

**Supplemental Digital Content**

**eTable 1. Study Population (unweighted frequencies).**

						COPD status among those 40-79 years of age with data on smoking history and completed pre- bronchodilator spirometry	
	Overall	40-79 years of age	Currently smokes	Previously smoked	Never smoked	COPD +	COPD -
Interviewed	30,442	10,536	2,149	3,049	5,328	--	--
Participated in MEC exam	29,353	10,219	2,089	2,960	5,163	--	--
Considered for pre- bronchodilator spirometry	23,433	10,219	2,089	2,960	5,163	--	--
<i>Excluded due to safety reason<sup>a</sup></i>	1,342	896	226	330	340	--	--
<i>Test unreliable / Partial Exam</i>	1,208	479	101	102	276	--	--
<i>Not done</i>	1,779	842	146	204	491	--	--
Completed pre- bronchodilator spirometry	19,104	8,002	1,616	2,324	4,056	1,270	6,726
<i>FEV<sub>1</sub>/FVC ratio &lt;0.7</i>	1,840	1,517	517	578	421	1,245	271
<i>FEV<sub>1</sub>/FVC ratio &lt;LLN</i>	1,858	948	405	316	227	821	127
Eligible for post- bronchodilator spirometry	2,769	1,606	541	589	475	1,263	342
<i>Excluded due to safety reason<sup>b</sup></i>	608	450	144	188	118	437	13
<i>Test unreliable / Partial Exam</i>	9	4	2	1	1	3	1
<i>Not done</i>	597	384	147	130	106	360	23

Completed post-bronchodilator spirometry	1,555	768	248	270	250	463	305
<i>FEV<sub>1</sub>/FVC ratio &lt;0.7</i>	518	459	176	173	110	459	0
<i>FEV<sub>1</sub>/FVC ratio &lt;LLN</i>	460	281	140	94	47	281	0

Abbreviations: CODP, chronic obstructive pulmonary disease; MEC, mobile examination center; FEV<sub>1</sub>, forced expiratory volume in one second; FVC, forced vital capacity; LLN, lower limit of normal

<sup>a</sup> Pre-bronchodilator spirometry exclusions included: current painful ear infection (only if 6-15 years of age); eye, open chest, or abdominal surgery in preceding 3 months; participant or household member with tuberculosis exposure; history of aneurysm or detached retina; history of collapsed lung; stroke or heart attack in preceding 3 months; breathing problems that requires supplemental oxygen; physical problem that precludes taking a deep breath; or history of hemoptysis in previous month.

<sup>b</sup> Post-bronchodilator spirometry exclusions included: recent use of a β2-adrenergic bronchodilator; previous adverse reaction to albuterol; current pregnancy or breastfeeding; congenital heart disease (only if 6-15 years of age); diagnosed major arrhythmia; elevated blood pressure for age; resting tachycardia; irregular pulse; currently taking certain medications (class 1 antiarrhythmic, implanted automatic defibrillator, monoamine oxidase inhibitor, anticonvulsant medication for epilepsy, or diuretic therapy without potassium supplementation/raising medication); currently taking tricyclic antidepressant (only if over 40 years of age and being treated for heart disease, kidney disease, or a thyroid disorder).

## eAppendix 1. Details on gradient boosting model.

### Study Design Variables

General formulas for estimating model parameters assume a simple random sampling (SRS) design. Unbiased estimation from data arising from other sampling design requires accommodation of sampling design variables into the estimation procedures. NHANES uses three design variables to describe the sampling procedures: observation weights, strata, and pseudo-primary sampling unit.<sup>1</sup> In the NHANES sampling process, U.S. counties are first grouped into strata (typically two counties) based on geographic and demographic similarity. This increases the observed variability of sampled units by ensuring that two similar counties will not both be sampled. Next, a population-weighted sample of strata and a population-weighted sample of counties (PSU clusters) within each strata are selected. Within each PSU cluster, individuals are randomly selected within city blocks and households to be sampled effectively at random.

### Estimation of population mean and variance from non-SRS sampling designs

Accommodation of sampling designs other than the SRS can be divided into two problems: weighted estimation and design-based variance estimation.<sup>2</sup> Weighted estimation, for instance the Horvitz-Thompson estimator for the mean, is sufficient to produce unbiased non-parametric estimators in a stratified sample.<sup>3</sup> Analogously, generalized linear regression models require weight-based adjustment for unbiased estimation of parameter values.<sup>4</sup> This estimation can be accomplished by applying inverse probability-of-sampling weights to non-parametric calculations or applying observation weights directly to the computation of the pseudo-likelihood. Parameter estimates corrected with design weights using these approaches are consistent even without specification of strata and clustering variables in complex survey designs.<sup>5</sup>

In contrast, estimation of the sampling variance of model parameters (used to estimate confidence intervals) in generalized linear regression must take observation strata and clustering variables into account because the variance-covariance matrix is only equal to the second derivative of the likelihood function under the assumption of simple random sampling.<sup>4</sup> For consistency, parameter variance must be calculated using alternate methods: (1) a robust sandwich estimator derived from a Taylor series approximation of the variance-covariance matrix summing variance components within clusters and strata, or (2) replication-based methods to estimate the design-based variance-covariance matrix.

### **Survey design adjustments in machine learning**

Like other statistical methods, machine learning algorithms can incorporate design variables to produce unbiased estimates for non-SRS sampling designs. Generally, these algorithms use multiple (parallel or serial) decision points to create classification schema for observed data using available covariates and identified cut points. With regard to survey sampling, machine learning algorithms can be divided into parametric and non-parametric categories. Parametric approaches, like classification and regression trees, rely on both model estimates and sampling variance to make learning decisions. Sampling variance of parameters is used to build a learning structure and also to simplify the learning structure to a more parsimonious form. For example, the parametric classification tree approach implemented in the *rmps* package makes branching decisions based on the relative statistical significance of potential classification schemes.<sup>6</sup>

Non-parametric approaches, like the gradient boosting approach used in this analysis, do not rely on parameter sampling variance to make learning decisions or as stopping criteria. Instead, these approaches rely on the relative magnitude of estimates to make learning decisions and use other methods like regularization to make decisions about what should and should not be included in the

learning structure. In the case of xgboost, tree growth decisions are made based on maximizing the loss function, which depends only on the absolute impurity within each of the leaf nodes produced by a hypothetical classification tree.<sup>7</sup> Regularization is handled by specifying the minimum improvement in the loss function necessary to create a new split as well as the observation weight necessary to retain a new split. Following this logic, we have incorporated NHANES sampling weights into the estimation procedure, but the learning process is invariant to the sampling variance of parameter weights (cluster and strata variables).

**eTable 2. Prevalence of COPD and classification parameters for questionnaire-based definitions, former and current smokers separately.**

Approach	Case Definition	Prevalence (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy <sup>a</sup> (%)
<i>Former smokers (n=2,324)</i>							
Gradient Boosting	<ul style="list-style-type: none"> <li>• Age ≥55, and any one of</li> <li>• Bronchodilator use, or</li> <li>• BMI &lt;32 and &gt;36 pack-year history, or</li> <li>• BMI &lt;32 and &gt;18 years organic dust exposure</li> </ul>	13 (11-16)	32 (26-39)	92 (90-93)	50 (44-57)	84 (81-86)	79 (77-82)
Mendy et al. <sup>8</sup>	<ul style="list-style-type: none"> <li>• Affirmative response to physician-diagnosed chronic bronchitis or emphysema</li> </ul>	9 (7-10)	16 (11-21)	94 (92-95)	41 (31-51)	80 (78-82)	77 (74-79)
Fessler et al. <sup>9</sup>	<ul style="list-style-type: none"> <li>• Affirmative response to physician-diagnosed chronic bronchitis or emphysema, or</li> <li>• ≥10 pack-year history of smoking and affirmative response to coughing up phlegm on most days for ≥3 consecutive months for ≥2 consecutive years</li> </ul>	11 (9-13)	20 (14-25)	92 (90-93)	39 (30-48)	80 (78-83)	76 (73-78)
<i>Current smokers (n=1,616)</i>							
Gradient Boosting	<ul style="list-style-type: none"> <li>• Age ≥55, and any one of</li> <li>• Bronchodilator use, or</li> <li>• BMI &lt;32 and &gt;36 pack-year history, or</li> <li>• BMI &lt;32 and &gt;18 years organic dust exposure</li> </ul>	17 (14-21)	38 (29-47)	92 (90-94)	69 (59-78)	76 (73-80)	75 (72-78)

Approach	Case Definition	Prevalence (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy <sup>a</sup> (%)
Mendy et al. <sup>8</sup>	<ul style="list-style-type: none"> <li>Affirmative response to physician-diagnosed chronic bronchitis or emphysema</li> </ul>	14 (11-17)	20 (13-26)	89 (86-91)	44 (33-55)	71 (68-74)	67 (64-70)
Fessler et al. <sup>9</sup>	<ul style="list-style-type: none"> <li>Affirmative response to physician-diagnosed chronic bronchitis or emphysema, or</li> <li>≥10 pack-year history of smoking and affirmative response to coughing up phlegm on most days for ≥3 consecutive months for ≥2 consecutive years</li> </ul>	24 (21-27)	35 (30-41)	82 (78-85)	47 (40-54)	73 (70-77)	67 (64-70)

Scale for Strength of Classification Parameters

The scale is a horizontal bar divided into three segments: white (0-25), light blue (25-75), and dark blue (75-100). The numbers 0, 50, and 100 are placed at the start, midpoint, and end of the bar respectively.

Abbreviations: COPD, chronic obstructive pulmonary disease; BMI, body mass index; PPV, positive predictive value; NPV, negative predictive value

<sup>a</sup> Accuracy is the proportion of true classifications (true positives + true negatives) among all classifications (true positives + true negatives + false positives + false negatives).

