Development and validation of a prediction score to assess the risk of incurring in COPD-related exacerbations: a population-based study in primary care

Francesco Lapi, Ettore Marconi, Francesco Paolo Lombardo, Iacopo Cricelli, Claudio Cricelli

1. *Health Search, Italian College of General Practitioners and Primary Care, Florence (Italy)*
2. *Italian College of General Practitioners and Primary Care, Florence (Italy)*
3. *Genomedics SRL, Florence (Italy)*
4. …

Corresponding Author:
Francesco Lapi, PhD
Health Search, Italian College of General Practitioners and Primary Care
Via del Sansovino 179, 50142, Florence, Italy
Phone/fax: +39 055 494900
Email: lapi.francesco@simg.it

CRediT authorship contribution statement

Francesco Lapi: Conceptualization, Methodology, Data curation, Supervision, Writing-Original draft-Preparation. Ettore Marconi: Data curation, Software, Writing-Original draft preparation. Francesco Paolo Lombardo: Writing-Reviewing and Editing. Iacopo Cricelli: Data curation, Software, Writing-Original draft-Preparation. Claudio Cricelli: Conceptualization, Supervision, Writing-Reviewing and Editing

Declaration of competing interests

FL and EM provided consultancies in protocol preparation for epidemiological studies and data analyses for GSK, Astra Zeneca, and Chiesi. FPL provided clinical for GSK, Astra Zeneca, and Chiesi. CC provided clinical and scientific consultances for epidemiological studies for GSK, Astra Zeneca, and Chiesi.

Acknowledgments

We would to thank Dr. Serena Pecchioli and Dr. Elisa Bianchini, PhD, who provided support for the preparation of the analyses plan and execution of statistical analyses. The study supported by an unconditional grant from Astra Zeneca.

Abstract

Background. Chronic obstructive pulmonary disease (COPD) is the fourth most important cause of death in high-income countries. Treatment misuse of COPD inhaled therapy, including suboptimal adherence (only 10% to 40% of patients reporting an adequate compliance) may shrink or even nullify the proven benefits of these medications. As such, an accurate prediction algorithm to assess the risk of COPD exacerbation might be relevant for general practictioners (GPs) to improve patient's therapy.

Methods. We formed a cohort of patients aged 45 years or older and diagnosed with COPD in the period between January 2013 to December 2021. Each patient was followed until the occurrence of COPD exacerbation up to the end of 2021. Up to 16 determinants were adopted to assemble the CopdEX(CEX)-Health Search(HS)core, which was therefore developed and validated in the related two sub-cohorts.

Results. In the study period, we idenfied 63763 patients aged 45 years or older, and diagnosed with COPD (mean age: 67.8 (SD:11.7); 57.7% males).When the risk of COPD exacerbation was estimated via CEX-HScore, its value was equal to 14.22% over a 6-month follow-up. Discrimination accuracy and explained variation were equal to 66% (95% CI: 65-67%) and 10% (95% CI: 9-11%), respectively. The calibration slope did not significantly differ from the unit.

Conclusions. The CEX-HScore was featured by fair prediction accuracy for prediction of COPD-related exacerbations over 6-month follow-up. Such a tool might therefore support GPs to enhance COPD patients' care and improve their outcomes by facilitating personalized approaches through a score-based decision support system.

Keywords: COPD, exacerbations, prediction score, primary care

Introduction

Affecting 4% to 10% of the adult population, chronic obstructive pulmonary disease (COPD) represents a frequent and burdensome condition that is associated with disability, poor quality of life, and higher mortality. Today, COPD is the fourth most important cause of death in high-income countries.[1,2] Despite the development and implementation of new symptomatic pharmacological treatments have improved health-related quality of life and survival of COPD patients, a treatment misuse – including suboptimal adherence – may shrink or even nullify their proven benefits. According to studies of pharmaco-utilization carried out in the general population, adherence to COPD inhaled therapy is low, with only 10% to 40% of patients reporting an adequate compliance.[3,4] Inappropriate use of inhaled costicosteroids (ICS)[5–7] continues to be an issue among COPD sufferers as well. In this context, the role of accurate prediction algorithms to assess the risk of COPD exacerbation might be crucial to tailor pantient's treatments. Recently, the The Acute Chronic Obstructive Pulmonary Disease (COPD) Exacerbation Prediction Tool (ACCEPT) algorithm was efficiently re-calibrated.[8,9] This model comprises these determinants: number of moderate and severe exacerbations in the previous 12 months, age, sex, smoking status, observed versus predicted forced expiratory volume in one second, St Geaorge Respiratory Questionnaire (SGRQ) score, body mass index (BMI), the use of domiciliary oxygen therapy, use of statins (representing cardiovascular disease risk), and type of inhaled COPD medications (as a surrogate for the severity of COPD). In specific, The ACCEPT 2.0 tool was well calibrated with positive or negative exacerbation history. Nevertheless, this score was develop using patients' data stemming from ECLIPSE[10] cohort, whose external validity might be limited by the fact that it included 2000 COPD patients being enlisted more than 20 years ago in 12 different countries. They were likely unable to represent the heterogeneity of COPD patients being cared by GPs, as those belonging to the Italian settings. Furthermore some determinants, such as the SGRQ, are not regularly registered in general practice, while patients' features might be useful to assess the prediction accuracy and clinical utility to assess the risk of COPD-related exacerbations. For instance, the co-diagnosis of asthma, the presence of depression and/or anxiety which might leads to nonadherence to medications, as well as cardiovascular and metabolic diseases whose pharmacotherapy might impair the concurrent use of medications for COPD.[5]

