STATISTICAL ANALYSIS OF RISK FACTORS Contributing to Chronic Obstructive Pulmonary Disease Prevalence In The Southeast United States

Erick Bora
Colorado School of Mines

ABSTRACT

Chronic Obstructive Pulmonary Disease (COPD), currently the fourth leading cause of death worldwide, is projected to become more problematic by 2030 unless appropriate actions are taken to reduce the major risk factors that contribute to COPD. However, the spatial distribution of individuals with COPD and the many potential factors that contribute to this disease make prevention methods difficult to discern at a national level. Thus, COPD must be evaluated at a local scale. This study analyzed the spatial distribution of COPD prevalence per county for the area south of the Ohio River and east of the Mississippi River which has experienced the highest COPD prevalence in the United States. Additionally, this study statistically analyzed the significance of socioeconomics, air quality, physical inactivity, household size, smoking, and occupation in predicting observed rates of COPD. Results show that high COPD rates clustered in eastern Kentucky, western Virginia, and southern Missouri, which correlated with areas of high smoking rates. Smokers, age, construction workers, healthcare workers, and physical inactivity were statistically significant in predicting high COPD prevalence. Thus, medical and healthcare professionals should focus on addressing these issues in this area of the country to reduce the risk of COPD.

Keywords: Chronic Obstructive Pulmonary Disease, Geographic Information Systems (GIS), Occupation, Regression, Risk Factors

Introduction

Chronic Obstructive Pulmonary Disease (COPD) is an umbrella term used to describe the progressive deterioration of lung strength and function due to a broad range of lung diseases, which include emphysema, chronic bronchitis, and asthma. COPD is a progressive respiratory disease, thus its severity increases overtime. Individuals affected by COPD suffer from the airways in their lungs becoming inflamed and thickened, thereby destroying the tissues where oxygen is exchanged, which makes it hard to breathe. Symptoms of COPD include shortness

of breath during everyday activities, a lingering cough (dyspnea), and fatigue (American Lung Association 2019). Additionally, overtime COPD weakens the rest of the body, increasing the likelihood for the individual to develop other serious chronic diseases including heart attacks and strokes.

According to the most recent Global Burden of Disease Estimate, approximately 174.5 million individuals worldwide have been diagnosed with COPD, which equates to approximately 2.3 percent of the world's population (Hart et al. 2018). In 2015, approximately 3.2 million deaths were caused by COPD (Hart et al. 2018). As of 2018, COPD is the fourth leading cause of death worldwide, behind cancer, heart disease, and stroke (Kauhl et al. 2018). By 2030, COPD is projected to become the third leading cause of death unless necessary steps are taken to reduce the underlying risk factors of COPD, specifically reduce smoking (WHO 2018).

COPD prevalence is higher in the Americas more than any other region of the world. In the United States, almost 15.7 million individuals have been diagnosed with COPD, which is approximately 6.4 percent of the country's total adult population (Wheaton et al. 2015). Within the United States, the area with the highest age adjusted prevalence of COPD is in the region to the south and east of the Mississippi and Ohio Rivers (Wheaton et al. 2015).

Many factors affect the severity and prevalence of COPD. Of most significance is cigarette smoking; approximately 80 percent of all COPD cases worldwide are caused by cigarette smoking (Liu et al. 2018). Cigarette smoke releases over 7,000 harmful chemicals and substances into the lungs which weakens the lung's defense against infection, narrows airway passages, causes swelling in air tubes, and destroys air sacs – all of which are contributing factors to COPD (American Lung Association 2019). Another contributor of COPD is air pollution and exposure to particulate matter or secondhand smoke. Nearly 20 percent of all COPD cases are among people who are non-smokers (Olloquequi et al. 2018). Biomass smoke, smoke caused from the burning of organic matter, is referred to as the main cause of COPD among non-smokers (Olloquequi et al. 2018). Particulate matter in the atmosphere caused by air pollution or biomass smoke has a similar effect on the lungs as smoking cigarettes in that the toxic particles clog lung airways and weaken the lungs' defense against infections.

COPD is most common in individuals aged 55 years old or older, which accounts for approximately 35 percent of all COPD cases in the United States (Wheaton et al. 2015). This is mainly because individuals are not diagnosed with COPD until after they have lost 50 percent of their lung function (Wheaton et al. 2015). Overtime, the lungs start to weaken, and individuals become more prone to other significant respiratory diseases. COPD causes progressive damage to a person's lungs which increases an individual's chance of developing other chronic diseases (Gaunt 2020). COPD tends to affect females (specifically elderly) more so than males – 8.9 million to 6.7 million cases in the United States – due to higher use of tobacco among women in higher income countries, like the United States (Wheaton et al. 2015). Women also have a higher risk of exposure to indoor air pollution from biomass smoke created from cooking or heating, which has caused female COPD cases to increase in recent years (WHO 2019). Female lungs are also more susceptible to lung damage from cigarette smoke and particulate matter because they are smaller than male lungs (American Lung Association 2019).

Additional factors that influence the likelihood of developing COPD include physical activity and living conditions. It has been well documented that increased physical activity in COPD patients can lead to fewer hospitalizations and reduce mortality rate (Boeselt et al. 2016). Due to progressively decreased lung strength, COPD patients have significantly reduced duration, intensity, and step count numbers due to the associated muscle weakness and limited mobility from COPD (Boeselt et al. 2016). While physical activity may be challenging for COPD patients, exercise is an important component of pulmonary rehabilitation to increase lung strength and function (Wheaton et al. 2015). Additionally, average household size may contribute to COPD. A study conducted in Germany (Kauhl et al. 2018) found that persons residing in a living community or shared apartments of unmarried people in steady relationships were at a high risk for exposure to secondhand smoke, and thus at higher risk for developing COPD.

Race has also played an important role in COPD prevalence. Between 2000 and 2014, the age-adjusted rates of death due to COPD declined by 21.1 percent in Non-Hispanic White (NHW) males and by 24.4 percent in black males, but increased in black females by 4.2 percent with no change in rates for NHW females (Ejike et al. 2019). It is believed that COPD prevalence in African American individuals will increase in the coming years if the prevalence of smoking continues to remain high in black individuals than white individuals (Ejike et al. 2019). In a study conducted by Chatila et al. (2004), showed black patients who were diagnosed with COPD were younger and had lower cumulative tobacco smoke exposure with similar lung function to their white counterparts. This is attributed to varying socio-economic factors including household size and median income. A study conducted by Hankinson et al. (1999) indicated that black individuals tend to have lower lung function than white individuals of the same age. This difference has been attributed in part to smaller trunk/leg ratios in black individuals. On average, black individuals smoke fewer cigarettes and start smoking at an older age than white individuals, but black individuals are more likely to die of smoking related diseases (i.e. COPD) than white individuals. It is still unclear what causes this trend, but black individuals appear to have greater nicotine intake from tobacco smoke and lower renal clearance of cotinine, which indicates recent exposure to tobacco smoke (Ejike et al. 2019). By comparison, in a U.S. study conducted by Wheaton et al. (2015), Asian individuals had the lowest age-adjusted COPD prevalence of any race (2.0 percent) compared to white (6.3 percent) and black individuals (6.5 percent).

