

**Prevalence, Risk Factors, and Mortality of COPD in Young People in the United
States
Online Data Supplement**

Additional detail on the method

Study population

In our analysis, we included only participants aged 20 to 50 in the baseline survey—that is, 9234 participants from the NHANES 2007 to 2012 datasets, given that sufficient personal information and examination results were available to observe the characteristic of COPD in young people. We excluded participants who had an interview but did not participate in the pulmonary function test (n= 1450). Finally, a total of 7654 young participants without spirometry-defined COPD and 130 young COPD patients were enrolled in our cohort for analysis (eFigure 1) and followed up to Dec 31, 2019.

This cohort study used only publicly available data and was exempt from institutional review board review. Moreover, this cohort study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline.

Assessments of COPD, covariates, and deaths

Personal FEV₁ predicted value and LLN of FEV₁/FVC were calculated with age, sex, race, and height according to the latest Global Lung Function Initiative (GLI) equations for spirometry.

The classification of GOLD stage was based on post-bronchodilator FEV₁ predicted value (i.e., GOLD1 FEV₁≥80%predicted; GOLD2 50%≤FEV₁<80%predicted; GOLD3 30%≤FEV₁<50%predicted; GOLD4 FEV₁<30%predicted).

Positive in bronchodilator response (BDR) was defined as an increase in FEV₁ of

12% predicted or more and 200 mL or more.

Asthma-COPD Overlap Syndrome (ACOS) was defined as COPD patients satisfied two major criteria (increase in FEV₁ of 15% or more and 400 mL or more, and a personal history of asthma).⁽¹⁾

The public-use linked mortality files provide follow-up data from the date of baseline through December 31, 2019. The primary outcome of our cohort study was all-cause mortality. Details of the public-use linked mortality files are available from the National Center for Health Statistics.

1.5% of data (2/130) on height was missed in the COPD group, and the average value of the corresponding group was used to fill this gap. We did not fill other gaps in the database.

Statistical analysis

The prevalence of COPD in young people was stratified by age and gender. We fitted linear regression models using age groups as continuous variables to evaluate trends over time. The association between each potential risk factor and the presence of COPD was determined using univariable logistic regression. And the significant risk factors determined in the univariable analysis were enrolled into the multivariable logistic regression model to determine the independent risk factors. The relationship between risk factors and young COPD were presented by Odd Ratio (OR).

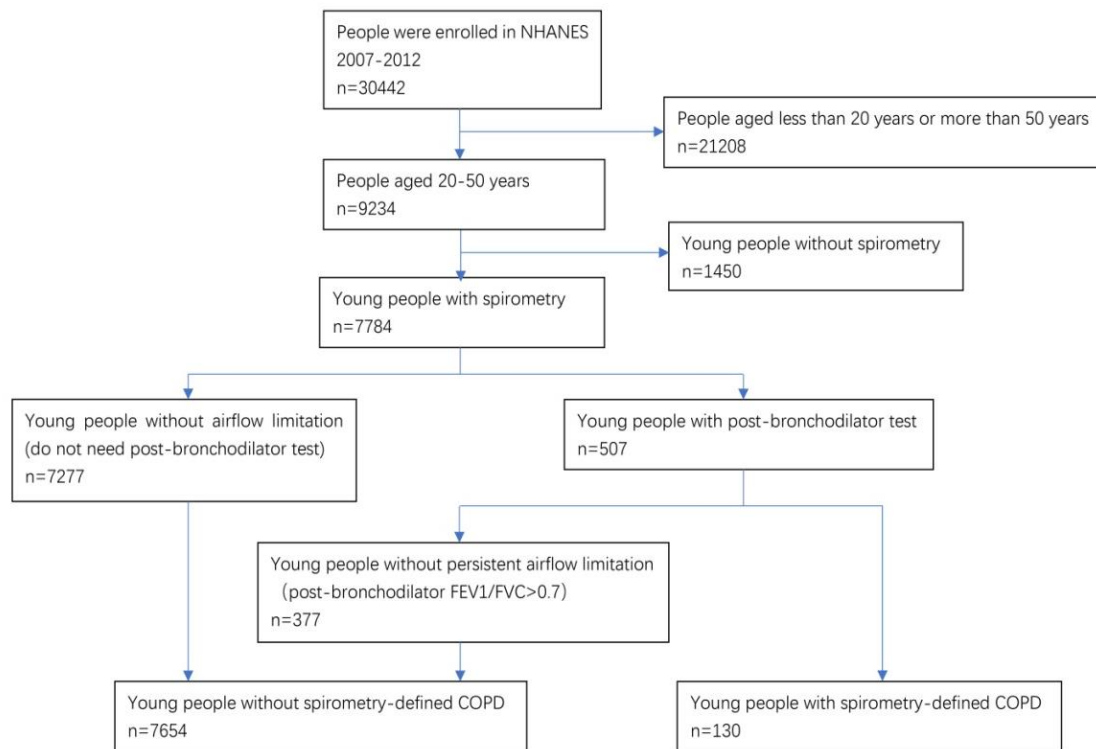
The propensity score method was used to account for the baseline imbalance. A logistic regression model was adapted to generate a propensity score for each participant based on confounding factors. Demographic characteristics (sex, age, and

