Prognostic models for outcome prediction in patients with chronic obstructive pulmonary disease: systematic review and critical appraisal

Vanesa Bellou,[1,2] Lazaros Belbasis,[1] Athanasios K Konstantinidis,[2] Ioanna Tzoulaki,[1,3,4] Evangelos Evangelou[1,3]

[1]Department of Hygiene and Epidemiology, University of Ioannina Medical School, Ioannina, Greece

[2]Department of Respiratory Medicine, University Hospital of Ioannina, University of Ioannina Medical School, Ioannina, Greece

[3]Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK

[4]MRC-PHE Center for Environment, School of Public Health, Imperial College London, London, UK

Correspondence to:
E Evangelou
vangelis@uoi.gr
(or @eevangelou on Twitter; ORCID 0000-0002-5488-2999)
Additional material is published online only. To view please visit the journal online.

ABSTRACT

OBJECTIVE
To map and assess prognostic models for outcome prediction in patients with chronic obstructive pulmonary disease (COPD).

DESIGN
Systematic review.

DATA SOURCES
PubMed until November 2018 and hand searched references from eligible articles.

ELIGIBILITY CRITERIA FOR STUDY SELECTION
Studies developing, validating, or updating a prediction model in COPD patients and focusing on any potential clinical outcome.

RESULTS
The systematic search yielded 228 eligible articles, describing the development of 408 prognostic models, the external validation of 38 models, and the validation of 20 prognostic models derived for diseases other than COPD. The 408 prognostic models were developed in three clinical settings: outpatients (n=239; 59%), patients admitted to hospital (n=155; 38%), and patients attending the emergency department (n=14; 3%). Among the 408 prognostic models, the most prevalent endpoints were mortality (n=209; 51%), risk for acute exacerbation of COPD (n=42; 10%), and risk for readmission after the index hospital admission (n=36; 9%). Overall, the most commonly used predictors were age (n=166; 41%), forced expiratory volume in one second (n=85; 21%), sex (n=74; 18%), body mass index (n=66; 16%), and smoking (n=65; 16%). Of the 408 prognostic models, 100 (25%) were internally validated and 91 (23%)

examined the calibration of the developed model. For 286 (70%) models a model presentation was not available, and only 56 (14%) models were presented through the full equation. Model discrimination using the C statistic was available for 311 (76%) models. 38 models were externally validated, but in only 12 of these was the validation performed by a fully independent team. Only seven prognostic models with an overall low risk of bias according to PROBAST were identified. These models were ADO, B-AE-D, B-AE-D-C, extended ADO, updated ADO, updated BODE, and a model developed by Bertens et al. A meta-analysis of C statistics was performed for 12 prognostic models, and the summary estimates ranged from 0.611 to 0.769.

CONCLUSIONS
This study constitutes a detailed mapping and assessment of the prognostic models for outcome prediction in COPD patients. The findings indicate several methodological pitfalls in their development and a low rate of external validation. Future research should focus on the improvement of existing models through update and external validation, as well as the assessment of the safety, clinical effectiveness, and cost effectiveness of the application of these prognostic models in clinical practice through impact studies.

SYSTEMATIC REVIEW REGISTRATION
PROSPERO CRD42017069247

WHAT IS ALREADY KNOWN ON THIS TOPIC

Historically, spirometry and age have been identified as the most important prognostic indicators in chronic obstructive pulmonary disease (COPD)

Global Initiative for Chronic Obstructive Lung Disease guidelines recommended use of multivariable prediction models to assess prognosis, instead of single predictors such as spirometry or history of exacerbations

No systematic overview has been published to summarise and critically appraise all multivariable prognostic models for outcome prediction in COPD patients

WHAT THIS STUDY ADDS

More than 400 prognostic models for outcome prediction in COPD patients exist, but only a minority have been externally validated and most were characterised by major drawbacks in the statistical analysis

Applying PROBAST showed that ADO, B-AE-D, B-AE-D-C, extended ADO, updated ADO, updated BODE, and a model developed by Bertens et al were derived in studies assessed as being at low risk of bias

Introduction

Chronic obstructive pulmonary disease (COPD) is a major public health problem. COPD accounts for at least 2.9 million deaths annually[1]; it is a leading cause of morbidity and mortality, and its prevalence is projected to increase over the coming years. Morbidity associated with the disease entails physician visits, emergency department visits, and hospital admissions,[2] all of which lead to a substantial economic burden. The greatest proportion of the costs is attributed to exacerbations of COPD.[2]

COPD is a fairly heterogeneous disease, and stratifying cases according to prognosis would raise the possibility of a precision medicine approach. For many years now, forced expiratory volume in one second (FEV_1) and age have been considered to be the most important prognostic indicators in COPD.[3] More recently, a wide variety of individual clinical factors have been also linked to prognosis of COPD.

Prognostic models, in general, have two distinct uses: they classify patients in groups with different prognosis

and estimate prognosis for individual patients. Although these are two different ways of looking at the same information, they differ fundamentally and the ultimate goal is to guide therapeutic and further diagnostic choices.[4] Use of a composite index to assess prognosis in COPD patients may provide a more comprehensive method of evaluation, incorporating a cluster of systemic manifestations of the disease.[5] Furthermore, in patients with COPD, multivariable prognostic models for various clinical outcomes could be used in clinical practice to assist decision making about hospital admission or admission to intensive care units and treatment strategy.[6]

Many prognostic models, combining multiple predictors for COPD related outcomes, have been developed. Global Initiative for Chronic Obstructive Lung Disease (GOLD) guidelines recommend the use of multivariable prediction models to assess the prognostic profile and facilitate follow-up of patients, instead of single predictors such as spirometry or history of exacerbations alone. Also, in the latest GOLD statement, the BODE index is proposed as a tool to determine who needs referral for consideration for lung transplantation.[7]

In this study, we aimed to systematically summarise the reported multivariable prognostic models developed for predicting subsequent outcomes in patients diagnosed as having COPD, to map their characteristics, and to examine whether they have undergone external validation. We used the Prediction model Risk Of Bias ASsessment Tool (PROBAST) to apply risk of bias assessment of the methodological features of the available studies developing or validating prognostic models. For prognostic models with multiple validation studies, we did a meta-analysis for performance and calibration of the models to obtain more accurate estimates.

Methods
We designed this systematic review according to the Checklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies (CHARMS) and the recent guidance by Debray et al.[8 9] A protocol for this study was published on PROSPERO (registration number CRD42017069247).

Literature search
We systematically searched PubMed from inception to 11 November 2018 to capture all studies developing and/or validating a prognostic model for clinical outcomes in COPD patients. On the basis of previous research,[10 11] we created the following search algorithm: (predict* OR progn* OR "risk prediction" OR "risk score" OR "risk calculation" OR "risk assessment" OR "c statistic" OR discrimination OR calibration OR AUC OR "area under the curve" OR "area under the receiver operator characteristic curve") AND ("chronic obstructive pulmonary disease" OR emphysema OR "chronic bronchitis" OR COPD). Two researchers (VB, LB) did the literature search independently, and discrepancies were resolved by a third researcher

(IT). We further hand searched the references of each eligible article for potential additional eligible studies.

Eligibility criteria
We included all studies that reported the development or validation of at least one multivariable model for predicting the risk for any clinical outcome in COPD patients. Table 1 shows a detailed description of the PICOTS for this review.[8 9] To consider a study as eligible, we followed the definition of prognostic model studies as proposed by the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statement.[12] Accordingly, it should specifically report the development, the update, or the external validation of a prognostic model used for making individualised predictions in COPD patients, either in its objectives or its conclusions. A study was also eligible if the development or update of a prognostic model could be deduced by the available information through the full text (for example, model presentation, measures of predictive performance for a multivariable model). Eligible outcomes included any possible clinical endpoint of COPD patients, such as mortality, exacerbations, and hospital admissions.

The eligible studies could report the development of multivariable models, the external validation of an existing model, and/or the update of an existing model. Updating of models may range from simple adjustment of the baseline risk/hazard or additional adjustment of predictors' weights by using the same or different adjustment factors to re-estimate predictors' weights to adding new predictors or removing existing predictors from the original model.[13] External validation studies aim to assess the predictive performance of an existing model in an independent population.[13] We included external validation studies that explicitly estimated and presented a measure of the model's performance. We also considered studies validating prediction models originally developed for other diseases in a COPD population. Also, eligible articles should report original research, study humans, and be written in English.

We excluded studies developing or validating diagnostic models to detect or exclude presence of COPD in patients with suspected COPD, studies examining only independent prognostic factors, methodological studies, and COPD case finding or screening studies. We also excluded studies that developed search algorithms to identify existing cases of COPD on the basis of administrative data. Given that prognostic models estimate a probability of a certain outcome for an individual patient over a specified time horizon, we excluded cross sectional studies because in this study design predictors and the outcome are measured concurrently. However, cohort studies that did an external validation of a model derived from a cross sectional study were deemed eligible.

Data extraction
To facilitate the data extraction process, three researchers (VB, LB, IT) constructed a standardised

Table 1 | Key items for framing aim, search strategy, and study inclusion and exclusion criteria for systematic review, following PICOTS guidance[8 9]

Item	Definition
Population	Patients diagnosed as having COPD
Intervention	Any prognostic model to predict any possible clinical outcome in COPD patients, to distinguish COPD patients with poor prognosis (ie, who will develop any unfavourable outcome), or to aid decision making in acute care and treatment planning in long term
Comparator	Not applicable
Outcomes	Any clinical outcome reported by prognostic models
Timing	Predictors measured at any timepoint in clinical course of COPD and preceding outcome; outcome measured in short term or long term without applying any specific limitation in prediction horizon
Setting	Patients visiting ambulatory healthcare facilities, patients admitted to hospital, or patients visiting emergency department

COPD=chronic obstructive pulmonary disease.

form by following recommendations in the CHARMS checklist.[8] Two researchers (VB, LB) independently extracted data. From all eligible articles, we extracted information on first author, year and journal of publication, and model name. From articles describing model development, we extracted the following information: study design, study population, geographical location, predicted outcome, definition of outcome, prediction horizon, definition of COPD, modelling method, method of internal validation, number of participants and number of events, number and type of predictors in final model, model presentation, and measures of predictive performance (discrimination, calibration, classification, overall performance). Potential measures of discrimination were C statistic and D statistic; potential measures of classification were sensitivity, specificity, positive and negative predictive value, and predictive accuracy; potential measures for overall performance were R^2 and Brier score; and potential measures for assessment of calibration were calibration plot, calibration-in-the-large, calibration slope, Hosmer-Lemeshow test, Harrell's E statistic, and calibration test.[14 15] Harrell's E statistic is defined as the absolute difference between smoothed observed outcomes and predicted probabilities.[14] Furthermore, we evaluated whether the authors reported only the apparent performance of a prognostic model or examined overfitting by using internal validation. Additionally, we examined whether a shrinkage of regression coefficients towards zero was performed in eligible studies and which method was used. We considered that the authors adjusted for optimism sufficiently if they re-evaluated the performance of a model in internal validation and performed shrinkage of model coefficients as well. We extracted information on whether the authors did decision curve analysis and net benefit analysis to evaluate the clinical usefulness of a model.[15 16] Moreover, for each eligible study, we examined whether the authors reported the presence of missing data on examined outcomes and/or variables included in prediction models; if so, we recorded how missing data were treated. We also extracted information on how continuous variables were handled and whether non-linear trends for continuous predictors were assessed by applying polynomials, fractional polynomials, or cubic splines. If the handling of continuous predictors

was not described explicitly, we scrutinised the full text and the tables of the respective papers to derive this information from the reported effect sizes. If this process was inconclusive, we described the handling of continuous predictors as unclear.