The effect of industry and job status on COPD prevalence has yet to be considered as a contributor to COPD in the United States. Larger urban areas are characterized by higher pollution levels, including particulates. A study in Germany indicates that a strong correlation exists between mining or steel manufacturing industries and air pollution (Aschan-Leygonie et al. 2013). Urban areas, which are more developed, generally have more industries and thus will likely have more particulate matter in the atmosphere compared to a rural setting. Therefore, certain occupations could be at a higher risk for developing COPD depending on exposure to higher levels of particulate matter in occupations where most of the time is spent outside (farmer, construction worker) and/or likelihood to smoke due to stress from an occupation (nurse or teacher).

Medical service involves care-taking of individuals who are sick or injured and any mistakes or errors could be costly and sometimes fatal (Familoni 2008). Not only are medical professionals prone to the same stresses as the general population, but these individuals must also cope with increased stress from their professions. According to a report published by the American Foundation for Suicide Prevention, on average, death by suicide is 70 percent most likely in male doctors than among other professionals and 250-400 percent higher among female doctors, with the major causes by stress and depression caused by the jobs (Familoni 2008). Early individual behavioral reactions among medical students may include onset or increased smoking or alcohol use as a way to cope with the increased stress (Familoni 2008). Medical students and professionals can also become "burnt out" which consists of emotional exhaustion, depersonalization of patients and colleagues, and low productivity/achievements which leads to an increase in stress and potential increase in smoking and alcohol use, deterioration of physical and mental health, and premature deaths (Familoni 2008).

Even though many educators feel their work is important and rewarding, a high degree of stress and "burn-out" is common in this profession. Occupational stress among teachers is associated with several factors including time pressure, discipline problems, lack of resources, lack of professional recognition, lack of support and the diversity of tasks required (Skaalvik & Skaalvik 2015). A majority of teachers tested and interviewed by Skaalvik & Skaalvik (2015) reported that workload and time pressure were extremely stressing. A typical day for a teacher can involve lesson planning, grading, meetings, and communication with students, parents and staff members. These factors accompanied by the low financial compensation can all contribute to added stress for educators who look to smoking as a form of stress release.

Both farmers and construction workers predominantly work in the outdoors, increasing their risk of exposure to harmful particulates in the atmosphere, the other major cause of COPD. A typical farmer can be subjected to various respiratory exposures on their farm including organic and inorganic dusts, bacteria, toxic gases like ammonia and hydrogen sulfide, and endotoxins (Elliot and Von Essen 2016). These particulates can enter the lungs and lead to COPD and other respiratory diseases overtime. The same is true for construction workers, who are exposed to various chemicals and particulates at a job site. According to Borup et al. (2017), the occupational exposure limit (OEL) for respirable dust is set at 5 mg/m3 in the United States, but this exposure limit was exceeded by 11 percent in the construction industry. As such, farmers and construction workers were included in this study given their association with a primary risk factor of COPD diagnosis.

Beneficiaries of this study include medical professionals and the health science communities who will gain a better understanding of which risk factors are most likely to contribute to COPD prevalence in certain areas of the United States. Knowing this information will allow doctors to better mitigate these risk factors and lower the risk of people developing COPD in the future. Understanding what major risk factors contribute to COPD development will allow medical professions to take the necessary precautions to reduce the severity of their disease and prescribe the proper treatment procedures.

The distribution of individuals suffering from COPD is high in the United States and previous research has shown that COPD prevalence is concentrated south and east of the Ohio and Mississippi Rivers. Unfortunately, the research to explain why COPD levels are high in these areas is limited. Therefore, the objective of the study was to analyze the spatial distribution of COPD prevalence and statistically determine the significance and spatial relationship of age, gender, smoking habits, air pollution, physical activity, household size, and job status to better understand the distribution patterns of COPD across the southeastern United States.

The United States is one of the most developed nations in the world, yet a limited number of studies have been conducted in the United States. Boeselt et al. (2016) evaluated the effect of physical activity as a treatment for COPD, rather than considering physical activity, or lack thereof, as a contributor to COPD prevalence. Additionally, COPD prevalence studies have used drastically variable sample sizes, with some studies as low as 20 patients (Boeselt et al. 2016). Studies that use larger sample sizes, such as this study, and use data on the general public (not just individuals with COPD) will increase the precision or confidence level in our conclusions regarding COPD prevalence distributions. Estimates always have some level of uncertainty, but a larger population size accounts for more variability, and will be more reflective of the entire population (Littler 2019).

Methods

Data for this study was collected from a variety of sources. Demographic data including population, gender, age, physical activity and household size were collected for each county from the U.S Census Bureau (2016). County data for adult smoking rates was collected from Data U.S.A (2017). Smoking data from 2017 was considered for this study, as it was the most recent data provided. In order for an individual to be included in this analysis, they must be considered a consistent smoker, to better understand the relationship between tobacco smoking and COPD crude rate. Crude COPD rates (COPD cases per 100,000 people) was provided by the Center for Disease Control (CDC) Wonder (2016). Average particulate matter (2.5 micrometers) was obtained from the U.S. Environmental Protection Agency (2018). Occupation data which included the total number of farmers, healthcare workers, construction workers, and education professionals per county were obtained from the U.S. Census Bureau (2018). County data was used for this analysis as the data was readily available through the U.S. Census Bureau and other sources. The goal of this analysis was to spatially and statistically determine where in the southeast United States COPD crude rates were created by various socioeconomic factors and county data better shows which regions of the county are affected rather than city data, as much of the population of the southeast U.S. resides in rural areas. According to the Centers for Disease Control and Prevention (CDC) (2019), adults living in rural areas are also more likely to be heavy smokers (more than fifteen cigarettes per day) than adults living in urban areas. Demographic, physical activity, smoking, COPD crude rate, particulate matter, and occupational data was organized and spatially analyzed by county using ESRI ArcGIS software.

The Tobacco Nation States are a 13-state region in the southeastern and Midwest United States stretching from Alabama to Michigan where tobacco smoking rates are higher than the rest of the nation. In the Tobacco Nation States, 21 percent of the adult population smokes, compared to 15 percent of adults in the rest of the U.S. These states have a lower household income (25 percent less than the typical U.S. resident) and cigarettes are cheaper compared to the rest of the country (Galvin 2019). Cigarettes are also cheaper in the Tobacco Nation States: $5.69 per pack compared with $7.05 in the rest of the United States (Galvin 2019). Particulate matter is also high in these states, specifically in the Appalachia region where the mountains were exposed to mountaintop removal mining which destroys the top of mountains to reach profitable coal seams. As a result, tons of particulate matter get released into the atmosphere. (Appalachian Voices 2017) Tobacco smoke and exposure to particulate matter are the most significant factors that contribute to COPD prevalence and are the reason why COPD rates are higher in the southeastern United States.

