Supplemental Material

From Pre-COPD to COPD: a Simple, Low cost, and ease of IMplementation (SLIM) risk calculator

Miguel J. Divo MD MPH, Congjian Liu PhD, Francesca Polverino MD PhD, Peter J. Castaldi MD, Bartolome R. Celli MD and Yohannes Tesfaigzi PhD

Content:

Figure E1. Flow diagram showing the selection process for participants of the COPDGene® (validation cohort), meeting the inclusion criteria for the current analysis.

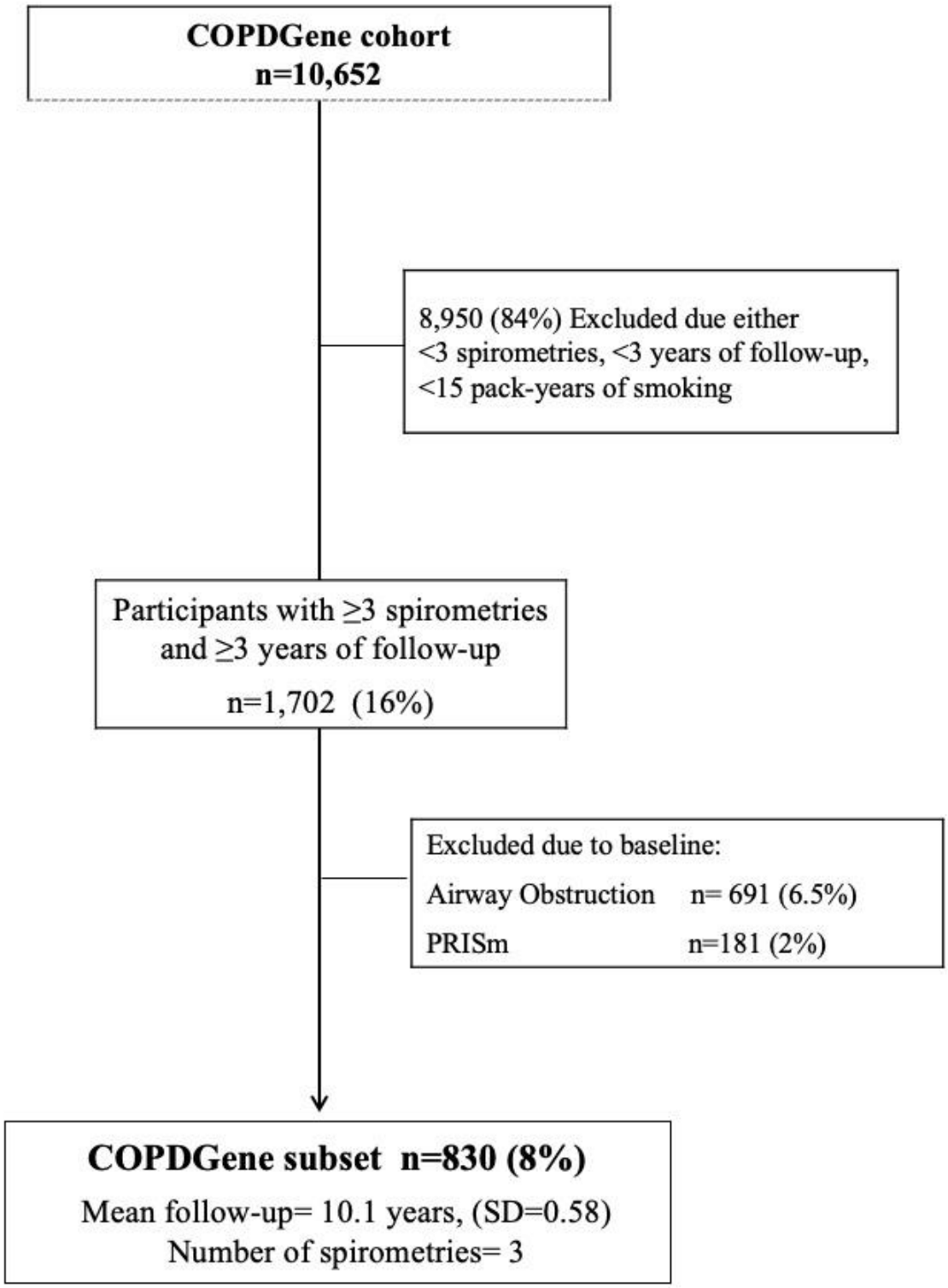


Figure E2. Flow diagram showing the selection process for participants in the Lovelace Smoker Cohort (LSC) meeting the inclusion criteria for the current analysis. Spirometric classification for normal, airway obstruction, and PRISm are based on criteria listed in Table E1-A.

