ORIGINAL RESEARCH

Predicting Severe Chronic Obstructive Pulmonary Disease Exacerbations

Developing a Population Surveillance Approach with Administrative Data

Hamid Tavakoli[1], Wenjia Chen[1], Don D. Sin[2], J. Mark FitzGerald[3], and Mohsen Sadatsafavi[1,3]; for the Canadian Respiratory Research Network

[1]Respiratory Evaluation Sciences Program, Collaboration for Outcomes Research and Evaluation, Faculty of Pharmaceutical Sciences, and [3]Institute for Heart and Lung Health, Department of Medicine, University of British Columbia, Vancouver, British Columbia, Canada; and [2]Centre for Heart Lung Innovation, St. Paul's Hospital, Vancouver, British Columbia, Canada

ORCID ID: 0000-0003-0574-7739 (H.T.).

Abstract

Rationale: Automatic prediction algorithms based on routinely collected health data may be able to identify patients at high risk for hospitalizations related to acute exacerbations of chronic obstructive pulmonary disease (COPD).

Objectives: To conduct a proof-of-concept study of a population surveillance approach for identifying individuals at high risk of severe COPD exacerbations.

Methods: We used British Columbia's administrative health databases (1997–2016) to identify patients with diagnosed COPD. We used data from the previous 6 months to predict the risk of severe exacerbation in the next 2 months after a randomly selected index date. We applied statistical and machine-learning algorithms for risk prediction (logistic regression, random forest, neural network, and gradient boosting). We used calibration plots and receiver operating characteristic curves to evaluate model performance based on a randomly chosen future date at least 1 year later (temporal validation).

Results: There were 108,433 patients in the development dataset and 113,786 in the validation dataset; of these, 1,126 and 1,136, respectively, were hospitalized for COPD within their outcome windows. The best prediction algorithm (gradient boosting) had an area under the receiver operating characteristic curve of 0.82 (95% confidence interval, 0.80–0.83), which was significantly higher than the corresponding value for the model with exacerbation history as the only predictor (current standard of care: 0.68). The predicted risk scores were well calibrated in the validation dataset.

Conclusions: Imminent COPD-related hospitalizations can be predicted with good accuracy using administrative health data. This model may be used as a means to target high-risk patients for preventive exacerbation therapies.

Keywords: chronic obstructive pulmonary disease; population surveillance; risk prediction; machine learning; big data

(Received in original form January 27, 2020; accepted in final form May 8, 2020)

Supported by the Canadian Respiratory Research Network.

Author Contributions: H.T. articulated and initiated the idea. H.T. and M.S. designed the study and created the data analysis plan. D.D.S. and J.M.F. provided feedback on the design. H.T. implemented all of the statistical and machine-learning analyses. W.C. provided feedback on machine-learning performance indicators. H.T. wrote the first draft of the manuscript. All authors critically commented on the manuscript and approved the final version. H.T. and M.S. are guarantors of the manuscript.

Correspondence and requests for reprints should be addressed to Mohsen Sadatsafavi, M.D., Ph.D., Faculty of Pharmaceutical Sciences, University of British Columbia, Vancouver, 2405 Wesbrook Mall, Vancouver, BC, V6T1Z3 Canada. E-mail: msafavi@mail.ubc.ca.

This article has a related editorial.

This article has an online supplement, which is accessible from this issue's table of contents at www.atsjournals.org.

Ann Am Thorac Soc Vol 17, No 9, pp 1069–1076, Sep 2020

DOI: 10.1513/AnnalsATS.202001-070OC
Internet address: www.atsjournals.org

Chronic obstructive pulmonary disease (COPD) is a common airway disorder that is associated with a high risk of morbidity and mortality (1). In Canada and many other countries, severe exacerbations of COPD are one of the leading causes of hospitalization (2). COPD is preventable and treatable, and as such it is considered "ambulatory care sensitive," indicating that with proper outpatient care, very few patients should require hospitalization for acute exacerbations (3, 4). In addition to

maintenance inhaler therapies that moderately reduce exacerbation risk, most exacerbations, if caught early in the process, can be managed in an outpatient setting (e.g., with systemic corticosteroids and/or antibiotics), reducing the risk of hospital admissions.

In this context, identifying individuals at high risk of severe exacerbations can provide valuable opportunities for prevention. Most previous risk-prediction models for COPD exacerbations were developed for use at point of care (5–8); however, an alternative approach is to use routinely collected health data (9). This approach is becoming increasingly relevant because of advances in health data collection, integration, and availability. Electronic health records and administrative or claims databases that are collected for clinical management or administrative purposes often encompass a wealth of information that can be used for predictions *en masse*. A distinct advantage of this approach is that risk prediction can be performed remotely and can be made arbitrarily complex to improve its accuracy. In this "population surveillance" approach, individuals who are identified as high-risk can then be contacted for preventive disease management.

Although administrative health data typically do not have nuanced clinical variables such as lung function metrics and symptom intensity, they can include other information, such as coexisting conditions, that are unaffected by patient recall. As such, these prediction models might "make up" for the lack of some conventional predictor variables with the availability of others. Our aim in this study was to develop and evaluate the prediction model for severe COPD exacerbations using administrative health data.

Methods

In reporting the results of this predictive modeling work, we have adhered to the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) statement (10).

