

1 **Machine Learning Prediction of Progression in FEV₁ in the COPDGene Study**

2

3 Adel Boueiz^{1,2}, Zhonghui Xu¹, Yale Chang³, Aria Masoomi³, Andrew Gregory¹, Sharon M.
4 Lutz⁴, Dandi Qiao¹, James D. Crapo⁵, Jennifer G. Dy³, Edwin K. Silverman^{1,2}, Peter J.
5 Castaldi^{1,6}, for the COPDGene investigators.

6

7 ¹Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston,
8 MA; ²Pulmonary and Critical Care Division, Department of Medicine, Brigham and Women's Hospital,
9 Harvard Medical School, Boston, MA; ³Department of Electrical and Computer Engineering, Northeastern
10 University, Boston, MA; ⁴Department of Population Medicine, Harvard Pilgrim Health Care Institute, Boston,
11 MA; ⁵Division of Pulmonary Medicine, Department of Medicine, National Jewish Health, Denver, CO;
12 ⁶Division of General Medicine and Primary Care, Brigham and Women's Hospital, Harvard Medical School,
13 Boston, MA.

14

15 **Corresponding Author:** Adel Boueiz, MD, MMSc, Channing Division of Network
16 Medicine, Brigham and Women's Hospital, 181 Longwood Avenue, Boston, MA, 02115,
17 Phone number: 617-525-2111; Email: adel.boueiz@channing.harvard.edu

18

19 **Authors' email addresses:** Adel Boueiz (adel.boueiz@channing.harvard.edu), Zhonghui Xu
20 (zhonghui.xu@channing.harvard.edu), Yale Chang (changyalee@gmail.com), Aria Masoomi
21 (masoomi.a@northeastern.edu), Andrew Gregory (andrew.gregory@channing.harvard.edu),
22 Sharon M. Lutz (sharon.m.lutz@gmail.com), Dandi Qiao
23 (dandi.qiao@channing.harvard.edu), James D. Crapo (CrapoJ@njhealth.org), Jennifer G. Dy
24 (jdy@ece.neu.edu), Edwin K. Silverman (ed.silverman@channing.harvard.edu), Peter J.
25 Castaldi (peter.castaldi@channing.harvard.edu)

26

1 **Running Head:** Predicting COPD progression

2 **Descriptor:** 9.03 COPD: Clinical Phenotypes

3

4 **Manuscript Word Count:** 2,993

5

6 **Keywords:** Chronic obstructive pulmonary disease; Disease progression; 5-year changes in

7 FEV₁; Prediction; Random forest machine learning.

8

9 **Funding Sources:** This work was supported by NHLBI K08 HL141601, R01 HL124233,

10 R01 HL126596, R01 HL147326, U01 HL089897, and U01 HL089856. The COPDGene

11 study (NCT00608764) is also supported by the COPD Foundation through contributions

12 made to an Industry Advisory Board that has included AstraZeneca, Bayer Pharmaceuticals,

13 Boehringer-Ingelheim, Genentech, GlaxoSmithKline, Novartis, Pfizer, and Sunovion.

14

15 **Supplement:** This article has an online data supplement, which is accessible from this issue's

16 table of content online.

17

18 **Notation of prior abstract publication/presentation:** Boueiz A, Chang Y, Cho MH,

19 DeMeo DL, Dy J, Silverman EK, Castaldi PJ. Machine Learning Prediction of 5-Year

20 Progression of FEV1 in the COPDGene Study. [abstract]. *American Journal of Respiratory*

21 *and Critical Care Medicine* 2018;197(A7430).

22

23

24

25

26

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

ABBREVIATION LIST

- AUC: Area under the curve
- BMI: Body mass index
- COPD: Chronic Obstructive Pulmonary Disease
- COPDGene study: Genetic Epidemiology of COPD study
- Δ FEV₁: Annualized five-year changes in FEV₁
- FEV₁: Forced expiratory volume in one second
- FVC: Forced vital capacity
- GOLD: Global Initiative for Chronic Obstructive Lung Disease spirometric grading system
- HU: Hounsfield units
- IQR: Interquartile range
- %LAA-950: Percent of CT scan low attenuation area below -950 HU at end-inspiration
- MMRC: Modified Medical Research Council
- NHW: Non-Hispanic White
- RF: Random forest
- RMSE: Root mean squared error
- ROC: Receiver operator characteristic
- SGRQ: St. George's Respiratory Questionnaire

1 **ABSTRACT**

2

3 **Background:** The heterogeneous nature of COPD complicates the identification of the
4 predictors of disease progression and consequently the development of effective therapies.
5 We aimed to improve the prediction of disease progression in COPD by using machine
6 learning and incorporating a rich dataset of phenotypic features.

7 **Methods:** We included 4,496 smokers with available data from their enrollment and 5-year
8 follow-up visits in the Genetic Epidemiology of COPD (COPDGene) study. We constructed
9 supervised random forest models to predict 5-year progression in FEV₁ from 46 baseline
10 demographic, clinical, physiologic, and imaging features. Using cross-validation, we
11 randomly partitioned participants into training and testing samples. We also validated the
12 results in the COPDGene 10-year follow-up visit.

13 **Results:** Predicting the change in FEV₁ over time is more challenging than simply predicting
14 the future absolute FEV₁ level. Nevertheless, the area under the ROC curves for the
15 prediction of subjects in the top quartile of observed disease progression was 0.70 in the 10-
16 year follow-up data. The model performance accuracy was best for GOLD1-2 subjects and it
17 was harder to achieve accurate prediction in advanced stages of the disease. Predictive
18 variables differed in their relative importance as well as for the predictions by GOLD grade.

19 **Conclusion:** This state-of-the art approach along with deep phenotyping predicts FEV₁
20 progression with reasonable accuracy. There is significant room for improvement in future
21 models. This prediction model facilitates the identification of smokers at increased risk for
22 rapid disease progression. Such findings may be useful in the selection of patient populations
23 for targeted clinical trials.

24

25

1 INTRODUCTION

2 Chronic obstructive pulmonary disease (COPD) is the second leading cause of
3 disability, the third leading cause of death, and the only major chronic disease continuing to
4 increase in mortality¹⁻⁴. Novel therapies that slow disease progression could result in an
5 improvement in COPD patients' health status and have a substantial impact on healthcare
6 utilization. The development of such therapies will be aided by improved tools for predicting
7 disease progression, enabling the selection of high-risk groups for targeted treatment.

