**PROBAST**

Study:

Home Monitoring of Asthma Exacerbations in Children and Adults With Use of an AI-Aided Stethoscope.

Step 2: Type of prediction study

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

Diagnostic

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

Development only

**What is the model of interest?**

Random forest

**What is the outcome of interest?**

Asthma exacerbations

Step 3: Assess risk of bias

**Domain 1: Participants**

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

Patients were recruited from the general Slavic population via advertisements and recruitment efforts conducted by medical centers. The inclusion criterion for children and adults was diagnosed asthma (primary or secondary care). In younger children, suspicion of asthma was also accepted. Any other comorbidities that might affect asthma assessment or influence the measured parameters constituted exclusion criteria.

**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?**

Y

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

Low

**Rationale of bias rating**

Reasonable eligibility criteria

**Domain 2: Predictors**

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

Pulse oximeters and peak flow meters (Mini Wright Peak Flow and Low Range [Clement Clarke International]), to collect SpO2 and PEF measurements. Other parameters were measured with the European Conformity (CE)-certified StethoMe stethoscope, which records auscultatory sounds from standard chest points (Figure 1) and transfers sound files wirelessly to a dedicated mobile phone application. The recordings were automatically analyzed by the AI module, and results (pathologic auscultatory sound intensities, HR, RR, I/E) were displayed in the application. In addition, the user provided other health state information in a survey in the mobile telephone application. We used StethoMe AI for automatic analysis of auscultatory recordings and aggregation of results for each examination

**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 assessed the same way.

**Domain 3: Outcome**

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

Data were analyzed by physicians who assessed the asthma exacerbation level (none, mild, moderate, severe) for each examination. The assessment was carried out at the examination level rather than the recording level. A total of 17 physicians with different specializations were involved in this process (2 internal medicine specialists, 4 pulmonologists, 9 pediatricians, 5 allergologists, 4 family medicine specialists [some had a double specialization]), each assessing examinations performed by participants assigned exclusively to them. Each physician took into account all of the information generated by the participant (Table 1), listened to each auscultation recording, and analyzed the spectrogram

**3.1 Was the outcome determined appropriately?**

PN

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

N

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

N

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

N

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

N

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

Y

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

High

**Rationale of bias rating**

Each physician was assigned specific participants, so different participants were evaluated by different physicians, rather than a consistent assessor across all cases. Physicians had access to all data generated by the participants, including predictors, during outcome assessment. Definition of mild and moderate and severe asthma unclear.

**Domain 4: Analysis**

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

A total of 149 patients participated in the study. Among the 6,442 complete examinations, which included a total of 41,872 recordings,

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

For this purpose, 12 feature sets were defined, each including patient age, gender, an identifier physician assessing the exacerbation level, and ≥1 user-registered parameters from Table 1. For each evaluated set, a random forest regressor (RFR) was fitted to the reference standard exacerbation levels mapped to a scale from 0 to 1.0 (none: 0; mild: 1/3; moderate: 2/3; severe: 1). Each RFR consisted of 100 decision trees trained using the squared error criterion and without constraining their maxi-mum depth. For quantitative evaluation of the performance of the RFR, we rescaled the model’s prediction labels from the 4-point scale to binary values by thresholding at 0.5 (values >0.5 were mapped to 1, and values <0.5 were mapped to 0) so as to predict the existence of at least moderate exacerbation level, as identified by the physician’s labels.

**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)**

Each feature set was evaluated in a 10-fold cross-validation experiment, and each experiment was repeated 20 times to estimate the feature set’s mean AUC and 95% CIs. In each fold of the cross-validation process, the data set was categorized into a training set comprising 90% of patients

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

AUC

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

None

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

6.4% did not meet quality criteria and were not analyzed further; exclusion was determined by StethoMe AI on the basis of inadequate quality of the majority of recordings for an examination.

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

Y

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

PN

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

Y

**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?**

U

**Risk of bias introduced by the analysis**

Low

**Rationale of bias rating**

There is some small risk as the outcome was treated as binary outcome although labeling contained four ordinal classes. Yet due to enough data and proper handling of validation and missing data, analysis domain may still be at low risk overall.

**Overall Risk of bias**

High