

# Association between health insurance access and chronic obstructive pulmonary disease (COPD) among adults with asthma in the United States

Mango Tango - PJ

2023-10-17

Total word count: XX

## 1 Methods (XX words)

### 1.1 Study populations

This study includes respondents from non-institutionalized United States (US) civilian adults aged 20-80 years with asthma, in all 50 states and Washington D.C. Data for the study was extracted from the National Health and Nutrition Examination Survey (NHANES). NHANES collects data through interviews and physical examinations to study the health and nutritional status of adults and children in the US. It has a publicly available database released by the National Center for Health Statistics (NCHS). NHANES uses a four-stage clustered sampling design and produces a nationally representative sample [1,2].

NHANES releases public data in two-year cycles, but the 2019-2020 cycle was not completed due to the coronavirus (COVID-19) pandemic. Therefore data collected from 2019 to March 2020 (pre-pandemic) were combined with NHANES 2017-2018 to get a nationally representative sample [2]. To increase the sample size for this study, the 2017-2020 pre-pandemic and 2015-2016 releases were combined for this study for a total of 5.2 years.

There are 9,971 records (respondents) in the 2015-2016 cycle and 15,560 in the 2017-March 2020 cycle. Therefore the full dataset contains 25,531 records. From the full data 21,766 non-asthma patient records were excluded. Further 1,593 respondents younger than 20 years old and older than 80 years, were removed. Missing values for the exposure and outcome were defined as exclusion criteria. The final analytic sample included 2,154 respondents (see Figure 1).

### 1.2 Study variables

The **outcome** of the study is if the participants have been diagnosed with chronic obstructive pulmonary disease (COPD). The outcome is based on questions, “*Have you ever been told by a doctor or other health professional that you had chronic obstructive pulmonary disease or COPD, emphysema, or chronic bronchitis?*” in the 2017-March 2020 cycle, and the individual questions for emphysema, chronic bronchitis, and COPD in the 2015-2016 cycle. Participants who responded “Yes” to either one of the 3 diseases in the 2015-2016 data were coded as “Yes”. Respondents who responded “No” to all diseases were coded as “No”.

The primary **exposure** of interest is the respondent’s access to health insurance. This was obtained through the question “*Are you covered by health insurance or some other kind of health care plan?*”, as a dichotomous outcome.

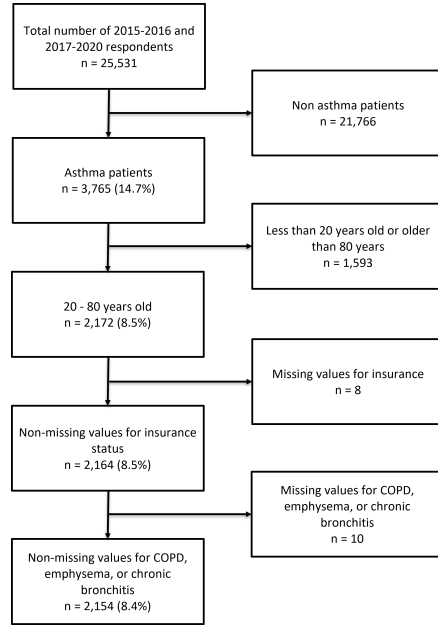


Figure 1: Flowchart of the study population and sample size for the analysis.

Confounders used in this analysis were selected based on the literature. These include the number of times the respondent received healthcare over past year (healthcare use), lung cancer, smoking status, number of smokers inside the home, and diabetes. **Healthcare use** was taken from the question “*During the past 12 months, how many times have you seen a doctor or other health care professional about your health at a doctor’s office, a clinic or some other place?*”.

**Lung cancer** was defined from the set of questions for the type of cancer. If the respondents selected lung cancer for any one of the corresponding variables they were considered to be diagnosed with lung cancer. If participants said “No” to the question, “*Have you ever been told by a doctor or other health professional that you had cancer or a malignancy of any kind?*”, they are defined as not having lung cancer.

**Smoking status** was defined from questions “*Do you now smoke cigarettes?*” and “*Have you smoked at least 100 cigarettes in your entire life?*”. If the respondents smoke every day or some days, they are considered “current smokers”. If the respondents are not smoking now but have smoked more than 100 cigarettes in their lifetime, they are considered as “former smokers”. If they haven’t smoked 100 cigarettes throughout life, they are considered “never smoked”. **Number of smokers inside the home** was directly obtained through the question “*Not counting decks, porches, or detached garages, how many people who live here smoke cigarettes, cigars, little cigars, pipes, water pipes, hookah, or any other tobacco product inside this home?*”.

Sex and age, in years, at the time of screening were used as other covariates in this analysis. For all above mentioned variables if the respondents answered “Refused” or “Don’t know”, the response was considered missing.

### 1.3 Statistical analyses

NHANES provides sampling weights for each respondent after accounting for the complex sampling design and oversampling criteria [1,2]. For the analysis to be representative of the population survey features were used in this analysis. Since multiple cycles were combined survey weights should be adjusted accordingly.