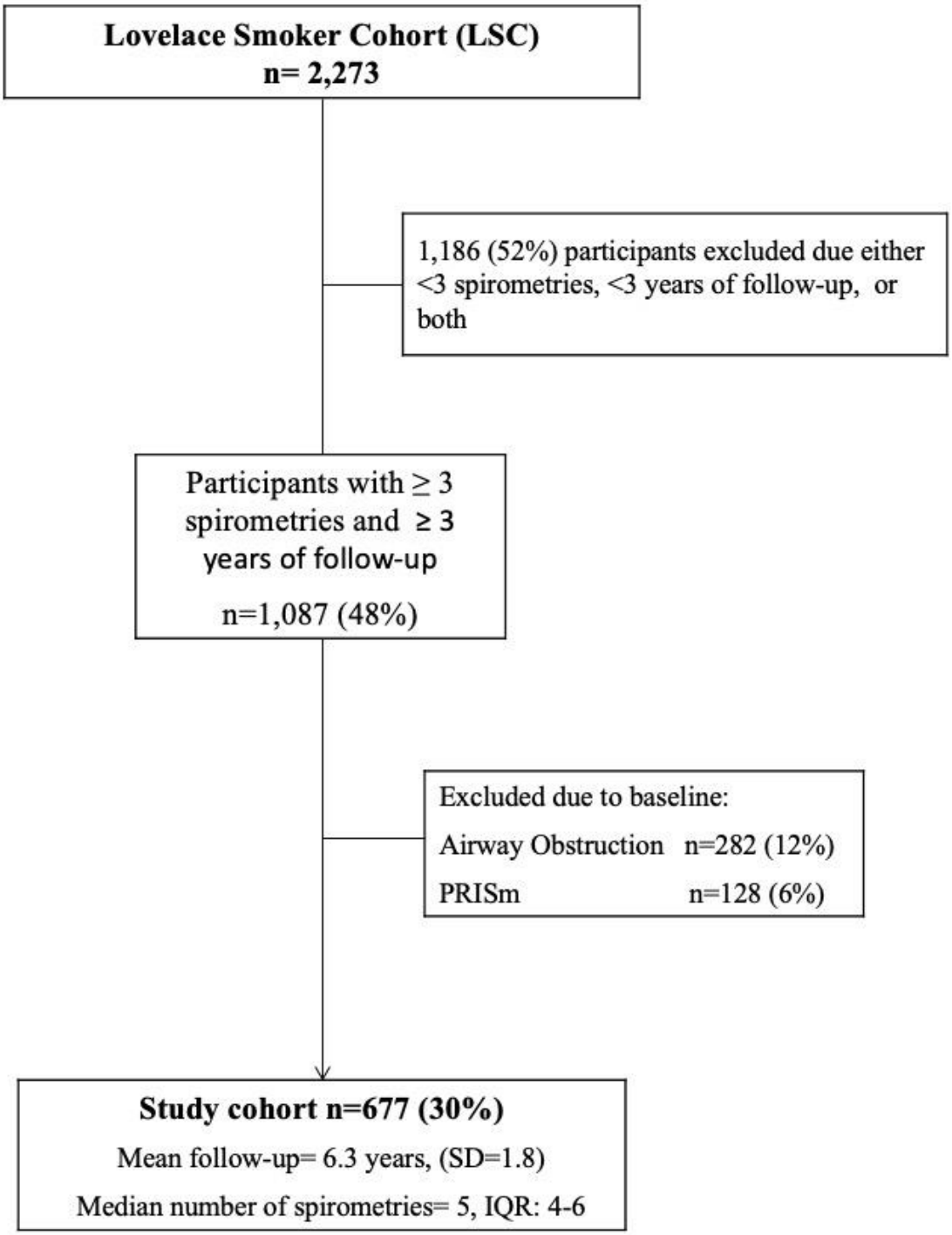


Figure E3. Scatterplot of the baseline FEV_1/FVC ratio vs. FEV_1 % predicted (GOLD classification) of the 677 subjects from the LSC study with normal baseline spirometry. The red dots represent