



Prediction Models in Chronic Obstructive Pulmonary Disease – Informing Practice and Improving Research

Milo A. Puhan¹, MD, PhD and Gerben ter Riet², MD, PhD

¹Epidemiology, Biostatistics and Prevention Institute (EBPI), University of Zurich, Zurich, Switzerland; ²Academic Medical Center, University of Amsterdam, Department of General Practice, Amsterdam, The Netherlands

ABSTRACT

Prediction models have the potential to advance personalised medicine through risk-stratified management, which optimises the benefits and harms of preventive and therapeutic interventions for individuals. Thereby, the risk that patients are either under- or over-treated can be minimised. In research, prediction models can be used to efficiently identify study subjects, to control for confounding and to study subgroup effects. In the area of chronic obstructive pulmonary disease (COPD), many prediction models have been developed to predict the risk of early death or exacerbations, the course of disease or for other outcomes. A prerequisite for their use is a careful development and external validation, which is, however, often lacking. This article describes the purposes of prediction models as well as their development and validation including novel approaches like multiple score comparison meta-analysis. It also describes research needs like the need for developing accurate and valid prediction models for exacerbations that support risk-stratified prevention of exacerbations. (BRN Rev. 2018;4(3):156-68)

Corresponding author: Milo Puhan, miloalan.puhan@uzh.ch

Key words: Chronic obstructive pulmonary disease. Prediction models. Prevention. Risk stratification. Treatment.

Correspondence to:

Milo A. Puhan, MD, PhD
Hirschengraben 84, Room HRS G29
CH-8001 Zurich, Switzerland
E-mail: miloalan.puhan@uzh.ch

Received in original form: 12-01-2018
Accepted in final form: 23-02-2018
DOI: 10.23866/BRNRev:2017-0018

PREDICTIONS ARE EVERYWHERE

Predictions are made every day and everywhere¹. Weather forecasts and predictions in sports are among the most popular predictions. Companies make business forecasts, public administrations predict the needs of their people and (re-)insurance companies predict the probability of natural disasters or the risks of untoward events of their customers. In the medical field, predictions play an important role on both the population and individual level. Public health agencies need to predict the burden of particular diseases and the prevalence of risk factors as well as the probability of infectious disease outbreaks²⁻⁴. In clinical medicine, predictions are used to inform patients about risks of disease or disease outcomes and to guide preventive and therapeutic interventions for individuals⁵.

DEFINITION AND PURPOSE OF PREDICTION MODELS

The most common (but not only) scientific basis for making predictions about health outcomes for individuals is prediction models. Prediction models combine a number of patient characteristics to predict the onset of a disease or some outcome of interest in patients with a particular (main) disease^{5,6}. Individual characteristics can be anything from the genotype to phenotypic (clinical manifestations and biomarkers) and socio-demographic information. Often, phenotypic predictors are among the strongest predictors since they incorporate life and disease history that lead to the particular phenotypic

manifestation. Exercise capacity, for example, is the strongest known predictor of mortality in patients with chronic obstructive pulmonary disease (COPD), presumably because the disease has substantial impact on the musculoskeletal system through physical inactivity and systemic inflammation, which in turn have a major impact on longevity⁷.

Prediction models fulfil purposes in both practice and research. By providing a probability of the outcome occurring within a given time frame, prediction models inform patients and caregivers about prognosis. Once a diagnosis is established, the important question for patients is: in what respects will this diagnosis (i.e. this disease) affect my future? This is, in essence, a prognostic question that can be informed by prediction models and it is often of value for patients and clinicians. Also, prediction models are arguably most useful if they help guide the choice of further actions (e.g. secondary or tertiary prevention)^{8,9}. However, prediction models alone cannot serve the purpose of risk-stratified management. In order to identify different risk strata in which the benefit harm balance varies, quantitative benefit harm assessment or decision models are needed^{8,9,11}. But once such risk strata are established, prediction models allow clinicians to assign individuals to these risk strata and thus guide the choice of interventions that provide more benefits than harms. For example, in cardiovascular prevention, risk scores such as the Framingham risk score are used to predict the risk of a heart attack over a specific period of time (e.g. ten years)¹². Persons at low risk (e.g. < 5%) are advised to maintain or adopt a healthy lifestyle, whereas in persons at greater risk drug interventions (e.g. low dose aspirin or

statins) may be acceptable since the absolute benefit in terms of risk reduction outweighs the side effects and cost that incur for the individual or society^{13,14}.

In the area of COPD, there are medications like roflumilast that reduce the risk of exacerbations but come with significant side effects and costs¹⁵. A quantitative benefit harm assessment showed that the benefits of a phosphodiesterase-4 inhibitor (roflumilast) only outweigh its harms if the risk for severe exacerbations requiring hospital admissions over one year exceeds 20%¹⁶. Patients at lower risk for mild and moderate exacerbations will have little benefit from the drug since the absolute risk reduction is small, while they may experience the gastrointestinal, neurological and psychiatric side effects of roflumilast. In order to identify patients at such levels of risk, prediction models are useful and often outperform the judgment of clinicians¹⁷.

There are multiple uses of prediction models in research. In randomised controlled trials (RCTs), prediction models help to identify those patients in whom the balance of benefits and harms is uncertain. For example, the risk of harms was well known for roflumilast before doing RCTs. It would not make sense or even be unethical to enrol patients at low risk of exacerbations since the harms of roflumilast will always outweigh the benefits in those patients. This is why the pivotal trials of roflumilast only enrolled patients with COPD who had at least one COPD exacerbation requiring systemic glucocorticosteroid or treatment in hospital in the year prior to enrolment¹⁸. However, there was not specific risk threshold that was pre-specified nor if such a risk refers to exacerbations of any

severity, or to only moderate or severe exacerbations.