Guidelines specify multiplying 2015-2016 weights by  $\frac{2}{5.2}$  and 2017-March 2020 weights by  $\frac{3.2}{5.2}$ . All variables used in this study were collected through the household questionnaire. Therefore interview weights were used in this analysis [2].

Descriptive statistics were obtained with medians, 1st and 3rd quartiles for numerical variables and percentages for categorical variables. Crude and adjusted odds ratios (OR) were obtained using logistic regression models. Adjustment variables used in this study are healthcare use, smoking status, number of smokers inside the home, and diabetes. 95% confidence intervals (CI) and two-sided p-values based on the Student t distribution were also obtained [3].

As the next steps of the analysis, missing values will be imputed using multiple imputation (MI) based on Fully Conditional Specification (FCS) [4]. FCS will be used in this data assuming that data are missing at random (MAR) [5]. Predictive mean matching will be used when imputing missing numeric variables, logistic regression imputation for dichotomous variables and polytomous regression imputation for categorical data with more than 2 levels [4].

All analysis were conducted using R version 4.3.1 [3]. Data manipulation was done using the `dplyr` package [6] and complex survey design features were incorporated using the `survey` package [7]. Multiple imputation was performed using the `mice` package [4].

## 2 Results (XX words)

Variables used in this study, grouped by levels of the outcome and accounting for the complex survey design, are summarized in Table 1. Respondents with diagnosed COPD are older (with a median age of 57) than respondents without COPD (median age 39). There are approximately 50% males and females without COPD while more females in the sample are diagnosed with COPD. The median number of healthcare visits in the past year is similar between the two response groups. The proportion of lung cancer among respondents in this sample is close to zero (2% in the COPD group and almost 0% in the non-COPD group). There are more current and former smokers with COPD and more respondents who have never smoked in the non-COPD group. There are more respondents with diabetes in the COPD group (24%) compared to the non-COPD group (10%).

Majority of the respondents didn't answer the number smokers inside the house, with 52% missing in the COPD group and 72% in the non-COPD group.

Table 1: Characteristics of the participants by COPD status. All proportions adjusted for the survey design.

Characteristic	Overall (n=2,154) <sup>1</sup>	Has COPD (n=508) <sup>1</sup>	Doesn't have COPD (n=1,646) <sup>1</sup>
Has insurance			
Yes	88%	91%	87%
No	12%	9%	13%
Age (years)	43.0 (29.0, 58.3)	57.0 (44.9, 65.0)	39.0 (27.0, 55.0)
Sex			
Male	42%	33%	45%
Female	58%	67%	55%
No. of healthcare visits	2.0 (1.0, 4.0)	3.0 (2.0, 6.0)	2.0 (1.0, 4.0)
Lung cancer			
Yes	0%	2%	0%
No	99%	97%	100%
Missing	0%	1%	0%
Smoking status			
Current smokers	21%	37%	17%

Former smokers	27%	31%	26%
Never smoked	51%	33%	56%
Missing	0%	0%	0%
No. of people who smoke inside			
0	20%	24%	19%
1	7%	13%	6%
2	4%	8%	3%
3	1%	2%	1%
Missing	68%	53%	72%
Has diabetes			
Yes	13%	24%	10%
No	85%	72%	88%
Borderline	2%	3%	2%
Missing	0%	0%	0%

<sup>1</sup>%; Median (IQR)

Crude OR for the outcome (COPD) and exposure (insurance status) was calculated for the complete data, containing 738 records. The OR was 1.55 with a 95% confidence interval of (1.12, 2.15). Therefore, the odds of being diagnosed with COPD is 1.6 times higher in US adults without health insurance, compared to US adults with health insurance.

Adjusted OR from the logistic regression model for the complete data are shown in Table 2. Lung cancer is not used for OR adjustment since the number of respondents with lung cancer are near zero (Table 1). After adjusting for the confounders and other variables, the OR (1.08) shows that there is no significant difference (p-value = 0.80) in the odds of being diagnosed with COPD between US adults with or without health insurance.

Table 2: Adjusted odds ratios. All proportions adjusted for the survey design.

Characteristic	OR <sup>1</sup>	95% CI <sup>1</sup>	p-value
Has insurance			
Yes	—	—	
No	1.10	0.58, 2.10	0.8
No. of healthcare visits	0.77	0.69, 0.86	<0.001
Smoking status			
Current smokers	—	—	
Former smokers	3.01	1.32, 6.83	0.010
Never smoked	4.15	2.06, 8.37	<0.001
No. of people who smoke inside	0.87	0.60, 1.25	0.4
Has diabetes			
Yes	—	—	
No	2.53	1.32, 4.85	0.007
Borderline	3.19	0.79, 12.8	0.10
Age (years)	0.94	0.93, 0.96	<0.001
Sex			
Male	—	—	
Female	0.54	0.28, 1.06	0.071

<sup>1</sup>OR = Odds Ratio, CI = Confidence Interval

## Results in progress

The reported OR uses the complete data with no missing values. This approach might not be suitable since there is a variable with high missing values. Therefore, multiple imputation would be used in the future submission to impute missing data. Before conducting the missing data analysis, data missing at random assumptions will be evaluated.

