

1 Retrospective development and evaluation of 2 prognostic models for exacerbation event prediction 3 in patients with Chronic Obstructive Pulmonary 4 Disease using data self-reported to a digital health 5 application

6 F. P. Chmiel^{1,*}, D. K. Burns¹, J. B. Pickering¹, Alison Blythin², T. M. A. Wilkinson^{2,3,4,†}, and
7 M. J. Boniface^{1,†}

8 ¹School of Electronics and Computer Science, University of Southampton, UK

9 ²my mhealth Limited, Bournemouth, UK

10 ³NIHR ARC Wessex, University of Southampton, Southampton, UK

11 ⁴University of Southampton Faculty of Medicine, Southampton, UK

12 *F.P.Chmiel@soton.ac.uk

13 †These authors contributed equally to this work.

14 **Background** Self-reporting digital applications provide a way of remotely monitoring and managing patients with chronic
15 conditions in the community. Leveraging the data collected by these applications in prognostic models could provide increased
16 personalisation of care and reduce the burden of care for people who live with chronic conditions. This study evaluated
17 the predictive ability of prognostic models for prediction of acute exacerbation events in people with Chronic Obstructive
18 Pulmonary Disease using data self-reported to a digital health application.

19 **Methods** Retrospective study evaluating the use of symptom and Chronic Obstructive Pulmonary Disease assessment test data
20 self-reported to a digital health application (myCOPD) in predicting acute exacerbation events. We include data from 2,374
21 patients who made a total of 68,139 self-reports. We evaluated the degree to which the different variables self-reported to the
22 application are predictive of exacerbation events and developed both heuristic and machine-learned models to predict whether the
23 patient will report an exacerbation event within three days of self-reporting to the application. The model's predictive ability
24 was evaluated on self-reports from an independent set of patients.

25 **Findings** Users self-reported symptoms and standard Chronic Obstructive Pulmonary Disease assessment tests display
26 correlation with future exacerbation events. Both a baseline model (AUROC 0.655 (95 % CI: 0.689-0.676)) and a machine-
27 learnt model (AUROC 0.727 (95 % CI: 0.720-0.735)) showed moderate ability in predicting exacerbation events occurring
28 within three days of a given self-report. While the baseline model obtained a fixed sensitivity and specificity of 0.551 (95 % CI:
29 0.508-0.596) and 0.759 (95 % CI: 0.752-0.767) respectively, the sensitivity and specificity of the machine-learned model can be
30 tuned by dichotomizing the continuous predictions it provides with different thresholds.

31 **Interpretation** Data self-reported to healthcare applications designed to remotely monitor patients with Chronic Obstructive
32 Pulmonary Disease can be used to predict acute exacerbation events with moderate performance. This could increase
33 personalisation of care by allowing pre-emptive action to be taken to mitigate the risk of future exacerbation events. It is
34 plausible future studies could improve the accuracy of these models by either the inclusion of symptom information recorded
35 with greater granularity or including variables not considered in our study, for example vital signs, information on activity, local
36 environmental data, and lifestyle information.

37 **Funding** This project was funded by the European Union's Horizon 2020 research and innovation programme under grant
38 agreement No 780495 BigMedilytics.

39 **Research in context**

40 **Evidence before this study** Avoiding exacerbations is desirable for COPD patients because it likely leads to both a higher
41 quality of and longer life. Several studies have attempted to risk-stratify patients according to their exacerbation risk. Generally,
42 these prognostic models predict a patient's exacerbation frequency to guide long-term care, as opposed to predicting acute
43 events to provide advanced warning of and the opportunity to mitigate the risk of symptom exacerbation. Prediction of patients
44 exacerbation frequency is important since it allows more informed decisions about care, for example in the decision to prescribe
45 pharmacological interventions (such as inhaled corticosteroids or phosphodiesterase-4 inhibitors) which reduce exacerbation
46 frequency but are strongly associated with harm. However, the accurate prediction of acute events is also of significant utility
47 since it would enable pre-emptive treatment strategies (as opposed to reactive care) which could prevent exacerbations and
48 reduce hospitalisation.

49 **Added value of this study** We demonstrated that data self-reported to a digital health application can be used to predict
50 short-term future exacerbation events. This outcome is distinct to the majority of studies, since we demonstrate it is likely acute
51 exacerbation events can be predicted several days in advance using data recorded by a digital health application as opposed to
52 simply the predicting a patient's yearly exacerbation frequency.

53 **Implications of all the available evidence** Our study could allow the implementation of new exacerbation mitigation strategies
54 aimed to avoid exacerbation events in the short-term. Increased level of care personalization will be facilitated by the integration
55 of prognostic models into clinically validated digital applications for managing COPD patients. Increasing the number of
56 variables available to these models would likely improve their discriminative ability further.

