From pre-COPD to COPD: a Simple, Low cost and easy to IMplement (SLIM) risk calculator

Miguel J. Divo [1], Congjian Liu [1], Francesca Polverino [2], Peter J. Castaldi [3,4], Bartolome R. Celli [1,5] and Yohannes Tesfaigzi [1,5]

[1]Pulmonary and Critical Care Division, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. [2]Pulmonary and Critical Care Medicine, Department of Medicine, Baylor College of Medicine Houston, Houston, TX, USA. [3]Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. [4]General Medicine and Primary Care, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. [5]B.R. Celli and Y. Tesfaigzi are senior authors and contributed equally to this study and manuscript.

Corresponding author: Miguel J. Divo (mdivo@bwh.harvard.edu)

Shareable abstract (@ERSpublications)
In middle-aged ever-smokers, a simple predictive model, with FEV_1/FVC, smoking history, BMI and chronic bronchitis, helps identify subjects at high risk of developing chronic airflow limitation
https://bit.ly/45mwLx3

Cite this article as: Divo MJ, Liu C, Polverino F, *et al*. From pre-COPD to COPD: a Simple, Low cost and easy to IMplement (SLIM) risk calculator. *Eur Respir J* 2023; 62: 2300806 [DOI: 10.1183/13993003.00806-2023].

This article has an editorial commentary: https://doi.org/10.1183/13993003.01470-2023

Received: 13 May 2023
Accepted: 9 Aug 2023

Abstract

Background The lifetime risk of developing clinical COPD among smokers ranges from 13% to 22%. Identifying at-risk individuals who will develop overt disease in a reasonable timeframe may allow for early intervention. We hypothesised that readily available clinical and physiological variables could help identify ever-smokers at higher risk of developing chronic airflow limitation (CAL).

Methods Among 2273 Lovelace Smokers' Cohort (LSC) participants, we included 677 (mean age 54 years) with normal spirometry at baseline and a minimum of three spirometries, each 1 year apart. Repeated spirometric measurements were used to determine incident CAL. Using logistic regression, demographics, anthropometrics, smoking history, modified Medical Research Council dyspnoea scale, St George's Respiratory Questionnaire, comorbidities and spirometry, we related variables obtained at baseline to incident CAL as defined by the Global Initiative for Chronic Obstructive Lung Disease and lower limit of normal criteria. The predictive model derived from the LSC was validated in subjects from the COPDGene study.

Results Over 6.3 years, the incidence of CAL was 26 cases per 1000 person-years. The strongest independent predictors were forced expiratory volume in 1 s (FEV_1)/forced vital capacity (FVC) <0.75, having smoked ⩾30 pack-years, body mass index (BMI) ⩽25 kg·m^2 and symptoms of chronic bronchitis. Having all four predictors increased the risk of developing CAL over 6 years to 85% (area under the receiver operating characteristic curve (AUC ROC) 0.84, 95% CI 0.81–0.89). The prediction model showed similar results when applied to subjects in the COPDGene study with a follow-up period of 10 years (AUC ROC 0.77, 95% CI 0.72–0.81).

Conclusion In middle-aged ever-smokers, a simple predictive model with FEV_1/FVC, smoking history, BMI and chronic bronchitis helps identify subjects at high risk of developing CAL.

Introduction

COPD is defined as a heterogeneous lung condition characterised by chronic respiratory symptoms (dyspnoea, cough, expectoration and/or exacerbations) due to abnormalities of the airways (bronchitis, bronchiolitis) and/or alveoli (emphysema) that cause persistent, often progressive, airflow obstruction [1]. COPD has a long latency period and once the diagnosis is established, modifying disease progression has proven difficult [2].

Cigarette smoking remains the leading cause of COPD. Yet, in this at-risk group, screening spirometry for early detection of COPD is discouraged in the absence of respiratory symptoms [3], creating a lack of

clinical spirometric data in young and middle-aged adults. Chronic airflow limitation (CAL) is often used as a surrogate marker of COPD incidence in long-term epidemiological studies [4–8]. From these studies, we have learned that the estimated incidence in ever-smokers ranges between 13% and 22%. Predicting who will develop CAL can provide an opportunity to inform those subjects at risk and potentially design effective disease-modifying interventions at early disease stages [4].

Different studies have shown that the risk for CAL is increased among subjects with poor lung growth and low maximal forced expiratory volume in 1 s (FEV_1) attained by the fourth decade of life [5, 9, 10], the amount of smoking [10], the presence of chronic mucous production [10, 11], a low baseline value of diffusing capacity of the lung for carbon monoxide (D_{LCO}) [12], a rapid rate of FEV_1 decline [13] or abnormalities on chest computed tomography (CT) [14]. However, except for chronic mucous production and smoking, which are easily obtainable, clinicians would need to perform yearly spirometry to determine FEV_1 slope, obtain a chest CT or measure D_{LCO} to identify possible cases [3].

We hypothesised that a more practical alternative could be integrating variables easily obtainable by any clinician and building a predictive model to identify ever-smokers with baseline normal spirometry more likely to develop CAL. We tested this hypothesis using the prospectively collected data from the Lovelace Smokers' Cohort (LSC). This longitudinal, well-characterised, observational cohort studies the factors associated with CAL development among ever-smokers. Analysis of data from the LSC has helped identify lung function trajectories [5, 13], spirometric variability over time [15] and individual risks for CAL development [16–18].

Methods

Study design, setting and population

The LSC recruited 2273 individuals from the Albuquerque area (NM, USA) aged 40–75 years with ⩾15 pack-years of smoking from 2001 to 2015. Subjects are followed every 18 months with questionnaires, anthropometric measurements and pulmonary function tests [18]. The Western (20031684) and Mass General Brigham (protocol 2020P003513) Institutional Review Boards approved the study, and all participants provided informed consent.

Eligibility criteria

To study CAL incidence, we included all current and ex-smokers with a post-bronchodilator FEV_1/forced vital capacity (FVC) ⩾0.7 and FEV_1 ⩾80% predicted, who were followed for at least 3 years and performed at least three spirometries measured at least 1 year apart. We used the 3 years and three spirometries criteria because the progression to COPD based on FEV_1 measurements can vary [19, 20] and three spirometries help establish a reliable progression pattern [15, 21].