8 Predictive models incorporate multiple sources of information to make patient-
9 specific predictions and are widely used in multiple areas of medical practice. Existing
10 models of disease progression in COPD have been limited in the scope of variables assessed
11⁵⁻⁹. COPD exhibits significant variation in clinical and radiographic presentation as well as
12 disease progression^{6,10-12}. This disease heterogeneity complicates the identification of the
13 predictors of COPD progression and limits the accuracy of predictive models. Furthermore,
14 COPD often progresses slowly over decades and true disease progression over short time
15 periods can be difficult to detect with existing measurements.

16 In this study, we aimed to improve the prediction of COPD progression by applying
17 machine learning to a rich dataset of clinical, demographic, patient-reported variables, and
18 imaging features in the COPDGene study. We hypothesized that deep phenotyping at the
19 initial study visit along with random forest modelling, which exploits complex non-linear
20 relationships and interactions among the risk factors, would facilitate the prediction of the
21 rates of disease progression as measured by FEV₁, a key aspect of COPD.

22

23 MATERIALS AND METHODS

24

25 *Study populations*

1 ***Derivation cohort - COPDGene Study Visit 1 and Visit 2:*** We analyzed 4,496
2 smokers with complete CT scans and relevant covariate data at the baseline visit (Visit 1) and
3 5-year follow-up visit (Visit 2) in the COPDGene cohort (NCT00608764,
4 www.copdgene.org).

5 ***Temporal validation cohort - COPDGene Study Visit 3:*** During the Phase 3 of the
6 COPDGene Study, enrolled subjects returned for their 10-year follow-up visit. At the time of
7 this analysis, 1,833 smokers had completed their 10-year follow-up visit and had available
8 10-year spirometric and radiographic data. To predict their outcome values at Year 10 (Visit
9 3), we entered their 5-year (Visit 2) predictor data into the models trained in the derivation
10 cohort.

11 The COPDGene study design, subject enrollment, and phenotype measurements have
12 been previously reported ¹³ and additional information is included in the Supplement.

13

14 ***Outcome variables***

15 We constructed models to predict annualized follow-up FEV₁ and five-year changes
16 in FEV₁ (Δ FEV₁). Δ FEV₁ (mL/year) was calculated by subtracting the Visit 1 value from the
17 Visit 2 value and dividing by the time between Visit 1 and Visit 2. Negative values represent
18 a lower value of the outcome at Visit 2 (i.e. worsening of the disease over the 5-year period
19 with greater loss of FEV₁).

20

21 ***Feature selection***

22 Candidate predictors consisted of 46 baseline demographic, clinical, physiologic, and
23 imaging variables that were available in the COPDGene population at Visit 1 and had
24 correlation coefficients of less than 0.90 with the other variables.

25

1 ***Training, testing, and validation samples***

2 We trained a prediction model for ΔFEV_1 in 4,496 subjects with data from
3 COPDGene Visit 1 and Visit 2 using a nested, 10-fold cross validation (CV) procedure. The
4 inner fold of CV was used for parameter tuning. In the outer fold, our studied derivation
5 cohort was randomly partitioned into ten mutually exclusive subsets (folds) of approximately
6 equal size, using nine folds for training and one fold for testing each time for ten times. This
7 entire procedure was repeated five times to account for the random variability of the
8 partitioning procedure and provide more accurate estimates of the performance. This repeated
9 resampling procedure created an ensemble of fifty models over which we averaged the
10 predictions, and we then validated the performance of this model using data from COPDGene
11 Visit 3 that had not been used in any aspect of the model training process (temporal
12 validation).

13

14 ***Random forest supervised machine learning***

15 Supervised random forest is an ensemble learning method that predicts outcomes by
16 fitting a series of decision trees and aggregating the results across trees. This method can
17 capture non-linear dependencies and has been shown to perform well for a range of tasks¹⁴.
18 It begins building each tree by randomly selecting participants for the tree with replacement
19 (bootstrap samples). Participants not selected in bootstrapping represent the out-of-bag set. For
20 each bootstrap sample, a decision tree is trained by recursive binary partition of the data until
21 the minimum node size is reached. At each node split, an optimal feature (and its split-point)
22 is identified from a randomly selected subset of features by minimizing a loss measure. The
23 random selection of features reduces the correlation between trees, leading to variance
24 reduction and improved generalization performance. It also allows a moderately informative
25 feature to assert its importance to the prediction. Once an ensemble of trees are grown, the

1 prediction for a new sample is made by aggregating predictions (e.g. averaging for regression
2 and majority vote for classification) from individual trees. In our study, we fixed the number
3 of trees at 500 and tuned the hyperparameters (the bootstrap sampling fraction, the minimal
4 node size and the number of features to use at each split) by minimizing root mean squared
5 error (RMSE) using a nested 10-fold cross-validation within the training data.

6

7 ***Random forest variable importance and their effects on the prediction***

8 We calculated variable importance scores as the aggregated increase in the mean
9 squared errors (IncMSE) of predictions estimated with out-of-bag samples when the values of
10 a given variable are randomly permuted^{15,16}. The larger the increase in prediction error when
11 permuted, the higher the variable importance score (IncMSE), and the more important the
12 variable is to the prediction. Since the “raw” permutation importance has better statistical
13 properties, the importance values were not normalized¹⁷. Therefore, they cannot be used to
14 compare variable importance across prediction tasks, but they can be used within the same
15 prediction task to rank variables by their contribution to the accuracy of the final model.

16

17 ***Prediction performance***

18 We assessed the accuracy of each prediction model using the RMSE and R-squared
19 metrics, indicators of the goodness of fit of a set of predictions to the observed values. For the
20 prediction of ΔFEV_1 , we also assessed the ability of the models to correctly identify subjects
21 in the top quartile of disease progression (i.e., greatest decline in FEV_1) as quantified by the
22 AUCs (areas under the receiver operator characteristic (ROC) curves).

23

24 ***Statistical analyses***

1 We performed a complete case analysis. Descriptive characteristics were reported
2 respectively as percentages and medians with interquartile ranges for categorical and
3 continuous variables. Variables were analyzed using the t-test for normally distributed
4 variables, the Wilcoxon rank sum test for non-normally distributed variables, and chi-square
5 tests for categorical variables. To identify differences in the quality of prediction and variable
6 importance in subjects with different levels of COPD severity, we also constructed prediction
7 models separately in various GOLD subgroups. All tests of significance were two-tailed with
8 a significance threshold of P-value < 0.05.

9

10 **RESULTS**

11

12 ***Subject characteristics***

13 In total, 4,496 COPDGene subjects (median age: 60; 51% men; 73% NHW) had
14 complete phenotypic data and were included in the analysis. The participant flow diagram is
15 shown in Figure 1.

