[1,4] Adrian J Lowe,[1,2] Chamara V Senaratna,[1,5] Graham L Hall,[6–8] Nick de Klerk,[6,8] Louise A Keogh,[9] Christine F McDonald,[4] Nilakshi T Waidyatillake,[1] Peter D Sly,[10] Deborah Jarvis,[11,12] Michael J Abramson,[13] Caroline J Lodge,[1,2,*] Shyamali C Dharmage[1,2,*]

[1]Allergy and Lung Health Unit, Centre for Epidemiology and Biostatistics, School of Population and Global Health, University of Melbourne, Melbourne, VIC, Australia; [2]Murdoch Children's Research Institute, Melbourne, VIC, Australia; [3]National Institute of Fundamental Studies, Kandy, Sri Lanka; [4]Department of Respiratory and Sleep Medicine, Institute for Breathing and Sleep, Austin Health, University of Melbourne, Melbourne, VIC, Australia; [5]Department of Community Medicine, University of Sri Jayewardenepura, Nugegoda, Sri Lanka; [6]Telethon Kids Institute, Perth, WA, Australia; [7]School of Physiotherapy and Exercise Science, Curtin University, Perth, WA, Australia; [8]Centre of Child Health Research, University of Western Australia, Perth, WA, Australia; [9]Centre for Health Equity, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, VIC, Australia; [10]Child Health Research Centre, The University of Queensland, Brisbane, QLD, Australia; [11]MRC-PHE Centre for Environment and Health, Imperial College London, London, UK; [12]Population Health and Occupational Diseases, National Heart and Lung Institute, Imperial College London, London, UK; [13]School of Public Health & Preventive Medicine, Monash University, Melbourne, VIC, Australia

*These authors contributed equally to this work

Correspondence: Shyamali C Dharmage
Allergy and Lung Health Unit, Centre for Epidemiology and Biostatistics, School of Population and Global Health, University of Melbourne, 207 Bouverie Street, Carlton, VIC 3010, Australia
Tel +61 3 8344 0737
Email s.dharmage@unimelb.edu.au

Abstract: Early identification of people at risk of developing COPD is crucial for implementing preventive strategies. We aimed to systematically review and assess the performance of all published models that predicted development of COPD. A search was conducted to identify studies that developed a prediction model for COPD development. The Checklist for Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies was followed when extracting data and appraising the selected studies. Of the 4,481 records identified, 30 articles were selected for full-text review, and only four of these were eligible to be included in the review. The only consistent predictor across all four models was a measure of smoking. Sex and age were used in most models; however, other factors varied widely. Two of the models had good ability to discriminate between people who were correctly or incorrectly classified as at risk of developing COPD. Overall none of the models were particularly useful in accurately predicting future risk of COPD, nor were they good at ruling out future risk of COPD. Further studies are needed to develop new prediction models and robustly validate them in external cohorts.

Keywords: COPD, early detection, predictors and risk prediction models

Introduction

COPD is a progressive debilitating lung condition with major impact on both morbidity and early mortality. Of global concern, COPD is projected to rank seventh in worldwide disease burden and as the third leading cause of death by 2030.[1] The burden of COPD is projected to rise in Africa and Asia due to the rising prevalence of smoking in these regions and high levels of air pollution.[2]

COPD carries a very poor prognosis with ~30% of those hospitalized for an exacerbation dying within 2 years[3] and an in-hospital case fatality rate of ~15%.[4] These high fatality rates may be directly related to under-diagnosis of early-stage COPD in the community, as those presenting to tertiary care generally have advanced disease.[5,6] There is an urgent need to implement strategies to better identify those at increased risk of COPD or those who have early disease.

There are several well-known risk factors for COPD including tobacco smoking;[7] occupational exposures to gases, dusts, and fumes;[8] and genetic factors, such as alpha-1-antitrypsin deficiency.[9] Smoking is strongly associated with COPD, however there is known individual variation in smoking effects on the lung with up to 30% of smokers never developing COPD.[10] There is increasing evidence that early life factors are related to reduced growth and early decline in lung function, both of which are associated with a greater risk of developing COPD at an earlier age.[11]

Current clinical practice guidelines in the US, UK, and Australia do not recommend screening asymptomatic adults for COPD using spirometry.[12] From the primary

prevention perspective, predicting those at increased risk of developing COPD can allow for implementation of interventions which may not only prevent COPD developing, but may also help preserve lung function and quality of life in those who do go on to develop COPD. Risk prediction tools that incorporate the best available evidence in order to stratify patients based on their individual risk profiles could facilitate early targeted interventions and management. These tools would also facilitate early detection, accurate diagnosis, and determination of prognosis.

There have been several reviews of predictive models for determining COPD prognosis[12,13] but no systematic reviews have investigated the potential of models for prediction of development of COPD. We aimed to systematically synthesize the evidence from studies that have investigated prediction models for subsequent development of COPD in those without a prior diagnosis.