Study measurements

Demographics, anthropometrics and smoking status were obtained by trained personnel at baseline and each visit. Body mass index (BMI) was calculated in $kg \cdot m^{-2}$. Dyspnoea was evaluated using the modified Medical Research Council dyspnoea scale (mMRC) [22], and quality of life using the St George's Respiratory Questionnaire (SGRQ) [23] and the adult American Thoracic Society Division of Lung Disease-78 questionnaire [24]. Chronic bronchitis was defined as the persistence of cough and phlegm for 3 months for at least 2 years. Self-reported comorbidities, including asthma, were ascertained during the initial visit.

Pulmonary function and spirometric classification

Spirometries were obtained following international guidelines [25] and interpretation of the post-bronchodilator results into the following categories: 1) normal: subject with FEV_1/FVC ⩾0.70 and FEV_1 ⩾80% predicted; 2) CAL: when FEV_1/FVC <0.70 with the severity of obstruction based on the Global Initiative for Chronic Obstructive Lung Disease (GOLD) spirometric classification [26]; and 3) preserved ratio impaired spirometry (PRISm): subjects with FEV_1/FVC ⩾0.70 and FEV_1 <80% predicted (supplementary table E1a) [27]. The National Health and Nutrition Examination Survey (NHANES) III provided the reference values [28].

Outcomes

To determine incident CAL, we included only those subjects with normal spirometry at baseline. Lung function trajectories were determined using all available spirometries rather than just the first and last measurements. Four possible trajectories were determined based on each visit's spirometric classification: 1) sustained normal spirometry included subjects with normal spirometry at every visit; 2) incident CAL included those who transitioned and remained in any spirometric GOLD stage in the subsequent visits;

3) incident PRISm included those who transitioned and remained in the PRISm classification; and 4) an "Unstable" category referring to subjects with fluctuating spirometric patterns without a clear trajectory (see supplementary table E2 for a graphic representation).

Statistical analysis

Summary statistics included means, standard deviation, medians and interquartile ranges for continuous variables and proportions for categorical variables. Chi-squared and Fisher's exact tests were used to analyse categorical variables, while the two-sample t-test was used for continuous variables.

Derivation of a prediction model for incident CAL

A multivariable logistic regression model was used to evaluate parameter estimates and odds ratios with 95% confidence intervals for factors associated with the incidence of CAL. Based on prior knowledge and clinical plausibility, we included: age [10], gender [10], Hispanic ethnicity [16], pack-years of smoking [10], current smoking status [10, 29], level of education, chronic bronchitis [11], SGRQ scores, mMRC score [30], history of asthma [10, 29, 31], have received a diagnosis of COPD, BMI, baseline post-bronchodilator FEV_1/FVC, FEV_1 % pred and FVC % pred as candidate predictors. The least absolute shrinkage and selection operator (LASSO) with cross-validation was used as the variable selection method for our final models. We calculated that a conservative sample size of at least 368 subjects was needed to minimise model overfitting and to target precise estimates based on the use of 15 candidate variables, an estimated incidence of 15% of the primary composite outcome and the expected Nagelkerke's R^2 of at least 0.57 [32].

To increase clinical applicability, a second model was derived. Continuous variables were dichotomised at threshold values calculated using a recursive partitioning decision tree with incident CAL as the outcome. Recursive partitioning splits continuous predictors by optimising the cutting value based on the LogWorth statistics [33]. Each model was evaluated for discrimination and calibration, and the Youden index field [34] was applied to choose the optimal probability threshold used to define a case.

External validation

The external validation population was drawn from eligible participants from the COPDGene study (www.COPDGene.org) [35] and none of the subjects were represented in both cohorts. COPDGene (ClinicalTrials.gov: NCT00608764) recruited 10 371 smokers with at least 10 pack-years of cigarette smoking at 21 US clinical centres between 2007 and 2011. COPDGene collects longitudinal data on study participants at 5-year intervals, with the 10-year study visit (Visit 3) ongoing. For this analysis, we included the 830 subjects who completed Visit 3 (having three spirometries) (supplementary figure E1). All coefficients from the derivation model in the LSC were precisely applied in this sample. We estimated the discrimination and calibration in the external validation cohorts by setting the same probability threshold chosen in the derivation cohort. COPDGene was approved by Institutional Review Boards at all study centres and all subjects signed written informed consent.

Sensitivity and secondary analysis

We repeated the analyses using the lower limit of normal (LLN) to define CAL, using the NHANES III values as the reference (supplementary table E1b) [28].

To better inform CAL evolution, we calculated each subject's FEV_1 and FVC slopes using linear regression and expressed them in mL per year [19]. The slopes of FEV_1 and FVC were compared between incident CAL and those with persistent normal spirometry. Also, we determined longitudinal changes in BMI, smoking quitting and relapse rates, symptoms progression by the mMRC, and SGRQ scores over the observation period using a mixed model for repeated measures. For these models, the group membership (incident CAL and persistent normal spirometry), time in years and the interaction were entered as the fixed effects, while each subject was the random effect. Previous reports have demonstrated that missingness in the LSC and COPDGene cohorts is completely at random; in this analysis, missingness was <25% and we used single imputation to complete the dataset. For hypothesis testing, $p \leqslant 0.05$ was deemed statistically significant. We used SAS JMP Pro version 16.1.0 (SAS Institute, Cary, NC, USA) and R version 4.0.3 (R Foundation for Statistical Computing, Vienna, Austria).

Results

Among the 1087 LSC participants with at least three spirometries and at least 3 years of follow-up, 677 (30%) had normal baseline spirometry and were included in this analysis (supplementary figure E2). These subjects had a mean age of 54 years, 51% were current smokers, 35 pack-years of smoking, 82% were female and 18% were Hispanic (table 1). On average, these subjects had five spirometries measured over

TABLE 1 Baseline characteristics of the Lovelace Smokers' Cohort (LSC) (derivation cohort) and the subjects of the COPDGene cohort (validation cohort)

	LSC (n=677)	COPDGene (n=830)
Age (years)	54±9	57±8
Female	571 (82)	435 (52)
Race/ethnicity		
White	508 (75)	558 (67)
Hispanic	125 (18)	
Black (African American)		272 (33)
BMI (kg·m^{-2})	28.0±5.4	28.9±5.6
Current smoker	346 (51)	453 (45)
Pack-years of smoking	35±18	38±19
Age of smoking initiation (years)	17±4	17±2
Pre-COPD#	273 (54)	416 (50)
Spirometries (n)	5 (4–6)	3
FEV_1/FVC	0.79±0.05	0.79±0.05
FEV_1 (L)	2.8±0.6	2.9±0.7
FEV_1 (% pred)	97±10	98±11
Follow-up (years)	6.3±1.8	10.1±0.6

Data are presented as mean±SD, n (%) or median (interquartile range). BMI: body mass index; FEV_1: forced expiratory volume in 1 s; FVC: forced vital capacity. All subjects in both cohorts had normal lung function at baseline, performed at least three spirometries, were followed for at least 3 years and had ⩾15 pack-years of cumulative smoking. #: pre-COPD is defined as having normal spirometric results and either chronic sputum production, cough or modified Medical Research Council dyspnoea scale (mMRC) score >2 points at baseline [41]; chest computed tomography or exacerbation-like events were not available. During the baseline evaluation, there were missing data for cough (n=69) and mMRC (n=43).

