

**Table S1.** ICD codes and semantic phrases considered in algorithm development

Inclusion Criteria for Cases	Additional Exclusion Criteria for Controls
ICD-9 Codes	ICD-9 Codes
491.xx: Chronic bronchitis	493.xx: Asthma
492.xx: Emphysema	516.31: Idiopathic pulmonary fibrosis
496.xx: Chronic airway obstruction, not elsewhere classified	135.xx: Sarcoidosis
162.xx: Neoplasm of the trachea, bronchus, and lung	
ICD-10 Codes	ICD-10 Codes
J41.xx: Simple and mucopurulent chronic bronchitis	J45.xx: Asthma
J42.xx: Unspecified chronic bronchitis	J84.112: Idiopathic pulmonary fibrosis
J43.xx: Emphysema	D86.xx: Sarcoidosis
J44.xx: Other chronic obstructive pulmonary disease	
C34.xx: Malignant neoplasm of bronchus and lung	
Problem list text	
COPD	
Chronic obstructive pulmonary disease	
Chronic obstructive lung disease	
Emphysema	
Chronic bronchitis	
Reactive airway(s) disease	
Exacerbation	
Oxygen	
O2	
Radiology report text	
Emphysema (excluding subcutaneous emphysema)	
Emphysematous changes	
Respiratory medications	
Accuneb	
Acridinium	

Advair
Albuterol
Alvesco
Anoro
Arcapta
Arformoterol
Asmanex
Atrovent
Beclomethasone
Beclovent
Breo
Brovana
Budesonide
Ciclesonide
Combivent
Daliresp
Dulera
Duoneb
Entocort
Flovent
Fluticasone
Foradil
Glycopyrrolate
Indaceterol
Ipratropium
Lasma
Levalbuterol
Mometasone
Olodaterol
Perforomist
ProAir
Proventil
Pulmicort
QVar
Robinul
Roflumilast
Salbutamol

Salmeterol
Serevent
Singulair
Spiriva
Symbicort
Theo
Theochron
Theodur
Theophylline
Tiotropium
Tudorza
Umeclidinium
Uniphyll
Vanceril
Ventolin
Vilanterol
Xopenox

**Table S2.** Numbers of cases, controls, indeterminates, and excludes identified among adults over 45 years of age in the Synthetic Derivative (N = 1,008,552).

<b>Case definition*</b>	<b>Cases, N</b>	<b>Control, N</b>	<b>Indeterminate, N</b>	<b>Exclude, N</b>
3+ ICD codes	28,520	544,056	23,091	28,319
ICD code+reg-expr	12,622	544,056	38,989	28,319
PFT	1,873	11,203	NA	NA

PFT = pulmonary function test, reg ex = regular expression

\*All case definitions also include the requirement that cases have records longer than 180 days.

**Table S3.** Demographic characteristics of COPD cases identified by PFT, 3+ ICD codes, and Code + regex algorithm.

<b>Variable</b>	<b>PFT cases* N = 1,873</b>	<b>3+ COPD ICD Cases N = 28,520</b>	<b>Code+Regex Cases N = 12,622</b>
Median age (IQR)	70 (62-76)	69 (60-76)	70 (62-77)
Sex (%)			
Female	911 (48.6)	13,356 (46.8)	6,163 (48.8)
Male	962 (51.4)	15,164 (53.2)	6,459 (51.2)
Unknown	0 (0.0)	0 (0.0)	0 (0.0)
Race (%)			
White	1,701 (90.8)	24,795 (86.9)	11,012 (87.2)
Black	135 (7.2)	2,684 (9.4)	1,249 (9.9)
Other	15 (0.8)	119 (0.4)	46 (0.4)
Unknown	22 (1.2)	922 (3.2)	315 (2.5)
Smoking status			
Ever	1,464 (78.2)	15,645 (54.9)	7,548 (59.8)
Never	402 (21.5)	2,900 (10.2)	1,034 (8.2)
Missing	7 (0.4)	9,975 (35.0)	4,040 (32.0)
Median record length (years) (IQR)	10.3 (5.1-16.1)	7.3 (3.2-13.0)	8.3 (4.0-13.8)
Number of respiratory meds** (median, IQR)	188 (72-385)	107 (32-277)	170 (56-387)
Top 5 phecodes by frequency (%)			

1	Other symptoms of respiratory system (88.7)	Chronic airway obstruction (100)	Chronic airway obstruction (100)
2	Chronic airway obstruction (78.0)	Hypertension (77.1)	Other symptoms of respiratory system (86.1)
3	Shortness of breath (71.8)	Other symptoms of respiratory system (75.4)	Hypertension (80.3)
4	Hypertension (71.8)	Essential hypertension (75.4)	Essential hypertension (78.9)
5	Essential hypertension (70.7)	Tobacco use disorder (63.0)	Shortness of breath (73.5)

\*Pulmonary function testing (PFT) cases are defined as individuals with a post-bronchodilator FEV<sub>1</sub>/FVC ratio < 0.7.