57 Introduction

58 Chronic Obstructive Pulmonary Disease (COPD) is a collection of progressive lung diseases, characterised by breathing
59 difficulties and an irreversible reduction of lung function. It is one of the most prevalent chronic conditions in the world
60 (in England 2.19 % of the population are expected to have a confirmed COPD diagnosis by 2030³) and, the absence of a
61 cure, means it represents a significant burden for patients who have to manage the condition on a daily basis^{1,2}. A key
62 characteristic of managing COPD is in mitigating the risk of ‘exacerbation events’, which can be defined as an acute, sustained
63 worsening of a patient’s condition that necessitates a change in medication⁴. Exacerbations accelerate lung function decline and
64 evidence suggests that the frequency of exacerbations increases with decreasing lung function^{5–7}. Minimizing the number of
65 exacerbation events can therefore have a significant impact on the prognosis for COPD patients. Currently, several methods exist
66 to help control exacerbation events, including pharmacological interventions, pulmonary rehabilitation, and self-management
67 programmes⁸. There is also an identified clinical need to predict exacerbation events in advance to personalize COPD treatment
68 and offer the opportunity to provide targeted pre-emptive interventions^{9,10}.

69 In recent years, the advent of mobile health applications has facilitated increased remote management and care of patients with
70 COPD^{11,12}. These applications allow the recording of temporally dense information about a patient’s condition which allow
71 (near) real-time monitoring of a patient’s symptoms, providing clinicians with a source of data to help them understand how the
72 patient is managing their condition and gain an insight into the patient’s exacerbation frequency and severity. For the patient,
73 digital health applications provide both an access point for educational content about their condition and the opportunity to
74 improve their self-care, leading to better long-term management¹³. In the context of COPD, there is an opportunity to increase
75 the efficacy of digital health applications further, by leveraging the data they collect to predict acute exacerbation events and
76 provide personalized alerts to the patient. These alerts could facilitate the following of a clinically validated, personalized
77 intervention programme in an attempt to mitigate the occurrence of an acute exacerbation event.

78 In this report we present a retrospective study making use of data collected by the myCOPD mobile application, an NHS
79 approved, clinically validated application for persons with a diagnosis of COPD¹³. The application assists with the management
80 of COPD by providing educational content alongside a digital, momentarily assessed symptom diary. Using the application,
81 users self-report on the COPD-related symptoms they are currently experiencing, as well as information which characterises
82 their long-term COPD status (i.e., the COPD Assessment Test). Using statistical analysis and machine-learning methods we
83 evaluate the effectiveness of exploiting this simple, self-reported information to predict exacerbation events in the near-future
84 and discuss how such predictions could be used to improve the long-term outcomes for people with COPD.

85 Methods

86 Dataset description

87 We used an anonymised extract of daily user self-reports submitted to the myCOPD application between 01/01/2017 and
88 31/12/2019 (inclusive). A single report features a self-assessed symptom score (Figure 1) which is a four-point scale ranging
89 from normal symptoms to a severe deterioration of symptoms requiring hospitalisation (we encode this as an ordinal variable
90 between 1 and 4 indicating increasing severity of COPD related symptoms, see Figure 1 b). Users also perform a COPD
91 Assessment Test (CAT) at regular (approximately monthly) intervals. The CAT is an eight question assessment and yields
92 a ‘score’ between 0 and 40, where higher values indicate a more severe impact of COPD on a user’s overall health¹⁴. It is
93 validated and is an accepted way of quantifying the burden of COPD on someone’s life^{15,16}. In addition to these scores, our
94 dataset also features additional demographic and lifestyle information self-reported to the application. These include patient
95 age, gender, current smoking status, and the number of years they have been smoking for.

96 Defining exacerbation events

97 We use a symptom-based definition for exacerbation events where an event is marked to have occurred if a patient self-reports
98 either the use of medication to control their symptoms or hospitalization resulting from their condition, corresponding to a
99 self-reported symptom score of 3 or 4 respectively (Figure 1)¹⁷.

100 Cohort selection and dataset segregation

101 In total 5,170 users were included in the extract who reported a total number of 94,882 reports in the study period. User
102 registration was incremental (i.e., not all users registered at the same time) throughout the period of the study and self-reports
103 were not necessarily submitted every day. To create our study cohort we followed a selection process outlined in Figure 2.
104 First, isolated self-reports (those in which a second report was not made within 3 days) were removed because the target
105 variable could not be reliably calculated (see ‘Predicting exacerbation events’). Next, reports from anomalous users (those
106 only reporting exacerbation events or entering self-reports before their registration date) were removed. After removal of these
107 reports we obtained our final study cohort, featuring 68,139 reports from 2,374 unique users.

Predicting exacerbation events

For each daily self-report, we created a binary variable which indicated whether a report is followed by an exacerbation event in the following three days. A three day window was chosen to be close enough to the future exacerbation event for any signal to be present in the data but sufficiently far from the event such that a range of pre-emptive actions could be available to patients. For the development of the prognostic models, we selected only reports in which the patient did not report an exacerbation event on the same day (N=49,122, Figure 2). We then randomly assigned all reports from 20 % of patients to a hold-out test set and the remaining to a training set. We created a baseline, heuristic model which uses a users most recently reported symptom score, assigning users a either a low risk (1.7 %) or heightened risk (7.2 %) of future exacerbation, where percentages (brackets) correspond to the mean three-day exacerbation rate for all reports in the training set with symptom scores of one or two respectively.