6.3 years of observation. For the COPDGene validation cohort, 830 participants met the inclusion criteria (supplementary figure E1). These subjects had a mean age of 57 years, 52% were female and 33% were African American (table 1).

Incidence of CAL in the LSC

Over the observation period, 110 subjects (16%) developed incident CAL, 489 (72%) maintained normal spirometries, 15 (2%) transitioned to PRISm and 63 (9%) demonstrated an "Unstable" pattern (figure 1). Thus, the incidence rate for CAL in the LSC was 26 cases per 1000 person-years. 19% of the symptomatic (n=70) and 13% of the asymptomatic (n=40) participants evolved to CAL (figure 1), and by the end of the observation period, 76 (69%) of the 110 with incident CAL were symptomatic, reflecting that six asymptomatic participants developed new symptoms during the observation period. The comparison of the baseline characteristics between subjects with and without incident CAL is presented in table 2 for the LSC cohort. As seen in supplementary figure E3, most of the subjects who developed CAL (red dots) were clustered close to the FEV_1/FVC 0.70 and FEV_1 80% predicted thresholds.

Derivation and validation of a prediction model for incident CAL

From the 15 candidate predictors for CAL incidence (supplementary table E4a), FEV_1/FVC, pack-years, age, BMI, FEV_1 % pred, FVC % pred, diagnosis of COPD, a history of chronic bronchitis and education level were the LASSO's selected candidate predictors. In the multivariate model, FEV_1/FVC, BMI, age, FEV_1 % pred and a history of chronic bronchitis were the best predictors of incident CAL in the derivation cohort. This model's discrimination and calibration characteristics are presented in table 3 (model 1). The prediction formula, graphic profiler for each predictor and predictor's contribution index table are shown in supplementary material C. To facilitate its practical use, we built a model in which the six continuous candidate variables (age, BMI, pack-years, FEV_1/FVC, FEV_1 % pred and FVC % pred) were dichotomised on thresholds determined by recursive partitioning. The resulting optimal split values were 55 years for age, 25 kg·m^{-2} for BMI, 30 pack-years for smoking, 0.75 for FEV_1/FVC, 100% for FEV_1 % pred and 95% for FVC % pred. We applied LASSO to select the new transformed variables (supplementary table E4b). We found that FEV_1/FVC 0.70–0.75, ⩾30 pack-years of cumulative smoking, BMI ⩽25 kg·m^{-2}, FEV_1 80–100% predicted and a history of chronic bronchitis were significant predictors of incident CAL (table 4: model 2); the model's characteristics are listed in table 3. However, a more parsimonious model was evaluated by excluding FEV_1 80–100% predicted from the previous model due to collinearity with FEV_1/FVC (table 4: model 3). In this simpler model, the 6-year probability of developing CAL in a

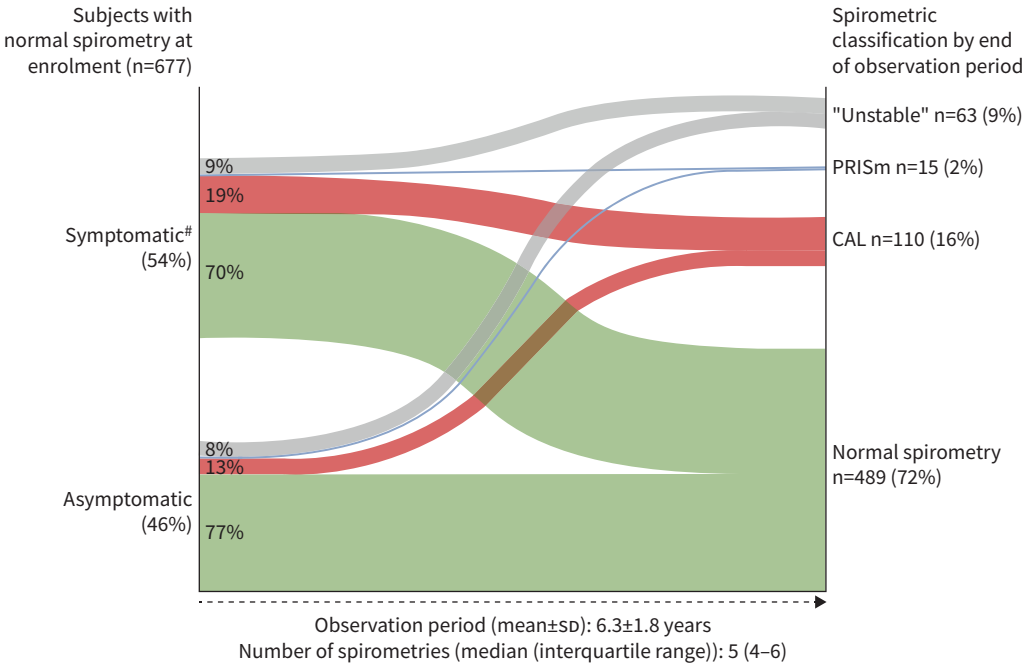

FIGURE 1 Cumulative incidence of subjects with chronic airflow limitation (CAL), preserved ratio impaired spirometry (PRISm), normal spirometry or those who fluctuate between these spirometric classifications ("Unstable") at the end of the observation period for those symptomatic and asymptomatic subjects who entered the study with preserved lung function in the Lovelace Smokers' Cohort. #: symptomatic is defined as having either chronic sputum production, cough or modified Medical Research Council dyspnoea scale (mMRC) score >2 points at baseline; chest computed tomography or exacerbation-like events were not available. During the baseline evaluation, there were missing data for cough (n=69) and mMRC (n=43).

subject with all four predictors is 85% compared with only 2% for those without any predictors (figure 2). As seen in table 3, the area under the receiver operating characteristic curve (AUC ROC) is 0.84 (95% CI 0.81–0.89). With the optimal threshold to classify a case as those with a probability of ⩾16% based on the probability formula presented in supplementary material D, sensitivity is 0.72, specificity is 0.85, positive predictive value is 0.52 and negative predictive value is 0.93, with a misclassification rate of 0.17.