Methods
Search strategy and selection criteria
EMBASE and Medline (PubMed) databases were systematically searched from inception to November 30, 2016 (Supplementary material). Reference lists of articles selected for full-text reading (30 articles) were manually searched for additional eligible articles. This review is reported in accordance with the recommendations set forth by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement[14] and was prospectively registered in PROSPERO systematic review registry (registration number 42017064447). Details of inclusion and exclusion criteria are given in Supplementary material.

Data extraction and critical appraisal
Two independent reviewers (GB and CS) ran the same search strategy in PubMed and EMBASE, independently screened titles and abstracts, assessed full-texts of eligible articles, and extracted data following the Checklist for Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies (CHARMS).[15,16] Development of the research questions for the systematic review was based on the CHARMS checklist (Table 1). For the selected studies, data were extracted for study design, derivation cohort, predictors, outcomes, and performance of prediction models in the model derivation and validation cohorts. Methods of risk of bias and applicability assessments are given in Supplementary material.

Table 1 Systematic review questions developed using the CHARMS checklist

Item	Systematic review characteristics
1. Type of prediction model	Prognostic model to predict development of COPD
2. Intended scope of the review	To identify adults (>18 years) who will or will not develop COPD in later adult life, to help in early detection, closer monitoring, and therapeutic decision making
3. Type of prediction modeling studies	Studies of prediction models developed with and without external validation in independent data. Prediction models developed for predicting risk of development of COPD and studies updating and validating previous risk prediction models
4. Target population	Adults in the general population, adults with high risk of development of COPD (smokers, asthmatics)
5. Outcome to be predicted	COPD diagnosis using spirometry by FEV_1/FVC <70% or by symptoms or by FEV_1/FVC < LLN
6. Time span	Predictors measured in adults and subsequent outcomes
7. Intended moment of using the model	Model to be used in early detection of COPD in general population or populations at risk. By primary care practitioners and respiratory physicians

Abbreviations: CHARMS, Checklist for Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies; FEV_1, forced expiratory volume in one second; FVC, forced vital capacity; LLN, lower limit of normal.

Results
Overview of the studies included in the review
Our search identified 4,481 non-duplicate records, from which 30 were selected for full-text review (Figure 1). Only four of these were selected for inclusion, data extraction, and synthesis. The four models developed by these studies differed on many characteristics including the derivation cohort (age, nationality, at-risk), predictors used, statistical methods, and COPD definitions[17–20] (Table 2).

Study type and population
All the model development studies included in this review concerned the development of original prediction models with the objective of developing a risk prediction tool for the development of COPD (Table 2). Of the four studies, one was from a general population cohort,[18] two were from electronic medical records databases (one from primary care data[20] and one from hospital record data of asthma patients),[19] and one was a hospital-based case–control study with participants selected from the respiratory medicine department.[17] One study did not report any specific method for the selection of the initial model predictors,[18] two studies used predictors previously
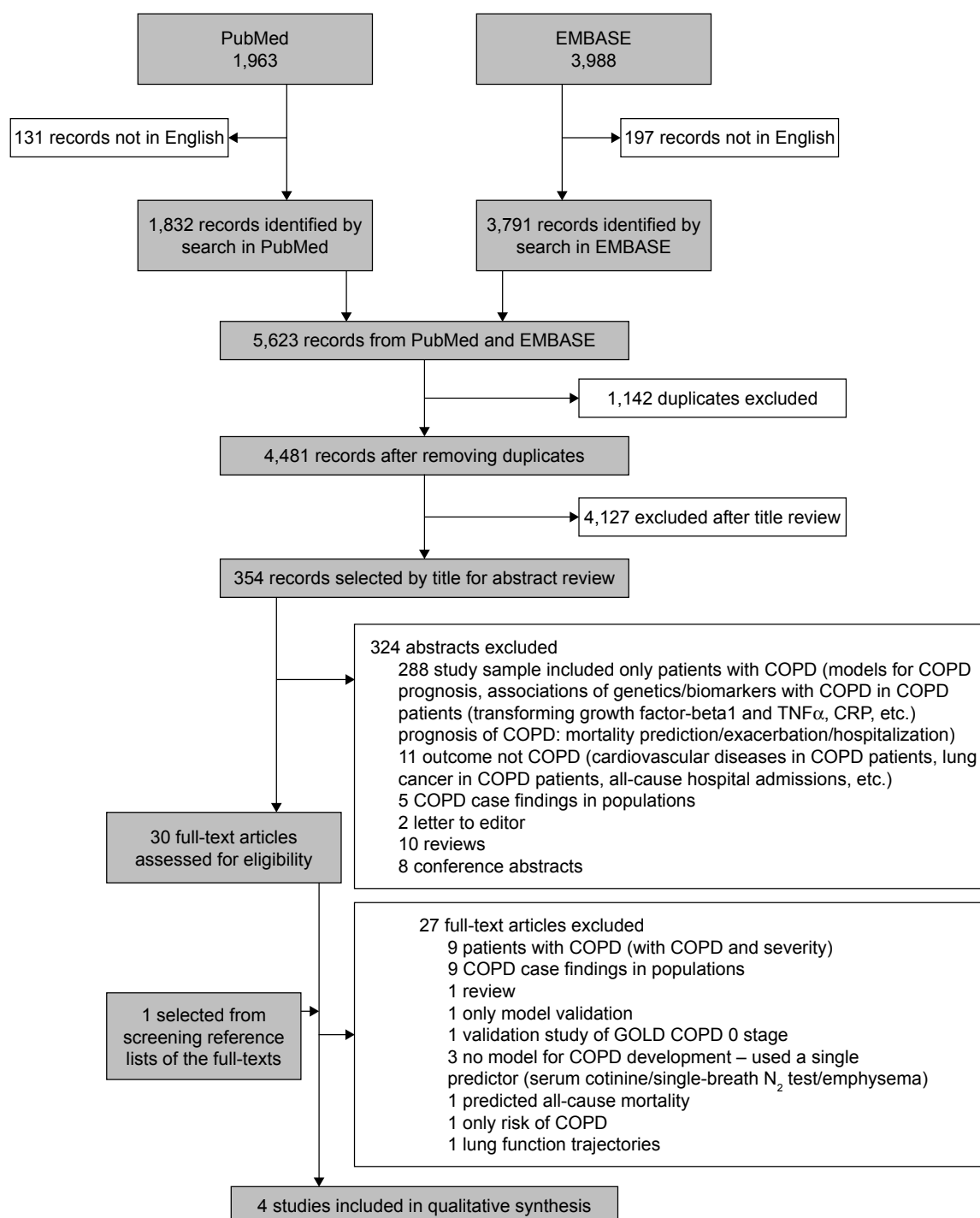
Figure 1 PRISMA flow diagram.
Abbreviations: CRP, C-reactive protein; TNFα, tumor necrosis factor α; GOLD, Global Initiative for Chronic Obstructive Lung Disease.