**eTable 3. Prevalence of COPD and classification parameters for questionnaire-based definitions, stratified by sex.**

Approach	Case Definition	Prevalence (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy <sup>a</sup> (%)
<b><i>Former and current smokers – female (n=1,618)</i></b>							
Gradient Boosting	<ul style="list-style-type: none"> <li>• Age ≥55, and any one of</li> <li>• Bronchodilator use, or</li> <li>• BMI &lt;32 and &gt;36 pack-year history, or</li> <li>• BMI &lt;32 and &gt;18 years organic dust exposure</li> </ul>	11 (9-14)	33 (25-41)	94 (92-96)	61 (51-72)	84 (81-87)	81 (78-84)
Mendy et al. <sup>8</sup>	<ul style="list-style-type: none"> <li>• Affirmative response to physician-diagnosed chronic bronchitis or emphysema</li> </ul>	13 (11-16)	22 (15-29)	89 (87-91)	35 (26-44)	81 (78-84)	75 (72-78)
Fessler et al. <sup>9</sup>	<ul style="list-style-type: none"> <li>• Affirmative response to physician-diagnosed chronic bronchitis or emphysema, or</li> <li>• ≥10 pack-year history of smoking and affirmative response to coughing up phlegm on most days for ≥3 consecutive months for ≥2 consecutive years</li> </ul>	17 (14-19)	28 (22-34)	86 (84-89)	36 (29-43)	81 (78-84)	74 (71-77)
<b><i>Never smokers – female (n=2,413)</i></b>							
Gradient Boosting	<ul style="list-style-type: none"> <li>• Age ≥55, and any one of</li> <li>• Bronchodilator use, or</li> <li>• BMI &lt;32 and &gt;36 pack-year history, or</li> <li>• BMI &lt;32 and &gt;18 years organic dust exposure</li> </ul>	4 (2-6)	10 (3-18)	96 (95-98)	17 (7-26)	94 (93-95)	91 (89-93)

<b>Approach</b>	<b>Case Definition</b>	<b>Prevalence (%)</b>	<b>Sensitivity (%)</b>	<b>Specificity (%)</b>	<b>PPV (%)</b>	<b>NPV (%)</b>	<b>Accuracy<sup>a</sup> (%)</b>
Mendy et al. <sup>8</sup>	<ul style="list-style-type: none"> <li>Affirmative response to physician-diagnosed chronic bronchitis or emphysema</li> </ul>	6 (5-7)	11 (4-18)	94 (93-96)	12 (3-22)	94 (92-95)	89 (87-91)
Fessler et al. <sup>9</sup>	<ul style="list-style-type: none"> <li>Affirmative response to physician-diagnosed chronic bronchitis or emphysema, or</li> <li>≥10 pack-year history of smoking and affirmative response to coughing up phlegm on most days for ≥3 consecutive months for ≥2 consecutive years</li> </ul>	6 (5-7)	11 (4-18)	94 (93-96)	12 (3-22)	94 (92-95)	89 (87-91)
<b><i>Former and current smokers – male (n=2,322)</i></b>							
Gradient Boosting	<ul style="list-style-type: none"> <li>Age ≥55, and any one of</li> <li>Bronchodilator use, or</li> <li>BMI &lt;32 and &gt;36 pack-year history, or</li> <li>BMI &lt;32 and &gt;18 years organic dust exposure</li> </ul>	18 (15-20)	36 (31-42)	89 (87-92)	58 (50-65)	78 (75-81)	74 (71-77)
Mendy et al. <sup>8</sup>	<ul style="list-style-type: none"> <li>Affirmative response to physician-diagnosed chronic bronchitis or emphysema</li> </ul>	9 (6-11)	15 (11-19)	94 (93-96)	52 (43-60)	73 (71-76)	72 (69-74)
Fessler et al. <sup>9</sup>	<ul style="list-style-type: none"> <li>Affirmative response to physician-diagnosed chronic bronchitis or emphysema, or</li> <li>≥10 pack-year history of smoking and affirmative response to coughing up phlegm on most days for ≥3</li> </ul>	15 (13-18)	27 (21-33)	89 (87-91)	51 (44-57)	75 (72-78)	71 (69-74)

Approach	Case Definition	Prevalence (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy <sup>a</sup> (%)
	consecutive months for ≥2 consecutive years						
<b><i>Never smokers – male (n=1,643)</i></b>							
Gradient Boosting	<ul style="list-style-type: none"> <li>• Age ≥55, and any one of</li> <li>• Bronchodilator use, or</li> <li>• BMI &lt;32 and &gt;36 pack-year history, or</li> <li>• BMI &lt;32 and &gt;18 years organic dust exposure</li> </ul>	4 (3-6)	10 (3-16)	96 (95-98)	20 (7-33)	92 (90-93)	88 (86-91)
Mendy et al. <sup>8</sup>	<ul style="list-style-type: none"> <li>• Affirmative response to physician-diagnosed chronic bronchitis or emphysema</li> </ul>	3 (2-4)	6 (0-12)	97 (96-98)	18 (0-37)	91 (89-93)	89 (87-91)
Fessler et al. <sup>9</sup>	<ul style="list-style-type: none"> <li>• Affirmative response to physician-diagnosed chronic bronchitis or emphysema, or</li> <li>• ≥10 pack-year history of smoking and affirmative response to coughing up phlegm on most days for ≥3 consecutive months for ≥2 consecutive years</li> </ul>	3 (2-4)	6 (0-12)	97 (96-98)	18 (0-37)	91 (89-93)	89 (87-91)
Scale for Strength of Classification Parameters							
0                          50                          100							

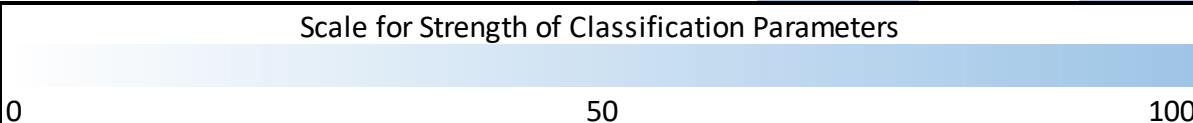
Abbreviations: COPD, chronic obstructive pulmonary disease; BMI, body mass index; PPV, positive predictive value; NPV, negative predictive value

<sup>a</sup> Accuracy is the proportion of true classifications (true positives + true negatives) among all classifications (true positives + true negatives + false positives + false negatives).

**eTable 4. Prevalence of COPD and classification parameters for questionnaire-based definitions, stratified by age.**

Approach	Case Definition	Prevalence (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy <sup>a</sup> (%)
<b><i>Former and current smokers – 40-59 (n=2,177)</i></b>							
Gradient Boosting	<ul style="list-style-type: none"> <li>• Age ≥55, and any one of</li> <li>• Bronchodilator use, or</li> <li>• BMI &lt;32 and &gt;36 pack-year history, or</li> <li>• BMI &lt;32 and &gt;18 years organic dust exposure</li> </ul>	6 (5-8)	18 (11-25)	96 (95-97)	52 (38-65)	84 (81-86)	82 (79-84)
Mendy et al. <sup>8</sup>	<ul style="list-style-type: none"> <li>• Affirmative response to physician-diagnosed chronic bronchitis or emphysema</li> </ul>	10 (8-12)	16 (11-22)	92 (90-94)	32 (22-42)	82 (80-85)	77 (75-80)
<b><i>Never smokers – 40-59 (n=2,427)</i></b>							
Gradient Boosting	<ul style="list-style-type: none"> <li>• Age ≥55, and any one of</li> <li>• Bronchodilator use, or</li> <li>• BMI &lt;32 and &gt;36 pack-year history, or</li> <li>• BMI &lt;32 and &gt;18 years organic dust exposure</li> </ul>	2 (1-3)	1 (0-2)	98 (97-99)	2 (0-5)	95 (94-96)	93 (92-95)

<b>Approach</b>	<b>Case Definition</b>	<b>Prevalence (%)</b>	<b>Sensitivity (%)</b>	<b>Specificity (%)</b>	<b>PPV (%)</b>	<b>NPV (%)</b>	<b>Accuracy<sup>a</sup> (%)</b>
Mendy et al. <sup>8</sup>	<ul style="list-style-type: none"> <li>Affirmative response to physician-diagnosed chronic bronchitis or emphysema</li> </ul>	4 (3-5)	8 (1-15)	97 (96-98)	11 (2-21)	95 (94-96)	92 (90-94)
Fessler et al. <sup>9</sup>	<ul style="list-style-type: none"> <li>Affirmative response to physician-diagnosed chronic bronchitis or emphysema, or</li> <li>≥10 pack-year history of smoking and affirmative response to coughing up phlegm on most days for ≥3 consecutive months for ≥2 consecutive years</li> </ul>	4 (3-5)	8 (1-15)	97 (96-98)	11 (2-21)	95 (94-96)	92 (90-94)
<b><i>Former and current smokers – 60-79 (n=1,763)</i></b>							
Gradient Boosting	<ul style="list-style-type: none"> <li>Age ≥55, and any one of</li> <li>Bronchodilator use, or</li> <li>BMI &lt;32 and &gt;36 pack-year history, or</li> <li>BMI &lt;32 and &gt;18 years organic dust exposure</li> </ul>	32 (28-37)	52 (45-59)	80 (76-84)	62 (56-67)	73 (69-77)	69 (66-73)
Mendy et al. <sup>8</sup>	<ul style="list-style-type: none"> <li>Affirmative response to physician-diagnosed chronic bronchitis or emphysema</li> </ul>	13 (10-15)	19 (15-24)	91 (89-94)	57 (49-66)	65 (62-69)	64 (61-68)
Fessler et al. <sup>9</sup>	<ul style="list-style-type: none"> <li>Affirmative response to physician-diagnosed chronic bronchitis or emphysema, or</li> <li>≥10 pack-year history of smoking and affirmative response to coughing up phlegm on most days for ≥3</li> </ul>	18 (15-21)	27 (22-33)	87 (85-90)	56 (49-64)	67 (63-70)	65 (61-68)

Approach	Case Definition	Prevalence (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy <sup>a</sup> (%)
	consecutive months for ≥2 consecutive years						
<b><i>Never smokers – 60-79 (n=1,629)</i></b>							
<p>Gradient Boosting</p> <ul style="list-style-type: none"> <li>• Age ≥55, and any one of</li> <li>• Bronchodilator use, or</li> <li>• BMI &lt;32 and &gt;36 pack-year history, or</li> <li>• BMI &lt;32 and &gt;18 years organic dust exposure</li> </ul>							
Mendy et al. <sup>8</sup>	<ul style="list-style-type: none"> <li>• Affirmative response to physician-diagnosed chronic bronchitis or emphysema</li> </ul>	10 (7-13)	18 (8-29)	92 (89-94)	25 (14-36)	88 (86-90)	82 (79-94)
Fessler et al. <sup>9</sup>	<ul style="list-style-type: none"> <li>• Affirmative response to physician-diagnosed chronic bronchitis or emphysema, or</li> <li>• ≥10 pack-year history of smoking and affirmative response to coughing up phlegm on most days for ≥3 consecutive months for ≥2 consecutive years</li> </ul>	7 (5-10)	9 (3-15)	93 (90-95)	16 (6-27)	87 (85-89)	82 (79-95)
Scale for Strength of Classification Parameters							
							

Abbreviations: COPD, chronic obstructive pulmonary disease; BMI, body mass index; PPV, positive predictive value; NPV, negative predictive value

<sup>a</sup> Accuracy is the proportion of true classifications (true positives + true negatives) among all classifications (true positives + true negatives + false positives + false negatives).