External validation of the prediction model

The LSC's simpler model with dichotomised variables was validated in the COPDGene cohort (n=830). Over the 10 years of follow-up, 146 subjects (18%) met the criteria for incident CAL (incidence rate of 18 cases per 1000 person-years). The comparison of the baseline characteristics between subjects with and without incident CAL in the COPDGene cohort is presented in supplementary table E3. We tested the model performance in this cohort using the derived prediction formula and the estimated probability of ⩾16% for a positive case. The results are displayed in table 3.

Sensitivity and secondary analyses

Analyses using spirometric classification by the LLN offered similar results to those obtained using the GOLD spirometric classification (supplementary material E). Finally, the longitudinal changes in lung function, smoking status, BMI and symptoms analyses are presented in supplementary material F.

Discussion

The present study shows that a combination of simple spirometry-derived parameters, FEV_1/FVC, combined with cumulative pack-year history of smoking, BMI and a history of chronic bronchitis provides a reliable estimate for the risk of developing CAL within 6 years in ever-smokers. The prediction model was first derived from the LSC and then externally validated in the COPDGene study, obtaining comparable results. Many of the model's individual components are known to increase the risk of CAL [10, 29, 36]; here, we replicated and combined them to build a single practical risk calculator able to

TABLE 2 Comparison of baseline characteristics between subjects with incident chronic airflow limitation (CAL) and those who maintained normal lung function at the end of the observation period in the Lovelace Smokers' Cohort

	Incident CAL (n=110)	Maintained normal lung function (n=489)	p-value
Demographic and anthropometric data			
Age (years)	58±9	53±9	<0.0001
Female	89 (81)	416 (85)	0.7533
Hispanic	15 (14)	101 (21)	0.1201
BMI (kg·m^{-2})	25.9±4.3	28.6±5.4	<0.0001
Height (cm)	165±8	166±9	0.1416
Follow-up (years)	6.5±1.9	6.2±1.8	0.3538
Spirometries (n)	5 (4–6)	5 (4–6)	0.4773
Exposure			
Age of smoking initiation (years)	18±5	17±4	0.0087
Current smoking	56 (51)	259 (53)	0.2348
Age of quitting smoking (years)[#]	51±8	45±9	<0.0001
Pack-years of smoking	42±18	33±17	<0.0001
Lung function			
FEV_1/FVC	0.74±0.03	0.80±0.04	<0.0001
FEV_1 (L)	2.67±0.61	2.85±0.61	0.0125
FEV_1 (% pred)	94±9	99±10	<0.0001
FVC (L)	3.61±0.84	3.56±0.77	0.9411
FVC (% pred)	100±4	99±5	0.0361
Symptoms			
mMRC score[¶]	1.33±1.15	1.09±1.22	0.0459
SGRQ total score	18±15	16±15	0.2120
Chronic bronchitis (%)	37	24	0.0206
Comorbidities			
History of asthma	19 (18)	61 (13)	0.1595
Told of having a diagnosis of COPD	11 (10)	11 (2)	0.0005

Data are presented as mean±SD, n (%) or median (interquartile range). BMI: body mass index; FEV_1: forced expiratory volume in 1 s; FVC: forced vital capacity; mMRC: modified Medical Research Council dyspnoea scale; SGRQ: St George's Respiratory Questionnaire. [#]: applies to those who quit smoking at baseline evaluation; [¶]: baseline data were missing in 15 subjects with incident CAL and 100 subjects who remained within the normal lung function range.

TABLE 3 Performance comparison of the derivation models (with continuous, dichotomic and dichotomic parsimonious variables) and their performance in the external validation cohort

	Derivation (Lovelace Smokers' Cohort)			Validation (COPDGene)
	Model 1	Model 2	Model 3	
Predictor's characteristics	Continuous variables	Continuous variables dichotomised	Continuous variables dichotomised (parsimonious model)	Validation with model 3 probability formula
Subjects (n)	599	599	599	830
CAL incidence (cases per 1000 person-years)	26	26	26	18
Predictors in the model (n)	5	5	4	4
Probability threshold for case assignment (%)	>15	>19	>16	>16
Sensitivity	0.89	0.75	0.72	0.79
Specificity	0.78	0.85	0.85	0.67
Positive predictive value	0.47	0.53	0.52	0.34
Negative predictive value	0.97	0.94	0.93	0.94
AUC ROC	0.91	0.86	0.84	0.77
Misclassification rate	0.20	0.17	0.17	0.31
Nagelkerke's R^2	0.52	0.41	0.39	

CAL: chronic airflow limitation; AUC ROC: area under the receiver operating characteristic curve.

TABLE 4 Estimates (odd ratios) and weight of predictors for the incidence of chronic airway limitation analysed by multivariate logistic regression in the Lovelace Smokers' Cohort

	OR (95% CI)	p-value	Total effect on model
Model 2			
FEV1/FVC <0.75	13.75 (8.15–23.17)	<0.0001	0.72
⩾30 pack-years cumulative smoking	3.38 (1.94–5.89)	<0.0001	0.15
BMI ⩽25 kg·m^{-2}	2.41 (1.43–4.06)	0.001	0.08
FEV1 % pred <100%	2.07 (1.17–3.67)	0.0130	0.05
Chronic bronchitis (yes)	1.89 (1.09–3.27)	0.0231	0.04
Model 3			
FEV_1/FVC <0.75	15.32 (9.14–25.69)	<0.0001	0.80
⩾30 pack-years cumulative smoking	3.38 (1.96–5.86)	<0.0001	0.15
BMI ⩽25 kg·m^{-2}	2.40 (1.43–4.03)	0.0009	0.07
Chronic bronchitis (yes)	1.87 (1.09–3.22)	0.0234	0.04

Model 2: continuous variables dichotomised and five predictors; model 3: continuous variables dichotomised and four predictors (parsimonious model). FEV_1: forced expiratory volume in 1 s; FVC: forced vital capacity; BMI: body mass index.

identify high-risk subjects likely to develop CAL. In addition, we then validated the model in a different cohort of at-risk smokers.