Data Sources and Study Design

We used administrative health databases containing information regarding health service use by all legal residents of British Columbia (BC, a Canadian province with a population of 4.8 million in 2015) (11) between January 1997 and December 2016. The following databases were used: demographics (12), vital statistics (13), hospitalizations (14), outpatient services (15), and medication dispensation records (16). Because the data are used for budget allocations and provider reimbursements, these databases have very low rates of missing or incorrect values (17). This study was approved by The University of British Columbia's Clinical Research Ethics Board (H13-00684). All inferences, opinions, and conclusions expressed in this work are those of the authors and do not reflect the opinions or policies of the data steward(s).

We first created a COPD cohort by applying a validated case definition (18). Individuals were marked as a patient with COPD if they had any of the *International Classification of Diseases, Ninth Revision* (ICD-9) codes of 491, 492, 496, or 493.2, or ICD-10 codes of J41, J42, J43, and J44 in the hospitalization dataset, or if they had three COPD-related outpatient visits with the above-mentioned ICD codes within a 12-month rolling period. A previous chart review established a sensitivity of 57.3% and a specificity of 96.9% for this algorithm (18).

From these large datasets, we created a development dataset, which we used to develop risk prediction models, and a validation dataset, which we used to model performance (calibration and discrimination). For the development dataset, we assigned a random date (referred to as the index date) to each individual after the date of their COPD diagnosis and between January 1, 2012, and December 31, 2014. This approach is aligned with the purpose of the algorithm, which is to be run on certain time intervals (a random date within the timeline of each individual's disease history).

We excluded patients who were not registered in the provincial insurance plan on their index date or had a long hospitalization stay ($\geq$60 d) within 6 months before the index date. During the immediate postexacerbation period, patients are at a high risk of readmission (19); however, patients are already under intensified care in the immediate postexacerbation episodes, and readmissions are likely a continuation of the original exacerbation (or its complications), rather than a *de novo* exacerbation episode. As such, individuals in the immediate postexacerbation period (defined as $\leq$30 d from COPD-related hospital discharge) were excluded.

Prediction Outcome

The primary outcome was the occurrence of at least one COPD-related hospitalization within the 2-month period after the index date (the outcome window). A COPD-related hospitalization was defined as a hospital admission with the most responsible diagnostic code for COPD, using the same ICD codes that were used for the COPD case definition.

Predictors

The assessment window, during which predictors were measured, was the 180-day period immediately before the index date. An initial set of the broad predictor categories was identified based on consensus among clinical experts within our team, and after a scoping review of predictors in published exacerbation risk models (5). The broad categories were translated into a candidate list of variables by the analyst. Predictors with low variability (e.g., binary variables that had the same value in >99% of cases) were removed. This process resulted in an initial broad list of predictors ($n = 83$). The list is provided in Table E1 of the online supplement. Count variables indicated the total number of events within the assessment period (e.g., the number of times a physician visit occurred owing to diabetes). All continuous variables were divided into their decile categories to protect against outlying values, which may overly influence predictions. Predictors for both the development and validation sets were measured within the last 6 months of the index date; however, because individuals with a 30-day history of severe exacerbations were excluded, this particular predictor was only measurable from Month 6 to 30 days before the index date.

Statistical Analysis

We examined several established statistical and machine-learning methods for risk prediction. In particular, we compared logistic regression (LR), random forest (RF), neural network (NN), and gradient-boosting (GB) methods (20). Using the development dataset, we trained models to estimate the probability of a severe exacerbation given the predictor values. All models were trained using a 10-fold cross-validation method. In this method, the development dataset was divided into training (90% of data) and test (10% of data) datasets. Each model was fitted in the

development dataset and its performance was evaluated in the test dataset (this was repeated 10 times; each time a different 10% was used as the test dataset). The best model was the one that had the highest average accuracy in the test dataset among 10 repetitions. For all algorithms, we applied downsampling to alleviate the imbalance in data due to the rarity of the outcome (21). We used SAS Enterprise (version 7.15) for cohort generation and *R* to fit the prediction algorithms and evaluate their performance.

A history of previous exacerbations is considered the best predictor of future exacerbations and forms the current basis of risk stratification in guidelines (22, 23). Therefore, to provide a basis for comparing the predictive power of the algorithms, we constructed a reference model that included the number of exacerbations (with moderate and severe as separate predictors) during the assessment window as the only predictors. Moderate exacerbation was defined based on a previously published algorithm that used outpatient visits and prescriptions for oral corticosteroids and antibiotics (24). To simplify the presentation, among the other models, the one with the best overall performance was identified as the "best" model, and its performance was considered vis-a-vis the reference model. Results that include all competing models are reported in the online supplement.

Examining Prediction Validity

To compare the performance of the models, we performed a temporal validation exercise based on the same source of data but at a different time point (10). Such temporal validation is the most relevant approach in this context given that the prediction model is intended to be used at future time points on the same dataset. External validation using an independent cohort was not considered because the prediction equations would not be exported to the data from other settings. Instead, other jurisdictions can use a similar methodology to develop a prediction algorithm that relies on the specifics of their own data. To perform temporal validation, we assigned a random date between January 1, 2015 and December 31, 2015 as the time point for new predictions, guaranteeing no time overlap between any of the development and validation time windows. Using the validation dataset, we evaluated the prediction algorithms in terms of discrimination, calibration, and potential utility as population surveillance tools.