The study area assigned for this analysis encompassed eight of the thirteen Tobacco Nation States: Alabama, Arkansas, Kentucky, Louisiana, Mississippi, Missouri, South Carolina, and Tennessee. The other five Tobacco Nation States (Michigan, Indiana, Ohio, Oklahoma, and West Virginia) were not included in this analysis as the goal of this study was to focus on the southeastern United States. The other states included in this analysis were Florida, Georgia, North Carolina, and Virginia. As mentioned, smoking is not the only cause of COPD. Particulate matter is also higher in the southeast United States than other areas of the country. According to the World Health Organization (2018), the guideline for annual average fine particulate matter is 10 $\mu g/m^3$. Much of the southeastern United States sits above these levels, with the average reading for a county in the southeast United States at 12.3 $\mu g/m^3$ (Data U.S.A. 2017).

According to the U.S. Census Bureau (2018) the total population of the twelve states included in this analysis is 89,247,684. 72.45 percent of this region is comprised of white individuals, 21.53 percent are African American, and 3.04 percent are Asian (U.S. Census Bureau 2018). The average adult smoking rate for this region is 20.01 percent (Data U.S.A. 2017). The average COPD crude rate by county is 55.01 percent (CDC 2016). Missouri (67.08 percent), Kentucky (67.01 percent) and Tennessee (61.82 percent) had the highest COPD crude rates by state in the region (CDC 2016). Every state in the region was above 41 percent, with Louisiana as the lowest in the region with a 41.69 percent crude rate (CDC 2016).

All demographic and occupational data was normalized in Esri ArcGIS. Percent of individuals over 55 years of age, those most at risk for COPD was calculated by dividing the total number of individuals over 55 years of age for each county by the county population. Since females have been more affected by COPD in recent years, percent of females were considered and calculated by dividing the total number of females (in all age groups) by the total population of the county. Physical activity data was normalized by dividing the total number of people deemed physically inactive (reporting no physical activity in last 30 days) by the total population of the county. Household sizes of three or more individuals were considered for this study as it is greater than the national average of 2.53 persons (Statista 2019). Rented and owned households with more than three individuals were divided by the total number of

housing units per county. Occupational data included the total number of individuals working as farmers, construction workers, healthcare workers and educational professionals which was divided by the total number of individuals employed over the age of sixteen to calculate the percentage of individuals employed as farmers, construction workers, healthcare workers and educational professionals for each county in the southeast United States.

To determine if COPD rates per county were clustering in the southeastern United States, the Hotspot Analysis tool was used. Results spatially identify where high and low COPD crude rates cluster at the 95 percent confidence interval. To determine the proper distance for the Hotspot Analysis, the Incremental Spatial Autocorrelation Tool was used, which relates different distance values to identify the peak z-score where spatial processes promote clustering.

Kauhl et al. (2018) conducted a spatial analysis to explain which populations were at greatest risk for COPD across northeast Germany. The factors tested included: age, migration background, household size, and area deprivation. Kauhl et al (2018) concluded that a global regression model was effective and therefore was used here. To explain the observed COPD crude rates per county, the Ordinary Least Squares Regression tool, which determines which explanatory variables were significant in predicting COPD crude rates per county at the 95 percent confidence interval. The global model produced coefficients for each explanatory variable in relation to the dependent variable. P-values were the probability that the spatial pattern was created by a random process (the null hypothesis). R^2 value (percentage of the dependent variable proven by the explanatory variables), Variance Inflation Factors (signify redundancy), the Jarque-Bera Statistic, which tests the distribution of the residuals from the regression and the Koenker Statistic which determines which factors were more significant in specific locations in the study area were all obtained through the OLS. The Spatial Autocorrelation tool plotted the residuals (errors) of the analysis and generated a Z-score and P-value to determine if the error was normally distributed or clustered.

The study conducted by Kauhl et al. (2018), used a Geographically Weighted Regression (GWR) to spatially analyze which regions of northeastern Germany were at greatest risk to COPD. A GWR local regression model was successful in predicting COPD cases in Northeastern Germany and therefore was used here. The Geographically Weighted Regression tool was used to construct a linear regression model on a local scale for each county within the southeastern United States to determine which explanatory variables were most significant at predicting COPD crude rates in certain regions of the southeast United States. In order to determine whether the residuals of the local model were random or clustered, the ESRI Spatial Autocorrelation tool was run again using the results of the GWR. A random distribution of the residuals was ideal in proving the validity of the analysis.

Results

High COPD crude rates with above 90 cases per 100,000 people were observed in Kentucky (Floyd, and Harlan counties), Missouri (Wayne and Hickory counties), and Arkansas (Woodruff and Fulton counties). Tampa, FL also had COPD crude rates above 90 (Figure 1). The thirteen risk factors tested to predict COPD crude rates yielded unique statistical results,

summarized in Table 1. The percentage of educators were well dispersed across the southeast United States. Counties with more than 9 percent of employed adults as educators were high in Missouri (Boone and Pulaski counties), Virginia (Abermarle and York counties), and Florida (Alachua and Lafayette counties) (Figure 2A). Healthcare workers were most numerous (greater than 9 percent) along the Mississippi River, in Missouri (Randolph and Madison counties), Mississippi (Rankin and Monroe counties), and Arkansas (Craighead and Saline counties) (Figure 2B). Farmers were not very numerous in the southeast United States (most of the area was less than 3 percent), but there were some counties with higher values namely in Georgia (Colquitt and Appling counties) Florida (Hardee county) and Arkansas (Lee county) where more than 9 percent of employed adults were farmers (Figure 2C). High levels of construction workers were more clustered, specifically in Louisiana (Cameron and De Soto counties), Arkansas (Stone and Montgomery counties) and Mississippi (Clark and Kemper counties) all had more than 9 percent of working populations as construction workers (Figure 2D).

Variable	Coefficient	P-Value	Robust_Pr	VIF
Intercept	-54.588846	Not Significant	0.000037	---
Average Particulate Matter (μg/m3)	-0.185737	Not Significant	0.407148	1.032230
Percent Smokers	2.235644	$p < 0.05$	0.000000	1.341076
Percent Healthcare Workers	0.714083	$p < 0.05$	0.004177	1.121923
Percent Farmers	0.501076	$p < 0.05$	0.027002	1.245767
Percent Construction Workers	0.154842	Not Significant	0.466612	1.197040
Percent Educators	-0.044576	Not Significant	0.856858	1.156721
Percent Households 3 or More	-0.176197	Not Significant	0.084837	2.458239
Percent Over 55	1.142801	$p < 0.05$	0.000000	3.423692
Percent Females	-0.108189	Not Significant	0.556365	1.586516
Percent White	0.918431	$p < 0.05$	0.000139	43.327382
Percent Black	0.053261	Not Significant	0.655329	39.880672
Percent Asian	-0.102691	Not Significant	0.241646	1.974076
Percent Physically Inactive	0.037783	Not Significant	0.385825	1.025974

Table 1. Raw data statistics for COPD Crude Rate and dependent variables tested in analysis.