Currently, most patients diagnosed with COPD are in their seventh or eighth decade of life when the management of these patients offers a limited impact on the natural progression of the disease. In COPD, as in other non-communicable chronic diseases [37–39], efforts are being made to identify patients with earlier stages of the disease. This concept has been adopted recently by the GOLD initiative to identify those individuals with a high likelihood of developing poorly reversible airflow limitation and labelled as

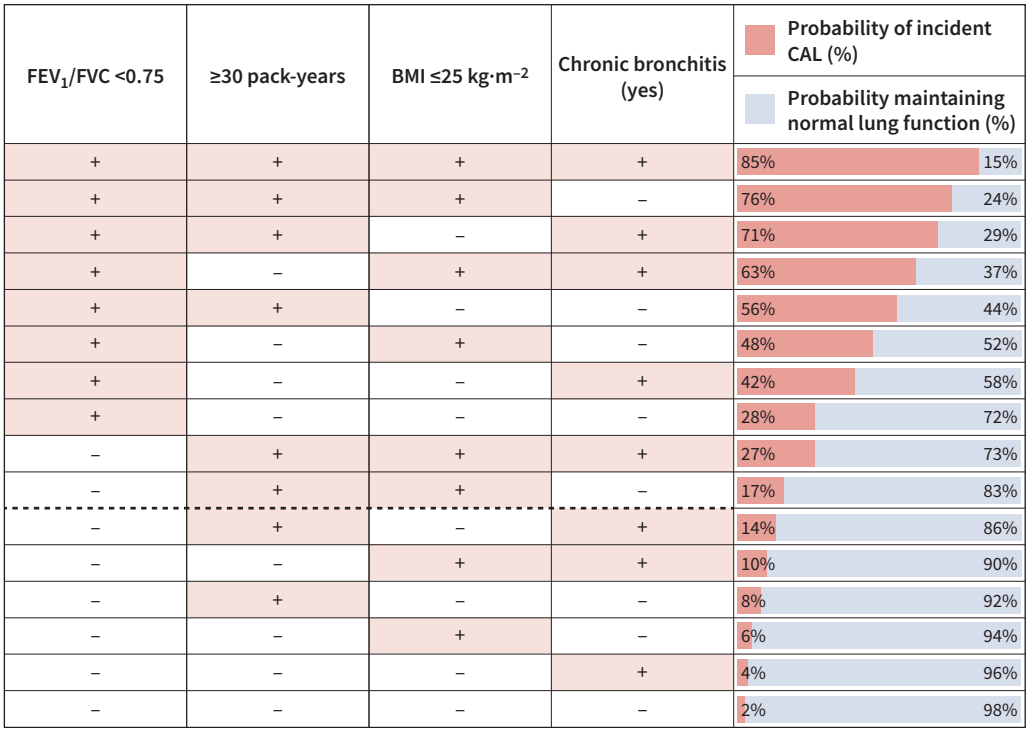

FIGURE 2 Prediction estimates for the incidence of chronic airway limitation (CAL). The probabilities were calculated with the prediction formula derived and presented in supplementary material D. The dashed line represents the cut-off threshold of >16% for case (CAL) assignment (see text for details). FEV_1: forced expiratory volume in 1 s; FVC: forced vital capacity; BMI: body mass index.

"pre-COPD" [1, 40]. Pre-COPD is defined by the presence of respiratory symptoms, structural lung lesions, physiological abnormalities (including low-normal FEV_1, gas trapping, hyperinflation, reduced D_{LCO} or rapid FEV_1 decline) without airflow limitation ($FEV_1/FVC \geqslant 0.7$) [1, 40]. In the LSC cohort, only 54% of the participants at baseline met the criteria for pre-COPD, albeit not having data from chest CT, D_{LCO} or exacerbation-like events (table 1). In our study, FEV_1/FVC between 0.70 and 0.75 was the strongest predictor for incident CAL (table 3), making a strong case to add this criterion to the pre-COPD definition. When FEV_1/FVC between 0.70 and 0.75 is combined with a cumulative smoking history of $\geqslant$30 pack-years, BMI $\leqslant$25 kg·m^{-2} and a history of chronic bronchitis, the probability of developing CAL reaches 85%. Using the colour-coded, easy-to-use chart (figure 2), or using the logistic regression formula in supplementary material C and D, integrated into an online or clinical decision calculator, the risk can be easily computed in many different settings.

Evidence from cohorts of smokers at risk has shown a relationship between the existence of symptoms (cough and phlegm), low FEV_1 % pred, rapid rate of lung function decline, low D_{LCO}, low BMI and emphysema detected with chest CT scans with incident CAL [10, 41, 42]. Currently, a chest CT is hard to justify if the subject does not meet lung cancer screening criteria, repeated spirometries over time to establish FEV_1 decline is time and resource consuming, while the D_{LCO} may not be readily available, particularly in low- or middle-income countries with resource-limited healthcare systems [43]. However, we recognise that the addition of chest CT and D_{LCO} could improve the accuracy of our model and expand case detection not only for CAL but also for more structural phenotypes as well. However, our model offers a practical and economical clinical tool for prognostication in ever-smokers or for the much-needed trials to prevent further progression to CAL as a case-finding tool for study recruitment, as exemplified in the algorithms presented in figure 3 and supplementary figure E4.

Having FEV_1/FVC between 0.70 to 0.75 suggests a dissociation between expiratory airflow (FEV_1) and lung volume (FVC) as described in the Tasmanian Longitudinal Health Study [44], where individuals with lower FEV_1/FVC trajectories from age 7 to 53 years had lower BMI, higher cumulative smoking and chronic sputum production [44]. Our observation could represent a snapshot of the evolution of some of these trajectories over 6 years in those who reach middle age and beyond.

Our results also support the concept of a dose–response relationship for cumulative smoking (pack-years) and also that of early quitting of smoking, which decreases the risk for CAL [9, 45]. A novel and interesting finding was the risk conferred by a lower BMI, a known predictor of worse outcomes in patients with established COPD [46–48] but less acknowledged as a risk factor for incident CAL [49]. Importantly, BMI did not change over 6.3 years, suggesting that, on average, lung disease progression was not accompanied by cachexia (supplementary material F). This indicates that the subjects were already in the lower BMI range at an earlier stage of life, an observation supported by longitudinal studies that included BMI measurements at younger age [50–52].