In Italy, as in other coutries with similar primary care settings and organization, the role of GPs in estimating the risk of COPD exacerbations is particularly relevant given the most recent update of GOLD guidelines,[11] which provided a substantial revision of COPD pharmacotherapy, along with the Italian NOTA 99 (Italian Medications Agency: AIFA),[12] which regulates the reimbursement criteria of COPD medications by extending the prescription of certain respiratory medications to GPs, reminding the role of mandatory registration of Forces Expiratory Volume (FEV1), and providing further indication for specialist's referral.

Given this background, we developed and validated a score to predict the risk of COPD exacerbation in primary care.

Methods

This score was developed and validated according to PROBAST indications[13] and TRIPOD statements.[14]

Data source

We adopted the Health Search Database (HSD), an Italian general practice data source in place since 1998, comprising data from computer-based patient records registered by a selected group of general practitioners (GPs), uniformly distributed across Italy[15] GPs voluntarily agreed to collect patient information and to attend specific training courses for data entry.

In HSD, data are linked through a unique encrypted identification number for each patient. All diagnoses are coded according to the International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM). Drug prescriptions comprise information on trade name, formulation, and active substance, and are coded according to the Anatomic Therapeutic Chemical (ATC).

To be considered for participation in epidemiological studies, GPs must meet up-to-standard quality criteria pertaining to the levels of coding.[15] The research validity of HSD has been demonstrated by several publications, including those providing prediction scores.[16–20]

Study population

We formed a cohort of patients aged 45 years or older and diagnosed with COPD (**ICD9CM** 491.2* or 496*) in the period between January 2013 to December 2021. Each patient was followed until the occurrence of the following event whichever came first: COPD exacerbation (event date; see 'Outome definition' paragraph), death, loss of contact with the GP, or the end of data availability (31 December 2021). The cohort was randomly splitted into two sub-cohorts containing approximately two-third and one-third of patients, respectively; these were referred to as the development and validation sub-cohort.

Outcome definition

Moderate and severe COPD exacerbation were operationally defined as follows: moderate COPD exacerbation was defined as **ICD9CM** (491.21) and/or the incident co-prescription, in the same day ± 3 days of an antibiotics (**ATC** J01*) and a corticosteroid drug (**ATC** H02*). A severe COPD exacerbation was defined when a hospitalization being associated with a diagnosis of COPD or related exacerbation, occurred.[21–23] Given the absence of reason of hospital admission in HSD, the related free-text being registered within 1 month the event date, were manually inspected and validated the presence of a COPD-related event.

Candidate determinants

The selection of the candidate determinants was based on the current literature,[8,24,25] our previous work,[4,26] and/or clinical bases. Every covariate examined for the occurrence of COPD exacerbations were operationally defined in the overall period preceding or on the index date as follows: along with age and sex, smoking habits (current, former or non-moker; last measurement), presence of overweight or obesity, prior diagnosis of asthma, cancer, osteoarthritis, gastro-esophageal reflux disease (GERD), diabetes mellitus, cardiovascular disease, atrial fibrillation, heart failure, depression and/or anxiety, polypharmacy and type of COPD pharmacotherapy.

Data analysis

Descriptive statistics were reported as means with standard deviations (SDs) and proportions with 95% confidence intervals (CI) for continuous and categorical variables, respectively. The incidence rate of COPD-exacerbation was calculated by dividing the number of events (numerator) by person–years (denominator) being cumulated during follow-up.

Using the development sub-cohort, a multivariable Cox regression was therefore estimated by including all the candidate determinants. The 'age' variable was mean-centered to minimized the potential effect of outliers on the risk estimates. The effect size of each candidate determinant was reported as adjusted hazard ratio (aHR) with 95% CI. A patient-specific score was derived through linear combination of the estimated coefficients, thus forming the CopdEX(CEX)-Health Search(HS)core. Then, by applying the scores to the validation sub-cohort, we evaluted the predictive accuracy of the scores by calculating the explained variance (pseudo-R^2) and Area Under the Curve (AUC) as performance and discrimination measures, respectively.[27–29] The calibration slope was calculated and formally (i.e. equivalence hypothesis) tested as well. To evaluate the burden of overfitting of the CEX-HScore, we provided optimism-corrected[30] pseudo-R^2 , AUC, intercept and calibration slope by bootstrapping 200 random samples from the entire (original) study cohort.[31] In addition, we provided the calibration measure by using the locally-weighted scatterplot smoothing (LOWESS) curve, in which predicted vs observed risks of outcomes were fitted.[32] We calculated the 6-month predicted risk of COPD exacerbation as a function of the cumulative baseline hazard and the linear predictors, i.e. the sum of the product between values for the predictors identified for the individual patient and beta coefficients for each candidate determinant.[33]