16

17 **Characteristics of “rapid FEV₁ progressors” in COPDGene**

18 To determine the characteristics of subjects who were “rapid FEV₁ progressors” in
19 COPDGene, we examined the characteristics of subjects in the top quartile of progression to
20 those in the bottom quartile (Table 1). Compared to subjects in the bottom quartile of Δ FEV₁,
21 those in the top quartile (“rapid FEV₁ progressors”) had a higher proportion of males with
22 less severe spirometric impairment at baseline but with higher exposure to smoking (pack-
23 years and percent of current smoking), more advanced radiographic disease (total
24 emphysema and gas trapping), more bronchodilator responsiveness, more dyspnea and
25 chronic bronchitis symptoms, and a lower rate of obesity and metabolic syndrome.

1 The median change in FEV₁ was -37 (IQR: -66, -9) mL/year (Figure 3). Fifty-seven
2 percent of the studied subjects had a rate of decline in FEV₁ of more than 30 mL/year over
3 the 5-year period and 7% had an increase in FEV₁ of more than 30 mL/year. Rapid FEV₁
4 progressors had a median change of -91 mL/year compared to 11 mL/year for slow
5 spirometric progressors (Table 1). When assessed according to the severity of airflow
6 limitation, the rate of FEV₁ decline was inversely related to the GOLD grade, with medians
7 of Δ FEV₁ of -46, -38, -31, -16 mL/year for GOLD 1-4, respectively.

8

9 ***Prediction performance for follow-up FEV₁ and 5-year change in FEV₁***

10 We constructed the prediction models using a nested cross-validation procedure and
11 we assessed the prediction performance in the COPDGene 10-year follow-up visit. A
12 schematic representation of our model is shown in Figure 2. The list of candidate predictors
13 is provided in Table 2. In the cross-validation testing samples, on average, 89.6% of the
14 variance in follow-up FEV₁ values were explained and the area under the ROC curves for the
15 prediction of subjects in the top quartile of observed disease progression was 0.97 (Table 3
16 and Figure 4). This high performance was maintained in the temporal validation with an R-
17 squared value of 0.91 and AUC of 0.98 (Table 3). For the prediction of the change in FEV₁
18 over time (Δ FEV₁), the average R-squared value was 0.15 and AUC was 0.71 in the testing
19 samples and respectively, 0.10 and 0.70 in the validation cohort.

20

21 ***Analysis of signal to noise ratio for 5-year change in FEV₁***

22 Changes in spirometric measures are more commonly used endpoints in COPD
23 clinical trials. Predicting future FEV₁ values is not the same as predicting the changes of
24 FEV₁ over the same period, since the Δ FEV₁ over a fixed time period generally contributes a
25 relatively small amount to the overall variance of FEV₁ at a given time point. Given the often

1 gradual rate of progression of COPD, five-years is a relatively short observation period, and
2 one of the concerns is that the signal to noise ratio in our progression variables is insufficient
3 for reliable prediction. To determine the signal-to-noise characteristics of our progression
4 variables, we calculated the expected signal-to-noise ratio using previously published values
5 of measurement error for FEV₁¹⁸. An important parameter in these calculations is the extent
6 of correlation in errors between the two study measurements. Since empiric data were
7 unavailable, we assumed independence between these errors; therefore, these estimates likely
8 represent a lower bound on the proportion of noise in these measures. We estimated that
9 measurement error accounted for at least 22% of the variance of Δ FEV₁ (calculations are
10 included in the supplement). Thus, the theoretical upper bound for prediction performance of
11 Δ FEV₁ was 78%.

12

13 ***Important predictors and their effects on prediction***

14 Figure 5 shows the ranking of the top-20 predictors based on their importance scores
15 in the random forest models. Several of the known COPD disease progression risk factors
16 were present as top-ranked risk factors in our models and other new predictors were
17 identified. The most important variables for FEV₁ progression included baseline spirometry,
18 CT-measured total lung volume, bronchodilator responsiveness, gas trapping, total
19 emphysema, and smoking exposure. Variables like the number of COPD exacerbations in the
20 prior year, selected comorbidities, and dyspnea scores were of less importance.

21

22 ***Prediction of COPD progression stratified by GOLD grade***

23 To determine whether progression was determined by different variables at different
24 GOLD spirometric grades, we examined the performance of random forest prediction models
25 for pre-specified subgroups of smokers stratified by GOLD grade (n= 4,496 (*Overall*), 499

1 (PRISm), 2,116 (GOLD 0), 1,318 (GOLD 1-2), and 563 (GOLD 3-4)). We observed
2 significant differences in predictive performance across these subgroups. The model
3 performance accuracy was best for GOLD 1-2 subjects and it was harder to achieve accurate
4 prediction in advanced stages of the disease. The area under the ROC curves for the
5 prediction of subjects in the top quartile of disease progression was 0.66 (GOLD 0), 0.73
6 (GOLD 1-2), and 0.58 (GOLD 3-4). The predictors of disease progression were also different
7 by GOLD grade (Figure 5). For instance, bronchodilator responsiveness seems to be less
8 important and emphysema and airway disease more important in the prediction of ΔFEV_1 in
9 subjects at more advanced stages of the disease.

10

11 **DISCUSSION**

12 This current study showed that the prediction of change in FEV_1 , which is more
13 relevant for disease progression, is more challenging than predicting absolute FEV_1 level.
14 Our prediction models for ΔFEV_1 represent the current state of the art for prediction of
15 prospective change in FEV_1 , but there is significant room for improvement in future models.
16 The most important predictive variables came from a wide range of clinical, spirometric, and
17 imaging features. Baseline spirometry, CT-measured total lung volumes, and bronchodilator
18 responsiveness dominated the prediction. In addition, the predictive performance and the
19 relative importance of predictors differed by GOLD grade.

20 Several screening tools are available to identify patients with undiagnosed COPD and
21 to predict outcomes in patients with COPD^{1,8,9,19-24}. While Zafari et al. and Chen et al.
22 developed and validated risk models to accurately predict lung function trajectory^{8,9}, our
23 study is the first to apply advanced machine learning methods, use an extensive set of
24 phenotypic measurements and comorbidities, predict not only the follow-up values but also
25 the more relevant “change” variables, and identify the relative importance of the predictors at

1 various stages of the disease. With respect to the outcomes evaluated in these two papers, our
2 predictive models gave similar performance for the prediction of future values of FEV₁. Our
3 study added the prediction of prospective changes in FEV₁ that were not reported in these
4 previously published studies. Predicting the change over time is more challenging than
5 simply predicting the future value, since the change typically represents a small proportion of
6 the overall variance in a given pair of FEV₁ measurements separated by five years or less.
7 However, it is important to assess the ability of models to predict prospective changes since
8 this is an important outcome for clinical trials.