Figure 1 shows the study flow of a potential RCT of roflumilast when added to other treatments such as long-acting bronchodilators with or without inhaled glucocorticosteroid and/or to pulmonary rehabilitation. In a first step, patients who had exacerbations of any severity in the previous two years are identified out of all patients with COPD (of, e.g. a clinical centre) for more detailed eligibility assessment. A prediction model that estimates the risk of a severe exacerbation (requiring a hospital admission) over the next twelve months is then applied to these patients in order to identify the group of patients who have a risk of at least 20%¹⁶ or whatever risk threshold is defined a priori. It is important to plan this prospectively in order to have all information needed for the prediction model, to pre-specify the risk threshold and to enrol only patients above that risk threshold. If done post hoc, for example, by looking at the effects and the benefit harm balance in a subgroup of patients defined after completion of a trial, the results may not be as valid and reproducible.

In addition, prediction models provide an efficient way to adjust for confounding, this is for differences between groups in terms of prognostically important characteristics that invalidate group comparisons. Prediction models are ideal to control for confounding, either through restriction, pre-stratification/matching or statistical adjustment, since they carry more prognostic information than single confounders and since it is statistically efficient to use only one overall variable ("confounder score") instead of at least one

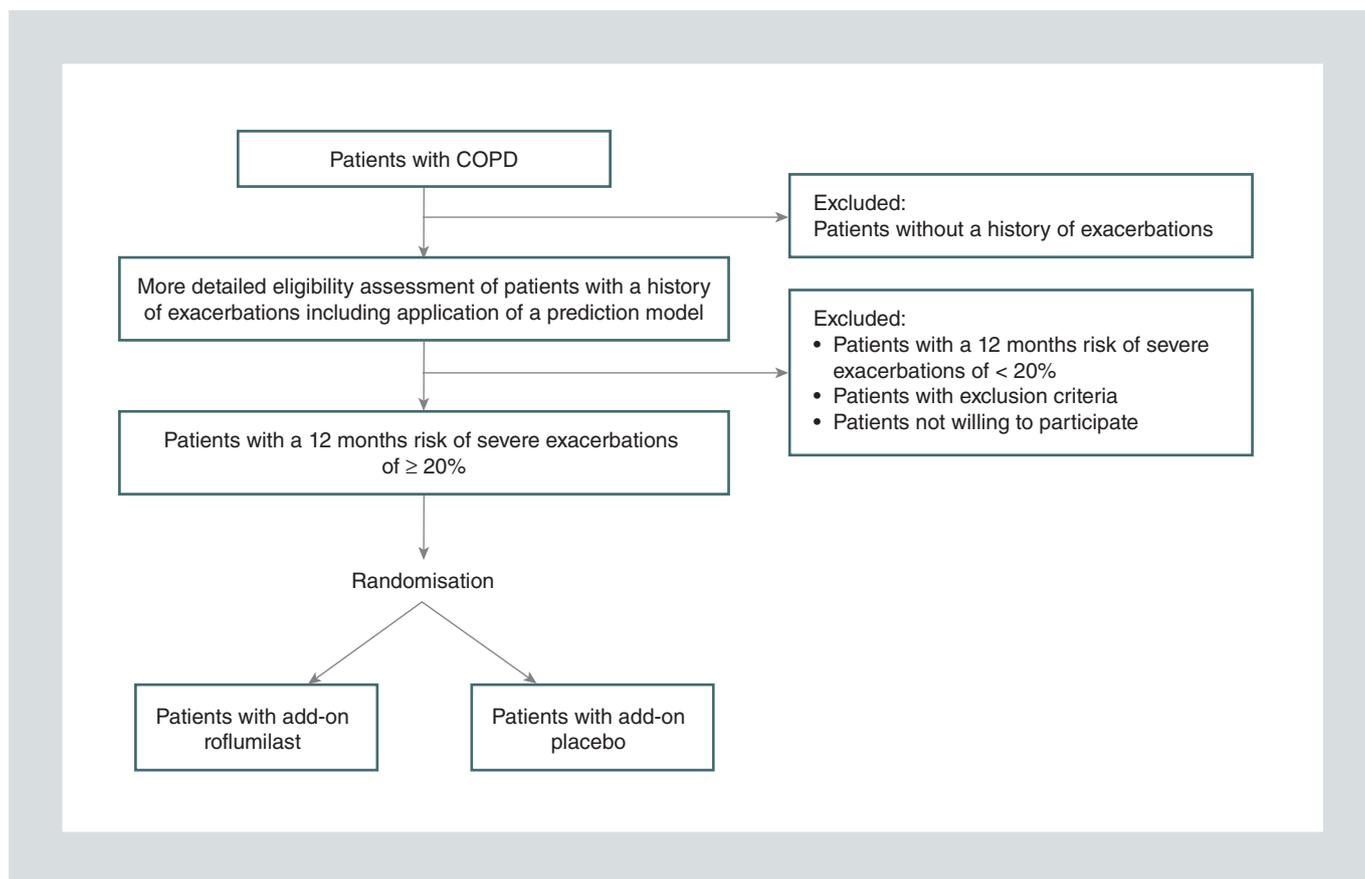


FIGURE 1. Use of a prediction model for assessment of eligibility in a randomised controlled trial. COPD: chronic obstructive pulmonary disease.

separate variable for each confounder. Finally, prediction models may also be used for subgroup analysis. It has been shown that it is more likely to identify subgroup effects if multivariable prediction models carrying much prognostic information are used than if single effect modifiers such as age or disease severity are used¹⁹.