Our supervised machine-learning models made use of patient demographics, lifestyle information, self-reported information, and aggregate features which summarizes a patients (recent) self-reporting history. A full schema of variables used by our models is presented in Supplementary Table 1. For our predictive models we used logistic regression and a Random Forest classifier each trained by 5-fold, grouped cross-validation at the user level, i.e., reports from a single user appear exclusively in either the training or validation fold. Missing CAT scores were forward-filled imputed at the user level where possible. All other missing values were filled using mean imputation within fold. Either target or ordinal encoding was used for all categorical variables (Supplementary Table 1). Model hyperparameters were optimized on the out-of-fold validation samples using Bayesian optimization via the Tree Parzen Estimator algorithm as implemented in the hyperopt Python library^{18,19}. Model performance was evaluated on the hold-out test set and 95 % confidence intervals were estimated using bootstrapping. To create a binary decision of exacerbation risk, model predictions were dichotimized with thresholds chosen to yield either a fixed specificity or the maximum Youden's J statistic on the test set²⁰.

Results

Our cohort self-reporting to the application featured 2,374 users in total, of which most (70.4 %) were between the ages of 60 and 79, inclusive (Table 1). Only 27.3 % of users reported their gender, with 419 males reporting compared to 231 females. A large fraction (49.7 %) of users reported their smoking history, with 86.5 % of those reporting being either a current or ex-smoker (Table 1). Out of the self-reports included in this study patients reported 5,906 self-reports which correspond to an exacerbation event in the context of this work, 8.7 % of total reports (Figure 2).

In Figure 3 we investigate the two metrics self-reported to the myCOPD application and their relation to self-reported exacerbation events. Panels a to d of Figure 3 display the correspondence between between symptom scores and the CAT results when self-reported on the same day. While for each symptom score users nearly report the full range of CAT scores, there is a clear correlation between CAT scores and symptom scores, with users reporting higher symptom scores more likely to also report a higher CAT score. For example, the mean CAT score reported when a user reports a symptom score of one is 13.5, which is significantly lower ($p < 0.001$) than the mean CAT score (19.5) when a user reports a symptom score of two. Such a correlation is to be expected, research has shown increased CAT scores correlate with an increased exacerbation frequency²¹, which in turn would lead to higher symptom scores being reported in our study.

In Figure 3 e and f, we evaluate the sensitivity of symptom scores and CAT results to future exacerbation events. We display how the mean of the two reported variables change in days preceeding (and following) the day in which users self-report their first exacerbation event to the application (not necessarily their first ever exacerbation event). By inspection of Figure 3 e we see that the mean reported symptom score in proximity of the first reported exacerbation event increases, indicating that in the days preceding their first exacerbation event users are increasingly likely to report a mild deterioration of symptoms (symptom score of 2). Changes in the mean symptom score calculated across all users can be seen several days in advance, suggesting that at least some users are observing a mild deterioration in symptoms several days in advance of exacerbation events. For days subsequent to users first self-reported exacerbation (right of the dashed line in Figure 3 e) the mean symptom score initially exceeds 2, showing that self-reported exacerbation events can be multi-day events - consistent with current understanding of exacerbation events⁴. Similarly, in Figure 3 f the reported mean CAT result is observed to increase in magnitude in the days preceeding an exacerbation event and then decrease (at a slower rate) following a reported event. Overall, these trends indicate there is potential in using these self-reported variables to predict at least a subset of exacerbation events in advance.

Next we discuss the baseline prognostic model, the model captures the key feature about exacerbation events observed in our data: persons reporting a mild deterioration of symptoms are significantly ($p < 0.001$) more likely to experience an exacerbation event in the next three days compared to those reporting normal symptoms, with a relative risk of 4.16 (95 % CI: 3.8-4.5). On the hold-out test set the baseline model obtained an AUROC of 0.655 (95 % CI: 0.689-0.676). In Figure 4, we display the Receiver Operating Characteristic (ROC) curve for our baseline model alongside the two machine-learned models, a logistic

160 regression model (solid grey line) which does not consider variable interactions and a random forest classifier (solid black line)
161 which does. The logistic regression model obtained an AUROC of 0.697 (95 % CI: 0.689-0.711) and the Random Forest model
162 0.727 (95 % CI: 0.720-0.735), on the hold-out test (Table 2). The significantly higher ($p < 0.001$) performance of the Random
163 Forest model suggests either interactions between variables are important in discriminating between reports associated with
164 exacerbation within three days or non-linear relations are present.