reported in the literature,[17,20] and the remaining study used a Bayesian network approach to identify predictors.[19]

Predictors included in all models

All four studies included smoking as a predictor (Table 3).[17–20] The studies by Guo et al,[17] Kotz et al,[20] and Himes et al[19]

all defined smoking as a binary variable; either ever or never smoker. Higgins et al[18] reported smoking in three ways: 1) any cigarette smoking (non-smoker, current smoker, ex-smoker), 2) cigarettes per day (categories of none; 1–19; 20+), and 3) change in cigarette smoking between the baseline and follow-up study.

Table 2 Study characteristics of the selected prediction models

Reference		Guo et al, 2015[17]	Kotz et al, 2014[20]	Himes et al, 2009[19]	Higgins et al, 1982[18]
Study location		**China**	**UK**	**USA**	**USA**
Predictive model derivation cohort	Source population	Pulmonary medicine department of a hospital	General population	Asthma patients from general population	General population
	Study design	Case–control study	Primary care electronic database	Electronic hospital medical records database	Longitudinal cohort study
	Recruitment period	January 2012–December 2013	1998–2008	1988–1998	1962–1965 re-examined 1978–1979
	Age at inclusion	≥40 years	35–74	≥18 years	16–64 years
	No participants in model development	682 (331 COPD cases and 351 controls)	480,903	N=9,349 patients (843 cases, 8,506 controls)	2,995
	Methods to select predictors	Not specified	Literature	A Bayesian network	Not specified
Predictive model validation cohort	Type of validation	Internal/external	Internal	Internal	External validation
	Validation cohort	N=30 COPD patients, n=20 healthy controls	247,755	N=992 patients (46 cases, 946 controls)	N=20 patients with OAD, asthma, CB, or none

Abbreviations: OAD, obstructive airway disease; CB, chronic bronchitis.

Table 3 Potential risk factors for COPD considered for inclusion in the COPD risk prediction models

Reference	Guo et al, 2015[17]	Kotz et al, 2014[20]	Himes et al, 2009[19]	Higgins et al, 1982[18]
Demographic and clinical characteristics				
Age	–	✓6	✓	✓
Sex	✓	✓	✓	✓
Race	✓1	–	✓	–
SES status	–	✓7	–	✓
Height, weight, or BMI	✓	–	✓	✓
Lung function	✓	–	–	✓10
Personal or FHx lung diseases	✓	✓8	✓9	✓11
Lifestyle factors				
Smoking history	✓	✓	✓	✓
Alcohol				✓
Early life factors				
History of RI in childhood	✓2			
Low birth weight	✓3			
Other				
Environmental pollution	✓4			
Biomarkers	✓5			✓12