Model Discrimination

The term "discrimination" refers to the capacity of a prediction algorithm to distinguish individuals at different risks of an outcome (25). We first evaluated the discrimination of the models by drawing the receiver operating characteristic (ROC) curve and evaluating its area under the curve (AUC) (26). The ROC evaluates the trade-off between the sensitivity and specificity of the model within the range of various thresholds for considering a patient at high risk of exacerbation (27). A nonparametric test was used to evaluate the differences in ROC curves among the models (28). Higher values of AUC indicate better discrimination power.

Model Calibration

The term "calibration" refers to the degree to which the predicted risks agree with the observed risks. To assess the calibration of the models, we drew a calibration plot by dividing the predicted risks into deciles and calculating the mean observed risk within each decile. The lines that connect the resulting points should be close to the diagonal line if the model is well calibrated.

The Utility of the Prediction as a Population Surveillance Method

In population-based predictions, where a large number of individuals are screened to capture relatively infrequent events, it is important for the prediction model to have a high positive predictive value so that not too many patients are flagged as high-risk. Frequently, the rate-limiting step in a population surveillance program is the resources that are required to follow such

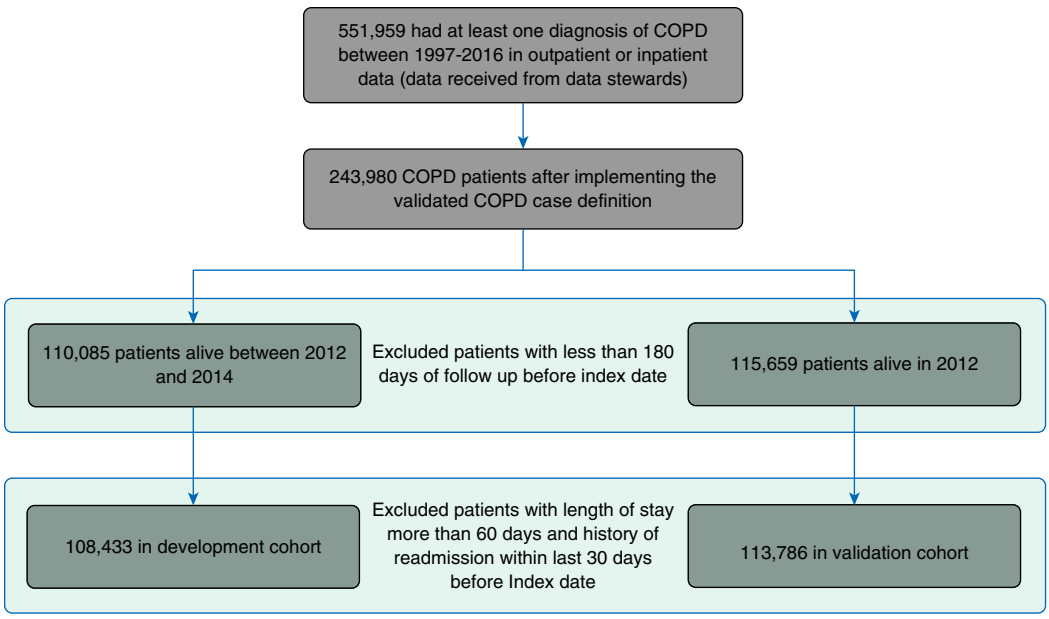

Figure 1. Flowchart of data extraction and cohort generation. COPD = chronic obstructive pulmonary disease.

individuals. However, the positivity threshold can be set high enough that a desired maximum number of individuals will be flagged as high-risk (as such, unlike clinical prediction models that are often bundled with fixed thresholds, surveillance models can have variable thresholds to give a desire number of high-risk cases). We visualized the performance of each algorithm in terms of the number of identified individuals who will experience the event when the threshold is placed such that a fixed number of individuals (e.g., 10,000) are flagged as high-risk at each model run. The more efficient the prediction algorithm, the higher the proportion of flagged individuals who will actually exacerbate, and thus the higher the yield of the population surveillance approach. On the other hand, at such high thresholds, the prediction algorithm will have low sensitivity (many patients who will be hospitalized in the next 2 mo will not be flagged). The precision-recall (PR) curve, which compares the sensitivity (recall) against the positive predictive value (precision) is a relevant graphical presentation of the performance of a prediction model in this context. In addition, when the outcome is rare, the area under the PR curve (AUCPR) remains a reliable summary statistic for comparing the performance of competing models (29). We therefore also present the PR curve and report its AUC.

Sensitivity Analysis

We implemented two sensitivity analyses. First, to examine the effect of penalizing the model against overfitting, we applied Least Absolute Shrinkage and Selection Operator (LASSO) (30). LASSO uses a penalization algorithm for coefficient estimation, and can result in the removal of some predictors from the model (30). In the second analysis, we considered the outcome to be the occurrence of death or severe exacerbations.

Results

Figure 1 provides a schematic illustration of cohort generation.

By applying the COPD case definition, we included 108,433 patients in the development dataset (49.2% female, average age on index date 70.7 [standard deviation = 12.2]). In the validation dataset, 113,786 individuals were included (49.4% female, average age on index date 71.2 yr [standard deviation = 12.0 yr]). In the development dataset, 1,126 patients were hospitalized for COPD within 60 days after their index date (1.0% of individuals); the corresponding value was 1,136 in the validation dataset (1.0%). The sociodemographic characteristics of the development and validation samples are provided in Table 1.

