**PROBAST**

Study:

A machine learning model for predicting acute exacerbation of in-home chronic obstructive pulmonary disease patients

Step 2: Type of prediction study

**Is the study a diagnostic or a prognostic study?**

Prognostic

**Is the study a development only, development and validation or validation only study?**

Development only

**What is the model of interest?**

Catboost

**What is the outcome of interest?**

AECOPD

Step 3: Assess risk of bias

**Domain 1: Participants**

**Describe the sources of data and criteria for participant selection**

In this study, we enrolled patients with COPD classified as type C or type D based on the GOLD guideline [1]. Patients meeting the following criteria were excluded from participation: (a) currently experiencing an AECOPD phase (as per GOLD 2018′s definition: acute deterioration of respiratory symptoms necessitating additional treatment) at the time of recruitment, (b) exhibiting other unstable or uncontrollable systemic diseases, (c) having an active, known, or suspected autoimmune disease, (d) being diagnosed with malignant tumors, (e) diagnosed with mental illness, (f) intending to relocate to another country in the near future, and (g) those whose participation might endanger themselves, disrupt study evaluation, or influence other participants due to cognitive impairment or other factors

**1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?**

Y

**1.2 Were all inclusions and exclusions of participants appropriate?**

PY

**Risk of bias introduced by selection of participants:**

Low

**Rationale of bias rating**

Reasonable eligibility criteria. Small risk of bias due to criterion a)

**Domain 2: Predictors**

**List and describe predictors included in the final model, e.g. definition and timing of assessment**

Each enrolled subject is provided with a portable spirometer and an electronic stethoscope, as demonstrated on the left side of Fig. 2. The portable spirometer is utilized to gather daily data on seven lung function measurements, which include Forced Vital Capacity (FVC), Forced Expiratory Volume in the first second (FEV1), Peak Expiratory Flow (PEF), Forced Expiratory Flow (FEF), Maximum Expiratory Flow Rate at 25 %, 50 %, or 75 % of pulmonary capacity. To ensure the accuracy of collected measurements, participants are requested to perform each measurement three times.

**2.1 Were predictors defined and assessed in a similar way for all participants?**

Y

**2.2 Were predictor assessments made without knowledge of outcome data?**

Y

**2.3 Are all predictors available at the time the model intended to be used?**

Y

**Risk of bias introduced by predictors or their assessment**

Low

**Rationale of bias rating**

Predictors are independent and applicable and assessed similarly.

**Domain 3: Outcome**

**Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination:**

In this study, our objective is to predict the onset of AECOPD for homebound COPD patients through remote monitoring. The input data comprises pulmonary functionality measurements obtained using patient-operable devices. The target is to determine whether the COPD status will exacerbate within the next k days (k = 1, 2, 3). Formally, considering the current day as T-1, each task k (where k = 1, 2, 3) involves binary prediction, aiming to forecast whether AECOPD will manifest by the end of day T + k-1. This prediction is based on the observational data collected from day T-5 to day T-1. Now, let’s formally establish the definitions of positive and negative samples in this study. An instance of AECOPD onset is categorized as a positive sample. For task k (where k = 1, 2, 3), the day of AECOPD occurrence is denoted as T + k-1. The spirometer and stethoscope features collected from day T-5 to T-1 are employed as input features, with the corresponding target label set to True. To train the machine learning models, it’s essential to include negative samples. For negative sampling, we identify 8-day time windows (from T′-5 to T′+2) that are not within a 30-day range from any AECOPD onset phase. Subsequently, for each task k (where k =1, 2, 3), a negative sample encompasses features from T′-5 to T′-1 as inputs, while the target label is set to False

**3.1 Was the outcome determined appropriately?**

Y

**3.2 Was a pre-specified or standard outcome definition used?**

Y

**3.3 Were predictors excluded from the outcome definition?**

Y

**3.4 Was the outcome defined and determined in a similar way for all participants?**

Y

**3.5 Was the outcome determined without knowledge of predictor information?**

Y

**3.6 Was the time interval between predictor assessment and outcome determination appropriate?**

Y

**Risk of bias introduced by the outcome or its determination**

Low

**Rationale of bias rating**

Forecasting of AECOPD based on data from prior days.

**Domain 4: Analysis**

**Describe number of participants, number of candidate predictors, outcome events and events per candidate predictor**

e observed 117 instances of AECOPD onset among the 66 patients. These 117 AECOPD events constitute the positive samples. By excluding the one-month period before and after the AECOPD onsets, we collected 468 negative samples from time windows without AECOPD events. In total, our dataset comprises 585 samples (117 positive and 468 negative), which can be categorized as a moderately sized dataset.

**Describe how the model was developed, predictor selection and risk group definition**

CatBoost.

**Describe whether and how the model was validated, either internally (cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants)**

To evaluate the models, a 5-fold cross-validation approach was adopted. On each fold, 20 % of the positive group and an equal proportion from the negative group were randomly selected as the test data, while the remaining samples constituted the training dataset.

**Describe the performance measures of the model, e.g. calibration, discrimination, classification, net benefit, and whether they were adjusted for optimism**

AUC, SEN, SPE, F1

**Describe any participants who were excluded from the analysis**

As illustrated in Fig. 5, the study encompasses a total of 250 enrolled patients, with 66 of them successfully completing the four follow-ups over a span of six months. Among these 66 patients, 32 are classified under group C, while 34 are categorized within group D

**Describe missing data on predictors and outcomes as well as methods used for missing data**

Data points with an absolute value exceeding 4 are treated as anomalies and omitted. Raw features with over 20 % missing values are dropped. For tree-based models, we retain missing values as-is, while for the Logistic model, we impute missing values with the mean value.

**4.1 Were there a reasonable number of participants with the outcome?**

PY

**4.2 Were continuous and categorical predictors handled appropriately?**

Y

**4.3 Were all enrolled participants included in the analysis?**

N

**4.4 Were participants with missing data handled appropriately?**

Y

**4.5 Was selection of predictors based on univariable analysis avoided?**

Y

**4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls)**

**accounted for appropriately?**

Y

**4.7 Were relevant model performance measures evaluated appropriately?**

Y

**4.8 Were model overfitting and optimism in model performance accounted for?**

Y

**4.9 Do predictors and their assigned weights in the final model correspond to the results**

**from multivariable analysis?**

Y

**Risk of bias introduced by the analysis**

Low

**Rationale of bias rating**

Amount of outcomes likely enough for generalizing results somehow. A lot of patients were excluded due to not enough follow-ups, however this criterion was pre-defined.

**Overall Risk of bias**

Low