PREDICTION MODELS ARE NOT ABOUT CAUSALITY

It is important to note that prediction of an event does not equal its aetiology. The goal of prediction models is to predict the outcome

of interest as accurately as possible irrespective whether or not the predictors in the model are causally related to the outcome. There has been some debate about whether age should or should not be included in prediction models²⁰. Age is included in many risk scores throughout medicine including the Framingham risk score where it contributes importantly to accurate prediction of cardiovascular events¹². Reasons against including age are, as is sometimes argued, that age is not modifiable and that including age, since it often is such a strong predictor of health events, leads to higher absolute risks and may trigger an intervention, although an elderly individual may not experience the benefit anymore. We

argue that the decision to treat or not to treat should not depend just on an absolute predicted risk of a single outcome (e.g. an exacerbation) but that a decision model or quantitative benefit harm assessment would show whether an elderly individual at high risk would still benefit from an intervention. This depends on when a (preventive) effect can be expected, on how strong the competing risk of death is and on other factors such as comorbidities, patient preferences or cost. Excluding age a priori because it is not modifiable, however, would be a missed opportunity when it contributes importantly to the prediction of the outcome of interest.

In addition, prediction does not need to directly guide decisions on preventive or therapeutic interventions by including only predictors that are modifiable. The latter may actually be misleading since acting on the modifiable risk factors does not guarantee that the probability of the outcome will change. For example, by modifying dyspnoea (part of both the Body mass index, airflow Obstruction, Dyspnoea and Exercise capacity [BODE] and Age, Dyspnoea and airflow Obstruction [ADO] index) it is not certain at all that mortality will be affected. In fact, there is no evidence to support that (short- or midterm) changes in prognostic scores lead to a modified risk of early death. Also, while many RCTs showed that long-acting bronchodilators and pulmonary rehabilitation meaningfully reduce dyspnoea, the effects on mortality are less certain²¹⁻²³. Thus, having risk scores like BODE and ADO as outcomes in RCTs is not that meaningful because changes in these scores may or may not translate into improvements in survival. Hence, the most important criterion for a

predictor to be included in a prediction model is whether it contributes to the prediction irrespective of it being causally related or modifiable.

PREDICTIONS IN CHRONIC OBSTRUCTIVE PULMONARY DISEASE – USEFUL PREDICTIONS AND CURRENT PREDICTION MODELS

In the area of COPD, there are now numerous prediction models to predict exacerbations with or without hospital admissions, the risk of early death (e.g. in the next three years) or health-related quality of life²⁴⁻²⁶. These prediction models can all serve the purpose of informing patients about prognosis and they can be used in research as described above. For clinical practice, prediction models that inform a specific decision are most useful. One could argue that prediction models for exacerbations are the most useful ones since they allow for risk-stratified prevention of exacerbations. Reducing the risk of exacerbations or at least shifting the severity of exacerbations from severe-to-moderate ones (i.e. avoiding exacerbations that require a hospital admission) is one of the major treatment goals in COPD²⁷⁻²⁸. Means to reduce the risk of exacerbations or to shift the severity of exacerbations include drug interventions (i.e. long-acting bronchodilators, inhaled glucocorticosteroid and roflumilast) and non-pharmacological interventions (i.e. exercise, enhancement of physical activity and COPD self-management education). Some of them come with harms and costs and are only indicated if patients have some risk of exacerbations. Valid prediction models for exacerbations would be very useful to guide the prescription of such interventions.

An additional approach to guide treatment selection is the use of biomarkers that predict response to treatment (e.g. the presence of blood eosinophils to predict the response to inhaled glucocorticosteroid). Such information may not guide treatments alone since it usually focuses on a specific treatment response and not on all benefit and harm outcomes. Information about (expected) response to treatment could easily be integrated into quantitative assessments of the benefit harm balance, this is by considering different treatment effects for those with or without expected response to treatment. The result of such an assessment would then inform above which outcome risks the benefits outweigh the harms for patients with or without likely response to treatment. Thereby, information on response to treatment and from prediction model act together to tailor treatments to individual patients.

Prediction models for mortality have the longest history in COPD. The BODE index sent the important message that the assessment of prognosis should go well beyond lung function²⁶. The ADO, the St. Georges Respiratory Questionnaire, Air-Flow obstruction and Exercise tolerance (SAFE), the Body mass index, severe Acute Exacerbations of COPD and Dyspnoea (B-AE-D) and other scores followed and included predictors like forced expiratory volume in one second (FEV_1) in percent predicted, dyspnoea, exercise capacity, age, previous exacerbations or body mass index. In a network meta-analysis of 26 COPD cohort studies, external validations and comparisons of these scores showed that the ADO has the best predictive properties of all scores. It is, interestingly, the only score that includes age²⁹.

A recent systematic review identified and critically appraised all models to predict exacerbations²⁴. Twenty-seven models were identified some of which were explicitly described as prediction models whereas others were multivariable models without explicit reference to the use as prediction models. It was difficult to compare the models and to identify the best models since the definition of exacerbations, time horizon, statistical methods and the measures of performance differed greatly. The greatest limitation was that only two out of 27 models were externally validated thwarting any judgment about the accuracy of most models in populations other than the cohort in which the models were developed. None of the models seemed ready to serve for risk-stratified prevention of exacerbations and there is a clear need for thoroughly developed and validated models for exacerbations.

Finally, there are also prediction models for health-related quality of life. These models help to identify patients who are at risk for a substantial decline and for whom the management may be optimised, for example through pulmonary rehabilitation. In a multicentre primary care cohort, prediction models were developed for each domain of the Chronic Respiratory Questionnaire as well as for its total score²⁵. Nomograms were provided to project health-related quality of life for 6 and 24 months. The Health, Activity, Dyspnoea and Obstruction (HADO) score is another score to predict health-related quality of life but its development was restricted to men³⁰.

In summary, a number of prediction models exist in the area of COPD. While some of

them are used in research, few of them have entered clinical practice. Important barriers for clinical use are that some predictors are not easily available, that most scores do not provide outcome probabilities that could be linked to decisions on prevention or treatment (i.e. do not inform risk-stratified management), and that most models were not internally, nor, very importantly, externally validated. The following sections review the most important steps for the development and validation of prediction models and highlight existing research gaps.