A history of chronic cough and phlegm, asthma or being an active cigarette smoker did not hold in our final model as strong predictors for CAL. Although this differs from the findings from the CARDIA, SPALDIA, UK MRC cohort and Copenhagen City Heart studies, the baseline spirometric values were not included in the prediction models of those studies [11, 53–55]. Our study agrees with that of the TESAOD cohort, where baseline spirometry was included in the model, and there, FEV_1/FVC at baseline was a strong predictor of future CAL [56].

There are several strengths to our study. The model is derived from a well-phenotyped cohort of at-risk subjects, with multiple spirometries over a median of >6 years. The final CAL predictors were externally validated in a second independent cohort with similar model performance. However, we also acknowledge important limitations. First, our model is far from perfect, with 40% of variance explained and acceptable performance in predicting this low probability event (26 cases per 1000 person-years); it is possible that the addition of chest CT imaging (emphysema or air trapping and dysanapsis) or low D_{LCO} could improve the model, but at the expense of more complexity and cost. Moreover, Smith *et al*. [57], using data from CanCOLD, showed that participants with dysanapsis assessed by chest CT had a higher incidence of CAL; interestingly, they also had lower baseline FEV_1/FVC, a finding indicating that perhaps these are surrogates' markers of the same process. Second, we cannot assume that those who maintained non-obstructed spirometry over 6 years of observation will remain normal if followed for a more extended period. Third, in addition to not having chest CT and D_{LCO}, we did not capture "exacerbation-like" events in our questionnaires; therefore, our estimate of 54% of participants meeting the pre-COPD definition is perhaps an underestimation (table 1). Fourth, the LSC cohort is mostly comprised of females (82%), with an important representation of Hispanics of Mexican origin (18%) and no African Americans, all of which

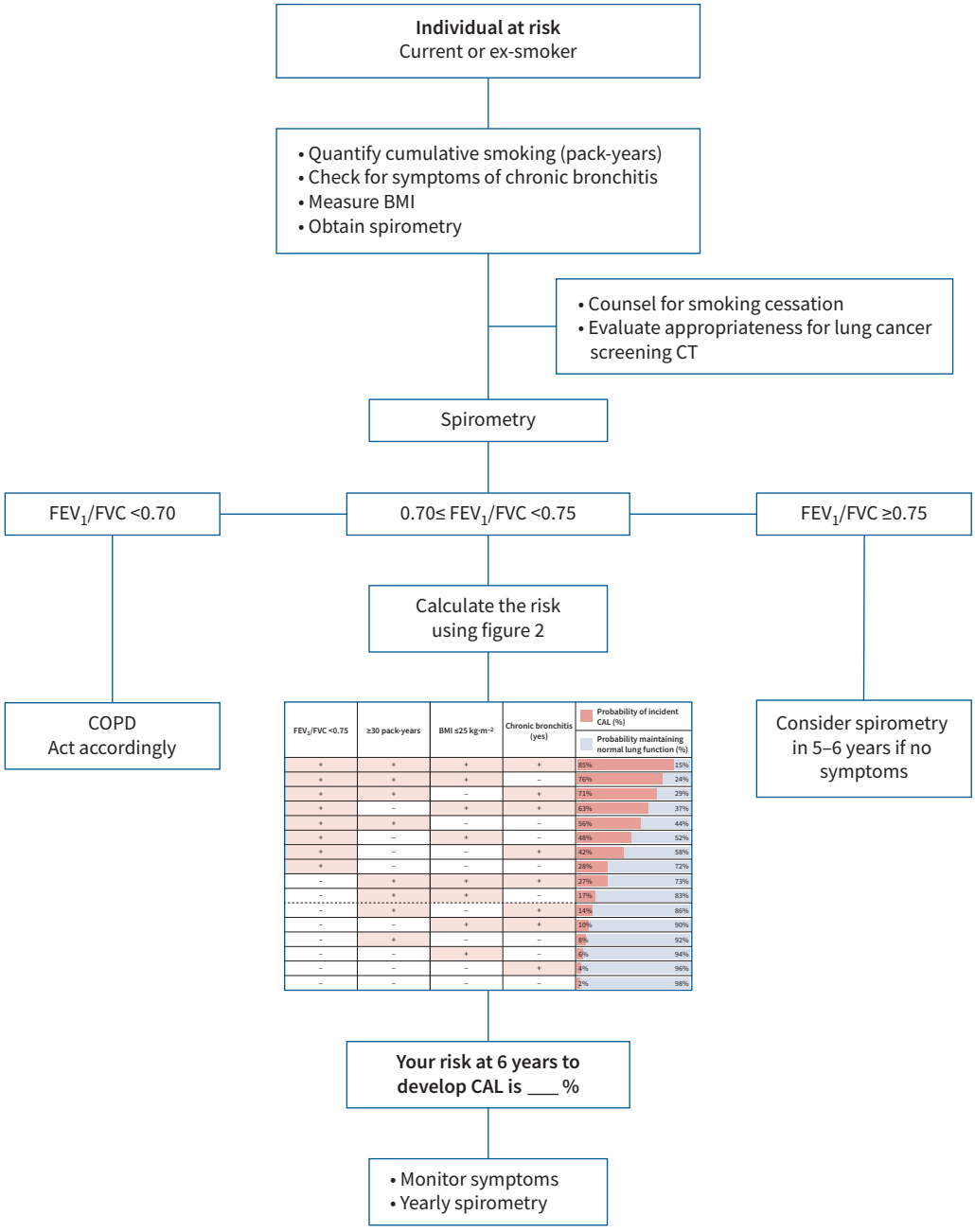

FIGURE 3 Proposed algorithm to apply our model to calculate the risk of incident chronic airway limitation (CAL) in the clinical setting. BMI: body mass index; CT: computed tomography; FEV_1: forced expiratory volume in 1 s; FVC: forced vital capacity.

affect external validation. We observed a drop in the AUC ROC in the validation cohort, which could be reflecting the baseline characteristics differences (table 1). However, an AUC ROC of 0.77 in the external validation cohort is good, particularly considering that we are predicting a low-incidence event. Importantly, the confidence intervals of both derivation and validation AUC ROC have some overlap, suggesting the truth is somewhere in the vicinity of the reported values (table 3).

In conclusion, in two different cohorts of at-risk smokers, we found that FEV_1/FVC 0.70–0.75, smoking history ⩾30 pack-years, BMI ⩽25 $kg \cdot m^{-2}$ and FEV_1 80–100% predicted provide a reasonable estimate for the risk of developing chronic airway obstruction. The variables included in the model are simple to obtain and provide an objective estimate to identify pre-COPD subjects.