To support clinical decision-making, we reported the CEX-HScore for those carrying at least one determinant, and sub-grouping them into three risk categories (low, intermediate and high risk), using cut-off points on the prognostic index determined by Cox's methods.[34,35]

As a sensitivity analysis, we attempted to extend the outcome estimation to 1-year event horizon.

Results

Over the study period, we idenfied 63763 patients diagnosed with COPD (mean age: 67.8 (SD:11.7); 57.7% males). Overall, the incidence rate of COPD moderate or severe exacerbations was equal to 11.3 case per 100

person-years. Among them, 11.2 and 0.12 cases per 100 person-years were the estimates for moderate and severe exacerbations, respectively. In this cohort, 42084 and 21679 COPD sufferers formed the development and validation sub-cohorts, respectively. The two sub-cohorts did not differ for what concerns demographic and clinical covariates. All univariate HRs and related 95% CI, for each candidate determinants, were indeed overlapped in the development and validation sub-cohorts. The occurrence of moderate and/or severe COPD exacerbations, we observed cumulative proportions of the outcome equal to 40 and 40.5%, in development and validation sub-cohorts (**Table 1**).

After forcing all the covariates in the multivariate model, we estimated the relationship between each candidate determinants and the occurrence of moderate/severe exacerbations. Namely, age (1% increase of COPD exacerbation for each additional year moving from mean age), smoking (1-2% increase), diagnosis of osteoarthritis (5% increase), gastroesophageal reflux disease (8% increase), and asthma (20% increase); the presence of polypharmacy (41% increase), history of prior moderate COPD exacerbations (90% increase) were statistically significant determinants as well. For what concerns the concurrent pharmacotherapies, prescripion of SABA, LAMA, ICS, and fixed combinations of LABA/ICS resulted associated with an increased risk of COPD exacerbation of 8, 8, 7, and 22%, respectively. Other candidate determinants, such as use of fixed LABA/LAMA combinations, showed a positive associations with the outcome although not statistically significant because of the reduced analysis power. In any case, on clinical bases, all candidate determinants were included in the multivariate model (**Table 2**). Then, the aforementioned beta coefficients were linearly combined to form an individual CEX-HScore for each COPD sufferer. When the risk of COPD exacerbation was estimated via CEX-HScore, its value was equal to 14.22% over a 6-month follow-up. Then, the score was applied to validation cohort by obtaining an AUC equal to 66% (95% CI: 65-67%) and pseudo-R^2 of 10% (95% CI: 9-11%). In terms of calibration, **Figure 1** depicts the calibration plot contrasting observed vs estimated COPD exacerbations over 6-month follow-up. In specific, the calibration slope was equal to 1.01 and the related test did not refuse the equivalence hypothesis vs "perfect" calibration.

The plot reported in **Figure 2** showed the correlation between observed and predicted risks for the validation su-cohort when the risk ranged 0-40%. In **Table 3** are reported the measures of prediction performance for the CEX-HScore, when they were calculated in the validation sub-cohort or after bootstrapping up to 200 random samples using the entire cohort. This approach allowed us to provide optimism-correted measures, so considering the effect of score overfitting. Namely, pseudo-R^2 and AUC were consistent between validation and bootstrap samples. For the letter, the coefficients stemming from each sample analysis were combined according to the Rubin's role.[36]

To provide GPs with risk thresholds which could be more easily applicable to their clinical practice, by means of the Cox method, we idenfied low, intermediate and high risk of COPD exacerbations related to cut-offs <9.4, between 9.4 and lower than 19, >=19%, respectively.

When we consiedered a 1-year follow-up for the sensitivity analysis, some overestimation was observed for those with high (>40%) predicted risk. Namely, the related AUC was equal to 63% (95% CI: 63-64%) the test for calibration slope resulted statistically significant (p value 0.001; **Supplementary Figure 1**).

Discussion

To our knowledge, this is the fist study which developed and validated a score to predict the risk of COPD-related exacerbations in primary care. The CEX-HScore was featured by good prediction accurary, as indicated by its discrimination and calibration measures. The identified determinants of COPD exacerbation aligned with previous research, so emphasizieng the multifactorial nature of this clinical entity. In particular, the strong association seen for the history of prior exacerbations confirmed the relavance of effective management for this patients' subgroup.