The highest rates of households with three or more individuals were highest in northeast Virginia (Chesapeake, York, and Loudoun counties), Georgia (Coweta, Walton, Bartow, and Chattahoochee counties) and southern Louisiana (Cameron and Plaquemines counties) was also among the highest rates for above average household sizes (Figure 3A). Across the entire southeast United States, a large majority of the counties had between 50 percent and 52 percent females. Some counties did have higher levels of females, mainly in Mississippi (Attala, Holmes, and Smith counties) (Figure 3B). Average particulate matter (2.5 μm) was highest (greater

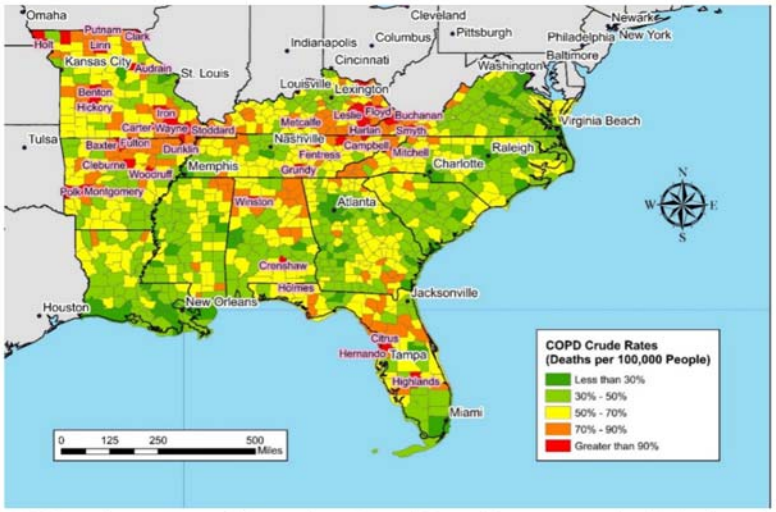

Figure 1. COPD crude rates in 2016 for southeast United States. Major cities are highlighted in white and counties within the highest class break are highlighted in pink.

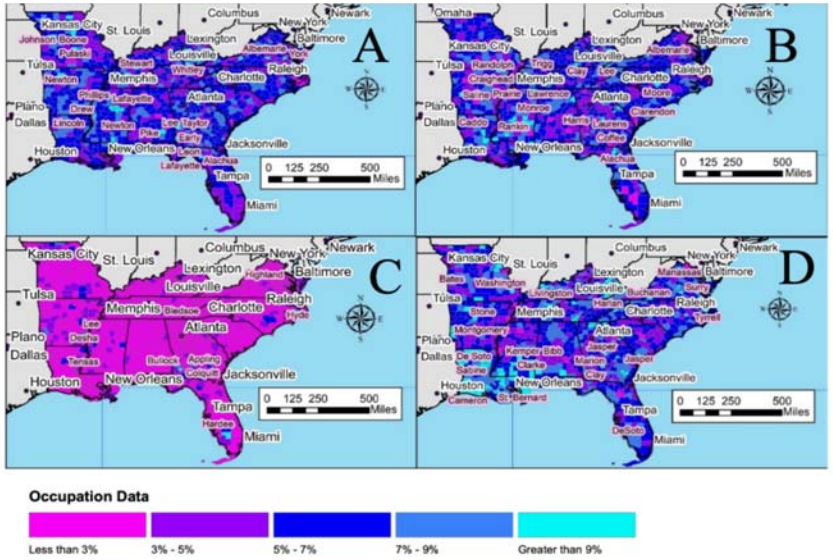

Figure 2. Distribution of occupations for educators (A), healthcare workers (B), farmers (C), and construction workers (D) in the southeast United States. Major cities are highlighted in white and counties within the highest class break are highlighted in pink.

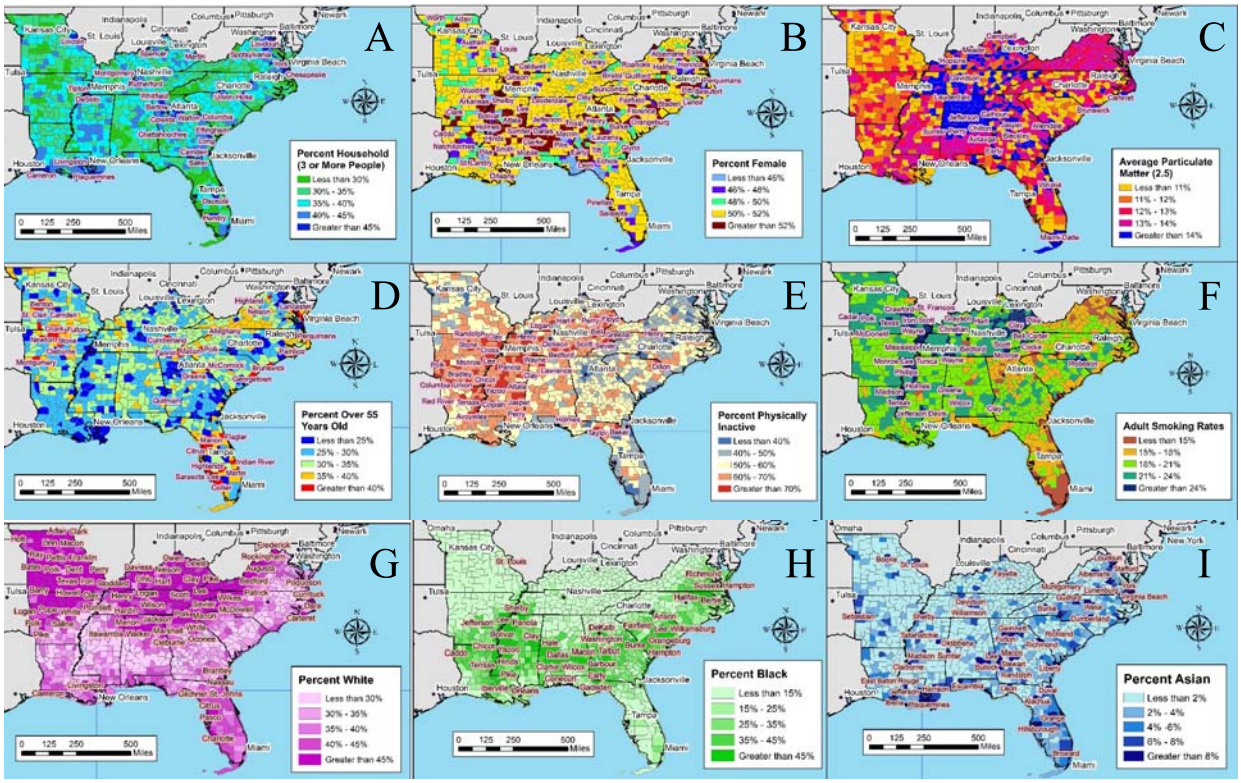

Figure 3. County data for percent of households with 3 or more individuals (A), percent females (B), average particulate matter (C), percent of people over 55 years (D), percent of people physically inactive (E), adult smoking rates (F), percent white (G), percent black (H), and percent Asian (I). Major cities are highlighted in white and counties within the highest class break are highlighted in pink.

than 14 percent) in northern Alabama (Jefferson and Perry counties) and Georgia (Calhoun and Jasper counties) (Figure 3C). The highest levels of individuals 55 years old and older were located in Florida (Sarasota and Marion counties), Missouri (Benton and St. Clair counties) and Arkansas (Ozark and Montgomery counties) (Figure 3D). The highest levels of physical inactivity in the southeast United States was clustered in Mississippi (Yazoo, Panola, and Attala counties), Louisiana (River, Tensas, and Union counties), and Arkansas (Polk, Monroe, and Stone counties) with more than 70 percent of individuals considered physically inactive (Figure 3E). The highest levels of adult smoking rates occurred in Kentucky (Clay, Hart, and Grayson counties) and along the Mississippi River in Louisiana (Madison county), Mississippi (Holmes county), Arkansas (Lee county) and Missouri (Wayne and Crawford counties) (Figure 3F).