In articles examining the performance of the same prediction model on various outcomes or multiple timepoints, we retained the prediction model referring to the outcome or timepoint mentioned as the primary analysis of the study. In cases in which a primary timepoint was not specified, we considered the prediction with the longest horizon as the primary analysis of the study, because longer follow-up would lead to a larger number of events. Whenever a study described a model's performance both in an overall sample and in specific subgroups of the population, we extracted the analysis on the total population.

From articles describing external validation of models, we extracted study population, geographical location, number of participants and events, the model's performance, and calibration. If an article described multiple models, we extracted data separately for each model. For each model externally validated in multiple articles, we included in our analysis only external validation studies with non-overlapping populations. Furthermore, we examined whether the research team performing the external validation was independent of the research team developing the prediction model.

Risk of bias assessment
We appraised the presence of bias in the studies developing or externally validating prognostic models by using PROBAST, which is a risk of bias assessment tool designed for systematic reviews of diagnostic or prognostic prediction models.[17 18] It contains a multitude of questions in four different domains: participants, predictors, outcome, and statistical analysis. Questions are answered with yes, probably yes, probably no, no, and no information, depending on the characteristics of the study. If a domain contains at least one question signalled as no or probably no, it is considered to be at high risk. To be considered at low risk, a domain should contain all questions answered with yes or probably yes. Overall risk of bias is graded as low risk when all domains are considered low risk, and overall risk of bias is considered high risk when at least one of the domains is considered high risk. Two

researchers (VB, LB) independently assessed risk of bias.

PROBAST describes the assessment of both development studies and external validation studies. Often, articles describe the development of multiple prognostic models using different populations or different statistical approaches. Hence, differences in the risk of bias assessment is expected among different prognostic models developed in the same article. For this reason, we chose to report the risk of bias assessment per developed prognostic model and not per article. Furthermore, articles may describe the external validation of multiple prognostic models in the same population or in multiple different populations. For this reason, we refer to external validation efforts and we report the risk of bias assessment per external validation effort.

Statistical analysis

We calculated and reported descriptive statistics to summarise the characteristics of the models. We calculated the median and interquartile range for continuous variables and the respective percentages for binary variables.

For the prediction models that were examined in more than two independent datasets (excluding the model development dataset), we did a random effects meta-analysis to calculate a summary estimate for models' performance and calibration. We also considered for the meta-analysis those prediction models that were internally validated through bootstrapping or cross validation and were externally validated in only two independent datasets. We followed a recently published framework for the meta-analysis of prediction models.[9 19] If a measure of uncertainty (standard error or 95% confidence interval) was not available for mean C statistic, we used a formula to approximate the standard error of mean C statistic based on number of events and number of participants.[9 19 20] We quantified between study heterogeneity by using the I^2 and τ^2 statistics.[21] We used R version 3.5.2 for the statistical analysis. For the meta-analysis of prediction models, we used the R package "metamisc."[19]

Patient and public involvement

No patients or participants were involved in setting the research question or the outcome measures, nor were they involved in developing plans for design or implementation of the study. No patients were asked to advise on interpretation or writing up of results. There are no plans to disseminate the results of the research to study participants or the relevant patient community.

Results

Of the 17 538 screened papers, 228 papers were eligible (fig 1). These articles described the development of 408 prognostic models in COPD patients, the external validation of 38 prognostic models, and the application of 20 prognostic models originally developed for health outcomes other than prognosis of COPD patients. One of the eligible papers was identified through the hand search of references from eligible articles.[22] The prognostic models were mainly developed in the US (n=91; 22%), Spain (n=57; 14%), and the UK (n=34; 8%), whereas 80 (20%) models were developed in multicentre studies from multiple countries. For the derivation cohorts, the median sample size was 409 (interquartile range 163-1033) and the median number of events was 63 (36-188). For the internal validation cohorts, the median sample size was 831 (225-4192) and the median number of events was 77 (40-370).

The eligible prognostic models were developed in a variety of clinical settings; 239 (59%) models were developed in an outpatient setting, and 155 (38%) models were developed on a sample of patients admitted to hospital; 14 (3%) prognostic models were developed for COPD patients attending the emergency department. The developed models focused on a wide range of clinical outcomes. The most commonly used endpoints were mortality (n=209; 51%), exacerbation (n=42; 10%), and readmission after an index hospital admission (n=36; 9%). Supplementary table A shows a summary of the predicted outcomes per clinical setting. Twenty four prognostic models focused on a composite outcome. The most commonly used predictors were age (n=166; 41%), FEV_1 (n=85; 21%), sex (n=74; 18%), body mass index (n=66; 16%), smoking (n=65; 16%), previous exacerbations (n=53; 13%), previous hospital admissions (n=50; 12%), BODE index (n=43; 11%), modified Medical Research Council (mMRC) dyspnoea scale (n=42; 10%), and Charlson comorbidity index (n=35; 9%). Supplementary table B shows the top predictors in the 408 prognostic models for COPD patients stratified by clinical setting. Figure 2 shows the predictors that were used in at least 20 models, and figure 3 shows the 10 most common predictors stratified by clinical setting. Below, we describe the methodological and clinical characteristics for a total of 408 prognostic models, based on clinical setting.

Prognostic models for outpatients

Most of the prognostic models (n=239; 59%) were developed on a sample of COPD patients examined in an outpatient facility (supplementary table C). For the derivation cohort, the median sample size was 431 (244-1000) and the median number of events was 63 (33-155). For the internal validation cohort, the median sample size was 249 (204-3468) and the median number of events was 150 (64-1642). The most common clinical endpoints examined by these models were mortality (n=124; 52%), exacerbation (n=40; 17%), spirometric indices (n=25; 10%), hospital admission (n=16; 7%), treatment failure during an acute exacerbation (n=8; 3%), and composite outcome (n=9; 4%). The most commonly used predictors in these models were age (n=105; 44%), FEV_1 (n=69; 29%), smoking (n=54; 23%), body mass index (n=51; 21%), sex (n=43; 18%), previous exacerbations (n=43;
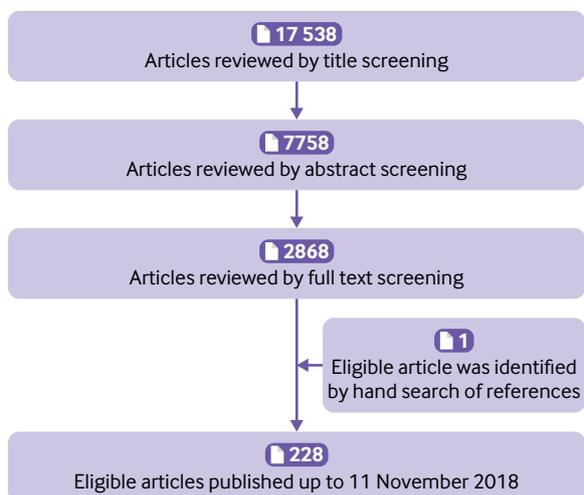
Fig 1 | Flowchart of literature search for prognostic models in patients with chronic obstructive pulmonary disease

18%), BODE index (n=43; 18%), previous hospital admissions (n=28; 12%), and diabetes mellitus (n=24; 10%).

A C statistic was reported for most (n=198; 83%) of these models, and the remaining 41 (17%) did not have a discrimination metric reported. For 172 prognostic models, only the apparent performance was reported in the development study. One prognostic model had temporal validation, and the remaining models had cross validation (n=28; 12%), bootstrapping (n=24; 10%), random split (n=12; 5%), or a combination of methods (n=2). Most (n=193; 81%) prognostic models were not calibrated; calibration was assessed for 46 prognostic models, and the most frequent method used was the Hosmer-Lemeshow test (n=35; 15%). Various modelling methods were applied, of which the most frequent were Cox regression (n=90; 38%), logistic regression (n=79; 33%), negative binomial

regression (n=21; 9%), and linear regression (n=16; 7%). For 12 prognostic models, shrinkage of regression coefficients was done to reduce overfitting. Application of a uniform shrinkage factor to all the regression coefficients was used for nine models, application of a penalised maximum likelihood method to estimate the regression coefficients was described for one prognostic model, and lasso regression was applied in two prognostic models to perform shrinkage for selection of predictors. For 17 prognostic models, a non-linear association between continuous predictors and predicted outcome was examined using the following methods: polynomials (n=7), restricted cubic splines (n=6), fractional polynomials (n=2), and Box-Tidwell transformation (n=2). A considerable number (n=178; 75%) of models did not have any type of model presentation, and only 24 (10%) reported the full regression equation. The most common type of presentation was sum score (n=30; 13%). Only one study performed decision analysis.[23] In this study, net benefit and decision curves are available for the updated ADO index. Net benefit is a category of decision analysis, comparing benefits and harms directly after transforming them on the same scale. Table 2 gives a detailed description of all the methodological characteristics.