9 The predictive accuracy of our models may potentially be further improved by
10 including additional predictors (such as DLCO, pulmonary vascular measures, and relevant
11 molecular biomarkers) and exploring other machine learning algorithms (such as deep
12 learning). At present, these models are not ready for clinical use but could be useful in the
13 design of COPD clinical trials to enrich the study populations by patients who are most likely
14 to experience rapid disease progression and benefit from therapeutic interventions. For
15 clinical use, better performing models that have been more extensively validated in multiple
16 additional and relevant target populations are necessary.

17 Rapid decline in lung function has previously been associated with a range of factors
18 such as smoking exposure, bronchodilator reversibility, higher baseline FEV₁, higher baseline
19 FVC, exacerbations in the prior year, low BMI, African American race, female sex,
20 emphysema, upper lobe emphysema predominance, and CT-detected small airway
21 abnormalities^{5,6,8,25-30}. Our study detected several of these known COPD disease progression
22 risk factors and identified other new predictors for FEV₁ decline. Our study is the first to our
23 knowledge to demonstrate that the patterns of predictors vary by GOLD spirometric grade.
24 The intriguing variations in the importance of different risk factors depending on the studied

1 subgroup may help inform further exploration of predictive risk factors and future
2 development of new risk prediction algorithms.

3 The relative unimportance of certain traditional risk factors such as COPD
4 exacerbations in the prior year, selected comorbidities, race, and sex in our machine learning
5 predictive models may be consistent with the disparate results from previous studies. For
6 example, although some publications have suggested a significant excess loss of FEV₁ for
7 each COPD exacerbation^{26,31,32}, others have reported minimal⁶ or no relationship³³. Such
8 discrepancy may also result from differences in methodology between studies as well as
9 differences sample size, study duration, study population, and variable definitions. The
10 relative unimportance of certain traditional risk factors in our models may also indicate that,
11 while these risk factors may attain statistical significance in some models, they do not
12 provide much additional predictive value after considering more important risk factors.

13 This study has a number of strengths. Analyses were performed within a well-
14 characterized cohort that included subjects at all stages of disease severity. In addition, by
15 focusing on prediction rather than the study of individual risk factors, our results provide
16 useful context regarding the relative importance of specific predictors. By constructing
17 models in subjects stratified by GOLD spirometric grade, we demonstrated that patterns of
18 optimal predictors vary by specific disease outcome and GOLD grade. Validation of our
19 findings in the temporal cohort represents another strength of our paper.

20 Our study also has limitations. We only used two measurements of lung function
21 separated by approximately 5 years. The large sample size available helped to overcome
22 some of the inherent challenges in low signal-to-noise ratio with studies of COPD
23 progression over a relatively short period of time. However, with longer follow-up and more
24 measurements, we will be better able to isolate measurement noise from real disease
25 progression which will result in greater predictive accuracy. Our analysis was based on

1 subjects who had completed their second study visit, and it is possible that patients who were
2 lost to follow-up differed from those available for analysis. Many of the patients with airflow
3 obstruction were receiving therapy for their disease. Although no existing pharmacotherapy
4 has been conclusively shown to affect the rates of disease progression, this still may have
5 influenced our results. However, we chose not to include pharmacotherapy data in these
6 analyses in order to reduce biases likely present in patient-reported pharmaco-epidemiologic
7 data ^{34,35}. Lastly, it is recognized that as the number of potential risk factors increases, the
8 complexity of the models can cause overfitting. We addressed this by appropriate
9 hyperparameter tuning and by evaluating the performance of our predictive models in cross-
10 validation and in the temporal cohort.

11 Random forest machine learning in conjunction with deep phenotyping improves the
12 prediction accuracy of COPD progression. The present study improves our ability to identify
13 patients at risk for rapid disease progression, and these models may be useful for the
14 development of targeted disease-modifying therapies.

15
16
17
18
19
20
21
22
23
24
25

FIGURE LEGENDS

1

2

3 **Figure 1.** Participants' flow diagram and general framework of the study.

4 **Figure 2.** Random forest modeling framework.

5 **Figure 3.** Histograms of change in FEV₁ (Δ FEV₁; mL/year) between Visit 1 and Visit 2 and

6 between Visit 2 and Visit 3.

7 **Figure 4.** Receiver operator characteristic (ROC) curves of the performance of the random

8 forest follow-up FEV₁ and 5-year change in FEV₁ models in correctly identifying subjects in

9 the top quartiles of spirometric progression in the COPDGene Visit 1 / Visit 2 cross-

10 validation testing samples. Solid lines represent the average performance, and colored dots

11 represent the performance in each of the sampling iterations.

12 **Figure 5.** Heatmaps of the top-20 predictors of Visit 2 FEV₁ (mL) (A) and change in FEV₁

13 (mL/year) (B). The x-axis contains the group assignments (All, PRISm, GOLD0, GOLD 1-2,

14 and GOLD 3-4). The y-axis includes the top-20 predictors ranked by their importance scores

15 in the predictive models built in the "All" group (decreasing order with the best predictors on

16 top). Darker shades of blue indicate a higher rank of the predictor. White cells indicate

17 variables that do not fall within the top-20 ranks. The sample sizes were (n= 4,496 (*All*), 499

18 (*PRISm*), 2,116 (*GOLD 0*), 1,318 (*GOLD 1-2*), and 563 (*GOLD 3-4*)). Abbreviations: BMI:

19 Body mass index; FEV₁: Forced expiratory volume in 1 second; FEF₂₅₋₇₅: Forced

20 expiratory flow at 25–75% of forced vital capacity (FVC); Bronchodilator responsiveness

21 (%) FEV₁: Percentage of subjects with post-bronchodilator increase in FEV₁ of at least 12%

22 from baseline; Bronchodilator responsiveness (%) FVC: Percentage of subjects with post-

23 bronchodilator increase in FVC of at least 12% from baseline; GOLD: Global Initiative for

24 Chronic Obstructive Lung Disease; SGRQ: St. George's Respiratory Questionnaire; MMRC:

25 Modified Medical Research Council; Adjusted Perc15 density: Cut off value in Hounsfield

1 units (HU) below which 15% of all voxels are distributed on a lung CT scan (per convention,
2 adjusted Perc15 density values are reported as the HU + 1000); Gas trapping (%): Percentage
3 of lung voxels with a density less than -856 HU at end exhalation; Pi10: Square root of the
4 wall area of a hypothetical airway of a 10-mm internal perimeter; % Segmental airway wall
5 thickness: Percentage of the wall relative to the total bronchial area for the segmental
6 airways.

7

8 **Acknowledgements:** The authors wish to thank the thousands of patients who participated
9 in the COPDGene study over the last 10 years.