Performance in the Validation Dataset

Model discrimination. Among the competing models in development set, the GB model had the highest performance, with an AUC of 0.88 (95% confidence interval [CI], 0.87–0.89), and the reference model had the lowest AUC (0.68; 95% CI, 0.67–0.69). For the rest of models, the AUCs were as follows: LR 0.82 (95% CI, 0.81–0.83), RF 0.83 (95% CI, 0.82–0.84), and NN 0.84 (95% CI, 0.83–0.85).

In the validation sample, the GB algorithm had the highest AUC (0.82; 95% CI, 0.80–0.83). All other models had slightly lower discriminatory performance: LR 0.80 (95% CI, 0.79–0.81), RF 0.81 (95% CI, 0.80–0.82), and NN 0.79 (95% CI, 0.78–0.81). In comparison, the reference model (only including previous exacerbation history) had an AUC of 0.68 (95% CI, 0.67–0.70).

Table 1. Sociodemographic characteristics and selected predictor values for individuals in the development and validation samples

Variables	Development Sample	Validation Sample
Sociodemographic variables		
Age, yr (SD)	70.8 (12.2)	71.2 (12.0)
Female, *n* (%)	53,295 (49.2%)	56,270 (49.4%)
Socioeconomic status, *n* (%)		
Quantile 1	27,553 (25.4%)	28,338 (24.9%)
Quantile 2	23,100 (21.3%)	24,236 (21.3%)
Quantile 3	20,428 (18.8%)	21,698 (19.1%)
Quantile 4	18,858 (17.4%)	20,051 (17.6%)
Quantile 5	16,399 (15.1%)	17,472 (15.4%)
Unknown	2,095 (1.9%)	1,991 (1.8%)
Gradient boosting top 20 variables (SD)*		
Years from diagnosis of COPD	5.7 (4.6)	6.6 (4.9)
Total number of dispensed medications (all causes)	4.4 (6.9)	4.2 (7.2)
Number of COPD exacerbations	0.2 (0.6)	0.2 (0.7)
Total number of SABA prescriptions	0.8 (1.7)	0.7 (1.5)
Total number of ICS/LABA prescriptions	0.7 (1.5)	0.8 (1.5)
Total number of LAMA prescriptions	0.4 (1.2)	0.4 (1.3)
Total number of oral corticosteroid prescriptions	0.4 (1.8)	0.4 (2.1)
Total number of SAMA prescriptions	0.5 (1.5)	0.5 (1.4)
Total number of antibiotic prescriptions	0.8 (2.1)	0.8 (2.2)
Length of stay in previous hospitalization	1.4 (5.7)	1.4 (5.5)
Socioeconomic status	2.9 (1.6)	2.9 (1.6)
Total number of hospitalizations	0.2 (0.5)	0.2 (0.6)
Total number of outpatient respiratory symptoms	0.4 (1.3)	0.4 (1.2)
Total number of outpatient visits for pneumonia	0.2 (1.5)	0.2 (1.7)
Total number outpatient visits	14.7 (15.6)	14.1 (14.9)
Total number of ICS prescriptions	0.3 (1.0)	0.2 (1.0)
Total number of outpatient visits for heart failure	0.3 (1.7)	0.3 (1.7)
Total number of outpatient visits for hypertension	0.5 (1.1)	0.5 (1.0)
Total number of outpatient visits for cardiovascular diseases	0.3 (1.3)	0.3 (1.2)

Definition of abbreviations: COPD = chronic obstructive pulmonary disease; ICS = inhaled corticosteroid; LABA = long-acting β-agonist; LAMA = long-acting muscarinic antagonist; SABA = short-acting β-agonist; SAMA = short-acting muscarinic antagonist; SD = standard deviation.
*Age as one of the top 20 variables is addressed in the sociodemographic variables.

The ROC contrast test showed that the GB model statistically outperformed other models (all pairwise P values < 0.001). As such, in what follows, only the best model (GB) and the reference model are compared. Table 1 also demonstrates the distribution of the top 20 variables in the GB model.

Figure 2 (left) provides the ROC curves for the best model and the reference model. Figure E1 shows the ROC curves for all models.

Model calibration. Figure 2 (right) provides the calibration plots in the validation dataset for the best model and the reference model. Both models were well calibrated in the validation sample, although the reference model had a narrower range of predicted risks than the best model. Figure E2 shows the calibration plot for all models. The GB algorithm behind the best model provides a quantitative index for the importance of each variable in explaining the variation of the outcome (31); this is shown for the top 20 variables in Figure 3.

Utility of the Models as Potential Tools for Population Surveillance

Figure 4 (left) provides the PR curves in the validation dataset. The AUCPR also showed that the best model had a better performance (0.07) than the reference model (0.05). The rest of the models had AUCPRs of ~0.06. Figure E3 presents the PR curves for all models.

Figure 4 (right) demonstrates the number of true cases (individuals who will experience severe exacerbations) as a function of the number of individuals who are flagged as high-risk, for the best model as well as the reference model. At very low or high thresholds, the two models performed similarly. However, within the range of thresholds that would result in identifying 5,000–40,000 individuals at high risk, the gain in utility of the best model becomes clear. For example, if the threshold is set such that 10,000 individuals would be flagged as high-risk, the best model results in detection of 30 more true cases than the reference model. The difference was at its maximum, with 108 more true cases at a cutoff of 20,000. Table E2 presents the number of true cases for all models.