The highest levels of white individuals were found in the northern region of the study area, namely Missouri (Linn, Iron, Franklin, and Perry counties), Kentucky (Stoddard, Hart, and Clay counties), Tennessee (Logan, Scott, and Wilson counties) and Virginia (Bedford and Rockingham counties). (Figure 3G) The lowest levels of white individuals were clustered in a band stretching from eastern Virginia, along the coastal Carolinas, into south-central Georgia, Alabama, and Mississippi with counties in this region showing populations where less than 30 percent of the individuals are white. The highest levels of black individuals occurred in nearly the same region of the lowest percentages of white individuals: North Carolina (Halifax and Anson counties), South Carolina (Lee and Orangeburg counties), Georgia (Dekalb and Talbot counties), Alabama (Dallas and Conecun counties) and Mississippi (Yazoo and Hinds counties (Figure 3H). The highest levels of Asian individuals were more scattered than the other two races. Populations with more than 8 percent Asian included Louisiana (Plaquemines and Iberia counties), Florida (Orange and Broward counties), and Georgia (Gwinnett and Fulton counties) (Figure 3I).

Clustering of high COPD crude rates were observed in eastern Kentucky, western Virginia, Orlando, Florida, Nashville, Tennessee, northern Arkansas, and northern/southeastern Missouri. (Figure 4) Areas where low COPD crude rates clustered include southern Louisiana, eastern Virginia, Miami, Florida, Atlanta, Georgia, Montgomery, Alabama and the coastal Carolinas (Figure 4).

The first trial of the global model explained 59.179 percent of observed COPD crude rates ($R^2 = 0.591790$), which may be indicative of key factors absent from the analysis (Table 2). This was further supported by the significance of the Jarque-Bera statistic, which indicated that the analysis was biased. However, the Variance Inflation Factors (VIFs) for percent whites and percent blacks were very high (Table 2), indicating redundancy between the two variables. This redundancy was likely attributed to a majority of the counties in the southeastern U.S. having similar percentages of white and black individuals, as observed in Figure 3G and Figure 3H. Percent blacks was removed from the second trial of the global model as there were a higher number of white individuals living in the study area. The second trial of the global model explained 59.1734 percent of observed COPD crude rates ($R^2 = 0.591734$), again indicating key factors may be absent (Table 3). The four most significant factors (given by the lowest robust probability values) were percent white, percent smokers, percent physically inactive and percent of people over 55. Percent smokers also had the highest coefficient in the global model, meaning percent of smokers was the most influential factor considered (Table 3).

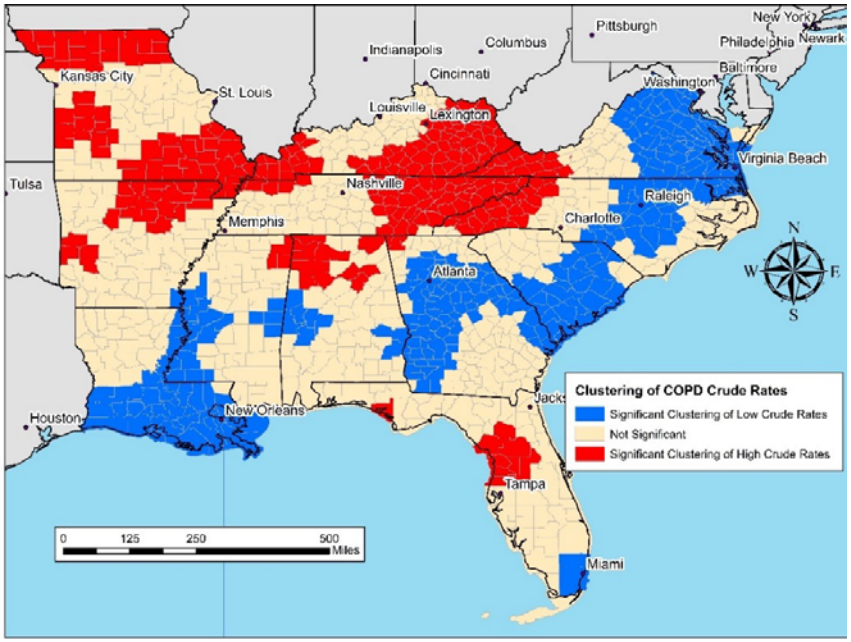

Figure 4. Identification of locations within the southeast where high and low COPD crude rates cluster significantly.

The global regression model determined percent smokers, percent healthcare workers, percent farmers, percent of individuals over 55 years old, and percent of white individuals were statistically significant in predicting COPD crude rates (Table 3). Average particulate matter (2.5μm), percent educators, percent of households with three or more individuals, percent females, and percent of Asian individuals had negative correlations to COPD crude rates (Table 3). Percent smokers, percent healthcare workers, percent farmers, percent construction workers, percent of individuals over 55 years old, percent white, and percent of people physically inactive were positively correlated (Table 3). All five significant variables had positive correlations with COPD crude rates. Interestingly, average particulate matter, a well-known risk factor of COPD in non-smokers was not identified as significant for the global model.

Results of the Spatial Autocorrelation yielded a Z-score of 19.3837 ($p < 0.05$). The high positive Z-score indicated that the distribution of the residuals from the global model were not a result of random chance; clustering of residuals of the global model was evident and further supports the need for additional independent variables to explain COPD crude rates.

The Geographically Weighted Regression was conducted using only the variables determined to be statistically significant in the global model. The local regression model explained 58.719 percent of observed COPD crude rates ($R^2 = 0.58719$), a slight increase

Intercept	-54.588846	Not Significant	0.000037	---
Average Particulate Matter (µg/m3)	-0.185737	Not Significant	0.407148	1.032230
Percent Smokers	2.235644	$p < 0.05$	0.000000	1.341076
Percent Healthcare Workers	0.714083	$p < 0.05$	0.004177	1.121923
Percent Farmers	0.501076	$p < 0.05$	0.027002	1.245767
Percent Construction Workers	0.154842	Not Significant	0.466612	1.197040
Percent Educators	-0.044576	Not Significant	0.856858	1.156721
Percent Households 3 or More	-0.176197	Not Significant	0.084837	2.458239
Percent Over 55	1.142801	$p < 0.05$	0.000000	3.423692
Percent Females	-0.108189	Not Significant	0.556365	1.586516
Percent White	0.918431	$p < 0.05$	0.000139	43.327382
Percent Black	0.053261	Not Significant	0.655329	39.880672
Percent Asian	-0.102691	Not Significant	0.241646	1.974076
Percent Physically Inactive	0.037783	Not Significant	0.385825	1.025974

R squared value = 0.576677

Table 2. Results from first trial of global multivariable regression analysis to explain observed COPD crude rates.