Prognostic models for patients admitted to hospital
One hundred and fifty five models were developed in patients admitted to medical wards, intensive care units, or rehabilitation centres (supplementary table D). The median sample size of the derivation cohort was 303 (102-920), and the median number of events was 67 (37-311). The median sample size of the internal validation cohort was 4131 (731-4840), and the median number of events was 333 (35-370). The most prevalent outcomes assessed were mortality (n=78; 50%), readmission after an index admission (n=36; 23%), failure of non-invasive ventilation (n=14; 9%),
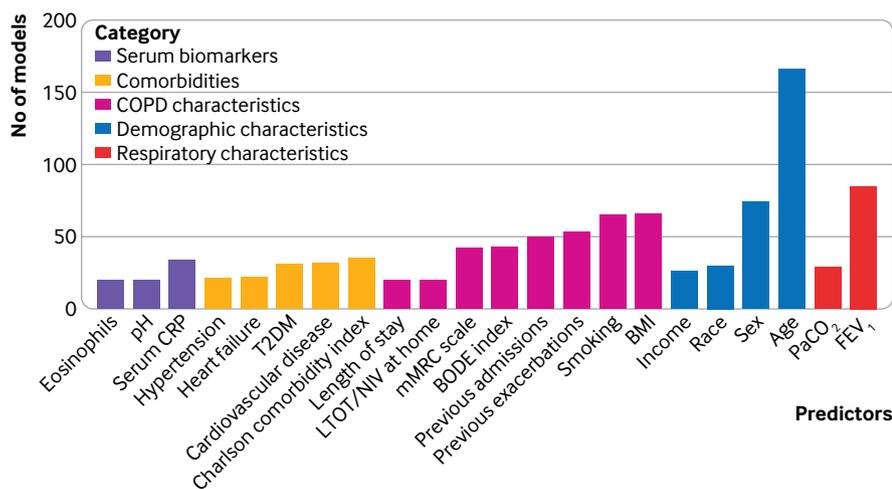

Fig 2 | Predictors included in at least 20 of 408 prognostic models for chronic obstructive pulmonary disease (COPD) patients by category of predictor. BMI=body mass index; CRP=C reactive protein; FEV$_1$=forced expiratory volume in one second; LTOT=long term oxygen therapy; mMRC scale=modified Medical Research Council dyspnoea scale; NIV=non-invasive ventilation; PaCO$_2$=partial pressure of carbon dioxide; T2DM=type 2 diabetes mellitus
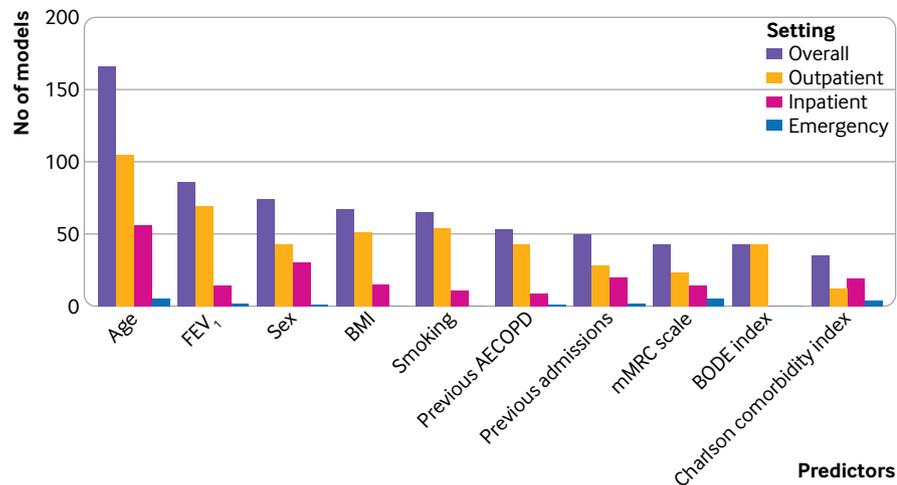
Fig 3 | 10 most frequently used predictors in 408 prognostic models for chronic obstructive pulmonary disease patients presented by clinical setting. AECOPD=acute exacerbation of chronic obstructive pulmonary disease; BMI=body mass index; FEV_1=forced expiratory volume in one second

and composite outcomes (n=13; 8%). The predictors encountered in most of the prognostic models were age (n=56; 36%), sex (n=30; 19%), partial pressure of carbon dioxide (n=24; 15%), previous hospital admissions (n=20; 13%), length of hospital stay (n=20; 13%), Charlson comorbidity index (n=19; 12%), pH (n=18; 12%), heart failure (n=16; 10%), body mass index (n=15; 10%), and serum albumin (n=15; 10%).

Of the 155 prognostic models, 31 (20%) were developed for patients admitted to intensive care units to predict mortality (n=22), weaning success (n=2), need for mechanical ventilation (n=6), and duration of mechanical ventilation (n=1). The most commonly used predictors were age (n=12), Glasgow or Japan Coma Scale (n=9), APACHE II (n=8), sex (n=6), pH (n=6), haemoglobin (n=6), serum albumin (n=6), heart failure (n=6), and hypertension (n=6).

A C statistic was reported for only 102 (66%) prognostic models; discrimination was not assessed for 53 (34%) models. One hundred and thirty one (85%) prognostic models did not have internal validation, and for the few models for which this was done, bootstrapping (n=9; 6%), random split (n=7; 5%), cross validation (n=3; 2%), or a combination of the aforementioned methods (n=2; 1%) was used. Three (2%) prognostic models had temporal validation. Calibration was not assessed for 116 (75%) prognostic models; the Hosmer-Lemeshow test (n=34; 22%) was the most frequently used method of calibration. Most of the prognostic models did not have a model presentation (n=104; 67%). A regression formula was available for 27 (17%) prognostic models. The most frequently used modelling methods were logistic regression (n=111; 72%) and Cox regression (n=21; 14%). For four prognostic models, shrinkage was applied to reduce overfitting. Application of a uniform shrinkage factor to all the regression coefficients was performed for two models, the penalised maximum likelihood approach was used in one model, and lasso shrinkage was applied for one model. For three

prognostic models, the non-linear association of predictors with the predicted outcome was considered using polynomials (n=1), fractional polynomials (n=1) and Box-Tidwell transformation (n=1). One study did a decision analysis after developing a prognostic model.[24]

Prognostic models for patients presenting to emergency department

Only 14 prognostic models were developed for patients who attend the emergency department (supplementary table E), with a median sample size of 1195 (871-1250) and a median number of events of 77 (40-137) in the derivation cohort. The median sample size of internal validation cohort was 1235 (266-1244), and the median number of events was 52 (29-66). The outcomes examined were mortality (n=7; 50%), change in physical activity (n=2), composite outcome (n=2), hospital admission (n=1), intensive care unit admission (n=1), and treatment failure after a visit to the emergency department for an acute exacerbation (n=1). Five of these models examined a long term prediction horizon (>1 month). The most prevalent variables included in these models were long term oxygen therapy or non-invasive ventilation at home (n=8; 57%), age (n=5; 36%), mMRC dyspnoea scale (n=5; 36%), Charlson comorbidity index (n=4; 29%), partial pressure of carbon dioxide (n=3; 21%), use of inspiratory accessory muscles and paradoxical breathing (n=3; 21%), and Glasgow or Japan Coma Scale (n=3; 21%).

An assessment of discrimination was not reported for three of these models, and a C statistic was reported for 11 models. Five models did not have any internal validation, and a random split of the dataset was used for eight models. Bootstrapping was used for internal validation of a single model. The most frequently used modelling method was logistic regression (n=10). A shrinkage procedure was not applied for any model.

Table 2 | Methodological characteristics of prognostic models developed for outcome prediction in patients with chronic obstructive pulmonary disease. Values are numbers (percentages*) unless stated otherwise

	Emergency department (14 models)	Inpatient setting (155 models)	Outpatient setting (239 models)	Overall (408 models)
Internal validation				
Non-random split	0	3 (2)	1 (<1)	4 (1)
Random split	8 (57)	7 (5)	12 (5)	27 (7)
Bootstrapping	1 (7)	9 (6)	24 (10)	34 (8)
Cross validation	0	3 (2)	28 (12)	31 (8)
Combination of methods	0	2 (1)	2 (1)	4 (1)
None	5 (36)	131 (85)	172 (72)	308 (75)
Modelling method				
Cox hazard model	0	21 (14)	90 (38)	111 (27)
Logistic regression model	10 (71)	111 (72)	79 (33)	200 (49)
Machine learning	2 (14)	3 (2)	2 (1)	7 (2)
Linear regression model	0	1 (1)	16 (7)	17 (4)
Generalised linear model	0	2 (1)	6 (3)	8 (2)
Negative binomial model	0	2 (1)	21 (9)	23 (6)
Weibull regression model	0	0	15 (6)	15 (4)
Other methods	2 (14)	4 (3)	5 (2)	11 (3)
More than one method	0	1 (1)	3 (1)	4 (1)
Not reported	0	10 (6)	2 (1)	12 (3)
Handling of missing data				
Imputation	4 (29)	13 (8)	18 (8)	35 (9)
No missing values	0	1 (1)	17 (7)	18 (4)
Exclusion of patients	7 (50)	28 (18)	47 (20)	82 (20)
Not reported	3 (21)	106 (68)	153 (64)	262 (64)
Inappropriate handling	0	7 (5)	4 (2)	11 (3)
Model discrimination				
C statistic	11 (79)	102 (66)	198 (83)	311 (76)
None	3 (21)	53 (34)	41 (17)	97 (24)
Model presentation				
Full equation	5 (36)	27 (17)	24 (10)	56 (14)
Sum score	3 (21)	16 (10)	30 (13)	49 (12)
Decision tree	2 (14)	1 (1)	3 (1)	6 (1)
Nomogram	0	1 (1)	3 (1)	4 (1)
Risk chart	0	3 (2)	0	3 (1)
More than one method	0	3 (2)	1 (<1)	4 (1)
None	4 (29)	104 (67)	178 (75)	286 (70)
Model calibration				
Hosmer-Lemeshow test	5 (36)	34 (22)	35 (15)	74 (18)
Calibration plot	0	1 (1)	4 (2)	5 (1)
More than one method	1 (7)	3 (2)	4 (2)	8 (2)
Other	0	1 (1)	3 (1)	4 (1)
None	8 (57)	116 (75)	193 (81)	317 (78)
Handling of continuous predictors				
Continuous	4 (29)	47 (30)	87 (36)	138 (34)
Categorical/dichotomous	10 (71)	87 (56)	64 (27)	161 (39)
Mixed handling	0	5 (3)	31 (13)	36 (9)
Not included	0	15 (10)	37 (15)	52 (13)
Unclear	0	1 (1)	20 (8)	21 (5)
Non-linearity				
Polynomials	0	1 (1)	7 (3)	8 (2)
Fractional polynomials	0	1 (1)	2 (1)	3 (1)
Restricted cubic splines	0	0	6 (3)	6 (1)
Box-Tidwell transformation	0	1 (1)	2 (1)	3 (1)
None	14 (100)	152 (98)	222 (93)	388 (95)

*Some percentages do not add up to 100% owing to rounding.

External validation studies

Of 408 prognostic models, 38 (9%) were externally validated at least once. However, only 12 (3%) models were externally validated by a fully independent research team. The prognostic models that were externally validated more than five times were ADO (17 cohorts), BODE (13 cohorts), BODEx (8 cohorts) and CODEX (7 cohorts).