Sensitivity Analysis

Variable selection with LASSO resulted in 15 predictors, and the amount of shrinkage was very small. Table E3 shows the list of selected predictors. The GB model still showed a slightly higher performance than LASSO (ROC contrast test $P = 0.02$), with an AUC of 0.81 (95% CI, 0.80–0.82). When the analyses were repeated with death and severe exacerbations as a composite outcome, the results remained largely the same, with the GB algorithm resulting in the best model (AUC = 0.82; 95% CI, 0.81–0.83). The reference model had an AUC of 0.64 (95% CI, 0.62–0.65). The ROC contrast test showed that the GB model had a statistically higher performance.

Discussion

We examined the predictability of severe COPD exacerbations requiring hospitalizations in routinely collected administrative health data. We compared different prediction methods, all applied to a 6-month assessment window, to predict severe exacerbations in the subsequent 2 months (excluding readmissions). We observed that an established machine-learning method (GB) narrowly outperformed other prediction algorithms and resulted in a prediction model with a high discrimination power (AUC = 0.82), which also showed robust calibration in the validation data. Compared with such a model, a history of previous exacerbations, the current standard for evaluating the risk of future exacerbations, had a much lower discriminatory performance (AUC = 0.68).

To demonstrate the potential utility of the model, we examined how many individuals who are labeled as high-risk will actually experience exacerbations according to different positivity thresholds. For example, if a health administrator plans to follow up (e.g., through phone calls or communication with care providers) with 1,000 individuals each time the model is run, the model will be able to capture 143 individuals who will exacerbate in the next 2 months (14.3% yield). This is equal to a number-needed-to-contact to detect one COPD hospitalization of seven. The lower the positivity threshold, the higher the number of individuals who will be identified as high-risk, as well as the number of individuals who will actually exacerbate. However, the proportion of identified high-risk individuals who will exacerbate will decline with lower thresholds. For example,

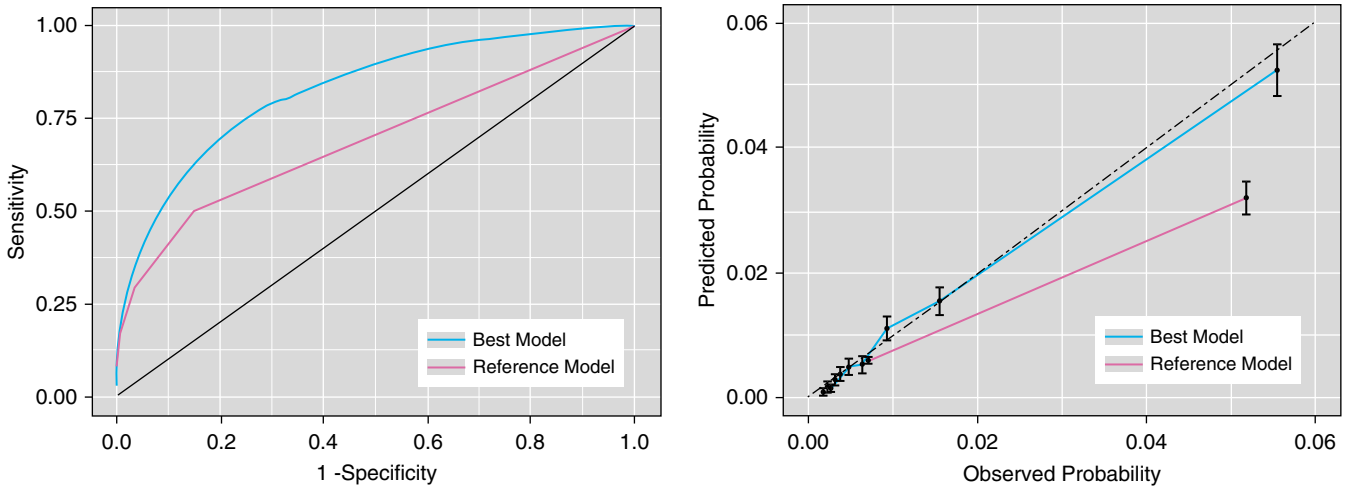

Figure 2. Left: receiver operating characteristic of validation data compared with the reference model. Right: calibration plot of the prediction algorithms in the validation dataset.