the 110 participants who developed CAO at the end of the observation period.

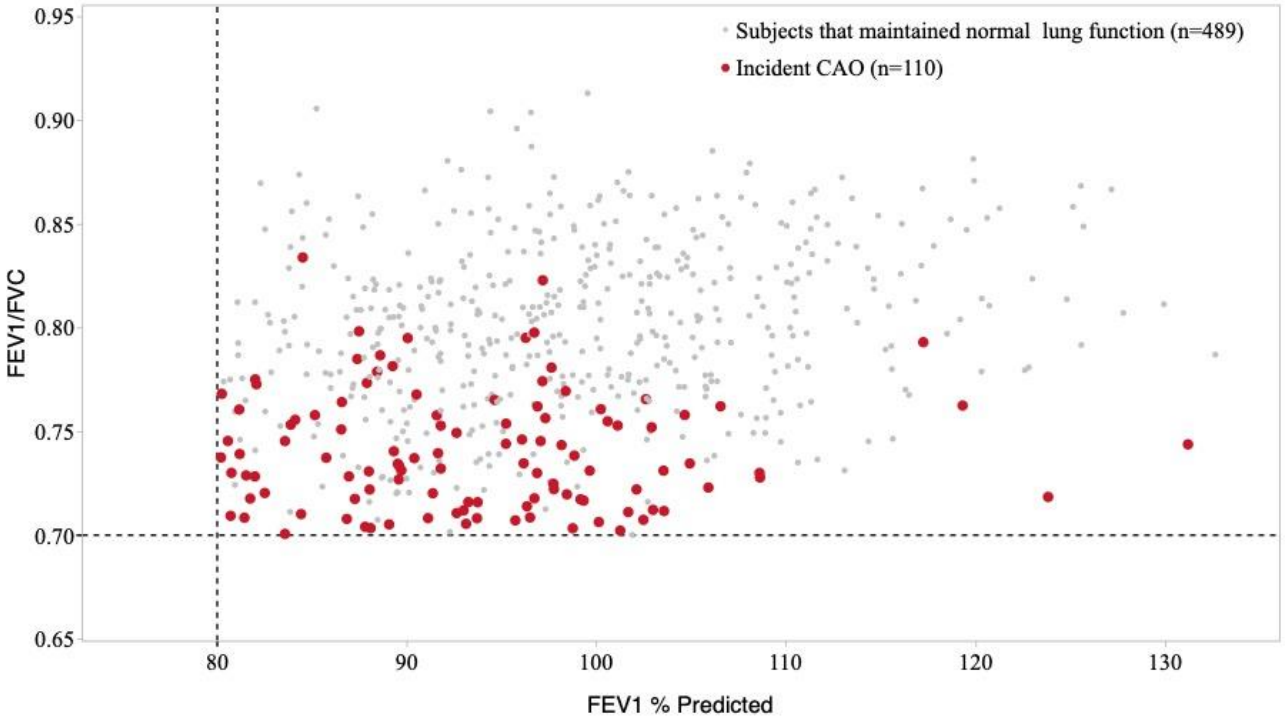


Figure E4. Proposed Algorithm to apply our model as a recruitment tool in research studies examining incident CAL.