Previous prediction tools, such as the ACCEPT algorithm,[9] provided findings which were generally consistent with ours. The most recent recalibration of this score, which led to the ACCEPT 2.0,[8] adopted the data from the ECLIPSE study,[10] a comprehensive three-year observational study with 1,803 patients and 2,117 COPD-related exacerbations. By applying non-parametric regression splines to predicted rates, the researchers fine-tuned the tool to enhance its calibration and predictive power. The ACCEPT 2·0, reported an AUC of 0.76 for predicting the outcome, surpassing the AUC of 0·68 achieved by the current standard of care, which relies solely on exacerbation history. In addition, it performed well regardless of an individual's exacerbation history. To enhance clinical usability and simplicity, the authors explored reduced versions of ACCEPT by removing predictors such as symptom scores and baseline medications. While there was a slight reduction in predictive power for the occurrence of any severe exacerbations, the clinical utility of these simplified models remains promising.

Even though the ACCEPT 2.0 adopted data from 12 countries, its size cannot guarantee an effective applincation in any primary care setting. In addition, the ACCEPT 2.0 was recalibrated using patients' information refferring to periods preceding 2008,[10] so exclung the more recent treatments and managements to COPD sufferers. For this reason, we attempted to obtain a country-specific tool using an Italian data source, representative and updated to 2022, which could be translated in to the same general practice setting.[37] By doing so, we were able to further exploit the availble EMRs for several concurrent conditions, which might be subsequently adopted for the implementation of a CDSS intended to GPs. The fact that we reported an AUC of 0.66 lower the dicrimination accuracy of the ACCEPT 2.0 (AUC: 0.76) is likely due to the greater heteorgeneity of our cohort, which has not been longitudinally created for research purposes. However, we obtained good calibration measures for a 6-month event horizon, so suggesting an acceptable value for AUC given the greater heterogeneity of our population-based setting.[38]

These findings are consistent with the broader trend in COPD research,[8,11,39,40] which seeks to develop personalized approaches to disease management. We found associations with comorbidities such as osteoarthritis, GERD, and asthma highlight the need for comprehensive patient assessments that consider both respiratory and non-respiratory conditions. This is particularly relevant in the context of polypharmacy, where careful medication management becomes critical to minimize potential interactions and side effects. The inclusion of concurrent pharmacotherapies sheds light on the potential role of specific medications in exacerbation risk: they could be the proxy of patient's severity, which is not easily identified because of the large number of missing data for spirometry, and reduced adherence to medications. Clinicians should

therefore weight the benefits and risks of medication regimens, taking into account their individual patients' profiles. As stated above, the strong association between a history of prior COPD exacerbations and future event risk underscores the importance of effective management and prevention strategies for this patients' subgroup.[8,9]

The use of CDSS based on CEX-HScore might aid the Italian GPs to be compliant with the GOLD[11] and NOTA 99[12] indications in several ways. First, the identification of COPD sufferers for whom the medications adherence is pivotal to reach the therapeutic target. Such an example, those reporting a >=19% (high) risk of COPD-related exacerbations might be periodically monitored to ensure pantient's compliance to pharmacotherapy and/or revise prescribing appropriateness. Second, the presence of an intermediate/high risk of exacerbations would indicate a positive history for this same conditions which could be inderectly identified via antibiotics/steroids administration. As such, the recreation of patient's profile should support GPs to revise the respiratory medications currently used as per the GOLD guidelines[11] and NOTA 99 indications.[12] In this respect, the use of ICS-containing medications is strictly related to the presence, number and severity of COPD exacerbations. Third, a better knowledge of the risk profile for COPD-related exacerbations might support GPs to decide for prescription for specailist referral.

This study has limitations. First, as in other primary care data souces, severe exacerbations might be underregistered because of fatal cases or misreported hospitalizations in those patients mainly in charge of specialists and/or older instituzionalized adults. Nevertheless, the predicted risk was based on good calibration,[38] and the moderate exacerbations were those which could be more frequently captured in this setting. Secondly, unlike ACCEPT 2.0[8] we did not test the CEX-HScore in COPD sufferers with no prior exacerbation. However, given the reliable calibration and the poor coding for this event (i.e.: exacerbations are inderectly captured via antibiotic/steorids prescriptions), the absence of the related beta coefficient might lead to misprediction in a real-word setting. Finally, the CEX-HScore was not tested in an external (indipendent) data source. Nevertheless, when these tools are developed using a representative data source in an attempt to apply this score in that same setting, the internal validity is sufficient.[37] Indeed, even though the CEX-HScore might be adopted by similar settings (e.g. countries with similar primary care organizations), its re-calibration (i.e. external validation) would be necessary to demonstrate the prediction accuracy in this different population.

Conclusion

We developed the CEX-HScore as a reliable tool in predicting COPD exacerbation in primary care. Even thought further validation and refinement may be necessary, this score has the potential to enhance the appropriateness of COPD care by facilitating personalized approaches using a score-based CDSS for Italian GPs.