\*\*Number of respiratory medications includes only mentions of medications on distinct days.

**Table S4.** Association of known clinical risk factors with COPD in each algorithm set and in the PFT set.

Risk factor	3+ code only		Code + regex		PFTs	
	OR	95% CI	OR	95% CI	OR	95% CI
Age	1.04	1.04-1.04	1.05	1.05-1.05	1.04	1.03-1.04
Sex	1.40	1.37-1.43	1.27	1.23-1.31	1.22	1.11-1.34
Ever smoking	9.22	8.87-9.57	12.36	11.61-13.16	3.36	3.00-3.75

OR: odds ratio; 95% CI: 95% confidence interval

**Table S5.** Comparison of demographics and clinical features for cases (post-bronchodilator FEV<sub>1</sub>/FVC < 0.7) and controls identified by PFTs.

<b>Characteristic</b>	<b>Cases* N = 1,873</b>	<b>Controls* N = 11,203</b>	<b>All PFT* N = 13,076</b>
Median age (IQR)	70 (62-76)	65 (57-72)	65 (57-73)
Sex (%)			
Female	911 (48.6)	6,023 (53.8)	6,934 (53.0)
Male	962 (51.4)	5,179 (46.2)	6,141 (47.0)
Unknown	0 (0.0)	1 (0.0)	1 (0.0)
Race (%)			
White	1,701 (90.8)	9,717 (86.7)	11,418 (87.3)
Black	135 (7.2)	1,201 (10.7)	1,336 (10.2)
Other	15 (0.8)	162 (1.4)	177 (1.4)
Unknown	22 (1.2)	123 (1.1)	145 (1.1)
Smoking status			
Ever	1,464 (78.2)	5,779 (51.6)	7,243 (55.4)
Never	402 (21.5)	5,359 (47.8)	5,761 (44.1)
Missing	7 (0.4)	65 (0.6)	72 (0.5)
Median record length (years) (IQR)	10.3 (5.1-16.1)	9.5 (4.5-15.5)	9.6 (4.6-15.6)
Number of respiratory meds** (median, IQR)	188 (72-385)	168 (66-372)	171 (67-374)
Top 5 phecodes by frequency (%)			
1	Other symptoms of respiratory system (88.7)	Other symptoms of respiratory system (84.5)	Other symptoms of respiratory system (85.1)