KEY ASPECTS FOR THE DEVELOPMENT OF PREDICTION MODELS

The development of a prediction model does not just include fitting a statistical model to some conveniently available data⁵. For a prediction model to be potentially meaningful and generalisable, a number of steps should be followed (Fig. 2). The first step is to determine the intended purpose(s) of a prediction model (i.e. specific clinical and/or research use). This includes the definition of the population (e.g. COPD patients) or a subpopulation (e.g. COPD patients on the intensive care unit or COPD patients in primary care) and of the specific outcome including its time horizon. For example, if exacerbations are to be predicted it should be specified if they are event- or symptom-based and if time to exacerbation, exacerbation counts or exacerbations within the specified time horizon (e.g. yes versus no within two years) will be modelled.

If a preventive or therapeutic decision is to be informed by the prediction model for risk-stratified management, the relevant prevention and

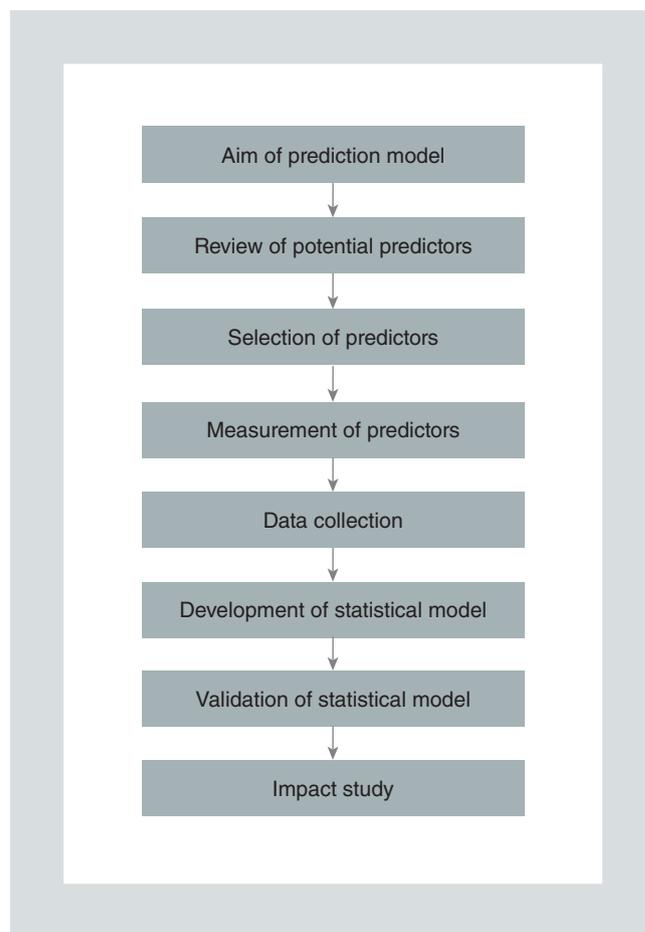


FIGURE 2. Overview of the development and validation process of prediction models.

treatment options for patients at different risks should be defined. Ideally, knowledge about the benefit harm balance for each of these options and for each risk group is available. Based on such knowledge, which is often absent, one could, for example, decide that COPD patients at low risk for a severe exacerbation requiring a hospital admission within the next year (e.g. < 10%), should receive one or two long-acting bronchodilators, that patients at moderate risk (e.g. 10-20%) should be offered additional pulmonary rehabilitation and that patients at higher risks (> 20%) may receive inhaled glucocorticosteroid in addition¹⁶.

After defining the purpose, the literature should be reviewed for already existing prediction models. If, as the recent systematic review showed, none of the existing prediction models, without or with external validation, serves the desired purpose, potential predictors for the prediction model should be identified from the literature. This may lead to a list of predictors that is likely too long both in terms of feasibility (i.e. measuring all of them in the development and validation cohorts) and statistical considerations (as a rule of thumb, ten events should be observed for each candidate predictor at the beginning of the statistical model development). Criteria to shorten the list are the strength of the predictor in previous studies (e.g. having suffered exacerbation previously is a much stronger predictor of exacerbations than body mass index), the availability and cost of the predictors in practice and the most valid and accurate way to measure a predictor.

The next step is planning and conduction of the actual study, which is typically a cohort study. Existing cohort studies are often used to develop prediction models³¹; this is both valid and efficient if these studies had the required follow-up, included the relevant population and measured the predictors and outcomes in a valid way. However, often researchers need to make compromises when using existing cohort studies, which may limit the usefulness and quality of prediction model. Prospectively planned cohort studies offer the advantage that all the necessary data can be collected and form an ideal basis to develop a prediction model. Finally, the statistical model, often a regression model, is developed. It is beyond the scope of this article to go into the details and there exists

excellent guidance for the development of prediction models^{5,32}. Note that the statistical analysis plan should already consider the external validation and the use of the prediction model in other cohorts of patients and, finally, in clinical practice and research. A common mistake is trying to fit the “perfect” model in the development cohort, which carries a great risk of over-fitting and over-optimism about the performance of the model outside the development cohort. The goal for the development of a prediction model should be to develop the model that performs best in both the development and validation cohort.

THE IMPORTANCE OF VALIDATION OF PREDICTION MODELS

It is a consistent observation across the medical field, that most of the prediction models are never validated. A systematic review from the cardiovascular field showed that just a third of the more than 300 existing models to predict cardiovascular events were externally validated³¹. The recent systematic review on prediction models in COPD showed that only two out of the 27 models for predicting exacerbations had been externally validated²⁴. External validation means that the prediction model is applied to a cohort of patients different from the development cohort³³. Very importantly, the regression model must not be developed again in the validation cohort but the original model from the development cohort (i.e. the coefficients and the intercept) must be applied to each patient of the validation cohort. The cohort may come from the same setting but different time period (i.e. temporal external validation) or

from different settings and countries (i.e. geographical external validation).