References

1. Edwards, S. C. *et al.* The burden of chronic obstructive pulmonary disease associated with maintenance monotherapy in the UK. *Int. J. Chron. Obstruct. Pulmon. Dis.* **11**, 2851–2858 (2016).
2. Miravitlles, M. *et al.* A review of national guidelines for management of COPD in Europe. *Eur. Respir. J.* **47**, 625–37 (2016).
3. Belleudi, V. *et al.* The impact of adherence to inhaled drugs on 5-year survival in COPD patients: a time dependent approach. *Pharmacoepidemiol. Drug Saf.* **25**, 1295–1304 (2016).
4. Vetrano, D. L. *et al.* Triple inhaled therapy in COPD patients: determinants of prescription in primary care. *Respir. Med.* **154**, (2019).
5. Vetrano, D. L. *et al.* Poor adherence to chronic obstructive pulmonary disease medications in primary care: Role of age, disease burden and polypharmacy. *Geriatr. Gerontol. Int.* (2017). doi:10.1111/ggi.13115
6. Lapi, F., Kezouh, A., Suissa, S. & Ernst, P. The use of inhaled corticosteroids and the risk of adrenal insufficiency. *Eur. Respir. J.* **42**, (2013).
7. Suissa, S., Patenaude, V., Lapi, F. & Ernst, P. Inhaled corticosteroids in COPD and the risk of serious pneumonia. *Thorax* **68**, (2013).
8. Safari, A. *et al.* ACCEPT 2·0: Recalibrating and externally validating the Acute COPD exacerbation prediction tool (ACCEPT). *EClinicalMedicine* **51**, 101574 (2022).
9. Adibi, A. *et al.* The Acute COPD Exacerbation Prediction Tool (ACCEPT): a modelling study. *Lancet Respir. Med.* **8**, 1013–1021 (2020).
10. Vestbo, J. *et al.* Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points (ECLIPSE). *Eur. Respir. J.* **31**, 869–873 (2008).
11. 2023 GOLD Report - Global Initiative for Chronic Obstructive Lung Disease - GOLD. Available at: https://goldcopd.org/2023-gold-report-2/. (Accessed: 9th September 2023)
12. Https://www.aifa.gov.it/nota-99. AIFA - NOTA 99.
13. Wolff, R. *et al.* PROBAST – A risk-of-bias tool for prediction-modelling studies. *The Global Evidence Summit* (2017).
14. Moons, K. G. M. *et al.* Transparent Reporting of a multivariable prediction model for Individual Prognosis or Disagnosis (TRIPOD): Explanantion and Elaboration. *Ann. Intern. Med.* **162**, W1-73 (2015).
15. Bianchini, E. *et al. VIII Report Health Search. Anno 2013/2014*. (2014).
16. Coloma, P. M. *et al.* Identification of acute myocardial infarction from electronic healthcare records using different disease coding systems: a validation study in three European countries. *BMJ Open* **3**, (2013).
17. Mazzaglia, G., Sessa, E. & Samani, F. Use of Computerized General Practice Database for epidemiological studies in Italy: a comparative study with the official national statistics. *J Epidemiol Community Heal.* **58**, (2004).
18. Sterrantino, C. *et al.* Burden of community-acquired pneumonia in Italian general practice. *Eur. Respir. J.* **42**, 1739–42 (2013).
19. Dentali, F. *et al.* Derivation and Validation of a Prediction Model for Venous Thromboembolism in Primary Care. *Thromb. Haemost.* **120**, 692–701 (2020).
20. Lapi, F. *et al.* Predicting the risk of severe COVID-19 outcomes in primary care: development and validation of a vulnerability index for equitable allocation of effective vaccines. *Expert Rev. Vaccines* 1–8 (2021). doi:10.1080/14760584.2022.2019582

21. Vetrano, D. L. *et al.* Triple inhaled therapy in COPD patients: determinants of prescription in primary care. *Respir. Med.* **154**, 12–17 (2019).

22. Kim, V. & Aaron, S. D. What is a COPD exacerbation? Current definitions, pitfalls, challenges and opportunities for improvement. *Eur. Respir. J.* **52**, (2018).

23. Kwak, M. S., Kim, E., Jang, E. J., Kim, H. J. & Lee, C. H. The efficacy and safety of triple inhaled treatment in patients with chronic obstructive pulmonary disease: A systematic review and meta-analysis using Bayesian methods. *Int. J. COPD* **10**, 2365–2376 (2015).

24. Backman, H. *et al.* Determinants of Severe Asthma - A Long-Term Cohort Study in Northern Sweden. *J. Asthma Allergy* **15**, 1429–1439 (2022).

25. Singh, N., Singh, P. N. & Hershman, J. M. Effect of calcium carbonate on the absorption of levothyroxine. *J. Am. Med. Assoc.* **283**, 2822–2825 (2000).

26. Vetrano, D. L. *et al.* Patterns of oral corticosteroids use in primary care patients with severe asthma. *Respir. Med.* **166**, (2020).

27. Van Calster, B. *et al.* A calibration hierarchy for risk models was defined: from utopia to empirical data. *J. Clin. Epidemiol.* **74**, 167–76 (2016).