race) and exposure history (smoking status, passive smoking, and occupational exposure) were introduced in this logistic regression model. Because of the large sample size of non-COPD groups, a 1:5 matching procedure (one COPD patient with five participants without COPD) was performed using nearest neighbor matching on the propensity score.

Cox proportional hazards models were applied to calculate the hazard ratios and corresponding 95% confidence intervals for all-cause mortality between COPD and non-COPD groups. The baseline of survival analysis was defined as the time when participants had their physical examinations. Months from baseline to the date of mortality, loss to follow-up, or 31 December 2019, was counted. For the main analyses, we assessed the hazard ratios (HR) of mortality for the COPD group compared with sex, age, and race-matched population. We also investigated the HR of mortality for the COPD group compared with sex, age, race, smoking, and occupational exposure matched population.

Reference

1. Soler-Cataluña JJ, Cosío B, Izquierdo JL, et al. Consensus document on the overlap phenotype COPD-asthma in COPD. *Archivos de bronconeumologia* 2012;48:331-337.

e-Figure 1. Flow chart of study population enrollment

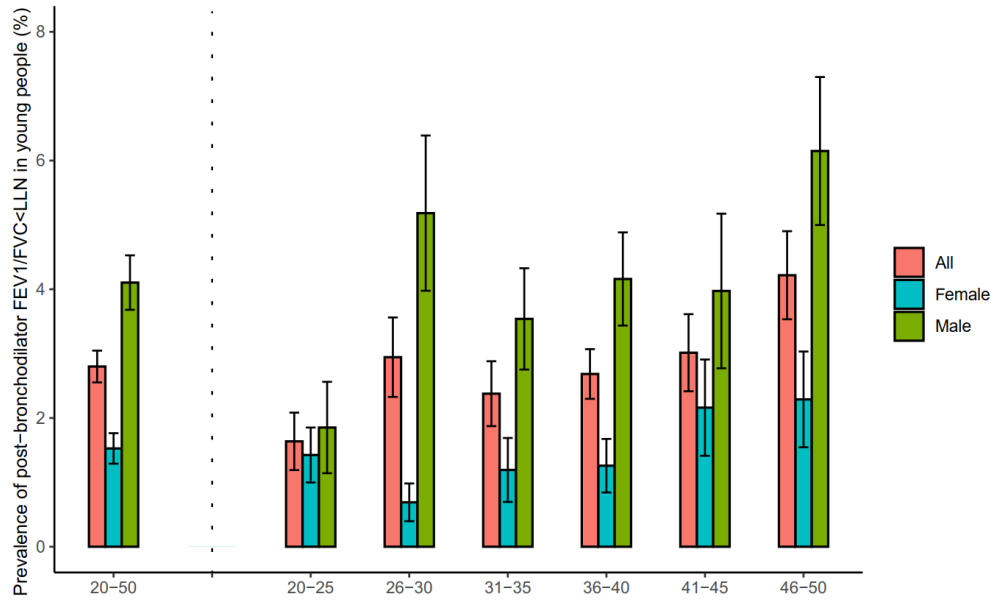
Abbreviations: NHANES, the National Health and Nutrition Examination Survey; FEV₁, forced expiratory volume in the first second; FVC, forced vital capacity; COPD, chronic obstructive pulmonary disease.

