Statistical Methods:

All data are expressed as means with standard deviations or medians with interquartile ranges (IQR). The primary outcome was airflow obstruction assessed using FEV_1/FVC. Secondary outcomes included five additional measures of disease: FEV_1, CT emphysema, CT gas trapping, functional capacity using six minute walk distance (6MWD), and respiratory morbidity using St. George's Respiratory Questionnaire (SGRQ). We analyzed the relationship between individual smoking variables (cigarettes/day, duration and the combined product pack-years) and the outcomes using two separate methods.

1. In the first method, using generalized regression models, we estimated the association of the primary and each of the secondary outcomes with the individual smoking variables (duration, cigarettes/day and pack-years) with outcomes as continuous variables. Each fixed-effects model included adjustments for age, race, sex, body-mass-index, CT scanner type, center, age of smoking onset and current smoking status. Quantitative imaging variables were not log-transformed. Pearson's correlation analysis was performed to assess the pairwise relationship of the three smoking variables. The r for duration and pack-years was 0.51 (p<0.001), cigarettes/day and pack-years was 0.83 (p<0.001) and duration and cigarettes/day was -0.004 (p=0.68). To estimate relative effect sizes without scale difference, we estimated and compared the standardized regression co-efficient of duration to that of cigarettes/day and pack-years.

2. Second, to improve clinical interpretability and applicability across a range of combinations of smoking duration and cigarettes/day, we also categorized the smoking variables of interest (smoking duration in years and average number of cigarettes smoked per day) by increments of 10 units to create consistent intervals for the smoking

variables. Increment intervals were set at 10 as cigarettes/day is commonly expressed in half to one pack a day, although we had data for the average individual cigarettes per day in our cohort. We grouped participants into 25 different combinations of smoking duration and cigarettes/day to create a 5 X 5 table (please see example for FEV_1/FVC in **Supplemental Table 1**). Of the 25 groups thus created, we excluded two groups as they had zero or insufficient cases (**Supplemental Table 1**). Within each of these groups, we calculated adjusted means of the outcome, after adjusting for age, race, sex, body-mass-index, CT scanner type, center, age of smoking onset and current smoking status. Similarly, adjusted means of outcomes were estimated separately for 5 groups of smoking pack-years in increments of 10 pack-years. We observed that with increasing duration across categories, the adjusted means outcomes changed in a consistent direction (increase in emphysema, gas trapping, SGRQ; and decrease in FEV_1, FEV_1/FVC and 6MWD). We plotted the adjusted means of outcomes in a graphical form, and drew the linear trends of these adjusted means to illustrate the trends and variance of the adjusted means of the outcomes over categorized cigarettes/day and duration respectively. Beacause we used categorized smoking variables, we were able to estimate the adjusted means of outcomes for 25 combinations of smoking duration and cigarettes/day, and observe the patterns of adjusted means over different levels of duration groups for a given level of cigarettes/day group, and similarly for cigarettes/day at different levels of duration groups. This would not be achievable by using continuous smoking variables. We estimated linear slopes of outcomes using univariate regression models to estimate the average change in the adjusted means of outcome per each change in category, going from least exposure to highest exposure. As the linear slopes estimated the average

change in the adjusted outcome per change in category, we did not have to correct for the size of the category interval (10 units for each smoking variable). In addition, linear trends were estimated for the adjusted means of outcomes over pack-years to compare the estimated linear slopes over pack-years. Steepness of linear slopes over pack-years was illustrated in a graphical form with linear slopes over duration and cigarettes/day for each of the outcomes. Residual plots were checked in order to validate estimated linear models. Results were determined to be statistically significant when the accompanying statistical test yields a probability of <0.05. All analyses were performed using SAS 9.4 (SAS Institute, Cary, NC).

Supplemental Table 1: Adjusted means of percent FEV$_1$/FVC by combined categories of cigarettes/day and smoking duration

		Smoking Duration				
		0 to 10	10.1 to 20	20.1 to 30	30.1 to 40	≥40.1
Cigarettes/day	0 to 10	* (n=0)	0.75 (n=28)	0.71 (n=189)	0.68 (n=547)	0.67 (n=383)
	10.1 to 20	0.77 (n=40)	0.77 (n=405)	0.72 (n=782)	0.67 (n=2159)	0.63 (n=1830)
	20.1 to 30	0.76 (n=22)	0.76 (n=115)	0.71 (n=323)	0.66 (n=838)	0.61 (n=721)
	30.1 to 40	0.76 (n=19)	0.78 (n=110)	0.71 (n=207)	0.64 (n=513)	0.59 (n=513)
	≥40.1	* (n=5)	0.74 (n=36)	0.68 (n=77)	0.66 (n=163)	0.68 (n=163)

FEV1 = Forced expiratory volume in the first second. FVC = Forced vital capacity.