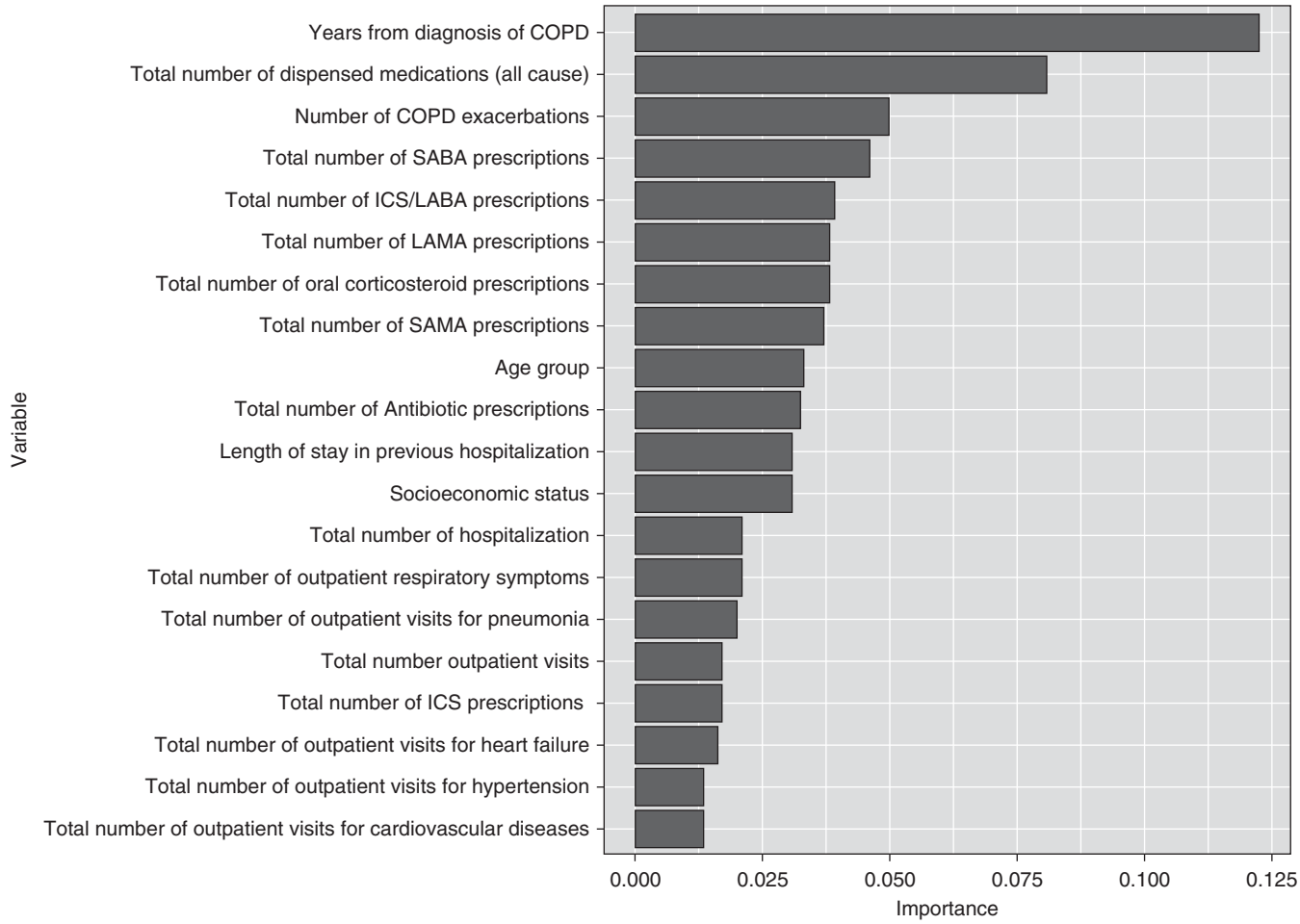

Figure 3. Relative importance of predictor variables (top 20) in the gradient-boosting model. COPD = chronic obstructive pulmonary disease; ICS = inhaled corticosteroid; LABA = long-acting β-agonist; LAMA = long-acting muscarinic antagonist; SABA = short-acting β-agonist; SAMA = short-acting muscarinic antagonist.

if the threshold is set such that 10,000 individuals are considered high-risk, 539 of them will exacerbate (5.4% yield), corresponding to a number-needed-to-contact of 19. In practice, changing the threshold can provide the flexibility within a health administration unit to scale such a surveillance program according to available resources.

We are aware of only one exacerbation prediction model, developed by Annavarapu and colleagues, that is based on administrative data and therefore conceivably could be used for population surveillance (5). Applied to 45,722 patients, 5,317 of whom experienced an exacerbation, the model included 29 variables, with the most important predictors being prior severe exacerbations and higher comorbidity scores. The AUC of this model was 0.77. The authors recommended a cutoff score that would be associated with a sensitivity of 17% and specificity of 98%. Of note, our model at such specificity will have a sensitivity of 23%. We do not believe we can compare the performance of our algorithm against those that were developed for readmissions (which typically occur within 21–30 d after discharge) (19), as the dynamics of readmissions will be different from those of *de novo* exacerbations. In addition, we believe that a population surveillance approach to readmission is less relevant given that readmitted patients will have recently interacted with care providers (thus, there would not be much to gain from alerting them or their care providers using a population surveillance approach). Compared with prediction algorithms based on routinely collected data, there are many more prediction algorithms based on clinical data that are applicable at point of care. Such models have access to more nuanced clinical features, such as lung function and symptoms. A recent systematic review identified 27 such models (32). Their performance in terms of AUC varied between 0.58 and 0.78. The high discriminatory power of our model compared with point-of-care prediction models, despite the lack of clinical variables, could be due to the inclusion of many more variables for prediction, in particular the burden of comorbidity.