10

11 **Author Contributions:** Drs. Boueiz and Castaldi had full access to all the data in the study
12 and take responsibility for the integrity of the data and the accuracy of the data analysis.
13 COPDGene investigators were instrumental in the design and implementation of the
14 COPDGene study and collected and analyzed data cited in this article. All authors have
15 reviewed, approved and endorsed all content and conclusions of this article.

16

17 **Declaration of Interest:** The COPDGene study is funded by National Heart, Lung, and
18 Blood Institute grants U01 HL089897 and U01 HL089856. The COPDGene study
19 (NCT00608764) is also supported by the COPD Foundation through contributions made to
20 an Industry Advisory Board that has included AstraZeneca, Bayer Pharmaceuticals,
21 Boehringer-Ingelheim, Genentech, GlaxoSmithKline, Novartis, Pfizer, and Sunovion. While
22 some individual authors of this manuscript were employed by one of the listed funders at the
23 time the work of this study was conducted, these employment relationships did not constitute
24 undue influence by funders. These funders have had no official role in the collection,
25 management, analysis and interpretation of the data or design and conduct of the study. All

1 authors have completed a Conflict of Interest form, disclosing any real or apparent financial
2 relationships including receiving royalties, honoraria or fees for consulting, lectures,
3 speakers' bureaus, continuing education, medical advisory boards or expert testimony;
4 receipt of grants; travel reimbursement; direct employment compensation. These disclosure
5 forms have been filed with the Chronic Obstructive Pulmonary Diseases: Journal of the
6 COPD Foundation Editorial Office and are available for review, upon request at COPDC@
7 njhealth.org.

REFERENCES

1. Guo YI, Qian Y, Gong YI, Pan C, Shi G, Wan H. A predictive model for the development of chronic obstructive pulmonary disease. *Biomed Rep.* 2015;3(6):853-863.
2. Jemal A, Ward E, Hao Y, Thun M. Trends in the leading causes of death in the United States, 1970-2002. *JAMA.* 2005;294(10):1255-1259.
3. Lozano R, Naghavi M, Foreman K, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet.* 2012;380(9859):2095-2128.
4. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Med.* 2006;3(11):e442.
5. Bhatt SP, Soler X, Wang X, et al. Association between Functional Small Airway Disease and FEV1 Decline in Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med.* 2016;194(2):178-184.
6. Vestbo J, Edwards LD, Scanlon PD, et al. Changes in forced expiratory volume in 1 second over time in COPD. *N Engl J Med.* 2011;365(13):1184-1192.
7. Vestbo J, Lange P. Natural history of COPD: Focusing on change in FEV1. *Respirology.* 2016;21(1):34-43.
8. Zafari Z, Sin DD, Postma DS, et al. Individualized prediction of lung-function decline in chronic obstructive pulmonary disease. *CMAJ.* 2016;188(14):1004-1011.
9. Chen W, Sin DD, FitzGerald JM, Safari A, Adibi A, Sadatsafavi M. An Individualized Prediction Model for Long-term Lung Function Trajectory and Risk of COPD in the General Population. *Chest.* 2019.
10. Han MK, Agusti A, Calverley PM, et al. Chronic obstructive pulmonary disease phenotypes: the future of COPD. *Am J Respir Crit Care Med.* 2010;182(5):598-604.

11. Lange P, Celli B, Agusti A, et al. Lung-Function Trajectories Leading to Chronic Obstructive Pulmonary Disease. *N Engl J Med*. 2015;373(2):111-122.
12. Martinez FD. Early-Life Origins of Chronic Obstructive Pulmonary Disease. *N Engl J Med*. 2016;375(9):871-878.
13. Regan EA, Hokanson JE, Murphy JR, et al. Genetic epidemiology of COPD (COPDGene) study design. *COPD*. 2010;7(1):32-43.
14. Touw WG, Bayjanov JR, Overmars L, et al. Data mining in the Life Sciences with Random Forest: a walk in the park or lost in the jungle? *Brief Bioinform*. 2013;14(3):315-326.
15. Svetnik V, Liaw A, Tong C, Culberson JC, Sheridan RP, Feuston BP. Random forest: a classification and regression tool for compound classification and QSAR modeling. *J Chem Inf Comput Sci*. 2003;43(6):1947-1958.
16. Breiman L. Random forests. *Mach Learn* 2001;45(1):5-32.
17. Strobl C, Boulesteix AL, Kneib T, Augustin T, Zeileis A. Conditional variable importance for random forests. *BMC Bioinformatics*. 2008;9:307.
18. Tweeddale PM, Alexander F, McHardy GJ. Short term variability in FEV1 and bronchodilator responsiveness in patients with obstructive ventilatory defects. *Thorax*. 1987;42(7):487-490.
19. Han MK, Steenrod AW, Bacci ED, et al. Identifying Patients with Undiagnosed COPD in Primary Care Settings: Insight from Screening Tools and Epidemiologic Studies. *Chronic Obstr Pulm Dis*. 2015;2(2):103-121.
20. Higgins MW, Keller JB, Becker M, et al. An index of risk for obstructive airways disease. *Am Rev Respir Dis*. 1982;125(2):144-151.

21. Himes BE, Dai Y, Kohane IS, Weiss ST, Ramoni MF. Prediction of chronic obstructive pulmonary disease (COPD) in asthma patients using electronic medical records. *J Am Med Inform Assoc.* 2009;16(3):371-379.
22. Kotz D, Simpson CR, Viechtbauer W, van Schayck OC, Sheikh A. Development and validation of a model to predict the 10-year risk of general practitioner-recorded COPD. *NPJ Prim Care Respir Med.* 2014;24:14011.
23. Matheson MC, Bowatte G, Perret JL, et al. Prediction models for the development of COPD: a systematic review. *Int J Chron Obstruct Pulmon Dis.* 2018;13:1927-1935.
24. Bellou V, Belbasis L, Konstantinidis AK, Tzoulaki I, Evangelou E. Prognostic models for outcome prediction in patients with chronic obstructive pulmonary disease: systematic review and critical appraisal. *BMJ.* 2019;367:l5358.
25. Casanova C, de Torres JP, Aguirre-Jaime A, et al. The progression of chronic obstructive pulmonary disease is heterogeneous: the experience of the BODE cohort. *Am J Respir Crit Care Med.* 2011;184(9):1015-1021.
26. Dransfield MT, Kunisaki KM, Strand MJ, et al. Acute Exacerbations and Lung Function Loss in Smokers with and without Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med.* 2017;195(3):324-330.
27. Hanrahan JP, Tager IB, Segal MR, et al. The effect of maternal smoking during pregnancy on early infant lung function. *Am Rev Respir Dis.* 1992;145(5):1129-1135.
28. Mohamed Hoesein FA, van Rikxoort E, van Ginneken B, et al. Computed tomography-quantified emphysema distribution is associated with lung function decline. *Eur Respir J.* 2012;40(4):844-850.
29. Nishimura M, Makita H, Nagai K, et al. Annual change in pulmonary function and clinical phenotype in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med.* 2012;185(1):44-52.