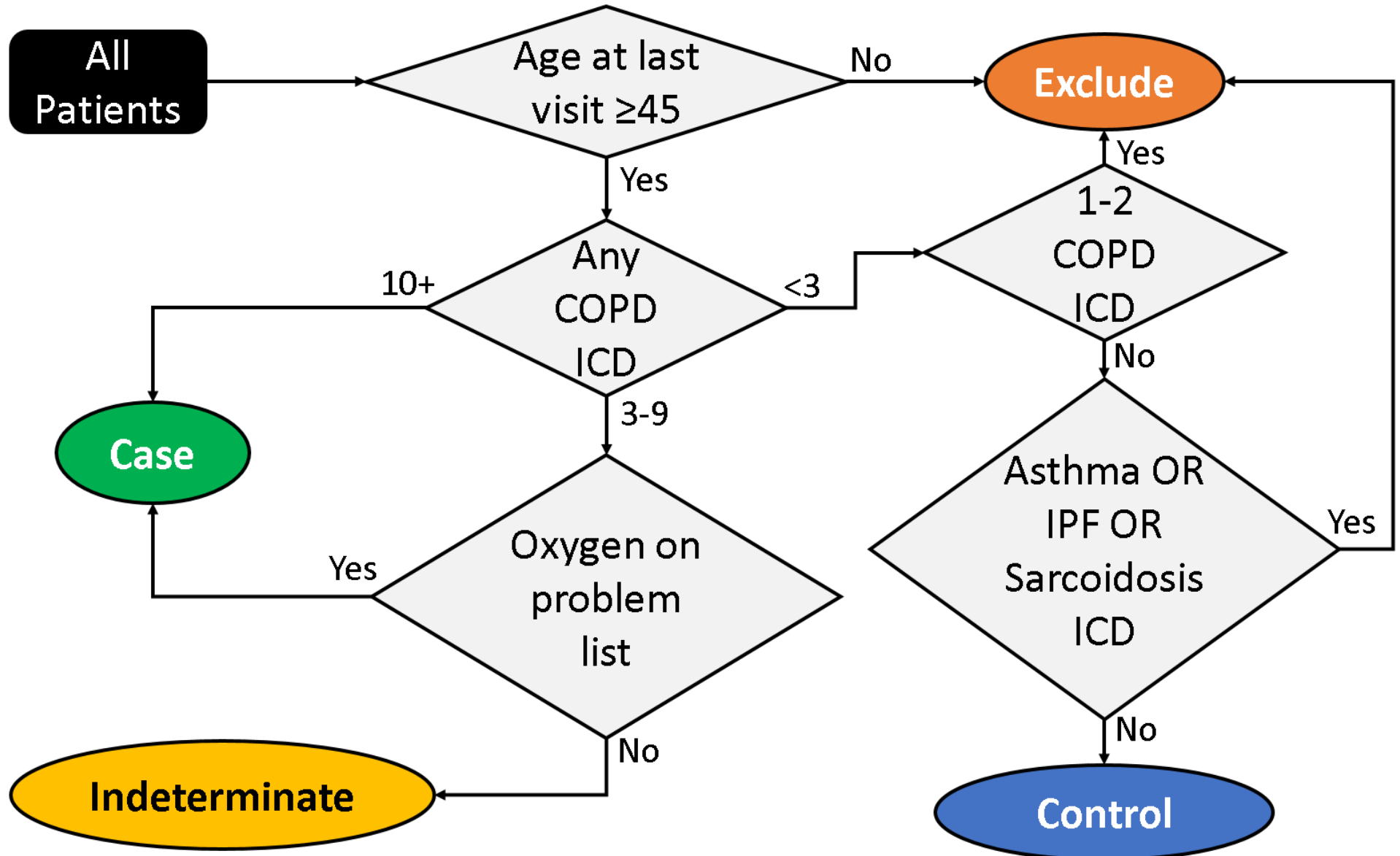


2	Chronic airway obstruction (78.0)	Hypertension (66.7)	Hypertension (67.1)
3	Shortness of breath (71.8)	Essential hypertension (65.6)	Essential hypertension (66.3)
4	Hypertension (71.8)	Shortness of breath (64.1)	Shortness of breath (65.2)
5	Essential hypertension (70.7)	Diseases of esophagus (54.9)	Disorders of lipid metabolism (55.2)

\*Cases, controls, and all PFT data sets were restricted to individuals with record lengths greater than 180 days to be comparable to those identified by the phenotyping algorithm.

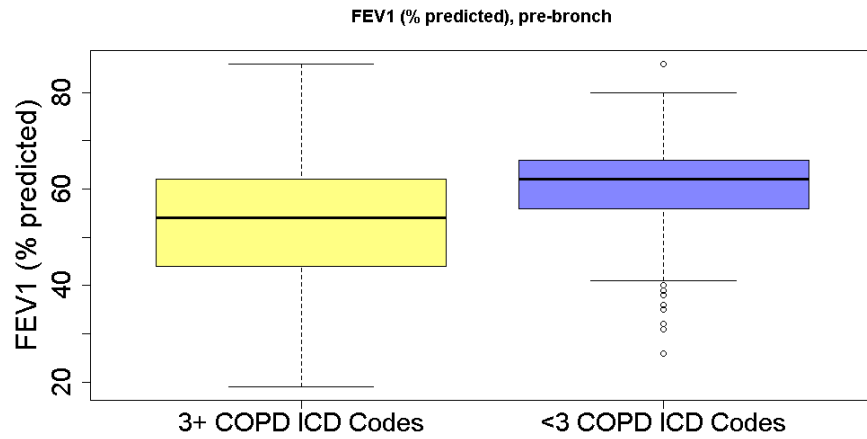
\*\*Number of respiratory medications includes only mentions of medications on distinct days.

Figure S1. Phenotyping algorithm with highest PPV in the chart review validation set.