The strengths of this study include its large development dataset, the use of a variety of machine-learning and statistical algorithms to find a model with high accuracy and robust calibration, and the use of temporal validation. The high level of

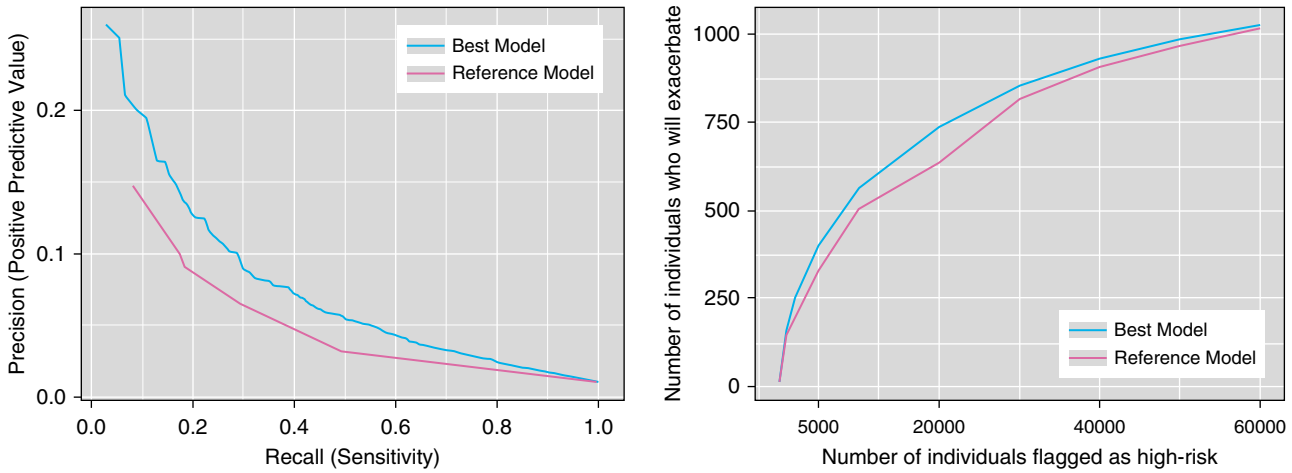

Figure 4. Left: precision-recall curve in the validation dataset for the best model (gradient boosting) and the reference model (logistic regression with history of exacerbations as the only predictor). Right: number of individuals who will exacerbate by different cutoff level (flagged by the prediction model).

accuracy (AUC = 0.82) compared with previous models for administrative health data (5) could be due to our use of a larger set of predictors (83 variables), more granular modeling of comorbid conditions, and examination of a variety of different prediction algorithms. Temporal validation provided assurances that a model that has "learned" from the data can preserve its performance for least 2 years. In practice, such a model can be constantly retrained and recalibrated to capture any temporal trends that might affect the accuracy of the predictions (e.g., trends in disease coding that might result in a change in exacerbation prevalence).

The limitation of this approach should also be considered. Results of this type provide information about the predictability of exacerbations based on routinely collected data, but the prediction model itself is not directly transferable to other jurisdictions. The performance characteristics can also vary with the amount of data that are available. For example, we did not have access to laboratory data (which can contain potentially important predictors such as blood eosinophil counts). It is likely that the prediction performance can be improved with such data. Finally, the "black box" nature of machine-learning algorithms (such as GB, which gave rise to the best model) might affect the acceptance of predictions among patients and care providers.

This study is a proof-of-concept demonstrating that a population surveillance approach for COPD exacerbations is plausible, but it did not consider the feasibility of this approach. The benefit of predicting exacerbations can only be realized by instituting disease management strategies in response to the prediction. If the identification of high-risk individuals does not trigger impactful actions (e.g., if adherence to preventive therapies is low), such a surveillance program will not be beneficial.

Conclusions

Severe COPD exacerbations are catastrophic but potentially preventable events. Preventing their occurrence will be associated with significant cost savings and gains in quality of life. Although vigilance against exacerbations at point of care has received much attention, a population surveillance approach based on routinely collected data can provide an independent path for prevention. Our findings show that exacerbations are indeed predictable in the routinely collected data. This finding may empower precision population health by enabling hospitals and care providers to identify high-risk individuals and implement population-based approaches to reduce this risk. In this work, we also propose a continuous learning protocol that will enable health authorities to use the available streaming data to identify high-risk patients based on the most updated data. ■

Author disclosures are available with the text of this article at www.atsjournals.org.

References

1 Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Med* 2006;3:e442.
2 Khakban A, Sin DD, FitzGerald JM, McManus BM, Ng R, Hollander Z, *et al*. The projected epidemic of chronic obstructive pulmonary disease hospitalizations over the next 15 years: a population-based perspective. *Am J Respir Crit Care Med* 2017;195: 287–291.
3 Bandurska E, Damps-Konstańska I, Popowski P, Jędrzejczyk T, Janowiak P, Świętnicka K, *et al*. Impact of integrated care model (ICM) on direct medical costs in management of advanced chronic obstructive pulmonary disease (COPD). *Med Sci Monit* 2017;23: 2850–2862.
4 Aaron SD, Vandemheen KL, Hebert P, Dales R, Stiell IG, Ahuja J, *et al*. Outpatient oral prednisone after emergency treatment of chronic obstructive pulmonary disease. *N Engl J Med* 2003;348:2618–2625.
5 Annavarapu S, Goldfarb S, Gelb M, Moretz C, Renda A, Kaila S. Development and validation of a predictive model to identify patients at risk of severe COPD exacerbations using administrative claims data. *Int J Chron Obstruct Pulmon Dis* 2018;13: 2121–2130.