30. Sun Y, Milne S, Jaw JE, et al. BMI is associated with FEV1 decline in chronic obstructive pulmonary disease: a meta-analysis of clinical trials. *Respir Res.* 2019;20(1):236.
31. Donaldson GC, Seemungal TA, Bhowmik A, Wedzicha JA. Relationship between exacerbation frequency and lung function decline in chronic obstructive pulmonary disease. *Thorax.* 2002;57(10):847-852.
32. Kanner RE, Anthonisen NR, Connett JE, Lung Health Study Research G. Lower respiratory illnesses promote FEV(1) decline in current smokers but not ex-smokers with mild chronic obstructive pulmonary disease: results from the lung health study. *Am J Respir Crit Care Med.* 2001;164(3):358-364.
33. Suzuki M, Makita H, Ito YM, et al. Clinical features and determinants of COPD exacerbation in the Hokkaido COPD cohort study. *Eur Respir J.* 2014;43(5):1289-1297.
34. Suissa S. Immortal time bias in pharmaco-epidemiology. *Am J Epidemiol.* 2008;167(4):492-499.
35. Wise L. Risks and benefits of (pharmaco)epidemiology. *Ther Adv Drug Saf.* 2011;2(3):95-102.

Table 1. Characteristics of the rapid spirometric progressors.

	Top quartile progressors (n = 1,124)	Bottom quartile progressors (n = 1,124)	P-value	
Age (years)	58.3 [51.9, 65.0]	59.7 [52.3, 66.6]	<0.001	
Male (%)	66.0%	51.2%	<0.001	
Non-Hispanic whites (%)	72.3%	69.6%	0.098	
Height (cm)	173.0 [166.7, 180.0]	170.0 [162.6, 177.5]	<0.001	
BMI	27.7 [24.5, 31.7]	29.3 [25.6, 33.9]	<0.001	
Pack-years of smoking	40.5 [29.9, 57.0]	38.0 [25.7, 52.5]	<0.001	
Current smoking (%)	56.1%	45.2%	<0.001	
Total emphysema (%LAA-950)	2.7 [0.7, 7.2]	1.7 [0.5, 5.0]	0.003	
U/L ratio	0.38 [0.00, 0.91]	0.41 [0.00, 0.92]	0.43	
Airway wall thickening (%)	49.6 [44.3, 55.2]	50.1 [44.6, 56.3]	0.09	
Pi10	2.1 [1.8, 2.5]	2.2 [1.9, 2.6]	0.01	
Gas trapping (%)	15.9 [7.6, 30.3]	12.3 [5.5, 25.0]	<0.001	
FEV ₁ (percent predicted)	86.4 [68.9, 99.7]	78.9 [60.9, 91.5]	<0.001	
FEV ₁ /FVC	0.71 [0.61, 0.79]	0.72 [0.62, 0.79]	<0.001	
Pre/Post- bronchodilator FEV ₁ (% change)	5.5 [1.8, 11.0]	2.8 [-1.4, 7.9]	<0.001	
Pre/Post- bronchodilator FVC (% change)	3.1 [-1.0, 8.8]	1.0 [-3.6, 6.7]	0.04	
MMRC dyspnea score				
	0	563 (50.1%)	529 (47.1%)	0.37
	1	171 (15.2%)	177 (15.7%)	
	2	143 (12.7%)	137 (12.2%)	
	3	166 (14.8%)	199 (17.7%)	
	4	81 (7.2%)	82 (7.3%)	
SGRQ score	31.8 ± 25.9	29.3 ± 24.5	0.03	
Chronic bronchitis (%)	21.1%	16.4%	0.004	
Metabolic syndrome (%)	16.0%	20.3%	0.01	
Obesity (%)	35.1%	45.0%	<0.001	

GOLD:	PRISm	75 (6.7%)	194 (17.3%)	<i><0.001</i>
	0	546 (48.6%)	468 (41.6%)	
	1	138 (12.3%)	65 (5.8%)	
	2	272 (24.2%)	221 (19.7%)	
	3	90 (8.0%)	138 (12.3%)	
	4	3 (0.3%)	38 (3.4%)	
5-year change in FEV ₁ (mL/year)		-91.0 [-117.0, -77.0]	11.0 [0.0, 32.0]	<i><0.001</i>

All values are from Visit 1.

BMI: Body mass index; FEV₁: Forced expiratory volume in 1 second; FVC: Forced vital capacity; MMRC: Modified Medical Research Council; SGRQ: St. George's Respiratory Questionnaire; Exacerbation frequency: Percent of subjects reporting at least one COPD exacerbation in the previous year; Metabolic syndrome: 3 of 4: BMI \geq 30 (measured), diabetes mellitus, hypertension, and high cholesterol (all self-report); Obesity: BMI \geq 30; GOLD: Global Initiative for Chronic Obstructive Lung Disease; PRISm: Preserved Ratio Impaired Spirometry.

Emphysema is defined as percent of CT low attenuation area below -950 Hounsfield units (HU) at end-inspiration using Thirona software (% LAA-950); U/L ratio: Ratio of %LAA-950 in upper lung third to %LAA-950 in lower lung third; Airway wall area percent is the percentage of the wall area compared with the total bronchial area for segmental airways; Pi10: Square root of the wall area of a hypothetical airway of 10-mm internal perimeter. "Change between Visit 1 and Visit 2 per year" variables are defined as (Value at Visit 2 - Value at Visit 1) / Time between Visit 1 and Visit 2 in years.

Variables are expressed as mean and standard deviation for continuous normally distributed variables, median and interquartile range (25th to 75th percentile) for continuous non-normally distributed variables, and percentages for categorical variables. P-values are obtained using t-test for the continuous normally distributed variables, Wilcoxon rank sum test for the continuous non-normally distributed variables, and chi-square test for the proportions. P-values < 0.05 are bolded and italicized.

Table 2. Variables included in the prediction algorithms.