Four prognostic models (DOSE index, SAFE index, mBODE% index, and COPD Severity Score) were developed in cross sectional studies, and these models were not described in the aforementioned sections. We retained only their external validation in cohort studies, of which there were 12 for DOSE index and one each for COPD Severity Score, SAFE index, and mBODE% index. Supplementary table F shows all the

external validation studies of the prognostic models for outcome prediction in COPD patients.[23 25-81]

Risk of bias assessment

We used PROBAST to assess the risk of bias of all studies developing or externally validating a prognostic model. In figure 4, we show a summary of the risk of bias assessment of developed models by domain. Seven prognostic models were assessed as being at low risk of bias, and all these models were developed for ambulatory COPD patients (ADO index, B-AE-D index, B-AE-D-C index, extended ADO index, updated BODE index, updated ADO index, and a model developed by Bertens et al). Table 3 shows the clinical setting, the predicted outcome and the time horizon, the events per variable number, the shrinkage method, and the optimism corrected C statistic for these seven prognostic models with low risk of bias. Table 4 shows the predictors included in these seven prognostic models. For one of these models (extended ADO index), a model presentation was not available. For the remaining six models, table 5 describes the model equations. Overall, 338 models were at low risk of bias for participants, 394 models were at low risk of bias for predictors, and 402 models were at low risk of bias for outcome, but only 10 models were at low risk of bias for statistical analysis.

We additionally assessed a total of 116 external validation efforts (fig 5). Of these efforts, only five were graded as being at low risk of bias according to PROBAST. These were one validation of the model developed by Bertens et al, one validation of DECAF, one validation of BAP-65, and two validations of PEARL. The remaining validation efforts were at high risk of bias.

Validation of prognostic models originally developed for other diseases

Twenty eight papers examined the predictive ability of 20 prediction models originally developed for diseases other than COPD (supplementary table G). Specifically,

these models are APACHE II and III, CHA_2DS_2-VASC, Charlson comorbidity index, CURB-65, CRB-65, CREWS, Elixhauser comorbidity index, Framingham risk score, GRACE, HOSPITAL, LACE, MDA, MODS, NEWS, NRS 2002, PSI, Salford-NEWS, SAPS, and SOFA. Overall, the predictive ability of these models was examined for mortality, exacerbation, hospital admission, failure of non-invasive ventilation, or identification of high cost patients.

Meta-analysis of prognostic models

Overall, we did 19 meta-analyses of C statistics for 12 prognostic models (ADO index, APACHE II, BOD index, BODE index, BODEx index, CODEX index, COTE index, CURB-65, DOSE index, LACE index, updated ADO index, and updated BODE index). For ADO index, APACHE II, BODE index, BODEx index, and CODEX index, we did two different meta-analyses for two distinct outcomes, whereas for DOSE index we did three different meta-analyses for three distinct outcomes. Eleven meta-analyses examined the risk of mortality, two meta-analyses examined the risk of acute exacerbation of COPD, five meta-analyses examined the risk of readmission or mortality, and one meta-analysis was focused on failure of non-invasive ventilation. I^2 estimates ranged from 0% to 96%, whereas τ^2 estimates ranged between 0 and 0.2605. In 12 meta-analyses of C statistics, we observed large between study heterogeneity ($I^2>50\%$). Summary C statistic estimates ranged from 0.611 for DOSE index in prediction of a composite outcome to 0.769 for APACHE II in prediction of mortality. Figure 6 shows a forest plot of all the meta-analyses, and table 6 shows the results of the meta-analyses of C statistics. We could not do meta-analysis of calibration measures, because they were not adequately reported in the external validation studies.

Discussion

Our systematic search yielded a detailed map of more than 400 prognostic models for the prediction of clinical outcomes in COPD patients. These models were developed in a wide range of clinical settings, including outpatient services, emergency departments, medical wards, intensive care units, and primary care structures. We identified seven prognostic models that were developed in studies at low risk of bias as assessed with PROBAST, and all these models were externally validated at least once. We complemented our systematic review and bias assessment with a meta-analysis of C statistics for 12 prognostic models.

Principal findings in context

Most of the prognostic models were developed in Western countries; more than half were developed in the US, Spain, and the UK. Although COPD is a quite prevalent chronic disease in low and middle income countries,[82] only a very small number of prognostic models were developed or validated in Asia, Africa, or South America. In the developing world, the main risk factors for COPD are history of tuberculosis and
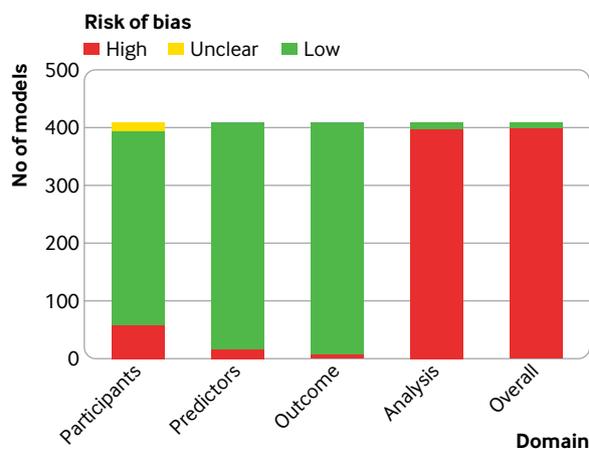

Fig 4 | Risk of bias assessment (using PROBAST) based on four domains across 408 prognostic models for outcome prediction in patients with chronic obstructive pulmonary disease

Table 3 | Characteristics of seven prognostic models for outcome prediction in chronic obstructive pulmonary disease patients that presented an overall low risk of bias

Reference	Model name	Clinical setting	Outcome	Predictors	Shrinkage methods	Handling of continuous predictors	EPV	Optimism corrected C statistic* (95% CI)
Puhan, 2009[49]	ADO	Outpatient	Mortality (3 years)	Age, FEV_1, mMRC	Uniform	Continuous	26.3	0.63†
Puhan, 2009[49]	Updated BODE	Outpatient	Mortality (3 years)	BMI, FEV_1, mMRC, 6MWD	Uniform	Continuous	19.8	0.61†
Puhan, 2012[23]	Updated ADO	Outpatient	Mortality (3 years)	Age, FEV_1, mMRC	Uniform	Continuous	311	0.73 (0.70 to 0.76)
Puhan, 2012[23]	Extended ADO	Outpatient	Mortality (3 years)	Age, FEV_1, mMRC, BMI, CVD, sex	Uniform	Continuous	155.5	0.74 (0.71 to 0.77)
Bertens, 2013[68]	NR	Outpatient	AECOPD (2 years)	FEV_1, previous exacerbations, smoking, vascular disease	Uniform	Continuous	17.5	0.66 (0.61 to 0.71)
Boeck, 2016[27]	B-AE-D	Outpatient	Mortality (2 years)	BMI, previous exacerbations, mMRC	Lasso	Continuous	18	0.63 (0.61 to 0.66)
Boeck, 2016[27]	B-AE-D-C	Outpatient	Mortality (2 years)	BMI, previous exacerbations, mMRC, serum copeptin	Lasso	Continuous	13.5	0.65 (0.57 to 0.72)

6MWD=6 minute walk distance test; AECOPD=acute exacerbation of chronic obstructive pulmonary disease; BMI=body mass index; CVD=cardiovascular disease; EPV=events per variable; FEV_1=forced expiratory volume in 1 second; mMRC=modified Medical Research Council dyspnoea scale; NR, not reported.
*Optimism corrected metric as reported in internal validation.
†Confidence intervals were not reported.

exposure to indoor air pollution.[83] Previous literature has shown a more favourable prognosis in COPD inflicted by biomass fuel than in smoking induced COPD.[84 85] We found only one paper reporting an external validation of BODEx index and COTE index in patients with COPD associated with biomass fuels[52]; however, this study was conducted in Spain. Our literature search indicates that currently developed prognostic models could not be generalised to developing countries, given that they have not been validated in these populations, except an external validation of BODE index in Brazilian population.[43]

Our systematic review showed several methodological pitfalls in the development of the models, which is also reflected in the risk of bias assessment. Only a quarter of the models were internally validated, and a tenth of the models were externally validated. The performance of a prognostic model is overestimated when simply determined in the sample of patients that was used to construct the model. Internal validation provides a more accurate estimate of model performance in new patients when it is properly performed—that is, using bootstrapping or cross validation techniques.[86] To ensure the generalisability of a prognostic model in populations with different characteristics, external validation studies are needed.[13] However, independent populations with large sample sizes of COPD patients and available COPD specific information (used as predictors in the prognostic models) can be hard to obtain to measure external validity. This necessitates the use of suitable internal validation techniques

to provide an optimism adjusted performance for the population in which the model was originally developed. Nevertheless, an evaluation of a model's performance in a different sample is not sufficient to overcome overfitting, and studies developing prognostic models should also apply shrinkage, which is a method to reduce overfitting by re-adjusting the regression coefficients.[87 88] Our systematic review showed that only a very small number of prognostic models performed shrinkage.

An important finding of our systematic review was that only a quarter of the models assessed calibration, which is the accuracy of absolute risk estimates—that is, it informs clinicians how similar the predicted absolute risk is to the true (observed) risk in groups of patients classified in different risk strata.[89] In addition, most of the models either did not report any method of handling missing data or performed a complete case analysis. Missing data often lead to biased estimates if not imputed, because they can distort the performance of a prediction model if the missingness of values is related to other known characteristics.[90] Additionally, in about half of the prognostic models, continuous predictors were dichotomised or categorised, and the non-linearity of continuous predictors was examined for only a small percentage of prognostic models. However, categorising continuous predictors into two or more categories has already been shown to lead to weaker prediction performance than analysing predictors on a continuous scale, owing to significant loss in information.[91] Additionally, non-linear associations can be efficiently modelled using restricted cubic splines or fractional polynomials.[92]

Another key factor is that discrimination and classification statistics that are usually reported in studies of prognostic models do not inform us about the clinical value of a model. Decision analysis is needed to evaluate whether the implementation of a prognostic model in clinical practice would be beneficial—that is, do more good than harm.[16] However, only two eligible studies did decision analysis.[23 24] Moreover, the applicability of a prediction model in clinical practice depends on the model presentation. In clinical practice, decision trees, sum scores, nomograms, and risk charts

Table 4 | Predictors included in prognostic models for outcome prediction in chronic obstructive pulmonary disease patients with low risk of bias

Model	mMRC	FEV_1	Age	BMI	Previous AECOPD	Additional predictors
Updated BODE	Yes	Yes	No	Yes	No	6MWD
ADO	Yes	Yes	Yes	No	No	-
Updated ADO	Yes	Yes	Yes	No	No	-
Extended ADO	Yes	Yes	Yes	Yes	No	Sex, CVD
Bertens, 2013	No	Yes	No	No	Yes	Smoking, vascular disease
B-AE-D	Yes	No	No	Yes	Yes	-
B-AE-D-C	Yes	No	No	Yes	Yes	Copeptin

6MWD=6 minute walk distance test; AECOPD=acute exacerbation of chronic obstructive pulmonary disease; BMI=body mass index; CVD=cardiovascular disease; FEV_1=forced expiratory volume in 1 second; mMRC=modified Medical Research Council dyspnoea scale.