**eTable 5. Prevalence of COPD and classification parameters for questionnaire-based definitions, stratified by race/ethnicity.**

Approach	Case Definition	Prevalence (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy <sup>a</sup> (%)
<b><i>Former and current smokers – non-Hispanic white (n=1,948)</i></b>							
Gradient Boosting	<ul style="list-style-type: none"> <li>• Age ≥55, and any one of</li> <li>• Bronchodilator use, or</li> <li>• BMI &lt;32 and &gt;36 pack-year history, or</li> <li>• BMI &lt;32 and &gt;18 years organic dust exposure</li> </ul>	17 (14-19)	37 (31-44)	91 (89-93)	62 (56-68)	78 (76-81)	76 (73-78)
Mendy et al. <sup>8</sup>	<ul style="list-style-type: none"> <li>• Affirmative response to physician-diagnosed chronic bronchitis or emphysema</li> </ul>	12 (10-14)	19 (15-23)	91 (88-93)	45 (37-52)	74 (71-76)	70 (68-72)
Fessler et al. <sup>9</sup>	<ul style="list-style-type: none"> <li>• Affirmative response to physician-diagnosed chronic bronchitis or emphysema, or</li> <li>• ≥10 pack-year history of smoking and affirmative response to coughing up phlegm on most days for ≥3 consecutive months for ≥2 consecutive years</li> </ul>	18 (16-21)	29 (24-34)	86 (84-89)	46 (40-52)	75 (73-77)	70 (67-72)
<b><i>Never smokers – non-Hispanic white (n=1,643)</i></b>							
Gradient Boosting	<ul style="list-style-type: none"> <li>• Age ≥55, and any one of</li> <li>• Bronchodilator use, or</li> <li>• BMI &lt;32 and &gt;36 pack-year history, or</li> <li>• BMI &lt;32 and &gt;18 years organic dust exposure</li> </ul>	5 (3-7)	10 (3-18)	96 (94-97)	17 (8-26)	92 (91-94)	88 (87-90)

<b>Approach</b>	<b>Case Definition</b>	<b>Prevalence (%)</b>	<b>Sensitivity (%)</b>	<b>Specificity (%)</b>	<b>PPV (%)</b>	<b>NPV (%)</b>	<b>Accuracy<sup>a</sup> (%)</b>
Mendy et al. <sup>8</sup>	<ul style="list-style-type: none"> <li>Affirmative response to physician-diagnosed chronic bronchitis or emphysema</li> </ul>	5 (4-6)	8 (3-14)	95 (94-97)	14 (5-23)	92 (90-93)	88 (86-90)
Fessler et al. <sup>9</sup>	<ul style="list-style-type: none"> <li>Affirmative response to physician-diagnosed chronic bronchitis or emphysema, or</li> <li>≥10 pack-year history of smoking and affirmative response to coughing up phlegm on most days for ≥3 consecutive months for ≥2 consecutive years</li> </ul>	5 (4-6)	8 (3-14)	95 (94-97)	14 (5-23)	92 (90-93)	88 (86-90)
<i><b>Former and current smokers – non-Hispanic black (n=901)</b></i>							
Gradient Boosting	<ul style="list-style-type: none"> <li>Age ≥55, and any one of</li> <li>Bronchodilator use, or</li> <li>BMI &lt;32 and &gt;36 pack-year history, or</li> <li>BMI &lt;32 and &gt;18 years organic dust exposure</li> </ul>	10 (8-13)	22 (15-28)	92 (90-95)	37 (26-48)	85 (82-87)	80 (77-83)
Mendy et al. <sup>8</sup>	<ul style="list-style-type: none"> <li>Affirmative response to physician-diagnosed chronic bronchitis or emphysema</li> </ul>	7 (5-9)	11 (6-16)	94 (92-96)	28 (16-39)	83 (80-85)	79 (76-82)
Fessler et al. <sup>9</sup>	<ul style="list-style-type: none"> <li>Affirmative response to physician-diagnosed chronic bronchitis or emphysema, or</li> <li>≥10 pack-year history of smoking and affirmative response to coughing up phlegm on most days for ≥3</li> </ul>	10 (8-12)	16 (11-22)	91 (89-94)	30 (21-39)	83 (81-85)	78 (75-80)

Approach	Case Definition	Prevalence (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy <sup>a</sup> (%)
	consecutive months for ≥2 consecutive years						
<b><i>Never smokers – non-Hispanic black (n=885)</i></b>							
Gradient Boosting	<ul style="list-style-type: none"> <li>• Age ≥55, and any one of</li> <li>• Bronchodilator use, or</li> <li>• BMI &lt;32 and &gt;36 pack-year history, or</li> <li>• BMI &lt;32 and &gt;18 years organic dust exposure</li> </ul>	3 (2-4)	10 (3-17)	98 (96-99)	23 (11-34)	94 (92-96)	92 (90-94)
Mendy et al. <sup>8</sup>	<ul style="list-style-type: none"> <li>• Affirmative response to physician-diagnosed chronic bronchitis or emphysema</li> </ul>	4 (3-5)	6 (0-12)	96 (95-97)	10 (1-19)	93 (91-95)	90 (88-92)
Fessler et al. <sup>9</sup>	<ul style="list-style-type: none"> <li>• Affirmative response to physician-diagnosed chronic bronchitis or emphysema, or</li> <li>• ≥10 pack-year history of smoking and affirmative response to coughing up phlegm on most days for ≥3 consecutive months for ≥2 consecutive years</li> </ul>	4 (3-5)	6 (0-12)	96 (95-97)	10 (1-19)	93 (91-95)	90 (88-92)
<b><i>Former and current smokers – Hispanic (n=879)</i></b>							
Gradient Boosting	<ul style="list-style-type: none"> <li>• Age ≥55, and any one of</li> <li>• Bronchodilator use, or</li> <li>• BMI &lt;32 and &gt;36 pack-year history, or</li> <li>• BMI &lt;32 and &gt;18 years organic dust exposure</li> </ul>	7 (6-9)	28 (22-35)	95 (93-97)	35 (26-44)	93 (92-95)	89 (86-91)

<b>Approach</b>	<b>Case Definition</b>	<b>Prevalence (%)</b>	<b>Sensitivity (%)</b>	<b>Specificity (%)</b>	<b>PPV (%)</b>	<b>NPV (%)</b>	<b>Accuracy<sup>a</sup> (%)</b>
Mendy et al. <sup>8</sup>	<ul style="list-style-type: none"> <li>Affirmative response to physician-diagnosed chronic bronchitis or emphysema</li> </ul>	5 (4-7)	16 (11-21)	96 (94-97)	27 (17-38)	92 (90-94)	88 (87-90)
Fessler et al. <sup>9</sup>	<ul style="list-style-type: none"> <li>Affirmative response to physician-diagnosed chronic bronchitis or emphysema, or</li> <li>≥10 pack-year history of smoking and affirmative response to coughing up phlegm on most days for ≥3 consecutive months for ≥2 consecutive years</li> </ul>	8 (6-9)	21 (16-27)	94 (92-95)	26 (17-35)	92 (91-94)	87 (86-89)
<b><i>Never smokers – Hispanic (n=1,142)</i></b>							
Gradient Boosting	<ul style="list-style-type: none"> <li>Age ≥55, and any one of</li> <li>Bronchodilator use, or</li> <li>BMI &lt;32 and &gt;36 pack-year history, or</li> <li>BMI &lt;32 and &gt;18 years organic dust exposure</li> </ul>	2 (1-3)	7 (0-14)	98 (98-99)	13 (0-26)	96 (95-97)	95 (94-96)
Mendy et al. <sup>8</sup>	<ul style="list-style-type: none"> <li>Affirmative response to physician-diagnosed chronic bronchitis or emphysema</li> </ul>	3 (2-4)	6 (1-12)	97 (96-99)	8 (0-16)	96 (95-97)	94 (93-95)
Fessler et al. <sup>9</sup>	<ul style="list-style-type: none"> <li>Affirmative response to physician-diagnosed chronic bronchitis or emphysema, or</li> <li>≥10 pack-year history of smoking and affirmative response to coughing up phlegm on most days for ≥3</li> </ul>	3 (2-4)	6 (1-12)	97 (96-99)	8 (0-16)	96 (95-97)	94 (93-95)

Approach	Case Definition	Prevalence (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy <sup>a</sup> (%)
	consecutive months for ≥2 consecutive years						
Scale for Strength of Classification Parameters							
		0	50	100			

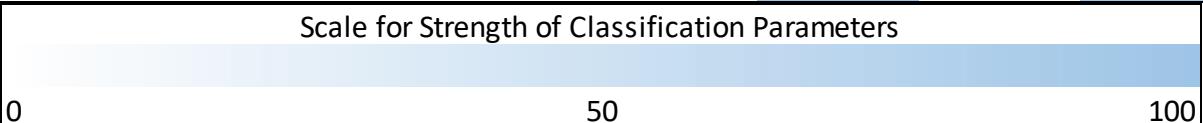
Abbreviations: COPD, chronic obstructive pulmonary disease; BMI, body mass index; PPV, positive predictive value; NPV, negative predictive value

<sup>a</sup> Accuracy is the proportion of true classifications (true positives + true negatives) among all classifications (true positives + true negatives + false positives + false negatives).

**eTable 6. Prevalence of COPD and classification parameters for questionnaire-based definitions, stratified by asthma status.**

Approach	Case Definition	Prevalence (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy <sup>a</sup> (%)
<b><i>Former and current smokers with asthma (n=529)</i></b>							
Gradient Boosting	<ul style="list-style-type: none"> <li>• Age ≥55, and any one of</li> <li>• Bronchodilator use, or</li> <li>• BMI &lt;32 and &gt;36 pack-year history, or</li> <li>• BMI &lt;32 and &gt;18 years organic dust exposure</li> </ul>	27 (22-32)	44 (32-55)	84 (79-88)	62 (54-70)	71 (63-79)	68 (62-75)
Mendy et al. <sup>8</sup>	<ul style="list-style-type: none"> <li>• Affirmative response to physician-diagnosed chronic bronchitis or emphysema</li> </ul>	28 (24-33)	36 (28-43)	76 (71-82)	49 (38-59)	66 (60-71)	61 (56-66)
Fessler et al. <sup>9</sup>	<ul style="list-style-type: none"> <li>• Affirmative response to physician-diagnosed chronic bronchitis or emphysema, or</li> <li>• ≥10 pack-year history of smoking and affirmative response to coughing up phlegm on most days for ≥3 consecutive months for ≥2 consecutive years</li> </ul>	34 (28-39)	44 (35-54)	73 (66-79)	50 (41-60)	68 (61-74)	62 (57-67)
<b><i>Never smokers with asthma (n=446)</i></b>							
Gradient Boosting	<ul style="list-style-type: none"> <li>• Age ≥55, and any one of</li> <li>• Bronchodilator use, or</li> <li>• BMI &lt;32 and &gt;36 pack-year history, or</li> <li>• BMI &lt;32 and &gt;18 years organic dust exposure</li> </ul>	17 (10-23)	35 (15-55)	86 (82-91)	31 (18-44)	88 (84-93)	79 (74-84)