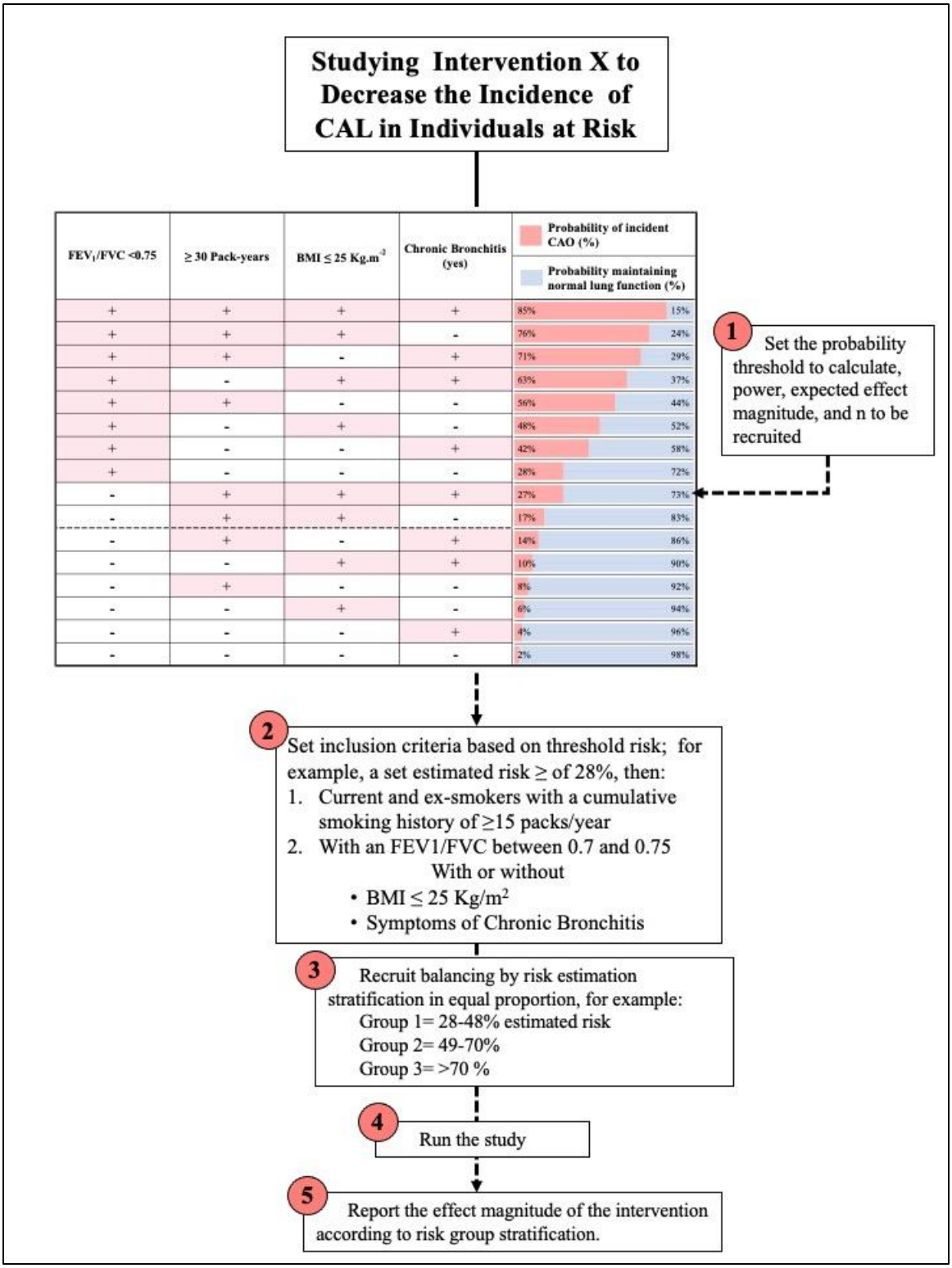


Table E1. Spirometric classification criteria.

A. Based on the Global Initiative for Chronic Obstructive Lung Disease (GOLD) 2023 guidelines(1), and Preserved Ratio Impaired Spirometry (PRISm)(2) .