Why is it so important to externally validate prediction models? A prediction model almost always performs worse in cohorts different from the development cohort. For example, the BODE index showed an area under the curve (AUC) of 0.74 in the development cohort but only of 0.67 and 0.62 in two validation cohorts^{26,34}. Or the recently optimized B-AE-D showed an AUC of 0.75 in the development cohort and of 0.65 in the validation cohort and an AUC of 0.62 in the multiple score comparison that involved 26 COPD cohorts and 16,000 COPD patients^{29,35}.

Poorer performance in cohorts other than the development cohort happens almost by definition since the procedures to develop a prediction model fit a model to the data at hand as much as possible (i.e. by minimising their residuals). That model will be less accurate in most validation cohorts because the latter are, on average, at higher or lower risk for the outcome (e.g. if the validation cohorts comes from a less or more specialised setting than the development cohort), because the specific predictors are more or less predictive in the validation cohort (e.g. due to different distributions of predictors and the outcome), or because poor statistical methods were used for the development of the prediction model.

Because of the great importance of external validation, ideally, developers of a prediction model should organize a collaboration with one or more other cohorts³⁶. Thereby, the prediction model can be externally validated right away and potential problems of generalisability identified. If the performance in the validation cohort is considered insufficient for a specific

clinical or research purpose, there is a useful 8-step framework to update the model starting with the smallest possible revision (i.e. adjusting the intercept to the validation cohort in order to adjust for differences in risks between the development and validation cohort) up to a completely new model (new predictors, new regression coefficients and standard errors)⁵. If only a minor update is needed (e.g. updating the intercept) the model may be ready for use in practice and research. If, however, a more extensive update is needed (e.g. re-estimating coefficients and adding new predictors), the updated model should be externally validated in yet another cohort.

The two most common metrics used for validation are the (statistical or graphical) comparison of predicted and observed risks to assess calibration and the AUC for discrimination. Figure 3 shows some examples from the large-scale validation of the ADO index using ten COPD cohort studies and 13,914 patients³⁶. The goal of that validation was to assess discrimination and calibration of the original ADO index³⁴. While discrimination was still high in the entire validation cohort (AUC 0.82) calibration was insufficient. The comparison of predicted and observed risks in the first part of the validation (upper part of figure 3, before re-calibration of intercept) shows that, overall, the predicted and observed risks were comparable (i.e. 10.0% versus 9.0%) but that within deciles of patients, the predicted and observed risks did not match. A simple update of the ADO index, i.e. an adjustment of the intercept, was not enough to improve the match of predicted and observed risks. Therefore, we performed a more extensive model update, which required another validation in a new set of cohort of COPD patients.

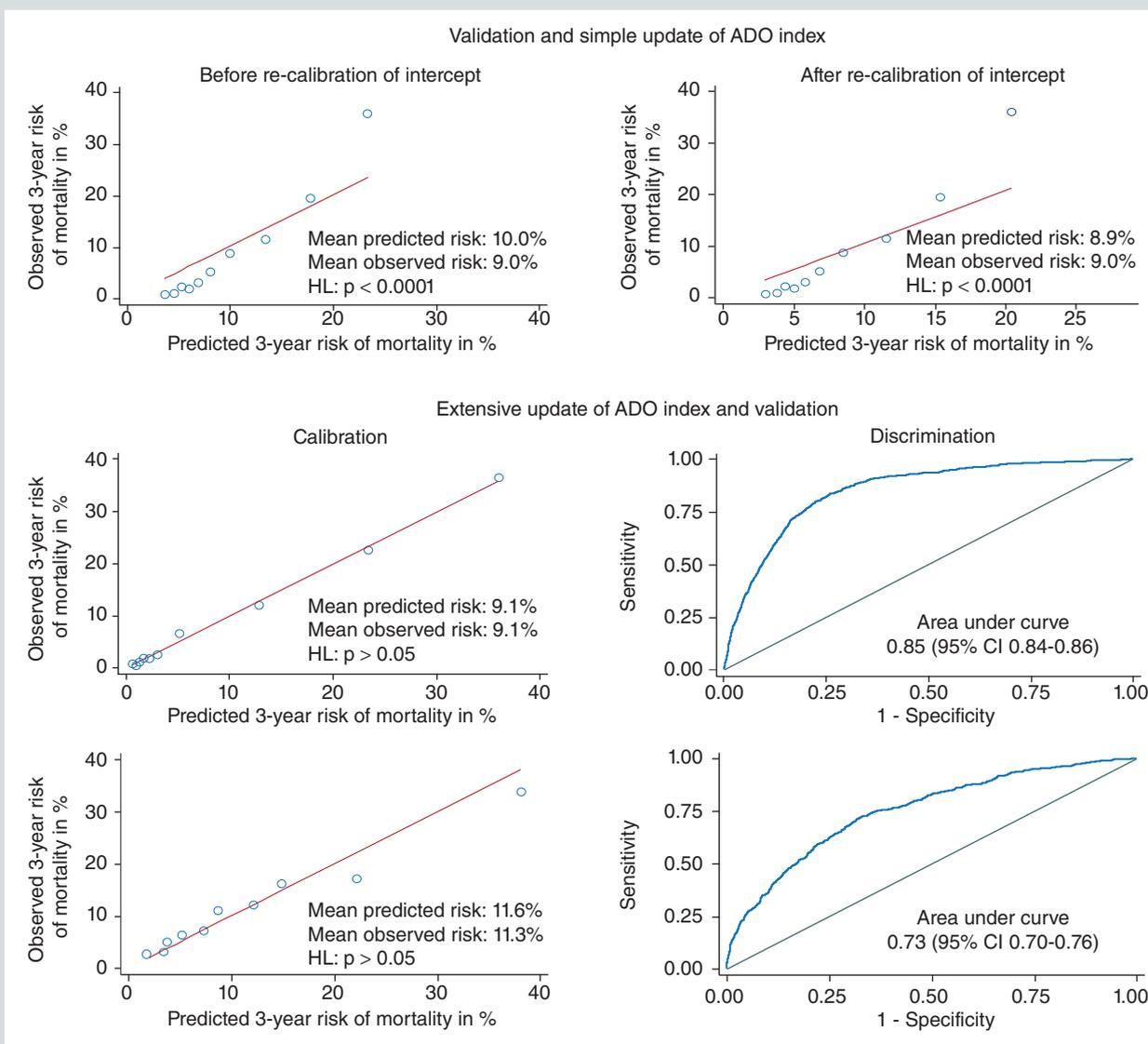


FIGURE 3. Validation, update and further validation of the Age, Dyspnoea and airflow Obstruction (ADO) index.