165 In Table 2 we present the sensitivity and specificity of the baseline model and the machine-learnt models evaluated on the
166 hold-out test set. While the baseline model is already dichotomized, for the machine-learnt models a threshold must be chosen
167 to binarize the continuous exacerbation risks they produce. The baseline model obtained a sensitivity of 0.551 (95 % CI:
168 0.508-0.596) with specificity of 0.759 (95 % CI: 0.752-0.767). While neither machine-learnt model significantly outperform the
169 baseline model at the same specificity (e.g., compare model a and e in Table 2), the tuning of the threshold used to dichotomize
170 the machine-learnt models predictions can lead to a range of sensitivity and specificities (compare models c, d and e in Table 2)
171 on the hold-out test which could be tuned to match different escalation policies and interventional strategies. For example, the
172 random forest model can be tuned to yield a sensitivity of 0.921 (95 % CI: 0.907-0.935) or 0.576 (95 % CI: 0.553-0.594) with
173 respective specificities of 0.250 (95 % CI: 0.246-0.254) or 0.750 (95 % CI: 0.749-0.751).

174 Discussion

175 Symptom information self-reported to the myCOPD application displayed correlation to the start of future exacerbation
176 events (Figure 2 e) and we found machine-learnt models utilizing this and other sources of self-reported data were able to
177 identify patients at risk of exacerbation within three days with moderate discriminative ability (AUROC 0.727 (95 % CI:
178 0.720-0.735)).

179 Our analysis has shown strong evidence that users self-report changes in symptoms prior to exacerbation events which could
180 be leveraged to predict future events. Machine-learnt models are well positioned to take advantage of these trends because
181 they can learn complex, non-linear relations between variables which are inaccessible to hand crafted models and, despite
182 the complexity of the models, they can be easily integrated into digital applications. This would greatly facilitate the uptake
183 of the model since they can be directly implemented in an active, digital ecosystem for patient care. Conversely, this is also
184 a limitation of machine-learnt models deployed in such a fashion since they will likely generalize poorly outside the digital
185 application for which they were designed.

186 The predictive performance of the machine-learnt models we trained are limited by the number of variables available within
187 our dataset. There are many factors not considered by our model which have the potential to be predictive of acute COPD
188 exacerbations. For example, various health and lifestyle factors (including comorbidity and socioeconomic status) not known to
189 our model are believed to impact COPD exacerbation frequency²². These could be acquired from a patient's medical record
190 or collected with questionnaires and used by the algorithm to further refine its predictions. Additionally, since self-reported
191 deterioration in symptoms can occur several days before an exacerbation (Figure 2 e), it is reasonable that symptom information
192 could be collected in a more granular manner in the days preceding an exacerbation event. This could be achieved with medical
193 devices (e.g. smart inhalers), wearables designed to monitor person's lung function and COPD related physiological and
194 behavioural variables, or through more granular self-reporting of symptoms through the digital application²³. While research
195 has shown that environmental factors (such as pollution and weather) localised to a user could potentially be predictive of future
196 COPD exacerbation events^{24,25}, available data sets from pollution monitoring stations are limited by spatial resolution, as are
197 proxies for exposure, such as distance of residence from roadside or other point sources. While such datasets may be appropriate
198 for population-level analyses, they are not well suited for understanding effects of pollution on a heterogeneous population at
199 the individual level. Furthermore, such approaches do not allow incorporation of individual movement, with error as soon as the
200 individual moves away from the monitoring/residential site, and potentially significant errors relating to differential exposures
201 in microenvironments (e.g. indoors at home, in vehicles). Further work is needed to study the chemical biological interaction
202 between individual movement, hyperlocal pollution monitoring, and the prediction of acute COPD exacerbation.

203 Our study has several other limitations. Firstly, while our self-reports were collected in a prospective fashion using momentary
204 assessments of patients COPD symptoms, reporting compliance and individual variation in reporting behaviours could still
205 be detrimental to our results, with exacerbation frequencies or severity being misreported²⁶. Imperfect reporting compliance
206 also acts to limit the amount of data available to our machine-learnt model which may limit their predictive ability. Reporting
207 compliance could be improved by increased clinician engagement but it is unlikely clinicians will have sufficient time to
208 monitor and engage with the patient self-reports to a sufficient level. Finally, we used a validation set featuring unique patients
209 that were extracted from the same study population as those in the training set and therefore may not meet all requirements to
210 be considered independent. A more robust validation set containing an external cohort of patients is desirable.

211 For the machine-learned models it is important to match their configuration to the escalation policy. If models are used as a binary
212 alert system, this is achieved by analysing the (so-called) sensitivity-specificity trade-off, considered in Table 2 or by inspection
213 of Figure 4. In Table 2 three configurations of the Random Forest model are chosen (models c, d, and e) which yield different
214 sensitivities and specificities. For example, model d in Table 2 uses a threshold chosen to obtain a specificity of 0.25 on the test
215 set and achieves a sensitivity of 0.921 (95 % CI: 0.907-0.935). This configuration could be appropriate if false positives are not
216 of significant concern (e.g., if the prescribed intervention plan is of little risk to the patient). Ultimately, different configurations
217 allow flexibility in the resulting escalation policies and is the key advantage of the machine-learned models compared to the
218 baseline model.