Variable	Coefficient	P-Value	Robust_Pr	VIF
Intercept	-52.323667	Not Significant	0.000013	---
Average Particulate Matter (µg/m3)	-0.173642	Not Significant	0.431911	1.016677
Percent Smokers	2.244506	$p < 0.05$	0.000000	1.311216
Percent Healthcare Workers	0.730434	$p < 0.05$	0.003062	1.092990
Percent Farmers	0.487461	$p < 0.05$	0.029833	1.220925
Percent Construction Workers	0.151874	Not Significant	0.000000	1.195379
Percent Educators	-0.042335	Not Significant	0.474297	1.156172
Percent Households 3 or More	-0.177144	Not Significant	0.864065	2.457263
Percent Over 55	1.150987	$p < 0.05$	0.000000	3.342736
Percent Females	-0.063669	Not Significant	0.694669	1.108897
Percent White	0.813894	$p < 0.05$	0.000000	1.280710
Percent Asian	-0.123118	Not Significant	0.094169	1.594279
Percent Physically Inactive	0.037136	Not Significant	0.393827	1.024749

R squared = 0.576604

Table 3. Results from second trial of global multivariable regression analysis to explain observed COPD crude rates.

Variable	Min. Coefficient	Max. Coefficient
Percent Farmers	0.336424	0.819818
Percent Healthcare Workers	0.504432	0.813643
Percent White	0.784995	0.892700
Percent Over 55	1.215704	1.496303
Percent Smokers	1.929614	2.634295

Table 4. Results from local multivariable regression analysis for observed COPD crude rates in the southeastern United States.

from the global model (Table 4). This improvement is likely attributed to the United States and the southeast region in particular having higher rates of adult smokers, particulate matter, and other factors influencing COPD prevalence compared to other areas of the world. The local model was produced to observe the correlation between percent of farmers, percent of healthcare workers, percent of white individuals, percent of people over 55 years old, and percent of adult smokers with COPD crude rate. The model showed percent farmers and percent healthcare workers had the weakest correlation with COPD crude rate in the study area as evidenced by the lower coefficients, while percent of individuals over 55 years old and percent smokers had the highest coefficients, indicating the strongest positive correlation (Table 4). For percentage of farmers, only Missouri and northern Arkansas exhibited a coefficient above 0.6 (Figure 5A). Percentage of healthcare workers was slightly more strongly correlated with COPD crude rate. The entire eastern half of the study area exhibited a coefficient between 0.6 and 0.8 (Figure 5B). Percent white exhibited a more positive coefficient between 0.8 and 1.0 in nearly the entire study area except northern Missouri, northern Kentucky, and northern Virginia, which all had coefficients between 0.6 and 0.8 (Figure 5C). Percent of individuals over 55 years old was more strongly correlated, with the entire study area having a coefficient between 1.2 and 1.4, except Missouri, Arkansas and western Louisiana, which had coefficients between 1.4 and 1.6 (Figure 5D). The model showed adult smoking rate had a very strong positive correlation in the entire study area, with every county exhibiting a coefficient above 1.6. (Figure 5E).

Discussion

Areas of clustering of low COPD crude rates occurred in cities including Atlanta, Miami, New Orleans, and Raleigh (Figure 4). These urban areas have easier access to hospitals and medical facilities than rural areas requiring less travel to receive proper treatments. These cities typically have hospitals that serve a larger number of people; more people may result in the hospitals more likely to conduct spirometry tests to diagnose patients with COPD and provide the proper treatments. Although the global model to explain COPD crude rates determined average particulate matter (2.5) was not statistically significant, particulate matter is considered to be the most likely risk factor of COPD in nonsmokers (Olloquequi et al. 2018). The insignificance of particulate matter in this study could be attributed to the fact that nonsmokers comprise a small population of COPD patients.

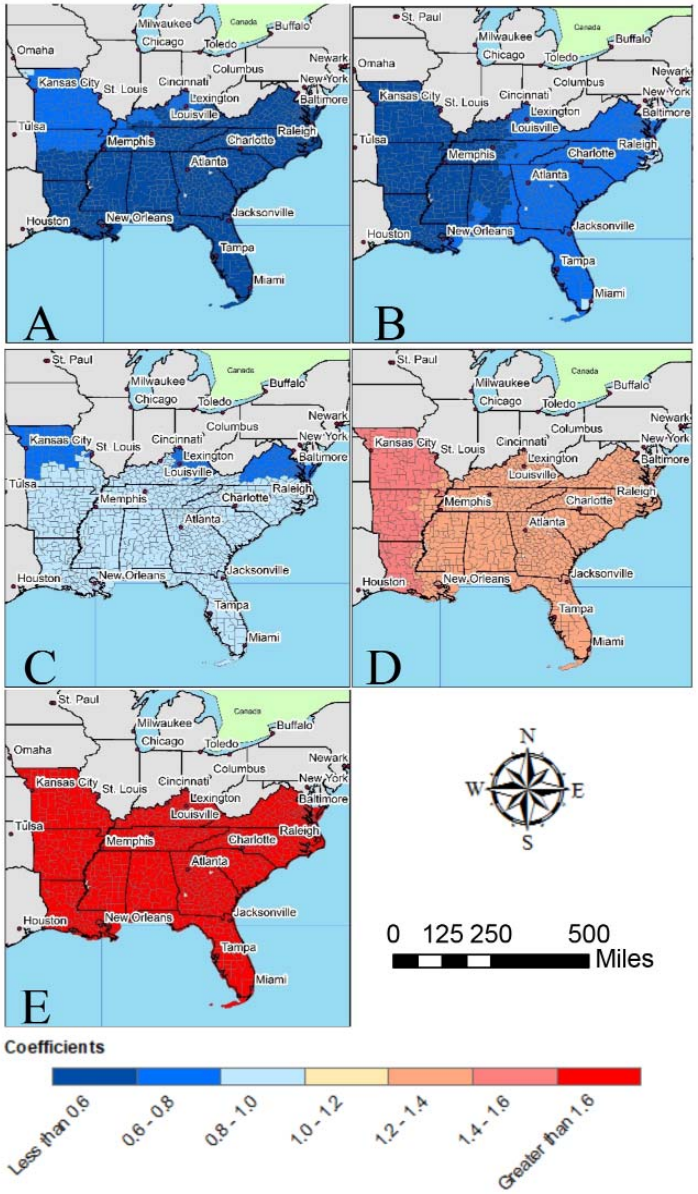

Figure 5. Results of the local regression analysis to show the correlation between COPD rates and farmers (A), healthcare workers (B), white individuals (C), individuals over 55 years of age (D), and adult smokers (E).