Adjusted least square means of percent CT emphysema are adjusted for age, race, sex, body-mass-index, scanner type, center, age of smoking onset and current smoking status.

*Numbers insufficient.

Supplemental Table 2: Smoking burden by cigarettes/day and duration across GOLD stages.

GOLD Stage	Pack-years of Smoking	Smoking Duration (years)	Cigarettes/day
0 (n = 1700)	31.9 (17.0)	30.2 (9.6)	21.6 (9.6)
1 (n = 333)	38.6 (19.8)	35.0 (10.1)	22.5 (10.5)
2 (n = 669)	43.2 (23.7)	36.5 (10.1)	24.9 (10.3)
3 (n = 435)	46.0 (20.7)	37.5 (9.5)	24.4 (10.8)
4 (n = 217)	46.0 (22.8)	37.7 (8.1)	22.7 (10.4)
PRISm (n = 442)	36.3 (20.2)	32.6 (9.6)	22.5 (10.9)

All values expressed as mean (standard deviation).

GOLD = Global Initiative for Chronic Obstructive Lung Disease. PRISm = GOLD unclassifiable or Preserved Ratio Impaired Spirometry.

Supplemental Table 3: Estimated linear trends of adjusted means of the outcomes over categorized Smoking Variables for participants ≥18 years of age (n= 3817)

	Smoking Duration		Cigarettes/day		Pack-years of Smoking	
	b(SE)	p value	b(SE)	p value	b(SE)	p value
FEV_1/FVC	-0.936 (0.077)	<0.001	-0.115 (0.216)	0.602	-0.964 (0.154)	0.008
FEV_1	-0.876 (0.105)	<0.001	-0.147 (0.216)	0.505	-0.948 (0.183)	0.014
CT Emphysema	0.902 (0.0.094)	<0.001	0.143 (0.217)	0.514	0.989 (0.086)	0.001
CT Gas Trapping	0.919 (0.086)	<0.001	0.229 (0.212)	0. 294	0.975 (0.128)	0.005
SGRQ	0.751 (0.144)	<0.001	0.537 (0.184)	0.008	0.966 (0.150)	0.008
6MWD	-0.845 (0.117)	<0.001	-0.307 (17.656)	0.208	--0.977 (0.124)	0.004

All values represent estimated standardized regression b (standard error, SE).

All outcome least square means adjusted for age, race, sex, body-mass-index, scanner type, center, age of smoking onset and current smoking status.

CT = computed tomography. FEV_1 = forced expiratory volume in the first second. FVC = forced vital capacity. SGRQ = St George's Respiratory Questionnare. 6MWD = six minute walk distance.

Supplemental Table 4: Estimated linear trends of adjusted means of the outcomes over categorized Smoking Variables for participants without adjustment for age of smoking onset (n= 10,187)

	Smoking Duration		Cigarettes/day		Pack-years of Smoking	
	b(SE)	p value	b(SE)	p value	b(SE)	p value
FEV$_1$/FVC	-0.922 (0.084)	<0.001	-0.207 (0.214)	0.342	-0.982 (0.109)	0.003
FEV$_1$	-0.909 (0.091)	<0.001	-0.215 (0.213)	0.323	-0.969 (0.143)	0.007
CT Emphysema	0.930 (0.080)	<0.001	0.183 (0.215)	0.404	0.990 (0.082)	<0.001
CT Gas Trapping	0.926 (0.082)	<0.001	0.229 (0.212)	0.293	0.980 (0.116)	0.004
SGRQ	0.828 (0.122)	<0.001	0.463 (0.194)	0.026	0.978 (0.121)	0.004
6MWD	-0.882 (0.103)	<0.001	-0.292 (0.209)	0.176	-0.982 (0.109)	0.003

All values represent estimated standardized regression b (standard error, SE). All outcome least square means adjusted for age, race, sex, body-mass-index, scanner type, center, and current smoking status.

CT = computed tomography. FEV$_1$ = forced expiratory volume in the first second. FVC = forced vital capacity. SGRQ = St George's Respiratory Questionnare. 6MWD = six minute walk distance.