219 In order to influence clinical outcome beneficially, a predictive model needs to enable a pre-emptive intervention with the likely
220 impact maximised the earlier this occurs. In the context of COPD exacerbations interventions may include increased use of
221 inhaled medication or additional administration of rescue packs of oral antibiotics and corticosteroids²⁸. The current standard
222 of care has been for patients to take these treatments when they are actually exacerbating, which creates a reactive model of
223 care requiring significant clinical deterioration to have occurred before an intervention is started. Our own work has shown that
224 early treatment of exacerbations is associated with improved clinical outcomes including faster recovery times²⁷. Therefore, if
225 presented appropriately this risk prediction model could enable patients to self-manage more effectively intervening before
226 life-threatening inflammation and infection can become established.

227 To conclude, data self-reported to a digital health application, designed for the management of people with COPD, can be
228 used to identify users at risk of exacerbation within three days with moderate discriminative ability (AUROC 0.727 (95 % CI:
229 0.720-0.735)). Further research utilizing additional linked data (particularly from medical devices such as smart inhalers and
230 environmental sensors) are expected to increase the accuracy of these models.

231 **Contributors**

232 FPC performed the analysis with support from DKB and MJB. FPC wrote the first draft of the manuscript. FPC, JBP, MJB
233 wrote the second draft of the manuscript. JBP obtained ethical and governance approvals. FPC, JBP, and MJB led the research
234 project at UoS. AB managed the data extraction at mymhealth. TMAW and AB provided clinical insight. FPC, MJB and
235 TMAW envisaged the research. All other authors contributed to future iterations of the manuscript.

236 **Acknowledgements**

237 We acknowledge support and discussions with respect to concept and data analysis from Dr B. Arbab-Zavar and Dr Zoheir
238 Sabeur.

239 **Ethical approval and data governance**

240 This work received ethics approval from the University of Southampton's Faculty of Engineering and Physical Science
241 Research Ethics Committee (ERGO/FEPS/52137) and was reviewed by the University of Southampton Data Protection Impact
242 Assessment panel (DPIA 0045), with the decision to support the research.

243 **Data Sharing**

244 Data will be made available upon reasonable request to persons with a university affiliation. Requestors will need appropriate
245 data protection, governance, and ethical review in place.

246 **Declaration of interests**

247 TMAW is Chief Science Officer and Co-Founder of mymhealth, the developer of the myCOPD application. AB is a Senior
248 Research Nurse and Clinical Trial Manager at mymhealth. All other authors declare no competing interests.

249 **References**

- 250 1. Lozano R, Naghavi M, Foreman K, et al. *Global and regional mortality from 235 causes of death for 20 age groups in 1990*
251 *and 2010: a systematic analysis for the Global Burden of Disease Study 2010*. The Lancet 2012; 380(9859):2095-128.
- 252 2. Miravittles M, Vogelmeier C, Roche N, et al. *A review of national guidelines for management of COPD in Europe*..
253 *European Respiratory Journal*. 2016; 47(2):625-37.
- 254 3. McLean S, Hoogendoorn M, Hoogenveen RT, et al. *Projecting the COPD population and costs in England and Scotland:*
255 *2011 to 2030*.. Scientific reports. 2016; 6(1):1-0.
- 256 4. Rodriguez-Roisin R. *Toward a consensus definition for COPD exacerbations*.. Chest. 2000; 117(5):398S-401S.