<b>Approach</b>	<b>Case Definition</b>	<b>Prevalence (%)</b>	<b>Sensitivity (%)</b>	<b>Specificity (%)</b>	<b>PPV (%)</b>	<b>NPV (%)</b>	<b>Accuracy <sup>a</sup> (%)</b>
Mendy et al. <sup>8</sup>	<ul style="list-style-type: none"> <li>Affirmative response to physician-diagnosed chronic bronchitis or emphysema</li> </ul>	23 (19-28)	29 (18-40)	78 (73-83)	19 (10-27)	86 (82-91)	71 (66-76)
Fessler et al. <sup>9</sup>	<ul style="list-style-type: none"> <li>Affirmative response to physician-diagnosed chronic bronchitis or emphysema, or</li> <li>≥10 pack-year history of smoking and affirmative response to coughing up phlegm on most days for ≥3 consecutive months for ≥2 consecutive years</li> </ul>	23 (19-28)	29 (18-40)	78 (73-83)	19 (10-27)	86 (82-91)	71 (66-76)
<i>Former and current smokers without asthma (n=3,406)</i>							
Gradient Boosting	<ul style="list-style-type: none"> <li>Age ≥55, and any one of</li> <li>Bronchodilator use, or</li> <li>BMI &lt;32 and &gt;36 pack-year history, or</li> <li>BMI &lt;32 and &gt;18 years organic dust exposure</li> </ul>	13 (11-15)	33 (27-39)	93 (91-94)	58 (51-65)	82 (80-84)	79 (77-81)
Mendy et al. <sup>8</sup>	<ul style="list-style-type: none"> <li>Affirmative response to physician-diagnosed chronic bronchitis or emphysema</li> </ul>	8 (6-10)	13 (10-17)	94 (92-95)	39 (30-48)	78 (76-80)	75 (73-77)
Fessler et al. <sup>9</sup>	<ul style="list-style-type: none"> <li>Affirmative response to physician-diagnosed chronic bronchitis or emphysema, or</li> <li>≥10 pack-year history of smoking and affirmative response to coughing up phlegm on most days for ≥3</li> </ul>	13 (11-15)	23 (19-27)	90 (88-92)	41 (35-48)	79 (77-81)	74 (72-76)

Approach	Case Definition	Prevalence (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy <sup>a</sup> (%)
	consecutive months for ≥2 consecutive years						
<b><i>Never smokers without asthma (n=3,608)</i></b>							
Gradient Boosting	<ul style="list-style-type: none"> <li>• Age ≥55, and any one of</li> <li>• Bronchodilator use, or</li> <li>• BMI &lt;32 and &gt;36 pack-year history, or</li> <li>• BMI &lt;32 and &gt;18 years organic dust exposure</li> </ul>	3 (2-3)	3 (1-5)	97 (97-98)	7 (2-13)	94 (92-95)	91 (90-93)
Mendy et al. <sup>8</sup>	<ul style="list-style-type: none"> <li>• Affirmative response to physician-diagnosed chronic bronchitis or emphysema</li> </ul>	2 (1-3)	1 (0-3)	98 (97-98)	4 (0-8)	93 (92-95)	91 (90-93)
Fessler et al. <sup>9</sup>	<ul style="list-style-type: none"> <li>• Affirmative response to physician-diagnosed chronic bronchitis or emphysema, or</li> <li>• ≥10 pack-year history of smoking and affirmative response to coughing up phlegm on most days for ≥3 consecutive months for ≥2 consecutive years</li> </ul>	2 (1-3)	1 (0-3)	98 (97-98)	4 (0-8)	93 (92-95)	91 (90-93)
Scale for Strength of Classification Parameters							
							
0                          50                          100							

Abbreviations: COPD, chronic obstructive pulmonary disease; BMI, body mass index; PPV, positive predictive value; NPV, negative predictive value

<sup>a</sup> Accuracy is the proportion of true classifications (true positives + true negatives) among all classifications (true positives + true negatives + false positives + false negatives).

**eTable 7. Prevalence of COPD and classification parameters for questionnaire-based definitions, Lower Limit of Normal and fixed ratio separately.**

Approach	Case Definition	Prevalence (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy <sup>a</sup> (%)
<b><i>Former and current smokers – fixed ratio (n=3,940)</i></b>							
Gradient Boosting	<ul style="list-style-type: none"> <li>• Age ≥55, and any one of</li> <li>• Bronchodilator use, or</li> <li>• BMI &lt;32 and &gt;36 pack-year history, or</li> <li>• BMI &lt;32 and &gt;18 years organic dust exposure</li> </ul>	15 (13-17)	36 (30-42)	92 (90-93)	59 (53-64)	81 (79-84)	78 (76-80)
Mendy et al. <sup>8</sup>	<ul style="list-style-type: none"> <li>• Affirmative response to physician-diagnosed chronic bronchitis or emphysema</li> </ul>	11 (9-12)	18 (14-22)	92 (90-93)	42 (35-49)	77 (75-79)	73 (72-75)
Fessler et al. <sup>9</sup>	<ul style="list-style-type: none"> <li>• Affirmative response to physician-diagnosed chronic bronchitis or emphysema, or</li> <li>• ≥10 pack-year history of smoking and affirmative response to coughing up phlegm on most days for ≥3 consecutive months for ≥2 consecutive years</li> </ul>	16 (14-18)	28 (23-32)	88 (86-90)	44 (38-49)	78 (77-80)	73 (71-75)
<b><i>Never smokers – fixed ratio (n=4,056)</i></b>							
Gradient Boosting	<ul style="list-style-type: none"> <li>• Age ≥55, and any one of</li> <li>• Bronchodilator use, or</li> <li>• BMI &lt;32 and &gt;36 pack-year history, or</li> <li>• BMI &lt;32 and &gt;18 years organic dust exposure</li> </ul>	4 (3-5)	10 (4-16)	96 (95-97)	18 (10-26)	93 (92-94)	90 (89-91)

<b>Approach</b>	<b>Case Definition</b>	<b>Prevalence (%)</b>	<b>Sensitivity (%)</b>	<b>Specificity (%)</b>	<b>PPV (%)</b>	<b>NPV (%)</b>	<b>Accuracy<sup>a</sup> (%)</b>
Mendy et al. <sup>8</sup>	<ul style="list-style-type: none"> <li>Affirmative response to physician-diagnosed chronic bronchitis or emphysema</li> </ul>	5 (4-6)	9 (4-13)	96 (95-97)	14 (7-21)	93 (92-94)	89 (88-91)
Fessler et al. <sup>9</sup>	<ul style="list-style-type: none"> <li>Affirmative response to physician-diagnosed chronic bronchitis or emphysema, or</li> <li>≥10 pack-year history of smoking and affirmative response to coughing up phlegm on most days for ≥3 consecutive months for ≥2 consecutive years</li> </ul>	5 (4-6)	9 (4-13)	96 (95-97)	14 (7-21)	93 (92-94)	89 (88-91)
<b><i>Former and current smokers – LLN (n=3,361)</i></b>							
Gradient Boosting	<ul style="list-style-type: none"> <li>Age ≥55, and any one of</li> <li>Bronchodilator use, or</li> <li>BMI &lt;32 and &gt;36 pack-year history, or</li> <li>BMI &lt;32 and &gt;18 years organic dust exposure</li> </ul>	16 (13-18)	36 (29-43)	89 (88-91)	45 (39-50)	85 (83-87)	79 (77-81)
Mendy et al. <sup>8</sup>	<ul style="list-style-type: none"> <li>Affirmative response to physician-diagnosed chronic bronchitis or emphysema</li> </ul>	11 (9-13)	20 (16-24)	91 (89-93)	35 (29-41)	83 (81-85)	78 (75-80)
Fessler et al. <sup>9</sup>	<ul style="list-style-type: none"> <li>Affirmative response to physician-diagnosed chronic bronchitis or emphysema, or</li> <li>≥10 pack-year history of smoking and affirmative response to coughing up phlegm on most days for ≥3</li> </ul>	17 (14-19)	33 (29-37)	87 (85-89)	38 (33-42)	85 (83-86)	77 (75-79)

Approach	Case Definition	Prevalence (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy <sup>a</sup> (%)
	consecutive months for ≥2 consecutive years						
<b><i>Never smokers – LLN (n=3,174)</i></b>							
Gradient Boosting	<ul style="list-style-type: none"> <li>• Age ≥55, and any one of</li> <li>• Bronchodilator use, or</li> <li>• BMI &lt;32 and &gt;36 pack-year history, or</li> <li>• BMI &lt;32 and &gt;18 years organic dust exposure</li> </ul>	4 (3-6)	12 (2-22)	96 (95-97)	11 (3-19)	96 (95-97)	92 (91-94)
Mendy et al. <sup>8</sup>	<ul style="list-style-type: none"> <li>• Affirmative response to physician-diagnosed chronic bronchitis or emphysema</li> </ul>	5 (4-6)	8 (2-14)	95 (94-97)	7 (2-13)	96 (95-97)	92 (91-93)
Fessler et al. <sup>9</sup>	<ul style="list-style-type: none"> <li>• Affirmative response to physician-diagnosed chronic bronchitis or emphysema, or</li> <li>• ≥10 pack-year history of smoking and affirmative response to coughing up phlegm on most days for ≥3 consecutive months for ≥2 consecutive years</li> </ul>	5 (4-6)	8 (2-14)	95 (94-97)	7 (2-13)	96 (95-97)	92 (91-93)
Scale for Strength of Classification Parameters							
0                          50                          100							

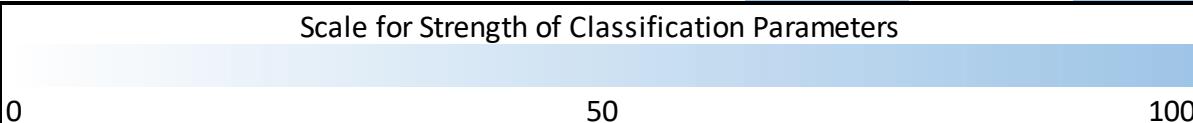
Abbreviations: COPD, chronic obstructive pulmonary disease; BMI, body mass index; PPV, positive predictive value; NPV, negative predictive value

<sup>a</sup> Accuracy is the proportion of true classifications (true positives + true negatives) among all classifications (true positives + true negatives + false positives + false negatives).

**eTable 8. Prevalence of COPD and classification parameters for questionnaire-based definitions, restricted by severity of COPD.**

Approach	Case Definition	Prevalence (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy <sup>a</sup> (%)
<i>Former and current smokers with severe<sup>b</sup> COPD (n=3,430)</i>							
Gradient Boosting	<ul style="list-style-type: none"> <li>• Age ≥55, and any one of</li> <li>• Bronchodilator use, or</li> <li>• BMI &lt;32 and &gt;36 pack-year history, or</li> <li>• BMI &lt;32 and &gt;18 years organic dust exposure</li> </ul>	14 (12-15)	45 (37-53)	92 (90-93)	48 (41-54)	91 (89-93)	85 (83-87)
Mendy et al. <sup>8</sup>	<ul style="list-style-type: none"> <li>• Affirmative response to physician-diagnosed chronic bronchitis or emphysema</li> </ul>	11 (9-13)	28 (23-33)	92 (90-93)	36 (29-43)	89 (87-90)	83 (81-85)
Fessler et al. <sup>9</sup>	<ul style="list-style-type: none"> <li>• Affirmative response to physician-diagnosed chronic bronchitis or emphysema, or</li> <li>• ≥10 pack-year history of smoking and affirmative response to coughing up phlegm on most days for ≥3 consecutive months for ≥2 consecutive years</li> </ul>	16 (14-18)	40 (35-45)	88 (86-90)	36 (30-41)	90 (88-91)	81 (79-83)
<i>Never smokers with severe<sup>b</sup> COPD (n=3,835)</i>							
Gradient Boosting	<ul style="list-style-type: none"> <li>• Age ≥55, and any one of</li> <li>• Bronchodilator use, or</li> <li>• BMI &lt;32 and &gt;36 pack-year history, or</li> <li>• BMI &lt;32 and &gt;18 years organic dust exposure</li> </ul>	4 (3-5)	22 (5-39)	96 (95-97)	12 (3-21)	98 (98-99)	95 (94-96)