Table 5 | Model equations of prognostic models for outcome prediction in chronic obstructive pulmonary disease patients with low risk of bias

Prediction model	Model equation
ADO	$y=-0.012\times FEV_1$ (% predicted)$+0.193\times mMRC+0.027\times age-3.436$
Updated ADO	$y=-0.288\times FEV_1$ (% predicted)$+0.2585\times mMRC+0.0703\times age-5.640$
Updated BODE	$y=-0.013\times BMI-0.005\times FEV_1$ (% predicted)$+0.146\times mMRC-0.005\times 6MWD+1.483$
B-AE-D*	$y=0.97\times(18.5\leq BMI<21)+1.45\times(BMI<18.5)+0.45\times$(previous severe exacerbations$=1$)$+1.22\times$(previous severe exacerbations≥2)$+0.97\times(mMRC=3)+1.67\times(mMRC=4)+constant$
B-AE-D-C*	$y=0.97\times(18.5\leq BMI<21)+1.45\times(BMI<18.5)+0.45\times$(previous severe exacerbations$=1$)$+1.22\times$(previous severe exacerbations≥2)$+0.97\times(mMRC=3)+1.67\times(mMRC=4)+0.50\times(20\leq copeptin<40)+1.58\times(copeptin\geq40)+constant$
Bertens et al.	$y=1.62\times$(presence of previous exacerbation)$-0.05\times FEV_1$ (% predicted, per 5% interval increase)$+0.12\times(2\times\log(pack\ years))+0.65\times$(presence of vascular disease)-1.33

6MWD=6 minute walk distance test; BMI=body mass index; FEV$_1$=forced expiratory volume in 1 second; mMRC=modified Medical Research Council dyspnoea scale.
*Constant of regression equation was not reported.

are commonly used in decision making. Sum scores and decision trees are more suitable for acute care settings, whereas risk charts and nomograms allow for a more detailed risk assessment and are more fitted for outpatient settings. However, more than two thirds of the developed models did not have any type of model presentation. Lack of presentation of a predictive tool does not allow its use in clinical practice. Additionally, lack of reporting of the regression formula in many of the prognostic models hinders future efforts for validation, update, and recalibration.[93]

The variables most commonly used in the development of prognostic models were age, FEV$_1$, sex, body mass index, smoking, previous exacerbations, previous hospital admissions, mMRC dyspnoea scale, BODE index, and Charlson comorbidity index. These variables are either anthropometric features, important factors in the natural progression of the disease, or markers of disease severity. They are easily measured, so they are available in settings where resources are limited (such as primary care) and in acute care facilities where prompt decisions need to be made (such as emergency departments). Another advantage of these predictors is their low risk for measurement bias, which leads to a smaller possibility of exposure misclassification. Finally, these variables have been identified as individual prognostic factors

in COPD.[3 94-98] However, we observed variability in the top predictors when the predictors were stratified by clinical setting. For example, in the prognostic models designed on the basis of COPD patients presenting at the emergency department, the most commonly used predictor was the use of long term oxygen therapy or non-invasive ventilation at home, which is uncommon in other settings. Also, smoking was a frequently used predictor only in models derived from outpatient settings, and it was only rarely used as a predictor in patients admitted to hospital. Furthermore, comorbid conditions, either in the form of multidimensional indices such as the Charlson comorbidity index, or as distinct conditions (for example, diabetes mellitus or cardiovascular disorders), were widely used and ranked among the most common predictors of clinical outcomes in all settings. Serum albumin and arterial blood gases were used almost exclusively as predictors in patients admitted to hospital and those visiting the emergency department.

The most extensively validated prognostic models were the BODE index and the ADO index.[49 99] The BODE index is the most established prognostic model in COPD and was developed to predict mortality.[99] In the GOLD statement, the BODE index is used in the prediction of mortality and in clinical decision making for lung transplantation and post-discharge follow-up of patients. The predictors included in the BODE index are body mass index, FEV$_1$, dyspnoea, and exercise capacity. Despite the lack of calibration in the original study of the BODE model, it has been validated and updated extensively in medical literature. The updated BODE index, a recalibration of the BODE index, is among the models with a low risk of bias.

The ADO index was based on the predictors used to develop the BODE index. It uses FEV$_1$, dyspnoea, and age.[49] The elimination of the six minute walking distance that was used in the BODE index was based on the rationale of developing a more easily applicable model, even by primary care physicians in settings with limited resources, rather than respiratory professionals alone. Despite the good predictive performance that the ADO index achieved in its development study, it showed poor calibration. This led to a recalibration of the ADO index in an independent population resulting in an updated ADO index, as well as an extended version of the recalibrated model with the addition of
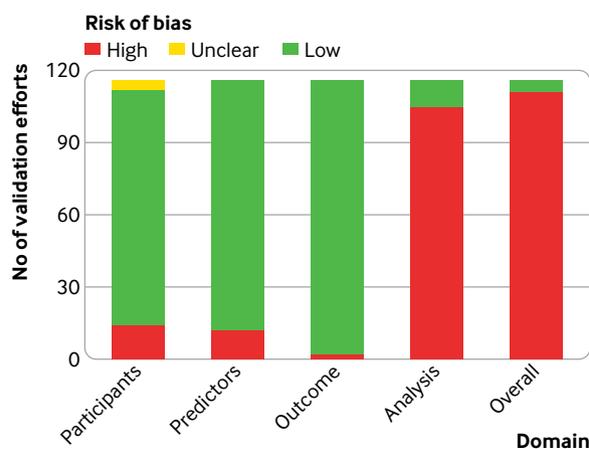

Fig 5 | Risk of bias assessment (using PROBAST) based on four domains across external validation studies of prognostic models for outcome prediction in patients with chronic obstructive pulmonary disease
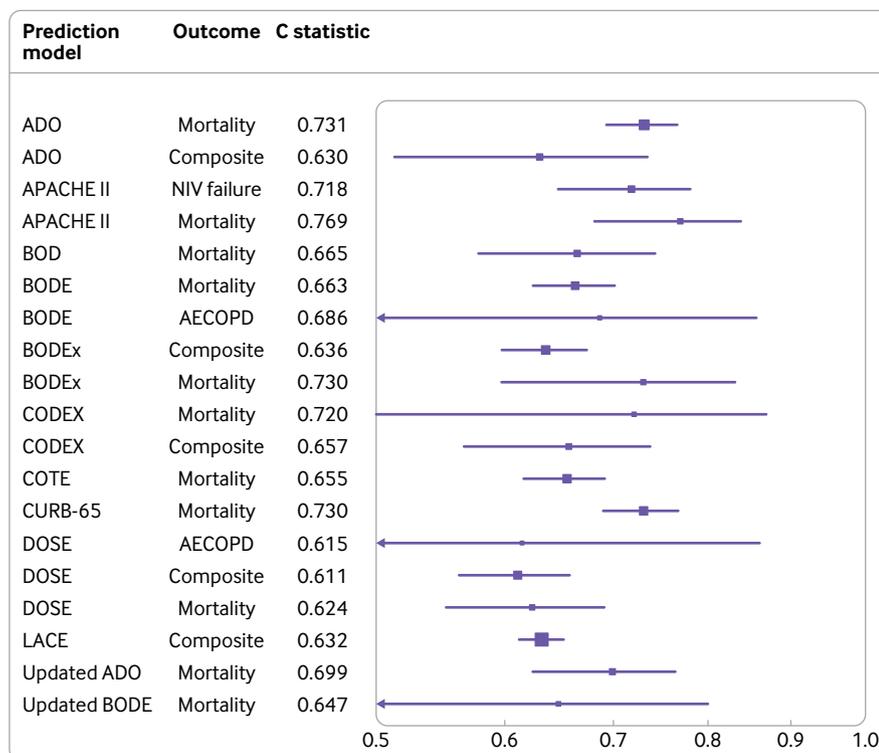
Prediction model	Outcome	C statistic
ADO	Mortality	0.731
ADO	Composite	0.630
APACHE II	NIV failure	0.718
APACHE II	Mortality	0.769
BOD	Mortality	0.665
BODE	Mortality	0.663
BODE	AECOPD	0.686
BODEx	Composite	0.636
BODEx	Mortality	0.730
CODEX	Mortality	0.720
CODEX	Composite	0.657
COTE	Mortality	0.655
CURB-65	Mortality	0.730
DOSE	AECOPD	0.615
DOSE	Composite	0.611
DOSE	Mortality	0.624
LACE	Composite	0.632
Updated ADO	Mortality	0.699
Updated BODE	Mortality	0.647

Fig 6 | Summary C statistic estimates for 19 meta-analyses of prognostic models for outcome prediction in patients with chronic obstructive pulmonary disease. AECOPD=acute exacerbation of chronic obstructive pulmonary disease

two variables. The ADO index, updated ADO index, and extension of updated ADO index had a low risk of bias and have been externally validated.