6 Santibáñez M, Garrastazu R, Ruiz-Nuñez M, Helguera JM, Arenal S, Bonnardeux C, *et al*. Predictors of hospitalized exacerbations and mortality in chronic obstructive pulmonary disease. *PLoS One* 2016; 11:e0158727.
7 Riley RD, Ensor J, Snell KIE, Harrell FE Jr, Martin GP, Reitsma JB, *et al*. Calculating the sample size required for developing a clinical prediction model. *BMJ* 2020;368:m441.
8 Adibi A, Sin DD, Safari A, Johnson KM, Aaron SD, FitzGerald JM, *et al*. The acute COPD exacerbation prediction tool (ACCEPT): a modelling study. *Lancet Respir Med* [online ahead of print] 13 Mar 2020; DOI: 10.1016/S2213-2600(19)30397-2.32178776
9 Collier R. WHO guidelines on ethical public health surveillance. *CMAJ* 2017;189:E977.
10 Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, *et al*. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015;162:W1–W73.
11 British Columbia. Population estimates, BC stats, Central Statistical Agency of the Province of British Columbia. 2015 [accessed 2010 Oct 29]. Available from: https://www2.gov.bc.ca/gov/content/data/statistics/people-population-community/population/population-estimates access date is 2020/06/04.
12 British Columbia Ministry of Health. Consolidation File (MSP Registration & Premium Billing). Data Extract. MOH (2014). Population Data BC; 2013 [accessed 2020 Jun 4]. Available from: https://www.popdata.bc.ca/data/population/consolidationfile.
13 BC Vital Statistics Agency. Data Extract BC Vital Statistics Agency (2013). Vital Statistics Deaths. Population Data BC; 2013 [accessed 2020 Jun 4]. Available from: https://www.popdata.bc.ca/data/population/vsdeaths.
14 Canadian Institute for Health Information. Discharge Abstract Database (Hospital Separations). Data Extract. MOH (2014). Population Data BC; 2014 [accessed 2020 Jun 4]. Available from: https://www.popdata.bc.ca/data/health/dad.
15 British Columbia Ministry of Health. Medical Services Plan (MSP) Payment Information File. Data Extract. MOH (2014). Population Data BC; 2014 [accessed 2020 Jun 4]. Available from: https://www.popdata.bc.ca/data/health/msp.
16 BC Ministry of Health. PharmaNet. Data Extract. Data Stewardship Committee (2013). BC Ministry of Health; 2014 [accessed 2020 Jun 4]. Available from: https://www.popdata.bc.ca/data/health/PharmaNet.
17 Williams J, Adger W. Inventory of studies on the accuracy of Canadian health administrative databases. Toronto: Institute for Clinical Evaluative Sciences; 1996.
18 Gershon AS, Wang C, Guan J, Vasilevska-Ristovska J, Cicutto L, To T. Identifying individuals with physician diagnosed COPD in health administrative databases. *COPD* 2009;6:388–394.
19 Gershon AS, Thiruchelvam D, Aaron S, Stanbrook M, Vozoris N, Tan WC, *et al*. Socioeconomic status (SES) and 30-day hospital readmissions for chronic obstructive pulmonary (COPD) disease: a population-based cohort study. *PLoS One* 2019;14:e0216741.
20 Bzdok D, Altman N, Krzywinski M. Statistics versus machine learning. *Nat Methods* 2018;15:233–234.
21 Maimon O, Rokach L. Data mining and knowledge discovery handbook, 2nd ed. New York, NY: Springer; 2010.
22 Vestbo J, Hurd SS, Agustí AG, Jones PW, Vogelmeier C, Anzueto A, *et al*. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med* 2013;187:347–365.
23 Kerkhof M, Freeman D, Jones R, Chisholm A, Price DB; Respiratory Effectiveness Group. Predicting frequent COPD exacerbations using primary care data. *Int J Chron Obstruct Pulmon Dis* 2015;10: 2439–2450.
24 Dhamane AD, Moretz C, Zhou Y, Burslem K, Saverno K, Jain G, *et al*. COPD exacerbation frequency and its association with health care resource utilization and costs. *Int J Chron Obstruct Pulmon Dis* 2015; 10:2609–2618.
25 Zelterman D. Applied multivariate statistics with R. New York: Springer International Publishing; 2015.
26 Bertens LCM, Reitsma JB, Moons KGM, van Mourik Y, Lammers JWJ, Broekhuizen BDL, *et al*. Development and validation of a model to predict the risk of exacerbations in chronic obstructive pulmonary disease. *Int J Chron Obstruct Pulmon Dis* 2013;8:493–499.
27 Saito T, Rehmsmeier M. Precrec: fast and accurate precision-recall and ROC curve calculations in R. *Bioinformatics* 2017;33:145–147.
28 DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44:837–845.
29 Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One* 2015;10:e0118432.
30 Tibshirani R. Regression shrinkage and selection via the lasso: a retrospective. *J R Stat Soc Series B Stat Methodol* 2011;73:273–282.
31 Hastie T, Tibshirani R, Friedman JH. The elements of statistical learning: data mining, inference, and prediction, 2nd ed. New York: Springer; 2009.
32 Guerra B, Gaveikaite V, Bianchi C, Puhan MA. Prediction models for exacerbations in patients with COPD. *Eur Respir Rev* 2017;26: 160061.