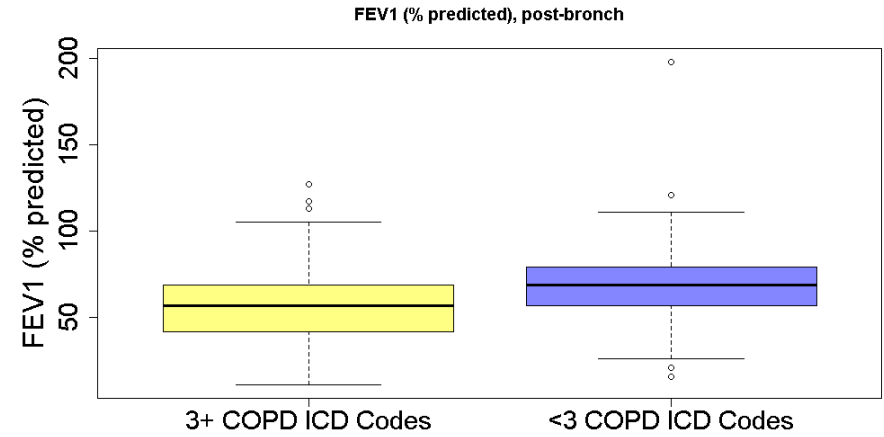


**Figure S2.** Comparison of PFT-defined cases with 3+ ICD codes for COPD and those with fewer than three ICD codes for COPD among individuals with record lengths greater than 180 days. A) Forced expiratory volume in 1 second, pre-bronchodilator. B) Forced expiratory volume in 1 second, post-bronchodilator. C) Forced expiratory volume in 1 second/forced vital capacity, pre-bronchodilator. D) Forced expiratory volume in 1 second/forced vital capacity, post-bronchodilator.

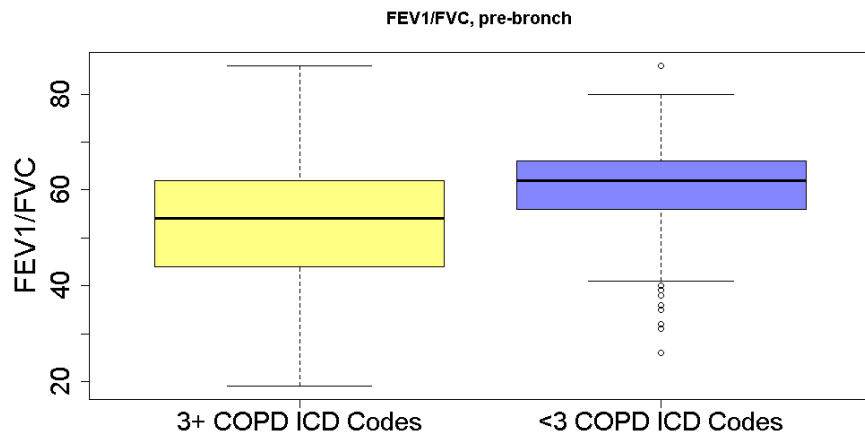
A)



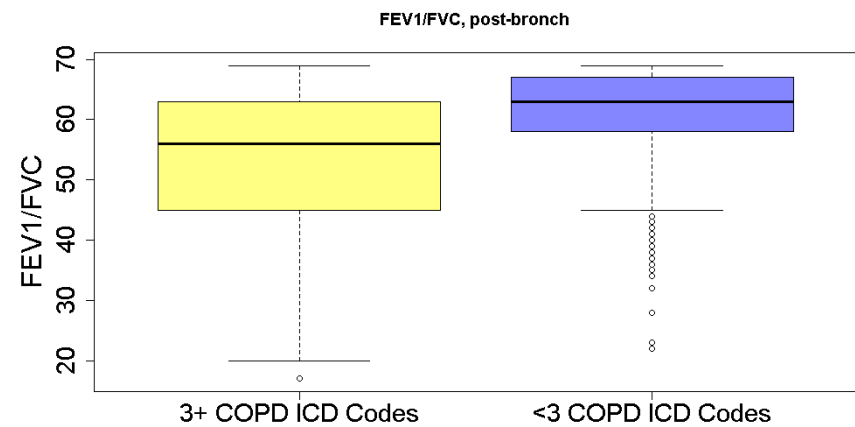
B)



C)



D)



**Figure S3.** Odds ratios from genetic associations of COPD cases (3+ code) and controls in BioVU compared with associations of select variants identified in a prior GWAS of COPD by Wain et al. (32)