Three additional prognostic models presented low risk of bias and were developed for the outpatient setting. The B-AE-D index, and its update, the B-AE-D-C index, were developed for stable COPD patients at GOLD stage II to IV to predict the risk of two year all cause mortality.[27] The prognostic model developed by Bertens et al was the only prognostic model at low risk of bias that was developed to predict the risk of future exacerbations at two years in stable COPD patients.[68]

An essential step before the application of prediction models in clinical practice is their external validation in independent populations with different clinical characteristics and comparison of performance among different prediction models to identify the models with the best discrimination and calibration. A large scale effort to externally validate and compare multiple prognostic models for COPD patients was recently published.[100] The researchers used network meta-analysis to compare the performance of eight multivariable prognostic models and two different GOLD classifications in 24 cohort studies. In this analysis, the updated ADO index had the best ability to predict three year mortality in patients with COPD, followed by the updated BODE index and e-BODE index. However, the researchers pointed out that the approach of network meta-analysis has not yet integrated the synthesis of calibration measures.[100]

Recommendations and policy implications

On the basis of the aforementioned methodological pitfalls, the following recommendations could be stated to improve the research on prognostic models for prediction of outcome in COPD patients. Firstly, model development studies should adjust for overfitting by doing internal validation (mainly through non-random split or resampling techniques such as bootstrapping) and using shrinkage techniques and should provide an optimism adjusted performance. Secondly, model calibration should be examined. If a prognostic model has poor calibration, efforts should be made to improve its calibration by updating it either through recalibration or through addition of new variables. Thirdly, researchers should apply imputation techniques when data are missing, and they should report the full equation of the prognostic model to allow its external validation and update by independent research teams. Fourthly, continuous predictors should not be dichotomised, and potential non-linear association with the outcome should be examined using fractional polynomials or restricted cubic splines.[88]

The vast majority of prognostic models predicted the risk for mortality. Other clinically important outcomes, such as risk for exacerbation, a very common outcome in randomised clinical trials for COPD treatment, attracted much less attention. Also, the predictive ability of existing models focused on European and North American populations and could not be easily generalised. Thus, external validation studies of existing models in other populations are needed.

Table 6 | Results of meta-analyses of C statistics for prognostic models in patients with chronic obstructive pulmonary disease

Model	Outcome	No of datasets	No of events/participants	Summary C statistic (95% CI)	I² (%)	τ²
ADO	Mortality	11	11 258/72 850	0.731 (0.692 to 0.766)	95	0.0659
ADO	Readmission or mortality	3	936/2417	0.630 (0.513 to 0.734)	81	0.0303
APACHE II	NIV failure	3	121/550	0.718 (0.647 to 0.780)	0	0
APACHE II	Mortality	7	NA/NA*	0.769 (0.681 to 0.838)	84	0.1654
BOD	Mortality	4	NA/NA*	0.665 (0.578 to 0.742)	63	0.0375
BODE	Mortality	8	847/6124	0.663 (0.624 to 0.701)	36	0.0124
BODE	AECOPD	3	156/428	0.686 (0.442 to 0.857)	71	0.1141
BODEx	Readmission or mortality	3	936/2417	0.636 (0.598 to 0.674)	0	0
BODEx	Mortality	4	1505/4963	0.730 (0.597 to 0.831)	93	0.1221
CODEX	Mortality	3	8359/53 975	0.720 (0.500 to 0.869)	96	0.1367
CODEX	Readmission or mortality	3	936/2417	0.657 (0.566 to 0.737)	67	0.0154
COTE	Mortality	4	8303/53 737	0.655 (0.616 to 0.692)	57	0.0090
CURB-65	Mortality	6	451/4250	0.730 (0.690 to 0.767)	24	0.0092
DOSE	AECOPD	3	NA/NA*	0.615 (0.291 to 0.861)	93	0.2605
DOSE	Readmission or mortality	3	936/2417	0.611 (0.562 to 0.658)	0	0
DOSE	Mortality	3	9390/56 546	0.624 (0.552 to 0.691)	85	0.0095
LACE	Readmission or mortality	4	1601/5079	0.632 (0.612 to 0.652)	0	1.96×10⁻⁶
Updated ADO	Mortality	4	NA/NA*	0.699 (0.624 to 0.764)	91	0.0419
Updated BODE	Mortality	3	149/723	0.647 (0.456 to 0.800)	48	0.0379

AECOPD=acute exacerbation of chronic obstructive pulmonary disease; NA=not available.
*For at least one dataset, number of events and/or participants was not reported.

External validation studies are not sufficient to guarantee the clinical utility of a prediction model. To select a prediction model for implementation in clinical practice, impact studies are needed.[13] These are randomised clinical trials applying a prognostic model in a clinical setting and assessing its clinical utility for decision making. However, we found only one impact study in the literature.[101] This study concluded that the DECAF score, a prognostic model that was initially developed for patients admitted to hospital with an exacerbation to predict in-hospital mortality,[102] is safe, clinically effective, and cost effective in the selection of COPD patients with an exacerbation that could be treated at home.[101 103]

Comparison with other studies
A previously published systematic review identified 15 prognostic models (either original models or updates of existing models) for stable COPD patients that were published up to September 2010.[6] This systematic review mainly focused on the description of clinical characteristics of prognostic models—that is, population characteristics and predictors. In contrast, our systematic review included a much broader spectrum of COPD patients by additionally detecting prognostic models for COPD patients admitted to hospital and for those visiting the emergency department. As a consequence, we captured a total of 408 prognostic models from various clinical settings. Furthermore, we reported a detailed presentation of methodological characteristics in multivariable prognostic models for outcome prediction in COPD patients. We additionally did a meta-analysis for prognostic models with multiple external validation studies, and we assessed the risk of bias by using PROBAST.

Strengths and limitations of study
The major strength of our study is that it provides an overall mapping of the available research on prognostic models for outcome prediction in COPD patients. We collected all published prognostic models used to forecast any clinical outcome that may occur in the course of COPD. We presented a detailed description of the characteristics of the developed models, as well as updates and validation studies of existing models. Another important aspect of our paper is the critical appraisal of prognostic models in COPD by using the PROBAST tool. We also did a meta-analysis of C statistics for prognostic models that were externally validated in multiple independent populations.

A limitation of our study is the inability to do meta-analysis of calibration measures for prognostic models, owing to poor reporting of calibration in the validation studies. Also, we observed large between study heterogeneity in the meta-analyses of C statistics. Potential sources of heterogeneity could be the differences in clinical setting, patients' characteristics, and time horizons across the validation studies, but we could not do meta-regression analyses or sensitivity analyses owing to the small number of external validation studies per prognostic model.[92]

Conclusions
Our paper constitutes a map of the research on multivariable prognostic models for outcome prediction in COPD patients, aiming to summarise their methodological characteristics, their calibration, and their performance. An abundance of prognostic models is available for patients with COPD, so deciding on which one to use in a specific setting or population can be challenging for healthcare professionals. Future prognostic research should steer towards recalibration or update of existing prognostic models with the

addition of new predictors to enhance their prognostic performance. Studies updating existing models should sufficiently estimate optimism adjusted performance and calibration measures by applying appropriate internal validation and should adjust for overfitting by applying shrinkage techniques. Future studies should also use multiple imputation to handle missing data as well as examine non-linearity of continuous predictors.

Moreover, to ensure the generalisability of prognostic models, validation studies in populations with different characteristics, with regards to setting and inclusion criteria, are needed. Prognostic tools with good calibration and external validity should inform clinical practice as well as be recommended by guidelines after they have undergone impact studies to examine the effect of using the model for a specific outcome in clinical practice.

1 Soriano JB, Rodríguez-Roisin R. Chronic obstructive pulmonary disease overview: epidemiology, risk factors, and clinical presentation. *Proc Am Thorac Soc* 2011;8:363-7. doi:10.1513/pats.201102-017RM

2 Global Initiative for Chronic Obstructive Lung Disease (GOLD). 2019 Global Strategy for the Diagnosis, Management and Prevention of COPD. 2019. https://goldcopd.org/gold-reports/.

3 Celli BR, Casanova Macario C. The use of multidimensional indices. In: Anzueto A, Heijdra Y, Hurst JR, eds. *Controversies in COPD*. European Respiratory Society, 2015: 143-60. doi:10.1183/2312508X.10019714

4 Altman DG, Royston P. What do we mean by validating a prognostic model?*Stat Med* 2000;19:453-73. doi:10.1002/(SICI)1097-0258(20000229)19:4<453::AID-SIM350>3.0.CO;2-5

5 Celli BR. Predictors of mortality in COPD. *Respir Med* 2010;104:773-9. doi:10.1016/j.rmed.2009.12.017

6 Dijk WD, Bemt Lv, Haak-Rongen Sv, et al. Multidimensional prognostic indices for use in COPD patient care. A systematic review. *Respir Res* 2011;12:151. doi:10.1186/1465-9921-12-151

7 Vogelmeier CF, Criner GJ, Martinez FJ, et al. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease 2017 Report: GOLD Executive Summary. *Eur Respir J* 2017;49:1700214. doi:10.1183/13993003.00214-2017

8 Moons KGM, de Groot JAH, Bouwmeester W, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Med* 2014;11:e1001744. doi:10.1371/journal.pmed.1001744

9 Debray TPA, Damen JAAG, Snell KIE, et al. A guide to systematic review and meta-analysis of prediction model performance. *BMJ* 2017;356:i6460. doi:10.1136/bmj.i6460

10 Geersing G-J, Bouwmeester W, Zuithoff P, Spijker R, Leeflang M, Moons KG. Search filters for finding prognostic and diagnostic prediction studies in Medline to enhance systematic reviews. *PLoS One* 2012;7:e32844. doi:10.1371/journal.pone.0032844

11 Damen JAAG, Hooft L, Schuit E, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. *BMJ* 2016;353:i2416. doi:10.1136/bmj.i2416

12 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. doi:10.1136/bmj.g7594

13 Moons KGM, Kengne AP, Grobbee DE, et al. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart* 2012;98:691-8. doi:10.1136/heartjnl-2011-301247

14 Steyerberg EW. *Clinical Prediction Models.* Springer New York, 2009. doi:10.1007/978-0-387-77244-8

15 Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010;21:128-38. doi:10.1097/EDE.0b013e3181c30fb2

16 Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *BMJ* 2016;352:i6. doi:10.1136/bmj.i6

17 Wolff RF, Moons KGM, Riley RD, et al, PROBAST Group†. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Ann Intern Med* 2019;170:51-8. doi:10.7326/M18-1376

18 Moons KGM, Wolff RF, Riley RD, et al. PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. *Ann Intern Med* 2019;170:W1-33. doi:10.7326/M18-1377

19 Debray TP, Damen JA, Riley RD, et al. A framework for meta-analysis of prediction model studies with binary and time-to-event outcomes. *Stat Methods Med Res* 2019;28:2768-86. doi:10.1177/0962280218785504

20 Snell KI, Ensor J, Debray TP, Moons KG, Riley RD. Meta-analysis of prediction model performance across multiple studies: Which scale helps ensure between-study normality for the C-statistic and calibration measures?*Stat Methods Med Res* 2018;27:3505-22. doi:10.1177/0962280217705678

21 Rücker G, Schwarzer G, Carpenter JR, Schumacher M. Undue reliance on I(2) in assessing heterogeneity may mislead. *BMC Med Res Methodol* 2008;8:79. doi:10.1186/1471-2288-8-79

22 Madkour AM, Adly NN. Predictors of in-hospital mortality and need for invasive mechanical ventilation in elderly COPD patients presenting with acute hypercapnic respiratory failure. *Egypt J Chest Dis Tuberc* 2013;62:393-400. doi:10.1016/j.ejcdt.2013.07.003