Notes: 1: All Chinese; 2: history of respiratory infections in childhood (yes/no); 3: low birth weight <2,500 g; 4: their place of residence and work environment; 5: SNPs genotyped rs2070600, rs10947233, rs1800629, rs2241712, rs1205; 6: age was categorized into 35–39, 40–44, 45–49, 50–54, 55–59, 60–64, and 65+ years; 7: Carstairs Index of Deprivation (coded 1= least deprived to 5= most deprived); 8: asthma; 9: all had a diagnosis of asthma; and 104 comorbidities were included in initial model development; 10: Vmax$_{50\%}$, FEV$_1$, FEV$_1$/FVC%; 11: frequent colds, chronic bronchitis, wheeze × cough × asthma × familial chronic bronchitis; 12: ABO blood group, Kell blood group.
Abbreviations: SES, socio economic status; BMI, body mass index; FHx, family history; RI, respiratory infections; SNPs, single nucleotide polymorphisms.

Predictors included in some models

One model included asthma as a predictor[20] and one study developed its model using a group of asthma patients.[19] Three studies included age as a predictor.[18–20] One used lung function (% predicted FEV$_1$)[18] and two studies used comorbidities (ie, "acute bronchitis and bronchiolitis", "pneumonia, organism unspecified", "shortness of breath", "respiratory distress or insufficiency", "diabetes mellitus", "acute upper respiratory infection", "viral and chlamydial infections", and "heart failure") or having respiratory infections as predictors (respiratory infections in childhood).[17,19] Included in the final model of Guo et al[17] were five single nucleotide polymorphisms (SNPs) (rs2070600, rs10947233, rs1800629, rs2241712, and rs1205) chosen out of a possible 76 identified in a preliminary genome wide association study (GWAS). Two studies developed models separately for men and women.[18,20]

Outcome definitions

The outcome definition of COPD varied between the different studies (Supplementary material). Two studies used spirometry to define COPD: one used the Global Initiative for Chronic Obstructive Lung Disease (GOLD) criterion of post-bronchodilator FEV$_1$/FVC ratio <0.70 as a cut-off[17] and one defined COPD as an FEV$_1$ <65% of the predicted value in combination with an FEV$_1$/FVC ratio <80%.[18] Two studies used disease classifications given in the electronic database systems which were based on clinical diagnosis with no further details provided.[19,20]

Critical appraisal

Risk of bias and applicability concerns in the selected studies were assessed using the CHARMS checklist (Table 4 and the details of CHARMS checklist [Supplementary material]).[15]

Of the four model development studies, three used long-term follow-up or long-term data collected in electronic databases, while one used cross-sectionally collected data. Although this latter study rated good quality based on reporting model development and validation, it lacked the ability to predict the risk of COPD in a temporal fashion.[17] Three of the four model development studies had a low risk of bias in the selection of participants and one had moderate risk of bias due to the use of a case–control study design, which generally leads to incorrect estimates of the model intercept or baseline hazard. For the selection of predictors, two studies had low risk and two had moderate risk of bias. One study rated moderate risk because 109 variables were considered leading to an event-per-variable ratio of <10,[19] potentially leading to model over-fitting. The other moderate risk study defined all variables as binary, potentially leading to the selection of spurious predictors and over-fitting. This study also did not provide detail on measurement and collection of predictor data.[17]

Outcome assessment in three of four studies had low risk of bias[18–20] as the outcomes were measured after predictor measurement. One had moderate risk of bias due to the synchronous measurement of outcome and predictors.[17] Risk of attrition was high in three studies that used a cohort design or electronic databases.[18–20] Two of the studies did not describe attrition and one study had more than 20% attrition. In the latter study, when baseline characteristics were compared between those followed-up and those who were not, lower smoking rates and better lung function were found in those followed-up.[18] All the studies had moderate risk of bias in relation to the analysis. No study accounted for over-fitting (ie, data for the model is well described but not applicable for new individuals) and optimism (ie, caused by over-fitting and leading to an over-estimate of the model's ability to predict the outcome in new individuals) by shrinkage or other methods; one study did not report on handling of missing data,[18] and only one study assessed calibration and discrimination.[20]

Development and presentation of the prediction model

At the model development stage, two studies selected predictors based on evidence from the literature and availability of predictors.[17,20] Another study a priori selected anthropometric predictors and smoking, and included all predictors of comorbidities recorded by the electronic database.[19] One study did not report on the criteria used to select predictors included in model development (Tables 3, 5, and Supplementary material).[18] In model building, two studies used stepwise multiple regression models,[17,18] one used Cox regression,[20] and another used Bayesian networks.[19]

Two of the studies that used electronic databases internally validated their models.[19,20] The case–control study model was externally validated.[17] The long-term cohort study validated the model using a small sample of hospital clinic patients (Table 3).[18] Three studies used complete case analysis, whereas the handling of missing data was not reported by the fourth.[17–20] To use COPD risk prediction tools in clinical practice, thresholds are needed to either rule