Another potential explanation for the insignificance of particulate matter is the wrong type of particulate matter was tested, specifically for farmers and construction workers. As mentioned, farmers and construction workers are exposed to particulates specific to their industry including asbestos, sulfides, construction dust, and other chemicals. COPD is still relatively high in these populations compared to other industries and percent of farmers was determined to be a significant factor in prediction of COPD prevalence in both the global model (Table 3) and the local model (Figure 5A), even though the correlation is the weakest among the five significant risk factors. A study in the United States concluded that the lifetime risk of COPD was twice as high among construction workers as among non-construction workers (Borup et al. 2017). A French study conducted by Guillien et al. (2015) concluded that in a study of 4,704 individuals and 3,787 farmers determined the population of farmers had a higher percentage of COPD prevalence than non-farmers. However, COPD prevalence depended on the farming activity and region. The same can be true for the United States and more data must be collected to develop a more definitive explanation for the relationship between farmers and construction workers with COPD prevalence.

Healthcare workers had a significant positive relationship with COPD crude rate in the global model (Table 3). In addition, there was a moderate positive correlation with COPD crude rate in the local model (Figure 5B). Nurses and doctors are among the most stressful professions, which can lead to increased smoking levels. Healthcare workers are also exposed to sick individuals on a daily basis. The likelihood of a healthcare worker catching an infection or disease is higher than other individuals. An infection can weaken the immune system and potentially increase the likelihood of that individual experiencing deteriorated lung function and becoming diagnosed with COPD. Instituting measures for medical professionals to take more frequent breaks or work shorter shifts could reduce their stress levels and hopefully reduce COPD prevalence. Better protection against the spread of infection should be given for healthcare workers including more personal protective equipment (PPE) and more frequent handwashing. More attention given to following organizational guidelines including knowledge, awareness and practice when treating patients for diseases is vital to protection of both the patient and the caretaker.

Percent whites had a significant positive relationship with COPD crude rate in the global model (Table 2) and had a moderately strong positive correlation with COPD crude rate in the local model for the southeastern United States (Figure 5D). This indicates race does play a factor in COPD prevalence, to an extent. It is somewhat surprising that white individuals were one of the more significant factors in predicting COPD crude rate given that black individuals in the United States have a higher prevalence and morbidity due to various biological, socioeconomic, and cultural factors (Ejike et al. 2019). This analysis tested only three races (with blacks being eliminated due to redundancy in the global model). An individual can be a variety of races or even multiracial. There is likely more substantial evidence to determine how race predicts COPD crude rates, but more research must be done to conclude this hypothesis.

In Missouri, Arkansas and Louisiana, individuals over 55 years of age were more strongly correlated with COPD crude rates than the rest of the study area (Figure 5C). Therefore, measures to ensure older individuals receive the proper diagnosis and treatments is necessary

to reduce COPD prevalence in this region and the southeast United States entirely. The most efficient way to diagnose COPD patients is using a spirometry test. Spirometry is a simple, noninvasive test that measures how strong an individual's lungs are and indicates the severity of airflow limitations (Skolnik et al. 2018). Underuse of spirometry is one of the largest contributors to underdiagnosis. Making spirometry tests mandatory during physical examinations for older individuals will help diagnosis those who may be affected with COPD and receive proper treatment. Another way older individuals can better protect themselves from developing COPD is by eating healthier and getting exercise, either self-enforced or with a caretaker. Taking medication properly helps as well because COPD is a progressive disease and can complicate with other diseases such as diabetes and asthma.

According to Figure 3F, adult smoking rates were highest in Kentucky, Missouri, and along the Mississippi River and lowest in Florida and the extreme southeast United States, however smoking rates in adults are high throughout most of the southeastern United States. The results of the local model identified similar trends in that COPD crude rates in the entire southeast United States correlated very strongly with adult smoking rates (Figure 5E). Smoking was determined to be the strongest risk factor influencing COPD crude rate in both the global model (Table 2) and local model (Figure 5F), as evidenced by the highest coefficients of all variables tested. Tobacco smoking is by far the most important risk factor in developing COPD (Liu et al. 2017). This is consistent with developing COPD; COPD weakens lung strength and leads to shortness of breath, which makes it harder to do everyday activities and contribute to physical inactivity (Liu et al. 2017). From the results of this analysis, medical professionals can determine which areas of the southeast United States where COPD is most prevalent, are affected from certain risk factors. In order to reduce the levels of smoking, certain measures should be taken to increase the tax on cigarettes to discourage people from buying them.

Future Experiments/Error

Future experiments should consider different types of air pollutants significant in COPD development considering particulate matter (2.5) was not significant on the global or local scale. Adjusting the age scale from 55 years and over to a smaller range could result in more significant results on the local scale. This is evident by the average prevalence rate for COPD being for people 75 years and older, the highest prevalence for any age group (Wheaton et al., 2015). Households with more than 3 individuals were grouped together in this study as a potential risk factor. A future experiment could yield promising results as a previous spatial and statistical analysis conducted by Kauhl et al. (2018), determined household size was a significant risk factor for a study in northeastern Germany. Studying other occupations could lead to future experiments and possibly determine which occupations are most at risk for COPD development, specifically in the southeastern United States where COPD is a large problem for the population. Testing other types of races and multiracial individuals would also yield more complete results in the relationship between race and COPD prevalence. A study analyzing how individuals already diagnosed with COPD cope with the disease as well as studying if these individuals develop more severe health problems from their diagnosis would help better understand the long-term effects of COPD.

Limitations

Data for this study was collected from several different years., due to the availability of the data. County data was not available for the same year of each risk factor, therefore data from years as close to the year of COPD crude rate was used for this analysis. While this does slightly alter the statistics for the risk factors, the discrepancy was kept as minimal as possible to conduct a study with valid results. County data was used in this analysis rather than individual or city data because county data was easily obtained for a large region like the southeastern United States. Individual data would have been more difficult to collect and analyze given the large population tested and major changes that could occur on a case-by-case basis. While this does essentially negate the effect of genetics, the third major predictor in COPD prevalence, a separate study focusing on how genetics predict COPD prevalence would yield better results. In addition, genetics would have been difficult to manipulate in a quantitative statistic to use in an OLS or GWR.

This study only provides a snapshot of COPD and its risk factors in the southeastern United States for one specific year. In order to best determine the risk factors and effects of COPD in any region, multiple years must be considered to assess the risk factors overtime, as well as follow how the health of an individual changes throughout their lifetime. People relocate constantly and therefore affect the demographic data of the places they live, distorting the data. For this reason, a study focusing on individual or household cases is imperative to assessing COPD prevalence.

Conclusions

This study is one of the first to test the spatial relationship between occupation and COPD crude rate in the United States. The selection of the occupation tested was determined by their relationship with the well-known causes of COPD (smoking and atmospheric particulate matter). While this study yielded important results, such as further validating smoking and old age contributes to COPD prevalence and indicating occupation does indeed help predict COPD prevalence, more research needs to be done to determine how particulate matter and other factors contribute to COPD prevalence. The methodology of this study proved useful in determining which risk factors helped predict COPD in the southeastern United States, which is important in understanding this disease, which will only continue to increase in cases unless certain measures are taken to prevent the deterioration of lung function. This study provides a useful starting point for assessing the effects of many risk factors to predict COPD in one of the most severely affected regions of the United States.