Acknowledgements: The LSC was recruited through Lovelace Scientific Resources (Albuquerque, NM, USA) under the direction of Darlene Harbour. We thank Adda Tewolde (independent contractor) for the figure preparation.

The Western (number 20031684) and Mass General Brigham (protocol number 2020P003513) institutional review boards approved the study, and all participants provided informed consent.

Author contributions: Conception and design: M.J. Divo, B.R. Celli, F. Polverino and Y. Tesfaigzi. Acquisition of data: Y. Tesfaigzi. Analysis and interpretation of data: M.J. Divo, B.R. Celli, C. Liu, P.J. Castaldi and Y. Tesfaigzi. Drafting the article or revising it critically for important intellectual content: M.J. Divo, B.R. Celli, C. Liu, P.J. Castaldi, F. Polverino and Y. Tesfaigzi. Final approval of the version to be submitted for revision: M.J. Divo, B.R. Celli, C. Liu, P.J. Castaldi, F. Polverino and Y. Tesfaigzi. All authors had full access to all the study data and accepted responsibility for submitting this work.

Conflict of interest: The authors declare that they have no potential conflicts of interest to disclose. F. Polverino is a Section Editor of the *European Respiratory Journal*.

Support statement: This work was supported by funding from the State of New Mexico (appropriation from the Tobacco Settlement Fund), unrestricted grants from the CHEST Foundation grant 7897 (to M.J. Divo), NIH grants HL068111 and HL134370 (to Y. Tesfaigzi), and NHLBI grants R01HL149744 (to F. Polverino), R01HL124233 (to P.J. Castaldi) and R01HL126596 (to P.J. Castaldi). The sponsors had no role in the design and conduct of the study, in the collection, analysis and interpretation of data, or the preparation, review or approval of the manuscript. Funding information for this article has been deposited with the Crossref Funder Registry.

References

1 Agusti A, Celli BR, Criner GJ, *et al.* Global Initiative for Chronic Obstructive Lung Disease 2023 Report: GOLD Executive Summary. *Eur Respir J* 2023; 61: 2300239.
2 Celli BR, Wedzicha JA. Update on clinical aspects of chronic obstructive pulmonary disease. *N Engl J Med* 2019; 381: 1257–1266.
3 US Preventive Services Task Force. Screening for chronic obstructive pulmonary disease: US Preventive Services Task Force recommendation statement. *JAMA* 2016; 315: 1372–1377.
4 Agustí A, Faner R. Lung function trajectories in health and disease. *Lancet Respir Med* 2019; 7: 358–364.
5 Lange P, Celli B, Agusti A, *et al.* Lung-function trajectories leading to chronic obstructive pulmonary disease. *N Engl J Med* 2015; 373: 111–122.
6 Allinson JP, Hardy R, Donaldson GC, *et al.* Combined impact of smoking and early-life exposures on adult lung function trajectories. *Am J Respir Crit Care Med* 2017; 196: 1021–1030.
7 Young AL, Bragman FJS, Rangelov B, *et al.* Disease progression modeling in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2020; 201: 294–302.
8 Ross JC, Castaldi PJ, Cho MH, *et al.* Longitudinal modeling of lung function trajectories in smokers with and without chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2018; 198: 1033–1042.
9 Kohansal R, Martínez-Camblor P, Agustí A, *et al.* The natural history of chronic airflow obstruction revisited: an analysis of the Framingham Offspring cohort. *Am J Respir Crit Care Med* 2009; 180: 3–10.
10 Higgins MW, Keller JB, Becker M, *et al.* An index of risk for obstructive airways disease. *Am Rev Respir Dis* 1982; 125: 144–151.
11 Allinson JP, Hardy R, Donaldson GC, *et al.* The presence of chronic mucus hypersecretion across adult life in relation to chronic obstructive pulmonary disease development. *Am J Respir Crit Care Med* 2016; 193: 662–672.
12 Harvey B-G, Strulovici-Barel Y, Kaner RJ, *et al.* Risk of COPD with obstruction in active smokers with normal spirometry and reduced diffusion capacity. *Eur Respir J* 2015; 46: 1589–1597.
13 Petersen H, Sood A, Polverino F, *et al.* The course of lung function in middle-aged heavy smokers: incidence and time to early onset of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2018; 198: 1449–1451.
14 Regan EA, Lynch DA, Curran-Everett D, *et al.* Clinical and radiologic disease in smokers with normal spirometry. *JAMA Intern Med* 2015; 175: 1539–1549.
15 Sood A, Petersen H, Qualls C, *et al.* Spirometric variability in smokers: transitions in COPD diagnosis in a five-year longitudinal study. *Respir Res* 2016; 17: 147.
16 Bruse S, Sood A, Petersen H, *et al.* New Mexican Hispanic smokers have lower odds of chronic obstructive pulmonary disease and less decline in lung function than non-Hispanic whites. *Am J Respir Crit Care Med* 2011; 184: 1254–1260.
17 Petersen H, Sood A, Meek PM, *et al.* Rapid lung function decline in smokers is a risk factor for COPD and is attenuated by angiotensin-converting enzyme inhibitor use. *Chest* 2014; 145: 695–703.