Table 4 Assessment of the risk of bias and applicability concerns based on the CHARMS checklist for the selected COPD risk prediction model studies

Measure	Reference	Guo et al, 2015[17]	Kotz et al, 2014[20]	Himes et al, 2009[19]	Higgins et al, 1982[18]
Risk of bias	Participant selection	M	L	L	L
	Predictor assessment	M	L	M	L
	Outcome assessment	M	L	L	L
	Attrition	N/A	H	H	H
	Analysis	M	M	M	M
Applicability concern	Participant selection	M	L	M	M
	Outcome	L	L	L	M
	Predictors	M	L	L	L
	Analysis	H	M	H	H
	Results	L	L	M	L
	Interpretation	L	L	L	L

Abbreviations: L, low risk of bias or applicability concern; M, moderate risk of bias or applicability concern; H, high risk of bias or applicability concern; CHARMS, Checklist for Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies; N/A, not available.

Table 5 Risk factors included in the final COPD risk prediction models

Reference	Guo et al, 2015[17]	Kotz et al, 2014[20]	Himes et al, 2009[19]	Higgins et al, 1982[18]
Demographic and clinical characteristics				
Age		✓6	✓	✓
Sex	✓1	✓*	✓	✓10
Race			✓	–
SES status		✓7	–	–
Height, weight or BMI			–	–
Lung function			–	✓
Personal or FHx lung diseases		✓8	✓9	–
Lifestyle factors				
Smoking history	✓	✓	✓	✓
Early life factors				
History of RI in childhood	✓2			
Low birth weight	✓3			
Other				
Environmental pollution	✓4			–
Biomarkers	✓5			

Notes: 1: All Chinese; 2: history of respiratory infections in childhood (yes/no); 3: low birth weight <2,500 g; 4: their place of residence and work environment; 5: SNPs genotyped rs2070600, rs10947233, rs1800629, rs2241712, rs1205; 6: age was categorized into 35–39, 40–44, 45–49, 50–54, 55–59, 60–64, and 65+ years; 7: Carstairs Index of Deprivation (coded 1= least deprived to 5= most deprived); 8: asthma; 9: asthma not included as a predictor, derivation and validation cohorts include patients diagnosed with asthma, but included eight comorbidities ("acute bronchitis and bronchiolitis", "pneumonia, organism unspecified", "shortness of breath", "respiratory distress or insufficiency", "diabetes mellitus", "acute upper respiratory infection", "viral and chlamydial infections", and "heart failure"); 10: males % Vmax$_{50}$ and females % FEV$_1$. *Models derived for males and females separately.

in or rule out COPD. However, only one[17] study reported clear cut-off points for determining COPD in their models (Supplementary material). One study provided a score chart with regression coefficients,[20] and one reported a predictive network based on Bayesian statistics (Supplementary material).[19]

Performance of the prediction model

The performance of the models was reported in several different ways including a narrative description,[18] graphical depictions,[19,20] and quantitative performance estimates.[17,19,20] Two studies reported calibration using the Hosmer–Lemeshow test; one did not report the *p*-value[18] and the other reported *p*=0.86, that is, no significant deviation between the observed and predicted events of COPD.[17] Two studies reported area under the receiver operating characteristic curve (ROC AUC); the discriminatory ability of the test to correctly classify those

with and without COPD. One study reported an overall ROC AUC of 0.83[19] and another study reported the ROC AUC for females as 0.85 and for males ROC AUC as 0.83.[20]

Of the four studies only one reported sensitivity, specificity, false positive, and false negative rates for a selected cut-off.[17] This model reported good sensitivity (83%) and specificity (85%). For two studies, we were able to calculate the sensitivity and specificity.[19] No study reported positive or negative predictive values. We calculated positive and negative likelihood ratios using the calculated or provided sensitivity and specificity for three studies.[17,19,20] The positive likelihood ratio ranged from 1.85 to 5.53 and negative likelihood ratio ranged from 0.04 to 0.22 (Table 6 and Supplementary material).

Applicability

Three of the four studies had a moderate risk of applicability issues related to selection of participants.[17–19] Two were in specific population groups which limited their models' wider applicability; one in a Chinese population including SNPs which were likely to be specific for Asian populations[17] and, another in asthmatics.[19] The third study was performed 30 years ago and so has lower applicability to current COPD risk.[18] One study had applicability issues related to selection of predictors. The classification of predictors as binary variables can lead to the selection of spurious predictors in the model and reduce applicability to new patients.[17]

All the studies had high or moderate applicability concerns in relation to the analysis. No study used methods to address the possibility of model overfitting. This may limit their applicability to be used in new patients.[17–20] Only one study assessed both calibration and discrimination.[20] Two studies reported calibration via a Hosmer–Lemeshow test[17,18] but neither of these studies reported any assessment of discrimination.