	Normal spirometry	GOLD 1	GOLD 2	GOLD 3	GOLD 4	PRISm
Postbronchodilator FEV_1/FVC	≥0.70	<0.70	<0.70	<0.70	<0.70	≥0.70
Postbronchodilator FEV_1 percent predicted	≥80 %	≥80 %	50-79 %	30-49%	<30 %	<80 %

B. Based on the Lower Limit of normal (LLN)

	Normal spirometry	CAL	PRISm
Postbronchodilator FEV_1/FVC	≥LLN	<LLN	≥LLN
Postbronchodilator FEV_1 percent predicted	≥LLN		<LLN

Table E2. Visual representation of the possible spirometric trajectories leading to the four possible outcomes. Each individual spirometry classification is based on the criteria displayed in Table E1.

	Baseline spirometry	Follow-up spirometry 1*	Follow-up spirometry n*	**Outcome classification**
1	N	N	N	**(preserved) Normal**
2	N	O	O	**Incident CAL**
2	N	N	O	**Incident CAOL**
3	N	P	P	**Incident PRISm**
3	N	N	P	**Incident PRISm**
4	N	O	N	**Unstable**
4	N	P	N	**Unstable**
4	N	P	O	**Unstable**

*For the Lovelace Smokers Cohort follow-up spirometry occurred at 18-months interval, in the COPDGene at 5-year interval.
Abbreviations: N= Normal spirometry, O= obstructive pattern, P= PRISm

Table E3. Comparison of baseline characteristics between subjects with incident CAL and those who maintained normal lung function at the end of the observation period in the COPDGene Cohort.

Characteristics	Incident CAO	Subject who remained in the normal lung function range	p-value
	146	684	
• Demographic and anthropometric data			
Age (years)	59 (9)	57 (8)	*0.0174*
Female sex (%)	68 (47%)	367 (54%)	*0.1200*
AA (%)	50 (34%)	22 (33%)	*0.6756*
BMI (Kg/m^2)	27.6 (5.5)	29.3 (5.6)	*0.0013*
Height (cm)	170 (9.5)	170 (9.2)	*0.5514*
Follow-up. mean (years)	10.2 (0.59)	10.08 (0.58)	*0.0827*
Number of spirometries (median, IQR)	3(3-3)	3(3-3)	
• Exposure			
Age of smoking initiation (years)	17 (5)	17 (5)	0.8354
Current smoking (%)	78 (53%)	325 (48%)	0.1946
Age of quitting smoking (years)*	51 (11)	47 (10)	*0.0050*
Pack-years of smoking	44 (25)	37 (18)	*0.0010*
• Lung function			
FEV_1/FVC	0.75 (0.04)	0.79 (0.05)	*<0.0001*
FEV_1 (L)	2.77 (0.68)	2.9 (0.67)	*0.0361*
FEV_1 % Predicted	94 (10)	98 (11)	*<0.0001*
FVC (L)	3.73 (0.96)	3.66 (0.87)	0.4384
FVC % Predicted	98 (11)	97 (11)	0.1784
• Symptoms			
mMRC score†	0.86 (1.23)	0.66 (1.08)	0.0719
SGRQ total	19 (19)	14 (16)	*0.0069*
Chronic bronchitis (%)	28 (19%)	69 (10%)	*0.0019*
• Comorbidities			
History of asthma (%)	27 (19%)	75 (11%)	0.2747
Told of having a diagnosis of COPD (%)	22 (15%)	36 (5%)	0.8727

* Applies to those who quit smoking at baseline evaluation.

Abbreviations: AA= African American, BMI= Body Mass Index, IQR= interquartile range, mMRC= modified Medical Research Council dyspnoea scale, SGRQ= St. Georges Respiratory Questionnaire

Table E4. Results from LASSO variable selection methods for the prediction of incident CAL.