Erick Bora received his Bachelor of Science from Eastern Connecticut State University, Willimantic, CT. He is currently a Master's Degree Student at the Colorado School of Mines, Golden, Colorado. Email: ebora@mines.edu. His research interests include Proterozoic structural geology, geomorphology, and geographic information systems.

References

American Lung Association 2019. *COPD causes and risk factors*. (https://www.lung.org/lung-health-diseases/lung-disease-lookup/copd/what-causes-copd (last accessed 29 June 2020).

Appalachian Voices 2017. *Mountaintop Removal 101*. http://appvoices.org/end-mountaintop-removal/mtr101/ (last accessed 21 August 2020).

Aschan-Leygonie, C., S. Baudet-Michel, M. Helene, & L. Sanders. 2013. Gaining a better understanding of respiratory health inequalities among cities: An ecological case study on elderly males in the larger French cities. *International Journal of Health Geographics* 12 (19):1-15. doi:10.1186/1476-072X-12-19

Boeselt, T., M. Spielmanns, C. Nell, J. H. Storre, W. Windisch, L. Magerhans, B. Beutel, K. Kenn, T. Greulich, P. Alter, C. Vogelmeier, A. R. Koczulla. 2016. Validity and usability of physical activity monitoring in patients with chronic obstructive pulmonary disease (COPD). *PLOS One* 11 (6):1-11. doi:10.1371/journal.pone.0157229

Borup, H., L. Kirkeskov, D. J. Hanskov, and C. Brauer. 2017. Systematic review: Chronic obstructive pulmonary disease and construction workers. *Occupational Medicine* 67 (1):199-204. doi:10.1093/occmed/kqx007

Centers for Disease Control and Prevention (CDC). 2016. *CDC Wonder*. https://wonder.cdc.gov/controller/datarequest/D140 (last accessed 23 July 2020).

Chatila, W. M., W. A. Wynkoop, G. Vance, and G. J. Criner. 2004. Smoking patterns in African Americans and whites with advanced COPD. *Chest* 125 (1):15–21.

Data U.S.A. 2017. *Adult Smoking by County*. https://datausa.io/map?enlarged=c-geomap-Z1X72Pg&groups=0-Z1X72Pg&measure=Z1NeNDq. (last accessed 8 July 2020).

Ejike, C. O., M. T. Dransfield, N. N. Hansel, N. Putcha, S. Raju, C. H. Martinez, and M. K. Han. 2019. Chronic obstructive pulmonary disease in America's black population. *American Journal of Respiratory and Critical Care Medicine* 200 (4):423-430.

Elliot, L., and S. Von Essen. 2016. COPD in farmers: What have we learnt. *European Respiratory Journal* 47 (1):16-18. doi:10.1183/13993003.01768-2015

Familoni, O. 2008. An Overview of Stress in Medical Practice. *African Health Sciences* 8 (1):6-7.

Galvin, G. 2019. *In "Tobacco Nation", Smoking Stubbornly Persists*. https://www.usnews.com/news/healthiest-communities/articles/2019-06-06/smoking-stubbornly-persists-in-tobacco-nation-states (last accessed 14 August 2020).

Gaunt A. 2020. *COPD in Seniors*.https://www.aplaceformom.com/planning-and-advice/articles/seniors-with-copd (last accessed 26 July 2022).

Guillien, A., M. Puyraveau, T. Soumagne, S. Guillot, F. Rannou, D. Marquette, P. Berger, S. Jouneau, E. Monnet, F. Mauny, et al. 2016. Prevalence and risk factors for COPD in farmers: A cross-sectional controlled study. *European Respiratory Journal* 47 (1):95-103.

Hankinson, J. L., J. R. Odencrantz, and B. K. Fedan. 1999. Spirometric reference values from a sample of the general U.S. population. *American Journal of Respiratory and Critical Care Medicine* 159 (1):179-187. doi:10.1164/ajrccm.159.1.9712108

Hart, J. E., S. T. Grady, F. Laden, B. A. Coull, P. Koutrakis, J. D. Schwartz, M. L. Moy, and E. Garshick. 2018. Effects of indoor and ambient black carbon and PM2.5 on pulmonary function among individuals with COPD. *Environmental Health Perspectives* 126 (12).

Kauhl, B., W. Maier, J. Schweikart, A. Keste, and M. Moskwyn. 2018. Who is where at risk for chronic obstructive pulmonary disease? A spatial epidemiological analysis of health insurance claims for COPD in northeastern Germany. *PLOS One* 13 (2):1-18. doi:10.1371/journal.pone.0190865

Liu, H., N. Wang, W. Chen, W. Liu, S. Wang, J. Lei, and H. Chen. 2018. Hospitalization trends in adult patients with COPD and other respiratory diseases in northeast China from 2005-2015. *BioMed Research International* 1-7. doi:10.1155/2018/1060497

Olloquequi, J., S. Jaime, V. Parra, E. Cornejo-Cordova, G. Valdivia, A. Agusti, and R. O. Silva. 2018. Comparative analysis of COPD associated with tobacco smoking, biomass smoke exposure or both. *BioMed Central* 19 (13):1-8.

Skaalvik, E. M., and S. Skaalvik. 2015. Job satisfaction, stress and coping strategies in the teaching profession - What do teachers say? *International Education Studies* 8 (3):181-192. doi:10.5539/ies.v8n3p181

Skolnik, N. M., A. G. Kaplan, T. M. Lake, M. L. Hayden, & N. A. Hanania. 2018. COPD management in the primary care setting. *Journal of Family Practice* 67 (10 Suppl):S27-S32

Statista Research Department. *Average number of people per household in the United States from 1960 to 2021*. 2019. https://www.statista.com/statistics/183648/average-size-of-households-in-the-us/ (last accessed 8 July 2020).

U.S. Census Bureau. 2016. *2018 Tiger Line/Shapefiles*. https://www.census.gov/geo/maps-data/data/tiger-data.html (last accessed 24 August 2020).

_____. 2018. *Explore census data*. https://data.census.gov/cedsci/advanced (last accessed 12 July 2020).

U.S. Environmental Protection Agency. 2018. *Outdoor Air Quality Data*. https://www.epa.gov/outdoor-air-quality-data/download-daily-data (last accessed 4 August 2020).

Wheaton, A. G., T. J. Cunningham, E. Ford, and J. B. Croft. 2015. Employment and activity limitations among adults with chronic obstructive pulmonary disease - United States, 2013. *Morbidity and Mortality Weekly Report* 64 (11):289-295.

World Health Organization (WHO). 2018. *Ambient (Outdoor)air pollution*. Accessed August 3, 2020. who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health (last accessed 3 August 3 2020).

_____. *Burden of COPD*. https://www.who.int/respiratory/copd/burden/en/ (last accessed 29 July 2020).