28. Steyerberg, E. W. *et al.* Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* **21**, 128–38 (2010).

29. Collins, G. S., Reitsma, J. B., Altman, D. G. & Moons, K. G. M. Transparent reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *J. Clin. Epidemiol.* **68**, 134–43 (2015).

30. Smith, G. C. S., Seaman, S. R., Wood, A. M., Royston, P. & White, I. R. Correcting for optimistic prediction in small data sets. *Am. J. Epidemiol.* **180**, 318–324 (2014).

31. Steyerberg, E. W. *et al.* Assessing the performance of prediction models : A framework for some traditional and novel measures. *Epidemiology* **21**, 128–38 (2010).

32. Cleveland, W. S. & Devlin, S. J. Locally weighted regression: An approach to regression analysis by local fitting. *J. Am. Stat. Assoc.* **83**, 596–610 (1988).

33. A, S. *et al.* Comparisons of risk prediction methods using nested case-control data. *Stat. Med.* **36**, 455–465 (2017).

34. Royston, P. & Altman, D. G. External validation of a Cox prognostic model: principles and methods. *BMC Med. Res. Methodol. 2013 131* **13**, 1–15 (2013).

35. Chang, C., Hsieh, M. K., Chang, W. Y., Chiang, A. J. & Chen, J. Determining the optimal number and location of cutoff points with application to data of cervical cancer. *PLoS One* **12**, (2017).

36. Li, Y., Bondarenko, I., Elliott, M. R., Hofer, T. P. & Taylor, J. M. G. Using multiple imputation to classify potential outcomes subgroups. *Stat. Methods Med. Res.* **30**, 1428–1444 (2021).

37. Ramspek, C. L., Jager, K. J., Dekker, F. W., Zoccali, C. & van Diepen, M. External validation of prognostic models: what, why, how, when and where? *Clin. Kidney J.* **14**, 49–58 (2021).

38. de Hond, A. A. H., Steyerberg, E. W. & van Calster, B. Interpreting area under the receiver operating characteristic curve. *Lancet Digit. Heal.* **4**, e853–e855 (2022).

39. Rojas-Reyes, M. X. imen., García Morales, O. M., Dennis, R. J. & Karner, C. Combination inhaled steroid and long-acting beta2-agonist in addition to tiotropium versus tiotropium or combination alone for chronic obstructive pulmonary disease. *Cochrane database Syst. Rev.* CD008532 (2016). doi:10.1002/14651858.CD008532.PUB3

40. Ferreira, A. J., Reis, A., Marçal, N., Pinto, P. & Bárbara, C. COPD: A stepwise or a hit hard approach? *Rev. Port. Pneumol. (English Ed.* **22**, 214–221 (2016).

Table 1. Characteristics of development and validation sub-cohorts.

Determinants	Development sub-cohort N=42084		Validation sub-cohort N=21679	
	N (%)	Univariate HR (95% CI)	N (%)	Univariate HR (95% CI)
Age, mean (SD)	67.75 (11.69)	1.01 (1.01 - 1.01)	67.88 (11.72)	1.01 (1.01 - 1.01)
Gender (male)	24102 (57.27)	0.98 (0.94 - 1.01)	12667 (58.43)	1.02 (0.97 - 1.07)
Smoking (non-smokers)	8779 (20.86)		4540 (20.94)	
Current	10741 (25.52)	1.21 (1.16 - 1.27)	5467 (25.22)	1.19 (1.12 - 1.28)
Former	8999 (21.38)	1.08 (1.03 - 1.13)	4769 (22)	1.1 (1.03 - 1.17)
Missing	13565 (32.23)	1.11 (1.05 - 1.17)	6903 (31.84)	1.1 (1.03 - 1.19)
Overweight or obesity (no)	8012 (19.04)		3960 (18.27)	
Yes	19375 (46.04)	1.03 (0.99 - 1.08)	10185 (46.98)	1 (0.94 - 1.06)
Missing	14697 (34.92)	1.02 (0.97 - 1.08)	7534 (34.75)	1.01 (0.94 - 1.08)
Co-morbidities				
Depression and anxiety	12584 (29.9)	0.98 (0.95 - 1.01)	6503 (30)	0.97 (0.92 - 1.02)
Cancer	14613 (34.72)	1 (0.97 - 1.04)	7689 (35.47)	0.98 (0.94 - 1.03)
Osteoarthritis	17555 (41.71)	1.03 (1 - 1.07)	9065 (41.81)	1.05 (1.01 - 1.1)
Gastro-esophageal reflux disease	11553 (27.45)	1.07 (1.03 - 1.11)	5939 (27.4)	1.08 (1.03 - 1.13)
Cardio and cerebrovascular diseases	11126 (26.44)	0.99 (0.96 - 1.03)	5891 (27.17)	0.96 (0.91 - 1)
Atrial fibrillation	3800 (9.03)	0.97 (0.92 - 1.03)	2035 (9.39)	0.98 (0.91 - 1.06)
Heart failure	3281 (7.8)	1.02 (0.96 - 1.08)	1755 (8.1)	0.95 (0.87 - 1.03)
Diabetes mellitus	9393 (22.32)	0.78 (0.75 - 0.81)	4969 (22.92)	0.81 (0.76 - 0.85)
Asthma	4949 (11.76)	1.25 (1.2 - 1.31)	2485 (11.46)	1.2 (1.13 - 1.28)
Polypharmacy	32727 (77.77)	1.48 (1.41 - 1.55)	16867 (77.8)	1.41 (1.32 - 1.5)
Prior moderate exacerbations	17480 (41.54)	1.95 (1.89 - 2.02)	9067 (41.82)	1.86 (1.77 - 1.94)
Prior severe exacerbations	318 (0.76)	1.08 (0.91 - 1.28)	179 (0.83)	1.22 (0.99 - 1.5)
COPD baseline pharmacotherapy*				
SABA	10266 (24.39)	1.07 (1.03 - 1.11)	5127 (23.65)	1.08 (1.02 - 1.13)
LABA	8246 (19.59)	1.05 (1.01 - 1.09)	4169 (19.23)	1.03 (0.97 - 1.08)
LAMA	16543 (39.31)	1.03 (0.99 - 1.06)	8535 (39.37)	1.08 (1.03 - 1.13)