<b>Approach</b>	<b>Case Definition</b>	<b>Prevalence (%)</b>	<b>Sensitivity (%)</b>	<b>Specificity (%)</b>	<b>PPV (%)</b>	<b>NPV (%)</b>	<b>Accuracy<sup>a</sup> (%)</b>
Mendy et al. <sup>8</sup>	<ul style="list-style-type: none"> <li>Affirmative response to physician-diagnosed chronic bronchitis or emphysema</li> </ul>	5 (4-6)	13 (4-22)	96 (95-97)	6 (1-12)	98 (97-99)	94 (93-95)
Fessler et al. <sup>9</sup>	<ul style="list-style-type: none"> <li>Affirmative response to physician-diagnosed chronic bronchitis or emphysema, or</li> <li>≥10 pack-year history of smoking and affirmative response to coughing up phlegm on most days for ≥3 consecutive months for ≥2 consecutive years</li> </ul>	5 (4-6)	13 (4-22)	96 (95-97)	6 (1-12)	98 (97-99)	94 (93-95)
<b><i>Former and current smokers with mild<sup>c</sup> COPD (n=3,397)</i></b>							
Gradient Boosting	<ul style="list-style-type: none"> <li>Age ≥55, and any one of</li> <li>Bronchodilator use, or</li> <li>BMI &lt;32 and &gt;36 pack-year history, or</li> <li>BMI &lt;32 and &gt;18 years organic dust exposure</li> </ul>	11 (9-12)	25 (20-31)	92 (90-93)	32 (25-38)	89 (87-91)	83 (81-85)
Mendy et al. <sup>8</sup>	<ul style="list-style-type: none"> <li>Affirmative response to physician-diagnosed chronic bronchitis or emphysema</li> </ul>	8 (7-10)	8 (4-11)	92 (90-93)	13 (8-17)	86 (85-88)	80 (78-82)
Fessler et al. <sup>9</sup>	<ul style="list-style-type: none"> <li>Affirmative response to physician-diagnosed chronic bronchitis or emphysema, or</li> <li>≥10 pack-year history of smoking and affirmative response to coughing up phlegm on most days for ≥3</li> </ul>	12 (10-14)	14 (9-20)	88 (86-90)	16 (11-20)	87 (85-89)	78 (76-80)

Approach	Case Definition	Prevalence (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy <sup>a</sup> (%)
	consecutive months for ≥2 consecutive years						
<i>Never smokers with mild<sup>c</sup> COPD (n=3,907)</i>							
Gradient Boosting	<ul style="list-style-type: none"> <li>• Age ≥55, and any one of</li> <li>• Bronchodilator use, or</li> <li>• BMI &lt;32 and &gt;36 pack-year history, or</li> <li>• BMI &lt;32 and &gt;18 years organic dust exposure</li> </ul>	4 (3-5)	5 (1-8)	96 (95-97)	6 (2-10)	95 (94-96)	92 (90-93)
Mendy et al. <sup>8</sup>	<ul style="list-style-type: none"> <li>• Affirmative response to physician-diagnosed chronic bronchitis or emphysema</li> </ul>	5 (4-5)	6 (0-12)	96 (95-97)	6 (0-12)	95 (94-96)	91 (90-93)
Fessler et al. <sup>9</sup>	<ul style="list-style-type: none"> <li>• Affirmative response to physician-diagnosed chronic bronchitis or emphysema, or</li> <li>• ≥10 pack-year history of smoking and affirmative response to coughing up phlegm on most days for ≥3 consecutive months for ≥2 consecutive years</li> </ul>	5 (4-5)	6 (0-12)	96 (95-97)	6 (0-12)	95 (94-96)	91 (90-93)
Scale for Strength of Classification Parameters							
							

Abbreviations: COPD, chronic obstructive pulmonary disease; BMI, body mass index; PPV, positive predictive value; NPV, negative predictive value

<sup>a</sup> Accuracy is the proportion of true classifications (true positives + true negatives) among all classifications (true positives + true negatives + false positives + false negatives).

<sup>b</sup> Severe COPD is defined as FEV<sub>1</sub>/FCV ratio below 0.6.

<sup>c</sup> Mild COPD is defined as FEV<sub>1</sub>/FCV ratio below 0.7 or LLN, while greater than 0.6.

**eTable 9. Prevalence of COPD and classification parameters for questionnaire-based definitions, correcting for outcome misclassification using a re-weighting approach.**

Approach	Case Definition	Prevalence (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy <sup>a</sup> (%)
<i>Current and former smokers (n=3,940)</i>							
Gradient Boosting	<ul style="list-style-type: none"> <li>• Age ≥55, and any one of</li> <li>• Bronchodilator use, or</li> <li>• BMI &lt;32 and &gt;36 pack-year history, or</li> <li>• BMI &lt;32 and &gt;18 years organic dust exposure</li> </ul>	17 (15-20)	36 (30-42)	91 (89-93)	63 (58-69)	76 (73-79)	74 (72-76)
Mendy et al. <sup>8</sup>	<ul style="list-style-type: none"> <li>• Affirmative response to physician-diagnosed chronic bronchitis or emphysema</li> </ul>	13 (10-15)	20 (16-24)	91 (89-93)	50 (44-56)	72 (69-74)	69 (67-71)
<i>Never smokers (n=4,056)</i>							
Gradient Boosting	<ul style="list-style-type: none"> <li>• Age ≥55, and any one of</li> <li>• Bronchodilator use, or</li> <li>• BMI &lt;32 and &gt;36 pack-year history, or</li> <li>• BMI &lt;32 and &gt;18 years organic dust exposure</li> </ul>	5 (3-6)	12 (5-18)	96 (95-97)	25 (15-34)	91 (89-92)	88 (86-89)

Approach	Case Definition	Prevalence (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy <sup>a</sup> (%)
Mendy et al. <sup>8</sup>	<ul style="list-style-type: none"> <li>Affirmative response to physician-diagnosed chronic bronchitis or emphysema</li> </ul>	6 (5-7)	12 (6-18)	95 (93-96)	20 (10-31)	91 (89-92)	86 (85-88)
Fessler et al. <sup>9</sup>	<ul style="list-style-type: none"> <li>Affirmative response to physician-diagnosed chronic bronchitis or emphysema, or</li> <li>≥10 pack-year history of smoking and affirmative response to coughing up phlegm on most days for ≥3 consecutive months for ≥2 consecutive years</li> </ul>	6 (5-7)	12 (6-18)	95 (93-96)	20 (10-31)	91 (89-92)	86 (85-88)

Scale for Strength of Classification Parameters

0	50	100
---	----	-----

Abbreviations: COPD, chronic obstructive pulmonary disease; BMI, body mass index; PPV, positive predictive value; NPV, negative predictive value

<sup>a</sup> Accuracy is the proportion of true classifications (true positives + true negatives) among all classifications (true positives + true negatives + false positives + false negatives).

**eTable 10. Prevalence of COPD and classification parameters for questionnaire-based definitions, using pre-bronchodilator values only.**

Approach	Case Definition	Prevalence (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy <sup>a</sup> (%)
<i>Current and former smokers (n=3,940)</i>							
Gradient Boosting	<ul style="list-style-type: none"> <li>• Age ≥55, and any one of</li> <li>• Bronchodilator use, or</li> <li>• BMI &lt;32 and &gt;36 pack-year history, or</li> <li>• BMI &lt;32 and &gt;18 years organic dust exposure</li> </ul>	15 (13-17)	32 (26-38)	92 (91-94)	64 (59-70)	76 (73-78)	74 (71-77)
Mendy et al. <sup>8</sup>	<ul style="list-style-type: none"> <li>• Affirmative response to physician-diagnosed chronic bronchitis or emphysema</li> </ul>	11 (9-12)	16 (13-20)	92 (90-93)	46 (39-53)	71 (69-73)	69 (67-71)
Fessler et al. <sup>9</sup>	<ul style="list-style-type: none"> <li>• Affirmative response to physician-diagnosed chronic bronchitis or emphysema, or</li> <li>• ≥10 pack-year history of smoking and affirmative response to coughing up phlegm on most days for ≥3 consecutive months for ≥2 consecutive years</li> </ul>	16 (14-18)	26 (21-30)	88 (86-90)	49 (44-55)	73 (70-75)	69 (67-71)
<i>Never smokers (n=4,056)</i>							
Gradient Boosting	<ul style="list-style-type: none"> <li>• Age ≥55, and any one of</li> <li>• Bronchodilator use, or</li> <li>• BMI &lt;32 and &gt;36 pack-year history, or</li> <li>• BMI &lt;32 and &gt;18 years organic dust exposure</li> </ul>	4 (3-5)	8 (4-13)	96 (95-97)	22 (15-30)	89 (88-91)	86 (85-88)

Approach	Case Definition	Prevalence (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy <sup>a</sup> (%)
Mendy et al. <sup>8</sup>	<ul style="list-style-type: none"> <li>Affirmative response to physician-diagnosed chronic bronchitis or emphysema</li> </ul>	5 (4-6)	8 (4-11)	96 (95-97)	19 (10-27)	89 (87-91)	86 (84-87)
Fessler et al. <sup>9</sup>	<ul style="list-style-type: none"> <li>Affirmative response to physician-diagnosed chronic bronchitis or emphysema, or</li> <li>≥10 pack-year history of smoking and affirmative response to coughing up phlegm on most days for ≥3 consecutive months for ≥2 consecutive years</li> </ul>	5 (4-6)	8 (4-11)	96 (95-97)	19 (10-27)	89 (87-91)	86 (84-87)

Scale for Strength of Classification Parameters

0	50	100
---	----	-----

Abbreviations: COPD, chronic obstructive pulmonary disease; BMI, body mass index; PPV, positive predictive value; NPV, negative predictive value

<sup>a</sup> Accuracy is the proportion of true classifications (true positives + true negatives) among all classifications (true positives + true negatives + false positives + false negatives).

**eTable 11. Prevalence of COPD and classification parameters for questionnaire-based definitions, excluding post-bronchodilator eligible participants with incomplete post-bronchodilator results.**

Approach	Case Definition	Prevalence (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy <sup>a</sup> (%)
<i>Current and former smokers (n=3,336)</i>							
Gradient Boosting	<ul style="list-style-type: none"> <li>• Age ≥55, and any one of</li> <li>• Bronchodilator use, or</li> <li>• BMI &lt;32 and &gt;36 pack-year history, or</li> <li>• BMI &lt;32 and &gt;18 years organic dust exposure</li> </ul>	10 (9-12)	27 (20-34)	92 (90-93)	31 (24-38)	90 (88-92)	84 (82-86)
Mendy et al. <sup>8</sup>	<ul style="list-style-type: none"> <li>• Affirmative response to physician-diagnosed chronic bronchitis or emphysema</li> </ul>	9 (7-10)	11 (7-14)	92 (90-93)	15 (10-20)	88 (87-90)	82 (80-84)
<i>Never smokers (n=3,853)</i>							
Gradient Boosting	<ul style="list-style-type: none"> <li>• Age ≥55, and any one of</li> <li>• Bronchodilator use, or</li> <li>• BMI &lt;32 and &gt;36 pack-year history, or</li> <li>• BMI &lt;32 and &gt;18 years organic dust exposure</li> </ul>	4 (3-5)	8 (1-15)	96 (95-97)	7 (1-12)	97 (96-98)	94 (92-95)

Approach	Case Definition	Prevalence (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy <sup>a</sup> (%)
Mendy et al. <sup>8</sup>	<ul style="list-style-type: none"> <li>Affirmative response to physician-diagnosed chronic bronchitis or emphysema</li> </ul>	4 (3-5)	6 (1-11)	96 (95-97)	4 (1-8)	97 (96-98)	93 (92-94)
Fessler et al. <sup>9</sup>	<ul style="list-style-type: none"> <li>Affirmative response to physician-diagnosed chronic bronchitis or emphysema, or</li> <li>≥10 pack-year history of smoking and affirmative response to coughing up phlegm on most days for ≥3 consecutive months for ≥2 consecutive years</li> </ul>	4 (3-5)	6 (1-11)	96 (95-97)	4 (1-8)	97 (96-98)	93 (92-94)

Scale for Strength of Classification Parameters

0                          50                          100

Abbreviations: COPD, chronic obstructive pulmonary disease; BMI, body mass index; PPV, positive predictive value; NPV, negative predictive value

<sup>a</sup> Accuracy is the proportion of true classifications (true positives + true negatives) among all classifications (true positives + true negatives + false positives + false negatives).