<p><u>Demographics:</u></p> <ul style="list-style-type: none"> • Age at study enrollment • Sex • Race • Body mass index (BMI) • Height • Pack-years of smoking • Current smoking • Age at smoking initiation
<p><u>Family history:</u></p> <ul style="list-style-type: none"> • Family history of COPD, chronic bronchitis, or emphysema
<p><u>Functional measures:</u></p> <ul style="list-style-type: none"> • Modified Medical Research Council (MMRC) dyspnea scale • St. George's Respiratory Questionnaire (SGRQ) • 6-minute walk distance
<p><u>COPD characteristics:</u></p> <ul style="list-style-type: none"> • Chronic bronchitis (Chronic cough and phlegm for ≥ 3 months/year for at least 2 consecutive years) • Blue Bloater (Chronic bronchitis, BMI ≥ 25, Resting oxygen saturation $< 90\%$) • Pink puffer (Emphysema $> 10\%$, BMI ≤ 20, Resting oxygen saturation $\geq 90\%$) • Number of COPD exacerbations over the prior year (Number of self-reported acute worsening of respiratory symptoms that required the use of antibiotics and/or systemic steroids in the previous year) • History of severe COPD exacerbation (Self-report of COPD exacerbation requiring an emergency department visit or

- hospital admission)
- Need for courses of systemic steroids
 - Poor exercise capacity (6-minute walk distance < 500 feet)
 - Hypoxemia (Resting oxygen saturation \leq 88%)
 - Severe early-onset COPD (Age \geq 55 years, FEV₁ $<$ 50% predicted)

Comorbidities:

- Diabetes mellitus (Self-report)
- Hypertension (Self-report)
- Dyslipidemia (Self-report)
- Pneumothorax (Self-report)
- Gastro-esophageal reflux disease (Self-report)
- Osteoporosis (Self-report)
- Coronary artery disease (Self-report of heart attack, coronary artery disease, angina, angioplasty, or coronary artery bypass graft)
- Congestive heart failure (Self-report)
- Peripheral vascular disease (Self-report)
- Metabolic syndrome (3 of 4: BMI \geq 30 (measured), self-reported diabetes mellitus, hypertension, and high cholesterol)
- Physician diagnosis of asthma before age 40 (Self-report)
- Asthma/COPD overlap (Self-report)
- Obstructive sleep apnea (Self-report)

Spirometry:

- Post-bronchodilator FEV₁
- Post-bronchodilator FVC
- FEV₁/FVC
- Post-bronchodilator FEF₂₅₋₇₅
- Pre/Post- bronchodilator FEV₁ (% change)
- Pre/Post- bronchodilator FVC (% change)

- GOLD

Radiology:

- Total emphysema (%LAA-950)
- Emphysema distribution (Upper over lower lung third %LAA-950 ratio)
- Gas trapping (Percentage of low attenuation area less than -856HU at end-expiration)
- CT-measured total lung volumes at end-inspiration
- Airway wall thickness (Obtained along the center line of the lumen, in the middle third of the airway segment, for one segmental airway of each lung lobe)
- Pi10 (Square root of the wall area of a hypothetical airway of 10-mm internal perimeter)

Table 3. Prediction performance of random forest in the cross-validation testing samples and temporal validation cohort.

	COPDGene Visit 1 / Visit 2 testing samples	COPDGene Visit 2 / Visit 3 temporal validation cohort
RMSE		
Follow-up FEV ₁	269.711 [259.252, 276.476]	236.742
Change in FEV ₁ (mL/year)	46.913 [45.647, 48.795]	52.289
R-squared		
Follow-up FEV ₁	0.896 [0.890, 0.903]	0.913
Change in FEV ₁ (mL/year)	0.147 [0.126, 0.173]	0.104
AUC		
Follow-up FEV ₁	0.974 [0.970, 0.979]	0.975
Change in FEV ₁ (mL/year)	0.707 [0.688, 0.724]	0.704

The derivation cohort (COPDGene Study Visit 1 and Visit 2) was randomly partitioned into training and testing samples using 10-fold cross validation. This procedure was repeated five times to account for the random variability of the partitioning procedure. This repeated resampling procedure created an ensemble of fifty models over which we averaged the predictions, and we then validated the performance of this model using data from COPDGene Visit 3 (temporal validation). To predict the outcome values at Year 10 (Visit 3), we entered the subjects' 5-year (Visit 2) predictor data into the models trained in the derivation cohort.

Variables are expressed as median and interquartile range (IQR) (25th to 75th percentile) when applicable.

AUC: Area under the ROC curve for prediction of subjects in the top quartile of COPD progression; FEV₁: Forced expiratory volume in one second; RMSE: Root mean square error.

ACKNOWLEDGEMENTS

COPDGene Investigators - Core Units:

Administrative Center: James D. Crapo, MD (PI); Edwin K. Silverman, MD, PhD (PI); Barry J. Make, MD; Elizabeth A. Regan, MD, PhD

Genetic Analysis Center: Terri H. Beaty, PhD; Peter J. Castaldi, MD, MSc; Michael H. Cho, MD, MPH; Dawn L. DeMeo, MD, MPH; Adel Boueiz, MD, MMSc; Marilyn G. Foreman, MD, MS; Auyon Ghosh, MD; Lystra P. Hayden, MD, MMSc; Craig P. Hersh, MD, MPH; Jacqueline Hetmanski, MS; Brian D. Hobbs, MD, MMSc; John E. Hokanson, MPH, PhD; Wonji Kim, PhD; Nan Laird, PhD; Christoph Lange, PhD; Sharon M. Lutz, PhD; Merry-Lynn McDonald, PhD; Dmitry Prokopenko, PhD; Matthew Moll, MD, MPH; Jarrett Morrow, PhD; Dandi Qiao, PhD; Elizabeth A. Regan, MD, PhD; Aabida Saferali, PhD; Phuwanat Sakornsakolpat, MD; Edwin K. Silverman, MD, PhD; Emily S. Wan, MD; Jeong Yun, MD, MPH

Imaging Center: Juan Pablo Centeno; Jean-Paul Charbonnier, PhD; Harvey O. Coxson, PhD; Craig J. Galban, PhD; MeiLan K. Han, MD, MS; Eric A. Hoffman, Stephen Humphries, PhD; Francine L. Jacobson, MD, MPH; Philip F. Judy, PhD; Ella A. Kazerooni, MD; Alex Kluiber; David A. Lynch, MB; Pietro Nardelli, PhD; John D. Newell, Jr., MD; Aleena Notary; Andrea Oh, MD; Elizabeth A. Regan, MD, PhD; James C. Ross, PhD; Raul San Jose Estepar, PhD; Joyce Schroeder, MD; Jered Sieren; Berend C. Stoel, PhD; Juerg Tschirren, PhD; Edwin Van Beek, MD, PhD; Bram van Ginneken, PhD; Eva van Rikxoort, PhD; Gonzalo Vegas SanchezFerrero, PhD; Lucas Veitel; George R. Washko, MD; Carla G. Wilson, MS

PFT QA Center, Salt Lake City, UT: Robert Jensen, PhD

Data Coordinating Center and Biostatistics, National Jewish Health, Denver, CO: Douglas Everett, PhD; Jim Crooks, PhD; Katherine Pratte, PhD; Matt Strand, PhD; Carla G. Wilson, MS