ICS	24545 (58.32)	1.06 (1.03 - 1.1)	12593 (58.09)	1.07 (1.02 - 1.13)
Fixed LABA/ICS combinations	19209 (45.64)	1.15 (1.11 - 1.18)	10015 (46.2)	1.22 (1.16 - 1.27)
Fixed LABA/LAMA combinations	250 (0.59)	1.05 (0.84 - 1.31)	134 (0.62)	1.23 (0.92 - 1.66)
Fixed LABA/LAMA/ICS combinations	114 (0.27)	1.2 (0.84 - 1.72)	70 (0.32)	0.65 (0.36 - 1.17)

**ever before ths date of COPD diagnosis (included)*

Table 2. Beta coefficients being estimated in the multivariate model for each candidate determinants.

Determinants	**Beta coeff. (95% CI)**
Age mean (SD)	0.009 (0.007 - 0.01)
Gender (male)	-0.025 (-0.058 - 0.008)
Smoking habits (non smoker)	
Current	0.195 (0.148 - 0.241)
Former	0.075 (0.027 - 0.123)
Missing	0.102 (0.049 - 0.155)
Overweight or obesity (no)	
Yes	0.032 (-0.01 - 0.074)
Missing	0.023 (-0.03 - 0.076)
Comorbidities	
Depression and anxiety	-0.021 (-0.055 - 0.013)
Cancer	0.005 (-0.027 - 0.037)
Osteoarthritis	0.031 (-0.001 - 0.064)
Gastro-esophageal reflux disease	0.068 (0.034 - 0.101)
Cardio and cerebrovascular diseases	-0.009 (-0.045 - 0.027)
Atrial fibrillation	-0.027 (-0.083 - 0.028)
Heart failure	0.016 (-0.045 - 0.077)
Diabetes mellitus	-0.253 (-0.292 - -0.214)
Asthma	0.222 (0.178 - 0.267)
Polypharmacy	0.39 (0.344 - 0.436)
Prior moderate exacerbations	0.669 (0.636 - 0.702)
Prior severe exacerbations	0.077 (-0.09 - 0.244)
COPD baseline pharmacotherapies	
SABA	0.071 (0.034 - 0.108)
LABA	0.046 (0.008 - 0.084)
LAMA	0.028 (-0.005 - 0.061)
ICS	0.062 (0.026 - 0.098)
Fixed LABA/ICS combinations	0.136 (0.102 - 0.169)
Fixed LABA/LAMA combinations	0.047 (-0.174 - 0.267)
Fixed LABA/LAMA/ICS combinations	0.181 (-0.178 - 0.54)

Figure 1. Calibration plot on expected (predicted) and observed risks of COPD moderate/severe exacerbation: risk estimation for 6-month follow-up.