Validation and simple update of the ADO index: The two plots show predicted versus observed risk in the validation cohort (i.e. in all patients from the ten cohorts except the one cohort in which the ADO was originally developed). Each circle represents 10% of patients that were ordered according to their 3-year predicted risk of death (i.e. deciles). For each patient in each decile the 3-year risk of death was calculated based on the original ADO index³⁴. For each decile this (average) predicted risk was compared to the observed risk in that decile. If calibration is perfect, all circles lie on the red line that indicates perfect agreement between predicted and observed risks.

These plots show that calibration was not good even after a small update (of adjusting the intercept to the risk of the validation cohort) and the statistical test (Hosmer Lemeshow [HL] test) showed that the predicted and observed risks differed statistically significantly.

Extensive update of the ADO index and validation: The two plots on the left compare again predicted versus observed risks. The upper plot shows calibration in the five cohorts in which the ADO index was newly developed (same predictors but new regression model). As expected (since the data were fitted to that cohort), the predicted and observed risks in the deciles of patients matched very well. This was also true in the validation cohort (the other five cohorts) where the circles representing deciles of patients cluster around the red line of perfect agreement. Only at higher risks of 3-year mortality the predicted risks were higher than the observed risks. The two plots on the right show the area under the curve. As it is commonly observed even in large samples like in the large-scale ADO validation, the area under the curve is substantially smaller in the validation cohort (*reproduced with permission from Puhan MA et al.*³⁶).

CI: confidence interval.

NOVEL DEVELOPMENTS FOR THE VALIDATION AND COMPARISON OF PREDICTION MODELS

Beside external validation, a comparison of existing prediction models in terms of their predictive properties is needed, but rarely done. Ideally, clinicians and researchers would be able to choose from existing prediction models based on the ability to accurately predict the outcome as well as based on the availability of the predictors and the cost of their collection. Until recently, no method was available to concurrently validate and compare prediction models. We developed the novel method of multiple score comparison that is based on network meta-analysis of RCTs. In multiple score comparison, different prediction models or the scores (usually simple scorings derived from the prediction models) can be validated and compared within and across a number of cohort studies. For example, based on the 26 cohort studies with more than 16,000 COPD patients, we evaluated ten scores to predict 3-year mortality. Between two and nine scores could be calculated for each of the 26 cohort studies (Fig. 4)³⁷.

The commonly used BODE index had a median AUC of 0.68. All other scores, except the ADO index, had a lower AUC. The ADO index (combining age, dyspnoea and airflow obstruction) showed a higher AUC (0.695) and a probability of being the score with the greatest discrimination of 96%²⁹. Multiple score comparison can be applied to any prediction models and scores (across the medical field) and fills an important gap since it provides a concurrent external validation and comparison of as many scores as desired^{38,39}.

GAPS AND RESEARCH NEEDS FOR PREDICTION MODELS IN CHRONIC OBSTRUCTIVE PULMONARY DISEASE

Prediction models, in particular to predict mortality, are commonly used in research but not in practice. A number of steps are missing to introduce them into practice and to further advance risk-stratified management, a form of personalised medicine. Firstly, it is currently unclear in what situation and for which decisions clinicians and patients would like to have outcome predictions. Qualitative studies could shed light on these needs and help define the potential purposes of prediction models in care for COPD patients. Secondly, when new prediction models are developed, a thorough development and validation process is a pre-requisite for their use in practice. Thus, there is a clear need to improve the standards in this area. Thirdly, if the information from prediction models is linked to prevention and treatment decisions, the risk categories, which would guide the selection of preventive and symptom-relieving interventions, need to be defined. For example, it needs to be explored at what outcome risk for exacerbations (e.g. > 20% risk of an exacerbation in the coming 12 months) adding an inhaled corticosteroid to long-acting bronchodilator(s) improves patient outcomes with acceptable side effects and costs. Quantitative benefit harm assessment is a way to define such AUC thresholds. Finally, impact studies are needed. Impact studies assess if the use of prediction models and risk-stratified management, respectively, ultimately lead to better patient outcomes. This can be done using modelling studies or, if resources allow, by randomised trials that compare health outcomes of patients who receive disease management strategies with versus without the use of prediction models.