18 Sood A, Petersen H, Blanchette CM, *et al.* Wood smoke exposure and gene promoter methylation are associated with increased risk for COPD in smokers. *Am J Respir Crit Care Med* 2010; 182: 1098–1104.
19 Casanova C, Aguirre-Jaime A, de Torres JP, *et al.* Longitudinal assessment in COPD patients: multidimensional variability and outcomes. *Eur Respir J* 2014; 43: 745–753.
20 Guerra S, Sherrill DL, Venker C, *et al.* Morbidity and mortality associated with the restrictive spirometric pattern: a longitudinal study. *Thorax* 2010; 65: 499–504.
21 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: 44–52.
22 Mahler DA, Wells CK. Evaluation of clinical methods for rating dyspnea. *Chest* 1988; 93: 580–586.
23 Jones PW, Quirk FH, Baveystock CM. The St George's Respiratory Questionnaire. *Respir Med* 1991; 85: Suppl. B, 25–31.
24 Ferris BG. Epidemiology standardization project (American Thoracic Society). *Am Rev Respir Dis* 1978; 118: 1–120.
25 Miller MR, Hankinson J, Brusasco V, *et al.* Standardisation of spirometry. *Eur Respir J* 2005; 26: 319–338.
26 Global Initiative for Chronic Obstructive Lung Disease (GOLD). Global Strategy for the Diagnosis, Management and Prevention of COPD. 2022. Available from: http://goldcopd.org/
27 Wan ES, Castaldi PJ, Cho MH, *et al.* Epidemiology, genetics, and subtyping of preserved ratio impaired spirometry (PRISm) in COPDGene. *Respir Res* 2014; 15: 89.
28 Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general U.S. population. *Am J Respir Crit Care Med* 1999; 159: 179–187.
29 Anthonisen NR, Connett JE, Kiley JP, *et al.* Effects of smoking intervention and the use of an inhaled anticholinergic bronchodilator on the rate of decline of FEV_1. The Lung Health Study. *JAMA* 1994; 272: 1497–1505.
30 Woodruff PG, Barr RG, Bleecker E, *et al.* Clinical significance of symptoms in smokers with preserved pulmonary function. *N Engl J Med* 2016; 374: 1811–1821.
31 Backman H, Jansson S-A, Stridsman C, *et al.* Chronic airway obstruction in a population-based adult asthma cohort – prevalence, incidence and prognostic factors. *Respir Med* 2018; 138: 115–122.
32 Riley RD, Ensor J, Snell KIE, *et al.* Calculating the sample size required for developing a clinical prediction model. *BMJ* 2020; 368: m441.
33 Sall J. Monte Carlo calibration of distributions of partition statistics. 2002. www.jmp.com/content/dam/jmp/documents/en/white-papers/montecarlocal.pdf Date last accessed: 15 August 2023.
34 Schisterman EF, Perkins NJ, Liu A, *et al.* Optimal cut-point and its corresponding Youden index to discriminate individuals using pooled blood samples. *Epidemiology* 2005; 16: 73–81.
35 Regan EA, Hokanson JE, Murphy JR, *et al.* Genetic epidemiology of COPD (COPDGene) study design. *COPD* 2010; 7: 32–43.
36 Vestbo J, Prescott E, Lange P. Association of chronic mucus hypersecretion with FEV_1 decline and chronic obstructive pulmonary disease morbidity. Copenhagen City Heart Study Group. *Am J Respir Crit Care Med* 1996; 153: 1530–1535.
37 McNiece KL, Poffenbarger TS, Turner JL, *et al.* Prevalence of hypertension and pre-hypertension among adolescents. *J Pediatr* 2007; 150: 640–644.
38 Olson DE, Rhee MK, Herrick K, *et al.* Screening for diabetes and pre-diabetes with proposed A1C-based diagnostic criteria. *Diabetes Care* 2010; 33: 2184–2189.
39 Steegers EA, Von Dadelszen P, Duvekot JJ, *et al.* Pre-eclampsia. *Lancet* 2010; 376: 631–644.
40 Celli B, Fabbri L, Criner G, *et al.* Definition and nomenclature of chronic obstructive pulmonary disease: time for its revision. *Am J Respir Crit Care Med* 2022; 206: 1317–1325.
41 Han MK, Agustí A, Celli BR, *et al.* From GOLD 0 to pre-COPD. *Am J Respir Crit Care Med* 2021; 203: 414–423.
42 Thomashow BM, Mannino DM, Tal-Singer R, *et al.* A rapidly changing understanding of COPD: World COPD Day from the COPD Foundation. *Am J Physiol Lung Cell Mol Physiol* 2021; 321: L983–L987.
43 Edelman Saul E, Guerra RB, Edelman Saul M, *et al.* The challenges of implementing low-dose computed tomography for lung cancer screening in low- and middle-income countries. *Nat Cancer* 2020; 1: 1140–1152.
44 Dharmage SC, Bui DS, Walters EH, *et al.* Lifetime spirometry patterns of obstruction and restriction, and their risk factors and outcomes: a prospective cohort study. *Lancet Respir Med* 2023; 11: 273–282.
45 Anthonisen NR, Connett JE, Murray RP. Smoking and lung function of Lung Health Study participants after 11 years. *Am J Respir Crit Care Med* 2002; 166: 675–679.
46 Schols AMWJ, Broekhuizen R, Weling-Scheepers CA, *et al.* Body composition and mortality in chronic obstructive pulmonary disease. *Am J Clin Nutr* 2005; 82: 53–59.
47 Celli BR, Cote CG, Marin JM, *et al.* The body-mass index, airflow obstruction, dyspnea, and exercise capacity index in chronic obstructive pulmonary disease. *N Engl J Med* 2004; 350: 1005–1012.
48 Celli BR, Locantore N, Tal-Singer R, *et al.* Emphysema and extrapulmonary tissue loss in COPD: a multi-organ loss of tissue phenotype. *Eur Respir J* 2018; 51: 1702146.
49 Harik-Khan RI, Fleg JL, Wise RA. Body mass index and the risk of COPD. *Chest* 2002; 121: 370–376.

50 Guerra S, Sherrill DL, Bobadilla A, *et al.* The relation of body mass index to asthma, chronic bronchitis, and emphysema. *Chest* 2002; 122: 1256–1263.

51 Breyer-Kohansal R, Faner R, Breyer M-K, *et al.* Factors associated with low lung function in different age bins in the general population. *Am J Respir Crit Care Med* 2020; 202: 292–296.

52 Divo MJ, Marin Oto M, Casanova Macario C, *et al.* Somatotypes trajectories during adulthood and their association with COPD phenotypes. *ERJ Open Res* 2020; 6: 00122-2020.

53 Probst-Hensch NM, Curjuric I, Pierre-Olivier B, *et al.* Longitudinal change of prebronchodilator spirometric obstruction and health outcomes: results from the SAPALDIA cohort. *Thorax* 2010; 65: 150–156.

54 Kalhan R, Dransfield MT, Colangelo LA, *et al.* Respiratory symptoms in young adults and future lung disease. The CARDIA lung study. *Am J Respir Crit Care Med* 2018; 197: 1616–1624.

55 Lindberg A, Jonsson AC, Ronmark E, *et al.* Ten-year cumulative incidence of COPD and risk factors for incident disease in a symptomatic cohort. *Chest* 2005; 127: 1544–1552.

56 Guerra S, Sherrill DL, Venker C, *et al.* Chronic bronchitis before age 50 years predicts incident airflow limitation and mortality risk. *Thorax* 2009; 64: 894–900.

57 Smith BM, Kirby M, Hoffman EA, *et al.* Association of dysanapsis with chronic obstructive pulmonary disease among older adults. *JAMA* 2020; 323: 2268–2280.