A. Variables selected with continuous variables left intact.

Selected Lambda=0.00933, Mean deviance= 0.5999

Explanatory variable
1. FEV_1/FVC ratio
2. Pack-years of smoking
3. BMI (Kg/m^2)
4. FEV_1 % predicted
5. Age (years)
6. History of COPD
7. Chronic bronchitis (yes)
8. High School education (yes)

B. Variables selected with continuous variables dichotomized (see methods section). Selected Lambda=0.00997, Mean deviance= 0.6874

Explanatory variable
1. FEV1/FVC ratio ≤0.75
2. Pack-years of smoking ≥30
3. BMI ≤ 25 (Kg/m^2)
4. FEV_1 between 80-100 % predicted
5. Age ≥ 55 years
6. History of COPD
7. Chronic bronchitis
8. FVC ≤ 95 % predicted
9. History of asthma
10. High School education

Abbreviation: BMI= Body Mass Index.

Section C. Result from the Logistic regression analysis for predicting incident CAL in the LSC with continuous variables.

1. **Prediction formula**

The following prediction formula results from the logistic regression analysis aimed to predict incident CAL in the LSC. The formula includes all significant predictors and correspondent coefficients needed to calculate in smokers the six-years probability of incident CAL.

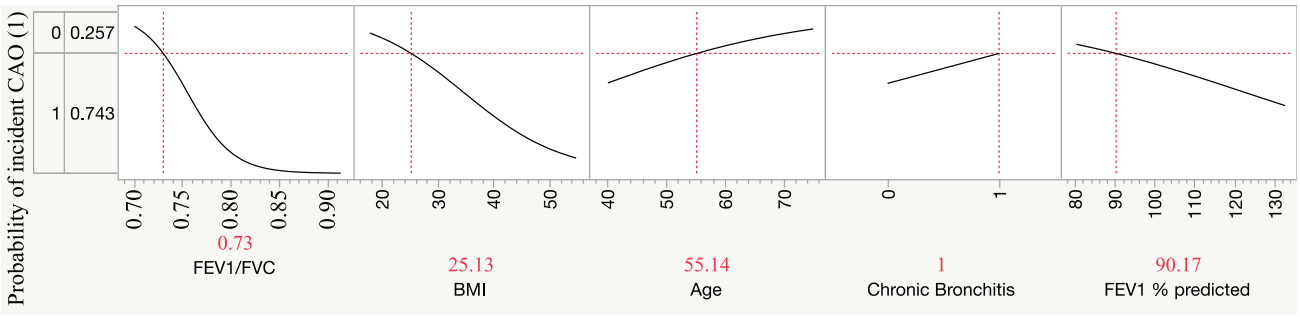
$$\text{Probability for incident CAO} = \frac{1}{1+ e^{(-Y)}}$$

Where **Y**:

Y= 34.09 + (-42.59) ∗ (FEV_1/FVC) + (-0.11) ∗ (BMI) + 0.05 ∗ (Age) + (-0.03) ∗ (FEV1 % predicted) + 0.84 ∗ (Chronic bronchitis, Yes=1, No=0)

2. **Graphic profiler**

The profiler is the graphic representation of the weight each predictor contributes across their value range in predicting the probability of incident CAL. As an example, for a 55 years-old individual with an FEV1/FVC=0.73, BMI of 25 $Kg.m^{-2,}$ who has chronic bronchitis and FEV_1 % predicted 90% (red dotted lines in each corresponding panel), the probability estimates for developing CAL at 6 years is 74% (left 2 columns, where first column represent group membership where 1= incident CAO 0= maintaining normal spirometry, and the 2nd column shows the calculated probability based on the parameters input.



3. **Contribution index table.**

The following table describes the contribution index of each parameter in predicting incident CAL in the derivation model derived from the logistic regression model.

Predictor	Weight (total contribution to the Variance)
FEV_1/FVC	0.81
BMI	0.31
Age	0.07
FEV_1 % predicted	0.05
Chronic Bronchitis	0.05

Section D. Prediction formula and contribution index table for the prediction of incident CAL in the LSC with dichotomic variables excluding FEV_1 between 80-100 % predicted (parsimonious model).