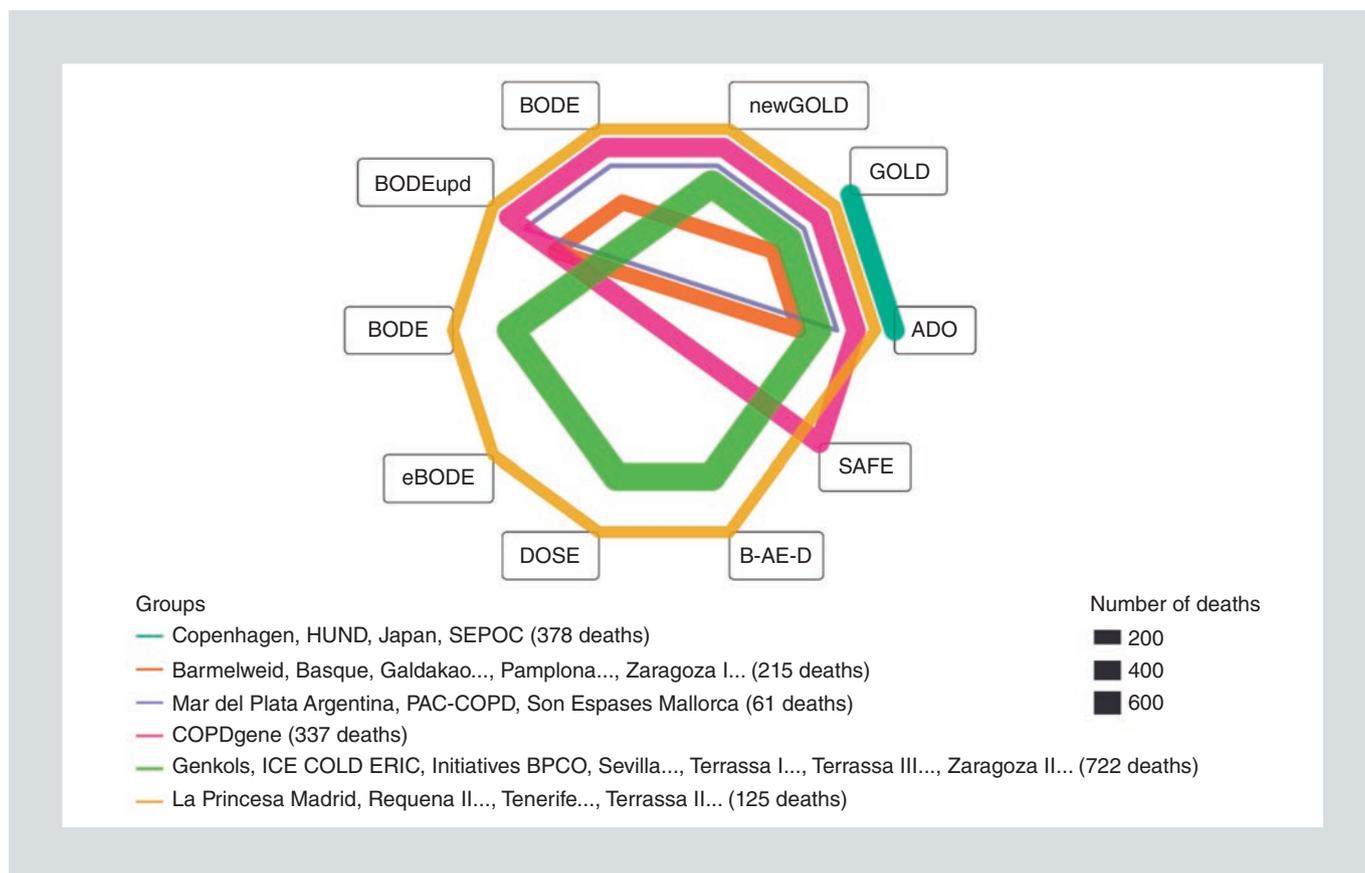


FIGURE 4. Simultaneous comparison of prognostic scores.

Network plot: The figure shows how many times the ten prognostic scores could be compared in the different cohort studies. Each closed polygon represents a group of cohort studies where the same prognostic scores were available. The thickness of the lines represents the total number of deaths in the specific group (*reproduced with permission from Haile SR²⁹*).

ADO: Age, Dyspnoea, airflow Obstruction; B-AE-D: Body mass index, severe Acute Exacerbations of COPD and Dyspnoea; BODE: Body mass index, airflow Obstruction, Dyspnoea and severe Exacerbations; BODEx: Body mass index, airflow Obstruction, Dyspnoea, severe acute Exacerbation of COPD; BODE upd.: BODE updated; e-BODE: severe acute exacerbation of COPD plus BODE; DOSE, Dyspnoea, Obstruction, Smoking and Exacerbation frequency; GOLD: Global Initiative for Chronic Obstructive Lung Disease; SAFE: St. Georges Respiratory Questionnaire, Air-Flow obstruction and Exercise tolerance.

CONCLUSIONS

Prediction models can serve a number of purposes in both clinical practice and research. They have the potential to advance personalised medicine through risk-stratified management, which optimises the benefits and harms of preventive and therapeutic interventions for individuals, thereby minimising the risk that patients are neither under- nor overtreated. A prerequisite for risk-stratified

management is to have a prediction model that is thoroughly developed, validated and proofed to perform best compared to other prediction models. In the area of COPD, the most intriguing area for risk-stratified management is the prevention of exacerbations, in particular severe exacerbations. But more accurate and widely applicable prediction models for exacerbations and risk-stratified management strategies still need to be developed.

CONFLICTS OF INTEREST

Dr. Milo Puhan and Dr. Gerben ter Riet are the principal investigators of the Barmelweid and ICECOLD ERIC cohorts, led the International COPD Cohorts Collaboration Working Group and are investigators of the 3CIA collaboration. They have developed the ADO index and the prediction Model for health-related quality of life and contributed to the development of the Multiple Score Comparison methodology. They have no financial conflicts of interest.