Discussion

From a systematic search of the existing literature, we identified only four models that aimed to predict an individual's future risk of COPD. The models differed significantly in predictors used, outcome definitions, and populations from which they were developed. Overall the models performed well using a limited number of predictors; age, sex, smoking, and lung function. However, few were validated in external populations and only one included novel risk factors such as genetic markers.

Three models were developed in Western countries and the fourth in China. China now consumes over a third of the world's tobacco[21] and Chinese men have a rapidly increasing death rate from tobacco-related causes.[22] This huge burden

Table 6 Performance of the COPD risk prediction tools based on derivation cohort

Reference	Guo et al, 2015[17]	Kotz et al, 2014[20]^		Himes et al, 2009[19],^^	Higgins et al, 1982[18]
		Females	**Males**		
Calibration (p-value)	0.86	–		–	**
Discrimination (ROC AUC and 95% CI)	–	0.845 (0.840–0.850)	0.832 (0.827–0.837)	0.83	–
Sensitivity (%)	0.83	0.85	0.85	0.98	–
Specificity (%)	0.85	0.71	0.68	0.47	–
False positive (%)	0.15	–	–	–	–
False negative (%)	0.16	–	–	–	–
Positive likelihood ratio	5.53	2.93	2.66	1.85	–
Negative likelihood ratio	0.2	0.21	0.22	0.04	–

Notes: ^Calculated using Youden index (= Sensitivity + Specificity − 1) for the best cut-off. ^^Measures of model performance calculated from ROC curve. **p-value for measure of calibration not reported.

of smoking-related disease is likely to also impact COPD prevalence in China and with similar smoking trends in other Asian countries is a burgeoning epidemic. Regional variance in smoking legislation between Asian and Western countries will also affect generalizability of COPD risk prediction models.

All studies included sex in their final models and two studies derived separate models for females and males.[18,20] One model identified that smoking in women led to a greater increase in risk of COPD than it did for men.[20] Evidence suggests that women are more susceptible to the risks of smoking and occupational exposures.[23,24] They receive a greater dose of toxin for the same amount of inhaled smoke because of their smaller airway size.[25] Another possible explanation is the role of hormonal factors during the transitional and postmenopausal periods which are associated with a more rapid decline in lung function and may increase the risk of COPD in older women.

Surprisingly only one model included lung function parameters. There is mounting evidence that low lung function predicts subsequent risk of COPD. We have shown that childhood lung function predicts COPD by middle age.[26] A study of three longitudinal cohorts found that, of those with COPD at the end of the follow-up period, 50% already had a low FEV_1 at the age of 40 years.[27] A study of lung function trajectories from the Tuscon cohort[28] found that those with the persistently low trajectory reached a maximal level of lung function that was 10% lower than those in the normal trajectory.[28] These data provide strong support for the use of lung function measures in prediction of subsequent COPD risk and also provide support for the examination of other early life factors as important potentially modifiable determinants of COPD risk.[29,30]

Smoking is an important risk factor for COPD and was included in all models. Smoking is known to increase the

risk of COPD and in clinical guidelines is a defining criterion for the diagnosis of COPD.[31] However, up to 45% of people with COPD are non-smokers and only 30% of smokers will develop the disease.[10] Most studies included smoking as a binary measure of ever or never smoking. However, Higgins et al[18] included a measure of change in smoking status, a useful predictor for clinical practice as it gives clinicians estimates based on the patient's current smoking status but also an estimate of risk associated with making changes to those characteristics. Apart from the likely increased precision of a model nuanced for smoking amount and duration, the inclusion of such predictors can enable clinicians to more realistically tailor risk management advice to the individual.

The inclusion of a measure of socioeconomic status that is transferable between different populations is difficult. Measures of socioeconomic status reflect a combination of factors including lifestyle, occupational, environmental, and demographic characteristics. Only one model included a measure of socioeconomic status, defined by a regional specific measure of deprivation. Alternative options include measures of educational level, income level, or number of people living in the home.

The model from the Chinese population considered air pollution exposure as a potential predictor, although this was not included in the final model. There is strong evidence that air pollution is associated with increased risk of COPD.[32,33] Other environmental factors such as occupational exposure,[32] were not considered by any study. Given the strong evidence supporting the role of exposure to dusts, gases, and fumes in the risk of COPD,[8] this is a risk factor that should be considered.

Beyond smoking, few models considered other lifestyle factors. One considered alcohol consumption, but it was not included in the final model. Only one model considered

childhood factors including childhood respiratory infections and low birth weight. Early life factors have important potential as future risk predictors. There is increasing evidence that low birth weight has an impact on asthma risk in middle life[34] and other early life factors and derived indices[30] have been shown to impact on risk of COPD into adulthood.