## eAppendix 2. SAS code for creating the analytic dataset and conducting the analysis.

```

*****  

* File Name: NHANES_COPD_Analytic_Code_<Date>_<Version>.sas  

*  

* Programmed by: Jesse Wilkerson  

* Program QC'd by: Matthew F. Bridge  

*  

* Description: Read in COPD data from eligible NHANES 2007-2012 participants,  

*               categorize/derive data, and run statistical analyses.  

*  

* External References:  

*  

* Type      Name                      Description of Reference  

* -----  -----  

* INPUT      NHANES_2007_12.sas7bdat    NHANES 2007-2012 data extract  

*             COPD_medications.sas7bdat   NHANES medication dataset  

*****/  

libname nhanes "C:\";  

data nhanes;  

set nhanes.NHANES_2007_12;  

/*Recode missing indicators as missing*/  

if SMD030 in(0,777,999) then SMD030=.; if SMD055 in(777,999) then SMD055=.;  

if SMD057 =999 then SMD057=.; if SMD650 in(777,999) then SMD650=.;  

if DMDEDUC2 in(7,9) then DMDEDUC2=.; if SMQ020 in(7,9) then SMQ020=.;  

if SMD030 in(0,777,999) then SMD030=.; if SMQ040 =7 then SMQ040=.;  

if SMQ050Q in(77777,99999) then SMQ050Q=.; if SMD055 in(777,999) then SMD055=.;  

if SMD057 =999 then SMD057=.; if SMD641 =99 then SMD641=.;  

if SMD650 in(777,999) then SMD650=.; if SMD410 in(7,9) then SMD410=.;  

if OCQ290G in(7,9) then OCQ290G=.; if OCD150 =7 then OCD150=.;  

if OCQ275 in(7,9) then OCQ275=.; if OCQ510 in(7,9) then OCQ510=.;  

if OCQ520 =9999 then OCQ520=.; if OCQ530 in(7,9) then OCQ530=.;  

if OCQ540 in(7777,9999) then OCQ540=.; if OCQ550 in(7,9) then OCQ550=.;  

if OCQ560 =9999 then OCQ560=.; if OCQ570 in(7,9) then OCQ570=.;  

if OCQ580 =9999 then OCQ580=.; if HSD010 =9 then HSD010=.;  

if MCQ010 in(7,9) then MCQ010=.; if MCQ025 =9999 then MCQ025=.;  

if MCQ035 =9 then MCQ035=.; if MCQ040 =9 then MCQ040=.;  

if MCQ051 =9 then MCQ051=.; if MCQ160G in(7,9) then MCQ160G=.;  

if MCQ180G =99999 then MCQ180G=.; if MCQ160K in(7,9) then MCQ160K=.;  

if MCQ170K =9 then MCQ170K=.; if MCQ180K =99999 then MCQ180K=.;  

if MCQ300B in(7,9) then MCQ300B=.; if CDQ010 in(7,9) then CDQ010=.;  

if PFQ061B in(7,9) then PFQ061B=.; if PFQ061C in(7,9) then PFQ061C=.;  

if RDQ031 in(7,9) then RDQ031=.; if RDD040 =99999 then RDD040=.;  

if RDQ050 in(7,9) then RDQ050=.; if RDD060 in(77777,99999) then RDD060=.;  

if RDQ070 in(7,9) then RDQ070=.; if RDQ080 =999 then RDQ080=.;  

if RDQ090 =9 then RDQ090=.; if RDQ100 =9 then RDQ100=.;  

if RDD120 =9999 then RDD120=.; if RDQ134 =9 then RDQ134=.;  

if RDQ135 =9 then RDQ135=.; if RDQ140 in(7,9) then RDQ140=.;  

if AGQ030 in(7,9) then AGQ030=.;  

/*Create oxygen indicator variable*/  

oxygen=.;  

if ENQ010=1 then oxygen=1;  

if ENQ010=2 then oxygen=0;

```

```

label oxygen = "Participant excluded from spirometry due to oxygen medication";

/*Create an indicator variable for being eligible for inclusion*/
If SPXNSTAT=1 /*Completed baseline spirometry*/
and SPXNQFVC in('A','B','C') /*Baseline FVC Quality exceeds or meets ATS data collection
standards*/
and SPXNQFV1 in('A','B','C') /*Baseline FEV1 Quality exceeds or meets ATS data collection
standards*/
and RIDAGEYR ge 40 /*Limit to individuals who are at least 40 years of age*/
then eligible=1;
else eligible=0;

/*Derive 6 year weights*/
WTMEC6YR=WTMEC2YR/3;
label WTMEC6YR = "Full Sample 6 Year MEC Exam Weight";

/*Derive Lower Limit of Normal values for spirometry data*/
*Males;
if riagendr=1 then do;
    if ridreth1=1 then do;i=80.925;a=-0.2186;end;*Mexican Americans;
    if ridreth1=3 then do;i=78.388;a=-0.2066;end;*Non-Hispanic White;
    if ridreth1=4 then do;i=78.822;a=-0.1828;end;*Non-Hispanic Black;
end;
*Females;
if riagendr=2 then do;
    if ridreth1=1 then do;i=83.044;a=-0.2248;end;*Mexican Americans;
    if ridreth1=3 then do;i=81.015;a=-0.2125;end;*Non-Hispanic White;
    if ridreth1=4 then do;i=80.978;a=-0.2039;end;*Non-Hispanic Black;
end;
lln_ref = (i + (ridgeyr*a))/100;
label lln_ref = "Lower Limit of Normal Reference Value";

/*Derive LLN1 eligible*/
LLN1_eligible=0;
if ridreth1 in (1,3,4) then LLN1_eligible=1;
label LLN1_eligible = "Eligible for LLN Analysis at Baseline";

/*Derive LLN1 indicator.*/
if -1 < SPXNFEV1/SPXNFVC < lln_ref then LLN1=1;
if SPXNFEV1/SPXNFVC ge lln_ref then LLN1=0;
if LLN1_eligible=0 then LLN1=.;
label LLN1 = "Baseline Spirometry Below Lower Limit of Normal";

/*Derive Fixed Ratio 1 indicator.*/
if -1 < SPXNFEV1/SPXNFVC < .7 then FixedRatio1=1;
if SPXNFEV1/SPXNFVC ge .7 then FixedRatio1=0;
label FixedRatio1 = "Baseline Spirometry Below Fixed Ratio 70%";

/*Derive COPD1 indicator*/
if LLN1=1 or FixedRatio1=1 then COPD1=1;
else COPD1=0;
label COPD1 = "COPD at Baseline";

if SPXBSTAT=1 then do;
    /*Derive LLN2 eligible*/
    LLN2_eligible=0;
    if ridreth1 in (1,3,4) then LLN2_eligible=1;

```

```

/*Derive LLN2 indicator.*/
if -1 < SPXBFEV1/SPXBFVC < lln_ref then LLN2=1;
if SPXBFEV1/SPXBFVC ge lln_ref then LLN2=0;
if LLN2_eligible=0 then LLN2=.;

/*Derive Fixed Ratio 2 indicator.*/
if -1 < SPXBFEV1/SPXBFVC < .7 then FixedRatio2=1;
if SPXBFEV1/SPXBFVC ge .7 then FixedRatio2=0;

/*Derive COPD2 indicator*/
if LLN2=1 or FixedRatio2=1 then COPD2=1;
else COPD2=0;
end;

if SPXBSTAT ne 1 or SPXBQFV1='D' or SPXBQFVC='D' then do;
    LLN2_eligible=.;
    LLN2=.;
    FixedRatio2=.;
    COPD2=.;
end;
label LLN2_eligible = "Eligible for LLN Analysis at 2nd Test";
label LLN2 = "2nd Test Spirometry Below Lower Limit of Normal";
label FixedRatio2 = "2nd Test Spirometry Below Fixed Ratio 70%";
label COPD2 = "COPD at 2nd Test";

/*Derive mild COPD*/
mild_COPD=.;
if -1 < SPXNFEV1/SPXNFVC < .6 then mild_COPD=0;
if .6 le SPXNFEV1/SPXNFVC < .7 then mild_COPD=1;

/*Derive additional respiratory variables*/
chronic_cough=0;
if rdq031=1 and rdd040 ge 2 then chronic_cough=1;
label chronic_cough = "Chronic Cough";

chronic_phlegm=0;
if rdq050=1 and rdd060 ge 2 then chronic_phlegm=1;
label chronic_phlegm = "Chronic Phlegm";

/*Derive additional covariate variables*/
race_eth=.;
if ridreth1 in(1,2) then race_eth=1;
if ridreth1 =3 then race_eth=3;
if ridreth1 =4 then race_eth=2;
if ridreth1 =5 then race_eth=4;
label race_eth = "Race/Ethnicity - Recode";

smoke=.;
if SMQ020=2 then smoke=0;
if SMQ020=1 and SMQ040=3 then smoke=1;
if SMQ020=1 and SMQ040 in(1,2) then smoke=2;
label smoke = "Smoking History";

pack_years=.;
if smoke=2 then do;
    if RIDAGEYR ne SMD030 then pack_years=(SMD650/20)*(RIDAGEYR-SMD030);