- 257 5. Hoogendoorn M, Feenstra TL, Hoogenveen RT, Al M, Rutten-van Mólken M. *Association between lung function and*
258 *exacerbation frequency in patients with COPD*. International journal of chronic obstructive pulmonary disease. 2010;
259 5:435.
- 260 6. Donaldson GC, Seemungal TA, Bhowmik A, Wedzicha JA. *Relationship between exacerbation frequency and lung*
261 *function decline in chronic obstructive pulmonary disease*. Thorax. 2002; 57(10):847-52.
- 262 7. GC Donaldson and JA Wedzicha. *Copd exacerbations 1: Epidemiology*. Thorax, 61(2):164–168, 2006.
- 263 8. Wedzicha JA, Seemungal TA. *COPD exacerbations: defining their cause and prevention*. The Lancet. 2007;
264 370(9589):786-96.
- 265 9. Adibi A, Sin DD, Safari A, et al. *The Acute COPD Exacerbation Prediction Tool (ACCEPT): a modelling study*. The
266 Lancet Respiratory Medicine. 2020.
- 267 10. Guerra B, Gaveikaite V, Bianchi C, Puhan MA. *Prediction models for exacerbations in patients with COPD*. European
268 Respiratory Review. 2017; 26(143):160061.
- 269 11. Sobnath DD, Philip N, Kayyali R, et al. *Features of a mobile support app for patients with chronic obstructive pulmonary*
270 *disease: literature review and current applications*. JMIR mHealth and uHealth. 2017; 5(2):e17.
- 271 12. Velardo C, Shah SA, Gibson O, et al. *Digital health system for personalised COPD long-term management*. BMC medical
272 informatics and decision making. 2017; 17(1):1-3.
- 273 13. North M, Bourne S, Green B, et al. *P238 A randomised controlled feasibility trial of an E-health platform supported care*
274 *vs usual care after exacerbation of COPD*. Thorax 2018; 73:A231.
- 275 14. Jones PW, Harding G, Berry P, Wiklund I, Chen WH, Leidy NK. *Development and first validation of the COPD Assessment*
276 *Test*. European Respiratory Journal. 2009; 34(3):648-54.
- 277 15. Dodd JW, Hogg L, Nolan J, et al. *The COPD assessment test (CAT): response to pulmonary rehabilitation. A multicentre,*
278 *prospective study*. Thorax. 2011; 66(5):425-9.
- 279 16. Gupta N, Morogan A, Pinto LM, Bourbeau J. *B43 COPD: SCREENING AND DIAGNOSTIC TOOLS: The Reliability,*
280 *Validity, Responsiveness And Minimum Clinically Important Difference Of The COPD Assessment Test: A Systematic*
281 *Review*. American Journal of Respiratory and Critical Care Medicine. 2014;189:1.
- 282 17. North, M, et al. *A randomised controlled feasibility trial of E-health application supported care vs usual care after*
283 *exacerbation of COPD: the RESCUE trial*. NPJ digital medicine. 2020; 3(1):1-8
- 284 18. Bergstra JS, Bardenet R, Bengio Y, Kégl B. *Algorithms for hyper-parameter optimization*. In Advances in neural
285 information processing systems 2011; 2546-2554.
- 286 19. Bergstra J, Yamins D, Cox D. *Making a science of model search: Hyperparameter optimization in hundreds of dimensions*
287 *for vision architectures*. In International conference on machine learning 2013; 115-123.
- 288 20. Youden WJ. *Index for rating diagnostic tests*. Cancer. 1950; 3(1):32-5.
- 289 21. Varol Y, Ozacar R, Balci G, Usta L, Taymaz Z. *Assessing the effectiveness of the COPD assessment test (CAT) to evaluate*
290 *COPD severity and exacerbation rates*. COPD: Journal of Chronic Obstructive Pulmonary Disease. 2014; 11(2):221-5.
- 291 22. Halpin DM, Miravitlles M, Metzdorf N, Celli B. *Impact and prevention of severe exacerbations of COPD: a review of the*
292 *evidence*. International journal of chronic obstructive pulmonary disease. 2017; 12:2891.
- 293 23. Walters EH, Walters J, Wills KE, Robinson A, Wood-Baker R. *Clinical diaries in COPD: compliance and utility in*
294 *predicting acute exacerbations*. International journal of chronic obstructive pulmonary disease. 2012; 7:427.
- 295 24. MacNee W, Donaldson K. *Exacerbations of COPD: environmental mechanisms*. Chest. 2000; 117(5):390S-7S.
- 296 25. Li J, Sun S, Tang R, Qiu H, Huang Q, Mason TG, Tian L. *Major air pollutants and risk of COPD exacerbations: a*
297 *systematic review and meta-analysis*. International journal of chronic obstructive pulmonary disease. 2016;11:3079.
- 298 26. Stone AA, Shiffman S. *Capturing momentary, self-report data: A proposal for reporting guidelines*. Annals of Behavioral
299 Medicine. 2002; 24(3):236-43.

- 300 **27.** Wilkinson TM, Donaldson GC, Johnston SL, Openshaw PJ, Wedzicha JA. *Respiratory syncytial virus, airway inflam-*
301 *mation, and FEV1 decline in patients with chronic obstructive pulmonary disease.* American journal of respiratory and
302 critical care medicine. 2006;173(8):871-6.
- 303 **28.** Global Initiative for Chronic Obstructive Lung Disease (GOLD) *Global Strategy for the Diagnosis, Management and*
304 *Prevention of COPD.* 2020; <https://goldcopd.org/> (Accessed 25 Nov 2020)