Basic models, including only a few selected predictors, may be easier to introduce into clinical practice, but they do not make the best use of all relevant clinical knowledge to predict COPD. However, the use of an extended model with a large number of clinical tests may be burdensome for patients, clinicians, and health resources, so a balance between the two is needed.

The main sources of bias related to the statistical analyses performed and how they were presented in the manuscripts. There was often insufficient information to fully assess the development and validation of the models. Objective measures of bias and performance did not rate well in any of the studies. We could not assess the performance of the prediction models quantitatively, because the characteristics of the derivation cohorts, the prediction models themselves, and reporting of model performance varied widely between the studies. Two models had good ability to discriminate between people who were correctly or incorrectly classified as having COPD (concordance statistic 0.830–0.845).[19,20]

The clinical usefulness of a model requires the determination of positive and negative likelihood ratios in order to determine the ability to rule in COPD versus ruling it out. The positive and negative likelihood ratios were presented in one study and calculable in two. The models had a small to moderately increased ability to detect future COPD risk (positive likelihood ratios ranged between 1.85 and 5.53), and a small to moderate chance of failing to identify someone at risk of future COPD (negative likelihood ratios ranged from 0.04 to 0.22). Overall these results suggest the models developed to date have poor discriminatory ability to predict future COPD risk.

The weak predictive ability of the models in this study highlights the need for future research and the development of more comprehensive models. The inclusion of early life and childhood factors could improve the discriminatory ability of the models. Model development strategies, including the comprehensive validation of all existing models and any newly developed models in a single external study population, are needed. This would allow for direct model comparison of predictors across all cohorts. Future COPD risk prediction models could incorporate new and better predictors including genetic risk prediction scores. Risk prediction model studies in cancer have shown that the addition of

susceptibility SNPs can improve the discriminatory power of established risk prediction models by more than 20%.[33] One study in this review used selected SNPs identified from a GWAS study of COPD, however the authors did not compare the discriminatory ability of the model with and without the inclusion of the SNPs and so it was not possible to determine how they improved its overall performamce.

In conclusion, our review identified only four models for the prediction of future COPD risk. The models included the most important known predictors of smoking, sex, and age, but pre-existing lung function measurements and other important predictors were not routinely included. Overall none of the models were particularly accurate at predicting future risk of COPD, nor were they good at ruling out future COPD risk. With the emergence of new evidence that low lung function can start in early life and that early life asthma can predict not only low lung function levels into adulthood, but also an increased risk of COPD, future COPD risk prediction models will need to incorporate some of these important early life risk factors.

Acknowledgment

Author contributions

MC Matheson is the guarantor of the paper. MC Matheson, SC Dharmage, CJ Lodge, G Bowatte, and CV Senaratna: conception, design, literature search, data interpretation and writing the manuscript.

JL Perret, AJ Lowe, GL Hall, PD Sly, N de Klerk, L Keogh, NT Waidyatillake, CF McDonald, D Jarvis, and MJ Abramson: conception, design, and writing the manuscript. All authors contributed toward data analysis, drafting and revising the paper and agree to be accountable for all aspects of the work.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Med*. 2006;3(11):e442.
2. Chan-Yeung M, Ait-Khaled N, White N, Ip MS, Tan WC. The burden and impact of COPD in Asia and Africa. *Int J Tuberc Lung Dis*. 2004; 8(1):2–14.

3. Singanayagam A, Schembri S, Chalmers JD. Predictors of mortality in hospitalized adults with acute exacerbation of chronic obstructive pulmonary disease. *Ann Am Thorac Soc*. 2013;10(2):81–89.

4. Hoogendoorn M, Hoogenveen RT, Rutten-van Molken MP, Vestbo J, Feenstra TL. Case fatality of COPD exacerbations: a meta-analysis and statistical modelling approach. *Eur Respir J*. 2011;37(3):508–515.

5. Abramson MJ, Schattner RL, Sulaiman ND, Del Colle EA, Aroni R, Thien F. Accuracy of asthma and COPD diagnosis in Australian general practice: a mixed methods study. *Prim Care Respir J*. 2012;21(2):167–173.

6. Toelle BG, Xuan W, Bird TE, et al. Respiratory symptoms and illness in older Australians: the Burden of Obstructive Lung Disease (BOLD) study. *Med J Aust*. 2013;198(3):144–148.

7. Samet JM. The surgeon generals' reports and respiratory diseases. From 1964 to 2014. *Ann Am Thorac Soc*. 2014;11(2):141–148.

8. Vestbo J, Hurd SS, Agusti AG, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med*. 2013;187(4): 347–365.

9. Stoller JK. Alpha-1 antitrypsin deficiency: an underrecognized, treatable cause of COPD. *Cleve Clin J Med*. 2016;83(7):507–514.

