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

Pramoda Jayasinghe

2023-10-15

Methods

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. The 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. NHANES has a publicly available database released by the National Center for Health Statistics (NCHS) [1,2].

To increase the sample size two data cycles were combined for this analysis. 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]. The 2017-2020 pre-pandemic and 2015-2016 cycles were used for this study for a total of 5.2 years.

ADD SAMPLE SIZES. The sample sizes for the study is shown in Figure 1.

Study variables

The outcome of the study is if the participants have been diagnosed with Chronic obstructive pulmonary disease (COPD). For this information 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 cycles were used. 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 respondents access to health insurance. This was obtained through the question "Are you covered by health insurance or some other kind of health care plan?".

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 "What kind of cancer?". If the respondents selected lung cancer for any one of the corresponding variables they were considered to be diagnosed with lung
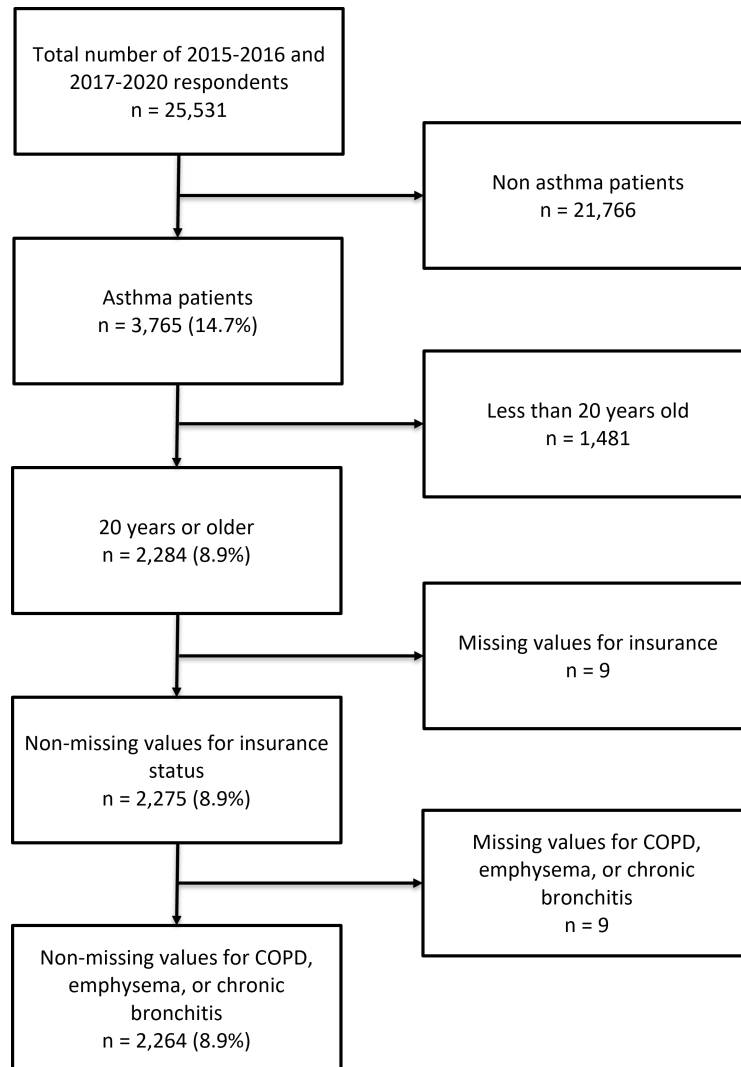
Figure 1: Flowchart of the study population and sample size for the analysis.

cancer. If participants said "No" to 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, 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.

Statistical analyses

NHANES uses a four-stage clustered sampling design and provides sampling weights for each respondent after accounting for the said sampling design and oversampling criteria [1,2]. To account for this complex sampling design, survey weights were used. Since multiple cycles were combined sample 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 design adjusted medians, 1st and 3rd quartiles for numerical variables and design adjusted percentages for categorical variables. Missing values with percentages less than 5% were imputed using multiple imputation based on Fully Conditional Specification (FCS) [3]. FCS was used in this data assuming that data are missing at random (MAR) [4]. Predictive mean matching was used when imputing missing numeric variables, logistic regression imputation for dichotomous variables and polytomous regression imputation for categorical data with more than 2 levels.

Odds ratios (OR) were obtained using a logistic regression model while adjusting for variables healthcare use, lung cancer, smoking status, number of smokers inside the home, and diabetes.

All analysis were conducted using R version 4.3.1 [5]. 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 [3].

Results

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

Characteristic	Overall[1]	Has COPD[1]	Doesn't have COPD[1]
Has insurance			
Yes	88%	92%	87%
No	12%	8%	13%
Age (years)	44.0 (29.9, 60.0)	57.0 (45.0, 67.0)	40.0 (27.0, 56.0)
Close relative with asthma			
Yes	44%	46%	44%
No	52%	49%	53%
Missing	4%	5%	3%
ED visits for asthma/past yr			
Yes	7%	15%	5%

No	49%	60%	46%
Missing	44%	24%	49%
Lung cancer			
Yes	0%	2%	0%
No	12%	21%	9%
Missing	88%	77%	91%
Smoked more than 100 cigarettes/life			
Yes	48%	66%	43%
No	52%	33%	57%
Missing	0%	0%	0%
No. of people who smoke inside			
0	19%	23%	18%
1	7%	13%	5%
2	4%	7%	3%
3	1%	2%	1%
Missing	69%	55%	72%

[1] %; Median (IQR)

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

Characteristic	OR[1]	95% CI[1]	p-value
Has insurance			
Yes	—	—	
No	73.9	0.14, 38,445	0.15
Age (years)	1.06	1.0, 1.12	0.069
Close relative with asthma			
Yes	—	—	
No	0.15	0.02, 1.07	0.056
ED visits for asthma/past yr			
Yes	—	—	
No	0.68	0.04, 12.8	0.8
Lung cancer			
Yes	—	—	
No	16,876,072	2,795,327, 101,884,965	<0.001
Smoked more than 100 cigarettes/life			
Yes	—	—	
No	1.61	0.15, 17.7	0.7
No. of people who smoke inside	0.67	0.25, 1.80	0.4

[1] OR = Odds Ratio, CI = Confidence Interval

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]   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.

[4]   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.

[5]   R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2023.

[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.

Appendix

Add appendix