REFERENCES

- Silver N. The signal and the noise: why so many predictions fail-but some don't. Penguin Press; 2012.
- WHO | Projections of mortality and burden of disease, 2002-2030. WHO [Internet]. World Health Organization; 2014; Available from: http://www.who.int/healthinfo/global_burden_disease/projections2002/en/
- Fischer LS, Santibanez S, Hatchett RJ et al. CDC Grand Rounds: Modeling and Public Health Decision-Making. *MMWR. Morb. Mortal. Wkly. Rep.* 2016;65:1374-7.
- Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Med.* 2006;3:e442.
- Steyerberg EW. Clinical prediction models: a practical approach to development, validation, and updating. Springer; 2009. ISBN 978-0-387-77244-8
- 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.
- Puhan M, Siebeling L, Zoller M, Muggensturm P, ter Riet G. Simple functional performance tests and mortality in COPD. *Eur Respir J.* 2013;42:956-63.
- Yu T, Vollenweider D, Varadhan R, Li T, Boyd C, Puhan MA. Support of personalized medicine through risk-stratified treatment recommendations - an environmental scan of clinical practice guidelines. *BMC Med.* 2013;11:7.
- Agustí A, Antó JM, Auffray C et al. Personalized respiratory medicine: exploring the horizon, addressing the issues. Summary of a BRN-AJRCCM workshop held in Barcelona on June 12, 2014. *Am J Respir Crit Care Med.* 2015;191:391-401.
- Guo JJ, Pandey S, Doyle J, Bian B, Lis Y, Raisch DW. A review of quantitative risk-benefit methodologies for assessing drug safety and efficacy-report of the ISPOR risk-benefit management working group. *Value Health.* 2010;13:657-66.
- Puhan M a, Singh S, Weiss CO, Varadhan R, Boyd CM. A framework for organizing and selecting quantitative approaches for benefit-harm assessment. *BMC Med Res Methodol.* 2012;12:173.
- Wilson PWF, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of Coronary Heart Disease Using Risk Factor Categories. *Circulation.* 1998;97:1837-47.
- Bibbins-Domingo K, Grossman DC et al. Statin Use for the Primary Prevention of Cardiovascular Disease in Adults. *JAMA.* 2016;316:1997-2007.
- Puhan MA, Yu T, Stegeman I, Varadhan R, Singh S, Boyd CM. Benefit-harm analysis and charts for individualized and preference-sensitive prevention: example of low dose aspirin for primary prevention of cardiovascular disease and cancer. *BMC Med.* 2015;13:250.
- Chong J, Leung B, Poole P. Phosphodiesterase 4 inhibitors for chronic obstructive pulmonary disease. *Cochrane Database Syst Rev.* 2017;9:CD002309.
- Yu T, Fain K, Boyd CM et al. Benefits and harms of roflumilast in moderate to severe COPD. *Thorax.* 2014;616-22.
- Moons KGM, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? *BMJ.* 2009;338:b375.
- Martinez FJ, Calverley PMA, Goehring U-M, Brose M, Fabbri LM, Rabe KF. Effect of roflumilast on exacerbations in patients with severe chronic obstructive pulmonary disease uncontrolled by combination therapy (REACT): a multicentre randomised controlled trial. *Lancet.* 2015;385:857-66.
- Hayward R a, Kent DM, Vijan S, Hofer TP. Multivariable risk prediction can greatly enhance the statistical power of clinical trial subgroup analysis. *BMC Med Res Methodol.* 2006;6:18.
- Ridker PM, Cook N. Should Age and Time Be Eliminated From Cardiovascular Risk Prediction Models? *Circulation.* 2005;111:57-658
- McCarthy B, Casey D, Devane D, Murphy K, Murphy E, Lacasse Y. Pulmonary rehabilitation for chronic obstructive pulmonary disease. *Cochrane Database Syst Rev.* 2015. 23;CD003793
- Puhan MA, Gimeno-Santos E, Cates CJ, Troosters T. Pulmonary rehabilitation following exacerbations of chronic obstructive pulmonary disease. *Cochrane Database Syst Rev.* 2016.12:CD005305.
- Scott DA, Woods B, Thompson JC et al. Mortality and drug therapy in patients with chronic obstructive pulmonary disease: a network meta-analysis. *BMC Pulm Med.* 2015;15:145.
- Guerra B, Gaveikaite V, Bianchi C, Puhan MA. Prediction models for exacerbations in patients with COPD. *Eur Respir Rev.* 2017;26:160061.
- Siebeling L, Musoro JZ, Geskus RB et al. Prediction of COPD-specific health-related quality of life in primary care COPD patients: a prospective cohort study. *NPJ Prim Care Respir Med.* 2014;24:14060.
- 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.
- GOLD 2017 Global Strategy for the Diagnosis, Management and Prevention of COPD - Global Initiative for Chronic Obstructive Lung Disease - Available from: <http://goldcopd.org/gold-2017-global-strategy-diagnosis-management-prevention-copd/>
- Wedzicha JA, Calverley PMA, Albert RK et al. Prevention of COPD exacerbations: a European Respiratory Society/American Thoracic Society guideline. *Eur Respir J.* 2017;50:1602265.
- Haile SR, Guerra B, Soriano JB, Puhan MA, 3CIA collaboration. Multiple Score Comparison: a network meta-analysis approach to comparison and external validation of prognostic scores. *BMC Med Res Methodol.* 2017; 17:172.
- Esteban C, Quintana JM, Aburto M, Moraza J, Capelastegui A. A simple score for assessing stable chronic obstructive pulmonary disease. *QJM.* 2006;99:751-9.
- Damen JAAG, Hooft L, Schuit E et al. Prediction models for cardiovascular disease risk in the general population: systematic review. *BMJ.* 2016;353:i2416.
- Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement. *Ann Intern Med.* 2015;162:55.
- Moons KGM, Kengne AP, Grobbee DE et al. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart.* 2012;98:691-8.
- Puhan M, 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.
- 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.
- Puhan MA, Hansel NN, Sobradillo P et al. Large-scale international validation of the ADO index in subjects with COPD: an individual subject data analysis of 10 cohorts. *BMJ Open.* 2012;002151.
- Guerra B, Haile SR, Lamprecht B et al. Large-scale external validation and comparison of prognostic models: an application to chronic obstructive pulmonary disease. *BMC Med.* 2018;16:33.
- Pennells L, Kaptoge S, White IR, Thompson SG, Wood AM, Emerging Risk Factors Collaboration RW et al. Assessing risk prediction models using individual participant data from multiple studies. *Am J Epidemiol.* 2014;179: 621-32.
- Riley RD, Ensor J, Snell KIE 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.