10. Salvi SS, Barnes PJ. Chronic obstructive pulmonary disease in non-smokers. *Lancet*. 2009;374(9691):733–743.

11. McGeachie MJ, Yates KP, Zhou X, et al. Patterns of growth and decline in lung function in persistent childhood asthma. *N Engl J Med*. 2016;374(19):1842–1852.

12. Guirguis-Blake JM, Senger CA, Webber EM, Mularski RA, Whitlock EP. Screening for chronic obstructive pulmonary disease: evidence report and systematic review for the US Preventive Services Task Force. *JAMA*. 2016;315(13):1378–1393.

13. Han MK, Steenrod AW, Bacci ED, et al. Identifying patients with undiagnosed COPD in primary care settings: insight from screening tools and epidemiologic studies. *Chronic Obstr Pulm Dis*. 2015;2(2):103–121.

14. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med*. 2009;151(4):264–269.

15. Moons KG, de Groot JA, Bouwmeester W, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Med*. 2014;11(10):e1001744.

16. Smit HA, Pinart M, Antó JM, et al. Childhood asthma prediction models: a systematic review. *Lancet Respir Med*. 2015;3(12):973–984.

17. Guo YI, Qian Y, Gong YI, Pan C, Shi G, Wan H. A predictive model for the development of chronic obstructive pulmonary disease. *Biomed Rep*. 2015;3(6):853–863.

18. Higgins MW, Keller JB, Becker M, et al. An index of risk for obstructive airways disease. *Am Rev Respir Dis*. 1982;125(2):144–151.

19. Himes BE, Dai Y, Kohane IS, Weiss ST, Ramoni MF. Prediction of chronic obstructive pulmonary disease (COPD) in asthma patients using electronic medical records. *J Am Med Inform Assoc*. 2009;16(3): 371–379.

20. Kotz D, Simpson CR, Viechtbauer W, van Schayck OC, Sheikh A. Development and validation of a model to predict the 10-year risk of general practitioner-recorded COPD. *NPJ Prim Care Respir Med*. 2014;24:14011.

21. Yang G. Marketing 'less harmful, low-tar' cigarettes is a key strategy of the industry to counter tobacco control in China. *Tob Control*. 2014; 23(2):167–172.

22. Chen Z, Peto R, Zhou M, et al. Contrasting male and female trends in tobacco-attributed mortality in China: evidence from successive nationwide prospective cohort studies. *Lancet*. 2015;386(10002): 1447–1456.

23. Petty T. The Rising Epidemic of COPD in Women: Why women are more susceptible; how treatment should differ. *Women's Health in Primary Care*. 1999;2:12.

24. Sin DD, Cohen SB, Day A, Coxson H, Paré PD. Understanding the biological differences in susceptibility to chronic obstructive pulmonary disease between men and women. *Proc Am Thorac Soc*. 2007;4(8): 671–674.

25. Chapman KR. Chronic obstructive pulmonary disease: are women more susceptible than men? *Clin Chest Med*. 2004;25(2):331–341.

26. Bui DS, Burgess JA, Lowe AJ, et al. Childhood lung function predicts adult COPD and asthma-COPD overlap syndrome (ACOS). *Am J Respir Crit Care Med*. 2017;196(1):39–46.

27. Lange P, Celli B, Agusti A, et al. Lung-function trajectories leading to chronic obstructive pulmonary disease. *N Engl J Med*. 2015;373(2): 111–122.

28. Berry CE, Billheimer D, Jenkins IC, et al. A distinct low lung function trajectory from childhood to the fourth decade of life. *Am J Respir Crit Care Med*. 2016;194(5):607–612.

29. Postma DS, Weiss ST, van den Berge M, Kerstjens HA, Koppelman GH. Revisiting the Dutch hypothesis. *J Allergy Clin Immunol*. 2015;136(3): 521–529.

30. Svanes C, Sunyer J, Plana E, et al. Early life origins of chronic obstructive pulmonary disease. *Thorax*. 2010;65(1):14–20.

31. Global Initiative for Asthma. *Asthma COPD and Asthma – COPD Overlap Syndrome (ACOS)*. 2014. Available from: http://goldcopd.org/asthma-copd-asthma-copd-overlap-syndrome/. Accessed October 31, 2017.

32. Cui L, Gallagher LG, Ray RM, et al. Unexpected excessive chronic obstructive pulmonary disease mortality among female silk textile workers in Shanghai, China. *Occup Environ Med*. 2011;68(12): 883–887.

33. Dite GS, MacInnis RJ, Bickerstaffe A, et al. Breast cancer risk prediction using clinical models and 77 independent risk-associated SNPs for women aged under 50 years: Australian breast cancer family registry. *Cancer Epidemiol Biomarkers Prev*. 2016;25(2):359–365.

34. Matheson MC, D Olhaberriague AL, Burgess JA, et al. Preterm birth and low birth weight continue to increase the risk of asthma from age 7 to 43. *J Asthma*. 2017;54(6):616–623.