1. **Prediction formula**

The following prediction formula is the result of the logistic regression analysis aimed to predict incident CAL in the LSC. The formula includes all significant predictors and correspondent coefficients needed to calculate in smokers the six-years probability of incident CAL.

$$\text{Probability for incident CAO} = \frac{1}{1+ e^{(-Y)}}$$

Where **Y**:

$$\mathbf{Y} = -3.6949 + 2.7294 * (0.70 < \text{FEV1/FVC} \leq 0.75, \text{Yes} = 1, \text{No} = 0) + 1.2189 * (\geq 30 pack - years, \text{Yes} = 1, \text{No} = 0) + 0.8762 * (BMI \leq 25 Kg/m2, \text{Yes} = 1, \text{No} = 0) + 0.6275 * (Chronic\ bronchitis, Yes = 1, No = 0)$$

2. **Contribution index table.**

The following table describes the contribution index of each parameter in predicting incident CAO in the derivation model derived from the logistic regression model.

Predictor	Weight (total contribution to the Variance)
FEV_1/FVC	0.81
Pack-years	0.15
BMI	0.07
Chronic Bronchitis	0.04

Section E. Analysis using spirometric classification by the Lower Limit of Normal (LLN)

Using the post-BD $FEV_1/FVC < LLN$ to define CAL, among the 1,085 subjects meeting the inclusion criteria, 761 had normal baseline spirometry, 221 were classified as obstructed, and 102 as PRISm. The baseline characteristics between subjects with incident CAL and those who maintained normal lung function based on LLN are presented below in Table E4.

Table E4. Baseline characteristics comparing subjects with incident CAL and those who maintained normal lung function based on LLN classification criteria.

Characteristics	Incident CAL by LLN	Subject who remained with preserved lung function by LLN	p-value
	102	616	
▪ **Demographic and anthropometric data**			
Age (years)	58 ± 2	54 ± 8	*0.0006*
Female sex (%)	82 (80%)	506 (82%)	0.6777
Hispanic (%)	16 (16%)	113 (18%)	0.5794
BMI (Kg/m^2)	26.5 ± 5.1	28.6 ± 5.7	*0.0002*
Height (cm)	165.2 ± 8.9	165.2 ± 8.6	0.9728
▪ **Exposure**			
Age of quitting smoking (years)*	(n=50) 51± 9	(n=307) 46 ± 9	*0.0003*
Current smoking (%)	54 (53 %)	303 (49%)	0.5218
Pack-years of smoking	41 ± 19	35 ± 18	*0.0060*
▪ **Lung function**			
FEV_1/FVC	0.72 ± 0.04	0.79 ± 0.05	*<0.0001*
FEV_1 (L)	2.55 ± 0.66	2.78 ± 0.63	*0.0010*
FEV_1 % Predicted	118 ± 11	126 ± 15	*<0.0001*
FVC (L)	3.52 ± 0.91	3.52 ± 0.80	0.9711
FVC % Predicted	124 ± 12	122 ± 14	0.1521
▪ **Symptoms**			
mMRC score†	1.22 ± 1.22	1.01 ± 1.21	0.4254
SGRQ total	17.7 ± 15.2	15.8 ± 14.5	0.2352
Chronic bronchitis (%)	36%	24%	*0.0097*
▪ **Comorbidities**			
History of asthma (%)	21%	13%	*0.0461*
Diagnosis of COPD given (%)	12%	3.74%	*0.0018*

* Applies to those who quit smoking at baseline evaluation.

† Baseline data missing in 19 subjects with incident CAO and 122 subjects who remained within normal lung function range.

Abbreviations: BMI= Body Mass Index, IQR= interquartile range, mMRC= modified Medical Research Council dyspnea scale, SGRQ= St. Georges Respiratory Questionnaire

Six variables were significantly associated with incident CAL: BMI, the FEV1/FVC, age and cumulative smoking (pack-years), history of asthma, and the FEV_1 % of LLN. Multivariate analysis showed that an $FEV_1/FVC<0.75$, a BMI < 25 $Kg.m^{-2}$, an asthma history, and > 30 pack-years of smoking were the best predictors for incident CAL (**Table E5**). The area under the ROC was 0.83 (95% CI, 0.79-0.86), and the generalized R^2 was 0.33. With the optimal probability threshold estimated at ≥17%, the sensitivity is 0.79, the specificity of 0.80; the Positive Predicted Value is 0.40, the Negative Predicted Value of 0.96, and a misclassification rate of 0.20.