A sensitivity analysis will be carried out with the “duration since last healthcare visit” replacing the number of visits, already included in the model.

## PICOT

	Description
P	Non-institutionalized United States civilian adults aged 20-80 years with asthma, in all 50 states and Washington D.C.
I	Individuals without access to health insurance
C	Individuals with access to health insurance
O	Chronic obstructive pulmonary disease (COPD)
T	NHANES 2015-March 2020 (pre-pandemic)

## Changes since the SAP

- The population was changed with people 80 years and older removed. This was done because NHANES reports 80 and above as 80.
- The variable “close relatives with asthma” was removed following reviewer feedback.
- “Frequency of healthcare visits” (HUQ051) was considered in the base case instead of “duration since last healthcare visit”, following reviewer feedback. The duration will still be used in the sensitivity analysis.
- SMQ040 (“Do you now smoke cigarettes?”) was combined with SMQ020 (“Smoked at least 100 cigarettes in life”) to define smoking status. The new definition is mentioned in Section 1.2.

## 3 References

- [1] Chen T-C, Clark J, Riddles MK, Mohadjer LK, Fakhouri THI. National Health and Nutrition Examination Survey, 2015-2018: Sample Design and Estimation Procedures. Vital and Health Statistics Series 2, Data Evaluation and Methods Research 2020:1–35.
- [2] Akinbam L, Chen T-C, Davy O, Ogden C, Fink S, Clark J, et al. National Health and Nutrition Examination Survey, 2017–March 2020 Prepandemic File: Sample Design, Estimation, and Analytic Guidelines. National Center for Health Statistics (U.S.); 2022. <https://doi.org/10.15620/cdc:115434>.
- [3] R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2023.
- [4] Buuren S van, Groothuis-Oudshoorn K. Mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software 2011;45:1–67. <https://doi.org/10.18637/jss.v045.i03>.
- [5] Liu Y, De A. Multiple Imputation by Fully Conditional Specification for Dealing with Missing Data in a Large Epidemiologic Study. International Journal of Statistics in Medical Research 2015;4:287–95. <https://doi.org/10.6000/1929-6029.2015.04.03.7>.
- [6] Wickham H, François R, Henry L, Müller K, Vaughan D. Dplyr: A Grammar of Data Manipulation. 2023.

- [7] Lumley T. Complex Surveys: A Guide to Analysis Using R: A Guide to Analysis Using R. John Wiley; Sons; 2010.

## 4 Appendix

Table 4: Characteristics of the participants by COPD status ignoring complex sampling design.

Characteristic	Overall (n=2,154) <sup>1</sup>	Has COPD (n=508) <sup>1</sup>	Doesn't have COPD (n=1,646) <sup>1</sup>
Has insurance			
Yes	86%	89%	85%
No	14%	11%	15%
Age (years)	47.0 (32.0, 61.0)	58.0 (46.0, 66.0)	42.5 (29.0, 58.0)
Sex			
Male	42%	40%	43%
Female	58%	60%	57%
No. of healthcare visits	2.0 (2.0, 4.0)	3.0 (2.0, 6.0)	2.0 (1.0, 3.0)
Lung cancer			
Yes	1%	3%	0%
No	99%	97%	100%
Missing	0%	1%	0%
Smoking status			
Current smokers	23%	37%	18%
Former smokers	24%	31%	22%
Never smoked	53%	32%	60%
Missing	0%	0%	0%
No. of people who smoke inside			
0	19%	22%	18%
1	9%	16%	7%
2	5%	7%	4%
3	1%	2%	1%
Missing	66%	53%	70%
Has diabetes			
Yes	18%	30%	14%
No	80%	66%	84%
Borderline	3%	4%	2%
Missing	0%	0%	0%

<sup>1</sup>%; Median (IQR)

Table 5: Adjusted odds ratios ignoring complex sampling design.

Characteristic	OR <sup>1</sup>	95% CI <sup>1</sup>	p-value
Has insurance			
Yes	—	—	
No	0.99	0.60, 1.67	>0.9
No. of healthcare visits	0.83	0.76, 0.90	<0.001
Smoking status			
Current smokers	—	—	

Former smokers	2.36	1.42, 3.99	0.001
Never smoked	4.60	2.83, 7.69	<0.001
No. of people who smoke inside	0.98	0.79, 1.22	0.9
Has diabetes			
Yes	—	—	
No	1.45	0.90, 2.33	0.13
Borderline	1.56	0.52, 5.13	0.4
Age (years)	0.95	0.93, 0.96	<0.001
Sex			
Male	—	—	
Female	0.78	0.53, 1.12	0.2

<sup>i</sup>OR = Odds Ratio, CI = Confidence Interval