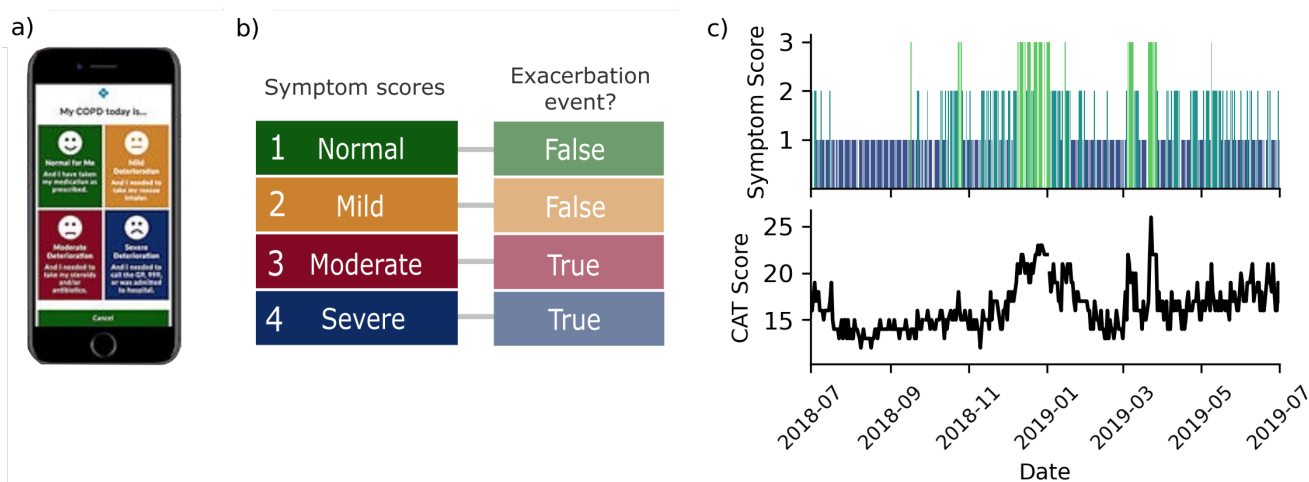


Figure 1. Self-reported symptom scores and COPD Assessment Test (CAT) results. Landing page of the myCOPD smartphone application where users self-report their daily symptom score. b) Symptom score rankings and classification of whether this score corresponds to an exacerbation event, as defined in the context of this work. c) Example user (with high reporting frequency) self-reporting timeline where the top panel displays self-reported symptom scores and the bottom panel self-reported CAT results.

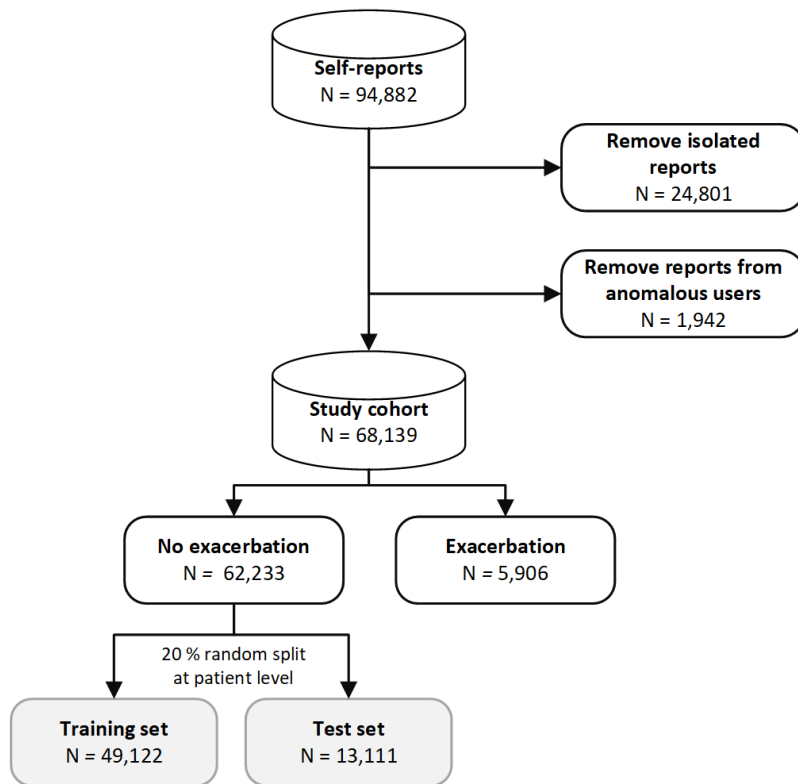


Figure 2. Selection of self-reports in our study cohort. Isolated reports (N=24,801) are those without a subsequent report in the following three days. Anomalous users (N=1,942) are those who only report exacerbation events or self-report to the myCOPD application before their registration date.