23 Puhan MA, Hansel NN, Sobradillo P, et al, International COPD Cohorts Collaboration Working Group. Large-scale international validation of the ADO index in subjects with COPD: an individual subject data analysis of 10 cohorts. *BMJ Open* 2012;2:e002152. doi:10.1136/bmjopen-2012-002152

24 Zhong X, Lee S, Zhao C, et al. Reducing COPD readmissions through predictive modeling and incentive-based interventions. *Health Care Manag Sci* 2019;22:121-39. doi:10.1007/s10729-017-9426-2

25 Abu Hussein N, Ter Riet G, Schoenenberger L, et al. The ADO index as a predictor of two-year mortality in general practice-based chronic obstructive pulmonary disease cohorts. *Respiration* 2014;88:208-14. doi:10.1159/000363770

26 Ansari K, Keaney N, Kay A, et al. Body mass index, airflow obstruction and dyspnea and body mass index, airflow obstruction, dyspnea scores, age and pack years-predictive properties of new multidimensional prognostic indices of chronic obstructive pulmonary disease in primary care. *Ann Thorac Med* 2016;11:261-8. doi:10.4103/1817-1737.191866

27 Boeck L, Soriano JB, Brusse-Keizer M, et al. Prognostic assessment in COPD without lung function: the B-AE-D indices. *Eur Respir J* 2016;47:1635-44. doi:10.1183/13993003.01485-2015

28 Echevarria C, Steer J, Heslop-Marshall K, et al. The PEARL score predicts 90-day readmission or death after hospitalisation for acute exacerbation of COPD. *Thorax* 2017;72:686-93. doi:10.1136/thoraxjnl-2016-209298

29 Esteban C, Arostegui I, Moraza J, et al. Development of a decision tree to assess the severity and prognosis of stable COPD. *Eur Respir J* 2011;38:1294-300. doi:10.1183/09031936.00189010

30 Jones RC, Price D, Chavannes NH, et al, UNLOCK Group of the IPCRG. Multi-component assessment of chronic obstructive pulmonary disease: an evaluation of the ADO and DOSE indices and the global obstructive lung disease categories in international primary care data sets. *NPJ Prim Care Respir Med* 2016;26:16010. doi:10.1038/npjpcrm.2016.10

31 Marin JM, Alfageme I, Almagro P, et al. Multicomponent indices to predict survival in COPD: the COCOMICS study. *Eur Respir J* 2013;42:323-32. doi:10.1183/09031936.00121012

32 Morales DR, Flynn R, Zhang J, Trucco E, Quint JK, Zutis K. External validation of ADO, DOSE, COTE and CODEX at predicting death in primary care patients with COPD using standard and machine learning approaches. *Respir Med* 2018;138:150-5. doi:10.1016/j.rmed.2018.04.003

33 Motegi T, Jones RC, Ishii T, et al. A comparison of three multidimensional indices of COPD severity as predictors of future exacerbations. *Int J Chron Obstruct Pulmon Dis* 2013;8:259-71. doi:10.2147/COPD.S42769

34 Ou C-Y, Chen C-Z, Yu C-H, Shiu CH, Hsiue TR. Discriminative and predictive properties of multidimensional prognostic indices of chronic obstructive pulmonary disease: a validation study in Taiwanese patients. *Respirology* 2014;19:694-9. doi:10.1111/resp.12313

35 Quintana JM, Esteban C, Unzurrunzaga A, et al, IRYSS-COPD group. Predictive score for mortality in patients with COPD exacerbations attending hospital emergency departments. *BMC Med* 2014;12:66. doi:10.1186/1741-7015-12-66

36 Waschki B, Kirsten A, Holz O, et al. Physical activity is the strongest predictor of all-cause mortality in patients with COPD: a prospective cohort study. *Chest* 2011;140:331-42. doi:10.1378/chest.10-2521

37 Zhang J, Rutten FH, Cramer MJ, Lammers JW, Zuithoff NP, Hoes AW. The importance of cardiovascular disease for mortality in patients with COPD: a prognostic cohort study. *Fam Pract* 2011;28:474-81. doi:10.1093/fampra/cmr024

38 Germini F, Veronese G, Marcucci M, et al, SIMEU Study Group. Validation of the BAP-65 score for prediction of in-hospital death or use of mechanical ventilation in patients presenting to the emergency department with an acute exacerbation of COPD: a retrospective multi-center study from the Italian Society of Emergency Medicine (SIMEU). *Eur J Intern Med* 2019;61:62-8. doi:10.1016/j.ejim.2018.10.018

39 Sangwan V, Chaudhry D, Malik R. Dyspnea, Eosinopenia, Consolidation, Acidemia and Atrial Fibrillation Score and BAP-65 Score, Tools for Prediction of Mortality in Acute Exacerbations of Chronic Obstructive Pulmonary Disease: A Comparative Pilot Study. *Indian J Crit Care Med* 2017;21:671-7. doi:10.4103/ijccm.IJCCM_148_17

40 Chan HP, Mukhopadhyay A, Chong PLP, et al. Prognostic utility of the 2011 GOLD classification and other multidimensional tools in Asian COPD patients: a prospective cohort study. *Int J Chron Obstruct Pulmon Dis* 2016;11:823-9. doi:10.2147/COPD.S96790

41 Crook S, Frei A, Ter Riet G, Puhan MA. Prediction of long-term clinical outcomes using simple functional exercise performance tests in patients with COPD: a 5-year prospective cohort study. *Respir Res* 2017;18:112. doi:10.1186/s12931-017-0598-6

42 Stolz D, Kostikas K, Blasi F, et al. Adrenomedullin refines mortality prediction by the BODE index in COPD: the "BODE-A" index. *Eur Respir J* 2014;43:397-408. doi:10.1183/09031936.00058713

43 Faganello MM, Tanni SE, Sanchez FF, Pelegrino NR, Lucheta PA, Godoy I. BODE index and GOLD staging as predictors of 1-year exacerbation risk in chronic obstructive pulmonary disease. *Am J Med Sci* 2010;339:10-4. doi:10.1097/MAJ.0b013e3181bb8111

44 Herer B, Chinet T. Acute exacerbation of COPD during pulmonary rehabilitation: outcomes and risk prediction. *Int J Chron Obstruct Pulmon Dis* 2018;13:1767-74. doi:10.2147/COPD.S163472

45 Horita N, Koblizek V, Plutinsky M, Novotna B, Hejduk K, Kaneko T. Chronic obstructive pulmonary disease prognostic score: A new index. *Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub* 2016;160:211-8. doi:10.5507/bp.2016.030

46 Moy ML, Teylan M, Danilack VA, Gagnon DR, Garshick E. An index of daily step count and systemic inflammation predicts clinical outcomes in chronic obstructive pulmonary disease. *Ann Am Thorac Soc* 2014;11:149-57. doi:10.1513/AnnalsATS.201307-243OC

47 Neo H-Y, Xu H-Y, Wu H-Y, Hum A. Prediction of Poor Short-Term Prognosis and Unmet Needs in Advanced Chronic Obstructive Pulmonary Disease: Use of the Two-Minute Walking Distance Extracted from a Six-Minute Walk Test. *J Palliat Med* 2017;20:821-8. doi:10.1089/jpm.2016.0449

48 Pedone C, Scarlata S, Forastiere F, Bellia V, Antonelli Incalzi R. BODE index or geriatric multidimensional assessment for the prediction of very-long-term mortality in elderly patients with chronic obstructive pulmonary disease? a prospective cohort study. *Age Ageing* 2014;43:553-8. doi:10.1093/ageing/aft197

49 Puhan MA, Garcia-Aymerich J, Frey M, et al. Expansion of the prognostic assessment of patients with chronic obstructive pulmonary disease: the updated BODE index and the ADO index. *Lancet* 2009;374:704-11. doi:10.1016/S0140-6736(09)61301-5

50 Strassmann A, Frei A, Haile SR, Ter Riet G, Puhan MA. Commonly Used Patient-Reported Outcomes Do Not Improve Prediction of COPD Exacerbations: A Multicenter 4½ Year Prospective Cohort Study. *Chest* 2017;152:1179-87. doi:10.1016/j.chest.2017.09.003

51 Golpe R, Suárez-Valor M, Veres-Racamonde A, et al. Octogenarian patients with chronic obstructive pulmonary disease: Characteristics and usefulness of prognostic indexes. *Med Clin (Barc)* 2018;151:53-8. doi:10.1016/j.medcli.2017.09.011

52 Golpe R, Mengual-Macenlle N, Sanjuán-López P, Cano-Jiménez E, Castro-Añón O, Pérez-de-Llano LA. Prognostic Indices and Mortality Prediction in COPD Caused by Biomass Smoke Exposure. *Lung* 2015;193:497-503. doi:10.1007/s00408-015-9731-9

53 Chan HP, Mukhopadhyay A, Chong PLP, et al. Role of BMI, airflow obstruction, St George's Respiratory Questionnaire and age index in prognostication of Asian COPD. *Respirology* 2017;22:114-9. doi:10.1111/resp.12877

54 Wildman MJ, Harrison DA, Welch CA, Sanderson C. A new measure of acute physiological derangement for patients with exacerbations of obstructive airways disease: the COPD and Asthma Physiology Score. *Respir Med* 2007;101:1994-2002. doi:10.1016/j.rmed.2007.04.002

55 Almagro P, Yun S, Sangil A, et al. Palliative care and prognosis in COPD: a systematic review with a validation cohort. *Int J Chron Obstruct Pulmon Dis* 2017;12:1721-9. doi:10.2147/COPD.S135657

56 Liu D, Peng S-H, Zhang J, Bai SH, Liu HX, Qu JM. Prediction of short term re-exacerbation in patients with acute exacerbation of chronic obstructive pulmonary disease. *Int J Chron Obstruct Pulmon Dis* 2015;10:1265-73.