```

```

        if RIDAGEYR = SMD030 then pack_years=(SMD650/20);
end;
if smoke=1 then do;
    if SMD055 ne SMD030 then pack_years=(SMD057/20)*(SMD055-SMD030);
    if SMD055 = SMD030 then pack_years=(SMD057/20);
end;
label pack_years = "Smoking History - Pack Years";

secondhand=0;
if SMD410=1 or OCQ290G=1 or OCQ275=1 then secondhand=1;
label secondhand = "Secondhand smoke exposure";

/*Self-rated health*/
hsd010_jw=.;
if hsd010 in(1,2) then hsd010_jw=1; /*Excellent or very good*/
if hsd010 =3 then hsd010_jw=2; /*Good */
if hsd010 in(4,5) then hsd010_jw=3; /*Fair or poor */

/*Wheezing disturbed sleep in past year*/
rdq090_jw=.;
if rdq090 =0 then rdq090_jw=0;
if rdq090 in(1,2) then rdq090_jw=1;

/*Amount of wheezing limits physical activity*/
rdq135_jw=.;
if rdq135 =1 then rdq135_jw=0;
if rdq135 in(2,3,4,5) then rdq135_jw=1;

/*Factor in skip logic*/
if smoke=0 then pack_years=0; /*No smoking means no pack-years*/
if mcq160k=2 then mcq170k=2; /*No ever chronic bronchitis means no current bronchitis*/
if mcq010=2 then do; /*No ever asthma means no*/
    mcq035=2; /*current asthma or */
    mcq051=2; /*asthma medication */
end;
if rdq070=2 then do; /*No wheeze in past year means no */
    rdq080=0; /*wheezing attacks or */
    rdd120=0; /*medical visits for wheezing or */
    rdq090_jw=0; /*wheezing disturbed sleep or */
    rdq100=2; /*wheeze during exercise or */
    rdq134=2; /*wheeze medication or */
    rdq135_jw=0; /*wheeze limiting physical activity*/
if ocq510=2 then ocq520=0; /*No exposure to mineral dusts means no years of mineral dust exposure*/
if ocq530=2 then ocq540=0; /*No exposure to organic dusts means no years of organic dust exposure*/
if ocq550=2 then ocq560=0; /*No exposure to exhaust fumes means no years of exhaust fume exposure*/
if ocq570=2 then ocq580=0; /*No exposure to other fumes means no years of other fume exposure*/
end;

/*Educational attainment*/
if dmdeduc2=1 then dmdeduc2=2; /*Recode less than 9th grade into less than high school graduate*/

/*Dichotomize age for subgroup analyses*/
age_2lev=.;
if ridgeyr < 60 then age_2lev=0;

```

```

if ridgeyr ge 60 then age_2lev=1;
if ridgeyr=. then age_2lev=.;

/*Hispanic ethnicity*/
race_2lev=.;
if race_eth=1 then race_2lev=0;
if race_eth in(2,3,4) then race_2lev=1;
if race_eth=. then race_2lev=.;

/*Capture pre-bronchodilator values for sensitivity analysis*/
COPD1b=COPD1;

/*Capture Fixed Ratio and LLN values seperately for sensitivity analysis*/
if FixedRatio2=0 then FixedRatio1=0;
if FixedRatio2=1 then FixedRatio1=1;

if LLN2=0 then LLN1=0;
if LLN2=1 then LLN1=1;

/*Override pre-bronchodilator values with post-bronchodilator values*/
if COPD2=0 then COPD1=0;
if COPD2=1 then COPD1=1;

current_asthma=.;
if mcq010=1 then do;
    if mcq035=1 then current_asthma=1;
    if mcq035=2 then current_asthma=0;
end;
if mcq010=2 then current_asthma=0;

/*Code in previously published definitions of COPD*/
mendy=.;
if MCQ160G=1 or MCQ160K=1 then mendy=1;
if MCQ160G=2 and MCQ160K=2 then mendy=0;

fessler=.;
fessler1=.;
if pack_years ge 10 and chronic_phlegm=1 then fessler1=1;
if pack_years < 10 or chronic_phlegm ne 1 then fessler1=0;
if pack_years=. and chronic_phlegm=. then fessler1=.;

if MCQ160G=1 or MCQ160K=1 or fessler1=1 then fessler=1;
if MCQ160G=2 and MCQ160K=2 and fessler1=0 then fessler=0;

if eligible ne 1 then delete;
in=1;
drop a i ENQ010;
run;

/*Merge in data on medication use and identify medication groups*/
data meds;set nhanes.COPD_medications;run;

proc sort data=nhanes;by seqn;run;
proc sort data=meds;by seqn;run;

data nhanes2;
merge nhanes meds;

```

```

by seqn;
if in ne 1 then delete;
if g1 ne 1 then g1=0;
run;

/*Removes participants with oxygen medication from the analysis*/

/*Code in Gradient Boosting definition of COPD*/
data nhanes3;
set nhanes2;
if oxygen=1 then delete;
GB_def=.;
if -1 < ridgeygr < 55 then GB_def=0;
if ridgeygr ge 55 then do;
    if g1=1 then GB_def=1;
    if -1 < bmxbmi < 32 and pack_years > 36 then GB_def=1;
    if -1 < bmxbmi < 32 and ocq540 > 18 then GB_def=1;
    if g1=0 and bmxbmi ge 32 then GB_def=0;
    if g1=0 and -1 < bmxbmi < 32 and -1 < pack_years le 36 and -1 < ocq540 le 18 then
        GB_def=0;
end;
in3=1;
run;

/*Stratify dataset by smoking status*/
data nonsmokers;
set nhanes3;
if smoke ne 0 then delete;
run;
data smokers;
set nhanes3;
if smoke < 1 then delete;
run;

/*Table 1 - Categorical Variables for Smokers*/
proc surveyfreq data=smokers;
    strata sdmvstra;
    cluster sdmvpsu;
    weight wtme6yr;
table riagendr race_eth dmdeduc2 hsd010_jw secondhand mcq160g mcq170k mcq010 agq030 chronic_cough
chronic_phlegm cdq010 rdq070 rdq090_jw rdq100 g1 rdq135_jw;
table copd1*(riagendr race_eth dmdeduc2 hsd010_jw secondhand mcq160g mcq170k mcq010 agq030
chronic_cough chronic_phlegm cdq010 rdq070 rdq090_jw rdq100 g1 rdq135_jw) / row;
run;

/*Table 1 - Categorical Variables for Non-smokers*/
proc surveyfreq data=nonsmokers;
    strata sdmvstra;
    cluster sdmvpsu;
    weight wtme6yr;
table riagendr race_eth dmdeduc2 hsd010_jw secondhand mcq160g mcq170k mcq010 agq030 chronic_cough
chronic_phlegm cdq010 rdq070 rdq090_jw rdq100 g1 rdq135_jw;
table copd1*(riagendr race_eth dmdeduc2 hsd010_jw secondhand mcq160g mcq170k mcq010 agq030
chronic_cough chronic_phlegm cdq010 rdq070 rdq090_jw rdq100 g1 rdq135_jw) / row;
run;

/*Table 1 - Continuous Variables for Smokers*/

```

```

proc surveymeans data=smokers q1 median q3;
  strata sdmvstra;
  cluster sdmvpsu;
  weight wtme6yr;
var ridgeyr indfmpir bmxbmi pack_years ocq520 ocq540 ocq560 ocq580;
domain COPD1;
run;

/*Table 1 - Continuous Variables for Non-smokers*/
proc surveymeans data=nonsmokers q1 median q3;
  strata sdmvstra;
  cluster sdmvpsu;
  weight wtme6yr;
var ridgeyr indfmpir bmxbmi pack_years ocq520 ocq540 ocq560 ocq580;
domain COPD1;
run;

/*Create main macro for error rate analysis*/
%macro error(group);
/*Analysis to get estimated COPD prevalence*/
proc surveyfreq data=&group.;
  strata sdmvstra;
  cluster sdmvpsu;
  weight wtme6yr;
table copd1 gb_def mendy fessler / cl;
run;
/*Analysis to get sensitivity and specificity*/
proc surveyfreq data=&group.;
  strata sdmvstra;
  cluster sdmvpsu;
  weight wtme6yr;
table COPD1*(gb_def mendy fessler)/ row cl;
run;
/*Analysis to get PPV, NPV, and accuracy*/
proc surveyfreq data=&group.;
  strata sdmvstra;
  cluster sdmvpsu;
  weight wtme6yr;
table (gb_def mendy fessler)*COPD1/ row cl;
run;
%mend;

/*Table 2*/
%error(smokers);
%error(nonsmokers);

/*Supplemental Table 1*/
/*Overall*/
data overall;
set nhanes.NHANES_2007_12;
run;
proc freq data=overall;
table RIDSTATR SPXNSTAT;
run;
data pre_eligible;
set overall;
if SPXNSTAT ne 1 then delete;

```

```

run;
proc freq data=pre_eligible;
table FixedRatio1 LLN1 SPXBSTAT;
run;
data post_eligible;
set pre_eligible;
if SPXBSTAT ne 1 then delete;
run;
proc freq data=post_eligible;
table FixedRatio2 LLN2;
run;

/*40-79 years of age*/
data age40_79;
set nhanes.NHANES_2007_12;
if ridgeyr < 40 then delete;
if ridgeyr > 79 then delete;
run;
proc freq data=age40_79;
table RIDSTATR SPXNSTAT;
run;
data pre_eligible40_79;
set age40_79;
if SPXNSTAT ne 1 then delete;
run;
proc freq data=pre_eligible40_79;
table FixedRatio1 LLN1 SPXBSTAT;
run;
data post_eligible40_79;
set pre_eligible40_79;
if SPXBSTAT ne 1 then delete;
run;
proc freq data=post_eligible40_79;
table FixedRatio2 LLN2;
run;

/*Currently smokes*/
data current;
set nhanes.NHANES_2007_12;
if ridgeyr < 40 then delete;
if ridgeyr > 79 then delete;
if smoke ne 2 then delete;
run;
proc freq data=current;
table RIDSTATR SPXNSTAT;
run;
data pre_eligible_current;
set current;
if SPXNSTAT ne 1 then delete;
run;
proc freq data=pre_eligible_current;
table FixedRatio1 LLN1 SPXBSTAT;
run;
data post_eligible_current;
set pre_eligible_current;
if SPXBSTAT ne 1 then delete;
run;

```

```

proc freq data=post_eligible_current;
table FixedRatio2 LLN2;
run;

/*Previously smoked*/
data former;
set nhanes.NHANES_2007_12;
if ridgeyr < 40 then delete;
if ridgeyr > 79 then delete;
if smoke ne 1 then delete;
run;
proc freq data=former;
table RIDSTATR SPXNSTAT;
run;
data pre_eligible_former;
set former;
if SPXNSTAT ne 1 then delete;
run;
proc freq data=pre_eligible_former;
table FixedRatio1 LLN1 SPXBSTAT;
run;
data post_eligible_former;
set pre_eligible_former;
if SPXBSTAT ne 1 then delete;
run;
proc freq data=post_eligible_former;
table FixedRatio2 LLN2;
run;

/*Never smoked*/
data never;
set nhanes.NHANES_2007_12;
if ridgeyr < 40 then delete;
if ridgeyr > 79 then delete;
if smoke ne 0 then delete;
run;
proc freq data=never;
table RIDSTATR SPXNSTAT;
run;
data pre_eligible_never;
set never;
if SPXNSTAT ne 1 then delete;
run;
proc freq data=pre_eligible_never;
table FixedRatio1 LLN1 SPXBSTAT;
run;
data post_eligible_never;
set pre_eligible_never;
if SPXBSTAT ne 1 then delete;
run;
proc freq data=post_eligible_never;
table FixedRatio2 LLN2;
run;

/*COPD+*/
data positive;
set nhanes.NHANES_2007_12;