Table E5. Predictors for the incidence of chronic airway obstruction (CAL). The model was built using the spirometric reclassification based on the Lower Limit of Normal (LLN).

Risk factors	**OR**	**(95% CI)**	p-value
FEV_1/FVC <0.75	13.65	7.87-23.65	<0.0001
BMI ≤ 25 $Kg.m^{-2}$	2.19	1.32-3.61	0.0023
History of Asthma (yes)	2.17	1.15-4.07	0.0161
≥ 30 Pack-years	2.02	1.21-3.38	0.0073
Current smoker(yes)	1.79	1.08-2.98	0.0245

Section F. Longitudinal changes in lung function, smoking status, BMI, and symptoms.

Over the 6.3 years of observation, subjects with incident CAO lost a mean (SD) FEV_1 of 31 ± 48 $ml \cdot year^{-1}$ compared to 10 ± 37 $ml \cdot year^{-1}$ in subjects who remained under the normal spirometry category($p<0.0001$). For the FVC, we observed a mean *increase* in 28 ± 47 $ml \cdot year^{-1}$ and 2 ± 47 $ml \cdot year^{-1}$ ($p<0.0001$), respectively. The SGRQ total score worsened by 15 points (95% CI, 12-17) in the CAO incident group, compared to only 4 points (95% CI, 3-5) ($p=0.0001$) in subjects remaining in the normal lung function range. The mMRC dyspnoea score difference noted at baseline (**Table 2**) was maintained throughout the observation period, and the proportion of tobacco quitters and relapses was similar in both groups.

G. TRIPOD checklist

TRIPOD Checklist: Prediction Model Development

Section/Topic	Item	Checklist Item	Page
Title and abstract			
Title	1	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	1
Abstract	2	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	4-5
Introduction			
Background and objectives	3a	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	6
	3b	Specify the objectives, including whether the study describes the development or validation of the model or both.	7
Methods			
Source of data	4a	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	7 & 9
	4b	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	7
Participants	5a	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	7
	5b	Describe eligibility criteria for participants.	7
	5c	Give details of treatments received, if relevant.	
Outcome	6a	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	8
	6b	Report any actions to blind assessment of the outcome to be predicted.	
Predictors	7a	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	9
	7b	Report any actions to blind assessment of predictors for the outcome and other predictors.	
Sample size	8	Explain how the study size was arrived at.	9
Missing data	9	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	11
Statistical analysis methods	10a	Describe how predictors were handled in the analyses.	9
	10b	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	9
	10d	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	9
Risk groups	11	Provide details on how risk groups were created, if done.	
Results			
Participants	13a	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	8,9, sm
	13b	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	11
Model development	14a	Specify the number of participants and outcome events in each analysis.	11
	14b	If done, report the unadjusted association between each candidate predictor and outcome.	
Model specification	15a	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	SM
	15b	Explain how to the use the prediction model.	12 Fig 3
Model performance	16	Report performance measures (with CIs) for the prediction model.	Table 3
Discussion			
Limitations	18	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	16
Interpretation	19b	Give an overall interpretation of the results, considering objectives, limitations, and results from similar studies, and other relevant evidence.	16
Implications	20	Discuss the potential clinical use of the model and implications for future research.	15
Other information			
Supplementary information	21	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	Yes SM
Funding	22	Give the source of funding and the role of the funders for the present study.	1

We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document.

References

1. Global strategy for the diagnosis, management, and prevention of Chronic Obstructive Pulmonary Disease (2023 Report). 2022. Available from: https://goldcopd.org/2023-gold-report-2/.

2. Wan ES, Castaldi PJ, Cho MH, Hokanson JE, Regan EA, Make BJ, Beaty TH, Han MK, Curtis JL, Curran-Everett D, Lynch DA, DeMeo DL, Crapo JD, Silverman EK. Epidemiology, genetics, and subtyping of preserved ratio impaired spirometry (PRISm) in COPDGene. *Respiratory research* 2014; 15: 219S-213.