e-Table 1. Demographic and clinical characteristics of FEV₁/FVC<LLN and non-FEV₁/FVC<LLN groups in young people in the NHANES

	level	non- FEV ₁ /FVC<LLN	FEV ₁ /FVC<LLN	<i>P</i>
Sex (%)	Female	63769064 (51.2)	998355 (27.6)	<0.001
	Male	60814631 (48.8)	2620270 (72.4)	
Age (mean [SD])		35.18 (9.11)	37.58 (8.89)	0.01
Race (%)	Mexican American	13665508 (11.0)	160247 (4.4)	<0.001
	Non-Hispanic Black	15566182 (12.5)	238975 (6.6)	
	Non-Hispanic White	76866589 (61.7)	2946060 (81.4)	
	Other Hispanic	8626742.8 (6.9)	40801.9 (1.1)	
	Other Race	9858674.5 (7.9)	232541.0 (6.4)	
BMI (mean [SD])		28.43 (6.86)	27.29 (5.67)	0.009
Family income to poverty (mean [SD])		2.82 (1.68)	2.81 (1.72)	0.938
Smoking status (%)	Current	30219189(24.3)	1875281(51.8)	<0.001
	Ever	20264204 (16.3)	687216 (19.0)	
	Never	74035961.4 (59.4)	1056129 (29.2)	
	Not available	64340.9 (0.1)	0.0 (0.0)	
Passive smoking (%)	Yes	19743733 (15.8)	1383781 (38.2)	<0.001
	No	104179812.3 (83.6)	2234844.7 (61.8)	
	Not available	660150.7 (0.5)	0.0 (0.0)	
Occupational exposure (%)	Yes	64529981 (51.8)	2444116 (67.5)	<0.001
	No	56174315 (45.1)	1142728 (31.6)	
	Not available	3879401 (3.1)	31782 (0.9)	
Self-reported asthma (%)	Yes	18019586 (14.5)	773199 (21.4)	0.009
	No	106428194 (85.4)	2823676 (78.0)	
	Not available	135915.6 (0.1)	21751 (0.6)	
Self-reported emphysema (%)	Yes	695925 (0.6)	64197 (1.8)	<0.001
	No	123846913 (99.4)	3520920 (97.3)	
	Not available	40858 (0.0)	33509 (0.9)	
Self-reported chronic bronchitis (%)	Yes	4872678 (3.9)	229370 (6.3)	0.122
	No	119610976 (96.0)	3378361 (93.4)	
	Not available	100042 (0.1)	10894 (0.3)	
Self-reported cancer (%)	Yes	3998777 (3.2)	267476 (7.4)	0.615
	No	120452171 (96.7)	3351150 (92.6)	
	Not available	132747.2 (0.1)	0.0 (0.0)	
Self-reported congestive heart failure (%)	Yes	682331.8 (0.5)	37463 (1.0)	0.076
	No	123845278 (99.4)	3581163 (99.0)	
	Not available	56085.4 (0.0)	0.0 (0.0)	

Shortness of breath (%)	Yes	3998777 (3.2)	267475.6 (7.4)	0.004
	No	33962621 (27.3)	1021948 (28.2)	
	Not available	78737027(63.2)	1876491 (51.9)	
Chronic cough (%)	Yes	4277865 (3.4)	267861 (7.4)	0.005
	No	41536352 (33.3)	1474274 (40.7)	
	Not available	78769479(63.2)	1876490.5 (51.9)	
Chronic sputum (%)	Yes	2934965 (2.4)	271778 (7.5)	<0.001
	No	40602862 (32.6)	1417412 (39.2)	
	Not available	81045869 (65.1)	1929435 (53.3)	
Wheezing or whistling (%)	Yes	2934965 (2.4)	271778 (7.5)	<0.001
	No	2934965 (2.4)	271778 (7.5)	
	Not available	2934965 (2.4)	271778 (7.5)	
Dry cough at night (%)	Yes	6072453 (4.9)	339703 (9.4)	0.027
	No	2934965 (2.4)	271778 (7.5)	
	Not available	2934965 (2.4)	271778 (7.5)	
preFEV ₁ (mean [SD])		3540 (816.29)	3162 (824.18)	<0.001
preFEV ₁ /FVC (mean [SD])		0.81 (0.07)	0.64 (0.07)	<0.001

Abbreviations: LLN, lower limit of normal; BMI, body mass index; preFEV₁, pre-bronchodilator forced expiratory volume in the first second; preFEV₁/FVC, pre-bronchodilator forced expiratory volume in the first second/forced vital capacity. NHANES results were estimated from statistical study weights. Data are presented as mean (SD) for continuous measures and n (%) for categorical measures. T-test was used for continuous variables, and chi square of Fisher exact test was used for categorical variables.

e-Figure 2. Prevalence of post-bronchodilator FEV₁/FVC < LLN in young people

Abbreviations: LLN, lower limit of normal.

Prevalence was estimated using sample weights from NHANES.