```

```

if ridgeyr < 40 then delete;
if ridgeyr > 79 then delete;
if smoke =. then delete;
if COPD2=1 then COPD1=1;
if COPD2=0 then COPD1=0;
if SPXNSTAT ne 1 then delete;
if COPD1 ne 1 then delete;
run;
proc freq data=positive;
table FixedRatio1 LLN1 SPXBSTAT;
run;
data post_eligible_positive;
set positive;
if SPXBSTAT ne 1 then delete;
run;
proc freq data=post_eligible_positive;
table FixedRatio2 LLN2;
run;

/*COPD-*/
data negative;
set nhanes.NHANES_2007_12;
if ridgeyr < 40 then delete;
if ridgeyr > 79 then delete;
if smoke =. then delete;
if COPD2=1 then COPD1=1;
if COPD2=0 then COPD1=0;
if SPXNSTAT ne 1 then delete;
if COPD1 ne 0 then delete;
run;
proc freq data=negative;
table FixedRatio1 LLN1 SPXBSTAT;
run;
data post_eligible_negative;
set negative;
if SPXBSTAT ne 1 then delete;
run;
proc freq data=post_eligible_negative;
table FixedRatio2 LLN2;
run;

/*Supplemental Table 2*/
data former;
set smokers;
if smoke ne 1 then delete;
run;
data current;
set smokers;
if smoke ne 2 then delete;
run;
%error(former);
%error(current);

/*Supplemental Table 3*/
data smokers_f;
set smokers;
if riagendr ne 2 then delete;

```

```

run;
data nonsmokers_f;
set nonsmokers;
if riagendr ne 2 then delete;
run;
data smokers_m;
set smokers;
if riagendr ne 1 then delete;
run;
data nonsmokers_m;
set nonsmokers;
if riagendr ne 1 then delete;
run;
%error(smokers_f);
%error(nonsmokers_f);
%error(smokers_m);
%error(nonsmokers_m);

/*Supplemental Table 4*/
data smokers_young;
set smokers;
if age_2lev ne 0 then delete;
run;
data nonsmokers_young;
set nonsmokers;
if age_2lev ne 0 then delete;
run;
data smokers_old;
set smokers;
if age_2lev ne 1 then delete;
run;
data nonsmokers_old;
set nonsmokers;
if age_2lev ne 1 then delete;
run;
%error(smokers_young);
%error(nonsmokers_young);
%error(smokers_old);
%error(nonsmokers_old);

/*Supplemental Table 5*/
data smokers_nhw;
set smokers;
if race_eth ne 3 then delete;
run;
data nonsmokers_nhw;
set nonsmokers;
if race_eth ne 3 then delete;
run;
data smokers_nhb;
set smokers;
if race_eth ne 2 then delete;
run;
data nonsmokers_nhb;
set nonsmokers;
if race_eth ne 2 then delete;
run;

```

```

data smokers_h;
set smokers;
if race_eth ne 1 then delete;
run;
data nonsmokers_h;
set nonsmokers;
if race_eth ne 1 then delete;
run;
%error(smokers_nhw);
%error(nonsmokers_nhw);
%error(smokers_nhb);
%error(nonsmokers_nhb);
%error(smokers_h);
%error(nonsmokers_h);

/*Supplemental Table 6*/
data smokers_asth;
set smokers;
if mcq010 ne 1 then delete;
run;
data nonsmokers_asth;
set nonsmokers;
if mcq010 ne 1 then delete;
run;
data smokers_noasth;
set smokers;
if mcq010 ne 2 then delete;
run;
data nonsmokers_noasth;
set nonsmokers;
if mcq010 ne 2 then delete;
run;
%error(smokers_asth);
%error(nonsmokers_asth);
%error(smokers_noasth);
%error(nonsmokers_noasth);

/*Supplemental Table 7*/
/*Create alternate macro for when a measure other than COPD1 is being used by sensitivity analysis*/
%macro error2(group,measure);
/*Analysis to get estimated COPD prevalence*/
proc surveyfreq data=&group.;
strata sdmvstra;
cluster sdmvpsu;
weight wtmec6yr;
table &measure. gb_def mendy fessler / cl;
run;
/*Analysis to get sensitivity and specificity*/
proc surveyfreq data=&group.;
strata sdmvstra;
cluster sdmvpsu;
weight wtmec6yr;
table &measure.*(gb_def mendy fessler)/ row cl;
run;
/*Analysis to get PPV, NPV, and accuracy*/
proc surveyfreq data=&group.;

```

```

strata sdmvstra;
cluster sdmvpsu;
weight wtmecc6yr;
table (gb_def mendy fessler)*&measure./ row cl;
run;
%mend;

data smokers_lln;
set smokers;
if lln1=. then delete;
run;
data nonsmokers_lln;
set nonsmokers;
if lln1=. then delete;
run;
%error2(smokers,fixedratio1);
%error2(nonsmokers,fixedratio1);
%error2(smokers_lln,lln1);
%error2(nonsmokers_lln,lln1);

/*Supplemental Table 8*/
data smokers_nomild;
set smokers;
if COPD1=1 and mild_COPD=1 then delete;
if COPD1=1 and mild_COPD=. then delete;
run;
data nonsmokers_nomild;
set nonsmokers;
if COPD1=1 and mild_COPD=1 then delete;
if COPD1=1 and mild_COPD=. then delete;
run;
data smokers_nosevere;
set smokers;
if COPD1=1 and mild_COPD=0 then delete;
if COPD1=1 and mild_COPD=. then delete;
run;
data nonsmokers_nosevere;
set nonsmokers;
if COPD1=1 and mild_COPD=0 then delete;
if COPD1=1 and mild_COPD=. then delete;
run;
%error(smokers_nomild);
%error(nonsmokers_nomild);
%error(smokers_nosevere);
%error(nonsmokers_nosevere);

/*Supplemental Table 9*/
data nhanes_full;
set set nhanes.NHANES_2007_12;
if ridgeyr < 40 then delete;
if ridgeyr > 79 then delete;
if RIDSTATR ne 2 then delete;
all=1;
run;

proc sort data=nhanes_full;by seqn;run;
proc sort data=nhanes3;by seqn;run;

```

```

data weight;
merge nhanes_full nhanes3;
by seqn;
if in3 ne 1 then in3=0;
if SPXBSTAT > 1 then SPXBSTAT =0;
run;

proc freq data=weight;
table SPXBSTAT copd2;
run;

/*2-Stage Re-weighting*/
/*1st Stage: Re-weight completed post-bronch to mimic eligible for post-bronch*/
data first;
set weight;
if SPXBSTAT=. then delete;
if mcq010=2 then mcq040=2;
if mcq035=2 then mcq040=2;
run;

proc surveylogistic data=first;
strata sdmvstra;
cluster sdmvpsu;
weight wtme6yr;
class race_eth riagendr mendy mcq040; /*mendy as a surrogate for self-reported COPD*/
model SPXBSTAT (ref='0') = ridgeyr race_eth riagendr mendy pack_years mcq040 / link=glogit;
output out=stage1 p=prob;
run;

data stage1a;
set stage1;
if SPXBSTAT=1 then w1=wtme6yr/prob;
if SPXBSTAT=0 then delete;
if prob=. then delete;
keep seqn w1;
run;

/*Transition: Merge new weights back into the full dataset*/
proc sort data=weight;by seqn;run;
proc sort data=stage1a;by seqn;run;

data second;
merge weight stage1a;
by seqn;
if w1 ne . then wtme6yr=w1;
if mcq010=2 then mcq040=2;
if mcq035=2 then mcq040=2;
mendy=.;
if MCQ160G=1 or MCQ160K=1 then mendy=1;
if MCQ160G=2 and MCQ160K=2 then mendy=0;
if smoke=0 then pack_years=0;
if SPXNSTAT ne 1 then SPXNSTAT=0;
run;

/*2nd Stage: Re-weight completed pre-bronch to mimic eligible for pre-bronch*/
proc surveylogistic data=second;

```

```

strata sdmvstra;
cluster sdmvpsu;
weight wtme6yr;
class race_eth riagendr mendy mcq040; /*mendy as a surrogate for self-reported COPD*/
model SPXNSTAT (ref='0') = ridgeyr race_eth riagendr mendy pack_years mcq040 / link=glogit;
output out=stage2 p=prob;
run;

data stage2a;
set stage2;
if SPXNSTAT=1 then w2=wtme6yr/prob;
if SPXNSTAT=0 then delete;
if prob=. then delete;
keep seqn w2;
run;

/*Transition: Merge new weights back into the analysis dataset*/
proc sort data=nhanes3;by seqn;run;
proc sort data=stage2a;by seqn;run;

data nhanes5;
merge nhanes3 stage2a;
by seqn;
if w2 ne . then wtme6yr=w2;
if mcq010=2 then mcq040=2;
if mcq035=2 then mcq040=2;
if COPD2=0 then COPD1=0;
if copd2=1 then copd1=1;
run;

data nonsmokers_rw;
set nhanes5;
if smoke ne 0 then delete;
run;
data smokers_rw;
set nhanes5;
if smoke < 1 then delete;
run;
%error(smokers_rw);
%error(nonsmokers_rw);

/*Supplemental Table 10*/
%error2(smokers,copd1b);
%error2(nonsmokers,copd1b);

/*Supplemental Table 11*/
data post_bronch_smokers;
set smokers;
if SPXBSTAT > 1 then delete;
run;
data post_bronch_nonsmokers;
set nonsmokers;
if SPXBSTAT > 1 then delete;
run;
%error(post_bronch_smokers);
%error(post_bronch_nonsmokers);

```

## References

- 1 Mohadjer, L. & R Curtin, L. Balancing sample design goals for the National Health and Nutrition Examination Survey. *Survey Methodology Statistics Canada* **34**, 119-126 (2008).
- 2 West, B. T., Sakshaug, J. W. & Aurelien, G. A. S. Accounting for Complex Sampling in Survey Estimation: A Review of Current Software Tools. **34**, 721, doi:<https://doi.org/10.2478/jos-2018-0034> (2018).
- 3 Horvitz, D. G. & Thompson, D. J. A Generalization of Sampling Without Replacement from a Finite Universe. *Journal of the American Statistical Association* **47**, 663-685, doi:10.1080/01621459.1952.10483446 (1952).
- 4 Heeringa, S. G., West, B. T. & Berglund, P. A. *Applied Survey Data Analysis*. 2nd Edition edn, (Chapman and Hall/CRC, 2017).
- 5 Binder, D. On the variances of asymptotically normal estimators from complex surveys. *International Statistical Review/Revue Internationale de Statistique*, 272-292 (1983).
- 6 Toth, D. rpms: Recursive Partitioning for Modeling Survey Data. *R package version 0.4.0* (2019).
- 7 Chen, T. & Guestrin, C. XGBoost: A Scalable Tree Boosting System. *eprint arXiv:1603.02754* (2016).
- 8 Mandy, A. *et al.* House Dust Endotoxin Association with Chronic Bronchitis and Emphysema. *Environ Health Perspect* **126**, 037007, doi:10.1289/EHP2452 (2018).
- 9 Fessler, M. B. *et al.* House Dust Endotoxin and Peripheral Leukocyte Counts: Results from Two Large Epidemiologic Studies. *Environ Health Perspect* **125**, 057010, doi:10.1289/EHP661 (2017).