Epidemiology Core, University of Colorado Anschutz Medical Campus, Aurora, CO: John E. Hokanson, MPH, PhD; Erin Austin, PhD; Gregory Kinney, MPH, PhD; Sharon M. Lutz, PhD; Kendra A. Young, PhD
Version Date: March 26, 2021

Mortality Adjudication Core: Surya P. Bhatt, MD; Jessica Bon, MD; Alejandro A. Diaz, MD, MPH; MeiLan K. Han, MD, MS; Barry Make, MD; Susan Murray, ScD; Elizabeth Regan, MD; Xavier Soler, MD; Carla G. Wilson, MS

Biomarker Core: Russell P. Bowler, MD, PhD; Katerina Kechris, PhD; Farnoush BanaeiKashani, PhD

COPD Gene Investigators - Clinical Centers:

Ann Arbor VA: Jeffrey L. Curtis, MD; Perry G. Pernicano, MD

Baylor College of Medicine, Houston, TX: Nicola Hania, MD, MS; Mustafa Atik, MD; Aladin Boriek, PhD; Kalpatha Guntupalli, MD; Elizabeth Guy, MD; Amit Parulekar, MD

Brigham and Women's Hospital, Boston, MA: Dawn L. DeMeo, MD, MPH; Craig Hersh, MD, MPH; Francine L. Jacobson, MD, MPH; George Washko, MD

Columbia University, New York, NY: R. Graham Barr, MD, DrPH; John Austin, MD; Belinda D'Souza, MD; Byron Thomashow, MD

Duke University Medical Center, Durham, NC: Neil MacIntyre, Jr., MD; H. Page McAdams, MD; Lacey Washington, MD

HealthPartners Research Institute, Minneapolis, MN: Charlene McEvoy, MD, MPH; Joseph Tashjian, MD

Johns Hopkins University, Baltimore, MD: Robert Wise, MD; Robert Brown, MD; Nadia N. Hansel, MD, MPH; Karen Horton, MD; Allison Lambert, MD, MHS; Nirupama Putcha, MD, MHS

Lundquist Institute for Biomedical Innovation at Harbor UCLA Medical Center, Torrance, CA: Richard Casaburi, PhD, MD; Alessandra Adami, PhD; Matthew Budoff, MD; Hans Fischer, MD; Janos Porszasz, MD, PhD; Harry Rossiter, PhD; William Stringer, MD

Michael E. DeBakey VAMC, Houston, TX: Amir Sharafkhaneh, MD, PhD; Charlie Lan, DO

Minneapolis VA: Christine Wendt, MD; Brian Bell, MD; Ken M. Kunisaki, MD, MS

Morehouse School of Medicine, Atlanta, GA: Eric L. Flenaugh, MD; Hirut Gebrekristos, PhD; Mario Ponce, MD; Silanath Terpenning, MD; Gloria Westney, MD, MS

National Jewish Health, Denver, CO: Russell Bowler, MD, PhD; David A. Lynch, MB

Reliant Medical Group, Worcester, MA: Richard Rosiello, MD; David Pace, MD

Temple University, Philadelphia, PA: Gerard Criner, MD; David Ciccolella, MD; Francis Cordova, MD; Chandra Dass, MD; Gilbert D'Alonzo, DO; Parag Desai, MD; Michael Jacobs, PharmD; Steven Kelsen, MD, PhD; Victor Kim, MD; A. James Mamary, MD; Nathaniel

Marchetti, DO; Aditi Satti, MD; Kartik Shenoy, MD; Robert M. Steiner, MD; Alex Swift, MD;
Irene Swift, MD; Maria Elena Vega-Sanchez, MD

University of Alabama, Birmingham, AL: Mark Dransfield, MD; William Bailey, MD; Surya P.
Bhatt, MD; Anand Iyer, MD; Hrudaya Nath, MD; J. Michael Wells, MD

University of California, San Diego, CA: Douglas Conrad, MD; Xavier Soler, MD, PhD; Andrew
Yen, MD

University of Iowa, Iowa City, IA: Alejandro P. Comellas, MD; Karin F. Hoth, PhD; John
Newell, Jr., MD; Brad Thompson, MD

University of Michigan, Ann Arbor, MI: MeiLan K. Han, MD MS; Ella Kazerooni, MD MS;
Wassim Labaki, MD MS; Craig Galban, PhD; Dharshan Vummidi, MD

University of Minnesota, Minneapolis, MN: Joanne Billings, MD; Abbie Begnaud, MD; Tadashi
Allen, MD

University of Pittsburgh, Pittsburgh, PA: Frank Sciorba, MD; Jessica Bon, MD; Divay Chandra,
MD, MSc; Joel Weissfeld, MD, MPH

University of Texas Health, San Antonio, San Antonio, TX: Antonio Anzueto, MD; Sandra
Adams, MD; Diego Maselli-Caceres, MD; Mario E. Ruiz, MD; Harjinder Singh

Figure 1

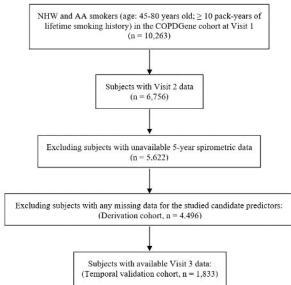


Figure 2

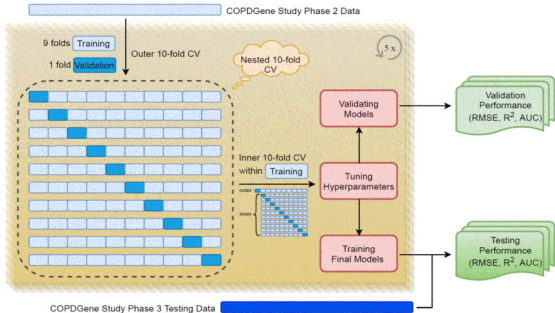
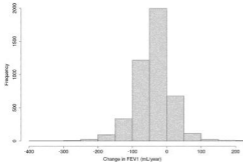


Figure 3

**Change in FEV1
between Visit 1 and Visit 2**



**Change in FEV1
between Visit 2 and Visit 3**

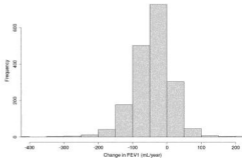
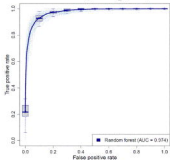


Figure 4

Follow-up FEV₁ (mL)



5-year change in FEV₁ (mL/year)

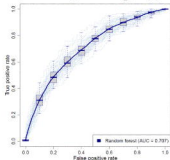


Figure 5

Follow-up FEV₁ (mL)

5-year change in FEV₁(mL/year)