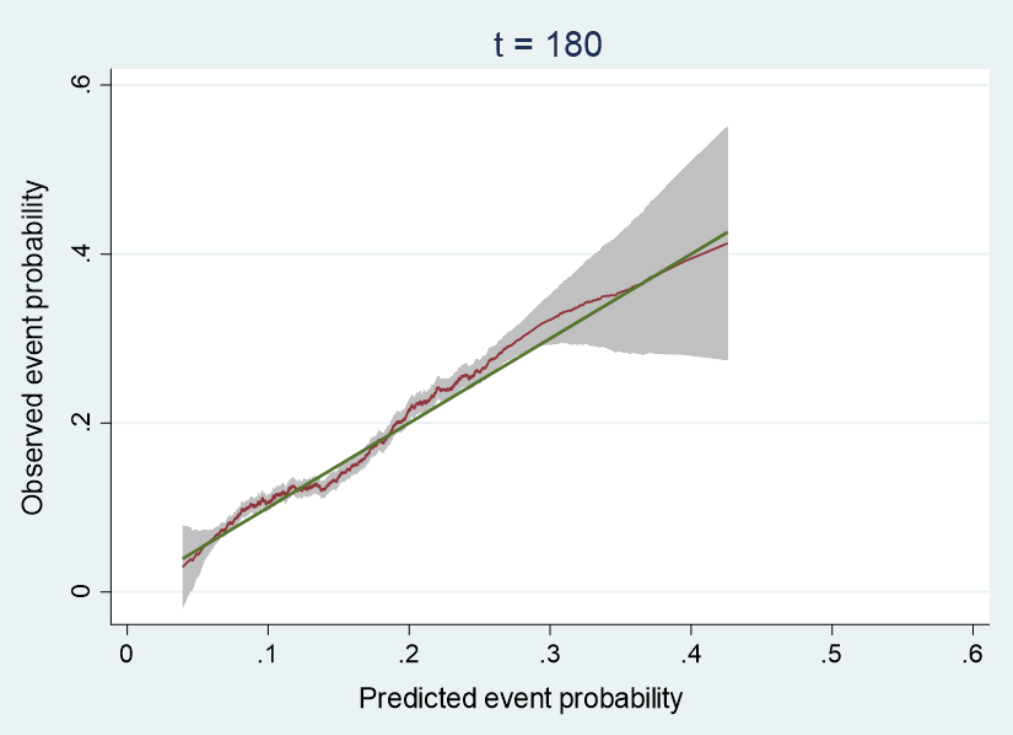

Figure 2. Calibration plot on expected (predicted) and observed risks of COPD moderate/severe exacerbation limited to a risk range between 0 to 40%.

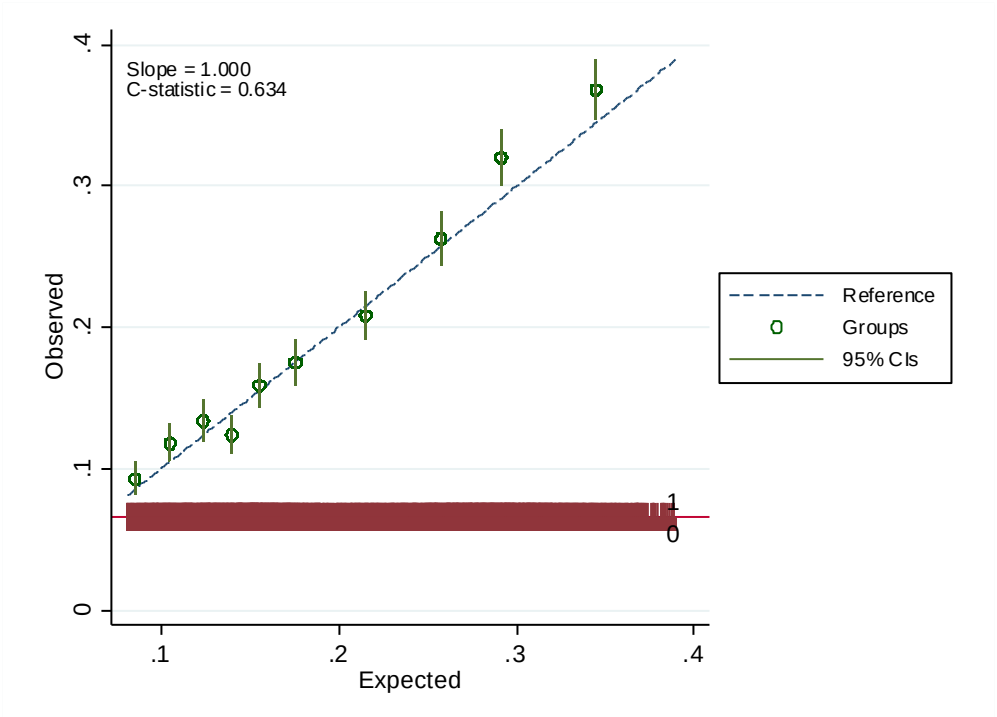

Table 3. Measures of prediction performances for the CEX-HScore in the validation sub-cohort and after bootstrapping.

	Validation sub-cohort (N=21679)	***Bootstrap samples (n=200)***
Event horizon: 1-year follow-up		
Explained variation		
Pseudo-R^2	0.098 (0.090 - 0.108)	0.108 (0.101 - 0.114)
Calibration		
AUC	0.656 (0.647- 0.666)	0.611 (0.606 - 0.616)
Calibration		
Slope	1.027 (0.945 - 1.109)	1.002 (0.970 - 1.029)
P-value	0.5138	

Supplementary Figure 1. Calibration plot on expected (predicted) and observed risks of COPD moderate/severe exacerbation: risk estimation for 1-year follow-up.