57 Miravitlles M, Izquierdo I, Herrejón A, Torres JV, Baró E, Borja J, ESFERA investigators. COPD severity score as a predictor of failure in exacerbations of COPD. The ESFERA study. *Respir Med* 2011;105:740-7. doi:10.1016/j.rmed.2010.12.020

58 Villalobos N, Davidson R, Ghori UK, Abdou Y, Abukhalaf J, Guillamet RV. External Validation of the COmorbidity Test. *COPD* 2017;14:513-7. doi:10.1080/15412555.2017.1354981

59 Hoogendoorn M, Feenstra TL, Boland M, et al. Prediction models for exacerbations in different COPD patient populations: comparing results of five large data sources. *Int J Chron Obstruct Pulmon Dis* 2017;12:3183-94. doi:10.2147/COPD.S142378

60 Echevarria C, Steer J, Heslop-Marshall K, et al. Validation of the DECAF score to predict hospital mortality in acute exacerbations of COPD. *Thorax* 2016;71:133-40. doi:10.1136/thoraxjnl-2015-207775

61 Jones RC, Donaldson GC, Chavannes NH, et al. Derivation and validation of a composite index of severity in chronic obstructive pulmonary disease: the DOSE Index. *Am J Respir Crit Care Med* 2009;180:1189-95. doi:10.1164/rccm.200902-0271OC

62 Rolink M, van Dijk W, van den Haak-Rongen S, Pieters W, Schermer T, van den Bemt L. Using the DOSE index to predict changes in health status of patients with COPD: a prospective cohort study. *Prim Care Respir J* 2013;22:169-74. doi:10.4104/pcrj.2013.00033

63 Esteban C, Quintana JM, Moraza J, et al. BODE-Index vs HADO-score in chronic obstructive pulmonary disease: Which one to use in general practice? *BMC Med* 2010;8:28. doi:10.1186/1741-7015-8-28

64 Esteban C, Quintana JM, Aburto M, et al. The health, activity, dyspnea, obstruction, age, and hospitalization: prognostic score for stable COPD patients. *Respir Med* 2011;105:1662-70. doi:10.1016/j.rmed.2011.05.005

65 Cote CG, Pinto-Plata VM, Marin JM, Nekach H, Dordelly LJ, Celli BR. The modified BODE index: validation with mortality in COPD. *Eur Respir J* 2008;32:1269-74. doi:10.1183/09031936.00138507

66 Stolz D, Meyer A, Rakic J, Boeck L, Scherr A, Tamm M. Mortality risk prediction in COPD by a prognostic biomarker panel. *Eur Respir J* 2014;44:1557-70. doi:10.1183/09031936.00043814

67 Antón A, Güell R, Gómez J, et al. Predicting the result of noninvasive ventilation in severe acute exacerbations of patients with chronic airflow limitation. *Chest* 2000;117:828-33. doi:10.1378/chest.117.3.828

68 Bertens LCM, Reitsma JB, Moons KGM, et al. Development and validation of a model to predict the risk of exacerbations in chronic obstructive pulmonary disease. *Int J Chron Obstruct Pulmon Dis* 2013;8:493-9. doi:10.2147/COPD.S49609

69 Bloch KE, Weder W, Bachmann LM, Russi EW. Model-based versus clinical prediction of the spirometric response to lung volume reduction surgery. *Respiration* 2004;71:611-8. doi:10.1159/000081762

70 Confalonieri M, Garuti G, Cattaruzza MS, et al, Italian noninvasive positive pressure ventilation (NPPV) study group. A chart of failure risk for noninvasive ventilation in patients with COPD exacerbation. *Eur Respir J* 2005;25:348-55. doi:10.1183/09031936.05.00085304

71 Connors AFJr, Dawson NV, Thomas C, et al. Outcomes following acute exacerbation of severe chronic obstructive lung disease.

The SUPPORT investigators (Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments). *Am J Respir Crit Care Med* 1996;154:959-67. doi:10.1164/ajrccm.154.4.8887592

72 Esteban C, Castro-Acosta A, Alvarez-Martínez CJ, Capelastegui A, López-Campos JL, Pozo-Rodriguez F. Predictors of one-year mortality after hospitalization for an exacerbation of COPD. *BMC Pulm Med* 2018;18:18. doi:10.1186/s12890-018-0574-z

73 Kerkhof M, Freeman D, Jones R, Chisholm A, Price DB, Respiratory Effectiveness Group. Predicting frequent COPD exacerbations using primary care data. *Int J Chron Obstruct Pulmon Dis* 2015;10:2439-50.

74 Lindenauer PK, Grosso LM, Wang C, et al. Development, validation, and results of a risk-standardized measure of hospital 30-day mortality for patients with exacerbation of chronic obstructive pulmonary disease. *J Hosp Med* 2013;8:428-35. doi:10.1002/jhm.2066

75 Murata GH, Gorby MS, Kapsner CO, Chick TW, Halperin AK. A multivariate model for the prediction of relapse after outpatient treatment of decompensated chronic obstructive pulmonary disease. *Arch Intern Med* 1992;152:73-7. doi:10.1001/archinte.1992.00400130097011

76 Stanford RH, Nag A, Mapel DW, et al. Claims-based risk model for first severe COPD exacerbation. *Am J Manag Care* 2018;24:e45-53.

77 Tabak YP, Sun X, Johannes RS, Hyde L, Shorr AF, Lindenauer PK. Development and validation of a mortality risk-adjustment model for patients hospitalized for exacerbations of chronic obstructive pulmonary disease. *Med Care* 2013;51:597-605. doi:10.1097/MLR.0b013e3182901982

78 Zafari Z, Sin DD, Postma DS, et al. Individualized prediction of lung-function decline in chronic obstructive pulmonary disease. *CMAJ* 2016;188:1004-11. doi:10.1503/cmaj.151483

79 Lau CS, Siracuse BL, Chamberlain RS. Readmission After COPD Exacerbation Scale: determining 30-day readmission risk for COPD patients. *Int J Chron Obstruct Pulmon Dis* 2017;12:1891-902. doi:10.2147/COPD.S136768

80 Roche N, Chavaillon J-M, Maurer C, Zureik M, Piquet J. A clinical in-hospital prognostic score for acute exacerbations of COPD. *Respir Res* 2014;15:99. doi:10.1186/s12931-014-0099-9

81 Almagro P, Soriano JB, Cabrera FJ, et al, Working Group on COPD, Spanish Society of Internal Medicine*. Short- and medium-term prognosis in patients hospitalized for COPD exacerbation: the CODEX index. *Chest* 2014;145:972-80. doi:10.1378/chest.13-1328

82 Ntritsos G, Franek J, Belbasis L, et al. Gender-specific estimates of COPD prevalence: a systematic review and meta-analysis. *Int J Chron Obstruct Pulmon Dis* 2018;13:1507-14. doi:10.2147/COPD.S146390

83 Bellou V, Belbasis L, Konstantinidis AK, Evangelou E. Elucidating the risk factors for chronic obstructive pulmonary disease: an umbrella review of meta-analyses. *Int J Tuberc Lung Dis* 2019;23:58-66. doi:10.5588/ijtld.18.0228

84 Pérez-Padilla R, Ramirez-Venegas A, Sansores-Martinez R. Clinical Characteristics of Patients With Biomass Smoke-Associated COPD and Chronic Bronchitis, 2004-2014. *Chronic Obstr Pulm Dis* 2014;1:23-32. doi:10.15326/jcopdf.1.1.2013.0004

85 Vestbo J, Celli B. Chronic obstructive pulmonary disease: different risk factors and different natural histories?*Am J Respir Crit Care Med* 2014;190:968-70. doi:10.1164/rccm.201409-1705ED

86 Steyerberg EW, Harrell FEJr, 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;54:774-81. doi:10.1016/S0895-4356(01)00341-9

87 Moons KGM, Kengne AP, Woodward M, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental

value of a new (bio)marker. *Heart* 2012;98:683-90. doi:10.1136/heartjnl-2011-301246

88 Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J* 2014;35:1925-31. doi:10.1093/eurheartj/ehu207

89 Alba AC, Agoritsas T, Walsh M, et al. Discrimination and Calibration of Clinical Prediction Models: Users' Guides to the Medical Literature. *JAMA* 2017;318:1377-84. doi:10.1001/jama.2017.12126

90 Rubin DB. Inference and missing data. *Biometrika* 1976;63:581-92. doi:10.1093/biomet/63.3.581

91 Collins GS, Ogundimu EO, Cook JA, Manach YL, Altman DG. Quantifying the impact of different approaches for handling continuous predictors on the performance of a prognostic model. *Stat Med* 2016;35:4124-35. doi:10.1002/sim.6986

92 Riley RD, Ensor J, Snell KI, 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. doi:10.1136/bmj.i3140

93 Bonnett LJ, Snell KIE, Collins GS, Riley RD. Guide to presenting clinical prediction models for use in clinical settings. *BMJ* 2019;365:l737. doi:10.1136/bmj.l737

94 Nishimura K, Izumi T, Tsukino M, Oga T. Dyspnea is a better predictor of 5-year survival than airway obstruction in patients with COPD. *Chest* 2002;121:1434-40. doi:10.1378/chest.121.5.1434

95 Schols AM, Slangen J, Volovics L, Wouters EF. Weight loss is a reversible factor in the prognosis of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 1998;157:1791-7. doi:10.1164/ajrccm.157.6.9705017

96 Landbo C, Prescott E, Lange P, Vestbo J, Almdal TP. Prognostic value of nutritional status in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 1999;160:1856-61. doi:10.1164/ajrccm.160.6.9902115

97 Yang H, Xiang P, Zhang E, et al. Is hypercapnia associated with poor prognosis in chronic obstructive pulmonary disease? A long-term follow-up cohort study. *BMJ Open* 2015;5:e008909. doi:10.1136/bmjopen-2015-008909

98 Lange P, Halpin DM, O'Donnell DE, MacNee W. Diagnosis, assessment, and phenotyping of COPD: beyond FEV_1. *Int J Obstruct Pulmon Dis* 2016;11:3-12.

99 Celli BR, Cote CG, Marin JM, et al. The body-mass index, airflow obstruction, dyspnea, and exercise capacity index in chronic obstructive pulmonary disease. *N Engl J Med* 2004;350:1005-12. doi:10.1056/NEJMoa021322

100 Guerra B, Haile SR, Lamprecht B, et al, 3CIA collaboration. Large-scale external validation and comparison of prognostic models: an application to chronic obstructive pulmonary disease. *BMC Med* 2018;16:33. doi:10.1186/s12916-018-1013-y

101 Echevarria C, Gray J, Hartley T, et al. Home treatment of COPD exacerbation selected by DECAF score: a non-inferiority, randomised controlled trial and economic evaluation. *Thorax* 2018;73:713-22. doi:10.1136/thoraxjnl-2017-211197

102 Steer J, Gibson J, Bourke SC. The DECAF Score: predicting hospital mortality in exacerbations of chronic obstructive pulmonary disease. *Thorax* 2012;67:970-6. doi:10.1136/thoraxjnl-2012-202103

103 Cook R, Thomas V, Martin R, NIHR Dissemination Centre. People with chronic obstructive pulmonary disease exacerbations prefer early discharge, then treatment at home. *BMJ* 2019;364:k5339. doi:10.1136/bmj.k5339

Supplementary tables