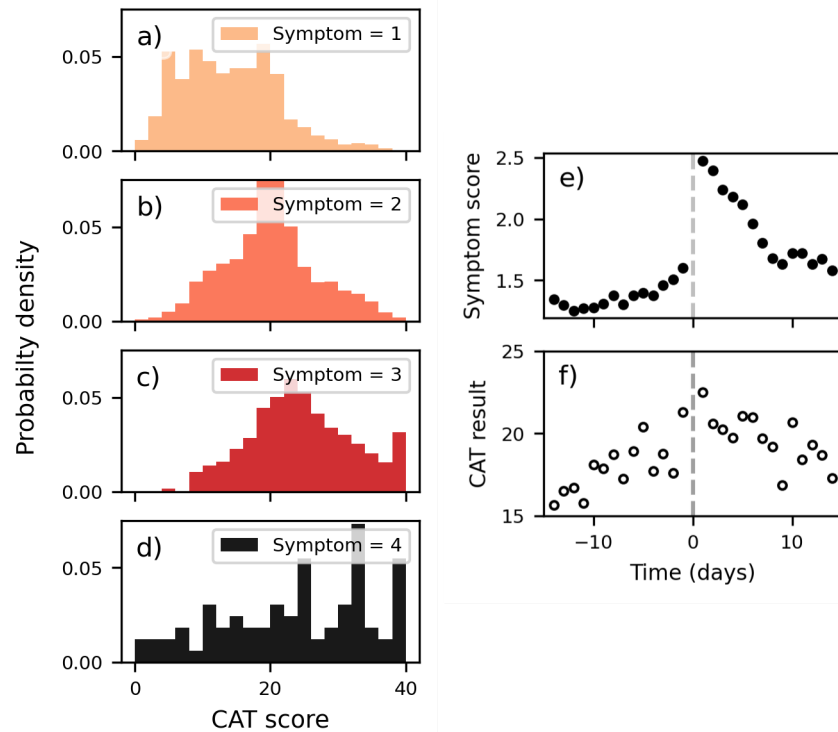


Figure 3. Self-reported symptom scores and results of COPD Assessment Tests for reports in our cohort. a-d) Displays the self-reported CAT result stratified by the self-reported symptom score (row) on the day of test completion. e) Mean self-reported symptom scores in the days preceding (and following) a day where a patient self-reports their first exacerbation event. f) Mean self-reported result of CAT in the days preceding (and following) a day where a patient self-reports their first exacerbation event. Grey dashed lines in all panels highlight the day of the first reported exacerbation event (time = 0 days). Panels e and f indicate that exacerbation events can be associated with a worsening of symptom scores and CAT results several days in advance of the event. The width of the observed ‘peaks’ (see panel e, right of dashed line) following the start of the exacerbation event demonstrate that exacerbation events can be multiple day events.

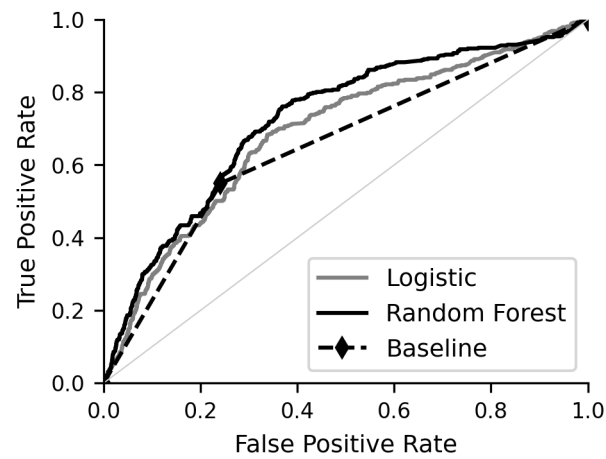


Figure 4. Receiver operating characteristic curve of our models evaluated on the patient hold-out test set. The baseline model (dashed line) has only one, non-trivial threshold for dichotomizing the prediction (diamond marker), whereas the machine learnt models have a large number of thresholds which needs to be optimized to suit the use-case (so called sensitivity-specificity trade-off).

	Patient count (%)
Age group	
Missing	10 (0.4)
19-29	7 (0.3)
30-39	39 (1.6)
40-49	89 (3.7)
50-59	325 (13.7)
60-69	791 (33.3)
70-79	881 (37.1)
80-89	212 (8.9)
90-99	15 (0.6)
100-110	5 (0.2)
Gender	
Missing	1,724 (72.6)
Female	231 (9.7)
Male	419 (17.6)
Smoking status	
Missing	1,217 (51.3)
Ex-smoker	843 (35.3)
Non-smoker	156 (6.6)
Smoker	158 (6.7)

Table 1. Patient demographics and smoking status in our cohort. All information was self-reported to the myCOPD application. Patient information relates to the time at which the patient first reported to the application (e.g., if there smoking status changed, this table summarizes the first reported status). There are 2,374 unique patients in total.

	Model	AUROC (95 % CI)	Threshold	Sensitivity (95 % CI)	Specificity (95 % CI)
a	Baseline model	0.655 (0.632-0.676)	N/A	0.551 (0.508-0.596)	0.759 (0.752-0.767)
b	Logistic Regression	0.697 (0.689-0.711)	Youden's J Statistic	0.708 (0.625-0.768)	0.644 (0.574-0.706)
c	Random Forest	0.727 (0.720-0.735)	Youden's J Statistic	0.755 (0.676-0.813)	0.629 (0.564-0.700)
d	Random Forest	-	Specificity = 0.25	0.921 (0.907-0.935)	0.250 (0.246-0.254)
e	Random Forest	-	Specificity = 0.75	0.576 (0.553-0.594)	0.750 (0.749-0.751)

Table 2. Model performances evaluated on the hold-out test set. The AUROC column denotes the area under the receiving operator curve (Figure 4) for each model. The three right-most columns display the sensitivity and specificity of models at predicting exacerbations with different thresholds used to dichotomize the predictions. The baseline model is already binary and only has one non-trivial configuration but the threshold used to dichotomize the machine learning models (b-e) can be tuned to suit the intended context of the model. The maximum of Youden's J statistic is used as a baseline criterion for dichotomizing the prediction (model b and c) and other cut-offs yielding fixed specificities are investigated for the Random Forest model.